結合模糊時間序列與Box-Jenkins模式在台北市失業率預測上之應用

邱志洲* 李天行* 林凡群**

摘 要

預測技術一直是決策者在決策過程中不可或缺的重要工具，且精確的預測結果可以提供決策者更多的資訊，有利於做出正確的決策與適當的反應。傳統的預測方法可以處理一般問題，但其預測模式常需要較嚴格的基本假設，使得預測模式的建構較為困難；而模糊時間序列模式並不需要強烈的基本假設，使得預測模式的建構更為簡單，也提供決策者更多的選擇。然而模糊時間序列模式僅有點估計值而缺乏區間的估計，使得決策者使用模式時，無法衡量預測值的可信度，大大的降低了預測模式的有效性。

失業率是勞動市場的重要指標之一，它能提供許多有關人力資源管理的資訊和政府決策的重要依據。因此，若能精確的預測失業率，對於決策者將會有很大的幫助。在本文中，嘗試先利用模糊時間序列模式根據台灣地區台北市自民國 72 年 1 月至民國 85 年 10 月間之月失業率進行預測，再針對預測後所得之殘差，以 ARIMA 分析的技術輔助建立信賴區間，希望能提供決策者一種新的預測方法，並使得傳統模糊時間序列預測模式在加上信賴區間的估計後，預測的結果更趨於完整，提供給使用者更多有用的判斷訊息。

關鍵詞：時間序列，預測，失業率，模糊，人力資源

* 管理研究所，輔仁大學，新莊，臺灣
** 應用統計研究所，輔仁大學，新莊，臺灣

1. 緒　論

預測技術一直是決策者在決策過程中不可或缺的重要工具，且精確的預測結果可以提供決策者更多有用的資訊，有利於做出正確的決策與適當的反應。傳統的預測方法可以用來處理一般的問題，但其預測模式常需要較嚴格的基本假設，使得預測模式的建構較為困難。Qiang 與 Chissom (1993) 針對歷史性的時間序列資料提出新的預測方法，定義模糊時間序列模型的基本架構。模糊時間序列模式的建構並不需要嚴格的基本假設，使得預測模式的建構更為簡單，也提供了決策者更多的選擇機會。但美中不足的是僅有點估計值而缺乏區間的估計，使得決策者建立預測模式時，無法衡量預測值的可靠範圍，降低了預測模式的可用性。

在本文中，嘗試將模糊時間序列模式針對台灣地區台北市月失業率資料所作的預測結果，再以 ARIMA 分析的技術輔助信賴區間的建立，希望能改善模糊時間序列模式預測時的缺陷，使預測的結果更趨於完整，提供決策者更有效的決策資訊。

本論文共分為六部分：第一部分為緒論，說明本文研究的動機與目的及論文的整體架構；第二部分的重點則是探討及回顧模糊時間序列模式與 ARIMA 分析技術的相關文獻；在第三部分中，則是介紹模糊時間序列模式的建構程序；於第四部分中，是說明 ARIMA 分析技術的建構流程；於第五部分的實證中，則說明了改良的結合模式提供了新預測值的預測信賴區間；最後，在第六部分提出本研究的結論。

2. 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 Volumn 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 (1965) 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) 則針對 Qiang 與 Chissom (1993) 所提的模糊時間序列模式，修改資料的處理方式和相關矩陣 (Relation Matrix) 的計算準則，經其證明該模式有較佳的預測能力。

Jurado 等人 (1995) 提出一種改善無母數預測模式的方法；使用無母數方法進行預測，並利用 ARIMA 模式針對預測所得之殘差值加以預測，而得到殘差之預

測值。再將無母數方法之預測值與 ARIMA 模式所得到的殘差預測值合併後，便可得到新的預測值。此外，利用 ARIMA 分析所產生的信賴區間，合併後新預測值的信賴區間便可成功的被建立起來。

3. 模糊時間序列模式

3.1 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【表一】)

表一、普通集合與模糊集合的數學定義

◎ 普通集合：

設 X 為論域，A 是 X 的任意子集，x 是 X 中之任意元素

$$C(x)=\begin{cases}1 & (x\in A)\\ 0 & (x\notin A)\end{cases}$$

◎ 模糊集合：

設 U 為論域，A 是 U 的任意模糊子集，u 是 U 中之任意元素

$$A=\left\{\frac{\mu_A(u)}{u}\,\middle|\, u\in U\right\}$$

※ μ 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)，$0\le\mu_A(u_i)\le 1$ 。

3.2 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。

3.2.1 定義符號

$Y(t)$ ：是一個模糊時間序列資料(t=0,1, 2, ...)。

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, ..., n)，可將其表為 $U=\{u_1, u_2, ..., u_n\}$; $u_i(t)$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

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

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

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

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).

3.2.2 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：

一、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。

二、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。

三、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。

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

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

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

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：

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

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

若我們想要預測第 t 期的的數據資料，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 $C^W(t)=F(t-1)=[u_{t-1,1}, u_{t-1,2},\ldots, u_{t-1,n}]$，運算矩陣 $O^W(t)$則為 $O^W(t)=[F(t-W), F(t-w+1), \ldots, F(t-2)]$。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

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

$$=\begin{bmatrix} u_{t-w,1} & u_{t-w,2} & \cdots & u_{t-w,n} \\ u_{t-w+1,1} & u_{t-w+1,2} & \cdots & u_{t-w+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{t-2,1} & u_{t-2,2} & \cdots & u_{t-2,n} \end{bmatrix}\otimes\begin{bmatrix} u_{t-1,1} & u_{t-1,2} & \cdots & u_{t-1,n} \end{bmatrix}$$

$$=\begin{bmatrix} u_{t-w,1}u_{t-1,1} & u_{t-w,2}u_{t-1,2} & \cdots & u_{t-w,n}u_{t-1,n} \\ u_{t-w+1,1}u_{t-1,1} & u_{t-w+1,2}u_{t-1,2} & \cdots & u_{t-w+1,n}u_{t-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{t-2,1}u_{t-1,1} & u_{t-2,2}u_{t-1,2} & \cdots & u_{t-2,n}u_{t-1,n} \end{bmatrix}=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{w-1,1} & R_{w-1,2} & \cdots & R_{w-1,n} \end{bmatrix}$$

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：$F(t)=[MAX(R_{11}, R_{21},..., R_{W-1,1}), MAX(R_{12}, R_{22},...,R_{W-1,2}),..., MAX(R_{1n}, R_{2n},..., R_{W-1,n})]$。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。

4. ARIMA模式分析技術的建構流程

4.1 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 $B^dY(t)=Y(t-d)$。$\Phi_p(B)$及$\theta_q(B)$則分別為 B 之多項式，可表示如下：

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

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

4.2 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)，其步驟說明如下：

步驟一、觀察資料結構：

將收集好的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)，並從 ACF 與 PACF 圖中觀察資料的特性，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：

將模式內的未知參數加以估計，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。

4.3 模糊時間序列模式信賴區間之建構

根據 Jurado 等人(1995)所提的理論為研究的基本架構，其先利用無母數的預測方法建立預測模式，得到預測值，並求出實際值與無母數預測值之殘差。再將殘差值利用 ARIMA 模式建立另一預測模式，得到殘差之預測值與預測信賴區間。最後將無母數方法之資料預測值與 ARIMA 模式所估計的殘差預測值結合後，便可得到新的預測值，此外我們可利用 ARIMA 模式中之信賴區間建構合併模式之預測區間。在本研究中，即是應用前述之概念，將模糊時間序列模式與 ARIMA 分析的技術結合，發展一種結合的預測模式，使此預測模式具備預測的信賴區間。

4.3.1 模糊時間序列模式與 ARIMA 分析方法之結合

利用 Jurado 等人 (1995) 所提的理論為研究的基本架構，使用模糊時間序列模式建立預測模式，得到預測結果，將預測模式簡單定義為

$$Y(t) = Fuzzy\,(Y\,(t-w+1)\,) + e\,(t)$$

其中，

$Y(t)$ ：為時間序列資料的第 t 期實際值。

$Fuzzy(Y(t-w+1))$ ：為模糊時間序列模式的第 t 期預測值。

$e(t)$ ：為模糊時間序列模式的第 t 期殘差值。

w ：為模糊時間序列模式的基準期數參數。

我們可再針對模糊時間序列模式的殘差值 $e(t)$，建立 ARIMA 的預測模式，得到殘差的 ARIMA 預測值 $\hat{e}(t)$，而定義合併模式的新預測值為

$$\hat{Y}(t) = Fuzzy\,(Y(t-w+1)) + \hat{e}(t)$$

其中，

$\hat{Y}(t)$ ：合併模式的新預測值。

$\hat{e}(t)$ ：為利用 ARIMA 方法，針對模糊時間序列模式預測之殘差值建立預測模式，所得的殘差預測值。

4.3.2 結合模式的 Box-Jenkins 預測區間

利用模糊時間序列模式建立預測模式後，針對預測所得之殘差值建立 ARIMA 模式，再藉由 Box-Jenkins 所提的方法中，我們可將由 ARIMA 方法所建構 $Y(t)$在信賴水準$(1-\alpha)$下的近似信賴區間，表示如下：

$$\hat{Y}(t) \pm Z_{\alpha/2}(\hat{\sigma}^{\,2}\sum_{j=0}^{k-1}\hat{\pi}_j^2)^{1/2}$$

其中，

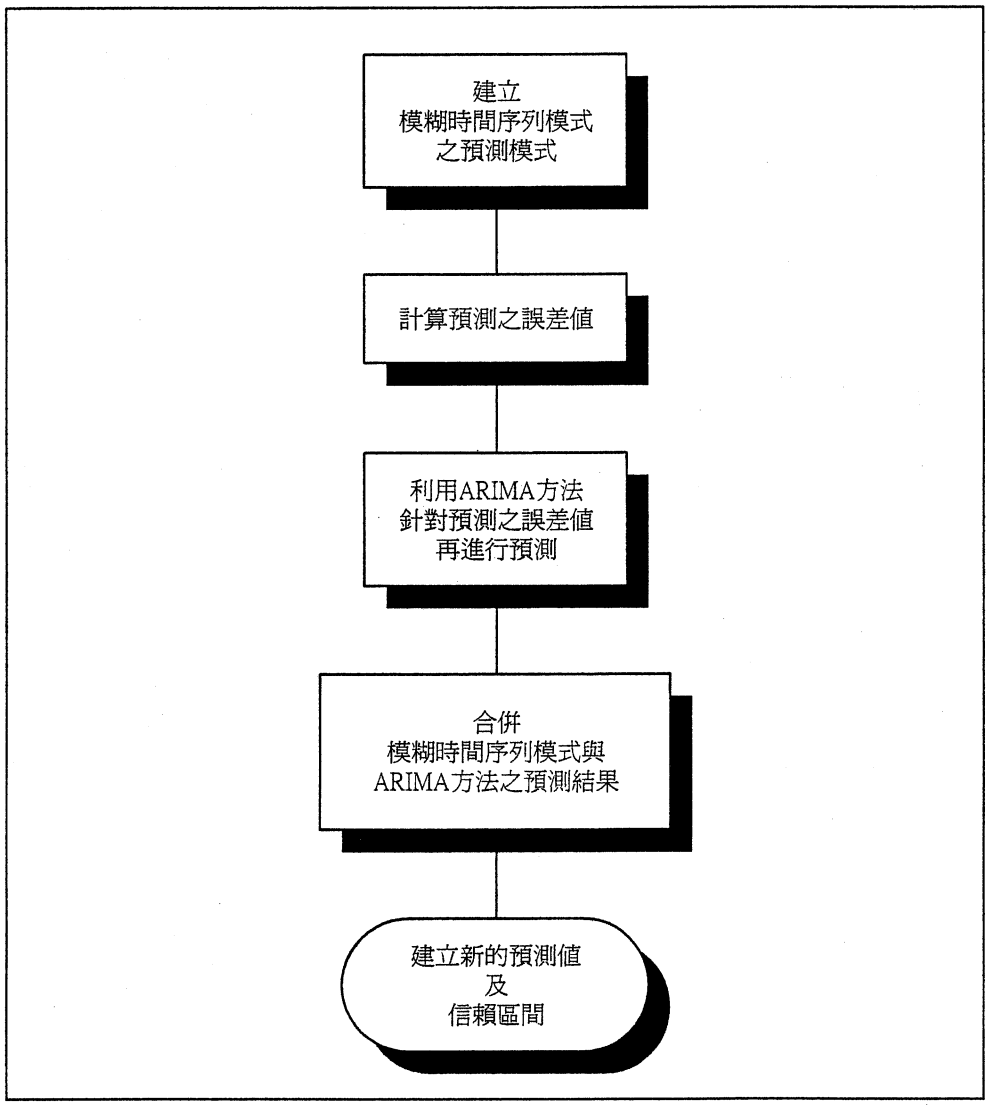
$Z_{\alpha/2}$ ：為標準常態下右尾機率為 $\alpha/2$ 之值。

$\hat{\sigma}^2$ ：為以 ARIMA 方法對模糊時間序列模式的殘差 $e(t)$，所求之估計變異數。

$\hat{\pi}_j$ ：為下列方程式所得之方程式π_j的係數估計值；其中，$(1-B)^d$代表d階差分之運算。且θ與φ為ARMA模式中自我迴歸及移動平均之係數。

$$\pi(B)=\frac{\theta(B)}{\varphi(B)(1-B)^d}$$

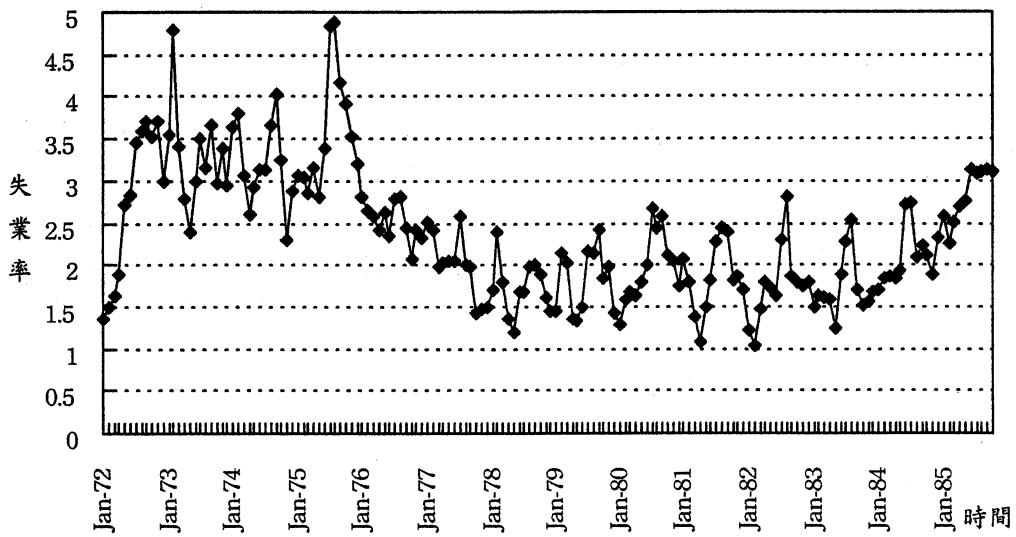
整個結合模式的建構流程可由下圖表之：



圖一、合併模式的建構流程

5. 實證研究

本文所使用的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(如圖二)；針對月失業率資料，利用模糊時間序列模式建立預測模式，得到所有預測值。再計算模糊時間序列模式預測結果與真實值間訓練資料殘差值，將這些殘差值資料分成兩部份，利用前面共 144 筆的資料當作 ARIMA 模式的訓練資料建立預測模式，並針對最後的 22 筆資料進行預測，得到殘差預測值與預測的信賴區間，再利用 Jurado 等人(1995)所提的理論，將兩個模式之預測值合併為一新的月失業率預測值，並計算其信賴區間。



圖二、台北市月失業率資料

5.1 建構台北市失業率的模糊時間序列預測模式

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【表二】。

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

u_1=[-1.40, -0.83)　　u_2=[-0.83, -0.26)　　u_3=[-0.26, +0.31)

u_4=[+0.31, +0.88)　　和　　u_5=[+0.88, +1.45]

表二、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

m_1= -1.115　　m_2= -0.545　　m_3= 0.025

m_4= 0.595　　和　　m_5= 1.165

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：

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 \qquad A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5$$

$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 \qquad A_4 = 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5$$

$$A_5 = 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5$$

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 $u_i\,(i=1,2,\ldots,5)$中，則此變動量的模糊變動量即定義為 $A_i\,(i=1,2,\ldots,5)$。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【表三】。

表三、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，應用第三節之計算程序，針對民國 72 年 8 月的月失業率進行預測可得以下之結果：

運算矩陣

$$O^6(^{72}/_{08})=\begin{bmatrix}A_3\\A_3\\A_3\\A_4\\A_3\end{bmatrix}=\begin{bmatrix}0&0.5&1&0.5&0\\0&0.5&1&0.5&0\\0&0.5&1&0.5&0\\0&0&0.5&1&0.5\\0&0.5&1&0.5&0\end{bmatrix}$$

基準矩陣

$$C^6(^{72}/_{08}) = [A_4] = [0 \quad 0 \quad 0.5 \quad 1 \quad 0.5]$$

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：

一、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。

二、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。

三、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

$$Y(^{72}/_{08}) = u_F m_4 = 1 \times 0.595 = 0.595$$

其中 $\quad u_F = MAX(0, 0, 0.5, 1, 0.25) = 1$

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（表四）：

表四、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%

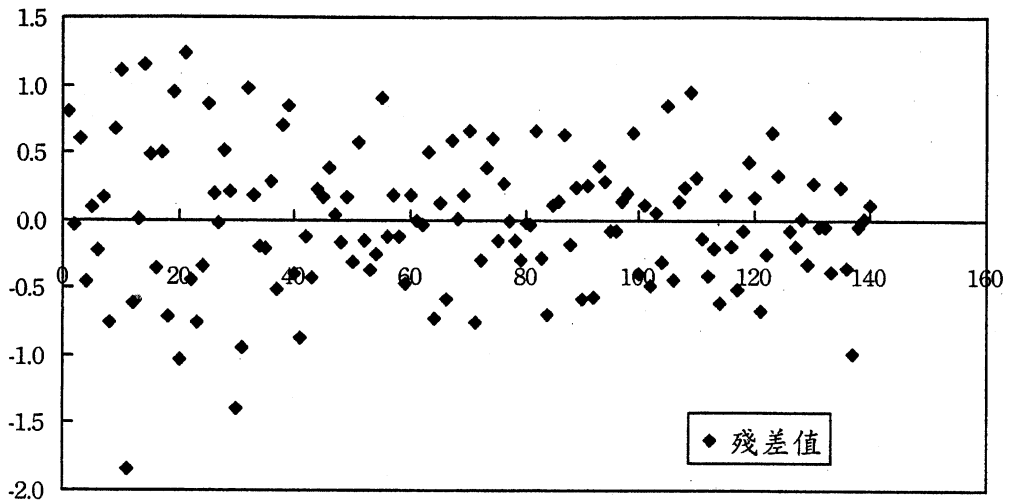
在比較不同模型參數 w 下所產生的預測模式後，我們發現在 w=3 時所產生的平均誤差較小。故選取 w=3 為本論文中，模糊時間序列模式之參數，確立模糊時間序列的預測模型。將 w=3 的模糊時間序列模式代入，計算最後 22 筆資料的失業率預測值與殘差值，得到如【表五】之結果：

表五、模糊時間序列模式預測結果(部份結果)

時間	真實值	預測值	殘差值
May–72	2.71	1.905	0.805
Jun–72	2.83	2.865	–0.035
Jul–72	3.46	2.855	0.605
Aug–72	3.60	4.055	–0.455
Sep–72	3.72	3.625	0.095
Oct–72	3.53	3.745	–0.215
Nov–72	3.72	3.555	0.165
Dec–72	2.99	3.745	–0.755
Jan–73	3.54	2.860	0.680
Feb–73	4.80	3.695	1.105
Mar–73	3.40	5.240	–1.840
Apr–73	2.79	3.400	–0.610
May–73	2.39	2.375	0.015

5.2 應用ARIMA分析方法對殘差值之預測

利用模糊時間序列預測模式中，所得到的月失業率預測值，計算從民國 72 年 5 月至民國 83 年 12 月間之預測殘差值，其值分布如下：



圖三、模糊時間序列模式殘差散佈圖

將模糊時間序列模式中，民國 72 年 5 月至民國 83 年 12 月間共 140 筆資料預測殘差值當作建立 ARIMA 模式的資料，並往後預測 22 筆殘差值及建立殘差的信賴區間。我們使用統計套裝軟體－SAS (Statistical Analysis System) 程式，來協助建立 ARIMA 預測模式。

觀察 SAS 報表中的 ACF 與 PACF 曲線圖，判斷模型的 p、d、q 值。我們選擇 p=2,3,5 及 q=1,12 為 ARIMA 模式之暫定模式，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。再由【表六】中觀知，此模式的 T-ratio 絕對值均大於 2，且由殘差自我相關檢驗中，得知此模式的殘差值可視為 White Noise，故我們選定這些模糊時間序列預測模式的殘差值之預測模式為 AR 模式 p=2,3,5，d=0 與 MA 模式 q=1,12 的 ARMA 模型。結合以上的分析結果，此預測模式可表示為：

$$(1+0.25757B^2+0.25546B^3+0.18298B^5)\,Y(t)=$$

$$(1-0.28209B+0.2292B^{12})a(t)$$

其中，

表六、殘差之ARIMA模式的檢定

Conditional Least Squares Estimation

Parameter	Estimate	Std Error	Approx. T Ratio	Lag
MA1,1	0.28209	0.08304	3.40	1
MA1,2	-0.22920	0.08304	-2.76	12
AR1,1	-0.25757	0.08435	-3.05	2
AR1,2	-0.25546	0.08321	-3.07	3
AR1,3	-0.18298	0.08681	-2.11	5

Variance Estimate = 0.22705658

Std Error Estimate = 0.47650455

AIC = 194.653476*

SBC = 209.361688*

Number of Residuals= 140

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations						
6	2.58	1	0.108	-0.022	0.045	-0.038	0.090	-0.014	-0.074
12	10.58	7	0.158	0.094	-0.118	-0.046	-0.024	0.164	0.013
18	15.03	13	0.306	-0.025	-0.004	-0.068	-0.098	-0.103	0.047
24	18.70	19	0.476	0.103	-0.037	-0.055	0.041	0.070	0.023

Model for variable X

No mean term in this model.

Autoregressive Factors

Factor 1: 1 + 0.25757 B**(2) + 0.25546 B**(3) + 0.18298 B**(5)

Moving Average Factors

Factor 1: 1 - 0.28209 B**(1) + 0.2292 B**(12)

$Y(t)$ ：為第 t 期的模糊時間序列預測模式的殘差值。

$a(t)$ ：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，以真實值逐筆代入往後預測 22 筆資料，可以得到預測結果如【表七】。

表七、殘差之ARMA模式預測結果

時間(月-年)	Fuzzy 預測值	Fuzzy 預測後之殘差值	ARIMA 對殘差之預測值
Jan-84	1.715	- 0.015	0.0418
Feb-84	1.725	0.115	0.1537
Mar-84	1.865	- 0.005	- 0.0391
Apr-84	1.885	- 0.045	- 0.0690
May-84	1.865	0.065	- 0.0922
Jun-84	1.955	0.775	0.1465
Jul-84	2.885	- 0.135	- 0.1274
Aug-84	2.775	- 0.675	- 0.2637
Sep-84	1.970	0.260	- 0.1548
Oct-84	2.255	- 0.145	0.0224
Nov-84	2.135	- 0.245	- 0.0176
Dec-84	1.915	0.405	0.0438
Jan-85	2.475	0.115	0.1088
Feb-85	2.615	- 0.355	- 0.0999
Mar-85	2.130	0.390	- 0.0268
Apr-85	2.545	0.145	- 0.0052
May-85	2.715	0.055	- 0.0902
Jun-85	2.795	0.335	- 0.0549
Jul-85	3.285	- 0.195	- 0.0980
Aug-85	3.115	0.005	- 0.2386
Sep-85	3.145	- 0.015	- 0.0355
Oct-85	3.155	- 0.055	- 0.0057

5.3 合併模式

利用 Jurado 等人(1995)所提的理論，將已建好的模糊時間序列預測模式的失業率預測值與 ARIMA 方法所得的殘差預測值合併而得到新的失業率預測值與信賴區間，結果如【表八】。

表八、合併模式之預測值與信賴區間

時間 (月-年)	失業率 真實值	Fuzzy 預測值	Fuzzy + ARIMA	Lower 90%	Upper 90%
Jan-84	1.70	1.715	1.7568	0.9730	2.5407
Feb-84	1.84	1.725	1.8787	1.0482	2.6771
Mar-84	1.86	1.865	1.8259	0.9612	2.6394
Apr-84	1.84	1.885	1.8160	0.9541	2.6565
May-84	1.93	1.865	1.7728	0.9310	2.6473
Jun-84	2.73	1.955	2.1015	1.2998	3.0195
Jul-84	2.75	2.885	2.7576	2.1146	3.8371
Aug-84	2.10	2.775	2.5113	1.8413	3.5645
Sep-84	2.23	1.970	1.8152	0.9310	2.6551
Oct-84	2.11	2.255	2.2774	1.3443	3.0701
Nov-84	1.89	2.135	2.1174	1.2701	2.9959
Dec-84	2.32	1.915	1.9588	1.0770	2.8029
Jan-85	2.59	2.475	2.5838	1.6182	3.3836
Feb-85	2.26	2.615	2.5151	1.7587	3.5242
Mar-85	2.52	2.130	2.1032	1.2421	3.0093
Apr-85	2.69	2.545	2.5398	1.6471	3.4168
May-85	2.77	2.715	2.6248	1.8199	3.5897
Jun-85	3.13	2.795	2.7401	1.9097	3.6797
Jul-85	3.09	3.285	3.1870	2.4011	4.1712
Aug-85	3.12	3.115	2.8764	2.2334	4.0035
Sep-85	3.13	3.145	3.1095	2.2621	4.0323
Oct-85	3.10	3.155	3.1493	2.2705	4.0408

再分別以平均絕對誤差(Mean Absolute Error，簡稱 MAE)與平均絕對誤差率（Mean Absolute Percentage，簡稱 MAPE）(公式整理如下)二種誤差衡量指標，來比較模糊時間預測模式與新模式的預測能力，可得到【表九】之結果。

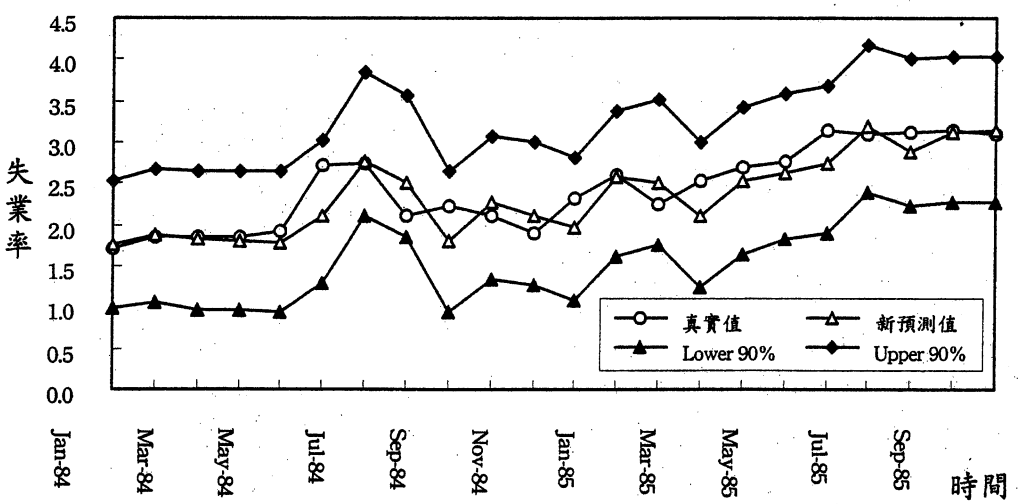
$$MAE = \frac{1}{m}\sum_{i=1}^{m} \left| \text{預測值}_i - \text{真實值}_i \right| \qquad MAPE = \left(\frac{1}{m}\sum_{i=1}^{m} \left| \frac{\text{預測值}_i - \text{真實值}_i}{\text{真實值}_i} \right| \right) * 100\%$$

表九、模式預測能力之衡量

	MAE	MAPE
Fuzzy Time Series	0.206818	0.08637884
合併模式	0.195578	0.08103827

由【表九】中發現，我們針對台北市月失業率資料使用模糊時間序列模式與合併模式進行預測，合併模式相對於原來的模糊時間序列模式有較佳的預測能力。

在合併模式下，針對台北市月失業率資料所建立在$(1-\alpha)=90\%$下的信賴區間預測圖如下圖：



圖四、合併模式在$(1-\alpha)=90\%$下的信賴區間

由【圖四】中可知，針對台北市失業率資料所建的信賴區間而言，在信賴水準$(1-\alpha)$為 90%下的信賴區間，包含了所有的實際月失業率資料值。由以上的結果可知，本研究應用模式合併的方法，改良了模糊時間序列模式的預測值，並且成功的為模糊時間序列模式建立了預測的信賴區間。

6. 結論

在本研究中，利用 Jurado 等人(1995)所提的理論，將已建好的模糊時間序列預測模式的失業率預測值與 Box-Jenkins 方法所得的殘差預測值合併而得到新的失業率預測值與信賴區間，改善模糊時間序列預測模式只有點估計的缺陷，使決策者可以從信賴區間中獲得預測的有效訊息。

由實證研究中可知合併的模式中，改良了原來模式的預測能力，且針對模式的新預測值建立了預測的信賴區間，使得模糊時間序列模式在新模式中有更加完整的表現。對於決策者而言，提供了一種新的預測方法，使得預測結果有較完整的預測資訊，讓決策者可以做出更正確而有效的判斷。

參考文獻

1. "中華民國台灣地區人力資源統計月報"，行政院主計處編印。

2. 吳柏林 (1994)，"時間數列分析導論"，雙葉書廊。

3. 林茂文 (1992)，"時間數列分析與預測"，華泰書局。

4. 吳柏林、陳雅玫 (1993)，"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。

5. Bowerman, O' Connell (1993), "Forecasting and Time Series: An Applied Approach", third edition, Duxbury, Belmont, California.

6. Box, G. E. P., and Jenkins, G. M. (1976), "Time Series Analysis , Forecasting and Control", Holden Day, Oakland, CA.

7. Funke, Michael (1992), "Time-series Forecasting of the German Unemployment Rate" , Journal of Forecasting, Vol. 11, pp.111-125.

8. Garcia-Jurado, I., Gonzalez-Manteiga, W., Prada-Sanchez, J. M., Febrero-Bande, M., and Cao, R. (1995), "Predicting Using Box-Jenkins, Nonparametric, and Bootstrap Techniques", Technometrics, Vol. 37, No. 3, pp.303-310.

9. Hwang, J. R., Chen, S. M., and Lee, C. H. (1995), "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.

10. SAS Institute Inc. (1993), "SAS/ETS User's Guide : Econometrics and Time Series Library" , Vol. 6, second edition, Cary, North Carolina.

11. Sims, Ch. (1980), "Macroeconomics and Reality", Econometrica, Vol. 48, pp.1-48.

12. Song, Q. and Chissom, B. S.(1993), "Fuzzy Time Series and Its Model", Fuzzy Sets and System, Vol. 54, No. 3, pp.269-277.

13. Song, Q. and Chissom, B. S. (1993), "Forecasting Enrollments with Fuzzy Time Series - Part I" , Fuzzy Sets and System, Vol. 54, No. 1, pp.1-9.

14. Song, Q. and Chissom, B. S. (1994), "Forecasting Enrollments with Fuzzy Time Series - Part II", Fuzzy Sets and System, Vol. 62, No. 1, pp.1-8.

15. Zadeh, L. A. (1965), "Fuzzy Sets", Information and Control, Vol. 8, pp.338-353.

Combining Fuzzy Time Series Approach with Box-Jenkins Methodology for Unemployment Rate Prediction in Taipei

Chih-Chou Chiu*, Tian-Shyug Lee*, Fan-Chyun Lin**

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 model with the Box-Jenkins time series technique. Analysis results demonstrate that the proposed method outperforms than the traditional fuzzy time series method.

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

* Graduate School of Business Administration, Fu-Jen Catholic University, Taiwan
** Graduate School of Applied Statistics, Fu-Jen Catholic University, Taiwan