

## 演化式多重組合羅吉斯迴歸模型—應用於信用評等

林萍珍

國立高雄應用科技大學金融系

柯博昌

國立高雄應用科技大學資訊管理系

游俊忠

寶碩財務科技股份有限公司

### 摘要

新巴塞爾協定建議銀行採用內部評等法自建授信系統以減少人為錯誤帶來的作業損失，並且能夠快速正確處理授信放款。目前被廣泛應用於信用風險預測模型的是羅吉斯迴歸，此模型雖然可以分成多類，但其S曲線通常以等距或經驗法則切割門檻值做信用分等，當違約機率產生變動時會造成評等等級變動的不對稱現象。因此，本研究提出演化式多重組合羅吉斯迴歸模型（EMCRS），每一個信用評等設計一個羅吉斯迴歸模型進行違約預測以決定評等的等級，希望能提供分類的準確性；模型的門檻值與預測變數是藉由遺傳演算法以非線性方式做最佳化；以此研究模型建立一套演化式多重組合羅吉斯迴歸信用評等系統。預測變數是擷取自財務變數、基本資料變數、會計師變數、總體經濟變數及公司治理變數季資料等五個構面進行預測。另外，其目標函數是以新巴塞爾協定建議的驗證方法對本研究模型進行違約預測效力、評等穩定性以及等級同質性三方向的模型驗證。實驗結果發現，EMCRS信用風險違約評等的預測效力以及等級同質性方面相較於其他四種評等模型（TCRI、區別分析、決策樹、羅吉斯）均有不錯的表現；多季的違約時間點實證中，反應出最近期的財報與公司治理等相關資訊的揭露對模型具有較佳的預測效力；代表穩定性指標的移轉矩陣率會隨著使用者設定的評等級數增加而呈下降的常態現象；具有關鍵因子含財務變數：每股淨值、常續性EPS、每股稅前淨利、總資產週轉率（次），其中常續性EPS被選擇的次數為最多，其次是公司治理變數的董監酬勞佔稅前淨利%，足見EMCRS評等模型受財務變數類與公司治理變數較具影響力。本文的發現對於建構信用評等模型有重要的貢獻與影響。

**關鍵字：**羅吉斯迴歸模型、遺傳演算法、新巴塞爾資本協定、信用評等、違約機率

# Evolutionary Multiple Combinatorial Logistic Regression Model Applied in Credit Rating

Ping-Chen Lin

Department of Finance and Information,  
National Kaohsiung University of Applied Sciences

Po-Chang Ko

Department of Information Management,  
National Kaohsiung University of Applied Sciences

Chun-Chung Yu

APEX International Financial Engineering Res. Tech. Co., Ltd

## Abstract

Some serious financial issues, such as the Asian Financial Crisis and Subprime Mortgage Crisis, have occurred in the last two decades. It is not self-evident that the real economy may suffer from the credit crunches as a result of the financial crises and bank inadequate management. Due to one of banks' major sources of profits being loan growth, especially in enterprise loans, it is important to manage and evaluate corporate financial risk effectively. Published in June 2004, Basel II is a well-known international initiative that requires banks to have a more risk sensitive framework. It establishes regulatory expectations for credit risk through the Internal Ratings Based (IRB) approach, which allows banks to assess key risk drivers as the primary capital calculation. In statics, the logistic regression is only suitable for probabilistic binary classification, but it cannot provide multiple classifications. Although cumulative logic regression (CLR) introduces a multi-class algorithm, it is hard to decide the thresholds in CLR. This paper proposes an evolutionary MCLR credit rating system (EMCRS) that uses an evolutionary approach to optimize multiple combinatorial logistic regression models. We implement GA to estimate non-stationary time-series data with dynamic non-linear searching capabilities. Finally, the EMCRS is verified by (1) capably predicting the default rate (e.g. KS, ROC, CAP), (2) rating stability (e.g. TM), and (3) grade homogeneity (e.g. CIER). The experimental results demonstrate that EMCRS has better competence to predict the enterprise default rate than TEJ. It is reasonable that rating stability will decrease if the number of ratings increases. Profitability, earnings per share, and management factors are critical for evaluating

the performance of EMCRS.

**Key words:** Multiple Combinatorial Logistic Regression Model, Genetic Algorithm, Basel II、Credit Rating、Default Probability

## 壹、緒論

近年來國際間的金融重大事件頻傳，例如亞洲金融風暴與美國次貸風暴等，這些風暴形成的原因與銀行金融機構的授信業務風險管理有著重要的關聯，銀行若資本計提（指依授信戶的信用評等等級給予不同的風險權數，進而計算其應計提的資本）不足，即可能發生重大損失。大部份銀行主要獲利來源為放款業務，而且多數都屬於企業放款，故企業違約風險的管理顯得格外重要。新巴塞爾協定（Basel II）所定義的信用風險是由貸方所產生的一種潛在損失風險；信用風險可經過客觀且系統化的方式轉換成信用等級，使得授信人員可以根據信用等級，快速且正確的制訂放款決策。Shin與Han（2001）指出銀行過去常仰賴授信人員的主觀評估信用評等，易造成人為判斷標準不一，使企業授信風險不能有效管理。

Basel II協定提到銀行若要有效控管信用風險必須系統化成為符合Basel規範的風險管理系統。Basel II推薦的信用風險衡量方法是內部評等法（Internal Rating Based Approach; IRB）。IRB法允許銀行依據本身策略需求，建置較具自主性和彈性的信用風險評等系統，頗獲各國金融機構推崇。在建置IRB系統過程中，銀行需考量違約機率（Probability of Default; PD）、違約損失率（Loss Given Default; LGD）、違約暴險額（Exposure at Default; EAD）以及有效到期期間（Maturity; M）這些因素。其中，PD為IRB的主要核心，故銀行更為重視對PD的評估預測。

決定信用風險評等系統績效的關鍵因素在於建構評等模型以及選擇預測變數，羅吉斯迴歸（Logistic Regression; LR）是目前被普遍應用在建置違約預測系統的統計模型。LR特性允許估計值落於正、負無限大之間，再透過Logit函數轉換成值域0到1的機率值，藉此門檻值協助判斷違約事件發生的程度。LR模型在應用上放寬對變數的假設限制，允許使用質化或量化變數型態；自變數不須符合常態分配或任何機率分配假設。傳統統計模型使用線性組合方式來挑選變數，例如逐步分析、因素分析等。在變數的挑選會以次序加入或剔除的方式來組合，依特定順序選擇變數的方式不易找到最佳預測變數組合。除了變數選擇是建構模型的重點外，模型門檻值的決定也是預測分類模型的重點。門檻值是將風險程度相同的授信戶做歸類的判斷依據，門檻值的決定也將會影響模型預測分類的效力。

本研究目的之一是藉由遺傳演算法挑選出在企業財務體質、公司自治以及總體環境等變數中最具預測效力的因素，提昇預測企業的信用風險等級的準確性。各家銀行的文化和特性不同，授信策略也會有所差異，故我們加入依銀行授信評等等級的設定來建立一套準確、有效的信用風險評等管理資訊系統，每一個類別的評等是個別建立模型與門檻值，以及藉由演化方式找尋最佳預測因果結合，期望能增添分類的效能，此為目的二。採用Basel II建議的違約預測效力驗證方法，分別是「接收者操作特性」（Receiver Operating Characteristic; ROC）曲線、「累積精確度」（Cumulative Accuracy Profiles; CAP）曲線、KS檢定；穩定性指標是「移轉矩陣」（Transition Matrix; TM）；等級區別

同質性是「條件資料熵比率」(Conditional Information Entropy Ratio; CIER)檢定。希望透過這樣的比較及驗證,提昇此模型的評等分類能力,此為目的三。最後,提出一個創新的評等模型,結合多重組合羅吉斯迴歸(Multiple Combinatorial Logistic Regression Model; MCLR)及遺傳演算法(Genetic Algorithm; GA)來建置出演化式多重組合羅吉斯迴歸信用評等系統(Evolutionary MCLR Credit Rating System; EMCRS),盼能有助於銀行管理企業授信風險以及建立符合銀行策略的內部評等系統,此為最重要的目的。

EMCRS系統需要一個穩定性佳且準確度高的分類評等模型,LR的擴充模型累積羅吉斯(Cumulative Logistic Regression; CLR),具有預測次序多類別(ordered multi-category)的特性。但CLR分等的門檻值不易區分,容易隨著樣本資料變化造成等級變動具有不對稱現象。為改善CLR模型的限制,本研究採用LR為基礎擴充為MCLR模型,依據評等等級數目建構個別的CLR評等模型,每一個評等等級有自己的預測模型以及門檻值,組合出MCLR模型。MCLR模型結合GA找出的最佳預測變數組合與各個等級的門檻值,建立一套有效的EMCRS系統,銀行可以依據本身的策略與特性找出適用且有效的企業信用風險評等模型。採用Basel II建議的驗證方法,分別以「違約預測力」、「穩定度」及「等級同質性」三方面做模型的驗證,並且與台灣經濟新報的信用風險指標(Taiwan Corporate Credit Rating Index; TCRI)、區別分析、決策樹、羅吉斯來做比較,希望透過這樣的比較及驗證,提昇此模型的穩定性與預測能力,提供銀行在授信放款業務一個較佳的投資決策參考。

## 貳、文獻回顧與探討

企業發生違約非突發事件,應有跡可循,Basel協定提出之後學術界與實務界對信用風險議題再度引起熱烈的討論。有必要對信用風險相關的議題包括違約預測模型特性、預測變數選擇問題與門檻值決定方法相關文獻進行探討;另外介紹遺傳演算法在信用風險相關文獻的討論,試圖從中發掘相關理論的優勢與限制以及整合應用的契機。

### 一、違約預測模型

根據Balcaen與Ooghe(2006)文獻整理違約預測方法分四大類:第一類,單變量分析(Univariate Analysis),使用單一變數,該變數與預測結果呈線性的關係;第二類,是風險指標模型(Risk Index Models),由Tamari(1966)提出簡單且直覺的分數系統,可包含不同的變數,每個變數都有分數,使用者主觀的分配變數權重彙總後再做預測;第三類,多元區別分析(Multiple Discriminant Analysis; MDA),由Altman(1968)應用MDA於企業倒閉預測,MDA限制使用的變數符合常態分配並互相獨立,變數建構出的線性方程式將產生一個區別分數介於 $-\infty$ 到 $\infty$ 的值域,藉由門檻值來判斷企業是否違約;第四類,條件機率模型(Conditional Probability Models; CPM),Hosmer與Lemeshow(1989)應用CPM發展出LR模型假設其機率呈現羅吉斯分配的機率模型,另假設變數和

違約機率之間呈線性關係的線性機率模型。

屬於條件機率模型的LR是由Ohlson (1980) 將LR應用在企業違約機率預測，並嘗試同時使用類別資料與數值型態的變數資料。Meyer與Pifer (1970) 和Westgaard與Van der Wijst (2001) 以及Tseng與Lin (2005) 等文獻應用LR在銀行倒閉預測，準確率較其他模型為佳。Baesens et al. (2003) 指出迴歸中，LR較能容許多元變數的誤差，其績效表現明顯的較佳於一般迴歸模型。Ugurlu與Aksoy (2006) 探討企業面對環境不確定性和動亂的情況下，比較MDA和LR模型對預測企業違約的效力，實驗結果指出LR模型優於MDA。Ko與Lin (2006) 以LR、NN、MDA預測企業發生財務危機，應用遺傳演算法與粒子群演算法最佳化變數以及各模型預測權重，實驗結果顯示整合模型的預測力最佳。

## 二、預測變數選擇相關文獻

信用風險預測變數包含多種類別，本節依據過去一些學者使用預測變數的類型進行探討做為本研究變數來源的參考。傳統風險預測模型挑選變數，通常是以線性搜尋的方式進行，例如：逐步分析法Donato et al. (1999)、因素分析法Deakin (1972)，其變數組合方式是以次序加入或剔除，較不易找到最佳變數組合。Altman (1968) 是最先使用了財務變數及經濟變數投入多元區別分析MDA模型預測公司倒閉。Ohlson (1980) 是首先利用多元羅吉斯迴歸模型 (Multiple Logistic Regression) 來建立PD預測模型，使用到的變數共有9個財務比率，包含流動比率、總負債是否大於總資產、總資產報酬率、負債比率、淨利變動率、資產規模、營運資金佔總資產比率、是否連續兩年為負淨利、營業活動淨現金流量/平均總負債。Laster (2003) 發表的文章中使用到的財務變數有七類：資本適足、負債財務槓桿、流動性資產 (liquidity)、績效表現、現金流、獲利能力 (profitability) 和規模大小 (size)，其在公司財務預測與評等上，有不錯的表現。Ugurlu與Aksoy (2006) 收集的預測變數有80個分成八類：獲利能力、流動性資產、償付能力 (solvency)、經濟危險程度 (degree of economic distress)、槓桿作用 (leverage)、效率 (efficiency)、波動度 (variability) 以及規模，並利用這八類的變數來有效預測新興市場的公司違約與否。

Keasey與Watson (1987) 以英國破產公司為研究樣本，來探討財務變數、非財務變數與中小企業破產之關聯。非財務變數包括公司治理變數、會計資訊變數、可能窗飾財報之變數及總體經濟變數4大構面。採逐步迴歸選取顯著變數，以Logit迴歸進行預測，發現公司交付財報時間愈不正常、董事會成員愈少、債權受銀行擔保者、最近一次財報受會計師認可且之前財報均未受會計師認可之中小企業發生財務危機機率愈大。Huyghebaert et al. (2000) 利用現金流量變數，如營運活動之現金流量、投資活動之現金流量及融資活動之現金流量、營運資金現金流量、公司規模大小取對數及產業變數等，透過Logit模型來評比現金流量模型與一般財務指標模型之預測能力。結果顯示，現金流量模型預測能力優於一般財務指標模型。

綜合各國內外學者的研究成果，可以發現變數選取攸關模型的解釋能力，包含財務變數及非財務變數的綜合模型解釋能力大多數優於僅考慮財務變數的財務變數模型，羅

吉斯迴歸模型對於預測企業財務危機，具有相當的效力。變數的選擇方法多是線性方法，如逐步迴歸或經驗法則等方法。若應用非線性方法做變數的最佳化或許可以獲得更好的預測結果。

### 三、門檻值相關文獻

Barniv et al. (2000) 文中強調門檻值是構成預測模型效力主要因素。門檻值為預測目標分類依據，若門檻值設定不恰當，容易提高錯誤率導致模型預測力下降。在巴塞爾銀行監管委員會 (The Basel Committee on Banking Supervision 2005) 中提到收受者操作特性 (Receiver Operating Characteristic; ROC) 驗證方法，即是針對模型門檻值區分預測效力進行量化的計算，ROC值愈高代表模型愈佳。Fu與Simpson (2002) 在累積羅吉斯分類以等距分等方式決定門檻，此方法採用均分法分配多種評等等級的門檻值，在S曲線的中間段的門檻值範圍較小而陡峭；曲線的頂端與底端的門檻值範圍較大而平滑，因為門檻值大小不一致較容易影響評等效力。Foreman (2003) 則是使用LR模型以一個常數或機率的平均值0.5來當門檻值，但是此種分法只適用於單一類別並不適用於多類別評等問題。由此看來，門檻值的決定對於信用評等模型的優劣佔有重要地位。

### 四、遺傳演算法

遺傳演算法乃Holland (1975) 所提出，它是人工智慧中有力 (powerful) 的一種技術，其本質上為一個機率性的演算法則 (probabilistic algorithm)，能在廣大的求解空間中，快速的搜尋最適解。GA以非線性搜尋最適預測變數集合應用上，具有良好的表現。Back et al. (1996) 以GA最佳化變數後再應用類神經網路 (Neural Network; NN)、MDA、LR等模型進行企業倒閉預測。Kim與Han (2003) 以資料探勘 (Data Mining) 的觀念探討專家經驗與GA所決定的變數配合的MDA、LR、NN以及決策樹 (Decision Tree) 進行倒閉預測，實驗結果顯示以GA所挑選的變數預測企業倒閉的效力較專家經驗為佳。Min et al. (2006) 和Wu et al. (2007) 結合GA與支援向量機 (Support Vector Machine; SVM) 做企業違約預測，研究結果指出藉由GA最佳化SVM參數後提昇預測的績效。

Shin與Lee (2002) 和Laitinen與Laitinen (2000) 兩篇文章均指出，雖然很多研究的實驗結果報導NN (或SVM) 的績效優於其他模型，但應用NN模型的缺點歸納為：第一，尋找變數與函數之間的非線性關係時容易產生過度配適 (overfitting) 問題；第二，藉由不同參數設定 (如神經元網路架構、學習函數、訓練函數等) 找尋合適的類神經網路成為一種藝術，意指以試誤法 (try and error) 設定相關參數以找尋類神經網路模型；第三，NN學習後的網路結構不易讓使用者了解和使用，對其結構規則可讀性不佳而產生黑箱 (black boxes) 作業的刻板印象；第四，其穩定性不足以及驗證困難使得實務界接受度不高。

綜合前述文獻的討論，本研究選擇以LR模型為基礎，並針對LR模型應用在分類評等

的限制做進一步的改善。改善的方向有：（1）提出多重組合羅吉斯迴歸，改善單一羅吉斯迴歸模型在同一條S曲線上決定門檻值的限制，不論依均等法或經驗法則在同一條曲線上決定門檻值均有不對稱性的問題；（2）不同的評等等級應該有各自的門檻值，亦即每一個等級可以擁有各自的羅吉斯迴歸與其所屬的門檻值；（3）結合GA最佳化的能力，協助LR挑選變數以及決定門檻值；（4）採用Basel II驗證方式，除預測效力外，進一步加入模型穩定性與同質性的證明，強化評等的效能。

## 參、研究方法

本研究提出演化式多重組合羅吉斯迴歸信用評等系統，本節即針對其研究架構的運作步驟進行說明；文中將闡述遺傳演算法的兩項重要工作為染色體編碼（包含變數與門檻值編碼）與適應函數設計；細部說明多重組合羅吉斯迴歸模型計算流程以及Basel II五種驗證方法。

### 一、研究架構

本研究提出EMCRS的研究架構說明共分成5個步驟見【圖1】：第一步驟，透過GA的染色體編碼方式進行第一代族群的初始化。第二步驟，此步驟分成5個動作：（1）每條染色體會依據兩部分進行解碼動作，包含預測變數及等級門檻值；（2）將解碼後的門檻值排序完即決定各等級順序；（3）代入訓練樣本及等級門檻值並利用最大概似法（MLE）來估計MCLR模型，即可求出違約機率，與各機率門檻比較，就可計算出該樣本機率歸類為何種信用風險等級；（4）評等結果再由Basel II的驗證方法（預測效力、穩定性以及等級同質性等）進行驗證，將不同的驗證方法產生的結果做彙總；（5）將驗證結果及變數數量分數（愈少愈高分）取倒數，即為染色體的適應函數值。第三步驟，進入選擇機制，此階段會依據適應函數值，依指定的選擇方法（如輪盤法或競爭法）選擇適應函數值較佳的染色體到交配池進行交配；第四步驟，以交配池裡的染色體，利用兩兩交配方式（如單點交配、均交配），進行部份基因的交換；第五步驟，每條染色體會依據突變率來決定染色體基因是否要進行突變。經由前面的步驟產生新一代的族群，並不斷重覆這樣的流程直到滿足設定的終止條件。EMCRS系統就是透過這樣的演化機制，來尋找最佳的評等估計模型、預測變數組合以及適當的門檻值，學習出最具驗證效力的信用風險模型。

在【圖2】中我們以迴歸殘差值 $\varepsilon$ 來代表樣本在MCLR所得到的預測值， $\varepsilon$ 的範圍是從負無窮大到無窮大，MCLR是由多個LR所互相累積而成，不同的 $\varepsilon$ 會對應不同的違約機率，當樣本預測出來的 $\varepsilon$ 落入二個門檻值 $\varepsilon_1^c$ 和 $\varepsilon_2^c$ 之間，則樣本的預測信用評等為第2級，而等級愈高則違約風險愈高，本研究模型會依據使用者的級別需求來產生適當並能有效區分企業的等級門檻值。在【圖2】是以評等四級為例。

MCLR不同於CLR的地方，在於CLR不易在累積機率分配S曲線上，區分多個門檻，

因為S曲線特性是從0到1，傳統採用均分法切割不同評等的門檻值，造成中間的線段BC是急劇上升如【圖3】。BC段在設定多個門檻值的評等時，因其違約機率值會過於接近，模型的預測規則容易受到樣本機率變動，使得預測樣本等級結果較敏感，造成模型的預測結果較不穩定，而線段AB和線段CD則較平緩，落入這兩線段的門檻，因其違約機率值間隔較大，樣本較容易被評等為該區段的等級。

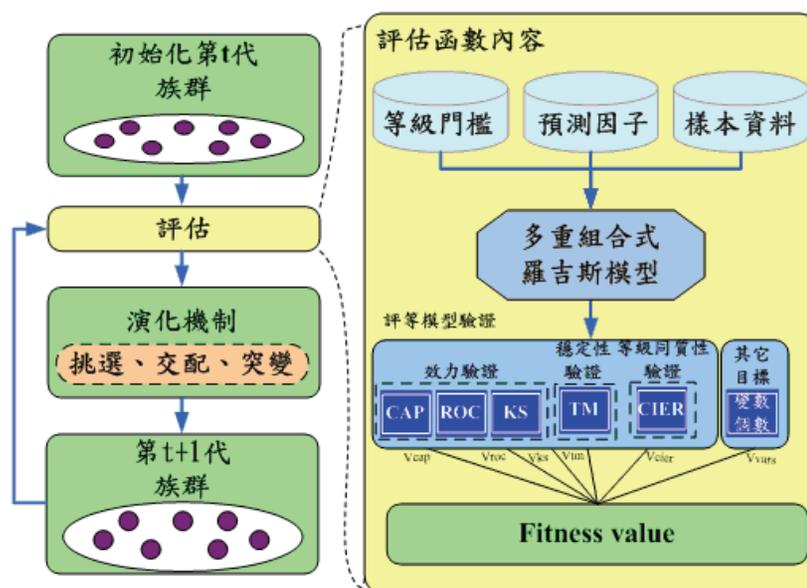


圖1：演化式多重組合羅吉斯迴歸信用評等系統（EMCRS）架構圖

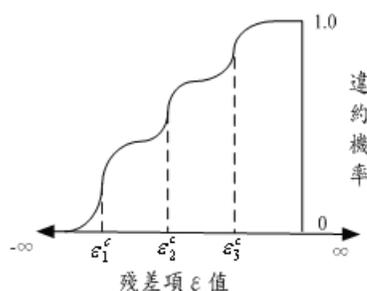


圖2：MCLR圖

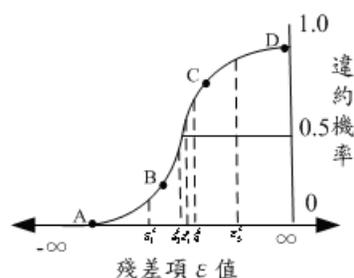


圖3：MCLR圖

## 二、染色體編碼

在【圖4】我們看到GA中染色體長度可分為二部分來解釋：（1）預測變數：由於本研究參考許多的文獻來挑選不同面向的各種變數，假設共有n個原始變數，因此在代表變數的部分是n bit，bit值為0則指該變數未被選擇，1即變數被選擇；（2）區分不同違約風險等級門檻值，EMCRS系統可依照使用者希望的分等數量如m等，來決定MCLR迴歸

的分等門檻數量，每個門檻值是以11 bit來表示（1 bit代表正負值、4 bit代表整數部分、6 bit代表小數部分），門檻值編碼共有  $(m-1) \times 11$  個bit，故染色體長度共有  $(m-1) \times 11$  個bit。

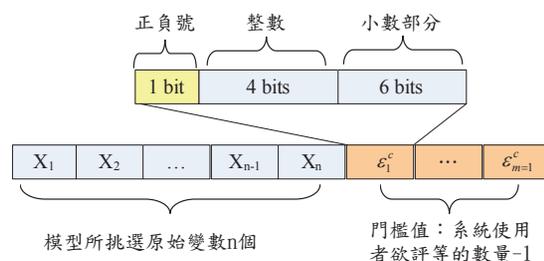


圖4：染色體編碼示意圖

染色體的編碼門檻值的部分，解碼之後會將門檻值進行由小到大的排序，即代表評等低到高（1...m）的門檻值，評等等級愈高，代表違約機率愈大。排序過程可見【圖5】示意圖。

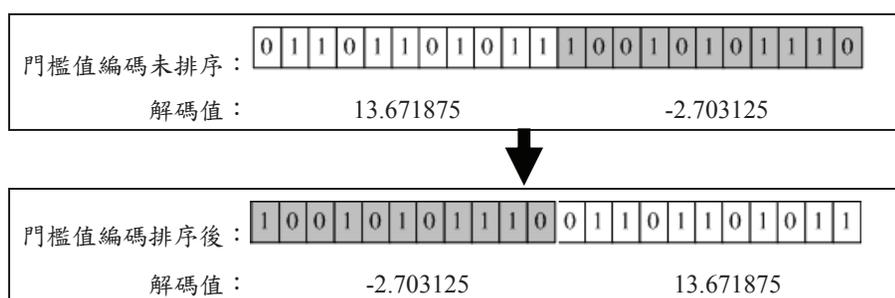


圖4：門檻值排序示意圖

### 三、多重組合羅吉斯迴歸模型

本研究是利用MCLR模型來預測企業信用風險等級，MCLR的累積機率分配曲線，是由多個LR間層層累積組合方式形成。

在LR方法中，首先假設自變數與因變數為線性關係如方程式（1）：

$$y_j^* = \beta_0 + \sum_{k=1}^n \beta_k x_{jk} + \varepsilon_j \quad (1)$$

當連續變數  $y_j^*$ ，其歸屬函數如方程式（ $y_j = 1$ ）（2）：

$$\begin{cases} y_j^* > 0, & y_j = 1 \\ y_j^* \leq 0, & y_j = 0 \end{cases} \quad (2)$$

可從 $P(y_j = 1|x_j)$ 推導得到MCLR中第 $m-1$ 個LR迴歸估計值 $P(y_j = m|x_j)$ ， $\varepsilon^c$ 表示MCLR中各LR的門檻值，其推導過程如方程式(3)：

$$\begin{aligned} P(y_j = m|x_j) &= P\left(\beta_0 + \sum_{k=1}^n \beta_k x_{jk} + \varepsilon_j > 0|x_j\right) \\ &= P\left(\varepsilon_j > -(\beta_0 + \sum_{k=1}^n \beta_k x_{jk})|x_j\right) \\ &= P\left(\varepsilon_j - \varepsilon_{m-1}^c > -(\beta_0 + \sum_{k=1}^n \beta_k x_{jk}) - \varepsilon_{m-1}^c|x_j\right) \end{aligned} \quad (3)$$

假設 $\varepsilon_j$ 的機率分佈是互相對稱，因此方程式(3)可以改寫成方程式(4)：

$$\begin{aligned} P(y_j = m|x_j) &= P\left(\varepsilon_j - \varepsilon_{m-1}^c < (\beta_0 + \sum_{k=1}^n \beta_k x_{jk}) + \varepsilon_{m-1}^c|x_j\right) \\ &= P\left(\varepsilon_j < (\beta_0 + \sum_{k=1}^n \beta_k x_{jk}) + 2\varepsilon_{m-1}^c|x_j\right) \end{aligned} \quad (4)$$

MCLR的累積違約機率計算方式為：假設MCLR欲分10等則需分成9個LR來組合，每個LR各佔有全部違約機率的 $\frac{1}{9}$ ，方程式(5)中 $m$ 是LR的等級數量， $\varepsilon_1^c < \varepsilon_2^c < \dots < \varepsilon_{m-1}^c$ ，每個 $\varepsilon^c$ 的位置是透過以GA找出最佳分界點，以提高違約機率分等的效力，則違約機率累積分佈函數如方程式(5)：

$$\begin{aligned} P(y_j = D|x_j) &= \frac{1}{m-1} \sum_{z=1}^{m-1} P(y_j = z|x_j) \\ &= \frac{1}{m-1} \sum_{z=1}^{m-1} \frac{1}{1 + e^{-\left(\beta_0 + \sum_{k=1}^n \beta_k x_{jk} + 2\varepsilon_z^c\right)}} \end{aligned} \quad (5)$$

為了推估方程式中各個參數 $\beta$ 值，我們利用最大概似法依據給定的樣本資料使概似函數最大化，它們的聯合機率函數可表示為各邊際機率函數的乘積見方程式(6)：

$$\ln L(\theta) = \ln\left(\prod_{n1} P_1 \prod_{n2} P_2 \prod_{n3} P_3 \prod_{n4} P_4 \dots \prod_{nm} P_m\right) \quad (6)$$

本為求出此事件機率最大值發生處，將已知之樣本資料代入 $L(\theta)$ 中，為計算方便，故先對 $L(\theta)$ 取對數可得其對數概似值見方程式(7)：

$$L(\theta) = \ln\left(\sum_{n1} P_1 + \sum_{n2} P_2 + \sum_{n3} P_3 + \dots + \sum_{nm} P_m\right) \quad (7)$$

為使 $L(\theta)$ 最大之估計值 $\beta$ 值，分別對方程式(7)式中 $\beta_0 \sim \beta_n$ 值做偏導數，並令其為0，則會產生 $n+1$ 條非線性方程式。在求出 $\beta_0 \sim \beta_n$ 之前，首先以泰勒展開式(Taylor's Expansion)將此 $n+1$ 條的非線性方程式轉成多項式，假設 $\chi$ 已知， $\Delta\chi$ 未知，則如方程式(8)：

$$f(\chi + \Delta\chi) = f(\chi) + \frac{f'(\chi)}{1!} \times (\Delta\chi) + \frac{f''(\chi)}{2!} \times (\Delta\chi)^2 + \dots \quad (8)$$

如果變動量 $\Delta\chi$ 趨近於0，則可只擷取前兩項，其餘省略，此即為牛頓法（Newton's Method）。為了求得其解，則其推導過程可見方程式（9）－（12）。

$$g = \nabla f = \begin{bmatrix} \frac{\partial f}{\partial \beta_0} \\ \frac{\partial f}{\partial \beta_1} \\ \vdots \\ \frac{\partial f}{\partial \beta_n} \end{bmatrix}, \beta = \begin{bmatrix} \beta_n \\ \beta_n \\ \vdots \\ \beta_n \end{bmatrix}, \Delta\beta = \begin{bmatrix} \Delta\beta_0 \\ \Delta\beta_1 \\ \vdots \\ \Delta\beta_n \end{bmatrix}, H = \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_0^2} & \frac{\partial^2 f}{\partial \beta_0 \partial \beta_1} & \dots & \frac{\partial^2 f}{\partial \beta_0 \partial \beta_n} \\ \frac{\partial^2 f}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 f}{\partial \beta_1^2} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \beta_0 \partial \beta_n} & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_n} & \dots & \frac{\partial^2 f}{\partial \beta_n^2} \end{bmatrix} \quad (9)$$

$$\Delta f = g^T \times \Delta\beta + \frac{1}{2} (\Delta\beta^T \times H \times \Delta\beta) \quad (10)$$

$$g + H \times \Delta\beta = 0 \quad (11)$$

$$\Delta\beta = -H^{-1} \times g \quad (12)$$

透過上列方程式取得 $\beta_0 \sim \beta_n$ 的值，再利用前面遺傳演算法得到 $\varepsilon_1^c \sim \varepsilon_{n-1}^c$ 和自變數 $x_j$ 代回（5）式即可以求出樣本j的違約機率 $P(y_j = D | x_j)$ ，而違約機率再與各個機率門檻比較，可得知樣本機率歸類為何種信用風險等級。

#### 四、驗證方法

本小節是介紹五種Basel II的驗證方法：（1）CAP曲線：用來評估模型的違約預測力；（2）KS檢定：檢定模型是否將違約和非違約有效的區分；（3）ROC曲線：檢定模型評等門檻的設定，是否能有效判斷樣本的等級；（4）CIER：檢定評等結果，是否具有等級同質性以及等級間差異性；（5）TM：了解模型的規則是否具有穩定性。以下將說明這5個指標的用法。

##### （一）CAP

CAP曲線為違約預測力之評估方法，其最終可求得一個準確值（Accuracy Ratio; AR），AR值愈接近1，則表示違約預測力越具有可信度。計算方法是將樣本的PD由高至低作排序（即由高違約的客戶至低違約的客戶），再根據累積PD繪出CAP曲線。假設現在樣本有300筆資料，總樣本的累積樣本比率為S，有D筆違約資料，N筆未違約資料，將它做PD的排序，由高至低，X軸為總樣本的累積比率，Y軸為違約樣本的累積比率。依據每個等級的樣本依序加入，並將累積至各等級總樣本的累積比率及違約樣本的累積比率繪成曲線見【圖6】。

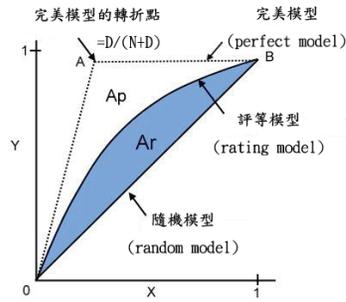


圖6：CAP圖

資料來源：Basel Committee on Banking Supervision (2005)

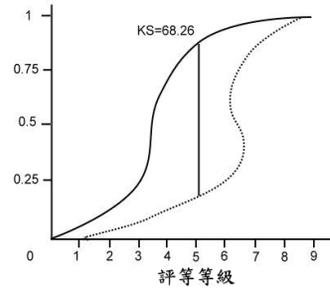


圖7：KS圖

資料來源：Basel Committee on Banking Supervision (2005)

CAP圖是由完美模型 (perfect model)、評等模型 (rating model)，以及隨機模型 (random model)。完美模型表示高違約機率客戶可以在X軸的前端將違約客戶全部捕捉到，代表模型的偵測能力最完美；評等模型所繪出的CAP曲線越接近完美模型曲線，則代表模型偵測能力越好；隨機模型的預測結果幾乎是隨機的，也就表示這個模型不具有偵測能力。CAP中重要的指標為AR，其計算方式就是透過上圖中Ar的面積佔有三角形OAB (Ap) 的面積之比率。N等於樣本大小，AR值越接近1，則表示模型偵測能力越好。Ar的計算見方程式 (13)。

$$AR = \frac{Ar}{Ap} \tag{13}$$

(二) KS

KS檢定的功用是檢定兩個連續型隨機變數是否具有相同的機率分配，其模型的檢定有三個步驟：(1) 將樣本的違約數與非違約數依據評分等級做累積加總；(2) 計算各評分等級的累積機率之差；(3) 找出最大累積機率之差，即為KS值。【圖7】的X軸為分級等級，Y軸為累積 (非) 違約佔總 (非) 違約之比率。其中，實線為非違約累積曲線，虛線為違約累積曲線，兩條線之間的最大差距為KS值。Mays (2001) 曾針對各種情況下的KS值，進一步說明其與模型鑑別能力之關係如【表1】。

表1：KS值表

KS值	模型的優劣程度
<20	模型不具鑑別能力
20~40	模型鑑別能力普通
41-50	模型鑑別能力好
51-60	模型鑑別能力非常好
61-75	模型鑑別能力優
>75	模型鑑別能力太好，應注意是否真實

### (三) ROC

ROC曲線是透過違約率高的客戶給予高的PD，以及給較低的得分；而非違約的客戶，則給予低的PD，並給較高的得分。依序將各等級的違約人數繪製一條違約客戶的等級分配曲線；另外將各等級的非違約人數繪製一條非違約客戶的等級分配曲線。這兩條曲線可得到違約與非違約客戶的等級分配圖見【圖8】：

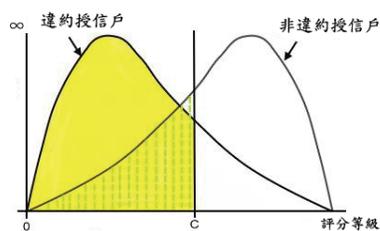


圖8：ROC等級分配圖

資料來源：Basel Committee on Banking Supervision (2005)

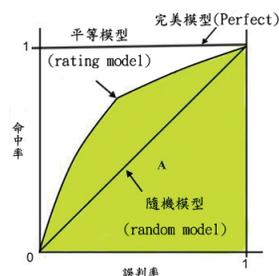


圖9：ROC圖

資料來源：Basel Committee on Banking Supervision (2005)

ROC做等級評分的目標是將違約與非違約兩群客戶完全的分開，兩群重疊部份愈少，模型預測愈準。【圖8】中C點的左邊是預測違約客戶群，右邊則是非違約的客戶群。當違約客戶小於門檻值C的部分（著色區域），表示那一部分違約客戶的分類是正確的；而非違約客戶小於門檻值C的部分（斑點區域），則表示那部分的非違約客戶被誤判了。繪製ROC曲線時，需要依據各級的門檻值C來計算出各門檻值的誤判率以及命中率。以X軸為誤判率，Y軸為命中率，繪製成ROC曲線。命中率HR(C)的計算如下方程式(14)：

$$HR(C) = \frac{H(C)}{N_D} \quad (14)$$

其中H(C)為在門檻值C下，真正被預測出來的違約客戶（著色區域）， $N_D$ 為樣本的違約樣本總數。誤判率FAR(C)的計算如方程式(15)：

$$FAR(C) = \frac{F(C)}{N_{ND}} \quad (15)$$

其中F(C)為在門檻值C下被誤判的非違約客戶（斑點區域）， $N_{ND}$ 為樣本的非違約樣本總數。依據各門檻值C的命中率及誤判率，可繪製出ROC曲線如【圖9】。計算AUC (Area Under Curve) 值，為【圖9】著色面積A÷正方形面積。完美模型的AUC值為100%，而沒有識別能力的隨機模型其AUC值為50%，故AUC值越高，則表示偵測違約的能力越好，AUC值則代表ROC檢定的結果。

### (四) CIER

CIER是用來評估評等模型的分等結果，是否具有等級同質性以及等級間差異性，首先會需要透過熵 (entropy) 的計算，以求得CIER值。我們定義發生率為p的信息熵H

(p) 為方程式 (16)。

$$H(p) = -(p \times \log(p) + (1-p) \times \log(1-p)) \quad (16)$$

可用下圖來說明信息熵H(p)跟發生率p的關係：

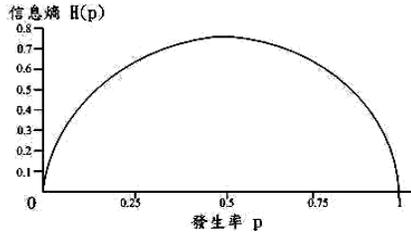


圖10：信息熵 H(p) 跟發生率p的關係

資料來源：Basel Committee on Banking Supervision (2005)

H(p) 若在p為50%的時候會達到最高點，則表示不確定程度處在最大的狀態。若是p等於0或1，則表示事件一定會發生。將所有的客戶依據其違約機率分成m等， $K=\{K_1, K_2, \dots, K_m\}$ ，若將D定義成違約客戶。ce(K<sub>i</sub>) 是以K<sub>i</sub>為前提下，評估不確定性的條件熵 (conditional entropy)，計算如方程式 (17)

$$ce(K_i) = -\{p(D|K_i) \times \log [p(D|K_i)] + [1 - p(D|K_i)] \times \log [1 - p(D|K_i)]\} \quad (17)$$

其中p(D|K<sub>i</sub>) 是第K<sub>i</sub>等的PD，若K<sub>i</sub>等有N<sub>Di</sub>個違約客戶，N<sub>NDi</sub>個未違約客戶，則p(D|K<sub>i</sub>) 計算如方程式 (18)：

$$P(D | K_i) = \frac{N_{Di}}{N_{Di} + N_{NDi}} \quad (18)$$

條件熵CE(K) 被視為各評等等級的加權ce(K<sub>i</sub>) 的平均值，則CE(K) 計算如方程式 (19)：

$$CE(K) = \frac{\sum_{i=1}^m (N_{Di} + N_{NDi}) \times ce(K_i)}{\sum_{i=1}^m (N_{Di} + N_{NDi})} \quad (19)$$

評估模型是否有效區分風險，就必須用到Kullback-Leibler距離 (Kullback-Leibler distance)，其算式為：Kullback-Leibler距離= H(p) - CE(K)，其中H(p) 為樣本違約機率的信息熵，而p計算如方程式 (20)。

$$p = \frac{\sum_{i=1}^m N_{Di}}{\sum_{i=1}^m (N_{Di} + N_{NDi})} \quad (20)$$

Kullback-Leibler距離介於0至 $H(p)$ 之間，距離越長則表示評等模型越能有效區分風險。可以利用Kullback-Leibler距離標準化，以產生CIER值，來評估評等模型的等級同質性，可見方程式(21)。CIER越接近1，表示評等模型所得到的結果，相同等級的樣本其性質相似，不同等的樣本差異很大，即等級內的樣本同質性高，等級與等級之間異質性大。

$$CIER = \frac{H(p) - CE(K)}{H(p)} \quad (21)$$

### (五) TM

TM是利用違約風險評等規則對於第 $n$ 期(如：季)樣本進行預測的結果並紀錄每個等級內的授信戶組成，再利用同樣的規則對第 $n+1$ 季的樣本預測評等，透過前後二季的評等結果來產生一個移轉矩陣，矩陣會表示出前季等級內的樣本有多少比率移轉到各個等級，舉例來說：由【表2】可見前季評等為第1級的樣本到了後季時，繼續維持在第1級的樣本只剩58%，其中有30%移轉到第2級且有14%移轉到第3級，其它等級以此類推，而矩陣對角線的集中度愈高，代表決定模型預測的穩定性愈高。本系統是利用對角線加總值作為本系統判斷穩定性的依據。

表2：移轉矩陣

	1	2	3	4	5	6	7	8
1	56%	30%	14%	0	0	0	0	0
2	0	42%	25.00%	14.00%	10.00%	9.00%	0	0
3	0	0	30.36%	48.21%	17.86%	3.57%	0	0
4	0	0	0.67%	12.66%	27.33%	38.67%	12.67%	8%
5	0	0	0	0	6.45%	8.06%	80.65%	4.84%
6	0	0	0	0	0	19.23%	46.15%	34.62%
7	0	0	0	0	0	0	15.38%	84.62%
8	0	0	0	0	0	0	0	100%

## 五、適應函數

在EMCRS系統中GA的適應函數計算分成四個步驟，首先會將染色體解碼出預測變數組合及評等門檻值；第二步是利用訓練樣本、選擇的變數組合及門檻值來產生MCLR迴歸式；第三步則是將訓練樣本帶入MCLR中得到違約風險預測評等結果；第四步是以Basel II建議的驗證方法檢定模型，適應計算如方程式(22)。

$$\text{Fitness value} = \frac{V_{\text{CAP}} + V_{\text{ROC}} + V_{\text{KS}} + V_{\text{TM}} + V_{\text{CIER}}}{V_{\text{vars}}} \quad (22)$$

其中， $V_{\text{CAP}}$ 、 $V_{\text{ROC}}$ 、 $V_{\text{KS}}$ 、 $V_{\text{TM}}$ 、 $V_{\text{CIER}}$ 是經由驗證方法的求得各個項目檢定所得到的值，例如： $V_{\text{CAP}}$ 代表代入訓練樣本求得的CAP指標值； $V_{\text{vars}}$ 計算是GA染色體解碼出來

被挑選出來的變數個數，當挑選到的變數數量愈多，其值愈大。由於 $V_{CAP}$ 、 $V_{ROC}$ 、 $V_{KS}$ 、 $V_{TM}$ 、 $V_{CIER}$ 的值都是愈大愈好，而 $V_{vars}$ 卻是相反，故我們將最後將各項 $V_{CAP}$ 、 $V_{ROC}$ 、 $V_{KS}$ 、 $V_{TM}$ 、 $V_{CIER}$ 值加總後的分數再除以 $V_{vars}$ ，就是染色體的適應函數值，適應函數值愈大則代表染色體愈優異，被遺傳到下一代的機率愈高。

## 肆、實驗結果分析

本研究之實驗目的是驗證演化式多重組合羅吉斯迴歸所產生的信用評等的預測效力與穩定度。實驗之前需要收集樣本資料來源，如預測變數以及違約與正常公司的樣本；接著要說明遺傳演算法做訓練時的參數設定；實驗設計的方向說明；最後是實驗結果與分析。

### 一、資料來源

本研究收集的原始變數種類包含基本資料變數（1）、財務變數（60）、公司治理變數（40）、會計師變數（1）及總體經濟變數（5）等共107個初始變數。本研究採用台灣經濟新報（TEJ）的TCRI為外部評等變數，做為EMCRS驗證的比較基準（benchmark），同時也加入了區別分析、羅吉斯迴歸以及決策樹做為比較模型。樣本資料源自於台灣經濟新報資料庫中不含金融業與建築業等特殊產業之外的一般產業上市上櫃公司共402家。以同產業、同期間、資產規模相似的正常公司與違約公司採取1:1的方式做樣本配對，違約定義參考自TEJ。樣本資料時間從1996年12月至2006年6月共39季的季資料。實驗樣本分配是將402家樣本中，每家樣本收集違約公司與配對的正常公司的違約前一季、違約前二季、違約前三季和違約前四季的四組樣本的變數資料，每一組透過亂數選擇出282家的訓練樣本和120家的測試樣本，分配比率為7：3，參考自Lin與McClean（2001）。受限於台灣證券市場不大的因素，每組樣本採隨機抽樣目的是希望做到近似交叉驗證（cross validation）的目標。

### 二、GA實驗參數

在參數的設計部分，我們參考Siriwardene與Perera（2006）從多組不同參數實驗，依據各組實驗結果決定最佳的GA參數，如【表3】可看出本研究系統在交配率0.8及突變率0.01的CIER, TM, CAP, ROC, KS的值都大於平均值與中位數，是EMCRS系統最佳化的一組參數。EMCRS系統的GA參數相關設定如【表4】。

表3：GA參數實驗

交配率	突變率	CIER	TM	CAP	ROC	KS
0.6	0.001	0.7010	0.3188	0.9000	0.9500	85.0000
0.6	0.004	0.7132	0.2501	0.9144	0.9572	88.3333
0.6	0.008	0.7143	0.3648	0.9125	0.9510	86.6667
0.6	0.01	0.7229	0.3190	0.9125	0.9563	88.3333
0.7	0.001	0.5890	0.2293	0.8775	0.9388	78.3333
0.7	0.004	0.7165	0.3165	0.9194	0.9597	86.6667
0.7	0.008	0.6614	0.2400	0.9061	0.9531	83.3333
0.7	0.01	0.6459	0.2949	0.8900	0.9450	85.0000
0.8	0.001	0.7370	0.3435	0.9142	0.9571	86.6667
0.8	0.004	0.7312	0.3010	0.9231	0.9615	86.6667
0.8	0.008	0.7297	0.3087	0.9314	0.9657	86.6667
0.8	0.01	0.7405	0.3554	0.9069	0.9535	88.3333
0.9	0.001	0.6095	0.3115	0.8644	0.9322	80.0000
0.9	0.004	0.6319	0.3192	0.8500	0.9250	85.0000
0.9	0.008	0.6809	0.4119	0.9092	0.9516	83.3333
0.9	0.01	0.5350	0.2627	0.8408	0.9204	75.0000
	平均值	0.6787	0.3092	0.8983	0.9488	84.5833
	最大值	0.7405	0.4119	0.9314	0.9657	88.3333
	最小值	0.5350	0.2293	0.8408	0.9204	75.0000
	標準差	0.0611	0.0479	0.0266	0.0132	3.8249
	中位數	0.7071	0.3140	0.9081	0.9533	85.8334

表4：GA參數設定

參數名稱	參數值
族群數 (population)	100
染色體選擇方式 (selection method)	競爭法
演化代數 (generation)	2000
突變率 (mutation rate)	0.01
交配率 (crossover rate)	0.8
交配點 (crossover point)	2
染色體長度 (chromosome length)	107+11* n (欲分級數量-1)

### 三、實驗設計

實驗設計朝三個方向進行：使用Basel II指標驗證五種評等模型的效能、模型配適度檢定以及MCLR預測模型解析。

- 評等模型的效能：透過五種不同評等方法（區別分析、羅吉斯、決策樹、TCRI以及EMCRS），同時利用不同違約時間點的樣本建構評等規則，接著我們以CIEF, CAP, ROC, KS, TM來驗證各評等方法的優劣。
- 模型配適度：羅吉斯與EMCRS模型依其均方差利用HL以及Brier分數來檢定評等模

型的配適度，其他評等模型（區別分析、決策樹、TCRI）沒有預測違約機率，因此此部份的實驗省略。HL以及Brier指標將在後續文章中說明。

- MCLR預測模型解析：針對EMCRS中變數挑選結果解析以及其估計的MCLR方程式說明。

### （一）Brier score

Brier是利用 $\hat{p}$ 預測機率與 $D$ 實際違約發生0到1之間的值，計算其誤差平方和， $n$ 為分組數，故當求得數字愈小代表預測效力愈佳。其定義式如方程式（23）：

$$\text{Brier} = \frac{\sum_{i=1}^n (\hat{P}_i - D_i)^2}{n} \quad (23)$$

### （二）Hosmer-Lemeshow指標（HL）

此方法是1989年Hosmer-Lemeshow提出HL指標，其功能類似皮爾遜 $\chi^2$ 統計量，使用者可以利用HL指標來檢定羅吉斯迴歸配適度見方程式如（24）。其中， $\bar{q}_i = \frac{1}{n} \sum_{j=1}^n q_j$ ， $q$ 是樣本的預測機率， $\hat{p}_i$ 則是每組組內樣本違約率的平均。若信用評用模型優良，則HL配適度會呈現卡方分配。

$$HL = \sum_{i=1}^k \frac{n_i (\bar{q}_i - \hat{p}_i)^2}{q_i (1 - q_i)} \quad (24)$$

## 四、實驗結果

EMCRS的實驗結果分析將朝評等模型的驗證比較、不同違約時間的評等效力以及最佳化後變數解析以及MCLR估計模型進行說明。

### （一）EMCRS與各評等模型效能比較

此實驗是將EMCRS系統所建構的評等結果進行驗證比較。我們利用Basel II建議的驗證方法來分析兩者的差異。驗證方向包含：違約預測力（如：KS檢定、ROC、CAP）、穩定度—移轉矩陣率（如：TM）及等級同質性（如：CIER）等三個構面總共5個檢定指標做驗證。EMCRS是採用GA做變數與各評等門檻值的挑選後再以Basel II的驗證方法做為目標函數進行最佳化。依不同違約時點，每個違約時點做10種等級的評等，並將每個違約時點各等級的驗證結果做平均。

由於在模型訓練時會利用到前後兩季來計算移轉矩陣，而違約前一季無法和下一季的現在時間點樣本來做計算，因此其他指標也配合只計算違約前四季、前三季、前二季的訓練樣本做實驗。【表5】是採用5個驗證指標針對區別分析、羅吉斯、決策樹、TCRI以及EMCRS五種評等方法，在三個違約時間點樣本分析其評等效力。實驗結果顯示EMCRS模型在各項驗證中，除了TM之外，幾乎都明顯優於其它模型。其中CIER的三個違約時間點表現而言，EMCRS的0.74, 0.71, 0.76是高於其它方法，CIER值愈接近1

表示模型的評等分類效果愈佳。CAP表現上則是EMCRS違約前三季為0.93最好，CAP值愈接近1表示評等模型分等的準確性愈高。ROC是EMCRS違約前三季的0.96較佳，ROC值愈接近1表示愈貼近完美模型。KS在EMCRS違約前四、二季為88.33表現最好，KS值愈高表示模型鑑別力愈好，當KS值大於75，表示模型鑑別能力極優良【表1】。另外，EMCRS在違約前四季的TM值為0.13較不如其它模型表現，可能是因為樣本離違約時間點較遠，所以評等的穩定性上變化較大，而違約前三季與前二季的TM值，是介於五種評等方法的中等。EMCRS所選擇的變數用來預測違約第四季後的結果，雖然準確度可以達到不錯的水準，但在評估當季及下一季（第三季）信用風險移轉上，評等的穩定度卻不理想。

表5：評等模型驗證比較表

違約時間點	評等方法	CIER	CAP	ROC	KS	TM
違約前四季	區別分析	0.3753	0.6439	0.8219	50.0000	0.2634
	決策樹	0.4078	0.7300	0.8650	58.3333	0.1529
	TCRI	0.2968	0.6039	0.8019	55.0000	0.5127
	羅吉斯	0.1465	-0.3242	0.3379	33.3333	0.0734
	EMCRS	0.7425	0.9039	0.9519	88.3333	0.1276
違約前三季	區別分析	0.6324	0.8503	0.9251	73.3333	0.3210
	決策樹	0.3752	0.2725	0.6363	45.0000	0.2878
	TCRI	0.3588	0.7069	0.8535	58.3333	0.2397
	羅吉斯	0.6324	0.8503	0.9251	73.3333	0.1458
	EMCRS	0.7082	0.9286	0.9643	85.0000	0.2951
違約前二季	區別分析	0.4942	0.7028	0.8514	73.3333	0.3615
	決策樹	0.3217	0.6522	0.8261	58.3333	0.2788
	TCRI	0.4158	0.7431	0.8715	61.6667	0.2936
	羅吉斯	0.7434	0.9253	0.9626	86.6667	0.1365
	EMCRS	0.7566	0.9150	0.9575	88.3333	0.2951

## （二）模型配適度檢定

羅吉斯迴歸和EMCRS系統對於樣本能夠計算其違約機率，因此我們從違約率配適度方面來衡量這兩個評等方法的檢定其預測效能。衡量方法是Brier分數以及HL值，兩者檢定的值都是取絕對值後，愈小愈好。有關Brier檢定方面：實驗結果顯示，除違約前四季的評等結果，EMCRS（0.1472）優於羅吉斯（0.6455）見【表6】，其餘兩季的結果兩種評等模型均在0.1左右，尤其是違約前2季的結果很接近0，表示違約時間點愈近Brier值愈小，表現愈好。分析HL統計量時需搭配卡方分配臨界值進行檢定比較。本研究的評等等級數目是10等，測試期的樣數為基礎，其卡方分配的自由度是8，在99%與95%信賴水準的單尾檢定臨界值分別為1.65與2.73。依據HL統計量的結果，EMCRS三季三季的評等配適度均屬良好。反之，羅吉斯的統計量則只有在違約前二季有在範圍內，其餘前三季與前四季均超出臨界值，HL檢定結果不理想。

表6：評等模型驗證比較表

違約時間點	評等方法	Brier	HL
違約 前四季	EMCRS	0.1472	1.4735
	羅吉斯	0.6455	405.5713
違約 前三季	EMCRS	0.1031	1.0017
	羅吉斯	0.1175	140.1330
違約 前二季	EMCRS	0.0832	0.9423
	羅吉斯	0.0578	10.0533

### (三) 評估模型分析

評估模型分析的部份，將針對EMCRS挑選到的變數與估計的MCLR方程式進行分析與說明，分述如下。

#### 1. EMCRS挑選後的變數分析

從EMCRS結果來分析，從【表7】被選到變數有財務變數及公司治理變數，其中來自財務變數且屬於獲利能力指標的有： $x_{19}$ 每股淨值A（元）、 $x_{21}$ 常續性EPS、 $x_{52}$ 總資產週轉率（次）、 $x_{56}$ 淨值週轉率（次）；而來自公司治理類的有： $x_{77}$ 自然人持股%； $x_{83}$ 董監質押%； $x_{84}$ 董監酬勞佔稅前淨利%。以上變數「常續性EPS」被選擇3次、「每股淨值A（元）」與「總資產週轉率（次）」被選擇2次。因此可知獲利能力的相關指標為影響信用風險評等很重要的變數。而公司治理變數則與自然人持股與董監事相關的變數較明顯，表示財務比率與公司治理變數對信用評等模型具有關鍵性的影響。依據挑選到變數可以發現，公司治理被選到的變數沒有重覆現象，隱含變數之間可能存在共線性問題，這或許是GA最佳化變數的一個限制。

本實驗另外發現，財務變數的償債能力與經營能力並未有顯著的影響力，檢視樣本之負債比、流動性及週轉率尚可，若財務報告未經窗飾，則此2型態之變數應不是造成本實驗樣本違約的主因。基本資料變數、會計師變數及總體經濟變數，在此次實證中亦不顯著。由於樣本中的公司，絕大多數成立10年以上，邁入成熟期，非因年輕根基不穩而產生違約。因此，基本資料變數不顯著；違約公司中有更換會計師的家數（4家）過少，使會計師變數不顯著；總體經濟變數的部分，有可能是因正常公司、違約公司面臨的總體經濟環境相同，故其不具解釋效果。

#### 2. 估計的MCLR方程式

本實驗的EMCRS在違約前二季表現較好，我們以此來說明MCLR評等方程式模型，MCLR的估計方程式（25）所求得殘差項如下：

$$\varepsilon_j = 1.71 + 2.06x_{19} - 21.47x_{21} - 10.1x_{52} - 21.27x_{84} \quad (25)$$

EMCRS產生的10個評等等級中的9個門檻值 $\varepsilon_1^c \sim \varepsilon_9^c$ 分別為-10.66, -6, -4.13, -0.98, 0.16, 3.28, 3.52, 7.84, 10.16見【圖11】，從圖形中可以看出此10個等級9個門檻值的分佈

尚屬均勻對稱。樣本將變數資料帶入方程式 (26) 求得機率值即可與各等級門檻比較，即可計算該樣本所屬的評等等級。

$$P(y_j = D|x_j) = \frac{1}{9} \sum_{z=1}^9 \frac{1}{1 + e^{-(\varepsilon_j + 2\varepsilon_z^c)}} \quad (26)$$

表7：EMCRS模型選擇預測因子表

季別	編號	名稱
違約前二季	X19	每股淨值A (元)
	X21	常續性EPS
	X77	自然人持股%
違約前三季	X21	常續性EPS
	X52	總資產週轉率 (次)
	X56	淨值週轉率 (次)
	X83	董監質押%
違約前二季	X19	每股淨值A (元)
	X21	常續性EPS
	X52	總資產週轉率 (次)
	X84	董監酬勞佔稅前淨利%

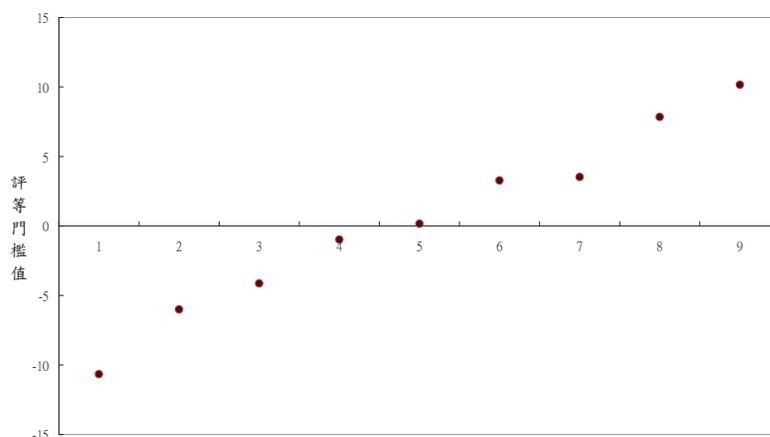


圖11：評等門檻值

## 伍、結論與未來建議

本研究提出以GA及MCLR模型整合來建構出的EMCRS系統，從實驗結果可發現，利用非線性搜尋方式的GA，能夠有效的找出最佳門檻值和預測變數。就理論創新而言，本研究將LR擴充為預測次序多類別的MCLR，應用於預測信用風險評等上，透過Basel II所建議的評等模型驗證方法來檢驗MCLR模型的評等效力。就銀行管理面來說，穩健有效的信用評等機制能強化銀行競爭力，銀行藉由信用風險進行評價，產生最適合的資

產組合，若資產的信用風險評等後被低估將產生實質虧損，反之高估風險將使產品訂價無競爭力。就企業管理面而言，企業藉由評等機制進行信用評等，可以了解企業本身的償債能力以及經營風險，改善公司財務體質，適時調整財務槓桿，協助公司獲得更佳的融資條件，創造經營效益。就投資人來說，對信用評等較低的投資標的可以先行出脫持股，可以選擇評等較高的企業進行投資決策，促進資本市場活絡，改善資訊不對稱投資現況。因此，本研究無論對銀行、企業或是投資大眾都有正面的啟示意義。

本研究實驗結果發現：（1）在「模型的評等分類效果」、「評等模型準確性」、「貼近完美模型」以及「模型鑑別力」方面EMCRS比其它評等方法更具有較佳的評等效力，說明GA與MCLR整合應用，獲得不錯的信用風險評等效果；（2）距離違約時間點愈遠等級同質性的效果較差，印證近期的財務資訊較能反應企業的營運狀況，其等級同質性也有中上程度的表現；（3）預測變數中被選到最多是財務變數類，包含有（每股淨值、常續性EPS、每股稅前淨利、總資產週轉率（次）），而常續性EPS被選擇的次數為最多；其次是公司治理變數（董監酬勞佔稅前淨利%），足見EMCRS評等模型受財務變數類的影響較大。

未來研究建議朝幾個方向進行：（1）配合其他統計之變數挑選方式（如因素分析、主成份分析或逐步回歸）事前檢定後再加入EMCRS驗證是否能改善評等效能；（2）樣本的數量增加對於模型的建立更具代表性；或分析更多的總體經濟、會計師變數等或許能找到相關的影響因子；（3）加入其它國際金融評等機構的信用評等模型做比較；（4）在建構評等模型時，不易兼顧所有構面（如同時要最具正確性、穩定性以及同質性等），依評等使用者（如銀行、企業或投資人）的需求偏好設定不同的權重，依不同偏好的優先順序最佳化其權重，期待建構更合適的模型，使評等模型發揮最大效用。

## 參考文獻

1. Altman, E. I. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance* (23:4), September 1968, pp. 589-609.
2. Back, B., Laitinen, T., and Sere, K. "Neural Networks and Genetic Algorithms for Bankruptcy Predictions," *Expert Systems with Applications* (11:4), 1996, pp. 407-413.
3. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring," *The Journal of the Operational Research Society* (54:6), June 2003, pp. 627-635.
4. Balcaen, S., and Ooghe, H. "35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems," *The British Accounting Review* (38:1), March 2006, pp. 63-93.
5. Barniv, R., Mehrez, A., and Kline, D. M. "Confidence Intervals for Controlling the Probability of Bankruptcy," *Omega* (28:5), October 2000, pp. 555-565.
6. Deakin, E. B. "A Discriminant Analysis of Predictors of Business Failure," *Journal of*

- Accounting Research* (10:1), 1972, pp. 167-179.
7. Donato, J. M., Schryver, J. C., Hinkel, G. C., Schmoyer Jr., R. L., Grady, N. W., and Leuze, M. R. "Mining Multi-Dimensional Data for Decision Support," *Future Generation Computer Systems* (15:3), April 1999, pp. 433-441.
  8. Foreman, R. D. "A Logistic Analysis of Bankruptcy within the US Local Telecommunications Industry," *Journal of Economics and Business* (55:2), March 2003, pp. 135-166.
  9. Fu, L., and Simpson, D. G. "Conditional Risk Models for Ordinal Response Data: Simultaneous Logistic Regression Analysis and Generalized Score Tests," *Journal of Statistical Planning and Inference* (108), November 2002, pp. 201-217.
  10. Holland, J. H. "Adaptation in Natural and Artificial Systems," *University of Michigan Press*, 1975, pp. 529-530.
  11. Hosmer, D.W., and Lemeshow, S. *Applied Logistic Regression*, John Wiley & Sons, New York, 1989.
  12. Huyghebaert, N., Gaeremynck, A., Roodhooft, F., and Van De Gucht, L. M. "New Firm Survival: The Effect of Start-up Characteristics," *Journal of Business Finance & Accounting* (27), 2000, pp. 627-651.
  13. Keasey, K., and Watson, R. "Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti's Hypotheses," *Journal of Business Finance & Accounting* (14), 1987, pp. 335-354.
  14. Kim, M.J., and Han, I. "The Discovery of Experts' Decision Rules from Qualitative Bankruptcy Data Using Genetic Algorithms," *Expert Systems with Applications* (25:4), November 2003, pp. 637-646.
  15. Ko, P.C., and Lin, P.C. "An Evolution-Based Approach with Modularized Evaluations to Forecast Financial Distress," *Knowledge-Based Systems* (19:1), March 2006, pp. 84-91.
  16. Laitinen, E. K., and Laitinen, T. "Bankruptcy Prediction: Application of the Taylor's Expansion in Logistic Regression," *International Review of Financial Analysis* (9:4), 2000, pp. 327-349.
  17. Laster, D. "Insurance Company Ratings," *Swiss Re sigma* (4), 2003, pp. 23-39.
  18. Lin, F.Y., and McClean, S. "A Data Mining Approach to the Prediction of Corporate Failure," *Knowledge-Based Systems* (14:3), June 2001, pp. 189-195.
  19. Mays, E. *Handbook of Credit Scoring*, Lessons Professional Publishing, 2001, pp 89-106.
  20. Meyer, P. A., and Pifer, H. W. "Prediction of Bank Failures," *Journal of Finance* (25:4), September 1970, pp. 853-868.
  21. Min, S.H., Lee, J., and Han, I. "Hybrid Genetic Algorithms and Support Vector Machines for Bankruptcy Prediction," *Expert Systems with Applications* (31:3), October 2006, pp. 652-660.

22. Ohlson, J. A. "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research* (18:1), 1980, pp. 109-131.
23. Shin, K.S., and Han, I. "A Case-Based Approach Using Inductive Indexing for Corporate Bond Rating," *Decision Support Systems* (32:1), November 2001, pp. 41-52.
24. Shin, K.S., and Lee, Y.J. "A Genetic Algorithm Application in Bankruptcy Prediction Modeling," *Expert Systems with Applications* (23:3), October 2002, pp. 321-328.
25. Siriwardene, N. R., and Perera, B. J. C. "Selection of Genetic Algorithm Operators for Urban Drainage Model Parameter Optimization," *Mathematical and Computer Modelling* (44), 2006, pp. 415-429.
26. Tamari, M. "Financial Ratios as a Means of Forecasting Bankruptcy," *Management International Review* (4), 1966, pp. 15-21.
27. The Basel Committee on Banking Supervision "Studies on the Validation of Internal Rating Systems" *Basel Committee on Banking Supervision Working Paper No. 14*, February 2005.
28. Tseng, F.M., and Lin, L. "A Quadratic Interval Logit Model for Forecasting Bankruptcy," *Omega* (33:1), February 2005, pp. 85-91.
29. Ugurlu, M., and Aksoy, H. "Prediction of Corporate Financial Distress in an Emerging Market: The Case of Turkey," *Cross Cultural Management: An International Journal* (13:4), 2006, pp. 277-295.
30. Westgaard, S., and Van der Wijst, N. "Default Probabilities in a Corporate Bank Portfolio: A Logistic Model Approach," *European Journal of Operational Research* (135:2), December 2001, pp. 338-349.
31. Wu, C.H., Tzeng, G.H., Goo, Y.J., and Fang, W.C. "A Real-Valued Genetic Algorithm to Optimize the Parameters of Support Vector Machine for Predicting Bankruptcy," *Expert Systems with Applications* (32:2), February 2007, pp. 397-408.

