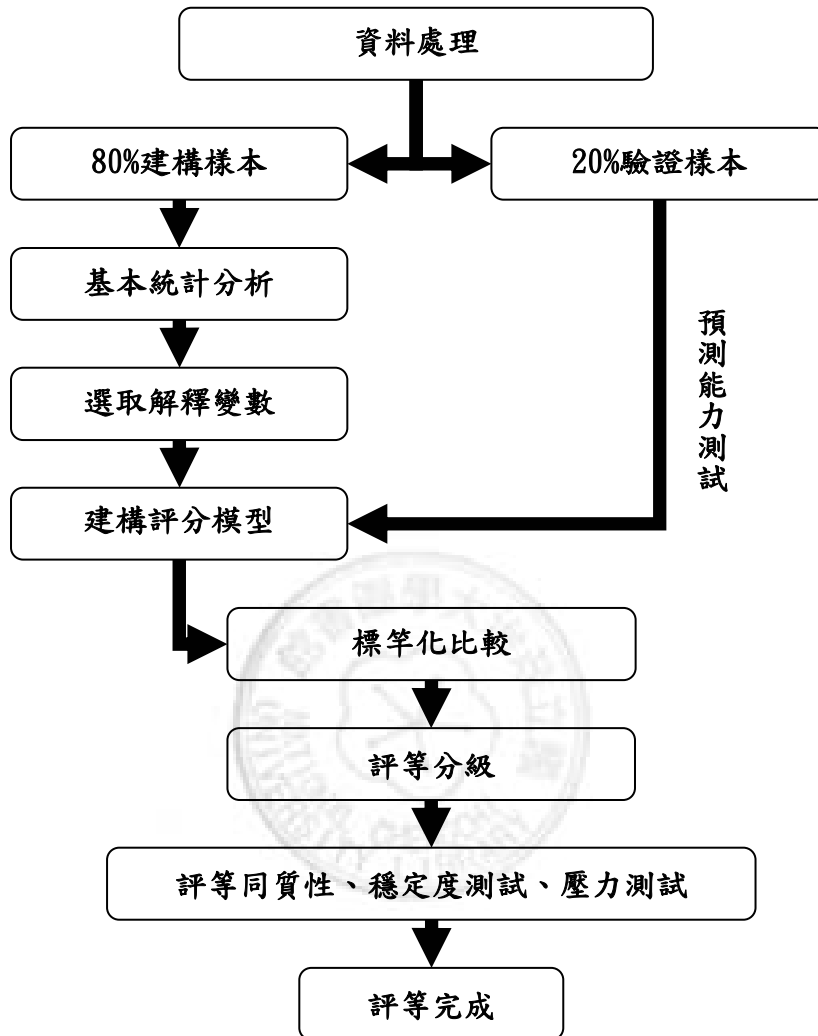


### 第三章 研究方法

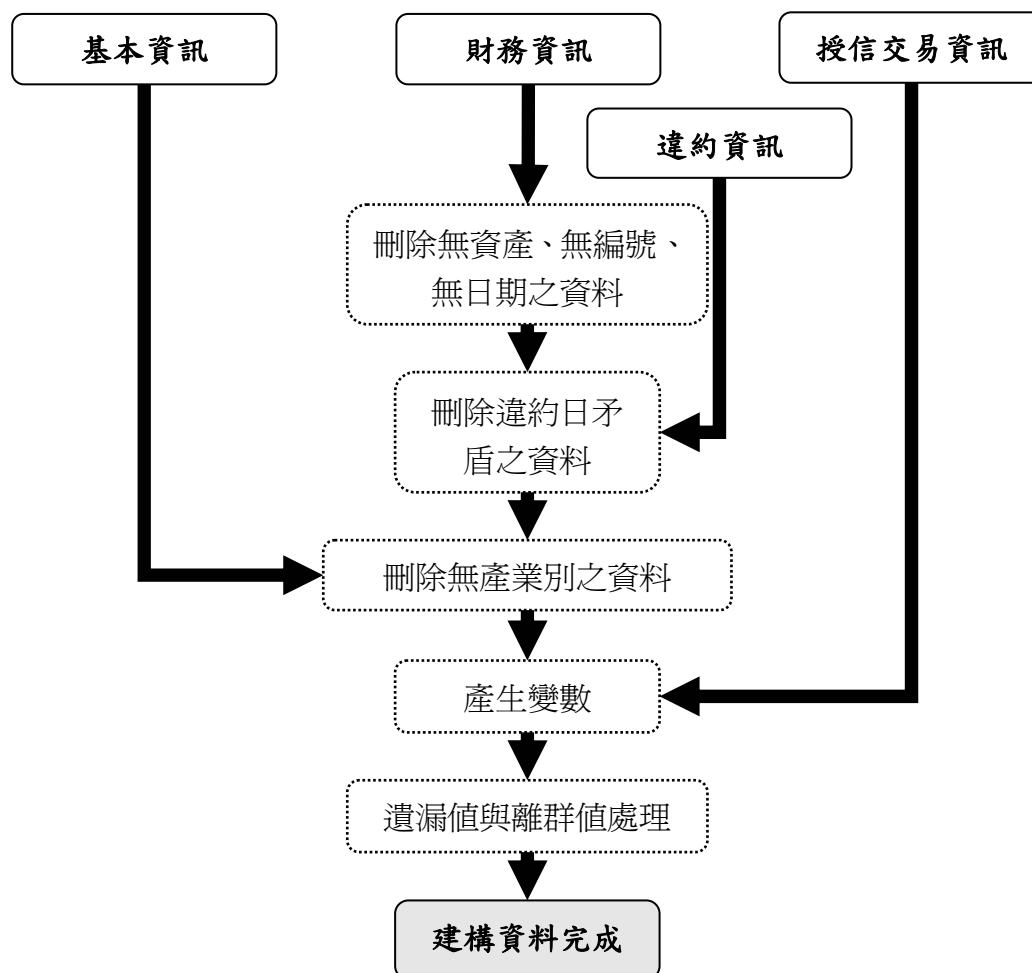
#### 第一節 建構模型程序



圖一：建構模型程序

## 第二節 資料處理

### 第一項 資料處理程序



圖二：資料處理程序

### 第二項 原始資料敘述與處理

模型建構之原始資料，包括 X 銀行 2002 至 2004 年具財務簽證之企業財務資訊，及聯合徵信中心(JCIC) 2003 至 2005 年企業與企業主之授信交易資訊。前者包括企業與企業主的基本資料及企業的財務報表；後者包括企業與企業主之授信交易行為資訊，如授信異常記錄，包含之資訊如下表。

表六：聯徵資料概要

聯合徵信中心資料	
企業戶聯徵資料	企業負責人聯徵資料
基本資料	基本資料
公司股票金額與股數資料	信用異常紀錄
信用異常紀錄	授信資料明細及還款紀錄資訊
授信資料明細及還款紀錄資訊	授信餘額變動資料
授信餘額變動資料	信用卡繳款資訊
逾期催收及呆帳資料	被查詢紀錄
退票異常紀錄	
拒絕往來紀錄	
被查詢紀錄	

處理原始資料的第一步就是處理來自X銀行內部之貸款企業戶的資訊。這些資訊包括了企業財務報表資料、基本資料與貸款違約資訊(日期)，分屬於不同的檔案中，我們以企業戶編號為鍵值作合併處理，再對合併後之資料集檢閱後做以下之處理。

- 刪除「總資產」欄位遺漏之資料：由於總資產是企業財務資訊的基礎，因此若總資產為遺漏值，則將該筆資料從建模資料集裡刪去。

- 刪除無財報公布日或企業戶編號之資料：不論在建模的過程中作合併或分析的處理時，兩者皆為重要之鍵值，因此若無財報公布日或企業戶編號，則將該筆資料從建模資料集裡刪去。

- 刪除違約日矛盾之資料：財務報表資訊與違約日期以企業戶編號為鍵值作合併，若企業戶的違約日期在財報公布日(資料日)之前，則不合常理，應將該筆資料從建模資料集裡刪去。

- 刪除無產業類別之資料：本研究乃針對企業型暴險建構內部評等模型，因此須將金融業等非企業型暴險之資料排除在母體外，若資料無產業類別便無從判別之，故刪除之。

- 資料日為西元2003至2004年：財務報表資訊之年度為2002至2004年，而企業與企業主之授信交易資訊之年度為2003至2005年，因此資料在合併處理後，用於

建模之資料年度為西元2003至2004年。

### 第三項 產生變數

初步資料處理完成後的下一步驟是產生變數，本研究將變數分為三大類：財務變數、授信交易變數、以及基本特性變數。其中財務變數再分為四小類：財務結構、經營效率、償債能力、以及獲利能力；授信交易變數分為企業戶交易資訊與負責人交易資訊兩類；基本特性變數則包含產業與規模變數。詳細變數定義與代碼請參閱附錄。

### 第四項 離群值與遺漏值處理

由於各個變數之離群值分佈不盡相同，因此須一一個別處理，處理方式係將離群值以該變數最大或最小值取代。遺漏值的處理係以該變數最大與最小值間隨機產生之數值填補。下表為模型變數之遺漏值比率與離群值比率列表，全變數的完整列表請參閱附錄。

表七：模型變數遺漏值比率與離群值比率

變數代碼	變數定義	遺漏值比率(%)	離群值比率(%)
B1	累計折舊 / 固定資產毛額	1.49	<0.01
F1	負責人是否有現金卡	0	<0.01
G2	長期借款是否增加	0	0
B7	營業收入 / 固定資產淨額	0	<0.01
C22	(現金及約當現金 + 短期投資) / 總資產	0	0
C29	銀行借款 / 總資產	0	0
C32	銀行長期借款 / (銀行長期借款 + 股東權益)	0	<1
D18	稅後淨利 / 總資產	0	0
E3	產業類別	0	0

## 第三節 基本統計分析

### 第一項 模型建構資料

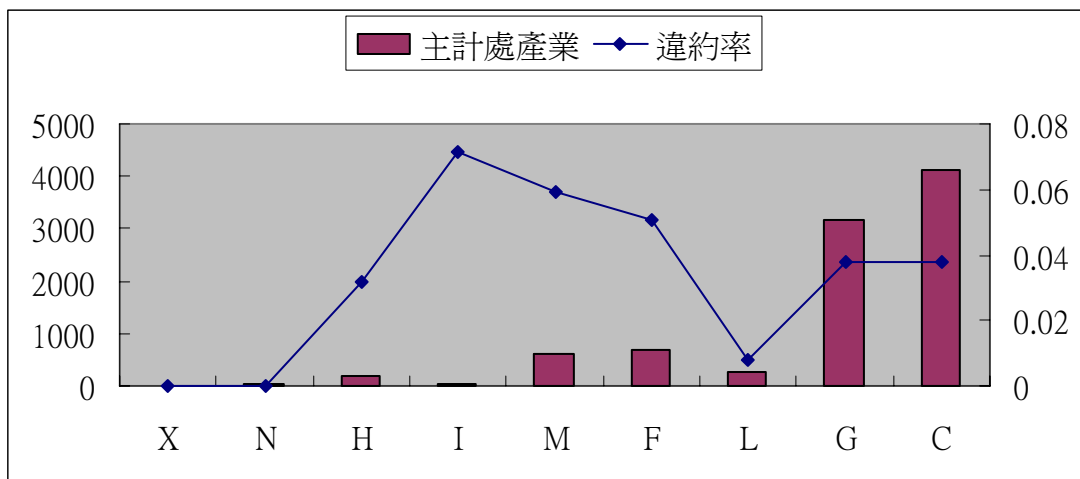
企業型暴險採用 IRB 基礎法，必須至少具備 5 年以上之資料用以估計違約機率，但依據過渡期間規定，銀行於 96 年起開始施行此架構之前，得具備 2 年以上之歷史資料，在爾後的 3 年過渡期間內，每年增加 1 年資料，因此本研究之模型建構資料雖僅包含 2 年資料，仍符合前述規定。

由陳錦村、江玉娟與朱育男(2006)的實證結果可知在不同產業所建構之信用評分模型，其解釋變數不盡相同，但由於本研究受限於資料數目較少(下表)，不足以依據產業類型分別建構模型，因此僅將產業因素以解釋變數的型式納入模型之中。

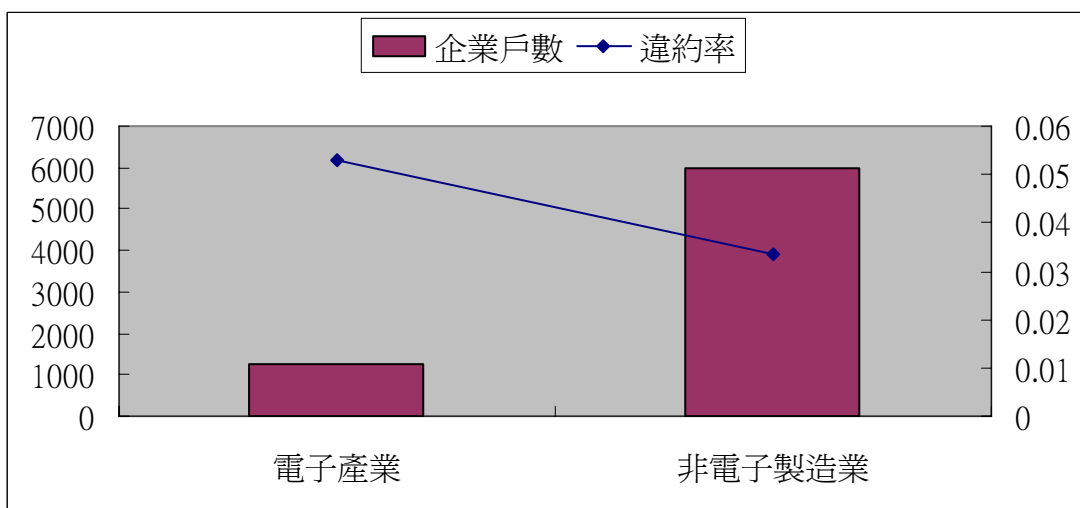
表八：資料違約分佈

企業戶	%	計數
非違約戶	96.07	8794
違約戶	3.93	360
總和	100	9154

利用主計處公佈之產業分類將建模資料分群(非營利企業為本研究自行分類)，接著檢視每一群組違約率之差異後(與下圖)，顯示不同產業群其違約率亦有不同，因此進一步將違約率相近之產業群作合併處理，較高之住宿及餐飲業、專業、科學及技術服務業、營造業併為一群，其餘產業類型另為一群。再者，由於電子產業之性質亦與一般製造業有所不同，例如陳錦村、江玉娟與朱育男(2006)的研究證實電子產業與傳統產業之評分模型有顯著不同，並且在檢視兩者違約率後，確實存在差異(如下圖)，因此本研究亦進一步將電子產業獨立於製造業之外。最後，再經過以上分析處理之後，便產生產業分群變數 E3(如下表與下圖)。



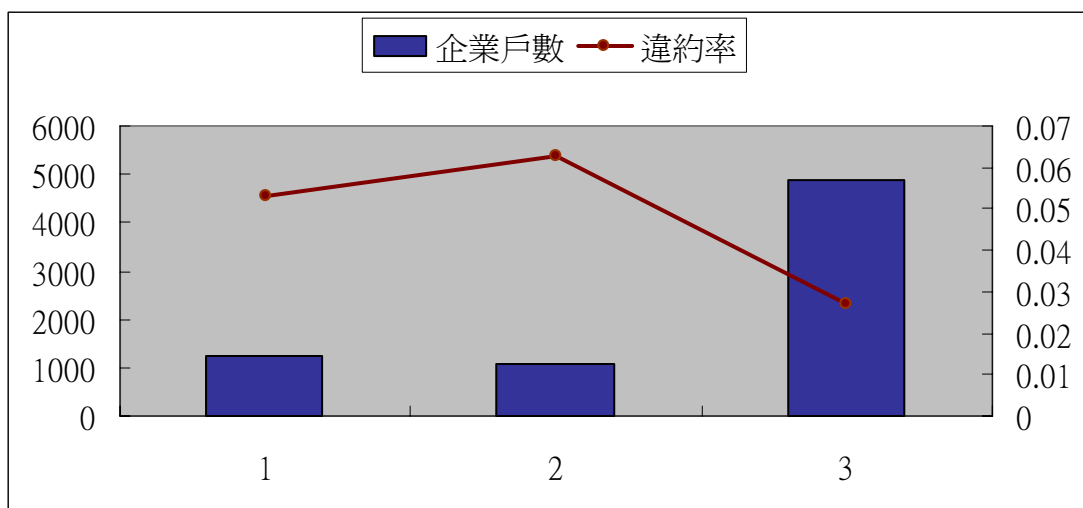
圖三：產業大類違約分佈



圖四：電子製造業違約分佈

表九：產業分類與再分類後違約分佈

產業類型	代碼	企業戶數	違約戶數	違約率	指標變數 E3	企業戶數	違約戶數	違約率
住宿及餐飲業	I	56	4	0.071429	2	1372	76	0.055394
專業、科學及技術服務業	M	610	36	0.059016				
營造業	F	706	36	0.050992	3	6204	184	0.029658
非營利	X	6	0	0				
支援服務業	N	48	0	0				
運輸及倉儲業	H	190	6	0.031579	1	1578	100	0.063371
不動產業	L	250	2	0.008				
批發及零售業	G	3164	120	0.037927				
製造業	C	4124	156	0.037827				

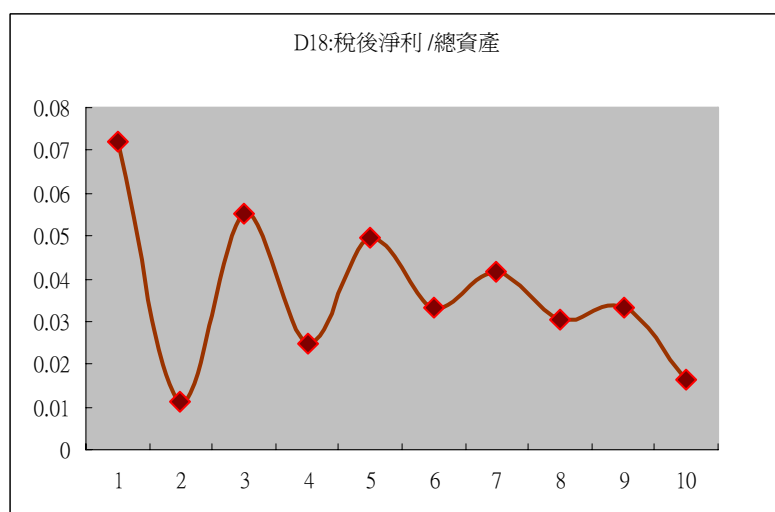


圖五：產業再分類後違約分佈

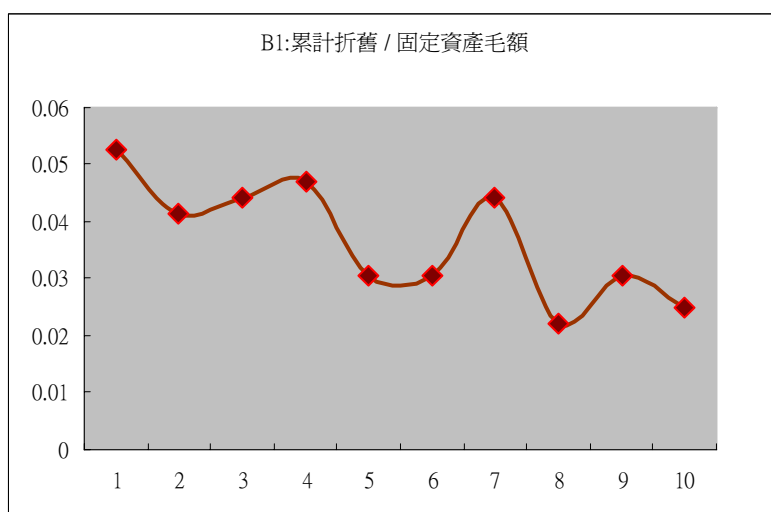
## 第二項 單變量分析

單變量分析在此係指針對個別變數與違約機率，探討兩者間是否存在單調性關係，以作為建構模型時選取變數的參考，亦可作為模型建構後初步判斷是否正確的依據。

如圖所示，若為連續型變數，則在排序後作十分位切割處理，再計算每一分位群組之違約戶比率，最後繪製成圖。本節僅對於模型中選取之解釋變數作分析，其中產業分群變數已於前節分析，其餘變數分析請參閱附錄。

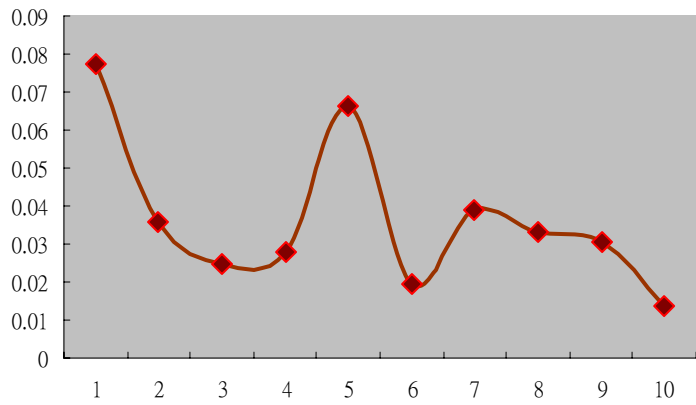


圖六：變數 D18 違約分佈



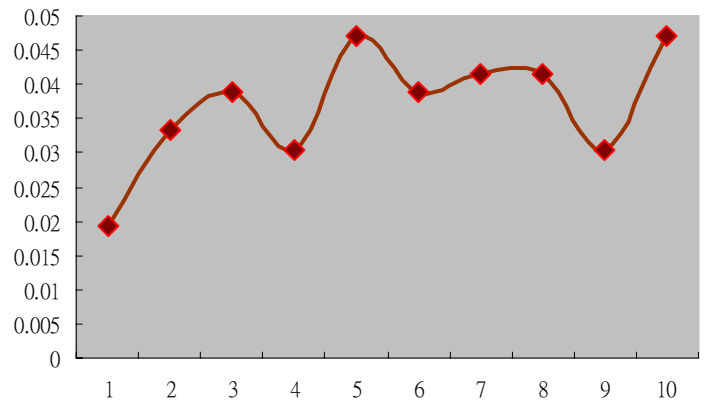
圖七：變數 B1 違約分佈

C22:(現金及約當現金 + 短期投資) / 總資產



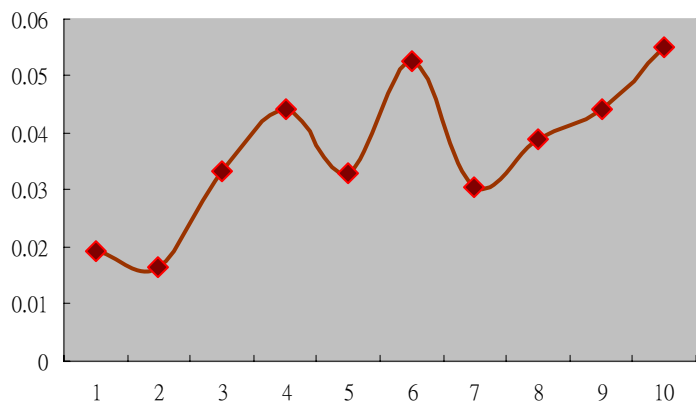
圖八：變數 C22 違約分佈

B7:營業收入 / 固定資產淨額



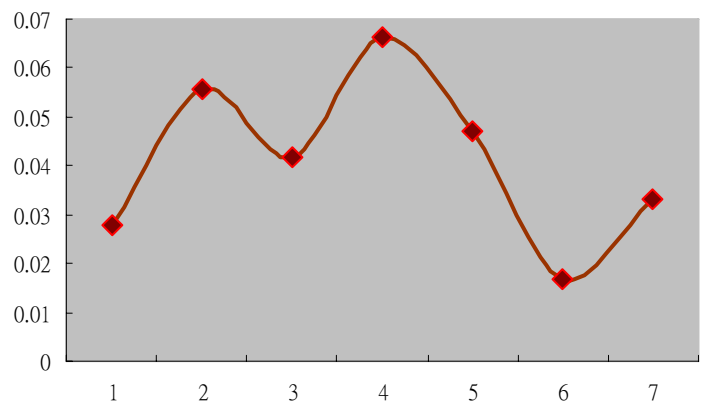
圖九：變數 B7 違約分佈

C29:銀行借款 / 總資產



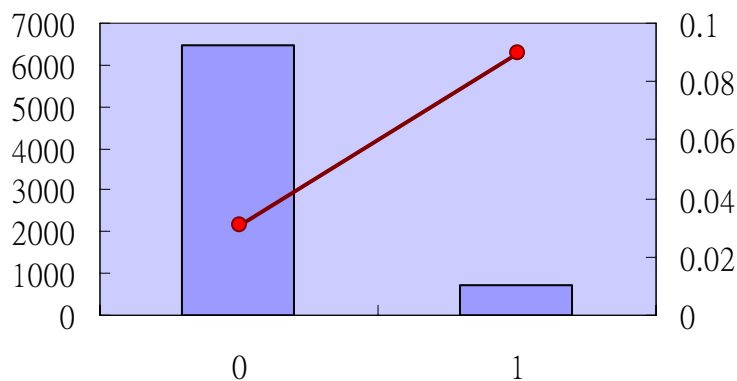
圖十：變數 C29 違約分佈

C32:銀行長期借款 / (銀行長期借款 + 股東權益)



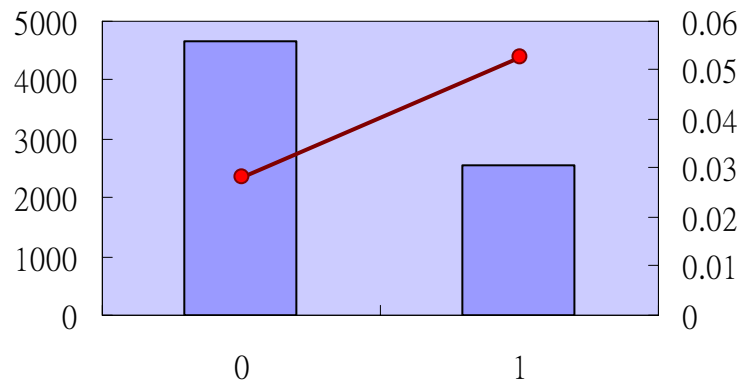
圖十一：變數 C32 違約分佈

F1:負責人是否持有現金卡 ● 違約率



圖十二：變數 F1 違約分佈

G2:長期借款是否增加 ● 違約率



圖十三：變數 G2 違約分佈



### 第三項 相關性分析

當參與模型的變數發生高度相關的情形便產生共線性(collinearity)的問題，有可能使參數估計量不穩定、參數估計量的正負符號不合理、模型顯著性測試結果不準確等問題，因此這些變數不應納入模型之中，例如，若模型包含了兩高度相關的財務結構變數，其中一個可能會和另一個變數的正負號相反，而且也會影響Wald檢定值以及兩變數的參數估計量，一般來說，解決的方法就是從這群變數中選取其一納入模型中。

我們通常利用相關係數的計算來檢驗變數間的相關性，由於本研究使用的變數多為連續型變數，其餘為指標變數(dummy variable)，因此藉由Pearson相關係數來計算之，但由於變數總數目相當多，所以僅就模型選入之變數計算並於下一章節陳述。計算方法如下：

假設兩個連續變數 $X$ 與 $Y$ ，平均數分別為 $\bar{X}$ 與 $\bar{Y}$ ，標準差分別為 $S_X$ 與 $S_Y$ 。Pearson相關係數 $r$ 為：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y}$$

其中 $X_i$ 和 $Y_i$ 是變數 $X$ 及變數 $Y$ 的第 $i$ 個值， $n$ 是資料的數目。

## 第四節 評分模型建構與驗證

### 第一項 羅吉斯迴歸模型

一般迴歸分析與羅吉斯迴歸模型不同的是，前者處理連續型應變數，但若迴歸模型之應變數呈現類別型式時，便無法滿足迴歸分析的基本假設，此時迴歸分

析可能就不適用；而後者則可處理二元或多元的類別型式變數，以本研究之信用評等模式為例，即預測企業是否違約，並進一步估計公司違約的機率，模型如下：

$$Y_i : i \text{ th 樣本之二元應變數值, } Y_i = \begin{cases} 1 \\ 0 \end{cases}$$

$$X_i : i \text{ th 樣本之解釋變數值, } X_i = (X_{1i}, X_{2i}, \dots, X_{pi})$$

$$P_i = P(Y_i = 1 | X_i)$$

$$\text{logit } P_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$$

$$\text{logit } P \equiv \log \frac{P}{1-P}$$

$$\text{i.e. } P_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}}}$$

$$\begin{aligned} \beta_i &= \text{logit } P(Y_i = 1 | X_i = r+1) - \text{logit } P(Y_i = 1 | X_i = r) \\ &= \ln \frac{P(Y_i = 1 | X_i = r+1) / P(Y_i = 0 | X_i = r+1)}{P(Y_i = 1 | X_i = r) / P(Y_i = 0 | X_i = r)} \end{aligned}$$

利用最大概似估計對參數  $\beta$  估計，概似函數為：

$$\begin{aligned} &\prod_{i=1}^k \binom{n_i}{c_i} p_{X_i}^{c_i} (1-p_{X_i})^{n_i-c_i} \\ &= \prod_{i=1}^k \binom{n_i}{c_i} \left( \frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}} \right)^{c_i} \left( \frac{1}{1 + e^{\alpha + \beta X_i}} \right)^{n_i-c_i} \end{aligned}$$

找出使以上之機率為最大的  $\alpha, \beta$ ，其中  $n_i$  為樣本數， $c_i$  為應變數為 1 的個數

## 第二項 預測能力測試

預測能力測試主要驗證評等模型是否能夠區別信用狀況好壞，並評估模型誤差率是否在合理可接受範圍內。檢驗的方法有數種，包括 Kolmogorov - Smirnov test、Gini 係數、CAP 等等方法，以下僅就本研究所使用的 ROC 方法以及預測分類矩陣做概述。

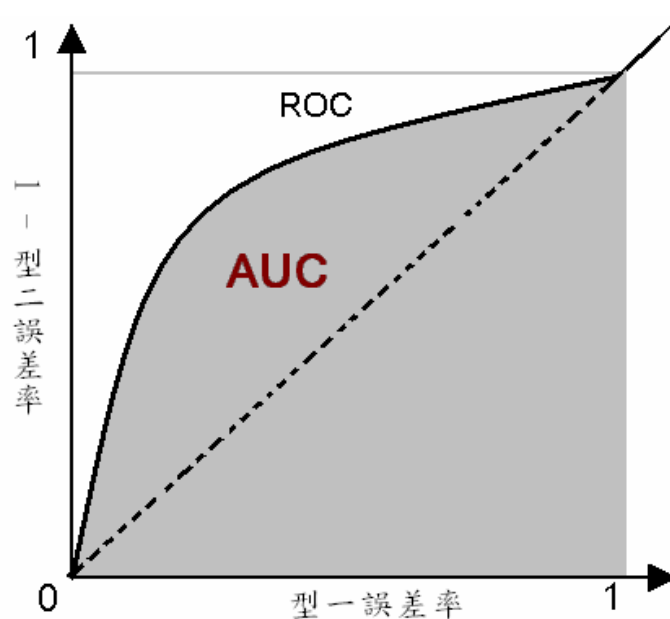
預測分類矩陣是一檢驗模型預測能力最簡易的方法，其方式係利用設定之截斷點  $c$  將模型估計之機率值分類，做為預測值，再與實際觀察值作比較。以羅吉

斯迴歸違約預測模型為例，若模型估計機率值大於截斷點( $c=0.5$ )，則預測值為違約( $\hat{Y}=1$ )；反之，若模型估計機率值小於截斷點，則預測值為非違約( $\hat{Y}=0$ )，

表十：預測分類矩陣

		預測		
		非違約( $\hat{Y}=0$ )	違約( $\hat{Y}=1$ )	總和
實際	非違約( $Y=0$ )	a	b	a+b
	違約( $Y=1$ )	c	d	c+d
	總和	a+c	b+d	a+b+c+d

如以上所述，當我們利用設定截斷點  $c$  建立預測分類矩陣時會產生型一誤差（實際為非違約戶被預測為違約戶）以及型二誤差（實際為違約戶被預測為非違約戶），ROC 曲線的繪製即依據不同截斷點產生的型一誤差率及型二誤差率而來，當曲線的凹向(concave)程度越強、越接近(0, 1)，代表模型愈能將違約戶與非違約戶區別，意即當 ROC 曲線下面積，AUC(Area Under Curve)，越大則模型的預測能力愈強。若 AUC 為 1，則代表模型可以完全區別違約戶與非違約戶；若 AUC 為 0.5，則代表隨機形式區別。



圖十四：ROC 曲線圖

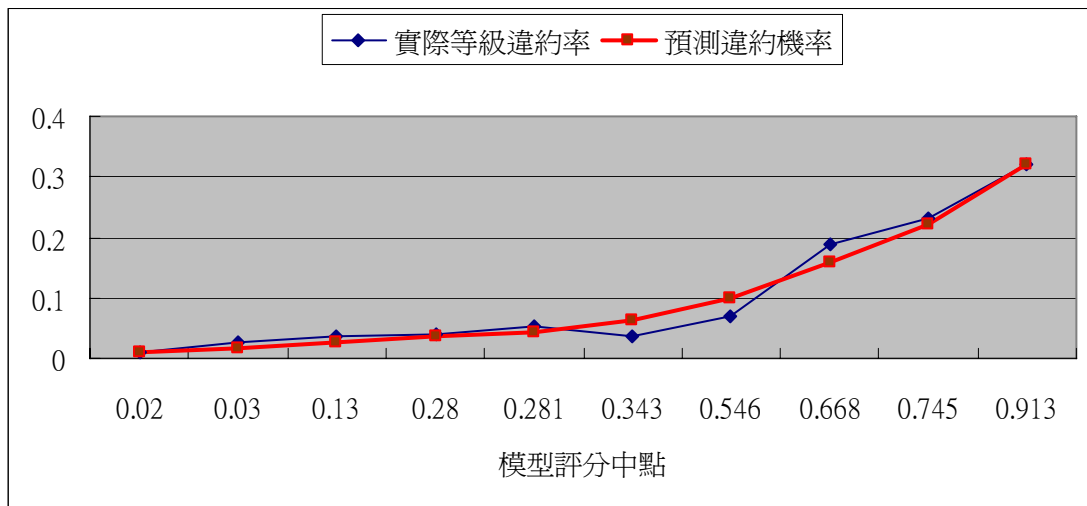
### 第三項 標竿化比較

本研究藉由比較相同樣本、不同模型方法(類神經網路)下建構之模型，以及外部模型的預測能力，來完成標竿化測試。外部模型係某知名管理顧問公司針對相同原始資料建構之模型，但本研究僅取得模型的預測分類矩陣結果，因此針對此項做為本研究的標竿化比較之一。

## 第五節 評等分級與其驗證

### 第一項 評等分級

由於羅吉斯模型係由重抽樣建置，可能使得產出之羅吉斯模型評分不能與母體實際違約率一致，倘若發生兩者不一致的情形便無法引入羅吉斯模型評分做為預測違約機率，因此須要進一步做調整。本研究所稱之調整，係指將模型評分調整至與母體實際違約率一致，本研究調整方法如下述。首先依模型評分分級後(圖例為 10 個評等等級)，計算每一評等等級評分中點，如下圖中的橫軸；接著計算每一等級實際的違約率，即每一評等的實際違約戶數目除以評等內企業戶數目，如下圖中藍色折線；再利用最小平方法估計出最佳曲線，如下圖中紅色曲線，爾後便可利用該估計式將羅吉斯模型評分對映至調整後之預測違約機率。



圖十五：預測違約機率調整圖例

### 第二項 評等穩定度測試

評等穩定度的測驗係觀察評等結果的變動趨勢，檢視其震盪的幅度是否合

理，其測試方法係建立評等轉移矩陣(transition matrix)。依據孫銘誼、王思芳(2004)的研究指出，建置轉移矩陣的方法包括了本研究所使用的分類統計方法(cohort approach)，以及連續統計方法(duration approach)。以下就分類統計方法做概述。

令 $P_{ij}(\Delta_t)$ 為經過 $\Delta_t$ 期後評等 $i$ 移轉至 $j$ 的機率，若期初時評等 $i$ 的企業授信戶數目為 $n_i$ 個，經過 $\Delta_t$ 期後轉為等級 $j$ 的授信戶有 $n_{ij}$ 個，則轉移機率估計值為 $P_{ij}(\Delta_t) = n_{ij} / n_i$ 。

### 第三項 等級同質性驗證

依據金管會公佈的「銀行自有資本與風險性資產計算方法說明」以及孫銘誼、王思芳(2004)的研究，評等等級同質性驗證的方法有 Binomial test、Granularity Adjustment、Moment Matching、CIER 等幾種，本研究僅使用 Binomial test 做為等級同質性驗證的方法。

一良好的評等分級應將不同信用品質的企業授信戶有效區隔，使信用品質相近的授信戶被分配至同一評等等級，因此同一評等等級的企業授信戶其差異最小，等級內的組內變異達到最小。故若假設同一評等等級之企業授信戶的違約機率相同，則會呈現二項式分配，違約的與否彼此獨立，企業授信戶的違約要素同質。

依據每一評等等級的估計違約機率與該等級內企業戶數目，利用 Binomial test 產生在信心水準  $\alpha$  下的違約數目臨界值，檢定之虛無與對立假設如下述。令 PD 為每一評等等級之估計違約機率。令 P 為每一評等等級之實際違約機率。因此，若評等等級中違約戶數目超出了臨界值，那麼在信心水準  $\alpha$  下，我們將拒絕實際違約機率等於估計違約機率之虛無假設。

虛無假設 H0: PD=P

對立假設 H1: PD≠P

$$d_{K, \alpha} = \min \left( d : \sum_{i=d}^K \binom{K}{i} P D^i (1 - P D)^{K-i} \leq \alpha \right)$$

#### 第四項 壓力測試

壓力測試係觀察未來經濟條件或對銀行信用暴險的可能變動下，對於內部評等模式的影響，藉以檢驗在不同情境下之信用評等模式的合理與穩定。最常用來作為情境測試的有經濟或產業衰退、市場風險事件及資金流量變化情形，但由於在資料上的限制，因此本研究僅以建模資料設立情境，檢驗可能出現的不同企業戶在模型中的評等情形。