

Scoring Day X

Свежие МЕТОДИКИ

Do scoring. Drive business





Сборник профессиональных материалов
для 10-й межотраслевой конференции

Scoring Day X

Содержание

Сергей АФАНАСЬЕВ, КБ «Ренессанс Кредит»

→ стр. 4

Анастасия СМЕРНОВА, КБ «Ренессанс Кредит»

Ильгиз АХМЕТСАФИН, КБ «Ренессанс Кредит»

Игорь МОЛОКАНОВ, КБ «Ренессанс Кредит»

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Банк России в Информационном письме от 18.03.2021 № ИН-03-36/14 рекомендовал кредитным организациям проводить оценку ожидаемых кредитных убытков в соответствии с лучшими международными и российскими практиками применения принципов МСФО (IFRS) 9. В статье описан процесс подготовки данных для разработки моделей LGD с учетом требований Базеля II и Положения № 483-П (ПВР).



Юрий ПОЛЯНСКИЙ, Банк России

→ стр. 24

Проблемы оценки качества моделей ПВР. Современные подходы к валидации моделей LGD

В ходе регуляторных ПВР-валидаций банков были неоднократно выявлены проблемы, связанные с оценкой прогностической силы моделей LGD. В статье приведено практическое статистическое доказательство правильного порядка расчета, конкретизированы и исследованы инструменты, применяемые при оценке дискриминационной силы и точности моделей LGD, даны рекомендации по оценке стабильности ПВР-моделей.



Владимир КОЗЛОВ, *raisk.ru*

→ стр. 39

Тотальный контроль и беззащитность: кого кредитовать в 2022 году

Розничное кредитование сопровождается существенными издержками на обработку больших данных. В системообразующих банках создаются озера данных, где хранится каждый клик пользователя. В банках поменьше аналитики принимают решения по нетиповым ссудам, используя в том числе общедоступную публичную информацию. Однако мы предлагаем обратить внимание и на иные затраты розничного банковского кредитования — операционные риски больших данных.



Михаил ПОМАЗАНОВ, ПАО «Промсвязьбанк»

→ стр. 49

Владимир ШИКИН, Национальное бюро кредитных историй (НБКИ)

Методика валидации эффективности риск-менеджмента розничного портфеля



В статье описаны методы количественной оценки качества дискриминации (одобрения) в сегментах розничного портфеля на основании текущих статистических данных об уровне отказа клиентам, обратившимся в банк, текущем уровне дефолтов и данных рынка (бюро кредитных историй). Проведен анализ экономической эффективности применяемого уровня одобрения заявок/отказа с учетом кредитного риска. На примере нескольких кейсов рассмотрены типичные причины слабости оценок качества дискриминации риск-процедур.



Дмитрий КУРЕННОЙ, ПАО «Промсвязьбанк»

→ стр. 60

Применение MATLAB Simulink для построения моделей оценки кредитного риска корпоративных заемщиков

В большинстве случаев методы моделирования, применяемые для оценки кредитного риска, не выходят за рамки классических методов машинного обучения. Однако спектр перспективных математических подходов и вычислительных инструментов может быть шире. В статье рассмотрены возможности пакета прикладных программ MATLAB для построения системно-динамических моделей, которые учитывают структуру компаний и могут применяться при обратном стресс-тестировании, а также при оценке вероятности дефолта.

Статьи и материалы в сборнике защищены Законом РФ «Об авторском праве и смежных правах». Перепечатка и распространение материалов запрещены без письменного согласия ООО «Регламент».

Банк России в Информационном письме от 18.03.2021 № ИН-03-36/14¹ рекомендовал кредитным организациям проводить оценку ожидаемых кредитных убытков в соответствии с лучшими международными и российскими практиками применения принципов МСФО (IFRS) 9. В статье описан процесс подготовки данных для разработки моделей LGD с учетом требований Базеля II и Положения № 483-П² (ПВР).

Сергей АФАНАСЬЕВ, КБ «Ренессанс Кредит», вице-президент, начальник управления статистического анализа

Анастасия СМЕРНОВА, КБ «Ренессанс Кредит», начальник отдела разработки и анализа эффективности скоринговых систем

Ильгиз АХМЕТСАФИН, КБ «Ренессанс Кредит», руководитель направления разработки рейтинговых систем

Игорь МОЛОКАНОВ, КБ «Ренессанс Кредит», главный аналитик направления разработки рейтинговых систем

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Процесс разработки и внедрения моделей в банке можно условно разбить на шесть фаз, установленных межотраслевым стандартом исследования данных CRISP-DM (рис. 1)³.

В этой части статьи мы разберем первые три фазы процесса разработки моделей LGD: постановку задачи, описание источников данных и подготовку выборки.

Постановка задачи

Статистические модели LGD позволяют давать количественную оценку кредитного риска, выраженную через прогнозируемый уровень потерь при дефолте. Без ограничения общности модели LGD могут применяться для различных целей:

¹ Информационное письмо Банка России от 18.03.2021 № ИН-03-36/14 «О рекомендациях по оценке ожидаемых кредитных убытков в соответствии с требованиями МСФО 9».

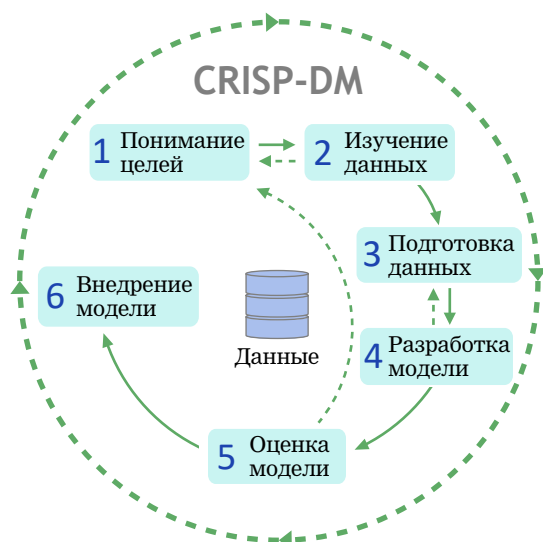
² Положение Банка России от 06.08.2015 № 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов».

³ Shearer C. The CRISP-DM model: the new blueprint for data mining // Journal of Data Warehousing. 2000. Vol. 5. P. 13-22.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Рисунок 1

Схема межотраслевого стандарта разработки моделей CRISP-DM



1) для оценки ожидаемых кредитных убытков (МСФО (IFRS) 9);
2) расчета нормативов достаточности капитала (ПБР, Базель II);
3) управления стратегиями взыскания просроченной задолженности¹.

Статистические модели LGD строятся на исторических данных, охватывающих кредиты с уже известными событиями дефолта и уровнем потерь. Таким образом, модели LGD разрабатываются с помощью статистических алгоритмов, относящихся к группе методов обучения «с учителем», когда по наблюдаемым кредитным требованиям известен уровень фактических потерь. Поскольку уровень потерь может быть разным, модели LGD строятся с помощью регрессионных алгоритмов и их модификаций.

Уровень потерь при дефолте можно оценивать для разных состояний кредитного требования, поэтому для целей ПБР и МСФО (IFRS) 9 традиционно строятся два типа моделей LGD:

1) LGD Will-Default — оценивает уровень потерь при дефолте кредитных требований, которые на момент применения модели не находятся в дефолте, но при этом исторически вышли в дефолт в течение 12 месяцев с отчетной даты наблюдения;

С учетом рекомендаций ПБР процесс построения модели LGD состоит из четырех этапов: разработка «ядра» модели, калибровка «ядра», оценка коэффициента спада, расчет надбавок консерватизма.

¹ В стратегиях взыскания просроченной задолженности используются модели прогноза показателя Recovery Rate, который определяется через фактический LGD по формуле: $RR = 1 - LGD$.

Сергей АФАНАСЬЕВ и др.

2) LGD In-Default — оценивает уровень потерь при дефолте кредитных требований, которые на момент применения модели уже вышли в дефолт.

С учетом рекомендаций ПБР процесс построения модели LGD состоит из четырех этапов:

- 1) разработка «ядра» модели — построение модели, ранжирующей кредитные требования в зависимости от их существенных характеристик, влияющих на ожидаемый уровень потерь при дефолте;
- 2) калибровка «ядра» модели — корректировка модельных значений для учета всей доступной информации, включая открытые дефолты, по которым невозможно однозначно определить значение целевой переменной;
- 3) оценка коэффициента спада — корректировка модельных значений для учета возможного ухудшения макроэкономических событий;
- 4) расчет надбавок консерватизма для учета неопределенности модельной оценки (модельного риска).

Изучение данных

Этот шаг критичен для предотвращения неожиданных проблем на протяжении следующей фазы (подготовки данных).

Поскольку модели LGD разрабатываются с помощью методов обучения «с учителем», данные разбивают на два типа:

- 1) данные о дефолтах — используются для определения целевой (зависимой) переменной;
- 2) характеристики заемщика/кредитного требования — используются для конструирования независимых признаков (факторов, переменных).

Данные о дефолтах

Модель LGD строится на основе фактической информации о событиях дефолта. Чтобы получить согласующиеся между собой результаты, необходимо использовать единое на всем периметре ПБР определение дефолта. Согласно Базелю II¹ и Положению № 483-П, дефолт считается реализовавшимся, когда произошло хотя бы одно из событий:

- должник более чем на 90 дней просрочил погашение любых существенных кредитных обязательств перед банком;
- банк считает, что должник не в состоянии полностью погасить кредитные обязательства без принятия банком таких мер, как

¹ Пункт 452 Базеля II.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

реализация обеспечения (если оно имеется). К индикаторам подобных событий могут относиться банкротство, смерть, реструктуризация, продажа долга и др.

Дефолт по розничным кредитам может определяться на уровне соответствующего инструмента, а не должника¹.

Данные для построения признаков

Для построения моделей LGD используется вся существенная, полная и доступная информация о заемщике и кредитном требовании. Для повышения качества моделей LGD банк может использовать внешние источники данных или внешние статистические модели при условии, что он может интерпретировать бизнес-смысл этих данных, а также понимает компоненты, использованные для построения внешних моделей.

Поскольку модели LGD применяются на всем жизненном цикле кредита, для их построения может использоваться разносторонняя апликативная и поведенческая информация о заемщике:

— социально-демографические характеристики: возраст, уровень образования, стаж на текущем месте работы, общий стаж, род деятельности, тип занятости, семейное положение, количество иждивенцев, наличие имущества и т.д.;

— информация о кредите: тип продукта, способ выплаты задолженности, продолжительность и величина кредитного требования, информация о залоге и пр.;

— внутренняя финансовая информация о заемщике: количество продуктов в банке, срок использования продуктов банка, суммарный доход, общая сумма задолженности и т.д.;

— информация о коммуникации банка с заемщиком: коммуникация с подразделениями службы взыскания, телемаркетинга и др.;

— данные из кредитных бюро: информация о количестве кредитов в других банках, просрочках по этим кредитам и пр.;

— поведенческая информация по кредиту: данные о совершенных платежах и просрочках на момент оценки кредита моделью LGD;

— данные из прочих внутренних и внешних источников: транзакции по картам и ДБО, скоринговые оценки операторов сотовой связи, данные из социальных сетей и т.п.

При работе с данными и отборе существенных признаков для построения моделей LGD необходимо проводить всесторонний анализ признаков на предмет их соответствия целям моделирования.

При отборе существенных признаков для построения моделей LGD необходимо исключать признаки, которые сильно зависят от внутренних процессов банка и по этой причине могут часто меняться и быть нестабильными.

¹ Пункт 455 Базеля II.

Сергей АФАНАСЬЕВ и др.

Например, необходимо исключать признаки, которые сильно зависят от внутренних процессов банка и по этой причине могут часто меняться и быть нестабильными.

Требования к данным

Для определения минимального объема выборки рекомендуется использовать статистическое правило “Number of Events Per Variable” (NEPV — количество событий на одну переменную). Согласно правилу NEPV, минимальный объем выборки для построения регрессионной модели определяется типом алгоритма и количеством признаков¹:

- для логистической регрессии на один признак (переменную) должно приходиться не менее 20 наблюдений. Например, если для построения модели используется 100 признаков, то минимальное количество наблюдений в выборке должно составлять 2000 (20×100);
- для дерева решений CHAID необходимо, чтобы на один признак приходилось не менее 50 наблюдений;
- для случайного леса, SVM и полносвязной нейронной сети на один признак должно приходиться не менее 200 наблюдений;
- для нестабильных алгоритмов, склонных к переобучению (таких как дерево решений CART), правило NEPV сформулировать сложно, поэтому необходим углубленный анализ статистической значимости и стабильности результатов работы модели.

Используемая для построения моделей LGD выборка должна включать в себя наиболее актуальную информацию. При этом в соответствии с требованиями Банка России период разработки моделей LGD должен составлять не менее пяти лет (по крайней мере для одного из источников), а также охватывать полный цикл деловой активности². В случае использования иного периода должна быть обоснована его достаточность.

В отношении кредитных требований к розничным заемщикам оценка уровня потерь при дефолте должна учитывать данные о реализованных потерях (отражающих условия экономического спада, если они выше средневзвешенного значения (по количеству наблюдений) уровня ожидаемых потерь за долгосрочный период, охватывающий полный цикл деловой активности) и оценки вероятности дефолта.

¹ Harrell F.E., Jr. Regression Modeling Strategies with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer, 2015.

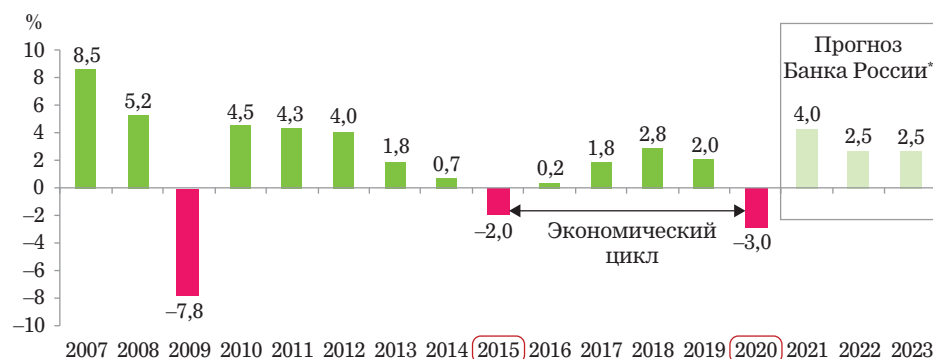
² Пункт 13.19 Положения № 483-П.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Длина экономического цикла определяется на основе статистического анализа, с помощью которого можно сделать выводы о макроэкономических условиях и характеристиках портфеля (рис. 2).

Рисунок 2

Пример определения экономического цикла по динамике темпов роста ВВП России



* Источник: interfax.ru/business/780390.

Обычно применяемая для целей моделирования продолжительность экономического цикла составляет 5–12 лет. При этом в соответствии с требованиями Положения № 483-П экономический цикл должен включать в себя как минимум период между двумя экстремумами экономических спадов.

При разработке «ядра» модели используются только те наблюдения, по которым наступило событие дефолта либо которые находились в дефолте на горизонте выборки. Так как по одному кредитному требованию на горизонте выборки может произойти несколько событий дефолта, статистической единицей для разработки модели является не кредитное требование, а событие дефолта на уровне кредитного требования.

Подготовка данных

Процесс подготовки данных может занимать большую часть времени проекта и включает в себя все основные этапы информационного цикла (рис. 3).

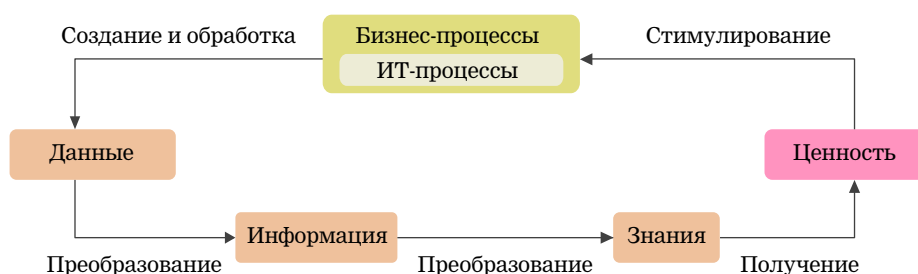
Сбор данных

К внутренним источникам данных относятся:

Сергей АФАНАСЬЕВ и др.

Рисунок 3

Информационный цикл создания и преобразования данных в банке (ISACA, COBIT 5)



- информационные системы, содержащие данные о кредитных заявках, в том числе отклоненных;
- информационные системы, содержащие данные о внутренней кредитной истории заемщика;
- информационные системы, содержащие данные о коммуникации с заемщиком (служба взыскания, телемаркетинг и др.);
- прочие информационные системы, содержащие данные о кредитах, заемщиках и кредитных заявках.

К внешним источникам относятся¹:

- бюро кредитных историй;
- официальные сервисы и базы данных государственных органов (ФМС, ПФР, ФССП, ФНС и др.);
- операторы сотовой связи, предоставляющие информационные сервисы для банков;
- и др.

Данные условно можно разделить на четыре группы:

- 1) книга дефолтов — содержит список всех используемых для построения модели статистических единиц, в который входят данные о дате открытия/закрытия события дефолта, причины открытия события дефолта, информация о связанных кредитных требованиях;
- 2) балансовая информация по кредитам, в том числе о величине задолженности, которая используется для расчета целевых значений;
- 3) информация о транзакциях, прошедших по счетам задолженности за период нахождения кредита в дефолте;

¹ Важным требованием к внешним источникам данных является возможность их валидации на стороне банка.

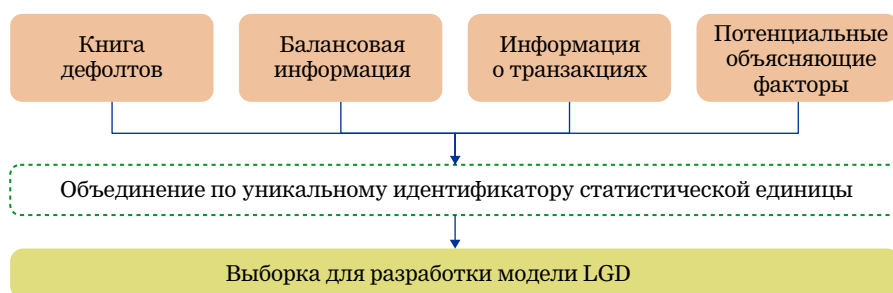
Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

4) потенциальные объясняющие факторы — расширенный список признаков, используемых для построения «ядра» модели.

При формировании выборки данные всех четырех групп объединяются по уникальному идентификатору статистических единиц и фильтруются для обеспечения соответствия периметру разработки и применения модели (рис. 4).

Рисунок 4

Схема подготовки выборки



Результатом сбора и обработки данных помимо подготовки выборки является формирование расширенного списка признаков, которые на фазе моделирования проходят этап аналитического отбора для попадания в финальную модель LGD. Все признаки, вошедшие в расширенный список, должны быть проанализированы на предмет доступности и качества данных, а также на соответствие бизнес-смыслу.

Агрегирование данных при реструктуризации

Для учета событий реструктуризации кредитов необходимо агрегировать данные по нескольким контрактам. В момент реструктуризации образуется новое кредитное требование, информация о котором используется для моделирования уровня потерь. Поскольку значения объясняющих факторов и денежные потоки учитываются до даты реструктуризации (дефолта) независимо в рамках нескольких контрактов, данную информацию необходимо привести к единой статистической единице:

- 1) всем контрактам, которые участвуют в реструктуризации (и родительским и дочерним), присваивается новый единый код дефолта;
- 2) дата начала дефолта по новому коду — это дата начала первого дефолта. Дата закрытия — это дата закрытия последнего дефолта (если событие не закрыто, дефолт остается открытым);

Сергей АФАНАСЬЕВ и др.

3) для разработки и калибровки модели (включая сбор потенциальных признаков) используется только новый дефолт с датой начала, равной первой дате дефолта.

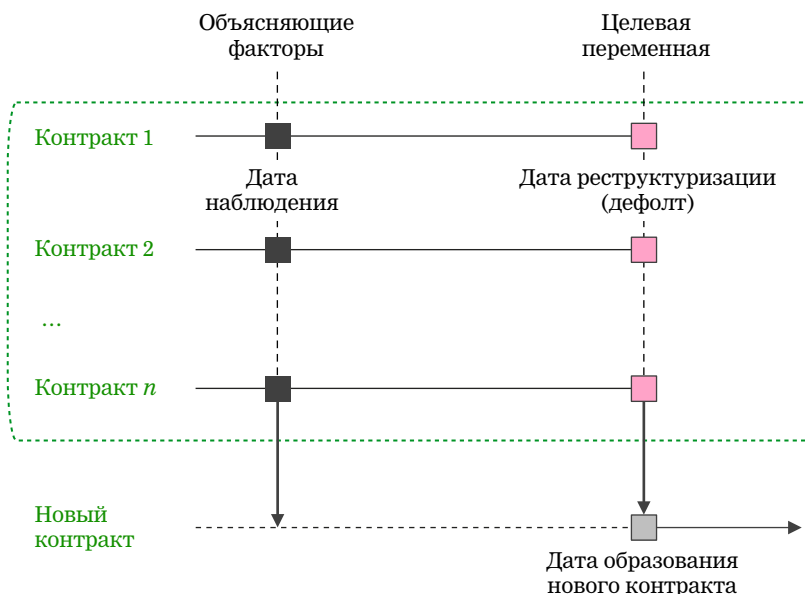
Для корректной оценки целевой переменной LGD при реструктуризации кредита необходимо агрегировать информацию по следующим сущностям:

- задолженность;
- уровень потерь;
- потенциальные объясняющие факторы (признаки).

На рис. 5 представлена схема агрегирования данных.

Рисунок 5

Схема агрегирования данных при реструктуризации кредитного требования



Агрегирование признаков на уровне кредита

Поскольку оценка LGD осуществляется на уровне кредитного требования, для корректного сбора потенциальных признаков необходимо агрегировать низкоуровневые признаки (например, объединять отдельные транзакции по кредиту в сумму платежей за период). Высокоуровневые признаки (город проживания, возраст заемщика и т.д.) не требуют дополнительной обработки.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Определение целевой переменной

Вне зависимости от функциональной формы модели для расчета значения LGD на уровне статистической единицы используется следующая целевая переменная:

$$LGD_{tot} = F(EAD, T(\cdot), f(\cdot), DC(\cdot), IC(\cdot), D(\cdot), MoC(\cdot)), \quad (1)$$

где $F(\cdot)$ — выбранная функциональная форма;

EAD — величина кредитного требования на момент дефолта (Exposure at Default);

$T(\cdot)$ — набор транзакций по каждой статистической единице, используемый для расчета целевой переменной¹;

$f(\cdot)$ — вектор объясняющих риск-факторов (признаков);

$DC(\cdot)$ — прямые дисконтированные расходы банка, отнесенные к конкретному кредитному требованию (Direct Costs);

$IC(\cdot)$ — косвенные недисконтированные расходы банка, направленные на сбор денежных средств, которые при этом не могут быть отнесены к конкретному кредитному требованию (Indirect Costs);

$D(\cdot)$ — вектор учета неблагоприятного эффекта от событий макроэкономического спада (Downturn);

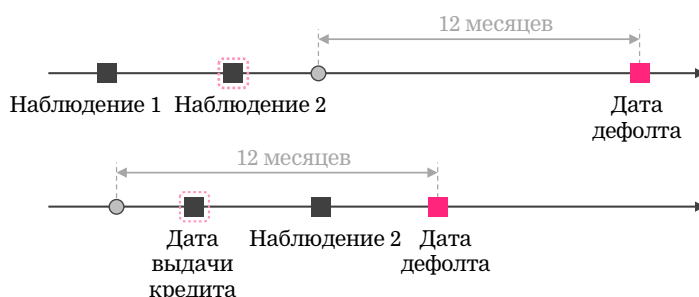
$MoC(\cdot)$ — надбавка консерватизма по результатам разработки модели (Margin of Conservatism).

Для моделей LGD Will-Default объясняющие факторы рекомендуется собирать одним из способов:

1) факторы собираются на дату за 12 месяцев до даты дефолта или наиболее раннюю дату, если дефолт произошел в течение 12 месяцев после выдачи (рис. 6);

Рисунок 6

Схема отбора наблюдений при дефолте, наступившем позднее года и ранее года после выдачи кредита



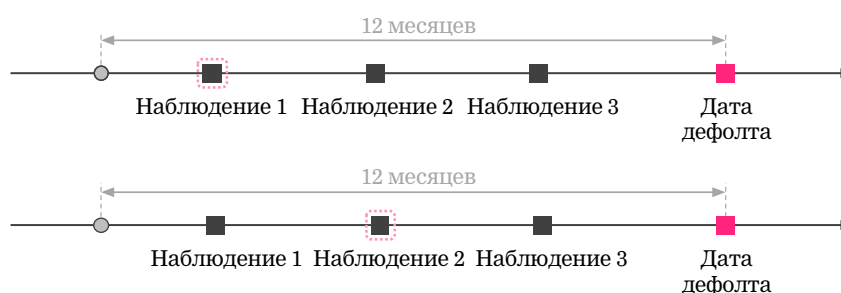
¹ Набор включает все доступные фактические транзакции начиная с момента дефолта.

Сергей АФАНАСЬЕВ и др.

2) собираются все доступные ежемесячные срезы на интервале от 3 до 12 месяцев до даты дефолта, а затем случайным образом отбирается один из срезов для разработки «ядра» модели (рис. 7). Если между датой выдачи кредита и началом дефолта меньше трех месяцев, то выбирается наиболее ранний срез после даты выдачи кредита.

Рисунок 7

Схема отбора случайного среза наблюдений



Модели LGD In-Default применяются к кредитам, которые уже находятся в дефолте. Поэтому при формировании целевой переменной для таких моделей необходимо оценивать будущие потери до момента закрытия дефолта. Для этого по каждому дефолту собирается несколько наблюдений после наступления события дефолта на определенные винтажи¹. Таким образом, количество наблюдений по каждому событию дефолта для моделей LGD In-Default определяется выбранной частотой винтажей и фактическим сроком до закрытия дефолта. Частота устанавливается аналитиком в зависимости от доступности данных и структуры портфеля.

На рис. 8 приведен пример сбора потенциальных признаков на определенные винтажи для различных статистических единиц до даты закрытия дефолта (данные по каждому контракту собираются либо на конец ближайшего месяца перед датой винтажа, либо на дату винтажа, если она приходится на конец месяца).

Определение наблюдений с фактом дефолта

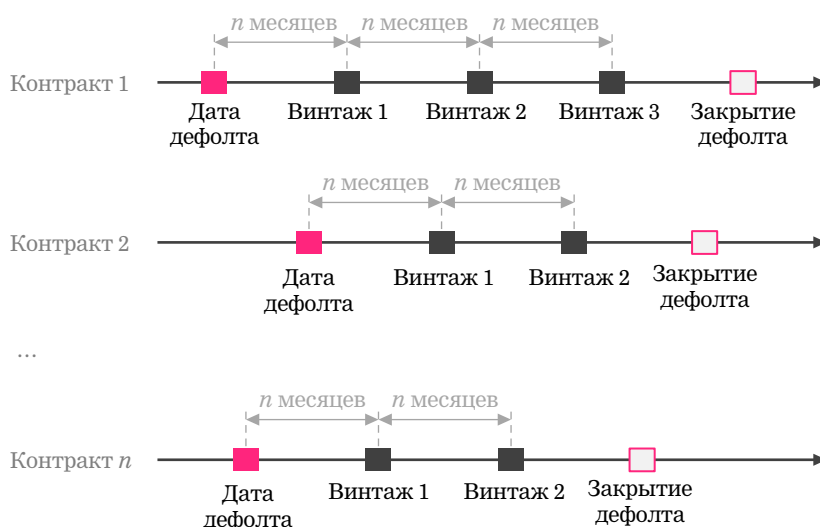
Для разработки моделей LGD используются все наблюдения, по которым произошло событие дефолта на горизонте выборки. Как

¹ Винтаж — это временной срез, определяемый временем нахождения кредитного требования в состоянии дефолта.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Рисунок 8

Пример определения винтажей для модели LGD In-Default

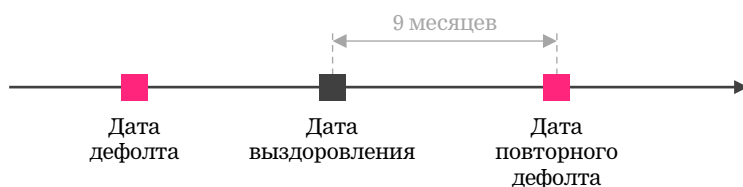


уже было сказано, по некоторым отобранным наблюдениям событие дефолта может наступать несколько раз. При этом в соответствии с рекомендациями ЕВА необходимо дополнительно анализировать время, прошедшее между окончанием одного дефолта и началом следующего¹.

Если для одного кредитного требования дефолты произошли на коротком промежутке времени, они объединяются в одно событие дефолта. Период, в течение которого происходит объединение дефолтов, обычно составляет 9–12 месяцев (рис. 9).

Рисунок 9

Схема объединения последовательных событий дефолта



¹ Guidelines on PD estimation, LGD estimation and treatment defaulted exposures. EBA, 20.11.2017. P. 6.1.1.

Сергей АФАНАСЬЕВ и др.

Для объединения событий дефолта выполняются следующие шаги:

- 1) проводится поиск дефолтов по контракту (или связанному контракту в случае реструктуризации);
- 2) рассчитывается расстояние между датой закрытия более раннего дефолта и датой открытия более позднего дефолта;
- 3) если расстояние между датами меньше 9 месяцев:
 - а) обоим дефолтам присваивается единый (новый) код дефолта;
 - б) дата начала дефолта по новому коду — это дата начала первого дефолта, дата закрытия — дата закрытия последнего дефолта (если событие не закрыто, дефолт остается открытым);
 - в) для разработки и калибровки модели (включая сбор потенциальных риск-факторов) используется только новый дефолт с датой начала, равной первой дате дефолта.

Расчет целевого значения

Расчет LGD для каждой статистической единицы осуществляется путем корректировки кредитного требования в момент дефолта на сумму транзакций после события дефолта:

$$LGD_i \cong 1 - \frac{\sum \text{tran}_i^{\text{disc}}}{EAD_i}, \quad (2)$$

где LGD_i — значение LGD для i -й статистической единицы;

$\sum \text{tran}_i^{\text{disc}}$ — сумма дисконтированных транзакций, связанных с денежными потоками в период нахождения кредита в дефолте, для i -й статистической единицы.

Для построения моделей обоих типов — LGD Will-Default и LGD In-Default — используются одни и те же статистические единицы. Однако для первой модели используются даты наблюдений до момента дефолта, для второй — даты наблюдений после дефолта.

Целевое значение LGD для кредитных требований «не в дефолте»

Целевое значение, используемое для разработки модели LGD Will-Default, рассчитывается по формуле:

$$LGD_i = 1 - \frac{\sum_{\text{DoD}}^{t=T} \text{tran}_{i,t}^{\text{disc}}}{EAD_{\text{DoD}}}, \quad (3)$$

где LGD_i — значение уровня потерь при дефолте для i -й статистической единицы;

DoD — дата дефолта (date of default);

Для построения моделей LGD Will-Default и LGD In-Default используются одни и те же статистические единицы. Однако для первой модели используются даты наблюдений до момента дефолта, для второй — даты наблюдений после дефолта.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

t — дата транзакции;
 T — дата последней фактической или смоделированной транзакции¹;
 $\sum_{DoD}^{t=T} \text{tran}_{i, DoD}^{\text{disc}}$ — сумма транзакций, дисконтированных к дате дефолта;
 EAD_{DoD} — размер задолженности для i -й статистической единицы на дату дефолта.

Если для расчета LGD используется CR-LR подход², то целевое значение рассчитывается по формуле:

$$LGD_i = CR_i \times LR_{i,c} + (1 - CR_i) \times LR_{i,t}, \quad (4)$$

где $LR_{i,c}$ и $LR_{i,t}$ — уровень потерь при дефолте для i -й статистической единицы для «выздоровевших» и не «выздоровевших» контрактов соответственно;

CR_i — вероятность «выздоровления» для i -й статистической единицы.

В зависимости от способа закрытия дефолта наблюдаемая компонента CR принимает значение 0 или 1, а компоненты $LR_{i,c}$ и $LR_{i,t}$ рассчитываются в соответствии с формулой (3).

Для построения модели LGD Will-Default необходимо использовать значения признаков, рассчитанные до наступления дефолта:

$$LGD_{tot,i} = f(X), \quad (5)$$

где $LGD_{tot,i}$ — значение LGD для недефолтных контрактов;

X — вектор признаков.

Целевое значение LGD для кредитных требований «в дефолте»

Для корректной оценки LGD по кредитным требованиям «в дефолте» необходимо учесть информацию о винтаже наблюдения (сроке нахождения в дефолте). В соответствии с требованиями Банка России переоценку LGD для дефолтных кредитов необходимо производить не реже чем раз в три месяца³.

В соответствии с описанной методикой период нахождения в дефолте разбивается на винтажи до момента закрытия дефолта (фактического или искусственного).

Для каждого винтажа до даты искусственного закрытия производится расчет целевой переменной по формуле:

¹ В случае калибровки по открытым дефолтам последней транзакцией считается последнее смоделированное погашение.

² Описание CR-LR подхода будет приведено во второй части статьи в следующем номере журнала.

³ Главы 13, 13.17 Положения № 483-П.

Сергей АФАНАСЬЕВ и др.

$$\text{LGD}_{i,V}^{\text{def}} = 1 - \frac{\sum_{t=V}^T \text{tran}_{i,t}^{\text{disc}}}{\text{EAD}_{i,V}}, \quad (6)$$

где $\text{LGD}_{i,V}^{\text{def}}$ — значение уровня потерь при дефолте для i -й статистической единицы на момент времени t ;

V — принимает значения винтажей нахождения в дефолте, на которых рассчитывается целевое значение;

T — дата последней фактической или смоделированной транзакции¹;

$\sum_{t=V}^T \text{tran}_{i,t}^{\text{disc}}$ — сумма транзакций, дисконтированных к дате винтажа;

$\text{EAD}_{i,V}$ — размер задолженности для i -й статистической единицы на винтаж V .

Если для расчета LGD используется CR-LR подход, то компоненты $\text{LR}_{i,c}$ и $\text{LR}_{i,t}$ рассчитываются по формуле (6), а для компоненты CR проставляется значение 0 или 1 в зависимости от факта «выздоровления» контракта.

Для модели, разрабатываемой на дефолтных наблюдениях, возможно:

1) построение новой модели для каждого нового винтажа — на каждый винтаж собираются потенциальные объясняющие признаки;

2) построение общей модели для всех винтажей — как и в первом варианте, объясняющие признаки на каждый винтаж собираются отдельно. В данном варианте в модель рекомендуется отдельно включать переменную «Винтаж наблюдения»;

3) построение новой модели для группы винтажей — данный вариант представляет собой гибрид двух перечисленных. Аналитик экспертно определяет временные интервалы, для которых будет применяться определенная модель. Затем, как и в первом варианте, для каждой группы винтажей строится отдельная модель (например, для винтажей 1–4 применяется модель $\text{LGD In-Default}_{1-4}$, а с 5-го винтажа применяется модель $\text{LGD In-Default}_{5+}$).

Для построения моделей LGD In-Default рекомендуется использовать признаки, рассчитанные на дату наблюдения:

$$\text{LGD}_{i,V}^{\text{def}} = f(X_V), \quad (7)$$

где $\text{LGD}_{i,V}^{\text{def}}$ — значение LGD для i -го контракта на винтаж t после даты наступления дефолта;

X_V — вектор признаков на винтаж V после даты наступления дефолта, собранных в соответствии с описанной методикой.

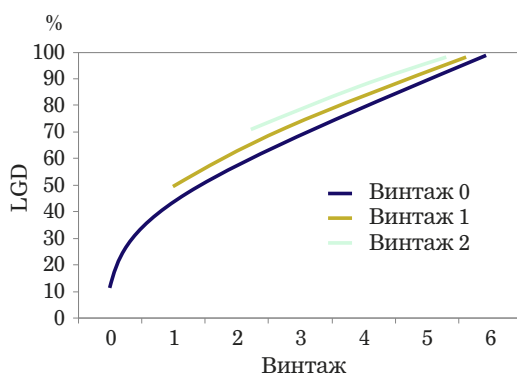
¹ В случае калибровки по открытым дефолтам последней транзакцией считается последнее смоделированное погашение.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Результат построения модели для кредитов «в дефолте» схематично представлен на рис. 10.

Рисунок 10

Иллюстрация модели для кредитных требований «в дефолте»



Дисконтирование денежных потоков

Чтобы учесть влияние времени на процесс возмещения долга, все денежные потоки для расчета целевой переменной необходимо дисконтировать по формуле:

$$\text{tran}_{i,t}^{\text{disc}} = \frac{\text{tran}_{i,t}}{(1+r)^{t-t_0}}, \quad (8)$$

где $\text{tran}_{i,t}$ — транзакция для i -й статистической единицы на момент времени t ;

r — ставка дисконтирования;

t — дата транзакции;

t_0 — дата, к которой производится дисконтирование (дата дефолта или дата отчетного наблюдения).

При разработке модели LGD можно использовать одну из следующих ставок дисконтирования:

1) безрисковую ставку + 5% (премия за дополнительный риск)¹. Для определения безрисковой ставки могут быть использованы государственные облигации (ГКО, ОФЗ), MIBOR/MIBID/MIACR/RUONIA или ключевая ставка Банка России;

2) кредитную ставку до наступления дефолта (включая первоначальную кредитную ставку);

¹ Guidelines on PD estimation, LGD estimation and treatment defaulted exposures. EBA, 20.11.2017. P. 6.3.1.3.

Сергей АФАНАСЬЕВ и др.

- 3) кредитную ставку после наступления дефолта;
- 4) эффективную процентную ставку.

Дисконтирование денежных потоков на основе безрисковой ставки + 5% позволяет учитывать стоимость фондирования на рынке в момент выхода заемщика в дефолт. Это более точный инструмент оценки стоимости денежных потоков в конкретных рыночных условиях в момент дефолта по сравнению с использованием процентной ставки по кредитному договору. Процентная ставка до наступления дефолта может существенно отставать от момента выхода в дефолт и быть оторвана от фазы экономического цикла и динамики процентных ставок.

Компоненты учитываемых денежных потоков

Целевое значение LGD рассчитывается путем учета денежных потоков, которые можно отнести к конкретному кредитному требованию.

Вне зависимости от методики построения и функциональной формы модели расчет целевого значения LGD должен учитывать следующие денежные потоки¹:

$$\text{tran} = \text{Principal} + \text{Interest}P + \text{Fee}P - \text{Principal}F + \text{Ceccion} - \text{DC} + \text{Balance}, \quad (9)$$

где tran — транзакции, учитываемые при расчете целевого значения;

Principal — платежи по основному долгу;

InterestP — платежи по процентам;

FeeP — платежи по штрафам;

PrincipalF — штрафы по основному долгу;

Ceccion — денежный поток в случае продажи кредита;

DC — прямые дисконтированные расходы банка, отнесенные к конкретному кредиту;

Balance — балансовое значение на момент «выздоровления» (в случае закрытия дефолта «выздоровлением»).

Выборка для разработки и калибровки модели

Результатом процесса сбора данных является выборка для разработки и калибровки модели, которая делится на подвыборки:

- обучающая выборка (train) — используется для обучения модели;
- тестовая выборка (test) — используется для проверки качества модели;
- выборка для калибровки — используется для приведения результатов моделирования к долгосрочному среднему значению.

Дисконтирование денежных потоков на основе безрисковой ставки + 5% — более точный инструмент оценки стоимости денежных потоков в конкретных рыночных условиях в момент дефолта по сравнению с подходом, при котором используется процентная ставка по кредитному договору.

¹ Guidelines on PD estimation, LGD estimation and treatment defaulted exposures. EBA, 20.11.2017. P. 6.3.1.1.

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

Обучающая и тестовая выборки

Обучающая и тестовая выборки включают в себя наблюдения, для которых доступна вся информация о реализации дефолта на рассматриваемом интервале времени. Наблюдения делятся между обучающей и тестовой выборками в соотношении от 60% : 40% для классического моделирования до 99% : 1% для глубоких нейронных сетей¹. Соотношение разбиения определяется аналитиком исходя из специфики задачи и характеристик выборки. Также в процессе обучения модели для настройки гиперпараметров может использоваться часть обучающей выборки для k -блочной перекрестной проверки.

Деление на обучающую и тестовую выборки осуществляется случайным образом либо с использованием стратифицированного отбора. В рамках метода стратифицированного отбора разделение выборки на подвыборки производится после классификации наблюдений по однородным группам (стратам). Случайный отбор наблюдений по стратам следует проводить таким образом, чтобы доля каждой группы совпадала с соответствующим показателем для исходной популяции. Чтобы разделить наблюдения по стратам, возможно использовать дату заявки, срок контракта, размер и тип задолженности и другие факторы.

При разделении выборок рекомендуется исключить возможность попадания кредитов одного клиента в разные выборки (обучающую и тестовую). Это необходимо, чтобы снизить зависимость между наблюдениями в разных выборках и, как следствие, снизить эффект переобучения модели.

Поскольку исключение данных может привести к существенным изменениям структуры выборки, рекомендуется анализировать репрезентативность полученной и исходной выборок относительно ключевых характеристик портфеля (уровень дефолта, социально-демографические факторы и т.д.). Анализ соответствия двух выборок можно проводить с помощью индекса стабильности популяции (PSI) или других статистических тестов.

Определение наблюдений, используемых для построения «ядра» и калибровки модели

Для построения «ядра» модели используются:

— закрытые дефолты — дефолты, по которым статус дефолта был снят в связи с «выздоровлением», полным погашением задолженности или списанием;

¹ Ng A. Neural Networks and Deep Learning (deeplearning.ai).

Сергей АФАНАСЬЕВ и др.

— открытые дефолты, которые можно искусственно признать закрытыми (Artificially Closed, AC). Это кредиты, находящиеся в дефолте (открытые дефолты) на последнюю дату выборки для разработки, по которым в будущем не ожидается материальных платежей, то есть значение целевой переменной может быть искусственно зафиксировано.

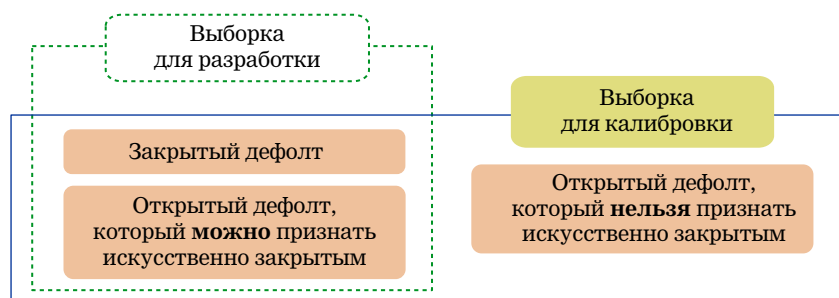
Для калибровки модели необходимо использовать всю доступную информацию:

- закрытые дефолты;
- все открытые дефолты, включая те, которые нельзя искусственно признать закрытыми.

Общая схема формирования выборок для разработки и калибровки модели LGD представлена на рис. 11.

Рисунок 11

Отбор наблюдений в выборки для разработки и калибровки



Для искусственного закрытия открытых дефолтов проводится количественный анализ динамики возмещений по закрытым дефолтам с целью определения периода эффективного сбора: срока нахождения в дефолте, после которого вероятность поступления материальных платежей по кредитному требованию незначительна.

Чтобы аналитически определить дату искусственного закрытия, после которой открытые контракты будут признаваться искусственно закрытыми, необходимо:

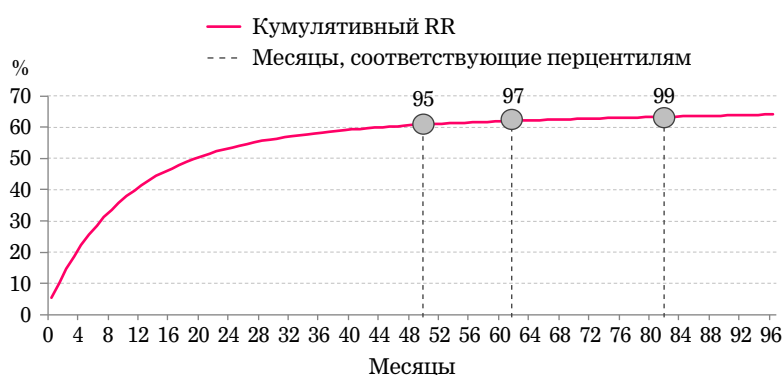
- 1) для закрытых дефолтов построить таблицу с кумулятивной долей погашения кредитных требований относительно всех погашений;
- 2) по всему портфелю определить срок, к которому будет выплачено не менее 95%/97%/99% от всех фактически выплаченных

Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных

платежей, используя дисконтированные или недисконтированные денежные потоки (рис. 12).

Рисунок 12

Анализ возмещений для определения даты искусственного закрытия



Если дефолт открыт дольше установленного периода, по кредитному требованию осуществляется искусственное закрытие дефолта, после чего наблюдение добавляется в выборку для обучения модели.

После успешного завершения этапа подготовки данных и формирования выборок можно переходить к следующей фазе цикла CRISP-DM — разработке моделей. О методах построения моделей LGD и сложностях, которые могут возникнуть на этой фазе, мы расскажем в следующей части статьи.

В ходе регуляторных ПБР-валидаций банков были неоднократно выявлены проблемы, связанные с оценкой прогностической силы моделей LGD. В статье приведено практическое статистическое доказательство правильного порядка расчета, конкретизированы и исследованы инструменты, применяемые при оценке дискриминационной силы и точности моделей LGD, даны рекомендации по оценке стабильности ПБР-моделей.

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD



Юрий ПОЛЯНСКИЙ,
*Банк России, начальник
отдела Департамента
банковского регулиро-
вания, ДВА, к.т.н.,
доцент*

Использование результатов внутренних оценок компонент кредитного риска в рамках БПБР¹ и ППБР² для расчета обязательных нормативов³ допускается, лишь когда банк подтвердит регулятору:

- соответствие методик ПБР общим научно-теоретическим подходам и современным банковским практикам;
- экономико-статистическую обоснованность методических, технологических и иных решений, принятых в ходе построения моделей ПБР;
- приемлемый уровень качества моделей ПБР в разрезе их ключевых характеристик;
- корректную и эффективную реализацию моделей ПБР в своих рейтинговых системах;
- соответствие специальным требованиям, установленным в нормативных актах Банка России.

Подтверждение приемлемого качества разработанных методик и моделей ПБР поэтапно осуществляется:

- 1) самими разработчиками в ходе процедуры внутреннего тестирования (верификация);

¹ Базовый подход на основе внутренних рейтингов.

² Продвинутый подход на основе внутренних рейтингов.

³ Basel III: Finalising post-crisis reforms. Bank for International Settlements, Basel Committee on Banking Supervision. December 2017; International Convergence of Capital Measurement and Capital Standards. Bank for International Settlements, Basel Committee on Banking Supervision. June 2006.

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

2) профильным подразделением банка, выполняющим углубленную внутреннюю оценку их качества после разработки и (или) усовершенствования (внутренняя валидация);

3) в рамках обязательной процедуры внешней оценки рейтинговых систем, официально выполняемой органом надзора для предоставления итогового разрешения на применение (регуляторная валидация).

Ключевые проблемы оценки качества моделей ПБР

Методология оценки компоненты PD (probability-of-default) неплохо разработана и достаточно успешно применяется передовыми банками. Как правило, наибольшую эффективность показывают методы моделирования PD, основанные на множественной logit-регрессии с ее последующим шкалированием на основе скоринговых алгоритмов.

Вместе с тем с началом разработки моделей оценки PD оказалось, что недостаточно просто построить ранжирующий модуль (ядро) модели, хорошо разделяющий наблюдения на 2 требуемые группы. Нужно также откалибровать алгоритм для его соответствия целевому значению — среднему уровню дефолтности (центральной тенденции), определенному на последнем историческом периоде: долгосрочном (through-the-cycle) или краткосрочном (point-in-time) в зависимости от принятой в банке калибровочной философии. Общий порядок такого «тюнинга» неплохо проработан банками и даже нормативно урегулирован¹.

Сложился набор показателей, позволяющих эффективно оценивать дискриминационную силу (ранжирующую способность) моделей PD: Gini (Accuracy Ratio, AR), AUROC (Area under Curve of Receiver Operating Characteristic), Колмогорова–Смирнова (K-S) и др.² Как правило, сейчас даже нет практической необходимости в совместном выполнении упомянутых количественных тестов в силу их тесной взаимосвязи и приемлемой надежности оценок и выводов.

Вместе с тем пока не разработаны действенные инструменты оценки другой характеристики модели PD — прогностической силы (точности), рассматриваемой как уровень соответствия прогноза и исторической центральной тенденции на выбранном уровне осреднения. Практически все повсеместно применяемые показатели и тесты

¹ Положение Банка России от 06.08.2015 № 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов».

² Studies on the validation of internal rating systems. Working Paper No. 14. Bank for International Settlements, Basel Committee on Banking Supervision. May 2005;
Битюцкий В., Патратий О. и др. Валидация. Ассоциация российских банков, Комитет по стандартам Базель II и управлению рисками. М., 2013.

Юрий ПОЛЯНСКИЙ

(биномиальный тест, тест Brier, тест Хосмера–Лемешова и др.) недостаточно эффективны по разным причинам, например:

- не способны выдавать однозначно и надежно трактуемый результат;
- как правило, не выдают конкретное число для его контроля в динамике с целью выявления проявившихся или накопившихся отклонений;
- не имеют устойчиво сложившихся критериев оценки и граничных значений;
- имеют частые ложные срабатывания (излишнюю консервативность);
- не имеют надежной статистической платформы (научного обоснования).

В такой ситуации для оценки модели в ходе процедур валидации банк и регулятор вынуждены не вычислять один надежный показатель, а выполнять несколько менее надежных и менее эффективных исследований, чтобы повысить общую достоверность итоговой оценки.

Аналогичные проблемы особенно характерны для моделей LGD (loss-given-default) и EAD (exposure-at-default). Здесь ситуация усугубляется отсутствием единых методических подходов к моделированию и эффективным инструментам оценки не только точности, но и дискриминационной силы моделей. Во многом это объясняется существенно иной (небинарной, непрерывной) природой их объясняемых переменных (LGD, CCF и пр.).

Оценка дискриминационной силы моделей LGD

Чтобы определить наиболее приемлемый вариант эффективной оценки дискриминационной силы моделей LGD, автором проведены количественные исследования, аналогичные ранее выполненным Bauke Maarse¹. Математическая формула показателя дискриминационной силы модели LGD теоретически обоснована² на основе площадей под реализованной (realized, R) и прогнозной (prognosed, P) кривыми Лоренца (L):

$$PR = \frac{\frac{1}{2} - \int_0^1 L^P(t)dt}{\frac{1}{2} - \int_0^1 L^R(t)dt} = \frac{\frac{1}{2} - AUC^P}{\frac{1}{2} - AUC^R}.$$

В указанной работе авторы предложили называть этот показатель Power Ratio (PR).

Из-за недостаточной эффективности инструментов оценки прогностической силы модели PD банк и регулятор вынуждены не вычислять один надежный показатель, а выполнять несколько менее надежных и менее эффективных исследований.

¹ Maarse B. Backtesting Framework for PD, EAD and LGD: Master Thesis. Rabobank International Quantitative Risk Analytics. July 16, 2012.

² Frontczak R., Jaeger M., Schumacher B. From Power Curves to Discriminative Power: Measuring Model Performance of LGD Models // Journal of Mathematical Finance. 2017. Vol. 7. P. 657-670.

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

Варианты расчета

Упорядочивание списка наблюдений и последующий расчет теоретически можно выполнять по разным столбцам используемого списка наблюдений:

- по уровням потерь при дефолте как долям от EAD (LGD^R и LGD^P);
- убыткам в стоимостном выражении, loss-at-default ($LAD^R = LGD^R \times EAD$ и $LAD^P = LGD^P \times EAD$).

Построение кривых и расчет площадей под ними банки осуществляют разными способами:

- а) обе кривые (реализованная и прогнозная) строятся по реализованным значениям (LGD^R или LAD^R);
- б) реализованная кривая строится по реализованным значениям (LGD^R или LAD^R), а прогнозная кривая — по прогнозным значениям LGD^P или LAD^P).

Расчеты для оси ОХ также встречаются в нескольких вариантах:

- а) доля (%) количества наблюдений от 1-го до текущего i -го (т.е. номер наблюдения) в общем количестве всех N наблюдений набора (от 1-го до последнего);
- б) доля (%) совокупного EAD от 1-го до текущего i -го наблюдения в совокупном EAD всех N наблюдений набора (от 1-го до последнего).

Возможно достаточно много комбинаций этих расчетных вариантов. Как показывают оценки, получаемые при этом результаты могут существенно отличаться, порождая не вполне корректные выводы о дискриминационной силе модели LGD. В ходе описываемых исследований рассмотрены наиболее адекватные варианты, перечисленные в табл. 1.

Чтобы выделить варианты, позволяющие наиболее точно оценить дискриминационную силу модели, задача рассмотрена в двух идеальных постановках, для которых значения показателя Powerstat заранее известны, что позволяет провести контрольные сравнения:

- задача 1 — ряд прогнозных значений LGD^P сформирован из реализованных значений LGD^R путем их представления в строго обратном порядке (идеальный Powerstat = -1);
- задача 2 — ряд прогнозных значений LGD^P сформирован из реализованных значений LGD^R путем их представления в случайном порядке (идеальный Powerstat = 0).

Для повышения надежности расчеты выполнены несколько (5–7) раз с последующим осреднением результатов.

Расчеты выполнялись в среде MS Excel на случайной выборке для розничного портфеля крупного коммерческого банка объемом 10 тыс. дефолтных наблюдений, по которым известны вызревшие значения LGD. Их отличительной особенностью является то, что в силу

Юрий ПОЛЯНСКИЙ

Таблица 1

Варианты расчета показателя Powerstat

№	По оси ОХ	По оси ОУ — доля (%) величины, совокупно накопленной от 1-го до i-го наблюдения, в ее совокупной сумме для всех (от 1-го до N-го) наблюдений			
		для факта (R)		для прогноза (P)	
		упорядочивание (по убыванию)	расчетная формула	упорядочивание (по убыванию)	расчетная формула
1	Доля (%) количества наблюдений от 1-го до текущего i-го в их общем количестве N: $X_i = i / N$	LGD_i^R	$Y_i = \frac{\sum_{k=1}^i LGD_k^R}{\sum_{k=1}^N LGD_k^R}$	LGD_i^P	$Y_i = \frac{\sum_{k=1}^i LGD_k^R}{\sum_{k=1}^N LGD_k^R}$
2		$LAD_i^R = LGD_i^R \times EAD_i$		$LAD_i^P = LGD_i^P \times EAD_i$	
3		LGD_i^R		LGD_i^P	
4		$LAD_i^R = LGD_i^R \times EAD_i$		$LAD_i^P = LGD_i^P \times EAD_i$	
5		LGD_i^R	$Y_i = \frac{\sum_{k=1}^i LAD_k^R}{\sum_{k=1}^N LAD_k^R}$	LGD_i^P	$Y_i = \frac{\sum_{k=1}^i LAD_k^R}{\sum_{k=1}^N LAD_k^R}$
6		$LAD_i^R = LGD_i^R \times EAD_i$		$LAD_i^P = LGD_i^P \times EAD_i$	
7		LGD_i^R		LGD_i^P	
8		$LAD_i^R = LGD_i^R \times EAD_i$		$LAD_i^P = LGD_i^P \times EAD_i$	

кредитной политики банка и функционирующих в нем процессов для значительной части дефолтного портфеля зафиксированы равные значения LGD^R (например, для 30% портфеля $LGD^R = 100\%$). Такая практическая особенность учтена далее при формулировке выводов.

Результаты

Результаты исследований отражены в табл. 2.

Таблица 2

Показатели Powerstat, полученные для контрольных задач

Варианты расчета	Порядок формирования прогнозных LGD из реализованных LGD	
	обратный (задача 1)	случайный (задача 2)
1	-0,94	0,03
2	-0,56	0,47
3	1,00	1,00
4	1,05	3,18
5	-0,92	-0,01
6	0,29	2,08
Контрольное значение	-1,00	0,00

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

Сравнение расчетных значений показателя Powerstat с контрольными значениями позволяет сделать следующие выводы:

а) расчеты по вариантам 3 и 4 (построение прогнозных кривых и расчет по прогнозным же значениям LGD или LAD) дали показатели Powerstat, максимально далекие от контрольных значений. Эти расчетные варианты полностью несостоятельны. Как для реализованных, так и для прогнозных кривых Лоренца необходимо вести расчеты по реализованным значениям LGD или LAD;

б) расчеты по вариантам 2, 4 и 6 (предварительное упорядочивание списка наблюдений по значениям LAD) также дали неудовлетворительные результаты. Это объясняется тем, что при умножении малых LGD на большие EAD получаются малопредсказуемые результаты (LAD), упорядочивание по которым снижает разделяющую силу алгоритмов;

в) расчеты по вариантам 1 и 5 (предварительное упорядочивание списка наблюдений по соответствующему LGD с расчетом по реализованным LGD или LAD) дали результаты, очень близкие к контрольному значению;

г) расчеты по варианту 5 показали повышенную нестабильность результатов, выражающуюся в повышенном разбросе значений показателя Powerstat в различных расчетных реализациях. Это связано с тем, что в различных расчетах наблюдения с равными реализованными LGD могут быть отсортированы по ним в разном порядке, что при умножении на их EAD может повлечь за собой некоторый разброс значений LAD и поэтому несколько изменять итоговые оценки;

д) целесообразно выполнять оценку как по варианту 1, так и по варианту 5 с консервативным итоговым выводом о точности модели.

Коэффициент ранговой корреляции

Для рассматриваемых случаев дополнительно выполнены расчеты коэффициента ранговой корреляции Спирмена (табл. 3), также применяющегося для оценки дискриминационной силы моделей LGD:

$$r_s = 1 - \frac{6}{N(N^2 - 1)} \sum_{i=1}^N (u_i - v_i)^2,$$

где u_i, v_i — ранги (номера по порядку) наблюдений при упорядочивании набора соответственно по реализованным и прогнозным LGD.

Анализ полученных результатов позволяет сделать следующие выводы:

а) коэффициент ранговой корреляции Спирмена пригоден для практической оценки дискриминационной силы моделей LGD как дополнительный инструмент исследований;

Юрий ПОЛЯНСКИЙ

Таблица 3

Коэффициент Спирмена, полученный для контрольных задач

Варианты расчета	Порядок формирования прогнозных LGD из реализованных LGD	
	обратный (задача 1)	случайный (задача 2)
1, 3, 5	-0,90	0,04
2, 4, 6	0,33	0,32
Контрольное значение	-1,00	0,00

б) подтверждается оптимальность вариантов 1 и 5;

в) применяемый в коэффициенте Спирмена расчет по рангам ведет к частичной потере точности оценки (в сравнении с Powerstat).

Оценка прогностической силы моделей LGD

Для оценки точности модели LGD в качестве одного из инструментов можно использовать показатель Expected Loss Shortfall (ELS):

$$ELS = 1 - \frac{\sum_{k=1}^N LGD_k^P \times EAD_k}{\sum_{k=1}^N LGD_k^R \times EAD_k}.$$

Он представляет собой взаимное соотношение прогнозного и реализованного совокупных убытков, вычисленных по всему набору данных, участвующему в оценке. Положительное значение ELS соответствует занижению прогноза по отношению к факту, то есть недооценке кредитного риска в разрезе совокупных убытков (красный сигнал светофора). Отрицательное значение свидетельствует о переоценке риска, что в рамках консерватизма обычно рассматривается как хороший результат (особенно для регулятора). Однако сильная переоценка также недопустима, уже для банка, поэтому обычно $ELS < -0,1$ (переоценка свыше 10%) рассматривается как желтый сигнал светофора, а $ELS < -0,2$ – как красный.

К сожалению, этот показатель сравнивает убытки лишь в целом по портфелю и не дает представления о взаимном соотношении положительных и отрицательных прогнозных ошибок (недооценок и переоценок), причем даже в целом по портфелю, не говоря уже об их сравнении в разрезе индивидуальных наблюдений. Условно говоря, если по одной части портфеля имеется существенная недооценка убытка, а по другой его части — равная переоценка, то эти отклонения взаимно компенсируют друг друга и дают идеальный итоговый

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

ELS = 0. Представляется, что в таком случае результат не может безоговорочно рассматриваться как хороший.

Чтобы получить представление о том, насколько в целом прогнозные значения далеки от фактических, можно оценить точность прогноза с применением показателя Mean Absolute Error (MAE), оценивающего среднюю абсолютную ошибку, в следующих его вариантах:

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{k=1}^N |\text{LGD}_k^R - \text{LGD}_k^P|, \\ \text{MAE}' &= \frac{\sum_{k=1}^N |\text{LGD}_k^R - \text{LGD}_k^P| \times \text{EAD}_k}{\sum_{k=1}^N \text{EAD}_k}. \end{aligned}$$

Невысокий $\text{MAE} < 0,1$ соответствует зеленому сигналу светофора, повышенный $0,1 \leq \text{MAE} \leq 0,2$ — желтому, высокий $\text{MAE} > 0,2$ — красному.

MAE дает представление о совокупном разбросе прогноза относительно факта, но все же без учета взаимного соотношения положительных и отрицательных ошибок.

Применяются и другие меры точности прогноза, например Mean Absolute Percentage Error (MAPE) — средняя абсолютная процентная ошибка:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|\text{LGD}_k^R - \text{LGD}_k^P|}{|\text{LGD}_k^R|}.$$

Однако, как отмечают многие исследователи¹, этот показатель, несмотря на его простоту и убедительность, демонстрирует серьезные недостатки при практическом применении. Например, он может давать нестабильные результаты, особенно в районе малых значений LGD_k^R .

Вывод: для итоговой оценки точности модели целесообразно рассматривать указанные (и иные) показатели совокупно, так как они дополняют друг друга. В этом и состоит один из аспектов проблемы оценки качества моделей LGD (а также моделей CCF, близких по сути).

Оценка стабильности моделей ПБР

Наряду с упомянутыми базовыми характеристиками моделей важна их стабильность, а также стабильность получаемого результата как по пространству (для разных выборок данных, разных частей

Показатель Expected Loss Shortfall сравнивает убытки лишь в целом по портфелю и не дает представления о взаимном соотношении положительных и отрицательных прогнозных ошибок (недооценок и переоценок).

¹ Hyndman R.J., Koehler A.B. Another look at measures of forecast accuracy // International Journal of Forecasting. 2006. Vol. 22. No. 4. P. 679-688;
Kim S., Kim H. A new metric of absolute percentage error for intermittent demand forecasts // International Journal of Forecasting. 2016. Vol. 32. No. 3. P. 669-679.

Юрий ПОЛЯНСКИЙ

портфеля), так и по времени (для разных исторических моментов и (или) периодов). Поэтому значимой частью оценки качества модели является исследование указанных аспектов ее стабильности, то есть способности сохранять структурно-функциональную целостность и требуемую эффективность в заданных рабочих границах применимости.

Вместе с тем, как показывает практика, банк может рассматривать стабильность в отношении не столько самой модели, ее факторов и результата, сколько в отношении портфеля, к которому применяется модель (т.е. как неизменность структуры сегмента ее применения). Для таких исследований зачастую ошибочно полагается вполне достаточным простое вычисление показателя PSI (Population Stability Index) в разрезе факторов модели:

$$PSI = IV = \sum_{i=1}^n IV_i = \sum_{i=1}^n \left\{ (p_{attr\ i}^{good} - p_{attr\ i}^{bad}) \times WOE_{attr\ i} \right\},$$

где $WOE_{attr} = \log_{10} \frac{p_{attr}^{good}}{p_{attr}^{bad}} = \log_{10} \left(\frac{n^{good}}{n^{bad}} \right) - \log_{10} \left(\frac{N^{good}}{N^{bad}} \right)$ — вес фактора Weight-of-Evidence, численно оценивающий качество разбиения им множества наблюдений на 2 выделяемые группы (например, на «хороших» и «плохих»);

$p_{attr}^{good} = \frac{n_{attr}^{good}}{N^{good}}$ — доля «хороших» наблюдений с данным атрибутом (n_{attr}^{good})

среди всех «хороших» наблюдений (N^{good});

$p_{attr}^{bad} = \frac{n_{attr}^{bad}}{N^{bad}}$ — доля «плохих» наблюдений с данным атрибутом (n_{attr}^{bad}) среди

всех «плохих» наблюдений (N^{bad}).

Стабильность структуры портфеля — несомненно, его важная характеристика. И все же это несколько иное свойство, относящееся к иной сущности. Стабильность модели вполне можно исследовать путем расчета упомянутых выше базовых характеристик применительно к разным частям портфеля (в т.ч. случайным выборкам) и разным периодам (или моментам), в том числе рассматривая их в виде динамических рядов, для которых есть самостоятельная продвинутая методология исследований.

Как показывает практика регуляторной валидации моделей ПВР в передовых российских банках, именно стабильность моделей труднее всего обеспечить и надежно оценить. В настоящее время эта проблема — одна из ключевых при разработке и внедрении методик и моделей ПВР.

Существенное влияние на стабильность моделей ПВР оказывают характер российской экономики, изменчивость кредитной политики банка, динамичность его внутренних процессов, качество его данных

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

и информационных систем. Подобные факторы можно попробовать учесть непосредственно в структуре модели, что находит отражение в так называемой **рейтинговой философии** банка. В зависимости от своих возможностей и планов банк решает, насколько ядро модели (т.е. ее ранжирующий модуль) будет чувствительно к изменениям экономической среды и технологического окружения, будут ли такие изменения учтены в составе методов и факторов, использованных в рейтинговой модели.

Как правило, российские банки выбирают наиболее простую ранжирующую философию ТТС (through-the-cycle), подразумевающую нечувствительность ядра модели к изменениям во внешней и внутренней среде банка, его нацеленность прежде всего на факторы, непосредственно характеризующие лишь профиль заемщика и (или) характеристики ссуды. Однако, как показали количественные исследования ЕВА¹, не менее трети европейских ПБР-банков уже внедрили продвинутые модели, разработанные по ранжирующей философии РИТ (point-in-time), то есть в той или иной степени чувствительные к внешнеэкономическим условиям.

Другим аспектом проблемы стабильности модели является выбранная при ее разработке **калибровочная философия**, определяющая цель и общий порядок калибровки (перекалибровки) модели, включая подходы к формированию структуры и длительности периодов калибровочной выборки. В российском банковском регулировании² пока нормативно определен лишь единственный тип калибровки — к долгосрочному среднему значению оцениваемой компоненты, то есть выполняемый по калибровочной философии ТТС. Вместе с тем некоторая часть (не более 20%) европейских ПБР-банков предпочли калибровку РИТ. Эти два типовых подхода фактически различаются лишь по длительности последнего исторического периода, используемого для расчета центральной тенденции, к которой приводится модель и отклонения от которой затем контролируются.

Для практического исследования стабильности моделей важно рассматривать поведение ключевых показателей качества в их динамике, то есть формировать классический временной ряд из последовательно наблюдаемых значений. При этом необходимо, чтобы эти значения были получены в примерно одинаковых условиях при

Как показывает практика регуляторной валидации моделей ПБР в передовых российских банках, труднее всего обеспечить и надежно оценить стабильность моделей ПБР.

¹ EBA report on IRB modelling practices. Impact assessment for the GLs on PD, LGD and the treatment of defaulted exposures based on the IRB survey results. European Banking Authority. November 20, 2017.

² Положение Банка России от 06.08.2015 № 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов».

Юрий ПОЛЯНСКИЙ

одинаковых параметрах расчета. Критериями оценки в таких случаях могут быть не только достигнутые значения рассматриваемого показателя (как обычной статической величины), но и его абсолютное и относительное изменения при применении к разным наборам или периодам (например, ежеквартальные темпы роста/падения).

Дополнительные характеристики моделей ПВР

Как правило, модели ПВР разрабатываются для оценки настоящего и (или) будущего состояния системы при ее функционировании в относительно стабильном режиме. Обычно они не учитывают пошаговую динамику изменения этого состояния под воздействием внешних факторов. Вместе с тем в реальности модели работают в крайне динамичной среде.

Поэтому важным направлением оценки качества моделей ПВР должно стать исследование их дополнительных характеристик (в т.ч. динамических¹), например таких, как:

- устойчивость — способность модели сохранять структурно-функциональную целостность и эффективность при резких и существенных (стрессовых, шоковых, кризисных) изменениях внешних и внутренних факторов;

- чувствительность — степень изменчивости результата под воздействием внешних и внутренних факторов различной направленности и интенсивности;

- сходимость — предсказуемость, логичность, непрерывность результатов прогноза, их стремление к некоему постоянному значению, пределу при постепенном изменении ключевых параметров модели (например, при расширении расчетного периода в прошлое).

На рис. 1 и 2 показаны изменения значений показателя ELS для моделей LGD Longrun (для ссуд, не находящихся в дефолте) и LGD In-default (для ссуд, находящихся в дефолте) при постепенном расширении периода оценки от расчетной даты вглубь истории. Видна существенная зависимость ELS от глубины учитываемой исторической статистики. На исторической глубине более 5–6 лет начинают наблюдаться снижение изменчивости оценок, их сходимости к некоторому долгосрочному уровню. Не случайно в соответствии с Базельскими рекомендациями для оценки качества моделей ПВР по различным сегментам Банк России требует использовать всю доступную

¹ Звонарёв С.В. Основы математического моделирования: Учебное пособие. Екатеринбург: Уральский федеральный университет, 2019;

Ермолова М.Д., Пенкас Г.И., Полянский Ю.Н. Исследование влияния модельного риска на точность оценок риск-взвешенных активов, полученных с помощью подхода на основе внутренних рейтингов // Управление финансовыми рисками. 2019. № 1.

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

Рисунок 1

Сходимость показателя ELS для модели LGD LR на горизонте оценки, расширяющемся вглубь истории

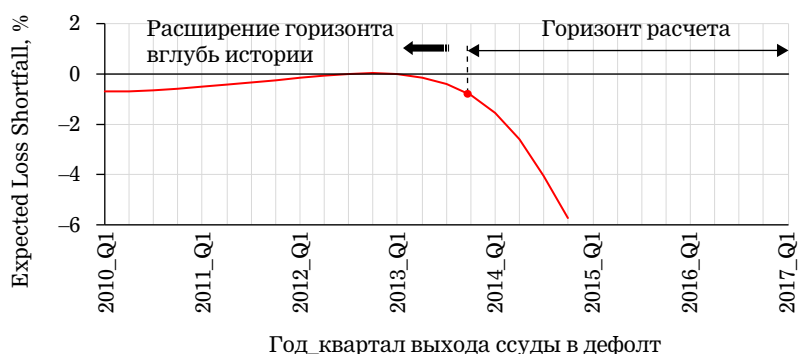
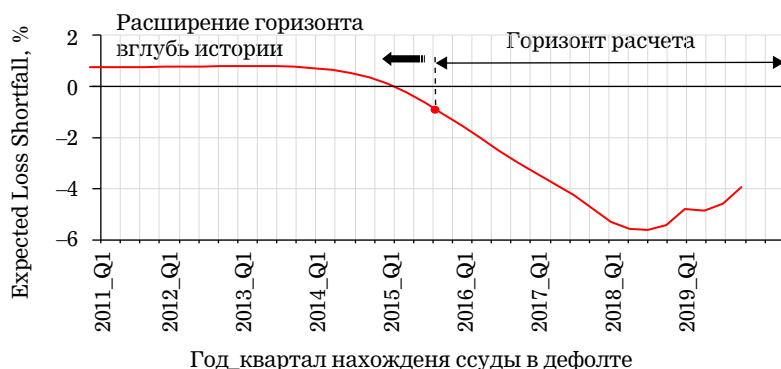


Рисунок 2

Сходимость показателя ELS для модели LGD ID на горизонте оценки, расширяющемся вглубь истории



банку существенную статистику, релевантную оцениваемому текущему кредитному портфелю, за исторический период, включающий экономический спад и охватывающий полный цикл деловой активности, но не менее 5 лет.

Однако подобная сходимость рассматриваемого показателя при расширении периода наблюдений имеет место не всегда. Это зависит от исторической статистики конкретного портфеля, а также от качества моделей.

Повышенная зависимость характеристик моделей (особенно точности) от выбранного периода исторической статистики и иных

Юрий ПОЛЯНСКИЙ

расчетных особенностей делает актуальной проблему устойчивости модели к резким и значительным (т.е. шоковым) изменениям во внешних и (или) внутренних условиях, а также к возможным скачкам иных параметров (например, качества и (или) состава данных), то есть робастности модели.

В настоящее время назрела необходимость в разработке методологии стресс-тестирования моделей ПВР и его регулярном проведении для определения допустимых границ применимости моделей, а также для мониторинга возможных нарушений этих границ.

Выбор параметров для расчета показателей

Как показали наблюдения автора в ходе процедур валидации рейтинговых систем передовых банков для получения разрешений на применение ПВР в целях расчета достаточности капитала, на значения показателей качества могут существенно влиять:

- выбранные для оценки исторические периоды, в том числе их глубина;
- порядок обработки особенных наблюдений (например, пропусков, выбросов, ошибок, нехарактерных наблюдений и иных отклонений);
- способ отбора наблюдений в оценочную выборку, в том числе их тип (ссуда, ссуда-на-дату, клиент и т.п.);
- тип прогнозных значений, используемых для оценки (например, после какого этапа моделирования результат участвует в оценке).

Визуальный анализ качества моделей ПВР

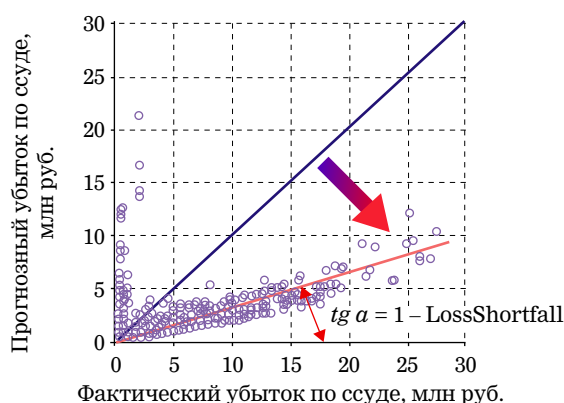
С учетом недостаточной эффективности некоторых показателей, не вполне устоявшихся правил и критериев оценки достаточно информативным может оказаться визуальный анализ с использованием продвинутых графиков. В ряде случаев наглядная картинка гораздо более убедительна для практики, чем некое число, алгоритм получения которого не вполне прозрачен, а критерии оценки и граничные значения могут быть спорными.

На рис. 3 показано облако точек, соответствующих парам значений индивидуальных убытков (реализованному и прогнозному), которое наглядно демонстрирует существенное занижение прогноза относительно факта для ссуд с суммой задолженности свыше 2–3 млн руб. (подавляющая часть портфеля). При этом для ссуд с малым остатком долга наблюдается значительное завышение. Как отмечено на графике, показатель Expected Loss Shortfall лишь подтверждает этот вывод.

Проблемы оценки качества моделей ПБР. Современные подходы к валидации моделей LGD

Рисунок 3

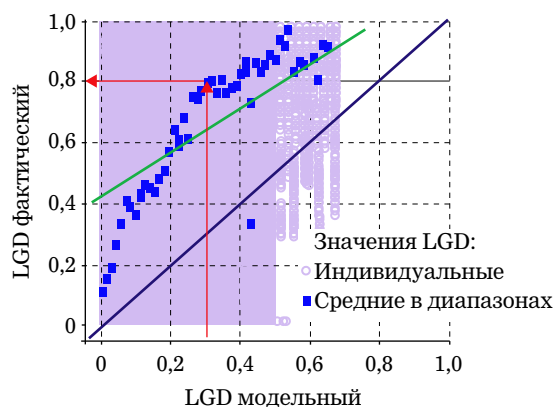
Распределение индивидуальных убытков в стоимостном выражении



Если визуализировать (рис. 4) распределение не индивидуальных убытков (в стоимостном выражении), а индивидуальных LGD (в процентах), может выявиться их существенный разброс по плоскости графика (множество круглых точек). Базельский комитет по банковскому надзору, предупреждая о подобных разбросах прогнозов индивидуальных LGD для розничных кредитных требований, рекомендует в ст. 331 (121) Базеля II/III оценивать компоненты кредитного риска не на индивидуальном, а на агрегированном уровне

Рисунок 4

Распределение LGD



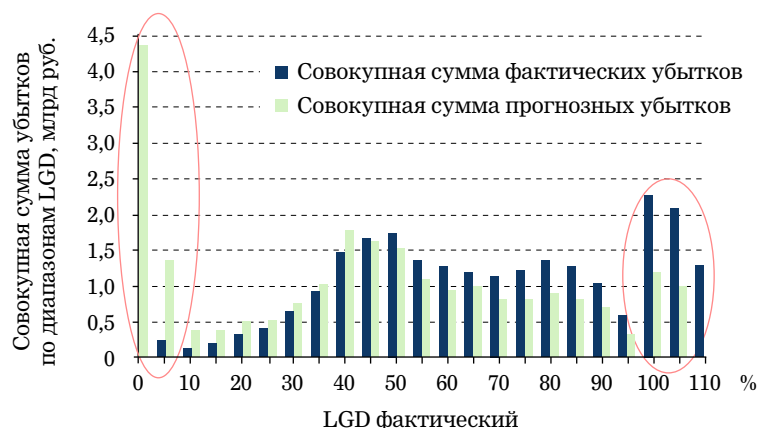
Юрий ПОЛЯНСКИЙ

(с осреднением прогноза по пулам). Такое осреднение индивидуальных LGD по их диапазонам позволяет рассмотреть картину распределения средних значений (квадратные точки на рис. 4), которая информативно иллюстрирует общее смещение прогнозов.

Качество прогнозных оценок можно также наглядно визуализировать в виде сравнения распределений фактических и прогнозных совокупных убытков по ссудам в разрезе перцентилей их фактических LGD (рис. 5).

Рисунок 5

Распределение совокупных сумм убытков



Таким образом, визуальный анализ качества моделей ПВР можно рассматривать как достаточно эффективный инструмент практических исследований в ходе тестирования и валидации.

Розничное кредитование сопровождается существенными издержками на обработку больших данных. В системообразующих банках создаются озера данных, где хранится каждый клик пользователя. В банках поменьше аналитики принимают решения по нетиповым ссудам, используя в том числе общедоступную публичную информацию. Однако мы предлагаем обратить внимание и на иные затраты розничного банковского кредитования — операционные риски больших данных.

Тотальный контроль и беззащитность: кого кредитовать в 2022 году

Что сейчас представляет собой процесс розничного кредитования? Конвейер, скоринг, запросы официальных документов в СМЭВ или иных подсистемах, получение согласий и обработка информации, далее принятие кредитного решения и процессинг. Мы привыкли к тому, что, задав вопрос о физическом лице или индивидуальном предпринимателе, получим корректный ответ. Это удобно, и это часть нашего бизнес-процесса. Но давайте посмотрим, как и с какими (даже теоретическими) издержками она реализована. Наш клиент — ради заботы о потребностях которого создаются департаменты и придумываются пайплайны — практически полностью «раздет» с точки зрения операционных рисков и кибербезопасности.

В процессе кредитования так или иначе задействован ряд государственных или квазигосударственных структур. Ими могут быть:

- МВД;
- ФМС;
- ФНС;
- ГИБДД;
- нотариат.

Эти ведомства уверенно идут по пути цифровизации — и открывают максимальный объем информации любому желающему. Давайте вспомним, что несколько лет назад не было такого количества спам-



Владимир КОЗЛОВ,
*FRM, консультант
по риск-менеджменту,
raisk.ru*

Владимир КОЗЛОВ

звонков. С чем связана такая активность? Может быть, с тем, что Роскомнадзор разместил в открытых данных базу всех DEF-номеров в разрезе по регионам и операторам? Теперь каждый, написав на Python псевдокод из нескольких строк вроде:

```
import Skype4Py
Call = skype.PlaceCall(userNumber)1,
```

может создать свою базу из сотен тысяч потенциальных «клиентов».

Возможно, что-то подобное нас ожидает и в области открытых данных при кредитовании? Однако здесь данные более чувствительны — речь идет об имуществе и официальных документах.

Мы попросили опытного кредитного аналитика, пожелавшего остаться анонимным, легально собрать информацию о каком-нибудь ИП, чтобы наши опасения развеялись и мы продолжали уверенно пользоваться всеми сервисами при кредитовании, не задавая лишних вопросов.

Однако вопросы появились.

Предлагаем проследить за его работой.

Заходим на reestr-zalogov.ru/search/index. Вводим в строку поиска фамилию и имя:

«Проваливаемся» в первую попавшуюся запись:

Федеральная нотариальная палата			Федеральный ресурс		
Дата регистрации	Номер уведомления о возникновении залога	Имущество	Залогодатель	Залогодержатель	
6	01.01.2021	Земельный участок. История изменений ²	Данные не представлены	Полное наименование: ООО "Сбербанк России"	Открытое акционерное общество "Сбербанк России"

¹ Расширенный пример: github.com/xapax/automatic-phonecalls/blob/master/call.py.

персональные данные \ информационная безопасность \ операционные риски

Тотальный контроль и беззащитность: кого кредитовать в 2022 году

Получаем данные паспорта и дату рождения:

Уведомление о возникновении залога движимого имущества
номер 2114-000-016027-000 от 05.12.2014 12:45:04 (время московское)
Состояние: Актуальное

1. Движимое имущество, переданное в залог

1.1 Иное имущество

1	ID	
	Описание иного имущества	погрузчик-экскаватор CATERPILLAR 434F, 2014 года выпуска, заводской № машины (рамы) С, двигатель № CRS, коробка передач № F, цвет желтый, паспорт самоходной машины и других видов техники ТТ 169996 от 30.04.2014 года

2. Сведения о залогодатель(ях)

2.1 Физическое лицо

1	Фамилия	Петров
	Имя	Павел
	Отчество	Дмитриевич
	Фамилия (латинскими буквами)	
	Имя (латинскими буквами)	
	Отчество (латинскими буквами)	
	Дата рождения	18.06.1982
	Документ, удостоверяющий личность	Паспорт гражданина Российской Федерации, 9 000000
	Адрес фактического места жительства в Российской Федерации	Удмуртская респ.
	Адрес электронной почты	

Проверяем, что паспорт все еще действует, на services.fms.gov.ru/:

Результат запроса

По Вашему запросу о действительности паспорта РФ 9 000000 № 9 000000 получен ответ о том, что данный паспорт «Среди недействительных не значится».

Параметры запроса

Поиск был произведен по следующим, указанным Вами параметрам:

Серия паспорта: 9 000000
Номер паспорта: 9 000000

[сообщить об ошибке](#)
[другой запрос](#)

Владимир КОЗЛОВ

Получаем ИНН на service.nalog.ru/inn.do¹:

Информация об ИНН найдена.

ИНН: XXXXXXXXXX

СВЕДЕНИЯ О ФИЗИЧЕСКОМ ЛИЦЕ

Фамилия*:

Имя*:

Отчество*: ☐ Нет отчества

Дата рождения*:

Место рождения:

Вид документа, удостоверяющего личность*:

Серия и номер документа*:

Дата выдачи документа:

* - Поля, обязательные для заполнения

Физическое лицо оказалось индивидуальным предпринимателем.

Яндекс

Поиск: Словари Видео Карты Маркет Новости Переводчик Календар Музыка Все

Найдено 191 результат

1 ИП П. П., село Первомайский...
rusprofile.ru > гл...
Основная вид деятельности является «Перевозка грузов неспециализированными автомобильными средствами». ИП П. П. зарегистрирован в налоговой 2014-01-30. Полный перечень юридических лиц, в которых... Читать еще >

2 ИП - П. П., село Первомайский, ИНН 1... ОГРН...
zachestnyibiznes.ru > компания...
По данным портала ЗАЧЕСТНЫЙБИЗНЕС, ИП П. П. зарегистрирован в налоговой 2014-01-30. Вы можете использовать ОГРН и использовать... при составлении... Читать еще >

3 ИП П. П., село Первомайский, ИНН 1... в село...
vypiska-nalog.com > запись...
Фирма с ИНН 1... зарегистрирована в налоговой 2014-01-30. Фирма с ИНН 1... надежная и стабильная, и НКОСБ риск закрытия или банкротства. Это основано на текущем рабочем статусе организации и возрасте... Читать еще >

4 ИП П. П., село Первомайский, ИНН 1...
basis.myseldon.com > компания...
При регистрации предпринимателя присвоены реквизиты: ИНН - 1..., ОГРН - 31... Основной вид деятельности - Перевозка грузов неспециализированными автомобильными средствами. Давать... Читать еще >

5 П. П., село Первомайский, ИП, ИНН 1...
5819.ru > Проверка контраентов > 1...
1... Перевозка грузов неспециализированными автомобильными средствами. еще 1 вид деятельности. Читать еще >

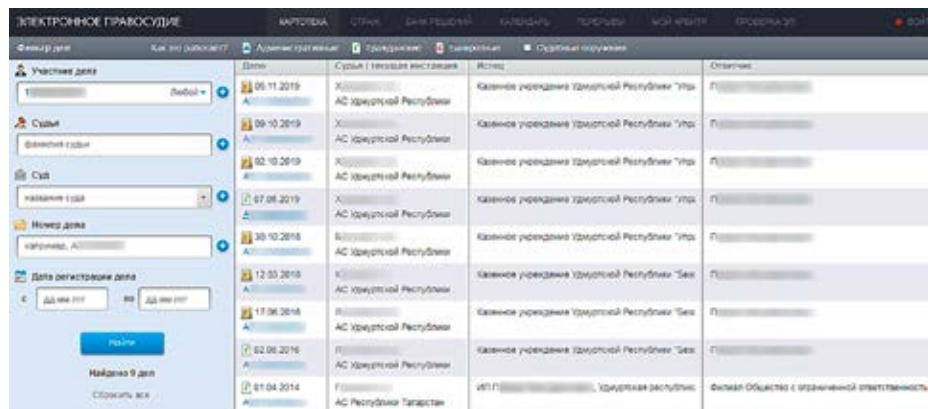
6 П. П., село Первомайский (ИНН: 1...) - участник...
sudact.ru > Арбитражные суды > арбитражные суды...
Последние судебные акты. Участие судебных дел. Последние документы. Читать еще >

¹ Необходимо поставить галочку, подтвердив, что предоставляешь данные о себе лично. Каков реальный правовой статус этих галочек, мы сможем понять только через несколько лет активного правоприменения.

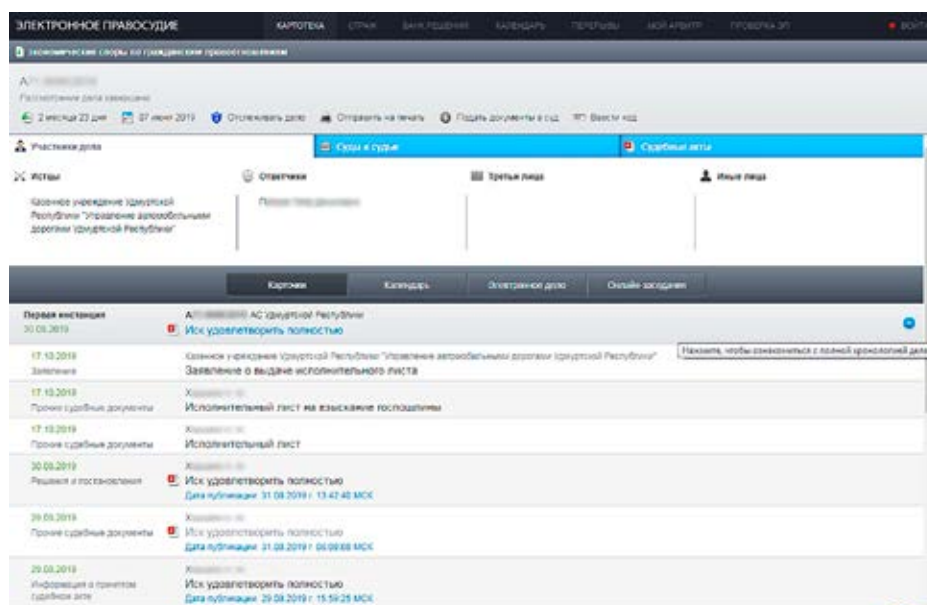
персональные данные \ информационная безопасность \ операционные риски

Тотальный контроль и беззащитность: кого кредитовать в 2022 году

Заходим на kad.arbitr.ru, вводим ИНН в строку поиска:



Смотрим информацию о судебном деле:



Нажимая на судебный документ, получаем данные о личном имуществе гражданина:

Суд установил. Как следует из материалов дела, 15 октября 2016 года на пункте весового контроля истца, расположенном на автомобильной дороге «Окружная а/д г.Ижевск», 6 км.+250, установлен факт нарушения

3
1157_924350
допустимых весовых параметров транспортным средством КАМАЗ 5100, государственный регистрационный номер Р-1000000, принадлежащим ответчику (свидетельство о регистрации 18 ХО №7000000), о чем был составлен акт о превышении транспортным средством установленных ограничений по общей массе и (или) нагрузке на ось № 443 от 15.10.2016.

Владимир КОЗЛОВ

Нашелся КАМАЗ 544108, теперь нам известны регистрационный номер и номер свидетельства о регистрации. Проверяем, что КАМАЗ принадлежит данному физическому лицу, на сервисе гибдд.рф/check/fines:

Проверка штрафов ГИБДД

Государственный регистрационный знак: **Р 544108**

Свидетельство о регистрации ТС: **1**

[интерактивный госномер очистить форму](#)

Результаты проверки

(ГРЗ: Р 544108, СТС: 1)

Проверка проведена 6 августа 2021 г. в 8:02:31

Проверка наличия информации о неуплаченных штрафах за правонарушения в области дорожного движения

Все корректно.

С помощью любого частного сервиса бесплатно узнаем VIN по гос-номеру:

Поиск по гос. номеру

Модель	КАМАЗ БЕЗ МОДЕЛИ
Цвет	ОРАНЖЕВЫЙ
Тип ТС	Грузовые автомобили тягачи седельные
Год	2011
Объем двигателя	11760.0
Номер двигателя	В 11760.0
Мощность двигателя	280.16лс.
Категория	C
VIN номер	X 11760.0
Номер кузова	2 11760.0
Серия и номер ПТС	1 11760.0
Кем выдан ПТС	ООО "ФИНАВТОСЕРВИС"
Владельцев по ПТС	2
С 20.02.2012 по 26.01.2020	Физическое лицо
С 26.01.2020 по настоящее время	Физическое лицо
Данные актуальны на 06.08.2021 08:17	
Найденные VIN номера: копировать VIN	
X 11760.0	
Тип проверки:	
История регистраций	

персональные данные \ информационная безопасность \ операционные риски

Тотальный контроль и беззащитность: кого кредитовать в 2022 году

Такие сервисы, как правило, составлены из ранее открытых данных на портале РСА¹.

Через сервис ГИБДД «Проверка автомобиля» (гибдд.рф/check/auto#...) выясняем последние события с автомобилем и его собственниками.

Нашлось ограничение:

Проверка проведена 6 августа 2021 г. в 8:21:25

Информация о наложении ограничения

Марка (модель) ТС:	Нет данных
Год выпуска ТС:	2011 г.
Дата наложения ограничения:	01.02.2021 г.
Регион инициатора ограничения:	Удмуртская Республика
Кем наложено ограничение:	Судебный пристав
Вид ограничения:	Запрет на регистрационные действия
Основание:	
Документ: 2	от 01.02.2021, А. [имя]
СМИ: 9	от 29.01.2021
ИП: 6	от 29.01.2021
Телефон инициатора:	+7 (917) 007-13-87
Ключ ГИБДД:	18# [ключ]

Переходим на сайт ФССП (fssp.gov.ru) по указанному номеру исполнительного производства:

Найдено записей: 1

Для участия в исполнительном производстве воспользуйтесь экраном <https://www.fssp.gov.ru/>

Должник (физ. лицо): ФИО, дата и место рождения, сер. лист, наименование, ср. адрес, фактический адрес	Исполнительное производство: (номер, дата возбуждения)	Реквизиты исполнительного документа (вид, дата, принятая орган, номер, наименование органа, выдавшего исполнительный документ)	Дата, причина окончания или прекращения ИП (статья, часть, пункт, основания)	Сервис	Предмет исполнения, сумма непогашенной задолженности	Отдел судебных приставов (наименование, адрес)	Судебный пристав-исполнитель, телефон для получения информации
П. [имя] П. [фамилия] 06.05.1988 ПОС. ПЕРВОМАЙСКИЙ ЗАВЬЯЛОВСКОГО Р-НА УДМУРТСКОЙ АССР	5 [номер] ИП от 29.01.2021	Акт по делу об административном правонарушении от 18.06.2020 № 1 [номер] ОГИБДД ММО МВД РОССИИ "САРАПУЛЬСКИЙ"		оплатить	Штраф ГИБДД: 17999.69 руб.	Завьяловское РОСП 427000, Россия, Респ. Удмуртская, . . с. Завьялово, ул. Чкалова, д. 20, .	А. [имя] В. [фамилия] +7 [номер]

Видим родственника нашего физического лица (вероятно, сына), также из пос. Первомайский.

Попробуем узнать номер его прав.

Нажмем кнопку «Оплатить», но выберем «Сформировать квитанцию»:

оплатить Штраф ГИБДД: 17999.69 руб.

госуслуги
Оплата любыми картами Квитанция

¹ Российский союз автостраховщиков.

Владимир КОЗЛОВ

В скачанной Excel-квитанции находим УИН:

The screenshot shows a payment receipt from the Federal Tax Service (ФНС России). The UIN (УИН) is highlighted in green in the top right corner. The receipt details include the amount of 17,999.00 RUB and the date of payment.

УИН вставляем в поисковую строку любого сервиса оператора платежей:

The screenshot shows the '10 money' website interface. The UIN is entered into the search bar, and the button 'Найти штраф' (Find fine) is visible. The website also has a 'Войти' (Login) button and a 'Создать кошелек' (Create wallet) button.

Получаем номер водительского удостоверения:

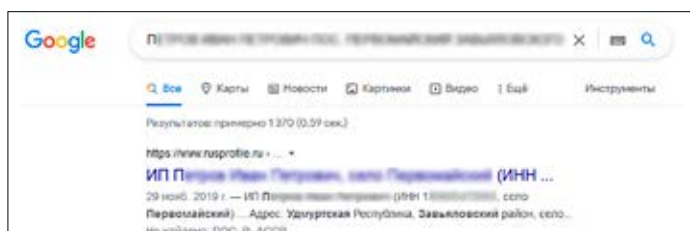
The screenshot shows the result of the search for the fine on the '10 money' website. The fine amount is 17,999.00 RUB, and the driver's license number is displayed. The website also shows the date of the fine (28 января 2021) and the document type (Постановление).

персональные данные \ информационная безопасность \ операционные риски

Тотальный контроль и беззащитность: кого кредитовать в 2022 году

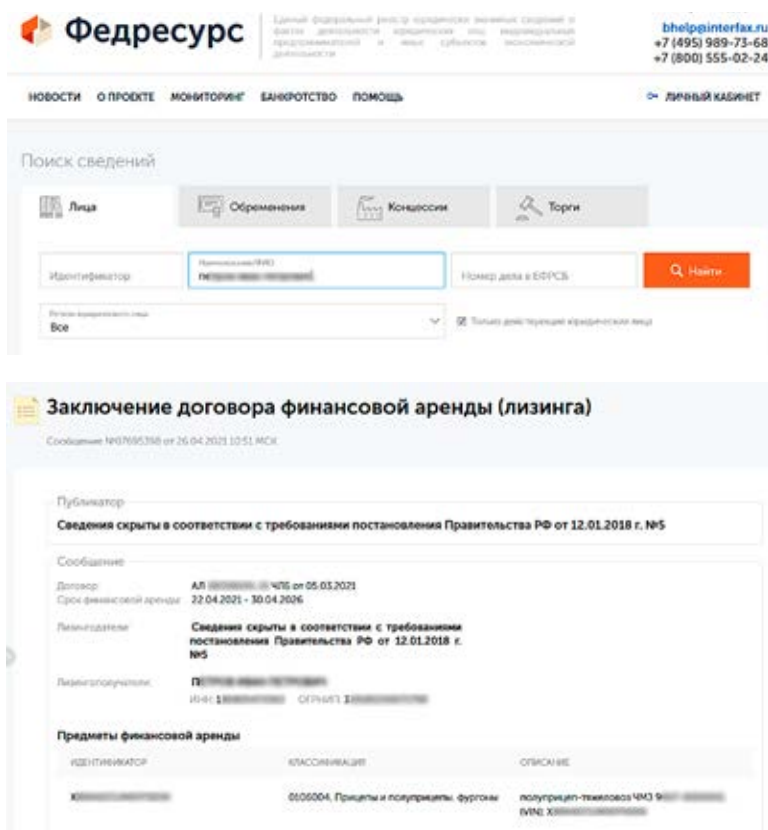
Поищем другое движимое имущество, принадлежащее данной семье.

Находим ИНН родственника нашего физического лица в первом сниппете любой поисковой системы:



Он также индивидуальный предприниматель.

На сайте Единого федерального реестра юридически значимых сведений о фактах деятельности юридических лиц, индивидуальных предпринимателей и иных субъектов экономической деятельности (fedresurs.ru) находим лизинговое имущество с VIN-кодом:




Нашелся полуприцеп-тяжеловоз.

Владимир КОЗЛОВ

На этом месте мы попросили нашего аналитика прекратить работу. Однако он указал на то, что информация о недвижимом имуществе и составе собственников, как известно, доступна любому, кто желает получить официальные выписки по кадастровому номеру. Кадастровый номер можно узнать по адресу, например адресу регистрации в паспорте.

Зная номер паспорта и то, что в 98,5% случаев паспорт выдают минимум за пять лет до заявленного года печати бланка (3–4-я цифра номера паспорта), а максимум — в течение трех лет после него¹, применяя требования о возрасте смены паспорта, после не более 700² попыток мы получим на официальном сервисе МВД РФ (сервисы.гувм.мвд.рф/info-service.htm?sid=2160) адрес регистрации и далее получим выписку из кадастра с указанием состава собственников.

В несколько опасном положении находятся наши заемщики — могут ответить специалисты по операционным рискам. Возможно, банкам стоит поднять вопрос о закрытии информации в Нотариальной палате (номер паспорта мы нашли именно там) и ряде сходных сервисов либо передавать информацию по закрытому протоколу, как в СМЭВ. В случае пассивности банков недалек день, когда на github появятся пара библиотек, которые по ФИО и какому-нибудь очевидному идентификатору вроде даты рождения найдут всех ближайших родственников и их имущество благодаря открытым и бесплатным государственным сервисам.

Кого же можно кредитовать в 2022 г.? Учитывая количество имеющихся открытых данных — любую подвыборку клиентов по любому имущественному/неимущественному сегменту. Только, «раздевая» наших клиентов с точки зрения информационной безопасности, не потеряем ли мы их прежде, чем успеем привлечь? 

[Вернуться в Содержание](#)

¹ habr.com/ru/company/hflabs/blog/478538/.

² Округленно.

В статье описаны методы количественной оценки качества дискриминации (одобрения) в сегментах розничного портфеля на основании текущих статистических данных об уровне отказа клиентам, обратившимся в банк, текущем уровне дефолтов и данных рынка (бюро кредитных историй). Проведен анализ экономической эффективности применяемого уровня одобрения заявок/отказа с учетом кредитного риска. На примере нескольких кейсов рассмотрены типичные причины слабости оценок качества дискриминации риск-процедур.

Михаил ПОМАЗАНОВ, ПАО «Промсвязьбанк», Дирекция рисков, руководитель по валидации, к.ф.-м.н.

Владимир ШИКИН, Национальное бюро кредитных историй (НБКИ), заместитель директора по маркетингу

Методика валидации эффективности риск-менеджмента розничного портфеля

Сегментация портфеля заявок



На первом этапе процедуры количественной оценки качества необходимо определить портфель заявок для анализа и сегментировать его. Отбор осуществляем по историческому периоду (винтаж) выдач кредитов, сегментацию — по статусу кредитной заявки. При определении исторического периода важно добиться исключения сезонных факторов, то есть период должен быть однородным с точки зрения их влияния. Так как основная контрольная переменная дискриминационных процедур в розничном кредитовании — это выход на просрочку свыше определенного количества дней (30+, 60+, 90+ и т.д.), исторический период продолжаем влево и вправо по временной шкале на то количество дней, которое определено в бизнес-правилах для данного типа кредитования. То есть (рис. 1):

- выбирается исследуемый сегмент заемщиков;
- выбирается винтажный период;
- предполагается равномерность выдач.

Кредитные заявки, попавшие в исследуемый портфель, сегментируем по статусу выдачи (табл. 1).

Методика валидации эффективности риск-менеджмента розничного портфеля

Рисунок 1

Модель винтажного портфеля для валидации

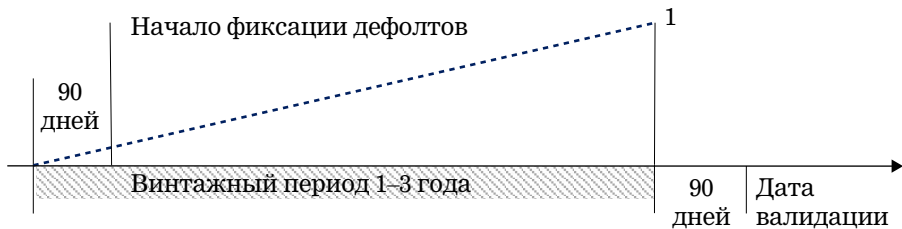


Таблица 1

Сегментация заявок по статусу выдачи кредита и группировка для включения в статистику валидации

Статус		Группа	Доли, %
Заявка одобрена и кредит выдан		A	20
Кредит не выдан	Заявка одобрена, но клиент отказался от получения	A'	10
	Отказ клиента в начале процедур	—	2
	Без финального решения	—	3
	Отказ по лимиту долговой нагрузки	—	15
	Отказ по минимальным требованиям (правилам)	—	8
	Отказ андеррайтинга	C	5
	Отказ по антифрод-системам	C	2
	Отказ по кредитной истории	C	20
	Отказ по скорингу	C	10
	Отказ по операционным ошибкам	C	5
Итого			100

Примечание: здесь и далее, если об этом не сообщено дополнительно, все цифровые значения являются модельными и неприменимы к показателям по отдельному банку, группе банков или сегменту рынка, известным авторам.

- В таблице:
- группа А учтена в статистике валидации, дает статистику дефолтов;
 - группа А' учтена в статистике валидации, не дает статистику дефолтов;
 - группа С учтена в статистике валидации, не дает статистику дефолтов;

Михаил ПОМАЗАНОВ Владимир ШКИН

— заявки без группы (—) не учтены в статистике валидации, так как по этим заявкам у банка нет свободы принятия решения об одобрении/отказе.

Таким образом, мы имеем возможность скорректировать первоначально сгенерированную выборку по эффективному числу заявителей на долю заемщиков, не взявших кредит. Для абстрактного примера, приведенного в табл. 1, корректировка приводит к результатам:

$$B = A + C \frac{A}{A + A'} = 48\%.$$

Коррекция рыночной доли дефолтности на долю «отказников», получивших кредит в другом банке

На следующем этапе мы должны сформулировать гипотезу: заявители, прошедшие обязательные формальные стадии отбора, имеют рыночные показатели дефолтности, скорректированные на статистическую долю попадающих на рынок после отказа банка.

Это означает, что если в банк приходит клиент и получает отказ по кредитной заявке, то существует вероятность, что этот клиент получит кредит в другом банке. Оценить масштаб такого явления помогает стандартизированный отчет НБКИ «Поведенческий анализ заемщиков после получения отказа», который позволяет узнать долю таких клиентов и, кроме того, качество обслуживания полученных кредитов такими «отказниками» на рынке.

Чтобы определить рыночную долю дефолтности в нужном нам сегменте кредитования и в исследуемых периодах выдачи, можно воспользоваться винтажными отчетами БКИ¹.

Таким образом, после сбора необходимых данных мы получаем возможность провести коррекцию рыночной доли дефолтности на долю кредитуемых рынком (референтной группой кредиторов) заявителей, получивших отказ в банке:

$$DR = DR(M) + \frac{A}{B} \times \left(\frac{1}{p} - 1 \right) \times (DR(M) - DR(A)),$$

где DR(M) — оценка среднерыночной доли дефолтов за винтажный период выдачи по данным БКИ (в нашем примере мы будем считать этот показатель равным 4,8%). Числитель показателя находится простым суммированием числа всех заемщиков исследуемого сегмента статистики БКИ, выпшедших на заданный

¹ В частности, НБКИ предоставляет винтажные таблицы в своем регулярном Национальном кредитном бюллетене, и кроме того, для получения более точных индикаторов может предоставить отдельный винтажный отчет по референтной группе кредиторов.

Методика валидации эффективности риск-менеджмента розничного портфеля

барьер просрочки (например, 90+), на всех отчетных периодах винтажного периода (см. рис. 1). Знаменатель — количество всех выданных кредитов;

p — среднестатистическая доля отказных претендентов, попадающая на кредитный рынок после отказа банка¹ (в нашем примере показатель принят равным 70%);

$DR(A)$ — доля дефолтов среди кредитруемых в нашей популяции заемщиков (3%).

Таким образом, после подстановки всех значений скорректированная доля дефолтов составит 5,12%.

Более подробно процедура и формула коррекции объясняются следующим.

Качество входящих в банк заявителей (рыночно-продуктовый сегмент банка) будет отличаться от качества тех, кто входит в статистику БКИ. Это связано с тем, что банк отказывает части заявителей, поэтому если бы заявители, которые не попадают в банк, не попадали никогда на кредитный рынок, то дефолтность DR заявителей банка должна была бы совпадать с дефолтностью $DR(M)$ на рынке (в БКИ). Но это не так. В то же время следует учесть, что не все заявители, которым отказывает банк, далее попадают на рынок и участвуют в статистике $DR(M)$. Поэтому необходимо делать коррекцию на долю заемщиков, получившую в банке отказ, но попадающую на кредитный рынок.

Необходимо провести исследование, опираясь на кредитный отчет БКИ по заявителям, которым банк отказал по кредитной заявке в статистически достаточный референтный период.

Для анализа выбирается несколько основных причин отказов, например:

- отказы по кредитной истории, в том числе из-за наличия микрофинансового кредита, исторической просрочки 180+, большого количества запросов в БКИ и пр.;

- отказы по скорингу;

- отказы по антифрод-политикам (QCC2 и FPS);

- отказы на этапе андеррайтинга.

Определяется референтный период так, чтобы на каждую причину отказа приходилось не менее 1000 заявителей, формируется запрос отчета в БКИ о наличии кредита у заявителей, оформленного в других банках после отказа вашего банка. Результатом этого исследования и будет параметр p (среднестатистическая доля отказных

¹ Этот важный параметр и устанавливается для интересующего сегмента банка из стандартизированного отчета НБКИ «Поведенческий анализ заемщиков после получения отказа».

Михаил ПОМАЗАНОВ Владимир ШИКИН

претендентов, попадающая на кредитный рынок после отказа банка). Этот параметр сугубо индивидуален для каждого банка, его рыночного и продуктового сегмента.

Пусть банк среди B заявителей кредитует количество A , то есть отказывает $B - A$, тогда частота (доля) дефолтов по всем входящим в банк заявителям будет вычисляться как

$$DR = \frac{A \times DR(A) + (B - A) \times DR(B - A)}{B},$$

где $DR(A)$, $DR(B - A)$ — дефолтность среди одобренных и отказных соответственно.

При этом часть p от $B - A$ участвует в статистике дефолтов БКИ. Учитывая, что рыночная дефолтность формируется среди всех одобренных банком и среди доли p отказных, получим частоту (долю) дефолтов на рынке в соотношении

$$DR(M) = \frac{A \times DR(A) + p \times (B - A) \times DR(B - A)}{A + p \times (B - A)}.$$

Исключая неизвестное $DR(B - A)$ из соотношений выше, получаем формулу коррекции рыночной доли дефолтности на долю кредитруемых рынком заявителей, получивших отказ в банке.

Валидация процедур риск-менеджмента

Вся наша популяция заявок может быть представлена в виде табл. 2, в которой все значения приведены к результатам работы процедур риск-менеджмента банка.

При этом, конечно, значения в более светлых ячейках получены нами в результате моделирования, а значения в остальных ячейках представляют собой объективные данные.

Таблица 2

Сегментация популяции в терминах риск-менеджмента

Результат решения риск-менеджмента	Дефолт	Нет дефолта	Всего
Риски «против»	$B \times DR - D$	$B \times (1 - DR) - A + D$	$B - A$
Кредит выдан	D	$A - D$	A
Итого	$B \times DR$	$B \times (1 - DR)$	B

Методика валидации эффективности риск-менеджмента розничного портфеля

Также из табл. 2 видны ошибки риск-менеджмента:

— 1-го рода — заявки были отклонены, но обслуживание аналогичных кредитов на рынке у «отказников» не сопровождалось дефолтом: $B \times (1 - DR) - A + D$;

— 2-го рода — принято положительное решение, но обслуживание кредита сопровождалось реализацией целевой переменной — дефолтом (D).

Для оценки эффективности риск-менеджмента в данном упражнении мы используем метод оценки индекса Gini, показывающего мощность дискриминационных процедур. Для этого по имеющейся в нашем распоряжении известной точке дефолтов необходимо построить CAP-кривую (cumulative accuracy profile).

Для восстановления CAP-кривой воспользуемся широко известной моделью Ван дер Бурга (Van der Burgt), которая имеет независимую переменную, являющуюся решением уравнения:

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

где k — параметр, показывающий эффективность (мощность) решений риск-менеджмента.

Результат восстановления CAP-кривой показан на рис. 2. При этом полученная кривая включает известную нам точку, координаты которой мы получили ранее:

— по оси «% всех заемщиков» — доля отклоненных заявок (риски «против») в общей популяции заемщиков;

— по оси «% всех дефолтов» — доля дефолтов в отклоненных заявках.

Индекс Gini, таким образом, вычисляется по формуле:

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

Требования к Gini (прогнозной мощности) дискриминационных процедур могут быть разными. Но в качестве базового ориентира из банковской практики розничного кредитования можно представить определенные рекомендации в отношении зональных оценок (табл. 3).

Таким образом, каждое полученное значение Gini всех совокупных процедур риск-менеджмента мы можем отнести к той или иной зоне. Говоря о совокупности, мы имеем в виду, что оценивается не только сама скоринговая модель, но и весь комплекс правил и процедур, применяемых риск-менеджерами для принятия решения «отказ» или «одобрение» (см. табл. 1).

Михаил ПОМАЗАНОВ
Владимир ШИКИН

Рисунок 2

Восстановление SAR-кривой дискриминационной точности риск-менеджмента

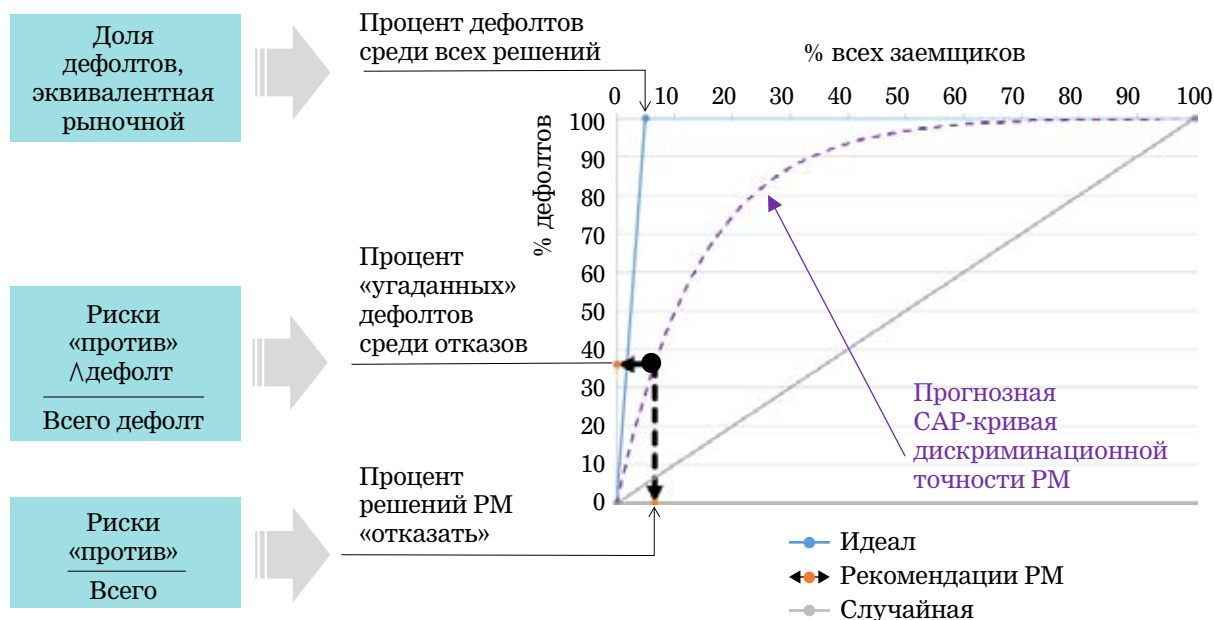


Таблица 3

Рекомендации в отношении зональных оценок с использованием метрики Gini

Модель	Диапазоны Gini, %		
	Красная зона	Желтая зона	Зеленая зона
Поведенческая	< 40	40–60	> 60
Аппликативная	< 35	35–55	> 55

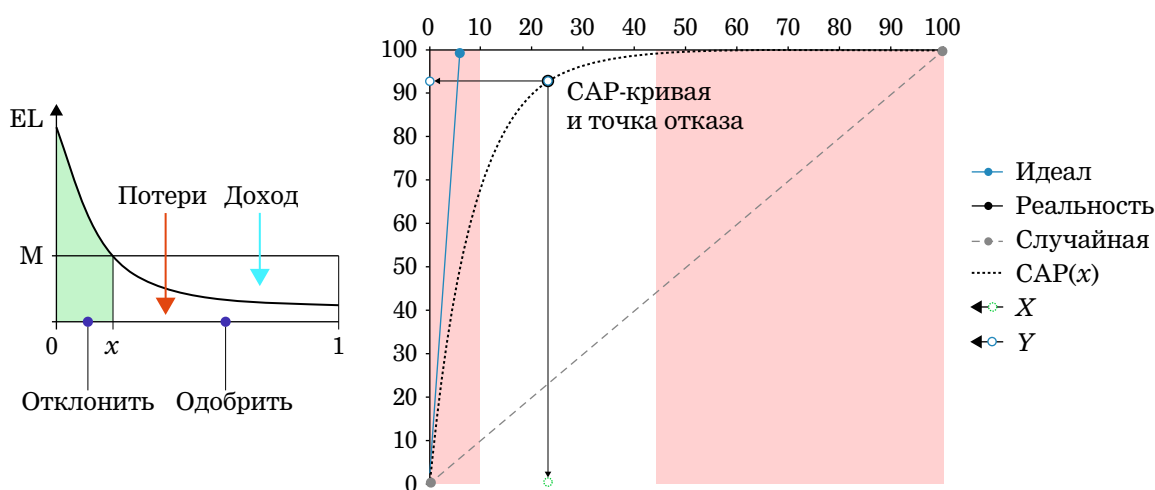
Следующим шагом валидации риск-менеджмента должна стать оценка экономической (коммерческой) эффективности: насколько точка «отказа» оправдана с точки зрения экономики розничного кредитования в банке. Понятно, что оптимальная точка дискриминации для «плохих» и «хороших» заемщиков должна соответствовать уровню потерь $EL(x)$, не превышающих маржинальную доходность (M) по кредитному продукту.

Схематично модель оценки потерь и формирования маржи можно представить в виде совмещения двух графиков, как на рис. 3.

Методика валидации эффективности риск-менеджмента розничного портфеля

Рисунок 3

Схематичное представление оценки маржи (М) и потерь (EL) по кредитному продукту



Здесь под уровнем потерь (EL) мы, конечно, имеем в виду не только дефолтность по одобренным кредитам:

$$PDx(x, Gini) = DF \times (1 - CAP(x, Gini)),$$

но и возможную прибыль по тем заявкам, по которым кредит был не получен клиентом:

$$P(x, Gini) = M - x \times M - PDx(x, Gini) \times LGD,$$

где М — норма NPV дохода винтажного периода с учетом всех затрат и сроков амортизации ссуд (кредитов);

x — точка отказа в процентах;

DF — доля дефолтов на входящем потоке, эквивалентная рыночному значению в данном сегменте кредитования.

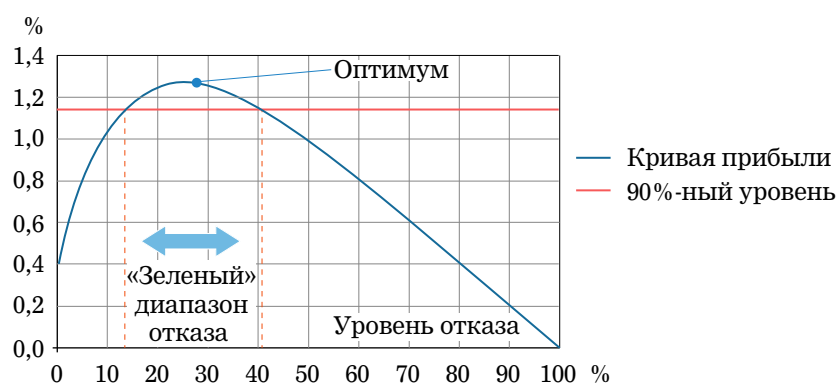
Прибыль после отказа $P(x, Gini)$ можно проиллюстрировать графиком (рис. 4), на котором выделены три зоны.

Например, первая левая зона соответствует тому, что если риск-менеджмент не будет отказывать таким клиентам, то дефолтность части этих клиентов будет превышать полученную прибыль и прибыль портфеля станет неоптимальной, а возможно, и отрицательной.

Михаил ПОМАЗАНОВ
Владимир ШИКИН

Рисунок 4

Схематичное представление кривой прибыли в координатах уровня отказов и доходности



Наша задача заключается в том, чтобы найти оптимальную точку и так называемый зеленый диапазон, при котором отказ должен стать экономически оправданным.

Для этого мы возвращаемся к зональному представлению метрики, в качестве которой на этот раз выступает уровень отказа (табл. 4).

Таблица 4

Зональная оценка уровня отказов

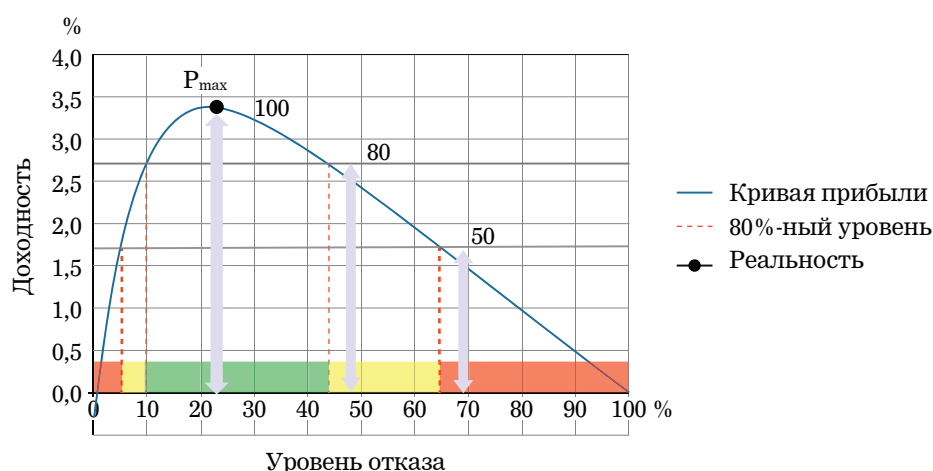
Метрика	Диапазоны значений недополученного дохода, %		
	Красная зона	Желтая зона	Зеленая зона
Уровень одобрения (отказа)	50–100 либо убыток	20–50	Менее 20

Таким образом, в зеленую зону входят значения уровня одобрения, при которых незаработанная прибыль теряется не более чем на 20%, в желтую — при которых теряется от 20 до 50%, в красную — при которых теряется более 50% незаработанной прибыли или генерируются убытки. При этом цветовая дифференциация в нашем примере достаточно условна: каждый банк может зонировать данную метрику, исходя из собственного опыта и целей. Предложенный подход можно считать относительно консервативным. Визуально зональное распределение представлено на рис. 5.

Методика валидации эффективности риск-менеджмента розничного портфеля

Рисунок 5

Зональное представление маржинальной доходности с учетом риска



Далее мы можем сформировать ключевые показатели доходности (КPY) с учетом риска:

— P_x — текущая NPV-доходность входящего потока заявителей, скорректированного на не взявших кредит заемщиков, на текущем уровне отказа x с учетом риска (она и оптимизируется);

— P_{max} — максимальная NPV-доходность скорректированного входящего потока с учетом риска;

— P_A — текущая NPV-доходность с учетом риска по популяции заемщиков, взявших кредит;

— P_{All} — NPV-доходность с учетом риска по всей популяции заявителей, за исключением не учитываемых в статистике. В табл. 1 эти заявители указаны без группы.

И в процессе валидации процедур риск-менеджмента мы имеем возможность оценить все кредитные продукты с помощью этих ключевых показателей. Таблица 5 является примером такой валидационной процедуры.

Анализ валидационных таблиц, полученных по данной методике, представляет собой отдельную работу, которую должны проводить не только риск-менеджеры, но и все подразделения банка, участвующие в создании и дистрибуции кредитного продукта. Между тем некоторые типовые случаи выделить можно. В табл. 6 представлены три часто встречающиеся на практике причины слабости дискриминационных процедур, для которых сформиро-

Михаил ПОМАЗАНОВ
Владимир ШИКИН

Таблица 5

Валидационная таблица оценки риск-менеджмента на различных кредитных продуктах

Продукт	Валидация, %					КРУ			
	Gini	Уровень отказа	Потеря доходности	Оценка Gini	Оценка уровня отказа	P_x	P_{max}	P_A	P_{All}
1	46	32	2	Желтый	Зеленый	1,6	1,7	2,4	0,1
2	85	28	1	Зеленый	Зеленый	3,1	3,2	4,4	2,3
3	33	63	15	Красный	Желтый	1,1	1,3	2,7	0,6

Таблица 6

Типичные причины слабости риск-менеджмента

Причина	Гипотеза по улучшению
Среди одобренных банком клиентов значительная часть оставляет предложение невостребованным или уходит в другой банк. Те, кто пользуется предложением, оказываются ниже качеством, чем в среднем по одобренным. Результат: снижение Gini	Сегментировать одобренных клиентов по кредитному качеству, предлагая лучшим из них более выгодные условия (Risk Based Pricing, RBP)
Пониженный (повышенный) уровень одобрения заявителей, снижение коммерческой эффективности продукта за счет роста ошибок I и II рода	Регулярная корректировка политики на основе запроса рыночных данных (по сегменту рынка, по референтной группе) в БКИ. Корректировка оптимального уровня отказа с учетом текущих ставок
Низкая дискриминационная сила кредитного процесса в определенных сегментах. Результат: обнуление Gini, необоснованно высокий уровень отказа	Доработка скоринговых моделей, внедрение сегментоориентированных моделей, тестирование и валидация кастомизированного и (или) отраслевого скоринга БКИ
Стагнация общего уровня коммерческой эффективности розничного кредитования/продукта	Регулярная валидация процедур риск-менеджмента, скоринговых моделей, системы правил (стоп-факторов), изучение поведения «отказников». Внедрение для риск-менеджеров и сотрудников кредитующих подразделений мотивационных инструментов повышения эффективности кредитного процесса

Вернуться
в Содержание

ваны гипотезы по повышению их эффективности. Формулирование этих гипотез, их отработка и повышение эффективности риск-менеджмента в конечном итоге и являются целью валидации.

В большинстве случаев методы моделирования, применяемые для оценки кредитного риска, не выходят за рамки классических методов машинного обучения. Однако спектр перспективных математических подходов и вычислительных инструментов может быть шире. В статье рассмотрены возможности пакета прикладных программ MATLAB для построения системно-динамических моделей, которые учитывают структуру компаний и могут применяться при обратном стресс-тестировании, а также при оценке вероятности дефолта.

Применение MATLAB Simulink для построения моделей оценки кредитного риска корпоративных заемщиков

Системная динамика представляет собой подход имитационного моделирования, предназначенный для исследования структуры и динамики сложных объектов. Системно-динамические модели — это системы дифференциальных уравнений. Однако на практике они представимы более просто: в виде потоковых диаграмм. Системно-динамическая модель компании позволяет оценить ее устойчивость к влиянию макроэкономических факторов.

Разработка системно-динамических моделей делится на два ключевых этапа:

- на первом описывается модель в виде потоковых диаграмм и определяются характеристики взаимного влияния ее частей;
- на втором этапе модель реализуется при помощи вычислительных инструментов и уточняется. Для этого применяется компьютерная симуляция, тестируются гипотезы о поведении объекта, а выявленные зависимости проверяются на тестовых данных.

Ключевыми элементами потоковых диаграмм являются накопители, потоки, связи, обратные связи и вспомогательные переменные.

Накопители (рис. 1) отражают текущее состояние системы. При этом они аккумулируют определенный материальный или нематериальный

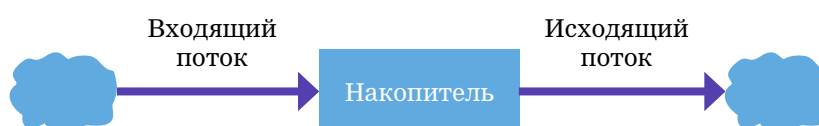


**Дмитрий
КУРЕННОЙ,**
*ПАО «Промсвязь-банк», главный
специалист отдела
моделирования*

Дмитрий КУРЕННОЙ

Рисунок 1

Простейшая потоковая диаграмма, включающая в себя два потока и один накопитель

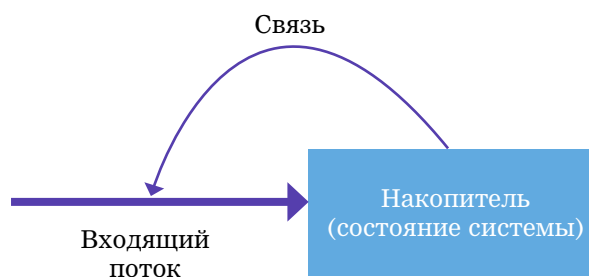


компонент и изменяются в результате влияния входящих и исходящих потоков. *Потоки* определяют изменение состояния системы во времени и описываются функциональными зависимостями от других элементов модели.

Связи отражают наличие влияния совокупности факторов на элемент модели. Факторы могут быть независимыми от других элементов системы или представлять определенную функциональную зависимость. *Обратная связь* (рис. 2) — разновидность связей, которая представляет собой механизм системы, усиливающий или стабилизирующий ее динамику путем косвенного влияния элементов на самих себя.

Рисунок 2

Элементарная обратная связь



Вспомогательные переменные обеспечивают возможность описания функциональных зависимостей и используются для представления зависимых или независимых элементов системы, влияющих на ее динамику, но не являющихся ни потоками, ни накопителями.

Динамическое поведение системы, то есть ее изменение во времени, является следствием интегрирования потоков в накопителях.

Применение MATLAB Simulink для построения моделей оценки кредитного риска корпоративных заемщиков

Инструменты имитационного моделирования в MATLAB Simulink

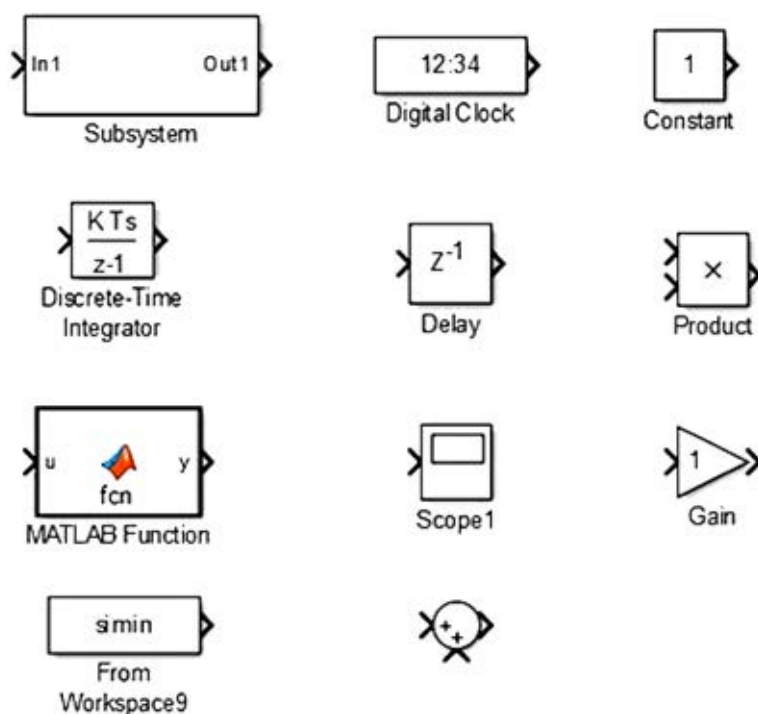
Преимущества MATLAB Simulink — наличие графической среды для создания моделей и широкий спектр инструментов для решения сопутствующих оптимизационных задач, например связанных с прогнозированием подаваемых на вход модели макроэкономических факторов.

Графическая среда моделирования позволяет представлять потоковые диаграммы в виде блок-схем. Основным элементам потоковой диаграммы соответствуют специальные блоки и сигналы Simulink (рис. 3), спецификации которых можно задавать в коде на языке MATLAB.

Элемент *Subsystem* (подсистема), представленный в виде отдельного блока, предназначен для корректного разбиения сложной модели на более простые составные части — подсистемы. Эти составные

Рисунок 3

Блоки MATLAB Simulink для представления потоковых диаграмм



Дмитрий КУРЕННОЙ

части связаны между собой входными (Inport) и выходными (Outport) обобщенными потоками.

Discrete-Time Integrator (интегратор) — ключевой элемент, реализующий процесс интегрирования потоков, который происходит в накопителях. Таким образом он формирует динамическое поведение системы и реализует концепцию накопителей потоковых диаграмм. По умолчанию в качестве метода численного интегрирования используется прямой метод Эйлера (Forward Euler). Выходной сигнал блока рассчитывается по формуле:

$$y(k) = y(k-1) + T \times u(k-1),$$

где y — выходной сигнал;

u — входной сигнал интегратора;

T — шаг дискретизации;

k — номер шага моделирования.

Элемент *MATLAB Function* (функционал) используется для реализации функциональных зависимостей, определяющих взаимодействие элементов модели между собой и соответствующих концепции потоков системно-динамической модели. При этом функция описывается в редакторе данного блока кодом на языке MATLAB.

Элемент *From Workspace* (элемент рабочей области) передает заданную в рабочей области MATLAB величину, например вспомогательную переменную, на входящие потоки компонент модели. При этом вспомогательная переменная становится входным сигналом для принимающей компоненты.

Элемент *Digital Clock* (часы) формирует дискретный временной сигнал, определяющий шаг моделирования.

Элемент *Delay* (задержка) позволяет задавать задержку входного сигнала.

Элемент *Scope* отвечает за построение графиков, что дает возможность наблюдать за изменениями входящих и исходящих сигналов. Результат работы этого блока можно сохранить в переменную рабочей области MATLAB.

Элемент *Sum* (сумматор) реализует операции сложения и вычитания входящих в него значений сигналов.

Элемент *Constant* (константа) задает постоянный во времени сигнал, по сути реализуя константную функцию.

Элемент *Product* (мультипликатор) реализует операции умножения и деления входящих значений сигналов.

Применение MATLAB Simulink для построения моделей оценки кредитного риска корпоративных заемщиков

Элемент *Gain* (множитель) умножает входной сигнал на фиксированный постоянный коэффициент.

Пример системно-динамической модели корпоративного заемщика в MATLAB Simulink

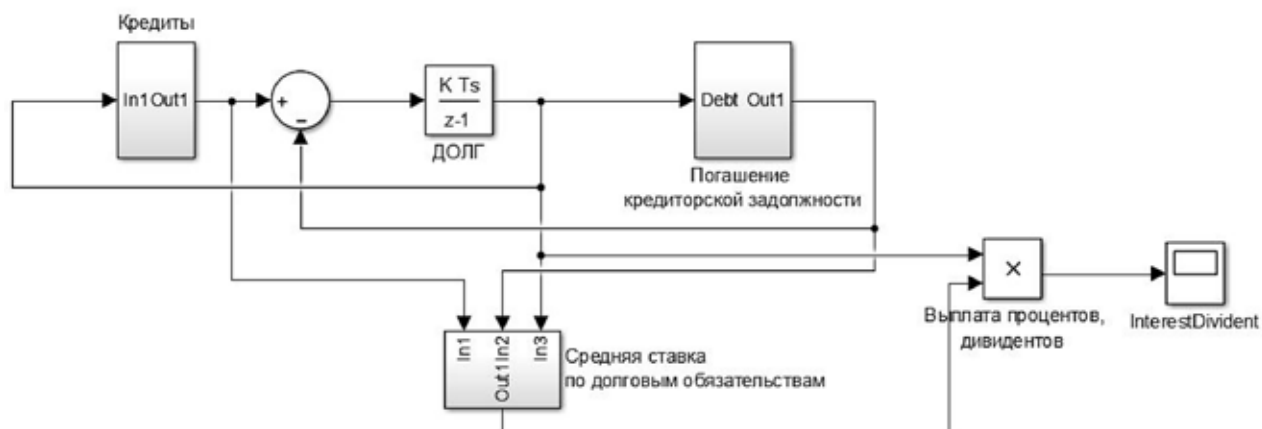
Объектом моделирования для нас будет нефтедобывающее и нефтеперерабатывающее предприятие. Макроэкономические факторы выступают в качестве входных переменных (сигналов), подаваемых на вход модели.

Модель строится на основе анализа финансовой отчетности и данных из открытых источников информации. При моделировании ставится цель воспроизвести структуру и динамику предприятия. Симуляция осуществляется на заданном временном горизонте, на котором оцениваются процессы в различных элементах модели. В результате делается вывод о ее устойчивости.

Структура модели отражена на рис. 4. Отдельно в качестве подсистем на рис. 5 и 6 представлены производственная и финансовая части, связанные с основной моделью блоками Subsystem. Стрелки указывают на наличие и направление связей между элементами модели.

Рисунок 4

Системно-динамическая модель нефтедобывающей и нефтеперерабатывающей компании в MATLAB Simulink



Дмитрий КУРЕННОЙ

Основой производственной части является интегратор «Запас нефти», для которого входными пополняющими сигналами (им соответствуют знаки «+» в сумматоре) служат объем добычи и объем закупок. Эти сигналы задаются простыми подсистемами, состоящими из функционалов, которые зависят от макроэкономических факторов (цены на нефть и нефтепродукты), и элементов задержки.

Истощают «Запас нефти» сигналы (им соответствуют знаки «-» в сумматоре) «Переработка» и «Продажа», определяемые простыми функционалами, связанными с макроэкономическими факторами.

Таким образом, запасы нефти формируются путем добычи и закупки, а используются в основном для переработки и отчасти для прямой продажи. Объем перерабатываемой нефти аккумулируется в соответствующем интеграторе и в процессе переработки распределяется по интеграторам получаемых из нее нефтепродуктов. После этого нефть для продажи и произведенные нефтепродукты в разном соотношении подаются на вход подсистемы, реализующей экспорт и соответствующую долларовую прибыль, и подсистемы, реализующей продажу продукции на внутреннем рынке и соответствующую рублевую прибыль.

Полученная в результате этого процесса прибыль выступает одним из входных сигналов ключевого интегратора всей системно-динамической модели компании — «Средства в рублях». Этот интегратор позволяет увидеть динамику средств, которыми располагает компания, и определить ее устойчивость к макроэкономическим факторам, которые выступают входными сигналами в большинстве перечисленных функционалов.

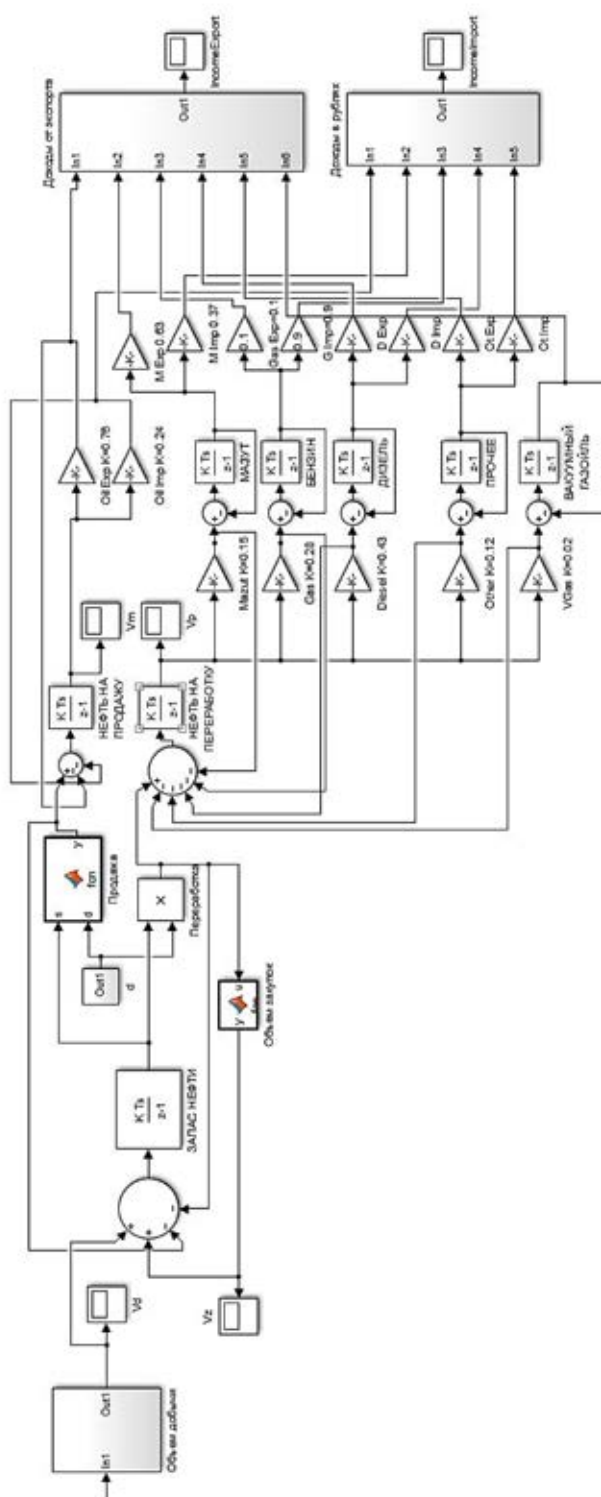
Динамическое поведение финансовой части модели задается интегратором, отражающим долг компании. В сумматор со знаком «+» передается подсистема «Кредиты» (привлекаются для обеспечения деятельности компании), со знаком «-» передается подсистема «Погашение кредиторской задолженности». В результате долг пополняется кредитами и истощается погашением кредиторской задолженности. При этом процесс погашения регулируется макроэкономическим фактором (ставка MosPrime), влияющим на динамику подсистемы «Средняя ставка по долговым обязательствам». Результат финансовой части модели напрямую влияет на количество средств, которыми располагает компания, а следовательно, регулирует ключевой интегратор «Средства в рублях».

Описанные производственная и финансовая части постоянно взаимодействуют. К примеру, объем прибыли определяет средства, которые могут быть потрачены на закупку нефти или инвестированы

Применение MATLAB Simulink для построения моделей оценки кредитного риска корпоративных заемщиков

Рисунок 5

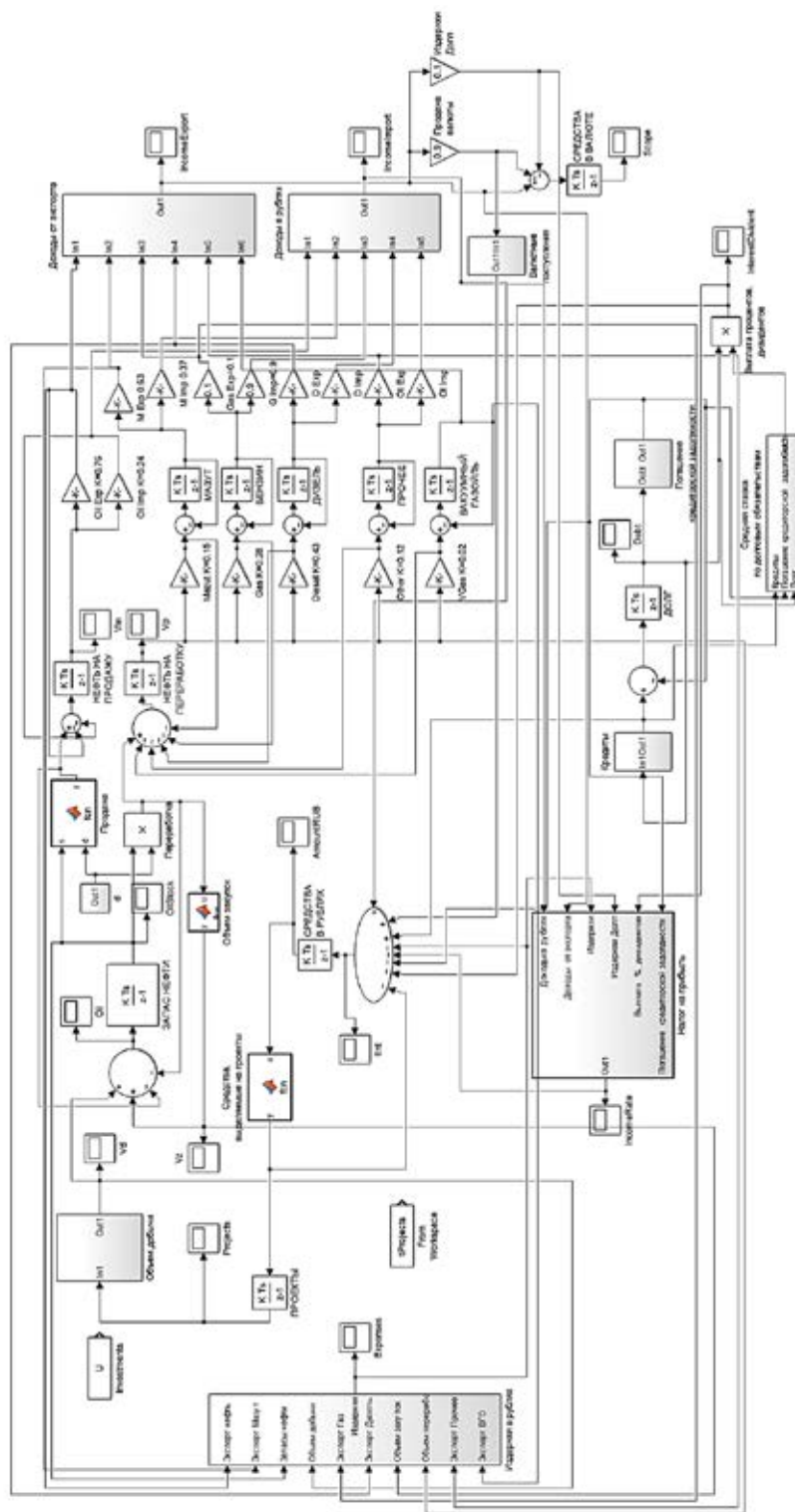
Производственная часть системно-динамической модели нефтедобывающей и нефтеперерабатывающей компании в MATLAB Simulink



Дмитрий КУРЕННОЙ

Рисунок 6

Финансовая часть системно-динамической модели нефтедобывающей и нефтеперерабатывающей компании в MATLAB Simulink



Применение MATLAB Simulink для построения моделей оценки кредитного риска корпоративных заемщиков

в увеличение добычи. Ключевой интегратор «Средства в рублях» аккумулирует в себе все процессы и величины, оказывающие влияние на средства компании. Они включают в себя издержки, истощающие интегратор: себестоимость производства и добычи, затраты на приобретение нефти, выплату налогов, таможенные пошлины, коммерческие, общехозяйственные и административные расходы, прочие расходы, выплату процентов, погашение кредиторской задолженности, транспортные расходы. Прибыль от производственной деятельности и привлекаемые кредиты пополняют «Средства в рублях».

Большинство указанных элементов заданы функциональными зависимостями от макроэкономических факторов: экспортных и внутренних цен на нефть, курса доллара к рублю, ставки привлекаемых и погашаемых кредитов MosPrime, базовой ставки на добычу полезных ископаемых, удельных себестоимостей общехозяйственных расходов, добычи и переработки нефти.

Динамика системно-динамической модели в MATLAB Simulink исследуется путем мониторинга изменения состояния ключевых интеграторов на заданном временном горизонте. Для прогноза устойчивости и оценки вероятности дефолта в качестве макроэкономических факторов подаются соответствующие прогнозные величины, а оценкой состояния компании может выступать значение основного интегратора «Средства в рублях». 