

# P2P Lending における集中投資リスクを考慮した投資意思決定モデル

プロフィットエンジニアリング研究

5216F019-5 山崎 慎一郎  
指導教員 大野 高裕

## An Investment Model Considering Concentration Risk in P2P Lending

YAMAZAKI Shinichiro

### 1. はじめに

近年、P2P Lending (Peer to Peer Lending) と呼ばれるインターネットを利用した個人間融資仲介サービスが注目を集めている。P2P Lending は、インターネットを利用することで仲介者が受け取る手数料を抑え、借手にとっては銀行借入よりも低い金利での借入を可能にし、貸手にとっては銀行預金よりも高い利率での運用を可能にしている。これらのメリットにより、アメリカや中国をはじめ世界中で急速に普及している。P2P Lending における貸手、運営会社、借手の関係を図 1 に示す。

しかし、P2P Lending には問題点もある。第一に、貸手と運営会社の間に情報の非対称性が存在する。借手の信用リスクは貸手が負うため、貸手は借手に関する情報を入手する必要がある。運営会社は、借手から借入の申し込みを受ける段階で借手に関する情報を入手するが、全ての情報が借手に開示されるわけではなく、貸手が入手できる借手に関する情報は限られている[1]。この点において、貸手と運営会社の間に情報の非対称性が存在し、貸手は借手の信用リスク評価に関して運営会社による格付けに頼らざるを得ない。

第二に、貸手と運営会社の間に利益相反関係が存在する。貸手は、借手の信用リスクについて客観的な評価に基づく格付けを望む。一方、運営会社は出資金額の一部を手数料収入として受け取るため、借手の信用リスクを負うことなく、多くの出資金額を集めようとして客観的でない格付けを行う可能性がある。この点において、貸手と運営会社の間に利益相反関係が存在し、貸手は格付けを全面的に信頼することができない。これらの問題に対処するため、運営会社による格付けを考慮せずに借手の信用リスク評価を試みる研究がなされている[1][2]。

P2P Lending の金融商品としての側面に着目すると、投資意思決定を行う際に価格変動リスクだけでなく集中投資リスクを考慮する必要がある。価格変動リスクとは、投資した金融商品の価格が変動するリスクであり、集中投資リスクとは、対象の偏った投資によりデフォルト時の損失が拡大するリスクである。各ローンについて借手の居住地や借入目的が異なるため、それらについて集中投資リスクを考慮することで、リターンを確保しつつリスクを抑えた投資意思決定が可能になると考えられる。



図 1. 貸手、運営会社、借手の関係

そこで本研究では、P2P Lending において、運営会社による格付けを考慮することなく、価格変動リスクに加えて集中投資リスクを考慮した投資意思決定モデルを構築する。

### 2. 従来研究

#### 2.1. Serrano et al. [1]

Serrano et al. [1] は、P2P Lending における借手のデフォルト・リスクに着目し、デフォルトに影響を与える要因を明らかにしている。この研究では、デフォルトの有無を目的変数、運営会社による借手の評価、ローンの特徴、借手の特徴、借手の信用履歴および借手の財政状態に関する各指標を説明変数としてロジスティック回帰分析を行い、運営会社による格付け以外にも多くの項目がデフォルトに影響していると結論づけている。

#### 2.2. Y. Guo et al. [2]

Y. Guo et al. [2] は、P2P Lending において過去のローンに関するデータ (以下、トレーニング・データ) から現在のローン (以下、テスト・データ) のリターンとリスクを予測し、それをもとに投資意思決定を行うモデルである Instance Based Model (以下、IOM) を確立している。IOM は、デフォルト確率の推定、リターンとリスクの予測、ポートフォリオの選択の 3 段階で構成される。

第 1 段階では、トレーニング・データを用いてロジスティック回帰分析によりパラメータを推定し、テスト・データのデフォルト確率を予測している。回帰式を次に示す。

$$\text{logit}(\hat{p}) = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3 + \widehat{\beta}_4 x_4 + \widehat{\beta}_5 x_5 \quad (1)$$

ここで  $\hat{p}$  は目的変数であるデフォルト確率、 $x_1 \sim x_5$  は説明変数、 $\widehat{\beta}_0 \sim \widehat{\beta}_5$  はパラメータである。説明変数には、借手のクレジット・スコア、過去 6 ヶ月間の借手の信用照会件数、持ち家の有無、対所得負債比率を用いている。

第 2 段階では、各テスト・データについてトレーニング・データとのデフォルト確率の差 (以下、デフォルト確率距離) を求め、それをもとに計算したウェイトを用いてトレーニング・データのリターンとリスクから、テスト・データのリターンとリスクを予測している。リターンとリスクの予測式を次に示す。

$$\mu_i = \sum_{j=1}^n w_{ij} R_j \quad (2)$$

$$\sigma_i^2 = \sum_{j=1}^n w_{ij} (R_j - \mu_i)^2 \quad (3)$$

ここで  $\mu_i$  と  $\sigma_i$  はテスト・データ  $i$  のリターンとリスク、 $w_{ij}$  はテスト・データ  $i$  とトレーニング・データ  $j$  のデフォルト

確率距離に基づくウェイト,  $R_j$ はトレーニング・データ  $j$  の実績リターンである。

第 3 段階では, テスト・データについて第 2 段階で予測されたリターンとリスクをもとに, 目標リターンを達成しつつリスクを最小化するようなポートフォリオの選択を行っている。具体的には, 次の最適化問題を解くことで各テスト・データに対する投資比率を求めている。

$$\min \sum_i \lambda_i^2 \sigma_i^2 \quad (4)$$

$$\text{Subject to } R^* = \sum_i \lambda_i \mu_i, \sum_i \lambda_i = 1, \lambda_i \geq 0 \quad (5)$$

ここで  $\lambda_i$  はテスト・データ  $i$  に対する投資比率である。

この研究では, IOM と運営会社による格付けに基づく投資意思決定モデル (RBM) をリターン (平均収益率), リスク (標準偏差), 効率性 (シャープ・レシオ) について比較している。シャープ・レシオは, リスクに見合うリターンが得られているか, すなわち投資の効率性を示す指標であり, 次の式で表される。

$$\text{シャープ・レシオ} = \frac{\text{リターン} - \text{無リスク利子率}}{\text{リスク}} \quad (6)$$

この研究ではリターン, リスク, シャープ・レシオの全てにおいて IOM が優れていると結論づけている。

### 3. 提案モデル

#### 3.1. 提案モデルの概要

Serrano et al. [1] では, デフォルトに影響する要因を明らかにするにとどまり, 投資意思決定については扱っていない。また, Y. Guo et al. [2] では, 運営会社による格付けを考慮しない投資意思決定モデルを構築しているが, ポートフォリオの選択プロセスにおいて価格変動リスクのみを考慮しており, 集中投資リスクを考慮していないという問題点がある。そこで本研究では Y. Guo et al. [2] の IOM を参考に, 価格変動リスクに加えて集中投資リスクを考慮した投資意思決定モデルを構築する。具体的には, トレーニング・データからテスト・データのリターンとリスクを予測し, それをもとに価格変動リスクと集中投資リスクを考慮した投資意思決定を行うモデルを構築する。

考慮する集中投資リスクの種類は複数考えられるが, 投資家に開示されている情報である必要があるため, 個別銘柄, 借手の居住地, 借入目的, 借手の職業に関する集中投資リスクに限定される。このうち, 借手の職業については, 開示されている職業の分類レベルが異なっておりデータの質に問題があるため, 個別銘柄, 借手の居住地, 借入目的の 3 種類の集中投資リスクを考慮する。個別銘柄に関する集中投資リスクを考慮することで, 特定のローンがデフォルトした場合の損失を軽減することができる。また, 借手の居住地に関する集中投資リスクを考慮することで, 災害などにより特定の地域の借手が返済不能になった場合などにおける損失を軽減することができると考えられる。さらに, 借入目的に関する集中投資リスクを考慮することで, 住宅価格の下落により住宅ローンを借入目的とする借手が同時にデフォルトした場合などにおける損失を軽減できると予想される。

#### 3.2. 提案モデルの詳細

##### 3.2.1. デフォルト確率の予測

まず, テスト・データについてデフォルトの有無を目的変数, デフォルトに影響する各項目を説明変数としてロジスティック回帰分析を行い, 各説明変数のパラメータを推定する。回帰式を次に示す。

$$\text{logit}(\hat{p}) = \widehat{\beta}_0 + \sum_{k=1}^K \widehat{\beta}_k x_k \quad (7)$$

ここで  $\hat{p}$  はデフォルト確率,  $x_1 \sim x_K$  は説明変数,  $\widehat{\beta}_0 \sim \widehat{\beta}_K$  はパラメータである。Serrano et al. [1] においてデフォルトに影響する要因が明らかにされているため, 本研究では Serrano et al. [1] で有意とされた説明変数を用いる。具体的には, 運営会社による借手の評価, ローンの特徴, 借手の特徴, 借手の信用履歴, 借手の財政状態に関する 14 の指標を説明変数とする。

続いて, 推定されたパラメータ  $\widehat{\beta}_0 \sim \widehat{\beta}_K$  を用いてトレーニング・データとテスト・データ両方のデフォルト確率を予測し, 各テスト・データについてトレーニング・データとのデフォルト確率距離を求める。

##### 3.2.2. リターンとリスクの予測

次に, 各テスト・データについてトレーニング・データとのデフォルト確率距離に基づくウェイトを算出する。デフォルト確率距離によって類似度を測定するため, デフォルト確率距離が短い方が大きく, 長い方が小さくなるようにウェイトを決める必要がある。そこで, 本研究では Y. Guo et al. [2] と同様にカーネル関数を用いる。テスト・データ  $i$  とトレーニング・データ  $j$  のデフォルト確率距離に基づくウェイトは次のように表される。

$$w_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{p_i - p_j}{h}\right)} \quad (8)$$

ここで  $p_i$  はテスト・データ  $i$  の予測デフォルト確率,  $p_j$  はトレーニング・データ  $j$  の予測デフォルト確率,  $K\left(\frac{p_i - p_j}{h}\right)$  はカーネル関数,  $h$  はパラメータである。カーネル関数は次の式で表される。

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \cdot \quad (9)$$

カーネル関数を用いるために必要なパラメータ  $h$  の推定には Silverman[3] の手法を用いる。推定の起点  $h_0$  を定め,  $0.25h_0 \sim 1.5h_0$  の間で交差検証誤差  $CV(h)$  を最小化する  $h$  を求める。 $h_0$  と  $CV(h)$  は次の式で表される。

$$h_0 = \left(\frac{4}{3n}\right)^{1/5} \sigma \cdot \quad (10)$$

$$CV(h) = \frac{1}{n} \sum_{j=1}^n \left[ \frac{\sum_{i=1, i \neq j}^m K\left(\frac{p_i - p_j}{h}\right) R_j}{\sum_{i=1, i \neq j}^m K\left(\frac{p_i - p_j}{h}\right)} - R_j \right]^2 \cdot \quad (11)$$

ここで  $n$  と  $m$  はそれぞれトレーニング・データとテスト・データの数,  $\sigma$  は  $R_j$  の標準偏差,  $R_j$  はトレーニング・データ  $j$  の実績リターンである。算定されたウェイト  $w_{ij}$  をもとに次の式を用いてテスト・データのリターン  $\mu_i$  とリスク  $\sigma_i$  を予測する。

$$\mu_i = \sum_{j=1}^n w_{ij} R_j \quad (12)$$

$$\sigma_i^2 = \sum_{j=1}^n w_{ij} (R_j - \mu_i)^2 \quad (13)$$

### 3.2.3. ポートフォリオの選択

テスト・データのリターンとリスクをもとにポートフォリオを選択する。従来研究では価格変動リスクを目的関数とし、最適化問題を解くことで価格変動リスクを最小化させるポートフォリオを求めている。本研究では、価格変動リスクに加え集中投資リスクを考慮するため、最小化すべき目的関数を改良する必要がある。個別銘柄に関する集中投資リスクについては全ての目的関数において考慮し、また、借手の居住地と借入目的に関する集中投資リスクは3種類の目的関数を比較する。個別銘柄と借手の居住地に関する集中投資リスクを考慮した目的関数を次に示す。

$$\text{居住地} \quad \sum_i \lambda_i^2 \sigma_i^2 \sum_r (\sum_{i_r} \lambda_{i_r}^2)^2 \quad (14)$$

ここで $\sum_{i_r} \lambda_{i_r}^2$ はポートフォリオに含まれる借手の居住地が $r$ であるローンについて、その中で個別銘柄をできるだけ分散させることを意図しており、その上で $\sum_r (\sum_{i_r} \lambda_{i_r}^2)^2$ はポートフォリオに含まれる借手の居住地をできるだけ分散させることを意図している。価格変動リスクと集中投資リスクの合成量を低減するような目的関数とするために、価格変動リスクと集中投資リスクを掛け合わせる構造とする。次に、個別銘柄と借入目的に関する集中投資リスクを考慮した目的関数を示す。

$$\text{借入目的} \quad \sum_i \lambda_i^2 \sigma_i^2 \sum_p (\sum_{i_p} \lambda_{i_p}^2)^2 \quad (15)$$

目的関数の基本的な構造は「居住地」と同様である。最後に、個別銘柄、借手の居住地、借入目的の全てに関する集中投資リスクを考慮した目的関数を示す。

$$\text{居住地} \times \text{借入目的} \quad \sum_i \lambda_i^2 \sigma_i^2 \sum_r (\sum_{i_r} \lambda_{i_r}^2)^2 \sum_p (\sum_{i_p} \lambda_{i_p}^2)^2 \quad (16)$$

価格変動リスクに「居住地」と「借入目的」の集中投資リスクを掛け合わせる構造となっている。続いて、先述の目的関数を用いてポートフォリオの選択をする際に解く最適化問題を示す。

$$\min (14) \text{ or } (15) \text{ or } (16) \quad (17)$$

$$\text{Subject to } R^* = \sum_i \lambda_i \mu_i, \sum_i \lambda_i = 1, \lambda_i \geq 0 \quad (18)$$

ここで $\lambda_i$ はテスト・データ $i$ に対する投資比率である。(18)式は順に、予測リターンを目標リターンに一致させる制約条件、投資比率の合計は1という制約条件、各投資比率は負になり得ないという制約条件を表している。

## 4. 検証

### 4.1. 検証方法

本研究の検証は、3.2.3で提案した3種類の目的関数を用いた投資意思決定モデルと従来研究で提唱されているIOM、運営会社による格付けに基づく投資意思決定モデルであるRBMの5つを比較することで行う。比較する

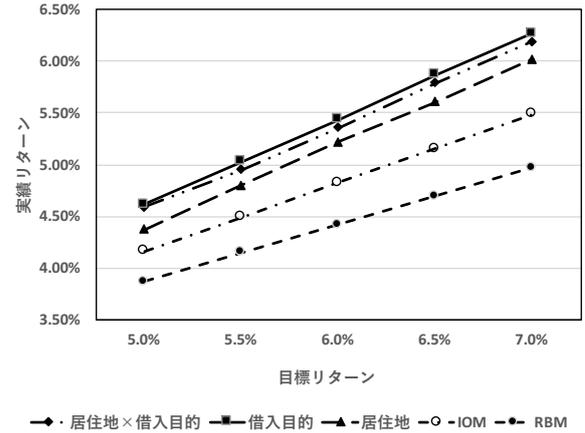


図 2. リターンの検証結果

指標は Y. Guo et al. [2] を参考にリターン (平均収益率)、リスク (標準偏差)、シャープ・レシオの3種類を用いる。具体的には、5%~7%の5種類の目標リターンを設定し、各投資意思決定モデルについて比較する指標を測定する。

### 4.2. 使用データ

本研究では、アメリカの大手 P2P Lending 運営会社である Lending Club が投資家向けに開示しているデータを用いる。検証には 2013 年 12 月に成立した 36 ヶ月ローン 1587 件をトレーニング・データ、テスト・データの2種類に分割して使用する。ポートフォリオの選択において使用する無リスク金利は Y. Guo et al. [2] と同様に、アメリカ国債の利回りを参考に 2.5%とする。

### 4.3. 検証結果

#### 4.3.1. リターンに関する検証

まず、各投資意思決定モデルのリターン (平均収益率) について検証を行う。検証結果を図 2 に示す。本研究の全ての提案モデルが IOM を上回っていることから、集中投資リスクを考慮したモデルの方が価格変動リスクのみを考慮したモデルと比較してリターンが大きくなるといえる。その理由としては、借手の居住地や借入目的を分散させるために、利率が高い (価格変動リスクが大きい) ローンにも投資する意思決定を行うためであると考えられる。また、運営会社による格付けを考慮しない投資意思決定モデル (本研究の提案モデルと IOM) が RBM を上回っていることから、運営会社による格付けを考慮しない投資意思決定モデルのほうが大きなリターンを得ることができるといえる。なお、全ての目標リターンについて実績が目標を下回っているが、その原因は返済が遅延していて現在未回収のローンについて、データ上はこれをデフォルトとして扱っている点にあると考えられる。

#### 4.3.2. リスクに関する検証

次に、各投資意思決定モデルのリスク (標準偏差) について検証を行う。検証結果を図 3 に示す。全体としては、全ての投資意思決定モデルについて、目標リターンが大きくなるほど、リスクも大きくなるという結果となった。本研究の全ての提案モデルが IOM よりもリスクが大きいことから、集中投資リスクを考慮するとリスクが小さくなるという仮説に反し、リスクが大きくなっている。

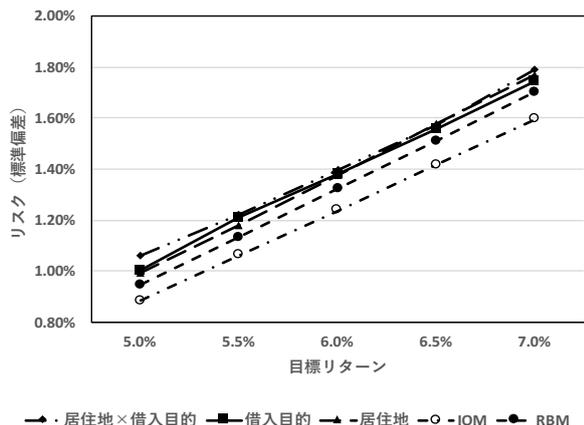


図 3. リスクの検証結果

リスクが増加する理由としては、借手の居住地や借入目的を分散させるために価格変動リスクが大きいローンにも投資する意思決定を行うためであると考えられる。

#### 4.3.3. 投資の効率性に関する検証

最後に、各投資意思決定モデルの効率性（シャープ・レシオ）について検証を行う。検証結果を図 4 に示す。本研究の全ての提案モデルが IOM よりもシャープ・レシオが大きいことから、集中投資リスクを考慮したモデルのほうが価格変動リスクのみを考慮したモデルよりも高効率な投資が可能になるといえる。また、運営会社による格付けを考慮しない投資意思決定モデル（本研究の提案モデルと IOM）が RBM を大きく上回っていることから、運営会社による格付けを考慮しない投資意思決定モデルのほうが高効率な投資が可能になるといえる。

#### 4.4. 考察

##### 4.4.1. 集中投資リスクを考慮する効果

集中投資リスクを考慮する効果を明らかにするために、本研究の提案モデルと IOM の比較を行う。本研究の検証により、集中投資リスクを考慮する投資意思決定モデルは、価格変動リスクのみを考慮した投資意思決定モデルと比較して、リターン、リスクともに増加するが、リターンの増加と比較してリスクの増加が抑えられ、投資の効率性は高まるという結果が得られた。リターンとリスクがともに増加する理由としては、集中投資リスクを考慮すると、借手の居住地や借入目的を分散させるために、価格変動リスクが大きいローン、すなわち利率が高いローンにも投資する意思決定がなされるからであると考えられる。リターンの増加と比較してリスクの増加が抑えられる理由としては、価格変動リスクが大きいローンに投資した場合であっても、集中投資リスクを考慮することで借手の居住地や借入目的が分散されているため、集中投資リスクを考慮せずに価格変動リスクが大きいローンに投資した場合よりもリスクの増加が少ないからであると考えられる。

##### 4.4.2. 考慮する集中投資リスクの比較

考慮する集中投資リスクの比較を行うために、本研究の提案モデル間の比較を行う。本研究の検証により、

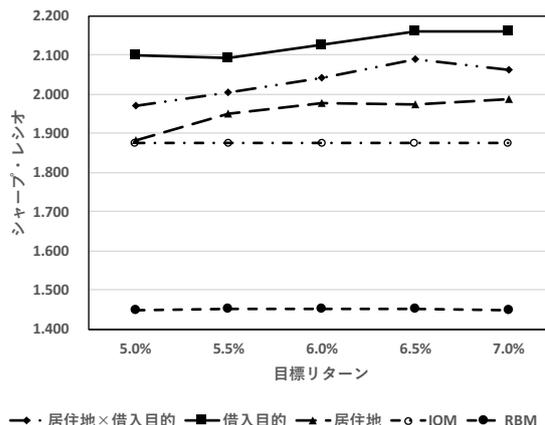


図 4. シャープ・レシオの検証結果

リターンについては「借入目的」が最も大きく、「居住地」が最も小さく、「借入目的×居住地」が 2 つのモデルの間となった。リスクについては 3 種類のモデルで大きな違いは見られず、投資の効率性についてはリターンと同様の結果となった。リターンと効率性について、「借入目的」が「居住地」を上回った理由としては、居住地の 50 種類（アメリカの州の数）は借入目的の 11 種類を大きく上回るため、借手の居住地を分散させるために、予測リスクに見合わない予測リターンのローンについても、投資を行う意思決定がなされたからであると考えられる。

複数の集中投資リスクを考慮した場合の影響について考察する。リターンと効率性の全てについて、「借入目的×居住地」が「居住地」と「借入目的」の中間の結果となった。その理由としては、「借入目的×居住地」は、借手の居住地に関する集中投資リスクと、借入目的に関する集中投資リスクを掛け合わせているため、両者の影響を均等に受けるからであると考えられる。

#### 5. おわりに

本研究では、P2P Lending において、価格変動リスクに加え集中投資リスクを考慮した投資意思決定モデルを構築し、価格変動リスクのみを考慮したモデルと比較して、投資の効率性を高めることができた。今後の課題としては、本研究の目的関数では価格変動リスクと集中投資リスクを掛け合わせるにより同等に扱っているが、より効率的な投資意思決定モデルを目指すために、その扱いの程度を変化させることが挙げられる。

#### 参考文献

- [1] C. Serrano-Cinca, B. Gutierrez-Nieto and L. Lopez-Palacios : "Determinants of Default in P2P Lending", <https://doi.org/10.1371/journal.pone.0139427>, (2015)
- [2] Y. Guo, W. Zhou, C. Luo, C. Liu, H. Xiong : "Instance-based credit risk assessment for investment decisions in P2P lending", *European Journal of Operational Research* 000, pp. 1-10 (2015)
- [3] B.W.Silverman : *Density Estimation for Statistics and Data Analysis*, (1986)