

# 結合變數篩選之最小平方支援向量機於財務危機預警模型之研究

葉清江

國立台北商業技術學院商學研究所

[yhinc@webmail.ntcb.edu.tw](mailto:yhinc@webmail.ntcb.edu.tw)

齊德彰

中國文化大學會計學系暨研究所

[chi58882000@yahoo.com.tw](mailto:chi58882000@yahoo.com.tw)

何秉劼

中國文化大學會計學系暨研究所

[stare7500@hotmail.com](mailto:stare7500@hotmail.com)

## 摘要

本研究嘗試以傳統的區別分析與 t 檢定結合較新之人工智慧演算法「最小平方支援向量機」來建立一個二階段的財務危機預警模型，並探討不同變數篩選方法與支援向量機所建立之模型效率差異。本研究發現，以 t-test 來進行變數篩選將有助於提升模型整體準確率，而最小平方支援向量機也較支援向量機優越。

關鍵字:財務危機、資料探勘、支援向量機、最小平方支援向量機、區別分析、t 檢定

## 壹、緒論

歷年來，許多大型企業所發生的財務危機事件，都讓社會付出高額的社會成本。因此學術與實務界無不積極想要建立一個成功的財務危機預警模型，來達到早期發現，及早因應，將損失降低到最小的防範機制。

財務危機之預警模型早從 1960 年代中期開始就有學者以傳統統計方法來研究，但傳統的分析方法受限於一些假設，像是自變數共線性問題以及殘差項存在自我相關等。因此近年來發展出許多沒有傳統統計方法建構模式時所需滿足的假設條件的人工智慧演算法來建立財務危機預警模型，如決策樹(Tam 和 Kiang, 1992)、類神經網路(Shah 和 Murtaza, 2000)和支援向量機(曾淑峰, 2008)等，其中支援向量機(Support Vector Machine, SVM)被認為是用來作為二元分類的良好方法(Pang-Ning Tan, 2005)。而在 1999 年 Suykens 等人基於支援向量機的演算法，導入最小平方失函數的概念來考慮訓練誤差，發展出最小平方支援向量機(Least Squares Support Vector Machine, LS-SVM)被認為有優於支援向量機的表現，但這部分的實證尚未有學者探討過。另外在建立財務危機預警模型的過程中，無論是人工智慧演算法或是傳統的統計方法，都需要花費大量的時間與人力來分析龐大的資料，當有過多的非攸關資訊時，會干擾模型準確率與解釋能力。因此，如何篩選出非攸關的資訊就變成了建立財務危機預警模型一項很重要的課題(Tsai, 2009)。綜上所述，本研究除了探討支援向量機與最小平方支援向量機所建立之財務危機預警模型的績效差異外，還以之前較多學者用來做為變數篩選之方法，區別分析與 t-test。來建立一個二階段財務危機模型，並探討不同的變數篩選搭配 SVM 以及 LS-SVM 所產生的績效差異。

本研究能提供，以支援向量機與不同的變數篩選方法來建立財務危機預警模型的一個基礎模型，作為日後建立財務危機預警模型的一項參考模型。本篇研究剩下的部分內容如下，第二部分將進行相關文獻的探討，第三部分將解說研究方法，第四部份探討實證結果，第五部分出具結論。

## 貳、文獻探討

### (一) 企業危機預警模型

預測企業財務危機發生與否的預警模型，最早的單變量分析至人工智慧演算法的今天，已經有眾多的模型產生。隨著新的人工智慧演算法的演進，與經濟環境不斷的變動，建構財務危機預警模型的研究亦隨之日益更新與改進。目前預警模型的相關研究如表一。

表一 國內外財務危機預警模型之相關研究

學者	變數選擇	樣本選取	研究方法	研究結果
Beaver(1966)	以以下六類 30 項財務比率： 一、 現金流量比率 二、 淨利比率	1954 年至 1964 年間共 79 家發生財務危機的公司。	單變量分析	最能預測財務危機的比率依序為： 一、 現金流量對總負債比率 二、 稅後淨利對總資產比率

	<p>三、 負債對總之產比率</p> <p>四、 流動資產對總負債比率</p> <p>五、 流動資產對流動負債</p> <p>六、 周轉率</p>			<p>三、 總負債對總資產比率。</p> <p>其中以危機發生前一年預測能力最高，為 87%。</p>
Altman(1968)	<p>22 個財務比率變數，分為五大類，分別為:流動性、獲利性、財務槓桿、償債能力及周轉能力。</p>	<p>美國 1946 至 1965 年間，33 家財務危機,33 家配對樣本，共 66 家樣本數。</p>	<p>以逐步迴歸將 22 財務比率篩選出五種財務比率指標變數，並構成區別函數，最後將此比率結合成綜合性指標，創造出 Z-score 模式。</p>	<p>預測財務危機前一年有良好的準確率，高達 95%，但是長期準確率不佳。</p>
Altman(1977)	<p>22 個財務比率變數，分為五大類，分別為:流動性、獲利性、財務槓桿、償債能力及周轉能力。</p>	<p>美國 1969 至 1975 年間，53 家財務危機,53 配對樣本共 106 家樣本數。</p>	<p>重建 Z-score 模型，限制樣本總資產在 20~100 萬美元，並由原始只關注特定產業再加入零售企業。</p>	<p>破產前一年準確率提高至 96%。</p>
Blum(1974)	<p>以現金流量的流動性、獲利性及變異性作為財務指標。</p>	<p>美國 1954 至 1968 年間，115 家財務危機發生前五年的財務資料，並配對樣本 115 家，共 230 家。</p>	<p>區別分析</p>	<p>以現金流量對總負債的比、淨值對總負債的比和速動資產淨值對存貨的比，三項變數最具區別能力。至前五年準確率可達 70%。</p>
Deakin(1972)	<p>以以下六類 30 項財務比率:</p> <p>一、 現金流量比率</p> <p>二、 淨利比率</p> <p>三、 負債對總之產比率</p> <p>四、 流動資產對總負債比率</p> <p>五、 流動資產對流動負債</p> <p>六、 周轉率</p>	<p>美國 1964 至 1970 年間，32 家財務危機發生前五年的財務資料，並隨機配對樣本 32 家，共 64 家。</p>	<p>區別分析</p>	<p>財務危機前兩年預測能力最佳，準確率可達 91%。</p>
Ohlson(1980)	<p>財務比率。</p>	<p>美國 1970 年至 1976 年間 105 家破產企業，並以 2058 家非財務危機企業為配對，總共 2163 家樣本。</p>	<p>邏輯斯回歸</p>	<p>其建立的模型之準確立，危機前一年為 96.12%、前兩年為 95.55%、前三年為 92.84%。</p>
Mensah,(1984)	<p>財務比率</p> <p>總體經濟指標</p>	<p>1972 年一月至 1980 年六月間，採礦、製造、零售</p>	<p>羅吉斯回歸</p>	<p>考慮經濟環境對企業發生財務危機的影響，並認為在不同的</p>

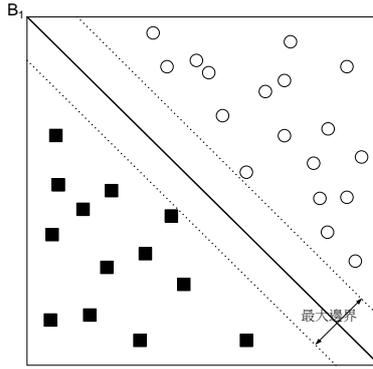
		和營建業共 110 組樣本。配對以資產和產業別為依據。		經濟環境下，應搭配不同的財務危機模型才有較好的準確率。
Odom 和 Sharda(1990)	比照 Altman(1968)22 個財務比率變數，分為五大類，分別為：流動性、獲利性、財務槓桿、償債能力及周轉能力。	1975 年至 1982 年美國 65 家危機公司，並配對 65 家正常公司為樣本共 130 家。	類神經網路 區別分析法	類神經網路準確率高於區別分析法。
Altman, Marco 和 Varetto(1994)	比照 Altman(1968)22 個財務比率變數，分為五大類，分別為：流動性、獲利性、財務槓桿、償債能力及周轉能力。	義大利 1982 年到 1992 年間，超過一千家，健康、財務困難的公司。	區別分析 羅吉斯回歸 類神經網路	三種方法所預測的準確率並無太大差異，不過作者特別指出類神經網路演算法的缺點，就是過度學習與變數加權沒有邏輯性可循。
邱志洲，簡德年和高凌菁(2003)	27 個財務指標變數和 7 個智慧資本指標變數	台灣 1997 年至 2000 年 35 家財務危機公司樣本，配對 35 家正常公司，共 70 家樣本。	區別分析結合類神經網路	發現財務危機除了受到傳統財務指標影響外，也受到智慧資本構面影響，並建立出結合區別分析的倒傳類神經網路，其準確率(88%)比單使用區別分析(84%)和倒傳類神經網路(80%)為高。
李天行和唐筱菁(2004)	財務比率和智慧資本指標變數	台灣 1998 年至 2000 年上市公司中 35 家財務危機公司為樣本，配對 70 家正常公司，共 105 家樣本。	多元適應性雲行回歸(multivariateadaptive regressionspines)結合類神經網路。	多元適應性雲行回歸(multivariateadaptive regressionspines)結合類神經網路其建模時間和準確率皆有提高。
林淑萍等人(2007)	財務比率	台灣 2000 年到 2004 年期末發生危機公司共 58 家，再依產業別隨機抽樣來選擇配對公司，以 1:3 的比例選出 174 家正常公司。	羅吉斯回歸。 資料包絡分析法(DataEnvelopmentAnalysis-DiscriminantAnalysis,DEA-DA)。 類神經網路。	DEA-DA 所建立的預警制度模型比類神經網路及邏輯斯回歸的模型有較佳的表現。
劉邦典，梁榮輝(2007)	5 項公司治理變數和 7 項經編碼後之會計師意見作為輸入變數。	取 20 家財務危機公司，以一比一配對方式選取 20 家正常營運公司，共 40 家公司樣本。	區別分析法和二元 Logistic 迴歸分析	公司財務危機發生前一季，二元 Logistic 迴歸預測模型之型二分類錯誤率優於區別分析法方法，最能有效地幫助投資人規避風險。
曾淑峰(2008)	財務狀況指標變數。	案例基層金融機構放款客	結合基因演算法之支	GA-SVM 所建立之財務危機預

	逾放情形指標變數。 帳戶情形指標變數。 逾催情形指標變數。 會計師意見指標變數。 獲利能力指標變數。 現金流量指標變數。	戶，共 323 家破產公司。 706 家無財務危機之公司。	援向量機(GA-SVM)。	警模型，優於 SVM、羅吉斯回歸與類神經網路之模型。
Tsai(2009)	資料庫中的變數	資料庫中的案例	多層感知器類神經網路以及以 t-test 和 stepwise, correlation matrix, 主成分分析與因素分析來進行變數篩選。	以 t-test 先做變數篩選有再進行建模會有較佳之準確率。
Yang 和 Da-wei(2010)	財務比率	中國上海證券交易所上中 2008 年至 2009 年之間，106 組上市公司為樣本。	約略集合理論 (roughsettheory,RST) 結合支援向量機 (supportvectormachine ,SVM)	結合約略集合理論的支援向量機有較佳的預測能力，另外發現支援向量機在財務預測上面優於類神經網路。
賴鈺城，李崑進，和李善玉 (2010)	財務比率與公司治理變數。	1996 年至 2008 年台灣電子業上市(櫃)公司。	邏吉斯回歸	加入公司治理變數後，預測模型準確率(94.1%)較只有財務比率(86.3%)的模型還要來得高。

資料來源:本研究整理

## (二) 支援向量機與最小平方法支援向量機

支援向量機是由(Vapnik,1995)研究團隊所發展出的一套人工智慧學習方法，其理論架構是以統計理論基礎並依循結構風險最小化原則(Structural Risk Minimization)，以最小化預測誤差的上界為目標之結構化風險。其功能有分類與預測，本研究因只用到分類功能，因此只做分類演算法的解說。分類問題上又分線性可分與非線性可分的分類方法，線性的支援向量機(線性可分)是尋找擁有最大邊界超平面的方法之一。圖一為最大邊界超平面示意圖，圖中黑色方塊為一分類群組，圓圈為另一分類群組，支援向量就是要找出中間 B1 黑色實線，其邊界(離兩邊群組最近之樣本)為最大。此黑色實線的決策函數為  $f(x) = x_i \cdot \omega + b$ ， $\omega$  為分類平面法向量;  $b$  為偏移值。則尋找最大間隔的限制式為式(1)。



圖一 最大邊界超平面式意圖

$$\text{Minimize } R(\omega, b) = \frac{1}{2} \|\omega\|^2$$

$$\text{s.t. } y_i((x_i \cdot \omega) + b) \geq 1 \quad i = 1, \dots, n \quad (1)$$

我們可以藉由 Lagrange 最佳化方法進行有限制式的二次規劃問題來求解。我們將 (3.1) 導入 Lagrange 乘子後可改寫為 Lagrange 函數為式(2)。

$$L(\omega, b; \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^n \alpha_i (y_i((\omega \cdot x_i) + b)) - \sum_{i=1}^n \alpha_i \quad (2)$$

(2) 式中  $\alpha_i$  為 Lagrange 乘子，對函數的各變數求偏微分，並令微分為 0，得到方程式 (3)。

$$\partial_{\omega} L = 0 \Rightarrow \omega = \sum_{i=1}^n \alpha_i y_i x_i$$

$$\partial_b L = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \quad (3)$$

將式(3.3)帶入式(3.2)，可轉換為 Lagrange 對偶型態函數如式(4)所示。

$$\text{Maximize } L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0 \quad i = 1, \dots, n \quad (4)$$

則  $\alpha^* = (\alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*, \dots, \alpha_n^*)^T$ ，將  $\alpha^*$  代入  $\omega^* = \sum_{i=1}^n \alpha_i^* y_i x_i$  與  $\sum_{i=1}^n \alpha_i^* [y_i(\omega \cdot x_i) + b - 1] = 0$  可求得  $\omega^*$  以及  $b^*$  值。最後，支援向量機支線性可分之問題決策函數可寫成式(5)

$$f(x) = \sum_{i=1}^n \alpha_i y_i (x \cdot x_i) + b \quad (5)$$

而遇到非線性可分的時候，可以以核函數(Kernel Function)來將原始資料轉換成較高維度特徵空間進行線性劃分，轉換為非線性問題時，支援向量機之決策函數可改寫為式(6)。核函數能將資料分布的原始空間，過一個非映射函數轉換至較高維度的特徵空間進行線性劃分。而現行常用之核函數包括，線性核函數、多項式和函數以及徑向基和函數(Radial Basis Function, RBF)。

$$\text{Maximize } L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j \varphi(x_i) \cdot \varphi(x_j)$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0 \quad i = 1, \dots, n$$

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x \cdot x_i) + b \quad (6)$$

而在 2001 年 Suykens(2001)等人以原始支援向量機導入最小平方失函數(Least Squares Loss Function)的概念，來考慮訓練誤差，其跟 SVM 最大的差異就是，SVM 只針對支援向量進行計算，而 LS-SVM 是將所有訓練樣本皆納入計算並且在求解過程等式情況下，可轉換成線性形式計算，SVM 卻只能透過複雜的對偶問題進行求解。因此大大提升了求解的速度。使其在商業的應用上更加有效率。最小平方支援向量機分類問題的決策函數如式(7)所示。

$$\begin{bmatrix} 0 & Y^T \\ y & ZZ^T + C^{-1}I \end{bmatrix} \cdot \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j) = y_i y_j K(x_i, x_j) \quad i = 1, \dots, n$$

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b \quad (7)$$

### (三) 區別分析與 t 檢定

傳統統計方法常使用區別分析來找出預測變數的線性組合，其可以判別那些變數對於兩組或兩組以上的樣本具有判別力，其數學原理乃是運用其組間差異平方和相對於組內插一平方和之比值為最大，但每一個線性組合與先前已經獲得的線性組合均不相關。其運作的程序乃是先檢定各組重心是否有差異再找出那些預測變數具有最大的區別能力，最後再根據新受測者預測變數的數值，將該受測者分派到其所應隸屬的群體。

t-test 可以判別兩組母體中對某一自變數是否存在著差異，本研究的兩個母體分別為，財務危機公司和非財務危機公司。因此只要找出本研究所使用之自變數在兩個母體中之平均數有差異的，就可用來做為判別有無發生財務危機的特徵變數。

## 參、研究方法

本研究先以 SVM 與 LS-SVM 建立未做變數篩選之財務危機預警模型，再以區別分析與 t-test 先做變數篩選後，再建立財務危機預警模型。然後比較各模型間的績效差異。

### 一、財務危機操作型定義

根據 Beaver(1966)所定義的企業財務危機，乃指一企業宣告破產、公司債務違約、未支付優先股股息時、銀行透支，就視為企業發生財務危機。而 Deaklin(1972)指出無償債能力、或清算的廠商則視為財務發生危機。Altman(1994)則指出，當企業投資報酬率低於其資金成本、無法償還到期的債務以及企業淨資產為負為財務危機公司。而賴鈺城(2010)等人，則認為公司發生，跳票擠兌、紓困及繼續經營疑慮時則為財務危機發生。因此以往學者都認為當企業償債能力，現金流動和獲利能力發生問題時，就會產生財務危機。此觀點與台灣經濟新報資料庫所整理的財務危機分類事件一致，因此本研究就以台灣經濟新報資料庫中的，企業在樣本選取期間，有被記錄發生跳票擠兌、倒閉破產、繼續經營疑慮、重整、紓困財危、接管、全額下市、財務吃緊停工、淨值為負之事件為公司發生財務危機之定義。

### 二、建立模型的變數解釋

本研究以台灣經濟資料庫所認定財務危機事件為定義，辨別公司是否為財務危機，如是則設虛擬變數為 1，否則設 0。而自變數主要國內外學者較長提出的變數為自變數。

變數定義與來源請見表三。

表三變數列表

指標構面	變數名稱	計算公式	文獻來源
經營能力	總資產週轉率(次)	營業收入淨額/平均資產總額	邱志洲等(2003)、(林淑萍,黃勁彥,&蔡昆霖,2007)、邱登裕(2007)、林豐騰(2009)
	應收帳款週轉率(次)	營業收入淨額/平均(應收帳款及票據+應收票據貼現)	邱志洲等(2003)、(李天行&唐筱菁,2004)、邱登裕(2007)、(林淑萍etal.,2007)、羅淑娟等(2009)、林豐騰(2009)、羅聖雅(2010)
	存貨週轉率(次)	營業成本/平均存貨	邱志洲等(2003)、(李天行&唐筱菁,2004)、邱登裕(2007)、(林淑萍etal.,2007)、羅淑娟等(2009)、林豐騰(2009)
	固定資產週轉率(次)	營業收入淨額/平均固定資產	邱志洲等(2003)、邱登裕(2007)
償債能力	流動比率	流動資產/流動負債*100	Beaver(1966)、Altman(1977)、邱志洲等(2003)、邱登裕(2007)、(林淑萍etal.,2007)、羅淑娟等(2009)、林豐騰(2009)、羅聖雅(2010)
	速動比率	(流動資產-存貨-預付款項-其他流動資產)/流動負債*100	Beaver(1966)、邱志洲等(2003)、(李天行&唐筱菁,2004)、邱登裕(2007)、羅淑娟等(林淑萍etal.,2007)、(2009)、林豐騰(2009)、
	負債比率	負債總額/資產總額*100	Beaver(1966)、邱志洲等(2003)、(李天行&唐筱菁,2004)、邱登裕(2007)、(林淑萍etal.,2007)、羅淑娟等(2009)、林豐騰(2009)、羅聖雅(2010)
	長期資金適合度	(淨值+長期負債)/固定資產*100	邱志洲等(2003)、(李天行&唐筱菁,2004)
	利息保障倍数	所得稅及利息費用前純益/本期利息支出	邱志洲等(2003)、(李天行&唐筱菁,2004)、邱登裕(2007)、(林淑萍etal.,2007)、羅淑娟等(2009)、羅聖雅(2010)
獲利能力	資產報酬率(A)%	(稅後淨利+利息支出*(1-25%))/平均資產總額*100	邱志洲等(2003)、(林淑萍etal.,2007)
	淨值報酬率—稅後	稅後淨利/平均淨值*100	邱志洲等(2003)、(林淑萍etal.,2007)
	營業利益率	營業利益/營業收入淨額*100	邱志洲等(2003)、(李天行&唐筱菁,2004)、(林淑萍etal.,2007)、羅淑娟等(2009)、
	每股盈餘	按當期之加權平均股數計算	邱志洲等(2003)、(林淑萍etal.,2007)、羅淑娟等(2009)、林豐騰(2009)
智慧資本指標	每人營收	營業收入/員工人數	邱志洲等(2003)、林豐騰(2009)
	每人營業利益	營業利益/員工人數	邱志洲等(2003)、林豐騰(2009)
	每人配備率	固定資產/員工人數	邱志洲等(2003)、李天行等(2004)、林豐騰(2009)

公司 治理 指標	董監持股率	全體董監事持股股數/公司發行在外 總股數	邱登裕等(2007)、羅淑娟等(2009)、林豐騰(2009)、賴鈺城等 (2010)
	董監質押比率	全體董監事質押股數/全體董監事持 股股數	邱登裕等(2007)、羅淑娟等(2009)、林豐騰(2009)、賴鈺城等 (2010)
	大股東持股比率	該公司之主管機關定義的大股東， 其持股佔公司總發行股份的比例。	林豐騰(2009)、邱登裕等(2007)、賴鈺城等(2010)

資料來源:本研究整理

### 三、樣本來源

本篇研究樣本以 2001 年至 2011 年底台灣非金融業<sup>1</sup>之上市及上櫃公司有發生在台灣經濟新報資料庫所記錄的財務危機事件之公司為樣本，其中多家公司還同時發生全額交割之紀錄，因其他財務危機事件較全額交割對財務危機之情況更適切，所以以其他危機事件日期為取樣標準，並剔除變數資料不齊全之樣本，再以一家危機公司配對 2 家正常公司為配對方法。採用至 2011 年底從未發生過任何財務危機事件的公司，來做為正常公司並進行隨機抽樣配對。最後危機公司之變數取財務危機前一年之財務比率與智慧資本指標變數資料進行建模，如前一年有發生全額交割事件則再往前一年。而正常公司變數取樣日期為 2010 年，樣本篩選過程詳見表四。

表四樣本篩選過程表

	樣本數(家)
2001 年~2011 年底台灣上市公司曾經發生財務危機事件之公司。	98
配對樣本	196
刪除資料有缺值之樣本	(44)
最後樣本數	250

### 四、建模流程

LS-SVM 和 SVM 建立分類模型時，需先將非線性資料轉換成高維度線性資料，本文使用逕向基和函數(RBF)來轉換非線性資料，所以必須先決定最佳參數，包括訓練誤差  $C$  以及核參數  $\gamma$ ，適當之參數組合，將直接影響預測模型的績效。本研究使用網格搜尋法(Grid Search)搭配交叉驗證(Cross Validation)來找尋最佳參數組合。

本文將建立 6 組模型，分別為未進行變數篩選之模型 2 組，以區別分析法篩選變數後之模型 2 組，最後以 t-test 判別最攸關變數篩選之模型 2 組。模型間差異見表五。

表五模型差異表

組別	模型	建模方法	使用變數
未篩選變數組	1	SVM	全體 19 個變數。
	2	LS-SVM	

<sup>1</sup> 因金融業適用法規及會計方法較不一樣，所以排除在樣本之外。

區別分析篩選變數組	3	區別分析+SVM	總資產週轉次數、負債比率、淨值報酬率-稅後、董監質押比率、每人配備率
	4	區別分析+LS-SVM	
t-test 找尋攸關變數組	5	t-test+SVM	總資產週轉次數、負債比率、資產報酬率、淨值報酬率-稅後、每股盈餘、每人營業利益、董監質押
	6	t-test+LS-SVM	

各模型流程為，先將所有樣本以 8:2 的比例，分成訓練資料與測試資料，再以訓練資料進行建模，並以測試資料進行準確率計算，再分別以訓練資料與測試資料之準確率與型一和型二誤差來探討模型效能。型一誤差為將非財務危機公司誤判為危機公司的機率，型二誤差為將財務危機公司誤判為非財務危機公司的機率。建置流程觀念圖見圖 2。

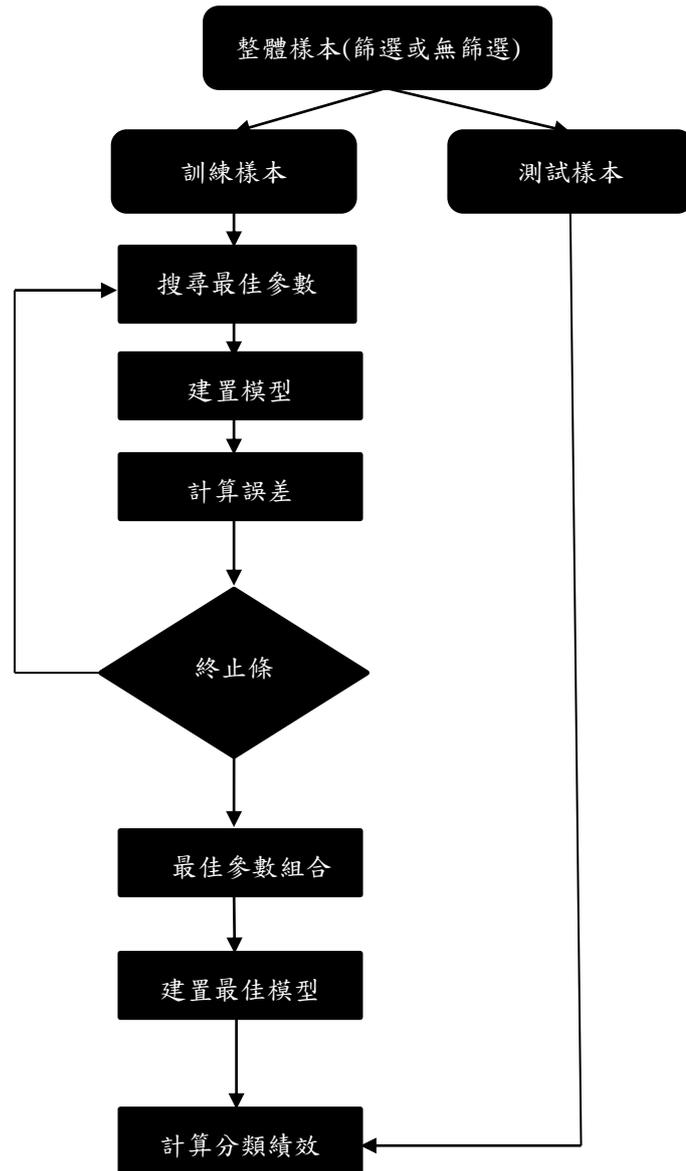


圖 2 建置流程觀念圖

#### 肆、實證結果

茲彙總 6 組模型的準確率於表六、表七、表八。表六顯示，在未篩選變數情況下，利用網格搜尋法所找尋的最佳參數對模型準確率和型一與型二誤差的影響。網格搜尋法所找到的參數為  $C=16.902$ ， $\gamma=0.547923$ ，在準確率方面，LS-SVM 明顯優於 SVM，而 SVM 有在型二誤差有較好的表現，LS-SVM 則是在型一誤差有較好之表現。

表六 未篩選變數組準確率與型一和型二誤差

未篩選變數組	SVM		LS-SVM	
	網格搜尋法	訓練組準確率	88.50%	訓練組準確率
$C=16.902$	測試組準確率	94.00%	測試組準確率	96.00%
$\gamma=0.547923$	Type1 誤差	12.35%	Type1 誤差	2.94%

	Type2 誤差	6.25%	Type2 誤差	17.50%
--	----------	-------	----------	--------

表七顯示，在區別分析篩選變數情況下，以總資產週轉次數、負債比率、淨值報酬率、董監質押比率、每人配備率搭配網格搜尋法找尋的最佳參數對模型準確率和型一與型二誤差的影響。網格搜尋法所找到的參數為  $C=2.0669$ ， $\gamma=0.25276$ ，雖然 SVM 在訓練組準確率有較高之表現，但是整體準確率仍然不如 LS-SVM，整體型一與型二誤差也與 LS-SVM 相當。

表七 區別分析法篩選變後組準確率與型一和型二誤差

區別分析法篩選變數組	SVM		LS-SVM	
	網格搜尋法	訓練組準確率	96.00%	訓練組準確率
$C=2.0669$ $\gamma=0.25276$	測試組準確率	90.00%	測試組準確率	96.00%
	Type1 誤差	5.29%	Type1 誤差	4.12%
	Type2 誤差	16.25%	Type2 誤差	13.75%

本研究使用逐步鑑別功能(stepwise discriminant function)，並且依據個變數之 Wilks' Lambda(U 統計量)做為刪減變數的準則與方法。

表八顯示，在 t 檢定篩選特徵變數情況下，以總資產週轉次數、負債比率、稅後息前、淨值報酬率、每股盈餘、每人營業利益、董監質押比率搭配網格搜尋法所找到的參數為  $C=3.56558$ ， $\gamma=13.004$ 。很明顯的，LS-SVM 整體準確率仍然高於 SVM，而在型一與型二誤差中，LS-SVM 整體來說，誤差機率較 SVM 來得大。

表八 t 檢定篩選特徵變數組準確率與型一和型二誤差

T 檢定篩選特徵變數組	SVM		LS-SVM	
	網格搜尋法	訓練組準確率	95.00%	訓練組準確率
$C=3.56558$ $\gamma=13.004$	測試組準確率	94.00%	測試組準確率	96.00%
	Type1 誤差	3.53%	Type1 誤差	2.65%
	Type2 誤差	8.75%	Type2 誤差	11.25%

本研究以 95% 的信賴水準來判別變數對於樣本是否為攸關變數。

SVM 與 LS-SVM 在三種建模方式下，搭配 T 檢定之變數篩選在整體準確率與型一與型二的誤差表現明顯優於未變數篩選與以區別分析篩選。而在 SVM 與 LS-SVM 的比較中，LS-SVM 在整體準確率與型一與型二的誤差表現明顯優於 SVM。但是 LS-SVM 與 SVM 都較容易將財務危機公司誤判為非財務危機公司(型二誤差都較型一誤差高)。

## 伍、結論

本研究除了進行支援向量機和最小平方方法支援向量機所建立之模型的比較外，更進一步的研究如果先經過特徵選取後的差異。整體來說，以 SVM 或 LS-SVM 來建立財務危機預警模型時，以 t-test 先進行變數篩選能全面地使模型準確率提高或是型一與型二誤差降低。至於 LS-SVM 和 SVM 在建模的優劣上，LS-SVM 在 6 組模型中，維持較高的準確率和較低的型一與型二誤差。未來在變數篩選結合支援向量機的相關議題的研究，可以再以其他類型的資料，如股價上升或下跌預測，和其他變數篩選方法來進行探討，研究哪種分類的資料用哪種變數篩選方法有較佳之結果顯現。

### 參考文獻:

1. 邱志洲、簡德年、高凌菁 (2003)。演化式類神經網路在企業危機診斷上之應用－智慧資本指標的考量。臺大管理論叢，14 (2)，1-22。
2. 李天行、唐筱菁 (2004)。整合財務比率與智慧資本於企業危機診斷模式之建構－類神經網路與多元適應性雲形迴歸之應用。資訊管理學報，11 (2)，161-189。
3. 林淑萍、黃劭彥、蔡昆霖 (2007)。企業危機預警模式之研究－DEA-DA、邏輯斯迴歸與類神經網路之應用。會計與公司治理，4 (1)，35-56。
4. 邱登裕、鍾典村、吳致遠、謝齊莊 (2007)。以 GA-SVM 法探討企業財務危機之研究。中華管理學報，8 (4)，61-85。
5. 劉邦典，梁榮輝(2007)。臺灣企業財務危機預警模型建構之研究。績效與策略研究 4(3):15-27。
6. 曾淑峰、江俊豪 (2008)。GA-SVM 組合式信用風險財務危機模型之研究。台灣金融財務季刊，9 (1)，1-25。
7. 葉宗翰 (2010)。運用支援向量機於按成本設計(dtc)預測系統之研究-以飛機結構系統研發為例。未出版之，國防大學中正理工學院國防科學研究所。
8. 賴鈺城、李崑進、李善玉 (2010)。公司治理下電子業之財務預警模型。華人前瞻研究，6 (1)，1-23。
9. Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23( 4), 589-609.
10. Altman, E. (1977). ZETA Analysis: a New Model to Identify the Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, 1, 29-54.
11. Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience). *Journal of Banking & Finance*, 18(3), 505-529.
12. Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 71-111.
13. Blum, M. (1974). Failing Company Discriminant-Analysis. *Journal of Accounting Research*, 12(1), 1-25.
14. Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10, 167-179.
15. Holland, J. H. (1975). *Adaptation in Natural and Artificial*. Cambridge.
16. Houghton, K., & Woodliff, D. (1987). Financial Ratios: The Prediction of Corporate 'Success' and Failure. *Journal of Business Finance & Accounting*, 14(4), 537-554.
17. Johnson, R. A. and Wichern, D. W. *Applied Multivariate Statistical Analysis*, Prentice-Hall Inc.: New York (1998).
18. Mensah, Y. (1984). An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models: A Methodological Study. *Journal of Accounting Research*, 22(1), 380-395.

19. Odom, M. D., & Sharda, R. (1990). A Neural Network Model for Bankruptcy Prediction. *IJCNN International Joint Conference*, 2, 163-168.
20. Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.
21. Pang-Ning Tan, M. S., Vipin Kumar (Ed.). (2005). *Introduction to Data Mining*: Addison Wesley.
22. Shah, J., & Murtaza, M. (2000). A Neural Network Based Clustering Procedure for Bankruptcy Prediction. *American Business Review*, 18(2), 80-86.
23. Suykens J. A. K., Vandewalle, J., & De Moor, B. (2001). Optimal Control by Least Squares Support Vector Machines. *Neural Networks*, 14(1), 23-35.
24. Suykens, J. A. K., & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, 9(3), 293-300.
25. Tam, K., & Kiang, M. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science*, 38(7), 926-947.
26. Tsai, C.-F. (2009). Feature selection in Bankruptcy Prediction. *Knowledge-Based Systems*, 22(2), 120-127.
27. Vapnik, V. (Ed.). (1995). *The Nature of Statistical Learning Theory*. New York: Springer.
28. Xie, C., Luo, C., & Yu, X. (2011). Financial Distress Prediction Based on SVM and MDA Methods: the Case of Chinese Listed Companies. *Quality and Quantity*, 45(3), 671-686.
29. Yang, M., & Da-wei, X. (2010). The Selection Method for Hyper-Parameters of Support Vector Classification by Adaptive Chaotic Cultural Algorithm. *International Journal of Intelligent Computing and Cybernetics*, 3(3), 449-462.

# The study of Combining Least Squares Support Vector Machine with feature selection in Prediction Model of Financial Distress

Ching-Chiang Yeh  
National Taipei College of Business  
[ychinc@webmail.ntcb.edu.tw](mailto:ychinc@webmail.ntcb.edu.tw)

Der-Jang Chi  
Chinese Culture University  
[chi58882000@yahoo.com.tw](mailto:chi58882000@yahoo.com.tw)

Ping-Chieh Ho  
Chinese Culture University  
[stare7500@hotmail.com](mailto:stare7500@hotmail.com)

## Abstract

This study tries to combine the Discriminant-Analysis and t-test with Least Squares Support Vector Machine to build a two-step prediction model of financial distress. We focused on the result variations of the different feature selection methods combining different SVM. We found that the feature selection of t-test had better accuracy in predicting the model and the LS-SVM was better than SVM in building prediction models of financial distress.

Keywords: Financial Distress, Data Mining, Support Vector Machine, Least Squares Support Vector Machine, Discriminant-Analysis, t-test