

2005年7月29日
SAS Forum ユーザー
学術総会 2005

データマイニングを用いた 中小企業の信用リスクの推定

～ 再構築(Re-Construct)アプローチ ～

UFJ 銀行 法人統括部
小 野 潔

(本報告は個人的見解です)

本研究のポイント

- 中小企業情報に基づく倒産モデル構築
- 再構築 (Re-Construct) アプローチ
- 120万社の中小企業の倒産率の算出
- 再構築モデルの構造
- 情報公開を優遇する加工変数
- 本モデルの有効限界

中小企業の倒産モデル

■ 金融機関の融資対象企業

- 基準例: 売上10億円以上、評点50数点以上
- 今、中小企業の与信対象を売上10億円未満に急速拡大へ

■ 日本の中小企業数 約500万社 (中小企業庁定義)

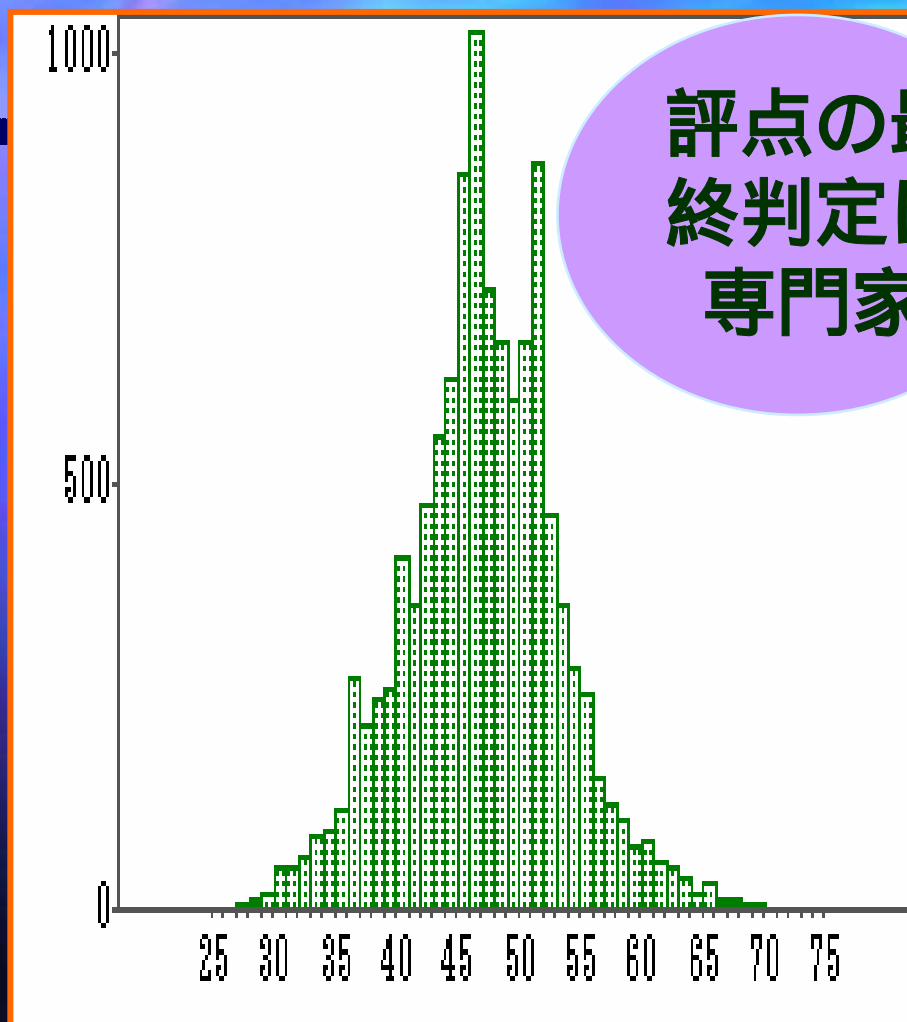
■ 情報ベンダーの企業情報(評点)数 約120～170万社

- 評点と倒産率の関係
- 帝国データバンク(TDB)の倒産率 約40万社
- 問題点: モデルの構造、説明変数がブラックボックス

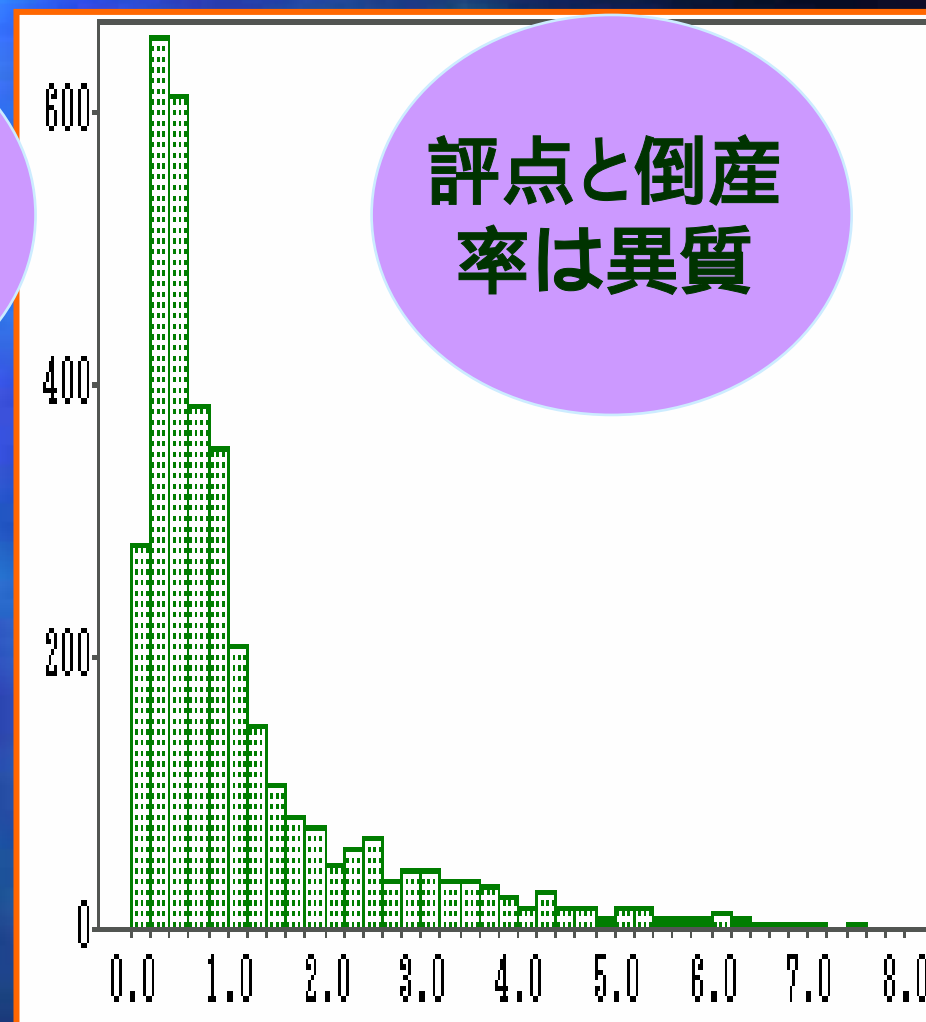
■ 倒産モデル

- 定量モデル 決算書ベース
- 定性モデル 精度は定量モデルより低い 各情報ベンダーが発表

TDBの評点と倒産率の度数分布



評点 120万社



CCRモデル 40万件

再構築 (Re-Construct) アプローチ

- 従来の倒産モデル構築法
 - DataBase: 2 ~ 3年分 Target: 倒産企業
- 再構築法: 予測倒産率から倒産モデルを推定
 - DataBase: 現時点(当月) Target: 倒産率(CCRモデル)
- 分析手法: 構造の解明 決定木
- 仮想倒産企業の定義
 - Black: 高倒産率の企業 White: 低倒産率の企業
- サンプルング比率の変更
 - 倒産企業の特徴を捉えるため Black数: White数=1:1
- 確信度から倒産確率へのベイズ変換
 - サンプルング比率変更による補正
- 変則バギング学習によるモデル安定化

再構築アプローチ (Re-Construct)の実例

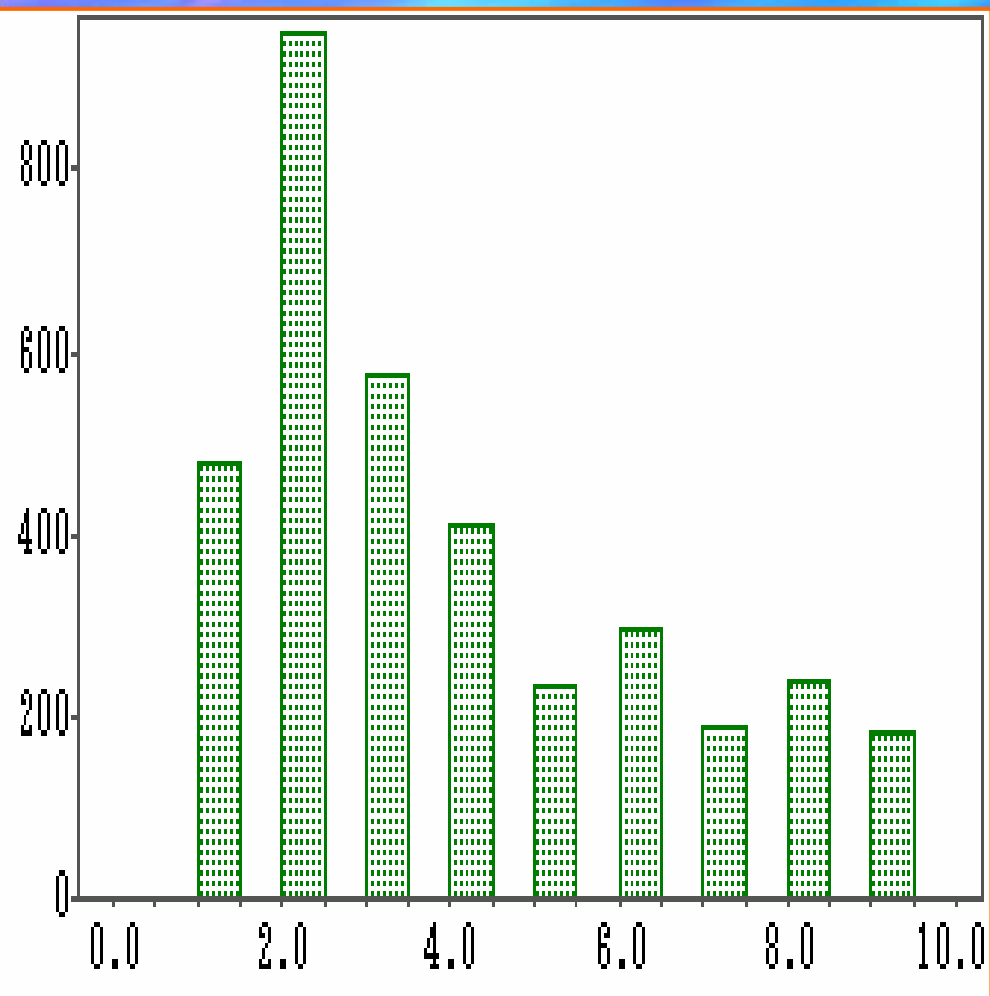
Black : 高倒産率の企業

White : 低倒産率の企業

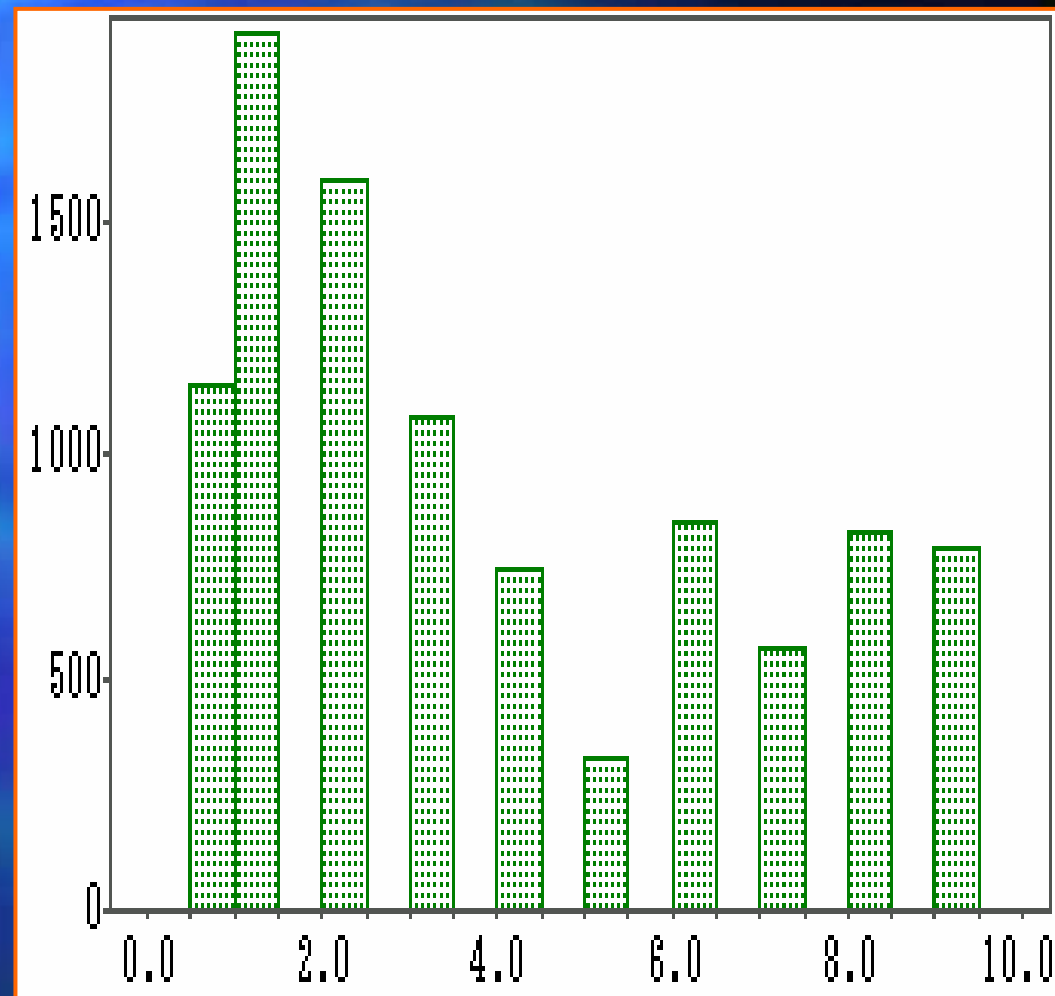
閾値	1.37%	2%	3%	4%	5%	6%	7%	8%	9%	10%
仮想倒産数	103,146	75,819	46,488	28,021	18,392	11,921	8,061	5,653	4,339	3,364
仮想 非倒産数	270,363	297,690	327,021	345,488	355,117	365,448	365,448	367,856	369,170	370,145
仮想倒産率	27.62%	20.30%	12.45%	7.50%	4.92%	3.19%	2.16%	1.51%	1.16%	0.90%
理論倒産率	2.85%	2.62%	2.09%	1.97%	1.68%	1.61%	1.59%	1.64%	1.52%	1.69%

変則バギング学習

倒産格付の分布比較



CCRモデル(TDB社)



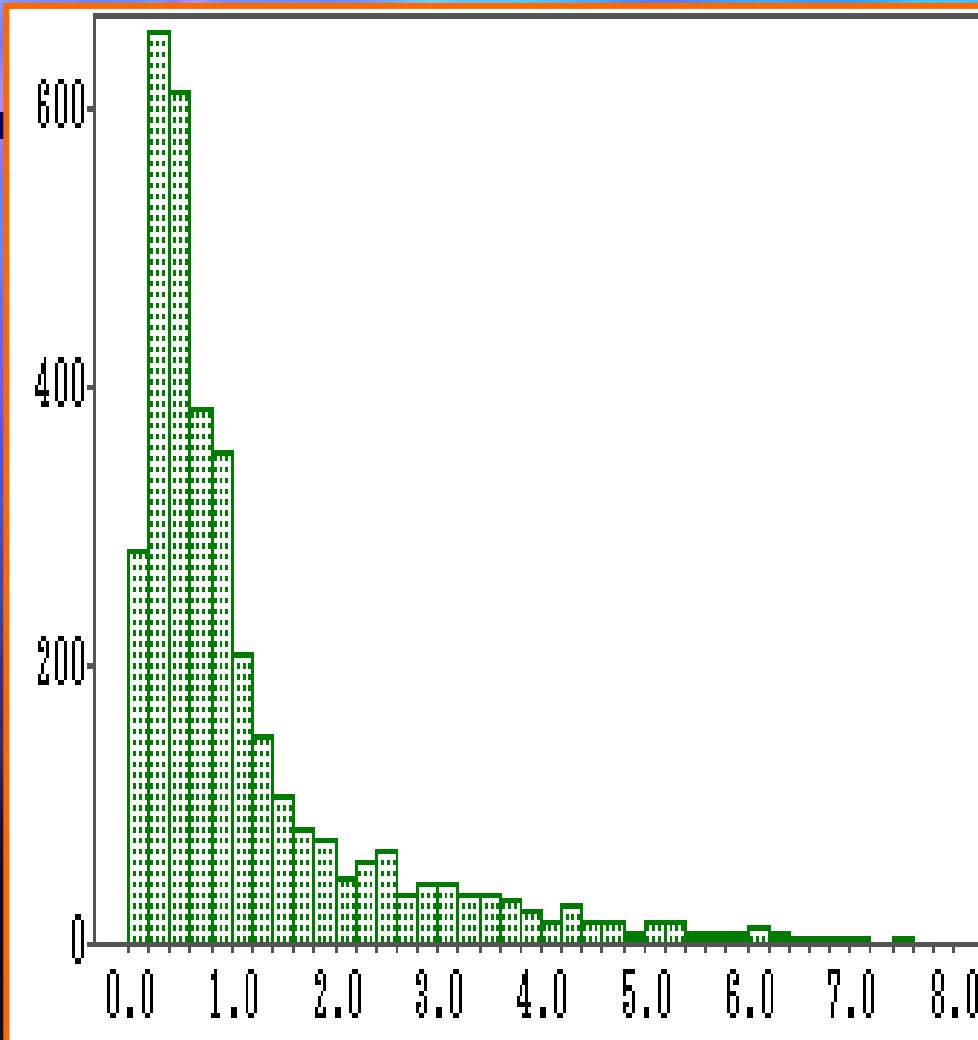
再構築モデル

決定木の構成要素

情報公開企業を
優遇する変数

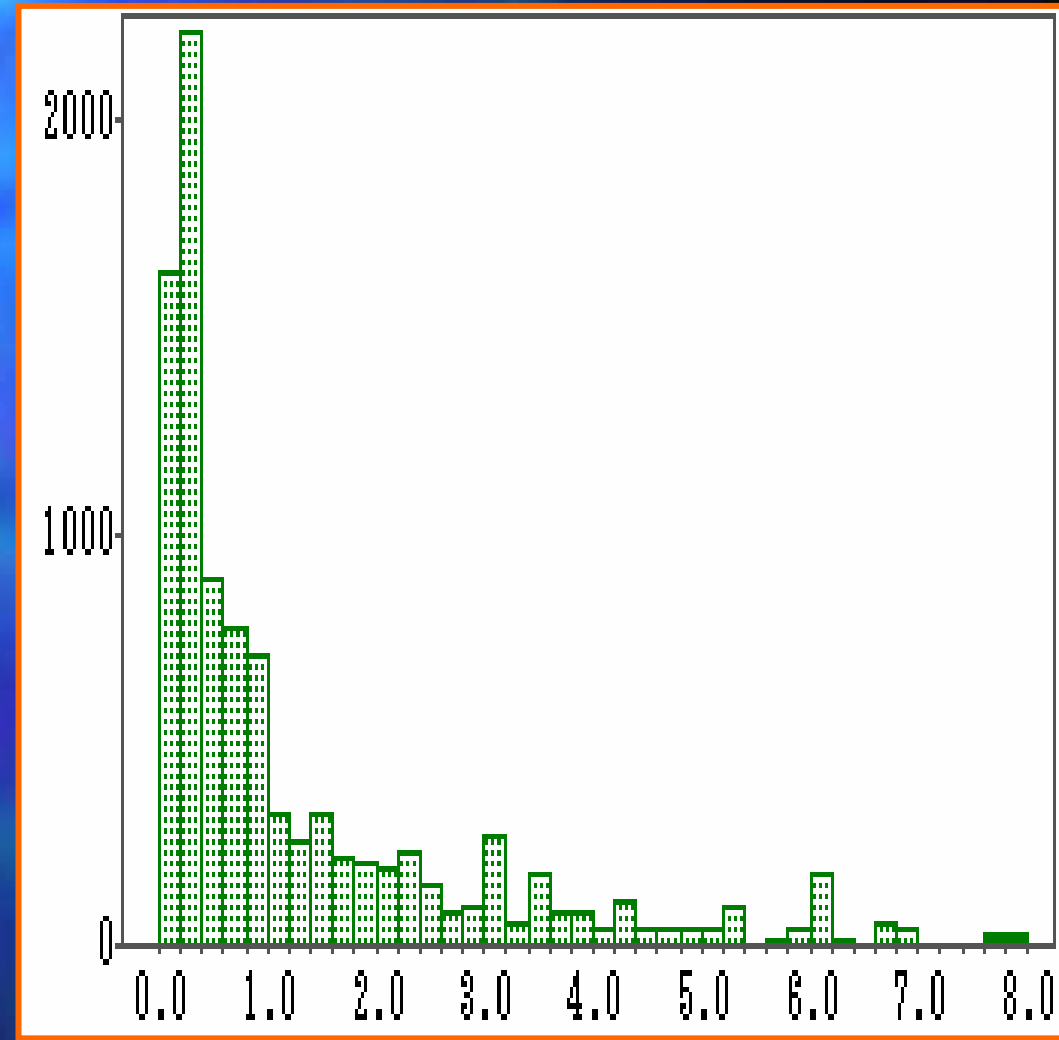
	変 数	平均分割度		変 数	平均分割度
1	評点	5.00	11	株主数	1.39
2	2分類P	3.70	12	最新期業績売上	1.22
3	最新期業績自己資本比率	3.20	13	税引後利益の正数P	1.02
4	税引後利益減少数P	3.10	14	全国業種ランキング相対位置P	0.93
5	法人申告所得額増加数P	2.34	15	前々期からの平均売上高伸率P	0.89
6	自己資本比率増加数P	2.30	16	最新期業績税引後利益	0.64
7	設立年数P	2.16	17	法人格コード	0.17
8	業績利益存在数P	2.01	18	前期売上高伸率P	0.12
9	決算書存在数P	1.54	19	前期平均売上高伸率P	0.07
10	最新期業績決算書有無P	1.41		(変数名にPがつくものは加工変数)	

倒産率の度数分布



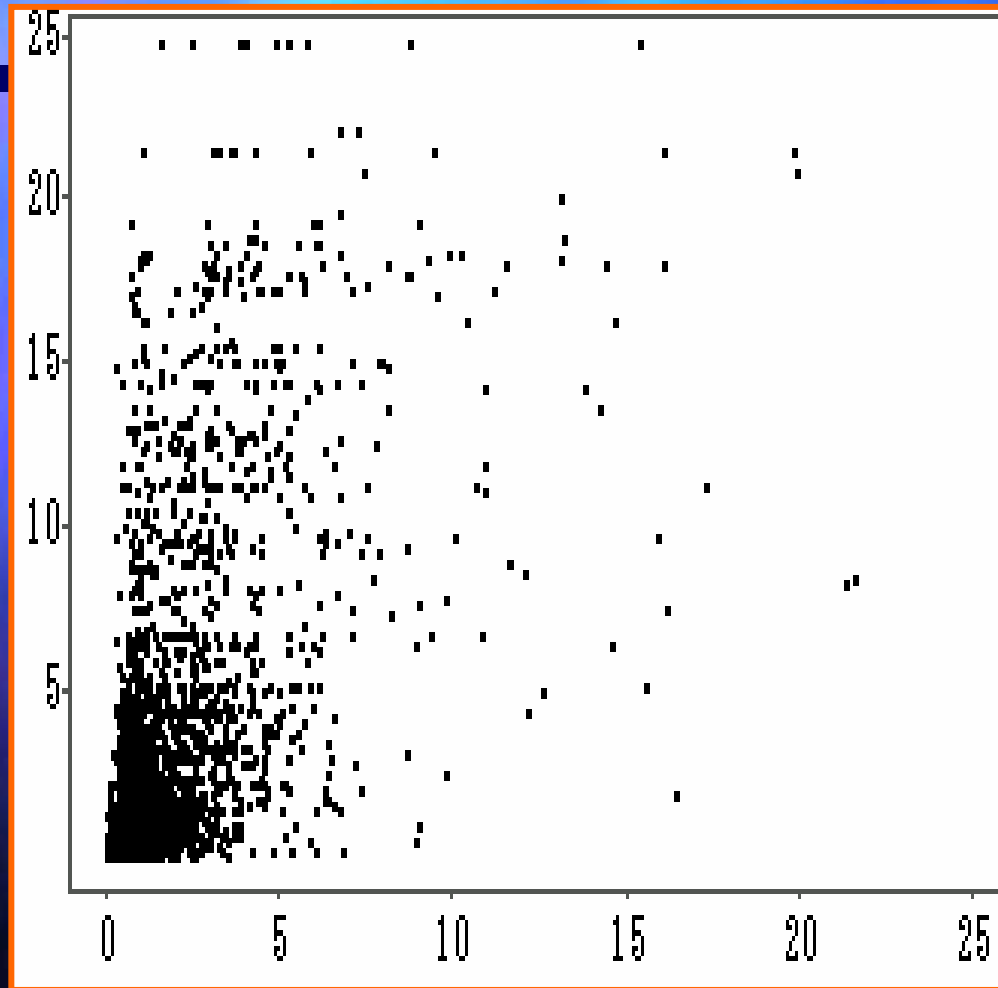
CCRモデル (40万社)

倒産率 (%)



再構築モデル (120万社)

再構築モデルとCCRモデルの関係

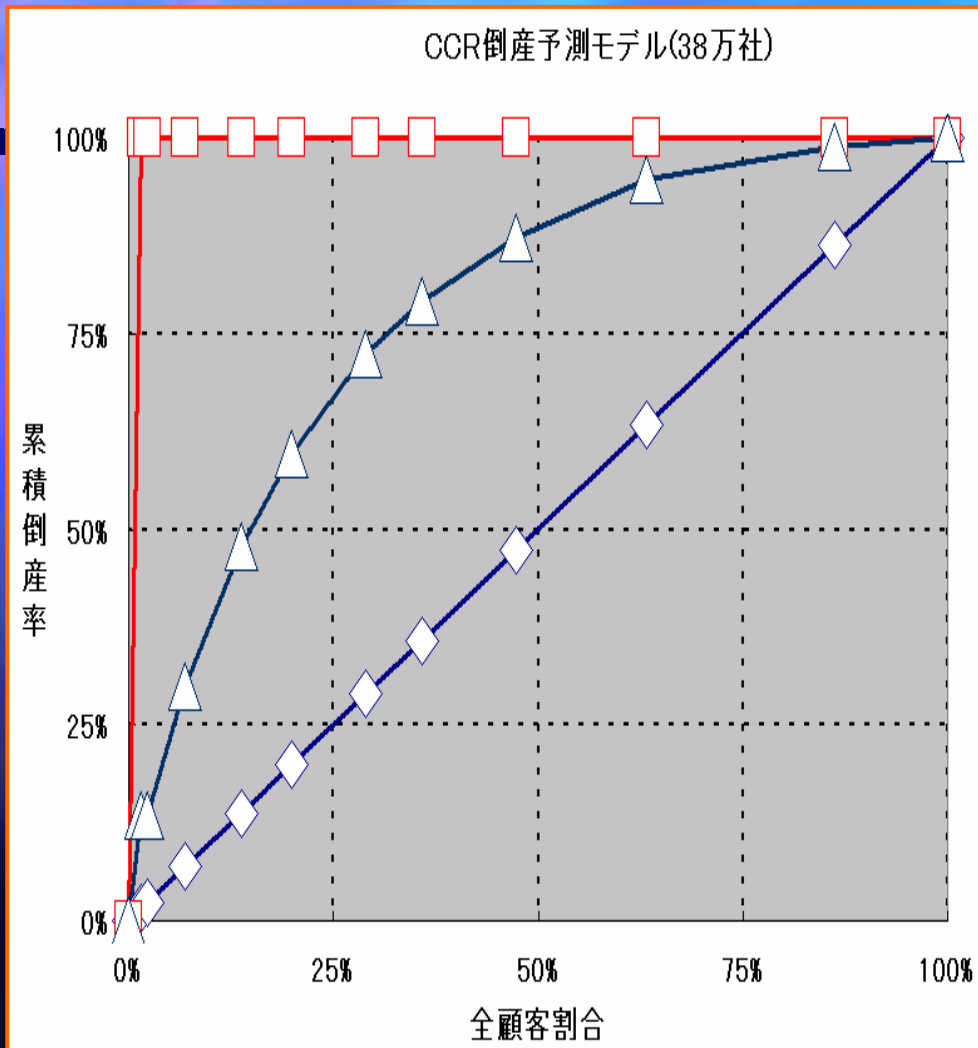


CCR倒産率

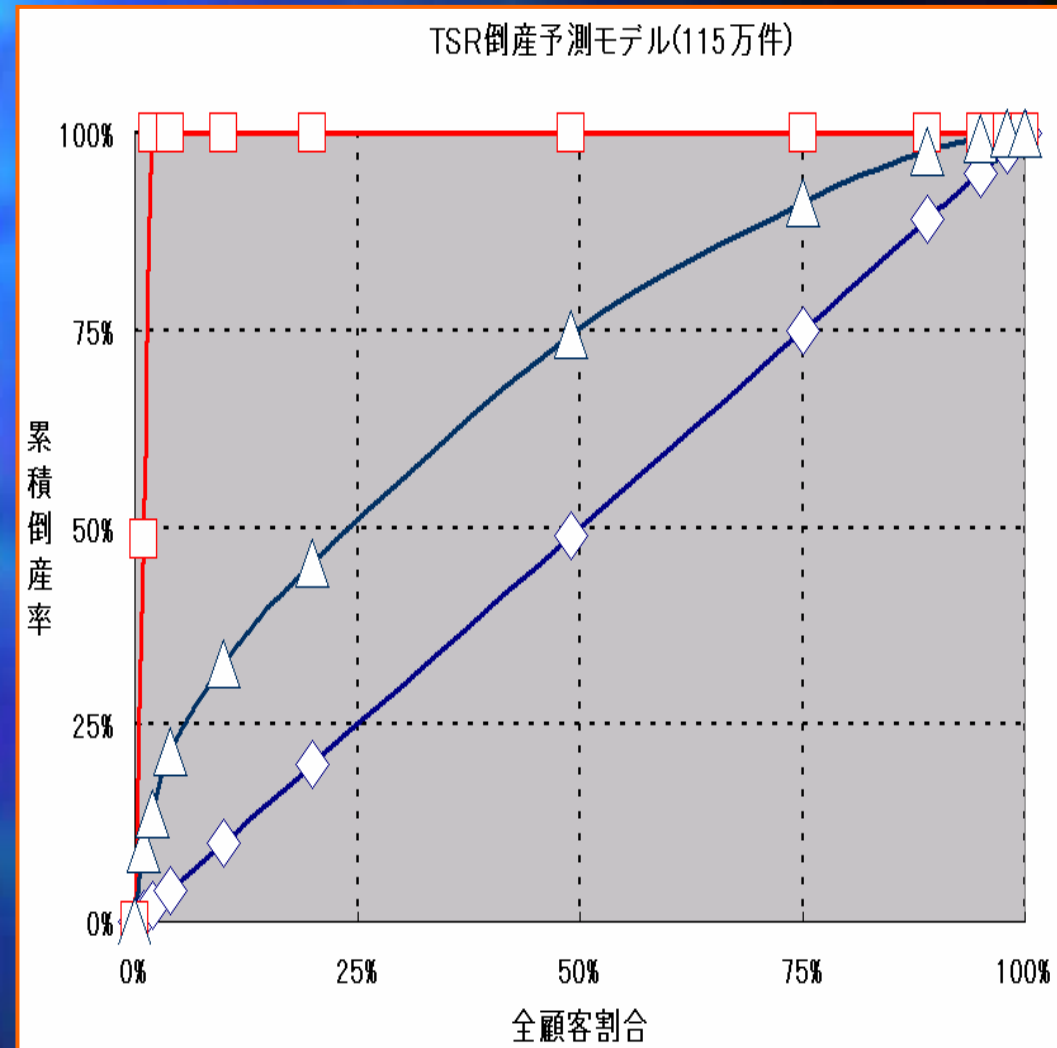
相関係数	CCRモデル	再構築モデル	評点
CCRモデル	1.00	0.60	-0.53
再構築モデル	0.60	1.00	-0.73
評点	-0.53	-0.73	1.00

格付のスペアマン相関係数
= 0.70

情報ベンダーの倒産モデルCap図



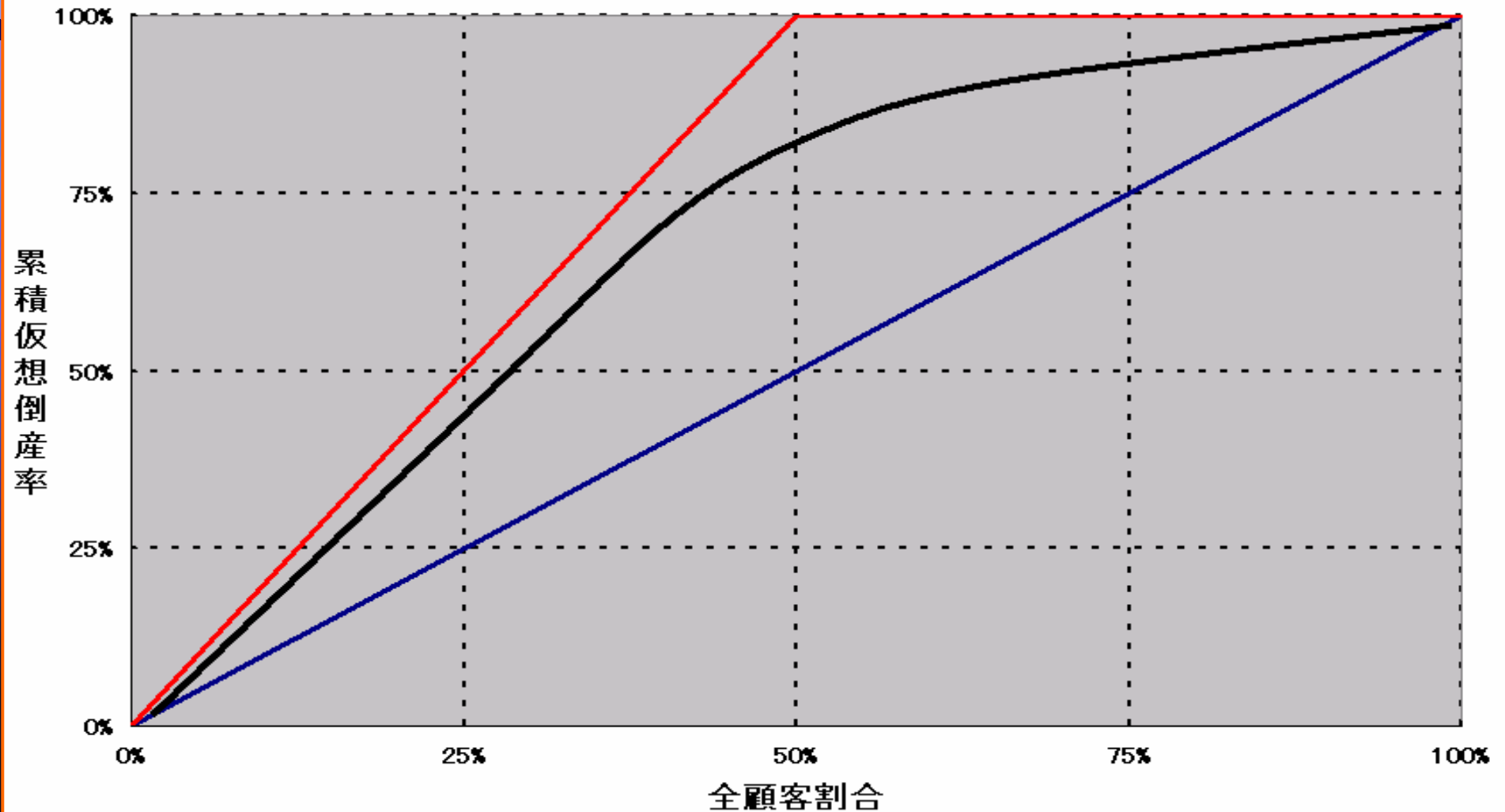
TDB社の倒産モデル
(40万社)



TSR社の倒産モデル
(115万社)

再構築モデルのCap図

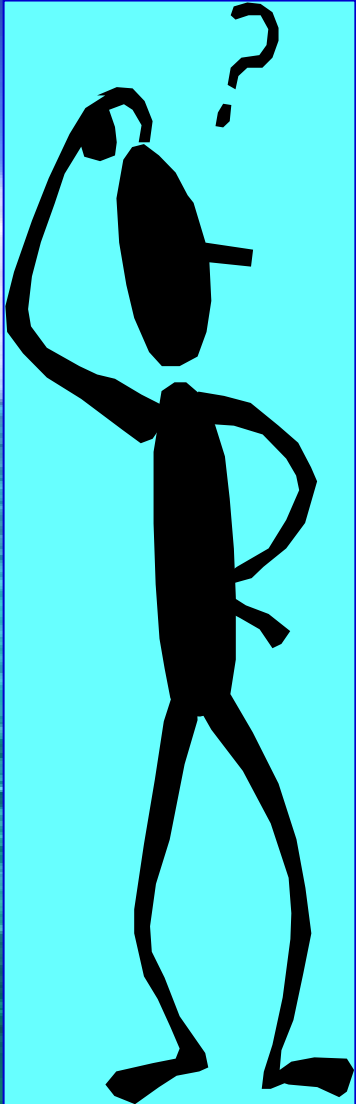
再構築モデル(120万社)



課 題

- 業種と地域の取り扱い
- 情報ベンダーの予測倒産モデルとの差異
 - 評点と企業照会回数の情報
- 再構築モデルの利用分野
 - マーケティング
 - 信用情報との融合

Question & Answer



Demonstrate How the Model works

