

**論文標題：建構台灣中小企業兩階段風險評估模型**

**研究生：吳佩珊/國立交通大學/工業工程與管理學系**

**指導教授：唐麗英/國立交通大學/工業工程與管理學系**

**張永佳/國立交通大學/工業工程與管理學系**

### **論文摘要：**

近年來金融市場不穩定與經濟不景氣，使得國內銀行或金融機構皆承受相當大的信用風險。為落實信用風險管理理念，根據國際清算銀行2004年公佈之新巴塞爾資本協定（The New Basel Capital Accord，簡稱為 Basel II）允許銀行及金融機構可以自行建立風險評估模型來衡量借款客戶的風險。現有的風險評估模型多是利用一個方法的分類判別模型，將借款客戶分成違約及不違約(default/non-default)兩類，利用這些判別模型在進行風險評估時，雖然有不錯的整體準確率，但常會出現某類借款客戶（如違約客戶）有準確率高，但另一類客戶之準確率偏低的情況。此種判定模型即使整體的準確率不錯，但由於對某類借款客戶之風險誤判機率過大，對於金融機構而言的實用性不高，無法直接應用到台灣之中小企業上。因此，本研究發展出一套二階段的判別模型，以改善這種準確率偏向某一類客戶的問題，進而提升現有判別模型判定之準確率。本研究利用邏輯斯迴歸(logistic

regression)與支持向量機(support vector machine, SVM)來建構此兩階段的風險評估模型，再依照最後所得之判定模式給予借款企業一個風險等級，以供銀行或金融機構能制定出最佳之放款策略。

【關鍵詞】風險評估、中小企業、邏輯斯迴歸、支持向量機

### 與中小企業關聯性：

台灣中小企業佔台灣企業數九成以上且提供了七成以上的就業機會，因此台灣中小企業不論是在穩定社會民生或是在總體經濟成長上，均扮演了重要的角色，中小企業外借資金管道大多是金融機構或個人借款，且中小企業外借資金之比率有逐年升高的趨勢。目前中、外文獻雖已發展出許多風險評估模型，但大多是針對上市、上櫃公司，少有以中小企業為研究對象，若以中外文獻所建議之上市、上櫃公司的風險評估模型直接應用到台灣之中小企業上，其模式之準確性可能不佳。中小企業之財務報表普遍不夠完備及透明，所以其財務資訊沒有完全的公信力，若純以財務資訊做為中小企業借款客戶之放款依據，可能無法完全作為風險衡量的資訊，本研究則透過一些非財務因素反映公司真實之營運狀況。最後，本研究以中小企業為研究對象，發展出適合中小企業風險評估之評估模型，以供銀行或金融機構能使用並制定出最佳之放款策略

## 壹、 研究動機

隨著國際金融市場環境的快速發展，企業面臨的風險比以往更為複雜，因此銀行等金融機構在企業倒閉或違約事件層出不窮的同時，必須發展一套有效的信用風險管理機制，才能永續經營。

台灣中小企業佔台灣企業數九成以上且提供了七成以上的就業機會，其每年的直接銷售產值的比例，更佔全台近三成，因此台灣中小企業不論是在穩定社會民生或是在總體經濟成長上，均扮演了重要的角色，但中小企業因其規模小，缺乏如上市或上櫃公司的融資管道，所以中小企業外借資金管道大多是金融機構或個人借款，且中小企業外借資金之比率有逐年升高的趨勢。近年來台灣經濟景氣低迷，產業大量外移，國內銀行等金融機構對企業之放款態度漸趨保守，使得中小企業在資金取得上更顯不易，此外，國內中小企業之會計制度多半不完整，且財務報表普遍不夠完備及透明，所以其財務資訊沒有完全的公信力，此導致中小企業與金融機構之間存在嚴重的資訊不對稱問題，使得金融機構在放款給中小企業時風險也相對增加。

現有之風險評估模型大多只使用一種分類方法來構建風險判別模型，即將借款之客戶分成違約(default)與不違約(non-default)兩類。然而利用這些判別模型在進行風險評估時，雖然有不錯的整體準確率，但有時會偏頗地將大部分客戶判定為同一類型客戶，造成不違約

客戶準確率過高及違約客戶準確率過低，或違約客戶準確率過高及不違約客戶準確率過低之現象，這種不對稱的判定情形會使金融機構誤判借款客戶之風險，有可能導致上億元之損失。

崔運驊[4] 利用邏輯斯迴歸建構中小企業風險評估模型，結果在訓練樣本及測試樣本上皆有不對稱準確率的情形發生，違約客戶準確率為88%與79%，但不違約客戶準確率只有47%與44%。黃啟峰[5]分別利用支持向量機、區別分析建構中小企業風險評估模式，經實例驗證亦有前述分類不對稱的準確率結果，測試樣本中不違約準確率為92%與80%，但其違約樣本準確率卻分別僅有28%與48%，有其中一極低的準確率情形。銀行或金融機構希望能經由放款來獲利，若為不違約客戶則透過正常放款，獲得正常的利息收入；若面臨可能違約之客戶則希望提高利率或是增加抵押條件，以增加利息收入及降低風險，所以風險評估不應偏重某一方。

## 貳、 研究目的

利用現有的風險評估模型進行違約與不違約客戶的判別時，常會產生一段模糊區域，而在此模糊區域的判別準確率太低通常是造成整體判別準確率不高的主因。因此針對台灣之中小企業借款者，本研究發展出一套兩階段的風險評估模型，以供銀行或金融機構準確地提升判定不違約與違約之借款客戶。本研究方法共分兩階段進行，第一階

段先判定可能違約與可能不違約的客戶，第二階段則是針對不易判別的模糊區域之借款客戶，再進行第二次分類，上述兩階段之判別模型應可有效降低整體分類準確率高而各分類準確率卻不對稱的情形，以提升借款客戶風險評估之準確率，確實達到管控風險之目的。

綜上所述，本研究之主要目的為針對中小企業借款客戶發展出一套有助於金融機構處理放款流程之兩階段風險判別程序，使金融機構在處理放款作業時，可針對借款者的各項財務及經營資訊進行風險評估，能快速有效區別出不違約及可能違約之借款客戶的風險等級，進而做出適當的放款決策，以有效降低放款給違約客戶之風險。

## 參、 文獻探討

本章先介紹金融機構授信原則，隨後探討風險評估與其常見之模型，最後介紹邏輯斯迴歸、支持向量機之理論基礎。

### 3.1 金融機構授信原則

金融機構在處理授信所面臨到最大的問題即是利潤與風險的權衡。金融機構若是採用過於嚴厲的放款標準，雖然可以避免違約所導致的呆帳損失，但也同時減少了放款所帶來之利潤。

金融機構在處理放款業務時，會針對不同的放款對象，採取不同之放款金額、期限或利率，但不論如何，銀行一定會遵守以下三點原

則[3]：安全性原則、流動性原則、獲利性原則。

## 3.2 風險評估

一般金融機構可依其自身之營運狀況對貸款者進行所謂之內部信用評等，內部信用評等常被用於銀行的信用風險管理，又可分為信用評分(credit scoring)的評分模型及信用評等(credit rating)的等級預測模型，等級預測模型將借款者分為多個信用等級，而評分模型則多為風險評估，是將借款者分為具有風險與不具風險兩類。風險評估模式，主要是預測企業財務是否發生危機，以信用評分的方式將借款對象區分為違約與不違約，也就是做二分類法的判別，以提供金融機構制定放款決策。

### 3.2.1 風險評估模型

風險評估模型是金融機構用來決定是否放款的依據，以評估在放款後該借款客戶是否有發生違約的可能，做二分法的區別，可供銀行用來作出判別借款或不借款的決定。而風險評估模型也從早期的人為判斷，演變為利用統計分析方法來建構模型，而後隨著電腦的快速發展與普及，人工智慧方法也逐漸應用在風險評估模型上，早期是以類神經網路來建構模型，近年有研究結果顯示使用支持向量機的分類效果可能優於使用統計方法與類神經網路[7]。目前國內外金融機構常用的風險評估模型如下[2]：

## 1. 區別分析 ( discriminant analysis, DA )

區別分析方法是根據樣本的特性建立區別函數，以區別函數值來對樣本進行分類。利用區別分析之優點為可同時考慮多項財務指標，較能衡量企業整體之績效；缺點則是區別分析變數需符合常態假設。

## 2. Logit 模型 (邏輯斯迴歸, logistic regression)

Logit 模型假設事件發生的機率服從 Logistic 分配，可解決自變數非常態的問題，所求得之機率值會落於 0 與 1 之間。模型可適用於非線性的情況，但須經轉換步驟才能求得機率值，而切割點的決定會影響到整個模式分類的準確度。

## 3. Probit 模型

Probit 模型假設事件發生的機率服從標準常態分配，採累加機率來進行資料轉換，以解決自變數非常態的問題，但轉換程序較為複雜，因此過去許多文獻多以 Logit 模型來建構風險評估模型。

## 4. 類神經網路 (neural network)

類神經網路模型是模擬人類大腦思考的方式建構出來的人工智慧系統，類神經網路並不受限於樣本為常態分配的假設，也無變數是否具共線性的問題，其結果會介於 0 與 1 之間，在信用風險的衡量下，即為公司之違約機率，但模型之運作時屬黑箱作業，無法解釋變數間的關係，因此不具統計理論。

## 5. 支持向量機 (support vector machine, SVM)

SVM 是近年所提出的基於統計學習理論人工智慧方法，2000 年後逐漸有學者應用到信用評分上，透過一超平面可將資料分類，並透過核心函數(kernel function)使原本非線性資料轉換成線性資料，以進行分類，提高分類的準確率。

### 3.2.2 風險評估模式相關文獻

Altman[6]逐步多元區別分析法構建模型，考慮不同財務比率分析項目間關連性對分析結果的影響，其研究結果顯示預測模式之短期預測準確率相當高，時間越長則預測準確率愈低。

Edmister[8]收集小型企業的財務比率變數，進行小企業的財務危機預測。他利用逐步多元區別分析建構風險評估模型，以區別分析值預測企業是否失敗，將所有企業區分為失敗、非失敗及灰色地帶，灰色地帶則由分析師以專業判斷企業是否失敗。

Desai 等人[7]分別利用類神經網路、區別分析及邏輯斯迴歸進行信用風險評分，結果顯示以類神經網路進行風險評估優於區別分析，而邏輯斯迴歸之分析結果則與類神經方法差異不大。

Baesens 等人[7]比較各種常用之信用評分分類方法，包含邏輯斯迴歸、區別分析、KNN(K-nearest neighbor)、類神經網路、決策樹、支持向量機等，並利用八種資料庫的信用評分資料進行分析，結果顯

示類神經網路與支持向量機有較佳的分類結果，而邏輯斯迴歸與判別分析在信用評分上也有不錯的效果。

Min和Lee[9]分別利用SVM和多元區別分析、邏輯斯迴歸、倒傳遞類神經網路(Back-propagation Network, BPN)對南韓的中小型重工業進行企業危機預測，研究結果顯示SVM有較好的預測結果，優於邏輯斯迴歸、多元區別分析與BPN。

由上述風險評估模型相關文獻可知，邏輯斯迴歸應用較廣，準確率較高且限制較少，是個相當簡便的統計模型。而支持向量機為近年所提出的人工智慧方法，且可選擇不同的核心函數，處理非線性資料，能提高分類上的準確率，且模型可以有合理的解釋能力，為較新的建構風險評估模型的方法。

### 3.3 邏輯斯迴歸

邏輯斯迴歸[1]是由 Berkson(1944)所提出，此法與迴歸分析方法類似，但可處理依變數為二元變數(binary variable)的資料。線性迴歸方法需要滿足依變數(Y)為連續性與常態分配之假設，而邏輯斯迴歸則沒有設定任何假設條件。

邏輯斯迴歸模式之依變數為二元變數時，表其結果只有成功( $Y = 1$ )或失敗( $Y = 0$ )兩種結果，則對其 Y 建立一般線性迴歸模式如下所示：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1)$$

其中  $X_k$  代表第  $k$  項自變數之值。但是以一般線性迴歸模式預測二元依

變數時，依變數的估計值可能會落在 (0,1) 之外，而與實際依變數情況不符。

邏輯斯迴歸提出一轉換函數稱為 Logit 函數，利用  $\ln \frac{p}{1-p}$  轉換予以線性化，主要模式可以(2)式或(3)式表示如下[1]：

$$Z = \ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \beta X \quad (2)$$

或

$$p = \frac{1}{1+e^{-Z}} \quad (3)$$

其中， $p$  為成功機率，成功機率與失敗機率相除得到優勝比  $\frac{p}{1-p}$  稱為勝算(odds)，而  $\ln \frac{p}{1-p}$  稱為勝算對數。換句話說，令  $X = [X_1, X_2, \dots, X_k]$ ，

假設  $X$  屬於成功的發生機率是 Logistic 分配。

$$p(Y=1|X) = F(Z) = X \text{ 屬於成功的機率} = \frac{1}{1+e^{-Z}} \quad (4)$$

$$p(Y=0|X) = F(Z) = X \text{ 屬於失敗的機率} = \frac{e^{-Z}}{1+e^{-Z}} \quad (5)$$

$Z$  為定義於  $[-\infty, \infty]$  的一個潛伏因變數，與  $Y$  值相關，令

$$\beta = [\beta_1, \beta_2, \dots, \beta_k]。$$

$$Z = \beta X = [\beta_1, \beta_2, \dots, \beta_k] \times X \quad (6)$$

$Z$  值與事件發生率  $p$  為正向關係，經過 Logit 函數轉換後，可確保  $p$  值落於 0 與 1 之間。

在應用邏輯斯迴歸時必須設定分割值(threshold value)，當迴歸模型所獲得的  $p$  值大於分割值時，則將此資料判別為成功，相反地，若  $p$  值小於分割值時，則判別為失敗。

### 3.4 支持向量機

支持向量機(Support Vector Machine, SVM)由 Vapnik(1995)[13]和其研究團隊所發展出來，可用於解決二分類問題，SVM 的學習方法是透過訓練資料的演算，計算出一個最理想的超平面(hyperplane)，再藉由此超平面分類，進行測試資料來判斷其準確率。

#### 3.4.1 支持向量機原理

SVM的原理主要運用超平面空間，將資料區分為兩類，在眾多的超平面空間分割中，希望兩類資料能夠分隔能越遠越好，即距離超平面的邊界值(margin)最大則該超平面即為最佳分割的超平面(optimal hyperplane)。

一般而言，大部分的分類方法是由所有的樣本資料決定分類模式之分類界限，而SVM決定分類界限之超平面則由使邊界值最大的樣本資料決定，則此樣本資料稱做支持向量(support vector)，SVM希望最大化超平面及支持向量的距離，以提升分類的準確性。

假設訓練樣本  $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), i = 1, \dots, n\}$ ，其中  $x_i \in R^n$  且  $y_i \in \{-1, +1\}$ 。若樣本資料為線性可以分類時，則必須存在

$(w, b)$  滿足式(10)和(11)滿足：

$$x_i \cdot w + b \geq +1 \quad \text{for } y_i = +1 \quad (10)$$

$$x_i \cdot w + b \leq -1 \quad \text{for } y_i = -1 \quad (11)$$

其中， $w$  為向量， $b$  為常數，則超平面為  $f(x) = x \cdot w + b$ ，而邊界值  $\rho(w, b)$

可以表示如下：

$$\rho(w, b) = \min_{\{x; y=1\}} \frac{x \cdot w + b}{\|w\|} - \max_{\{x; y=-1\}} \frac{x \cdot w + b}{\|w\|} = \frac{2}{\|w\|} \quad (12)$$

其中 $\|w\|$ 表向量距離，則最佳超平面即使邊界值 $\frac{2}{\|w\|}$ 最大化，最大化邊界值即倒數的最小化，等同於最小化 $\frac{1}{2}\|w\|$ ，可以表示如下：

$$\begin{aligned} \underset{w, b}{Min} \quad & \frac{1}{2} w^T w \\ \text{subject to} \quad & y_i (x_i \cdot w + b) \geq 1 \quad i = 1, \dots, n \end{aligned} \quad (13)$$

將此最佳化問題轉換為拉格蘭茲(Lagrange)問題如下：

$$L = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i (y_i (x_i \cdot w + b) - 1) \quad \text{where } \alpha_i \geq 0 \quad (14)$$

此問題必須符合以下兩個限制式：

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \quad (15)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (16)$$

將(15)式和(16)式代入(14)式中，可得新的線性規劃模式如下：

$$\begin{aligned} \underset{\alpha_i}{Max} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \text{subject to} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \quad i = 1, \dots, n \end{aligned} \quad (17)$$

求解此模式，可得最佳解 $\alpha_i^*$ ，而 $w$ 最佳解 $w^* = \sum_{i=1}^n \alpha_i^* y_i x_i$ ，即可求得

最佳分類之界限。因此，最佳超平面決策函數 $f(x) = \text{sgn}(\langle w^* \cdot x \rangle + b^*)$ 可

以表示如下：

$$f(x) = \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i^* \langle x_i, x \rangle + b^* \right) \quad (18)$$

### 3.4.2 核心函數

當資料為線性無法分類的資料時，則可以透過核心函數(Kernel Function)的轉換，將資料轉換到高維度的特徵空間(feature space)中。

原本低維度線性不可分割的資料，則可透過核心函數轉換，找到線性可分割的高維度特徵空間，在此空間即可使用線性的方式找出最佳分割之超平面。

而原本線性不可分之最佳化問題，則透過核心函數 $\phi$ （如(19)式）取代原始資料，轉換為線性可分的最佳化問題，如(20)式所示。

$$(\phi(x_i) \cdot \phi(x_j)) := k(x_i, x_j) \quad (19)$$

$$\begin{aligned} \text{Max} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \quad i = 1, \dots, n \end{aligned} \quad (20)$$

經由上述步驟，最後可得一決策函數如(21)式，來判斷資料屬於何種分類。

$$\begin{aligned} f(x) &= \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i^* \langle \phi(x), \phi(x_i) \rangle + b^* \right) \\ &= \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i^* \langle k(x, x_i) \rangle + b^* \right) \end{aligned} \quad (21)$$

不同的核心函數對不同的資料有不同程度的分類結果，常見的SVM核心函數類型包含線性函數(Linear)、多項式(Polynomial)、徑向基函數(Radial Basis Function, RBF) S型函數(Sigmoid)等，其中以RBF函數（如(22)式）最為廣泛使用，且實證效果多半較其他核心函數為

佳[9][10]。

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (22)$$

## 肆、 研究方法

本研究所建構之風險評估模型之過程主要包括四大步驟。首先為資料之收集與整理；接下來建構兩階段的風險評估模型，利用邏輯斯迴歸先進行第一階段之風險評估，將無法有效分類之資料，再以執行風險評估模型的第二階段判別，在此階段中，利用 SVM 進行第二階段之風險評估；最後則依照風險評估模型給予客戶一個合理的信用風險等級，以利金融機構進行放款的程序。

### 4.1 資料收集與整理

本研究之研究對象為國內中小企業之借款客戶，而中小企業所提供的財務報表不如上市上櫃公司般具有公信力，且有時企業會因借款需要會美化其財務報表，使得其可信度更為降低，因此若純以財務資訊做為中小企業借款客戶之放款依據，可能無法完全作為風險衡量的資訊。因此，本研究希望透過一些非財務因素，如員工忠誠度、公司之經營展望等來反映公司真實之營運狀況，由專業人員對企業狀況進行評分作為非財務資訊，以利有效判斷借款企業之風險。除此之外，經濟景氣好壞也是同時影響中小企業是否得以繼續生存的條件之一，故總體經濟因素也是影響借貸風險高低的重要預測變數。由此可

知，同時考量財務、非財務資訊及總體經濟指標等三大資訊變數建構風險評估模型，應可大幅提升模型之準確性。

因此本研究所建構之兩階段風險評估模型將同時考慮上述三大類變數，當作建構模型的自變數。

#### 4.2 建立兩階段風險評估模型

當資料數據收集完成並整理後，即進行兩階段風險評估模型之構建。本研究利用逐步邏輯斯迴歸為第一階段的判別方法，以 SVM 建構第二階段之風險評估模型。如本論文前所述，邏輯斯迴歸應為目前應用在風險評估模型中最为簡單好操作的統計方法，且對於模型解釋能力佳；SVM 則可以透過核心函數，解決非線性的資料問題，因此本研究將邏輯斯迴歸無法處理的資料再利用 SVM 方法進行分析。

由於中小企業資料複雜，本研究利用兩階段的判別方法，以提升現有風險評估模型僅用一種分類方法的判別正確率。本研究第一階段是針對邏輯斯迴歸所得之判別方程式，決定最佳切點，現有文獻應用邏輯斯迴歸時，皆以整體準確率最高的機率值作為最佳切點，利用此最佳切點，將樣本資料區分為兩類，但由各文獻中之判別結果發現，此情況易發生不對稱的準確率情形，(如圖 4-1)，即某類樣本之準確率與另類樣本之準確率可能相差過大，但整體準確率卻相當不錯。

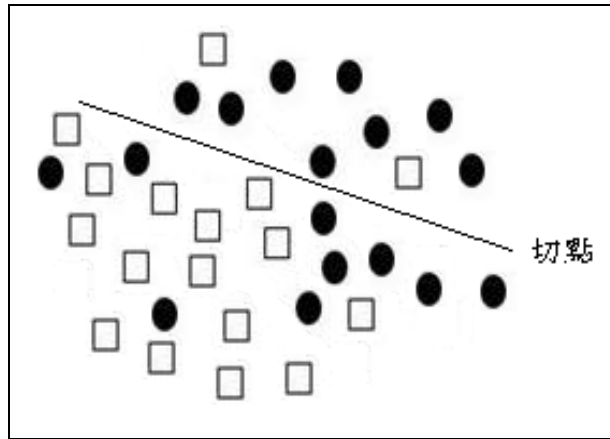


圖 4-1 現有風險評估模型概念示意圖

因此本研究提出以利用兩個切點方式，將樣本資料區分為不違約、違約及模糊區，在兩個可明確區分的不違約與違約資料的區域，表示其提供之企業資訊足以進行判斷借款客戶是否違約，而模糊區內的資料則代表不易判別借款客戶是否違約，因此需對此模糊區域資料進行第二階段之判別，以得到正確的判定結果；而第二階段中再針對模糊區內的資料，以 SVM 建構風險評估模型。圖 4-2 為兩階段風險評估模型概念示意圖。

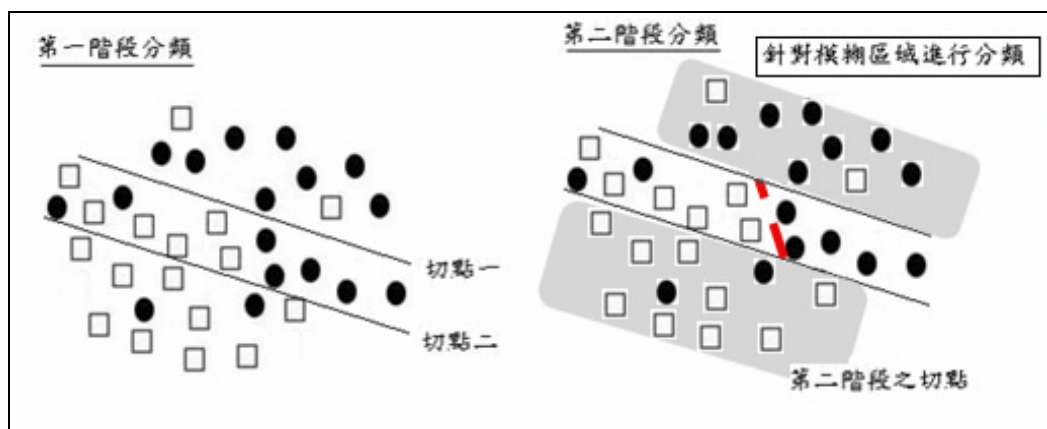


圖 4-2 兩階段風險評估模型概念示意圖

#### 4.2.1 第一階段風險評估模型

此階段以邏輯斯迴歸為構建判別模式的方法。本研究以逐步邏輯斯迴歸法，挑選解釋力較佳的自變數，以形成簡潔且解釋能力強的適配良好模型。利用此適配模型作為第一階段之判別，以簡化原本複雜的資料，再建構第二階段判別模型之建構。建構第一階段風險評估模型之步驟如下：

1. 決定訓練樣本，以等比例的不違約及違約樣本建構模型。由於利用兩切點的方式進行分類，若資料筆數不均，則在決定切點上容易發生偏誤的情形，而無法準確的決定最佳切點，因此本研究以等比例之樣本組合建構風險評估模型，對不違約樣本資料進行隨機抽樣。
2. 根據訓練樣本利用逐步邏輯斯迴歸建構風險評估模型。
3. 決定最佳兩切點，切點以 80%精確率(precision)<sup>1</sup>作為切割值設定標準，容許 20%的誤判率為考量，即 10 筆樣本中允許 2 筆為誤判樣本。以邏輯斯迴歸建立判別方程式最後可得到一界於 0 至 1 的機率值，越接近 0 為違約樣本的機率越高，反之，接近 1 則為不違約樣本；所以將切點設定為由 0 及 1 開始，誤判率超過 20%即停止，則該機率值即為切點。

---

<sup>1</sup> 本研究以 80%精確率作為切點的設定標準，認為可以透過此切點，同時保有準確率的預測效果與判別較多的資料的優勢，以利第二階段風險評估模型的建構。

4. 兩切點外之樣本即為第一階段判定不違約及違約樣本，而切點內之模糊區域則繼續進行第二階段風險評估。

#### 4.2.2 第二階段風險評估模型

本研究以支持向量機並結合 RBF 核心函數作為第二階段風險評估方法，且利用 10-fold 交叉驗證之方式訓練學習，以建構判別模型，因此針對進行過第一階段風險評估模型無法判斷是否違約之模糊區資料進行模型建構。

## 伍、 研究結果

本研究利用台灣某金融機構所提供之中小企業借款歷史資料作為模式建構之說明，並用以驗證本研究模型的有效性。

### 5.1 資料收集與整理

本研究蒐集台灣某金融機構 2001 年至 2006 年之中小企業借款資料來驗證本研究所提出之風險評估模型。本研究針對所收集的已結案之借款客戶資料，刪除含有遺失值之借款客戶資料，並將落於各變數平均值之四倍標準差外的值視為離群值並予以刪除，共得 3438 筆資料，其中違約案件共 225 筆，不違約案件共 3213 筆。本研究以 2001 年至 2005 年的五年資料作為模式建構之訓練樣本，2006 年資料作為測試樣本。由於在所有資料的 3438 筆資料中，不違約案件發生的比例約佔 93%，若直接用所有資料來建構模型，會有違約/不違約案件

極度不對稱的情形發生，因此本研究針對建模之訓練樣本中不違約樣本進行抽樣，挑選與違約樣本等比例之資料組合(詳細資料如表 5-1 所示)，使得評估模型能夠更準確的判別出兩類型樣本是否違約。

表 5-1 樣本資料組合

2001 年 ~2005 年	所有資料			訓練樣本	
	不違約	3213	3438	225	450
違約	225	225			
2006 年	所有資料			測試樣本	
	不違約	93	124	93	124
	違約	31		31	

本研究依案例機構所提供之變數包括 16 個財務變數、15 個非財務變數與 1 個經濟變數 (詳見表 5-2)，而各變數評分標準由該金融機構制定。其中財務變數由申請借款企業所提供之財務報表中取得，非財務變數則為該金融機構派遣專業審查員至申請借款之企業實地訪問後，對該企業進行評分，而經濟變數中之經濟指標則是萃取評分年度的多項總體經濟指標而來。

表 5-2 變數說明表

財務變數		非財務變數	
代碼	變數名稱	代碼	變數名稱
F1	自有資金率	X1	公司歷史
F2	負債比率	X2	內部員工忠誠
F3	長期資金 / 固定資產	X3	背景
F4	流動比率	X4	經營理念
F5	速動比率	X5	資力
F6	債務償債能力-DSR	X6	財報可信
F7	淨值週轉數	X7	法規政策

F8	平均收款天數	X8	經濟因素
F9	平均銷貨天數	X9	產業展望
F10	毛利率	X10	經營展望
F11	淨利率	X11	銷售
F12	淨值報酬率	X12	公司行銷網健全
F13	每股淨利成長率	X13	市場開拓能力
F14	營業額成長率	X14	經營團隊
F15	週轉率	X15	同業客戶評價
F16	Ln(淨值/相對規模)	經濟變數	
		E1	經濟指標

## 5.2 建立兩階段風險評估模型

利用本研究方法建立兩階段之風險評估模型過程如下：

### 1. 第一階段風險評估模型

此階段利用統計軟體 SAS 8.0 進行逐步邏輯斯迴歸。利用所收集到的五年資料進行模式建構，並導入邏輯斯迴歸，利用逐步篩選法來選取重要變數，其建構之判別模式如下：

$$Z = \ln \frac{p}{1-p} \quad (23)$$

$$= 0.8542 + 0.5626 * \text{負債比率} - 0.765 * \text{長期資金/固定資產} + 0.2755 * \text{速動比率} + 0.9654 * \text{淨值週轉率} - 0.5361 * \text{平均銷貨天數} - 0.2474 * \text{營業額成長率} + 0.3952 * \text{Ln(淨值/相對規模)} + 0.613 * \text{公司歷史} + 0.3293 * \text{內部員工忠誠} + 0.225 * \text{財報可信} + 0.3639 * \text{經濟因素}$$

$$p = \frac{1}{1 + e^{-Z}} \quad (24)$$

將各案件資料代入式(23)中所得之 Z 值可透過式(24)轉換成機率值 p

值， $p$  值可視為樣本的不違約機率，愈接近 0 表示此樣本資料發生違約的機率越高，反之，愈接近 1 則為不違約樣本可能性愈大；將切點設定為由 0 及 1 開始，精確率低於 80% 即停止（如表 5-3 中第一階段之判別結果為不違約樣本的樣本數共 81 筆，其精確率為 80.25%），則該機率值即為切點，在此判別模式中，決定之兩最佳切點分別為 0.36 與 0.69，以此兩切點作為第一階段風險評估模型之判別分類界限，以樣本資料所得之  $p$  值進行判別，若  $p$  值低於 0.36 則判定此樣本為違約樣本，代表此樣本資料發生違約機率大，是不違約樣本的可能性低於 0.36；若  $p$  值高於 0.69 則代表此樣本資料為不違約樣本，不違約的可能性有 0.69 以上；而  $p$  值落於 0.36 至 0.69 的樣本資料則視為不易判別的模糊區域資料，將繼續進行第二階段之風險評估模型建構。

表 5-3 第一階段風險評估模型判別結果

原始樣本 \ 判別結果	不違約	違約
	不違約	65
違約	16	98
精確率	80.25%	80.33%

## 2. 第二階段風險評估模型

此階段利用 Weka3.5.6 軟體作為 SVM 之輔助分析工具。將未能

判別資料繼續進行第二階段之 SVM 模式建構。兩階段風險評估模式

結果如表 5-4 所示，並和只使用邏輯斯迴歸或 SVM 結果進行比較。

表 5-4 三種方法之風險評估模型建構結果

方法	預測 原始	判別結果						
		不違約	違約	準確率	整體準確率			
邏輯斯迴歸	不違約	143	82	63.6%	64.9%			
	違約	76	149	66.2%				
SVM	不違約	120	105	53.3%	52.0%			
	違約	111	114	50.7%				
兩階段	預測 原始	判別結果				兩階段判定準確數	兩階段準確率	整體準確率
		第一階段 (邏輯斯迴歸)		第二階段 (SVM)				
	不違約	違約	不違約	違約	154	68.4%		
	65	24	89	47				
	不違約	16	98	60	51	149	66.2%	
違約	65	24	89	47	154	68.4%		

從表 5-4 中可看出以本研究所提出之兩階段風險評估模型的整體準確率不僅高於僅使用邏輯斯迴歸或 SVM 方法所建構的模型準確率，其對於違約與不違約兩組的判別正確率相較於邏輯斯迴歸或 SVM 單一方式也較對稱，故可證實本研究所提出之兩階段風險評估模型可降低利用此兩種單一方法所建構的風險評估模型不對稱性的情形，並可提高模型之整體準確率。

### 模型測試結果

整體而言，本研究所提出之兩階段方法預測能力可達 62.9%，而邏輯斯迴歸與 SVM 方法建構的模型預測能力僅有 48%與 56%，且兩

階段模型可以同時保有較佳的預測能力，不會偏重於其中某類樣本的預測，依上述結果可以驗證本研究提出之兩階段風險評估模型確實有良好的預測效果與再現性，其結果如表 5-5。

表 5-5 三種方法之風險評估模式測試結果

方法	預測 原始	判別結果						
		不違約	違約	準確率	整體準確率			
邏輯斯迴歸	不違約	34	59	36.6%	48.4%			
	違約	5	26	83.9%				
SVM	不違約	56	37	60.2%	56.5%			
	違約	17	14	45.2%				
兩階段	預測 原始	判別結果						
		第一階段 (邏輯斯迴歸)		第二階段 (SVM)		兩階段判 定準確數	兩階段 準確率	整體 準確率
	不違約	違約	不違約	違約				
	不違約	10	28	44	11	54	58.1%	62.9%
	違約	2	18	5	6	24	77.4%	

### 模型驗證

為了驗證本研究方法確實可提高準確率與降低不對稱性，對不違約樣本進行五次抽樣並和違約樣本配對，利用上述方法建構五組不同的模型，同時也比較單獨只使用邏輯斯迴歸或 SVM 方法的風險評估模型判別相同訓練樣本的結果，並將此十五組模型的準確率結果進行完全隨機集區設計的變異數分析(Analysis of Variance, ANOVA)檢定，檢定不同的實驗方法對於模型的準確率和對稱性是否有顯著差異。結果如表 5-6 所示。

表 5-6 三種方法模型判別與檢定結果

檢定項目	不違約樣本 準確率(%)			違約樣本 準確率(%)			整體 準確率(%)			準確率 差距(%)		
	邏輯斯	SVM	兩階段	邏輯斯	SVM	兩階段	邏輯斯	SVM	兩階段	邏輯斯	SVM	兩階段
1	49.33	55.11	68.00	79.11	50.22	60.89	64.22	52.67	64.44	29.78	4.89	7.11
2	72.00	53.33	64.89	59.56	55.11	60.00	65.78	54.22	62.44	12.44	1.78	4.89
3	51.11	58.22	61.78	71.56	59.11	57.33	61.33	58.67	59.56	20.44	0.89	4.44
4	68.00	57.78	61.33	63.11	60.44	63.56	65.56	59.11	62.44	4.89	2.67	2.22
5	63.56	53.33	68.44	66.22	50.67	66.22	64.89	52.00	67.33	2.67	2.67	2.22
Mean	60.80	55.56	64.89	67.91	55.11	61.60	64.36	55.33	63.24	14.04	2.58	4.18
P-value	0.157741			0.036706*			0.003762*			0.030058*		

由檢定結果顯示在訓練樣本的模型建構上，三種方法在模型違約準確率、整體準確率與模型判別對稱性上確實有顯著差異，其 P 值 (P-value) 皆小於 0.05，本研究所提出之建構兩階段模型之方法顯著優於 SVM 方法；在模型判別的對稱性中，兩階段模型也顯著優於邏輯斯迴歸模型。<sup>P-value<0.05:\*</sup>

三種方法中以本研究所提出之兩階段方法建構模型的準確率最為穩健，能夠同時擁有較高之準確率（平均模型整體準確率有 63.24%）及較對稱性的優點（違約與不違約樣本平均準確率的差距只有 4.18%），且針對不同樣本組所建構的模型變異也較小。由此可知，本研究所提出之兩階段模型確實有效且穩健可行。

將前述十五組樣本所建構的模型進行模型測試，比較其模型的預

測能力與再現性好壞，由檢定結果可知，三種方法在模型測試上四個檢定項目（不違約樣本、違約樣本、整體準確率與模型對稱性）中皆有顯著差異，其結果列於表 5-7。

表 5-7 三種方法模型測試與檢定結果

檢定項目	不違約樣本 準確率(%)			違約樣本 準確率(%)			整體 準確率(%)			準確率 差距(%)		
	邏輯斯	SVM	兩階段	邏輯斯	SVM	兩階段	邏輯斯	SVM	兩階段	邏輯斯	SVM	兩階段
1	22.58	59.14	54.84	96.77	54.84	80.65	41.13	58.06	61.29	74.19	4.30	25.81
2	59.14	59.14	60.22	70.97	64.52	61.29	62.10	60.48	60.48	11.83	5.38	1.08
3	36.56	63.44	59.14	87.10	38.71	67.74	49.19	57.26	61.29	50.54	24.73	8.60
4	32.26	54.84	52.69	80.65	54.84	80.65	44.35	54.84	59.68	48.39	0.00	27.96
5	36.56	60.22	58.06	83.87	45.16	77.42	48.39	56.45	62.90	47.31	15.05	19.35
Mean	<b>37.42</b>	<b>59.35</b>	<b>56.99</b>	<b>83.87</b>	<b>51.61</b>	<b>73.55</b>	<b>49.03</b>	<b>57.42</b>	<b>61.13</b>	<b>46.45</b>	<b>9.89</b>	<b>16.56</b>
P-value	<b>0.002540*</b>			<b>0.001934*</b>			<b>0.007039*</b>			<b>0.006698*</b>		

P-value<0.05:\*

從表 5-7 中可看出邏輯斯迴歸模型容易將其預測結果偏向其中某類樣本上，在違約樣本上有高度預測能力，平均預測能力有 83.9%，但不違約樣本預測準確率僅 37.4%，其預測結果不佳。SVM 模型的違約樣本預測能力不佳，為三種方法中較差的預測結果；而本研究所提出之兩階段模型預測結果平均模式整體預測準確率有六成以上 (61.13%)，對稱情形明顯優於邏輯斯迴歸模型的方法，且對於不違約樣本(平均預測準確率 57%)與違約樣本(平均預測準確率 74%)皆可以有效判別，因此推論本研究所提出之兩階段風險評估模型在預測能

力上同樣較其他兩方法所建構之風險評估模型有較好的表現。

## 陸、 結論

本研究所建構之兩階段風險評估模型主要以台灣中小企業為研究範圍，因此可作為台灣金融機構進行放款決策的依據。

本研究利用台灣某金融機構近年放款歷史資料進行實證分析，驗證本研究所提出之兩階段風險評估模型確實能達到降低判別模型的不對稱性，可有效提高整體模型準確率及預測借款企業好壞。

目前尚未見有中外文獻利用兩階段之風險評估模型做為金融機構放款決策的依據，因此本研究具有創新性。本研究所建構之兩階段風險評估模型主要以台灣中小企業為研究範圍，因此可作為台灣金融機構進行放款決策的依據。利用邏輯斯迴歸與 SVM 所建立的判別模型，方法簡單，利用軟體操作則實用性高，實用容易，因此對於金融機構而言，甚具應用價值。

## 參考文獻

- [1] 王濟川、郭志剛，「Logistic迴歸模型-方法與應用」，五南出版社，第二版，2004。
- [2] 沈大白、張大成、劉宛鑫，「信用評等模型之簡介」，中國商銀月刊，2000。
- [3] 翁霓譯，「商業銀行管理政策」，幼獅文化事業公司與台北銀行，

- 1987。
- [4] 崔運驊，「應用償還率與風險評估模式建構金融機構放款評等流程」，國立交通大學工業工程與管理學系碩士論文，2007。
- [5] 黃啟峰，「利用風險評估與存活期預測模式構建台灣中小企業信用評等流程」，國立交通大學工業工程與管理學系碩士論文，2006。
- [6] Altman, E. I., "Financial Ratio Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, Vol. 23, pp.588-609, 1968.
- [7] Baesens, B., Van, G.T., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen J, "Benchmarking state-of-the-art classification algorithms for credit scoring", *Journal of the Operational Research Society*, 54, pp.627–635, 2003.
- [8] Desai, V.S., Crook, J.N., Overstreet, G.A., "A comparison of neural networks and linear scoring models in the credit union environment.", *European Journal of Operational Research*, Vol.95, No.1, pp.24–37, 1996.
- [9] Edmister, R.O., "An empirical test of financial ratio analysis for small business failure prediction", *The Journal of Financial and Quantitative Analysis*, pp.1477-1493, 1972.
- [10] Min, J.H., Lee, Y.C., "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters", *Expert Systems with Applications*, Vol.28, No.4, pp.603-614, 2005.
- [11] Shin, K.S., Lee, T.S., Kim H.J., "An application of support vector machines in bankruptcy prediction model", *Expert Systems with Applications*, Vol.28, No.1, pp.127-135, 2005.
- [12] Vapnik, V., "Support-vector networks," *Machine Learning*, Vol. 20, No.3, pp. 273-297, 1995.