

銀行口座の入出金情報に基づく個人の行動特性を考慮した カードローンの途上管理に関するモデル構築

上武 治紀* 枇々木 規雄†

概要 一般的にカードローンは、法人向け貸出と比較して金額が小口であり管理コストを抑えるため、カードローン契約後は延滞の発生や極度額の増額の申し込みなどがない限り特段のアフタフォロー等は実施していないことが多い。実際、2019年9月に金融庁 [1] より発表された「銀行カードローンのフォローアップ調査結果」によれば、2017年3月の全銀協による「銀行による消費者向け貸付けに係る申し合わせ」[11] の効果もあり融資審査態勢の見直し等の業務運営の改善に向けた取組みは進んでいるとの評価を受けた一方、融資実行後の途上管理については、能動的に顧客の変化やその予兆を把握しようとする動きが鈍く、取組みが不十分であるとの評価を受けており、途上管理態勢の構築を速やかに進める必要がある。年収証明書の定期的な取得などは、カードローン契約後に顧客の変化やその予兆を把握するのに効果があると思われるが、現実的には実施は困難である。したがって、その他の方法で実際の借入・返済に関する行動特性を把握することができれば一定の効果が見られると考えられる。そこで、本研究では個人の借入・返済に関する行動特性を捉えるための一つのアプローチとして、カードローン利用者が借入・返済に利用している銀行口座の入出金データを分析する。入出金データを利用してカードローンのデフォルトを分析およびモデル化している唯一の先行研究である上武他 [3] はカードローンの審査を目的としたモデルを構築しており、カードローン利用者の途上管理を目的として、借入・返済に関する個人の行動特性を考慮してカードローンのデフォルトを分析・評価する研究は筆者らが知る限り存在しない。モデルを構築するために、審査モデルでも利用した説明変数のほかに借入・返済行動に関する借入回数や約定返済率などの変数を入出金データから生成し、デフォルトとの関連を分析する。そして、これらを説明変数候補として、デフォルトを評価するロジットモデルの構築を試みる。モデルの有用性を検証するために、2013年4月～2018年12月の約6000万件の入出金データを用いてモデルを構築する。ある基準日から過去18カ月間の入出金データで、基準日以降の12カ月以内にデフォルトするか否かを評価するモデルに対して、AR値を用いてモデルの序列性能の精度を評価したところ、AR値が50%を超える水準となり、一定の精度を確保することができた。

キーワード: 信用リスク, データ分析, 入出金情報

1. はじめに

銀行カードローンは貸金業法改正による総量規制の対象外であることなどを受けて2017年にかけて増加していたが、全国銀行協会による各会員銀行に対する申し合わせ [11] 以降、銀行カードローン残高は減少に転じている。また、2020年以降は新型コロナウイルスによる消費活動の低迷により残高は更に減少している。銀行が今後カードローン残高を増加させるには全国銀行協会による申し合わせの趣旨を踏まえ、顧客の利便性と保護の両面に十分配慮し、審査態勢等を構築する必要がある。

一般的に、カードローンの審査は、個人の収入や家族状況、勤務状況などの借入希望者のカードローン申込時の属性情報によるスコアリングモデルによって行われているが、上武他 [3] は、銀行口座の入出金データを利用して個人の行動特性を考慮したカードローンのデ

* 株式会社りそな銀行リスク統括部 本論文の内容は著者の所属組織の公式見解を示すものではなく、全て個人の見解である。

† 慶應義塾大学理工学部管理工学科

フォルトを評価するモデルを構築し、カードローンの審査に実用可能な水準の精度が確保できたことを示している。この個人の行動特性に着目したものとして、増井 [4] はソーシャルメディア上の情報から利用者の生活状況、性格、交友関係を読み取り借入希望者の「人となり」を把握し、消費性ローンの審査に活用しようとしているが、対象が限定的であり、実用化は困難であると考えられる。一方、上武他 [3] は、銀行口座の入出金データから分かる個人の行動特性を表す代表的な変数として、「ピーク残高」および「手数料支払回数」に関する変数を挙げている。毎月の銀行口座の取引の中で、通常は給与振込などのまとまった入金があったときに最も残高が大きくなりやすいが、その直後に資金を引き出すという行動特性を表すために「ピーク残高」に関する変数を導入している。また、中村 [6] は時間外の ATM の利用にかかる手数料に関して「何の抵抗もなく払っている。その姿勢こそが問題だ。」と指摘しているが、それを躊躇なく行うという行動特性を表すために、「手数料支払回数」に関する変数を導入している。これらの変数は銀行口座の平均残高に匹敵する説明力を有しており、審査態勢の強化に資するものといえる。

一方で、カードローンは法人向け貸出と比較して金額が小口であり、審査時と同様に管理コストを抑えるため、カードローン契約後は延滞の発生や極度額の増額の申し込みなどが無い限り特段のアフタフォロー等は実施していないことが多く、契約後の途上管理態勢は不十分であると考えられる。実際、2019 年 9 月に金融庁より発表された「銀行カードローンのフォローアップ調査結果」[1] によれば、融資審査態勢の見直し等の業務運営の改善に向けた取組みは進んでいるとの評価を受けた一方で、融資実行後の途上管理については、能動的に顧客の変化やその予兆を把握しようとする動きが鈍く、取組みが不十分であるとの評価を受けており、途上管理態勢の構築を速やかに進める必要がある。

年収証明書の定期的な取得などは、カードローン契約後に顧客の変化やその予兆を把握するのに効果があると思われるが、返済状況が悪化した場合などでない限り、現実的には実施は困難である。したがって、その他の方法で実際の借入・返済に関する行動特性を把握することができれば一定の効果が得られると考えられる。そこで、本研究では個人の借入・返済に関する行動特性を捉えるための一つのアプローチとして、カードローン利用者が借入・返済に利用している銀行口座の入出金データを分析する。

入出金データを利用してカードローンのデフォルトを分析およびモデル化している唯一の先行研究¹である上武他 [3] はカードローンの審査を目的としたモデルを構築しており、カードローン債務者の途上管理を目的として、借入・返済に関する個人の行動特性を考慮して分析・評価した研究は、筆者らの知る限り存在しない。本研究の主な貢献は以下の 2 点である。

- (1) 銀行口座の入出金データからカードローン契約後の借入・返済に関する個人の行動特性を把握する。
- (2) カードローンの途上管理を目的として、ローンの借入・返済に関する個人の行動特性および審査モデルの説明変数（個人の行動特性・取引状態・属性）を用いて、個人のカードローンのデフォルト確率を推定するロジットモデルを構築する。

本論文の構成は次の通りである。2 節では個人のデフォルトを評価するための問題の設定を示す。3 節ではモデル構築の概要を説明する。4 節では構築したモデルの評価精度を示し、その有効性について考察する。5 節では本論文をまとめ、結論と今後の課題を述べる。

¹銀行の口座情報（入出金データ）を利用した企業のデフォルト分析を行っている論文としては、根本他 [7]、本田他 [2]、三浦他 [5]、Norden and Weber[8]、Puri *et al.*[9]、Yao *et al.*[10] などがある。

2. 問題の設定

カードローンの途上管理に焦点を当てたときの行動特性を考慮し、個人のデフォルトを評価するための問題設定を行う。途上管理のモデルにおいても、カードローン審査時のモデルで利用した変数も候補とする一方で、モデル化の方法においては同様に行う点と相違する点が存在する。そこで、2.1 節では、本研究の対象とする途上管理との比較のために、カードローン審査時の問題設定およびモデル化の概要を示す。2.2 節では入出金データの概要、2.3 節ではデフォルト確率を推定するロジットモデルで使用する説明変数候補を説明する。2.4 節ではデフォルトの判定基準とそのためのデータ観測期間を示す。

2.1. 審査時の問題設定およびモデル化の概要 (上武他 [3])

上武他 [3] の研究を簡単に紹介し、本研究との関連も説明する。入出金データとして、2012 年 8 月から 2017 年 12 月までのカードローン契約前の摘要コードで分類した約 760 万件の明細データを用いる。具体的には、りそな銀行のカードローンの契約がある顧客 (カードローン契約者) に対するデフォルトデータと非デフォルトデータを 1 対 2 の割合でサンプル抽出している。実際の割合は非デフォルトデータの方がかなり多く、実際の割合で分析すると、非デフォルトを当てることに重きが置かれるようにモデル構築されるからである。デフォルトの判定基準は、延滞回数が一定以上になったことで判定することもあるが、実務に即してカードローンの契約者が実際に代位弁済となったときにデフォルトと判定している。後述するが、本研究でも摘要コード、デフォルト/非デフォルトの割合、デフォルトの判定基準は同様である。ただし、審査モデルであるので、カードローン契約時点を基準年月として、基準年月以前の一定期間を入出金観測期間、以降の一定期間をデフォルト観測期間としており、これらは本研究とは異なる。

カードローン契約時点までに入手可能な個人の行動特性を表す変数 (17 種類) と個人の取引状態・属性を表す変数 (6 種類) にカテゴリーを分類し、これらを用いてモデル構築を行っている。最終的には、個人の行動特性を表す 3 つの変数 (ピーク残高指数、時間外手数料支払回数、入金額対給与比率) と個人の取引状態・属性を表す 2 つの変数 (平均残高、年齢) の計 5 つの変数を用いたモデル (以降、審査時の 5 変数モデル) を構築し、クロスバリデーションやバックテストを行い、モデルの頑健性を検証している。本研究でも、2 つのカテゴリーに分類して変数を取り上げ、その中にこれらの 5 つの変数も変数候補として用いている。

2.2. 入出金データの概要

次に、本研究で用いた入出金データについて説明する。2013 年 4 月から 2018 年 12 月にりそな銀行のカードローンの契約がある顧客から入出金データをサンプル抽出する。ただし、カードローンの全体に占めるデフォルトの割合は非常に少なく、2.1 節で説明したように、非デフォルトデータとして、デフォルト先の 2 倍程度をランダムに抽出した約 7,200 件 (内デフォルト約 2,400 件、非デフォルト約 4,800 件) の入出金データをデータセットとする。その顧客に対する口座情報の明細単位で約 6 千万件の入出金データを用いる。入出金の明細データには 1 件毎に「摘要コード」といわれる取引を分類するコードがある。主な摘要コードを表 1 に示す。

入出金データを表 1 によって分類した上で加工し、月次での金額や件数への集計を行う。これは、審査時であっても途上管理であっても変わらないため、明細単位のデータを月次単位で集計したものを説明変数として用いる。

表 1: 入出金データの摘要コード

入金	給与、現金、振込、賞与
出金	現金、振替、送金、手数料(時間外、コンビニ、送金) 公共料金(電気・ガス・水道・NHK・電話)、保険料

2.3. 使用する変数

口座情報および入出金データを加工し、途上管理としてデフォルトを評価するモデルで使用する説明変数候補を 21 種類作成した。この中には、前述したように上武他 [3] の審査時の 5 変数モデルの 5 つの変数も含まれている。ここで、口座情報とは新規で預金口座を作成したときに顧客から得られる情報であり、具体的には年齢や給与振込設定の有無などである。

上武他 [3] と同様、変数の特徴を表すために下記の 2 つのカテゴリーへ分類し、途上管理においては、カテゴリーの違いによりどの程度デフォルトに影響するのかを調べる。

(A) 個人の行動特性を表す変数 (17 種類)

(B) 個人の取引状態・属性を表す変数 (4 種類)

(A) 個人の行動特性を表す変数

① 手数料支払回数

手数料支払回数は審査時の 5 変数モデルの説明変数であり、説明力が高い変数である。これは、時間外の ATM の利用にかかる手数料を躊躇なく支払うという行動特性を表しており、このような行動はカードローン契約後も変わらないと考えられる。そこで、途上管理においても、時間外手数料支払回数を説明変数の候補とする。

② ピーク残高指数

ピーク残高指数は審査時の 5 変数モデルの説明変数であり、説明力が高い変数である。これは、給与振込などでまとまった入金があったときに最も残高が大きくなりやすく、その直後に資金を引き出すという行動特性を表しており、資金に余裕がない借り手はカードローン契約後も変わらないと考えられる。そこで、途上管理においても、ピーク残高指数を説明変数の候補とする。

③ 入金額対給与比率

個人口座への入金は給与がメインであり、それ以外の入金は多くないのが普通である。不特定な入金や、その入金に対する出金が多く観測されると、本人の他の口座からの入金、もしくは他からの借入である可能性が高いと考えられる。入金額対給与比率は審査時の 5 変数モデルの説明変数であり、不特定な入金が途切れたときにデフォルトの懸念があると考えられるため、途上管理においても説明変数の候補とする。

④ 契約から初回借入までの月数

カードローン契約後にすぐに借入を行わない人の生活は、比較的安定していて、将来の一時的な資金不足に備えて契約したと考えられる。一方で、すぐに借入を行う人の資金需要は切迫していて、生活は不安定であると想定される。そこで、契約から初回借入までの月数を説明変数候補とする。

⑤ 借入回数 (2 種類)

借入回数が多いほど、資金需要が切迫していたり利用が計画的でなかったりと、生活は不

安定であると想定される。具体的には、月平均借入回数、初回借入月の借入回数を説明変数候補とする。

⑥ 返済回数

常に借入残高があり、資金に余裕がない利用者の返済回数は当初定められた月1回に近いと考えられるが、資金に余裕があれば定められた返済日以外にも追加で返済することができ、月平均の返済回数は1よりも大きくなると想定される。一方、借入残高がゼロとなる月もある利用者であれば、月平均の返済回数は1より低くなる。したがって、返済回数は返済行動の特徴が現れる指標と考えられ、具体的には月平均返済回数を説明変数候補とする。

⑦ 約定返済率

約定返済とは、カードローン契約時に定めた日に借入を返済することであり、観測期間中の約定返済回数を総返済回数で除した値を約定返済率と定義する。全体の返済回数に占める約定返済回数が多いほど、追加で返済する余裕がないと考え、約定返済率を説明変数候補とする。

⑧ 自動返済率

自動返済²とは、借入を行った、その当日中に借入返済用の預金口座に入金したことによって生じる返済のことである。例えば、公共料金の引き落としがある日に預金口座に残高不足が生じると、カードローンの借り入れが自動的に発生するが、当日中に入金すれば、自動的に返済に充当され、自動返済として取り扱われる。計画性があれば、前日までに預金口座へ入金すると考えると、自動返済は不注意による借入・返済の行動特性が現れると考えられる。そこで、観測期間中の自動返済回数を総返済回数で除した値を自動返済率と定義し、説明変数候補とする。

⑨ 手数料込借入率

カードローン借入の際、預金口座の残高が借入金額＋ATM利用手数料となっている場合³、手数料も含めて借り入れることになり預金口座に余裕がないと考えられる。具体的には、観測期間中の預金口座残高が手数料込となっている件数を借入回数で除した値を手数料込借入率と定義し、説明変数候補とする。

⑩ 借入後残高ゼロ率

手数料込借入率(⑨)と似ているが、借入後の口座残高がゼロとなった日数が多いほど余裕がないと考えられる。そこで、観測期間中の借入後の口座残高がゼロとなった日数を借入した日数で除した値を借入後残高ゼロ率と定義し、説明変数候補とする。

⑪ 入金回数/入金金額(6種類)

借入返済用の預金口座への入金回数や金額が少ないほど、資金に余裕がないと考えられる。現金や振込など入金の種別によっても異なることが想定される。具体的には、月平均現金入金回数、月平均現金入金金額、月平均振込入金回数、月平均振込入金金額、月平均入金回数、月平均入金金額を説明変数候補とする。

(B) 個人の取引状態・属性を表す変数

① 平均残高

途上管理においても審査時と同様に、預金残高の水準が高ければ資金に余裕があると考えられるので、審査時の5変数モデルの説明変数として採用されている期間中の平均預金残高

²金融機関のサービスによっては、他の返済の意味で使用される場合もあることを留意されたい。

³例えば、10万円を必要資金として借りる場合でも、借入後の残高が10万110円となっている場合などが相当する。

を説明変数候補とする。

② 年齢

途上管理においても、年齢は銀行口座を作成する際に取得する属性情報として代表的な変数であり、審査時の5変数モデルの説明変数として採用されている年齢を説明変数候補とする。

③ 給与振込(2種類)

途上管理においても審査時と同様に、給与振込の設定があれば、安定的に入金が見込める点で重要な変数と判断し、審査時の5変数モデルの説明変数として採用されている給与振込の有無を説明変数候補とする。また、途上管理においては、入出金データの観測期間も長くなり、その期間中に計測可能な月平均の給与振込回数も説明変数候補とする。

2.4. デフォルトの判定基準と観測期間

デフォルトの判定基準として、延滞回数が一定以上になったことをもって判定することもあるが、本研究では実務に即して審査時と同様、カードローンの契約者が実際に代位弁済となったときにデフォルトと判定する。カードローン契約後の途上管理という観点で分析を行うため、入出金観測の過去Xカ月以内に契約され、基準日から過去Yカ月分の入出金データ(カードローン契約後)を用いて、基準日からZカ月以内のデフォルトを予測する問題を設定する。ただし、入出金期間に延滞が継続している場合などは明らかな予兆が検知可能なため、基準日から4カ月間は除外期間とし、この間にデフォルトしたデータは利用しない。

図1に示すように、入出金観測開始時点から過去Xカ月間を契約期間、基準日から過去Yカ月間を入出金観測期間、基準日からZカ月間をデフォルト観測期間と呼ぶ。ただし、カードローンの全体に占めるデフォルトの割合は非常に少なく、非デフォルトデータはデフォルト先の約2倍程度のデータをランダムに抽出したものを使用した。また、紙面の都合上、3.2節以降の分析では、 $X=6$ 、 $Y=18$ の場合の結果のみを示す⁴。



図 1: 入出金およびデフォルト観測期間

3. モデル構築

本節では、2.3節で述べた入出金データから生成した説明変数を用いてロジットモデルを構築する。そこで3.1節では本研究で用いるロジットモデルを説明する。また、3.2節では、2.3節で生成した各説明変数の効果を調べるために、単変量ロジットモデルによる分析結果を示す。

3.1. ロジットモデル

上武他 [3] と同様に、本研究においてもロジットモデルを用いる。個人の信用リスクの分野では住宅ローンの審査などにも一般的なモデルとして採用されている例が多い。

⁴(X,Y) の組み合わせとして (6,6), (12,6), (12,12), (6,12), (3,12), (6,18) を試したが、(6,18) の組み合わせが最も AR 値が高くなった。

カードローン利用者 i の途上管理における入出金データを用いて作成した N 個の変数 $x_{ij}(i = 1, \dots, n; j = 1, \dots, N)$ を使用して、ロジットモデルを構築し、パラメータ $\alpha_j(j = 1, \dots, N)$ は最尤法によって推定する。ここで、 p_i は利用者 i のデフォルト確率、 n はカードローン利用者数を表す。このとき、 z_i が大きいほど利用者 i のデフォルト確率は低くなる。

$$p_i = \frac{1}{1 + \exp(z_i)}, \quad z_i = \ln \left(\frac{1 - p_i}{p_i} \right) = x_0 + \sum_{j=1}^N \alpha_j x_{ij} \quad (i = 1, \dots, n) \quad (1)$$

説明変数の中にはデフォルトとの関係が非線形のものも含まれているため、説明変数ごとに決定木モデルを用いてカテゴリー化を行う。決定木の構成は、エントロピーに基づくゲインによって決定する。決定木の深さは変数の特徴や途上管理への適用を考慮し、個別に設定している。各カテゴリーをスコア化した値を説明変数の値とし、デフォルト率が高い場合に、変数の値が大きくなるようにスコア化する。モデルの評価精度を計測する指標としては、AR (Accuracy Ratio) 値を用いる。

3.2. 単変量ロジット分析

2.3 節で述べた説明変数候補に対して、個々の変数を単独で用いた場合の効果を調べるために、単変量ロジットモデルによる分析結果を示す。2.2 節に記載したデータから表 2 の通りサンプル抽出したデータを対象として、変数ごとにロジットモデルを作成し、AR 値を計測した。また、デフォルト観測期間 Z は上武他 [3] と同様に、12 カ月、15 カ月、18 カ月の場合についても算出した。

表 2: 対象データ

	デフォルト観測期間		
	12 カ月	15 カ月	18 カ月
デフォルト	約 1,400 件	約 1,900 件	約 2,400 件
非デフォルト	約 2,800 件	約 3,800 件	約 4,800 件
全体	約 4,200 件	約 5,700 件	約 7,200 件

デフォルト観測期間 12 カ月の分析結果を表 3 に示す。審査時も使用した説明変数については、対象データが異なるため一概に比較することはできないが、参考のためデフォルト期間 12 カ月の審査時の AR 値を上武他 [3] から引用し、右端に示す。デフォルト観測期間に関わらず最も AR 値が高かった変数は、審査時と同様に個人の状態を表すカテゴリー変数 (B) の平均残高であった。カードローンの返済は預金から支払われることになるため、預金残高の水準が高ければそれだけデフォルトのリスクは減少するというのは直感的に納得できる結果である。一方で個人の行動特性を表すカテゴリー (A) の中では、借入後残高ゼロ率の AR 値が高かったが、これは残高の影響が大きいと考えられる。次に AR 値が大きいのは月平均借入回数と月平均返済回数である。これは利用者の行動特性を表す変数であり、残高に関係する変数に次いでデフォルトに与える影響が大きいことを表している。行動特性を表す変数のうち、振込入金に関する変数は全体的に AR 値が低いことが分かる。審査時の説明変数として採用されていたカテゴリー (A) の説明変数は、途上管理において必ずしも上位に位置していない。給与振込有無は「有」「無」それぞれ相応のデータがあるものの、審査時と同様に AR 値が低い。通常カードローンの契約をする際には、「継続的な収入がある」ことが必

要とされるため、対象の銀行口座に給与振込が設定されていなくても他銀行の口座では給与振込が設定されていることが想定され、給与振込の有無のみではデフォルトの判定は困難であることが分かる。ただ、月平均給与振込回数は年齢よりも AR 値が高く、相応の精度があることが確認できる。給与振込回数がゼロであれば他の口座に給与振込があることも考えられるが、そうでなければ、収入の安定性を表す指標として、給与振込の有無よりも精度が高くなったと考えられる。

表 3: 単変量ロジット分析結果

カテゴリー (2.3 節)	変数 (カテゴリー内 AR 値を降順)	上武他 [3] AR 値
(A) 個人の 行動特性	借入後残高ゼロ率	30.8%
	月平均返済回数	
	月平均借入回数	
	時間外手数料支払回数	
	手数料込借入率	
	契約から初回借入までの月数	31.0%
	初回借入月の借入回数	
	ピーク残高指数	
	月平均入金金額	
	月平均現金入金回数	
	月平均現金入金金額	3.1%
	約定返済率	
	自動返済率	
	入金額対給与比率	
	月平均入金回数	
(B) 個人の取引 状態・属性	月平均振込入金回数	34.8%
	月平均振込入金金額	
	平均残高	
	月平均給与振込回数	
	年齢	17.3%
	給与振込有無	0.0%

4. モデルの評価

入出金データから生成した変数候補の 21 個を用いてデフォルトモデルを構築し、その分析結果を示す。4.1 節ではロジットモデルのパラメータ推定結果を、4.2 節ではアウトオブサンプルテストの結果を述べる。

4.1. パラメータ推定結果

3.2 節と同じ対象データを用いて、パラメータを推定した結果を表 4 に示す。2.3 節で述べた説明変数候補の 21 個に対して、AIC を評価基準として変数増減法によるステップワイズ法を用いてロジットモデルの変数選択を行う。3.2 節でスコア化した説明変数は値が大きいほどデフォルトとなる可能性が高くなるため、パラメータの符号条件はプラスである。また、スコアは標準化しており、パラメータの値が大きいほど説明変数がデフォルトへ与える影響が大きいことを表している。なお、多変量ロジット分析で計算されたスコアがタイデータとなることは単変量ロジット分析と比べると少ないが、3.2 節と同様に、AR 値はその最大と最小の平均値を示す。

表 4: 各デフォルト観測期間のモデル構築結果

No (※ 1)	説明変数	パラメータ (p 値)(※ 2)		
		12 カ月	15 カ月	18 カ月
A1	時間外手数料支払回数	0.2989 (<.0001)	0.3045 (<.0001)	0.3292 (<.0001)
A2	ピーク残高指数	0.1949 (0.0004)	0.2754 (<.0001)	0.2349 (<.0001)
A3	入金額対給与比率	—	0.1960 (0.0064)	0.2551 (<.0001)
A4	契約から初回借入までの月数	—	—	0.0978 (0.0771)
A5	月平均借入回数	0.2536 (0.0046)	0.2389 (0.0004)	0.2126 (0.0019)
A6	月平均返済回数	0.2017 (0.0273)	0.3194 (<.0001)	0.2976 (0.0001)
A7	初回借入月の借入回数	0.1359 (0.0131)	0.1042 (0.0275)	0.1280 (0.0026)
A8	約定返済率	0.3887 (0.0002)	0.3308 (<.0001)	0.4167 (<.0001)
A9	自動返済率	0.1444 (0.1264)	—	—
A10	手数料込借入率	—	—	—
A11	借入後残高ゼロ率	0.2163 (0.0007)	0.1581 (0.0029)	0.2032 (<.0001)
A12	月平均現金入金回数	0.1553 (0.0159)	—	—
A13	月平均現金入金金額	—	—	0.0783 (0.1052)
A14	月平均振込入金回数	—	0.0789 (0.1002)	—
A15	月平均振込入金金額	—	—	0.0918 (0.0401)
A16	月平均入金回数	—	—	—
A17	月平均入金金額	0.1234 (0.0947)	0.0866 (0.1419)	0.1062 (0.0682)
B1	年齢	0.2154 (<.0001)	0.1570 (0.0017)	0.1142 (0.0059)
B2	平均残高	0.3451 (<.0001)	0.3351 (<.0001)	0.2583 (<.0001)
B3	給与振込有無	-0.1181 (0.0565)	-0.1806 (0.0029)	-0.2001 (0.0002)
B4	月平均給与振込回数	0.2130 (0.0002)	0.2095 (0.0002)	0.1456 (0.0037)
AR 値		56.7%	56.0%	55.7%

(※ 1) 2.3 節のカテゴリーを番号の前に付けている。

(※ 2) 「—」はステップワイズ法で選択されなかった変数を表す。

まずはじめに、デフォルト観測期間が12カ月の結果を述べる。最もパラメータが大きかったのは約定返済率 (A8) である。これは、3.2 節の単変量ロジット分析では AR 値がそれほど高くなかったこととは対照的な結果である。次に大きかった変数が平均残高 (B2) である。これは、3.2 節の単変量ロジット分析からも想定された結果である。また、契約から初回借入までの月数 (A4) や手数料込借入率 (A10) は、単変量分析の AR 値で相応に上位に位置していたが、多変量ロジット分析では変数として採用されなかった。給与振込有無 (B3) は表 5 に示すように、契約から初回借入までの月数 (A4) は月平均返済回数 (A5) との相関係数が 0.425、手数料込借入率 (A10) は時間外手数料支払回数 (A1) との相関係数が 0.445、借入後残高ゼロ率 (A11) との相関係数が 0.413、と比較的高いことが原因と考えられる。

また、その他の説明変数については、概ね単変量ロジット分析で AR 値が低い説明変数はパラメータの値が小さかったり、変数として選択されない傾向となっている。給与振込有無のパラメータがマイナスとなっている。これは、月平均給与振込回数 (B4) と給与振込有無 (B3) の相関が高い (0.328) ことによる多重共線性の影響であると考えられる。

次に、12 カ月のデフォルト観測期間と比較のため、15 カ月および 18 カ月の結果についても述べる。説明変数の選択結果やパラメータの傾向は、平均残高 (B2) や約定返済率 (A8) のパラメータが高くなるなど 12 カ月の場合と概ね同様の傾向がみられる。ただ、デフォルト観測期間が 12 カ月から 15 カ月、18 カ月と長期化するにつれて平均残高 (B2) のパラメータが小さくなる一方で、時間外手数料支払回数 (A1) のパラメータは大きくなる。本研究では、カードローン契約後の情報を用いてデフォルト確率を推定するモデルを用いているが、デフォルト観測期間が長期化するほど年齢や預金残高などの個人の取引状態・属性よりも個

人の行動特性に依存する説明変数が重要となることが分かる。

表 5: 変数間の相関係数

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	B1	B2	B3	B4
A1	1	0.087	-0.038	0.113	0.291	0.16	0.185	0.059	0.12	0.445	0.1	0.06	0.08	-0.098	-0.065	0.16	-0.1	0.202	0.008	-0.076	0.043
A2	0.087	1	0.145	0.089	0.16	0.14	0.066	0.042	0.104	0.164	0.252	0.127	0.071	-0.002	0.017	0.025	0.073	0	0.233	0.1	0.164
A3	-0.038	0.145	1	0.06	0.048	0.071	0.034	-0.067	-0.004	0.017	0.145	0.373	0.171	-0.02	-0.163	-0.085	0.214	-0.064	0.212	0.572	0.493
A4	0.113	0.089	0.06	1	0.425	0.589	0.304	0.203	0.243	0.221	0.234	0.243	0.16	-0.014	-0.011	0.232	-0.012	-0.079	0.217	-0.002	-0.012
A5	0.291	0.16	0.048	0.425	1	0.615	0.362	0.117	0.265	0.449	0.423	0.339	0.179	-0.075	-0.035	0.358	-0.109	-0.021	0.202	-0.085	-0.001
A6	0.16	0.14	0.071	0.589	0.615	1	0.291	0.391	0.401	0.307	0.313	0.314	0.218	0.009	0.029	0.336	0.047	-0.074	0.256	-0.007	0.01
A7	0.185	0.066	0.034	0.304	0.362	0.291	1	0.073	0.133	0.243	0.213	0.166	0.107	-0.073	-0.026	0.16	-0.055	0.025	0.122	-0.061	-0.01
A8	0.059	0.042	-0.067	0.203	0.117	0.391	0.073	1	0.387	0.099	0.077	-0.052	0.084	0.045	0.064	0.016	0.174	-0.02	0.131	-0.024	0.015
A9	0.12	0.104	-0.004	0.243	0.265	0.401	0.133	0.387	1	0.186	0.195	0.094	0.069	-0.012	0.024	0.07	0.03	-0.038	0.118	-0.057	-0.003
A10	0.445	0.164	0.017	0.221	0.449	0.307	0.243	0.099	0.186	1	0.413	0.11	0.114	-0.049	-0.025	0.164	0.002	0.092	0.142	-0.016	0.047
A11	0.1	0.252	0.145	0.234	0.423	0.313	0.213	0.077	0.195	0.413	1	0.197	0.178	0.133	0.105	-0.005	0.261	0.017	0.464	0.118	0.14
A12	0.06	0.127	0.373	0.243	0.339	0.314	0.166	-0.052	0.094	0.11	0.197	1	0.373	0.105	0.074	0.342	0.081	-0.05	0.25	0.208	0.148
A13	0.08	0.071	0.171	0.16	0.179	0.218	0.107	0.084	0.069	0.114	0.178	0.373	1	0.168	0.09	0.134	0.244	0.019	0.246	0.102	0.07
A14	-0.098	-0.002	-0.02	-0.014	-0.075	0.009	-0.073	0.045	-0.012	-0.049	0.133	0.105	0.168	1	0.232	-0.158	0.444	-0.011	0.342	0.189	0.032
A15	-0.065	0.017	-0.163	-0.011	-0.035	0.029	-0.026	0.064	0.024	-0.025	0.105	0.074	0.09	0.232	1	-0.077	0.437	0.067	0.306	-0.108	-0.089
A16	0.16	0.025	-0.085	0.232	0.358	0.336	0.16	0.016	0.07	0.164	-0.005	0.342	0.134	-0.158	-0.077	1	-0.249	-0.034	-0.051	-0.16	-0.099
A17	-0.1	0.073	0.214	-0.012	-0.109	0.047	-0.055	0.174	0.03	0.002	0.261	0.081	0.244	0.444	0.437	-0.249	1	0.047	0.6	0.398	0.276
B1	0.202	0	-0.064	-0.079	-0.021	-0.074	0.025	-0.02	-0.038	0.092	0.017	-0.05	0.019	-0.011	0.067	-0.034	0.047	1	0.002	-0.083	0.015
B2	0.008	0.233	0.212	0.217	0.202	0.256	0.122	0.131	0.118	0.142	0.464	0.25	0.246	0.342	0.306	-0.051	0.6	0.002	1	0.292	0.209
B3	-0.076	0.1	0.572	-0.002	-0.085	-0.007	-0.061	-0.024	-0.057	-0.016	0.118	0.208	0.102	0.189	-0.108	-0.16	0.398	-0.083	0.292	1	0.328
B4	0.043	0.164	0.493	-0.012	-0.001	0.01	-0.01	0.015	-0.003	0.047	0.14	0.148	0.07	0.032	-0.089	-0.099	0.276	0.015	0.209	0.328	1

4.2. アウトオブサンプルテスト

本研究では、オーバーフィッティング⁵の可能性を考えて、2.2 節のデータからモデル構築に使用していないデータをサンプル抽出したものをアウトオブサンプルデータとして使用する。実際に運用することを想定してアウトオブサンプルデータの AR 値を評価した結果を表 6 に示す。アウトオブサンプルデータの AR 値をみると、いずれのデフォルト観測期間においても AR 値が 50% を超える水準となった。また、インサンプルデータの AR 値と比較するとデフォルト観測期間が 15 カ月以降、多少 AR 値の乖離がみられ、オーバーフィッティングの可能性がみられる。

表 6: アウトオブサンプルデータの対象データおよび AR 値

	デフォルト観測期間		
	12 カ月	15 カ月	18 カ月
デフォルト	約 700 件	約 850 件	約 1,200 件
非デフォルト	約 1,400 件	約 1,900 件	約 2,400 件
全体	約 2,100 件	約 2,750 件	約 3,600 件
AR 値 (アウトオブサンプル)	53.6%	51.2%	51.5%
(インサンプル)	56.7%	56.0%	55.7%

4.3. モデルの再構築

表 4 で構築したモデルで採用された変数で、相関が高い変数や、他の変数との相関の影響でパラメータの推定値がマイナスになったと思われる変数があった。そこで、それらの変数を説明変数候補から除いてモデルの再構築を行う。結果を表 7 に示す。いずれのデフォルト観測期間においても、表 4 の結果と比較してデフォルト観測期間が 12 カ月から 15 カ月、18 カ月と長期化するにつれて平均残高 (B2) のパラメータが小さくなり、月平均返済回数 (A6) のパラメータが大きくなるなど個人の取引状態・属性よりも個人の行動特性に依存する説明

⁵インサンプルデータの当てはまりはよく、アウトオブサンプルデータに対する当てはまりが極端に悪くなる場合、オーバーフィッティングが懸念される。

変数が重要となることが明確になった。また、いずれのパラメータも 0.01%水準で統計的に有意であることも確認できる。AR 値も再構築前と比較してほぼ同水準を維持することができた。インサンプルとアウトオブサンプルの AR 値の差はあまり変わらなかった。

表 7: 各デフォルト観測期間のモデル構築結果

No	説明変数	パラメータ (p 値)		
		12 カ月	15 カ月	18 カ月
A1	時間外手数料支払回数	0.3005 (<.0001)	0.2955 (<.0001)	0.3276 (<.0001)
A2	ピーク残高指数	0.1949 (0.0004)	0.2754 (<.0001)	0.2349 (<.0001)
A3	入金額対給与比率	0.1898 (0.0005)	0.2742 (<.0001)	0.2277 (<.0001)
A5	月平均借入回数	0.2698 (0.0015)	0.2339 (0.0004)	0.2035 (0.0022)
A6	月平均返済回数	0.2528 (0.0047)	0.3388 (<.0001)	0.3688 (<.0001)
A7	初回借入月の借入回数	0.1405 (0.0098)	0.1048 (0.0255)	0.1431 (0.0006)
A8	約定返済率	0.4291 (<.0001)	0.3332 (<.0001)	0.4228 (<.0001)
A11	借入後残高ゼロ率	0.2223 (0.0004)	0.1638 (0.0019)	0.2249 (<.0001)
B1	年齢	0.2174 (<.0001)	0.1649 (0.0010)	0.1224 (0.0027)
B2	平均残高	0.4057 (<.0001)	0.3688 (<.0001)	0.3224 (<.0001)
B4	月平均給与振込回数	0.2224 (<.0001)	0.2620 (<.0001)	0.2083 (<.0001)
AR 値	(インサンプル)	56.0%	55.4%	54.5%
	(アウトオブサンプル)	53.4%	50.6%	51.3%

5. まとめ

本研究では、銀行口座の入出金データから個人の行動特性を把握し、それを説明変数にしてカードローン契約後の途上管理を目的とした個人のカードローンのデフォルトを評価するモデルを構築した。いずれのデフォルト観測期間においても、延滞の情報を入れずにアウトオブサンプルデータで AR 値が 50%を超える結果となり、カードローンのデフォルトモデルとしては比較的高い水準を得ることができた。

本研究によらずとも予測できる結果であるが、審査時と途上管理いずれにおいても銀行口座の平均残高がデフォルトに大きく影響することがあらためて確認できた。本研究では、それに加えて途上管理においては、約定返済率や月平均借入回数、月平均返済回数などカードローン利用後のみに観測される行動特性を表す説明変数の影響が大きいことが分かった。また、審査時に行動特性を表す説明変数として代表的な時間外手数料支払回数は本研究においても相応の説明力を有している。途上管理においては、これらの説明変数を組み合わせて、より高い精度をもつモデルを構築し、デフォルトを評価することが可能であると考えられる。今後の課題としては、以下の点が挙げられる。

- (1) 本研究では、4.3 節において相関が高い変数や、パラメータの推定値がマイナスになった変数を除いてモデル構築を行い、アウトオブサンプルデータにおいても相応の AR 値が得られ、モデルの頑健性を確認することができた。ただし、より頑健性を確かめるためには、クロスバリデーションを行い、検証する必要がある。
- (2) 本研究で設定した契約期間、入出金観測期間は、最も AR 値が高い組み合わせを用いて分析を実施した。しかし、実務への応用を考えた場合には、様々な期間の組み合わせに対する感度分析を実施し、頑健な期間の抽出の検証が必要である。

参考文献

- [1] 金融庁 (2019), 銀行カードローンのフォローアップ調査結果について.
<https://www.fsa.go.jp/news/r1/ginkou/20190918/002.pdf>
- [2] 本田大悟, 大古田俊介, 井實康幸 (2018). 「入出金データを用いた企業デフォルト予測」, 『日本金融・証券計量・工学学会, 2017 年冬季大会予稿集』, 192–203.
- [3] 上武治紀, 吉田博哉, 枇々木規雄 (2020), 「銀行口座の入出金情報に基づく個人の行動特性を考慮したカードローンのデフォルト分析とモデル化」, 日本統計学会誌, **49**(2), 217–240.
- [4] 増井智則 (2014), 「ソーシャルメディアを活用した新たなリテール金融与信管理の可能性」, 三菱総合研究所, コラム・レポート.
<https://www.mri.co.jp/opinion/column/uploadfiles/20140929report.pdf>
- [5] 三浦 翔, 井實 康幸, 竹川 正浩 (2020). 「入出金情報を用いた信用リスク評価 —機械学習による実証分析—」, 日本統計学会誌, **49**(2), 187–216.
- [6] 中村芳子 (2014), 『20 代のいま、やっておくべきお金のこと』ダイヤモンド社.
- [7] 根本直子, 吉野直行, 大久保豊, 稲葉大明, 柳澤健太郎 (2019). 「銀行口座動態情報を活用した中小企業の信用リスク分析」, 『現代ファイナンス』, **40**, 49–71.
- [8] Norden, L. and Weber, M. (2010). Credit line usage, checking account activity, and default risk of bank borrowers, *The Review of Financial Studies*, **23**(10), 3665–3699.
- [9] Puri, M., Rocholl, J. and Steffen, S.(2017). What do a million observations have to say about loan defaults? Opening the black box of relationships, *Journal of Financial Intermediation*, **31**, 1–15.
- [10] Yao, J., Chapira, M. L. and Margaryan, M. (2017). Checking account activity and credit default risk of enterprises: An application of statistical learning methods, *arXiv:1707.00757v1*.
- [11] 全国銀行協会 (2017), 銀行による消費者向け貸付けに係る申し合わせ.
<https://www.zenginkyo.or.jp/fileadmin/res/news/news290336.pdf>