

微、小型企業信用評分模型

丁正中 / 金融聯合徵信中心風險研究組

微、小型企業信用評分模型(以下簡稱本模型)已於去(97)年底完成模型建構，並於聯徵中心外部研究平台提供會員機構使用。本模型的評分範圍主要為「資本額」與「過去12個月授信總餘額最大值」均小於新台幣三千萬之公司組織，其中，依據風險行為的不同，本模型又區分為「有延遲還款記錄」、「無延遲還款記錄—製造業」與「無延遲還款記錄—非製造業」三個區隔。回顧本模型的研發過程中，有些對於會員機構或後續研發企業信用評分模型時，頗值得參考的部分，本文將其彙整並做說明，以增進會員機構對本模型的瞭解。

變數產生、處理與篩選流程的標準化與自動化

利用資料倉儲(Data Warehouse)累積長期且大量的歷史資料，提供分析資料來源已是相當普遍的做法，但要將成千上萬的資料，加工後產生變數，大致上有兩種方法，一種方法是先設定主題再挑選相關資料欄位，產生特定變數，另一種是設定一些標準後，經由自動化的方式衍生相關變數。前者優點為能集中分析焦點，節省資料處理的成本，缺點是設定主題與挑選欄位時，人為操控成分較高，且產生變數時僅使用資料庫的部分資料，變數缺乏完整性；後者優點為產生變數時，可涵蓋較完整的資料，並透過標準化與自動化的流程，產生一系統性的架構，以減少人為操控產生的錯

誤，但是缺點則是必須付出較多資料處理的成本，且當變數產生過多時，將會增加後續挑選變數的複雜度，另外，此方法亦無法顧及特殊狀況的處理。

本模型在變數設計上，採用後者方法，希望藉由標準化與自動化的流程，效率性及系統性地產生變數，使變數涵蓋性較為完整，而針對此方法的缺點，本模型則利用變數篩選流程的設計，以及學者專家意見的諮詢，加以補強。以下本文將依序說明本模型在變數產生與處理，以及變數篩選與替換時的處理方法：

一、變數產生與處理

產生變數之前，首先必須要瞭解資料倉儲上的詮釋資料(Metadata)，以快速掌握與

資料相關的資訊，例如：資料來源、資料欄位定義、更新週期、是否經過運算或轉換等。其次，決定產生變數的類型，使變數產生過程中能顧及各種面向，本模型經過多次內部會議討論及專家諮詢後，決定下列八大變數類型：

1. 企業基本資料變數：係指該企業的基本資料，如：成立至今月數、資本額。
2. 繳款行為類變數：係指在觀察期間內，企業授信戶對於各類授信貸款以及票據信用行為之還款情況，如：延遲還款之嚴重程度、延遲還款發生之次數等。
3. 負債總額類變數：評估該企業授信戶之負債程度，包括負債總餘額、各類負債之額度使用率等。
4. 負債型態類變數：除評估企業授信戶之負債程度外，另考量不同授信科目下的借款情形，如：無擔保之授信餘額等。
5. 負債變動類變數：係評估企業授信戶之負債變動程度。
6. 信用歷史長度類變數：係指在觀察期間內，企業授信戶之信用歷史長度，如：最早有授信未逾期餘額距今的月數等。
7. 信用往來情況類變數：係指在觀察期間內，企業授信戶的信用情形。如：授信往來家數等。
8. 新信用查詢紀錄變數：係指在觀察期間內，銀行查詢企業授信戶的信用狀況。如：新業務查詢次數等。

最後，為使變數有一致性的處理標準及衍生原則，例如：缺失值處理、函數計算等，本模型乃根據不同的變數單位，設定不同的變數處理原則，並將相同來源的變數，透過不同期

間及函數的組合，產生同一系列的衍生變數。變數產生後，分類存放於各資料超市（Data Mart），方便分析人員做後續分析。

二、變數之篩選與替換

經過標準化與自動化流程產生的變數，雖較為完整，但也增加後續挑選變數的複雜度，此時，變數篩選的設計就顯現出其重要性，本模型在此步驟中，採取「分階段篩選方式」，除可逐步縮小變數個數，以避免誤刪的情況發生，也可將合理性及替換性納入篩選流程中。目前本模型做法大致如下：首先，計算每一變數對於目標變數（違約與否）的區隔力，並據此設定篩選條件，排除不符合該條件之變數，其次，針對同一類型之衍生變數，選取區隔能力較佳的變數，並依據本模型所設定的八大面向，在每一類型下各自進行因素分析與篩選，另外，進行跨期間變數區隔能力的測試，以確認變數的穩定性，完成備選變數清單（約80個變數）的建立。而後，經由內外部專家的諮詢，逐一檢視變數的合理性，並選定變數較難以直覺解釋的部分，蒐集專家意見後，嘗試產生新的變數，以彌補自動化產生變數的缺點。最後，根據備選變數清單，嘗試利用統計的分析，以及同類型變數的替換，完成評分模型的建置。

值得一提的是，雖然本模型在設計時認為，八大變數類型都要放入一個以上的變數，但是，本模型的結果顯示，信用歷史長度類變數雖可勉強放入模型，但因模型已具相當程度的區隔能力，此變數對於模型區隔能力的提升並不明顯，又經內部研究後認為，該類變數與

企業成立至今的月數有相當程度的關聯，故決定暫不將此類變數放入模型，而是列入後續每月監控的變數清單中，以觀察此類變數在未來是否有區隔能力提升的情況。

企業負責人與保證人信用變數的處理

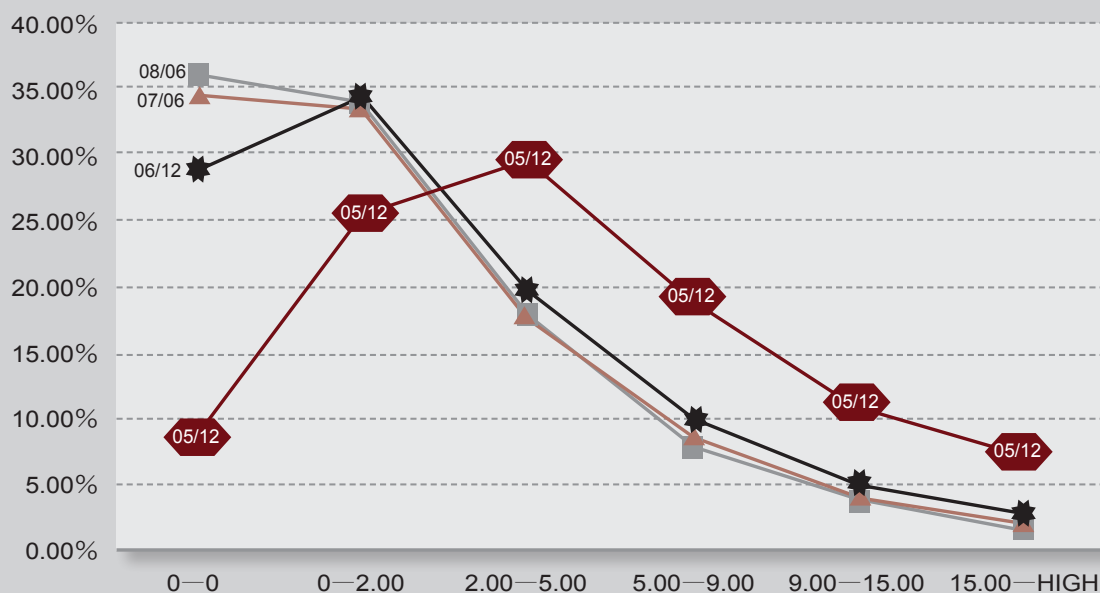
請參照本刊物97年10月專題報導〈微、小型企業評分模型對負責人信用表現之處理方式〉。

加強變數與模型之穩定性

評分模型的假設為未來的狀況與過去相類似，可利用歷史資料建構的模型套用到現在，但實際狀況是，外在狀況不可能都不改變，變

數在不同時間點也會有所變化，只是變化的程度是否大到使模型失效。本模型在變數篩選流程中，特別加入變數跨期區隔能力的測試，希望藉此篩選出相對穩定的模型變數，另外，對於區隔能力強，但跨期區隔能力穩定性不佳的變數，本模型亦嘗試透過一些方法（如：組合其他類型的變數），來提升變數的穩定性。以過去12個月新業務被查詢次數為例，經多期觀察該變數的分配（詳圖一）可知，在2005年12月之後，該變數出現結構性的改變，但因該變數為本模型區隔能力較強的變數，分配的改變將對評分結果產生重大影響，因此，本模型乃決定以2006年12月的資料為基礎，並嘗試加入授信資料，進行組合變數的創造，以增加該變數的穩定性。

圖一 新業務被查詢次數分配比較



表一 微、小型企業評分模型結果

資料年/月	總樣本數	違約率	AUC	KS	PSI
2006/12	83098	4.07%	0.8636	54.14	-
2007/03	81864	3.74%	0.8618	54.17	0.0023
2007/06	80314	3.50%	0.8560	53.15	0.0052
2007/09	78925	3.34 %	0.8447	51.67	0.0091
2008/09	73190	-	-	-	0.0430

在模型完成之後，聯徵中心進行建模時點外測試（Out- of-time Validation），以確認模型不會因為評分時間不同而失效，目前本模型在不同期間的測試結果如表一所示，由區隔能力指標KS值（Kolmogorov-Smirnov Statistic）及AUC值¹（Area Under Curve）可知，本模型的區隔能力僅出現微幅下降的情況，而由穩定性指標PSI（Population Stability Index）顯示，本模型評分分配的變動仍在可接受的範圍之內。由上可知，本模型透過變數與模型之跨期穩定性分析，確認模型在不同的評分時間時，仍可維持正常運作。

結語

運用標準化與自動化的流程進行變數的產生，可將建模方法紀律化，以減少人為操作

的失誤，並可減少分析人員計算統計數據的負擔，而透過各面向變數的篩選及專家諮詢，可強化變數的完整性與合理性。本模型雖未放入企業負責人相關信用表現，但是，聯徵中心仍強調負責人的好壞對於微、小企業的信用表現具有相當程度的影響力，因此，會員機構在未來應用評分結果時，應同時參考企業負責人相關資料，再做決策判斷。而聯徵中心除了會針對現有的微、小型企業信用評分模型進行持續監控外，未來將嘗試針對部分群體進行分析，例如：特殊產業，以確認模型的通用性。

本模型目前僅於外部研究平台提供服務，未來將與中型企業模型整合，提供會員機構在評估未公開發行公司的信用表現時，有外部參考與比較的指標。

¹ AUC值為Receiver Operating Characteristic 曲線下之面積。