Gamma 回帰によるデフォルト債権回収額の推計

今井 健太郎, 尾藤 剛[1]

本研究では、国内の金融機関におけるデフォルト債権回収の実データを用いて、回収額の推計モデルの構築を行った。従来の回収推計モデルには推計対象を回収率とするものが多く、0%と 100%にサンプルが集中する bimodal 型の分布を扱うための様々な手法が提案されてきた。本研究では、推計対象を回収率ではなく回収額とすることで、bimodal 型分布の推計を回避した。モデルの有効性の検証にあたっては、被説明変数をそれぞれ回収額と回収率とする二種類の推計モデルを構築し、平均二乗誤差などを用いて比較を行った。この結果、推計対象を回収額とした Gamma 回帰モデルの有効性を確認できた。

1. はじめに

1.1. 回収額の推計

金融機関が信用リスク（貸出資産の損失可能性）を評価する際には、予想損失額（EL:Expected Loss）の推計を通じて行うのが通常である。EL は、デフォルト確率（PD:Probability of Default）、デフォルト時損失率（LGD:Loss Given Default）、およびデフォルト時貸出残高（Exposure at Default）の 3 つのパラメータにより、EL=EAD×PD×LGD の形に分解できる。また、LGD の代わりに回収率（RR:Recovery Rate）を用いると、RR=1-LGD により、EL=EAD×PD×(1-RR)とあらわすことができる。このうちの PD については、統計的手法に基づく様々な推計モデルが多くの金融機関にて実用化されている一方、LGD（RR）については、推計モデルの理論研究が一部で進んでいるものの、ことに長期間の回収活動による回収額を推計対象とするワークアウト LGD についての研究事例は、世界的にも数少ない状況にある。このため、PD 推計モデルにおける logistic 回帰のような実務上のデファクトスタンダードも、LGD 推計においてはいまだ確立されていない。

このため実務上は、LGD（RR）の期待値に代えて、担保や保証といった事前に評価額が定まっている保全額の割合（保全率）を用いることが多い。このとき、担保・保証によらない回収（信用回収）を推計上はゼロとするため、結果として回収額が過小推計となり、EL も過度に保守的な評価になることが、金融機関の信用リスク管理における技術面での大きな課題となっている。

LGD 推計を困難にしている要因としては、個別金融機関だけでは統計分析に耐えるデータ件数の確保が難しいこと、デフォルト後に正常復帰する貸出先の扱い、根担保・根保証に代表される日本特有の貸出慣行、時に 5 年・10 年もの長期にわたる回収期間の長さなどが挙げられる。また先行研究の多くが示すように、実際の回収率の頻度分布が 0%と 100%に集中する bimodal 型の形状をとることも、モデルによる推計を難しくしている要因の 1 つと考えられる。

Asarnow and Edwards(1995)は、米国シティバンクのデータを用いて回収率の分布が bimodal 型であることを示している。また、Felsovalyi and Hurt(1998)も、ラテンアメリカのシティバンクのデータから同様に bimodal 型を示している。Franks et al.(2004)は、英仏独の 10 銀行における 1984~2003 年の回収データが bimodal 型の分布に従うことを示している。

国内の先行研究においても回収率が bimodal 型を示す報告が見られ、伊藤・山下(2007)は、信用保証協会の求償債権の累積回収データから、「累積回収率が 0%の付近と累積回収率 100%の付近で標本が集中」する bimodal 型であることを示している。また、尾木他(2012)は日本政策金融公庫のデータを分析

[1] 日本リスク・データ・バンク株式会社 E-mail:db@riskdatabank.co.jp
本稿に示されている意見は筆者の個人的なものであり,日本リスク・データ・バンク株式会社の公式見解を示すものではありません

し、特に保全のない債権については回収率が 0%と 100%に集中する bimodal 型であることを示している。

一方で、川田・山下(2012)は、3 つの銀行から集めたデータにて回収終了時点の LGD を分析し、「双峰型の分布（LGD が 0%付近と 100%付近に集中する）ものとは異なっている」として、「LGD=0(%)付近にデータが集中して」存在し、「LGD=100%という例はまれ」と報告している。

bimodal 型の分布に対して、線形回帰にて回収率を推計するモデルを構築すると、実際の回収率が基本的には 0%から 100%の範囲をとるにもかかわらず、モデルは負の回収率をとりうる問題が発生する。また、0%と 100%に集中する bimodal 型の特徴的な分布を表現するためには、”Due to the boundaries of the dependent variable, one cannot use ordinary least square(OLS) regression”（Dermine and Carvalho(2006)）、「回収率のヒストグラムをよく表すような確率分布を当てはめる必要がある」（森平(2009)）といった指摘がなされている。

この問題に対応するため Gupton, Stein(2005)は、Moody's 社の LossCalc™ V2 で採用した方法として、”These recoveries, however, are not normally distributed. An alternative that better approximates the prices in our dataset is Beta distribution.”、すなわち Beta 分布による近似を試みている。なお、同社の後継モデルと見られる LossCalc™ V3.0 は、”Simple Approach”として推計値が 0 から 1 の範囲に入るような”final transform”を実施した OLS による推計を行っている（Douglas, Irina(2009)）。他には、Grippa, Iannoti, Leandri(2005)が F-logit (fractional logit)回帰を、Dermine, Carvalho(2006)が log-log 変換を試みている。

日本国内では、伊藤・山下(2008)が、0%、0%超 50%未満、50%超 100%未満、100%の 4 つのカテゴリに分けた順序ロジットモデルを提唱している。森平(2009)は、経験分布は 0%と 100%の範囲に限定されるが、潜在的な回収率は負の値や 100%超の値もとりうることを仮定した両側 Tobit モデルを提唱している。川田・山下(2012)は、正常復帰確率モデル、毀損発生確率モデル、LGD モデルを多段的に組み合わせ、LGD モデルには logit 変換した回収率を線形回帰で推計する方法を提唱している。また尾木他(2012)は、回収率が 0%、0%超 100%未満、100%の 3 クラスに分け、順序ロジットモデルによる LGD 推計をおこなっており、そこでは回収源泉に着目し、債務と担保の状況から債務を 100%カバーした担保付、0%超 100%未満をカバーした一部担保、無担保無保証の 3 種類の保全別にモデルを推計している。

本研究で主として扱う回収額推計モデルは、被説明変数を従来の回収率に代えて回収額（Recovery Amount）とすることを特徴としている。これにより、0%と 100%にサンプルが集中する bimodal 型の分布の問題を回避できるほか、被説明変数と説明変数それぞれに金額を用いることで、両者の関係を後述する乗算の形式で表せるのも、実務で一般的な「掛け目」の発想に近いというメリットを有している。

債権回収の分野で被説明変数を回収額（または損失額）にしたモデルは、筆者の調査範囲内ではみつけられなかったが、損害保険の請求額の推計を中心とする保険数理の分野では、率ではなく額を推計対象としたモデルが用いられている。推計モデルには一般化線形モデル（GLM:Generalized Linear Model）のうちの Gamma 回帰モデルが欧州を中心に使われており、Anderson, et al.(2007)によると”Today GLMs are widely recognized as industry standard method for pricing passenger auto and other personal lines and small commercial lines insurance in the European Union and many other markets.”とされている。

本研究では、GLM の Gamma 回帰モデルの枠組みを用いて回収額推計モデルを構築するのとあわせて、回収率推計モデルを含む数種類のモデル構築を行い、実績値と推計値の誤差や相関によって性能を比較することで、Gamma 回帰型モデルが推計精度の上でも有効であることを報告する。

1.2.　回収源泉別の推計

本研究の回収額推計モデルのもう一つの特徴としては、保証および担保による回収（保全回収）と、保証・担保によらない回収（非保全回収、または信用回収）を区別し、非保全回収のみをモデルの推計対象としたことが挙げられる。本研究においては、回収源泉を区別せず全ての回収額（回収率）を一括で推計対象とする従来型の手法をシングル（single）型のアプローチ、回収源泉別に分けてそれぞれ推計したものを最後に合算して回収額（回収率）とする手法をマルチ（multi）型のアプローチとそれぞれ呼び分けることとする[2]。

図 1 は回収源泉別に、担保・保証としての評価額と実際の回収額との関係を示した散布図である。図左側は、保証、および換金性の高い預金担保、手形・売掛債権担保の合計額を「優良担保」として合算して横軸に、この優良担保を源泉として実際に回収した金額を縦軸に、それぞれ対数プロットしたものである。また、図中央は優良担保以外の有価証券や不動産、その他担保の合計額を「一般担保」として横軸に、この一般担保を源泉に実際に回収した金額を縦軸にそれぞれプロットしている。そして図右は、デフォルト時貸出残高（EAD）から、優良担保と一般担保を差し引いた「非保全額」を横軸に、優良担保と一般担保以外による回収額を「信用回収」として縦軸に、それぞれプロットしたものである。なお、回収額が 0 円の場合は便宜上、0.1 円として 10^{-1} にプロットしている。一方、図 2 は回収源泉別の回収率を、左から順に優良担保、一般担保、信用についてそれぞれヒストグラムにて集計している。

両図から優良担保回収、一般担保回収、信用回収の傾向がそれぞれ異なるのは明らかである。たとえば優良担保の場合、回収率は 0%もしくは 100%に 4 割前後が集中しているが、これは優良担保の性質上、金融機関が回収行動を実行すれば 100%、しなければ 0%ということを意味しており、これを予測するならば二項ロジットモデルを用いるのが適当と考えられる。一方で実行するかしないかの判断は、最終的には金融機関の自由意志によるところであり、そもそも推計ロジックへの実務上の必要性は想定しがたい。また一般担保の場合も、回収率 0%に 6 割程度が集中して存在し、次いで 100%に 1 割程度が存在している。こちらも回収行動の実行有無については二項ロジットモデルが適当ではあるものの、実務上の必要性については優良担保と同様である[3]。一方、信用回収の部分は、0%近傍と 100%にサンプルが集中する bimodal 型ではあるものの、その割合は 1 割強であり、0%と 100%以外に広く存在しているほか、100%超にも一定の割合で存在していることが分かる。

したがって、回収源泉別に回収額の経験分布の特徴を生かすことができるマルチ型のアプローチが回収の推計には適切であると考え、本研究の基本的な推計手法として採用した。特に、信用回収のモデル化の可能性に注目し、また bimodal 型分布の問題を回避できる回収額を被説明変数とすることで、推計精度の向上を図るものとする[4]。

2.　統計モデル

2.1.　GLM の概略

McCullagh and Neldr(1989)に従い、まず線形モデル(Linear Model)を考える。

$\boldsymbol{Y}$ を構成する n 個の成分は、互いに独立で期待値 μ を持つと仮定する。μ はパラメータ $\beta_1,\ldots,\beta_p$ と共変量 $\boldsymbol{x}$ を用いて以下のように記述できる。

---

[2] 100%カバーの担保付債券、一部担保付債券、無担保無保証債券でモデルを分けた尾木他[2012]の手法はこれに近いが、一部担保付のモデルでは回収源泉別に回収額を 1 つにまとめて推計しており、本分析における分類ではシングル型となる。

[3] 担保・保証の価値を推計するモデルの必要性を否定するものではない。

[4] 実務上は、担保・保証によらない回収（=信用回収)を一律にゼロとみなす手法が現時点では多く採用されており、この部分を推計するモデルの構築が課題となっている。

$$\boldsymbol{\mu} = \sum_{1}^{p} \boldsymbol{x}_j \beta_j$$

μ を n×1、X は n×p、β は p×1 の行列とすると $\mu = X\beta$ であり、線形モデルの場合、Y は独立な正規分布で分散 σ^2 は一定であるという仮定のもとで、次の式を導くことができる。

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} \; where \; \boldsymbol{\mu} = \boldsymbol{X\beta}$$

上記をまとめると、線形モデルには以下の 3 つの仮定がある。

(LM1) Random component： Y の成分は互いに独立で正規分布に従い、分散 σ^2 は均一である。

(LM2) Systematic component：共変量 $\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}$ を組み合わせて、線形予測子(linear predtictor) $\boldsymbol{\eta}$ が次の式で与えられる。

$$\boldsymbol{\eta} = \boldsymbol{X\beta}$$

(LM3) Link Function：Random component と Systematic component とはリンク関数で結びついており、それは次のような恒等式である。

$$E[\boldsymbol{Y}] \equiv \boldsymbol{\mu} = \boldsymbol{\eta}$$

これに対し一般化線形モデル（GLM）は線形モデルの LM1 と LM3 を拡張したものである。

(GLM1) Random component：Y の成分は互いに独立で、指数型分布族[5]に含まれる分布であればよい。

(GLM2) Systematic component：X を組み合わせることで、線形予測子（linear predtictor） $\boldsymbol{\eta}$ が次の式で与えられる。

$$\boldsymbol{\eta} = \boldsymbol{X\beta}$$

(GLM3) Link Function：Y と μ は、微分可能で単調な関数 g をリンク関数として用いて、次のように関連付けられる[6]。

$$E[\boldsymbol{Y}] \equiv \mu = g^{-1}(\boldsymbol{\eta})$$

または

$$g(E[\boldsymbol{Y}]) \equiv g(\boldsymbol{\mu}) = \boldsymbol{\eta}$$

である。

2.2. Gamma 分布

実際の回収額の分布形状は 0 円から 100 億円を超えるものも存在する右に裾が長い分布となっているため（図 3）、本研究では回収額推計モデルとして、指数型分布族の 1 つである Gamma 分布を仮定した。Gamma 分布の確率密度は以下のように定義され、その値は 0 を超える連続量である。

$$\mathrm{f(y)} = \frac{r^s}{\Gamma(s)} y^{s-1} \exp(-ry)$$

s を shape パラメータ、r（または 1/r）を rate パラメータ（または scale パラメータ）と呼ぶ。

2.3. モデル式

平均的な回収額 μ_i を、ある財務指標 z_i の単調増加関数として、以下の式で表現する。

$$\mu_i = Az_i^b$$

---

[5] 確率変数 Y がパラメータ θ を持つとき、確率分布が以下の式で記述されるなら、その分布は指数型分布族に属する。

$$\mathrm{f(y;\theta)} = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

例えば a(y)=y, b(θ)= μ/σ^2, c=$-2\mu/\sigma^2-1/2*\log(2\pi\sigma^2)$, d=$-y^2/2\sigma^2$ の場合、f(y; θ)=$1/(2\pi\sigma^2)^{1/2} \exp[-1/2\sigma^2(y-\mu)^2]$となり、正規分布の確率変数である。

$A = exp(a)$とおいてから、全体を指数関数でまとめる。

$$\mu_i = exp(a)\, z_i^b = exp(a + b \log z_i)$$

さらに両辺を対数にとる。

$$log\, \mu_i = a + b\, log\, z_i$$

ここで$x_i = logz_i$とすれば線形予測子$\mu = a + bx_i$ で、リンク関数gはlogになっていることが分かる。

上記を踏まえ、被説明変数を回収額、誤差項がGamma分布、リンク関数をlogとする推計モデルを構築し、これをMolde1とした。本研究では、この他に比較対象として7つのモデルを構築した。表 1に構築したモデルの概略を示す。

表 1 構築モデル

No.	モデル式	誤差構造 ε	被説明変数 Y
1	$\mathrm{log(E(Y))} = \mathrm{a} + \sum b_j x_j + \mathrm{log}\xi$	Gamma	Recovery Amount
2	$\mathrm{log(E(Y))} = \mathrm{a} + \sum b_j x_j + \mathrm{log}\xi$	Normal	Recovery Amount
3	$\mathrm{E(Y)} = \mathrm{a} + \sum b_j x_j$	Normal	Recovery Amount
4	$\mathrm{E(log(Y))} = \mathrm{a} + \sum b_j x_j$	Normal	Recovery Amount
5	$\mathrm{E(Y)} = \mathrm{a} + \sum b_j x_j$	Normal	Recovery Rate
6	$\mathrm{E(Y)} = \mathrm{a} + \sum b_j x_j$	Normal	Min(Recovery Rate, 1)
7	$\mathrm{E(Y)} = \mathrm{a} + \sum b_j x_j$	Normal	Logit(Recovery Rate)
8	$\mathrm{E(Y)} = \mathrm{a} + \sum b_j x_j$	Normal	Beta(Recovery Rate)

No.1は被説明変数Yが回収額で、リンク関数にlog、誤差構造にGamma分布を仮定したものである。No.2も同様に被説明変数Yは回収額で、リンク関数にlog、誤差構造には正規分布を仮定したもので、この2つは最尤法でパラメータを推定し、説明変数xには対数化したもののみを使った。なおξはデフォルト時の貸出残高で、説明変数と同様に対数を取っているが、係数は1であり、この項はoffset項と呼ばれている。

No.3からNo.8は線形回帰によるものだが、GLMの枠組みで説明するならば、誤差項に正規分布、リンク関数には恒等式を用いたことになる。これらも最尤法でパラメータを推定するが、結果は最小二乗法によるものと同じである。No.3は回収額を、No.4は対数化した回収額を、No.5は回収率を、それぞれ被説明変数Yにして線形回帰したものである。

No.6は100%超の回収率を100%に丸め処理したものを被説明変数に用いた。

No.7は回収率RRをlog((1-RR)/RR)でロジット変換したものであるが、回収率が100%超の場合に変換ができないため、99.99%に置き換えて変換している。またNo.8は回収率RRをベータ変換したもので、ベータ分布の形状パラメータにはモデル構築用全データから推計したものを用いた。回収率が100%超の取り扱いはNo.7と同様である。No.3からNo.8の説明変数xには対数化したものに限定せず、対数化しないものも候補に加えた。

3. 使用したデータ

分析に用いたデータは、日本リスク・データ・バンク株式会社（RDB）の「デフォルト債権回収データベース」から抽出した。同データベースには日本の 19 の金融機関（2014 年 1 月末時点）が参加しており、各金融機関が拠出した債務者単位で約 4 万 5 千件の回収データを格納している。

本研究では、債務者単位での回収額／率を推計するモデル構築方法の比較を目的としており、具体的には「金融機関が本格的に回収行動に移った後の」回収額／率を推計対象とした。いわゆる「デフォルト（債務不履行）後」とは異なる条件としたのは、回収額／率が、債務者の属性によらず金融機関の意思決定に左右される要素を極力排除するためである。主なものとしては以下のケースが挙げられる。

- 正常先に復帰して債務不履行を解消（実際に回収できた金額は 0 円でも回収率は 100%）
- 条件変更を実施して長期間据え置き（回収率は 0%）
- バルクセール（債権一括売却）による処分（回収率は個別債務者の属性に依存しない）

実際に行ったデータの絞り込み過程は以下の通りである。

(1) 実質破綻先以下に限定

実データによると、債務者の信用状態が実質破綻先以下になった後に、本格的な回収行動が実行に移されることが多く見られた。反対に要管理先や破綻懸念先の段階では、金融機関は回収行動をとらず「様子見」をしていることも多い。そこで本研究では、実質破綻先以下のデータに対象を限定し、さらに実質破綻先に該当してから 36 か月間の回収金額が 10 万円以下のような少額の場合には、金融機関はまだ本格的な回収行動に入っていないと考え、これもモデル構築データから除外した。

(2) 正常復帰していない

モデルの推計対象は実際に回収できた額／率であるが、債務者が正常先に復帰した場合、実際の回収金額が 0 円の場合でも全額回収とみなすことが多い。一方で、正常化の有無は、金融機関側の意思決定や規制上のルール変更に大きく影響を受けるため、ここでは正常復帰した先を構築データから除外した。

(3) バルクセールでない

(4) デフォルト基準時点における貸出残高が 1 百万円以上

最終的なモデル構築用データは N=4,367 件となった。

4. モデル構築

4.1. 説明変数の候補

モデル構築のための説明変数は、デフォルト直前の決算期の財務指標、デフォルト時点で得られる貸出情報、属性情報から 150 程度を候補指標とした。

表 2 使用した説明変数候補

類型	具体的な変数名
財務情報	売上高、総資産、自己資本、有利子負債、現預金比率、自己資本比率、など
貸出情報	貸出残高、貸出シェア、規模ダミー、メイン行ダミー、保全状態、など
属性情報	業種、金融機関クラス、など

規模ダミーは、保全部分を含む全貸出残高を基準に、1 億円未満、10 億円未満 1 億円以上、10 億円以上の

3 つのクラスのカテゴリ変数である。また保全状態は、優良担保なし且つ一般担保なし（無担保無保証）、優良担保のみあり、一般担保のみあり、優良担保あり且つ一般担保ありの 4 区分のクラス変数とした。金融機関クラスは、金融機関による回収率の水準の違いに応じて、3 つのグループに集約したクラス変数としたものである。

4.2. モデル構築手順

表 1 の 8 種類全てについてモデルの推計を試みたが、説明変数を全てのモデルで共通にした場合、特定のモデルが不利になる恐れがある。そこで、最初に 3 つの共通変数モデルを推計したのち、モデルごとに精度を改善するように変数を 3 つ追加し、最終的に 6 変数でのモデルを構築した。具体的な手順は、まずカテゴリ変数である規模ダミー、メイン行ダミー、業種、貸出シェアダミー、保全状態、金融機関クラスの 6 つのカテゴリ変数でモデル推計を行い、対数尤度が高かった規模ダミー、保全状態、金融機関クラスの 3 つの変数を全モデルの共通変数とした。次に、財務情報、および貸出情報の候補変数から対数尤度が高くなるように 3 つの指標を追加した。各モデルに採用された指標名[9]、各指標の p 値、対数尤度を切片項のみモデル、共通変数のみのモデル、最終モデルの 3 通で計算し表 3 にまとめた。

Model1、Model2 の規模ダミーはいずれも p 値が 10%水準を上回ったが、これは規模の効果として offset 項(ξ)を入れたためと考えられる。また、Model6～8 も規模ダミーは有意とならなかった。さらに Model6 は保全状態ダミーのいずれも有意とはならなかった。それ以外の変数については、いずれも 5%以下の有意水準であった。

4.3. モデルの評価

被説明変数がモデルによって回収額または回収率で異なり、さらにそれらを変換しているため、決定係数、対数尤度や AIC などでは、単純にはモデル間の能力比較ができない。ここでは、平均二乗誤差（MSE）、平均絶対値誤差（MAE）、および相関の 3 指標にて評価することとした。また、回収額推計モデルの場合には、推計回収額を EAD で割ったものを推計回収率とし、逆に回収率推計モデルの場合には、推計回収率に EAD をかけたものを推計回収額とした。なお被説明変数を対数変換、logit 変換、beta 変換したものの場合には、それぞれ再変換を行い回収額、回収率に戻している。

具体的な比較手法としては、モデルの精度と頑健性をみるために 10 交叉検証法を用いた。10 交叉検証とは、データセットを 10 分割し、そのうちの 9 つのサブセットでモデル構築を行い、残る 1 つのサブセットで MSE、MAE、相関の各計算を行い、これを 10 回繰り返す方法である。各統計量が安定して高いモデルが、精度、頑健性の双方にすぐれると評価できる。なお、モデル構築の際には、表 3 の結果から有意水準が 5%以下にならなかった説明変数（たとえば Model1 であれば規模ダミー）を外した。10 交差検証によって得られた MSE,MAE を図 4 に、Spearman の順位相関を図 5 にそれぞれ示した。

図 4 のパネル左側 2 列は回収額の誤差、右側 2 列は回収率の誤差の各評価結果である。パネル最上段は EAD の規模を考慮しない全データでの結果をまとめている。

回収額の誤差の評価結果によると、MAE では Model1、2 の値が低く、また MSE では Model2、3、次いで Model1 の結果が低くなっており、回収率を被説明変数とした Model5～8 の結果よりも推計の誤差が小さいモデルであることが分かる。一方、回収率の誤差の評価結果は、MAE では Model6～8、次いで Model1 の結果が低く、MSE では Model1 と 6 が同水準で低い値となっている。

---

[9] 指標名の後ろの添え字は変数を変換した方法を示すものである。1 は基準化したのちに対数変換、2 は基準化、3 は全ての変数が 0 超になるようにした後に対数変換、4 は 0 超になるように変換したものである。

回収額を被説明変数にしたModel1が、率に直してもMSE、MAEが小さくモデル性能は良好といえる一方で、回収率を被説明変数にしたModel5～8は、額に直すと誤差が大きくなりやすいのが分かる。

2段目から4段目はEADの規模別に集計したもので、回収額のMAE、MSEの結果から、回収率推計モデルのModel5～8は、規模の小さい先においては値が小さく、Model1よりも良い結果が得られている。一方で規模の大きい先ではModel1が優位となる。回収率のMAE、MSEの結果においては、規模によって傾向に明確な差があるとはいえず、Model1はModel5～8と同水準であった。

図 5の順位相関係数の結果をみると、回収額に対して最も高い相関を示したのはModel1、次いでModel4、6であった。また、回収率に対する相関ではModel7、8が高い相関を示すものの、Model1もそれに準ずる水準となった。

5. 結論

本研究では、RDB のデフォルト債権回収データベースの実データを用いて、回収額または回収率推計モデルの構築手法の比較を行った。先行研究では回収率を推計するモデルが多かったが、本研究では回収額を推計するモデルを構築して性能を比較した結果、回収額を Gamma 回帰で推計するモデルが、回収額／率それぞれの推計で安定的に高い性能を示すことを確認した。これは、回収率を推計する従来型のモデルでは、回収率が 0%と 100%に集中する bimodal 型の分布を推計するために、複雑なモデル化や変数変換を行うのに対して、額を推計対象としたことにより、Gamma 分布に従うという仮定によるモデル推計を簡素な枠組みにて実現した成果と考える。

また Gamma 回帰モデルは、特に EAD の大きい先における回収額／率の推計で、他の推計モデルと比べて良好な性能を示すことも確認できた。これも額を推計対象にしたことの効果と考えられ、回収金額の絶対的な大きさが重要な金融機関実務において望ましいモデルの性質といえよう。

残された課題としては、Gamma 回帰モデルにて回収額が過大推計となっている債務者が多数存在している点であり、これらについてはモデル推計精度に改善の余地が残る。現時点では、ゼロ強調型（Zero Inflated Gamma）の適用が一つの手段として考えられる。

6. 参考文献

Anderson, Feldblum, Modlin, D.Schirmacher, E.Schirmacher, Thandi(2004)"A Practitioner's Guide to Generalized Linear Models", *Casualty Actuarial Society Discussion Paper Program*

Annette J.Dobson(2008)[著], 田中豊, 森川敏彦, 山中竹春, 冨田誠[訳]『一般化線形モデル入門 原著第２版』,共立出版

Asarnow and Edwards(1995) "Measuring LGD on commercial loans: an 18-year internal study", *Journal of Commercial Lending* , Vol. 77, No. 7 ,pp.11-23

Dermine and Carvalho(2006) "Bank loan losses-given-default: A case study", *Journal of Banking & Finance*, pp.1243-1291

Dwyer, Korablev(2009) "Moody'z KMV LossCalc™" V3.0" , *Moody's Analytics, manuscript*

Felsovalyi and Hurt(1998) "Measuring Loss on Latin American Defaulted Bank Loans: A 27-Year Study of 27 Countries",*Journal of Lending & Credit Risk Management*, Vol. 81, No. 2, (October 1998), pp. 41-46.

Franks, Servigny, Davydenko(2004) "A comparative analysis of the recovery process and recovery rates for private companies in the UK, France and Germany", *Standard and Poor's Risk Solutions*

Grippa, Iannoti, Leandri(2005)" Recovery rates in the banking industry: stylised facts emerging from the Italian experience" in *Recovery risk : the next challenge in credit risk management* edited by Edward, Altman, Resti, Sironi, Risk Books

Gupton, Stein(2005)" LossCalc v2: Dynamic prediction of LGD", *Moodys KMV Investors Services*

McCullagh and Nelder(1989)"Generalized Linear Models, Second Edition",Chapman & Hall

伊藤有希, 山下智志(2007)「中小企業に対する債権回収率の実証分析」金融庁 FSA リサーチ・レビュー 2007 第 4 号

上野大(2006)「バーゼルⅡにおける LGD の扱い」『債権回収率・LGD モデルシンポジウム』

尾木研三, 戸城正浩, 枇々木規雄(2012)「小企業向け保全回収率モデルの構築と実証分析」JAFEE 冬季大会予稿集、pp.33-44

川田章広, 山下智志(2012)「回収実績データに基づく LGD の要因分析と多段階モデルによる LGD および EL 推計」金融庁金融研究センター ディスカッションペーパー DP2012-6

久保拓弥(2012)『データ解析のための統計モデリング入門』,岩波書店

尾藤剛(2011)『ゼロからはじめる信用リスク管理』,きんざい,p.297

三浦翔, 山下智志、江口真透(2010)「内部格付手法における回収率・期待損失の統計型モデル : 実績回収率データを用いた EL・LGD 推計」金融庁 FSA リサーチ・レビュー2010 第 6 号

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,pp.137-158

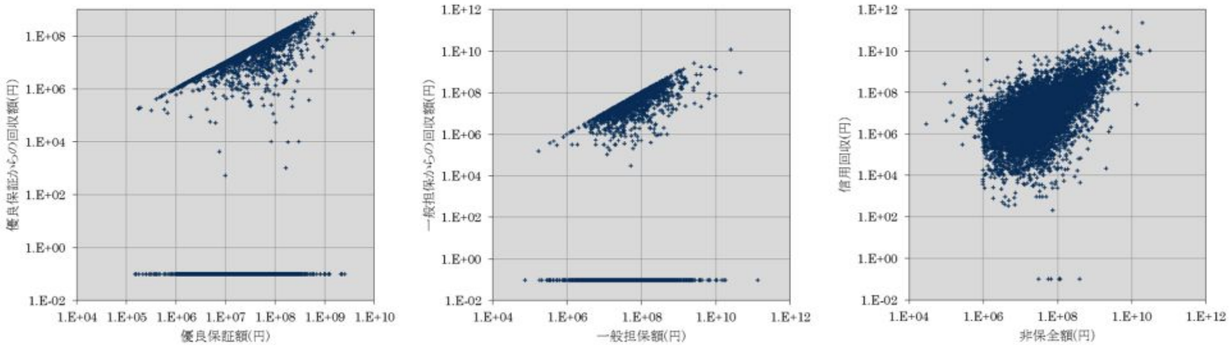

図 1 回収源泉別の回収額
左から優良担保による回収金額、一般担保による回収金額、信用による回収金額
（注）回収額が 0 円の場合は回収額 0.1 円=1.0E-1 にプロットしている。優良担保、一般担保、信用貸出残高の額が 0 円以下は除いているため、N=4955(優良担保)、3121（一般担保)、9926（信用貸出）である。

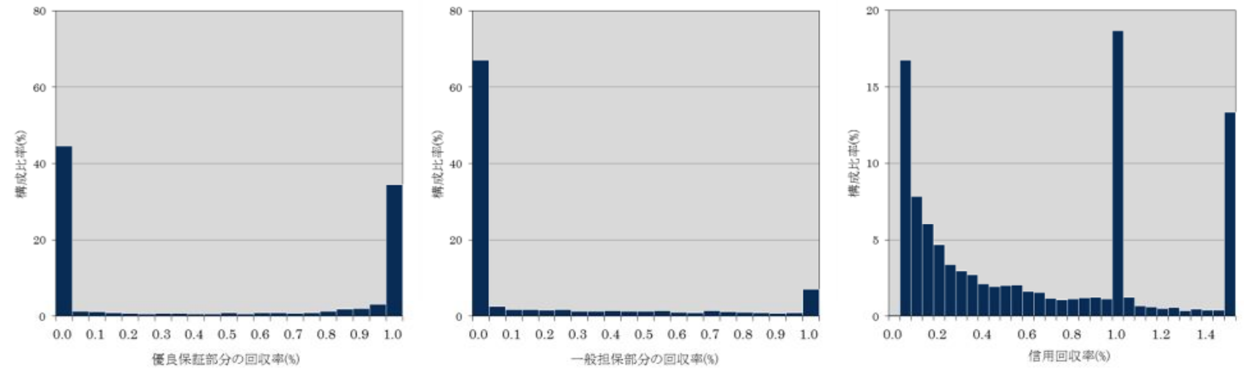

図 2 回収源泉別の回収率ヒストグラム
左から優良担保による回収率、一般担保による回収率、信用による回収率
（注）信用回収率が 150%を超えるものは 150%に集約している。

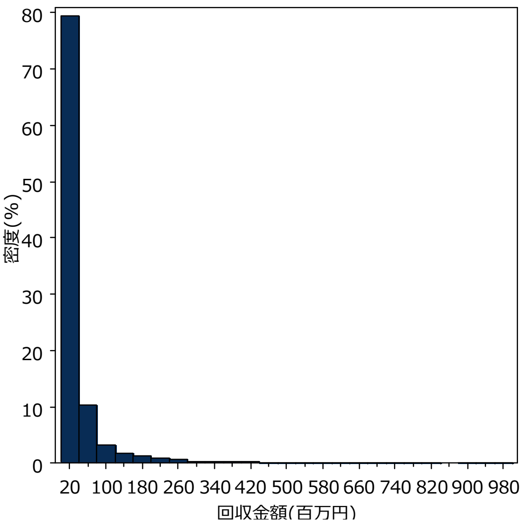

図 3 信用回収額のヒストグラム

表 3 モデル推計結果

	model1		model2		model3		model4	
	GLM, Gamma, RA		GLM, Normal, RA		LM, RA		LM, log(RA)	
変数名	係数	p 値	係数	p 値	係数	p 値	係数	p 値
定数項	0.884	0.0003	-0.168	<.0001	63.631	<.0001	-0.478	0.0266
規模小	-0.152	0.1396	-0.018	0.9913	27.936	0.0003	0.327	0.0146
規模中	0.004	0.9451	-0.023	0.8675	33.167	<.0001	0.262	0.0014
金融機関 1	-0.474	<.0001	-0.844	<.0001	-144.084	<.0001	-0.466	<.0001
金融機関 2	-0.432	<.0001	-0.377	<.0001	-129.465	<.0001	0.176	0.0194
有担保有保証	-1.394	<.0001	-0.176	<.0001	26.683	0.0149	0.338	0.0063
無担保無保証	0.218	0.0075	0.081	<.0001	56.489	<.0001	0.055	0.6269
保証のみ有	-1.202	<.0001	-0.138	<.0001	46.111	<.0001	0.344	0.0037
貸出残高 P_1	-0.097	<.0001						
現預金対貸出残高 P 比率_1	0.052	<.0001						
優良保証比率_1	0.196	<.0001						
純有利子負債対貸出残高比率_1			0.183	<.0001				
自己資本比率_1			-0.071	<.0001				
有形固定資産対負債比率_1			0.030	<.0001				
純有利子負債シェア相当分_4					0.368	<.0001		
流動資産シェア相当分_4					-0.254	<.0001		
経常運転資金_4					0.004	<.0001		
貸出残高 P_3							0.650	<.0001
現預金シェア相当分_2							0.900	<.0001
収益弁済対象有利子負債シェア相当分_3							-0.031	<.0001
対数尤度　定数項のみ	-19703.4		-30901.4		-30901.4		-9166.6	
対数尤度　共通モデル	-17169.4		-27880.7		-30715.6		-8467.7	
対数尤度	-17068.2		-27027.8		-27663.0		-8050.6	

	model5		**model6**		**model7**		**model8**	
	LM, RR		LM, limited,RR		LM, RR logit		LM, RR Beta	
変数名	係数	p 値	係数	p 値	係数	p 値	係数	p 値
定数項	0.369	<.0001	0.4897	<.0001	0.852	0.0142	-0.155	0.0341
規模小	-0.268	<.0001	0.0218	0.5294	0.073	0.8638	-0.069	0.3851
規模中	-0.077	0.0084	0.0199	0.2793	0.206	0.3616	0.050	0.2706
金融機関 1	-0.277	<.0001	-0.294	<.0001	-3.284	<.0001	-0.666	<.0001
金融機関 2	-0.195	<.0001	-0.203	<.0001	-2.163	<.0001	-0.468	<.0001
有担保有保証	-0.250	<.0001	-0.0474	0.1227	-0.536	0.1562	-0.023	0.7528
無担保無保証	0.015	0.7182	0.0341	0.1841	0.682	0.031	0.141	0.0273
保証のみ有	-0.096	0.0431	0.0158	0.5994	0.011	0.9765	0.120	0.0702
優良保証比率_1_2	0.278	<.0001						
保証担保合計__4	0.001	<.0001						
貸出残高 P_1	-0.097	<.0001						
貸出残高 P_1			-0.0473	<.0001				
現預金_1			0.0200	<.0001				
優良保証比率_1_2			0.1671	<.0001				
貸出残高 P_1					-0.726	<.0001		
現預金_1					0.291	<.0001		
優良保証比率_1_2					2.083	<.0001		
貸出残高 P_1							-0.138	<.0001
現預金対貸出残高 P 比率_1							0.045	<.0001
保証担保合計__4							0.001	<.0001
対数尤度　定数項のみ	-4087.6		-2224.4		-13209.0		-6080.9	
対数尤度　共通モデル	-3875.6		-1764.3		-12764.5		-5576.6	
対数尤度	-3646.2		-1664.7		-12629.4		-5462.6	

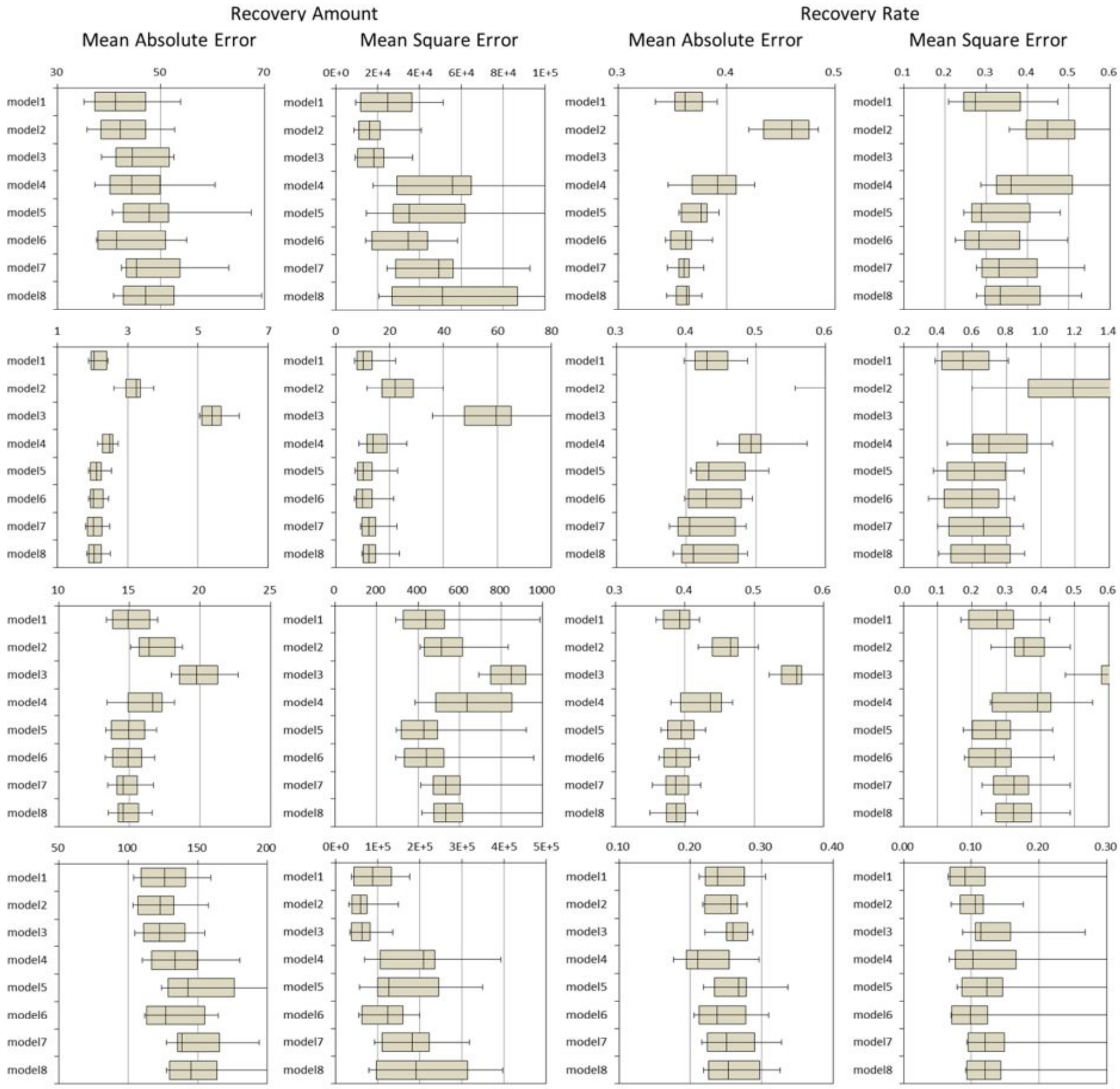

図 4 推計回収率および推計回収額の平均絶対誤差 MAE、平均二乗誤差 MSE

図左側のパネルから回収額の平均絶対誤差（MAE)、平均二乗誤差（MSE)、回収率の MAE、MSE。最上段のパネルは全データ、二段目は EAD 額が 1000 万円以下、三段目は 1000 万円超 1 億円以下、四段目は 1 億円超のデータ

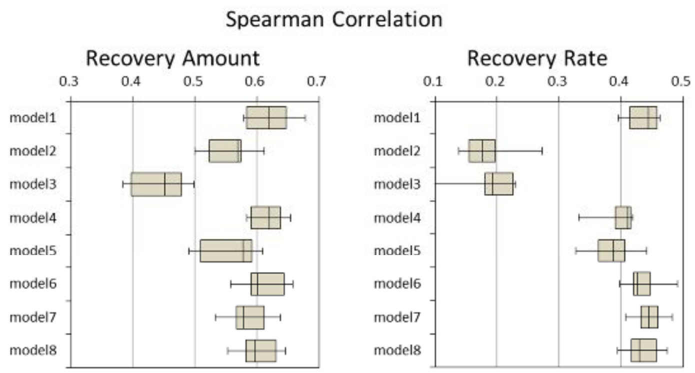

図 5 推計回収率および推計回収額の Spearman 順位相関