

## 模糊迴歸參數估計及在景氣對策信號之分析應用

吳柏林

曾能芳

政治大學應數研究所 政治大學統計研究所

### 摘 要

傳統的迴歸是假設觀測值的不確定性來自於隨機，模糊迴歸 (fuzzy regression) 則是假設不確定性來自多重隸屬現象。模糊迴歸採用樣本模糊數  $(x_i, Y_i)$  來對模糊迴歸參數進行估計，其中  $Y_i$  為觀測模糊數。一般模糊參數  $A$  的估計方式是採用線性規劃，求出適當的區間，來將觀測模糊數  $Y_i$  的分佈範圍全部覆蓋住。我們認為此法不能真實地表達出樣本所蘊含的資訊，本研究將另行設立一套模糊參數估計方法，此法對樣本的解釋方式將更為合理，且估計的過程也比線性規劃簡便。迴歸常用來建構經濟和財務的模型，而此種模型經常帶有模糊的特質，例如景氣循環、不規則趨勢等，本文將針對台灣景氣指標進行實務分析，以此說明模糊迴歸模式的實用性。

關鍵詞：模糊迴歸參數估計，三角形隸屬度函數， $h$  截集，最小平方法。

## 1. 前言

在社會科學研究過程中，我們常使用迴歸來描述自變數 $X$ 和因變數 $Y$ 之間的因果關係。迴歸參數的估計是迴歸分析的主要研究課題。傳統迴歸分析均假設觀察值由實際因變數影響加上隨機誤差而產生的。也就是說，因變數是一種帶有不確定性的隨機變數。不過在很多的實務應用上，往往無法滿足這樣的假設。例如有些因變數 $Y$ 的觀察值並不是以單一數值的型式存在，而是帶有多重隸屬的特性。這些觀察值雖然帶有不確定性，但這種不確定的特性是來自於模糊現象，而非隨機現象。

以1998年6月新台幣對美元的匯率為例，統計數字為新台幣/美元=34.8。但這數據只是單一時間點的數值資料，它並不能完全準確地反應出在1998年6月整個時間區間下，新台幣對美元匯率的變動狀況。是故新台幣對美元的匯率這個數字的本身就具有不確定性與模糊性。在這些情況下，如果利用假性的“精確值”，可能誤導模型的建構，也可能擴大預測結果和實際狀態之間的誤差。因此，具有模糊性的數值資料，在計量方法的演算過程中，是否符合人類的邏輯推論與歸納的原則，值得吾人深思。

數值模糊的特性，在經濟的實証研究過程中，應予以慎重考慮。近年來模糊統計和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解釋的一種自然發展結果。傳統的數字資料有需求過度及過度解釋的危險，如果應用模糊的數值資料，較能避免這樣的危險發生。尤其是在人文社會科學領域的研究中，必須慎重的考慮數值模糊的特性運用在預測過程中。近年來，一般也認為在人文社會科學的測度理念裡，模糊統計和模糊相關性的使用可說是數值模式的推廣。

模糊理論(fuzzy theory)目前已被廣泛地應用在統計各種領域。例如模糊決策分析(fuzzy decision making analysis)方面有Wierzchon (1982)，以及Werners (1987)等；模糊統計分類(fuzzy clustering)方面有Gath與Geva (1989)，以及Von Cutsem與Gath (1993)等；模糊迴歸分析方面有Tanaka與Ishibuchi (1993)，以及Yang與Ko (1997)等；模糊時間數列(fuzzy times series)的分析方面有Song與Chisson (1993)，吳柏林、張鈿富與廖敏治(1996)，以及Wu與Hung (1998)等。至於應用在社會調查研究方面截至目前為止，也發表了許多可供

參考的著作如Dubois與Prade(1992)，以及吳柏林與楊文山(1997)。這些研究均獲得不錯的結果。

爲了改變傳統對統計資料的“隨機”觀點，以另一種角度來看待不確定性的問題，我們考慮用模糊觀點來取代隨機觀點。而若由模糊觀點套用迴歸分析來處理模糊資料問題，就稱爲模糊迴歸分析。模糊迴歸分析最早是由Tanaka, Uejima, 與Asai(1980)提出。一般模糊迴歸能夠處理的樣本型式有兩種：一種是實數值樣本 $(x_i, y_i)$ ，其中 $x_i$ 爲實數向量， $y_i$ 爲觀測到的實數值；一種是模糊數(fuzzy number)樣本 $(x_i, Y_i)$ ，其中 $Y_i$ 非屬實數的型式，而是觀測到的模糊數。不論是實數值樣本或是模糊數樣本，想要估計模糊迴歸參數，估計方法大都是利用線性規劃的技巧，來尋求一組能將實數值樣本 $(x_i, y_i)$ 或模糊數樣本 $(x_i, Y_i)$ 全部包括住的區間，再將三角形隸屬函數(triangular membership function)套用在此區間上，如此便求得一組模糊迴歸估計式。Tanaka, Uejima, 與Asai(1982)對線性規劃求取模糊參數的程序有詳盡的介紹。

以上提過，其實可以採用實數值樣本 $(x_i, y_i)$ 來估計模糊迴歸參數，這是因爲使用者相信觀測值 $(x_i, y_i)$ 內含有強烈的隸屬度資訊。不過這樣的說法亦較欠缺說服力，因爲隸屬度是分佈在區間上的，而樣本 $(x_i, y_i)$ 則是以實數值的型式存在，在“實數線”上能否隱含正確的隸屬度資訊是值得商榷的。要得到正確的估計結果，最好是收集具有代表性的隸屬函數樣本，所謂具有代表性的隸屬函數樣本應該是以模糊數的形式存在。所以我們較支持用樣本模糊數 $(x_i, Y_i)$ 來估計模糊參數。至於如何利用樣本模糊數來估計模糊參數，本文將對此問題作詳盡的介紹。

本文內容編排如下。第2節說明模糊迴歸架構。第3節說明模糊迴歸模式的參數估計。第4節做實例探討。最後，第5節是結論。

## 2. 模糊迴歸的架構

考慮具有模糊特性資料時，參數本身可能具有不確定性與模糊性。因此我們首先定義模糊數如下。

定義 2.1 { 模糊數 }：令  $A$  爲一個論域， $A$  的元素設爲

$$\{I_i : I_i \in N, i = 1, 2, \dots, K\},$$

若為佈於論域  $A$  之一述句  $X$  的隸屬度函數以

$$u_A(X) = \{u_1(X), u_2(X), \dots, u_K(X)\}$$

表示，則述句  $X$  的模糊數可表示為

$$X = u_A(X)A = u_1(X)I_1 + u_2(X)I_2 + \dots + u_K(X)I_K。$$

**例 2.1**{您一天吃幾碗飯?}：一天吃幾碗飯以模糊數表示為  $X$ ，集合  $A$  可視為整數論域，即可以吃的碗數，令為  $A = \{0, 1, 2, 3, 4, 5, \dots\}$ ，設您一天吃幾碗飯模糊數隸屬度函數  $X$  為  $u_A(X) = \{u_0(X) = 0.1, u_1(X) = 0.2, u_2(X) = 0.4, u_3(X) = 0.2, u_4(X) = 0.1, u_5(X) = 0\}$ 。因此一天吃幾碗飯模糊數可表示為

$$X = u_A(X)A = 0.1I_0(0) + 0.2I_1(1) + 0.4I_2(2) + 0.2I_3(3) + 0.1I_4(4)。$$

線性模糊迴歸模式常表示成

$$\tilde{Y}(x_i) = A_0 + A_1x_{1i} + A_2x_{2i} + \dots + A_px_{pi}。$$

這裡  $x_i = (1, x_{1i}, x_{2i}, \dots, x_{pi})'$  為自變數向量， $\tilde{Y}(x_i)$  為模糊因變數， $A_m$  ( $m = 1, 2, \dots, p$ ) 為模糊參數。一般是假設模糊參數  $A_m$  具有三角形隸屬度函數，因此，模糊因變數  $\tilde{Y}(x_i)$  的隸屬度函數也屬三角型式。在利用樣本模糊數  $(x_i, Y_i)$  來估計模糊參數  $A_m$  之時，可使用線性規劃技巧估計出  $\tilde{Y}(x_i)$  隸屬度的分佈區間；首先加總所有自變數  $x_i$  之下  $\tilde{Y}(x_i)$  的區間長度，此區間的長度總和要為最小，而且該區間得包含住所有的樣本模糊數，如此便可求得  $\tilde{Y}(x_i)$  隸屬度的分佈區間，更進一步整理出模糊參數  $A_m$  的估計式。

以上用線性規劃估計模糊參數的程序當中，我們以為仍有些許不妥之處，何以要硬性要求  $\tilde{Y}(x_i)$  的估計區間得包含住所有的樣本模糊數  $Y_i$ ？樣本模糊數  $Y_i$  真有蘊含這樣的意義嗎？在本研究中，我們將放寬此項要求，重新解

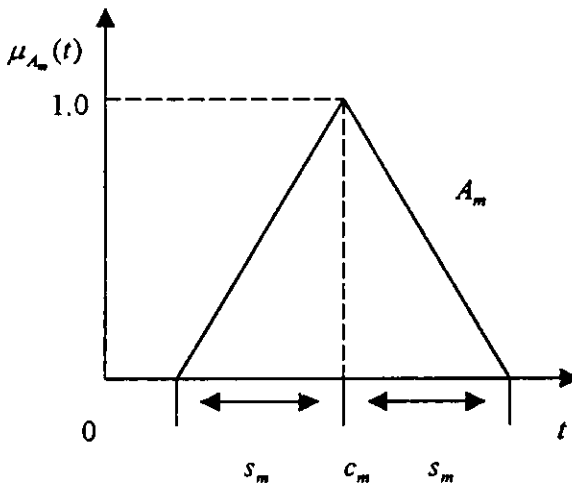
釋樣本模糊數  $Y_i$  在模糊迴歸中的定位。針對此項前提，分別以最小平方方法和無母數方法重新對模糊參數進行估計，我們也因此獲得一套比線性規劃更為簡便的估計程序。Savic 與 Pedrycz (1991) 也曾提出以最小平方方法估計模糊參數的方法，不過該估計程序是由最小平方方法和線性規劃共同組成，與本研究設定的理念並不相同，兩者分屬不同的架構。

對模糊參數  $A_m$  進行估計時，一般是先假設  $A_m$  的隸屬函數為三角型式。 $A_m$  的隸屬函數表達如下：

$$\mu_{A_m}(t) = \max\left\{1 - \frac{|t - c_m|}{s_m}, 0\right\}, \quad -\infty < t < \infty,$$

其中  $c_m$  為三角形的中點， $s_m$  是三角形的分佈半徑。而  $c_m, s_m$  和隸屬函數  $\mu_{A_m}(t)$  的關係如圖 2.1 所示。

圖 2.1：  $c_m$  與  $s_m$  在  $\mu_{A_m}(t)$  中的位置與分佈



為了表示方便起見，我們可令模糊參數  $A_m = \langle c_m, s_m \rangle$ ，則前述的模糊迴歸模式可改寫成

$$\tilde{Y}(x_i) = \langle c_0, s_0 \rangle + \langle c_1, s_1 \rangle x_{1i} + \langle c_2, s_2 \rangle x_{2i} + \cdots + \langle c_p, s_p \rangle x_{pi}。$$

因此，模糊因變數  $\tilde{Y}(x_i)$  的隸屬函數也屬三角型式，可表示成

$$\mu_{\tilde{Y}(x_i)}(t) = \max\left\{1 - \frac{|t - \sum_{m=0}^p c_m x_{mi}|}{\sum_{m=0}^p s_m |x_{mi}|}, 0\right\}, \quad -\infty < t < \infty.$$

至於此式的證明細節可參照 Tanaka, Uejima, 與 Asai(1982) 的研究。

在上述的模糊迴歸模式中，模糊參數  $A_m$  是未知的，這是需要我們估計的部分。但先前已假設  $A_m$  的隸屬函數為三角型態，因此，只要能夠估計出三角形的中點  $c_m$  和半徑  $s_m$ ，便可獲至  $A_m$  隸屬函數的估計值。

### 3. 模糊迴歸模式的參數估計

當輸出結果受到人類判斷或測量不精準的影響時，此輸出結果幾乎不可能以單純的數值型態存在，這種情況下我們考慮用模糊數來表達觀測到的樣本。令觀測到的樣本為  $(x_i, Y_i)$ ， $i = 1, 2, \dots, n$ ，其中  $x_i$  為自變數向量， $Y_i$  為觀測到的模糊數。樣本模糊數  $Y_i$  的隸屬函數也屬三角型式，可表達成

$$\mu_{Y_i}(t) = \max\left\{1 - \frac{|t - y_i|}{r_i}, 0\right\}, \quad -\infty < t < \infty.$$

這裡  $y_i$  為三角形隸屬度函數的中點， $r_i$  則是三角形隸屬度函數的分佈半徑，可令  $Y_i = \langle y_i, r_i \rangle$ 。我們得利用  $y_i, r_i$  對未知的模糊參數  $A_m = \langle c_m, s_m \rangle$  進行估計。在進行估計之前，一般是先定義模糊樣本  $Y_i$  和模糊因變數  $\tilde{Y}(x_i)$  之間的配置衡量值。配置衡量值如下。

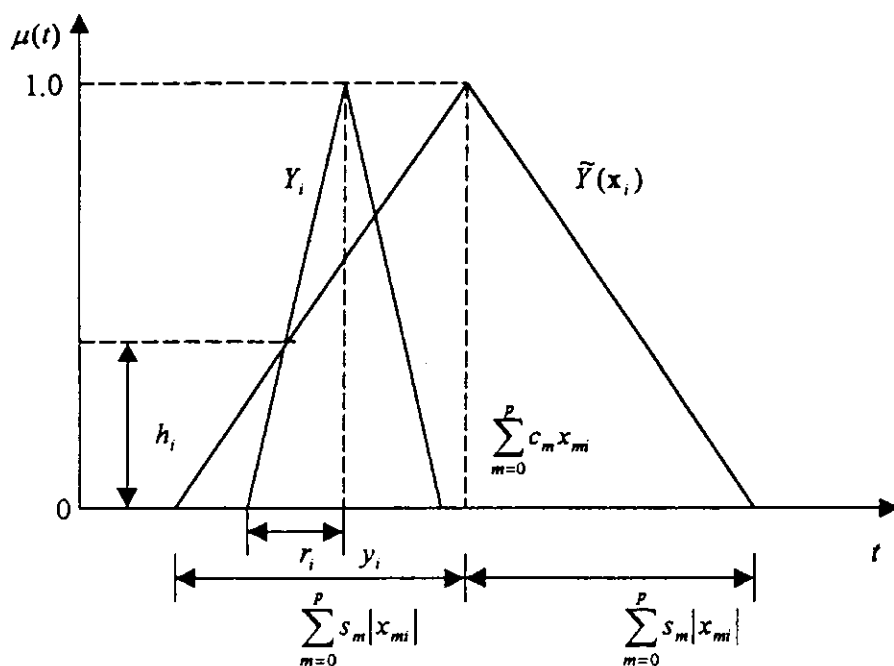
**定義 3.1:** 令  $(x_i, Y_i)$  為一組模糊樣本， $\tilde{Y}(x_i)$  為模糊因變數。若  $[Y_i]^h$  與  $[\tilde{Y}(x_i)]^h$  分別為  $Y_i$  與  $\tilde{Y}(x_i)$  的  $h$  截集(h-cut)，即

$$[Y_i]^h = \{t : \mu_{Y_i}(t) \geq h\},$$

$$[\tilde{Y}(x_i)]^h = \{t : \mu_{\tilde{Y}(x_i)}(t) \geq h\}.$$

則在第  $i$  個樣本之下，能使得  $[Y_i]^h \subset [\tilde{Y}(x_i)]^h$  成立的最大  $h$  值稱為第  $i$  個樣本配置衡量值，以  $h_i$  表示。

樣本配置衡量值是用來衡量樣本值與估計值之間配置程度的重要參考指標。若樣本配置衡量值  $h_i$  愈大，代表樣本  $Y_i$  和因變模糊數  $\tilde{Y}(x_i)$  之間契合(配置)得愈好。 $Y_i, \tilde{Y}(x_i)$  和  $h_i$  的關係如圖 3.1 所示。

圖 3.1 :  $Y_i$ ,  $\tilde{Y}(x_i)$ , 和  $h_i$  的對應圖


性質 3.1 : 在定義 3.1 條件下, 配置衡量值  $h_i$  可改寫成

$$h_i = 1 - \frac{|y_i - \sum_{m=0}^P c_m x_{m_i}|}{\sum_{m=0}^P s_m |x_{m_i}| - r_i}.$$

在估計  $c_m, s_m$  之時, 我們常會限制所有的  $h_i$  得大於某特定值  $H$ , 即

$$h_i \geq H, \quad i = 1, 2, \dots, n,$$

其中  $H$  由使用者自行設定。藉此可保證樣本  $Y_i$  和因變模糊數  $\tilde{Y}(x_i)$  之間的契合程度皆能保持在水準  $H$  之上。由此限制式可進一步推演出兩條限制式。對每一個  $i = 1, 2, \dots, n$ , 滿足

$$y_i \geq \sum_{m=0}^P c_m x_{m_i} - (1-H) \sum_{m=0}^P s_m |x_{m_i}| + (1-H)r_i,$$

$$y_i \leq \sum_{m=0}^p c_m x_{mi} + (1-H) \sum_{m=0}^p s_m |x_{mi}| - (1-H)r_i。$$

再者，我們希望估計得到的模糊因變數  $\tilde{Y}(x_i)$ ，其三角形隸屬函數的半徑（如圖 3.1 所示）愈小愈好，如此隸屬度的資訊才能更集中。因此，將所有  $x_i$  之下的三角形隸屬函數的半徑加總，作為模糊參數估計的目標函數，其定義如下。

**定義 3.2：**模糊迴歸參數目標函數 (objective function)  $O_f$  定義為

$$O_f = \sum_{i=1}^n \sum_{m=0}^p s_m |x_{mi}|,$$

其中  $\sum_{m=0}^p s_m |x_{mi}|$  為  $\tilde{Y}(x_i)$  之隸屬函數的半徑， $\sum_{i=1}^n \sum_{m=0}^p s_m |x_{mi}|$  則是所有  $x_i$  之下代表  $\tilde{Y}(x_i)$  之隸屬函數半徑的總和。

在前述的兩條限制式條件下，我們希望找到可使  $O_f$  為極小的  $c_m, s_m$ 。一般是採用線性規劃來解出  $c_m, s_m$ ，如此便尋得模糊參數  $A_m = \langle c_m, s_m \rangle$  的估計式。

以上曾作出“ $h_i \geq H, i = 1, 2, \dots, n$ ”的限制，這是因為研究者相信：樣本  $Y_i$  會呈現出相當的代表性，其與因變模糊數  $\tilde{Y}(x_i)$  之間的類似（契合）程度至少要大於某個數值  $H$ ，所以在此假設每組樣本  $Y_i$  和因變模糊數  $\tilde{Y}(x_i)$  之間的配置衡量值  $h_i$  均要大於  $H$ 。也由於定義了  $h_i$ ，使得  $Y_i$  的  $h$  截集  $[Y_i]^h$  必定會為  $\tilde{Y}(x_i)$  的  $h$  截集  $[\tilde{Y}(x_i)]^h$  所包含，因此，模糊迴歸的估計其實是在求取一組能將全部樣本概括住的分佈區間，並將此估計區間設計成  $\tilde{Y}(x_i)$  隸屬度的散佈範圍 (support)。

針對上述的估計程序，我們想要質疑：何以能信任樣本  $Y_i$  和  $\tilde{Y}(x_i)$  之間的配置衡量值  $h_i$ ？若採用了配置衡量值  $h_i$ ，便表示  $[Y_i]^h$  僅能落在  $[\tilde{Y}(x_i)]^h$  的內部（因為  $[Y_i]^h \subset [\tilde{Y}(x_i)]^h$ ）。一旦出現離羣值 (outlier)，仍依照以上的估計方法，找出一組正好包住這些樣本的區間，則會得一組範圍甚寬的區間。由於散佈範圍甚廣，這樣的區間其參考價值並不大。我們願意相信樣本  $Y_i$  是具備相當代表性的，不過具有代表性的樣本其  $h$  截集一定得落在  $\tilde{Y}(x_i)$  的  $h$  截集之內嗎？這種假設似乎不真穩健性，正確的說法應該是：具有代表性的



樣本  $Y_i$  其  $h$  截集  $[Y_i]^h$  應該要儘量接近模糊因變數  $\tilde{Y}(x_i)$  的  $h$  截集  $[\tilde{Y}(x_i)]^h$ ，即  $[Y_i]^h \approx [\tilde{Y}(x_i)]^h$ ，其中容許  $[Y_i]^h$  不完全座落在  $[\tilde{Y}(x_i)]^h$  之內的情況發生。

根據以上的觀點，我們認為使用  $h_i$  來估計模糊參數太過冒險，無法真實地表達出樣本所蘊含的資訊，我們得重新解釋樣本  $Y_i$  所蘊含的意義。之前曾設立樣本模糊數  $Y_i = \langle y_i, \tau_i \rangle$  屬三角型隸屬函數，所以樣本  $Y_i$  的隸屬度是分佈在區間  $[y_i - \tau_i, y_i + \tau_i]$  上，其中  $y_{Li} = y_i - \tau_i$  為左端點， $y_{Ri} = y_i + \tau_i$  為右端點。樣本  $Y_i$  的左右端點不太可能與  $\tilde{Y}(x_i)$  的左右邊界完全契合，這是由於測量誤差導致樣本和  $\tilde{Y}(x_i)$  之間有所差池。不過，兩者之間仍有緊密的關係。我們認為樣本的左右端點  $y_{Li}, y_{Ri}$  隸屬於  $\tilde{Y}(x_i)$  的程度有一定的水準，設立此水準為  $H$ ，因此假設  $y_{Li}, y_{Ri}$  隸屬於  $\tilde{Y}(x_i)$  的隸屬度會儘量地接近  $H$ ，其間的關係如下所列示

$$\mu_{\tilde{Y}(x_i)}(y_i - \tau_i) \approx H \quad \text{與} \quad \mu_{\tilde{Y}(x_i)}(y_i + \tau_i) \approx H。$$

所謂儘量接近  $H$ ，是可以容許隸屬度大於  $H$  或小於  $H$ ，而不是僅一味地要求均要大於  $H$ 。

根據以上對樣本的解釋，本研究重新來對模糊參數作估計，則可得到一套更符合直覺、簡便的估計程序。運用最小平方法，以所有樣本的左端點  $\{(x_i, y_{Li}) : i = 1, 2, \dots, n\}$  配置一迴歸直線，稱之為左端直線估計式

$$y_L(x_i) = \hat{L}_0 + \sum_{m=1}^p \hat{L}_m x_{mi}, \quad i = 1, 2, \dots, n,$$

其中  $L_m$  為迴歸係數。再以所有的右端點  $\{(x_i, y_{Ri}) : i = 1, 2, \dots, n\}$  配置右端直線估計式

$$y_R(x_i) = \hat{R}_0 + \sum_{m=1}^p \hat{R}_m x_{mi}, \quad i = 1, 2, \dots, n,$$

其中  $R_m$  為迴歸係數。之前已假設樣本的左右端點隸屬於  $\tilde{Y}(x_i)$  的隸屬度會儘量地接近  $H$ ，所以我們以這兩條直線函式來估計  $\tilde{Y}(x_i)$  之  $H$  截集的左右邊界，即

$$[\tilde{Y}(x_i)]^H = [\hat{L}_0 + \sum_{m=1}^p \hat{L}_m x_{mi}, \hat{R}_0 + \sum_{m=1}^p \hat{R}_m x_{mi}]。$$

如此可導出  $\tilde{Y}(x_i)$  的隸屬函數為

$$\mu_{\tilde{Y}(x_i)}(t) = \max\{1 - |t - \sum_{m=0}^p (\frac{\hat{R}_m + \hat{L}_m}{2})x_{mi}| / \sum_{m=0}^p (\frac{\hat{R}_m - \hat{L}_m}{2(1-H)})|x_{mi}|, 0\},$$

其中  $x_{i0} = 1$ 。而且模糊參數  $A_m$  的隸屬函數為

$$\mu_{A_m}(t) = \max\{1 - |t - (\frac{\hat{R}_m + \hat{L}_m}{2})| / (\frac{\hat{R}_m - \hat{L}_m}{2(1-H)}), 0\}。$$

亦即

$$A_m = \langle \frac{\hat{R}_m + \hat{L}_m}{2}, \frac{\hat{R}_m - \hat{L}_m}{2(1-H)} \rangle,$$

其中  $\frac{\hat{R}_m + \hat{L}_m}{2}$  是三角形底邊中點， $\frac{\hat{R}_m - \hat{L}_m}{2(1-H)}$  是分佈半徑。如此便得到一組估計的模糊迴歸模式

$$\begin{aligned} \tilde{Y}(x_i) = & \langle \frac{\hat{R}_0 + \hat{L}_0}{2}, \frac{\hat{R}_0 - \hat{L}_0}{2(1-H)} \rangle + \langle \frac{\hat{R}_1 + \hat{L}_1}{2}, \frac{\hat{R}_1 - \hat{L}_1}{2(1-H)} \rangle x_{1i} + \\ & \langle \frac{\hat{R}_2 + \hat{L}_2}{2}, \frac{\hat{R}_2 - \hat{L}_2}{2(1-H)} \rangle x_{2i} + \cdots + \langle \frac{\hat{R}_p + \hat{L}_p}{2}, \frac{\hat{R}_p - \hat{L}_p}{2(1-H)} \rangle x_{pi}。 \end{aligned} \quad (1)$$

由上可看出，本研究以最小平方法找出模糊迴歸估計式，其演算程序要比線性規劃更形簡化。

在前面的“ $h_i \geq H, i = 1, 2, \dots, n$ ”的限制中，傳統的線性規劃方法是將  $H$  看作樣本與母體的配置衡量值。但在本研究中，我們用的最小平方法所設的  $H$ ，是指樣本端點  $y_{Li}, y_{Ri}$  隸屬於因變數  $\tilde{Y}(x_i)$  的程度。顯然，其端點隸屬於  $\tilde{Y}(x_i)$  的程度本就不應太高。若  $H$  設得愈大，表示樣本的端點與因變數的端點差距愈大，樣本和母體的隸屬函數反而契合得愈差。其實樣本和母體的契合程度還得由實際情況來判斷。一般而言， $H$  儘可能不要設得過大。因為過大的  $H$  值將導致  $\tilde{Y}(x_i)$  之隸屬函數的分佈範圍過於寬廣，有違一般自然經驗法則。

之前我們已改變原模糊估計程序對樣本的解釋方式，舊有的線性規劃程序對樣本的觀念是很保守的，它規定樣本必須落在設定的範圍之內。而我們採用的解釋方式已將樣本的設限放寬，這種前提很適合套用在抽樣態度較開放的程序上，不過，此等抽樣程序將造就較大的樣本離異性，出現離羣值的可能性也跟著大增。受到離羣值的影響，以最小平方法求得的估計式將會產生嚴重的偏頗，以致無法精確估算出模糊參數  $A_m = \langle c_m, s_m \rangle$ 。也就是說，最

小平方法的估計程序不夠穩健(robustness)，易受離羣值的影響，此為最小平方方法的缺點。

在傳統的統計迴歸分析中，無母數迴歸(non-parametric regression)可用來化解離羣值對估計式的干擾。這是一種穩健的估計程序，例如Agee與Turner(1979)利用葛蘭史密斯(Gram-Schmidt)垂直化(orthogonalization)技巧建立了一套無母數迴歸的參數估計程序。為了因應模糊迴歸可能產生離羣值的狀況，本研究將引進無母數迴歸的估計原則，借用Agee與Turner的方法對模糊參數  $A_m = \langle c_m, s_m \rangle$  進行估計。假設  $\tilde{Y}(x_i)$  之  $H$  截集的左邊界為

$$y_L(x_i) = L_0 + \sum_{m=1}^p L_m x_{mi}, \quad i = 1, 2, \dots, n。$$

此處不再用最小平方方法來估計迴歸參數  $L_m$ 。Agee與Turner利用葛蘭史密斯垂直化技巧將自變數  $x_{mi}$  轉換成  $z_{mi}$ ， $x_{mi}$  垂直化成  $z_{mi}$  的程序如下：

$$z_{1i} = x_{1i}, \quad i = 1, 2, \dots, n；$$

$$z_{mi} = x_{mi} - \sum_{k=1}^{m-1} r_{mk} z_{ki}, \quad \text{其中 } m \geq 1 \text{ 且 } i = 1, 2, \dots, n；$$

$$r_{mk} = \left( \sum_{i=1}^n x_{mi} z_{ki} \right) / \left( \sum_{i=1}^n z_{ki}^2 \right)。$$

則形成以下的迴歸函式

$$y_L(z_i) = L'_0 + \sum_{m=1}^p L'_m z_{mi}, \quad i = 1, 2, \dots, n。$$

我們將藉此轉變後的迴歸函式求得  $L'_m$  的估計值，再反轉  $L'_m$ ，最後得到  $L_m$  的估計值。估計  $L'_m$  的程序如下：

- (1) 對於每個  $m$ ，首先設  $L'_m = 0, m = 1, 2, \dots, p$ ；
- (2) 令  $\delta L'_m = \text{med}\{b_{ij}(m) = \frac{y_{Lj} - y_{Li}}{z_{mj} - z_{mi}} : z_{mi} < z_{mj}, 1 \leq i < j \leq n\}$ ；
- (3)  $L'_m \leftarrow L'_m + \delta L'_m$ ；
- (4)  $y_{Li} \leftarrow y_{Li} - \delta L'_m z_{mi}$ ；
- (5) 重複步驟(2)至(4)，直到  $L'_m$  收斂，最後的結果即是  $L'_m$  的估計量  $\tilde{L}'_m$ ；
- (6) 設  $\tilde{L}'_0 = \text{med}\{y_{Li}\}$ 。

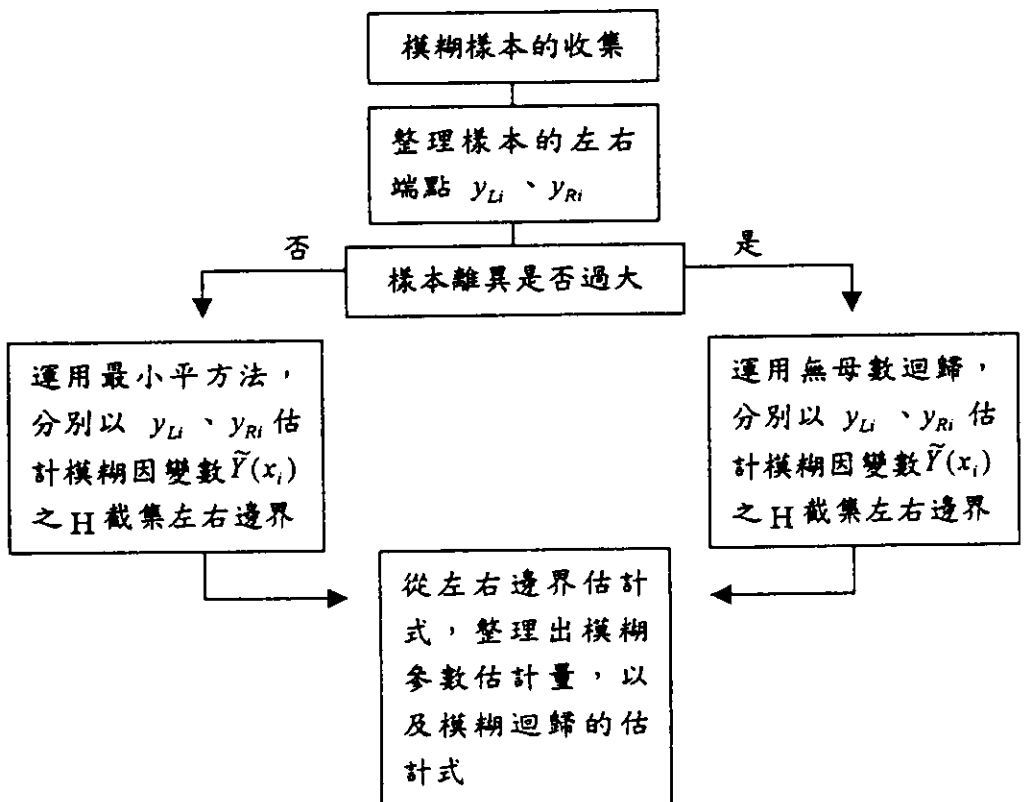
根據以上的步驟可求得  $\tilde{L}'_0, \tilde{L}'_1, \tilde{L}'_2, \dots, \tilde{L}'_p$ 。由於曾經應用葛蘭史密斯垂直化的步驟，最後  $L_m$  的估計量  $\tilde{L}_m$  可表達如下：

$$\tilde{L}_p = \tilde{L}'_p ;$$

$$\tilde{L}_{p-m} = \tilde{L}'_{p-m} - \sum_{i=0}^{m-1} r_{p-i,p-m} \tilde{L}_{m-i}, \quad m = 1, 2, \dots, p-1 ;$$

$$\tilde{L}_0 = \tilde{L}'_0 .$$

圖 3.2：模糊參數估計方法流程圖



藉由這些  $\tilde{L}_0, \tilde{L}_1, \dots, \tilde{L}_p$ ，即可求出  $\tilde{Y}(x_i)$  之  $H$  截集的左邊界。同理，仍運

用此法求取 $\tilde{Y}(x_i)$ 之 $H$ 截集的右邊界。最後將左右邊界的估計式代入式(1)中，便得到以無母數原則導出的模糊迴歸估計式。

本研究根據樣本離異的大小，決定採用最小平方法，或是無母數方法來進行模糊參數估計。但要如何判斷模糊樣本離異之大小？傳統的迴歸分析經常採用統計量

$$t_i = \frac{y_i - \hat{y}_{i(i)}}{\sqrt{\text{var}[y_i - \hat{y}_{i(i)}]}}$$

作為判定離羣值的準則，其中 $\hat{y}_{i(i)}$ 表捨去第 $i$ 個樣本之後，所配置出第 $i$ 個樣本的預測值(predicted value)。若 $|t_i|$ 大於某特定值，則稱第 $i$ 個樣本為離羣值，我們將借用此項準則，用來判斷模糊樣本是否離羣。以所有模糊樣本的左端點 $\{(x_i, y_{Li}) : i = 1, 2, \dots, n\}$ 配置一迴歸直線，並計算每個端點的 $t_i$ 。同理依此計算樣本右端點的 $t_i$ 。若第 $i$ 個模糊樣本的左端點或右端點其 $|t_i| > 3$ ，則稱此樣本為模糊離羣值。如果模糊離羣值的個數超過總樣本數的 $1/10$ ，就稱該組樣本的離異性過大，此時可決定採用無母數估計方法，來消弭離羣值對估計式的干擾。

我們將以上模糊參數估計的觀念流程用圖3.2列示。

#### 4. 實例探討

在收集資料時，由於時間的遲延，有時僅能收集到單一時間點的數值資料，例如：一個月公佈一次的景氣動向指標、一季公佈一次的國內生產毛額(GDP) ... 等。利用此等方式收集到的資料，其表達出的資訊看似真實，實際上它僅表達出某一範圍區間的特定位置，它並沒有清楚地記錄下整個範圍的變動狀況，譬如說：景氣動向雖是一個月僅記錄一次，其實在一個月三十天當中，景氣動向每天都在變化。

這類型的資料是先累積整個範圍下的變動，再加以處理，最後以單一數值的面貌呈現。我們以為此資料並非具有單一數值的意義，它實際所表達的應是某一範圍區間的數值分佈狀況，也就是說，該資料具有模糊的特性。在此情況下，如果利用單一時間點的資料來建構模型，則無法呈現該經濟行為的真實面貌，也可能擴大預測結果和實際狀態間的誤差，在實証研究過程中

，應予以慎重考慮。有鑑於此，本文乃嘗試以模糊迴歸分析，研究台灣景氣對策信號和各項金融面、實質面指標之間的因果關係。爲了加強國家經濟建設的推動，分析國內外經濟動向，由行政院經濟建設委員會每月報導臺灣之景氣指標，其中包括景氣對策信號。景氣對策信號係由貨幣供給M1B、放款金額、票據交換金額、股價指數、製造業新接訂單、海關出口值、工業生產指數、製造業成品存貨率、非農業部門就業率等九項變動率指標編製而成。綜合當期的九項指標，計算其景氣綜合判斷分數後，以數值落入的範圍決定其景氣對策燈號。共有五種信號燈表示景氣狀況，「紅燈」表示景氣過熱，「黃紅燈」表示景氣微熱，「綠燈」表示景氣穩定，「黃藍燈」表示景氣欠佳，「藍燈」表示景氣衰退。

景氣對策信號是具有模糊特性的資料，我們選用模糊迴歸來建構此濟模型。九項指標當中，我們挑選貨幣供給年增率 $X_1$ 、海關出口值年增率 $X_2$ 、工業生產指數年增率 $X_3$ 三項指標作爲自變數，景氣對策燈號則是模糊因變數 $Y$ 。本文資料係從民國79年10月收集至民國87年1月，以月爲單位，共計88筆。

景氣對策信號本是由九項指標依特定公式所組成，上述三項指標與景氣對策信號間的關係似乎已確立，何必再研究這三項指標和景氣對策信號間的關係呢？我們建構此迴歸模式的目的，是重新想以模糊的觀點，呈現出各項景氣指標的另一種面貌。景氣對策信號的表示法不光只是紅、黃紅、綠、黃藍、藍五種而已，它應表示成受自變數影響的模糊集合，如此才能清楚呈現景氣對策信號的含意。再者，不將九項指標全部列爲自變數，是想以更精簡的方式來表達指標間的因果關係，這才符合迴歸模式的精神。

設立景氣對策信號 $\tilde{Y}(x_i)$ 的迴歸模式爲

$$\tilde{Y}(x_i) = A_0 + A_1x_{i1} + A_2x_{i2} + A_3x_{i3},$$

其中 $x_{i1}$ 爲貨幣供給年增率， $x_{i2}$ 爲海關出口值年增率， $x_{i3}$ 爲工業生產指數年增率（單位皆爲%），而 $A_j = \langle c_j, s_j \rangle$ 爲模糊迴歸係數。對於88筆資料，我們將景氣對策信號視爲樣本模糊數 $Y_i$ ，並依表4.1的燈號與景氣分數的對應關係，將樣本燈號化成分數區間，如此便得到各筆資料的左右端點 $y_{iL}, y_{iR}$ 。

我們想看看這組模糊樣本的離異狀況如何，以便決定該採用最小平方方法

，或是無母數方法來估計模糊參數。計算每個樣本左、右端點的 $t_i$ 值之後，發現有表4.2中的數筆資料有較大的 $|t_i|$ 值。

表4.1：燈號與景氣分數的對應關係

燈號	景氣分數範圍
紅	38 - 45
黃紅	32 - 37
綠	23 - 31
黃藍	17 - 22
藍	9 - 16

表4.2：有較大 $|t_i|$ 值的資料

時間	燈號	左端點之 $t_i$	右端點之 $t_i$
83年9月	黃紅	2.26	1.20
84年6月	黃紅	2.18	1.25
84年10月	藍	-2.89	-2.38
84年12月	藍	-2.50	-1.97
85年5月	藍	-2.20	-1.77

不過，表4.2中的 $|t_i|$ 值仍未大於3，未有離羣值發生。因此，我們認為這組樣本的離異並不大，建議採用最小平方方法來估計模糊迴歸模式。

我們曾經提過，樣本的端點隸屬於 $\tilde{Y}(x_i)$ 的程度值 $H$ 應由實際情況來判斷。在景氣信號的實例當中，樣本與母體的契合程度到底達到多少？即 $H$ 值應該設成多少，才能符合實際狀況？根據以往的經驗， $H$ 通常介於0.1至0.5間

。本例中，我們取一個保守值  $H = 0.3$  為準；也就是說，希望樣本的左右端點隸屬於  $\tilde{Y}(x_i)$  的隸屬度會盡量地接近 0.3。

所以，根據樣本的左右端點  $y_{iL}, y_{iR}$ ，可以得到  $\tilde{Y}(x_i)$  之  $H$  截集的左右邊界估計式。運用最小平方法，以所有樣本的左端點  $(x_i, y_{iL})$  配置一迴歸直線，得到左端直線函式

$$y_L(x_i) = \sum_{j=0}^3 L_j x_{ij} = 14.5 + 0.448x_{i1} + 0.116x_{i2} + 0.182x_{i3}。$$

此迴歸配置的相關資訊給在表 4.3。

表 4.3： $y_L$  模型之迴歸參數相關統計量值

迴歸參數	估計係數 $\hat{L}_j$	標準差 $\hat{\sigma}(\hat{L}_j)$	T-Value	P-Value
$L_0$	14.500	0.6461	22.45	0.000
$L_1$	0.448	0.0500	8.96	0.000
$L_2$	0.116	0.0321	3.63	0.000
$L_3$	0.182	0.0711	2.56	0.012
$\hat{\sigma} = 3.057$			$R^2 = 61.0\%$	

再以樣本的右端點  $(x_i, y_{iR})$  配置左端直線函式，得到

$$y_R(x_i) = \sum_{j=0}^3 R_j x_{ij} = 20.2 + 0.550x_{i1} + 0.125x_{i2} + 0.211x_{i3}。$$

此迴歸配置的相關資訊說明在表 4.4 中。

根據式 (3.13) 模糊參數  $A_j$  的計算公式，可求得

$$A_0 = \langle 17.3687, 4.0919 \rangle, \quad A_1 = \langle 0.4992, 0.0727 \rangle,$$

$$A_2 = \langle 0.1205, 0.0057 \rangle, \quad A_3 = \langle 0.1963, 0.0206 \rangle。$$

如此便得到一組估計的模糊迴歸模式

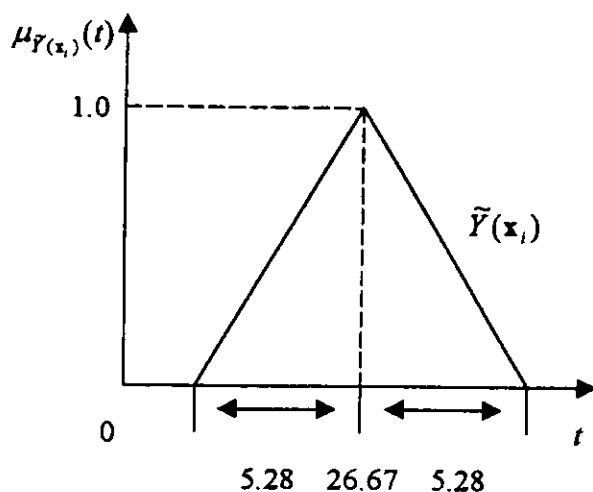


$$\tilde{Y}(x_i) = (17.3687, 4.0919) + (0.4992, 0.0727)x_{i1} + (0.1205, 0.0057)x_{i2} + (0.1963, 0.0206)x_{i3} \circ$$

表 4.4 :  $y_R$  模型之迴歸參數相關統計量值

迴歸參數	估計係數 $\hat{R}_j$	標準差 $\hat{\sigma}(\hat{R}_j)$	T-Value	P-Value
$R_0$	20.200	0.6779	29.85	0.000
$R_1$	0.550	0.0524	10.48	0.000
$R_2$	0.125	0.0336	3.70	0.000
$R_3$	0.211	0.0745	2.83	0.006
$\hat{\sigma} = 3.208$			$R^2 = 67.0\%$	

圖 4.1 : 民國 86 年 10 月台灣景氣對策信號之隸屬函數



我們以此模糊函式來表示景氣對策信號  $\tilde{Y}(x_i)$  和貨幣供給年增率  $x_1$ 、海關出口值年增率  $x_2$ 、以及工業生產指數年增率  $x_3$  之間的因果關係。

有了以上的模糊迴歸函式後，未來對景氣對策信號的表示方法便不再侷限於紅、黃紅、綠、黃藍、藍這五種顏色，也就是說，我們看待景氣對策信號時，不再僅有“顏色”的觀念而已。舉例說明，民國86年10月的貨幣供給年增率 $x_1$ 為12.87，海關出口值年增率 $x_2$ 為5.9，工業生產指數年增率 $x_3$ 為11.05(單位皆為%)，當時的景氣對策信號原以綠燈表示。若將三項自變數指標代入式(4.2)，可得到當期景氣對策信號 $\tilde{Y}(x_i)$ 的隸屬函數為

$$\mu_{\tilde{Y}(x_i)}(t) = \max\{1 - |t - 26.27|/5.28, 0\},$$

並將其列示於圖4.1。圖4.1中的橫軸 $t$ 代表景氣分數，不過它是每個時段皆存在的景氣分數，而非經建會所定義每個月計算一次的景氣分數。我們重新用此隸屬函數來代表民國86年10月的景氣對策信號。此如此一來，便脫離了單純的燈號表示法，此隸屬函數更能充分地表示出景氣信號所隱含的資訊。

## 5. 結論

本研究提出有別於傳統藉由線性規劃估計迴歸參數的方法，因為線性規劃是一種較注重數值運算的估計程序，其觀點與統計學之精神有相當的差異。因此我們應用模糊統計分析與演算法，重新解釋樣本 $Y_i$ 所蘊含的意義。如此一來，樣本的解釋將更符合實際狀況。文中採用最小平方法求得模糊因變數的左右邊界，進而導出模糊迴歸的估計式，應是一套較符合統計原則的估計程序。

比起一般線性規劃的估計技巧，本研究提供的最小平方法亦能充分配合樣本的特性，而且運算程序也較為簡易。雖然最小平方法易受離羣值的影響，有時可能會產生較大的配適誤差。因此我們也考慮引用無母數迴歸估計技巧，以降低離羣值的影響力，增加模糊參數估計的穩健性。因此，比起傳統方法，本研究提出的估計方法更能適切地描述出 $X, Y$ 之間的模糊因果關係。

在景氣對策信號的實例分析當中，最後導出模糊迴歸估計式，可以看出景氣對策信號和三項自變數在結構上的相關程度。我們也藉此式表達出景氣對策信號所隱含的模糊意識。在此例中，也容易看出模糊迴歸較傳統的迴歸具有更進一步的解釋能力。再者，以模糊的觀念來解釋景氣信號的資料結構

，過去某些真實且重要、但常讓人忽略的資訊便可清楚地呈現出來。

另一方面，在傳統的迴歸中，最小平方方法所求出的估計式具有最佳線性不偏性質(best linear unbiased estimate)，但本研究引用最小平方方法所求得的模糊迴歸估計式是否也具有類似的性質？而在模糊架構中，又該如何重新定義最佳線性不偏性質這類的估計特性呢？本文中我們雖提供了一套完善的迴歸參數估計程序，但是對此估計程序的統計性質並未詳細推導其數理特性。所以未來研究的重點將可針對此方向發展。

最後，在景氣信號的實例中，該模糊迴歸估計式雖有著不錯的解釋能力，卻無法對景氣對策信號作有效的預測。因為迴歸的預測，得先掌握自變數的結果，再依據自變數的資訊來對因變數作預測。而景氣對策信號和三項自變數的結果幾乎是同時產生，這時還有必要用自變數去猜測景氣對策信號的結果嗎？為了改進對景氣對策信號的預測能力，未來我們可考慮改以模糊時間序列分析（亦即自相關的模糊迴歸）來建構景氣對策信號，並發展一套較佳的模糊預測的工具。

#### 參考文獻

- 吳柏林、張鈿富、廖敏治(1996)。模糊時間數列與台灣地區中學教師需求人數之預測。國立政治大學學報，73，287-312。
- 吳柏林、楊文山(1997)。模糊統計在社會調查分析的應用。社會科學計量方法發展與應用，楊文山主編：中央研究院中山人文社會科學研究所，289-316。
- Agee, W. S. and Turner, R. H. (1979). Application of Robust Regression to Trajectory data reduction. In Robustness in Statistics (R. L. Launer and G. N. Wilkinson, eds). London: Academic Press.
- Dubois, D. and Prade, H. (1992). Evidence, knowledge, and belief functions. International Journal of Approximate Reasoning, 6, 295-319.
- Gath, I. and Geva, A. (1989). Fuzzy clustering for the estimation of the pa-

- rameters of the components of mixtures of normal distributions. *Patt. Recog. Lett.* 9, 77-86.
- Savic, D.A. and Pedrycz, W. (1991). Evaluation of fuzzy linear regression models. *Fuzzy Set and Systems*, 23, 51-63.
- Song, Q. and Chisson, B. S. (1993). Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54, 269-277.
- Tanaka, H., Uejima, S., and Asai, K. (1980). Fuzzy linear regression models. *International Congress on Applied Systems Research and Cybernetics*. Aculpoco, Mexico.
- Tanaka, H., Uejima, S., and Asai, K. (1982). Linear regression analysis with fuzzy models. *IEEE Trans., Systems Man Cybernet*, 12, 903-907.
- Tanaka, H. and Ishibuchi, H. (1993). An architecture of neural networks with interval weights and its application to fuzzy regression analysis. *Fuzzy Sets and Systems*, 57, 27-39.
- Von Cutsem, B. and Gath, I. (1993). Detection of outliers and robust estimation using fuzzy clustering. *Computational Statistics and Data Analysis*, 15, 47-61.
- Werners, B. (1987). An interactive fuzzy programming system. *Fuzzy Sets and Systems*, 23, 131-147.
- Wierzchon, S. T. (1982). Application of fuzzy decision-making theory to coping with ill-defined problems. *Fuzzy Sets and Systems*, 7, 1-18.
- Wu, B. and Hung, S. (1998). A fuzzy identification procedure for nonlinear time series: with example on ARCH and bilinear models. *Fuzzy Set and System*. To appear.
- Yang, M. and Ko, C. (1997). On cluster-wise fuzzy regression analysis. *IEEE Trans., Systems Man Cybernet*, 27, 1-13.

[民國87年6月24日收稿, 87年11月20日修訂, 87年12月10日接受]

## On Fuzzy Regression Models and Its Applications

Berlin Wu and Nen-Fon Cheng

Department of Mathematics and Department of Statistics  
Chengchi University

### ABSTRACT

In this paper, we propose a parameter estimation method for fuzzy regression models by using the fuzzy number and the method of least squares. Fuzzy regression models are frequently applied in economic or financial modeling. These models exhibit certain kind of linguistic requirements, such as the business cycle and the diversity trend. We take the linguistic prediction as our illustration example for demonstration. Empirical results demonstrate that our estimation procedure can determine fuzzy regression models effectively.

Key words and phrases: Fuzzy regression, fuzzy parameter, triangular membership function, h-cut, methods of least squares.