

大規模データによるデフォルト確率の推定 — 中小企業信用リスク情報 データベースを用いて —

高橋 久尚^{1,2} ・ 山下 智志^{3,4}

(受付 2002 年 4 月 1 日; 改訂 2002 年 10 月 15 日)

要 旨

信用リスク管理という観点から、企業の倒産(デフォルト)確率を正確に把握することが重要である。しかし、従来中小企業の倒産確率については、データの整備が遅れ、まとまったデータの入手には困難があった。しかし今般、銀行、信用保証協会、中小企業庁の協力を得て設立された CRD が信用リスクおよび財務データの共有化と一元管理を行い、データベースの構築を行ったことから、大量の中小企業データの整備が進み、これらのデータに基づく研究が可能となった。そこで、線形ロジットモデルを用いて、中小企業の財務データから倒産確率を求めた。説明変数の検定には t 検定を用いた。一般に、中小企業の財務データは欠損値(欠測値)が多く含まれており、そのため多くの統計的困難を生じる。この点を独自の方法で回避することを試みた。本稿では、特に企業の規模による倒産確率の違いについて考察した。

キーワード：信用リスク、ロジットモデル、倒産確率、中小企業。

1. 導入

企業の倒産(デフォルト)判別、倒産(デフォルト)確率の研究は、金融の分野において重要なテーマである。特に近年、銀行の不良債権問題と絡んで、信用リスク管理という観点から重要なテーマの一つとなっている。信用リスク管理のためには、倒産を予想するだけでなく倒産確率を正確に予想し、信用リスクに見合うリターンを確保することが必要である。この観点から様々な手法が考え出され、様々なデータに対して適用され、検証が行われている(Saunders (1999), 木島・小守林(1999) 森平(1999))。

データに関して言えば、大企業のもと中小企業を使ったものと大きく分けることができる。財務指標に関しては、前者は情報公開が進み、まだ不十分な点があるとは言え、多くの企業で信用リスクを分析するための正確な情報が得られ、また整備、蓄積されるようになってきた。一方、中小企業の財務指標に関して言えば、決まった情報公開の形が整備されていないこともあり、情報の入手は困難であり、信用リスクを分析するためのデータの整備、蓄積が大

¹ 総合研究大学院大学 数物科学研究科統計科学専攻：〒106-8569 東京都港区南麻布 4-6-7

² 金融庁 総務企画局政策課：〒100-8967 東京都千代田区霞ヶ関 3-1-1

³ 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

⁴ CRD 運営協議会：〒104-0031 東京都中央区京橋 2-12-2 第二ぬ利彦ビル 5F

企業に比べて進んでいない。

このことから近年、中小企業のデータを整備、蓄積する団体、中小企業信用リスク情報データベース運営協会(CRD 運営協議会)が設立された。そして昨今、この CRD 運営協議会から情報の提供を得ることができ、大量の中小企業データに基づく研究を始めた。その分析結果を報告する。分析には線形ロジットモデルを用いたが、その理由は線形ロジットモデルが一般的に知られた手法の一つであるという筆者の認識からである。

本稿の構成は次のようになっている。第 2 章で分析に用いた方法について述べる。一般に、中小企業の財務データは欠損値(欠測値)が多く含まれており、そのため多くの統計的困難を生じる。この点を回避するため、ロジットモデルに 0-1 のダミー変数を新たに取り込んだ。また、この方法はある企業の得られた財務指標に欠損値があるという状態を、その企業のデフォルトを予測する上で価値ある情報であると捕らえたものとも言える。

第 3 章では分析に用いた CRD のデータの性質について述べる。また、デフォルトの定義についても述べる。最後に、第 4 章で分析の結果を示し、選ばれた変数に対して考察する。また、企業の規模による倒産確率と説明変数の違いについて議論する。

2. 分析の手法

2.1 モデル

分析は線形ロジットモデルを用いる(木島・小守林(1999), 土木学会土木計画学研究委員会(1995))。標本データとして N 社の企業が存在する場合を考えて、企業 $i(i = 1, 2, \dots, N)$ の財務データを $\mathbf{x}^i = (x_1^i, x_2^i, \dots, x_M^i)^t$ とする。このとき、企業 i の倒産確率 p_i を

$$(2.1) \quad p_i = \frac{1}{1 + \exp(\beta_1 x_1^i + \dots + \beta_M x_M^i + \beta_{M+1})} \\ = \frac{1}{1 + \exp(\boldsymbol{\beta} \cdot \mathbf{x}^i + \beta_{M+1})}$$

と表せるものとする。但し、パラメータ $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_M)$ である。

パラメータの推定法には最尤法を用いる(坂元 他(1983))。変数 θ を用いて同時確率密度関数 $f(x_1, \dots, x_N | \theta)$ が、確率変数 (X_1, \dots, X_N) に対して与えられているものとする。関数 $f(x_1, \dots, x_N | \theta)$ を θ の関数と見なすとき、この関数を尤度と呼び、一般に $L(\theta)$ で表す。各確率変数が独立である場合、同時確率密度関数は

$$L(\theta) \equiv f(x_1, \dots, x_N | \theta) = f(x_1 | \theta) f(x_2 | \theta) \cdots f(x_N | \theta)$$

と各 X_i の確率密度関数の積で書くことができる。上式の両辺の対数をとると対数尤度 $l(\theta)$ が得られる。

$$l(\theta) = \ln f(x_1, x_2, \dots, x_N | \theta) = \sum_{i=1}^N \ln f(x_i | \theta)$$

上式を最大化することにより、最適な変数の組を得る。

この最尤法を用いてパラメータ $(\boldsymbol{\beta}, \beta_{M+1})$ の推定をする。企業倒産は企業ごとに独立に起こると仮定すると、尤度関数は

$$(2.2) \quad L(\boldsymbol{\beta}, \beta_{M+1}) = \prod_i p_i^{\delta_i} (1 - p_i)^{1 - \delta_i}$$

となる。但し、

$$\delta_i = \begin{cases} 1 & \text{企業 } i \text{ が倒産(デフォルト)しているとき} \\ 0 & \text{その他} \end{cases}$$

である．

よって，対数尤度は

$$(2.3) \quad \begin{aligned} l(\beta, \beta_{M+1}) &= \sum_i \ln p_i^{\delta_i} (1 - p_i)^{1 - \delta_i} \\ &= \sum_i \left\{ \delta_i \ln \frac{1}{1 + \exp(\beta \cdot \mathbf{x}^i + \beta_{M+1})} + (1 - \delta_i) \ln \frac{\exp(\beta \cdot \mathbf{x}^i + \beta_{M+1})}{1 + \exp(\beta \cdot \mathbf{x}^i + \beta_{M+1})} \right\} \end{aligned}$$

である．この対数尤度を最大化することにより，最適な変数の組 $\hat{\beta}, \hat{\beta}_{M+1}$ を得る．

以上は，一般的に行われているロジットモデルの手法であると思われる．今回，以上の方法を基に以下の様な改良を行った．式(2.1)を

$$(2.4) \quad \begin{aligned} p_i &= \frac{1}{1 + \exp(\beta_1 x_1^i + \cdots + \beta_M x_M^i + \beta_{M+1} + \gamma_1 y_1^i + \cdots + \gamma_k y_k^i)} \\ &= \frac{1}{1 + \exp(\beta \cdot \mathbf{x}^i + \beta_{M+1} + \gamma \cdot \mathbf{y}^i)} \\ x_k^i &= \begin{cases} 0 & \text{企業 } i \text{ の財務データ } k \text{ が得られないとき} \\ x_k^i & \text{その他} \end{cases} \\ y_j^i &= \begin{cases} 1 & \text{企業 } i \text{ の項目 } j \text{ が欠損値のとき} \\ 0 & \text{その他} \end{cases} \end{aligned}$$

とした．ただし， $\gamma = (\gamma_1, \dots, \gamma_k)$ ， $\mathbf{y}^i = (y_1^i, \dots, y_k^i)^t$ である．一般のロジットモデルとの比較で言うと，このモデルは，財務データ x_k^i が得られない場合は零と置き，代わりにダミー変数 y_j^i を 1 としたものである．つまり，CRD のデータには欠損値があり，すべての企業の分析データが完全な形で得られる訳ではない．そこで欠損値により分析データが作れなかった場合，その変数(財務データ)は零と置き，そのような事態が発生した場合は， y_j^i を 1 と置く．

このモデルは，ある企業の得られた財務指標に欠損値があるという状態を，その企業のデフォルトを予測する上で有益な情報であると捕らえ，積極的にデータの欠損という状態を分析に取り入れたものと捕らえることができる．今回，計算時間の関係から，二つの変数 y_1 ((受取)手形に関するデータが得られないか，または値が零の企業に対して 1 となる．)と， y_2 (y_1 以外でデータが欠損しているか，または値が零のためにわり算ができないことにより，分析項目の変数が作れなかったもの．例えば，前年度のデータがないもの等に対して 1 となる．)だけを用いた．

このような欠損値によるバイアス効果は，パラメータ β に対して影響があるという処理も可能だが，変数の数がもともと多くこれ以上モデルを複雑にしないためにもダミー変数により対応することとした．

2.2 検定と変数選択

t 検定を用いて変数選択を行う．まず，式(2.3)からヘッセ行列を計算する．

$$\nabla^2 l(\hat{\beta}, \hat{\beta}_{M+1}) = \begin{pmatrix} \partial^2 l(\hat{\beta}, \hat{\beta}_{M+1}) / \partial \beta_1^2 & \cdots & \partial^2 l(\hat{\beta}, \hat{\beta}_{M+1}) / \partial \beta_1 \partial \beta_{M+1} \\ \vdots & & \vdots \\ \partial^2 l(\hat{\beta}, \hat{\beta}_{M+1}) / \partial \beta_1 \partial \beta_{M+1} & \cdots & \partial^2 l(\hat{\beta}, \hat{\beta}_{M+1}) / \partial \beta_{M+1}^2 \end{pmatrix}$$

ただし, $\hat{\beta}, \hat{\beta}_{M+1}$ は最尤推定値である.

$\hat{\beta}_i$ が正規分布をしているものとする, t 値をヘッセ行列の逆行列の対角成分 $[\nabla^2 l(\hat{\beta}, \hat{\beta}_{M+1})]_{kk}^{-1}$ を用いて

$$t_k = \frac{\hat{\beta}_k}{\sqrt{[\nabla^2 l(\hat{\beta}, \hat{\beta}_{M+1})]_{kk}^{-1}}}$$

と表せる(Rao(1973)).

t_k の絶対値が 1.96 以上ならば, β_k は 95% の信頼度で倒産確率に影響を与える変数である. 逆に 1.96 以下ならば信頼度 95% で倒産確率に影響を与えない変数である. そこで, $|t_k|$ の値が 1.96 以下の変数を除き, 再び同じ手順で残された変数に対して t_k の値を求め, 全ての $|t_k|$ の値が 1.96 以下になったところで計算を終了する. 最後に残された変数を選ばれた説明変数として扱う. t 検定以外に以下の検定を行い, 選ばれた変数の確からしさを確認する.

全ての説明変数の値が零でないかどうかの検定をする. 帰無仮説 $H: \beta_1 = \cdots = \beta_{M+1} = 0$ のもとで, サンプル数の多い極限で漸近的に

$$-2\{l(0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1})\}$$

が自由度 $M+1$ の χ^2 分布に従うことが知られている(Rao(1973)). したがって, 自由度 $M+1$ の χ^2 分布の α 超過確率の値を χ_α^2 とすると, $-2\{l(0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1})\} < \chi_\alpha^2$ のとき有意水準 α で帰無仮説 H を採択し $-2\{l(0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1})\} > \chi_\alpha^2$ ならば棄却する.

決定された説明変数の適合度をみるために適中率 Hit R を計算する.

$$\text{Hit } R = \frac{\sum_i \delta_{i, S_i}}{N}$$

ただし, $\delta_{i, j}$ は Kronecker のデルタ関数

$$\delta_{i, j} = \begin{cases} 1 & i = j \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

また,

$$S_i = \begin{cases} 1 & p_i > .5 \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

である. 更に, 倒産企業のみの中中率

$$\text{Hit } R_d = \frac{\sum_i \delta_{i, S_i}}{N_d} \quad \text{和は倒産企業のみとする}$$

及び非倒産企業の適中率

$$\text{Hit } R_n = \frac{\sum_i \delta_{i,S_i}}{N_n} \quad \text{和は非倒産企業のみでとる}$$

を計算する。ただし、 N_d, N_n はそれぞれ倒産企業件数、非倒産企業件数である。
尤度比または McFadden 係数といわれる

$$\rho^2 = 1 - \frac{l(\hat{\beta}, \hat{\beta}_{M+1})}{l(0, 0)}$$

を計算する。さらに、自由度で補正した

$$\hat{\rho}^2 = \left(\frac{N - M + 1}{N} \right) \rho^2$$

を計算する。

3. データ

データはすべて中小企業信用リスク情報データベース運営協会のものである。CRD は会員（信用保証協会、政府系金融機関、民間金融機関）が有する取引先中小企業の財務非財務データやデフォルトデータを全て匿名の形で取り込み、蓄積するとともに、それらデータを基に統計的な分析を行い、蓄積された企業データや統計的分析の結果導かれる統計的情報など中小企業の信用リスクの定量化に資する情報を、個別企業を識別することができない形でユーザに提供する機関である(CRD(2001))。CRD の中小企業の定義は中小企業基本法(表 1)に基づくものになっている。

提供されたデータは大別すると、法人の財務データと企業性個人の財務データであったが、法人のデータのみを用いた。年にもよるが、企業数は 20 万件程度である。

データの項目については、使用した法人データについては、表 2 に示してあるような項目数がある。残念ながらすべての項目が埋まっていない企業が 4 割ぐらいある。ただし、今回分析に用いた財務指標を作るのに支障をきたしたのは 1 割以下である。

表 1. 中小企業の定義「改正中小企業法（平成 11 年 12 月 3 日に公布施行）。

	製造業その他	卸業者	小売業	サービス業
資本金	3 億円以下	1 億円以下	5 千万円以下	5 千万円以下
従業員	300 人以下	100 人以下	50 人以下	100 人以下

- 資本金基準と従業員基準のどちらか一方を満たせば中小企業となる。
- 政令によりゴム製造業は、資本金 3 億円以下または従業員 900 人以下、旅館業は資本金 5 千万円以下または従業員 200 人以下、ソフトウェア業情報処理サービス業は、資本金 3 億円以下または従業員 300 人以下を中小企業とする。

表 2. 項目。

項目数の指標	属性・定性 決算書属性	決算書			データ 項目数
		B/S	P/L 他	合計	
全体のデータ項目数	9	62	31	93	102
必須項目数	9	26	13	39	48

表3. 企業規模区分.

業種	指標	1	2	3	4
		区分毎の基準			
建設業	総資本額	0~3億円	3~7億円	7~13億円	13億円以上
製造業	従業員数	1~20人	21~50人	51~100人	101人以上
卸売業	従業員数	1~5人	6~20人	21~50人	51人以上
小売業	従業員数	1~5人	6~10人	11~20人	21人以上
運輸通信業	従業員数	1~20人	21~50人	51~100人	101人以上
不動産業	従業員数	1~20人	21~50人	51~100人	101人以上
サービス業	従業員数	1~5人	6~10人	11~20人	21人以上
飲食店	従業員数	1~5人	6~10人	11~20人	21人以上

表4. 財務指標.

総資本当期利益率	固定負債キャッシュフロー倍率
総資本営業利益率	有利子負債キャッシュフロー倍率
総資本利払後事業利益率	現預金費率
総資本経常利益率	支払準備率
企業利潤率	預借率
自己資本当期利益率	当座比率
自己資本営業利益率	流動比率
自己資本利払後事業利益率	正味運転資本額
自己資本経常利益率	純運転資本額
売上高総利益率	正味運転資本率
売上高当期利益率	売上債券対買入債務比率
売上高営業利益率	棚卸資産対買入債務比率
売上高利払後事業利益率	棚卸資産対正味運転資本率
売上高経常利益率	受取手形割引率
総資本回転率	受取手形割引高回転日数
固定資産回転率	固定負債対有形固定資産比率
有形固定資産回転率	固定比率
流動資産回転日数	固定長期適合率
売上債券回転日数	自己資本額
受取手形回転日数	自己資本比率
売掛金回転日数	負債比率
買入債務回転日数	有利子負債額
支払手形回転日数	有利子負債対自己資本比率
買掛金回転日数	借入金依存度
棚卸資産回転日数	有利子負債利率
売上高原価率	インスタントカバレッジ
売上高販管費率	減価償却率
売上高減価償却費率	前年比増収率
売上高純金利負担率	総資本増減率
売上高支払利息・割引料率	固定資産増減率
営業利益支払利息率	利払後事業利益増加率
償却前内部留保率	経常利益増加率
投融資利回り	当期利益増加率
レバレッジ倍率	自己資本増加率
1人あたり売上高	借入金月商倍率
1人あたり利払後事業利益	経常収支比率
1人あたり総資本	債務償還年数
実質利益額	
キャッシュフロー	

デフォルトについて関係する項目は、初回延滞発生年月日(原則3月以上の延滞)、実質破綻発生年月日(民間金融機関において自己査定上の区分と同様)、破綻発生年月日(法的、形式的な破綻)、代弁発生年月日(信用保証協会において代弁済が実行された年月日)がある。これら一つでも発生していればその企業はデフォルト(倒産)したものとみなした。更に上記の項目の内、最も時期の早い日付をデフォルトした時点とした。CRDのデータとしては、デフォルト発生後の企業データも含まれているが、今回の分析にはそれらのデータは使わなかった。

今回の分析では、企業の規模別の分析を行うが、分析の際の企業区分はCRDの区分(表3)に従った。

4. 結果と考察

財務指標として解析に用いたものは、表4に示したものである。表5は、2000年と1999年のデータを規模に関わらず、すべて用いて分析した結果である。そして、同じデータから各規模のデータのみを取り出して分析を行った結果を表6, 7, 8, 9に示す。表で額に関する項目は、 β の値を1000倍して示してある。

当然のことながら、選ばれた変数は全て t 値が 1.98 より大きいものである。しかし、 t 値が十分大きいからと言って、その説明変数がデフォルト確率に大きく影響を与えるとは限らない。 t 値が大きいと言うのは、 β が零でないことが十分に確からしいことを表しているにすぎないからである。そこで、企業 i の注目している説明変数を x_j^i とするとき

表 5. 結果：全て。

	β	t 値
β_{M+1}	$3.24 \times 10^{+0}$	93.26
預借率	$3.19 \times 10^{+0}$	42.45
減価償却率	8.79×10^{-1}	17.24
総資本回転率	9.48×10^{-2}	11.92
自己資本額	7.79×10^{-8}	8.53
総資本利払後事業利益率	8.24×10^{-1}	6.63
1人あたり利払後事業利益	6.73×10^{-7}	5.16
当期利益増加率	8.36×10^{-5}	3.65
経常収支比率	1.22×10^{-2}	2.83
売上債券対買入債務比率	7.94×10^{-4}	2.18
売上高減価償却費率	4.07×10^{-1}	2.11
γ_2	-8.08×10^{-2}	-3.17
総資本営業利益率	-7.06×10^{-1}	-5.16
正味運転資本額	-6.83×10^{-9}	-6.76
γ_1	-3.28×10^{-1}	-11.50
受取手形割引率	-6.49×10^{-1}	-17.10

企業件数: 倒産 9555 件, 非倒産 400548 件, 合計 410103 件

説明変数数: 16

$$l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma}) = -4.28 \times 10^{+4}$$

$$-2\{l(0, 0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma})\} = 4.83 \times 10^{+5}$$

$$\text{Hit } R = 0.977, \text{ Hit } R_n = 1.00, \text{ Hit } R_d = 0.00$$

$$\rho^2 = 0.850, \bar{\rho}^2 = 0.850$$

表 6. 結果：規模 1 .

	β	t 値
β_{M+1}	$3.40 \times 10^{+0}$	70.63
預借率	$2.74 \times 10^{+0}$	31.88
減価償却率	6.88×10^{-1}	12.45
総資本回転率	8.51×10^{-2}	8.90
自己資本額	2.92×10^{-7}	6.66
キャッシュフロー	6.84×10^{-7}	6.58
1人あたり総資本	8.42×10^{-8}	5.46
正味運転資本額	7.78×10^{-8}	4.14
当期利益増加率	1.06×10^{-4}	3.07
実質利益額	5.40×10^{-8}	2.86
固定長期適合率	1.39×10^{-4}	2.14
有形固定資産回転率	-3.52×10^{-5}	-2.19
γ_2	-1.22×10^{-1}	-2.58
有利子負債キャッシュフロー倍率	-1.06×10^{-5}	-3.11
純運転資本額	-1.46×10^{-7}	-5.17
γ_1	-5.02×10^{-1}	-10.44
受取手形割引率	-5.61×10^{-1}	-11.32

企業件数: 倒産 6295 件, 非倒産 249259 件, 合計 255554 件

説明変数数: 17

$$l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma}) = -2.80 \times 10^{+4}$$

$$-2\{l(0, 0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma})\} = 2.98 \times 10^{+5}$$

$$\text{Hit } R = 0.975, \text{ Hit } R_m = 1.00, \text{ Hit } R_d = 0.00$$

$$\rho^2 = 0.842, \bar{\rho}^2 = 0.842$$

$$\langle \beta_j x_j \rangle \equiv \frac{\beta_j \sum_i x_j^i}{N}$$

の値が十分の大きさであるか, また説明変数の標準偏差がどの程度であるかを吟味する必要がある. この点に留意して, いくつかの選ばれた説明変数について具体的に考察をする.

選ばれた説明変数はほとんどが財務比率に関するものである. このことから, 「規模が小さければデフォルトしやすい」という仮説が正しくないことが分かる. ただし, 自己資本額が説明変数に入っていることから「規模が小さければ倒産しやすい」という仮説を完全には否定出来ない. しかし, 自己資本額の平均は規模 1 から 4.22×10^4 , 6.00×10^4 , 1.56×10^5 , 9.16×10^5 であり(実際の自己資本額はこの値の千倍), β の値が 10 のマイナス 6 乗から 8 乗と小さいことから, 倒産確率にあまり効いていない. このことから企業の倒産を見極める上で, 規模に関する変数は重要でない結論できる.

預借率 (= 現預金 / (借入金 + 受取手形割引高)) が説明変数として上位にあがっているのは, 現預金が多く, 借入金が少ない会社は倒産しにくいことを意味しており, 常識に一致する. β は 3 程度の値であり, 有意水準 1% でお互いに値が異なるという仮説を棄却することができる. 説明変数の平均の値が規模 1 から順に 0.59, 0.99, 1.0, 1.7 であるから, デフォルト確率にかなり効いている. 標準偏差に関して言えば, 81, 107, 44, 140 である. このことから, 預借率は倒産企業を見分ける上でかなり有用な変数である.

総資本回転率は, β の値が規模 3 の企業において 1.03 と大きく, そのほかでは 0.1 のオー

表 7. 結果：規模 2 .

	β	t 値
預借率	$2.73 \times 10^{+0}$	13.83
総資本利払後事業利益率	$1.98 \times 10^{+1}$	10.19
有利子負債利率	$5.80 \times 10^{+0}$	9.77
借入金依存度	9.21×10^{-1}	9.32
償却前内部留保率	5.93×10^{-1}	8.43
自己資本比率	6.40×10^{-1}	7.44
β_{M+1}	$6.43 \times 10^{+0}$	6.81
減価償却率	8.58×10^{-1}	5.83
総資本回転率	1.52×10^{-1}	5.62
売上高減価償却費率	$3.80 \times 10^{+0}$	4.11
現預金費率	$1.01 \times 10^{+0}$	3.84
売上高総利益率	$-3.05 \times 10^{+0}$	-3.27
売上高原価率	$-3.72 \times 10^{+0}$	-3.97
γ_1	-3.92×10^{-1}	-5.40
売掛金回転日数	-2.54×10^{-3}	-6.71
受取手形割引率	-6.41×10^{-1}	-7.64
流動資産回転日数	-7.58×10^{-4}	-7.87
総資本営業利益率	$-1.98 \times 10^{+1}$	-10.13

企業件数: 倒産 1678 件, 非倒産 72938 件, 合計 74616 件

説明変数数: 18

$$l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma}) = -7.20 \times 10^{+3}$$

$$-2\{l(0, 0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma})\} = 8.90 \times 10^{+4}$$

$$\text{Hit } R = 0.977, \text{ Hit } R_n = 1.00, \text{ Hit } R_d = 0.00$$

$$\rho^2 = 0.861, \bar{\rho}^2 = 0.861$$

ダー程度である。正確に言えば、規模 1 と 2 の間では有意水準 1% で値が異なるという仮説が棄却され、それ以外の規模との間では互いに有意水準 1% で値が異なるという仮説が採択される。説明変数の平均は規模 1 から順に 0.84, 1.8, 1.7, 1.7 であり、標準偏差は 18, 3.7, 2.9, 3.0 である。このことから、規模 1 (小規模) の企業について言うと、標準偏差が大きいためデフォルト確率に良く効いている。しかし、規模 2 の企業については平均が $\langle \beta_j x_j \rangle \sim .2$ 程度で分散が 3.7 ということであまりデフォルト確率に効いていない。規模 3 と 4 の企業については、 $\langle \beta_j x_j \rangle \sim 1$ であるから、この規模の企業についてはデフォルト確率に良く効いていることが分かる。この様に、総資本回転率は企業の規模によってデフォルト確率に対する効き方に違いがある。

減価償却率については、 β の値が 0.688, 0.858, 2.26, 3.47 と規模とともに大きくなる傾向がある。規模 1 と 2, 3 と 4 の間では有意水準 1% で値が異なるという仮説が棄却されるが、その他の間では 1% で値が異なるという仮説が採択される。説明変数の平均は規模 1 から順に 0.060, 0.17, 0.16, 0.14 であり、標準偏差が 0.14, 0.32, 0.30, 0.26 である。このことから、減価償却率は倒産確率にはあまり効いていない変数であることが分かる。

(受取)手形に関するデータ欠損フラグ γ_1 については、規模の小さい規模 1, 2 の企業に負で効き、規模 3 には正で効いて、規模 4 の企業については説明変数に選ばれていないことが面白い。これは、規模の小さい企業では、手形による決済をしている(手形を受け取っている)ことがある程度企業の信用力を表していると言える。しかし、ある程度の規模の企業について言

表 8. 結果：規模 3 .

	β	t 値
総資本回転率	$1.03 \times 10^{+0}$	20.41
自己資本額	1.41×10^{-6}	9.34
売上高減価償却費率	$1.60 \times 10^{+1}$	8.99
預借率	$3.86 \times 10^{+0}$	8.00
借入金月商倍率	$5.96 \times 10^{+0}$	7.78
減価償却率	$2.26 \times 10^{+0}$	6.77
γ_1	5.23×10^{-1}	6.04
現預金費率	$4.93 \times 10^{+0}$	5.95
1人あたり売上高	4.20×10^{-7}	3.63
償却前内部留保率	3.72×10^{-1}	3.26
前年比増収率	6.06×10^{-1}	3.12
当期利益増加率	2.95×10^{-4}	2.87
受取手形割引率	2.44×10^{-1}	2.42
流動資産回転日数	-8.25×10^{-4}	-3.09
実質利益額	-1.32×10^{-7}	-3.29

企業件数: 倒産 797 件, 非倒産 37293 件, 合計 38090 件

説明変数数: 15

$$l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma}) = -3.70 \times 10^{+3}$$

$$-2\{l(0, 0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma})\} = 4.54 \times 10^{+4}$$

$$\text{Hit } R = 0.979, \text{ Hit } R_n = 1.00, \text{ Hit } R_d = 0.00$$

$$\rho^2 = 0.860, \hat{\rho}^2 = 0.859$$

うと、手形による決済は日常的であり、デフォルトと関係しない。更に、規模 3 の企業の様にある大きさの企業については倒産し易いことさえある。このことから、変数 γ_1 は、デフォルト確率に非線形に効く変数であることが分かる。

以上から、中小企業の倒産を説明するための変数は、規模によらないものもあるが、規模に依存しその説明力が変わるものもあることが分かった。

選択された説明変数と β の符号を見ると、財務指標分析の常識と一致しないものがあるが見受けられる。例を挙げるならば、固定長期適合率である。一般に固定長期適合率は優良企業ほど値が低いとされているが、今回の結果では値が低いほど倒産しやすいという結果になっている。これは倒産寸前の企業は、資金繰りが悪化し、資産の売却を進めるために起こる現象としてとらえられる(白田 1999)。この様に、倒産直前の企業においては財務指標分析の常識が通用しないことが知られている。そのため、今回得られた結果について、その説明変数の符号条件を確かめることは難しい。しかし、筆者が確認した限りでは、明らかにおかしいと思われるものはなかった。尚、最終的に残らなかった変数の中で、一般に用いられる財務指標であるが分母がマイナスとなる可能性がある指標(例えば自己資本利益率)については、符号条件が常識と異なることが多いことが分かった。

今回の分析に用いた財務指標はその分布が正規分布をしていることを仮定しているものではなく、更に言えば、仮定できるものでもない。財務指標で使われる指標のいくつかは、正規分布をしていないことはよく知られた事実である(Takayasu and Okuyama(1998))。表 10, 11, 12, 13, 14 に相関係数を載せておくが、これらの解釈は上記の理由から難しい。しかし、ある程度の正規性を仮定するならば、相関の高い変数のペアがあることには問題がある。Logit

表 9. 結果：規模 4 .

	β	t 値
自己資本比率	$2.54 \times 10^{+0}$	20.82
借入金依存度	$2.71 \times 10^{+0}$	20.22
預借率	$3.46 \times 10^{+0}$	11.12
総資本回転率	4.61×10^{-1}	10.49
減価償却率	$3.47 \times 10^{+0}$	6.91
総資本利払後事業利益率	$1.99 \times 10^{+1}$	6.41
売上高減価償却費率	$7.02 \times 10^{+0}$	5.43
支払準備率	6.42×10^{-1}	4.28
自己資本額	5.07×10^{-8}	4.13
経常収支比率	2.62×10^{-1}	3.63
売上債券対買入債務比率	2.74×10^{-2}	2.58
当期利益増加率	6.02×10^{-4}	2.57
正味運転資本額	-3.60×10^{-8}	-3.98
流動比率	-5.23×10^{-2}	-4.40
総資本営業利益率	$-1.91 \times 10^{+1}$	-6.13

企業件数: 倒産 747 件, 非倒産 38386 件, 合計 39133 件

説明変数数: 15

$$l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma}) = -3.34 \times 10^{+3}$$

$$-2\{l(0, 0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1}, \hat{\gamma})\} = 4.75 \times 10^{+4}$$

$$\text{Hit } R = 0.980, \text{ Hit } R_n = 1.00, \text{ Hit } R_d = 0.00$$

$$\rho^2 = 0.877, \bar{\rho}^2 = 0.877$$

モデルのような一般化線形モデルにおいて、相関の高い変数を説明変数に用いるといわゆる多重共線性の問題が発生し、予測パラメータが不安定となる。そのため、相関の高い変数のパラメータの値や符号条件については信用できず、またその考察も意味のあるものではない。しかし、一般化線形モデルを予測に用いる場合、予測される企業の財務データが、この相関表から有意に相違しないデータである限り、多重共線性の影響は予測結果に影響しない。例えば、ある財務変数が常に他の財務変数の 2 倍程度であった場合、予測される企業の財務変数も同じく 2 倍の関係を持っているならば、予測結果に影響しないということである。このような場合一方の変数を恣意的に除く方法も考えられるが、より正確には主成分分析や共分散構造を用いた潜在変数モデル等を用いて、合理的に多重共線性を除く分析を行うべきである。この手法については、今後の研究により明らかにしたい。

$-2\{l(0, 0) - l(\hat{\beta}, \hat{\beta}_{M+1})\}$ の値から、どの分析結果においても有意水準 1% で帰無仮説 H: すべての説明変数が零であることを棄却できる。

尤度比 ρ 及び自由度補正をした $\bar{\rho}$ であるが、今回の分析では分析に用いた企業数が多いため両者はほとんど変わらないものとなっている。またこれらの値は、1 に近いほどモデルの適合度が高いことを表している。更に、0.2-0.4 程度で十分高い適合度であると言われている。このことから今回の結果は十分に高い適合度を示していると言える。

適中率に関しては、本分析のように非デフォルト企業が極端に多い場合、確率 50% 点で 2 分類する方法は不適切であるかもしれない。今後、エントロピー理論を応用したモデルの評価方法について検討したいと考えている。

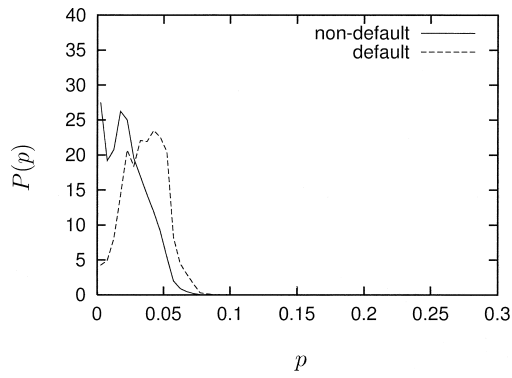


図 1. 倒産確率(全て).

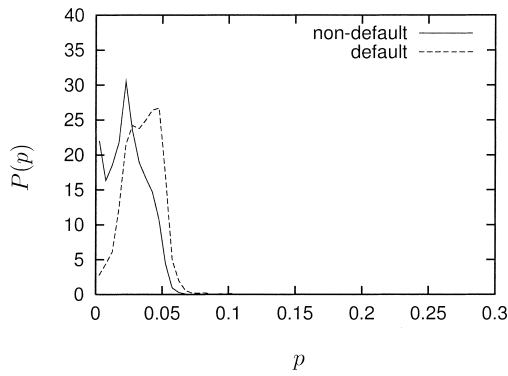


図 2. 倒産確率(規模 1).

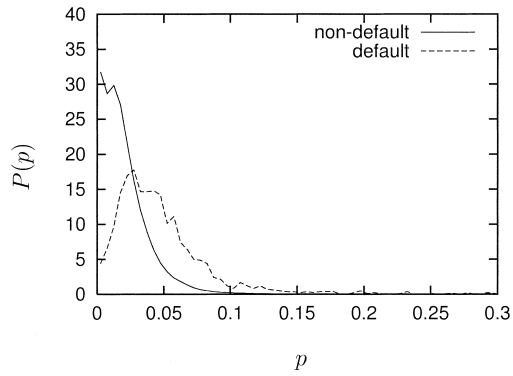


図 3. 倒産確率(規模 2).

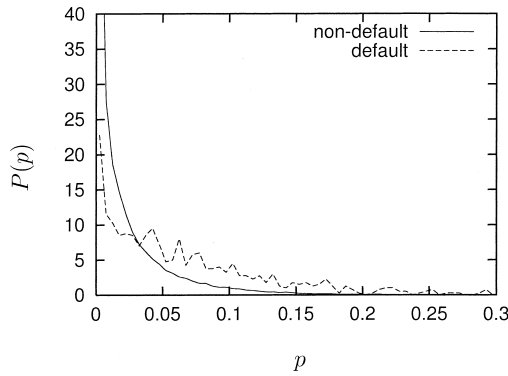


図 4. 倒産確率(規模 3).

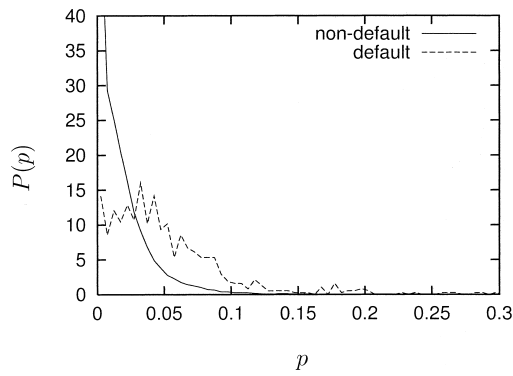


図 5. 倒産確率(規模 4).

補足としてデフォルト企業と非デフォルト企業の分布をデフォルト確率を横軸にとった企業の密度グラフにまとめた(図 1, 2, 3, 4, 5). 実線が非デフォルト企業の密度を表しており, 波線がデフォルト企業の密度を表している. グラフから分かるようデフォルト企業は非デフォル

ト企業よりもデフォルト確率が高い企業の割合が多い。また、分析データの非デフォルト企業の割合が数パーセントであることに対応して、デフォルト企業もデフォルト確率が数パーセントのところによく集まっていることが分かる。

5. まとめ

本論文の目的は、大規模データを用いたデフォルト確率の推定である。我が国においてはこれまで信用リスクを分析することを目的とした中小企業データがなく、大規模データを用いた信用リスクに関する研究は近年初めて可能になった研究である。大規模なデータを用いる利点は、

- ・ データセグメントを行うとき、各セグメントのサンプル数の減少による分析精度の低下を回避できる、
- ・ より多くの説明変数を安定的にモデルに取り込むことができる、
- ・ 未発見の新たな構造の発見が期待できる、

ことである。一般的な言い方をすれば、より多くの情報を取り出すことができるということである。

今回の大規模データをもとに行った分析の結果、次の結論を得た。

- ・ 中小企業のデフォルトに対して預借率、減価償却率などが説明力が高い。
- ・ 預借率などは規模によらず説明力があるが、規模に依存しその説明力が変わる総資本回転率やデータ欠損値フラグなどの指標もある。
- ・ 分析データの手形に関する欠測値からデフォルトを予測する上で有益な情報を取得することができる。

逆に大規模データ故の論点・問題点も明らかになった。例えば、採用できる説明変数が多いことによる多重共線性の問題や、計算時間などのテクニカルな問題である。また、セグメントによるデメリットが少ないことから、規模以外のセグメントの可能性も検討しなければならない。さらに、説明変数を多く取り込むのではなく、より複雑な構造をもったモデルを検討することもできる。

そこで、今後の課題として以下の点を指摘したい。

- ・ 業種別などのさらに細かいセグメントを検討し、最適セグメントを探索する。
- ・ 欠損値フラグについて 2 種類提案したが、さらに有効なフラグを探索する。
- ・ 指標間の相関関係を考慮し、因子分析や LESREL モデルなどの手法を用いて説明力の高いモデルを構築する。
- ・ データのバイアスについて、今後本研究の結果と他の研究を比較検討し明らかにする。

謝 辞

有益な議論をしていただいた金融庁のかたがたに感謝したい。さらに、レフリーの有益なコメントに感謝する。

参 考 文 献

- CRD(2001). 中小企業信用リスク情報データベース(CRD)整備事業成果報告書, CRD, 東京.
- 土木学会土木計画学研究委員会 編(1995). 『非集計行動モデルの理論と実際』, 土木学会, 東京.
- 木島正明, 小守林克哉(1999). 『信用リスク評価の数理モデル』, 朝倉書店, 東京.
- 森平爽一郎(1999). 信用リスク測定と管理 — 第二回: 定性的従属変数回帰分析による倒産確率の推定 —, 証券アナリストジャーナル, **11**, 81–101.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.
- 坂元慶行, 石黒真木夫, 北川源四郎(1983). 『情報量統計学』, 共立出版, 東京.
- Saunders, A. (1999). *Credit Risk Measurement*, Wiley, New York.
- 白田佳子(1999). 『企業倒産予知情報の形成』, 中央経済社, 東京.
- Takayasu, H. and Okuyama, K. (1998) Country dependence on company size distributions and a numerical model based on competition and cooperation, *Fractal*, **6**, 67–79.

Estimation of Probability of Default Using Credit Risk Database

Hisanao Takahashi

(Department of Statistical Science, The Graduate University for Advanced Studies;
Financial Services Agency)

Satoshi Yamashita

(The Institute of Statistical Mathematics; CRD)

It is important in credit risk management to determine the probability of bankruptcy. Few reliable analyses of bankruptcy have been developed for small and medium-sized enterprises because of the delay in developing of databases to capture credit risks for the enterprises. Recently, a large-scale database for estimating credit risks for such enterprises has become available as “Credit Risk Database”. In this paper, we estimate the probability of bankruptcy by applying the logit model to the data from this database. We use the t value to evaluate the significance of the model’s parameters. We discuss the differences in explanatory factors of credit risk depending on the enterprise scale.