金融機関の信用リスク管理における デフォルト率推定のための共変量の自動設計

Automatic Design of Covariates for Estimating Default Probability Utilized for Credit Risk Management of Financial Institutes

信用リスク管理は金融機関にとって最も重要な業務の一つである。企業あるいは個人が債務不履行になる確率をデフォルト率という。デフォルト率の推定に用いられるモデルの一つに二項ロジットモデルがある。このモデルは高精度にデフォルト率を推定できるが，その共変量の設計が困難であるという課題がある。一般に共変量は専門家の試行錯誤によって設計される。本論文では，この共変量を自動設計する手法を提案する。提案手法では，遺伝的アルゴリズムによって共変量を設計し，各共変量の係数を準ニュートン法の一つであるBFGS 法によって求める。数値実験によって，本手法により優れた共変量を設計できることを確認した。

手塚　大　Tezuka Masaru
伊藤　洋一　Ito Yoichi
宗形　聡　Munakata Satoshi

1. はじめに

金融機関の倒産は，国内外に対して大きな社会的影響を与える。銀行が事業を継続しそれにより経済の安定に寄与するために，正確で適切な信用管理が必要とされている。

国際決済銀行による自己資本比率規制（BIS 規制）[1]では，リスクに対して十分な資本を保有するよう銀行に要求している。さらに新 BIS 規制では，各行ごとの状況にあわせた信用リスク計測の導入によるリスク評価の精緻化を求めている。

新 BIS 規制では信用リスクの計測方法を，その精緻度によって「標準的手法」「基礎的内部格付法」「先進的内部格付法」に分類している[2]。

標準的手法は格付機関による格付や，与信先の分類（大企業，中堅企業，中小企業，個人，住宅ローン）によってリスク量を定める方法である。しかし本手法より精緻な内部格付法に移行することが強く望まれている。

取引先がデフォルト状態，すなわち債務を履行できない状態になる確率をデフォルト率という。基礎的内部格付法では銀行が独自に取引先を格付けし，格付に基づいて取引先のデフォルト率を推定しこれをリスク評価に用いる。

先進的内部格付法は，デフォルト率に加えてデフォルト時損失額（担保による保全や，清算による回収の金額や回収できる確率などを考慮）を推定しリスク評価を行う方法である。

銀行等の金融機関では，金融庁の金融検査マニュアル[3]にもとづき自己査定を行い，この中で取引先を六つの債務者区分「正常先」，「要注意先」，「要管理先」，「破綻懸念先」，「実質破綻先」，「破綻先」に分類している。なお，実務上は要管理先を除いた 5 区分としている場合が多い。この場合，要管理先は，要注意先を細分化した中の一区分として位置づけられる。

具体的にどのような状態をデフォルトとするかは金融機関や格付機関により異なるが，自己査定で破綻懸念先以下の区分をデフォルトとして扱う金融機関が多い。しかし，最近は銀行が企業の経営に介入し支援することで改善させるリレーションシップ・バンキングの強化により，破綻懸念先の債務不履行が減ってきているため，実質破綻先以下をデフォルトとして扱う金融機関も増えてきている。

前述のように新 BIS 規制での基礎的および先進的内部格付法では，内部格付に基づいてデフォルト率を推定し，リスク量を評価する必要がある。デフォルト率の推定には内部格付の遷移マトリクスを用いる手法が広く使われている[4]。しかし，遷移マトリクスでは取引先の個別の状況を考慮することが難しいなどの問題がある。

取引先ごとの個別の財務状況にもとづく定量的なデフォルト率推定方法に二項ロジットモデルがある。こ

のモデルは，取引先の個別状況をモデルで考慮できるため，内部格付の遷移マトリクスよりも高い精度でデフォルト率が推定できる。しかし，モデルで用いる共変量の設計が困難であるという課題がある。本報告では，共変量を自動設計する手法を提案し，精度の高い二項ロジットモデルによるデフォルト率の定量的推定を実現する。

2．デフォルト率の推定

2.1　項目反応理論と二項ロジットモデル

二項ロジットモデルは，項目反応理論に基づいて構築されたモデルである。項目反応理論では，興味の対象の状態を表す共変量の和が，ある閾値より大きく（あるいは小さく）なると，その対象が反応を引き起こすと考える。デフォルト率の場合には，ある企業の経営状況を表す財務数値の（重み付きの）和がある閾値以下になると，経営状況の悪化への反応として債務不履行状態になると考える。状態を表す全ての共変量が観測可能とは限らないので観測できない共変量は誤差として扱う。

取引先の数を n，各取引先の状態を表す観測可能な共変量の数を m とする。取引先 i の j 番目の共変量を x_{ij} $(i=1,...,n,\ j=1,...,m)$ とする。また，j 番目の共変量の重みを b_j とする。項目反応理論から，取引先 i は

$$\sum_{j=1}^{m} b_j x_{ij} + \varepsilon < \theta$$

となった時にデフォルトとなる。ここで，θ は閾値，ε は観測できない共変量を表す誤差である。

簡単のために，閾値 θ を $-b_0$ とおき，x_{i0} を導入する。x_{i0} は全ての i で 1 とする。取引先 i のデフォルト率は，

$$\begin{aligned} d_i &= P\left(\sum_{j=1}^{m} b_j x_{ij} + \varepsilon < \theta\right) \\ &= P\left(\varepsilon < \theta - \sum_{j=1}^{m} b_j x_{ij}\right) \\ &= P\left(\varepsilon < -\sum_{j=0}^{m} b_j x_{ij}\right) \\ &= P\left(\varepsilon < -\mathbf{b}^T \mathbf{x}_i\right) \end{aligned}$$

となる。ここで，$\mathbf{b} = (b_0, b_1, ..., b_m)^T$ および $\mathbf{x}_i = (x_{i0}, x_{i1}, ..., x_{im})^T$.

多くの分野で誤差に正規分布を用いるのが一般的であるが，正規分布は閉じた形で表せないため扱いにくい。ロジスティック分布は正規分布に似た形状で，閉じた形で表現できて解析的に扱いやすいという特徴を持つ。そこで，誤差 ε がロジスティック分布 5) にしたがうとする。

ロジスティック分布の分布関数は，パラメータ μ，s を用いて

$$F(z; \mu, s) = \frac{1}{1 + e^{-(z-\mu)/s}}.$$

となる。ε が，$\mu = 0$，$s = 1$ のロジスティック分布に従うとすると，取引先 i のデフォルト率は

$$\begin{aligned} d_i &= F(-\mathbf{b}^T \mathbf{x}_i; 0, 1) \\ &= \frac{1}{1 + e^{\mathbf{b}^T \mathbf{x}_i}} \end{aligned} \tag{1}$$

となる。式 (1) を二項ロジットモデルと呼ぶ。

2.2　共変量の設計

一般に財務諸表に記載されている項目をもとに取引先の状況を表す共変量を設計する。財務諸表とは，貸借対照表，損益計算書，キャッシュフロー計算書などを指す。その他，企業に関して公開されている情報も利用できる。

これら諸数値を共変量として直接用いることはほとんどない。例えば次式のように，何らかの計算式で変換した値を共変量として用いる。

$$x_{i1} = \frac{(\text{流動資産})_i - (\text{流動負債})_i}{(\text{総資産})_i} \tag{2}$$

なお，表記「$(\text{科目})_i$」は取引先 i の当該勘定科目の数値を表す。この例では，総資産に対する比とすることで，企業の大きさによらずに財務状況を表す共変量としている。

共変量を計算するための数式は，専門家が経験にもとづいて，あるいは試行錯誤によって設計している。この共変量の設計が，二項ロジットモデルの導入が困難となる原因である。

3．デフォルト率の推定方法の提案

提案手法の全体像を図 1 に示す。債務者区分（デフォルト/非デフォルト），財務諸表データ，およびその他の利用可能なデータを入力とし，モデルの共変量を遺伝的アルゴリズム (GA)6) を用いて自動設計する。また，モデルの係数を準ニュートン法を用いて計算する。これらの共変量と係数を用いて，二項ロジットモデルによってデフォルト率を推定する。

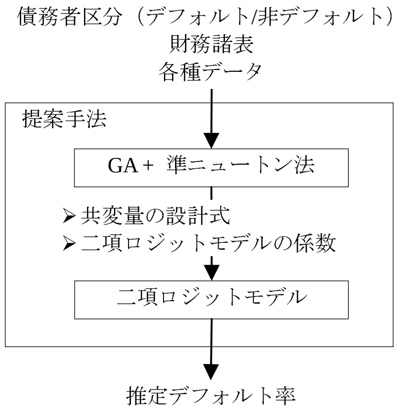

図１　デフォルト率の推定

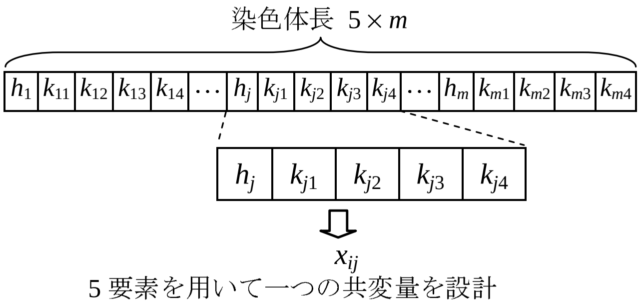

図２　染色体表現

3.1　遺伝的アルゴリズムによる共変量の自動設計

財務諸表などの利用可能なデータに t の項目があるとする。取引先 i の k 番目の財務データ項目の値を y_{ik} $(i = 1, ..., n, k = 1, ..., t)$ とする。

後の計算で 0 による割り算の発生を防ぐため，$y_{i1}, y_{i2}, ..., y_{iu}$ $(u \leq t)$ は 0 になることがなく，$y_{iu+1}, ..., y_{it}$ は 0 になりうるように各項目に添え字を付ける。

本報告で提案する染色体表現を図 2 に示す。モデルで使用する共変量の数を m とすると，染色体の長さを $5 \times m$ とする。これは，5 個の遺伝子座 $h_j, k_{j1}, k_{j2}, k_{j3}, k_{j4}$ が一組となって，j 番目の共変量を設計するためである。

各遺伝子座の値の範囲は，$j = 1, ..., m$ について

$$\begin{aligned} h_j &\in \{1, 2, ..., 6\}, \\ k_{j1}, k_{j2} &\in \{1, 2, ..., t\}, \\ k_{j3}, k_{j4} &\in \{1, 2, ..., u\} \end{aligned} \tag{3}$$

とする。

$$x_{ij} = \begin{cases} \dfrac{y_{i(k_{j1})} + y_{i(k_{j2})}}{y_{i(k_{j3})} + y_{i(k_{j4})}} & \text{if } h_j = 1, \\[2ex] \dfrac{y_{i(k_{j1})} - y_{i(k_{j2})}}{y_{i(k_{j3})} + y_{i(k_{j4})}} & \text{if } h_j = 2, \\[2ex] \dfrac{y_{i(k_{j1})} + y_{i(k_{j2})}}{y_{i(k_{j3})}} & \text{if } h_j = 3, \\[2ex] \dfrac{y_{i(k_{j1})} - y_{i(k_{j2})}}{y_{i(k_{j3})}} & \text{if } h_j = 4, \\[2ex] \dfrac{y_{i(k_{j1})}}{y_{i(k_{j3})} + y_{i(k_{j4})}} & \text{if } h_j = 5, \\[2ex] \dfrac{y_{i(k_{j1})}}{y_{i(k_{j3})}} & \text{if } h_j = 6. \end{cases} \tag{4}$$

染色体をもとに，式 (4) のように共変量が設計される。分母に来る k_{j3} と k_{j4} の値の範囲は $1, ..., u$ であり，前述のように $y_{i1}, ..., y_{iu}$ は 0 とならないようにしてあるので，分母が 0 になることはなく，従って 0 による割り算を避けることができる。

第一世代では全ての個体の染色体は式 (3) の範囲内でランダムに初期化される。次の世代の生成には，一様交叉 [7] を用いる。突然変異では，各要素の値を値の範囲内でランダムに変化させる。

3.2　適応度関数

取引先 i がデフォルトになったときに 1，それ以外で 0 を取る変数 δ_i を用いると，二項ロジットモデルの対数尤度関数は

$$L(\mathbf{b}; \mathbf{x}_1, ..., \mathbf{x}_n) = \sum_{i=1}^{m} (\delta_i \log d_i + (1 - \delta_i) \log(1 - d_i)) \tag{5}$$

となる。二項ロジットモデルの係数 $\mathbf{b}$ は尤度が最大になるように設定される。したがって，染色体によって表される共変量の適応度は，次式となる。

$$f(\mathbf{x}_1, ..., \mathbf{x}_n) = \arg\min_{\mathbf{b}} L(\mathbf{b}; \mathbf{x}_1, ..., \mathbf{x}_n). \tag{6}$$

この対数尤度関数のヘッセ行列は半負定値なので関数は凸である。したがって大域収束性は保証されているが，非線形関数のため解析的に解くことはできない。したがって，準ニュートン法などの反復解法によって解く。提案手法ではブロイデン-フレッチャー-ゴールドファルブ-シャノ (BFGS) 法 [8] を用いる。

表１　パラメータ設定

集団数	500	エリート数	10
世代数	500	トーナメント数	2
交叉率	0.6	ϵ (BFGS 終了条件)	1.0×10^{-6}
突然変異率	0.2		

4．数値実験

240 企業の財務データ 60 項目からなるデータ 3 年分を用いて，提案手法の評価実験を行った。データ項目の数 t は 180 (= 60 項目 ×3 年) となる。

ただし今回は債務者区分情報を入手できず，代わりに別途推定されたデフォルト値の情報を使用した。そこで，入手できたデータにより提案手法を評価するために，実験の中では適応度関数を以下のように修正して用いた。

提案手法で推定されたデフォルト率と，上記別途推定されたデフォルト値の平均二乗誤差は次式で表される。

$$g\left(\mathbf{b};\mathbf{x}_1,...,\mathbf{x}_n\right)=\frac{1}{n}\sum_{i=1}^{n}\left(d_i-\hat{d}_i\right)^2 \qquad (7)$$

ここで，$d_i, \hat{d}_i$ はそれぞれ，提案手法および信頼できる方法で推定された既取引先 i のデフォルト率である。

二項ロジットモデルの係数 $\mathbf{b}$ は式 (7) が最小になるように設定される。したがって，染色体によって表される共変量の適応度は，次式となる。

$$f\left(\mathbf{x}_1,...,\mathbf{x}_n\right)=\arg\min_{\mathbf{b}} g\left(\mathbf{b};\mathbf{x}_1,...,\mathbf{x}_n\right). \qquad (8)$$

実験では BFGS 法 [8] を用いて適応度を計算した。ヘッセ行列の逆行列の計算を効率化するためにシャーマン-モリソン公式を適用した。ステップ長はアルミホ条件 [9] を満たすように設定した。図 3 に係数 $\mathbf{b}$ を求める手順を示す。

図 3 のステップ 2 で用いる，g の b_j に関する偏微分は次式になる。

$$\frac{\partial g}{\partial b_j}=\frac{2}{n}\sum_{i=1}^{n}\frac{e^{\mathbf{b}^T\mathbf{x}_i}\left(d_i-\dfrac{1}{1+e^{\mathbf{b}^T\mathbf{x}_i}}\right)x_{ij}}{\left(1+e^{\mathbf{b}^T\mathbf{x}_i}\right)^2}.$$

クロス・バリデーション法 [10] によって提案手法の評価を行った。フォールド数は 20 を用いた。

ステップ１　探索の出発点 $\mathbf{b}_0$ を $\mathbf{0}$ に設定。ヘッセ行列の逆行列 H_0 を単位行列として初期化。

ステップ２　探索方向を次式で計算。

$$\mathbf{d}_\tau=-H_\tau\nabla g(\mathbf{b}_\tau).$$

ここで，

$$\nabla g(\mathbf{b})=\left(\frac{\partial g}{\partial b_0},...,\frac{\partial g}{\partial b_m}\right)^T$$

ステップ３　探索ステップ長 α_τ を，アルミホ条件を満たすように決定。

ステップ４　探索点を次式で更新。

$$\mathbf{b}_{\tau+1}=\mathbf{b}_\tau+\alpha_\tau\mathbf{d}_\tau,$$

ここで，$\nabla g(\mathbf{b}_{\tau+1})=\mathbf{0}$ または $|\mathbf{b}_{\tau+1}-\mathbf{b}_\tau|<\epsilon$ であれば, $\mathbf{b}_{\tau+1}$ が解 $\hat{\mathbf{b}}$ となる。ϵ は終了条件を制御するパラメータ。

ステップ５　ヘッセ行列の逆行列を次式で更新。

$$H_{\tau+1}=H_\tau-\frac{H_\tau\mathbf{y}_\tau\mathbf{s}_\tau^T+\mathbf{s}_\tau\left(H_\tau\mathbf{y}_\tau\right)^T}{\mathbf{s}_\tau^T\mathbf{y}_\tau}+\left(1+\frac{\mathbf{y}_\tau^T H_\tau\mathbf{y}_\tau}{\mathbf{s}_\tau^T\mathbf{y}_\tau}\right)\frac{\mathbf{s}_\tau\mathbf{s}_\tau^T}{\mathbf{s}_\tau^T\mathbf{y}_\tau}$$

ここで，

$$\mathbf{s}_\tau=\mathbf{b}_{\tau+1}-\mathbf{b}_\tau$$
$$\mathbf{y}_\tau=\nabla g(\mathbf{b}_{\tau+1})-\nabla g(\mathbf{b}_\tau)$$

ステップ６　ステップ 2 から探索を繰り返す。

図３　BFGS 法による $\mathbf{b}$ の推定

提案手法は Microsoft Visual C++ 2005 で実装され，Intel Pentium IV 3.6GHz および 1Gbytes の RAM を搭載した PC を用いた。OS は Windows XP professional SP2 を使用した。数値実験で用いたパラメータ設定を表 1 に示す。共変量の自動設計を 1 回実行するのに約 3 時間 26 分（平均 206.3 分，分散は 2.748）を要した。なお，計算時間は学習に用いる企業の数に依存する。

比較のために文献 [11] で紹介されている共変量によるデフォルト率の推定も行った。

表 2 に実験結果を示す。MSE は平均二乗誤差を示し，小さいほど推定精度が高い。VAR は誤差の分散を示し，推定の安定性を示す。小さいほど安定した推定となる。COR は相関係数を示し，1.0 に近いほど推定精度が高い。提案手法で自動設計した共変量を用いると，高精度，高安定に推定できた。

表２　テストデータでのデフォルト率の推定精度

	提案手法で自動設計	文献記載の共変量
MSE	2.00×10^{-2}	2.41×10^{-2}
VAR	3.57×10^{-4}	7.23×10^{-4}
COR	0.41	0.26

5．おわりに

本報告では，二項ロジットモデルによるデフォルト率推定に用いる共変量を，自動設計する手法を提案した。共変量設計のための問題エンコード（染色体表現）法を提案し，これを GA によって最適化する。二項ロジットモデルの係数は準ニュートン法の一つ, BFGS 法によって求める。数値実験によって，提案手法によって自動設計した共変量を用いると，高精度にデフォルト率を推定できることを示した。

本技術を，金融分野向けの新しいソリューションを支える技術として確立し，事業拡大につなげていく予定である。

参考文献

1) Basle Committee on Banking Supervision, “Amendment to the Capital Accord to Incorporate Market Risks”, Bank for International Settlements, 1996.
2) 日本銀行, “新 BIS 規制案の概要,” 日本銀行 2004 年 10 月 12 日公表資料, 2004
3) 金融庁, 預金等受入金融機関に係る検査マニュアル, 金検第 79 号, 2007
4) 日本銀行金融機構局, “内部格付制度に基づく信用リスク管理の高度化”, リスク管理高度化と金融機関経営に関するペーパーシリーズ, 2005
5) M. Evans N. Hastings and B. Peacock, *Statistical Distributions*, Wiley Interscience, 2000.
6) T. Bäck, D. B. Fogel and Z. Michalewicz, *Handbook of Evolutionary Computation*, Institute of Physics Publishing and Oxford University Press, 1997.
7) C. R. Reeves, *Modern Heuristic Techniques for Combinatorial Problems*, Halsted Press, 1993.
8) P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, Elsevier, 1983.
9) J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Verlag, 1999.
10) R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” *Proc. Int. Joint Conf. on Artificial Intelligence IJCAI 1995*, 1995.
11) 木島 正明, 小守林 克哉, 信用リスク評価の数理モデル, 朝倉書店, 1999.

（執筆者紹介）

手塚　大　　1994 年入社
研究開発部
意思決定，リスク分析，最適化技術の研究，開発
tezuka@hitachi-to.co.jp

伊藤　洋一　　1989 年入社
金融第 2 部
融資支援システムの開発，拡販
you-ito@hitachi-to.co.jp

宗形　聡　　2003 年入社
研究開発部
統計，数理的アプローチによる業務診断，意思決定技術の研究，開発
munakata@hitachi-to.co.jp