

---

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

---

В предыдущих двух статьях мы описали этапы определения целей, подготовки данных и разработки «ядра» модели LGD. В этой статье разберем последние три этапа цикла разработки — калибровку «ядра» модели, расчет надбавок консерватизма и оценку качества финальной модели — и расскажем, как проанализировать влияние оценок LGD на портфель.

**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

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

### Калибровка модели

В соответствии с Положением № 483-П<sup>1</sup> оценка уровня потерь при дефолте должна соответствовать долгосрочному среднему значению. Для этого проводится калибровка модели, включающая в себя следующие этапы:

- подготовку выборки для калибровки;
- расчет долгосрочного среднего значения (Long Run Average Loss Given Default, LRA LGD);
- проведение калибровки.

Для моделей LGD “Will-Default” и LGD “In-Default” применяется единый подход. Для оценки целевой переменной при калибровке

---

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

## Сергей АФАНАСЬЕВ и др.

модели используется целевое значение LGD, рассчитанное по формуле<sup>1</sup>:

$$LGD_i \cong 1 - \frac{\sum tran_i^{disc}}{EAD_i}, \quad (1)$$

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

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

### Подготовка выборки для калибровки

В выборку для калибровки необходимо включать все доступные наблюдения, в том числе открытые дефолты, которые не были использованы для построения «ядра» модели, так как по ним ожидаются материальные платежи, а целевые значения не могли быть зафиксированы. Поэтому для расчета долгосрочного среднего значения требуется расчет прогноза денежных потоков по данным кредитам. При этом в выборку для калибровки необходимо включать данные за период длиной не менее пяти лет, охватывающие полный экономический цикл. Для определения экономического цикла могут использоваться разные методы. Далее описан один из них.

В выборку, используемую для расчета долгосрочного среднего значения, необходимо включать данные за период, охватывающий полный цикл деловой активности, который включает в себя как минимум один период между пиками экономического спада. Спад определяется путем анализа макроэкономических показателей.

Алгоритм для определения экономического спада:

- 1) сначала осуществляется экспертный выбор макроэкономических факторов в соответствии с модельным сегментом;
- 2) затем строятся графики изменения макроэкономических факторов;
- 3) потом вычисляется тренд, который позволяет определить границы изменения макроэкономических условий.

Период спада определяется с помощью проведения линии отсечения в пределах определенных ранее границ изменения фактора. Такой анализ проводится для всех рассматриваемых макроэкономических факторов, в результате выбираются пики экономического спада. Если для получения более консервативных результатов

Алгоритм для определения экономического спада: (1) экспертный выбор макроэкономических факторов в соответствии с модельным сегментом; (2) построение графиков изменения макроэкономических факторов; (3) вычисление тренда, который позволяет определить границы изменения макроэкономических условий.

<sup>1</sup> Определение моделей "Will-Default" и LGD "In-Default" и методика расчета целевой переменной приводились в первой части статьи (Риск-менеджмент в кредитной организации. 2021. № 3).

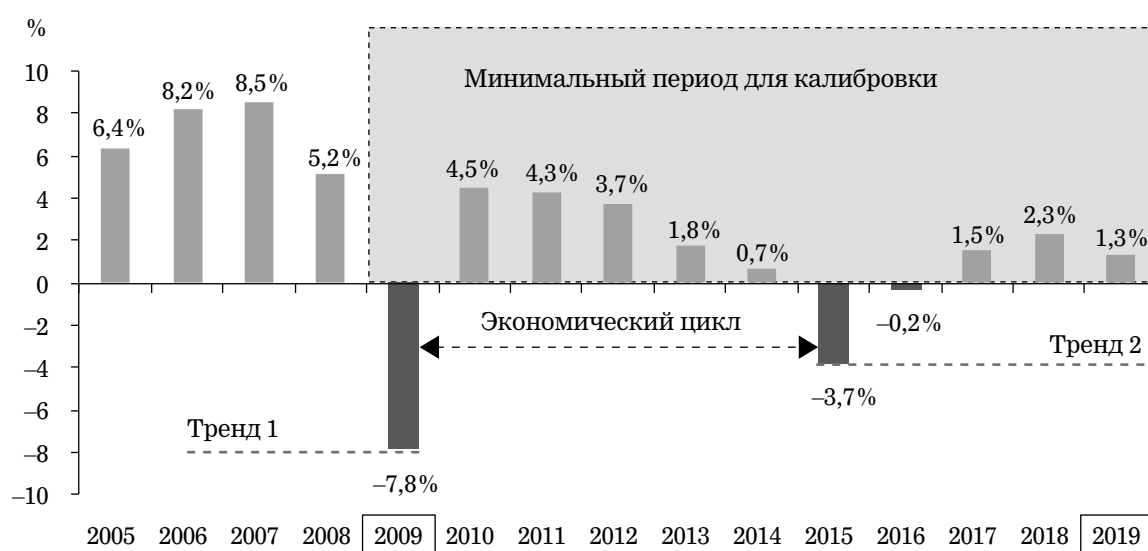
## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

используются несколько факторов, для определения начала цикла выбирается самая ранняя точка пика первого спада, окончания — последняя точка пика спада.

По результатам такого анализа определяется временной горизонт между двумя пиками спада, который необходимо использовать как минимальный период формирования выборки для калибровки модели LGD. На рис. 1 представлен пример определения границ экономического цикла через анализ темпов роста ВВП России.

Рисунок 1

### Пример определения экономического цикла через анализ темпов роста ВВП России



Если у банка недостаточно исторических данных для охвата минимального периода калибровки модели, возможно несколько подходов к анализу данных об уровнях дефолтов:

- 1) на основе собственных данных (наиболее точный метод);
- 2) на основе оценочных данных: если для периода экономического спада нет достаточного количества данных, возможно применение метода экстраполяции.

Обоснование подхода определения периода для калибровки, а также выбора макроэкономических факторов отражается в отчете о разработке модели LGD.

Сергей АФАНАСЬЕВ и др.

## Расчет долгосрочного среднего значения

Для расчета долгосрочного среднего значения LRA LGD<sup>1</sup> и калибровки модели необходимо рассчитать значение целевой переменной LGD для всех наблюдений.

### Подход "Gross"

В подходе "Gross"<sup>2</sup> будущие денежные потоки прогнозируются для открытых дефолтов на уровне оцененных модельных бакетов в соответствии с шагами, описанными далее.

*Шаг 1.* Все закрытые и «условно закрытые» дефолты распределяются по периодам, в течение которых договор находился в дефолте.

*Шаг 2.* Рассчитывается количество дефолтов в каждом винтаже, чтобы определить вероятность дожития дефолта до  $j$ -го винтажа:

$$n_{i,j} = \frac{N_j}{N_i}, \quad (2)$$

где  $n_{i,j}$  — вероятность дожития до  $j$ -го винтажа при нахождении на  $i$ -м винтаже;

$N_j$  — количество открытых дефолтов на  $j$ -м винтаже;

$N_i$  — количество открытых дефолтов на  $i$ -м винтаже.

Винтажи для списанных кредитов учитываются по времени нахождения в дефолте без учета даты списания.

*Шаг 3.* Рассчитываются маргинальные уровни возмещения (доли погашений от оставшегося долга) для каждого винтажа:

$$\text{MRR}_{k,j} = \frac{R_{k,j}}{\text{EAD}_k - \sum_{j=1}^{j-1} R_{k,j}}, \quad (3)$$

где  $\text{MRR}_{k,j}$  — маргинальные уровни возмещения для  $k$ -й статистической единицы  $j$ -го винтажа;

$R_{k,j}$  — сумма денежных потоков для  $k$ -й статистической единицы на  $j$ -м винтаже, включая прямые расходы и дополнительные снятия и исключая баланс при выздоровлении;

$\text{EAD}_k$  — сумма задолженности для  $k$ -й статистической единицы на дату начала дефолта.

<sup>1</sup> LRA LGD (Long Run Average Loss Given Default) — долгосрочное среднее значение уровня потерь при дефолте.

<sup>2</sup> Подходы "Gross" и "Cure/Loss Rate" описаны во второй части статьи (Риск-менеджмент в кредитной организации. 2021. № 4).

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

Расчет  $MRR_{k,j}$  проводится для всех закрытых и «условно закрытых» дефолтов, после чего полученные результаты усредняются:

$$\overline{MRR}_j = \frac{\sum_k^{k=n} MRR_{k,j}}{n}, \quad (4)$$

где  $\overline{MRR}_j$  — среднее значение MRR для  $j$ -го винтажа;

$n$  — количество наблюдений, использованных для усреднения MRR по  $j$ -му винтажу.

**Шаг 4.** Рассчитываются доли погашений от оставшегося долга с учетом вероятности дожития для каждого винтажа:

$$SRR_{i,j} = \overline{MRR}_j \times n_{i,j}, \quad (5)$$

где  $SRR_{i,j}$  — доля погашений от оставшегося долга в  $j$ -м винтаже с учетом вероятности дожить до  $j$ -го винтажа, находясь в  $i$ -м винтаже.

В результате формируется матрица, представленная в табл. 1.

Таблица 1

### Пример матрицы с расчетом доли погашений от оставшегося долга с учетом вероятности дожития

Винтаж	Винтаж в дефолте				
	1	2	3	...	T
1	$SRR_{1,1}$	$SRR_{1,2}$	$SRR_{1,3}$	...	$SRR_{1,T}$
2		$SRR_{2,2}$	$SRR_{2,3}$	...	$SRR_{2,T}$
3			$SRR_{3,3}$	...	$SRR_{3,T}$
...	...	...	...	...	...
T					$SRR_{T,T}$

**Шаг 5.** Рассчитываются денежные потоки для каждого открытого дефолта по тем винтажам, по которым неизвестна информация о платежах:

$$REC_j = \prod_{l=v}^{j-1} [(1 - SRR_l) \times SRR_j \times \widetilde{EAD}], \quad (6)$$

где  $REC_j$  — денежные потоки для открытого дефолта за  $j$ -й винтаж;

$SRR_l$  — доля погашений от оставшегося долга в  $l$ -м винтаже с учетом вероятности дожития до  $l$ -го винтажа для открытого дефолта;

$v$  — винтаж (срок нахождения в дефолте на последнюю дату наблюдения);

$\widetilde{EAD}$  — задолженность на дату последнего доступного винтажа  $v$ .

## Сергей АФАНАСЬЕВ и др.

Для учета полученных возмещений и срока, проведенного в дефолте на последнем наблюдаемом винтаже, необходимо обобщить формулу (6). Итоговая формула для денежных потоков в периоде, равном винтажу, имеет вид:

$$REC_v = (\overline{EAD} - R) \times SRR_v \times \frac{m-t}{m}, \quad (7)$$

где  $R$  — платежи, совершенные в последнем доступном винтаже;

$t$  — период нахождения контракта в дефолте в последнем доступном периоде (в днях);

$m$  — количество дней в винтаже.

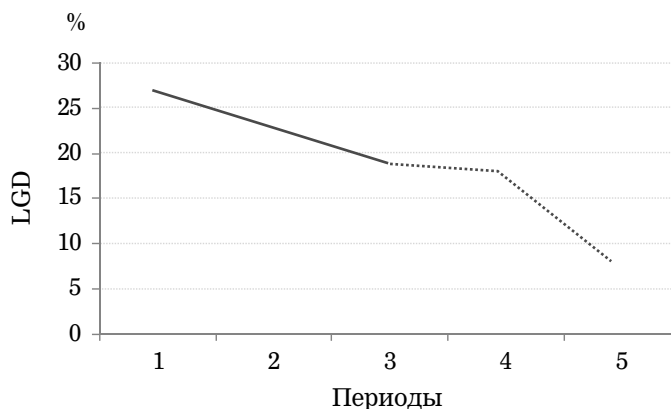
Таким образом, полная формула расчета денежных потоков для всех периодов, идущих после периода, равного винтажу, имеет вид:

$$REC_j = \prod_{l=v+1}^{j-1} [(1 - SRR_l)] \times SRR_j \times \left( (\overline{EAD} - R) \times (1 - SRR_v \times \frac{m-t}{m}) \right). \quad (8)$$

**Шаг 6.** Итоговый LGD с учетом дисконтированных денежных потоков, полученных на предыдущем шаге, рассчитывается для каждого открытого дефолта (рис. 2).

Рисунок 2

### Пример «достроения» целевого значения LGD для открытых дефолтов



#### Подход "Cure/Loss Rate"

При использовании подхода "Cure/Loss Rate" калибровка модели производится в два этапа.

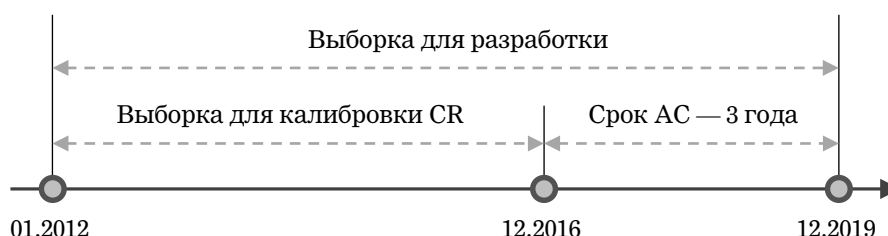
## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

**Шаг 1.** Модель CR калибруется с целью соответствия предсказываемых значений CR долгосрочному среднему уровню выздоровлений. Для оценки долгосрочного значения CR используется выборка для разработки модели LGD, ограниченная таким образом, чтобы флаг выздоровления мог быть проставлен однозначно. Для этого используется временной горизонт от начальной даты выборки до самой поздней даты, с которой открытые дефолты признавались искусственно закрытыми.

На рис. 3 приведен пример для выборки, собранной за период с 2012 г. по 2019 г., со сроком искусственного закрытия три года.

Рисунок 3

### Пример определения периода для расчета долгосрочного среднего значения CR



**Примечание:** срок АС — срок искусственного закрытия (Artificially Closed).

Долгосрочный средний уровень выздоровлений вычисляется по формуле:

$$\overline{CR} = \frac{\sum_{i=1}^n CR_i}{n}, \quad (9)$$

где  $\overline{CR}$  — долгосрочный средний уровень выздоровлений;

$CR_i$  — среднегодовой уровень выздоровлений  $i$ -го года, определяемый как доля выздоровевших дефолтов среди всех вышедших в дефолт на  $i$ -м году;

$n$  — количество лет в периоде выборки для расчета центрального значения.

После проведенных расчетов принимается одно из двух решений:

1. Используется результат модели в непрерывной форме.
2. Результаты модели разделяются на несколько бакетов (рекомендуется не более семи бакетов) с помощью одного из подходов:

## Сергей АФАНАСЬЕВ и др.

1) применение метода *k-means*:

- параметром кластеризации является модельное значение вероятности выздоровления;
- оптимальное количество бакетов определяется экспертно;
- для контроля результатов может быть использован показатель  $R^2$ , оцениваемый через сравнение фактического уровня выздоровлений и центральных значений сформированных бакетов;

2) применение метода *дерева решений*:

- целевой переменной является флаг выздоровления, а объясняющим фактором — модельная оценка вероятности выздоровления;
- проводится максимизация коэффициента Джини:

$$\text{Gini} = 1 - \sum_k p_k^2 \rightarrow \max, \quad (10)$$

где  $p_k$  — доля наблюдений в  $k$ -м бакете;

— определяется оптимальная глубина дерева с помощью кросс-валидации;

— по результатам построения дерева решений некоторые бакеты объединяются с целью исключения групп с малым числом наблюдений и получения экспоненциального роста CR от разряда к разряду.

После оценки центральных значений CR проводится оценка верхних границ. Верхние границы рассчитываются как среднее геометрическое соответствующих центральных значений по формулам:

$$\text{CR}_k^{\text{upper}} = \sqrt{\overline{\text{CR}}_k \times \overline{\text{CR}}_{k+1}}; \quad (11)$$

$$\text{CR}_k^{\text{upper}} = \text{CR}_{k+1}^{\text{lower}}, \quad (12)$$

где  $\overline{\text{CR}}_k$  — центральное значение CR для  $k$ -го класса;

$\text{CR}_k^{\text{upper}}$  — значение верхней границы  $k$ -го класса;

$\text{CR}_k^{\text{lower}}$  — значение нижней границы  $k$ -го класса.

**Шаг 2.** Пересчитываются модельные значения LGD с учетом открытых дефолтов, которые не были искусственно закрыты. Оценка LGD по этим дефолтам проводится по такому алгоритму:

1. К открытым дефолтам, которые не были искусственно закрыты, применяется модель LGD “In-Default”: для статистической единицы оценивается модельное значение LGD, соответствующее фактическому винтажу этой статистической единицы на самую позднюю дату наблюдения в выборке.



## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

2. Для каждой  $i$ -й статистической единицы рассчитывается ожидаемый последний платеж на последний доступный срез выборки для разработки:

$$CF_{i,T} = EAD_{i,T} (1 - LGD_{model,i}), \quad (13)$$

где  $LGD_{model,i}$  — модельное значение LGD “In-Default” для  $i$ -й статистической единицы;

$T$  — последний доступный срез выборки для разработки;

$EAD_{i,T}$  — значение задолженности для  $i$ -й статистической единицы на дату  $T$ ;

$CF_{i,T}$  — ожидаемый денежный поток на дату искусственного закрытия  $T$ .

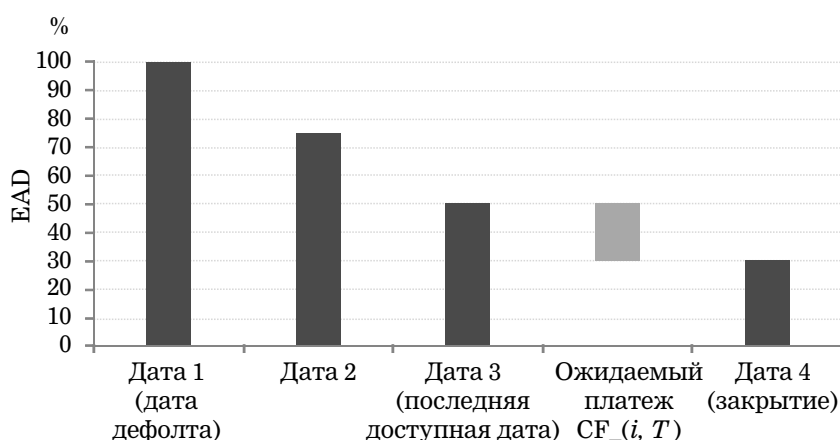
3. Полученный ожидаемый денежный поток  $CF_{i,T}$  используется для оценки набора целевых значений LGD по  $i$ -й статистической единице:

- для LGD “Will-Default” рассчитывается одно целевое значение;
- для LGD “In-Default” рассчитывается набор целевых значений по необходимым винтажам для  $i$ -й статистической единицы.

Полученные целевые значения используются для пересчета модельных значений и расчета долгосрочных средних значений LGD. На рис. 4 приведен пример учета  $CF_{i,T}$  для расчета LGD.

Рисунок 4

### Пример учета ожидаемого денежного потока



При использовании результатов моделей CR или LR в непрерывной форме модельные значения будут также непрерывными. Однако в случае деления результатов модели CR на бакеты, а также при использовании дерева решений или “Look-up” таблицы для построения

Сергей АФАНАСЬЕВ и др.

модели LR, модельные значения будут фиксированными и будут определяться правилами пересечения бакетов LR–CR. Например, если по результатам разработки получено два бакета для LR<sub>c</sub> модели, два бакета для LR<sub>t</sub> модели и четыре бакета для модели CR, то модельные значения будут оцениваться согласно табл. 2.

Таблица 2

**Пример определения модельного значения  
при разделении модели на бакеты**

Бакет LR <sub>c</sub>	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
Бакет LR <sub>t</sub>	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
Бакет CR	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
LGD model, %	15	18	24	30	20	25	37	42	20	23	31	39	26	33	48	55

**Методология калибровки модели LGD**

Калибровка модели LGD применяется относительно целевых переменных модели. Методология калибровки, описанная далее, может использоваться как при первой разработке модели, так и для повторной калибровки. Анализ выполняется на уровне сегментов, выделенных на стадии разработки модели, или с учетом дополнительных калибровочных сегментов.

Если модель построена с помощью дерева решений, “Look-up” таблицы или бакетов LR–CR, то каждое значение LGD на уровне полученных бакетов калибруется с помощью коэффициента количественной оценки риска (*Risk Quantification Factor*). Данный коэффициент определяется как отношение средневзвешенного (по количеству наблюдений) реализованных значений LGD на выборке для калибровки ( $avg\_Realized_{MC}$ ) к средневзвешенному (по количеству наблюдений) оценок LGD на выборке для разработки ( $avg\_Estimated_{MC}$ ):

$$Risk\ Quantification\ Factor_{MC} = \frac{avg\_Realized_{MC}}{avg\_Estimated_{MC}}. \quad (14)$$

Коэффициент количественной оценки риска на уровне компонент модели может быть как больше, так и меньше единицы.

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

Итоговое значение LGD для  $k$ -го бакета определяется по формуле:

$$MC_{\text{Calibrated},k} = MC_k \times \text{Risk Quantification Factor}_{MC,k}, \quad (15)$$

где MC — модельная компонента (*model component*).

Для моделей, построенных с помощью линейной и логистической регрессии, коэффициент количественной оценки риска рассчитывается и применяется на полной выборке без разделения на сегменты.

### Расчет надбавок

Консервативный подход к оценке кредитного риска предполагает, что модельные оценки LGD должны учитывать влияние внешних макроэкономических изменений и внутренних модельных рисков. Для этого рассчитываются следующие надбавки:

- 1) надбавка экономического спада;
- 2) учет косвенных расходов;
- 3) надбавки консерватизма категорий А, В и С.

### Надбавка экономического спада

Модельное значение LGD должно учитывать возможные условия экономического спада, если такая оценка является более консервативной. Это влияние учитывается за счет надбавки экономического спада, расчет которой состоит из пяти этапов:

- 1) макроэкономический анализ;
- 2) определение периода спада;
- 3) анализ восстановлений;
- 4) определение периода для оценки эффекта;
- 5) количественная оценка воздействия спада.

Далее приводится пример расчета надбавки экономического спада в соответствии с концепцией, рекомендованной European Banking Authority (ЕБА)<sup>1</sup>.

### Макроэкономический анализ

Для выявления экономического спада строятся графики с динамикой макроэкономических факторов и рассчитывается тренд, который позволяет определить границы изменения показателя. Возможен анализ следующих факторов<sup>2</sup>:

В рамках консервативного подхода к оценке кредитного риска рассчитываются надбавки: (1) надбавка экономического спада; (2) учет косвенных расходов; (3) надбавки консерватизма категорий А, В и С.

<sup>1</sup> Guidelines for the estimation of LGD appropriate for an economic downturn ('Downturn LGD estimation').

<sup>2</sup> Список не ограничивается перечисленными факторами и может быть изменен или расширен.

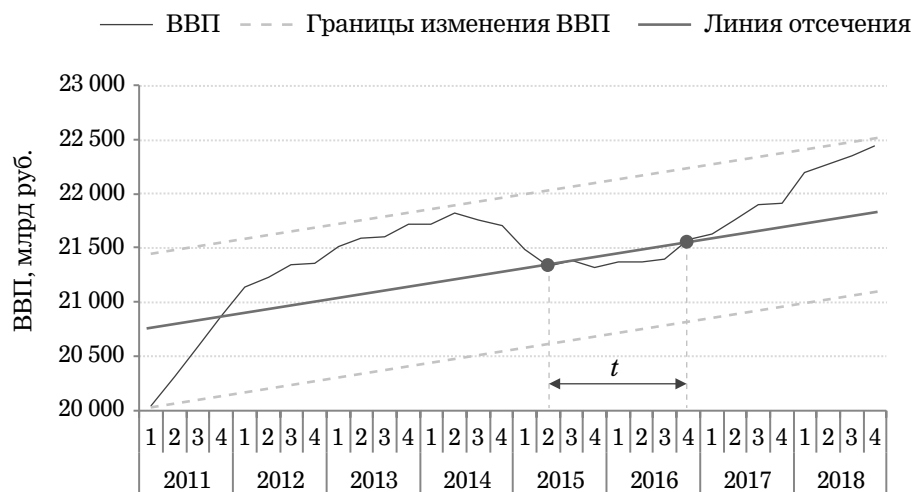
Сергей АФАНАСЬЕВ и др.

- уровень ВВП;
- уровень безработицы;
- рыночный уровень дефолтов;
- индексы цен рынка недвижимости.

Период спада определяется путем проведения линии отсечения в области пика спада. Данный анализ проводится для всех рассматриваемых факторов. По результатам анализа выбирается минимальный период спада, при этом длина спада должна составлять не менее одного года и может быть увеличена для получения более консервативных оценок (рис. 5).

Рисунок 5

### Пример определения периода экономического спада через анализ ВВП России



Линии отсечения для определения периода спада определяются следующим образом:

- 1) строится линейный тренд для макроэкономического фактора;
- 2) строятся линии максимума и минимума (границы изменения макроэкономического фактора) путем параллельного переноса линии тренда до момента касания с графиком макроэкономического фактора;
- 3) далее линии максимума и минимума взвешиваются (пропорция определяется экспертно). Полученная прямая является линией отсечения, а длина отрезка, образованного пересечением с графиком макроэкономического фактора, является периодом спада  $t$ .

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

### Анализ динамики сборов и определение периода спада

После определения периода экономического спада  $t$  необходимо проанализировать значения LGD с целью выявления периода, на котором будет оцениваться эффект влияния спада на значения LGD.

Для этого проводится анализ динамики сборов с целью определения периода, в котором наблюдалось максимальное значение LGD. На горизонте доступных данных после начала экономического спада рассчитывается набор средних значений LGD за периоды длины  $t$  для статистических единиц, которые вышли в дефолт в данных периодах (периоды расчета сдвигаются на месяц).

Таким образом, от даты начала периода спада строится динамика скользящих средних значений LGD по временным промежуткам, длина которых равна  $t$ :

$$\overline{\text{LGD}}_{t_j} = \frac{\sum_{i \in t_j} \text{LGD}_i}{n}, \quad (16)$$

где  $\overline{\text{LGD}}_{t_j}$  — среднее значение LGD для статистических единиц, начало дефолта по которым попадает в интервал  $t_j$ ;

$\text{LGD}_i$  — LGD для  $i$ -й статистической единицы, начало дефолта по которой попадает в интервал  $t_j$ ;

$n$  — количество  $i$ -х статистических единиц, начало дефолта по которым попадает в интервал  $t_j$ .

Далее выбирается период  $t_j$ , которому соответствует максимальное значение  $\overline{\text{LGD}}_{t_j}$ . По результатам проведенного анализа определяется период  $\{T_j, T_{j+t}\}$ , на котором будет оцениваться эффект от экономического спада.

### Количественная оценка эффекта

Для учета влияния неблагоприятных экономических событий на модельное значение LGD рассчитывается надбавка экономического спада. Вне зависимости от методики построения модели LGD на данном этапе проводится расчет коэффициента корректировки для учета экономического спада через отношение средних значений в период спада и на протяжении всего цикла.

Если модель LGD строилась с помощью дерева решений или “Look-up” таблицы, для каждого  $k$ -го бакета пересчитывается среднее значение LGD. Для этого в каждом бажете для наблюдений, попавших в период  $[T_j, T_{j+t}]$ , усредняются значения целевой переменной.

Для учета влияния неблагоприятных экономических событий на модельное значение LGD рассчитывается надбавка экономического спада. Проводится расчет коэффициента корректировки для учета экономического спада через отношение средних значений в период спада и на протяжении всего цикла.

## Сергей АФАНАСЬЕВ и др.

Полученное значение параметра  $\overline{\text{LGD}}_{k\{T_j, T_{j+t}\}}$  является модельным значением параметра кредитного риска после учета фактора макроэкономического спада. Сам фактор макроэкономического спада  $D_k$  для каждого бакета рассчитывается по формуле:

$$D_k(\text{LGD}) = \max \left( 1, \frac{\overline{\text{LGD}}_{k\{T_j, T_{j+t}\}}}{\text{LGD}_k^{\text{EST}}} \right), \quad (17)$$

где  $\text{LGD}_k^{\text{EST}}$  — модельное значение LGD в  $k$ -м бакете.

Для непрерывных модельных предсказаний LGD коэффициент спада применяется как мультипликатор модельного значения и рассчитывается на полной выборке:

$$D(\text{LGD}) = \max \left( 1, \frac{\overline{\text{LGD}}_{\{T_j, T_{j+t}\}}}{\text{LGD}^{\text{EST}}} \right), \quad (18)$$

где  $\overline{\text{LGD}}^{\text{EST}}$  — среднее модельное значение LGD;  
 $\overline{\text{LGD}}_{\{T_j, T_{j+t}\}}$  — среднее значение LGD на периоде спада  $\{T_j, T_{j+t}\}$ .

В финальных модельных значениях LGD должны учитываться косвенные расходы, которые могут не относиться к конкретному кредитному требованию.

### Учет косвенных расходов

В финальных модельных значениях LGD должны учитываться косвенные расходы, которые могут не относиться к конкретному кредитному требованию.

Учет косвенных расходов проводится после применения надбавки экономического спада, поскольку косвенные расходы не могут быть учтены при расчете целевой переменной. На основе принципов существующей системы учета расходов все материальные косвенные расходы распределяются по соответствующим кредитным требованиям.

Косвенные расходы могут аллоцироваться на различные подразделения и источники:

- расходы юридического подразделения;
- расходы подразделения взыскания;
- расходы на ИТ-системы;
- ФОТ департамента ИТ;
- ФОТ фронт-офиса;
- и др.

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

Уровень косвенных расходов может быть оценен по формуле:

$$IC = \frac{\sum \frac{Cost_i}{EAD_{NPL}}}{N} \times Share, \quad (19)$$

где  $\sum Cost_i$  — сумма косвенных расходов по продукту;

$EAD_{NPL}$  — уровень NPL задолженности по продукту на начало года;

$N$  — количество периодов на рассматриваемом временном интервале;

$Share$  — доля продукта в кредитном портфеле.

Также для учета косвенных расходов можно использовать взвешенные по величине NPL портфеля значения расходов и другие статистические методы.

### Надбавки консерватизма

В соответствии с Положением № 483-П для минимизации недооценки компонентов кредитного риска при использовании моделей количественной оценки риска на основе ПБР банк применяет консервативный подход, состоящий в применении надбавок к рассчитанным компонентам кредитного риска и (или) величине кредитного риска<sup>1</sup>. Согласно рекомендациям ЕБА<sup>2</sup>, необходимо оценивать надбавки консерватизма следующих категорий:

- *категория А*: выявленные недостатки данных и методологии;
- *категория В*: изменения в стандартах андеррайтинга, в аппетите к риску, в политиках сбора и возмещения и в любом другом источнике дополнительной неопределенности;
- *категория С*: стандартная ошибка модели.

Итоговая надбавка консерватизма рассчитывается как сумма надбавок категорий А, В и С:

$$LGD_{i \text{ after MOC}} = LGD_i \times (1 + MOC_A + MOC_B + MOC_C), \quad (20)$$

где  $LGD_{i \text{ after MOC}}$  — итоговая оценка уровня потерь при дефолте для наблюдения  $i$  после применения надбавок консерватизма;

$LGD_i$  — оценка уровня потерь при дефолте для наблюдения  $i$  до применения надбавок;

$MOC_A$  — сумма надбавок консерватизма<sup>3</sup> категории А;

Надбавка консерватизма категории А применяется для устранения неопределенности, которая может возникнуть в связи с недостатками в качестве данных и методологии.

<sup>1</sup> Главы 1, 1.20 Положения № 483-П.

<sup>2</sup> EBA Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures. EBA/GL/2017/16. European Banking Authority. 20.11.2017. Chapter 4.4. P. 57.

<sup>3</sup> MOC — Margin of Conservatism.

## Сергей АФАНАСЬЕВ и др.

МОС<sub>В</sub> — надбавка консерватизма категории В;  
МОС<sub>С</sub> — надбавка консерватизма категории С.

### Расчет надбавки консерватизма категории А

Надбавка консерватизма категории А применяется для устранения неопределенности, которая может возникнуть в связи с недостатками в качестве данных и методологии. Поэтому для принятия решения о необходимости установления надбавки консерватизма проводится анализ качества данных, предсказательной способности и точности калибровки модели. Если результат качественной или количественной оценки неудовлетворительный, может быть установлена надбавка консерватизма категории А. Надбавка консерватизма категории А может состоять из суммы нескольких независимых компонент.

При количественной оценке надбавки консерватизма категории А может быть использован подход сравнения качества исходной модели с качеством консервативной модели, полученной путем внесения корректировок в исходную модель (или в исходную выборку) согласно перечню указанных недостатков исходной модели (выборки). Таким образом, надбавка консерватизма категории А может быть рассчитана по общей формуле:

$$\text{МОС}_A = \max \left( 0; \frac{\text{Quality}_{\text{Corrected}}}{\text{Quality}_{\text{Production}}} - 1 \right), \quad (21)$$

где  $\text{Quality}_{\text{Corrected}}$  — качество консервативной (скорректированной) модели, полученной после внесения корректировок в исходную (эксплуатируемую) модель или выборку (согласно указанному перечню выявленных количественных и качественных недостатков исходной модели/выборки);

$\text{Quality}_{\text{Production}}$  — качество исходной модели, внедряемой в эксплуатацию.

Подход количественной оценки, основанный на формуле (21), соответствует бизнес-смыслу, поскольку скорректированная консервативная модель либо улучшает качество предсказания целевого события (тогда к прогнозам эксплуатируемой модели применяется надбавка), либо ухудшает его (тогда надбавка к прогнозам эксплуатируемой модели не применяется).

В качестве показателей  $\text{Quality}_{\text{Corrected}}$  и  $\text{Quality}_{\text{Production}}$  используется метрика качества, отражающая точность предсказания целевого события. Выбор метрики качества зависит от типа задачи и категории оцениваемой надбавки. По умолчанию для расчета надбавки консерватизма категории А рекомендуется использовать показатель



## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

*Powerstat* как межбанковский стандарт оценки качества предсказательной способности модели LGD:

$$\text{МОС}_A = \max \left( 0; \frac{\text{Powerstat}_{\text{Corrected}}}{\text{Powerstat}_{\text{Production}}} - 1 \right). \quad (22)$$

В зависимости от выборок (Test, OOT, Validation), на которых рассчитываются скорректированные показатели *Powerstat*, надбавка консерватизма категории А может быть двух типов:

1) надбавка за снижение устойчивости модели (in-time): *Powerstat*<sub>Production</sub> рассчитывается на исходной тестовой выборке, а *Powerstat*<sub>Corrected</sub> — на скорректированной тестовой выборке (in-time);

2) надбавка за ухудшение стабильности модели: *Powerstat*<sub>Production</sub> рассчитывается на исходной тестовой выборке, а *Powerstat*<sub>Corrected</sub> — на скорректированной валидационной выборке (out-of-time).

В случаях, когда нет возможности рассчитать стандартную метрику качества модели, в качестве показателей для оценки изменений могут выбираться характеристики портфеля или модели:

1) фактические значения LGD по портфелю — когда нет возможности рассчитать прогнозные значения LGD для портфеля (например, когда нет полных данных для оценки LGD);

2) прогнозные значения LGD — когда недоступны фактические значения целевой переменной LGD (например, когда дефолты не «вызрели»).

В остальных случаях при расчете надбавки консерватизма категории А необходимо использовать метрику для оценки качества модели, поскольку надбавка применяется к прогнозным значениям за ухудшение качества модельных предсказаний, которое может быть оценено только при наличии двух компонент — прогнозного значения LGD и фактического значения LGD.

### Расчет надбавки консерватизма категории В

Надбавка консерватизма категории В применяется для устранения неопределенности, которая может возникнуть в связи с изменениями внутренней и внешней среды (например, политик, стратегии банка и т.д.).

Надбавка консерватизма категории В может состоять из суммы нескольких компонент:

1) надбавки, связанные с изменением внутренней среды (например, стратегии банка);

Надбавка консерватизма категории В применяется для устранения неопределенности, которая может возникнуть в связи с изменениями внутренней и внешней среды (например, политик, стратегии банка и т.д.).

## Сергей АФАНАСЬЕВ и др.

2) надбавки, связанные с изменением внешней среды (экономическими кризисами, изменением клиентского потока или рыночной конъюнктуры и др.).

Важным критерием для суммирования компонент надбавки является их независимость: нужно исключить дублирование коррелированных компонент.

При количественной оценке надбавки консерватизма категории В может быть использован подход сравнения качества модели, рассчитанной на исходных данных, с качеством модели, рассчитанной на измененных данных, полученных путем внесения корректировок в исходную выборку (согласно перечню выявленных недостатков)<sup>1</sup>:

$$\text{МОС}_B = \max \left( 0; \frac{\text{Quality}_{\text{old\_sample}}}{\text{Quality}_{\text{correct\_sample}}} - 1 \right), \quad (23)$$

где  $\text{Quality}_{\text{correct\_sample}}$  — качество консервативной модели, полученной после внесения корректировок в исходную (некорректную) выборку;

$\text{Quality}_{\text{old\_sample}}$  — качество модели на исходной (некорректной) выборке.

Количественный подход оценки надбавки консерватизма категории В соответствует бизнес-смыслу, поскольку модель на измененных консервативных данных либо ухудшает качество предсказания целевого события (тогда к прогнозам исходной модели применяется надбавка), либо улучшает его (тогда надбавка к прогнозам исходной модели не применяется).

В качестве показателей  $\text{Quality}_{\text{correct\_sample}}$  и  $\text{Quality}_{\text{old\_sample}}$  используется метрика качества, отражающая точность предсказания целевого события (фактического LGD), например *Powerstat*:

$$\text{МОС}_B = \max \left( 0; \frac{\text{Powerstat}_{\text{old\_sample}}}{\text{Powerstat}_{\text{correct\_sample}}} - 1 \right). \quad (24)$$

В зависимости от выборок (Test, OOT, Validation), на которых рассчитываются скорректированные показатели *Powerstat*, надбавки консерватизма категории В могут разбиваться на типы:

1) надбавка за снижение устойчивости модели (in-time):  $\text{Powerstat}_{\text{old\_sample}}$  рассчитывается на исходной тестовой выборке, а  $\text{Powerstat}_{\text{correct\_sample}}$  — на скорректированной тестовой выборке (in-time);

Надбавка консерватизма категории С рассчитывается как стандартная ошибка модели.

<sup>1</sup> В зависимости от ситуации показатели качества модели на старой и новой выборках могут меняться местами в числителе и знаменателе. Важно, чтобы применяемая формула правильно отражала суть изменения качества работы модели в реальных и скорректированных условиях.

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

2) надбавка за ухудшение стабильности модели:  $Powerstat_{old\_sample}$  рассчитывается на исходной тестовой выборке, а  $Powerstat_{correct\_sample}$  — на скорректированной валидационной выборке (out-of-time).

В случаях, когда нет возможности рассчитать стандартную метрику качества модели LGD, для оценки изменений могут выбираться характеристики портфеля или модели, такие как:

1) фактические значения LGD по портфелю — когда нет возможности рассчитать прогнозные значения LGD для портфеля (например, когда нет полных данных для оценки LGD);

2) прогнозные значения LGD — когда недоступны фактические значения целевой переменной (например, когда дефолты не «вызрели»).

В остальных случаях при расчете надбавки консерватизма категории В необходимо использовать метрику для оценки качества модели LGD, поскольку надбавка применяется к прогнозным значениям за ухудшение качества модельных предсказаний, которое может быть оценено только при наличии двух компонент — прогнозного LGD и фактического LGD.

### Расчет надбавки консерватизма категории С

Надбавка консерватизма категории С рассчитывается как стандартная ошибка модели:

$$SE_{LGD} = \frac{S_{LRA\_LGD}}{\sqrt{N}}, \quad (25)$$

где  $SE_{LRA\_LGD}$  — стандартное отклонение фактических значений LGD от долгосрочного среднего значения реализованного LGD (LRA LGD) на выборке для калибровки;

$N$  — количество наблюдений в выборке.

Итоговые модельные оценки необходимо скорректировать на мультипликатор консерватизма, который выражается через надбавку консерватизма:

$$MOC_{SE} = \frac{SE_{LGD}}{LRA\_LGD}, \quad (26)$$

где  $MOC_{SE}$  — мультипликатор консерватизма;

$SE_{LGD}$  — стандартная ошибка модели LGD.

Дискриминационная способность модели LPD оценивается с помощью трех метрик: (1) Loss Shortfall; (2) Powerstat; (3) коэффициента корреляции Спирмена.

### Оценка качества модели

По результатам разработки модели LGD, а также в процессе ее построения необходимо оценивать ее качество. Оценка точности модели LGD проводится после калибровки модели, расчета корректировки

## Сергей АФАНАСЬЕВ и др.

экономического спада, учета косвенных расходов и оценки надбавок консерватизма.

Оценка качества «ядра» модели проводится до калибровки, корректировки на экономический спад, учета косвенных расходов и надбавок консерватизма. Однако оценку ранжирующей способности «ядра» модели следует проводить и после применения всех корректировок, в случае, когда распределение значений финальной модели LGD немонотонно изменилось по сравнению со значениями «ядра» модели.

### Дискриминационная способность модели

Дискриминационная способность модели LPD оценивается с помощью трех метрик<sup>1</sup>:

- коэффициент *Loss Shortfall* — оценивает точность модели;
- показатель *Powerstat* — оценивает предсказательную способность модели;
- коэффициент корреляции Спирмена — оценивает ранжирующую способность модели.

Коэффициент *Loss Shortfall* оценивает точность модели и показывает, насколько общие фактические потери превышают общие смоделированные.

### Коэффициент *Loss Shortfall*

Коэффициент *Loss Shortfall* показывает, насколько общие фактические потери превышают общие смоделированные. Потери определяются для наблюдений, находящихся в состоянии дефолта, как произведение LGD (фактических или смоделированных) и EAD (фактических). Коэффициент *Loss Shortfall* рассчитывается по формуле:

$$\text{Loss\_Shortfall} = 1 - \frac{\sum_{i=1}^n \text{Loss}_i \times \text{EAD}_i}{\sum_{i=1}^n O\text{Loss}_i \times \text{EAD}_i}, \quad (27)$$

где  $O\text{Loss}_i$  — взвешенные по величине задолженности фактические значения LGD в  $i$ -м бакете;

$\text{Loss}_i$  — взвешенные по величине задолженности смоделированные значения LGD в  $i$ -м бакете;

$\text{EAD}_i$  — суммарная фактическая величина кредитного требования в  $i$ -м бакете;

$n$  — количество модельных групп (при использовании “Look-up” таблицы или дерева решений) или количество наблюдений (при использовании линейной регрессии).

<sup>1</sup> Перечисленные метрики часто используются на практике, но не ограничиваются приведенным списком.

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

Модель LGD имеет приемлемую точность, когда общие фактические потери не превышают смоделированные ( $Loss\ Shortfall \leq 0$ ). Если значение коэффициента  $Loss\ Shortfall$  лежит в интервале  $[-0,2; 0]$ , то качество модели LGD рассматривается как удовлетворительное.

Для оценки достоверности рассчитанного показателя (доверительных интервалов) коэффициент  $Loss\ Shortfall$  может быть рассчитан на выборках, сформированных с использованием бутстреп-метода (случайно сгенерированных выборок из исходной выборки).

### Показатель Powerstat

Метрика  $Powerstat$  рассчитывается как отношение площадей под кривыми фактических и модельных убытков<sup>1</sup>. Расчет метрики  $Powerstat$  осуществляется по следующему алгоритму (рис. 6):

1. Для оценки кривой фактических убытков (идеальная модель) наблюдения упорядочиваются по убыванию фактических значений LGD “Will-Default” и LGD “In-Default”. Для каждого наблюдения указывается соответствующая накопленная доля от общего числа наблюдений и от общей суммы фактических значений LGD “Will-Default” и LGD “In-Default”. Далее оценивается площадь ( $S_A + S_B$ ) между построенной кривой фактических убытков и диагональю, соответствующей случайной модели.

2. Для оценки кривой модельных убытков наблюдения упорядочиваются по убыванию модельных значений LGD “Will-Default” и LGD “In-Default”. Для каждого наблюдения указывается соответствующая накопленная доля от общего числа наблюдений и от общей суммы фактических значений LGD “Will-Default” и LGD “In-Default”. Далее оценивается площадь  $S_A$  между построенной кривой модельных убытков и диагональю, соответствующей случайной модели.

3. Показатель  $Powerstat$  рассчитывается как отношение площади  $A$  к площади  $(A + B)$ . Чем больше значение показателя  $Powerstat$ , тем выше предсказательная способность модели.

Для подхода “Cure/Loss Rate” метрика  $Powerstat$  рассчитывается с учетом всех трех компонент  $CR$ ,  $LR_c$  и  $LR_f$ , поскольку компонента  $CR$  вносит отдельный вклад в предсказательную способность модели, который не учитывается при индивидуальной оценке компонент  $LR_c$  и  $LR_f$ .

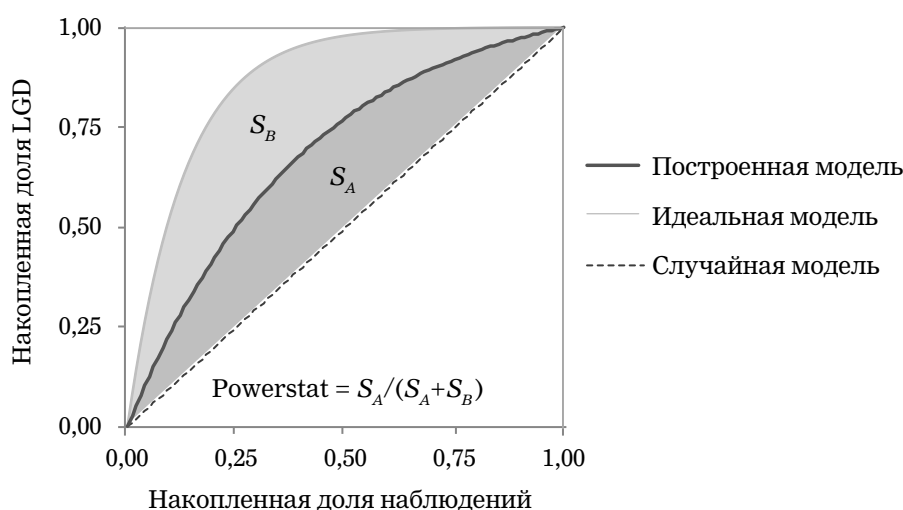
Для оценки предсказательной способности моделей LGD используется метрика  $Powerstat$ , которая рассчитывается как отношение площадей под кривыми фактических и модельных убытков.

<sup>1</sup> Li D., Bhariok R., Keenan S., Santilli S. Validation techniques and performance metrics for loss given default models // Journal of Risk Model Validation. 2009. Vol. 3. No. 3. DOI: 10.21314/JRMV.2009.045.

Сергей АФАНАСЬЕВ и др.

Рисунок 6

### Пример расчета показателя Powerstat через кривые фактических и модельных LGD



С помощью коэффициента корреляции Спирмена оценивается ранжирующая способность модели.

### Коэффициент корреляции Спирмена

С помощью коэффициента корреляции Спирмена оценивается ранжирующая способность модели. Для этого необходимо упорядочить наблюдения по фактическим и прогнозным значениям LGD, после чего каждому наблюдению присваивается порядковый номер в зависимости от фактического и прогнозного значения LGD. Коэффициент корреляции Спирмена рассчитывается по формуле:

$$\rho = 1 - 6 \frac{\sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)}, \quad (28)$$

где  $x_i, y_i$  — порядковые номера наблюдений в соответствии с фактическими и прогнозными значениями LGD соответственно;

$n$  — количество наблюдений в выборке.

Коэффициент корреляции Спирмена может принимать значения от  $-1$  до  $1$ . Если коэффициент равен  $-1$ , то между фактическими и прогнозными значениями существует строго обратная зависимость. Если коэффициент равен  $1$ , то между фактическими и прогнозными значениями существует строгая прямая зависимость. Соответственно чем выше значение коэффициента корреляции Спирмена, тем выше предсказательная способность модели.

## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

### Репрезентативность данных

На разных этапах процесса моделирования могут использоваться не все доступные в банке данные. В таких случаях необходимо проводить анализ репрезентативности данных путем сравнения распределений двух выборок:

1) генеральной совокупности — полной выборки наблюдений, ограниченной выбранным периодом, зависящим от целей задачи (например, выборка для разработки/калибровки, охватывающая период не менее пяти лет с двумя экономическими спадами);

2) подвыборки — выборки, сформированной из генеральной совокупности с помощью одного из методов равновероятностного отбора (Equal Probability of Selection Method, EPSEM): простой случайный выбор, стратифицированный отбор, систематический отбор каждого  $k$ -го по времени наблюдения и др.<sup>1</sup>

Для анализа репрезентативности выборок необходимо выбрать ключевые характеристики, по которым будет проводиться сравнение двух выборок. К таким характеристикам могут относиться социально-демографические показатели и наиболее сильные переменные модели.

В качестве метрики для сравнения двух распределений можно использовать индекс стабильности популяции (Population Stability Index, PSI):

$$PSI = \sum_{i=1}^n (a_i - b_i) \ln \frac{a_i}{b_i}, \quad (29)$$

где  $i$  — порядковый номер интервала/квантиля разбиения;

$n$  — количество интервалов/квантилей разбиения (разбиение проводится по значению LGD);

$a_i$  и  $b_i$  — значения плотностей распределений для выборок  $A$  и  $B$  соответственно.

При использовании стандартных полностью воспроизводимых методов равновероятностного отбора (EPSEM) и достаточного количества наблюдений в выборках (генеральной совокупности и подвыборке), удовлетворяющих критериям NEPV (количество событий на одну переменную), тесты на проверку репрезентативности можно не проводить.

К воспроизводимым методам равновероятностного отбора могут относиться методы, реализованные в открытых библиотеках Python и R, с открытым исходным кодом и документацией, описывающей применяемую методологию.

Внедрение новой модели LGD может оказать существенное влияние на параметры риска, поэтому после разработки модели рекомендуется проводить анализ влияния полученных оценок LGD на изменение взвешенной по риску величины активов (RWA).

<sup>1</sup> Хили Дж. Статистика. Социологические и маркетинговые исследования. 6-е изд. ДиаСофт; СПб.: Питер, 2005. С. 179-184.

Сергей АФАНАСЬЕВ и др.

## Анализ влияния оценок LGD на портфель

Внедрение новой модели LGD может оказать существенное влияние на параметры риска, поэтому после разработки модели рекомендуется проводить анализ влияния полученных оценок LGD на изменение взвешенной по риску величины активов (Risk-Weighted Assets, RWA).

Взвешенная по риску величина активов рассчитывается по формуле<sup>1</sup>:

$$RWA = K_{ПВР} \times EAD, \quad (30)$$

где EAD — сумма кредитных требований, подверженная риску дефолта;

$K_{ПВР}$  — коэффициент риска, рассчитываемый по формулам (31) и (32).

В продвинутом подходе на основе внутренних рейтингов для кредитов, находящихся в дефолте, коэффициент риска рассчитывается по формуле:

$$K_{ПВР} = 12,5 \times \max \{ 0; LGD_{\text{default}} - EL^* \}, \quad (31)$$

где 12,5 — множитель, обратный 8%;

$LGD_{\text{default}}$  — уровень потерь при дефолте, рассчитанный по модели;

$EL^*$  — наилучшая оценка ожидаемых потерь.

Для кредитов, находящихся не в дефолте, коэффициент риска рассчитывается по формуле:

$$K_{ПВР} = 12,5 \times LGD \times \left[ \Phi \times \left( \frac{\Phi^{-1}(PD) + \sqrt{R} \times \Phi^{-1}(\gamma)}{\sqrt{1-R}} \right) - PD \right] \times MA, \quad (32)$$

где 12,5 — множитель, обратный 8%;

$\Phi^{-1}(\gamma)$  — обратная функция стандартного нормального распределения ( $\gamma = 99,9\%$ );

MA (Maturity Adjustment) — коэффициент корректировки на срок погашения: для розничных кредитных требований MA = 1;

R — показатель корреляции, учитывающий вероятность совместных дефолтов различных активов.

Для кредитов, обеспеченных залогом жилого помещения, показатель корреляции  $R = 0,15$ .

Для возобновляемых кредитных требований  $R = 0,04$ .

<sup>1</sup> Главы 3, 5 Положения № 483-П (в ред. от 15.04.2020).



## LGD-модели для розничного кредитования. Часть 3: калибровка, расчет надбавок и оценка качества модели

Для остальных кредитов показатель корреляции рассчитывается по формуле:

$$R = d \times \left[ R_{\min} \times \left( \frac{1 - e^{-k \times PD}}{1 - e^{-k}} \right) + R_{\max} \times \left( 1 - \frac{1 - e^{-k \times PD}}{1 - e^{-k}} \right) \right] - \Delta R, \quad (33)$$

где  $d$  — повышающий коэффициент: для крупнейших финансовых организаций  $d = 1,25$ , для остальных финансовых организаций  $d = 1$ ;

$k$  — коэффициент зависимости корреляции от PD: для розничных кредитов  $k = 35$ , для других кредитов  $k = 50$ ;

$R_{\min}$  и  $R_{\max}$  — минимальный и максимальный пороги корреляции: для розничных кредитных требований  $R_{\min} = 0,03$  и  $R_{\max} = 0,16$ ;

$\Delta R$  — поправка корреляции: для розничных кредитных требований  $\Delta R = 0$ .

Анализ влияния полученных оценок LGD на изменение RWA проводится для текущего портфеля, соответствующего области применения модели LGD.

По результатам построения модели весь процесс разработки документируется (готовится паспорт модели). Дальнейшая работа в рамках цикла ПВР включает в себя:

- 1) валидацию модели;
- 2) аудит разработки и валидации модели;
- 3) внедрение модели;
- 4) мониторинг качества модели;
- 5) перекалибровку модели (не реже одного раза в год);
- 6) перестроение модели (через 2–3 года). 