

金融風險管理季刊
民95，第二卷，第一期，97-114

中小企業違約機率的預測—考慮極端值

沈中華

政治大學金融系教授

忻維毅

政治大學經濟研究所研究助理

1. 前言

中小企業是我國經濟發展重要的一環，在台灣總體企業中，中小企業在家數、就業人數的比重分別為97.8%、77.18%，但如何對中小企業融資進行評比與預測違約機率的文獻卻比較少，而現有的研究的作法一般都是使用Logit模型或Probit模型來對中小企業融資做評價與預測。

例如，蘇紋慧(2000)以因素分析法及區別分析法探討中小企業信用評估模式，實證結論為：一、利用中小企業財務報表評估其安全性的正確率為58%至72%之間，故中小企業的財務資料可信度低。二、利用徵信報告的非財務資訊的確可以提高區別能力。三、結合財務報表資料與徵信報告資料之後，其區別率為75%至83%之間。但在利用這些模型時並未考慮中小企業融

資的資料中會存在極端值的影響，實際上極端值常常出現在橫剖面資料，而且對於實證結果往往有很大地影響，因此極端值必須要被謹慎處理。

例如在1998年Levine與Zervos這兩位學者於The American Economic Review期刊中，刊登了一篇研究報告“Stock markets, Banks and Economic Growth”，該研究報告試圖要回答一個問題：「Do well-functioning stock markets and banks promote long-run economic growth?」為了回答這個問題，Levine與Zervos使用47個國家，1976到1993年的資料來進行實證，最後的實證結果，認為市場的流動性(Market Liquidity)以及銀行的發展程度(Bank Development)與經濟的成長(Economic Growth)有著顯著的正向關係¹，然而在2002年Zhu、Ash與Pollin這三位學者於

¹ 在此僅簡短說明其實證結果，對於該研究報告有興趣的讀者請翻閱該研究報告。Levine, Ross & Zervos, Sara (1998), “Stock markets, Banks, and Growth,” *American Economic Review*, Vol.88 (3), 537-558.

Political Economy Research Institute所發表的 Working paper “Stock Market Liquidity and Economic Growth: A Critical Appraisal of The Levine/ Zervos Model” 中，重新做了相同的實證，但使用其他幾種不同的方式來處理極端值之後，所得到的結果顯示，除了亞洲的5個國家(台灣、韓國、香港、新加坡、泰國)以外，市場的流動性與經濟的成長並沒有顯著的關係。

所以，在進行計量模型的估計時，對於極端值的處理應當特別謹慎，因為當極端值沒有被適當處理時，對於整個估計的結果會產生極大的影響，以致於造成最後對於估計結果解釋上的偏誤。

在橫剖面資料裡面，常常會發現極端值，這些極端值除了可能是該觀察值本身奇特的性質所造成，也可能是資料登錄時發生錯誤，或是遺漏值 (Missing Data)，若這些極端值沒有經過適當的處理，則這些極端值將會對最後的估計結果產生很大的影響，進而導致以整個估計結果向外預測的準確率降低。

Robust Logit Regression (穩健羅吉斯迴歸)，是Logit Model (羅吉斯模型)再加上極端值 (Outliers) 的處理，沈中華與林公韻 (2005) 曾使用同樣方法，但他們針對的是上市與上櫃的企業資料，而本文則否，我們認為如果上市、上櫃企業資料會有極端值的存在，則中小企業資料中存在極端值的現象應該更明顯，但是過去的研究在估計中小企業違約機率時，卻忽略極端值，而本文所使用的方法：Robust Logit

Regression，是處理極端值的Logit方法。

2. 文獻回顧

2.1 Robust Regression(穩健迴歸)

在穩健迴歸估計量(Robust Estimators) 中，必須要提到一個重要的名詞：容離值(Breakdown Point, Donoho & Huber, 1983)。

現假設 Z 樣本集合 n 筆資料中，有 m 筆資料被任意值所取代，這代表有 m 筆資料是與原樣本群不相關的值，可當作是極端值的概念，而被 m 筆資料混和取代後的樣本集合以 Z' 表示；以 $bias(m; T, Z)$ 表示 Z 樣本中有 m 筆資料被任意取代後，在進行迴歸估計，所得到的最大誤差值，其定義可寫為：

$$bias(m; T, Z) = \sup_z \|T(Z') - T(Z)\| \quad (1)$$

若 $bias(m; T, Z)$ 是無限大(Infinite)，表示這 m 筆極端值對於 T 有極大的影響，讓估計量 T 的結果沒有效率了，所以對於樣本 Z 而言，估計量 T 的容離值(Breakdown Point，以 $\epsilon_n^*(T, Z)$ 表示)為：

$$\epsilon_n^*(T, Z) = \min \left\{ \frac{m}{n}; bias(m; T, Z) \text{ is infinite} \right\} \quad (2)$$

即 $\epsilon_n^*(T, Z)$ 是使得估計量 $T(Z')$ 任意 (Arbitrarily)遠離原估計量 $T(Z)$ 的極端值占所有樣本數的比例。對於使用最小平方方法的LS (Least Sum of Squares)，其 $\epsilon_n^*(T, Z) = \frac{1}{n}$ ，

只要有一個極端值的存在，便會使得估計量 $T(Z')$ 任意遠離原估計量 $T(Z)$ 。所以當 $\epsilon_n^*(T, Z)$ 越高，表示在估計量 $T(Z)$ 中可容忍的極端值比例越大，故該估計量 $T(Z)$ 便能較不受極端值的影響， $T(Z)$ 也就越穩健(Robust)。

Rousseeuw 在 1984 年所提出的 LMS (Least Median of Square, 最小中位數平方)，其容離值為 50%，而 50% 是所有穩健迴歸估計量中可達到的最高容離值，這表示極端值對於 LMS 的影響很小。其他後期所發展的穩健迴歸估計量還有 LTS (Least Trimmed Squares, 最小消去平方法, Rousseeuw, 1983 & 1984)、S-Estimators (Rousseeuw & Yohai, 1984) 等。

有關 Robust Regression 的相關研究文獻，多出自於統計及精算類別，且內容多以演算方法上的推導及證明，實際在財務方面運用的相關研究則是相當地少。

1973 年 Fama 與 MacBeth 估計資本定價模型 (Capital Asset Pricing Model, CAPM) 中的 beta 係數，是使用最小平方方法來估計；呂倩如 (2002) 認為以最小平方方法產生的估計值很容易受離群值 (Outlier) 影響，改以穩健估計量中的 LTS 來估計 CAPM 中的 beta 係數，其研究結果顯示出，穩健估計相對於最小平方方法就 CAPM 有較佳的解釋力。

在某些情況下，當資料中只存在一個離群值時，該離群值可以經由殘差圖或是 Cook's Distance、Welsch's Distance 輕易地被偵測出來而剔除。然而在較為複雜的情況下，資料中可能存在多個離群值，這些離

群值並無法經由殘差圖而偵測出來，即產生所謂的遮蔽效果 (Masking Effect)，Cook 與 Welsch 先將所有的資料納入迴歸，將離群值一個一個地偵測出來而後剔除，所以在不斷進行資料的配適時，樣本資料會逐漸減少，但使用這個方法，仍然有可能會產生遮蔽效果。相對這種「Backward Procedure」，Atkinson 提出一個「Forward Procedure」--步進搜尋演算法 (Forward Search Algorithm)，這個方法是先取出一小部分非離群值的資料進行配適，然後將其他資料一個一個地加入，再利用穩健估計量去判斷出哪組資料集合能得到最佳的配適效果，利用這個方法，不僅可以偵測出多重極端值，亦可以評估這些極端值對於該配適模型的影響。

Atkinson (1994) 所提出的步進搜尋演算法可分為三個步驟：第一、選出起始的子集合。若有 p 個變數，則一開始選出觀察值個數為 $p+1$ 個 (可以更多，如 $p+2$ 、 $p+3$ 個) 的子集合，若有 n 個觀察值，那麼便有 $\binom{n}{p+1}$ 組子集合。所選出的起始子集合，其中不應含有極端值--所選出的起始子集合，其經過配適後所產生的殘差值平方中位數，要是所有子集合中殘差值平方中位數最小的。第二、在步進搜尋演算法的進行過程中，每次多增加一個觀察值後，會形成新的子集合，假設為 $m+1$ 個，而這新的子集合內的觀察值，必須是前次迴歸配適後殘差值平方最小的 $m+1$ 個觀察值。如此步驟重複進行，直到所有 n 個觀察值都被加入為止。第三、檢視每一階段的步進搜尋演算法。隨著殘

差值平方最小的 $m+1$ 個觀察值一個一個地被加入，在每個階段的步進搜尋演算法，可藉由計算殘差值和相對於子集合的元素個數圖，看出在不同子集合下每個觀察值其殘差變動的情形，以用來判定離群值。

吳秉勳(2000)發現利用LMS的步進搜尋演算法，可以快速而有效率地找到資料中的多重離群值。范少華(2003)也用步進搜尋演算法來確認資料中的多重離群值，並可以解決多重離群值時常遇到的遮蓋效應(Masking Effect)。本研究則選用LMS穩健估計量及步進搜尋演算法的概念，進行Robust Logiti Regression的估計。

3. 模型架構

3.1 資料

本研究需要「失敗」企業及「未失敗企業」(健全公司)的資料來進行PD的估計，而此企業為中小企業，但和經濟部中小企業處所定義之中小企業不同，本研究之中小企業範圍為未上市及上櫃之企業。本研究將「失敗」定義為中小企業在財務狀況上發生跳票，且後來使債權銀行發生逾放的情況。而未「失敗」公司定義為和銀行往來借貸而沒有發生跳票、產生逾期放款以及遭催收情形。

而關於失敗企業的定義，大致上可分為兩類，法定失敗及企業財務狀況失敗。在法定失敗的定義方面，Booth(1983)定義「企業被股票交易所除名」為失敗企業，Taffler(1983)定義「經法院下令清算而被接

管資產及債權」為失敗企業，而國內在企業預警模型方面的研究，則定義「上市公司變更交易方式為全額交割股」為法定失敗。由於企業失敗應該由實質面去衡量，故亦有從企業財務狀況來定義失敗企業，如Deakin(1972)定義「經歷過倒閉、無力償還債務或是清算的企業」為失敗企業，Blum(1974)定義「無能力償債事件發生(公司已進入破產程序，或債權人同意減少其負債)的企業」為失敗企業，Scott, W, R., (1997)與Bahnsen, P.R. 和 J.W. Bartley(1992)則與Blum類似，定義「不足抵償債務，或無能力償債」為失敗企業。

所認定的失敗企業資料是自財團法人金融聯合徵信中心(以下簡稱聯徵中心)樣本時間由民國83年到92年的中小企業財務報表資料。由於未上市上櫃企業不同於上市上櫃企業，不需定期提供財務報表，而是中小企業需向銀行借貸時才提供，因此在資料的選取上便容易產生財務資料和跳票的年度時間相差過遠的情況發生，如此財務報表上的資料對跳票的解釋能力就因時間的久遠失去意義。所以我們在篩選出事中小企業財務報表時，乃採用跳票年度及前兩年之財務報表資料，故一家跳票中小企業最多可以擷取出三年度「失敗」的財務報表資料，而最少只能擷取出一個年度的財務報表，也可能沒有。最後我們隨機選取了334家失敗企業，共500筆「失敗」財務資料。

而未失敗中小企業樣本，我們隨機選取這十年內的中小企業，和銀行多次往來

皆無跳票、遭催收的情況發生，共317家企業，2,000筆財務資料。雖然這樣選取的樣本資料，是屬於追蹤資料 (Panel Data) 的形式，但是在本研究的處理上，將之視為橫剖面資料 (Cross-Sectional Data)，故總計出事公司資料 (即觀察值為1的公司) 500筆，未出事公司資料 (即觀察值為0的公司) 2,000筆，二者合計2,500筆。

3.2 估計方式的進行步驟

本節估計方法乃參考沈中華與林公韻 (2005)，在此簡略說明。首先將樣本資料分成樣本內 (In Sample) 資料與樣本外 (Out Sample) 資料兩部分，隨機取80%的原始資料作為樣本內資料 (計2,000筆)，剩下的20% (計500筆) 則為樣本外資料。使用這80%的樣本內資料所得到的估計參數後，再分別對樣本內以及樣本外資料進行的預測。

使用的估計步驟說明如下：第一、隨機選取所有樣本資料的80%作為樣本內資料，接下來的估計步驟皆針對樣本內資料所進行。第二、從樣本內資料任選用 $k+1$ (本研究選取所有樣本內資料數的1/3為起始樣本數量) 個觀察值，以Logit Model估計得 $\hat{\beta}^{(k+1)}$ ，並以此 $\hat{\beta}^{(k+1)}$ 計算所有 N 個觀察值的預測值：

$$\hat{y}_i = F(x_i \hat{\beta}^{(k+1)}) \quad i=1, \dots, N \quad (3)$$

接著計算預測觀察值的正確機率 $p^{(k+1),i}$ ：

$$p^{(k+1),i} \begin{cases} = \hat{y}_i & , \text{ if } y_i = 1 \\ = 1 - \hat{y}_i & , \text{ if } y_i = 0 \end{cases}$$

對應著正確機率，也可以計算觀察值的錯誤機率 $e^{(k+1),i}$ (類似殘差的概念) 為：

$$e^{(k+1),i} = 1 - p^{(k+1),i} \quad (4)$$

將這些 $e^{(k+1),i}$ 由小排到大後： $e^{(k+1),1} < e^{(k+1),2} < \dots < e^{(k+1),N}$ ，再取出這些 $e^{(k+1),i}$ 的中位數，並令其為 $e^{k+1,med}$ 。

第三、進行步進搜尋演算法。多增加一個觀察值，即取用 $k+1$ 個觀察值，但這 $k+1$ 個觀察值是 $\langle \text{Step 1} \rangle$ 中， $e^{(k+1),i}$ 排序後最小的 $k+1$ 個觀察值 (代表選用殘差最小的 $k+1$ 的觀察值，等於是排除了極端值)，即 $e^{(k+1),1}, e^{(k+1),2}, \dots, e^{(k+1),k+2}$ 這 $k+1$ 個觀察值，用這 $k+1$ 個觀察值再進行估計，得 $\hat{\beta}^{(k+2)}$ 。重複第一步驟中對所有 N 個觀察值的 $e^{(k+2),i}$ 進行排序，取出中位數，並令其為 $e^{(k+2),med}$ 。

第四、重複第二步驟，再增加一個觀察值，取步驟二中 $e^{(k+2),i}$ 排序後最小的 $k+2$ 個觀察值 (代表選用殘差最小的 $k+2$ 的觀察值，等於是排除了極端值)，再進行估計而得 $\hat{\beta}^{(k+3)}$ ，同樣地再對所有 N 個觀察值的 $e^{(k+3),i}$ 進行排序，取出中位數，並令其為 $e^{(k+3),med}$ 。如此逐次加入一個觀察值，直到所有觀察值都被加入為止，如此可再得到 $\hat{\beta}^{(k+4)}, \hat{\beta}^{(k+5)}, \dots, \hat{\beta}^N$ ，以及每次 $e^{(k+1),i}$ 的中位數 $e^{(k+4),med}, e^{(k+5),med}, \dots, e^{N,med}$ 。

第五、 $e^{*,med} = \min[e^{(k+1),med}, e^{(k+2),med}, \dots, e^{N,med}]$
 $e^{*,med}$ 配對的 $\hat{\beta}^*$ 就是我們最後要的 $\hat{\beta}^*$ 。

第六、可重複步驟一，但取用不同的 $k+1$ 個觀察值。

在進行上述的進行步驟中，會出現一些問題，所以在程式的設計上，必須注意樣本中1(出事公司)與0(未出事公司)的比例。

在實際的狀況下，出事公司的樣本遠比未出事公司的樣本來得少，在進行估計時，若兩者的數量比例差距太大，則估計時很容易出現無法收斂狀況，針對這個問題，本研究在所有樣本的選取上，出事公司與未出事公司的比例為1:4，而在步驟一中，隨機選取的樣本內資料與樣本外資料，出事公司與未出事公司的比例亦都維持為1:4。

其次，起始樣本數量太少，則無法進行估計。在Atkinson(1994)所提出的步進搜尋演算法中，起始樣本的數量為 $p+1$ 個(p 為變數數量)，在本研究中，使用這樣的起始樣本數量過少，無法進行Logit Regression。而本研究取所有樣本資料的1/3，來作為起始樣本數量。

第三，進行步進搜尋演算法所選用的 $e^{(k+1),i}$ 排序後最小的 $k+n$ 個觀察值，有可能會選到全部都是1(出事公司)或全部都是0(未出事公司)的觀察值，由於Logit的估計必須要有1與0的觀察值，當這種情況發生時，估計將無法進行，所以在程式的設計上，要注意每次步進搜尋演算法所選出的排序後最小的 $k+n$ 個觀察值，都必須要選到

1與0的觀察值。本研究在避免這個問題的設計上，分別選1觀察值中殘差最小的 a 個與0觀察值中殘差最小的 b 個，在一開始進行步進搜尋演算法所選取的樣本數較少時，讓 $a:b=1:1$ 。但由於樣本資料中1與0的觀察值比例為1:4，所以進行步進搜尋演算法後，隨著所選取的樣本數的逐漸增加， $a:b$ 的比率也會漸漸地接近1:4。不管所選用 $a:b$ 的比例為何，最主要的目的，是要讓每次步進搜尋演算法所選取的樣本中都會選到1與0的觀察值。

第四，避開選取收斂性不佳的 $e^{*,med}$ ，在進行步進搜尋演算法時，每一次都會產生一個 $e^{k+1,med}$ ，最後從中選取 $e^{*,med}$ 而其所配對的 $\hat{\beta}^*$ 則是最後要的估計結果。為了避免選到收斂性不佳的 $e^{*,med}$ ，所以在每次步進搜尋演算法所得到的估計結果時，必須要同時考慮估計結果的收斂性。倘若在其收斂性不佳，則要避開選取該次所產生的 $e^{k+1,med}$ 。

最後，進行步進搜尋演算法的前幾次，無法收斂。使用RATS進行步進搜尋演算法時，在前幾次無法得到收斂的估計結果，所以在前幾次所得到 $e^{k+1,med}$ ，並不會是最後的 $e^{*,med}$ 大概要納入約一半數量的樣本內資料時，估計結果才開始收斂。

4. 估計變數的選取及調整

我們採用逐步迴歸的方式，剔除不顯著及線性重合的財務變數，我們篩選出六個最具顯著的財務變數；分別為流動資

產、現金及約當現金、總資產、利息支出、營業毛利、融資活動淨現金流入。

而變數的處理，我們採用以財務比率的方式，讓所有的財務變數以比率的方式呈現。又因為Logit Model應變數的值最高為1，而各企業總資產此項變數多過於龐大，會使模型總資產參數係數十分小，故我們對所有企業總資產此項變數取LOG來進行估計。建立的模型如下：

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \quad (5)$$

y ：違約與否

$$X_1 = \text{LOG}(\text{總資產})$$

$$X_2 = \frac{\text{流動資產}}{\text{總資產}}$$

$$X_3 = \frac{\text{現金及約當現金}}{\text{總資產}}$$

$$X_4 = \frac{\text{利息支出}}{\text{總資產}}$$

$$X_5 = \frac{\text{營業毛利}}{\text{總資產}}$$

$$X_6 = \frac{\text{融資活動淨現金流入}}{\text{總資產}}$$

當判別分數的 y 值愈小，代表公司越健全，也就是其違約風險愈低，反之，若 y 值

很大或是為1，就帶表示公司越脆弱，其違約風險也愈高。

5. 信用模型預測效力的比較方法

本文在模型間的預測效力比較上，使用三種方式：CAP 曲線 (Cumulative Accuracy Profiles)、ROC (Relative or Receiver Operation Characteristic) 曲線、KS (Kolmogorov- Smirnov)。

5.1 CAPs (Cumulative Accuracy Profiles) 曲線

模型間的表現好壞，可以直接從CAP的圖形上來看。

我們舉例說明CAP曲線的繪製，假設現在有500家公司(即500個觀察值)，其中有50家為出事公司，450家為未出事公司，在這個例子中， $N + D = 500$ ， $D = 50$ ， $N = 450$ 。經過模型配適後，這500家公司各有模型配適後的違約機率，將這些違約機率由大排到小之後，取前5%家公司(即 $x=5$ ，代表前25家公司)，發現模型所配適的違約機率為95%以上且實際上觀察值也是1的公司(即實際上為出事公司)的數目總計有4個，這代表當模型以95%為截斷點 (Cut Point)²時，模型可以偵測出8% ($\frac{4}{50} \times 100\% = 8\%$ ，即 $y(5\%) = 8\%$)的違約公司；

² 截斷點的意思是：決定公司為出事公司或是未出事公司的門檻值。在本文中，若以95%為截斷點，則表示若違約機率在95%以上，則判定該公司為出事公司；若違約機率在95%以下，則判定該公司為未出事公司。

接著再取前10%家公司(即 $x=10$ ，代表前50家公司)，發現模型所配適的違約機率為90%以上且實際上觀察值也是1的公司(即實際上為出事公司)的數目總計有9個，這代表當模型以90%為截斷點(cut point)時，模型可以偵測出18% ($\frac{9}{50} \times 100\% = 18\%$ ，即 $y(5\%) = 18\%$)的違約公司，如此進行，直到取到100% (即取完所有500家公司)為止，以 $x\%$ (即5%、10%、...、100%)為橫軸， $y(x\%)$ (即8%、18%、...、100%)為縱軸，則可繪製出CAP曲線。

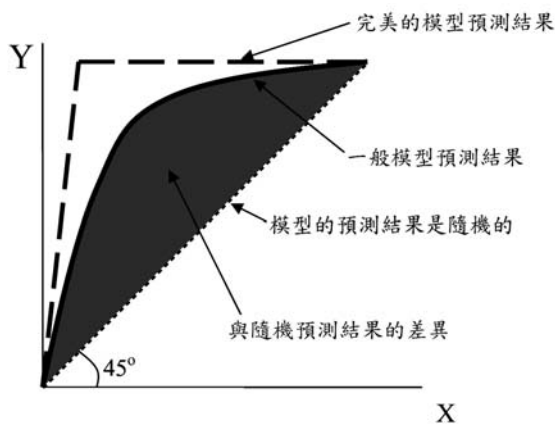


圖1 CAP曲線

圖1中的實線為依模型配適結果所繪製的CAP曲線，45度角的粗點虛線表示模型所偵測出的違約公司比例與被偵測的公司數目(經排序後)占所有公司數目的比例相等，這代表這個模型預測結果幾乎是隨機的，也就是表示這個模型並不具有任何的偵測能力。最理想的模型應該對所有實際上出事公司所配適出的違約率都很高，所

以可以在X軸越前端的部分(即違約率越高的部分)捕捉到越多的出事公司，也就是在X軸越前端的部分， y 值越大越好，也就是CAP曲線越往左上，CAP曲線下的面積越大，則代表模型的偵測能力越好，圖1中的灰色長條虛線表示完美的模型預測結果。

5.2 ROC(Relative or Receiver Operation Characteristic)曲線

在討論ROC曲線之前，要先定義表1中的幾個名詞：

表1 交叉分類表

實際情況 模型判定	出事公司 ($y_i = 1$)	未出事公司 ($y_i = 0$)
出事公司 ($\hat{y}_i = F(x_i, \hat{\beta}^*) \geq \text{截斷點}$)	TP% × D	FP% × N
未出事公司 ($\hat{y}_i = F(x_i, \hat{\beta}^*) < \text{截斷點}$)	FN% × D	TN% × N

表1中TP為True Positive之意，TP%即實際上違約且模型的估計結果也判定違約的比率；FN為False Negative之意，FN%即實際上違約但模型的估計結果卻未判定是違約的比率，這種錯誤稱為型一錯誤(Type I Error)，在這樣的定義之下， $TP\% + FN\% = 1$ 。FP為False Positive之意，FP%即實際上未違約但模型的估計結果卻判定為違約的比率，這種錯誤稱為型二錯誤(Type II Error)；

TN為True Negative之意，TN%即實際上未違約且模型的估計結果也判定未違約的比率，這樣的定義之下， $FP\% + TN\% = 1$ 。

依據不同的截斷點，可以得到不同的交叉分類表，而ROC曲線就是在繪製FP%與TP%之間的關係。舉例來說，同樣地，假設現在有500家公司(即500個觀察值)，其中有50家為出事公司，450家為未出事公司，在這個例子中， $N + D = 500$ ， $D = 50$ ， $N = 450$ 。經過模型配適後，這500家公司各有模型配適後的違約機率，將這些違約機率由大排到小，然後先要找到一個截斷點 $c\%$ ，在這個截斷點之下，FP%(型二誤差，代表實際上是未違約公司，但模型的估計結果卻判定為違約的比例)為5%，即以 $c\%$ 為截斷點時，在450家未違約公司中，有22.5 ($450 \times 5\% = 22.5$)家公司被判定為違約(因為公司數目必須以整數出現，故實際上可取23家或22家)，找到這個截斷點 $c\%$ 之後，再依據這個截斷點，計算TP%。若在50家違約公司中，以 $c\%$ 為截斷點時，該模型可以捕捉到4個違約公司(即觀察值為1且估計的違約機率在 $c\%$ 以上的樣本公司)，則代表TP%為8% ($4 \div 50 = 8\%$)。接著進行重複的動作，即對應10%、15%、...、100%的型二誤差，找到不同的截斷點 $c\%$ ，再根據不同的截斷點找到不同的 $TP_1\%$ 、 $TP_2\%$ 、...、 $TP_{20}\%$ 。以型二錯誤為橫軸，即5%、10%、15%、...、100%，縱軸為 $TP_1\%$ 、 $TP_2\%$ 、...、 $TP_{20}\%$ ，則可繪製出ROC曲線，如圖2。

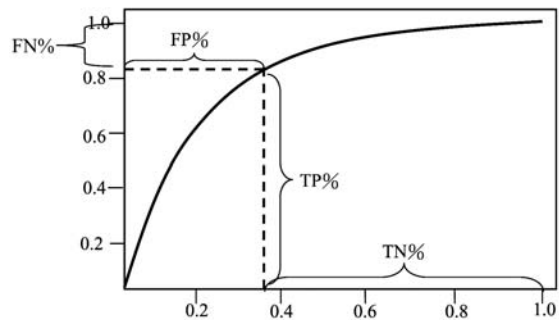


圖2 ROC曲線

由圖2 ROC曲線來看，在同一個TP%之下，FP%越小，則代表所犯的型二錯誤越小，代表模型的檢驗效力越好，因此，ROC曲線越往左上，ROC曲線下的面積越大，則代表模型越好。

5.3 K-S (Kolmogorov-Smirnov)

K-S是繪製無違約公司的預測值的累積分配和違約公司的預測值的累積分配圖形，如果這兩個累積機率分配的差異越大，代表模型對於無違約公司與違約公司的鑑別度越大，則代表模型的檢驗效力越好。

6. 實證結果

6.1 變數估計結果

我們以傳統Logit以及Robust Logit Regression兩種模型來討論實證結果。這六

表2 解釋變數基本統計量

變數名稱	觀察個數	平均值	標準差	最小值	最大值
X ₁	2,500	12.18804	1.167761	7.917172	17.00461
X ₂	2,500	0.563441	0.26679	0.001023	1
X ₃	2,500	0.07832	0.104914	0	0.763735
X ₄	2,500	0.028678	0.023728	0	0.584242
X ₅	2,499	0.187217	0.190505	-1.16432	2.779161
X ₆	2,499	0.031864	0.223765	-5.08683	1.269564

個變數中，我們預期總資產、流動資產、現金及約當現金、營業毛利越高越不容易倒閉，而利息支出與融資活動淨現金流入越高則易倒閉。

表2為此六個變數的基本統計量，我們可以看出總資產觀察值的差異性最大，而利息支出觀察值的差異性最小。

我們可由圖3到圖8中各解釋變數的直方圖來了解各變數的觀察值的分佈情形；在圖3為總資產直方圖中，我們將總資產取log值來處理，以減少此變數分配的差異性，但取log值後，總資產的變異仍就是此六項變數中差異性最大的，但仍可看出此變數有集中分配的趨勢。

圖4流動資產則沒有明顯的趨勢分佈，有點略成均勻分配的趨勢，而此變數在六項解釋變數中差異性第二小。而圖5現金與約當現金、圖6利息支出呈現出明顯右偏情況，其中圖6利息支出解釋變數為六項解釋變數中變異性最小的。圖7營業收入、圖8融資活動淨現金流入則有集中趨勢的現象，圖形類似鐘型分配。

在表3為傳統Logit模型及Robust Logit Regression模型的實證結果；在Logit模型及Robust Logit Regression模型中，六個變數的係數符號，總資產、流動資產、利息支出與融資活動淨現金流入符號為正，其中總資產與流動資產越高越容易倒閉和我們一開始預期相反，是因當中小企業規模擴大時與銀行往來更密切頻繁，所以承受的營業風險也跟著增加？是因為如此或其他原因需要再深入研究。而現金與約當現金越高則越不容易倒閉，符合我們的預期，如果中小企業資金充沛則可定時還款，如此將不會造成違約。而利息支出越高則越容易倒閉符合我們預期，中小企業利息支出負擔龐大則代表財務槓桿過高，一旦營運上不順利，很容易就會發生財務危機產生違約。而營業收入越高越不容易倒閉也符合我們的預期，如果中小企業營業收入高，代表資金回收快，則比較沒有付息的壓力，而不容易違約。最後融資活動淨現金流入越高則越容易違約，則也說明了中小企業自有資金太少而財務槓桿越高則越

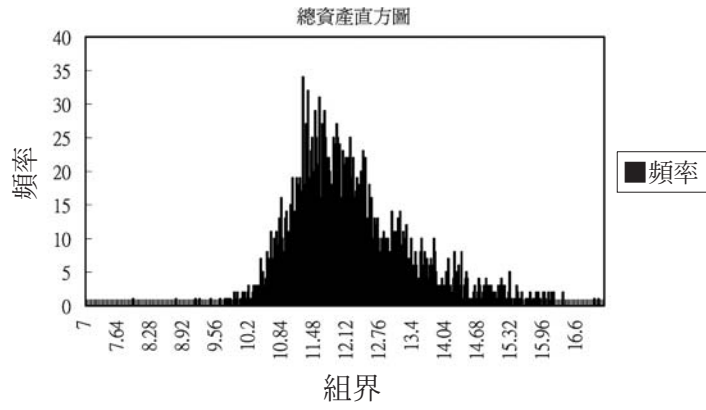


圖3 總資產直方圖

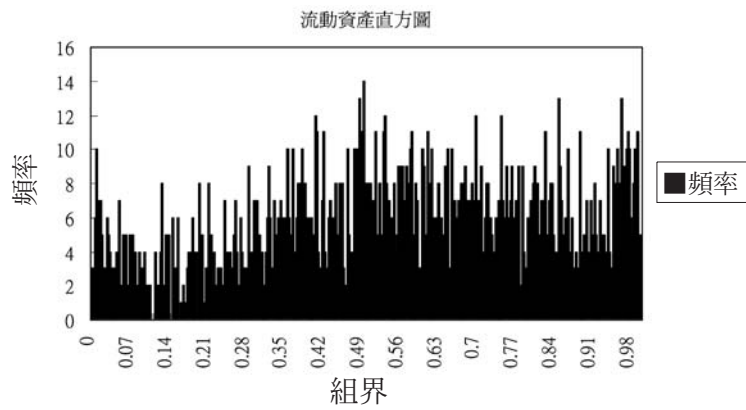


圖4 流動資產直方圖

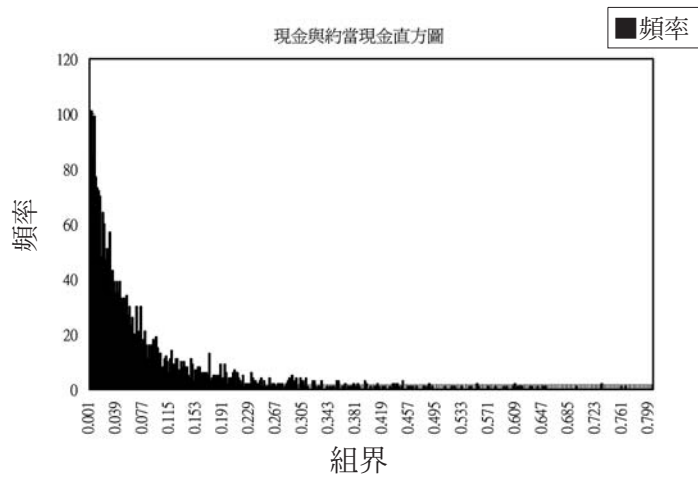


圖5 現金與約當現金直方圖

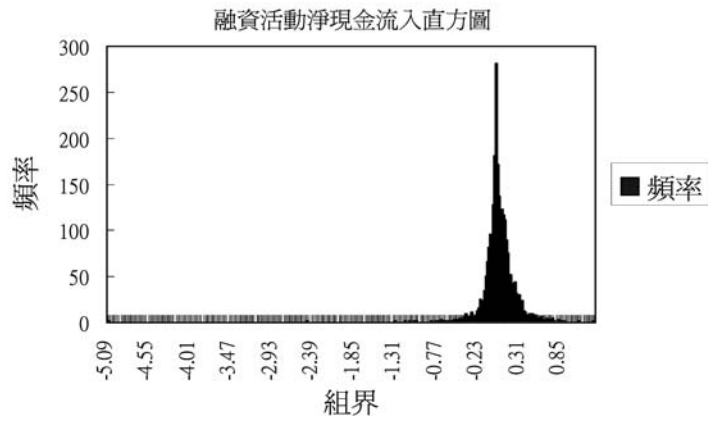
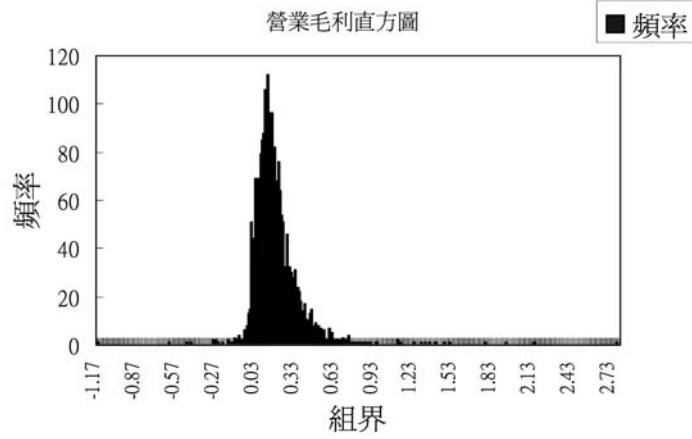
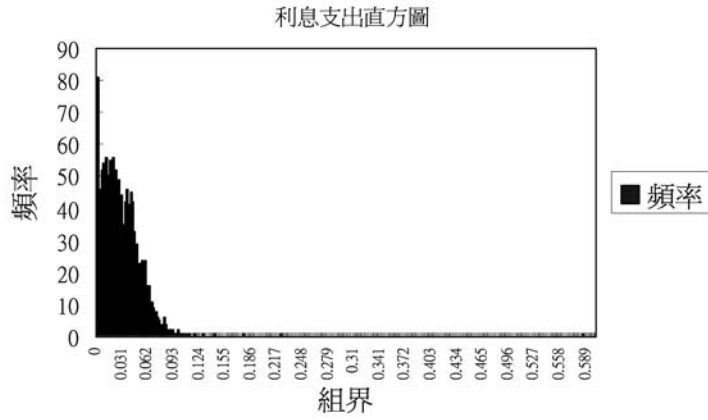


表3 Logit與Robust Logit模型實證結果

變數名稱		logit		Robust Logit	
		估計係數	T值	估計係數	T值
constant		-6.537	-8.274	-9.660	-7.366
Log(總資產)	X ₁	0.237	4.250	0.298	3.178
流動資產/總資產	X ₂	2.413	8.979	5.289	11.584
現金與約當現金/總資產	X ₃	-4.860	-4.539	-8.821	-6.654
利息支出/總資產	X ₄	41.517	12.726	140.510	15.961
營業收入/總資產	X ₅	-2.137	-4.200	-8.800	-9.372
融資活動淨現金流入/總資產	X ₆	1.801	5.351	5.681	9.591
樣本數		2,000		1,534	

表4 兩個模型的Pseudo-R-square

	Logit模型	Robust Logit Regression模型
Pseudo-R-square	0.1851972	0.5937802

容易發生財務危機。

在Robust Logit Regression模型中，共剔除極端值樣本466個，而再與Logit模型變數的顯著效果上比較，模型的流動資產變數顯著效果下降，而總資產、營業毛利及融資活動淨現金流入變數的顯著性提升。

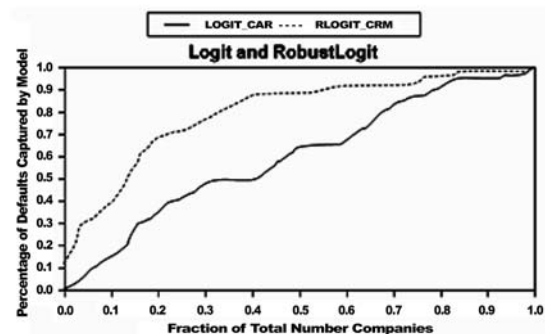
在Logit與Robust Logit Regression在Pseudo-R-square比較中可發現經過刪除離群值後的Robust Logit Regression模型得解釋能力上升；由0.1851972上升至0.5937802。

6.2 預測效力的比較

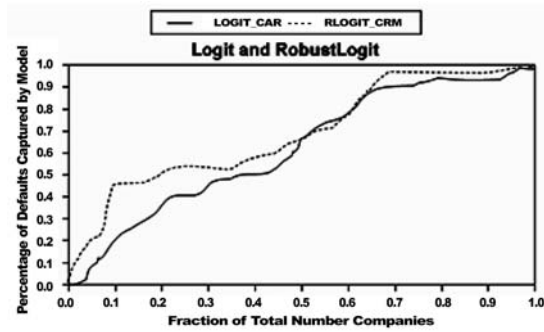
在模型預測效力的部分，本研究會分

別針對樣本內(2,000筆觀察值)與樣本外(500筆觀察值)對二個模型的預測效力做比較：

(一)CAPs (Cumulative Accuracy Profiles)曲線



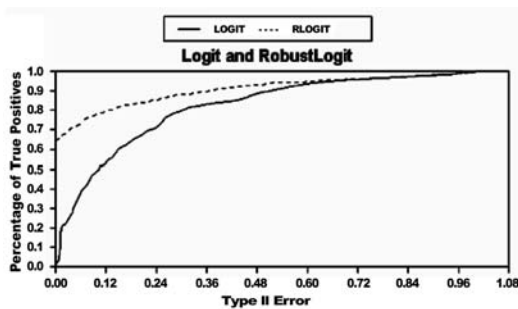
CAP曲線(樣本內)



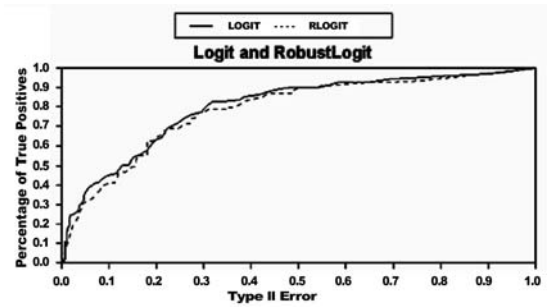
CAP曲線(樣本外)

由前所述CAP曲線越往左上，則代表模型的預測效力越好，而虛線為Robust Logit Regression模型所產生的曲線，Logit模型為實線，虛線在實線左上方，故可以很明顯地看出針對樣本內資料或樣本外資料，使用Robust Logit Regression的模型預測效力比傳統Logit的預測效力好。

(二) ROC (Relative or Receiver Operation Characteristic)曲線



ROC曲線(樣本內)

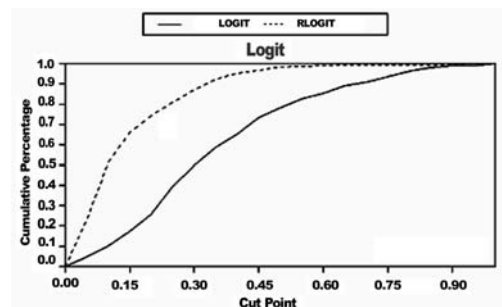


ROC曲線(樣本外)

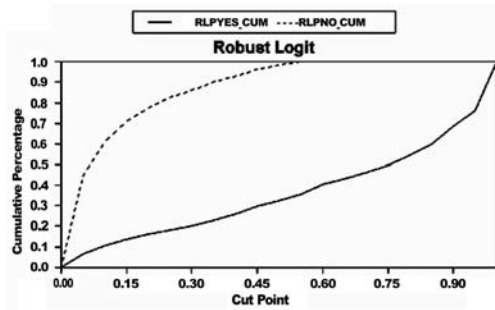
ROC曲線越往左上，則代表模型的預測效力越好，而虛線為Robust Logit Regression模型所產生的曲線，Logit模型為實線，樣本內資料中Logit面積為0.81153，Robust Logit面積為0.90563，Robust Logit準確率較佳。而使用樣本外資料Logit面積為0.80251，Robust Logit面積為0.78692，兩者準確率差不多，使用Robust Logit Regression並沒有讓模型的預測效力有明顯的提升效果，其模型的提升效果難以判斷。

(三) K-S (Kolmogorov-Smirnov)

(1) K-S(樣本內)

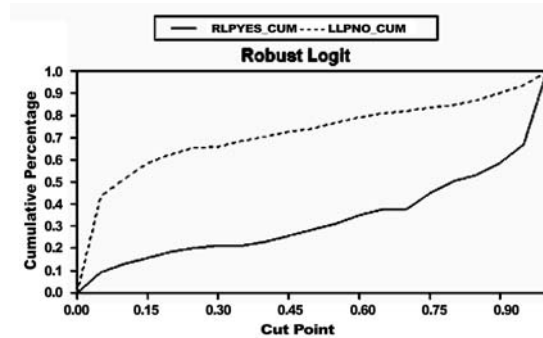


K-S(樣本內)



K-S(樣本內)

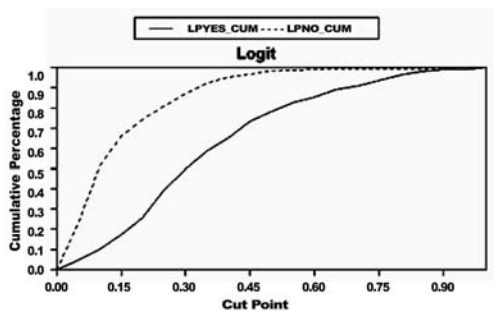
圖中的實線表示違約公司預測值的累積機率分配，虛線則表示無違約公司預測值的累積機率分配，如果這兩個累積機率分配的差異越大，代表模型越好。代表針對樣本內資料，可以很明顯地看出針對樣本內資料，使用Robust Logit Regression的模型預測效力比傳統Logit的預測效力好。



K-S(樣本外)

圖中的實線表示違約公司預測值的累積機率分配，虛線則表示無違約公司預測值的累積機率分配，如果這兩個累積機率分配的差異越大，代表模型越好。代表針對樣本外資料，可以很明顯地看出針對樣本外資料，使用Robust Logit Regression的模型預測效力比傳統Logit的預測效力好。

(2) K-S (樣本外)



K-S(樣本外)

6.3 模型偵測出的極端值

在本研究中，一開始隨機抽取的2,000個樣本內資料中，有400個出事公司資料($y_i=1$)及1,600個未出事公司資料($y_i=0$)。Robust Logit Regression模型最後選出了1,534個觀察值，其中有390個為出事公司資料，1,144個為未出事公司資料，也就是在模型二中，有456個未出事公司以及10個出事公司被視為極端值。

本研究使用Robust Logit Regression所偵測出的極端值，大部分皆是屬於未出事公司，代表這些未出事公司在該模型下所估算求出的違約機率偏高，與實際狀況

表5 樣本內及樣本外模型效力比較

比較方法	CAP	ROC	KS
Robust Logit Regression模型 是否優於Logit模型(樣本內)	○	○	○
Robust Logit Regression模型 是否優於Logit模型(樣本外)	○	—	○

($y_i=0$)不符，因而被視為極端值，這表示 Robust Logit Regression的模型中所選出的極端值是傾向型二錯誤(實際上為未出事公司，而模型判定為出事公司)的觀察值，所以是屬於比較保守的模型。

7. 研究結論與未來研究建議

本文所提出的 Robust Logit Regression，是一個考慮極端值後的Logit Model，而本研究在樣本內資料的實證結果顯示，使用Robust Logit Regression對於整個模型的預測效力的確有提升的效果，這符合本文的目的及預期。這表示在資料中，極端值的存在的確降低模型的預測效力，所以在進行PD的預測時，是必須要考慮極端值的影響的，而本文所提出的這個PD估計方法，的確可以消彌極端值所造成的不良影響，進而提升模型的預測效力。我們將樣本內(In Sample)資料、樣本外(Out of Sample)資料的實證結果列如表5，表中的○符號代表「是」，—符號代表「無法判斷」。

從表5可知，針對樣本內資料，模型預

測效力的檢定方法中，都顯示使用Robust Logit Regression的模型能夠產生較佳的模型預測效力，而針對樣本外資料，在CAP曲線及KS圖中，顯示使用Robust Logit Regression的模型能夠產生較佳的模型預測效力。最後，在模型的解釋能力上，Robust Logit Regression模型的Pseudo-R-square值大於傳統Logit模型，表示相較於傳統Logit法，使用Robust Logit Regression能讓模型的解釋能力提升。

本文在變數的選取上，僅以中小企業財務報表中的六個變數作為解釋變數，但未上市上櫃中小企業的財務報表不如上市櫃企業嚴謹，故財務資料可能存在部份偏差。而中小企業大部分為初始草創之企業，因此負責人的信用狀況及經營能力對於營運及償債能力有顯著的影響，故如能採納負責人個人信用紀錄為變數列入模型將有助於模型的預測能力。另外，本文總資產解釋變數係數為負，和預期結果相反也是一個值得深入探討的方向。此外，對於違約機率變數的選取，仍有其他面向的考量，如結合會計基礎與市場基礎之變數外，另行加入總體經濟變數後。對於不同

的產業，亦有其較為特殊的變數需要特別考慮。

本文所採取的穩健估計量為LMS，而尚其他的穩健估計量如LTS、S-Estimators等值得探討。另外除了本文殘差的定義外尚有其他殘差的定義方式，依據不同殘差的定義方式，在進行步進搜尋演算法時所求得的殘差便不同，後續的研究可採用不同的殘差定義方式，或許可以使得模型得到更好預測效力。在出事公司與健全公司的樣本配對上，本研究採用1：4的比例建構模型，但實質上可能可藉由研究決定比例的多寡而建構更佳之模型。套入Robust Logit Regression模型計算違約機率的觀察值，必須是「非極端值」，但如何定觀察值屬於「極端值」或「非極端值」，是一個非常值得探討的議題。

參考文獻

- 蘇紋慧 (2000)，「中小企業信用評估模式之研究~以中小型製造業為例」，國立中山大學財務管理學系研究所。
- 林公韻 (2005)，「違約機率之預測-Robust Logistic Regression」，國立政治大學金融學系研究所
- 吳秉勳 (2000)，「Detection of Outliers with Data Transformation」，國立政治大學統計學系研究所。
- 呂倩如 (2002)，「On the CAPM from the Views of Robustness and Longitudinal Analysis」，國立政治大學統計研究所。
- 范少華 (2003)，「Robust Diagnostic for the Logistic Regression Model With Incomplete Data」，政治大學統計學系研究所。
- Atkinson, A.C. (1994), "Fast Very Robust Methods for the Detection of Multiple Outliers," *Journal of the American Statistical Association*, 89, 1329-1339.
- Andong Zhu, Michael Ash and Robert Pollin (2002), "Stock Market Liquidity and Economic Growth: A Critical Appraisal of The Levine/ Zervos Model," working paper series, number 47, Political Economy Research Institute.
- Black, F. and M. Scholes. (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, May-Jun, 637-654.
- Blum, M.(1974), "Failing company discriminate analysis," *Journal of Accounting Research* 12, 72-102.
- Booth, P. J. (1983), "Decomposition Measures and the Prediction of Financial Failure," *Journal of Business Finance and Accounting* 10, 67-85.
- Bahnsen, P.R. and J.W. Bartley (1992), "The Sensitivity of Failure Prediction Models to Alternative Definitions of Failure," *Advances in Accounting*, 255-278.
- Deakin, E. B. (1972), "A Discriminant Analysis of Predictors of Business Failure," *Journal of Accounting Research* 10, 167-179.
- Donoho, D.L. and Huber, P.J. (1983), "The notion of breakdown point. In Bickel, P.J., Doksum, K.A. and Hodges, J.L. Jr. (eds)," *A Festschrift for Erich L. Lehmann*, 157-184, Belmont, California. Wadsworth.
- Levine, Ross and Zervos, Sara (1998), "Stock markets, Banks, and Growth," *American Economic Review*, Vol.88(3), 537-558.
- Rousseeuw P.J. (1983), "Regression techniques with high breakdown point," *The Institute of Mathematical Statistics Bulletin*, 12, 155.

- Rousseuw P.J. (1984), "Least median of squares regression," *Journal of the American Statistical Association*, 79, 871-880.
- Rousseuw P.J. and Yohai V.J. (1984), "Robust regression by means of S-estimators," *Robust and Nonlinear Time Series Analysis* (Eds) W. H. Franke and R.D. Martin Springer Verlag, New York, 256-272.
- Scott, W, R., (1997), *Financial Accounting Theory*. Ontario: prentice-Hall. Canada Inc.
- Taffler, R.J. (1983), "The Assessment of Company Solvency and Performance Using a Statistical Model," *Accounting and Business Research*, 295-308.