

УДК 336.717.061, 519.21

*М.А. Широбокова***МОДЕЛЬ ОЦЕНКИ РИСКА ДЕФОЛТА НА ВСЕМ ПРОТЯЖЕНИИ ЖИЗНИ КРЕДИТА**

Вопрос управления кредитным риском в банковской сфере рассматривается на основе построения скоринговой модели, оценивающей индивидуальный риск заемщика и выступающей базой для оценки совокупного риска кредитного портфеля. При этом чаще всего под индивидуальным риском заемщика понимается просрочка по основному долгу или процентам в течение 90 дней и более в течение первого года жизни кредита. Основным недостатком данного подхода является неучтенный факт того, что с течением жизни кредита происходит изменение индивидуальной вероятности выхода в дефолт. Исследуется актуальная для коммерческих банков проблема оценки кредита не только на первый год жизни кредита, но и на весь срок его жизни. Согласно стандарту МСФО (IFRS) 9, объем формируемых резервов должен зависеть от динамики уровня риска, которая определяется с помощью оценки риска дефолта кредита на всем протяжении жизни кредита (lifetime PD). Предложена модель получения динамической оценки вероятности выхода в дефолт по кредиту на основе его данных о качестве исполнения обязательств по кредитному договору (future engineering) и фазе макроэкономического цикла. Модель строится на основе метода логистической регрессии с регуляризацией, качество которой определяется через расчет коэффициента Джини. Построение модели оценки вероятности выхода в дефолт производится на примере регионального розничного банка.

Ключевые слова: кредитный риск, соглашение Базель II, МСФО 9, поведенческий скоринг, вероятность дефолта, логистическая регрессия с регуляризацией.

В условиях повышенной неопределенности на финансовых рынках управление кредитным риском в банковском риск-менеджменте является первостепенной задачей. В соответствии с Базельским соглашением [3] основой управления кредитным риском выступает формирование резервного капитала, соответствующего рассчитанным ожидаемым потерям по портфелю. Поэтому именно корректность оценки ожидаемых потерь становится наиболее актуальной.

Согласно Базель II и IFRS 9 оценка ожидаемых потерь кредитного портфеля производится на основе *IRB* подхода к кредитному риску, который при соблюдении банком определенных минимальных условий и требований по раскрытию информации позволяет опираться на свои собственные внутренние оценки компонентов риска при расчете покрытия капиталом определенного риска [3]. Согласно соглашению Базель II, расчет ожидаемых потерь *EL* (Expected Loss) по кредиту производится по следующей формуле:

$$EL = PD \cdot EAD \cdot LGD, \quad (1)$$

где *PD* (Probability of Default) — вероятность наступления дефолта;

EAD (Exposure At Default) — подверженность кредитному риску, представляющая собой экономическую оценку стоимости активов, подверженных риску в момент объявления дефолта;

LGD (Loss Given Default) — потери в случае дефолта, отражающие долю невозвратного убытка при дефолте.

При этом чаще всего в качестве *PD* применяется индивидуальная оценка вероятности выхода в дефолт, которая рассчитывается на основе скоринговой модели. Согласно рекомендациям соглашения Базель II, дефолтом считается просрочка по основному долгу или процентам в течение 90 дней и более в течение первого года жизни кредита, а в качестве *PD* выбирается вероятность возникновения просрочки по основному долгу или процентам более 90 дней в течение первого года жизни кредита (здесь и далее в качестве года будет приниматься значение 365 календарных дней). Основной проблемой при использовании текущего определения становится неучтенный факт того, что с течением жизни кредита происходит изменение индивидуальной вероятности выхода в дефолт. В связи с чем вопрос оценки кредита не только на первый год жизни кредита, а на весь срок его жизни становится все более актуальным [8]. Данное положение закреплено в стандарте IFRS 9 [1], согласно которому объем формируемых резервов должен зависеть от динамики уровня риска, которая определяется с помощью оценки риска дефолта кредита на всем протяжении жизни кредита (lifetime estimation).

Целью данной статьи является построение модели оценки вероятности *PD_t* выхода в дефолт на некоторый момент времени (срез) *t* с учетом качества исполнения обязательств по кредитному дого-

вору и фазы макроэкономического цикла. Динамическую оценку PD_t риска дефолта кредита предлагается строить с помощью модели поведенческого скоринга (behavior scoring), которая позволяет оценить характеристики заемщика на основе его кредитной истории по уже выданному кредиту.

Рассмотрим построение модели оценки вероятности PD_t выхода в дефолт на некоторый момент времени (срез) t на примере регионального розничного банка. Пусть для каждого кредита имеется оценка PD_0 скоринговой модели, определяющая вероятность выхода в дефолт в течение первого года жизни кредита и рассчитанная на момент выдачи кредита на основе анкетных характеристик заемщика. Тогда построение оценки PD_t производится путем коррекции базового значения PD_0 некоторой моделью поведенческого скоринга. Определением дефолта для построения скоринговой модели оценки PD_t будет являться вероятность возникновения просрочки по основному долгу или процентам более N дней в течение следующего года жизни кредита от даты среза t . То есть в данном случае имеет место условная вероятность того, что событие дефолта произойдет в интервале $[t, t + 365]$ при условии того, что на момент оценки событие дефолта не произошло: $P(t < T \leq t + 365 | T > t)$, где t – момент среза, T – момент выхода на просрочку. Данная оценка PD_t рассчитывается на основе скоринговой модели, целевой функцией которой служит индикатор выхода в дефолт в интервале $[t, t + 365]$. В рамках данной работы и в соответствии с IFRS 9 моментами среза t имеет смысл выбирать последнее число месяца оценки (здесь и далее последнее число месяца включительно).

В качестве исходных данных для построения скоринговой модели оценки PD_t служат данные кредитного портфеля по состоянию на последнее число месяца за некоторый период (например, 2012-01-31, 2012-02-29, 2012-03-31, ...). Для каждой даты среза t формируется портфель действующих кредитов: открытых кредитов, по которым не произошло событие дефолта и по которым известны последующие фактические значения индикатора выхода в дефолт в интервале $[t, t + 365]$. Портфели кредитов для каждого момента среза t в единую выборку. Разделение на обучающую выборку и выборку для валидации производится на основе случайного отбора кредитов в отношении 70/30 таким образом, что каждый кредит попадает только в обучающую выборку или выборку для валидации. Стоит отметить, что при использовании простого случайного отбора наблюдений будем получать очевидное переобучение модели: коэффициент Gini на тестовой выборке может оказаться выше, чем на обучающей [11]. Корректное построение обучающей и валидирующей выборок позволит получать более стабильные модельные оценки, что после внедрения модели исключит возникновение непредсказуемых ситуаций при оценке поведенческой вероятности дефолта кредитного портфеля на некоторый срез (при условии неизменной структуры портфеля и стабильной макроэкономической ситуации).

Переменные, которые будут использоваться для построения модели поведенческого скоринга, строятся на основе кредитной истории по каждому кредиту. Поставленный вопрос выбора переменных, которые будут корректно отражать качество исполнения обязательств по кредитному договору (future engineering), является наиболее актуальным. Введем следующие обозначения: пусть PR – множество значений количества дней просрочки, которое допускает клиент, а M – множество значений количества месяцев, за которое производится оценка наличия просрочки PR , тогда $pr \in PR$, $m \in M$.

Тогда в качестве переменных предлагается использовать следующие количественные оценки:

- количество выходов на просрочку pr за первых m месяцев жизни кредита – данные переменные позволяют оценить качество обслуживания кредитов в первый год жизни кредита;
- количество выходов на просрочку pr за последних m месяцев жизни кредита (отсчет месяцев производится с даты среза обратным счетом) – позволяют оценить качество обслуживания кредитов в последний год жизни кредита перед датой среза, таким образом, оценивается наиболее свежая кредитная история по кредиту;
- доля общего числа выходов на просрочку pr за весь период жизни кредита от срока кредита в месяцах;
- доля общего числа выходов на просрочку pr за весь период жизни кредита от числа месяцев, которые прожил кредит к дате среза;
- доля максимального числа месяцев, когда кредит находился в непрерывной просрочке PR , от числа месяцев, которые прожил кредит к дате среза.

Аналогично рассчитываются переменные на основе суммы, вышедшей на просрочку, по кредитам:

- отношение суммы, вышедшей на просрочку pr за первых m месяцев жизни кредита, к общей сумме кредита;
- отношение суммы, вышедшей на просрочку pr за последних m месяцев жизни кредита, к общей сумме кредита, отсчет месяцев производится с даты среза обратным счетом;
- отношение общей суммы, вышедшей на просрочку pr за весь период жизни кредита, разделенной на срок кредита в месяцах, к общей сумме кредита;
- отношение общей суммы, вышедшей на просрочку pr за весь период жизни кредита, разделенной на число месяцев, которые существует кредит к дате среза, к общей сумме кредита;
- отношение средней суммы, вышедшей на просрочку, когда кредит находился в непрерывной просрочке pr , к общей сумме кредита.

Описанные переменные можно отнести к блоку переменных, изменяющихся с течением жизни кредита. Дополнительно стоит учесть тот факт, что историческая выборка кредитного портфеля по состоянию на некоторый срез t содержит в себе кредиты с разным сроком жизни. Используя в качестве переменных кредитную историю по кредитам, получаем, что для построения модели имеются неполные, ограниченные по времени данные, а именно данные, цензурированные справа, то есть имеется некоторая поведенческая история клиента по кредиту до момента среза, но нет информации о том, каким будет поведение клиента в следующий период $[t, T]$, где t – момент среза, T – срок кредита. В связи с этим вводится дополнительная переменная – факт дожития кредита до некоторого момента времени.

Помимо блока переменных, изменяющихся с течением жизни кредита, можно выделить блок переменных, фиксирующихся в момент выдачи кредита, а именно:

- доля первоначального взноса от суммы выданного кредита;
- доля первоначального взноса от стоимости товара;
- скоринговый балл PD_0 , определяющий вероятность выхода в дефолт в течение первого года жизни кредита и рассчитанный на момент выдачи кредита на основе анкетных характеристик заемщика.

Таким образом, для каждого договора рассчитываются переменные, на основе которых строится поведенческая скоринговая модель оценки выхода в дефолт. Из имеющегося многообразия методов построения модели в текущем примере в качестве метода построения модели был использован метод логистической регрессии с регуляризацией. Формула расчета скорингового балла PD :

$$PD = w \cdot \frac{1}{1 + e^{-z}}, \quad (2)$$

где z рассчитывается как сумма константы и баллов, соответствующих характеристикам заемщика и их сочетаниям: $z = b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + b_0$, где

$x_1, x_2, \dots, x_n$ – значения переменных,

$b_1, b_2, \dots, b_n$ – коэффициенты при переменных,

b_0 – некоторая константа,

w – поправочный коэффициент на макроэкономический цикл.

Значение z определяется в соответствии с рассчитанными переменными, являющимися характеристиками заемщика. Наложение дополнительных ограничений на вектор весов в модели производится с помощью регуляризации модели. Данный подход заключается в том, чтобы выбирать веса, малые по абсолютной величине в среднем [5], что приводит к снижению неустойчивости модели, то есть зависимости от обучающих данных. В качестве схемы регуляризации используется метод эластичной сети, совмещающий модели лассо и гребневой регрессии. Классическая задача регрессии на примере расчета скорингового балла PD ставится следующим образом:

$$|y - PD|^2 \rightarrow \min_B, \quad (3)$$

где PD – оцененное по модели значение скорингового балла,

B – вектор коэффициентов модели логистической регрессии,

y – целевая функция.

Тогда введение регуляризаторов типа $\lambda_1 |B|_1$ и $\lambda_2 |B|_2^2$ преобразует исходную задачу в следующий вид:

$$|y - PD|^2 + \lambda_1 |B|_1 + \lambda_2 |B|_2^2 \rightarrow \min_B. \quad (4)$$

При $\lambda_1 = 0$ задача сводится к частному случаю регуляризации — модели гребневой регрессии, при $\lambda_2 = 0$ — модели лассо. Таким образом, применение регуляризации позволяет произвести отбор наиболее значимых из большого числа используемых переменных.

Значение поправочного w коэффициента на макроэкономический цикл является некоторой константой в отдельно взятом периоде, расчет которой выходит за рамки текущей статьи. Если с учетом поправочного коэффициента величина PD становится больше 1, то PD принимается равным 1.

Качество при построении модели оценивается с помощью графика ROC-кривой и соответственного расчета площадь под ROC-кривой и коэффициента $Gini$.

Рассмотрим построение модели оценки вероятности PD_t выхода в дефолт на момент времени (срез) t на примере регионального розничного банка. В качестве исходных данных для построения скоринговой модели оценки PD_t служат данные по кредитному портфелю по состоянию на последнее число месяца за период с 2012-01-01 по 31-10-2016 с учетом факта «дожития» кредита до каждого исследуемого среза. Построение модели производится на наиболее свежем срезе портфеля: состояние портфеля на срез 31-10-2016 и его последующей фактической оценкой по состоянию на 31-10-2017. Для каждой даты среза был сформирован портфель действующих кредитов: открытых кредитов по состоянию на дату среза и с текущей просрочкой по основному долгу или процентам не более $N = 15$ дней. Портфели, полученные на каждый момент среза, были объединены в единую обучающую выборку согласно описанной методике [11]. Объем обучающей выборки составил 151586 кредитов, доля дефолтных кредитов в котором 0.04146 по состоянию на 31-10-2017. Объем валидирующей выборки – 66310, доля дефолтных кредитов – 0.04340. Объем тестовой выборки – 22038, доля дефолтных кредитов – 0.03825. Разделение выборки на обучающую и валидирующую произведено на основе описанного выше правила. Для каждой из выборок были рассчитаны переменные, приведенные ранее, полагая $PR = \{1+, 15+, 30+, 90+\}$ и $M = \{1, 3, 6, 9, 12\}$, и на их основе была построена модель логистической регрессии с регуляризацией. Коэффициент $Gini$ на обучающей выборке составил 0.7163, на валидирующей выборке – 0.7024, на тестовой выборке представляющей фактическую оценку по состоянию на 31-10-2017 – 0.6943 (рис.).

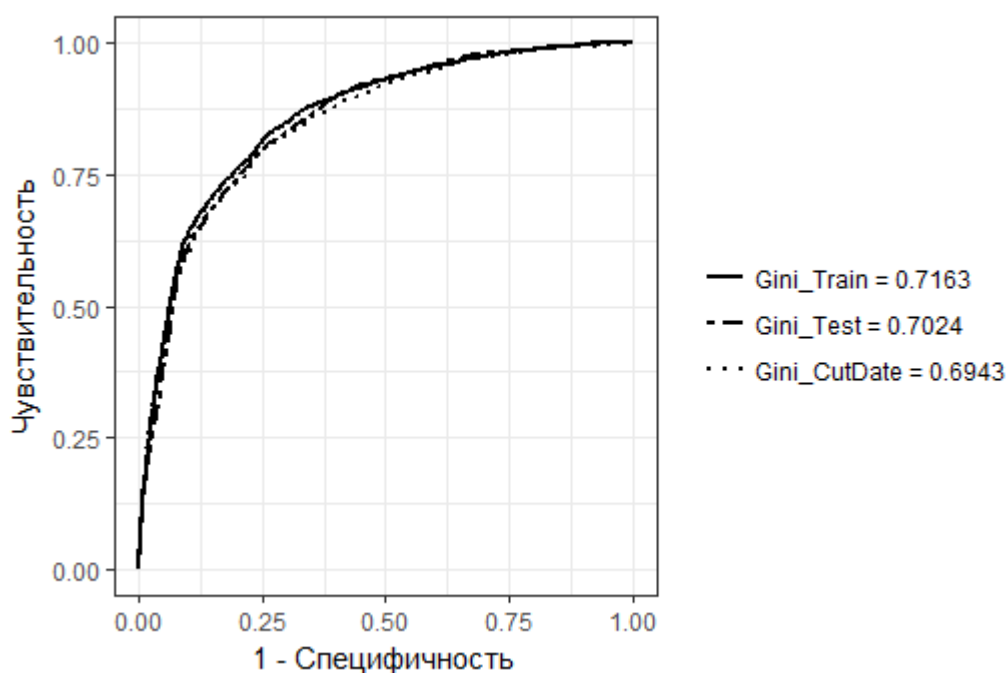


Рис. График ROC-кривой

Таким образом, в данной работе было приведено решение проблемы оценки кредита не только на первый «год жизни» кредита, а на весь срок его жизни (lifetime estimation). На основе построения модели поведенческого скоринга (behavior scoring). Рассмотренный пример построения модели оценки вероятности PD_t выхода в дефолт на некоторый момент времени (срез) t с учетом качества исполнения обязательств по кредитному договору, оцененных по описанным переменным (future engineering) и фазы макроэкономического цикла применен на примере регионального розничного банка. Полученные оценки качества модели свидетельствуют о ее высокой прогностической силе. Подобная оценка риска дефолта кредита на всем протяжении жизни кредита удовлетворяет требованиям стандарта IFRS 9 и позволяет на своей основе более точно формировать объем резервов по портфелю в зависимости от динамики уровня риска.

СПИСОК ЛИТЕРАТУРЫ

1. О требованиях к системе управления рисками и капиталом кредитной организации и банковской группы: указание Банка России от 15.04.2015 г. № 3624-У // Вестн. Банка России. 2015. № 1. С. 15-35. URL: <https://www.cbr.ru/publ/Vestnik/ves150615051.pdf> (дата обращения: 20.08.2017).
2. Энциклопедия финансового риск-менеджмента / под ред. канд. экон. наук А.А. Лобанова и А.В. Чугунова. М.: Альпина Паблишер, 2009. 932 с.
3. Алескеров Ф.Т., Андриевская И.К., Пеникас Г.И., Солодков В.М. Анализ математических моделей Базель II / 2-е изд., испр. М.: Физматлит, 2013. 296 с.
4. Международная конвергенция измерения капитала и стандартов капитала: Уточненные рамочные подходы / Базельский комитет по банковскому надзору. Банк международных расчетов. 2004. 266 с.
5. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А.А. Слинкина. М.: ДМК Пресс. 2015. 400 с.
6. Банных А.А., Летчиков А.В. Методика оценки кредитного риска заемщика с применением скоринга бюро кредитных историй // Вестн. Удм. ун-та. Сер. Экономика и право. 2013. Вып. 4. С. 5-9.
7. Банных А.А., Летчиков А.В. Методика расчета экономического капитала на покрытие непредвиденных потерь по портфелю потребительских кредитов // Вестн. Удм. ун-та. Сер. Экономика и право. 2015. Т. 25, вып. 1. С. 18-24.
8. Гуцин К., Козлов А. Новый подход к оценке поведенческой вероятности дефолта // Методический журнал «Риск-менеджмент в кредитной организации». 2017. №3(27). С. 35-42.
9. Лётчиков А.В. Расчет индивидуальной рискованной маржи при выдаче потребительского кредита // Математические методы и интеллектуальные системы в экономике и образовании: материалы Всерос. заоч. науч.-практ. конф. Ижевск: Изд-во ИЭиУ УдГУ, 2016. С. 7-9.
10. Лётчиков А.В., Маркова А.А. Прогнозная оценка убытков по кредитному портфелю на основе миграционной модели // Математические методы и интеллектуальные системы в экономике и образовании: материалы Всерос. заоч. науч.-практ. конф. Ижевск: Изд-во ИЭиУ УдГУ, 2015. С. 10-12.
11. Широбокова М.А. Обработка данных для построения модели оценки поведенческой вероятности дефолта // Математические методы и интеллектуальные системы в экономике и образовании: материалы Всерос. заоч. науч.-практ. конф. / под ред. А. В. Лётчикова. Ижевск: Изд-во ИЭиУ УдГУ, 2017. С. 26-30.
12. Широбокова М.А., Летчиков А.В. Сравнение методов калибровки скоринговой модели при прогнозировании логистической регрессией // Вестн. Удм. ун-та. Сер. Экономика и право. 2017. Т. 27, вып. 2. С. 74-79.

Поступила в редакцию 21.01.2018

М.А. Shirobokova

MODEL OF EVALUATING THE DEFAULT CREDIT RISK THROUGHOUT THE WHOLE LIFE OF THE LOAN

The issue of credit risk management in the banking sector is researched on the basis of building a scoring model which is used to evaluate an individual's credit risk for a borrower and to estimate the overall risk of the loan portfolio. In this case, most often the individual risk of the borrower is the delay in the principal debt or interest for 90 days or more during the first year of life of the loan. The main drawback of this approach is the unaccounted fact of a change in the individual probability of a default. The article explores the actual for commercial banks problem of assessment credit not only for the first credit life year, but for the whole period of its life. This condition is fixed in the IFRS 9 standard. Also according to IFRS 9, the amount bank reserves should depend on the dynamics of the level of risk, which is determined by the estimation of the default credit risk throughout the life of the loan (lifetime PD). The article suggests a lifetime PD model constructing on the data of the quality of credit history payment (future engineering) and the phase

of the macroeconomic cycle. To solve the problem the method of logistic regression with regularization is used in the article. The quality of the model is estimated by calculating the Gini coefficient. The given calculations of probability of default were made on the data provided by the regional retail bank.

Keywords: credit risk, Basel II, IFRS 9, behavioral scoring, probability of default, lifetime PD, future engineering, logistic regression with regularization.

Широбокова Маргарита Александровна, аспирант
ФГБОУ ВО «Удмуртский государственный университет»
426034, Россия, г. Ижевск, ул. Университетская, 1 (корп. 4)
E-mail: shirobokova.margarita@mail.ru

Shirobokova M.A., postgraduate student
Udmurt State University
Universitetskaya st., 1/4, Izhevsk, Russia, 462034
E-mail: shirobokova.margarita@mail.ru