

# 財務諸表データに対する外れ値処理と信用リスク評価モデリング

高橋 淳一

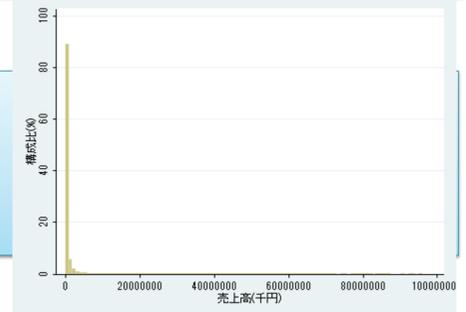
総合研究大学院大学 統計科学専攻 博士後期課程5年

## 問題意識

- 中小企業の経営データを大量に集積したデータベースであるCRD(Credit Risk Database)は、過去20年以上の中小企業の決算書データを2000万件以上蓄積。
  - 中小企業決算書データに関しては、外れ値や異常値が多く、信用リスク評価をモデリングする際には、扱いに注意が必要となる。
- ⇒ 本研究では、中小企業決算書データに対して、モデル精度を向上させるような、外れ値や異常値を適切に処理する方法を提案する。

## 決算書データの性質

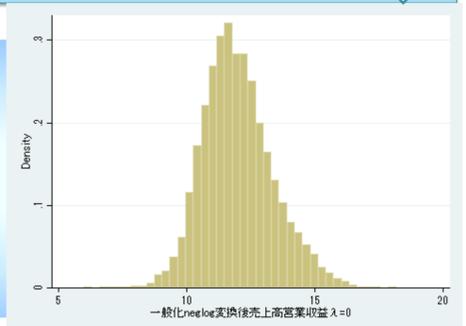
- 決算書データを集積した時、多くの財務項目の分布は、非常に裾の長い分布となっている（右図は売上高の分布）。
- ⇒ 信用リスクモデリングで良く利用される最尤法は、外れ値の影響を強く受けるため、分布の補正が必要となる。



## 分布の補正方法

- 【折り畳み処理】各財務指標  $x_j (j=1,2,\dots,8)$  に対して分布の上下  $\theta\%$  点の値  $\bar{x}_j^\theta, \underline{x}_j^\theta$  を設定し、上下限値  $\bar{x}_j^\theta, \underline{x}_j^\theta$  超の値を  $\bar{x}_j^\theta, \underline{x}_j^\theta$  に折り畳み処理をする。
- 【一般化neglog変換】大きく偏った分布を補正するため、一般化neglog変換を実施。

変換後の分布



## ※一般化neglog変換とは

- 正値の対数変換を一般化したBox-Cox変換と同様に、正負の両方の値に対して対数変換できるneglog変換を一般化したものを一般化neglog変換と呼ぶ。
- 一般化neglog変換の変換率を  $\lambda$  とした時、変換式は以下のようになる。

$$\psi(\lambda_j, x_j) = \begin{cases} \text{sgn}(x_j) \times \frac{(|x_j|+1)^{\lambda_j} - 1}{\lambda_j} & (\lambda_j \neq 0) \\ \text{sgn}(x_j) \times \ln(|x_j|+1) & (\lambda_j = 0) \end{cases}$$

## 比較分析のステップ

- 12000件の完全データを用いて、以下の3パターンで分析を行い、AUCの比較を行った。



比較分析時の折り畳み処理と一般化neglog変換を同時に実施するときの変換式は以下の通り。

$$\psi(\lambda_j, x_j^\theta) = \begin{cases} \text{sgn}(x_j^\theta) \times \frac{(|x_j^\theta|+1)^{\lambda_j} - 1}{\lambda_j} & (\lambda_j \neq 0) \\ \text{sgn}(x_j^\theta) \times \ln(|x_j^\theta|+1) & (\lambda_j = 0) \end{cases}$$

パターン1では $\lambda=1$ で固定、パターン2では $\lambda=0$ で固定し、それぞれ $\theta$ を最適化している。パターン3では $\lambda$ と $\theta$ を同時に最適化している。

## 計算結果とインプリケーション

### 計算結果

最適な  $(\lambda^*, \theta^*)$

j	財務指標名	$\lambda^*$	$\theta^*$
1	総資本当期利益率(ROA)	0.32	-*
2	総資本回転率	-2.45	5.9
3	棚卸資産回転日数	0.62	0.3
4	支払準備率	-0.52	3.1
5	現預金比率	-0.69	2.5
6	自己資本比率	0.26	-
7	デットキャパシティレシオ	-0.31	-
8	預借率	-0.27	1.8
9	売上高支払利息割引料率	-0.12	6.2
10	流動資産その他流動資産比率	0.59	3.9

※ $\theta^*$ の“-”は、最適計算の結果、折り返しなし( $\theta = 0$ )となったことを意味する。

### パターン別AUC比較

パターン	AUC
1 変換なし	0.7865
2 neglog変換	0.8078
3 一般化neglog変換	0.8121

### 【インプリケーション1】

一般化neglog変換が、通常のneglog変換よりもAUC向上という面で有効な変数変換の手法であることが確認できた。

### 【インプリケーション2】

AUCを向上させる一般化neglog変換の変換率は、必ずしも標準正規分布に近づける変換率とは限らないことが確認された（下図参照）。

