

類神經網路應用於擬定汽車保險費率

Applying Artificial Neural Network to Automobile Insurance Ratemaking

余清祥^{*}

黃泓智^{**}

陳志昌^{***}

Jack C. Yue

Hong-Chih Huang

Chi-Chung Cheng

摘 要

汽車保險是與消費者關係最為密切的財產保險，但或許因為國人對汽車保險的認知不足，至今仍存在不合理現象。例如：近年汽車車體損失險的投保率下降且損失率逐年上升，其原因或可歸咎於現行的保費不見得反映實際的風險，但此有違精算費率精神的現象若持續下去，勢必對汽車保險的財務健全有不良影響。本文採用國內某產險公司 1999 年至 2002 年汽車車體損失保險資料，探討保費收入與理賠支出的關係，希冀在滿足保費均衡的原則下，尋求較小變異數的預測方法，以降低風險。本文考量過去用於產險的最小誤差估計法，以及根據經驗建構模型的類神經網路法，比較這兩種方法何者較能降低分類的誤差與縮小個體的誤差，以期保費收入與理賠支出兩者間有較小的差異。

實證結果顯示，現行國內車體損失險不完全符合保費均衡原則，其間仍存在保險補貼。而在模型配適上，最小誤差估計法計較能改善收支不平衡的現象；而類神經網路法的加減費系統具有較大加減幅度，更能有效區分高低風險群組，降低不同危險群組間的補貼現象，並在跨年度的資料中具有較小的誤差變異。

關鍵詞：汽車車體損失保險、最小誤差估計法、類神經網路

^{*} Professor, Department of Statistics, National Chengchi University
政治大學統計系教授

^{**} Associate Professor, Department of Risk Management and Insurance, National Chengchi University
政治大學風險管理與保險系副教授

^{***} Fu-Ban Insurance Company
富邦產物保險公司

Abstract

Among all casualty insurances, auto insurance is the most common in our daily lives. Despite its popularity, there exists awkward phenomenon in auto insurance business. For example, the insured rate of Automobile Material Damage Insurance is going downward but the loss ratio is climbing upward. By charging corresponding premium based on individual risks, we can attract low risk entrants and collect reasonable premiums from the highly risk groups. To further illustrate the concept, we aim to take Automobile Material Damage Insurance for example, to study the most efficient estimator of the future claim. In this study, we compare Minimum Biased Estimate, a previously used method, and Artificial Neural Network (ANN). The reason for including the ANN is due to the fact that the relationship of loss experience (input) and future claim estimation (output) is similar to how the human brain performs. The comparison is based on achieving the minimum error of classes or individuals, using the data between 1999 and 2002 from an insurance company.

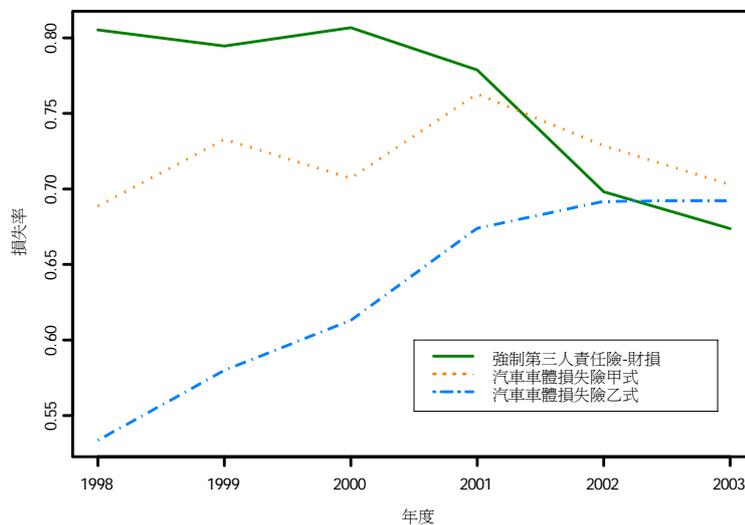
We found that cross subsidization exists in Automobile Material Damage Insurance. In addition, the new rate produced by minimum bias estimate can alleviate the unbalance between the premium and loss. However the ANN classification rating can allocate those premiums more fairly, where 'fairly' means that higher premiums are paid by those insured with greater risk of loss and vice-versa. Also, the ANN is more efficient than the minimum bias estimator in the panel data.

Keyword : Automobile Material Damage Insurance 、 Minimum Biased Estimate 、 Artificial Neural Network

1. 緒論

汽車保險是財產保險中與一般消費者關係較為密切的商品，在台灣的汽車保險始於民國 40 年。雖然汽車保險開始的時間較早，其中也經過不少法令及變革，例如：保費計算從以往的從車主義變革為兼採從車及從人主義，但汽車保險市場仍存在不合理的現象。其中原因或許與汽車保險的商品類型較多，一般民眾不易區分差異外，保險公司計算保費未完全依照保費公平原則，使得保戶與保險公司間過度依賴議價決定保費，也是不能排除的原因之一。

圖 1. 1998~2003 年強制汽車責任與任意車體險損失率比較圖



自民國 90 年起連續兩次調整費率¹，已將一直居高不下的強制汽車責任險損失率，由 80% 大幅降低至 70% (圖 1)。但車體損失險的損失率卻未出現類似下降的現象，乙式車體損失險反而急遽上升，自 1999 年至 2003 年的五年間，損失率上升已逾 15%，在 2003 年時更超過責任險的損失率。由於強制險屬於基本保障及強制購買，較容易依被保險人的風險訂價，也不容易產生逆選擇。車體損失險則較為不同，費率不單單反應被保險人的風險，也需計入被保障的車輛成本，加上在費率過高時，消費者可選擇不購買或暫緩購買，因此費率計算難度較高。以 1998 年至

¹分別為 90.6.19 與 91.12.20 台財保字第 0900750591 號、第 0910751592 號令。

2003 年這六年的甲乙式投保率為例(表 1)，一般認為投保率下降的原因是保費過高。當然，投保率下降不等於保費計算不合理，保險公司需要確定的是計算出來的費率能反映仍然購買保險的客戶，不希望計算費率時為了補貼高風險的被保險人，而使得低風險客戶因為費率不合理而不續保，產生逆選擇的問題。

表 1 1998 至 2003 年自小客貨車甲乙式車體險投保率

年別	小客車	小貨車	自小客貨	投保甲乙式	投保率
	自用(1)	自用	總計	車輛數(2)	
1998	4,433,195	650,592	5,083,787	410,705	8.089%
1999	4,401,730	618,943	5,020,673	391,351	7.795%
2000	4,608,960	643,796	5,252,756	400,318	7.621%
2001	4,720,641	665,718	5,386,359	364,741	6.772%
2002	4,888,050	690,750	5,578,800	364,668	6.537%
2003	5,071,981	717,915	5,789,896	394,016	6.805%

資料來源： 1. 交通部統計 (台閩地區機動車輛登記數)

2. 保險事業發展中心，汽車業務統計年報。

在費率擬訂的過程中，精算人員需要確定現在收取的保費(估計值)能負擔未來的理賠，亦即收支平衡，在統計上稱為不偏(Unbiased)估計量，並且希望這個估計值的變異數愈小愈好。由被保險人的特性來決定不偏估計量的方法很多，一般以線性關係(即以加入的原因決定加成的比例)較易計算及解釋，實務上大多採用線性模式，但實證上較複雜的非線性關係配適效果較佳，因此近年非線性模型的使用比例有增加的趨勢。

為了比較線性與非線性模型何者更適合台灣的车險市場，本文的線性模型選擇過去曾用於美國車險資料的最小誤差估計法，非線性模型則採用近年常見的類神經網路(Artificial Neural Network；簡稱 ANN)，因為保費收入與賠賠支出的關係計算，類似大腦中輸入與輸出的回饋關係。過去其他領域的應用上，類神經網路在求取不偏估計的過程中，多

半能兼顧最小變異數的特性，本研究將驗證在縮小保費收入與賠款支出的前提下，使用類神經網路是否能降低實際觀察值與預測值之間的誤差。

本文的研究目標是探討目前汽車車體損失保險中，保費收入與賠款支出的關係，在滿足不偏性的要求下，尋求降低預測誤差變異數的方法，實證資料將採用國內某產險公司 1999 年至 2002 年的資料。下一節將先整理國內外產險費率的相關研究；第三節介紹本文使用的最小誤差估計法及類神經網路，以及本文對兩模型的修正方法；第四節為實證分析比較；最後一節為討論與結論。

2. 文獻回顧

汽車保險費率釐訂中，通常分析方法可概分為損失金額的分配函數為已知或未知，也就是參數模型或無母數方法。由於損失金額往往會集中在某些數值，且一般為右偏(Right Skewed)，若套上參數模式，因為統計理論的限制，合適的分配模式往往不易獲得，因此本文將假設分配函數未知，以自由分配(Distribution Free)方法考慮費率釐訂。Bailey and Simon (1960)提出的最小偏誤(Minimum Bias)估計法，可算是最早由資料的角度計算車險費率的方法。為了使整體費率達到收支均衡，保費收入與理賠支出的誤差應該愈小愈好，藉由降低別個分類因素間的誤差來達到整體費率的不偏性，因為條件式的個數永遠會比分類因素多一項，因此以疊代(Iteration)的方式求取模型的解析解，做為加減費係數的估計值。

關於分類因素與平均保費之間的關係，除 Almer(1957)提出的乘法模型，以及 Bailey(1960)提出的加法模型外，為了提高分類因素的配適度，DuMouchel(1983)提出混合模式(Hybrid Model)。之後 Brown(1988)以函數型態將分類因素與平均保費的關係一般化，在假設損失金額的分配函數已知的情況下，以最大概似法(Maximum Likelihood)及最小平方方法(Least Squares Estimate)計算各分類係數，文中提到以損失率為解釋變數(Dependent Variable)估計的結果優於損失金額。梁正德(民 82)以國內強制汽車第三人責任保險比較貝里氏(Bailey)與布朗式(Brown)提出的估計方法，結果建議國內在分類費率釐訂時可參考布朗式估計法，與其他國家的研究結果類似 (Anderson et al., 2003)。

在從車因素的保費計算中，若其中含有從人因素影響的成份，則兩項因素將造成雙重懲罰的情形。會發生這樣的現象，導因於最小誤差法的獨立性假設，因此在考慮分類因素間的交叉影響下，有學者提出新的方法來計算費率(例如：Guo, 2003；Dugas et al., 2003)，包括決策樹(Decision Trees)、支持向量機(Support Vector Machines)、以及類神經網路。心理學家 McCulloch(1943)與數學家 Pitts 共同提出了神經元最早的數學模式，Rosenblatt(1958)首先引用感知機(Perceptron)觀念來模擬大腦的感知與學習兩大能力，Minsky(1969)證明感知機無法處理互斥或(Exclusive OR; XOR)問題，直至 Rumelhart(1986)借由加入隱藏層補足了感知機缺點，再度喚起研究風潮。類神經網路於產險領域最初應用在預測產險公司是否失卻清償能力(Brockett et al., 1994)，Brodsky 等人(1999)則以類神經網路預估勞工補償保險的賠案處理時間，Cristina(2002)在個人汽車保險費率擬訂上，比較廣義線性模型、類神經網路、分類與迴歸樹等方法，研究結果肯定類神經網路較不受離群值(Outlier)與遺漏值(Missing Value)的影響，但可能有過度配適(Overfitting)的危險，存在不易了解與解釋的黑盒子，在簡單又好用的標準下，建議選擇廣義線性模型。Dugas(2003)等人探討統計學習理論在汽車保險費率擬訂上的應用，文中提出廣義線性模型、類神經網路、決策樹與支向機等模型，研究結果類神經網路能反映因素間的複雜關係且具有預測能力。因此，Mano 與 Dugas 兩人對於類神經網路在汽車保險費率釐訂上的應用有著截然不同的看法：Mano 著重在模型的解釋能力；Dugas 則強調模型的預測能力。

依求取損失金額不偏估計量的目標，在滿足自由分配的假設下，本文將比較貝里氏最小偏差估計與類神經網路兩種方法。選擇這兩種方法的原因，貝里氏最小偏差估計是最早符合的研究方法，且為國內現行制度，而類神經網路在最小平方法下，是滿足觀察值與估計值的最佳映射函數。另外，雖然加型與乘型均符合不偏性的要求，但乘型的有效性不如加型(Weisberg,1982)，故採用貝里氏加型模型以做為比較²。另外、採用梁正德(民 82)以的研究建議(貝里氏乘型、布朗式常態分配(乘型)、布朗式最小平方估計(乘型)為最佳)，故採用布朗式最小平方法與貝里氏模

² 在美國加型、乘型或混合式均有保險公司使用，歐洲則以加型居多

型相互比較。

依 Bailey 提出的理論，估計費率與實際損失間之平均誤差需為零，採用最小平方法借由反覆的疊代，找出收斂的參數估計值，雖仍滿足不偏性，但其變異數已非最小。另一方面，採用類神經網路計算加減費係數，則在於將估計費率與實際損失間的誤差，倒向分配予各網路輸入層權重與隱藏層權重，使得估計值與觀察值之間的誤差最小。兩種方法的主要差異在於誤差的來源。貝里氏透過降低分類誤差的估計值來達到整體誤差的最小化；類神經網路則直接縮小個別資料的誤差估計值(表 2)。類神經網路採用的最陡坡降法算是牛頓法(Newton's Method)的特例。

表 2 貝里氏與類神經網路演算式比較表

模型	最小誤差估計	類神經網路模型
誤差函數	$AD = \sum n_{ij}(r_{ij} - \Phi(x_i, y_j))$	$SSE = E(e^2) = E[(t - a)'(t - a)]$
最適化過程	$\Delta x = x_{n+1} - x_n = -\frac{f(x)}{f'(x)}$	$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{i,j}^m}$
方法	牛頓(Newton-Raphson)法	最陡坡降 (gradient steepest descent) 法

3. 實證模型

本文採用最小誤差法及類神經網路二個方法，比較最小化分類誤差或最小化個體誤差二個準則下，何者對整體誤差的降低具有較小變異，本節首先描述資料現況，接著說明本文對研究模型設定條件。最後以單年度的預測誤差及續保年度的收支均衡原則，做為衡量模型的標準。

3.1 資料描述與分析

本文以 1999 年新投保汽車車體損失險甲、乙、丙三式合計，且續保至 2002 年的被保險人(共 9104 筆)為研究對象，因現行賠款記錄係數，係依被保險人過去三年的理賠記錄計算，且甲乙丙三式適用同一分類係數表，故在此三年內加入及退出的被保險人不在研究對象之內。保費收入與賠款支出均以保單年度做為業績歸屬的基礎，限定為己車己開(駕駛

人與被保險人同一人)，以理賠金額計次，整合後的資料記錄如下：

表 3 資料量彙整表

年度	1999	2000		2001		2002	
承保資料量	30.8MB	33.2MB		34.2MB		36.2MB	
理賠資料量	9.15MB	10.3MB		10.7MB		10.8MB	
當年度個人戶 (人數)	39,824	41,814		40,676		50,915	
當年度新投保 個人戶 (人數)	20,973	21,176		23,785		29,195	
1 年續保戶 (人數)		16,252	40.80%	17,934	42.89%	19,769	48.60%
2 年續保戶 (人數)	1999 ~ 2001	11,010	27.64%	2000 ~2002		12,715	30.41%
3 年續保戶 (人數)	1999 年新保戶三年續約至 2002 年					9,104	22.86%
9,104(100%)	三年皆投保甲式					1,263	13.87%
	三年皆投保乙式					2,251	24.72%
	三年皆投保丙式					3,614	39.69%
	三年中混合投保					1,976	21.70%

現行制度下，當年度無理賠記錄計為-1 點，發生 1 次理賠計為 0(不加費)，2 次得 1 點並以此類推，每點加減費幅度 20%。採用投保前三年的損失經驗累計賠款記錄點數，各組內賠款平均數大於中位數(圖 1)，表示理賠金額呈右偏分配，若以平均數做為衡量風險的指標易受離群值影響。依資料分賠款記錄點數在 1 點及 1 點以下的群組，過去三年所收取的保費收入足以支付賠款成本外(例如：1 點的平均保費收入 59,631 大於平均理賠支出 48,478)，賠款記錄點數 2 點以上者，過去三年所收的保費則不足以支付賠款(圖 2)(例：3 點的平均保費收入 66,007 小於平均理賠支出 88,304)。

圖 1. 累計賠款記錄點數理賠金額基礎統計量

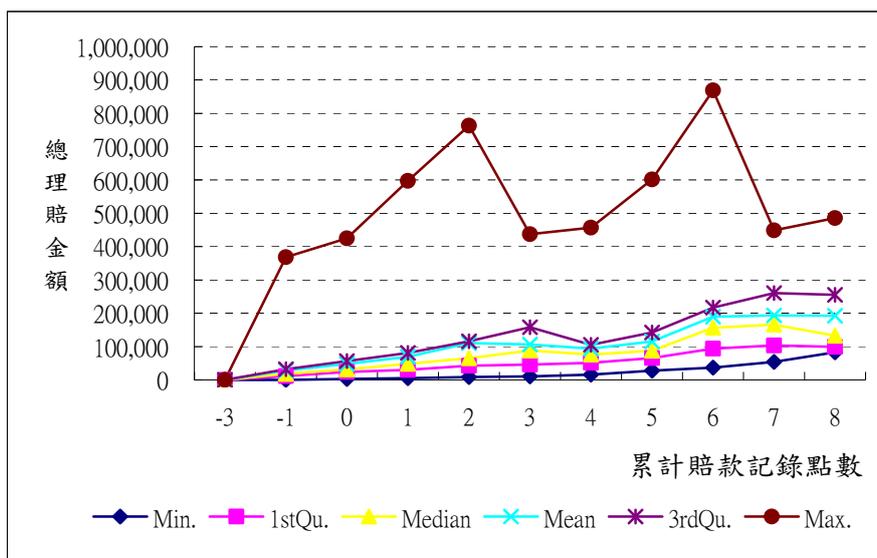
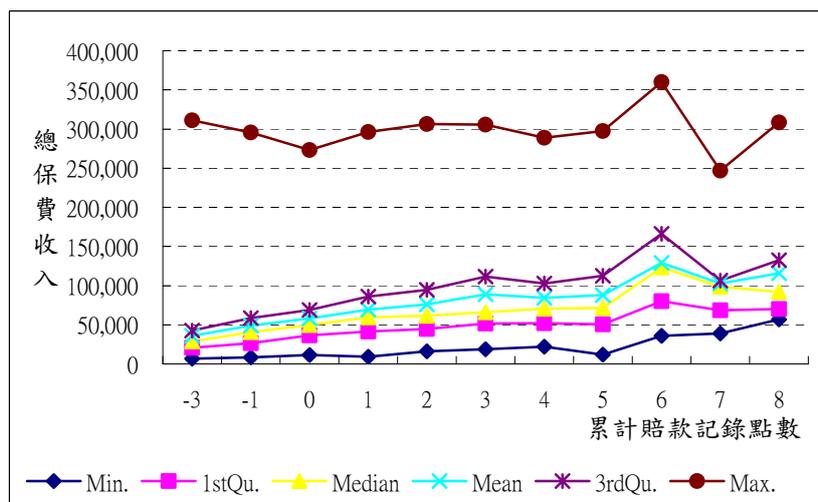
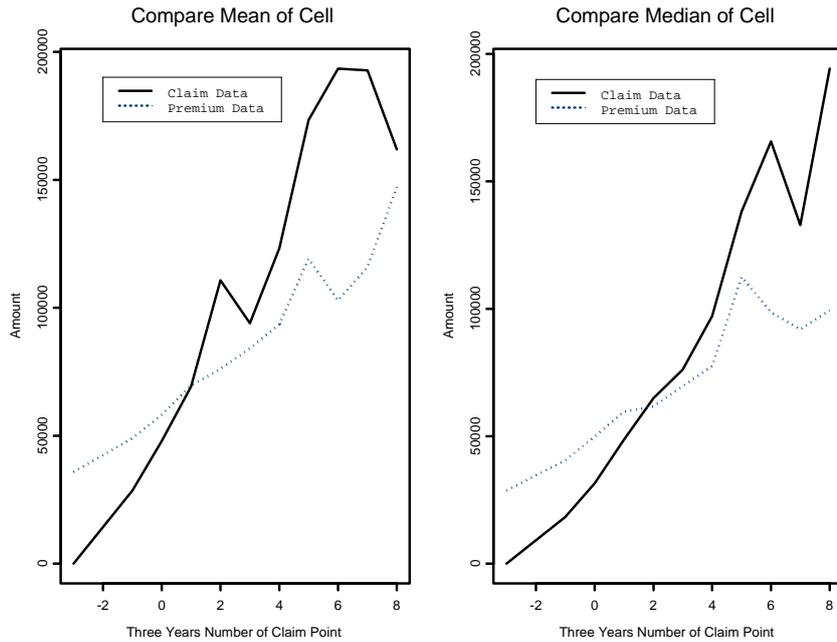


圖 2. 累計賠款記錄點數總保費收入基礎統計量



由於各群組中的賠款金額呈右偏分配(圖 1)，採用中位數取代平均數或可減輕異常值的影響(圖 3)，各群組中的平均賠款與平均保費的差距，隨累計賠款記錄的增加而增加，由於高賠款記錄群組的保費低估，而造成損失率上升的現象。

圖 3. 2002 年投保前三年理賠記錄點數、支出與保費收入比較圖



3.2 卡方最小偏差估計模型設計

3.2.1. 賠款記錄係數

雖然分類類別增加會使得資料過少，但若目標變數隨著某變數呈現線性變化，則可使用迴歸模型，例如：魏長賢(民 83)觀察到純保費與賠款記錄點數呈線性關係。本文以過去三年累計賠款記錄點數所繳付的總保費為自變數，應變數為過去三年所發生的總理賠金額進行迴歸分析，迴歸係數即代表各累計賠款記錄點數群所需調整的比例。

3.2.2. 費率代號係數

由資料分析中發現新車與發照年份四年以上者發生理賠的機率較高，並非為直線關係。目前費率代號係數表依貝里氏乘法模型，按發照年份與費係代號計算各組的最適純保費，並以費率代號第七等級(車價 55 萬至 60 萬)為基準，依線性關係求得最後從車因素費率係數表。為符合資料現況，費率代號係數的損失頻率與損失幅度計算平均損失成本，再依線性調整各代號群係數，再上貝里氏加型模式估計發照年份與性別的分類係數(以民國 90 年為基準)，依此計得當年度發照的各級費率代號係數表。

3.3 類神經網路模型

3.3.1 網路設計

設計網路首先需決定網路架構，一般問題用一層，複雜的問題選用兩層，根據影響的重要性，使用類神經網路前，網路設計的步驟為：首先、以單一隱藏層試行，其次、固定網路學習率為 0.5 選定隱藏層單位元。第三、決定疊代次數。第四、固定隱藏層與疊代次數決定網路學習率。第五、條件不變下衡量是否需再加入隱藏層。最後、依試行結果參考經驗法則以決定網路變數設定。依嘗試錯誤法設定本文類神經網路為單層隱藏層、三個隱藏層單位元、學習率為 0.01、疊代次數 50 次，試行過程如下所述。

(i) 選擇疊代次數

若網路訓練過早停止，則無法達到收斂的最佳狀態。輸入值採現行加減費因素，選用一層隱藏層、固定學習率為 0.5 的條件下，以疊代次數 500 次試行，發現疊代次數 32 次趨於收斂。因此依試行結果隱藏層中選擇三個單位元，並設定疊代次數為 50 次。

(ii) 選擇網路學習率

學習率在類神經網路中扮演著調整權重值大小的角色，功能類似於轉彎時的方向盤，當轉角大時需多打幾圈，轉角小時只需要稍微轉動一下即可，以 0 至 1 間隔 0.001 測試學習率，誤差隨學習率上升而上升(圖 5)，為避免矯枉過正，本文將學習率設定為 0.01。

(iii) 選擇隱藏層數

本文採經驗法則(輸入與輸出層單位元數總和的一半)設定處理單位元。關於隱藏層的設定，簡單的模型由於參數少，對於目標問題可能產生較大的近似誤差，相對的，估計誤差則較小；複雜的模型由於參數較多，近似目標問題的誤差較小，但估計誤差較大，以 1999 年續保資料訓練類神經網路權重值，並以 2000、2001 年續保資料為測試資料集。比較誤差變異數後，顯示複雜的模型不一定表現就比較好，應該針對問題選

擇適合的模型。

3.3.2 類神經網路驗證

為建立網路權重，一般以隨機的方式，抽取的資料的三分之二做為訓練資料集，並以剩餘的三分之一留做測試資料集，以防止網路產生過度配適的現象。關於網路的測試以均方誤差(Mean Square Error)為判斷標準，其定義如下：

$$MSE = \sqrt{\frac{\sum_P^M \sum_j^N (T_j^P - Y_j^P)^2}{M \cdot N}} \quad (5)$$

T_j^P :第 p 個範例的第 j 個輸出單位元的目標輸出值

Y_j^P :第 p 個範例的第 j 個輸出單位元的推論輸出值

M :範例數目 N :輸出層處理單位元數

各輸入單位元對於輸出值的解釋標準，則以相關係數定義如下：

$$\rho = \frac{\left(\sum_P^M \sum_j^N T_j^P \cdot Y_j^P \right) - n \cdot \mu_t \cdot \mu_y}{(n-1) \cdot \delta_t \cdot \delta_y} \quad (6)$$

$\mu_t \delta_t$:目標輸出值的平均數與標準差; y 表推論輸出值

3.3.3 類神經網路輸出值在加減費上的應用

若不對網路輸出值加以修正，以類神經網路估計費率將形同自己保險(Self Insurance)。由於訓練網路之初，按資料逐筆修正權重值，輸出是個別被保險人經驗損失率，而不是在承保團體下的經驗費率，故對於類神經網路在加減費的應用，本文修正定義如下：

$$\text{保險費} = BP \times \{1 + (Y - \mu(Y) \cdot \text{Max}(Y)) \times f\} = BP \times \{[\Delta \text{Loss Ratio} \cdot f]\} \quad (7)$$

BP：基本保費；Y：網路輸出值；Max(Y)：網路輸出最大值

$\mu(Y)$ ：網路輸出值期望值；f：預期損失率

意即個人經驗與平均損失率的差異需考慮分攤比例，例如一被保險人今年的損失經驗高於平均損失率 60%，但因為加入風險集合故只需加

費 $0.6f$ ，而 $0.6X(1-f)$ 的比例則需由自己承擔，借由公式 7 使得類神經網路將風險程度(網路輸出值)轉換為純保費基礎。另外，由於調整費率的因素需與計算保險費所考量的因素一致，故本文在類神經網路的費率調整因素加入性別、年齡、發照年份、費率代號、賠款記錄等分類因素，及保險種類、自負額等純保費因素。

本文使用的方法及相關參數的選擇，詳見表 4 整理。

表 4 模型比較表

項 目 \ 模型	貝里氏最小偏差法 布朗氏最小平方法	類神經網路模型
模型經驗 資料期間	三年 (1999~2002)	一年
最小化誤差準則	以降低分類群組的估計誤差，達到整體估計誤差的不偏性	以降低個體的估計誤差，達到整體估計誤差的不偏性
單一年度 (衡量估計誤差)	由 1999 年投保並續保至 2002 年的經驗資料，計算分類費率，據此估計 1999(2000、2001)各年度的平均理賠金額	以 1999(2000、2001)年的承保資料，預測滿期後的損失率
續保三年度 (衡量收支平衡)	以上述經驗資料計算出的分類費率，配適 1999(2000、2001)	以 1999 年資料訓練模型，預測滿期後(續保一年、二年、三年)的損失率

4. 實證結果

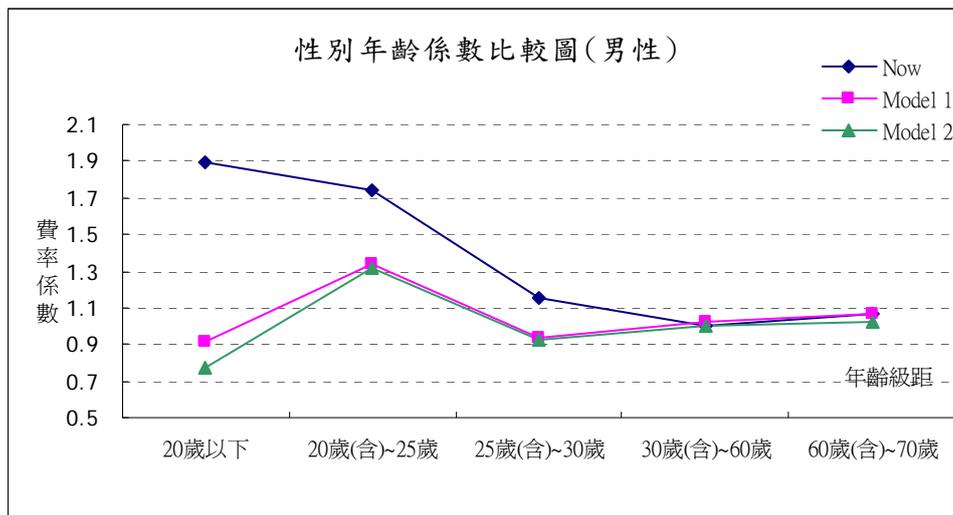
本節使用產險公司的 1999 至 2002 年汽車車體損失保險甲乙丙式資料為依據，以分類費率模型做為對照組，類神經網路做為實驗組進行模型比較。本節首先按現行方法重新推計分類係數(最小偏差法計為模型 1、布朗氏最小平方法計為模型 2)，並與現況進行顯著性檢定。接著進行類神經網路(模型 3)在單一年度的交叉驗證結果。最後比較兩種模型在單一年度的估計誤差變異數，以及在續保年度(滿足收支平衡的條年下)反映不同風險程度的機制。

4.1. 最小誤差分類係數

4.1.1 加減費係數

(i) 年齡性別係數

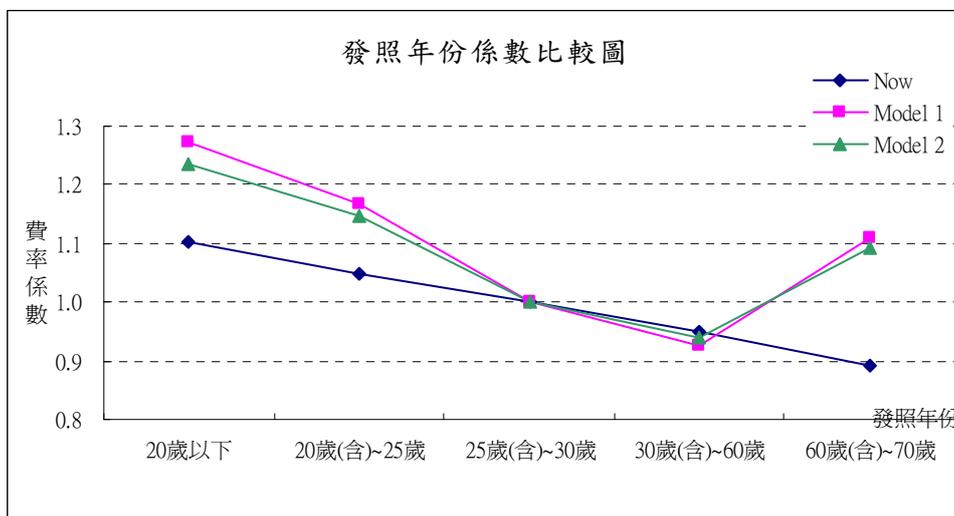
圖 4. 男性年齡係數比較圖



男性第 4 級(30 至 60 歲)為基準，以模型 1 及模型 2 進行疊代，平均絕對差(Average absolute difference: ADD) 0.001 為標準，男性 30 歲以下的年齡係數較現行制度為低(圖 4)，60 歲以下的年齡係數較現行制度高，女性的性別係數為 1.083，而現行女性性別係數為男性的 0.9 倍，因此女性被保險人較不會選擇模型 1 與模型 2。

(ii) 發照年份

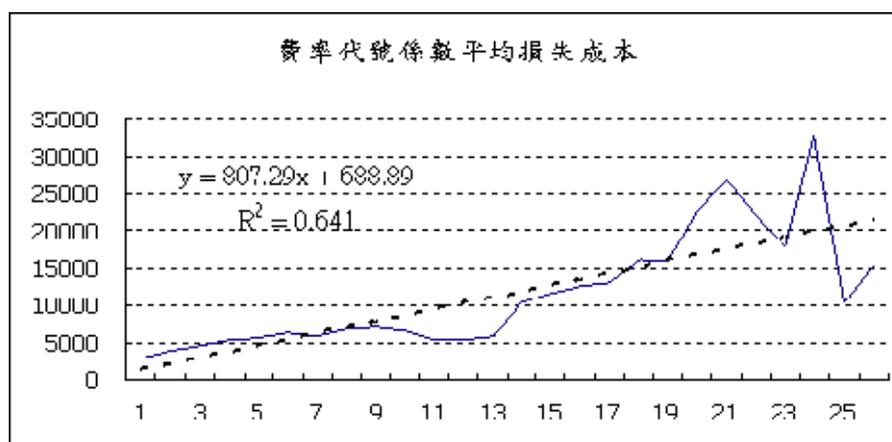
圖 5. 發照年份費率代號係數比較圖



以發照年份及性別的貝里氏加型模式估計，計得發照年份費率係數，新費率能反映資料呈現的資訊(圖 5)，而模型 2 對新車(0-2 年)的費率係數略高於模型 1。

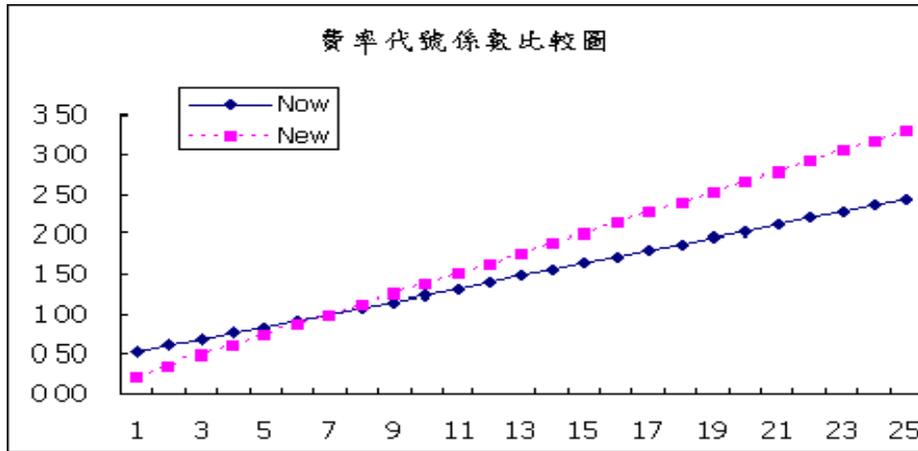
(iii) 車價

圖 6. 2001 費率代號平均損失成本迴歸分析圖



按各車價群組總保費收入為自變數，各群中平均理賠成本為因變數，得到 28 個虛擬變數進行迴歸分析，而後線性調整迴歸係數(判定係數 0.64)，比較費率代號係數(圖 7)，新係數在第 7 車價等級以下較低於現行係數，對於車價等級 7 以下的車主較有吸引力，新係數對於新車及發照年份四年以上者高於現行係數，配合車價代號費率表，於第 7 車價等級以上不論發照年份皆高於現行費率係數。

圖 7. 費率代號係數比較圖



(iv) 賠款記錄係數

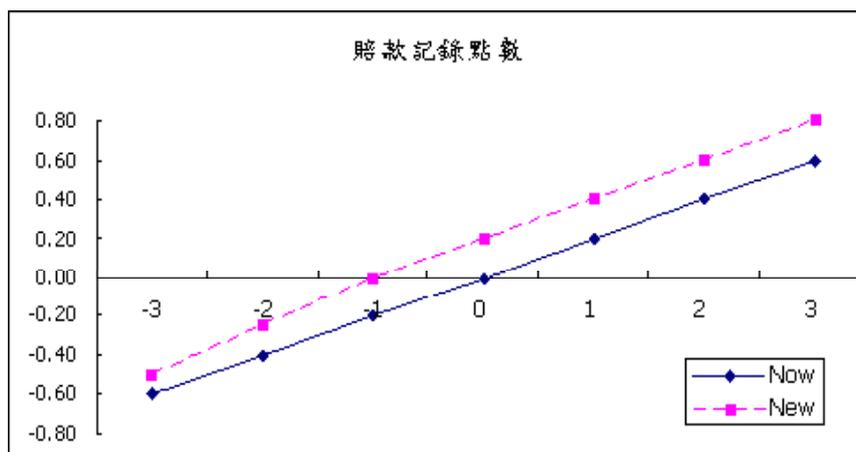
以過去三年累計賠款記錄點數所繳付的總保費為自變數，應變數為過去三年所發生的總理賠金額進行迴歸分析(公式 8)：

$$y = 12990 - 0.24x_{-3} + 0.30x_{-1} + 0.53x_0 + 0.74x_1 + 1.06x_2 + 0.98x_3 + 0.92x_4 + 1.01x_5 + 1.28x_6 + 1.64x_7 + 1.38x_8 \quad (8)$$

$$y = 26199 - 0.49x_{-3} + 0.03x_{-1} + 0.27x_0 + 0.45x_1 + 0.80x_2 + 0.67x_3 + 0.62x_4 + 0.71x_5 + 1.00x_6 + 1.34x_7 + 1.05x_8 \quad (9)$$

判定係數 44.94%，顯示在基本保費 12,990 下，賠款群組-3 點所繳付的保費與賠款的變動為-24%，若以各群組內的平均賠款總金額為應變數式(9)，判定係數 73.52%，基本保費若為 26,199，則 3 年無肇事者所繳付的保費與賠款的變動為-49%，1 年無肇事者為 0%，0 年無肇事者 +27%，比較在現行制度下，3 年無肇事者可減費-60%，1 年無肇事者可減費-20%，肇事記錄 1 點者需加費 20%，是故該公司在現行制度下，需調高各賠款記錄加減費係數+20%，顯示保費有低估的可能。

圖 8. 賠款記錄係數比較圖



3.1.2 顯著性檢定

依據 A 公司經驗資料推計各加減費係數，採無母數 Wilcoxon 符號順序檢定，資料與市場制度的差異是否顯著，觀察前資料(現況)與觀察後資料(推計結果)差異是否顯著？虛無假設：觀察前後差異的中位數=0。

(i) 年齡性別係數

採用無母數檢定的優點在於不需對母體的分配做任何假設，但在母體分配已知的情況下，則有母數方法的效率較佳，本文假設母體分配未知的情況下，以現行年齡性別係數為觀察前資料，比較推計結果，各年度均拒絕觀察前後無差異的假設，以 Z 值推論觀察前的係數顯著地大於觀察後的係數，符合本文推計結果。

表 5. 年齡性別係數檢定結果

年齡性別係數	Wilcoxon signed-rank test	
Alternative hypothesis : mean is not equal to 0		
signed-rank normal statistic with correction Z		
1999	51.1628	p-value = 0
2000	53.0796	p-value = 0
2001	52.1305	p-value = 0

(ii) 費率代號係數

表 6. 費率代號係數檢定結果

費率代號係數	Wilcoxon signed-rank test	
alternative hypothesis : mean is not equal to 0		
Signed-rank normal statistic with correction Z		
1999	-79.3983	p-value = 0
2000	-76.7633	p-value = 0
2001	-80.4955	p-value = 0

發照年份 2002 年為第一年的費率代號係數表做為觀察前資料，比較推計結果，各年度均拒絕觀察前後無差異的假設，以 Z 值推論觀察前的係數顯著地小於觀察後的係數，整體而言，觀察前的費率代號係數，小於觀察後的費率代號係數。

(iii) 賠款記錄係數

表 7. 賠款記錄係數檢定結果

賠款記錄係數	Wilcoxon signed-rank test	
alternative hypothesis : mean is not equal to 0		
signed-rank normal statistic with correction Z		
1999	-93.0915	p-value = 0
2000	-77.7722	p-value = 0
2001	-71.3486	p-value = 0

依公式(9)修正後的費率代號係數，經檢定在各年度中拒絕觀察前後無差異的假設，以 Z 值推論觀察前的係數顯著地小於觀察後的係數，故調高賠款記錄係數的同時，第一節現況分析的結果顯示保費有低估的可能，經過迴歸分析可推估需要調整多少，以 Z 值推論觀察前的係數顯著地小於觀察後的係數。

3.2 類神經網路加減費係數

交叉驗證各年度的類神經網路，(表 11)誤差期望值接近於零的情況下，誤差變異數平均約為 0.02，各輸入值與輸出值的相關性約 0.54，顯示尚有其他具影響力的因素未加入模型中，進一步衡量各變數的重要性，以刪除變數而增加的錯誤率，做為衡量輸入值對模式的影響力，(表

12)，賠款記錄最具影響力，保險種類次之，其餘各因素的重要性差異不大。

表 8. 類神經網路誤差彙整表

Data Set		From 1999	From 2000	From 2001	
		To 2000	To 2001	To 2002	
ANN Model	Train Error	0.05822	0.03229	0.03428	
	Test Error	0.03548	0.03550		
	Predict Error	0.03523			
Model	Data Mean	Data S.D.	Error Mean	Error S.D.	Correlation
1999	0.01997	0.06786	0.00231	0.03247	0.5186
2000	0.02701	0.06304	0.00049	0.01985	0.5807
2001	0.03828	0.07144	0.00144	0.02402	0.6246

表 9. 類神經網路輸入值重要性分析表

加減費因素	性別	年齡	發照年份	賠款記錄	費率代號	保險種類	自負額
錯誤率	0.05662	0.05641	0.05659	0.06533	0.05601	0.05751	0.05624
重要性	3	5	4	1	7	2	6

3.3 模型比較

本文的評估標準有二：對於單一年度，在估計誤差期望值等於零的前提下，預測誤差的變異數應該愈小愈好；對於續保年度，各賠款記錄點數群組的平均損失率的變動應該愈小愈好，代表著保費收入隨著賠款支出的變動而調整。

表 10. 分類係數彙整表

賠款記錄係數：						
-3	-2	-1	0	1	2	3
-0.5	-0.25	0	0.20	0.40	0.60	0.80
發照年份係數：						
0	1	2	3	4		
1.206	1.108	1.000	0.879	1.053		
性別年齡係數：						
	1	2	3	4	5	6
男性	0.915	1.342	0.94	1.000	1.071	1.322
女性	0.992	1.454	1.019	1.109	1.160	1.432

3.3.1 單年度比較

模型 1、2 以 2002 投保前三年為經驗期間，預測 2002 投保滿期後理賠金額，模型 3 以 1999 年為訓練資料集，2000 及 2001 年的資料做測試，模型 2 誤差變異數在現行制度下較模型 1 小(表 14)，模型 3 的預測誤差變異數較模型 1、2 小。由於測試資料集中，能從訓練學出通則的資料佔大多數，使得測試資料的誤差變異反而較，反映於本文的研究對象為續保三年的保戶，意即續保團體內的特徵有趨於穩定的現象。

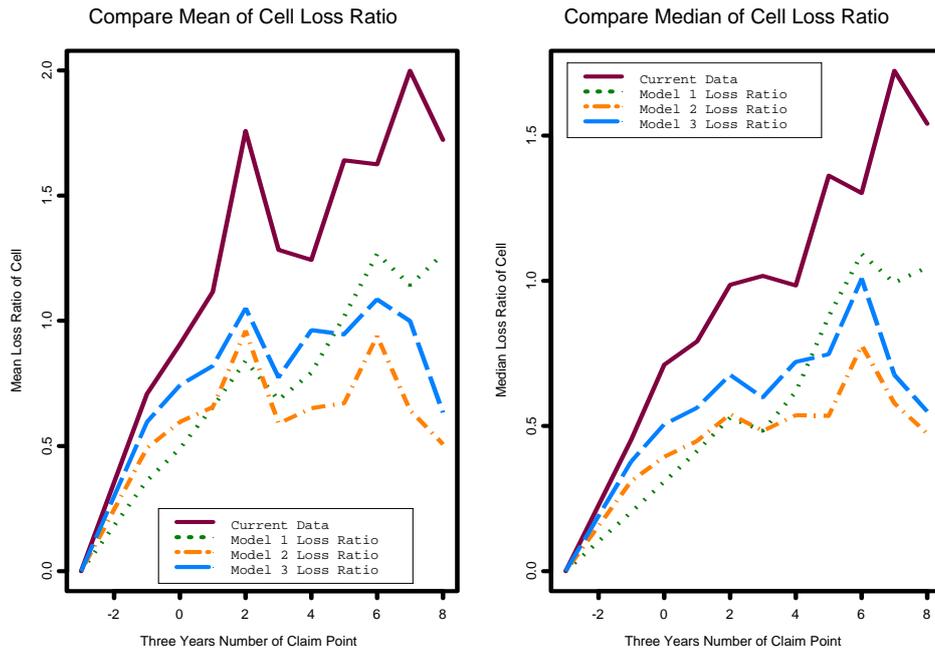
表 11. 預測誤差變異數模型比較表

模型 1	1999 ~ 2002			2002
誤差變異數	0.69065	0.57128	0.62221	0.69948
模型 2	1999 ~ 2002			2002
誤差變異數	0.22149	0.38790	0.41995	0.25777
模型 3	訓練資料集 1999	測試資料集 2000	測試資料集 2001	預測資料集 2002
誤差變異數	0.00336	0.00123	0.00222	0.00190

3.3.2 續保年度比較

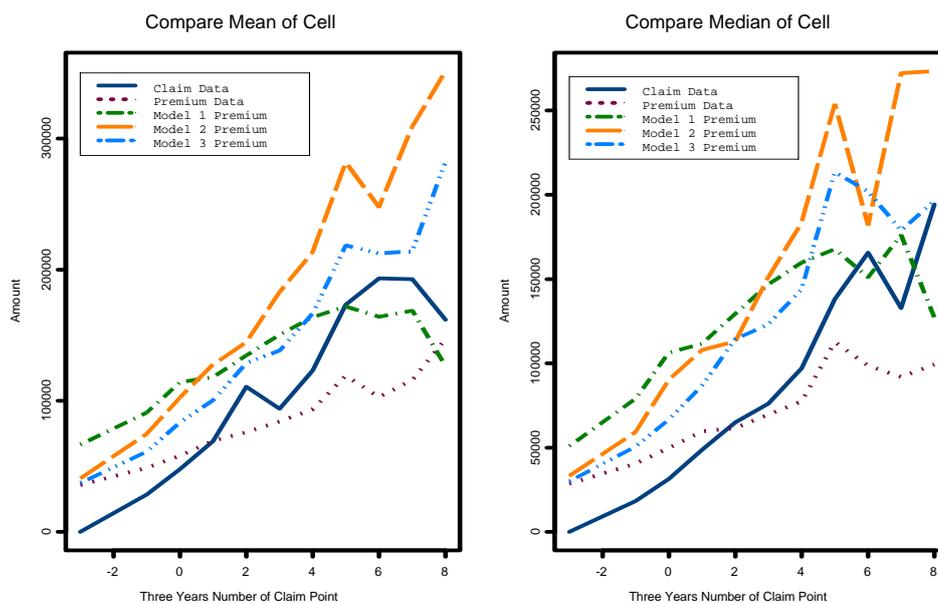
符合精算公平的模型不應只能反映單一年度，對於跨年度的資料也應該要有相當反應能力。對於不同風險分類組間的損失率應該愈平穩愈好，意即保費隨風險程度的大小做調整，以符合保險收支均等原則。

圖 9. 各累計賠款記錄群損失率比較圖



比較 2002 年投保前三年各賠款記錄群組中的平均損失率與損失率中位數，模型 1 相較於模型 2、模型 3 反映風險程度較差(圖 11)，模型 2、模型 3 各賠款記錄群組中的損失率較平穩，另外、比較投保前三年的收支平衡狀況，模型 1 隨著賠款記錄的上升，保費短收的情況愈嚴重(圖 12)，模型 1 對低風險群組減費較少，對於高風險群組加費不足，產生低風險補賠高風險的狀況，模型 2 相對收取較高保費，模型 3 對於對低風險群組減費較多，加費幅度較模型 2 少，且足以支付高風險群組賠款支出，對於賠款支出的配適，模型 3 較佳，綜上所述，模型 3 優於模型 2 優於模型 1。

圖 10. 各累計賠款記錄群收支平衡比較圖



4. 結論與建議

汽車保險在我國產險市場占有舉足輕重的地位，其費率之釐訂自然不可忽視。有鑑於此，本文使用產險公司的 1999 至 2002 年汽車車體損失保險甲乙丙式資料為依據，採用類神經網路與現行最小誤差估計法，比較在降低分類的誤差與縮小個體的誤差，何者較能達到縮小保費收入與賠款支出的差異。

研究結果顯示，車體損失險存在保費補貼，對風險較低的保戶不盡公平，其中依過去三年累計賠款記錄(圖 3)，累積點數在 1 點以下者的保費收入足以支付賠款成本外，累計賠款記錄點數 2 點以上的群組，保費收入則不足以支付賠款，據此在整個承保團體未虧損的情況下，存在低風險補貼高風險的情形。另外，以最小誤差估計法計算的新費率，可以改善收支不平衡的現象，但對於應該減費的低風險保戶，以及應該加費的高風險保戶，以類神經網路推計的加減費系統具有較大加減幅度，因此更能有效的區分高低風險群組，降低不同危險群組間的補貼現象，並在跨年度的資料中具有較小的誤差變異。

以精算公平的觀點，需依據風險程度的高低計收保險費，各群組間的損失率應該愈接近愈好。據此模型 3 與模型 2 較模型 1 平穩(圖 9)，模

型 2 與模型 3 較能依據實際經驗的高低風險程度計收保費(圖 10)，因此高風險者需加費較高，而造成排擠高風險者的刮脂效果(圖 10 過去三年累積理賠次數 5 次以上者)，但由於不考慮損失金額，因此這樣的排擠效果反而有助於抑制小額賠款的發生，達到損害防止及控制損失的目的。因此在不改變現況的情形下，模型 2 優於模型 1。而就收支平衡原則與加減費的比較上，模型 3 減費幅度較大、模型 2 次之、模型 1 最小；模型 2 的加費幅度最大、模型 3 次之、模型 1 加費不足。故整體而言，模型 3 優於模型 2 優於模型 1。

過去認為只要能保持每個各別分類的不偏性，就能達到整體費率的不衡。由文獻探討可得知，貝里氏提出最小誤差估計法的同期(1960)，正是類神經網路被證明只能處理線性問題的受挫期。因此認為從降低個體的估計誤差來最小化整體估計誤差，是不可能也不可行的。

自從類神經網路被加入隱藏層之後，使得降低個體的估計誤差變成可能，應該重新思考最小化整體估計誤差的可行性。依本文研究結果，類神經網路在求取不偏估計的過程中，能兼顧最小變異數的特性，可在滿足收支平衡的條件下，依據風險程度的高低計收保險成本。以下經本文研究分析後之心得與發現提出幾點建議：

(i) 配適其他資料：

由於本文採 1999 至 2002 年三年資料實證，除此外並無其他年份的對照，故應該再持續觀察該公司 2003 年後的車險資料，以檢視本文研究結果，確定類神經網路在分類費率的實證應用；另由於資料採礦的結果，常只適用於採礦的資料，意即單一公司的經驗，為深入了解研究結果的可行性，應以其他公司資料加以配適，探討本文研究結果是否可以推廣至其他公司的損失經驗，以確認研究結果的可行性。

(ii) 檢視現行分類：

現行年齡係數將 30 至 60 歲分為同一級距可能不合乎實際情況。由於現行年齡級距的分類採自民國 85，距今有八年之久，社會的發展與經濟狀況可能已經與當時不同，諸如因高等教育的普及導致進入就業市場的年齡較過去為晚、近年來實施周休二日使得小家庭用車時數增加...等

因素，皆是影響個人對於車輛使用時數與性質的轉變，因此有必要重新檢視現行分類因素的級距以反映現況。

參考文獻

1. 林進田、吳瑞雲，強制汽車責任保險加減費系統之探討，保險專刊，第 48 輯，86 年。
2. 林進田等著，高等產險精算理論與實務，89 年。
3. 梁正德，再談我國汽車保險從人因素係數之釐訂，保險專刊，第 32 輯，82 年。
4. 陳強，「汽車保險費率釐訂精算報告」，台北市產物保險同業公會，84 年。
5. 賴曜賢，實用財產及責任保險費訂定原理，88 年。
6. 魏長賢，「肇事記錄對汽車保險費率影響之探討」，逢甲大學統計與精算研究所碩士論文，83 年。
7. Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., and Thandi, N., 2003, A Practitioner's Guide to Generalized Linear Models, *Casualty Actuarial Society Forum*.
8. Bailey, R. and LeRoy, S., 1960, Two Studies in Automobile Insurance Ratemaking, *Proceedings of the Casualty Actuarial Society*.
9. Brocket, P. and Cooper, W., 1994, A Neural Network Method for Obtaining An Early Warning Of Insurer Insolvency, *Journal of Risk & Insurance*, Vol. 61(3), 402-424.
10. Brown, R., 1988, Minimum Bias with Generalized Linear Models, *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society*, LXXV.
11. Charles, D., et al., 2003, Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking, *Casualty Actuarial Society Forum*.
12. Cristina, M. and Elena, R., 2002, A Discussion of Modeling Techniques for Personal Lines Pricing, *Transactions*, 27th ICA.
13. McClelland, T.L. and Rumbelhart, D.E., 1986, Parallel Distributed

- Processing, *MIT Press and the PDP Research Group*.
14. McCulloch, W.S. and Pitts, W., 1943, A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*.
 15. Minsky, M.L. and Papert, S.A., 1969, Perceptrons, *Combridge MA, MIT Press*.
 16. Rosenblatt, F., 1958, The Perceptron: A Probabilistic Model for Information Storage and organization in the Brain, *Psychological Review*, Vol.65, 386-408.
 17. Brodsky, J., Chudova, D., and Speights, D. B., 1999, Using Neural Networks to Predict Claim Duration in the Presence of Right Censoring and Covariates, *Casualty Actuarial Society Forum*.
 18. Weiberg, H. I. and Tomberlin, T. J., 1982, A Statistical Perspective on Actuarial Methods for Estimating Pure Premiums from Cross-classified Data, *Journal of Risk & Insurance*, Vol. 49(4), 539-563.