

國立政治大學應用數學系

碩士學位論文

應用基因演算法決定 SETAR 門檻區間及其

應用

Use genetic  
algorithms to determine the SETAR  
threshold interval and Its Applications

碩士班學生：江增堂 撰

指導教授：吳柏林 博士

中華民國 一百零二年 一月 十七日

## 謝誌

首先要感謝我的指導教授 吳柏林老師，在我寫作論文時遇到困難時給予我方向以及鼓勵，老師總是用很幽默的方式教導了我很多事情，不論是作研究或是做人處事都讓我受益良多。再來，我要感謝口試委員劉老師以及溫老師在口試時給我的建議，讓此論文可以更臻完善。

接著我要感謝我的父母親，在求學的一路上給我最大的支持跟資源，讓我在安逸的環境下盡情學習我想要學習的數學，每每看到他們臉上新長出細紋而不曾說辛苦，總讓我感念在心中。在這最後半年的研究生生涯，雙親身體受到些病痛，所幸在家人同心協力、互相扶持之下，順利度過難關，我要特別感謝我的哥哥跟弟弟，在我很多時候由於忙於學業論文而沒有辦法照料雙親，他們會體恤我並且照應父母，我愛我的家人，我今日以及往後有任何成就都是因為有家人做我最強力的後盾。

在政大的研究生涯當中，結交到了很多好朋友，而讓我生活充滿樂趣跟精彩。李沛承承受了我的一切心事並且給我最忠實的忠告，也常常帶給大家歡樂，我很幸運可以交到這樣的朋友；陳俊愷學弟是經由打籃球結識的，他就像我大學時期的樣子，我常常可以從他身上找尋到我有時候幾乎快要忘記的青春熱血的模樣，阿愷在歡送我的歡送會上真情流露，讓我一輩子不會忘記。其他還要感謝的還有李李仁學長、瑋慈麻吉學姐、YOYO、高裕哲、劉宇恩、張芷菱、阿國學長、兵兵學姐，讓研究室很歡樂也讓我解放很多壓力，特別要感謝林澤佑大神同學，如果不是他的耐心教導，我也不會畢業。還有球隊當中的學長學弟們，高世光、小豬，每次搭計程車上山練球讓我印象難忘，也一起打贏了很多比賽，咬牙練習的過程尤其讓我很快樂，大家一起朝向一個目標努力的感覺很美好。

再來我想要感謝我在台北的家教家人，恩浩一家人以及子寧一家人，他們提供了我在台北的經濟來源，並且把我當成他們家裡的一份子，我很常受到他們實質上的贊助以及精神上的關心，恩浩媽媽讓我每次家教有吃又有拿，有什麼紀念品也會帶一份給我，不管是生活上遇到任何問題恩浩媽媽都會幫我處理，真的很感謝她；子寧媽媽打的果汁很好喝，也會常常關心我，很熱情也對我很好。我在家教的時候，常常會以他們的家庭氣氛為目標，思考以後我所組成的家庭如何像兩位家教家裡面一樣和氣融洽，並且培養出下一代正確的價值觀跟面對壓力所應該具有的正面思考。這兩個家庭就像是我在台北的家人，我以後如果有什麼成就也一定會好好地與他們分享。

很多老朋友在我成長學習過程當中也給予我很多幫助，李昆洲同學是最好的高中朋友，有任何事情他都會聽我說，也很挺我，還有阿訥、子祥、宥辰、芭樂、家誠、阿慶、小孩子、阿暴學長、仁甫、肥羊、歪頭、砲灰、志儒等等，我也想要謝謝你們，精彩且充實了我的生活。

最後我要非常感謝啤酒，在我心情低落的時候陪著我，聽我講心中的挫折也給我很多協助，只要有你在，我認為任何事情都可以迎刃而解；開心的時候也陪著我四處遊玩，讓我釋放了壓力進而可以繼續面對挫折跟逆境。謝謝你，那些酸甜苦辣鑄成了生命中每一個深刻的記憶，我都十分感念跟感謝。

謝謝上蒼，發生在我身上的一切都是祂的恩典，我會常常心存感激並且幫助更多需要幫助的人。

## 摘要

近年來，面對傳統線性時間序列的預測問題，有許多技術上的改良而被大量廣泛的使用，但是線性模式往往無法處理常常發生結構改變(structural changes)的問題，這使得非線性(nonlinearity)時間序列轉折點的研究越來越受到重視，利用非線性時間序列解決實例更可以貼近真實情況。再者，隨著模糊理論的蓬勃發展以及區間軟計算(soft computing)的成熟，相較於點估計預測方法所需的嚴格假設，區間估計方法的假設寬鬆許多並且能符合實際情況，可以提供給決策者更彈性的選擇。本文將應用基因演算法(genetic algorithms)針對模糊區間資料(fuzzy data)作模糊分析(fuzzy analysis)，找出資料轉折的門檻區間(threshold interval)，藉此發展出非線性的區間門檻自迴歸模式(interval SETAR model)，最後以台股為例，建構出門檻自迴歸模型與傳統區間 ARIMA 模式比較，藉此探討其預測方法的效率評估與準確性。

關鍵字：非線性、區間軟計算、模糊分析、基因演算法、門檻自迴歸、門檻區間



## Abstract

In recent years, in the face of traditional linear time series forecasting problems, there are many technical improvements and widely used. But linear model are often unable to deal with the problem often happens structural changes, which makes the nonlinear turning point for the study of the time series more and more attention. Use nonlinear time series more close to the real situation. Moreover, with the fuzzy theories flourish and soft computing mature, compared to the point estimate methods required strict assumptions, interval estimation method which without many assumptions can meet the actual situation. It can be provided to decision-makers more flexibility of choice. In this paper, the application of genetic algorithms for fuzzy data to identify structural changes interval (threshold interval), so as to develop the nonlinear range threshold autoregressive mode (interval SETAR model), and finally, for example, the Taiwan stock market, construct a threshold autoregression model with the traditional interval ARIMA model to investigate the prediction method efficiency and accuracy.

Keyword: nonlinear 、 soft computing 、 fuzzy analysis 、 genetic algorithms 、 SETAR 、 threshold interval



## 目錄

1.前言.....	1
2 研究理論與方法.....	3
2.1 門檻自迴歸模式.....	3
2.2 區間型門檻自迴歸模式.....	5
2.3 門檻區間自迴歸模式.....	7
2.4 基因演算法.....	8
2.5 模糊時間數列基因演算法.....	11
3.實證分析.....	13
3.1 資料分析.....	13
3.2 建構SETAR門檻區間.....	16
3.3 用外生多變數建構門檻轉換模式.....	18
4. 結論與建議.....	21



## 1.前言

傳統研究中，利用時間序列分析資料以及預測的種種方法已經在許多領域被廣泛的使用，然而使用上的限制的條件像是穩定性(stationary) 的隨機過程就顯得十分嚴格，但是時間數列資料常常存在著非線性的現象，為了解釋非線性的時間數列資料，許多學者提出了門檻模式(TAR model)、自迴歸條件異質變異數模式(ARCH model)、GARCH 模式、雙線性模式(Bilinear model)、排列自迴歸(Arranged autoregression)(Tsay,1989)等等，而被拿來適用的研究領域也很廣泛，像是農業(Machado,1995)、經濟(Stock and Watson,1996)、財金(Hansen,1999、2000)等等。

Tong 在 1983 年提出的門檻自迴歸模式(threshold autoregressive model)，即是這類問題的最初解決模型，門檻自迴歸模式可以較不受影響的解釋時間數列資料，比起一般線性 ARIMA 傳統模式可以描述的好得多，因此受到重視並且更多學者投入推廣發展。早期的 Tong and Lim(1980)建構門檻自迴歸模式分析加拿大山貓(Canadian lynx)數量，爾後也有 Tiao and Tasy(1994)研究美國自 1947 年至 1991 年間調整後的 GNP(really gross national production)資料，並且分別以線性 AR 模型及 TAR 模型進行分析比較。在非穩定型的資料建構方面，普遍已經可以接受透過轉折區間(change point)的概念，也就是將非線性資料依其結構改變的情況區分成兩個或是多個區域，然後每個區域內資料可服從某個穩定型的時間數列模式，見 Wu and Chen(1995)。這個概念提供了另一種可行的研究方向，針對這類問題，Nyblom(1989)、Bleaney(1990)及 Ploberger and Kramer(1992)等多位學者均提出不同轉折發生的檢定方法，用來確認資料是否有結構性的轉變，透過種種文獻可以發現利用門檻自迴歸模式不僅有較佳的配適度也能忠實地呈現資料型態。

而在過去傳統的時間數列預測當中，使用的資料以及預測數往往都是單一值，這在運用上會有所不方便以及失去彈性，像是日常生活中的每日氣溫、每日美金對台幣的匯率、每年使用的汽油量等等，都是由區間資料來描述以及預測，才能忠實地表達出其真實情形。基於上述這些原因，越來越多學者透過電腦軟計算發展出區間預測的演算法，早期的 Chatfield(1993)、Wu and Chen(1995)用修正後的中心化累加平方和法及模糊熵方式尋找出轉折區間，Tseng、Tzeng、Yu and Yuan(2001)即透過模糊 ARIMA 模式預測美金與台幣之間的匯率，Liu(2007)的研究透過隸屬度函數改良對模糊時間序列的影響，Hsu(2008)針對區間時間序列作分析及預測，區間的預測方法很多也具有一定效率，也更貼近真實生活中的種種不確定語言。

延續上方理論以及概念，本研究將從統計模式理論以及基因演化的概念進行分析，引用 Holland(1975)所提出的基因演算法(Genetic Algorithm)處理資料。基因演算法是一種以目標為導向的平行多點搜尋技術，並非單一點搜尋法，主要應用在最佳化問題上，見 Goldberg(1989)，近年來已廣泛與模糊理論結合而成為人工智慧的重要演算法之一，見 Loraschi, Tettamanzi, Tomassini and Verda(1995)。

綜合上述，本文將以區間資料為例，使用非線性模式來配適模型，透過基因演算法找出每代適合資料之統計模式的演化，這樣的時變基因模式比較能符合社會演變的脈動。在此研究中有別於一般基因演算法先對研究參數進行基因編碼的運算過程，而是先定義適應函數來取得最佳解再尋找模式參數，接著根據資料本身特性及對資料精準度的需求，訂定轉折決策規則，來判斷時間序列資料是否有結構性的改變，尋找出轉折區間以後再將資料分群，分別配適最佳模式。最後，利用台灣每週股價的最高值及最低值當資料，以本文提出的規則來尋找出走勢變化的轉折區間並且配適模式，以此驗證本文所提出的分析以及預測方法。

本文第一章為前言。第二章為 SETAR 模式介紹以及基因演算法的理論及方法。第三章為以台灣股價為實例。第四章為結論與建議。



## 2 研究理論與方法

### 2.1 門檻自迴歸模式

門檻自迴歸模式(threshold autoregressive models, TAR)是由 Tong(1978)首度提出，用來處理傳統的單根檢定方法無法探討資料中不對稱且非線性的關係。此模型的基本原理為依據時間數列當中的觀察值的走勢變化，觀察並找出資料中的延誤參數(delay parameter)以及結構轉變的點，稱之為轉折點(change point)，亦或稱之為門檻值(threshold)。而被門檻值劃分開來的資料則會自成一個系統，進而可以針對不同的系統分別建立模式以符合資料型態。舉例而言，股票市場會因為政經情勢不同或是發生某些重大事件而在短時間內突然持續上漲或下跌，例如：來台觀光人數會隨著政府開放政策而有很明顯的差異、生物族群會受到環境影響而突然銳減等等，這些具有上下起伏或是震動的資料很明顯的存在著轉折點。在此特別要注意的是資料具有轉變型態以及趨勢(trend)這兩種情況在資料走勢上的明顯不同，見圖 2.1 及圖 2.2。

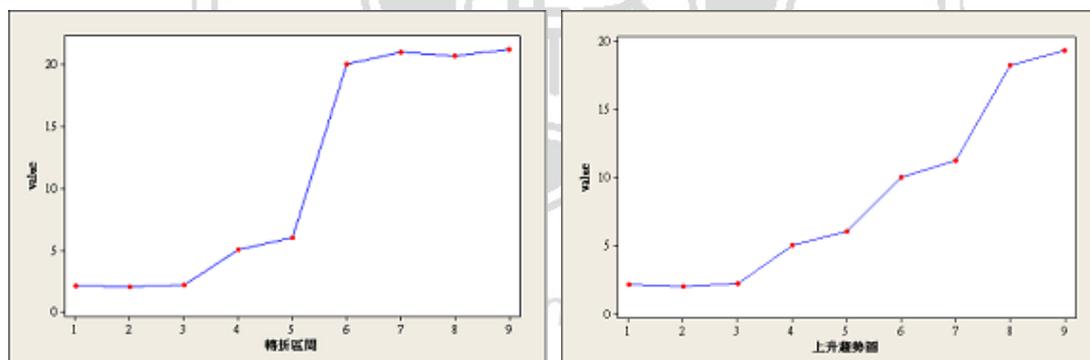


圖 2.1：具有轉折型態的資料型態

圖 2.2：具有上升趨勢的資料型態

Tong and Lim(1980)根據數個線性自迴歸模式(AR models)及切換結構的組成繼續發展此模型，利用門檻變數於門檻值不同之區間來判定變數自我迴歸之相關情形之研究。大抵來說，門檻自迴歸模式可以分為兩類：第一種是以「時間」為門檻值，針對時點作切割來描述資料具有轉折型態(piecewise in time)，第二種則是以「變數」作為門檻值，用變數本身去解釋資料轉折(piecewise in variable)，而門檻自迴歸模式亦稱為 SETAR 模式(Self-Excited TAR)。

以一個兩階段的門檻自迴歸(TAR(2;  $k_1$ ,  $k_2$ ), 模式)為例，其表示方式如下：

$$Y_t = \begin{cases} \Phi_{1,0} + \sum_{i=1}^{k_1} \Phi_{1,t} Y_{t-i} + a_{1,t} & Y_{t-d} \leq r \\ \Phi_{2,0} + \sum_{i=1}^{k_2} \Phi_{2,t} Y_{t-i} + a_{2,t} & Y_{t-d} > r \end{cases} \quad (2.1)$$

(2.1)式中的  $k_1$ 、 $k_2$ ：兩個自迴歸模式的階次(order);

$\Phi_{1,i}$ 、 $\Phi_{2,i}$ ：自迴歸係數

$a_{1,t}$ 、 $a_{2,t}$ ：白干擾項(white noise)

d：延誤參數，及控制門檻發生之處

t：期數或時間

r：為一常數，稱之為門檻值

另外也有以時間點分段的 TAR 模式，其表示法如下：

$$Y_t = \begin{cases} \Phi_{1,0} + \sum_{i=1}^{k_1} \Phi_{1,t} Y_{t-i} + a_{1,t} & t \leq s \\ \Phi_{2,0} + \sum_{i=1}^{k_2} \Phi_{2,t} Y_{t-i} + a_{2,t} & t > s \end{cases} \quad (2.2)$$

(2.2)式中的  $k_1$ 、 $k_2$ ：兩個自迴歸模式的階次(order);

$\Phi_{1,i}$ 、 $\Phi_{2,i}$ ：自迴歸係數

$a_{1,t}$ 、 $a_{2,t}$ ：白干擾項(white noise)

t：期數或時間

s：為某時期或某一時點

由上述式子可知，門檻自迴歸模式會透過門檻值進行切換(switch)。而更進一步的，區間模糊資料也被應用在相似的模式下進行延伸研究，透過區間資料可以使分析具有彈性並且符合實際情況。

## 2.2 區間型門檻自迴歸模式

傳統上針對時間數列都是單點對單點的預測方式，使用上侷限於單點資料，在充滿模糊資訊的社會中並不實用。所以很多學者利用了模糊資料來建構模式以符合問題，區間型門檻自迴歸模式即為常見的模式，以下介紹並且定義區間型門檻自迴歸模式，首先在此必須先定義區間時間數列：

### 定義 2.1 區間時間數列(Interval time series; Wu, 2006)

令  $\{Y_t\} = [a_t, b_t] = (c_t; r_t), t = 1, 2, \dots, n$ ，則稱  $\{Y_t\}$  為一個區間時間數列，其中  $c_t = \frac{a_t + b_t}{2}$ ， $r_t = \frac{b_t - a_t}{2}$ ，每一個  $\{Y_t\}$  都稱之為區間模糊數。

透過區間時間數列定義可以建構區間門檻自迴歸模式，而建構門檻自迴歸模式有很多種，以下介紹以左右端點各自獨立以及以取中間點與半徑所定義出來的門檻自迴歸模式：

### 定義 2.2 左右端點法之區間門檻自迴歸模式(Hsu, 2008)

令  $\{Y_t\} = [a_t, b_t] = (c_t; r_t), t = 1, 2, \dots, n$  為一個區間時間數列，則可以分別處理並且定義如下：

$$a_t = \begin{cases} \Phi_{a,1,0} + \sum_{i=1}^{k_1} \Phi_{a,1,t} a_{t-i} + \varepsilon_{a,1,t} & a_{t-d_1} \leq S_1 \\ \Phi_{a,2,0} + \sum_{i=1}^{k_2} \Phi_{a,2,t} a_{t-i} + \varepsilon_{a,2,t} & a_{t-d_1} > S_1 \end{cases} \quad (2.3)$$

$$b_t = \begin{cases} \Phi_{b,1,0} + \sum_{i=1}^{k_1} \Phi_{b,1,t} b_{t-i} + \varepsilon_{b,1,t} & b_{t-d_2} \leq S_2 \\ \Phi_{b,2,0} + \sum_{i=1}^{k_2} \Phi_{b,2,t} b_{t-i} + \varepsilon_{b,2,t} & b_{t-d_2} > S_2 \end{cases} \quad (2.4)$$

(2.3)、(2.4)式中的  $\Phi$  為自迴歸模式係數， $S$  為門檻值， $a_{t-d_1}$ 、 $b_{t-d_2}$  為延誤參數，

$\varepsilon_{a,1,t}$ 、 $\varepsilon_{a,2,t}$ 、 $\varepsilon_{b,1,t}$ 、 $\varepsilon_{b,2,t}$  為白干擾項(white noise)，則

$$\hat{Y}_t = E[Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1] = [\hat{a}_t, \hat{b}_t]$$

$$\hat{a}_t = E[a_t | a_{t-1}, a_{t-2}, \dots, a_1] = \begin{cases} \Phi_{a,1,0} + \sum_{i=1}^{k_1} \Phi_{a,1,t} a_{t-i} & a_{t-d_1} \leq S_1 \\ \Phi_{a,2,0} + \sum_{i=1}^{k_2} \Phi_{a,2,t} a_{t-i} & a_{t-d_1} > S_1 \end{cases} \quad (2.5)$$

$$\hat{b}_t = E[b_t | b_{t-1}, b_{t-2}, \dots, b_1] = \begin{cases} \Phi_{b,1,0} + \sum_{i=1}^{k_1} \Phi_{b,1,t} b_{t-i} & b_{t-d_1} \leq S_2 \\ \Phi_{b,2,0} + \sum_{i=1}^{k_2} \Phi_{b,2,t} b_{t-i} & b_{t-d_1} > S_2 \end{cases} \quad (2.6)$$

上述方法並沒有將區間的長度值影響考慮進去，故以下有以中心點以及半徑建構出來的門檻自迴歸模式：

### 定義 2.3 中心點及半徑法之區間門檻自迴歸模式(Hsu, 2008)

令  $\{Y_t\} = [a_t, b_t] = (c_t; r_t), t = 1, 2, \dots, n$ ，則稱  $\{Y_t\}$  為一個區間時間數列，利用中心點跟半徑的概念則可定義如下：

$$c_t = \begin{cases} \Phi_{c,1,0} + \sum_{i=1}^{k_1} \Phi_{c,1,t} c_{t-i} + \varepsilon_{c,1,t} & c_{t-d_1} \leq S_1 \\ \Phi_{c,2,0} + \sum_{i=1}^{k_2} \Phi_{c,2,t} c_{t-i} + \varepsilon_{c,2,t} & c_{t-d_1} > S_1 \end{cases} \quad (2.7)$$

$$r_t = \begin{cases} \Phi_{r,1,0} + \sum_{i=1}^{k_1} \Phi_{r,1,t} r_{t-i} + \varepsilon_{r,1,t} & r_{t-d_2} \leq S_2 \\ \Phi_{r,2,0} + \sum_{i=1}^{k_2} \Phi_{r,2,t} r_{t-i} + \varepsilon_{r,2,t} & r_{t-d_2} > S_2 \end{cases} \quad (2.8)$$

(2.6)、(2.7)式中的  $\Phi$  為自迴歸模式係數， $S$  為門檻值， $c_{t-d_1}$ 、 $r_{t-d_2}$  為延誤參數， $\varepsilon_{c,1,t}$ 、

$\varepsilon_{c,2,t}$ 、 $\varepsilon_{r,1,t}$ 、 $\varepsilon_{r,2,t}$  為白干擾項(white noise)，則

$$\hat{c}_t = E[c_t | c_{t-1}, c_{t-2}, \dots, c_1] = \begin{cases} \Phi_{c,1,0} + \sum_{i=1}^{k_1} \Phi_{c,1,t} c_{t-i} & c_{t-d_1} \leq S_1 \\ \Phi_{c,2,0} + \sum_{i=1}^{k_2} \Phi_{c,2,t} c_{t-i} & c_{t-d_1} > S_1 \end{cases} \quad (2.9)$$

$$\hat{r}_t = E[r_t | r_{t-1}, r_{t-2}, \dots, r_1] = \begin{cases} \Phi_{r,1,0} + \sum_{i=1}^{k_1} \Phi_{r,1,t} r_{t-i} & r_{t-d_2} \leq S_2 \\ \Phi_{r,2,0} + \sum_{i=1}^{k_2} \Phi_{r,2,t} r_{t-i} & r_{t-d_2} > S_2 \end{cases} \quad (2.10)$$

## 2.3 門檻區間自迴歸模式

上述的自迴歸模式已能大幅度的符合真實區間資料的型態並且得以使用，但是對於選取某個單點值來當作門檻值依然存在著點估計的種種問題，故本研究將延續上述概念，先對區間資料反模糊化後再透過基因演算法(Genetic algorithm)建構轉折區間，最後再將資料群皆還原成原本的區間資料，得到一門檻轉折區間，將轉折資料尋找出來並且以此區分資料群再接著分別建構最佳模式。特別一提的是，門檻轉折區間的資料群常常具有動盪或是波動的資料型態，在本研究中皆視為不穩定的時間序列資料，將其資料群收集起來且不加以建模以達到可以完全區隔低門檻值區間資料以及高門檻值區間資料的功能。

傳統時間數列的模式選取準則，像是 AIC(Akaike information criteria)、SBC(Schwartz bayesian information criterion) 及 CAT(criterion for autoregressive transfer functions)都可以用來評估自迴歸模式的優劣(吳柏林,1995)，而在本文中將針對基因演算法演算出的每一個候選門檻區間計算 AIC 值以得到最合適的門檻自迴歸模式。

首先必須先定義區間資料反模糊化：

### 定義 2.4 區間反模糊化(吳柏林, 2012)

若  $A = [a, b]$  為一區間模糊數， $c = \frac{(a+b)}{2}$  為區間中心點， $L = |b - a|$  為區間長度，則此區間的反模糊化定義為  $S = c + \frac{L}{2\ln(e+c)}$  (2.11)

例 2.1: 設某組區間資料  $A_1 = [2,8]$ ，則區間中心點為  $\frac{(2+8)}{2} = 5$ 、區間長度為  $8 - 2 = 6$ ，則  $S = 5 + \frac{6}{2\ln(e+5)} = 6.468$ ，另一組區間資料  $A_2 = [0,10]$ ，則區間中心點為

$\frac{(0+10)}{2} = 5$ 、區間長度為  $10 - 0 = 10$ ，則  $S = 5 + \frac{10}{2\ln(e+5)} = 7.447$

由例子可以發現中心點相同的兩個區間會根據區間的長度而受影響，也就是必須考慮區間長度所佔的比重，故這裡將區間長度除上自然對數，但是影響還是過大，所以再將其除以二，進而得到此轉換公式。

有了反模糊化公式，可以將區間資料先反模糊化後搭配基因演算法，透過準則尋找出資料中的那些轉折點，接著把那些轉折點對照回原本的區間資料，建構出轉折區間。

所以必須利用下一節所介紹的基因演算法的準則去建構出門檻區間，用來取代傳統的門檻值，使得時間序列的資料更具有可用性跟延伸性。

## 2.4 基因演算法

基因演算法是一種由大自然中生物體自然進化的基本架構所激發出來的方法，它是由 J.H. Holland 博士於 1975 年首次提出，運用了達爾文博士 (Dr.Charles Darwin) 的「物競天擇，適者生存」的原理，結合電腦計算機程式發展「演化」(Evolution) 機制，例如：遺傳、選擇、交配、突變等等，從原始族群中透過競爭選出較好的物種進行交換資訊，用以產生優良的下一代，持續這個演化過程，直到找到最具有競爭力及適應能力的後代 (Mitchell, 1996)。

傳統上，必須將資料轉換成二元數字列(binary string)來表示，這個角色就像是生物體內的染色體(chromosome)，像是: {01110}、{11010}這兩種字串即為兩種染色體，而其中的 0 與 1 即是染色體當中的基因。這些最初的字串被稱之為親代(parents)，然後透過交配生出了子代(child)，接著透過為達到最佳化目的所定義出來的適應函數(fitness function)，計算該數字串的適應度(fitness)。唯有適應能力越強的染色體才有越多機率存活下來，並且再繁殖出下一代的子女。也就是說，適應度越高的染色體有較高的機會被選擇 (selection)來進行複製(reproduction)，複製後，再根據交配率及突變率對字串進行隨機性交配(crossover)及突變(mutation)，以孕育出新的下一代子女。如此一代接著一代，周而復始地進行演化過程即為基因演算法的概念及步驟。

以下為基因演算法的名詞介紹以及計算程序:

### 選擇(selection)與複製(reproduction)

選擇的機制要透過適應函數，在適應函數中有優良表現的染色體會被留下，而根據基因演算法可以決定染色體被選擇跟複製的機率，所以下一代族群必會有更高適應度的染色體，透過進化機制來提昇下一代族群的整體品質。至於複製的個數，則決定於選擇的方法，以下有兩種選擇方式(Goldberg and Deb,1991)

#### (1).輪盤法(roulette wheel proportionate)

這個方法是依照每個染色體其適應度占全體族群的比例而定。即第  $i$  個染色體被選擇的機率為:

$$P_i = \frac{f_i}{\sum_{i=1}^N f_i} \quad (2.12)$$

$N$  為族群的數目， $f_i$  為第  $i$  個染色體的適應度。由此可以發現(3.12)只是簡單的比例公式，而每個染色體的複製個數為  $n_i = N \times P_i$ 。

#### (2).排序法(ranking selection with roulette wheel)

排序法是將  $N$  個染色體的適應度按照大小排列，然後將適應度轉換成排序值 (rank)，最大的稱為  $N$ ，最小的稱為 1。接著再根據輪盤法來選擇與複製染色體，此方法的機率為相對機率而不是絕對機率，因此在染色體適應度相差不大時，也能選擇出較高適應度的染色體。

### 交配(crossover)與突變(mutation)

生物體與生俱來的有性生殖，提供了交換資料的途徑，使得族群中各式各樣的染色體可以透過隨機性的交配過程，產生新的子代，新一代的子代中的染色體既遺傳了親代的某些特性(即基因)，也同時擁有部分差異，不過大抵上整體表現會比親代來的優秀，所以可以非常有效率的來尋找出問題的最佳解。

交配的頻率越是交配率的大小而定，交配率越高，新的染色體進入族群的速度就越快，而找到最佳解的速度就越快。但是，相同地，好的染色體被族群被剔除的速度也可能越快，而喪失了交配的功能。因此，交配率的大小必須按照每個實例不同而有所不同。以下即為交配率 0.5 的例子(見圖 2.3)，上方兩列字串表示親代，下方產生出來的為子代:

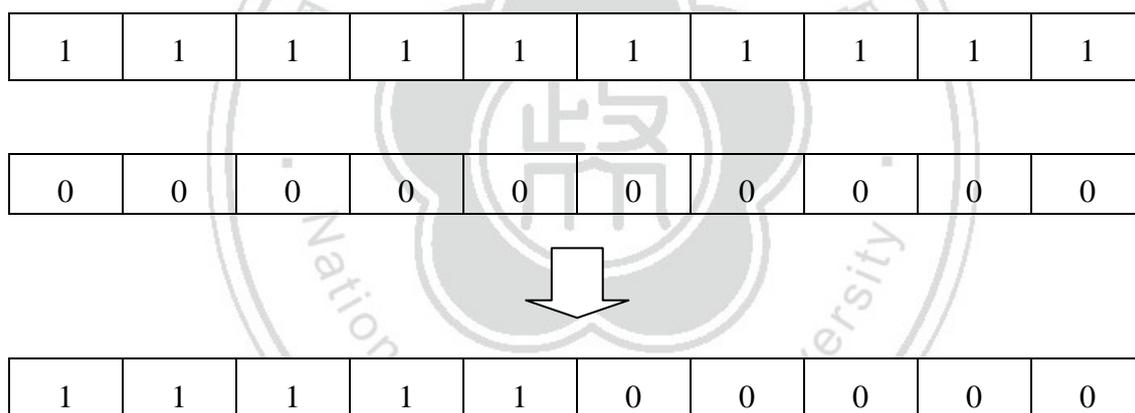


圖 2.3：遺傳模式中交配率為 0.5 的字串，其中最上方字串稱之為父親，中間字串為母親，最下方的字串為子代

而染色體當中如果有某些基因完全相同時，其值經過了複製以及交配仍然會不變，這麼一來就會限制了新的染色體進入族群的機會，以致於找不到最佳解。為避免發生這種情況，必須在演算法當中加入突變的狀況以確保物種的多變性，作法即為下圖所示(見圖 2.4)。在自然界當中的突變機率非常低，所以在這個演算法當中也會將突變率訂的很低，一方面避免與一般的隨機性搜尋無異，另一方面也更貼近自然界當中的生物遺傳機制。

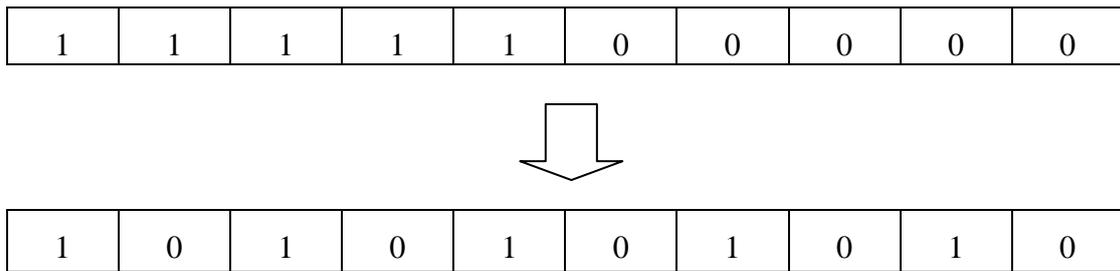


圖 2.4：遺傳模式中突變率為 0.5 的字串

透過圖 2.5 可以更清楚基因演算法的過程：

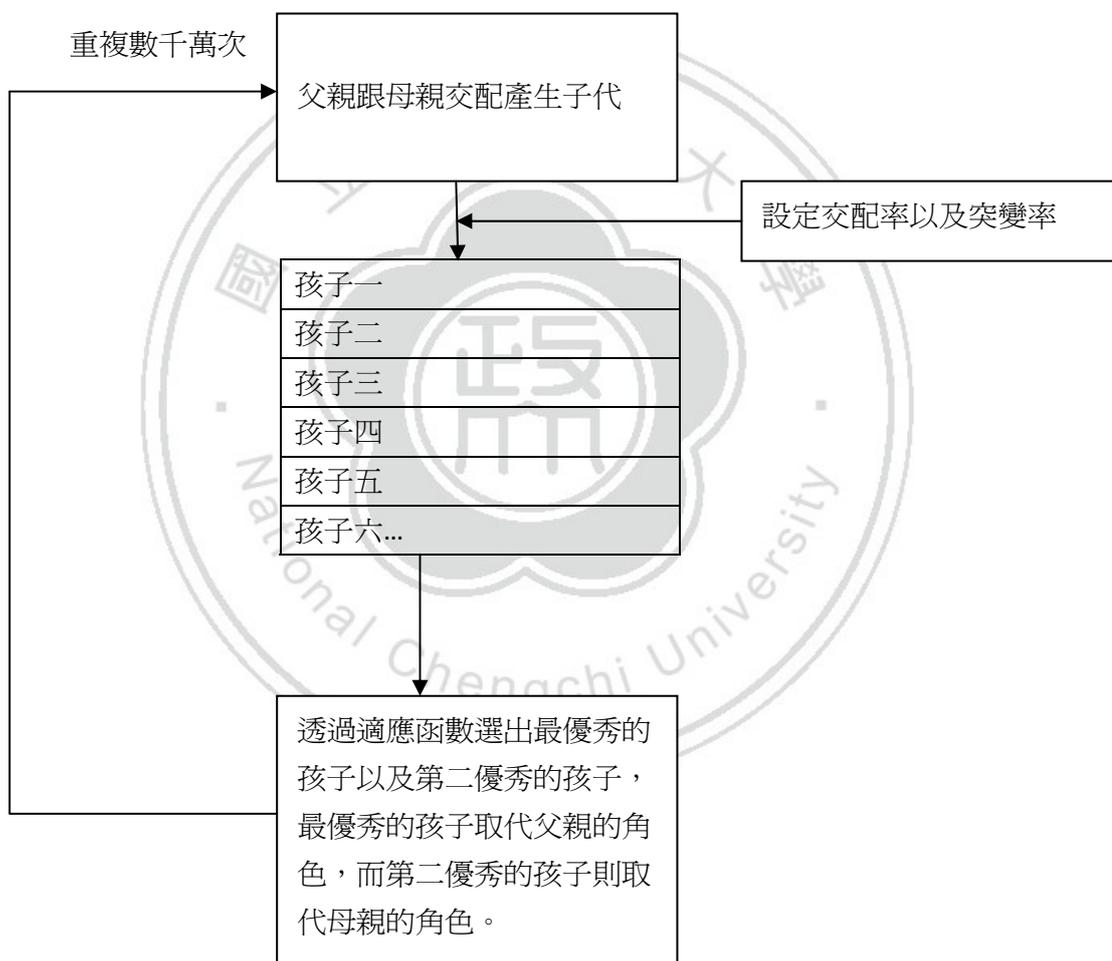


圖 2.5：基因演算法流程圖

## 2.5 模糊時間數列基因演算法

以時間數列資料建構模式的方法發展至今已經有相當多高效率的方法，本研究將針對模糊資料輔以基因演算法，找出門檻轉折區間，建立最佳模式，其流程敘述如下：

- 步驟 1：先將模糊區間資料反模糊化。
- 步驟 2：編碼並且隨機產生數個染色體的原始族群。
- 步驟 3：給定適應函數，接著透過基因演算法找出最佳解。
- 步驟 4：利用門檻區間轉折準則找出資料中的轉折值資料群，也就是門檻值。
- 步驟 5：最後將這些門檻值還原成原始的區間資料，利用統計方法建構出門檻區間。
- 步驟 6：由於轉折區間值具有震盪不穩定的性質，故在此扣除掉門檻區間資料群，將高數值時間數列資料跟低數值時間數列資料分別建模。

區間資料在進行時間數列分析時會面臨到如何分辨資料的結構存在轉折，這也是本研究的重點，本研究以二元數字列的數即代表投資與否，也就是“1”即代表投資信號(buying sign)，“0”即代表不投資，針對這種投資動作與否，以下列出幾個判定門檻轉折的準則。

### 決策準則

- (1). 投資信號必須連續出現六次以上。  
例如:第 10,11,12,13,14,15 天都出現“1”這個投資信號，故這六天都為買點。
- (2). 兩串投資信號中間夾著一個以下不投資信號，仍然視為一串投資信號。  
例如:第 2,3,4,5,6,8,9,10,11,12 天出現“1”這個投資信號，可以發現第 7 天並不是投資信號，但是在此我們仍然把信號連貫，也就是將第 7 天也視為投資信號。
- (3). 三串投資信號中間夾著兩個以下不投資信號，仍然視為一串投資信號。  
例如:第 1,2,3,4,5,7,8,9,10,11,12,14,15,16,17,18 天出現“1”這個投資信號，可以發現第 6 天以及第 13 天皆不是投資信號，但是在此我們仍然把信號連貫。

本研究透過上述兩個準則，可以將低門檻資料群以及高門檻資料群區隔開來，最後分別取左端點跟右端點的平均，即可建構出門檻區間，再對照回原始資料，將其兩段資料分別透過 ARIMA 模式尋求最適係數，以得到最佳模式，可見下式(2.13)。

$$f_t = \begin{cases} \Phi_{f,1,0} + \sum_{i=1}^{k_1} \Phi_{f,1,t} f_{t-i} + \varepsilon_{f,1,t} & f_{t-d_1} < S_1 \\ \Phi_{f,2,0} + \sum_{i=1}^{k_2} \Phi_{f,2,t} f_{t-i} + \varepsilon_{f,2,t} & f_{t-d_1} > S_2 \end{cases} \quad (2.13)$$

當中 $[S_1, S_2]$ 即為轉折區間

$f_t$ 為反模糊化後的數值，

$\Phi$ 為自迴歸模式係數，

$S$ 為門檻值，

$f_{t-d_1}$ 為延誤參數， $\varepsilon_{f,1,t}$ 、 $\varepsilon_{f,2,t}$ 為白干擾項(white noise)。



### 3.實證分析

#### 3.1 資料分析

台灣股票市場為一個明顯的非線性時間數列資料，且其資料易受到政府政策的制定以及世界金融環境的景氣等等所影響，各個因子所影響的程度和層面也有所不同，股市時間序列資料經常也會經過某些時點震盪過後而又趨於緩和，也就是資料有進行轉折的事實。大抵來說，情況分為兩種，第一種是由數值較高的時間數列資料群透過轉折區間下降到較低的时间數列資料群，而第二種就是由數值較低的时间數列資料群透過轉折區間上升到較高的時間數列資料群，在股市分析中，第二種為令人感興趣的情形，倘若可以掌握資料何時準備上升即可進行有效的投資而從中獲得報酬。

本研究分析了三檔股票，即台灣水泥公司(Taiwan Cement Corp, TCC)、台灣積體電路製造股份有限公司(Taiwan Semiconductor Manufacturing Company, TSMC)、華南金控股份有限公司(Hua Nan Financial Holdings CO., LTD.)，這三檔股票分別代表了傳統產業、電子業以及金融業，分別利用基因演算法找出轉折區間並且進行分析。

資料來源為台灣證券交易所，資料期間自西元 2010 年 1 月 1 日起至西元 2011 年 12 月 31 日的每周資料，這裡將每周的資料最高值及最低值建構出區間資料進行分析，總共 102 筆資料，如圖 3.1、圖 3.2、圖 3.3 所示，而原始詳細數據見表 3.1，可以清楚看到資料皆有轉折的現象。

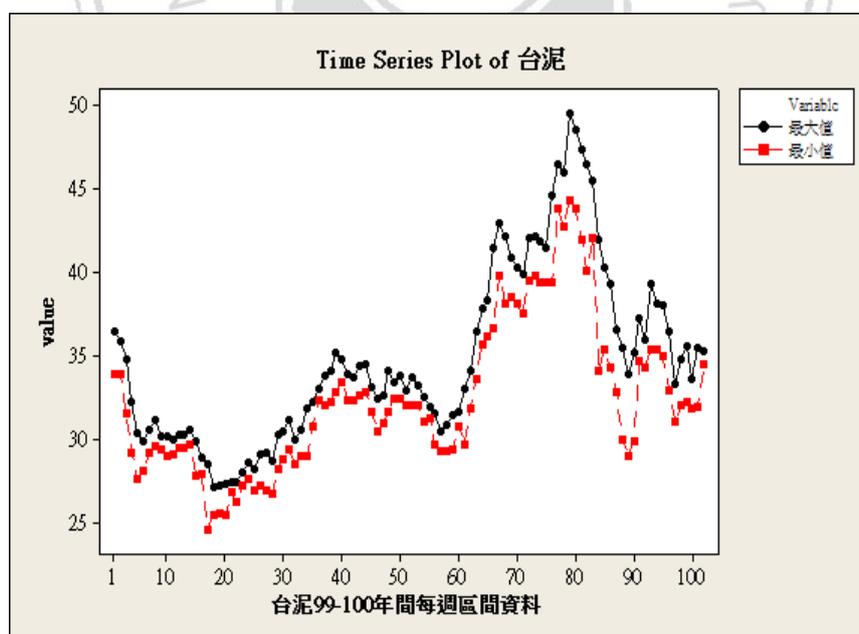


圖 3.1 台泥 2010-2011 年間資料走勢

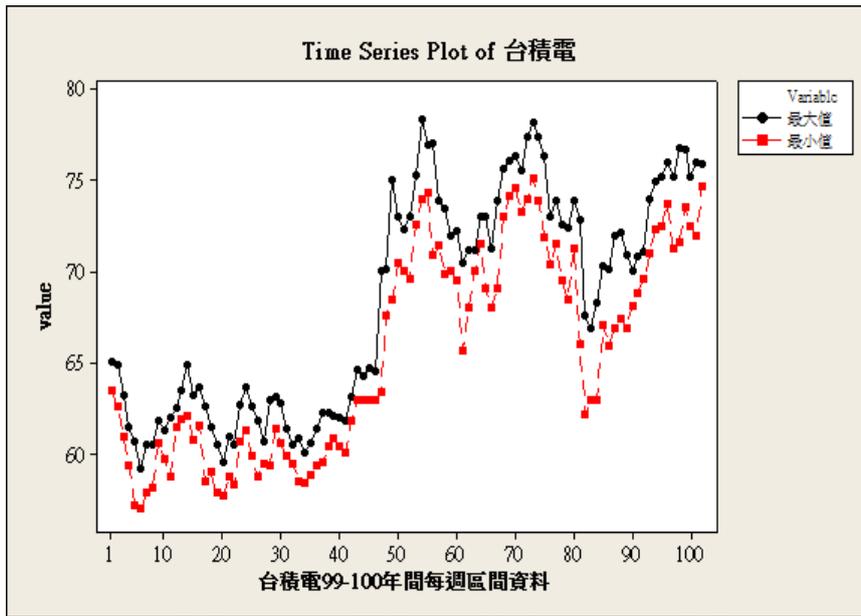


圖 3.2 台積電 2010-2011 年間資料走勢

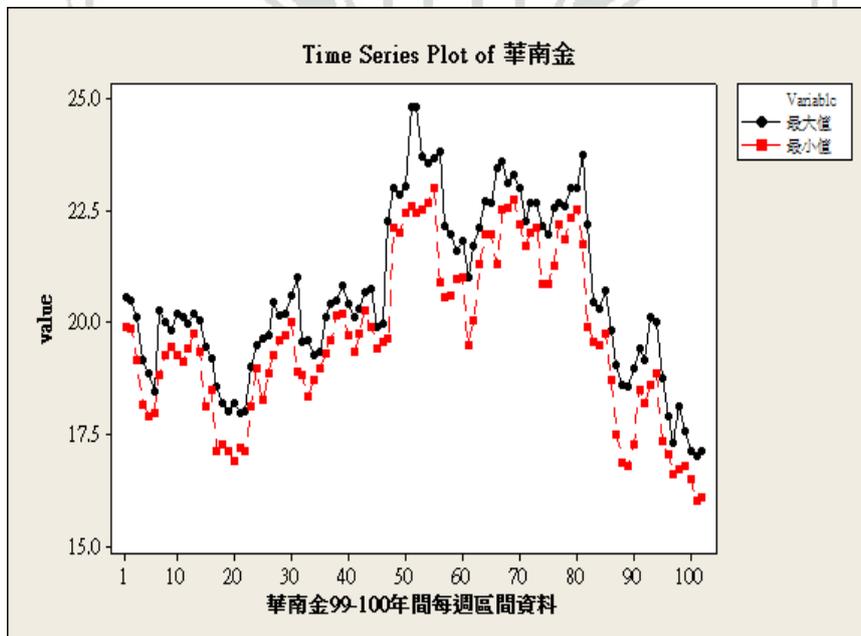


圖 3.3 華南金 2010-2011 年間資料走勢

表 3.1 台泥、台積電、華南金 99-100 年間原始資料

年月日	台泥		台積電		華南金	
	最高價(元)	最低價(元)	最高價(元)	最低價(元)	最高價(元)	最低價(元)
2010/1/8	36.4	33.9	65.1	63.5	20.55	19.9
2010/1/15	35.8	33.9	64.9	62.6	20.5	19.85
2010/1/22	34.75	31.55	63.2	61	20.1	19.15
2010/1/29	32.2	29.2	61.5	59.4	19.15	18.15
2010/2/6	30.35	27.65	60.7	57.2	18.85	17.9
2010/2/10	29.85	28.1	59.2	57	18.45	17.95
2010/2/26	30.55	29.2	60.5	57.9	20.25	18.8
2010/3/5	31.1	29.6	60.5	58.2	20	19.25
2010/3/12	30.2	29.35	61.8	60.6	19.8	19.45
2010/3/19	30.2	29	61.3	59.7	20.2	19.25
2010/3/26	30	29.05	62	58.8	20.1	19.1
2010/4/2	30.25	29.45	62.5	61.5	19.95	19.4
2010/4/9	30.25	29.5	63.5	61.9	20.2	19.75
2010/4/16	30.55	29.65	64.9	62.1	20.05	19.35
2010/4/23	29.85	27.8	63.2	60.8	19.45	18.1
2010/4/30	28.85	27.9	63.7	61.6	19.2	18.5
2010/5/7	28.45	24.6	62.6	58.5	18.55	17.1
2010/5/14	27.1	25.5	61.5	59	18.2	17.25
2010/5/21	27.2	25.6	60.5	57.9	18	17.1
2010/5/28	27.35	25.5	59.6	57.7	18.2	16.9
2010/6/4	27.45	26.85	61	58.8	17.95	17.2
2010/6/11	27.45	26.2	60.5	58.3	18	17.1
2010/6/18	28	27.25	62.7	60.7	19	18.1
2010/6/25	28.6	27.6	63.7	61.3	19.5	18.95
2010/7/2	28.25	26.9	62.6	59.9	19.65	18.25
2010/7/9	29.05	27.25	61.8	58.8	19.7	18.85
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....
2011/12/9	35.6	32.2	76.7	73.5	17.55	16.8
2011/12/16	33.55	31.85	75.2	72.5	17.1	16.5
2011/12/23	35.5	31.95	76	72	17	16
2011/12/30	35.25	34.5	75.9	74.7	17.1	16.1

### 3.2 建構 SETAR 門檻區間

由於區間資料在目前可以使用的門檻模式有限制，所以在此仍然必須將區間資料透過模糊理論跟概念處理，進而建構 SETAR 門檻區間模式。而第一步就是要先將區間資料反模糊化，本研究的處理方式是透過 2.10 式的反模糊化公式進行運算，將每週區間資料中長度影響也加入進來，見表 3.2:

表 3.2 台泥、台積電、華南金反模糊化後的數值

年月日	台泥	台積電	華南金	年月日	台泥	台積電	華南金
2010/1/8	35.49	64.49	20.33	2010/6/4	27.24	60.17	17.70
2010/1/15	35.11	64.02	20.28	2010/6/11	27.01	59.67	17.70
2010/1/22	33.60	62.36	19.78	2010/6/18	27.73	61.94	18.70
2010/1/29	31.13	60.70	18.81	2010/6/25	28.25	62.79	19.31
2010/2/6	29.39	59.37	18.53	2010/7/9	28.41	60.66	19.41
2010/2/10	29.23	58.37	18.28	2010/7/16	28.39	60.24	20.04
2010/2/26	30.07	59.52	19.76	2010/7/23	28.01	61.63	19.96
2010/3/5	30.56	59.63	19.75	2010/7/30	29.52	62.45	20.03
2010/3/12	29.90	61.34	19.68	2010/8/6	29.88	61.96	20.40
2010/3/19	29.77	60.69	19.88	2010/8/13	30.48	60.83	20.29
2010/3/26	29.66	60.79	19.76	2010/8/20	29.47	60.12	19.30
2010/4/2	29.96	62.12	19.76	2010/8/27	30.03	59.99	19.18
2010/4/9	29.98	62.89	20.05	.....	.....	.....	.....
2010/4/16	30.23	63.83	19.81	.....	.....	.....	.....
2010/4/23	29.12	62.29	19.00	.....	.....	.....	.....
2010/4/30	28.51	62.90	18.96	2011/12/2	33.79	74.80	17.63
2010/5/7	27.10	61.04	18.06	2011/12/9	34.37	75.47	17.30
2010/5/14	26.54	60.55	17.88	2011/12/16	32.94	74.16	16.90
2010/5/21	26.64	59.52	17.70	2011/12/23	34.22	74.46	16.67
2010/5/28	26.70	58.88	17.77	2011/12/30	34.98	75.44	16.77

接著定義適應函數以便讓基因演算法進行演算，在此由於設定為投資兩年為限，所以將兩年期最終的價格扣掉購買點的價格除上所有購買點的平均值再取百分比作為適應函數，這麼一來可以先扣除掉比最終價格高的那些資料群，雖然可能在資料當中那些資料群也存在著轉折行為，但是對於兩年期投資而言最終並不會得到正報酬，所以我們將以扣除，也就是可以由此公式代入值後得到負值則表示資料最終往較低的區域移動，這是本研究不感興趣的部份。

#### 定義 3.1 適應函數

$$\frac{F(T) - x(t)}{A(t)} \times 100\% \quad (3.1)$$

$F(T)$  為兩年期最後年終的價格，為一定值;

$x(t)$  為時點  $t$  時的價格;

$A(t)$  為所有購買點的價格算術平均。

另外本研究也加入了一些限制式，譬如從低門檻群到高門檻群的資料點距離必須大於某些值，在此每個時點如要購買則顯示 1，不購買則顯示 0，如此進行基因演算法。在此研究中將經過五萬次演化過程，在此為了增加其生物的多樣性，演算法中的突變機率設定為 5%，每一個世代產生一千個子代，其中有百分之五的突變物種，最後再利用給定隨機切割法進行交配過程，也就是在隨機給定切割點後，往前的字串皆遺傳自父親，之後的字串皆來自母親。可以得到最佳的後代，如表 3.3:

表 3.3 三檔股票的最佳後代

年月日	台泥	台積電	華南金	年月日	台泥	台積電	華南金
適應值	2.78	6.56	-13.06				
2010/1/29	0	1	1	2010/6/25	1	0	1
2010/2/6	1	1	1	2010/7/2	1	0	1
2010/2/10	1	1	0	2010/7/9	0	0	1
2010/2/26	0	1	0	2010/7/16	0	0	0
2010/3/5	0	1	0	2010/7/23	1	0	1
2010/3/12	0	1	0	2010/7/30	1	0	0
2010/3/19	0	1	1	2010/8/6	1	0	1
2010/3/26	0	1	1	2010/8/13	0	0	0
2010/4/2	0	1	0	2010/8/20	1	0	0
2010/4/9	1	1	0	2010/8/27	0	0	1
2010/4/16	1	1	1	2010/9/3	1	1	1
2010/4/23	1	1	1	.....	...	...	...
2010/4/30	1	1	1	.....	...	...	...
2010/5/7	1	1	1	.....	...	...	...
2010/5/14	1	0	1	2011/12/9	1	0	1
2010/6/11	1	1	1	2011/12/16	0	0	0
2010/6/18	1	0	1	2011/12/23	0	1	1

根據上一章制定的轉折決策準則，可以得到以下結論：對於台泥，2010 年 4 月 9 日至 7 月 2 日為購買點，接著為 2010 年 7 月 23 日至 9 月 10 日，還有 2010 年 12 月 3 日至

3月11日;而對於台積電,2010年1月8日至2010年5月7日還有2010年9月24日至2011年1月7日;最後是華南金,2010年4月16日至8月6日以及2011年9月2日至11月25日。特別要注意的是華南金呈現負值,將資料走勢圖拿出來對照會發現華南金的走勢中到兩年期的最後是整個走勢中的最低資料群,也就是表達出資料在這兩年間大致上由高門檻轉化成低門檻,故在這兩年無正報酬,不過基因演算法仍然可取出最大值,讓損失降到最低。而另外兩檔股價可以由資料原始圖形觀察發現圖形在通過這些區間以後開始上升,進入到較高的資料群當中。接著把購買點群按照數值大小排列後取最後三分之一的部份資料對照回去原始資料,再將原始資料的左右端點分別取算術平均,即可得到轉折區間,見下表 3.4:

表 3.4 三檔股票之轉折日期及區間

	台泥	台積電	華南金
轉折區間日期	10/12/3-11/3/11	10/01/08-10/05/07	10/04/16-10/08/06
轉折區間	[ 32.18 , 33.6]	[ 62.2 , 64.22]	[ 19.48 , 20.19]

### 3.3 用外生多變數建構門檻轉換模式

在上一章已經得到了轉折區間,然而許多非線性時間數列相關問題,會受到多個變數的牽動而受影響,如產品銷售量受到廣告成本以及優惠折扣方案等等影響、生出率受到結婚率及政府政策影響等等、本國與外國的匯率受到本國及外國的利率變動以及相對生產力的影響等等。由於本研究中分割的是數值而非時點,所以前後期的關係無法在此得到解釋,也就是透過門檻數值區分,可能會因為某些區域的資料群上下躍動的振幅太大而會導致時間序列的關聯性削減,所以在此考慮加入新的變數到模式當中,也就是利用主要影響因素-外生變數加以解釋。透過建構含外生多變數之門檻自迴歸模式,可以加強時間序列中的關聯並增加其解釋能力,對於函數的配適也會準確許多。其定義如下:

#### 定義 3.2 含外生多變數之門檻自迴歸模型

令  $x_t$  為一個時間數列,  $y_t$  為其外生對應的變數,  $r$  為門檻值,含外生多變數之門檻自迴歸模型為

$$x_{t+1} = \begin{cases} a_1 + b_1 x_t & y_t > r \\ a_2 + b_2 x_t & y_t \leq r \end{cases} \quad (3.2)$$

會影響股價的因子相當多也相當複雜,像是成交量、報酬率、股價淨值比等等,而在本研究當中加入的外生變數為台灣加權指數,在此處會選擇台灣加權指數的原因是因為兩者相關係數極高,用來解釋股價也能得到最小的 *AIC* 值。

由上表 3.4 可以得知每檔股票的轉折區間,故按照先前設定的方式,將轉折區間資料群刪減過後,加入外生變數來建構轉折模式,並且利用 Akaike(1974)提出的 *AIC* 準則

來選取模式當中的參數，其定義如下：

$$AIC(M) = -2\ln[\text{最大概似函數}] + 2M$$

$M$  為模型中參數的個數，

$$\text{當中概似函數的形式爲 } \ln L = -\frac{n}{2} \ln 2\pi\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} S(\phi, \mu, \theta)$$

式中  $S(\phi, \mu, \theta) = \sum_{t=1}^n \varepsilon_t^2(\phi, \mu, \theta)$ ，故在  $\phi, \mu, \theta$  已知之下化簡求得

$$AIC(M) = n \ln \sigma_\varepsilon^2 + 2M$$

在迴歸模型建構過程中，每多增加一個變項可以降低殘差值的和，但也會增加預測的複雜性，同樣地，再時間數列模型建構過程中，模型的複雜度增加也許可以降低對資料的變異性，卻也增加了偏差性。在此將各個候選模式的參數表列如下並且選取最小的  $AIC$  值以建構出最好的門檻自迴歸模式。

台泥:

低資料群模式參數	ARIMA(1,0,0)	ARIMA(2,0,0)
AIC 值	93.44	93.26
高資料群模式參數	ARIMA(0,0,1)	ARIMA(1,0,1)
AIC 值	10.32	76.55

$$x_{t+2} = 5.046 + 0.91x_{t+1} + 0.99x_t + \varepsilon_t \quad y_t < 7953.19$$

$$x_t = 36.17 + \varepsilon_t - 0.94\varepsilon_{t-1} \quad y_t > 8303.24$$

其中  $y_t \in [7953.19, 8303.24]$  為轉折區間， $y_t$  為台灣加權指數 (3.3)

台積電:

高資料群模式參數	ARIMA(1,0,0)	ARIMA(2,0,0)
AIC 值	65.92	59.82

$$x_{t+1} = 2.42 + 1.24x_t - 0.28x_{t-1} + \varepsilon_t, \quad y_t > 8269.04$$

其中  $y_t < 8269.04$  為轉折區間， $y_t$  為台灣加權指數 (3.4)

**華南金:**

低資料群模式參數	ARIMA(1,0,0)	ARIMA(2,0,2)
AIC 值	-15.02	-10.78
高資料群模式參數	ARIMA(1,0,0)	ARIMA(2,0,0)
AIC 值	-5.16	-13.79

$$x_{t+1} = 3.07 + 0.83x_t + \varepsilon_t, \quad y_t < 7355.29$$

$$x_{t+2} = 2.01 + 0.53x_{t+1} + 0.37x_t + \varepsilon_t, \quad y_t > 7668.16$$

其中  $y_t \in [7355.29, 7668.16]$  為轉折區間， $y_t$  為台灣加權指數 (3.5)

在本研究中利用了許多變數當門檻值做測試，最後發生使用台灣加權指數可以得到較佳的 AIC 值，此外也可以透過建構模式了解各檔股票的資料走勢，例如台泥在七千九百點到八千三百點附近震盪，投資人可以趁這段時間進行投資，進一步地透過歷史資料可以發現國際原物料在那些時點附近較為便宜，故影響到了台泥的股價，因此台泥股價準備看漲。而台積電的時間序列資料剛好在本研究的一開始呈現低迷，之後受到國內外經濟成長以及需求增加而大幅向上攀升，到達八千兩百點左右後趨於穩定。最後是華南金，在調查期間之初，金融業的大環境較其餘類股來的穩定，並且也是在七千三百點到七千六百點附近震盪後上漲，而在資料調查的末期受到美國金融業的影響來到低谷。

透過以上的建構方式可以得到轉折區間，投資人可以利用歷史資料尋找應該進場的投資時點或是門檻值，構成有效率的投資行為。此外，此研究建構出的模式有將區間的影響加入考慮，期能利用模糊理論及概念將台灣股價分別進行分析，利用轉折區間有效率地將資料群給分開來，以獲得最適模式。

#### 4. 結論與建議

在非線性時間數列資料分析中，Manski(1990)曾指出，單一數值資料有過度需求及解釋的危險，如果應用模糊的數值資料便能大大降低如此的危險。也就是說，利用點估計中嚴格的假設，可能會造成因果判定偏差、決策失誤以及擴大實際資料與預測結果的差異等等。而在時間數列資料當中，資料的模糊性相關問題可說是傳統時間數列在建構時不易克服的難題。因此必須針對區間資料作分析，利用模糊理論，結合較具穩健性與符合實際狀況的模糊統計分析，並且使用模糊且樸實的資料來建構模式作為預測，將有效的改善傳統時間數列資料分析中所面臨的問題。

本文利用定義模糊關係，透過反模糊化的概念以及利用基因演算法進行建構轉折區間，藉此用來改變傳統使用的門檻值，同時也可以尋找出資料中不穩定以及震盪的資料群，這些資料群在某些時點當中的不確定性會影響配適模型的準確度。接著由於本研究是利用數值點作切割，所以時間數列的相關性減弱，必須加入外生變數將資料型態完整呈現。從實證分析我們可以看見，配適出來的式子較可以表達出原始資料的原始面貌。

在本文中尚有很多待討論解決的問題，亦是未來值得研究的方向，例如反模糊化式子的定義、如何定義轉折區間、基因演算法在模式中的選擇等等。



## 參考文獻

- 吳柏林(1995)，時間數列分析導論，華泰書局，台北。
- 吳柏林(2005)，模糊統計導論方法與應用，五南出版社，台北。
- 吳柏林、阮亨中(2000)，模糊數學與統計應用，俊傑書局，台北。
- 吳柏林、林玉鈞(2002)，模糊時間數列分析與預測-以台灣地區加權股價指數為例，應用數學學報，第 25 卷，第 1 期，頁 67-76。
- 程友梅(1995)，轉移型時間序列的認定。國立政治大學統計系碩士論文。
- 張新發(1996)，遺傳演算法在門檻自迴歸模式(d,r)值估計的應用。國立政治大學統計系碩士論文。
- 楊亦農(2009)，時間序列分析：經濟與財務上之應用，雙葉書廊，台北。
- F.-M. Tseng and G.-H. Tzeng (2002) a fuzzy seasonal ARIMA model for forecasting. *Fuzzy sets and systems*, 126(3), 367-376.
- H. T. Nguyen and B. Wu (2006) *Fundamentals of Statistics with Fuzzy Data*. New York : Springer.
- Hansen, B.E. (1997). *Inference in TAR Models*, Studies in Nonlinear Dynamics and Econometrics, 2, 1-14.
- Hsu, H.L. (2008). *Evaluating forecasting performance for interval data*. Computers and Mathematics with Applications 56, 2155-2163.
- Hsu, H. L. (2011). *Interval Time Series Analysis with Forecasting Efficiency Evaluation*, Doctorial Thesis, Department of Mathematical Science, National Chengchi University, Taipei, Taiwan.
- Kumar, K. and Wu, B. (2001). Detection of change points in time series analysis with fuzzy statistics, *International Journal of Systems Science* 32(9), 1185-1192.
- Ludermir, T. B. (2008). Forecasting models for interval-valued time series. *Neurocomputing* 71, 3228-3238.

- M. Bleaney, N. Gemmell, R. Kneller (1989) Testing the endogenous growth model: public expenditure, taxation, and growth over the long run.
- M. Khashei, S.R. Hejazi and M. Bijari (2008) A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy sets and systems*, 159, 769-786.
- S.K. Chang (2007) On the Testing Hypotheses of Mean and Variance for Interval Data. *Management Science & Statistical Decision*, 4(2), 63-69.
- Tong, H. & Lim, K. S. (1980). *Journal of the Royal Statistical Society, Series B*, "Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion)", 42, 245-292.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press.
- Tseng, F.M., Tseng, G.H., Yu, H.C., and Yuan, B.C. (2001). *Fuzzy ARIMA model for forecasting the foreign exchange market*. *Fuzzy sets and systems* 118(1), 9-19.
- V. Kreinovich, H. T. Nguyen and B. Wu (2007) *On-line algorithms for computing mean and variance of interval data, and their use in intelligent systems*. *Information Sciences*, 177, 3228-3238.
- Wu, B and Hung, S. (1999). *Fuzzy Sets and Systems*. A fuzzy identification procedure for nonlinear time series with example on ARCH and bilinear models. 108, 275-287.
- Wu, B. (2011). *Efficiency Evaluation in Time Management for School Administration with Fuzzy Data*, Technical Report, Department of Mathematical Science, National Chengchi University, Taipei, Taiwan.
- Zhou H. D. (2005). *Nonlinearity or structural break – data mining in evolving financial data sets from a Bayesian model combination perspective*. *Proceedings of the 38<sup>th</sup> Hawaii International Conference on System Sciences*, Hawaii, U.S.A.