

## 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，在第一節中將闡明使用追蹤資料進行研究的原因，並接著說明選用 OLS 模型的原因及其基本架構。第二節中將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，將陳述於第三節中。

### 第一節 研究方法說明

如前所述，本研究的目的是在於瞭解社會福利支出對所得分配之影響，並決定以台灣各縣市的追蹤資料為觀察對象進行研究，藉此得到最佳的估計。然而，擇定追蹤資料並非隨意挑選，在本節中便將對該擇定理由加以說明，並更深入地探討其下的 OLS 模型。

#### 一、追蹤資料的使用

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，而且又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，藉此降低參數估計上的偏誤。因此，若想要分析某觀察群體長期性的決定因素，使用追蹤資料較能分析出正確且嚴謹的結果，而這正符合本研究的要求。

依本文所使用之追蹤資料特性而論，選用固定效果分析法 (fixed effects approach) 應是較合適的。但由於觀察點個數有限，相對於太多要估計的

代表縣市別之參數，線性重合（multicollinearity）現象致使迴歸效果不佳，<sup>14</sup>因而改用簡單的 OLS 估計法。

## 二、OLS 模型

簡單線性迴歸方程式表示為：

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \quad (12)$$

變數  $Y_t$  稱作應變數， $X_t$  稱作自變數或解釋變數， $\varepsilon_t$  稱為隨機干擾項（stochastic disturbance），也稱為誤差項（error term）， $\beta_1$  和  $\beta_2$  是未知迴歸係數，下標  $t$  代表  $t$  個觀察值。如果樣本資料按照時間先後順序排列，稱為時間序列資料（time-series data），下標  $t$  代表時間，大多屬總體經濟資料；如果樣本資料是在某個特定時點，針對某特定地區或群體取樣得到，稱作橫斷面資料（cross-section data）， $t$  代表觀察對象，大多屬個體經濟資料。此外，有些資料同時包含時間序列與橫斷面兩種性質，稱為縱橫資料或追蹤資料。此時需使用兩個下標，例如  $i$  與  $t$ ，分別代表觀察對象和時間。變數  $Y_t$  和  $X_t$  透過資料蒐集過程可被觀察到，但  $\varepsilon_t$  無法觀察到。

由於  $Y_t$  是隨機變數  $\varepsilon_t$  的函數，故  $Y_t$  也是隨機變數，其機率分配取決於  $\varepsilon$ 。迴歸模型（12）隱含對應每個  $X$ ，有整個  $Y$  的機率分配與之對應。迴歸模型中包含隨機干擾項的理由有三，成為它與數學方程式的主要差異。1. 它代表所有無法控制的因素。2. 真正的迴歸模型可能為曲線式，而研究者卻使用線性迴歸式，則隨機干擾項亦包含近似誤差在內。3. 可代表人們的一些突發行為。

<sup>14</sup> Hyclak (1996) 利用美國 200 城市地區的追蹤資料，針對勞動需求的結構性改變對失業率的影響進行研究。劉鶯釗、黃智聰（2002）探討造成台灣縣市別失業率差異的決定因素，並且進一步分析影響各地區女性失業率的因素。高安邦、黃智聰、潘俊男（2005）係探討中國大陸地方政府效率與吸引外資之間的關係。此三篇文章在模型估計方法的選用上，皆存在與本文雷同的問題。

完整迴歸模型設定，除了（12）式外，還包括下列五點假設，統稱為基本假設或古典假設：

1. 常態性： $\varepsilon$  具有常態分配。
2. 零平均數： $\varepsilon$  的期望值為零，即  $E(\varepsilon_t)=0$ 。
3. 變異數齊一性： $\varepsilon_t$  的變異數是常數，不隨觀察值而改變，即  $\text{Var}(\varepsilon_t)=\sigma^2$ 。
4. 無自我相關：任一兩個不同隨機干擾項的共變數皆為零。即  $\text{Cov}(\varepsilon_t, \varepsilon_s)=0, \forall t \neq s$ 。
5. 解釋變數  $X$  非為隨機變數，在重覆抽樣過程中其值固定不變。而且， $\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 \neq 0$ ，當樣本數  $T$  增加至無限大時， $\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 \neq 0$  為有限數值。

完整簡單線性迴歸模型的設定，需包含（12）式及五點基本假定，合稱古典常態線性迴歸模型（classical normal linear regression model）。前三個假設聯合起來，指誤差項  $\varepsilon_t$  是具有常態分配的隨機變數，平均數為零，變異數為未知常數；第四個假設強調不同  $\varepsilon$  之間，相互獨立。前四個假設聯合起來，指出不同  $\varepsilon$  為相同且彼此獨立的隨機變數。最後一個假設，限定解釋變數  $X$  為非隨機變數，意指其為可控制或可預測者，則  $E(\varepsilon_t X_s)=X_s E(\varepsilon_t)=0, \forall t, s$ ；因為  $X$  非為隨機變數，故假設其值在不同樣本中固定不變。所有基本假設將被用來推導迴歸係數的估計式和其性質，已知誤差項為平均數零的常態分配，它的變異數  $\sigma^2$  未知。因此，在迴歸模型中，實際共有三個未知數，包含  $\beta_1$ 、 $\beta_2$  及  $\sigma^2$ 。估計迴歸模型中的三個參數，可視同估計應變數  $Y$  的機率分配參數，其中  $\beta_1$  和  $\beta_2$  與  $Y$  的平均數有關，則  $\sigma^2$  為  $Y$  的變異數。

## 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

本研究主要目的在研究社會福利支出對所得分配的影響，是否有確實達成其政策目標，並考慮其他可能影響所得分配的因子，期望建立一個能更完整解釋所得分配變動的實證模型。為更深入了解改善所得分配的具體可行政策，本研究將分兩個模型，來探討何種社會福利支出形式較能改善所得分配，第一個模型的社會福利支出變數 ( $SUB_{t-1}$ )，放的是各縣市年度社會福利支出決算審定數；第二個模型的社會福利支出變數 ( $SUB_{t-1}$ )，則改放各縣市社會福利支出中的低收入戶生活扶助金額。其中，研究對象為台灣23個縣市；模型一的研究期間為1994年至2004年，共有253個樣本數；<sup>15</sup>模型二的研究期間亦為1994年至2004年，共有253個樣本數。此外，為了對影響縣市所得分配不均度的因素做更精準的分析，本文之研究變數除了地方及全國的經濟變數外，還列入產業變數、婦女勞動參與變數、教育程度變數與區域變數一併考量。

本文依據現有的文獻，決定解釋變數包含實質社會福利決算審定數或實質低收入戶扶助、實質平均每戶所得、實質平均每戶所得平方、工業就業人口比例、服務業就業人口比例、教育程度變異數、婦女勞動參與率、北部虛擬變數、中部虛擬變數、東部虛擬變數、2001年虛擬變數。本文所使用的OLS估計法，以數學式表

$$\begin{aligned} Gini = & \beta_0 + \beta_1 \log(SUB_{t-1}) + \beta_2 I_{i,t-1} + \beta_3 ISQ_{i,t-1} + \beta_4 INDUS_{i,t-1} + \beta_5 SERVICE_{i,t-1} \\ & + \beta_6 EDUVAR_{i,t-1} + \beta_7 WOMAN_{i,t-1} + \gamma_1 NORTH + \gamma_2 MIDDLE \\ & + \gamma_3 EAST + \alpha_1 Y2001 + \varepsilon_{i,t} \end{aligned} \quad (13)$$

在文獻回顧中，部分文獻證實社會福利支出有可能改變所得分配，但

<sup>15</sup> 因政府編定的各縣市決算審定數的彙編僅從1997年開始。

這種關係卻未必是單向的。Alesina and Rodrik (1994) 以及Persson and Tabellini (1994) 認為，所得的不均也迫使政府採取某些重分配政策。因此，為了避免內生性 (endogeneity) 的計量問題，上述的解釋變數皆為前一期的數據。另外，各類有關金額的解釋變數，皆以2001年的物價指數加以平減，轉換為實質變數。

## 二、本研究之資料來源

以相關文獻決定的變數資料來源如下：模型一的社會福利決算審定數變數資料來自監察院審計部的《直轄市及縣市地方決算審核結果綜合報告》，而模型二的低收入戶生活扶助變數資料來自內政部統計處的《中華民國內政統計年報》。代表經濟發展變數各縣市每戶平均所得 ( $I_{t-1}$ ) 及其平方項 ( $ISQ_{t-1}$ ) 的資料來自行政院主計處中部辦公室的《中華民國台灣地區個人所得調查報告》；代表產業結構轉變的工業人口比例 ( $INDUS_{t-1}$ ) 及服務業人口比例 ( $SERVICE_{t-1}$ )、代表婦女勞動參與率 ( $WOMAN_{t-1}$ ) 及代表教育程度 ( $EDUVAR_{t-1}$ ) 的變數資料則來自行政院主計處的《中華民國台灣地區人力資源調查統計年報》。

### 第三節 實證變數假設

本研究之應變數為代表各縣市家庭所得不均度的吉尼係數。而此吉尼係數是根據 Mookherjee and Shorrocks (1982) 所提出吉尼係數分解公式計算出來的。本文依行政院主計處之家庭收支調查原始資料，就各縣市家庭的可支配所得，採吉尼係數以衡量所得分配不均等程度，並探討各類因素結構變動趨勢及其影響效果。本文為便於研究分析，利用過去十年間所得資料，佐以實證分析，研討各縣市家庭所得不均度與各類因素變動之相關程度，進而剖析各所得來源分配不均之主要導因及可能採取之因應之道，以改善持續惡化的所得分配不均。

張漢裕 (1975) 指出以 Kuznets 的倒 U 型曲線來解釋台灣所得分配的情形，則所得分配的反轉點會出現在 1968 年。前半段其間的家庭所得分布趨勢為集中化，而在後段其間的趨勢則趨向均等化。因為各縣市缺乏代表經濟成長的 GDP 或工業指數資料，故本研究以與經濟成長連動的前一年各縣實質每戶平均所得 ( $I_{t-1}$ ) 來代表經濟成長的概念，並嘗試放入其平方項 ( $ISQ_{t-1}$ )，期望找出其代表的經濟成長階段，來驗證台灣的所得分配現況是否符合 Kuznets 的倒 U 型曲線假說。

本研究亦放入工業人口比例 ( $INDUS_{t-1}$ ) 及服務業人口比例 ( $SERVICE_{t-1}$ ) 代表產業結構轉變變數。郭婉容 (1976) 探討 1960 及 1970 年代台灣地區的所得分配狀況，根據吉尼係數與最窮和最富 20% 家庭組別之相對所得所計算之 Oshima 指標，發現台灣的家庭所得分配在 1964 年至 1972 年間有逐步均等化的趨勢，且非農業部門內的分配均等化的現象，有助於全國吉尼係數的下降。而對此結果的主要解釋為當時勞動密集工業的快速，吸收了大部分機會成本較低的勞動力，進而縮小非農業部門薪資所得的差距，促使均等化的家庭所得分配的產生。劉錚錚 (1978) 利用產業三角圖，探討初級、次級和三及產業的所得分配。結論指出：1964 年為一

轉捩點，從當年開始，台灣開始進入經濟快速成長的階段，產業結構快速轉變，每人之實質所得快速提高，而產業間勞動平均生產力的差距快速縮小，且為物價波動平穩之時期。本研究用來自《中華民國台灣地區人力資源統計年報》的各縣市工業人口比例及服務業人口比例，希望驗證 1994 年至 2004 年這十一年間，產業結構變動對所得分配不均是否也有如文獻所述的影響？進而對政府的產業決策提出正確的建議。

多數文獻認為教育程度的差異有可能是影響家庭所得不均的重要因素之一。林金源、朱雲鵬（2000）曾利用主計處家庭收支調查資料，針對個人工作報酬不均度的長期變化進行研究，該研究指出個人所得分配的變化與勞動者的特性息息相關，他們發現性別與相關變數在 1976 年至 1996 年間是最重要的平均化因子。主要的原因是「男性與女性的所得差距隨年紀增長而擴大的現象」已逐漸消失。另外，不同教育程度間的報酬差距逐步擴大，使教育變成 1983 年之後工作報酬不均化的重要因子，且其程度與日俱增。本研究利用教育程度變異數（ $EDUVAR_{t-1}$ ）來檢視，一個縣市平均教育程度的集中與否，會對家庭所得不均造成何種影響。

近年來，兩性平權的概念逐漸受到重視，婦女勞動參與逐漸普遍，而對其研究的文章亦趨增加。婦女勞動參與究竟是擴大或縮小家戶間所得的差異？Mincer（1974）、Thurow（1975）、Smith（1979）及 Lehrer and Nerlove（1984）研究結論皆指出，若丈夫的所得水準較低時，婦女勞動參與可縮小家戶間所得的差異。張鶯釗與張清溪（1987）使用變異係數的平方做為不均度的衡量指標，用以分析婦女勞動報酬對家庭所得分配所造成的影響。他們使用 1977 年至 1985 年的台灣地區家庭收支調查報告上的資料從事實證研究，結論指出妻子的勞動報酬對家庭所得分配是具有均等化的效果。本研究亦將婦女勞動參與（ $WOMAN_{t-1}$ ）此項變數放入，期望找出近年來婦女勞動參與對家庭所得不均度的影響。

台灣依行政區劃分，可分成四個主要區域，分別為北部地區、中部地區、南部地區及東部地區。<sup>16</sup>Kim (2003) 以所得來源及區域分析韓國所得不均度的來源。此種不均度指數分解法提供較好的觀點來討論所得不均度的實際決定因素及區域和所得群體間的內部關連性。再者，這些指數顯示整體所得不均度在區域內及區域間的差異，可當作區域或國家以平衡區域發展為目標時的表現方法。為考慮地理區域及可能的區域文化差異，故放入代表區域的三個虛擬變數 (*NORTH*、*MIDDLE*、*EAST*)，希望能找出現狀下台灣各區域的所得分配狀況及差異。

根據各縣市吉尼係數資料分析，發現 2001 年的各縣市吉尼係數明顯跳升，主要可能原因為當年度經濟成長率為負的，經濟環境的惡化擴張了各縣市的所得差距。故將此年設一虛擬變數 (*Y2001*)，觀察此年對家庭所得分配不均的可能影響。各變數定義、基本統計量與預期影響皆呈現於表 11。

---

<sup>16</sup> 北部區域包含台北市、基隆市、新竹市、台北縣、宜蘭縣、桃園縣、新竹縣；中部區域包含台中市、苗栗縣、台中縣、南投縣、彰化縣、雲林縣；南部區域包含高雄市、台南市、嘉義市、嘉義縣、台南縣、高雄縣、屏東縣、澎湖縣。東部區域包含花蓮縣、台東縣。



表 11：各變數定義、基本統計量與預期影響

變數符號	變數定義	平均數 (標準差)		預期影響
		模型一	模型二	
$SUB_{t-1}$	模型一：前一年之各地區實質社會福利支出決算審定數 (新台幣萬元)	323422.78	14894.21	
	模型二：前一年實質低收入戶生活扶助 (新台幣萬元)	(515581.09)	(19963.44)	
$I_{t-1}$	前一年之各地區實質每戶平均所得 (新台幣萬元)	77.93 (14.70)	77.93 (14.70)	+
$ISQ_{t-1}$	前一年之各地區實質每戶平均所得的平方工實質平均工資 (新台幣萬元)	6287.63 (2471.20)	6287.63 (2471.20)	-
$INDUS_{t-1}$	前一年之各地區工業就業人口比例 (%)	35.99 (9.31)	35.99 (9.31)	-
$SERVICE_{t-1}$	前一年之各地區服務業就業人口比例 (%)	52.31 (12.23)	52.31 (12.23)	-
$EDUVAR_{t-1}$	前一年之各地區教育程度變異數 (%)	11.76 (1.82)	11.76 (1.82)	+
$WOMAN_{t-1}$	前一年之各地區婦女勞動參與率 (%)	45.15 (5.58)	45.15 (5.58)	-
$NORTH$	如果該地區位於北部區域則設為 1。否則設為 0。	0.3 (0.46)	0.3 (0.46)	-
$MIDDLE$	如果該地區位於中部區域則設為 1。否則設為 0。	0.26 (0.44)	0.26 (0.44)	-
$EAST$	如果該地區位於東部區域則設為 1。否則設為 0。	0.26 (0.44)	0.26 (0.44)	+
$Y2001$	如果該年為 2001 年則設為 1。否則設為 0。	0.09 (0.28)	0.09 (0.28)	+

資料來源：1993-2004 年《中華民國內政統計年報》、《中華民國台灣地區人力資源統計年報》、《中華民國台灣地區個人所得調查報告》。

- 註：1. 所有變數皆為落後一期 (上一年) 的資料。  
 2. 所有金額變數皆以 2001 年之物價水準平減。  
 3. 模型一的觀察值為 1994 年至 2004 年，共 253 個觀察值；模型二的觀察值為 1994 至 2004 年，共 253 個觀察值。

## 第四節 本章小結

本研究的目的是在於瞭解台灣社會福利支出及低收入戶生活扶助對台灣各縣市所得分配的影響，並決定以台灣各縣市的追蹤資料為觀察對象進行研究，藉此得到最佳的估計。本章先詳細說明本研究使用之研究設計。首先，在第一節中將闡明本研究使用追蹤資料進行研究的原因及優點，並說明選用 OLS 模型的原因及其基本架構。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

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，因此不但擁有時間序列的動態性質，而且又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，能觀察到變數間原來無法觀察到的動態變化。此外，尚有降低估計上的偏誤、減少共線性的問題以及提供更完整的訊息等好處。因此，符合本研究觀察群體長期性的要求。依本文所使用之追蹤資料特性而論，選用固定效果分析法（fixed effects approach）應是較合適的。但由於觀察點個數有限，相對於太多要估計的代表縣市別之參數，線性重合（multicollinearity）現象致使迴歸效果不佳，因而改用簡單的 OLS 估計法。

### 二、實證模型設定與資料來源

（一）在實證模型設定方面，本研究主要分為兩個模型，來深入探討何種社會福利支出形式較能改善所得分配，第一個模型的社會福利支出變數（ $SUB_{t-1}$ ），放的是各縣市年度社會福利支出決算審定數，研究對象為台灣 23 個縣市，研究期間為 1994 年至 2004 年，共有 253 個樣本數；第

二個模型的社會福利支出變數 ( $SUB_{t-1}$ )，則改放各縣市社會福利支出中的低收入戶生活扶助金額，研究對象同為台灣 23 個縣市，研究期間亦為 1994 年至 2004 年，共有 253 個樣本數。根據文獻整理，放入的其他自變數皆包括實質平均每戶所得、實質平均每戶所得平方、工業就業人口比例、服務業就業人口比例、教育程度變異數、婦女勞動參與率、北部虛擬變數、中部虛擬變數、東部虛擬變數、2001 年虛擬變數。

(二) 在實證資料來源方面，模型一的社會福利決算審定數變數資料來自監察院審計部的《直轄市及縣市地方決算審核結果綜合報告》，而模型二的低收入戶生活扶助變數資料來自內政部統計處的《中華民國內政統計年報》。經濟發展變數的資料來自行政院主計處中部辦公室的《中華民國台灣地區個人所得調查報告》；產業結構轉變、婦女勞動參與率及教育程度的變數資料則來自行政院主計處的《中華民國台灣地區人力資源調查統計年報》。

### 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：負相關的變項有實質平均每戶所得平方、工業就業人口比例、服務業就業人口比例、婦女勞動參與率、北部虛擬變數及中部虛擬變數；正相關的變項有實質平均每戶所得、教育程度變異數、東部虛擬變數以及 2001 年虛擬變數。至於社會福利支出變項是否真能改善所得分配，與其呈負向關係，則無法預知。