

Методологический подход к моделированию величины потерь в случае дефолта в рамках оценки кредитного риска

Девайкина Анастасия Сергеевна 

Аспирант

Новосибирский государственный университет, г. Новосибирск, Российская Федерация

E-mail: i.nuriev@g.nsu.ru

Ибрагимов Наимджон Мулабоевич  *

Доктор экономических наук, доцент

Новосибирский государственный университет, г. Новосибирск, Российская Федерация;

Институт экономики и организации промышленного производства СО РАН, г. Новосибирск, Российская Федерация;

Новосибирский государственный технический университет, г. Новосибирск, Российская Федерация

E-mail: naimdjon.ibragimov@nsu.ru

КЛЮЧЕВЫЕ СЛОВА

кредитный риск, доля потерь при дефолте, мультимодальное распределение, интерпретируемость, непараметрические модели, гибридные модели

АННОТАЦИЯ

Доля потерь банка в случае реализации дефолта заемщика – это безусловно критически важный показатель, без точного прогнозирования которого невозможно корректно оценить уровень кредитного риска, адекватно сформировать резервы под ожидаемые кредитные убытки и рассчитать достаточный объем капитала, необходимого для защиты финансовой устойчивости учреждения от непредвиденных потерь, связанных с невозвратом кредитов. Этот показатель, известный как Loss Given Default (LGD), в частности отражает долю невозвращённого долга относительно его суммы и напрямую влияет на уровень резервирования, нормативные требования и стратегическое планирование банковской деятельности. Прогнозирование доли потерь при дефолте однозначно требует учета ряда специфических свойств данных, таких как мультимодальность и асимметричность распределения, ограничение значений интервалом от 0% до 100%+, нелинейность влияния объясняющих факторов. В статье проанализированы исследования, посвящённые моделированию доли потерь в случае дефолта в рамках оценки кредитного риска. Описаны метрики качества модели в разрезе четырёх ключевых критериев: точности, дискриминационной способности, стабильности и интерпретируемости. Сделаны выводы и даны рекомендации по разработке и внедрению моделей в банковские процессы. Главный итоговый вывод заключается в том, что гибридные модели, сочетающие преимущества параметрических и непараметрических подходов, а также интерпретируемые непараметрические модели набирают популярность благодаря способности сохранять как точность прогнозов, так и интерпретируемость. Тем не менее, параметрические модели остаются базовым инструментом анализа и бенчмарком для оценки прироста качества более сложных моделей.

JEL codes: G32; G33; C51; C58

DOI: <https://doi.org/10.52957/2221-3260-2026-6-44-58>

Благодарность: Исследование выполнено по плану НИР ИЭОПП СО РАН, проект 7.5.2.1 «Модели и методы исследования социально-экономического развития многоуровневой пространственной системы Российской Федерации в современных условиях», № 1025031900028-1.

Для цитирования: Девайкина, А.С. Методологический подход к моделированию величины потерь в случае дефолта в рамках оценки кредитного риска / А.С. Девайкина, Н.М. Ибрагимов. – Текст : электронный // Теоретическая экономика. – 2026. – № 6. – С.44-58. – URL: <http://www.theoreticaleconomy.ru> (Дата публикации: 30.06.2026)

Введение

Для минимизации потерь в случае осуществления кредитного риска коммерческому банку требуется формировать резерв под ожидаемые кредитные убытки и надежно оценить капитал для защиты от непредвиденных кредитных убытков. В соглашении Базель II (2006) [4] предлагается использовать индивидуальную оценку кредитного риска для каждого заемщика и отражается, что в этих целях банку следует оценить три параметра:

- Вероятность дефолта (PD - Probability of Default) – вероятность, что контрагент не выполнит обязательства, как правило, в течение ближайших 12 месяцев.
- Долю потерь при дефолте (LGD - Loss Given Default) – процент кредитного требования, который будет потерян банком в случае реализации дефолта.
- Величину кредитного требования на момент дефолта (EAD - Exposure at Default) – рыночная стоимость требования.

Центральный Банк Российской Федерации в 2025 году опубликовал Положение № 845-П от 02.11.2024 [36]. В документе указано, что банк, включенный в перечень системно значимых кредитных организаций на 1 января 2025 года, должен к 1 января 2030 года применять внутренние методики и модели на основе внутренних рейтингов для оценки не менее 50% от суммы кредитных требований.

Данные требования по обязательному применению внутренних моделей количественной оценки кредитного риска в ближайшем будущем существенно повышают необходимость разработки и внедрения в российских банках моделей оценки компонентов кредитного риска. Разработка качественных моделей PD, LGD и EAD становится не просто задачей совершенствования рискового управления, но и необходимостью для выполнения требований регулятора и поддержания стабильности всей банковской системы. Таким образом, проблема разработки и внедрения надежных интерпретируемых моделей оценки потерь от реализации кредитного риска становится актуальной для российских банков, особенно для системно значимых организаций, которым необходимо соответствовать новым требованиям Центрального Банка.

Моделирование компонента доли потерь при дефолте (LGD) является особо сложной задачей, поскольку необходимо учитывать специфические свойства данных о потерях при дефолте, такие как:

- Мультимодальность и асимметричность распределения. В зависимости от портфеля может наблюдаться большая концентрация наблюдений у нулевых потерь, либо у полных потерь, а также довольно часто распределение имеет одновременно оба этих пика [28]. Это приводит к проблеме дисбаланса данных и усложняет выбор функциональной формы модели для точной оценки доли потерь при дефолте.
- Ограничение значений доли потерь при дефолте интервалом $[0,1]$, что предполагает, что банк не взыскивает и не теряет больше непогашенной суммы. Однако, в некоторых случаях значения доли потерь при дефолте все-таки могут превышать 100%, т.к. банк, помимо потерь по долгу заемщика, несет издержки на взыскание. Кроме того, доля потерь может оказаться ниже нулевых из-за процентных платежей или высокого залога. Например, Schmit, Stuyck (2002) [40] указывают, что до 59% дефолтных лизинговых контрактов в их выборке имеют уровень возврата, превышающий 100%.
- Нелинейное влияние различных риск-факторов на целевую переменную.

Требования к моделям оценки доли потерь при дефолте

Для того чтобы использовать разработанную модель прогнозирования доли потерь при дефолте в оценке капитала в соответствии с Положением № 845-П (2024) [36] банку следует оценивать количественные характеристики качества модели.

1) Точность

В научных исследованиях качество модели чаще всего определяют именно по точности. Метрики точности измеряют степень расхождения между прогнозами модели и фактическими данными. Можно выделить следующие наиболее популярные метрики для определения точности моделей оценки доли потерь при дефолте у исследователей [5, 22, 26, 33, 47]:

- MAE (средняя абсолютная ошибка);
- MSE (средний квадрат ошибок);
- RMSE (среднеквадратическая ошибка);
- R^2 и R^2 скорректированный

Чем меньше значение метрик ошибок прогноза MAE, MSE, RMSE, тем точнее модель справляется с задачей оценки целевой переменной. Чем выше значение R^2 и R^2 скорректированного, тем большую долю вариации зависимой переменной объясняет модель.

В Положении № 845-П (2024) [36] требуется использовать метрику Loss Shortfall, которая показывает, насколько предсказанные значения убытков отличаются от наблюдаемых значений. Показатель рассчитывается с учетом величины задолженности по ссуде на дату, что позволяет оценить точность оценки потерь банка в рублях, а не в долях. Такой подход более практико-ориентирован и учитывает финансовую значимость ошибок в прогнозировании убытков для банка.

2) Дискриминационная способность

Оценка дискриминационной способности является более сложной по сравнению с оценкой точности. Для оценки дискриминационной способности модели на практике банков, а также в научной литературе [7, 19, 45] популярны следующие методы:

- Ранговая корреляция Спирмена,
- Ранговая корреляция Кендалла,
- SomersD.

Данные метрики являются мерами измерения ранговой ассоциации между рядами и основываются на определении согласующихся (с увеличением значения переменной X увеличивается значение переменной Y) и несогласующихся (с увеличением значения переменной X уменьшается значение переменной Y) пар значений. Коэффициент может принимать значения от -1 до 1, при этом чем ближе значение коэффициента по модулю к 1, тем сильнее связь между рядами (если коэффициент стремится к -1 – связь обратная, если к 1 – связь прямая).

Следует отметить, что для моделей оценки вероятности дефолта (PD) на практике используется метрика оценки дискриминационной способности ROC-AUC на основе ROC-кривой. Но такой подход не может быть применен к оценке моделей прогноза доли потерь при дефолте (LGD), поскольку целевая переменная не является бинарной. В работе Frontczak et al. (2017) [13] представлено обоснование альтернативной метрики оценки дискриминационной способности на основе кривых Лоренца, подходящей для моделей LGD - PowerStat. На практике существуют различные подходы к расчету данной метрики - на разных составах выборки, с разным порядком сортировки элементов. В исследовании Polianskii (2021) [35] определен более предпочтительный алгоритм расчета. В статье Kazianka et al. (2021) [23] сравниваются различные способы оценки дискриминационной способности моделей LGD, подчеркиваются недостатки метрик, основанных исключительно на парных сравнениях, и предлагается метрика VUS, которая обобщает концепцию ROC-кривой для многоклассовой классификации.

3) Стабильность

Стабильность модели доли потерь при дефолте часто оценивается отсутствием значимого падения метрик точности или дискриминационной способности на out-of-time выборке, либо на out-of-sample выборке, либо на k подвыборках доступного набора данных [10, 16, 31, 33, 44, 46].

Для оценки стабильности модели используют критерий согласия Колмогорова-Смирнова, который позволяет определить, относятся ли две выборки к одной и той же генеральной совокупности, а также индекс стабильности популяции, с помощью которого можно оценить

насколько сильно изменилось распределение [20, 29].

Для подтверждения стабильности модели некоторые исследователи проверяют согласованность результатов моделирования с результатами повторного моделирования на скорректированных выборках. Например, Dermine, de Carvalho (2006) [11] перестроили модель после исключения из выборки 10% крупнейших кредитов и зафиксировали согласованность новых результатов с результатами базовых спецификаций, подтверждая статистическую значимость риск-факторов модели. Leow, Mues (2012) [26] провели переоценку модели после удаления из обучающей выборки всех наблюдений с повторными дефолтами заемщика и убедились в стабильности модели.

4) Интерпретируемость

Ключевым требованием к моделям кредитного риска является их интерпретируемость. Интерпретируемость обеспечивает прозрачность оценок и вызывает доверие у регулирующего органа, поэтому банки как правило в моделях не применяют «черные ящики». Это в свою очередь может затруднять достижение высокой статистической эффективности модели [1].

Функциональные формы моделей оценки доли потерь при дефолте

В данном разделе представлен обзор ключевых подходов к моделированию доли потерь при дефолте, описанных в научной и бизнес-литературе. Эти методы можно разделить на параметрические, непараметрические и гибридные, в зависимости от предположений о данных и функциональной форме модели. Параметрические модели предполагают строго заданную функциональную форму и распределение ошибок, что делает их интерпретируемыми, но ограниченными при работе со сложными распределениями. Непараметрические модели не делают предположений о функциональной форме и распределении данных, полагаясь на алгоритмы, которые выявляют зависимости напрямую из данных. Гибридные модели сочетают элементы данных подходов, ослабляя предположения параметрических моделей: комбинируют линейные элементы с нелинейными, что позволяет моделировать сложные зависимости при сохранении части параметрической структуры и делает их адаптивными к разнообразным данным. Каждая категория предлагает определенные возможности и ограничения для моделирования доли потерь при дефолте в зависимости от структуры данных, задач анализа, строгости требований к интерпретации алгоритма отнесения ссуды к категории риска, имеющихся временных и профессиональных ресурсов на разработку модели. Примеры исследований по формам модели приведены в Таблице 1.

Таблица 1 – Примеры исследований по формам моделей

Категория	Форма модели	Примеры исследований
Параметрические	Линейная регрессия	Grunert, Weber (2009) [14], Qi, Yang (2009) [38], Bellotti, Crook (2009) [6], Calabrese (2012) [9], Loterman et al. (2012) [28], Dahlin, Storkitt (2014) [10], LaCour-Little, Zhang (2014) [25], Bonini, de Carvalho (2016) [8], Nazemi, Fabozzi (2018) [32], Kaposty et al. (2019) [22], Bellotti et al. (2020) [5], Li et al. (2021) [27]

Категория	Форма модели	Примеры исследований
	Тобит-регрессия	Bellotti, Crook (2009) [6], Calabrese (2012) [9], Kaposty et al. (2019) [22], Bandyopadhyay (2022) [3]
	Бета-регрессия	Loterman et al. (2012) [18], Jaber et al. (2021) [18]
	Гамма-регрессия	Sigrist и Stahel (2011) [41]
Непараметрические	Деревья решений, деревья регрессий	Bellotti, Crook (2009) [6], Qi, Zhao (2011) [39], Loterman et al. (2012) [18], Dahlin, Storkitt (2014) [10], Hurlin et al. (2018) [17], Kaposty et al. (2019) [22], Bellotti et al. (2020) [5], Li et al. (2021) [27], Vlasenko (2022) [45]
	Ансамблевые подходы	Miller, Tows (2017) [30], Hurlin et al. (2018) [17], Nazemi, Fabozzi (2018) [32], Kaposty et al. (2019) [22], Bellotti et al. (2020) [5], Li et al. (2021) [27]
	Метод ближайших соседей	Bellotti et al. (2020) [5], Ptak-Chmielewska et al. (2023) [37]
	Регрессия опорных векторов	Yao et al. (2014) [46], Loterman et al. (2012) [28], Nazemi, Fabozzi (2018) [32], Nazemi et al. (2018) [34], Hurlin et al. (2018) [17], Bellotti et al. (2020) [5], Li et al. (2021) [27]
	Нейронные сети	Loterman et al. (2012) [28], Hurlin et al. (2018) [17], Kaposty et al. (2019) [22], Bellotti et al. (2020) [5], Li et al. (2021) [27]
Гибридные	Квантильная регрессия	Somers, Whittaker (2007) [42], Kruger, Rosch (2017) [24]
	Сплайн регрессия	Loterman et al. (2012) [28], Miller (2017) [31], Bellotti et al. (2020) [5]
	Модель выживания Кокса	Joubert et al. (2021) [21], Ptak-Chmielewska et al. (2023) [37]
	Конечные смеси	Baixauli, Alvarez (2010) [2], Kaposty et al. (2019) [22]

Категория	Форма модели	Примеры исследований
	Многоступенчатые модели	Calabrese (2012) [9], Loterman et al. (2012) [28], Dahlin, Storkitt (2014) [10], Gurtler, Hibbeln (2013) [15], Tong et al. (2013) [44], Tanoue et al. (2017) [43], Yao et al. (2017) [47], Fan et al. (2023) [12]
	Слияние решений на основе нечеткой логики	Nazemi (2017) [33]

Источник: составлено автором.

Параметрические модели

1) Линейная регрессия

Линейные регрессионные модели широко представлены для оценки LGD в литературе благодаря их простоте и интерпретируемости. Предполагается линейная зависимость между целевой переменной и объясняющими переменными, что позволяет получить интуитивно понятные результаты и оценить влияние отдельных факторов на целевую переменную.

Например, в исследовании Bonini, de Carvalho (2016) [8] линейная регрессия использовалась для оценки LGD по дефолтным корпоративным и розничным кредитам группы коммерческих банков под надзором ЕЦБ. Среди наиболее значимых факторов выделены: сегмент кредитования, наличие обеспечения, географическое положение и характер урегулирования. Модель показала объясняющую способность на уровне скорректированного $R^2 = 31\%$, что подчеркивает ограниченные возможности линейных моделей в учете сложных зависимостей.

В исследовании Grunert and Weber (2009) [14] метод линейной регрессии OLS использовался для анализа факторов, влияющих на коэффициенты возврата по займам немецких компаний. Результаты показали, что наличие обеспечения и тесное взаимодействие заемщика с банком положительно влияют на коэффициенты возврата, при этом более крупные компании демонстрируют меньший уровень возврата. Модель продемонстрировала объясняющую способность на уровне скорректированного $R^2 = 33\%$.

В работе Qi and Yang (2009) [38] линейная регрессия OLS применялась для анализа факторов, влияющих на долю потерь при дефолте по ипотечным кредитам. Ключевым предиктором модели выступил текущий уровень LTV, оказавший сильное влияние на LGD. Дополнительно учитывались такие факторы, как размер кредита, возраст займа, тип недвижимости, цель кредита (покупка или рефинансирование), а также макроэкономическая среда и законы о взыскании долгов в разных штатах. Модель продемонстрировала высокую по сравнению с аналогичными исследованиями объясняющую способность (скорректированный $R^2 = 61\%$). При этом исключение текущего LTV серьезно понижает метрику до 14.5%.

LaCour-Little and Zhang (2014) [25] в статье использовали линейную регрессию OLS для моделирования доли потерь при дефолте по кредитам на покупку жилья в США. Основное внимание было уделено различиям между секьюритизированными и портфельными кредитами. Выявлено, что секьюритизированные кредиты демонстрируют уровень LGD выше на 7–9%. При этом полученная линейная модель характеризуется слабой объясняющей способностью (скорректированный $R^2 = 6.4\%$), что указывает на необходимость поиска альтернативных подходов к моделированию и расширению набора объясняющих переменных.

В современных исследованиях авторы все чаще предлагают более продвинутые методы моделирования, которые могут учитывать сложную структуру данных LGD, используя при этом

линейную регрессию как базовую модель для сравнения и оценки улучшений метрик качества от использования предлагаемых идей.

2) Тобит-регрессия

Некоторые исследователи в качестве альтернативы линейной регрессии используют Тобит-регрессию, которая подходит для анализа переменных, область значений которых ограничена. Следовательно, тобит-регрессия позволяет преодолеть проблему оценок за пределами допустимого интервала значений доли потерь при дефолте и обеспечивает более реалистичные прогнозы. При этом сохраняются основные недостатки, присущие линейным моделям.

Тобит-регрессия применялась, например, в исследовании потерь индийских банков Vandyopadhyay (2022) [3]. Результаты показали, что значимыми факторами являются наличие обеспечения, регион, размер займа, тип кредита, доля стоимости залога в сумме долга. Модель обеспечила псевдо R^2 на уровне 44%.

3) Бета-регрессия

Бета-регрессия подходит для моделирования переменных из интервала значений (0,1); позволяет моделировать асимметричные и бимодальные распределения. Бета-регрессия предполагает, что зависимая переменная распределена по бета-распределению с плотностью вероятности, определяемой параметрами среднего и точности. Среднее значение связано с объясняющими переменными через функцию связи, например, логистическую функцию.

Например, в исследовании Jaber et al. (2021) [18] авторы использовали бета-регрессию с различными функциями связи для анализа доли потерь при дефолте в банковском секторе Иордании. Результаты показали, что probit-связь обеспечивает наилучшие прогнозы LGD с $R^2 = 52\%$. Макроэкономические переменные (ВВП и уровень инфляции) включены для оценки параметра среднего, в то время как характеристики клиента (политика ценообразования на услуги, операционная эффективность и коэффициент ликвидности) используются для оценки параметра точности.

Стоит учитывать, что бета-регрессия не может работать со значениями 0 и 1, может быть чувствительной к набору предикторов, оценка параметров модели может быть сложной при высоком количестве факторов, поэтому в исследованиях рекомендуется использование штрафных методов для стабилизации модели.

4) Гамма-регрессия

Гамма-регрессия подходит для моделирования распределений, сдвинутых вправо, например, LGD, где большинство значений близки к 0, но есть редкие наблюдения близкие к 1. Гамма-регрессия позволяет учитывать как среднее значение LGD, так и дисперсию, зависящие от объясняющих переменных.

В статье Sigrist, Stahel (2011) [41] предложена модель, основанная на цензурованном гамма-распределении, для прогнозирования доли потерь при дефолте в страховых облигациях. Модель предполагает, что латентная переменная, распределенная по гамма-распределению, трансформируется в наблюдаемую переменную с учетом ограничения интервала (по аналогии с Тобит регрессией).

Непараметрические модели

1) Деревья решений и деревья регрессий

Методы на основе деревьев разбивают исходную выборку на подмножества с помощью последовательных условий, что позволяет учесть нелинейность зависимости целевой переменной от риск-факторов и сохранить интерпретируемость результатов.

Например, Vlasenko (2022) [45] адаптировал дерево решений для моделирования доли потерь при дефолте в корпоративном сегменте. На первом уровне деление происходит по совокупной долговой нагрузке. На последующих уровнях участвует коэффициент финансовой устойчивости, коэффициент краткосрочной ликвидности и уровень задолженности. Ранговая корреляция

Спирмена реальных и прогнозных значений LGD составляет 49%, что указывает на умеренную тесноту связи.

В исследовании Qi and Zhao (2011) [39] проведено сравнение шести методов моделирования доли потерь при дефолте, включая параметрические и непараметрические подходы. Анализ проводился на данных Moody's Ultimate Recovery Database, содержащих 3 751 наблюдение по корпоративным дефолтам в США с задолженностью на момент дефолта более 50 млн долларов. Результаты показали, что дерево регрессии обеспечило наилучшую предсказательную способность среди непараметрических методов, а также превзошла параметрические.

Недостатком подхода на основе деревьев является проблема нестабильности и неустойчивости к выбросам, необходимость тщательной настройки гиперпараметров и дискретный набор оценок.

2) Алгоритмы ансамблей

Для моделирования доли потерь при дефолте часто используют следующие ансамблевые алгоритмы:

- Случайный лес: Использует ансамбль деревьев решений, обученных на случайных подвыборках наблюдений и признаков, что снижает переобучение и улучшает усредненный прогноз модели.

- Градиентный бустинг: Использует последовательное обучение деревьев, где каждое последующее дерево учитывает ошибки предыдущего.

Таким образом, непараметрические модели ансамблей способны достичь более высокого качества прогнозов по сравнению с одиночным деревом. При этом деревья решений и деревья регрессий являются интерпретируемыми в отличие от ансамблевых моделей, интерпретация которых затруднена.

В исследовании Miller and Tows (2017) [30] представлен многоэтапный подход с использованием случайных лесов при моделировании доли потерь при дефолте для лизинговых контрактов, который разделяет LGD на две составляющие: связанную с активами и прочую. Данные включают 1 493 дефолтных немецких лизинговых контракта. Результаты показали, что многоэтапный подход с использованием случайных лесов значительно улучшает точность прогнозов (MAE=6.7%) по сравнению с линейной регрессией OLS (MAE=27.6%) и позволяет более эффективно управлять процессом взыскания.

Kaposty et al. (2019) [22] в статье сравнивают различные подходы к моделированию и в результате приходят к выводу, что алгоритм случайного леса лучше всего справляется с задачей прогнозирования доли потерь при дефолте на данных по немецким лизинговым контрактам.

Исследование Bellotti et al. (2020) [5] направлено на прогнозирование коэффициентов восстановления по необслуживаемым кредитам базы данных европейского агентства по списанию долгов. Сравниваются 20 различных моделей оценки коэффициентов восстановления. Лучшие результаты демонстрируют ансамблевые алгоритмы на основе деревьев решений - случайные леса и бустинг.

3) Метод ближайших соседей

Метод ближайших соседей использует для прогнозирования LGD наблюдения из тренировочного набора данных, которые наиболее близки по метрике расстояния. Предсказание основывается на средних значениях LGD ближайших соседей.

Метод ближайших соседей применяется в статье Ptak-Chmielewska et al. (2023) [37] для прогнозирования доли потерь при дефолте, обеспечивая более высокую точность ($R^2 = 25\%$) на имеющихся данных по сравнению с другими подходами на выборке незавершенных дефолтов.

Серьезным недостатком метода ближайших соседей является необходимость при прогнозировании сравнивать каждое наблюдение со всеми объектами выборки разработки, что при условии работы с большими данными может стать долгим процессом, требующим больших вычислительных мощностей.

4) Регрессия опорных векторов

Модель регрессии опорных векторов использует ядра для преобразования данных в более высокое измерение, где разделение классов становится линейным. Она считается более стабильной, чем дерево регрессии, но страдает от необходимости настройки гиперпараметров, и не является интерпретируемой.

Исследование Yao et al. (2014) [46] представляет собой применение регрессии опорных векторов для моделирования доли потерь при дефолте корпоративных облигаций на базе данных Moody's Ultimate Recovery Database за 1985-2012 годы. Результаты показывают, что подход значительно превосходит традиционные модели в точности прогнозов.

В исследовании Nazemi et al. (2018) [34] использована модель SVR в нескольких вариациях для прогнозирования коэффициента восстановления корпоративных облигаций. Дополнительно для уменьшения размерности и повышения интерпретируемости данных применялись методы главных компонент. Использование SVR в сочетании с обширным набором макроэкономических переменных и факторов, специфичных для отрасли и инструмента, позволило добиться значительного улучшения точности прогнозирования по сравнению с методами линейной регрессии и дерева регрессий.

В статье Hurlin et al. (2018) [17] сравниваются шесть моделей для прогнозирования доли потерь при дефолте: регрессия дробного отклика, дерево решений, случайный лес, градиентный бустинг, нейронная сеть и регрессия опорных векторов. Эти модели тестировались на данных кредитных и лизинговых контрактов Международного банка. SVR продемонстрировала лучшие результаты с точки зрения метрик MSE и MAE, однако при использовании асимметричных функций потерь, учитывающих регуляторные капитальные требования, модель случайного леса показала большую устойчивость и точность.

5) Нейронные сети

Нейросети используют многослойную архитектуру, чтобы выявлять сложные зависимости и нелинейные паттерны в данных. В контексте моделирования доли потерь при дефолте нейронные сети часто встречаются в качестве одной из альтернатив в сравнительных исследованиях форм моделирования [5, 17, 22, 27]. Как правило, они уступают ансамблевым моделям, но превосходят или сопоставимы по точности с параметрическими моделями. Их применение ограничено сложностью интерпретации и необходимостью значительных вычислительных ресурсов.

Гибридные модели

1) Квантильная регрессия

Квантильная регрессия моделирует зависимости LGD для различных квантилей (5%, 50%, 95%) с использованием асимметричной функции потерь.

Somers, Whittaker (2007) [42] в исследовании указывают, что если обеспечение не покрывает стоимость дефолтного кредита, то банк несет убытки, поэтому оценка нижнего хвоста стоимости недвижимости имеет гораздо большее значение для оценки вероятных потерь, чем оценка среднего значения (при котором в большинстве случаев убытки не возникают). Авторы используют метод квантильной регрессии для прогнозирования распределения стоимости недвижимости, которое затем используется для расчета оценок потерь при дефолте по розничным ипотечным обеспеченным кредитам.

В исследовании Kruger and Rosch (2017) [24] авторы предложили использовать метод квантильной регрессии для анализа данных Global Credit Data по дефолтным кредитам малого и среднего бизнеса. Подход продемонстрировал превосходство над традиционными методами в прогнозировании доли потерь при дефолте. Значимыми риск-факторами модели стали тип обеспечения, старшинство долга и изменение S&P 500.

2) Сплайн-регрессия

Если квантильная регрессия позволяет исследовать различия во взаимосвязях между

объясняющими переменными на разных уровнях распределения целевой переменной, то сплайн-регрессия позволяет модели быть нелинейной через разделение на сегменты диапазона значений объясняющей переменной.

Например, в исследовании Miller (2017) [31] сплайн-регрессия используется для моделирования LGD лизинговых контрактов на основе следующих характеристик: размер долга, возраст актива, тип клиента и макроэкономические факторы.

3) Модель выживания Кокса

Можно моделировать процесс восстановления дефолтного кредита как процесс выживания с помощью полупараметрической модели Кокса.

В статье Ptak-Chmielewska et al. (2023) [37] модель пропорциональных рисков Кокса используется для прогнозирования доли потерь при дефолте с учетом временной динамики восстановления задолженности. Joubert et al. (2021) [21] представили адаптацию метода анализа выживания для моделирования LGD в рамках стандарта МСФО 9. Используя розничный портфель южноафриканского банка, авторы получили прогноз доли потерь при дефолте при различных макроэкономических сценариях и оценили ожидаемые кредитные потери.

4) Конечные смеси

Модели конечной смеси — это модели для описания сложных распределений данных, состоящих из нескольких кластеров. Они представляют собой комбинацию нескольких вероятностных распределений (компонент), каждая из которых описывает отдельный кластер или подгруппу в данных. Итоговое распределение — взвешенная сумма всех компонент.

В статье Baixauli, Alvarez (2010) [2] предложена модель смеси распределений для оценки LGD и расчета кредитного риска. Модель описывает распределение потерь как комбинацию нескольких бета-распределений, где каждое представляет группу с разным уровнем риска (низкий, средний, высокий).

5) Многоступенчатые модели

Активно применяются подходы, разделяющие сложную задачу на более простые подзадачи для обеспечения гибкости и точности прогнозов с сохранением большей интерпретируемости полученных результатов.

Например, в исследовании Tong et al. (2013) [44] модель состоит из двух компонентов: логистической регрессии для оценки вероятности нулевых потерь и гамма-регрессии для моделирования распределения суммы потерь при их наличии. Подход применен на дефолтных ипотечных кредитах в Великобритании. Модель показала лучшие результаты по сравнению с традиционными методами.

В работе Calabrese (2012) [9] представлена совместная аддитивная регрессия на основе бета-распределения, которая сочетает логистическую регрессию для предсказания крайних значений (0 и 1) и бета-регрессию для LGD в интервале (0, 1). Описанная структура позволяет добиться более высокой точности по сравнению с OLS и тобит-регрессиями.

Tanoue et al. (2017) [43] описали многоэтапную модель, которая применяется для прогнозирования доли потерь при дефолте с учетом специфики японских банковских кредитов. В модели используется логистическая регрессия на первых двух этапах: для оценки вероятности полного восстановления ($LGD=0$) и вероятности возникновения потерь ($LGD>0$). На третьем этапе применяется логит-преобразование для прогнозирования уровня потерь при дефолте ($LGD|LGD>0$) с последующим обратным преобразованием для получения значений в пределах $[0,1]$. Каждый этап использует разные наборы факторов и коэффициентов. Многоэтапная модель превзошла OLS и тобит- и бета регрессию ($R^2=32\%$).

В статье Gurtler and Hibbeln (2013) [15] авторы предложили двухэтапный подход для моделирования LGD клиентов немецкого банка. На первом этапе применяется логистическая регрессия для оценки вероятности, что дефолтный кредит завершится списанием, а не

восстановлением. Далее для каждого типа дефолта (списание/восстановление) строятся отдельные линейные регрессионные модели для прогнозирования уровня LGD.

В исследовании Yao et al. (2017) [47] предложен двухэтапный подход к моделированию доли потерь при дефолте, который использует методы опорных векторов. На первой стадии применяется регрессия опорных векторов для разделения случаев, когда коэффициенты восстановления равны 0 или 1, от остальных наблюдений. Затем на второй стадии применяется регрессионный метод, например, OLS или SVR, для прогнозирования значений LGD в диапазоне (0,1). Такой подход улучшает точность прогнозов по сравнению с традиционными методами, использующими логистическую регрессию в первой стадии.

Fan et al. (2023) [12] представили многоступенчатую модель, которая сочетает алгоритмы кластеризации на первом этапе, прогнозирование полных и частичных потерь с использованием машинного обучения на втором этапе и регрессию для оценки частичных потерь на третьем этапе. Авторы предлагают использовать метод SHAP для предоставления интерпретируемого объяснения влияния каждого фактора на прогнозы.

6) Слияние на основе нечеткой логики

Метод представляет собой интеграцию решений нескольких алгоритмов, где веса адаптивно определяются для каждого случая. Такой подход позволяет улучшить точность прогнозирования доли потерь при дефолте, учитывая неопределенность и разнородность данных.

В исследовании Nazemi (2017) [33] предложена модель прогнозирования потерь при дефолте корпоративных облигаций, основанная на слиянии решений с использованием нечеткой логики. Базовые модели объединяются через нечеткую систему правил. Для повышения точности прогнозов модель использует главные компоненты, выделенные из макроэкономических индикаторов, а также характеристики облигаций и отраслевые показатели.

Заключение

Моделирование доли потерь при дефолте (LGD) для оценки кредитного риска является сложной задачей, требующей внимания к научным изысканиям и регуляторным требованиям, выбора среди широкого спектра разнообразных подходов к разработке. Проведенный обзор литературы, а также опыт разработки моделей на практике позволяют сформулировать ключевые рекомендации по моделированию компонента LGD в рамках оценки кредитного риска.

1. Погоня за конкретной характеристикой качества модели, например, точностью, может негативно отражаться на других аспектах качества модели - дискриминационной способности, стабильности, интерпретируемости. Важно соблюсти баланс с учетом всех критериев и требований.

2. Ключевым требованием к моделям кредитного риска является интерпретируемость. Но это не означает, что использование "черных ящиков" для улучшения качества прогнозов вовсе недоступно. Они могут быть использованы либо в качестве ориентира, либо на предварительных этапах разработки для определения сложных нелинейных зависимостей в данных, которые затем можно преобразовать для включения в интерпретируемую форму модели.

3. Для успешного выполнения регуляторных требований банкам полезно ориентироваться на гибридные модели, которые способны учитывать сложное распределение целевой переменной и взаимосвязи между объясняющими переменными, сохраняя высокий уровень интерпретируемости.

4. Порой более простые методы построения модели на практике могут оказаться предпочтительными, даже если они уступают продвинутому решению, если взглянуть с точки зрения баланса между затраченными ресурсами и качеством прогнозов.

СПИСОК ЛИТЕРАТУРЫ

1. Baesens B., Smedts K. (2023). Boosting credit risk models. *The British Accounting Review*. DOI: 10.1016/j.bar.2023.101241
2. Baixauli S., Alvarez S. (2010). The Role of Market-Implied Severity Modeling for Credit VaR. *Annals of Economics and Finance*, Vol. 11, No. 2, pp. 337–353.
3. Bandyopadhyay A. (2022). Loan level loss given default (LGD) study of Indian banks. *IIMB Management Review*, Vol. 34, pp. 168–177. DOI: 10.1016/j.iimb.2022.06.003
4. Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards*.
5. Bellotti A., Brigo D., Gambetti P., Vrins F. (2020). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting*. DOI: 10.1016/j.ijforecast.2020.06.009
6. Bellotti T., Crook J. (2009). Loss Given Default models for UK retail credit cards. CRC working paper 09/1.
7. Bijak K., Thomas L.C. (2015). Modelling LGD for unsecured retail loans using Bayesian methods. *Journal of the Operational Research Society*, Vol. 66, pp. 342–352. DOI: 10.1057/jors.2014.9
8. Bonini S., de Carvalho G. (2016). Econometric Approach for Basel II Loss Given Default Estimation: from Discount Rate to Final Multivariate Model.
9. Calabrese R. (2012). Estimating bank loans loss given default by generalized additive models.
10. Dahlin F., Storkitt S. (2014). Estimation of Loss Given Default for Low Default Portfolios. *Royal Institute of Technology*.
11. Dermine O., de Carvalho C.N. (2006). Bank loan losses-given-default: A case study. *Journal of Banking & Finance*, Vol. 30, pp. 1219–1243. DOI: 10.1016/j.jbankfin.2005.05.005
12. Fan M., Wu T.-H., Zhao Q. (2023). Assessing the Loss Given Default of Bank Loans Using the Hybrid Algorithms Multi-Stage Model. *Systems*, Vol. 11, 505. DOI: 10.3390/systems11100505
13. Frontczak R., Jaeger M., Schumacher B. (2017). From Power Curves to Discriminative Power: Measuring Model Performance of LGD Models. *Journal of Mathematical Finance*, Vol. 7, pp. 657–670. DOI: 10.4236/jmf.2017.73034
14. Grunert J., Weber M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance*, Vol. 33, pp. 505–513. DOI: 10.1016/j.jbankfin.2008.09.002
15. Gurtler M., Hibbeln M. (2013). Improvements in loss given default forecasts for bank loans. *Journal of Banking & Finance*, Vol. 37, pp. 2354–2366. DOI: 10.1016/j.jbankfin.2013.01.031
16. Han C., Jang Y. (2012). Effects of debt collection practices on loss given default. *Journal of Banking and Finance*.
17. Hurlin C., Leymarie J., Patin A. (2018). Loss functions for Loss Given Default model comparison. *European Journal of Operational Research*. DOI: 10.1016/j.ejor.2018.01.020
18. Jaber J.J., Ismail N., Ramli S.N.M., Albadareen B., Hamadneh N.N. (2021). Estimating Loss Given Default Based on Beta Regression. *Computers, Materials & Continua*, Vol. 66, No. 3. DOI: 10.32604/cmc.2021.014509
19. Jacobs Jr. M. (2024). Modeling Ultimate Loss-Given-Default and Time-to-Resolution on Corporate Debt. *Journal of Financial Risk Management*, Vol. 13, pp. 426–459. DOI: 10.4236/jfrm.2024.132020
20. Jacobs M. (2010). An Option Theoretic Model for Ultimate Loss-Given-Default with Systematic Recovery Risk and Stochastic Returns on Defaulted Debt. *Forthcoming in the Proceedings of the 2010 3rd Annual Joint Bank for International Settlements, World Bank and European Central Bank Public Investors Conference*.
21. Joubert M., Verster T., Raubenheimer H., Schutte W.D. (2021). Adapting the Default Weighted Survival Analysis Modelling Approach to Model IFRS 9 LGD. *Risks*, Vol. 9, 103. DOI: 10.3390/risks9060103
22. Kaposty F., Kriebel J., Loderbusch M. (2019). Predicting loss given default in leasing: A closer look at models and variable selection. *International Journal of Forecasting*. DOI: 10.1016/j.ijforecast.2019.05.009

23. Kazianka H., Morgenbesser A., Nowak T. (2021). Assessing the discriminatory power of loss given default models. *Journal of Applied Statistics*. DOI: 10.1080/02664763.2021.1910936
24. Kruger S., Rosch D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking and Finance*, Vol. 79, pp. 42–56. DOI: 10.1016/j.jbankfin.2017.03.001
25. LaCour-Little M., Zhang Y. (2014). Default Probability and Loss Given Default for Home Equity Loans. *Economics Working Paper* 2014-1.
26. Leow M., Mues C. (2012). Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. *International Journal of Forecasting*, Vol. 28, pp. 183–195. DOI: 10.1016/j.ijforecast.2011.01.010
27. Li K., Zhou F., Li Z., Yao X., Zhang Y. (2021). Predicting loss given default using post-default information. *Knowledge-Based Systems*, Vol. 224, 107068. DOI: 10.1016/j.knosys.2021.107068
28. Loterman G., Brown I., Martens D., Mues C., Baesens B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, Vol. 28, pp. 161–170. DOI: 10.1016/j.ijforecast.2011.01.006
29. Maarse B. (2012). Master Thesis Backtesting Framework for PD, EAD and LGD. Rabobank International Quantitative Risk Analytics.
30. Miller P., Tows E. (2017). Loss Given Default Adjusted Workout Processes for Leases. *Journal of Banking and Finance*. DOI: 10.1016/j.jbankfin.2017.01.020
31. Miller P. (2017). Modeling and Estimating the Loss Given Default of Leasing Contracts.
32. Nazemi A., Fabozzi F.J. (2018). Macroeconomic variable selection for creditor recovery rates. *Journal of Banking and Finance*, Vol. 89, pp. 14–25. DOI: 10.1016/j.jbankfin.2018.01.006
33. Nazemi A., Fatemipour F., Heidenreich K., Fabozzi F.J. (2017). Fuzzy Decision Fusion Approach for Loss-Given-Default Modeling. *European Journal of Operational Research*. DOI: 10.1016/j.ejor.2017.04.008
34. Nazemi A., Heidenreich K., Fabozzi F.J. (2018). Improving corporate bond recovery rate prediction using multi-factor support vector regressions. *European Journal of Operational Research*. DOI: 10.1016/j.ejor.2018.05.024
35. Polianskii Iu. (2021). Problemy otsenki kachestva modelei PVR. *Sovremennye podkhody k validatsii modelei LGD. Risk-menedzhment v kreditnoi organizatsii*, № 3.
36. Polozhenie Banka Rossii 845-P 02.11.2024. (2024). O poriadke rascheta velichiny kreditnogo riska bankami s primeneniem bankovskikh metodik upravleniia kreditnym riskom i modelei kolichestvennoi otsenki kreditnogo riska.
37. Ptak-Chmielewska A., Kopciuszewski P., Matuszyk A. (2023). Application of the kNN-Based Method and Survival Approach in Estimating Loss Given Default for Unresolved Cases. *Risks*, Vol. 11, 42. DOI: 10.3390/risks11020042
38. Qi M., Yang X. (2009). Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Finance*, Vol. 33, pp. 788–799. DOI: 10.1016/j.jbankfin.2008.09.010
39. Qi M., Zhao X. (2011). Comparison of modeling methods for Loss Given Default. *Journal of Banking & Finance*, Vol. 35, pp. 2842–2855. DOI: 10.1016/j.jbankfin.2011.03.011
40. Schmit M., Stuyck J. (2002). Recovery Rates in the Leasing Industry.
41. Sigrist F., Stahel W.A. (2011). Using the censored gamma distribution for modeling fractional response variables with an application to loss given default. *Astin Bulletin*, Vol. 41, No. 2, pp. 673–710. DOI: 10.2143/AST.41.2.2136992
42. Somers M., Whittaker J. (2007). Quantile regression for modelling distributions of profit and loss. *European Journal of Operational Research*, Vol. 183, pp. 1477–1487. DOI: 10.1016/j.ejor.2006.08.063
43. Tanoue Y., Kawada A., Yamashita S. (2017). Forecasting loss given default of bank loans with multi-stage model. *International Journal of Forecasting*, Vol. 33, pp. 513–522. DOI: 10.1016/j.ijforecast.2016.11.005
44. Tong E.N.C., Mues C., Thoma L. (2013). A zero-adjusted gamma model for mortgage loan loss given default. *International Journal of Forecasting*, Vol. 29, pp. 548–562. DOI: 10.1016/j.ijforecast.2013.03.003

45. Vlasenko M. (2022). Postroenie modelei LGD dlia korporativnogo segmenta v ramkakh A-IRB-podkhoda s ispolzovaniem algoritma derevov reshenii. Bankauski viesnik, № 4 (705).
46. Yao X., Crook J., Andreeva G. (2014). Support vector regression for loss given default modelling. European Journal of Operational Research. DOI: 10.1016/j.ejor.2014.06.043
47. Yao X., Crook J., Andreeva G. (2017). Enhancing Two-Stage Modelling Methodology for Loss Given Default with Support Vector Machines. European Journal of Operational Research. DOI: 10.1016/j.ejor.2017.05.017

Methodological approach of loss given default modeling in credit risk assessment

Devaykina Anastasia Sergeevna

Postgraduate Student

Novosibirsk State University, Novosibirsk, Russian Federation

E-mail: i.nuriev@g.nsu.ru

Ibragimov Naimdzhon Mulaboevich

Doctor of Economic Sciences, Associate Professor

Novosibirsk State University, Novosibirsk, Russian Federation;

Institute of Economics and Industrial Engineering of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation;

Novosibirsk State Technical University, Novosibirsk, Russian Federation

E-mail: naimdjon.ibragimov@nsu.ru

KEYWORDS

credit risk, Loss Given Default, multimodal data distribution, interpretability, non-parametric models, hybrid models

ABSTRACT

The bank's loss given default (LGD) – the proportion of losses incurred when a borrower defaults – is a critically important metric. Without its accurate forecasting, it is impossible to correctly assess the level of credit risk, adequately form reserves for expected credit losses, or calculate the sufficient amount of capital required to safeguard the institution's financial stability against unexpected credit losses arising from loan defaults. This metric reflects the proportion of unpaid debt relative to its total amount and directly influences the level of provisioning, regulatory requirements, and strategic planning of banking operations. Predicting the losses given default requires considering several specific data properties, including multimodality, asymmetric distribution, constraints within the interval $[0,1]$, and nonlinear relationships with explanatory factors. This article reviews studies focused on modeling the share of default losses within the credit risk assessment. Model quality metrics are discussed based on four key criteria: accuracy, discriminative ability, stability, and interpretability. The article provides with recommendations for model development and implementation. Hybrid models that combine the advantages of parametric and non-parametric approaches, as well as interpretable non-parametric models, are becoming increasingly popular in credit risk assessment due to their ability to offer both predictive accuracy and interpretability. Nevertheless, parametric models remain the base analytical tool and benchmark against which the performance gains of more sophisticated models can be compared.
