

Финансовая экономика

МОДЕЛИРОВАНИЕ УРОВНЯ ПОТЕРЬ ПО РОЗНИЧНОМУ ПОРТФЕЛЮ НЕОБЕСПЕЧЕННЫХ КРЕДИТОВ БАНКОВ

Фролова Дарья Дмитриевна
студентка магистратуры,
МГУ имени М.В. Ломоносова, экономический факультет
(г. Москва, Россия)

Аннотация

Грамотное моделирование потерь по розничному кредитному портфелю в случае дефолта (Loss given default – LGD) является важной задачей коммерческих банков. Актуальность исследования обусловлена активным расширением кредитования физических лиц за последние 5 лет. Увеличение объема розничного портфеля с 17,6 до 36,9 трлн р. с 1 января 2020 г. по 1 января 2025 г. способствовало росту кредитного риска российских банков. Дополнительные сложности возникают в связи с высоким уровнем ключевой ставки Банка России и ужесточением макропруденциальной политики в сегменте необеспеченных кредитов физических лиц.

В работе описаны основные подходы к моделированию LGD. На основе данных по необеспеченным потребительским кредитам одного из российских банков за период с декабря 2020 г. по март 2025 г., которые содержат более 700 тыс. наблюдений, предложена двухэтапная методика оценки LGD. Подход включает в себя винтажный анализ и применение моделей машинного обучения градиентного бустинга (XGBoost и LightGBM) для оценки и прогнозирования потерь при дефолте с учетом новых макроэкономических факторов: курса доллара к рублю на рынке Форекс и среднемесячной заработной платы работников в РФ.

Ключевые слова: потери при дефолте по кредитным обязательствам, розничный портфель, коммерческие банки, необеспеченные кредиты физических лиц, моделирование, винтажный анализ, машинное обучение.

JEL коды: C53, G17, G21.

Для цитирования: Фролова Д.Д. Моделирование уровня потерь по розничному портфелю необеспеченных кредитов банков // Научные исследования экономического факультета. Электронный журнал. 2025. Том 17. Выпуск 4. С. 136-150. DOI: 10.38050/2078-3809-2025-17-4-136-150.

Введение

Моделирование уровня потерь при дефолте (Loss Given Default, далее – LGD) представляет собой ключевую задачу коммерческих банков в условиях макроэкономической нестабильности и регуляторных изменений. Точность оценки LGD существенно влияет на резервирование под возможные убытки и выбор стратегий взыскания задолженности, определяя финансовую устойчивость и прибыльность банка. Особое значение данный процесс приобретает для розничного кредитного портфеля, характеризующегося устойчивым ростом за последние 5 лет (см. рис. 1), высоким кредитным риском и ужесточением требований Банка России, включая повышение ставок по необеспеченным кредитам и высокую ключевую ставку.



Рисунок 1. Динамика объема розничных кредитов и просроченной задолженности по ним за период с 1 января 2020 г. по 1 января 2025 г., млрд р. (составлено автором на основе (Статистические показатели банковского..., 2025))

В научной литературе моделирование LGD активно изучалось. Авторы применяли линейные (метод наименьших квадратов, бета-регрессия) и нелинейные методы (нейронные сети, деревья регрессии), а также винтажный анализ. Однако выбор оптимального метода остается дискуссионным и зависит от характеристик кредитного портфеля, экономического цикла, бизнес-целей банка, политики взыскания просроченной задолженности и распределения LGD. При этом банки, использующие подход на основе внутренних рейтингов, обязаны обеспечивать прозрачность и высокую дискриминационную силу моделей в соответствии с требованиями регулятора.

Для построения моделей LGD применяются внутренние данные банков (социально-демографические характеристики заемщиков, заявочные и поведенческие факторы) и внешние источники (макроэкономические индикаторы, данные бюро кредитных историй, уровень дефолтности в экономике и др.). Выбор предикторов определяется спецификой портфеля, доступностью данных, экономической ситуацией и корреляцией факторов, при этом предпочтение отдается стабильным и независимым от внутренних процессов банка переменным.

Цель работы – предложить модель оценки и прогноза уровня потерь при дефолте необеспеченных потребительских кредитов. В работе описана двухэтапная методика оценки и прогнозирования LGD с использованием социально-демографических, поведенческих и макроэкономических предикторов для данных по розничным краткосрочным необеспеченным кредитам одного из российских банков: расчет показателя LGD и его прогнозирование для

месяцев дефолта с помощью таких моделей машинного обучения «градиентного бустинга», как XGBoost и LightGBM. Данные модели характеризуются высокой точностью оценок, что позволяет рекомендовать их для применения в банковской практике.

1. Обзор литературы

Корректная оценка уровня потерь при дефолте должна соответствовать требованиям Базельского комитета и Банка России. Согласно Положению № 590-П¹, LGD отражает долю невозвращенных средств при дефолте заемщика, дополняя такие параметры кредитного риска, как вероятность дефолта (PD) и сумма под риском (EAD). Базель III и Положение № 845-П² ЦБ подчеркивают необходимость валидации внутренних моделей LGD, учета экономических убытков, волатильности в кризисы и данных минимум за 5 лет.

Под дефолтом понимается наличие просрочки платежей по основному долгу и начисленных процентов более чем на 90 дней, а также уверенность кредитной организации в неспособности заемщика погасить свои обязательства³. При этом параметр LGD является обратным ставке восстановления (Recovery rate, далее – RR). Последняя отражает уровень возврата средств в случае дефолта по кредитному обязательству ($LGD = 1 - RR$).

В целом можно выделить следующие шесть этапов моделирования потерь при дефолте (Rantanen, 2023):

- 1) сбор данных из хранилища;
- 2) анализ данных и выбор временного окна моделирования;
- 3) определение возможных предикторов LGD по статистическим и бизнес-критериям;
- 4) выбор метода моделирования и факторов;
- 5) тестирование и проверка модели;
- 6) реализация и мониторинг выбранной модели.

Что касается обзора литературы, то исследователи предлагают разнообразные подходы к моделированию LGD и его предикторам. Так, многие авторы подчеркивают значимость макроэкономических показателей, таких как уровень безработицы и динамика ВВП, для оценки LGD. Т.В. Ишмуратова предлагает модели, интегрирующие индивидуальные характеристики заемщиков и внешние экономические индикаторы, что повышает точность прогнозов (Ишмуратова, 2018). Аналогичный подход развивают другие авторы (Bellotti и Crook, 2012), применяя логистическую регрессию и методы машинного обучения для учета временных рядов макроэкономических данных. Их исследование показывает, что включение таких переменных, как уровень банковских процентных ставок и безработицы, улучшает предсказательную способность моделей в условиях нестабильности финансовых рынков. Авторы также сравнивают различные методы, включая МНК, модель Tobit и деревья решений, делая вывод о том, что МНК демонстрирует наиболее высокую эффективность для бимодального распределения LGD.

¹ Положение Банка России от 28 июня 2017 г. N 590-П «О порядке формирования кредитными организациями резервов на возможные потери по ссудам, ссудной и приравненной к ней задолженности».

² Положение Банка России от 02.11.2024 N 845-П «О порядке расчета величины кредитного риска банками с применением банковских методик управления кредитным риском и моделей количественной оценки кредитного риска».

³ Basel Committee on Banking Supervision // «Basel III: Finalising post-crisis reforms» BCBS Publications. 2017. URL: <https://www.bis.org/bcbs/publ/d424.htm> (дата обращения: 14.07.2025).

Т.В. Ишмуратова предлагает альтернативный подход, основанный на бета-трансформации и бинарной трансформации. По мнению автора, данные методы лучше учитывают бимодальное распределение LGD для необеспеченных кредитов. Точность прогнозов оценивается с помощью статистики Колмогорова–Смирнова и индекса Джини. В то же время другие акцентируют внимание на гибридных алгоритмах, интегрирующих данные о поведении заемщиков после дефолта, что повышает точность моделей (Fan et al., 2023). К. Li и другие подчеркивают значимость данных о реструктуризации долга и частичных выплатах (Li et al., 2021). Данный подход дополняется другими исследователями, которые предлагают использовать временные показатели выживаемости (time-varying survival scores) для динамического моделирования LGD (Li et al., 2023).

G. Loterman и другие сравнивают 24 регрессионных метода, включая машины опорных векторов (LSSVM), нейронные сети (ANN), многомерные аддитивные регрессионные сплайны (MARS) и МНК с различными преобразованиями (Loterman et al., 2012). Результаты исследования указывают на превосходство нелинейных моделей, таких как LSSVM и ANN, в точности прогнозов, однако их применение ограничено высокими вычислительными затратами. Loterman дополняет анализ, подчеркивая важность предобработки данных, включая обработку пропущенных значений и выбор релевантных факторов, таких как социально-демографические характеристики и уровень долговой нагрузки, для предотвращения переобучения (Loterman, 2013).

Российские исследователи, в частности С. Афанасьев и другие, выделяют два типа моделей LGD в соответствии с МСФО 9: LGD-Will-Default (для кредитов, не вошедших в статус дефолта) и LGD-In-Default (для дефолтных кредитов) (Афанасьев и др., 2021). Для второго типа применяется винтажный анализ, в рамках которого собираются наблюдения после наступления события дефолта. Авторы также описывают этапы разработки моделей: создание ядра, калибровка, оценка коэффициента спада и расчет надбавок консерватизма. Помимо этого, подчеркивается важность качества данных, что подтверждает Т. Schuermann, указывающий на необходимость учета типа кредита и информации о взыскании (Schuermann, 2004). Другие исследователи акцентируют внимание на исторических данных о взыскании, включая макроэкономические факторы и отраслевые характеристики, которые имеют слабую взаимную корреляцию, но высокую статистическую значимость (Gupton et al., 2002).

К. Li и другие предлагают использовать аппликативные и поведенческие баллы, такие как соотношение кредита к доходу, кредитный рейтинг и история просрочек, для повышения предсказательной способности моделей (Li et al., 2023). С. Афанасьев с соавторами дополняют этот подход, включая социально-демографические характеристики, поведенческую информацию и данные кредитных бюро. Другие отмечают гибкость сокращенных моделей LGD, использующих кредитные рейтинги и рыночные индикаторы (Bakshi et al., 2022). Также исследователи предлагали в качестве предикторов для моделирования LGD следующие переменные: реальный рост ВВП, доходность фондового рынка, индикатор экономической стабильности, уровень безработицы в стране и изменения долгосрочных процентных ставок (Galow et al., 2024).

Кроме того, регуляторные стандарты, а именно МСФО 9 и Базель III, существенно влияют на моделирование LGD. В одном из исследований подчеркивается, что МСФО 9 требует точного учета ожидаемых кредитных убытков, что усложняет процесс оценки LGD, но повышает точность прогнозов (Galow et al., 2024).

шает прозрачность расчетов (Porretta et al., 2020). С. Афанасьев и соавторы детализируют требования Банка России, включая единое определение дефолта. Авторы также определяют минимальный объем выборки согласно правилу «Number of Events Per Variable» (NEPV) и рекомендуют использовать безрисковую ставку с премией за риск для дисконтирования денежных потоков, что учитывает рыночные условия.

S. Caivano предлагает эконометрический подход с учетом макроэкономических факторов и корректной ставки дисконтирования (Caivano, 2016), а другие исследователи развивают методологию корректировки ставки восстановления в соответствии с требованиями сопоставимости финансовой отчетности согласно МСФО 16 (Delgado-Vaquero et al., 2022). М.В. Помазанов акцентирует внимание на минимизации остаточного риска в условиях ограниченных данных, предлагая баланс между точностью и устойчивостью моделей (Помазанов, 2021).

Таким образом, литература в области моделирования LGD демонстрирует высокую степень разработанности темы, предлагая широкий спектр методов – от статистических (МНК, Tobit) до нелинейных (нейронные сети, LSSVM). Однако актуальным остается вопрос об оптимальном сочетании данных и адаптации методов и алгоритмов к локальным рынкам, в особенности российскому, где ограниченность данных о дефолтах и конфиденциальность информации создают дополнительные сложности. Разработка моделей, устойчивых к экономическим шокам и соответствующих регуляторным стандартам, представляется важной задачей коммерческих банков.

2. Описание используемых данных

Для оценки и прогнозирования LGD были использованы закрытые данные по дефолтам заемщиков розничного портфеля краткосрочных необеспеченных потребительских кредитов одного из российских банков. Количество наблюдений составляет 782 317. Период выборки: с декабря 2020 г. по март 2025 г.

По розничному портфелю была предоставлена выгрузка по следующим параметрам кредитов и факторам:

- отчетная дата;
- уникальный номер договора;
- дата рождения заемщика;
- дата выдачи кредита;
- дата первой просрочки;
- дата последнего платежа по основному долгу;
- сумма последнего платежа по основному долгу в рублях;
- срок договора в месяцах;
- сумма выдачи в рублях;
- среднемесячный доход в рублях;
- задолженность основного долга на отчетную дату в рублях;
- количество дней просрочки в днях.

3. Методология и модель:

Методология предполагает два этапа. В рамках первого производится количественная оценка уровня потерь по розничным необеспеченным кредитам банка, реализуемая в Excel.

На втором этапе осуществляется прогнозирование LGD с помощью моделей машинного обучения XGBoost и LightGBM посредством использования языка программирования Python.

На первом этапе применения методологии производится последовательный расчет необходимых для оценки LGD показателей по розничному кредитному портфелю в Excel. Для этого обозначим следующие формулы (Помазанов, 2021):

$$R = \sum_{t=0}^{\infty} \frac{P_t - C_t}{(1+q)^t}, \quad (1)$$

где R – сумма восстановления,

P_t – восстановительные платежи в момент времени t от даты дефолта,

C_t – затраты на издержки банка по взысканию задолженности.

$\frac{P_t - C_t}{(1+q)^t}$ – дисконт-фактор со ставкой q .

$$LGD = 1 - RR, \quad (2)$$

где LGD (*Loss Given default*) – доля от суммы, подверженной кредитному риску, которая может быть потеряна в случае дефолта контрагента (в %),

RR (*Recovery rate*) – ожидаемый процент возврата по просроченным выплатам.

$$RR_{avg} = \frac{1}{n} \sum_{i=1}^n \frac{R_i}{E_i}, \quad (3)$$

где R_i – сумма поступлений средств в счет погашения задолженности заемщика i , дисконтированных к дате дефолта,

E_i – EAD заемщика i ,

EAD (*Exposure at default*) – сумма основного долга, начисленных процентов, штрафов и прочих начислений к отчетному периоду до дефолта.

Для целей настоящего исследования вводятся следующие определения: месяц дефолта (Month of Default, далее – MoD) – месяц, в котором заемщик впервые перешел в состояние дефолта, и дата дефолта – дата первого выхода в дефолт. Моделирование уровня потерь при дефолте (LGD) осуществляется с использованием винтажного анализа, в котором под винтажом понимается временной срез, определяемый длительностью нахождения кредитного обязательства в состоянии дефолта. Под последним понимается просрочка платежей по основному долгу и процентам на 90 и более дней, что соответствует Положению № 590-П от 28.06.2017 Банка России, и банкротство физического лица.

Оценка LGD для необеспеченных краткосрочных потребительских кредитов включает следующие этапы.

1. Определение категории просрочки («DPD bucket») 91–120 дней, где значение «1» присваивается кредитам с просрочкой 91–120 дней (первый выход в дефолт), а «0» – в остальных случаях. Расчет выполняется с использованием функции Excel «ЕСЛИ» с множественными условиями, фиксируя винтаж для последующего анализа.

2. Установление даты дефолта по уникальному номеру кредитного договора с применением функции «МИНЕСЛИ».

3. Вычисление MoD как разницы между отчетной датой и датой дефолта в месяцах.

Расчет даты дефолта выполняется для каждого кредитного обязательства, которое характеризуется уникальной датой выхода в дефолт. Это необходимо для агрегирования суммы задолженности портфеля, перешедшего в дефолт на определенную дату.

4. Формирование сводной таблицы, где строки соответствуют датам дефолта, столбцы – MoD, а значения – сумме основного долга в рублях, для учета поступлений от заемщика в погашение просроченной задолженности.

5. Расчет уровня восстановления (Gross Recovery Rate, RR) и LGD:

- обработка аномалий (например, от продажи кредита) путем замены выбросов средними значениями двух соседних ячеек;
- заполнение пропущенных значений средними за шесть соседних месяцев;
- дисконтирование RR с использованием средневзвешенной ставки по краткосрочным потребительским кредитам, рассчитанной Банком России⁴;
- вычисление LGD в % для каждой даты дефолта как единица за вычетом суммы дисконтированных RR;
- агрегирование LGD по последним 12 месяцам для получения общего показателя по портфелю необеспеченных кредитов, составившего 75%.

Наряду с вычислением общего уровня потерь при дефолте (LGD) за последние 12 месяцев, определены LGD для всех дат дефолта и соответствующих месяцев после дефолта (MoD). Это обеспечивает достаточный объем данных для моделирования показателя с использованием методов машинного обучения в Python.

С целью построения прогноза указанного параметра необходимо сформировать выборку данных, включающую значения LGD и его возможные предикторы: макроэкономические факторы и различные характеристики кредитного портфеля. Далее требуется построить корреляционную матрицу факторов с целью выбора наиболее подходящей модели.

В качестве предикторов LGD использовались следующие факторы.

• Макроэкономические индикаторы:

- курс USD/RUB (индекс 40329 в cbonds);
- ключевая ставка Банка России (%)⁵;
- средневзвешенная ставка по краткосрочным кредитам физических лиц (%, ЦБ);
- уровень инфляции (ИПЦ, %, Росстат⁶, на конец предыдущего месяца);
- среднемесячная номинальная заработная плата в РФ (р.)⁷;
- реальные располагаемые доходы населения (% к прошлому году, ежеквартально, индекс 98646, cbonds);

⁴ Банк России. Процентные ставки по кредитным и депозитным операциям кредитных организаций в рублях: URL: https://www.cbr.ru/statistics/bank_sector/int_rat/1124/ (дата обращения: 10.03.2025).

⁵ Банк России. Ключевая ставка Банка России: URL: https://www.cbr.ru/hd_base/keyrate/ (дата обращения: 03.04.2025).

⁶ Федеральная служба государственной статистики. Цены, инфляция: URL: <https://rosstat.gov.ru/statistics/price> (дата обращения: 17.04.2025).

⁷ Федеральная служба государственной статистики. Рынок труда, занятость и заработная плата: URL: https://rosstat.gov.ru/labor_market_employment_salaries (дата обращения: 17.04.2025).

- ВВП в текущих ценах (ежеквартально, индекс 93926, cbonds).
- Характеристики кредитного портфеля:
 - средний возраст заемщика;
 - средний срок договора в месяцах;
 - среднемесячный доход физических лиц;
 - среднее количество дней просрочки.

Корреляционная матрица факторов модели представлена на рис. 2. Анализ указывает на высокую корреляцию средних значений количества дней просрочки и уровня безработицы с большинством предикторов. При наличии проблемы мультиколлинеарности факторов целесообразно применять машинное обучение. Это позволит минимизировать ошибки прогнозирования LGD, так как при использовании классических методов, например линейной регрессии, существует высокая вероятность незначимости факторов и низких значений R-квадрата. Поэтому для реализации второго этапа методологии предлагается применять современные модели машинного обучения градиентного бустинга: Extreme Gradient Boosting (далее – XGBoost) и Light Gradient Boosting Machine (далее – LightGBM).

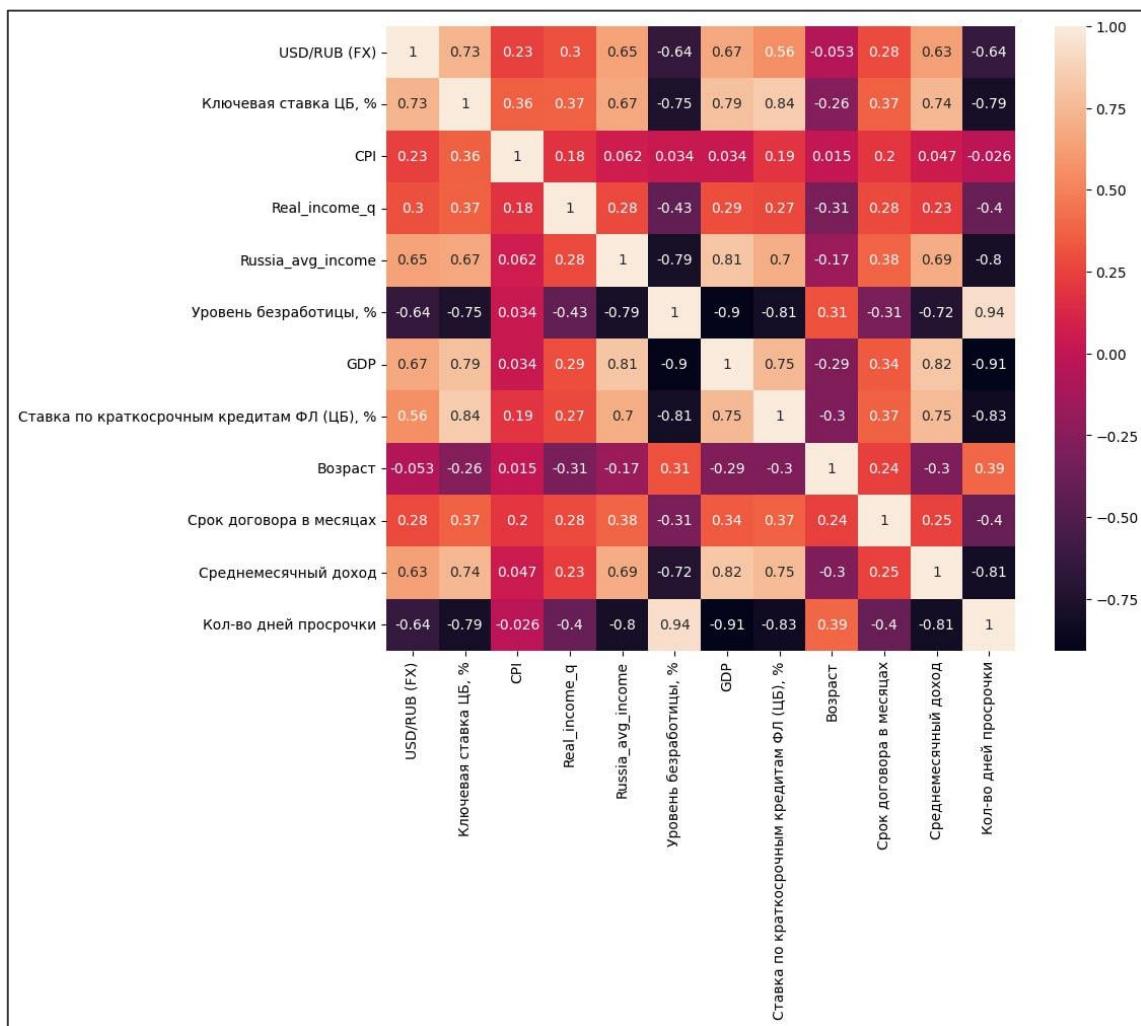


Рисунок 2. Матрица корреляций факторов⁸ (составлено автором)

⁸ Наименования: CPI – ИПЦ на товары и услуги по РФ к концу предыдущего месяца, %; GDP – ВВП в текущих ценах, ежеквартальный; Real_income_q – реальные располагаемые денежные доходы населения, в % к соответствующему периоду прошлого года (ежеквартальный); Russia_avg_income – среднемесячная номинальная начисленная заработка работников организаций (РФ) (рассчитано автором).

XGBoost и LightGBM представляют собой алгоритмы машинного обучения, основанные на градиентном бустинге, которые эффективно применяются для задач регрессии, классификации и прогнозирования, включая моделирование уровня потерь при дефолте (LGD). Их работа заключается в последовательном построении ансамбля деревьев решений, где каждое последующее дерево корректирует ошибки предыдущих, минимизируя заданную функцию потерь (например, среднеквадратическую ошибку).

XGBoost предполагает добавление регуляризации (L1 и L2) для предотвращения переобучения. Алгоритм итеративно добавляет деревья, минимизируя остатки ошибок, посредством использования уровневого роста дерева («level-wise»), т. е. построение дерева по уровням. Преимуществом XGBoost является учет нелинейных зависимостей между факторами.

LightGBM, в отличие от XGBoost, использует гистограммный подход для разбиения данных, что ускоряет обучение. В алгоритме применяется разбиение по принципу «leaf-wise» (рост дерева по лучшему листу). Преимуществом данной модели является оптимизация памяти.

Для построения указанных моделей выборка данных была поделена на тестовую и обучающую в соотношении 20%: 80%. Иллюстрация моделей XGBoost и LightGBM содержится на рис. 3 и 4 соответственно.

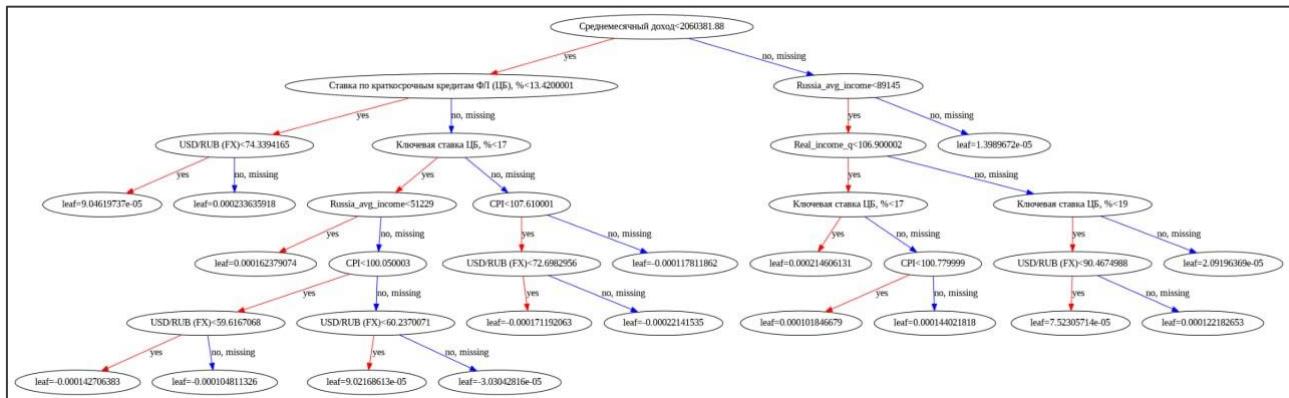


Рисунок 3. Иллюстрация XGBoost (Дерева решений) (составлено автором)

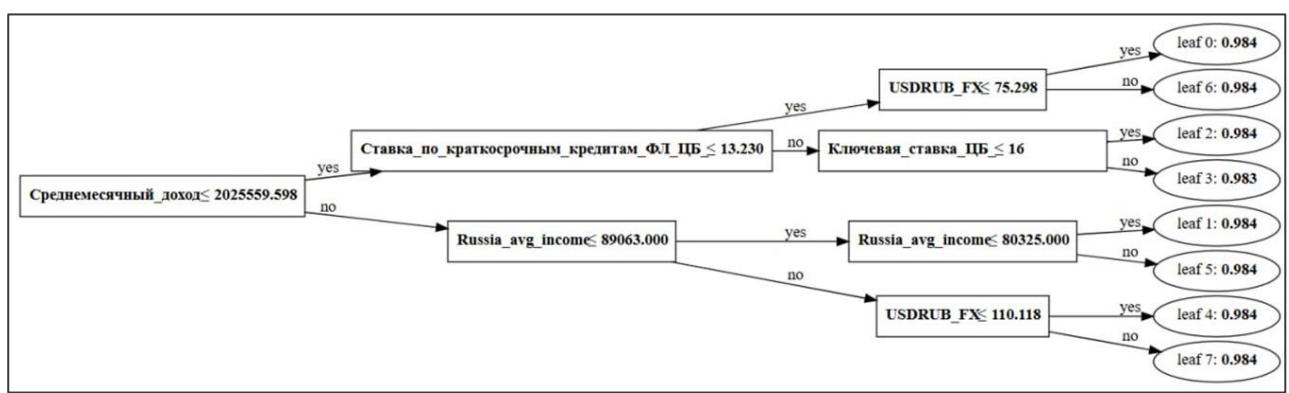


Рисунок 4. Иллюстрация LightGBM (составлено автором)

4. Результаты

В рамках первого этапа методологии итоговый уровень потерь при дефолте (LGD) по портфелю необеспеченных кредитов физических лиц за последние 12 месяцев составил 75% (см. рис. 5). Этот показатель характерен для розничного кредитования и показывает, что при

текущей структуре портфеля банк может потерять 75% суммы задолженности в сегменте небеспеченных кредитов, перешедших в дефолт. На основе данного значения банк способен оценивать потребность в формировании резервов на возможные потери по ссудам.

Рисунок 5. Расчет уровня восстановления (RR), дисконтированных уровней потерь при дефолте и итоговый LGD по портфелю на 12 месяцах (составлено автором)

В рамках второго этапа методологии были применены модели XGBoost и LightGBM с целью прогнозирования LGD для розничного портфеля необеспеченных потребительских кредитов. Оценка точности алгоритмов была произведена с использованием метрик средней абсолютной ошибки, средней абсолютной процентной ошибки и среднеквадратической ошибки.

Результаты вычисления указанных метрик представлены в табл. 1 (см. ниже). Анализ показывает, что обе модели характеризуются низким уровнем ошибок. Это свидетельствует о высокой точности прогнозных значений уровня потерь при дефолте (LGD) по необеспеченным потребительским кредитам, что подтверждает целесообразность применения данных моделей для банков в случае наличия проблемы мультиколлинеарности факторов.

Таблица 1

Результаты оценки точности моделей XGBoost и LightGBM

Показатель	Результат для модели XGBoost		Результат для модели LightGBM	
	Тестовая выборка	Обучающая выборка	Тестовая выборка	Обучающая выборка
MSE	0,0244%	0,0238%	0,0244%	0,0238%
MAE	1,18%	1,14%	1,18%	1,14%
MAPE	1,22%	1,17%	1,22%	1,17%

Источник: составлено автором.

Также были выявлены топ-10 факторов по уровню силы в предложенных моделях. Результат по модели XGBoost представлен на рис. 6. Анализ показал, что для данного алгоритма самыми важными предикторами являются среднемесячный доход, ставка по краткосрочным кредитам физических лиц, средний возраст заемщика и ключевая ставки Банка России.

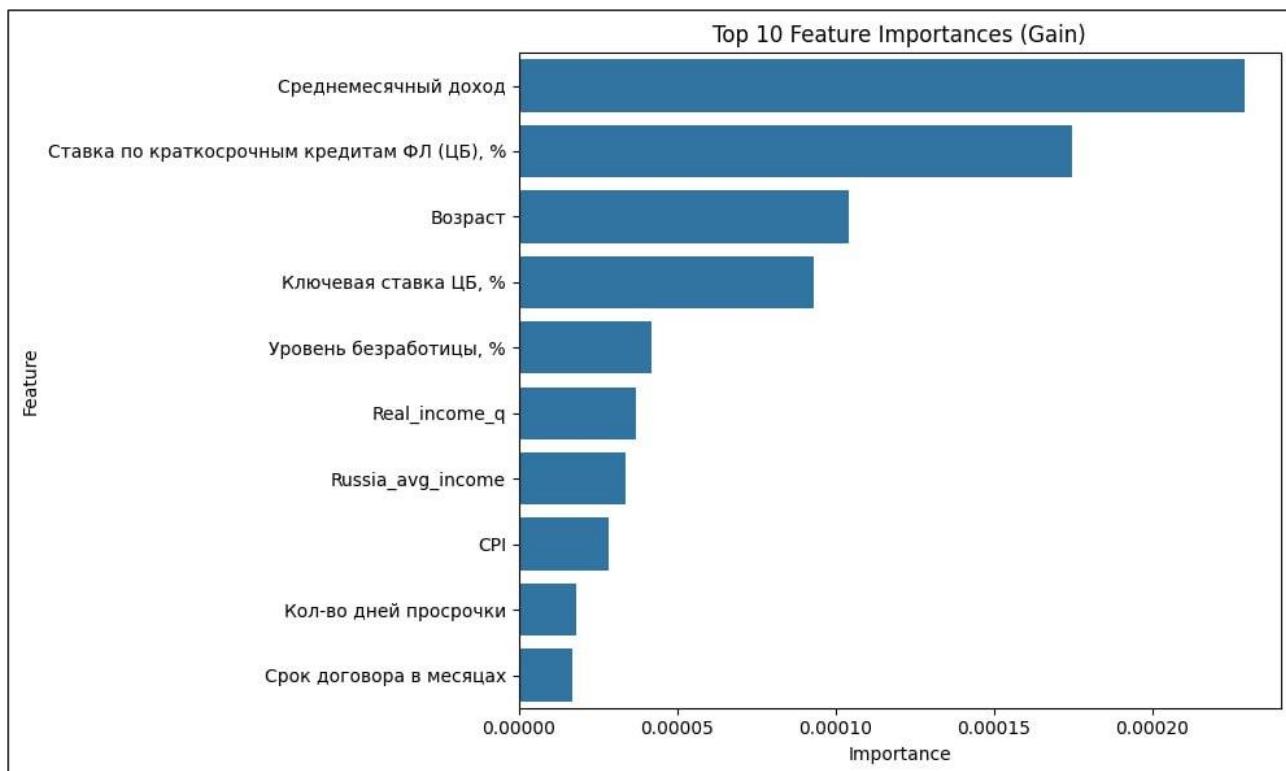


Рисунок 6. Оценка значимости факторов для модели XGBoost (составлено автором)

Для модели LightGBM самую высокую силу показали следующие факторы: курс доллара на Форексе, средний возраст заемщика, среднемесячная начисленная заработка по РФ, срок кредитного договора в месяцах, а также индекс потребительских цен. Иллюстрация топ-10 предикторов представлена на рис. 7.

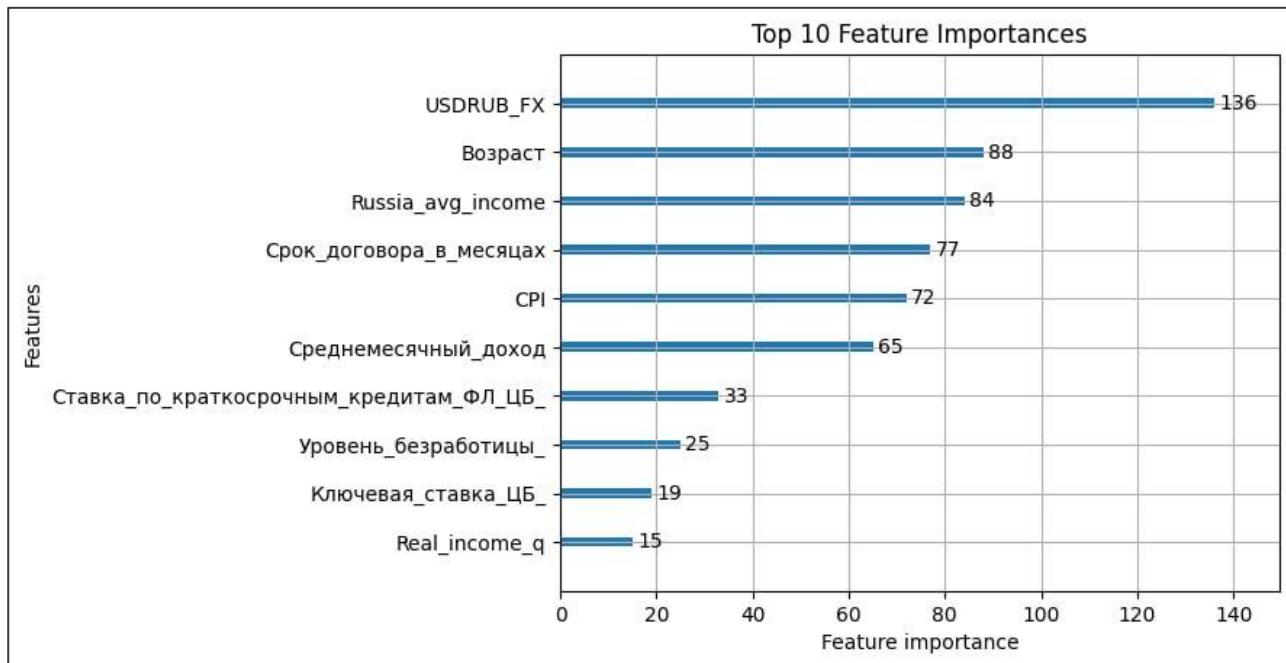


Рисунок 7. Оценка значимости факторов для модели LightGBM (составлено автором)

Таким образом, использованные модели машинного обучения демонстрируют высокую точность прогнозных значений LGD по кредитным обязательствам на дату дефолта, что подтверждается низким уровнем ошибок. Проведен анализ значимости факторов, в результате которого ключевыми предикторами являются среднемесячный доход заемщика, его возраст и

уровень процентных ставок. Предложенные модели и факторы могут быть эффективно использованы коммерческими банками для оценки LGD.

Заключение

Оценка и прогнозирование уровня потерь при дефолте (LGD) являются одними из ключевых задач кредитной организации с целью обеспечения соответствия регуляторным требованиям Базеля, стандартов МСФО и Банка России, а также грамотного формирования резервов на возможные убытки. Исследования в данной области выделяют разнообразные методы моделирования LGD, включая линейные (МНК, бета-регрессия) и нелинейные (нейронные сети, деревья решений) подходы и винтажный анализ.

При этом моделирование осложняется бимодальностью распределения параметра LGD, ограниченностью доступных данных о дефолтах, необходимостью обработки данных и выбором наиболее значимых предикторов. Важным является также соответствие банковских процессов требованиям регулятора и обеспечение прозрачности оценок показателя потерь при дефолте с высокой точностью моделей прогнозов.

В исследовании предложена двухэтапная методика оценки LGD на основе данных по необеспеченным розничным кредитам российского банка. В рамках первого этапа посредством винтажного анализа рассчитываются значения LGD в каждом месяце дефолта и суммарный показатель потерь при дефолте на горизонте 12 месяцев. На втором этапе осуществляется прогнозирование параметра благодаря моделям машинного обучения XGBoost и LightGBM, которые успешно решают проблему мультиколлинеарности факторов. В качестве предикторов LGD были использованы различные макроэкономические показатели (например, курс USD/RUB, инфляция, ВВП) и такие характеристики портфеля, как средний возраст заемщика, доход и количество дней просрочки. Обе модели показали низкие ошибки, подтверждая целесообразность их применения в банковской практике. Дальнейшие исследования могут быть направлены на выявление новых предикторов и адаптацию моделей к ужесточающейся макропруденциальной политике ЦБ в области необеспеченных потребительских кредитов и высоким процентным ставкам по кредитам в России.

Список литературы

Афанасьев С., Молоканов И., Ахметсафин И., Смирнова А. Разработка LGD-моделей для розничного кредитования. Ч. 1: подготовка данных // Риск-менеджмент в кредитной организации. 2021. №. 3. С. 4–23.

Ишмуратова Т.В. Математическое моделирование потерь в случае наступления дефолта на уровне кредитного обязательства // Вестник Чувашского университета. 2018. № 3. С. 192–205.

Помазанов М.В. Моделирование потерь в случае дефолта в концепции минимизации остаточного риска // Управление финансовыми рисками. 2021. №. 3. С. 170–187.

Bakshi G., Gao X., Zhong Z. Decoding default risk: A review of modeling approaches, findings, and estimation methods // Annual Review of Financial Economics. 2022. Vol. 14. No. 1. P. 391–413.

Bellotti T., Crook J. Loss given default models incorporating macroeconomic variables for credit cards // International Journal of Forecasting. 2012. Vol. 28. No. 1. P. 171–182.

Caivano S.B.G. Econometric approach for Basel III Loss Given Default Estimation: from discount rate to final multivariate model. 2016.

Delgado-Vaquero D., Morales-Diaz J., Zamora-Ramírez C. IFRS 16 incremental borrowing rate: Comparability issues and a methodology proposal for loss given default adjustment // Accounting in Europe. 2022. Vol. 19. No. 2. P. 287–310.

Fan M., Wu T.H., Zhao Q. Assessing the loss given default of bank loans using the Hybrid Algorithms Multi-Stage model // Systems. 2023. Vol. 11. No. 10. P. 505.

Galow B., Georgescu O.M., Ponte Marques A. Loss-Given-Default and Macroeconomic Conditions // ECB Working Paper. 2024. No. 2024/2954.

Gupton G.M. et al. LossCalc: Model for predicting loss given default (LGD). 2002.

Li A., Li Z., Bellotti A. Predicting loss given default of unsecured consumer loans with time-varying survival scores // Pacific-Basin Finance Journal. 2023. Vol. 78. P. 101949.

Li K. et al. Predicting loss given default using post-default information // Knowledge-Based Systems. 2021. Vol. 224. P. 107068.

Loterman G. et al. Benchmarking regression algorithms for loss given default modeling // International Journal of Forecasting. 2012. Vol. 28. No. 1. P. 161–170.

Loterman G. Predicting loss given default. Ghent University, 2013.

Porretta P. et al. Credit risk management in bank. Impacts of IFRS 9 and Basel 3 // Risk Governance and Control: Financial Markets and Institutions. 2020. Vol. 10. No. 2. P. 29–44.

Rantanen J. Modeling loss given default on unsecured consumer loans: case study of a Finnish financial institution. 2023.

Schuermann T. What do we know about loss given default? 2004.

Статистические показатели банковского сектора Российской Федерации» // Банк России: официальный сайт: URL: https://www.cbr.ru/statistics/bank_sector/review/ (дата обращения: 07.04.2025).

LOSS GIVEN DEFAULT MODELLING FOR A BANK'S RETAIL PORTFOLIO OF UNSECURED LOANS

Daria D. Frolova

Master's student,

*Lomonosov Moscow State University, Faculty of Economics
(Moscow, Russia)*

Abstract

Proper modelling of loss given default (LGD) is an important task for commercial banks. The relevance of the study is due to the active expansion of retail lending over the past 5 years. The increase in the volume of the retail portfolio from 17.6 to 36.9 trillion rubles from the 1st of January until the 1st of January 2025 contributed to an increase in the credit risk of Russian commercial banks. Additional difficulties arise due to the high level of the Bank of Russia's key interest rate and the tightening of macroprudential policy in the segment of unsecured retail loans. The paper describes the main approaches to LGD modelling. Based on data on unsecured consumer loans from one of the Russian banks for the period from December 2020 to March 2025, which contain more than 700,000 observations, a two-stage LGD assessment methodology has been proposed. The approach includes vintage analysis and the use of machine learning models (XGBoost and LightGBM) to assess and predict LGD considering new macroeconomic factors: the USD/RUB exchange rate on the Forex market and the average monthly salary of Russian employees.

Keywords: Loss Given Default, retail portfolio, commercial banks, unsecured retail loans, modelling, vintage analysis, machine learning.

JEL: C53, G17, G21.

For citation: Frolova, D.D. (2025) Loss Given Default Modelling for a Bank's Retail Portfolio of Unsecured Loans. Scientific Research of Faculty of Economics. Electronic Journal, vol. 17, no. 4, pp. 136-150. DOI: 10.38050/2078-3809-2025-17-4-136-150.

References

- Afanas'ev S., Molokanov I., Akhmetsev I., Smirnova A. Razrabotka LGD-modeley dlya roznichnogo kreditovaniya. Ch. 1: podgotovka dannykh. Risk-menedzhment v kreditnoy organizatsii. 2021. №. 3. S. 4–23. (In Russ.).
- Ishmuratova T.V. Matematicheskoe modelirovaniye poter' v sluchae nastupleniya defolta na urovne kreditnogo obyazatel'stva. Vestnik Chuvashskogo universiteta. 2018. № 3. S. 192–205. (In Russ.).
- Pomazanov M.V. Modelirovaniye poter' v sluchae defolta v kontseptsii minimizatsii ostatochnogo riska. Upravlenie finansovymi riskami. 2021. №. 3. S. 170–187. (In Russ.).

Bakshi G., Gao X., Zhong Z. Decoding default risk: A review of modeling approaches, findings, and estimation methods. *Annual Review of Financial Economics*. 2022. Vol. 14. No. 1. P. 391–413.

Bellotti T., Crook J. Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting*. 2012. Vol. 28. No. 1. P. 171–182.

Caivano S.B.G. Econometric approach for Basel III Loss Given Default Estimation: from discount rate to final multivariate model. 2016.

Delgado-Vaquero D., Morales-Díaz J., Zamora-Ramírez C. IFRS 16 incremental borrowing rate: Comparability issues and a methodology proposal for loss given default adjustment. *Accounting in Europe*. 2022. Vol. 19. No. 2. P. 287–310.

Fan M., Wu T.H., Zhao Q. Assessing the loss given default of bank loans using the Hybrid Algorithms Multi-Stage model. *Systems*. 2023. Vol. 11. No. 10. P. 505.

Galow B., Georgescu O.M., Ponte Marques A. Loss-Given-Default and Macroeconomic Conditions // ECB Working Paper. 2024. No. 2024/2954.

Gupton G.M. et al. LossCalc: Model for predicting loss given default (LGD). 2002.

Li A., Li Z., Bellotti A. Predicting loss given default of unsecured consumer loans with time-varying survival scores. *Pacific-Basin Finance Journal*. 2023. Vol. 78. P. 101949.

Li K. et al. Predicting loss given default using post-default information. *Knowledge-Based Systems*. 2021. Vol. 224. P. 107068.

Loterman G. et al. Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*. 2012. Vol. 28. No. 1. P. 161–170.

Loterman G. Predicting loss given default. Ghent University, 2013.

Porretta P. et al. Credit risk management in bank. Impacts of IFRS 9 and Basel 3. *Risk Governance and Control: Financial Markets and Institutions*. 2020. Vol. 10. No. 2. P. 29–44.

Rantanen J. Modeling loss given default on unsecured consumer loans: case study of a Finnish financial institution. 2023.

Schuermann T. What do we know about loss given default? 2004.

Статистические показатели банковского сектора Российской Федерации». Bank Rossii: ofitsial'nyy sayt: Available at: https://www.cbr.ru/statistics/bank_sector/review/ (accessed: 07.04.2025). (In Russ.).