

消費者信用貸款違約風險評估模型 之研究-以 CART 分類與迴歸樹建模

On the Research of Default Risk Model of Consumer Credit Loan—Using CART (Classification and Regression Tree)

梁德馨 *Te-Hsin Liang*

私立輔仁大學統計資訊系

Department of Statistics and Information Science,
Fu Jen Catholic University

葉建良 *Chien-Liang Yeh*

玉山銀行金控總部

Department of Financial Holding
E. Sun Bank

摘要

本研究旨在找出影響消費者信用貸款違約之重要變數，並藉以建立消費者信用貸款違約風險評估模型，以期做為銀行在信用貸款風險管理上之參考依據。本研究以相關分析、因素分析及主成份分析來選擇及整合對信用貸款違約有重要影響力之變數，並以分類與迴歸決策樹（CART）進行預測違約模型之建構，最後經由模型預測能力指標、ROC 曲線與 Cumulative Lift 曲線之綜合比較後，選取出消費者信用貸款違約風險預測之最適模型～「違約逾期相關程度法-決策樹」模型。研究結果顯示，此模型整體的預測正確率可達 77%以上。

關鍵字：消費者信用貸款、違約風險、預期違約率、分類與迴歸決策樹

Abstract

This research is for the purpose of discovering the important variables which will affect the consumer credit loan. So as to and establishes the default risk model of consumer credit loan. This research uses the correlation analysis, factor analysis and principle component analysis to integrate those important effective variables. And the CART (Classification and Regression Tree) method is adopted to build up a default risk model of consumer credit loan. Three kinds of variable selection models including “correlated with default”, “factor analysis”, and “principle analysis” are proposed and compared. By comparing the Accuracy rate, ROC Curve, and Cumulative Lift Curve of those models, the “correlated with default” model is chosen to be the adaptable model. The Accuracy rate of this chosen model is high than 77 percent.

Key Words : Consumer Credit Loan, Default Risk, The Probability of Expect Default, CART

壹、緒論

台灣近年來貸款業務隨著新銀行的開放而競爭激烈。在追求市場佔有率及獲利之下，各類貸款更是以多元化且較寬鬆審核方式出現。在惡性競爭之下，使得放款業務市場呈現飽和，各家銀行皆出現逾放比過高且獲利不佳之狀況。而隨著台灣地區消費習慣的改變及國人消費意識抬頭，國內的消費金融市場出現了極大的需求。是以，各銀行逐漸將獲利的矛頭轉向消費金融業務，其中又以消費性信用貸款為盛，所相中的無非是其不同於以往的放款型態，存在著顧客廣、金額小、回收快及利差大等誘因。由圖 1 可得知近年來消費貸款市場的蓬勃發展。

消費金融放款為銀行放款業務的熱門首選業務，其中又以消費者小額信用貸款最受矚目。消費者小額信用貸款主要強調申請便利及核貸迅速，銀行必須

在最短時間內，有效對顧客的信用狀況與違約風險進行衡量評估。但其申請對象分布於社會各階層，故所面臨的違約風險迥異且不易掌控。因而，各銀行特別強調個人的信用狀況評估與違約風險的測量，亦即必須重視違約逾期管理（薛兆亨，1991），期能透過有效辨識顧客違約風險程度給予不同風險屬性的顧客適當的授信調整，以降低銀行業者的違約損失。

如前述所言，銀行能成功發展消費者小額信用貸款的主要關鍵因素，在於事先能對顧客的信用狀況加以辨識及有效管理違約風險。隨著新巴塞爾新資本協定（Basel II）之提出及實施，國內金融機構須在風險管理上追上國際標準，皆著手嘗試建立內部模型，以期降低需計提資本，使資金運用更具效率。因此如何估計風險值(Value-at-Risk, VaR)變成顯學。

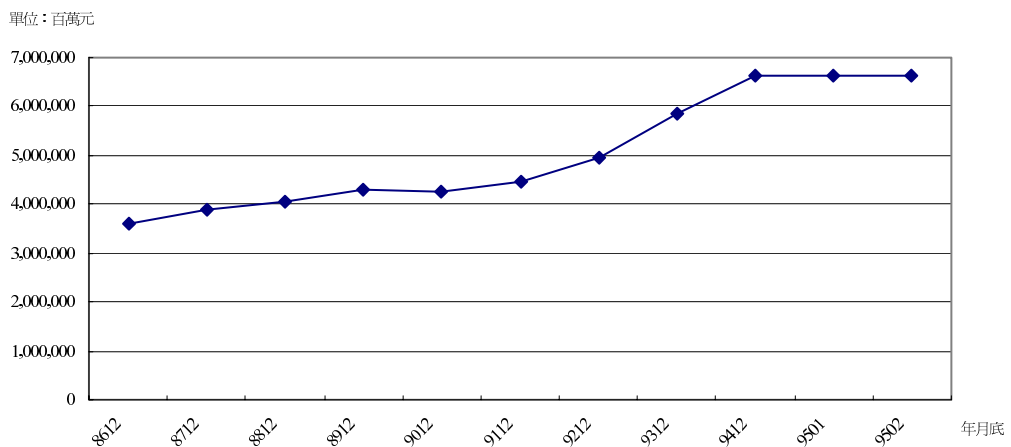


圖 1 消費貸款餘額變化情形¹

金融風險一般可分為市場風險 (Market Risks)、信用風險 (Credit Risks)、流動性風險 (Liquidity Risks)、作業風險 (Operational Risks) 及法律風險 (Legal Risks) 五類 (林邵杰，2003)。這五類風險中，信用風險是唯一完全由顧客行為所產生之風險，因此必須透過針對顧客的違約行為進行分析，才能預估及防範信用風險之發生。

近年來學者嘗試將市場風險管理中之估計風險值(Value-at-Risk, VaR)的

¹ 資料來源：中央銀行統計資料消費者貸款及建築貸款餘額，2007 http://www.cbc.gov.tw/economic/statistics/total_index.asp

概念應用在信用風險的管理上，利用顧客違約機率以及違約後可能損失，或者顧客信用評等被降級的機率以及降級後可能產生的損失，計算出信用風險所導致的最大可能損失，此最大可能損失即為信用風險值（Credit-at-Risk, Credit VaR）。在新巴塞爾新資本協定中估計信用風險值的方法有三，分別為標準法（Standardized method）、基本內部評等法（Internal Rating Based: Foundation, IRB）及進階內部評等法（Internal Rating Based: Advanced）。其中標準法是允許銀行採用外部評等機構之評等結果作為風險權數，而採用此法較無法依顧客特性來降低銀行資本提列的金額。內部評等法則可依銀行顧客的風險等級，來分別提列最低資本，因此相對於標準法而言，較有可能精算出最低或最適的資本提列金額。在信用風險值的衡量模型中，有 Creditmetrics、KMV、CreditPortfolioView 等，各模型雖各有優缺點及特色，各模型皆必須立內部信用評等系統，並計算違約機率（Probability of Default, PD）（林邵杰，2003）。相較於標準法而言，內部評等法之技術障礙存更高，必須針對貸款者違約機率（Probability of Default, PD）提出計算標準，以向主管機關證明該銀行有能力運用內部評等法進行信用風險資本提列（張大成，2003）。本研究即是針對違約機率之預測及計算進行研究。

有關違約機率之計算方式，根據 Hull（2003）的分類，主要可分為債券價格法（Bond Price Approach）、歷史評等資料法（Historical Rating Data Approach）及選擇權理論法（Option Theory Approach）。台灣現行的債券市場並不是十分發達，故債券評等法之實用性不高；選擇權理論法較試用於法人金融之公司違約機率估計；歷史評等資料法則需有完整且品質良好的顧客歷史資料，隨著資料庫及資訊科技技術的快速發展及進步，近年來台灣各銀行大多已整併出完整且品質不錯的顧客資料庫，因此若銀行欲採用內部評等法進行資本提列，歷史評等資料法則是較佳的選擇（張大成，2003）。在違約機率計算方法之歷史評等資料法中，目前較常被採用的信用評等系統，主要可分為二大類：一為專家意見法，二為統計模型法（張大成，2003）。而專家意見法通常是由金融機構地方或分行放款主管所決定，其較屬於主觀判斷的結果，不符合新巴塞爾新資本協定的規範。因此本研究將以統計模型法進行違約機率之計算及預測。

目前國內之相關研究，最常採用的統計預測模型有區別分析（Discriminate Analysis）、邏輯斯迴歸（Logistic Resgrssion）、類神經網路（Neural Network）及決策樹（Decision Tree）等方法。而區別分析及類神經網路雖然常可以做出很準確地預測，但區別分析之資料要求需要符合常態性、獨立性及均質性等假

設，在實務上不易達到（陳順宇，2005），且區別分析的模型無法解釋變數間共線性的問題（Henley & Hand, 1996; 莊瑞珠，2007）；而類神經網路其計算的過程為一黑箱作業（黃宇翔、毛紹睿，2005；謝邦昌，2005），無法明確地掌握自變數對違約機率之影響，因此無法符合新巴塞爾新資本協定的要求，亦較無法透過模式去解讀顧客的特質及行為；故此二種方法較少被銀行界正式採用。邏輯斯迴歸及決策樹法則是目前業界較為採用的二種統計預測模型法，此二種方法皆可透過模式去解讀顧客的特質及行為。邏輯斯迴歸法在選用建模變數時，其建模變數之尺度雖不設限，但若資料為間斷變數型態時，則必須將其轉換為虛擬變數（Dummy Variable），而當間斷變數個數較多時，會造成模型的維度過大等現象，而且分判點之決定亦會影響模型的預測力（范哲銘，2003）；但決策樹法則不受此限，且決策樹在建模過程中一次僅取一個建模變數（謝邦昌，2005），可有架構地找到不同風險等級下之重要影響因素組合，其產生的訊息最容易解讀。因為本研究的資料型態包含連續及間斷二類，所以本研究採用決策樹法進行違約風險之預測。

此外，由於銀行除了本身所擁有的顧客資料外，尚可透過聯合徵信中心（JCIC）取得相關顧客信用資料，所以在資訊較以往充足的情況下，各家銀行更是積極地著手建立違約風險模型，以區辨出銀行所面臨的顧客風險大小，降低銀行的違約損失（Loss Given Default, LGD）。然而目前大多研究者在建立違約風險評估模型時所選用的建模變數，多為顧客基本資料或少部分的信用狀況資料，能完整考慮顧客於聯合徵信中心的信用狀況資料（後續稱信用相關整合變數）的研究極少。而本研究除將採用顧客基本資料或行內信用狀況資料外，將大量採用顧客之聯合徵信中心的信用狀況資料，進行違約風險之預測，以期能更完整地分析具有違約風險顧客，以了解其基本特質及在行內及他行信用狀況。

在上述研究背景下，本研究與國內某家金控銀行（以下簡稱 A 銀行）合作，運用其既有之顧客資料庫，建立消費者小額信用貸款違約風險評估模型，除可做為計算 VaR 值時風險權數之參考基礎外，並期望能有效的辨識現有顧客之違約風險程度高低，並用以預測新顧客之違約風險程度高低，以適時規避或控管違約風險較高之顧客族群，進而能篩選出信用狀況良好之顧客特性，加以積極爭取；另外，亦可同時提升審核及放款效率，為銀行爭取更高之利潤。故基於上述本研究之主要目的如下：

1. 針對信用貸款顧客基本資料及顧客於聯合徵信中心之相關信用資料進行剖析，以期找出對違約風險具影響力之變數。

- 2.透過分類與迴歸決策樹法(以下簡稱 CART, Classification and Regression Tree)進行違約風險模型之建構,以期找出較適變數組合,並且將顧客分群,以預估各群顧客之預期違約機率;另外,亦針對所估計之預期違約率進行最適切點分析之探討。

貳、相關研究探討

一、消費性信用貸款與信用風險

個人消費性信用貸款從消費者角度所下之定義²:「凡信用良好之顧客,為小額投資理財或週轉之目的向銀行申請免擔保之短/中期無擔保放款,稱為個人消費性信用貸款」。若以銀行經營角度觀之,則可以定義為:「銀行為提高放款收益率,增加消費者各階段之產品服務,提供經銀行徵信審核合格之顧客,申請免擔保及免保證人之小額融資,並在有效控制之風險程度內追求利潤之極大,謂之個人消費性信用貸款」。

所謂信用風險(Credit Risk),又稱為授信風險,係由於債權人於「現在」將錢貸與債務人,而「期待」債務人於「未來」還本付息的行為。其中「期待」及「未來」充滿風險性和不確定性,因為錢貸放出去後,必須經過一段時間之後,才能收回本息,而在這未來的一段時間裡,充滿著許多不確定因素,所以「信用風險」伴隨著「授信」而存在(馬芳資, 1994)。換言之,信用風險是指交易對手不願意或無法履行契約中規定義務時所產生的風險。信用風險為雙方承諾交易完成後,其某一方未遵守合約完成物品、服務或者是金融資產的交付(Daniels & VanHoose, 1999)。整體而言,金融機構的信用風險係由違約風險、暴露風險及回收風險所構成。其中所謂違約風險(Default risk)係由違約事件的發生機率來描述,銀行面對的違約行為涵蓋尚未履行付款義務的違約、交易對手未遵守契約規定的違約及經濟違約三種層面。本研究所探討的信用風險即為銀行面對的違約行為涵蓋尚未履行付款義務的違約風險。(陳木在、陳錦村, 2001)。

消費者貸款風險的評估因素歸納為三項。其一為穩定性(Stability),指

² 參考汪海清、黃景泰、謝維國、楊培宏、王南豪, 1999, 消費者貸款實務, 台北市: 台灣金融研訓院。

申貸人職業特性、任職期間及個人本身特質所作的考慮(劉長寬, 2003)。此即為本研究之顧客基本資料類之變數。其二償債能力(Ability to Pay), 包括對於申貸者所得來源、所得水準及持續性負債多寡以及申貸金額之綜合考慮。其三為還款意願(Willing to Pay)指運用所得以償還借款之意願, 過去的還款紀錄可做為參考指標, 銀行於承做貸款之前會先行查證申請人有無債信不良或違約紀錄, 若有則設法予以婉拒(劉長寬, 2003)。後二項即為本研究所採用之聯徵信用相關類變數。

信用風險評估方法常見的有以下五種方法, 分別為: 經驗法則(Rules of Thumb)、信用評等制度(Credit Grading System)、信用評分制度(Credit Scoring System)、混合評等與評分制度及專家系統法, 各評分方法之定義及其優缺點(龔昶元, 1998; 莊瑞珠, 2007)。綜合各種評估方法, 可知「信用評分制度」及「專家系統」方法較為客觀。「混合評等與評分制度」雖然結合「信用評等」和「信用評分制度」之優點, 但複雜度高; 而「專家系統」在方法建立的過程較為繁複, 並且實際執行上仍有許多困難有待克服, 成本也較高。因此五種方法中以「信用評分制度」較具實用性。故本研究根據信用評分制度之觀念, 建立消費者小額信用貸款風險評估模型。

二、建模方法

在建模方法之使用上, 近幾年在消費者小額信用貸款之相關研究中, 有部分研究者皆採用邏輯斯模型建立風險評估模型(陳鴻文, 2002; 范哲銘, 2003; 戴堅, 2004; 林勉今, 2004; 許育嘉, 2005; 張維仁, 2005; 羅金川, 2004), 部分學者採用決策樹模型(Glorfeld & Hardgrave, 1996; Joos, et. al., 1998; 吳振晃, 2003), 有部分學者採用類神經網路(江世傑, 2001; 吳振晃, 2003; 洪榮隆, 2003)及Logit模型(林建州, 2001; 蔡明憲, 2002; 張明哲, 2003), 極少數研究採用描述性統計法進行信用貸款風險之研究(陳宗豪, 2000; 何貴清, 2002)。

在上述各方法中, 若僅採用描述性統計法進行信用貸款風險似乎過分簡單, 而且較無法同時將多個變數的訊息一併考量分析, 以進行違約機率之預測。區別分析及類神經網路雖然常可以做出很準確地預測, 但區別分析之資料要求需要符合常態性、獨立性及均質性等假設, 在實務上不易達到(陳順宇, 2005), 且區別分析的模型無法解釋變數間共線性的問題(Henley & Hand, 1996;

莊瑞珠, 2007); 而類神經網路其計算的過程為一黑箱作業(黃宇翔、毛紹睿, 2005; 謝邦昌, 2005), 無法明確地掌握自變數對違約機率之影響, 因此無法符合新巴塞爾新資本協定的要求, 亦較無法透過模式去解讀顧客的特質及行為; 此外, 類神經網路之演算效率常不佳, 江世傑(2001)就曾僅用 30 筆資料進行消費者貸款違約之預測模型訓練, 其訓練模型之時間竟高達 20 分鐘。故區別分析及類神經網路此二種方法較少正式被銀行界正式採用。而邏輯斯迴歸雖然預測能力極佳, 且無區分分析模型之限制, 但是當其建模變數為間斷型變數資料時, 則必須將其轉換成虛擬變數始得以進行建模, 容易形成維度過大的現象, 而且分判點之決定亦會影響模型的預測力(范哲銘, 2003)。

Glorfeld & Hardgrave (1996) 運用區別分析、Ratchet 口袋演算法 (Pocket Algorithm with Ratchet)、邏輯斯迴歸、決策樹 ID3、遞迴分割分析 (Recursive Partition Analysis)、倒傳遞類神經網路 (Standard Back-propagation)、串接網路迴歸 (Cascaded Network Regression)、記憶推理法 (Cased-based Reasoning, CBR) 八種方法, 針對美國阿肯色州西北國際銀行消費者放款資料共 40 筆, 其中核可資料 20 筆, 駁回資料 20 筆, 進行放款研究。其結果顯示, 決策樹 ID3 在模型訓練階段準確率為 100%, 排名第一。Joos, et. al. (1998) 運用決策樹針對比利時某家大型銀行 1991 至 1993 年的借款人信用資料共計 3,187 筆(其中正常戶 2,473 筆, 違約戶 714 筆), 研究結果顯示決策樹的準確率為 81.3%。吳振晃(2003)以資料採礦技術進行銀行消費者貸款授信之研究, 採用決策樹及類神經法, 針對 58,535 筆(正常戶 49,298 筆, 違約戶 9,237 筆, 違約率為 18.74%)資料進行是否發生逾期違約之預測, 結果發現決策樹法其預測正確性較類神經法為佳。Ikizler & Guvenir (2002) 以決策樹 C4.5 法進行銀行放款顧客資料之逾期還款預測, 其資料分為訓練資料集 1300 筆(其中 909 筆為正常戶, 391 筆為逾期還款戶), 測試資料集 143 筆, 其研究結果發現當借款期間大於 8 年且職業為自由業且無車子者, 其可能發生逾期還款的機率為 80.9%。

由上述研究可發現, 決策樹模型在與其他模型相較時, 其預測的準確率極佳。決策樹法在建模過程中一次僅取一個建模變數, 可有架構地找到不同風險等級下之重要影響因素組合, 其產生的訊息最容易解讀(謝邦昌、易丹輝, 2003; 謝邦昌, 2005)。另外, 當建模變數為間斷型變數型態時, 決策樹法無須將其轉換為虛擬變數, 因為本研究的資料包含大量之間斷型變數, 故本研究將採用決策樹法進行違約風險之預測。

三、變數共線性之問題

另外，在建立預測模型時，若變數間出現高相關性之共線性現象，會使被選入模型的變數有相互的連帶影響性，而使模式的預測力失真（林真真、鄒幼涵，1993）。上述研究在變數之選用時較未考量變數間之共線性問題，而因素分析（Factor Analysis）常可用於將變數依相關性分群（陳順宇，2005；Sharma, 1996），是以，本研究為了在眾多共線性高的變數中選擇出對信用貸款違約最具影響性之變數，提出了三種篩選及整合法用以解決變數間共線性的問題，這三種方法分別為相關程度法、因素篩選法及因素主成份法（馮志剛，1996；鄭廳宜，1999）。

四、建模變數之選擇

在建模變數之選擇上，絕大部分的研究主要採用的建模變數皆為顧客的基本特性相關資料。因為文獻較多，本研究將對逾期違約具顯著影響力之基本資料類變數彙整如表 1 所示。由表 1 可得知，各研究的結果所得的結果不盡相同，其中「月（年）所得或其衍生性變數」及「性別」為最多研究指出之重要影響變數，其他對逾期違約具影響力之變數依序為「學歷」、「工作年資」、「年齡」、「職業」、「婚姻及家庭狀況」、「居所及地區別」、「服務公司或其分類」及「職稱或職稱」另外，陳宗豪（2000）及范哲銘（2003）的研究中發現「是否願意為家人所知」亦為對信用貸款違約之顯著影響性變數。

除了顧客的基本特性相關資料之外，亦有眾多研究採用聯徵中心之資料作為預測信用貸款違約之建模變數，其中又以「他行查詢次數或家數」（陳宗豪，2000；蔡明憲，2002；陳鴻文，2002；范哲銘，2003；洪榮隆，2003；戴堅，2004；羅金川，2004）為最常被各研究中提及。亦有一些研究指出，「信用卡張數及額度」（陳宗豪，2000；何貴清，2002；陳鴻文，2002；洪榮隆，2003）為對信用貸款違約之顯著影響變數。其他如「是否使用循環利息」（陳鴻文，2002；戴堅，2004）、「現金卡張數」（戴堅，2004）及「信用卡現欠金額」（陳宗豪，2000），亦有學者提出其為信用貸款違約具顯著影響性之變數。

表 1 對逾期違約具顯著影響力之基本資料類變數彙整表

變數名稱	曾提列該變數對逾期違約具顯著影響力之研究
性別	葉秋南，1997；陳宗豪，2000；江世傑，2001；蔡明憲，2002；何貴清，2002；陳鴻文，2002；范哲銘，2003；洪榮隆，2003；張明哲，2003；林勉今，2004；許育嘉，2005；Liang, Lin & Liao, 2007
月(年)所得或其衍生性變數	陳宗豪，2000；江世傑，2001；林建州，2001；蔡明憲，2002；何貴清，2002；吳振晃，2003；洪榮隆，2003；林勉今，2004；許育嘉，2005；Steenackers & Goovates, 1989；Ikizler & Guvenir, 2002
學歷	江世傑，2001；林建州，2001；蔡明憲，2002；范哲銘，2003；羅金川，2004；戴堅，2004；許育嘉，2005；Liang, Lin & Liao, 2007
工作年資	陳宗豪，2000；何貴清，2002；陳鴻文，2002；范哲銘，2003；林勉今，2004；羅金川，2004；Steenackers & Goovates, 1989
年齡	葉秋南，1997；陳宗豪，2000；江世傑，2001；蔡明憲，2002；張明哲，2003；Steenackers & Goovates, 1989；Liang, Lin & Liao, 2007
職業	葉秋南，1997；陳宗豪，2000；江世傑，2001；蔡明憲，2002；洪榮隆，2003；Steenackers & Goovates, 1989；Ikizler & Guvenir, 2002；Liang, Lin & Liao, 2007
婚姻及家庭狀況	何貴清，2002；蔡明憲，2002；林勉今，2004；張維仁，2005；Ikizler & Guvenir, 2002；Liang, Lin & Liao, 2007
居所及地區別	何貴清，2002；林勉今，2004；張維仁，2005；Ikizler & Guvenir, 2002；Steenackers & Goovates, 1989；Liang, Lin & Liao, 2007
服務公司或其分類	林建州，2001；張明哲，2003；林勉今，2004；張維仁，2005；Steenackers & Goovates, 1989
職稱或職稱	陳宗豪，2000；江世傑，2001；林建州，2001；何貴清，2002
是否願意為家人所知	陳宗豪，2000；范哲銘，2003

資料來源：本研究整理

此外，一些研究採用了信用貸款產品本身相關之資料進行貸款違約之預測，並指出「貸款期間」(范哲銘，2003；Steenackers & Goovates, 1989；Ikizler & Guvenir, 2002；Liang, Lin & Liao, 2007)、「貸款餘額」(范哲銘，2003；張維仁，2005；Ikizler & Guvenir, 2002)、「借款用途」(陳宗豪，2000；吳振晃，2003)、「貸款利率」(范哲銘，2003；林勉今，2004)、「攤還方式」(吳振晃，2003；Liang, Lin & Liao, 2007)、「擔保內容」(吳振晃，2003)、「是否有保證人」、「是否自動轉帳繳款」(Liang, Lin & Liao, 2007)等為具顯著影響性變數。

亦有少數研究發現「擁有的不動產總值」(葉秋南, 1997; 陳宗豪, 2000; 江世傑, 2001; Steenackers & Goovates, 1989)、「過去借款及保證狀況或個數」(江世傑, 2001; Steenackers & Goovates, 1989)、「支票存款開戶數」(何貴清, 2002)、「負債比率」、「信用狀況」(林勉今, 2004; 許育嘉, 2005)、「是否願意接受交叉銷售活動」(Liang, Lin & Liao, 2007) 等為對信用貸款違約具顯著影響性之變數。

由前述的文獻探討及彙整可發現, 在消費者信用貸款的相關文獻中, 大多研究者於主要影響變數中都採用了顧客基本資料, 其中尤以性別、收入、學歷、工作年資、年齡及職業等為最常被採用且具有顯著影響力之變數。另外, 有部分研究採用取自於聯徵中心之個人信用相關變數, 做為重要顯著影響變數。是以本研究主要考慮採用此兩大類變數作為建模變數, 總共採用了 25 個基本資料變數³, 40 個來自於聯徵中心的信用相關變數⁴, 所採用的變數幾乎囊括了上述各文獻所提及的各類對信用貸款違約之顯著影響性變數, 相較於過去國內外各研究而言, 本研究所考量的變數最為廣泛完整。

本研究根據上述文獻探討, 擬將採用顧客之基本資料、聯徵狀況中的個人信用相關資料為建模考慮變數, 再利用 CART (Classification and Regression Tree) 建立消費貸款信用風險評估模型。

³ 25 個基本資料變數為：性別、學歷、年齡、現任年收入、現住房屋型態(居住狀況)、戶籍地是否與現居地相同、近親關係、保證人婚姻狀況、是否提供近親姓名、婚姻狀況、婚姻與小孩關係、現任公司地址(縣市)、戶籍地址(縣市)、現居地(縣市)、是否提供戶籍地電話、是否提供現居地電話、是否提供行動電話、是否提供現任公司電話、是否提供朋友電話、現任公司營業項目、營業型態、工作層級、聘僱關係、現任職稱、現任年資。

⁴ 40 個聯徵信用相關變數為：正卡是否強停、附卡是否強停、繳款狀況、曾經授信異常註記、發放銀行(活卡)家數、首張卡持卡月數、活卡持卡月數、正卡強停距現在月數、附卡強停距現在月數、正卡張數、附卡張數、有效卡普卡張數、有效卡金卡張數、有效卡白金卡張數、有效卡普卡張數比例、有效卡金卡張數比例、有效卡白金卡張數比例、具預借現金的信用卡張數、信用總額度、最高信用額度、最低信用額度、平均信用額度、目前持卡之總掛帳總金額、信用總額度與總掛帳總金額的差值、過去 N 月無消費次數比率、過去 N 月全額繳清次數比率、過去 N 月循環無逾期次數比率、過去 N 月逾期次數比率、過去 N 月循環無逾期月數比率、過去 N 月逾期月數比率、循環掛帳比率、預借現金比率、預借現金月數比率、逾期次數、最近授信異常距現在月數、循環無逾期最後一次距今月數、預借現金最後一次距今月數、逾期最後一次距今月數、最近三個月他行查詢銀行家數、最近他行查詢距現在天數。

五、建模資料量之探討

張仁哲（1982）以某金融機構對個人消費貸款者為研究對象，隨機抽樣 100 個優良顧客與 33 個不良顧客進行研究。Glorfeld & Hardgrave（1996）針對美國阿肯色州西北國際銀行消費者放款資料共 40 筆，其中核可資料 20 筆，駁回資料 20 筆，進行放款研究。江世傑（2001）針對台灣中部某銀行進行消費者貸款信用評等之研究，其所採用的資料僅 30 個個案。林建州（2001）亦曾以國內某金融銀行為研究對象，抽取正常戶 200 件，違約戶 200 件，共 400 件建立信用風險審核模式。陳鴻文（2002）曾針對台灣南部某銀行顧客進行信用貸款違約之研究，其研究資料採用正常顧客 150 件，呆帳顧客 150 件。范哲銘（2003）利用羅吉斯迴歸及區別分析，針對銀行體系與行銷體系的授信案件，各抽取正常繳息之案件 150 件，滯欠案件 150 件，共計 600 個樣本資料，而其用以建模變數共有 16 個。張明哲（2003）以國內某金融機構為對象，蒐集樣本逾期戶 157 件、正常戶 469 件，共計 626 件。戴堅（2004）以國內某金融機構為對象，抽取樣本正常戶 150 件、違約戶 150 件，共計 300 件進行個人消費性信用貸款之授信評量模式建構。林勉今（2004）以某銀行消費者貸款案件為研究對象，以 2001 年 1 月到 12 月底於該銀行辦理消費貸款案件為抽樣母體，為了建立模型，由正常貸款抽取 231 件，逾期案件抽取 231 件，共計 462 件。羅金川（2004）以國內某一銀行之個人消費性信用貸款為資料抽樣對象，總共抽樣出 300 個樣本進行建模研究，其中 150 個為不良戶，150 個為正常戶。

上述各研究所獲取的模型雖然預測的正確率皆不錯，但其所採用的建模資料集資料筆數少，且樣本中正常戶與非正常戶（違約戶）的占比大多各占一半。而在銀行授信管理實務上，正常戶相較於違約戶應為較稀少事件（Rare Event），而該研究在建模資料集中過分放大違約之占比，會使所建構出之模型預測力佳，但實際上有誇大預測能力之嫌。若該等模式若要於實務上採用，仍需進行調整及修正。依據 Zavgren（1985）、Shi（2006）及 Liang, Lin & Liao（2007）的研究指出，若遇稀少事件而使樣本結構無法與資料庫結構一致時，則正常戶與違約戶在樣本資料中之占比，亦非以 1:1 為最佳組合。因綜觀國內外多變相關研究，其資料大多僅幾百筆，易會造成模型分類效果或被選取出之建模變數不具代表性。本研究為避免預測力被誇大不實及結論不具代表性等現象，與國內某銀行進行合作研究，取得小額信貸有效資料共計 10,285 筆，其中有 1,672 筆為違約逾期戶，8,613 筆為正常繳交每期攤還金額戶，該等資料與該銀行實際資料結構相符。

參、研究設計

六、實證分析流程架構

本研究將以 A 銀行 2003 年 6 月至 2004 年 6 月已核貸之貸款人所申辦之消費者小額信用貸款相關資料為研究標的，有效資料共計 10,285 筆，其中有 1,672 筆為違約逾期戶及 8,613 筆為正常繳交每期攤還金額戶。其整體違約逾期率為 16.26%。

本研究之實證分析流程架構圖如下圖 2 所示。本研究首先探討 A 銀行信用貸款個人戶之基本資料及信用相關整合變數與是否違約逾期之關係，並進行建模變數之篩選，再將資料集分割成訓練和測試資料集後，運用 CART 分析方法建立信用貸款違約風險評估模型。

七、名詞操作型定義

(一)違約逾期戶定義

本研究違約逾期戶的定義為在 2004 年 6 月 30 日以前曾經有逾期天數超過 90 天以上未繳足最低應繳款者。

(二)預期違約率

本研究將 CART 模型葉子節點上的違約逾期比例定義為預期違約率。以最終葉子節點上的違約逾期比例，估計每一個顧客的違約機率，同一葉子節點內的顧客其預期違約率皆相同。

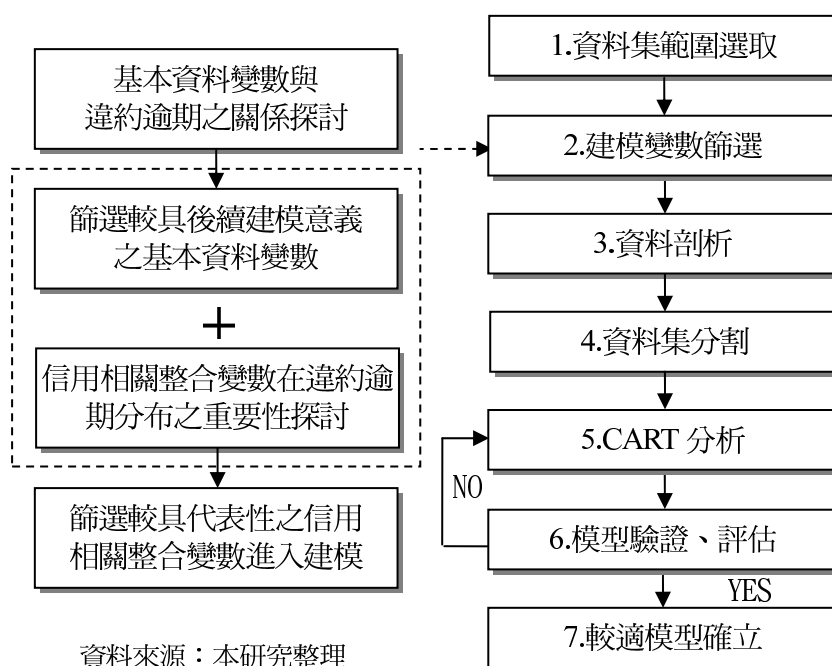


圖 2 實證分析流程架構圖

(三)建模變數

1.基本資料變數

係指 A 銀行透過信用貸款顧客所填具之貸款申請書中所取得之個人相關基本資料。

2.信用相關整合變數

係指 A 銀行透過聯合徵信中心所取得之個人信用狀況資料,以及 A 銀行自行將信用狀況資料再行產生之整合性變數。

八、建模方法：分類與迴歸樹(Classification And Regression Tree)

本研究主要是利用 CART 來分析信用貸款顧客資料集透過二元遞迴切割的基本原理過程，CART 會依照自變數與其相對應的屬性，將訓練資料集切割成數個兩兩相對的類別，並將其切割過程彙總成一連串的規則(樹狀結構圖)；在其切割運算過程中，會產生末端節點數最多的最大樹狀結構，然後，CART

本身會再依照使用者所設定之的準則來進行決策樹的修剪，直到滿足停止修剪的準則為止；在 CART 的運算過程中，雖然較傳統決策樹在切割時一旦達到設定之準則即停止切割，所耗費的運算時間久但能避免資訊隱藏的缺點；並且，CART 以樹狀結構呈現分析結果，讓使用者清楚了解到模型的重要變數及其切割的效果。（Quinlan, 1993; Berry & Linoff, 1997; SAS, 2005）

CART 分析流程依續大致可分為三個步驟⁵：

(一)步驟一：建構最大樹狀結構

CART 切割點的選取準則是每個節點的異質性(Heterogeneity)高低。評估節點異質性的指標有三種，分別是 Twoing，Entropy 及 Gini，其中 Gini Index 較為普遍，是以本研究將採用 Gini Index 做為評估節點異質性的指標。Gini Index 的計算方式如下：

$$Gini(t) = 1 - \sum_i p_i^2$$

其中 t 為已知節點 t；Pi 為第 t 個節點中第 i 個類別的機率。當 Gini 值越小時代表在該切割點所產生的子節點中組成份子越單純，即該切割點判別不同類別差異的能力越好，因此所造成之誤判率也越低。

(二)步驟二：修剪樹狀結構並產生子樹群

在找出最大樹結構後，CART 便開始修剪樹狀結構以降低誤判率，並找出最適當的決策樹，決定最適決策樹所依據的準則為誤判率(Error Rate)，其計算方式如下：

$$\text{節點誤判率} = \frac{\text{節點中分類錯誤的資料筆數}}{\text{節點中總資料筆數}}$$

CART 會計算最大樹與所有子樹的誤判率，而最適的決策樹是具有最低誤判率的樹狀結構。在決策樹修剪過程中亦會考量重代估計值(Resubstitution Estimate)。當重代估計值越小時，代表末端節點個數越多，

⁵ 參考下列：陳姍霓，2004，整合類神經網路、多元適應性雲形迴歸與分類迴歸樹於信用評等模式之建構-以房屋貸款為例，輔仁大學管理學研究所碩士論文

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, **Classification and Regression Trees**, New York: Chapman & Hall.

即樹狀結構中切割點越多，樹狀結構也越大。另外，本研究設定每個葉子節點樣本數大於等於 30。

(三)步驟三：挑選最佳樹狀結構

完成前述步驟後，將測試樣本再次代入所有可能的樹狀結構並計算每個樹狀結構的誤判率，最後再選取最小誤判率的樹狀結構為最佳樹狀模型。

肆、違約風險評估模型建構及預測

一、建模變數選取

本研究首先依據資料型態將變數區分為間斷及連續二類，在間斷類之變數中，將依卡方獨立性檢定、Goodman and Kruskal tau 及 Cramer's V 值選，取出對消費者信用貸款違約逾期具有顯著影響力之變數用以建模。在連續型之變數中，將初步依獨立樣本平均數 T 檢定及 Eta Squared 值中，選取出對違約逾期具有顯著影響力之變數；接著考量顯著性變數之間的共線性狀況，運用三種不同方式(違約逾期相關程度法、因素分群篩選法及因素後主成份法)進行變數選取或縮減，並採用選取或縮減後之變數用以建模。整體步驟如圖 3。

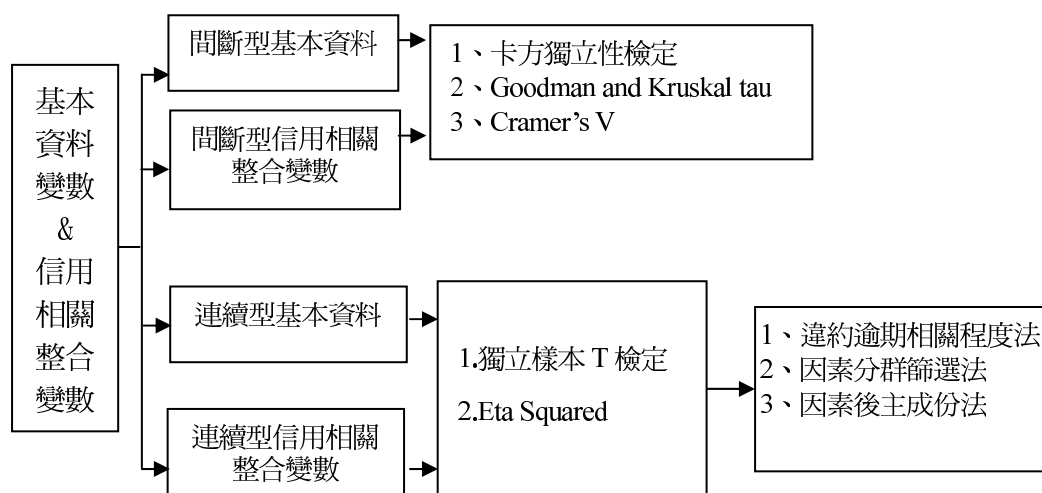


圖 3 建模變數選取過程示意圖

(一)間斷型建模變數選取

1.間斷型基本資料變數之篩選

本研究所用資料庫中間斷型基本資料變數共有 22 個。透過卡方獨立性檢定(Weiss,2002)，並考慮 Goodman and Kruskal tau 值與 Cramer's V 值(Kass,1980)之後，最後選取出本研究用以建模的基本資料變數共 8 個，如表 2 所示。

2.間斷型信用相關整合變數之篩選

間斷型信用相關整合變數共有 4 個，分別為「正卡是否強停」、「附卡是否強停」、「繳款狀況」及「曾經授信異常註記」等。本研究亦依上述步驟進行篩選，最後選用了「正卡是否強停」及「繳款狀況」此兩變數投入建模(表 2)。

(二)連續型建模變數選取

1.連續型基本資料變數之篩選

本研究所用資料庫中連續型基本資料變數有「年齡」、「年資」及「年收入」等 3 個。藉由平均數檢定，並輔以 Eta Squared 值 (Kass, 1980) 判讀，最後本研究僅選取「年齡」變數用以建模。

2.連續型信用整合變數之篩選

本研究之連續型信用相關整合變數共計有 36 個，本研究篩選出在是否違約逾期上存在顯著差異($P\text{-value} \leq 0.05$)之變數共計 25 個(表 3)。

接著再將上述所篩選出的 25 個連續型信用相關整合變數，以「違約逾期相關程度法」、「因素分群篩選法」及「因素後主成份法」，進一步篩選及整合信用相關變數。詳細過程說明如後。

(1)違約逾期相關程度法

本研究所篩選出之 25 個變數進行兩兩配對相關性檢定，將變數間相關係數大於等於 0.4 視為相似屬性的變數。若變數與是否違約逾期之 Eta Squared 值大於 0.1 則將其選入做為建模變數。針對 Eta Squared 值小於 0.1 之變數，再由相似屬性的變數中，選取其與是否違約逾期之 Eta Squared 值最大者，做為代表用以建模的變數(表 3)。

表 2 間斷型建模變數與目標變數之檢定結果

項目		Pearson Chi-Square	P-value	Goodman and Kruskal tau	Cramer's V
基本資料變數	性別	187.3423	0.0000	0.0182	0.1350
	學歷	138.2016	0.0000	0.0134	0.1159
	現住房屋型態(居住狀況)	16.4499	0.0115	0.0016	0.0400
	戶籍地是否與現居地相同	11.0054	0.0041	0.0011	0.0327
	近親關係	17.9635	0.0013	0.0017	0.0418
	現任公司地址(縣市)	98.3093	0.0000	0.0096	0.0978
	是否提供戶籍地電話	4.8045	0.0284	0.0005	0.0216
	現任職稱	27.1641	0.0013	0.0026	0.0514
信用相關整合變數	正卡是否強停	11.9546	0.0025	0.0012	0.0341
	繳款狀況	34.1891	0.0000	0.0033	0.0577

表 3 連續型建模變數與目標變數之檢定結果

項目		相似 ¹ 屬性分群	變異數同質性檢定 P-value	平均數檢定 P-value	Eta Squared	違約逾期相關程度法之選擇變數
基本資料	1.年齡	—	0.0000	0.0000	0.0103	—
信用相關整合變數	循環掛帳比率	2	0	0	0.4878	●
	過去 N 月循環無逾期次數比率	2	0.0703	0	0.2965	●
	預借現金比率	2	0	0	0.2899	●
	平均信用額度	1	0	0	0.1964	●
	信用總額度與總掛帳總金額的差值	4	0	0	0.1772	●
	信用總額度	4	0	0	0.1659	●
	過去 N 月全額繳清次數比率	S	0	0	0.1551	●
	目前持卡之總掛帳總金額	2	0	0	0.1014	●
	過去 N 月無消費次數比率	S	0	0	0.0279	●
	最近三個月他行查詢銀行家數	S	0	0	0.0611	●
	過去 N 月逾期次數比率	S	0.0093	0.0246	0.0507	●
	首張卡持卡月數	1	0	0	0.0265	●
	活卡持卡月數	1	0	0	0.0233	
	有效卡金卡張數比例	1	0	0	0.0117	
	有效卡白金卡張數比例	1	0	0	0.006	
	有效卡白金卡張數	1	0	0.0024	0.0038	
	有效卡金張數	1	0	0	0.0033	
	發放銀行(活卡)家數	1	0.299	0.0365	0.0028	
	具預借現金的信用卡張數	2	0	0	0.0147	●
	預借現金月數比率	2	0.0061	0	0.0134	
	最近他行查詢距現在天數	3	0	0.0038	0.0279	●
	循環無逾期最後一次距今月數	3	0	0	0.0054	
預借現金最後一次距今月數	3	0	0	0.005		
最高信用額度	4	0.0095	0	0.0475	●	
最低信用額度	4	0.0122	0	0.0269		

1:將變數相關程度大於 0.4 者畫分為同一群，S 表示該變數與其他變數之相關係數皆未大於 0.4

另外,若某一連續型信用相關整合變數與其他變數間之相關係數皆未大於等於 0.4,但其與是否違約逾期存在顯著相關者,則此變數將被視為具獨特變數,直接列入建模變數中。依上述原則,本研究最後共選擇了 15 個變數參與建模(表 3)。

(2)因素分群篩選法

藉由因素分析可將變數區隔成數群內部相關性高的變數族群。本研究將從每一群中挑選符合設定條件之變數作為建模之自變數,以期能由相關性高之自變數中,選取或整合出具代表性之建模變數(Sharma, 1996)。本研究將所篩選出之 25 個變數進行因素分群,其中因素萃取方法採用主成份法,旋轉方法為變異最大旋轉法(Varimax)。並以共通性大於等於 0.4 者做為變數是否選擇進入因素分群的準則。在 25 個變數中僅有「過去 N 月逾期次數比率」其共通性的萃取是小於 0.4,故此變數將被獨立考慮進入建模變數中。(表 4)

因素分群結果主要將變數區分為五群,而在最後一群中僅有「最近三個月他行查詢銀行家數」變數單獨存在,故本研究亦將之選擇獨立考慮進入建模變數;剩餘的四群因素分群結果本研究將之分別命名如下「持卡行為群」、「借款行為群」、「信用狀況群」及「時間影響群」等四群。

對同一群內的變數中,本研究將選取其與是否違約逾期之檢定 t 值及 Eta Squared 值最大者做為代表用以建模的變數。在「持卡行為群」中選出變數為「平均信用額度」;「借款行為群」中選出變數為「循環掛帳比率」;「信用狀況群」中選出變數為「信用總額度與總掛帳總金額的差值」;「時間影響群」中選出變數為「最近他行查詢距現在天數」。

(3)因素後主成份法

由於主成份分析可將多個相關的變數簡化成為少數幾個獨立的線性組合的主成份指標(Sharma, 1996)。故本研究將利用因素分析將變數分群,再逐一將每群內的變數透過主成份分析創造總指標,其主要目的除了整合自變數間之相關性並降低自變數個數外,尚可避免將可用的變數資訊遺漏。

表 4 因素分群結果

因素	變數	Eta Squared	因素負荷	特徵值	累積解釋變異量
持卡行為	首張卡持卡月數	0.0265	0.9040	7.5912	23.8087
	平均信用額度	0.1964	0.8761		
	活卡持卡月數	0.0233	0.8685		
	發放銀行(活卡)家數	0.0028	0.8592		
	有效卡金張數	0.0033	0.8345		
	有效卡金卡張數比例	0.0117	0.8158		
	有效卡白金卡張數	0.0038	0.6830		
	有效卡白金卡張數比例	0.0060	0.6455		
借款行為	過去 N 月循環無逾期次數比率	0.2965	0.8623	4.1394	44.6572
	預借現金月數比率	0.0134	0.8608		
	循環掛帳比率	0.4878	0.8470		
	預借現金比率	0.2899	0.8033		
	具預借現金的信用卡張數	0.0147	0.8033		
	目前持卡之總掛帳總金額	0.1014	0.7541		
信用狀況	信用總額度與總掛帳總金額的差	0.1772	0.7978	2.1216	56.2765
	最高信用額度	0.0475	0.6964		
	信用總額度	0.1659	0.6779		
	過去 N 月全額繳清次數比率	0.1551	0.5789		
	最低信用額度	0.0269	0.5629		
	過去 N 月無消費次數比率	0.0279	0.5286		
時間影響	最近他行查詢距現在天數	0.0279	0.8503	1.9863	65.2936
	循環無逾期最後一次距今月數	0.0054	0.8438		
	預借現金最後一次距今月數	0.0050	0.7229		
因素 5	最近三個月他行查詢銀行家數	0.0611	0.4918	1.1842	70.9286

前述因素分析已將變數分成「持卡行為群」、「借款行為群」、「信用狀況群」及「時間影響群」等四群，接著將嘗試運用主成份分析來整合各群內高相關性之變數。在針對相關矩陣求取特徵值 (Eigenvalue) 與特徵向量 (Eigenvector) 後，保留特徵值大於等於 1 的主要成份，做為建模變數。表 5 顯示透過主成份分析，可將

因素所歸類的四群變數整合出六個主成份指標。若再加上「最近三個月他行查詢銀行家數」此一單獨成群變數及共同性小於0.4的「過去 N 月逾期次數比率」，最後本研究最後共萃取出八個建模變數。

(4)變數篩選結果之彙整

三種連續型信用相關整合變數篩選結果彙整表 6。

表 5 連續型信用卡相關整合建模變數-因素後主成份

主成份指標名稱		主成份方程式	Eigenvalue	累積解釋變異
持卡行為	主成份指標 1	$=-0.3730*發放銀行(活卡)家數-0.3836*首張卡持卡月數-0.3683*活卡持卡月數-0.3615*有效卡金卡張數-0.3153*有效卡白金卡張數-0.3418*有效卡金卡張數比例-0.2957*有效卡白金卡張數比例-0.3791*平均信用額度$	5.5773	82.2169
	主成份指標 2	$=-0.1140*發放銀行(活卡)家數-0.1713*首張卡持卡月數-0.2310*活卡持卡月數-0.1632*有效卡金卡張數+0.5856*有效卡白金卡張數-0.3336*有效卡金卡張數比例+0.6501*有效卡白金卡張數比例-0.0277*平均信用額度$	1.0000	
借款行為	主成份指標	$=-0.4174*具預借現金的信用卡張數-0.3804*目前持卡之總掛帳總金額-0.4091*過去 N 月循環無逾期次數比率-0.4093*循環掛帳比率-0.3955*預借現金比率-0.4357*預借現金月數比率$	4.2273	70.4544
信用狀況	主成份指標 1	$=-0.5020*信用總額度-0.5090*最高信用額度-0.3827*最低信用額度-0.4628*信用總額度與總掛帳總金額的差值-0.2268*過去 N 月無消費次數比率-0.2774*過去 N 月全額繳清次數比率$	2.9395	67.7386
	主成份指標 2	$=-0.3063*信用總額度-0.3243*最高信用額度-0.1303*最低信用額度+0.1331*信用總額度與總掛帳總金額的差值+0.6800*過去 N 月無消費次數比率+0.5513*過去 N 月全額繳清次數比率$	1.1248	
時間影響	主成份指標	$=-0.6124*循環無逾期最後一次距今月數-0.5251*預借現金最後一次距今月數-0.5909*最近他行查詢距現在天數$	2.0580	68.5995
最近三個月他行查詢銀行家數				
過去 N 月逾期次數比率				

表 6 三種連續型信用相關整合變數建模變數一覽表

違約逾期相關程度法	因素分群篩選法	因素後主成份法
過去 N 月逾期次數比率	過去 N 月逾期次數比率	過去 N 月逾期次數比率
最近三個月他行查詢銀行家數	最近三個月他行查詢銀行家數	最近三個月他行查詢銀行家數
平均信用額度	平均信用額度	持卡行為主成份指標 1
信用總額度與總掛帳總金額差值	信用總額度與總掛帳總金額差值	持卡行為主成份指標 2
循環掛帳比率	循環掛帳比率	借款行為主成份指標
信用總額度	最近他行查詢距現在天數	信用狀況主成份指標 1
過去 N 月無消費次數比率		信用狀況主成份指標 2
過去 N 月全額繳清次數比率		時間影響主成份指標
目前持卡之總掛帳總金額		
過去 N 月循環無逾期次數比率		
預借現金比率		
首張卡持卡月數		
具預借現金的信用卡張數		
最近他行查詢距現在天數		
最高信用額度		

二、資料集描述與分割

本研究目標變數為「是否違約逾期」，違約逾期資料筆數占整體資料集之比例為 16.26%。本研究將違約逾期定義為「成功」之事件進行建模及預測，即定義違約逾期為 1，定義非違約逾期為 0。為確保建模及測試資料集的代表性，本研究以「性別」與「是否違約逾期」二變數為分層變數，將原始資料集切割為 70%之訓練及 30%之測試資料集，並以卡方齊一性檢定驗證。由表 7 得知，二資料集在「性別」與「是否違約逾期」二變數之分配結構上與全資料集無差異(P-value 皆大於 $\alpha=0.05$)。

表 7 資料集分割結果與齊一性檢定結果

資料集	類別	計數	是否違約逾期		總計	齊一性檢定 P-value	
			否	是		性別	違約逾期比例
原始資料集	女	人數	3,906	456	4,362	--	--
		總和百分比	37.98	4.43	42.41		
	男	人數	4,707	1,216	5,923		
		總和百分比	45.77	11.82	57.59		
	總計	人數	8,613	1,672	10,285		
		總和百分比	83.74	16.26	100.00		
70% 訓練資料集	女	人數	2,717	319	3,036	0.7111	0.8956
		總和百分比	37.75	4.43	42.18		
	男	人數	3,311	850	4,161		
		總和百分比	46.01	11.81	57.82		
	總計	人數	6,028	1,169	7,197		
		總和百分比	83.76	16.24	100.00		
30% 測試資料集	女	人數	1,189	137	1,326	0.5434	0.9866
		總和百分比	38.50	4.44	42.94		
	男	人數	1,396	366	1,762		
		總和百分比	45.21	11.85	57.06		
	總計	人數	2,585	503	3,088		
		總和百分比	83.71	16.29	100.00		

三、模型建構結果

本研究利用 70%訓練資料集進行模型之建構，將基本資料及信用相關整合變數等自變數利用 CART 建立模型。其中模型一包含基本資料及信用相關整合變數之「違約逾期相關程度法」等自變數，模型二包含基本資料及信用相關整合變數之「因素分群篩選法」等自變數，模型三則包含基本資料及信用相關整合變數之「因素後主成份法」等自變數。在三種模型建構完成後，本研究除了將比較各模型在 70%訓練資料集與 30%之測試資料集中之預測與判斷能力，另將利用 ROC 及 Cumulative Lift 兩基準來比較決定本研究之最適模型。其模型建構示意圖如圖 4 所示。

(一)各模型重要建模變數

運用前述研究方法中所提及的 CART 方法及步驟，本研究三種決策樹模型所使用到之建模變數、變數使用次數與重要性指標彙整如表 8 所示。

(二)決策樹模型結果比較

1.決策樹預測能力綜合比較

為進行決策樹預測能力綜合之比較，本研究首先針對各決策樹所得的預期違約率設定一個切點，以便能將資料區分成預測違約逾期及預測非違約逾期二類。在考量整體預測正確率⁶、違約逾期⁷及非違約逾期的預測能力⁸後，本研究設定最適預期違約率切點值的判定準則為：整體預測正確率能達 75%以上、非違約逾期預測率 80%以上、違約逾期誤判率⁹約 50%左右。在此準則下各模型的切點值主要仍介於 0.20-0.25 之間(表 9 的第一欄)。為探討各模型的預測能力，本研究以混淆矩陣來表示決策樹模型在最適預期違約率切點下分類正確性結果。(表 9)

⁶ 整體預測正確率(Accuracy rate) (比率越高表示模型之整體預測準確率越高)

$$= \frac{\text{預測違約逾期且實際違約逾期人數} + \text{預測非違約逾期且實際非違約逾期人數}}{\text{總人數}} \times 100\%$$

⁷ 違約逾期預測率 (Recall rate) (此一比率表示模型預測為壞帳且實際為壞帳之人數佔實際總壞帳人數的比例，此比率越高表示模型中所能抓出之壞帳能力越好)

$$= \frac{\text{預測為違約逾期但實際為違約逾期的人數}}{\text{實際違約逾期總人數}} \times 100\%$$

⁸ 非違約逾期預測率(True Negative rate)(此一比率表示模型預測出好帳且實際為好帳之人數佔好帳實際總人數的比例，此比率越高表示模型中所能抓出的好帳人數愈多)

$$= \frac{\text{預測非違約逾期且實際非違約逾期人數}}{\text{實際非違約逾期總人數}} \times 100\%$$

⁹ 違約逾期誤判率 (False Negative rate) (此比率亦為一般混淆矩陣中所指的型一誤差，亦即代表損失成本之機率值，一般情況而言，此比率越低表示越好)

$$= \frac{\text{預測為非違約逾期但實際為違約逾期的人數}}{\text{實際違約逾期總人數}} \times 100\%$$

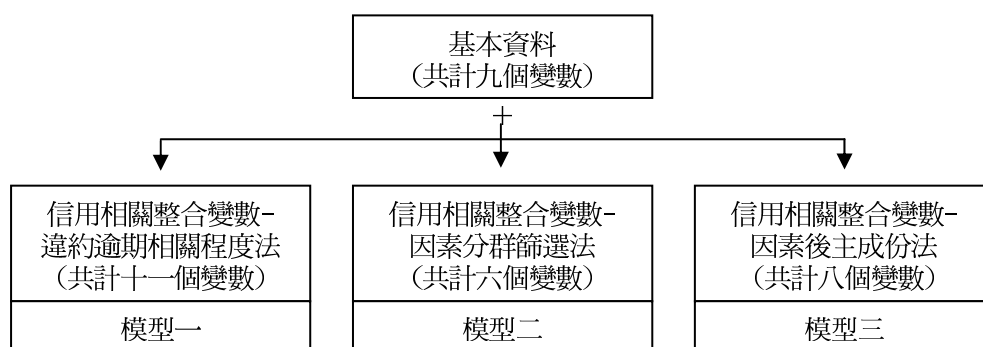


圖 4 模型建構示意圖

表 8 決策樹模型之建模變數、使用次數與重要性指標整理表

違約逾期相關程度法			因素分群篩選法			因素後主成份法		
變數	使用次數	重要性指標	變數	使用次數	重要性指標	變數	使用次數	重要性指標
最近三個月他行查詢銀行家數	5	1.0000	最近三個月他行查詢銀行家數	6	1.0000	最近三個月他行查詢銀行家數	3	1.0000
信用總額度與總掛帳總金額的差值	5	0.7025	信用總額度與總掛帳總金額的差值	2	0.6053	信用狀況主成份指標 1	6	0.7062
性別	2	0.3846	性別	1	0.3550	借款行為主成份指標	4	0.4026
學歷	2	0.3688	現任公司地址(縣市)	3	0.3315	性別	1	0.3835
現任公司地址(縣市)	3	0.3302	學歷	1	0.3048	信用狀況主成份指標 2	1	0.2108
過去 N 月全額繳清次數比率	5	0.3105	平均信用額度	3	0.2878	現任公司地址(縣市)	2	0.2081
具預借現金的信用卡張數	2	0.1772	循環掛帳比率	3	0.2613	學歷	1	0.1952
平均信用額度	1	0.1576						

表 9 決策樹訓練資料集模型預測能力綜合比較

模型與預期違約率切點 ¹	目標變數	預測/決策值	目標變數條件百分比		預測/決策值條件百分比		樣本數		總合百分比		整體分類正確性	
			訓練	測試	訓練	測試	訓練	測試	訓練	測試	訓練	測試
違約逾期相關程度法 0.2316	0	0	88.93%	88.84%	83.28%	83.48%	5,020	2,158	69.75%	69.88%		
	1	0	11.07%	11.16%	53.46%	53.88%	625	271	8.68%	8.78%	77.31	77.40
	0	1	64.95%	64.80%	16.72%	16.52%	1,008	427	14.01%	13.83%	%	%
	1	1	35.05%	35.20%	46.54%	46.12%	544	232	7.56%	7.51%		
因素分群篩選法 0.2149	0	0	88.59%	88.99%	82.66%	82.24%	4,983	2,126	69.24%	68.85%		
	1	0	11.41%	11.01%	54.92%	52.29%	642	263	8.92%	8.52%	76.56	76.62
	0	1	66.48%	65.67%	17.34%	17.76%	1,045	459	14.52%	14.86%	%	%
	1	1	33.52%	34.33%	45.08%	47.71%	527	240	7.32%	7.77%		
因素後主成份法 0.2154	0	0	88.69%	88.68%	82.48%	81.20%	4,972	2,099	69.08%	67.97%		
	1	0	11.31%	11.32%	54.23%	53.28%	634	268	8.81%	8.68%	76.52	75.58
	0	1	66.37%	67.41%	17.52%	18.80%	1,056	486	14.67%	15.74%	%	%
	1	1	33.63%	32.59%	45.77%	46.72%	535	235	7.43%	7.61%		

1:最適預期違約率切點值的判定準則為：整體預測正確率能達 75%以上、非違約逾期預測正確率 80%以上、違約逾期誤判率約 50%左右。

以測試資料結果而言，整體預測正確率在三種模型中以「違約逾期相關程度法-決策樹」的 77.40%較優，然就違約逾期預測率而言，則以「因素分群篩選法-決策樹」較佳有 47.71%正確性；若再以代表型一誤差的違約逾期誤判率而論，仍以「因素分群篩法-決策樹」最低 (52.29%)。(表 9)

本研究所設立的三個決策樹模式在整體預測正確率、違約逾期預測率及違約逾期誤判率上的差異極小，顯示三個決策樹預測能力是難分軒輊的。

2.測試資料集 ROC 曲線圖¹⁰

本研究將 30%之測試資料集代入所建構之模型後，各決策樹模型之 ROC 曲線圖如圖 5。在所建構的三種決策樹 ROC 曲線圖中，以

¹⁰ROC 曲線圖的橫軸 1-特異度(Singular)等於本研究之 1-違約逾期誤判率(型一誤差)，而其縱軸的敏感度(Sensitive)則代表著違約逾期預測率；當曲線愈凹向下或曲線下面積 (AUC) 愈大者表示模型在預測能力可容忍的錯誤中表現愈佳。

「違約逾期相關程度法-決策樹」在圖裡所呈現的情況較佳。

可由 ROC 曲線圖中 1-特異度=0.2 之後的曲線可得知，「違約逾期相關程度法-決策樹」其敏感度皆明顯高於其他兩者；而「因素分群篩選法-決策樹」及「因素後主成份法-決策樹」兩模型結果在 ROC 曲線圖中幾乎重疊，故較無法判斷其優劣。

3.測試資料集 Cumulative Lift 曲線圖¹¹

由圖 6 可知，「因素分群篩選法-決策樹」在前 15 個百分位數時，明顯地其抓取違約逾期的能力是較優於其他兩模型，而在超過前 20 個百分位數後，則以「違約逾期相關程度法-決策樹」所得之測試結果較佳，此點與利用 ROC 曲線圖判斷所得之結果是一致的。

(三)最適模型選取

本研究所建構之各決策樹模型之預測結果彼此差異並不大，很難斷定模型的優劣。但在考量 ROC 曲線圖及 Cumulative Lift 曲線圖中模型判斷之結果後，本研究選擇最適之決策樹預測模型為「違約逾期相關程度法-決策樹」，其最終葉子節點數為 26 個、最大深度為 10 層。該模型的整體預測正確率達 77%以上；違約逾期預測率為 46.12%，其違約逾期誤判率為 53.88%，此決策樹結構圖見附錄一所示。而圖中被黑色粗虛線框選取的葉子節點為大於等於預期違約率最適切點 (0.2316) 之葉子節點，亦即當預期違約率大於 0.2316 時，顧客將被判定為會產生預期違約的行為。這些節點的詳細分類規則如附錄二所示。

¹¹ Cumulative Lift 之定義為將預期違約率由大至小排序後，在各分位數中預測為違約逾期且實際是違約逾期人數占所有違約逾期人數比例除以整體違約逾期率，也就是預測模型中所能抓到的違約逾期率相對於整體違約逾期率的倍數。

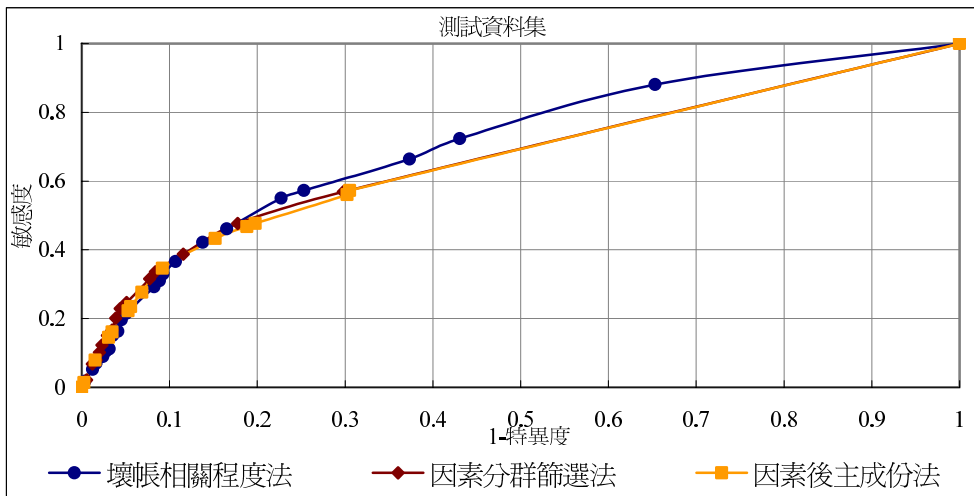


圖 5 決策樹測試資料集 ROC 曲線圖

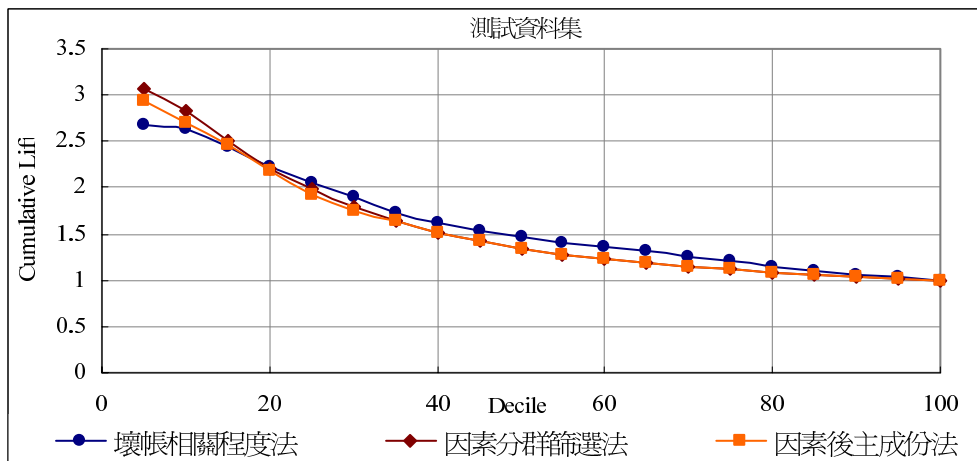


圖 6 決策樹測試資料集 Cumulative Lift 曲線圖

(四)重要研究發現

本研究資料庫的整體違約逾期比例為 16.26%，建模訓練資料集的違約逾期比例為 16.24%。由附錄一及附錄二的「違約逾期相關程度法-決策樹」及其訓練資料集預期違約率分類情形可以得知，本研究預期違約率高於 0.50 之族群共有 305 人，占訓練資料集(7,197 人)的 4.24%；在此群的 305 人中，實際為違約逾期的人數有 182 人，占訓練資料集中所有違約逾期人數(1,169 人)的 15.57%。預期違約率介於 0.30-0.49 之族群，共

有 510 人，占有訓練資料集的 7.09%；在此群的 510 人中，實際為違約逾期的人數亦為 182 人，占訓練資料集中所有違約逾期人數的 15.57%。預期違約率介於 0.23-0.39 之族群，共有 737 人，占有訓練資料集的 10.24%；在此群的 737 人中，實際為違約逾期的人數亦為 180 人，占訓練資料集中所有違約逾期人數的 15.40%。

由上述數據顯示，運用本模式可在預測違約率排名約 5%的顧客中，抓取到實際違約逾期顧客的 15%以上；在預測違約率排名大約前 11%的顧客中，可抓取到實際違約逾期顧客的 30%以上；在預測違約率排名約 20%的顧客中，可抓取到實際違約逾期顧客的 47%以上。

伍、結論與建議

一、研究結論

本研究針對信用貸款顧客基本資料及顧客於聯合徵信中心之相關信用資料進行剖析，以期找出對違約風險具影響力之變數。並採用分類與迴歸樹 (CART) 建立信用貸款違約風險之預測模型。以下即分別彙整說明本研究的發現及結論。

(一)對違約風險具影響力之變數

本研究所考慮二大類可能影響信用貸款違約風險變數，總共有 65 個；其中基本資料變數共 25 個，信用相關整合變數共 40 個。經由卡方獨立性檢定或獨立樣本 T 檢定後，篩選出對違約逾期具影響力的變數共 36 個，包含基本資料變數 9 個，分別為：性別、年齡、學歷、現住房屋型態(居住狀況)、戶籍地是否與現居地相同、近親關係、現任公司地址(縣市)、是否提供戶籍地電話、現任職稱；信用相關整合變數共 27 個，分別為：正卡是否強停、繳款狀況、發放銀行(活卡)家數、首張卡持卡月數、活卡持卡月數、有效卡金張數、有效卡白金卡張數、有效卡金卡張數比例、有效卡白金卡張數比例、具預借現金的信用卡張數、信用總額度、最高信用額度、最低信用額度、平均信用額度、目前持卡之總掛帳總金額、信用總額度與總掛帳總金額的差值、過去 N 月無消費次數比率、過去 N 月全額繳清次數比率、過去 N 月循環無逾期次數比率、過去

N 月逾期次數比率、循環掛帳比率、預借現金比率、預借現金月數比率、循環無逾期最後一次距今月數、預借現金最後一次距今月數、最近三個月他行查詢銀行家數、最近他行查詢距現在天數。

各銀行在進行信用貸款違約風險評估時，可優先考慮上述 36 個變數。在建構信用貸款違約風險建模分析資料庫時，建議亦應納入此 36 個變數。在過去國內外研究中，未有研究能如此完整地建議及提供建立違約預測模型時可採用的起啓變數。

(二)決策樹模型之預測能力探討

1. 精簡變數能力

本研究運用分類與迴歸樹(CART)建立信用貸款違約風險之預測模，最後所選擇的最適模型為「違約逾期相關程度法-決策樹」，透過該模型篩選出的建模變數依其重要度排序依次為最近三個月他行查詢家數、信用總額度與總掛帳總金額之差值、性別、學歷、現任公司地址、過去 N 月全額繳清次數比率、等 8 個。各銀行在進行信用貸款違約風險評估時，在節省時間及成本考量下，建議可精簡地採用上述 8 個變數篩選申請者，並進行必要之信用貸款違約風險估計。

本研究決策樹模型所選取的建模變數中，三個月他行查詢家數、性別、學歷等三個變數為過去研究中常被採用且具有顯著影響力之變數，信用總額度與總掛帳總金額之差值、過去 N 月全額繳清次數比率、具預借現金的信用卡張數、平均信用額度及現任公司地址等五個變數為過去研究中較少被提及的變數。因為本研究的資料量夠大，且抽樣比率皆與實際資料庫相近，故本研究所發掘的這些顯著性變數應具有一定的代表性，可做為未來研究者或銀行界選用時之考量變數。

2. 違約逾期的抓取能力

依據本研究的資料庫，若不採用任何風險預測模型隨機猜測顧客的違約逾期狀況，平均每 6 至 7 個顧客才有可能抓取到一個違約逾期顧客。而採用本研究的「違約逾期相關程度法-決策樹」進行顧客的篩選，在經篩選出的不好顧客中，約每 3 個¹²顧客就有一個為實際違

¹² $544 \div 1552 = 0.35 = 1/2.85 \approx 1/3$ ，其中 544 為模式預測為違約逾期且實際為違約逾期的人數；1552 為模式預測為違約逾期的人數。

約逾期顧客；且在預期違約率前 20%的顧客中，即可抓取到近五成的實際違約逾期顧客。由此可知，本研究之決策樹模型可以有效地判讀出高違約逾期族群，節省銀行授信時所花的人力及時間。換言之，運用本研究的模型，可以有效地規避或控管違約風險較高之顧客族群，進而能篩選出信用狀況良好之顧客特性，加以積極爭取；另外，亦可同時提升審核及放款效率，為銀行爭取更高之利潤。

3. 違約逾期者的特徵

透過本研究之「違約逾期相關程度法-決策樹」建模分類規則可知，最近三個月他行查詢銀行家數愈高者發生違約逾期情形也愈高；信用總額度與總掛帳總金額的差值愈大者發生違約逾期情形則是愈低；性別為男性者發生違約逾期情形相對於女性為高；學歷愈低者其發生違約逾期之情形亦是相對較學歷較高者為高；過去 N 月全額繳清次數比率愈高者則發生違約逾期的情形愈低；具預借現金的信用卡張數愈多者則發生違約逾期情形愈高；平均信用額度愈高者發生違約逾期情形則愈低。

二、未來研究建議

本研究採用分類與迴歸樹(CART)建立信用貸款違約風險之預測模型，建議未來研究者，可參考本研究之建模變數選取原則，但採用其他建模方式，如邏輯斯迴歸、區別分析等，進行信用貸款違約風險預測模型之建構，更進一步或可嘗試比較各種不同建模方式之預測效果。

另，本研究是採用信用風險評估方法中的信用評分制度(Credit Scoring System)來進行研究，未來研究者亦可依循參考本研究之結果，進一步嘗試建構信用評等或混合評等與評分制度之風險評估模型。

參考文獻

- 中央銀行統計資料消費者貸款及建築貸款餘額
http://www.cbc.gov.tw/economic/statistics/total_index.asp, accessed on March, 15, 2007.
- 江世傑, 2001, 模糊類神經網路在消費性貸款之應用, 成功大學工業管理學系碩士論文。
- 何貴清, 2002, 消費者小額信用貸款之信用風險研究-以一商業銀行顧客為例, 中山大學人力資源管理研究所碩士論文。
- 吳振晃, 2003, 資料採礦技術於銀行授信之應用---以消費者貸款為例, 中國文化大學資訊管理研究所碩士在職專班碩士論文。
- 汪海清、黃景泰、謝維國、楊培宏、王南豪, 1999, 消費者貸款實務, 台北市: 台灣金融研訓院。
- 林建州, 2001, 銀行個人消費信用貸款授信風險評估模式之研究, 中山大學財務管理學系研究所碩士論文。
- 林邵杰, 2003, 「風險值 (VaR) 在金融風險管理的角色」, 台灣金融財務季刊, 4 輯 1 期: 39~54。
- 林勉今, 2004, 消費性貸款授信風險評估之研究-以 X 銀行為例, 大同大學事業經營研究所碩士論文。
- 林真真、鄒幼涵編著, 1993, 迴歸分析, 台北市: 華泰書局, 二刷。
- 范哲銘, 2003, 銀行體系與行銷體系信用貸款授信模型的研究, 高雄第一科技大學財務管理研究所碩士論文。
- 洪榮隆, 2003, 消費性貸款信用風險之分析-應用類神經網路, 高雄第一科技大學風險管理與保險所碩士論文。
- 莊瑞珠, 2007, 「邏輯斯迴模型運用在女性信用卡評分制度之研究」, 輔仁管理評論, 14 卷 1 期: 127~154。
- 馬芳資, 1994, 信用卡信用風險預警範例學習系統之研究, 國立政治大學資訊管理研究所碩士論文。
- 張仁哲, 1982, 我國信用卡現代化問題之研究, 國立政治大學企業管理研究所碩士論文。
- 張大成, 2003, 「違約機率與信用評分模型」, 台灣金融財務季刊, 4 輯 1 期: 19~37。
- 張明哲, 2003, 個人消費信用貸款授信模型之研究-以國內某金融機構為例, 雲林科技大學財務金融系碩士班碩士論文。
- 張維仁, 2005, 小額消費性貸款-信用評估模型之探討, 國立彰化師範大學會計學企業高階管理所碩士論文。
- 許育嘉, 2005, 銀行消費性小額信用貸款逾期戶之探討, 元智大學管理研究所碩士論文。
- 陳木在、陳錦村, 2001, 商業銀行風險管理, 台北市: 新陸出版社。
- 陳姍霓, 2004, 整合類神經網路、多元適應性雲形迴歸與分類迴歸樹於信用評等模式之建構-以房屋貸款為例, 輔仁大學管理學研究所碩士論文。
- 陳宗豪, 2000, 消費者小額信用貸款之信用風險研究-甄選的觀點, 中山大學人力資源

- 管理研究所碩士論文。
- 陳順宇，2005，多變量分析，台北市：華泰書局。
- 陳鴻文，2002，個人小額信用貸款授信模式之個案研究，國立高雄第一科技大學財務管理研究所碩士論文。
- 黃宇翔、毛紹睿，2005，「由已訓練類神經網路擷取成本敏感之分類規則」，Journal of e-Business，7卷3期：275~292。
- 馮志剛，1996，公營與民營金融機構個人擔保與信用放款授信評估之研究，高雄工學院管理科學研究所碩士論文。
- 葉秋南，1997，美國金融業務風險管理，台北市：財團法人金融聯合徵信中心編輯委員會。
- 鄭廳宜，1999，信用卡授信審核之實證研究，朝陽科技大學財務金融研究所碩士論文。
- 劉長寬，2003，應用 Logit 模型於消費者擔保貸款違約行為之實證研究，朝陽科技大學財務金融系碩士班碩士論文。
- 蔡明憲，2002，金融機構消費信用貸款授信評量模式，中山大學財務管理學系研究所碩士論文。
- 戴堅，2004，個人消費性信用貸款授信評量模式之研究，中正大學國際經濟研究所碩士論文。
- 薛兆亨，1991，「信用管理政策之訂定(一)剖析信用管理」，會計研究月刊，66期：98~106。
- 謝邦昌、易丹輝，2003，統計資料分析-以 STATISTICA6.0 為例，台北縣：中華資料採礦協會。
- 謝邦昌，2005，資料採礦與商業智慧—SQL Server 2005，台北市：鼎茂圖書出版股份有限公司。
- 羅金川，2004，個人消費性信用貸款授信風險之評估--以國內某銀行消費金融部門之案件為例，中興大學高階經理人碩士在職專班碩士論文。
- 龔昶元，1998，「Logistic Regression 模式應用於信用卡風險審核之研究—以國內某銀行信用卡中心為例」，台北銀行月刊，28卷9期：35~49。
- Berry, M. J. A. and Linoff, G. S., 1997, **Data Mining Techniques: for marketing, sales, and customer support**, New York: John Wiley & Sons, Inc..
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, **Classification and Regression Trees**, New York, Chapman & Hall.
- Daniels, J. P. and VanHoose, D. D., 1999, **International monetary and financial economics**, Cincinnati, Ohio: South-Western College Pub.
- Glorfeld, L. W. and Hardgrave, B. C., 1996, "A Performance Comparison of Eight Inductive Decision Models for Predicting Commercial Loan Creditworthiness," **Computer and Operations Research**, Vol. 23, No.10, 933-944..
- Henley, W. E. and Hand, D. J., 1996, "A K-Nearest-Neighbour Classifier For Assessing Consumer Credit Risk," **The Statistician**, Vol. 45, No. 1, 77-95.
- Hull, J., 2003, **Option, Futures, and Other Derivatives**, Fifth edition, Upper Saddle River, N.J.: Prentice-Hall Press.
- Ikizler, N. and Guvenir, H. A., 2002, **Mining Interesting Rules for Banking Loan Data**,

Ph.D. Dissertation, Bilkent University.

- Joos, P., Vanhoof, K., Ooghe, H., and Sierens, N., 1998, "Credit classification: A comparison of logit models and decision trees.", **10th European Conference on Machine Learning, Workshop notes: Application of machine learning and datamining in finance**. TU hemnitz, Dorint-Parkhotel, Chemnitz, Germany.
- Kass, G. V., 1980, "An exploratory technique for investigating large quantities of categorical data," **Applied Statistics**, Vol. 29, No. 2, 119-127.
- Liang, T. H., Lin, J. B., and Liao, J. J., 2007, "The Best Oversampling Proportion and Variables Selection Procedure in the Predictive model of Probability of Default of Personal Consumer Loans.", in **Proceedings 10th Joint Conference on Information Sciences (CIEF2007)**, Salt Lake City, Utah, U.S.A..
- Quinlan, J. Ross, 1993, **C4.5 Programs for Machine Learning**, Morgan Kaufmann Publishers.
- SAS, 2005, **Applying Data Mining Techniques Using SAS Enterprise Miner: Course Notes**, USA: SAS Institute Inc..
- Sharma, S., 1996, **Applied Multivariate Techniques**, USA: John Wiley & Sons, Inc..
- Shi, X. J., 2006, "Optimal Sample Pairing and Critical Value of Logistic Default Risk Modeling: The China Case," **Application of Statistics and Management**, Vol. 25, No.6, 675-682.
- Steenackers, A. and Goovates, M. J., 1989, "A Credit Scoring Model for Personal Loans," **Insurance: Mathematics and Economics**, Vol. 8, No.1, 31-34.
- Weiss, N. A., 2002, **Elementary Statistics**, 5th edition, USA: Pearson Education Inc..
- Zavgren, C. V., 1985, "Assessing the vulnerability to failure of American industrial firms: a logistic analysis," **Journal of Business Finance and Accounting**, Vol. 12, No.1, 19-45.

附錄 2：最適切點下「違約逾期相關程度法-決策樹」訓練資料集違約逾期分類情形表

節點	分類規則	分割變數	分割準則	結果
1.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 58 預期違約率： 67.2%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	≥ 8.5	
	AND	平均信用額度	$< 159,5000$	
	AND	現任公司地址（縣市）	宜蘭縣、基隆市、彰化縣、南投縣、雲林縣、嘉義市、臺南市、臺南縣及高雄縣	
2.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 66 預期違約率： 63.6%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	≥ 8.5	
	AND	平均信用額度	$< 159,5000$	
	AND	現任公司地址（縣市）	臺北市、臺北縣、桃園縣、新竹市、新竹縣、苗栗縣、高雄市及屏東縣	
	AND	最近三個月他行查詢銀行家數	< 11.5	
	AND	過去 N 月全額繳清次數比率	< 1.92	
3.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 41 預期違約率：58.5%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	< 8.5	
	AND	信用總額度與總掛帳總金額的差值	$< 87,204.5$	
	AND	信用總額度與總掛帳總金額的差值	$\geq -51,500$	
	AND	具預借現金的信用卡張數	≥ 1.5	
	AND	最近三個月他行查詢銀行家數	< 7.5	

附錄 2：最適切點下「違約逾期相關程度法-決策樹」訓練資料集違約逾期分類情形表

節點	分類規則	分割變數	分割準則	結果
	AND	信用總額度與總掛帳總金額的差值	<88,000	
4.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 35 預期違約率：57.1%
	AND	信用總額度與總掛帳總金額的差值	<258,500	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	<8.5	
	AND	信用總額度與總掛帳總金額的差值	<87,204.5	
	AND	信用總額度與總掛帳總金額的差值	<-51,500	
5.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 30 預期違約率：56.7%
	AND	信用總額度與總掛帳總金額的差值	<258,500	
	AND	性別	男性	
	AND	學歷	大學/專科及研究所（碩士）以上	
	AND	過去 N 月全額繳清次數比率	≥ 7.16	
	AND	過去 N 月全額繳清次數比率	<11.93	
6.	IF	最近三個月他行查詢銀行家數	<5.5	樣本數： 39 預期違約率：53.8%
	AND	信用總額度與總掛帳總金額的差值	<172,500	
	AND	最近三個月他行查詢銀行家數	≥ 3.5	
	AND	學歷	高中職及國中以下	
	AND	性別	男	
	AND	現任公司地址（縣市）	宜蘭縣、台中市、台中縣、南投縣、雲林縣、嘉義市、嘉義縣、臺南縣及其他	
	AND	具預借現金的信用卡張數	≥ 0.5	
	AND	過去 N 月全額繳清次數比率	<3.845	

附錄 2：最適切點下「違約逾期相關程度法-決策樹」訓練資料集違約逾期分類情形表（續一）

節點	分類規則	分割變數	分割準則	結果
7.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 36 預期違約率： 52.8%
	AND	信用總額度與總掛帳總金額的差值	$\geq 258,500$	
	AND	性別	女性	
	AND	現任公司地址（縣市）	宜蘭縣、基隆市、彰化縣、南投縣、雲林縣、嘉義市、臺南市、臺南縣及高雄縣	
	AND	過去 N 月全額繳清次數比率	≥ 0.415	
8.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 54 預期違約率： 46.3%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	≥ 8.5	
	AND	平均信用額度	$< 159,5000$	
	AND	現任公司地址（縣市）	臺北市、臺北縣、桃園縣、新竹市、新竹縣、苗栗縣、高雄市及屏東縣	
	AND	最近三個月他行查詢銀行家數	< 11.5	
	AND	過去 N 月全額繳清次數比率	≥ 1.92	
9.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 41 預期違約率： 41.5%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	< 8.5	
	AND	信用總額度與總掛帳總金額的差值	$< 87,204.5$	
	AND	信用總額度與總掛帳總金額的差值	$\geq -51,500$	
	AND	具預借現金的信用卡張數	≥ 1.5	
	AND	最近三個月他行查詢銀行家數	< 7.5	
	AND	信用總額度與總掛帳總金額的	$\geq 88,000$	

附錄 2：最適切點下「違約逾期相關程度法-決策樹」訓練資料集違約逾期分類情形表（續一）

節點	分類規則	分割變數	分割準則	結果
		差值		
10.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 48 預期違約率： 37.5%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	≥ 8.5	
	AND	平均信用額度	$< 159,5000$	
	AND	現任公司地址（縣市）	臺北市、臺北縣、桃園縣、新竹市、新竹縣、苗栗縣、高雄市及屏東縣	
11.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 40 預期違約率： 35%
	AND	信用總額度與總掛帳總金額的差值	$\geq 258,500$	
	AND	性別	女性	
	AND	現任公司地址（縣市）	宜蘭縣、基隆市、彰化縣、南投縣、雲林縣、嘉義市、臺南市、臺南縣及高雄縣	
	AND	過去 N 月全額繳清次數比率	< 0.415	
12.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 197 預期違約率： 33.5%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	< 8.5	
	AND	信用總額度與總掛帳總金額的差值	$< 87,204.5$	
	AND	信用總額度與總掛帳總金額的差值	$\geq -51,500$	
	AND	具預借現金的信用卡張數	< 1.5	

附錄 2：最適切點下「違約逾期相關程度法-決策樹」訓練資料集違約逾期分類情形表（續二）

節點	分類規則	分割變數	分割準則	結果
13.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 30 預期違約率： 33.3%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	≥ 8.5	
	AND	平均信用額度	$\geq 159,500$	
14.	IF	最近三個月他行查詢銀行家數	< 5.5	樣本數： 42 預期違約率： 33.3%
	AND	信用總額度與總掛帳總金額的差值	$< 172,500$	
	AND	最近三個月他行查詢銀行家數	≥ 3.5	
	AND	學歷	高中職及國中以下	
	AND	性別	男	
	AND	現任公司地址（縣市）	宜蘭縣、台中市、台中縣、南投縣、雲林縣、嘉義市、嘉義縣、臺南縣及其他	
	AND	具預借現金的信用卡張數	≥ 0.5	
	AND	過去 N 月全額繳清次數比率	≥ 3.845	
15.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 58 預期違約率： 31.0%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	大學/專科及研究所（碩士）以上	
	AND	過去 N 月全額繳清次數比率	≥ 7.16	
	AND	過去 N 月全額繳清次數比率	≥ 11.93	
16.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 41 預期違約率： 29.3%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	< 8.5	
	AND	信用總額度與總掛帳總金額的差值	$< 87,204.5$	
	AND	信用總額度與總掛帳總金額的差值	$\geq -51,500$	
	AND	具預借現金的信用卡張數	≥ 1.5	

附錄 2：最適切點下「違約逾期相關程度法-決策樹」訓練資料集違約逾期分類情形表（續二）

節點	分類規則	分割變數	分割準則	結果
	AND	最近三個月他行查詢銀行家數	≥ 7.5	
17.	IF	最近三個月他行查詢銀行家數	< 5.5	樣本數： 61 預期違約率： 26.2%
	AND	信用總額度與總掛帳總金額的差值	$< 172,500$	
	AND	最近三個月他行查詢銀行家數	≥ 3.5	
	AND	學歷	高中職及國中以下	
	AND	性別	男	
	AND	現任公司地址（縣市）	宜蘭縣、台中市、台中縣、南投縣、雲林縣、嘉義市、嘉義縣、臺南縣及其他	
AND	具預借現金的信用卡張數	< 0.5		
18.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 87 預期違約率： 25.3%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	高中職及國中以下	
	AND	最近三個月他行查詢銀行家數	< 8.5	
	AND	信用總額度與總掛帳總金額的差值	$\geq 87,204.5$	
19.	IF	最近三個月他行查詢銀行家數	≥ 5.5	樣本數： 263 預期違約率： 24.3%
	AND	信用總額度與總掛帳總金額的差值	$< 258,500$	
	AND	性別	男性	
	AND	學歷	大學/專科及研究所（碩士）以上	
	AND	過去 N 月全額繳清次數比率	< 7.16	
20.	IF	最近三個月他行查詢銀行家數	< 5.5	樣本數： 285 預期違約率： 23.2%
	AND	信用總額度與總掛帳總金額的差值	$< 172,500$	
	AND	最近三個月他行查詢銀行家數	≥ 3.5	
	AND	學歷	高中職及國中以下	
	AND	性別	男	
	AND	現任公司地址（縣市）	臺北市、臺北縣、桃園縣、新竹市、新竹縣、苗栗縣、高雄市及屏東縣	

作者簡介

梁德馨

淡江大學管理科學博士畢業，現任輔仁大學管理學院統計資訊系副教授（1997~迄今），並兼任中華資料採礦協會秘書長及中華民國民意測驗協會理事。主要研究以調查及抽樣為主，近年來著重在各類應用統計相關研究，研究主要方向包含資料採礦之銀行風險模型預測、各類行銷研究及網路行為相關研究。發表期刊包括 INFORMATION SCIENCES 2007 - Proceedings of the 10th Joint Conference (IE 收錄)、中山管理評論、數據分析及創新與管理。

E-mail: stat1013@mail.fju.edu.tw

葉建良

輔仁大學應用統計研究所碩士，碩士論文主要研究領域為信用風險預測模型之建構。現職為玉山商業銀行股份有限公司金控總部專員，主要從事有關 Basel II 消金信用風險分析之研究。

E-mail: matt-07155@email.esunbank.com.tw