統計数理(2018)
第 66 巻 第 2 号 295–317

［原著論文］

B-スプライン及びAdaptive Group LASSOに基づく正則化非線形ロジットモデルによるデフォルト確率の推定

高部 勲[1,2]・山下 智志[3]

（受付 2017 年 12 月 31 日；改訂 2018 年 6 月 9 日；採択 6 月 20 日）

要　　旨

企業の過去のデフォルトデータを基にデフォルト確率予測モデルを構築する際には線形な 2 項ロジットモデルが用いられることが多いが，これについては従前から，(1)企業の信用スコアと財務指標との間の非線形性に対する考慮が不十分であり，また(2)多くの説明変数の候補からの変数選択に莫大な計算時間がかかるというという 2 つの課題についての指摘がある．本稿では，このような非線形性と変数選択という 2 つの課題を同時に解決することを目的として，(1)B-スプラインに基づく非線形・ノンパラメトリック回帰モデル及び(2)Adaptive Group LASSO に基づく効率的な変数選択という 2 つの手法を組み合わせることにより，従前の手法よりも効果的かつ効率的なデフォルト確率予測モデルの構築を試みた．複数の銀行のデータを統合した独自のデータベースを用いてデフォルト確率予測モデルの構築を行った結果，本稿で提案したモデルは，t 値・p 値に基づく変数選択や単純な LASSO と比較して，いずれの期間においても最も説明変数の数が少なくなっており，より効率的な変数選択を行うことができた．また AR 値などの指標の観点から，推定精度が向上していることが確認された．

キーワード：信用リスク， B-スプライン， Adaptive Group LASSO.

1. 導入

金融機関の信用リスク管理を考える際に，個別企業のデフォルト確率や倒産確率の予測精度の向上は重要な課題となっている．企業のデフォルト確率の予測モデルには，企業価値や債券価格を確率過程で記述するモデル(Merton, 1974; Duffie and Singleton, 1999)や，多変量判別分析に基づくモデル(Altman, 1968; 白田, 2008)などがあるが，金融機関における実務では企業の過去のデフォルトに関するデータを基にデフォルト確率を予測するモデルを構築することが多く，その際には線形な 2 項ロジットモデルがよく用いられている(尾木, 2017; 山下・三浦, 2011; 森平, 2009; Martine, 1977; Engelmann and Raumeier, 2006)．しかし線形な 2 項ロジットモデルについては従前から，以下の 2 つの課題があることが指摘されている．

(1)企業の信用スコアと各種財務指標との間の非線形性に対する考慮が不十分

[1] 総合研究大学院大学 複合科学研究科統計科学専攻：〒 190–8562 東京都立川市緑町 10–3
[2] 総務省統計局：〒 162–8668 東京都新宿区若松町 19–1
[3] 統計数理研究所：〒 190–8562 東京都立川市緑町 10–3

(2)多くの説明変数(各種財務指標)の候補からの変数選択に莫大な計算時間がかかる

(1)の非線形性に関する課題について，従来の研究ではロジットモデルの説明変数として 2 次以上の多項式などを用いることにより対処している場合が多い．しかしそのようなモデルでは，非線形かつ多様な変動を把握するには限界があると考えられる．また，(2)の変数選択の課題については，t 値・p 値や AIC を基にしたステップワイズな変数選択により対処している事例が多いが，説明変数として用いる財務指標の数が多くなると，比較対象となるモデルの数も指数的に増大し，計算時間の面で限界があることから，より効率的な変数選択の手法が必要とされている．

上記の 2 つの課題に対して個別に対処している先行研究は存在するものの(これらの先行研究の具体的な内容については 2 節で示す)，これらの課題を同時に考慮したデフォルト確率予測モデルの事例については，調べた限りでは存在しない．そこで本稿では，これらの課題を同時に解決することを目的として，以下の 2 つの手法を組み合わせることにより，従前の結果と比較して，より精度の高いデフォルト確率予測モデルの効率的な構築を試みた．

(1)B-スプラインに基づく非線形・ノンパラメトリック回帰モデル
(2)Adaptive Group LASSO に基づく効率的な変数選択

本研究では複数の銀行のデータを統合した独自のデータベースを用いて，中小企業を対象としたデフォルト確率予測モデルの構築を行う．本研究において提案した手法はデフォルト確率との非線形な関係の合理的・効率的な構築に寄与するものであり，各種財務指標に基づく与信判断などにも資すると考えられる．

2. 先行研究と課題

2.1 ロジットモデルに基づくデフォルト確率予測モデル

企業の過去のデフォルトに関するデータを基にデフォルト確率予測モデルを構築する際に，個別企業に関する大規模なデータが活用できる場合には，線形な 2 項ロジットモデルが利用されることが多い．このような研究としては，中小企業信用リスク情報データベース協会(CRD 協会)のデータを用いた高橋・山下 (2002)や，日本政策金融公庫のデータを用いた尾木他 (2015)などがある．

線形な 2 項ロジットモデルは，企業 i $(1 \leq i \leq n)$ のデフォルト確率を P_i，対応する財務指標を x_{ij} $(1 \leq j \leq p)$ とした場合，以下のように表現される．

$$\log \frac{P_i}{1-P_i} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \tag{2.1}$$

式(2.1)の左辺は P_i のロジット変換である．式(2.1)は以下のように表現することもできる．

$$\begin{aligned} P_i &= \frac{1}{1+\exp(-Z_i)} \\ Z_i &= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \end{aligned} \tag{2.2}$$

ここで Z_i は企業 i の信用スコアを表しており，一般的にこの数値が大きくなるほど企業の信用力が低くデフォルト確率 P_i が高くなる．信用スコア Z_i を基に企業のデフォルトの可能性を予測することができる．

上記の 2 項ロジットモデルにおける回帰係数 $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ は最尤法により推定する．

具体的には以下の対数尤度 $L_1(\beta)$ を最大化するような $\hat{\beta}$ を回帰係数の推定値とする．

$$
\begin{aligned}
L_1(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \log[{P_i}^{\delta_i}(1-P_i)^{1-\delta_i}] \\
&= \sum_{i=1}^{n} [\delta_i \log(P_i) + (1-\delta_i)\log(1-P_i)] \qquad (2.3) \\
\delta_i &= \begin{cases} 1 & (\text{企業 } i \text{ がデフォルトしている場合}) \\ 0 & (\text{企業 } i \text{ が非デフォルトである場合}) \end{cases}
\end{aligned}
$$

P_i には，式(2.2)で表されるデフォルト確率を代入する．このような単純な 2 項ロジットモデルに基づくデフォルト確率予測については，回帰係数の推定が容易である一方，次節以降に示すような課題があることが指摘されている．

2.2 財務指標と信用スコアとの非線形な関係

式(2.1)による 2 項ロジットモデルでは線形なモデルを仮定している．しかし財務指標によっては信用スコアとの間に非線形な関係があることが指摘されている（Dwyer et al., 2004; 白田, 2008）．このような場合に線形なモデルを用いると，信用スコアと財務指標との関係を適切にモデリングすることができず，デフォルト確率の予測精度が低下するおそれがある．図 1 は，今回使用するデータ(詳細は 4.1 節を参照)のうち，2005 年から 2013 年までの期間に関して，いくつかの財務指標と実績デフォルト率のロジット変換値との関係を示したものである．具体的には各財務指標の大きさの順に企業を並べ，それらを財務指標の大きさに応じて 200 のクラスに分割し，各クラスにおける実績デフォルト率のロジット変換値をプロットしたものである．その際に歪みの大きい一部の財務指標に対して対数変換又は neglog 変換を行っており，さらにそれらの値が 0 から 1 の範囲に収まるように線形変換を行っている．なお neglog 変換は以下のように定義されるもので，対数変換を負の値に拡張した変換となっている(森平, 2009; 山下・三浦, 2011)．

$$
\mathrm{neglog}(x) = \begin{cases} \log(x+1) & (\ x \geq 0) \\ -\log(-x+1) & (\ x < 0) \end{cases} \qquad (2.4)
$$

また図 1 には併せて線形ロジットモデルによる予測値(点線)と，後述の式(3.1)の B-スプラインに基づく非線形ロジットモデルについて，単変量のモデルを各財務指標に当てはめて推定した予測値(実線)を示している．その際の B-スプラインの基底の数は，AIC に基づき選択した．図 1 から，財務指標によっては実績デフォルト率のロジット変換値との間に明らかに非線形な関係があり，線形なモデルではこれらの変動に対応できていないことがわかる．

式(2.1)で示した 2 項ロジットモデルにおいて非線形な効果を扱う場合には，各財務指標の多項式や対数，平方根などの項を導入することが考えられる(Hosmer et al., 2013)．しかしそれらの関数形のどれが正しいかを事前に知ることはできないため，このような方法では各種財務指標の非線形な影響の把握には限界があると考えられる．

財務指標との関係を多項式モデルのような形であらかじめ設定するのではなく，ノンパラメトリック回帰モデルの手法を用いてデータから柔軟に曲線関係を推定している先行研究も存在する．Berg (2007)では，一般化加法モデル(Generalized Additive Model, Hastie and Tibshirani, 1990)の枠組みを企業の倒産確率モデルに導入することにより，データから柔軟な形で財務指標と倒産確率との非線形な関係を推定している．そして従来の判別分析モデルや線形な 2 項ロ

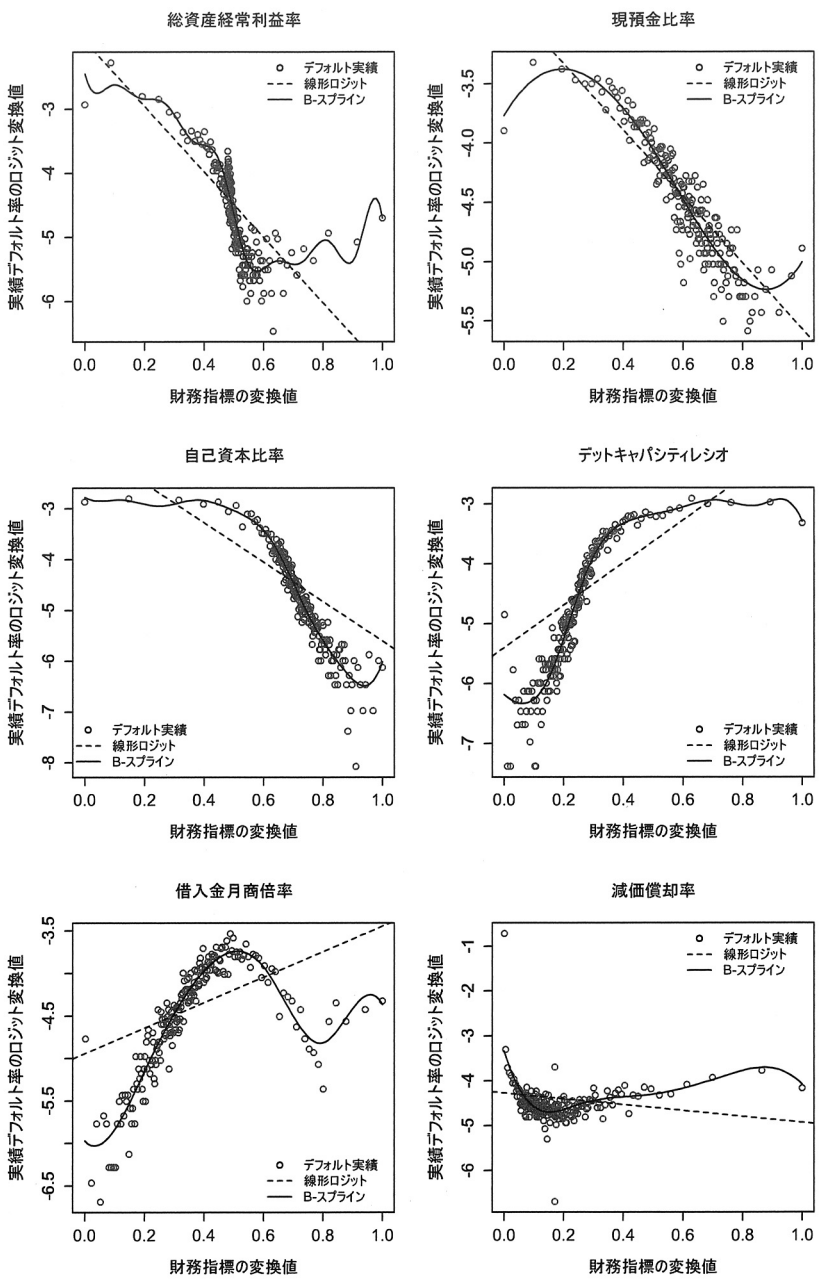

図 1. 実績デフォルト率の状況(図中の「現預金比率」,「デットキャパシティレシオ」及び「借入金月商倍率」については neglog 変換を行っている．また全ての財務指標について，0 から 1 の範囲に収まるように線形変換を行っている.)．

ジットモデルと比較して，AR 値の観点から推定精度が向上したと報告している．Giordani et al.（2014)では 2 次の自然スプラインを用いた非線形ロジットモデルを用いて個別企業の倒産

確率を分析しており，線形ロジットモデルと比較して，AR 値や疑似決定係数の観点からモデルの精度が向上し，倒産確率と各種財務指標との非線形な関係を適切に捉えることができたと報告している．山内 (2010)では財務指標を離散化したスコアリングテーブルに基づき，遺伝的アルゴリズムにより多目的最適化問題を解くことによって非線形なモデルを推定している．ただしこれらの先行研究では，いずれも非線形なモデルの構築のみに焦点を当てており，各種財務指標の中から適切なものを選択するという変数選択の観点は考慮されておらず，具体的な計算に入る前の段階でモデルに導入する財務指標の種類を，事前にある程度限定している．

2.3 複数の財務指標に関する変数選択

デフォルト確率予測モデルを構築する場合，説明変数として用いられる財務指標の候補の数は主要なものだけでも数十程度あり，場合によっては 100 を超えることもある．これらの財務指標の全ての組合せに基づくモデルを推定して比較を行う場合，対象となるモデルの数が非常に多くなるため，モデル構築にかなりの時間を要する．例えば候補となる財務指標の数が 50 である場合，$2^{50}\,(\fallingdotseq 10^{15})$ 通りのモデルの候補が考えられる．これらの候補の中から AIC 等の基準に基づくステップワイズな方式によりモデル選択を行った場合，現実的な計算時間で推定を行うことは困難であることから，より効率的な変数選択の手法が必要となる．しかし従来のデフォルト確率予測モデルの構築においては t 値・p 値を用いた単純な変数の絞込みや，何らかの先験的な知見に基づく事前の財務指標の選択が行われているのが実情である．

これに関して近年，回帰係数の推定と変数選択を同時に実行できる LASSO(Least Absolute Shrinkage and Selection Operator)に関する研究が発展しており（Tibshirani, 1996; Hastie et al., 2015; 冨岡, 2015)，この方法を適用した企業のデフォルト確率や倒産確率の推定に関する研究も行われるようになってきている（Amendola et al., 2012; Perederiy, 2009; Tian et al., 2015)．LASSO に基づくロジットモデルでは，式(2.3)の対数尤度 $L_1(\boldsymbol{\beta})$ に，$L1$ ノルムに基づく正則化項を加えた以下の罰則付きの対数尤度 $L_2(\boldsymbol{\beta})$ の最大化を行うことにより，回帰係数 $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ の推定を行う．

$$L_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\delta_i \log(P_i) + (1 - \delta_i) \log(1 - P_i)] - \lambda \sum_{j=1}^{p} |\beta_j| \tag{2.5}$$

式(2.5)の最大化は，回帰係数 $\boldsymbol{\beta}$ の範囲に $\sum_{j=1}^{p} |\beta_j| \leq t$ という制約を加えた下での式(2.3)の最大化と同値である(λ と t は 1 対 1 に対応)．なお定数項 β_0 にはこのような制約を課さないのが一般的である(Hastie et al., 2015)．$L1$ ノルムに基づく正則化項の下では，値の小さい回帰係数が 0 になりやすくなる傾向があり，この性質が回帰係数の推定と説明変数の選択を同時に行うことを可能としている．ここで λ は正則化項の効果を調整するチューニングパラメータであり，交差検証法により決定することが多い．

Prederiy (2009)は企業の倒産予測における変数選択の問題について，LASSO に基づく 2 項ロジットモデルを用いて対処した先駆的な研究であり，効率的な変数選択により計算量の削減を達成するとともに，モデルの予測精度も向上したと報告している．ただし単純な線形ロジットモデルに LASSO を適用するにとどまっており，財務指標との非線形な関係を考慮しておらず，最終的に選択された財務指標の数も多くなっている．また Amendola et al. (2012)や Tian et al. (2015)では，Cox 比例ハザードモデルと LASSO を組み合わせて企業の倒産確率の長期予測を行っているが，これらの研究においても同様に単純な線形ロジットモデルが用いられており，財務指標との間の非線形な関係を考慮したモデルとはなっていない．

3. 非線形・正則化ロジットモデルに基づくデフォルト確率予測モデルの構築

3.1 本研究の目的

これまでに述べたように，信用スコアと財務指標との間の非線形性及び変数選択の問題については，双方ともモデルの構築に当たり重要な課題であるが，それぞれの課題に個別に対応する研究事例はあるものの，これらを同時に考慮したモデルに関する研究については，調べた限りでは存在しない．本研究ではこれらの課題に対し，(1)B-スプラインに基づく非線形・ノンパラメトリック回帰モデルの導入，及び(2)Adaptive Group LASSO に基づく合理的な変数選択の適用という 2 つの手法を組み合わせたデフォルト確率予測モデルを提案する．

3.2 B-スプラインに基づく非線形モデル

まず財務指標との間の非線形性を考慮したモデリングについて検討する．これについてはスプラインに基づく非線形な項を導入することにより対応する．スプラインは，説明変数に関するデータが含まれる区間をいくつかの小区間に分割し，各区間において区分的な多項式モデルを当てはめる方法である(小西, 2010; 山下・安道, 2006; 桜井, 1981)．説明変数とデフォルト確率との複雑な関係を単一の多項式モデルで把握するのではなく，隣り合う各区間における多項式モデルを滑らかに接続することにより，非線形な構造に対処する方法となっている．

本稿では B-スプラインに基づく方法を検討する．B-スプラインは局所的な台を持つスプライン関数であり，複数の多項式を滑らかに接続して基底関数を構成する．B-スプラインの導入により，特定の関数形を仮定せずに，財務指標とデフォルト確率との間の非線形な関係をデータから柔軟に推定することが可能となる．B-スプラインに基づく非線形ロジットモデルは，式(2.2)における信用スコア Z_i を以下の式(3.1)で置き換えることで得られる．

$$Z_i = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}), \quad f_j(x_{ij}) = \sum_{k=1}^{m_j} \beta_{jk}\phi_k(x_{ij}) \tag{3.1}$$

ここで f_j $(1 \le j \le p)$ は各財務指標に対応する非線形関数であり，ϕ_k $(1 \le k \le m_j)$ は B-スプラインの基底を表している．図 2 は B-スプラインに基づく非線形回帰のイメージを示したものである．左側の図が各説明変数に対する B-スプラインの基底を表しており(基底の数は 9 に設定)，右側の図はこれらの基底に基づく非線形回帰モデルの予測値を示している．このように非線形な基底を組み合わせることで，データから柔軟に関数を推定することが可能となる．

本稿では先行研究(Huang et al., 2010)に基づき B-スプラインの次数は 3 次とし，基底の計

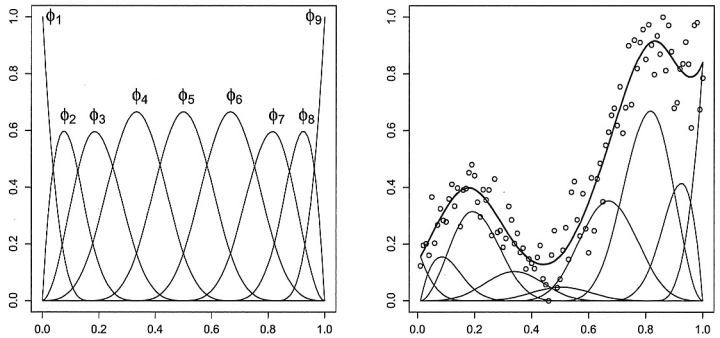

図 2. B-スプラインに基づく非線形回帰のイメージ　左図：B-スプラインの基底　右図：B-スプラインに基づく非線形モデル(点がサンプルデータ，太線が予測値)．

算には R の bs 関数を用いている．B-スプラインを構築するに当たり，区間を分割する節点を設定する必要がある．節点の位置については等間隔に設定している．また節点の数については，これを 5 から 15 の範囲で変化させて各財務指標に対して単変数の非線形ロジットモデルを当てはめ，AIC に基づき財務指標ごとにその数を事前に決定している．

3.3 Group LASSO に基づく変数選択

B-スプラインに基づく非線形モデルでは，財務指標ごとに基底を複数個用意して滑らかな非線形の曲線を表現する．このとき 1 つの財務指標に対して複数の基底が対応することになるため，変数選択の際にはこれらの複数の基底をまとめてモデルに取り込む，あるいはモデルから除去する必要がある．このように複数の変数をグループとしてまとめて扱い，変数選択を行う方法として，Group LASSO がある（Meier et al., 2008; Hastie et al., 2015）．

Group LASSO では，式(2.5)における $L1$ ノルムによる正則化項の代わりに，$L2$ ノルム $\|\boldsymbol{\beta}_j\|_2$ $(= \sqrt{{\beta_{j1}}^2 + {\beta_{j2}}^2 + \cdots + {\beta_{jm_j}}^2})$ による正則化項を用いた以下の $L_3(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p)$ を最大化することにより，回帰係数 β_0 及び $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm_j})$ の推定値である $\hat{\beta}_0$ 及び $\hat{\boldsymbol{\beta}}_j$ $(1 \leq j \leq p)$ を得る手法である．これによりグループ単位での回帰係数の推定と変数選択を同時に行うことが可能となる．

$$L_3(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p) = \sum_{i=1}^{n} [\delta_i \log(P_i) + (1 - \delta_i) \log(1 - P_i)] - \lambda \sum_{j=1}^{p} \sqrt{m_j} \|\boldsymbol{\beta}_j\|_2 \tag{3.2}$$

P_i の式に含まれる信用スコア Z_i には，式(3.1)を代入する．なお，Yuan and Lin (2006)では式(3.2)のように，Group LASSO の重みにはグループのサイズの平方根を用いることが推奨されている．

ここで式(3.1)に関して，例えばある項 $f_j(x_{ij})$ に定数 C を加え，別の項 $f_k(x_{ik})$（あるいは定数項）から定数 C を引いても同一の信用スコア Z_i が得られることから，非線形関数の一意性が保証されないことになる．そこで非線形関数の一意性のために，Huang et al. (2010) に基づき，以下の制約を課す．

$$\sum_{i=1}^{n} \sum_{k=1}^{m_j} \beta_{jk} \phi_k(x_{ij}) = 0 \tag{3.3}$$

上記の制約については，ϕ_k を以下のように変換した新たな基底 ψ_{jk} を用いることで対応できる．

$$\bar{\phi}_{jk} = \frac{1}{n} \sum_{i=1}^{n} \phi_k(x_{ij}), \qquad \psi_{jk}(x_{ij}) = \phi_k(x_{ij}) - \bar{\phi}_{jk} \tag{3.4}$$

スプラインと Group LASSO を組み合わせたモデルを遺伝子分野の研究に応用した事例として，Huang et al.（2010），Meier et al.（2009）がある．本稿では Huang et al.（2010）の方法をベースとしつつ，次節に示すような調整を行った上で，デフォルト確率予測モデルを構築している．

3.4 Multistep Adaptive Group LASSO に基づく変数選択

LASSO や Group LASSO では正則化項にかかるチューニングパラメータ λ を変化させることで回帰係数にかかる制約の強さをコントロールすることができるが，全ての回帰係数に同一のパラメータ λ を適用している点は改良の余地がある．そこで回帰係数の大きさの逆数を罰則とすることで絶対値の小さな係数により大きな罰則を課し，効率的に変数を選択する方法が Adaptive Group LASSO である（Bühlmann and van de Geer, 2011; Huang et al., 2010）．

Adaptive Group LASSO では，既に得られている推定値 $\hat{\boldsymbol{\beta}}_j$ を用いて計算した ω_j を基に，以下の $L_4(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p)$ を最大化することにより回帰係数の推定を行う．

$$L_4(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p) = \sum_{i=1}^{n} [\delta_i \log(P_i) + (1-\delta_i)\log(1-P_i)] - \lambda \sum_{j=1}^{p} \sqrt{m_j}\omega_j \|\boldsymbol{\beta}_j\|_2 \tag{3.5}$$

$$\omega_j = \begin{cases} \|\hat{\boldsymbol{\beta}}_j\|_2^{-1} & (\|\hat{\boldsymbol{\beta}}_j\|_2 > 0) \\ \infty & (\|\hat{\boldsymbol{\beta}}_j\|_2 = 0) \end{cases}$$

ここで $\omega_j = \infty$ となる場合には，対応する変数をモデルから取り除くこととする．

本稿では変数の選択をより効率的に行うために，Adaptive Group LASSO を複数回適用する方法を用いる（以下ではこれを Multistep Adaptive Group LASSO と呼ぶ）．具体的には以下の手順により，係数を推定する．

(1) まず，Group LASSO を適用し，係数の推定値 $\hat{\beta}_0$ 及び $\hat{\boldsymbol{\beta}}_j$ $(1 \leq j \leq p)$ を得る．
(2) 得られた係数 $\hat{\boldsymbol{\beta}}_j$ を基に重み ω_j を計算し，Adaptive Group LASSO を適用して，係数の推定値 $\hat{\beta}_0^*$ 及び $\hat{\boldsymbol{\beta}}_j^*$ $(1 \leq j \leq p)$ を得る．
(3) 得られた係数 $\hat{\boldsymbol{\beta}}_j^*$ を基に，再度重みを計算し，Adaptive Group LASSO を適用して係数の最終的な推定値を求める．

今回の分析では計算のコストを考慮して，阪本 他 (2010) の設定を参考に，Multistep Adaptive Group LASSO における反復回数を 2 回に設定している．これらの計算の際には Adaptive Group LASSO の計算を比較的容易に行うことが可能であり，かつ高速な計算アルゴリズム (Groupwise Majorization Descent) を採用している R のパッケージ `gglasso` (Yang and Zou, 2015) を使用してモデルの構築及びパラメータの推定を行った．

4. 分析結果

4.1 データ

本稿の分析では，複数の銀行の債権に関する 2005 年から 2014 年までの統合データを用いている．またデフォルトの定義に関しては，企業の債権者区分が破たん懸念先以下に遷移する状況（破懸基準）をデフォルトとして扱っている．このデータを，モデルの構築に用いる期間と，構築したモデルの評価（バックテスト）を行う期間（アウトオブタイム）に分割して分析を行う．なお推定を行う期間の違いによって最適なモデルや結果の評価が影響を受ける可能性もあることから，分析に当たっては以下の表 1 に示すように，期間の分割の仕方を変えた 4 種類のデータセットを用意し，各データセットを対象としてモデルの構築を行い，結果を比較した．分析に用いた財務指標の一覧は表 2 に示している．

4.2 モデル構築及びパラメータ推定の際の設定

モデルの構築に当たっては，以下の設定の下でパラメータの推定等を行った．

表 1．分析に用いたデータセットの種類．

データセット	モデル構築期間	バックテスト期間
データセット 1	2005 年 〜 2013 年	2014 年
データセット 2	2005 年 〜 2012 年	2013 年 〜 2014 年
データセット 3	2005 年 〜 2011 年	2012 年 〜 2014 年
データセット 4	2005 年 〜 2010 年	2011 年 〜 2014 年

表 2. 分析に用いた財務指標の一覧.

(1) エクスポージャー	(21) 金利対現金預金比率	(41) 期末役員従業員数
(2) 融資シェア	(22) 自己資本比率	(42) 資産合計土地比率
(3) 総資産営業利益率	(23) 純資産倍率	(43) 資産合計有形固定資産比率
(4) 総資産経常利益率	(24) 固定長期適合率	(44) 資産合計前払費用比率
(5) ROA	(25) 固定比率	(45) 資産合計流動資産比率
(6) ROE	(26) 借入金依存度	(46) 資産合計固定資産比率
(7) 売上高総利益率	(27) デットキャパシティレシオ	(47) 資産合計支払手形比率
(8) 売上高営業利益率	(28) 預借率	(48) 資産合計受取手形比率
(9) 売上高経常利益率	(29) 借入金月商倍率	(49) 資産合計当座資産比率
(10) 売上高当期利益率	(30) 売上高支払利息割引料率	(50) 資産合計棚卸資産比率
(11) 売上事業キャッシュフロー率	(31) 有利子負債利子率	
(12) 総資産回転率	(32) 現金預金対利子割引料率	
(13) 売上債権回転日数	(33) 支払利息割引料対総利益率	
(14) 棚卸資産回転日数	(34) 事業キャッシュフロー有利子負債比率	
(15) 売上高有形固定資産率	(35) インタレストカバレッジ	
(16) 買入債務回転日数	(36) 事業キャッシュフロー金利負担率	
(17) 流動比率	(37) 減価償却率	
(18) 当座比率	(38) 売上高減価償却率	
(19) 支払準備率	(39) 売上高	
(20) 現預金比率	(40) 資産合計	

変数変換の適用：財務指標によっては売上高のように，少数の企業が非常に大きな値をとるような右に歪んだ分布となる場合がある．このように歪みの強い変数については，対数変換又は neglog 変換を適用し，変数の安定化を図った．その上で，さらに全ての変数に対し，0 から 1 の範囲に収まるように線形変換を行った．

はずれ値への対応：財務指標によっては，上記の変換を行ってもなお，はずれ値が存在することがある．そこで，はずれ値の影響を軽減するため，財務指標を大きさの順にソートし，分布の上下 1% で折返し処理(上下 1% を超える値に対して上下 1% における値を代入)を行った．

欠測値への対応：財務指標によっては，欠測値が存在することがある．そのような場合には中央値を代入して補完を行った．なお今回のデータセットでは欠測値がそれほど多くないため(全体の 5% 程度)，欠測値補完による分析結果への影響は，それほど大きくないと考えられる．

フラグ(ダミー)変数の導入：業種別，銀行別に関するフラグ変数を導入した．なお，これらのフラグ変数には LASSO の罰則を課していない．

チューニングパラメータ λ の決定：Adaptive Group LASSO を適用する際に，チューニングパラメータ λ を決定する必要がある．これについては AUC に基づく 5 重交差検証法により最小となる値を求め，これをベースとして最終的に 1 標準誤差ルール（Hastie et al., 2015; 川野他, 2018)により λ を決定した．AUC については 4.3 節を参照．

4.3 複数の手法に基づくモデルの比較・検証方法

本稿では，パラメータの推定と変数(財務指標)の選択に関する以下の 5 つのモデルについて，各種指標により比較を行った．

(1)**線形モデル + p 値に基づく変数選択[モデル 1]**：線形な 2 項ロジットモデルを基に，2 段階で変数の選択を行う．具体的には，まず全ての変数を用いて推定を行い，p 値が 0.1 以上の変数をモデルから除外する．そして再度パラメータの推定を行い，p 値が 0.05 以上の変数をモデルから除外して，最終的なモデルを決定した．

(2)**線形モデル + LASSO[モデル 2]**：線形な 2 項ロジットモデルを基に，式(2.5)に基づき

パラメータの推定及び変数の選択を行った．LASSO による推定には R のパッケージ `glmnet` (Friedman et al., 2010)を用いた．

(3)**線形モデル + Multistep Adaptive LASSO**[モデル 3]：線形な 2 項ロジットモデルを基に，以下の式(4.1)に基づく Adaptive LASSO を 2 回適用することにより，パラメータの推定及び変数の選択を行った．

$$L_5(\beta_0,\beta_1,\beta_2,\ldots,\beta_p)=\sum_{i=1}^{n}[\delta_i\log(P_i)+(1-\delta_i)\log(1-P_i)]-\lambda\sum_{j=1}^{p}\omega_j|\beta_j| \tag{4.1}$$
$$\omega_j=\begin{cases}|\beta_j|^{-1} & (|\beta_j|>0)\\ \infty & (|\beta_j|=0)\end{cases}$$

(4)**B-スプライン + Group LASSO**[モデル 4]：B-スプラインに基づく 2 項ロジットモデルを基に，式(3.2)に基づく Group LASSO を適用することにより，パラメータの推定及び変数の選択を行った．

(5)**B-スプライン + Multistep Adaptive Group LASSO**[モデル 5]：B-スプラインに基づく 2 項ロジットモデルを基に，式(3.5)に基づく Multistep Adaptive Group LASSO により，パラメータの推定及び変数の選択を行った．

上記の方法により推定したモデル間の比較に用いる各種指標の定義については以下のとおりである（尾木, 2017; 山下・三浦, 2011; 森平, 2009; Engelmann and Raumeier, 2006）．

AUC(**Area Under the Curve**)：AUC は，ROC 曲線(Receiver Operatorating Characteristic curve)の下側部分の面積で定義される指標である．AUC はモデルの順位性(信用スコアの低い(高い)企業ほどデフォルト率が高く(低く)なっているか)を評価するための指標であり，この値が大きいほどデフォルトの予測精度が高いといえる．AUC の計算には R の `pROC` パッケージを用いた．

AR 値(**Accuracy Ratio**)：AR 値は，CAP(Cumulative Accuracy Profiles)曲線の下側面積から計算される統計量である．AR 値と AUC との間には，AR 値 $=2\text{AUC}-1$ という関係があり，これらは同等な統計量であるが，信用リスクモデルの評価には AR 値を用いることが多い．

疑似決定係数(**Pseudo R^2**)：疑似決定係数は，$1-(L_{opt}/L_{init})$ で表される統計量であり，マクファーデンの決定係数とも呼ばれる．ここで L_{init} は定数項のみのロジットモデルの推定を行った場合の対数尤度であり，L_{opt} は財務指標を用いたロジットモデルの推定を行った場合の対数尤度である．疑似決定係数はインサンプルにおけるモデルのデータへの当てはまりを表す指標であり，この値が大きいほど当てはまりが良いといえる．

ブライアスコア：ブライアスコアは，$(1/n)\sum_{i=1}^{n}(P_i-\delta_i)^2$ で表される統計量である．ここで P_i は企業 i のデフォルト確率であり，δ_i は企業 i がデフォルトしていれば 1，非デフォルトであれば 0 となる定数である．ブライアスコアはモデルの一致性(推定されたデフォルト確率と実際のデフォルト率がどの程度近いか)を表す指標であり，この値が小さいほど一致性が高いといえる．

4.4 推定結果

期間の分割の仕方を変えた 4 つのデータセットを対象に分析を行い，説明変数として選択された財務指標について示したものが，表 3 から表 6 である．

全てのデータセットにおいて，提案手法(モデル 5)が，選択された変数の数が最も少なくなっている．また，線形モデル + LASSO(モデル 2)と提案手法(モデル 5)について，各データセットにおいて選択された変数をまとめたものが表 7 である．

表 3. 各推定方法における変数選択の結果：データセット 1.

	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5
(1) エクスポージャー		✓	✓	✓	✓
(2) 融資シェア	✓	✓		✓	
(3) 総資産営業利益率	✓	✓	✓	✓	
(4) 総資産経常利益率	✓	✓	✓	✓	✓
(5) ROA	✓	✓	✓	✓	
(6) ROE	✓	✓	✓		
(7) 売上高総利益率					
(8) 売上高営業利益率	✓				
(9) 売上高経常利益率	✓	✓	✓	✓	
(10) 売上高当期利益率			✓	✓	✓
(11) 売上事業キャッシュフロー率	✓				
(12) 総資産回転率	✓	✓	✓		
(13) 売上債権回転日数	✓	✓		✓	✓
(14) 棚卸資産回転日数	✓	✓			
(15) 売上高有形固定資産率	✓				
(16) 買入債務回転日数	✓	✓		✓	✓
(17) 流動比率	✓			✓	✓
(18) 当座比率	✓			✓	
(19) 支払準備率	✓	✓	✓	✓	
(20) 現預金比率				✓	✓
(21) 金利対現金預金比率	✓		✓	✓	
(22) 自己資本比率				✓	✓
(23) 純資産倍率	✓	✓	✓	✓	✓
(24) 固定長期適合率					
(25) 固定比率					
(26) 借入金依存度		✓			
(27) デットキャパシティレシオ		✓		✓	✓
(28) 預借率	✓				
(29) 借入金月商倍率	✓	✓		✓	✓
(30) 売上高支払利息割引料率	✓			✓	
(31) 有利子負債利子率	✓	✓	✓	✓	✓
(32) 現金預金対利子割引料率	✓	✓	✓	✓	✓
(33) 支払利息割引料対総利益率	✓	✓	✓	✓	✓
(34) 事業キャッシュフロー有利子負債比率	✓	✓	✓	✓	✓
(35) インタレストカバレッジ	✓	✓	✓	✓	
(36) 事業キャッシュフロー金利負担率					
(37) 減価償却率	✓	✓	✓	✓	✓
(38) 売上高減価償却率	✓	✓	✓	✓	
(39) 売上高	✓			✓	
(40) 資産合計	✓			✓	✓
(41) 期末役員従業員数	✓				
(42) 資産合計土地比率	✓	✓		✓	
(43) 資産合計有形固定資産比率	✓	✓	✓	✓	✓
(44) 資産合計前払費用比率	✓	✓	✓	✓	✓
(45) 資産合計流動資産比率					
(46) 資産合計固定資産比率	✓	✓	✓		
(47) 資産合計支払手形比率			✓	✓	
(48) 資産合計受取手形比率	✓	✓	✓	✓	✓
(49) 資産合計当座資産比率	✓		✓		
(50) 資産合計棚卸資産比率	✓	✓	✓	✓	✓
選択された変数の数 (ダミー変数除く)	38	29	25	34	21

モデル 1：線形モデル + p 値に基づく変数選択
モデル 2：線形モデル + LASSO
モデル 3：線形モデル + Multistep Adaptive LASSO
モデル 4：B-スプライン + Group LASSO
モデル 5：B-スプライン + Multistep Adaptive Group LASSO

表 4. 各推定方法における変数選択の結果：データセット 2.

	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5
(1) エクスポージャー		✓	✓	✓	✓
(2) 融資シェア	✓	✓		✓	
(3) 総資産営業利益率	✓	✓	✓	✓	
(4) 総資産経常利益率	✓	✓	✓	✓	
(5) ROA	✓	✓	✓	✓	✓
(6) ROE	✓	✓	✓		
(7) 売上高総利益率				✓	
(8) 売上高営業利益率	✓	✓			
(9) 売上高経常利益率	✓	✓	✓	✓	✓
(10) 売上高当期利益率					
(11) 売上事業キャッシュフロー率	✓				
(12) 総資産回転率	✓	✓	✓		
(13) 売上債権回転日数	✓	✓		✓	✓
(14) 棚卸資産回転日数	✓	✓	✓		
(15) 売上高有形固定資産率	✓				
(16) 買入債務回転日数	✓	✓	✓	✓	✓
(17) 流動比率	✓			✓	✓
(18) 当座比率	✓	✓	✓	✓	
(19) 支払準備率	✓	✓	✓	✓	
(20) 現預金比率				✓	✓
(21) 金利対現金預金比率	✓	✓	✓	✓	
(22) 自己資本比率				✓	✓
(23) 純資産倍率	✓	✓	✓	✓	✓
(24) 固定長期適合率		✓			
(25) 固定比率					
(26) 借入金依存度		✓			
(27) デットキャパシティレシオ		✓		✓	✓
(28) 預借率	✓				
(29) 借入金月商倍率	✓	✓		✓	✓
(30) 売上高支払利息割引料率	✓	✓		✓	
(31) 有利子負債利子率	✓	✓	✓	✓	✓
(32) 現金預金対利子割引料率	✓	✓	✓	✓	✓
(33) 支払利息割引料対総利益率	✓	✓	✓	✓	✓
(34) 事業キャッシュフロー有利子負債比率	✓	✓	✓	✓	✓
(35) インタレストカバレッジ	✓	✓	✓		
(36) 事業キャッシュフロー金利負担率					
(37) 減価償却率		✓	✓	✓	✓
(38) 売上高減価償却率		✓	✓	✓	
(39) 売上高	✓	✓		✓	
(40) 資産合計	✓		✓	✓	✓
(41) 期末役員従業員数	✓	✓			
(42) 資産合計土地比率	✓	✓	✓	✓	
(43) 資産合計有形固定資産比率	✓	✓	✓	✓	✓
(44) 資産合計前払費用比率	✓	✓	✓	✓	✓
(45) 資産合計流動資産比率					
(46) 資産合計固定資産比率	✓	✓	✓		
(47) 資産合計支払手形比率		✓	✓	✓	
(48) 資産合計受取手形比率	✓	✓	✓	✓	✓
(49) 資産合計当座資産比率	✓		✓		
(50) 資産合計棚卸資産比率	✓	✓	✓	✓	✓
選択された変数の数 (ダミー変数除く)	36	37	29	33	21

モデル 1：線形モデル + p 値に基づく変数選択
モデル 2：線形モデル + LASSO
モデル 3：線形モデル + Multistep Adaptive LASSO
モデル 4：B-スプライン + Group LASSO
モデル 5：B-スプライン + Multistep Adaptive Group LASSO

表 5. 各推定方法における変数選択の結果：データセット 3.

	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5
(1) エクスポージャー		✓	✓		✓
(2) 融資シェア	✓	✓		✓	
(3) 総資産営業利益率	✓	✓	✓	✓	
(4) 総資産経常利益率		✓	✓	✓	
(5) ROA	✓	✓	✓	✓	✓
(6) ROE	✓	✓	✓		
(7) 売上高総利益率					✓
(8) 売上高営業利益率	✓				
(9) 売上高経常利益率	✓	✓	✓	✓	✓
(10) 売上高当期利益率		✓	✓		
(11) 売上事業キャッシュフロー率	✓				
(12) 総資産回転率	✓	✓	✓		
(13) 売上債権回転日数	✓	✓		✓	✓
(14) 棚卸資産回転日数	✓	✓			
(15) 売上高有形固定資産率	✓				
(16) 買入債務回転日数	✓	✓	✓	✓	✓
(17) 流動比率	✓			✓	✓
(18) 当座比率	✓				
(19) 支払準備率	✓	✓	✓	✓	
(20) 現預金比率				✓	✓
(21) 金利対現金預金比率	✓		✓	✓	
(22) 自己資本比率				✓	✓
(23) 純資産倍率	✓	✓	✓	✓	✓
(24) 固定長期適合率					
(25) 固定比率					
(26) 借入金依存度		✓	✓		
(27) デットキャパシティレシオ		✓		✓	✓
(28) 預借率	✓				
(29) 借入金月商倍率	✓	✓		✓	✓
(30) 売上高支払利息割引料率	✓		✓	✓	
(31) 有利子負債利子率	✓	✓	✓	✓	✓
(32) 現金預金対利子割引料率	✓	✓	✓	✓	✓
(33) 支払利息割引料対総利益率	✓	✓	✓	✓	✓
(34) 事業キャッシュフロー有利子負債比率	✓	✓	✓	✓	✓
(35) インタレストカバレッジ	✓	✓	✓		
(36) 事業キャッシュフロー金利負担率					
(37) 減価償却率		✓	✓	✓	✓
(38) 売上高減価償却率		✓	✓	✓	
(39) 売上高	✓				
(40) 資産合計	✓		✓	✓	✓
(41) 期末役員従業員数	✓	✓			
(42) 資産合計土地比率	✓	✓			
(43) 資産合計有形固定資産比率	✓	✓	✓	✓	✓
(44) 資産合計前払費用比率	✓	✓	✓	✓	✓
(45) 資産合計流動資産比率					
(46) 資産合計固定資産比率	✓	✓	✓		
(47) 資産合計支払手形比率			✓	✓	
(48) 資産合計受取手形比率	✓	✓	✓	✓	✓
(49) 資産合計当座資産比率	✓		✓		
(50) 資産合計棚卸資産比率	✓	✓	✓	✓	✓
選択された変数の数 (ダミー変数除く)	35	31	29	28	22

モデル 1：線形モデル + p 値に基づく変数選択
モデル 2：線形モデル + LASSO
モデル 3：線形モデル + Multistep Adaptive LASSO
モデル 4：B-スプライン + Group LASSO
モデル 5：B-スプライン + Multistep Adaptive Group LASSO

表 6. 各推定方法における変数選択の結果：データセット 4.

	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5
(1) エクスポージャー		✓	✓		
(2) 融資シェア		✓		✓	
(3) 総資産営業利益率	✓		✓		
(4) 総資産経常利益率		✓	✓	✓	
(5) ROA	✓	✓	✓	✓	✓
(6) ROE		✓	✓		
(7) 売上高総利益率					✓
(8) 売上高営業利益率	✓				
(9) 売上高経常利益率	✓	✓	✓	✓	✓
(10) 売上高当期利益率		✓	✓		
(11) 売上事業キャッシュフロー率	✓				
(12) 総資産回転率	✓		✓		
(13) 売上債権回転日数	✓	✓		✓	✓
(14) 棚卸資産回転日数	✓	✓			
(15) 売上高有形固定資産率	✓				
(16) 買入債務回転日数	✓	✓		✓	✓
(17) 流動比率		✓		✓	✓
(18) 当座比率	✓				
(19) 支払準備率	✓		✓		
(20) 現預金比率				✓	
(21) 金利対現金預金比率	✓		✓		
(22) 自己資本比率				✓	✓
(23) 純資産倍率	✓	✓	✓	✓	✓
(24) 固定長期適合率					
(25) 固定比率					
(26) 借入金依存度		✓	✓		
(27) デットキャパシティレシオ		✓		✓	✓
(28) 預借率					
(29) 借入金月商倍率	✓	✓		✓	✓
(30) 売上高支払利息割引料率	✓		✓	✓	
(31) 有利子負債利子率	✓	✓	✓	✓	✓
(32) 現金預金対利子割引料率	✓	✓	✓	✓	✓
(33) 支払利息割引料対総利益率	✓	✓	✓	✓	✓
(34) 事業キャッシュフロー有利子負債比率	✓	✓	✓	✓	
(35) インタレストカバレッジ	✓	✓	✓	✓	
(36) 事業キャッシュフロー金利負担率					
(37) 減価償却率		✓		✓	✓
(38) 売上高減価償却率		✓	✓		
(39) 売上高	✓			✓	
(40) 資産合計	✓		✓	✓	✓
(41) 期末役員従業員数	✓				
(42) 資産合計土地比率	✓	✓			
(43) 資産合計有形固定資産比率	✓	✓	✓	✓	✓
(44) 資産合計前払費用比率	✓	✓	✓	✓	✓
(45) 資産合計流動資産比率					
(46) 資産合計固定資産比率	✓	✓	✓		
(47) 資産合計支払手形比率			✓	✓	
(48) 資産合計受取手形比率	✓	✓	✓	✓	✓
(49) 資産合計当座資産比率			✓		
(50) 資産合計棚卸資産比率	✓			✓	✓
選択された変数の数 (ダミー変数除く)	30	27	26	26	19

モデル 1：線形モデル + p 値に基づく変数選択
モデル 2：線形モデル + LASSO
モデル 3：線形モデル + Multistep Adaptive LASSO
モデル 4：B-スプライン + Group LASSO
モデル 5：B-スプライン + Multistep Adaptive Group LASSO

表 7. モデル 2 及びモデル 5 において選択された変数.

モデル	モデル 2				モデル 5			
データセット	1	2	3	4	1	2	3	4
(1) エクスポージャー	✓	✓	✓	✓	✓	✓	✓	
(2) 融資シェア	✓	✓	✓	✓				
(3) 総資産営業利益率	✓	✓	✓					
(4) 総資産経常利益率	✓	✓	✓	✓	✓			
(5) ROA	✓	✓	✓	✓		✓	✓	✓
(6) ROE	✓	✓	✓	✓				
(7) 売上高総利益率							✓	✓
(8) 売上高営業利益率		✓						
(9) 売上高経常利益率	✓	✓	✓	✓		✓	✓	✓
(10) 売上高当期利益率			✓	✓	✓			
(11) 売上事業キャッシュフロー率								
(12) 総資産回転率	✓	✓	✓					
(13) 売上債権回転日数	✓	✓	✓	✓	✓	✓	✓	✓
(14) 棚卸資産回転日数	✓	✓	✓	✓				
(15) 売上高有形固定資産率								
(16) 買入債務回転日数	✓	✓	✓	✓	✓	✓	✓	✓
(17) 流動比率				✓	✓	✓	✓	✓
(18) 当座比率		✓						
(19) 支払準備率	✓	✓	✓					
(20) 現預金比率					✓	✓	✓	
(21) 金利対現金預金比率		✓						
(22) 自己資本比率					✓	✓	✓	✓
(23) 純資産倍率	✓	✓	✓	✓	✓	✓	✓	✓
(24) 固定長期適合率		✓						
(25) 固定比率								
(26) 借入金依存度	✓	✓	✓	✓				
(27) デットキャパシティレシオ	✓	✓	✓	✓	✓	✓	✓	✓
(28) 預借率								
(29) 借入金月商倍率	✓	✓	✓	✓	✓	✓	✓	✓
(30) 売上高支払利息割引料率		✓						
(31) 有利子負債利子率	✓	✓	✓	✓	✓	✓	✓	✓
(32) 現金預金対利子割引料率	✓	✓	✓	✓	✓	✓	✓	✓
(33) 支払利息割引料対総利益率	✓	✓	✓	✓	✓	✓	✓	✓
(34) 事業キャッシュフロー有利子負債比率	✓	✓	✓	✓	✓	✓	✓	
(35) インタレストカバレッジ	✓	✓	✓	✓				
(36) 事業キャッシュフロー金利負担率								
(37) 減価償却率	✓	✓	✓	✓	✓	✓	✓	✓
(38) 売上高減価償却率	✓	✓	✓	✓				
(39) 売上高		✓						
(40) 資産合計					✓	✓	✓	✓
(41) 期末役員従業員数		✓	✓					
(42) 資産合計土地比率	✓	✓	✓	✓				
(43) 資産合計有形固定資産比率	✓	✓	✓	✓	✓	✓	✓	✓
(44) 資産合計前払費用比率	✓	✓	✓	✓	✓	✓	✓	✓
(45) 資産合計流動資産比率								
(46) 資産合計固定資産比率	✓	✓	✓	✓				
(47) 資産合計支払手形比率		✓						
(48) 資産合計受取手形比率	✓	✓	✓	✓	✓	✓	✓	✓
(49) 資産合計当座資産比率								
(50) 資産合計棚卸資産比率	✓	✓	✓		✓	✓	✓	✓
選択された変数の数 (ダミー変数除く)	29	37	31	27	21	21	22	19

モデル 2：線形モデル + LASSO
モデル 5：B-スプライン + Multistep Adaptive Group LASSO

線形モデル + LASSO(モデル 2)ではデータセットによって(特にデータセット 2 とそれ以外で)選択される変数が大きく異なる場合があるのに対し，提案手法(モデル 5)による推定結果で

表8．各推定方法における推定結果の比較(太字は最も良いもの)．

データセット1	交差検証法					アウトオブタイム				
	モデル1	モデル2	モデル3	モデル4	モデル5	モデル1	モデル2	モデル3	モデル4	モデル5
AUC	0.84064	0.83690	0.83469	0.84803	**0.84923**	0.88175	0.87759	0.87140	0.88360	**0.88417**
AR値	0.68127	0.67380	0.66938	0.69606	**0.69847**	0.76350	0.75517	0.74285	0.76724	**0.76835**
ブライアスコア	0.01189	0.01192	0.01193	0.01181	**0.01180**	0.00663	0.00666	0.00680	**0.00658**	**0.00658**
疑似決定係数	0.15752	0.15080	0.14901	0.16742	**0.17023**			—		
サンプルサイズ			642,025					67,250		
デフォルト件数			7,980					458		
デフォルト率			1.24%					0.68%		

データセット2	交差検証法					アウトオブタイム				
	モデル1	モデル2	モデル3	モデル4	モデル5	モデル1	モデル2	モデル3	モデル4	モデル5
AUC	0.83760	0.83574	0.83235	0.84571	**0.84626**	0.86960	0.86801	0.84470	**0.87380**	0.87319
AR値	0.67519	0.67147	0.66471	0.69141	**0.69253**	0.73919	0.73603	0.68934	**0.74763**	0.74637
ブライアスコア	0.01233	0.01236	0.01238	**0.01225**	**0.01225**	0.00742	0.00743	0.00766	**0.00739**	**0.00739**
疑似決定係数	0.15529	0.15209	0.14801	0.16586	**0.16774**			—		
サンプルサイズ			572,772					136,503		
デフォルト件数			7,392					1,046		
デフォルト率			1.29%					0.77%		

データセット3	交差検証法					アウトオブタイム				
	モデル1	モデル2	モデル3	モデル4	モデル5	モデル1	モデル2	モデル3	モデル4	モデル5
AUC	0.83516	0.83137	0.83062	0.84118	**0.84379**	0.86500	0.86266	0.85120	0.87050	**0.87116**
AR値	0.67032	0.66273	0.66123	0.68236	**0.68758**	0.73001	0.72533	0.70245	0.74097	**0.74231**
ブライアスコア	0.01253	0.01257	0.01258	0.01246	**0.01244**	0.00866	0.00869	0.01158	**0.00861**	**0.00861**
疑似決定係数	0.15340	0.14729	0.14617	0.16109	**0.16577**			—		
サンプルサイズ			500,270					209,005		
デフォルト件数			6,563					1,875		
デフォルト率			1.31%					0.90%		

データセット4	交差検証法					アウトオブタイム				
	モデル1	モデル2	モデル3	モデル4	モデル5	モデル1	モデル2	モデル3	モデル4	モデル5
AUC	0.83042	0.82684	0.82450	0.83693	**0.83994**	0.86212	0.85889	0.84480	0.86570	**0.86571**
AR値	0.66085	0.65368	0.64900	0.67386	**0.67987**	0.72424	0.71778	0.68965	0.73136	**0.73143**
ブライアスコア	0.01275	0.01279	0.01280	0.01268	**0.01266**	0.00935	0.00938	0.01647	0.00930	**0.00929**
疑似決定係数	0.14938	0.14310	0.14101	0.15686	**0.16215**			—		
サンプルサイズ			427,697					281,578		
デフォルト件数			5,708					2,730		
デフォルト率			1.33%					0.97%		

は，選択された変数にそれほど大きな違いはなく，安定した推定結果となっている．

交差検証法(モデル構築期間)及びバックテスト(アウトオブタイム)における推定結果を示したものが表8である．

データセット2(モデル構築期間：2005年～2012年)のアウトオブタイムのサンプルにおけるAUC及びAR値を除いて，いずれのデータセットにおいても，提案手法(モデル5)が最も良い性能を示しており，他のモデルと比較して，AR値や疑似決定係数などの観点から推定精度が向上していることがわかる．

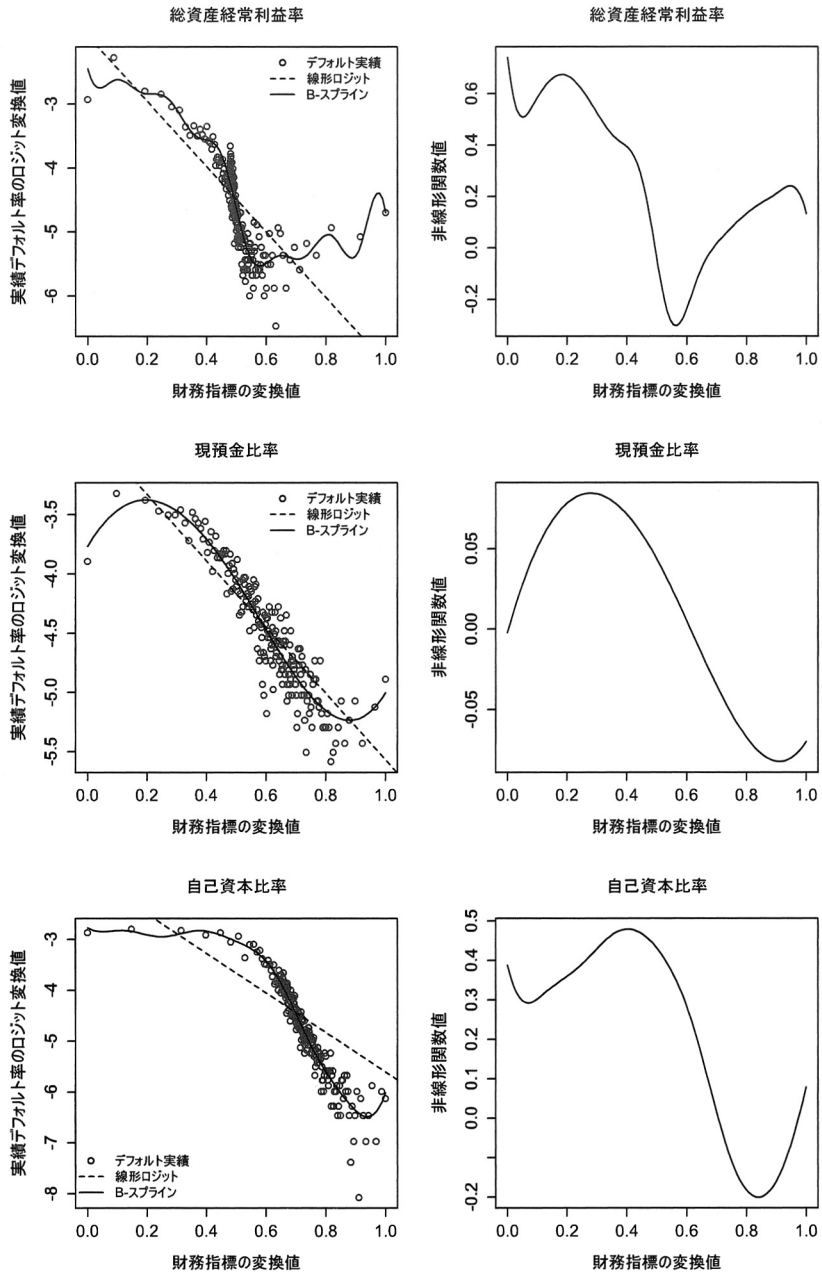

図 3. 非線形関数の推定結果(1)左列：図 1 再掲　右列：非線形関数の推定値(図中の「現預金比率」については neglog 変換を行っている．また全ての財務指標について，0 から 1 の範囲に収まるように線形変換を行っている．)．

提案手法(モデル 5)に基づき，データセット 1(モデル構築期間：2005 年〜2013 年)に対して推定された一部の財務指標に関する非線形関数(式(3.1)における f_j)を示したものが図 3 及び図 4 である．実績デフォルト率との比較のため，図 1 を再掲している．推定された非線形関数

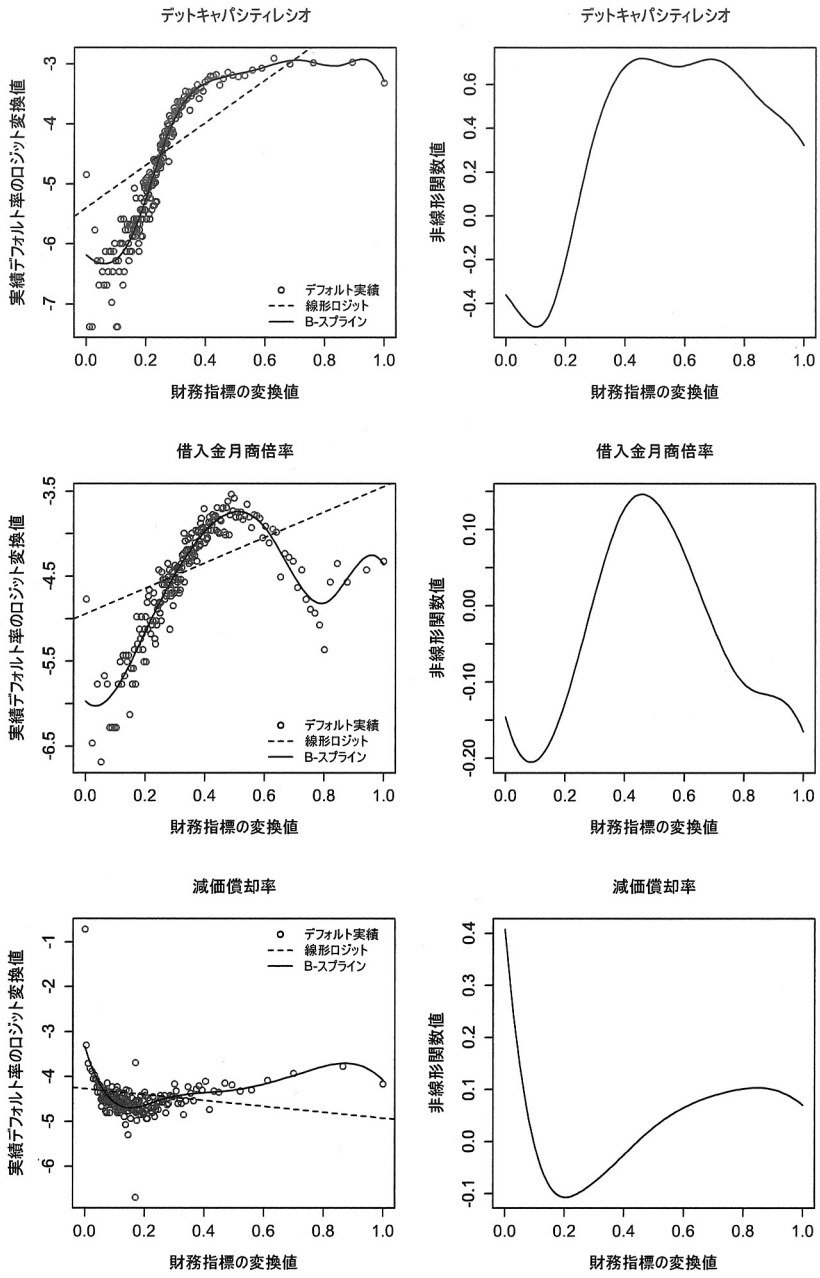

図 4. 非線形関数の推定結果(2)左列：図 1 再掲　右列：非線形関数の推定値(図中の「デットキャパシティレシオ」及び「借入金月商倍率」については neglog 変換を行っている．また全ての財務指標について，0 から 1 の範囲に収まるように線形変換を行っている.)．

(右の列)は，実績デフォルト率の変動(左の列)を，ある程度捉えていることがわかる．ただし横軸で 0 又は 1 に近い領域では，サンプルサイズが小さいため，変動に幅があることに注意する必要がある．

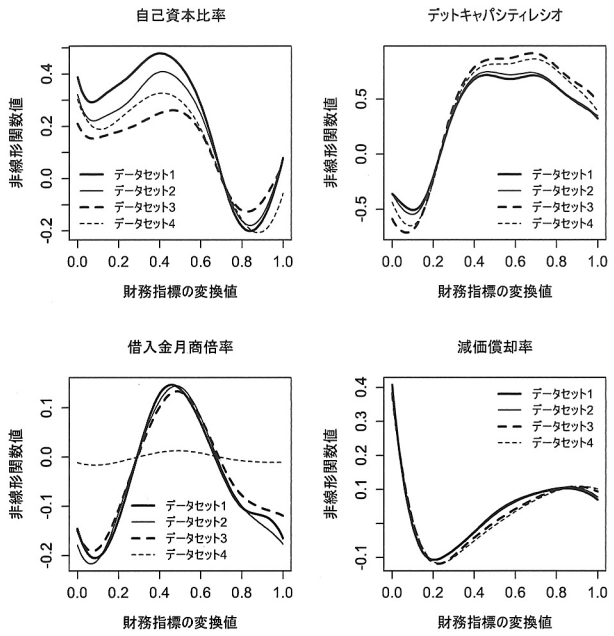

図 5. 各データセットにおける非線形関数の推定結果(図中の「デットキャパシティレシオ」及び「借入金月商倍率」については neglog 変換を行っている．また全ての財務指標について，0 から 1 の範囲に収まるように線形変換を行っている.)．

図 1 に示した財務指標の中で，提案手法(モデル 5)において，全てのデータセットで変数として選択されている「自己資本比率」,「デットキャパシティレシオ」,「借入金月商倍率」及び「減価償却率」の 4 つの財務指標について，各データセットから推定された非線形モデルの予測値を重ねて表示したものが図 5 である．図 5 をみると，モデルを構築する際に用いるデータの期間の違いによって，推定される非線形関数の水準は異なるものの，期間が異なっても非線形関数の形状には大きな違いはないことがわかる．なお，借入金月商倍率に関しては，データセット 4(モデル構築期間：2005 年〜2010 年)において，非線形関数の値が他のデータセットの場合と比較して 0 に近く，フラットに近い形状であるものの，上昇・下降のパターンは他のデータセットの場合と同様である．

5. 考察

5.1 モデルの精度

本稿では複数の銀行データを統合したデータベースを基に，B-スプラインに基づく非線形モデル及び Multistep Adaptive Group LASSO に基づく変数選択の手法を導入したデフォルト確率予測モデルの構築を行った．このようにして得られたモデルは，t 値・p 値に基づく変数選択や単純な LASSO による方法と比較して，どの期間のデータセットにおいても最も変数が少なくなっており，選択された変数の種類に大きな変動がなく，効率的かつ安定的な変数選択を行うことができた．さらに AR 値などの各種指標を用いて比較を行った結果，本稿で提案したモデルが最も推定精度が高く，当てはまりの良いモデルであることが確認された．B-スプラインに基づく非線形モデルの導入により，信用スコアと財務指標との非線形な構造を捉えること

が可能となり，モデルの推定精度が向上したと考えられる．さらに Multistep Adaptive Group LASSO に基づく変数選択の手法を導入することにより，よりコンパクトなモデルを推定することが可能となり，モデルの安定性が向上したことで，アウトオブタイムにおける推定精度の向上につながったものと考えられる．

5.2 財務指標の選択

表 3 から表 6 において，提案手法(モデル 5)の説明変数として，異なるデータセットで複数回選択された変数を見ると，利益，回転率，短期支払能力といった総合的な収益性の面から「ROA」,「売上高経常利益率」,「売上債権回転日数」,「買入債務回転日数」,「流動比率」,「現預金比率」といった，実務でもよく用いられる代表的な財務指標が選択されている．これに対してデフォルト予測や与信判断に直接的に関係すると考えられる借入・資産の面からは「デットキャパシティレシオ」,「借入金月商倍率」,「有利子負債利子率」,「現金預金対利子割引料率」,「自己資本比率」,「減価償却率」などのほか,「資産合計」やこれに占める各種資産の割合など，多くの財務指標が選択されている．収益性に関する指標を代表的なものに絞りつつ，借入・資産に重点を置くという，メリハリのある変数選択が行われている．

5.3 推定された非線形関数の形状

提案手法(モデル 5)に基づき推定された，主な財務指標の非線形関数の形状について考察する．総資産経常利益率は高い方が望ましいが，資金の必要性から総資産を処分する際に高くなる可能性もあり，極端に高すぎる又は低すぎる値は望ましくないと考えられる．自己資本比率は高い方が，デットキャパシティレシオ(有利子負債と融資の担保にできる資産との比)は低い方が望ましいが，どちらもある程度の水準を満たしていればよい指標であり，一定値以上(以下)で頭打ちになると想定される．減価償却率については，早目に償却した方が安全である一方，逆に償却が進むと経費計上分が減少してしまうという観点もある．図 3 及び図 4 における非線形関数の形状には，これらの関係が表れていると考えられる．

一部の財務指標について，モデル構築に用いるデータの期間が異なる場合における非線形関数の形状の変化を見ると，図 5 に示すように，推定される非線形関数の水準は異なるものの，期間が異なっても非線形関数の形状には大きな違いはなく，安定していることが示された．このようにして推定された各財務指標の非線形関数を用いることで，財務指標ごとに信用スコアが急激に変化する点や最も高くなる点などを判別することが可能となり，与信判断に資する情報が得られるものと期待される．

本研究において提案したデフォルト確率予測モデルは，財務指標と信用スコアとの非線形な関係が「見える」モデルの合理的・効率的な構築に寄与するものであり，各種財務指標に基づいて与信判断・審査等を行う金融実務において，有益であると考えられる．

6. 今後の課題

今後の課題として，以下の点が挙げられる．今回の分析では計算のコストを考慮して，Multistep Adaptive Group LASSO における反復回数を 2 回としたが，反復回数を多くすることがモデルの推定精度の改善に寄与するかという点に関しては検討の余地が残されている．

また，今回の手法を，より大規模なデータセットに対して分析を行うことが考えられる．具体的には，複数のデータベースを結合して得られた大規模なデータベースに対して適用することで，より多くの変数から効率的に非線形な構造を抽出できると考えられる．

謝　辞

本研究は科研費(16H02013 及び 15H03390)の助成を受けています．また改稿に当たり，有益なコメントをいただいた 2 名の査読者に感謝申し上げます．

参 考 文 献

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, **23**, 589–609.

Amendola, A., Restaino, M. and Sensini, L. (2012). Dynamic statistical models for corporate failure prediction in Italy, *Journal of Modern Accounting and Auditing*, **8**, 1214–1224.

Berg, D. (2007). Bankruptcy prediction by generalized additive models, *Applied Stochastic Models in Business and Industry*, **23**, 129–143.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data*: *Methods, Theory and Applications*, Springer, Berlin.

Duffie, D. and Singleton, K. J. (1999). Modeling term structures of defaultable bonds, *Review of Financial Studies*, **12**, 687–720.

Dwyer, D. W., Kocagil, A. E. and Stein, R. M. (2004). The Moody's KMV EDF RiskCalc v3.1 Model: Next generation technology for predicting private firm risk, Moody's KMV Company, San Francisco.

Engelmann, B. and Raumeier, R. (2006). *The Basel* II *Risk Parameters*: *Estimation, Validation and Stress Testing*, Springer, Berlin.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, **33**, 1–22.

Giordani, P., Jacobson, T., von Schedvin, E. and Villani, M. (2014). Taking the twists into account: Predicting firm bankruptcy risk with splines of financial ratios, *Journal of Financial and Quantitative Analysis*, **49**, 1071–1099.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman & Hall/CRC, Boca Raton, Florida.

Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity*: *The Lasso and Generalizations*, Chapman & Hall/CRC, Boca Raton, Florida.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied Logistic Regression*: *Third Edition*, Wiley, New York.

Huang, J., Horowitz, J. L. and Wei, F. (2010). Variable selection in nonparametric additive models, *Annals of Statistics*, **38**, 2282–2313.

川野秀一, 松井秀俊, 廣瀬慧 (2018).『スパース推定法による統計モデリング』, 共立出版, 東京.

小西貞則 (2010).『多変量解析入門：線形から非線形へ』, 岩波書店, 東京.

Martin, D. (1977). Early warning of bank failure: A logit regression approach, *Journal of Banking and Finance*, **1**, 249–276.

Meier, L., van de Geer, S. and Bühlmann, P. (2008). The group lasso for logistic regression, *Journal of the Royal Statistical Society Series B*, **70**, 53–71.

Meier, L., van de Geer, S. and Bühlmann, P. (2009). High-dimensional additive modeling, *Annals of Statistics*, **37**, 3779–3821.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance*, **29**, 449–470.

森平爽一郎 (2009).『信用リスクモデリング：測定と管理』, 朝倉書店, 東京.

尾木研三 (2017).『スコアリングモデルの基礎知識：中小企業融資における見方・使い方』, 金融財政事情研究会, 東京.

尾木研三, 戸城正浩, 枇々木規雄 (2015). 小規模企業向け保善別回収率モデルの構築と実証分析,『ファイナンスとデータ解析(ジャフィー・ジャーナル：金融工学と市場計量分析)』(日本金融・証券計量・工学学会 編), 168–201, 朝倉書店, 東京.

Perederiy, V. (2009). Bankruptcy prediction revisited: Non-traditional ratios and lasso selection, European University Viadrina, Working Paper 16, Frankfurt.

阪本亘, 高橋史朗, 竹内正弘 (2010). 正則化法を用いたロジスティック回帰モデルによる多次元データでの変数選択手法に関する研究, 数理解析研究所講究録, **1703**, 32–52.

桜井明 (1981).『スプライン関数入門：情報処理の新しい手法』, 東京電機大学出版局, 東京.

白田佳子 (2008).『倒産予知モデルによる格付けの実務』, 中央経済社, 東京.

高橋久尚, 山下智志 (2002). 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, 統計数理, **50**, 241–258.

Tian, S., Yu, Y. and Guo, H. (2015). Variable selection and corporate bankruptcy forecasts, *Journal of Banking and Finance*, **52**, 89–100.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B*, **58**, 267–288.

冨岡亮太 (2015).『スパース性に基づく機械学習』, 講談社, 東京.

山下智志, 安道知寛 (2006). 時間依存共変量を用いたハザードモデルによるデフォルト確率期間構造の推計手法, 統計数理, **54**, 23–38.

山下智志, 三浦翔 (2011).『信用リスクモデルの予測精度：AR 値と評価指標』, 朝倉書店, 東京.

山内浩嗣 (2010). 多目的遺伝的アルゴリズムを用いたスコアリングモデルのチューニング,『定量的信用リスク評価とその応用(ジャフィー・ジャーナル：金融工学と市場計量分析)』(日本金融・証券計量・工学学会 編), 24–54, 朝倉書店, 東京.

Yang, Y. and Zou, H. (2015). A fast unified algorithm for solving group-lasso penalized learning problems, *Statistics and Computing*, **25**, 1129–1141.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society*: *Series B*, **68**, 49–67.

Estimation of Default Probability Using Regularized Nonlinear Logit Model with B-spline and Adaptive Group LASSO

Isao Takabe[1,2] and Satoshi Yamashita[3]

[1]Department of Statistical Science, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies
[2]Consumer Statistics Division, Statistics Bureau, Ministry of Internal Affairs and Communications
[3]The Institute of Statistical Mathematics

Linear binomial logit models are widely used for the assessment and evaluation of a company's default probability based on a company default database. Previous studies have been criticized on the following bases: (1) insufficient attention to nonlinear relationships between default probabilities and financial indicators; and (2) too much time required for variable selection from many candidates for regressors in the models. In this study, we aimed to solve these problems simultaneously by combining the following techniques: (1) nonlinear and nonparametric logistic regression model based on the B-spline; and (2) reasonable variable selection using adaptive group LASSO. We constructed a default probability prediction model using datasets of multiple periods, based on our own database of data from Japanese banks. The proposed model achieved more effective performance than models in other related studies. Compared with the method using t-statistic (p-value) or simple LASSO, our proposed method had the smallest number of explanatory variables in any period, and achieved more efficient variable selection. Moreover, estimation accuracy was improved from the viewpoint of AR (accuracy ratio) value.

Key words: Credit risk, B-spline, adaptive group LASSO.