

# АНАЛИЗ СКОРИНГОВЫХ МОДЕЛЕЙ КРЕДИТНОГО РИСКА

Д.С. Петроченкова, В.М. Дуплякин

*Самарский государственный аэрокосмический университет*

*им. академика С.П. Королева, Самара, Россия*

Конкуренция на рынке кредитования населения заставила банки вести более агрессивную кредитную политику, направленную на увеличение кредитного портфеля за счет привлечения в короткие сроки широкого круга заемщиков. Эта задача была решена за счет упрощения процедуры кредитования (сокращен перечень необходимых документов для получения ссуды, ликвидирован институт поручительства). Привлекая клиентов таким образом, банки приняли на себя дополнительные кредитные риски, которые, реализовавшись, привели к росту просроченной ссудной задолженности.

Уменьшение кредитных рисков – актуальная задача, стоящая перед коммерческими банками. Опыт зарубежных банков показывает, что методы оценки рисков, основанные на математических моделях, являются наиболее объективными.

Модели оценки кредитного риска, использующие математические алгоритмы называются *скоринговыми моделями*. Скоринг физических лиц представляет собой сложную математическую систему оценки, основанную на различных характеристиках клиентов, таких как личный доход, возраст, семейное положение, профессия и многих других. Они являются входными переменными модели, классифицирующей потенциальных заемщиков. В результате анализа переменных, поступающих на вход скоринговой системы, на выходе системы скоринга получается интегрированный показатель, который и оценивает степень кредитоспособности заемщика по ранговой шкале: «хороший» заемщик или «плохой» заемщик. Данные для формирования скоринговых систем получают из вероятностей возвратов кредитов отдельными группами заемщиков, полученными из анализа кредитной истории тысяч людей.

В последние пять лет такой метод оценки кредитного риска стали использовать российские банки. Однако полноценный скоринг, основанный на методах математического моделирования, а не банальные опросные листы, на основании которых потенциальному заемщику присваиваются баллы, присутствует еще не везде; а существующие скоринговые модели

нуждаются в значительной доработке.

При построении модели необходимо принимать во внимание такие факторы, как изменения экономической ситуации в стране, кредитные истории новых клиентов банка и особенности экономики каждого из регионов. Последний аспект имеет большое значение; очевидно, что разница в экономическом положении различных регионов РФ порождает необходимость формирования отдельной скоринговой модели для каждого региона.

В данной статье представлены результаты работы по построению математической модели риска невозврата кредита заемщиком – жителем Самарской области по одному из видов кредитных карт Банка «Стандарт».

Для синтеза модели использованы данные из 819 кредитных досье заемщиков, получивших кредит в Банке в июне 2008 года и имеющих кредитную историю в 18 месяцев. 18 месяцев является типичным показательным периодом, в течение которого большинство счетов проявляют себя с позиции дефолта. Дефолт рассматривается как наличие у клиента трёхкратной задолженности более 30 дней, 2 раза более 60 дней или 1 раз более 90 дней.

При построении модели учитывалось 18 признаков, влияющих на поведение активного объекта управления. Проведенный корреляционный анализ позволил выявить факторы, находящиеся в тесной линейной зависимости, а также устранить мультиколлинеарность. В результате было принято решение оставить в модели следующие факторы:

- пол;
- возраст;
- наличие детей;
- место работы;
- должность;
- наличие недвижимости;
- коэффициент долговой нагрузки.

Коэффициент долговой нагрузки – вторичный, т.е. расчетный признак, который показывает, какую часть доходов заемщик тратит на обслуживание ссудной задолженности, при условии, что он полностью реализует кредитный лимит карты. Он определяется по формуле:

$$K_d = \frac{\text{Ежемесячный\_доход}}{\text{Лимит} \cdot 10\%}. \quad (1)$$

В рассматриваемой модели результирующий показатель является качественной переменной, определяющей одно из двух возможных состояний описываемого ей объекта: либо до истечения исследуемого периода наступает дефолт, либо дефолт не наступает.

Таким образом, вектор исходных статистических данных зависимой переменной будет содержать только дихотомические, бинарные признаки, т.е. его компоненты  $y_i$  могут принимать только одно из двух значений: «0» (дефолта нет) или «1» (дефолт наступил). Для исследования статистической зависимости между переменными в подобных случаях строят некоторую специальную регрессионную модель зависимости вероятности  $P\{y = 1 | \tilde{X}\}$  от линейной формы  $\theta^T \tilde{X}$ .

При построении классической линейной модели ( $y_i = \theta_1 x_{i1} + \dots + \theta_p x_{ip} + \varepsilon_i, i = 1, n$ ) возникает ряд проблем, а именно:

1. Вероятность любого события меньше 1 и больше 0, линейная функция может принимать также отрицательные или превышающие единицу значения.
2. Проблема гетероскедастичности, осложненная еще и тем, что в выражения для дисперсий  $\varepsilon_i$  входит неизвестный вектор параметров  $\theta$ .

Чтобы решить первую проблему, для моделирования значений  $P\{y_i = 1 | X_i\}$  подбирают функции, область значений которых ограничивается отрезком  $[0;1]$ , а линейная форма  $X^T \theta$  играет роль аргумента этой функции, т.е.

$$P\{y = 1 | X_i\} = F(X_i^T \theta), \quad (2)$$

причем функции  $F(z)$  должны удовлетворять следующим требованиям:

$$\left. \begin{aligned} &F(z) \text{ монотонно } \uparrow \text{ по } z \\ &0 \leq F(z) \leq 1 \\ &F(z) \rightarrow 1 \text{ при } z \rightarrow \infty \\ &F(z) \rightarrow 0 \text{ при } z \rightarrow 0 \end{aligned} \right\}. \quad (3)$$

Модели типа (2) при ограничениях (3) называют *моделями бинарного*

*выбора*. Наиболее распространенными среди них являются логит- и пробит-модели, которым соответствуют функция стандартного логистического распределения и функция стандартного нормального распределения.

В данном случае для моделирования риска невозврата кредита было принято решение использовать стандартную логит-модель. Для оценки параметров модели использован метод максимального правдоподобия. Полученная модель кредитного риска выглядит следующим образом:

$$Z = \begin{cases} 1, \frac{e^{\theta^T \tilde{x}_j}}{1 + e^{\theta^T \tilde{x}_j}} > 0,5 \\ 0, \frac{e^{\theta^T \tilde{x}_j}}{1 + e^{\theta^T \tilde{x}_j}} \leq 0,5 \end{cases} \quad (5) \quad \text{где}$$

$\theta = (-2,297; 0,103; 0,642; 0,094; 0,114; 0,918; 0,888; 0,178)^T$  - вектор значений параметров модели.

Наиболее часто используемым показателем качества модели является коэффициент детерминации, однако в случае бинарных зависимых переменных представляется необходимым определить какую-то иную меру качества представленной модели. Такой мерой представляется сравнение результатов, которые дает сформированная нами модель, с результатами работы модели, используемой Банком для данного продукта. Из 819 клиентов, которые получили положительную оценку действующей модели, 215 оказались «плохими» заемщиками (дефолт произошел); в результате Банк понес убыток, который составил 4 283 тыс. руб. В 604 случаях оценка оказалась правильной. Таким образом, оценка верна в 73,7% случаев.

Представленная модель, учитывающая социально-экономические особенности Самарской области, дала положительную оценку 726 заемщикам; из них 574 оказались «хорошими» клиентами, а 152 – «плохими». Оценка оказалась верной в 77,8% случаев.

Оценка финансовых результатов применения модели на данном этапе представляется затруднительной. В первую очередь, отдельного рассмотрения требует вопрос о том, к чему относить полученный эффект: к количеству клиентов или к периоду использования модели? В данном случае был выбран первый вариант, как более рациональный. Кроме того, необходимо принимать во внимание упущенную выгоду – процентный

доход, который недополучит Банк, отказав в предоставлении кредита благонадежным клиентам. Точную сумму дохода привести не представляется возможным, поскольку эта информация является конфиденциальной. С учетом используемой по данному продукту процентной ставки (36% годовых, проценты начисляются на остаток задолженности ежемесячно) была получена условная сумма 965 тыс. руб.

Правильное определение 63 из 215 «плохих» заемщиков в данном случае позволило бы избежать убытка в размере 1317 тыс. руб. В то же время, из 93 клиентов, которым модель предсказала дефолт, 63 оказались «плохими» и 30 – «хорошими». Условная сумма упущенной выгоды - 665 тыс. руб. Таким образом, можно говорить о том, что эффект от применения модели составляет 652 тыс. руб. или 0,796 тыс. руб. на одного клиента.

Безусловно, приведенные оценки являются ориентировочными и требуют уточнения на основе использования более представительной статистики.