

模糊統計分類在台灣地區失業率分析與預測之應用

吳柏林

許毓云

國立政治大學應用數學研究所

摘 要

失業率是各國經濟發展所應重視的指標之一。過高的失業率會影響整個就業市場及社會秩序。近年來，由於台灣工資高漲，加上大量引進外勞而導致失業率升高，已造成社會的嚴重問題。此一問題若未獲得改善，不僅形成人力資源的閒置，且將造成嚴重的社會失調。

本文即以失業率代表勞動市場供需變數之指標，嘗試以模糊時間數列分類與辨識的方法，應用平均累加模糊熵 (*average of the sum of fuzzy entropies*)，來認定台灣地區失業率走勢與結構改變。由實證結果發現，應用模糊統計分析方法比傳統中的測度方法能得到更合理的解釋，且預測結果可以提供決策者更多的資訊，做出正確的決策。

關鍵詞：失業率、模糊時間數列、模糊分類法。

1. 前言

失業率是各國經濟發展所應重視的指標之一。過高的失業率會影響整個就業市場及社會秩序。更進一步地，失業對整個經濟體系而言，是沒有善用資源的表徵；對個人而言，則是物質與精神生活的無以寄託，因而將引發人際與社會關係的危機。因此，失業問題便一直是決策當局所重視的主要課題之一，而失業現象自然也成為經濟學者們從事理論研究的重要標的。

根據定義，失業率係指勞動力中失業人口所佔之比率。主計處乃根據資料標準週(行政院主計處電腦網站, 1998)之失業人數與勞動力而得到失業率資料。目前勞動力調查對失業者的工作意願是定義為“想工作且正在找工作”。然而這種狹窄的失業定義，往往低估了失業人口以及失業率的計算。尤其是在不景氣的時段內，許多失業的人因為工作機會的減少而失去在勞動市場內積極尋找工作信心與興趣，而成為所謂的“失志勞工(discouraged workers)”。除此之外，由於資料收集的誤差、時間的遲延(time lag)，使得單一度量的數值，形式上看似精確，但它實際上所表達的應該是某一範圍區間的可能值。例如，民國85年台灣地區的失業率為2.60%，但這數據只是一統計估計值，它並不能完全準確地反應民國85年台灣地區之失業情況。是故，失業率這個數字的本身就具有不確定性與模糊性。在此情況下，失業率數值的模糊特性，在經濟的實証研究過程中，應予以慎重考慮。

模糊理論(fuzzy theory)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。因此，模糊理論目前已被廣泛地應用於各種領域。例如航空、機械、醫學、電力、地質...等等，近來也逐漸應用在社會科學方面。而Clymer、Corey與Gardner(1992)提出離散事件的機場起落模糊控制方法，Cutsem與Gath(1993)曾提出利用模糊分類的程序來偵測離羣值及穩健地估計參數，Hathaway與Bezdek(1993)則對模糊迴歸模式進行參數估計與分類。Yoshinari，Pedrycz與Hirota(1993)經由模糊分類方法建立模糊模式，Romer與Kandel(1995)在統計推論上採用模糊分割的理論及可能性理論。此外，Ramakrishnan和Rao(1992)及Park和Kim(1996)在其所發表的論文中均提出模糊權重相加法則及其註解，但對於權重的如何決定仍未說明，或者仍然延用過去主觀經驗的定義方式。

近年來，在人文社會科學的測度理論裡，模糊統計與模糊相關性日漸受到重視，這應是複雜的人文社會現象無法以傳統數值模型充分合理解釋的一種自然發展結果。有鑑於此，本文乃嘗試以模糊時間數列分析，來認定台灣地區失業率走勢與結構改變。這對於了解台灣的經濟波動將深具意義。

本文主要分為四節。第一節探討模糊理論與相關文獻，以獲取發展理論架構的依據，並說明分析失業率之目的。第二節說明模糊時間數列分析與理論方法，以為實務應用的基礎。第三節為實證分析。最後，本文結論給在第四節。

2. 模糊統計分析

2.1 模糊時間數列的型態識別

傳統上，時間數列的型態識別與分類(pattern recognition and classification)分析方法，大部份應用在定態的(stationary)隨機過程。但近年來非線性時間數列之模型分類與建構的研究越來越受到重視，而根據時間數列的走勢型態可以作為判斷事件發生的基礎，並進而對未來的走勢作有效的預測。但傳統的辨識方法常因前提假設的限制(如資料是否定態，自我相關程度等)，而失去辨識準確度的穩健性。再者，由於一些辨認模型結構性改變的方法，常受制於時間數列本身的非定態及干擾過程的不確定影響，因此在以時間數列方法辨認模型結構性改變的問題時，首先必須考慮轉折(或結構變動)的認定。

轉折可以解釋為，在一個定態的時間數列模型中，一個或多個參數發生改變。像Chow (1960)，Nyblom (1989)，Bleaney (1990)及Ploberger與Kramer (1992)等很多學者，均曾提出不同之轉折發生的檢定方法。Lin與Terasvirta (1994)則提出應用改良的累加平方和(cumulative sum)方法與推廣的拉格蘭治乘數檢定方法(Lagrangian multiplier test)，來解決參數一致性檢定的問題。雖然有各種方法被提出用以檢定轉折的發生，但對於轉折點的定義似乎仍無一明確的標準。因為一具有明顯突發性改變的時間數列，其轉折點的認定相當簡單，甚至用目視就能分辨出來。但對於大部份趨勢(trend)或非線性的時間數列而言，要確定某一點為轉折點是相當複雜的。其實“轉折”一詞本身在

語意學(semantics)上的認定也頗為分歧。因為有很多的現象(包括轉折),我們很難以一般對或錯(true-false)的二值邏輯來予以認定,而模糊邏輯剛好可以突破這方面的限制,這亦是我們考慮以模糊邏輯來辨識與分類一時間數列的原因。因此,在本文中我們應用模糊羣落中心的隸屬度函數(membership function)及平均累加模糊熵辨識演算法,來檢定非線性時間數列的轉型問題。

在建構線性時間數列模型的程序上,結構轉變的檢定先行於單根的檢定。而目前常見有關結構性轉變的檢定,有拉格蘭治乘數檢定,雙光譜檢定(bispectrum test),概似率基準檢定(likelihood ratio based test)以及排列自我迴歸檢定(arranged autoregression test)等方法。其中,Jaikonen與Luukkonen(1988),及Guegan與Pham(1992)曾分別利用拉格蘭治乘數檢定方法來檢定自激門檻自我迴歸(self-excited threshold autoregression),門檻自我迴歸(threshold autoregression),自我迴歸條件非均變異(autoregressive conditional heteroskedasticity)及雙直線性(bilinear)等模型。Tsay(1991)也曾利用排列自我迴歸方法檢定門檻自我迴歸模型。不過上述這些檢定方法是針對特定的非線性時間數列模式設計的,因此僅對特定的模式具有相當的檢定力,但若檢定其他的模式,則檢定力將會降低(De Gooijer與Kumer, 1992)。

由於傳統對結構轉變的時點皆以單一轉折點來表示,但轉型並非一蹴可幾,通常會呈緩慢現象,因此若以單一的轉折點來解釋轉型的現象,似乎不太恰當。所以找一轉折點,不如找一轉折區間,在求出轉折區間後,在對資料建構模式之認定、估計、診斷、預測才有意義。因此我們考慮以模糊理論來探討轉折,進而求得一轉折區間。

2.2 模糊集合(邏輯)

世界上的許多事物,包括人類的語言、思維與決策,都存在有模糊和非量化的特質。尤其是在人的行為中,更具有許多不明確的偏好。有時硬要把不十分確定的現象,用二分法強行分類,就會產生誤謬的結論。在此背景下,模糊集合(邏輯)的概念由Zadeh(1965)首先提出,其乃參考人腦對動態環境所使用模糊度量與分類的原理,對多元複雜的模糊現象,給予較為穩健

描述的解決方法。

模糊理論乃擴展傳統二值邏輯的觀念與運算方法，利用隸屬度函數來表示事物的模糊現象。在傳統的集合論中，一元素只能是屬於某集合，或是不屬於該集合。但是，在模糊集合中的元素，其隸屬程度可能只有部份是屬於某集合。例如，“景氣好”這一名詞，究竟在何種情況之下才是景氣好，並無一明確界限。但模糊邏輯可以顯示出A景氣指標90%的好，而B景氣指標30%的好。在模糊邏輯中，隸屬度的全距通常設定在0到1之間，因此可以將景氣的程度依0到1之間的數值來定，這使“景氣好”此一語言變項，成為一可能性(possibility)的分布，分布的平均值代表人們對“景氣好”評定的數值。

要對模糊的現象予以適當的集合描述，必須應用新的邏輯假設，即論域(指被討論之事物全體的變數，有時亦稱為空間或定義域)上的對象由屬於某一集合到不屬於某一集合是漸進而非突變的。因此，隸屬度函數的設定在模糊理論的測度與分類上，扮演相當重要的角色。因此我們利用這些介於[0, 1]的實數，即可標記人們對被討論之事物的看法。

應用模糊理論檢定是否發生轉折區間時，首先應將時間數列模糊化，然後再對模糊時間數列進行分類。在分類的過程中，我們會應用到模糊羣落中心、隸屬度及模糊熵等觀念。因此將其定義如下：

定義 2.1 (模糊時間數列羣落中心)：設 X_t 為一時間數列， $t = 1, 2, \dots, N$ ， k 為一正整數。若存在一集合 $C = \{C_i \in R; i = 1, 2, \dots, k\}$ ，使得 X_t 中的元素 x_t 與 C 中的元素 C_i 的距離(一般為歐基里得距離)加權平方和最短，即

$$\text{Min} \sum_{t=1}^N \sum_{i=1}^k (\mu_{it})^m \|x_t - C_i\|^2,$$

則稱集合 $C = \{C_i \in R; i = 1, 2, \dots, k\}$ 為時間數列 X_t 的羣落中心。上式中， C_i 表示時間數列 X_t 的第 i 個羣落中心， m 為羣落權數， N 為樣本數。 μ_{it} 為時間數列 X_t 中的元素 x_t 對 C_i 的隸屬度。

定義 2.2 (時間數列羣落隸屬度)：設 X_t 為一時間數列， $t = 1, 2, \dots, N$ ，且 C_1 與 C_2 為時間數列 X_t 的兩個羣落中心。令 μ_{it} ， $i = 1, 2$ 表示時間數列 X_t 中的元素 x_t 對 C_1 、 C_2 的隸屬度，則定義隸屬度為

$$\mu_{it} = \frac{1}{\|x_t - C_i\|}, \quad i = 1, 2, \dots, k \quad \|\cdot\| \text{表示歐基里得距離。}$$

上式中之 μ_{it} 若大於 1，則令 $\mu_{it} = 1$ 。

定義 2.3 (模糊熵)：設 X_t 為一時間數列， $t = 1, 2, \dots, N$ ， μ_{it} 為時間數列 X_t 中的元素 x_t 對羣落中心 C_i 的隸屬度。 x_t 對羣落中心 C_i 的模糊熵定義為：

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

在模糊識別過程中，模糊熵是類似於熱力學中能量的含蘊標準單位，但它和傳統之熵的意義不同。在定義模糊熵時並非使用機率論(probability)的觀念，而是使用可能性理論的觀念。模糊熵表示模糊集合的平均內部訊息量，此訊息量是作為對模糊集合所描述的對象進行分類時，分類的判定標準。因此，將一時間數列化為起因於內部可能訊息量的度量(以模糊熵表示)，可用以有效地辨認此時間數列模型是否有結構性改變的發生。此外，並利用 t 個時間的平均累加模糊熵來觀察模糊熵的訊息變化情形，並據以做為模型轉折分類的標準。平均累加模糊熵定義如下：

$$MS\delta(x_t) = \frac{1}{t} \sum_{i=1}^t \delta(x_i)。$$

經由觀察平均累加模糊熵的走勢，便可以判定此時間數列是否有結構性改變。換言之，若平均累加模糊熵發生顯著的羣聚(連串)變化時，即意謂模型發生結構性變化；反之，若模型無結構性的變化，則平均累加模糊熵的走勢是相當穩定，而不會有顯著的羣聚變化。此外，無論是自然或人文社會科學的研究，時間數列羣落中心的認定相當主觀與分歧，若事先無特定的評量尺度，必會增加實証分析的困難度。因此，模糊分類通常會設定一門檻水準(threshold level) λ ， $0 < \lambda < 1$ ，來決定分類的標準。根據實証經驗， λ 值不應選取太大，否則會造成無法分類；但也不應太小，否則會有分類過多的情形發生，通常我們視情況選取 λ 在 0.1 至 0.001 之間。

利用 λ 門檻水準將欲檢定之模糊時間數列予以分類後，若相對於每個羣落中心的分類結果皆一致，則不用再對此分類結果做調整。但若在某些時點發

生分類不相同的情形時，則以相同時點之隸屬度最高者的分類結果為標準，其原因在於隸屬度最高者表示在某一時點對某個羣落中心有最高的歸屬度，則其分類結果將會比隸屬度較低者更為準確。

在完成統一分類結果的步驟之後，接著要應用模糊統計來檢定分類的轉型是否具有統計顯著性。模型結構改變的檢定法則為： H_0 ：模型無結構改變， H_1 ：模型有結構改變。決策標準為：設定判定水準 α （通常取 α 在0.1至0.05之間），且 α 值的選取應考慮時間數列的樣本數 N 。因此，我們以 $[\alpha N]$ （ $[\alpha N]$ 為小於或等於 αN 的最大整數）作為分類是否顯著的判定標準。若連串的樣本數大於 $[\alpha N]$ ，則此一連串的樣本歸屬於同一類組。當分類的組數超過一組時，則拒絕 H_0 ：模型無結構改變，表示此一時間數列之模型在 λ 顯著分類水準下發生結構性改變，進而找出其轉折區間。

本文我們取兩個羣落中心，此乃根據一般的實証分析經驗及時間數列的走勢研判的(Wu與Chen, 1999)，其分類的步驟如下：

步驟1：先找出時間數列 X_t 的2個羣落中心，並決定時間數列 X_t 對2個羣落中心相對之隸屬度 μ_{it} ， $i = 1, 2$ 。

步驟2：計算出時間數列 X_t 其所對應的模糊熵 $\delta(x_t)$ 、平均累加模糊熵 $MS\delta(x_t)$ 及此數列的中位數 $Median(MS\delta(x_t))$ 。

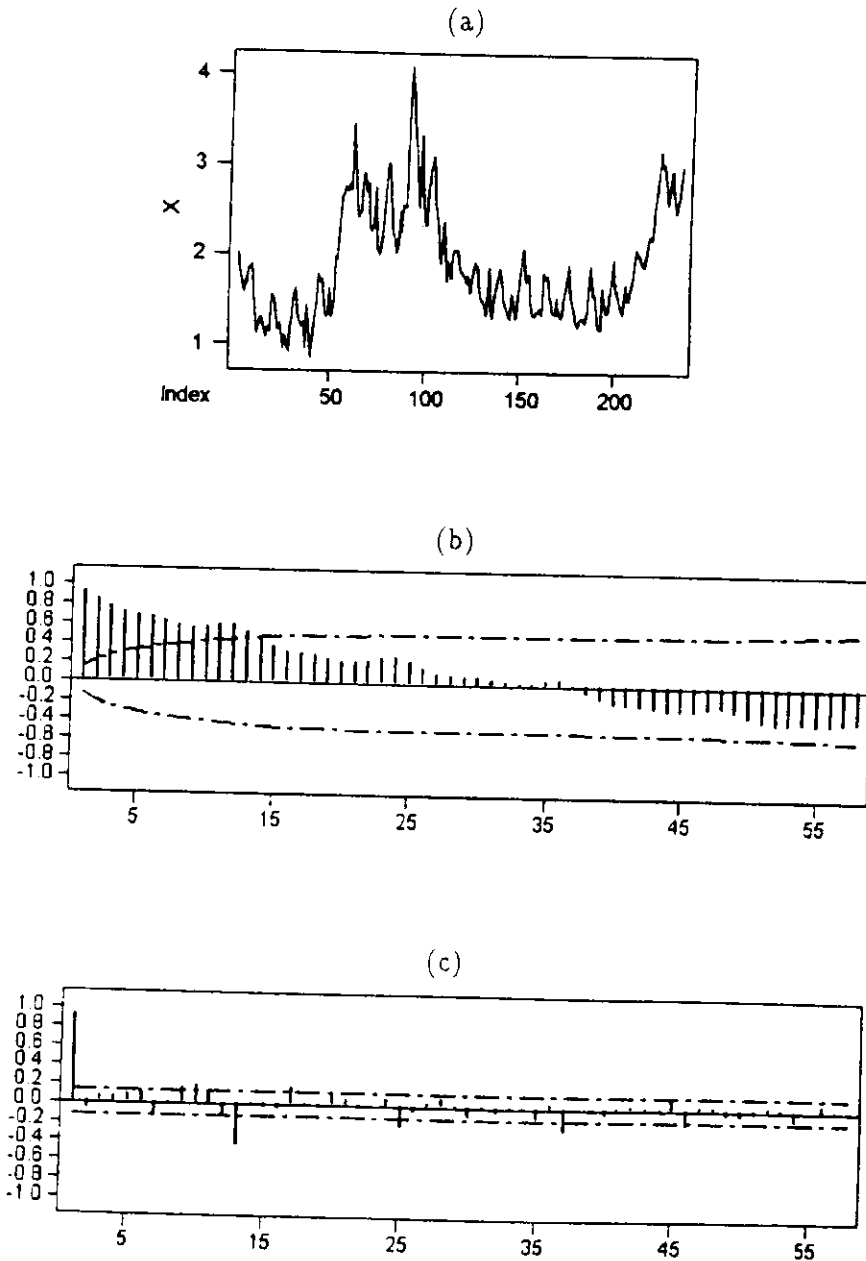
步驟3：選取適當的 λ 門檻水準，將時間數列 X_t 對應的平均累加模糊熵 $MS\delta(x_t)$ 數列進行分類，以1表示第一類組，其平均累加模糊熵 $MS\delta(x_t)$ 落在區間 $[0, Median(MS\delta(x_t)) - \lambda]$ ；以2表示第二類組（轉折集），其平均累加模糊熵 $MS\delta(x_t)$ 落在區間 $[Median(MS\delta(x_t)) - \lambda, Median(MS\delta(x_t)) + \lambda]$ ；以3表示第三類組，其平均累加模糊熵 $MS\delta(x_t)$ 落在區間 $(Median(MS\delta(x_t)) + \lambda, 1]$ 。

步驟4：若分類結果不一致，則對此分類結果做調整。若分類皆相同，則跳至步驟5。

步驟5：選取適當的判定水準 α ，若連串的樣本數大於 $[\alpha N]$ ，則此一連串的

樣本歸屬於同一類組。當分類的組數超過一組時，表示此一時間數列發生結構性改變。進而找出其轉折區間。

圖1. (a)台灣地區失業率走勢圖與其(b)ACF圖以及(c)PACF圖



3. 實證分析與結果

3.1 資料分析

本文資料從民國67年元月至民國86年8月以月為單位的失業率資料(以變數 X 表示)，共計236筆。繪出此時間數列的走勢圖，並參考其ACF(autocorrelation function)圖、PACF(partial autocorrelation function)圖(見圖1)。

3.2 以SARIMA 模式建構

由失業率的樣本ACF圖，可以看出此資料有明顯的季節特性。在本節中，我們使用季節時間數列(seasonal autoregressive and integrated moving average, SARIMA)模式進行資料分析。因為在時差12、24、36期ACF值達高峰，且在時差7、19、31期ACF值達谷底。故我們考慮以下三種候選季節時間數列的模式，比較它們的殘差值及AIC；為了與其它模式預測效果作一比較，我們保留最後4筆資料，接著利用MINTAB套裝軟體，並根據AIC準則，由各候選模式中選取AIC值最小的配適模式。

可能適合的模式	MSE	AIC
$(2, 0, 0) \times (1, 0, 1)_{12}$	0.02951	-809.342
$(1, 0, 0) \times (1, 0, 1)_{12}$	0.02982	-808.918
$(0, 0, 1) \times (1, 0, 1)_{12}$	0.09900	-530.531

因此我們選取的最適模式為SARIMA(2, 0, 0)(1, 0, 1)₁₂，其相關數據如下：

最適模式	$(1 - 0.8335B - 0.1364B^2)(1 - 0.9970B^{12})X_t$ $= (1 - 0.9505B^{12})\varepsilon_t$
MSE	0.02951
AIC	-809.342

3.3 以模糊分類法分類並建立一門檻 ARIMA 模式

在本節中，我們嘗試以模糊分類法來分類，以期能建立更完善的模式。由分析的過程中，我們認為可將此數列分為兩類，分類的步驟如下：

圖2. 平均累加模糊熵的走勢圖

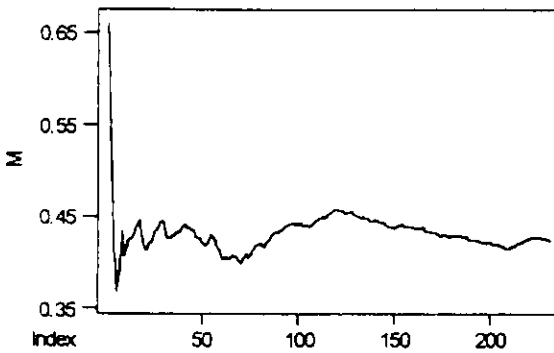
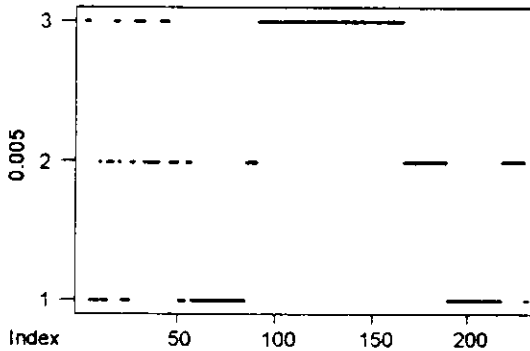


圖3. 圖2中的平均累加模糊熵，以 $\lambda = 0.005$ 所得出的分類圖



步驟1. 利用MINITAB 套裝軟體的k組平均法(k-means method)找出兩個羣落中心 C_1, C_2 ($C_1 = 1.5773, C_2 = 2.7398$)。並定義 X_t 對 C_i 的相對隸屬度函數 μ_{it} 為：

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

步驟2. 計算 X_t 對應的模糊熵 $\delta(x_t)$ 、平均累加模糊熵 $MS\delta(x_t) = \frac{1}{t} \sum_{i=1}^t \delta(x_i)$

及此數列的中位數 $Median(MS\delta(x_t)) = 0.429763$ ，其中平均累加模糊熵的走勢圖如圖 2。

步驟 3. 取門檻值為 $\lambda = 0.005$ ，此處 $\lambda = 0.005$ 為相當嚴格的門檻值。一般 λ 的選取，皆考慮 $0.001 \leq \lambda \leq 0.1$ 。再依據理論方法中的分類法，繪成如圖 3 的分類圖。

取 $\lambda = 0.005$ 時，轉型區間為 80 年 11 月至 82 年 9 月 (167-189 期)，我們依此轉型期，建立一個新的門檻時間數列 (autoregressive and moving average, ARIMA) 模式如下：

$$(1 - 0.9403B)(1 - B^{12})X_t = (1 - 0.8208B^{12})\alpha_t, \quad 1 \leq t \leq 90,$$

$$(1 + 0.3386B)(1 - B)(1 - B^{12})X_t = (1 - 0.9517B^{12})\beta_t, \quad 91 \leq t \leq 166.$$

$$(1 - 0.7650B - 0.0261B^2)(1 - 0.9937B^{12})X_t = (1 + 0.6929B^{12})\gamma_t, \quad 167 \leq t \leq 189,$$

$$(1 - 0.9932B^{12})(1 - B)X_t = (1 - 0.0929B)(1 - 0.8541B^{12})\delta_t, \quad 190 \leq t \leq 232,$$

其中 $\alpha_t \sim WN(0, \sigma_\alpha^2)$, $\beta_t \sim WN(0, \sigma_\beta^2)$, $\gamma_t \sim WN(0, \sigma_\gamma^2)$, $\delta_t \sim WN(0, \sigma_\delta^2)$ 。

3.4 預測結果的比較與分析

模式建構之後，自然地我們會關心其預測能力。我們以 SARIMA 模式與模糊分類法配合門檻模式所得預測結果如下：

	真實值	SARIMA (2, 0, 0)(1, 0, 1) ₁₂	模糊分類法 ARIMA(0, 1, 1)	模糊分類法 SARIMA (0, 1, 1)(1, 0, 1) ₁₂
86年5月	2.51	2.69	2.57	2.71
86年6月	2.67	2.78	2.60	2.90
86年7月	2.85	3.04	2.62	3.20
86年8月	3.03	3.18	2.64	3.37
MSE		0.02951	0.172268	0.073853

其中模糊分類法配合門檻ARIMA模式是以轉折區間之後所配適模式來作預測，而模糊分類法配合門檻SARIMA模式亦是以轉折區間之後所配適模式來作預測。

結果我們發覺在三種方法中，以模糊分類法配適模式(季節性)來預測所得的值其MSE也不至於過大(在三者中僅大於SARIMA模式配適的MSE)；但以其所得資訊較SARIMA模式少(只根據82年10月以後的資料來配適預測)，能有如此的配適結果是相當令人滿意的。從另一個角度來看，也證明了我們的假設：一含結構性轉變的時間數列，若能找出其發生結構性轉變的轉折區間，對於我們欲對資料做進一步的分析與預測，當有更好的效果。

4. 結論

嚴重的失業帶來整個社會貧困與衰退的現象，是一國經濟發展的最大阻礙。然而，目前除了日本、台灣等有令人滿意的低失業水準外，大部分的國家幾乎都籠罩在高失業率的陰霾下；尤其歐美等先進國家，面對每下愈況的失業情形，均不得不喟然提出失業乃是“第一大公敵”。有鑑於失業的日趨嚴重，本文乃以模糊分類法建立一門檻SARIMA模式以期比傳統方法有更好的配適效果與最佳的預測能力。

本文所討論的方法，是針對數列資料隨平均值改變的方向來考慮並加以分類。若是數列本身的特性是變異數不變，而平均值隨時間改變的話，以我們的方法所找出的轉型區間效果會非常的好。但若是數列隨時間改變，其變異數亦改變的話，就很難找出其轉型期。事實上，一筆時間數列的資料，它既是隨時間而改變的資料，光靠有限的經驗與智慧，其實是無法完全準確預測未來所要發生的事。再者，由過去的文獻中我們發現，僅對於非季節性的時間數列資料提出轉折點(或轉折區間)的辨識方法；然而對於季節性的時間數列資料卻未加以深入的探討。因此，如何決定季節性資料的轉折區間，值得吾人深思。是故，我們提出的方法與結果只能提供類似的訊息。

然而，失業現象除了經濟因素的影響外，亦受到非經濟因素的影響。例如，農業及自營作業的吸收能力、廠商不輕易解雇員工、卻志工作效果及已婚婦女的勞動供給與主要家計負責人所得呈負相關等因素。在台灣除這些現

象外，尚有家族企業、自由設立攤販、中小企業多等等。再加上失業保險制度的缺乏，所以台灣的失業率並不能完全反應真正的失業情形。

事實上，由於失業現象的錯綜複雜，當一國遭遇高失業率時，要尋求出到底是由那種經濟變數所引起或區分出是屬於何種失業型態並不容易，因此政策上如何適當的應用也就不像理論那麼單純。唯有好好瞭解自己國家的經濟結構、發展型態與社會背景等，才能掌握失業的真正成因，也才能施以對策，杜絕後患之憂！

參考文獻

- Bleaney, M. (1990). Some comparisons of the relative power of simple tests for structure change in regression models. *Journal of Forecasting*, 9, 437-444.
- Chow, G.C. (1960). Testing for equality between sets of coefficients in two linear regression. *Econometrics*, 28, 291-605.
- Clymer, J., Corey, P., and Gardner, J. (1992). Discrete event fuzzy airport control. *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 343-351.
- Cutsem, B.V. and Gath, I. (1993). Detection of outliers and robust estimation using fuzzy clustering. *Computational Statistics and Data Analysis*, 15, 47-61.
- De Gooijer, J.G. and Kumar, K. (1992). Some recent developments in nonlinear time series modeling, testing, and forecasting. *International Journal of Forecasting*, 135-156.
- Guegan, D. and Pham, T.D. (1992). Power of the score test against bilinear time series models. *Statistics Sinica*, 2, 157-169.

- Hathaway, R.J. and Bezdek, J.C. (1993). Switching regression models and fuzzy clustering. *IEEE Transactions of Fuzzy Systems*, 1, 195-204.
- Lin, C.F. and Terasvirta, T. (1994). Testing the constancy of regression parameters against continuous structure change. *Journal of Econometrics*, 62, 211-218.
- Saikkonen, P. and Luukkonen, R. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75, 491-499.
- Nyblom, J. (1989). Testing for the constance of parameters over time. *Journal of the American Statistical Association*, 84, 223-230.
- Park, K.S. and Kim, S.H. (1996). A note on the fuzzy weighted additive rule. *Fuzzy Sets and Systems*, 77, 315-320.
- Ploberger, W. and Kramer, W. (1992). The CUSUM-test with OLS residuals. *Econometrics*, 60, 271-285.
- Ramakrishnan, R. and Rao, C.J.M. (1992). The fuzzy weighted additive rule. *Fuzzy Sets and Systems*, 46, 177-187.
- Romer, C. and Kandel, A. (1995). Fuzzy partitions of the sample space and fuzzy parameter hypotheses. *IEEE Transactions on Systems, Man and Cybernetics*, 25, 1314-1321.
- Tsay, R.S. (1991). Detecting and modeling non-linearity in univariate time series analysis. *Statistics Sinica*, 1, 431-451.
- Wu, B. and Chen, M.H. (1999). Use of fuzzy statistical technique in change periods detection of nonlinear time series. *Applied Mathematics and Computation*, 99, 241-254.
- Yoshinari, Y., Pedrycz, W., and Hirota, K. (1993). Construction of fuzzy models through clustering techniques. *Fuzzy Sets and Systems*, 54, 157-165.

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

[民國87年11月9日收稿, 88年2月1日修訂, 88年3月5日接受]

Fuzzy Statistical Cluster Analysis and Its Application to Taiwan's Unemployment

Berlin Wu and Yi-Yun Hsu

Department of Applied Mathematics

Chengchi University

ABSTRACT

The percentage of unemployment is one of the targets that every country emphasizes in the social-economic development. The soaring percentage of lasting unemployment would surely influence the wholesome society. Taiwan has been making an amazing progress in economy and thus are ranked as one of the well-developed country. The imbalanced development of the economy, however, has caused the problem of the sources of manpower. The increasing rise of unemployment manifests the imbalance of the allocation of manpower.

This paper uses the percentage of unemployment as the indication of the supply-and-need variable in labor market. Applying the fuzzy time series classification method, we make an analysis of the variations in Taiwan labor market and seek a further solution to lower the percentage of unemployment. Seeing that the factors to evaluate are prone to change, we should take the fuzzy weight into account. Finally, instead of the application of traditional methods of measuring, fuzzy statistical classification is expected to reflect the current circumstances.

Key words and phrases : Unemployment, fuzzy time series, fuzzy classification.