一、 研究主題

1.1 前言

全民健康醫療保險自民國八十四年三月一日開辦至今已經過九個年頭，當初由公保、勞保、農保等制度蛻變而來的健保制度經過眾人共同打拼、努力經營、篳路藍縷、從無到有，到全民滿意度超越 70%、再到健保雙漲引起軒然大波，再三的突顯出現今政經環境下全民健保永續經營的不易，遂引發本人進行「健保局北區分局財務收支預測模式研究」的動機。

1.2 說明

全民健康醫療保險制度乃是社會保險制度的一環，當初建立的最大宗旨就在透過保險制度的設計，援用大多數人的力量幫助少數、弱勢的族群獲得一定程度以上的醫療照顧；並由中央健康保險局(以下簡稱健保局)成立保險人，國民為被保險人（依不同屬性分為一至六類），一般公司行號、農漁職業工會、地方政府（視被保險人屬性歸屬而定）為投保單位。健保局向投保單位及被保險人依既定的費率收取健保費並支付給全國的特約醫療院所，在收入方面：健保收入主要來自投保對象、投保單位及政府，其主要決定因素為投保對象之種類與其薪資；在支出方面：保險給付已於去年出現重大變革，總額給付費用佔了絕大多數金額，造成以往估算之準則是否適用是值得爭議的。

九年來因人口老化、醫療器材日益昂貴、醫療費用增加、健保費率調整不易、地方政府財務吃緊、保費收入減少、人民就醫習慣不善、醫療資源浪費等不利因素相乘下，全民健保財務出現收入與支出不平衡的嚴重危機。若能分別就收入面與支出面進行個別的分析研究，在眾多預測理論或工具中找尋一有效又實用的方法，建立保費收入與保險支出兩者的預測分析模式，提供使用者財務收支預測的參考依據。

1.3 計劃

邀請健保局北區分局財務單位同仁參與，提供保費收入與醫療費用支出的相關專業知識，採用健保局發展資料分析系統的主要工具— SAS。它具有多種不同功能的模組，經過適當的組合應用，可滿足從使用者介面設計、連結及擷取資料庫的介面、各種常用的統計分析模組、複雜的程式語言能力到各種圖形圖表製作等功能。經過訪談、資訊收集、預測理論研究學習、系統分析設計、程式設計、過濾檢核、預測資料集建立、程式測試、錯誤修正等系統發展步驟，期能完成本研究目標。

1.4 各章節簡要說明

第二章 資訊蒐集與探討：

蒐集與本論文相關領域有關的知識，如資料採礦的意義、時間數列分析法等相關資料。

第三章 研究方法說明：

進一步說明時間數列的意義、時間數列分析法的內容及本研究所用到的各種預測分析模式的細節。

第四章 研究資料來源與前置處理：

說明本研究實驗的資料來源及相關的前置處理過程、方法。

第五章 資料採礦及運算：

詳細說明本研究實驗及運算的細節，各種預測模式分析的內容均以圖形化方式呈現。

第六章 結果評估與解釋：

針對本研究實驗的結果，進行各種預測分析模式特性的比較，並對實驗的資料特性加以評估與解釋。

第七章 結論建議與未來展望：

說明本研究獲得的結論與建議、未來應用及改進的方向。

二、 資訊蒐集與探討

2.1 資料採礦的意義

資料採礦（Data mining） 一詞，有人稱為知識發現 (Knowledge Discovery in Database，KDD)、資料發掘(Information Discovery)、知識萃取(Knowledge Extraction)等。在各種文獻的中譯名稱有： 資料探勘、資料挖掘、資料採礦、資料挖礦等，中華資料採礦協會(Chung-Hua Data Mining Socie ty，CDMS)譯為『資料採礦』。該協會指出：資料採礦最早是由Fayyad 於1991 年提出，主要目的是從龐大的資料中找出規則。Berry [1] 則認為資料採礦是從大量的資料中，利用自動或半自動的方式，從中分析找出有意義的關聯或法則。Greenfeld [2] 則認為資料採礦是知識發現的整個過程。Hand [3] 的說法是： Data mining is the process of seeking interesting or valuable information in large data bases.。

學者Peacock [4] 從狹義和廣義的角度來定義資料採礦：

狹義的資料採礦指的是自動發現隱藏在資料中有意義但不明顯的模式，所謂有意義即指有可能會改變策略或戰略，甚至影響到組織目標。在方法論上是強調其發現的過程。

廣義的資料採礦，則強調其欲研究或測試發現資料彼此間的關係，故使用統計方法、建立假設，研究並確認彼此間的關係以支持在狹義的資料採礦中發現的模式。

最廣義的定義即為資料採礦和資料庫知識發現同義，包含獲取內部與外部資料、資料的轉換、整理、格式化、分析、辨認、賦予資料含義、建立及執行決策支援系統與工具，使其結果能發揮效用。

綜合以上的定義：資料採礦即是在資料庫中，利用各種不同的分析工具與方法，將以往累積的大量歷史資料，進行整理分析、歸納、預測及整合，找出有價值的隱藏事件進一步加以分析，粹取出有用的資訊或是使用者有興趣的樣式（Patterns）與知識，提供管理階層作為決策的依據。

從以上定義中，對於資料的採礦可歸納出幾項重點：

1. 通常是大量的資料才會需要應用資料採礦的技術。
2. 資料採礦是多步驟的，對於不同類的資料，應採取適當的演算法。
3. 資料採礦的目的是找出資料之間的關聯性與發現其樣式，這些結果或知識是先前未知但有意義的。
4. 資料採礦的結果可做為預測及決策時的參考。

2.2 資料採礦和知識發掘的關係

資料採礦是所謂知識發掘(Knowledge Discovery) 的一部份，目前資訊科技和運算工具已大量應用在知識發掘和資料採礦的領域中，圖2-1 就是所謂KDD 之過程，吾人可以清楚的瞭解其中的關係 [5]：

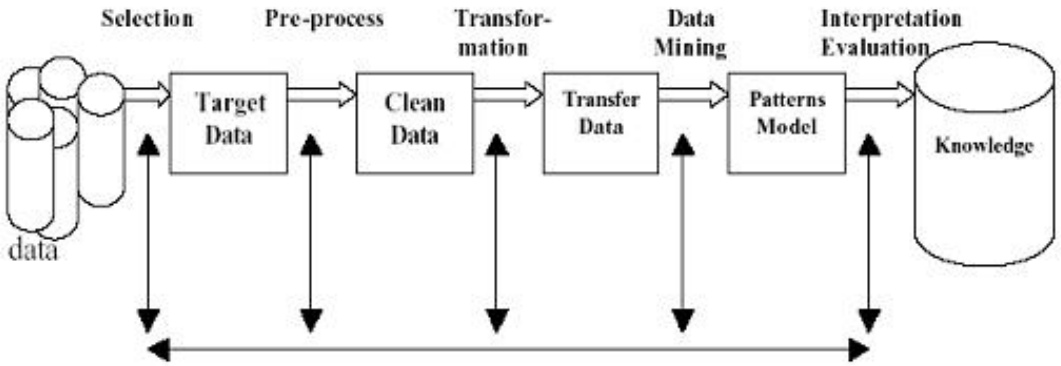


圖 2-1 知識發現的過程

對於資料採礦的進行，有學者提出一系列的步驟供吾人參考 [5],[6]：

1. 瞭解資料與進行的工作
2. 吸收相關知識與技術 (Acquisition)
3. 整合與檢查資料 (Integration and checking)
4. 去除錯誤或不一致的資料 (Data cleaning)
5. 發展分析模式與假設 (Model and hypothesis development)
6. 實際進行資料採礦工作
7. 測試及檢核採礦使用的資料 (Testing and verification)
8. 解釋與應用資料 (Interpretation and use)

2.3 資料採礦的預測工具－時間數列分析

時間數列分析理論自Yule 與Slutsky 兩位教授於1920 年代發明以來，其主要目的之一是對於時間數列的資料作一預測，因為過去、現在、未來之間往往具有密不可分的關係，現在的現象常常是以前事實演變而來，並且會影響未來之變化。所謂時間數列係指「以時間順序型態出現之一連串觀察值的集合… 對某動態系統(Dynamic System)隨時間連續觀察所產生有順序的觀察值之集合」 [7]。意即，是找出在連續的一段時間中，事件發生的相似性。以採用相關的分析方法，探討當時間改變時，其他屬性值的變化，進一步達到預測的目的。

三、 研究方法說明

本章中前三節先說明時間數列的意義、種類及時間數列分析法的基本觀念，第四節討論離群值的觀念及種類，第五節詳細描述本研究中使用的各種時間數列分析預測模式內涵，包含：平均預測法、移動平均預測法、加權移動平均預測法、迴歸分析法、自我迴歸整合移動平均預測法等。

3.1 時間數列的意義

一般而言，所謂時間數列(Time Series)係指以時間順序型態出現之一連串觀測值集合，或更確切地說，對某動態系統(Dynamic System)隨時間連續觀察所產生有順序的觀測值之集合。假若這種集合屬於連續型(Continuous)，則稱為連續型時間數列，假若這種集合屬於離散型(Discrete)，則稱為離散型時間數列。離散型時間數列一般可從二方面產生，一可從連續型時間數列中抽樣(Sampling)而得，另可由在一個期間內對某一變量累積(Accumulating)而得。

吾人將一個離散型時間數列在時間t1,t2,..,tN時之觀測值，記為 $Z(t_1),Z(t_2),..,Z(t_N)$。本研究僅考慮等長時隔之離散型時間數列，故又以 $Z_1,Z_2,..,Z_t,..,Z_N$代表在等長時隔$t_0+h, t_0+2h,..., t_0+th,..., t_0+Nh$之觀測值，$t_0$為時間之起始點，h代表時間之間隔，則Zt可稱為在時間t之觀測值 [7]。

3.2 平穩型數列與非平穩型數列

由前述吾人可知：時間數列是指對一隨時間動態變化、連續觀察所產生之一有次序觀測值的集合。而時間數列模式的基本假設是，所預測之時間數列產生於一隨機過程(Stochastic Process)，因此，在建立時間數列分析模式時，必須先考慮資料是否平穩，若觀察該時間數列發現具有顯著特徵，即所觀測的值係在同一固定水準或固定區域之間上下變動，且整個數列概略地隨時間變化依然存有這種特徵，且其統計特性（例如平均數或變異數）不隨時間之改變而變化，即屬平穩型數列（Stationary Series），如此方能配適適當的模式；若其呈現出一種漂浮無定向的情形，即歸類為無定向數列或非平穩型數列（NonStationary Series），必須先透過差分的方法將其轉換為平穩型的時間數列。

實務上，大部份的資料以非平穩型的數列居多，對於如何分辨非平穩型數列的方法如下：

1. 該數列具有非常明顯的向上或向下的趨勢。
2. 數列的變異程度隨著時間遞增或遞減。
3. 數列同時具有上二項的特徵。

如何將無定向型數列轉化為平穩型時間數列，首先吾人觀察於圖 3-1 中之時間數列顯示在平均水準與斜率上不同之無定向型數列，該數列經取第一次差分(即 Zt - Zt-1)後，變為僅在水準上不同之無定向型數列，如圖 3-2；再取第二次差分[(Zt - Zt-1)-(Zt-1 - Zt-2)]，則轉為平穩型數列，如圖 3-3。因此，吾人可知，對一無定向型數列，經取連續的差分後，終將變為一穩定型數列。

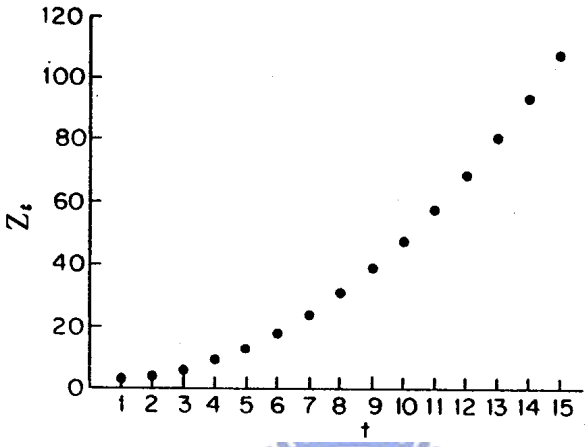


圖3-1 原始數列資料分布圖（差分前）

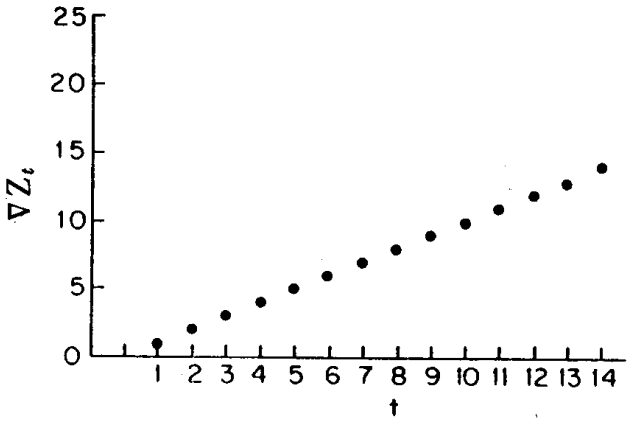


圖 3-2 一階差分資料分布圖

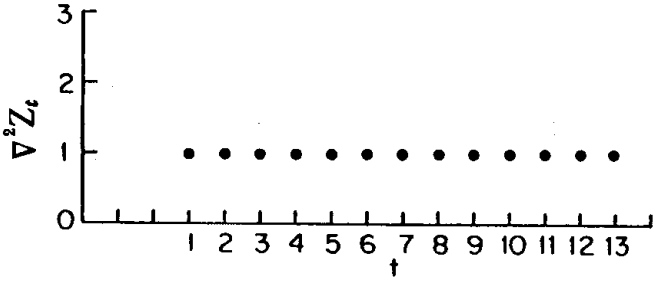


圖 3-3二階差分資料分布圖 [7]

差分運算

吾人定義差分運算子(Difference Operator)▽為

$$\nabla Z_t = Z_t - Z_{t-1} = (1-B)Z_t \quad \text{.....(3.1)}$$

因此，▽與後移運算子 B 之關係為▽ =l－B ，所以，高階之差分可以表示為

$$\nabla^2 = (1-B)^2 \text{ , } \nabla^3 = (1-B)^3 \text{ , ... , } \nabla^d = (1-B)^d \text{ ,}$$

例如:第二次差分可表示為

$$\begin{aligned} \nabla^2 Z_t &= (1-B)^2 Z_t = (1-2B+B^2)Z_t \\ &= Z_t - 2Z_{t-1} + Z_{t-2} \end{aligned} \quad \text{.....(3.2)}$$

若有 n 個觀測值之時間數列{Zt}，經取第 d 次差分後，將轉成為有 n - d 個觀測值之新時間數列 $\{W_t\} = \{\nabla^d Z_t\}$ 。一般而言，欲獲得無定向型時間數列之模式，係假設原始數列經取第 d 次差分(d > 0)後可轉為平穩型數列。

3.3 時間數列分析簡介

時間數列分析理論是Yule 與Slutsky 兩位教授於1920年提出，其主要目的之一是對時間數列的資料作一預測，因為過去、現在和未來之間常常具有密不可分的關係，現在的現象往往是經由以前的事例演變而來，並且還有可能會影響未來情況之變化 [7]。

時間數列分析與預測係一計量方法(quantitative method)，隨著電腦科技的發展，在醫療保健、經濟、社會、人口、環保、管理控制與經營規劃等領域的應用，愈來愈受到重視，預測技術成為甚具參考價值的決策依據。因為透過既有數據的分析，觀察未來供需的趨勢和結構的變化，配合經營模式的調整，將可充分發揮決策運作的效率，提高經營管理水準和獲得最佳經濟效益目標。

預測方法基本上可以分為兩種基本型態：定性方法（Qualitative methods）與定量方法（Quantitative methods），前者通常以專家意見為主，依據過去經驗或特殊感官功能，對未來的事件做本質、特性的預測。後者則是將歷史事件，轉化為時間序列資料趨勢圖，並判別出他們的特徵，以數理方法模式化後再進行量的預測 [8]。

由前述可知，時間數列分析與預測模式就是從過去的觀測值，建立一種適合的模型以預測未來的走勢；也就是說，要找出在一段連續的時間中，事件發生的相似性，採用適當的數理分析方法，探討當時間改變時，其他屬性值產生的變化，達到預測的目的。

吳柏林老師在其所著「時間數列分析導論」指出幾個作決策過程時必須考量的因素 [8]：

1.需要何種型式的預測

預測的型式有三種：點預測、區間預測及等第（rank）預測。

2.預測期間多長

這要按資料與決策的性質，可能需要預測時間點只有幾天或幾週，也有可能長達數月甚至數年。

3.有多少項目需要預測

整體而言，不須對影響系統之每項變數作預測，過多變數的預測，反而會模糊了系統目標，在多變量模式建立過程中，五個變數之系統結構已相當複雜。

4.預測要精確到甚麼程度

預測得精確度關係到管理決策的品質，但精確度較高的預測，相對付出的成本與時間亦較高。

5.系統結構的轉變

由於系統結構性的轉變（structure change），導致需求或供給的時間數列走勢與過去迥異，預測者須配合動態變化的歷史演變，建構符合目前狀況之模式，若自限於過去的經驗則難以對新市場的變遷作一準確之預測。

3.4 離群值之探討

時間數列觀測值有時會受干擾事件之影響，諸如戰爭或罷工之發生、經濟或政治之危機、天氣之突變、甚至鍵入或記錄之錯誤等，這些干擾事件會產生可疑的觀測值造成與數列中其他觀測值的不一致性，這些異常的觀測值稱為離群值(Out Liers)。

假設干擾事件之發生時間與原因是可知道的，則可利用介入模式來分析其影響效應，可是在一般實際應用上，干擾事件之發生時間通常均不可知且離群值可能會使資料分析的結果產生不可靠甚至無效的情形，因此，離群值的處理，在時間數列分析中，也是一個重要的課題。Fox [9] 於 1972 年首先提出時間數列離群值之偵測研究並介紹其模式，其後陸續有許多學者進行研究。離群值包括有相加性離群值(Additive Outlier,簡稱 AO)、創新性離群值(Innovational Outlier,簡稱 IO)、水平移動(Level Shift，簡稱 LS)、及暫時性變動(Temporary Change,簡稱 TC)等。

3.4.1 相加性離群值(AO)

當一種事件的效應僅影響時間數列的一個時期，其最常發生的情況為資料記錄錯誤(如實際值為 3.5 但誤記為 35)所產生的相加性離群值，假設該離群值發生在時點 t=T，則模式表示為

$$Y_t = N_t + \omega_A I_t(T) \quad \cdots\cdots\cdots(3.3)$$

式中 $I_t(T) = 1$ ，當 t = T 時；

$I_t(T) = 0$ ，當 t ≠ T 時，

ω_A 表示 Yt 在時間 t = T 所產生的變動量。

吾人假設在 t=30 有相加性離群值且 ω_A=6，則此種結果所產生的數列，如圖 3-5 所示。吾人可發現該數列除在 t =30 之觀測值有遽增(如圖虛線部份)外，其餘之觀測值均與圖 3-4 相同，整個數列可說是除第 30 點外均沒有改變。

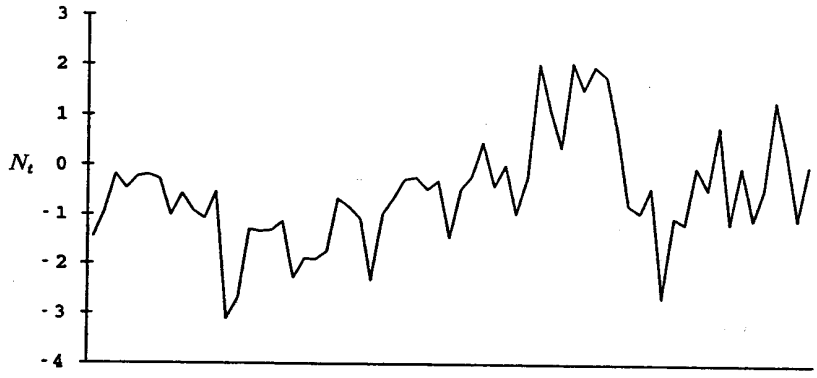


圖 3-4 原始數列資料分布圖

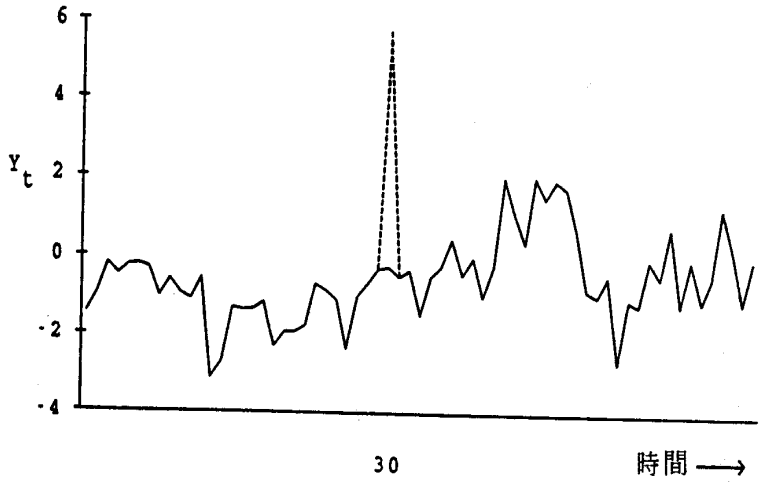


圖 3-5 具 AO 性質的數列資料分布圖 [7]

3.4.2 創新性離群值(IO)

當一個數列有創新性離群值發生，在該時點之後的觀測值均會受到影響。觀測數列之模式可表示為

$$Y_t = N_t + \frac{\theta(B)}{\phi(B)}\omega_I I_t(T) \quad \text{.....(3.4)}$$

或

$$Y_t = \frac{\theta(B)}{\phi(B)}(a_t + \omega_I I_t(T)) \quad \text{.....(3.5)}$$

吾人比較 AO 及 IO 兩種模式，發現 AO 僅受觀測值 Nt 影響，而 IO 受變動項 a_t 影響，因此，AO 僅影響一個觀測值 Yt，而 IO 自 t ≧T 起一段時期根據模式之 ϕ 權重影響其間的觀測值 Yt。

圖 3-6 係表示具 IO 效應的數列 Yt，本例 IO 影響自 t = 30 至 47 期，IO 之影響效應會漸漸消失，最後 Yt 與 Nt 數列間變成沒有差異。

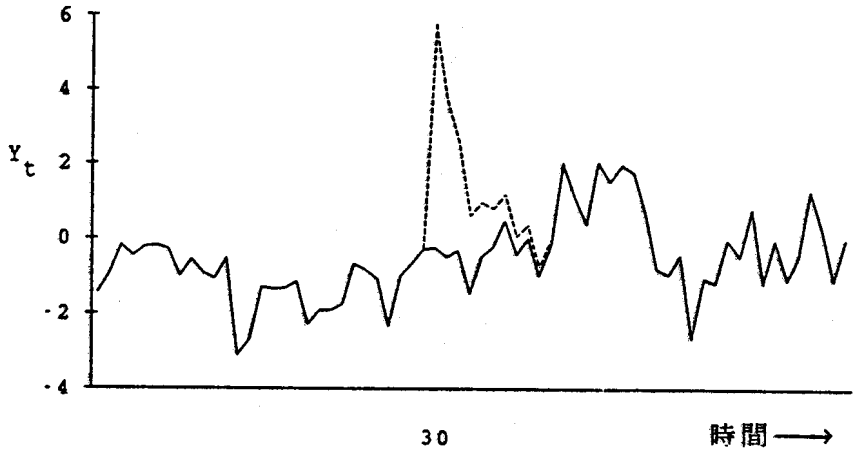


圖 3-6 具 IO 性質的數列資料分布圖

3.4.3 水平移動(LS)

當一種事件的效應對時間數列在已知的期間會呈現永久性的影響，當數列有一種水平性移動發生，此種模式可表示為

$$Y_t = N_t + \frac{1}{1-B}\omega_L I_t(T) \quad \text{.....(3.6)}$$

可改寫為

$$Y_t = N_t + \omega_L S_t(T) \quad \text{.....(3.7)}$$

式中 St(T)稱為階段函數(Step Function)其 St(T)=0 當 t < T , St(T) = l 當 t ≧ T。與 AO 模式比較可知，AO 僅在 t = T 時影響 Yt，LS 係自 t = T 起永久影響 Yt。

圖 3-7 說明模式 LS 之影響，在時點 t = 30 有水平移動現象且$\omega_L = 6$，由該圖很明顯地看出在 t < 30，Yt 與 Nt 數列兩者相同，在 t ≧ 30 Yt 數列值較 Nt 數列值為高(如虛線部份)，其值為$\omega_L = 6$

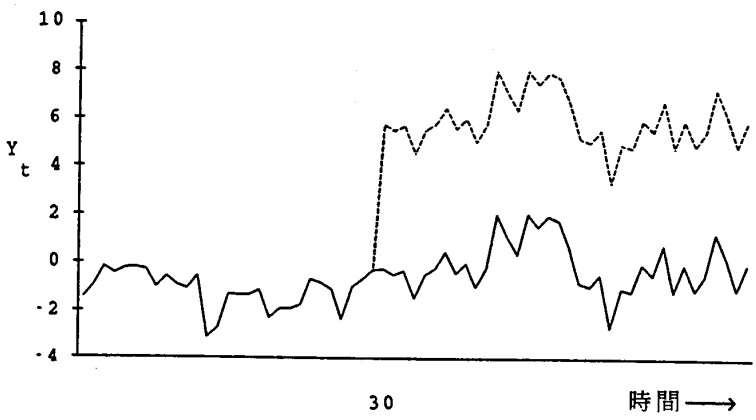


圖 3-7 具 LS 性質的數列資料分布圖

3.4.4 暫時性變動

相加性離群值與水平移動分別代表時間數列受到一種事件影響的兩種不同型態，對相加性離群值而言，此種變動僅會影響一個時期，對水平移動而言，此種變動將會影響到未來所有期間。吾人需考慮一種事件的效應對時間數列起始有影響，然後漸漸消失其影響，此種情況稱為暫時性變動(Tc)，具有一遞減因子(設為 δ)，此種模式為

$$Y_t = N_t + \frac{1}{1-\delta B}\omega_C I_t(T), \quad 0 < \delta < 1 \qquad \text{.....(3.8)}$$

由(3.8)式可知，AO 與 LS 模式為其特例，當 δ=0 時與 AO 模式相同，當 δ=1 時與 LS 模式相同。

3.5 研究方法說明

本研究中使用的時間數列預測模式，有下列幾種：

1. 平均預測法 （The Averaging Forecasting Method）
2. 移動平均法 （The moving average Forecasting Method）
3. 加權移動平均法（weighted moving average Method）
4. 迴歸分析法 （regressive analysis method）
5. 自我迴歸移動平均整合模式ARIMA
   （Autoregressive Integrated Moving Average Model）

3.5.1 平均預測法

這是最簡單、最直覺的預測方法，即下期的預測值為目前所有觀測值總和的平均值，並且餘此類推以計算下下期的預測值，此預測模式的最大優

點是簡單易懂、容易計算，缺點是當時間數列資料集本身變異程度稍大時，其預測準確度的表現差異性頗大。其數學模式如下：

$$Y_{t+1}=(Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3} + \ldots.. + Y_2 + Y_1) / N \quad \ldots..(3.9)$$

Y_t，Y_{t-1}，……，Y_1 分別為在時間 t，t-1，…..，1 時，Y 的實際值

Y_{t+1}，為在時間 t+1 時，Y 的預測值

N 為觀測值總個數，故每個觀測值對下期預測值的貢獻程度均為 1/ N

3.5.2 移動平均預測法

是平均預測法的改良，加入 window size 的觀念，吾人可自行決定下期預測值是由最近 n 期的觀測值加總後計算平均值以求得，此即代表下期預測值是與最近 n 期的觀測值相關，每筆觀測值對下期預測值的貢獻程度均為 1 / n，一般而言，此法比平均預測法有彈性，當 window size n 值等於觀測值總個數 N 時，其預測的結果與平均預測法相同。其數學模式如下：

$$Y_{t+1} = (Y_t + Y_{t-1} + Y_{t-2} + \ldots Y_{t-n+1})/n \quad ..(3.10)$$

Y_t，Y_{t-1}，……，Y_{t-n+1} 分別為在時間 t，t-1，…..，$t-n+1$ 時，Y 的實際值

Y_{t+1} 為在時間 t+1 時，Y 的預測值

n 為選取最近的觀測值個數，意即 window size 數

3.5.3 加權移動平均法

此法又係移動平均法的改良，除了可選取 window size 外，還可進一步針對每一個被選取的觀測值給予設定不同的權值比重（weighting，數學符號記為ω），運用此法，其參與計算的觀測值對下期預測值的貢獻度相較於平均預測法的 1/N ，或移動平均預測法的 1/n，又更具彈性（在此，N 值代表在平均預測法中全體觀測值的總個數；n 值 代表在移動平均預測法

與本預測法中部分觀測值的個數，意即 window size 的意思）。其數學模式如下：

$$Y\hat{} = \sum_{t=1}^{n} \omega_t Y_t \text{，} \sum_{t=1}^{n} \omega_t = 1 \qquad \text{…(3.11)}$$

$Y\hat{}$ 為下期預測值；Y_t 為在時間 t 時，Y 的實際值，t=1，…..，n

ω_t 為在時間 t 時，Y_t 的權值比重（weighting），且所有 ω_t 的總和 = 1，

當每一個觀測值的權值都相等時，其預測結果與前述之移動平均法相同。本研究中所採取的 weighting 方法，係以指數遞減的方法計算，其數學模式為：

$$Y\hat{}_{t+1} = \omega_0 Y_t + \omega_1 Y_{t-1} + \omega_2 Y_{t-2} + \omega_3 Y_{t-3} + \cdots \qquad \text{…(3.12)}$$

設定

$\omega_0 = \omega$

$\omega_1 = \omega(1-\omega)^1$

$\omega_2 = \omega(1-\omega)^2$

……

上（3.12）式可表示為

$$Y\hat{}_{t+1} = \omega Y_t + \omega(1-\omega)Y_{t-1} + \omega(1-\omega)^2 Y_{t-2} + \omega(1-\omega)^3 Y_{t-3} + \cdots \qquad (3.13)$$

可得

$$Y\hat{}_{t+1} = \omega Y_t + (1-\omega)[\ \omega Y_{t-1} + \omega(1-\omega)Y_{t-2} + \omega(1-\omega)^2 Y_{t-3} + \cdots] \qquad (3.14)$$

可得

$$Y\hat{}_{t+1} = \omega Y_t + \omega(1-\omega)Y\hat{}_t \qquad \text{….(3.15)}$$

以下分別為 weight = 0.5，0.25，0.125 時，權值變化的情形：

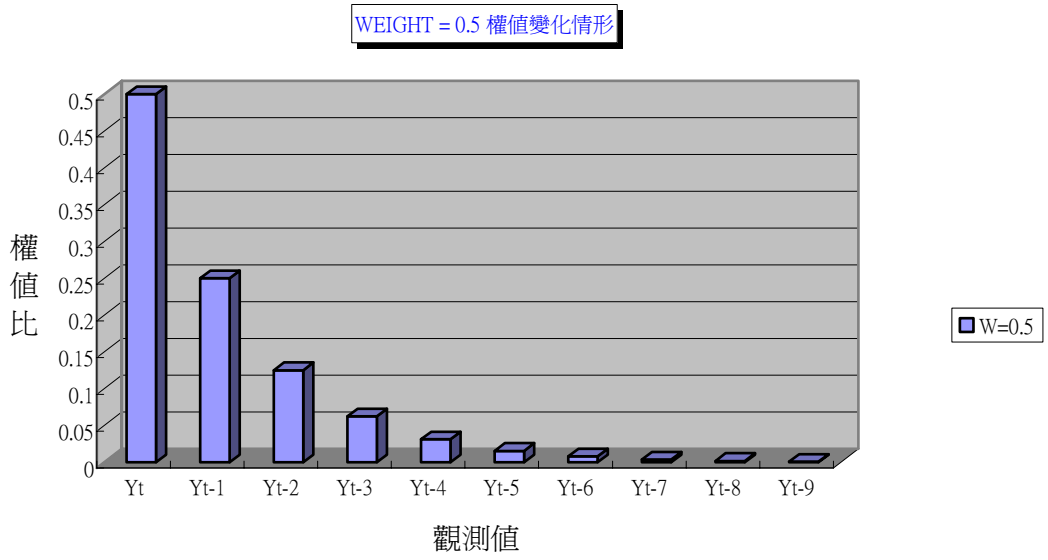


圖 3-8　weight = 0.5 時，權值變化的情形

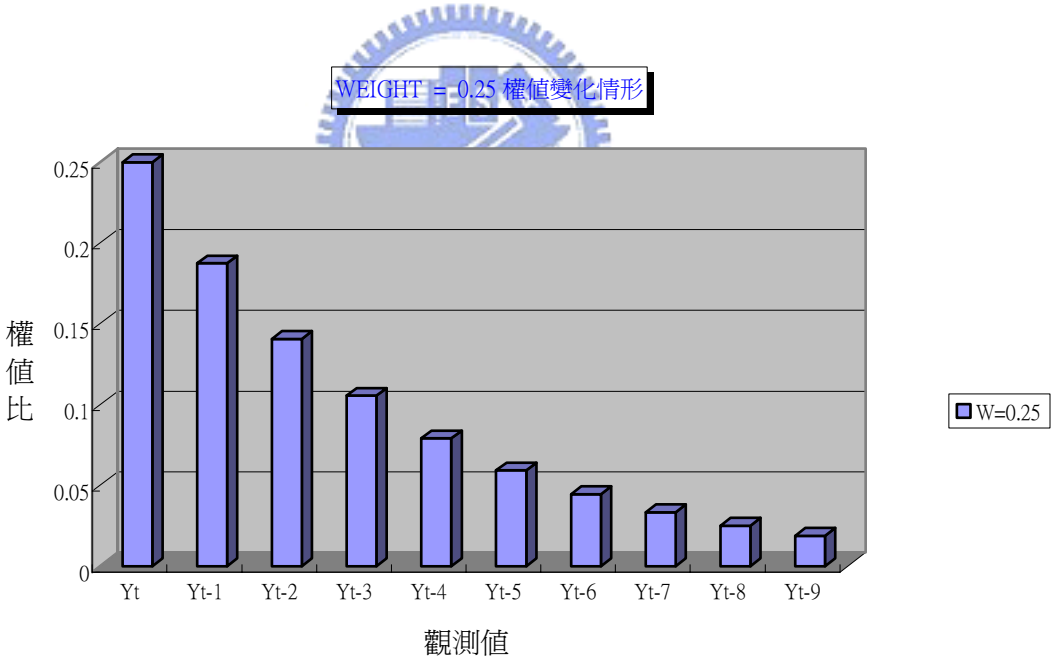


圖 3-9　weight = 0.25 時，權值變化的情形

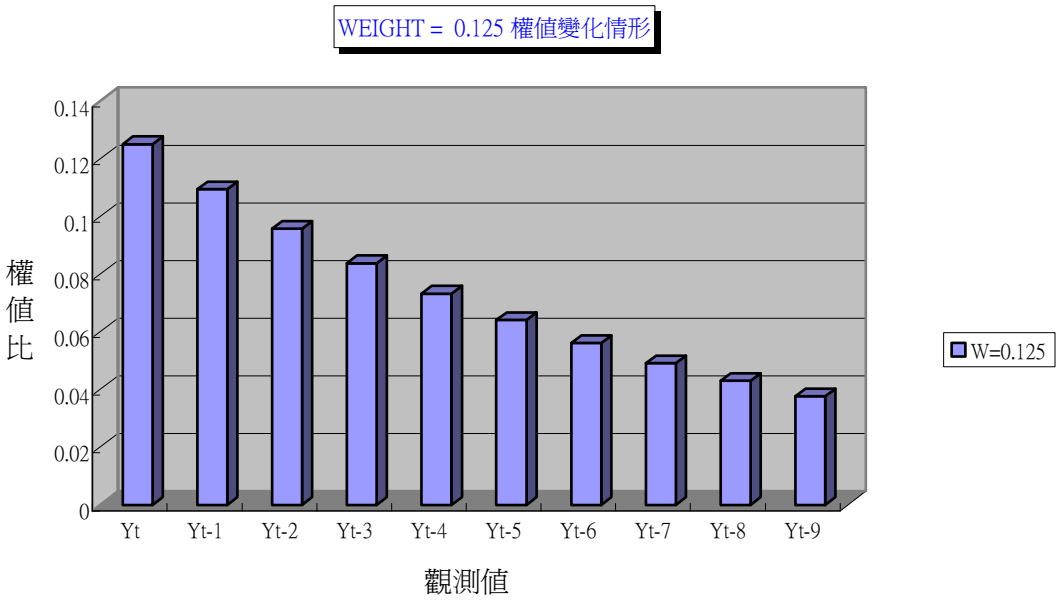


圖 3-10　weight =0.125 時，權值變化的情形

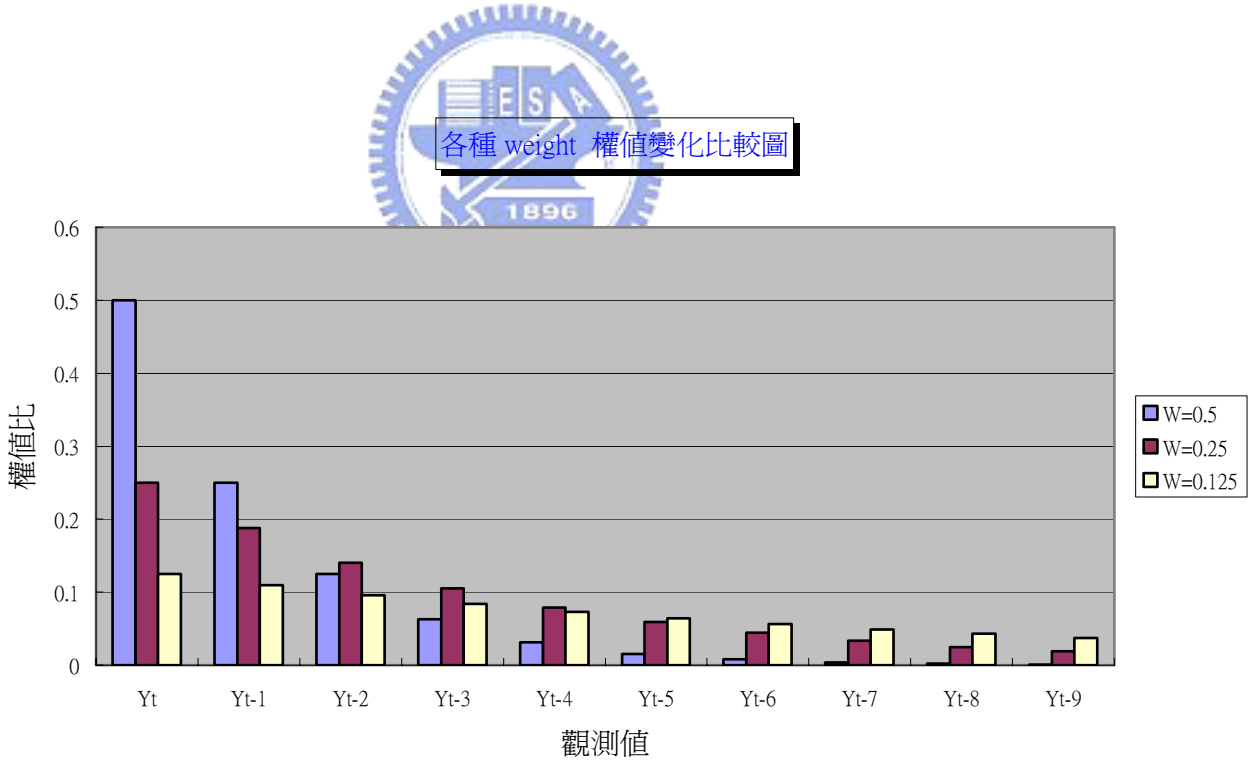


圖 3-11　各種 weight 權值變化比較圖

3.5.4 迴歸分析模式

利用雙變項資料(bivariate-data)通常可以研究相關(Correlation)和預測(Prediction)的問題。相關是指兩個變項之間關聯的強度。瞭解相關通常有二種方式，一為繪製資料散佈圖，另為計算相關係數 [10] 。而當兩變項有相關存在時，則可進行簡單迴歸分析，通常可由一個自變項 X，來預測一個依變項 Y 。

基本的簡單迴歸分析模式，其公式記為 ：

$$Y = A + \beta X \quad \ldots(3.16)$$

A 為截距，β 為斜率，也稱為自變項 X 的迴歸係數

在本研究中，可考慮將時間列為自變項 X ，將欲求得預測值的時間數列視為依變項 Y 代入模式中，如此便可用此模式隨著時間的遞移變化，獲得該時間數列的預測值。一般而言， 簡單迴歸分析建立在某些統計假設的基礎上，其中一個最主要的假設是：每一個誤差項彼此間是不相關的，換言之：假設所有的觀測值其誤差項的變異程度均相同，然而，在時間數列中，其迴歸誤差項卻往往是與時間相關的，這點違反了簡單迴歸分析模式的基本假設。

故欲以迴歸分析方法來分析時間數列時，需考慮其誤差項往往具有與時間相關的特性，也就是說其誤差項具有序列相關(serially correlated)或自我相關(autocorrelated)。當誤差項具有自我相關特性時，對其參數最小平方估計的有效性會有不利的影響，並且會對其標準差的估計產生偏誤；由於簡單迴歸分析模式遵守誤差項彼此間獨立的假設，所以會有底下三種情形必須考慮 [11]：

1. 導致其參數顯著性及預測值信賴區間的統計檢定錯誤。
2. 其迴歸係數估計的有效性比不上有增加誤差項自我相關性質的迴歸模式。
3. 因為這些迴歸的誤差項彼此相關，若在迴歸模式中加以考慮誤差項的相關性，可改善預測的準確度。

因此將簡單的迴歸模式加以改良，將誤差項的自我迴歸性質導入模式中成為：

$$Yt = \beta Xt + Vt \quad \ldots.(3.17)$$

$$Vt = -\varphi_1 Vt\text{-}1 - \varphi_2 Vt\text{-}2 - \ldots - \varphi_m Vt\text{-}m + \varepsilon t \quad \ldots(3.18)$$

式中的 ε t ~ IN (0，σ^2) 表示每一個 ε t 都符合平均值為 0，變異數為 σ^2 的獨立性常態分配，

藉由同時估計迴歸係數 β 及自我迴歸誤差係數 φ i 便可修正簡單迴歸模式造成分析結果偏差的現象。

3.5.5 Box-Jenkins ARIMA 模式

以 ARIMA 模式對資料進行建模工作，若同時採用自我迴歸與移動平均的方式，一般稱為 (p，d，q) 階的整合自我迴歸移動平均模式 (Autoregressive Integrated Moving Average Model of Order (p,d,q)，簡稱 ARIMA(p,d,q)。有五種型式之時間數列模式 [7]：

1. 自我迴歸模式(AR: Autoregressive Model)：

p階自我迴歸模式AR(p)：係指時間數列在第t期的觀測值(Zt)，是由p個前期觀測值為自變數，所形成的迴歸方程式。即觀測值Zt（因變數）可表示為p個前期觀測值（自變數）以及當期干擾項之迴歸模式。模式為：

$$Z_t = C + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t \quad \text{....(3.19)}$$

稱為自我迴歸乃因當期之觀測值Zt為依同一數列諸個前期觀測值 Zt-1,Zt-2,....,Zt-p之迴歸。AR(P)過程可藉後移運算子B，將其改寫為

$$Z_t = C + (\phi_1 B^1 + \phi_2 B^2 + \cdots + \phi_p B^p) Z_t + a_t \quad \text{……(3.20)}$$

或

$$(1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_p B^p) Z_t = C + a_t \quad \text{……(3.21)}$$

假設令

$$\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_p B^p \quad \text{……(3.22)}$$

上式可記為

$$\phi_p(B)Z_t = C + a_t$$

2. 移動平均模式（MA: Moving Average Model）：

q階移動平均模式MA(q)：係指時間數列在第t 期的觀測值(Zt)，是由q個前期的干擾項(Disturbances)所形成之移動平均方程式，即觀測值Zt 可表示為q期的隨機干擾項(Disturbances)之移動線性組合。模式為：

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad \ldots(3.23)$$

即該式為以Zt當作因變數，而以其前q個時期的干擾項當作自變數之迴歸模式。式中 $-\theta_1, -\theta_2, \cdots, -\theta_q$ 為一有限集合的權數，權數的負號僅為方便計算。利用後移運算子，MA(q)的過程可表示為

$$\begin{aligned} Z_t &= \mu + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t \\ &= \mu + \theta_q(B)a_t \end{aligned} \quad \ldots.(3.24)$$

式中 $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$

3. 自我迴歸移動平均模式(Autoregressive-moving Average Models 簡稱 ARMA 模式) ：

p，q 階移動平均模式ARMA(p,q)，為自我迴歸模式AR(p)和移動平均模式 MA(q) 兩者包含在一個模式中，兼具AR(p)與MA(q)之特性。模式為：

$$\begin{aligned} Z_t &= C + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t \\ &\quad + \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \end{aligned} \quad \ldots(3.25)$$

或

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)Z_t = C+(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)a_t \quad \text{....(3.26)}$$

即

$$\phi_p(B)Z_t = C+\theta_q(B)a_t \quad \text{....(3.27)}$$

Zt 為t 期的值，式中(ϕ1,ϕ2,.., ϕp)稱為自我迴歸參數，(θ1, θ2,.., θq)稱為移動平均參數，C為一常數。

4. 自我迴歸模式整合移動平均模式(ARIMA： Auto Regressive Integrated Moving Average Model)：

所謂的整合(integrated)即指當時間數列為非平穩型時，需對其進行差分使其數列成為平穩型，因此ARMA(p,q)模式經過d 次差分後所形成的數列，即為ARIMA(p,d,q)：

$$(1-\phi_1 B-.....-\phi_p B^p)\left[(1-B)^d Z_t-\mu\right]=(1-\theta_1 B-.....-\theta_q B^q)a_t \quad \text{…(3.28)}$$

或

$$(1-\psi_1 B-.....-\psi_{p+q}B^{p+d})Z_t = C+(1-\phi_1 B-.....-\phi_q B^q)a_t \quad \text{…(3.29)}$$

其中

$$1-\psi_1 B-....-\psi_{p+q}B^{p+d} = (1-\phi_1 B-....-\phi_p B^p)(1-B)^d$$

$$C=(1-\phi_1-.....-\phi_p)\mu \quad \text{…..(3.30)}$$

經過差分d 次後所形成的平穩數列，(3.28)也可改寫:

$$Z_t = C+\phi_1 Z_{t-1}+....+\phi_{t-p}Z_{t-p-d}+a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2}-.....-\theta_q a_{t-q} \quad \text{…(3.31)}$$

5. 季節性相乘模式（Seasonal ARIMA Model：Multiplicative Models）

不同年度相同季節(或月份)之觀測值可能彼此具有相依的關係，若將此種季節性或週期性因素納入考慮，則可將前述之時間數列過程擴充為 Seasonal ARIMA(p,d,q) (P,D,Q)s:

$$(1-\Phi_1 B^s - .... - \Phi_p B^{Ps})\left[(1-B^s)^D Z_t - \mu\right] = (1-\Theta_1 B^s - .... - \Theta_Q B^{Qs})a_t \quad (3.32)$$

Φ和Θ為固定之seasonal autoregressive 和moving average 之參數，s 為週期，與ARIMA(p,d,q) 公式(3.28) 和seasonal ARIMA(P,D,Q)公式(3.32)，產生:

$$(1-\phi_1 B^s - ... - \phi_p B^p)(1-\Phi_1 B^s - .... - \Phi_p B^{Ps})(1-B)^d (1-B^s)^D Z_t$$

$$= C + (1-\theta_1 B - .... - \theta_q B^q)(1-\Theta_1 B^s - ... - \Theta_Q B^{Qs})a_t \quad (3.33)$$

3.5.6 Box-Jenkins ARIMA 模式的建構步驟

1970 年Box 與Jenkins 提出進階的建模技術，主要分成四個步驟[7]:

1. 模式的暫定 (Model Identification)
2. 參數的估計 (Parameter Estimation)
3. 診斷檢定 (Diagnostic Checking)
4. 預測 (Forecasting)

茲分述如下：

1. 模式的暫定 (Model Identification)

Box and Jenkins 提供模式暫定步驟的檢定準則：

(1)時間數列繪圖觀察 (Time plot) ：畫出時間數列圖，並觀察其趨勢等特徵，初步檢查資料是否包含異常值或有資料錯誤情形，觀察此資料是否受到結構性變化影響。

(2)自我相關函數 (ACF: Autocorrelation Function) ：判定原數列為平穩型 (stationary) 或非平穩型數列 (nonstationary)，若數列為非平穩型，其ACF會維持許多期的相關，且ACF 的值通常是很緩慢的遞減到0 (die-out)。

(3)差分（Differencing）：即是判定ARIMA (p,d,q) 模式中d 的階數。

(4)ACF 和PACF：判定 p 及 q 的階數，當一數列已為平穩型或差分後成為平穩型，可利用自我相關函數(ACF)及偏自我相關函數(PACF)作為判定 p 及 q 階數的工具，判斷之標準如下表所示：

表3-1：AR 和MA 中p 與q 值的判定

Model	ACF	PACF
AR(p)	呈“指數遞減“或正負相間遞減的形式(die down)	p 期後”截斷” (cuts off after lag p)
MA(q)	q 期後”截斷” (cuts off after lag q)	呈“指數遞減”或正負相間遞減的形式(die down)

＊表中“截斷”的意義為樣本的ACF 與PACF 只有少數幾階的顯著。

2. 參數的估計 (Parameter Estimation)

當模式暫定完後，接著進行參數估計的工作，ARIMA 參數估計方法有二種，一種為最小平方法，一種為最大概似法。所謂最小平方法是計算求得參數的實際值與估計值間之差的平方和為最小，但由於真正的參數值無法得知，所以當樣本數較小時(n＜70)，大多數都利用此法來計算，可得到最佳或最有效的估計值。當樣本數較大時(n＞＝70)，較適用也常用最大概似法(Likelihood Method)，此方法的推導與最小平方法極類似，是求取估計值使得模型的概似函數得到極大值，當此法應用到時間數列時，根據其對初始殘差項不同的處理方式，常用的方法包含以下兩種：

(1)條件概似法(Conditional Likelihood Method)：

令 a0 = a-1 = … .=a1-q = 0

(2)正確概似法(Exact Likelihood Method)：

由資料一同估計a0， a-1，… ，a1-q 與參數。

3. 診斷檢定 (Diagnostic Checking)

模式的檢定在建立模式的過程中，可幫助吾人尋找修正模式的方向、額外得到無法由模式解釋的資訊。主要的工作是在檢定誤差項 at 數列是否符合常態分配且互相獨立的假設，亦即是否為 iid (independent and identical distribution)，也就是假設為 ’白噪音’ (white noise)數列。若檢定的結果為否，則表示此模式不適合且必須加以修正；若檢定的結果為是，則表示此模型的殘差數列應為互相獨立的常態分配，其平均數為 0，變異數為一定值。檢定的方式有許多種，最簡單的就是畫出殘差項的時間數列圖，和偵測殘差數列的ACF，倘若殘差為白噪音數列，則顯示數列並沒有任何顯著的自我相關，另一個檢定殘差為 Q 統計量檢定（Q statistics Ljung & Box, 1978），也就是當n 很大時，Q 統計量會服從自由度k-p-q 的卡方χ2 分配，統計量表示如下：

$$Q = n(n+2)\sum_{k=1}^{k} \frac{r_k^2(\hat{a}_t)}{n-k} \qquad (3.34)$$

式中n 是ˆat 之實際個數，k 為計算殘差自我相關值個數，p 和q 為模式的

參數個數。

最後還要注意的一點就是參數模式的精簡原則(Principle of Parsimony)，意即所用到的參數越少越好。如果通過上述標準之檢驗，就可以利用此模型來進行後續的預測，否則必須放棄此模型，重新回到第一階段的模型暫定步驟，反覆進行，直到模型配適為止。

4. 預測(Forecasting)

一旦找到配適的模式後，就可以利用此模式進行預測，而預測誤差的大小，就是判斷最佳預測模式的重要準則，一般的取決標準可用平均絕對百分誤差（MAPE）或根均方誤差（RMSE）等統計量來衡量預測準確性，本研究中主要採用 MAPE 作為預測精確度的判斷依據。

(1)平均絕對百分誤差(MAPE , Mean Absolute Percentage Error)

定義第 l 期之預測誤差為：

$$e_l = Y_{t+l} - Y_t(l)$$

t 為預測起始點，Y_{t+l} 為 t+l 期之實際值，$Y_{t(l)}$ 為以 t 為起始點第 l 期之預測值，預測到第 n 期的 MAPE:

$$MAPE = \frac{1}{n}\sum_{t+l}^{n}\frac{|e_l|}{Y_{t+l}} \times 100\%$$

(2)根均方誤差（RMSE , Root Mean Square Percentage Error）

定義第 l 期之預測誤差為：

$$e_l = Y_{t+l} - Y_t(l)$$

t 為預測起始點，Y_{t+l} 為 t+l 期之實際值，$Y_{t(l)}$ 為以 t 為起始點第 l 期之預測值，預測到第 n 期的 RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_l^2}$$

一般根據 MAPE 之大小，將模式預測之能力分成四個等級：

< 10	高精確度預測
10 ~ 20	良好的預測
20 ~50	合理的預測
> 50	不正確的預測

Box & Jenkins 在1970 年代把模式建構步驟的四個階段修正得非常有系統，下圖即為本研究所欲採用模式建立之流程圖：

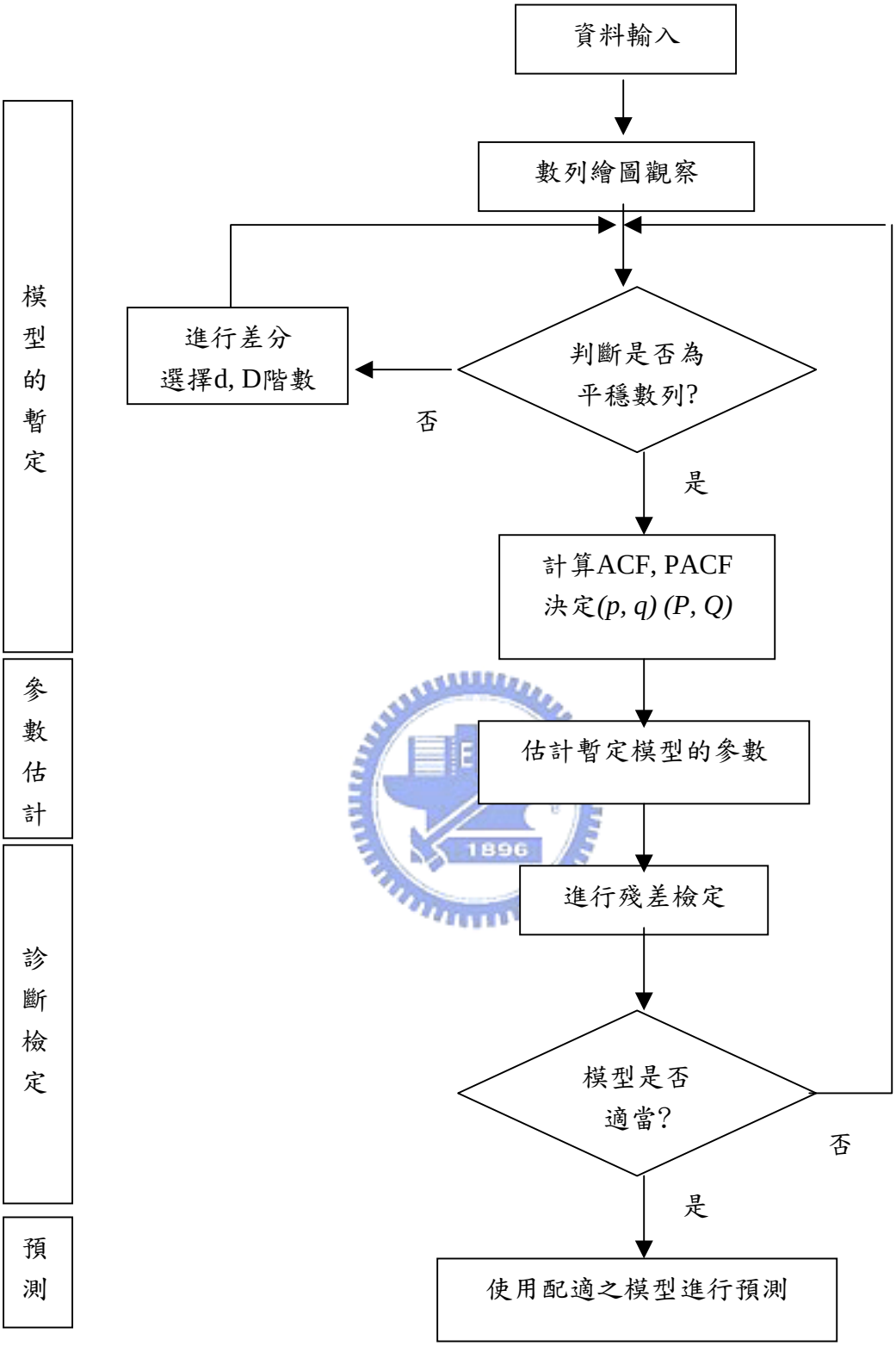


圖3-12 運用Box and Jenkins 模式建立的流程圖 [7]

四、 研究資料來源與前置處理

本章第一節先說明資料來源與分類，第二節說明資料欄位定義，第三節說明資料轉換前置處理過程，第四節說明資料轉換前後檔案格式。

4.1 資料來源說明

中央健康保險局每個月都會產生新的保費收入及醫療費用支出的數據資料，它們是隨著時間遞移產生的時間數列資料，適合用來作為本研究中實驗的資料來源，而且其資料本質一為收入數列，一為支出數列，除可滿足本研究使用各種不同預測模式比較外，尚可比較收入與支出數列特性之不同，可謂一舉數得。本研究使用之資料為中央健康保險局北區分局真實的業務資料，包括保費收入及醫療費用保險支出兩大類，由於本研究蒐集資料期間正逢健保局二代承保財務系統上線後，故保費收入部分有一代及二代的分別，細節如後：

1. 保費收入資料來源
   (1) 應收保費檔 [1]
   (2) 中斷保費檔 [2]
   (3) 滯納金檔 [3]

2. 保險支出來源
   (1) 受理點數檔：門診、藥局受理點數檔 [4]，住診受理點數檔 [5]
   (2) 暫付點數檔 [6]
   (3) 核減點數檔 [7]
   (4) 核付點數檔 [8]

4.2 資料欄位定義

1 保費收入 = 應收保費 + 中斷保費 + 滯納金

2 保險支出 = 核付點數 + 暫付點數

暫付點數 = 受理點數 * 95%

已核付資料 = 若受理點數檔與核付點數檔內均有資料，則算已核付資料

未核付資料 = 若受理點數檔有資料但核付點數檔無資料，則算未核付資料

---

[1] 應收保費檔名=Fbbrc , fbbtrc , fbbprc_6 (一代承保) , F02h_prc , f02h_urc (二代承保)

[2] 中斷保費檔=Ubt_id_cut_d (一代承保) , U07t_cut_dtl , u07h_cut (二代承保)

[3] 滯納金檔=Fbbpar_6 , fbbtar , fbbar (一代承保) , F02h_uar , f02h_par (二代承保)

[4] 門診、藥局受理點數檔=Pbb_op_tlist

[5] 住診受理點數檔=Pbb_hp_tlist

[6] 暫付點數檔=Fdbprep

[7] 核減點數檔=Fdbaprv

[8] 核付點數檔=Fdbaprv

4.3 資料轉換與前置處理

主要是以 SAS 程式分別從一、二代承保系統的資料庫及醫療系統的資料庫（均為 Oracle 8 database），將前述的多項 table 讀取、過濾、處理缺失值、加總並輸出到 NT Server 上，分別產生可供以後預測分析使用的保費收入 SAS 資料集(income_plot , income)及醫療費用支出 SAS 資料集(exp_plot , expand)，程式流程如圖 4-1：

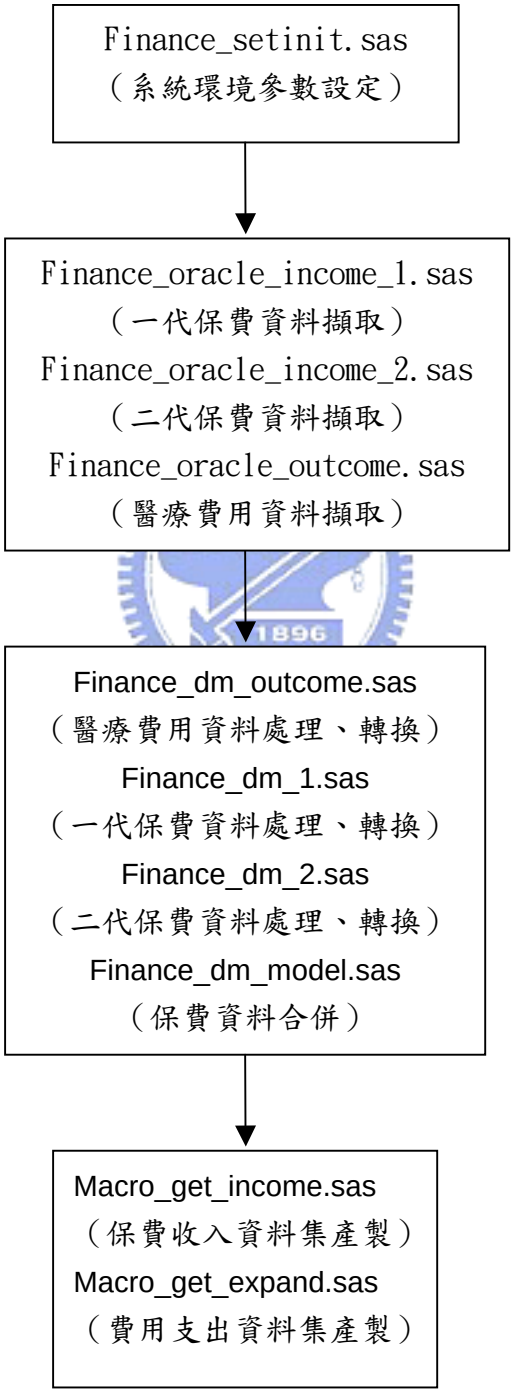


圖 4-1 前置處理程式流程圖

4.4 相關檔案格式

表 4-1 保費收入資料集(income_plot，轉換前) 檔案格式

-----Alphabetic List of Variables and Attributes-----						
#	Variable	Type	Len	Pos	Format	Label
4	AR20_AMT	Num	8	24		滯納金
3	CUT_AMT	Num	8	16		中斷保費
2	UNIT_AMT	Num	8	8		應收保費
1	fee_ym2	Num	8	0	DATE9.	保費年月
5	tot_a	Num	8	32		保費收入

表 4-2 保費收入資料集(income，轉換後) 檔案格式

-----Alphabetic List of Variables and Attributes-----						
#	Variable	Type	Len	Pos	Format	Label
1	prem_ym2	Num	8	0	DATE9.	保費年月
2	unit_amt	Num	8	8		保費收入

表 4-3 費用支出資料集(exp_plot，轉換前）檔案格式

-----Alphabetic List of Variables and Attributes-----							
#	Variable	Type	Len	Pos	Format	Informat	Label
4	APRVPAY_QTY	Num	8	24			簽付點數
1	FEE_YM2	Num	8	0	DATE9.		費用年月
3	SUBTRACT_QTY	Num	8	16	10.	10.	核減點數
2	T_APPL_QTY	Num	8	8	11.	11.	受理點數
5	prepay_amt	Num	8	32	11.	11.	暫付點數
6	tot_a	Num	8	40			費用支出

表 4-4 費用支出資料集(expand，轉換後）檔案格式

-----Alphabetic List of Variables and Attributes-----						
#	Variable	Type	Len	Pos	Format	Label
2	exp_amt	Num	8	8		費用支出
1	prem_ym2	Num	8	0	DATE9.	費用年月

五、 資料採礦及運算

本研究中將以前述所提的六種預測模式（平均預測、移動平均預測、加權移動平均預測、迴歸分析、自我迴歸整合移動平均、季節性自我迴歸整合移動平均），分別針對保費收入及費用支出兩組資料集進行預測分析、以圖形化方式觀察每組模式預測值趨勢，綜合比較每組模式平均絕對百分誤差(MAPE)圖形的變化情形，作為判斷預測模式準確度的參考。

5.1 保費收入預測

本節中共使用六種預測模式進行分析（平均預測、移動平均預測、加權移動平均預測、迴歸分析、自我迴歸整合移動平均、季節性自我迴歸整合移動平均），茲分述如后：

5.1.1 平均預測模式

程式名稱=average.sas
輸入數列=保費收入資料集 income_plot(保費年月,實際金額)
處理條件=觀測值全體
輸出數列=保費收入平均預測資料集
(保費年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-1-1、5-1-2、5-1-3、5-1-4）

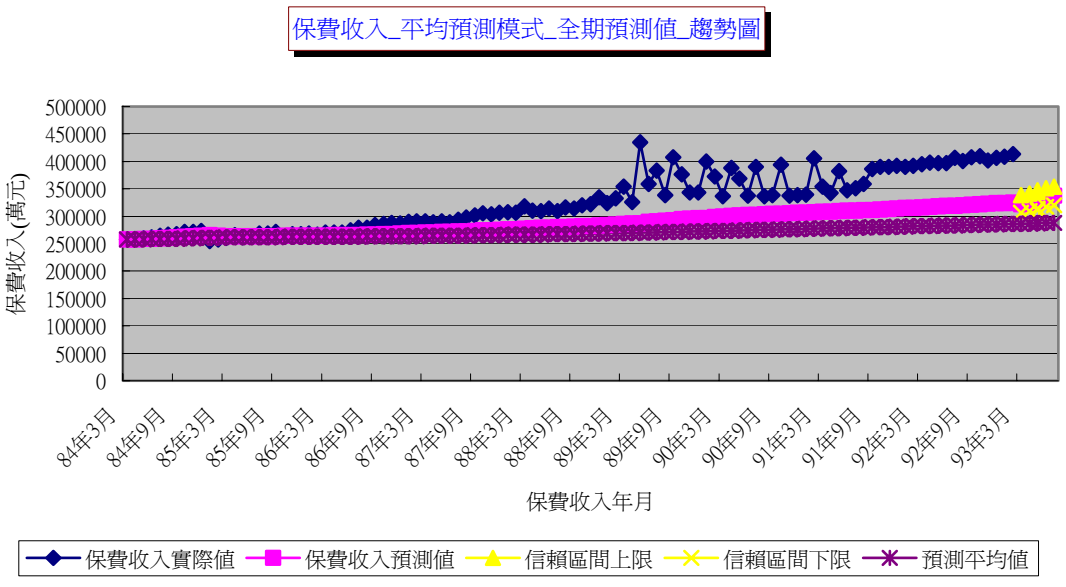


圖 5-1-1 保費收入_平均預測模式_全期預測值_趨勢圖

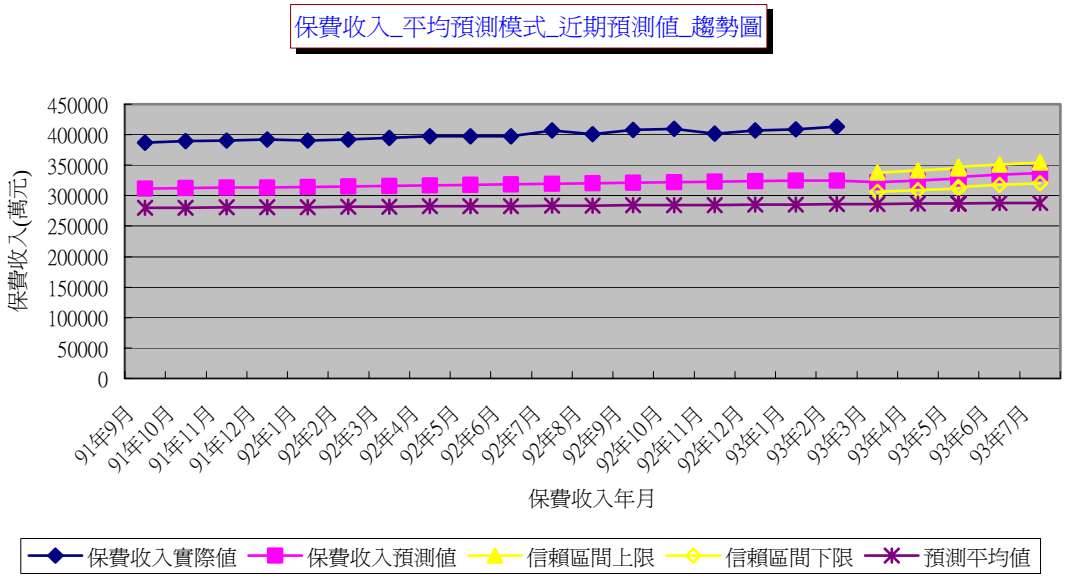


圖 5-1-2 保費收入_平均預測模式_近期預測值_趨勢圖

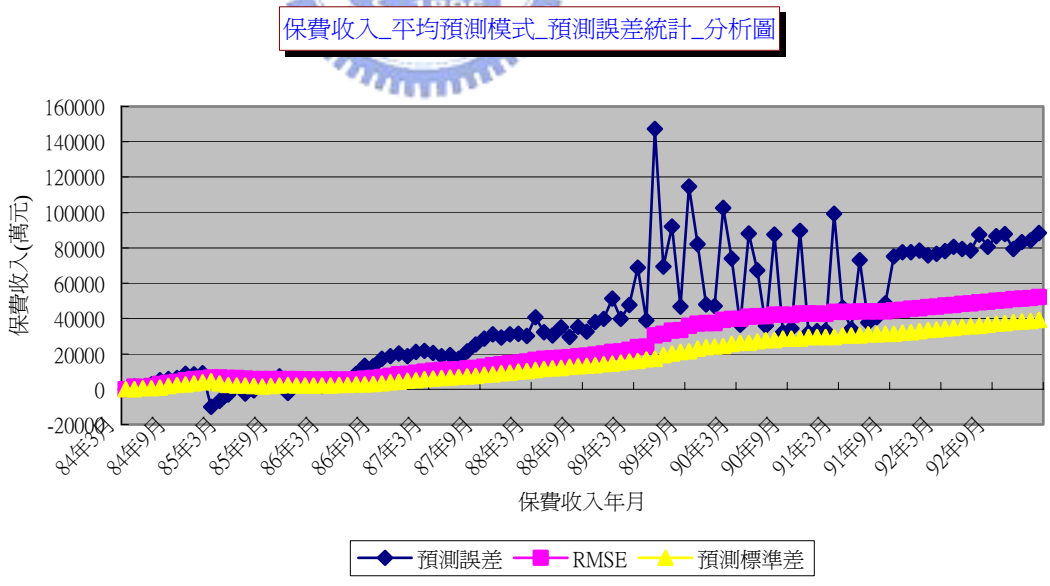


圖 5-1-3 保費收入_平均預測模式_預測誤差統計_分析圖

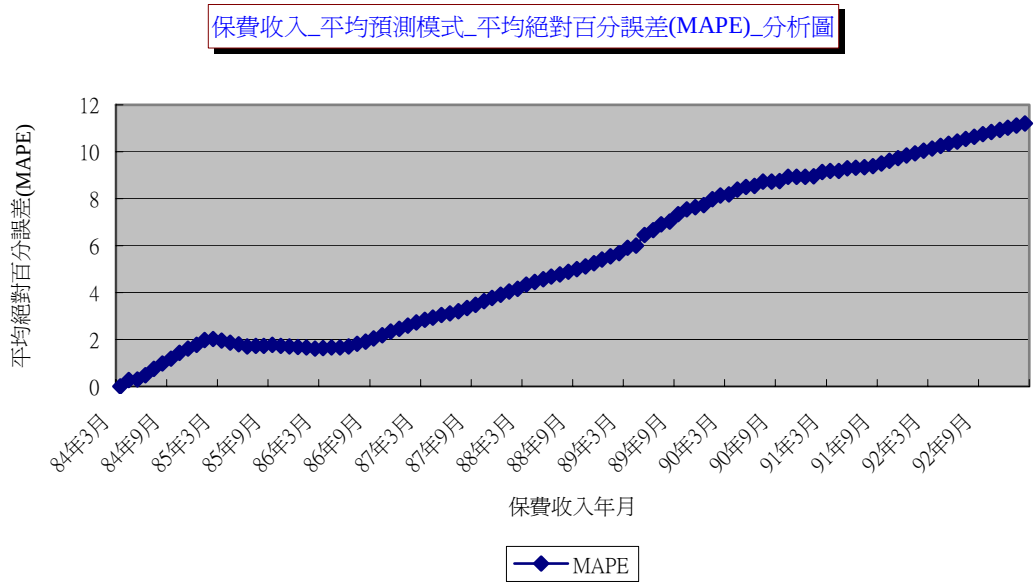


圖 5-1-4 保費收入_平均預測模式_平均絕對百分誤差(MAPE)_分析圖

5.1.2 移動平均預測模式

程式名稱=macro_moving.sas

輸入數列=保費收入資料集 income_plot(保費年月,實際金額)

處理條件=觀測值全體或部分

Window size=(12~108,increase by 12)輸出數列=保費收入移動平均預測資料集

(保費年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-1-5、5-1-6）

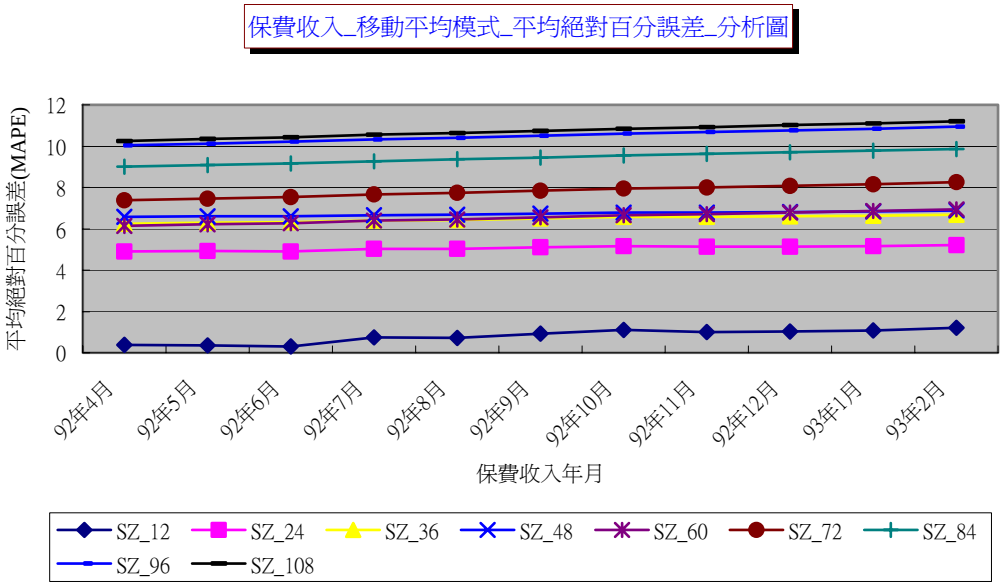


圖 5-1-5 保費收入_移動平均模式_分析圖

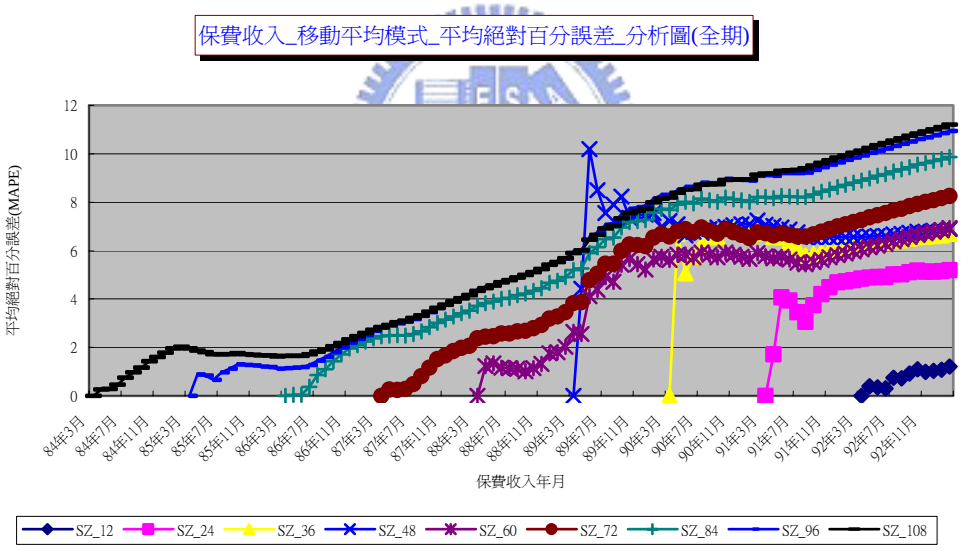


圖 5-1-6 保費收入_移動平均模式_分析圖(全期)

5.1.3 加權移動平均預測模式

程式名稱=macro_weight.sas

輸入數列=保費收入資料集 income_plot(保費年月,實際金額)

處理條件=觀測值全體或部分

Window size= SZ(12~108,increase by 12)

Weight = WT(0.1~1.0,increase by 0.1)

輸出數列=保費收入加權移動平均預測資料集

(保費年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-1-7、5-1-8）

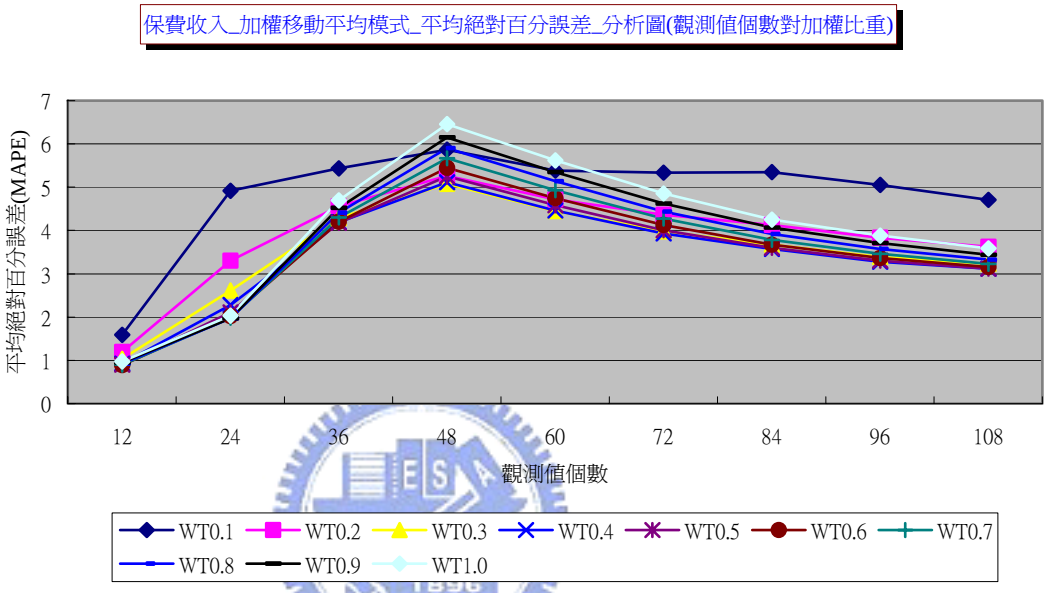


圖 5-1-7 保費收入_加權移動平均模式_分析圖(觀測值個數對加權比重)

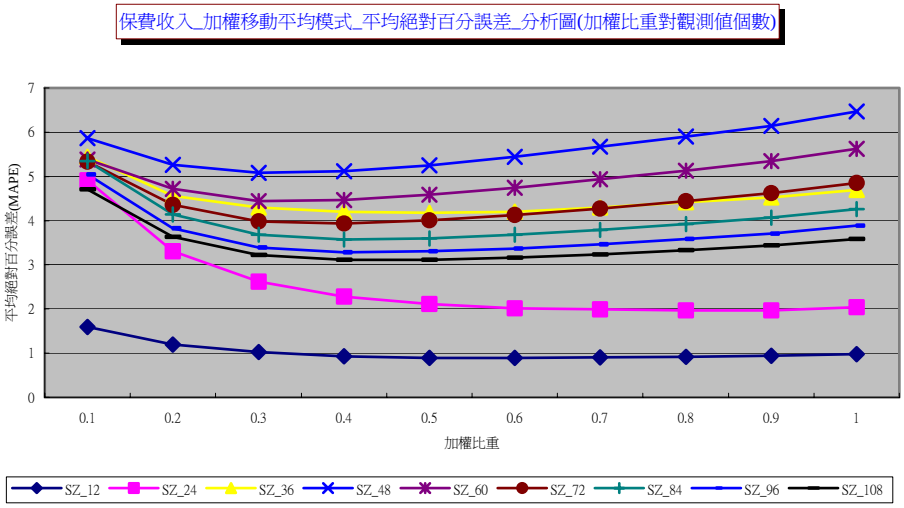


圖 5-1-8 保費收入_加權移動平均模式_分析圖(加權比重對觀測值個數)

5.1.4 迴歸分析模式

程式名稱=macro_autores.sas

輸入數列=保費收入資料集 income_plot(保費年月,實際金額)

處理條件=觀測值全體或部分

迴歸項選擇=

YX = 保費收入(Y)與保費年月(X)

YY^ = 保費收入(Y)與前期保費收入(Y^)

YX_顯著相關項=保費收入(Y)與保費年月(X) 顯著相關項

相關項個數= NLAG(2~24,increase by 2)

觀測項個數= SZ(12~108,increase by 12)

輸出數列=保費收入迴歸分析預測資料集

(保費年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖(請參考:圖 5-1-9、5-1-10、5-1-11、5-1-12、5-1-13、5-1-14)

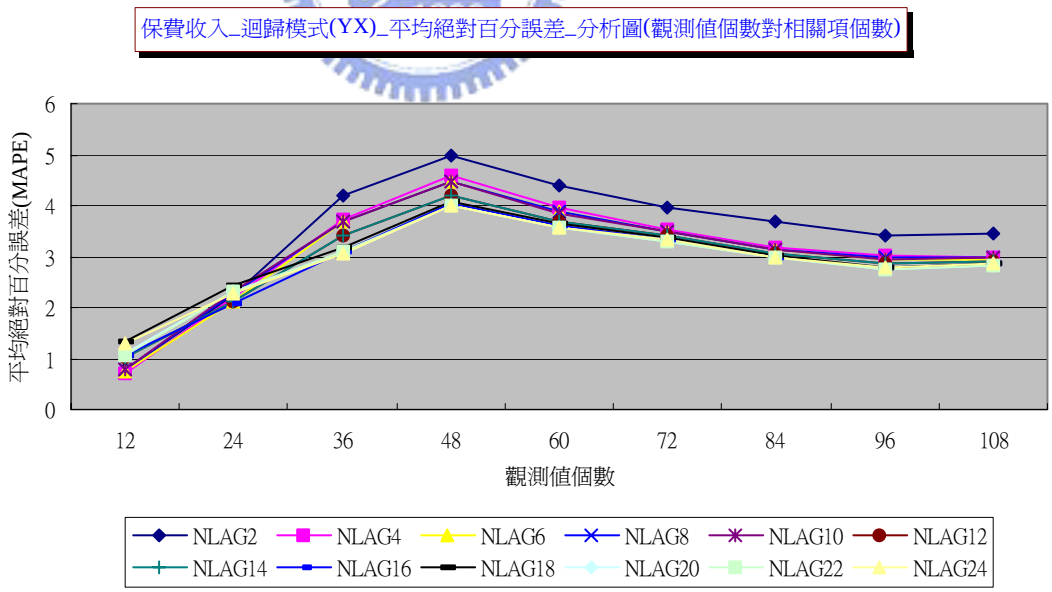


圖 5-1-9 保費收入_迴歸模式(YX)_分析圖(觀測值個數對相關項個數)

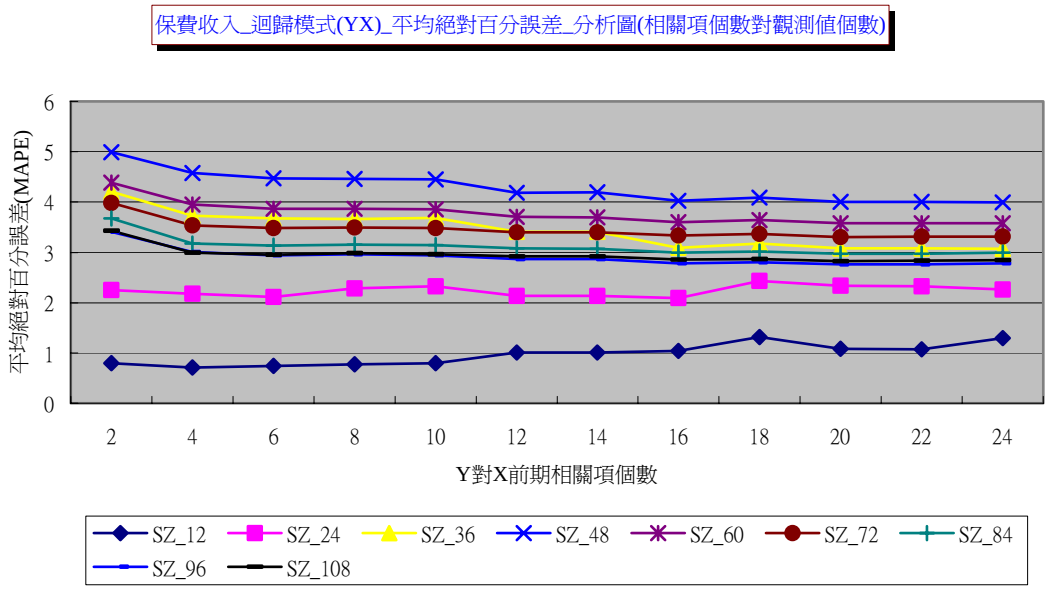


圖 5-1-10　保費收入_迴歸模式(YX)_分析圖(相關項個數對觀測值個數)

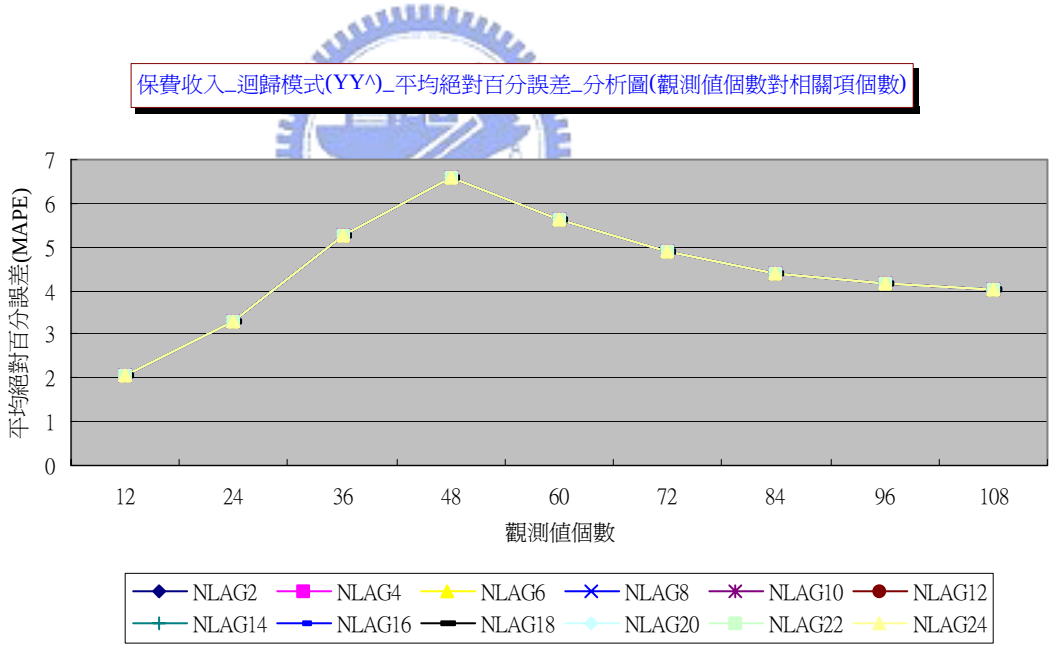


圖 5-1-11　保費收入_迴歸模式(YY^)_分析圖(觀測值個數對相關項個數)

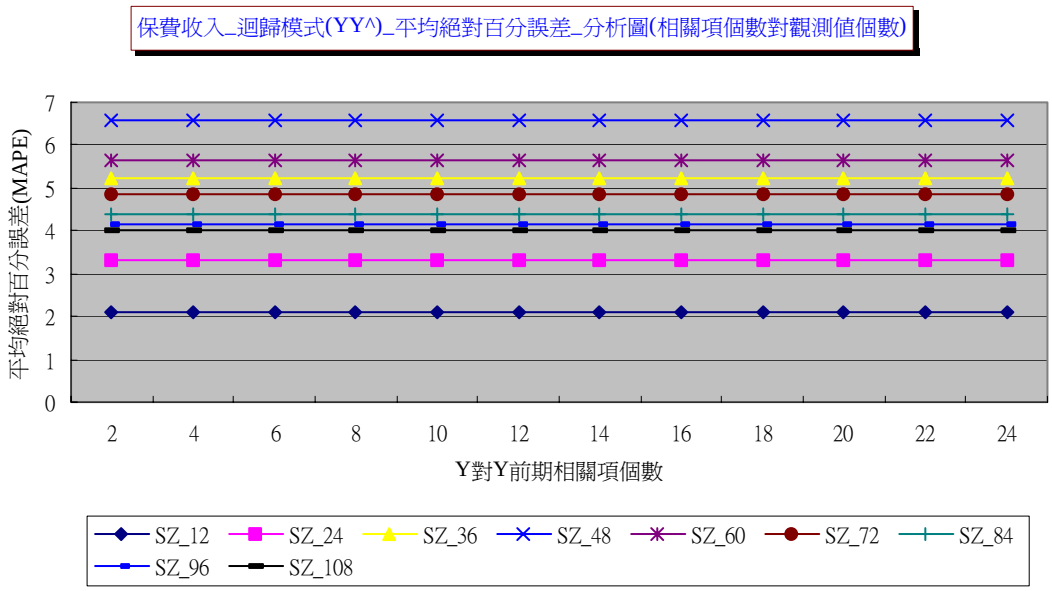


圖 5-1-12　保費收入_迴歸模式(YY^)_分析圖(相關項個數對觀測值個數)

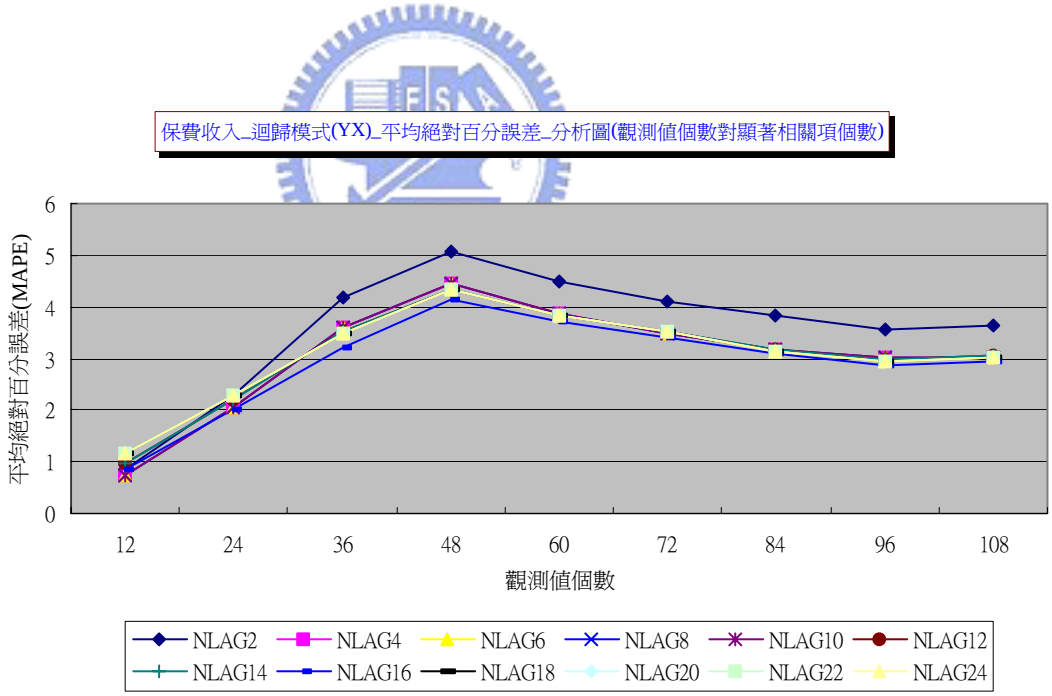


圖 5-1-13　保費收入_迴歸模式(YX)_分析圖(觀測值個數對顯著相關項個數)

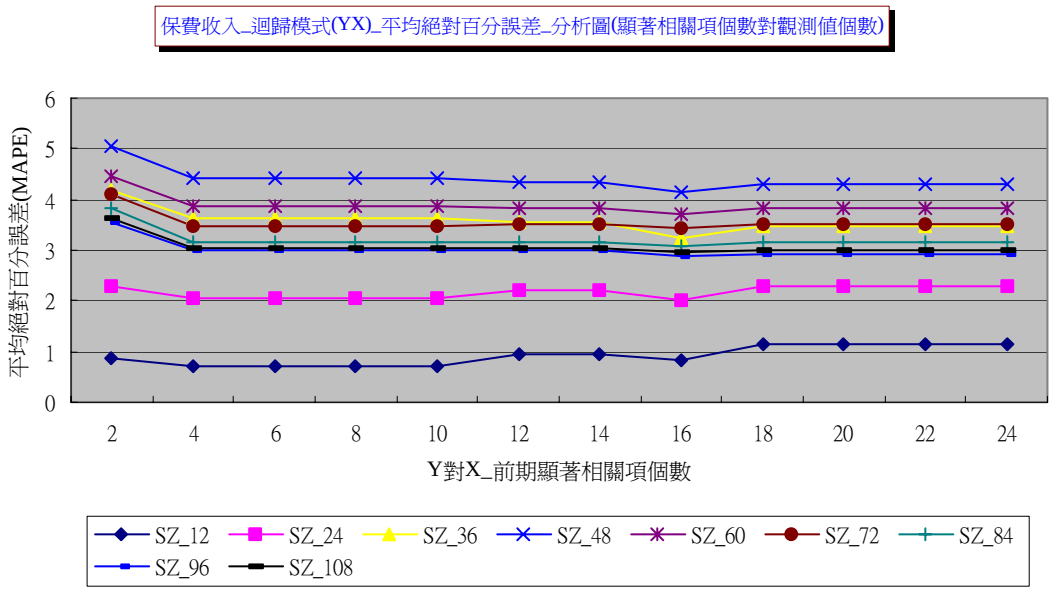


圖 5-1-14 保費收入_迴歸模式(YX)_分析圖(顯著相關項個數對觀測值個數)

5.1.5 自我迴歸整合移動平均模式

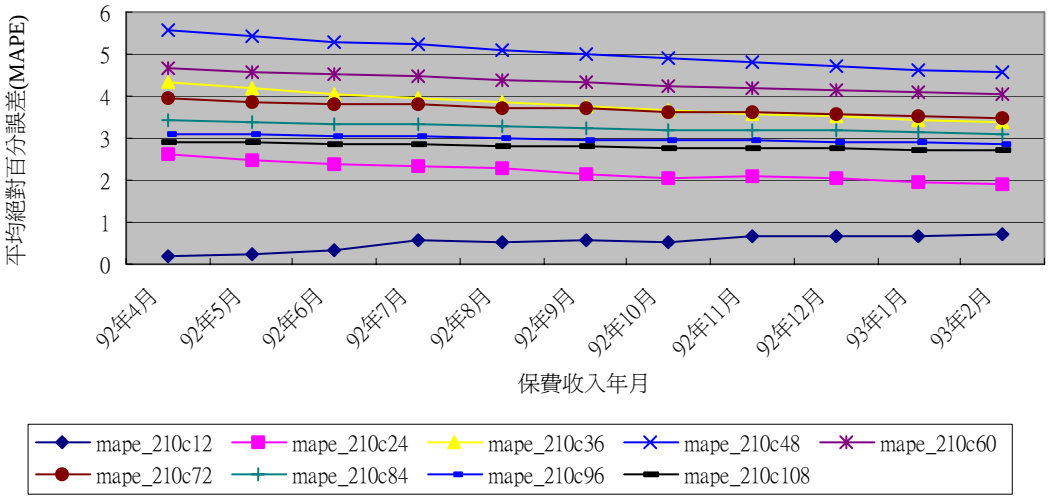
程式名稱=macro_arima_count.sas，get-arima-select.sas
輸入數列=保費收入資料集 income_plot(保費年月,實際金額)
處理條件=觀測值全體或部分
p d q 階次選擇=p(0~3),d(1,2),q(0~3)
觀測值個數= c (12~108,increase by 12)

使用 DO Loop 迴圈 (p * d * q * c),共產生 288 種組合情形可供判斷分析其預測準確度,因組合情形過多,無法以肉眼比較,故撰寫程式以觀測值個數的角度來分析挑選最適之階次,並在最後得到一建議的 p d q 階次組合。

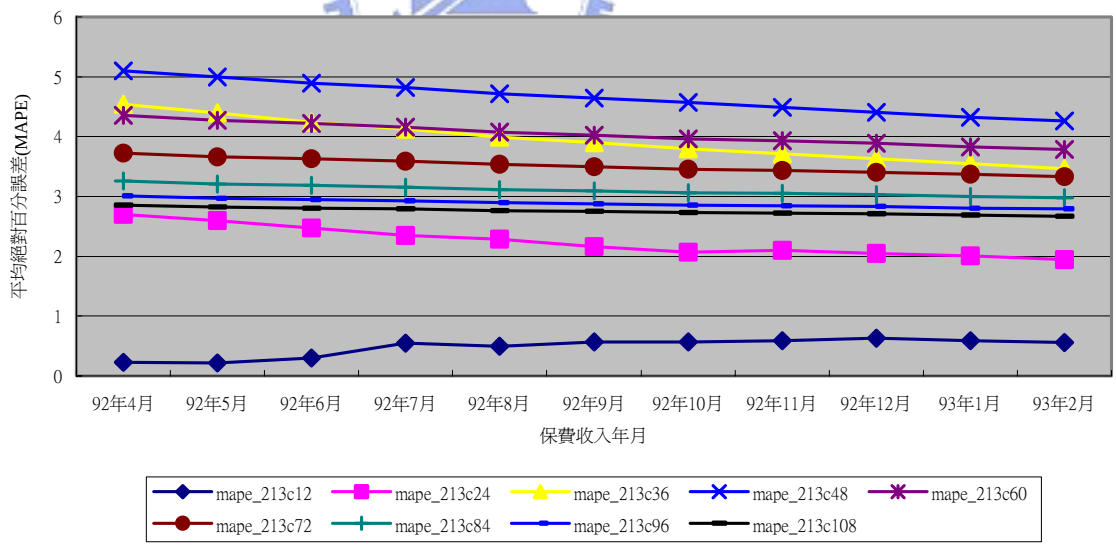
輸出數列=保費收入自我迴歸整合移動平均預測資料集
(保費年月,實際金額,預測金額,預測誤差各項相關統計值)

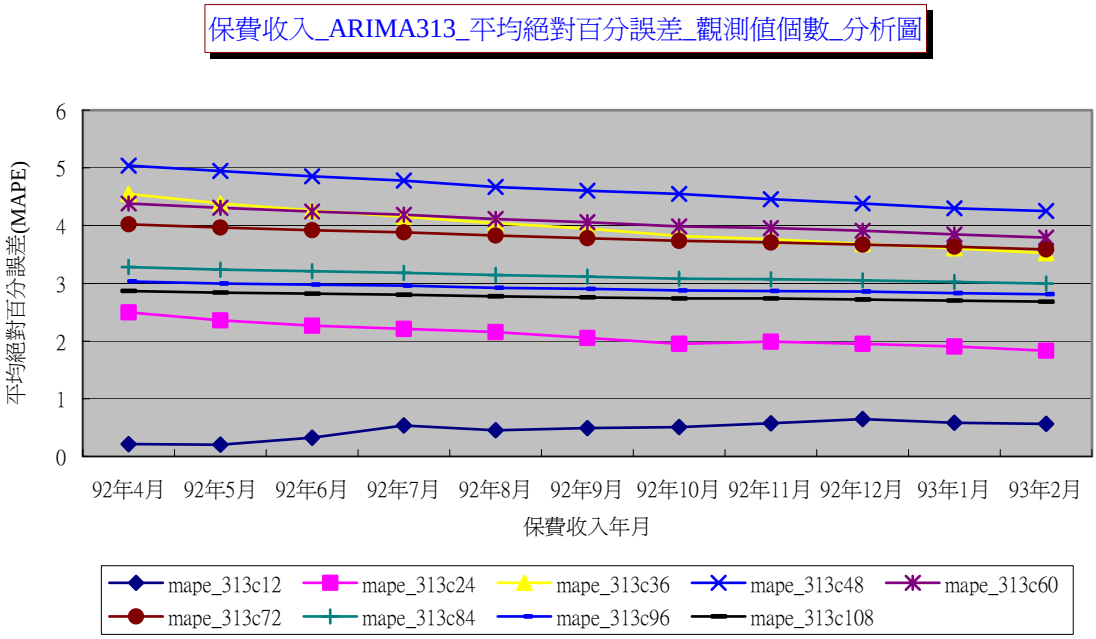
各項統計分析圖（請參考：圖 5-1-15）

保費收入_ARIMA210_平均絕對百分誤差_觀測值個數_分析圖



保費收入_ARIMA213_平均絕對百分誤差_觀測值個數_分析圖





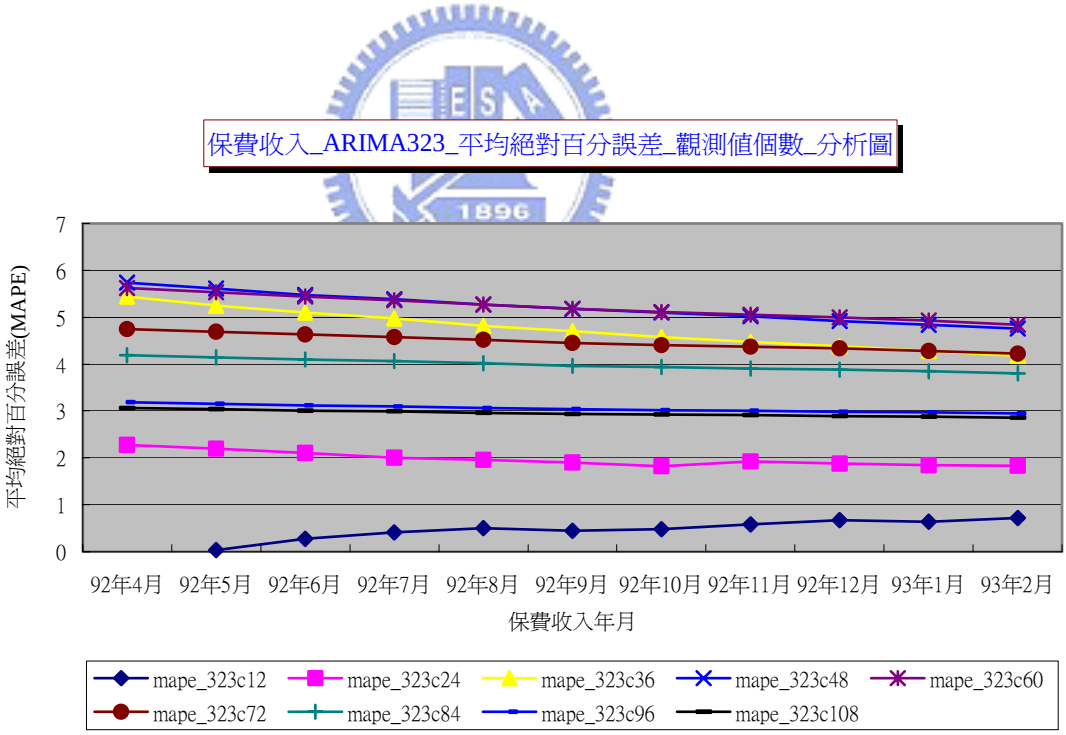


圖 5-1-15　保費收入_自我迴歸整合移動平均模式_分析圖(ARIMA)

5.1.6 季節性自我迴歸整合移動平均模式

程式名稱=macro_arima_cx_season.sas，get-arima-select.sas

輸入數列=保費收入資料集 income_plot(保費年月,實際金額)

處理條件=觀測值全體或部分

pdq 階次選擇=p(0~3),d(1,2),q(0~3)

PDQ 階次選擇=P(3,4,12),D(1,2),Q(3,4,12)

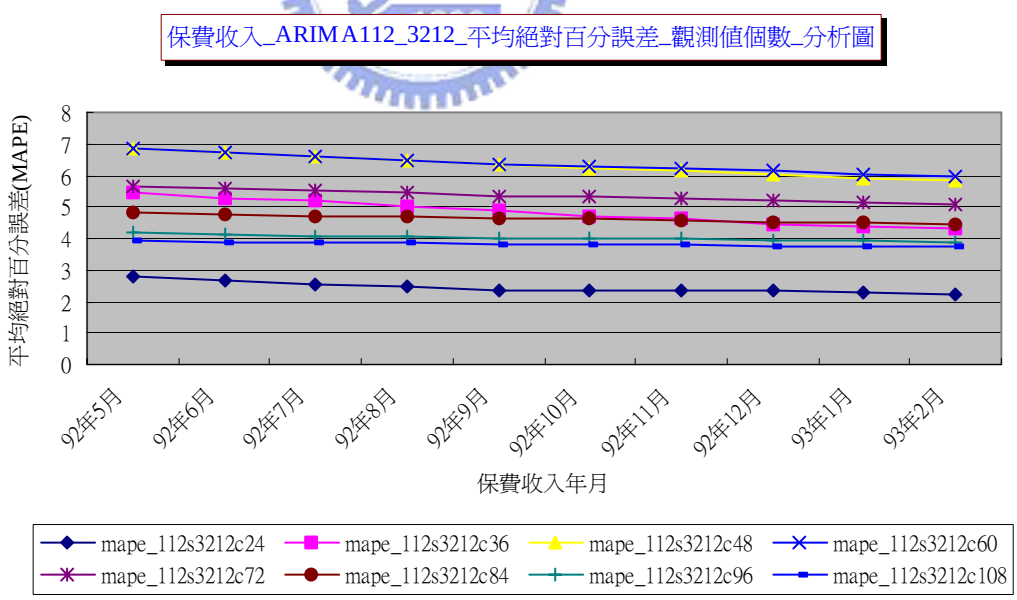
觀測值個數= C (12~108,increase by 12)

使用 DO Loop 迴圈 (p * d * q * P * D * Q * C),共產生 5184 種組合情形可供判斷分析其預測準確度,因組合情形相當龐大,無法以肉眼比較,故撰寫程式以觀測值個數的角度來分析挑選最適之階次,並最後得到一建議的 p d q 及 P D Q 階次組合。

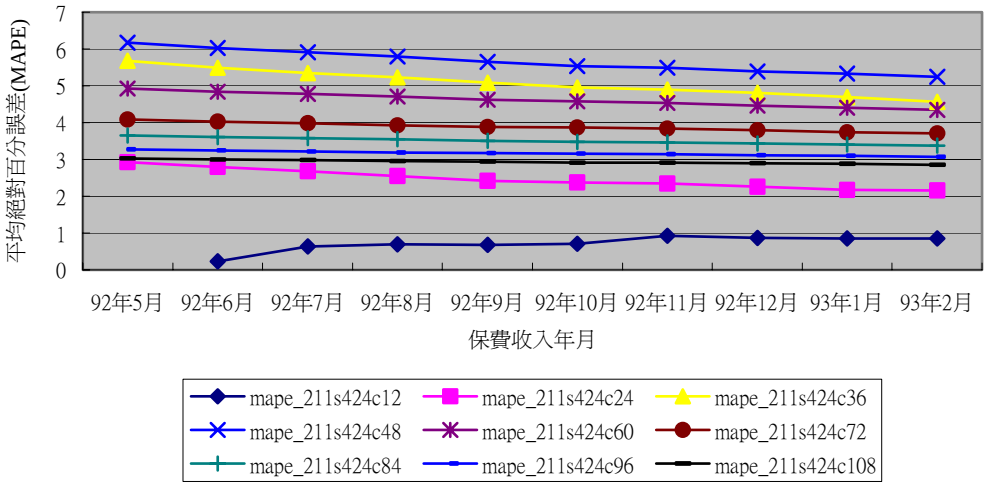
輸出數列=保費收入季節性自我迴歸整合移動平均預測資料集

(保費年月,實際金額,預測金額,預測誤差各項相關統計值)

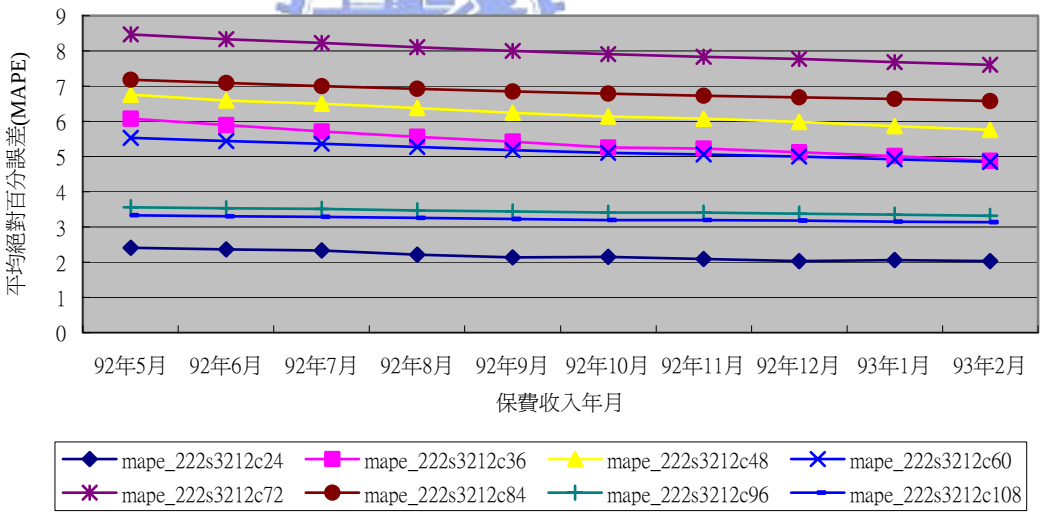
各項統計分析圖（請參考：圖 5-1-16）



保費收入_ARIMA211_424_平均絕對百分誤差_觀測值個數_分析圖



保費收入_ARIMA222_3212_平均絕對百分誤差_觀測值個數_分析圖



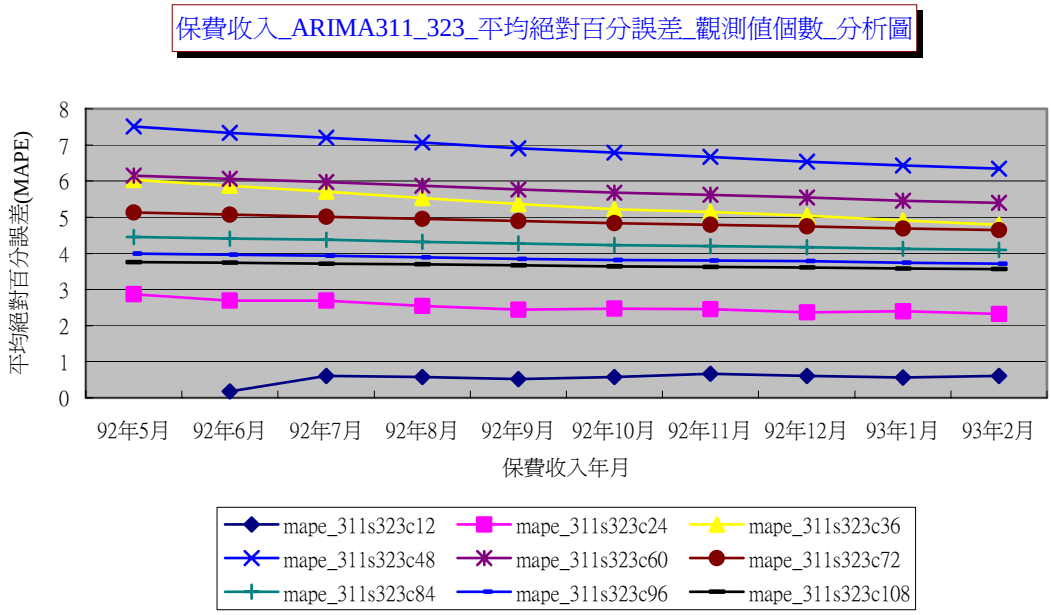


圖 5-1-16　保費收入_自我迴歸整合移動平均模式_分析圖(SEASON ARIMA)

5.2　保費收入六種預測模式綜合比較

程式名稱=MACRO-COMPARE-6.SAS
輸入項目=六項預測模式產生的輸出數列資料集中最適合的階次組合
運算條件=分別針對六項分析模式給特定的參數組合，如 window size，weight，迴歸項選擇，pdq，PDQ 階次，觀測值個數 C
輸出數列=六種保費收入預測模式最佳項組合資料集(保費年月,實際金額,六種預測金額,六種預測誤差各項相關統計值(依觀測值個數分類))、六種模式平均絕對百分誤差_全期數列資料集

各項統計分析圖（請參考：圖 5-2-1、5-2-2、5-2-3、5-2-4、5-2-5、5-5-6、5-2-7、5-2-8、5-2-9、5-2-10、5-2-11、5-2-12、5-2-13、5-2-14）

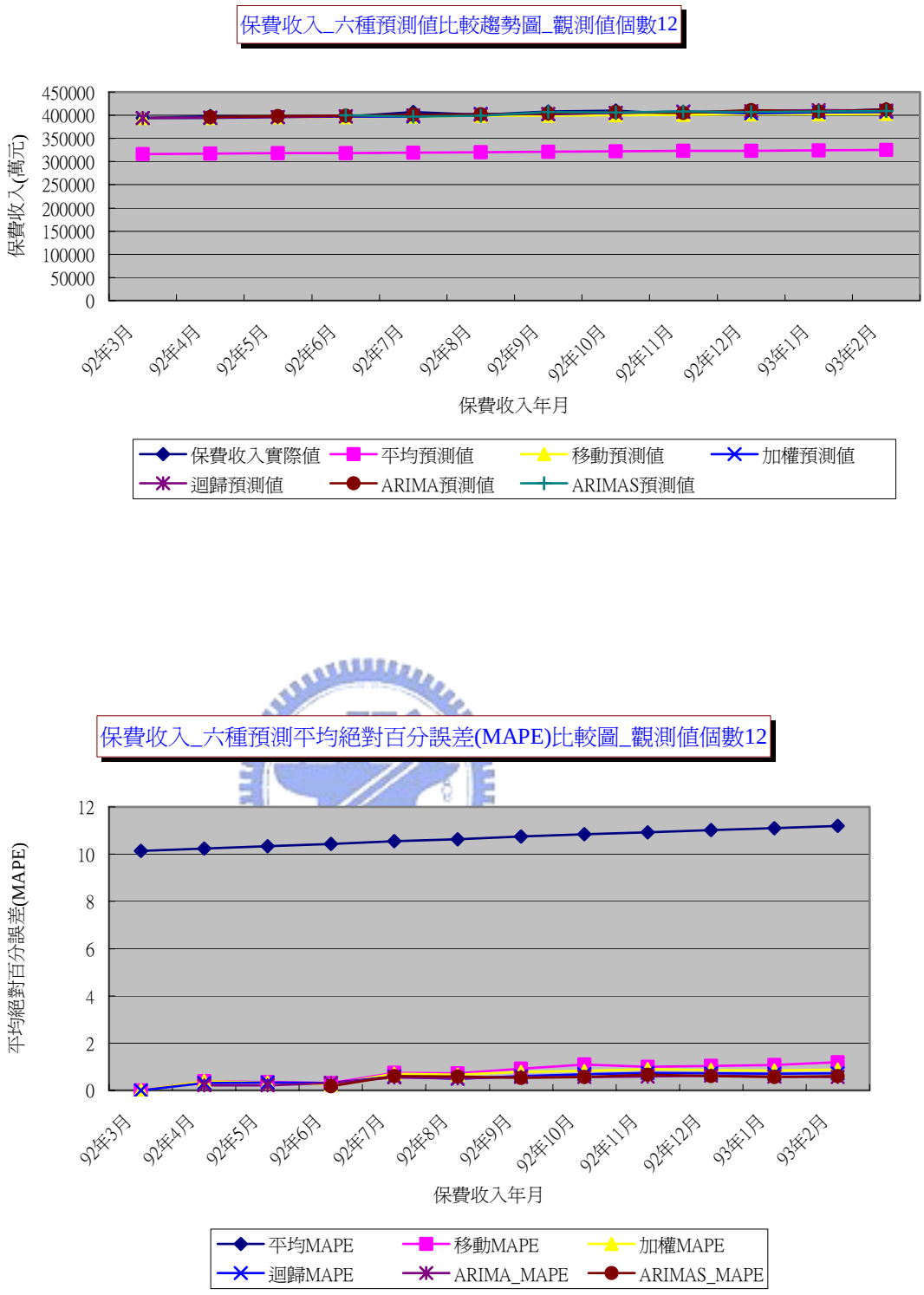


圖 5-2-1　保費收入_六種預測模式比較圖(觀測值個數 12)

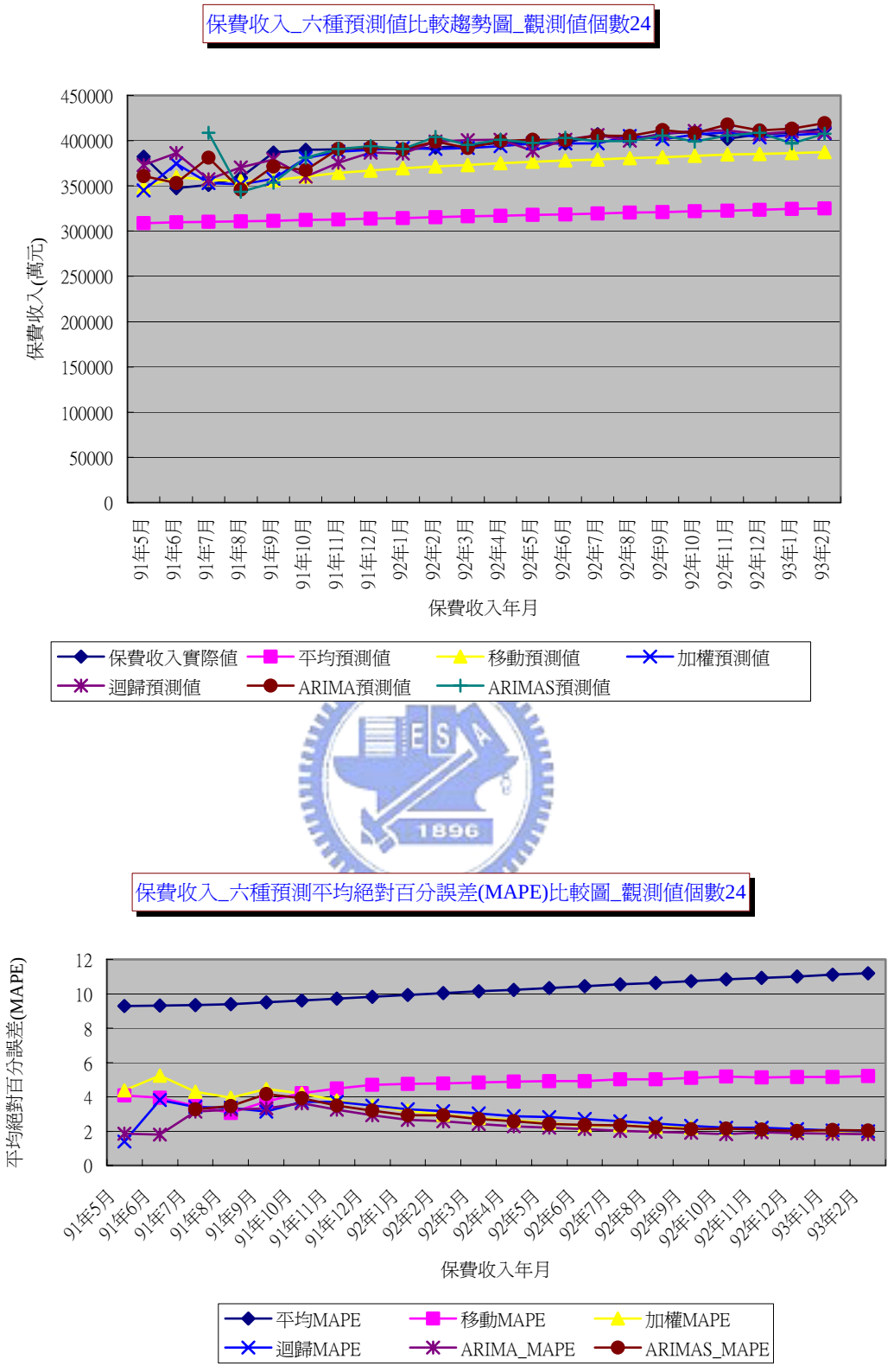


圖 5-2-2　保費收入_六種預測模式比較圖(觀測值個數 24)

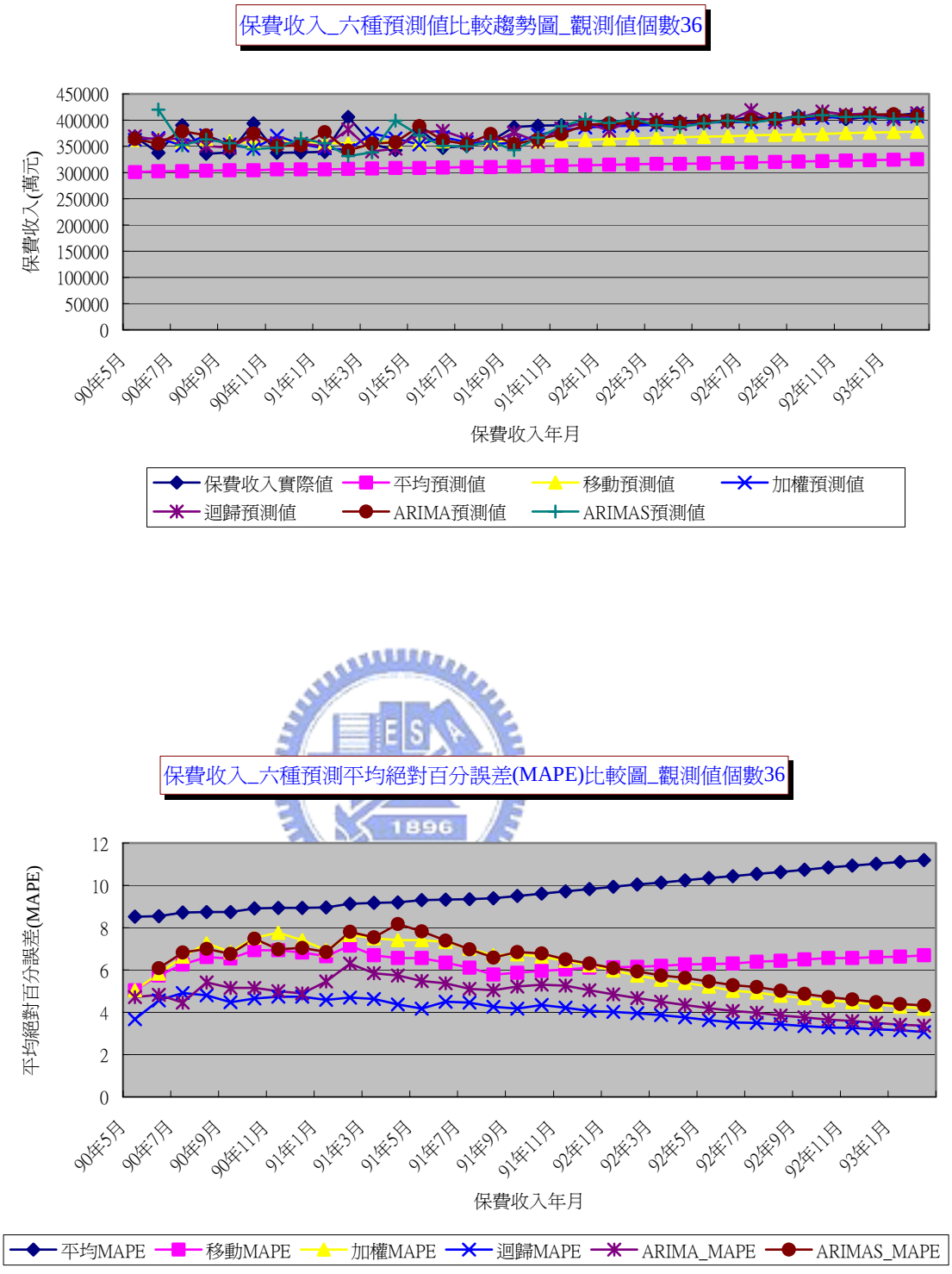


圖 5-2-3　保費收入_六種預測模式比較圖(觀測值個數 36)

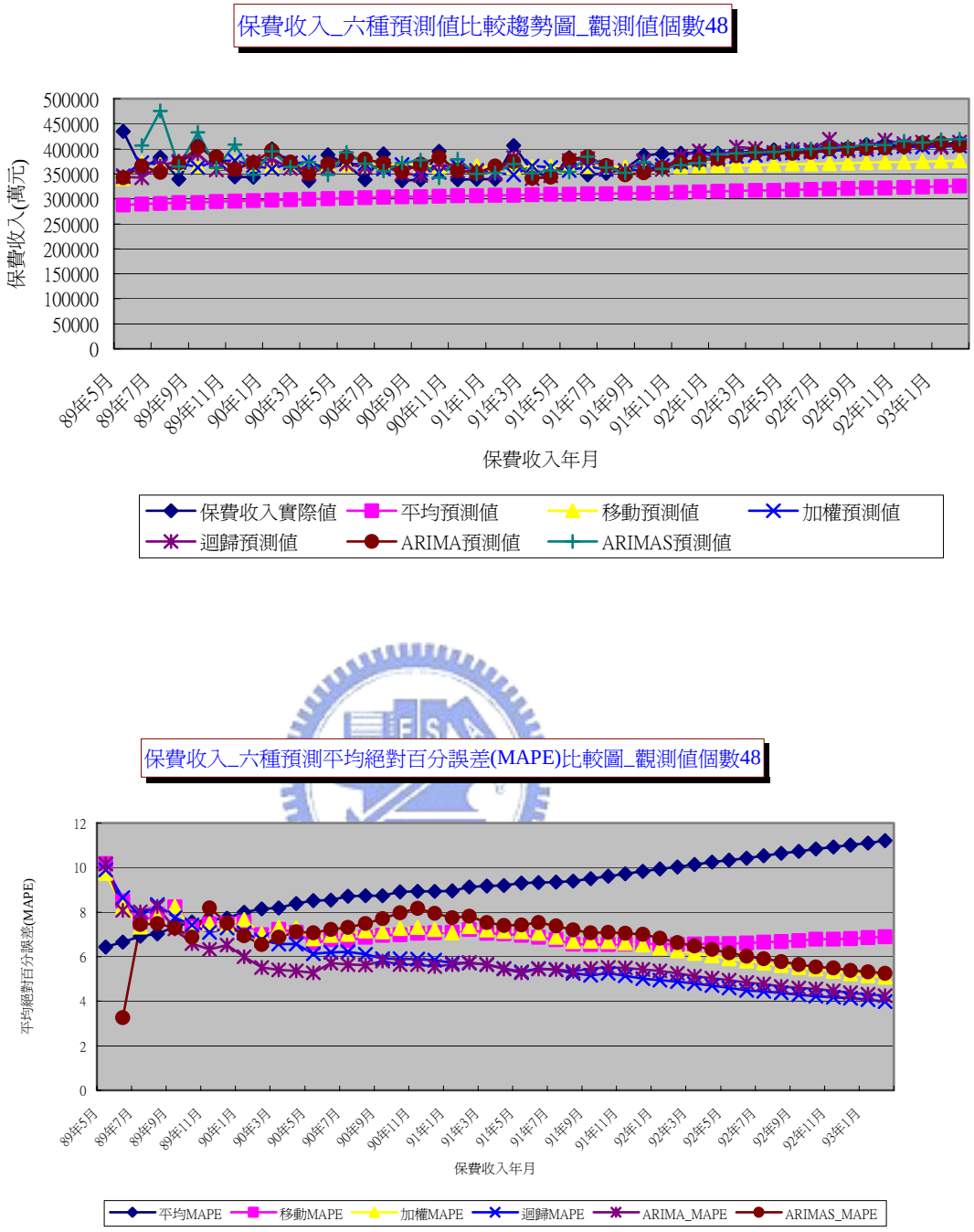
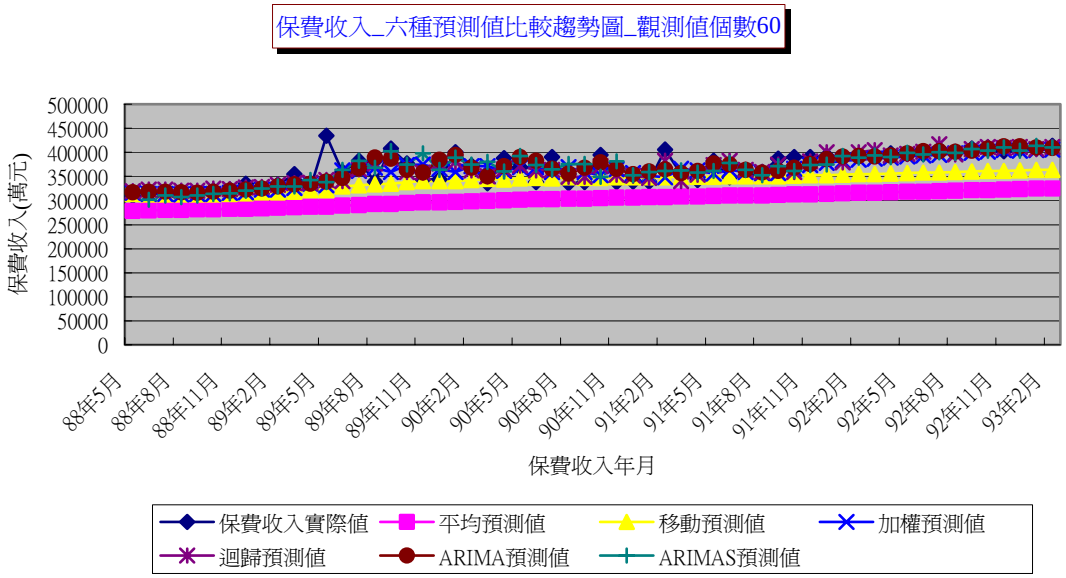


圖 5-2-4　保費收入_六種預測模式比較圖(觀測值個數 48)



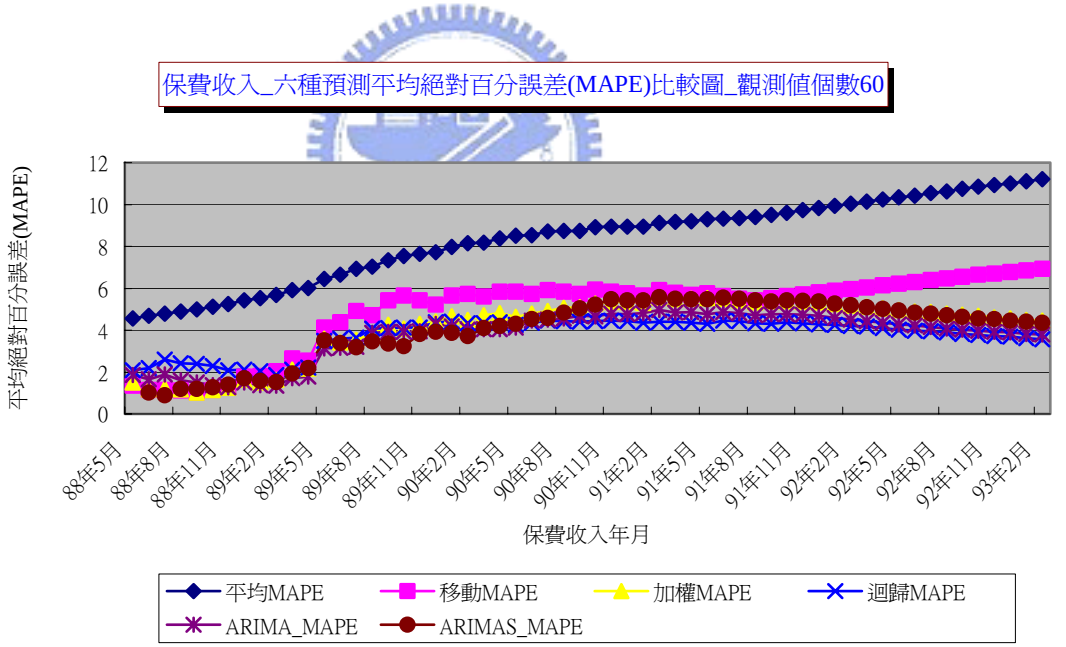
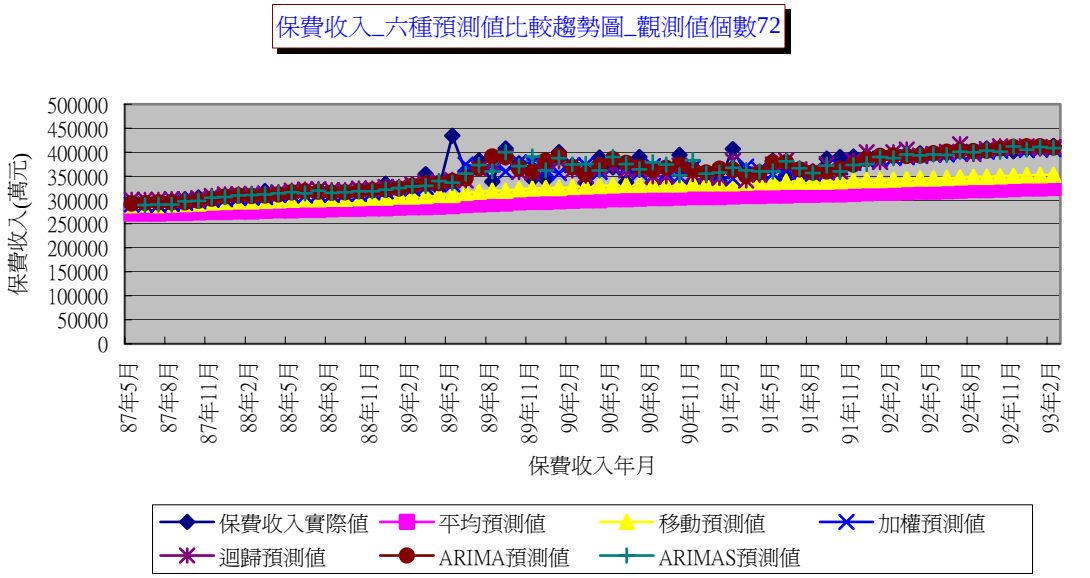


圖 5-2-5　保費收入_六種預測模式比較圖(觀測值個數 60)



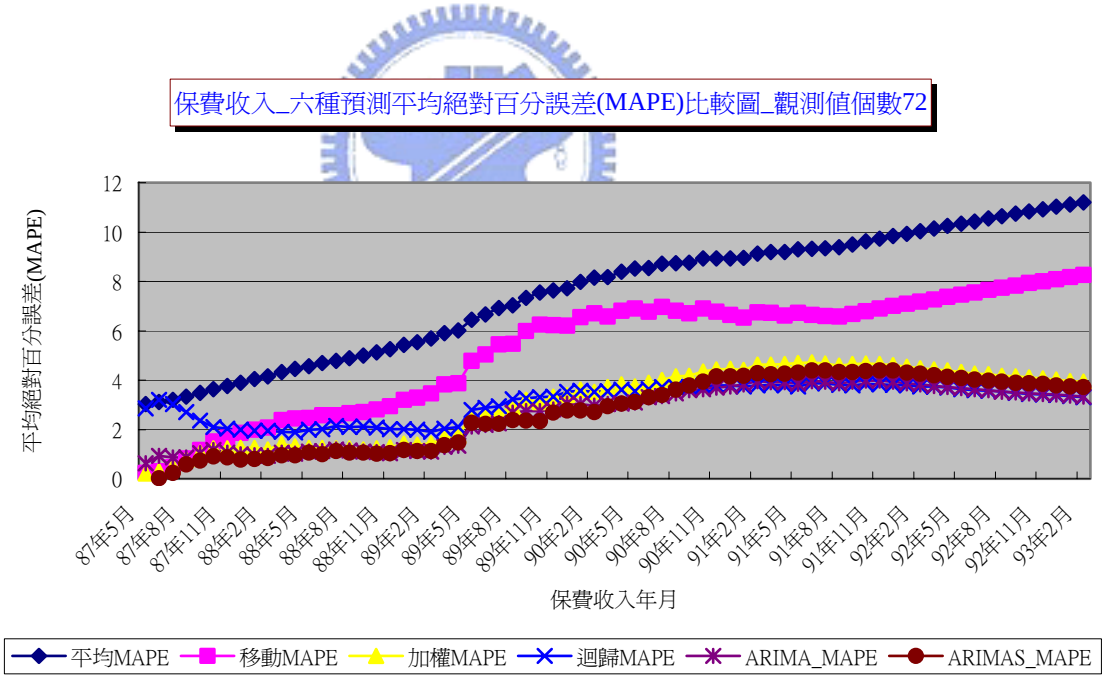
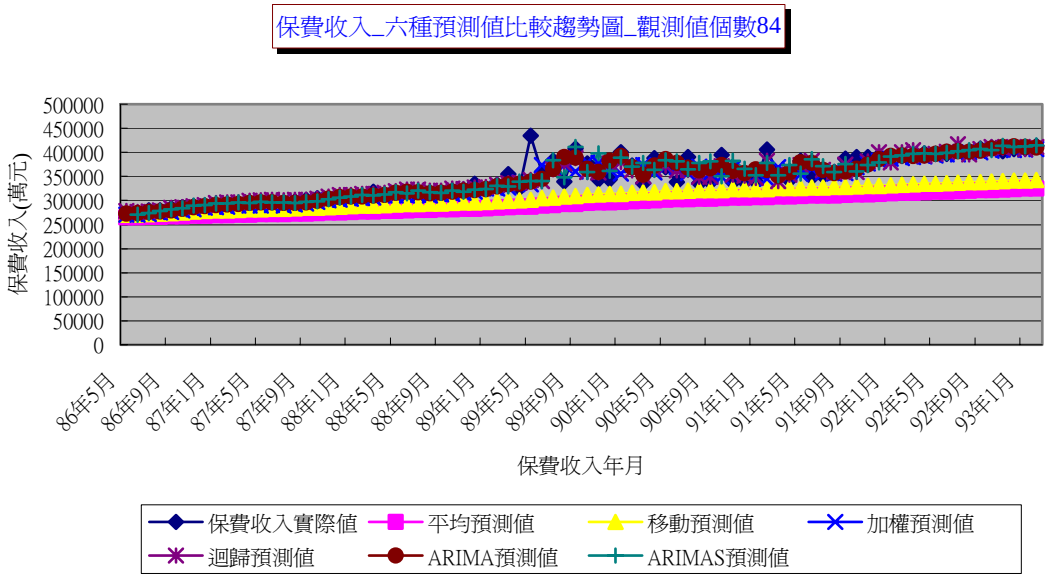


圖 5-2-6　保費收入_六種預測模式比較圖(觀測值個數 72)



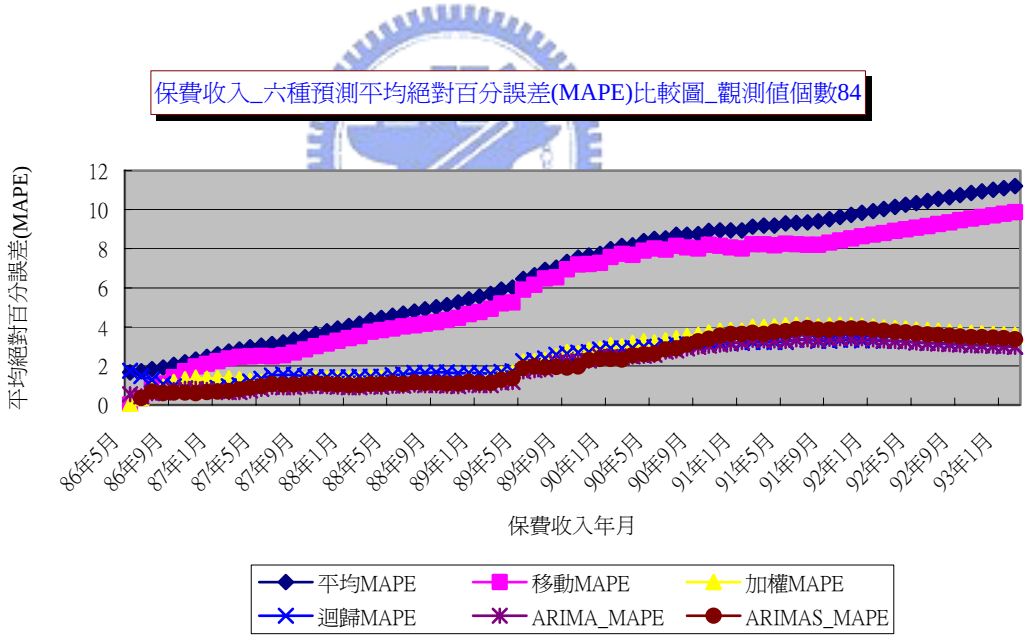
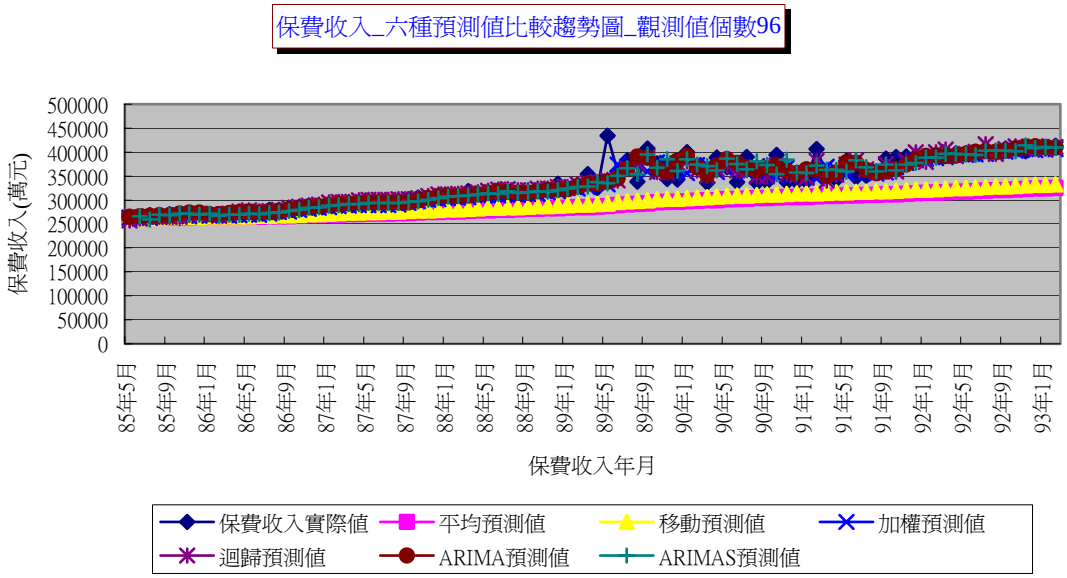


圖 5-2-7　保費收入_六種預測模式比較圖(觀測值個數 84)



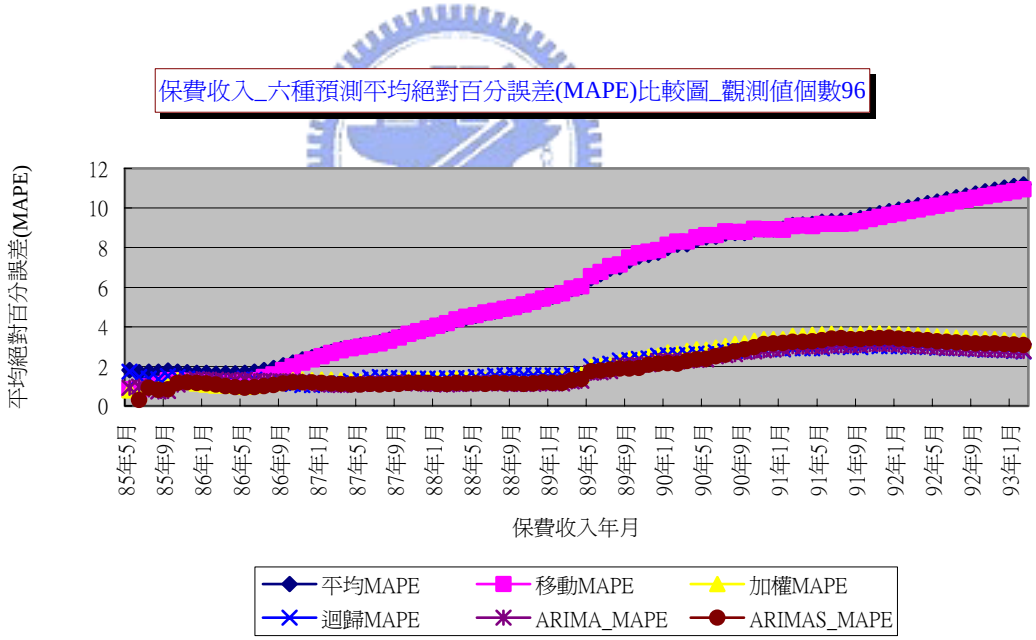
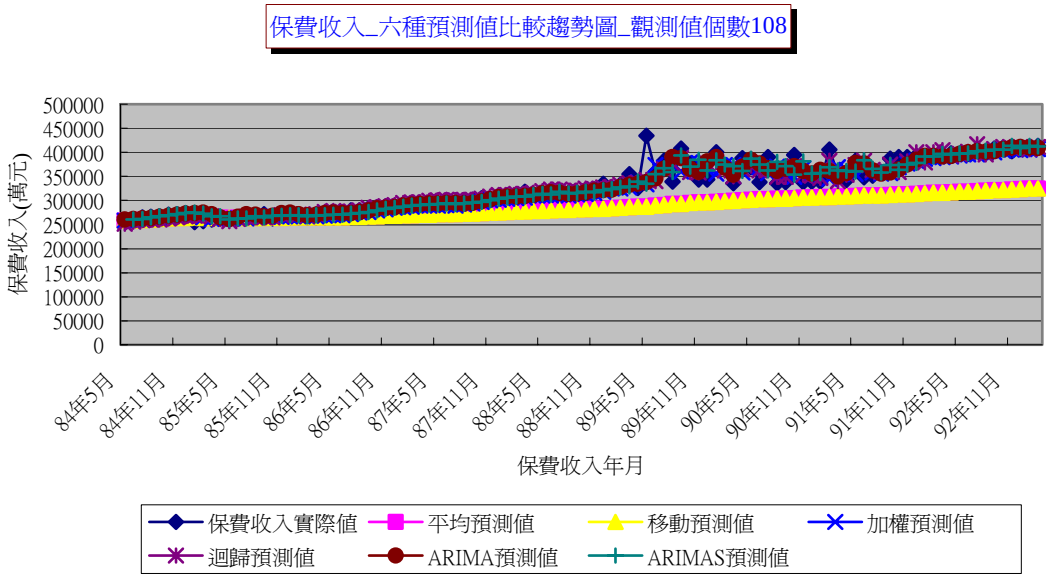


圖 5-2-8　保費收入_六種預測模式比較圖(觀測值個數 96)



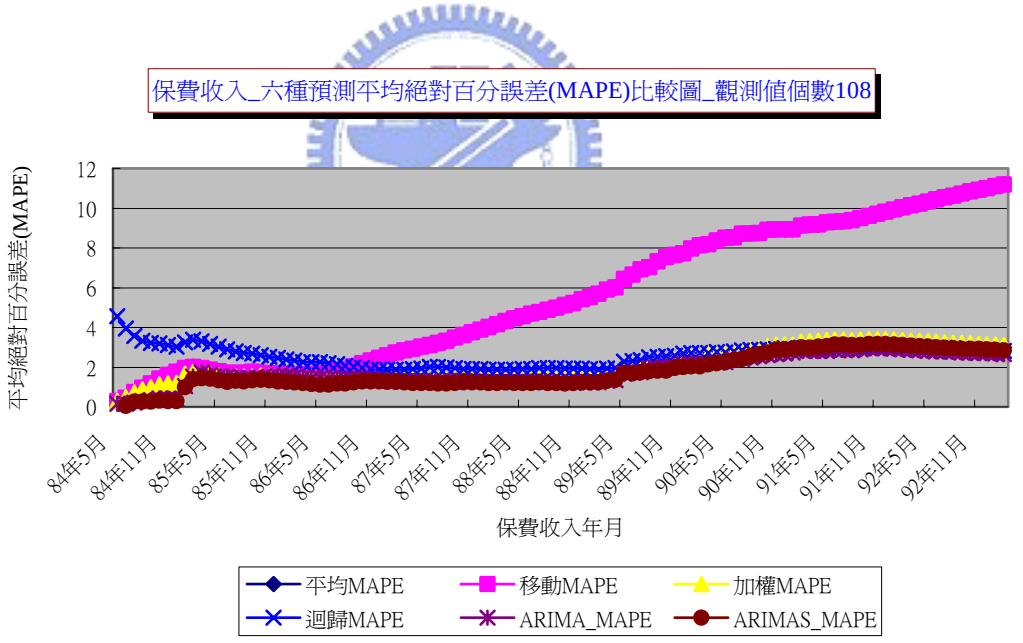


圖 5-2-9　保費收入_六種預測模式比較圖(觀測值個數 108)

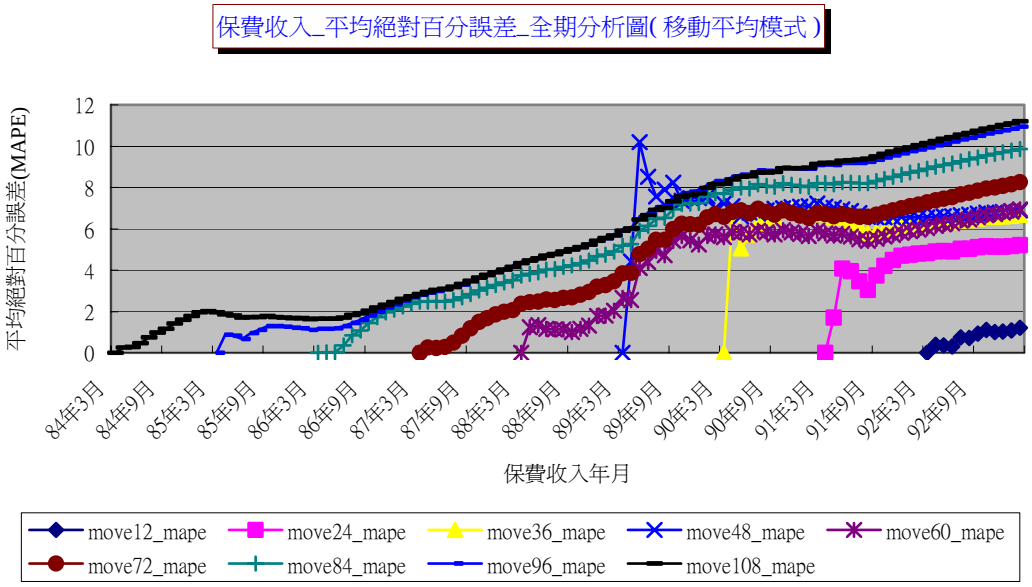


圖 5-2-10　保費收入_全期分析圖(移動平均模式)

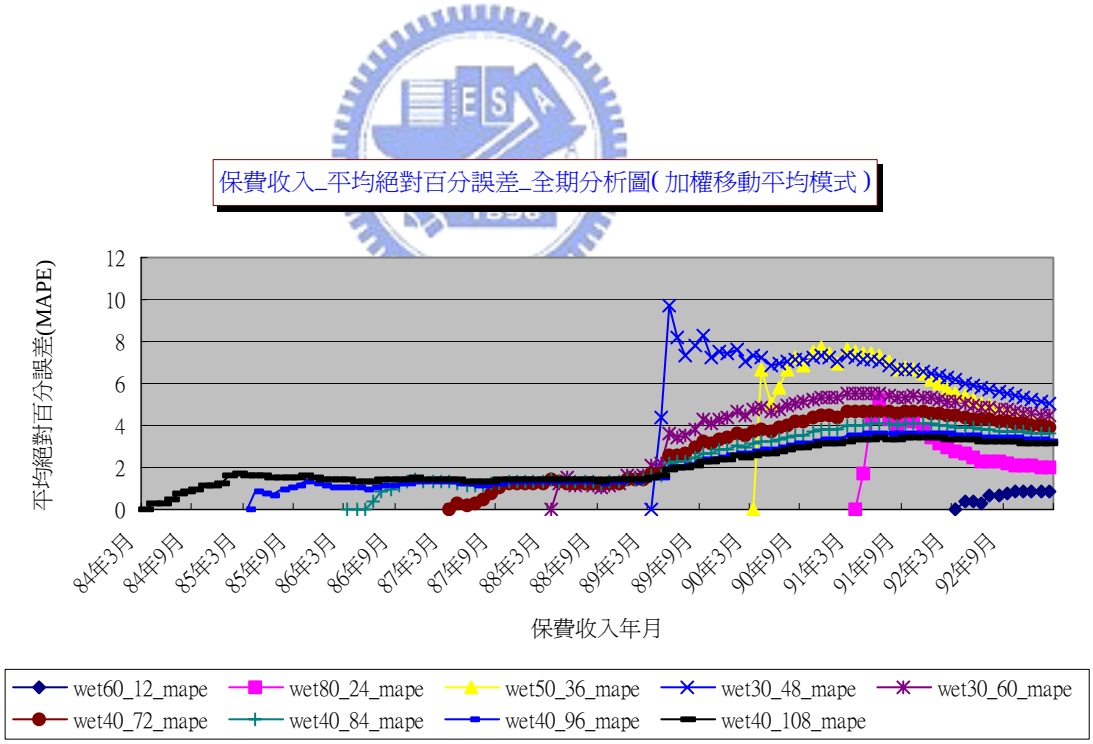


圖 5-2-11　保費收入_全期分析圖(加權移動平均模式)

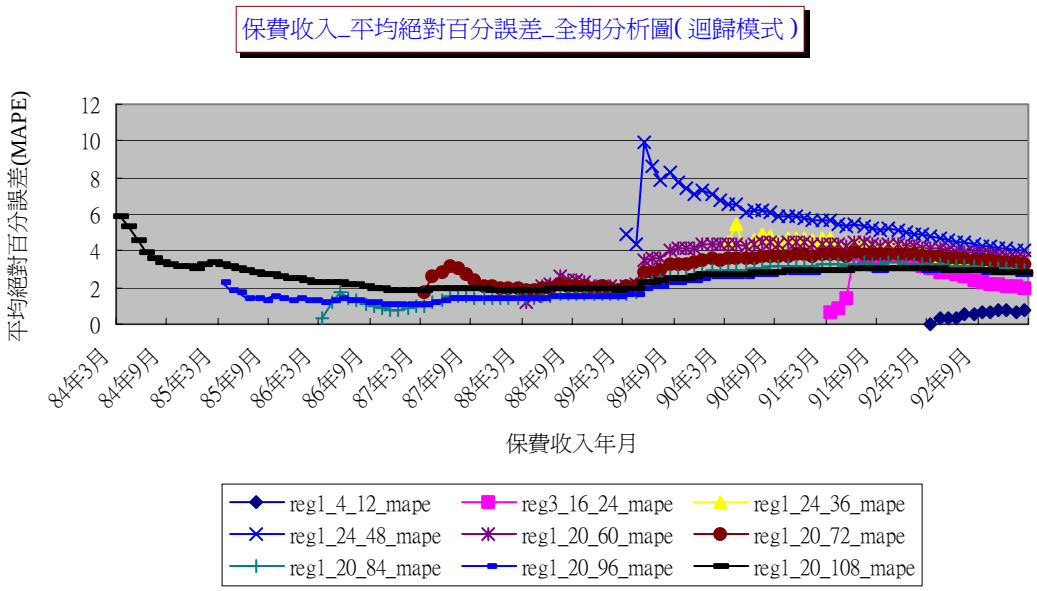


圖 5-2-12　保費收入_全期分析圖(迴歸模式)

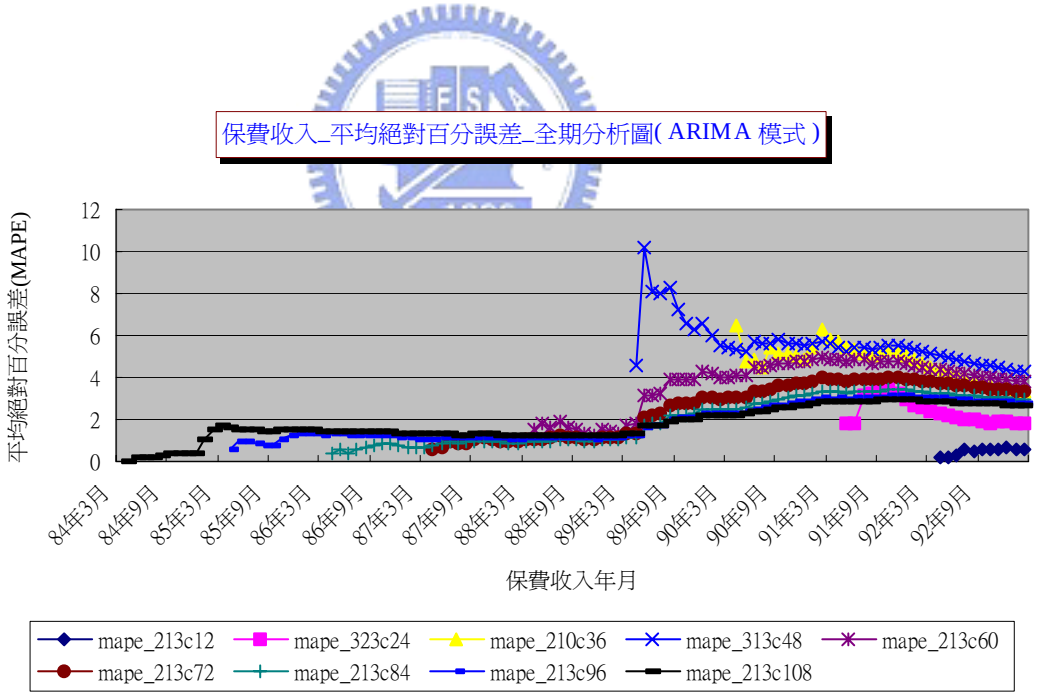


圖 5-2-13　保費收入_全期分析圖(ARIMA 模式)

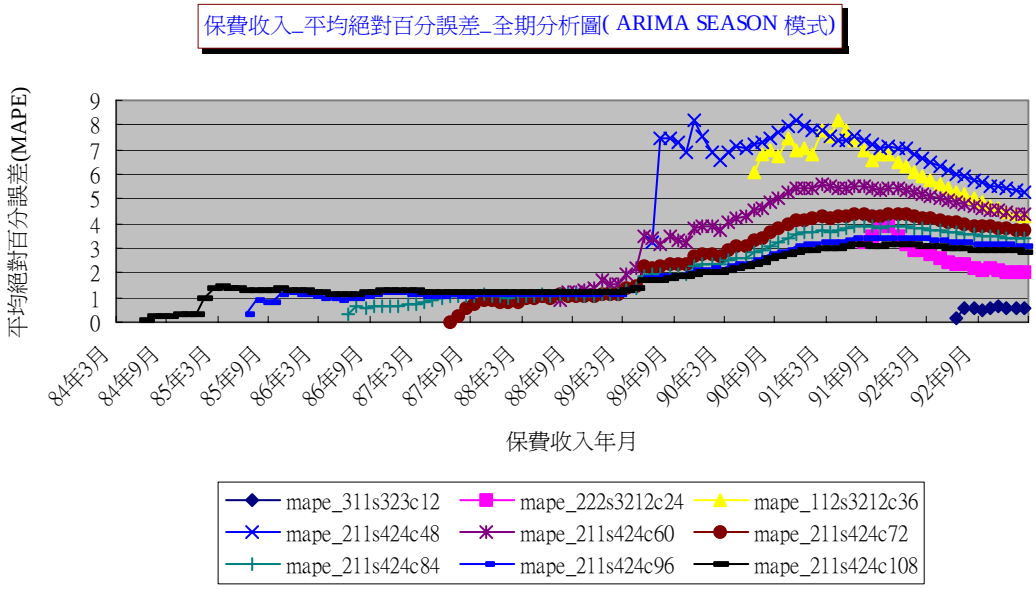


圖 5-2-14　保費收入_全期分析圖(SEASON ARIMA 模式)

5.3　費用支出預測

本節中共使用六種預測模式進行分析（平均預測、移動平均預測、加權移動平均預測、迴歸分析、自我迴歸整合移動平均、季節性自我迴歸整合移動平均），茲分述如后：

5.3.1 平均預測模式

程式名稱=exp-average.sas
輸入數列=費用支出資料集 exp_plot(費用年月,實際金額)
處理條件=觀測值全體
輸出數列=費用支出平均預測資料集
(費用年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-3-1、5-3-2、5-3-3、5-3-4）

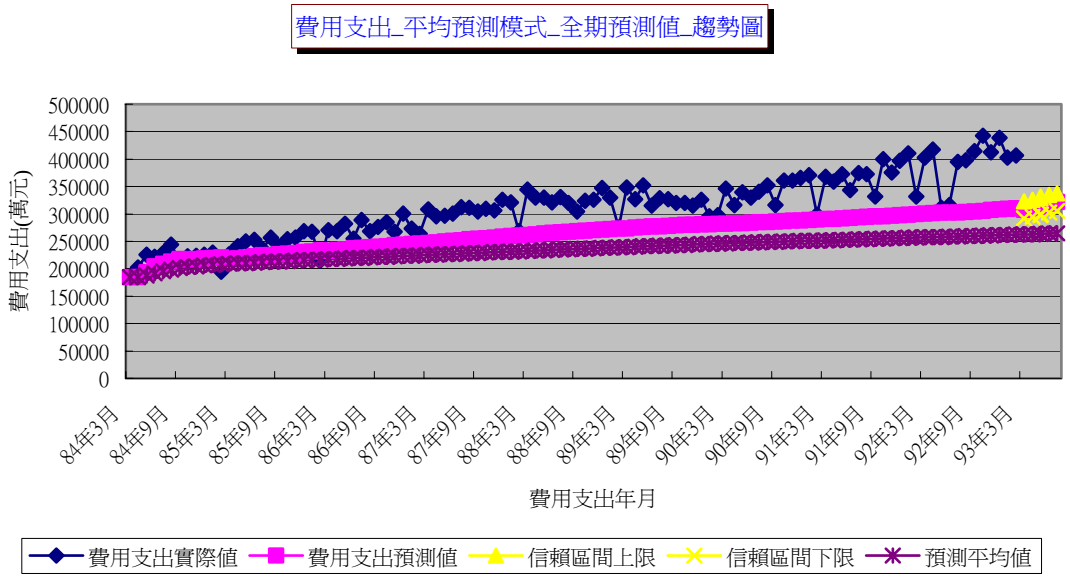


圖 5-3-1　費用支出_平均預測模式_全期預測值_趨勢圖

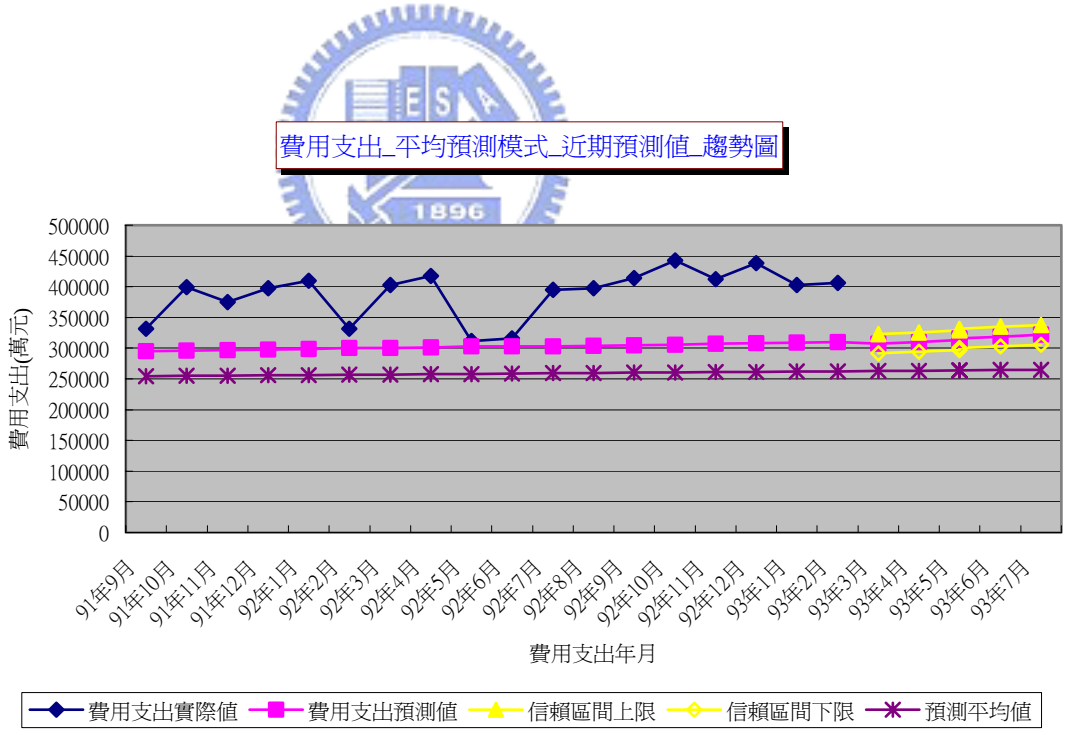


圖 5-3-2　費用支出_平均預測模式_近期預測值_趨勢圖

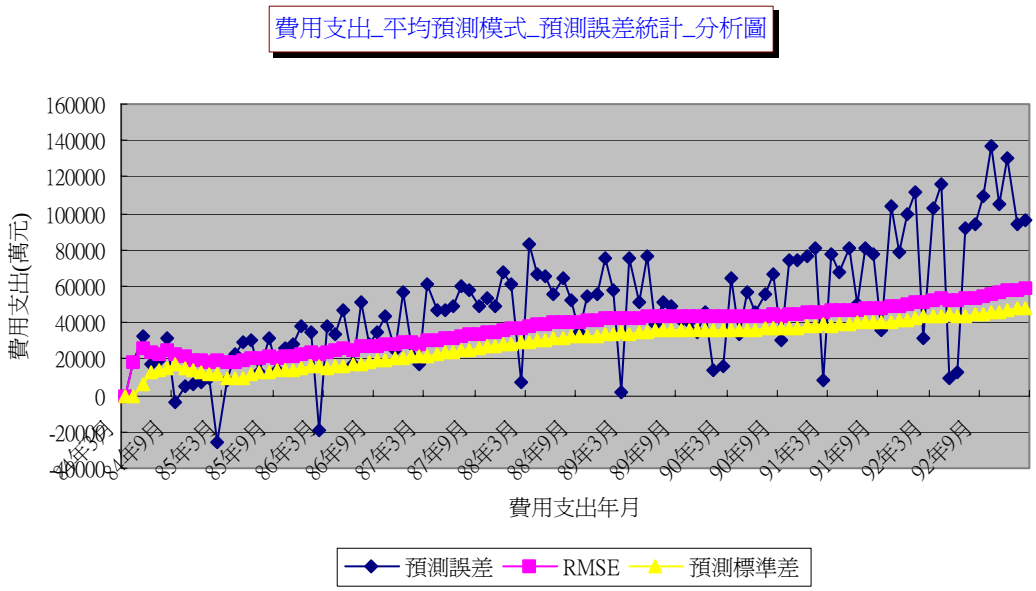


圖 5-3-3　費用支出_平均預測模式_預測誤差統計_分析圖

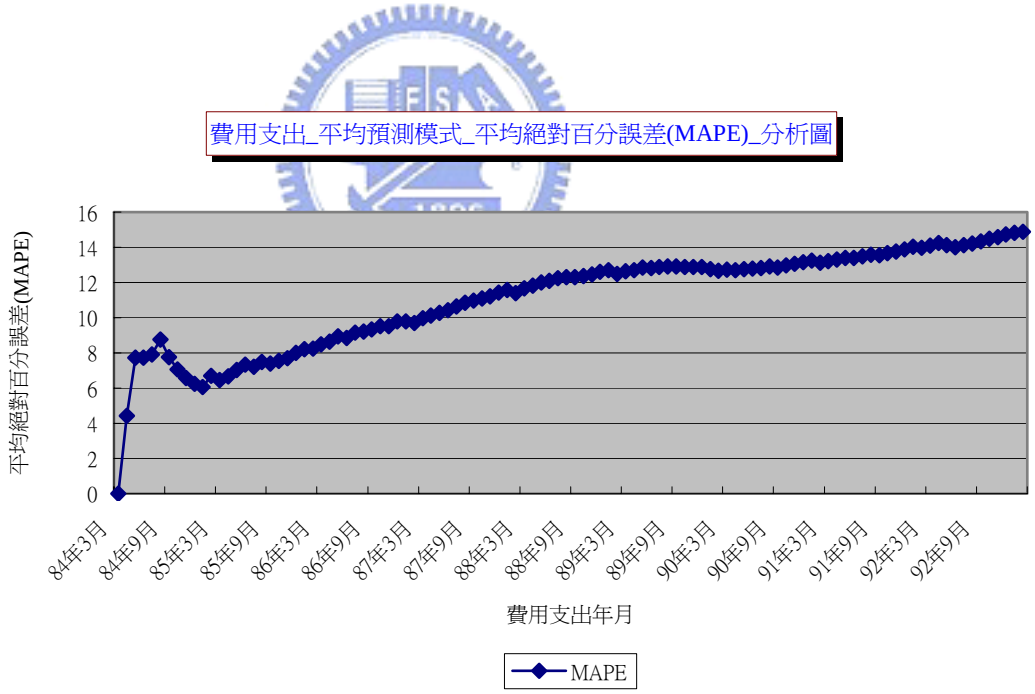


圖 5-3-4　費用支出_平均預測模式_平均絕對百分誤差(MAPE)_分析圖

5.3.2 移動平均預測模式

程式名稱=exp-macro_moving.sas
輸入數列=費用支出資料集 exp_plot(費用年月,實際金額)
處理條件=觀測值全體或部分
Window size= SZ(12~108,increase by 12)輸出數列=費用支出移動平均預測資料集
(費用年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-3-5、5-3-6）

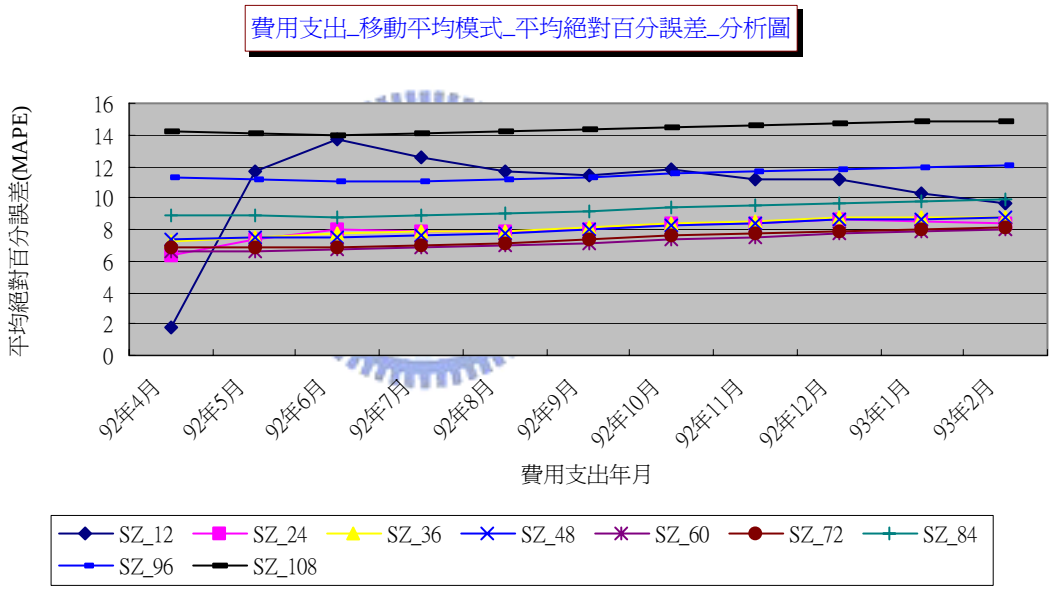


圖 5-3-5　費用支出_移動平均模式_分析圖

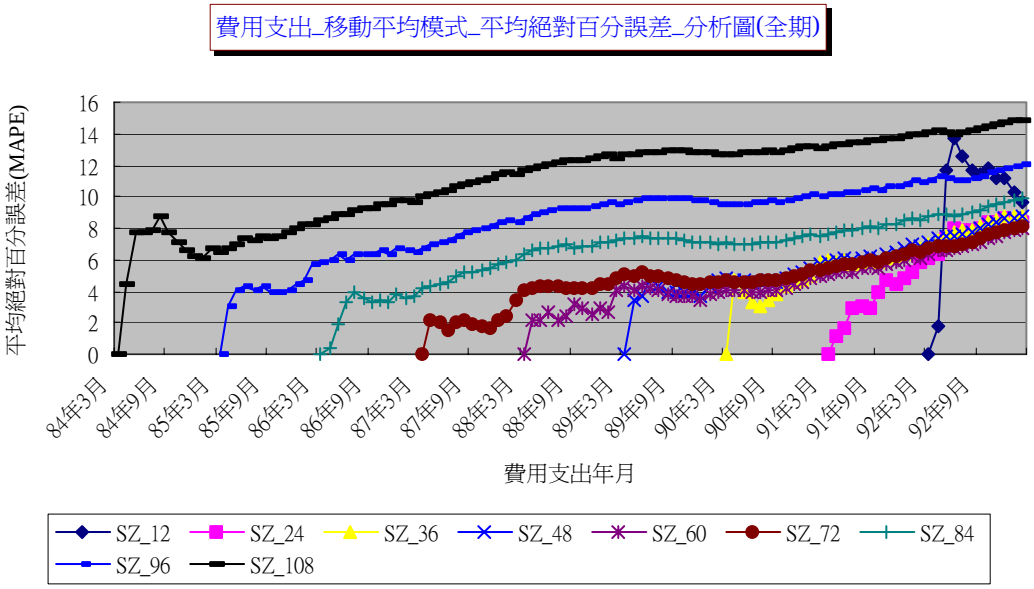


圖 5-3-6　費用支出_移動平均模式_分析圖(全期)

5.3.3 加權移動平均預測模式

程式名稱=exp-macro_weight.sas

輸入數列=費用支出資料集 exp_plot(費用年月,實際金額)

處理條件=觀測值全體或部分

Window size= SZ(12~108,increase by 12)

Weight = WT(0.1~1.0,increase by 0.1)

輸出數列=費用支出加權移動平均預測資料集

(費用年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-3-7、5-3-8）

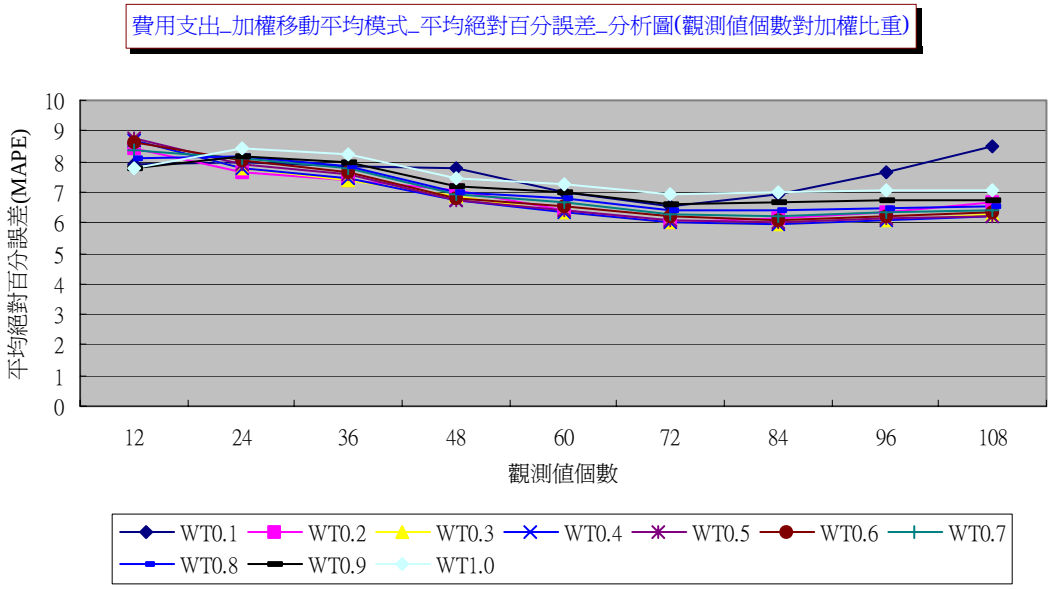


圖 5-3-7　費用支出_加權移動平均模式_分析圖(觀測值個數對加權比重)

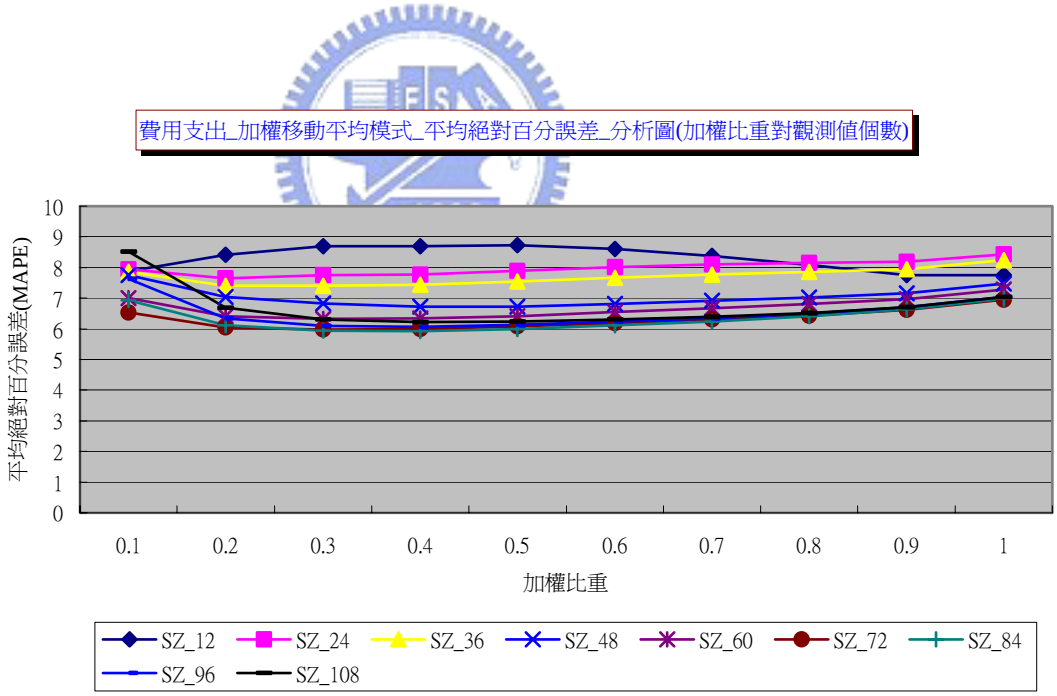


圖 5-3-8　費用支出_加權移動平均模式_分析圖(加權比重對觀測值個數)

5.3.4 迴歸分析模式

程式名稱=exp-macro_autores.sas

輸入數列=費用支出資料集 exp_plot(費用年月,實際金額)

處理條件=觀測值全體或部分

迴歸項選擇=

YX = 費用支出(Y)與費用年月(X)

YY^ = 費用支出(Y)與前期費用支出(Y^)

YX_顯著相關項=費用支出(Y)與費用年月(X) 顯著相關項

相關項個數= NLAG(2~24,increase by 2)

觀測項個數= SZ(12~108,increase by 12)

輸出數列=費用支出迴歸分析預測資料集

(費用年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖(請參考:圖 5-3-9、5-3-10、5-3-11、5-3-12、5-3-13、5-3-14)

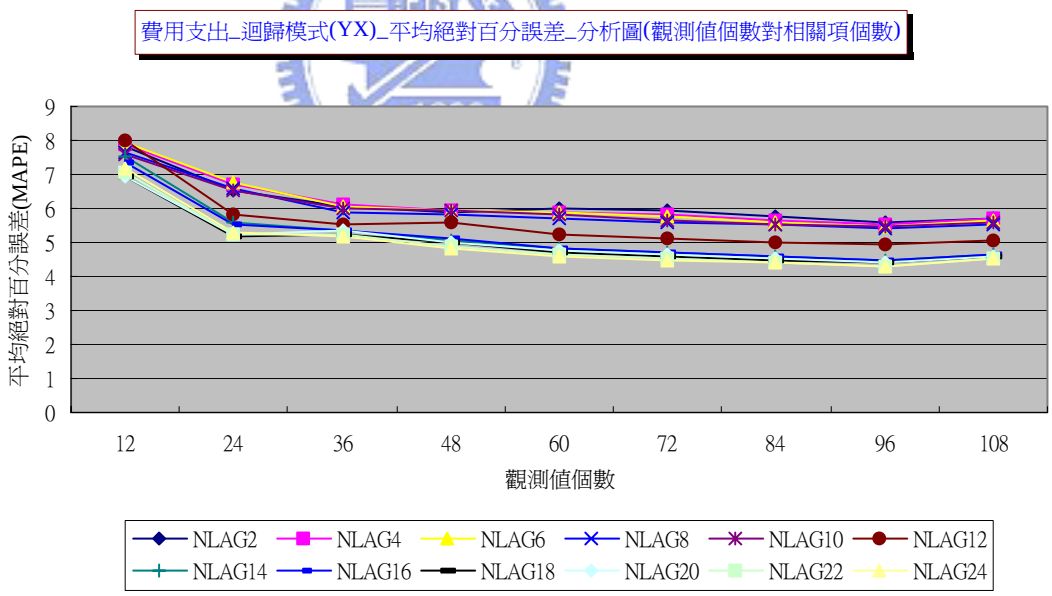


圖 5-3-9 費用支出_迴歸模式(YX)_分析圖(觀測值個數對相關項個數)

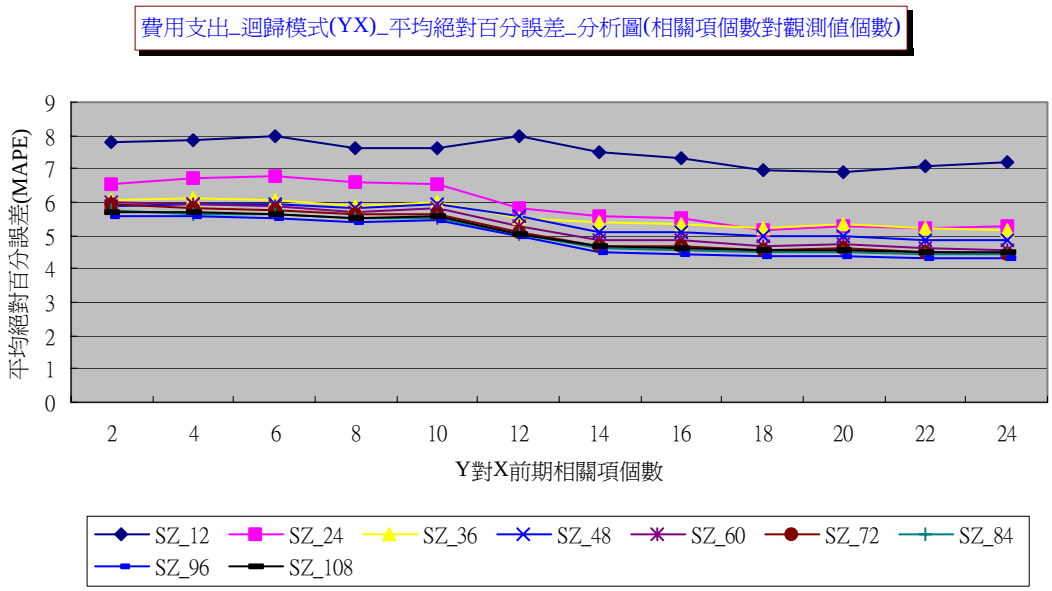


圖 5-3-10　費用支出_迴歸模式(YX)_分析圖(相關項個數對觀測值個數)

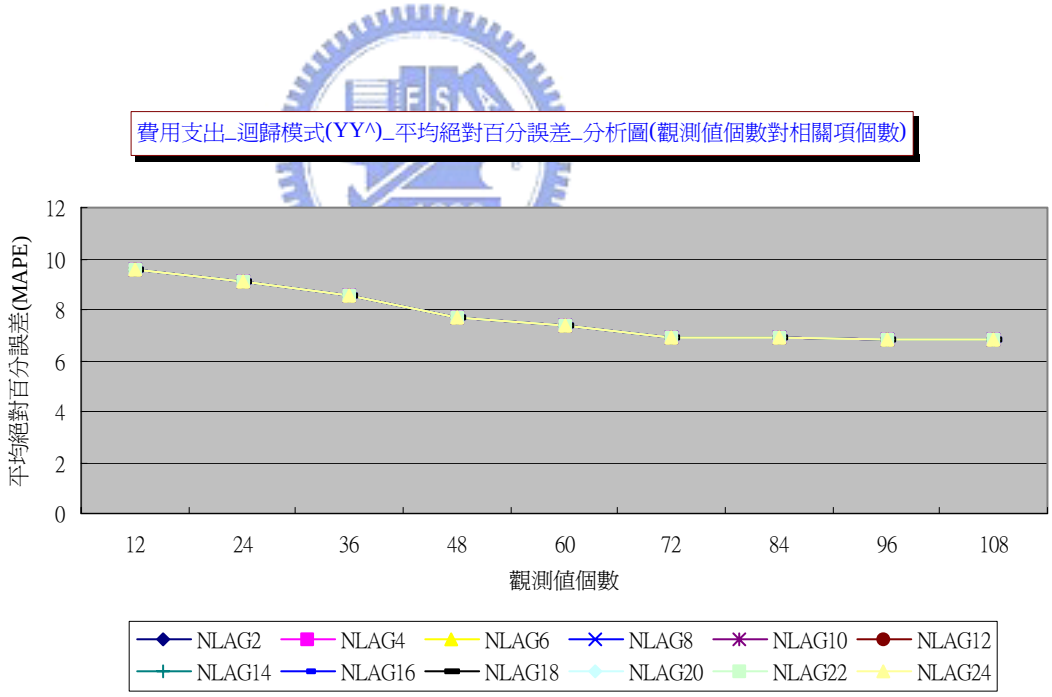


圖 5-3-11　費用支出_迴歸模式(YY^)_分析圖(觀測值個數對相關項個數)

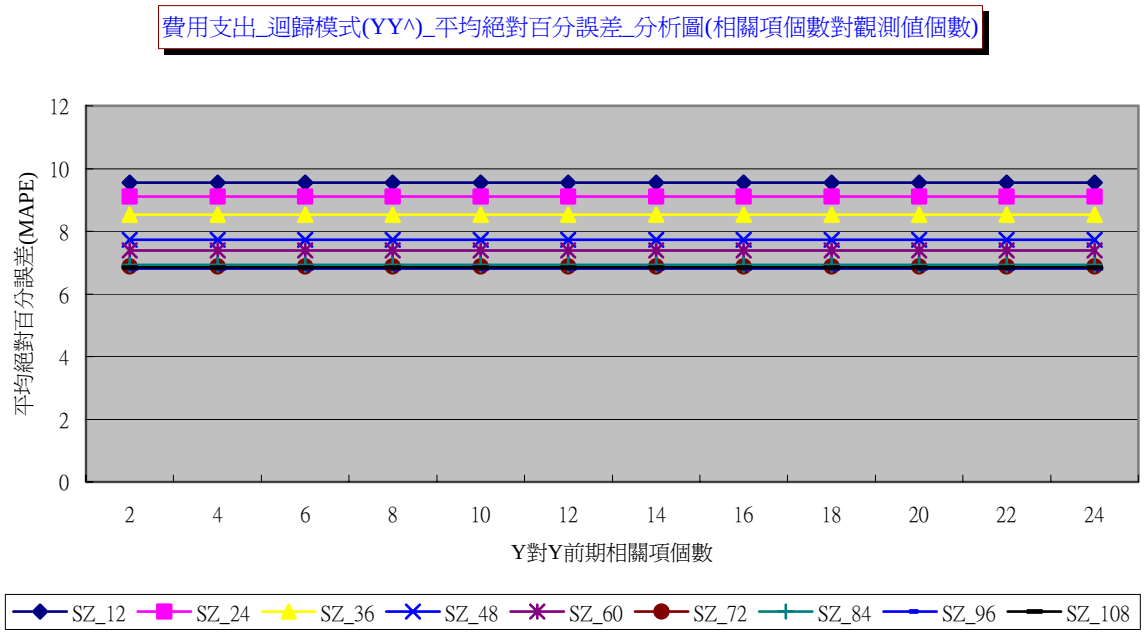


圖 5-3-12　費用支出_迴歸模式(YY^)_分析圖(相關項個數對觀測值個數)

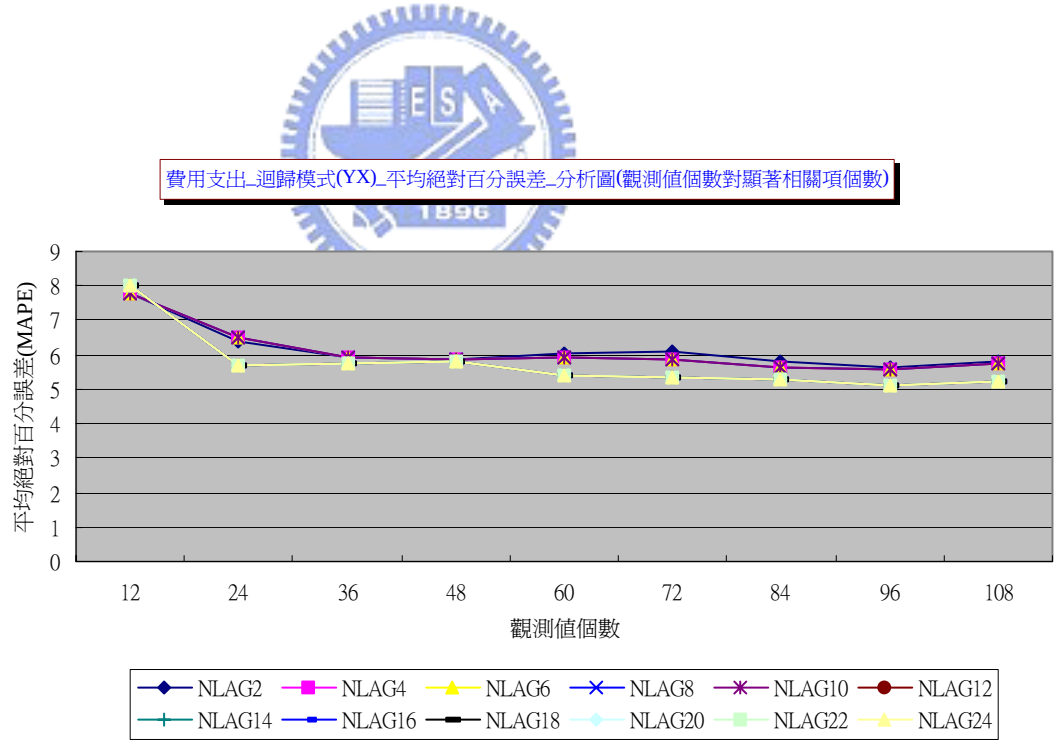


圖 5-3-13　費用支出_迴歸模式(YX)_分析圖(觀測值個數對顯著相關項個數)

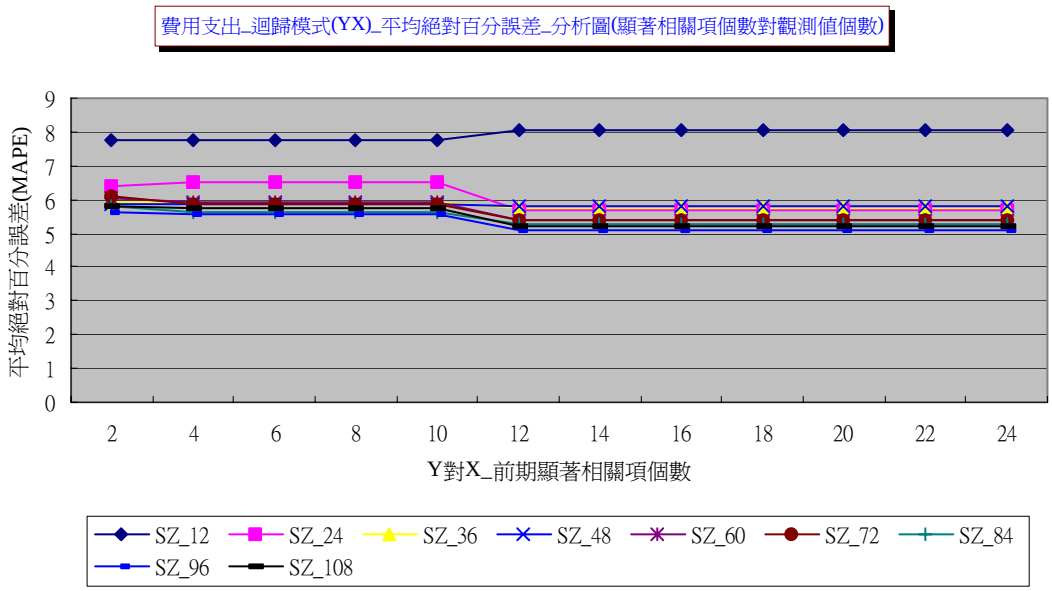


圖 5-3-14　費用支出_迴歸模式(YX)_分析圖(顯著相關項個數對觀測值個數)

5.3.5 自我迴歸整合移動平均模式

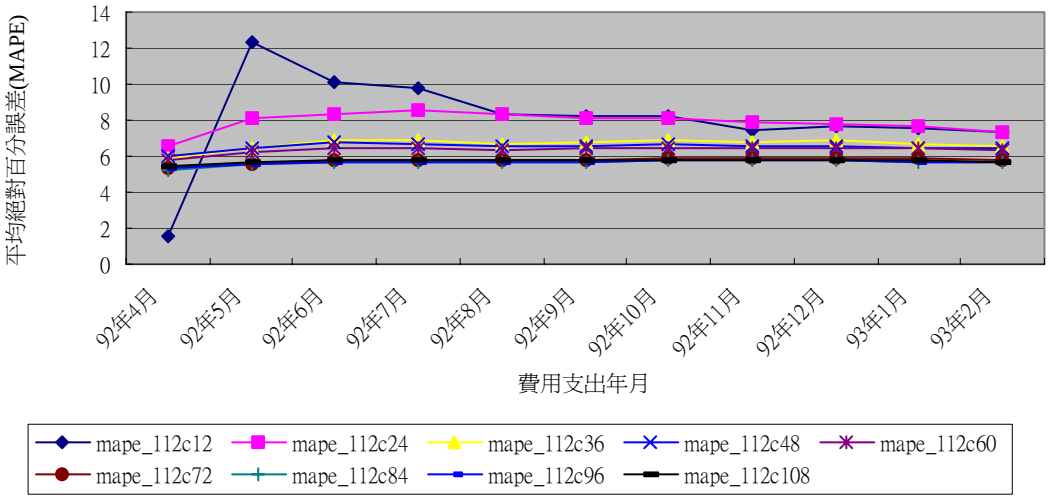
程式名稱=exp-macro_arima_count.sas，get-arima-select.sas
輸入數列=費用支出資料集 exp_plot(費用年月,實際金額)
處理條件=觀測值全體或部分
p d q 階次選擇=p(0~3),d(1,2),q(0~3)
觀測值個數= C (12~108,increase by 12)

使用 DO Loop 迴圈 (p * d * q * c),共產生 288 種組合情形可供判斷分析其預測準度,因組合情形相當龐大,無法以肉眼比較,故撰寫程式以觀測值個數的角度來分析挑選最適之階次,並最後得到一建議的 p d q 階次組合。

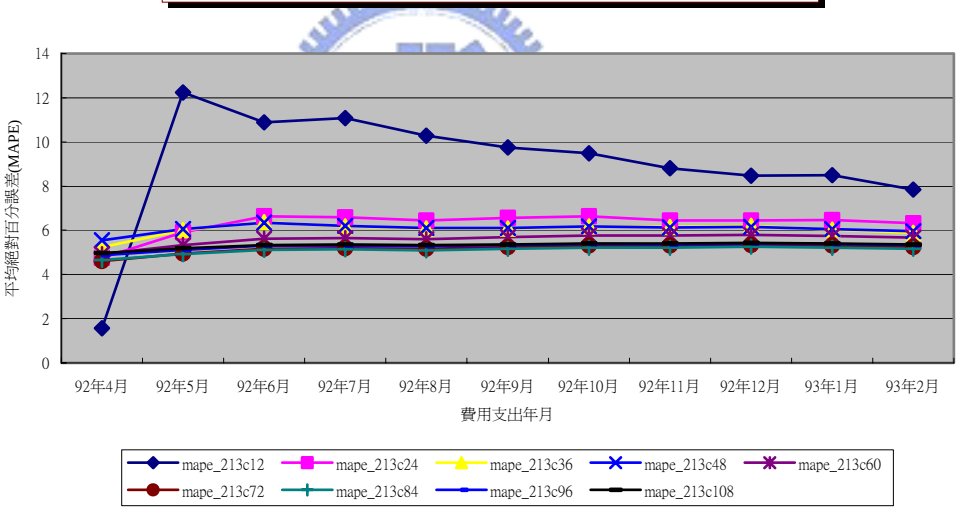
輸出數列=費用支出自我迴歸整合移動平均預測資料集
(費用年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-3-15）

費用支出ARIMA112預測_平均絕對百分誤差_觀測值個數_分析圖



費用支出ARIMA213預測_平均絕對百分誤差_觀測值個數_分析圖



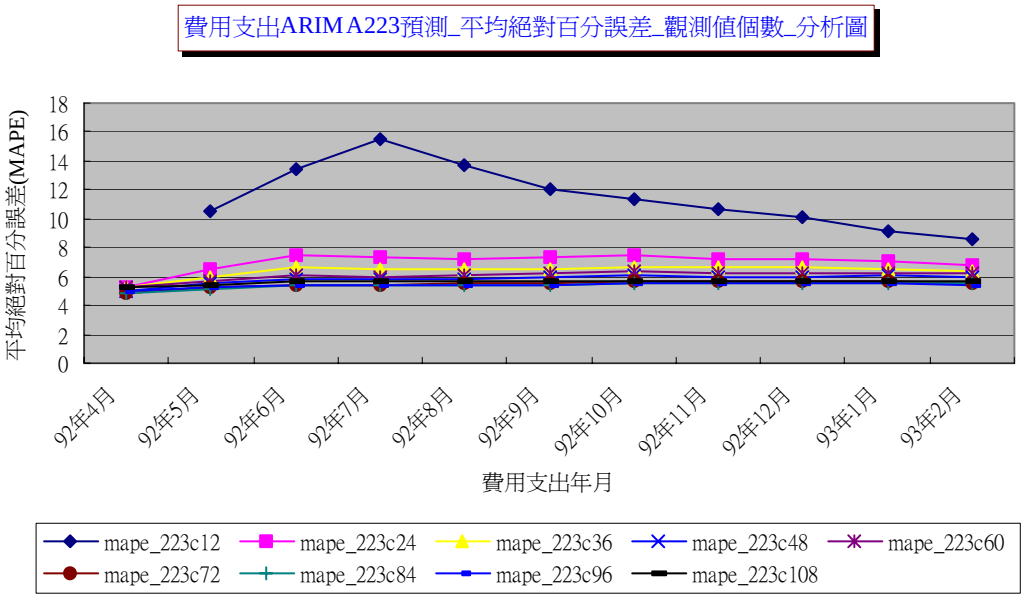


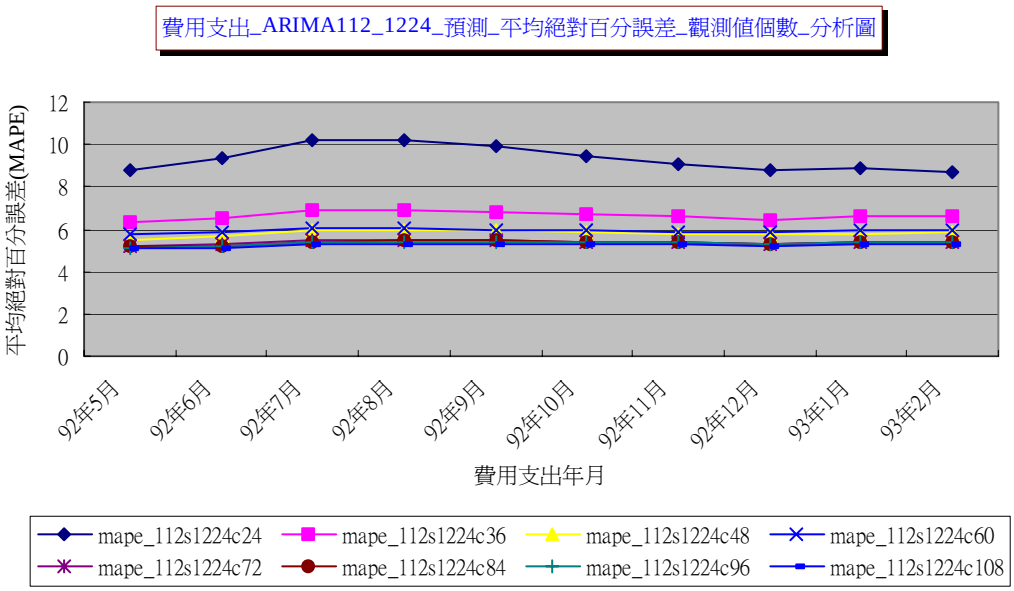
圖 5-3-15　費用支出_自我迴歸整合移動平均模式_分析圖（ARIMA）

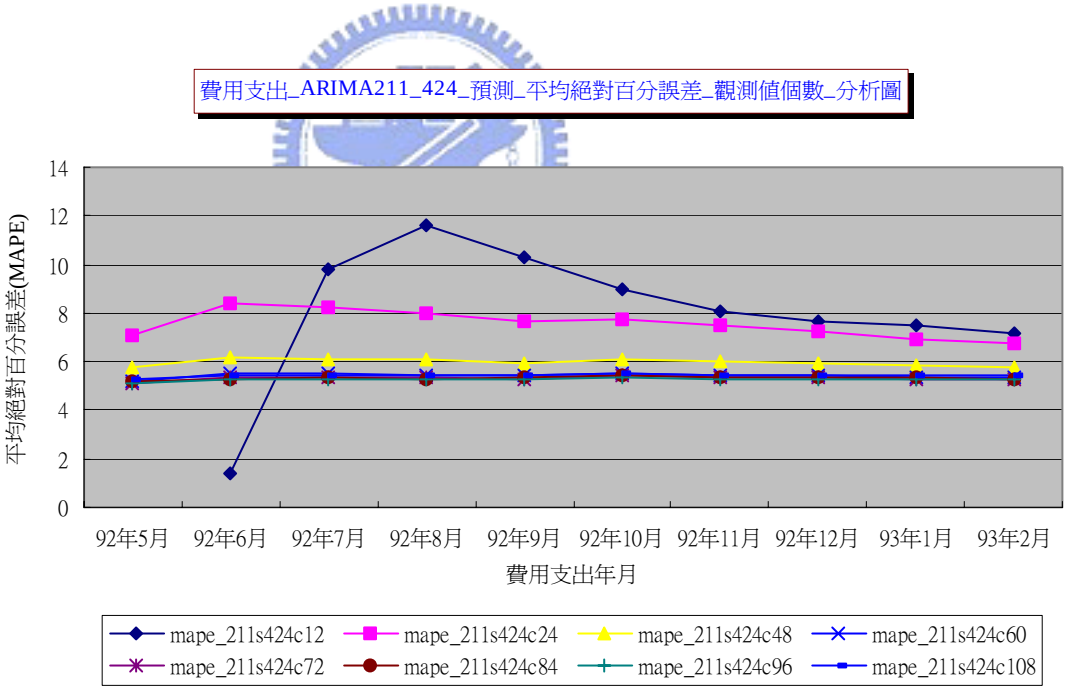
5.3.6 季節性自我迴歸整合移動平均模式

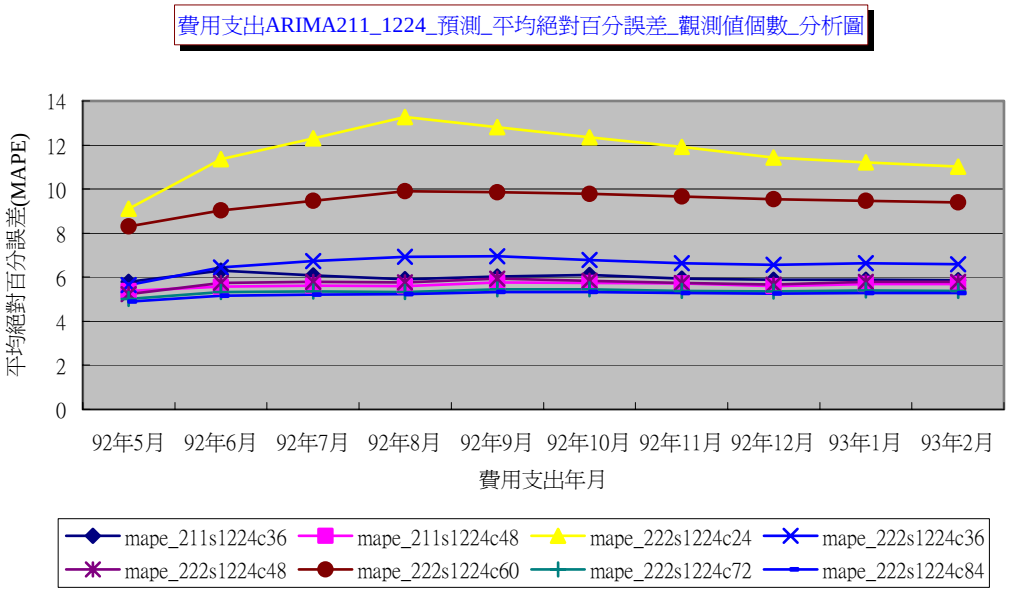
程式名稱=exp-macro_arima_cx_season.sas，get-arima-select.sas
輸入數列=費用支出資料集 exp_plot(費用年月,實際金額)
處理條件=觀測值全體或部分
pdq 階次選擇=p(0~3),d(1,2),q(0~3)
PDQ 階次選擇=P(3,4,12),D(1,2),Q(3,4,12)
觀測值個數= C (12~108,increase by 12)

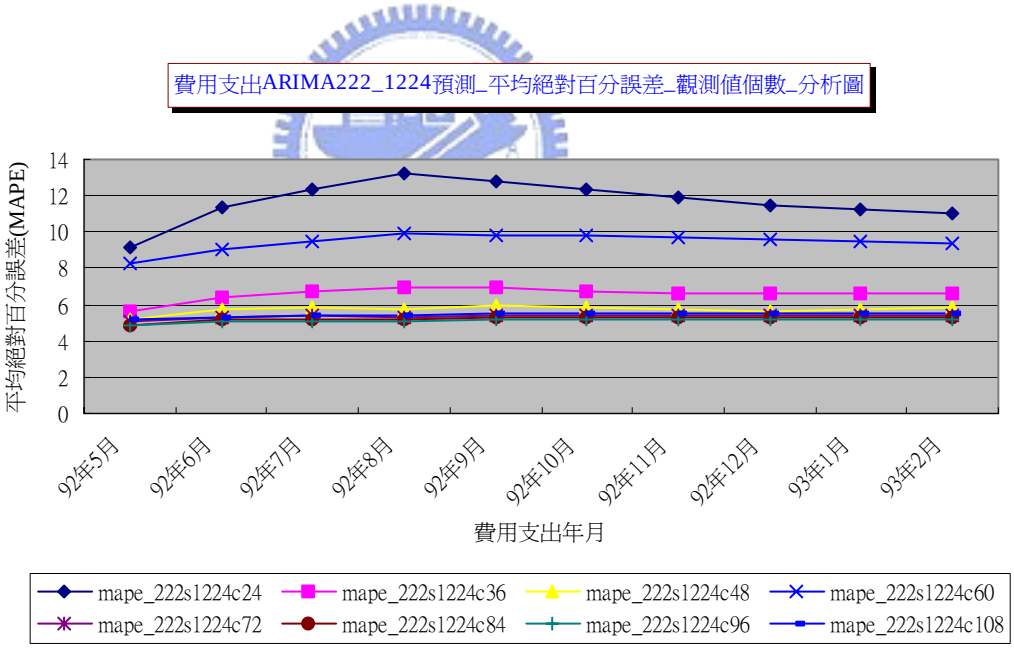
使用 DO Loop 迴圈 (p * d * q * P * D * Q * C),共產生 5184 種組合情形可供判斷分析其預測準確度,因組合情形過於龐大,無法以肉眼比較,故撰寫程式以觀測值個數的角度來分析挑選最適之階次,並最後得到一建議的 p d q 及 P D Q 階次組合。
輸出數列=費用支出季節性自我迴歸整合移動平均預測資料集
(費用年月,實際金額,預測金額,預測誤差各項相關統計值)

各項統計分析圖（請參考：圖 5-3-16）


費用支出_ARIMA112_1224_預測_平均絕對百分誤差_觀測值個數_分析圖
平均絕對百分誤差(MAPE)
12
10
8
6
4
2
0
92年5月
92年6月
92年7月
92年8月
92年9月
92年10月
92年11月
92年12月
93年1月
93年2月
費用支出年月
mape_112s1224c24
mape_112s1224c36
mape_112s1224c48
mape_112s1224c60
mape_112s1224c72
mape_112s1224c84
mape_112s1224c96
mape_112s1224c108


費用支出_ARIMA211_424_預測_平均絕對百分誤差_觀測值個數_分析圖
平均絕對百分誤差(MAPE)
14
12
10
8
6
4
2
0
92年5月
92年6月
92年7月
92年8月
92年9月
92年10月
92年11月
92年12月
93年1月
93年2月
費用支出年月
mape_211s424c12
mape_211s424c24
mape_211s424c48
mape_211s424c60
mape_211s424c72
mape_211s424c84
mape_211s424c96
mape_211s424c108


費用支出ARIMA211_1224_預測_平均絕對百分誤差_觀測值個數_分析圖
平均絕對百分誤差(MAPE)
14
12
10
8
6
4
2
0
92年5月
92年6月
92年7月
92年8月
92年9月
92年10月
92年11月
92年12月
93年1月
93年2月
費用支出年月
mape_211s1224c36
mape_211s1224c48
mape_222s1224c24
mape_222s1224c36
mape_222s1224c48
mape_222s1224c60
mape_222s1224c72
mape_222s1224c84


費用支出ARIMA222_1224預測_平均絕對百分誤差_觀測值個數_分析圖
平均絕對百分誤差(MAPE)
14
12
10
8
6
4
2
0
92年5月
92年6月
92年7月
92年8月
92年9月
92年10月
92年11月
92年12月
93年1月
93年2月
費用支出年月
mape_222s1224c24
mape_222s1224c36
mape_222s1224c48
mape_222s1224c60
mape_222s1224c72
mape_222s1224c84
mape_222s1224c96
mape_222s1224c108

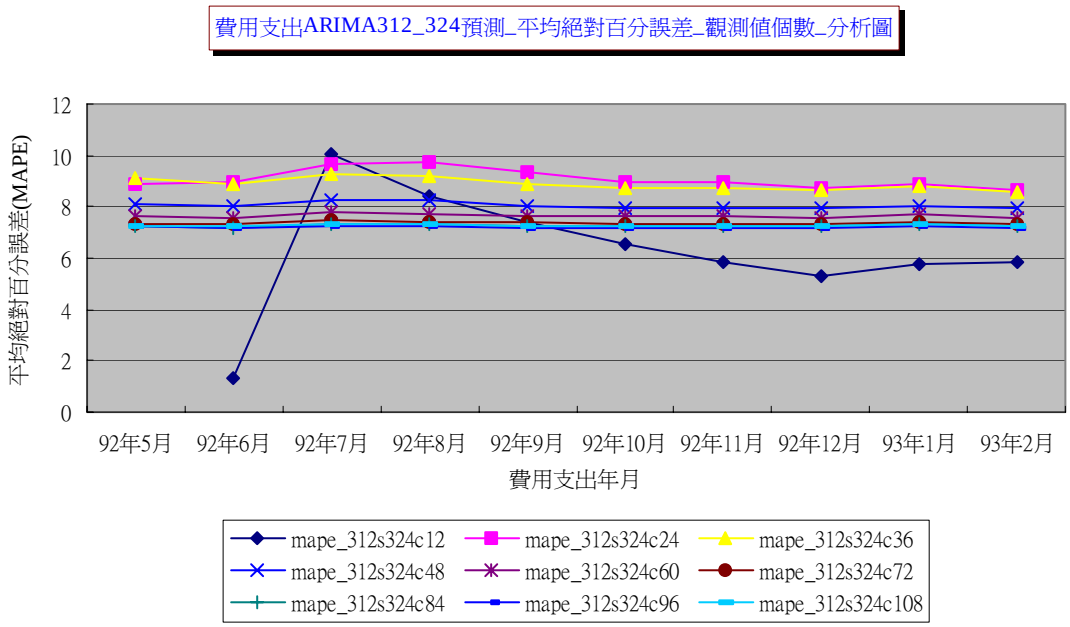


圖 5-3-16 費用支出_自我迴歸整合移動平均模式_分析圖 (SEASON ARIMA)

5.4 費用支出六種預測模式綜合比較

程式名稱=exp-MACRO-COMPARE-6.SAS
輸入項目=六項預測模式產生的輸出數列資料集中最適合的階次組合
運算條件=分別針對六項分析模式給特定的參數組合，如 window size，weight，迴歸項選擇，pdq，PDQ 階次，觀測值個數 c
輸出數列=六種費用支出預測模式最佳項組合資料集(費用年月,實際金額,六種預測金額,六種預測誤差各項相關統計值(依觀測值個數分類))、六種模式平均絕對百分誤差_全期數列資料集

各項統計分析圖（請參考：圖 5-4-1、5-4-2、5-4-3、5-4-4、5-4-5、5-4-6、5-4-7、5-4-8、5-4-9、5-4-10、5-4-11、5-4-12、5-4-13、5-4-14）

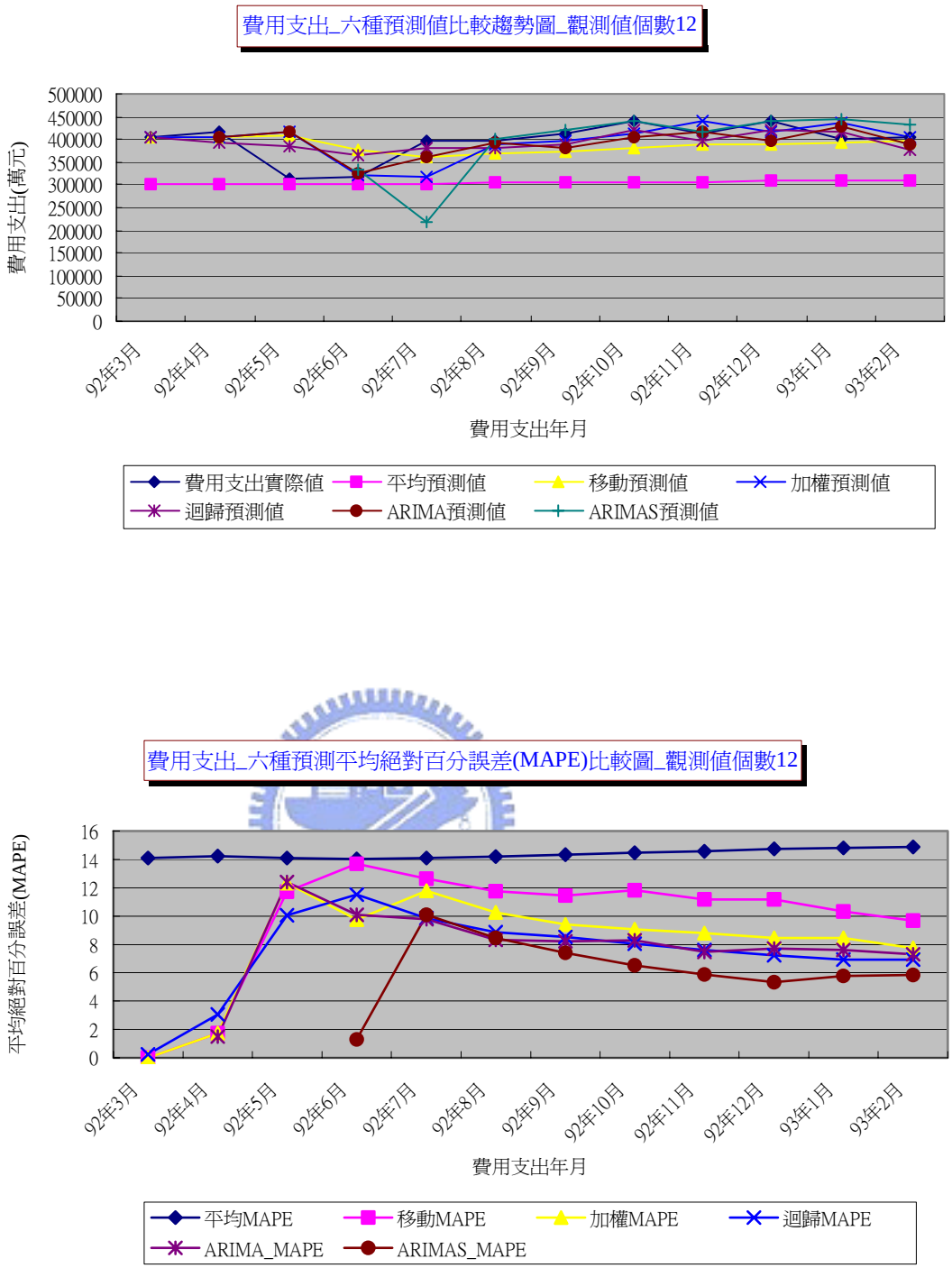
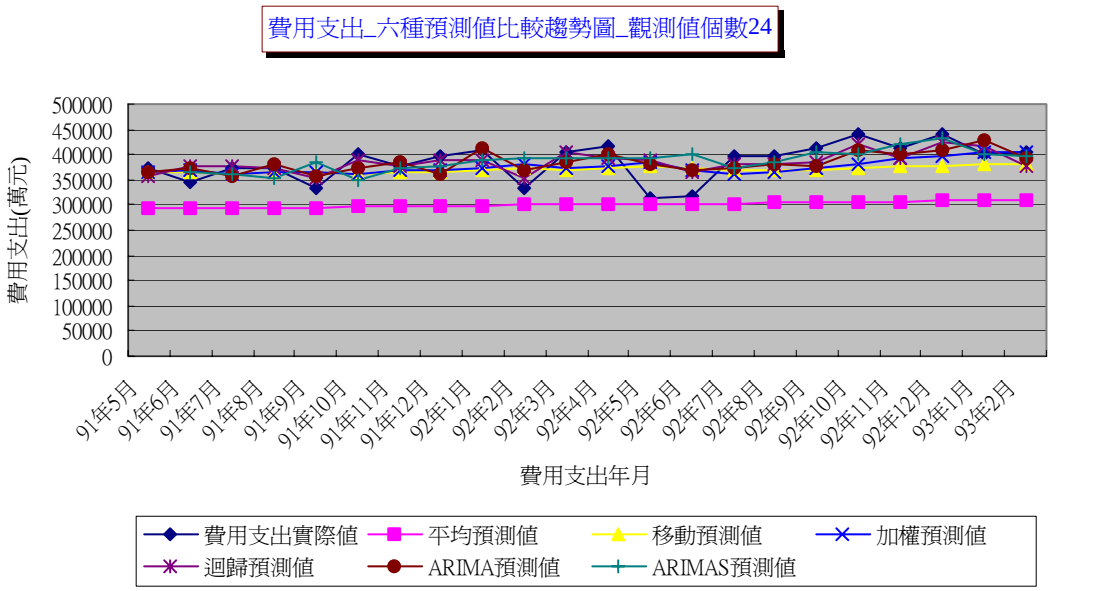


圖 5-4-1　費用支出_六種預測模式比較圖(觀測值個數 12)



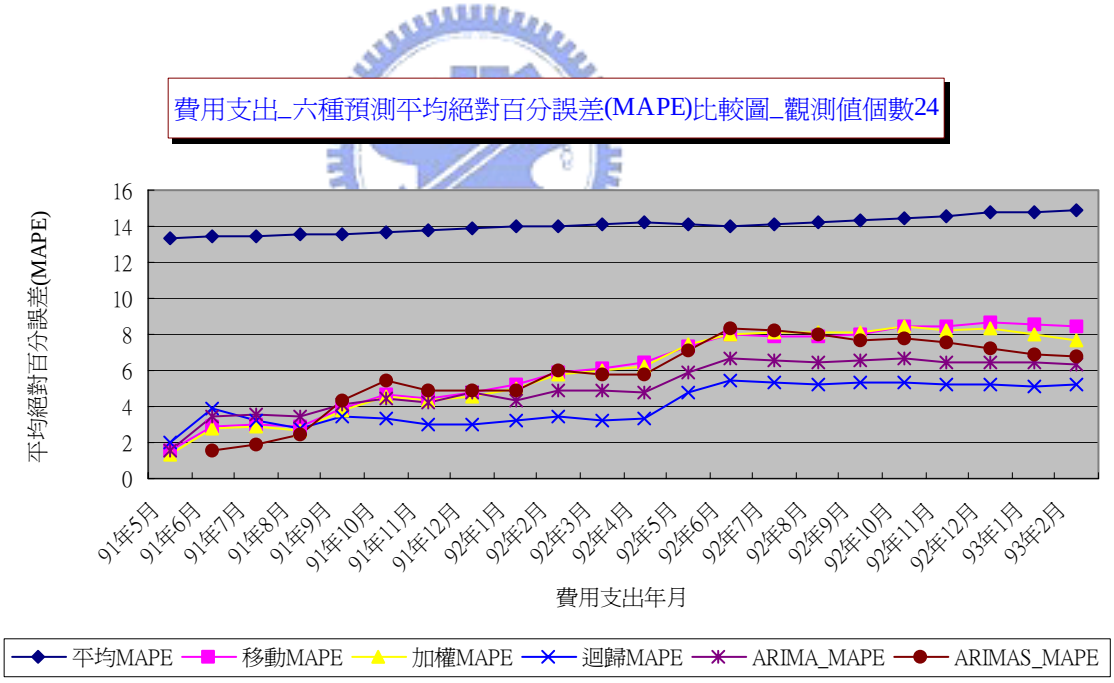


圖 5-4-2　費用支出_六種預測模式比較圖(觀測值個數 24)

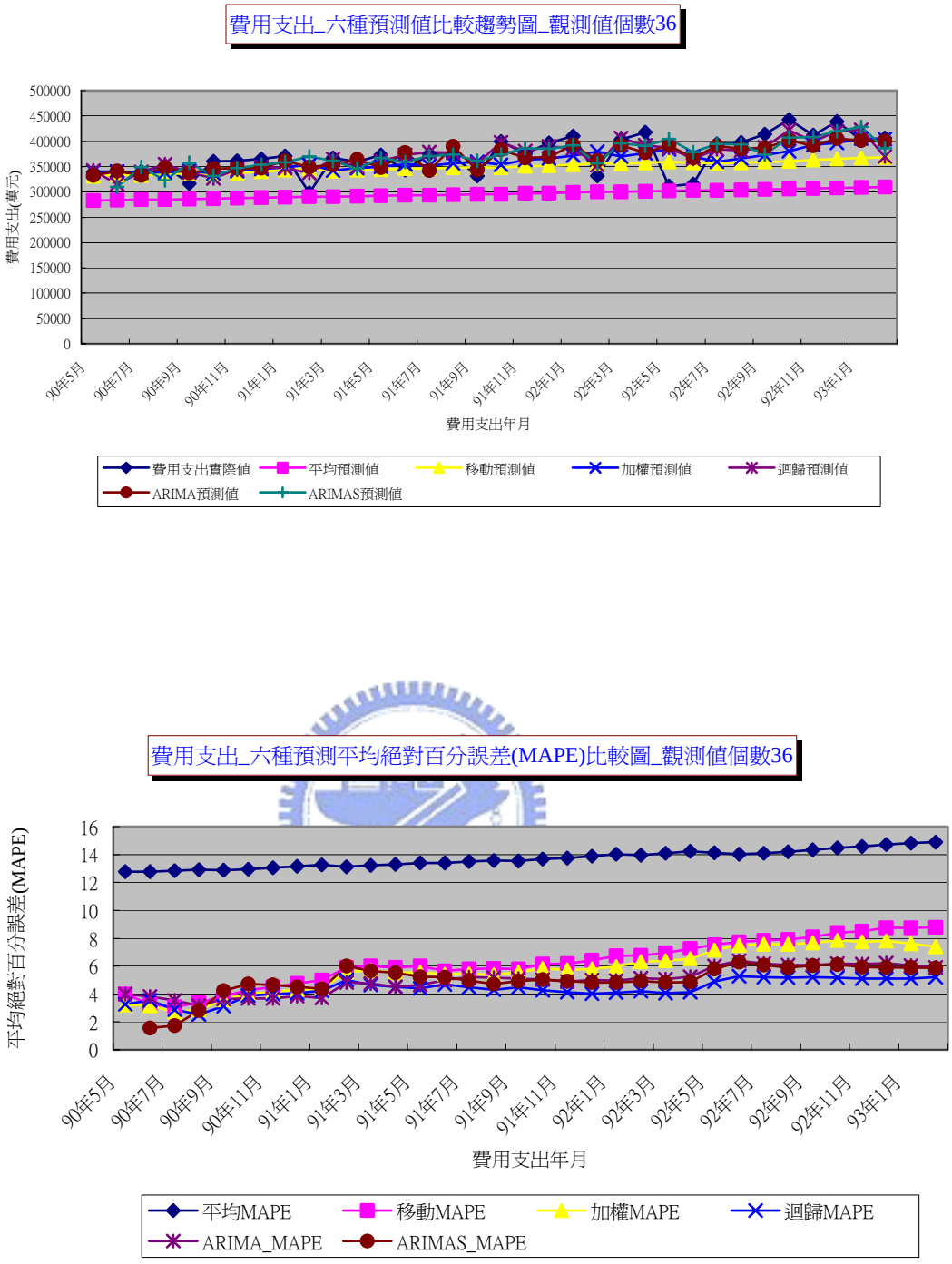
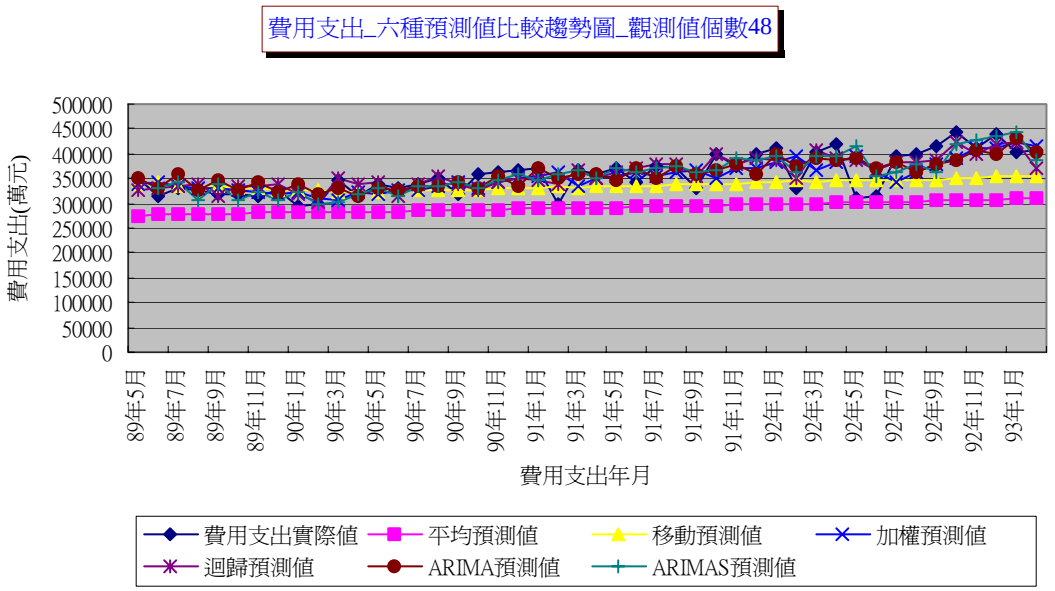


圖 5-4-3　費用支出_六種預測模式比較圖(觀測值個數 36)



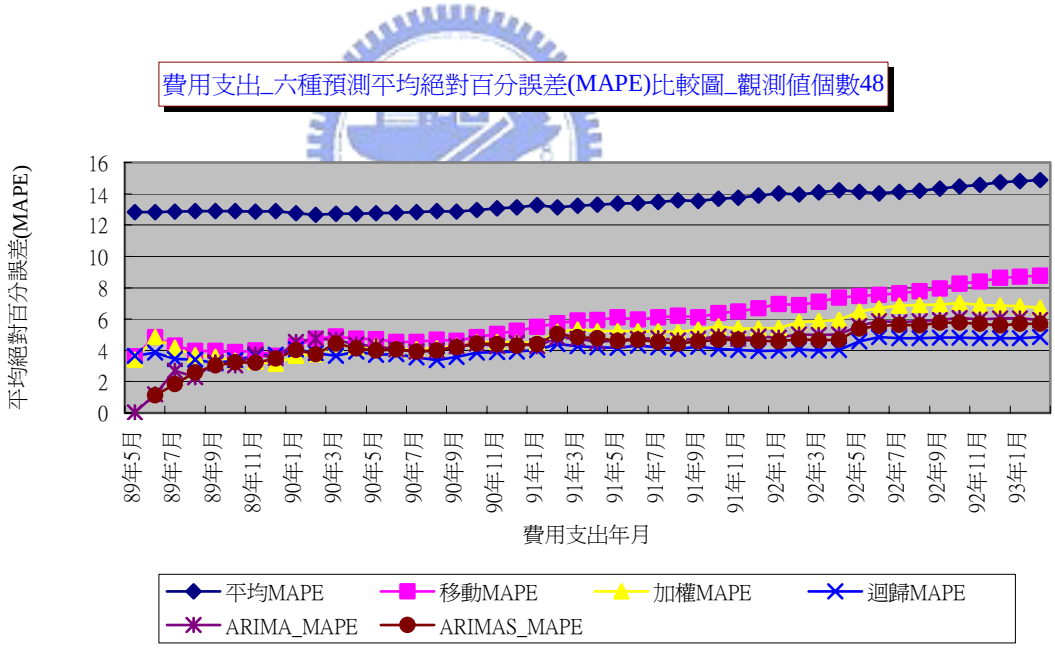


圖 5-4-4　費用支出_六種預測模式比較圖(觀測值個數 48)

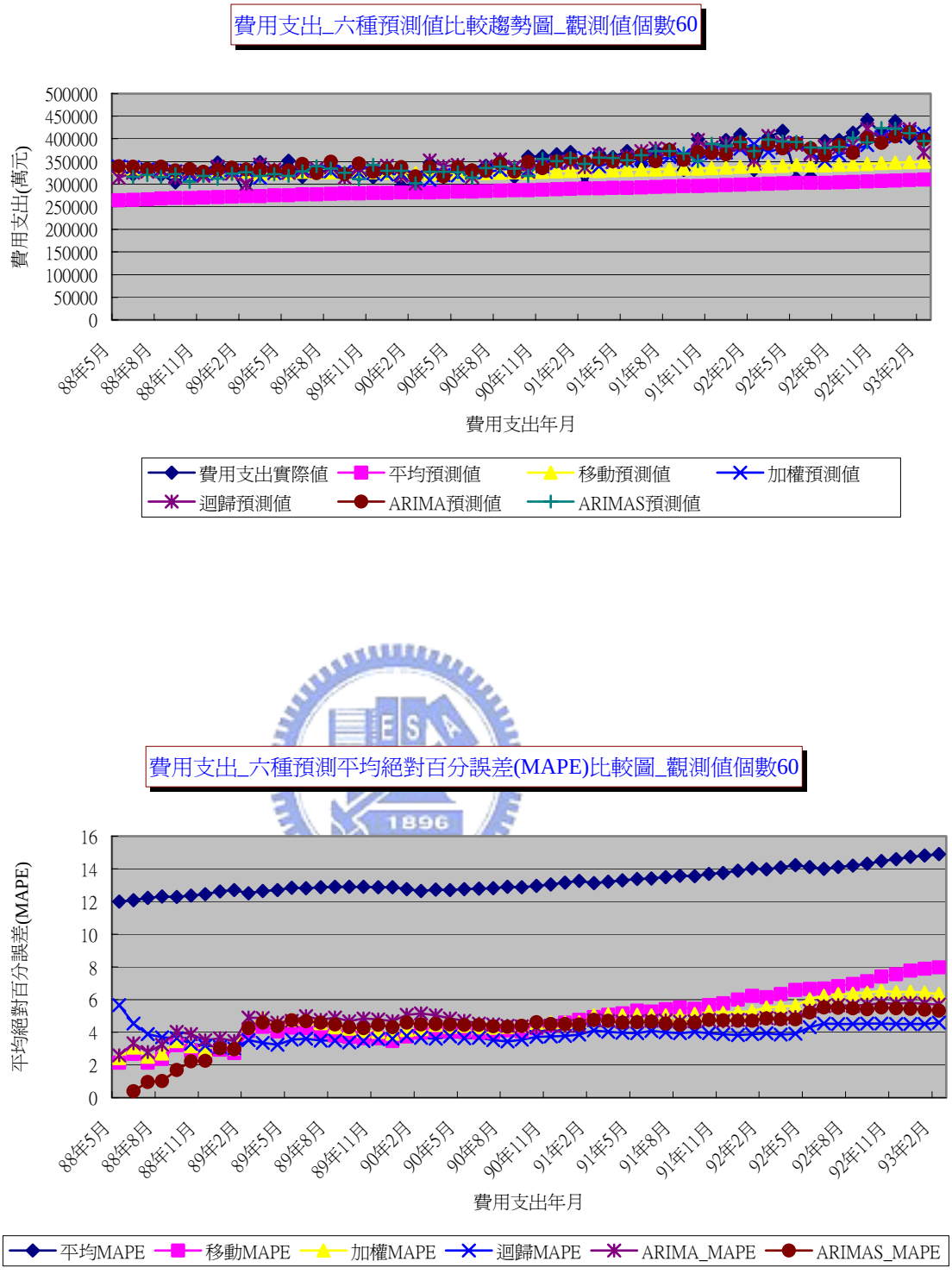
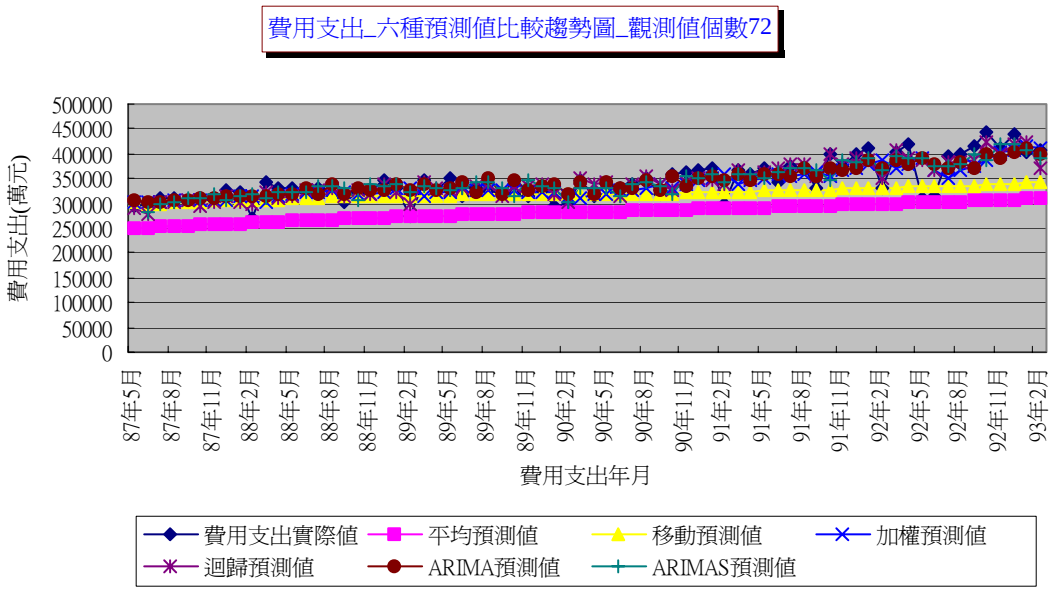


圖 5-4-5　費用支出_六種預測模式比較圖(觀測值個數 60)



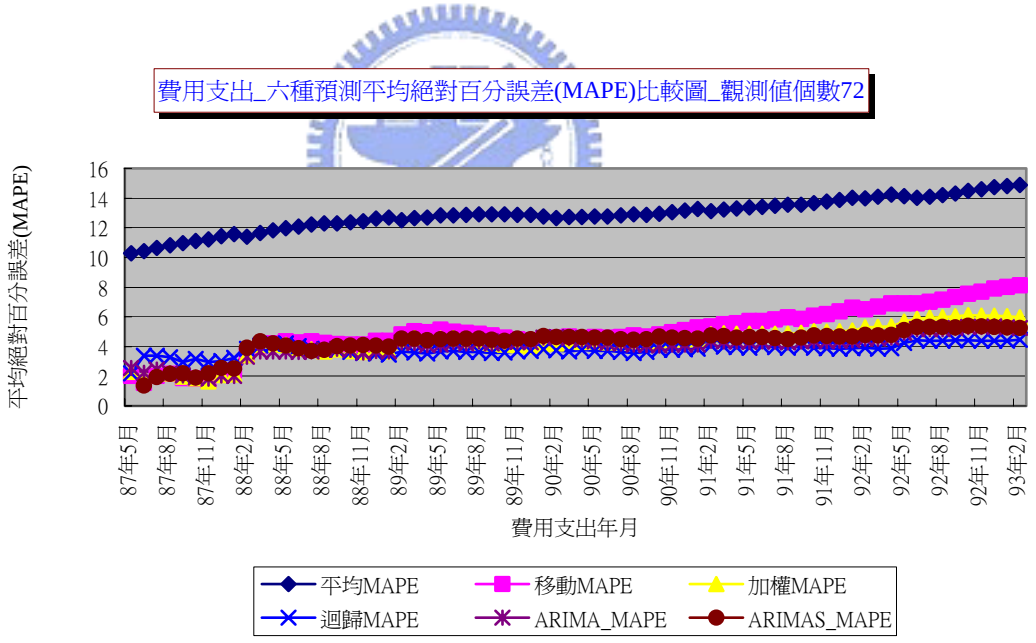


圖 5-4-6　費用支出_六種預測模式比較圖(觀測值個數 72)

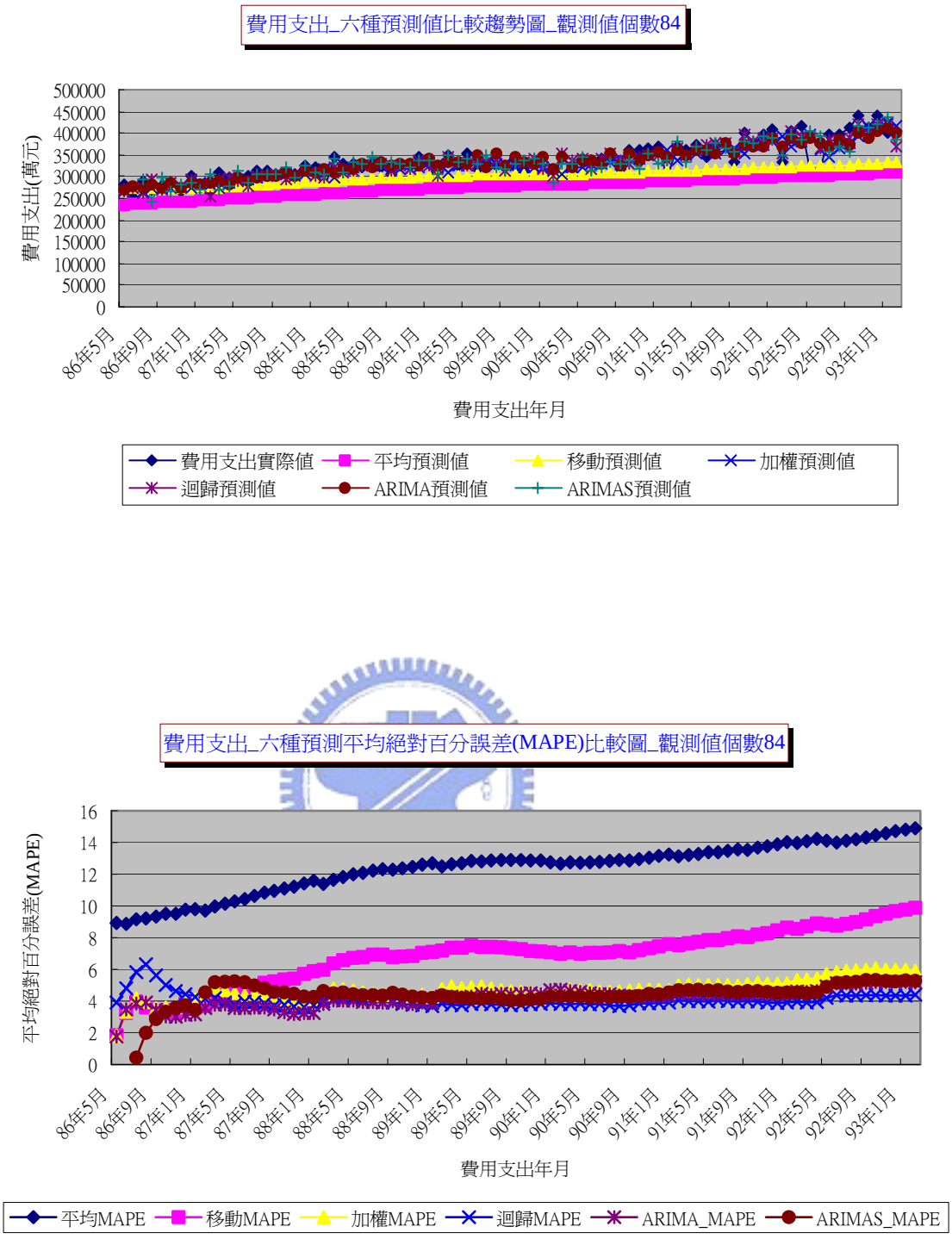


圖 5-4-7　費用支出_六種預測模式比較圖(觀測值個數 84)

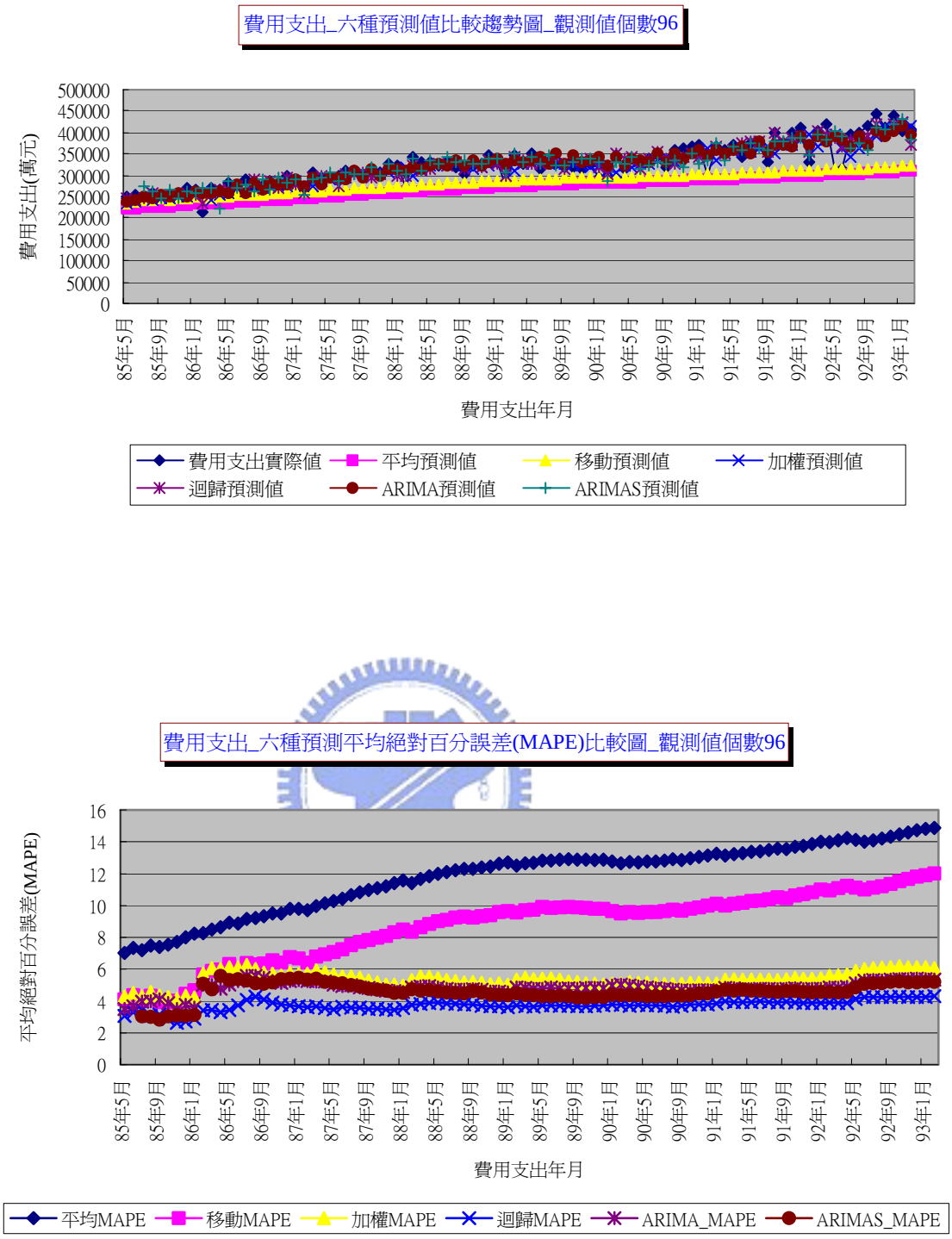


圖 5-4-8　費用支出_六種預測模式比較圖(觀測值個數 96)

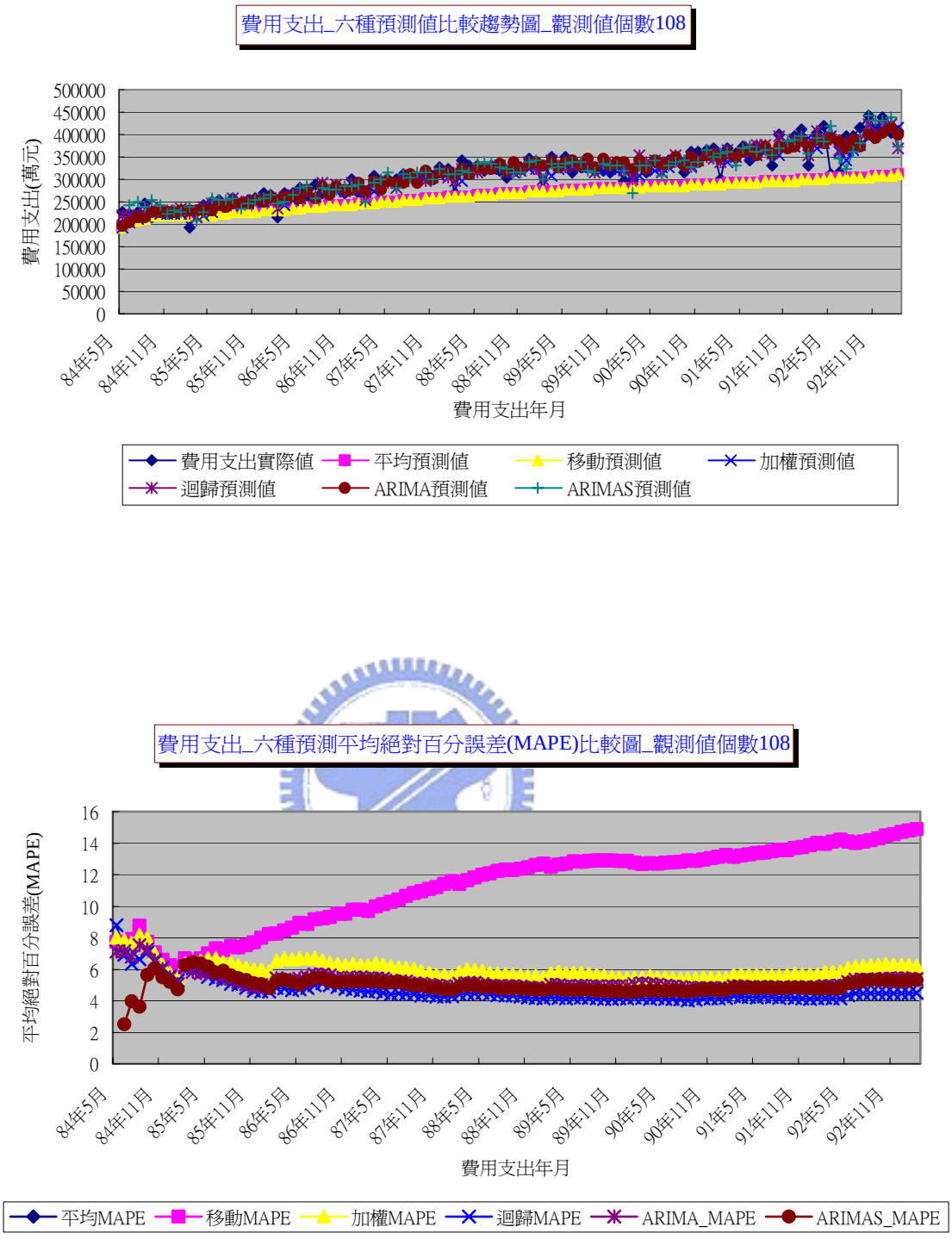


圖 5-4-9　費用支出_六種預測模式比較圖(觀測值個數 108)

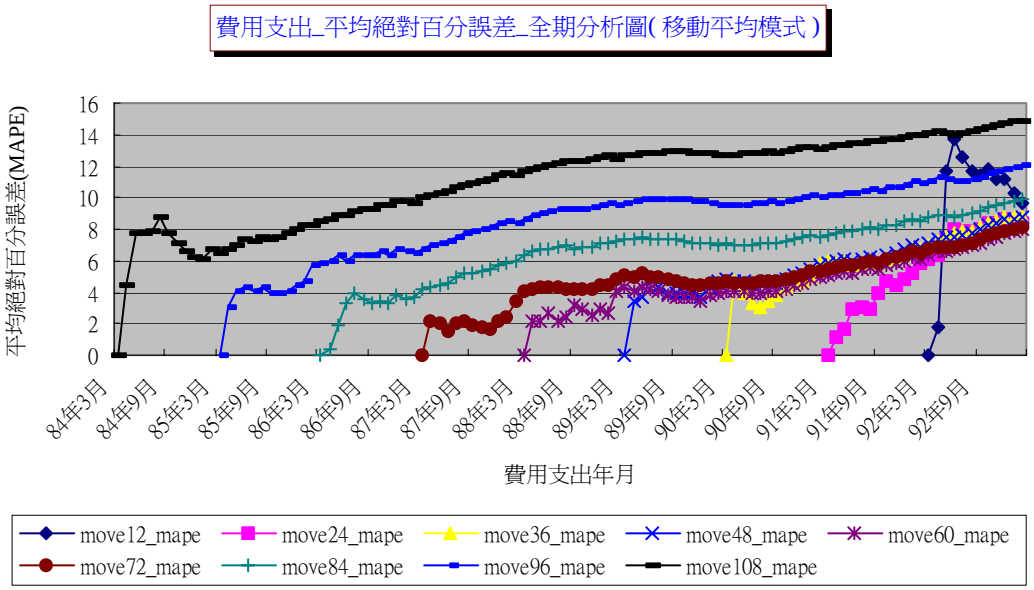


圖 5-4-10　費用支出_全期分析圖(移動平均模式)

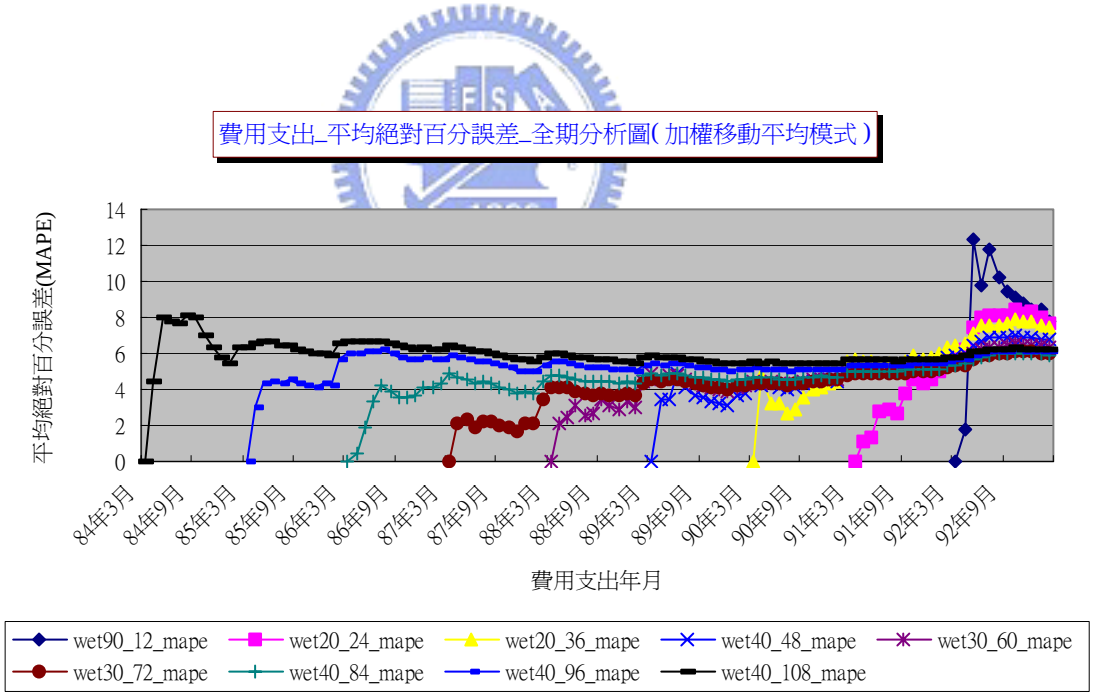


圖 5-4-11　費用支出_全期分析圖(加權移動平均模式)

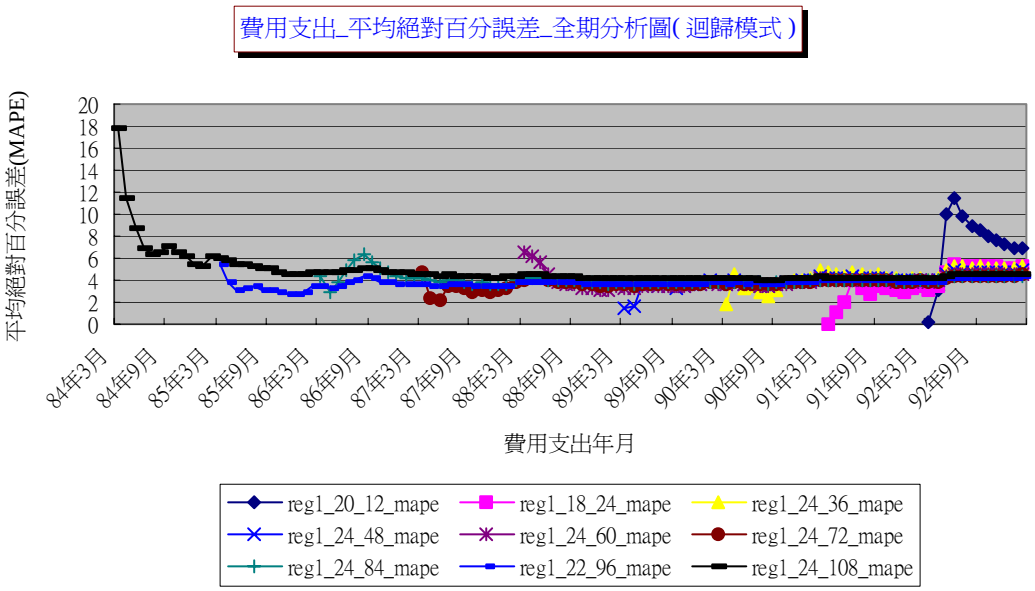


圖 5-4-12　費用支出_全期分析圖(迴歸模式)

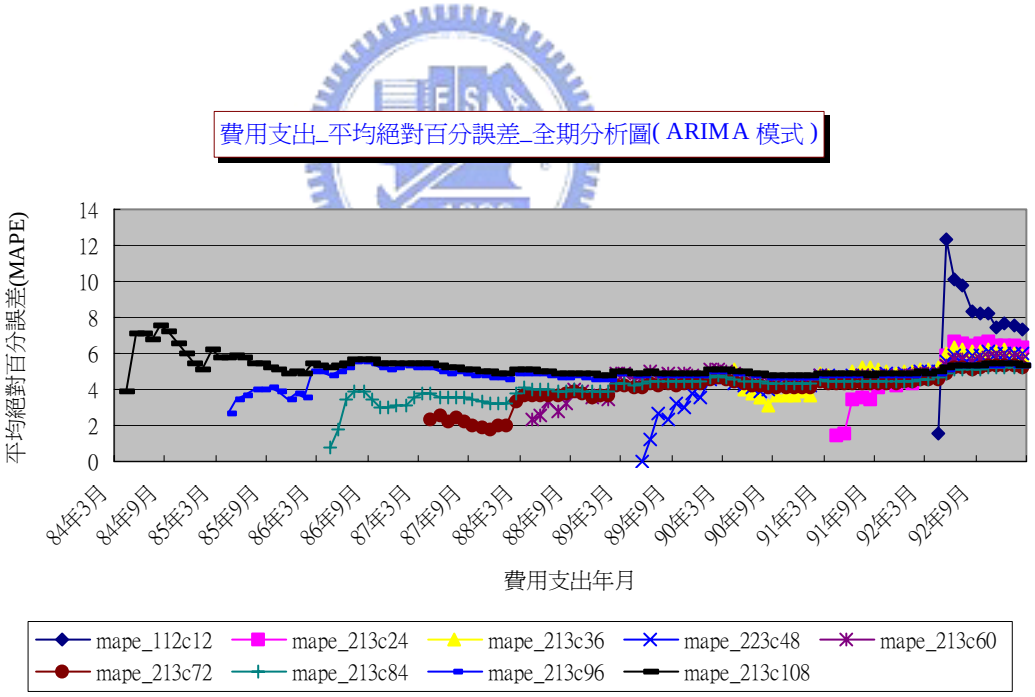


圖 5-4-13　費用支出_全期分析圖(ARIMA 模式)

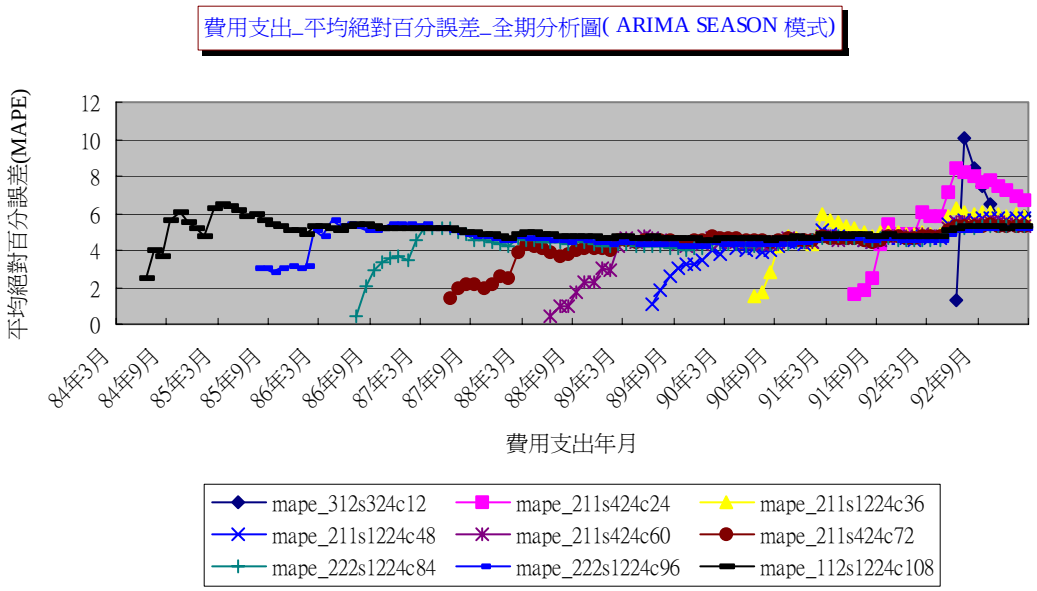


圖 5-4-14　費用支出_全期分析圖(ARIMA SEASON 模式)

六、 結果評估與解釋

依據前述資料採礦及運算的結果，吾人發現各預測模式均有不同的特性，茲依各模式分述如下：

6.1 平均預測模式

觀察圖 5-1-1、5-1-2、5-1-3、5-1-4、5-3-1、5-3-2、5-3-3、5-3-4 可知，平均預測模式是最簡單,最容易理解的方法，因其只運用平均值的概念將全部的觀測值計算其平均值以作為預測值，處理方式較無彈性，預測準確度較差，本研究中保費收入預測值的 MAPE=11.201 ，費用支出預測值的 MAPE=14.888，兩者的 MAPE 值均過高。

6.2 移動平均預測模式

本模式預測準確度與觀測值個數(window size)相關，觀察圖 5-1-5、5-1-6、5-2-10、5-3-5、5-3-6、5-4-10 發現，大體而言，當觀測值個數越少，預測準確度越高，當觀測值個數越多時,預測準確度反而降低，當觀測值全部使用時，其結果與平均預測模式相同。

本研究中因須與其他模式比較，觀測值個數需>=24，故觀測值個數=12 應視為特例需另行考量；吾人也發現費用支出的 MAPE 普遍大於保費收入的 MAPE 值，顯示費用支出數列的變異性高於保費收入數列。因篇幅有限，MAPE 值僅列出最小的二組。

保費收入部分：

Window size=12，MAPE=1.202

Window size=24，MAPE=5.201

費用支出部分：

window size=60，MAPE=7.978

window size=70，MAPE=8.133

6.3 加權移動平均預測模式

本模式預測準確度與觀測值個數(window size)及權值(weight)相關，觀察圖 5-1-7、5-1-8、5-2-11、5-3-7、5-3-8、5-4-11 發現，大體而言，保費收入數列，在觀測值個數相同的情形下，權值介於 0.3~0.5 之間，預測準確度較高。在權值相同的形下，觀測值個數介於 12~24 間有最好的預測準確度，觀測值=48 時準確度最低，隨著觀測值個數增加，準確度又慢慢提高。

保費收入部分，在 window size 48 期時有最高的 MAPE 值，經探討其原因係中央健保局自民國 88 年起，開始進行中斷保費收繳作業有關，且其高峰期集中在 89、90、91 年的部分月份，故造成相對的 window size 為 48、

60、72、84 期時，均有較高的 MAPE 值。研究結果顯示：中央健保局收繳中斷保費的政策，的確改變了保費收入預測的行為模式，保費收入預測準確度的變異程度因而增加。

費用支出數列，在觀測值個數相同的情形下，權值介於 0.3~0.5 之間，預測準確度較高。在權值相同的形下，隨著觀測值個數增加，準確度又慢慢提高。

本研究中因須與其他模式比較，觀測值個數需>=24，故觀測值個數=12 應視為特例需另行考量；吾人也發現費用支出的 MAPE 普遍大於保費收入的 MAPE 值，顯示費用支出數列的變異性高於保費收入數列。因篇幅有限，MAPE 值僅列出較佳的數組。

保費收入部分：

window size=24，weight=0.8，MAPE=1.970

window size=108，weight=0.4，MAPE=3.114

window size=96，weight=0.4，MAPE=3.284

費用支出部分：

window size=84，weight=0.4，MAPE=5.919

window size=72，weight=0.3，MAPE=5.981

window size=96，weight=0.4，MAPE=6.067

6.4 迴歸分析預測模式

6.4.1 Y 對 X 迴歸模式

本模式預測準確度相關於觀測值個數(window size)及相關項個數(nalg)，觀察圖 5-1-9、5-1-10、5-2-12、5-3-9、5-3-10、5-4-12 發現，整體而言，保費收入數列的觀測值個數介於 12~24 間有最好的預測準確度，觀測值=48 時準確度最低，隨著觀測值個數增加，準確度又慢慢提高；相關項個數越少，準確度越差，相關項個數越多，準確度越高。

在觀測值個數相同的情形下，相關項個數增加，可提升預測準確度。

在相關項個數相同時，觀測值個數介於 12~24 間有最好的預測準確度，觀測值=48 時準確度最低，隨著觀測值個數增加，準確度又慢慢提高；此現象與前述加權移動平均情形相同，中斷保費收繳是主要原因。

費用支出數列整體而言，觀測值個數越高，準確度越高；相關項個數越多，準確度越高。

在觀測值個數相同時，相關項個數越多，準確度越高。

在相關項個數相同時，觀測值個數越多，準確度越高。

6.4.2 Y 對 Y 的前項值 迴歸模式

本模式預測準確度只與觀測值個數(window size)相關，吾人觀察圖 5-1-11、5-1-12、5-3-11、5-3-12 發現，整體而言，保費收入數列的觀測值個

數介於 12~24 間有最好的預測準確度，觀測值=48 時準確度最低，隨著觀測值個數增加，準確度又慢慢提高；此現象與前述加權移動平均情形相同，中斷保費收繳是主要原因。

特別的是，相關項個數的改變對預測準確度沒有影響。

費用支出數列整體而言，預測準確度只與觀測值個數(window size)相關，觀測值個數越高，準確度越高。

特別的是，相關項個數的改變對預測準確度沒有影響。

6.4.3 Y 對 X(顯著相關)迴歸模式

本模式預測準確度相關於觀測值個數(window size)及相關項個數(nalg)，觀察圖 5-1-13、5-1-14、5-3-13、5-3-14 發現，此模式特性與 Y 對 X 迴歸模式 不管是保費收入數列或費用支出數列均非常類似。

吾人也發現費用支出的 MAPE 普遍大於保費收入的 MAPE 值，顯示費用支出數列的變異性高於保費收入數列。因篇幅有限，MAPE 值僅將具有代表性的列出。

保費收入部分：

type=3，window size=24，nlag=16，MAPE=1.997

type=1，window size=24，nlag=16，MAPE=2.095

type=1，window size=96，nlag=20，MAPE=2.759

費用支出部分：

type=1，window size=96，nlag=22，MAPE=4.300

type=3，window size=96，nlag=12，MAPE=5.086

type=2，window size=108，nlag=24，MAPE=6.835

6.5 自我迴歸整合移動平均(ARIMA)預測模式

本模式預測準確度與觀測值個數(obs count)及 p d q 階次相關，由於組合情形非常複雜，需實際計算才能得知，觀察圖 5-1-15、5-2-13、5-3-15、5-4-13 可知，當觀測值越多，預測準確度越高，MAPE 的變異程度降低，當觀測值越少，預測準確度越低，MAPE 的變異程度升高。當 window size 48 時有最高的 MAPE 值，此現象與前述加權移動平均情形相同，中斷保費收繳是主要原因。

費用支出的 MAPE 普遍大於保費收入的 MAPE 值，顯示費用支出數列的變異性高於保費收入數列。因篇幅有限，MAPE 值僅將具有代表性的列出。

保費收入部分：

pdq=213，window size=24，MAPE=1.95

pdq=213，window size=108，MAPE=2.671

pdq=213，window size=84，MAPE=2.975

費用支出部分：

pdq=213，window size=84，MAPE=5.174

pdq=213，window size=60，MAPE=5.679

pdq=213，window size=24，MAPE=6.315

6.6 季節性自我迴歸整合移動平均(SEASON ARIMA)預測模式

本模式預測準確度與觀測值個數(obs count)及 p d q 、P D Q 階次相關，組合情形比 ARIMA 更複雜，需實際計算才能得知，觀察圖 5-1-16、5-2-14、5-3-16、5-4-14 可知，當觀測值越多，預測準確度越高，MAPE 的變異程度降低，當觀測值越少，預測準確度越低，MAPE 的變異程度升高。當 window size 48 時有最高的 MAPE 值，此現象與前述加權移動平均情形相同，中斷保費收繳是主要原因。

費用支出的 MAPE 普遍大於保費收入的 MAPE 值，顯示費用支出數列的變異性高於保費收入數列。因篇幅有限，MAPE 值僅將具有代表性的列出。

保費收入部分：

pdq=211，PDQ=424，obs size=108，MAPE=2.854

pdq=213，PDQ=424，obs size=96，MAPE=3.072

pdq=213，PDQ=424，obs size=72，MAPE=3.71

費用支出部分：

pdq=112，PDQ=1224，obs size=108，MAPE=5.303

pdq=112，PDQ=1224，obs size=96，MAPE=5.392

pdq=112，PDQ=1224，obs size=84，MAPE=5.415

6.7 研究結果

經過前述資料採礦、運算、結果評估等步驟，再針對六種預測模式綜合分析，觀察圖 5-2-1~5-2-9 及圖 5-4-1~5-4-9，吾人得出研究結果如後：

1. 經由比較北區健保局的保費收入與費用支出兩者的 MAPE 值，吾人發現費用支出數列的 MAPE 值遠大於保費收入的 MAPE 值，顯示費用支出數列本身的變異程度遠大於保費收入數列。

2. 當觀測值個數 N <= 24，經由判讀各種預測模式的 MAPE 值，可知各種預測模式的表現均不差，但隨著觀測值個數變大， ARIMA 模式的 MAPE 值小於其他模式的 MAPE 值，重要的是其 MAPE 的變化區間較其他模式為小，代表其預測值的變異程度比其他模式都小，因此 ARIMA 預測模式的穩定性及準確度均優於其他預測模式。

3. 值得注意的是，window size 48 時，保費收入的加權模式、迴歸模式、ARIMA 模式及季節性 ARIMA 模式都有最高的 MAPE 值，應與中央健保局進行中斷保費收繳作業有關。高峰期集中在 89、90、91 年的部分月份，因此其相對應的 window size 48、60、72、84，均有較高的 MAPE 值。研究結果顯示：屬不確定因素的中斷保費收入，影響了保費收入預測的行為模式，預測準確度的穩定性因而降低。

4. 若單純以各預測模式的 MAPE 值判斷，預測能力的優劣評估為 ：
ARIMA 預測法 > 迴歸預測法 > 季節性 ARIMA 預測法 > 加權移動平均預測法 > 移動平均預測法 > 平均預測法

5. 考慮在相同的實驗環境下（CPU：Pentium IV 2.4G ，RAM：256 DDR），運用本研究的方法，執行六種預測模式所需的磁碟空間及運算時間如表 6-1：

表 6-1 各種預測模式運算時間比較

預測模式	估計磁碟空間（MB）	估計運算時間
平均預測法	0.3	6 秒
移動平均法	0.36	28 秒
加權移動平均法	4.3	2 分 25 秒
迴歸分析法	4.2	2 分 15 秒
ARIMA 分析法	4.1	2 分 20 秒
季節性 ARIMA 分析法	52.5	96 分

觀察表 6-1 可知磁碟空間需求最高者為 52.5MB（季節性 ARIMA 分析法），其他分析模式空間需求均很低，以今日電腦硬體規格而言，足可應付本研究所需的磁碟空間不成問題。

若比較各模式的運算時間，則差異性頗大，最高者為 96 分（季節性 ARIMA 分析法），其他如 加權移動平均法、迴歸分析法、ARIMA 分析法三者均約需 2~3 分鐘。故若以本研究所發展的方法，同時考慮運算時間，並搭配預測能力(MAPE)來做預測效能（efficiency）的比較，則六種預測模式的效能（efficiency）以 ARIMA 分析法最優，迴歸分析法次之，加權移動平均法再次之；而平均及移動平均法雖然運算時間較少，但預測能力與前三者相去甚遠，季節性 ARIMA 分析法則因運算時間太長，導致預測效能與前三者相差甚遠。

七、 結論與未來展望

本研究以六種不同的預測分析模式分別針對保費收入與保險費用支出進行實驗結果顯示，吾人的確可將時間數列預測分析模式，應用在健保局的保費收入及保險費用支出預測的領域當中；而分析顯示，保險費用支出數列的變異性遠高於保費收入數列，顯見醫療費用成長的控制仍是健保局財務收支平衡最大的關鍵。

7.1 結論與建議

雖然目前總額預算制度已實施，但如何嚴加監控醫療費用黑洞、避免醫療資源浪費、提高醫療服務品質，仍是當務之急；而善用適當的時間數列預測分析方法，或可提昇健保局財務收支預測的能力，提供經營管理決策的參考。

由研究實驗得知，若單純考慮預測準確度（比較MPAE值大小），當觀測值個數 N < 24時，建議可依次採用ARIMA預測法、迴歸預測法、加權移動平均預測法、移動平均法等方法；當觀測值個數多於24筆以上時，建議可依次採用 ARIMA預測法、迴歸預測法、季節性 ARIMA 預測法、加權移動平均預測法，其中以 ARIMA預測法的預測準確度最高，預測能力表現最穩定。

但若考量預測效能（同時考慮運算時間長短及比較 MPAE 值大小），建議依次採用 ARIMA 預測法、迴歸預測法、加權移動平均預測法等三種方法。

7.2 未來應用與展望

時間數列預測分析除應用在保費收入與醫療費用支出之預測外，尚可運用於健保局其他數據資料預測之研究，例如重大傷病醫療費用成長預估（如：洗腎病患增加，洗腎費用大幅成長）、總額預算制度的點值預測，或收支餘絀數列預測等。而本研究的實驗資料均來自於中央健康保險局北區分局，故研究的觀點與涵括的範圍，自以北區分局為主，未來希有賢達人士能繼續進行全局性的財務收支預測分析研究，俾能提供使用者更全面、廣泛之應用。