

第四章 研究設計

從過去的文獻我們可以發現，影響地區恩格爾係數的因素眾多，且各項因素之間又存在著錯綜複雜的關係，學者們都作了相當深入的研究。在中國大陸政府宣示了「全面建設小康社會」之後，關於恩格爾係數的話題越來越受到重視，然而對中國大陸恩格爾係數整體的實證研究仍相對缺乏，且各方學者眾說紛紜。有鑑於此，本文將在本章建立實證模型，並說明相關變數假設與資料來源。

第一節 實證模型設定

本碩士論文的研究目的在於釐清影響中國大陸恩格爾係數變動因素的分析，並加入其他因素加以探討，以其對於掌握恩格爾係數的變動有更深入的了解。

一、追蹤資料(panel data)

本碩士論文將使用追蹤資料(panel data)的方法來分析資料，意即同時考慮橫斷面資料(cross-section data)與時間序列(time-series data)資料，兼有兩者的優點。橫斷面資料，即同一時間上不同區域的數據，若僅以橫斷面進行分析，容易因經濟個體本身存在特殊特性，而產生異質變異(Heteroscedasticity)。時間序列資料則是同一區域，不同時間點之資料，由於只考慮到相關變數的時間序列資料，可能發生序列相關(Serial Correlation)的現象。

發生序列相關或者異質變異時，雖然仍具有不偏性及一致性，但不具有有效性，且均非最佳線性不偏估計式(Best Linear Unbiased Estimate, BLUE)，因而可能對信賴區間與假設檢定等統計推論產生偏差。因此為了避免時間序列或橫斷面資料分析無法同時比較時間變動及個體差異，並避免忽略某些變數而產生估計偏誤或無效，故在此採用追蹤資料模式分析。

Hsiao(1986)的研究發現追蹤資料有以下的優點：

1. 通過對不同橫面單元不同時間觀察值的結合，追蹤資料分析能有效增加樣本數以及模型的自由度。
2. 降低變數間共線性(Collinearity)的問題。
3. 能減低個體加總所產生的偏誤。
4. 能顯示時間序列與橫斷面資料所無法單獨檢定的效果。
5. 可以控制經濟個體的異質性。追蹤資料能夠分析橫斷面和時間序列資料中不能控制的涉及地區與時間為常數的狀況。

二、雙因子固定效果模型(Two-Way Fixed Effect Model)

固定效果模型(Fixed Effect Model)與隨機效果模型(Random Effects Model)為追蹤資料最常被採用之模式。若樣本來自特定母體，且個體特性不隨時間不同而改變時，使用固定效果模型可強調個體差異性；若樣本是隨機抽樣自母體，則較適合使用隨機效果。

本碩士論文研究目標為中國大陸城鄉恩格爾係數的差異，為避免估計上的偏誤，在迴歸估計式上，就必須將各省市地區的個別特質效果 (Individual-Specific Effect) 予以考慮。又基於捕捉時間動態效果 (Time-Specific Effect) 之考量，故本碩士論文採用能夠兼顧兩者的雙因子固定效果模型 (Two-Way Fixed Effects Model) 作為估計方法。

雙因子固定效果模型其基本型式如下：

$$Y_{it} = \alpha_i + \theta_t + \beta_0 + \sum_{k=1}^k \beta_k + X_{kit} \varepsilon_{it} \quad (2)$$

θ_t 代表第 t 期時間虛擬變數的係數，下標 i 代表觀察單位， $i = 1, \dots, N$ ， t 是觀察時間 $i = 1, \dots, t$ ， $\beta = (\beta_1 \dots \beta_k)'$ 代表 $(K-1) \times 1$ 行向量， X_{it} 為對應之自變數向量，包含截距項即為上式。

在實證模型方面，本碩士論文以 ENG 代表恩格爾係數，探討何種因素會影響中國大陸恩格爾係數的變動。函數模型呈現如下：

$$ENG = f(RI, DEPR, IF, PIRE, CLO, MED, EDU, FUR, CPI) \quad (3)$$

上列式子中，ENG 代表恩格爾係數，RI 代表實質收入，DEPR 代表依賴扶養比 (非勞動人口/勞動人口)，IF 代表食品物價指數，PIRE 代表房地產價格指數，CLO 代表衣著支出，MED 代表醫療支出，EDU 代表教育支出，FUR 代表耐用品消費，CPI 代表物價指數。式(3)可進一步表示如下：

$$ENG_{it} = \alpha_i + \theta_t + \beta_1 IR_{it} + \beta_2 DEPR_{it} + \beta_3 IF_{it} + \beta_4 PIRE_{it} + \beta_5 CLO_{it} + \beta_6 MED_{it} + \beta_7 EDU_{it} + \beta_8 FUR_{it} + \beta_9 CPI_{it} + \varepsilon_{it} \quad (4)$$

其中， α_i 代表地區特質效果， θ_t 代表時間特質效果， ε_{it} 代表殘差項； i 代表30個省市。 $i=\{$ 北京、山西、山東、浙江、河南、黑龍江、吉林、寧夏、河北、陝西、甘肅、天津、湖南、新疆、上海、江蘇、青海、廣東、重慶、四川、貴州、湖北、遼寧、福建、江西、雲南、廣西、安徽、海南 $\}$ 。其中，西藏因為資料缺失並未列入在本碩士論文的研究範圍內。 t 代表時間， $t=\{$ 1997, 1998, 1999, ..., 2005, 2006, 2007 $\}$ 。本碩士論文共採用30個省市地區、11年，共330個樣本資料。

比較城鎮與農村影響因素的不同，為本碩士論文的重要研究目的，故在模型估計上將分為城鎮與農村作實證研究。

在估計迴歸式時，若遺漏或包含部相關之變數，這些變數將會透過殘差值嚴重影響整個檢定結果，而殘差也可能出現自我相關現象。此時可對模型所包含的變數作顯著性檢定。

第二節 實證變數假設

實質收入。在恩格爾最原始的研究中，收入就是影響恩格爾係數高低的唯一因素，即「收入上升會導致恩格爾係數的下降」，收入為界定人民生活水平的重要指標，後來 Chenery and Syrquin(1975)的研究也證實了這一點。在中國，雖然傳出許多質疑的聲浪，但不可否認收入仍是決定恩格爾係數變動的重要變數。所以，可以預期實質收入與恩格爾係數之間為負的關係。

人口依賴比。一個省份依賴人口越多，自然會對勞動人口造成負擔，因為依賴人口沒有收入，他們的食物支出來源將轉嫁到勞動人口的身上。一地的依賴人口越多，勞動人口就必須支出更多的食物消費，造成恩格爾係數上升。所以本碩士論文預期人口依賴比與恩格爾係數為正的關係。

食品價格指數。食品價格指數代表著食物的價格，如果一地的食品消費總量維持不變，食品價格上升，就會導致恩格爾係數的上升；反之，總量不變、價格下降，將導致恩格爾係數的下降。故食品價格指數與恩格爾係數的預期為正的關係。

房地產價格指數。根據之前的資料顯示，農村的居民在收入增加後，比城鎮的居民更會把收入投入到住房支出。隨著收入的增加，農村居民的住房支出也隨之上升，這有可能對於食品支出產生了排擠的作用。加上農村居民缺乏如城鎮居民般住房的補貼，故本碩士論文預期房地產價格指數與恩格爾係數為負的關係。

衣著消費。城鎮居民比農村居民在此一消費項目上支出較多，代表城鎮居民更願意花錢在衣著上，為了服裝而縮衣節食，與食物支出有排擠作用。故本碩士論文預期衣著消費與恩格爾係數為負的關係。

醫療支出。隨著政策的改變，醫療支出成為近幾年城鄉居民消費結構的新熱點，同時醫療支出也是必要支出，因此城鄉居民都無可避免的在這一支出項目上增加消費金額。故本碩士論文預期醫療支出與恩格爾係數為負的關係。

教育支出。和醫療支出一樣，教育支出也是無可避免的消費項目，尤其受到中華文化「一舉成名天下知」觀念的影響，父母對於這個支出項目也是不能省的。故本碩士論文預期醫療支出與恩格爾係數為負的關係。

耐用品消費。從「老三件」(手錶、自行車、縫紉機)到「新四件」(彩電、洗衣機、錄音機、電冰箱)，耐用品消費一直是家庭支出結構重要的一環，范劍平、張少龍(2005)甚至把恩格爾係數降低歸咎於儲蓄買耐用品，可見耐用消費品對食品消費有排擠作用。故本碩士論文預期耐用消費品與恩格爾係數為負的關係。

物價指數。物價指數的漲跌與民生息息相關，物價指數的漲幅如果大於食品物價指數，一般生活物品支出價格就會相對較低，可能造成假性的恩格爾係數下降。如果與食品物價指數對照，當年食品物價指數的漲幅超過物價指數，就會造成恩格爾係數的上升。故本碩士論文預期物價指數與恩格爾係數為負的關係。以上所述的各項因素、基本統計量與預期影響如下表(4-1)、(4-2)所示。

表 4-1：實證變數定義及統計量與預期效果(城鎮)

符號	變數說明	平均值	標準差	預期影響
IR	各地區實質收入 (單位：元人民幣)	7889.44	3042.51	—
CPI	各地區物價指數 (前一年為 100)	101.90	5.13	—
IF	各地區食品價格指數 (前一年為 100)	101.65	5.20	+
DEPR	各地區依賴比。15 歲以下和 65 歲以上佔勞動人口比(單位：%)	39.94	7.96	+
FUR	耐用品消費支出 (單位：元人民幣)	219.12	100.68	—
PIRE	房地產價格指數 (前一年為 100)	107.44	9.74	—
MED	醫療消費支出 (單位：元人民幣)	407.04	199.05	—
EDU	教育消費支出 (單位：元人民幣)	434.46	178.31	—
CLO	衣著消費支出 (單位：元人民幣)	621.88	195.34	—
ENG	恩格爾係數。 食物支出佔總支出比例 (單位：%)	39.47	4.87	

資料來源：中國統計年鑑、作者自行整理。

表 4-2：實證變數定義及統計量與預期效果(農村)

符號	變數說明	平均值	標準差	預期影響
IR	各地區實質收入 (單位：元人民幣)	2891.08	1355.16	—
CPI	各地區物價指數 (前一年為 100)	101.9	5.13	—
IF	各地區食品價格指數 (前一年為 100)	101.65	5.20	+
DEPR	各地區依賴比。15 歲以下和 65 歲以上佔勞動人口比(單位：%)	39.94	7.96	+
FUR	耐用品消費佔支出 (單位：元人民幣)	99.9	65.50	—
PIRE	房地產價格指數 (前一年為 100)	107.44	9.74	—
MED	醫療消費支出 (單位：元人民幣)	129.00	87.85	—
EDU	教育消費佔總支出 (單位：%)	235.19	141.87	—
CLO	衣著消費支出 (單位：元人民幣)	129.33	65.40	—
ENG	恩格爾係數。食物支出佔總消費支出比例。 (單位：%)	48.24	7.49	

資料來源：中國統計年鑑、作者自行整理。

第三節 資料來源

本碩士論文採用 1997 至 2007 年共 11 年的年資料，加上除西藏以外的 30 個省市資料，共 330 個樣本數進行實證研究。資料來源主要是中國大陸的國家統計局，以及《中國統計年鑑》，以各地的統計局、統計年鑑為補充。如前文所述，回顧國內外相關影響恩格爾係數研究的文獻後，發現實質收入、依賴比、食品價格指數、房地產價格指數、衣著支出、醫療支出、教育支出、耐用品支出以及物價指數均為造成恩格爾係數變動的因素。每一變數的資料來源分述如下：

實質收入、依賴比、食品價格指數、房地產價格指數、衣著支出、醫療支出、教育支出、耐用品支出以及物價指數，從 1997 年開始，中國大陸國家統計局均有將上述資料作為統計普查的項目，只有西藏有所缺漏。

第四節 本章小結

本碩士論文主要的研究目的有以下兩點：一為探討影響中國大陸恩格爾係數變動的因素為何，二為這些因素影響的程度又是多少。經過文獻回顧篩選出實質收入、依賴比、食品價格指數、房地產價格指數、衣著支出、醫療支出、教育支出、耐用品支出以及物價指數作為影響的重要因素，截取 1997 至 2007 年共 11 年的年資料，以及除西藏的 30 個省市共 330 筆資料，對影響恩格爾係數的因素進行實證分析。首先，在第一節中經過分析比較，建立本碩士論文的實證模型，接著在第二節提出模型設定的相關變數假設，最後本碩士論文的資料來源在第三節中一併說明。本章整理的重點內容如下。

一、實證模型設定

本碩士論文採用追蹤資料進行實證分析，即同時考慮橫斷面資料 (cross-section data) 與時間序列 (time-series data) 資料，兼有兩者的優點，可以避免時間序列或橫斷面資料分析無法同時比較時間變動及個體差異，並避免忽略某些變數而產生估計偏誤或無效。同時為避免估計上的偏誤，在迴歸估計式上必須將各省市地區的個別特質效果予以考慮。又基於捕捉時間動態效果之考量，故採用能夠兼顧兩者的雙因子固定效果模型作為估計方法。

實證模型設定方面，以恩格爾係數為依變數，而以實質收入、依賴比、食品價格指數、房地產價格指數、衣著支出、醫療支出、教育支出、耐用品支出以及物價指數，為了探討城鎮與農村影響的因素是否不同，故各用一個模型來進行實證分析。

二、實證變數假設

對實證模型設定完畢之後，本碩士論文緊接著為實證變數進行假設，並預測最後的研究結果。實質收入與恩格爾係數之間預期為負的關係；人口依賴比與恩格爾係數預期為正的關係；食品價格指數與恩格爾係數的預期為正的關係；房地產價格指數與恩格爾係數預期為負的關係；衣著消費與恩格爾係數預期為負的關係；醫療支出與恩格爾係數預期為負的關係；預期醫療支出與恩格爾係數預期為負的關係；預期耐用消費品與恩格爾係數為負的關係；預期物價指數與恩格爾係數為負的關係。

三、資料來源

本碩士論文採用 1997 至 2007 年共 11 年的年資料，加上除西藏的 30 個省市資料，共 330 個樣本數進行實證研究。資料來源主要是中國大陸的國家統計局，以及《中國統計年鑑》，以各地的統計局、統計年鑑為補充。從 1997 年開始，中國大陸國家統計局均有將上述名目作為統計普查的項目，除了西藏有缺失以外，都相當完整。