

第四章 研究設計

在開始進行實證研究前，本章將先詳細說明本研究使用之研究設計。首先，第一節中將闡明使用追蹤資料進行研究的原因，並比較其下的固定效果模型和隨機效果模型（random-effect model）之異同和取捨方法。²⁵接著，於第二節建立本研究的實證模型，並說明實證資料之來源。至於本研究實證模型之實證變數假設，將陳述於第三節中。最後，於第四節進行本章內容之小結。

第一節 研究方法說明

如前所述，本研究的目的是在於瞭解台灣各縣市的社會福利支出，對人口遷移是否存在著影響效果；倘若遷移決策的確會受到社會福利支出的影響，又此具有影響力的變數對男性或女性遷移者的影響是否一致？於是本研究針對台灣自1995年至2005年，23個縣市的追蹤資料，利用雙因子固定效果模型加以分析，藉此得到最佳的估計。

所謂追蹤資料是指針對某一特定調查對象組群，鎖定這些組群持續一段時間所得到的各種資料。由於是時間序列資料（time-series data）和橫斷面資料（cross-section data）的合併使用，因此不但擁有時間序列的動態性質，並能兼顧橫斷面資料可以表達不同樣本間特性的優點。和橫斷面資料相較之下，不但得觀察到變數間原來無法觀察到的動態變化，甚者可在固定效果模型中以虛擬變數（dummy variable）的形式代表一些難以衡量的質化變數（qualitative data），以分析觀察個體的固定效果，藉此降低參數估計上的偏誤。此外，亦能避免時間序列資料的樣本數過少及自由度過小的問題，以及減少共線性的產生。因此，若欲分析某觀察群體長期性的決

²⁵本文研究方法主要參閱 Greene（2003）及黃台心（2005）。

定因素，使用追蹤資料較能得出正確且嚴謹的結果，而這正符合本研究的要求。以下將更深入地探討追蹤資料下的固定效果模型和隨機效果模型。

一、固定效果模型

假設一追蹤資料有N個橫斷面單位以及T個時期，因此總觀察變數M=NT。以一般傳統迴歸式表現如下：

$$Y_{it} = X_{it}\beta + \varepsilon_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (4.1)$$

其中 X_{it} 代表解釋變數，是 $1 \times k$ 的向量。 ε 為隨機殘差項，且 $E(\varepsilon_{it}) = 0$ ， $E(\varepsilon_{it}^2) = \delta \varepsilon^2$ 。如果每個誤差項加總為零，可使用一般最小平方法（OLS）來估計。如果干擾因子會對所有年度的個別橫斷面資料造成影響，各年度間的誤差項 ε_{it} 和 ε_{is} ($t \neq s$) 將具有高度相關。相對的，如果干擾因子會影響某年度的所有橫斷面資料，則各地區間的 ε_{it} 和 ε_{jt} ($i \neq j$) 將具有高度相關。若使用傳統的計量方法，例如OLS，來估計追蹤資料很有可能會造成相當大的偏誤。

在固定效果模型中，認為不同觀察單位間的差異，可被不同個別效果（individual effect）解釋，視每個截距項為待估計未知常數，故也稱為共變異數模型（Covariance Model）。固定效果模型的迴歸式可表示為：

$$Y_{it} = \phi_i + \sum_{k=1}^K \beta_k X_{kit} + \varepsilon_{it}, \quad i=2, \dots, N, t=2, \dots, T \quad (4.2)$$

式中下標 i 代表第 i 個觀察單位， t 代表第 t 期時間，至於 k 則代表第 k 個外生變數（exogenous variables）。 Y_{it} 為應變數， X_{it} 為對應之自變數向量， β_k 為斜率項變數； ϕ_i 為截距項變數，即個別效果。可將第(4.2)式改寫成：

$$Y_i = X_i\beta + J\phi_i + \varepsilon_i \quad (4.3)$$

令 Y_i 與 X_i 分別為 $T \times 1$ 和 $T \times (K-1)$ 向量和矩陣，是第 i 觀察單位的應變數與自變數樣本觀察值， ε_i 為對應之隨機干擾項向量， $J_T = (1 \ 1 \ \dots \ 1)'$ 為 $T \times 1$ 向量，其中每個元素皆為一。式中：

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix} \quad X_i = \begin{bmatrix} X_{1i1} & X_{2i1} & \dots & X_{Ki1} \\ X_{1i2} & X_{2i2} & \dots & X_{Ki2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1iT} & X_{2iT} & \dots & X_{KiT} \end{bmatrix} \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \quad (4.4)$$

將全部 $N \times T$ 樣本堆疊起來成為

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \beta + \begin{bmatrix} J & 0 & \dots & 0 \\ 0 & J & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (4.5)$$

或

$$Y = [d_1 \ d_2 \ \dots \ d_N \ X] \begin{bmatrix} \phi \\ \beta \end{bmatrix} + \varepsilon \quad (4.6)$$

式中 d_i 對應 i 單位的虛擬變數行向量，即第(4.5)式等號右邊第一個矩陣的第 i 行，令 $D = [d_1 \ d_2 \ \dots \ d_N]$ 代表 $NT \times N$ 矩陣，並進一步將第(4.6)式簡化為：

$$Y = X\beta + D\phi + \varepsilon \quad (4.7)$$

此模型通常被稱作最小平方虛擬變數模型 (least squares dummy variable model)，簡稱 LSDV 模型。LSDV 模型屬於古典迴歸模型。若觀察單位數 N 不大，第(4.7)式直接可以最小平方法估計；相對的，如果觀察單位成千上萬，應設法先消除固定效果，則可容易估計出係數向量 β ，進

而再估計固定效果參數。一般而言，多以組內估計法（within group estimation）或組間估計法（between group estimation）進行。而所謂的「一組」即為一觀察單位，可能是一個人、家戶、廠商、產業或國家。

（一）、組內與組間估計法

考慮一複迴歸方程式：

$$Y_{it} = \sum_{k=1}^K \beta_k X_{kit} + \phi_i + \varepsilon_{it} \quad (4.8)$$

定義各組平均值如下：

$$\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it} = \phi_i + \sum_{k=1}^K \beta_k \bar{X}_{ki} + \bar{\varepsilon}_i \quad (4.9)$$

將(4.8)、(4.9)兩式相減，得到：

$$Y_{it} - \bar{Y}_i = \sum_{k=1}^K \beta_k (X_{kit} - \bar{X}_{ki}) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (4.10)$$

以全體母數的平均數（ \bar{Y} 和 \bar{X} ）為基準，計算總動差矩陣：

$$S_{xx}^{total} = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \bar{X})(X_{it} - \bar{X})' \quad S_{xy}^{total} = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \bar{X})(Y_{it} - \bar{Y})' \quad (4.11)$$

其中組內動差矩陣的定義為：

$$S_{xx}^{within} = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i)' \quad S_{xy}^{within} = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \bar{X}_i)(Y_{it} - \bar{Y}_i)' \quad (4.12)$$

定義組間動差矩陣：

$$S_{xx}^{between} = \sum_{i=1}^N T(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' \quad S_{xy}^{between} = \sum_{i=1}^N T(\bar{X}_i - \bar{X})(\bar{Y}_i - \bar{Y})' \quad (4.13)$$

可得：

$$S_{xx}^{total} = S_{xx}^{within} + S_{xx}^{between} \quad S_{xy}^{total} = S_{xy}^{within} + S_{xy}^{between} \quad (4.14)$$

因此，經由最小平方法可估計出三個係數向量 β ，分別為全體、組內以及組間估計式：

$$\hat{\beta}^{total} = [S_{xx}^{total}]^{-1} S_{xy}^{total} = [S_{xx}^{within} + S_{xx}^{between}]^{-1} [S_{xy}^{within} + S_{xy}^{between}] \quad (4.15)$$

$$\hat{\beta}^{within} = [S_{xx}^{within}]^{-1} S_{xy}^{within} \quad (4.16)$$

$$\hat{\beta}^{between} = [S_{xx}^{between}]^{-1} S_{xy}^{between} \quad (4.17)$$

將組內及組間估計係數的加權平均，可得固定效果參數的估計式：

$$\hat{\beta}^{total} = F^{within} \hat{\beta}^{within} + F^{between} \hat{\beta}^{between} \quad (4.18)$$

$$F^{within} = [S_{xx}^{within} + S_{xx}^{between}]^{-1} S_{xx}^{within} = I - F^{between} \quad (4.19)$$

(二)、時間效果

LSDV模型稍加擴充，即可納入時間效果，第(4.2)式可改為：

$$Y_{it} = \sum_{k=1}^K \beta_k X_{kit} + \phi_i + \varphi_t + \varepsilon_{it} \quad (4.20)$$

式中 φ_t 代表第t期時間虛擬變數的係數，類似的係數本應有T個，唯其中任意一個必須標準化為零，以避免線性重合問題。另一種模型設定方法為：

$$Y_{it} = \sum_{k=1}^K \beta_k X_{kit} + \beta_0 + \phi_i + \varphi_t + \varepsilon_{it} \quad (4.21)$$

式中 β_0 是截距項，為避免線性重合的問題，故需加入限制條件

$\sum_i \phi = \sum_t \varphi = 0$ 。定義：

$$Y_{it}^* = Y_{it} - \bar{Y}_i - \bar{Y}_t + \bar{Y}, \quad X_{it}^* = X_{it} - \bar{X}_i - \bar{X}_t + \bar{X} \quad (4.22)$$

其中，時間固定效果與全體樣本平均數為：

$$\bar{Y}_t = \frac{1}{N} \sum_{i=1}^N Y_{it} \quad \bar{Y} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Y_{it} \quad \bar{X}_t = \frac{1}{N} \sum_{i=1}^N X_{it} \quad \bar{X} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it}$$

利用第(4.21)式可證明下式成立：

$$Y_{it}^* = \sum_{k=1}^K \beta_k X_{kit}^* + \varepsilon_{it} \quad (4.23)$$

欲估計斜率項係數 β ，運用最小平方法估計第(4.23)式；至於截距項及其他虛擬變數之係數，需使用下列方法得到：

$$\hat{\beta}_0 = \bar{Y} - \bar{X}'\hat{\beta} \quad (4.24)$$

$$\hat{\phi}_i = (\bar{Y}_i - \bar{Y}) - \sum_{k=1}^K \hat{\beta}_k (\bar{X}_i - \bar{X}) \quad i=1\dots N \quad (4.25)$$

$$\hat{\phi}_t = (\bar{Y}_t - \bar{Y}) - \sum_{k=1}^K \hat{\beta}_k (\bar{X}_t - \bar{X}) \quad t=1\dots T \quad (4.26)$$

(三)、固定效果的係數檢定

最後，關於模型是否存在固定效果之問題，可用F統計量來檢定。

令 虛無假設 $H_0: \phi_1 = \phi_2 \dots = \phi_N$

對立假設 H_1 ：各項不全相等

檢定結果若接受 H_0 ，只需估計一個截距項，意味此縱橫資料的N個觀察單位，每單位T期資料，可被視作有NT個觀察值的橫斷面或時間數列樣本，因而喪失追蹤資料的特性，此時使用共同迴歸模型較適當。相對的，當F統計量大於其對應的臨界值時將拒絕虛無假設，表示有固定效果的存在，則採用固定效果模型較佳。檢定統計量定義為：

$$F(N-1, NT-N-K+1) = \frac{(SSE_R - SSE_U)/(N-1)}{SSE_U/(NT-N-K+1)} \quad (4.27)$$

第(4.27)式中 SSE_R 來自於受限制的殘差平方和，即假設第(4.2)式中 $\beta_{it} = \beta_1$ 後，估計該是所得到之殘差平方和，也就是共同迴歸模型的殘差平方和； SSE_U 來自於不受限制模型的殘差平方和，即直接估計第(4.2)式中得到的殘差平方和； $(N-1)$ 代表虛無假設裡限制條件的個數； $(NT-N-K+1)$ 是不受限制模型的自由度。在虛無假設為真的情形下，假定統計量 F 是自由度 $(N-1)$ 和 $(NT-N-K+1)$ 的 F 分配隨機變數。

最後，值得注意的是，固定效果模型的使用將產生一些附帶問題，例如自由度的消耗以及並未明確指出造成迴歸估計式平移之原因。

二、隨機效果模型 (Random Effects Model)

隨機效果模型又稱為誤差成分模型 (error component model)。之前由於對模型的認知不足，懷疑有遺漏變數，因此在模型中加入虛擬變數而成為固定效果模型。但當同樣從干擾項著手，卻假設未觀察到的個體效果和時間效果與其他自變數無關時，不同觀察單位擁有不同的特定隨機變數，此即為隨機效果模型。其迴歸式可表示為：

$$Y_{it} = \sum_{k=1}^K \beta_k X_{kit} + u_i + \varepsilon_{it} \quad (4.28)$$

等號右邊第一項 β_1 為截距項； u_i 是隨機變數，屬第 i 觀察單位特有，不隨時間而改變。此模型通常有如下假設：

$$E(u_i) = E(\varepsilon_{it}) = 0 \quad , \quad E(\varepsilon_{it}^2) = \sigma_\varepsilon^2 \quad , \quad E(u_i^2) = \sigma_u^2 \quad (4.29)$$

$$E(u_i \varepsilon_{jt}) = 0 \quad \text{對所有 } i, t \text{ 和 } j$$

$$E(\varepsilon_{it} \varepsilon_{js}) = 0 \quad \text{若 } t \neq s \text{ 或 } i \neq j$$

$$E(u_i u_{js}) = 0 \quad \text{若 } i \neq j$$

與固定效果模型相似的是， μ_i 以及 ε_{it} 分別代表了個體效果和時間效果。至於其相異處，則隨機效果模型假設 μ_i 和 ε_{it} 為隨機變數，可視為 N 個個體和 T 個期間是由一龐大母體所抽出的隨機樣本。假設只考慮個體效果，於是對於 i 個體可將模型改寫為：

$$y_i = X_i\beta + Ju_i + \varepsilon_i \quad (4.30)$$

根據上述假設，組合誤差向量 $Ju_i + \varepsilon_i$ 的平均數等於零向量，共變數矩陣為：

$$\begin{aligned} \Omega &= E(Ju_i + \varepsilon_i)(Ju_i + \varepsilon_i)' \\ &= E(u_i^2)JJ' + JE(\gamma_i\varepsilon_i') + E(\varepsilon_i\gamma_i)J' + E(\varepsilon_i\varepsilon_i') \\ &= \sigma_\gamma^2 JJ' + \sigma_\varepsilon^2 I_T \\ &= \begin{bmatrix} \sigma_\gamma^2 + \sigma_\varepsilon^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\gamma^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\gamma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 + \sigma_\varepsilon^2 \end{bmatrix} \end{aligned} \quad (4.31)$$

其中， I_T 為 $T \times 1$ 之向量，內容全為1。而由於任意 i 與 j 觀察單位彼此相互獨立，因此，所有 NT 個樣本誤差項共變數矩陣 V 成為區塊對角矩陣 (block diagonal matrix)

$$\begin{aligned} V &= E(\gamma \otimes J + \varepsilon)(\gamma \otimes J + \varepsilon)' \\ &= \sigma_\gamma^2 I_N \otimes JJ' + \sigma_\varepsilon^2 I_{NT} \\ &= I_N \otimes (\sigma_\gamma^2 JJ' + \sigma_\varepsilon^2 I_T) \\ &= I_N \otimes \Omega \\ &= \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega \end{bmatrix} \end{aligned} \quad (4.32)$$

式中符號 \otimes 代表寇雷克(Kronecker product)，經由第(4.31)式可以得到：

$$\Omega^{-1/2} = I_T - \frac{\theta}{T} JJ' \quad \theta = 1 - \frac{\sigma_\varepsilon}{\sigma} \quad \sigma^2 = T\sigma_\gamma^2 + \sigma_\varepsilon^2 \quad (4.33)$$

如果 σ_γ^2 與 σ_ε^2 為已知，應採用一般化最小平法（Generalized least squares，簡稱GLS）進行估計。而實際上 Ω 往往是無法事先得知，需要先對 σ_γ^2 和 σ_ε^2 進行估計，在針對模型的參數進行估計與檢定。這樣的過程便是可行的一般化最小平方法（Feasible Generalized Least Squares，FGLS）。

至於隨機效果模型的檢定，除了可利用F統計量檢定外，另可採LM統計量（Lagrange multiplier test）檢定方法。²⁶其檢定方式如下：

令虛無假設 $H_0 : \sigma_\mu^2 = 0$

對立假設 $H_1 : \sigma_\mu^2 \neq 0$

$$\lambda = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left[\sum_{t=1}^T e_{it} \right]^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right]^2 = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N (T\bar{e}_i)^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \quad (4.34)$$

其中， e_{it} 為 OLS 模型的殘差值。當 LM 統計量大於其對應的臨界值時將拒絕虛無假設，表示有隨機效果的存在。

三、Hausman 檢定

前述之固定效果模型和隨機效果模型的選擇標準，目前為止看法迥異，最簡單的方法是以「樣本有無透過抽樣過程」來分辨，若樣本為母體或無透過抽樣過程，則採用固定效果模型較佳。至於計量方法上，則通常透過 Hausman（1978）的方法檢定。由於固定效果模型和隨機效果模型最大的差異，在於隨機效果模型中的 u_i 與解釋變數間是否具有相關性。若有，則固定效果模型之估計將具一致性和有效性，而隨機效果模型之估計將不具一致性，故應採用固定效果模型；反之，隨機效果模型之估計將具

²⁶ LM 統計量檢定方法可參閱 Breusch and Pagan（1980）。

一致性和有效性，而固定效果模型之估計式將具一致性但不具有有效性，故應採用隨機效果模型。其檢定方式如下：

令虛無假設 $H_0: E(u_i X_{it})=0$ ，即 u_i 與解釋變數間不具相關性；

對立假設 $H_A: E(u_i X_{it})\neq 0$ ，即 u_i 與解釋變數間具相關性。

Hausman 的檢定統計量可表示如下：

$$H = (b_{fix} - b_{ran})'(M_{fix} - M_{ran})^{-1}(b_{fix} - b_{ran}) \sim X^2(K) \quad (4.35)$$

第(4.35)式中， H 統計量呈 χ^2 分配， K 為其自由度。而 b_{fix} 與 b_{ran} 分為固定係數與隨機係數的參數估計值， M_{fix} 與 M_{ran} 分別為其共變異矩陣。當 $H < X^2(K)$ 時無法拒絕虛無假設，即兩模型之估計無差異，或 u_i 與解釋變數間不具相關性，則選擇隨機效果模型；反之，當 $H > X^2(K)$ 時拒絕虛無假設，即兩模型之估計有差異，或 u_i 與解釋變數間具相關性，則應選擇固定效果模型為佳。

第二節 實證模型設定與資料來源

一、本研究之實證模型設定

如前所述，本文使用結合橫斷面與時間序列的追蹤資料，其優點是相較於橫斷面與時間序列兩種資料，追蹤資料包含更多的資訊。如此，不但有助於提高樣本數與自由度，使估計結果較為準確；而實證模型的採用也將更具有更多的選擇。而在回顧眾多探討人口遷移決定因素的文獻後，歸納出社會福利支出、可支配所得、地方公共支出、產業結構、失業率、重大犯罪率、教育服務，以及選舉年等，都可能是影響人口遷移決策的重要決定因素。由於本文探討社會福利支出對人口遷移之影響、以及對男性遷移和女性遷移的影響，故將分別進行三個迴歸式的估計。在綜合以上各重要變數假設的探討後，可設出下列的關係式：

$$MR = f(SW, I, GE, CRI, EDU, INSER, UNEM, ELE) \quad (4.36)$$

第(4.36)式的意義為，人口淨遷移率 (MR) 為每人可享有社會福利支出決算審定數 (SW)、每戶可支配所得 (I)、每人可享有除了社會福利支出以外的公共支出服務 (GE)、犯罪率 (CRI)、教育水準 (EDU)、工業及服務業比例 ($INSER$)、失業率 ($UNEM$)、以及選舉年 (ELE) 的函數。第(4.36)式可以進一步表示為：

$$MR_{it} = \beta_{0i} + \psi_t + \beta_1 \log(SW)_{i,(t-1)} + \beta_2 \log(I)_{i,(t-1)} + \beta_3 GE_{i,(t-1)} + \beta_4 CRI_{i,(t-1)} + \beta_5 EDU_{i,(t-1)} + \beta_6 INSER_{i,(t-1)} + \beta_7 UNEM_{i,(t-1)} + \beta_8 ELE_{i,t} + \varepsilon_{i,t} \quad (4.37)$$

在第(4.37)式中，表示取自然對數 \log 後的值， i 表示縣市， t 代表時間，至於 ε 則為殘差項。因本研究欲進一步探討社會福利支出對人口遷移之影響、以及對男性遷移和女性遷移的影響，被解釋變數 $MR_{i,t}$ 可分別以 $NMR_{i,t}$ 、 $MNMR_{i,t}$ 和 $FNMR_{i,t}$ 表示。 $NMR_{i,t}$ 代表第 i 個縣市在第 t 年的總人口淨遷移

率， $MNMR_{i,t}$ 代表第 i 個縣市在第 t 年的男性人口淨遷移率， $FNMR_{i,t}$ 代表第 i 個縣市在第 t 年的女性人口淨遷移率。此外，為避免社會福利支出與人口遷移、以及其他解釋變數間潛藏的內生性問題 (endogeneity)，除了選舉年虛擬變數 (ELE) 外，其餘解釋變數皆以前一期變數取代。²⁷

此外，由於各縣市可能存在個別特質，使各遷移者在進行遷移決策時產生不同的區位選擇。為了衡量這些個別特質的型態，因此本研究在估計的實證模型中，加入區域特質的個別效果 (region-specific effect) β_{0i} 。另外，為了捕捉時間趨勢的效果，本研究也考慮隨時間經過，產生對被解釋變數影響的時間效果 ψ_t 。因此，第(4.37)式的實證估計模型為雙因子固定效果模型。若假設這些個別特質效果，是以隨機變動形式來影響各遷移者在各縣市的區位選擇。 $\beta_{0i} = \bar{\beta}_0 + u_i + \lambda_t$ ， $u_i \sim \text{iid}(0, \sigma_u^2)$ 且 $\lambda_t \sim \text{iid}(0, \sigma_\lambda^2)$ ，則實證估計模型將改為隨機效果模型。至於使用哪種模型來估計較為適當，本文將以前述 Hausman (1978) 的方法檢定。

二、本研究之資料來源

至於實證資料來源方面，由於觀察台灣中央與地方的社會福利支出，在1994年後有大幅成長之態勢。因此，涵蓋年份從1995年到2005年，共11年的資料。至於追蹤對象方面，以台灣23個縣市為研究標的。此乃因福建省金門縣及連江縣之人文地理環境特殊，中央政府辦理之例行性統計調查，多未將金門縣及連江縣納入統計調查範圍，故將研究範圍設定為臺灣地區，包括臺北市、高雄市等23縣市。在整合上述的人口遷移理論與相關實證文獻，並逐一檢討各項可能的影響人口遷移的決定因素後，本文由歸

²⁷根據「公職人員選舉罷免法」(2005年6月22日修正)第15條規定，有選舉權人在各該選舉區繼續居住四個月以上者，為公職人員選舉各該選舉區之選舉人。而台灣的直轄市及縣市長選舉均發生於年底，因而推論選舉誘發的遷移多發生於年初，故以當年度的選舉年為虛擬變數。

納出前述可能影響人口遷移的主要原因，並加以驗證。茲將每一變數的預計的資料來源，將分述如下。

由於縣市政府在執行社會福利政策時，乃以戶籍資料作為認定縣(市)民之依據。基於此，本研究將使用內政部戶政司的戶籍遷移資料作為應變數。各縣市的人口遷移率，以及男性和女性遷移率來自於《中華民國台閩地區人口統計》。而有鑒於各縣市的社會福利政策不盡相同，且涵蓋範圍廣泛，故以各縣市平均每人享有的社會福利支出金額，作為衡量各縣市的社會福利水準。各縣市的人均享有社會福利決算數來自於《直轄市及縣市地方決算審核結果綜合報告》；每戶可支配所得乃根據《中華民國台灣地區家庭收支調查報告》；每人享有地方公共支出（不含社會福利支出）來自於各縣市的《統計要覽》；各縣市的重大犯罪率來自於《中華民國臺灣地區警政統計年報》；產業結構（從事工業及服務業就業人口之比例）以及失業率來自於《中華民國台灣地區人力資源統計調查年報》；至於代表教育服務的國中小師生比乃根據《中華民國教育統計》而得；選舉年虛擬變數來自於中選會公佈歷年選舉時間為依據。以上各變數都以2001年為基期的消費者物價指數（CPI）加以平減，以消除物價波動的影響。

第三節 實證變數假設

由前所述，影響人口遷移的因素大致可分為經濟因素、非經濟因素、以及遷移者特質。以下針對上述幾項變數，加以詳細說明其對人口遷移的影響。

由於本研究欲瞭解當一縣市的社會福利政策，較相鄰縣市完善時，是否會吸引較多的人口遷入。依Sjaastad (1962) 之建議，遷移行為之研究，以遷移率作為應變數較以遷移人口數量更為適當。在迴歸分析中，若以遷移率做為應變數，則可以隨機機率來解釋遷移行為，並可避免因地區間人口規模差異所引起可能誤差。是故，本研究採用淨人口遷移率作為迴歸模型中的應變數，即遷入人口扣減遷出人口後的淨遷移人口數，除以該縣市的年中人口數。而為了確保實證結果的完整性，本研究預計除了採用人口淨遷移率外，尚將使用男性及女性人口的淨遷移率，並分別將其加入實證模型進行實證估計，以進一步探討男性或女性人口的福利遷移傾向是否一致。

至於自變數方面，Gramlich and Laren (1984) 發現，參與AFDC計畫的人較可能從福利較低的地區遷移到福利較高的地區。Meyer (1998) 指出社會福利的接受者會自動往高社會福利支出的州遷移，即使並非出於社會福利利益的動機。Blank (1988) 發現單親媽媽通常會選擇居住於有較高的AFDC及勞動市場來源所得的地區。Enchautegui (1997) 發現社會福利對於單親媽媽與接受公共救助的婦女的遷移有很大的影響；McKinnish (2005) 亦支持社會福利利益的確會影響遷移決定。因此，本文預期每人可享有社會福利支出決算審定數 (SW) 對縣市之人口淨遷移率 ($MR_{i,t}$) 有正向之影響。

而就經濟因素而言，劉小蘭及劉念華 (1995) 的研究指出，台灣地區

及各區域人口淨遷移皆與就業成顯著正相關。洪嘉瑜等（2003）也實證地區的相對就業機會、相對薪資與相對產業結構，對人口遷入有正向關係。因此，本文預期失業率（*UNEM*）的估計值應為負值。此外，王素英（1986）指出，聚集經濟與非農就業機會的擴張是吸引人口遷入的拉力因素，即支持產業結構對人口遷移具有影響力。因此，此外，本文預期從事工業及服務業之人口比例（*INSER*）的估計值應為正值。另外，Cebula（1978）支持人口遷移會受到實質所得差異之影響；而Berger and Blomquist（1992）認為工資水準會提高家戶發生遷移的機率，以及決定遷移地點的選擇。是故，本文預期每戶可支配所得（*I*）的估計值應為正值。至於，支持在探討地方公共財（租稅型態與地方公共支出水準）對於人們居住區位之選擇，多由Tiebout（1956）所提之「以足投票」理論出發。²⁸而在公共支出部分，根據薛立敏等（2003）指出，能夠讓遷入者直接感受到地方政府公共支出水準的部分，主要集中於當地的教育資源、治安與交通狀況、及地方衛生情形。因此，社會福利支出以外的公共支出服務（*GE*）預期對縣市之人口淨遷移率（ $MR_{i,t}$ ）有正向之影響。²⁹

此外，影響人口遷移的非經濟因素方面，尚有犯罪率、教育水準、與選舉年等因素。Cebula and Vedder（1973）指出，犯罪水準高將不利於人口遷入，甚至易促使人口向外流出。Cushing（1993）亦認為犯罪率會阻礙人口遷入。Cebula（1978）指出，不論白人或非白人，均對地方的教育水準有相當的敏感度。此外，每逢選舉年（特別是縣市長的選舉），台灣各縣市就會產生大規模人口遷移的現象。因此，本文預期犯罪率（*CRI*）、教育水準（*EDU*）、以及選舉年（*ELE*）的係數估計值應分別為負值、正值、

²⁸鑒於所得稅為中央稅，各地的所得稅率差異不大，故採地方稅（包括地價稅、土地增值稅、使用牌照稅、房屋稅、娛樂稅、印花稅、契稅）作為租稅變數。經初步檢定發現，各縣市的地方稅與每戶可支配所得、產業結構間具有高度相關（係數分別為0.85936和0.81625）。為避免自變數間具高度相關，故本研究不討論租稅變數。

²⁹為了避免自變數間有高度相關，因此，本研究的公共支出變數為扣除社會福利支出的每人可享公共支出水準。

以及正值。

此外，就遷移者特質而言，Ravenstein (1889) 曾提出婦女的遷移率比男性高，男性有較長距離的遷移。王素英 (1986)，發現人口遷移具有選擇性，即年齡結構愈輕，教育程度愈高，或為女性則愈具遷移傾向。因男性與女性的遷徙行為並不相同，因此本文特別針對男性和女性遷移率分別加以估計。以上所述各項實證變數的說明與基本統計量，均彙整列於表 5 之中。

表 5：實證變數之基本統計量

變數名稱	符號	變數說明	平均值	標準差	預期影響方向
人口淨遷移率	NMR_t	模型一：該年各地區人口淨遷移率(‰)	-1.55	7.25	
男性淨遷移率	$MNMR_t$	模型二：該年各地區男性淨遷移率(‰)	-1.34	6.74	
女性淨遷移率	$FNMR_t$	模型三：該年各地區女性淨遷移率(‰)	-1.81	7.93	
人均社會福利決算審定數	SW_{t-1}	前一年之各地區平均每人社會福利支出決算審定數(新台幣千元)	3.51	2.86	+
每戶每年可支配所得	I_{t-1}	前一年之各地區每戶每年可支配所得(新台幣萬元)	81.16	14.74	+
每人享有地方公共支出	GE_{t-1}	前一年之各地區每人享有地方公共支出(扣除社福支出的公共支出)。(新台幣千元)	27.33	10.10	+
產業結構	$INSER_{t-1}$	前一年之各地區工業以及服務業就業人口比例(%)	88.79	9.76	+
失業率	$UNEM_{t-1}$	前一年之各地區失業率(%)	3.30	1.32	-
重大犯罪率	CRI_{t-1}	前一年之各地區重大犯罪率(件/十萬人)	241.93	162.03	-
教育服務	EDU_{t-1}	前一年之各地區教育服務(%)	21.74	7.21	+
選舉年虛擬變數	ELE_t	選舉年虛擬變數。如果該年遇有直轄市或縣市長選舉則設為 1。否則設為 0。	0.26	0.44	+

資料來源：《中華民國台閩地區人口統計》、《直轄市及縣市地方決算審和結果綜合報告》、主計處《中華民國台灣地區家庭收支調查報告》、各縣市《統計要覽/年報》、《中華民國台灣地區人力資源統計調查年報》、《中華民國台灣地區警政統計年報》、《中華民國教育統計》。

註：1. 淨遷移率 = (遷入人數 - 遷出人數) * 1000 / 年中人口數。年度中人口數(人) = (前一年底人口數 + 該年底人口數) / 2。

2. 平均每戶可支配所得(萬元/戶)。公式：可支配所得 / 總戶數。

3. 前一年之各地區每人享有地方公共支出，基於避免與其他自變數有高度相關，故採扣除社會福利支出後的公共支出為資料變數。

4. 適逢會計年度變更，88 年度(含)以前係指上年 7 月 1 日至當年 6 月 30 日止，89 年度係指 88 年下半年及 89 年，90 年度以後與曆年相同。為使比較基礎一致，本研究對政府歲出金額，皆以加權調整為歷年制。

5. 重大刑案發生率(或稱犯罪率)：指每十萬人口(期中人口)中刑案發生件數(或以萬人計算)。精省前(1998 年前)，各縣市刑案發生數不包括台灣省警政廳所屬機關(即不包括警政廳/處所屬鐵路警察局、公路警察大隊及基隆、台中、高雄、花蓮四個港警所資料)。公式：重大刑案發生數 * 100000 / 年中人口數。重大刑案細分為暴力犯罪(包含故意殺人、強盜、搶奪、擄人勒贖、強制性交、恐嚇取財、重傷害等)、重大竊案、汽車失竊。

6. 金額變數皆經物價平減(以 2001 年物價為基期)。

第四節 本章小結

本研究的目的是在於瞭解1995至2005年間，台灣地區各縣市地方社會支出，對人口遷移的影響為何。本章先詳細說明本研究使用之研究設計。首先，在第一節中將闡明使用追蹤資料進行研究的原因，並比較其下的固定效果模型和隨機效果模型之異同和取捨方法。接著，第二節中建立本研究的實證模型和說明實證資料之來源。至於依據本研究實證模型設立之實證變數假設，則陳述於第三節中。本章經整理重點如下：

一、研究方法說明

為了同時擁有時間序列的動態性質，並且兼顧橫斷面資料表達不同樣本間特性的優點，本研究之設計採用結合橫斷面資料和時間序列資料的追蹤資料。除了能觀察到因使用橫斷面資料，變數間原來無法觀察到的動態變化，甚者得使用固定效果模型來分析觀察個體的固定效果，瞭解個體的特性之外，尚有減少共線性的問題以及降低估計上的偏誤等好處。因此，符合本研究觀察群體長期性的要求。其中，追蹤資料模型又可分為固定效果模型和隨機效果模型，透過Hausman（1978）的方法檢定，決定本文採用固定效果模型較佳

二、實證模型設定與資料來源

（1）在實證模型設定方面，本研究主分要分為三個模型，來深入探討社會福利支出對人口遷移、男性及女性遷移者的影響是否一致。模型一以各縣市人口淨遷移率為應變數，模型二與模型三的應變數分別為，各縣市男性淨遷移率與女性淨遷移率。根據文獻整理，放入的其他自變數皆包括每年各縣市的每人可享有社會福利支出決算審定數、每年各縣市的每戶可支配所得、每人可享有除了社會福利支出以外的公共支出服務、每年各縣

市的重大犯罪率、每年各縣市的教育服務、每年各縣市從事工業及服務業的比例、每年各縣市的失業率，以及選舉年虛擬變數。此外，在估計的實證模型中，加入區域特質的個別效果，以衡量各縣市可能存在個別特質的型態；並考慮時間效果，以捕捉時間趨勢對被解釋變數之影響。因此，本研究之實證估計模型為雙因子固定效果模型。

(2) 在實證資料來源方面，由於台灣中央與地方的社會福利支出，在1994年後有大幅成長之態勢，故本文使用1995年到2005年，共11年台灣縣市別的追蹤資料。上述實證模型中的各變數資料來自於《中華民國台閩地區人口統計》、《直轄市及縣市地方決算審核結果綜合報告》、《中華民國台灣地區家庭收支調查報告》、各縣市的《統計要覽》、《中華民國台灣地區警政統計年報》、《中華民國台灣地區人力資源統計調查年報》，以及《中華民國教育統計》。另外，選舉年虛擬變數則參考中選會網站。

三、實證變數假設

實證模型設定後，本研究對實證變數進行假設。經由過去相關文獻的探討後，本文初步的假設為：負相關的變項有各縣市的重大犯罪率和失業率；正相關的變項有各縣市平均每人社會福利支出、每戶可支配所得、教育服務、產業結構以及每人享有公共支出（不含社會福利支出）。至於選舉年虛擬變數，亦預期為正相關。