

Оценка финансовых потоков от взыскания в рамках коллекторского скоринга

З.В. Максименко

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия

e-mail: maximenkozv@gmail.com

Г.Р. Бикбулатова

Отдел финансирования государственной
службы и оплаты труда
Министерство финансов Республики
Башкортостан
Уфа, Россия

e-mail: guzel-sabiryanova@yandex.ru

Л.Ф. Розанова

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия

e-mail: rozanova_lara@mail.ru

К.А. Фёдорова

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия

e-mail: kseniya-666@mail.ru

Аннотация¹

Резкий рост просроченной задолженности по розничному кредитованию – в настоящее время одна из важнейших проблем финансово-кредитной системы как России, так и западных стран. Оптимизировать деятельность по взысканию долгов возможно с помощью систем коллекторского скоринга. В статье рассматривается одна из основных задач коллекторского скоринга – оценка финансовых потоков от взыскания задолженности и вероятности ее взыскания.

1. Введение

Возникновение просроченной задолженности — на сегодняшний день одна из самых распространенных проблем в банковской сфере в России. Причинами невозврата кредита заемщиком может быть как сложившаяся тяжелая жизненная ситуация, например, потеря источника доходов или работы, так и умышленное уклонение от уплаты кредита. Одним из наиболее эффективных инструментов взыскания долгов является система коллекторского скоринга. Коллекторский скоринг – это система определения приоритетных направлений работы по возврату просроченной задолженности на основе анализа проблемного портфеля, практическим результатом чего будет построение оптимальной стратегии работы с каждым должником [4].

Труды третьей международной конференции
"Информационные технологии интеллектуальной
поддержки принятия решений", 18 - 21 мая, Уфа,
Россия, 2015

Данная статья посвящена решению первоочередной задачи коллекторского скоринга – классификации должников и определению вероятной суммы взыскания с них. В результате классификации выделяется категория безнадежных клиентов, которые не могут или не хотят возвращать долг. Это не значит, что коллекторы не должны с ними работать – как показывает практика, даже для таких долгов характерен минимально возможный уровень взыскания. Результаты скоринга позволяют выбрать наилучший способ взыскания в отношении каждого должника и делают действия коллекторского подразделения/агентства более оправданными.

2. Описание исследования

Целью исследования является оценка размера суммы возврата с каждого должника и вероятности взыскания задолженности. Объектом исследования выступают физические лица – потребители услуг розничного кредитования (заемщики банков, действующих на территории Российской Федерации), имеющие просроченную задолженность по кредитам, находящихся в работе коллекторского подразделения банка.

Вероятность и сумма возврата долга зависит от многих факторов, часть из которых известна и может быть оценена статистически. Для исследования были отобраны следующие факторы, характеризующие выданный кредит, просрочку по кредиту, а также личность заемщика – всего 26 исходных показателей, включающих финансовую информацию и обезличенные социально-демографические данные. Также для исследования предоставлены значения двух целевых переменных. Первая принимает значение 1, если заемщик производит выплаты по

долгу, и 0, если заемщик не производит выплаты по долгу. Вторая представляет среднюю ежемесячную сумму взыскания с каждого должника, причем для неплательщиков она равна 0.

По заданному набору данных необходимо разработать аналитическую модель, которая для каждого займа позволит спрогнозировать вероятность того, что заемщик будет производить выплаты по долгу и определить вероятную сумму взыскания.

Традиционный метод нахождения вероятной суммы взыскания в данной задаче предполагает построение цензурированной линейной модели регрессии или Тобит-I (частично наблюдаемая переменная – вероятная сумма взыскания), где принятие должником решения платить или не платить определяется самой суммой долга, которую он намеревается выплатить. Однако правильнее бы было рассмотреть другую модель, в которой процесс принятия решения о сумме выплаты отделен от процесса принятия решения «платить/не платить». Для этого в данном исследовании будет применяться модель Тобит-II [5].

Получить состоятельные и асимптотически эффективные оценки параметров модели Тобит-II можно, используя метод максимального правдоподобия, при котором соответствующая функция правдоподобия максимизируется по всем возможным значениям параметров модели. Однако чаще такую модель оценивают, используя простую в вычислительном отношении двухшаговую процедуру Хекмана [5]. В данном исследовании на первом этапе будет строиться модель бинарного выбора попадания в круг неплательщиков, а на втором этапе, модель, позволяющую определить для должников, делающих взносы, вероятную ежемесячную сумму погашения задолженности.

3. Математическое описание модели Тобит-II

В модели Тобит-II имеются две латентные переменные, удовлетворяющие следующим моделям:

$$z^* = z^T c + u, \quad (1)$$

$$y^* = x^T b + \varepsilon, \quad (2)$$

где x^T , z^T – факторы модели, b, c – векторы коэффициентов при факторах, ε , u – векторы ошибок.

1 этап. Модель бинарного выбора для определения «попадания» клиента в группу осуществляющих платежи. z^* определяет «будет/не будет платить».

2 этап. Модель для определения суммы взыскания при условии попадания должника в группу осуществляющих платежи. Если выбирается «не попадание», то u не наблюдается (равна нулю), так как платежей нет.

$$y = \begin{cases} y^*, & g = 1 \\ 0, & g = 0 \end{cases}, \quad (3)$$

$$g = \begin{cases} 1, & z^* > 0 \\ 0, & z^* \leq 0 \end{cases}. \quad (4)$$

Предполагая, что случайные ошибки моделей латентных переменных коррелированы и связаны соотношением: $\varepsilon = \sigma_{\varepsilon u} u + v$, $\sigma_{\varepsilon u}$ – коэффициент корреляции между величинами ε и u ,

$$\begin{aligned} \Rightarrow E(y|g=1) &= x^T b + \sigma_{\varepsilon u} E(u|u > -z^T c) = \\ &= x^T b + \sigma_{\varepsilon u} \frac{f(z^T c)}{F(z^T c)} = x^T b + \sigma_{\varepsilon u} \lambda(z^T c) \end{aligned} \quad (5)$$

где F и f – соответственно функция распределения и плотность либо стандартного нормального ($F(x) \equiv \Phi(x)$), либо логистического ($F(x) \equiv \Lambda(x)$), либо экстремального распределения ($F(x) \equiv E(x)$), $\lambda(z^T c)$ – «лямбда Хекмана» [1].

Таким образом на втором этапе для каждого должника оценивается линейная модель:

$$y_i = x_i^T b + \sigma_{\varepsilon u} \lambda_i + \eta_i \quad (6)$$

4. Построение и анализ модели бинарного выбора

Набор переменных, использующийся для построения модели, и их ожидаемое влияние на вероятность погашения долга представлены в таблице 1. Общее количество исходных данных после корректировки составило 2320 наблюдений. Традиционно 70% наблюдений выборки применялись для обучения, 30% – для тестирования.

На этапе предварительного анализа исходных данных исключены переменные Дата выхода в просрочку и Остаток основного долга в рублях как значительно коррелирующие с другими переменными. На этапе построения моделей исключены переменные R3, R8 и Сумма начисленной госпошлины в рублях, как однозначно определяемые моделью (все наблюдения имеют одинаковое значение целевой переменной).

В результате моделирования (было построено и проанализировано около 50 моделей) выявлено, что наилучшим образом модель описывает логистическая регрессия. Функция стандартного логистического распределения:

$$F(u) = \Lambda(u) = \frac{e^u}{1 + e^u}, \text{ где } u = c_0 + c_1 z_1 + c_2 z_2 + \dots + c_n z_n, \quad (7)$$

где c_0 – независимая константа модели, c_j – параметры модели, z_j – значение j -й независимой переменной, $j = \overline{1, n}$.

Таблица 1. Описание переменных

Описание переменной	Тип переменной	Обозначение	Ожидаемый эффект на вероятность возврата долга
Целевая переменная - вероятность возврата долга	Целевая	TARGET	-
Дата выдачи кредита (дата заключения договора на выдачу)	Количественная	DATE_OF_LOAN_ISSUE	Неясный
Размер кредита, выраженный в рублях	Количественная	LOAN_SIZE	Отрицательный
Размер ежемесячного платежа в рублях	Количественная	MONTHLY_PAYMEN	Отрицательный
Число месяца, в которое производится ежемесячная выплата	Количественная	MONTHLY_PAY_DATE	Неясный (возможно незначимый)
Срок кредита в месяцах по договору	Количественная	LOAN_DURATION	Положительный
Дата окончания кредита по договору	Количественная	LOAN_EXPIRY_DATE	Неясный
Годовая процентная ставка по договору	Количественная	INTEREST_RATE	Отрицательный
Тип залога	Бинарная	PLEDGE_TYPE	Положительный
Количество поручителей	Количественная	SURETY_AMOUNT	Положительный
Тип долга (кредиты МСБ, экспресс-кредит, потребительский, кредитная карта, автокредит)	Категориальная	CREDIT_MSB CREDIT_EXPRESS CREDIT_POTREB CREDIT_CARD CREDIT_AVTO	Неясный
Дата выхода в просрочку	Количественная	DATE_OF_DEFAULT	Положительный
Количество дней в просрочке с даты выхода в просрочку	Количественная	DAYS_DELAY	Отрицательный
Общая сумма долга в рублях	Количественная	TOTAL_DEBT	Отрицательный
Остаток основного долга в рублях	Количественная	PRINCIP_BALANCE	Неясный
Сумма начисленных штрафов в рублях	Количественная	SUM_OF_PENALTIE	Отрицательный
Сумма начисленной комиссии в рублях	Количественная	SUM_COMMISSION	Неясный (возможно незначимый)
Сумма начисленной госпошлины в рублях	Количественная	S_TAX_AMOUNT	Неясный (возможно незначимый)
Начисленные проценты в рублях	Количественная	PERCENTS	Неясный
Сумма прочих начислений в рублях	Количественная	SUM_OF_OTHER	Неясный (возможно незначимый)
Дата рождения заемщика	Количественная	DOB	Неясный
Пол заемщика (0- женщина, 1- мужчина)	Бинарная	GENDER	Неясный
Регион фактического места проживания должника (центральный, южный, северо-западный, дальневосточный, сибирский, уральский, приволжский, северо-кавказский)	Категориальная	R1 R2 R3 R4 R5 R6 R7 R8	Неясный
Город фактического проживания должника (тип населенного пункта по количеству населения)	Категориальная	G1 G2 G3 G4 G5 G6 G7	Неясный
Индикатор наличия указанного телефона для связи с должником	Бинарная	TEL_FLG	Положительный
Дата последнего платежа	Количественная	LAST_PAYMENT_DATE	Отрицательный
Тип безнадежности	Бинарная	TYPE_OF_HOPELESS	Отрицательный

Оценивание параметров модели производилось с помощью метода максимального правдоподобия в ППП EViews. Построенная модель приведена на рис.1.

Гипотеза о значимости коэффициентов модели проверялась с помощью теста отношения

правдоподобия (LR) на уровне значимости 0,05. Адекватность подобранной модели реальному процессу была проверена с помощью теста Хосмера-Лемешоу [3] при уровне значимости 0,05. Тест показал, что модель адекватна, и может быть использована в аналитических целях (рис.2).

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: TARGET									
Method: ML - Binary Logit (Quadratic hill climbing)									
Date: 02/23/15 Time: 01:31									
Sample (adjusted): 1 2320									
Included observations: 2320 after adjustments									
Convergence achieved after 12 iterations									
Covariance matrix computed using second derivatives									
Variable	Coefficient	Std. Error	z-Statistic	Prob.					
DATE_OF_LOAN_ISSUE	0.013326	0.005942	2.242746	0.0249					
LOAN_EXPIRY_DATE	-0.014586	0.005962	-2.446605	0.0144					
LAST_PAYMENT_DATE	-0.000138	3.99E-05	-3.446795	0.0006					
SURETY_AMOUNT	0.940263	0.210032	4.476755	0.0000					
LOAN_DURATION	0.430274	0.177042	2.430356	0.0151					
SUM_OF_PENALTIE	0.000132	5.05E-05	2.617602	0.0089					
PERCENTS	-2.42E-05	4.75E-06	-5.099013	0.0000					
MONTHLY_PAYMEN	0.000133	2.75E-05	4.828407	0.0000					
TYPE_OF_HOPELESS	-1.429203	0.568776	-2.512770	0.0120					
GENDER	-0.229621	0.100561	-2.283386	0.0224					
DOB	-3.85E-05	1.33E-05	-2.898206	0.0038					
G2	-0.446686	0.181294	-2.463878	0.0137					
G4	0.350077	0.120850	2.896793	0.0038					
G7	-0.326921	0.125274	-2.609651	0.0091					
R2	-0.713590	0.344304	-2.072558	0.0382					
R6	-1.468401	0.611864	-2.399883	0.0164					
C	57.52592	7.081312	8.123625	0.0000					
McFadden R-squared	0.177360	Mean dependent var	0.374569						
S.D. dependent var	0.484116	S.E. of regression	0.430652						
Akaike info criterion	1.102750	Sum squared resid	427.1172						
Schwarz criterion	1.144878	Log likelihood	-1262.190						
Hannan-Quinn criter.	1.118102	Restr. log likelihood	-1534.315						
LR statistic	544.2509	Avg. log likelihood	-0.544047						
Prob(LR statistic)	0.000000								
Obs with Dep=0	1451	Total obs	2320						
Obs with Dep=1	869								

Рис. 1. Результаты построения модели бинарного выбора

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Goodness-of-Fit Evaluation for Binary Specification Andrews and Hosmer-Lemeshow Tests Grouping based upon predicted risk (randomize ties)									
Quantile of Risk	Low	High	Actual	Dep=0 Expect	Actual	Dep=1 Expect	Total Obs	H-L Value	
1	0.0224	0.1558	202	206.525	30	25.4747	232	0.90303	
2	0.1563	0.2076	193	189.591	39	42.4089	232	0.33531	
3	0.2076	0.2426	179	179.764	53	52.2356	232	0.01444	
4	0.2430	0.2771	170	171.902	62	60.0985	232	0.08120	
5	0.2771	0.3156	162	163.063	70	68.9370	232	0.02332	
6	0.3158	0.3615	163	153.718	69	78.2816	232	1.66092	
7	0.3615	0.4165	145	142.271	87	89.7288	232	0.13533	
8	0.4167	0.5049	118	125.807	114	106.193	232	1.05845	
9	0.5049	0.7055	97	94.3181	135	137.682	232	0.12850	
10	0.7068	1.0000	22	24.0400	210	207.960	232	0.19312	
Total			1451	1451.00	869	869.000	2320	4.53362	
H-L Statistic	4.5336		Prob. Chi-Sq(8)		0.8061				
Andrews Statistic	4.5938		Prob. Chi-Sq(10)		0.9166				

Рис. 2. Результаты теста Хосмера-Лемешоу

Интерпретация результатов моделирования в силу нелинейности модели проводится на основе предельных (маржинальных) эффектов (табл. 2). Предельный коэффициент каждого объясняющего фактора z_j , $j=1,...,k$ является непрерывным и зависит от значения остальных факторов и определяется по формуле:

$$\frac{\partial P(z_j^*=1)}{\partial z_j} = c_j \cdot F'(z_j^T c) = c_j \cdot f(z_j^T c), \quad (8)$$

где f – плотность вероятности.

Для логит-модели:

$$\frac{\partial P(z_j^*=1)}{\partial z_j} = c_j \cdot \Lambda'(z_j^T c) = c_j \cdot \lambda(z_j^T c), \quad (9)$$

$$\text{где } \lambda(u) = \frac{e^u}{(1+e^u)^2}.$$

По выполненным расчетам можно сделать следующие выводы:

1. Наибольшее положительные влияние на вероятность возврата долга оказывают показатели: количество поручителей, срок кредита в месяцах по договору, город фактического проживания должника с населением 100-450 тыс. чел. Наличие поручителя по кредиту увеличивает вероятность возврата долга на 17,3%, что является вполне предсказуемым результатом. Каждый дополнительный месяц срока кредита повышает вероятность возврата на 7,91% – это может объясняться как наличием дополнительного времени для возврата, так и меньшим среднемесячным платежом, что в целом снижает нагрузку на бюджет должника. Для проживающих в городах с численностью населения 100-450 тыс. чел. вероятность возврата повышается на 6,44%. Вероятно, в таких городах представлено малое количество банков, соответственно меньше возможностей взять кредит, особенно с плохой кредитной историей, поэтому дисциплина исполнения кредитных обязательств более высока. Так же города такого масштаба как правило обслуживают предприятия сферы добычи и переработки полезных ископаемых и являются достаточно богатыми.
2. Существенно снижают вероятность возврата долга показатели: тип безнадёжности (смерть заемщика) на 26,3%, регион фактического места проживания должника – уральский и южный на 25,02% и 13,13% соответственно. Также уменьшается вероятность возврата долга, если заемщик проживает в городе-миллионнике на 8,22% или в малых поселениях (село, деревня, остров) на 6,01%. Для больших городов это может быть связано с более дорогим уровнем жизни наряду с высокими зарплатами, с большей закрепованностью населения. В случае изменения финансовой кредитоспособности или каких-либо непредвиденных обстоятельств заемщик оказывается не в состоянии выплатить очередной долг. Для малых же поселений характерна низкая финансовая и юридическая грамотность, а также трудности доступа к отделениям банка для своевременной оплаты долга, что влечет низкий уровень кредитной дисциплины.
3. Также существенное влияние оказывает пол должника. Выявлено, что мужчины хуже возвращают долги. Так если заемщик мужского пола, вероятность возврата снижается на 4,22%.

Таблица 2. Маржинальные эффекты для параметров модели

Параметр	Коэффициент logit-модели	Маржинальный эффект
DATE_OF_LOAN_ISSUE	0,01333	0,2453%
DOB	-0,00004	-0,0010%
GENDER	-0,22962	-4,2260%
LAST_PAYMENT_DATE	-0,00014	-0,0030%
LOAN_DURATION	0,43027	7,9190%
LOAN_EXPIRY_DATE	-0,01459	-0,2680%
MONTHLY_PAYMENT	0,00013	0,0025%
PERCENTS	-0,00002	-0,0004%
SUM_OF_PENALTIES	0,00013	0,0024%
SURETY_AMOUNT	0,94026	17,3050%
TYPE_OF_HOPELESS	-1,4292	-26,3040%
R2 Южный ФО	-0,71359	-13,1330%
R6 Уральский ФО	-1,4684	-25,0250%
G2 Города-миллионники	-0,44669	-8,2210%
G4 Города с нас. 100-450 тыс. чел.	0,35008	6,4430%
G7 Малые поселения (село, деревня, остров)	-0,32692	-6,0170%

Оценка качества классификации проводилась на основе анализа ROC-кривых, а также расчете производных от нее: показателя AUC (площади под кривой) и коэффициента Джини [3Ошибка!

Источники ссылки не найден.]. Графики ROC-кривых приведены и значения показателей качества модели на рис. 3 и в табл.3.

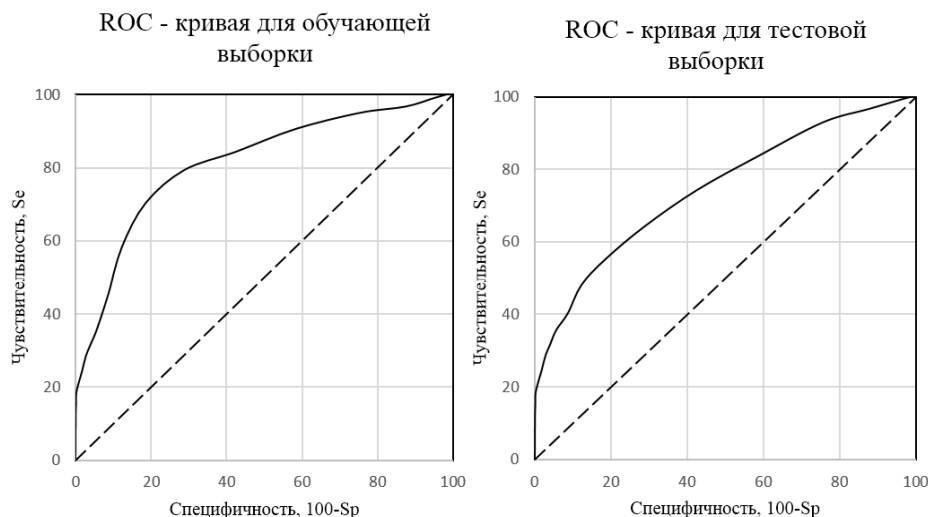


Рис. 3. Анализ ROC-кривых

Таблица 3. Результаты оценки качества модели

Коэффициенты	Для обучающей выборки	Для тестовой выборки
AUC	0,8074	0,7461
Коэф. Джинни= $2 \cdot AUC (AUC=0,5)$	0,6148	0,4922

Чем ближе график ROC-кривой проходит к левому верхнему углу, тем качественнее модель. Если кривая совпадает с диагональю, то модель бесполезна.

Значение площади под кривой $AUC=1$ соответствует идеальному, а $AUC=0,5$ – бесполезному классификатору. Значения от 0,7 до 0,9 показывают хорошую классификационную способность модели. Также интерпретируется и коэффициент Джинни – чем ближе его значение к 1, тем выше прогностическая способность модели [2].

ROC-анализ также позволяет выбрать оптимальный порог вероятности разделения должников на тех, кто будет и не будет возвращать долг. В табл. 4 приведен фрагмент массива точек «Чувствительность-Специфичность».

Таблица 4. Определение порога отсечения

Для обучающей выборки										
Порог отсечения	...	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	...
Se+Sp	...	132,24	141,82	150,52	151,74	145,95	136,71	130,14	127,81	
Se-Sp	...	48,85	26,86	8,13	11,04	29,52	45,93	58,56	64,06	...
Для тестовой выборки										
Se+Sp	...	125,24	131,82	135,52	136,74	135,95	131,71	130,14	127,81	...
Se-Sp	...	41,85	16,86	6,87	26,04	39,52	50,93	58,56	64,06	...

Для определения оптимального порога необходимо задать критерий его определения, т.к. в разных задачах существует своя оптимальная стратегия. Критериями выбора порога отсечения могут выступать [2]: максимум чувствительности и специфичности (Se+Sp), баланс между чувствительностью и специфичностью (|Se-Sp|) и другие подходы. По результатам анализа в качестве порога разделения должников выбрано значение 0,4.

Полученные на базе данной бинарной модели результаты в дальнейшем применяются для оценки линейной модели для определения суммы взыскания с каждого должника.

Проводимое исследование носит прикладной характер и имеет практическую ценность, т.к. позволяет провести оценку финансовых потоков от взыскания, оптимизацию работы с должником (исходя из рассчитываемой вероятности положительного исхода), максимизировать прибыль от взысканий по портфелю за счет правильного выбора стратегии работы с отдельными должниками.

5. Заключение

В статье рассмотрена задача оценки финансовых потоков от взыскания просроченной задолженности. Предложено решение с применением модели Тобит-II, позволяющее определить вероятность того, что заемщик будет производить выплаты по долгу, разделить должников на группы плательщиков/не плательщиков и оценить вероятную сумму взыскания с каждого должника. Результаты исследования могут служить обоснованием для планирования работы с должником, позволят выявить безнадежные долги, и в соответствии с этим эффективно перераспределить коллекторские ресурсы и снизить операционные издержки по взысканию.

Список используемых источников

1. Heckman James. "Sample selection bias as a specification error". *Applied Econometrics* 2013; 31(3): 129-137.
2. *Логистическая регрессия и ROC-анализ - математический аппарат.* URL:

<http://www.basegroup.ru/library/analysis/regression/logistic/>

3. Максименко З.В., Лакман И.А., Розанова Л.Ф. "Информационная поддержка формирования стратегии взыскания просроченной задолженности на основе коллекторского скоринга". В сборнике: *Информационные технологии и системы.* Труды Четвертой Международной научной конференции. Отв. ред. Ю. С. Попков, А. В. Мельников. Челябинск, 2015. С. 139-140.
4. Михмель П.С., Довгий Н.В. "Коллекторский скоринг (collection scoring)". *Банковское дело.* 2013. №2. С.65-71.
5. Носко В.П. *Эконометрика для начинающих (Дополнительные главы).* –М.: ИЭПП, 2005. 379 с.