Факультет экономических наук | Школа финансов | Москва, 2024

Разработка системы моделей оценки вероятности дефолта для корпоративных заемщиков в отрасли «Металлургия» с целью резервирования и ценообразования

Автор: Часовникова Мария Александровна

Научный руководитель: Моргунов Алексей Владимирович, к.э.н.

Почему модель оценки вероятности дефолта?

- С 6.08.2015 действует **Положение Банка России 483-П** «О порядке расчета величины кредитного риска на основе внутренних рейтингов»
- PD (Probability of Default) – одна из **компонент** оценки кредитного риска в ПВР-подходе
- Более **точная** и **диверсифицированная** оценка рисков по клиентам и более **эффективное** использование капитала банка
- С 2025 года ЦБ планирует обязать **все СЗКО** перейти на ПВР-подход*

Почему металлургия?

- **Более половины ВВП РФ** от строительства, промышленности и энергетики, полностью зависящих от металлургии
- **Вызовы извне:** ограничение доступа к рынкам сбыта, технологиям, разрушение торговых цепочек
- **Особый фокус** на отрасль в рамках Стратегии развития металлургической промышленности РФ до 2030 года
- Развитие **внутреннего рынка** за счет средних металлургических предприятий

**Источник: Frank Media*

Цель исследования заключается в разработке системы моделей оценки вероятности дефолта малых и средних российских корпоративных заемщиков металлургической отрасли.

Систематизировать подходы к построению моделей PD для различных сегментов

Собрать данные финансовой и бухгалтерской отчетностей металлургических компаний с 2013 по 2022 год включительно

Разработать логит-модели на основе отобранных риск-факторов

Оценить дискриминационную и прогнозную силу моделей и привести их к существующим стандартам

Откалибровать модели с учетом и без учета макроэкономической конъюнктуры

Построить рейтинговую мастер-шкалу на основе полученных оценок вероятности дефолта

Провести процедуру валидации моделей на отчетностях компаний за 2022 год

Моделирование различных риск-сегментов

Авторы и год	Объект исследования	Итоговые риск-факторы
Жевага и др. (2016)	Российские промышленные компании	Коэффициент автономии, Объем выручки, Заемные средства/EBITDA, EBITDA/Выручка, EBITDA/Проценты к уплате
Тотьмянина, (2014)	Российские строительные компании	Рентабельность затрат, Доля запасов в оборотных активах, Коэффициент автономии
Zheng et. al., (2018)	Американские строительные компании	Доля рабочего капитала в активах, Коэффициент автономии, ROA, Оборачиваемость кредиторской задолженности
Altman & Sabato, (2007)	Малые и средние предприятия	EBITDA/Активы, ROA, Доля денежных средств в активах

Почему используется логит в данном исследовании?

- Количество дефолтов в выборке незначительно
- Данные однородны с экономической точки зрения

В таких случаях модель логистической регрессии обеспечивает **минимальный уровень переобучения**

Методология. Логит-модель

$$PD = \frac{1}{1+\exp\left\{-\left(\vec{a}\times\vec{x}^{T}+b\right)\right\}},$$

$$LogL = \sum_{k=1}^{N}\left[Y_i \cdot \ln\left(PD_i\right)+\left(1-Y_i\right)\cdot\ln\left(1-PD_i\right)\right] \rightarrow min$$

где:

$\vec{x}^{T}$ – вектор-столбец нормированных значений риск-факторов;

$\vec{a}$ – вектор-строка регрессионных коэффициентов при нормированных риск-факторах;

b – свободный член регрессии.

Y_i – наблюдаемая бинарная переменная, принадлежащая множеству {0;1}, фиксирующая факт дефолта;

PD_i – вероятность дефолта заемщика i, полученная с помощью логистической регрессии.

Гипотезы

H1. Макроэкономические показатели оказывают **значительное влияние** на оценку риска дефолта металлургических компаний в разные моменты экономического цикла *(на основе (Карминский, 2015; Тотьмянина, 2014; Моргунов, Жевага, 2016).*

H2. Компании в различных подотраслях металлургии, в зависимости от стадии металлургического процесса, имеют различные специфические структуры баланса. Поэтому **подотрасли будут обладать разной подверженностью к дефолту.**

H3. Риск-факторы обладают определенной **логикой** влияния на вероятность дефолта, соответствующей **экономическому смыслу и практическим результатам.**

Сбор и обработка данных

Spark Interfax

Ежегодные данные с 2013 по 2022

Панельная структура
(1888 компаний всего, 530 дефолтов, 17322 строк)

54 Балансовых и финансовых показателя

Флаг дефолта ставится на год n, если дата дефолта принадлежит интервалу: **[01.09.n; 01.09.n+1)**

- Логарифмизация
- Логистическая трансформация
- Нормализация

2013-2021 гг. – обучение

2022 год - валидация

Как собрать данные из СПАРК Интерфакс?

Выгрузить данные можно только с университетского компьютера

Запросить логин и пароль через Яндекс-форму. Консультирование: jesda@hse.ru)

Ознакомиться с руководством пользователя

В разделе «Выборки» выбрать интересующий субъект (банки, компании, ИП и т.д.)

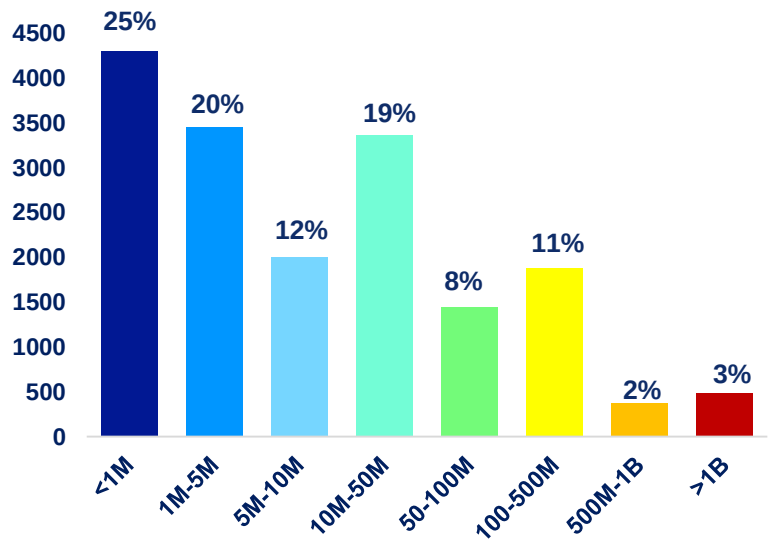
Настроить фильтры: отрасль по ОКВЭД, статус, год отчетности

Выбрать интересующие бухгалтерские и финансовые показатели

! Одна выгрузка ограничена 2500 компаниями, скачиваем их отчетности отдельными файлами за каждый год, меняя фильтр

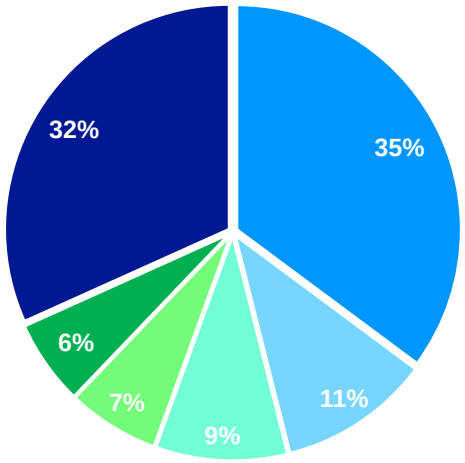
! Даты банкротства собираем вручную из столбца «Важная информация»

Распределение наблюдений по величине активов, руб.

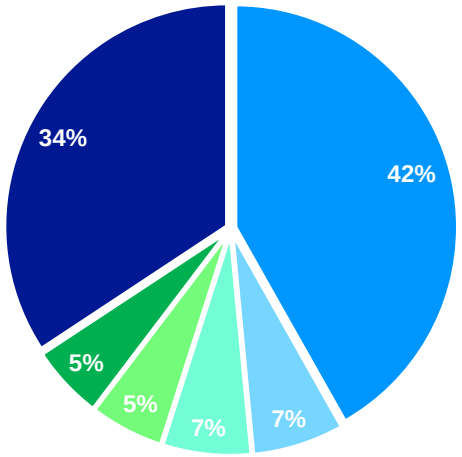


В выборке преобладают металлургические компании **малого и среднего** размеров. Самая крупная подотрасль – **строительные конструкции**. Все подотрасли обладают схожим уровнем дефолтности.

Распределение отраслей по всем наблюдениям, %



Распределение отраслей в дефолтах, %



- **Строительные конструкции**
- **Обработка металлических изделий**
- **Производство дверей и окон**
- **Добыча платиновых руд**
- **Обработка чистых металлов**
- **Остальные отрасли**

Индустриальный фактор (H2)

	Металлодобыча	Готовые изделия
Добыча	• Производство **чугуна, стали и ферросплавов** • **Добыча и обогащение** железных руд • Производство **свинца, цинка и олова** • **Добыча руд** цветных металлов	• Производство строительных металлических **конструкций** • Производство металлических **дверей и окон** • Производство металлических **цистерн** • Производство **изделий** из проволоки, цепей и пружин • Производство **замков и петель**
Первичная обработка	• **Литье** чугуна • Производство **стали в слитках** • **Литье** стали	• Производство **изделий** из проволоки • и пружин • Производство ножевых **изделий** • и столовых **приборов** • Производство **радиаторов**

Трансформация **Weight of Evidence**

$$WOE_i = ln\left(\frac{\left(N_{good_i}/N_{good_{all}}\right)}{\left(N_{bad_i}/N_{bad_{all}}\right)}\right)$$

↑WOE ↓PD

Группа	WOE
Металлодобыча	**1.23**
Изделия	**-0.81**

14 риск-факторов, отобранных на основании максимального Джини в однофакторном анализе (max 0.51, min 0.36):

Риск-фактор	Группа	Интерпретация связи с вероятностью дефолта (Н3)
Коэффициент концентрации собственного капитала (автономии), %	Леверидж	**Отрицательная.** Чем **меньше** собственных средств у компании, тем более она зависима от кредиторов и тем **больше** вероятность дефолта.
Рентабельность активов (ROA), %	Эффективность	**Отрицательная.** Чем **больше** сумма чистой прибыли, которой компании принес рубль стоимости активов, тем эффективнее используются активы и тем **меньше** вероятность дефолта.
Логарифм активов, RUB	Масштаб	**Отрицательная.** Чем **больше** абсолютное значение активов компании, тем больше ее финансовые возможности и выше вероятность поддержки со стороны банков и государства и **меньше** вероятность дефолта.
Доля рабочего капитала в активах компании, %	Ликвидность	**Отрицательная**. Чем **меньше** у компании собственных средств для финансирования текущей деятельности, тем **больше** вероятность дефолта.
Оборачиваемость кредиторской задолженности, разы	Оборачиваемость	**Отрицательная.** Чем **больше** раз компания может рассчитаться со своей кредиторской задолженностью в год, тем **меньше** вероятность дефолта.

Выбор оптимальной модели до калибровки:

- **Качество:** Максимальный коэффициент Gini
- **Значимость:** p-value меньше 0.05
- **Устранение мультиколлинеарности:** VIF меньше 2
- **Интерпретируемость:** логичные знаки коэффициентов (должны быть **отрицательными** из-за логит- и WOE-трансформации)

	coef.	st. error	z	P>|z|
const	-4.2264	0.077	-55.099	0.000
Industry	-0.1977	0.056	-3.512	0.000
ROA	-0.4886	0.065	-7.538	0.000
Autonomy	-0.4987	0.075	-6.608	0.000
NWC/Assets	-0.1859	0.067	-2.791	0.005

Gini	0.5686	**Max VIF**	1.77
AUROC	0.7843	**Max p-value**	0.005

Процедура калибровки Through-the-Cycle

- **Through-the-Cycle** калибровка используется для выделения **идиосинкратического** риска заемщика, например, для задачи **определения RWA** (Risk-weighted Assets) для расчета Достаточности собственного капитала
- Подбор **свободного коэффициента** модели, чтобы при нем средняя полученных оценок равнялась центральной тенденции по портфелю

$$Avg(PD) = \frac{\sum_{i=1}^{n} DR_i}{n} = \frac{\sum_{i=1}^{n} \frac{N_{def_i}}{N_{def_i} + N_{nondef_i}}}{n} = CT,$$

где DR_i – частота дефолта в год i, n – количество лет, по которым присутствуют отчетности заемщиков в портфеле, N_{def_i} – количество дефолтных наблюдений в год i, N_{nondef_i} – количество недефолтных наблюдений в год i, CT – центральная тенденция по портфелю.

Результат калибровки:

$$Average(DR_i) = 2\% = Average(PD)$$

$$a_{0_old} = -4.2 \rightarrow a_{0_new} = -4.4$$

Снижение базовой вероятности дефолта после очищения от временных колебаний

Процедура калибровки Point-in-Time

- **Point-in-Time** калибровка используется для выделения **текущего состояния экономики**, например, в задачах **ценообразования и резервирования**
- Собирается датасет потенциальных макрофакторов с 2013 по 2022 гг., а оценки PD, полученные TTC-моделью, преобразовываются в логит риск-фактор:

$$logit(PD_{TTC_i}) = \ln\left(\frac{PD_{TTC_i}}{1 - PD_{TTC_i}}\right)$$

Риск-фактор	Gini	Экономический смысл (НЗ)	Соответствие модельной логике
Средняя годовая ставка ЦБ РФ, %	0.21	$\uparrow r \ \uparrow PD$	Нет
Годовой прирост инфляции, %	0.06	$\uparrow \pi \uparrow PD$	Нет
Годовой прирост реального ВВП в ценах 2016 года, %	0.01	$\uparrow Y \downarrow PD$	Нет
Отток населения из РФ, тыс. чел.	**0.35**	$\uparrow \text{отток } n \uparrow PD$	Да
Индекс физического объема экспорта, %	0.09	$\uparrow Ex \ \downarrow PD$	Нет
Средний курс доллара к рублю, руб.	0.25	$\uparrow USDRUB \ \downarrow PD$	Нет

Выбор оптимальной модели после калибровки:

- **Качество:** Максимальный коэффициент Gini
- **Значимость:** p-value меньше 0.05
- **Устранение мультиколлинеарности:** VIF меньше 2
- **Интерпретируемость:** логичные знаки коэффициентов

	coef.	st. error	z	P>|z|
const	-7.1349	0.052	18.437	0.000
Logit(PD TTC)	0.9604	0.001	10.196	0.000
Population outflow	0.0074	0.316	-22.613	0.000

Gini	0.6666	**Max VIF**	1.00
AUROC	0.8333	**Max p-value**	0.000

Интерпретация связи с фактором Отток населения

- В 2022 году Россию покинули **100** тысяч IT-специалистов*
- Примеры задач, решаемых IT-специалистами в **металлургических компаниях**: оптимизация резервов, сопровождение и оптимизация работы оборудования, генерация рецептур, сопровождение цифровых систем (MES, ERP и т.д.)
- В основном, в 2022 г. в выборке дефолтнули **малые предприятия** (у 81% дефолтнувших компаний величина активов ниже 150 млн руб.)
- **Размер оклада** в небольшой компании создает **меньше стимулов** у работника остаться в компании по сравнению с крупной корпорацией
- Уход IT-специалиста из небольшой компании может вызвать **серьезный операционный риск** ввиду низкой численности IT-команды

**Источник: Интерфакс*

Оценка качества моделей

- Валидация осуществлялась на отложенной отчетности за **2022** год для заемщиков из обучаемой выборки
- В 2022 год шок. Средний DR 2022 года – **8%** против среднего DR 2013-2021, равного **2%**
- Рекордный отток населения из РФ – более **670** тыс. человек

AUROC < 0.5, Gini < 0.4, ΔGini > 0.2*

AUROC < 0.7, Gini < 0.4, ΔGini > 0.2

AUROC > 0.7, Gini > 0.4, ΔGini < 0.2

Model	Gini	ΔGini (train vs test)	AUROC	Result
PD TTC train	0.57	0.06	0.78	Валидация пройдена по всем критериям
PD TTC test	0.51		0.75	
PD PIT train	0.67	0.16	0.83	Валидация пройдена по всем критериям
PD PIT test	0.51		0.75	

**Источник: Пеникас Г. (2022), Валидация Комитет по стандартам Базель II и управлению рисками (2013)*

Валидация – разработка рейтинговой шкалы

- Для полученных оценок на обучаемой выборке была построена **рейтинговая шкала** и проведены **биномиальный** тест и тест **Хосмера-Лемешева** на соответствие распределений средних DR и дискретных PD на каждом интервале рейтинга
- Ввиду сильного ухудшения DR на валидации шкала оказалась неприменима к валидационной выборке и была **оптимизирована**
- Корректировка шкалы оправдана **шоком после февраля 2022 г.**

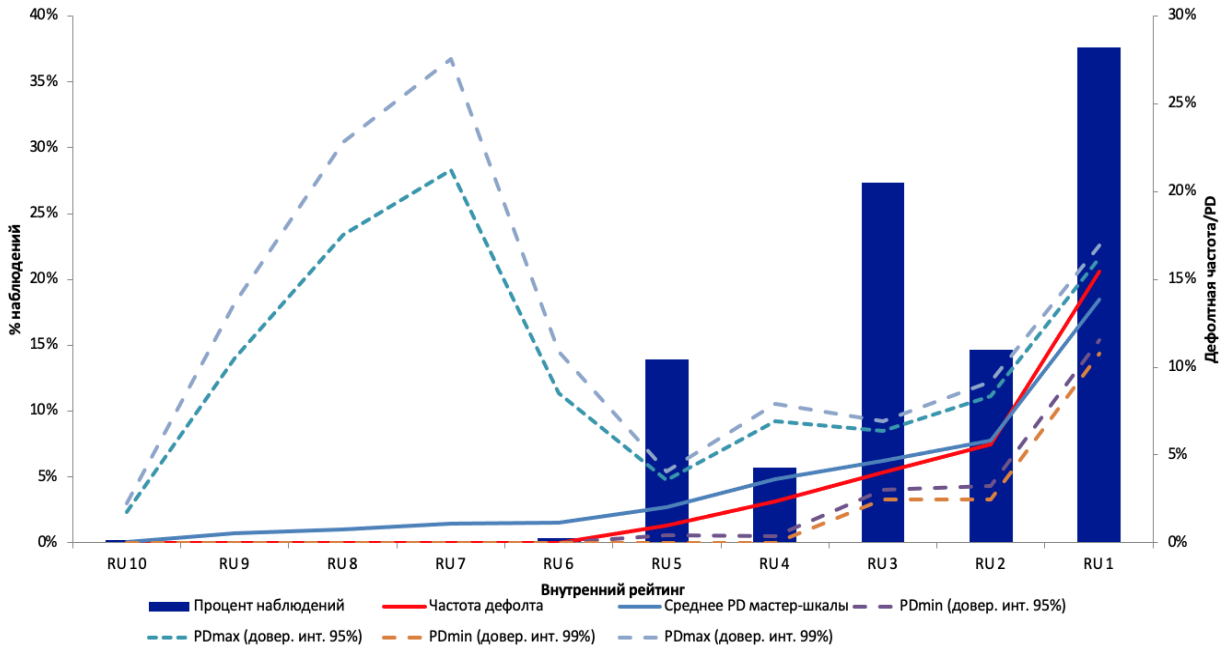
Дискр. PD	Нижняя граница	Верхняя граница	Рейтинг
0.03%	0.00%	0.275%	RU 10
0.52%	0.275%	0.63%	RU 9
0.74%	0.63%	0.905%	RU 8
1.07%	0.905%	1.11%	RU 7
1.15%	1.11%	1.575%	RU 6
2,00%	1.575%	3.5%	RU 5
3.63%	3.5%	3.85%	RU 4
4.68%	3.85%	5.52%	RU 3
5.81%	5.52%	6.06%	RU 2
13.87%	6.06%	100.00%	RU 1

Результаты

H1. Макропоказатель **«Отток населения из РФ»** при росте на 1 тыс. чел. в среднем при прочих равных на 0.0074 снижает Z-дискриминант логит-модели, что приводит к **увеличению вероятности дефолта.**

H2. Принадлежность компании к группе компаний, относящимся к **готовым изделиям**, в среднем при прочих равных приводит к **увеличению вероятности дефолта.**

H3. Экономическому смыслу **соответствует** модельная логика факторов: Рентабельность активов (ROA), %; Коэффициент концентрации собственного капитала (автономии), %; Доля рабочего капитала в активах компании, %; Индустриальный фактор, Отток населения из РФ.



Большая часть оценок, построенных на отчетностях 2022 года, сконцентрирована в наихудших рейтинговых интервалах, из-за значительного ухудшения отчетностей малых компаний и макрофактора

Ограничения

- Используется лишь одна эконометрическая модель
- Разработанные модели формируют годовые оценки вероятности дефолта
- Обучение на определенном макротренде, после кризисного оттока населения 2022 года влияние фактора **может ослабеть** ввиду адаптации предприятий к отечественному оборудованию и переквалификации кадров

Решение

- Сравнительный анализ качества моделей различных типов: например, случайного леса и нейронных сетей
- Разработка модели, формирующей оценки вероятности дефолта в течение срока жизни финансового актива (PD Lifetime)
- Повторное проведение процедуры валидации на данных последующих лет, включение дополнительных риск-факторов, характеризующих близость к дефолту заемщиков, которые позволят повысить точность ранжирования

HSE UNIVERSITY • EST 1992 •

Логистическая трансформация

Чем **ближе** трансформированное значение к 1, тем **меньше** вероятность дефолта. Таким образом, регрессионные коэффициенты при трансформированных риск-факторах должны иметь **отрицательный знак.**

Логистическая трансформация в случае, когда с **ростом** фактора **снижается** уровень кредитного риска заемщиков с экономической точки зрения:

$$Factor_{tr} = \frac{1}{1 + exp\{-Slope \times (Factor - Midpoint)\}},$$

Логистическая трансформация в случае, когда с **ростом** фактора **повышается** уровень кредитного риска заемщиков с экономической точки зрения:

$$Factor_{tr} = 1 - \frac{1}{1 + exp\{-Slope \times (Factor - Midpoint)\}},$$

Коэффициент трансформации (Slope) находится согласно условию:

$$\frac{1}{1 + exp\{-Slope \times (Factor_{95\%} - Midpoint)\}} = 0.95$$

$*\ Midpoint - (Factor_{5\%} + Factor_{95\%})/2$ *риск-фактора, где* $Factor_{5\%}$ *и* $Factor_{95\%}$ *– 5% и 95% перцентили риск-фактора соответственно.*

Все риск-факторы, отобранные на основании максимального Джини в однофакторном анализе (max 0.51, min 0.36):

Леверидж:

- Коэффициент концентрации собственного капитала (автономии), %;
- Коэффициент концентрации заемного капитала,%;
- Коэффициент соотношения заемных и собственных средств, %;
- Коэффициент обеспеченности собственными оборотными средствами, %

Эффективность:

- Рентабельность активов (ROA), %;
- Рентабельность продаж, %;
- Рентабельность затрат, %;
- Чистая норма прибыли, %;
- Соотношение валовой прибыли к активам компании, %

Масштаб: Логарифм активов, RUB

Ликвидность: Доля рабочего капитала в активах компании, %

Оборачиваемость:

- Оборачиваемость кредиторской задолженности, разы;
- Оборачиваемость дебиторской задолженности, разы;
- Коэффициент оборачиваемости совокупных активов, %

Результаты теста Хосмера-Лемешева

$$T_k = \sum_{i=0}^{k} \frac{(PD_i - DR_i)^2}{\left(\frac{1}{n_i}\right) PD_i(1 - PD_i)},$$

Тест Хосмера-Лемешева был применен к разработанной мастер-шкале **на обучении и на валидации.** Количество степеней свободы для выполнения теста равно 10, так как наблюдения распределены в **10** рейтинговых разрядах.

По результатам прохождения теста на уровне значимости 10% принимается гипотеза о том, что дефолтные частоты по рейтингам распределены по вероятностному распределению, получаемому на основании модели. То есть Хи-квадрат тест выполнен на доверительном уровне **91%**, **валидация пройдена.**

Рейтинг	Количество наблюдений	Вероятность дефолта (PD мастер-шкалы)	Частота дефолта (DR)	Расчет статистики хи-квадрат
RU 10	4	0,03%	0,00%	0,00
RU 9	2	0,52%	0,00%	0,01
RU 8	1	0,74%	0,00%	0,01
RU 7	1	1,07%	0,00%	0,01
RU 6	8	1,15%	0,00%	0,09
RU 5	306	2,00%	0,98%	1,62
RU 4	126	3,63%	2,38%	0,56
RU 3	603	4,68%	3,98%	0,66
RU 2	322	5,81%	5,59%	0,03
RU 1	829	13,87%	15,44%	1,71
Общая статистика Хи-квадрат				**4,71**
Доверительный уровень для теста Хи-квадрат				**91%**

Результаты биномиального теста

$$int_PD_i^{min,\alpha} = \max\left\{PD_i - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{PD_i(1-PD_i)}{n_i}}; 0\right\}$$

$$int_PD_i^{max,\alpha} = min\left\{PD_i + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{PD_i(1-PD_i)}{n_i}}; 1\right\}$$

Биномиальный тест был применен к разработанной мастер-шкале **на обучении и валидации.** По результатам прохождения теста на уровне значимости 1% дефолтные частоты по всем рейтингам попадают в рассчитанные доверительные интервалы. То есть биномиальный тест выполнен на доверительном уровне **99%, валидация пройдена.**

Рейтинг	Количество наблюдений	Вероятность дефолта (PD мастер-шкалы)	Частота дефолта (DR)	$PD_i^{min,0,01}$	$PD_i^{max,0,01}$
RU 10	4	0,03%	0,00%	0,00%	2,26%
RU 9	2	0,52%	0,00%	0,00%	13,62%
RU 8	1	0,74%	0,00%	0,00%	22,82%
RU 7	1	1,07%	0,00%	0,00%	27,57%
RU 6	8	1,15%	0,00%	0,00%	10,86%
RU 5	306	2,00%	0,98%	0,00%	4,06%
RU 4	126	3,63%	2,38%	0,00%	7,92%
RU 3	603	4,68%	3,98%	2,46%	6,90%
RU 2	322	5,81%	5,59%	2,45%	9,17%
RU 1	829	13,87%	15,44%	10,78%	16,96%