

摘要

近年來，預測技術的創新與改進愈來愈受到重視。對於預測效率評估的要求也愈來愈高。尤其在經濟建設、人口政策、經營規畫、管理控制等問題上，預測更是決策過程中不可或缺的重要資訊。目前有關模糊時間數列分析與預測效率評估並不多見。主要是模糊殘差值的測量相當困難。有鑑於此，本文提出以模糊距離來進行效率評估。並且從不同的角度來探討預測的準確度。實證研究顯示，藉由中心點與區間長度的整合測度，可以得到一個合理的評估結果。這對於財務金融的模糊數據分析與未來市場的走勢將深具意義。

關鍵詞：模糊時間數列分析、預測、整合測度、效率評估。



目錄

1. 前言.....	3
2. 區間模糊數與預測效率分析.....	5
2.1 模糊時間數列.....	5
2.2 常見的區間時間數列預測模式.....	6
2.3 預測效率評估.....	9
3. 研究方法.....	12
3.1 加權時間數列法.....	12
3.2 加權模糊時間數列法.....	16
4. 實證分析.....	17
4.1 資料來源.....	17
4.2 加權模糊時間數列法.....	17
4.3 左右端點 k 階區間移動平均法.....	22
4.4 比較「加權模糊時間數列法」及「左右端點 k 階區間移動平均法」 的測量誤差：.....	27
5. 結論.....	28
參考目錄.....	29

1. 前言

預測是人類思維中很重要的行為，人類依據過去的經歷，計畫將來的行為，如同人類念歷史，從古人的經驗，得到新的啟發一般。傳統的點預測已經無法滿足人類充滿不確定與灰色地帶的心理模式，所以發展出區間預測。目前有關時間數列分析與預測方法多著重在點預測，應用區間時間數列分析並不多見(Wu 2005)。但在實務上，區間預測更有效率，且更符合投資者需求。以台股為例，目前仍以收盤價作為判斷股市漲跌的依據，然而台股每天從股市開盤至收盤的 4.5 個小時，股價呈現持續地波動震盪，投資者也是在這段時間交易買賣，若單純以收盤價作為漲跌依據，不免流於以管窺天、以蠡測海的侷限。再者，台股設有 7% 漲跌幅的限制，而上海、深圳股市也有漲跌幅 10% 的規定，股價在漲跌幅範圍內震盪，所以區間時間數列更能精準地顯示出股價的趨勢，區間的波動過程，比起點對點的預測，更可以幫助我們做出正確的決策。又如預測氣象時，氣象播報員給我們每天氣溫的區間，根據太陽照射角度不同，我們知道氣溫預測的對低點通常出現在凌晨，氣溫預測的最高點出現在中午，這也是將區間時間數列運用在日常生活上的例子。

模糊理論(Zadeh 1965)已經成功地運用在預測方面，如預測：氣溫、股價、匯率等等，在 Zadeh 介紹模糊集合之後，Song 與 Chissom (1993)也發展了模糊時間數列的定義，以及用模糊關係式發展模糊時間數列的模型。Kunhuang (2001)曾經研究區間長度對模糊時間數列預測結果的影響，Chen (1996)也提出了一個簡單的模糊時間數列預測模式，來預測學生註冊人數，並且得到很好的預測效果。Wu and Hung (2006)提出模糊認定法則，以作為 ARCH 模式族與 Bilinear 模式族的決策判定標準。Chen (2002)發表高階的模糊時間數列 (high-order fuzzy time series)來預測阿拉巴馬大學的學生註冊人數。Chen 等人 (2006)對 TAIEX 的預測提出了加權模糊時間數列模型。

至於區間效率評估的部分，隨著軟計算的興起，區間計算方法的發展越來越受到重視，Chatfield(1993)是區間預測計算的先驅之一，Huarng(2001)研究區間長度對模糊時間數列預測結果的影響，Tseng and Tzeng (2002) 對模糊的季節性 ARIMA 預測模型作估計，而 Chang(2007)也為區間資料的平均數與變異數訂定假設檢定法則，Vladik, Hung and Wu (2006) 提出網路即時的統計計算方法。現在也有更多學者，依循前人的腳步，不斷推陳出新，發展更適合如今變動社會的區間預測模型，以及對區間效率評估的研究。

許多的模糊時間數列模式，為了得到較精確的預測結果，也都有如下的缺點：

1. 在建構模糊關係時，投入相當多的公式作計算
2. 計算的成本高

有鑑於此，本研究推廣 Pathak and Singh(2011)的模糊時間數列預測模式，將其推展為區間模糊時間數列預測模型，並且透過台股的五日、七日、十日移動平均線，來預測台股每日股價的最高點、最低點所構成的區間，並分析不同預測模式的效率的效率。

本文共分為五章，第一章探討相關方法的文獻，以獲取發展理論架構的依據，並說明分析研究台股的動機。第二章說明模糊時間數列的理論與分類，第三章架構新的加權模糊時間數列預測模式，並且比較幾種常間的模糊時間數列的預測效率。第四章以台股股價作為實證分析。本文結論在第五章。

2. 區間模糊數與預測效率分析

2.1 模糊時間數列

本研究針對加權模糊時間數列的預測與績效評估進行探討，並將模糊時間數列推廣成加權模糊時間數列。在定義加權模糊時間數列之前，需要先行定義一般的模糊時間數列。

定義 2.1 模糊時間數列(Hsu, Tse and Wu 2003)

設 $\{X_t \in R, t = 1, 2, \dots, n\}$ 為一時間數列且 Ω 為其論域。給定 Ω 的一個次序分割(Ordered Partition Set), $\{P_i; i = 1, 2, \dots, r, \cup_{i=1}^r P_i = \Omega\}$, 其相對應之語言變數為 $\{L_i, i = 1, 2, \dots, r\}$ 。若在 $\{L_i, i = 1, 2, \dots, r\}$ 上相對於 X_t 的模糊集合 FX_t 有隸屬度函數為 $\{\mu_i(X_t); 0 \leq \mu_i(X_t) \leq 1, i = 1, 2, \dots, r\}$, 則我們稱 $\{FX_t\}$ 為相對於 $\{X_t\}$ 上的一個模糊時間數列且表示為

$$FX_t = \frac{\mu_1(X_t)}{L_1} + \frac{\mu_2(X_t)}{L_2} + \dots + \frac{\mu_r(X_t)}{L_r}$$

其中"+"表示連結符號, $\frac{\mu_r(X_t)}{L_r}$ 表示 X_t 相對於語言變數 L_i 及其隸屬度函數 $\mu_i(X_t)$ 的對應關係, $\mu_i: R \rightarrow [0, 1]$, 且對所有 $t=1, 2, \dots, r$, $\sum_{i=1}^r \mu_i(X_t) = 1$ 。

為了方便起見, 以後我們利用其相對應各語言變數的隸屬度來表示 FX_t , 且簡寫為 $FX_t = (\mu_1, \mu_2, \dots, \mu_r)$ 。

例 2.1 設一時間數列 $\{X_t\} = \{0.7, 1.9, 2.7, 4.2, 3.5, 3.1, 4.4, 3.7\}$ 。觀察其全距取其論域為 $\Omega = [0, 5]$, 並選擇一個次序分割為 $\{(0, 1), [1, 2), [2, 3), [3, 4), [4, 5)\}$; 令論域

中各分割集合之相對應的語言變數為：很低 = $L_1 \propto [0,1)$ ，低 = $L_2 \propto [1,2)$ ，中等 = $L_3 \propto [2,3)$ ，高 = $L_4 \propto [3,4)$ ，很高 = $L_5 \propto [4,5)$ ，其中“ \propto ”表示“相對於”。再針對各個分割集合取其平均數為 $\{m_1 = 0.5, m_2 = 1.5, m_3 = 2.5, m_4 = 3.5, m_5 = 4.5\}$ 作為其相對於語言變數的典型值。因為 X_1 介於 0.5 與 1.5 之間，則可得相對於 X_1 的模糊時間數列為 $FX_1 = (0.8, 0.2, 0, 0, 0)$

$$\frac{1.5 - 0.7}{1.5 - 0.5} = 0.8 \in L_1, \quad \frac{0.7 - 0.5}{1.5 - 0.5} = 0.2 \in L_2$$

同理，我們可得到相對於 $\{X_t\}$ 的模糊時間數列 $\{FX_t\}$ 。

很低	低	中等	高	很高	
$FX_1 = ($	0.8,	0.2,	0,	0,	0)
$FX_2 = ($	0,	0.6,	0.4,	0,	0)
$FX_3 = ($	0,	0,	0.8,	0.2,	0)
$FX_4 = ($	0,	0,	0,	0.3,	0.7)
$FX_5 = ($	0,	0,	0,	1,	0)
$FX_6 = ($	0,	0,	0.4,	0.6,	0)
$FX_7 = ($	0,	0,	0,	0.1,	0.9)
$FX_8 = ($	0,	0,	0,	0.8,	0.2)

2.2 常見的區間時間數列預測模式

傳統的時間數列預測方法，例如：AR，MA，ARMA，ARIMA……，是點對點的預測，但是區間時間數列無法以點對點的方式來預測，以下列舉幾項常見的區間時間數列預測模型。

定義 2.2 區間時間數列(Wu 2006)

令 $\{X_t = [a_t, b_t] = (m_t, l_t u_t)_{LR} | t = 1, 2, 3, \dots\}$ ，為一個區間時間數列，因為 $X_t = [a_t, b_t] = (m_t, l_t u_t)_{LR}$ 皆為區間資料，則稱 $\{X_t\}_{t=1}^n$ 為一區間時間數列。

定義 2.3 左右端點法之 ARIMA 模式 (Hsu 2008)

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

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

$$b_t = \alpha + \beta_1 b_{t-1} + \dots + \beta_{p_r} b_{t-p_r} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \dots - \varphi_{q_r} \varepsilon_{t-q_r}, \text{ 其中}$$

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

則 $\widehat{X}_t = E[X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_1] = [\widehat{a}_t, \widehat{b}_t]$ ，其中

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

$$\widehat{b}_t = E(b_t | b_{t-1}, b_{t-2}, \dots, b_1) = \alpha + \beta_1 b_{t-1} + \dots + \beta_{p_r} b_{t-p_r}$$

定義 2.4 中心點及半徑法之 ARIMA 模式 (Hsu 2008)

令 $\{X_t = [c_t, r_t], t = 1, 2, \dots, n\}$ 為一個區間時間數列，

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

$$r_t = \alpha + \beta_1 r_{t-1} + \dots + \beta_{p_r} r_{t-p_r} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \dots - \varphi_{q_r} \varepsilon_{t-q_r},$$

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

則 $\widehat{X}_t = E[X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_1] = [\widehat{c}_t - \widehat{r}_t, \widehat{c}_t + \widehat{r}_t]$ ，其中

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

$$\widehat{r}_t = E(r_t | r_{t-1}, r_{t-2}, \dots, r_1) = \alpha + \beta_1 r_{t-1} + \dots + \beta_{p_r} r_{t-p_r}$$

定義 2.5 左右端點 k 階區間移動平均法(Hsu 2008)

$$\text{令 } \widehat{m}_t = \frac{m_{t-1} + \dots + m_{t-k}}{k}, \widehat{l}_t = \frac{l_{t-1} + \dots + l_{t-k}}{k}, \widehat{u}_t = \frac{u_{t-1} + \dots + u_{t-k}}{k}, t = k+1, \dots$$

$$\text{則 } \widehat{X}_t = E[X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-k}] = (\widehat{m}_t, \widehat{l}_t, \widehat{u}_t)_{LR} \circ$$

定義 2.6 左右端點 k 階加權區間移動平均法(Hsu 2008)

$$\text{令 } \widehat{m}_t = \sum_{i=t-k}^{t-1} p_i m_i, \widehat{l}_t = \sum_{i=t-k}^{t-1} p_{Li} l_i, \widehat{u}_t = \sum_{i=t-k}^{t-1} p_{Ri} u_i, t=k+1, \dots$$

$$\text{其中 } p_i = \frac{f_i}{\sum_{j=t-k}^{t-1} f_j}, f_i = \frac{\|\Omega\|}{\|X_i\|}, p_{Li} = \frac{\|X_i\|_L}{\sum_{j=t-k}^{t-1} \|X_i\|_L}, p_{Ri} = \frac{\|X_i\|_R}{\sum_{j=t-k}^{t-1} \|X_i\|_R},$$

$$\text{其中 } i = t-k, t-k+1, \dots, t-1, \Omega = \left[\min_{t-k \leq j \leq t-1} \{a_j\}, \max_{t-k \leq j \leq t-1} \{b_j\} \right],$$

$$\text{則 } \widehat{X}_t = E[X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-k}] = (\widehat{m}_t, \widehat{l}_t, \widehat{u}_t)_{LR}$$

例 2.2 令 $\{X_t\} = \{[1,2], [2,4], [3,4], [4,6], [3,7]\}$ 且 $l = u$,

$$X_1 = (1.5, 0.5, 0.5)_{LR}, X_2 = (3, 1, 1)_{LR}, X_3 = (3.5, 0.5, 0.5)_{LR},$$

$$X_4 = (5, 1, 1)_{LR}, X_5 = (5, 2, 2)_{LR},$$

根據**定義 2.5**，若以左右端點 5 階區間移動平均法，可得

$$\begin{aligned} \widehat{X}_6 &= \left(\frac{1.5 + 3 + 3.5 + 5 + 5}{5}, \frac{0.5 + 1 + 0.5 + 1 + 2}{5}, \frac{0.5 + 1 + 0.5 + 1 + 2}{5} \right)_{LR} \\ &= (3.6, 1, 1)_{LR} = [3.6 - 1, 3.6 + 1] = [2.6, 4.6] \end{aligned}$$

$$\text{因爲 } \Omega = [1, 7], \text{ 所以 } \|\Omega\| = 6, f_1 = \frac{6}{1}, f_2 = \frac{6}{2}, f_3 = \frac{6}{1}, f_4 = \frac{6}{1}, f_5 = \frac{6}{4},$$

可得

$$p_1 = \frac{\frac{6}{1}}{\frac{6}{1} + \frac{6}{2} + \frac{6}{1} + \frac{6}{2} + \frac{6}{4}} = 0.31, p_2 = 0.15, p_3 = 0.31, p_4 = 0.15, p_5 = 0.08$$

則

$$\widehat{m}_6 = 0.31 \times 1.5 + 0.15 \times 3 + 0.31 \times 3.5 + 0.15 \times 5 + 0.08 \times 5 = 2.79,$$

因為假設 $l = u$,

$$\text{所以 } p_{L1} = p_{R1} = \frac{0.5}{0.5 + 1 + 0.5 + 1 + 2} = 0.1, p_{L2} = p_{R2} = 0.2,$$

$$p_{L3} = p_{R3} = 0.1, p_{L4} = p_{R4} = 0.2, p_{L5} = p_{R5} = 0.4,$$

則

$$\hat{I}_6 = \hat{u}_6 = 0.1 \times 0.5 + 0.2 \times 1 + 0.1 \times 0.5 + 0.2 \times 1 + 0.4 \times 2 = 1.3,$$

根據**定義 2.6**，若以左右端點 k 階加權區間移動平均法，可得

$$\hat{X}_6 = (2.79, 1.3, 1.3)_{LR} = [1.69, 4.09]$$

2.3 預測效率評估

預測結果好壞，是預測者最關心的問題，在一般的點預測中有許多評估標準，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。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(Wu 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文基礎，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，藉此分析最後的預測結果。

定義 2.7 反模糊值(Wu 2012)

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，則 $[a_t, b_t]$ 之反模糊值為

$$c_t + \frac{|b_t - a_t|}{2 \ln(e + c_t)}$$

其中 c_t 為中心點， e 為自然指數。其中 \ln 內項我們取 $e + c_t$ ，因為 c_t 若小於1， $\ln(c_t)$ 無意義，先會考慮 $\ln(1 + c_t)$ ，但是對於太小的 c_t ，如 $1 < c_t < 2$ ， $\ln(1 + c_t)$ 仍然過小。

定義 2.8 區間距離(distance of interval) (Wu 2012)

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，預測區間為 $\hat{X}_t = [\hat{a}_t, \hat{b}_t] = [\hat{c}_t; \hat{r}_t]$ ，則區間距離

$$d(X_t, \hat{X}_t) = |c_t - \hat{c}_t| + \left| \frac{|b_t - a_t|}{2 \ln(e + c_t)} - \frac{|\hat{b}_t - \hat{a}_t|}{2 \ln(e + \hat{c}_t)} \right|$$

定義 2.9 區間平均誤差(interval mean error) (Hung and Wu 2006)

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，預測區間為 $\hat{X}_t = [\hat{a}_t, \hat{b}_t] = [\hat{c}_t; \hat{r}_t]$ ， $\hat{\varepsilon}_t = d(X_t, \hat{X}_t)$ 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 k 代表往前預測期數。

例 2.3 大學生預測薪資如下表

往前期數	預測薪資	實際薪資
1	[3,4]	[3,6]
2	[2,6]	[4,5]

3	[3,5]	[2,6]
4	[5,7]	[5,8]
5	[4,5]	[3,8]

則根據**定義 2.7**、**定義 2.8** 可知反模糊化值：

$$X_1 = [3,4] = (3.5; 0.5), \hat{X}_1 = [3,6] = (4.5; 1.5)$$

$$d(X_1, \hat{X}_1) = |3.5 - 4.5| + \left| \frac{1}{2 \ln(e + 3.5)} - \frac{3}{2 \ln(e + 4.5)} \right| = 1.49$$

$$X_2 = [2,6] = (4; 2), \hat{X}_2 = [4,5] = (4.5; 0.5)$$

$$d(X_2, \hat{X}_2) = |4 - 4.5| + \left| \frac{4}{2 \ln(e + 4)} - \frac{1}{2 \ln(e + 4.5)} \right| = 0.77$$

$$X_3 = [3,5] = (4; 1), \hat{X}_3 = [2,6] = (4; 2)$$

$$d(X_3, \hat{X}_3) = |4 - 4| + \left| \frac{2}{2 \ln(e + 4)} - \frac{4}{2 \ln(e + 4)} \right| = 0.52$$

$$X_4 = [5,7] = (6; 1), \hat{X}_4 = [5,8] = (6.5; 1.5)$$

$$d(X_4, \hat{X}_4) = |6 - 6.5| + \left| \frac{2}{2 \ln(e + 6)} - \frac{3}{2 \ln(e + 6.5)} \right| = 0.98$$

$$X_5 = [4,5] = (4.5; 0.5), \hat{X}_5 = [3,8] = (5.5; 2.5)$$

$$d(X_5, \hat{X}_5) = |4.5 - 5.5| + \left| \frac{1}{2 \ln(e + 4.5)} - \frac{5}{2 \ln(e + 5.5)} \right| = 1.93$$

$$\text{則根據 } \mathbf{定義 2.9}, \text{IME} = \frac{1}{5}(1.49 + 0.77 + 0.52 + 0.98 + 1.93) = 1.138$$

上面概述了常用的區間模糊時間數列，以及區間效率分析的相關內容，下一章將闡述本研究所提出的區間預測方法為「加權模糊時間數列法」。

3. 研究方法

下面我們推廣 Pathak and Singh (2011) 的加權時間數列預測方法，成為加權模糊時間數列以進行區間預測。

3.1 加權模糊時間數列法

本節我們提出加權模糊時間數列法，根據 Pathak and Singh (2011)，以區間中出現的次數多少來決定權重：

步驟 1：定義論域 $U = \{D_{max}, D_{min}\}$ ，並且將其分割七個長度相等的區間，

以表 3.1 為例， $U = [13\ 000, 20\ 000]$ ， $u_1 = [13\ 000, 14\ 000]$ ，

$u_2 = [14\ 000, 15\ 000]$ ， $u_3 = [15\ 000, 16\ 000]$ ，

$u_4 = [16\ 000, 17\ 000]$ ， $u_5 = [17\ 000, 18\ 000]$ ，

$u_6 = [18\ 000, 19\ 000]$ ， $u_7 = [19\ 000, 20\ 000]$

步驟 2：分別計算七個區間中，每一個區間發生的次數，如表 3.2。

步驟 3：再重新分割區間長度：次數發生最多的區間，將此區間再細分為四

等分；找出次數發生第二多的區間，將此區間細分為三等分；找出次

數發生第三多的區間，將此區間細分為二等分，並且將沒有次數發生

的區間捨棄。於是我們得到新區間為：

$u_1 = [13\ 000, 13\ 500]$ ， $u_2 = [13\ 500, 14\ 000]$ ，

$u_3 = [14\ 000, 15\ 000]$ ， $u_4 = [15\ 000, 15\ 250]$ ，

$u_5 = [15\ 250, 15\ 500]$ ， $u_6 = [15\ 500, 15\ 750]$ ，

$u_7 = [15\ 750, 16\ 000]$ ， $u_8 = [16\ 333, 16\ 667]$ ，

$u_9 = [16\ 667, 17\ 000]$ ， $u_{10} = [18\ 000, 18\ 500]$ ，

$u_{11} = [18\ 500, 19\ 000]$ ， $u_{12} = [19\ 000, 20\ 000]$

步驟 4：定義重新分割之後的模糊集合，並給予對應的語言變數，以及隸屬
度函數，為了簡化起見，我們不列出 membership value 為 0 的部分，
如下：(語言變數 A_i 呈現於表 3.1)

$$A_1 = \text{very}^4 \text{ few} = \frac{1}{u_1} + \frac{0.5}{u_2}, A_2 = \text{very}^3 \text{ few} = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3}$$

$$A_3 = \text{very}^2 \text{ few} = \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4}, A_4 = \text{very few} = \frac{0.5}{u_3} + \frac{1}{u_4} + \frac{0.5}{u_5}$$

$$A_5 = \text{few} = \frac{0.5}{u_4} + \frac{1}{u_5} + \frac{0.5}{u_6}, A_6 = \text{moderate} = \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7}$$

$$A_7 = \text{many} = \frac{0.5}{u_6} + \frac{1}{u_7} + \frac{0.5}{u_8}, A_8 = \text{very many} = \frac{0.5}{u_7} + \frac{1}{u_8} + \frac{0.5}{u_9}$$

$$A_9 = \text{too many} = \frac{0.5}{u_8} + \frac{1}{u_9} + \frac{0.5}{u_{10}}, A_{10} = \text{too many}^2 = \frac{0.5}{u_9} + \frac{1}{u_{10}} + \frac{0.5}{u_{11}}$$

$$A_{11} = \text{too many}^3 = \frac{0.5}{u_{10}} + \frac{1}{u_{11}} + \frac{0.5}{u_{12}}, A_{12} = \text{too many}^4 = \frac{0.5}{u_{11}} + \frac{1}{u_{12}}$$

步驟 5：建立模糊趨勢關係，若 $X_t \in A_i, X_{t+1} \in A_j$ ，則建立模糊關係 $A_i \rightarrow A_j$ 。將資料以模糊趨勢關係分為 12 群，建立模糊規則資料庫，如表 3.3。

步驟 6：建立模糊權重，The fuzzy output 是根據之前幾年的 fuzzy input 而來。

Case 1：若第 n 年的 fuzzy input 是 A_i ，則第 n+1 年的 fuzzy output 將是 A_i ，(如 1971, 1972, ...)。

Case2：若第 n 年的 fuzzy input 是 A_i ，而且從之前的幾年，我們得到更多的模糊趨勢關係 $A_i \rightarrow A_j, A_i \rightarrow A_k, \dots$ ，則第 n+1 年的 fuzzy output 將是 (A_j, A_k, \dots) ，(如 1983, 1984, 1988)。

步驟 7：Output values 即是 fuzzy output 所在區間的中心點，若 fuzzy output 兩個以上，則取 fuzzy output 所在區間的中心點的平均數，如 1983 年 fuzzy output $A_5 = [15\ 250, 15\ 500]$ 及 $A_6 = [15\ 500, 15\ 750]$ ，則 output values 為兩個區間的中心點的平均數，即是 15 500。

步驟 8：預測，圖 3.1 為實際值與預測值的走勢圖。

表 3.1 Historical data (Pathak and Singh 2011)

Year	Actual data	Fuzzy input	Fuzzy output	calculated enrollment
1971	13,055	A1		
1972	13,563	A2	A1	13,250
1973	13,867	A2	A2	13,750
1974	14,696	A3	A2	13,750
1975	15,460	A5	A3	14,500
1976	15,311	A5	A5	15,375
1977	15,603	A6	A5	15,375
1978	15,861	A7	A6	15,625
1979	16,807	A9	A7	15,875
1980	16,919	A9	A9	16,833
1981	16,388	A8	A9	16,833
1982	15,433	A5	A8	16,500
1983	15,497	A5	A5,A6	15,500
1984	15,145	A4	A5,A6	15,500
1985	15,163	A4	A4	15,125
1986	15,984	A7	A4	15,125
1987	16,859	A9	A9	16,833
1988	18,150	A10	A8,A9	16,667
1989	18,970	A11	A10	18,125
1990	19,328	A12	A11	18,750
1991	19,337	A12	A12	19,500
1992	18,876	A11	A12	19,500

表 3.2 Frequency of data (Pathak and Singh 2011)

區間	次數
[13,000, 14,000]	3
[14,000, 15,000]	1
[15,000, 16,000]	9
[16,000, 17,000]	4
[17,000, 18,000]	0
[18,000, 19,000]	3
[19,000, 20,000]	2

表 3.3 模糊規則資料庫(Pathak and Singh 2011)

Serial No.	Fuzzy logical relationship groups(模糊關係群)
1	A1→A2(1971→1972)
2	A2→A2(1972→1973) , A2→A3(1973→1974)
3	A3→A5(1974→1975)
4	A4→A4(1984→1985) , A4→A7(1985→1986)
5	A5→A4(1983→1984) , A5→A5(1975→1976 , 1982→1983) , A5→A6(1976→1977)
6	A6→A7(1977→1978)
7	A7→A9(1978→1979 , 1986→1987)
8	A8→A5(1981→1982)
9	A9→A9(1979→1980) , A9→A8(1980→1981) , A9→A10(1987→1988)
10	A10→A11(1988→1989)
11	A11→A12(1989→1990)
12	A12→A11(1991→1992) , A12→A12(1990→1991)

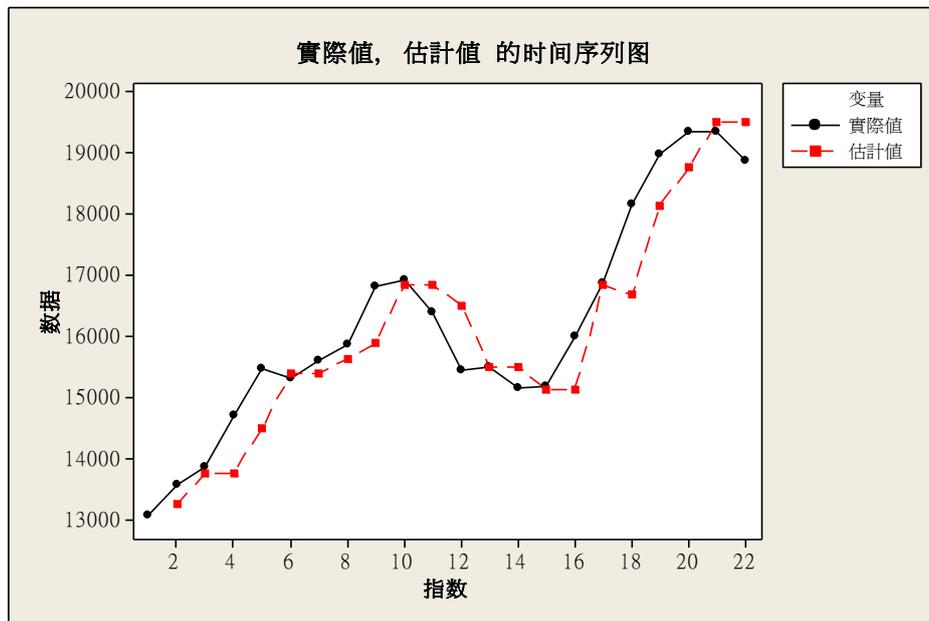


圖 3.1. 時間數列之實際值與估計值走勢圖

3.2 加權模糊時間數列法

我們將加權時間數列法，推廣成加權模糊時間數列法：

$\{X_t = [a_t, b_t], t = 1, 2, \dots, n\}$ 為一個區間時間數列，其預測的步驟如下：

步驟 1：令左端點成一時間數列 $\{a_t\}_{t=1}^n$ ，並進行一次加權模糊時間數列法得

預測值 \hat{a}_{t+1} 。

步驟 2：令右端點成一時間數列 $\{b_t\}_{t=1}^n$ ，並進行一次加權模糊時間數列法得

預測值 \hat{b}_{t+1} 。

步驟 3：得預測區間 $\hat{X}_{t+1} = [\hat{a}_{t+1}, \hat{b}_{t+1}]$ 。

模式已經建構完成，下一章我們將以加權模糊時間數列及五日、七日、十日移動平均線預測法，預測每日台股的最高價與最低價所形成的區間，並探討各種預測方法的效率及誤差。

4.實證分析

4.1 資料來源

資料來源為台灣證券交易所，收集 2011/07/01 至 2011/11/21 共 121 筆台灣上市公司股價指數每日的最高價、最低價，並且考慮以第三章的預測模式，來預測 2011/11/22 至 2011/11/24 這三期的最高價、最低價所構成的區間，先將這 121 筆資料標記為 $\{X_t\}_{t=1}^{121}$ ，下圖為走勢圖

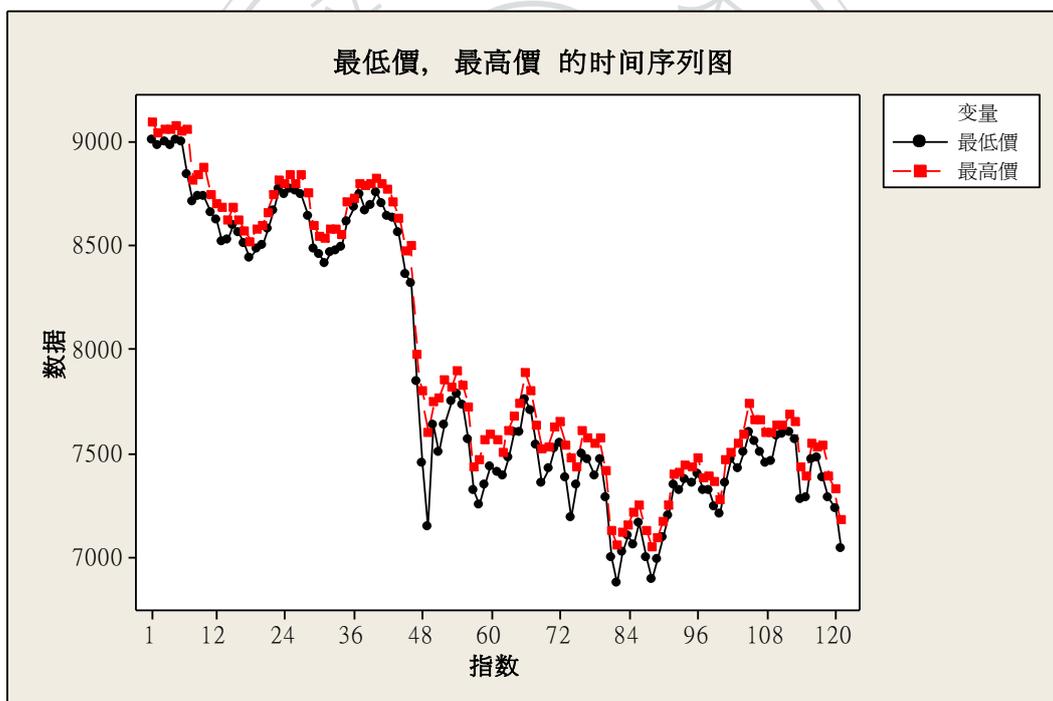


圖 4.1 2011/07/01 至 2011/11/21 台股上市股價每日最高價、最低價走勢圖

我們先使用 121 期來預測 3 期，預測方式如下：

4.2 加權模糊時間數列法

(1)最低價

已知期間內，最低價落在 $[6877.12, 9007.53]$ 之間，我們將論域 U 設定為

[6800,9040]，先分爲七個區間，並計算各區間發生的次數，如表 4.1：

表 4.1 最低價的原分割區間的各區間發生的次數

區間	次數
[6800, 7120]	10
[7120, 7440]	32
[7440, 7760]	31
[7760, 8080]	2
[8080, 8400]	2
[8400, 8720]	28
[8720, 9040]	16

之後根據區間所發生的次數作新的分割，第一多次數的區間爲[7120, 7440]，再分割四等分：[7120, 7200]，[7200, 7280]，[7280, 7360]，[7360, 7440]；第二多次數的區間爲[7440, 7760]，再分割爲三等分：[7440, 7546]，[7546, 7653]，[7653, 7760]；第三多次數的區間爲[8400, 8720]，再分割爲二等分：[8400, 8560]，[8560, 8720]。得到新的分割，如表 4.2：

表 4.2 最低價的新分割區間的各區間發生的次數

區間	次數
A1=[6800, 7120]	10
A2=[7120, 7200]	4
A3=[7200, 7280]	5
A4=[7280, 7360]	13
A5=[7360, 7440]	10
A6=[7440, 7546]	15
A7=[7546, 7653]	12
A8=[7653, 7760]	4
A9=[7760, 8080]	2

A10=[8080, 8400]	2
A11=[8400, 8560]	14
A12=[8560, 8720]	14
A13=[8720, 9040]	16

則模糊趨勢關係如表 4.3：

表 4.3 最低價的模糊趨勢關係

Date	Actual data	Fuzzyinput	Fuzzy output	calculated enrollment
07/01	8999.84	A13		
07/02	8977.67	A13	A13	8880
07/03	8997.35	A13	A13	8880
...
09/28	7099.73	A1	A1	6960
09/29	7061.01	A1	A1	6960
09/30	7163.75	A2	A1	6960
10/03	6998.37	A1	A1, A2	7120
...
11/21	7042.64	A1	A1, A2	7120

由於 2011/11/21 的真實值 7042.64 屬於 A1，再由資料顯示 A1 的模糊趨勢關係為 A1→A1 (09/28→09/29)、A1→A2 (09/29→09/30)，所以我們選擇 A1=[6800, 7120]以及 A2=[7120, 7200]兩個區間中點的平均數 7120 為 \hat{a}_{122} 的預測值；又由於 2011/11/22 的真實值 6970.72 屬於 A1，同理，7120 為 \hat{a}_{123} 的預測值；又由於 2011/11/23 的真實值 6806.43 屬於 A1，同理，7120 為 \hat{a}_{124} 的預測值。故最低價三期的預測值分別為：

$$\hat{a}_{122} = 7120, \hat{a}_{123} = 7120, \hat{a}_{124} = 7120$$

(2)最高價

已知期間內，最高價落在[7047.99,9089.47]之間，我們將論域 U 設定為 [7000,9100]，將論域先分為七個區間，並計算各區間發生的次數，如表 4.4：

表 4.4 最高價的原分割區間的各區間發生的次數

區間	次數
[7000, 7300]	13
[7300, 7600]	36
[7600, 7900]	25
[7900, 8200]	1
[8200, 8500]	2
[8500, 8800]	30
[8800, 9100]	14

之後根據區間所發生的次數作新的分割，第一多次數的區間為[7300, 7600]，再分割四等分：[7300, 7375]，[7375, 7450]，[7450, 7525]，[7525, 7600]；第二多次數的區間為[8500, 8800]，再分割為三等分：[8500, 8600]，[8600, 8700]，[8700, 8800]；第三多次數的區間為[7600, 7900]，再分割為二等分：[7600, 7750]，[7750, 7900]。得到新的分割為表 4.5：

表 4.5 最高價的新分割區間的各區間發生的次數

區間	次數
A1=[7000, 7300]	13
A2=[7300, 7375]	2
A3=[7375, 7450]	12
A4=[7450, 7525]	7
A5=[7525, 7600]	15
A6=[7600, 7750]	16

A7=[7750, 7900]	9
A8=[7900, 8200]	1
A9=[8200, 8500]	2
A10=[8500, 8600]	10
A11=[8600, 8700]	6
A12=[8700, 8800]	14
A13=[8800, 9100]	14

則模糊趨勢關係如表 4.6：

表 4.6 最高價的模糊趨勢關係

Date	Actual data	Fuzzy input	Fuzzy output	calculated enrollment
07/01	9089.47	A13		
07/02	9036.27	A13	A13	8950
07/03	9052.18	A13	A13	8950
...
10/06	7170.92	A1	A1	7150
10/07	7253.09	A1	A1	7150
10/11	7398.71	A2	A1	7150
...
11/21	7183.47	A1	A1, A2	7300

由於 2011/11/21 的真實值 7183.47 屬於 A1，再由資料顯示 A1 的模糊趨勢關係為 A1→A1(10/06→10/07)、A1→A2 (10/07→10/11)，所以我們選擇 A1=[7000, 7300]以及 A2=[7300, 7375]兩個區間中點的平均數 7300 為 \hat{b}_{122} 的預測值；又由於 2011/11/22 的真實值 7082.4 屬於 A1，同理，7300 為 \hat{b}_{123} 的預測值；又由於 2011/11/23 的真實值 6981.54 屬於 A1，同理，7300 為 \hat{b}_{124} 的預測值。故最高價三期的預測值分別為：

$$\hat{b}_{122} = 7300, \hat{b}_{123} = 7300, \hat{b}_{124} = 7300$$

所以，

$$\hat{X}_{122} = [\hat{a}_{122}, \hat{b}_{122}] = [7120, 7300] = (7210; 90)$$

$$\hat{X}_{123} = [\hat{a}_{123}, \hat{b}_{123}] = [7120, 7300] = (7210; 90)$$

$$\hat{X}_{124} = [\hat{a}_{124}, \hat{b}_{124}] = [7120, 7300] = (7210; 90)$$

4.3 左右端點 k 階區間移動平均法

接下來以五日、七日、十日移動平均線，分別對最低價、最高價作區間的預測：

(1) 五日移動平均線法，以下為 2011/06/01 至 2011/11/21 的每日最低價的五日平均線，由 Minitab 軟體預測出

$$\hat{a}_{122} = 7284.24, \hat{a}_{123} = 7183.31, \hat{a}_{124} = 7067.57$$

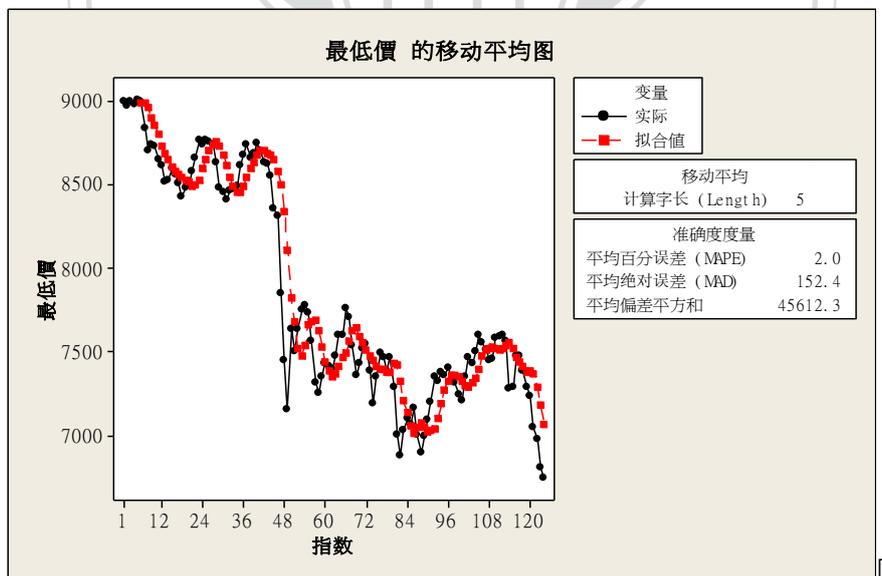


圖 4.2 最低價的五日移動平均線

接著 2011/06/01 至 2011/11/21 的每日最高價的五日平均線，由 Minitab 軟體預測出

$$\hat{b}_{122} = 7394.88, \hat{b}_{123} = 7305.61, \hat{b}_{124} = 7193.58$$

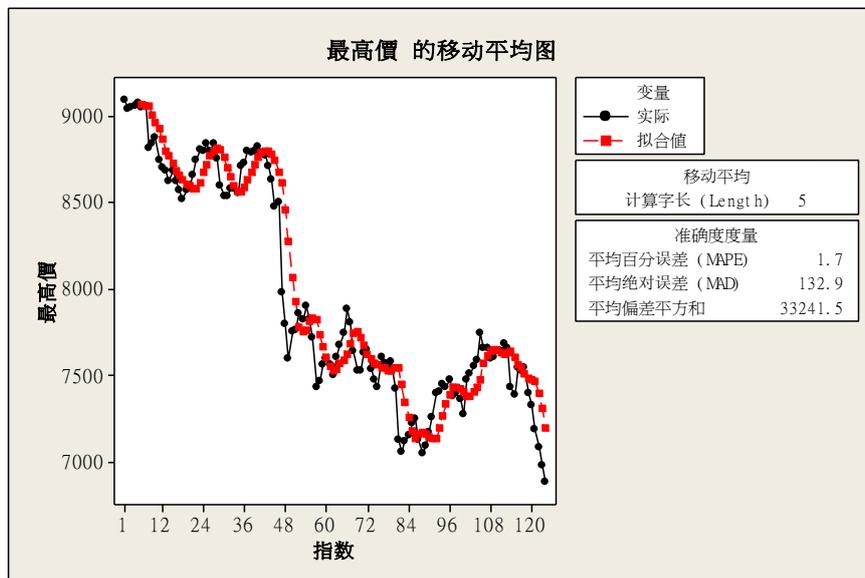


圖 4.3 最高價的五日移動平均線

所以，以五日平均線預測法，得到：

$$\hat{X}_{122} = [\hat{a}_{122}, \hat{b}_{122}] = [7284.24, 7394.88] = (7339.56; 55.32)$$

$$\hat{X}_{123} = [\hat{a}_{123}, \hat{b}_{123}] = [7183.31, 7305.61] = (7244.46; 61.15)$$

$$\hat{X}_{124} = [\hat{a}_{124}, \hat{b}_{124}] = [7067.57, 7193.58] = (7130.575; 63.005)$$

(2) 七日移動平均線法，以下為 2011/06/01 至 2011/11/21 的每日最低價的七日

平均線，由 Minitab 軟體預測出

$$\hat{a}_{122} = 7311.69, \hat{a}_{123} = 7266.29, \hat{a}_{124} = 7171.19$$

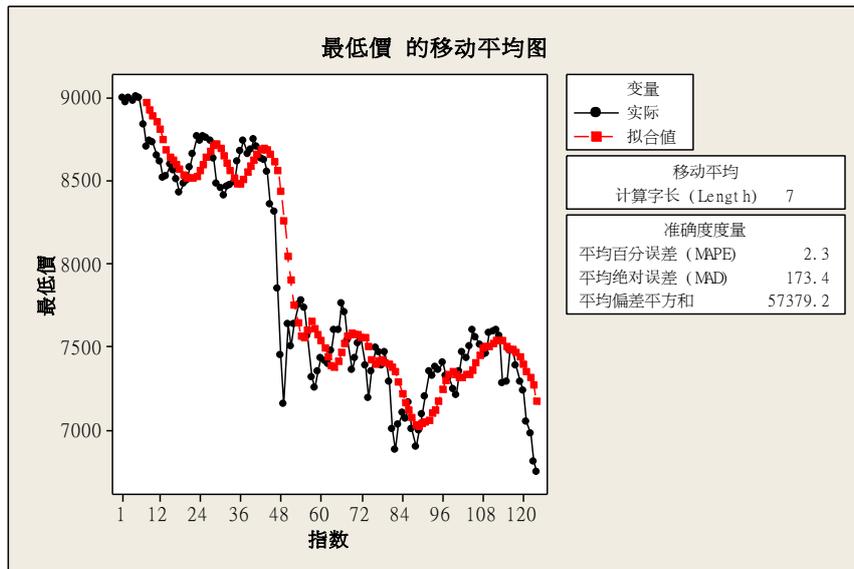


圖 4.4 最低價的七日移動平均線

接著 2011/06/01 至 2011/11/21 的每日最高價的七日平均線，由 Minitab 軟體預測出

$$\hat{b}_{122} = 7414.95, \hat{b}_{123} = 7371.58, \hat{b}_{124} = 7291.19$$

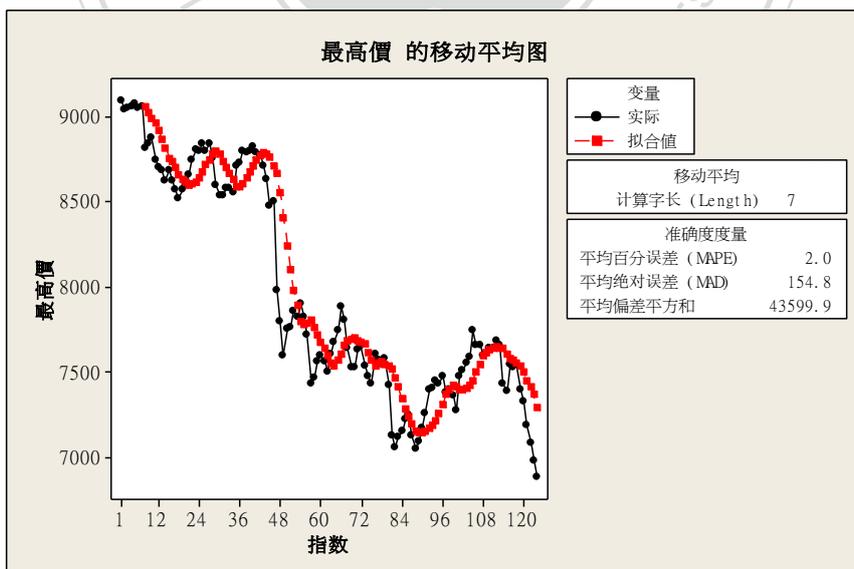


圖 4.5 最高價的七日移動平均線

所以，以七日平均線預測法，得到：

$$\hat{X}_{122} = [\hat{a}_{122}, \hat{b}_{122}] = [7311.69, 7414.95] = (7363.32; 51.63)$$

$$\hat{X}_{123} = [\hat{a}_{123}, \hat{b}_{123}] = [7266.29, 7371.58] = (7318.935; 52.645)$$

$$\hat{X}_{124} = [\hat{a}_{124}, \hat{b}_{124}] = [7171.19, 7291.19] = (7231.19; 60)$$

(3) 十日移動平均線法，以下為 2011/06/01 至 2011/11/21 的每日最低價的十日平均線，由 Minitab 軟體預測出

$$\hat{a}_{122} = 7361.93, \hat{a}_{123} = 7298.93, \hat{a}_{124} = 7223.38$$

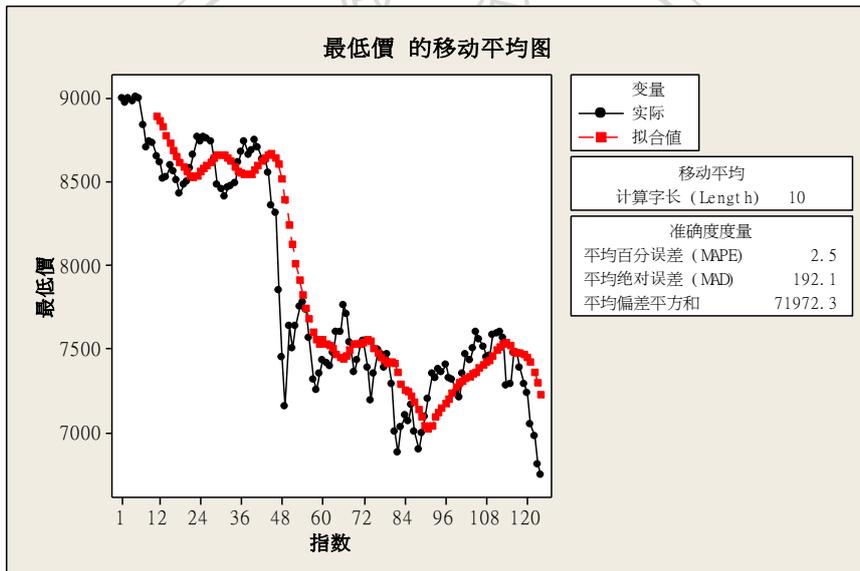


圖 4.6 最低價的十日移動平均線

接著 2011/06/01 至 2011/11/21 的每日最高價的十日平均線，由 Minitab 軟體預測出

$$\hat{b}_{122} = 7467.62, \hat{b}_{123} = 7407.55, \hat{b}_{124} = 7340.29$$

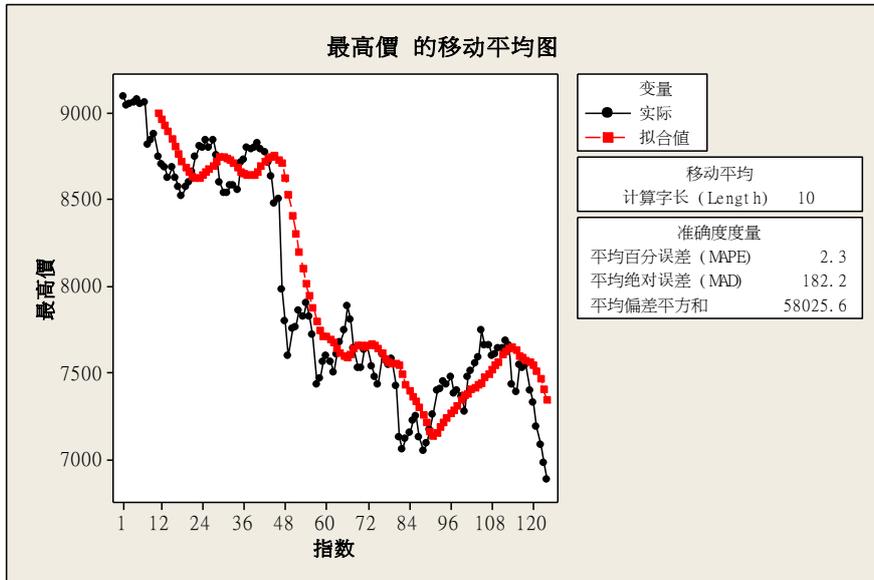


圖 4.7 最高價的十日移動平均線

所以，以十日平均線預測法，得到：

$$\hat{X}_{122} = [\hat{a}_{122}, \hat{b}_{122}] = [7361.93, 7467.62] = (7414.775; 52.845)$$

$$\hat{X}_{123} = [\hat{a}_{123}, \hat{b}_{123}] = [7298.93, 7407.55] = (7353.24; 54.31)$$

$$\hat{X}_{124} = [\hat{a}_{124}, \hat{b}_{124}] = [7223.38, 7340.29] = (7281.835; 58.455)$$

4.4 比較「加權模糊時間數列法」及「左右端點 k 階區間移動平均法」的測量誤差：

	加權模糊時間 數列法預測區 間	左右端點 5 日 區間移動平均 法	左右端點 7 日 區間移動平均 法	左右端點 10 日 區間移動平均 法
X_{122} = [6970,7082]	\hat{X}_{122} = [7120,7300]	\hat{X}_{122} = [7284,7394]	\hat{X}_{122} = [7311,7414]	\hat{X}_{122} = [7361,7467]
X_{123} = [6806,6981]	\hat{X}_{123} = [7120,7300]	\hat{X}_{123} = [7183,7305]	\hat{X}_{123} = [7266,7371]	\hat{X}_{123} = [7298,7407]
X_{124} = [6744,6885]	\hat{X}_{124} = [7120,7300]	\hat{X}_{124} = [7067,7193]	\hat{X}_{124} = [7171,7291]	\hat{X}_{124} = [7223,7340]
IME	300.3	325.1	390.8	436.3

我們發現「加權模糊時間數列法」的誤差比「左右端點 k 階區間移動平均法」小的許多。

5. 結論

本文應用加權模糊時間數列法，制定一個新的區間模糊時間數列來預測台股上市公司的最低價、最高價，本研究模型和其他研究者最大的差異在於，利用美日台股上市公司指數的波動振盪區間，找出其最低價及最高價，建立模糊趨勢，以達區間預測的功效。此亦符合現實的股市操作現場：由於台股有漲停板與跌停板的設定範圍，所以每日的漲幅或跌幅有一定的範圍在，許多股市分析師往往以收盤價及均線來預測股價指數，但遇到股價跌跌不休、或是接連上漲時，平均線往往落後於現實值。然而，本研究方法，得到更高的預測效率。由於區間預測是由傳統的點預測所發展出來的預測方式，既可以解決點預測無法處理的問題，又不違背傳統點預測的精神。

除了股價的問題適合用區間預測的概念之外，還有許多自然現象，如：每日氣溫的走勢，社會現象，如：房價、外匯市場、黃金市場也適合。但對於預測分割區間的選取，則會因為預測期間的長短而略有不同，意即不同預測其間會有不同的分割區間長度。再者，因為影響台股的因素，除了可衡量的經濟因素之外，尚有政治因素、政府干預、投資者的預期心理變數等等，這些皆是無法量化的變數，因此在未來可找出更有效率的方式將其量化。

參考目錄

- [1] 吳柏林 2005 模糊統計導論與應用。五南書局。
- [2] 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。
- [3] 吳柏林 1995 時間數列分析導論。華泰書局。
- [4] 林茂文 1992 時間序列分析與預測。華泰書局。
- [5] 林原宏 2006 模糊統計。五南書局。
- [6] 楊奕農 2009 時間序列分析-經濟與財務上之應用。雙葉書廊。
- [7] Chang, S. K. (2007). "On the Testing Hypotheses of Mean and Variance for Interval Data," *Management Science & Statistical Decision*, Vol. 4, No. 2, pp. 63-69.
- [8] Chatfield, C. (1993). "Calculating Interval Forecasts," *Journal and Business & Economic Statics*, Vol. 11, No. 2, pp. 121-135.
- [9] Chen, S. M. (1996). "Forecasting Enrollments Based on Fuzzy Time Series," *Fuzzy Sets and Systems*, Vol. 81, No. 3, pp. 311-319.
- [10] Chen, S. M. (2002). "Forecasting Enrollments Based on High Order Fuzzy Time Series," *Cybernetics and Systems: An International Journal*, Vol. 133, No. 1, pp. 1-16.
- [11] Chen, S. M. and Hsu, C. C. (2004). "A New Method to Forecast Enrollment Using Fuzzy Time Series," *International Journal of Applied Science and Engineering*, Vol. 3, No. 2, pp. 234-244.
- [12] Cheng, C. H., Chen, T. L., and Chiang C. H. (2006). "Trend-Weighted Fuzzy Time Series Model for TAIEX Forecasting," *Proceeding of the 13th International Conference on Neural Information Processing*, Part-III, Lecture Notes in Computer Science, Hong Kong, Vol. 4234, pp. 469-477.
- [13] Huarng, K. (2001). "Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series," *Fuzzy Sets and Systems*, Vol. 123, No. 3, pp.387-394.
- [14] Hsu, H. L. (2008). "Interval Time Series Analysis with Forecasting Efficiency Evaluation," Doctorial Thesis, Department of Mathematical Science, National Chengchi University, Taipei, Taiwan.
- [15] Hsu, Y.Y., Tse, S.M. and Wu, B. (2003). "A New Approach of Bivariate Fuzzy Time Series Analysis to the Forecasting of a Stock Index," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 11, No. 6, pp. 671-690.
- [16] Kreinovich, V., Nguyen H. T., and Wu, B. (2006). "On-Line Algorithms for

- Computing Mean and Variance of Interval Data, and their Use in Intelligent Systems," *Information Science*, Vol. 177, pp. 3228-3238.
- [17] Pathak, H. K. and Singh, P. (2011). "A New Bandwidth Interval Based Forecasting Method for Enrollments Using Fuzzy Time Series," *Scientific Research*, Vol. 2, pp. 504-507.
- [18] Song, Q. and Chissom, B. S. (1993). "Forecasting Enrollment with Fuzzy Time Series-Part I," *Fuzzy Sets and Systems*, Vol. 54, No. 1, pp. 1-9.
- [19] Tseng, F.-M. and Tzeng, G.-H. (2002). "A Fuzzy Seasonal ARIMA Model for Forecasting," *Fuzzy Sets and Systems*, Vol. 126, No. 3, pp. 367-376.
- [20] Wu, B. and Hung, S. (2006). "A Fuzzy Identification Procedure for Nonlinear Time Series with Example on ARCH and Bilinear Models," *Fuzzy Sets and Systems*, Vol. 108, pp. 275-287.
- [21] Zadeh, L. A. (1965). "Fuzzy Sets," *Information Control*, Vol. 8, No. 3, pp. 338-353.

