

國立政治大學應用數學系

碩士學位論文

區間 SETAR 模式的建構分析與預測

Interval SETAR Modelling and
Forecasting Evaluation.

碩士班學生：廖育琳 撰

指導教授：吳柏林 博士

中 華 民 國 一 百 年 六 月

摘要

雖然傳統線性時間數列在預測上已被廣泛的使用，但是在一般的時間數列中或多或少都會有結構改變(structural changes)的現象，我們往往很難找到一簡單的線性模式來詮釋資料中普遍存在的非線性(nonlinearity)結構，同時隨著模糊理論的興起與區間軟計算(soft computing)的發展，區間預測(interval forecasting)已成為未來研究的重點。本文應用模糊分類法(fuzzy classification)，找出結構改變的位置，藉此發展出非線性的區間門檻自迴歸模式(interval SETAR model)，再以「來臺觀光客人數」與「新臺幣兌美元匯率」作為實例，建構兩種區間門檻自迴歸模式與區間 ARIMA 模式並比較之，結果顯示兩種非線性的預測效果都比線性模式好。

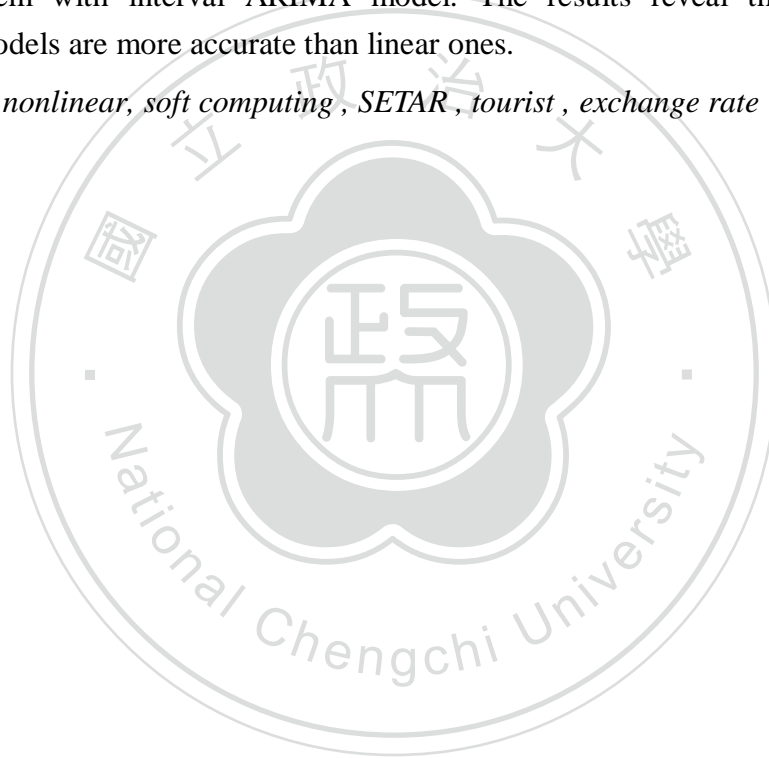
關鍵字：非線性、區間軟計算、門檻自迴歸、觀光客、匯率



Abstract

Although the traditional linear time series has been widely used in forecasting, there are, more or less, some structural changes in general time series. It is difficult to find a simple linear model to interpret the nonlinear structure which exists in many cases. In the mean time, with the rise of fuzzy theory and the development of interval soft computing, interval forecasting has been the research topic in the future. Therefore, this study is going to apply the fuzzy classification to point out the position of structural changes, so as to develop nonlinear interval SETAR model. Furthermore, "the number of tourists to Taiwan" and "the exchange rate of NTD to USD" will be taken as examples in order to construct two kinds of interval SETAR models and compare them with interval ARIMA model. The results reveal that these two nonlinear models are more accurate than linear ones.

Keywords : *nonlinear, soft computing, SETAR, tourist, exchange rate*



目錄

第一章 前言.....	1
第二章 研究方法.....	4
2.1 門檻自迴歸模式.....	4
2.2 模糊隸屬度與模糊熵分類法.....	8
2.3 區間 ARIMA 模式、區間門檻自迴歸模式.....	10
2.4 預測效率評估.....	12
第三章 實證分析-來臺觀光客人數.....	16
3.1 資料來源.....	16
3.2 以區間型 ARIMA 模式建構.....	18
3.3 平均累加模糊熵分類.....	19
3.4 以區間型門檻自迴歸模式建構.....	22
3.5 預測結果比較與分析.....	28
第四章 實證分析-新臺幣兌美元匯率.....	30
4.1 資料來源.....	30
4.2 以區間型 ARIMA 模式建構.....	32
4.3 平均累加模糊熵分類.....	33
4.4 以區間型門檻自迴歸模式建構.....	36
4.5 預測結果比較與分析.....	42
第五章 結論.....	44
參考文獻.....	46

第一章 前言

傳統線性時間數列在預測上已被廣泛的使用，但是其限制條件如：穩定性(stationarity)、線性(linearity)的隨機過程，卻常常困擾著實證的學者，在一般的時間數列中或多或少都會有結構改變(structural changes)的現象，我們往往很難找到一合適的線性模式來配適，由於時間數列的資料普遍存在有非線性(nonlinearity)的現象，使得非線性的領域在近幾年受到極大的重視，舉凡社會(Abraham, 1987, Macmillan, 1995, Roger 與 Wilson, 1996)、農業(Machado, 1995)、通訊(Szmania, 1989)，經濟(Perron 與 Vogelsang, 1992, Dufour 與 Ghysels, 1996, Stock 與 Watson, 1996)(程友梅, 1995)，並同時發展出了許多非線性的模式，比如說門檻自迴歸模式(threshold autoregressive models; Tong, 1983)、雙線性模式(Bilinear models), (Subba-Rao 與 Gabr, 1984)、指數型自迴歸模式(Exponential autoregressive models; Haggan, 1981)等...。這些模式當中，門檻自迴歸模式因具有許多線性 ARIMA 模式所不能描述的特性而受到重視，從早期 Tong 與 Lim (1980)利用門檻自迴歸模式分析加拿大山貓成長趨勢與太陽黑子出現情形，以及 Byers 與 Peel (1995)以門檻自迴歸模式預測六個國家的工業生產指數，和傳統線性分析的方法比較之下，發現利用門檻自迴歸模式來分析不僅能有較佳的配適度也更能呈現出資料的發展趨勢。

隨著近年來全球經濟蓬勃發展，休閒旅遊產業越來越受到重視，觀光業已是目前世界各國公認最重要的無煙囪工業之一，除了可以提升國際知名度、展現國家文化內涵外，在創造就業機會及活絡經濟發展上亦具有明顯的效益。根據世界觀光組織(World Tourism Organization, WTO)的分析報告指出，「觀光」已成為許多國家賺取外匯的首要來源，在全球各國的外匯收入中約有超過百分之八來自觀光收益，高居所有國際貿易種類之冠，而 WTO 更預估至 2020 年全球觀光人數將突破五十六億二百萬人次，全球觀光收益也將達到二兆美元。因此近年來我國陸續推出「觀光客倍增計畫」、「旅行臺灣年」、「觀光拔尖領航方案」、「旅行臺灣・感動 100」等各種規劃(中華民國觀光年報, 2010)，希望藉由增加觀光客來增加外匯

並振興經濟，然而不論是本地觀光業者或政府對於觀光環境之投資與政策制定，皆有賴於相關單位對於觀光客人數的掌握，若能準確的評估將可事先對各種軟硬體環境做出完善的籌劃，不僅可避免各種資源的不敷使用或浪費閒置的情形，更能促進我國觀光產業的正向發展，因此對於來臺觀光人口數作一準確的預測實屬必要。

Michael and Charles (1997)指出「匯率」為衡量兩國產品與勞務相對價格的重要指標，說明了兩國之間的產品競爭力。對於全球任何開放經濟體系而言，國內經濟活動與國外經濟情勢的關連性十分密切，而各國的匯率在這個國內與國外經濟活動的交互作用中，扮演著極為重要的角色，而我國自 1979 年外匯市場開放運作以來，由於對美貿易順差持續增加，引起美方的強烈關注，再加上國際政治因素及貿易金融自由化的考量，促使我國的中央銀行採取「管理浮動匯率制度」(managed floating exchange rate)政策，以減少政府對於外匯市場的干預，將貨幣視作一種商品，讓自由的市場的供需機制來決定其價值，自此之後，為使臺幣符合市場真實價格，新臺幣對美元的匯率已從 1985 年的 40.43 元經過多次波動，漸漸攀升到目前約 29 元的水準(張新發, 1996)，而未來的走勢將如何的做變化，相信絕對是經濟、金融、學術各界所關注的焦點。

在過去傳統的時間數列預測中，使用的資料以及作出的預測往往是以單一的數值為主，然而在人類的生活中有許多的事物皆存在一不明確的範圍，如氣溫的高低是以早晨、正午或半夜的溫度為準？或者新臺幣對美元的匯率是以開盤、收盤還是平均值為主呢？由於僅用單一數值形式收集來的資料，其建立的模式並不足以描述每日或每月的發展趨勢，因此隨著軟計算的興起，越來越多的學者開始重視區間預測的方法，Chatfield (1993)是區間計算預測的先驅之一，Tseng, Tzeng, Yu, and Yuan (2001)將模糊 ARIMA 模式應用於臺幣與美元間的匯率，Lin (2007)研究隸屬度函數及區間長度改良對模糊時間序列預測之影響，Chen (2008)發表區間型模糊數的迴歸分析，Ludermir (2008)將區間預測建立在時間數列模型上，Hsu (2008)探討區間時間序列預測及分析，Liao (2009)研究多變量模糊時間數列分析與

轉折區間的檢測，Ro (2010)提出區間模糊相關係數，隨著近年來學著們的努力，區間預測的領域已漸漸受到了重視。

對於「觀光客」和「匯率」這兩種資料而言，容易受到國際局勢變化、或政府政策推行因素的影響，可能具有結構改變的特性，這是一般傳統線性 ARIMA 模式難以描述及處理的地方(吳柏林, 1991)，因此我們希望使用非線性模式來配適「來臺觀光客人數」與「新臺幣兌美元匯率」，不過自 1983 年 Tong 提出門檻自迴歸模式以來，文獻中大多將其用於建構單點的模式，關於區間資料的研究反而較少，然而在現實生活中有許多的資料都是以區間來呈現及預測的，就比方說因為一天中的氣溫有高有低，氣象預報會說明天最高溫、最低溫幾度之間的範圍，因此由於許多的資料本身即具有不確定性及模糊性，若我們利用假性的「精確值」來做分析，就有可能造誤導預測模式的建構((吳柏林, 2005)，對於「來臺觀光客人數」或「新臺幣兌美元匯率」亦為如此，一年當中每個月份的旅客人次不可能完全相同，每年都有所謂的淡、旺季之分；一天之內的匯率也有高有低，而非單一的匯價，所以若要預測的話也不該武斷的只採取單點值的資料，而應是若干個不同範圍的區間資料才是。因此本文的目標即是以模糊理論為基礎，突破傳統一般點對點的時間數列模式，建構出一套以區間資料為主的門檻自迴歸模式，希望不僅能兼顧資料的模糊性、並同時解決可能隱藏在資料中的結構性改變的問題。

本文主要分為四大部份。第一章說明研究動機、探討相關文獻並提出研究方向；第二章引進區間運算及模糊理論的觀念，並以此為基礎建構區間型門檻自迴歸模式，並定義區間預測之效度評估標準；第三章和第四章以「來臺觀光客人數」與「新臺幣兌美元匯率」為原始區間資料，分別以區間 ARIMA 與兩種區間 TAR 模式建構並比較其預測效果，最後第五章為結論。

第二章 研究方法

2.1 門檻自迴歸模式

由於傳統的單根檢定無法探討資料中可能存在的不對稱非線性關係，因此就有門檻自迴歸模式(threshold autoregressive models, TAR)的出現，此模式最早是由 Tong (1983)所提出，它的基本原理就是把時間數列中的各觀察值，依其走勢變化的情形，找出一個資料結構轉變的地方，稱之為轉捩點(change point)或門檻值(threshold)，將數列資料清楚的劃分為數個狀態(regime)，各自形成一個系統，服從不同的線性自迴歸(autoregressive, AR)模式。門檻模式本身有幾項明顯的特徵，作為一般判斷識別的標準，如有周期循環性、上下起伏的震動、或突然向上、向下的現象，這些特徵是傳統 ARIMA 模式難以描述的地方(吳柏林, 1995)，因此這種模式常常用來解決時間數列中資料不對稱的狀況，在我們的生活中就有許多結構改變的現象是依據某些門檻所產生的，舉例來說，股價持續上漲，會因投資人獲利回吐，而導致股價下跌；數量過於龐大的生物族群，會因環境資源的侷限，而導致物種逐漸的減少，諸如此類...。因此近年來非線性方法逐漸受到了重視，TAR 模式已被廣泛的利用到財務金融及社會科學等非線性的領域上，如經濟成長率、加權股票指數、匯率、出生率、死亡率、醫療健保支出、毒品查獲量等...。以下我們將詳細介紹門檻自迴歸模式及其建構方法。

1. 模式基本介紹

基本上門檻自迴歸模式是由幾個線性自迴歸模式將資料分成數個不同的狀態，各個狀態之間再利用門檻值(threshold parameter)區分為兩個以上的狀態，而一般的門檻值主要分為兩類：一種是以「時間」作為狀態變化的分水嶺(piecewise in time)，在某個時點前後構成兩種不同的自迴歸模式；或者是利用模式中「變數」作為結構改變的依據(piecewise in variable)，於此門檻值上下各自形成不同的系統，此類亦可稱為 SETAR 模式(Self-Excited TAR)，以下即為兩類門檻自迴歸模式的表

示方法，其中(2.1)是以時間點分段的 TAR 模式，其表示法如下：

$$Y_t = \begin{cases} \Phi_{1,0} + \sum_{i=1}^{K_1} \Phi_{1,i} Y_{t-i} + a_{1,t} & \text{if } t \leq s \\ \Phi_{2,0} + \sum_{i=1}^{K_2} \Phi_{2,i} Y_{t-i} + a_{2,t} & \text{if } t > s \end{cases} \quad (2.1)$$

其中 K_1, K_2 ：為兩個自迴歸模式的階次(order)

$\Phi_{1,i}, \Phi_{2,i}$ ：為自迴歸係數

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

t 為期數或時間

s 為某期或某一時點

另外 2.2 為 Tong (1983)所提出的兩階段之門檻自迴歸模式 SETAR(2; K_1, K_2)，如下列所示：

$$Y_t = \begin{cases} \Phi_{1,0} + \sum_{i=1}^{K_1} \Phi_{1,i} Y_{t-i} + a_{1,t} & \text{if } Y_{t-d} \leq r \\ \Phi_{2,0} + \sum_{i=1}^{K_2} \Phi_{2,i} Y_{t-i} + a_{2,t} & \text{if } Y_{t-d} > r \end{cases} \quad (2.2)$$

其中 K_1, K_2 ：為兩個自迴歸模式的階次(order)

$\Phi_{1,i}, \Phi_{2,i}$ ：為自迴歸係數

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

d ：為控制門檻發生處之變數，稱之為延誤參數

r ：為一常數，稱之為門檻值

所謂的門檻自迴歸模式即是利用門檻值的特性來作切換(switch)的動作，當模式處於某個特殊的時點前後，或者是當其中的延誤參數(delay parameter)高於或低於某一門檻值時，原本的 AR 模式就會被切換到另一個不同狀態的 AR 模式，數個不同的狀態之間便可透過門檻值彼此切換，亦即所謂的「片斷線性」(piecewise linear)的函數(楊奕農, 2009)。

2. 模式建構方法

雖然一般以時間分段的 TAR 模式而言，我們只要找到結構改變的時點再分別配適線性迴歸模式即可，但是對於以「變數」分段的 SETAR 而言，將一時間數列分成兩段不同的系統卻不是很容易，如何找出該模式中的延誤參數(d)、以及門檻值(r)是建構模式的一大難題，因此 Tong (1983)提出了一個利用 AIC(Akaike Information Criteria)準則(Akaike, 1974)作為選擇參數依據的建構方法。接下來我們先介紹何謂 AIC 準則，再以 SETAR(2 ; k_1, k_2)為例，說明詳細的建構步驟。

定義 2.1 AIC 準則(Akaike Information Criteria; Akaike, 1974)

設 y_t 為一樣本數為 n 之時間數列，若以 p 階自迴歸模式配適， K 為待估參數總數、 $\ln(RSS(p)/n)$ 為殘差均方和取對數，則其 AIC 值定義如下：

$$AIC(p) = \ln(RSS(p)/n) + 2K/n$$

SETAR(2 ; k_1, k_2) 模式建構步驟：

第一步：首先令 L 為模式的最高階次(Tong 建議 $L=[n^\alpha]$ ， $\alpha < 0.5$ ， n 為樣本數)，

事先給定某組 d 和 r 值，再將觀察值由小至大依序排列，接著根據往前 d 期的觀察值 y_{t-d} 高於或低於門檻值 r ，將所有的觀察值分成兩個部份，分別以最小平方方法(The ordinary least squares, OLS)作參數的估計，配適恰當的線性自迴歸模式，並以

$$AIC(k_i) = \ln(RSS(k_i/n_i)) + 2(k_i)/n_i \quad i=1, 2$$

(其中 n_i 為該段的樣本數、 $RSS(k_i)$ 為殘差平方和、 k_i 為待估參數總數)

作為選擇階次的依據，使得

$$AIC(\hat{k}_i) = \min_{0 \leq k_i \leq L} \{AIC(k_i)\} \quad i=1, 2$$

最後再計算其和

$$AIC(d, r) = AIC(\hat{k}_1) + AIC(\hat{k}_2)$$

第二步：考慮門檻值可能發生的位置 $\{r_1, r_2, r_3 \dots r_s\}$ ，並固定延誤參數 d ，重複第一個步驟得到

$$AIC(d, \hat{r}) = \min_{r \in \{r_1, r_2, r_3 \dots r_s\}} \{AIC(d, r)\}$$

其中 \hat{r} 即為 r 的估計值。

第三步：由於不同的 d 值會造成樣本數的變化，為了不使樣本數的多寡影響到 AIC 值的比較，因此我們利用下式來做調整

$$AIC(d) = AIC(d, \hat{r}) / (n - d)$$

以此作為估計 d 的依據，以上一步中的門檻值 \hat{r} 作為基礎，在 $d \leq L$ 的範圍內搜尋，重複第一與第二步驟一直到找到使 $AIC(d)$ 最小的 \hat{d} 值為止。

完成了上述三個步驟後，可以求得 $\{\hat{k}_1, \hat{k}_2, \hat{r}, \hat{d}\}$ 四個值，即分別為兩個自迴歸模式的階次 k_1, k_2 、門檻值 r 以及延誤參數 d ，這時所得到的模式，即是所謂的門檻自迴歸模式，兩段 Y_t 皆為線性迴歸模式，其參數隨過去的值變動。

然而由前述門檻自迴歸的演算法可得知，若自迴歸模式的個數為 M ，最大階次為 L ，門檻值的個數為 $S(S \geq M-1)$ ，延誤參數的個數為 T ，則演算過程所需的計算的模式個數為：

$$L^M \times \binom{S}{M-1} \times T = L^M \times \frac{S(S-1)(S-2) \dots (S-M+2)}{1 \cdot 2 \cdot 3 \dots (M-1)} \times T$$

故當門檻值的個數 S 很大且自迴歸模式個數 M 在某個程度時，所需計算模式個數將非常大(阮正治, 1996)，這對是門檻自迴歸模式建構時的一個很大的困難，Tsay (1989)亦表示：(1)門檻變數的認定與相對門檻值估計是很困難的。(2)此類模式目前並沒有一既定的建構程序。因此若我們能修正模式建構的程序，在開始計算前避免主觀的選取門檻值，利用有系統的方式檢定模式可能的結構改變之處，在第一步驟即時給定可信度較高的門檻候選值，除了較為客觀外，亦可避免掉一些盲目的計算，相信必能提升演算過程中的效率。目前學界已提出許多關於結構轉變的檢定，例如 Lagrangian multiplier (LM) test; Guegan 與 Pham, 1992、Likelihood

ratio-based test; Chan Tong, 1986、Bispectrum test; Hinich, 1982，然而這些檢定方法是針對特定的非線性時間數列所設計的，僅對特定的模式具有相當的檢定力，若是要檢定其他的模式，則檢定力將會降低(Gooijerm 與 Dumar, 1992)，因此爲了提高此模型的適用性，我們在下一節中將會引進 Zadeh (1965)所提出的模糊集合理論(Fuzzy Set Theory)，利用平均累加模糊熵 (吳柏林, 1999)有系統的來蒐尋模式中的可能的結構轉變點。

2.2 模糊隸屬度與模糊熵分類法

人類的思維主要來自對於自然與社會現象的認知意識，而人類的知識語言也會因本身的主觀意識時間環境和研判事情的角度不同而具有模糊性，因此在許多社會科學的範疇中，我們無法將某些現象簡單分成非此即彼的關係，比如說若我們將一群人的心情簡單的分成「快樂」與「不快樂」與兩類，這樣的劃分法很明顯有不合理之處，因爲人的心情並非是二元的現象，而是有各種不同情緒程度連續性之特性，因此爲了解釋這些現實生活中具有模糊性的事物，Zadeh (1965)發表了模糊集合理論，解決了許多傳統二元邏輯所無法描述的現象，重新以隸屬度(membership grade)來呈現元素與集合之間的關係，例如，人們身高 200 公分絕對屬於高，則其隸屬度當屬於 1，而身高 180 公分或 178 公分的隸屬度則約等於 0.8，這表示身高 180 公分或 178 公分屬於高的程度有 0.8 之多，因此隸屬度表示論域中某一元素屬於集合的程度，一般通常會以介於 0 到 1 之間函數來描述這樣的關係，即爲所謂的隸屬度函數，當我們要分析時間數列中是否有發生結構轉變時，便是運用模糊隸屬度與模糊熵等觀念作分類，進而找出數列中可能的轉折之處，其定義如下：

定義 2.2 模糊隸屬度(Fuzzy Membership Grade; Wu, 1999)

令一時間數列 $Y_t, t = 1, 2, \dots, n$ ， C_1 與 C_2 爲其兩個群落中心，令 μ_{it} ， $i=1, 2$ 表示時間數列 Y_t 中的元素 y_t 對 C_1 、 C_2 的隸屬度，則定義隸屬度爲

$$\mu_{it} = 1 - \frac{|y_t - C_i|}{\sum_{i=1}^2 |y_t - C_i|}$$

定義 2.3 模糊熵(Fuzzy Entropy; Wu, 1999)

令一時間數列 $Y_t, t = 1, 2, \dots, n$, μ_{it} 表 y_t 對群落中心 $C_i, (i=1, 2, \dots, k)$ 的隸屬度，則 y_t 的模糊熵定義為

$$\delta(y_t) = -\left(\frac{1}{k}\right) \sum_{i=1}^k [\mu_{it} \ln(\mu_{it}) + (1 - \mu_{it}) \ln(1 - \mu_{it})]$$

定義 2.4 平均累加模糊熵(Mean Cumulated Fuzzy Entropy; Wu, 1999)

令一時間數列 $Y_t, t = 1, 2, \dots, n$, $\delta(y_t)$ 為其模糊熵，則定義平均累加模糊熵為

$$MS\delta(y_t) = \frac{1}{t} \sum_{i=1}^t \delta(y_i)$$

所謂熵原是熱力學中蘊含能量的標準單位，在模糊理論中定義模糊熵來測量模糊集合的不確定性，以資訊理論(Information Theory)的觀點來說，它代表來模糊集合的平均內部訊息量，此訊息量是作為對以模糊集合描述之對象進行分類時的判斷標準，運用模糊熵可有效判斷一時間數列是否有結構性改變的發生，利用 t 個時間的平均累加模糊熵來觀測模糊熵的訊息變化情形，並以此來作為模型轉折分類的標準，意即若是當 $MS\delta(y_t)$ 發生顯著的群據變化時，表示模型正在發生結構性改變，我們便可以透過觀察 $MS\delta(y_t)$ 的走勢來了解改變發生之處，進而選定合適的門檻候選值。以下便是吳柏林與張建瑋 (1999) 根據一般的時間數列實證分析經驗所提出的模糊熵分類法。

平均累加模糊熵分類法：

第一步：先利用 k-means method(Sharma, 1996) 找出時間序列 $\{Y_t\}$ 的 2 個群落中心，並決定 $\{Y_t\}$ 對 2 個群落中心的隸屬度 $\mu_{it}, i = 1, 2$ ，其中

$$\mu_{it} = 1 - \frac{|y_t - C_i|}{\sum_{i=1}^2 |y_t - C_i|}$$

第二步：計算出對應的模糊熵 $\delta(y_t)$ 、平均累加模糊熵 $MS\delta(y_t) = \frac{1}{t} \sum_{i=1}^t \delta(y_i)$

及此數列的中位數 $Median(MS\delta(y_t))$

第三步：取適當的一門檻值 λ ，將 $\{Y_t\}$ 對應的平均累加模糊熵 $MS \delta(y_t)$ 數列進行分類。若 $MS \delta(y_t)$ 落在區間 $[0, Median(MS \delta(y_t)) - \lambda]$ ，則以 1 表示第一組；若 $MS \delta(y_t)$ 落在區間 $[Median(MS \delta(y_t)) - \lambda, Median(MS \delta(y_t)) + \lambda]$ ，則以 2 表示第二組；若 $MS \delta(y_t)$ 落在區間 $[Median(MS \delta(y_t)) + \lambda, 1]$ ，則以 3 表示第三組。

第四步：若分類結果不一致，則對此分類結果作調整；若分類皆相同，則跳過至第五步。

第五步：選取適當的判定水準 α ，若連串的樣本數大於 $[\alpha n]$ ，則此連串樣本屬於同一組，當分類的組數超過一組時，表示此數列發生結構性改變。

2.3 區間型 ARIMA 模式 & 區間型門檻自迴歸模式

傳統時間數列為 $\{Y_t, t = 1, 2, \dots, n\}$ ，其預測方式為 $\hat{Y}_t = E(Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1)$ ，為一般點對點之間的預測，使用上只能侷限於單時點的資料，若是樣本為區間的形式，則無法以傳統的方法來進行預測，使的一般時間的數列在應用上受到了限制，為了增加其廣泛性及適用性，以下先引進區間模糊數的概念，介紹區間時間數列，最後再分別以 ARIMA 以及 TAR 法針對左右端點、中心點半徑兩種區間表式法來定義區間型式的時間數列。

定義 2.5 區間模糊數(Fuzzy Number of Interval; Wu, 2006)

令 $X = [a, b] = (c; r)$ ，其中 $c = \frac{(a+b)}{2}$ 為區間 X 之中心， $r = \frac{(b-a)}{2}$ 為區間長度的半徑，

若 a, b, c, r 皆為隨機變數，則稱 X 為區間模糊數。

定義 2.6 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。

定義 2.7 左右端點法之 ARIMA 模式(Tsu, 2007)

令 $\{Y_t=[a_t, b_t], t=1, 2, \dots, n\}$ 為一個區間時間數列，

$$a_t = \theta + \phi_1 a_{t-1} + \dots + \phi_{p_c} a_{t-p_c} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_{q_c} \varepsilon_{t-q_c}$$

$$b_t = \alpha + \beta_1 b_{t-1} + \dots + \beta_{p_c} b_{t-p_c} + \varepsilon_t - \eta_1 \varepsilon_{t-1} - \dots - \eta_{q_c} \varepsilon_{t-q_c} \quad \text{其中 } \varepsilon_t \text{ 為白干擾項，}$$

則 $\hat{Y}_t = E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_1] = [\hat{a}_t, \hat{b}_t]$ ，其中

$$\hat{a}_t = E[a_t | a_{t-1}, a_{t-2}, \dots, a_1] = \theta + \phi_1 a_{t-1} + \dots + \phi_{p_c} a_{t-p_c}$$

$$\hat{b}_t = E[b_t | b_{t-1}, b_{t-2}, \dots, b_1] = \alpha + \beta_1 b_{t-1} + \dots + \beta_{p_c} b_{t-p_c}$$

定義 2.8 中心點及半徑法之 ARIMA 模式(Tsu, 2007)

令 $\{Y_t=(c_t; r_t), t=1, 2, \dots, n\}$ 為一個區間時間數列，

$$c_t = \theta + \phi_1 c_{t-1} + \dots + \phi_{p_c} c_{t-p_c} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_{q_c} \varepsilon_{t-q_c}$$

$$r_t = \alpha + \beta_1 r_{t-1} + \dots + \beta_{p_c} r_{t-p_c} + \varepsilon_t - \eta_1 \varepsilon_{t-1} - \dots - \eta_{q_c} \varepsilon_{t-q_c} \quad \text{其中 } \varepsilon_t \text{ 為白干擾項，}$$

則 $\hat{Y}_t = E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_1] = [\hat{c}_t - \hat{r}_t, \hat{c}_t + \hat{r}_t]$ ，其中

$$\hat{c}_t = E[c_t | c_{t-1}, c_{t-2}, \dots, c_1] = \theta + \phi_1 c_{t-1} + \dots + \phi_{p_c} c_{t-p_c}$$

$$\hat{r}_t = E[r_t | r_{t-1}, r_{t-2}, \dots, r_1] = \alpha + \beta_1 r_{t-1} + \dots + \beta_{p_c} r_{t-p_c}$$

定義 2.9 左右端點法之區間門檻自迴歸模式

令 $\{Y_t=[a_t, b_t], t=1, 2, \dots, n\}$ 為一個區間時間數列，其中

$$a_t = \begin{cases} \Phi_{a,1,0} + \sum_{i=1}^{K1} \Phi_{a,1,i} a_{t-i} + \varepsilon_{a,1,t} & \text{if } a_{t-d1} \leq S_1 \\ \Phi_{a,2,0} + \sum_{i=1}^{K2} \Phi_{a,2,i} a_{t-i} + \varepsilon_{a,2,t} & \text{if } a_{t-d1} > S_1 \end{cases}$$

$$b_t = \begin{cases} \Phi_{b,1,0} + \sum_{i=1}^{K1} \Phi_{b,1,i} b_{t-i} + \varepsilon_{b,1,t} & \text{if } b_{t-d2} \leq S_2 \\ \Phi_{b,2,0} + \sum_{i=1}^{K2} \Phi_{b,2,i} b_{t-i} + \varepsilon_{b,2,t} & \text{if } b_{t-d2} > S_2 \end{cases}$$

Φ 為自迴歸係數， S 為門檻值， a_{t-d1}, b_{t-d2} 為延誤參數， $\varepsilon_{a,1,t}, \varepsilon_{b,1,t}$ 為白干擾項

則 $\hat{Y}_t = E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_1] = [\hat{a}_t, \hat{b}_t]$ ，其中

$$\hat{a}_t = E[a_t | a_{t-1}, a_{t-2}, \dots, a_1] = \begin{cases} \Phi_{a,1,0} + \sum_{i=1}^{K1} \Phi_{a,1,i} a_{t-i} & \text{if } a_{t-d1} \leq S_1 \\ \Phi_{a,2,0} + \sum_{i=1}^{K2} \Phi_{a,2,i} a_{t-i} & \text{if } a_{t-d1} > S_1 \end{cases}$$

$$\hat{b}_t = E[b_t | b_{t-1}, b_{t-2}, \dots, b_1] = \begin{cases} \Phi_{b,1,0} + \sum_{i=1}^{K1} \Phi_{b,1,i} b_{t-i} & \text{if } b_{t-d2} \leq S_2 \\ \Phi_{b,2,0} + \sum_{i=1}^{K2} \Phi_{b,2,i} b_{t-i} & \text{if } b_{t-d2} > S_2 \end{cases}$$

定義 2.10 中心點及半徑法之區間門檻自迴歸模式

令 $\{Y_t = (c_t; r_t), t=1, 2, \dots, n\}$ 為一個區間時間數列, 其中

$$c_t = \begin{cases} \Phi_{c,1,0} + \sum_{i=1}^{K1} \Phi_{c,1,i} c_{t-i} + \varepsilon_{c,1,t} & \text{if } c_{t-d1} \leq S_1 \\ \Phi_{c,2,0} + \sum_{i=1}^{K2} \Phi_{c,2,i} c_{t-i} + \varepsilon_{c,2,t} & \text{if } c_{t-d1} > S_1 \end{cases}$$

$$r_t = \begin{cases} \Phi_{r,1,0} + \sum_{i=1}^{K1} \Phi_{r,1,i} r_{t-i} + \varepsilon_{r,1,t} & \text{if } c_{t-d2} \leq S_2 \\ \Phi_{r,2,0} + \sum_{i=1}^{K2} \Phi_{r,2,i} r_{t-i} + \varepsilon_{r,2,t} & \text{if } c_{t-d2} > S_2 \end{cases}$$

Φ 為自迴歸係數, c_{t-d1}, r_{t-d2} 為延誤參數, S 為門檻值, $\varepsilon_{c,1,t}, \varepsilon_{r,1,t}$ 為白干擾項

則 $\widehat{Y}_t = E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_1] = [\widehat{c}_t - \widehat{r}_t, \widehat{c}_t + \widehat{r}_t]$

$$\widehat{c}_t = E[c_t | c_{t-1}, c_{t-2} \dots c_1] = \begin{cases} \Phi_{c,1,0} + \sum_{i=1}^{K1} \Phi_{c,1,i} c_{t-i} & \text{if } c_{t-d1} \leq S_1 \\ \Phi_{c,2,0} + \sum_{i=1}^{K2} \Phi_{c,2,i} c_{t-i} & \text{if } c_{t-d1} > S_1 \end{cases}$$

$$\widehat{r}_t = E[r_t | r_{t-1}, r_{t-2} \dots r_1] = \begin{cases} \Phi_{r,1,0} + \sum_{i=1}^{K1} \Phi_{r,1,i} r_{t-i} & \text{if } c_{t-d2} \leq S_2 \\ \Phi_{r,2,0} + \sum_{i=1}^{K2} \Phi_{r,2,i} r_{t-i} & \text{if } c_{t-d2} > S_2 \end{cases}$$

2.4 預測效率評估

預測結果的好壞, 是預測者最關心的問題, 在一般的點預測中有許多的評估標準、如平均百分誤差(MPE)、平均誤差平方和(MSE)、平均絕對誤差(MAE)、平均絕對百分誤差(MAPE)等等...。然而對於區間預測而言, 亦需要一套可供比較的評估準則, 然而由於在文獻中各種區間運算並沒有統一的規則, 因此對於區間距離的測量尚無完備之定義(吳柏林, 2010), 所以目前針對區間預測的效度評估還未有一既定的法則, 所以本節將以模糊理論為基礎, 將區間資料反模糊化(defuzzification)轉換為合適的實數值, 計算兩區間之間的距離, 估計可能的誤差範圍, 藉此分析最後的預測結果。

定義 2.11 區間反模糊化值(Wu, 2010)

若 $A = [a, b]$ 為一區間模糊數, $c = \frac{a+b}{2}$ 為區間中心點、 $\|A\| = b-a$ 為區間長度,

則此區間的反模糊化值定義為 $RA = c + [1 - \frac{\ln(1 + \|A\|)}{\|A\|}]$

定義 2.12 區間距離(Wu, 2010)

若 $A_1=[a_1, b_1]$ 、 $A_2=[a_2, b_2]$ 為兩個區間模糊數， $c_i=\frac{a_i+b_i}{2}$ ， $i=1, 2$ 分別為兩區間中心點、 $\|A_i\|=b_i - a_i$ ， $i=1, 2$ 分別為兩區間長度，則兩區間之距離為

$$d(A_1, A_2) = |c_1 - c_2| + \left| \frac{\ln(1+\|A_1\|)}{\|A_1\|} - \frac{\ln(1+\|A_2\|)}{\|A_2\|} \right|$$

例 2.1：設有兩組區間資料分別為 $X_1=[2,4]$ 、 $X_2=[3,7]$ 與 $Y_1=[3,9]$ 、 $Y_2=[7,9]$

其中心點距離皆為 2，無法比較出兩組距離遠近，因此我們以定義 2.11、2.12 分別計算其區間距離如下：

X_1 與 X_2 距離： $X_1=[2,4]$ 之區間中心為 3、區間長度為 2

$$\text{可得 } RX_1 = 3 + \left[1 - \frac{\ln(1+2)}{2}\right] = 3.5493$$

$X_2=[3,7]$ 之區間中心為 5、區間長度為 4

$$\text{可得 } RX_2 = 5 + \left[1 - \frac{\ln(1+4)}{4}\right] = 5.5976$$

$$\text{則 } d(X_1, X_2) = |3 - 5| + \left| \frac{\ln(1+2)}{2} - \frac{\ln(1+4)}{4} \right| = 2.1469$$

Y_1 與 Y_2 距離： Y_1 的區間中心為 6、區間長度為 6

$$\text{可得 } RY_1 = 6 + \left[1 - \frac{\ln(1+6)}{6}\right] = 6.6757$$

Y_2 的區間中心為 8、區間長度為 2

$$\text{可得 } RY_2 = 8 + \left[1 - \frac{\ln(1+2)}{2}\right] = 8.4507$$

$$\text{則 } d(Y_1, Y_2) = |6 - 8| + \left| \frac{\ln(1+6)}{6} - \frac{\ln(1+2)}{2} \right| = 2.225$$

故我們可以得到 $d(X_1, X_2) < d(Y_1, Y_2)$ ，即 X_1 與 X_2 兩區間距離較近。

定義 2.13 區間平均誤差(mean error of interval, IME)

令 $\{X_t = [a_t, b_t] = (c_t, r_t), t=1, 2, \dots, n\}$ 為一個區間時間數列，區間預測值為 $\hat{X}_t = [\hat{a}_t, \hat{b}_t] = (\hat{c}_t, \hat{r}_t)$ ， $\hat{\varepsilon}_t = d(\hat{X}_t, X_t)$ 為預測區間與實際區間之誤差，則定義區間平均誤差為

$$IME = \frac{1}{k} \sum_{t=n+1}^{n+k} \hat{\varepsilon}_t$$

其中 n 表當期時間、 k 表往後預測期數

定義 2.14 區間平均百分誤差(mean percent error of interval, IMPE)

令 $\{X_t = [a_t, b_t] = (c_t, r_t), t=1, 2, \dots, n\}$ 為一個區間時間數列，區間預測值為 $\hat{X}_t = [\hat{a}_t, \hat{b}_t] = (\hat{c}_t, \hat{r}_t)$ ， RX_t 為區間 $[a_t, b_t]$ 的反模糊化值， $\hat{\varepsilon}_t = d(\hat{X}_t, X_t)$ 為預測區間與實際區間之誤差，則定義區間平均百分誤差為

$$IMPE = \frac{100}{k} \sum_{t=n+1}^{n+k} \frac{\hat{\varepsilon}_t}{RX_t} \%$$

其中 n 表當期時間、 k 表往後預測期數

例 2.2：某地區溫度預測如下表

往前期數	實際溫度	預測溫度
1	[21, 31]	[20, 28]
2	[23, 27]	[24, 32]

則根據定義 2.11、2.12 可知

反模糊化值：

$x_1 = [21, 31]$ 其區間中心為 26、區間長度為 10

$$\text{則 } RX_1 = 26 + \left[1 - \frac{\ln(1+10)}{10}\right] = 26.7602$$

$\hat{x}_1 = [20, 28]$ 其區間中心為 24 區間長度為 8

$$\text{則 } R\hat{x}_1 = 24 + \left[1 - \frac{\ln(1+8)}{8}\right] = 24.7253$$

$x_2=[23,27]$ 其區間中心爲 25、區間長度爲 4

$$\text{則 } Rx_2 = 25 + [1 - \frac{\ln(1+4)}{4}] = 25.5976$$

$\widehat{x}_2=[24,32]$ 其區間中心爲 28、區間長度爲 8

$$\text{則 } R\widehat{x}_2 = 28 + [1 - \frac{\ln(1+8)}{8}] = 28.7253$$

區間距離：

$$\widehat{\epsilon}_1 = d(x_1, \widehat{x}_1) = |26 - 24| + \left| \frac{\ln(1+10)}{10} - \frac{\ln(1+8)}{8} \right| = 2.0349$$

$$\widehat{\epsilon}_2 = d(x_2, \widehat{x}_2) = |25 - 28| + \left| \frac{\ln(1+4)}{4} - \frac{\ln(1+8)}{8} \right| = 3.1277$$

則根據定義 2.13、2.14 可知區間平均誤差(IME)與區間平均百分誤差(IMPE)爲

$$IME = \frac{1}{k} \sum_{t=1}^{n+k} \widehat{\epsilon}_t = \frac{1}{2} (2.0349 + 3.1277) = 2.5813$$

$$IMPE = \frac{100}{k} \sum_{t=n+1}^{n+k} \frac{\widehat{\epsilon}_t}{Rx_t} \% = \frac{100}{2} \left(\frac{2.0349}{26.7602} + \frac{3.1277}{25.5976} \right) \% = 9.91\%$$

第三章 實證分析-來臺觀光客人數

3.1 資料來源

資料來源為我國交通部觀光局從西元 1979 年至西元 2010 年共 32 年來臺觀光客人數的月別統計資料，依據各年一到十二月每月不同的來臺觀光客人數，以月為單位選取該年當中最高人次、最低人次的月份，作為當年來臺觀光客人數的最大值、最小值，再分別以左右端點法、及中心點半徑法形成兩組各 32 筆的區間資料，圖 3.1~3.2 分別為其走勢圖，詳細數據如表 3.1 所示。(民國九十二年四月由於我國爆發 SARS 疫情，被列為旅遊紅燈區，故來臺人次數驟然減少，此年數據不列入計算)。

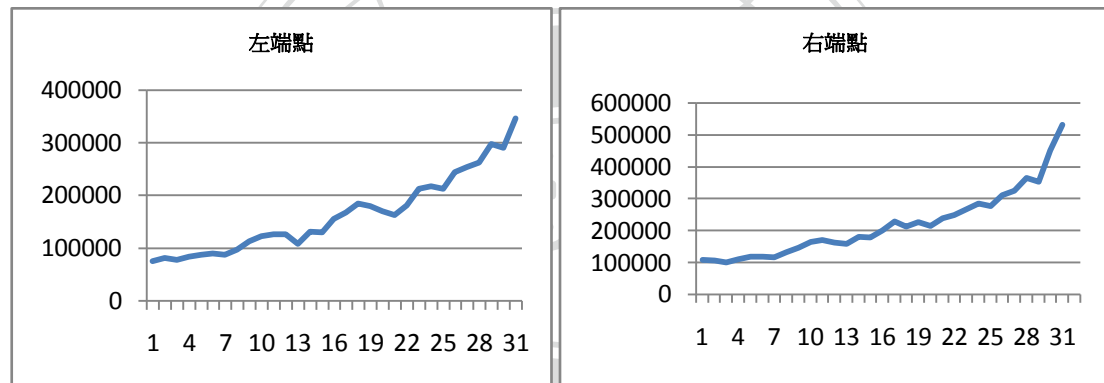


圖 3.1：來臺旅遊人次區間左、右端點走勢圖

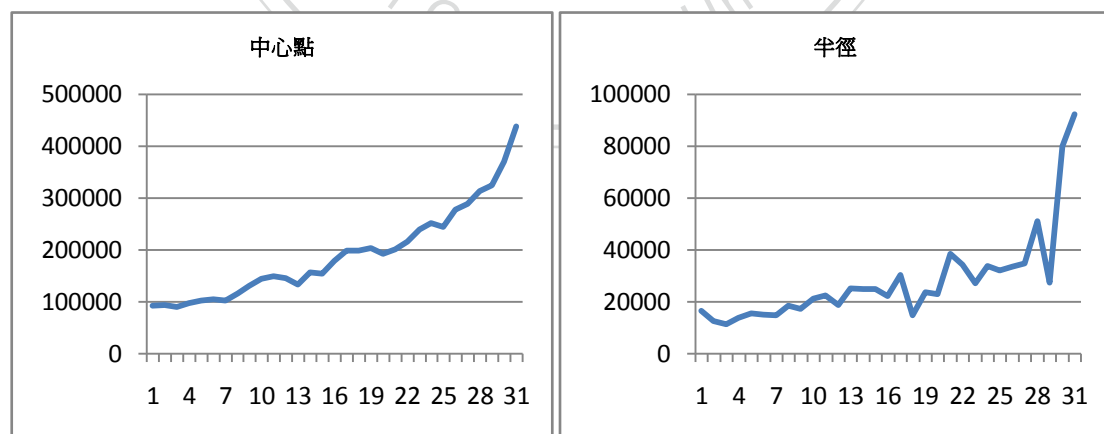


圖 3.2：來臺旅遊人次區間中心點、半徑走勢圖

表 3.1：來臺觀光旅客人數原始資料(1979~2010)

年份		單月來台旅遊人次		區間表示法	
西元	民國	單月最多	單月最少	左右端點法	中心點半徑法
1979	68	108059	75336	[75336,108059]	(91698;16362)
1980	69	105890	80928	[80928,105890]	(93409;12481)
1981	70	100878	78336	[78336,100878]	(89607;11271)
1982	71	111096	83461	[83461,111096]	(97279;13818)
1983	72	118054	86955	[86955,118054]	(102505;15550)
1984	73	119473	89701	[89701,119473]	(104587;14886)
1985	74	117034	87686	[87686, 117034]	(102360;14674)
1986	75	133426	96795	[96795,133426]	(115111;18315)
1987	76	147693	113312	[113312,147693]	(130503;17191)
1988	77	164811	122282	[122282,164811]	(143547;21265)
1989	78	171599	126773	[126776,171599]	(149188;22412)
1990	79	163715	126420	[126420,163715]	(145068;18648)
1991	80	157962	107532	[107532,157962]	(132747;25215)
1992	81	180962	131258	[131258, 180962]	(156110;24852)
1993	82	178962	129408	[129408, 178962]	(154185;24777)
1994	83	200433	155972	[155972,200433]	(178203;22231)
1995	84	228808	168140	[168140,228808]	(198474;30334)
1996	85	213496	184081	[184081,213496]	(198789;14708)
1997	86	227591	180176	[180176,227591]	(203884;23708)
1998	87	215647	169908	[169908,215647]	(192778;22870)
1999	88	238957	162282	[162282,238957]	(200620;38338)
2000	89	249352	180922	[180922,249352]	(215137;34215)
2001	90	266176	212103	[212103,266176]	(239140;27037)
2002	91	285303	217600	[217600,285303]	(251452;33852)
2003	92	259861	40256	[40256,259861]	(150059;109803)
2004	93	276680	212854	[212854,276680]	(244767;31913)
2005	94	311245	244252	[244252,311245]	(277749;33497)
2006	95	323931	254347	[254347,323931]	(289139;34792)
2007	96	363916	261799	[261799,363916]	(312858;51059)
2008	97	352038	297442	[297442,352038]	(324740;27298)
2009	98	449806	290099	[290099,449806]	(369953;79854)
2010	99	530594	345981	[345981,530594]	(438288;92307)

3.2 以區間型 ARIMA 模式建構

我們利用 2.3 節中的定義 2.7 及 2.8，以表 3.1 中的區間資料為依據，分別觀察左端點、右端點與中心點、半徑四個時間數列的 ACF 圖與 PACF 圖，以 1920 年 Box 與 Jenkins 所提出的「自我迴歸移動平均」整合模式建構程序，以 ADF 單根檢定(Augmented Dickey-Fuller Test)驗證模式是否為定態(stationary)，適當的選擇模式中的差分階數(d)、自我迴歸級數(p)、移動平均級數(q)，分段配適出其個別的 ARIMA(p,d,q)模式，最後再綜合左、右端點與中心點、半徑兩兩合成不同的區間時間數列。

以左右端點法建構

左端點之 ARIMA 模式：

$$(1-0.0045B)(1-B)a_t = 0.0019 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

右端點之 ARIMA 模式：

$$(1-0.0152B)(1-B)b_t = 0.0126 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

左右端點法之 ARIMA 模式：

$$[a_t, b_t] = \begin{cases} (1-0.0045B)(1-B)a_t = 0.0019 + \varepsilon_t \\ (1-0.0152B)(1-B)b_t = 0.0126 + \varepsilon_t \end{cases} \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

以中心點半徑法建構

中心點之 ARIMA 模式：

$$(1-0.0095B)(1-B)c_t = 0.0089 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

半徑之 ARIMA 模式：

$$(1-0.0021B)(1-B)r_t = 0.0007 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

中心點半徑法之 ARIMA 模式：

$$(c_t, r_t) = \begin{cases} (1-0.0095B)(1-B)c_t = 0.0089 + \varepsilon_t \\ (1-0.0002B)(1-B)r_t = 0.0007 + \varepsilon_t \end{cases} \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

3.3 以平均累加模糊熵分類

由於在建構區間 SETAR 模式前需事先給定門檻值，因此在本節中我們希望先以平均累加模糊熵分別對左、右端點，中心點、半徑作分類，分析各數列中的結構轉變之處，藉此客觀的找出可能的門檻候選值。

第一步：先利用 k-means method(Sharma, 1996)找出時間序列 $\{Y_t\}$ 的 2 個群落中心，並決定 $\{Y_t\}$ 對 2 個群落中心的隸屬度 $\mu_{it}, i = 1, 2$

表 3.2：左、右端點與中心點、半徑的兩個群落中心

	左端點	右端點	中心點	半徑
群落中心 C_1	134306	191195	162751	28444
群落中心 C_2	259609	351077	305343	45734

第二步：計算出對應的模糊熵 $\delta(y_t)$ 、平均累加模糊熵 $MS\delta(y_t) = \frac{1}{t} \sum_{i=1}^t \delta(y_i)$

及此數列的中位數 $Median(MS\delta(y_t))$ ，繪成如圖 3.3、3.5、3.7、3.9 之走勢圖

表 3.3：左、右端點與中心點、半徑的平均累加模糊熵中位數

	左端點	右端點	中心點	半徑
$Median(MS\delta(y_t))$	0.042632	0.039048	0.040814	0.041448

第三步：取適當的一門檻值 λ ，將 $\{Y_t\}$ 對應的平均累加模糊熵 $MS\delta(y_t)$ 數列進行分類。若 $MS\delta(y_t)$ 落在區間 $[0, Median(MS\delta(y_t)) - \lambda]$ ，則以 1 表示第一組；若 $MS\delta(y_t)$ 落在區間 $[Median(MS\delta(y_t)) - \lambda, Median(MS\delta(y_t)) + \lambda]$ ，則以 2 表示第二組；若 $MS\delta(y_t)$ 落在區間 $[Median(MS\delta(y_t)) + \lambda, 1]$ ，則以 3 表示第三組，依據理論中之分類法，繪成如圖 3.4、3.6、3.8、3.10 之分類圖

第四步：選取適當的顯著水準 α ，此時取 $\alpha = 0.2$ 若連串的樣本數大於 $[32\alpha] = 6$ 時我們才算分類成功，反知將視為轉折型式歸納分組，當分類的組數超過一組時，表示此數列發生結構改變，進而找出其可能的門檻轉折點。

觀察圖 3.3~3.10 可以發現左端點與右端點的轉型期皆發生在 25 期(92 年)到 28 期(95 年)；中心點發生在 23 期(90 年)到 26 期(93 年)；半徑則發生在 24 期(91 年)到 27 期(94 年)，將轉型期中的各期人數作為門檻候選值如下表：

表 3.4：左、右端點與中心點、半徑的門檻候選值

門檻候選值	左端點	右端點	中心點	半徑
r_1	217600	259861	239140	33852
r_2	212854	276680	251452	31913
r_3	244252	311245	244767	33497
r_4	254347	323931	277749	34792

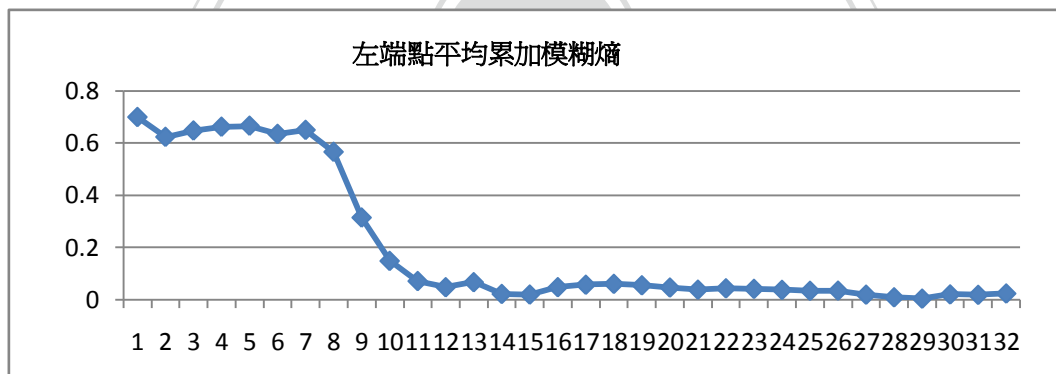


圖 3.3 左端點之平均累加模糊熵走勢圖

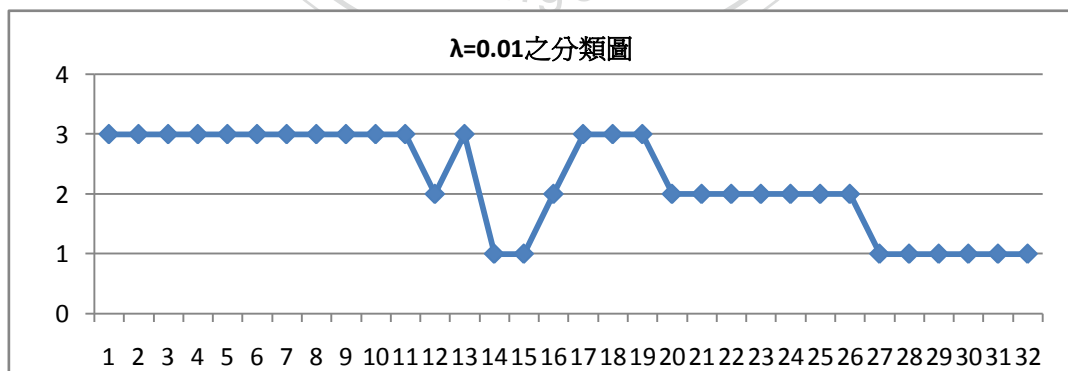


圖 3.4 左端點以 $\lambda = 0.01$ 所作出之分類圖

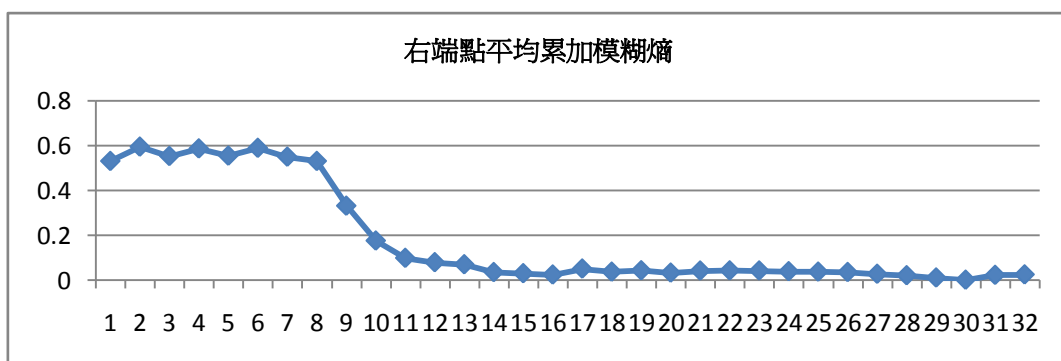


圖 3.5 右端點之平均累加模糊熵走勢圖

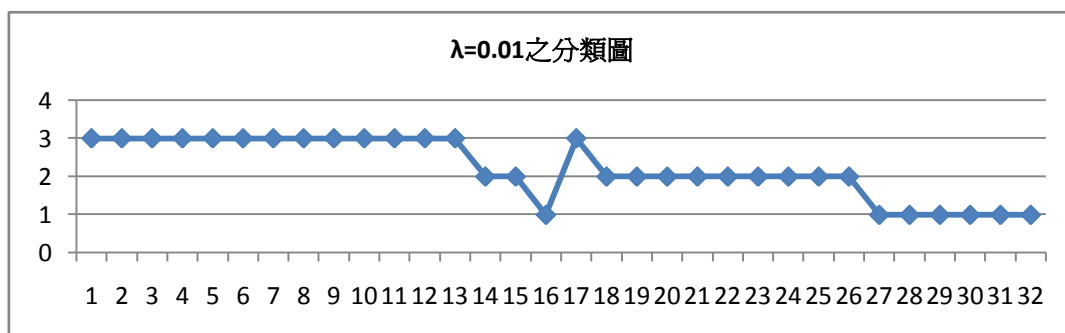


圖 3.6 右端點以 $\lambda=0.01$ 所作出之分類圖

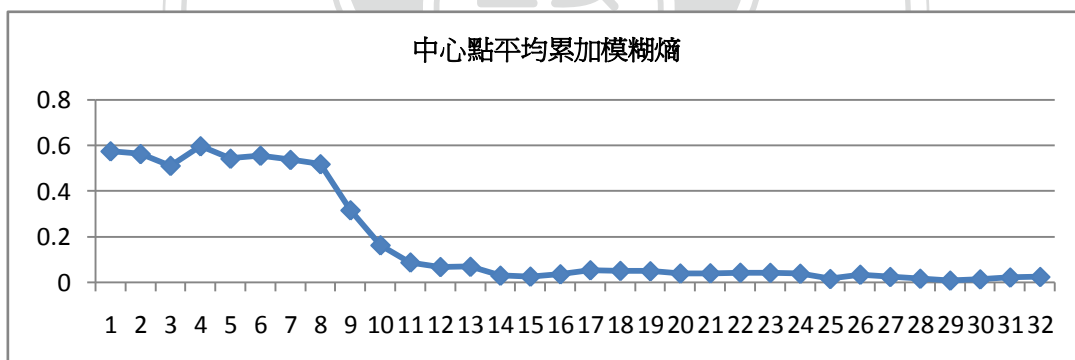


圖 3.7 中心點之平均累加模糊熵走勢圖

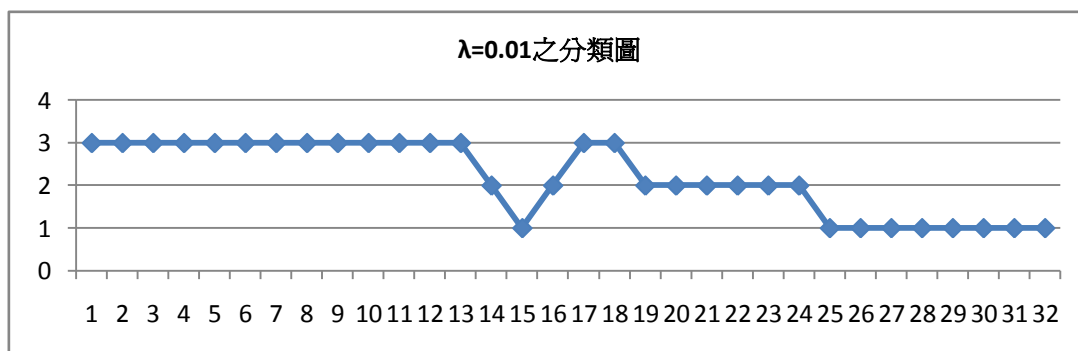


圖 3.8 中心點以 $\lambda=0.01$ 所作出之分類圖

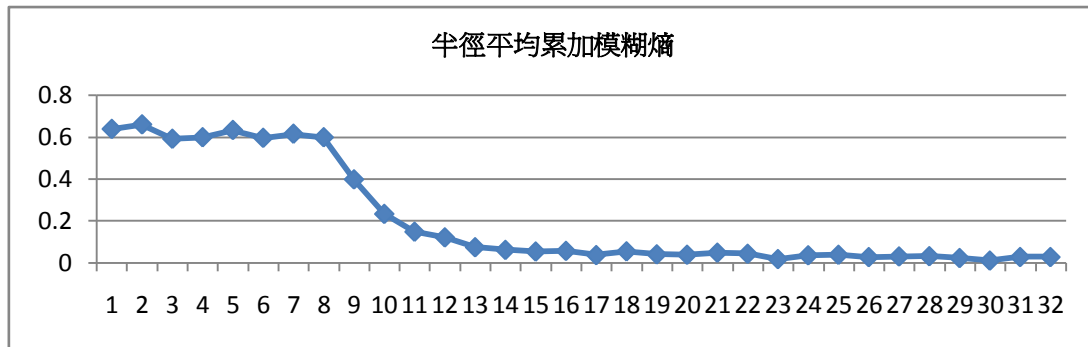


圖 3.9 半徑之平均累加模糊熵走勢圖

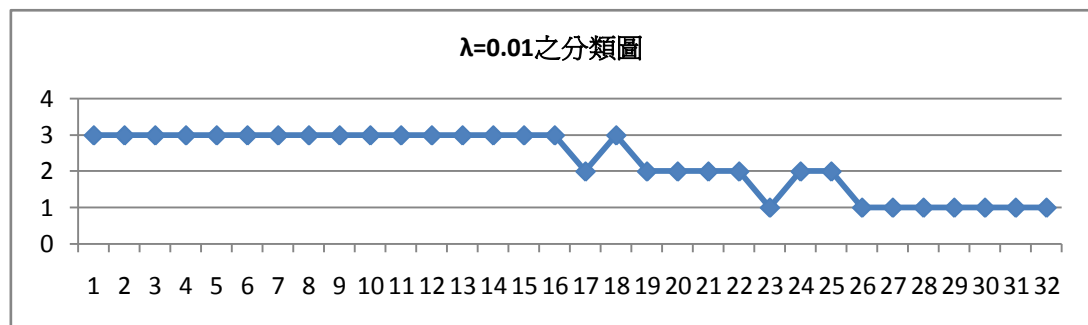


圖 3.10 半徑以 $\lambda=0.01$ 所作出之分類圖

3.4 以區間型門檻自迴歸模式建構

本節針對左右端點、中心點半徑兩種區間模糊數的定義方式，把左端點、右端點、中心點、半徑假設為四個獨立的時間數列，先利用上節所找出的門檻候選值，並以 Tong 所建議的 $L=[32^{0.5}]=5$ 作為模式最大可能階次，依 3.2 所述之步驟建構，分別計算各種可能的 AIC 值，以 AIC 最小的延誤參數和門檻值配適出 SETAR 模式；另外再用可能影響觀光的某一重大歷史事件作為一特殊時點，前後分段形成不同的自迴歸模式，最後我們再分別將左右端點、中心點半徑兩兩合成不同的區間型門檻自迴歸模式。

1. 區間型 SETAR 模式

以左右端點法建構

左端點之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 3.5：以不同的門檻值與延誤參數計算左端點之 AIC 值

$\{r=r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
$AIC(1, r_1)$	$AIC(2, r_1)$	$AIC(3, r_1)$	$AIC(4, r_1)$	$AIC(5, r_1)$
44.31288	46.19569	47.96844	50.36749	51.66678
$AIC(1, r_2)$	$AIC(2, r_2)$	$AIC(3, r_2)$	$AIC(4, r_2)$	$AIC(5, r_2)$
46.52504	46.32696	49.36457	49.36529	52.99967
$AIC(1, r_3)$	$AIC(2, r_3)$	$AIC(3, r_3)$	$AIC(4, r_3)$	$AIC(5, r_3)$
48.01642	45.90493	48.63794	50.99874	52.66647
$AIC(1, r_4)$	$AIC(2, r_4)$	$AIC(3, r_3)$	$AIC(4, r_3)$	$AIC(5, r_3)$
48.45365	46.40886	49.36277	51.69997	53.67112

比較上列各 AIC 值發現 $(d, r)=(1, 217600)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到左端點的門檻自迴歸模式如下：

$$a_t = \begin{cases} 0.0009a_{t-1} + \varepsilon_t & \text{if } a_{t-1} < 217600 \\ 0.0120a_{t-1} + \varepsilon_t & \text{if } a_{t-1} \geq 217600 \end{cases} \quad \varepsilon_t \text{ 為白干擾項}$$

右端點之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 3.6：以不同的門檻值與延誤參數計算右端點之 AIC 值

$\{r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
AIC(1, r_1)	AIC(2, r_1)	AIC(3, r_1)	AIC(4, r_1)	AIC(5, r_1)
45.93630	46.99371	47.62359	50.36471	52.31479
AIC(1, r_2)	AIC(2, r_2)	AIC(3, r_2)	AIC(4, r_2)	AIC(5, r_2)
46.42516	47.29922	48.69736	50.93149	53.62179
AIC(1, r_3)	AIC(2, r_3)	AIC(3, r_3)	AIC(4, r_3)	AIC(5, r_3)
46.68541	47.61810	49.36712	49.67812	52.34167
AIC(1, r_4)	AIC(2, r_4)	AIC(3, r_4)	AIC(4, r_4)	AIC(5, r_4)
47.16106	47.74531	49.33258	51.64187	53.41796

比較上列各 AIC 值發現 $(d, r) = (1, 259861)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到左端點的門檻自迴歸模式如下：

$$b_t = \begin{cases} 0.0038b_{t-1} + \varepsilon_t & \text{if } b_{t-1} < 259861 \\ 0.0590b_{t-1} + \varepsilon_t & \text{if } b_{t-1} \geq 259861 \end{cases} \quad \varepsilon_t \text{ 為白干擾項}$$

左右端點法之區間型 SETAR 模式：

綜合以上左端點及右端點之 SETAR 模式合成區間型 SETAR 模式如下：

$$[a_t, b_t] = \begin{cases} a_t = \begin{cases} 0.0009a_{t-1} + \varepsilon_t & \text{if } a_{t-1} < 217600 \\ 0.0120a_{t-1} + \varepsilon_t & \text{if } a_{t-1} \geq 217600 \end{cases} \\ b_t = \begin{cases} 0.0038b_{t-1} + \varepsilon_t & \text{if } b_{t-1} < 259681 \\ 0.0590b_{t-1} + \varepsilon_t & \text{if } b_{t-1} \geq 259681 \end{cases} \end{cases} \quad \varepsilon_t \text{ 為白干擾項}$$

以中心點半徑法建構

中心點之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 3.7：以不同的門檻值與延誤參數計算中心點之 AIC 值

$\{r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
AIC(1, r_1)	AIC(2, r_1)	AIC(3, r_1)	AIC(4, r_1)	AIC(5, r_1)
43.62987	44.63219	46.35179	47.64439	48.32719
AIC(1, r_2)	AIC(2, r_2)	AIC(3, r_2)	AIC(4, r_2)	AIC(5, r_2)
44.32179	44.99867	46.35298	47.00929	48.32985
AIC(1, r_3)	AIC(2, r_3)	AIC(3, r_3)	AIC(4, r_3)	AIC(5, r_3)
45.66378	45.11197	47.99876	46.16327	49.22279
AIC(1, r_4)	AIC(2, r_4)	AIC(3, r_4)	AIC(4, r_4)	AIC(5, r_4)
45.61987	46.33217	47.33214	47.63477	50.11147

比較上列各 AIC 值發現 $(d, r)=(1, 277749)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到中心點的門檻自迴歸模式如下：

$$c_t = \begin{cases} 0.008c_{t-1} + \varepsilon_t & \text{if } c_{t-1} < 277749 \\ 0.014c_{t-1} + \varepsilon_t & \text{if } c_{t-1} \geq 277749 \end{cases} \quad \varepsilon_t \text{ 為白干擾項}$$

半徑之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 3.8：以不同的門檻值與延誤參數計算半徑之 AIC 值

$\{r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
AIC(1, r_1)	AIC(2, r_1)	AIC(3, r_1)	AIC(4, r_1)	AIC(5, r_1)
25.21444	26.31487	27.14679	27.3125	29.33341
AIC(1, r_2)	AIC(2, r_2)	AIC(3, r_2)	AIC(4, r_2)	AIC(5, r_2)
26.14279	27.31288	27.31449	28.31499	28.31620
AIC(1, r_3)	AIC(2, r_3)	AIC(3, r_3)	AIC(4, r_3)	AIC(5, r_3)
27.61549	27.31468	27.31497	27.9986	29.31561
AIC(1, r_4)	AIC(2, r_4)	AIC(3, r_4)	AIC(4, r_4)	AIC(5, r_4)
27.31963	27.31468	28.11007	28.31497	30.14458

比較上列各 AIC 值發現 $(d, r) = (1, 33497)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到半徑的門檻自迴歸模式如下：

$$r_t = \begin{cases} 0.0005r_{t-1} + \varepsilon_t & \text{if } r_{t-1} < 33497 \\ 0.0067r_{t-1} + \varepsilon_t & \text{if } r_{t-1} \geq 33497 \end{cases} \quad \varepsilon_t \text{ 為白干擾項}$$

中心點半徑之區間型 STAR 模式：

綜合以上中心點及半徑之 SETAR 模式合成區間型 SETAR 模式如下：

$$(c_t, r_t) = \begin{cases} c_t = \begin{cases} 0.008c_{t-1} + \varepsilon_t & \text{if } c_{t-1} < 277749 \\ 0.014c_{t-1} + \varepsilon_t & \text{if } c_{t-1} \geq 277749 \end{cases} \\ r_t = \begin{cases} 0.0005r_{t-1} + \varepsilon_t & \text{if } r_{t-1} < 33497 \\ 0.0067r_{t-1} + \varepsilon_t & \text{if } r_{t-1} \geq 33497 \end{cases} \end{cases} \quad \varepsilon_t \text{ 為白干擾項}$$

2. 區間型 TAR 模式

由於民國 37 年國共內戰爆發，爲了確保臺灣主權，全國開始進入動員戡亂(戒嚴)時期，當時政府對於觀光採取保守的態度，針對所有國內外出入境的旅客皆有所限制，一直到民國 80 年召開第一屆國民大會，一致通過廢止「動員戡亂臨時條款」後，終於宣佈終止長達近四十多年的動員戡亂時期，自此之後，我國的觀光產業才真正獲得發展的空間，因此我們以民國 80 年(第 13 期)爲重大歷史事件的時間轉折點，在此時點前後分別以 AIC 準則配適自迴歸模式，建構左右端點法及中心點半徑之區間型 TAR 模式。

以左右端點法建構

左端點之 TAR 模式：

$$a_t = \begin{cases} (1 - 0.0009B)(1 - B)a_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.0120B)(1 - B)a_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \quad \begin{matrix} B \text{ 為倒退算子 } t \text{ 為期數} \end{matrix}$$

右端點之 TAR 模式：

$$b_t = \begin{cases} (1 - 0.0038B)(1 - B)b_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.0590B)(1 - B)b_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \quad \begin{matrix} B \text{ 為倒退算子 } t \text{ 為期數} \end{matrix}$$

左右端點法之區間型 TAR 模式：

最後綜合以上左端點及右端點之 TAR 模式合成區間型 TAR 模式如下

$$[a_t, b_t] = \begin{cases} a_t = \begin{cases} (1 - 0.0009B)(1 - B)a_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.0120B)(1 - B)a_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \\ b_t = \begin{cases} (1 - 0.0038B)(1 - B)b_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.0590B)(1 - B)b_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \end{cases}$$

以中心點半徑法建構

中心點之 TAR 模式：

$$c_t = \begin{cases} (1 - 0.008B)(1 - B)c_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.014B)(1 - B)c_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \quad \begin{matrix} B \text{ 為倒退算子 } t \text{ 為期數} \end{matrix}$$

半徑之 TAR 模式：

$$r_t = \begin{cases} (1 - 0.0005B)(1 - B)r_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.0067B)(1 - B)r_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \quad \begin{matrix} B \text{ 為倒退算子 } t \text{ 為期數} \end{matrix}$$

中心點半徑法之區間型 TAR 模式：

最後綜合以上中心點及半徑之 TAR 模式合成區間型 TAR 模式如下：

$$(c_t, r_t) = \begin{cases} c_t = \begin{cases} (1 - 0.008B)(1 - B)c_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.014B)(1 - B)c_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \\ r_t = \begin{cases} (1 - 0.0005B)(1 - B)r_{t-1} + \varepsilon_t & \text{if } t < 13 \\ (1 - 0.0067B)(1 - B)r_{t-1} + \varepsilon_t & \text{if } t \geq 13 \end{cases} \end{cases}$$

3.5 預測結果比較與分析

我們以 2.4 節中所定義的區間平均誤差 (IME)、區間平均絕對誤差和(IMPE)作為模式評估依據，分別對所建構的區間時間數列進行分析，詳如下列各表：

表 3.9 以 IME、IMPE 評估左右端點法之預測結果

保留 期數	實際區間	左右端點法		
		區間 ARIMA	區間 SETAR	區間 TAR
2008	[297442,352038]	[324759,365197]	[309187,357819]	[288943,347614]
2009	[290099,449806]	[269812,398711]	[283199,441339]	[298961,451135]
2010	[345981,530594]	[345988,543512]	[355976,532371]	[336865,552899]
IME		11967.24	11967.24	5812.61
IMPE		10.21%	10.21%	4.32%

表 3.10 以 IME、IMPE 評估中心點半徑法之預測結果

保留 期數	實際區間	中心點半徑法		
		區間 ARIMA	區間 SETAR	區間 TAR
2008	(324740;27298)	(34127;25471)	(324579;26042)	(330112;27035)
2009	(369953;79854)	(369214;73985)	(364671;78687)	(357862;80051)
2010	(438288;92307)	(453986;97869)	(443271;92632)	(446782;93671)
IME		12276.11	6533.43	5599.34
IMPE		11.57%	6.33%	4.01%

由以上各表中的比較，我們可以藉由 IME、IMPE 為模式優劣比較工具，發現中心點半徑法的預測效果與左右端點法差不多，而三種模式之預測能力依序為區間 TAR > 區間 SETAR >> 區間 ARIMA，表示兩種門檻自迴歸模式都比單純的區間 ARIMA 法好得多，這正符合我們的假設，表示若能找出發生結構改變之處，對於掌握資料走勢或是做進一步的預測分析都會有較好的結果。

第四章 實證分析-新臺幣兌美元匯率

4.1 資料來源

資料來源為「財團法人台北外匯市場發展基金會」，搜集自 2011 年 2 月 16 日到 3 月 31 日 (扣除非交易日)共 31 筆新臺幣兌美元的單日統計資料，依據每日最高、最低匯價作為單日匯率的最大值、最小值，再分別以左右端點法、及中心點半徑法形成兩組各 31 筆的區間資料，圖 4.1~4.2 分別為其走勢圖，詳細數據如表 4.1 所示。

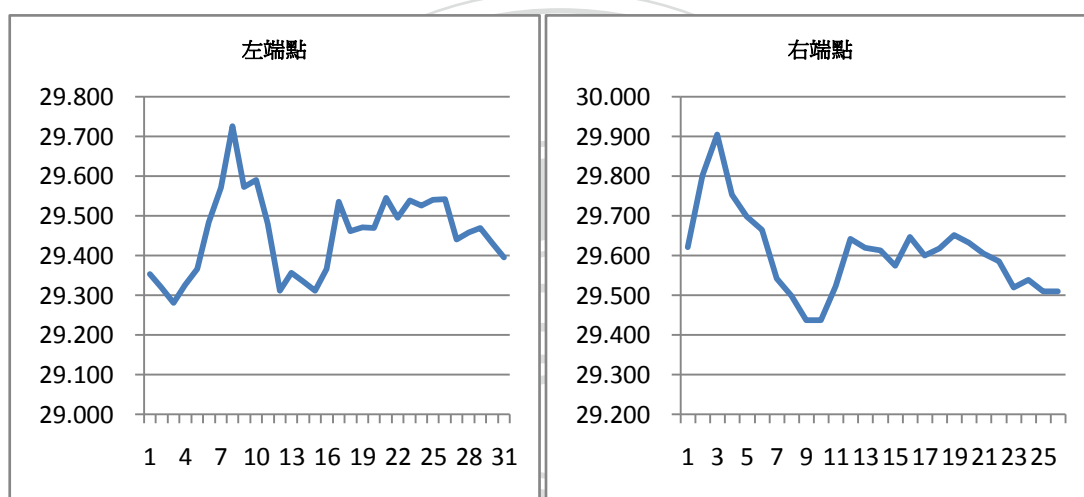


圖 4.1：新臺幣兌美元匯率區間左、右端點走勢圖

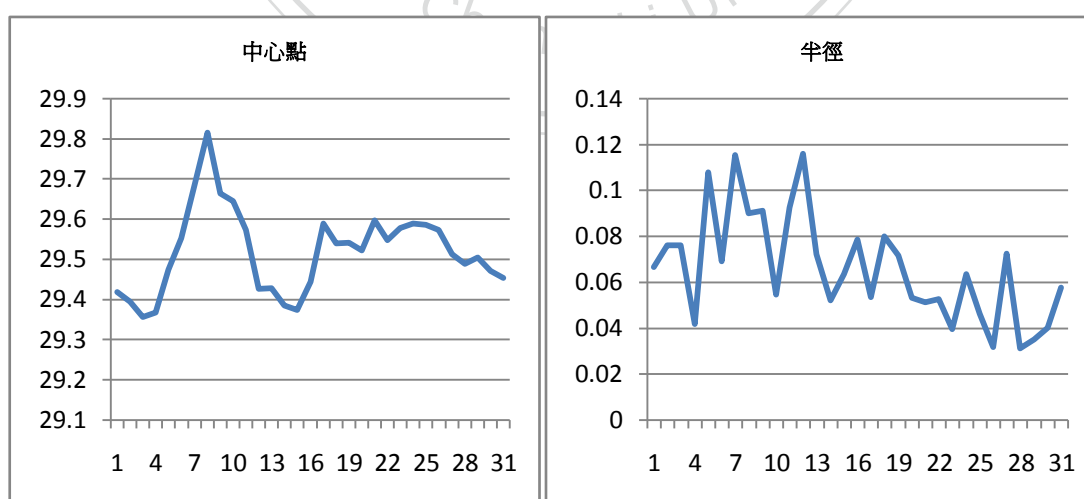


圖 4.2：新臺幣兌美元匯率區間中心點、半徑走勢圖

表 4.1：新臺幣兌美元原始資料(2011/2/16~3/31)

日期	美元兌新臺幣每日匯率		區間表示法	
	最高匯價	最低匯價	左右端點法	中心點半徑法
2011/2/16	29.479	29.470	[29.470, 29.479]	(29.4185;0.0665)
2011/2/17	29.470	29.432	[29.432, 29.470]	(29.3940;0.0760)
2011/2/18	29.432	29.408	[29.408, 29.432]	(29.3560;0.0760)
2011/2/21	29.408	29.366	[29.366, 29.408]	(29.3665;0.0415)
2011/2/22	29.366	29.557	[29.557, 29.366]	(29.4740;0.1080)
2011/2/23	29.557	29.572	[29.572, 29.557]	(29.5520;0.0690)
2011/2/24	29.572	29.760	[29.572, 29.760]	(29.6855;0.1155)
2011/2/25	29.760	29.754	[29.754, 29.760]	(29.8150;0.0900)
2011/3/1	29.754	29.572	[29.572, 29.754]	(29.6630;0.0910)
2011/3/2	29.699	29.590	[29.590, 29.699]	(29.6445;0.0545)
2011/3/3	29.665	29.480	[29.480, 29.665]	(29.5725;0.0925)
2011/3/4	29.542	29.310	[29.310, 29.542]	(29.4260;0.1160)
2011/3/7	29.499	29.355	[29.355, 29.499]	(29.4270;0.0720)
2011/3/8	29.437	29.333	[29.333, 29.437]	(29.3850;0.0520)
2011/3/9	29.437	29.310	[29.310, 29.437]	(29.3735;0.0635)
2011/3/10	29.522	29.365	[29.365, 29.522]	(29.4435;0.0785)
2011/3/11	29.642	29.535	[29.535, 29.642]	(29.5885;0.0535)
2011/3/14	29.620	29.460	[29.460, 29.620]	(29.5400;0.0800)
2011/3/15	29.613	29.470	[29.470, 29.613]	(29.5415;0.0715)
2011/3/16	29.574	29.468	[29.468, 29.574]	(29.5210;0.0530)
2011/3/17	29.647	29.545	[29.545, 29.647]	(29.5960;0.0510)
2011/3/18	29.600	29.495	[29.495, 29.600]	(29.5475;0.0525)
2011/3/21	29.617	29.538	[29.538, 29.617]	(29.5775;0.0395)
2011/3/22	29.652	29.525	[29.525, 29.652]	(29.5885;0.0635)
2011/3/23	29.632	29.540	[29.540, 29.632]	(29.5860;0.0460)
2011/3/24	29.605	29.542	[29.542, 29.605]	(29.5735;0.0315)
2011/3/25	29.585	29.440	[29.440, 29.585]	(29.5125;0.0725)
2011/3/28	29.519	29.457	[29.457, 29.519]	(29.4880;0.0310)
2011/3/29	29.539	29.469	[29.469, 29.539]	(29.5040;0.0350)
2011/3/30	29.510	29.430	[29.430, 29.510]	(29.4700;0.0400)
2011/3/31	29.510	29.395	[29.395, 29.510]	(29.4525;0.0575)

4.2 以區間型 ARIMA 模式建構

我們利用 2.3 節中的定義 2.8 及 2.9，以表 4.1 中的區間資料為依據，分別觀察左端點、右端點與中心點、半徑四個時間數列的 ACF 圖與 PACF 圖，以西元 1920 年 Box 與 Jenkins 所提出的「自我迴歸移動平均」整合模式建構程序，以 ADF 單根檢定(Augmented Dickey-Fuller Test)驗證模式是否為定態(stationary)，適當的選擇模式中的差分階數(d)、自我迴歸級數(p)、移動平均級數(q)，配適出其個別的 ARIMA(p,d,q)模式，最後再綜合左、右端點與中心點、半徑合成兩組不同的區間時間數列。

以左右端點法建構

左端點之 ARIMA 模式：

$$(1-0.7335B)a_t = 0.1280 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

右端點之 ARIMA 模式：

$$(1-0.7894B)b_t = 0.1171 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

左右端點法之 ARIMA 模式：

$$[a_t, b_t] = \begin{cases} (1-0.7335B)a_t = 0.1280 + \varepsilon_t \\ (1-0.7894B)b_t = 0.1171 + \varepsilon_t \end{cases} \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

以中心點半徑法建構

中心點之 ARIMA 模式：

$$(1-0.7868B)c_t = 0.0075 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

半徑之 ARIMA 模式：

$$(1-0.2943B)r_t = 0.1799 + \varepsilon_t \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

中心點半徑法之 ARIMA 模式：

$$(c_t, r_t) = \begin{cases} (1-0.7868B)c_t = 0.0075 + \varepsilon_t \\ (1-0.2943B)r_t = 0.1799 + \varepsilon_t \end{cases} \quad B \text{ 為倒退算子 } \varepsilon_t \text{ 為白干擾項}$$

4.3 以平均累加模糊熵分類

由於在建構區間 SETAR 模式前需事先給定門檻值，因此在本節中我們希望先以平均累加模糊熵分別對左、右端點，中心點、半徑作分類，分析各數列中的結構轉變之處，藉此客觀的找出可能的門檻候選值。

第一步：先利用 k-means method(Sharma, 1996)找出時間序列 $\{Y_t\}$ 的 2 個群落中心，並決定 $\{Y_t\}$ 對 2 個群落中心的隸屬度 μ_{it} , $i = 1, 2$

表 4.2：左、右端點與中心點、半徑的兩個群落中心

	左端點	右端點	中心點	半徑
群落中心 C_1	29.3618	29.4923	29.4270	0.0653
群落中心 C_2	29.5281	29.6607	29.5944	0.0663

第二步：計算出對應的模糊熵 $\delta(y_t)$ 、平均累加模糊熵 $MS\delta(y_t) = \frac{1}{t} \sum_{i=1}^t \delta(y_i)$

及此數列的中位數 $Median(MS\delta(y_t))$ ，繪成如圖 4.3、4.5、4.7、4.9 之走勢圖

表 4.3：左、右端點與中心點、半徑的平均累加模糊熵中位數

	左端點	右端點	中心點	半徑
$Median(MS\delta(y_t))$	0.034503	0.031965	0.032264	0.046285

第三步：取適當的一門檻值 λ ，將 $\{Y_t\}$ 對應的平均累加模糊熵 $MS\delta(y_t)$ 數列進行分類。若 $MS\delta(y_t)$ 落在區間 $[0, Median(MS\delta(y_t)) - \lambda]$ ，則以 1 表示第一組；若 $MS\delta(y_t)$ 落在區間 $[Median(MS\delta(y_t)) - \lambda, Median(MS\delta(y_t)) + \lambda]$ ，則以 2 表示第二組；若 $MS\delta(y_t)$ 落在區間 $[Median(MS\delta(y_t)) + \lambda, 1]$ ，則以 3 表示第三組，依據理論中之分類法，繪成如圖 4.4、4.6、4.8、4.10 之分類圖

第四步：選取適當的顯著水準 α ，此時取 $\alpha = 0.2$ 若連串的樣本數大於 $[31\alpha] = 6$ 時我們才算分類成功，反知將視為轉折型式歸納分組，當分類的組數超過一組時，表示此數列發生結構改變，進而找出其可能的門檻轉折點。

觀察圖 4.3~4.10 可以發現左端點的轉型期皆發生在 21 期(3/17)到 24 期(3/22)；右端點發生在 22 期(3/18)到 25 期(3/23)；中心點發生在 23 期(3/21)到 26 期(3/24)；半徑則發生在 19 期(3/15)到 22 期(3/18)，將轉型期中的各期匯價作為門檻候選值如下表：

表 4.4：左、右端點與中心點、半徑的門檻候選值

門檻候選值	左端點	右端點	中心點	半徑
r_1	29.545	29.600	29.5775	0.0715
r_2	29.495	29.617	29.5885	0.0530
r_3	29.538	29.652	29.5860	0.0510
r_4	29.525	29.632	29.5735	0.0525

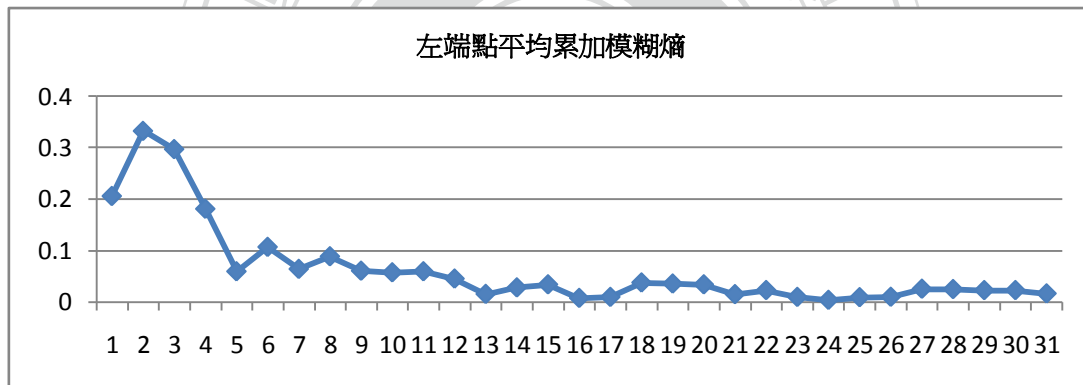


圖 4.3 左端點之平均累加模糊熵走勢圖

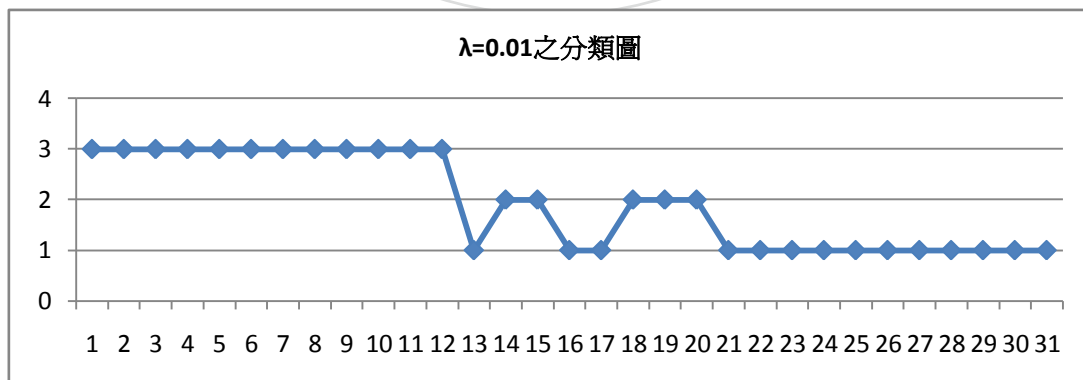


圖 4.4 左端點以 $\lambda=0.01$ 所作出之分類圖

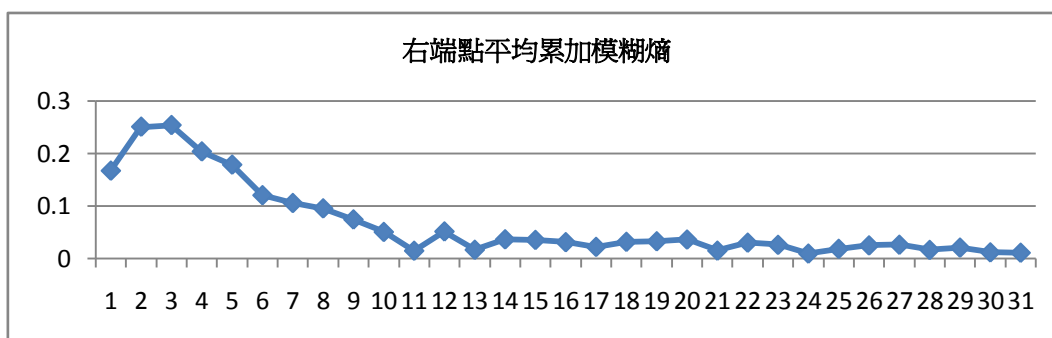


圖 4.5 右端點之平均累加模糊熵走勢圖

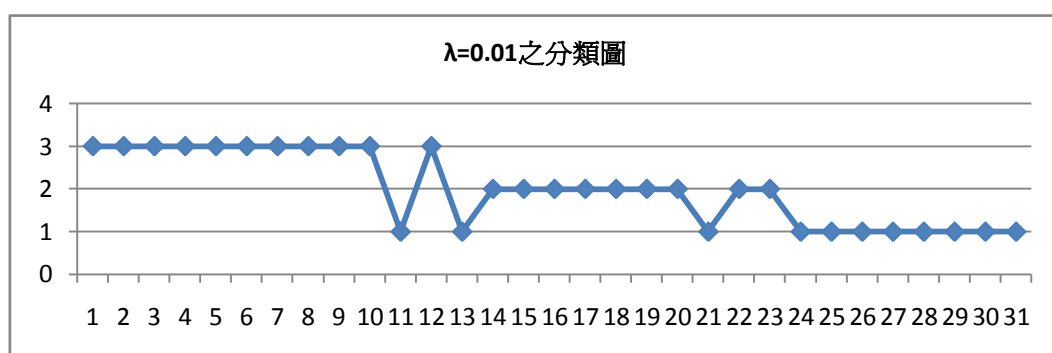


圖 4.6 右端點以 $\lambda=0.01$ 所作出之分類圖

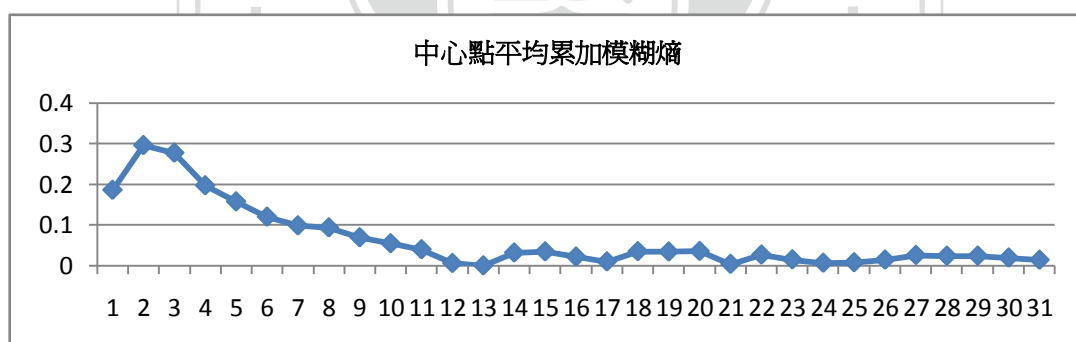


圖 4.7 中心點之平均累加模糊熵走勢圖

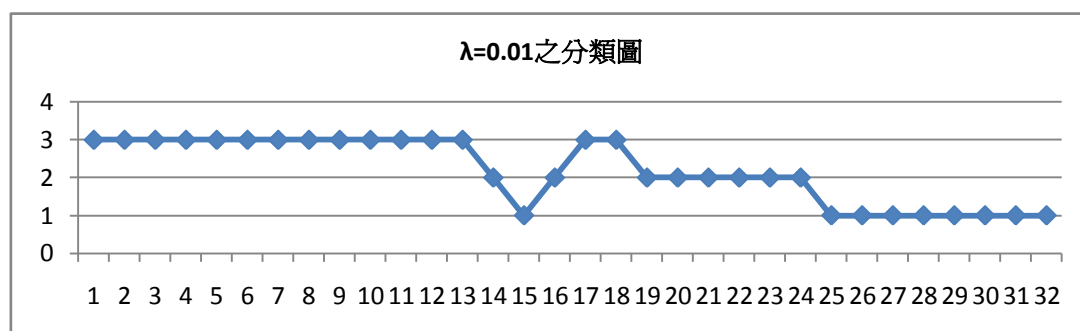


圖 4.8 中心點以 $\lambda=0.01$ 所作出之分類圖

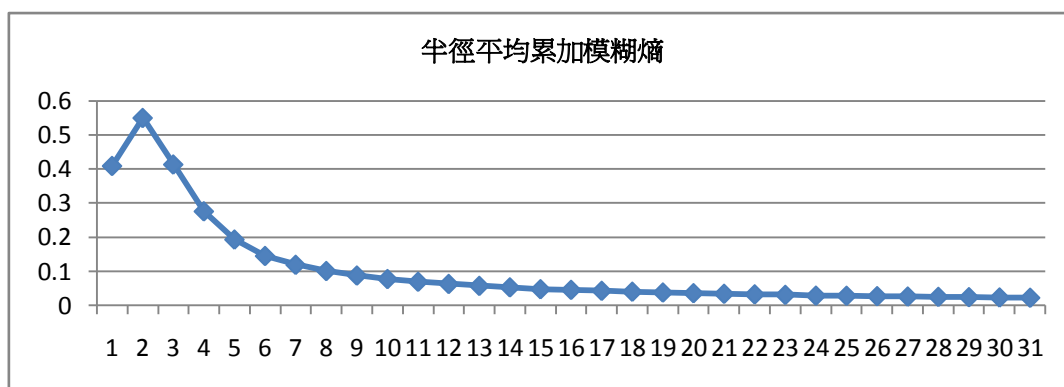


圖 4.9 半徑之平均累加模糊熵走勢圖

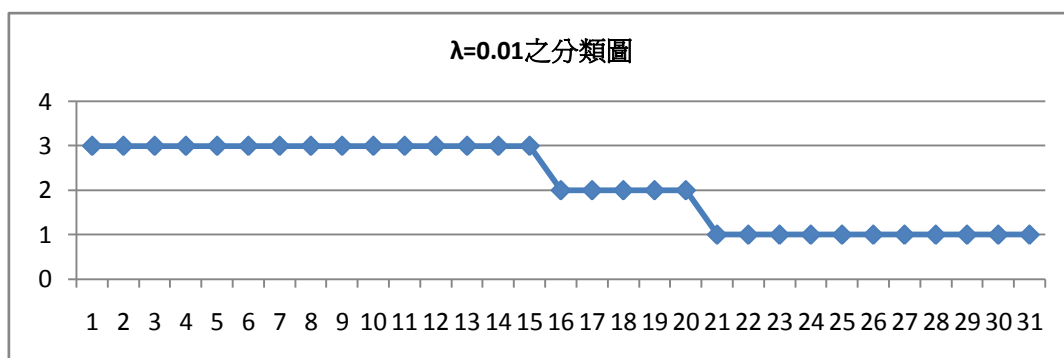


圖 4.10 半徑以 $\lambda=0.01$ 所作出之分類圖

4.4 以區間型門檻自迴歸模式建構

本節針對左右端點、中心點半徑兩種區間模糊數的定義方式，把左端點、右端點、中心點、半徑假設為四個獨立的時間數列，先利用上節所找出的門檻候選值，並以 Tong 所建議的 $L=[32^{0.5}]=5$ 作為最大可能階次，依 3.2 所述之步驟建構，分別計算各種可能的 AIC 值，以 AIC 最小的延誤參數和門檻值配適出 SETAR 模式；另外再用可能影響匯率的某一重大歷史事件作為一特殊時點，前後分段形成不同的自迴歸模式，最後我們再分別將左右端點、中心點半徑兩兩合成不同的區間型門檻自迴歸模式。

1. 區間型 SETAR 模式

以左右端點法建構

左端點之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 4.5：以不同的門檻值與延誤參數計算左端點之 AIC 值

$\{r=r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
$AIC(1, r_1)$	$AIC(2, r_1)$	$AIC(3, r_1)$	$AIC(4, r_1)$	$AIC(5, r_1)$
-4.5712	-3.7518	-2.9721	-1.1204	0.3217
$AIC(1, r_2)$	$AIC(2, r_2)$	$AIC(3, r_2)$	$AIC(4, r_2)$	$AIC(5, r_2)$
-4.6982	-3.7624	-2.2227	-1.7691	0.7156
$AIC(1, r_3)$	$AIC(2, r_3)$	$AIC(3, r_3)$	$AIC(4, r_3)$	$AIC(5, r_3)$
-4.1798	-3.6617	-2.6603	-1.5454	0.0147
$AIC(1, r_4)$	$AIC(2, r_4)$	$AIC(3, r_4)$	$AIC(4, r_4)$	$AIC(5, r_4)$
-4.1142	-3.7744	-2.1476	-1.0249	0.6725

比較上列各 AIC 值發現 $(d, r)=(1, 29.495)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到左端點的門檻自迴歸模式如下：

$$a_t = \begin{cases} 0.7167a_{t-1} = 0.0034 + \varepsilon_t & \text{if } a_{t-1} < 29.495 \\ 0.9241a_{t-1} = 0.0073 + \varepsilon_t & \text{if } a_{t-1} \geq 29.495 \end{cases}$$

右端點之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自

迴歸模式再加總之：

表 4.6：以不同的門檻值與延誤參數計算右端點之 AIC 值

$\{r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
$AIC(1, r_1)$	$AIC(2, r_1)$	$AIC(3, r_1)$	$AIC(4, r_1)$	$AIC(5, r_1)$
-4.2671	-3.1674	-2.0198	-1.0463	0.0176
$AIC(1, r_2)$	$AIC(2, r_2)$	$AIC(3, r_2)$	$AIC(4, r_2)$	$AIC(5, r_2)$
-4.7843	-3.1845	-2.6179	-1.0171	0.9736
$AIC(1, r_3)$	$AIC(2, r_3)$	$AIC(3, r_3)$	$AIC(4, r_3)$	$AIC(5, r_3)$
-4.1836	-3.0548	-2.3667	-1.0369	0.3047
$AIC(1, r_4)$	$AIC(2, r_4)$	$AIC(3, r_4)$	$AIC(4, r_4)$	$AIC(5, r_4)$
-4.1873	-3.9644	-2.0112	-1.0176	0.9916

比較上列各 AIC 值發現 $(d, r)=(1, 29.617)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到左端點的門檻自迴歸模式如下：

$$b_t = \begin{cases} 0.6347b_{t-1} = 0.0012 + \varepsilon_t & \text{if } b_{t-1} < 29.617 \\ 0.7145b_{t-1} = 0.0041 + \varepsilon_t & \text{if } b_{t-1} \geq 29.617 \end{cases}$$

左右端點法之 SETAR 模式：

綜合以上左端點及右端點之 SETAR 模式合成區間型 SETAR 模式如下：

$$[a_t, b_t] = \begin{cases} a_t = \begin{cases} 0.7167a_{t-1} = 0.0034 + \varepsilon_t & \text{if } a_{t-1} < 29.495 \\ 0.9241a_{t-1} = 0.0073 + \varepsilon_t & \text{if } a_{t-1} \geq 29.495 \end{cases} \\ b_t = \begin{cases} 0.6347b_{t-1} = 0.0012 + \varepsilon_t & \text{if } b_{t-1} < 29.617 \\ 0.7145b_{t-1} = 0.0041 + \varepsilon_t & \text{if } b_{t-1} \geq 29.617 \end{cases} \end{cases}$$

以中心點半徑法建構

中心點之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 4.7：以不同的門檻值與延誤參數計算中心點之 AIC 值

$\{r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
$AIC(1, r_1)$	$AIC(2, r_1)$	$AIC(3, r_1)$	$AIC(4, r_1)$	$AIC(5, r_1)$
-3.7261	-2.0184	-1.0544	-0.3147	0.0041
$AIC(1, r_2)$	$AIC(2, r_2)$	$AIC(3, r_2)$	$AIC(4, r_2)$	$AIC(5, r_2)$
-3.6174	-2.3016	-1.6154	-0.2168	0.2716
$AIC(1, r_3)$	$AIC(2, r_3)$	$AIC(3, r_3)$	$AIC(4, r_3)$	$AIC(5, r_3)$
-3.4651	-2.0174	-1.7459	-0.9863	0.9862
$AIC(1, r_4)$	$AIC(2, r_4)$	$AIC(3, r_4)$	$AIC(4, r_4)$	$AIC(5, r_4)$
-3.1264	-2.3967	-1.3349	-0.8741	0.1145

比較上列各 AIC 值發現 $(d, r)=(1, 29.5775)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到中心點的門檻自迴歸模式如下：

$$c_t : \begin{cases} 0.7843c_{t-1} = 0.0167 + \varepsilon_t & \text{if } c_{t-1} < 29.5775 \\ 0.6516c_{t-1} = 0.0037 + \varepsilon_t & \text{if } c_{t-1} \geq 29.5775 \end{cases}$$

半徑之 SETAR 模式：

首先將所有的觀察值由小至大依序排列，利用上節中所找出的門檻候選值 $r=r_1, r_2, r_3, r_4$ 配合 $d=1, 2, 3, 4, 5$ 將樣本分為兩個不同的系統，再以 AIC 值為參

數選擇的依據，分別配適門檻值前後兩段樣本，找出使兩段的 AIC 值最小的自迴歸模式再加總之：

表 4.8：以不同的門檻值與延誤參數計算半徑之 AIC 值

$\{r_1, r_2, r_3, r_4\} \times \{d=1, 2, 3, 4, 5\}$				
$AIC(1, r_1)$	$AIC(2, r_1)$	$AIC(3, r_1)$	$AIC(4, r_1)$	$AIC(5, r_1)$
-6.1644	-6.2115	-5.3216	-4.3012	-4.0031
$AIC(1, r_2)$	$AIC(2, r_2)$	$AIC(3, r_2)$	$AIC(4, r_2)$	$AIC(5, r_2)$
-6.8947	-7.3462	-5.3044	-4.69444	-3.1161
$AIC(1, r_3)$	$AIC(2, r_3)$	$AIC(3, r_3)$	$AIC(4, r_3)$	$AIC(5, r_3)$
-7.4721	-6.6465	-5.9367	-4.6327	-3.1544
$AIC(1, r_4)$	$AIC(2, r_4)$	$AIC(3, r_4)$	$AIC(4, r_4)$	$AIC(5, r_4)$
-7.6271	-7.4567	-5.3217	-4.9125	-3.1749

比較上列各 AIC 值發現 $(d, r)=(1, 0.0525)$ 時為最小，我們以其為最終的延誤參數和門檻值，可以得到半徑的門檻自迴歸模式如下：

$$r_t : \begin{cases} 0.0816 r_{t-1} = 0.0836 + \varepsilon_t & \text{if } r_{t-1} < 0.0525 \\ 0.0071 r_{t-1} = 0.0016 + \varepsilon_t & \text{if } r_{t-1} \geq 0.0525 \end{cases}$$

中心點半徑法之區間型 SETAR 模式：

綜合以上中心點及半徑之 SETAR 模式合成區間型 SETAR 模式如下：

$$(c_t, r_t) = \begin{cases} c_t = \begin{cases} 0.7843c_{t-1} = 0.0167 + \varepsilon_t & \text{if } c_{t-1} < 29.5775 \\ 0.6516c_{t-1} = 0.0037 + \varepsilon_t & \text{if } c_{t-1} \geq 29.5775 \end{cases} \\ r_t = \begin{cases} 0.0816 r_{t-1} = 0.0836 + \varepsilon_t & \text{if } r_{t-1} < 0.0525 \\ 0.0071 r_{t-1} = 0.0016 + \varepsilon_t & \text{if } r_{t-1} \geq 0.0525 \end{cases} \end{cases}$$

2. 區間型 TAR 模式

全球第三大經濟體－日本，於 2011 年 3 月 11 日時間 13:46 發生芮氏規模 9.0 前所未見的強烈地震，並引發巨大海嘯浪潮，除造成無數重大災情外，更引發了核電廠爆炸幅射外洩事件，導致當日東京日經指數(Nikkei225)重挫 1,015.3 點，創下歷年有史以來第三大跌幅，亞洲股價幾乎呈現一面倒的趨勢，臺灣加權股價指數(TWII)亦慘跌了 285.24 點，雖然日本政府為穩定匯率，緊急投注 12 兆日幣穩住日圓水準，但世界各國由於預期心理，紛紛大舉出脫日股或日幣，對於身為鄰近國家的臺灣造成了不小的影響(NHK, TSEC)，因此我們以此為重大歷史事件的時間轉折點，在此時點前後分別以 AIC 準則配適自迴歸模式，建構左右端點法及中心點半徑之區間型 TAR 模式。

以左右端點法建構

左端點之 TAR 模式：

$$a_t = \begin{cases} 0.8212a_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.3241a_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \quad t \text{ 為期數}$$

右端點之 TAR 模式：

$$b_t = \begin{cases} 0.4133b_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.9812b_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \quad t \text{ 為期數}$$

左右端點法之區間 TAR 模式：

綜合以上左端點及右端點之 TAR 模式合成區間型 TAR 模式如下：

$$[a_t, b_t] = \begin{cases} a_t = \begin{cases} 0.8212a_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.3241a_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \\ b_t = \begin{cases} 0.4133b_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.9812b_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \end{cases} \quad t \text{ 為期數}$$

以中心點半徑法建構

中心點之 TAR 模式：

$$c_t : \begin{cases} 0.6617c_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.3211c_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \quad t \text{ 為期數}$$

半徑之 TAR 模式：

$$r_t = \begin{cases} 0.0021r_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.0046r_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \quad t \text{ 為期數}$$

中心點半徑法之區間 TAR 模式：

綜合以上中心點及半徑之 TAR 模式合成區間型 TAR 模式如下：

$$(c_t, r_t) = \begin{cases} c_t = \begin{cases} 0.6617c_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.3211c_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \\ r_t = \begin{cases} 0.0021r_{t-1} + \varepsilon_t & \text{if } t < 17 \\ 0.0046r_{t-1} + \varepsilon_t & \text{if } t \geq 17 \end{cases} \end{cases} \quad t \text{ 為期數}$$

4.5 預測結果比較與分析

我們將 2.4 節中所定義的區間平均誤差 ϵ 此作為模式評估依據，分別對所建構的區間時間數列進行分析，詳如下列各表：

表 4.9 以 IME、IMPE 評估左右端點法之預測結果

保留 期數	實際區間	左右端點法		
		區間 ARIMA	區間 SETAR	區間 TAR
3/29	[29.469, 29.539]	[29.621, 30.617]	[29.331, 29.636]	[29.315, 29.634]
3/30	[29.430, 29.510]	[28.301, 29.889]	[29.681, 29.624]	[29.671, 29.368]
3/31	[29.395, 29.510]	[29.361, 29.631]	[29.634, 29.671]	[29.614, 29.634]
IME		0.237	0.054	0.087
IMPE		12.38%	5.36%	6.12%

表 4.10 以 IME、IMPE 評估中心點半徑法之預測結果

保留 期數	實際區間	中心點半徑法		
		區間 ARIMA	區間 SETAR	區間 TAR
3/29	(29.5040;0.0350)	(29.6714;0.0476)	(29.4751;0.0041)	(29.6315;0.0720)
3/30	(29.4700;0.0400)	(29.1473;0.0432)	(29.5671;0.0132)	(29.6712;0.0630)
3/31	(29.4525;0.0575)	(29.6712;0.0627)	(29.6127;0.0474)	(29.3716;0.0887)
IME		0.324	0.012	0.074
IMPE		14.65%	5.89%	6.42%

由以上各表中的比較，我們可以藉由 IME、IMPE 為模式優劣比較工具，發現中心點半徑法的預測結果與左右端點法差不多，而三種模式之預測能力依序為區間 SETAR > 區間 TAR > 區間 ARIMA，表示兩種門檻自迴歸模式都比單純的區間 ARIMA 法好得多，這正符合我們的假設，表示若能找出發生結構改變之處，對於掌握資料走勢或是做進一步的預測分析都會有較好的結果。

第五章 結論

古典拓樸學中「點」只是一種「區間」的特例，但在以往發展出的各種時間數列模式中，不論是線性或者非線性大多都以單時點的資料為主，文獻中對於區間型式時間數列則較少探討，而且由於目前社會現象多變、人類情感複雜，簡單的線性模式往往無法符合人類現實生活的需求，因此本研究中提出了一個有別於傳統時間數列的方法，以門檻自迴歸模式來配適區間資料，建構出一套非線性型式的區間時間數列，將點預測中門檻轉折點的概念推廣到區間模式的建構上，如此一來，除了保留了本身區間的模糊性，亦能呈現出資料中結構改變的特性，同時也使得一般單點值的門檻自迴歸成為本模式的一種特例，不但包含了轉折的精神，也將樣本做了更貼切的詮釋，適用範圍則更加的廣泛。

雖然過去已有專家學著針對「來臺觀光客人數」與「新臺幣兌美元匯率」提出研究，然而以非線性的區間模式來預測並不多見，因此本文透過建立兩者之區間資料，分別以非線性的門檻自迴歸以及傳統線性 ARIMA 模式來比較，最後以模糊理論為基礎，計算區間誤差進行預測效度評估比較，研究結果我們可以發現在三種模式的預測中以「左右端點」或以「中心點半徑」兩種區間定義法之效果差異並不大，但是在預測結果上「非線性」確實優於「線性」，而且對於資料呈現向上增加趨勢的「來臺觀光客人數」而言，使用以「時間點」分段的 TAR 模式較佳；然而對於資料呈現上下擺盪趨勢的「新臺幣兌美元匯率」而言，使用以「變數」區分的 SETAR 模式較佳，不過就整體而言，若我們以 IMPE 為評估標準比較可以發現，在三種模式中「來臺觀光客人數」的預測皆較「新臺幣兌美元匯率」準確，我們歸咎其原因，或許是因為近年來臺灣並未發生重大的天災或疫情，再加上政府的大力推廣觀光產業，使得「來臺觀光客人數」呈現持續向上成長的趨勢，其走勢顯得較為單純，然而相對於觀光客人數的穩定，台幣的升貶與強弱則受到其它國家的匯率、國際投資熱錢流向、外銷競爭力與外匯存底餘額等諸多因素的影響，造成「新臺幣兌美元匯率」的走勢相對地顯得十分地波動，太

多可能的變數增加了預測上的誤差，因此未來若能考慮與其有關的變項，或許會有更準確的預測結果。總之若我們要能精確預測區間資料未來的變化，就必需掌握資料的屬性，考慮其相關的變數，找出其中可能的改變轉折之處，並選擇恰當的配適模式，如此才能有效的預測其未來的走向。

隨著全球經濟開放，國際間貿易頻繁，使的各國間匯率波動幅度甚巨，對於各層面產生的連動效應，造成之影響十分地廣泛，若政府或企業能掌握其未來走勢，則對於促進國家經濟穩定或規避投資風險將有正面的助益，而台灣因為本身具有豐富的自然資源及多樣化人文景觀，再加上處於亞太地區的交通樞紐位置，國外旅客來台觀光或在台轉機相當的便利及頻繁，若能充分把握契機，透過觀光客人數的預測，政府才能有效的完成相關規劃，使旅遊市場的供需達到平衡，創造可觀的經濟效益。

未來可能研究方向與建議：

- 一、本研究中的區間資料皆假設為均勻分配，中心點定為左、右端點平均值，然而實際情形應該並非如此單純，未來若能考量每個區間中的不同分配，如常態分配、指數分配等，將其納入預測模式當中，相信應該能更貼近區間資料真實的變化。
- 二、本文假設左、右端點之間、中心點、半徑之間互為獨立，並未考量其間可能的相關性，若將兩者間的關係納入考慮，則模式該如何建構？
- 三、除了門檻自迴歸模式，目前還有許多非線性的方法，如雙線性模式、指數型自迴歸模式等等，嘗試其它的預測法或許會有不同的結果。
- 四、「觀光客人數」與「匯率」兩者之間是否具有關聯性？能否以轉換函數模式 (transfer function model) 建構兩個區間資料之間的預測模式？
- 五、生活中有許多的區間資料都可能具有結構改變的特性，未來或許能將本研究模式應用在其它的領域上，如天氣預報、股價預測等等。

參考文獻

中文部分：

- [1]. 交通部觀光局 (2002)。觀光客倍增計畫。
- [2]. 交通部觀光局 (2002)。觀光政策白皮書。
- [3]. 交通部觀光局 (2007)。旅遊台灣年計畫。
- [4]. 交通部觀光局 (2010)。中華民國年觀光年報。
- [5]. 吳柏林 (1995) 時間數列分析與導論。台北：華泰書局。
- [6]. 吳柏林、張建瑋 (1996)。非線性時間數列的分類與預測。第三屆三軍官校基礎學術研討會論文集 98-214。
- [7]. 吳柏林 (2000)。模糊數學與統計應用。台北：俊傑書局。
- [8]. 吳柏林 (2005)。模糊統計導論:方法與應用。台北：五南出版社。
- [9]. 沈中華 (2000)。40 分鐘學會匯率危機預測。台北：新陸書局。
- [10]. 李榮謙 (1999)。國際貨幣與金融。台北：智勝文化。
- [11]. 阮正治 (1996)。遺傳演算法在非線性時間數列結構改變之分析與應用。國立政治大學統計系碩士論文。
- [12]. 林茂文 (1992)。時間序列分析與預測。台北：華泰書局。
- [13]. 林原宏 (2006)。模糊統計。台北：五南出版社。
- [14]. 程友梅 (1995)。轉折型時間序列的認定。國立政治大學統計系碩士論文。
- [15]. 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。
- [16]. 楊奕農 (2006)。時間序列分析-經濟與財務上之應用。台北：雙葉書廊。
- [17]. 賈昭南 (2002)。國際金融實務與理論。台北：華泰文化。

英文部分：

- [1]. Chang, S.K. (2007). On the Testing Hypotheses of Mean and Variance for Interval Data. *Management Science and Statistical Decision* 4(2), 63-69.
- [2]. Chatfield, C. (1993). Calculating Interval Forecasts. *Journal and Business & Economic Statistics* 11(2), 121-135.
- [3]. Chen, S.M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems* 81, 311-319.
- [4]. Huarng, K. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy sets and systems* 123(3), 387-394.
- [5]. Hsu, H.L. (2008). Evaluating forecasting performance for interval data. *Computers and Mathematics with Applications* 56, 2155-2163.
- [6]. Kashia, M., Hejaz, S.R. and Bijari, M. (2008). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy sets and systems* 159, 769-786.
- [7]. Kreinovich, V., Nguyen, H.T. and Wu, B. (2007). On-line algorithms for computing mean and variance of interval data, and their use in intelligent systems. *Information Sciences* 177, 3228-3238.
- [8]. Ludermit, T.B. (2008). Forecasting models for interval-valued time series. *Neurocomputing* 71, 3344-3352.
- [9]. Nguyen, H.T. and Wu, B. (2006). *Fundamentals of Statistics with Fuzzy Data*. New York : Springer.
- [10]. Römer, C. and Kandel, A. (2000). Statistical tests for fuzzy data. *Fuzzy sets and systems* 72(1), 1-26.
- [11]. Tong, R.M. (1978). Synthesis of fuzzy models for industrial processes. *International Journal of General Systems* 5(4), 143-162.

- [12]. Tsay, R.S. (1991). Detecting and modeling non-linearity in univariate time series Analysis. Statistica Sinica 1(2), 431-451.
- [13]. 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.
- [14]. Tseng, F.M. and Tseng, G.H. (2002). A fuzzy seasonal ARIMA model for forecasting. Fuzzy sets and systems 126(3), 367-376.

