

# 穩健性住宅租金模式之探討－異常點之分析\*

## Robust Housing Rent Modeling Outlier Analysis\*

林秋瑾\*\*

Vickey C.C. Lin\*\*

### 摘要

國內外特徵住宅價格法 (Hedonic Housing-Price) 運用於房價及房租的研究文獻相當多，但在諸多研究中可以發現到大部分研究忽視將所選取之變數資料中的異常點修正，而所謂異常點是指資料本身之值與平均值之間差異甚多的點，此異常點的存在足以影響整體統計資料的解釋能力，進而影響以特徵住宅價格法估計求得之房地產價格及房屋租金的可信度，本文乃基於前述之理由，以房屋租金實證方式來探討特徵住宅價格法中異常點判定及修正的重要性。本文以實證方式來探討異常點判定與修正之重要性，因此採用「行政院主計處台灣地區勞動力調查住宅專案調查」之問卷資料為分析對象，並將資料透過統計電腦軟體 SAS 系統而轉換為衡量異常點的指標加以探討，並以單一異常點，複異常點及多重異常點的衡量標準判定資料中之異常點，接著將所選出之異常點應用穩健性方法加以修正，而運用 SAS 之統計分析方式來比較異常點修正前後整體統計資料的變數對房屋租金解釋能力，修正異常點後整體統計資料變數解釋能力均增加，且有些變數由不顯著的情形變為顯著。因此，我們認為異常點訊息掌握是不容忽視，且有其修正之必要性，本文最後選取異常點修正的最佳模式，供擬研究房價及房租者參考。

### ABSTRACT

We established the rational real estate evaluation models.

First, This study was based on the Hedonic price theory to build an explanatory model of residential properties rent in terms of the observed traits of each housing or land unit in an urban area.

Second, this study robustified the strategies for the housing rent model in the presence outlier.

Third, this explored the comparison of several internal analysis of robustness of least square for give samples are pointed out.

Finally, The methodology using in this study was illustrated by building the robust Hedonic housing rent model in the Taipei Metropolitan Area.

\* 本文乃接受行政院國科會補助研究 (NSC-83-031-H-004-049)，特以至謝。

\*\* 國立政治大學地政系副教授



## 一、前言

國內外特徵住宅價格法 (Hedonic Housing-Price) 運用於房價及房租的研究文獻相當多；國外如 Rosen (1974)、Belsley D. A. ; Huh E. ; Welsch R. E. (1980) , Daniel & Jarier (1984) , Graves、Murdoch、Thayer & Waldman (1988) 等；國內如辜炳珍及劉瑞文 (1979) 、林元興 (1979、1988) 、林祖嘉 (1990、1993) 、張梅英 (1992) 張金鶲及范垂爐 (1993) 、張麗姬 (1994) 等。但在諸多研究中可以發現到大部分之研究忽視將所選取之變數資料中的異常點修正，而所謂異常點是指資料本身之值與平均值之間差異甚多的點，此異常點的存在足以影響整體統計資料的解釋能力，本文乃基於前述之理由，將以實證方式來探討異常點的重要性。

上述內容中曾說明到本文將以實證方式來探討異常點判定與修正的重要性，本文所採用的資料是1988及1989二個年度「行政院主計處台灣地區勞動力調查住宅專案調查」之問卷資料，以台北縣轄市之三重市、板橋市、新莊市、中和市、永和市租賃之住宅為分析對象，二個年度合記的有效樣本達六百多筆以上，此樣本數在統計理論上應屬大樣本資料，當資料龐大、變數眾多，對於異常點之修正更不容忽視（註1）。基於資料在大樣本下，該如何客觀的修正差異性大的觀察值，引發研究本文之動機。

本文之研究方法為將資料透過統計電腦軟體 SAS 系統而轉換為衡量異常點的資料，並擬以單一異常點，複異常點及多重異常點的衡量標準判定資料中之異常點，接著擬將所選出之異常點加以修正，並運用 SAS 之統計分析方式比較異常點修正前後整體統計資料變數解釋能力，若修正異常點後整體統計資料變數解釋能力叫異常點修正前高，則我們認為異常點的存在是不容忽視且其修正是重要的，本文最後擬選取異常點修正的最佳模式，供擬研究房價及房租者參考。

本文第二節為模式之建立與資料分析，第三節為異常點的判斷，第四節為異常點修正之重要性，第五節為最適特徵住宅租金模式的選取，第六節為結論。

## 二、模式之建立與資料分析

本文中我們將以都市區域中住宅單位的特徵、環境及可及性等因素考量來建立一個住宅租金的解釋模型，試圖透過此模型的建立來分析異常點之問題。此方法即為都市經濟學中之特徵價格模式，而 Griliches (1971), Ball (1973), Quigley (1979) 等人曾於文獻中討論過（註2）。本文將依此建立住宅租金的理論模型；並透過資料分析而建立本研究之實證模式、由實證模式的建立來了解異常點之問題。

### (一) 模式之建立



假設下例模式為特徵住宅租金模式：

$$Y_i = \alpha + \sum_{j=1}^r \sum_{j=1}^{p_i} \beta_{ij} X_{ij} + \varepsilon_i \quad \dots \quad (2-1-1)$$

$$j=1, 2, 3, \dots, p_i,$$

$$i=1, 2, \dots, r$$

而

$y_i$  = 第  $i$  個觀察值 (住宅租金)

$X_{ij}$  = 第  $i$  個觀察值之第  $j$  個 Nonstochastic 獨立變數 (住宅屬性)

$\varepsilon_i$  = 第  $i$  個觀察值之干擾誤差

在模式 (2-1-1) 中，假設  $\varepsilon_i$  為皆獨立且具相同的分配之隨機干擾項 (Random Disturbances)，平均數為 0，變異數為  $\sigma^2$ 。本文所討論的問題是當  $\sigma^2$  未知時，如何估計  $\alpha$  與  $\beta$  的問題。當  $\varepsilon_i$  為常態分配時，許多方法可以用來估計住宅屬性之隱含租金  $\alpha$  與  $\beta$ 。在文獻中，這些估計式分別是：普通最小平方估計式(OLS)，一般化最小平方估計式(GLS)，最大概似估計式(MLE)。經由蒙地卡羅實證，當  $\varepsilon_i$  為常態分配時，OLS、MLE、GLS 之  $\alpha$  與  $\beta$  估計式同具不偏性但 OLS 之  $\alpha$  與  $\beta$  估計式較具有效性，亦即符合最佳線性不偏最小變異之估計式(BLUE)。當  $\varepsilon_i$  為非常態分配時，且當  $\sigma^2$  未知時，將如何估計  $\alpha$  與  $\beta$  的問題！Tan 與 Tabatabai(1988) 與 Tiku、Tan 與 Balakrishnan (1986) 的研究顯示，最小平方法之估計式是基於常態假設的隨機干擾項形成的；在許多情形下當隨機干擾項母體不是常態時，建立另一個穩健性估計方法，此方法比常態估計法更有效；而當母體是常態時，此方法幾乎與常態估計法一樣有效。本研究藉由特徵住宅租金模式探討非常態干擾項母體時之問題。

## (二) 資料分析

本文所採用的資料是 1988 及 1989 二個年度「行政院主計處台灣地區勞動力調查住宅專案調查」之問卷資料，以台北縣之三重市、板橋市、新莊市、中和市、永和市租賃之住宅為分析對象，二個年度合記的有效樣本達六百多筆以上，問卷資料含七十個變數，本文基於資料限制，將研究範圍界定於台北縣地區，並從問卷資料中選取二十三個變數，且擬將資料變數分為住宅本身特性、住宅附近公共設施項目至住宅距離、住宅環境品質及住宅區位四類而說明如下：

### (A) 住宅本身特性：(註3)

1. ROOM1：住宅內的房間數。
2. ROOM2：住宅內的大廳數，包含客廳及餐廳。
3. FLOOR：住宅內的樓地板面積，住宅內總面積之坪數，但不包括陽台面積。
4. USEA：住宅用途採用虛擬變數，若虛擬變數為 1 表示為住家專用，為 0 則表示其他，用途如住家間店舖、住家兼工廠、住家兼其他用途等。



5. HSTRU A：住宅結構採用虛擬變數，若虛擬變數為1表示為鋼筋混泥土造，為0 則表示其他，如加強磚造，磚造，木造及石造等。
6. HAGE：屋齡，指住宅建造竣工至主計處進行住宅狀況調查期間的年齡。
7. HSTYPE A：住宅類型A，為虛擬變數，若HSTYPE A = 0 則為五樓以下之公寓；若HSTYPE A = 1，則為其他，如中式或西式（日式）獨院式之房屋或雙拼，連棟式之房屋或六樓以上之公寓（大廈）。
8. HSTYPE B：住宅類型B，為虛擬變數，若HSTYPE B = 0 則為五樓以下之公寓，或六樓以上之公寓（大廈），若HSTYPE B = 1，則為其他，如中式或西式（日式）獨院式之房屋或雙拼，連棟式之房屋。
9. HSTYPE C：住宅類型C，為虛擬變數，若HSTYPE C = 0 則為中式或西式（日式）獨院式之房屋，若HSTYPE C = 1，則為其他，如五樓以下之公寓，或六樓以上之公寓（大廈）或雙拼連棟式之房屋。
10. RENT：房租，指每月所應繳的租金，此處依問卷資料將房租分為九個層級的內容說明如下：
 

(1) 1,000元以下	(4) 3,000—3,999元
(2) 1,000—1,999元	(5) 4,000—4,999元
(3) 2,000—2,999元	(6) 5,000—5,999元
(7) 6,000—6,999元	(8) 7,000—7,999元
(9) 8,000—8,999元以上	

(B) 住宅附近公共設施項目至住宅距離（以公里為單位）：

1. DPMS：距離住宅地點內的國小距離。
2. DHGS：距離住宅地點內的國中距離。
3. DMKT：距離住宅地點最近的菜市場距離。
4. DRES：距離住宅地點休閒場所的距離，此處之休閒場所係指公園，綠地，堤防用地，運動場地等。公園綠地不包括自宅庭院，而以主觀認定可供休閒，散步遊樂之場所；運動場則包括對外開放之公私立學校。

(C) 住宅環境品質指標，均採用虛擬變數（Dummy Variable）：

1. TRAFFA：居住地點的交通便利程度，若TRAFFA = 0，則交通便利；若TRAFFA = 1，則交通狀況不佳。
2. NOISEA：居住地點受到噪音之干擾程度，若NOISEA = 0，則沒有噪音之干擾；若NOISEA = 1，則有噪音干擾。
3. AIRA：居住地點空氣污濁異味之情形，若AIRA = 0，則沒有空氣污染；若AIRA = 1，則為空氣污染。
4. DROWNA：住宅及其四周最近五年內是否曾遭淹水，若DROWNA = 0，則表示近



- 五年內沒有淹過水；若 DROWN A = 1，則為曾有淹過水。
5. W A T E R A：居住地點之飲水衛生情形，包括飲水是否有異味，水色不清或有異物，若 W A T E R A = 0，則為飲水衛生情況良好；若 W A T E R A = 1，則為水質不良。
  6. S A T I S A：對住宅狀況的整體滿意程度。若 S A T I S A = 0，則表示滿意程度高；若 S A T I S A = 1，則表示滿意程度低。

(D) 住宅區位，均採虛擬變數 (Dummy Variable)：

1. D1：代表三重市；若 D1 = 1，則為三重市，若 D1 = 0，則表示其他地區。
2. D2：代表板橋市；若 D2 = 1，則為板橋市，若 D2 = 0，則表示其他地區。
3. D3：代表新莊市；若 D3 = 1，則為新莊市，若 D3 = 0，則表示其他地區。
4. D4：代表永和市；若 D4 = 1，則為永和市，若 D4 = 0，則表示其他地區。

將上述四類共二十三個變數，依各分類分別做逐步迴歸 (Stepwise Regression) 選取較佳之變數，以避免變數間有線性重合的存在，經過逐步迴歸 (Stepwise Regression) 的選取變數及經驗判斷，除了因變數 RENT 外，在第一類中共選取五個變數—FLOOR、HAGE、HUSEA、HSTRU A、HSTYPEB；在第二類中亦選取四個變數—DPMS、DHGS、DMKT、DRES，捨醫院及郵局變數不用因其與選擇租屋關聯較小；而在第三類中則選取五個變數—TRAFFA、NOISEA、AIRA、DROWNA、WATERA，滿意度變數因其樣本回答代表性問題影響迴歸估計參數因而捨棄不用。（註4）

本研究避免線性重合的存在，將第二類選取變數依典型相關分析就樣本變數合併為 V1【V1 = (- 0.0138 \* DPMS) + (0.3691 \* DHGS) + (0.3895 \* DMKT) + (0.9524 \* DRES)】，且將第三類選取變數合併為 V2【V2 = (0.0006 \* WATERA) + (- 0.0336 \* AIRA) + (0.8147 \* TRAFFA) + (- 0.5096 \* DROWNA) + (- 0.1186 \* NOISEA)】（註5），並加入第四類之區位變數及時間因素分別以 D 及 T(T1：代表時間；若 T1 = 1，則為 T = 1988年，若 T1 = 0，則表示 T = 1989年)。選取模式之過程中、初步以三個模式（模式一、模式二、模式三）為基礎模式（註6）；三個模式所採用之變數分別合計為十二、十一及九個變數。這些模式之變數的內容及其樣本數，平均數及標準差的資料由表 2.1 說明之。

由表 2.1 可明確的看出各模式所使用的變數個數，而整體的樣本資料達 614 筆，至於 1988 及 1989 年樣本資料主要分布的那些地區及該地區所選取的筆數為何則列表說明如表 2.2。



表 2·1 台北縣—五縣轄市：房租及房租變數資料分析  
樣本數：614

模式一		模式二		模式三			
變數	平均數	標準差	變數	標準差	變數	平均數	標準差
RENT	5.06	1.630	RENT	5.06	1.630	RENT	5.06
FLOOR	23.08	7.136	FLOOR	23.08	7.136	FLOOR	23.08
HAGE	12.65	6.535	HAGE	12.65	6.535	HAGE	12.65
HSTRU A	0.64	0.481				HSTRU A	0.64
HTYPEB	0.25	0.434	HTYPEB	0.25	0.434	ROOM1	2.45
V1	3.29	1.458	V1	3.29	1.458	ROOM2	1.18
V2	-0.42	0.328	V2	-0.42	0.328	HTYPA	0.26
D1	0.23	0.419	D1	0.23	0.419	DRES	1.91
D2	0.30	0.458	D2	0.30	0.458		
D3	0.12	0.328	D3	0.12	0.328		
D4	0.12	0.330	D4	0.12	0.330	D4	0.12
T1	0.52	0.500	T1	0.52	0.500	T1	0.52

資料來源：本研究整理

表 2·2 1988及1989年樣本資料分布地區及筆數

城市名稱	1988 年	1989 年
三重市	83	56
板橋市	88	95
新莊市	32	43
永和市	39	37
中和市	79	62
小計	321	293
合計		614

資料來源：本研究整理



### 三、異常點的判斷

由常態機率圖（NP-PLOT）之圖3.1及圖3.2中看出確有遠離四十五度角線之樣本點擬似異常點存在。為確定那些樣本點是屬於異常點，本文運用SAS系統而轉換為衡量異常點的資料，並運用判定異常點標準公式，確定異常點，這些判斷異常點的標準計算公式分述如下：（註7）

- (1) R S T U D E N T (R) — 以數值之絕對值大於 2 為異常點。
- (2) H A T D I A G . H . (H) — 以數值之絕對值大於  $(2P/N)$  為異常點。
- (3) C O V R A T I O (C) — 以數值之絕對值大於  $(1+3P/N)$  或小於絕對值  $(1-3P/N)$  為異常點。
- (4) D F F I T S (D) — 以數值之絕對值大於  $(2 * \sqrt{P/N})$  為異常點。
- (5) D F B E T A S (B) — 以數值之絕對值大於  $(2 / (\sqrt{N}))$  為異常點。
- (6) COOK'S (k) — 以數值之絕對值大於 1.00 為異常點。

上述的六個公式是一般常用衡量異常點的標準計算公式。式中N表示樣本數，P為自變數個數， $\sqrt{P}$ 表示為平方根，若以 SAS 系統而轉換為衡量異常點的資料之絕對值大於異常點的標準計算公式之值，則稱此點為異常點，現擬用上述之公式，計算本文所採用資料的異常點衡量標準值，這些標準值如表 3 所示。

表 3 1988年、1989年之異常點衡量標準值

衡量值 模式	N	P	R	H	C	D	B	k
模式一	614	12	±2	±0.039	$\geq 1.059$ $\leq 0.941$	±0.280	±0.081	±1.00
模式二	614	11	±2	±0.036	$\geq 1.054$ $\leq 0.946$	±0.268	±0.081	±1.00
模式三	614	9	±2	±0.029	$\geq 1.044$ $\leq 0.956$	±0.242	±0.081	±1.00

資料來源：本研究整理

應用表 3 之衡量標準值，分別採用單一異常點，複異常點及多重異常點的衡量標準判定資料中之異常點，以作為修正異常點的依據。



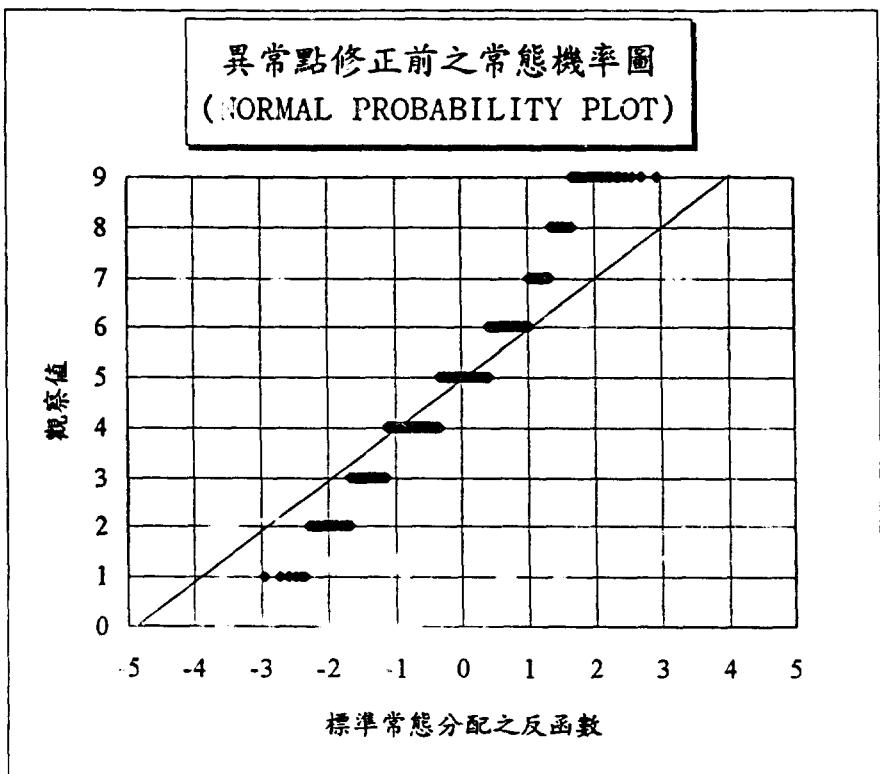


圖3.1 異常點修正前之常態機率圖

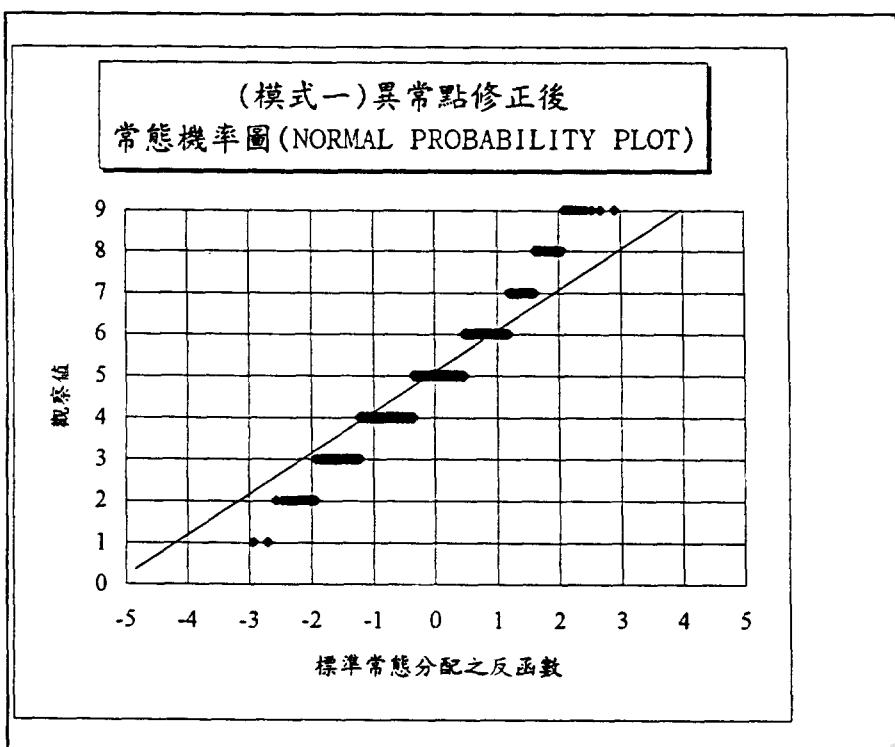
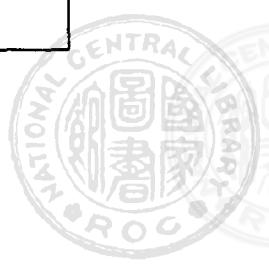


圖3.2 模式一異常點修正後之常態機率圖



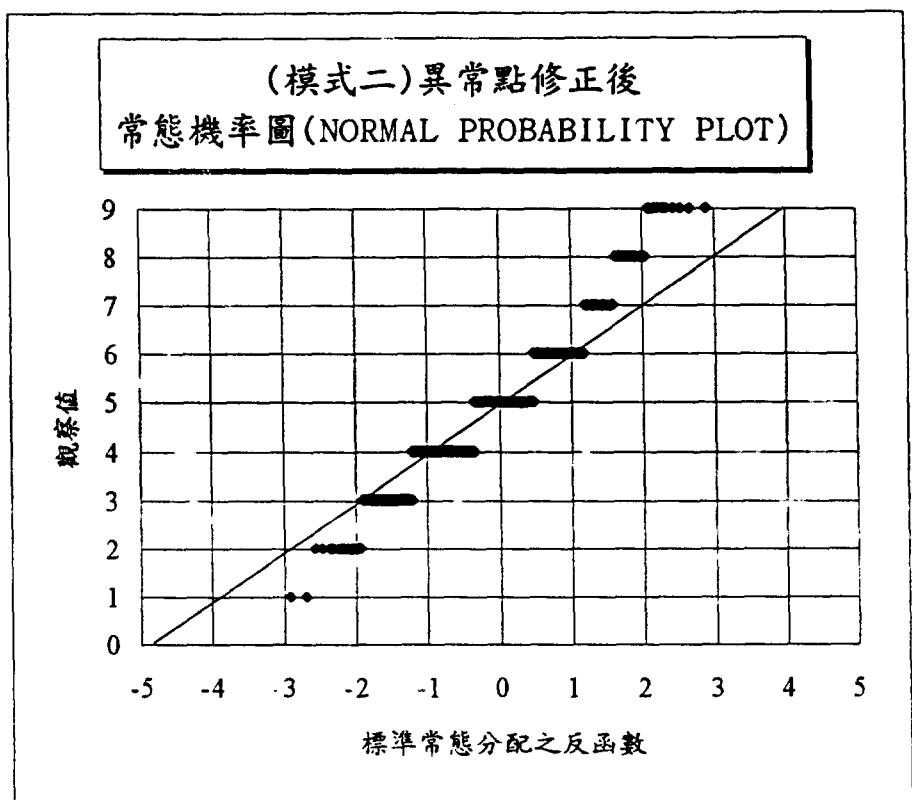


圖3.3 模式二異常點修正後之常態機率圖

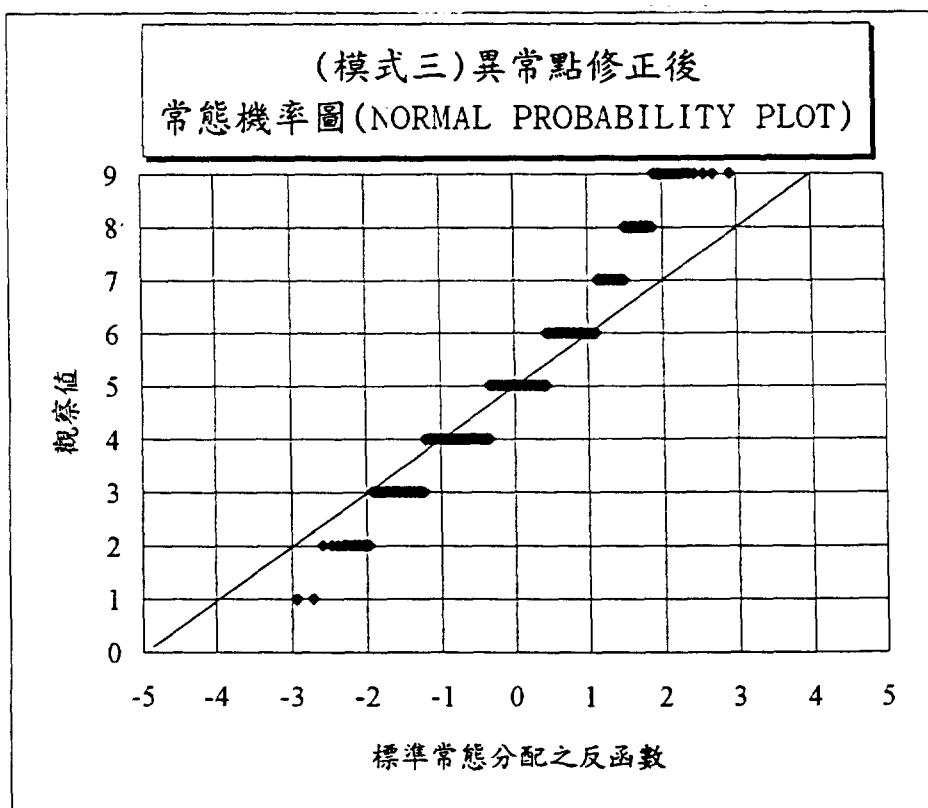


圖3.4 模式三異常點修正後之常態機率圖



表 3-1 樣本資料分布地區及筆數

單一異常點判斷標準

異常點判斷標準	R	H	D	COV
城市名稱				
三重市	12	3	15	8
板橋市	13	2	14	12
新莊市	7	8	12	8
永和市	5	4	8	3
中和市	5	1	6	6
合計	42	18	55	37

表 3-2 樣本資料分佈時間及筆數

單一異常點判斷標準

異常點判斷標準	R	H	D	COV
時間				
1988	22	13	34	15
1989	20	5	21	22
合計	42	18	55	37

表 3-3 樣本資料分布地區及筆數

複異常點判斷標準

異常點判斷標準	R/D	R/COV	R/H	D/COV	D/H
城市名稱					
三重市	11	7	2	7	2
板橋市	9	11	0	8	0
新莊市	7	3	0	3	0
永和市	5	3	2	3	2
中和市	3	5	0	4	1
合計	35	29	4	25	5

資料來源：本研究整理



表 3-4 樣本資料分佈時間及筆數

複異常點判斷標準

異常點判斷標準	R/D	R/COV	R/H	L COV	D/H
時間					
1988	20	16	2	14	4
1989	15	13	2	11	1
合計	35	29	4	25	5

表 3-5 樣本資料分布地區及筆數

多重異常點判斷標準

異常點判斷標準	R/C/D	R/H/D	R/H/COV	D/COV/H
城市名稱				
三重市	7	2	2	2
板橋市	8	0	0	0
新莊市	3	0	0	0
永和市	3	2	1	1
中和市	3	0	0	1
合計	24	4	3	4

表 3-6 樣本資料分佈時間及筆數

多重異常點判斷標準

異常點判斷標準	R/C/D	R/H/D	R/H/COV	D/COV/H
時間				
1988	14	2	2	3
1989	10	2	1	1
合計	24	4	3	4

資料來源：本研究整理



表 4 樣本資料之統計值(排序後)表續

標準	D	D	R	R	R	COV	COV	COV	D
	模式二	模式一	模式三	模式一	模式二	模式三	模式二	模式一	模式三
統計值									
N	558	559	576	572	571	582	578	577	556
P	11	12	9	12	11	9	11	12	9
O	56	55	38	42	43	32	36	37	58
R-SQR	0.443	0.438	0.432	0.435	0.428	0.393	0.389	0.388	0.381
Adj-R	0.432	0.426	0.423	0.423	0.417	0.384	0.377	0.375	0.371
DW	2.012	1.971	1.874	1.950	1.985	1.887	1.871	1.903	1.947

標準	H	原始資料	K	H	H	原始資料	K	原始資料	K
	模式三	模式三	模式三	模式一	模式二	模式二	模式二	模式一	模式一
統計值									
N	564	614	614	596	597	614	614	614	614
P	9	9	9	12	11	11	11	12	12
O	40	0	0	18	17	0	0	0	0
R-SQR	0.322	0.297	0.2966	0.300	0.293	0.286	0.2856	0.286	0.286
Adj-R	0.311	0.286	0.2861	0.283	0.280	0.273	0.273	0.272	0.272
DW	1.898	1.900	1.9	1.880	1.872	1.912	1.912	1.915	1.915

標準	R/D	R/D	R/COV	R/COV	R/D	D/COV	D/COV	D/COV	R/H
	模式二	模式一	模式三	模式二	模式三	模式二	模式一	模式三	模式三
統計值									
N	577	579	587	585	586	588	589	593	605
P	11	12	9	11	9	11	12	9	9
O	37	35	27	29	28	26	25	11	9
R-SQR	0.4149	0.4109	0.3948	0.3937	0.3894	0.388	0.3749	0.3682	0.3473
Adj-R	0.4035	0.3984	0.3854	0.3821	0.3799	0.3763	0.3664	0.3584	0.3375
DW	1.951	2.05	1.918	1.91	1.875	1.903	1.925	1.869	1.89

註：N 表示樣本數，P 為自變數個數，O 為異常點個數



表 4 樣本資料之統計值(排序後)續完

標準	D/H	R/COV	COV/H	R/H	R/H	D/H	D/H	COV/H	COV/H
	模式三	模式一	模式三	模式二	模式一	模式一	模式二	模式二	模式一
統計值									
N	589	585	590	610	610	609	608	604	603
P	9	12	9	11	12	12	11	11	12
O	25	29	24	4	4	5	6	10	11
R-SQR	0.3395	0.3919	0.3201	0.3049	0.3055	0.3052	0.3003	0.2941	0.2963
Adj-R	0.3292	0.3248	0.3095	0.2921	0.2916	0.2912	0.2874	0.281	0.2792
DW	1.887	1.897	1.869	1.889	1.927	1.889	1.891	1.873	1.878

標準	R/D/COV	R/D/COV	R/D/COV	D/R/H	D/H/COV	R/COV/H	D/R/H/COV	D/R/H
	模式二	模式一	模式三	模式三	模式三	模式三	模式三	模式二
統計值								
N	589	590	596	605	605	609	605	610
P	11	12	9	9	9	9	9	11
O	25	24	18	9	9	5	9	4
R-SQR	0.3878	0.3792	0.3642	0.3473	0.3294	0.3263	0.3263	0.3049
Adj-R	0.3762	0.3663	0.3545	0.3375	0.3193	0.3161	0.3161	0.2921
DW	1.923	1.932	1.911	1.89	1.87	1.888	1.888	1.889

標準	D/R/H	R/COV/H	D/R/H/COV	D/H/COV	R/COV/H	D/R/H/COV	D/H/COV
	模式一	模式二	模式二	模式二	模式一	模式一	模式一
統計值							
N	610	611	610	610	611	610	610
P	12	11	11	11	12	12	12
O	4	3	4	4	3	4	4
R-SQR	0.3053	0.2966	0.2966	0.2963	0.2973	0.2973	0.2971
Adj-R	0.2916	0.2837	0.2837	0.2834	0.2832	0.2832	0.2829
DW	1.891	1.886	1.886	1.887	1.888	1.888	1.883

註：N 表示樣本數，P 為自變數個數，O 為異常點個數



表 5-1 模式一樣本原始資料之模式及單一異常點衡量標準之DFFITS(D)  
其修正異常點後得之模式 — 統計資料

變數	原始資料			(D)		
	參數值	t 值	t 之 P 值	參數值	t 值	t 之 P 值
INTERCEP	5.47 *	17.55	0.0001	4.69 *	17.98	0.0001
FLOOR	0.07 *	8.42	0.0001	0.10 *	12.87	0.0001
HAGE	-0.03 *	-3.17	0.0016	-0.03 *	-3.29	0.0011
HUSEA	-1.34 *	-10.15	0.0001	-1.37 *	-12.19	0.0001
HSTRU A	0.09	0.60	0.5471	0.04	0.38	0.7046
HTYPEA	-0.33 *	-2.11	0.0349	-0.22	-1.80	0.0729
V1	0.07	1.10	0.2716	0.07	1.32	0.1887
V2	0.01	0.10	0.9216	-0.06	-0.49	0.6248
T1	-0.87 *	-6.23	0.0001	-0.88 *	-7.95	0.0001
D1	-0.20	-1.10	0.2710	0.05	0.32	0.7529
D2	-0.34 *	-2.11	0.0352	-0.21	-1.72	0.0867
D3	-0.36	-1.77	0.0770	-0.55 *	-3.34	0.0009
D4	0.24	1.17	0.2409	0.55 *	3.42	0.0007
	Adj R-sq	0.27		Adj R-sq	0.43	
	樣本數	614		樣本數	559	

\* 在顯著性水準為 0.05 之下顯著變數

資料來源：本研究整理



表 5-2 模式二樣本原始資料之模式及單一異常點衡量標準之DFFITS(D)  
其修正異常點後得之模式 — 統計資料

變 數	原始資料			(D)			
	參數值	t 值	t 之 P 值	參數值	t 值	t 之 P 值	
INTERCEP	5.49 *	17.85	0.0001	4.91 *	19.31	0.0001	
FLOOR	0.07 *	8.41	0.0001	0.09 *	12.28	0.0001	
HAGE	-0.03 *	-3.15	0.0017	-0.03 *	-3.55	0.0004	
HUSEA	-1.33 *	-10.12	0.0001	-1.37 *	-12.12	0.0001	
HTYPEB	-0.30 *	-1.84	0.0663	-0.25 *	-1.97	0.0492	
V1	0.07	1.04	0.2972	0.08	1.53	0.1142	
V2	0.02	0.13	0.8931	-0.08	-0.72	0.4771	
T1	-0.82 *	-7.00	0.0001	-0.91 *	-9.87	0.0001	
D1	-0.19	-1.05	0.2950	0.002	0.01	0.9914	
D2	-0.34 *	-2.15	0.0319	-0.21	-1.76	0.0906	
D3	-0.36	-1.77	0.0770	-0.52 *	-3.13	0.0018	
D4	0.24	1.16	0.2456	0.52 *	3.27	0.0012	
	Adj R-sq 0.28			Adj R-sq 0.44			
	樣本數 614			樣本數 558			

\* 在顯著性水準為 0.05 之下顯著變數

資料來源：本研究整理



表 5-3 模式三樣本原始資料之模式及單一異常點衡量標準之RSTUDENT(R)  
其修正異常點後得之模式 — 統計資料

變 數	原始資料			(R)			
	參數值	t 值	t 之 P 值	參數值	t 值	t 之 P 值	
INTERCEP	5.29 *	18.22	0.0001	4.69 *	19.16	0.0001	
FLOOR	0.10 *	8.73	0.0001	0.11 *	11.58	0.0001	
HAGE	-0.03 *	-2.96	0.0032	-0.03 *	-2.99	0.0029	
HUSEA	-1.19 *	-8.79	0.0001	-1.30 *	-11.33	0.0001	
ROOM1	-0.21 *	-2.18	0.0293	-0.13 *	-1.59	0.1126	
ROOM2	-0.34 *	-2.71	0.0070	-0.23	-2.16	0.0310	
HTYPEA	-0.40 *	-2.59	0.0098	-0.33 *	-2.56	0.0108	
DRES	0.10 *	2.08	0.0378	0.09 *	2.32	0.0205	
T1	-0.84 *	-7.19	0.0001	-0.87 *	-9.01	0.0001	
D4	0.47 *	2.76	0.0060	0.80 *	5.65	0.0001	
	Adj R-sq	0.29		Adj R-sq	0.42		
	樣本數	614		樣本數	576		

\* 在顯著性水準為 0.05 之下顯著變數

資料來源：本研究整理



### (一) 單一異常點的衡量標準：

此處選取較長用的四個單一異常點的衡量標準，此四個異常點的衡量標準為 R、H、C、D，現擬用上述四個單一異常點的衡量標準以模式一為例、分別找出樣本資料之異常點。（註8）

1. R S T U D E N T (R) 如表3.1所示依 R S T U D E N T 之單一異常點的衡量標準，從表3.1中可以找出樣本資料中有42個異常點的存在，並由此表可了解異常點形成之區位分布情形。從表3.2中可以發現到1988年樣本資料中有22個異常點的存在，而在1989年樣本資料中有20個異常點的存在。
2. H A T D I A G . H. (H) 如表3.1所示依 H A T D I A G . H. 之單一異常點的衡量標準，從表3.1中可以找出樣本資料中有18個異常點的存在，並由此表可瞭解異常點形成之區位分布情形。而從表3.2中可以發現到1988年樣本資料中有13個異常點的存在，而在1989年樣本資料中有5個異常點的存在。
3. D F F I T S (D) 如表3.1所示，依單一異常點的衡量標準之 D I F F S 從表3.1中可以找出樣本資料中有55個異常點的存在，並由此表可了解異常點形成之區位分布情形。而從表3.2中可以發現到1988年樣本資料中有34個異常點的存在，而在1989年樣本資料中有21個異常點的存在。
4. C O V R A T I O (C) 如表3.1所示，依單一異常點的衡量標準之 C O V R A T I O 從表3.1中可以找出樣本資料中有37個異常點的存在，並由此表可了解異常點形成之區位分布情形。而從表3.2中可以發現到1988年樣本資料中有22個異常點存在，而在1989年樣本資料中有15個異常點的存在。

### (二) 複異常點的衡量標準：

此處選取複異常點的衡量標準為 RSTUDENT-DIFFS (R/D)、RSTUDENT-COVRATIO (R/CO)、RSTUDENT-HAT DIAG. H. (R/H) 以及 HAT DIAG. H. -DIFFS (H/D) 之倆倆判斷異常點的衡量標準，現擬用上述三個衡量標準分別找出各不同區位及1988與1989樣本資料之異常點分布情形，如表3.3及表3.4所示。

### (三) 多重異常點的衡量標準：

多重異常點的衡量方式採用 RSTUDENT-COVRATIO-DIFFS (R/CO/D)、RSTUDENT-HATDIAG. H. -DIFFS (R/H/D)、RSTUDENT-HATDIAG. H. -COVRATIO (R/H/CO)，此種判斷異常點的衡量標準，現擬用此種衡量標準分別找出各不同區位及1988與1989年樣本資料之異常點分布情形，如表3.5及表3.6所示。



#### (四) 小結：

從上面資料的內容可以明顯看出不同異常點的衡量標準將產生不同的異常點，究竟應該如何選取異常點的衡量標準，才不致於造成過度修正或修正不足的情形產生，本文將在下二節討論之。

### 四、異常點修正之重要性

從上一節中可以看出異常點的判定方法有很多種，如單一異常點，複異常點及多重異常點的判定方法，但前述判定法將得到不大一致的異常點及筆數，有的方法所求出之筆數多，有的筆數少，究竟應用何種方法修正才不會有過度或不足現象發生，則為本節及下節探討的重點所在。

一般修正異常點之方法有以統計理論基礎之穩健性（ROBUST）迴歸估計方式如Herb's之M—迴歸估計方式，Tiku's之迴歸估計方式及其他統計法來修正異常點，而本文是採用除去法來處理異常點，採用除去法的原因乃基於本文所採用之資料均為大樣本的緣故。（註9）

此外，本節擬以一般統計常用衡量模式的標準，分別以三類型之模式，來探討異常點除去前後，對整個模式的解釋能力是否有所改變，並以此推估異常點除去法的適當性。而一般統計常用衡量模式的標準，及指以下之值，RSQUARE，Adj-RSQUARE及Durbin-Watson等SAS系統所顯示之統計值，整理如表4所示。

從表4中可以明顯看出，無論是採用單一異常點，複異常點及多重異常點的衡量標準除去資料中之異常點，其所得到之統計值除少數外，均較原始資料統計值優異，亦即異常點除去後，對整個模式的解釋能力增強，由此可知在資料為大樣本之情況下，異常點之修正乃是重要的。

### 五、最適特徵住宅租金模式的選取

上一節主要是考慮到有異常點存在時是否需要修正的問題，經過實證研究的結果認為經過異常點修正後的統計值與原始資料統計值相比較，則異常點除去修正後對整個模式的解釋能力大多增加，由此可見不可因大樣本資料而忽視異常點的存在，但在上一節中並沒考量到要選取何種方法以修正異常點，才不會有過之與不及的現象產生，本節擬從表4中找出何者為異常點修正之最佳模式。

從表4中可以看出異常點筆數除去多寡，與整個模式的解釋能力優劣並無相關，而是因單一異常點，複異常點及多重異常點的衡量標準而有所不同。以單一異常點為例，在五種單一異常點衡量標準下，以D F F I T S (D) - 異常點的衡量標準，修正異常點後得最佳模式，但在複異常點之下，則以R S T U D E N T - D I F F S (R/D) 得到修正異常點後之最佳模式，

而在多重異常點下，則以 R / C / D 為最佳模式。

將單一異常點修正後之最佳模式（D）、複異常點修正後之最佳模式（R / D）多重異常點之最佳模式（R / C / D），三者之統計值相比較，則可以發現到以單一異常點衡量標準之 DIFFTS，其修正異常點後得之最佳模式對整個模式的解釋能力超過複異常點衡量標準之 R / D 及多重異常點衡量標準之 R / C / D，其修正異常點後得之最佳模式對整個模式的解釋能力，由此可知異常點修正後之最佳模式為單一異常點衡量標準之 DIFFTS，藉其修正異常點後可得最佳模式。

模式一、模式二中單一異常點衡量標準之 DIFFTS 及模式三中單一異常點衡量標準之 RSTUDENT，其修正異常點後得之模式雖然對整個模式的解釋能力最佳的異常點修正方法，但其對個別參數之估計值及是否顯著之解釋能力有否較原始資料佳則無從判斷，故將列表討論之。

從表 5.1、表 5.2 及表 5.3 中樣本原始資料之模式 t 值，及修正異常點後得之模式之 t 值，可以看出原本不顯著之變數經過修正後之變化，而參數估計值因符號修正前後均一致，故較無顯著之差異。大體而言，模式一、模式二及模式三中以單一異常點衡量標準修正異常點後得之模式，其中以模式二然對整個模式的解釋能力最佳；其可為異常點修正後之最適特徵住宅租金模式，並可作為估計租金模式之參考。

從上表之結果，更可以肯定大樣本資料更必須進行異常點的修正。

## 六、結論

一般人認為基於中央極限定理，資料在大樣本之下，其在統計上是屬於近似常態分配而可以忽視異常點的存在，但經本文實證結果顯示不僅小樣本資料必須做異常點的修正甚至於大樣本亦必須做異常點的修正，經過異常點修正後整個模式及各自變數之間的解釋能力均增加了，故異常點修正是有存在之必要性。

文中由於資料屬大樣本之緣故，故以異常點除去法修正模式，而異常點的修正方式應可由統計基礎之穩健性 (Robust) 迴歸估計方式如 Herb's 之 M 一迴歸估計方式，Tiku's 之迴歸估計方式及其他統計法來修正異常點，留待後續研究。又本文以房租模式來探討異常點之問題，而台灣之房租變動幅度較小；若以房價模式應可獲得更進一步探討之課題。（註 10）



## 註釋

註 1：參考文獻 Daniael Pena Javier Ruig-Castill(1984)。

以小樣本中處理異常點之方法有較多之文獻探討（如 Hoaglin, Mosteller and Tukey (1983)；Tiku, Tan and Balakrishnan (1986)之穩健性估計法）且亦可以無母數法簡易處理，本研究不予討論；僅討論易為人忽視大樣本資料處理異常點之方法。

註 2：參考文獻 Thibodeau T.G. (1992).

註 3：參考文獻林祖嘉（1990）出租住宅類型之分類

註 4：實證中、依經驗或逐步迴歸、所選取變數組合成三類較佳之模式，並分別以不同異常點的橫量標準，衡量、判定樣本資料之異常點。由於篇幅有限，經篩選後，將部分之結果列於本文中。

註 5：依典型相關分析法將第二類選取樣本變數合併為 V1；且將第三類選取變數合併為 V2。  
期間之關聯係數由最適標準化之正典相關係數得之。

註 6：同註4。

註 7：參考文獻 Belsley D. A.; Huh E.; Welsch R.E. (1980)。

註 8：同註4。

註 9：以小樣本中處理異常點之方法有較多之文獻探討（如 Hoaglin, Mosteller and Tukey (1983)；Tiku, Tan and Balakrishnan (1986)之穩健性估計法）且亦可以無母數法簡易處理，本研究不予討論；僅討論易為人忽視大樣本資料處理異常點之方法。

註10：異常點非以除去法修正房價模式所為之結果參考文獻 林秋瑾, (1994) "穩健性特徵房屋價格模式之研究---異常點分析" 行政院國家科學委員會研究計畫成果報告.民國八十三年十月。



## 參 考 文 獻

英 文 部 分 :

Belsley D. A., Huh E., Welsch R.E.

1980 Regression Diagnistics : Identifying Influential Data and Source of Collinearity John Wiley and Sons : New York.

Daniel Pena Javier Ruiz - Castill .

1984 "Robust Methods of Building Regression Model - An Application to the Housing Sec" Jounal of Business and Economic Statistics Vol 2, No1, pp.10-20.

David C. Hoaglin, F. Mosteller and Tukey J. W.

1983 Understanding Robust and Data Analysis John Wiley and Sons : New York.

Graves Phil, Murdoch J. C., Thayer Mark A. and Waldman D.

1988 "The Robustness of Hedonic Price Estimation Urban Air Quality" Land Economics, Vol 64 No.3, August pp.220-233.

Rosen S.

1974 "Hedonic Prices and Implicit Markets : product Differentiation in pure Competition," Jounal of Political Economy 82 pp. 34-55.

SAS User's Guide : Basics, Version 5ed, Cary, NC : SAS Institute Inc., 1985.

SAS User's Guide : SAS/ETS, Version 6ed, Cary, NC : SAS Institute Inc., 1990.

SAS User's Guide : Statistics (1,2) Version 6ed, Cary, NC : SAS Institute Inc., 1990.

Tan W.Y. and Tabatabai M.A.

1988 "A modified Winsorized Regression Procedure for Linear Models," Jour. Statist. Comput. Simul. Vol.30, pp99-313.

Terry E. Dielman

1991 Applied Regression Analysis for Business and Economics Pws-Kent : Boston.

Tiku M.L. ,Tan W.Y. and Balakrishnan N.

1986 Robust Inference Marcel Dekker; New york.



Thibodeau T.G.

1992 "Residential Real Estate Price: 1974-1983", From the Standard Metropolitan Statistical Area Annual Housing Surveys, edited by Richd B. Clemmer, Centeral Michigan University, Blackstone Company.

中文部分：

林元興

1988 <台北市地價之計量分析> 《國立政治大學學報》(39): 117-166。

林元興

1988 <不動產估價計量方法初探> 《國立政治大學學報》(58): 173-187。

林秋瑾

1994 《穩健性特徵房屋價格模式之研究---異常點分析》行政院國家科學委員會研究計畫成果報告，民國八十三年十月。

林祖嘉

1990 <台灣地區房租與房價關係之研究> 《台灣銀行季刊》第43卷(1): 227-243。

林祖嘉 林素菁

1993 <台灣地區環境品質與公共設施對房價及房租的影響之分析> 《住宅學報》第一期: 21-45。

張金鶴 范垂爐

1993 <房地產真實交易價格之研究> 《住宅學報》第一期: 75-91。

張梅英

1992 <台灣地區都市地價變動分析> 《經社法制論叢》第十期, 81/7: 301-327。

張麗姬

1994 <從遠期契約和現貨的角度論預售屋和成屋的價格關係—以台北市為例> 《住宅學報》第二期: 67-85

辜炳珍及劉瑞文

1979 《房地產價格指數查編之研究》，行政院主計處統計專題研究報告。

