

УДК 336.717.061

*М.А. Широбокова, А.В. Лётчиков***ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ К РАСЧЕТУ РЕНТАБЕЛЬНОСТИ СОБСТВЕННОГО КАПИТАЛА БАНКА ПО КРЕДИТНОМУ ДОГОВОРУ**

Требования к более точной оценке индивидуального риска заемщика усложнились с введением Базель II и IFRS 9. Все более часто такую оценку риска производят с помощью построения скоринговых моделей, однако, как правило, критерием качества построенных моделей выступает коэффициент *Gini*, и совсем не исследуется влияние моделирования на финансовую составляющую, а именно на рентабельность капитала, которая выступает базой для ведения бизнеса в области кредитования. В связи с этим в статье предлагается методика оценки рентабельности капитала без учета риска и ее усложнение за счет принятия ко вниманию индивидуального риска заемщика. Само построение динамической модели оценки кредитного риска в статье рассматривается на основании моделей выживания, построенных методами машинного обучения. Проблема учета цензурированных данных решается с помощью специфического построения переменных для модели и методов, которые учитывают цензурирование: логистической регрессии, модели пропорциональных рисков Кокса, модели случайного леса выживаемости. На примере данных регионального коммерческого банка оценивается рентабельность собственного капитала и производится их сравнение в зависимости от выбора модели оценки риска. В результате исследования делается вывод о необходимости применения методики расчета рентабельности собственного капитала с учетом риска, оцененным методом машинного обучения.

Ключевые слова: кредитный риск, коммерческий банк, рентабельность собственного капитала, вероятность дефолта, логистическая регрессия, модель Кокса, модель случайного леса выживаемости.

DOI: 10.35634/2412-9593-2020-30-2-230-235

Оценка индивидуального риска заемщика, основанная на применении скоринговой карты, на практике дает возможность ранжировать кредитные заявки по риску конечных потерь, однако не учитывает финансовые показатели кредитного договора, такие как сумма кредита, ставка кредитования, длительность кредита. Поскольку для высокорискованных кредитов, как правило, банки предлагают более высокие ставки кредитования, для банка может быть выгодно выдача более рискованных и одновременно более доходных кредитов. При этом эффективность кредитного договора определяется маржой банка, вычисляемой как предполагаемый процентный доход банка по кредиту в год за минусом трансфертных платежей и риск-маржи, рассчитанной из конечных потерь на годовой период. Однако такой подход не учитывает эффективность использования собственного капитала и поэтому требует использования других финансовых показателей.

В этом случае важным показателем эффективности инвестиционных вложений банка является рентабельность собственного капитала (*ROE*), показывающая размер годовой прибыли, полученной на единицу собственного капитала банка. Существует несколько алгоритмов расчета этого показателя, но наиболее точным является метод, основанный на построении потока возвращаемых сумм, включающих в себя потоки прибыли после выплаты налогов и стоимости привлеченного капитала. При этом при расчете возвращаемых сумм используются внешние для кредитного договора финансовые показатели такие, как структура капитала банка, ставка налога на прибыль и трансфертная цена привлеченного капитала, включающая в себя расходы по привлечению капитала. Одной из проблем применения такого метода является использование довольно продвинутых численных методов для вычисления показателя *ROE*, что в принципе преодолимо существующими информационными и вычислительными ресурсами.

Другой проблемой, которая возникает при расчете *ROE* на основе потока возвращаемых сумм, является учет риска потерь по кредитному договору. Поскольку практически невозможно построить неслучайный поток возвращаемых сумм для дефолтных кредитных договоров, одним из возможных подходов к решению этой проблемы является рассмотрение дефолта как страхуемого кредитного события. В этом случае на каждый кредитный договор рассчитывается ожидаемая величина возможных конечных потерь, на сумму которой производится резервирование собственного капитала. За счет этих средств в момент дефолта производится возмещение потерь собственного и привлеченного капиталов. При таком алгоритме расчета является важным достаточно точная индивидуальная оценка вероятности дефолта по кредитному договору в фиксированный момент времени. Поэтому интерес-

ным на сегодняшний день является исследование влияния различных методов расчета временной функции вероятности дефолта на оценку рентабельности собственного капитала по индивидуальному договору кредитования.

На текущий момент одной из приоритетных задач в банковском секторе является управление кредитным риском, для оценки которого все чаще применяются скоринговые модели, позволяющие индивидуально оценить риск заемщика на основе имеющейся информации. При этом в соответствии с развитием банковского сектора и наличием законодательной основы в виде рекомендаций Базель II и IFRS 9 [1; 2] становится интересным вопрос оценки кредита не только на первый год жизни кредита, а на весь срок его жизни, учитывая факт того, что с течением жизни кредита происходит изменение индивидуальной вероятности выхода в дефолт. Однако зачастую для моделирования значения PD используются только те кредиты, которые прожили год и более от даты оценки риска. В этом случае кредиты, которые были закрыты (вне зависимости от причины закрытия: наступление срока погашения по договору, досрочное погашение, реструктуризация при одновременной выдаче нового кредита и др.) в течение рассматриваемого года не учитываются при моделировании либо считаются не дефолтными. Таким образом, часть полезной информации, способной дать положительный эффект на качество модели, теряется или используется в ненадлежащем виде. Для того чтобы получить наиболее корректную оценку индивидуального кредитного риска на весь срок его жизни, необходимо использовать в том числе «неполные» данные, которые называются цензурированными [3]. В случае с кредитным скорингом имеет место цензурирование справа [4]: на конец периода наблюдения известен факт возникновения дефолта каждого клиента, однако произойдет ли дефолт за пределами исследования неизвестно. Методика построения моделей методами машинного обучения с учетом цензурированных данных была исследована ранее в [5], где выбор модели производился на основе ROC -кривой и значения коэффициента $Gini$. Однако, помимо оценки качества модели как таковой, отдельный интерес представляет влияние способа моделирования на рентабельность капитала, особенно при усложнении модели и учете цензурированных данных, что и послужило целью настоящего исследования.

Определим денежный поток по отдельному договору кредитования. Пусть D – сумма выдачи по кредиту, величина основного долга в момент выдачи (руб.), T – длительность кредита (месяцев), R – ставка по договору (% годовых). Обозначим через $C_D(t)$ поток возвращаемых сумм основного долга в месяц t (руб.), $t = 0, 1, \dots, T$. Тогда $D(t)$ – величина основного долга в конце месяца t (руб.), $t = 0, 1, \dots, T$ рассчитывается по формулам: $D(0) = D - C_D(0)$, $D(t) = D(t-1) - C_D(t)$, $t = 1, \dots, T$. Соответственно, поток договорных процентов на основной долг $C_R(t)$ (руб.), $t = 1, \dots, T$ вычисляется как $C_R(t) = \frac{R}{12} \cdot D(t-1)$. В итоге $C(t)$ – поток возвращаемых сумм по кредиту (руб.), $t = 0, 1, \dots, T$, определяется следующим образом: $C(0) = C_D(0)$, $C(t) = C_D(t) + C_R(t)$, $t = 1, \dots, T$.

Для расчета рентабельности собственного капитала без учета риска построим денежные потоки возврата собственного капитала и чистой прибыли по следующему алгоритму.

В начале рассматриваемого кредитного договора первичная сумма долга $D(0)$ считается инвестиционным капиталом, из которого собственный капитал банка составляет $w \cdot D(0)$, а привлеченное фондирование – $(1-w) \cdot D(0)$, где w – доля собственного капитала в структуре капитала банка. В каждый месяц t из возвращаемой суммы $C_D(t)$ основного долга по кредитному договору выделяется сумма на возврат собственного капитала пропорционально доли w . Таким образом, строится поток возврата собственного капитала $RC(t) = w \cdot C_D(t)$. Поскольку оставшаяся сумма идет в погашение привлеченного капитала, долг по привлеченному капиталу к концу месяца t составит $(1-w) \cdot D(t)$.

Для расчета потока чистой прибыли строится поток трансфертных процентов, определяемый по формуле $C_{tr}(t) = \frac{Rt}{12} \cdot (1-w) \cdot D(t-1)$, где Rt – трансфертная ставка (% годовых). Чистая прибыль из получаемых процентов определяется как получаемые проценты по договору за минусом трансфертных процентов и налога на прибыль: $NP(t) = (1-n) \cdot (C_R(t) - C_{tr}(t))$, где n – ставка налога на прибыль (% годовых).

Для расчета рентабельности капитала строится денежный поток $K(0), K(1), K(2), \dots, K(T)$ по следующему правилу: $K(0) = -w \cdot D(0)$, $K(t) = RC(t) + NP(t)$, $t = 1, \dots, T$. Расчетная величина рентабельности собственного капитала ROE находится как внутренняя ставка доходности денежного потока $K(0), K(1), K(2), \dots, K(T)$, умноженная на 12, и является решением уравнения:

$$\sum_{t=0}^T K(t) \cdot \left(1 + \frac{ROE}{12}\right)^{-t} = 0. \#(1)$$

Для расчета рентабельности собственного капитала с учетом риска ключевым параметром выступает оценка функции вероятности дефолта. В отличие от классического подхода [1; 6; 7] под дефолтом будем понимать выход в просрочку более 15 дней по кредитному договору. Тогда основные функции, необходимые для оценки показателя ROE , будут следующие: $P(t)$ – вероятность первого выхода в дефолт на t -том месяце жизни договора, $Q(t)$ – вероятность невыхода в дефолт в течение t месяцев.

На каждый месяц t ($t = 1, \dots, T$) определяется $Y(t)$ – размер условной страховой выплаты в момент t , погашающей долг по собственному капиталу и привлеченному капиталу с начисленными трансфертными процентами: $Y(t) = D(t-1) \cdot \left(1 + (1-w) \cdot \frac{Rt}{12}\right)$. С учетом вычисленных вероятностей рассчитывается средняя величина Y выплат по рисковому событию дефолта, дисконтированных на момент выдачи кредита:

$$Y = P(1) \cdot Y(1) \cdot v + P(2) \cdot Y(2) \cdot v^2 + \dots + P(T) \cdot Y(T) \cdot v^T, \#(2)$$

где $v = \frac{1}{(1+Rt)^{1/12}}$ – дисконтирующий множитель, Rt – трансфертная ставка (в годовых). Соответственно, резервируемая на момент выдачи сумма RP ожидаемых потерь по кредитному договору равна $RP = Y \cdot LGD$, где LGD – коэффициент уровня потерь при дефолте.

Для расчета рентабельности собственного капитала с учетом риска строится усредненный денежный поток $U(0), U(1), U(2), \dots, U(T)$ по следующему правилу: $U(0) = K(0) - RP$, $U(t) = Q(t) \cdot K(t) + P(t) \cdot w \cdot D(t-1)$, $t = 1, \dots, T$. Расчетная величина рентабельности капитала ROE с учетом риска находится как внутренняя ставка доходности денежного потока $U(0), U(1), U(2), \dots, U(T)$, умноженная на 12, и является решением уравнения:

$$\sum_{t=0}^T U(t) \cdot \left(1 + \frac{ROE}{12}\right)^{-t} = 0. \#(3)$$

Как видно из выписанного, применение построенной методики расчета рентабельности капитала с учетом риска требует адекватной оценки кредитного риска, рассчитанной индивидуально по каждому заемщику. Как показывает практика, существует несколько методов, позволяющих оценить риск с учетом цензурированных данных [5; 8]. В рамках настоящего исследования рассмотрим логистическую модель, модель Кокса и случайный лес выживаемости.

Логистическая регрессия

Для логистической регрессии формула расчета скорингового балла p основывается на вычислении логарифмов отношения шансов дефолтного и не дефолтного кредитов. На основе найденного с помощью логистической регрессии значения PD для применения формулы (3) расчета рентабельности капитала с учетом риска к оценке кредитного риска важной является оценка вероятностей дефолта $P(t)$ на каждый момент времени t . Согласно исследованию [9], статистические данные о выходах в дефолт (просрочку более 15 дней) показывают, что распределение случайной величины τ месяца выхода в дефолт адекватно моделируется геометрическим распределением. Тогда следующие величины определяются так: $PD = p$ – вероятность выхода в 15+ в течение года, $q = (1 - PD)^{1/12}$ – вероятность невыхода в 15+ в течение месяца, $P(t) = (1 - q) \cdot q^{t-1}$ – вероятность выхода в 15+ на t -том месяце жизни договора, $Q(t) = q^t$ – вероятность невыхода в 15+ в течение t месяцев.

Модель пропорциональных рисков Кокса

Модель пропорциональных рисков Кокса является моделью анализа выживаемости, для которой функция выживания однозначно определяется функцией риска $h(t|x)$, в рассматриваемом случае кредитного скоринга понимаемая как условная вероятность того, что событие дефолта произойдет в бесконечно малом интервале $[t, t + dt]$, при условии того, что на момент t событие дефолта не произошло. Тогда взаимосвязь между вероятностью выживания $P(t)$ и уровнем риска $h(t|x)$ записывается следующим образом: $S(0) = 1$ – вероятность того, что до момента $t = 0$ дефолта не случилось, $S(1) = (1 - h(1))$ – вероятность того, что до момента $t = 1$ дефолта не случилось, ... $S(t) = S(t-1)(1 - h(t)) = (1 - h(1)) \cdot \dots \cdot (1 - h(t))$ – вероятность того, что до момента t дефолта не случилось, или $Q(t)$ – вероятность невыхода в 15+ в течение t месяцев. Тогда $P(t) = (1 - h(1)) \cdot \dots \cdot (1 - h(t-1)) \cdot h(t)$ – вероятность того, что дефолт в 15+ случится на t -том месяце жизни договора.

Случайный лес выживаемости

Одним из методов реализации анализа выживаемости является случайный лес выживаемости, который показывает вероятность наступления события, а также вероятностные характеристики времени до его наступления. Случайный лес выживаемости представляет собой ансамбль деревьев решений, как и случайный лес, однако для расщепления промежуточных узлов дерева применяется логранговый критерий. Среднее значение функции риска по всем деревьям является усредненным ответом по всему ансамблю деревьев. Тогда: $P(t) = \overline{H_B(t|x)}$ – вероятность выхода в 15+ на t -том месяце жизни договора, $Q(t) = P(t) \cdot (1 - P(t))^{t-1}$ – вероятность невыхода в 15+ в течение t месяцев, где $H_B(t|x)$ – ответ каждого отдельного дерева при заданных параметрах t и x , а B – число деревьев.

На основе описанных методов расчета оценки риска применим методику расчета рентабельности собственного капитала на базе регионального розничного банка, для этого предварительно построим модели оценки PD . В качестве исходных данных для построения 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, то есть целевой функцией служила оценка выхода в дефолт 15+ в течение последующих 12 месяцев от даты среза. Для каждой даты среза был сформирован портфель действующих кредитов: открытых кредитов по состоянию на дату среза и с текущей оценкой бинарного признака дефолта по кредиту. Для указанной выборки на основе кредитной истории были рассчитаны переменные, а также произведено разделение выборки на обучающую и валидирующую в соответствии с [11]. Качество построенных моделей оценивалось с помощью расчета площади под ROC -кривой и коэффициента $Gini$. Графики ROC -кривых для построенных моделей следующие.

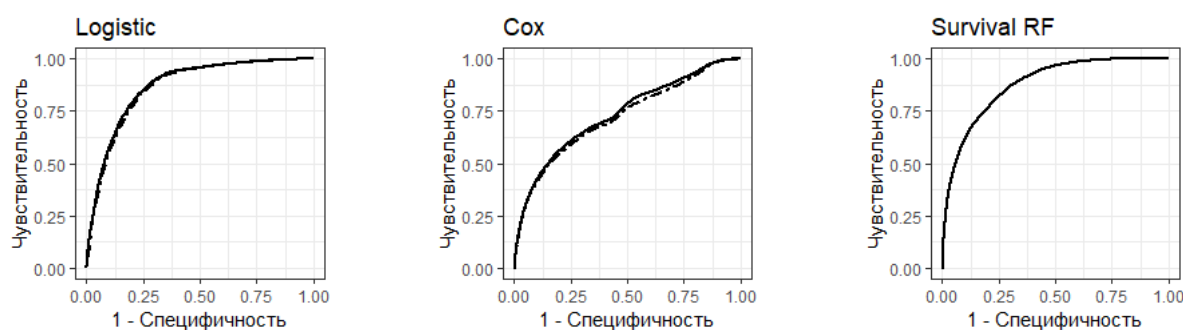


Рис. 1. Графики ROC -кривых для моделей логистической регрессии, модели Кокса, случайного леса выживаемости

Соответствующие значения коэффициента $Gini$ для обучающей выборки, валидирующей выборки и тестовой выборки представлены в табл. 1.

Таблица 1

Значения коэффициента $Gini$ для моделей логистической регрессии, Кокса и случайного леса выживаемости

	$Gini$ на обучающей выборке	$Gini$ на валидирующей выборке	$Gini$ на тестовой выборке
Логистическая регрессия	0.6832	0.6883	0.6949
Модель Кокса	0.4687	0.4399	0.4479
Случайный лес выживаемости	0.7544	0.7538	0.7570

По графикам ROC -кривых и значениям коэффициентов $Gini$ можно сделать вывод, что наиболее точной и в то же время стабильной является модель случайного леса выживаемости. При это модель Кокса дает наихудшую оценку, что связано с тем, что основным предположением модели Кокса является пропорциональность в изменении риска с течением времени, что оказалось не верным в случае оценки кредитного риска. Однако интерес представляет не только оценка качества самой модели, но и расчет рентабельности капитала с учетом риска.

Для каждого отдельного кредита $i \in N$ определим его индивидуальную рентабельность ROE_i с учетом риска, расчеты произведем на наиболее свежем срезе портфеля: открытые кредиты на 31-10-2018. Всего в выборку попало $N=39260$ кредитов, которые были выданы с 01-01-2012 по 31-10-2018. По каждому договору кредитования имеются описанные ранее входные параметры модели, а также индивидуальные данные заемщика. В соответствие с целевой функцией сравним полученные оценки показателя ROE в зависимости от $P(12)$, вероятности того, что произойдет выход в дефолт 15+ в течение следующего года, которая была оценена логистической моделью, моделью Кокса и моделью случайного леса выживаемости.

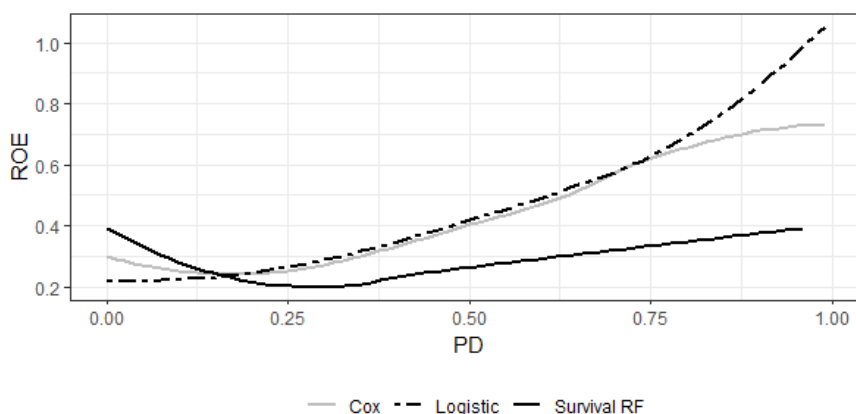


Рис. 2. Графики зависимости ROE от оценки риска, рассчитанной моделью логистической регрессии, моделью Кокса, моделью случайного леса выживаемости

Наблюдается явная зависимость: при увеличении риска по кредиту увеличивается рентабельность собственного капитала. При оценке моделями Кокса и случайного леса выживаемости для высокорискованных кредитов ROE увеличивается вплоть до значений близких к единице, что действительно возможно за счет более высоких процентных ставок по кредиту, однако неверно с точки зрения оценки риска, так как данные кредиты, вероятнее всего, станут дефолтными или вовсе невозвратными. При этом по модели случайного леса выживаемости для высокорискованных кредитов рентабельность собственного капитала также растет, но менее быстрыми темпами и при этом положительна, что более объективно оценивает инвестиционную эффективность кредитного договора.

Таким образом, данный подход позволяет наглядно оценить размер рентабельности собственного капитала в зависимости от модели оценки риска, а в дополнении к критерию качества *Gini* обосновывает необходимость использования более сложных моделей машинного обучения для оценки кредитного риска и является дополнительным инструментом для выбора окончательной модели. Дальнейший интерес данного исследования заключается в оценке кредитных договоров с досрочным погашением и их влиянием на общую рентабельность капитала.

СПИСОК ЛИТЕРАТУРЫ

1. Приказ Минфина России от 27.06.2016 № 98н «О введении документов Международных стандартов финансовой отчетности в действие на территории Российской Федерации и о признании утратившими силу некоторых приказов Министерства финансов Российской Федерации» (Зарегистрировано в Минюсте России 15.07.2016 № 42869). URL: <http://publication.pravo.gov.ru/Document/View/0001201607180025>.
2. Энциклопедия финансового риск-менеджмента / под ред. канд. экон. наук А.А. Лобанова и А.В. Чугунова. М.: Альпина Паблишер, 2009. 932 с.
3. Кокс Д.Р., Оукс Д. Анализ данных типа времени жизни / пер. с англ. М.: Финансы и статистика, 1988. 191 с.
4. Man R. Survival analysis in credit scoring: A framework for PD estimation / Twente: University of Twente, 2014. URL: <https://pdfs.semanticscholar.org/b4e3/ee5a66e180ba6d3cc7174ee232799cfd1831.pdf>.
5. Лётчиков А.В., Матвеев Р.Ю., Широбокова М.А. Решение проблемы цензурированных данных при моделировании оценки индивидуального кредитного риска // Вестн. Удм. ун-та. Сер. Экономика и право. 2019. Т. 29, вып. 1. С. 34-41.
6. Алескеров Ф.Т., Белоусова В.Ю., Сердюк М.Ю., Солодков В.М. Стереотипы поведения российских банков // Банковское дело. 2008. №7. С. 44-50.

7. Алескеров Ф.Т., Андриевская И.К., Пеникас Г.И., Солодков В.М. Анализ математических моделей Базель II. 2-е изд., испр. М.: ФИЗМАТЛИТ. 2013. 295 с.
8. Воронцов К.В. Лекции по линейным алгоритмам классификации. Электрон. док., 2009. URL: <http://www.machinelearning.ru/wiki/images/6/68/voron-ML-Lin.pdf>.
9. Лётчиков А.В. Расчет индивидуального кредитного риска с применением модели геометрического распределения // Вестн. Удм. ун-та. Сер. Экономика и право. 2018. Т. 28, вып. 2. С. 208-213.
10. Широбокова М.А. Модель оценки риска дефолта на всем протяжении жизни кредита // Вестн. Удм. ун-та. Сер. Экономика и право. 2018. Т. 28, вып. 2. С. 228-233.
11. Ishwaran H. The effect of splitting on random forests / The Author(s), 2014. URL: <https://link.springer.com/content/pdf/10.1007%2Fs10994-014-5451-2.pdf>
12. Mogensen U.B., Ishwaran H., Gerds, T.A. Evaluating random forests for survival analysis using prediction error curves. University of Copenhagen, 2012. URL: https://ifsv.sund.ku.dk/biostat/annualreport/images/4/4d/Research_Report_10-8.pdf.
13. Груздев А.В. Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес. М.: ДМК Пресс, 2018. 642 с.
14. Чистяков С.П. Случайные леса: обзор // Тр. Карельского науч. центра РАН / Ин-т прикладных матем. исслед. Карельского науч. центра РАН. Петрозаводск, 2013. С. 125-126.

Поступила в редакцию 28.02.2020

Широбокова Маргарита Александровна, старший преподаватель

E-mail: shirobokova.margarita@mail.ru

Лётчиков Андрей Владимирович, доктор физико-математических наук, профессор

E-mail: letchi@udm.ru

ФГБОУ ВО «Удмуртский государственный университет»

426034, Россия, г. Ижевск, ул. Университетская, 1 (корп. 4)

M.A. Shirobokova, A.V. Letchikov

APPLICATION OF MACHINE LEARNING METHODS TO THE CALCULATION OF BANK'S RETURN ON EQUITY UNDER A LOAN AGREEMENT

DOI: 10.35634/2412-9593-2020-30-2-230-235

The requirements for a more accurate assessment of the individual risk of a borrower became more complicated with the introduction of Basel II and IFRS 9. Such risk assessment is more and more often carried out using the construction of scoring models, however, as a rule, the Gini coefficient acts as a quality criterion for the constructed models, and the influence of modeling on financial component, namely on the return on equity, which acts as the basis for doing business in the field of lending, is not investigated at all. In this regard, the article proposes a methodology for assessing the return on equity without taking into account risk and its complication by taking into account the individual risk of a borrower. The construction of a dynamic model for assessing credit risk in the article is considered on the basis of survival models constructed by machine learning methods. The problem of accounting for censored data is solved using specific construction of variables for the model and methods that take into account censorship: logistic regression, Cox proportional risk model, random survival forest model. On the example of the data of a regional commercial bank, the return on equity is estimated and compared, depending on the choice of a risk assessment model. The result of the study is the conclusion that it is necessary to apply the methodology for calculating the return on equity taking into account risk assessed by the machine learning method.

Keywords: credit risk, commercial bank, return on equity, probability of default, logistic regression, Cox model, random survival forest model.

Received 28.02.2020

Shirobokova M.A., Senior lecturer

E-mail: shirobokova.margarita@mail.ru

Letchikov A.V., Doctor of Physics and Mathematics, Professor

E-mail: letchi@udm.ru

Udmurt State University

Universitetskaya st., 1/4, Izhevsk, Russia, 426034