

**БАНКОВСКИЕ РИСКИ:  
ТЕОРИЯ, ПРАКТИКА,  
МЕТОДОЛОГИЯ**

## **ROC-АНАЛИЗ И КАЛИБРОВКА СКОРИНГОВЫХ МОДЕЛЕЙ НА ОСНОВЕ МЕТРИК ТОЧНОСТИ ВТОРОГО ПОРЯДКА**

В статье представлены новые метрики точности скоринговых моделей второго порядка, которые показывают целевое предпочтение скоринга: диагностировать «хорошие» объекты (заемщиков) или выделять «плохие» при неизменной прогнозной силе, определяемой общепринятой метрикой первого порядка — индексом Джини.

**КЛЮЧЕВЫЕ СЛОВА:** скоринговая модель, рейтинг, калибровка, вероятность дефолта, ROC-кривая, индекс Джини, прогнозная сила, валидация

### **ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ**

Самый популярный метод валидации скоринговых (рейтинговых) моделей, используемый в настоящее время, — это анализ характеристик кривой профиля совокупной точности (Cumulative Accuracy Profile, CAP) и его сводная статистика, коэффициент точности. Подробное объяснение этого метода можно найти в работе Дж. Собхарта и др. [10] Для построения CAP-кривой все полученные значения кредитного скоринга (рейтинга) упорядочиваются по возрастанию по всей совокупности объектов скоринга (юридических, физических лиц) за определенный период или за совокупность периодов. Должно быть известно, наступил ли дефолт по объекту в течение анализируемого периода (как правило, года) с даты расчета рейтинга. Эффективная рейтинговая модель должна иметь предсказательную силу, позволяющую отделять дефолтеров от «хороших» объектов, помещая первых в начало списка.

Для каждого возможного числового значения рейтинга от 0 до 100 (например, 10) по оси Y откладывается доля дефолтов с меньшим рейтингом



**Помазанов Михаил Вячеславович** — к. ф.-м. н., руководитель подразделения валидации блока «Риски» ПАО «Промсвязьбанк». Автор более 25 научных работ, в том числе двух монографий (г. Москва)

(т.е. доля дефолтов от объектов с рейтингом ниже 10), по оси  $OX$  — доля объектов с меньшим рейтингом (т.е. доля объектов с рейтингом ниже 10 от всех объектов). Абсолютное значение рейтинга (рейтинговый разряд) на графике  $САР$  может не отражаться. У эффективной рейтинговой модели  $САР$ -кривая должна быть максимально выпукла над диагональю и максимально возрастать в начале оси  $OX$ . Если  $САР$ -кривая близка к диагонали, это означает, что рейтинг никак не характеризует вероятность дефолта (неэффективная скоринговая модель, Random).

Показателем Джини (Gini index), или коэффициентом точности скоринговой модели (Accuracy Ratio, AR), называется отношение площади между  $САР$ -кривой и диагональю к площади между идеальной  $САР$ -кривой и диагональю (рис. 1):

$$AR = \frac{a_R}{a_P}.$$

Величина  $AR$  измеряется в диапазоне от 0 (неэффективная система) до 1 (идеальная система). Если  $AR$  меньше нуля (может быть до  $-1$ ), то рейтинговая модель хуже, чем ее отсутствие, либо рейтинговые объекты упорядочены в обратном порядке.

Идеальная  $САР$ -кривая будет задана в виде треугольника с вырезом с основанием  $D$  (см. рис. 1).  $D$  — доля дефолтов среди всех рейтингованных объектов обучающей выборки.

Математически:

$$AR = \frac{2 \int_0^1 C(x) dx - 1}{1 - D}, \quad (1)$$

где  $C(x)$  — значение  $САР$ -кривой для координаты  $x$  объекта скоринга, которая обозначает долю объектов, упорядоченных по рейтингу таким образом, что рейтинг объекта  $x$  выше или равен доли  $x$  всех объектов рейтинга, но ниже, чем доли  $1 - x$  соответственно.

В силу определения  $C(0) = 0, C(1) = 1$ .

Из построения  $САР$ -кривой следует, что вероятность дефолта в обучающей выборке объектов  $x \in [0, 1]$  задана следующим соотношением:

$$PD(x) = D \times \frac{d}{dx} C(x), \quad (2)$$

где  $\frac{d}{dx}$  — первая производная.

Для построения кривой рабочей характеристики приемника (Receiver Operating Characteristic, ROC) по оси  $OX$  откладывается доля недефолтных объектов, упорядоченных по возрастанию рейтинга (скоринга), поэтому на рисунке ROC-кривой отсутствует треугольник с основанием  $D$ . Идеальная ROC-кривая будет совпадать с осью  $OY$  до значения 1, поэтому из формулы для показателя Джини формально исчезнет зависимость от доли дефолтов  $D$ .

$$AR = 2 \times AUC - 1, \quad (3)$$

где  $AUC = \int_0^1 R(g) dg$  — площадь под ROC-кривой (Area under ROC Curve);

$g$  — доля «хороших» (без дефолта) объектов.

Метрика  $AUC$  широко и давно применяется в различных задачах диагностики бинарных событий [4]: от проблем отделения сигнала от шума в радиотехнике [6] и медицинской диагностики [11, 14] до алгоритмов искусственного интеллекта [3] и кредитного скоринга, используемого в риск-менеджменте [8].

Показатель Джини, вычисленный по ROC-кривой, оказывается инвариантен по отношению к  $D$ . Действительно, доля «хороших» объектов будет определяться следующим образом:

$$g(x) = \frac{x - \int_0^x PD(s) ds}{1 - D}. \quad (4)$$

Очевидно, что  $g(x) \in [0, 1]$ .

При этом:

$$\begin{aligned} C(x) &= R(g(x)), \\ R(0) &= 0, R(1) = 1. \end{aligned} \quad (5)$$

Тогда:

$$\begin{aligned} AR &= 2 \int_0^1 R(g(x)) dg(x) - 1 = \\ &= 2 \int_0^1 C(x) (1 - PD(x)) dx - 1 = \end{aligned}$$

Рис. 1. CAP-кривая

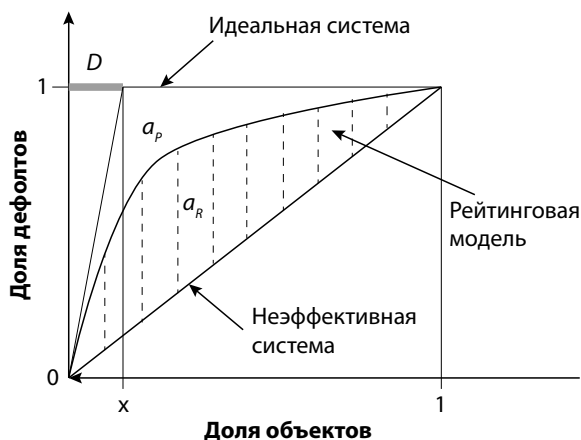
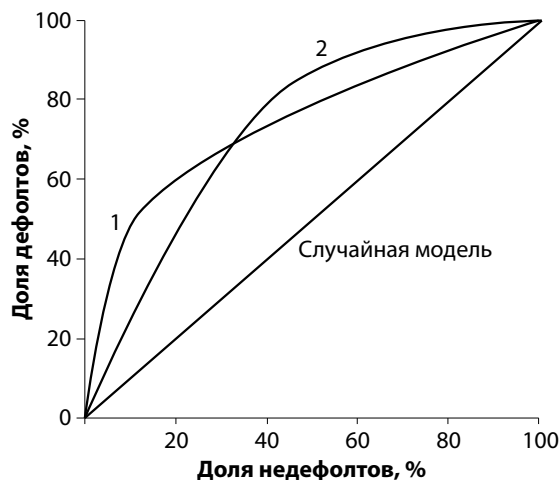


Рис. 2. ROC-кривые двух моделей одинаковой мощности (Джини)



$$\begin{aligned}
 &= \frac{2 \int_0^1 C(x) dx - 2D \int_0^1 C(x) \times C'(x) dx}{1-D} - 1 = \\
 &= \frac{2 \int_0^1 C(x) dx - D - (1-D)}{1-D} = \frac{2 \int_0^1 C(x) dx - 1}{1-D},
 \end{aligned}$$

т.е. формулы (1) и (3) эквивалентны.

Метрика Джини является наиболее известной и общепринятой в кредитном риск-менеджменте, она досконально изучена, в работе Б. Энгельмана и др. [5] даны оценки статистической точности этой метрики на ограниченных данных измерений. Однако метрика Джини не характеризует особенности скоринговых моделей, которые заключаются в том, чтобы лучшим образом определять «плохие» объекты и не лучшим образом различать «хорошие», или наоборот.

На рис. 2 изображены ROC-кривые двух скоринговых моделей, имеющих одинаковые показатели точности AR. Однако кривая 1 характеризует рейтинговую модель 1, которая лучше, чем модель 2, отделяет «плохие» объекты. Наоборот, кривая 2 показывает, что модель 2 лучше отделяет «хорошие» объекты.

Математически в этом легко убедиться, если обратить внимание на геометрический смысл формулы (2), которая определяет вероятность дефолта как тангенс угла наклона касательной для CAP-кривой, умноженный на нормирующую дефолтность.

На практике целевая необходимость построения преимущественно левой модели (т.е. модели 1) или преимущественно правой модели (2) возникает достаточно часто. Например, если стратегия направлена на максимизацию кредитования одобренных компаний на общих условиях (залог, маржа, короткий срок и т.п.), то разумно, что предпочтение такой «кредитной фабрики» будет отдано модели 1. В случае долгосрочного кредитования на льготных условиях или кредитования в условиях спада экономики разумное предпочтение будет отдано модели 2, поскольку для такой стратегии требуется отобрать лучшие компании с минимальным уровнем дефолтности.

Для разделения этих целей необходимо ввести дополнительные метрики, которые позволят различать особенности скоринговой модели

с левым и правым предпочтением выпуклости ROC-кривой.

### ЛЕВЫЙ И ПРАВЫЙ КОЭФФИЦИЕНТЫ ТОЧНОСТИ СКОРИНГОВОЙ МОДЕЛИ

Предполагается, что построена скоринговая (рейтинговая) модель, для которой известна ROC-кривая  $R(x)$ .

Приведем определения.

Левым коэффициентом точности (Left AR) скоринговой модели называется следующая метрика:

$$LAR = \int_0^1 LJ(c)dc,$$

где  $LJ(c) = 2 \frac{\int_0^c R(x)dx}{c \times R(c)} - 1$ .

Правым коэффициентом точности (Right AR) скоринговой модели называется метрика:

$$RAR = \int_0^1 RJ(c)dc,$$

где  $RJ(c) = 2 \frac{\int_c^1 (R(x) - R(c))dx}{(1-c) \times (1-R(c))} \times R'(c) - 1$ .

Другое тождественное определение для Right AR, как нетрудно показать, можно дать в следующем виде:

$$RAR = \int_0^1 RJ(c)dc,$$

где  $RJ(c) = 2 \frac{\int_c^1 (R(x) - 1)dx}{(1-c) \times (1-R(c))} \times R'(c) + 1$ .

Если проинтегрировать последнее выражение по частям, положив в качестве дифференцируемой части  $u = \int_c^1 (R(x) - 1)dx$ , то можно получить альтернативное представление для RAR в виде:

$$RAR = 1 - 2 \times \int_0^1 (1 - R(c)) \times \left( \int_0^c \frac{R'(x)}{(1-x) \times (1-R(x))} dx \right) dc.$$

Такое представление дает более устойчивый к неполноте данных численный аналог.

Нормировка  $LAR$  и  $RAR$  выбрана таким образом, чтобы при  $AR = 0$  (случайная скоринговая модель)  $LAR = RAR = 0$ . Действительно, при нуле-

вом  $AR$ ,  $R(x) = x$  и  $LAR[random] = 2 \int_0^1 \frac{\int_0^c xdx}{c^2} dc - 1 = 0$ .

Аналогично получится и для  $RAR$ .

Правый интегральный Джини ( $RAR$ ) является представлением левого интегрального Джини ( $LAR$ ), и наоборот, если сделать симметричное обратное преобразование ROC-кривой, а именно:

$$\begin{cases} y = 1 - R(x) \\ Q = 1 - x, \end{cases} \quad (6)$$

тогда  $Q(y)$  будет правой ROC-кривой (типа кривой 2, см. рис. 2), если  $R(x)$  было левой ROC-кривой (типа кривой 1, см. рис. 2), и наоборот.

Действительно,

$$I = \int_0^1 \frac{\int_0^c Q(y)dy}{c \times Q(c)} dc.$$

После замены переменных формулы (6), при которых  $dy = -R'(x)dx$ ,  $y|_0^c \rightarrow x|_1^{x(c)}$ , где  $x(c) = R^{-1}(1 - c)$  (обратная функция  $R$ ),  $I$  примет следующий вид:

$$\begin{aligned} I &= \int_0^1 \frac{-\int_1^{x(c)} (1-x)R'(x)dx}{c \times Q(c)} dc = \\ &= \int_0^1 \frac{R(x)|_{x(c)}^1 - xR(x)|_{x(c)}^1 + \int_{x(c)}^1 R(x)dx}{c \times Q(c)} dc = \\ &= \int_0^1 \frac{\int_{x(c)}^1 R(x)dx - (1-x(c))R(x(c))}{c \times Q(c)} dc = \\ &= \int_0^1 \frac{\int_{x(c)}^1 (R(x) - R(x(c)))dx}{c \times Q(c)} dc. \end{aligned}$$

Проведем следующую замену переменной:  $z = x(c)$ , тогда:

$$x(0) = R^{-1}(1) = 1,$$

$$x(1) = R^{-1}(0) = 0.$$

Далее:

$$R(z) = 1 - c,$$

$$dc = -R'(z)dz,$$

$$c = 1 - R(x(c)) = 1 - R(z),$$

$$Q(c) = 1 - x(c) = 1 - z.$$

Окончательно выражение  $I$  получается в виде:

$$I = \int_0^1 \frac{\int_{x(W)}^1 (R(x) - R(x(c))) dx}{c \times Q(c)} dc =$$

$$= \int_0^1 \frac{\int_z^1 (R(x) - R(z)) dx}{(1 - R(z)) \times (1 - z)} R'(z) dz.$$

Таким образом, после отображения формулы (6) определение  $LAR$  переходит в определение  $RAR$ . В силу тождественности преобразований верно и обратное.

### ЧИСЛЕННЫЕ ФОРМУЛЫ ДЛЯ ЛЕВОГО И ПРАВОГО КОЭФФИЦИЕНТА ТОЧНОСТИ

Практическое применение целевых метрик предполагает их расчет на скоринговой модели, которая валидируется на точечных данных подсчетов количества измерений и количества дефолтов. После упорядочения всех дефолтов и недефолтов по возрастанию рейтинга (показателя скоринга) производится их подсчет для вычисления точечной оценки метрик на обучающей выборке.

Пусть число упорядоченных измерений рейтинга равно  $N$ ,  $i = 1, \dots, N$ , среди них число последующих дефолтов  $N_d$ , число недефолтов  $N_g = N - N_d$ .

Индекс дефолта:

$$d_i = \begin{cases} 1, & \text{если дефолт,} \\ 0, & \text{если недефолт.} \end{cases}$$

$$\text{Координаты } R(k) = \frac{\sum_{i=1}^k d_i}{N_d}, \quad g(k) = \frac{k - \sum_{i=1}^k d_i}{N_g}$$

будут формировать ROC-кривую в осях  $OY$ ,  $OX$  соответственно.

Если полагать  $g(0) = 0$ ,  $R(0) = 0$ , точечным числовым аналогом определений стандартного

показателя Джини  $AR$ ,  $LAR$  и  $RAR$  будут следующие выражения:

$$AR = 2 \sum_{k=1}^N \left[ \frac{R(k) + R(k-1)}{2} \times \right. \\ \left. \times (g(k) - g(k-1)) \right] - 1,$$

$$LAR = 2 \sum_{k=1}^N \left[ \frac{g(k) - g(k-1)}{g(k) \times R(k)} \times \right. \\ \left. \sum_{s=1}^k \frac{R(s) + R(s-1)}{2} \times \right. \\ \left. \times (g(s) - g(s-1)), \right. \\ \left. \begin{array}{l} \text{если } g(k) \times R(k) \neq 0, \\ 0, \text{ если } g(k) \times R(k) = 0 \end{array} \right] - 1, \quad (7)$$

$$RAR = 1 - 2 \sum_{k=1}^N \left[ \frac{(1 - R(k)) \times (g(k) - g(k-1)) \times \right. \\ \left. \sum_{s=1}^k \frac{R(s) - R(s-1)}{1 - \frac{g(s) + g(s-1)}{2}} \times \left( 1 - \frac{R(s) + R(s-1)}{2} \right) \right. \\ \left. \begin{array}{l} \text{если } (1 - g(k)) \times (1 - R(k)) \neq 0, \\ 0, \text{ если } (1 - g(k)) \times (1 - R(k)) = 0 \end{array} \right].$$

Для наиболее точного с практической точки зрения расчета этих показателей следует формировать множество  $\Omega$  случайных подвыборок  $\omega \in \Omega$  обучающей выборки, а затем проводить усреднение точечных метрик для каждой подвыборки  $\omega$ :

$$LAR = \frac{1}{\dim(\Omega)} \sum_{\omega \in \Omega} LAR_{\omega},$$

$$RAR = \frac{1}{\dim(\Omega)} \sum_{\omega \in \Omega} RAR_{\omega},$$

$$AR = \frac{1}{\dim(\Omega)} \sum_{\omega \in \Omega} AR_{\omega}.$$

Однако следует иметь в виду ограничения вычислительной мощности. Количество операций, необходимое для расчета  $AR$ , пропорционально  $N$ , для расчета  $LAR$  и  $RAR$  требуется  $N^2 / 2$  на каждый показатель, т.е. при росте мощности выборки в десять раз время вычисления  $LAR$  возрастет в 50 раз и т.п. По этой причине рекомендуется выборку  $ROC(x_i)$  с точечным (непрерывным) рейтингом равномерно прореживать до  $N = 100-200$  точек.

### КАЛИБРОВОЧНАЯ ЦЕЛЕВО-НЕЙТРАЛЬНАЯ ROC-КРИВАЯ

Для понимания того, насколько метрики  $LAR$  и  $RAR$ , характеризующие заданную для исследования скоринговую модель и имеющие заданный показатель дискриминирующей силы  $AR$ , отличаются от нейтрального значения, необходимо предложить нейтральную ROC-кривую, обладающую следующими свойствами.

1. Нейтральная ROC-кривая должна быть однопараметрической (обозначим этот параметр как  $\beta$ ).

2. Для искомой кривой должно выполняться равенство  $LAR(\beta) = RAR(\beta)$ .

3. Нейтральная  $ROC = R(x)$  должна быть непрерывной рациональной функцией, не выходящей за пределы верхнего левого угла (см. рис. 1) и соблюдающей граничные условия  $R(0) = 0, R(1) = 1$ .

4. Множество значений  $ranAR(\beta) = [0, 1)$ , и существует обратная функция  $\beta = AR^{-1}([0, 1))$ .

Из требования 2 следует, что порождающая ROC-кривую функция должна быть симметрична относительно диагонали  $y = 1 - x$  (45%) на плоскости  $XOY$ . Самой очевидной рациональной кривой, не выходящей за границы прямого угла  $XOY$ , является функция  $Y = \frac{\alpha}{x}$ , где  $\alpha$  — параметр. Эта функция

будет повернута относительно искомой на 90 градусов. Для того чтобы кривая проходила через точки  $(0, 1)$  и  $(1, 0)$ , оставаясь симметричной относительно диагонали  $y = x$ , достаточно параметризовать ее сдвиг вдоль диагонали в виде  $y + \beta = \frac{\alpha}{x + \beta}$ . Учитывая граничные условия  $y(1) = 0$

и  $y(0) = 1$ , легко установить, что  $\alpha = \beta \times (1 + \beta)$ . Чтобы получить искомую кривую  $R(x)$ , симметричную относительно диагонали  $y = 1 - x$ , достаточно сделать преобразование  $R(x) = 1 - y$ . В итоге получим искомую ROC-кривую, обладающую свойствами 1–3, в следующем виде:

$$R(x) = (1 + \beta) \frac{x}{x + \beta}. \tag{8}$$

Расстояние от левого верхнего угла прямоугольника ROC-кривой до самой выпуклой точки  $(x^*, R(x^*))$ , для которой  $R'(x^*) = 1$ , будет определяться следующим образом:

$$\rho = \frac{\sqrt{2}}{1 + \sqrt{\frac{1 + \beta}{\beta}}},$$

при  $\beta \rightarrow 0, \rho \rightarrow 0$ , т.е. найденная ROC-кривая будет стремиться к идеальной кривой (идеальной скоринговой модели).

При  $\beta \rightarrow \infty, \rho \rightarrow \frac{\sqrt{2}}{2}$  она соответствует случайной ROC-кривой, т.е. из построения очевидно ожидать, что требование 4 также будет выполнено (рис. 3).

Расчет показателя точности  $AR$  скоринговой модели по формуле (3), определяющей найденную нейтральную ROC-кривую (формула (8)), не составляет труда:

$$\begin{aligned} AR &= 2(1 + \beta) \int_0^1 \frac{x}{x + \beta} dx - 1 = \\ &= 2(1 + \beta) \left( 1 - \beta \times \ln\left(1 + \frac{1}{\beta}\right) \right) - 1. \end{aligned} \tag{9}$$

Как нетрудно убедиться,  $AR \in (0, 1)$  при  $\beta \in (0, \infty)$  и требование 4 выполнено.

Формула для показателя точности второго порядка ( $LAR = RAR$ ) не имеет окончательного выражения в элементарных функциях:

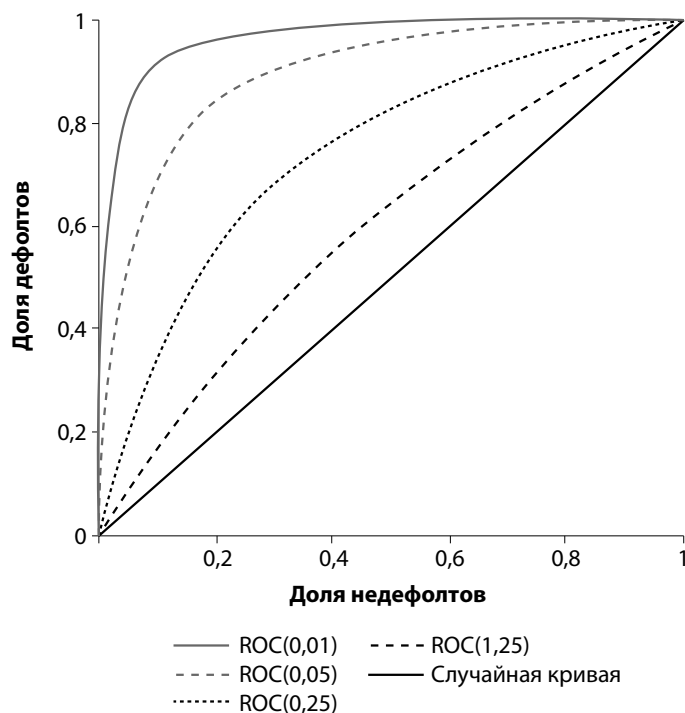
$$\begin{aligned} LAR = RAR &= 2\beta \times \\ &\times \left( (1 + \beta) \ln\left(1 + \frac{1}{\beta}\right) - 1 + \ln(\beta) \times \right. \\ &\times \left. \ln\left(1 + \frac{1}{\beta}\right) + PL\left(0, \frac{1}{\beta}\right) \right) + 1, \end{aligned} \tag{10}$$

где  $PL(a, b) = \int_a^b \frac{\ln(x)}{1 + x} dx$  вычисляется путем численного интегрирования.

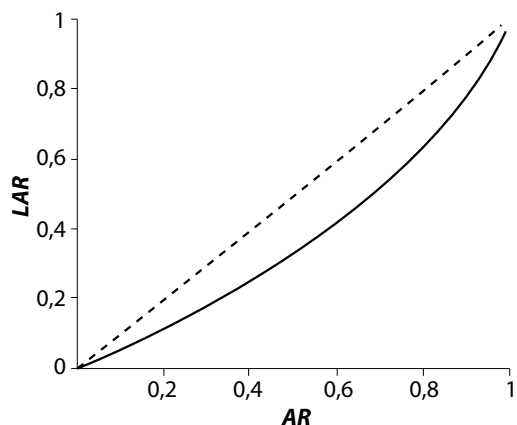
Зависимость  $LAR$  от  $AR$  для предложенной целево-нейтральной ROC-кривой представлена на рис. 4.

С практической точки зрения для анализа (валидации) ROC-кривой, характеризующей исследуемую

**Рис. 3.** Целево-нейтральные ROC-кривые для различных параметров  $\beta$



**Рис. 4.** Зависимость  $LAR$  от  $AR$  для целево-нейтральной ROC-кривой



рейтинговую модель, необходимо для ее показателя точности  $AR$  найти параметр  $\beta$  из уравнения (9) (численно). Затем нужно рассчитать  $LAR = RAR$  по формуле (10) и сравнить его с  $LAR/RAR$  ROC-кривой исследуемой модели, вычисленный по формулам из предыдущего раздела. Полезно также визуальное сравнение ROC-кривой с нейтральной ROC (формула (8)).

Следующим шагом для полученной целево-нейтральной ROC будет построение соответствующей CAP-кривой, а также формулы для калибровочной вероятности дефолта  $PD(x)$ .

Пусть задана ROC-кривая в виде  $ROC = R(g)$ , вероятность дефолта  $PD$  обозначается как  $p(x)$ , где  $x$  — координата всех измерений (fraction of all obligors).

Доля «хороших» объектов по оси  $Ox$  ROC-кривой, равной  $g(x)$ , определяется по формуле (4), из которой следует, что:

$$p(x) = 1 - g'(x) \times (1 - D), \quad (11)$$

где  $D$  — доля дефолтов.

В то же время согласно формулам (2) и (5):

$$p(x) = \frac{D}{1-D} R'(g) \times (1 - p(x)).$$

Объединяя эти два тождества, получим дифференциальное уравнение на  $g(x)$ :

$$g'(x) = \frac{1}{1-D + D \times R'(g)}.$$

Это уравнение имеет граничное значение  $g(0) = 0$  и решается прямым интегрированием, из которого получается следующее тождество:

$$(1 - D) \times g + D \times R(g) = x, \quad (12)$$

задающее неявную функцию  $g(x)$ , зависящую от формы ROC-кривой  $R(g)$ .

После решения тождества (12) относительно  $g(x)$  в явном виде вероятность дефолта  $PD = p(x)$  вычисляется согласно формуле (11), а форма CAP-кривой — формуле (5) как  $CAP(x) = R(g(x))$ . Решение уравнения (11) для целево-нейтральной ROC-кривой (формула (8)) дает решение для  $g(x)$  в виде:

$$g(x, \beta, D) = \frac{x - \beta - D + \sqrt{(x - \beta - D)^2 + 4\beta(1-D)x}}{2(1-D)}. \quad (13)$$

Подстановка формулы (13) в (11) дает зависимость вероятности дефолта  $PD(x, \beta, D)$  в виде:

$$PD(x, \beta, D) = \frac{1}{2} \left( 1 - \frac{x + \beta - D - 2\beta D}{\sqrt{(x - \beta - D)^2 + 4\beta(1-D)x}} \right), \quad (14)$$

где  $x$  — координата объекта (квантиль) в выборке всех объектов.

На рис. 5 представлены калибровочные кривые  $PD(x)$  (формула (14)) для целево-нейтральных скоринговых моделей разных показателей точности  $AR$  при среднем уровне дефолта  $D = 5\%$ .

## ТРИАНГУЛЯЦИЯ РОС-КРИВОЙ

Простейшей геометрической иллюстрацией ROC-кривой, обладающей несимметричными

свойствами скоринговой модели, является треугольная ROC-кривая (рис. 6).

Такая ROC-кривая задается только двумя параметрами  $a, d$ , причем  $a \in [0, 1 - d]$ . Очевидно, что показатель Джини  $AR = d$ . Для нахождения  $LAR$  и  $RAR$  необходимо представить ROC-кривую в виде функции:

$$ROC(x) = \begin{cases} x \frac{a+d}{a}, & x \leq a; \\ x \frac{1-a-d}{1-a} + \frac{d}{1-a}, & x > a. \end{cases} \quad (15)$$

Затем следует воспользоваться формулами  $LAR$  и  $RAR$  для определения *Left AR*, *Right AR*.

Опуская достаточно громоздкие, но стандартные вычисления, получаем:

$$\begin{aligned} LAR(a, d) &= a \times \ln(a) - \frac{1-a}{1-a-d} \times \\ &\quad \times (a+d) \times \ln(a+d); \\ RAR(a, d) &= (1-a-d) \times \ln(1-a-d) - \\ &\quad - \frac{a+d}{a} \times (1-a) \times \ln(1-a). \end{aligned} \quad (16)$$

В силу свойства симметричности  $LAR$  и  $RAR$  (формула (6)) при симметричном обратном преобразовании кривых должно наблюдаться тождество  $LAR$  и  $RAR$  (формулы (15)). Если параметр  $a$  для  $LAR$  представить в виде  $a(LAR) = x \times (1 - d)$ ,  $a(RAR) = (1 - x) \times (1 - d)$ , то нетрудно убедиться, что формулы (16) станут тождественными.

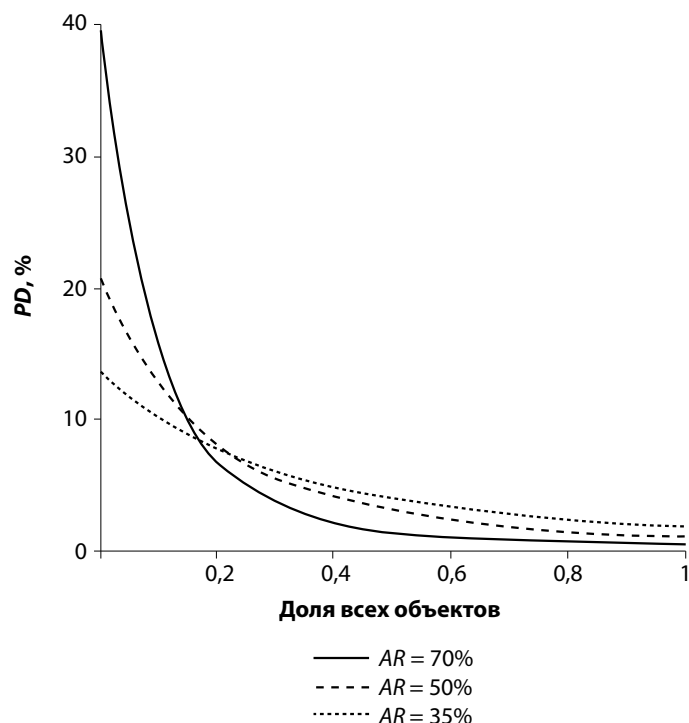
На рис. 7 представлены графики метрик  $LAR(x, d)$  и  $RAR(x, d)$  при нормированном значении  $a = x \times (1 - d)$ .

Видно, что метрики симметричны относительно обращения оси  $x \rightarrow 1 - x$ .

Из формул (16) следуют предельные значения метрик  $LAR$  и  $RAR$  для заданной точности скоринговой модели  $AR$ . Если значение  $a$  справа устремить к нулю ( $a \rightarrow 0 + \varepsilon$ ), то получим ROC-кривую, максимально смещенную влево, с максимальным  $LAR$  и минимальным  $RAR$ , если  $a \rightarrow 1 - d - \varepsilon$ , то получится наоборот (минимальный  $LAR$ , максимальный  $RAR$ ). Нетрудно вычислить пределы формул (16):



**Рис. 5.** Калибровочные зависимости вероятности дефолта от доли всех объектов скоринга для различных показателей точности  $AR$



$$\begin{aligned} \min(LAR, RAR) &= AR + (1 - AR) \times \ln(1 - AR); \\ \max(LAR, RAR) &= -\frac{AR \times \ln(AR)}{1 - AR}. \end{aligned} \quad (17)$$

Треугольные метрики  $a(LAR)$  и  $a(RAR)$ , полученные из заданной ROC-кривой с известным показателем Джини ( $AR = d$ ), имеют очевидный геометрический смысл. Если решить трансцендентные уравнения (15), то они будут параметризовать ROC-кривую заданной скоринговой модели, указывая размер и протяженность трех ее основных зон вероятности дефолта, условно разделяя их по принципу светофора на

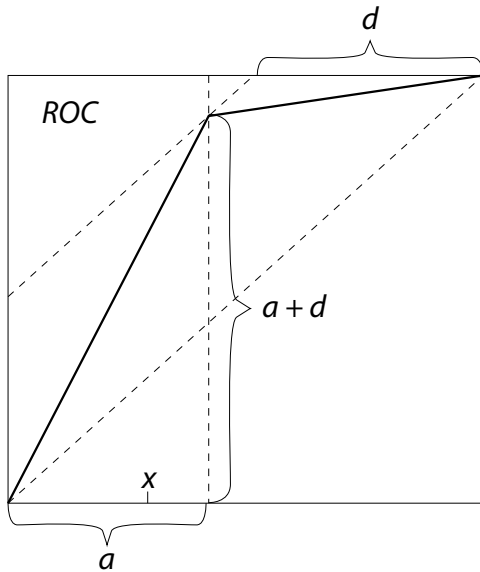
«красную»  $(0, a(LAR)]$ , «желтую»  $(a(LAR), a(RAR)]$  и «зеленую»  $(a(RAR), 1]$  по оси X («Доля недефолтов») (рис. 8).

При этом степень раскраски «светофора» будет определяться наклоном треугольной ROC-кривой (формула (15)), кратным ожидаемой вероятности дефолта  $PD$  всей популяции<sup>1</sup>. В «красной», «зеленой» и «желтой» зонах соответственно:

$$\begin{aligned} PD_{red} &= PD \times \mu DL; \\ PD_{green} &= PD \times \mu DR; \\ PD_{yellow} &= PD. \end{aligned}$$

<sup>1</sup> Это не строгое определение, но при  $PD \rightarrow 0$  оно становится точным. — Прим. авт.

Рис. 6. Треугольная ROC-кривая



Мультипликаторы  $\mu_{DL}$ ,  $\mu_{DR}$  левых и правых зон ROC-кривой (среди популяции недефолтных объектов) рассчитываются следующим образом:

$$\begin{aligned} \mu_{DL} &= \frac{a(LAR) + AR}{a(LAR)}; \\ \mu_{DR} &= \frac{1 - a(RAR) - AR}{1 - a(RAR)}. \end{aligned} \tag{18}$$

При этом  $\mu_{DL} > 1$ ,  $\mu_{DR} < 1$ .

Чем выше мультипликатор  $\mu_{DL}$ , тем выше левое предпочтение ROC-кривой. Чем ниже  $\mu_{DR}$ , тем выше правое целевое предпочтение при равном  $AR$ . При увеличении общей дискриминационной точности  $AR$  оба показателя будут вести себя монотонно:  $\mu_{DL}$  — расти,  $\mu_{DR}$  — падать. Параметры  $AR$ ,  $\mu_{DL}$ ,  $\mu_{DR}$  могут быть также альтернативной параметризацией ROC-кривой вместо  $AR$ ,  $LAR$ ,  $RAR$ .

Таким образом,  $AR$ ,  $LAR$  и  $RAR$  дают информативный «слепок» с ROC-кривой заданной скоринговой модели, геометрически отображающийся

в параметры  $d = AR$ ,  $a(LAR)$ ,  $a(RAR)$  соответственно (см. рис. 6).

В Приложении 1 представлены варианты триангуляционных калибровок мультипликаторов левых и правых зон для двух ROC-кривых реальных скоринговых моделей левого и правого целевого предпочтения. ROC-кривые показаны на рис. 9. Как и следовало ожидать, левый мультипликатор  $\mu_{DL}$  для модели левого целевого предпочтения существенно больше, чем для модели правого, несмотря на то что правая скоринговая модель имеет заметно больше прогнозной мощности (дискриминационной точности)  $AR$ . Аналогичные выводы можно сделать и для  $\mu_{DR}$ , только  $PD$  будет снижаться в правой части недефолтной популяции объектов, упорядоченных по рейтингу.

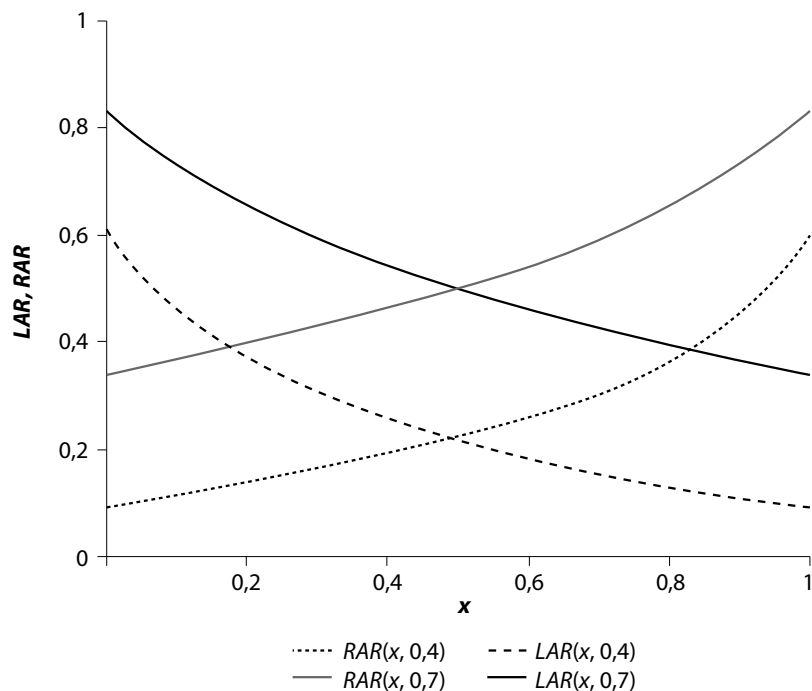
### ОБЩАЯ ФОРМУЛА МОДЕЛИ ROC-КРИВОЙ ДЛЯ СКОРИНГОВЫХ МОДЕЛЕЙ ПРАВОГО ИЛИ ЛЕВОГО ЦЕЛЕВОГО ПРЕДПОЧТЕНИЯ

В работе С. Сэтчела и В. Ся вводится понятие популяционной ROC-кривой для построения соответствующих аналитических моделей и анализируются ROC-кривые, ведомые несколькими известными распределениями (распределением Вейбулла, логистическим и нормальным), а также их комбинациями [9]. Авторы предлагают пользователю самому сделать выбор наиболее эффективной модели ROC-кривой среди предложенного ими многообразия, не предполагая однозначности этого выбора, что нивелирует практическую ценность подхода.

Рассмотрим наиболее практичную модель ROC-кривой, предложенную в работе М. ван дер Бюргта [13]. Она проста для анализа и рекомендуется для калибровки скоринговых моделей [12], применяется при практической оценке эффективности риск-менеджмента [1]. В качестве модели ROC-кривой берется однопараметрическая модель:

$$ROC(x) = \frac{1 - e^{-kx}}{1 - e^{-k}}$$

Рис. 7. Метрики LAR и RAR для треугольных ROC-кривых



Примечание:  $d = 0,4, 0,7$  ( $AR = 0,4, 0,7$ ).

Показатель Джини ( $AR$ ) для нее вычисляется по формуле:

$$AR = 2 \times \left( \frac{1}{1 - e^{-k}} - \frac{1}{k} - \frac{1}{2} \right).$$

Решая это трансцендентное уравнение относительно  $k$ , строим функцию  $k(AR)$ .

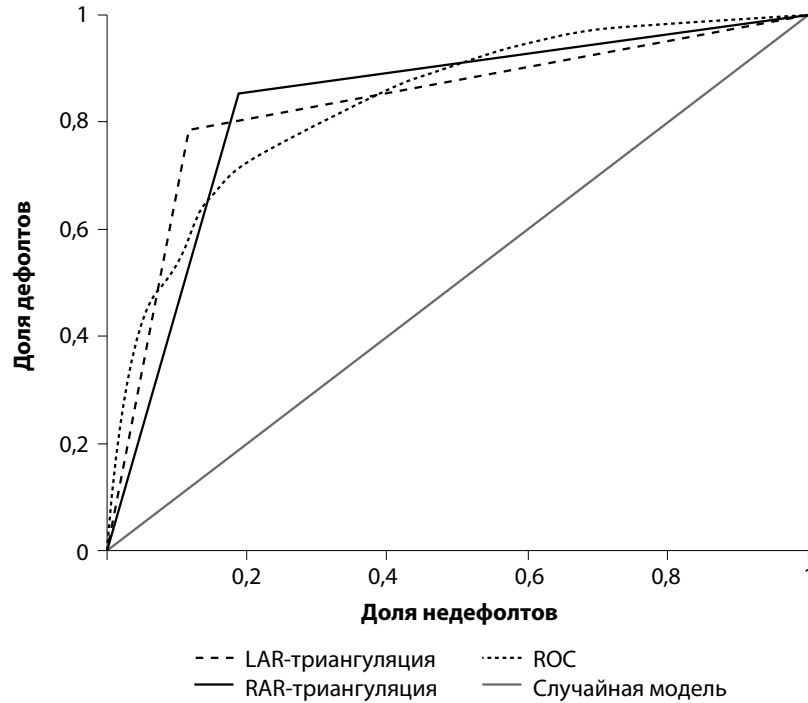
Используя определения из раздела «Левый и правый коэффициенты точности скоринговой модели», вычисляем  $LAR(k)$ ,  $RAR(k)$ , а также  $k(AR)$ , строим зависимости  $LAR(AR)$ ,  $RAR(AR)$  для кривой ван дер Бюргта, представленные на рис. 10.

Зависимости  $\min(LAR, RAR)$  и  $\max(LAR, RAR)$  от  $AR$  установлены соотношениями (17). Из графиков, представленных на рис. 10, видно, что кривая ван дер Бюргта имеет исключительно правое

целевое предпочтение, поскольку на диапазоне показателя Джини  $AR = (0, 1)$  всегда  $RAR > LAR$ , поэтому калибровать скоринговые модели левого целевого предпочтения с помощью данной кривой нельзя.

Возникает очевидный и закономерный вопрос о том, какую универсальную модель ROC-кривой использовать для калибровки произвольной скоринговой модели, имеющей как правое предпочтение, так и левое. Понятно, что для практической эффективности кривая должна быть не более чем двухпараметрической и в предельных случаях достигать предельных уровней (формулы (17)) для показателей точности  $LAR$  и  $RAR$ , а также сводиться к вырожденным треугольным кривым (см. рис. 6) для  $a = 0$  (левая вырожденная ROC) и  $a = 1 - d$  (правая вырожденная ROC).

**Рис. 8.** Триангуляция ROC-кривой скоринговой модели



Примечание:  $AR = 0,667$ ,  $LAR = 0,53$ ,  $RAR = 0,486$ ,  $(\alpha(LAR) = 0,116$ ,  $\alpha(RAR) = 0,185$ ).

Предлагается строить гиперболическую кривую по аналогии с построением целево-нейтральной кривой (формула (8)).

Сначала построим кривую, расположенную в главном угле системы координат XOY (рис. 11а), затем проведем преобразование  $X' = X$ ,  $Y' = 1 - Y$ , как представлено на рис. 11. Для этого преобразования получается правая ROC-кривая (рис. 11б). Получить левую ROC-кривую можно из правой ROC-кривой, сделав симметричное преобразование  $X'' = 1 - Y'$ ,  $Y'' = 1 - X'$  (рис. 11в). Целево-нейтральная кривая (формула (8)) получается при  $d = 1$ .

Кривая на рис. 11а должна иметь две асимптоты, одна параллельна оси OX, другая — гипотенузе левого треугольника  $(1 - d, 0, Y)$ , а также проходить через точки  $(0, 1)$  и  $(1, 0)$ . Первые два условия

задают вид ROC-кривой в четырехпараметрической форме:

$$y + \beta = \frac{\alpha}{(y + \beta) \times (1 - d) + x + \beta_1},$$

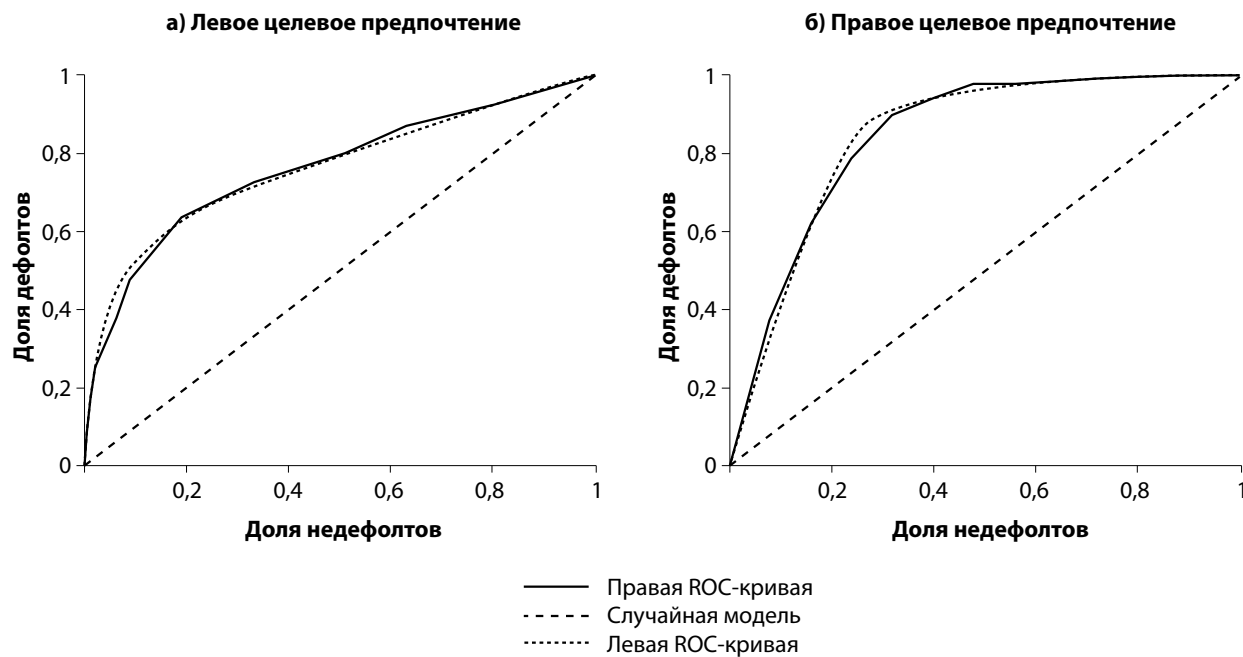
при выполнении условий пересечения:

$$\begin{cases} x = 0, y = 1 \\ x = 1, y = 0 \end{cases} \Rightarrow \begin{cases} \beta_1 = \beta \times d - (1 - d) \times (1 - \beta) \\ \alpha = \beta \times (1 - \beta) \times d, \end{cases}$$

откуда получается правая ROC-кривая в неявном двухпараметрическом виде:

$$\begin{aligned} x &= y \times (1 - d) - \beta \times d + \\ &+ \frac{\beta \times (1 - \beta) \times d}{1 + \beta - y} = \\ &= \frac{y \times (1 + \beta - d - (1 - d) \times y)}{1 + \beta - y}. \end{aligned} \tag{19}$$

Рис. 9. Калибровка модельными ROC-кривыми реальных ROC-кривых двух скоринговых моделей



Из формулы (19) хорошо видна асимптота, параллельная прямой  $y = \frac{1}{1-d}x + b$ .

Используя интегрирование неявной функции  $AUC^- = \int_0^1 x(y) \times dy$ , нетрудно найти:

$$AR = 1 - 2 \times AUC^-, \quad (20)$$

$$AR = d \times \left( 2(1+\beta) \left( 1 - \beta \times \ln \left( 1 + \frac{1}{\beta} \right) \right) - 1 \right).$$

Приведенное далее выражение (21) соответствует формуле (9), построенной для целево-нейтральной ROC-кривой, при  $d = 1$ .

Левая ROC получается после замены  $1 - y \rightarrow x$ ,  $1 - x \rightarrow y$  и вычисляется уже в явном двухпараметрическом виде:

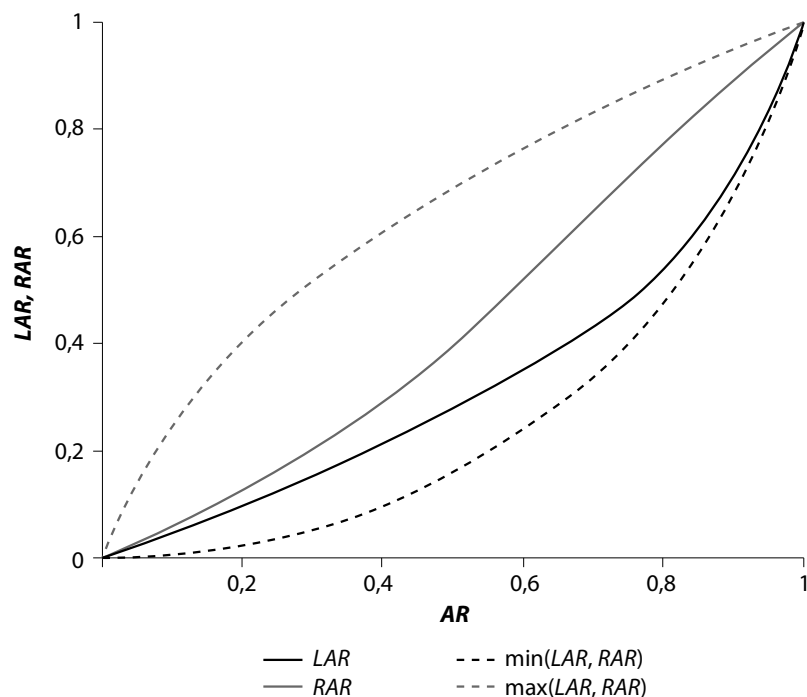
$$y = x \times (1-d) + d \times (1-\beta) - \frac{\beta \times (1-\beta) \times d}{x + \beta} = \frac{x \times (\beta + d + (1-d) \times x)}{x + \beta}. \quad (21)$$

Нетрудно проверить, что  $AR$  для левой ROC вычисляется по той же формуле (20).

Очевидно, что при  $\beta \rightarrow 0$  кривая (21) вырождается в прямую  $y = d + (1-d) \times x$ , соответствующую значению  $a = 0$  (см. рис. 6). Аналогично при  $\beta \rightarrow 0$  из формулы (19) получится и вырожденная правая ROC-кривая для  $a = 1 - d$  (см. рис. 6), обладающая максимальным  $RAR$  для заданного  $AR = d$ .

Для калибровки скоринговой модели, заданной произвольной ROC-кривой, потребуется вычислить коэффициенты точности первого и второго порядка, т.е. Джини ( $AR$ ) и  $LAR$ ,  $RAR$  соответственно. Вычисления проводятся по численным формулам (7), являющимся числовыми аналогами соответствующих определений показателей точности. Затем следует определить целевое предпочтение скоринговой модели. Для этого достаточно сравнить  $LAR$  и  $RAR$ . Если  $LAR > RAR$ , то целевое предпочтение ROC-кривой является левым и в качестве калибровочной функции (модели

**Рис. 10.** Зависимость левой точности *LAR* и правой *RAR* от Джини *AR* для кривой ван дер Бюргта



ROC-кривой) следует выбрать формулу (21), а если наоборот, то (19). В случае левого предпочтения мы используем для калибровки значение *LAR*, а в случае правого берется рассчитанное значение *RAR*. Обобщенно можно полагать, что выбранное значение *sAR* (side *AR*), равное  $\max(LAR, RAR)$ , является главной характеристикой ROC-кривой. Вторая главная характеристика — это *AR*. Эти два параметра позволят однозначно вычислить константы  $\beta, d$  модельной ROC-кривой после определения ее целевого предпочтения (левого или правого).

Поскольку модели правой и левой ROC-кривой являются симметричными, не имеет значения, какую кривую использовать для вычисления индекса второго порядка точности, для левой это будет *LAR*, для правой — *RAR*, а формулы будут идентичны.

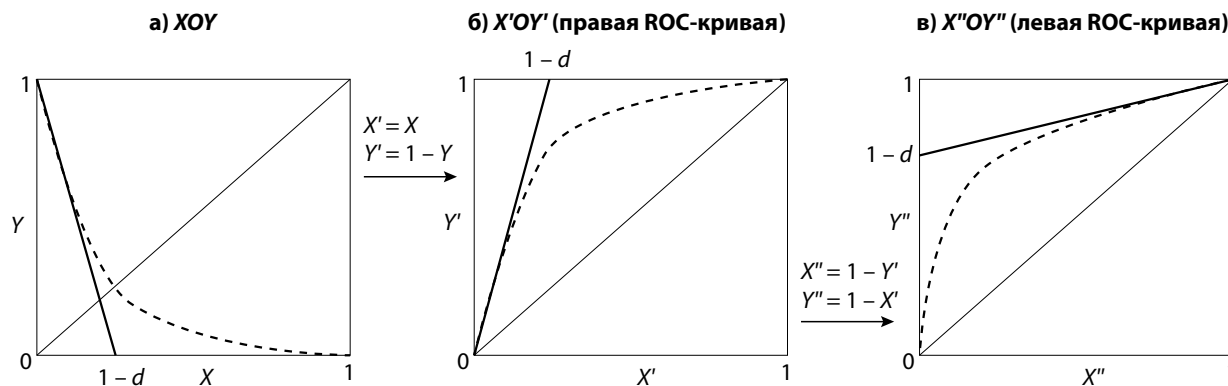
По этой причине для упрощения и без того существенно громоздких вычислений достаточно вывести формулу  $LAR(\beta, d)$  для левой ROC-кривой:

$$LAR(\beta, d) = 2 \times \int_0^1 \int_0^x \frac{y(s, \beta, d) ds}{y(x, \beta, d)x} dx - 1 = 2 \times \int_0^1 \frac{(x + \beta) \times A(x, \beta, d)}{x^2 \times (x(1-d) + \beta + d)} dx - 1,$$

где  $A(x, \beta, d) = d \times (1 + \beta) \times x + \frac{x^2}{2} (1 - d) - d \times \beta \times (1 + \beta) \times \ln\left(1 + \frac{x}{\beta}\right)$ .

Опуская объемные вычисления и учитывая, что полученная формула  $sAR(\beta, d)$  будет универсальна для левой и правой ROC, результат можно записать следующим образом:

Рис. 11. Преобразование XOY в X'OY' (правая ROC-кривая) и далее в X''OY'' (левая ROC-кривая)



$$\begin{aligned}
 sAR(\beta, d) &= d \times \frac{1+\beta}{d+\beta} \times \\
 &\times \left( \frac{d-\beta+2\beta \times d}{1-d} \times \ln\left(\frac{1+\beta}{d+\beta}\right) + 2\beta \times \right. \\
 &\times \left( (1+\beta) \times \ln\left(1+\frac{1}{\beta}\right) - 1 + \frac{d \times (1+\beta)}{d+\beta} \times \right. \\
 &\times \left. \left( \ln\beta \times \ln\left(1+\frac{1}{\beta}\right) + \ln\left(\frac{d(1+\beta)}{(1-d) \times \beta}\right) \times \right. \right. \\
 &\times \ln\left(\frac{1+\beta}{d+\beta}\right) + PL\left(0, \frac{1}{\beta}\right) + \\
 &\left. \left. \left. + PL\left(\frac{(1-d) \times \beta}{d \times (1+\beta)}, \frac{1-d}{d}\right) \right) \right) \right), \tag{22}
 \end{aligned}$$

где  $PL(a, b)$  определено в формуле (10). Таким образом можно решить систему уравнений с двумя неизвестными  $\beta, d$  для  $AR$  (20) и  $sAR$  (22) и определить параметры целево-ориентированной ROC-кривой.

Минимальное значение  $sAR = sAR_0$  будет достигаться для целево-нейтральной ROC-кривой при значении параметра точности первого порядка  $AR$ , равного калибруемой ROC-кривой. Минимальное значение  $sAR_0 = LAR = RAR$  определяется по

формуле (10), которая является пределом выражения (22) при  $d \rightarrow 1$ . В то же время максимальное значение  $sAR (sAR_{max})$  вычисляется по формуле (17). Это позволит табулировать решения  $\beta, d$  уравнений (10), (22) в зависимости от:

$$\begin{aligned}
 AR &\in (0, 1); \\
 p &= \frac{sAR - sAR_0}{sAR_{max} - sAR_0} \in (0, 1),
 \end{aligned}$$

где  $sAR$  — вычисленный из ROC-кривой главный параметр точности второго порядка заданной скоринговой модели после определения ее целевого предпочтения.

В Приложении 2 даны табулированные матрицы параметров  $\beta, d$ , являющиеся решениями уравнений (20), (22) для равноотстоящих узловых  $AR$  и  $p$ , а также строковые нормировочные параметры  $sAR_{max}$  и  $sAR_0$ , зависящие только от  $AR$ . Применение таблиц, представленных в Приложении 2, позволит форсировать точечное решение нелинейных уравнений (20), (22), что, конечно, снизит точность для конкретных данных. Однако наличие естественной статистической ошибки при построении ROC-кривых для ограниченной данными разрабатываемой скоринговой модели даст приемлемый допуск для ошибки округления до

приведенных данных при применении элементарной линейной интерполяции значений  $\beta, d$  по трем ближайшим узловым точкам.

Пример для двух реализованных на практике скоринговых моделей представлен в Приложении 1. В таблицах Приложения 1 даны узловые точки ROC-кривой, над которыми производится расчет  $AR, LAR, RAR$ . Определяется правое либо левое целевое предпочтение, затем решаются уравнения (20), (22) и выдаются калибровочные константы  $\beta, d$  для построения модельной ROC-кривой. Построенные модельные ROC-кривые, реализованные по формулам (19), (21) в соответствии с выявленным целевым предпочтением скоринговых моделей, показаны на рис. 6.

При недостатке узловых точек наблюдений (например, по фиксированным рейтинговым разрядам) для приемлемой точности калибровки реальных ROC-кривых модельными аналогами рекомендуется равномерное распределение точек наблюдений по оси недефолтов либо по оси дефолтов.

### КАЛИБРОВКА СКОРИНГОВЫХ МОДЕЛЕЙ НА ВЕРОЯТНОСТЬ ДЕФОЛТА С УЧЕТОМ ИХ ЦЕЛЕВОГО ПРЕДПОЧТЕНИЯ

Задача калибровки функции вероятности дефолта для скоринговых (рейтинговых) моделей является ключевой для кредитного риск-менеджмента. От нее зависит принятие кредитных или инвестиционных решений. Калибровка конкретной скоринговой модели не должна осуществляться без учета ее целевого предпочтения, поскольку при калибровке на привычную общепризнанную зависимость это может привести к существенной модельной ошибке, порождаемой искаженным распределением компаний по рейтинговым разрядам, и, как следствие, к неверным решениям и финансовым потерям кредитора или инвестора.

В рамках представленного подхода возможно выстроить формулы калибровки, учитывающие целевое предпочтение скоринговой модели.

Калибровка целево-нейтральной кривой уже была проведена в соответствующем разделе статьи, поэтому в дальнейших рассуждениях достаточно ее скорректировать с учетом двух параметров модельной ROC-кривой заданного предпочтения, а не одного. Как и в разделе «Калибровочная целево-нейтральная ROC-кривая», предполагается, что задана ROC-кривая в виде  $ROC = R(g)$ , вероятность дефолта  $PD$  обозначается как  $p(x)$ , где  $x$  — координата двух фракций дефолта и недефолта,  $g = g(x)$  — доля недефолтных заемщиков в выборке  $x$  всех объектов (дефолтных и недефолтных), упорядоченных по возрастанию скоринговой оценки (рейтинга). Уровень дефолта  $D$  всей популяции предполагается заданным.

Левая ROC-кривая задана уравнением (21), которое имеет явный вид, поэтому из соотношения (12) будет следовать уравнение на  $g(x)$ , которое следует решить:

$$(1 - D) \times g \times (g + \beta) + D \times g \times ((1 - d) \times g + \beta + d) = x \times (g + \beta).$$

После раскрытия скобок это уравнение превращается в квадратное и имеет единственное непротиворечащее граничным условиям решение:

$$g(x) = \frac{x - \beta - D \times d + \sqrt{DL(x, \beta, d, D)}}{2 \times (1 - D \times d)},$$

где  $DL(x, \beta, d, D) = (x - \beta - D \times d)^2 + 4 \times x \times \beta \times (1 - D \times d)$ .

Учитывая формулу (11), для  $p(x) = 1 - g'(x) \times (1 - D)$  после недолгих преобразований можно представить результат калибровки ROC-кривой левого предпочтения на вероятность дефолта  $PD$  в следующем виде:

$$PDL(x, \beta, d, D) = \frac{1}{2 \times (1 - D \times d)} \times \left. \begin{aligned} & \times (1 + D - 2 \times D \times d - (1 - D) \times \\ & \times \frac{x + \beta - D \times d \times (1 + 2 \times \beta)}{\sqrt{DL(x, \beta, d, D)}} \end{aligned} \right\} \quad (23)$$

Формула (23) является конечной и явной формулой для калибровки скоринговой модели



левого предпочтения, которая вырождается в целево-нейтральную калибровочную формулу (14) при  $d \rightarrow 1$ .

Сложнее получить калибровку для  $PD$  скоринговой модели правого предпочтения из-за неявной (формула (19)) формы ее модельной ROC-кривой. Имеется уравнение правой ROC-кривой  $y = R(x)$ , заданное неявно в виде  $x = S(R)$ , где  $S()$  — известная функция (формулы (19), (12)) имеет вид:

$$(1 - D) \times g + D \times R(g) = x,$$

откуда:

$$R(g) = \frac{x - (1 - D) \times g}{D},$$

$$S(R(g)) = g = S\left(\frac{x - (1 - D) \times g}{D}\right).$$

Последнее соотношение превращается в квадратное уравнение, но с более сложной параметрической структурой:

$$D \times g \times ((1 + \beta) \times D - x + (1 - D) \times g) =$$

$$= (x - (1 - D) \times g) \times ((1 + \beta - d) \times$$

$$\times D - (1 - d) \times (x - (1 - D) \times g)).$$

Это уравнение успешно, но более громоздко, чем в случае левого предпочтения, решается для  $g(x)$ :

$$g(x) = \frac{x \times (D + 2 \times (1 - D) \times (1 - d))}{2 \times (1 - D) \times (1 - d + D \times d)} -$$

$$- \frac{D \times (1 + \beta - (1 - D) \times d)}{2 \times (1 - D) \times (1 - d + D \times d)} +$$

$$+ \frac{\sqrt{DR(x, \beta, d, D)}}{2 \times (1 - D) \times (1 - d + D \times d)},$$

где  $DR(x, \beta, d, D) = (x \times (D + 2 \times (1 - D) \times (1 - d)) - D \times (1 + \beta - (1 - D) \times d))^2 + 4 \times x \times (1 - D) \times (1 - d + D \times d) \times ((1 + \beta - d) \times D - (1 - d) \times x)$ .

После применения формулы (11), проведения упрощающих преобразований получается искомая калибровка правой ROC-кривой на вероятность дефолта  $PD$  в виде:

$$PDR(x, \beta, d, D) = \frac{D}{2 \times (1 - d + D \times d)} \times \quad (24)$$

$$\times \left( 1 - D \times \frac{x + \beta \times (2 \times (1 - D) \times d - 1) + (1 - D) \times d - 1}{\sqrt{DR(x, \beta, d, D)}} \right).$$

При  $d \rightarrow 1$  формула правой калибровки (24) также превращается в целево-нейтральную калибровку (14). В этом нетрудно убедиться несмотря на обилие параметров и кажущуюся сложность.

Таким образом, получены окончательные явные формулы для калибровки на вероятность дефолта целево-ориентированных скоринговых моделей. Для левого предпочтения  $PDL(x, \beta, d, D)$  задается формулой (23), для правого  $PDR(x, \beta, d, D)$  — формулой (24).

На рис. 12 представлены варианты калибровки скоринговых моделей с левым и правым целевым предпочтением при постоянной точности дискриминации  $AR = 50\%$  и средним уровнем дефолта  $D = 5,5\%$ . Сверху представлены графики

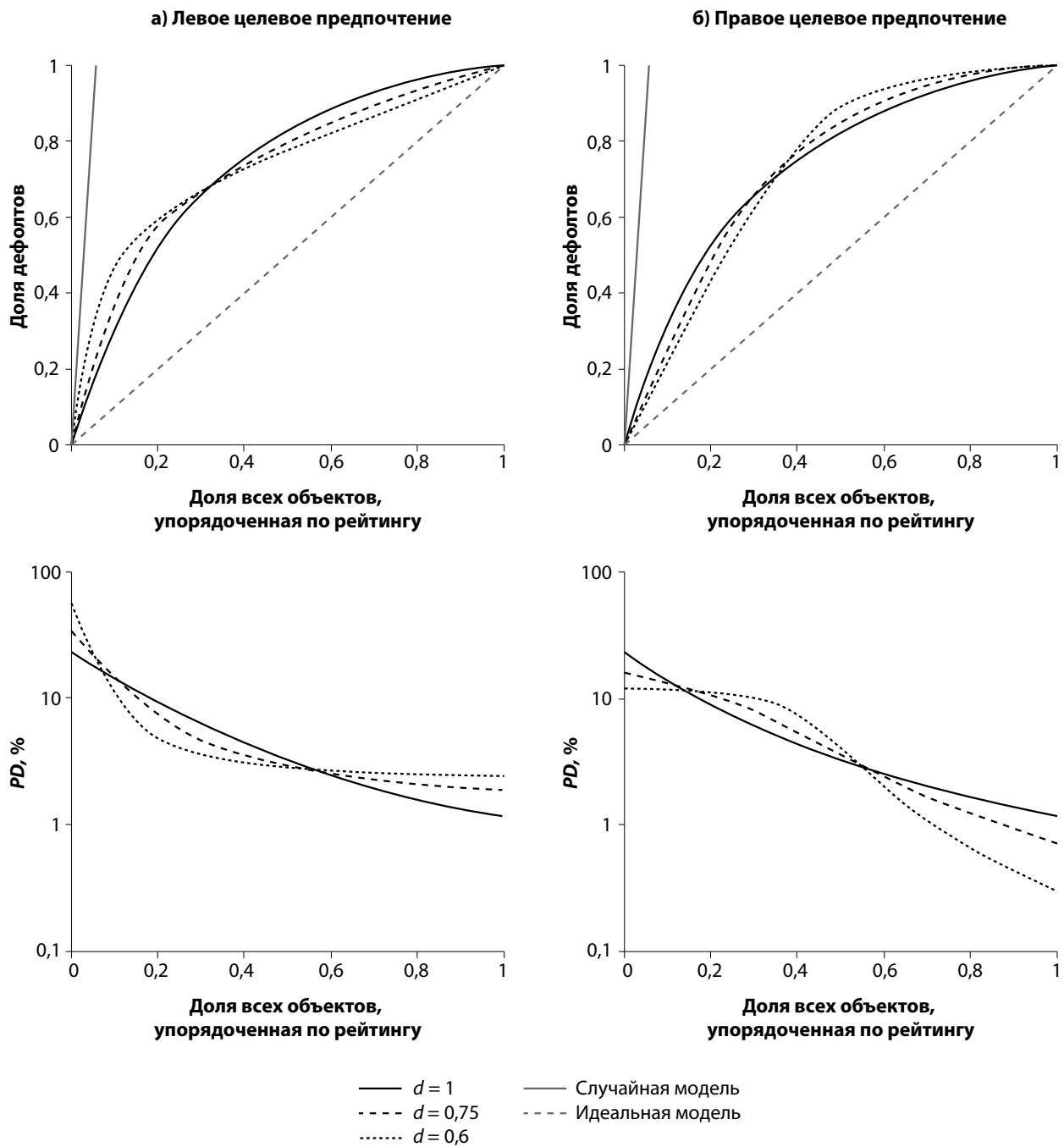
$$CAP(x) = \frac{\int_0^x PD(s) ds}{D} = \frac{x - (1 - D) \times g(x)}{D} \text{ (из формулы$$

(4)), снизу —  $PD(x)$  в логарифмическом масштабе. Предпочтением целевой ориентации управляет параметр  $d$ , при  $d = 1$  ROC-кривая становится целево-нейтральной.

Из графиков на рис. 12а  $PD(x)$  видно, насколько чувствительность вероятности дефолта  $PD(x)$  в начале интервала аргумента  $x$  для моделей левого предпочтения выше нейтральной. В то же время наблюдается низкая чувствительность вероятности дефолта к положению (рейтингу) во второй половине интервала аргумента  $x$ , т.е. модель хорошо различает «плохие» объекты рейтинга. Для моделей правого предпочтения, наоборот, вероятность дефолта в первой половине интервала аргумента  $x$  не чувствительна к рейтингу, однако при приближении к максимальному рейтингу в его правой части она существенно падает по сравнению с нейтральным вариантом скоринговой модели той же мощности  $AR$ , т.е. правая модель эффективно различает «хорошие» объекты рейтинга.

Скоринговая (рейтинговая) модель калибруется на вероятность дефолта относительно

Рис. 12. Результат калибровки ROC-кривых



Примечание: параметр точности скоринговых моделей  $AR = 0,5$ , параметр дефолтности  $D = 0,055$ .

перцентиль  $x$  рейтингуемого объекта в опорной выборке, участвующей в калибровке. Для окончательного решения задачи калибровки необходимо сделать последний шаг — оценить этот перцентиль  $x(R)$ , зависящий от рейтингового балла  $R$ . Если распределение рейтингового балла по частотам близко к нормальному с известными средним  $\mu$  и стандартным отклонением  $\sigma$ , то сделать это очень просто:

$$x(R) = N(R, \mu, \sigma).$$

Если распределение более сложное, то однозначной рекомендации, какой функцией фиттировать перцентиль  $x(R)$ , быть не может, но можно использовать удачные решения, например предложенные в работе Е. Кануласа и др. [7] Обзор содержательного опыта решения проблемы фиттирования распределений представлен в статье А. Арампатзиса и др. [2], где обсуждаются гипотезы универсализации распределений и приближений к нормально-экспоненциальным моделям. Из опыта автора одна из наиболее часто удовлетворительных моделей фиттирования распределения скоринг-баллов — это скошенно-нормальная четырехпараметрическая модель  $x(R) = N(\alpha \times |R - \beta|^\gamma \times \text{sign}(R - \beta) + \delta)$ , которую надо приближать к наблюдениям, используя широко известный критерий согласия Пирсона.

## ЗАКЛЮЧЕНИЕ

В настоящей работе представлены две новые специализированные метрики точности скоринговых моделей второго порядка, которые показывают целевое предпочтение моделей. Левая ( $LAR$ ) и правая ( $RAR$ ) метрики имеют как интегральное представление, так и численное для расчета по точной ROC-кривой. Метрики лежат в основе триангуляции ROC-кривой, komponуя ее свойства в три экономически очевидных параметра, причем триангуляция может быть проведена как для рейтинговой модели в целом, так и ее компонентов в частности. С использованием метрик  $LAR/RAR$  разработано двухпараметрическое универсальное гиперболическое семейство модельных ROC-кривых, которое изометрически гомеоморфно всем возможным практическим ROC-кривым, определенным тремя главными метриками  $AR, LAR, RAR$ . Это позволило разработать универсальные формулы калибровки функции вероятности дефолта скоринговых моделей, учитывающие целевые предпочтения моделей, левую, правую и нейтральную. В приложениях даны численные примеры показателей, рассчитанные из конкретных скоринговых моделей, а также калибровочные таблицы, которые позволят избежать громоздких вычислений на практике, где допустима ограниченная точность.

## ПРИЛОЖЕНИЕ 1.

Пример калибровки модели ROC-кривой для двух реализованных на практике скоринговых моделей

Левая ROC-кривая.

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$x_i$	0	0,000	0,000	0,001	0,003	0,005	0,014	0,024	0,061	0,090	0,191	0,332	0,510	0,630	0,800	1
$ROC_i$	0	0,027	0,080	0,097	0,133	0,159	0,230	0,257	0,381	0,478	0,637	0,726	0,800	0,870	0,920	1

Показатели точности дискриминации первого и второго порядка.

<i>AR</i>	<i>LAR</i>	<i>RAR</i>
0,523	0,509	0,391

Параметры модельной ROC-кривой.

$\beta$	$d$
0,0348	0,634

Параметры триангуляции ROC-кривой.

$\alpha(LAR)$	$\mu DL$	$\alpha(RAR)$	$\mu DR$
0,077	7,75	0,312	0,240

Правая ROC-кривая.

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$x_i$	0	0,08	0,16	0,24	0,32	0,4	0,48	0,56	0,64	0,72	0,8	0,88	0,96	1
$ROC_i$	0	0,377	0,626	0,786	0,9	0,95	0,973	0,98	0,988	0,991	0,995	0,998	0,999	1

Показатели точности дискриминации первого и второго порядка.

<i>AR</i>	<i>LAR</i>	<i>RAR</i>
0,69	0,415	0,676

Параметры модельной ROC-кривой.

$\beta$	$d$
0,0147	0,764

Параметры триангуляции ROC-кривой.

$\alpha(LAR)$	$\mu DL$	$\alpha(RAR)$	$\mu DR$
0,216	4,19	0,265	0,061

**ПРИЛОЖЕНИЕ 2.**

Таблицы параметров для целево-ориентированных скоринговых моделей

Параметры нормировки.

<b>AR, %</b>	<b>15</b>	<b>20</b>	<b>25</b>	<b>30</b>	<b>35</b>	<b>40</b>	<b>45</b>	<b>50</b>	<b>55</b>	<b>60</b>	<b>65</b>	<b>70</b>	<b>75</b>	<b>80</b>
<b>sAR<sub>0</sub></b>	0,081	0,111	0,142	0,175	0,210	0,246	0,285	0,326	0,369	0,415	0,465	0,517	0,574	0,636
<b>sAR<sub>max</sub></b>	0,335	0,402	0,462	0,516	0,565	0,611	0,653	0,693	0,731	0,766	0,800	0,832	0,863	0,893

Параметр *d*.

<b>p / AR, %</b>	<b>15</b>	<b>20</b>	<b>25</b>	<b>30</b>	<b>35</b>	<b>40</b>	<b>45</b>	<b>50</b>	<b>55</b>	<b>60</b>	<b>65</b>	<b>70</b>	<b>75</b>	<b>80</b>
<b>10</b>	0,424	0,528	0,611	0,679	0,735	0,78	0,818	0,85	0,877	0,9	0,92	0,937	0,952	0,964
<b>20</b>	0,2927	0,383	0,464	0,536	0,6	0,656	0,707	0,751	0,791	0,826	0,858	0,886	0,911	0,933
<b>30</b>	0,2356	0,3132	0,386	0,455	0,518	0,577	0,631	0,681	0,727	0,77	0,808	0,844	0,877	0,907
<b>40</b>	0,2042	0,2731	0,34	0,404	0,465	0,523	0,578	0,63	0,679	0,725	0,769	0,81	0,848	0,883
<b>50</b>	0,1847	0,2475	0,3092	0,369	0,428	0,484	0,539	0,592	0,642	0,69	0,737	0,781	0,823	0,863
<b>60</b>	0,1717	0,2301	0,288	0,345	0,401	0,456	0,51	0,562	0,613	0,663	0,711	0,757	0,802	0,845
<b>70</b>	0,1628	0,218	0,2729	0,327	0,381	0,435	0,487	0,539	0,59	0,64	0,689	0,737	0,784	0,83
<b>80</b>	0,1566	0,2094	0,262	0,3145	0,367	0,419	0,47	0,521	0,572	0,622	0,672	0,721	0,77	0,818
<b>90</b>	0,1524	0,2034	0,2545	0,3055	0,356	0,407	0,458	0,508	0,559	0,609	0,659	0,709	0,758	0,807

Параметр  $\beta$ .

<b>p / AR, %</b>	<b>15</b>	<b>20</b>	<b>25</b>	<b>30</b>	<b>35</b>	<b>40</b>	<b>45</b>	<b>50</b>	<b>55</b>	<b>60</b>	<b>65</b>	<b>70</b>	<b>75</b>	<b>80</b>
<b>10</b>	0,49	0,44	0,38	0,32	0,272	0,229	0,19	0,157	0,128	0,103	0,082	0,063	0,047	0,033
<b>20</b>	0,229	0,218	0,201	0,181	0,161	0,141	0,122	0,104	0,087	0,072	0,059	0,046	0,035	0,0257
<b>30</b>	0,122	0,121	0,115	0,108	0,099	0,089	0,079	0,069	0,059	0,05	0,042	0,034	0,0261	0,0193
<b>40</b>	0,069	0,07	0,068	0,065	0,061	0,056	0,051	0,045	0,04	0,034	0,0289	0,0237	0,0188	0,0141
<b>50</b>	0,039	0,04	0,04	0,039	0,037	0,035	0,032	0,029	0,0258	0,0226	0,0194	0,0162	0,013	0,01
<b>60</b>	0,0215	0,0226	0,0228	0,0225	0,0217	0,0205	0,0192	0,0176	0,0159	0,0142	0,0123	0,0104	0,0085	0,0066
<b>70</b>	0,011	0,0117	0,012	0,012	0,0117	0,0112	0,0106	0,0099	0,009	0,0081	0,0072	0,0061	0,0051	0,004
<b>80</b>	0,0048	0,0052	0,0054	0,0054	0,0054	0,0052	0,005	0,0047	0,0043	0,004	0,0035	0,00308	0,00259	0,00209
<b>90</b>	0,00141	0,00154	0,00161	0,00165	0,00164	0,00162	0,00156	0,00149	0,0014	0,00129	0,00117	0,00104	0,00089	0,00073

## ЛИТЕРАТУРА

1. Помазанов М.В. Концепция мотивации эффективного управления кредитными рисками // Финансы и кредит. — 2020. — №26(11). — С. 2567–2593. — <https://doi.org/10.24891/fc.26.11.2567>.
2. Arampatzis A., Robertson S.E., Kamps J. (2011). «Modeling score distributions in information retrieval». *Information Retrieval Journal*, Vol. 14(1), pp. 26–46. — DOI: 10.1007/s10791-010-9145-5.
3. Cortes C., Vapnik V. (1995). «Support-vector networks». *Machine Learning*, Vol. 20(3), pp. 273–297.
4. Egan J.P. (1975). *Signal Detection Theory and ROC Analysis*. New York: Academic Press.
5. Engelmann B., Hayden E., Tasche D. (2003). *Measuring the Discriminative Power of Rating Systems*. — [https://www.researchgate.net/publication/23755139\\_Measuring\\_the\\_Discriminative\\_Power\\_of\\_Rating\\_Systems](https://www.researchgate.net/publication/23755139_Measuring_the_Discriminative_Power_of_Rating_Systems).
6. Hanley J.A., McNeil B.J. (1982). «The meaning and use of the area under a ROC curve». *Radiology*, Vol. 143, pp. 27–36.
7. Kanoulas E., Dai K., Pavlu V., Aslam J.A. (2010). «Score distribution models: assumptions, intuition, and robustness to score manipulation». *SIGIR '10: Proceedings of the 33d International ACM SIGIR Conference on Research and Development in Information Retrieval*. — <https://doi.org/10.1145/1835449.1835491>.
8. Miura K., Yamashita S., Eguchi S. (2010). «Area under the curve maximization method in credit scoring». *The Journal of Risk Model Validation*, Vol. 4(2), pp. 3–25. — DOI: 10.21314/JRMV.2010.057.
9. Satchel S., Xia W. (2008). «Analytic models of the ROC curve: applications to credit rating model validation». In: Christodoulakis G., Satchell S. (Eds.). *The Analytics of Risk Model Validation*. New York: Academic Press.
10. Sobehart J., Keenan S., Stein R. (2000). *Benchmarking Quantitative Default Risk Models: a Validation Methodology*. — <http://www.rogermstein.com/wp-content/uploads/53621.pdf>.
11. Swets J. (1988). «Measuring the accuracy of diagnostic systems». *Science*, Vol. 240, pp. 1285–1293.
12. Tasche D. (2009). *Estimating Discriminatory Power and PD Curves When the Number of Defaults is Small*. — <https://arxiv.org/abs/0905.3928>.
13. Van der Burgt M. (2008). «Calibrating low-default portfolios, using the cumulative accuracy profile». *Journal of Risk Model Validation*, Vol. 1(4), pp. 17–33. — DOI:10.21314/JRMV.2008.016.
14. Zou K.H., O'Malley A.J., Mauri L. (2007). «Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models». *Circulation*, Vol. 115, pp. 654–657. — DOI: 10.1161/circulationaha.105.594929.