

國立政治大學應用數學系

碩士學位論文

相對移動率應用在區間時間序列預測及

其效率評估

The Application of Relative Moving
Ratio for Forecasting and Performance
Evaluation in Interval Time Series

碩士班學生：李治陞 撰

指導教授：劉明郎 博士

吳柏林 博士

中華民國一百零一年七月三日

國立政治大學應用數學系

李治陞 君所撰之碩士班學位論文

相對移動率應用在區間時間序列預測及

其效率評估

The Application of Relative Moving
Ratio for Forecasting and Performance
Evaluation in Interval Time Series

業經本委員會審議通過

論文考試委員會委員：

指導教授：

系主任：

中華民國 101 年 7 月 3 日

謝辭

隨著論文的完成，學生生涯也將告一個段落。此時，要感謝的人太多，能讓我順利完成這一篇論文。

首先要感謝的便是我的指導教授劉明郎教授及吳柏林教授，若沒有吳老師提希與照顧，劉老師耐心的指導，逐字逐句的修改論文。這一篇論文是不可能完成的。再來，如果沒有丁媽媽在英文摘要上給予我很大的助力，坦白說，我根本寫不出來。另外，在研究室裡首先要感謝的便是亮哥及家盛，除了給予我在學問上很大的指導外，同時在百忙之中爆肝陪我玩世紀帝國讓我度過快樂的暑假。沛承，小哈、kinki 也陪我瘋狂的玩了許久的魔獸三國，再來就是感謝 yoyo 的貼心，能讓我有許多煩惱時，幫我處理許多雜物使我可以無後顧之憂。感謝班上許許多多的同學，認識大家真的太好了。

同時，還要感謝我母親給予我的支持與鼓勵。雖然囉嗦了點，但是如果沒有她，我不可能來政大唸書，更不可能畢業。

最後便是感謝智婷，給了我很多鼓勵與幫忙，在我忙得焦頭爛額時，很有耐心的陪在我身邊，從大學時期，便花了很多時間幫助我完成許多瘋狂的事情，及幫助我度過很多次的難關與挫折。

要感謝的人真的太多，短短一篇文章，是不可能表達對所有幫助過我的人有無上的感謝。

治陞

2012.7.23

相對移動率應用在區間時間序列預測及其效率評估

李治陞

摘要

時間序列是用來預測未來趨勢的一種重要技術，然而在實務上建構時間序列模型時，參數很難有效估計。原因可能來自於時間序列本身的模糊性質，而導致參數的不確定性使得預測結果產生極大誤差。如果將參數模糊化引進時間序列的模型中，往往過於複雜。本論文提出相對移動率為新的模糊時間序列建構方法，讓原本具有模糊性質的時間序列經由反模糊化(defuzzification)後，以點估計的方式估計起始中心點，經由適當的修正調整為較佳的中心點以及半徑，建立有效的區間時間序列。並將相對移動率引進門檻自迴規模型中，取代原有之門檻值設定，並建立區間時間序列。最後，我們使用台灣加權股價指數為例，以本論文所提出之方法進行區間預測及效率評估。

關鍵字：模糊時間序列、反模糊化、區間預測、相對移動率、門檻自迴規模型

The Application of Relative Moving Ratio for Forecasting and Performance Evaluation in Interval Time Series

Li, Chih-Sheng

Abstract

The time series is an important technology that is used to predict future trends, however in the real world, parameter is difficult to estimate effectively when we construct a time series model due to the of the fuzzy property of the times series data. The estimated parameters in the time series will cause a big error due to the uncertainty of fuzzy data. It is too complex to introduce the fuzzy parameters into the time series model. In this thesis, we propose *relative moving ratio* as a new criteria in constructing procedure of an interval time series. We defuzzify a fuzzy data and use point estimation to obtain an initial center, then we adjust the center and radius making it more appropriately. The resulting center and radius is then become an interval time series that can be use to forecast an interval data. We also apply relative moving ratio in threshold autoregressive models by replacing the threshold in constructing interval time series. Finally, in empirical studies chapter, we use Taiwan weighted Stock Index as examples to evaluate the performance of the proposed two methods in building the interval time series.

Keywords : fuzzy time series, defuzzification, interval prediction, relative moving ratio, threshold autoregressive models.

目錄

謝辭	iv
摘要	v
Abstract	vi
目錄	vii
第一章 緒論	1
第二章 模糊時間序列分析	3
2.1 序言	3
2.2 模糊時間序列模型	7
2.3 模糊時間序列模型之預測方法	9
第三章 模糊時間序列模型之效率評估	14
3.1 模糊區間之定位及測量辦法	14
3.2 模糊時間序列模型之預測指標	16
第四章 實證分析	18
第五章 結論	35
參考文獻	36

第一章 前言

時間數列的發展上，自從 Box 與 Jenkins 在 1976 年發展 ARIMA 以來，已經引起廣大學者的興趣。例如在時間數列分析導論(吳柏林，1994)中將 ARIMA 應用在空氣汙染、人口成長率、出生率、死亡率、就業率...等。但是若是用 ARIMA 分析隨機漫步上的時間數列資料，會因為假設條件太嚴苛，如資料的穩定性雖然可以經過差分解決，但是變異數和期望值的固定卻是無法解釋隨著時間改變而使得變異數改變的時間序列資料，所以非線性時間數列成為近來研究的重心。非線性的模式有許多種類，例如門檻自迴歸模式(threshold autoregressive models)(Tong, 1980)，雙線性模式(bilinear models, Subba-Rao 與 Gabr, 1980)、指數自迴歸模式(exponential autoregressive models)(Haggan, 1980)。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。但是，光是如此，仍然不能解決期預測準確性的問題，於是模糊時間序列便因此而誕生，因為模糊集合理論本身具有語言變數(linguistic variables)蘊含特性，這種特性可以減少在處理不確定性問題時可能造成的困擾，將模糊邏輯應用在時間序列的分析過程時，第一步是要考慮如何結合語言變數的分析法，以解決資料模糊性的問題。

Tong (1978)提出邏輯檢查法(logical examination method)利用決策表來描述模糊式，但是此方法很難擴展到多變數的系統。為了獲得更精確的模糊模式，Graham 與 Newell(1989)，提出了具有學習能力的方法去修正模糊模式，但此方法相當麻煩。因此大部分的學者，常採用模糊關係方程式求解來分析。Song 與 Chissom (1993)利用模糊關係方程式，提出詳盡的模糊間序列的建構過程及模式的理論架構。但是關於模糊時間序列階次的認定，到目前為止並未有一個有效的辦法，在以往的文獻中都是直接給定的，這方法太過於主觀，理論基礎也相當薄弱。因此預測結果往往會有極大誤差產生。所以，本論文將提出一種相對移動率方法，將模糊時間序列的資料反模糊化之後，希望能夠建立一個簡單的預測模型，能夠有更好的預測結果，同時也希望相對移動率能應用在門檻自迴規模型中。

在門檻自迴規模型中，「轉折」乃結構轉變的考量。當所採用的樣本期間較長或是樣本期間有重大的因素發生轉變，則考慮是否因為這些因素變動，而使得

過去的數據和現在的數據之間的關係發生變化 (楊奕農, 2009)。門檻自迴歸模式和結構轉變模式非常相似, 過去 Chow (1960) 在線性迴歸模型下, 檢定單一已知結構改變時間點的轉折是否顯著。由於 Chow (1960) 以「時間」為結構改變的轉折點, 主觀認定結構改變的時間, 方法上較不客觀, 且當模型中重要資料在短時間內持續發生大幅度變化時, 可能無法診斷出模型的結構改變。以「變數」當轉折點可避免上述缺失, 關於這方面的研究, Tong (1980) 首先提出以「變數」為結構轉變的轉折點之 TAR 模式。Kumar 與 Wu (2001) 發現在非線性的時間序列藉由模糊邏輯的概念可以有效找出結構轉折。Zhou (2005) 提出整合貝氏 (Bayesian) 行的結構轉變分析與轉折點的判斷方法。雖然眾多學者針對結構轉變做分析, 其研究結構轉變使用的方法不計其數, 但在數理的推論過程都相當的繁瑣且轉折點的定義似乎仍無一明確的標準, 因此本論文將以相對移動率, 當轉折點進行 TAR 模式預測, 採用「多變數」, 非單一變數, 希望能夠得到更有效率的預測結果。

由於模糊時間序列的模型的計算往往過於複雜, 因此本文希望能透過返模糊化的過程中, 建構一個簡單但是有效的預測模糊時間序列模型, 並且把新的模型中的參數套入門檻自回歸模型中, 比較其有效性。

本論文在第二章從何謂模糊理論開始做一個簡單的介紹, 進而給一個新的模糊區間預測模型的新方法, 並把這種方法和門檻自回歸模型做一個整合。第三章從兩組區間資料的距離開始研究, 給一個效率性的分析。第四章是蒐集從民國一百年六月一號到十二月三十一號的開盤價及收盤價後, 用第二章所提出的方法來檢討其效率性。第五章則是總結。

第二章 模糊時間序列分析

模糊集合理論是在 1965 年由美國加州大學柏克萊分校 Zadeh 教授提出的。他認為以模糊集合理論為基礎，可以用數學的方法來解決生活中原本無法解決的模糊現象。現在，模糊理論已經被應用到許多的領域。如：模糊時間序列(fuzzy time series)、模糊類神經網路(fuzzy neural network)模糊控制(fuzzy control)、模糊邏輯(fuzzy logic)等，以下先簡單介紹模糊集合理論。

第一節 序言

如果要做模糊時間序列的預測，必須要有幾樣工具。我們要先弄清楚什麼是模糊理論、隸屬度函數、模糊時間數列及區間模糊時間數列。在這一節中，我們首先要先探討什麼是模糊理論，並且給隸屬度函數一些簡單的介紹外，再給模糊時間數列及區間模糊時間數列一些定義。

人類的思維只要是來自於自然現象和社會現象的認知意識，所以，人類的知識語言會因為本身的主觀意識和所接收的資訊不同而具有模糊性。例如，當有人說他今天心情好時，他對心情好的認知為何？要怎麼樣測量呢？如對人的好感度描述等。於是為了描述事物特徵的模糊性進而發展了模糊集合理論。

模糊理論的概念，主要在強調個人喜好程度不需要非常清晰或數值精確，因此對人類而言，模糊模式比直接指定單一物體的一個值，較適合評估物體間的多元性。在這種情況下時，“非此即彼”的二元邏輯是無法界定的。在此，模糊集合理論為描述這類邊界不清的事物提供了一套有效的方法，使人們有系統化、公式化的手段處理這類事物的能力，能夠在模糊環境中解決問題，做出正確的決策。

在傳統集合理論中有個重要的概念，即特徵函數(Characteristic function)。對 A 的子集合 A 而言，定義 $\chi_A(x)$ 為 A 集合的特徵函數

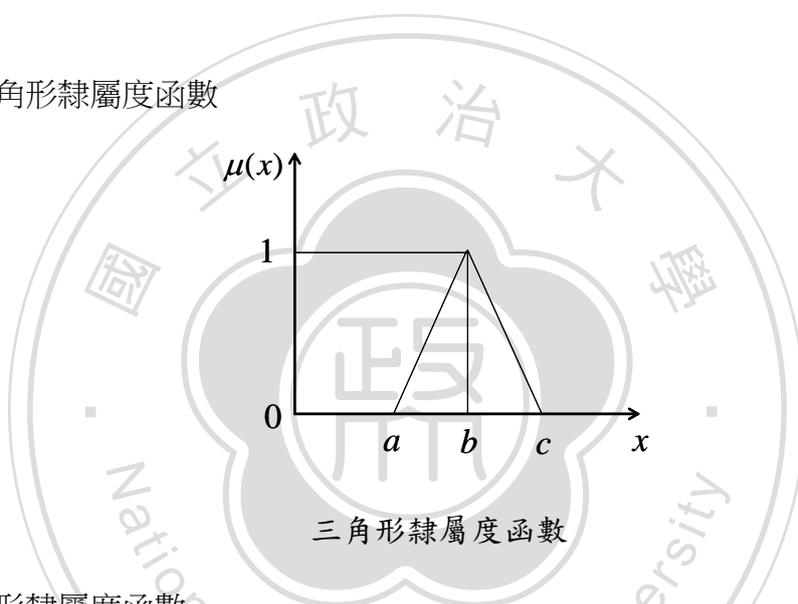
$$\chi_A(x) = \begin{cases} 1 & \text{若 } x \in A \\ 0 & \text{若 } x \notin A \end{cases} \quad (2.1)$$

如上公式(2.1)所述，若 x 為 A 的元素，則其特徵函數 $\chi_A(x) = 1$ ；反之若 x 不為 A 的元素，則其特徵函數為 $\chi_A(x) = 0$ 。其值域可表示成 $\chi_A(x) = \{0, 1\}$ 。

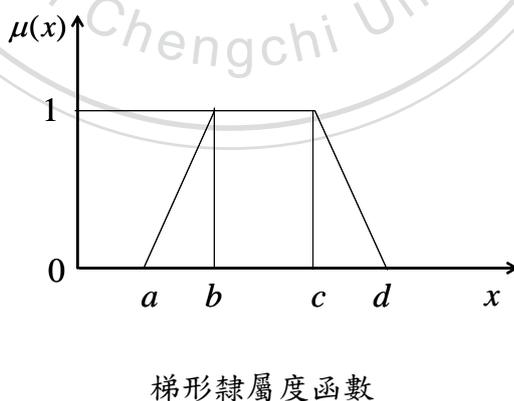
Zadeh (1965)教授將傳統集合理論特徵函數從二元運算，推廣為可從 0 到 1 之間的任何值作選擇，這種特徵函數，稱為隸屬度函數(membership function)。隸屬度函數是模糊理論中最基本的概念，他不僅可以描述模糊集合的性質，更可以對模糊集合進行量化，並利用精確的數學方法分析和處理人類模糊性資訊。隸屬度函數可以分為離散型(discrete type)和連續型(continuous type)兩種，離散型隸屬度函數是直接給予有限模糊集合中每一個元素的隸屬度，且用向量的型式表示。連續型隸屬度函數一般常用的有三角形、梯形、高斯函數、不規則形…等等。

函數定義的表現，可以是無限模糊集合的元素及其隸屬度之間的關係，也可以是有限模糊集合的元素和隸屬度之間的關係。常見的連續型隸屬度函數圖形如下：

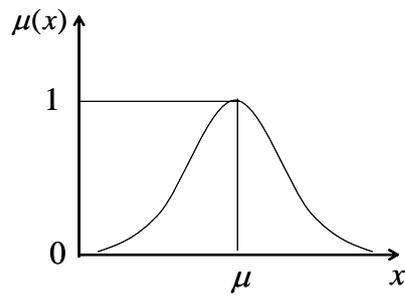
1. 三角形隸屬度函數



2. 梯形隸屬度函數

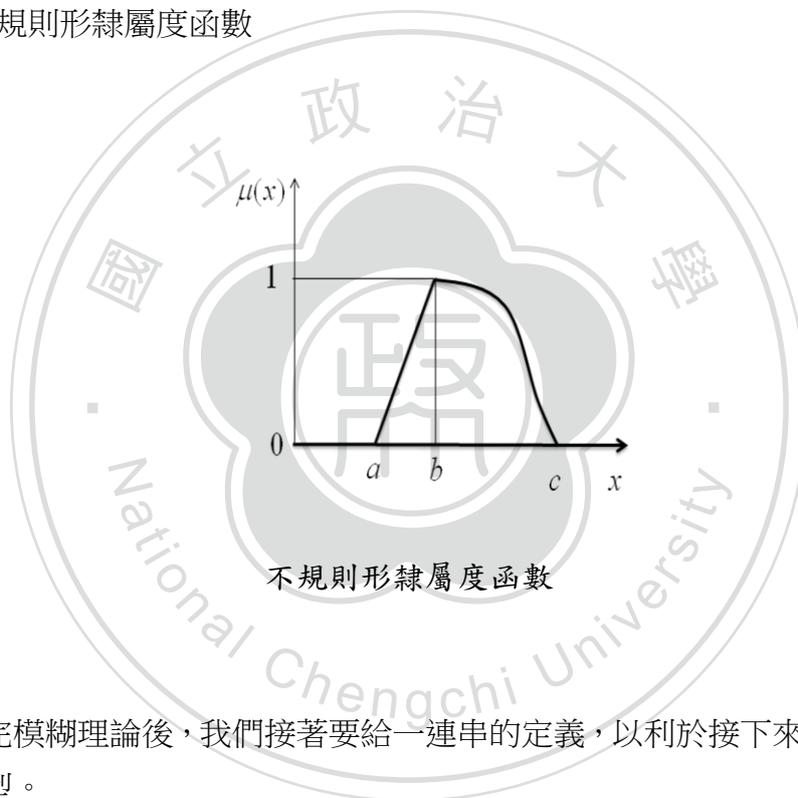


3. 高斯隸屬度函數



高斯隸屬度函數

4. 不規則形隸屬度函數



不規則形隸屬度函數

介紹完模糊理論後，我們接著要給一連串的定義，以利於接下來建立區間時間序列模型。

定義 2.1 (Wu, 2006)

$X = [a, b]$ 若 a 和 b 是隨機變數，則區間 $[a, b]$ 為模糊隨機區間。

定義 2.2 (Wu, 2006)

若 $X = [a, b]$ 為一個布於實數的模糊隨機區間， $c = \left(\frac{a+b}{2}\right)$ 為模糊隨機區間

X 之圓心，且 $r = (\frac{b-a}{2})$ 為模糊隨機區間 X 之半徑，則模糊隨機區間 X 可以表示為 $X = (c; r)$ ，其中 a, b, c 皆為實數。

例 2.1

令 $X = [1, 5]$ 為一個布於實數的模糊隨機區間，則 $X = [1, 5] = (3; 2)$ 。 □

定義 2.3 (Wu, 2006)

令 $X = (c; r)$ 布於實數 R 之隨機區間，則模糊隨機區間 X 之期望值定義為 $E[X] = E(c; r) = (E(c); E(r))$ ，其中 $E(c)$ 和 $E(r)$ 是為隨機變數 c 和 r 的期望值。

定義 2.4 (Wu, 2006)

令 $\{X_t = [a_t, b_t] = (c_t; r_t), t = 1, 2, \dots, n\}$ 為一模糊時間數列，因 X_t 皆為區間模糊數則稱 $\{X_t\}_{t=1}^n$ 為一區間模糊時間數列。

定義 2.5 (Song 與 Chissom, 1993)

令 $Y(t) (t = \dots, 0, 1, 2, \dots)$ 為一個論域且為 R 的部分集合，且令 $f_i(t) (i = 1, 2, \dots)$ 為定義在論域 $Y(t)$ 中之一個模糊集合。令 $F(t)$ 為 $f_i(t) (i = 1, 2, \dots)$ ，則 $F(t)$ 為 $Y(t) (t = \dots, 0, 1, 2, \dots)$ 的一個模糊時間序列。

定義 2.6 (Song 與 Chissom, 1993)

令 $F(t)$ 為模糊時間序列，且存在一個模糊關係 (fuzzy relation) $R(t, t-1)$ ，使得 $F(t) = F(t-1) \otimes R(t, t-1)$ ，其中 \otimes 是個運算子。

定義 2.7 (Song 與 Chissom, 1993)

令 $F(t)$ 為模糊時間序列，且令 $F(t)$ 由 $F(t-1), F(t-2), \dots, F(t-n)$ 導的，則其 n 階模糊邏輯關係可被表示為 $F(t-n), \dots, F(t-2), F(t-1) \rightarrow F(t)$ 其被稱為單因子 n 階模糊時間序列(one-factor n -order fuzzy time series)預測模式。

例 2.4

令 X_t 為一區間時間序列且 $X_t = [t-2, t+2] = (t; 2)$ ，則 $\{X_t = [t-2, t+2] = (t; 2), t = 1, 2, \dots, n\} = \{[-1, 3], [0, 4], [1, 5], \dots, [n-2, n+2]\}$ 為一個區間模糊時間數列。 □

第二節 模糊時間序列模型

Song 與 Chissom (1993)根據模糊集合理論首先提出模糊時間序列的定義。因為以此方法預測美國阿拉巴馬州的學年度學生註冊人數需要非常大量且複雜的運算，所以 Chen (1996)提出了一個簡單的模糊時間序列預測模式來預測學生註冊人數，得到很好的預測效果。而 Chen (2002)則又提出如何利用高階模糊時間序列(high-order fuzzy time series)之預測方法用以預測學生註冊人數。進行預測時往前看前 n 年的資料，產生高階模糊邏輯關係(high-order fuzzy logical relationship)產生更高的參考價值且預測值也更準確。

Tseng 與 Tzeng (1999)將模糊理論與傳統的 ARIMA 模式結合，提出 Fuzzy ARIMA 模式，並針對具季節性循環之時間序列，提出 Fuzzy Seasonal ARIMA 模式，對季節性的時間資料，提供了不同的預測方式，並與傳統的 Seasonal ARIMA 模式做比較後，確實得到了更精準的預測值。

吳柏林與林玉鈞(2002)則提出了模糊時間數列的屬性預測，作者認為資料的收集會因為時間的延遲及變數之間的交互影響，使得看起來是單一的數值，實際上可能表達的是一個區間範圍的可能值，故其明確的建立模糊規則資料庫與模糊

關係，以人工智慧的方式將模糊數值轉換，並利用所建立的演算方法對台灣加權股票日指數，建構預測模式。

傳統常用的時間數列方法如 AR，MR，ARMR，ARIMA 等等，是點對點的預測，但區間時間數列無法以傳統的方法來預測，所以下列提供幾種常見的區間時間數列的預測模型。

1. 中心點及半徑 ARMR 法(Hsu 2008)

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，且

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

$$r_t = \eta + \phi_1 r_{t-1} + \dots + \phi_{pr} r_{t-pr} + \varepsilon_t - \eta_1 \varepsilon_{t-1} - \dots - \eta_{qc} \varepsilon_{t-qr}$$

其中 $\varepsilon_t \sim N(0, \sigma^2)$ ，則

$$E(c_t | c_{t-1}, c_{t-2}, \dots, c_1) = \theta + \phi_1 c_{t-1} + \dots + \phi_{pc} c_{t-pc}$$

$$E(r_t | r_{t-1}, r_{t-2}, \dots, r_1) = \eta + \phi_1 r_{t-1} + \dots + \phi_{pr} r_{t-pr}$$

且令 $\hat{c}_t = E(c_t | c_{t-1}, c_{t-2}, \dots, c_1)$ ， $\hat{r}_t = E(r_t | r_{t-1}, r_{t-2}, \dots, r_1)$ ，則區間時間數列的預測為

$$E(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}) = [\hat{c}_t - \hat{r}_t; \hat{c}_t + \hat{r}_t] = (\hat{c}_t; \hat{r}_t)$$

2. 中心點及半徑 k 階區間移動平均法 (Hsu, 2008)

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，且

$$\hat{c}_t = \frac{c_{t-1} + \dots + c_{t-k}}{k}, \quad \hat{r}_t = \frac{r_{t-1} + \dots + r_{t-k}}{k}, \quad t = k+1, k+2, \dots$$

則區間時間數列的預測值為

$$E\{X_t | X_{t-1}, X_t, \dots, X_k\} = [\hat{c}_t - \hat{r}_t; \hat{c}_t + \hat{r}_t] = (\hat{c}_t; \hat{r}_t)$$

3. 左右端點 ARMAR 法(Hsu, 2008)

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，且

$$a_t = \theta + \phi_1 c_{t-1} + \dots + \phi_{pc} c_{t-pc} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \phi_{pc} \varepsilon_{t-pc}$$

$$b_t = \eta + \varphi_1 r_{t-1} + \dots + \varphi_{pr} r_{t-pr} + \varepsilon_t - \eta_1 \varepsilon_{t-1} - \dots - \eta_{qc} \varepsilon_{t-qr}$$

其中 $\varepsilon_t \sim WN(0, \sigma^2)$ ，則

$$E(a_t | a_{t-1}, a_{t-2}, \dots, a_1) = \theta + \phi_1 a_{t-1} + \dots + \phi_{pc} a_{t-pc}$$

$$E(b_t | b_{t-1}, b_{t-2}, \dots, b_1) = \eta + \varphi_1 b_{t-1} + \dots + \varphi_{pr} r_{t-pr}$$

且令 $\hat{a}_t = E(a_t | a_{t-1}, a_{t-2}, \dots, a_1)$, $\hat{b}_t = E(b_t | b_{t-1}, b_{t-2}, \dots, b_1)$ ，則區間時間數列的預測為

$$E(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}) = [\hat{a}_t, \hat{b}_t]$$

4. 左右端點 k 階區間移動平均法(Hsu, 2008)

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，且

$$\hat{a}_t = \frac{(a_{t-1} + a_{t-2} + \dots + a_{t-k})}{k}$$

$$\hat{b}_t = \frac{(b_{t-1} + b_{t-2} + \dots + b_{t-k})}{k}$$

其中 $t = k+1, k+2, \dots$ ，則區間時間數列的預測值為

$$E(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}) = [\hat{a}_t, \hat{b}_t]$$

第三節 模糊時間序列模型的預測方法

在這一節中，我們要利用移動平均值去尋找出中心點，並決定適當的半徑。因為如果我們單純的把 $\frac{(\max\{p_t\} + \min\{p_t\})}{2}$ 當成區間預估的圓心，這樣並不是最

好的。所以這一節我們要將圓心調整到適當的位子。要想調整圓心，首先要定義相對移動率。

定義 2.8：相對移動率(relative moving ratio)

令 $X_t = (c_t; r_t)$ 是一個區間時間序列， $\bar{r} = \frac{\sum_{k=0}^i r_{t-k}}{i}$ 為半徑，則第 i 階對第 j 階之相對移動率定義為 $\alpha_t = \frac{m_{t_i}}{m_{t_j}} - 1$ ，其中 $m_{t_i} = \frac{1}{i} \sum_{k=0}^i c_{t-k}$ 是以 c_t 為圓心之第 i 階之移動平均且 $0 < i < j$ 。

例 2.2.1：令 $\{ X_1 = [0, 2], X_2 = [1, 4], X_3 = [2, 7], X_4 = [3, 4], X_5 = [5, 8] \}$ 為一區間時間序列，則第 3 階對第 5 階之相對移動率為

$$\alpha_t = \frac{m_3}{m_5} - 1 = \frac{4.83}{3.6} - 1 = 0.342 \quad \square$$

接下來便利用相對移動率及中心點及半徑 k 階區間移動平均法(Hsu 2008)提出一種新的區間預估方法。本論文將移動平均數轉換成相對移動率之後，提出一個模糊時間數列的預測方法，在上述中，我們已經定義相對移動率，令參數 α_t 為相對移動率，當 $\alpha_t \geq 0$ 時，我們將第 $t+1$ 期 \hat{c} 上移至 \tilde{c} ，反之亦然，如公式(2.2)所表示。

$$\tilde{c} = \begin{cases} \hat{c}_t + \hat{r}_t & \text{若 } \alpha_t \geq 0 \\ \hat{c}_t - \hat{r}_t & \text{若 } \alpha_t < 0 \end{cases} \quad (2.2)$$

利用公式(2.2)，我們可以得到一個預測的模糊時間區間

$$X_{t+1} = (\tilde{c}; \hat{r}_t)$$

我們將上述及定義 2.8 整合之後就可以得到模糊區間預測模型。

定義 2.9 (相對移動率區間預測法)

令 $X_t = (c_t; r_t)$ 是一個區間時間序列， α_t 為相對移動率，
 $\hat{c}_t = \frac{c_{t-1} + \dots + c_{t-k}}{k}$ ， $\hat{r}_t = \frac{r_{t-1} + \dots + r_{t-k}}{k}$ ， $t = k+1, k+2, \dots$ ，若 $\alpha_t \geq 0$ 則模糊區間
 時間序列的預測結果為 $E\{X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}\} = [\hat{c}_t; \hat{c}_t + 2\hat{r}_t] = (\tilde{c}; \hat{r}_t)$ ，若 $\alpha_t < 0$
 則模糊區間時間序列的預測結果為 $E\{X_t | X_{t-1}, X_t, \dots, X_k\} = [\hat{c}_t - 2\hat{r}_t; \hat{c}_t] = (\tilde{c}; \hat{r}_t)$ 。

例 2.3.1

令時間序列 $\{X_t = [1, 2], [2, 4], [3, 4], [4, 6], [3, 7]\}$ ，其中 $X_1 = [1, 2] = (1.5; 0.5)$ ，
 $X_2 = [2, 4] = (3; 1)$ ， $X_3 = [3, 4] = (3.5; 0.5)$ ， $X_4 = [4, 6] = (5; 1)$ ，
 $X_5 = [3, 7] = (5; 2)$ ，則第 3 階對第 5 階之相對移動率
 $\alpha_t = \frac{m_3}{m_5} - 1 = \frac{4.5}{3.6} - 1 = 0.25 > 0$ ， $\hat{r}_t = 1$ 且 $\tilde{c} = 5 + 2 = 7$ ，所以預測的區間為
 $\hat{X}_6 = [6, 8] = (7; 1)$ 。 □

接下來，本文將粗略的介紹門檻自回歸模式並嘗試著將相對移動率帶入門檻值中。現在研究非線性的門檻值自回歸模型可以分為兩大類，一類是以「時間」作為結構改變的轉折點 (piecewise in time)，探討當模型中解釋變數發生結構改變的時間點前後，與被解釋變數之間的關係是否有所不同。另一類則是以 Tong (1980) 所提出門檻迴歸方法為基礎，以「變數」為結構改變的轉折點 (piecewise in variable) 分析在解釋變數門檻值之上與之下，與被解釋變數之間的關係是否有所不同。由於以「時間」為轉折點的分析方法，必須先主觀的認定發生結構性改變的時間點，方法上較不客觀，因此所得出的結論也就相當分歧。此外，當解釋變數在短期間內持續地發生大幅度的變化時，以時間為結構改變轉折點的模型可能會無法診斷出模型的結構改變。相較之下，以變數為轉折點的分析方法，可以避免上述的缺失。

從 Tong 提出了門檻自我迴歸模型之後，它即成為非線性時間序列上最為常用的模型之一。以下將介紹門檻自迴歸模式及其建構方法。另外，將把相對移動率放入由 Tong (1980) 所提出門檻自迴歸模式中。

由Tong (1980)所提出門檻自迴歸模式，它的基本觀念是把每個時間觀察值所組成的序列中，依其變化情形，找出一轉捩點(change point)或門檻值(threshold)，來清楚劃分為二個區域(regime)，且各自成一系統，服從不同的自迴歸模式。門檻自迴歸模式本身有幾項明顯特徵，作為一般判別的標準，如具有週期循環性、上下起伏的震盪、突然向上跳升或下降的現象，這些特徵是一般傳統線性模式難以描述的地方。Tong發展的門檻自迴歸模型是以「變數」為結構轉變的轉折點，模型中不同的結構是透過以門檻變數大於或小於某一門檻值來表示。本文便是將門檻值用相對移動率(relative moving ratio)帶入嘗試建立一個上升趨勢及下降趨勢的時間數列模型。

在建立模型前，首先要探討一個落後 n 期，以自身落後一期($n=1$)為門檻變數的兩結構門檻自迴歸模型可表示為：

$$X_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1}X_{t-1} + \cdots + \varphi_{n,1}X_{t-n} + \varepsilon_{1t} & \text{若 } X_{t-1} \geq \alpha_t \\ \varphi_{0,2} + \varphi_{1,2}X_{t-1} + \cdots + \varphi_{n,2}X_{t-n} + \varepsilon_{2t} & \text{若 } X_{t-1} < \alpha_t \end{cases} \quad (1)$$

其中， ε_{1t} 與 ε_{2t} 為誤差項，且 $\varepsilon_{it} \sim N(0, \sigma_i^2)$ ， $E(\varepsilon_{1t} | \Omega_{t-1}) = 0$ ， $E(\varepsilon_{2t} | \Omega_{t-1}) = 0$ ， $i=1,2$ ； Ω_{t-1} 表示前一期的訊息集合； α_t 為門檻值。上述門檻自迴歸模型表示當門檻變數不小於門檻值 α_t 時，其適用迴歸式(1)式；當門檻變數小於門檻值 α_t 時，其適用迴歸式(2)式。當 $\varepsilon_{it} \sim N(0, \sigma_i^2)$ 成立時，(1)與(2)式可改寫為

$$X_t = (\varphi_{0,1} + \varphi_{1,1}X_{t-1} + \cdots + \varphi_{n,1}X_{t-n} + \varepsilon_{1t})I_t + (\varphi_{0,2} + \varphi_{1,2}X_{t-1} + \cdots + \varphi_{n,2}X_{t-n} + \varepsilon_{2t})(1-I_t) + \varepsilon_t$$

其中 I_t 為指標函數定義為

$$I_t = \begin{cases} 1 & \text{若 } X_{t-1} \geq \alpha_t \\ 0 & \text{若 } X_{t-1} < \alpha_t \end{cases}$$

且 ε_t 為誤差項， $\varepsilon_t = \varepsilon_{1t} + \varepsilon_{2t}$ 。

本論文中將討論多變量門檻自迴歸模型，Tong (1980)的原始門檻回歸模型並未考慮其他的解釋變數只考慮單一變數，找到時間數列之轉折點，但許多非線性問題，會受到多個變數的牽動而影響，本文探討的台灣股價加權指數受到價格和政策成交量…等等影響。本文將單純的把變數減少為價格，並利用相對移動率取代外生多變數之門檻自迴歸模式。

定義 2.10

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

$$X_{t+1} = \begin{cases} \varphi_{0,1} + \varphi_{1,1}X_t + \varepsilon_{1,t+1} & \text{若 } f(y, z) \geq \alpha_t \\ \varphi_{0,2} + \varphi_{1,2}X_t + \varepsilon_{2,t+1} & \text{若 } f(y, z) < \alpha_t \end{cases}$$

例 2.3.2

令 X_t 為一個時間數列， y_t 、 z_t 為其外生對應的時間數列變數， α_t 為門檻值，外生變數函數 $f(y, z) = \frac{y_t}{z_t}$ 為相除效果，即

$$X_{t+1} = \begin{cases} 2 + 0.8X_t + \varepsilon_{1,t+1} & \text{若 } \frac{y_t}{z_t} \geq 1 \\ 4 + 0.3X_t + \varepsilon_{2,t+1} & \text{若 } \frac{y_t}{z_t} < 1 \end{cases}$$

□

例 2.3.2 將門檻值設為 $\alpha_t = 1$ ，是將定義 2.8 中的 $\alpha_t - 1 > 0$ 移項得到的。

第三章 模糊時間序列的效率評估

第一節 模糊區間的定位及測量辦法

在作效率評估之前，我們要先給區間的反模糊化及區間之間的距離作一些定義。並做一些調整及修改，Wu (2011)將區間的距離給了以下的定義並舉了例子。

定義 3.1 (Wu, 2011)

令 $X = [a, b] = (c; r)$ 為區間模糊數，定義區間模糊數之反模糊化為

$$RX = c + \frac{\ln(1 + |b - a|)}{|b - a|}$$

定義 3.2 (Wu, 2011)

令 U 為一個論域，且 $\{X_i = [a_i, b_i], i = 1, 2\}$ 為 U 中的一個時間數列，其中原心

$c_i = \frac{a_i + b_i}{2}$ ，且半徑 $r_i = \frac{b_i - a_i}{2}$ ，則 X_i 和 X_j 的距離定義為

$$d(X_i, X_j) = |c_i - c_j| + \left| \frac{\ln(1 + |b_i - a_i|)}{|b_i - a_i|} - \frac{\ln(1 + |b_j - a_j|)}{|b_j - a_j|} \right|$$

例 3.1

令兩區間集合 $X_1 = [2, 5]$ ， $X_2 = [3, 7]$ 則 $X_1 = (3.5; 1.5)$ ， $X_2 = (5; 2)$ ，且 X_1 和 X_2 的距離為

$$d(X_1, X_2) = |3.5 - 5| + |\ln(1 + |1.5 - 2|)| = 1.9 \quad \square$$

但是這種定義對區間中的元素數字越大是越不精確的。如下例 3.2

例 3.2

令兩區間集合 $X_1 = [7000, 7010]$ ， $X_2 = [5790, 6210]$ 則 $X_1 = (7005; 5)$ ， $X_2 = (6000; 210)$ ，且 X_1 和 X_2 的距離為

$$d(X_1, X_2) = |7005 - 6000| + \left| \frac{\ln(1 + 10)}{10} - \frac{\ln(1 + 420)}{420} \right| = 1005.225 \quad \square$$

由例 3.2，我們可以發現，當兩個區間的圓心距離越遠，我們幾乎都是用圓心的距離來表示兩個區間的距離。針對此種狀況，我們再給以下一個例子。

例 3.3

令兩區間集合 $X_1 = [6000, 7500]$ ， $X_2 = [7300, 8000]$ 。則 $X_1 = (6750; 750)$ ， $X_2 = (7650; 350)$ ，且 X_1 和 X_2 的距離為

$$d(X_1, X_2) = |7650 - 6750| + \left| \frac{\ln(1+1500)}{1500} - \frac{\ln(1+700)}{700} \right| = 900.004 \quad \square$$

我們將例子 3.1 和 3.2，3.3 做一個比較，我們會發現定義 3.2 將距離計算分成兩個部分，可以觀察出，第一項是圓心的距離，第二項是干擾項，且將干擾項定為圓心距離的百分。所以我們可以做出一個結論就是當我們用定義 3.2 計算區間距離時，如果區間裡的元素很大時，不管區間大小，區間是否重疊，我們皆只要計算圓心距離就好。這似乎和現實不大符合。除此之外，我們可以大膽猜測，隨著區間元素的數字放大，用定義 3.2 去計算的結果誤差會越大。相反的，如果區間元素的數值為個位數，則定義 3.2 是一個好的區間距離衡量方式。由於本文探討的區間為加權指數區間，數字元素非常大，為了避免有誤差的情形，我們將定義 3.1 和定義 3.2 做了些微調整，得到定義 3.3 及定義 3.4。

定義 3.3

令 $X = [a, b] = (c; r)$ 為區間模糊數，定義區間模糊數之反模糊化為

$$RX = c + \frac{|b-a|}{\ln(e+|b-a|)}$$

定義 3.4

令 U 為一個論域，且 $\{x_i = [a_i, b_i], i = 1, 2\}$ 為 U 中的一個時間數列，其中原心

$c_i = \frac{a_i + b_i}{2}$ ，且半徑 $r_i = \frac{a_i - b_i}{2}$ ，則 X_1 和 X_2 的距離定義為

$$d(X_i, X_j) = |c_i - c_j| + \left| \frac{|b_i - a_i|}{\ln(e+|b_i - a_i|)} - \frac{|b_j - a_j|}{\ln(e+|b_j - a_j|)} \right|$$

有了定義 3.4，使用例 3.1，3.2 與 3.3 相同的資料，我們重新定義兩區間的距離分別為例 3.4，3.5 與 3.6。

例 3.4

令兩區間集合 $X_1 = [2, 5]$, $X_2 = [3, 7]$ 則 $X_1 = (3.5; 1.5)$, $X_2 = (5; 2)$, 且 X_1 和 X_2 的距離為

$$d(X_1, X_2) = |3.5 - 5| + \left| \frac{|5-2|}{\ln(e+3)} - \frac{|7-3|}{\ln(e+4)} \right| = 1.88 \quad \square$$

用定義 3.1 的計算方法為 1.9, 但是用修改後的方法計算值為 1.517, 兩者差距不大, 這是因為區間裡的元素都是個位數的。所以可以印證前述的猜測。

例 3.5

令兩區間集合 $X_1 = [7000, 7010]$, $X_2 = [5790, 6210]$ 則 $X_1 = (7005; 5)$, $X_2 = (6000; 210)$, 且 X_1 和 X_2 的距離為

$$d(X_1, X_2) = |7005 - 6000| + \left| \frac{10}{\ln(e+10)} - \frac{420}{\ln(e+420)} \right| = 1070.53 \quad \square$$

例 3.6

令兩區間集合 $X_1 = [6000, 7500]$, $X_2 = [7300, 8000]$ 。則 $X_1 = (6750; 750)$, $X_2 = (7650; 350)$, 且 X_1 和 X_2 的距離為

$$d(X_1, X_2) = |7650 - 6750| + \left| \frac{1500}{\ln(e+1500)} - \frac{700}{\ln(e+700)} \right| = 998.27 \quad \square$$

比較定義 3.2 及定義 3.4 計算的結果, 我們可以發現, 當區間元素的數值變大時, 區間之間的距離也變大了, 這是比較符合現實的。因為隨著區間元素的數值變大, 區間之間的干擾項應該也會變大。這避免了干擾項為圓心距離百分比的情況。所以在第四章, 我們做實證分析時計算區間的距離將採用定義 3.4 的計算方式。

第二節 區間預估模型的效率評估

在這一章節中, 我們要考慮的是區間模糊預測模型的有效性。當然, 隨著各種區間模糊預測的模型出現, 探討其有效性的方法就越來越多。Hsu (2011) 中提到 MSEI, 概念就是把區間的位置, 長度找出來, 再求取兩區間的長度差。如果

我們沒有定義 3.4，因為本文探討的對象為加權指數，所以也可以考慮用這種方式去做區間之間的距離及效率的評估。最直覺的想法便是，如果區間之間的距離越小，表示我們的區間模糊預測模型越準確，但是怎麼樣才算是一個好的預測模型呢？我們需要一種指標來告訴我們模糊預測的精確性，因為如果區間內的元素如果太大，相對的，利用定義 3.4 的公式，得到的距離是會越大的，這是無法直覺的辨別預測的精確性。所以，如果透過指數函數希望能把區間之間的距離能夠用一個介於 [0, 1] 之間的表示方法來評斷。因此本文引用了 Wu (2011) 中的 *IOE*。

定義 3.3 (Wu, 2011)

令 U 為一個論域且 $OI = (c_o, r_o)$ ，為 U 中區間觀測值， $EI = (c_e, r_e)$ 為 U 中區間預測值。則區間觀測值與區間預測值之模糊距離指標定義為

$$IOE = \exp \left\{ - \left[\left| \frac{c_o - c_e}{c_e} \right| + \ln \left(1 + \left| \frac{r_o - r_e}{r_e} \right| \right) \right] \right\}$$

其中 c_o 和 c_e 為區間預測值及區間觀測值之圓心。

為了讓模糊區間預測模型的有效性有一個準則，我們將把 *IOE* 定義於 0 到 1 之間。當 *IOE* 值越高時，表示模糊區間預測模型的方法是越有效率的。當 *IOE* = 0，我們可以觀察出是完全無效率的。如果 *IOE* = 1，代表模糊區間預測模型的建構是完全有效率的。

第四章 實證分析

本文收集從民國 100 年 6 月 1 號至民國 100 年 12 月 30 號的 150 筆歷史資料開盤價格與收盤價格做為區間，分別以點預測，模糊相對移動率預測模型，以及利用相對移動率為門檻值之 ARIMA 自迴規模型三種方法來預測民國 101 年 1 月 2 號的開盤價格與收盤價格的區間預測，並利用第三章的定義 3.2 以及定義 3.3 比較其誤差距離及預估之準確性。以下將這 150 筆資料標記為 $\{X_t\}_{t=1}^n$ ，下圖 4.1 為其走勢圖

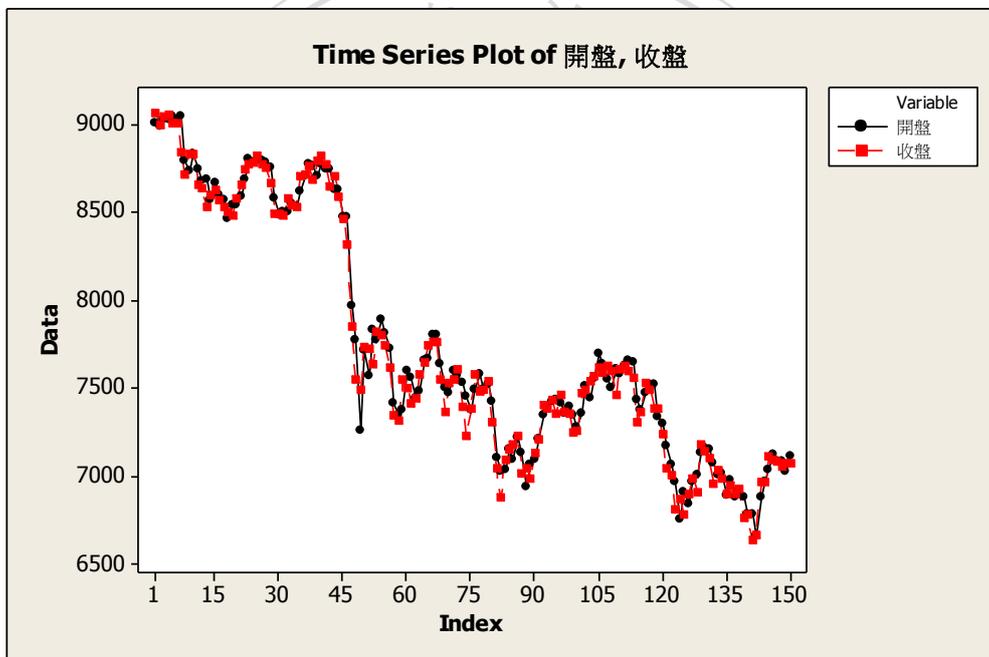


圖 4.1：民國 100 年 6 月 1 號至民國 100 年 12 月 30 號之指數走勢圖

在做預測分析前，為了表現出預估誤差的重要性，先簡單的用 MINITAB 中的 moving average 指令來預測下一期的模糊區間預測，這種預測方法和本文所提供的預測方法是息息相關的，在稍後介紹移動平均波動率時，本文會再詳細介紹其相關性。而圖形 4.2 是用收盤價的 10 日移動平均線來預測未來一期的收盤價，並用 95% 的信賴區間來表示其收盤價可能落入之範圍，如圖形 4.2 所示。

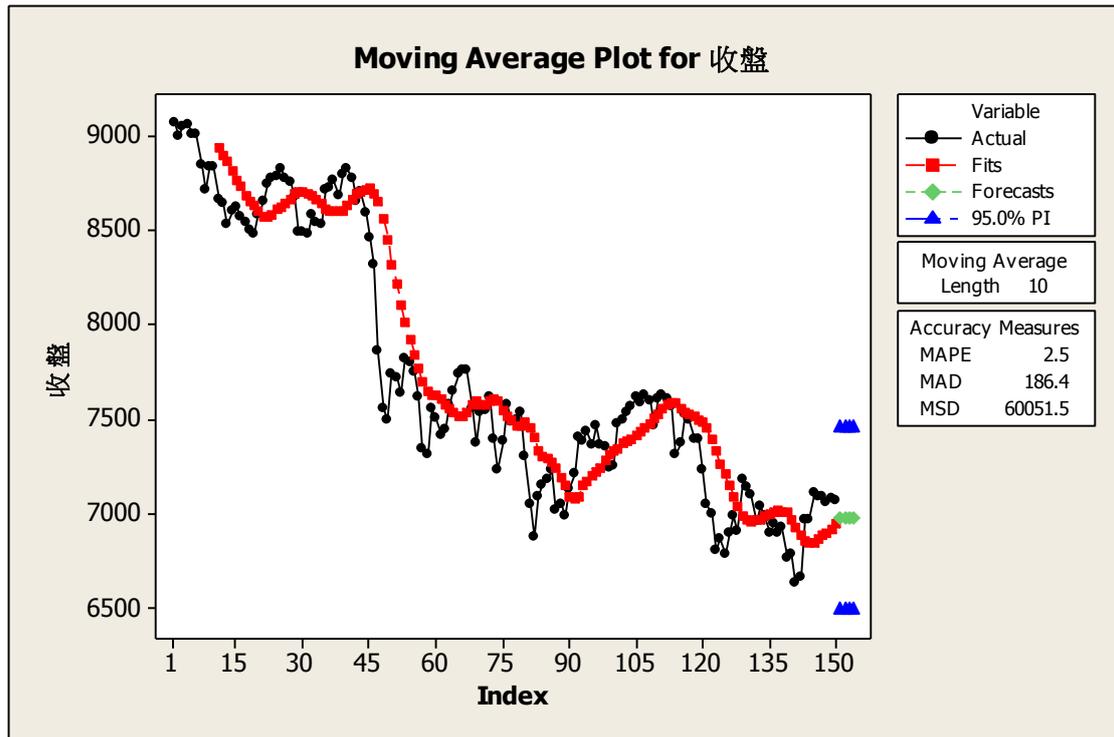


圖 4.2：收盤價之 10 日平均線的預測結果

在圖形 4.2 中，充分的顯示出移動平均的一個很重要的缺陷，就是預估出來的範圍太廣泛了。最後一期的可能收盤價範圍從 6700 到 7500 左右，已經有 800 點左右的誤差，這在現實世界中，是不被投資大眾所接受的。當然，不能否認的，移動平均所創造出來的模糊區間預測還是非常重要的。本文便是利用類似的概念，但是修改其範圍的誤差所創造出來的模型。

為了預測分析方便，我們先把一些建立模型時所需要的基本參數呈列於下表 4.1。

表 4.1 模型建立所需參數

日期	開盤價	收盤價	最高價	最低價	圓心價	半徑
100/06/01	8999.84	9062.35	9089.47	8999.84	9031.10	31.255
100/06/02	8981.71	8991.36	9036.27	8977.67	8986.54	4.825
100/06/03	9024.54	9046.28	9052.18	8997.35	9035.41	10.870
100/06/07	9025.21	9057.10	9059.19	8981.44	9041.16	15.945
100/06/08	9044.85	9007.53	9070.25	9007.53	9026.19	18.660
100/06/09	9020.60	9000.94	9045.44	8996.08	9010.77	9.830
100/06/10	9042.35	8837.82	9053.39	8837.82	8940.09	102.265
100/06/13	8791.67	8712.95	8811.97	8703.52	8752.31	39.360
100/06/14	8737.24	8829.21	8836.51	8737.24	8783.23	45.985
...
100/12/19	6780.64	6633.33	6780.64	6609.11	6706.99	73.655
100/12/20	6654.67	6662.64	6696.93	6646.30	6658.66	3.985
100/12/21	6878.63	6966.48	6966.48	6878.63	6922.56	43.925
100/12/22	6968.22	6966.35	6989.59	6940.85	6967.29	0.935
100/12/23	7035.10	7110.73	7122.28	7035.10	7072.92	37.815
100/12/26	7125.04	7092.58	7125.67	7078.06	7108.81	16.230
100/12/27	7085.50	7085.03	7107.67	7043.64	7085.27	0.235
100/12/28	7086.10	7056.67	7093.47	7036.92	7071.39	14.715
100/12/29	7026.86	7074.82	7074.82	6998.97	7050.84	23.980
100/12/30	7109.85	7072.08	7139.03	7054.79	7090.97	18.885

我們先使用 150 期來預測下一期的結果，預測方式分為三種且預測結果如 1：點預測：

利用公式(2.3.2)我們可以很容易的求出一月二號的收盤預估價格應該為

$$\hat{X}_{151} = E\{X_{151}|X_{150}, X_{149}, \dots, X_1\} = 7733.03$$

但是民國 101 年 1 月 2 號收盤價為 6952.21，很顯然的，利用點預測來預測未來的指數就有一段不小的差距。同理，利用公式(2.3.2)我們可以求出民國 101 年 1 月 2 號的開盤預估價格應該為

$$\hat{X}_{151} = E\{X_{151}|X_{150}, X_{149}, \dots, X_1\} = 7750.92$$

民國 101 年 1 月 2 號開盤價為 7071.35，所以我們可以結論出如果光用點預測，是很容易產生出及大的誤差。當然，我們也可以用類似移動平均的概念來計算點預測。例如從我們的資料庫中，把最後 5 天的收盤價考慮就好，其他的捨去，可以得到。

$$\hat{X}_6 = E\{X_6|X_5, X_4, \dots, X_1\} = 7076.24$$

這個數字，就和實際收盤價接近許多。但是如同上述，其誤差範圍太廣泛了，因此，如何利用移動波動率來修飾誤差範圍將是本文稍後研究的重心。同理，用 5 日平均開盤價格去預估民國 100 年 1 月 2 號的開盤價格結果為

$$\hat{X}_6 = E\{X_6|X_5, X_4, \dots, X_1\} = 7086.67$$

比用 150 日去估計還要好很多。我們把用 5 日平均開盤價及收盤價的預測結果及其誤差利用圖形表示在圖 4.2 和圖 4.3 中。

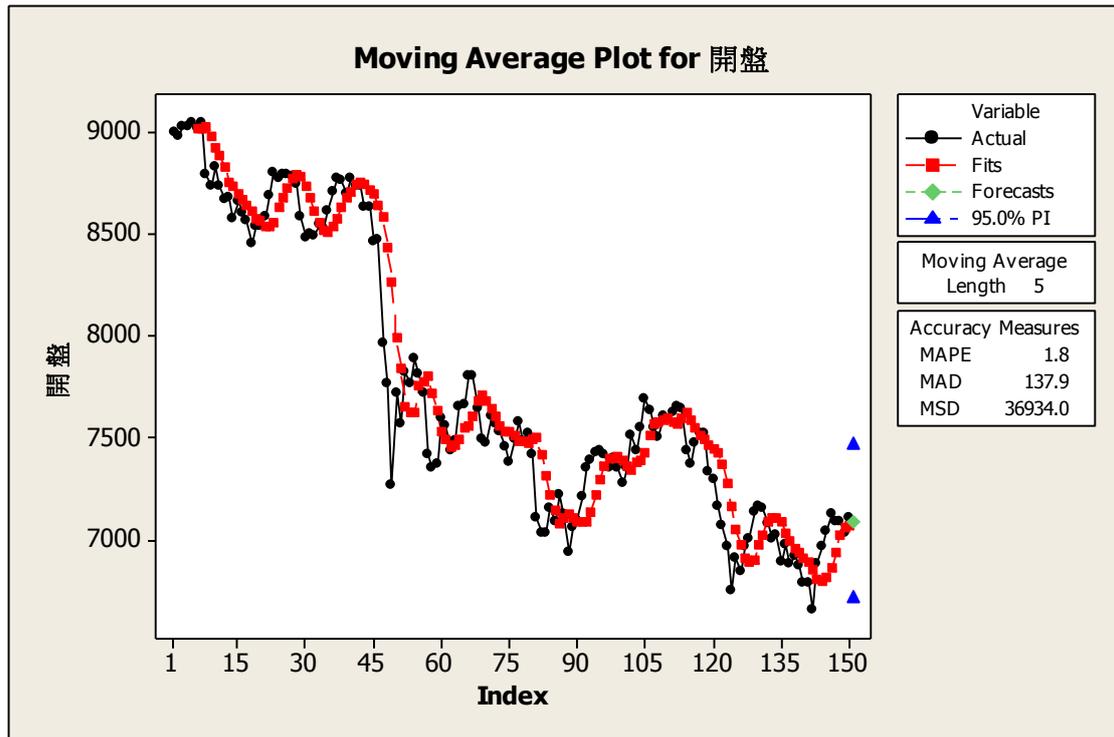


圖 4.3：平均 5 日開盤價預測結果

2：相對移動率：

討論移動平均波動率的實證分析時，我們先從移動平均線開始介紹，台灣的移動平均線是使用一段時間的收盤價格來計算平均值，這和上述點預測是類似的概念，且預測天數通常是用 12 天，24 天，及 144 天為基準。由於本文的資料庫只有 150 天，所以將以 12 天及 24 天為例子，把 12 天及 24 天的平均價格由表 4.2 呈列出來並用圖 4.5 表示。

表 4.2：12 天及 24 天收盤價格平均線的比較

日期	12 天收盤價格平均	24 天收盤價格平均
100/07/05	8612.79	8750.87
100/07/06	8637.27	8740.96
100/07/07	8651.92	8731.88
100/07/08	8662.62	8719.52
100/07/11	8670.84	8703.21
100/07/12	8667.35	8681.69
100/07/13	8666.35	8660.32
100/07/14	8666.55	8645.47
...
100/12/21	6870.46	6932.76
100/12/22	6871.30	6921.62
100/12/23	6877.78	6924.46
100/12/26	6886.92	6928.31
100/12/27	6902.90	6939.92
100/12/28	6911.86	7056.67
100/12/29	6926.74	7074.82
100/12/30	6939.20	7072.08

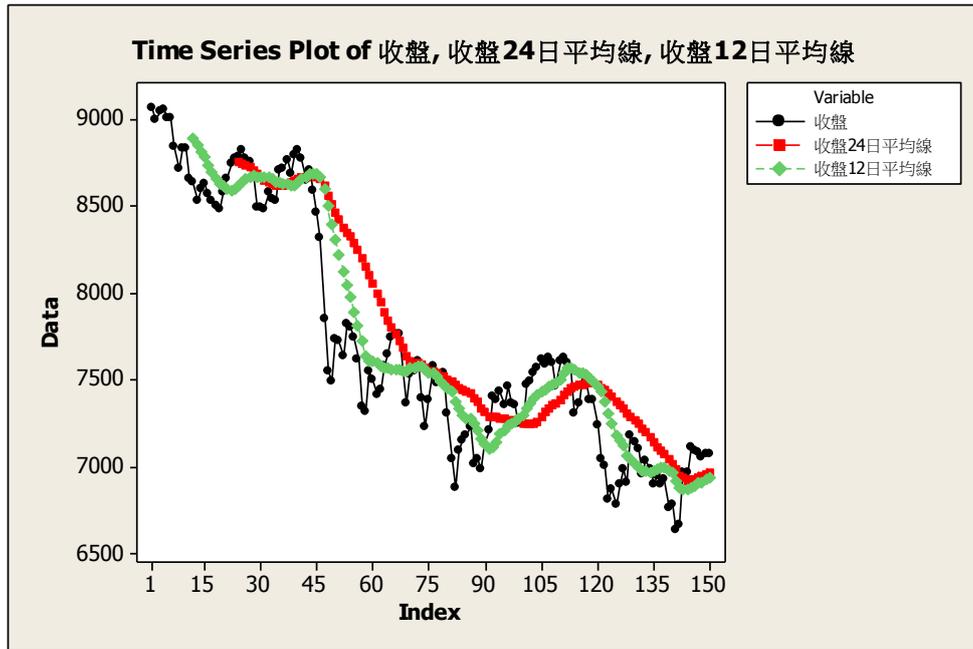


圖 4.5：12 天及 24 天價格平均線及實際值

由圖所知，隨著平均天數的增加，平均線會越趨於圓滑，這也代表離真實的股價收盤數誤差會越來越大。相反的，如果選取的平均數太小，則波動太大，這並非是一種好的預估模式。所以，如何選取一個有效的樣本參數去估計真實的情況是一門很值得研究的學問。在找出相對移動率之前，我們必須根據觀測期數的開盤價與收盤價找出一系列的圓心及半徑。如上表 4.1 所示。於是根據定理 2.5，我們有

$$\{X_t\} = \{ [8999.84, 9062.35], [8981.71, 8991.36], \dots, [7109.85, 7072.08] \}$$

其中

$$X_1 = [8999.84, 9062.35] = (9031.10; 31.255)$$

$$X_2 = [8981.71, 8991.36] = (8986.54; 4.825)$$

⋮

$$X_{150} = [7109.85, 7072.08] = (7090.97; 18.885)$$

很顯然的，如果我們取估計的參數為 150，我們的區間估計誤差會很大。這道理和移動平均的誤差是一樣的。這也是我們為何在討論相對移動率時，要先研

究移動平均線的原因。在移動平均線裡，如同上述所示，平均天數的選取會影響移動平均線的好壞，同理，如何選取一個有效的觀測天數，對相對移動率而言也會有很直接的影響，在本文中，我們簡單的用 5 天 10 天來當作我們估計的參數。會想用 5 天及 10 天作為參數，是因為越短期的預估(例如預測單位為一天)，則股價振幅會越大，較容易失真。所以，本文希望能在短期投資當中，讓振幅不要這麼大，卻又不失其精確性，所以採用 5 個交易日及 10 個交易日的天數套入相對移動率區間預估參數。在表 4.3 中我們利用定理 2.11 求出相對移動率其中 \hat{r}_t 的參數為 5。

表 4.3：相對移動率的各種參數

日期	5 日圓心 均價(m_5)	10 日圓心 均價(m_{10})	相對移動 率(α_t)	移動平均 半徑(\hat{r}_t)
100/06/15	8863.81	8943.95	-0.0089596	39.729
100/06/16	8801.09	8910.55	-0.0122848	46.304
100/06/17	8743.82	8877.27	-0.0150328	29.386
100/06/20	8714.35	8834.23	-0.0135698	36.361
100/06/21	8674.66	8788.59	-0.0129635	29.733
100/06/22	8636.61	8750.21	-0.0129800	33.774
100/06/23	8614.03	8707.56	-0.0107000	28.620
...
100/12/21	6778.33	6855.44	-0.0112479	35.479
100/12/22	6807.91	6852.09	-0.0064478	24.707
100/12/23	6865.68	6870.15	-0.0006495	32.063
100/12/26	6946.05	6884.79	0.0088970	20.578
100/12/27	7031.37	6904.35	0.0183969	19.828
100/12/28	7061.14	6919.73	0.0204346	19.828
100/12/29	7077.85	6942.88	0.0194396	18.595
100/12/30	7081.46	6973.57	0.0154707	14.809

接下來，本文將利用上述的參數去預估民國 101 年 1 月 2 號至 1 月 6 號的平均開盤價格，並探討其準確性。

$$\hat{r}_t = \frac{\sum_{i=1}^5 r_i}{5} = \frac{16.23 + 0.235 + 14.715 + 23.98 + 18.885}{5} = 14.809$$

其中 $t_1 = 5$ ， $t_2 = 10$ 則相對移動率 α_t 為

$$\alpha_t = \frac{m_{t_1}}{m_{t_2}} - 1 = \frac{7081.46}{6973.57} - 1 > 0$$

因為 $\alpha_t > 0$ ，利用公式(2.3.3)，我們可以得到一個新的圓心 \tilde{c}

$$\tilde{c} = \hat{c} + \hat{r} = 7090.97 + 14.809 = 7105.779$$

有了新圓心之後，再利用公式(2.3.4)，我們可以得到一個預測的模糊區間

$$\hat{X}_{t+1} = (\tilde{c}; \hat{r}_t) = (7105.779; 14.809)$$

比較需要注意的是，這組新的模糊預測區間為下 5 個交易日的模糊預測區間，並非下一個交易日，因為半徑的參數 \hat{r}_t 為 5，這代表的意義是最近 5 日交易日中的振幅，所以代表下 5 個交易日時最有可能產生之振幅，也就是 \hat{r}_t 。如果想要預測的模糊區間的時間單位為下一交易日，只要改參數 t_1 、 t_2 和半徑的時間參數就可以，建議是把時間縮短找出合適的預估振幅。本文在此不再贅述。

在民國 101 年 1 月 2 號至 1 月 6 號之平均開盤價格為 7071.35，平均收盤價格為 7120，這和我們用相對移動率所預測的模糊區間幾乎吻合。圖 4.6 將每日的模糊預測區間顯示出來。

接下來，我們要利用第三章的公式來探討相對移動率模型的有效性。本文將直接利用定義 3.2 來求民國 101 年 1 月 2 號至 1 月 6 號那一週的模糊預測價格區間及實際價格區間之間的距離，並利用定義 3.3 求出其 IOE。

$$X_{EI} = (\tilde{c}; \hat{r}_t) = (7105.78; 14.81) = [7090.97, 7120.58]$$

$$X_{OI} = (c; r) = (7095.5; 24.5) = [7071, 7120]$$

則實際股價區間以及預測股價區間的距離為

$$d(X_{OI}, X_{EI}) = 10.25 + \left| \frac{29.61}{\ln(e + 29.61)} - \frac{49}{\ln(e + 49)} \right| = 14.15$$

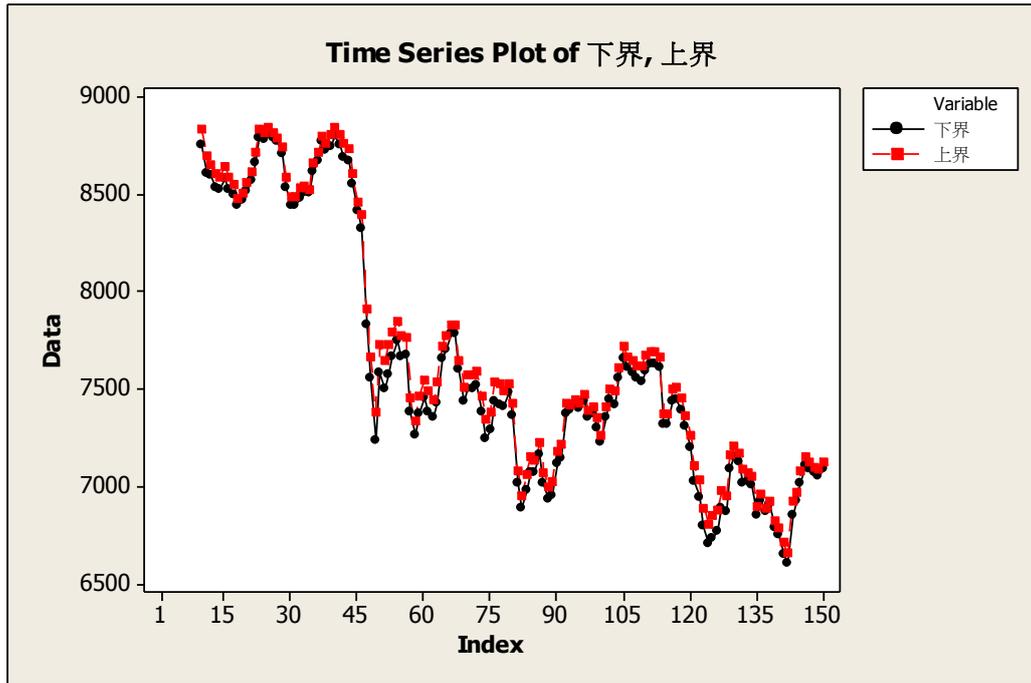


圖 4.6：相對移動率之預估結果

就實際考量而言，我們利用相對移動率所預測出來的區間，和民國 101 年 1 月 2 號至 1 月 6 號的平均開盤價格與平均收盤價格之區間相差只有 20 點。有了模糊預測價格區間及實際價格區間之間的距離之後，接下來本文便要測試模糊相對移動率預測模型在預測民國 101 年 1 月 2 號至 1 月 6 號的平均開盤價格與平均收盤價格之效率性，根據定義 3.3

$$OI = (c_o; r_o) = (7095.5; 24.5) = [7071, 7120]$$

$$EI = (c_e; r_e) = (7105.78; 14.81) = [7090.97, 7120.58]$$

則相對移動率所預測區間效率性

$$IOE = e^{-\left(\frac{|c_o - c_e| + \ln(1 + \frac{|r_o - r_e|}{r_e})}{c_e}\right)} = 0.603$$

由於 IOE 是介於 0~1 之間且越靠近 1 預測效率越高。就預測民國 101 年 1 月 2 號至 1 月 6 號這一週的預測而言呢，效率算是好的。本文利用表 4.4 來表示從民國 100 年 6 月 15 號起每 5 天的實際區間和利用模糊移動平均波動率所預測區間之間的距離及其 IOE 。

表 4.4：模糊移動平均波動率之預測距離及效率評估

日期	實際圓心	預測圓心	距離	IOE
100/06/15	8832.66	8792.93	77.692	0.678045
100/06/16	8697.14	8650.84	56.835	0.703371
100/06/17	8653.78	8624.39	26.785	0.606569
100/06/20	8604.92	8568.56	47.544	0.652864
100/06/21	8584.78	8555.05	31.827	0.840819
100/06/22	8642.45	8608.68	37.76	0.865146
100/06/23	8584.22	8555.60	31.7531	0.842029
...
100/12/23	7072.92	7040.86	38.13	0.505261
100/12/26	7108.81	7129.39	22.69	0.730781
100/12/27	7085.27	7105.10	23.71	0.732541
100/12/28	7071.39	7085.38	30.630	0.519463
100/12/29	7050.84	7069.44	13.881	0.597833
100/12/30	7090.97	7105.78	28.697	0.604827

3：門檻自迴規模型：

此方法為移動平均波動率之應用，利用移動平均波動率做為門檻值，代入 ARIMA 模型中，解出預測值之後在利用 \hat{r}_t 將數值模糊化為預測結果。建立模型前，要先決定落後期數。在建構二個具有相同參數之模型時，建構過程中每多增加一個變項可以降低殘差值的和，可是會增加將來預測的複雜性，同樣地，在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。為了估計模型的品質並避免模型參數過度配適的情形，Akaike (1973) 提出了以懲罰多餘參數的 AIC (Akaike information criteria) 準則，其定義如下：

$$AIC(M) = -2 \ln L(\hat{\beta}) + 2M$$

此處的 M 為模型中參數 β 的個數， L 為 β 的概似函數(likelihood function)， $\hat{\beta}$ 為 β 的最大概似估計量，對含有 n 個有效觀測值之 ARIMA 模式其概似函數為

$$\ln L(\beta) = -\frac{n}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n \varepsilon_t^2(\phi, \mu, \theta)$$

其中 ϕ , μ , θ 已知下可簡化求得

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

最佳模式之選取是由 $AIC(M)$ 之最小值所決定。除此之外，也可以用 ACF 及 PACF 之圖形判定。由圖 4.7 , 4.8 表示出 AR 及 MR 之落後期數為各為 1。

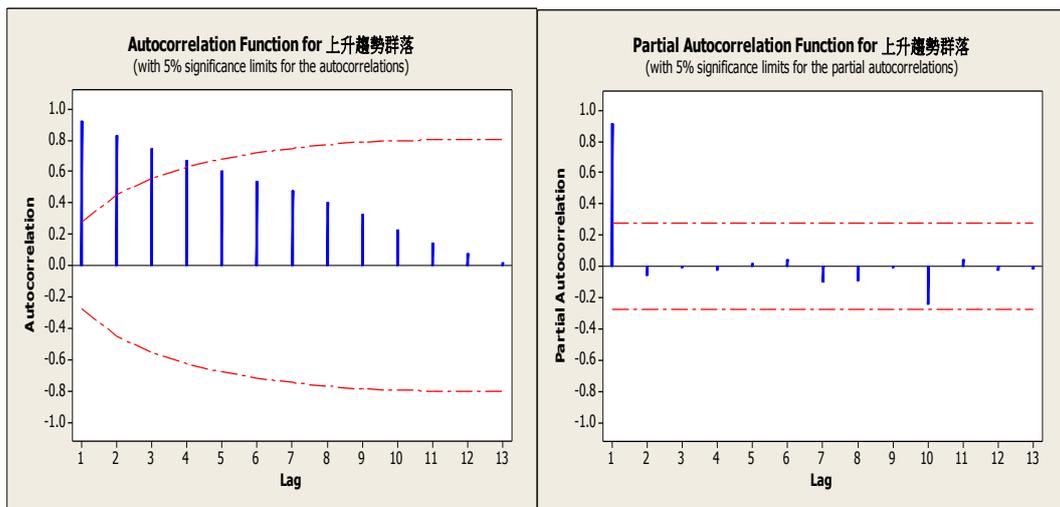


圖 4.7：上升趨勢群落之 ACF 及 PACF 圖形

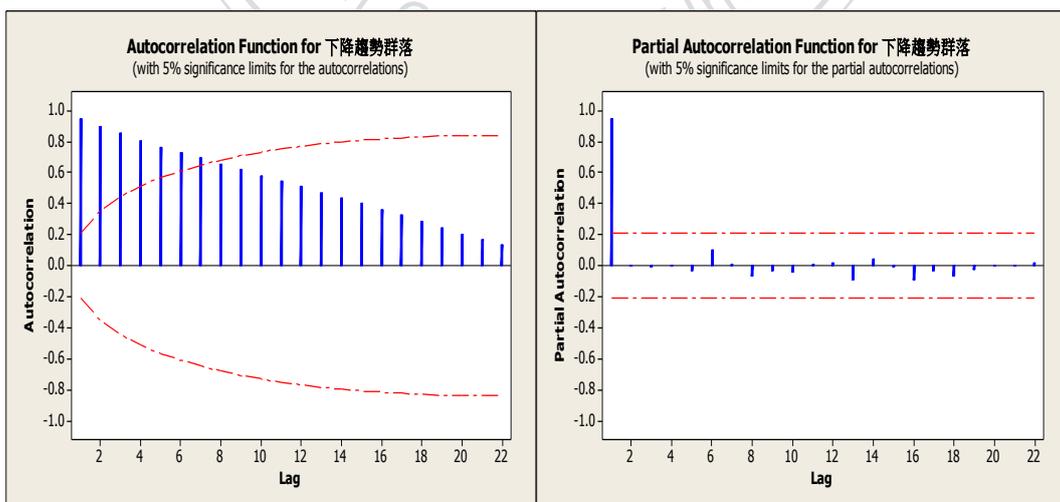


圖 4.8：下降趨勢群落之 ACF 及 PACF 圖形

決定落後期數後根據定義 2.11，可以得到下列的模型 A。

模型 A：

令 X_t 為一個時間數列， $y_t = m_5$ ， $z_t = m_{10}$ 為其外生對應的變數， $\alpha_t = 1$ 為門檻值，則含外生多變數之門檻自迴歸模型為

$$X_{t+1} = \begin{cases} \phi_{1,1} + \phi_{1,2}X_t + \varepsilon_{1t+1} & \text{若 } \frac{y_t}{z_t} \geq \alpha_t \\ \phi_{2,1} + \phi_{2,2}X_t + \varepsilon_{2t+1} & \text{若 } \frac{y_t}{z_t} < \alpha_t \end{cases}$$

有了模型之後，需要將資料利用門檻值分類，本文利用相對移動率的經濟意義將資料分成兩類，第一類組為上升趨勢群落，第二類組為下降趨勢群落，分別以阿拉伯數字 1 及 2 代表第一類組及第二類組，分組之結果以表 4.5 來表示。



表 4.5 台灣加權指數分類表

日期	台灣加權指數分類
100/06/15	2
100/06/16	2
100/06/17	2
100/06/20	2
100/06/21	2
100/06/22	2
100/06/23	2
100/06/24	2
100/06/27	2
100/06/28	2
100/06/29	2
100/06/30	1
...	...
100/12/15	2
100/12/16	2
100/12/19	2
100/12/20	2
100/12/21	2
100/12/22	2
100/12/23	2
100/12/26	1
100/12/27	1
100/12/28	1
100/12/29	1
100/12/30	1

分類完成後，要先檢定資料是否穩定。由於股價的走勢圖為隨機漫步(random walk)，所以分類完後並不是穩定的狀態，以圖 4.9 顯示上升趨勢群落及下降趨勢群落的趨勢圖。

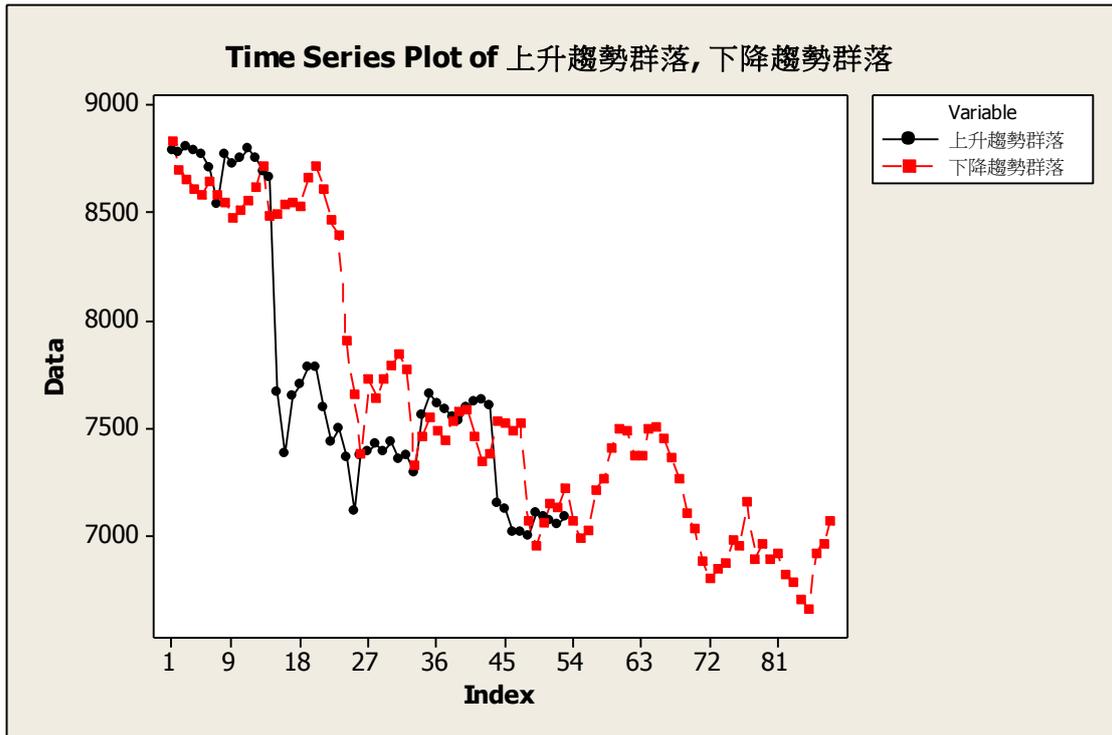


圖 4.9 上升趨勢群落及下降趨勢群落之走勢

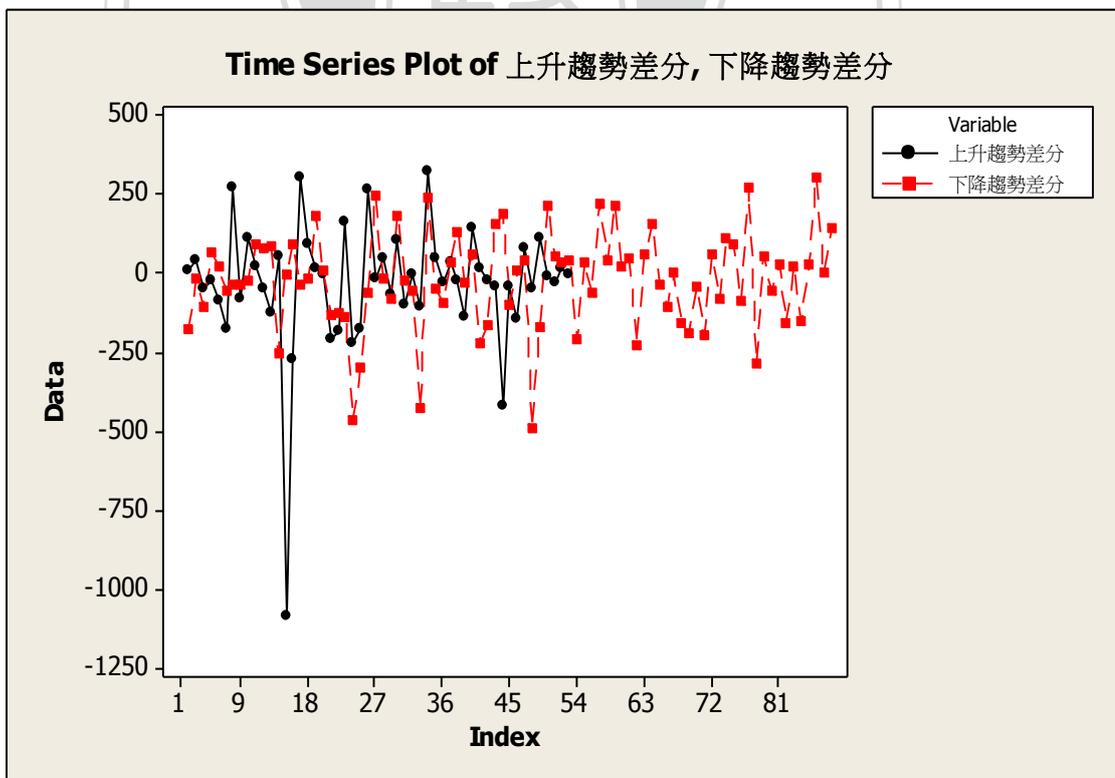


圖 4.10 差分過後第一類及第二類圖形

由於分類過後的資料是不穩定，因此我們將利用差分(difference)使其期望值穩定，將資料轉換成穩定型態。如果差分一次後並不能使資料呈顯穩定狀態，則繼續差分直到資料呈現穩定狀態為止。圖 4.10 表示差分一次後的圖形。

差分完畢後，資料呈現弱穩定狀態，於是我們利用 MINITAB 求出模型係數。可以得到下列門檻值自迴規模型。

$$\Delta X_{t+1} = \begin{cases} -4.2399 + 0.897\Delta X_t + \varepsilon_{1t} + 1.0298\varepsilon_{1t+1} & \text{若 } \frac{m_5}{m_{10}} \geq \alpha_t = 1 \\ -3.3067 + 0.849\Delta X_t + \varepsilon_{2t} + 0.9736\varepsilon_{2t+1} & \text{若 } \frac{m_5}{m_{10}} < \alpha_t = 1 \end{cases}$$

利用上述模型求出來的預測值此時還是點預測，此時我們還要利用表 4.3 中的移動平均半徑將數值模糊化，就可以得到模糊預測時間序列。跟利用模糊移動平均波動率預測模型一樣，所預估的模糊時間數列為下 5 日交易日的平均價格，本文把上述模型點預測之結果利用圖 4.11 及 4.12 表示。

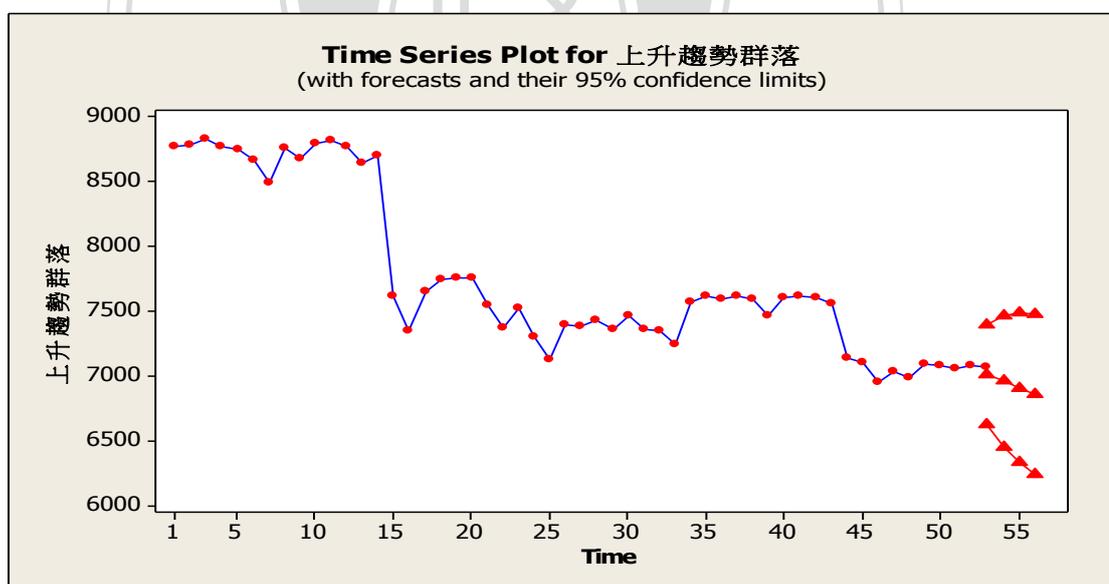


圖 4.11 門檻自我回歸模上升趨勢型預測結果

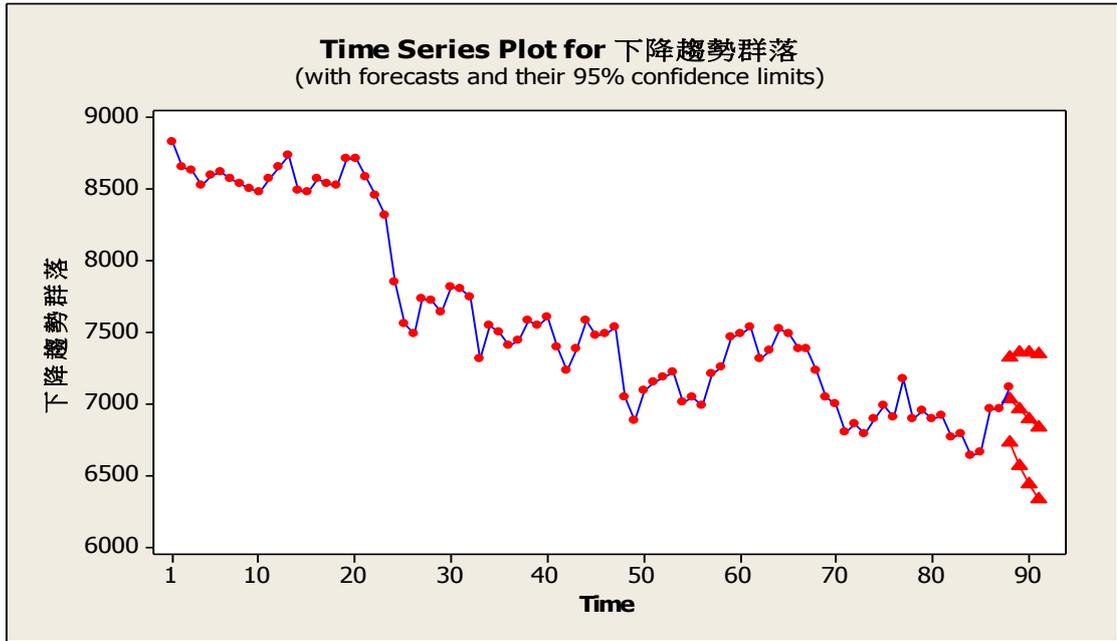


圖 4.12 門檻自我回歸模上升趨勢型預測結果

由表 4.5，民國 100 年 12 月 30 日之分類為上升趨勢群落。因此，利用模型求出的預測圓心為 7070.36。得到預測的圓心後，再利用表 4.3 之移動平均半徑 \hat{r}_t 把數值模糊化，根據表 4.3， $\hat{r}_t = 14.809$ 。所以我們得到民國 101 年 1 月 2 號至民國 101 年 1 月 6 號的平均價格預測區間為

$$X_{EI} = (\tilde{c}; \hat{r}_t) = (7070.36; 14.81) = [7055.55, 7085.17]$$

民國 101 年 1 月 2 號至民國 101 年 1 月 6 號之實際價格為

$$X_{OI} = (c; r) = (7095.5; 24.5) = [7071, 7120]$$

根據定義 3.2 預測價格區間和實際價格區間之距離為

$$d(X_{OI}, X_{EI}) = \left| 7070.36 - 7095.5 \right| + \left| \frac{29.62}{\ln(e + 29.62)} - \frac{49}{\ln(e + 49)} \right| = 29.04$$

根據定義 3.3 預測價格區間和實際價格區間之效率為

$$IOE = e^{-\left(\left| \frac{25.14}{7070.36} \right| + \ln \left(1 + \left| \frac{9.69}{14.81} \right| \right) \right)} = 0.602$$

這結果和利用模糊移動平均波動率模型所預測出來的結果成效是很接近的。

第五章 結論

由於人類的思考複雜，社會上各種現象的影響，傳統的點估計預測已經不敷使用，模糊區間預測是由傳統的點預測所發展出來的預測方式，他既可以解決一些點預測無法處理的問題，又不違背傳統點預測的精神。雖然點預測方式有其較強的理論基礎。但是，數據收集程序可受一些不可預知的因素，而降低準確性。如果我們利用這種人為的準確性做因果分析，這可能導致偏差因果關係的判斷，誤導決策。

本文應用模糊時間數列法，提出一個相對移動率模糊時間序列預測模型來預測加權指數的開盤價及收盤價，此種方法解決了傳統上時間序列模型複雜的計算量，也可以和門檻自迴歸模型搭配使用，除此之外，本論文也將 Wu(2011)提出區間時間數列的距離，做了一些修改，解決了在修正項調整影甚小的結果。本論文模型的最大特色為，藉著調整估計的參數來達到區間預測的功效。方法不僅簡單便利，也可以藉著參數的改變，對於短期或中長期區間預測也有一定的預測水準。

本論文提出以區間數據，在常見的區間模糊時間數列方法外，再提供一種新的預測方式。但有關於影響圓心的選取，會因其相對移動率的權重不同而不同。因此，在未來如何選取一個恆常而且有效的相對移動率，為後續研究者所探討的。又因為影響加權股價指數的因素，除了可衡量的經濟因素外，還有政治因素、政府干預、投資者的預期心理變數等，這些皆是無法量化的變數。再加上，如何選取一個有效的估計參數也是一門很大的學問。因此，在未來可以考慮如何採取有效的衡量方法將其量化之外，還有以研究如何選取一個有效的估計參數。

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，第一期，頁67-76。
- 楊奕農(2009)，時間序列分析：經濟與財務上之應用，雙葉書廊，台北。
- Akaike, H. (1973). Information theory and an extension of maximum likelihood principle, *Second International Symposium on Information Theory* 1, 267-281.
- Box, G. P. and Jenkins, G. M. (1976). *Time series analysis forecasting and control*. San Francisco: Holden-Day.
- Byers, J. D. and Peel, D. A. (1995). Evidence on volatility spillovers in the interwar floating exchange rate period based on high/low prices, *Applied Economics Letters* 2(10), 394-396.
- Chow, G. C. (1960), Tests of equality between sets of coefficients in two linear regressions, *Econometrica* 28(3), 591-605.
- Donald W. K. A. and Werner P. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62(6), 1383-1414.
- Graham, B. P. and Newell, R. B. (1989). Fuzzy adaptive control of a first-order process. *Fuzzy sets and system* 31, 47-65.
- Haggan V. and Ozaki T. (1980). Amplitude-dependent exponential AR model fitting for non-linear random vibrations, in *Time Series*, O. D. Anderson ed., North-Holland, Amsterdam.
- 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.
- Subba R. T. and Gabr M. (1980). A test for linearity of stationary time series analysis, *Journal of Time Series Analysis* 1(1), 145-158.
- Tong, R. M. (1978). Synthesis of fuzzy models for industrial processes. *Int. J. Gen.* 4, 143-162.
- Tong H. and Lim K. S. (1980). Threshold Autoregressive, Limit Cycles and Cyclical Data (with Discussion), *Journal of the Royal Statistical Society. Series B* 42(3),

245-292.

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.

Zadeh, L. A. (1965). Fuzzy sets, *Information and Control* 8, 338-353.

Zhou H. D. (2005). Nonlinearity or structural break - data mining in evolving financial data sets from a Bayesian model combination perspective, *Proceedings of the 38th Hawaii International Conference on System Sciences*, Hawaii, U.S.A.

