

# 時間數列分析之轉換函數模式 在學生數預測上之應用

馬 信 行\*

## Abstract

In order to improve the predictions made by the single-variate ARIMA time series analysis, the transfer function model is introduced to forecast the number of population, the student number of elementary schools as well as of junior high schools in the ROC. In deciding the optimal orders of  $\delta(B)$  and  $w(B)$ , all possible combinations of  $r=0,1,2,3$ ,  $s=0,1,2,3$  and additionally the logical inference were considered and tried. The criteria used in checking the adequacy of different models were: standard error estimate, AIC, SBC, autocorrelation function of residuals, inverse autocorrelation function of residuals, partial autocorrelation function of residuals, cross correlation function of residuals with prewhitened input series, and the difference between the predicted value and the real value. It was found that, in determining the b period, which is that the effect of independent series on dependent series begins to appear, in addition to the cross correlation function

---

\* 作者為本校教育系教授。

between two series, the logical inference should also be included in consideration. The results of the study show that with the prerequisite of high correlation between two series, the transfer function model can be fruitfully applied to quantitative predictions in the educational planning.

### 摘要

本研究以二元時間數列分析（轉換函數模式）來預測國小學生數，國中學生數，及年中人口數。預期能有更準確的預測。在 $w(B)$  與  $\delta(B)$  之秩的決定上，本研究採 $s=0, 1, 2, 3; r=0, 1, 2, 3$  之組合，加上實際推理的考量，一一嘗試，以尋求最佳模式。選擇模式所用指標有：殘差的標準誤， $AIC$ ,  $SBC$ ，殘差的自我相關，殘差的逆自我相關，殘差的偏自我相關，殘差與「白化後的自變數列」之交叉相關函數，及預測值與實際值之差。在嘗試錯誤過程中發現， $b$  期（自變數列對依變數列的影響開始顯現的期）的決定不能光靠自變數列與依變數列的交叉相關，而應加上實際推理之考量。本研究顯示，在兩數列有高相關的前提下，轉換函數模式可有效的應用於教育計量上的預測，這對於教育計畫的科學化將有重大的貢獻。

馬信行（民67 a）曾經以單變量時間數列預測我國各級學校未來學生數，同時也預測了我國未來人口數及出生數。當時也遭遇到一些困難，例如對我國國小學生數及國中學生數的預測，不敢冒然預測十二年。因為國中，小學受出生數的影響很大，依據單變量時間數列模式，國中學生數在1997年可能達1,600,139人，這顯得不合邏輯。因為一年出生人數以40萬人計算，充其量（全部就學）三年的國中生頂多也達120萬人。不可能達160萬人（頁131）。本研究擬以轉換函數模式（transfer function models）來克服這些困難。因為國小學生數受出生數影響很大，故以出生數為自變數列（independent variable series 或稱 exogenous variable，或 input variable），以國小學生數為依變數列（dependent variable series

## 時間數列分析之轉換模式在學生數預測上之應用

，或稱 endogeneous variable，或 output variable)。另外以國小畢學生為自變數列，以國中學生數為依變數列，因為最近國小升國中的升學率也幾達99.9%。

在討論轉換模式之前，先就單變量時間數列預測之情形加以檢討。表1是馬信行（民76 a）在三年前所作預測，至今已有三個實際值出現。預測準確度以相對誤差來表示。所謂相對誤差是（預測值－實際值）／實際值。

**表1 單變量時間數列分析預測準確度之檢核〔預測值取自馬信行（民76 a），實際值取自教育部（民78）〕**

預測項目 年度	年 中 人 口 數 (B 案)			出 生 人 口 數 (B 案)		
	預 測 值	實 實 值	相 對 誤 差	預 測 值	實 實 值	相 對 誤 差
1985	19143872	19135283	0.000449	368015	345053	0.0665
1986	19412120	19356331	0.002882	366850	308187	0.1903
1987	19677500	19563611	0.005821	365371	313062	0.1671
1988	19940040	19788212	0.007673	363898	341054	0.067

預測項目 年度	幼 稚 園 學 生 數 (A 案)			國 小 學 生 數 (B 案)		
	預 測 值	實 實 值	相 對 誤 差	預 測 值	實 實 值	相 對 誤 差
1986	236861	238428	0.006572	2360008	2364438	0.001874
1987	242526	250179	0.030590	2392271	2400614	0.003475
1988	248553	248498	0.000221	2419498	2407166	0.005123

預測項目 年度	國 中 學 生 數			高 中 學 生 數 (A 案)		
	預 測 值	實 實 值	相 對 誤 差	預 測 值	實 實 值	相 對 誤 差
1986	1047302	1052993	0.005405	199191	200599	0.007019
1987	1032696	1053923	0.020141	203957	206019	0.010009
1988				209728	208994	0.003512

表 1 [續 1]

預測項目 年度	高 職 學 生 數			專科與大學部學生(A案)		
	預測值	實際值	相對誤差	預測值	實際值	相對誤差
1986	438856	437924	0.002128	430660	429211	0.003376
1987	457042	447328	0.021716	444674	449543	0.010831
1988	467615	444232	0.052273	458229	479189	0.043741

預測項目 年度	碩士班學生數(A案)			博士班學生數(A案)		
	預測值	實際值	相對誤差	預測值	實際值	相對誤差
1986	11747	11294	0.0401	2066	2143	0.0359
1987	12832	12426	0.0327	2355	2695	0.1262
1988	13885	14119	0.0166	2650	3222	0.2122

預測項目 年度	特殊教育學生數(B案)			普通補校(B案)		
	預測值	實際值	相對誤差	預測值	實際值	相對誤差
1986	2938	3039	0.0332	61113	61768	0.0106
1987	2985	2948	0.0132	60939	59104	0.0310
1988	2987	2998	0.0037	60778	56910	0.0680

預測項目	職業補校		
年 度	預測值	實際值	相對誤差
1986	220156	243931	0.0975
1987	226551	238963	0.0519
1988	233133	242784	0.0397

表 1 中的年中人口數，B案比較與實際值接近（馬信行，民 76 a，頁 125）。因為 A案有常數項，有常數項的模式較能反映過去的成長率，沒有常數項的模式較能反映最近的趨勢。最

近出生數有減緩的趨勢，故B案較適合。1985年的相對誤差是0.000449，比行政院經建會人力規劃小組（民73）之預測更為準確。其低推估之相對誤差為0.008765。出生數的預測較不準確，相對誤差超過5%。不過行政院經建會人力規劃小組預測的更不準，1985年低推估的相對誤差就達0.10997，1986年達0.2265，高推估與中推估的相對誤差更大（參閱馬信行，民76a，頁127）。原因可能是我國在1985年通過優生保健法，對人口減緩有立竿見影之效，而1988年是龍年，出生數又見回升。但在1988年家庭計畫單位也怕出生人口減少後造成人口斷層現象而改變宣傳口號為「三個孩子不算多」。到底1988年出生人口的回升是政策使然或是民間的龍年迷信，有待日後驗證。幼稚園由於尚未全部入學，故不受出生數的影響。A案較接近實際值。1986年的相對誤差只有0.006572。可算滿意。國小學生數（B案）在1986年的相對誤差也僅達0.001874。國中學生數預測的相對誤差雖沒超過5%，但預測方向是逐年下降，可是實際值是已有回升現象，故模式有待改變。高中學生數預測的相對誤差三年內也都在1%以內。高職學生數包含了以往的師範生，但不包含五專前三年的學生，因為教育部的教育統計把五專前三年視為專科學校學生。高職學生數三年的相對誤差也在5%以內。專科與大學部學生數，碩士班學生數預測值的相對誤差都在5%以內，惟博士班1987與1988預測值的相對誤差超過5%。也就是這兩年來博士班學生數在國內大幅成長，其成長率大於以往的成長率。特殊學校學生數三年預測的相對誤差也都在5%以內。普通補校（包含國小補校與高中補校），1988年的預測相對誤差超過5%，達6.8%。職業補校（含專科補校及空大）預測的相對誤差，1986及1987兩年都超過5%。最近由於延長國民教育的政策性改變，補校學生數大幅成長，導致預測值有低估的現象。

綜合以上的檢討，預測相對誤差三年內皆在5%以內的項目計有：年中人口數，幼稚園學生數，國小學生數，國中學生數，高中學生數，專科與大學部學生數，碩士班學生數，特殊教育學生數等。國小學生數由於就學率已達99.9%，而國小畢業生升入國中者也達99%。如以出生數當自變數列，國小學生數當依變數列，用時間數列分析的轉換函數模式來預測國小學生數，是否會有更精確的預測？同樣的，以國小畢業生數為自變數列，以國中學生數為依變數列，用轉換模式來預測國中學生數，是否可避免預測方向相反的現象？

Helmer & Johansson (1977) 應用轉換函數模式來了解廣告對銷售的影響。他們用40

年的數據來確定模式，另保留14年的數據，用來比較各種預測方法所預測出來的預測值的準確性。預測準確性以殘差（實際值與預測值之差）的均方和（mean square error）來判斷，殘差均方和愈小則表示預測愈準確。結果發現在預測上，轉換函數模式比單變量有更小的殘差均方和。與其他的計量經濟方法（例如 Palda 使用計量經濟方法所建立的累積效果模式）相較，亦有較高的預測力，即殘差均方和較小。Emstead (1977) 也嘗試以領先綜合指標當自變數列，來預測季的股票價格，他認為 Box & Jenkins 的轉換函數模式可有效的應用於很多商業與經濟方面的預測。Hopwood & McKeown (1981) 比較單變量與多變量的預測力，單變量是以公司的每股所得 (earnings per share) 數列，多變量是以每股所得為依變數列，以市場所得指標 (market earnings index) 為自變數列，建立轉換函數模式。所取樣本有267 個公司。每個公司有自1962到1977年的每股所得數列與市場所得指標數列的連續性資料。保留四年（1974到1977）的資料用來檢定預測力。預測力是以（實際值－預測值）/ 實際值。他們稱之為平均絕對誤差百分率。結果發現，轉換函數模式的預測力比單變量的時間數列模式為高，也就是預測誤差較小。

Downing, Jack, & Westley (1883) 研究 1949 年 10 月到 1959 年 10 月間共 522 週的 Chattahoochee 河上游平均流量（每分鐘立方公尺）。在 1955 年 10 月開始建造 Sidney Lanier 水壩，水壩建好之後河流的平均流量亦大減。從未建水壩之前每分鐘 58.1 立方公尺降到平均每分鐘 33.0 立方公尺。未建水壩之前河流平均流量變異很大，顯然受降雨量的影響。而建完水壩之後，變異就較小，顯然水壩有調節流量的功能。如將流量數據分為兩部分——建水壩之前與之後，則發現在建水壩之前，用轉換函數模式（以每週平均流量為依變數列，以降雨量為自變數列）來預測每週平均流量會比用單變量時間數列模式更準確，但在建完水壩之後，降雨量對河水平均流量的影響就消失了。這可顯示水壩確有調節雨量的功能。

在教育上時間數列的應用尚很少，轉換函數模式的應用更罕見。本文擬嘗試引進轉換函數模式來預測學生量的成長，預期對教育計畫的科學化有推動作用。

## 方 法

## 時間數列分析之轉換模式在學生數預測上之應用

轉換函數模式的方法仍以 Box & Jenkins (1976) 敘述得最詳細，另外 Vandaele (1983) 及 Montgomery & Weatherby (1980) 亦有按步就班的介紹。為便於在國內教育界推廣，茲將轉換函數模式的方法簡述於下。

## 一、轉換函數模式的一般模式

轉換函數模式的一般模式是

這很像迴歸方程，事實上，其意義也近似迴歸方程，只不過迴歸方程裏的變項是隨機變項，不能有自我相關 (autocorrelation)，因為有自我相關的話，會使迴歸係數的標準誤增大，當相關趨近於 1 時，標準誤會趨於無限大，導致原本有顯著的迴歸係數，變為沒顯著（參閱馬信行，民76 b），而轉換函數模式要將有自我相關的數列先轉化為無自我相關的數值。公式(1)中的  $v(B)$  是多項式。 $v(B) = v_0 + v_1 B + v_2 B^2 + \dots$ ， $B$  是後移算子 (backward shift operator)， $BX_t = X_{t-1}$ ， $B^2 X_t = X_{t-2}$ ， $X_t$  是自變數列第  $t$  期的觀察值， $Y_t$  是依變數列第  $t$  期的觀察值。 $N_t$  是殘差數列（或稱干擾數列）第  $t$  期的數值。 $b$  是經過  $b$  期之後， $X_t$  對  $Y_t$  所發生的作用開始顯現。 $v(B)$  稱為衝擊反應函數 (impulse response function)。實際上， $v(B)$  是兩個固定多項式的比值 (ratio)。

其中， $w(B) = w_0 - w_1 B - w_2 B^2 - \dots - w_s B^s$

$$\delta(B) \equiv \delta_0 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r$$

<sup>路</sup>3. 標準化爲 1，也不會失去其一般性（Vandaele, 1983, 頁260），故常寫爲

$$\delta(B) \equiv 1 - \delta_1 B - \delta^2 B_2 - \cdots - \delta_r B^r \circ$$

$N_t$  是一個隨機變數，包含著所有無名的干擾因子，它可能有，也可能沒有自我相關。 $X_t$  與  $Y_t$  如果有趣勢（不平穩），則必須先差分為平穩的數列之後才能進行分析。差分 (differencing) 是數列中前後兩個觀察值相減，即  $(1-B)Y_t = Y_t - Y_{t-1}$ ，或  $(1-B)X_t = X_t - X_{t-1}$ 。由於  $X_t$ ,  $Y_t$ ，及  $N_t$  都可能有自我相關，故必須化為沒有自我相關，這種過程稱為白化 (prewhiten)。因為  $X_t$  是自變數列，故先求  $X_t$  的單變量 ARIMA (Autoregressive Integrated Moving Average) 模型。

grated Moving Average) 模式，以之來自化  $X_t$ ,  $Y_t$ , 及  $N_t$ 。首先是白化  $X_t$ ,

$$\phi_x(B)X_t = \theta_x(B)\alpha_t$$

其中  $\alpha_t$  是經過白化以後的自變數列。 $\phi_x(B)$  是自變數列自我迴歸的多項式參數， $\theta_x(B)$  是自變數列的移動平均的多項式參數。

然後以自變數列的單變量 *ARIMA* 模式來白化依變數列及殘差數列，這是因為所持假設是：依變數列的變化是受自變數列的影響。

$$\phi_x(B)Y_t = \theta_x(B)\beta_t$$

由公式(1)的 $Y_t$ 代入公式(4)，暫且不考慮 $b$

$$\beta_t = \phi_{\hat{x}}(B) \theta_{\hat{x}}^{-1}(B) [v(B)X_t + N_t]$$

$$= \phi_{\tilde{x}}(B) \theta_{\tilde{x}}^{-1}(B) v(B) X_t + \phi_r(B) \theta_r^{-1}(B) N_t$$

再將公式(3)代入

## 二、衝擊反應函數的求法

將公式(5)等號兩邊各乘以  $\alpha_{t-k}$ ，並取期望值，則得  $\beta_t$  與  $\alpha_{t-k}$  的共變異數。

$$E[\beta_t \alpha_{t-h}] \equiv E[u(B)\alpha_t \alpha_{t-h} + \phi(B)\theta^{-1}(B)N_{t+h-t-1}]$$

因為  $N_t$  與  $\alpha_{t-k}$  是獨立的（不相關的），故  $E[N_t X_{t-k}] = 0$

在公式(6)中， $\alpha_t$ 與 $\alpha_{t-k}$ 都是經白化過的，沒有自我相關的白雜音 (white noise)，故 $\alpha_t$ 與 $\alpha_{t-k}$ 是獨立的，其共變異數為零。故 $E[\alpha_t\alpha_{t-k}] = 0$ ，但當 $k=0$ 時， $E[\alpha_t\alpha_{t-k}] = E[\alpha_t\alpha_t] = \sigma^2$ 。

時間數列分析之轉換模式在學生數預測上之應用

基於時間數列的變異數同質性的假設：各期變異數等於母羣體的變異數，也就是  $E[\alpha_k \alpha_k] =$

$$E[\alpha_{t-k}\alpha_{t-k}^2] = \sigma_\alpha^2$$

公式(6)中， $E[\beta_t \alpha_{t-k}]$  也可用  $COV_{\alpha\beta}(k)$  來表示， $COV$  是 covariance 的縮寫，公式(6)可寫成  $COV_{\alpha\beta}(k) = v_k \sigma_\alpha^2$ ，移項後得

根據相關係數的公式， $\alpha_{t-k}$  與  $\beta_t$  之相關係數是

$$\rho_{\alpha t-k \beta t} = \frac{COV_{\alpha \beta}(k)}{\sigma_{\alpha} \sigma_{\beta}}$$

其中  $\sigma_\alpha$  為  $\alpha_{t-k}$  的標準差， $\sigma_\beta$  是  $\beta_t$  的標準差。 $\rho_{\alpha_{t-k}\beta_t}$  也可用  $\rho_{\alpha\beta}(k)$  來表示。

將公式(8)代入公式(7)，得

公式(9)是以母羣體的參數來表示，如以樣本來估計，則

其中  $v_k$  為第  $k$  期的衝擊反應函數， $r_{\alpha\beta}(k)$  為  $\alpha_{t-k}$  與  $\beta_t$  的交叉相關函數 (cross correlation function, 簡稱  $ccf$ )， $s_\alpha$  為自變數列白化後的標準差， $s_\beta$  為依變數列白化後的樣本標準差，公式(10)中， $r_{\alpha\beta}(k)$ ， $s_\alpha$ ，及  $s_\beta$  都可從數列的觀察值求得，所以  $v_k$  也可間接的從公式(10)計算出來。

### 三、從衝擊反應函數求 $w(b)$ 與 $\delta(b)$

如果暫且不管  $N_t$ ，公式(1)及公式(2)可導出

$Y_t = \frac{w(B)}{\delta(B)} X_{t-b}$ , 也可寫成

將公式(11)與公式(2)比較，如果現在考慮  $b$ ，則

$$v(B) = \frac{w(B)B_b}{\delta(B)}$$

將各個多項式展開，並移項，得

因為 $\delta_r B^r$ 及 $w_s B^s$ 以後的項都不會發生作用，或僅發生些微作用，故可忽視。

將公式(12)等號左邊乘式相乘，可得出一個有系統性的式子。

$$(1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r)(v_0 + v_1 B + v_2 B^2 + \cdots + v_j B^j) \\ = v_0 + (v_1 - \delta_1 v_0)B + (v_2 - \delta_1 v_1 - \delta_2 v_0)B^2 + \cdots + (v_j - \delta_1 v_{j-1} - \delta_2 v_{j-2} - \cdots - \delta_r v_{j-r})B^j$$

因為  $\delta_{r+1}$  及以後的諸項被忽視而當作零，故  $\delta_r v_{j-r}$  以後可忽視，而當作零。

將公式(12)等號右邊的也乘開，得

因為公式(13)與公式(14)相等，故找  $B$  係數相對應者，發現：當  $j < b$  時，及  $j > b + s$  時，兩公式沒有對應項。

當  $j < b$  時， $v_j$  還沒發生作用，衝擊反應函數為零，即  $v_j = 0$ 。當  $j = b$  時，有一對應項。

$$(v_j - \delta_1 v_{j-1} - \delta_2 v_{j-2} - \cdots - \delta_r v_{j-r}) B^j = w_0 B^b$$

$$v_j - \delta_1 v_{j-1} - \delta_2 v_{j-2} - \cdots - \delta_r v_{j-r} = w_0$$

移項後得

當  $j = b + s$  時，則是最後一項相對應。

$$(v_j - \delta_1 v_{j-1} - \delta_2 v_{j-2} - \cdots - \delta_r v_{j-r}) B^j = -w_s B^{b+s}$$

$$v_j - \delta_1 v_{j-1} - \delta_2 v_{j-2} - \dots - \delta_r v_{j-r} = -w_s = -w_{i-h},$$

因  $j = b + s$ ，故  $s = j - b$ ，上式移項得

當  $b < j < b + s$  時，也各有其對應項，其對應情形與當  $j = b + s$  時同，故適用公式(16)。故

## 時間數列分析之轉換模式在學生數預測上之應用

公式(16)適用於當  $b+1 \leq j \leq b+s$  時，當  $j > b+s$  時，因為沒有對應項，故

$v_j - \delta_1 v_{j-1} - \delta_2 v_{j-2} - \cdots - \delta_r v_{j-r} = 0$ ，移項得

因為  $r$  與  $s$  不易得知，但一般來講， $r$  與  $s$  在 2 或以下，已足夠用來建立適當模式 (Box & Jenkins, 1976, 頁386)。

公式(15)到公式(17)中的  $v_j$  可依公式(10)計算出來。

求 $\delta_1, \delta_2, \dots, \delta_r$ 之值的方法，一般是先建立公式(17)的式子，後用聯立方程式解出 $\delta_1, \delta_2, \dots, \delta_r$ 之值，後將 $\delta_1, \delta_2, \dots, \delta_r$ 之值代入公式(16)，求出 $w_1, w_2, \dots, w_s$ 之值，也將 $\delta_1, \delta_2, \dots, \delta_r$ 之代值入公式(15)求 $w_0$ 之值。

到底  $\delta_r$  之  $r$  是等於 0, 1, 或 2,  $w_s$  之  $s$  是等於 0, 1, 或 2, 可分別嘗試組合, 作出九個模式,  $(r, s) = (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2)$ , 比較這九個模式, 那個模式的平均標準誤最小, 或用其他指標, 選出最佳模式 (參閱 Box & Jenkins, 1976, 頁349-351)。

#### 四、轉換函數模式之計算舉例

爲解釋上面所列計算過程，茲舉一例說明之，原始分數轉載於 Box & Jenkins (1976, 頁532-533)，它包括三行資料， $t$  是期數， $X$  是瓦斯氣（甲烷）進氣率（每分鐘立方呎）， $Y$  是廢氣中 $\text{CO}_2$  百分率。由於原始數列相當平穩，顯示出不必差分，自變數列的單變量 ARIMA 模式經辨認是  $AR(3)$  的模式，其參數經電腦計算出，是  $\phi_1=1.97$ ,  $\phi_2=-1.37$ ,  $\phi_3=0.34$ ，用此模式來自化自變數列，得

$$\alpha_t = (1 - 1.97B + 1.37B^2 - 0.34B^3)X_t \ ,$$

白化後的數列之變異數 ( $s_a^2$ ) 為 0.0353，標準差是其平方根， $s_a = 0.188$ 。

以自變數列的 $AR(3)$ 來自化依變數列，得

$$\beta_t = (1 - 1.97B + 1.37B^2 - 0.34B_3)Y_t$$

$\beta_t$  的標準差為 0.358， $\alpha_{t-k}$  與  $\beta_t$  的交叉相關函數前 10 期列於表 2。衝擊反應函數也列在第四列，第三列是檢定交叉相關係數是否達顯著的標準誤，它等於  $1/\sqrt{n}$ ，其中  $n$  是樣本數，本例

的樣本數為 270， $v_k$  是從交叉相關係數， $s_\alpha$ ，及  $s_\beta$  求出的，如果  $k=4$ ，則根據公式(10)

$$v_4 = \frac{0.358}{0.188}(-0.33) = -0.63。$$

表 2 從白化後的瓦斯爐進氣量及廢氣中二氯化碳濃度之時間數列所估計出來的交叉相關函數及衝擊反應函數

期 數	0	1	2	3	4	5	6	7	8	9	10
$r_{\alpha\beta}(k)$	-0.01	0.05	-0.03	-0.28	-0.33	-0.46	-0.27	-0.17	-0.03	0.03	-0.05
$\hat{\sigma}(r)$	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
$v_k$	-0.02	0.10	-0.06	-0.53	-0.63	-0.88	-0.52	-0.32	-0.06	0.06	-0.10

(取自 Box & Jenkins, 1976, 頁382)

從表 2，可看出  $r_{\alpha\beta}(k)$  從第 3 期才開始顯著，也就  $X_{t-3}$  的甲烷進氣率會影響  $Y_t$  的廢氣中  $\text{CO}_2$  濃度，所以  $b=3$ 。現在  $v_k$  為已知，可用公式(15)到公式(17)求  $\delta_r$  與  $w_s$  之參數估計值。

①當  $j < 3$  時， $v_j = 0$

②當  $j = 3$  時，公式(15)適合，

$$v_j = \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots + \delta_r v_{j-r} + w_0$$

由於  $v_2 = v_1 = v_0 = 0$ ，故  $v_3 = w_0$

③當  $4 \leq j \leq s+3$  時公式(16)適合，設  $w_2$  為最高秩 (order)，因為一般假定  $s=2$  已足夠。

$$v_j = \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots + \delta_r v_{j-r} - w_{j-b}$$

$$v_4 = \delta_1 v_3 - w_1$$

$$v_5 = \delta_1 v_4 - \delta_2 v_3 - w_2$$

④當  $j > s+b$  時，公式(17)適用。

$$v_j = \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots + \delta_r v_{j-r}，假定 \delta_2 為最高秩。$$

$$v_6 = \delta_1 v_5 + \delta_2 v_4$$

$$v_7 = \delta_1 v_6 + \delta_2 v_5$$

時間數列分析之轉換模式在學生數預測上之應用

$v_3$  到  $v_7$  以表 2 的數據代入。首先從  $v_6$  與  $v_7$  的聯立方程式解出  $\delta_1$  與  $\delta_2$

$$\begin{cases} -0.52 = 0.88\delta_1 - 0.63\delta_2 \\ -0.32 = -0.52\delta_1 - 0.88\delta_2 \end{cases}$$

解出  $\delta_1 = 0.57, \delta_2 = 0.02$ 。

$$w_0 = v_3 = -0.53$$

$$v_4 = \delta_1 v_3 - w_1, w_1 = 0.57(-0.53) - (-0.63) = 0.33$$

$$v_5 = \delta_1 v_4 + \delta_2 v_3 - w_2, w_2 = 0.57(-0.63) + 0.02(-0.53) - (-0.88) = 0.51$$

其次計算干擾數列  $N_t$ ，我們仍然以自變數列的單變量 ARIMA 模式來白化  $N_t$ ，令白化過的干擾數列為  $\epsilon_t$ 。

$$\epsilon_t = (1 - 1.97B + 1.37B^2 - 0.34B^3)N_t$$

之後找  $\epsilon_t$  的 ARIMA 模式，發現 MA(1) 適合，因為白化過的殘差之 acf 第一期達顯著。

$$\epsilon_t = (1 - \theta_1 B)a_t$$

在單變量模式中，MA(1) 係數可從相關係數求出 (Vandaele, 1983, 頁43)。

$$r_1 = \frac{-\theta_1}{1 + \theta_1^2}$$

而  $r_1 = -0.38$  (電腦求出)， $-0.38 = \frac{-\theta_1}{1 + \theta_1^2}$ ，得出  $0.38\theta_1^2 - \theta_1 + 0.38 = 0$

以  $X = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$  的公式解出一元二次方程式，得  $\theta_1 = 0.45$  或  $\theta_1 = 2.17$ 。

一般 ARIMA 的參數都小於 1，故取  $\theta_1 = 0.45$

$$\epsilon_t = (1 - 1.97B + 1.37B^2 - 0.34B^3)N_t = (1 - 0.45B)a_t$$

$$a_t = \frac{1 - 1.97B + 1.37B^2 - 0.34B^3}{1 - 0.45B}N_t = (1 - 1.52B + 0.69B^2)N_t$$

$$N_t = \frac{1}{1 - 1.52B + 0.69B^2}a_t$$

其中 1.52 相當於  $\phi_1$ ；-0.69 相當於  $\phi_2$ 。

$v(B)$  與  $N_t$  都求出之後，便可代入公式(1)。

$$\begin{aligned}
 Y_t &= v(B)X_{t-b} + N_t = \frac{w(B)}{\delta(B)}X_{t-b} + N_t \\
 &= \frac{w_0 - w_1B - w_2B}{1 - \delta_1B - \delta_2B^2}X_{t-b} + \frac{1}{1 - \phi_1B - \phi_2B^2}a_t \\
 &= \frac{-0.53 - 0.33B - 0.51B^2}{1 - 0.57B - 0.02B^2}X_{t-3} + \frac{1}{1 - 1.52B + 0.69B^2}a_t
 \end{aligned}$$

這是模式的起始估計值 (initial estimates)，再以疊代法 (iteration) 求，直到最後收斂為止，所得的參數估計值便是最後模式的參數估計值 (Box & Jenkins, 1976, 頁396)，疊代法所得的最後模式是

$$Y_t = \frac{-(0.53 + 0.37 + 0.51B^2)}{1 - 0.57B - 0.01B^2}X_{t-3} + \frac{1}{1 - 1.53B + 0.63B^2}a_t$$

此式可供預測之用，在預測之前，先將式子通分，展開，在參數中，數值太小的，可省略，參數間相關高的 (.8 以上)，可省去其中一個。在  $\delta(B)$  中  $-0.01B^2$  很小，可省略，通分後得

$$\begin{aligned}
 Y_t(1 - 2.1B + 1.5021B^2 - 0.3591B^3) &= [-0.53 + 0.4409B - 0.2829B^2 \\
 &\quad + 0.5724B^3 - 0.3213B^4]X_{t-3} + (1 - 0.57B)a_t, \quad \text{展開後移項得}
 \end{aligned}$$

$$\begin{aligned}
 Y_t &= 2.1Y_{t-1} - 1.5021Y_{t-2} + 0.3591Y_{t-3} - 0.53X_{t-3} + 0.4409X_{t-4} \\
 &\quad - 0.2829X_{t-5} + 0.5472X_{t-6} - 0.3213X_{t-7} + a_t - 0.57a_t
 \end{aligned}$$

如果要往後預測  $l$  期，便把  $t$  改為  $t+l$  即可。

$$\begin{aligned}
 Y_t(l) &= [Y_{t+l}] = 2.1[Y_{t+l-1}] - 1.5021[Y_{t+l-2}] + 0.3591[Y_{t+l-3}] \\
 &\quad - 0.53[X_{t+l-3}] + 0.4409[X_{t+l-4}] - 0.2829[X_{t+l-5}] \\
 &\quad + 0.5472[X_{t+l-6}] - 0.3213[X_{t+l-7}] + [a_{t+l}] - 0.57[a_{t+l-1}]。
 \end{aligned}$$

其中  $Y_{t+l-1}$  到  $Y_{t+l-3}$ ;  $X_{t+l-3}$  到  $X_{t+l-7}$  都是數列的觀察值， $a_{t+l}$  與  $a_{t+l-1}$  如未知，則視為零。預測方法與單變量 ARIMA 的預測方法同，單變量 ARIMA 的預測參閱馬信行 (民76, 頁122-123)。

## 五、模式適合性檢定

一般轉換函數是否適合，判斷標準有二：a) 殘差的自我相關函數不達顯著，b) 白化

後的自變數列與殘差之交叉相關函數不達顯著，檢定交叉相關係數的標準誤是與檢定偏自我相關係數 (partial autocorrelation, 簡稱 *pacf*) 或逆自我相關 (inverse autocorrelation, 簡寫成 *acf*<sup>-1</sup>) 的標準誤一樣的，是等於  $1/\sqrt{n}$ ， $n$  為樣本數。檢定殘差的自我相關的標準誤是與檢定單變量 *ARIMA* 的自我相關係數的標準誤一樣，

$$SE_{r_{aa}(k)} \approx \{1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)\}^{1/2} \quad q = k - 1$$

其中  $n$  為殘差數， $SE_{r_{aa}(k)}$  是第  $k$  期殘差自我相關的標準誤。另外，較粗糙的檢定殘差的 *acf* 是否達顯著的方法是檢定第一期到第  $k$  期的殘差 *acf* 是否達顯著。

$$Q = m \sum_{k=1}^K r_{aa}^2(k)$$

其中  $m = n - \max(r, s + b) - p$ ， $K$  為期數， $\max(r, s + b)$  為取  $r$  與  $s + b$  兩者中最大的一項， $p$  為 *AR*( $p$ ) 的秩， $n = N - d$ ， $N$  為觀察值， $d$  為差分的秩。檢定  $Q$  值是否達顯著，要以  $\chi^2_{(k-p-q)}$  的臨界值來比較，其中  $k$  是期數， $p$  與  $q$  是殘差的 *ARIMA*( $p, q$ ) 之秩，這可看出  $\chi^2$  的自由度與  $w_s$  及  $\delta_r$  之秩無關。

檢定自變數列與殘差交叉相關是否達顯著，所用標準誤是  $SE_{r_{aa}(k)} = 1/\sqrt{n}$ ，其中  $n$  是殘差數。此外，也可用粗糙的方法檢定第零期到第  $k$  期的所有交叉相關係數是否達顯著，

$$S = m \sum_{k=0}^K r_{aa}(k)$$

其中  $m = n - \max(r, s + b) - p$ ，此  $m$  的意義與計算  $Q$  值時之  $m$  一樣， $r_{aa}(k)$  為  $a_{t-k}$  與  $a_t$  在第  $k$  期的交叉相關係數，檢定  $S$  值是否達顯著，要用  $\chi^2_{(k+1-(r+s+1))}$  的臨界值，其中  $r$  與  $s$  分別為  $\delta_r$  與  $w_s$  之秩，這可看出  $\chi^2$  之自由度與殘差 *ARIMA*( $p, q$ ) 之秩無關。

事實上，模式適合並不能說預測一定準確。在實用上，最佳的模式是預測值與實際值之差為最小。但這要等未來實際值出現後才能知道。本研究同時用幾種指標來考慮：a) 殘差數列的標準誤 (standard error estimate), AIC (Akaike's Information Criterion)，及 SBC (Schwartz's Bayesian Criterion)。這三個值都是愈小愈好，b) 殘差的 *acf*, *acf*<sup>-1</sup>，及 *pacf*。這三個指標如皆未達顯著，則便可接受，c) 白化後的自變數列與殘差的交叉相關函數皆不達顯著，則可接受，d) 從最後一年開始預測，產生最後一期的預測值，與最

後一年的實際值比較，殘差愈小者愈佳。本研究使用教育部電算中心在政大電算中心之工作站之終端機，所用程式為 SAS 裏的 *ETS*。因為  $\delta_r$  與  $w_s$  之秩難以精確判斷，故  $r$  與  $s$  皆以  $0 \sim 3$ ，一一嘗試，以找出最佳模式。

## 結 果

### 一、國民小學學生數的預測

我國自民國39年到77年的出生數趨勢如圖1。我國國小學生數之趨勢如圖2。

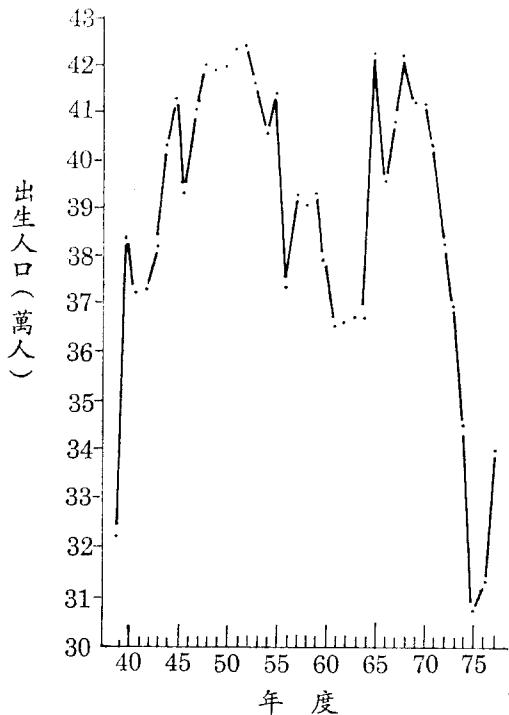


圖 1 我國出生人口趨勢 [資料來源：行政院衛生署等，民78，頁54]

國小學生數受出生數的影響是顯然的，因為我國國小就學率已達漸近線 100 %，所以出生後的活存率大約等於在學率。本研究以出生數為自變數列，以國小學生數為依變數列來預測往後十二年的國小學生數。

時間數列分析之轉換模式在學生數預測上之應用

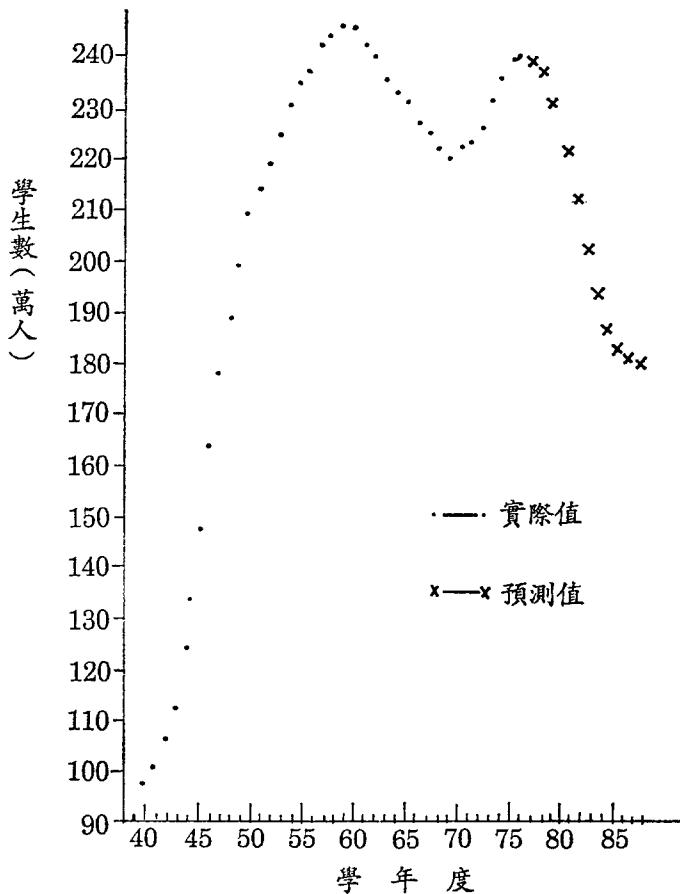


圖 2 我國國民小學學生數趨勢〔實際值資料

來源：取自教育部，民78，頁18-20〕

首先看兩個數列的單變量數列的 *ARIMA* 過程。國小學生數從圖 2 可看出並不平穩，故須差分。差分之後，自我相關函數成指數下降 (exponential decay)，而偏自我相關函數在第一期達顯著，故國小學生數的單變量 *ARIMA* 模式是  $p=1, d=1, q=0$ 。出生數從圖 1 很難看出有逐漸上升或下降的趨勢。其自我相關函數顯出不須差分。因為自我相關函數成指數下降，偏自我相關函數則在第一期達顯著，故判斷 *ARIMA* 的模式  $p=1, d=0, q=0$ 。因為此模式所產生的殘差之自我相關，逆自我相關 (inverse autocorrelation)，及偏自我相關 (partial autocorrelation) 皆不達顯著，表示模式適合。電腦算出  $\phi_1 = 0.9545748$ 。整個模式可寫成  $(1 - 0.9545748B)X_t + a_t$ 。我們要用這模式來白化自變數列、依變數列、與干擾

數列。白化後自變數列與依變數列的交叉相關 (cross correlation function, 簡稱  $ccf$ ) 在第 3, 第 6, 及第 12 期達顯著。從實際情形來看，第 3 期與第 12 期達顯著應該是資料的偶然性。因為出生數對國小學生數的影響不可能發生在出生後三年及十二年。十二年已是國小畢業。追查其原因是民國 77 年的國小學生數停滯不前，如將民 77 年的資料不計，則  $ccf$  只有第 6 期達顯著，較合乎常理。這也許是由於觀察值尚不足 50 個，模式有點不穩定之故。所以本研究檢定模式仍以  $b=6$  為主。 $s$  與  $r$  皆從 0 到 3 一一嘗試。本來是用 0 到 2 即可，但為慎重起見，以 overfitting (過度適合) 的方式來尋找。基於實際情形的考慮，出生數在 6 年後會影響國小學生數 ( $b=6$ )，但影響期可能有 6 年，因為要讀 6 年才畢業，所以也考慮  $w(B) = w_0 - w_1 - w_2 - w_3 - w_4 - w_5 - w_6$  來嘗試，看看是否模式更適合。從這些模式選出最佳模式後再看看是否需要截距 (intercept，也就是常數項)。故把常數項取掉，看模式是否更適合。另外建立單變量 ARIMA 的模式，來與轉換模式比較。模式適合性的比較，主要用一些指標：a) 殘差標準誤的估計值 (standard error estimate)，標準誤愈小愈佳，b) AIC (Akaike's information criterion)，其值愈小愈佳，c) SBC (Schwartz's Bayesian criterion)，其值也是愈小愈佳，d) 殘差的  $acf$  不顯著，殘差的  $acf$  達顯著的話，表示干擾數列模式不適合，表 3 第 5 列裏如果是 1，則表示第 1 期達顯著，餘類推；e) 殘差的  $acf^{-1}$  不達顯著最好，不過如果顯著，其嚴重性不如  $acf$  達顯著之不能忍受， $pacf$  與  $acf^{-1}$  的意義相近似 (SAS Institute Inc., 1984, 頁 139)；f) 殘差的  $pacf$  不達顯著最好，但其嚴重性亦不如殘差的  $acf$  達顯著者；g) 殘差數列與自變數列的  $ccf$  不達顯著，如達顯著，則表示自變數列模式不適合，表 3 第 8 列中，如果是 6，則表示  $ccf$  第 6 期達顯著，如註明 1—5 則表示第 1 到第 5 期以粗糙的檢定法， $Q$  值的  $X^2$  達顯著，餘類推；h) 預測值與實際值的差愈小愈好，本研究僅提供民國 77 年的實際值以與預測值比較，一般來講，在預測初期誤差較小的話，以後幾期的誤差也可能會較小，Hopwoop (1980) 的研究顯示轉換模式如在前三期的預測比單變量模式為準確的話，則其餘到第十期的預測也會較準確。根據以上指標，分別檢定各種模式，結果發現模式 13： $\left[\frac{1, 2, 3}{1}\right]$  最好，因為標準誤第二低，而 AIC 與 SBC 最低，殘差的  $acf$ ,  $acf^{-1}$ ,  $pacf$  皆不達顯著。模式 9： $\left[\frac{1, 2}{1}\right]$  雖然誤差最低，

時間數列分析之轉換模式在學生數預測上之應用

表3 以轉換函數模式預測國小學生數，各種模式適合性之比較

比較〔出生數之 ARIMA,  $p=1, d=0, q=0$ ; 國小學生數之 ARIMA,  $p=1, d=1, q=0; b=6$ 〕

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$	$\textcircled{1} \left[ \begin{array}{c} 0 \\ -1 \end{array} \right]$	$\textcircled{2} \left[ \begin{array}{c} 0 \\ 1, 2 \end{array} \right]$	$\textcircled{3} \left[ \begin{array}{c} 0 \\ 1, 2, 3 \end{array} \right]$	$\textcircled{4} \left[ \begin{array}{c} -1 \\ 0 \end{array} \right]$	$\textcircled{5} \left[ \begin{array}{c} -1 \\ -1 \end{array} \right]$	$\textcircled{6} \left[ \begin{array}{c} 1 \\ -1, 2 \end{array} \right]$	$\textcircled{7} \left[ \begin{array}{c} 1 \\ 1, 2, 3 \end{array} \right]$	
殘差的標準誤	17861	16200	15795	16278	17914	17003	16456	
AIC	721	693	670	715	722	697	673	
SBC	726	700	678	721	729	705	683	
達顯著的期	殘差的acf	2.	2.	ns.	ns.	ns.	1.	
	殘差的 $acf^{-1}$	2.3.5.	2.	2.	2.	2.3.5.	1.	
	殘差的pacf	ns.	2.	ns.	ns.	ns.	4.	
殘差與自變數列之ccf達顯著的期	6.	2.6.	6.	6.	6.	6.	1-5 6.	
77年實際值	2407166							
預測值	77年	2430875	2431898	2431242	2431372	2431869	2431677	2431251
	78年	2445407	2449839	2449856	2447003	2449321	2449021	2450169
	79年	2447781	2459908	2462320	2451082	2456669	2460334	2461592
	80年	2430972	2457372	2464640	2437640	2448284	2457954	2462343
	81年	2385235	2432035	2449952	2396951	2415358	2433533	2442527
	82年	2335150	2411667	2440894	2352849	2380339	2414615	2426018
	83年	2284411	2399613	2438762	2310056	2346734	2397782	2418566
	84年	2233957	2387244	2440077	2267827	2313570	2379154	2412138
	85年	2184013	2372606	2439050	2226138	2280887	2362315	2402493
	86年	2134615	2358260	2435862	2184963	2248658	2345071	2392375
	87年	2085756	2345073	2432632	2144278	2216863	2328510	2383616
	88年	2037415	2332286	2430766	2104063	2185483	2312116	2375821

表3 [續1]

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$		$\textcircled{8} \left[ \begin{smallmatrix} 1, 2 \\ 0 \end{smallmatrix} \right]$	$\textcircled{9} \left[ \begin{smallmatrix} 1, 2 \\ 1 \end{smallmatrix} \right]$	$\textcircled{10} \left[ \begin{smallmatrix} 1, 2 \\ 1, 2 \end{smallmatrix} \right]$	$\textcircled{11} \left[ \begin{smallmatrix} 1, 2 \\ 1, 2, 3 \end{smallmatrix} \right]$	$\textcircled{12} \left[ \begin{smallmatrix} 1, 2, 3 \\ 0 \end{smallmatrix} \right]$	$\textcircled{13} \left[ \begin{smallmatrix} 1, 2, 3 \\ 1 \end{smallmatrix} \right]$	$\textcircled{14} \left[ \begin{smallmatrix} 1, 2, 3 \\ 1, 2 \end{smallmatrix} \right]$
殘差的標準誤	16009	12399	14682	15236	15002	13790	15271	
AIC	692	677	688	669	667	663	669	
SBC	699	686	698	680	675	672	681	
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.	ns.	ns.
	殘差的 $acf^{-1}$	ns.	2.5.	ns.	2.	2.	ns.	ns.
	殘差的pacf	ns.	9.	ns.	ns.	ns.	ns.	ns.
殘差與自變數列之ccf達顯著的期	6.	1-5 1-11	1-5 1-11	1-5 6.	6.	6.	6.	
77年實際值	2407166							
預測值	77年	2432998	2419199	2430261	2428966	2428901	2426174	2429232
	78年	2452606	2413630	2445242	2441330	2445620	2435602	2446205
	79年	2466385	2394572	2452742	2444732	2454953	2432311	2455357
	80年	2468813	2362334	2449807	2433608	2452563	2412601	2452856
	81年	2451573	2312669	2429856	2402654	2432889	2371593	2434737
	82年	2441059	2280112	2419823	2377910	2421141	2332129	2421753
	83年	2435988	2272775	2423214	2362442	2419142	2301972	2416038
	84年	2430649	2285708	2435049	2357766	2426031	2286205	2420001
	85年	2425594	2316014	2454552	2359442	2432112	2280077	2428090
	86年	2420809	2361188	2480222	2365382	2438367	2280982	2439636
	87年	2416284	2419053	2511027	2372523	2444787	2287014	2453243
	88年	2412005	2487740	2546054	2380113	2451363	2296778	2468294

時間數列分析之轉換模式在學生數預測上之應用

表 3 [續 2]

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$		$\textcircled{15} \left[ \begin{array}{l} 1, 2, 3 \\ 1, 2, 3 \end{array} \right]$	$\textcircled{16} \left[ \begin{array}{l} 1, 2, 3 \\ 1 \end{array} \right]$ 無截距	$\textcircled{17} 3 \left[ \begin{array}{l} 1, 3 \\ 1 \end{array} \right]$	$\textcircled{18} \left[ \begin{array}{l} 1, 2, 3, 4, 5, 6 \\ 1 \end{array} \right]$	$\textcircled{19} \left[ \begin{array}{l} 1, 2, 3, 4, 5, 6 \\ 1 \end{array} \right]$ 無截距	$\textcircled{20}$ 單變量， $pdq = (1, 0, 0)$ 無截距
殘差的標準誤	15956	13444	15352	7160	6967	22492	
AIC	673	660	735	563	561	870	
SBC	685	669	744	576	573	872	
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.	ns.
	殘差的 $acf^{-1}$	2.	ns.	ns.	ns.	ns.	2.
	殘差的pacf	ns.	ns.	ns.	ns.	ns.	ns.
殘差與自變數列之ccf達顯著的期	6.	6.	ns.	ns.	ns.		
77年的實際值	2407166			2407166	2407166		
預測值	77年	2430193	2422051	2425356	2408619	2408763	2434665
	78年	2449471	2420018	2428336	2403390	2403644	2466715
	79年	2462535	2397018	2427984	2377851	2377903	2496883
	80年	2465422	2345760	2414935	2319772	2319119	2525278
	81年	2454342	2258460	2381995	2228862	2226831	2252005
	82年	2449569	2165064	2346750	2130323	2126410	2577162
	83年	2451501	2076834	2311419	2033922	2027915	2600841
	84年	2461540	2005454	2276503	1952463	1944264	2623129
	85年	2473982	1946555	2242100	1885045	1874374	2644107
	86年	2488480	1897941	2208216	1839344	1825906	2663854
	87年	2503934	1857795	2174833	1824514	1808168	2682440
	88年	2520018	1824619	2141930	1822798	1803326	2699934

但殘差的  $acf^{-1}$  與  $pacf$  有達顯著者，且殘差數列與自變數列的交叉相關函數 1—5，1—11 達顯著，顯示模式不適合。但從模式 13 與 9 可看出  $\delta(B)$  皆是  $1 - \delta_1 B$ 。模式 13 的殘差數列與自變數列之  $ccf$  粗糙檢定結果，有第 6 期達顯著，顯出可能  $b=6$  並不適合。將模式 13 的常數項取掉顯出標準誤， $AIC$  及  $SBC$  都有些微下降，可能常數項是不需要的。為了殘差與自變數列第 6 期  $ccf$  達顯著之故，再試將  $b$  定為 3。模式 17： $3\{ \frac{1}{1}, \frac{3}{1} \}$  表示  $b=3$ ，模式是  $\lfloor \frac{w_0 - w_1 B - w_3 B^3}{1 - \delta_1 B} \rfloor$  因為  $w_2$  與  $w_3$  有高相關，故去掉  $w_2$ ，結果發現模式反而更糟。因此再依實際考慮  $b=6$ ， $w_s = w_0 - w_1 B - w_2 B^2 - w_3 B^3 - w_4 B^4 - w_5 B^5 - w_6 B^6$ ，因為出生數從第 6 年開始對國小學生數影響（滿 6 歲入學），且連續影響 6 年（讀 6 年），結果發現模式 18 的標準誤， $AIC$  及  $SBC$  皆下降很多，殘差數列與自變數列的  $ccf$  皆不再顯著。把模式 18 的常數項取掉，得模式 19，顯出模式更佳。所以最後的模式是模式 19。（粗框的那欄）。以之與單變量的國小學生數列預測相比，轉換模式的標準誤， $AIC$  及  $SBC$  皆較單變量為佳。民國 77 年的預測值也比單變量的更接近實際值。未來的預測值是否比單變量的預測更準確，要待時間來考驗。初步的估算，民 72 年到民 77 年這六年的出生數分別為 382153, 369725, 345043, 308187, 313063, 341054 總共為 2059234。這筆數目便約為六年後國小一年級到六年級的學生數，即約民 83 年的學生數，從這點看，模式 19 比單變量更為接近。電腦算出了模式 19 的參數值。

$$\begin{aligned} \text{整個模式是：} (1-B)Y_t = & [(0.723418 - 1.32418B + 0.005142B^2 - 0.055329B^3 \\ & + 0.0264798B^4 - 0.022563B^5 - 0.539537B^6)/(1 - 0.49521B)]X_{t-6} \\ & + [1/(1 - 0.47026B)]a_t。 \end{aligned}$$

有些  $w(B)$  的參數雖然很小， $t$  值也小於 1，尤其是  $w_2$  到  $w_5$ 。但由於參數間的相關都在 .5 以下，並不大，故皆保留各參數。

另外我國的出生數近十二年變動很大，首先是民國六十五年的龍年，遽然上升，造成教育計劃上的困擾，之後，民 72 年才慢慢回降，民 74 年元月一日通過優生保健法，使墮胎合法化，更使得出生數下降到一年 30 萬。之後家庭計畫單位因怕人口斷層，加上人口老化，使得每人所負擔的老人人口比例加大，因此提出「三個孩子不算多」的口號，加上民 77 年是龍年，使

## 時間數列分析之轉換模式在學生數預測上之應用

出生數在民77年又回升到一年34萬。以後我國出生數如何，實難以預料。因此在此以出生數列作單變量預測。由於不經差分的數列  $acf$  成指數下降， $acf^{-1}$  及  $pacf$  都是第一期達顯著，故模式判斷為  $p=1, d=0, q=0$ 。在此模式下，殘差的  $acf$ ,  $acf^{-1}$ ，及  $pacf$  皆不達顯著。自民77年到民88年的預測值分別為：77(313796), 78(314497), 79(315167), 80(315805), 81(316415), 82(316997), 83(317553), 84(315805), 85(318589), 86(319073), 87(319534), 88(319974)。不過筆者對此出生數的預測抱極度懷疑的態度，因為實際的出生數從沒這麼平穩過，似乎預測值不能反映實際情形。筆者認為出生數的預測需待觀察值超過50個以後才能找到適當的模式，也許會有季節性，例如十二年一週期等。筆者在此尚不敢揣測。

## 二、國中學生數之預測

在實際推理上國中學生數受國小畢業生數的影響，圖3與圖4有相似的趨勢，故本研究嘗試以國小畢業生數為自變數列，以國中學生數為依變數列，作轉換函數模式分析。國小畢業生數列差分後的  $acf$  只有第一期達顯著，而  $acf^{-1}$  第7期， $pacf$  第1及第7期達顯著，故模式判斷為： $p=1,7, d=1, q=1$ 。國中學生數差分後， $acf$  成指數下降，且第1, 2期達顯著，逆自我相關函數第1, 8期達顯著，偏自我相關函數也是第1, 8期達顯著，故模式判斷為： $p=1,8, d=1, q=1,2$ 。（因為這模式的殘差之自我相關函數沒達顯著，逆自我相關函數及偏自我相關函數都在第7期達顯著。如用  $p=1,8, d=1, q=0$ ，則殘差的自我相關，逆自我相關，及偏自我相關函數皆在第7期達顯著。用  $p=1,8, d=1, q=2$ ，則殘差的自我相關及逆自我相關函數在第7期達顯著，偏自我相關函數在第7, 17期達顯著。而且模式  $p=1,8, d=1, q=1,2$  的殘差標準誤亦較其他兩個模式為少。參數的  $t$  值也都在1.5以上。常數項的  $t$  值達4.31，故不捨棄）。國小畢業生數的單變量模式電腦計算的結果是：

$$(1 - 0.519284B - 0.317083B^7)X_t = 10822.8 + (1 - 0.129584B)a_t$$

以此公式來白化自變數列，依變數列，及干擾數列。白化以後的自變數列與依變數列之交叉相關函數在第-8, 3, 6, 10期達顯著。第負8期達顯著表示，依變數列影響自變數列，這顯然與實際不符。國小畢業生在  $t-3$ 期發生變化後才導致國生數發生變化，這也與實際情形不符。依實際推理，國小畢業生在  $t-1$ 期發生變化，就導致國中生第  $t$  期發生變化。故  $b=1$  才

合實際情形。交叉相關之所以發生這種情形，可能是我國國小畢業生升學率在民75年以後才達99%以上，而在民39年只有32%，民50年有54%，民60年有81%（教育部，民78，頁29）。當升學率沒達100%時，國中學生數受國小畢業生數的影響就不完全。筆者仍以實際考慮，令 $b=1$ 來分析，將 $w_s$ 中的 $s$ 以0到3， $\delta_r$ 中的 $r$ 以0到3分別組合來嘗試。結果如表4。表4顯出，殘差標準誤最小的模式是模式4，AIC及SBS也是最小，殘差的自我相關，偏自我相關，及逆自我相關函數，殘差數列與白化後的自變數列之交叉相關函數都不達顯著。

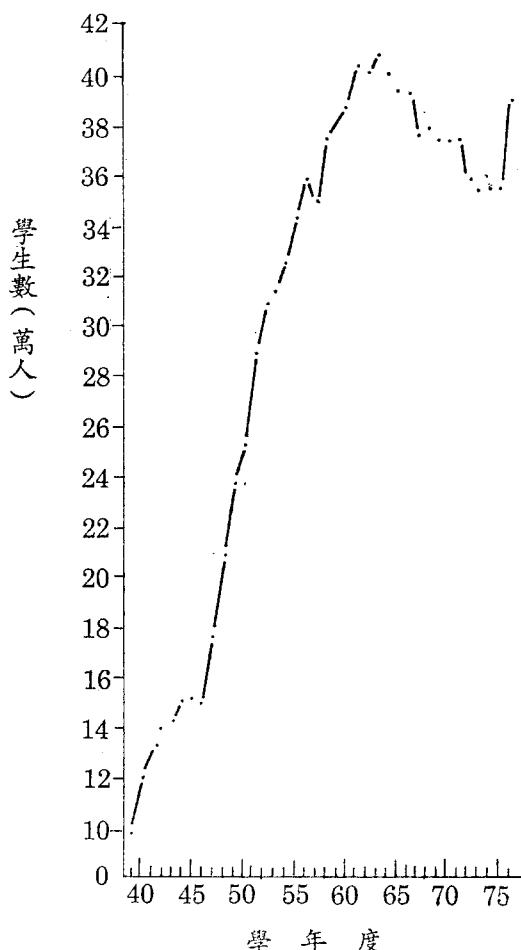


圖 3 我國國民小學畢業學生數趨勢

[資料來源：教育部，民78，頁22-24]

### 時間數列分析之轉換模式在學生數預測上之應用

其次適合的模式是模式 1。其殘差的標準誤， $AIC$  及  $SBC$  比模式 4 略高，殘差的自我相關，逆自我相關，偏自我相關，殘差與白化後的自變數列之交叉相關函數也都不達顯著。但模式 4 在民76年的預測值比實際值大很多，而模式 1 的預測值較為接近，故考慮採用模式 1 為最佳模式。把模式 1 的常數項取掉，其殘差的標準誤卻大增，且殘差的自我相關，逆自我相關，及偏自我相關皆達顯著，故不用。與單變量比較，則發現模式 1 轉換函數模式的殘差標準誤， $AIC$ ,  $SBC$  都比單變量為小，且單變量的殘差之逆自我相關及偏自我相關函數有達顯著。雖然民76年的預測值以單變量較為接近，由於殘差標準誤， $AIC$ ,  $SBC$  較大，且逆自我相關及偏自我相關達顯著，往後的預測可能會比轉換函數模式不準。由於國小畢業生數的

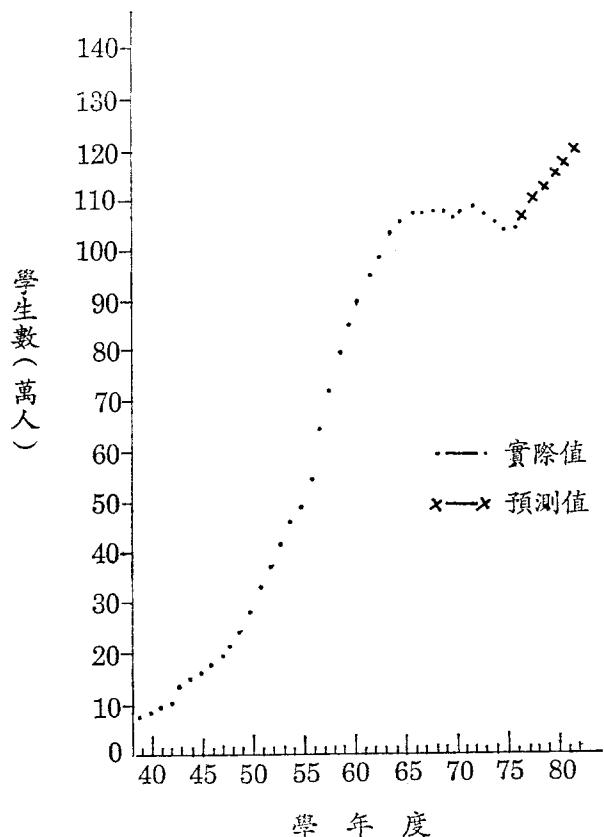


圖 4 我國國民中學（包含以前的初中、職）學生數趨勢  
〔實際值資料來源：教育部，民78，頁18-22〕

表 4 以轉換函數模式預測國中學生數，各種模式適合性之比較

〔國小畢業生數之 ARIMA,  $p=1, 7, d=1, q=1$ ; 國中學生數  $p=1, 8, d=1, q=1, 2; b=1$ 〕

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$		① $\left[ \frac{1, 2, 3}{0} \right]$	② $\left[ \frac{1, 2, 3}{1} \right]$	③ $\left[ \frac{1, 2, 3}{1, 2} \right]$	④ $\left[ \frac{1, 2, 3}{1, 2, 3} \right]$	⑤ $\left[ \frac{1, 2}{0} \right]$	⑥ $\left[ \frac{1, 2}{1} \right]$	⑦ $\left[ \frac{1, 2}{1, 2} \right]$
殘差的標準誤	10932	11069	9748	9577	12898	12283	10546	
AIC	714	716	708	707	746	744	734	
SBC	728	731	724	725	759	758	749	
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.	ns.	ns.
	殘差的 $acf^{-1}$	ns.	ns.	6.	ns.	ns.	ns.	6.
	殘差的pacf	ns.	ns.	12.	ns.	ns.	ns.	ns.
殘差與自變數列之ccf達顯著的期	ns.	ns.	ns.	ns.	10.	5.	ns.	
76年的實際值	1053923			1053923				
預測值	76年	1065089	1065970	1085108	1081924	1055830	1059485	1091859
	77年	1084704	1087927	1129775	1107854	1065892	1071428	1155947
	78年	1105494	1111315	1166700	1128508	1078382	1086184	1203876
	79年	1129348	1138064	1204550	1151403	1092384	1103508	1244048
	80年	1152590	1164571	1243040	1176294	1106089	1120174	1279925
	81年	1175487	1190953	1277957	1197587	1119941	1137387	1309502
	82年	1199157	1218372	1313494	1218858	1134916	1156213	1335806

時間數列分析之轉換模式在學生數預測上之應用

表 4 [續 1]

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$		$\left[ \frac{1, 2}{1, 2, 3} \right]$	$\left[ \frac{-1}{0} \right]$	$\left[ \frac{-1}{1} \right]$	$\left[ \frac{1}{1, 2} \right]$	$\left[ \frac{1}{1, 2, 3} \right]$	$\left[ \frac{0}{0} \right]$	$\left[ \frac{0}{1} \right]$
殘差的標準誤		9274	13668	13892	14132	16630	13698	13597
AIC		705	772	773	753	743	793	771
SBC		721	783	786	767	758	802	782
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.	ns.	ns.
	殘差的 $acf^{-1}$	6.	ns.	ns.	ns.	ns.	ns.	ns.
	殘差的pacf	ns.	ns.	ns.	ns.	ns.	16.	ns.
殘差與自變數列之ccf達顯著的期		ns.	3.9.	3.9.	1-5 1-11	1-5 1-11	3.9.	1-5 1-11
民76年的實際值		1053923						
預測值	76年	1084584	1056534	1056815	1054494	1086729	1053347	1056590
	77年	1129597	1071710	1072105	1066369	1093988	1062629	1701766
	78年	1163115	1095047	1095523	1085754	1151197	1078818	1094975
	79年	1194480	1122558	1122996	1108818	1180377	1098444	1122226
	80年	1225508	1153654	1153843	1134439	1266707	1121956	1152832
	81年	1251786	1191936	1191242	1166318	1330320	1152903	1190235
	82年	1277454	1239979	1237826	1206770	1461680	1240986	1237009

表 4 [續 2]

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$		$\textcircled{15} \left[ \frac{0}{1, 2} \right]$	$\textcircled{16} \left[ \frac{0}{1, 2, 3} \right]$	$\textcircled{17} \left[ \begin{matrix} 1, 2, 3 \\ 0 \end{matrix} \right]$ 無截距	單變量 $\textcircled{18} p=1, 8,$ $q=1, 2,$ $d=1$	單變量 $\textcircled{19} p=1, 7,$ $q=1, 2,$ $d=1$	單變量 $\textcircled{20} \left[ \begin{matrix} 1, 2, 3 \\ 0 \end{matrix} \right]$ $p=1, 7, q=1, 2,$ $d=1$
殘差的標準誤		13659	13768	24084	13722	14743	11091
<i>AIC</i>		750	730	766	814	819	715
<i>SBC</i>		763	743	778	822	827	729
達顯著的期	殘差的 $acf$	ns.	ns.	1, 2, 3	ns.	ns.	ns.
	殘差的 $acf^{-1}$	ns.	5.	2.	7.	4.	ns.
	殘差的 $pacf$	ns.	ns.	1, 2, 8	7.	ns.	ns.
殘差與自變數列之 $ccf$ 達顯著的期		1-5 1-11	1-5 1-11	ns.			ns.
民76年的實際值		1053923					
預測值	76年	1059413	1056388	1057848	1051253	1050292	1067145
	77年	1079050	1071853	1060980	1056457	1051535	1086150
	78年	1107512	1095088	1063079	1067612	1059636	1105205
	79年	1140577	1122415	1065667	1081610	1067460	1147834
	80年	1177362	1151846	1067484	1099737	1068995	1166965
	81年	1220451	1187200	1065523	1125089	1066149	1184728
	82年	1272637	1230551	1064312	1158916	1064436	1206541

## 時間數列分析之轉換模式在學生數預測上之應用

升學率以往並未達 100%，故無法很有信心的將之當作自變數列。筆者嘗試以出生數為自變數列，但亦未見改善。出生數列與國中學生數列之交叉相關函數皆沒達顯著。然因最近幾年國中學生數已幾達全民就學，故受出生數的影響很大。在民國73年以後我國出生人數有銳減的現象，這將導致民85年以後國中學生數的下降，故表 4 的預測值只預測 6 期，到民82年為止。以後要靠逐年更新資料才能測出較精確的值。模式 1 的各個參數電腦算出為：

$$(1-B)Y_t = 168649 + [(0.66482 + 0.743153B + 0.674114B^2 - 0.290887B^3)$$

$$(1-B)]X_{t-1} + [(1 + 0.27144B + 0.312796B^2)/(1 - 0.70638B + 0.21439B^8)]a_t。$$

因為國中學生數的 *ARIMA* 模式以  $p=1, 8, d=1, q=1, 2$  時，殘差的逆自我相關及偏自我相關皆在第 7 期達顯著，故嘗試以  $p=1, 7, d=1, q=1, 2$  進行單變量的檢定與預測，結果得模式 19。殘差的偏自我相關不再達顯著，但殘差的標準誤卻有增加。再試修改模式 1。將  $p=1, 8$  改為  $p=1, 7$ ，其餘不變，結果得模式 20，發現並沒多大改善，且殘差的標準誤有點增加，故仍以模式 (1) 為最佳模式。

### 三、我國年中人口數的預測

人口的預測是各項經建預測之母，所以我國經建計畫一定要先列出人口成長之預測。在此擬以出生數為自變數列，以年中人口為依變數列，嘗試各種轉換函數模式，看是否會比單變量更為準確。出生數列的 *ARIMA* 模式仍然是  $p=1, d=0, q=0$ 。而人口數列經一次差分後，自我相關函數的第一期達顯著，逆自我相關及偏自我相關的第一期也達顯著，故模式判斷為  $p=1, d=1, q=1$ 。以出生人口的單變量模式  $1 - 0.954574 B$  來白化自變數列與依變數列。白化後的自變數列與依變數列之交叉相關皆未顯著。照理每年的出生數對人口的影響可有七八十年，但  $w_s$  之  $s$  不可能到 70，故從 0 到 3， $s_r$  之  $r$  從 0 到 3 分別嘗試不同組合模式，結果得表 5。

表 5 發現大多數的模式都適合（即殘差的自我相關，逆自我相關，偏自我相關，殘差與自變數列的交叉相關皆不達顯著）。但比較各適合模式中，以模式 9 之殘差標準誤為最小，但其 *AIC* 與 *SBC* 却比模式 15 的更大。比較民77年的殘差（實際值與預測值之差）則模式 15 較接近。故以模式 15 為甲案，模式 9 為乙案，所作預測標於圖 5。

表5 以轉換函數模式預測我國年中人口數，各種模式適合性之比較  
性之比較〔出生數之 ARIMA;  $p=1, d=0, q=0$ ; 人口數之 ARIMA;  $p=1, d=1, q=1; b=1$ 〕

模式 $\left(\frac{w(B)}{\delta(B)}\right)$	$\textcircled{1}\left(\begin{array}{c} 0 \\ 0 \end{array}\right)$	$\textcircled{2}\left(\begin{array}{c} 1 \\ 0 \end{array}\right)$	$\textcircled{3}\left(\begin{array}{c} 1, 2 \\ 0 \end{array}\right)$	$\textcircled{4}\left(\begin{array}{c} 1, 2, 3 \\ 0 \end{array}\right)$	$\textcircled{5}\left(\begin{array}{c} 0 \\ -1 \end{array}\right)$	$\textcircled{6}\left(\begin{array}{c} -1 \\ 1 \end{array}\right)$	$\textcircled{7}\left(\begin{array}{c} 1, 2 \\ -1 \end{array}\right)$	
殘差的標準誤	29807	29540	28086	28694	29200	29618	27953	
AIC	894	871	845	824	870	872	845	
SBC	901	879	854	834	878	882	856	
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.	ns.	
	殘差的 $acf^{-1}$	ns.	ns.	ns.	ns.	ns.	3.	
	殘差的pacf	ns.	ns.	ns.	ns.	ns.	ns.	
殘差與自變數列之ccf達顯著的期	ns.	ns.	ns.	ns.	ns.	ns.	1-5	
77年的實際值	19788212							
預測值	77年	19801966	19802348	19767408	19787173	19797689	19798466	19778019
	78年	20050101	20059781	19971660	20030411	20044958	20046016	19986604
	79年	20301722	20323011	20183552	20283444	20296233	20296979	20196276
	80年	20554841	20588359	20397461	20538265	20549007	20548938	20407084
	81年	20808816	20854697	20612556	20794080	20802611	20801377	20618944
	82年	21063419	21121668	20828609	21050605	21056862	21054248	20831779
	83年	21318595	21389150	21045532	21307749	21311700	21303786	21045533
	84年	21574271	21657092	21263276	21565469	21567097	21561419	21260157
	85年	21830430	21925463	21481802	21823735	21823027	21815758	21475610
	86年	22087046	22194242	21701073	22082523	22079467	22070598	21691853
	87年	22344099	22463410	21921056	22341807	22336396	22325927	21908851
	88年	22601569	22732948	22141717	22601506	22593790	22581728	22126569

## 時間數列分析之轉換模式在學生數預測上之應用

表 5 [續 1]

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$	$\textcircled{8} \left[ \frac{1, 2, 3}{1} \right]$	$\textcircled{9} \left[ \begin{matrix} 0 \\ 1, 2 \end{matrix} \right]$ 乙案	$\textcircled{10} \left[ \begin{matrix} 1 \\ 1, 2 \end{matrix} \right]$	$\textcircled{11} \left[ \begin{matrix} 1, 2 \\ 1, 2 \end{matrix} \right]$	$\textcircled{12} \left[ \begin{matrix} 1, 2, 3 \\ 1, 2 \end{matrix} \right]$	$\textcircled{13} \left[ \begin{matrix} 0 \\ 1, 2, 3 \end{matrix} \right]$	$\textcircled{14} \left[ \begin{matrix} 1 \\ 1, 2, 3 \end{matrix} \right]$	
殘差的標準誤	29099	27937	28310	27674	28662	28961	29349	
AIC	825	844	846	845	825	824	826	
SBC	838	854	857	858	839	835	838	
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.	ns.	
	殘差的 $acf^{-1}$	ns.	ns.	ns.	ns.	ns.	ns.	
	殘差的pacf	ns.	ns.	ns.	11.	ns.	ns.	
殘差與自變數列之cof達顯著的期	1-5	ns.	ns.	ns.	ns.	ns.	1-5	
77年的實際值	19788212							
預測值	77年	19790749	19771544	19771107	19769031	19789995	19791252	19802959
	78年	20039423	19964474	19965632	19969209	20044034	20038020	20072252
	79年	20299226	20144563	20151147	20181428	20306611	20302589	20354924
	80年	20561791	20319282	20334661	20411891	20572470	20585428	20643840
	81年	20825514	20495335	20520942	20649778	20840475	20882395	20935179
	82年	21089944	20676962	20711954	20884333	21110300	21185386	21227543
	83年	21354942	20865891	20907673	21114399	21381809	21485808	21520265
	84年	21620451	21061925	21107092	21344406	21654911	21777873	21813366
	85年	21886440	21263715	21308981	21577254	21929527	22060290	22106763
	86年	22152884	21469466	21512317	21812311	22205585	22335917	22400538
	87年	22419762	21677451	21716334	22047781	22483014	22609866	22694612
	88年	22687054	21886318	21921000	22282963	22761749	22887085	22989014

表 5 [續 2]

模式 $\left[ \frac{w(B)}{\delta(B)} \right]$		$\textcircled{15} \left[ \begin{smallmatrix} 1, 2 \\ 1, 2, 3 \end{smallmatrix} \right]$ 甲案	$\textcircled{16} \left[ \begin{smallmatrix} 1, 2, 3 \\ 1, 2, 3 \end{smallmatrix} \right]$	$\textcircled{17} \left[ \begin{smallmatrix} 1, 2 \\ 0 \end{smallmatrix} \right]$ 無截距	$\textcircled{18} \left[ \begin{smallmatrix} 1, 2 \\ 1, 2, 3 \end{smallmatrix} \right]$ 無截距	單變量 $\textcircled{19} p=1, d=1, q=1$
殘差的標準誤	28284	29885	28320	27662	34376	
AIC	824	828	844	822	904	
SBC	838	844	852	834	909	
達顯著的期	殘差的acf	ns.	ns.	ns.	ns.	ns.
	殘差的 $acf^{-1}$	ns.	ns.	ns.	11.	ns.
	殘差的pacf	ns.	ns.	ns.	11.	ns.
殘差與自變數列之ccf達顯著的期		ns.	ns.	2.	ns.	
77年的實際值						
預測值	77年	19783411	19792579	19777059	19803651	19773588
	78年	20019455	20054841	20006874	20053431	19984218
	79年	20270343	20327110	20252658	20312103	20195444
	80年	20532739	20601427	20503964	20572080	20407215
	81年	20796073	20877269	20757687	20837644	20619484
	82年	21053661	21155066	21012645	21101792	20832208
	83年	21308042	21434402	21268378	21370246	21045349
	84年	21565142	21715025	21524698	21635920	21258870
	85年	21826694	21996975	21781522	21905733	21472738
	86年	22089607	22280224	22038806	22172087	21686923
	87年	22350792	22564677	22296520	22442890	21901399
	88年	22610501	22850274	22554642	22709731	22116140

### 時間數列分析之轉換模式在學生數預測上之應用

模式15與模式9的殘差，標準誤， $AIC$ ，及 $SBC$ 皆比單變量（模式19）為小，照理往後之預測應較為準確。另外把模式15的截距取消，發現殘差的逆自我相關與偏自我相關都在第11期達顯著。把模式9的截距取消，則殘差與自變數列之交叉相關第2期達顯著。故模式17與18仍然不足取。雖然其殘差標準誤有改善。模式15之參數經電腦計算結果是：

$$(1-B)Y_t = -42467 + [(0.321571 + 0.278412B + 0.307889B^2)/(1-0.725487B) \\ + 0.725266B^2 - 0.085192B^3]X_{t-1} + [(1-0.999998B)/(1-0.27282B)]a_t.$$

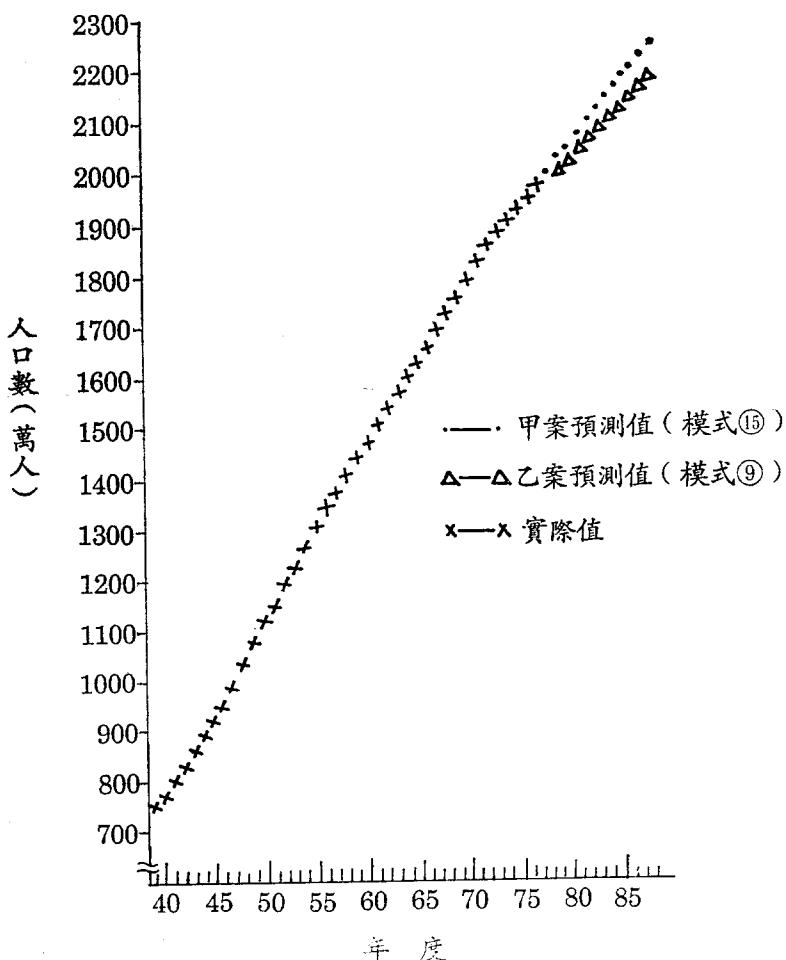


圖 5 我國年中人口數趨勢 [資料來源：行政院衛生署等，民78，頁54]

本文模式的公式是依一般的模式： $y_t = c + [w(B)/\delta(B)]x_{t-b} + [\theta(B)/\phi(B)]a_t$ 。而寫成的〔參閱 Edlung, 1984，頁298，公式(1)〕。其中 $y_t$ 與 $x_t$ 等於 $(1-B)Y_t$ 與 $(1-B)X_t$ 。

## 討 論

本研究嘗試用轉換函數模式來預測我國國小學生數，國中學生數，及年中人口數。在預測國小之學生數及年中人口數時皆以出生數為自變數列，在預測國中學生數時以國小畢業生數為自變數列。從白化過的自變數列與依變數列之交叉相關來看，這兩個自變數列並非好的自變數列，因為交叉相關函數與實際推論者不太吻合。原因可能是自變數列前半段的影響力與後半段的不同。前半段的影響力較小，因為就學率在前半段較低，另外，觀察值只有38期，不足所須的50期。這會造成模式不穩定。因此本研究的三種轉換函數模式預測皆以實際推論為判斷原則，而不依交叉相關函數。在衝擊反應函數中的 $w(B)$ ，則顯出所求得的模式皆能與實際推論判斷吻合。大凡 $w_s$ 中 $s$ 的秩是反映自變數列對依變數列發生作用的期數，例如國小畢業生對國中學生數發生作用的時間有三年（從國中一年級到國小三年級），故 $s$ 的秩為3。同樣的，出生數對國小學生數影響年數有6年，故 $s=6$ 。

本研究發現在方法論（methodology）方面有幾點值得提出討論的：

1. 檢定模式的指標有多種，但尚無一個唯一可靠的指標可供使用者遵循。一般來講殘差標準誤愈小， $AIC$ 與 $SBC$ 也愈小，則愈好，但常發現有例外，如表5之模式15與模式9就顯出這種不一致。最可靠的當然是預測值與實際值最接近者，但如模式不適合，則最後一期預測較準確，並不能保證以後也是較準確。本研究以殘差與白化後的自變數列之交叉相關函數為第一考慮，凡是交叉相關函數達顯著之模式皆被捨棄。其次是殘差的自我相關，逆自我相關，及偏自我相關函數不顯著。再次則比較殘差的標準誤， $AIC$ 及 $BCS$ 這三個值皆最小者，便可選為最佳模式。最後也考慮預測值與實際值的誤差。本文在預測時，在 $FORECAST$ 指令下用 $BACK=1$ 。以最後一年實際觀察值開始預測，故最後一年不但有觀察值，也有預測值，可資比較。

2. 參數估計值，原則上有高相關者最好省略其中一個，如果參數值的 $t$ 值太低者也可省略。

## 時間數列分析之轉換模式在學生數預測上之應用

但到底相關要多高才算高，參數值的  $t$  值要多低才算低，尚無定論。Heyse & Wei (1985) 的研究中把參數值沒比其標準誤大 1.5 倍者皆取消，但這也是依照經驗法則而已。筆者在嘗試錯誤過程中有時發現，把  $t$  值小於 1 的參數取消，卻導致模式更不適合。Montgomery & Weatherby (1980) 也發現有時  $\phi_1$  與  $\phi_2$  有高相關，但如去掉  $\phi_2$ ，卻導致殘差有達顯著者，因而兩個參數都得保留。

3. 模式適合性的判斷標準，有些研究者用  $R^2$  來當指標。例如 Bowning et al. (1983) 除了用殘差的標準誤外，也用  $R^2$  來判斷幾種模式的適合性，但 SAS 並沒提供此指標。
4. 在從事預測時，實務管理的經驗，直覺，實驗上的證據，及環境變遷與政策改變等問題都要在選擇模式時加以考慮（參閱 Helmer & Johansson, 1977）。本文在決定  $b$  的期數時，皆以實際考量來下決斷。
5. 在目前有關轉換函數模式的文獻對於  $w(B)$  與  $\delta(B)$  之秩如何決定皆語焉不詳，包括 Box & Jenkins (1976) 的書亦然。為解決此問題，可用地氈式搜索法，將可能的組合一一嘗試，然後比較各模式的適合性。這是本文所用的方法。看起來很笨拙，但覺得並不見得浪費時間。
6. 轉換函數模式是以一個自變數列及一個依變數列來預測依變數列，希望使依變數列的預測更加準確。事實上影響依變數列的，可能不只一個。像迴歸有多元迴歸一樣，轉換函數模式也可用多個自變數列來進行預測。目前已有多位研究者討論到多元時間數列分析 (multiple time-series analysis)，例如 Liu & Hanssens (1982); Liu (1987); Hillmer, Larcker, & Schroeder (1983); Newbold (1983); Montgomery & Weatherby (1980)；及 Riise & Tjøstheim (1984) 等。一般要作轉換函數模式時，自變數列與依變數列的相關要大才可以，愈大愈佳。Hillmer, Larcker, & Schroeder (1983) 發現多元轉換函數模式比二元轉換函數模式更能準確預測，而二元轉換函數模式的預測又比單變量 ARIMA 的預測更準。本文的轉換函數模式是指二元轉換函數模式。然而 Riise & Tjøstheim (1984) 的研究發現由於兩個數列的後段期相關低，導致二元轉換模式的預測誤差較單變量的為大。所以選擇自變數列時要特別謹慎。一般多變量的轉換函數模式公式是：

$$y_t = c + \sum_{j=1}^m \frac{w_j(B)}{\delta_j(B)} x_{j,t-b} + \frac{\theta(B)}{\phi(B)} a_t.$$

其中  $j$  是自變數列的數目，可有  $m$  個自變數列。目前 SAS 的 ETS 裏的 ARIMA 也可作多變量的轉換函數模式。

7. 時間數列模式分析無論是單變量 ARIMA 或是多變量轉換函數模式，基本上有其弱點：即影響依變數列的政策忽然被引進時，最初幾期便無法準確預測，並且可能會有方向相反的預測。例如我國優生保健法使墮胎合法化，造成出生數的銳減。延長以職業教育為主的國民教育使職業補習學校學生數遽增。對於這些政策性的影響，時間數列分析便無法準確預測出數字來。即使用時間數列的介入法 (intervention)，也要在政策發生後有一些觀察值可運算，才能診斷出模式，進而預測往後的數字。

研究的目的在了解 (understanding)，預測 (predicting)，及控制 (controlling)。在自然科學裏，科學家已進入控制階段，但教育科學裏大部分仍停留在了解階段。筆者祈望教育科學能早日進入預測階段，因此想借本文拋磚引玉，引發教育界的研究者投入這方面的研究。隨著研究人數的增多，方法與技術會更臻完善，預測自會更精確。對我國教育政策與教育計畫的擬定將有很大助益與貢獻。

## 參考書目

### 一、中文部份

行政院經建會人力規劃小組。（民73）。中華民國臺灣地區73年至100年人口推估。臺北。

行政院衛生署，臺灣省衛生處，臺北市衛生局，高雄市衛生局（民78）。中華民國七十七年衛生統計：公務統計。

馬信行。（民76）。我國各級學校未應學生數之預測。政大學報，第56期，頁111~147(a)。

馬信行。（民76），簡介迴歸分析。國教研究，1, 60-69, (b)。

教育部。（民78）。中華民國教育統計。臺北。

### 二、英文部份

Box, G.E.P. & Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Downing, D.J., Pack, D.J. & Westley, G.W. (1983). A diverting structure's effects of a river

時間數列分析之轉換模式在學生數預測上之應用

- flow time series. *Management Science*, 29, 225-236.
- Edlung, P. (1984). Identification of the multi-input Box-Jenkins transfer function model. *Journal of Forecasting*, 3, 297-308.
- Helmer, R.M. & Johansson, J.K. (1977). An exposition of the Box-Jenkins transfer function analysis with an application to the advertising-scales relationship. *Journal of Marketing Research*, 14, 227-239.
- Heyse, J.F. & Wei, W.W.S. (1985). Modeling the advertising-sales relationship through use of multiple time series techniques. *Journal of Forecasting*, 4, 165-181.
- Hillmer, S. C., Larcker, D. F., & Schroeder, D. A. (1983). Forecasting accounting data: A multiple time-series analysis. *Journal of Forecasting*, 2, 389-404.
- Hopwood, W.S. (1980). The transfer function relationship between earnings and market-industry indices: An empirical study. *Journal of Accounting Research*, 18, 77-90.
- Hopwood, W. S. & McKeown, J.C. (1981). An evaluation of univariate time-series earnings models and their generalization to a single input transfer function. *Journal of Accounting Research*, 19, 313-322.
- Liu, L.M. & Hanssens, D.M. (1982). Identification of multiple input transfer function models. *Communication in Statistics. A* 11, 297-314.
- Liu, L.M. (1987). Sales forecasting using multi-equation transfer function models. *Journal of Forecasting*, 6, 223-238.
- Montgomery, D.C. & Weatherby, G. (1980). Modeling and forecasting time series using transfer function and intervention methods. *AIEE Transactions*, 12, 289-307.
- Newbold, P. (1983). ARIMA model building and the time series analysis approach to forecasting. *Journal of Forecasting*, 2, 23-35.
- Riise, T. & Tjøstheim, D. (1984). Theory and practice of multivariate ARMA forecasting. *Journal of Forecasting*, 3, 309-317.
- SAS Institute Inc. (1984). SASIETS user's guide, (5. edition). North Carolina: SAS Institute Inc.
- Umstead, D.A. (1977). Forecasting stock market prices. *The Journal of Finance*, 32, 427-441.
- Vandaele, W. (1983). *Applied time series and Box-Jenkins models*. New York: Academic Press.