住宅學報　第三期　民國八十四年一月　學術論著　第73頁—98頁
JOURNAL OF HOUSING STUDIES NO. 3, JAN. 1995 RESEARCH pp. 73—98

空屋率的模型選擇及其穩定性：遺傳規畫的應用*

On the Model Selection and Its Stability of the Natural Vancancy Rates of Housing: An Application of Genetic Programming

陳樹衡**　林祖嘉***　葉佳炫****
Shu-Heng Chen**, Chu-Chia Lin***, Chia-Hsuan Yeh****

摘　要

本文應用人工智慧領域新發展的遺傳規畫重新對林祖嘉等(1994)自然空屋率的估計進行研究。雖然遺傳規畫所得到的估計值與論文研究相似，但是對自然空屋率高峰期的出現卻有不同。另外，遺傳規畫也發現空屋率背後的模型，可能有穩定性的問題。這點也許與房地產市場在資料的涵蓋期間，正好是由衰退步入繁榮的變化有關。

ABSTRACT

This paper discusses the applications of genetic programming to the empirical study of the natural rates of vancancy in Taiwan's housing market. The genetic programming paradigm, a new approach developed in artificial intelligence, is an automatic model search process and is very promising in treating the issue of model selection. By using the model in Lin et. al. (1994) as a benchmark, we explore the advantages of this approach by demonstrating two things: firstly, how genetic programming can be used to investigate the robustness of a given model; secondly, how genetic programming can be used to detect the potential non-linearity and structural stability in the model. Our findings are two-fold. First of all, using the data running from 1981 to 1988, we find that the 2SLS model considered in Lin et. al. is pretty robust at least in the sense of linearity. However, if we exclude 1981 and add 1989 to our data set, the model is not robust any more. Our further analyses suggest that business cycles in the housing market might affect our estimate of natural vancancy rates and should be taken into account in future studies.

＊本文初稿曾發表於淡江大學第二屆產業經濟學術研討會。在此，特別感謝陳忠榮教授及黃台心教授所提供的意見。除此，並感謝三位匿名評審人對本文所提供的修正方向。當然，本文一切錯誤仍由作者自行負責。
** 政治大學經濟系副教授
*** 政治大學經濟系教授
**** 政治大學經濟系博士班研究生

一、研究動機與方法

模型選擇在計量經濟學中一直是個議論不定的問題，其影響所及，不僅在於模型設定的本身，而且也連帶地左右到模型穩定性的問題。Koza (1992a)對模型選擇提供了一個新的思維架構，在其架構中，所謂好的模型必須要能通過達爾文“物競天擇，適者生存”的驗証。因此，它在分析方法上的創新是將每一個計量模型都賦予一個進化的空間，這點和傳統計量模型的被設定或侷限在一定範圍有很大的差別。

不僅如此，遺傳規畫(Genetic Programming)也提供了程序資訊理論一條可以實踐的路。程序資訊理論係發展於1964年左右，它是對資訊理論中有關資訊這個概念的反省與重新定義。在1964年以前的資訊理論，也就是所謂的“夏濃資訊理論”(Shannon Information Theory)，是將資訊定義在機率論的基礎上，因此機率論就自然成了資訊理論的基礎。1964年崛起的程序資訊理論則以計算理論中程式(algorithm)的概念來定義資訊(註1)，而將每一個物件(object)所蘊涵的資訊定義成能描述該物件最短程式的長度，此一長度也稱程序複雜度。這個突破與夏濃資訊理論的差別在於後者對資訊的量化只能用於隨機變數，而前者則可應用在任何物件上。這個區別使得資訊理論有其獨立的地位，而不再只是機率論中的一個附屬概念。

程序複雜度不僅為資訊這個概念提出了一個新的定義，而且也更具體地將Occam's razor的精神帶入到歸納性學習的領域中來。Occam's razor是說如果有兩個或兩個以上的理論可以解釋同一現象，則愈簡單的理論便是愈好的理論。而程序資訊理論正好可以為“什麼是簡單？”提出一個具體的定義，而將Occam's razor詮釋成：歸納性學習的意義是要從資料中找到能描述該資料的最短程式或模型，或者簡稱為“最短描述法則”(Minimum Description Length Principle, MDLP)。然而，要將最短描述法則真正應用到模型選擇上來，必須要解決的問題是如何去“尋找”此一最短描述的模型或法則。除非“尋找”的問題能解決，否則“最短描述法則”就只能陳諸高閣。有關這點，Rissanen (1989)便提到：

"However, the theory gives no guidance to the practical construction of programs, let alone of the shortest one, which in fact turns out be non-computable. Accordingly, the theory had little or no direct impact on statistical problems." (p.i)

針對這個“尋找”的困難，遺傳規畫所提供的就正是一種出路。遺傳規畫是將達爾文進化論中複製、交配與突變的機制運作在一群由程式所組成的社會中。這裡的每一個程式都代表著一個模型，而每一個模型又是用人工智慧學者McCarthy在1950年代所發展出的LISP (List Programming)的語法(syntax)來表示。LISP的語法可以將每一個模型表達成一個層級深度不同的樹(註2)，而遺傳規畫便是利用這種樹狀結構來設計它的進化(註3)，使得一個模型都像樹一般地有生命現象，可以生長、停滯與死亡。進化從描述長度較短的模型開始，視需要來決定複雜化的程度。因此，它是由簡入繁地盡可能尋找最短描述的模型，所以它也是對程序資訊理論中最短描述法則的一種實踐。

在這樣的認識下，本文的目的是想透過一個具體的應用，來說明遺傳規畫怎麼樣可以影響到傳統計量經濟分析。在這裡，我們將以林祖嘉等(1994)一文中對空屋率的研究作為起點，然後說明遺傳規畫在模型選擇及模型穩定性上的應用。在前者，我們將可看到：在既定模型的估計下，遺傳規畫如何能對此一估計予以一種敏感度評估(sensitivity analysis)。而在後者，我們將看到遺傳規畫如何能在不設定模型的情況下，來探討穩定性的問題。由於這只是非常初步的研究，因此我們也會在文中討論現階段所遭遇到的一些困難及其未來可改進的方向(註4)。

林祖嘉等(1994)一文對臺灣地區空屋率之研究中發現(註5)：臺灣地區70年至78年的平均實際空屋率為14.68%，遠高於其估計到的自然空屋率4.59%。同時，4.59%的自然空屋率也較美國一些實證結果為低(註6)。所以，究竟是臺灣地區的實際空屋率真的遠高於自然空屋率，還是臺灣地區的自然空屋率被低估，值得進一步的研究。

在模型的設定上而言，林祖嘉等(1994)用方程式(1)來代表自然空屋率的決定：

$$V_t^n = d_0 + d_1 M_t + d_2 D_t, \tag{1}$$

其中V_t^n、M_t及D_t分別代表自然空屋率、遷徙率及房價離散程度(註7)，且d_1、d_2皆為正值。

換句話說，自然空屋率(V_t^n)與遷徙率(M_t)及房價離散程度(D_t)之間有一個恆定的關係。只要知道d_0, d_1及d_2，我們就可以從M_t及D_t中得知觀察不到的V_t^n。但是，d_0, d_1及d_2這些係數顯然無法由方程式(1)得到。所幸，實際空屋率與自然空屋率有一個如方程式(2)的關係，即

$$V_t - V_t^n = c_1 P_t + c_2 V_{t-1}, \tag{2}$$

也就是實際空屋率與自然空屋率之間的差距，可以用房價(P_t)及上一期的實際空屋率(V_{t-1})來決定。就某些意義上而言，方程式(2)可以視為方程式(1)的動態調整式。若再將方程式(2)中的V_t^n移至等號右邊，並將方程式(1)代入，我們便可得到：

$$\begin{aligned} V_t &= V_t^n + c_1 P_t + c_2 V_{t-1} \\ &= d_0 + d_1 M_t + d_2 D_t + c_1 P_t + c_2 V_{t-1}, \end{aligned} \tag{3}$$

用(3)式做迴歸方程式的基礎，便可以得到d_0, d_1及d_2之估計，而得到對V_t^n之估計。

總而言之，以上作法基本上是將實際空屋率視為自然空屋率與動態調整的一個線性組合。這種想法顯然是方程式(4)的一種線性特殊化。

$$\begin{aligned} V_t &= f\left(V_t^n, V_{t-1}, P_t\right) + u_t \\ &= f\left(V_t^n\left(M_t, D_t\right), V_{t-1}, P_t\right) + u_t, \end{aligned} \tag{4}$$

至於方程式(4)能否簡化成方程式(3)的形式，在計量上而言是個模型設定或選擇的問題。如果我們對$f(.)$的形式為已知，而且$f(.)$可以涵蓋到方程式(3)，則我們自可使用Hendry“從一般而特定”(from general to specific)的檢定方式(註8)。可是計量上最難獲得的訊息就是$f(.)$的具體形式。因此，本文在不先驗認定任何函數形式的前提下，嘗試以Koza (1992a)在人工智慧領域中所提出的遺傳規畫(Genetic Programming)來設計一個函數自動尋找的過程(註9)。

二、模型選擇與空屋率之估計

本文所使用的遺傳規畫是以林祖嘉等(1994)的聯立方程估計為基礎，也就是以該文(8)與(9)兩式為基礎。為了便於思考，我們將其轉錄於下：

$$P_t = \alpha_0 + \alpha_1 V_t + \alpha_2 Y_t + \alpha_3 N_t + \alpha_4 V_{t-1} + \varepsilon_{1t} \tag{5}$$

$$V_t = \beta_0 + \beta_1 P_t + \beta_2 M_t + \beta_3 D_t + \beta_4 V_{t-1} + \varepsilon_{2t}, \tag{6}$$

而其中$\beta_0 = d_0$, $\beta_2 = d_1$, $\beta_3 = d_2$, Y_t為家計單位的所得，N_t為家計單位的數目。

我們先仿照林祖嘉等(1994)一文的計量方法，對其做70年到77年的估計(註10)，然後根據其估計結果做對78年的預測。之所以如此，是為了了解遺傳規畫是否產生過度配適(overfitting)的問題(註11)。所以，我們在最後仍要以其預測的結果拿來與2*SLS*或3*SLS*再做一個比較，以做為其可信任性的參考。

表一是聯立方程式(5)與(6)，分別用2*SLS*與3*SLS*及民國70到77年的資料所得出的結果。其中係數符號正負的部份都與林祖嘉等(1994)一文相符，而且除了D_t (房價離散程度)外，係數也均為顯著。至於這樣的模型是否具有競爭力？亦即在一個設定的競爭環境中，當這個模型在遭遇到其他模型時(註12)，能否經得起考驗便是本文首先所要探索的。

為此，一個模型競爭的環境由遺傳規畫發動(註13)。表二是該環境的參數設定，這些參數基本上可讓我們選擇演化(競爭)的規模及演化的過程。在規模方面包含了競爭對手的多寡(族群個體總數)，及開放競爭的時間(子孫代數)。而在過程方面則已包含了基本結構(函數集合與終點集合)、創造(完整法與成長法)，及演化運作元(複製、交配、突變與新生嬰兒)的進化方式。而競爭力的衡量是以誤差平方和(*SSE*)來代表，因此，表二明顯地揭櫫了一個競爭的環境(註14)。

表三則是整個競爭後的結果。表三中第二欄中的第一小欄SSE_1是代表到第200代時，最佳模型對70年到77年的空屋率之配適誤差平方和。從這一欄中，我們可以看出在七次模擬中，只有在第六次時，2*SLS*模型存活了下來。而在其他各次模擬中，2*SLS*都遭到淘汰的命運。

表一 結構方程式之迴歸估計(民國70-77年)

解釋變數	2SLS V_t	3SLS V_t
常數項	3.106610 (6.930)**	3.111272 (6.940)**
P_t	-1.471608 (-7.920)**	-1.530827 (-8.262)**
M_t	0.265525 (8.272)**	0.269438 (8.397)**
D_t	0.002161 (0.625)	0.004692 (1.378)
V_{t-1}	0.903974 (34.460)**	0.900350 (34.340)**
$\bar{R}^2$	0.9088	0.8694
F-value	456.852	305.542
System Weighted R^2	—	0.9129
Num. of Observations	184	184

附註：
1 括弧內為t值。
2 有**者表示該係數在95%顯著水準下，顯著的異於0。

表二 模型競賽環境之參數設定表

族群個體總數	500
完整法創造個體的數目	50
成長法創造個體的數目	50
函數集合的元素	{+, -, ×, %, *Exp*, *Log*, *Sin*, *Cos*}
終點集合的元素	$\{P_t, M_t, D_t, V_{t-1}, R\}$
複製的個數	50
新生嬰兒的個數	100
突變的個數	100
突變的機率	0.2
個體長度的限制	17
葉子點被選到的機率	0.5
子孫代數	200
指數定義域的上限值	1700
保留最佳個體數目	2
競爭力的衡量標準	誤差平方和(*SSE*)

表三 各次競賽的結果

模 擬	進化到第200代的SSE (70-77, 78)		
	SSE_1	Pred SSE	SSE_T
house01	252.84	191.48	444.33
house02	252.84	191.48	444.33
house03	249.15	164.38	413.53
house04	251.27	190.44	441.71
house05	248.34	192.79	441.14
house06	260.19	347.73	607.92
house07	252.70	190.42	443.12
2*SLS*	260.19	347.73	607.92
3*SLS*	267.68	336.13	603.82
OLS	220.00	159.30	379.31

附註：
1 SSE_1：對70年到77年真實空屋率之配適誤差平方和。
2 Pred SSE：對78年真實空屋率之預測誤差平方和。
3 SSE_T = SSE_1 + Pred SSE。

以下則是各次競賽到第200代時的最佳模型。

house01:

$$V_t = ((((3.106610 - (1.471608+P_t)) + (0.265525 * M_t)) + (0.002161 * V_{t-1})) + (0.903974 * V_{t-1}))$$

house02:

$$V_t = ((((3.106610 - (1.471608+P_t)) + (0.265525 * M_t)) + (0.002161 * V_{t-1})) + (0.903974 * V_{t-1}))$$

house03:

$$V_t = ((((3.106610 - (1.471608+P_t)) + (0.265525 * M_t)) + 0.265525) + (0.903974 * V_{t-1}))$$

house04:

$$V_t = ((((3.106610 - (1.471608+P_t)) + (0.265525 * M_t)) + (0.002161 * V_{t-1})) + (0.903974 * ((SinV_{t-1} \% V_{t-1})+V_{t-1})))$$

house05:

$$V_t = (((((3.106610 - (1.471608+P_t)) + (0.265525 * M_t)) + (0.002161 * V_{t-1})) + (0.903974 * V_{t-1})) - (Cos(V_{t-1}-(V_{t-1}+P_t) \% V_{t-1}))$$

house06:

$$V_t = ((((3.106610 - (1.471608 * P_t)) + (0.265525 * M_t)) + (0.002161 * D_t)) + (0.903974 * V_{t-1}))$$

house07:

$$V_t = ((((3.106610 - (1.471608+P_t)) + (0.265525 * M_t)) + (0.002161 * V_{t-1})) + (0.903974 * Log(D_t+ExpLog(Log\ (V_{t-1} + (Log(ExpV_{t-1}+ExpV_{t-1}) + ExpV_{t-1})) + ExpV_{t-1}))))$$

從所列出的七個模擬的結果中，我們可以看到除了模擬5及7有一些非線性的傾向外(註15)，其餘的模擬基本上都是由原有模型的一些簡單突變或交配所產生的一些“線性進化”(註16)。譬如說，模型1相對於原有模型(2*SLS*)而言只是運作了一次簡單的突變，而模型3則是再加上了一次交配。另外值得注意的是在原來2*SLS*中不顯著的房價離散度(D_t)係數，在七次模擬中幾乎是遭到完全出局的命運。除了模擬6及7外，沒有任何一次模擬建議留下D_t這個變數(註17)。所以，就這點而言，遺傳規畫所建議的結果與統計上的t檢定是一致的。

再者，由於非線性模型並沒有成功地在競爭中存活下來，因此過度配適的問題似乎並不值得擔心。這點也可以從表三中的“Pred *SSE*”一欄中看出。“Pred *SSE*”是用70年到77年所產生的模型來預測78年的真實空屋率的誤差。我們可以看出除了模擬6的模型是保有原來2*SLS*外，其餘的模型在對78年23個行政區的真實空屋率的預測誤差平方和上，較3*SLS*小了近50%以上。其中預測最好的house03更不到3*SLS*的一半，其餘幾個模型的預測誤差平方和也都維持在190左右(註18)。就這個結果來看，遺傳規畫的方法並沒有衍生出過度配適的問題，因此我們便可以此模型作為推估自然空屋率的基礎。

然而遺傳規畫七次模擬所留下的最佳方程式，並非每一條都具有方程式(4)的一種可分割的形式(註19)。以模擬7所留下的“house07”而言，D_t便與V_{t-1}一同出現在該方程式非線性的部份。因此對方程式(7)來說，我們並沒有辦法從真實空屋率中分離出自然空屋率。而“house07”也是在七個方程式中唯一對林祖嘉(1994)自然空屋率設定形式提出質疑的。但是，如果我們接受house01到house05的建議，認為D_t不重要的話，而將其從自然空屋率的定義中刪除，則house07所產生的非遞延性就不會對自然空屋率的估計產生困擾。如果是這樣的話，house01到house07對自然空屋率所選擇的函數形式便如表四所列。

根據表四我們得出了三條對自然空屋率的估計方程式(7)、(8)及(9)。從這三條方程式中，我們可以看出house06就是2*SLS*對自然空屋率的估計最高，再來是house03。而表五至七則是分別根據方程式(7)、(8)、(9)所得到對自然空屋率的估計。從這三個表中，我們可以發現民國70年到民國77年的自然空屋率之平均值，分別是從最低的4.35%到最高的5.98%。而同樣期間的平均值在林祖嘉等(1994)是4.47% (註20)。若是我們用預測最好的house03，即方程式(8)來做比較的標準，則林祖嘉等(1994)的結果(4.47%)相對於表六而言(4.61%)是顯得略低。比較方程式(8)與(10)，我們可以看出house03與林祖嘉等(1994)在自然空屋率估計上的差異。在遷徙率比較高的

表四　估計自然空屋率的函數

house01		
house02		
house04	$V_t^n = 1.635002 + 0.265525 * M_t$	(7)
house05		
house07		
house03	$V_t^n = 1.900527 + 0.265525 * M_t$	(8)
house06	$V_t^n = 3.106610 + 0.265525 * M_t + 0.002161 * D_t$	(9)
林祖嘉等	$V^n = 1.9253 + 0.1586 * M_t + 0.009 * D_t$	(10)

表五　各縣市自然空屋率之估計，民國70-77年，方程式(7)

單位：%

	70	71	72	73	74	75	76	77	平均
台灣地區	4.70	4.30	4.42	4.27	4.22	4.29	4.33	4.25	4.35
台北市	8.21	7.26	7.80	7.11	7.23	7.33	7.68	7.36	7.50
高雄市	6.96	5.88	6.22	5.50	5.86	5.89	5.81	5.98	6.01
台灣省	4.43	4.08	4.18	4.08	4.00	4.07	4.10	4.02	4.12
基隆市	5.40	4.98	4.84	4.81	4.55	4.71	5.13	4.92	4.92
台中市	7.33	6.33	6.99	6.48	6.38	6.51	6.60	7.13	6.72
台南市	6.96	7.08	6.75	6.61	6.31	6.34	6.00	5.66	6.46
台北縣	6.53	6.02	6.13	5.67	5.79	5.72	7.04	5.51	6.05
宜蘭縣	3.48	3.36	3.83	3.44	3.03	3.43	3.38	3.54	3.44
桃園縣	6.38	5.36	5.46	5.01	4.92	4.68	5.06	4.93	5.23
新竹市	5.15	4.94	4.32	4.63	4.74	4.93	5.48	5.12	4.91
新竹縣	3.44	3.21	3.41	3.59	3.09	2.97	2.97	3.26	3.24
苗栗縣	3.37	2.92	2.99	3.33	3.25	3.16	3.14	2.76	3.12
台中縣	4.27	4.03	4.13	3.77	4.28	4.12	3.86	3.88	4.04
彰化縣	3.43	3.09	3.21	3.29	3.37	3.26	3.10	3.06	3.23
南投縣	3.53	3.00	3.28	2.95	2.95	2.93	3.00	3.10	3.09
雲林縣	2.71	2.76	3.22	2.87	2.68	2.87	2.72	2.99	2.85
嘉義市	4.71	5.02	4.89	4.42	4.15	5.14	4.82	4.86	4.75
嘉義縣	2.99	2.76	2.58	3.01	2.61	2.57	2.70	2.69	2.74
台南縣	3.49	3.45	3.35	3.60	3.58	3.30	3.60	3.35	3.47
高雄縣	4.39	4.01	4.16	4.15	3.81	3.83	3.87	4.30	4.07
屏東縣	3.46	3.15	3.48	3.03	3.26	3.14	3.06	3.14	3.22
台東縣	4.43	3.59	3.20	3.36	3.89	3.65	3.55	3.44	3.64
花蓮縣	4.57	3.81	3.81	3.80	3.95	4.00	4.01	3.66	3.95
澎湖縣	3.02	2.78	3.68	3.83	3.33	4.13	3.10	3.18	3.38

表六　各縣市自然空屋率之估計，民國70-77年，方程式(8)

單位：%

	70	71	72	73	74	75	76	77	平均
台灣地區	4.97	4.56	4.69	4.54	4.48	4.55	4.60	4.52	4.61
台北市	8.47	7.52	8.06	7.38	7.50	7.60	7.94	7.63	7.76
高雄市	7.22	6.15	6.48	5.76	6.13	6.15	6.07	6.24	6.28
台灣省	4.69	4.34	4.44	4.34	4.26	4.33	4.37	4.29	4.38
基隆市	5.67	5.25	5.11	5.07	4.81	4.97	5.40	5.19	5.18
台中市	7.60	6.60	7.25	6.75	6.64	6.77	6.86	7.40	6.98
台南市	7.22	7.35	7.02	6.87	6.58	6.61	6.27	5.92	6.73
台北縣	6.80	6.28	6.40	5.94	6.06	5.98	7.30	5.77	6.32
宜蘭縣	3.75	3.62	4.09	3.71	3.29	3.69	3.65	3.80	3.70
桃園縣	6.64	5.63	5.73	5.28	5.18	4.94	5.32	5.20	5.49
新竹市	5.41	5.20	4.58	4.90	5.01	5.20	5.74	5.39	5.18
新竹縣	3.70	3.48	3.68	3.86	3.35	3.24	3.24	3.53	3.51
苗栗縣	3.63	3.19	3.25	3.59	3.52	3.42	3.40	3.02	3.38
台中縣	4.54	4.30	4.40	4.04	4.55	4.39	4.13	4.14	4.31
彰化縣	3.69	3.35	3.48	3.55	3.64	3.53	3.36	3.33	3.49
南投縣	3.79	3.26	3.55	3.22	3.21	3.19	3.27	3.37	3.36
雲林縣	2.97	3.02	3.48	3.14	2.94	3.14	2.99	3.26	3.12
嘉義市	4.98	5.29	5.16	4.68	4.42	5.40	5.08	5.12	5.02
嘉義縣	3.26	3.02	2.85	3.27	2.88	2.84	2.97	2.95	3.01
台南縣	3.75	3.72	3.62	3.86	3.84	3.57	3.87	3.62	3.73
高雄縣	4.66	4.27	4.42	4.42	4.08	4.09	4.14	4.57	4.33
屏東縣	3.72	3.41	3.75	3.30	3.52	3.41	3.32	3.40	3.48
台東縣	4.70	3.86	3.46	3.63	4.16	3.91	3.81	3.71	3.91
花蓮縣	4.83	4.08	4.07	4.07	4.22	4.26	4.27	3.93	4.22
澎湖縣	3.28	3.04	3.94	4.09	3.59	4.40	3.36	3.44	3.64

縣市，如台北市、高雄市、基隆市、台中市與台南市等，house03所估計的自然空屋率一般都會高於林祖嘉等(1994)。反之，遷徙率比較低的縣市如宜蘭縣、雲林縣與嘉義縣等，則林祖嘉等對自然空屋率的估計會較house03為高。

此外，自然空屋率出現的高檔期在我們的結果中也與林祖嘉等(1994)不同。林祖嘉等的高檔期是出現在民國76年至民國78年，這段時間是房地產景氣的復甦到繁榮期。而我們估計的高檔則傾向在民國70年至72年，亦即是房地產景氣從衰退走到低迷的時候。造成這個差異的主要原因可能來自於房價離散程度。由於遺傳規畫所建議的模型多將房價離散程度排除在自然空屋率的定義外，所以當房地產從復甦到繁榮時房價離散程度的提高，以及房地產從衰退走向低迷時房價離散程度的降低，都對遺傳規畫中所建議自然空屋率沒有影響，但卻會影響到2*SLS*估計的自然空屋率。在景氣時，它會使2*SLS*所估計的自然空屋率上升。而在不景氣

表七 各縣市自然空屋率之估計，民國70-77年，方程式(9) (2*SLS*)

單位：%

	70	71	72	73	74	75	76	77	平均
台灣地區	6.32	5.90	6.04	5.87	5.82	5.94	6.01	5.95	5.98
台北市	9.83	8.95	9.48	8.76	8.90	9.20	9.44	9.13	9.21
高雄市	8.64	7.62	7.88	7.25	7.57	7.62	7.48	7.75	7.73
台灣省	6.05	5.68	5.79	5.67	5.59	5.70	5.77	5.72	5.75
基隆市	6.94	6.53	6.47	6.38	6.11	6.45	6.75	6.54	6.52
台中市	8.93	8.03	8.67	8.14	8.06	8.36	8.54	8.95	8.46
台南市	8.61	8.79	8.39	8.26	8.05	7.97	7.72	7.44	8.15
台北縣	8.18	7.60	7.77	7.29	7.37	7.35	8.72	7.20	7.69
宜蘭縣	5.07	4.94	5.44	4.99	4.56	5.05	4.99	5.20	5.03
桃園縣	7.94	6.97	7.19	6.62	6.49	6.25	6.73	6.61	6.85
新竹市	6.76	6.51	5.91	6.23	6.38	6.60	7.27	6.94	6.58
新竹縣	5.13	4.77	4.96	5.15	4.73	4.57	4.53	4.88	4.84
苗栗縣	4.96	4.53	4.68	4.92	4.84	4.74	4.75	4.44	4.73
台中縣	5.88	5.65	5.77	5.39	5.92	5.85	5.51	5.49	5.68
彰化縣	5.00	4.67	4.84	4.88	5.02	4.89	4.71	4.69	4.84
南投縣	5.15	4.57	4.86	4.49	4.50	4.63	4.67	4.76	4.70
雲林縣	4.27	4.34	4.80	4.44	4.25	4.45	4.39	4.67	4.45
嘉義市	6.48	6.64	6.52	6.14	5.74	6.80	6.60	6.61	6.44
嘉義縣	4.60	4.31	4.15	4.56	4.15	4.12	4.30	4.29	4.31
台南縣	5.17	5.03	4.99	5.16	5.18	4.90	5.22	4.99	5.08
高雄縣	5.93	5.65	5.80	5.73	5.37	5.44	5.51	5.91	5.67
屏東縣	5.02	4.72	5.07	4.59	4.80	4.70	4.66	4.83	4.80
台東縣	6.04	5.20	4.77	4.90	5.46	5.26	5.22	5.16	5.25
花蓮縣	6.39	5.47	5.40	5.35	5.51	5.60	5.68	5.56	5.62
澎湖縣	4.53	4.32	5.21	5.36	4.88	5.73	4.76	4.86	4.96

時，它會使2*SLS*估計的自然空屋率下降。因此，就造成了遺傳規畫與2*SLS*在自然空屋率高低檔期估計上的差異。雖然，就經濟意義上而言，價格離散程度會影響到搜尋的密度，進而影響到空屋率，而遺傳規畫中不採納此一變數的原因也許並非否定該一變數的重要性，而可能是房價離散度與自然空屋率之間有一更複雜的關係，這自然是未來值得再深一步研究的課題。

最後，從各表中，我們也可以發現各縣市的自然空屋率以台北市為最高。但是，自然空屋率是個均衡的概念，而檢視市場空屋問題的嚴重性應以失衡嚴重的程度來斷定。而失衡程度的指標可以自然空屋率與實際空屋率的差距來衡量，當兩者的差距愈大，則市場失衡的情況就愈來愈嚴重，反之亦然。為此，我們在表八與表九中分別算出林祖嘉等(1994)與house03的自然空屋率與實際空屋率的比例。該一比例愈接近1，則失衡程度愈緩和，否則愈嚴重。若以

表八 自然空屋率／實際空屋率，林祖嘉等(1994)

	70	71	72	73	74	75	76	77	平均
台灣地區	0.33	0.28	0.30	0.28	0.29	0.31	0.35	0.39	0.32
台北市	0.55	0.49	0.53	0.50	0.50	0.65	0.72	0.82	0.60
高雄市	0.45	0.39	0.37	0.37	0.36	0.36	0.34	0.43	0.38
台灣省	0.31	0.27	0.29	0.27	0.27	0.29	0.33	0.37	0.30
基隆市	0.27	0.23	0.25	0.22	0.19	0.25	0.25	0.28	0.24
台中市	0.32	0.30	0.31	0.29	0.28	0.35	0.40	0.40	0.33
台南市	0.35	0.35	0.32	0.31	0.33	0.31	0.34	0.38	0.34
台北縣	0.32	0.25	0.26	0.24	0.22	0.25	0.36	0.37	0.28
宜蘭縣	0.21	0.17	0.20	0.18	0.16	0.20	0.21	0.25	0.20
桃園縣	0.26	0.20	0.27	0.24	0.24	0.23	0.32	0.34	0.26
新竹市	0.27	0.24	0.23	0.26	0.27	0.29	0.37	0.40	0.29
新竹縣	0.36	0.28	0.31	0.33	0.36	0.35	0.36	0.47	0.35
苗栗縣	0.39	0.37	0.44	0.41	0.39	0.41	0.46	0.53	0.43
台中縣	0.21	0.21	0.23	0.23	0.26	0.31	0.30	0.32	0.26
彰化縣	0.21	0.19	0.22	0.20	0.26	0.24	0.24	0.27	0.23
南投縣	0.28	0.23	0.25	0.22	0.22	0.30	0.31	0.33	0.27
雲林縣	0.31	0.32	0.35	0.31	0.31	0.31	0.38	0.44	0.34
嘉義市	0.33	0.26	0.26	0.25	0.20	0.24	0.27	0.29	0.26
嘉義縣	0.27	0.22	0.23	0.23	0.22	0.21	0.25	0.29	0.24
台南縣	0.36	0.28	0.31	0.29	0.32	0.32	0.36	0.40	0.33
高雄縣	0.32	0.32	0.34	0.30	0.28	0.31	0.34	0.41	0.33
屏東縣	0.38	0.35	0.42	0.35	0.35	0.38	0.39	0.47	0.39
台東縣	0.43	0.38	0.32	0.31	0.39	0.41	0.44	0.48	0.40
花蓮縣	0.43	0.29	0.27	0.23	0.24	0.25	0.28	0.35	0.29
澎湖縣	0.22	0.21	0.24	0.27	0.23	0.24	0.25	0.26	0.24

此來看，我們會發現台北市儘管擁有最高的自然空屋率，但卻是失衡程度最低的城市。而自然空屋率並不是很高的宜蘭縣，反而是失衡程度最嚴重的城市。這個現象無論是在林祖嘉等(1994)或是house03中，皆是一致的。失衡的程度通常也反映房屋市場的價格機能。據此而言，台北市的房屋市場仍顯得相當地具有效率。當然，對房屋市場效率在各縣市之間差異的認識，也有待未來更深入的研究。

三、模型選擇與穩定性

在前一節中，我們利用遺傳規畫來進化一個既定的模型，並據此對自然空屋率予以重估。然而，在這個過程中，我們發現用來作為基準模型的2*SLS*或3*SLS*不僅在對78年的預測上表現很差，而且係數估計值也與林祖嘉等(1994)利用70年到78年整段時間的估計有所差異。因此我們不禁懷疑78年這年的資料室否有其“獨特性”。或者，說得更透徹點，70年到78年的

表九 自然空屋率／實際空屋率，方程式(8)

	70	71	72	73	74	75	76	77	平均
台灣地區	0.35	0.30	0.31	0.30	0.30	0.31	0.33	0.36	0.32
台北市	0.69	0.56	0.63	0.60	0.59	0.65	0.80	0.89	0.68
高雄市	0.52	0.40	0.42	0.36	0.38	0.37	0.37	0.43	0.41
台灣省	0.32	0.28	0.29	0.29	0.28	0.29	0.31	0.34	0.30
基隆市	0.34	0.28	0.27	0.25	0.22	0.23	0.28	0.30	0.27
台中市	0.41	0.33	0.36	0.33	0.31	0.34	0.35	0.41	0.36
台南市	0.42	0.40	0.37	0.36	0.35	0.36	0.36	0.37	0.37
台北縣	0.37	0.30	0.31	0.28	0.27	0.29	0.41	0.38	0.33
宜蘭縣	0.22	0.18	0.21	0.19	0.17	0.19	0.20	0.23	0.20
桃園縣	0.32	0.23	0.27	0.27	0.28	0.26	0.33	0.35	0.29
新竹市	0.30	0.28	0.25	0.28	0.28	0.31	0.35	0.36	0.30
新竹縣	0.31	0.30	0.33	0.36	0.32	0.33	0.36	0.44	0.34
苗栗縣	0.39	0.34	0.36	0.40	0.39	0.41	0.43	0.42	0.39
台中縣	0.23	0.22	0.23	0.23	0.26	0.28	0.29	0.33	0.26
彰化縣	0.22	0.18	0.20	0.20	0.23	0.22	0.22	0.25	0.22
南投縣	0.27	0.22	0.25	0.23	0.22	0.24	0.26	0.29	0.25
雲林縣	0.30	0.29	0.35	0.30	0.29	0.30	0.30	0.37	0.31
嘉義市	0.30	0.29	0.28	0.23	0.22	0.25	0.24	0.27	0.26
嘉義縣	0.25	0.22	0.22	0.24	0.21	0.20	0.23	0.25	0.23
台南縣	0.31	0.29	0.28	0.30	0.32	0.31	0.35	0.36	0.32
高雄縣	0.38	0.32	0.34	0.34	0.31	0.31	0.33	0.43	0.35
屏東縣	0.40	0.35	0.42	0.36	0.38	0.39	0.37	0.39	0.38
台東縣	0.46	0.37	0.33	0.34	0.43	0.40	0.40	0.40	0.39
花蓮縣	0.36	0.27	0.28	0.25	0.27	0.27	0.26	0.24	0.28
澎湖縣	0.24	0.21	0.27	0.31	0.24	0.26	0.21	0.22	0.25

整段資料有無所謂的結構變遷。如果有的話，其變遷的時點可能發生在那裡。

為了了解這點，我們將原始資料重新分割成71年到78年為一段訓練期，而將70年作為測試期，然後依照第二節的步驟重新再執行了七次遺傳規畫的模擬。表十是我們用林祖嘉等(1994)的方法所產生的2SLS與3SLS估計結果。根據表十所產生的2SLS與3SLS，我們將其放進遺傳規畫的進化環境中令其進化。表十一則是七次進化模擬的結果。七次競賽到第200代時的最佳模型則列於house21至house27。

表十　結構方程式之迴歸估計(民國71-78年)

解釋變數	2SLS V_t	3SLS V_t
常數項	1.197250	1.158652
	(2.852)**	(2.761)**
P_t	-0.597560	-0.671457
	(-4.221)**	(-4.832)**
M_t	0.101943	0.092591
	(3.534)**	(3.233)**
D_t	0.007065	0.010986
	(2.528)**	(4.580)**
V_{t-1}	0.935777	0.938836
	(35.914)**	(36.065)**
$\bar{R}^2$	0.9251	0.9055
F-value	566.130**	439.172**
System Weighted R^2	—	0.9226
Num. of Observations	184	184

附註：
1 括弧內為t值。
2 有**者表示該係數在95%顯著水準下，顯著的異於0。

表十一　各次競賽的結果

模擬	進化到第200代的*SSE* (71-78, 70)		
	SSE_1	Pred *SSE*	SSE_T
house01	195.70	114.26	309.96
house22	192.84	105.00	297.84
house23	196.47	119.46	315.93
house24	193.46	103.75	297.21
house25	196.41	119.42	315.84
house26	195.56	116.46	312.02
house27	189.96	111.80	301.77
2 SLS	219.34	72.36	291.71
3 SLS	231.34	76.48	307.82
OLS	193.22	100.70	293.92

附註：
1 SSE_1：對71年到78年真實空屋率之配適誤差平方和。
2 Pred *SSE*：對70年真實空屋率之預測誤差平方和。
3 SSE_T = SSE_1 + Pred *SSE*。

以下則是各次競賽到第200代時的最佳模型。

house21

$$V_t = (((((((((V_{t-1} \% (-6.386446 \% (V_{t-1}+V_{t-1}))) + D_t) \% M_t) \% P_t) + V_{t-1}) \% (-6.386446 \% V_{t-1})) + V_{t-1}) \% M_t) \% V_{t-1}) + V_{t-1})$$

house22

$$V_t = (Cos\ (CosCos\ (CosP_t \% (SinP_t \% (P_t * Cos\ ((Log\ (LogV_{t-1}-M_t) \% V_{t-1}) * (CosCos\ (7.419459 + (ExpV_{t-1}-V_{t-1})) \% ((V_{t-1} \% (V_{t-1} \% V_{t-1})) * LogSin\ (LogV_{t-1}-M_t))))))) + -2.321577) + V_{t-1})$$

house23

$$V_t = ((SinV_{t-1} \% V_{t-1}) + ((SinV_{t-1} \% V_{t-1}) + ((Sin\ 5.121078 \% 5.121078) + V_{t-1})))$$

house24

$$V_t = (((((((V_{t-1}+V_{t-1}) \% V_{t-1}) + V_{t-1}) + Exp\ (P_t-(V_{t-1} - ((M_t-(V_{t-1}-V_{t-1})) \% V_{t-1})))) \% D_t) + V_{t-1}) - (P_t \% M_t))$$

house25

$$V_t = (V_{t-1}-SinSinSin\ (0.330715 * ExpCosP_t))$$

house26

$$V_t = (V_{t-1}- (LogP_t \% 6.786166))$$

house27

$$V_t = (V_{t-1} * LogExpCosSinLog\ ((CosSinV_{t-1}+-3.116343) + (-3.116343 - (V_{t-1}-(8.830683 \% -2.481308)))))$$

拿這些結果與第二節的結果相比，我們可以觀察到一些有趣的現象。

第一、從前一節遺傳規畫所產生的最佳模型(house01-house07)中，我們發現“非線性進化”的程度很有限。但是，當我們將70年的資料去除，而放入78年後，我們發現此時的模型(house21-house27)都有相當明顯非線性化的進化。七個最佳模型中沒有一個是原有模型(2*SLS*或3*SLS*)的線性進化。也就是說，在進化了200代之後，2*SLS*或3*SLS*的面貌已經完全不復見了。以方程式(4)而言，70年到77的資料大體上對$V_t = f\,(V_t^n\,(M_t, D_t, V_{t-1}, P_t)$仍主張維持在一個線性的形式，任何非線性化的進化頂多只能扮演一個邊陲的角色，基本線形的格局仍可以維持。也就是70年77年的資料並沒有賦予$f(.)$一個可以高度非線性化發展的空間。但是70年資料剔

除，與78年的資料加入卻立刻賦予$f(.)$一個很容易非線性化的空間。從house21-house27的形式來看，這個影響大到可以使原來的線性結構解體。因此，從這個意義上來說，78年資料確實很獨特。對原有的線性結構而言，78年的獨特反應了一個可能存在的結構性變遷。

第二、比較表三與表十一，我們不難發現在用70年到77年的資料來預測78年時，其困難度顯然高於用71年到78年的資料來預測70年。從表三中可以看到前者預測的最低值是house03的164.38。而在表十中，預測最壞的house25也不過119.43。至於最好的2*SLS*卻不到house03的一半。這點表示71年到78年相對於70年的資訊，遠較70年到77年相對於78年的資訊豐富。而這也反應了78年這年的獨特性。因為它的獨特，所以我們能從70年到77年學到有關它的部份就相對於較我們能從71年到78年學到有關70年的部份為少。

第三、78年的獨特性還可以從表十一中SSE_1與Pred *SSE*的比較中看出。雖然非線性化的最佳模型的SSE_1是從house27的189.97到house23的196.47，都領先於2*SLS*的219.34。但是在“Pred *SSE*”上，七個模型對70年自然空屋率的預測，從house22的105.00到house25的119.46都比2*SLS*的72.37與3*SLS*的76.48高了許多。這表示了非線性的進化在樣本內雖然有必要，但是他的延伸性卻不理想。這種現象也可以由78年的獨特性找到答案。因為78年的獨特，引進了非線性模型進化的空間，這個空間於是“誤導”遺傳規畫的競賽去往“迎合”78年的獨特性去發展。也就是說，在這種情況下，最佳的模型不僅要考慮到71年以降的那個線性結構，還要考慮到78年的獨特性。2*SLS*與3*SLS*由於其“線性的限制”，並不容易做到上述要求的“面面討好”，因此在競賽中很容易的被能夠調整曲度的非線性模型而淘汰。但是，由於78年的獨特性，這些非線性模型在被“誤導”學習的情況下，對70年的預測反而大大不如原有不去刻意迎合78年的線性模型，2*SLS*或3*SLS*。

此外78年的獨特性也影響到了我們對自然空屋率的定義及其估計。在七個最佳模型中，只有house21與house24建議變數D_t的使用。另外，被用來定義自然空屋率的兩個變數M_t與D_t在house23, 25, 26與27中都雙雙落榜，而即使在house21, 22與24中，(M_t, D_t)與(P_t, V_{t-1})的可分性也不存在。這種不可分性的出現將使得我們無法再像前面一樣來估計自然空屋率。

以上的分析雖然建議78年的資料有其獨特性，但是以上分析並表示這個獨特性只是從78年才開始。衡諸現實，房地產的最近一次景氣循環是從76年起開始復甦，而在78年到達峰頂。因此，相對於70年到75年的衰退，76年到78年的繁榮正好可以將資料分割成兩個不同階段。因此76年與77年也有可能已經開始了這個獨特性。只是在“多數決”的前提下，76年與77年相對於70年到75年只有一個邊陲的地位，它們的影響只能視為“噪音”。而當78年再加入後，它們的力量才匯集成一股主流，將原來的線性結構解體。若是如此，則我們的空屋率模型就隨著景氣循環而產生結構變遷。這點在往後空屋率的研究上應有很重要的意義。因為它告訴我們景氣循還可能影響到實際空屋率與自然空屋率的高低，所以在我們進行推估時，必須將景氣循環考慮進來。否則，將會高估或低估了自然空屋率。此外，景氣循環時通常有價格持續的上漲或持續的下跌的現象。因此，我們的發現也可以說是當價格在持續的上漲或

持續的下跌時，它對自然空屋率可能有不同的影響，而這點只是人類預期行為(或理性預期)下一種很自然的現象。今後，怎麼將理性預期或學習放進自然空屋率模型中並進而了解其與景氣循環或是結構變遷的關係，是有待未來研究須深入的方向。

四、結論

本文利用遺傳規畫發現林祖嘉等(1994)所估計的空屋率應屬合理的範圍。雖然遺傳規畫所建議的空屋率高峰期與他們的估計稍有出入，但是在空屋率估計上一個更基本的問題，就是結構性變遷。因為房地產市場從70年到78年正好歷經了完整的一次循環，所以在我們資料只能涵蓋到一次循環的限制下，衰退期與繁榮期可能須視為兩個不同的階段。而我們利用遺傳規畫所作的簡單分析中，的確也有傾向支持這點。特別是遺傳規畫發現到71年到78年對於70年的資訊，相對於70年到77年對於78年的資訊要來得豐富。這點也正與這段時期內，衰退長達六年，而繁榮卻只有三年相互印証。因此，遺傳規畫的發現似乎並不離譜。值得注意的是，遺傳規畫是在不限定模型下來看結構變遷的問題，這與傳統計量有很大的不同。至於76年是否可視為結構變遷的轉折點，將是後續研究需要更深入的問題。

附　錄

附錄一、遺傳規畫的範例

以下我們便以一個具體的範例來說明遺傳規畫的執行環境及其過程。對遺傳規畫執行環境及其過程可以分兩部份來說明：第一是原始森林的創造，第二是森林的演化。

A.1 原始森林的創造

原始森林也就是遺傳規畫中第0代的森林。而由於森林是樹的集合，因此遺傳規畫的第一步就是建立一個“植樹”的程式。又因為樹是由原子(atom) (註21)所長成，植樹的程式很自然地便是由於下列兩部份所組成，第一．原子的建立，第二．原子的長成(樹的建立)。

原子：

由於原子是組成函數的基本單位，所以原子可以是原始函數(initial function)，而也可以是常數(R)或變數(X)。所以，遺傳規畫的基礎點便是一個函數集合與終點集合。茲舉例如下：

函數集合：$F = \{+, -, \times, \%, EXP, RLOG, Sin, Cos\}$ (註22)

終點集合：$T = \{X, R\}$ (註23)

$|F|$: F中元素的個數。

$|T|$: T中元素的個數。

植樹的基本規則：

在函數集合確定後，函數所要銜接的原子數目便可以依函數的本質來決定。譬如說：

1. 遇到+, -, ×, %自動要接二個原子。
2. 遇到*EXP*, *RLOG*, *Sin*, *Cos*只接一個原子。
3. 遇到終點集合中的元素，則停止。

終點集合中的元素是預備將來要配合輸入的資料。

交叉對分法：

在植樹的基本規則建立後，原始森林的建立便可依寇札(1992)所建立的交叉對分法(ramped half-and-half)所產生。這種方法基本上是在給定樹深度的限制下，讓樹依照植樹基本規則得以自由或半自由地成長。所謂自由成長，也就是並不要求樹長到給定的深度，而可以在這個深度之前自由停止。這種方法也叫成長法(grow method)。而半自由成長則要求樹一定要長到既定的深度。所以交叉對分法就是將這兩種方法等比重地使用於給定的各種深度中。譬如說，假設原始森林的規模是500棵樹，則交叉對分法的執行便可以如下的方式進行：

1. 深度最長為6 (註24)
2. 100個深度為2: 50個採取完整法(full method)創造，50個採取成長法(grow method)創造

   100個深度為3: 50個採取完整法創造，50個採取成長法創造

   100個深度為4: 50個採取完整法創造，50個採取成長法創造

   100個深度為5: 50個採取完整法創造，50個採取成長法創造

   100個深度為6: 50個採取完整法創造，50個採取成長法創造

至於不論是完整法或成長法，每棵樹的每個原子皆是依均勻分配由函數集合(F)、終點集合(T)或是函數集合與終點集合($F \cup T$)的聯集中產生。詳細說明如下：

完整法：完整的樹。每一片葉子的深度皆等於設定的最深之深度K。

建立方式：

1. 當深度＜K時，從F中選取，$P = 1/|F|$。
2. 當深度＝K時，從T中選取，$P = 1/|T|$。

成長法：每一片葉子的深度皆不大於設定的最深之深度K

建立方式：

1. 當深度＜K時，從$F \cup T$中選取，$P = 1/|F|+|T|$。
2. 當深度＝K時，從T中選取，$P = 1/|T|$。

A.2 森林的演化

在原始森林產生後，遺傳規畫便將達爾文物競天擇的原則引進到森林的演化中。然而在使用“適者生存”的原則之前，我們先要回答什麼叫做“適者”？當然，適者的答案依問題之不同而不同。以預測而言，我們可以把它定義成對資料配適的能力。也就是一棵樹(函數或模型)對資料的配適能力愈強，則其適應力便愈強。依照這個精神，我們定義幾個與“適者”

相關而具體的概念：

定義：

1. 原始適合性(raw fitness) = $r(i,G) = SSE(i,G) = \Sigma^{N}_{j=1}[V(i,G;j)-S(j)]^2$，

   其中，

   $V(i,G;j)$：在第G代中，對於環境j，函數(樹) i的估計值(G=0,..., 50; i=1,..., 500)。

   $S(j)$：第j個環境下的觀察值(j=1,..., N)。

2. 調節適合性(adjusted fitness) = $a(i,G) = 1/[1+r(i,G)]$

3. 正規適合性(normalized fitness) = $n(i,G) = a(i,G)/\Sigma_{k=1}^{M=500}a(k,G)$

4. $SSR(i,G) = \Sigma^{N}_{j=1}[V(i,G;j) - (1/N)\ \Sigma^{N}_{j=1}S(j)]^2$

5. $R^2 = SSR(i,G)/[SSR(i,G) + r(i,G)]$

基於正規適合性所象徵的相對適應能力，遺傳規畫的演化便可依下列程序進行：

1. 計算每一個個體的原始適合性、調節適合性 正規適合性與。並依照正規適合性的大小順序給予每一個個體一個序號(1,...,500)，若有相同的正規適合性，則任意給予序號。

2. 執行複製和交配的工作。

   (A) 複製：從500個個體(Gen = 0)中選取50個個體(Gen: = Gen+1)。

   選取方式：每一個個體皆附予一個被選取到的機率為$P(i,G) = n(i,G)$可重覆被選取到。

   (B) 交配：

   (a) 從500個個體中選取2個個體。

   選取方式：與複製相同。

   (b) 進行交配

   (1) 先測試每個個體的原子總數A，及葉子總數Y。

   (2) 每個原子被選到的機率P。

   當原子不是葉子時，P=0.5/(A-Y)。

   當原子是葉子時，P=0.5/Y。

   (3) 以其被選取到的機率進行選取交配點(crossover point)。對二個個體雙方皆獨立進行此步驟。

   將被選到的原子及其以下的子樹(subtree)進行雙方互換。

   若互換後的個體之深度皆不大於17，則為新一代個體(Gen = Gen+1)。

   若互換後的個體之深度大於17，則進行調整：

   i. 只有一個個體的深度大於17，則任選父母之一方取代此一個體(P=0.5)，成為新一代個體。另一深度不大於17的個體則成為新一代個體。

   ii. 二個個體的深度皆大於17，則父母雙方皆取代此二個體，成為新一代個體(Gen = Gen+1)。

3. 再進行1的步驟共224次(即共執行225次，共選出了450個個體(225對)進行交配)。

4. 另外亦可再加入突變與新生嬰兒的運作(註25)：

[A]突變：

其運作方式是在每代中依正規適合性$n(i,G) = P(i,G)$選取個體參與突變。參與突變的個體，其每一個節點(含葉子點)會發生突變的機率為p_m，突變的發生在各個節點中獨立進行互不相干。當某一節點被選為要發生突變時，便在此節點的突變集合中，隨機選取一個元素代替原有的元素。當每個節點皆完成了突變的運作後，此一新的個體便成為下一代族群中的個體。

突變集合：

{+, -, ×, %}

{*RLOG*, *EXP*, *Sin*, *Cos*}

{*X*, *R*}

[B]新生嬰兒：

在每一代中，皆加入一些根據成長法隨機創造出的個體。

5. 計算出新一代500個個體的原始適合性、正規適合性、調節適合性與R^2及給予序號，此時即完成了第一代個體的工作。

再重覆執行步驟二中2到5的全部工作，完成下一代的工作，直到終止判斷基準(Termination Criterion)被達到為止。

終止判斷基準：(1) $Gen = 50$或(2)有個體的原始適合性為0。

此程式須保存任一代的任一個體之原始適合性、調節適合性與R^2。

在完成了一代中的所有工作後，若此時有一個體滿足終止基準，則不再執行創造下一代的工作。

附錄二

表2-1 林祖嘉等(1994)結構方程式之迴歸估計(民國70-77年)

解釋變數	2SLS P_t	2SLS V_t	3SLS P_t	3SLS V_t
常數項	-2.2950**	1.9253**	-2.2963**	1.9059**
	(-6.213)	(4.724)	(-6.293)	(4.677)
V_t	0.6369**	—	0.6372**	—
	(5.014)		(5.054)	
P_t	—	-0.8535**	—	-0.9431**
		(-6.154)		(-6.887)
M_t	—	0.1586**	—	0.1554**
		(5.941)		(5.824)
D_t	—	0.0090**	—	0.0131**
		(3.106)		(4.835)
Y_t	0.1360**	—	0.1361**	—
	(16.120)		(16.713)	
N_t	0.0024**	—	0.0024**	—
	(6.705)		(7.140)	
V_{t-1}	-0.6304**	0.9053**	-0.6307**	0.9055**
	(4.991)	(34.942)	(-5.032)	(34.951)
$\bar{R}^2$	0.7117	0.9035	0.7899	0.8511
F-value	128.118**	483.449**	194.631**	295.410**
System Weighted R^2		0.9025		
Num. of Observations	207	207		

附註：有**者表示該係數在99%顯著水準下，顯著地異於0。

表2-2 林祖嘉等(1994)各縣市自然空屋率之估計，民國70-78年

單位：%

	70	71	72	73	74	75	76	77	78	平均
台灣地區	4.61	4.30	4.44	4.21	4.21	4.53	4.71	4.82	5.52	4.59
台北市	6.68	6.52	6.80	6.20	6.36	7.60	7.19	7.06	8.53	6.99
高雄市	6.27	6.03	5.73	5.90	5.84	5.95	5.56	6.28	7.35	6.10
台灣省	4.43	4.11	4.27	4.03	4.03	4.32	4.55	4.65	5.28	4.41
基隆市	4.53	4.34	4.72	4.39	4.17	5.36	4.80	4.70	4.80	4.65
台中市	6.00	6.04	6.29	5.90	5.97	7.07	7.66	7.22	8.64	6.76
台南市	6.10	6.51	5.91	5.91	6.26	5.62	5.92	6.14	6.51	6.10
台北縣	5.86	5.13	5.52	5.14	5.01	5.25	6.37	5.53	6.05	5.54
宜蘭縣	3.71	3.56	4.01	3.45	3.09	3.86	3.76	4.17	5.03	3.85
桃園縣	5.25	4.94	5.71	4.73	4.56	4.28	5.15	5.10	6.19	5.09
新竹市	4.85	4.47	4.23	4.45	4.72	5.01	6.10	6.03	6.02	5.10
新竹縣	4.30	3.32	3.39	3.57	3.78	3.42	3.21	3.75	3.89	3.63
苗栗縣	3.64	3.48	4.04	3.65	3.54	3.45	3.66	3.86	4.19	3.72
台中縣	4.30	4.22	4.35	4.06	4.44	4.91	4.30	4.08	5.07	4.42
彰化縣	3.56	3.40	3.76	3.59	4.00	3.82	3.64	3.72	7.55	4.12
南投縣	3.92	3.32	3.54	3.09	3.18	4.05	3.90	3.86	4.06	3.66
雲林縣	3.08	3.27	3.51	3.19	3.12	3.30	3.72	3.93	4.76	3.54
嘉義市	5.50	4.75	4.73	5.05	4.07	5.13	5.62	5.51	6.14	5.17
嘉義縣	3.55	3.07	3.03	3.21	2.89	2.90	3.32	3.34	3.97	3.25
台南縣	4.29	3.59	3.92	3.63	3.86	3.68	3.92	3.95	5.67	4.06
高雄縣	3.93	4.31	4.39	3.99	3.69	4.04	4.23	4.31	4.86	4.19
屏東縣	3.52	3.42	3.73	3.22	3.26	3.34	3.55	4.12	4.32	3.61
台東縣	4.38	3.89	3.41	3.32	3.77	3.96	4.26	4.43	3.78	3.91
花蓮縣	5.76	4.34	3.89	3.62	3.76	4.07	4.51	5.69	6.18	4.65
澎湖縣	2.95	3.01	3.50	3.56	3.39	4.15	3.90	4.12	3.30	3.54

附註：依3*SLS*係數推估

註　釋

註1：程序資訊理論的基礎是由Solomonoff (1964)、Kolmogorov (1965)與Chaitin (1966)三位學者在不同時間獨立地建立起來。有關它爾後的發展與應用請參考Chaitin (1987)。Cover與Thomas (1991)一書中並對程序資訊理論有專章(第七章)介紹。至於Kolmogorov (1983, a.b)對程序資訊理論與機率論之間的關係更有深入淺出的闡述。

註2：有關這點請參考陳樹衡與葉佳炫(1994)。

註3：請參見附錄一。

註4：分析空屋率的模型有很多種，例如Rosen and Smith (1983)的單一方程式估計模型，例如林祖嘉等(1994)採用的聯立方程式模型。該模型主要利用房屋下濾(filtering-down)的觀念，討論住宅所有權轉移及空屋產生等現象，例如Cooke and Hamilton (1984)。由於本研究的遺傳規畫法是在探討如何選擇一個最適的估計模型，如2*SLS*、3*SLS*或非線性模型，而非比較市場均衡模型分析法與空屋鏈分析法之差異。所以，本文不擬對空屋鏈之分析法做探討，但可以確定的是：只要空屋鏈的模型能轉換成計量模型來估計，則我們就可以用遺傳規畫法來選擇一個最是的計量模型。

註5：林祖嘉等(1994)一文中採用的空屋率係以民國69年與79年兩年的戶口普查資料為基準，其間70年到78年再用台電公司電表不足底度的戶數加以調查計算而得。

註6：請參見林祖嘉等(1994)對於美國空屋率相關研究與估計之文獻回顧。

註7：在林祖嘉等(1994)一文中，遷徙率指移出與移入戶數佔全部戶數的比例，房價離散程度則以房屋總價的標準差代表。

註8：有關Hendry的計量經濟哲學，可參考Darnell & Evans (1990)、Charemza & Deadman (1992)及Hendry (1993)。

註9：遺傳規畫作為人工智慧在統計或歸納學習上的新工具，其理論基礎可以奠基在程序資訊理論(algorithmic information theory)上。陳樹衡、葉佳炫(1994)曾從"預測難易"的角度來說明遺傳規畫與程序資訊理論之間如何存在著一個自然的連結。而有關遺傳規畫在模型選擇應用上的其他文獻，可參見Koza (1992b), Chen and Yeh (1994)。

註10：原作者是用70年到78年的資料來估計，為便於比較我們將該文估計結果列於附錄二中的表2-1。

註11：因為通常在統計模型的設定中，我們都將資料產生的機制(data generating process, DGP)分成有規律與沒有規律的兩部份。而在學習的過程中，我們當然希望我們的學習只受到規律的牽引而沒有受到非規律部份的干擾。否則，若過分重視對非規律部份的配適，則我們產生的模型或許能夠十分成功地解釋過去，但在預測未來時，卻會表現得很差。這種"只能解釋過去，而不能預測未來"的現象，在歸納性學習中，便稱為過度配適。過度配適在神經網路的學習上，是個十分普遍的問題。晚近程序資訊理論的

發展對這個問題提出了一些解決的途徑。其中最有名的便是Rissanen的最短描述法則(minimum description length principle, MDLP)。有關這點可參考Rissanen (1989)及Weigend et. al. (1992)。

註12：例如較一般的非線性模型。

註13：有關對遺傳規畫如何發動一個競爭進化過程的詳細說明，請參考陳樹衡、葉佳炫(1994)。另外，本文用來執行遺傳規畫的程式是由Pascal 4.0所寫成。為使讀者便於參考我們將該文所使用之範例列於本文附錄當中。

註14：然而，由於本文是對單一方程式，而非對整個系統方程設定競爭的環境，因此，比較成功的挑戰應該是來自於模型的"非線性的進化"而非"線性的進化"。因為若是線性的進化，我們已經知道*OLS*在單一方程式中的*SSE*自然較2*SLS*為佳。然而如果先驗的知識若是肯定系統方程式較單一方程式為重要時，遺傳規畫的設計應將單一方程排斥在進化或競賽之外。然而，不論是系統與系統的競賽，或系統與單一方程式的聯合競爭，雖然比較適合於本題的環境，但是由於在程式的設計上還有一些沒有辦法克服的問題，我們只能將這個有趣而重要的應用留待未來的研究。所以，在這個略欠嚴謹的基礎上，本文只能視為對遺傳規畫在系統模型應用上的初步性探索。

註15：在機率論上，我們知道若X、Y是$L^2(\Omega,F,P)$機率空間上的隨機變數且$(_Y^X)$為聯合常態分配，則$E(Y|X)$必為線性的形式。但是在一般的情況下，$E(Y|X) = a+bX+\phi(X)$，其中$\phi(X)$是隨機變數X的Borel函數。然而目前我們仍然不清楚遺傳規畫在尋找$\phi(X)$的效率如何，但是身為一個尋找的指導原則，遺傳規畫應可用於引導神經網路的進化(見Koza (1992a), Chap 19)，而我們預期這樣的應用將可增加對$\phi(X)$尋找的效率。

註16：此點正是我們在前面註解中提到的。如果遺傳規畫所帶動的是"線性化的進化"，則*OLS*將比2*SLS*更為適合。針對這點我們也曾經以*OLS*作為原始模型而放入競賽中，結果發現*OLS*的生存能力十分強韌，在我們所有的模擬當中，*OLS*都能存活下來，而且在*OLS*的基礎上"非線性的進化"也很難產生。但是，*OLS*所估計預測的自然空屋率，在某些行政區中，竟然有負的。所以，若是我們將自然空屋率必須為非負數作為一個競賽的限制，則*OLS*將無法存活下來。

註17：但是模擬7所產生的最佳方程式並不支持林祖嘉等(1994)一文所定義的自然空屋率，這點在後面會再提到。

註18：至於為什麼2*SLS*或3*SLS*的模型在預測上會較遺傳規畫所建議的模型或*OLS*差這麼大是個很有趣的問題。從模型的形式來看，house01-05及house07均建議一個較低的價格係數及一個較低的常數，且多數模型都不主張放入D_t。所以，預測的差異可能來自於P_t與D_t係數上的不穩定或是結構方程式在錯誤設定下的嚴重低效率。關於後者，是對結構方程式或單一方程式取捨時最關鍵的問題。通常在結構方程式設定不正確的情況下，其漸近效率可能較單一方程式要來得差。

註19：即$V_t^n(.)$與$f(.)$之間透過變數M_t、D_t有一個遞延的(recursive)結構。

註20：其實表七只是將林祖嘉等(1994)的資料中去除掉78年後所得到的結果。但是，其對自然空屋率在台灣地區跨年平均值的估計為5.98，較林祖嘉等(1994)的4.59高了許多，這反應出78年資料的獨特性。這點我們在第三節中還會深入討論。

註21：在此我們並不稱其為葉子，因為葉子是用來代表樹的最底一層。所以對中間層我們仍稱其為原子。另外這種稱法也符合LISP (List Programming)語言的用法。

註22：函數集合中函數的運作皆按照一般的方式，唯獨除法(%)與對數(*RLOG*)為了便於程式之進行，已將其修正為：

%：令$k/0 = 1, \forall\ k \in R$

$$\mathrm{RLOG} = \begin{cases} 0 & x=0 \\ LOG(x) & x>0 \\ LOG(-x) & x<0 \end{cases}$$

註23：終點集合中的常數係由隨機產生。

註24：此處的意義為從簡單的模型入手。

註25：此時經由交配所產生的個體便不是450個了，有一些是經由突變與新生嬰兒所產生。

參考文獻

林祖嘉　張金鶚　彭建文

1994 〈台灣地區空屋率與房價調整之均衡分析〉，發表於“八十二年度經濟學門專題計畫研究成果發表會”，國科會人文處主辦。

陳樹衡　葉佳炫

1994 〈遺傳規畫與科技預測〉，袁建中主編《1994年中華民國科技管理研討會論文集》，377-391，中華民國科技管理學會。

Chaitin, G. J.

1966 "On the Length of Programs for Computing Finite Binary Sequences", Journal of the Association of Computing Machinery, 13: 547-569.

Chaitin, G. J.

1987 Information Randomness & Incompleteness: Papers on Algorithmic Information Theory, 2nd ed. Singapore: World Scientific Publishing Co. Pte. Ltd.

Charemza, W. W. and D. F. Deadman

1992 New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression, UK: Edward Elgar.

Chen, S. and C. Yeh

1994 "On the Competitiveness of the Quantity Theory of Money: A Natural-Selection Test Based on Genetic Programming", Conference on Forecasting, Prediction and Modelingin Statistics and Econometrics and Regional Meeting of International Society for Bayesian Analysis, Hsinchu, Taiwan, December 12-14.

Cooke, T., and B. W. Hamilton

1984 "Evolution of Urban Housing Stocks: A Model Applied to Baltimore and Houston", Journal of Urban Economics, 17: 317-337.

Darnell, A. C. and J. L. Evans

1990 The Limits of Econometrics, Edward Elgar.

Hendry, D. F.

1993 Econometrics: Alchemy or Science?, UK: Blackwell.

Kolmogorov, A. N.

1965 "Three Approaches to The Definition of The Concept "Amount of Information"", Information Transmission. 1: 3-11.

1983a "Combinatorial Foundations of Information Theory and the Calculus of Probabilities", Russian Mathematical Surveys. 38:4, 29-40.

1983b "On Logical Foundations of Probability Theory", in K. Ito and J. V. PrRokhorov (eds.), Probability Theory and Mathematical Statistics. Lecture Notes in Mathematics. Vol. 1021, Springer-Verlag: 1-5.

Koza, J. R.

1992a "A Genetic Approach to Econometric Modelling", in Economics and Cognitive Science. ed. Bourgine, P. and B. Walliser. UK: Pergamon Press.

1992b Genetic Programming: On the Programming of Computers by Means of Natural Selection. USA: The MIT Press.

Rissanen, J.

1989 Stochastic Complexity in Statistical Inquiry. Singapore: World Science Publishing Co. Pte. Ltd.

Rosen, K. T., and L. B. Smith

1983 "The Price-Adjustment Process for Rental Housing and the Natural Vacancy Rate", American Economic Review. 73: 779-786.

Solomonoff, R. J.

1964 "A Formal Theory of Inductive Inference. Part II", Information and Control. 7: 224-254.

Thomas M. Cover and Joy A. Thomas

1991 Elements of Information Theory. USA: John Wiley and Sons, Inc.

Weigend, A. S., B. A. Huberman, and D. E. Rumelhart

1992 "Predicting Sunspots and Exchange Rates with Connectionist Networks", in Nonlinear Modeling and Forecasting. ed. M. Casdagli and S. Eubank Addison-Wesley.