

經濟論文  
中央研究院經濟研究所  
34:4(2006),541-590

# 台灣地區企業信用評分系統的 建置、驗證和比較

鍾經樊\*

中央研究院經濟研究所  
國立政治大學經濟學系  
國立中山大學經濟學研究所

黃嘉龍

國立台灣大學經濟學系博士生

黃博怡

實踐大學財務金融學系

謝有隆

國立政治大學經濟學系碩士

**關鍵詞:** 信用評分、驗證統計量、變量轉換與篩選、評分模型建置與驗證  
**JEL 分類代號:** C50, C53, G32

---

\* 聯繫作者: 鍾經樊, 中央研究院經濟研究所, 台北市 115 南港區研究院路二段 128 號。電話: (02) 2782-2791 分機 188; 傳真: (02) 2785-3946; E-mail: cfchung@econ.sinica.edu.tw。

## 摘 要

本文根據台灣新報資料庫所載數千家公開發行企業的三年財務報表資料,估計以違約與否為應變量、財務比例為解釋變量的Logit迴歸模型,並根據估計結果建置適用於企業信貸的信用評分卡。本文除詳細描述信用評分卡的建置過程外,特別強調信用評分卡雖可視為一個預測虛擬變量的計量經濟模型,但卻又和典型的實證經濟研究有很大的不同,這是因為不論在估計Logit模型之前還是在獲得迴歸係數之後,都必須採行諸多繁複的操作和分析,尤其重要的是建置信用評分卡前一定要保留部分樣本以驗證信用評分卡的預測績效。本文根據十多項驗證統計量就建模樣本之後的兩年資料來評估信用評分卡的預測績效,並證明我們所建置的信用評分卡在所有的績效指標上都優於台灣經濟新報所發佈的基本評等。在2006年底實施「新巴塞爾資本協定」的前夕,不論是對即將採行內部評等法以衡量信用風險的銀行,還是對政府金融監理機關,本文都有很適時的實用價值。由經濟學理來說,信用評分模型的發展有助於解決借貸市場中先天存在之訊息不對稱所造成的逆選擇和道德危機問題,從而增加借貸市場供給和需求兩方面的效率,更進而影響貨幣政策傳導機制,深具政策意義。

## 1. 前言

銀行預測一個借款戶未來還款行為的最佳指標，是其過去的借貸記錄、財務狀況、乃至於人格特質，所謂的「信用評分」（credit scoring），便是將多個能夠預測借款戶未來行為的解釋變量加以整合，以統計方法或是資料探勘方法（data mining）形成一個可相互比較的量化指標，以顯示借款戶在未來一特定期間內違約的可能性。<sup>1</sup> 銀行的借款戶可分為企業和消費者兩類，前者借款的目的是生產，而後者則是為了消費，借款目的不同，銀行審核這兩類借款戶的方式當然也有所不同，因此銀行多將其信貸組織分為企業金融（或稱法人金融）和消費金融（或稱個人金融）兩個部門。其中消費金融部門的內容較為繁雜，包括信用卡、現金卡、小額信貸、乃至於車貸、房貸等，為快速有效的審核數量龐大的消費貸款申請，信用評分是一個完全不可或缺的工具，至於企業金融部門中，則是以類似消費金融的中小企業信貸審核最需要信用評分。消費貸款和中小企業貸款的共同特徵是貸款規模較小，因此常被合稱為零售型貸款，本文所討論的信用評分可適用於所有的零售型貸款。

近年來美歐先進國家之消費金融得以急速擴張的一個重要原因是信用評分技術的迅速發展，銀行不論大小，多對其消費信貸產品建置了專屬的信用評分系統。信用評分在消費金融上的成功，自然也延伸至前此以傳統徵信方法審批貸款、或採廣設分行建立個人關係網絡以篩選客戶的中小企業信貸。本文主旨就是展示建置和驗證企業金融之信用評分系統的步驟，而這些步驟當然也適用於建置消費金融的信用評分系統。

國際清算銀行（Bank for International Settlement）屬下的巴塞爾銀行監督管理委員會（Basel Committee on Banking Supervision）自 1999 年 6 月頒佈「新巴塞爾資本協定」第一版草案後，即廣徵建言而多次修正其內容，於 2004 年

---

<sup>1</sup> 簡介信用評分的歷史及其製作過程的國內文獻可見丁正中（2004）和阮正治與江景清（2004）。信用評分和所謂的「信用評等」或「債信評等」（credit rating）都是對信用風險的評估和預測，但兩者之間不同之處在於，債信評等通常是對發行債券的機構（大多是能夠發行公司債的大型上市上櫃公司，或是發行公債的中央及地方政府）所進行的違約風險評估，其服務對象是投資者，而信用評分則是對借款戶（大多是消費者或是無法發行公司債的中小企業）所進行的違約風險評估，其服務對象是銀行。

6月發佈「資本衡量及適足性國際規範」(簡稱「新巴塞爾資本協定(Basel II)」),預訂自2006年底開始實施,要求由「最低資本需求」(minimum capital requirement)、「監理審查程序」(supervisory review process)及「市場紀律」(market discipline)三個支柱來強化銀行的風險管理。在「最低資本需求」方面,提出衡量信用風險、作業風險、及市場風險的準則,其中針對信用風險還分為標準法(standardized approach)、及內部評等法(internal ratings-based approach, IRB)兩種方法供銀行選擇使用。在「監理審查程序」方面,要求監理機關需對銀行適足資本的計提進行質量性評估,以進行必要的干預,而在「市場紀律規範」方面,則提出資訊公開揭露的多項規定,以使市場得以發揮監督的功能。信用評分系統在衡量信用風險之內部評等法中佔有核心的地位,因為信用評分不僅能協助銀行區分借款戶的信用等級,更能直接預測借款戶的違約機率(probability of default, PD),而這些違約機率就將是銀行計算其應計提資本的基本參數。採用內部評等法的銀行除須對其各項信貸產品建置信用評分系統,且為滿足「監理審查程序」和「市場紀律規範」的要求,尚須對政府監理機關以及社會大眾提出信用評分系統的驗證結果。在新巴塞爾資本協定實施的前夕,對信用評分系統建置與驗證的研究有其迫切性。

由經濟學理來說,信用評分模型的發展有助於解決借貸市場中先天存在之訊息不對稱所造成的逆選擇(adverse selection)和道德危機(moral hazard)問題:金融機構唯有在準確測度借款戶的債信後,才能正確選擇最值得貸款的人以避免逆選擇問題。也只有當借款戶的借貸記錄不論好壞都會影響其債信,進而影響未來續借的可能性時,借款戶才會盡力維護其債信,從而避免了道德危機問題。相關文獻請見 Klein (1992)、Vercammen (1995)、Padilla and Pagano (1997, 2000)、Jappelli and Pagano (1993, 2002)。

信用評分容許金融機構降低控管大量借款戶信用風險的成本,並因而得以掌控其信用風險全貌,從而更擴大其授信對象,從供給面增加借貸市場的效率(Mester, 1997; Jappelli and Pagano, 2002),沒有信用評分的發展,消費貸款過去四十年來的高速成長是不可能的。信用評分也容許借款戶更了解本身的信用及其市場價值,進而得以修補和改進信用,這自然有助於借款戶尋求更恰適的借款條件,促進借貸市場的競爭,從需求面增加借貸市場的效率。借貸市場供給和需求兩方面的效率提升可進而影響貨幣政策傳導機制,

深具經濟政策上的重要性。此外，信用評分有助於金融機構將信貸證券化，增加信貸次級市場形成的可能性，更進一步增加借貸市場的效率。而政府也可利用信用評分以更有效率的監理金融機構對消費貸款以及中小企業貸款的信用風險，這事實上也是「新巴塞爾資本協定」的建議。

台灣近年來中小企業放款因壞帳居高不下以及同業間過度競爭致使獲利縮小，銀行也遵循國際趨勢大幅轉進風險較高的消費金融業務，引進信用評分系統控管伴隨而來的信用風險乃無可避免的措施。而更因新巴塞爾資本協定實施在即，本土銀行多以此為提升風險管理能力的契機，引進信用評分系統遂瞬爾蔚為風潮。台灣本土銀行發展信用評分的歷程遠短於歐美先進國家，大型行庫多聘雇國際風險管理顧問建置以消費金融為主的信用評分系統，而中小型銀行則以雇用能複製歐美技術的本地金融顧問公司為主，也有銀行經由電腦資訊系統業者的仲介購買歐美日先進銀行所使用的風險管理軟體，嘗試自行組建信用評分系統。然而我們必須指出，當歐美先進國家的信用評分已朝向貸款使用率、貸款期中審查、提前還款解約、追繳欠款、預准貸款、簽帳審核、跨售、促銷、行銷、保險、客戶支出和付款行為模式、住屋抵押貸款、乃至風險定價等多類型應用時，台灣本土銀行的信用評分仍侷限在以信用卡審核為主的消費金融初級應用上。歐美先進國家的銀行大都已擁有獨立自主的專業技術，可針對不同信貸商品自行製作信用評分模型。而台灣本土銀行卻只能委外製作或是零星外購信用評分系統，所得到的只是成本以百萬元計但卻缺乏時效與精確度的短期應急措施。我們認為，對於有意和國際金融業接軌的銀行，強化風險控管人才的專業素質，自行建置本土信用評分系統才是長遠的做法，本文內容對金融業界有志於研發專屬信用評分系統者應有所助益。

國內學術界對信用評分的專業論文並不多見，相關實證研究大多是就一中小型借款戶資料，根據 Logit 或 Probit 模型或是資料探勘技術（以神經網絡以及決策樹方法為主）估計一個以違約與否為應變量的迴歸模型，再採用一至二個標準的模型配適統計量比較估計模型的表現，這樣的研究和實際可行之信用評分相差甚遠，也少見改進信用評分的具體建言。而國內各銀行在外部顧問協助下根據本身信貸資料所建置的信用評分系統，則多屬商業機密，其內容外界不得而知，也無法評估其效能。此外台灣唯一的信用通報機構「財團法人金融聯合徵信中心」，雖也嘗試利用本身所儲備來自全台各行庫

的信用資料對多類信貸產品建置通用信用評分,然其性質和銀行信用評分系統類似,內容並不全然對外公佈。對信用評分這樣一個有重大實用和政策價值的應用計量經濟課題,我們希望能夠激發更多的學術研究,也希望在實務上銀行業者以及政府主管機關能夠參考本文所提多項驗證指標,檢視現行信用評分系統能否達到我們所認同的品質水準。

本文將根據台灣新報資料庫所載公開發行企業(不只是中小企業)的財務報表以及違約與否的六年資料,經過一系列資料整理、變量轉換和篩選的步驟,得出三組 8 到 9 個解釋變量組合,以違約與否虛擬變量為應變量,進行 Logit 迴歸模型的估計,再對迴歸係數估計值進行適當的線性轉換,導出對各解釋變量之各個可能數值的對應評分。表 1 和表 2 所示為本文所導出之三個信用評分卡之一。<sup>2</sup>

表 1 的信用評分卡有四個特點需要進一步說明:(1) 雖然各個財務解釋變量的可能數值都是連續的,但信用評分的計算卻是以分組方式進行,這是信用評分卡的標準操作形式。正因為信用評分卡的這種特殊形式,信用評分的計算得以大為簡化及電腦化,但其建置過程則趨於複雜,尤其是對每一個解釋變量數值之分組數與組界點的決定,均成為信用評分卡建置過程中的難題。本文貢獻之一就是建議一套決定組數以及組界點的步驟。(2) 每一個解釋變量之分組評分的相對大小可顯示該解釋變量之數值對違約可能性的影響,除此之外,我們還可對不同解釋變量的評分進行跨變量的比較,直接評估不同變量對違約可能性的影響。(3) 解釋變量中之「有息負債利率」、「稅後資產報酬率」、和「存貨週轉率」的數值分組,包含缺漏值組,這個特殊分組大為降低了建置及使用信用評分卡時對資料的要求,我們不再要求所有解釋變量都要有完整的樣本值,這有助於維護樣本不致因缺漏值而遭到

---

<sup>2</sup> 在此我們將借之簡述信用評分卡的使用方式:根據表 1 和表 2 我們將可對任一申請信用貸款的企業非常快速的計算出其信用評分、信用等級、以及違約機率,計算步驟如下:我們可就該企業的「短期借款」、「借款依存度」、「每股盈餘」、「有息負債利率」、「稅後資產報酬率」、「長期負債/淨值」、與「存貨週轉率」七個財務變量之財報值,依照表 1 分別給予評分。例如,若該企業的短期借款是 1 億元,則可得到 17.66 分;每股盈餘是 1.5 元,可得 30.18 分;存貨週轉率是 5,可得到 -0.62 分等等。將各變量所得評分加總再加上基準評分 701.09 便可得到該企業的信用評分,我們可再依照表 2 的信用評等表,找出該企業的信用等級,以及對應的違約機率。例如,若一企業的信用評分是 700 分,則該企業便屬於第 5 等級,預期的違約機率是 3.2%。由各級別的違約機率可看出,信用評分越大信用風險就越低,而在 8 個信用等級中,第 1 等級的信用評分最高信用風險最低,而第 8 等級的信用評分最低信用風險最高。

表 1 信用評分卡

基準評分: 701.09

分組	組界點	組評分	分組	組界點	組評分
1. 短期借款 $\times 10^{-8}$			5. 稅後資產報酬率 (%)		
1	$x = 0.00$	47.82	1	$-\infty < x \leq 0.00$	-13.68
2	$0.00 < x \leq 0.63$	33.04	2	$0.00 < x \leq 11.00$	3.88
3	$0.63 < x \leq 2.25$	17.66	3	$11.00 < x \leq \infty$	48.71
4	$2.25 < x < \infty$	-29.94	4	缺漏值	29.17
2. 借款依存度 (%)			6. 長期負債 / 淨值		
1	$-\infty < x \leq 19.50$	47.20	1	$-\infty < x \leq 0.00$	23.94
2	$19.50 < x \leq 34.11$	30.48	2	$0.00 < x \leq 78.31$	24.08
3	$34.11 < x \leq 62.56$	13.26	3	$78.31 < x \leq 93.72$	2.04
4	$62.56 < x \leq 174.17$	-15.04	4	$93.72 < x \leq 100.00$	-13.54
5	$174.17 < x < \infty$	-38.72	5	$100.00 < x < \infty$	-40.08
3. 每股盈餘			7. 存貨週轉率 (次)		
1	$-\infty < x \leq -1.05$	-32.08	1	$0.00 < x \leq 0.35$	-54.04
2	$-1.05 < x \leq 0.00$	-11.55	2	$0.35 < x \leq 3.18$	-13.78
3	$0.00 < x \leq 1.14$	5.33	3	$3.18 < x \leq 6.13$	-0.62
4	$1.14 < x \leq 1.61$	30.18	4	$6.13 < x \leq 15.73$	25.05
5	$1.61 < x \leq 2.30$	53.94	5	$15.73 < x < \infty$	32.09
6	$2.30 < x \leq \infty$	63.05	6	缺漏值	35.34
4. 有息負債利率 (%)					
1	$x = 0.00$	69.00			
2	$0.00 < x \leq 5.38$	48.20			
3	$5.38 < x \leq 5.97$	21.71			
4	$5.97 < x \leq 7.59$	-9.05			
5	$7.59 < x \leq 8.14$	-24.13			
6	$8.14 < x < \infty$	-12.82			
7	缺漏值	104.64			

刪除。(4) 各解釋變量的分組評分和變量數值之間除了單調遞增或遞減的關係外,「有息負債利率」的分組評分呈現 U 字形分布。解釋變量數值和評分之間的這種非線性關係,完全是根據 Logit 模型的迴歸係數估計結果所推導出來的,而之所以能夠由線性迴歸係數推導出非線性關係的關鍵就是在解釋變量數值的分組上。我們要明確指出,信用評分卡的這四個特點都和解釋變量數值的分組結構有關,本文的一個重點就是要強調信用評分卡的分組計分特質,並詳細解釋如何對解釋變量數值進行分組。我們尤其要強調,Logit 迴歸模型的估計結果雖然是信用評分卡的核心,但信用評分卡的建置卻不是一個標準的計量經濟學應用,這是因為信用評分卡的目的不在於估計結果的分

表 2 信用評等

評等	評分區間	違約率 (%)
1	941 – 1,000	0.0
2	846 – 940	0.2
3	797 – 845	0.6
4	725 – 796	1.5
5	696 – 724	3.2
6	627 – 695	7.1
7	519 – 626	23.0
8	0 – 518	62.9

析而在於對新樣本的預測, 這使得建置信用評分卡的分析重點迥異於典型的計量經濟研究, Logit 模型的估計 (或是資料探勘技術的使用) 只是建置信用評分卡的過程中一小步, 不論是在估計 Logit 模型之前還是之後, 我們還必須進行諸多深入的操作和分析, 這些操作和分析才是建置信用評分卡的主題, 均屬建置信用評分卡的必要步驟, 每一步驟亦皆有其理論上的必要性。

我們要將信用評分再進一步加以粗分等 (如表 2) 的原因, 除了因為粗分等具有簡化的作用而能讓使用者更易於了解外,<sup>3</sup> 我們還可因此對每一信用評等的等級指定單一違約機率, 並據此將所有企業的信用風險予以分等。根據新巴賽爾資本協定, 分等違約機率是銀行評估信用風險、計算計提資本、乃至控管總體風險的重要參數。也就是說, 對於新巴賽爾資本協定的執行而言, 表 2 的內容遠較表 1 更為重要。然而我們也要強調, 每一信用評等等級都跨越一整個區段的信用評分, 但對應的違約機率只有一個, 這表示信用評分不同的企業會有相同或是非常接近的違約可能性, 實際狀況是否如此是需要仔細檢驗的。

本文共分五節, 在下一節中我們將簡單回顧多種評估信用評分系統的統計方法, 我們只有在知道如何評估信用評分系統之後, 才得以了解信用評分卡建置過程中諸多步驟的意涵。在該節所討論統稱為「驗證統計量」的 11 個來自多個不同領域的統計量可分為四大類: 「辨識度」、「序別相關」、「信息理論」、與「違約率配適度」, 我們將以統一的符號解釋各個統計量在功能上的異同與關連。對信用評分系統的架構乃至評估方式有了認識之

<sup>3</sup> 由表 1 所計算出的信用總評分大致是在 350 分和 950 分之間, 信用評分雖然很精確, 但大多數人可能更習慣於使用類似債券評等之 1 到 8 等 (或其他分級數) 的信用評等, 對評等等級所代表的意義比較有直覺, 對等級的升降也比較敏感, 過於細緻的信用評分反而不易解讀。



後，在第三節中我們才能深入說明如何根據台灣新報資料庫所載數千家公開發行企業的三年財務報表資料，建置信用評分卡的步驟，說明的重點將是資料的整建手續、變量的轉換與篩選方法、以及係數估計值的轉換公式等議題，在這一節中我們將根據三種不同的原則建置三組不同的信用評分卡。在第四節中我們將對第三節所擬議的三組信用評分卡進行測試與驗證，我們將分別就建模樣本和驗證樣本計算第二節所提出的 11 種驗證統計量，並據之評估各信用評分卡的績效，我們也將更進一步和台灣經濟新報所發佈的基本評等指標進行比較，嘗試證明以客觀系統方式建置的信用評分卡能夠較以經驗主觀方式建置的評等指標有更好的績效。在第五節中我們將總結本文所獲得多項實證結果。

## 2. 評估信用評分卡的統計方法

假設  $d$  代表違約與否的虛擬變量，亦即  $d = 1$  表示違約，而  $d = 0$  表示未違約。建置信用評分系統的目的是預測  $d$ 。信用評分系統的主體是信用評分的評分值，我們將以變量  $z$  來代表這個評分值，使用信用評分  $z$  預測  $d$  的方式通常是：先決定一截斷點 (cutoff) 數值  $\bar{z}$ ，只要  $z$  的數值小於或等於  $\bar{z}$ ，則我們便預測違約的發生。

根據信用評分  $z$  與對應的截斷點  $\bar{z}$  以預測違約與否  $d$  將可得到如表 3 所示的四種結果：

表 3 預測結果

	$d = 0$ (未違約)	$d = 1$ (違約)
$z > \bar{z}$ (預測未違約)	正確的預測未違約	錯誤的預測未違約 (型二誤差)
$z \leq \bar{z}$ (預測違約)	錯誤的預測違約 (型一誤差)	正確的預測違約

對表 3 中所包含的兩種錯誤，我們可採用統計學中假設檢定的術語，將左下角的「錯誤的預測違約」稱為「型一誤差」，而右上角的「錯誤的預測未違約」則稱為「型二誤差」。表 3 的功能除了列舉信用評分的預測結果，也有助於我們進一步評估信用評分的優劣。直覺上來說，信用評分的優劣將視型

一誤差和型二誤差之「可能性」的大小而定：顯然的，型一誤差和型二誤差的可能性越小信用評分就越好。

在接下來的四個小節中我們將根據巴賽爾銀行監理委員會 (BCBS, 2005) 第 14 號文件的建議，有系統的說明多種評估信用評分系統績效的統計量。相關文獻尚有 Sobehart et al. (2000)，比較完整的說明可參見鍾經燮 (2005)。

## 2.1 辨識度

要有系統的分析信用評分系統的優劣，我們必須假設違約與否虛擬變量  $d$  與信用評分  $z$  都是隨機變量，它們所有的性質都是由一個聯合分配所決定。由於  $d$  是虛擬變量，所以必然就是 Bernoulli 隨機變量，其分配可以違約可能性  $P(d = 1)$  代表（我們將以  $p$  來表示這個可能性）。至於信用評分  $z$  的分配，不論  $z$  是連續型隨機變量還是間斷型隨機變量，都和 Bernoulli 隨機變量  $d$  屬於極不相同的隨機變量，因此  $z$  和  $d$  的聯合分配不易表示，比較常用的做法是採用信用評分  $z$  的兩個條件累積分配函數  $F_0(\bar{z}) = P(z \leq \bar{z} | d = 0)$  和  $F_1(\bar{z}) = P(z \leq \bar{z} | d = 1)$  來代表  $z$  和  $d$  的聯合分配。有了這兩個條件分配函數，我們就很容易的定義如下的兩個重要比率：就任一給定的截斷點  $\bar{z}$ ，我們將稱呼  $F_0(\bar{z})$  為「誤查率」(false alarm rate)， $F_1(\bar{z})$  為「實查率」(hit rate)，因為前者表示在給定未違約的狀況下，錯誤預測違約的可能性，而後者表示在給定違約的狀況下，正確預測違約的可能性。實查率又稱為「敏感度」(sensitivity)，而「1- 誤查率」則稱為「特定度」(specificity)。我們要強調，根據累積分配函數的單調遞增性質可知，實查率和誤查率都是截斷點  $\bar{z}$  的單調遞增函數，也就是說，實查率和誤查率會隨著截斷點  $\bar{z}$  數值的增加而增加。實查率和誤查率所代表的便是信用評分系統的「辨識度」(discrimination)，若信用評分系統的實查率高而誤查率低，則我們就說這個評分系統的辨識度很高，而辨識度越高的評分系統就越優異。

只要信用評分系統是合理的，在每一個截斷點上的實查率都應該要比對應的誤查率為大。我們首先要指出，信用評分系統辨識度的高下就看實查率  $F_1$  要比誤查率  $F_0$  大多少而定，文獻中常以「ROC 線」(receiver-operating-characteristic) 來呈現實查率與誤查率之間的差異，<sup>4</sup> 所謂的 ROC 線便是就任一截

<sup>4</sup> Swets (1988) 對 ROC 線分析的各種應用做了相當有趣的介紹，亦請參見 Hanley and Mc-

斷點  $\bar{z}$ ，將數值較小的誤查率  $F_0(\bar{z})$  訂為橫座標值（即  $x$  座標值），並將數值較大的實查率  $F_1(\bar{z})$  訂為縱座標值（即  $y$  座標值）所導出的曲線（請見圖 1）。<sup>5</sup>

ROC 線雖然能夠很直覺的顯示信用評分系統的整體表現，但圖形終究不如數字簡明方便，所以有必要推導出一些能夠顯示 ROC 線特質的統計量，其中最重要的無疑是「AUC 值」（area under the curve 值），乃 ROC 線和橫座標軸之間的面積。AUC 值越大信用評分系統就越好，其最大值是 1。當 ROC 線和 45 度線重合時，AUC 值等於 0.5 時，對應的評分系統無辨識度可言。

除了 AUC 值外，我們還可由 ROC 線導出一個較次要的統計量：KS 統計量，乃 ROC 線和 45 度線之間的最大垂直距離。

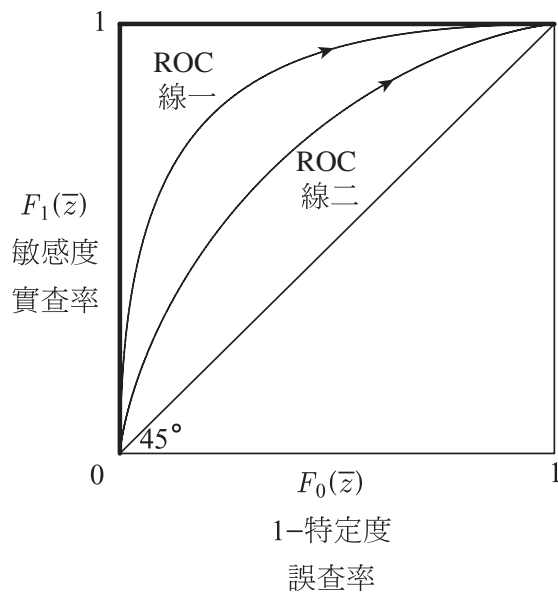


圖 1 ROC 線

Neil (1982)。對 ROC 線分析的理論和應用做出最主要的貢獻多為生物醫學統計學者，請見 Pepe (2003)。

<sup>5</sup> ROC 線上各點所代表的意義是：要達到任一給定的實查率（高度）所必須接受的誤查率（寬度）。隨著截斷點的遞增，實查率（高度）和對應的誤查率（寬度）均會遞增，亦即會如同圖 1 中 ROC 線上的箭頭所示由左下方往右上方移動，也就是說，隨著實查率（高度）要求的提高，誤查率（寬度）成本也必然增加，而優劣評分系統之間的差別完全是在於誤查率（寬度）成本增加速度的快慢。完美信用評分系統的 ROC 線將和單位正方形值左邊和上邊重合，如圖 1 中的粗黑線所示。而最低辨識度之 ROC 線則將和 45 度線重合。合理的 ROC 線應該在 45 度線之上，ROC 線和 45 度線之垂直距離越大，對應信用評分系統就越好（在圖 1 中對應 ROC 線一的信用評分系統要比對應 ROC 線二的信用評分系統為好）。

## 2.2 信用評分與違約與否的序別相關

根據實查率  $F_1(z)$  和誤查率  $F_0(z)$  的對比所建立的辨識度, 可說是評估信用評分系統優劣的最主要指標, 但採用信用評分  $z$  在給定違約  $d$  下之條件分配函數  $F_1(z)$  和  $F_0(z)$  來分析  $z$  和  $d$  之間的關係, 在概念上卻不是非常直覺易懂, 在這一小節中我們將考慮  $z$  和  $d$  的相關性, 並直接根據這個相關性的大小來判斷的信用評分系統優劣。很顯然的, 高信用評分  $z$  就表示  $d = 0$  (不違約) 的可能性高, 而低信用評分  $z$  則表示  $d = 1$  (違約) 的可能性高, 因此, 合理的信用評分  $z$  和違約與否虛擬變量  $d$  之間應該存在一個負向關係, 所以直覺上我們可就  $z$  和  $d$  之間負向關係的強弱來評估信用評分  $z$  的優劣:  $z$  和  $d$  的負向關係越強, 信用評分  $z$  就越好。但在這裡我們也要特別強調, 因為信用評分  $z$  和違約與否虛擬變量  $d$  的絕對數值無意義只有相對大小才有意義, 所以我們不應該使用普通的相關係數(共變數除以二變異數乘積之開根號)來衡量  $z$  和  $d$  之間的負向關係, 而必須採用兩個隨機變量排序的相關係數, 這些統稱為「序別相關」的統計量將不受信用評分  $z$  之任何單調遞增轉換的影響。

最簡單的序別相關統計量是 Spearman 的「序號相關係數」(rank correlation coefficient)。此外, 還有兩個更符合序別相關概念的「Kendall's  $\tau_a$ 」和「Kendall's  $\tau_b$ 」統計量。若將 Kendall's  $\tau_a$  統計量視為信用評分  $z$  和違約與否虛擬變量  $d$  之間的共變數, 則對應的相關係數便是所謂的 Kendall's  $\tau_b$  統計量。

## 2.3 信息理論

我們可借用信息理論(information theory)中的一些概念, 以做為評估信用評分系統的另一類方法, 在本小節中我們將先簡單的回顧信息理論, 較完整的討論可參見 Cover and Thomas (1991)。

信息理論簡單來說就是一套分析隨機變量所呈現之不確定性的理論。假設  $f(x)$  是隨機變量  $x$  的邊際密度函數, 而  $f(y|x)$  是給定隨機變量  $x$  之條件下隨機變量  $y$  的條件密度函數, 則我們可定義隨機變量  $x$  的「熵」(entropy)

如下:  $H(x) \equiv - \int_{-\infty}^{\infty} f(x) \ln f(x) dx$ , 以測度  $x$  的不確定性; 我們還可定義給定隨機變量  $x$  之條件下隨機變量  $y$  的「條件熵」如下:  $H(y|x) \equiv - \int_{-\infty}^{\infty} [ \int_{-\infty}^{\infty} f(y|x) \cdot \ln f(y|x) dy ] f(x) dx$ , 以進一步測度給定隨機變量  $x$  的信息後, 隨機變量  $y$  所還有的不確定性。現再考慮任一密度函數  $g$ , 我們可定義「信息值」(information value, IV 值)如下:

$$IV(f, g) = \int_{-\infty}^{\infty} [f(x) - g(x)] \cdot \ln \frac{f(x)}{g(x)} dx,$$

以測度兩個密度函數  $f$  和  $g$  之間的「距離」。只要  $f$  和  $g$  不相等, 信息值就一定是正的, 且信息值越大, 則兩個密度函數  $f$  和  $g$  之間的距離就越大。

假設  $f$  和  $g$  是有相同變異數  $\sigma^2$  的兩個常態分配密度函數, 則其差別只在於期望值, 可分別以  $\mu_1$  和  $\mu_2$  表示, 經過簡單運算可知  $f$  和  $g$  之間的信息值是  $(\mu_1 - \mu_2)^2 / \sigma^2$ 。根據這個結果, 我們定義任兩個密度函數  $f$  和  $g$  之間的「離散統計量」(divergence statistic) 為

$$DS = \frac{(\mu_1 - \mu_2)^2}{\left(\frac{\sigma_1^2 + \sigma_2^2}{2}\right)}, \quad (1)$$

其中  $\mu_1$  和  $\sigma_1^2$  是密度函數  $f$  的期望值和變異數, 而  $\mu_2$  和  $\sigma_2^2$  則是密度函數  $g$  的期望值和變異數。

給定兩個隨機變量  $x$  和  $y$ , 我們可證明  $H(x) \geq H(x|y)$ , 這意味著只要  $y$  和  $x$  不獨立,  $y$  的信息便有助於降低  $x$  的不確定性, 所以我們可進一步研究隨機變量  $y$  之信息所能降低隨機變量  $x$  之不確定性的比例:

$$CIER(x|y) = \frac{I(x; y)}{H(x)} = 1 - \frac{H(x|y)}{H(x)}.$$

這個比例稱為「條件熵比率」(conditional information entropy ratio, CIER)。

接下來我們將針對間斷型信用評分  $z$  定義對應的信息值。假設間斷型信用評分  $z$  有  $K$  個可能數值  $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_K$ , 則其條件密度函數可以如下方式表示:  $f_0(\bar{z}_k) \equiv P(z = \bar{z}_k | d = 0) = \pi_{0k}$  和  $f_1(\bar{z}_k) \equiv P(z = \bar{z}_k | d = 1) = \pi_{1k}$ , 這

裡的  $\pi_{0k}$  和  $\pi_{1k}$  均為機率值, 且  $\sum_{k=1}^K \pi_{0k} = \sum_{k=1}^K \pi_{1k} = 1$ 。信用評分  $z$  的信息所測度的就是  $f_0$  和  $f_1$  之間的距離:

$$IV(f_0, f_1) = \sum_{k=1}^K (\pi_{0k} - \pi_{1k}) \cdot \ln \left( \frac{\pi_{0k}}{\pi_{1k}} \right). \quad (2)$$

信用評分的辨識度(以 AUC 值為代表)可以說是對其條件累積分配函數之間距離的衡量, 而信息值則是對其條件密度函數之間距離的衡量, 兩者概念上雖然相似, 但數值大小卻可能非常不同, 因此就衡量信用評分優劣的功能而言, 信息值是一個獨立於辨識度的不同測度。

間斷型信用評分  $z$  相對於違約與否虛擬變量  $d$  的條件熵比率也可類似的方法加以定義, 這裡便不再複述, 我們只強調, 條件熵比率顯示的是信用評分  $z$  之信息所能降低違約與否虛擬變量  $d$  之不確定性的比例, 其優點是其數值介於 0 和 1 之間, 較易相互比較, 但缺點是會受到違約率  $p$  影響。

信息值和條件熵比率雖然都源自於信息理論, 但兩者的意義卻相當不同。條件熵比率顯然比較接近迴歸模型的  $R^2$  概念, 著重的是信用評分對違約與否虛擬變量的解釋能力, 而信息值的概念比較簡單, 基本上只是在衡量兩個密度函數之間的距離。

這裡我們要針對信息值定義(2)中的對數值項  $\ln(\pi_{0k}/\pi_{1k})$  做更進一步的分析, 這個數值稱為「證據權重」(weight of evidence, WOE), 它是分組違約率  $p_k$  和總違約率  $p$  的函數:

$$WOE \equiv \ln \left( \frac{\pi_{0k}}{\pi_{1k}} \right) = \ln \left[ \frac{\frac{p}{(1-p)}}{\frac{p_k}{(1-p_k)}} \right]. \quad (3)$$

由此式可見, 證據權重是分組違約率  $p_k$  的單調遞減轉換, 當總違約率  $p$  小於、等於、或大於分組違約率  $p_k$  時, 則證據權重便是負的、零、或正的。當分組違約率  $p_k$  趨近其極端值 0 或 1 時, 則證據權重會快速的趨近  $\infty$  或  $-\infty$ 。

## 2.4 違約率的配適度

假設一個信用評分  $z$  能夠產生對違約率  $p$  的預測值  $q$  (數值介於 0 和 1 之間的機率值), 則我們便可根據均方差以及相關的統計量的概念來衡量違約率的「配適度」(fitness), 這將是另一種衡量信用評分系統優劣的方法。此外, 由於間斷型信用評分的作用之一是將樣本分組, 且對同一個分組內所有的樣本均指定以相同的違約率預測值, 對於這種做法顯然有必要檢測其正確性, 此即所謂的「校準度」(calibration)分析, 以查驗同一分組內所有樣本的違約率預測值的確相同, 且等於所指定的數值。<sup>6</sup>

最直觀的配適度統計量是「Brier 分數」, 以均方差形式直接顯示違約與否虛擬變量和違約率預測值之間的配適度。假設  $(d_i, q_i)$ ,  $i = 1, 2, \dots, n$ , 是違約與否虛擬變量  $d$  和對應違約率  $p$  之預測值  $q$  的樣本, 則「Brier 分數」(Brier score)的定義是

$$b = \frac{1}{n} \sum_{i=1}^n (d_i - q_i)^2,$$

其數值介於 0 和 1 之間, 越低越好。但我們也必須指出 Brier 分數的缺點是其絕對大小的意義難以評估, 這是因為 Brier 分數會受樣本違約率  $\hat{p}$  (亦即  $d_i$  的樣本平均數) 大小的影響: 假設以固定數值  $\hat{p}$  做為所有個別違約率的預測值:  $q_i = \hat{p}$ , 則 Brier 分數便縮減為  $\hat{p}(1 - \hat{p})$ , 這個數值於  $\hat{p} = 0.5$  時達到極大值 0.25, 而當  $\hat{p}$  接近 0 或 1 時會趨近於 0。也就是說, Brier 分數很受樣本違約率  $\hat{p}$  大小的影響: 當  $\hat{p}$  很接近 0 或 1 時, Brier 分數會特別小而失去比較意義。

若能將樣本分組, 我們還可計算「分組 Brier 分數」以進一步的衡量信用評分系統的分組配適度。假設樣本可分  $K$  組, 各分組樣本數分別為  $n_k$ ,  $k = 1, 2, \dots, K$ , 為簡化符號, 我們將  $d_i$  和  $q_i$  樣本的下標依照該樣本所屬分組分別改寫為  $d_{ki}$  和  $q_{ki}$ ,  $i = 1, 2, \dots, n_k$ ,  $k = 1, 2, \dots, K$ , 並以  $\bar{q}_k = n_k^{-1} \sum_{i=1}^{n_k} q_{ki}$  表示違約率預測值的分組平均, 則「分組 Brier 分數」的定義是  $b^* = (1/n) \sum_{k=1}^K \sum_{i=1}^{n_k} (d_{ki} - \bar{q}_k)^2$ 。給定各分組的樣本違約率  $\hat{p}_k = n_k^{-1} \sum_{i=1}^{n_k} d_{ki}$ , 則分組 Brier

<sup>6</sup> 有些文獻稱前者為同質性分析, 後者才是校準度分析, 本文將之統稱為校準度分析。

分數可分解為  $b^* = (1/n) \sum_{k=1}^K n_k \hat{p}_k (1 - \hat{p}_k) + (1/n) \sum_{k=1}^K n_k (\bar{q}_k - \hat{p}_k)^2$ , 其中第一項只是  $d_{ki}$  的組內變異數, 和違約率預測值的配適度無關。由第二項可知, 分組 Brier 分數顯示的其實是違約率預測值分組平均  $\bar{q}_k$  的配適度。我們尚可以將這第二項改寫成所謂的 Hosmer-Lemeshow 卡方配適統計量 (Hosmer-Lemeshow  $\chi^2$  goodness-of-fit) 或簡稱 HL 卡方配適統計量 (請見 Hosmer and Lemeshow, 2000):

$$HL \equiv \sum_{k=1}^K \frac{n_k (\bar{q}_k - \hat{p}_k)^2}{\bar{q}_k (1 - \bar{q}_k)}. \quad (4)$$

若信用評分系統配適良好, 則 HL 卡方配適統計量會呈卡方分配, 其自由度需視樣本而定: 若所使用的樣本就是製作信用評分系統的樣本, 則自由度是  $K - 2$ ; 但若所使用的樣本和製作評分系統的樣本不是同一個樣本, 則自由度增加為  $K$ 。我們可利用 HL 卡方配適統計量以假設檢定的方式查驗分組後的配適度是否合理。

#### 2.4.1 校準度檢定

一個間斷型信用評分系統對於屬於同一分組中的所有樣本, 通常都會指定某一特定的違約率。對於同一分組中之樣本是否具有給定之違約率的問題, 我們可以進行假設檢定的手續以判斷其真偽, 此即所謂的校準度檢定。

為推導出校準度的假設檢定統計量, 我們必須對屬於同一分組之違約與否虛擬變量  $d$  做出兩組基本假設: (1) 假設  $d$  的數值是根據某一隨機指標  $\delta$  是否跨越某一門檻值  $c$  而定, 若屬於第  $k$  分組的樣本  $d_i$  都具有相同的違約率  $p_k$ , 則此  $p_k$  必然等於  $P(\delta_i \leq c_i)$ 。現若假設隨機指標  $\delta_i$  具標準常態分配, 並以  $\Phi$  表示標準常態的累積分配函數, 則  $p_k = \Phi(c_i)$ 。因此門檻值  $c_i$  便可以  $\Phi^{-1}(p_k)$  表示。(2) 當違約與否虛擬變量  $d$  的某些樣本被歸為同一分組且具有相同的違約機率時, 則這些樣本必然都受到一些共同因素的影響, 因此  $d$  的這些樣本便不太可能是彼此獨立。我們可假設這些樣本 (以  $d_i$  代表) 之間的相關性是肇因於對應隨機指標  $\delta_i$  具有如下的「單因子結構」:

$$\delta_i = \sqrt{\rho} x + (\sqrt{1 - \rho}) \varepsilon_i,$$



其中  $x$  乃是所有指標  $\delta_1, \delta_2, \dots, \delta_{n_k}$  所共有的「系統風險因子」, 而  $\varepsilon_i$  則是彼此互相獨立、專屬於指標  $\delta_i$  的「特有風險因子」。我們假設係數  $\rho$  是一個介於 0 和 1 之間的分數, 標示系統風險因子  $x$  和特有風險因子  $\varepsilon_i$  之間的相對重要性。我們假設系統風險因子  $x$  和特有風險因子  $\varepsilon_i$  之間也是互相獨立的, 且為確保指標  $\delta_i$  是一標準常態隨機變量, 我們更進一步假設  $x$  和  $\varepsilon_i$  均是標準常態隨機變量。

根據大數法則我們可進一步證明, 當第  $k$  分組的樣本數  $n_k$  趨近  $\infty$  時, 則分組樣本違約率  $\bar{d} \equiv n_k^{-1} \sum_{i=1}^{n_k} d_i$  會收斂到如下的一個隨機變量:

$$\bar{d} \xrightarrow{\text{a.s.}} E(d_i|x) = P(d_i = 1|x) = \Phi \left[ \frac{\Phi^{-1}(p_k) - \sqrt{\rho}x}{\sqrt{1-\rho}} \right].$$

關於單因子結構模型及上述收斂結果的進一步分析, 可見 Bluhm et al. (2003)。對於任一給定的機率值  $q_k$ , 我們可利用這個結果來檢定  $p_k$  是等於  $q_k$ , 還是大於  $q_k$ 。假設  $\bar{d}$  的樣本值是  $\bar{d}_0$ , 則在虛無假設  $H_0 : p_k = q_k$  成立時, 我們可得

$$P(\bar{d} > \bar{d}_0) \approx \Phi \left[ \frac{\Phi^{-1}(q_k) - (\sqrt{1-\rho})\Phi^{-1}(\bar{d}_0)}{\sqrt{\rho}} \right]. \quad (5)$$

只要這個機率小於給定的檢定容忍度 (size), 我們便拒絕虛無假設  $H_0 : p_k = q_k$ , 並接受對立假設  $H_0 : p_k > q_k$ 。

採用此校準度檢定最重要的前提是, 屬於同一分組的樣本  $d_i$  之間有相關性, 且其相關性可以「單因子結構」表示。若  $d_i$  的樣本間沒有相關性, 且彼此獨立的話, 則可採用二項分配檢定, 假設檢定的步驟可以大為簡化。值得一提的是, 校準度檢定的檢定力 (power) 大多不很高, 不太容易拒絕虛無假設。

## 2.5 小結

我們在本節中總共介紹了如表 4 所列的 11 種驗證統計量: 分屬於辨識度、序別相關、信息理論、和違約率配適度四個類別, 可對信用評分系統的不同

表 4 驗證統計量

辨識度	序別相關	信息理論	違約率配適度
AUC 值	Spearman's $\rho_s$	信息值 *	Brier 分數
KS 值	Kendall's $\tau_a$	離散統計量	HL 配適統計量 *
	Kendall's $\tau_b$	條件熵比例 *	校準度檢定 *

註: \* 只適用於間斷型信用評分的驗證統計量。

層面進行評估。

最後我們要指出,表 4 所列的驗證統計量不僅可用來評估信用評分系統的優劣,還可用來篩選信用評分卡內的解釋變量,也就是說,我們可將任何和違約有關的個別解釋變量視為一個只包含一個解釋變量的信用評分,再利用上述驗證統計量來測度該解釋變量和違約可能性之間的關係,逐一篩選出最能預測違約可能性的一組解釋變量,並據以建置包括這些解釋變量的信用評分卡。

### 3. 信用評分卡的建置

建置企業金融信用評分卡的過程不外乎就是利用企業借款戶過去的財務、產業、經營等各方面的資料,找出與其違約可能性有關的一組解釋變量,估計一個以違約與否虛擬變量作為應變量之迴歸模型(例如 Logit 模型或是 Probit 模型),再根據估計結果推導出各個解釋變量和違約之間的量化關係,以預測企業借款戶在未來某一特定期間內違約的可能性。信用評分卡的製作過程可分為五個步驟:(1)資料的整理;(2)解釋變量的轉換;(3)解釋變量的篩選;(4)製作信用評分卡;(5)測試和驗證信用評分卡。其中前四個步驟將分別在接下來的四個小節進行,最後一步測試和驗證則在第 4 節詳述。

信用評分卡基本上就是預測一個虛擬變量的計量經濟模型,但我們也要明確的指出,信用評分卡的建置卻又和典型的實證經濟研究有很大的不同,雖然迴歸模型是信用評分卡的核心,但估計迴歸模型的目的並非要研究違約的形成機制或因果關係,也不是要衡量解釋變量對違約的邊際影響,而完全是為了要能對違約可能性做出準確的預測。因此,諸如對迴歸係數估計值之

顯著性的檢定、對模型設定的檢討等典型計量經濟分析，通常都不會在建置信用評分卡的過程中出現。相反的，建置信用評分卡的統計分析幾乎全部集中在預測準確度的評估上，因此，在建置信用評分卡時，除了用來製作信用評分卡的樣本（即所謂的建模樣本）外，一定要保留部分樣本（稱為驗證樣本）以驗證信用評分卡的預測績效，驗證可說是建置信用評分卡之過程中絕不可少的環節。

和建置信用評分卡相關的文獻早期以學術界對公司財務危機預警模型的研究為主，Altman（1968）使用多變量統計學中的「區別分析」（discriminant analysis）來研究公司財務變量和財務危機之間的關係，類似的學術研究持續了相當長一段時間，近年來仍陸續有相關論文在學術期刊出現。這類論文通常著重在諸如方法論、公司財務理論等學術性課題，例如陳業寧等（2004），其實證研究和具實用價值之信用評分卡尚有一段距離。真正以建置和驗證信用評分卡為主題的文獻通常都是業界研發人員所著，例如Mays（1998, 2001）、Falkenstein et al.（2000）、和Bailey（2004a, 2004b）。Thomas et al.（2002）和de Servigny and Renault（2004）二書則以一更廣泛的架構探討信用評分系統。

### 3.1 資料整理

根據台灣經濟新報 1998 年至 2003 年個別企業財務報表資料，除了選用以營業收入、資產額、長短期借款額為主的多個財報變量外，再由原財報資料定義出 80 個財務比率，做為信用評分卡解釋變量的候選者。<sup>7</sup> 應變量是違約與否虛擬變量，我們直接採用台灣經濟新報的原有定義。根據台灣經濟新報資料除了可知企業的違約年份，還可推知企業違約時點是 3 月、6 月、9 月、或 12 月。我們所使用的資料涵蓋的違約時點是從 1999 年 12 月到 2004 年 9 月。我們將以 1998 年至 2000 年的前三年資料做為建模樣本，再以 2001 年至 2002 年的後兩年資料做為驗證樣本。對於建模樣本中的未違約正常戶，其財報解釋變量的樣本值將採用建模樣本所涵蓋之三年中最晚近之 2000 年的資

---

<sup>7</sup> 台灣經濟新報資料涵蓋範圍廣泛，不是只包含對銀行有借貸關係的企業，所以個別銀行建置信用評分卡時所常遭遇的「拒絕樣本」問題較小。所謂拒絕樣本問題是指被銀行拒絕貸款申請的企業無法確知其違約與否，以致建置信用評分卡所使用的樣本有截斷（truncation）問題。

料,對於屬於驗證樣本中的正常戶,其財報解釋變量樣本值將採用2002年的資料。

對於違約戶之財報解釋變量樣本值時點的選取,為避免 Ohlson (1980) 所指出之「財報時間點」的問題,我們不選用違約發生同年的財報資料,而改採違約發生之前一至二年之間的解釋變量樣本值,原因是財報常因需要會計師的簽證而造成公佈時間長達半年以上延宕,所以財報資料最快也只能反應企業半年前的財務狀況。表5列舉所有違約戶違約發生時點、以及我們所採用之財務解釋變量樣本值時點,如前所述,解釋變量樣本值時點都選在違約發生之前的一至二年之間。

由表5可知,在2002年12月或之後違約的企業,在建模樣本中是正常戶,但在驗證樣本中則成為違約戶。我們要特別指出驗證樣本的兩項特點:(1)那些在2002年12月之前違約而成為建模樣本之違約戶的樣本,將會被排除在驗證樣本外,也就是說,除了2001年之後才出現的新樣本外,驗證樣本中正常戶和違約戶在建模樣本中均是正常戶。(2)在2004年9月和2005年9月之間違約的企業,對應財務解釋變量的樣本值時點將是2003年。因此若缺2005年的資料,則無法確認2003年的企業樣本哪些是違約戶哪些不是,所以驗證樣本正常戶的樣本時點最遲也只能是2002年,2003年之後的樣本將不會以正常戶的形式出現在我們的驗證樣本內。

本文建置信用評分卡的過程中容許解釋變量包含缺漏值,但我們要求缺漏值的比例必須是在2%到30%之間(對缺漏值的處理詳情請見第3.2小節),我們將刪除缺漏值比例超過30%的解釋變量,對於缺漏值比例小於2%的解釋變量,我們則只刪除缺漏值樣本但保留該解釋變量。

由台灣經濟新報可獲得六年21,031筆樣本,刪除重複錯誤樣本後,再經由上述的資料處理方式,得到的樣本數如表6所示。

我們要特別指出建模樣本的違約率遠大於驗證樣本中的違約率,主要原因是台灣地區的企業違約情況在2001年後有所改善,且建模樣本的違約率乃三年累積比率,而驗證樣本則是兩年累積比率。驗證樣本和建模樣本另一個不同之處在於,驗證樣本的違約戶絕大數都是在2002年之前三年均維持不違約的企業,也就是說,驗證樣本的違約戶是一群由好變壞的企業。這群企業和建模樣本中的違約戶相比,除了數目少了很多之外,在違約原因上

表 5 違約戶財報資料的擷取時點

	違約時點	財務變量 樣本時點	違約戶數
建模樣本	1999 年 12 月、2000 年 3 月、2000 年 6 月、2000 年 9 月	1998 年	67
	2000 年 12 月、2001 年 3 月、2001 年 6 月、2001 年 9 月	1999 年	120
	2001 年 12 月、2002 年 3 月、2002 年 6 月、2002 年 9 月	2000 年	105
驗證樣本	2002 年 12 月、2003 年 3 月、2003 年 6 月、2003 年 9 月	2001 年	27
	2003 年 12 月、2004 年 3 月、2004 年 6 月、2004 年 9 月	2002 年	39

表 6 樣本結構

	總戶數	正常戶數	違約戶數	違約率 (%)
建模樣本	2,519	2,290	229	9.09
驗證樣本	2,223	2,157	66	2.97

也應有很大的差異。我們由此推測，由建模樣本所製作出的信用評分卡在驗證樣本中的表現可能不會太好，對於這個問題我們將在第 4 節有較詳盡的討論。

### 3.2 解釋變量的轉換

為了得到如同表 1 形式的信用評分卡，特別是希望在對各個財務解釋變量計算信用評分時，可以方便的分組計分方式直接評分，我們在執行迴歸模型的估計之前，必須將所有的解釋變量（不論這些變量的數值是連續形式的還是間斷形式的）都加以「序別化」，使之成為序別變量 (ordinal variable)。所謂的序別變量是指變量值只有有限個（通常是 2 到 8 個）、且這有限個變量值之間可比較大小的變量。解釋變量經過序別化後的變量值不僅可比較大小，還具違約機率的意義。這裡我們就詳細說明對一個連續的解釋變量  $x$  進行序別化的步驟：

步驟一：給定一個正整數  $m$ ，找出  $m - 1$  個由小到大的組界點  $c_1, c_2, \dots, c_{m-1}$ ，將連續變量  $x$  的可能數值分割為  $m$  組。

步驟二：根據  $x$  變量樣本值的分組，將樣本分組，並計算各分組的違約比例： $p_1, p_2, \dots, p_m$ 。

步驟三: 創造序別變量  $x^*$  如下: 當  $x$  變量的樣本值落在  $(c_{k-1}, c_k)$  分組中時, 則將序別變量  $x^*$  的對應樣本值設定為  $p_k$ , 所得到  $x^*$  變量便將是一個只有  $m$  個不同樣本值的序別變量。

關於分組違約比例  $p_k$ , 我們有如下的觀察: 若連續變量  $x$  和違約可能性之間呈正相關, 則分組違約比例傾向於單調遞增:  $p_1 \leq p_2 \leq \dots \leq p_m$ 。若連續變量  $x$  和違約可能性之間呈負相關, 則分組違約比例傾向於單調遞減:  $p_1 \geq p_2 \geq \dots \geq p_m$ 。若連續變量  $x$  和違約可能性之間的線性關係並不明顯, 亦即分組違約比例雖大致呈單調趨勢, 但也有不少例外時, 我們尚可採用 LOWESS (locally weighted scatterplot smoothing) 方法進行平滑化修正, 以減少離群值的影響, 強化分組違約比例的單調趨勢, 凸顯序別變量和違約可能性之間的平滑關係(或是沒有關係)。<sup>8</sup>

在定義序別變量  $x^*$  的步驟三中, 我們尚可將分組違約比例  $p_k$  依照 (3) 式轉換為分組「證據權重」,<sup>9</sup> 甚或是任何其他的單調轉換, 但不論是採取何種定義, 序別變量  $x^*$  之  $m$  個可能變量值除可相互比較外, 也可直接顯示出和違約可能性的關係。

### 3.2.1 分組數和組界點的決定

對解釋變量進行序別化之前, 必須先決定分組數  $m$  與個組界點  $c_1, c_2, \dots, c_{m-1}$ , 本文採用反覆試誤方式搜尋能使 AUC 值極大化的分組數和組界點。

我們要指出, 就如同任何「資料探勘」方法一樣, 利用電腦以反覆試誤方式搜尋分組數和組界點極可能產生「過度配適」或「過度學習」的問題: 即所獲得的分組數和組界點太受建置資料中特有資料特性的影響, 以致使用驗證資料進行驗證時, 會產生遠不如建置資料績效的預測結果。為了降低發生「過度配適」問題的可能性, 我們還對照樣本數的大小對分組數做了限制: 僅在二至十組之間搜尋, 要求所有的分組至少有一個違約戶樣本。此外, 我們更採用自體抽樣的方式求取組界點的信賴區間, 要求各組界點的信賴區

<sup>8</sup> 減少離群值的做法可能也有其缺點: 違約可視為異常事件, 而異常事件無疑會和某些財務變量之產生離群樣本值有密切關係, 任何降低離群值影響的做法都等同於拋棄某些重大信息。但我們的態度是採取一個較保守的態度, 我們認為引入不正確的離群值因而導致信用評分卡偏誤的後果, 要比未能充分利用信息為嚴重。

<sup>9</sup> 分組證據權重是分組違約比例  $p_k$  的單調遞減轉換, 當總違約比例  $p$  小於 (大於) 分組違約比例  $p_k$  時, 分組證據權重便是負的 (正的)。

間儘可能彼此不重疊。

分組數和組界點非常重要，不適當的分組數和組界點不僅會扭曲序別化解釋變量和違約可能性之間的關係，還會對解釋變量的篩選造成非常不利的影響，進而成為建置信用評分卡失敗的主因。這是因為分組數和組界點的決定必須和解釋變量的篩選聯結起來做整體的考量，也就是說，在篩選解釋變量的過程中，我們不僅要決定是否要保留某一個解釋變量，也要同時決定其分組數和組界點。

### 3.2.2 轉換解釋變量的目的

序別化解釋變量除了可簡化信用評分卡的使用外，還有六項優點：(1) 避免過度配適：序別化解釋變量本身是一種將變量值模糊化的機制，所以有助於避免過度配適及捕捉應變量中主要的趨勢。(2) 降低離群值的影響：<sup>10</sup> 不管原變量  $x$  的樣本值如何離散，序別化後的變量  $x^*$  之可能數值只有  $m$  個，絕不會有過大或過小的樣本值。若序別變量值  $x_k^*$  是分組異常比例  $p_k$ ，則其數值都應該在 0 和 1 之間，若序別變量值  $x_k^*$  是證據權重，則其數值都應該都在  $\pm 2$  之間。過小的離群值將全轉變成  $x_1^*$  值，過大的離群值則將都轉變成  $x_m^*$  值，降低離群值的影響，有助於穩定迴歸模型的估計。(3) 妥善利用缺漏值：由於缺漏值可自行歸為一類而得以保存，還可自然的融入序別變量的定義中。(4) 可呈現解釋變量和違約可能性之間的非線性關係，尤其是 U 字形與倒 U 字形的二次式關係，對於此點我們將在第 3.4.3 小節有更多的解釋。(5) 強化解釋變量評分的意義：序別化解釋變量的分組評分除了可跨組比較外，也容許變量之間的直接比較，使用者可因此了解各解釋變量的貢獻及其相對重要性。(6) 有助於解釋變量的篩選：若序別化解釋變量值是分組證據權重，則其和違約可能性之間必然是負向關係；但若序別化解釋變量值是分組違約比例，則其和違約可能性之間必然是正向關係，估計迴歸模型所得到的迴歸係數估計值之正負符號自然也應該滿足這個單調關係。若迴歸係數估計值的符號有所不符，就表示對應的序別化解釋變量和違約可能性之間有異常現象，例如兩個或多個序別解釋變量之間發生了線性重合 (multicollinearity) 問題，此時

---

<sup>10</sup> 雖然很多借款戶的財務變量在發生違約前會有離群值的現象，但離群值不見得就一定是違約的前兆。我們很希望能夠更有效的利用離群值，但卻也無法忽視直接納入離群值對估計過程的可能負面影響，在找到更好的方法以萃取離群值的特有信息前，我們選擇以分組的方式降低離群值的衝擊。

便須考慮刪除這個異常的解釋變量。換言之, 刪除迴歸係數估計值之符號出問題的序別化解釋變量將是篩選解釋變量的一個步驟。

### 3.3 解釋變量的篩選

我們將依循兩套不同的方法來篩選序別化解釋變量, 其間的差別在於是否考慮財務比率的分類結構。財務報表分析的教科書多將財務比率變量分為短期償債能力、長期償債能力(財務結構)、經營績效、獲利能力、與現金流量五大類型, 篩選財務比率變量時可在每一類型中各挑出一些最能預測違約可能性的變量, 以獲得一平衡的財務比率的組合。我們當然也可忽略財務比率類型的平衡, 完全依賴第2節所介紹的各種驗證統計量來選取最能預測違約可能性的解釋變量組合。

不論是否考慮財務比率類型, 我們都將以 Logit 迴歸模型為基礎, 並以「正向選入法」做為篩選解釋變量的基本架構, 也就是說, 我們將在所有候選的序別化解釋變量中, 逐一將表現最佳的新變量選入組合, 而每次挑選新變量時, 不會將已選入組合的變量再排除。在這種篩選變量的過程中, 最關鍵之處就是選入新變量的標準。<sup>11</sup> 我們對於選擇組合的第一個變量與其後逐次選入的新變量設定兩套不同的標準: 挑選起始變量的主要依據是 AUC 值, 但也同時參考 Spearman 序別相關係數、信息值、與條件熵比例。<sup>12</sup> 由於在所有的序別化解釋變量中「借款依存度」的四個驗證統計量正好都是最高的, 所以借款依存度是起始變量的第一選擇。我們亦嘗試其他的起始變量, 但無法得到表現更好的組合。

從第二個序別化解釋變量開始, 挑選每一輪新變量的手續都是相當費時的反覆試誤過程, 若不考慮財務比率類型, 則對尚未選入的序別化解釋變量

---

<sup>11</sup> 「正向選入法」似乎要比「反向淘汰法」(即將所有的候選解釋變量全部納入後, 再逐一淘汰表現不佳者)簡單。

<sup>12</sup> AUC 值代表辨識度, Spearman 係數代表序別相關, 而信息值與條件熵比例則代表信息理論值。由於現階段尚未建立對各企業之違約率的估計值, 所以無法採用屬於違約率配適度的驗證統計量。我們之所以選用了超過一個由信息理論導出的驗證統計量是因為信息值有一個結構上的缺點, 其數值很受分組數及組界點選擇的影響。信息值數值會隨著分組數的增加而增加, 當分組數過多以致一些分組違約率過於接近零時, 信息值的數值就會趨向無限大, 換言之, 信息值很可能產生一些並無實質意義的過大數值, 因此在使用信息值之前, 一定要確定它們的數值是建立在一個合理的分組數上。



逐一考量,只保留能夠滿足下列四項條件的變量:(1) 該序別化解釋變量在建模樣本中的 AUC 值應和在驗證樣本中的 AUC 值相當接近,其間的差異不得超過 0.1;(2) 所有序別化解釋變量在 Logit 迴歸模型中的迴歸係數估計值必須是負的,且對應的  $t$  統計量必須大於 0.5,而貢獻度則必須大於 3%;(3) 各序別化解釋變量之各組評分(尤其是缺漏值組)相對大小的區隔必須明顯而合理;(4) 根據信用評分卡對建模樣本中的企業進行評分後排序,並將評分由高到低排序後的建模樣本分為三組,各組的分組違約率必須呈單調遞增趨勢。在所有滿足這四項條件的序別化解釋變量中,我們最後便挑選一個能使對應信用評分卡在建模樣本中所得到的 AUC 值最大,同時又能在驗證樣本中得到一個合理的 AUC 值者(其 AUC 值的排名在驗證樣本中應屬最好的前五分之一)<sup>13</sup>。

若要考慮財務比率類型,則挑選新變量的手續將分成兩個階段,在第一個階段中,我們將從短期償債能力、長期償債能力、經營績效、獲利能力、與現金流量五大類型中各挑選出一個序別化變量,選入的條件只採前述四條件的第一項條件和第二項條件而已,其他兩項條件就不再使用,也就是說,我們將放寬選入變量的要求,以期從每一個財務類型都能夠順利的選入一個變量。一旦選入五個變量後便進入第二個階段,此時對新變量的選取便恢復原來四項條件的標準。

上述篩選解釋變量的手續充分顯現信用評分卡的建置和一般迴歸模型的分析有根本上的差異,例如,我們並不依照一般假設檢定所採用較嚴苛之顯著水準為 99% 或 95% 的臨界值來評估  $t$  統計量,卻非常重視係數估計值符號不正確者,我們也不考慮標準計量經濟分析所常用的卡方檢定統計量(或是類似  $R^2$  的配適度統計量)來評估整個 Logit 模型估計的優劣,而是採用一般計量經濟分析所不常用的 AUC 值來衡量信用評分卡的表現。

根據上述篩選手續,我們得到兩組序別化解釋變量,考慮財務比率類型的一組稱為 A 組,不考慮財務比率類型的一組稱為 B 組。A 組包括 9 個序別化解釋變量,其中 7 個是分屬四類財務比率類型的財務比率:屬於投資報酬類型的「常續性淨值報酬率」;屬於資本結構及長期償債能力類型的「借款

---

<sup>13</sup> 篩選變量時很容易出現過度配適的狀況(請見 Dwyer, 2005),我們在篩選變量的過程中除了比較在建模樣本中的 AUC 值外,還不斷檢驗由驗證樣本所得到 AUC 值,進行這種交互驗證(cross-validation)手續有助於避免過度配適的問題。

依存度」和「長期負債對淨值比率」；屬於經營績效類型的「每股營業額」、「存貨週轉率」、和「內部保留比率」；以及屬於短期償債能力類型的「速動比率」；此外還包括一個不屬於任何類型的「用人費用率」；以及一個非財務比率的「短期借款」。值得注意的是,縱使我們放寬了篩選變量的條件,仍然無法選入屬於五種財務類型中現金流量類型的解釋變量。不考慮財務比率類型的B組包括7個序別化解釋變量,其中6個是財務比率,分別是屬於投資報酬類型的「稅後資產報酬率」;屬於資本結構及長期償債能力類型的「借款依存度」和「長期負債對淨值比率」;屬於經營績效類型的「每股盈餘」、和「存貨週轉率」;與不屬於任何類型的「有息負債利率」,此外尚包括一個非財務比率的「短期借款」。屬於現金流量類型與短期償債能力類型的解釋變量則未能被納入。

除了自選的兩組解釋變量組合,我們還參考台灣經濟新報所發佈之「企業信用風險指標」(Taiwan Corporate Credit Risk Index, TCRI)所選用的10個變量,包括「借款依存度」、「利息支出率」、「常續性淨值報酬率」、「速動比率」、「收款月數」、「售貨月數」、「總資產報酬率」、及「營業利益率」等8個財務比率,再加上兩個非財務比率:「營業收入淨額」與「資產總額」。將這些解釋變量序別化後再進行Logit迴歸模型的估計,發現「總資產報酬率」與「營業利益率」之迴歸係數估計值的符號不正確,分別刪除之便得到包含8個變量的C組序別化解釋變量。

我們在表7中列舉上述三組解釋變量,由表中我們可清楚看出,「借款依存度」是三組所共有的變量,「長期負債對淨值比率」、「存貨週轉率」、和「短期借款」是A、B兩組所共有的變量,「速動比率」和「常續性淨值報酬率」是A、C兩組所共有的變量,而B、C兩組則沒有其他共同的變量。我們要強調,三組信用評分卡是根據不同的篩選原則從80個財務變量(有興趣的讀者可向作者索取這些財務變量的名單)中所挑選出的解釋變量組合,候選變量數目相當多,篩選方式也不同,但得到的三組信用評分卡(尤其是我們所擬議的前二組)竟有相當多的共同變量,顯然在諸多財務變量中是有一些比較和違約風險相關的變量。Altman and Narayanan (1997)瀏覽文獻中二十個國家的財務預警模型,請參見該文所列舉之各模型所採用的財務變量。最後我們要指出,由台灣新報資料我們亦可辨識每一個樣本企業所屬產業別,我們亦曾嘗試將產業別變量選入信用評分卡中,但發現其影響非常微弱後刪除之。

表 7 篩選後解釋變量

A 組	B 組	C 組
1. 借款依存度 <sup>b</sup>	1. 借款依存度 <sup>b</sup>	1. 借款依存度 <sup>b</sup>
2. 長期負債對淨值比率 <sup>b</sup>	2. 長期負債對淨值比率 <sup>b</sup>	2. 利息支出率 <sup>e</sup>
3. 存貨週轉率 <sup>c</sup>	3. 存貨週轉率 <sup>c</sup>	3. 資產總額 <sup>f</sup>
4. 短期借款 <sup>f</sup>	4. 短期借款 <sup>f</sup>	4. 營業收入淨額 <sup>f</sup>
5. 速動比率 <sup>d</sup>	5. 有息負債利率 <sup>e</sup>	5. 速動比率 <sup>d</sup>
6. 常續性淨值報酬率 <sup>a</sup>	6. 每股盈餘 <sup>c</sup>	6. 常續性淨值報酬率 <sup>a</sup>
7. 內部保留比率 <sup>c</sup>	7. 稅後資產報酬率 <sup>a</sup>	7. 售貨月數 <sup>c</sup>
8. 每股營業額 <sup>c</sup>		8. 收款月數 <sup>c</sup>
9. 用人費用率 <sup>e</sup>		

註: 1. a : 屬於投資報酬類型的解釋變量。

b : 屬於資本結構及長期償債能力類型的解釋變量。

c : 屬於經營績效類型的解釋變量。

d : 屬於短期償債能力類型的解釋變量。

e : 不屬於任何類型的解釋變量。

f : 非財務比例的解釋變量。

2. 三組均未考慮的是屬於現金流量類型的解釋變量。

### 3.3.1 控制變量

在進行迴歸模型的估計時，除了序別化解釋變量外，通常還需要加入一些非關借款戶個別特質因而不會在信用評分卡中出現的特殊解釋變量，以控制建置資料中不屬於借款戶個別差異的異質性，或是建置資料與未來適用資料之間的總體差異，我們將稱呼這些特殊解釋變量為「控制變量」。正如在任何迴歸模型估計過程中遺漏重要解釋變量所造成的「遺漏變量偏誤」(omitted-variable bias)，未能妥善處理控制變量將導致對序別化解釋變量評分的偏誤。我們也要指出，由於控制變量不會出現在信用評分卡中，所以通常也不予以序別化轉換，以簡化估計手續。

不同的樣本會有不同的貸款存續期間，而貸款存續期間通常和違約可能性有直接關連，<sup>14</sup> 此外，在不同時期啟始的貸款（因而造成不同的貸款存續期間），面對的核貸規則乃至准貸條件都可能有所不同。因此貸款存續期間代

<sup>14</sup> 這裡我們提出作用正好相反的兩種可能關連：(1) 貸款存續年數等同於暴露在風險之下的時間，存續年數越長違約的可能性就較大，也就是說，違約可能性和貸款存續期間成正比；(2) 貸款存續期間代表接受考驗的耐力，存續期間越長越表示經得起考驗，就越不會違約。也就是說，違約可能性和貸款存續期間成反比。

表一項重要的樣本異質性, 但又不屬於借款戶的個別特質, 乃一典型的控制變量。當建置資料包含了不同啟始時點的貸款樣本時, 有必要將各貸款樣本的存續期間當做控制變量納入迴歸模型的估計中, 以消除不同貸款存續期間對違約可能性的影響。<sup>15</sup>

至於最常用來控制建置資料與未來適用資料之間總體差異的變量, 是諸如經濟成長率、失業率、貨幣總供給量等反應景氣循環的總體經濟變量。加入這些控制變量的目的自然是要消除經濟景氣對違約可能性的影響, 這在使用跨越景氣循環的資料以建置信用評分卡時特別必要。

在本文中, 不論是篩選變量還是製作最後的信用評分卡時, 我們都將引入一個簡單的控制變量: 自 1998 年至 2000 年間在建模樣本中持續出現的年數, 有 1、2、和 3 三種可能數值。我們必須承認, 雖然這樣的一個變量是我們由台灣經濟新報資料所能找到最簡單而又有意義的控制變量, 但它的意義卻並不是十分明確: 對正常戶而言這個變量顯現的是貸款存續期間,<sup>16</sup> 對違約戶而言這個變量的意義反而比較接近總體經濟狀況: 樣本值 1、2、和 3 分別表示不同的違約年份, 因而代表不同年份之總體經濟狀況對違約戶的影響。由於 1998 年至 2000 年都是台灣企業經營比較艱辛的年份, 越晚違約 (即樣本值為 3 時) 代表企業控制信用風險的能力越高。在建模樣本中對應控制變量值 1、2、和 3 的三個分組違約率分別是 18.8%、6.21%、和 3.77%, 顯示貸款存續期間的變量值越大, 違約可能性就越小, 我們因此預期這個控制變量在迴歸模型中的係數應傾向於是負的。

### 3.4 製作信用評分卡

一旦確認所有的解釋變量, 我們便需進行迴歸模型的估計以獲得這些解釋變量在信用評分卡中所佔的權數 (亦即評分), 由於應變量是違約與否虛擬變量, 標準的線性迴歸模型並不合適, 而 Probit 模型乃至高度非線性的神經

---

<sup>15</sup> 舉例來說, 若擔保品價值比例是信用評分卡中的一個解釋變量, 但假設做為准貸條件一部分的擔保品要求近年來逐漸放寬, 反應在資料中的現象將是新貸款的擔保品價值比例傾向於比較低, 但新貸款的違約可能性會因為觀察期間較短而顯得比較低, 若不將貸款存續期間做為一個控制變量, 所做出的信用評分卡就會錯誤的顯示擔保品價值比例越低其評分就越高。

<sup>16</sup> 樣本值 1 和 2 所呈現的是正確的貸款持續年數, 但樣本值 3 代表的則是「截堵的」(censored) 貸款持續期間, 真實的貸款持續年數可能是 3 也可能多於 3。

網絡 (neural network) 模型都要較線性迴歸模型更為合適, 但業界最普遍的還是 Logit 模型。本研究也採用 Logit 迴歸模型, 主要的原因是 Logit 迴歸模型具有一個其他模型所沒有比較不受樣本違約率大小影響的重要優點, 也就是說, 縱使建模樣本中違約戶和正常戶的比例和母體的並不相符,<sup>17</sup> Logit 迴歸模型的估計結果仍然是一致的 (consistent), 詳情請見 Hosmer and Lemeshow (2000, Ch. 6.3)。由於我們所使用的建模樣本和驗證樣本有很不相同的違約率, 我們認為台灣新報資料各年度的違約率並不穩定, 使用 Logit 模型應有助於緩解這個問題。

在本節中我們將說明如何將 Logit 模型中序別化解釋變量的係數估計值轉變成評分。首先我們將前一小節所決定之三組序別化解釋變量 (以及控制變量) 的 Logit 模型估計結果列於表 8, 各個序別化解釋變量的個別驗證統計量 (AUC 值、信息值、條件熵比例、Spearman 序別相關係數  $\rho_s$ ) 也附列於表 8。

由於序別化解釋變量採用的分組證據權重, 因此係數估計值必然就都是負值, 而控制變量和常數項的係數估計值則可正可負。表 8 各欄統計量中唯一尚未定義的貢獻度, 其定義是: 給定  $m$  個序別化解釋變量, 假設第  $j$  個序別化解釋變量的第  $i$  個樣本值是  $x_{ji}$ 、樣本平均數是  $\bar{x}_j$ 、對應的迴歸係數估計值是  $b_j$ , 則第  $j$  個序別化解釋變量的貢獻度是

$$\text{貢獻度}_j \equiv \frac{\frac{1}{n} \sum_{i=1}^n |b_j x_{ji} - b_j \bar{x}_j|}{\frac{1}{n} \sum_{i=1}^n |b_1 x_{1i} - b_1 \bar{x}_1| + \dots + \frac{1}{n} \sum_{i=1}^n |b_m x_{mi} - b_m \bar{x}_m|}$$

這裡的  $n$  是樣本數。由於第  $j$  個序別化解釋變量的第  $i$  個樣本值的信用評分將大致等於  $-b_j x_{ji}$  (詳情見下一小節的說明), 貢獻度所測度的是一個序別化解釋變量之信用評分的樣本絕對離均差, 佔總信用評分的樣本絕對離均差中的比例, 貢獻度除了受迴歸係數估計值  $b_j$  大小的影響外, 也受樣本值離均差

<sup>17</sup> 由於大多數貸款類型的違約率都不高, 建置信用評分卡時常會刻意提高違約戶樣本的比例, 有些甚至達到違約戶和正常戶各半的比例, 企圖以此改進迴歸模型的估計效率, 但本文並無這種樣本比例問題。

表 8 各解釋變量的驗證統計量及 Logit 模型估計結果

解釋變量	係數估計	t 統計量	貢獻度	AUC 值	$\rho_s$	信息值	條件熵比
A 組							
借款依存度	-0.550	-4.32	20.32	0.794	0.303	1.397	0.162
每股營業額(元)	-1.119	-5.50	17.55	0.639	0.143	0.237	0.027
短期借款	-0.590	-4.51	16.71	0.723	0.230	0.773	0.091
常續性淨值報酬率	-0.329	-2.66	10.02	0.744	0.250	0.800	0.095
長期負債/淨值	-0.418	-3.18	9.40	0.701	0.207	0.589	0.076
用人費用率	-1.034	-3.76	9.29	0.599	0.101	0.099	0.011
存貨週轉率(次)	-0.384	-2.41	6.24	0.653	0.157	0.329	0.041
內部保留比率	-0.182	-1.19	5.34	0.660	0.173	0.711	0.068
速動比率	-0.143	-1.17	5.13	0.737	0.251	0.816	0.094
控制變量	-1.237	-12.19					
截距項	0.337	1.59					
B 組							
有息負債利率	-0.642	-4.19	19.47	0.685	0.188	0.876	0.095
短期借款	-0.571	-4.56	17.06	0.723	0.230	0.773	0.091
每股盈餘	-0.492	-3.33	16.51	0.738	0.243	0.797	0.090
借款依存度	-0.379	-3.46	14.93	0.794	0.303	1.397	0.162
存貨週轉率(次)	-0.732	-4.93	12.37	0.653	0.157	0.329	0.041
長期負債/淨值	-0.469	-3.67	11.10	0.701	0.207	0.589	0.076
稅後資產報酬率	-0.255	-1.47	8.55	0.708	0.227	0.924	0.084
控制變量	-1.161	-11.94					
截距項	0.220	1.06					
C 組							
借款依存度	-0.586	-5.76	24.54	0.794	0.303	1.397	0.162
常續性淨值報酬率	-0.633	-4.94	20.98	0.739	0.244	0.803	0.095
資產總額	-0.507	-3.13	15.49	0.665	0.180	0.350	0.041
售貨月數	-0.584	-4.54	9.86	0.654	0.157	0.362	0.046
營業收入淨額	-0.674	-2.00	8.28	0.583	0.088	0.099	0.012
收款月數	-0.636	-2.80	7.96	0.602	0.107	0.119	0.016
速動比率	-0.184	-1.53	7.40	0.737	0.251	0.815	0.094
利息支出率	-0.109	-0.82	5.50	0.763	0.269	1.314	0.133
控制變量	-1.158	-12.02					
截距項	0.200	0.98					

(即樣本變異程度)的影響。<sup>18</sup>我們還要強調,貢獻度只涵蓋了計入信用評分卡的序別化解釋變量,並不被考慮常數項和控制變量。

表 8 中各組序別化解釋變量是依照貢獻度的大小排列,以凸顯各個統計

<sup>18</sup> 如同在第 3.2.1 小節的解釋,解釋變量經過序別化後,其樣本值不是分組違約比率,就是分組證據權重,所以可跨變量的直接比較,上述貢獻度的定義因此才有意義。

量大小排序之間的異同對比。我們可由表 8 中看出迴歸係數估計值、 $t$  統計量、貢獻度、乃至於四個驗證統計量均有不同的排序, 經過仔細比對我們也注意到, 表 8 所列舉的七個統計量可分成三群, 每一群內的統計量都有極為類似的排序: 迴歸係數估計值、 $t$  統計量、與貢獻度構成第一群, AUC 值與  $\rho_s$  是第二群, 這顯示辨識度與序別相關係數之間有很密切的關係; 第三群包括信息值與條件熵比例, 顯示由信息理論所推導出來的統計量所呈現也是很類似的消息。此外我們也要指出, 就迴歸模型的估計結果而言, 係數估計值和對應的  $t$  統計量之間存在一個接近正比的關係是一個相當不尋常的現象, 這顯然是由所有解釋變量均以同一種序別化轉換方式轉變成可互相比較的證據權重所造成的。

上述三群統計量的另一個特點是不同群之統計量所呈現的排序會有相當大的差別, 我們尤其關注貢獻度的排序和個別驗證統計量的排序之間有很大的差異, 由於前者考慮了所有解釋變量之間的交互影響而後者並未考慮, 這個結果除了顯示個別驗證統計量的侷限性外, 也提醒我們解釋變量間的相關性會對評分造成重大影響。

### 3.4.1 評分的定義

再次假設序別化解釋變量的數目是  $m$  個, 第  $j$  個序別化解釋變量的第  $i$  個樣本值是  $x_{ji}$ , 對應的迴歸係數估計值是  $b_j$ , 此外, 假設  $a_0$  和  $a_1$  分別代表常數項和控制變量的迴歸係數估計值, 則根據 Logit 模型的設定, 可得「不違約險算比」(odds ratio)之對數值的估計 ( $\hat{p}_i$  表示對第  $i$  個樣本之違約可能性  $p_i$  的估計值)如下:

$$\ln \left( \frac{1 - \hat{p}_i}{\hat{p}_i} \right) = -(a_0 + a_1 \cdot w_i + b_1 x_{1i} + \cdots + b_m x_{mi}), \quad (6)$$

這裡的  $w_i$  乃是控制變量的第  $i$  個樣本值。現若給定一個控制變量值  $w^*$ , 則我們可以此給定值  $w^*$  替代樣本值  $w_i$  並將上述等式等號右邊部分分解如下:

$$\tilde{z}_0 = -(a_0 + a_1 \cdot w^*), \quad (7)$$

和

$$\tilde{z}_{ji} = -b_j x_{ji}, \quad j = 1, 2, \dots, m, \quad (8)$$

再加總為

$$\tilde{z}_i = \tilde{z}_0 + \tilde{z}_{1i} + \dots + \tilde{z}_{mi}. \quad (9)$$

由於  $\tilde{z}_i$  和違約可能性  $\hat{p}_i$  大致成反比, 所以可視為第  $i$  個借款戶的信用評分,  $\tilde{z}_i$  越高違約可能性就越低。

相對於總評分  $\tilde{z}_i$ , 其分項  $\tilde{z}_{ji} = -b_j x_{ji}$  代表的是第  $i$  個借款戶在第  $j$  個解釋變量上所得到的評分, 由於迴歸係數估計值  $b_j$  值固定,  $\tilde{z}_{ji}$  的任何變化所反應的就只是  $x_{ji}$  之分組證據權重值 (或違約比例) 之間的差異, 也就是說, 分組證據權重值間的差異越大, 評分  $\tilde{z}_{ji}$  間的差異就會越大。我們還要指出, 不同的分項評分  $\tilde{z}_{1i}, \dots, \tilde{z}_{mi}$  之間也可相互比較, 當一個分項評分的變異程度要比其他的為大時, 通常就表示對應的序別化解釋變量之相對重要性較高, 早先對貢獻度的定義便是建立在這個想法上。

總評分  $\tilde{z}_i$  除了以其相對數值顯示信用的相對好壞外, 由 (6) 式得知其絕對數值尚有不違約險算比對數值的解釋, 根據這個解釋, 我們可進一步對總評分  $\tilde{z}_i$  做一個不變動其相對大小的線性轉換, 但要求調整後的總評分和違約機率之間滿足如下兩個對應: (1) 當違約機率等於 0.0003 時, 對應的總評分要等於 1,000; (2) 當違約機率等於 0.9997 時, 對應的總評分要等於 0。經過適當的演算後, 我們可推導出滿足這兩個條件的線性轉換公式如下:<sup>19</sup>

$$\begin{aligned} z_i &= 500 + 61.641 \cdot \tilde{z}_i \\ &= (500 + 61.641 \cdot \tilde{z}_0) + 61.641 \cdot \tilde{z}_{1i} + \dots + 61.641 \cdot \tilde{z}_{mi} \\ &\equiv z_0 + z_{1i} + \dots + z_{mi}, \end{aligned} \quad (10)$$

<sup>19</sup> 若違約機率等於兩個給定值  $p_a$  和  $p_b$  時, 我們要求對應的總評分要分別等於給定的數值  $Z_a$  和  $Z_b$ , 則必須對總評分  $\tilde{z}_i$  做如下的線性轉換:

$$z_i = Z_b + \left( \frac{Z_a - Z_b}{\tilde{Z}_a - \tilde{Z}_b} \right) (\tilde{z}_i - \tilde{Z}_b),$$

其中  $\tilde{Z}_a = \ln[(1 - p_a)/p_a]$  和  $\tilde{Z}_b = \ln[(1 - p_b)/p_b]$ 。



這裡的  $z_0 (\equiv 500 + 61.641 \cdot \tilde{z}_0)$  將是所謂的「基準評分」，而  $z_{ji} (\equiv 61.641 \cdot \tilde{z}_{ji})$  則是第  $j$  個別序別化解釋變量的評分。調整後的總評分  $z_i$  等於基準評分加上所有  $m$  個別序別化解釋變量之評分的總和，而  $z_i$  除了滿足前述有關違約機率與評分之間的兩個條件外，尚有如下兩項有用的性質：當不違約險算比等於 1 時（即不違約和違約可能性都是 0.5，故  $\tilde{z}_i = 0$ ），則總評分  $z_i$  將正好等於 500。且每當總評分  $z_i$  增加 42.727 分時，則不違約險算比就增加一倍。

在附錄的附表 1 中，我們列出 B 組信用評分卡中的基準評分（標題下方）、各序別化解釋變量的分組組界點、對應分組評分、在建模樣本與驗證樣本中的各分組戶數、分組戶數佔總戶數比例（樣本率）、以及分組違約率、以及由建模樣本與驗證樣本之分組樣本率和分組違約率所分別求得的信息值 (PSI)，<sup>20</sup> 我們將在下一小節（第 3.4.2 小節）對這些結果做一比較完整的說明。計算基準評分時所有借款戶的控制變量數值都假設為  $w^* = 3$ ，亦即當我們根據擬議的信用評分卡對新借款戶計算評分時，將假設它是處於建模樣本之貸款持續期間為 3 的狀況。

在接下來的附表 2 和附表 3 中我們分別列出兩種信用評分卡的 8 級信用評等，其第二欄（標題是分等等界）所列出的就是信用評分組界值，此外還有建模樣本與驗證樣本中各分等戶數、分等戶數佔總戶數比例（樣本率）、與三種分等違約率，和附表 1 一樣，附表 2 和 3 也包含了由建模樣本與驗證樣本之分組樣本率和分組違約率所分別求得的信息值 (PSI)，我們將在第 3.4.3 小節對這些信用評等的結果詳加解釋。至於檢定統計量欄所列數值則是校準度檢定的結果，我們將在第 4 節解釋其意義。本文第 1 節之表 1 和表 2 所列舉之作為範例的信用評分卡內容，就是擷取自附表 1 和附表 3 之 B 組信用評分卡的評分及評等結果。

### 3.4.2 評分的分析

由附表 1 可知，信用評分卡的設計大為簡化了對各解釋變量評分的計算方式：對於解釋變量的任何樣本值，只要找出所屬的數值區段，便可當即查對出對應評分，甚至對於一些解釋變量（例如存貨週轉率）的缺漏值都可予以評分。

根據解釋變量由小而大的數值區段以及對應評分我們可看出，每一個解釋變量和評分之間是正關係還是負關係，在表 9 中我們依照三組信用評分卡

<sup>20</sup> 為節省篇幅我們未列出將 A 組和 C 組信用評分卡的評分結果，有興趣的讀者可向作者索取。

表 9 解釋變量和信用評分的關係

A 組	正關係	每股營業額	常續性淨值報酬率	存貨週轉率	速動比率
	負關係	借款依存度	短期借款	長期負債 / 淨值	
	倒 U 關係	用人費用率	內部保留比率		
B 組	正關係	每股盈餘	存貨週轉率	稅後資產報酬率	
	負關係	短期借款	借款依存度	長期負債 / 淨值	
	正 U 關係	有息負債利率			
C 組	正關係	常續性淨值報酬率	售貨月數	速動比率	
	負關係	借款依存度	資產總額		
	倒 U 關係	營業收入淨額	收款月數	利息支出率	

中各解釋變量和評分之間的關係將解釋變量加以分類。各解釋變量和評分之間除了呈現相當合理的正負線性關係外,還有非線性的 U 字或倒 U 字關係。我們要強調,能夠有效的捕捉到解釋變量和評分之間的這種非線性關係必須歸功於解釋變量的序別化轉換,若未對解釋變量進行序別化轉換,則解釋變量和評分之間的任何非線性關係都將簡化(很可能是嚴重的扭曲)為線性關係。

相對於多個解釋變量的評分呈現倒 U 字形分布,只有一個解釋變量的評分是 U 字形分布,可見對多數財務變量而言,出現極端值不論大小都隱含違約風險的增加,信用評分也就隨之呈現倒 U 字形分布。

如前所述,能夠保留缺漏值並給予評分是序別化解釋變量的優點,但我們要強調,雖然缺漏值的評分方式和正常值幾乎完全一樣,<sup>21</sup>我們通常還是會將缺漏值視為非正常狀態下的產物,應儘可能的了解缺漏值產生的原因,以進一步評估其評分的合理性。原則上缺漏值的評分不應太過極端,尤其不應是分組評分中的最高者。很不幸的,B 組信用評分卡中三個包含缺漏值的解釋變量中的兩個(有息負債利率和存貨週轉率)都被指定了最高的分組評分,尤其是有息負債利率的評分較最低分組評分多了幾乎 130 分。對此我們的看法是有息負債利率和存貨週轉率作為財務比例的分母,前者是負債額而後者是存貨額,皆有可能是零值,進而造成缺漏值,也就是說,這兩個財務比

<sup>21</sup> 早先我們在第 3.1 小節(第 560 頁)提過,缺漏值比例過高以致超過 30% 的變量將被刪除,而對於缺漏值比例低於 2% 的變量則會刪除對應缺漏值的樣本。

例變量產生缺漏值的最可能原因是代表企業狀況良好的零負債或零存貨，對應缺漏值的高分組評分相對而言還算合理。

經過上述的比對和分析後，各個序別化解釋變量的分組評分都還算是在可接受的範圍內，在第4節中我們將進一步的對擬議的三組信用評分卡進行更完整的測試和驗證。

附表1中分組評分之右的八欄數字所顯示的是，在建模樣本和驗證樣本中各分組之樣本分布與違約率的對比，我們除了可檢視各分組的樣本戶數是否過於不均（跨列的比較）之外，還可比較建模樣本和驗證樣本的樣本分布乃至分組違約率之間是否過於不同（跨欄的比較）。為了便於後者的分析，我們在最後兩欄列舉了樣本率和違約率的信息值：給定  $\pi_{0j}$  和  $\pi_{1j}$  分別是建模樣本和驗證樣本第  $j$  評分組的樣本率（或是違約率），則該二欄位的各個數值是  $(\pi_{0j} - \pi_{1j}) \cdot \ln(\pi_{0j}/\pi_{1j})$ ，而最底端的數值便是信息值，乃其上數值的總和，這裡的信息值越小就表示建模樣本和驗證樣本之間的差異越小，亦即財務解釋變量的樣本值分布以及和違約變量的關係越穩定，為明確指出這裡的信息值和之前做為驗證統計量之信息值有極為不同的意義，我們將現今的信息值改稱為「母體穩定指標」（population stability index 或簡稱為 PSI）。

我們還可進一步將 A 組、B 組、和 C 組信用評分卡中所有財務解釋變量的母體穩定指標加總，以比較這三組信用評分卡的表現。A 組、B 組、和 C 組之樣本率母體穩定指標的總和分別是 0.398、1.104、和 0.319，顯示 B 組所包含之財務解釋變量的樣本分布並不穩定，這主要是有息負債利率（之第二分組）所造成的問題。A 組、B 組、和 C 組之違約率母體穩定指標的總和分別是 3.69、3.083、和 6.562，所以 B 組所包含之財務解釋變量在違約率上反而比較穩定。

### 3.4.3 評等的分析

附錄的附表2和附表3是根據建模樣本的評分結果分別製作的兩種八等級的評等表。評等表也是信用評分卡的一部分，其主要功能便是根據每一個企業的信用評分再給予該企業以一個信用評等，要求違約可能性相同或接近的企業都歸屬於同一個評等等級，所以我們可對屬於同一評等等級的所有企業都指定以一個相同的違約機率。（注意：評等等級數越低，違約可能性就越低，也就是說，評等等級正好和信用評分值成反比。）我們在第一節曾經指出，就

執行新巴賽爾資本協定而言, 評等表要比信用評分表更為重要。

附表 2 的分等方式是將建模樣本企業依照其信用評分加以排序後均分八組並隨之決定其七個分界點, 而附表 3 的分等方式則是根據第 3.2.1 小節所述搜尋組界點的方法來找出能極大化 AUC 值的七個分界點。<sup>22</sup>

附表 2 和附表 3 分別包含了兩種對應各評等的違約率: 實際違約率和預測違約率, 實際違約率乃各分等中違約戶佔分等戶數的實際樣本比率, 要進一步了解預測違約率的意義前, 則要先知道根據 Logit 迴歸模型所隱含之不違約險算比和信用評分的關係 (6) 式, 我們可由信用評分  $z_i$  反推出建模樣本中各企業之違約機率的估計值  $\hat{p}_i$ , 由於信用評分包含了控制變量的數值, 對控制變量的不同設定便會產生不同的違約機率估計值。若控制變量的數值是採建模樣本的樣本值, 則對應的違約機率估計值將稱為「預測違約率」; 若控制變量的數值是給定的數值 3 (請見第 3.3.1 小節), 則對應的違約機率估計值將稱為「基準違約率」。附表 2 和附表 3 中的預測違約率乃各評等所屬企業之預測違約機率的平均值。

預測違約率將是我們對每一評等等級所指定的違約率, 對未來借款戶計算其信用評分時, 如同我們在第 3.3.1 小節所言, 會將控制變量的數值設定為 3, 附表 1 中所列的基準評分是根據控制變量的這個設定值所定義的, 對未來借款戶進行評等時, 每一評等等級的對應違約率也將就是這個預測違約率。

預測違約率還將是我們進行校準度檢定的根據, 更具體的說, 不論是建模樣本還是驗證樣本, 我們將檢定的虛無假設是: 屬於同一個評等等級之所有企業的違約機率都等於對應等級之預測違約率, 也就是說, 在使用校準度檢定統計量的  $p$  值公式 (5) 式時, 對應的  $q_k$  將就是第  $k$  評等等級的預測違約率。舉例來說, 對 A 組信用評分卡在附表 2 中的第 3 級評等而言, 虛無假設是: 屬於該評等等級之企業 (建模樣本的 315 個企業或是驗證樣本的 237 個企業) 的違約機率都等於 0.027 (即 A 組信用評分卡第 5 級評等的預測違約率)。附表 2 和附表 3 之標題為「檢定統計量」欄下的數值便是對 (5) 式的計算結果, 只要對應數值大於給定的檢定容忍度 (像是 1% 或是 5%), 我們便不拒絕對應的虛無假設, 表示該評等等級內所有企業的違約機率均等於給定的預

<sup>22</sup> 決定附表 3 中的第 1 級評等和第 8 級評等時, 我們刻意的將其樣本數大致減少到其他評等等級的一半, 原因是我們認為信用風險最大與最小的極端情形都應只是所有企業中的少數。

測違約率。<sup>23</sup> 注意：在校準度檢定統計量的公式中有一個尚未設定之  $\rho$  參數，乃同一評等等級各企業之信用風險因子間的相關係數，由於很難在進行假設檢定時同時估計這個  $\rho$  參數，所以我們將主觀給定其數值為 0.25。由於新巴賽爾資本協定中對各項資產風險之間相關係數的規定其數值最高者是 0.24，我們所採用的 0.25 算是一個相當保守的數值。

根據附表 2 和附表 3 的檢定結果可知，不論是建模樣本還是驗證樣本，也不論採用哪一個常用的檢定容忍度，我們都不能拒絕三組信用評分卡中任何評等等級內之違約率的校準，這個結果除了顯示我們所建置的附表 2 和附表 3 均正確的區隔企業違約可能性外，也暗示校準度檢定統計量的檢定力可能不是很高。

附表 2 和附表 3 中最右邊兩欄之樣本率母體穩定指標和違約率母體穩定指標所顯示的是建模樣本和驗證樣本在樣本分布和違約狀況的對比，其定義和附表 1 同，可見前一小節的解釋。

由於附表 2 的評等是由均分建模樣本所得，對於這個評等表中的樣本分布，我們直覺上還可提出如下的兩點要求：(1) 在驗證樣本中各評等等級的戶數應大致相等；(2) 建模樣本和驗證樣本各評等的實際違約率應該隨著等級由 1 至 8 遞增而遞增（最起碼不能遞減）。關於前一點我們根據附表 2 倒數第 4 欄發現，在驗證樣本中不論是哪一組信用評分卡，屬於前三評等的戶數均遠多於屬於後三評等的戶數，這種不均勻的現象乃驗證樣本的違約率低於建模樣本違約率所造成的副作用。對於第二點要求我們發現，在建模樣本中 A 組和 B 組信用評分卡能夠符合預期，但 C 組則否（第 2 等級和第 3 等級的違約率倒錯）。在驗證樣本中 B 組信用評分卡能夠符合預期，但 A 組和 C 組則否。我們認為相對於驗證樣本，在建模樣本中不應容許實際違約率發生倒錯的現象，一旦發生便有必要重建信用評分卡。換言之，我們將認定 C 組信用評分卡有嚴重缺陷，A 組信用評分卡也有輕微的缺陷，只有 B 組信用評分卡是唯一完全符合要求的。

雖然附表 2 的評等表容許我們進行更多的分析以評比信用評分卡的表現（所以在建置信用評分卡的過程中應該製作附表 2 以執行這些分析），但在實際操作時，我們建議還是應該採用附表 3 的評等表做為最終的產出（附表

<sup>23</sup> 有些評等等級的樣本違約機率等於 0，例如 A 組信用評分卡在建模樣本的第 1 級和第 2 級，則因為校準度檢定無法用來檢定各企業的違約機率等於 0 的假設，將缺對應的檢定統計量。

2 將被視為建置過程中的一項中間產出而已)。第 1 節中的表 2 所複製的便是附表 3, 我們在下一節對信用評分卡所進行的測試和驗證也都是建立在附表 3 上。最後我們還要指出, 分等方式並不限於根據建模樣本之樣本數的分等方式而已, 我們可根據預測違約率的大小進行分等: 例如: 我們可要求預測違約率大於 0.3 的樣本併為一等, 預測違約率介於 0.1 和 0.3 的樣本分為兩等, 預測違約率小於 0.1 的樣本分為四等。本文雖未採用這種分等方式, 但要強調這種分等方式也許更合乎信用評分卡使用者的預期與需求。

#### 4. 信用評分卡的測試和驗證

我們將三組信用評分卡的測試和驗證結果列於表 10。對於表 10 所列的結果我們有三點說明: (1) 當我們進行測試或驗證時, 計算信用評分所使用的控制變量數值將是實際樣本值, 而不是未來將信用評分卡應用到新資料時所設定的控制變量值<sup>3</sup>。(2) 由於屬於信息理論的信息值、條件熵比、和條件變異比以及屬於違約率配適度的 HL 配適統計量均只適用於間斷型信用評分系統, 所以它們都是根據附表 3 的評等結果(而非信用評分結果)所計算而得。(3) 隸屬於違約率配適度的 Brier 分數與 HL 配適統計量的計算需要違約率的預測值, 我們將採用前一小節(第 576 頁)所定義之預測違約機率作為違約率的預測值。

由表 10 可知, 除了最不穩定的信息值外, 所有驗證統計量(HL 配適統計量除外)都顯示三組信用評分卡在建模樣本中的表現均是以 A 組為最佳, 以 C 組為最差。由於驗證樣本的總違約率遠低於建模樣本的總違約率(見表 6), 三組信用評分卡在驗證樣本中的表現都遠遜於在建模樣本的表現。我們特別注意到 Brier 分數在驗證樣本的表現反而較建模樣本的表現為好, 其原因也仍在於總違約率上的差異(Brier 分數很受總違約率的影響, 和總違約率有很明顯的反比關係)。若再進一步的觀察三組信用評分卡在驗證樣本中的優劣排序, 除了信息值外, 所有的驗證統計量(HL 配適統計量除外)均顯示 B 組最佳, C 組最差。至於 HL 配適統計量的使用則需根據卡方分配的臨界值進行假設檢定, 由於附表 3 之評等等級的總數是 8, 適用建模樣本的卡方分配自由度是 6, 對應信賴水準為 95% 和 99% 的單尾檢定臨界值分別是 12.6 與

表 10 信用評分卡的測試和驗證統計量

	建模樣本			驗證樣本		
	A 組	B 組	C 組	A 組	B 組	C 組
辨識度						
AUC 值	0.914	0.910	0.903	0.794	0.823	0.792
KS 值	0.676	0.676	0.662	0.477	0.514	0.463
序別相關						
Spearman's $\rho_s$	-0.412	-0.408	-0.402	-0.173	-0.190	-0.172
Kendall's $\tau_a$	-0.137	-0.136	-0.133	-0.034	-0.037	-0.034
Kendall's $\tau_b$	-0.337	-0.334	-0.328	-0.141	-0.155	-0.140
信息理論						
信息值	2.558	3.065	2.827	1.351	0.801	1.135
離散統計量	3.974	3.768	3.608	1.398	1.819	1.384
條件熵比	0.386	0.378	0.360	0.141	0.175	0.126
違約率配適度						
Brier 分數	0.053	0.055	0.056	0.029	0.028	0.029
HL 配適統計量	1.806	2.072	1.527	12.940	9.776	4.847

16.8, 而適用驗證樣本的卡方分配自由度是 8, 對應信賴水準為 95% 和 99% 的單尾檢定臨界值分別是 15.5 與 20.1, 所以我們根據 HL 配適統計量所得到的檢定結論是, 三組信用評分卡之評等表的配適度均為良好, 在這一方面我們無法區別三組信用評分卡的高下。

綜合上述的討論我們發現, A 組和 B 組的表現似乎是在伯仲之間, 觀諸這兩組信用評分卡有四個共同的解釋變量, 這樣的結果顯然並不令人驚訝。財務類型平衡的 A 組顯然在建模樣本中有稍佳的測試績效, 而未考量財務類型平衡的 B 組則在驗證樣本中有較優的驗證表現, 對於財務類型的平衡是否有助於提升信用評分卡的問題, 我們無法做出定論。但 A 組的 9 個解釋變量要比 B 組的 7 個解釋變量為多則意味著, 擴增解釋變量雖有助於改進建模樣本的測試結果, 但不見得能夠提升驗證樣本的驗證表現。

由於 A 組和 B 組不論是測試還是驗證均優於 C 組, 顯然我們可經由反覆試誤找出優於台灣經濟新報根據長年經驗所建議的財務變量組合。但我們也必須強調, 這三組信用評分卡的表現雖有高下之分, 但差異卻不是非常顯著, 我們將就此做出如下推論: 大體而言我們是可以穩定的由不超過十個財報變量有效的預測違約可能性, 本文所擬議的三組信用評分卡也都具有精準預測的能力。

#### 4.1 與 TCRI 評分指標的比較

在本小節中我們將進一步將我們所擬議之三組信用評分卡和台灣經濟新報所推出之「台灣企業信用風險指標」(TCRI 指標)做一比較。我們所考量的 TCRI 指標是其第一階段指標(將簡稱 TCRI 指標 I), 這一階段的指標分為 9 等級, 其建置過程中只考慮財務報表變量, 未涉及台灣經濟新報所添加對非財務面的其他規定, 所以缺漏值也最少。和我們之前所使用的樣本相比, TCRI 指標 I 在建模樣本中少了 304 筆, 包括 190 筆違約戶, 而在驗證樣本中則少了 28 筆資料, 包括 4 筆違約戶, 詳情請見表 11。我們將簡稱有 TCRI 指標 I 的企業樣本為 TCRI 樣本。由於 TCRI 樣本在建模樣本部分漏失的大多數是違約戶樣本, 而在驗證樣本部分則多漏失正常戶樣本, 所以建模樣本的違約率降至一個較驗證樣本的違約率還要低的水準。

針對 TCRI 樣本的測試和驗證結果列於表 12,<sup>24</sup> 由於建模樣本違約率大幅降低, 三組信用評分卡在建模樣本中的表現均明顯惡化, 且其排序亦發生變動(信息值例外), 原來排名第二的 B 組上升為第一, 但原來最差的 C 組仍維持掛尾。在驗證樣本方面, 雖然樣本違約率也下降, 但因幅度較小, 所有的驗證統計量卻反而有所改善, 三組信用評分卡之間的績效排序(信息值例外)也未改變, 仍然以 B 組為最優而 C 組為最差, B 組信用評分卡的穩定度由此可見一斑。

表 12 中最重要的結果自然是在於擬議的三組信用評分卡和 TCRI 指標 I 之間的比較: TCRI 指標 I 在 11 個驗證統計量都明顯落後, 特別有趣的是, C 組信用評分卡完全就是根據 TCRI 指標 I 所採用的十個財務變量(事實上 C 組信用評分卡只用了八個), 只是對財務變量進行了序別化並予以不同的權數便產生了很明顯的改進, 這充分顯現對變量的轉換以及 Logit 模型估計的重要性。

---

<sup>24</sup> 因為 TCRI 指標 I 並未提供對應的違約率, 所以我們無法計算屬於違約率配適度的兩個統計量。



表 11 TCRI 樣本結構

	總戶數	正常戶數	違約戶數	違約率 (%)
建模樣本	2,215	2,176	39	1.76
驗證樣本	2,195	2,133	62	2.82

表 12 TCRI 和信用評分卡的比較

	建模樣本				驗證樣本			
	TCRI	A 組	B 組	C 組	TCRI	A 組	B 組	C 組
辨識度								
AUC 值	0.768	0.815	0.824	0.807	0.768	0.806	0.835	0.805
KS 值	0.424	0.532	0.575	0.524	0.421	0.500	0.526	0.500
序別相關								
Spearman's $\rho_s$	0.123	-0.144	-0.148	-0.140	0.155	-0.176	-0.192	-0.175
Kendall's $\tau_a$	0.019	-0.022	-0.022	-0.021	0.029	-0.034	-0.037	-0.033
Kendall's $\tau_b$	0.107	-0.117	-0.121	-0.114	0.134	-0.144	-0.157	-0.143
信息理論								
PSI	0.662	1.054	0.770	0.841	0.514	0.898	0.966	1.214
離散統計量	1.122	1.656	1.776	1.514	1.194	1.551	2.016	1.545
條件熵比	0.108	0.150	0.181	0.133	0.130	0.162	0.188	0.135
違約率配適度								
Brier 分數	-	0.029	0.029	0.029	-	0.027	0.026	0.028
HL 配適統計量	-	8.717	9.263	5.410	-	14.232	9.551	4.829

## 5. 結論

在本文中我們詳細說明並利用台灣經濟新報之企業財報資料以實例展示建置和驗證信用評分卡的手續, 本文強調信用評分卡的建置雖然是建立在 Logit 迴歸模型的估計結果上, 但重點卻不是一般實證經濟分析所重視的模型估計過程以及估計結果分析, 而是在估計之前對變量的轉換和篩選, 以及估計之後對估計值的轉換和驗證。

本文的重要實證結果有五: (1) 我們所擬議的三組信用評分卡都顯著的優於 TCRI 指標 I, 包括使用完全相同財報變量的 C 組, 這有力的顯示客觀的

統計方法是要比主觀的經驗法則能更有效的辨識違約風險,不論是業界還是學界都應該繼續研究改進建置與驗證信用評分卡的統計方法。(2) 一個解釋變量在整個信用評分卡的貢獻和其個別驗證統計量的高下有很大的差異,這表示解釋變量之間的相關性是會對信用評分卡產生很重大的影響,所以篩選解釋變量不只是挑選哪一個解釋變量的問題而已,而是要挑選出一個可互相配合的解釋變量組合,這使得篩選解釋變量成為建置信用評分卡的過程中最困難也是最關鍵的一步。我們在文中建議了多重的篩選標準,雖然這些篩選標準可能不見得是最完美的,但我們認為多層次的篩選標準要比單一標準更有可能找出好的解釋變量組合。(3) 在擬議的三組信用評分卡中,未考量財務變量類型平衡且變量數目最少的B組卻展現出最佳最穩的表現,這意味著財務變量類型的平衡乃至信用評分卡中變量的數目並非信用評分卡績效的主要決定因素。(4) 在我們所考量的11種驗證統計量中,除了信息值有數值不穩的現象外,其他的驗證統計量,不論是我們所推介的AUC值還是業界所偏好的KS統計量,均顯現極為類似的評比結果(尤其是辨識度和序別相關),所以採用哪一種驗證統計量做為測試和驗證的工具似乎並不是那麼重要。(5) 較諸文獻中對缺漏值的多種處理方法,本文的處理方式比較合理,但有些缺漏值仍然被指定了過高的評分,在這方面未來還有改進的空間。

建置和驗證信用評分卡的過程中牽涉到無數個相當主觀的細部設定和調整,有些決定無可避免的會產生相當大的後果,舉例來說,序別化解釋變量時所採用的分組數、違約時點和資料擷取時點之間的時差(見表5)、控制變量的選取、缺漏值的處理、非財報資料的使用等,都可能對信用評分卡產生很大的影響。由一次實證研究我們實在是無法盡知這許多主觀設定的後果,我們因此期盼有更多的相關研究能夠公佈發表,這對「新巴賽爾資本協定」實施後,銀行管控本身的信用風險,乃至政府監理機關維護整體金融安定都有很重大的意義。

在本文中我們使用了11個驗證統計量來佐證信用評分卡預測違約的績效,但我們不得不指出,銀行更在意的應該是由借款而來的獲利能力,信用評分卡優劣的最終裁判顯然應該是「能否改進使用者的獲利」,我們所建立的信用評分卡和TCRI之熟優熟劣,自也應以使用者的獲利為比較準則。但苦於缺少銀行對借款額、利率定價,以及還款期間的資料,因而無法就這個觀點進行深入的比較,事實上,由於資料的闕如,文獻中迄今也尚未出現這類的

研究,最多只能間接證明信用評分卡並非以極大獲利為目標(例如 Roszbach, 2004; Boyes et al., 1989)。若能超脫由統計學所推導出的驗證統計量,收集相關資料並發展出建立在財務理論上的驗證統計量,將是一個非常有價值的未來努力方向。

附表 1 B 組信用評分卡

基準評分: 701.09

分組	組界點	組評分	建模樣本			驗證樣本			比較	
			分組 戶數	樣本率	違約率	分組 戶數	樣本率	違約率	樣本率 PSI	違約率 PSI
1. 有息負債利率(%)										
1	$x = 0.00$	69.00	233	0.092	0.017	273	0.123	0.011	0.009	0.003
2	$0.00 < x \leq 5.38$	48.20	425	0.169	0.033	1,206	0.543	0.019	0.437	0.008
3	$5.38 < x \leq 5.97$	21.71	229	0.091	0.066	203	0.091	0.034	0.000	0.021
4	$5.97 < x \leq 7.59$	-9.05	741	0.294	0.127	345	0.155	0.058	0.089	0.054
5	$7.59 < x \leq 8.14$	-24.13	217	0.086	0.159	44	0.020	0.068	0.096	0.077
6	$8.14 < x < \infty$	-12.82	532	0.211	0.120	131	0.059	0.069	0.194	0.028
7	缺漏值	104.64	142	0.056	0.007	21	0.009	0.048	0.086	0.079
									0.910	0.270
2. 短期借款 $\times 10^{-8}$										
1	$x = 0.00$	47.82	559	0.222	0.025	569	0.256	0.014	0.005	0.006
2	$0.00 < x \leq 0.63$	33.04	558	0.222	0.038	510	0.229	0.010	0.000	0.037
3	$0.63 < x \leq 2.25$	17.66	543	0.216	0.057	486	0.219	0.035	0.000	0.011
4	$2.25 < x < \infty$	-29.94	859	0.341	0.190	658	0.296	0.055	0.006	0.167
									0.011	0.222
3. 每股盈餘										
1	$-\infty < x \leq -1.05$	-32.08	467	0.185	0.227	402	0.181	0.067	0.000	0.195
2	$-1.05 < x \leq 0.00$	-11.55	406	0.161	0.136	346	0.156	0.049	0.000	0.089
3	$0.00 < x \leq 1.14$	5.33	749	0.297	0.081	672	0.302	0.022	0.000	0.077
4	$1.14 < x \leq 1.61$	30.18	208	0.083	0.048	191	0.086	0.021	0.000	0.022
5	$1.61 < x \leq 2.30$	53.94	204	0.081	0.022	195	0.088	0.005	0.001	0.025
6	$2.30 < x < \infty$	63.05	485	0.193	0.010	417	0.188	0.005	0.000	0.003
									0.001	0.412
4. 借款依存度(%)										
1	$-\infty < x \leq 19.50$	47.20	841	0.334	0.013	840	0.378	0.008	0.005	0.002
2	$19.50 < x \leq 34.11$	30.48	303	0.120	0.026	302	0.136	0.020	0.002	0.002
3	$34.11 < x \leq 62.56$	13.26	466	0.185	0.054	430	0.193	0.023	0.000	0.026
4	$62.56 < x \leq 174.17$	-15.04	694	0.276	0.160	553	0.249	0.067	0.003	0.081
5	$174.17 < x < \infty$	-38.72	215	0.085	0.344	98	0.044	0.061	0.027	0.490
									0.038	0.601
5. 存貨週轉率(次)										
1	$0.00 < x \leq 0.35$	-54.04	189	0.075	0.249	139	0.063	0.029	0.002	0.473
2	$0.35 < x \leq 3.18$	-13.78	544	0.216	0.119	463	0.208	0.039	0.000	0.089
3	$3.18 < x \leq 6.13$	-0.62	641	0.254	0.092	550	0.247	0.042	0.000	0.039
4	$6.13 < x \leq 15.73$	25.05	663	0.263	0.054	705	0.317	0.020	0.010	0.034
5	$15.73 < x < \infty$	32.09	299	0.119	0.047	332	0.149	0.006	0.007	0.084
6	缺漏值	35.34	183	0.073	0.044	34	0.015	0.147	0.092	0.124
									0.111	0.844
6. 長期負債/淨值										
1	$x = 0.00$	23.94	860	0.341	0.042	787	0.354	0.009	0.000	0.051
2	$0.00 < x \leq 0.08$	24.08	455	0.181	0.042	406	0.183	0.025	0.000	0.009
3	$0.08 < x \leq 0.20$	2.04	435	0.173	0.085	385	0.173	0.026	0.000	0.070
4	$0.20 < x \leq 0.59$	-13.54	559	0.222	0.138	482	0.217	0.041	0.000	0.118
5	$0.59 < x < \infty$	-40.08	210	0.083	0.286	163	0.073	0.117	0.001	0.151
									0.002	0.398
7. 稅後資產報酬率(%)										
1	$-\infty < x \leq 0.00$	-13.68	669	0.266	0.193	636	0.286	0.066	0.001	0.136
2	$0.00 < x \leq 11.00$	3.88	1,339	0.532	0.072	1,193	0.537	0.020	0.000	0.067
3	$11.00 < x \leq \infty$	48.71	446	0.177	0.004	381	0.171	0.000	0.000	0.024
4	缺漏值	29.17	65	0.026	0.015	13	0.006	0.000	0.029	0.110
									0.031	0.336

附表 2 信用評分卡的第一種評等分布

評等	分等等界	建模樣本				驗證樣本				比較		
		預測 違約率	分等 戶數	樣本率	實際 違約率	檢定 統計量	分等 戶數	樣本率	實際 違約率	檢定 統計量	樣本率 PSI	違約率 PSI
<b>A 組信用評分卡</b>												
1	900 - 1,000	0.001	314	0.125	0.000	-	346	0.156	0.003	0.069	0.007	0.017
2	843 - 899	0.003	315	0.125	0.000	-	414	0.186	0.002	0.235	0.024	0.011
3	794 - 842	0.006	315	0.125	0.010	0.165	360	0.162	0.006	0.258	0.010	0.002
4	747 - 793	0.012	315	0.125	0.010	0.333	316	0.142	0.032	0.101	0.002	0.026
5	702 - 746	0.027	315	0.125	0.038	0.212	237	0.107	0.034	0.234	0.003	0.000
6	649 - 701	0.056	315	0.125	0.054	0.349	261	0.117	0.077	0.240	0.000	0.008
7	577 - 648	0.136	315	0.125	0.137	0.382	218	0.098	0.051	0.719	0.007	0.085
8	0 - 576	0.487	315	0.125	0.479	0.510	71	0.032	0.183	0.755	0.127	0.285
											0.180	0.434
<b>B 組信用評分卡</b>												
1	900 - 1,000	0.001	314	0.125	0.000	-	431	0.194	0.000	-	0.031	0.000
2	837 - 899	0.003	315	0.125	0.003	0.198	444	0.200	0.000	-	0.035	0.017
3	789 - 836	0.006	315	0.125	0.010	0.173	364	0.164	0.017	0.099	0.010	0.004
4	742 - 788	0.014	315	0.125	0.016	0.241	280	0.126	0.029	0.129	0.000	0.008
5	696 - 741	0.029	315	0.125	0.016	0.467	295	0.133	0.044	0.195	0.000	0.028
6	645 - 695	0.060	315	0.125	0.070	0.293	203	0.091	0.074	0.272	0.011	0.000
7	573 - 644	0.140	315	0.125	0.114	0.470	170	0.076	0.088	0.557	0.024	0.007
8	0 - 572	0.475	315	0.125	0.498	0.453	36	0.016	0.250	0.624	0.223	0.171
											0.333	0.235
<b>C 組信用評分卡</b>												
1	871 - 1,000	0.001	314	0.125	0.000	-	373	0.168	0.000	-	0.013	0.000
2	826 - 870	0.004	315	0.125	0.010	0.094	344	0.155	0.003	0.275	0.006	0.008
3	781 - 825	0.007	315	0.125	0.003	0.443	398	0.179	0.015	0.132	0.019	0.019
4	738 - 780	0.015	315	0.125	0.022	0.195	297	0.134	0.027	0.157	0.001	0.001
5	695 - 737	0.030	315	0.125	0.025	0.353	275	0.124	0.026	0.351	0.000	0.000
6	646 - 694	0.062	315	0.125	0.051	0.402	286	0.129	0.073	0.278	0.000	0.008
7	580 - 645	0.138	315	0.125	0.149	0.351	202	0.091	0.084	0.560	0.011	0.037
8	0 - 579	0.470	315	0.125	0.467	0.499	48	0.022	0.125	0.852	0.182	0.451
											0.232	0.525

附表 3 信用評分卡的第二種評等分布

評等	分等等界	建模樣本				驗證樣本				比較		
		預測 違約率	分等 戶數	樣本率	實際 違約率	檢定 統計量	分等 戶數	樣本率	實際 違約率	檢定 統計量	樣本率 PSI	違約率 PSI
A 組信用評分卡												
1	935 - 1,000	0.001	157	0.062	0.000	-	165	0.074	0.000	-	0.002	0.000
2	867 - 934	0.002	337	0.134	0.000	-	417	0.188	0.005	0.082	0.018	0.031
3	806 - 866	0.004	363	0.144	0.003	0.336	461	0.207	0.002	0.386	0.023	0.000
4	740 - 805	0.012	447	0.177	0.011	0.291	435	0.196	0.030	0.106	0.002	0.019
5	688 - 739	0.031	345	0.137	0.038	0.254	258	0.116	0.027	0.342	0.003	0.004
6	635 - 687	0.069	326	0.129	0.061	0.385	263	0.118	0.088	0.269	0.001	0.010
7	512 - 634	0.219	386	0.153	0.223	0.408	214	0.096	0.084	0.798	0.026	0.136
8	0 - 511	0.658	158	0.063	0.658	0.542	10	0.004	0.200	0.988	0.153	0.545
											0.229	0.745
B 組信用評分卡												
1	941 - 1,000	0.000	157	0.062	0.000	-	244	0.110	0.000	-	0.027	0.000
2	846 - 940	0.002	427	0.170	0.002	0.192	574	0.258	0.000	-	0.037	0.011
3	797 - 845	0.006	315	0.125	0.003	0.360	351	0.158	0.006	0.241	0.008	0.002
4	725 - 796	0.015	476	0.189	0.017	0.258	467	0.210	0.036	0.111	0.002	0.014
5	696 - 724	0.032	199	0.079	0.020	0.443	178	0.080	0.045	0.224	0.000	0.020
6	627 - 695	0.071	426	0.169	0.063	0.383	272	0.122	0.077	0.317	0.015	0.003
7	519 - 626	0.230	361	0.143	0.241	0.399	132	0.059	0.129	0.687	0.074	0.070
8	0 - 518	0.629	158	0.063	0.639	0.517	5	0.002	0.200	0.983	0.201	0.510
											0.364	0.630
C 組信用評分卡												
1	903 - 1,000	0.001	157	0.062	0.000	-	190	0.085	0.000	-	0.007	0.000
2	833 - 902	0.003	421	0.167	0.002	0.259	472	0.212	0.002	0.279	0.011	0.000
3	781 - 832	0.007	366	0.145	0.008	0.226	453	0.204	0.013	0.143	0.020	0.002
4	744 - 780	0.014	273	0.108	0.018	0.223	265	0.119	0.019	0.217	0.001	0.000
5	686 - 743	0.031	419	0.166	0.024	0.382	356	0.160	0.037	0.267	0.000	0.006
6	638 - 685	0.070	301	0.119	0.066	0.363	278	0.125	0.076	0.320	0.000	0.001
7	516 - 637	0.209	424	0.168	0.210	0.412	204	0.092	0.093	0.748	0.046	0.095
8	0 - 515	0.624	158	0.063	0.639	0.505	5	0.002	0.200	0.982	0.201	0.510
											0.287	0.615

## 參考文獻

- 丁正中 (2004), 「消費金融信用風險研究—信用評分概述」, 《信用資訊》, 4, 1–32。
- 阮正治與江景清 (2004), 「台灣企業信用評分模型建置與驗證」, 《信用資訊》, 4, 1–16。
- 陳業寧, 王衍智與許鴻英 (2004), 「台灣企業財務危機之預測—信用評分法與選擇權評價法孰優?」, 《風險管理學報》, 6, 155–179。
- 鍾經樊 (2005), 「評分系統的驗證」, 未發表論文。
- Altman, E. I. (1968), “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *Journal of Finance*, 123, 589–609.
- Altman, E. I. and P. Narayanan (1997), “An International Survey of Business Failure Classification Models,” *Financial Markets, Institutions and Instruments*, 6, 1–57.
- Bailey, M. (2004a), *Credit Scoring: The Principles and Practicalities*, 2nd Edition, London: White Box Publishing.
- Bailey, M. (2004b), *Consumer Credit Quality: Underwriting, Scoring, Fraud Prevention and Collections*, London: White Box Publishing.
- Basel Committee on Banking Supervision (2005), “Studies on the Validation of Internal Rating Systems,” *Working Paper*, No. 14.
- Bluhm, C., L. Overbeck, and C. Wagner (2003), *An Introduction to Credit Risk Modeling*, Boca Raton, FL: Chapman and Hall/CRC.
- Boyes, W. J., D. L. Hoffman, and S. A. Low (1989), “An Econometric Analysis of the Bank Credit Scoring Problem,” *Journal of Econometrics*, 40, 3–14.
- Cover, T. M. and J. A. Thomas (1991), *Elements of Information Theory*, New York: Wiley.
- de Servigny, A. and O. Renault (2004) *Measuring and Managing Credit Risk*, New York: McGraw-Hill.
- Dwyer, D. W. (2005), “Examples of Overfitting Encountered When Building Private Firm Default Prediction Models,” one report from Moody’s KMV, New York.
- Falkenstein, E., L. V. Carty, and A. Boral (2000), “RiskCalc for Private Companies: Moody’s Default Model,” one report from Moody’s Investors Service, New York.
- Hanley, J. A. and B. J. McNeil (1982), “The Meaning and Use of the Area under an ROC

- Curve,” *Radiology*, 143, 29–36.
- Hosmer, D. W. and S. Lemeshow (2000), *Applied Logistic Regression*, 2nd Edition, New York: Wiley.
- Jappelli, T. and M. Pagano (1993), “Information Sharing in Credit Markets,” *Journal of Finance*, 48, 1693–1718.
- Jappelli, T. and M. Pagano (2002), “Information Sharing, Lending and Defaults: Cross-Country Evidence,” *Journal of Banking and Finance*, 26, 2023–2054.
- Klein, D. B. (1992), “Promise Keeping in the Great Society: A Model of Credit Information Sharing,” *Economics and Politics*, 4, 117–136.
- Mays, E. (1998), *Credit Risk Modeling: Design and Application*, New York: AMACOM.
- Mays, E. (2001), *Handbook of Credit Scoring*, New York: AMACOM.
- Mester, L. J. (1997), “What’s the Point of Credit Scoring?” *Business Review*, September, 3–16.
- Ohlson, J. M. (1980), “Financial Ratios and the Probabilistic Prediction of Bankruptcy,” *Journal of Accounting Research*, 18, 109–131.
- Padilla, A. J. and M. Pagano (1997), “Endogenous Communication among Lenders and Entrepreneurial Incentives,” *Review of Financial Studies*, 10, 205–236.
- Padilla, A. J. and M. Pagano (2000), “Sharing Default Information as a Borrower Discipline Device,” *European Economic Review*, 44, 1951–1980.
- Pepe, M. S. (2003), *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford: Oxford University Press.
- Roszbach, K. (2004), “Bank Lending Policy, Credit Scoring, and the Survival of Loans,” *Review of Economics and Statistics*, 86, 946–958.
- Sobehart, J. R., S. C. Keenan, and R. M. Stein (2000), “Benchmarking Quantitative Default Risk Models: A Validation Methodology,” one report from Moody’s Investors Service, New York.
- Swets, J. (1988), “Measuring the Accuracy of Diagnostic Systems,” *Science*, 240, 1285–1293.
- Thomas, L. C., D. B. Edelman, and J. N. Crook (2002), *Credit Scoring and Its Applications*, Philadelphia, PA: SIAM.
- Vercammen, J. A. (1995), “Credit Bureau Policy and Sustainable Reputation Effects in Credit Markets,” *Economica*, 62, 461–478.



## DEVELOPING, VALIDATING, AND COMPARING CREDIT SCORING SYSTEMS IN TAIWAN

**Ching-Fan Chung**\*

Institute of Economics  
Academia Sinica,  
Department of Economics  
National Chengchi University,  
and Institute of Economics  
National Sun Yat-Sen University

**Jia-Long Huang**

Department of Economics  
National Taiwan University

**Bor-Yi Huang**

Department of Finance and Banking  
Shih Chien University

**Yiu-Long Hsieh**

Department of Economics  
National Chengchi University

**Keywords:** Credit scoring, Statistical validation, Variable transformation and selection, Credit scorecard development and validation

**JEL classification:** C50, C53, G32

---

\* Correspondence: Ching-Fan Chung, Institute of Economics, Academia Sinica, Nankang, Taipei 115, Taiwan. Tel: (02) 2782-2791 ext. 188; Fax: (02) 2785-3946; E-mail: cfchung@econ.sinica.edu.tw.

## ABSTRACT

*Using financial report data of thousands of public-issued companies over three years drawn from the TEJ (Taiwan Economic Journal) database, we develop credit scorecards for business lending based on the estimates of a Logit regression model. This paper emphasizes that the in-depth analysis and transformation of the original data before the estimation as well as testing and validation of the estimated model are much more important in developing credit scorecards than in a typical empirical econometric analysis. In particular, reserving a large percentage of data for the validation purpose is absolutely necessary when developing credit scorecards. Using two additional years of data as the validation sample, we show that our credit scorecards perform much better than the basic credit rating provided by the TEJ in terms of several performance statistics. Before implementing the new Basel Capital Accord at the end of 2006, this paper presents a timely contribution to both the banks which intend to adopt internal rating based approach and the governmental agencies which regulate the banking industry in managing credit risks.*