



# XVII Russia Risk Conference

Москва  
20 - 21 октября

2021

## Сопоставительный анализ – ключевой аппарат валидации для ПВР

*Михаил Помазанов, к.ф.-м.н  
Руководитель по валидации,  
Дирекция рисков,  
ПАО Промсвязьбанк*



[psbank.ru](https://psbank.ru)

- Рейтинговая модель соответствует критерию высокой прогнозной точности, а факторы, включенные в модель, являются статистически значимыми;
- внутренняя оценка производится на основе статистических данных, а не основана исключительно на экспертном суждении;
- оценки компонентов кредитного риска являются точными и основаны на существенных факторах, оказывающих влияние на изменение параметров, по которым определяются значения компонентов кредитного риска;
- полученные оценки компонентов кредитного риска пересматриваются не реже одного раза в год в сроки, определяемые внутренними документами банка, а также при поступлении существенной информации;
- полученные количественные оценки компонентов кредитного риска соответствуют исторически наблюдаемым значениям, рассчитанным за долгосрочный период, по которому у банка имеются данные, охватывающий как минимум полный цикл деловой активности (период между пиками экономического спада). Для различных сегментов кредитных требований цикл деловой активности может различаться

### 14.2. В рамках проведения внутренней валидации банк должен:

- не реже одного раза в год осуществлять сопоставительный анализ рассчитанных значений вероятности дефолта, полученных в результате применения моделей, используемых в рейтинговой системе, с фактической частотой реализованных дефолтов заемщиков (далее - сопоставительный анализ) для каждого разряда рейтинговой шкалы (портфеля однородных кредитных требований);
- при проведении сопоставительного анализа использовать внутреннюю статистическую информацию за период времени, охватывающий как минимум один полный цикл деловой активности;
- в случае недостаточности внутренней статистической информации использовать внешнюю статистическую информацию;
- отражать во внутренних документах и обновлять как минимум раз в год данные для проведения сопоставительного анализа;
- в случае если фактическая частота реализованных дефолтов заемщиков, отнесенных к одному разряду рейтинговой шкалы, выходит за рамки рассчитанных по модели значений, **анализировать причины таких отклонений и разработать предложения по их устранению**, зафиксировав их в отчете. Порядок сопоставительного анализа и дальнейших действий в таких случаях должен быть отражен во внутренних документах банка. Этот порядок учитывает фазу цикла деловой активности;



### Биномиальный тест

Пусть имеется  $k$  событий дефолта среди  $n$  заемщиков данного класса, наблюдаемая частота дефолта будет  $DF = \frac{k}{n}$ . Фиксируется уровень доверия  $\alpha$  и строится доверительный интервал для модельной оценки величины вероятности дефолта  $PD$

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

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

$N^{-1}$  – обратное нормальное распределение.

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

Отвечает на вопрос, смещена ли оценка среднего  $PD$  относительно факта с учетом статистической погрешности

### Тест Хи-квадрат (Хосмер-Лемешоу)

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

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  $i$ ,  $\theta_i$  – количество заемщиков в дефолте в рейтинговом интервале  $i$ . При  $n_i \rightarrow \infty$  для всех  $i$ ,  $T_k$  стремится к распределению  $\chi_{k-1}^2$ .

$$\Omega_\alpha = [T_k - \Delta_\alpha, T_k + \Delta_\alpha]$$

**Утверждается, что тест Хи-квадрат отвергает гипотезу о состоятельности оценки  $PD$  в рейтинговых разрядах с уровнем достоверности  $\alpha$  если  $k - 1 \notin \Omega_\alpha$ .**

Позволяет проверять нулевую гипотезу о совпадении распределения частот дефолтов с заданным распределением  $PD$  в рейтинговых разрядах

## Недостатки «классики»:

- **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, особенно в условии неполноты статистики дефолтов?**

## Предлагаем

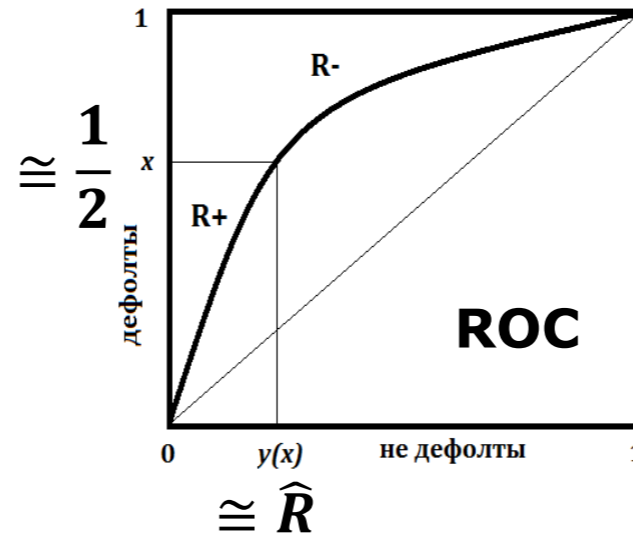
Пусть сопоставительная база кредитного портфеля, содержащая рейтинги (эквивалентно PD) компаний и событий дефолта, содержит  $n$  полных лет. Фиксируются не дефолтные рейтинги (разряды) на начало каждого периода  $RN_i^n, i \in 1 \dots N_n$ , за которыми не следовал дефолт в течение года, а также рейтинги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

### **Определение.**

Медианой дефолтов называется средний рейтинговый разряд  $\hat{R}$ , такой, что количество дефолтов  $RD_d^n < \hat{R}$  равнялось количеству  $RD_d^n > \hat{R}$ .

В случае невозможности 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.

## Разбиение статистики измерений на два множества



Оценка ожидаемой вероятности дефолта в подмножествах согласно мастер-шкале банка

$$\begin{aligned}
 PD^+ &= E_{r \in R^+} [PD(r)] \\
 &> PD &= E_{r \in R} [PD(r)] \\
 &> PD^- &= E_{r \in R^-} [PD(r)]
 \end{aligned}$$

## Первое, что смотрим: Биномиальный тест в подмножествах

Сопоставительный тест не пройден на уровне доверия  $\alpha$ , если нарушено хотя бы одно условие:

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

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

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

а также к тесту на отношение:

$$T2. \frac{PD^+}{PD^-} \in \left[ \frac{P^+}{P^-} \cdot (1 - \Delta_\alpha), \frac{P^+}{P^-} \cdot (1 + \Delta_\alpha) \right]$$

В случае не прохождения T1 выше верхней границы доверительного интервала – рейтинговая система исторически переоценивает риск с уровнем значимости  $\alpha$ , (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.

Уровень доверия рекомендуется общий  $\alpha = 90\%$

## Последовательность тестирования методики

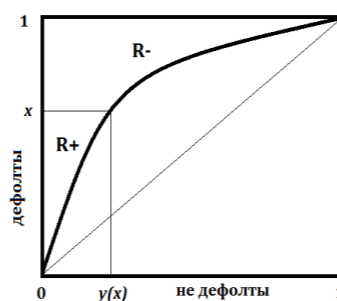
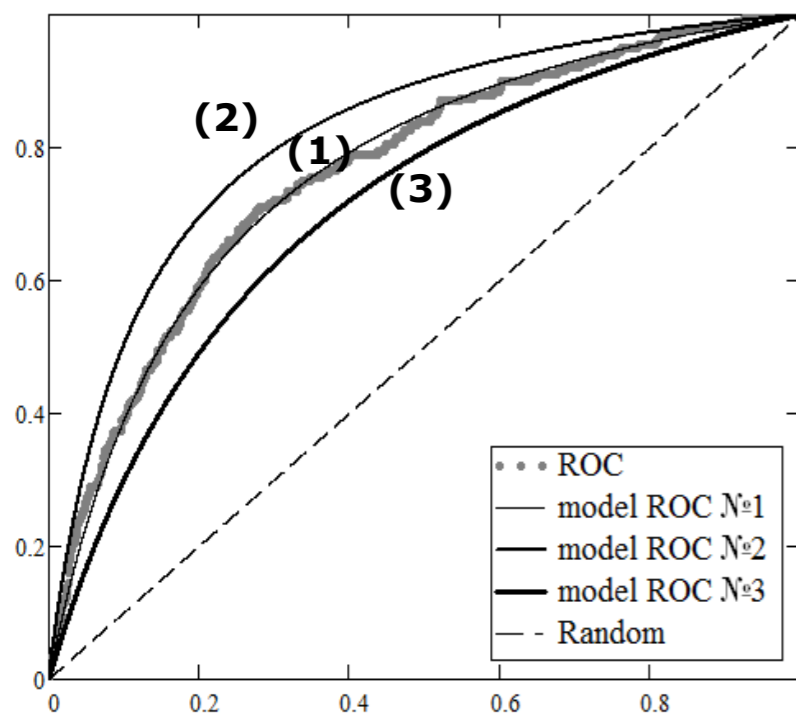
- Рейтинг  $RN_i \in N(0, 1), i \in 1 \dots N$  не дефолтных компаний генерируется стандартным нормальным распределением, а рейтинг дефолтных компаний генерируется настраиваемым нормальным распределением  
 $RD_d \in N(-m, \sigma), d \in 1 \dots D$  с параметрами  $m, \sigma$ .
- После генерации выборок рейтингов  $RN_i, RD_d$  и упорядочивания их по возрастанию, возможно построить ROC-кривую и определяется показатель Джини
- Полагая, число дефолтов четным, медианный рейтинг дефолтов оценивается как  $\hat{R} = RD_{\frac{D}{2}}$
- $PD^-, PD^+, PD$  оцениваются как средние значения калиброванной вероятности дефолта в подмножествах ниже и выше медианы  $\hat{R}$
- Оцениваются наблюдаемые частоты дефолтов  
 $P^- = \frac{D}{2 \cdot N^-}, P^+ = \frac{D}{2 \cdot N^+}, P = \frac{D}{N+D}$   
 имеющие стандартные отклонения  
 $\delta P^- = \sqrt{\frac{P^- \cdot (1 - P^-)}{N^-}}, \delta P^+ = \sqrt{\frac{P^+ \cdot (1 - P^+)}{N^+}}, \delta P = \sqrt{\frac{P \cdot (1 - P)}{N}}$
- Сравниваются с калибровочными  $PD^-, PD^+, PD$  с уровнем доверия  $\alpha = 90\%$  для четырех разных сценариев для проверки сопоставительной гипотезы о несостоятельности калибровки рейтинговой модели.

## Сценарии тестирования

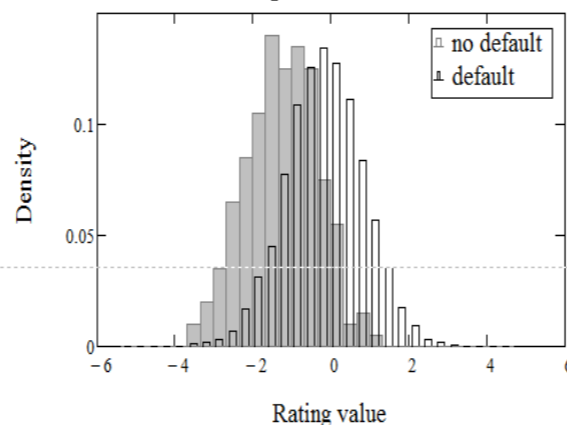
- ✓ **Сценарий № 1:** генерация измерений – случайная ROC-кривая, близкая к симметричной, с сопоставительным количеством дефолтов  $D=200$ , не дефолтов  $N=6000$ , имеющая показатель Джини **AR=53.0%**; вероятность дефолта принимается несмещенной и равна опорной частоте дефолтов  $p = P = \frac{D}{N+D}$
- ✓ **Сценарий № 2:** генерация измерений – сценарий №1; калибровочный показатель Джини увеличен на «четыре сигма» **AR = 53.0% + 4σ<sub>AR</sub>**
- ✓ **Сценарий № 3:** генерация измерений – сценарий №1; калибровочный показатель Джини уменьшен на «четыре сигма» **AR = 53.0% - 4σ<sub>AR</sub>**
- ✓ **Сценарий № 4:** генерация измерений – случайная ROC-кривая, имеющая выраженное «левое» предпочтение имеющая показатель Джини **AR=62.5%**; калибровка стандартная по показателю Джини **AR=62.5%**; вероятность дефолта также остается несмещенной



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



### «Плохие» и «хорошие»

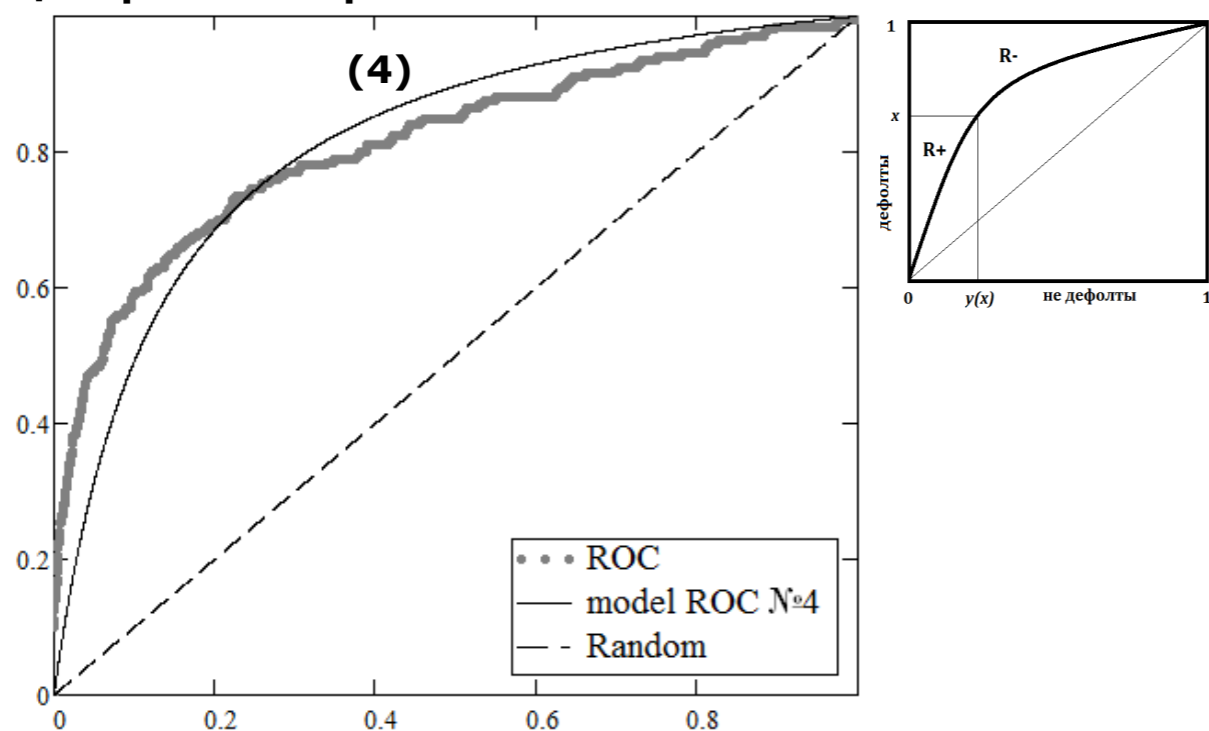


Сценарий №2	Параметры	Значение	Результат теста на уровне доверия $\alpha = 90\%$ . Сопоставительная гипотеза:
1	$PD^-$	1.47%	<b>Отвергается на уровне <math>\alpha</math></b>
	$P^-$	[1.6%, 2.2%]	
2	$PD^+$	12.2%	<b>Отвергается на уровне <math>\alpha</math></b>
	$P^+$	[8.4%, 11.5%]	
T1	$PD$	3.22%	<b>Не отвергается</b>
	$P$	[2.9%, 3.6%]	
T2	$\frac{PD^+}{PD^-}$	8.29	<b>Отвергается на уровне <math>\alpha</math></b>
	$w = \frac{P^+}{P^-}$	[4.08, 6.53]	
<b>Итог...</b>	Тест сопоставительного анализа не пройден		

Сценарий №1	Параметры	Значение	Результат теста на уровне доверия $\alpha = 90\%$ . Сопоставительная гипотеза:
1	$PD^-$	1.88%	<b>Не отвергается</b>
	$P^-$	[1.6%, 2.2%]	
2	$PD^+$	10.08%	<b>Не отвергается</b>
	$P^+$	[8.4%, 11.48%]	
T1	$PD$	3.22%	<b>Не отвергается</b>
	$P$	[2.9%, 3.6%]	
T2	$\frac{PD^+}{PD^-}$	5.35	<b>Не отвергается</b>
	$w = \frac{P^+}{P^-}$	[4.08, 6.53]	
<b>Итог...</b>	Гипотеза состоятельности сопоставительного анализа не отвергается на уровне $\alpha = 90\%$		

Сценарий №3	Параметры	Значение	Результат теста на уровне доверия $\alpha = 90\%$ . Сопоставительная гипотеза:
1	$PD^-$	2.25%	<b>Отвергается на уровне <math>\alpha</math></b>
	$P^-$	[1.6%, 2.2%]	
2	$PD^+$	8.21%	<b>Отвергается на уровне <math>\alpha</math></b>
	$P^+$	[8.4%, 11.5%]	
T1	$PD$	3.2%	<b>Не отвергается</b>
	$P$	[2.9%, 3.6%]	
T2	$\frac{PD^+}{PD^-}$	3.66	<b>Отвергается на уровне <math>\alpha</math></b>
	$w = \frac{P^+}{P^-}$	[4.08, 6.53]	
<b>Итог...</b>	Тест сопоставительного анализа не пройден		

## Генерированная ROC-кривая измерений сценарий калибровки № 4



Сценарий №4	Параметры	Значение	Результат теста на уровне доверия $\alpha = 90\%$ . Сопоставительная гипотеза:
1	$PD^-$	2.0%	Не отвергается
	$P^-$	[1.5%, 2.0%]	
2	$PD^+$	16.6%	Отвергается на уровне $\alpha$
	$P^+$	[18.3%, 24.5%]	
T1	$PD$	3.2%	Не отвергается
	$P$	[2.9%, 3.6%]	
T2	$\frac{PD^+}{PD^-}$	8.0	Отвергается на уровне $\alpha$
	$w = \frac{P^+}{P^-}$	[9.68, 15.50]	
Итог...	Тест сопоставительного анализа не пройден		

### Анализ данных

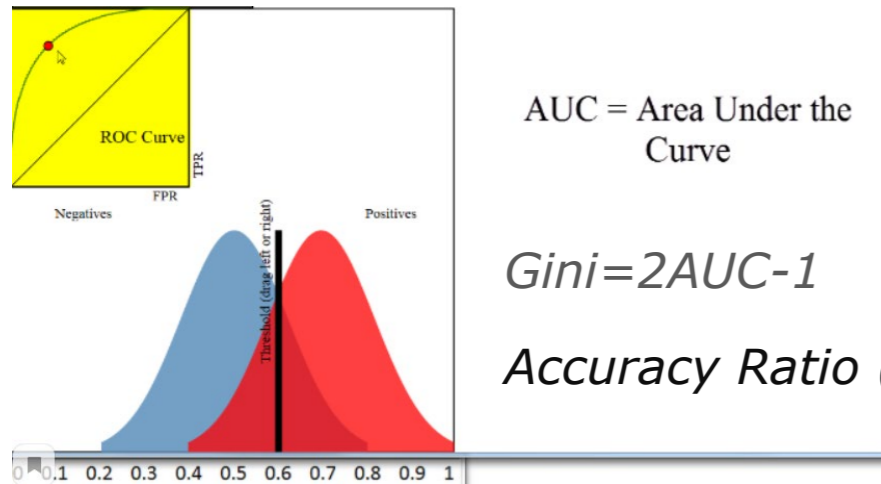
**Сценарий калибровки №1:** показывает, что при условии несмещённой калибровочной вероятности дефолтов и не смещенной мощности рейтинговой модели сопоставительный тест показывает на заданном уровне доверия отсутствие оснований для отвержения сопоставительной гипотезы

**Сценарий калибровки №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»)

**Сценарий калибровки №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»

**Сценарий калибровки №4:** показывает неадекватность формулы калибровки 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.

## Метрика точности дискриминации первого порядка

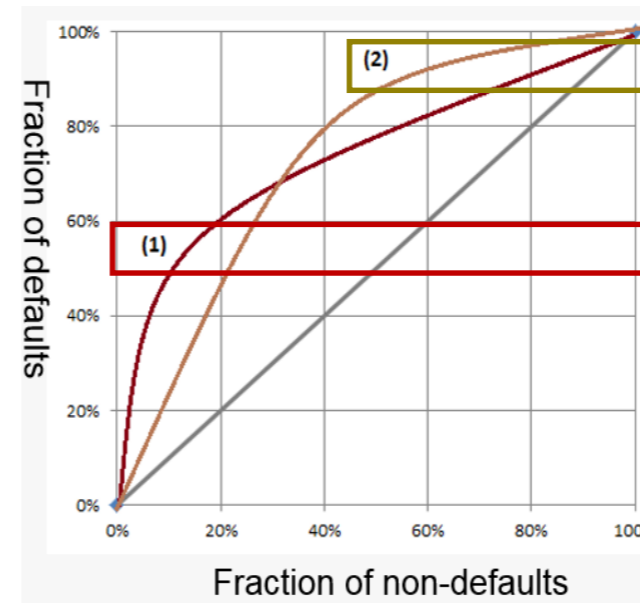


AUC = Area Under the Curve

$Gini = 2AUC - 1$

Accuracy Ratio (AR)

## Метрика точности дискриминации второго порядка



Right Accuracy Ratio (RAR)

$RAR > LAR$

Left Accuracy Ratio (LAR)

$LAR > RAR$

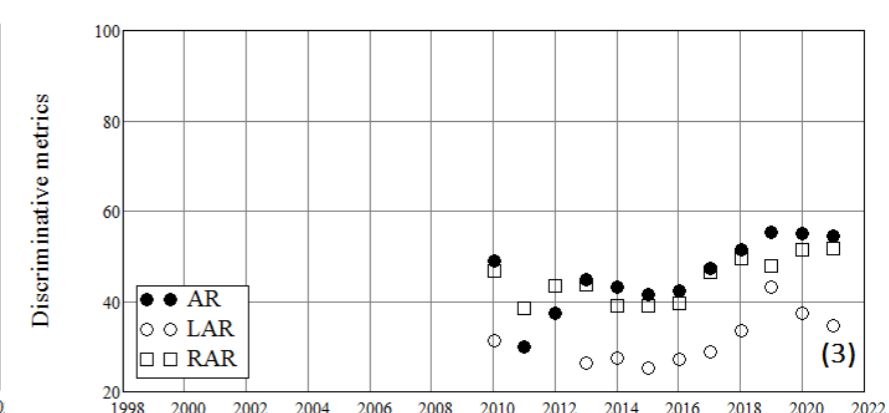
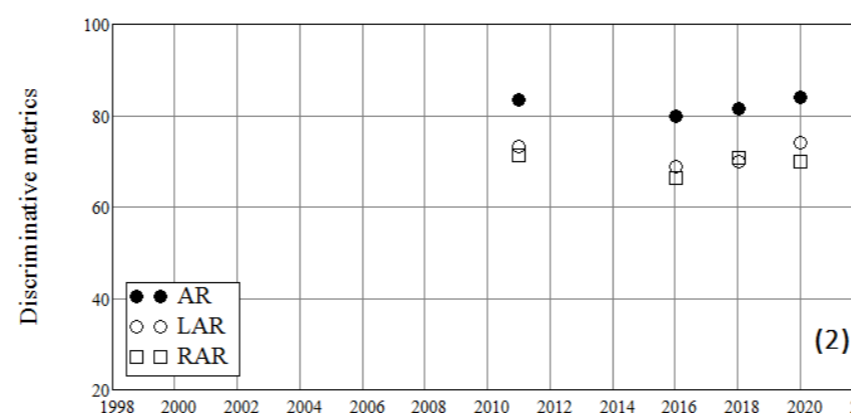
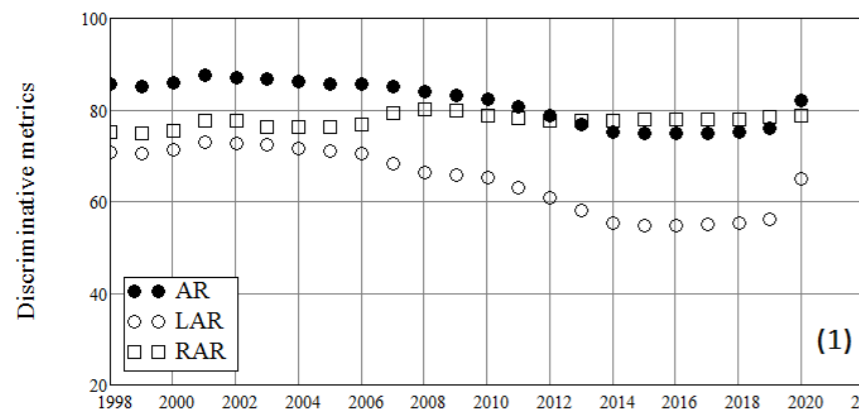
**Несимметричные ROC**

## Три рейтинговых агентства

- (1) – Moody's
- (2) – Fitch
- (3) – Expert RA

## Данные

- Вероятности дефолта в рейтинговых разрядах
- Данные о количестве рейтингуемых компаний в каждом разряде



## Анализ данных

Moody's не уступает Fitch по вмененному индексу Джини, однако с точки зрения целевых предпочтений рейтинг Moody's «видит» хорошие корпоративные компании лучше и отличает «плохие» хуже, чем рейтинг Fitch, т. е. имеет правые предпочтения. Кривая ROC Fitch близка к симметричной. Дискриминационная сила суверенного «Эксперт РА» значительно ниже, чем у представителей «большой тройки», но по оценкам метрик второго порядка рейтинг «Эксперт РА» также лучше «видит» хорошие, имея преимущественно правые предпочтения. Различие целевых предпочтений рейтинговых механизмов у разных рейтинговых агентств усложняет задачу сравнения их рейтинговых шкал.



**Table 1. Generated Models Parameters**

Preference	AR (%)	LAR (%)	RAR (%)	$\sigma_{\max}(\%)$
Neutral	49.7	37.1	34.4	4.6
Left	41.2	51.1	10.2	4.7
Right	48.9	13.2	53.2	4.6

AR, RAR, RAR – метрики точности модели первого и второго порядка, рассчитанные по бинарным данным

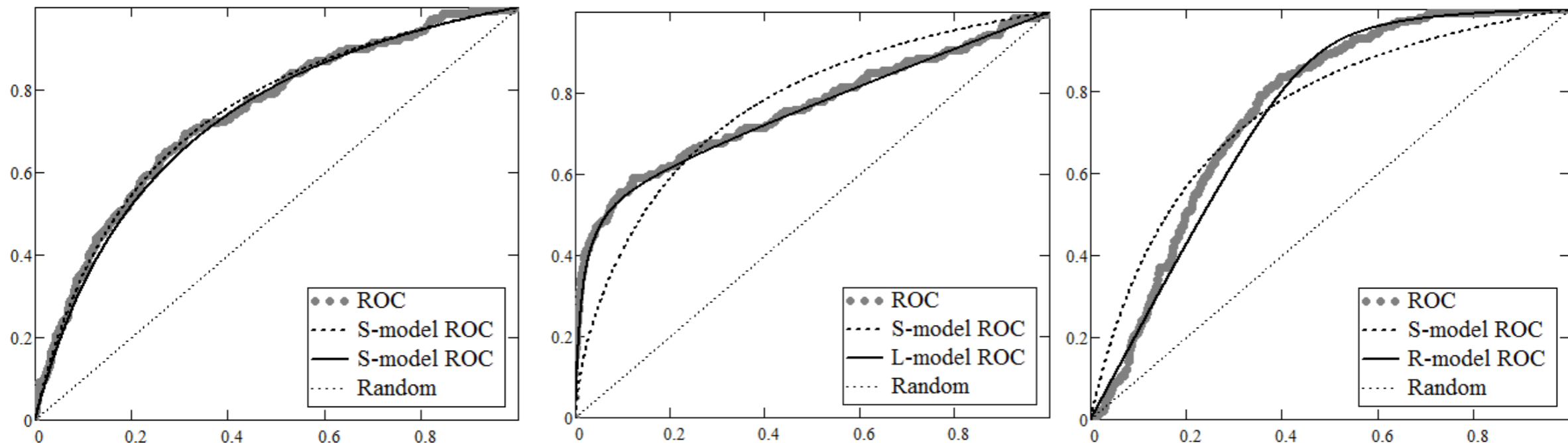
$AR_{imp}$ ,  $LAR_{imp}$ ,  $RAR_{imp}$  – метрики точности модели первого и второго порядка, рассчитанные по данным вероятности дефолта (калибровка)

**Table 2. Results (%)**

Preference	Calibration method	AR	LAR	RAR	distL	distR	Decision
Neutral	Symmetr.	0.2	4.7	2.3	-16	-7.4	Apply
Left	Symmetr.	0.2	25.5	15.1	-79	99	Adjust
Right	Symmetr.	0.22	18.6	21.8	68	-75	Adjust
Left	Left-prefer.	0.50	0.56	1.7	4	4.4	Apply
Right	Right-prefer.	0.32	4.3	0.17	10	8.5	Apply

distL, distR – процент искажения вероятности дефолта в левом (худшие) и правом (лучшие) сегменте заемщиков

## Результат использования универсальной калибровки, учитывающей предпочтение рейтинговой модели



# Спасибо!

**Помазанов Михаил  
Вячеславович**

[romazanovmv@psbank.ru](mailto:romazanovmv@psbank.ru)



[psbank.ru](http://psbank.ru)