

創業企業向け信用リスクモデルにおける人的要因の有効性 — 創業時の年齢と斯業経験年数の効果 —

05000227 日本政策金融公庫 国民生活事業本部 尾木 研三 OGI Kenzo
日本政策金融公庫 国民生活事業本部 *峰下 正博 MINESHITA Masahiro
日本政策金融公庫 国民生活事業本部 内海 裕一 UTSUMI Yuichi
01505910 慶應義塾大学理工学部 枇々木 規雄 HIBIKI Norio

1. はじめに

創業前の企業は決算書の情報が無い。尾木ら[1]は創業前の企業の非財務情報だけを用いて信用リスクモデル（ロジスティック回帰モデル）を構築し、実務で利用可能な精度であることを明らかにした。ただ、重要な人的変数である「創業時の年齢」や「斯業経験年数」については、データ数が少なく、オーバーフィッティングの可能性があるので、ダミー変数を用いている。

この点について、現場の経験則や先行研究を踏まえると、仮説として、デフォルトに対して線形もしくは下に凸型の曲線が考えられる。そこで、本研究では、その後蓄積した約11万社のデータを用いて二つの人的変数とデフォルトとの関係性を詳細に見直し、モデルの精度向上を目指す。さらに、創業にはベストなタイミング（旬）があるといわれていることから、二つの変数を組み合わせた新たな「旬」変数の有効性を確認する。

分析の結果、「創業時の年齢」は4次関数、「斯業経験年数」は2次関数で近似できることがわかった。また、飲食店において創業の「旬」ダミー変数が有意であることを確認した。これらの変数をモデルに投入すると、尾木ら[1]のモデル（以下、OUHモデル）に比べてAR値が5.62%ポイント上昇した。

2. データの概要

本研究の使用データは、日本政策金融公庫が2011年度から2018年度までに融資した創業企業のデータDB1と、2017年以降に取得を開始した情報を保管したDB2である。概要を表1に示す。*印はOUHモデルで採用された変数である。

表1 使用データの概要

データベース	融資時期	データ数	主な変数
DB1	2011年4月～ 2019年3月	108,718	*創業者のクレジットヒストリー *融資金額、*資本金 など
DB2	2017年1月～ 2019年3月	33,550	*創業時の年齢、*斯業経験年数 *業種、*創業者の手持ち資金 など

3. ロジスティック回帰モデルの構築

3.1 モデルの概要

データベース(DB)によって使用できる変数が異なるため、尾木ら[1]と同様にDBごとにモデルを構築し、それぞれのモデルから算出されるスコアを加重する統合モデルを構築する。まず、DB1が保有する創業企業*i*の*X*個の変数（主に創業者のクレジットヒストリーに関する変数） $f_{1,i,x}(i=1,\dots,N_1;x=1,\dots,X)$ を用いてモデル1を構築し、最尤法によってパラメータ $\beta_{1,x}(x=1,\dots,X)$ を推定する。ここで、 $p_{1,i}$ は融資した創業企業*i*が融資後2年以内にデフォルトする確率、*u*はモデル番号、 N_1 はDB1の創業企業数を表す。 $z_{u,i}$ が大きいほどデフォルト確率は低くなる。

$$p_{1,i} = \frac{1}{1 + e^{z_{1,i}}}, z_{1,i} = \ln\left(\frac{1 - p_{1,i}}{p_{1,i}}\right) = \alpha_{1,0} + \sum_{x=1}^X \beta_{1,x} f_{1,i,x}$$

次に、DB2が保有する創業企業*i*の*J*個の人的要因に関する変数

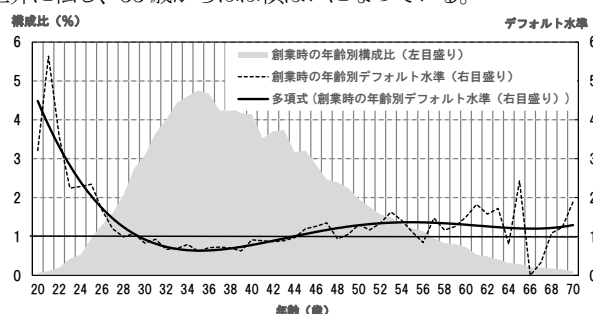
$h_{2,i,j}(i=1,\dots,N_2;j=1,\dots,J)$ 、*K*個の金融要因に関する変数 $f_{2,i,k}(i=1,\dots,N_2;k=1,\dots,K)$ 、*M*個の業種要因に関する変数 $g_{2,i,m}(i=1,\dots,N_2;m=1,\dots,M)$ を使用して、モデル2を構築し、パラメータ $\beta_{2,j}(j=1,\dots,J)$ 、 $\gamma_{2,k}(k=1,\dots,K)$ 、 $\delta_{2,m}(m=1,\dots,M)$ を推定する。 N_2 はDB2の創業企業数を表す。

$$p_{2,i} = \frac{1}{1 + e^{z_{2,i}}}, z_{2,i} = \ln\left(\frac{1 - p_{2,i}}{p_{2,i}}\right) = \alpha_{2,0} + \sum_{j=1}^J \beta_{2,j} h_{2,i,j} + \sum_{k=1}^K \gamma_{2,k} f_{2,i,k} + \sum_{m=1}^M \delta_{2,m} g_{2,i,m}$$

精度改善のポイントは変数選択である。変数次第でモデルの精度が左右され、有効な変数を一見つけただけで性能が大きく向上することもある。OUHモデルでは、「創業時の年齢」「斯業経験年数」についてダミー変数を使用している。本研究では、二つについて詳細な分析を行い、モデル2の精度向上に貢献する変数を見つける。その際、実務での利用を念頭に、先行研究だけではなく、審査員の経験則との整合性も重視する。

3.2 創業時の年齢の分析

まず、「創業時の年齢」について、審査員の経験則として「創業時の年齢は、若過ぎても、高齢でもいけない」というものがある。先行研究をみると、創業時の年齢は創業後のパフォーマンスに対してマイナス方向に線形で有意であるとする研究が多い。ただ、玄田[2]は、独立開業時の年齢の2乗が有意で、40歳くらいが最適であることを示している。図1に創業時の年齢別デフォルト水準を示す。これをみると、35歳まで低下したあと上昇に転じ、55歳からほぼ横ばいになっている。



注：デフォルト水準は全体のデフォルト率を「1」として指数化した値

図1 創業時の年齢別の構成比とデフォルト水準

そこで、年齢別デフォルト水準の3次から5次関数の3パターンを作成し、OUHモデルの40歳未満ダミーとAR値を比較した。表2のとおり、4次関数を変数に採用する。

表2 年齢関数（*n*次多項式）のスコアのAR値

3次関数	4次関数	5次関数	40歳未満ダミー
16.7%**	17.6%**	17.6%	6.5%**

注：**はすべての*n*次項が1%水準で有意。5次関数は3次項が非有意。

3.3 斯業経験年数の分析

次に、斯業経験年数について、経験則として「斯業経験年数は長いほど良いが、長すぎてもいけない」というものがある。先行研究は、創業後のパフォーマンスに対してプラス方向に線形に有意であるとする研究が多いが、玄田[2]は、年数の2乗が経済的成功に対して有意であることを示しており、審査員の経験則と整合的な結果となっている。図2に斯業経験年数別デフォルト水準を示す。これを見ると、18.51年まで低下したあと上昇に転じており、経験則や玄田[2]と整合的な結果となった。

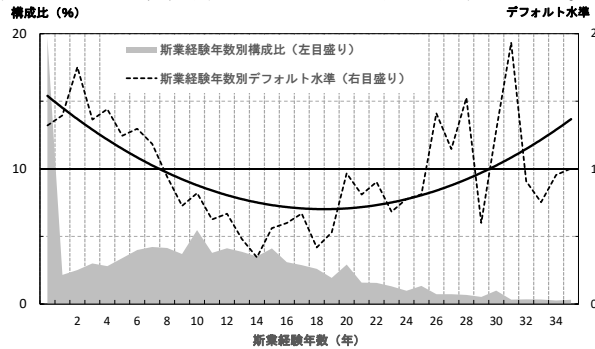


図2 斯業経験年数別の構成比とデフォルト水準

斯業経験年数とデフォルトの関係について、斯業経験年数別デフォルト水準の1次から3次の3パターンを作成し、OUHモデルの5年以内ダミーとAR値を比較した。結果は表3のとおりで、2次関数が最も高いため、これを変数に採用する。

表3 斯業経験年数関数 (n次多項式) のスコアのAR値

1次関数	2次関数	3次関数	5年以内ダミー
13.8%**	19.5%**	18.7%	13.2%**

注：**はすべてのn次項が1%水準で有意。3次関数は1次項が非有意。

3.4 創業の旬の分析

審査員の経験則として「若いころから経験を積みながら資金を貯め、機が熟したタイミングで創業するパターンが成功しやすい」というものがある。玄田[2]は、独立開業には「旬」があり、20年くらい経験を積み、40歳くらいで開業するケースが経済的成功を収めやすいことを示した。図3に二つの変数をクロスした飲食店の散布図を示す。全体の図をみると、対角線上に並んでいる企業とそうでない企業があることがわかる。

飲食店や小売業のように参入障壁が低く競争が激しい業種の場合、業種全体のデフォルト率は高くても、一定の経験を積んで暖簾分けのような形で「旬」に創業した企業は相対的にデフォルト率が低い可能性がある。とりわけ、飲食店は創業企業全体の3割近くを占めており、より精緻な分析が必要である。

そこで、デフォルト水準が「1」よりも高い細分類業種に属する企業「DF高」と、低い細分類業種に属する企業「DF低」に分けた。右図のDF低をみると、対角線上が濃くなっており、「旬」で創業している企業が多く含まれていることがわかる。

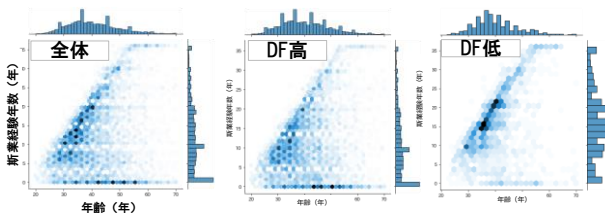


図3 飲食店の散布図

次に、表4のとおり、創業時の年齢と斯業経験年数をそれぞれ5歳刻みにしてデフォルト水準をみてみると、創業時の年齢は30歳から45歳、斯業経験年数は11年から25年のゾーンに、デフォルト水準が低い企業が集中している(濃い網掛け部分)。

表4 飲食店の「旬」とデフォルト水準

創業時の年齢	創業時の年齢									
	20歳～	20歳～	25歳～	30歳～	35歳～	40歳～	45歳～	50歳～	55歳超	
0年	0.0	0.4	2.2	1.0	0.7	1.6	1.2	1.7	0.7	
1～5年	0.0	2.1	1.2	1.0	0.9	1.4	0.9	1.2	0.8	
6～10年	—	1.4	1.4	0.8	0.8	1.1	1.4	1.3	1.7	
11～15年	—	—	1.0	0.5	0.5	0.7	1.1	1.1	1.9	
16～20年	—	—	—	2.3	0.6	0.6	0.9	0.9	0.9	
21～25年	—	—	—	—	1.6	0.5	0.9	1.8	0.6	
26～30年	—	—	—	—	—	0.5	1.5	1.0	2.3	
30年超	—	—	—	—	—	—	2.0	0.8	1.8	

注：データ欠けと数値が薄いセルは件数構成比が1%未満のゾーン

4. モデルの構築と精度データの概要

モデル1の変数は、创业者のクレジットヒストリーに関する7変数が有意となり、AR値は38.2%となった。モデル2の変数は、人的要因で、創業時の年齢(4次関数)と斯業経験年数(2次関数)、飲食店における創業の「旬」ダミー(表4の濃い網掛け部分とそれ以外)が、金融要因で1変数、業種要因で8変数が選択され、AR値は42.7%となった。

最後にモデル1とモデル2のスコアを用いて統合モデルのモデル3を構築する。まず、DB1とDB2の両方のデータを保有する33,550社のデータを用いて、創業企業ごとにモデル1のスコア $z_{1,i}^*$ とモデル2のスコア $z_{2,i}^*$ を算出する。次に、各スコアを説明変数としたロジスティック回帰モデル(モデル3)を構築する。モデル3の式は以下のとおりである。

$$p_{3,i} = \frac{1}{1 + e^{z_{3,i}}} , z_{3,i} = \alpha_3 + \omega_1 z_{1,i}^* + \omega_2 z_{2,i}^*$$

モデル3の結果を表5に示す。モデルの精度を表すAR値はアウトオブサンプルで5.62%ポイント上昇した。

表5 統合モデルのAR値

	①OUH	②本研究	②-①
インサンプル	48.06%	53.06%	4.99%
アウトオブサンプル	48.60%	54.23%	5.62%

5. おわりに

本研究では、創業企業向けモデルで重要な変数となる人的要因とデフォルトとの関係を、実務や理論との整合性を踏まえて詳細に分析した。多くの先行研究は線形関係をベースに分析しているが、変数とデフォルトとの関係は線形であるという保証はない。この点に十分注意して分析した結果、変数とデフォルトとの非線形の関係の有意性が明らかになったほか、「創業の旬」という新変数を見つけることにも成功し、モデルの精度を向上させることができた。本研究の成果が信用リスクモデルの精度改善に取り組む金融機関の参考になれば幸いである。

参考文献

- 尾木研三, 内海裕一, 枇々木規雄(2017)「創業企業の信用リスクモデル」『リスク管理・保険とヘッジ(ジャフィー・ジャーナル: 金融工学と市場計量分析)』, 133-156, 朝倉書店。
- 玄田有史(2001)「独立の旬: 開業のためのキャリア形成」, 『国民生活金融公庫「新規開業実態調査」の再分析(1991年～2000年)』, SSJ Data Archive Research Paper Series, SSSDA-17