卒論「倒産企業の判別」

この卒論は昭和60年11月18日に提出したものです。

〈昭和42年以降に倒産した企業の財務指標による判別関数の作成と考察〉

①目的

判別関数を使って倒産企業と非倒産企業を判別することが主たる目的。8つの財務比率のうち倒産と非倒産の群を引き離すのにどの指標をどのような割合で結び付け関数を作れば良いかを考えるものであり、1つ1つの指標を独立させて考えるわけではない。

②8つの財務指標

1. 売上高経常利益率　2. 売上高対支払利息率　3. 総資本回転率
4. 売上債権回転率　5. 自己資本比率　6. 流動比率　7. 当座比率
8. 固定比率

③倒産企業の算出

昭和42年以降、昭和60年までの39社を選出。これらはａ更正の申立、ｂ銀行取引の停止、ｃ自己破産の申立、ｄ和議の申立、ｅ会社整理の申立のいずれかで上場廃止になった一部、二部上場の企業。

④非倒産企業

上記の倒産企業に相対する非倒産企業には1. ほぼ同じ時期、2. 規模(資本金)が大きく違わない、3. 業種が同じで事業内容が似ている企業を選び、同じく8つの財務指標を算出した。

⑤判別関数

何十通りものシミュレーションを行い、一番有効であった判別関数式は次の通り。

判別関数＝自己資本比率＋26.187×総資本回転率＋0.838
×売上高経常利益率＋0.151×流動比率－63.136

この値がマイナスであれば倒産企業として判別される。

倒産企業37社のうち26社を倒産と判別、11社を非倒産と判別した。

正しい判別率は70.3％である。

非倒産企業35社のうち8社を倒産と判別、27社を非倒産と判別した。正しい判別率は77.1％である。

⑥銀行の判別度

当時は数十社という少ない企業数で、表面上の決算書で分析したのでこの程度が限界でしたが、それでも、中心線近辺では別紙のグラフの通り正誤が入り乱れるものの、ある幅以上は間違いがなくなっています。

大手銀行はデータ量も違えば資産の中身も実態ベースで分析するため、かなり精緻なものと思われます。

分析の理論と分析方法

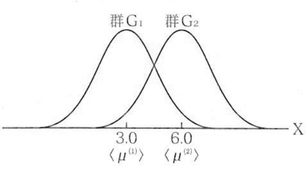
〔分析の理論〕

まず初めに金利負担率という指標を例に、基本的な考え方を明らかにする。ここでは、金利負担率は正規分布するという仮定におくことにする。

非倒産の企業群を群G_1とし、倒産の企業群を群G_2で表す。群G_1の金利負担率の平均$\mu^{(1)}$を3.0％とし、群G_2の金利負担率の平均$\mu^{(2)}$を6.0％とする。また、それぞれの群の分散は等しく$\delta^2 = 1.5^2$であるとする。図1はその密度関数を図示したものである。

非倒産の群G_1の平均$\mu^{(1)}$と倒産の群G_2の平均$\mu^{(2)}$とが完全に離れていれば、判別において誤ることはないが、ほとんどの場合には図1のように一部重なり合っている。この重なり合っている部分の値をとる企業は群G_1からであると判定しても群G_2からであると判定しても、過ちを犯す確率をゼロにすることはできない。そこで今、金利負担率という指標に基づいて、この誤って判定する確率を最小にするような判別の基準を見出すことを考える。

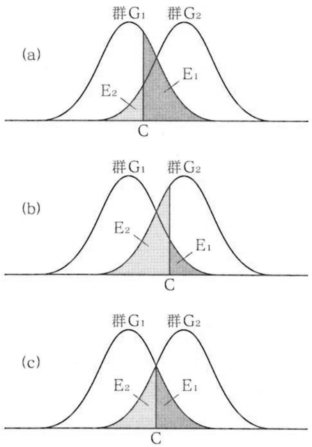
図1：非倒産群G_1と倒産群G_2の金利負担率Xの分布



判別の方法としては、データXが
X＜Cならば X は群G_1からと判定、
X＞Cならば X は群G_2からと判定、
が考えられよう。

図２は、いく通りかの判別の基準点Cに対し、誤判別の確率を図示したものである。面積E_1は本当は群G_1に属するのに、群G_2からのものと判定してしまう誤りを犯す確率、面積E_2は本当は群G_2に属するのに、群G_1からのものと判別してしまう誤りを犯す確率である。

図２：群G_1を誤ってG_2と判定してしまう確率E_1と、群G_2を誤ってG_1と判定してしまう確率E_2



誤判別の確率の和は、

$$E_1 + E_2 \quad \cdots\cdots \quad (1)$$

である。最良の判別の基準値は、この誤判別の確率の和を最小にするものである。図２では(c)のCが、最良の判別点C_0ということになる。

その点C_0は、

$$C_0 = \frac{\text{群}G_1\text{の平均}+\text{群}G_2\text{の平均}}{2} = \frac{\mu^{(1)}+\mu^{(2)}}{2} \quad \cdots\cdots \quad (2)$$

であり、この場合の誤判別の確率は、

$$E_1 = E_2 \quad \cdots\cdots \quad (3)$$

になっている。

判別の基準点C_0から、それぞれの群の平均$\mu^{(1)}$、$\mu^{(2)}$までの距離の平方は、それぞれ

$$(\text{判別点}-\text{群}G_1\text{の平均})^2 = (C_0-\mu^{(1)})^2 \quad \cdots\cdots \quad (4)$$
$$(\text{判別点}-\text{群}G_2\text{の平均})^2 = (C_0-\mu^{(2)})^2 \quad \cdots\cdots \quad (5)$$

である。それを分散δ^2で割ったものが、マハラノビス平方距離の1変数の場合の定義式である。つまり群G_1の平均$\mu^{(1)}$から判別点までのマハラノビス平方距離D_1^2は、

$$D_1^2 = \left(\frac{\text{判別点}C_0-\text{群}G_1\text{の平均}}{\text{標準偏差}}\right)^2 = \left(\frac{C_0-\mu^{(1)}}{\delta}\right)^2 \quad \cdots\cdots \quad (6)$$

とかけ、群G_2の$\mu^{(2)}$からC_0までのマハラノビス平方距離D_2^2は、

$$D_2^2 = \left(\frac{\text{判別点}C_0-\text{群}G_2\text{の平均}}{\text{標準偏差}}\right)^2 = \left(\frac{C_0-\mu^{(2)}}{\delta}\right)^2 \quad \cdots\cdots \quad (7)$$

とかける。金利負担率Xで、誤判別の確率の和を最小とする最良の判別点C_0は、

$$C_0 = \frac{3.0+6.0}{2} = 4.5 \quad \cdots\cdots \quad (8)$$

であり、$\delta = 1.5^2$であったから、それぞれのマハラノビス平方距離は、

$$D_1^2 = \left(\frac{4.5-3.0}{1.5}\right)^2 = 1.0 \quad \cdots\cdots \quad (9)$$
$$D_2^2 = \left(\frac{4.5-6.0}{1.5}\right)^2 = 1.0 \quad \cdots\cdots \quad (10)$$

になる。逆に最良の判別点C_0は、

$$D_1^2 = D_2^2 \quad \cdots\cdots \quad (11)$$

を満たす点として求められる。

正規分布の場合には、

$$\frac{X-\text{平均}}{\text{標準偏差}} \quad \cdots\cdots \quad (12)$$

は平均0、分散1の標準正規分布に従う。点Xから平均μまでのマハラノビス距離Dは、

$$D=\frac{X-\mu}{\delta} \quad \cdots\cdots \quad (13)$$

であるから、これは標準正規分布N(0、1)に従うことになる。

金利負担率の場合はD＝1.0であるから、本当は群G_1に属するのに群G_2からと判定してしまう誤りの確率は、標準正規分布表から0.159となる。同様に群G_2に属するのに群G_1からのものと判定してしまう誤りの確率も0.159である。2群の平均間のマハラノビス距離を判別効率と言い、

$$D=\frac{\mu^{(1)}-\mu^{(2)}}{\delta} \quad \cdots\cdots \quad (14)$$

とかける。分散$\delta^2=1.5^2$としたときの平均の差に対する判別効率と誤判別の確率を表に示す。

判別効率と誤判別の確率($\delta^2=1.5^2$)

平均$\mu^{(1)}-\mu^{(2)}$の値	判別効率D	誤判別の確率
1.0	0.667	0.371
1.5	1.000	0.309
2.0	1.333	0.251
2.5	1.667	0.203
3.0	2.000	0.159

図3：それぞれの群の平均$\mu^{(1)}\cdot\mu^{(2)}$から最良の判別点C_0までのマハラノビス距離

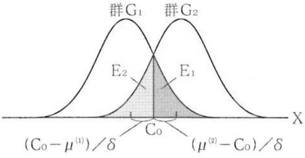


図4：金利負担率による誤判別の確率E_1

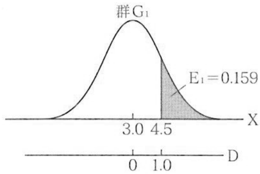
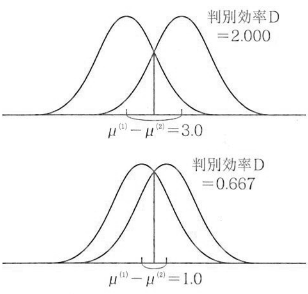


図5：2群の差が1.0と3.0のときの判別効率Dと誤判別の確率



以上、論じられた判別式は線形判別式と呼ばれ、線形判別式の理論は、データが次の２つの前提を満たす条件の下で構成されている。

1．多変量正規分布に従う。

2．両群の分散共分散行列は等しい。

これらの前提条件が満たされていないときは、得られた判別式の信頼性について、より注意深く検討しなければならない。

多変量正規分布という前提条件のズレは、誤判別の確率の大きさのズレになって現れてくる。正規分布が満たされていないときは、誤判別の確率E_1とE_2とは等しくなくなり、誤判別の確率の片方が大きくなったり小さくなったりするというアンバランスが生じる。

同時に、期待される誤判別の確率の和E_1+E_2が、得られた判別式を用いるときに最小になるという保証はなくなる。多変量正規分布から著しくズレる場合には、判別式の信頼性について慎重に検討しなければならない。これが満たされているものとして、前提条件２の説明に移る。

分散共分散行列が等しくない場合には、理論的に導出される判別式は線形判別式

$$Z= A_0+A_1X_1+A_2X_2+ \cdots +A_PX_P \quad \cdots\cdots (15)$$

ではなく、X_1^2、X_2^2、X_1X_2、…等を含む２次式になる。判別式が線形になるのはこの前提条件によるのである。例えば、分散が倍以上も違っているような場合には、線形判別式を用いることに慎重でなければならない。

さて、先程の説明では指標として金利負担率をとり上げ、平均・分散は既知であるとして説明した。実際には、これらの値は未知であり、データから推定することになる。

〔ＷｉｌｋｓのΛ統計量〕

ｑ変量を用いて、ｇ個の群がどの程度判別されるかを表す１つの指標として、ＷｉｌｋｓのΛ統計量がある。

これはq変量についての全体及び群内の平方和積和行列

$$T = \sum_{k=1}^{g}\sum_{i=1}^{nk}(X_i^{(K)} - \bar{X})(X_i^{(K)} - \bar{X})^1、(\bar{X} = {}^1\!/_h \sum_{K=1}^{g}\sum_{i=1}^{nK} X_i^{(K)}) \quad \cdots\cdots (16)$$

$$W = \sum_{K=1}^{g}\sum_{i=1}^{nK}(X_i^{(K)} - \bar{X}^{(K)})(X_i^{(K)} - \bar{X}^{(K)})^1 \quad \cdots\cdots (17)$$

の行列式の比

$$\Lambda = |W| \,/\, |T| \quad \cdots\cdots (18)$$

で定義される。Λは0と1の間の値をとり、0に近いほど、よく判別されていることを表す。

〔各変量の寄与の検定〕

ある特定の変量が判別に寄与しているかどうか検討したい。今、q個の変量X^*がg群の判別に用いられているとする。X^*に含まれていない変量X_iを追加したときの判別力の増加をΛ統計量を用いて、

$$\Lambda(X_i | X^*) = \Lambda(X^*、X_i) / \Lambda(X^*) \quad \cdots\cdots (19)$$

で測ることができる。右辺の分子は変量X^*とX_iを両方用いたとき、分母は変量X^*のみを用いたときのΛを表し、$\Lambda(X_i | X^*)$は偏Λ(partial Λ)統計量と呼ばれる。

このとき、

$$F = \frac{n - g - q}{g - 1} \cdot \frac{1 - \Lambda(X_i | X^*)}{\Lambda(X_i | X^*)} \quad \cdots\cdots (20)$$

は新しく加えた変量X_iが判別に寄与しないという仮説の下で、自由度(g－1、n－g－g)のF分布に従うことが知られている。式(20)のFは、X^*を共変量として変量X_iについて、共分散分析を行ったときの群間の差を検定するF統計量に等しい。

判別分析に利用される変量Xの中の特定の変量の寄与を検定したいときには、検定したい変量をX_i、残りの変量をX^*とおいて式(20)の統計量を求めれば良い。

〔変量の選択〕

先の検定（各変量の寄与の検定）を利用して、逐次的に変数選択を行うことができる。今、変量X^*を用いて判別分析が行われているとする。X^*に含まれていない、ある変量X_iを加えるべきか否かについては$\Lambda(X_i|X^*)$を計算して、

$$F_0 = \frac{n-g-q}{g-1} \cdot \frac{1-\Lambda(X_i|X^*)}{\Lambda(X_i|X^*)} \geqq Fin \quad \cdots\cdots \quad (21)$$

ならばX_iを加え、そうでなければ加えないことにすれば良い。ここに、Finは指定された限界値である。

また、X^*に含まれているある変量X_kを除くべきか否かについては、$\Lambda(X_k|X^*_{(K)})$を計算（ただし、$X^*_{(K)}$はX^*からX_Kを除いた〈$g-1$〉個の変量を表す）し、

$$F_0 = \frac{n-g-q+1}{g-1} \cdot \frac{1-\Lambda(X_k|X^*_{(K)})}{\Lambda(X_k|X^*_{(K)})} < Fout \quad \cdots\cdots \quad (22)$$

ならばX_iを除き、そうでなければ除かないことにすれば良い。Foutも Finと同様、指定された限界値である。