Препринт «Финансы и кредит» №12, 2021

Financial Control

BENCHMARKING TEST FOR THE CALCULATED VALUES OF DEFAULT PROBABILITIES OBTAINED BY APPLICATION OF RATING MODELS

Mikhail V. POMAZANOV

Ph.D., Associate Professor, School of Finance, Faculty of Economic Sciences, National Research University Higher School of Economics, Moscow, Russia
m.pomazanov@hse.ru
ORCID: 0000-0003-3069-1511

JEL Classification: C58, G17, G28

Abstract

Importance Validation of the consistency of rating model forecasts.
Objectives To provide rating model developers and validators with a practical fundamental test for benchmarking the calculated default probabilities resulting from the application of the models used in the rating system.
Methods The classical interval approach of testing statistical hypotheses, focused on the subject area of calibration of rating systems.
Results In addition to the generally accepted tests for the correspondence of the predicted probabilities of default of credit risk objects to the historically realized values, a new statistical test is proposed that corrects the shortcomings of the generally accepted ones, focused on "diagnosing" the consistency of the implemented discrimination of objects by the rating model. Examples of recognizing the reasons for a negative test result and negative consequences for lending are given while maintaining the current settings of the rating model. The proposed method, in addition to the bias in the assessment of the total frequency of defaults in the loan portfolio, makes it possible to objectively reveal the inadequacy of discrimination against borrowers with a calibrated rating model, to diagnose the "disease" of the rating model. Moreover, this does not require the completeness of statistics in each rating category, which expands the scope of applicability of comparative analysis on historical data with a small number of defaults that occurred during the validation period.
The scope of the results is the process of internal validation by the bank of its own rating models, which is required by the Bank of Russia for approaches based on internal ratings.

Conclusions and Relevance It is concluded that the new practical benchmark test allows, at a given level of confidence and available historical data, to reject the hypothesis about the consistency of assessing the probability of default by the rating model, and the test has the advantage of practical interpretability, based on its results, it is possible to draw a conclusion about the direction of the model correction.

Keywords: credit risk, probability of default, benchmarking, validation, statistical test, Gini, ROC curve

Финансовый контроль

СОПОСТАВИТЕЛЬНЫЙ ТЕСТ ДЛЯ РАССЧИТАННЫХ ЗНАЧЕНИЙ ВЕРОЯТНОСТЕЙ ДЕФОЛТА, ПОЛУЧЕННЫХ В РЕЗУЛЬТАТЕ ПРИМЕНЕНИЯ РЕЙТИНГОВЫХ МОДЕЛЕЙ

Михаил Вячеславович ПОМАЗАНОВ

к.ф.-м.н., доцент, Школа финансов, Факультет экономических наук, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия
m.pomazanov@hse.ru
ORCID: 0000-0003-3069-1511
SPIN-код: 1763-5033

УДК 336.71(075.8)

JEL Classification: C58, G17, G28

Аннотация

Предмет. Валидация состоятельности прогнозов рейтинговых моделей.

Цели. Дать разработчикам и валидаторам рейтинговых моделей практичный фундаментальный тест для сопоставительного анализа рассчитанных значений вероятности дефолта, полученных в результате применения моделей, используемых в рейтинговой системе.

Методология. Классический интервальный подход проверки статистических гипотез, ориентированный на предметную область калибровки рейтинговых систем.

Результаты. В дополнение к общепринятым тестам на соответствие прогнозных вероятностей дефолта объектов кредитного риска реализованным исторически значениям предложен новый статистический тест, исправляющий недостатки общепринятых, ориентированный на «диагностику» состоятельности реализованной дискриминации объектов рейтинговой моделью. Даны примеры распознавания причин отрицательного результата тестирования и негативных последствий для кредитования при сохранении текущих настроек рейтинговой модели. Предложенный метод, кроме смещенности оценки общей частоты дефолтов в кредитном портфеле, позволяет объективно выявить неадекватность дискриминации заемщиков калиброванной рейтинговой моделью, диагностировать «болезнь» рейтинговой модели. Причем для этого не требуется полнота статистики в каждом рейтинговом разряде, что дает расширение области применимости сопоставительного анализа на исторических данных с небольшим количеством дефолтов, произошедших за валидационный период.

Областью применения результатов является процесс проведения внутренней валидации банком собственных рейтинговых моделей, требуемый Банком России к подходам, основанным на внутренних рейтингах.

Выводы. Сделан вывод, что новый практичный сопоставительный тест позволяет на заданном уровне доверия и доступных исторических данных отвергнуть гипотезу о состоятельности оценки вероятности дефолта рейтинговой моделью, причем тест обладает преимуществом практической интерпретируемости, по его результатам можно сделать вывод о направлении коррекции модели.

Ключевые слова: кредитный риск, вероятность дефолта, сопоставительный анализ, валидация, статистический тест, Джини, ROC-кривая

Введение

На основании п.14.2 Положения Банка России от 6 августа 2015 г. N 483-П[1] "О порядке расчета величины кредитного риска на основе внутренних рейтингов", в рамках проведения внутренней валидации банк должен не реже одного раза в год осуществлять сопоставительный анализ рассчитанных значений вероятности дефолта (PD), полученных в результате применения моделей, используемых в рейтинговой системе, с фактической частотой реализованных дефолтов заемщиков для каждого разряда рейтинговой шкалы. Требования Банка России к подходам, основанным на внутренних рейтингах (ПВР), являются результатом переработки и адаптации к Российской банковской системе требований Международного регулятора[2] к Продвинутому подходу, а также рекомендаций методов валидации ПВР-моделей[3].

Статистические тесты прогноза вероятности дефолта в рейтинговых моделях представлены в широком списке научных работ по направлению валидации. В работе [1] обсуждаются техники, которые можно использовать для удовлетворения количественных нормативных требований, однако их целесообразность зависит от конкретных условий, при которых они применяются.

В контексте работы [2] предлагается простой механизм для сравнения эффективности основных рейтинговых агентств и других систем оценки рейтингов, таких как внутренние рейтинговые системы коммерческих банков в рамках режима Базель II, представлен простой критерий проверки оценок PD, который поможет в мониторинге эффективности различных систем оценки кредита, участвующих в оценке приемлемого обеспечения, лежащего в основе операций денежно-кредитной политики Евросоюза.

В работе [3] представлены оценки долгосрочной вероятности дефолта, согласующиеся с подходом оценки экономического капитала, лежащего в основе Базель-2. Авторы обнаружили, что использование простого среднего значения частоты дефолтов в качестве оценки может привести к недооценке долгосрочной вероятности дефолта, предлагаются альтернативные способы установления доверительных интервалов (CI). В большинстве случаев использование CI, построенных на основе предложенных оценок максимального правдоподобия, приводит к меньшему количеству ошибок при проверке гипотез.

[1] https://cbr.ru/faq_ufr/dbrnfaq/doc/?number=483-П
[2] BCBS, Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version, 2006
[3] BCBS, Studies on the Validation on Internal Rating Systems, 2005

В статье [4] представлен обобщенный подход, как использовать набор односторонних многомерных тестов для фактического обнаружения недооценки кредитного риска рейтинговыми моделями. Существующие односторонние многомерные тесты основаны на оценке рейтинговых характеристик в каждом из рейтинговых классов системы отдельно. Новизна представленных тестов заключается в совместной оценке результатов во всех рейтинговых классах. Показано, что новые тесты превосходят установленный односторонний многомерный тест Вестфал-Волнгер [5] с точки зрения мощности для различных рейтинговых портфелей Standard & Poor's.

Классический статистический подход (биномиальный тест) для проверки гипотез может быть легко модифицирован для целей калибрации рейтинговой модели. В этом случае предполагается независимость событий возникновения дефолта заемщиков, что в общем случае верно, поскольку рейтинговые системы, как правило, используются для оценки вероятности дефолта некредитных организаций, либо физических лиц, степень взаимосвязи между которыми, в отличие от рынка межбанковского кредитования, довольно низкая. Пусть имеется k событий дефолта среди n заемщиков данного класса, наблюдаемая частота дефолта будет $p^* = \frac{k}{n}$. Фиксируется уровень доверия $\boldsymbol{\alpha}$ и строится доверительный интервал для модельной оценки величины вероятности дефолта PD

$$\Omega_\alpha = \left[p^* - \boldsymbol{t_\alpha}\sqrt{\frac{p^*\cdot(1-p^*)}{\boldsymbol{n}}}, p^* + \boldsymbol{t_\alpha}\sqrt{\frac{p^*\cdot(1-p^*)}{\boldsymbol{n}}}\right],$$

где $\boldsymbol{t_\alpha = N^{-1}\left(\frac{1+\alpha}{2}\right)}$,

N^{-1}– обратное нормальное распределение.

Утверждается, что биномиальный тест отвергает гипотезу о состоятельности оценки PD с уровнем достоверности $\boldsymbol{\alpha}$ если $PD \notin \Omega_\alpha$.

Необходимо отметить, что биномиальное распределение может быть аппроксимировано нормальным распределением если $k > 10$ $u\, n - k > 10$.

Тест Хи-квадрат (Хосмер-Лемешоу) [6] позволяет проверять нулевую гипотезу о совпадении распределения событий с некоторым заданным распределением. События предполагаются независимыми и одинаково распределенными, результаты каждого из событий должны быть взаимоисключающими. Хи-квадрат вычисляется следующим образом:

$$T_k = \sum_{i=0}^{k} \frac{(n_i p_i - \theta_i)^2}{n_i p_i (1-p_i)},$$

где $p_0, \dots, p_k$ – модельно оцененные вероятности дефолта в k+1 рейтинговом разряде, при этом требуется несмещенность PD, т.е.

$\sum_{i=0}^{k} n_i p_i = \sum_{i=0}^{k} \theta_i$,

где n_i – количество заемщиков в рейтинговом интервале i,

θ_i – количество заемщиков в дефолте в рейтинговом интервале i.

В соответствии с центральной предельной теоремой при $n_i \to \infty$ для всех i, T_k стремится к распределению χ^2_{k-1}. Значение p-value теста Хи-квадрат может служить как мера точности оценки вероятности дефолта: чем ближе p-value к нулю, тем хуже оценка.

В случаях, когда ожидаемое значение в знаменателе оказывается малым (что означает либо малую вероятность, либо малое количество наблюдений), нормальная аппроксимация

полиномиального распределения может не работать, тогда, возможно, стоит использовать G-тест [7]. Который асимптотически имеет такое же распределение $G_k \to \chi^2_{k-1}$,

$$G_k = 2\sum_{i=0}^{k} \theta_i \cdot ln\left(\frac{\theta_i}{n_i p_i}\right)$$

Как показано в монографии [8], достаточное число дефолтов для практически значимого совпадения результатов двух вышеупомянутых тестов
$\sum_{i=0}^{k} \theta_i \approx 1000.$
Банкам предлагается самостоятельно устанавливать пороговые значения теста Хосмер-Лемешоу.
Следующий общепринято используемый тест носит имя Шпигельхальтера [9]. Критерий не предполагает установки конечного числа рейтинговых разрядов, в каждом из которых предполагается существенным асимптотическое требование большого числа измерений. Тест Шпигельхальтера основан на вычислении величины среднеквадратичной ошибки MSE
$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)^2,$
где p_i – модельно оценённая вероятность дефолта заемщика $i, i = 1 \dots N$, а y_i наблюдаемый индикатор его дефолта на сопоставительной генеральной совокупности

$$y_i = \begin{cases} 1, & \text{если заемщик } i \text{ в дефолте} \\ 0, & \text{если заемщик } i \text{ не в дефолте} \end{cases}.$$

Рассматриваются гипотезы:

$H_0: P(y_i = 1) = p_i \ \forall_i$ – нулевая гипотеза

$H_1: P(y_i = 1) \neq p_i$ хотя бы для одного i – альтернативная гипотеза

При справедливости нулевой гипотезы не трудно показать, что

$$E(MSE) = \frac{1}{N}\sum_{i=1}^{N} p_i \cdot (1 - p_i),$$

$$D(MSE) = \frac{1}{N^2}\sum_{i=1}^{N}(1 - 2p_i)^2 p_i(1 - p_i).$$

Учитывая асимптотическую нормальность распределения значения

$$Z = \frac{MSE - E(MSE)}{\sqrt{D(MSE)}} \in \mathbf{N}(0,1), N \to \infty,$$

можно сформировать доверительный интервал для реализованных бинарных измерений $\{y_i\}$:

$$\Omega_\alpha = \left[\frac{\sum_{i=1}^{N}(y_i - p_i)^2}{\sqrt{\sum_{i=1}^{N}(1-2p_i)^2 p_i(1-p_i)}} - t_\alpha, \frac{\sum_{i=1}^{N}(y_i - p_i)^2}{\sqrt{\sum_{i=1}^{N}(1-2p_i)^2 p_i(1-p_i)}} + t_\alpha\right]$$

Утверждается, что тест Шпигельхальтера отвергает гипотезу H_0 о состоятельности оценки p_i с уровнем достоверности $\boldsymbol{\alpha}$, если

$\frac{\sum_{i=1}^{N} p_i \cdot (1-p_i)}{\sqrt{\sum_{i=1}^{N} (1-2p_i)^2 p_i (1-p_i)}} \notin \Omega_\alpha$.

Общим недостатком представленных тестов на состоятельность оценки модельных PD является отсутствие интерпретируемости для понимания того, что не так с рейтинговой моделью, если тест не пройден. Не является ли отвержение нулевой гипотезы результатом статистической погрешности, особенно для теста Хосмер-Лемешоу? Для теста Шпигельхальтера встает вопрос, почему рассматривается именно среднеквадратическое отклонение (пространственная метрика L_2), почему не L_1 (разность модулей или абсолютное отклонение) или L_4 и т.п.? Может быть другая конструкция подобного теста «спасет» модель? Или на любую модель можно всегда найти такую конструкцию, которая эту модель «завалит»?

В следующем пункте представлен и обоснован тест, в котором вопросы сингулярного воздействия статистической погрешности или сомнительности конструкции не возникают. По отрицательным результатам теста можно поставить общий «диагноз болезни» рейтинговой модели, предложить «лечение». Это демонстрируется на примерах.

Основной результат

Пусть сопоставительная база кредитного портфеля, содержащая рейтинги (эквивалентно PD) компаний и событий дефолта, содержит n полных лет. Фиксируются не дефолтные рейтинги (разряды) на начало каждого периода $\boldsymbol{RN_i^n, i \in 1 \ldots N_n}$, за которыми не следовал дефолт в течение года, а также рейтинги $\boldsymbol{RD_d^n, d \in 1 \ldots D_n}$ за которыми следовал дефолт в течение года**,** $\boldsymbol{D = \sum_n D_n}$**.**

Определение. Медианой дефолтов называется средний рейтинговый разряд $\hat{R}$, такой, что количество $\boldsymbol{RD_d^n} < \hat{R}$ равнялось количеству $\boldsymbol{RD_d^n} > \hat{R}$. В случае не возможности обеспечить точное равенство, находится разряд $\hat{R}$, в котором достигается минимум модуля разницы количеств дефолтов. В случае не единственности решения (например, 3 разряда $\hat{R}$), выбирается средний или максимально близкий к среднему.

Формируется три множества разрядов:

$R^- = \{\boldsymbol{RN_i^n, RD_d^n: RN_i^n} < \hat{R}, \boldsymbol{RD_d^n} < \hat{R}\}$,

$R^+ = \{\boldsymbol{RN_i^n, RD_d^n: RN_i^n} > \hat{R}, \boldsymbol{RD_d^n} > \hat{R}\}$ и совокупный

$R = \{R^- \cup R^+ \cup \{\boldsymbol{RN_i^n, RD_d^n: RN_i^n} = \hat{R}, \boldsymbol{RD_d^n} = \hat{R}\}\}$.

Очевидно, что среднее

$PD^+ = E_{r \in R^+}[PD(r)] > PD = E_{r \in R}[PD(r)] > PD^- = E_{r \in R^-}[PD(r)]$,

где $PD(r)$ рассчитанное (модельное) значение вероятности дефолта для рейтинга r (либо присвоенное в соответствии с принятой в Банке мастер-шкалой).

Определяется количество число измерений $N^-, N^+, N = \sum_n (\boldsymbol{D_n + N_n})$ и частота дефолтов P^-, P^+, P в каждом из трех множеств рейтинговых разрядов R^-, R^+, R, оценивается стандартное отклонение статистической погрешности

$\delta P^- = \sqrt{\frac{P^- \cdot (1-P^-)}{N^-}}$, $\delta P^+ = \sqrt{\frac{P^+ \cdot (1-P^+)}{N^+}}$, $\delta \mathrm{P} = \sqrt{\frac{\mathrm{P} \cdot (1\text{-}\mathrm{P})}{\mathrm{N}}}$ соответственно.

Сопоставительный тест не пройден на уровне доверия α, если нарушено хотя бы одно условие:

1) $\boldsymbol{PD^+ \in [P^+ - t_\alpha \cdot \delta P^+, P^+ + t_\alpha \cdot \delta P^+]}$

2) $\boldsymbol{PD^- \in [P^- - t_\alpha \cdot \delta P^-, P^- + t_\alpha \cdot \delta P^-]}$

Где $\boldsymbol{t_\alpha = N^{-1}\left(\frac{1+\alpha}{2}\right)}$, N^{-1}– обратное нормальное распределение.

Причину отрицательного теста можно выяснить, если обратиться к тесту на общее среднее:

T1. $\boldsymbol{PD \in [P - t_\alpha \cdot \delta P, P + t_\alpha \cdot \delta P]}$,

а также к тесту на отношение:

$$\textbf{T2.}\ \frac{PD^+}{PD^-} \in \left[\frac{P^+}{P^-} \cdot \frac{1-t_\alpha \cdot \sqrt{\frac{4}{D} - t_\alpha{}^2 \cdot \frac{4}{D^2}}}{1 - t_\alpha{}^2 \cdot \frac{2}{D}},\quad \frac{P^+}{P^-} \cdot \frac{1+t_\alpha \cdot \sqrt{\frac{4}{D} - t_\alpha{}^2 \cdot \frac{4}{D^2}}}{1 - t_\alpha{}^2 \cdot \frac{2}{D}}\right]. \qquad (1)$$

В случае не прохождения T1 выше верхней границы доверительного интервала – рейтинговая система исторически переоценивает риск с уровнем значимости $\boldsymbol{\alpha}$, (желтая зона), T1 ниже нижней границы – недооценивает соответственно (красная зона). В случае не прохождения T2 выше верхней границы – рейтинговая система исторически переоценивает свою дискриминационную способность, в случае T1 ниже нижней – недооценивает соответственно. В случае не прохождения обоих тестов – следуют оба вывода, возможны комбинации итогового вывода с учетом первых двух биномиальных тестов. Уровень доверия $\boldsymbol{\alpha}$ можно выбрать разным для T1 и T2, но рекомендуется общий $\boldsymbol{\alpha = 90\%}$ ($\boldsymbol{t_\alpha = 1.64}$).

Обоснование бенчмарка медианы дефолтов

Основные два теста, а также тест T1 – суть стандартные биномиальные тесты, активно применяющиеся на практике сопоставительного анализа. Тест T2 не является очевидным и общепринятым, необходимо его обоснование.

Пусть отношение двух случайных величин, измеренных с погрешностью, равно случайной величине

$$w = \frac{\tilde{\xi}}{\tilde{\mu}} = \frac{\xi + \delta\tilde{\xi}}{\mu + \delta\tilde{\mu}},$$

где $\xi, \delta\tilde{\xi}, \mu, \delta\tilde{\mu}$ – это средние не случайные величины и их центрированные случайные погрешности соответственно, дисперсии $D\left(\delta\tilde{\xi}\right) = \delta\xi^2,\ D(\delta\tilde{\mu}) = \delta\mu^2$.

Первые приближения разложений в ряд Тейлора для математических ожиданий и дисперсий равны соответственно

$$E\left(\frac{\tilde{\xi}}{\tilde{\mu}}\right) \cong \frac{\xi}{\mu} \cdot \left(1 + \frac{\delta\mu^2}{\mu^2}\right),$$

$$\delta w = \sqrt{D\left(\frac{\tilde{\xi}}{\tilde{\mu}}\right)} \cong \frac{\xi}{\mu} \cdot \sqrt{\frac{\delta\xi^2}{\xi^2} + \frac{\delta\mu^2}{\mu^2}}, \qquad (2)$$

где δw – стандартное отклонение.

Известная трансформация Гири-Хинкли [10,11] при определённых ограничениях [12] на $\frac{\delta\xi}{\xi} > 0.005, \frac{\delta\mu}{\mu} < 0.39,$

при отсутствии корреляции, на девяностопятипроцентном доверительном уровне дает стандартную нормальную статистику для преобразованной величины $z(w)$

$$z(w) = \frac{w \cdot \mu - \xi}{\sqrt{\delta\xi^2 + w^2 \cdot \delta\mu^2}} \in N(0,1), \qquad (3)$$

которая позволяет построить доверительный интервал для w.

Отношение $w = \frac{P^+}{P^-}$ имеет погрешность, стандартное отклонение которой дается формулой (2), поэтому очевиден вопрос, какое разбиение множеств R^+ и R^- применить так, чтобы свести погрешность измерения w к минимуму? Согласно (2) относительная погрешность задается метрикой

$$L = \left(\frac{\delta P^+}{P^+}\right)^2 + \left(\frac{\delta P^-}{P^-}\right)^2,$$

которая и будет оптимизироваться.

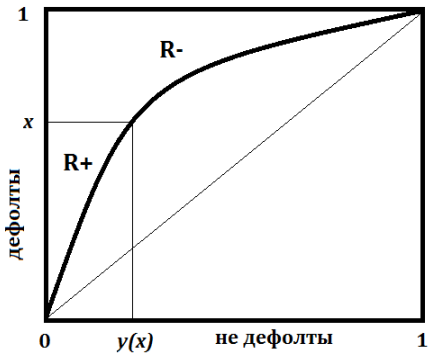
Пусть рейтинговая модель, на основании которой рейтинговые оценки R^+ и R^- распределяются по рейтинговым разрядам, имеет ROC-кривую, изображенную на рис. 1.

Рисунок 1

Разбиение статистики измерений на два множества

Figure 1

Splitting measurement statistics into two sets



Источник: подготовлено автором.
Source: prepared by the author.

Набор измерений рейтинговых оценок разбивается на два множества и R^- координатами ROC-кривой (x – по оси ординат, y(x) – по оси абсцисс)

Одна из самых простых моделей ROC-кривой [13] , которая не имеет правого или левого предпочтения дискриминации, задается формулой

$$x(y) = (1+\beta)\frac{y}{y+\beta}. \qquad (4)$$

Метрика Джини для этой модели ROC-кривой определяется соотношением

$$AR = 2(1+\beta)\left(1-\beta\cdot\ln\left(1+\frac{1}{\beta}\right)\right) - 1. \qquad (5)$$

Тогда,

$$y(x) = \frac{\beta\cdot x}{1+\beta-x}\ , \quad 1-y(x) = \frac{(1+\beta)\cdot(1-x)}{1+\beta-x}.$$

Для P^+, P^- имеем выражения

$$P^+ = P\cdot\frac{x}{y(x)+P\cdot x} = P\cdot\frac{1+\beta-x}{\beta+P\cdot(1+\beta-x)}\ , \qquad (6)$$

$$P^- = P\cdot\frac{1-x}{1-y(x)+P\cdot(1-x)} = P\cdot\frac{1+\beta-x}{1+\beta+P\cdot(1+\beta-x)}.$$

Квадраты стандартных отклонений статистических ошибок P^+, P^- оцениваются по формулам:

$$(\delta P^+)^2 = \frac{P^+\cdot(1-P^+)}{N\cdot(y+P\cdot x)} = \frac{P\cdot\beta}{N}\cdot\frac{(1+\beta-x)^2}{x\cdot(\beta+P\cdot(1+\beta-x))^3}, \quad (7)$$

$$(\delta P^-)^2 = \frac{P^-\cdot(1-P^-)}{N\cdot(1-y+P\cdot(1-x))} = \frac{P\cdot(1+\beta)}{N}\cdot\frac{(1+\beta-x)^2}{(1-x)\cdot(1+\beta+P\cdot(1+\beta-x))^3}.$$

Учитывая $D = P\cdot N$, (6), (7)

$$\left(\frac{\delta P^+}{P^+}\right)^2 = \left(D\cdot x\cdot\left(1+P\cdot\frac{1+\beta-x}{\beta}\right)\right)^{-1}, \left(\frac{\delta P^-}{P^-}\right)^2 = \left(D\cdot(1-x)\cdot\left(1+P\cdot\frac{1+\beta-x}{1+\beta}\right)\right)^{-1}.$$

Метрика $L(x) = \left(\frac{\delta P^+}{P^+}\right)^2 + \left(\frac{\delta P^-}{P^-}\right)^2 = \frac{1}{D}\left(\frac{1}{x}+\frac{1}{1-x}\right) - P\cdot K(x,P,\beta)$. При $P \ll \beta$, метрика $L(x)$ достигает минимума в точке, близкой к $x = 0.5$, и $L(0.5) = \frac{4}{D} - O(P) < \frac{4}{D}$. При не самом высоком и не самом низком показателе Джини AR=50%, из (5) получается параметр $\beta = 0.24$.

При $P = 0$ имеем

$$L(0.5) = \frac{4}{D}, \left(\frac{\delta P^+}{P^+}\right)^2 = \left(\frac{\delta P^-}{P^-}\right)^2 = \frac{2}{D}. \quad (8)$$

Эту оценку можно использовать как умеренно консервативную при $0 \neq P \ll \beta$ для оценки доверительного интервала $w = \frac{P^+}{P^-}$, основываясь на преобразовании Гири-Хинкли (3). Последнее дает интервал на уровне доверия α

$$w_{min,max} = \frac{P^+}{P^-}\cdot\frac{1\mp t_\alpha\cdot\sqrt{\left(\frac{\delta P^+}{P^+}\right)^2+\left(\frac{\delta P^-}{P^-}\right)^2 - t_\alpha{}^2\cdot\left(\frac{\delta P^+}{P^+}\right)^2\cdot\left(\frac{\delta P^-}{P^-}\right)^2}}{1-t_\alpha{}^2\cdot\left(\frac{\delta P^-}{P^-}\right)^2},$$

Который при $x = 0.5$ и $P \ll 1$ консервативно с параметрами (8), оценивается как

$$w_{min} = \frac{P^+}{P^-}\cdot\frac{1-t_\alpha\cdot\sqrt{\frac{4}{D}-t_\alpha{}^2\cdot\frac{4}{D^2}}}{1-t_\alpha{}^2\cdot\frac{2}{D}}, w_{max} = \frac{P^+}{P^-}\cdot\frac{1+t_\alpha\cdot\sqrt{\frac{4}{D}-t_\alpha{}^2\cdot\frac{4}{D^2}}}{1-t_\alpha{}^2\cdot\frac{2}{D}}. \quad (9)$$

При этом оценки среднего и стандартного отклонения $w = \frac{P^+}{P^-}$ (2) будут

$E(w) = \frac{P^+}{P^-}\cdot\left(1+\frac{2}{D}\right)$ и $\delta w = \frac{P^+}{P^-}\cdot\sqrt{\frac{4}{D}}$ соответственно.

Численные примеры калибровок рейтинговых моделей с различными результатами сопоставительных тестов

Настраивается компьютерная симуляция рейтинговой модели на принципе разделения выборок дефолтов и не дефолтов на оси рейтинга. Не умаляя общности рейтинг фиксируется рейтинговым баллом и не ограничивается определенным количеством рейтинговых разрядов. Рейтинг $RN_i \in N(0,1), i \in 1\ldots N$ не дефолтных компаний генерируется стандартным нормальным распределением, а рейтинг дефолтных компаний генерируется настраиваемым нормальным распределением $RD_d \in N(-m,\sigma), d \in 1\ldots D$с параметрами m,σ. Такой метод фитирования ROC-кривой относится к нормальным смесям и, как показано авторами [14], оказывается близким к наблюдаемому на практике. Параметры m,σ позволяют настроить дискриминационную мощность модели (AR), а также ее предпочтение («левое» или «правое»). Оказалось продуктивно параметризовать mфункцией $m(Z,\sigma) = Z\cdot\sqrt{\frac{1}{N}+\frac{\sigma^2}{D}}$.

После генерации выборок рейтингов RN_i, RD_d и упорядочивания их по возрастанию, возможно построить ROC-кривую и определить показатель Джини по формулам [15] :

ROC-кривая в точке $x_i = 0 \dots \frac{i}{N} \dots 1$

$$ROC_0 = 0, ROC_i = \frac{1}{D} \cdot \sum_{d=1}^{D} \delta_{RN_i}(RD_d) ,$$

$$AUC = \frac{1}{N \cdot D} \cdot \sum_{i=1}^{N} \sum_{d=1}^{D} \delta_{RN_i}(RD_d), \qquad (10)$$

$$AR = 2 \cdot AUC - 1,$$

где функция $\delta_u(w) = \begin{cases} 1, \text{если } u > w \\ \frac{1}{2} \text{ если } u = w. \\ 0, \text{если } u < w \end{cases}$

Асимптотическая, упрощенная формула оценки стандартного отклонения AR при условии $\frac{D}{N-D} \ll 1, D \gg 1$ вычисляется [16] как

$$\sigma_{AR} = \sqrt{\frac{(1-AR)^2 \times (1+AR)}{D \times (3-AR)}}. \qquad (11)$$

Точная формула калибровки симметричной ROC-кривой (4) задается [13] соотношением

$$PD(x, \beta, p) = \frac{1}{2}\left(1 - \frac{x+\beta-D-2\beta p}{\sqrt{(x-\beta-p)^2+4\beta(1-p)x}}\right), \qquad (12)$$

где x координата объекта (квантиль) в выборке всех объектов $R = \{RN_i \cup RD_d\}, i = 1 \dots N, d = 1 \dots D,\ dim(R) = N + D.$

Полагая, число дефолтов четным, медианный рейтинг дефолтов оценивается как $\hat{R} = RD_{\frac{D}{2}}$.

Выборки R^- и R^+ по определению:

$R^- = \{RN_i \cup RD_d : RN_i \le \hat{R}, RD_d \le \hat{R}, \},$

$R^+ = \{RN_i \cup RD_d : RN_i > \hat{R}, RD_d > \hat{R}, \},$

$i = 1 \dots N, d = 1 \dots D.$

$\boldsymbol{PD^-}$, $\boldsymbol{PD^+}$, $\boldsymbol{PD}$ оцениваются как средние значения калиброванной вероятности дефолта (12):

$$\boldsymbol{PD^-} = \frac{1}{N^-}\left(\sum_{RN_i \in R^-} PD(N(RN_i, r, d), \beta, p) + \sum_{,RD_d \in R^-} PD(N(RD_d, r, d), \beta, p)\right),$$

$$\boldsymbol{PD^+} = \frac{1}{N^+}\left(\sum_{RN_i \in R^+} PD(N(RN_i, r, d), \beta, p) + \sum_{,RD_d \in R^+} PD(N(RD_d, r, d), \beta, p)\right),$$

$$\boldsymbol{PD} = \frac{1}{N+D}\left(\sum_{RN_i \in R} PD(N(RN_i, r, d), \beta, p) + \sum_{,RD_d \in R} PD(N(RD_d, r, d), \beta, p)\right),$$

где $N^- = \dim(R^-), N^+ = \dim(R^+)$, $N(Y, r, d)$ – кумулятивное нормальное распределение со средним $r = E(R)$ и стандартным отклонением $d = st.dev(R)$, параметр β вычисляется по показателю Джини из уравнения (5), p – калибровочная вероятность дефолта.

Наблюдаемые частоты дефолтов $\boldsymbol{P^-} = \frac{D}{2 \cdot N^-}$, $\boldsymbol{P^+} = \frac{D}{2 \cdot N^+}$, $\boldsymbol{P} = \frac{D}{N+D}$, имеющие стандартные отклонения $\delta P^- = \sqrt{\frac{P^- \cdot (1-P^-)}{N^-}}$, $\delta P^+ = \sqrt{\frac{P^+ \cdot (1-P^+)}{N^+}}$, $\delta P = \sqrt{\frac{P \cdot (1-P)}{N}}$, будут сравниваться с калибровочными $\boldsymbol{PD^-}$, $\boldsymbol{PD^+}$, $\boldsymbol{PD}$ с уровнем доверия $\alpha = 90\%$ для четырех разных сценариев для проверки сопоставительной гипотезы о несостоятельности калибровки рейтинговой модели.

Сценарий № 1: генерация измерений – случайная ROC-кривая, близкая к симметричной, с сопоставительным количеством дефолтов D=200, не дефолтов N=6000, $Z = 13, \sigma = 1.0$ имеющая показатель Джини AR=53.0%; калибровка (12): вероятность дефолта (параметр p, (12)) принимается несмещенной и равна опорной частоте дефолтов $p = P = \frac{D}{N+D}$, параметр β определяется по показателю Джини AR=53.0% из уравнения (5).

Сценарий № 2: генерация измерений – сценарий №1; калибровка (12): вероятность дефолта $p = P = \frac{D}{N+D}$, параметр β определяется по показателю Джини $\mathrm{AR} = 53.0\% + 4\sigma_{AR}$ из уравнения (5), $\sigma_{AR} = 2.6\%$ (11).

Сценарий № 3: генерация измерений – сценарий №1; калибровка (12): вероятность дефолта $p = P = \frac{D}{N+D}$, параметр β определяется по показателю Джини $\mathrm{AR} = 53.0\% - 4\sigma_{AR}$ из уравнения (5), $\sigma_{AR} = 2.6\%$.

Сценарий № 4: генерация измерений – случайная ROC-кривая, имеющая выраженное «левое» предпочтение, с сопоставительным количеством дефолтов D=200, не дефолтов N=6000, $Z = 13, \sigma = 1.5$ имеющая показатель Джини AR=62.5%; калибровка (12): вероятность дефолта также остается несмещенной $p = \frac{D}{N+D}$, параметр β определяется по показателю Джини AR=62.5% из уравнения (5).

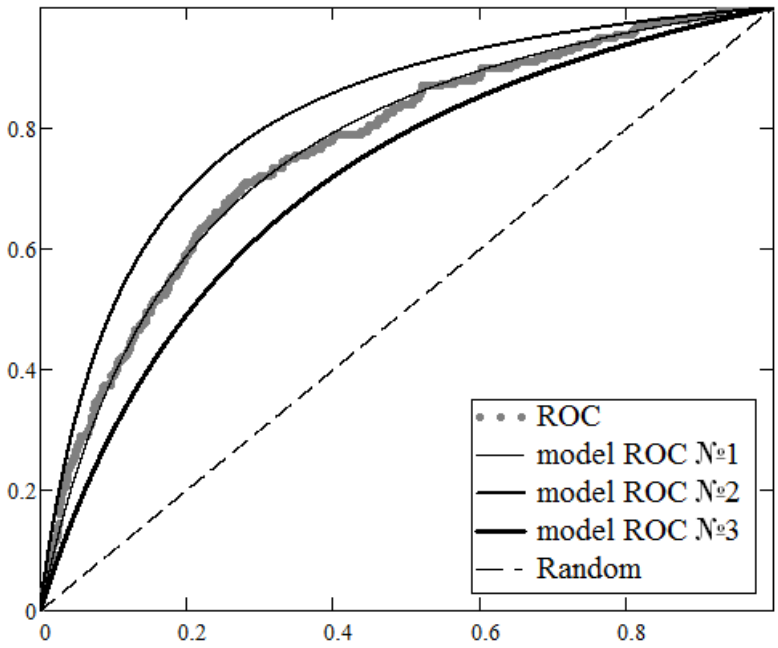
На рис. 2. представлены ROC-кривые калибровки и генерации в рамках сценариев № 1-3.

Рисунок 2

Генерированная ROC-кривая измерений и три сценария калибровки № 1, 2, 3

Figure 2

Generated ROC measurement curve and three calibration scenarios no. 1, 2, 3



Источник: подготовлено автором.
Source: prepared by the author.

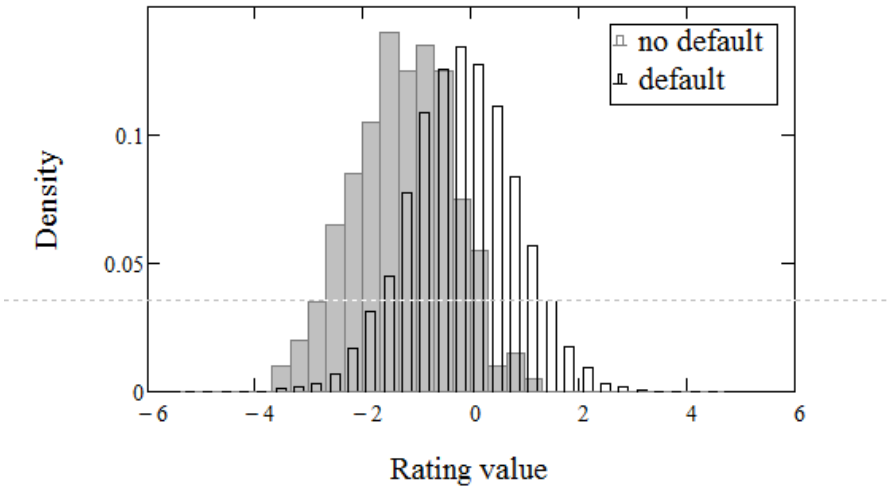
Плотности распределения «плохих» и «хороших» объектов представлены на рис. 3.

Рисунок 3

Плотности распределения дефолтных и не дефолтных объектов по рейтинговому баллу в сценариях № 1-3

Figure 3

Density of distribution of default and non-default objects by rating score in scenarios no. 1-3



Источник: подготовлено автором.
Source: prepared by the author.

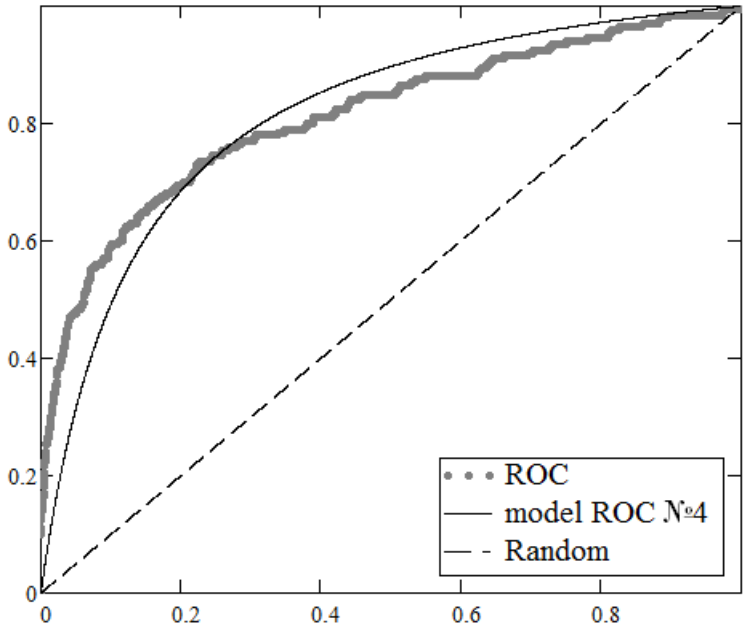
На рис. 4 представлены ROC-кривые калибровки и генерации в рамках сценария № 4.

Рисунок 4

Генерированная ROC-кривая измерений и сценарий калибровки № 4

Figure 4

Generated ROC Measurement Curve and Calibration Scenario no. 4



Источник: подготовлено автором.
Source: prepared by the author.

В табл. 1–4 представлены результаты сопоставительного теста для рассчитанных значений вероятности дефолта, полученных в результате применения модели (12) для сценариев № 1–4.

Таблица 1

Сопоставительный тест для сценария № 1

Table 1

Benchmark test for scenario no. 1

Тест	**Параметры**	**Значение**	**Результат теста на уровне доверия $\alpha = 90\%$. Сопоставительная гипотеза:**
1	$\boldsymbol{PD^-}$	1.88%	Не отвергается
	$\boldsymbol{P^-}$	[1.6%, 2.2%]	
2	$\boldsymbol{PD^+}$	10.08%	Не отвергается
	$\boldsymbol{P^+}$	[8.4%, 11.48%]	

T1	$\boldsymbol{PD}$	3.22%	Не отвергается
	$\boldsymbol{P}$	[2.9%, 3.6%]	
T2	$\frac{\boldsymbol{PD}^{+}}{\overline{\boldsymbol{PD}^{-}}}$	5.35	Не отвергается
	$\boldsymbol{w} = \frac{\boldsymbol{P}^{+}}{\overline{\boldsymbol{P}^{-}}}$	[4.08, 6.53]	
Итог…	Гипотеза состоятельности сопоставительного анализа не отвергается на уровне $\alpha = 90\%$		

Источник: подготовлено автором.
Source: prepared by the author.

Таблица 2

Сопоставительный тест для сценария № 2

Table 2

Benchmark test for scenario no. 2

Тест	**Параметры**	**Значение**	**Результат теста на уровне доверия $\alpha = 90\%$. Сопоставительная гипотеза:**
1	$\boldsymbol{PD}^{-}$	1.47%	Отвергается на уровне α
	$\boldsymbol{P}^{-}$	[1.6%, 2.2%]	
2	$\boldsymbol{PD}^{+}$	12.2%	Отвергается на уровне α
	$\boldsymbol{P}^{+}$	[8.4%, 11.5%]	
T1	$\boldsymbol{PD}$	3.22%	Не отвергается
	$\boldsymbol{P}$	[2.9%, 3.6%]	
T2	$\frac{\boldsymbol{PD}^{+}}{\overline{\boldsymbol{PD}^{-}}}$	8.29	Отвергается на уровне α
	$\boldsymbol{w} = \frac{\boldsymbol{P}^{+}}{\overline{\boldsymbol{P}^{-}}}$	[4.08, 6.53]	
Итог…	Тест сопоставительного анализа не пройден		

Источник: подготовлено автором.
Source: prepared by the author.

Таблица 3

Сопоставительный тест для сценария № 3

Table 3

Benchmarking test for scenario no. 3

Тест	**Параметры**	**Значение**	**Результат теста на уровне доверия $\alpha = 90\%$. Сопоставительная гипотеза:**
1	$\boldsymbol{PD}^{-}$	2.25%	Отвергается на уровне α
	$\boldsymbol{P}^{-}$	[1.6%, 2.2%]	
2	$\boldsymbol{PD}^{+}$	8.21%	Отвергается на уровне α
	$\boldsymbol{P}^{+}$	[8.4%, 11.5%]	
T1	$\boldsymbol{PD}$	3.2%	Не отвергается

	$\boldsymbol{P}$	[2.9%, 3.6%]	
T2	$\frac{\boldsymbol{PD}^{+}}{\boldsymbol{PD}^{-}}$	3.66	Отвергается на уровне α
	$\boldsymbol{w} = \frac{\boldsymbol{P}^{+}}{\boldsymbol{P}^{-}}$	[4.08, 6.53]	
Итог…	Тест сопоставительного анализа не пройден		

Источник: подготовлено автором.
Source: prepared by the author.

Таблица 4

Сопоставительный тест для сценария № 4

Table 4

Benchmark for scenario no. 4

Тест	**Параметры**	**Значение**	**Результат теста на уровне доверия $\alpha = 90\%$. Сопоставительная гипотеза:**
1	$\boldsymbol{PD}^{-}$	2.0%	Не отвергается
	$\boldsymbol{P}^{-}$	[1.5%, 2.0%]	
2	$\boldsymbol{PD}^{+}$	16.6%	Отвергается на уровне α
	$\boldsymbol{P}^{+}$	[18.3%, 24.5%]	
T1	$\boldsymbol{PD}$	3.2%	Не отвергается
	$\boldsymbol{P}$	[2.9%, 3.6%]	
T2	$\frac{\boldsymbol{PD}^{+}}{\boldsymbol{PD}^{-}}$	8.0	Отвергается на уровне α
	$\boldsymbol{w} = \frac{\boldsymbol{P}^{+}}{\boldsymbol{P}^{-}}$	[9.68, 15.50]	
Итог…	Тест сопоставительного анализа не пройден		

Источник: подготовлено автором.
Source: prepared by the author.

Анализ данных табл. 1 показывает, что при условии несмещённой калибровочной вероятности дефолтов и не смещенной мощности рейтинговой модели сопоставительный тест показывает на заданном уровне доверия отсутствие оснований для отвержения сопоставительной гипотезы. В табл. 2 калибровочная модель была параметризована с заведомым превышением дискриминирующей способности (на «четыре сигма») относительно генерированной модели, при условии сохранения несмещённости PD. Результат теста оказался отрицательным из-за того, что модель существенно занижает вероятность дефолта компаний (объектов) в области ниже медианы дефолтов (т.е. неправомерно улучшает «хороших») и завышает PD в области выше медианы (т.е. неправомерно ухудшает «плохих»). Табл. 3, в которой тестировалась модель с заведомым понижением дискриминирующей способности (на «четыре сигма») относительно генерированной модели, показывает противоположный негативный результат. Т.е. модель № 3 неправомерно ухудшает «хороших» и улучшает «плохих».

С точки зрения кредитного бизнеса применение модели № 2 приведет к недооценке риска заемщиков с рейтингом выше среднего и переоценке риска заемщиков кредитоспособности ниже среднего. Это, очевидно, может привести к убыткам из-за недооценки стоимости риска (резервов) кредитного портфеля («хороших») и к упущенной выгоде из-за необоснованно жесткой политики по отношению к соискантам финансирования у кого кредитоспособность ниже среднего. Применение модели № 3 приведет к недооценке риска «плохих» заемщиков, а значит к прямым кредитным убыткам из-за дефолтов, а также к упущенной выгоде из-за неоправданно повышенных требований к «хорошим» (кредитоспособным) заемщикам в условиях конкуренции.

Модель № 4, тестирование которой представлено в табл. 4, показывает неадекватность формулы калибровки (12), которая не учитывает явное левое предпочтение (хорошо «видит плохих»). Это также приводит к недооценке риска заемщиков кредитоспособности ниже средней даже при учете релевантности параметризации дискриминирующей силы и несмещенности вероятности дефолта. Адекватные формулы для калибровки модели левого или правого предпочтения представлены в работе [13].

Заключение

В данной работе предложен новый практичный сопоставительный тест, позволяющий на заданном уровне доверия отвергнуть на исторических данных гипотезу о состоятельности оценки вероятности дефолта рейтинговой моделью. Задача сопоставительного анализа рассчитанных значений вероятности дефолта, полученных в результате применения моделей, используемых в рейтинговой системе, с фактической частотой реализованных дефолтов заемщиков должна периодически решаться в рамках проведения внутренней валидации ПВР.

Предложенный метод, кроме смещенности оценки общей частоты дефолтов в кредитном портфеле, позволяет объективно выявить неадекватность дискриминации заемщиков калиброванной рейтинговой моделью, диагностировать «болезнь» рейтинговой модели. Причем для этого не требуется полнота статистики в каждом рейтинговом разряде, которой на практике почти никогда нет, что дает расширение области применимости сопоставительного анализа на исторических данных с небольшим количеством дефолтов, произошедших за валидационный период.

С практической точки зрения в любой рейтинговой модели, применяемой к крупным компаниям-объектам риска, присутствуют экспертные коррекции рейтинга, которые также экспертно регламентируются. Такие коррекции будут приводить к искажению дискриминирующей мощности как в сторону повышения, так, возможно, и понижения. Предложенный тест поможет вовремя обнаружить, в том-числе, несостоятельность таких коррекций либо вообще несостоятельность общей калибровки.

Список литературы

1. *Tasche, D.* Validation of internal rating systems and PD estimates // The Analytics of Risk Model Validation, 2008, pp. 169–196.

2. *González, Fernando & Coppens, François & Winkler, Gerhard.* The performance of credit rating systems in the assessment of collateral used in Eurosystem monetary policy operations // Occasional Paper Series 65, 2007, European Central Bank.
3. *Miu, P. and Ozdemir, B.* Estimating and Validating Long-Run Probability of Default with Respect to Basel II Requirements // Journal of Risk Model Validation, 2008, no. 2, pp. 1–39.
4. *Sauer, Stephan & Coppens, François & Mayer, Manuel & Millischer, Laurent & Resch, Florian & Schulze, Klaas.* Advances in multivariate back-testing for credit risk underestimation // Working Paper Series 1885, 2016, European Central Bank.
5. *P.H. Westfall and R.D. Wolnger.* Multiple tests with discrete distributions // The American Statistician, 51:3{8, 1997.
6. *Hosmer, David W.; Lemeshow, Stanley*. Applied Logistic Regression // New York: Wiley. 2013. ISBN 978-0-470-58247-3.
7. *Sokal, R. R.; Rohlf, F. J.* Biometry: The Principles and Practice of Statistics in Biological Research (Second ed.) // New York: Freeman. 1981. ISBN 978-0-7167-2411-7
8. *McDonald, John H.* Small numbers in chi-square and G–tests". Handbook of Biological Statistics (3rd ed.) // Baltimore, MD: Sparky House Publishing, 2014, pp. 86–89.
9. *Spiegelhalter, D.* Probabilistic prediction in patient management and clinical trails // Statistics in Medicine, 1986, vol. 5, pp. 421-433.
10. *Geary, R. C.* The Frequency Distribution of the Quotient of Two Normal Variables // Journal of the Royal Statistical Society, 1930, vol. 93, pp. 442–446.
11. *Hinkley, D.V.* On the Ratio of Two Correlated Normal Random Variables // Biometrika, 1969, vol. 56, pp. 635–639.
12. *Hayya, Jack; Armstrong, Donald; Gressis, Nicolas.* A Note on the Ratio of Two Normally Distributed Variables // Management Science, 1975, no. 21 (11), pp. 1338–1341.
13. *Помазанов М.В.* ROC-анализ и калибровка скоринговых моделей на основе метрик точности второго порядка // Управление финансовыми рисками. 2021. № 2. С. 100–121. DOI: 10.36627/2221-7541-2021-2-2-100-121
14. *Hong, Chong-Sun, Lee, Won-Yong.* ROC Curve Fitting with Normal Mixtures // The Korean Journal of Applied Statistics, 2011, vol. 24, issue 2, pp. 269–278 https://doi.org/10.5351/KJAS.2011.24.2.269
15. *Engelmann, B., Hayden, E., and Tasche, D.* Measuring the discriminative power of rating systems // Risk, 2003, pp. 82–86.
16. *Hanley, A., and B. McNeil.* The Meaning and Use of the Area Under a Receiver Operating Characteristics (ROC) Curve // Diagnostic Radiology, 1982, no. 143, pp. 29–36.

Информация о конфликте интересов

Я, автор данной статьи, со всей ответственностью заявляю о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи. Настоящее заявление относится к проведению научной работы, сбору и обработке данных, написанию и подготовке статьи, принятию решения о публикации рукописи.

References

1. Tasche, D. Validation of internal rating systems and PD estimates. *The Analytics of Risk Model Validation,* 2008, pp. 169–196.
2. González, Fernando & Coppens, François & Winkler, Gerhard. The performance of credit rating systems in the assessment of collateral used in Eurosystem monetary policy operations. *Occasional Paper Series 65,* 2007, European Central Bank.
3. Miu, P. and Ozdemir, B. Estimating and Validating Long-Run Probability of Default with Respect to Basel II Requirements. *Journal of Risk Model Validation,* 2008, no. 2, pp. 1–39.
4. Sauer, Stephan & Coppens, François & Mayer, Manuel & Millischer, Laurent & Resch, Florian & Schulze, Klaas. Advances in multivariate back-testing for credit risk underestimation, *Working Paper Series 1885, 2016,* European Central Bank.
5. P.H. Westfall and R.D. Wolnger. Multiple tests with discrete distributions. *The American Statistician,* 51:3{8, 1997.
6. Hosmer, David W.; Lemeshow, Stanley. Applied Logistic Regression. *New York: Wiley*. 2013. ISBN 978-0-470-58247-3.
7. Sokal, R. R.; Rohlf, F. J.. Biometry: The Principles and Practice of Statistics in Biological Research (Second ed.). *New York: Freeman.* 1981. ISBN 978-0-7167-2411-7
8. McDonald, John H. Small numbers in chi-square and G–tests". Handbook of Biological Statistics (3rd ed.). *Baltimore, MD: Sparky House Publishing,* 2014, pp. 86–89.
9. Spiegelhalter, D. Probabilistic prediction in patient management and clinical trails. *Statistics in Medicine,* 1986,Vol. 5, pp. 421-433.
10. Geary, R. C., The Frequency Distribution of the Quotient of Two Normal Variables. *Journal of the Royal Statistical Society,* 1930, vol. 93, pp. 442–446.
11. Hinkley, D.V. On the Ratio of Two Correlated Normal Random Variables, *Biometrika,* 1969, vol. 56, pp. 635–639.
12. Hayya, Jack; Armstrong, Donald; Gressis, Nicolas. A Note on the Ratio of Two Normally Distributed Variables. *Management Science,* 1975, no. 21 (11), pp. 1338–1341.
13. Pomazanov M.V. [ROC Analysis and Calibration of Scoring Models Based on Second Order Accuracy Metrics]. *Upravlenie finansovymi riskami = Financial Risk Management,* 2021, no. 2, pp. 100–121. (In Russ.) DOI: 10.36627/2221-7541-2021-2-2-100-121
14. Hong, Chong-Sun, Lee, Won-Yong. ROC Curve Fitting with Normal Mixtures. *The Korean Journal of Applied Statistics,* 2011, vol. 24, issue 2, pp. 269–278 https://doi.org/10.5351/KJAS.2011.24.2.269
15. Engelmann, B., Hayden, E., and Tasche, D. Measuring the discriminative power of rating systems. *Risk,* 2003, pp. 82–86.
16. Hanley, A., and B. McNeil. The Meaning and Use of the Area Under a Receiver Operating Characteristics (ROC) Curve. *Diagnostic Radiology,* 1982, no. 143, pp. 29–36.

Conflict-of-interest notification

I, the author of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.