

# 格付予測モデルの構築における サンプルセレクションバイアス

宋 明 子\*

格付を予測モデルを構築するにあたり、モデルの目的変数である格付は観測されるものもあれば観測されないものもある、むしろ観測されない企業の方が多い。この時、格付が観測される企業のデータのみで格付予測モデルを構築すると、サンプルセレクションバイアスの存在の可能性を排除することができない。本稿では、サンプルセレクションバイアスの有無を確認する手法として、格付が観測されるか否かの事象と格付が観測される場合どの格付が付与されているかの二つの事象を同時にモデル化する相関を持つ二変量順序プロビットモデルを提案し、推定される相関係数よりサンプルセレクションバイアスの有無を確認した。モデルの推定結果より、上記の二つの事象には高い相関があることが判明し、格付が観測されるデータは比較的高い格付が付与される確率が高いことが確認できた。言い換えると、格付が観測されるか否かの事象と格付が観測される場合どの格付が付与されているかの二つの事象に相関がある場合、格付が観測されるか否かの事象を無視して、格付が観測されるデータのみでモデル構築を行うと、モデルから算出される予測格付はサンプルセレクションバイアスを含んだ格付であり、実態より高めの格付が算出される傾向があることが確認できた。

キーワード：格付，サンプルセレクションバイアス，順序ロジット，二変量順序プロビット

## 1. イントロダクション

本節では、金融機関の信用リスク計量化に用いられている格付及び格付予測モデルについて概要を説明した上で、格付予測モデルを構築するにあたっての本研究の新規性を明確にし、本

研究での主な分析手法及び分析結果をまとめる。

### 1.1. 格付と格付予測モデル

近年格付は信用リスクを表す重要な指標として注目され活用されている。特に、金融機関にとって大型倒産も発生しているなか貸出先の信用リスクをより精緻に把握することは重要な課題となっている。金融機関において格付がどうして重要な指標となっているかまた格付の活用方法などについては、炭谷（2008）を参照。

その前に格付とは何か、格付投資情報センター編（1998, p.17）では以下のように定義している。「格付とは、債券などの元本、利息が

---

\*株式会社金融工学研究所：〒103-0027 東京都中央区日本橋1-4-1 日本橋一丁目三井ビルディング19F

本稿を作成するにあたり、ご指導いただいた横浜国立大学の小林正人教授に感謝いたします。なお、本稿で示されている内容および意見は筆者個人に属し、金融工学研究所の公式見解を示すものではありません。

期限とおり、約定とおりに支払えるか、その安全度を等級（以降本文中ではノッチと呼称）で示したものである。安全度はアルファベットで表すが、最上級のトリプル A (AAA) からダブル A (AA)、シングル A (A)、トリプル B (BBB)、ダブル B (BB) と段階的に下がるに従って、安全度が低下する。安全度を表す格付符号は格付機関によって異なる。日本にある主な格付機関は、ムーディーズ、スタンダード・アンド・プアーズ、フィッチレーティング、格付投資情報センター（以降本文中では R&I と呼称）、日本格付研究所などがある。それぞれの格付機関の格付の定義および特徴については、黒沢（2007）と黒沢（2009）を参照。

信用格付は信用リスク管理において欠かせない重要な指標であるため、格付予測モデルの研究および構築も盛んでいる。現在金融機関において最も広く使用されている信用リスク計量モデルは大きく分けて 2 つに分類される。一つは、個別企業の倒産確率を推定するモデルであり、二項ロジットモデルが最も利用されている。もう一つは、個々の企業や債券の信用格付を推定するモデルであり、順序ロジットモデルが最も利用されている。二項ロジットおよび順序ロジットなどの詳細については、木島・小守林（2006）を参照。

個別企業の倒産確率を推定するモデルは、大きく分けて二種類のモデルに区分される。一つは、企業の過去の財務データを用いて将来倒産する確率を求める統計モデルである。統計モデルについては、Altman（1968）の Z スコアモデルからはじめ、様々なバリエーションがある。山下・川口（2003）の非上場企業の過去の財務データを用いて将来デフォルトする確率を求める二項ロジットモデル、山下・安道（2006）の非上場企業のデフォルト率の期間構造を表現したハザードモデルなどがある。もう一つは、市場情報を用いて将来デフォルトする確率を求める構造型モデルである。代表的な構造型モデルには、Merton（1974）の上場企業の株価などを用いて企業の資産価値の変動を予測し、将来

負債を下回る確率を計算する Merton モデルなどがある。

格付予測モデルは主に順序ロジットモデルを用いて構築されるが、モデルの説明変数にどのような情報を用いるかによりいくつかの発展がある。まずは説明変数として企業の決算情報から得られる財務指標を用いるのが最も普通である。安川（2003）ではいくつかの格付機関で重視されている財務指標を用いて格付を予測するモデルを構築している。実際格付機関が格付を付与する際には、企業の財務指標以外にもたくさんの定性情報などを用いるのが普通である。そのため、近年では財務指標と定性情報の両方を用いて格付予測モデルを構築する試みもある。また成松（2007）では、企業の諸財務指標に個別企業のデフォルト確率を説明変数に追加して、二段階による格付予測モデルを構築している。これらの先行研究はいずれも目的変数は公表されている格付機関の格付を用いて分析しており、多くの研究は説明変数に更なる情報を追加したり、モデルの構造を二段階にしたりすることで、モデルの予測精度を上げることを目的としている。モデルの目的変数である格付についてはデータの収集を含め特段議論は行っていない。

## 1.2. 本研究の目的と新規性

どのようなモデルでもモデル構築に用いるデータの違いにより、構築されたモデルの出力結果が変わる可能性がある。通常の場合モデル構築は入手可能な観測されるデータを用いて行う。モデル構築に用いる観測されるデータがある大きな母集団から抽出された一部のサンプルである時は、モデル構築用データにサンプルセクションバイアスが含まれる可能性があるため、サンプルセクションバイアスの存在有無を確認する必要がある。

信用スコアリングモデルにおいてサンプルの選択にバイアスが存在することは知られており、推定量の不偏性や一致性に影響が生じること指摘されている。詳細については、

Banasik *et al.* (2003) 及び Greene (1998) を参照。Banasik *et al.* (2003) では補助投与者のデータ、Greene (1998) ではクレジットカードローンのデータを用いて、それぞれのスコアリングモデルにおけるサンプルセレクションバイアスの存在を確認した。どの論文もモデル構築に用いるデータがある大きな母集団から抽出される場合、そのデータの抽出過程も同時にモデル化することで、サンプルセレクションバイアスの有無を確認し調整する必要があることを主張している。同様な考え方より、非上場企業の倒産確率を推定する信用スコアリングモデルの構築においても、サンプルセレクションバイアスが存在する可能性はある。なぜならモデル構築に用いられる観測されるデータは金融機関が融資を行っている企業のみであり、融資を行っていない企業或いは融資を実行する前の審査において不合格となった企業についてはほとんどの情報がデータとして登録されていないため、これらの情報はモデルに反映されない。モデル構築に用いられるデータにある一定の傾向がある場合、例えば金融機関が融資を行っている企業は融資を行っていない企業と比べて比較的信用力が高い企業であると考えられる場合、モデル構築データにはサンプルセレクションバイアスが存在する可能性がある。ただし、金融機関が融資を行っていない企業のデータについては金融機関ごとにデータの登録基準などが異なり、データの登録が不十分であるとのデータ利用上の制約もある。

格付予測モデルの構築におけるサンプルセレクションバイアスについても、倒産確率を推定するモデルの場合と同様に、格付が観測される企業は日本全国の企業のうち極一部であり、格付が観測されない企業と比べて一定の傾向がある場合、例えば格付が観測される企業は格付が観測されない企業に比べて比較的信用力が高いなどの傾向がある場合、モデル構築データにはサンプルセレクションバイアスが存在する可能性がある。つまり、ある母集団から抽出した標本の一部しかモデル構築に利用しないことにな

り、標本選択問題が起きていると考えられる。標本選択においてバイアスが存在する場合、モデルの推定結果にもバイアスがかかり、正しい予測格付が算出されないことになる。このような格付予測モデルの構築における標本選択問題については先行研究がなく、本稿で初めて目的変数である格付に注目し、格付予測モデルを構築するに当たり、モデル構築に用いるデータにサンプルセレクションバイアスがあるかどうかを確認し、サンプルセレクションバイアスがある場合の対処方法について提案する。

以下ではまず格付機関が公表している格付のみを用いて格付予測モデルを構築する際にどうしてサンプルセレクションバイアスが生じるかについて直観的に説明する。通常統計モデルを構築する際にモデル構築に用いる標本の抽出においては、「無作為抽出」と「作為抽出」がある。統計では標本は無作為に抽出されたと仮定する場合が多い。なぜなら無作為抽出法では、全体の一部を調べるだけで、母集団全体の情報がつかめるという利点があるからである。「作為抽出」だと、標本の抽出過程においてある種の作為が生じ、統計的な結果にバイアスが生じてしまう。これを通常サンプルセレクションバイアスという。例えば、格付予測モデルの構築を例にあげると、日本全国の上場企業を対象とした場合、格付が付与されている企業よりも付与されていない企業の方が数的には多い。この時、格付を取得している企業にある一定の傾向がある場合標本選択は単純な無作為抽出ではなくなり、標本選択においてバイアスがかかっている可能性がある。格付が付与されている企業に一定の傾向があるとのことは、格付を取得する企業はある程度信用力が高い企業であると思われるほか、実際に格付の取得を希望しても投資不適格と呼ばれるBBゾーン以下の格付しか取得できない可能性がある場合には、そもそも格付の公表を希望しないと思われることから、格付を取得している企業には何らかの傾向があると推察される<sup>1)</sup>。実際R&I格付を取得している企業に対して、格付別の分布を確認したところ、シ

シングル A ゾーンが全体の 47% 程度を占めていることが確認でき、R&I の格付を取得している企業は比較的に高い格付の企業が多いことが確認できる。一方で、日本全国の上場企業を対象とした場合、上場企業の格付の平均ゾーンがシングル A ゾーンとは考えられない<sup>2)</sup> ため、R&I の格付を取得している企業は日本全国の上場企業から無作為にランダムに抽出した標本であるとは考えにくいと思われ、R&I の格付を取得している企業にはそれなりに高い格付が付与される傾向があると考えられる。このように格付を取得している企業にある一定の傾向がある場合、格付を取得している企業のみでモデル構築を行うと、構築したモデルからの予測格付は、サンプルセレクションバイアスを含んだ格付となり、予測結果には偏りが生じ、実態よりも甘くまたは保守的に格付が予測されることになる。

本稿では格付予測モデルの構築におけるサンプルセレクションバイアスの有無を確認し、サンプルセレクションバイアスが存在する場合、推定格付はどれぐらいの偏りを持つかを確認するために、以下のような二つのモデルを構築した。

一つは、格付機関が公表している格付（ここでは R&I の公表格付を用いた）のみを用いて構築した順序ロジットモデルである。

もう一つは、日本全国の上場企業を対象とし、格付取得の有無と格付を取得した場合どの格付が付与されるかの二つの事象を同時にモデル化した二変量順序プロビットモデルである。二変量順序プロビットモデルにおいては、格付を取得しているか否かの事象と格付を取得した場合どの格付が付与されるかの事象に相関構造を仮定することでより現実に近い仮定の下でモ

デル構築を行うことが可能となると考え、上記の二つの事象に相関を仮定してモデルの推定を行った。もし推定される相関係数がゼロなら、二つの事象は独立であるとのことで、サンプルセレクションバイアスは存在せず、格付を取得している企業のみで順序ロジットモデルの構築を行っても問題ないと解釈できる。もし推定される相関係数がゼロではない場合、二つの事象には相関があるとのことになり、格付を取得していない企業も含んで構築した二変量順序プロビットモデルの方が手法としてより適切であり、R&I の格付を取得したとの条件の下で、どの格付が付与されたかを求めることにより、サンプルセレクションバイアスを調整した予測格付を推定することが可能となる。

実際上記の二つのモデルを推定し、モデルの予測格付を比較した結果、順序ロジットモデルは格付を取得している企業のみでデータでモデル構築を行っているため、R&I 格付と推定格付の一致率は、相関を仮定した二変量順序プロビットモデルの結果と比べて 2% 程度高い結果となったが、二変量順序プロビットモデルも十分なパフォーマンスを示した。さらに順序ロジットモデルの推定格付は R&I の格付との乖離が非常に大きい格付が数件あったが、二変量順序プロビットモデルの推定格付はそのような R&I の格付との乖離が大きい格付が比較的少なかった。また二変量順序プロビットモデルに仮定していた相関係数については 0.6 以上の高い相関が確認され、格付を取得している企業は高い格付が付与される確率が高くなる傾向があることが確認できた。さらに、順序ロジットモデルの推定格付と二変量順序プロビットモデルの推定格付を比較すると、順序ロジットモデルの推定格付が二変量順序プロビットモデルの推定格付に比べて高くなる傾向が確認できた。つまり、二つの事象（格付取得の有無と格付を取得した場合どの格付が付与されるか）に相関があるにも関わらず、相関を無視して格付を取得している企業のデータのみで順序ロジットモデルの構築を行うと、推定結果にはバイアスがかかり、

1) 現在は、金商法により、一旦格付の取得を希望した企業は格付結果の如何に関わらずその結果が公表されることになっている。

2) ㈱金融工学研究所の推計では、BBB ~ BB ゾーンが平均的なゾーンである。

サンプルセレクションバイアスを調整した二変量順序プロビットモデルよりも高い格付が付与されることが確認できた。

本稿での結果は、あくまで本稿で用いた5つの財務指標より構築されたモデルにおける結論であり、モデルの構築に用いるデータまたはモデルに用いる財務指標の組み合わせが変われば、異なる結果が導かれる可能性もなくはない。そのため、格付推計モデルの構築を行う際には、説明変数の組み合わせが決まった後に、順序ロジットモデルと二変量順序プロビットモデルの両方を構築し、相関の有無および構築したモデルの推定格付を比較することで、サンプルセレクションバイアスの有無を確認することが望ましいと考える。

本稿の構成は以下のとおりである。まず、2節では、一般的な格付予測モデルの構築に用いられることが多い順序ロジットモデルおよび本稿で主に扱う二変量順序プロビットモデルについてモデル式と尤度関数を示す。次に、3節では、分析に用いるデータ及び説明変数について解説する。続いて、4節では格付を取得している企業のみで構築した（サンプルセレクションバイアスを含んだ）順序ロジットモデルと格付を取得しているかしていないかの情報も含んだ日本全国の上場企業を対象として構築した二変量順序プロビットモデルを推定し、両モデル精度がどのぐらい異なるかを検証する。最後に、5節では本研究における主な分析結果をまとめる。

## 2. モデル

離散データの分析にはロジットモデルとプロビットモデルが最も広く用いられている。これらのモデルは目的変数を取る値によって、企業の倒産分析などに用いられる二項ロジット（プロビット）モデル（デフォルトしたら1、しなかったらゼロなどの二つの値をとる）、格付の分析などに用いられる順序ロジット（プロビット）モデル（通常格付は複数の記号に区分されており、格付には順序がついている、例えばAAAはAより

信用力が高いなど）などに区分される。

本稿で主に扱う格付のような順序付離散データの分析には順序ロジットモデル及び順序プロビットモデルが最も広く利用されている。両モデルの違いは仮定する誤差項の分布の違いであり、順序ロジットモデルは誤差項にロジスティック分布を仮定し、順序プロビットモデルは誤差項に正規分布を仮定している。仮定している誤差項の分布には違いがあるものの、モデルパラメータの推定値には大きな乖離がなく、どのモデルを用いても整合的な推定結果が得られる。実際金融分野の実務においては順序ロジットモデルの方が関数形がわかりやすく説明が容易であることと、モデルパラメータ推定も比較的容易であることから順序プロビットモデルよりも広く利用されている。そのため本稿でも順序ロジットモデルを用いてモデル構築を行う。順序ロジットモデルのモデル式及び尤度関数については、2.1節で解説する。

モデルの構築に用いられるデータは様々な形で制限されている。例えば本稿で扱う格付を例とすると、日本全国の上場企業を母集団とした場合、格付が観測される企業はその母集団の極一部である。このような場合母集団（日本全国の上場企業）と抽出した標本（格付が観測される企業）が同質であれば標本抽出のバイアスは存在しないかもしれないが、抽出した標本にある一定の傾向などがある場合標本抽出にバイアスがかかる可能性がある。本稿では標本抽出のバイアスの有無を確認するために、日本全国の上場企業を対象に、格付が観測されるかされないか、格付が観測される企業はどの格付が付与されているかの二つの事象を同時にモデル化する。この時目的変数は格付だけではなく、格付が観測されるかされないかも目的変数となり、目的変数として二つが仮定される。そのためモデルとして二変量順序ロジットモデルと二変量順序プロビットモデルが候補となる。本稿ではさらに格付が観測されるかされないか、格付が観測される企業はどの格付が付与されているかの二つの事象に相関があるか否かを確認すべ





ここで、 $P_{i,jk}$  の計算は以下のとおり。

① R&I の格付を取得していない企業  

$$P(Y_{i1} = 0) = \Phi(\tau - \alpha_1 X_{i1} - \dots - \alpha_m X_{im})$$

② R&I の格付 1 を取得している企業  

$$P(Y_{i1} = 1, Y_{i2} = 1) = \Phi(\mu_1 - \beta_1 R_{i1} - \dots - \beta_r R_{ir})$$

$$- N(\mu_1 - \beta_1 R_{i1} - \dots - \beta_r R_{ir}, \tau - \alpha_1 X_{i1} - \dots - \alpha_m X_{im}, r)$$

③ R&I の格付  $j$  ( $j=2 \sim J-1$ ) を取得している企業  

$$P(Y_{i1} = 1, Y_{i2} = j) = \Phi(\mu_j - \beta_1 R_{i1} - \dots - \beta_r R_{ir})$$

$$- \Phi(\mu_{j-1} - \beta_1 R_{i1} - \dots - \beta_r R_{ir})$$

$$- N(\mu_j - \beta_1 R_{i1} - \dots - \beta_r R_{ir}, \tau - \alpha_1 X_{i1} - \dots - \alpha_m X_{im}, r)$$

$$+ N(\mu_{j-1} - \beta_1 R_{i1} - \dots - \beta_r R_{ir}, \tau - \alpha_1 X_{i1} - \dots - \alpha_m X_{im}, r)$$

④ R&I の格付  $J$  を取得している企業  

$$P(Y_{i1} = 1, Y_{i2} = J) = 1 - \Phi(\mu_{J-1} - \beta_1 R_{i1} - \dots - \beta_r R_{ir})$$

$$- \Phi(\tau - \alpha_1 X_{i1} - \dots - \alpha_m X_{im})$$

$$+ N(\mu_{J-1} - \beta_1 R_{i1} - \dots - \beta_r R_{ir}, \tau - \alpha_1 X_{i1} - \dots - \alpha_m X_{im}, r)$$

ここで、 $\Phi(x)$  は標準正規分布の累積分布関数で、 $N(x, y, r)$  は、相関を持つ二変量標準正規分布の累積分布関数である。上記の  $P_{i,jk}$  をすべての企業  $n$  について足し合わせることでこのモデルにおける対数尤度関数が求まる。本稿では、この尤度関数を用いてモデルのパラメータの最尤推定量を求める。

モデルから算出される予測格付について、考え方として無条件の期待値つまり、式 (5) の  $Z_{i2} = \beta_1 R_{i1} + \dots + \beta_r R_{ir}$  をもとに算出する方法と、R&I の格付を取得している条件の下で、どの格付が付与されたかという、条件付き期待値である  $E[y_2 | y_1 \text{ は観測される}]$  をもとに予測格付を算出する二通りの考え方がある。

本稿で扱う目的変数のように、観測される目的変数がある母集団からの一部の標本であり、さらにある一定の傾向を持っている場合、母集団からの標本抽出にバイアスが生じる可能性がある。この場合は、標本抽出のバイアスによる偏

りを調整した条件付き期待値を用いるのが適切である。これの理論根拠については Greene (2003, pp.781) をもとに、以下で詳細を解説する。

観測される目的変数に切断がある場合の二変量正規分布の積率については以下の定理がある。その前に、切断というのは、標本データが研究対象のより大きい分布のある部分のみから抽出されたとき生じるものである。本稿の例では、モデル構築に用いる標本データである R&I の格付を取得している企業は、より大きい母集団である日本上場企業全体のデータの一部から抽出される標本であり、切断が生じている可能性があることになる。

ここでは Greene (2003, pp.781) の定理 22.5 をそのまま掲載する。Greene (2003, pp.781) の定理 22.5 によると、観測される目的変数に切断がある場合の二変量正規分布の積率は以下のとおりとなる。

定理 22.5 :  $y$  と  $z$  が平均が  $\mu_y$  と  $\mu_z$ 、標準偏差が  $\sigma_y$  と  $\sigma_z$ 、相関が  $\rho$  の二変量正規分布に従うとすると、一次および二次の条件付き積率は以下のように計算される。

$$E[y | z > a] = \mu_y + \rho \sigma_y \lambda(\alpha_z),$$

$$\text{Var}[y | z > a] = \sigma_y^2 [1 - \rho^2 \delta(\alpha_z)].$$

ここで、

$$\alpha_z = \frac{(a - \mu_z)}{\sigma_z}, \lambda(\alpha_z) = \frac{\phi(\alpha_z)}{[1 - \Phi(\alpha_z)]}, \delta_z = \lambda(\alpha_z) [\lambda(\alpha_z) - \alpha_z]$$

である。また  $\phi(a_z)$  は標準正規分布の確率密度関数、 $\Phi(a_z)$  は標準正規分布の累積分布関数を表し、 $\lambda(a_z)$  は逆ミルズ比と呼ばれるもので、詳細は、Heckman (1979) を参照。

上記の定理 22.5 を本稿における二変量順序プロビットモデルに当てはめると、以下のようになる。

まず、標本選択を決定する式、つまり R&I の格付を取得しているか否かのモデル式は、式 (3) と式 (4) のとおりである。ここで、R&I

の格付を取得している企業が選択された標本となる。次に、主要な研究対象となる格付予測モデル式は、式 (5) と式 (6) のとおりである。R&I の格付を取得しているか否かの規則は、式 (3) の  $z_{i1}$  がゼロ ( $\tau = 0$ ) よりも大きいときのみ観測される。式 (3) と式 (5) の誤差項は平均ゼロ、分散 1 の標準正規分布に従い、誤差項の間の相関は  $r$  であると仮定する。これらを Greene (2003, pp.781) の定理 22.5 に代入すると、R&I の格付を取得している条件の下で、どの格付を取得したかのモデルの予測格付を表す条件付き期待値は、以下のとおりである。

$$\begin{aligned} E[y_2 | y_1 \text{は観測される}] &= E[y_2 | z_1 > 0] \\ &= E[y_2 | e_1 > -\alpha_1 X_1 - \dots - \alpha_m X_m] \\ &= \alpha_1 X_1 + \dots + \alpha_m X_m + E[e_2 | e_1 > -\alpha_1 X_1 - \dots - \alpha_m X_m] \\ &= \alpha_1 X_1 + \dots + \alpha_m X_m + r\lambda(\alpha_{e_1}) \end{aligned}$$

ここで、 $\alpha_{e_1} = -\alpha_1 X_1 - \dots - \alpha_m X_m$  と

$$\lambda(\alpha_{e_1}) = \frac{\phi(-\alpha_1 X_1 - \dots - \alpha_m X_m)}{\Phi(-\alpha_1 X_1 - \dots - \alpha_m X_m)}$$

である。

$r\lambda(\alpha_{e_1})$  はサンプルセレクションバイアスの調整項であり、相関係数がゼロの場合サンプルセレクションバイアスが存在しないためこの項はゼロとなり、予測格付は式 (5) の  $Z_{i2} = \beta_1 R_{i1} + \dots + \beta_r R_{im}$  をもとに算出する無条件の予測格付と一致する。また相関係数がゼロではない場合、 $r\lambda(\alpha_{e_1})$  の分だけサンプルセレクションバイアスが存在するとのことになる。

### 3. 使用データ

本節では、2004 年から 2009 年までの上場企業を対象とし、R&I の格付を取得している企業及び取得していない企業を区別した上で、2.1 節と 2.2 節の順序ロジットモデル及び二変量順序プロビットモデルを構築し、両モデルの精度及び推定格付がどの程度異なるかを分析する。

#### 3.1. データ

モデル構築に当たって、財務データとし

表 1. 分析基準日と対象決算書

分析基準日	対象決算書	R&I 格付
2004 年 9 月	2003 年 6 月～2004 年 5 月	2004 年 9 月 30 日
2005 年 9 月	2004 年 6 月～2005 年 5 月	2005 年 9 月 30 日
2006 年 9 月	2005 年 6 月～2006 年 5 月	2006 年 9 月 30 日
2007 年 9 月	2006 年 6 月～2007 年 5 月	2007 年 9 月 30 日
2008 年 9 月	2007 年 6 月～2008 年 5 月	2008 年 9 月 30 日
2009 年 9 月	2008 年 6 月～2009 年 5 月	2009 年 9 月 30 日

て 2004 年度から 2009 年度までの上場企業の財務データを対象とした。財務データは日経 NEEDS のデータベースから取得した。また格付データとして 2004 年度から 2009 年度までの R&I の格付を取得している企業を対象とした。

モデル構築に用いるデータを作成するためには、財務データと格付データを紐付ける必要があるが、紐付ルールは以下のとおりとした。まず、分析基準日として 2004 年 9 月から 2009 年 9 月を設け、次に、分析基準日から 4 ヶ月の決算書の未入手期間を置いてそこから対象期間内 (1 年間) にある直近の決算書を取ってきた。最後に、上記で確定した決算書に 9 月 30 日の R&I の格付を紐付けた。具体的な分析基準日と対象となる決算書及び R&I 格付との対応関係は表 1 のとおりである。

モデル構築に当たっては、業種別に企業の特徴が異なることより、業種別のモデルを構築するのが一般的なやり方である。本稿では、いくつかの業種の中から標本数が最も多い製造業のデータを用いてモデル構築を試みる。

製造業における R&I の格付を取得している企業数は 1,648 件であり、取得していない企業数は 8,427 件である。うち、R&I の格付を取得している企業について、格付別企業の割合を示したのが表 2 である。

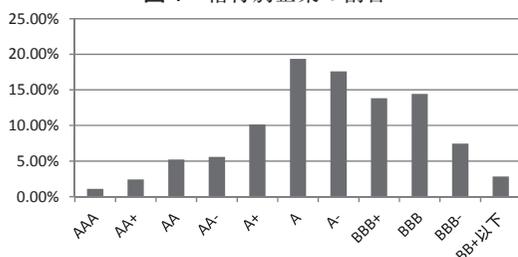
表 2 および図 1 の R&I の格付を取得している企業の格付別分布をみると、シングル A ゾーンを中心 (48% 弱) に、トリプル B ゾーンの格付が概ね 36% 程度を占める分布となっていることが確認できる。

しかし、日本の製造業の上場企業全体を母集団と考えた場合、その上場企業全体の格付の中

表2. 格付別企業数

格付	企業数	割合
AAA	18	1.092%
AA+	40	2.427%
AA	86	5.218%
AA-	92	5.583%
A+	167	10.133%
A	319	19.357%
A-	290	17.597%
BBB+	228	13.835%
BBB	238	14.442%
BBB-	123	7.464%
BB+ 以下	47	2.852%
計	1,648	100.00%

図1 格付別企業の割合



心がシングル A ゾーンとは直観的に考えにくい。そのため、R&Iの格付を取得している企業は、日本の製造業の上場企業全体からランダムに抽出した標本ではなく、ある一定の傾向を持つ標本であり、サンプルセレクションバイアスを含む可能性があることもある程度推察される。

### 3.2. モデルの説明変数

格付を予測するモデルの構築において説明変数をどのように選択するかは格付分析の重要な研究課題であり、一般的にはより広範囲の財務指標から格付を分析するために、規模・安全性・収益性・効率性などを表す財務指標が用いられている。

本稿ではこの問題に深く立ち入ることは避け、先行研究で用いられているいくつかの指標を参考として、信用力のモデル構築に最も多く用いられる代表的な5つの財務指標を用いて分析を進めることとする。具体的には、規模指標

として自己資本、安全性指標として自己資本比率、収益性指標として総資本事業利益率 (ROA)、効率性指標として営業運転資本回転期間、その他指標として経常収支総資産比率の5指標を用いてモデル構築を行う。

モデルの説明変数である財務指標には指標値が大きいほど信用力が高いと解釈できる指標と指標値が小さいほど信用力が高いと解釈できる指標がある。今般モデル構築に用いる5指標のうち、効率性指標以外の指標は、指標値が大きいほど信用力が高いと考えられる指標であり、効率性指標である営業運転資本回転期間は指標値が小さいほど信用力が高いと考えられる指標である。本稿ではモデル構築における説明変数の符号条件<sup>3)</sup>をよりわかりやすくするために、効率性指標である営業運転資本回転期間については、逆数を取ることで指標値が大きいほど信用力が高く格付も高くなると解釈できる指標に変換し、他の指標との向き(意味合い)を合わせた。

次頁以降では具体的な説明変数の算式を示す。なお説明変数の加工については詳細は割愛するが、外れ値処理や対数変換などを行っている。

表4に上記の説明変数に関する統計量を示す。表の各種統計量はすべての指標の対数変換後の値に対して算出している。

また説明変数間の相関係数は表5のとおり。表5より各変数間には極端に高い相関はなく、総資本事業利益率 (ROA) と経常収支総資産比率に0.52程度の相関が確認された。

次節では、上記の5つの説明変数を用いて、順序ロジットモデル及び二変量順序プロビットモデルのモデルの構築を行う。

3) 指標の意味合いが異なる指標、つまり指標値が大きいほど信用力が高いと解釈される指標と指標値が小さいほど信用力が高いと解釈される指標ではパラメータの符号が異なる。本稿では指標値が大きいほど信用力が高いと解釈できる指標のパラメータにはマイナスの符号制約を設けた。

表 3. 説明変数の算出

No.	区分	指標名	算式
1	規模	自己資本	資本合計
2	安全性	自己資本比率	資本合計 / 資産計
3	収益性	総資本事業利益率 (ROA)	(営業損益 + 受取利息・配当金) / 総資産
4	効率性	営業運転資本回転期間逆数	1 / ((受取手形 + 売掛金 + 棚卸資産 - 支払手形 - 買掛金 + 割引手形) / (売上高 / 決算月数))
5	その他	経常収支総資産比率	(経常収入 - 経常支出) / 総資産

表 4. 説明変数の統計量

No.	区分	指標名	平均	標準偏差	中央値	最小値	最大値
1	規模	自己資本	2.94	1.31	2.82	0.38	5.22
2	安全性	自己資本比率	1.71	0.38	1.77	0.57	2.20
3	収益性	総資本事業利益率 (ROA)	1.47	1.19	1.74	-1.82	2.96
4	効率性	営業運転資本回転期間逆数	1.72	1.17	1.52	0.91	14.08
5	その他	経常収支総資産比率	0.54	0.37	0.58	-0.48	1.11

表 5. 説明変数間の相関係数

指標名	自己資本	自己資本比率	売上高利益率	営業運転資本回転期間逆数	経常収支総資産比率
自己資本	1				
自己資本比率	0.23	1			
総資本事業利益率 (ROA)	0.27	0.39	1		
営業運転資本回転期間逆数	0.06	-0.17	-0.13	1	
経常収支総資産比率	0.22	0.19	0.52	0.22	1

#### 4. 分析結果

本節では、3節のデータ及び説明変数を用いて、2節の順序ロジットモデルと二変量順序プロビットモデルの推定を行い、両モデルの推定結果について比較分析を行う。

##### 4.1. 順序ロジットモデルの推定

ここでは、3.1節のデータと3.2節の説明変

表 6. 順序ロジットモデルのパラメータ推定値

No.	区分	指標名	推定値
1	規模	自己資本	-2.90
2	安全性	自己資本比率	-3.12
3	収益性	総資本事業利益率 (ROA)	-0.22
4	効率性	営業運転資本回転期間逆数	-0.25
5	その他	経常収支総資産比率	-1.31

表 7. 格付閾値

格付	順序ロジットモデル閾値
aaa	-26.66
aa+	-25.39
aa	-24.28
aa-	-23.59
a+	-22.64
a	-21.11
a-	-19.64
bbb+	-18.19
bbb	-16.08
bbb-	-14.01
bb+ 以下	-14.01 超

数である5つの財務指標を用いて、順序ロジットモデルのパラメータ推定を行う。

モデルの目的変数であるR&I格付については、投資適格の格付はそのまま用い、投機適格であるBB+以下の格付は一つの格付にまとめて推定を行う。

順序ロジットモデルのパラメータ推定値は表6のとおり。

格付と説明変数の符号関係をみると、今般モデル構築においてすべての財務指標に対して、指標値が大きいほど信用力が高いと解釈されるように指標変換を行っているため、推定されるパラメータの符号はすべて「-」となっている。また目的変数である格付において、「aaa」は「1」、「aa+」は「2」、…、のように信用力が高いほど数値が小さくなるように設定した。そのため、2.1節の式(1)における $Z_i$ の値が小さいほど高い格付が付与されることになる。

予測格付を算出するに当たっての格付閾値は表7のとおりである。

表 8. 二変量順序プロビットモデルのパラメータ推定値

No.	区分	指標名	格付有無 モデル推 定値	格付モデ ル推定値	相関係数
1	規模	自己資本	34.04	-136.92	-0.61
2	安全性	自己資本比率	-	-158.32	
3	収益性	総資本事業利 益率 (ROA)	-	-	
4	効率性	営業運転資本 回転期間逆数	-	-7.88	
5	その他	経常収支総資 産比率	-	-61.31	

## 4.2. 二変量順序プロビットモデルの推定

ここでは、3.1 節の製造業の全上場企業のデータを用いて、R&I の格付を取得しているか否かを判別するモデルと R&I の格付を取得している企業はどの格付が付与されるかの二つのモデルを同時に推定する。二変量順序プロビットモデルの推定においては、誤差項に相関を仮定して推定を実施する。モデルの説明変数としては、順序ロジットモデルと同様に 3.2 節の 5 変数を用いる。

R&I の格付を取得しているか否かを判別するモデルの説明変数は、すべての組み合わせによる推計結果の中から説明変数が有意かつ R&I 格付と推定格付の一致率が最も高い組み合わせを選択した。なお、R&I の格付を取得している企業はどの格付が付与されるかのモデルについては、3.2 節の 5 変数すべてを用いた。

最終的なモデルの説明変数及び推定値は表 8 のとおりである。

格付とモデルの説明変数の関係であるが、まず、R&I の格付を取得しているか否かを判別するモデルにおいては、規模を表す自己資本がモデルに採用された。解釈としては規模が大きい企業ほど積極的に格付を取る傾向があることになる。つまり 2.2 節の式 (3) の  $Z_{i1}$  の値が大きいほど格付を取得する傾向があるとの解釈になる。次に、R&I の格付を取得している企業はどの格付が付与されるかのモデルについては順序ロジットモデルの場合と同様にすべての説明変数のパラメータ推定値は「-」となった。つま

表 9. 偏り調整前後の Z 値【全体】

	偏り調整前	偏り	偏り調整後
平均	-7.20	-0.32	-7.52
標準偏差	2.11	0.14	2.06

表 10. 偏り調整前後の Z 値【R&amp;I 格付を取得している企業】

	偏り調整前	偏り	偏り調整後
平均	-9.54	-0.29	-9.83
標準偏差	1.16	0.10	1.13

り 2.2 節の式 (5) の  $Z_{i2}$  の値が小さいほど高い格付が付与されることになる。採用変数をみると収益性の売上高利払後事業利益率が欠落しているが、これは経常収支総資産比率との相関が原因であると考えられる。最後に、 $Z_{i1}$  と  $Z_{i2}$  の間に -0.61 の相関があるとのことは、 $Z_{i1}$  の値が大きい企業は  $Z_{i2}$  の値が小さくなる傾向にあり、結果として R&I の格付を取得している企業は比較的に高い格付が付与される傾向があるとの解釈になる。

R&I の格付を取得しているとの条件の下で、どの格付を取得したかは、2.2 節の定理 22.5 により、「無条件の期待値」に、サンプルセレクションバイアスによる偏りである「相関係数×逆ミルズ比」を足した条件付き期待値となる。以下では、偏りを調整する前と調整した後の結果を比較することで、どの程度の偏りがあったかを確認する。

表 9 と表 10 における「偏り調整前」はサンプルセレクションバイアスを考慮しない無条件の期待値である式 (5) の「 $Z_{i2} = \beta_1 R_{i1} + \dots + \beta_r R_{in}$ 」を示し、「偏り」はサンプルセレクションバイアスによる偏りを示すもので、「相関係数×逆ミルズ比」で計算され、「偏り調整後」は R&I の格付を取得しているとの条件の下で、どの格付を取得したかの条件付き期待値を示すもので、「 $Z_{i2} = \beta_1 R_{i1} + \dots + \beta_r R_{in} + \text{相関係数} \times \text{逆ミルズ比}$ 」で計算される。

サンプルセレクションバイアスによる偏りは、「相関係数×逆ミルズ比」で計算されるが、逆ミルズ比は常に正の値をとり、相関係数は本稿では -0.61 だったため、偏りは負の値をとる。

表 11. 格付閾値

格付	二変量順序プロビットモデル閾値
aaa	-13.25
aa+	-12.72
aa	-12.22
aa-	-11.88
a+	-11.41
a	-10.66
a-	-9.94
bbb+	-9.25
bbb	-8.28
bbb-	-7.43
bb+ 以下	-7.43 超

つまり、偏りを調整することで、偏り調整後の  $Z_{i2}$  の平均は偏り調整前平均である  $Z_{i1}$  より小さくなり、標準偏差も小さくなっていることが確認できた。

予測格付を算出するに当たっての格付閾値は表 11 のとおりである。

#### 4.3. モデル推定結果の比較

ここでは、4.1 節と 4.2 節にて構築した順序ロジットモデルの推定格付と二変量順序プロビットモデルの推定格付を用いて、モデルの目的変数である R&I 格付との比較を実施する。

R&I の格付を取得している 1,648 件について、R&I の格付と順序ロジットモデルの推定格付、二変量順序プロビットモデルの推定格付の推移を示したのが表 12 と表 13 である。表中の縦の格付は R&I の格付を表し、横の格付は推定モデルの格付を表す。なお、表現をわかりやすくするために、R&I の格付は大文字（例えば AAA）、順序ロジットモデルの推定格付は小文字（例えば aaa）、二変量順序プロビットモデルの推定格付は小文字太字（例えば **aaa**）で表す。

表 12 より、R&I の格付を取得している企業について、R&I の格付と順序ロジットモデルの推定格付を比較した結果、全体でみた場合 34.10% が一致しており、78.82% が 1 ノッチ以内に、91.44% が 2 ノッチ以内に収まっていることが確認できる。

格付別でみた場合、上位格と最下位格ではサ

表 12. 格付推移：R&amp;I 格付→モデル推定格付（順序ロジットモデル）

	aaa	aa+	aa	aa-	a+	a	a-	bbb+	bbb	bbb-	bb+ 以下	計
AAA	0	0	2	5	4	6	1	0	0	0	0	18
AA+	0	0	5	11	13	11	0	0	0	0	0	40
AA	0	0	3	19	16	44	4	0	0	0	0	86
AA-	0	0	4	21	19	36	11	1	0	0	0	92
A+	0	0	1	32	44	73	16	1	0	0	0	167
A	0	0	1	10	61	150	92	5	0	0	0	319
A-	<i>I</i>	<i>2</i>	0	2	10	106	100	62	7	0	0	290
BBB+	0	0	0	0	0	27	82	74	39	5	1	228
BBB	<i>I</i>	0	0	0	0	5	24	57	123	22	6	238
BBB-	0	0	0	0	0	0	10	17	41	42	13	123
BB+ 以下	0	0	0	0	0	1	5	11	15	10	5	47
計	2	2	16	100	167	459	345	228	225	79	25	1,648

2 ノッチ上げ	104	6.31%
1 ノッチ上げ	393	23.85%
一致	562	34.10%
1 ノッチ下げ	344	20.87%
2 ノッチ下げ	104	6.31%

表 13. 格付推移：R&amp;I 格付→モデル推定格付（二変量順序プロビットモデル）

	aaa	aa+	aa	aa-	a+	a	a-	bbb+	bbb	bbb-	bb+ 以下	計
AAA	0	0	0	0	5	7	6	0	0	0	0	18
AA+	0	0	0	0	8	29	3	0	0	0	0	40
AA	0	0	0	1	11	50	23	1	0	0	0	86
AA-	0	0	0	0	10	45	32	5	0	0	0	92
A+	0	0	0	0	16	87	58	5	1	0	0	167
A	0	0	0	1	5	116	144	50	3	0	0	319
A-	0	0	0	0	2	32	124	95	37	0	0	290
BBB+	0	0	0	0	0	6	46	90	72	11	3	228
BBB	0	0	0	0	0	0	16	31	126	54	11	238
BBB-	0	0	0	0	0	0	2	17	29	43	32	123
BB+ 以下	0	0	0	0	0	0	2	11	15	11	8	47
計	0	0	0	2	57	372	456	305	283	119	54	1,648

2 ノッチ上げ	57	3.46%
1 ノッチ上げ	154	9.34%
一致	523	31.74%
1 ノッチ下げ	495	30.04%
2 ノッチ下げ	223	13.53%

ンプル数が少ないため推定結果にばらつきが存在することも確認できる。特に R&I 格付が「A-」と「BBB」の一部のサンプルにおいては、推定格付がかなり高いことが確認できる（表中の斜体太字部分）。また 1 ノッチ以内に収まって

いる格付の中で、1ノッチ上げが23.85%、1ノッチ下げが20.87%、2ノッチ以内に収まっている格付の中で、2ノッチ上げが6.31%、2ノッチ下げが6.31%を占めており、順序ロジットモデルからの推定格付はノッチ上げの割合が比較的多く、R&Iの格付より高めに推定される格付が比較的多いことが確認できる。

目的変数であるR&Iの格付は財務情報以外にも各種定性情報等によって付与されている。本稿では限られた5つの財務指標のみでR&Iの格付を当てているため、実際のR&I格付とモデル推定格付の一致率はさほど高くない。

表13より、R&Iの格付を取得している企業について、R&Iの格付と二変量順序プロビットモデルの推定格付を比較した結果、全体でみた場合31.74%が一致しており、71.12%が1ノッチ以内に、88.11%が2ノッチ以内に収まっていることが確認できる。また1ノッチ以内に収まっている格付の中で、1ノッチ上げが9.34%、1ノッチ下げが30.04%、2ノッチ以内に収まっている格付の中で、2ノッチ上げが3.46%、2ノッチ下げが13.53%を占めており、二変量順序プロビットモデルからの推定格付はノッチ下げの割合が比較的多く、R&Iの格付より保守的に推定される格付が多いことが確認できる。

二変量順序プロビットモデルから算出される予測格付は、順序ロジットモデルの場合に比べて、R&I格付との一致率が2%程度低いですが、その結果については以下のとおり解釈できる。順序ロジットモデルはR&Iの格付を取得している企業のみで構築したモデルであり、一致率が高いのは当然のことかもしれない。一方で、二変量順序プロビットモデルは、二つのモデルを同時に推定しており、R&I格付を取得しているか否かもモデル化しているため、このモデルにおいても誤差が存在し、その分R&I格付との一致率が若干低くなっていると考えられる。それにしても二変量順序プロビットモデルの一致率も順序ロジットモデルと比べて大きな差異はなく、相応のパフォーマンスであると考えられる。

また二変量順序プロビットモデルからの推定

表14. 格付推移：順序ロジットモデルの推定格付  
→二変量順序プロビットモデルの推定格付

	aaa	aa+	aa	aa-	a+	a	a-	bbb+	bbb	bbb-	bb+以下	計
aaa	0	0	0	0	0	0	1	0	1	0	0	2
aa+	0	0	0	0	2	0	0	0	0	0	0	2
aa	0	0	0	1	15	0	0	0	0	0	0	16
aa-	0	0	0	0	35	65	0	0	0	0	0	100
a+	0	0	0	1	4	162	0	0	0	0	0	167
a	0	0	0	0	1	141	317	0	0	0	0	459
a-	0	0	0	0	0	4	134	207	0	0	0	345
bbb+	0	0	0	0	0	0	4	92	132	0	0	228
bbb	0	0	0	0	0	0	0	6	149	70	0	225
bbb-	0	0	0	0	0	0	0	0	1	48	30	79
bb+以下	0	0	0	0	0	0	0	0	0	1	24	25
計	0	0	0	2	57	372	456	305	283	119	54	1,648

2ノッチ上げ	0	0.00%
1ノッチ上げ	18	1.09%
一致	592	35.92%
1ノッチ下げ	954	57.89%
2ノッチ下げ	80	4.85%

格付は順序ロジットモデルの推定格付のようにR&Iの格付との乖離が大きい格付はなかったことも特徴の一つである。サンプルセレクションバイアスを調整することにより、R&Iの格付により近い推定格付が得られたと考えられる。

最後に、順序ロジットモデルの推定格付と二変量順序プロビットモデルの推定格付について比較する。表14の縦の格付は順序ロジットモデルの推定格付、横の格付は二変量順序プロビットモデルの推定格付を表す。

表14より、R&Iの格付を取得している企業について、順序ロジットモデルの推定格付と二変量順序プロビットモデルの推定格付を比較した結果、全体でみた場合35.92%が一致しており、94.90%が1ノッチ以内に、99.76%が2ノッチ以内に収まっている。また1ノッチ以内に収まっている格付の中で、1ノッチ上げが1.09%、1ノッチ下げが57.89%、2ノッチ以内に収まっている格付の中で、2ノッチ上げが0.00%、2ノッチ下げが4.85%を占めている。つまり、順序ロジットモデルの推定格付に比べて二変量順序プロビットモデルの推定格付はノッチ下げの割合が大きいことより、二変量順序プロビットモデ

ルは順序ロジットモデルに比べて予測格付が保守的に算出されることが確認できる。

## 5. 結論

前節までは R&I の格付を予測するモデルを構築するにあたり、上場企業の製造業のデータを用いて二つのモデルを構築した。一つは、R&I の格付を取得している企業のみで構築した順序ロジットモデルである。もう一つは、製造業の上場企業全体のデータを用いて、R&I の格付を取得しているか否かのモデルと R&I の格付を取得している企業はどの格付が付与されているかのモデルを同時に推定する相関を仮定した二変量順序プロビットモデルである。二変量順序プロビットモデルの相関係数の推定値からサンプルセレクションバイアスの有無を確認し、サンプルセレクションバイアスがある場合、二つのモデルから算出される予測格付にどのような傾向があるかについて確認した。結論をまとめると以下のとおりである。

通常モデルは、どのようなサンプルデータを用いて構築されたかによって出力結果が変わってくる。特に、モデル構築に用いるデータがある大きな母集団から抽出された一部のサンプルである時は、サンプルセレクションバイアスの存在有無を確認する必要がある。本稿で扱う目的変数である R&I の格付を取得している企業は、日本全上場企業のうちの一部の標本であることから、モデルの構築に用いるデータのサンプルセレクションバイアスの有無を確認するのが望ましい。そこでサンプルセレクションバイアスの有無を確認する手法として、R&I 格付を取得しているか否かの事象と R&I 格付を取得している場合にはどの格付が付与されているかの二つの事象に相関を仮定し、相関を持つ二変量順序プロビットモデルを構築し、推定される相関係数よりサンプルセレクションバイアスの有無を確認した。推定される相関係数がゼロであれば、上記の二つの事象は独立であるとのことで、サンプルセレクションバイアスは存在せず、R&I 格付を取得している企業のみでモデ

ル構築を行うことに問題はない。一方で、相関係数がゼロではない場合は、上記の二つの事象には関連があるとの解釈になり、その場合は上記の二つの事象をそれぞれモデル化し、二変量順序プロビットモデルを構築してサンプルセレクションバイアスによる偏りを調整した予測格付を算出するのが手法として適切であると考えられる。

モデルの推定結果より、二変量順序プロビットモデルの相関係数が 0.6 以上となり、R&I の格付を取得している企業は比較的高い格付が付与される確率が高いことが確認できた。言い換えると、R&I の格付を取得している企業のみで順序ロジットモデルを構築すると、推定される格付は実態より高めに予測される傾向があるとのこととなる。ただし、本稿の結論は、あくまで限られた 5 つの財務指標による分析結果であり、モデルの推定に用いる標本が変わるまたは説明変数の組み合わせが変わる場合、結論が異なる可能性もある。そのため、モデル構築を行うにあたっては、モデルに使用する説明変数候補が決まったら、順序ロジットモデルと相関を仮定した二変量順序プロビットモデルを構築して、サンプルセレクションバイアスの有無を確認することが適切であると考えられる。

## 参考文献

- 格付投資情報センター編 (1998). 『格付の知識』日経文庫
- 木島正明, 小守林克哉 (2006). 『信用リスク評価の数理モデル』朝倉書店
- 黒沢義孝 (2007). 『格付会社の研究』東洋経済新報社
- 黒沢義孝 (2009). 『格付情報のパフォーマンス評価』梓出版社
- 炭谷健志 (2008). 「銀行のリスク管理と格付け: 事業法人等への与信を中心に」『年報財務管理研究』19. pp.1-7. 日本財務管理学会
- 成松恭多 (2008). 「SME 格付けの概要と日本の中堅・中小企業市場における格付けの役割」『年報財務管理研究』19. pp.8-14. 日本財務管理学会

- 安川武彦 (2003). 「順序離散データモデルの拡張による格付けデータの分析」, 筑波大学ビジネス科学研究科, 2003, 博士論文
- 山下智志, 川口昇 (2003). 「大規模データベースを用いた信用リスク計測の問題点と対策 (変数選択とデータ量の関係)」『金融庁金融研究研修センター』, ディスカッションペーパー
- 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. pp.23-38.
- Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, Vol.23, No.4, pp.189-209.
- Banasik, J., Crook, J., Thomas, L. (2003), Sample selection bias in credit scoring models, *Journal of the Operational Research Society*, Vol.54, No.8, pp.822-832.
- Greene, W.H. (2003). *Econometric analysis*, 5th edition. Upper Saddle River, NJ: Prentice Hall.
- Greene, W.H. (1998). Sample selection in credit-scoring models. *Japan and the World Economy*, Vol.10, pp.299-316.
- Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, Vol.47 No.1, pp.153-161.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion), *J. Roy. Statist. Soc. Ser. B*,42, pp.109-142.
- Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance*, Vol.29, No.2, pp.449-470.

(株式会社金融工学研究所)  
(査読付投稿論文 2014年10月受理)