

台北市失業率之預測 - 模糊時間序列之應用

邱志洲* 李天行* 林凡群**

*輔仁大學企業管理學系

**輔仁大學應用統計研究所

(收稿日期：86 年 7 月 16 日；第一次修正：86 年 10 月 16 日；
接受刊登日期：86 年 11 月 5 日)

摘要

失業率是勞動市場的重要指標之一，它可以反映出短期的人力移動，並可以此做為人力資源供需的指標，提供許多有關人力資源管理的資訊和市場決策的重要依據。因此，若能精確的預測失業率，對於決策者將會有很大幫助。在本研究中，嘗試以模糊時間序列 (Fuzzy Time Series) 的分析方法建立短期失業率的預測模式，以期能更精確的預測這個重要指標。

我們將台北市從民國 72 年 1 月至民國 85 年 10 月的月失業率資料，作為實證的時間序列資料，同時利用模糊時間序列分析與 ARIMA 模型 (Autoregressive Integrated Moving Average Model) 建立短期失業率的預測模式，並將兩者之間的預測結果加以評估比較。由實證結果中顯示，本研究所使用的模糊時間序列模型顯著有較好的預測能力。

關鍵詞彙：時間序列，預測，失業率，模糊，人力資源

壹 緒論

失業率是勞動市場的重要指標之一，它可以反映出短期的人力移動，並可以此做為人力資源供需的指標，提供許多有關人力資源管理的資訊和市場決策的重要依據，影響的層面大可至一個國家的經濟政策，小至一個公司的人員需求政策。而失業人口的產生，表示在整個經濟體系中，無法有效的運用人力資源，形成了一種嚴重的資源浪費。大量的失業人口，對於整個社會所產生的各種嚴重負面影響，將造成許多無法估計的社會成本，例如犯罪率的增加、家庭問題的產生等等。若能預測出失業率的高低，將可提供決策者一些重要的訊息，並提早訂定出有效的因應政策。如此，應可避免人力資源的浪費及社會成本的增加。

本研究嘗試利用模糊時間序列方法針對台北市的月失業率資料，建構短期的失業率預測模式。以行政院主計處所公佈臺灣地區台北市自民國 72 年 1 月至民國 85 年 10 月間，共 166 筆的月失業率數據為實證的時間序列資料，並將模糊時間序列方法與 ARIMA 模型的預測結果做分析比較。希望能藉由

模糊時間序列方法中的無模式假設 (model free) 特性 (吳柏林 et al. 1996)，彌補傳統統計方法之不便。本論文共分為五部分：在第一部分中，說明了本研究的動機與目的；第二部分的重點，是介紹有關失業率預測與模糊時間序列方法的一些相關文獻探討。在第三部分中，則介紹模糊時間序列模型 (Fuzzy Time Series Model) 的基本架構和預測模式的建構流程。第四部分中，則以台灣地區台北市之失業率為實證資料，同時利用模糊時間序列模型和 ARIMA 模型建構預測模式，並加以分析評估。最後，在第五部分中，則根據實證研究中的結果提出說明。

貳 文獻探討

Sims (1980) 針對經濟預測模式提出一些批評，建議使用 VAR 模式 (Pure Vector Autoregressive Model)，並認為此模式可排除一些不符實際的限制，且估計過程較為簡單、清楚。而 Funke (1992) 則針對德國的失業率使用二種預測模式加以預測：一為單變量時間數列模式，僅就失業率資料本身進行探討；另一種方法，則多考慮了三種相關的經濟指標變數：工業生產指數 (Index of Total Industrial Production)、新訂貨量指數 (Index of the Volume of New Order) 與未來生產趨勢的抽樣資料 (Survey Data on the Future Tendency of Production)，利用這三種變數與失業率資料共同建立多變量時間數列模式。但不論是 VAR 模式抑或時間數列模式，其對於建立模式的變數與資料都需有很嚴謹的假設，如穩態假設等。若資料不符合假設，則需經過變數轉換或差分的處理方式，使資料能夠合乎預測模式的限制。這些模式的假設，使得決策者在使用上，需要面對更複雜的資料處理工作及犧牲更多重要的訊息。

吳柏林及陳雅玫 (1993) 曾針對臺灣地區失業率利用時空數列分析方法進行預測及討論，因其認為發生資料的時間和空間因素之重要性相同，若使用向量時間序列模式並無法討論系統中之空間相關 (Spatial Autocorrelation)，故考慮地區與地區間的時空動態關係。其假設台北市、高雄市與台灣省之間有地緣關係存在，故使用時空數列分析方法分析臺灣地區勞動市場的異動情形，並與單變量時間數列方法及狀態空間分析模式做比較，證明時空數列模式有較好的預測能力。文中也提到，不論是 Box-Jenkins 建立的時間序列分析模式或是上面所提到的時空數列方法，其資料均需具備可逆性 (Invertibility)

與穩定性 (Stationarity) 方可使用。所以，若使用無需樣本假設的無模式 (Model Free) 估計法，應可以消除一些建立預測模式的障礙。

自模糊集 (Fuzzy Set) 的理論架構被 Zadeh (1965) 提出後，它在理論和應用上均有顯著的成就。人們開始利用這個理論架構創造新的研究方法，希望藉此理論的特性來處理一些較無法量化的資料。Qiang 與 Chissom (1993) 以 Zadeh 之模糊理論為基礎，針對歷史性的時間序列資料提出新的預測方法，定義了模糊時間序列模型的基本架構，並以美國 Alabama 大學註冊資料為例，說明預測模式建構的流程與方法。而 Hwang 等人 (1995) 針對 Qiang 與 Chissom (1993) 的模糊時間序列模式中的資料處理方式及相關矩陣 (Relation Matrix) 的計算準則進行修改，以相同的實證資料建立預測系統，並證明了新模式有較佳的預測能力。

參 模糊時間序列模式

一、模糊集 (Fuzzy Set)

在傳統數學中，對「概念性」變數必須給予明確的定義，一個元素是否屬於某個集合，是絕對化的二值邏輯，也就是答案只有「是」或「否」。但在許多的實際狀況中，並不是所有的「概念性」變數都能有如此明確的定義。

模糊理論承認事物存在模糊性，但它在描述事物的模糊性時，並不模糊，而是力求精確。它利用一定的數值來表現模糊的程度，對於集合中的每一個元素賦予一個介於 0 和 1 之間的數值，來表示此元素對於該集合的所屬程度。若該數值等於 1 時，表示該元素完全屬於該集合；若為 0 時，則表示完全不屬於，用這種方法定義的集合即稱為模糊集合。(參照下頁表一)

二、模糊時間序列模型之介紹

本研究所使用的模糊時間序列模型是取自 Hwang 等人 (1995) 所提出修改過的模糊時間序列模型。在本小節中，將介紹模糊時間序列模式中的重要符號及預測模型的建構方法。

表一 普通集合與模糊集合的數學定義

<p>普通集合：</p> <p>設 X 為論域，A 是 X 的任意子集，x 是 X 中之任意元素</p> $C(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$ <p>模糊集合：</p> <p>設 U 為論域，A 是 U 的任意模糊子集，u 是 U 中之任意元素</p> $A = \left\{ \frac{\mu_A(u)}{u} \mid u \in U \right\}$ <p>稱為 A 的隸屬函數，$\mu_A(u_i)$ 稱為元素 u_i 的隸屬度 (Grade of Membership)，$0 \leq \mu_A(u_i) \leq 1$。</p>
--

(一) 定義符號

$Y(t)$ ：是一個模糊時間序列資料 ($t=0, 1, 2, \dots$)。

U ：是包含所有已處理時間序列資料的論域，也就是所有討論的資料均包含在其中。

u_i ：是將 U 平均分割成數個等距的區間 u_i ($i=0, 1, 2, \dots, n$)，可將其表為 $U = \{u_1, u_2, \dots, u_n\}$ ； $u_i(t)$ 則定義為 $Y(t)$ 的一個模糊集合。

$F(t)$ ：為 $Y(t)$ 的模糊時間序列 (Fuzzy Time Series)，定義表示為

$$F(t) = \{u_1(t), u_2(t), \dots, u_n(t)\}.$$

μ_A ：定義為模糊集 A 的隸屬函數 (Membership Function)，其數學對應關係可表示為 $\mu_A : U \rightarrow [0, 1]$ 。

A ：為論域 U 的模糊集 (Fuzzy Set)，定義為

$$A = \frac{\mu_A(u_1)}{u_1} + \frac{\mu_A(u_2)}{u_2} + \dots + \frac{\mu_A(u_n)}{u_n}$$

其中， $\mu_A(u_i)$ 定義為 u_i 在模糊集 A 中的隸屬度，對應關係為

$$\mu_A(u_i) \in [0, 1].$$

$C^W(t)$ ：為基準矩陣，且定義為 $C^W(t) = F(t-1)$ 。

$R^W(t, t-1)$ ：為模糊集合的模糊關係運算子 (Fuzzy Operator)，代表第 t 期與第 $(t-1)$ 期模糊集合之間的模糊關係 (Fuzzy Relationship)。

(二)模式的修改準則

在介紹模糊時間序列模式之前，先將 Hwang 等人 (1995) 提出的新模式中所修改的部份，整理出三個基本準則：

- (1)認為在整個時間序列資料中，當期的變化量與前幾期的變化量有關，且在前幾個有關的期數當中，以前一期的資料變動與當期的關係最為密切。
- (2)若前幾期的變化量有遞增的趨勢，則本期的變化量應該也會遞增；若遞減，則本期亦然。
- (3)以前一期的資料變化量為預測當期的標準，計算前幾期有關的期數與前一期的資料變化量之間的模糊關係 $R_w(t, t-1)$ (Fuzzy Relationships)，並以前一期的資料變動為一個預測標準，預測當期的資料變化量。

綜觀以上三個準則可知，新模式與原來的模糊時間序列模式最大的不同，在於模式建構前的資料處理：在修改過的模式中，輸入資料捨棄使用原始資料而改用序列資料的變動 (Variation)，也就是將資料先經過一階差分的處理後再放入預測模式中，藉此方法解決時間序列資料的非穩態問題 (Non-stationarity)。其最大的貢獻在於資料符合穩態假設 (參照表二) 時，會出現一些資料特質，有助於預測模式建立的穩定性與簡單化。

表二 穩態假設

穩態的假設：

平均值為常數，每一時期的期望值都相同，故資料可合併估計期望值。

變異數為常數，資料可合併估計變異數。

(三)模糊時間序列模式的介紹與應用

模糊時間序列模式的建立，主要是將時間序列資料運用模糊理論轉換為模糊集合，再利用模糊集合的隸屬度觀察前後期的相關趨勢，藉過去資料的變動趨勢確定預測模式，然後進行預測。其應用流程，詳述如下：

首先，計算所有原始資料的變動量，亦即對原始資料進行一階差分的運算處理，並找出最大的變動量 D_{MAX} 和最小的變動量 D_{Min} 。取適當的正整數 D_1 、 D_2 ，訂定預測模型的資料變動量討論區域 U ，使得 $U=[D_{Min}-D_1, D_{MAX}+D_2]$ 。並將討論區域 U 分割成若干個等寬的變量討論區間 u_1, u_2, \dots ，

u_n ，使得 $U = \{ u_1, u_2, \dots, u_n \}$ ，並且給定模糊集合 A 在 u_i 的隸屬函數為 $\mu_A(u_i)$ ，故模糊集合 A 可被表示為：

$$A = \frac{\mu_A(u_1)}{u_1} + \frac{\mu_A(u_2)}{u_2} + \dots + \frac{\mu_A(u_n)}{u_n} \quad \text{或}$$

$$A = \{ \mu_A(u_1), \mu_A(u_2), \dots, \mu_A(u_n) \}$$

其次，將經過差分處理的資料 $Y(t)$ ，利用隸屬函數轉換為模糊集合 $F(t)$ ，使得模糊集合 $F(t)$ 定義為：

$$F(t) = \frac{\mu_A(u_1(t))}{u_1} + \frac{\mu_A(u_2(t))}{u_2} + \dots + \frac{\mu_A(u_n(t))}{u_n} \quad \text{或}$$

$$F(t) = \{ \mu_A(u_1(t)), \mu_A(u_2(t)), \dots, \mu_A(u_n(t)) \}$$

若我們想要預測第 t 期的數據資料，必須先決定要使用過去多少期的歷史資料當作預測基準，也就是認定過去 w 期的歷史資料會對當期資料產生影響。此一變數 w 的決定，至今仍無確切的理論方法，通常會先使用不同的基準期數 w 來進行預測，進而比較不同 w 下所產生預測的平均誤差，最後選取具備較小預測誤差的參數 w 為最佳的預測模式。

在選取最佳的預測模式後，可以決定參數 w 的值。此時，若想預測第 t 期的模糊集合，則可以將第 t 期之前的 w 期當作預測的基期。將第 $(t-1)$ 期的模糊集合資料當作計算第 t 期模糊集合的標準矩陣，而前面共 $(W-1)$ 期的模糊集合則形成一個運算矩陣 (Operation Matrix)。

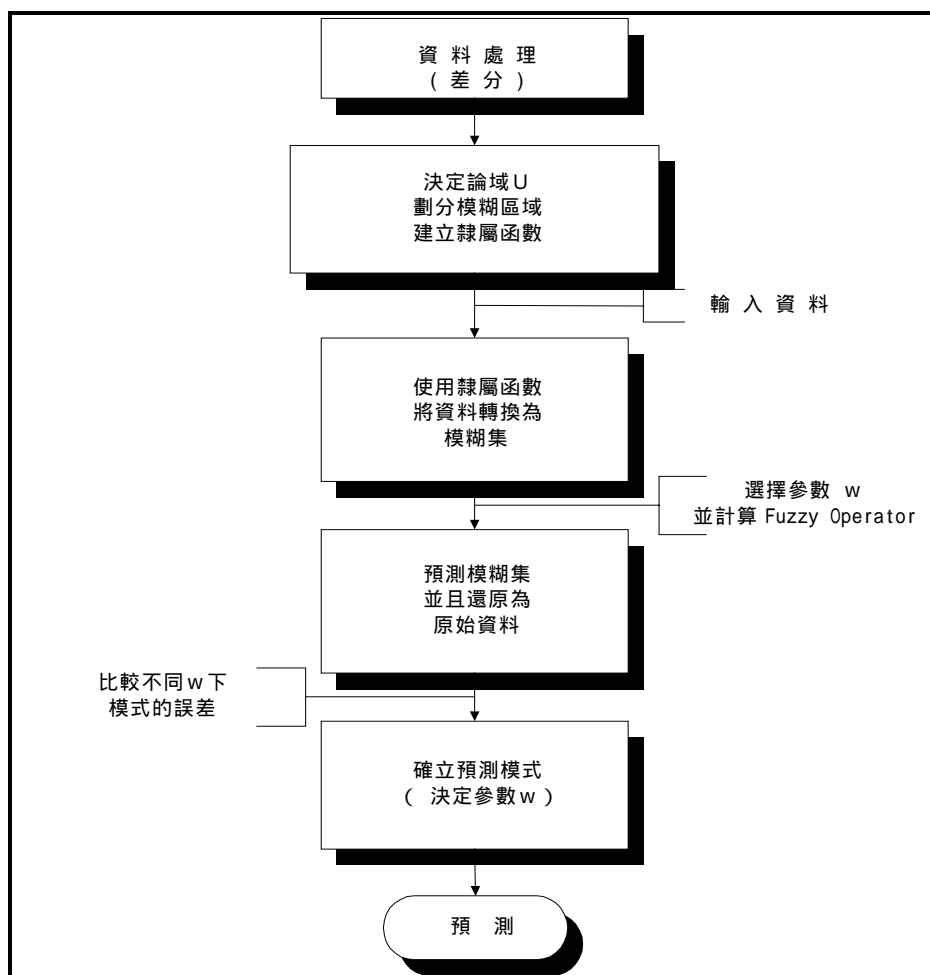
針對第 t 期而言，基準矩陣 $C^W(t)$ 為 $C^W(t) = F(t-1) = [u_{t-1,1}, u_{t-1,2}, \dots, u_{t-1,n}]$ ，運算矩陣 $O^W(t)$ 則為 $O^W(t) = [F(t-w), F(t-w+1), \dots, F(t-2)]$ 。其中，參數 n 為分割資料論域之區間數目。利用運算矩陣 $O^W(t)$ 與基準矩陣 $C^W(t)$ ，計算模糊時間序列第 t 期模糊集合相關矩陣 $R^W(t)$ ，並可將 $R^W(t)$ 展開如下

$$R^W(t) = O^W(t) \otimes C^W(t)$$

$$= \begin{bmatrix} u_{t-w,1} & u_{t-w,2} & \Lambda & u_{t-w,n} \\ u_{t-w+1,1} & u_{t-w+1,2} & \Lambda & u_{t-w+1,n} \\ M & M & O & M \\ u_{t-2,1} & u_{t-2,2} & \Lambda & u_{t-2,n} \end{bmatrix} \otimes [u_{t-1,1} \quad u_{t-1,2} \quad \Lambda \quad u_{t-1,n}]$$

$$= \begin{bmatrix} u_{t-w,1}u_{t-1,1} & u_{t-w,2}u_{t-1,2} & \Lambda & u_{t-w,n}u_{t-1,n} \\ u_{t-w+1,1}u_{t-1,1} & u_{t-w+1,2}u_{t-1,2} & \Lambda & u_{t-w+1,n}u_{t-1,n} \\ M & M & O & M \\ u_{t-2,1}u_{t-1,1} & u_{t-2,2}u_{t-1,2} & \Lambda & u_{t-2,n}u_{t-1,n} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \Lambda & R_{1n} \\ R_{21} & R_{22} & \Lambda & R_{2n} \\ M & M & O & M \\ R_{w-1,1} & R_{w-1,2} & \Lambda & R_{w-1,n} \end{bmatrix}$$

在以 $R^W(t)$ 為基礎下，我們可定義出第 t 期的預測模糊集合 $F(t)$ ，如下： $F(t)=[\text{MAX}(R_{11}, R_{21}, \dots, R_{w-1,1}), \text{MAX}(R_{12}, R_{22}, \dots, R_{w-1,2}), \dots, \text{MAX}(R_{1n}, R_{2n}, \dots, R_{w-1,n})]$ 。最後將模糊集合 $F(t)$ 還原為原始的資料格式時，即可以得到第 t 期的預測數據。綜合而言，模糊時間序列模式的建構流程可被表示為圖一：



圖一 模糊時間序列模式建構流程

肆 ARIMA模式分析技術的建構流程

一、ARIMA模式之介紹

ARIMA 分析方法是由 Box 及 Jenkins 於 1970 年所發展出來的，他們認為影響時間序列資料變動的主要因素可分為兩種：可藉由序列中的歷史資料來推論未來的趨勢，則稱此序列符合自我迴歸過程 (Autoregressive Process, AR Process)；若現期的不規則變異可以藉由過去的不規則變異所估計出，則稱此序列符合移動平均過程 (Moving Average Process, MA Process)。其模式可定義如下：

$$\Phi_p(B)(1-B)^d Y(t) = \mu + \theta_q(B)a(t)$$

其中

$Y(t)$ ：為一穩定 (Stationary) 狀態的時間數列值。

μ ：為常數項。

$a(t)$ ：為干擾項， $a(t)$ 符合白色噪音 (White Noise) 的假設，即期望值為零、變異數為一固定常數的假設。

上面所定義的模式可以 ARIMA(p,d,q)表之，其中 p、d 與 q 為非負整數，分別代表自我迴歸 (Autoregression) 的階數差距、差分之階數 (Order) 及移動平均 (Moving Average) 的階數差距，而 B^d 則稱為 d 階後移運算子 (Backward-shift Operator)，即 $B^d Y(t) = Y(t-d)$ 。 $\Phi_p(B)$ 及 $\theta_q(B)$ 則分別為 B 之多項式，可表示如下：

$$\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p$$

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

二、ARIMA模式之分析流程

Box 與 Jenkins 所提出的 ARIMA 模式建構程序為一種試誤遞迴過程 (Trial and Error Iterative Process) (如【圖二】)，其步驟說明如下：

步驟一、觀察資料結構

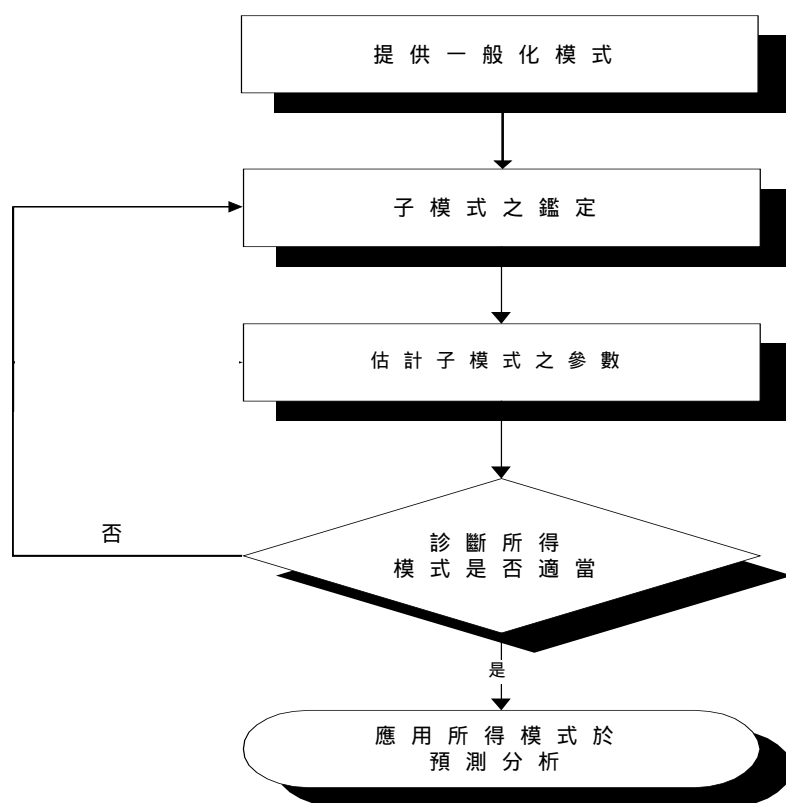
將時間序列資料依序排列，計算其自我相關函數 (Autocorrelation Function：ACF) 與偏自我相關函數 (Partial Autocorrelation Function：PACF)，並參考 ACF 與 PACF 圖的特性，考慮所擬採用的有用模式。

步驟二、提出暫定模式

利用較粗略的方法鑑定這些模式的子類型，來推測一個合適並合乎精簡原則之子類型模式為一暫定模式。

步驟三、模型的診斷與檢定

估計模式內的未知參數，並利用檢定方法判定所得的模式是否適當，若模式並不適當則必須重新鑑定、估計與診斷直到檢定至一適當模式為止。待適當模式被確認後，便可利用此模式加以預測。

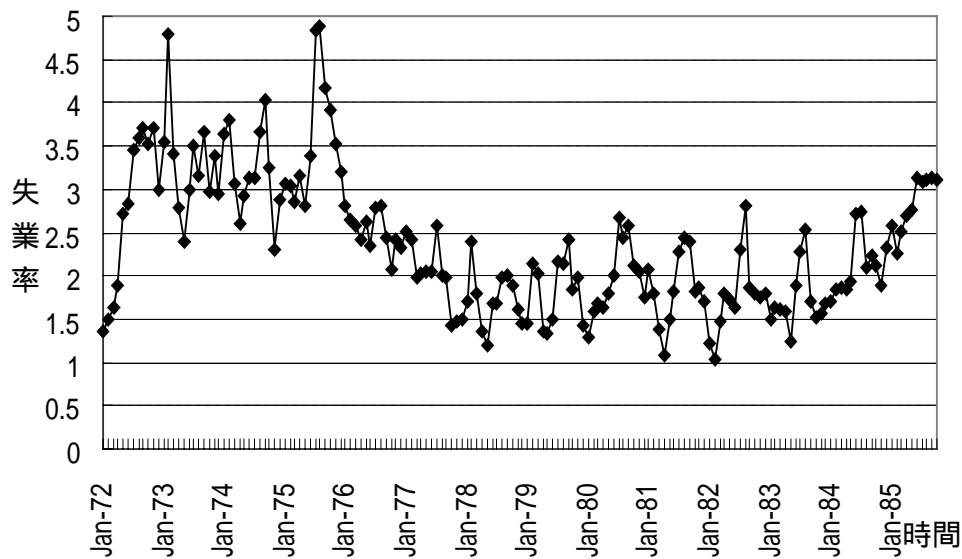


圖二 ARIMA 模式的建構流程

伍 實證研究

本文所使用的時間序列資料為行政院主計處所公佈的臺灣地區台北市自民國 72 年 1 月至民國 85 年 10 月間，共 166 筆的月失業率資料，資料分布趨勢如圖三。為了評估模糊時間序列模式之預測能力，我們將數列的前 144 筆

資料，分別應用模糊時間序列方法及 ARIMA 模型建構短期的月失業率預測模式，並使用最後的 22 筆資料根據所建立的預測模式進行預測。



圖三 台北市月失業率資料

一、ARIMA模型

在本研究中，我們使用統計套裝軟體 - SAS (Statistical Analysis System)，來協助建立 ARIMA 預測模式。將台北市月失業率共 166 筆的資料，分割成前後兩組：擷取前面的 144 筆資料，當作 SAS 建立 ARIMA 預測模式的訓練資料，並往後預測 22 筆資料來測試模型的預測能力。

由 SAS 的自我相關函數 (Autocorrelation Function: ACF) 曲線圖中 (如表三)，發現原始時間序列資料有均值不穩定的狀態。

嘗試將原始資料經過一階差分後，再觀察差分後的 ACF (如表四) 與部份自我相關函數 (Partial Autocorrelation Function: PACF) (如表五)，得知經過差分轉換後的資料呈現穩定的狀態。除此之外，藉由觀察 ACF 與 PACF 曲線圖，我們選擇 AR 模式 $p=12$ 、 $d=1$ 與 MA 模式 $q=2,3$ 為 ARIMA 之暫定模式，並將此模式與其他暫定模式比較，發現此暫定模式有較小的 AIC 值。且由表六中觀之，其預測的殘差可視為 White Noise，故我們選定 AR 模式 $p=12$ 、 $d=1$ 與 MA 模式 $q=2,3$ 為其 ARIMA 模型。(由 PACF 圖中發現，差分後之資料仍具有季節性效應，相關之 SARIMA 模式曾被列舉探討，但因其 AIC 值皆較暫定模式之 AIC 值為大，故本文中仍選擇暫定模式為最佳模式。)

表三 原始資料的 ACF 曲線圖

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0	0.671302	1.00000												*****									
1	0.572130	0.85227										.		*****									
2	0.474387	0.70667									.	.		*****									
3	0.412723	0.61481									.	.		*****									
4	0.391358	0.58298									.	.		*****									
5	0.372890	0.55547									.	.		*****									
6	0.360787	0.53744									.	.		*****									
7	0.355152	0.52905									.	.		*****									
8	0.322042	0.47973									.	.		*****									
9	0.309710	0.46136									.	.		*****									
10	0.336179	0.50079									.	.		*****									
11	0.378646	0.56405									.	.		*****									
12	0.388432	0.57862									.	.		*****									
13	0.349962	0.52132									.	.		*****									
14	0.311455	0.46396									.	.		*****									
15	0.274213	0.40848									.	.		*****									
16	0.252463	0.37608									.	.		*****									

"." marks two standard errors

表四 一階差分後的 ACF 曲線圖

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0	0.189712	1.00000												*****									
1	-0.0010627	-0.00560										.		.									
2	-0.035465	-0.18694									****		.										
3	-0.038073	-0.20069									****		.										
4	0.0067091	0.03536									.		*	.									
5	-0.0067148	-0.03539									.		*	.									
6	-0.0039299	-0.02072									.		.										
7	0.025781	0.13590									.		***	.									
8	-0.018438	-0.09719									.		**	.									
9	-0.040275	-0.21230									****		.										
10	-0.014657	-0.07726									.		**	.									
11	0.027190	0.14332									.		***	.									
12	0.053846	0.28383									.		*****										
13	0.0087046	0.04588									.		*	.									
14	-0.010843	-0.05716									.		*	.									
15	-0.019524	-0.10291									.		**	.									
16	-0.023654	-0.12469									.		**	.									
17	-0.0057850	-0.03049									.		*	.									

"." marks two standard errors

表五 一階差分後的 PACF 曲線圖

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	-0.00560																						
2	-0.18698									****													
3	-0.21034									****													
4	-0.01112																						
5	-0.12079									**													
6	-0.07220									*													
7	0.11691											**											
8	-0.15102									***													
9	-0.21457									****													
10	-0.10715									**													
11	-0.01811																						
12	0.21457											****											
13	0.08692											**											
14	0.02186																						
15	0.01278																						
16	-0.11530									**													
17	-0.06894									*													
18	0.00853																						
19	0.05614											*											
20	-0.04445									*													
21	-0.01202																						
22	0.03624											*											
23	0.03794											*											
24	0.05590											*											

表六 ARIMA 模式之檢定

Conditional Least Squares Estimation									
Approx.									
Parameter	Estimate	Std Error	T Ratio	Lag					
MA1,1	0.21286	0.07959	<u>2.67</u>	2					
MA1,2	0.26291	0.07995	<u>3.29</u>	3					
AR1,1	0.24084	0.08445	<u>2.85</u>	12					
Variance Estimate = 0.16464047									
Std Error Estimate = 0.40575913									
<u>AIC</u> = 150.81377*									
<u>SBC</u> = 159.702303*									
Number of Residuals = 143									
* Does not include log determinant.									
Autocorrelation Check of Residuals									
To	Chi	Autocorrelations							
Lag	Square	DF	Prob						
6	3.95	3	0.267	-0.093	0.000	0.045	0.088	-0.062	-0.065
12	11.22	9	0.261	0.110	-0.096	-0.131	-0.033	0.086	-0.008
18	13.08	15	0.596	0.019	-0.008	-0.028	-0.065	-0.064	0.043
24	16.66	21	0.732	0.096	-0.071	-0.049	0.052	0.042	0.010
Model for variable X									
No mean term in this model.									
Period(s) of Differencing = 1.									
Autoregressive Factors									
Factor 1: 1 - 0.24084 B**(12)									
Moving Average Factors									
Factor 1: 1 - 0.21286 B**(2) - 0.26291 B**(3)									

此預測模式可表示為：

$$(1 - 0.24084B^{12})(Y(t) - Y(t - 1)) = (1 - 0.21286B^2 - 0.2691B^3)a(t)$$

$Y(t)$ ：為原始的月失業率資料。

$a(t)$ ：為干擾項， $a(t)$ 符合白色噪音 (White Noise) 的假設，即期望值為零、變異數為一固定常數的假設。

根據此一預測模式，我們針對民國 84 年 1 月後的 22 個台北市月失業率值進行預測。在本篇文章中，我們為了與模糊時間序列模式的預測基準一致，故預測時均以實際之失業率歷史資料代入 ARIMA 預測模式計算，其結果可整理如下：

表七 ARIMA 之預測值

時間(月-年)	實際值	ARIMA	時間(月-年)	實際值	ARIMA
Jan-84	1.70	1.64	Dec-84	2.32	1.83
Feb-84	1.84	1.73	Jan-85	2.59	2.42
Mar-84	1.86	1.82	Feb-85	2.26	2.58
Apr-84	1.84	1.81	Mar-85	2.52	2.10
May-84	1.93	1.72	Apr-85	2.69	2.54
Jun-84	2.73	2.06	May-85	2.77	2.70
Jul-84	2.75	2.77	Jun-85	3.13	2.82
Aug-84	2.10	2.61	Jul-85	3.09	3.08
Sep-84	2.23	1.73	Aug-85	3.12	2.85
Oct-84	2.11	2.30	Sep-85	3.13	3.06
Nov-84	1.89	2.14	Oct-85	3.10	3.04

二、模糊時間序列模型

將臺灣地區台北市自民國 72 年 1 月至民國 85 年 10 月間，共 166 筆的月失業率歷史資料按時間順序排列，我們可計算月與月間的失業率變動量，其部份結果如下表。

由所有的變動數值中，我們找出最大變動量 D_{MAX} 和最小的變動量 D_{Min} 分別為 +1.45 與 -1.4，。取 D_1 、 D_2 皆為 0，定義變動量的討論區域 U 為 $U = [-1.4, +1.45]$ 後，我們可以將論域 U 分為

$$\begin{array}{lll}
 u_1 = [-1.40, -0.83) & u_2 = [-0.83, -0.26) & u_3 = [-0.26, +0.31) \\
 u_4 = [+0.31, +0.88) & \text{和} & u_5 = [+0.88, +1.45]
 \end{array}$$

表八 失業率變動量 (部份結果)

時間 (月-年)	失業率(月)	變動量	時間 (月-年)	失業率(月)	變動量
Jan-72	1.35	-	Dec-72	2.99	-0.73
Feb-72	1.49	0.14	Jan-73	3.54	0.55
Mar-72	1.64	0.15	Feb-73	4.80	1.26
Apr-72	1.88	0.24	Mar-73	3.40	-1.40
May-72	2.71	0.83	Apr-73	2.79	-0.61
Jun-72	2.83	0.12	May-73	2.39	-0.40
Jul-72	3.46	0.63	Jun-73	3.00	0.61
Aug-72	3.60	0.14	Jul-73	3.50	0.50
Sep-72	3.72	0.12	Aug-73	3.16	-0.34
Oct-72	3.53	-0.19	Sep-73	3.67	0.51
Nov-72	3.72	0.19	Oct-73	2.97	- 0.70

五個等寬的區間，並計算得各區間中點分別為

$$\begin{array}{lll}
 m_1 = -1.115 & m_2 = -0.545 & m_3 = 0.025 \\
 m_4 = 0.595 & \text{和} & m_5 = 1.165
 \end{array}$$

有關論域的分割與寬度的設定，理論上並無一定的依據。在本研究中，為簡化作業的考量使用五個等寬的區間分割之。針對討論區域 u_1, u_2, u_3, u_4 ，與 u_5 ，，定義模糊集合 A_1, A_2, A_3, A_4 與 A_5 ，分別表示如下：

$$\begin{array}{l}
 A_1 = \frac{1}{u_1} + 0.5 \frac{1}{u_2} + 0 \frac{1}{u_3} + 0 \frac{1}{u_4} + 0 \frac{1}{u_5} \quad A_2 = 0.5 \frac{1}{u_1} + 1 \frac{1}{u_2} + 0.5 \frac{1}{u_3} + 0 \frac{1}{u_4} + 0 \frac{1}{u_5} \\
 A_3 = 0 \frac{1}{u_1} + 0.5 \frac{1}{u_2} + 1 \frac{1}{u_3} + 0.5 \frac{1}{u_4} + 0 \frac{1}{u_5} \quad A_4 = 0 \frac{1}{u_1} + 0 \frac{1}{u_2} + 0.5 \frac{1}{u_3} + 1 \frac{1}{u_4} + 0.5 \frac{1}{u_5} \\
 A_5 = 0 \frac{1}{u_1} + 0 \frac{1}{u_2} + 0 \frac{1}{u_3} + 0.5 \frac{1}{u_4} + 1 \frac{1}{u_5}
 \end{array}$$

根據前面所定義的隸屬函數與模糊集合，所有的月失業率變動量可模糊化 (Fuzzify) 並轉換為變動量的模糊集合，換言之即將失業率變動量對應至討論區域，例如失業率變動量若包含於 u_i ($i=1,2,\dots,5$) 中，則此變動量的模糊變動量即定義為 A_i ($i=1,2,\dots,5$)。根據前面所述之流程，以此類推，即可得到各個模糊變動量，其部份結果如表九。

表九 失業率模糊變動量 (部份結果)

時間 (月-年)	變動量	模糊變動量	時間 (月-年)	變動量	模糊變動量
Jan-72	-	-	Dec-72	-0.73	A_2
Feb-72	0.14	A_3	Jan-73	0.55	A_4
Mar-72	0.15	A_3	Feb-73	1.26	A_5
Apr-72	0.24	A_3	Mar-73	-1.40	A_1
May-72	0.83	A_4	Apr-73	-0.61	A_2
Jun-72	0.12	A_3	May-73	-0.40	A_2
Jul-72	0.63	A_4	Jun-73	0.61	A_4
Aug-72	0.14	A_3	Jul-73	0.50	A_4
Sep-72	0.12	A_3	Aug-73	-0.34	A_2
Oct-72	-0.19	A_3	Sep-73	0.51	A_4
Nov-72	0.19	A_3	Oct-73	-0.70	A_2

在求得模糊變動量後，我們必須選擇適當的模型參數 w ，來計算每期的運算矩陣 $O^w(t)$ 和基準矩陣 $C^w(t)$ 。此處若以 $w=6$ 為例，當針對民國 72 年 8 月的月失業率進行預測時，其計算過程可表示為

運算矩陣

$$\begin{aligned}
 O^6(72/08) &= \begin{bmatrix} \text{fuzzy variation of the unemployment rate of } 72/02 \\ \text{fuzzy variation of the unemployment rate of } 72/03 \\ \text{fuzzy variation of the unemployment rate of } 72/04 \\ \text{fuzzy variation of the unemployment rate of } 72/05 \\ \text{fuzzy variation of the unemployment rate of } 72/06 \end{bmatrix} \\
 &= \begin{bmatrix} A_3 \\ A_3 \\ A_3 \\ A_4 \\ A_3 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 & 0.5 & 0 \end{bmatrix}
 \end{aligned}$$

基準矩陣

$$\begin{aligned}
 C^6(72/08) &= [\text{fuzzy variation of the unemployment rate of } 72/07] \\
 &= [A_4] = [0 \quad 0 \quad 0.5 \quad 1 \quad 0.5]
 \end{aligned}$$

則相關矩陣可被表示為

$$R^6(72/08) = O^6(72/08) \otimes C^6(72/08) = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.25 & 1 & 0.25 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

依定義我們可得到預測的模糊變動量 $F(72/08)$ 為

$$F(72/08) = [0 \quad 0 \quad 0.5 \quad 1 \quad 0.25]$$

最後，我們可以將預測的模糊變動量還原為一般的資料形態。在本文中，我們使用了三個還原模糊集合的準則：

- (1) 若預測的模糊集合當中，僅有一個最大的隸屬度，則取這一隸屬度所在區間的區間中點與隸屬度的乘積作為預測的變動量。
- (2) 若預測的模糊變動量當中，有超過一個以上的最大隸屬度，則取這幾個最大隸屬度所在的區間中點之平均值與隸屬度的乘積作為預測的變動量。
- (3) 若預測的模糊變動量當中，所有的隸屬度皆為 0，則預測變動量取為 0。

根據這些轉換準則，我們可將模糊集合還原為一般的資料形態。以前面的 $F(72/08)$ 為例，應用模糊時間序列模式，可求得失業率值為

$$Y(72/08) = u_F m_4 = 1 * 0.595 = 0.595$$

$$\text{其中 } u_F = \text{MAX}(0, 0, 0.5, 1, 0.25) = 1$$

根據上述的模型建構流程，我們針對參數 w 分別使用不同的數值建構其預測模型，並計算模型的平均誤差。所有結果，整理如下表：

表十 模糊時間序列模式在不同參數 w 下的模式比較

參數	w=3	w=4	w=5	w=6
平均誤差	16.561%	16.573%	17.342%	18.055%

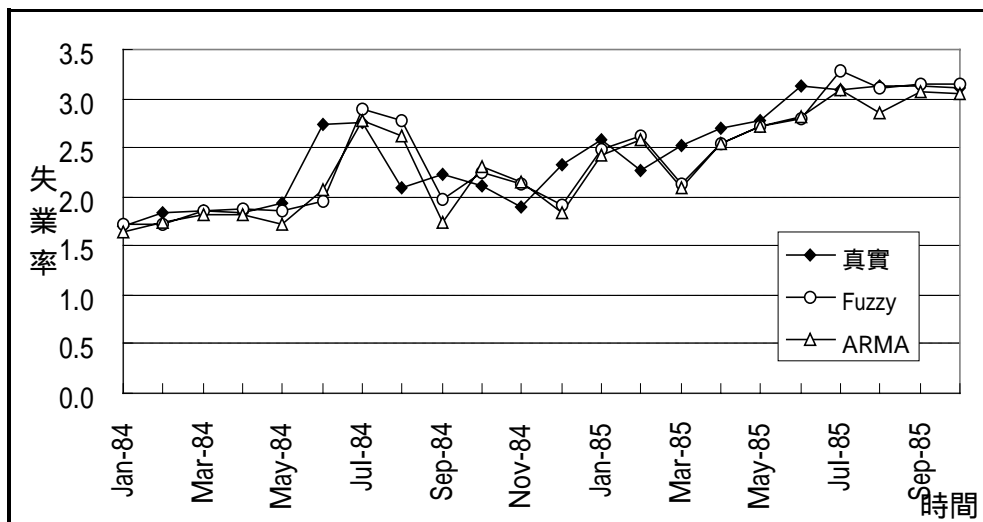
在比較不同模型參數 w 下所產生的預測模式後，我們發現在 $w=3$ 時所產生的平均誤差較小。故選取 $w=3$ 為本論文中，模糊時間序列模式之參數，其最後之預測結果及真實失業率值可整理如表十一。

表十一 模糊時間序列模式之預測值

時間 (月-年)	失業率(月) 實際值	失業率(月) 預測值	時間 (月-年)	失業率(月) 實際值	失業率(月) 預測值
Jan-84	1.70	1.71	Dec-84	2.32	1.91
Feb-84	1.84	1.72	Jan-85	2.59	2.47
Mar-84	1.86	1.86	Feb-85	2.26	2.61
Apr-84	1.84	1.88	Mar-85	2.52	2.13
May-84	1.93	1.86	Apr-85	2.69	2.54
Jun-84	2.73	1.95	May-85	2.77	2.71
Jul-84	2.75	2.88	Jun-85	3.13	2.79
Aug-84	2.10	2.77	Jul-85	3.09	3.28
Sep-84	2.23	1.97	Aug-85	3.12	3.11
Oct-84	2.11	2.25	Sep-85	3.13	3.14
Nov-84	1.89	2.13	Oct-85	3.10	3.15

三、ARIMA及模糊時間序列模式預測結果之分析

藉由前面之描述，我們可以建構出 ARIMA 與模糊時間序列模式的預測模型，並計算得其預測結果。我們將這兩種模式的預測結果，以折線圖表示並與真實值作比較（如圖四）。



圖四 預測結果之分析

由圖四中可知，模糊時間序列模式與 ARIMA 模式在針對台北市月失業率的預測上，均有不錯的預測表現。分別以均方誤差 (Mean Square Error, 簡稱 MSE), 均方根誤差 (Root Mean Square Error, 簡稱 RMSE), 平均絕對誤差 (Mean Absolute Error, 簡稱 MAE) 與平均絕對誤差率 (Mean Absolute Percentage, 簡稱 MAPE) 四種誤差衡量指標 (其中 MAE 與 MAPE 的公式整理如下)，來比較這兩種模型的預測能力，可得到表十二之結果。

$$MAE = \frac{1}{m} \sum_{i=1}^m |\text{預測值}_i - \text{真實值}_i| \quad MAPE = \left(\frac{1}{m} \sum_{i=1}^m \left| \frac{\text{預測值}_i - \text{真實值}_i}{\text{真實值}_i} \right| \right) * 100\%$$

表十二 模式預測能力之衡量

	MSE	RMSE	MAE	MAPE
ARIMA	0.089405	0.299006	0.222401	9.394647
Fuzzy Time Series	0.088585	0.297632	0.206818	8.637884

由表十二中可知，若我們使用模糊時間序列模式與 ARIMA 模式同時對台北市月失業率資料進行預測，在這兩種確定的預測模式下利用真實值逐筆的對月失業率做預測，發現模糊時間序列模式相對於 ARIMA 模式有較好的預測能力。

陸 結論

本研究中，嘗試利用模糊時間序列方法針對台北市的月失業率資料，建構短期的失業率預測模式。我們以臺灣地區台北市自民國 72 年 1 月至民國 85 年 10 月間的月失業率資料為實證的歷史資料，分別以模糊時間序列模型與 ARIMA 模型建構適當的預測模型，並對於不同模式的預測能力分析評估。

我們以 MSE, RMSE, MAE 與 MAPE 四種不同的誤差指標來衡量這兩種模型的預測能力，並且很清楚的顯示出模糊時間序列模式相對於 ARIMA 模型，有較佳的預測能力。這對於決策者而言，不外乎提供了一種新的預測方法，並且又可以獲得較好的預測結果，提供更正確的資訊，使得決策者可以做出更正確的判斷。

參考文獻

- 中華民國台灣地區人力資源統計月報，行政院主計處編印。
- 吳柏林，*時間數列分析導論*，雙葉書廊，1994。
- 林茂文，*時間數列分析與預測*，華泰書局，1992。
- 吳柏林、陳雅玫，「台灣地區失業率的時空數列分析與預測」，*人力資源學報*，Vol. 3，1993。
- 吳柏林、張鈿富、廖敏治，「模糊時間數列與臺灣地區中學教師需求人數之預測」，*國立政治大學學報*，Vol. 73, 1996，pp.287-312。
- Bowerman, O'Connell, *Forecasting and Time Series: An Applied Approach*，third edition, Duxbury, Belmont, California, 1993.
- Box, G. E. P., and Jenkins, G. M., *Time Series Analysis, Forecasting and Control*, Holden Day, Oakland, CA. 1976.
- Funke, Michael, "Time-series Forecasting of the German Unemployment Rate", *Journal of Forecasting*, Vol. 11, 1992, pp.111-125.
- Hwang, J. R., Chen, S. M., and Lee, C. H., "A New Method for Handling Forecasting Problems Based on Fuzzy Time Series", Discussion Paper, Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, R. O. C. 1995.
- SAS Institute Inc., *SAS/ETS User's Guide: Econometrics and Time Series Library*, Vol. 6, second edition, Cary, North Carolina, 1993.
- Sims, Ch., "Macroeconomics and Reality", *Econometrica*, Vol. 48, 1980, pp.1-48.
- Song, Q. and Chissom, B. S., "Fuzzy Time Series and Its Model", *Fuzzy Sets and System*, Vol. 54, No. 3, 1993, pp.269-277.
- Song, Q. and Chissom, B. S., "Forecasting Enrollments with Fuzzy Time Series - Part I", *Fuzzy Sets and System*, Vol. 54, No. 1, 1993, pp.1-9.
- Song, Q. and Chissom, B. S., "Forecasting Enrollments with Fuzzy Time Series - Part II", *Fuzzy Sets and System*, Vol. 62, No. 1, 1994, pp.1-8.
- Zadeh, L. A., "Fuzzy Sets", *Information and Control*, Vol. 8, 1965, pp.338-353.

Prediction of Taipei's Unemployment Rate - Using Fuzzy Time Series Methodology

CHIH-CHOU CHIU*, TIAN-SHYUG LEE* AND FAN-CHYUN LIN**

**Department of Business Administration, Fu-Jen Catholic University.*

***Institute of Applied Statistics, Fu-Jen Catholic University.*

ABSTRACT

This study presents a novel semiparametric prediction system for the Taipei's unemployment rate series. The prediction method incorporated into the system consists of a fuzzy time series model that estimates the trend, as well as a Box-Jenkins prediction of the residual series. In terms of the adaptability of the Box-Jenkins method, the prediction intervals of the system can be successfully constructed. The extensive studies are performed on the robustness of the built fuzzy model using different specified model basis. To demonstrate the effectiveness of our proposed method, the Taipei's monthly unemployment rate from Feb. 1983 to October 1996 was evaluated using a fuzzy time series technique. Analysis results demonstrate that the proposed method outperforms than the traditional time series methodology.

Keywords: fuzzy time series, unemployment rate, time series analysis, model basis