

第三章 資料特性與模型建立

第一節 資料說明

利用台灣不動產成交行情公報（成交公報）的資料，該資料是由吉家網自民國 93 年起結合太平洋房屋、中信房屋、住商不動產與信義房屋等四大房仲業者的交易資料所提供的不動產價格資訊，該資料並提供不動產基本特徵屬性，剛好為本研究進行模型設定時所用，故選擇此資料庫之資料進行實證分析，選定成交價格作為市場價格的表徵。本研究所採用的樣本期間自 2004 年 Q1 至 2005 年 Q2 共計 18 個月，就可取得資料做實證分析，並取時間範圍較完整的資料以減少時間落差的問題。經過初步篩選剔除有缺失值之樣本及過於不合理的樣本（例如：坪數和房間數、衛浴數不成比例者或是樓層數過高等不合邏輯之樣本），得到共 3707 筆台北市大廈資料。

在資料樣本的處理方面，本研究採取剔除 5 分位數以下及 95 分位數以上的樣本，僅保留此二分位量之內的樣本。這樣的作法在進行橫斷面資料型態的實證分析時相當常見，許多的實證分析文獻亦採用這樣的方法。之後並依據文獻上的建議，針對離迴歸式較遠的樣本點作異常點刪除的動作，林秋瑾（1996，2004）經過實證顯示 DFFITS 法之表現較 R-student、Covratio、Cook'D 等方法為佳，故本研究採用 DFFITS 法做為異常點篩選的準則。經過樣本資料處理之後，共有台北市大廈資料 3270 筆。對資料樣本進行以上處理的主要目的，乃是為了剔除限定價格的個案，而單純研究正常價格的特徵價格模型；且由於原始資料數目堪稱充裕，因此將資料經過上述的處理，可以得到較整齊的統計資料。

根據國際估價準則（2005）指出，¹以市場為基礎的估價，通常是使用自市場取得之價格資料。而價格是指要價、出價或支付財貨勞務的總額，為一歷史事實。故本研究選定以成交公報中之成交價格作為市場價格的表徵。成交公報的特色在於其為真實的交易價格，且其為結合數大房屋仲介公司之交易資料所得之資訊，可降低不同仲介公司市場占有率不等之可能造成偏誤的情況。在住宅需求動向調查中發現，²民眾透過仲介購屋的比例提高，這也使得成交公報的資料更具代表性。

¹ 國際估價準則（International Valuation Standards, IVSC, 7th）。

² 此調查是由經建會委託財團法人國土規劃及不動產資訊中心所進行，一季公佈一次調查結果，是國內唯一有關購屋需求方面的調查。不同於許多由建商公布的供給面統計，台灣住宅需求調查主要調查國內有意購屋民眾的購屋傾向與偏好。

第二節 研究範圍與資料敘述統計

一、研究範圍及限制

本研究主要以台北市的住宅為研究對象。由於不同型態之住宅，其住宅屬性影響價格之意義也有所不同，故本研究擬獨立出台北市的大廈類型做討論。同時，主要是在大量估價的領域之下解決研究問題，研究限制為在有限資料變數下，就現有資料變數作建模選擇。

二、資料樣本之敘述統計

表 3-1 為本研究樣本的敘述統計，將樣本資料（共 3270 筆）依照變數分別列出平均數、標準差、最小值與最大值。由於本研究採用對成交總價進行分量研究的方式，對於成交總價的資料狀況需要進一步掌握，故將樣本資料中成交總價的各分量分配、偏態與峰度情形列於表 3-2。

表 3-1 樣本敘述統計

變數（單位）	平均數	標準差	最小值	最大值
成交總價（萬元）	1130.76	577.43	300.00	8955.00
坪數（坪）	41.12	15.21	11.91	173.94
房間數（間）	3.07	0.87	0.00	9.00
廳數（廳）	1.94	0.35	0.00	4.00
衛浴數（間）	1.80	0.56	0.00	5.00
屋齡（年）	17.48	8.07	0.00	38.70

表 3-2 大廈住宅成交總價的敘述統計

基本敘述 統計狀況 與分量值	偏態	峰度	10% 機率值	25% 機率值	50% 機率值	75% 機率值	90% 機率值
	2.78	22.30	580	748	1000	1350	1800

由資料的敘述統計以及大廈成交總價的各分量分配、偏態與峰度情況中可以看出大廈的坪數平均為 41.12 坪，標準差為 15.21 坪；成交總價介於 300 萬及 8955 萬元之間，平均為 1130.76 萬元，標準差為 577.43 萬元，由偏態與峰度係數可以看出資料型態為一高狹峰、右偏分配。

第三節 實證模型設定

在本研究的實證部份，採用 Rosen (1974) 的特徵價格模型，函數型態依文獻建議設定為半對數線性模型。以住宅成交總價的對數值做為被解釋變數。模型形式如下：

$$\ln(TP_i) = \alpha + \sum_{j=1}^k \beta_j X_{ij} + \phi_i \quad (9)$$

其中， TP_i 為第 i 筆資料的成交總價；

α 為第 i 筆資料的截距項；

β_j 為第 i 筆資料的第 j 個特徵之隱含價格，即各特徵屬性之迴歸係數；

X_{ij} 為第 i 筆資料的第 j 個特徵屬性（包括坪數、坪數平方、房間數、衛浴數、屋齡、屋齡平方、樓層虛擬變數、車位虛擬變數、區位虛擬變數以及時間虛擬變數）；

ϕ_i 為第 i 筆資料的誤差項。

本文的模型設定如上述，採 SAS 與 STATA 進行計量工作，並分別以普通最小平方迴歸以及分量迴歸進行估計。

第四節 變數選取與描述

實證模型中所採用的變數分為面積變數、內部屬性變數、屋齡變數、樓層變數、車位變數、區位變數以及時間變數等類型。在變數選擇方面，主要是配合文獻整理出對房價有顯著影響的各項不動產特徵，並搭配不動產估價時所須注意的勘估標的所在之區位狀況以及交易時間狀況進行模型中的變數選取。

在預期符號說明的部分，最小平方特徵價格模型的變數預期符號可以藉由文獻的整理做判斷。然而，分量迴歸模型的部分，因為相關研究文獻缺乏，故以過去的觀察以及經驗對變數的預期符號做判斷及解釋。

一、應變數：大廈成交總價取自然對數

傳統上不動產交易即是以總價進行，故模型中價格變數選取如同一般文獻設定為不動產總價。模型的函數型態設定綜合文獻的結論，採用半對數線性模型，故將成交總價取自然對數。

房地產的估價模型中，有許多變數是屬於屬質變數，代表不動產本身的部分特徵或異質性，此亦為國內外文獻採用特徵價格方程式來估計不動產價格的原因之一。不動產與其他財貨不同，具有不可分割的特性，一旦購買房地產，消費者不僅僅購買該房地產本身，也同時購買其實質屬性、周圍環境、公共設施、鄰里關係等。因此，在購屋時，消費者購買的是整個不動產，而無法分割不動產特徵而進行交易。此外，住宅總價是由面積乘以單價所決定，由於面積的認定基準不同，因此利用單價模型進行實證研究會產生效果差異。傳統上不動產的交易是以總價進行，且考慮不動產的不可分割性，價格變數選取如同一般文獻設定為不動產總價。

二、自變數：

（一）面積變數：坪數、坪數平方

指建物之樓地板面積。在區隔出不同的消費者需求的同時，住宅成交總價受面積單價的影響極大。在其他變數不變之下，住宅面積越大，總價越高，故預期符號為正向變動。由於坪數可能存在報酬遞減現象，故加入平方項，對於住宅價格的影響預期符號為負。

在坪數可能存在報酬遞減的預期之下，分量迴歸模型中，預期低價位時坪數對於大廈價格的影響為正；偏高價位之大廈，坪數對於大廈價格的

影響可能受報酬遞減現象而呈現正向但遞減的關係。由於低價位大廈相對而言坪數較小，小面積住宅增加一坪對於使用者的效用較大，而高總價大面積的大廈增加一坪對於使用者的效用可能漸弱，故預期偏高價位的大廈將存在坪數的報酬遞減現象。

(二) 內部屬性變數：房間數、衛浴數

根據國內住宅特徵價格之相關文獻，住宅房間數、衛浴數越多，將使住宅總價越高，預期符號為正向變動。Sirmans, Macpherson and Zietz (2005) 的研究針對過去國外使用特徵價格法估計住宅價格的文獻，整理出將房間數與衛浴數納入模型中的文獻均達 40 篇，並得到房間數與衛浴數對於住宅價格的影響均為正向變動的結論。在少數文獻中，房間數對於價格的影響為負向變動；而幾乎全部的文獻均顯示衛浴數對於價格的影響為正向變動。

在控制其他變數不變之下，增加房間數可能導致房間的面積變小而存在使用者的偏好問題，此現象預期對於高總價、大面積大廈的使用者而言，將出現較明顯之房間數對於大廈價格影響力的遞減效果。是以，分量迴歸模型中，預期低價位大廈之房間數對於大廈價格的影響為正；偏高價位而坪數較大之大廈，預期房間數對於大廈價格的影響為正向而遞減的效果。在衛浴數的部分，由建造成本較高且使用者需求的觀點，預期對於大廈價格的影響在各價位的大廈均為正。

(三) 屋齡變數：屋齡、屋齡平方

住宅的屋齡越高越易產生折舊，故屋齡對於住宅價格的影響預期符號為負。然而，隨著屋齡增加可能存在維護更新以保存不動產價值的行為；此外，即使有折舊的現象產生，仍有地價支撐著成交價格的可能。由於一般房產的價值中，房屋價值僅佔三成而土地價值佔了七成，隨著不動產存在時間的拉長，將出現房屋折舊、土地增值的現象。故加入屋齡平方項，對於價格影響的預期符號為正。

由於較高價位的大廈，是因為具有面積大、區位佳的條件，相對的使用者必定為具相當經濟能力並對於居住品質有相當要求者。因此推論，高價位大廈的使用者對於住宅的管理維護將更要求，故於分量迴歸模型中，預期低價位大廈屋齡對於大廈價格的影響較大；偏高價位之大廈，屋齡對於大廈價格的影響較小。

(四) 樓層虛擬變數：是否為一樓、是否為頂樓

對於一樓及頂樓的偏好乃國內不動產之特殊性，變數反映樓層的差別是否影響成交價。一樓的住宅因為可及性佳，價格通常高於其他樓層；而頂樓因為景觀與隱私性較佳，亦會有較高的售價。故對於價格的影響預期符號皆為正。

由於有能力負擔高價位住宅的使用者對於可及性、住宅景觀以及隱私性等居住品質的要求較高。是以，於分量迴歸模型中，預期偏高價位之大廈，使用者對於一樓及頂樓更具偏好，因而特徵對於大廈價格的影響較大。

(五) 車位虛擬變數

在其他條件不變下，含有停車位的住宅預期價格會較高。以無車位為基準組，相對於無車位，含有停車位的住宅對於價格影響的預期符號為正。由於預期較具經濟能力者，對於住宅包含停車位的條件將更要求。是以，分量迴歸模型中，預期越高價位之大廈，特徵對於大廈價格的影響越大。

(六) 區位虛擬變數

由於相關研究文獻指出住宅次市場切割的重要性，³然而，由於本研究使用分量迴歸模型進行估計，在避免樣本選擇偏誤的前提下，將不進行次市場的切割，而是利用虛擬變數的方式討論次市場之間的關係，以劃分次市場的方式區隔區位對於價格的影響。

以郵遞區號做為次市場劃分的基礎，並且以「影響住宅價格的因素同樣會影響次市場劃分」的原則選取變數，進行方式以階層集群分析法劃分次市場。以歐基里得直線距離平方做為測量區間，同時以 Ward's method 為集群。選取變數根據文獻整理影響不動產價格的因素以及可取得的資料，整理出的變數包括：每坪單價、建物面積、屋齡。

次市場劃分的結果，舊市區中心包括中山、中正、松山等三個行政區；新市區中心包括信義、大安等兩個行政區；東郊包括內湖、南港等兩個行政區；北郊包括士林、北投等兩個行政區；早期開發與南郊包括萬華、文山、大同等三個行政區。以早期開發與南郊區為基底，主要是由於萬華區、大同區屬於台北市較早發展的地帶而現況呈現較為老舊市區的樣貌，而文山區較位於台北市的外圍南郊，以階層集群分析法進行次市場劃分的結果

³ Bourassa, Hoesli and Peng (2003) 研究顯示在大量估價的目的下，次市場的定義必要性會使住宅特徵的預測準確度最佳化。

將上列三個行政區劃設為同一次市場。以價位較低的早期開發與南郊區作為基準組，故位於其他各次市場其對於價格影響的預期符號皆為正，在最小平方迴歸以及分量迴歸模型中均有同樣的預期。

(七) 時間虛擬變數

此變數表示資料進行交易的期間，由於不動產估價中，對於估價期日的部分相當重視，因而放入此變數作為控制變數。從台灣房地產景氣動向季報發現，資料的交易期間房地產景氣呈現些微下滑，但景氣均為綠燈。在景氣波動屬於緩和的走勢下，為配合估價中交易期間對於價格的影響，研究經測試設定半年放一個時間虛擬變數。共放 2 個虛擬變數(2004 年 7~12 月、2005 年 1~6 月)，以交易於 2004 年 1~6 月間進行的樣本為基準組。