

# Оценка вероятности отзыва лицензии банка с использованием методов регуляризации

В исследовании используются открытые данные регулятора для предсказания вероятности отзыва банковской лицензии. Применение регуляризации позволяет учесть весь объем данных, что предполагает автоматический выбор оптимальной модели. В статье оценен ряд моделей, демонстрирующих высокую прогнозную силу. Предложенная методика может быть использована участниками финансового рынка для непрерывной оценки надежности контрагентов.

**А. И. ВОТИНОВ,**  
лаборант-исследователь  
Центра перспективного  
финансового планирова-  
ния, макроэкономического  
анализа и статистики  
финансов, НИФИ  
Минфина, лаборант-  
исследователь – сотруд-  
ник Научно-учебной лабо-  
ратории макроструктурно-  
го моделирования эконо-  
мики России, НИУ ВШЭ



**О**ценка вероятности дефолта конкретного экономического агента – важная задача и для самого субъекта, и для его контрагентов. Для успешного планирования деятельности кредитная организация нуждается в корректной оценке собственных рисков. Для контрагентов интерес заключается в минимизации кредитного риска, сопряженного с дефолтами по обязательствам. Разработка моделей, позволяющих прояснить сущность дефолтов, является для экономических субъектов насущной прикладной задачей.

## Подход решает две задачи

Окончание деятельности кредитной организации в отечественной банковской системе связано, скорее, не с фактом дефолта, а с отзывом лицензии со стороны регулятора. При моделировании вероятности этого события следует учитывать, что речь идет не о некотором экономическом процессе, а о принятии решения регулятором, обладающим более подробными данными о состоянии и деятельности банка. Поэтому моделирование вероятности отзыва лицензии с использованием данных из открытой банковской статистики тесно связано с восстановлением той внутренней информации, на основе которой Банк России принимает решение об отзыве лицензии.

**Н. П. ПИЛЬНИК,**  
кандидат экономических  
наук, доцент департа-  
мента прикладной эконо-  
мики факультета эконо-  
мических наук НИУ  
ВШЭ, старший научный  
сотрудник Научно-  
учебной лаборатории  
макроструктурного  
моделирования эконо-  
мики России, НИУ ВШЭ



Большинство методов оценки вероятности отзыва банковской лицензии предполагают использование заранее определенных факторов, потенциально влияющих на финансовое положение отдельно взятого агента. Таким образом, можно выделить две задачи: выявление предсказывающих факторов и прогнозирование с их помощью возможной ситуации. Предложенный ниже подход, основанный на LASSO-регуляризации и техниках машинного обучения, решает обе задачи одновременно.

В статье анализ вероятности отзыва банковской лицензии дается на основе открытых данных регулятора. Современный период истории отечественной банковской системы характеризуется большим количеством отзывов лицензий на осуществление банковской деятельности. Основным стимулом этих действий явилось оздоровление системы в целом, заключающееся в ликвидации слабых и неэффективных участников рынка. Сочетание LASSO-регуляризации и официальных данных Банка России позволяет провести оптимальный отбор факторов и оценить модель. Методология анализа универсальна и может быть применена к любой задаче бинарной классификации.

### Обзор литературы по предсказанию вероятности наступления события

Проблема оценки кредитоспособности контрагентов была актуальна всегда. Одна из самых ранних попыток использовать статистические методы для оценки кредитоспособности была предпринята еще полвека назад [1]. Тогда применялся дискриминантный анализ с целью выявить, какой финансовый показатель в лучшей степени позволяет классифицировать фирмы на обанкротившиеся и нет. В результате было установлено, что наибольшей дискриминантной способностью обладает показатель «отношение денежных потоков к долгу». Существенным недостатком теории была попытка автора объяснить явление дефолта с помощью одной переменной, игнорируя его сложную природу.

Следующим логическим шагом в вопросе оценки кредитоспособности было использование множественного дискриминантного анализа (MDA), предложенного в [2]. Авторы работы оценили модель сразу по 5 различным финансовым показателям на данных 66 различных фирм. Ошибка оцененной классификационной модели составила около 5%, если были использованные данные за год до наступления дефолта. Разумеется, ошибка классифика-

ции возрастала при увеличении временного лага между наблюдениями и событием.

Существенный недостаток дискриминантного анализа заключается в том, что он никак не оценивает вероятность дефолта. Полученные при анализе так называемые Altman's z-score могут принимать любые значения, что позволяет классифицировать банки по принципу «0–1». Эту проблему решают модели бинарного выбора, основанные на логистическом (logit) и нормальном (probit) распределениях. Такой подход дает возможность не только непосредственно оценить вероятность дефолта, но и провести тест Вальда для оценки статистической значимости регрессоров.

Различные примеры использования бинарных моделей можно найти в [3, 4, 5] и др. Так, например, в работе [3] оценена 9-факторная логистическая регрессия. Результатом оценки является так называемый Ohlson O-Score, который по сути является логарифмом отношения шансов (odds ratio). Этот показатель служит альтернативой Altman's z-score и может быть преобразован в оценку вероятности дефолта. Работа [4] посвящена сравнению логистической регрессии и дискриминантного анализа. Выявлено, что логистическое оценивание является более робастным, и при некоторых априорных предположениях оценки дискриминантного анализа оказываются состоятельными и асимптотически эффективными. Работа [5] является в какой-то степени энциклопедической. В ней можно найти не только подробный обзор существующих статистических методов и примеров их использования, но и сравнение их предсказательной силы.

Для регулирования кредитных, операционных, рыночных и прочих рисков существует ряд рекомендаций Базельского комитета по банковскому надзору. Они предполагают выполнение тех или иных нормативов, обеспечивающее финансовую стабильность кредитной организации. За систематическое невыполнение нормативов предусмотрен отзыв регулятором лицензии на осуществление банковской деятельности. Отечественная практика перехода на Базель обладает рядом особенностей [6]. Например, переход к Базелю III осуществляется без окончательного внедрения Базеля II. Также отмечается, что часть отечественных банков просто не готова перейти на новые стандарты управления риском.

В работе [7] оцениваются причины отзыва лицензий на осуществление банковской деятельности с

**Abstract.** In this study open data is used to estimate the probability of license withdrawal. The use of regularization allows taking into account the entire amount of data suggesting automatic selection of the optimal model. A number of predictive models are estimated. The methodology could be implemented by market actors in purpose to continuous monitoring of counterparties' financial stability.

**Keywords.** Bank, license withdrawal, assets, liabilities, balance lists, bank normative, regularization.

**Ключевые слова.** Банк, отзыв лицензии, активы, пассивы, оборотная ведомость, нормативы, регуляризация.

IV квартала 2008-го по II квартал 2015 г. Основные причины отзыва лицензии – финансовая несостоятельность кредитной организации и «отмывание денежных средств». Был выявлено, что микропоказатели балансовой отчетности лучше предсказывают отзыв лицензии за «отмывание денежных средств», а макроэкономические – по причине финансовой несостоятельности. Показано также, что оценка мультиномиальной регрессии не улучшает качество полученных оценок.

В работе [8] оценивается вероятность дефолта кредитных организаций с использованием логистической регрессии. Авторы показали, что включение полиномов от объясняющих переменных позволяет не только учесть нелинейные зависимости между финансовыми показателями, но и значительно улучшить качество прогнозов модели. Также показано, что значимое влияние имеют индекс монопольной власти Лернера и макроэкономические факторы.

Несмотря на ряд преимуществ над дискриминантным анализом, бинарные модели обладают рядом недостатков. Во-первых, природа дефолта сложна и нелинейна. Финансовое состояние отдельно взятого банка может нелинейно зависеть от показателей. Существует множество различных подходов к моделированию дефолтов, которые позволяют исправить этот недостаток. Классические нелинейные алгоритмы включают в себя искусственные нейронные сети, метод опорных векторов, деревья решений, случайные леса, байесовские сети и различные комбинации этих методов. Хороший обзор существующей литературы по классификационным моделям предложен в [9], где, в частности, рассматриваются особенности применения нелинейных моделей.

Во-вторых, стандартные модели бинарного выбора подразумевают заранее обусловленный набор регрессоров. По этой причине модель может быть либо недоопределена и тогда сильно исказит результат, либо переопределена, что может привести к неэффективности оценок. Таким образом, использование заранее определенного набора регрессоров некорректно.

Одним из решений этой проблемы является LASSO<sup>1</sup>-регуляризация. Такой статистический метод предполагает ограничение на параметры модели, которое приводит к самоотбору коэффициентов. В результате в итоговую модель войдет лишь часть переменных, обладающих наибольшей прогнозной силой. В работе [10] представлена модель оценки вероятности дефолта банка, основанная на LASSO-регуляризации и Elastic-net-регуляризации, служащей общим случаем первой.

Наконец, традиционные методы не учитывают структуру банковского сектора. Авторы работы [11] анализируют топологию банковского сектора Российской Федерации и демонстрируют, что дефолт одного банка может запустить целую цепь дефолтов. Схожий результат был получен при анализе банковского сектора Бразилии [12]. Исследования продемонстрировали, что финансовая стабильность ключевых агентов, обладающих наибольшим количеством экономических связей, является более важным фактором систематического риска по сравнению со «средней» финансовой стабильностью. Влияние банковских кризисов на экономику, и на бюджет в частности, описано в работе [13].

Ниже применен подход, являющийся наиболее типичным для машинного обучения. Оценена логистическая регрессия с LASSO-регуляризацией, где количество регрессоров будет на порядок превосходить типичное для классического анализа. Использование такого подхода позволяет идентифицировать показатели, которые наилучшим образом предсказывают отзыв лицензии банков, а не использовать предопределенный заранее набор регрессоров.

### Описание данных

В целях мониторинга и поддержания стабильности банковской системы регулятор России требует от финансовых организаций своевременно предоставлять информацию о текущем состоянии и о деятельности в виде отчетов установленной формы. Часть отчетов в соответствии с письмами Банка России № 165-Т<sup>2</sup>, № 72-Т<sup>3</sup> и № 17-Т<sup>4</sup> на ежемесячной основе (и ежеквартальной для формы 102) публикуются в открытом доступе на сайте ЦБ РФ. Основным преимуществом данных является их полнота и объем. Например, форма 101 содержит около 1800 различных счетов с информацией по входящим и исходящим значениям, а также оборотам по кредиту и дебету в рублях и иностранной валюте.

Остановимся более подробно на содержании указанных форм отчетности.

Форма 101 «Данные оборотной ведомости по счетам бухгалтерского учета» содержит информацию по счетам и оборотам активных и пассивных счетов банков. В отчетах по этой форме для каждого банка можно найти информацию о депозитах, кредитах, резервах и прочих банковских счетах, где в качестве контрагентов могут выступать коммерческие организации, банки по МБК, регулятор, государственные организации, иностранные компании, домохозяйства, НКО, другие финансовые

<sup>1</sup> Least absolute shrinkage and selection operator.

<sup>2</sup> «О раскрытии информации кредитными организациями по формам 0409101 «Оборотная ведомость по счетам бухгалтерской отчетности учета кредитной организации» и 0409102 «Отчет от прибылях и убытках кредитной организации»».

<sup>3</sup> «О раскрытии информации кредитными организациями по формам 0409134 «Расчет собственных средств (капитала)» и 0409135 «Информация об обязательных нормативах»».

<sup>4</sup> «О раскрытии информации кредитными организациями по формам 0409123 «Расчет собственных средств (капитала) (Базель III)»».

организации. Более того, данные можно агрегировать по срокам обязательств – до востребования, до 30/90/180 дней, до 1 года, до 3 лет и более 3 лет. Всего данные по форме 101 можно агрегировать почти по 200 различным показателям, как это было сделано, например, в работе [14].

Форма 102 «Отчет о прибылях и убытках кредитной организации» публикуется ежеквартально и содержит информацию о структуре доходов и расходов банка. В данном отчете можно найти информацию о процентных платежах по кредитам и депозитам различным экономическим агентам, а также о доходах и расходах от другой деятельности банка: результаты операций с ценными бумагами и валютной, различные комиссии и выплаченные налоги.

Форма 134 «Расчет собственных средств (капитала)» отражает состояние собственного капитала кредитной организации. Форма имеет многоуровневую структуру, содержащую подробную информацию о структуре капитала: уставному капиталу, эмиссионному доходу, части нераспределенной прибыли и т. д. Форма 123 «Расчет собственных средств (капитала) (Базель III)» является в каком-то смысле аналогичной форме 134, но соответствует стандартам отчетности Базеля III.

Форма 135 «Информация об обязательных нормативах» содержит данные о нормативах и тех показателях кредитной организации, которые необходимы для расчета нормативов. Значения нормативов регламентируются Инструкцией Банка России № 139-И «Об обязательных нормативах банков» и являются целевыми показателями для регулирования рисков банковской системы.

Используемые для оценки микропоказатели можно разбить на три группы: финансовые, балансовые и оборотные коэффициенты банков. Финансовые коэффициенты включают в себя показатели ROE и ROA, балансовые – долю балансовых агрегатов в общих активах<sup>5</sup>, оборотные коэффициенты – отношение оборотов (входящих, исходящих и всего) по счету к остатку по тому же счету на конец периода.

Банки с низкими и/или отрицательными показателями ROE/ROA обладают менее привлекательным финансовым положением и неэффективным управлением, что может стать дополнительным аргументом в пользу решения об отзыве лицензии. Балансовые коэффициенты отражают, например, структуру кредитного портфеля банка, которая напрямую влияет на риски банка. Резкое увеличение оборотов по тем или иным счетам может свидетельствовать о наличии различных экономических процессов и операций, которые служат индикаторами финансового положения организации. Отметим, что в модели используются логарифмы переменных, где это возможно, чтобы сократить разброс значений.

Для корректного учета влияния показателей, описывающих текущее состояние банка, на веро-

ятность того, что у него будет отозвана лицензия, следует учитывать общее экономическое состояние внешней (для банковской системы) экономической среды. Для этого в модели используются следующие показатели макроэкономики:

■ *темпы прироста ВВП к предыдущему кварталу с учетом сезонной корректировки.* Поскольку данные по ВВП доступны только в квартальной разбивке, в исследовании предполагается, что в течение каждого месяца соответствующего квартала ВВП рос одинаковыми темпами;

■ *индекс промышленного производства.* Этот показатель позволяет учитывать помесечную динамику реального сектора и используется в качестве дополнения к квартальному темпу прироста ВВП;

■ *объем экспорта в долларах.* В условиях сильной зависимости российской экономики от экспортной деятельности наличие такого показателя позволяет учесть в модели влияние этого фактора;

■ *обменный курс доллара к рублю.* Важность этого показателя связана также с существенной долларовой частью в активах и пассивах российской банковской системы;

■ *инфляция (индекс потребительских цен).* Поскольку в исследовании используется большое число номинальных финансовых показателей, для их сопоставления с реальными темпами роста необходима возможность учета ценовой компоненты;

■ *ставка РЕПО как показатель политики, проводимой Банком России.* Этот показатель выбран, чтобы избежать противоречия между показателями ключевой ставки и ставки рефинансирования в период с 2010 г.;

■ *темпы прироста денежной массы M2 по отношению к соответствующему месяцу прошлого года.* Показатель отражает состояние денежной системы экономики, частью которой является и банковская система.

Информация об отзывах лицензий кредитных организаций взята с сайта Банка России. Время наблюдений охватывает период с июня 2010-го по сентябрь 2016 г. Всего за наблюдаемый период было 343 случая отзыва лицензий. Общее число рассмотренных кредитных организаций за весь период – 958, всего около 40 тыс. наблюдений. Если учитывать, что банки с отзывными лицензиями перестают публиковать информацию, соотношение данных становится равным 1:10, т. е. на каждое наблюдение банка с отзывной лицензией приходится 10 наблюдений кредитных организаций, у которых лицензия не отзывалась.

## Методология анализа с помощью LASSO-регуляризации

### Логистическая модель с регуляризацией

Классическая логистическая модель имеет вид:

<sup>5</sup> В качестве прокси для активов используется сумма всех активных балансовых счетов по балансовой отчетности.



$$P(Y = 1|x) = f(\beta^T x), \quad (1)$$

где  $f(\cdot)$  – функция распределения логистического распределения (сигмоид), а  $Y$  – бинарная переменная. Оценка вектора коэффициентов  $\beta$  производится с помощью метода максимального правдоподобия, где функция правдоподобия имеет вид:

$$L = \prod_{i=1}^n P(Y = y_i | X = x_i) = \prod_{i=1}^n f(\beta^T x_i)^{y_i} (1 - f(\beta^T x_i))^{1-y_i}. \quad (2)$$

Чаще всего удобнее оптимизировать логарифмическую функцию правдоподобия следующего вида:

$$\log L = \sum_{i=1}^n \left( y_i \log f(\beta^T x_i) + (1 - y_i) \log(1 - f(\beta^T x_i)) \right). \quad (3)$$

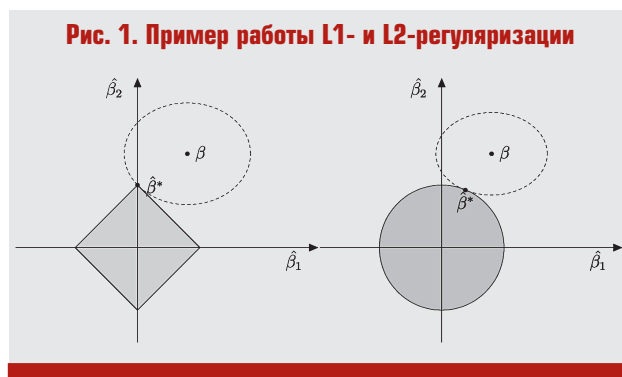
Идея регуляризации заключается в условной максимизации функции правдоподобия, где условие накладывается на параметры модели. Обычно ограничение накладывают в виде L1- и/или L2-нормы, которые известны под названиями LASSO- и RIDGE-регуляризации соответственно. Для получения оценок  $\beta_{LASSO}$  достаточно максимизировать функционал с L1-нормой:

$$\log L_p = \log L - \lambda \|\beta\|_1 = \log L - \lambda \sum_{i=1}^n |\beta_i|. \quad (4)$$

При использовании L2-нормы вместо  $\|\beta\|_1$  используется  $\|\beta\|_2^2$ , представляющая собой сумму квадратов всех параметров модели.

Основным преимуществом LASSO-регуляризации является самоотбор регрессоров. В общем случае подбор оптимального набора регрессоров является NP-сложной задачей. Существуют методы отбора переменных, заключающиеся в поочередном включении/исключении регрессоров, но такой подход обладает рядом недостатков. Например, число регрессоров может быть значительно больше числа наблюдений, что делает невозможной процедуру поочередного исключения регрессоров. Также процедура «жадного» включения переменных может привести к локальному максимуму, который не обязательно будет близок к глобальному. LASSO-регуляризация позволяет решить эти проблемы.

Чем выше значение показателя регуляризации  $\lambda$ , тем меньше переменных включается в итоговое уравнение регрессии. Самоотбор происходит из-за негладкости ограничения, что можно видеть на рис. 1. Действительно, из-за ограничения по L1-норме возможны угловые решения, в которых некоторые переменные обращаются в ноль. В общем случае это неверно для L2-нормы и других гладких норм. Из-за негладкости нормы невозможно в общем случае вывести аналитическую оценку для параметра  $\beta$ . Тем не менее первая норма является



выпуклой, что позволяет применять методы численной оптимизации для нахождения оценок. В целом LASSO-регуляризацию можно применять для отбора переменных, после чего можно оценивать стандартную логистическую регрессию.

При LASSO-оценивании можно строить так называемые регуляризационные пути. По своей сути таким путем является график зависимости оценок параметров модели  $\beta$  от параметра регуляризации  $\lambda$ . График этого типа позволяет оценить влияние регуляризации на оценки параметров модели. При большей регуляризации в модели учитывается меньшее количество коэффициентов, их норма становится меньше. Напротив, по мере уменьшения параметра  $\lambda$  все больше коэффициентов попадают в исходную модель. Заметим, что при  $\lambda \rightarrow 0$  оценки LASSO-регрессии совпадают с оценками модели без регуляризации.

### Кросс-валидация как подход к оцениванию параметров модели

Чтобы можно было на качественном уровне оценить эффективность работы построенной модели, часто разбивают выборки на обучающую и тестовую. Оценка модели происходит на большей обучающей подвыборке с последующей проверкой на тестовой. Такая стратегия оценивания модели позволяет уменьшить эффект переобучения.

Процедура оценивания модели по обучающей подвыборке также имеет свои тонкости. В общем случае модель может зависеть от внешних параметров, не определяемых при оценивании. В нашем случае таким параметром является регуляризирующий множитель  $\lambda$ . Разумеется, желательно подобрать такое значение параметра  $\lambda$ , которое будет максимизировать качество модели. Тем не менее оптимизация метрики качества по параметру на обучающей выборке может привести к эффекту переобучения.

Одной из стратегий подбора оптимального значения  $\lambda$  является процедура кросс-валидации. Процедура заключается в разбиении обучающей выборки на  $q$  непересекающихся подвыборок с последующим поочередным обучением модели на  $(q - 1)$  выборках и в тестировании на оставшейся одной. После  $q$  повторений данной процедуры получают усредненную метрику качества при заданном значе-

нии параметра. Оптимальным значением параметра признается тот, для которого среднее значение оценки качества является наилучшим.

Оценка качества классификационной (бинарной) модели обычно происходит с помощью матрицы смежности (confusion matrix). В табл. 1 значения TP (True positive) и TN (True negative) – это число правильно классифицированных случаев отзывов и неотзывов соответственно. Значения FP (False positive) и FN (False negative) показывают количество неверно классифицированных случаев.

Таблица 1

Матрица смежности

Истинные \ Предсказанные	Предсказанные	
	1	0
1	TP	FN
0	FP	TN

Классификация объекта происходит в соответствии с правилом:

$$default_i = \begin{cases} 1, & \hat{p} \geq c \\ 0, & \hat{p} < c \end{cases} \quad (5)$$

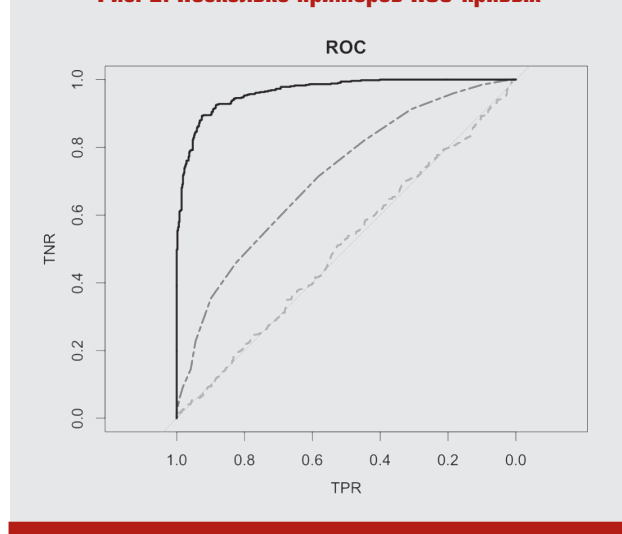
Показатель  $\hat{p}$  предсказанная моделью вероятность отзыва лицензии банка,  $c$  – уровень-отсечение принятия решения. Таким образом, для разных значений параметра будет разное качество классификационной модели. Например, при  $c = 0$  все банки будут признаны несостоятельными.

Чтобы оценить качество прогнозной силы модели, нужно построить некоторую метрику, которая не зависит от параметра. Такой метрикой может быть ROC-кривая. Она отражает выбор между долей правильно предсказанных позитивных исходов (True positive rate) и долей правильно предсказанных негативных исходов (True negative rate). Так, например, при  $c = 0$  не было упущено ни одного отзыва лицензии (TPR = 1), но признаны финансово несостоятельными все стабильные банки (TNR = 0). Таким образом, мы можем рассчитать точки (TPR, TNR) для каждого значения уровня-отсечения, которые и образуют ROC-кривую.

Показатель AUC (Area Under roC) равен площади под ROC-кривой. Классификационная способность модели тем лучше, чем ближе AUC к 1. Классификатор признается бесполезным при AUC, равном 0,5. На рис. 2 показан пример нескольких ROC-кривых на сгенерированных данных. Темная кривая соответствует почти идеальному классификатору, светлая пунктирная – наихудшему.

Таким образом, в работе применяется процедура оценивания модели с LASSO-регуляризацией и кросс-валидацией для выбора показателей, которые будут включены в итоговую модель. После отбора показателей производится оценивание классической логистической регрессии с последующим анализом качества модели.

Рис. 2. Несколько примеров ROC-кривых



### Балансировка выборки

Как было сказано выше, используемая выборка является сильно несбалансированной – количество наблюдений с отзывом лицензии занимает незначительную долю во всей выборке. Главное следствие использования несбалансированной выборки заключается в ухудшении качества оценки параметров. Чтобы получить лучшее качество оценивания, необходимо сбалансировать выборку. Этого можно добиться несколькими способами.

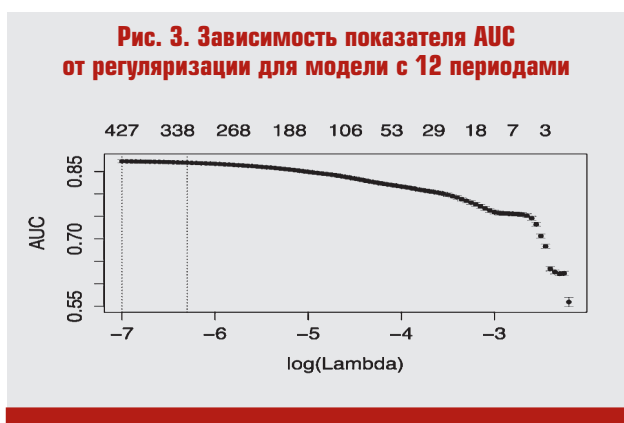
Во-первых, можно сократить подвыборку наблюдений без отзыва лицензий. Во-вторых, можно искусственно увеличить количество наблюдений с отзывами с помощью случайного дублирования данных. Очевидно, что эти способы не являются совершенными: в первом случае теряется информация, а во втором повышается риск переобучения – ситуации, когда модель хорошо описывает имеющиеся данные, но обладает низкой прогнозной силой. Можно добиться некоторого равновесия путем совмещения двух способов балансирования выборки.

Для оценивания модели необходимо определить значения зависимой переменной  $y_i$ , которая отвечает за состояние банка. В рамках статьи используется 4 подхода: зависимая переменная принимает значение 1 за 12, 6, 3 и 1 период до объявления об отзыве лицензии. Выбор количества периодов влияет на интерпретацию результатов. Так, в случае выбора 12 периодов отклик модели интерпретируется как вероятность отзыва лицензии у банка в течение ближайшего года.

При оценке всех моделей использовался подход балансировки данных, при котором количество наблюдений с единичной зависимой переменной удваивалось случайным образом, а количество наблюдений с нулевой зависимой переменной случайным образом сокращалось так, чтобы выборка была сбалансированной. Доля обучающей выборки составляет 70% от общей. Число подвыборок при кросс-валидации равно 10.

### Оценивание модели

Отметим, что, хотя LASSO-регуляризация помогает произвести отбор регрессоров, выбор модели остается сложной задачей. Добавление переменных в модель почти всегда увеличивает значение показателя AUC на обучающей выборке. На рис. 3 представлена зависимость метрики AUC от значения показателя регуляризации (вверху отмечено количество регрессоров, включенных в модель). Действительно, количество регрессоров положительно влияет на качество классификации. Однако модель, включающая в себя такое большое количество показателей, сложно интерпретируема и непрактична. Более того, может возникнуть ситуация переобучения, при которой модель, например, с 300 регрессорами плохо работает на тестовой выборке.



Кривая зависимости метрики AUC от параметра регуляризации содержит как резкие перепады, так и относительно стабильные участки. Мы будем пользоваться следующим критерием выбора оптимальной модели. В качестве итоговой будет признана наилучшая модель, содержащая в себе не более 10 объясняющих переменных. Во-первых, такое количество переменных можно назвать относительно небольшим, им можно легко дать экономическую интерпретацию. Во-вторых, при таком количестве переменных риск переобучения невелик.

В табл. 2 представлены результаты оценивания логистических регрессий после отбора параметров с помощью LASSO-регуляризации. Все коэффициенты значимы на любом разумном уровне значимости. Отметим, что из-за небольшого количества регрессоров ситуации переобучения не возникло. Модели одинаково хорошо классифицируют как обучающую, так и тестовые выборки. Чем меньше периодов до отзыва лицензии принято за единицу, тем лучше характеристика качества модели. Этот результат представляется вполне предсказуемым, так как чем ближе банк к несостоятельности, тем более выражены факторы нестабильности.

Отметим также, что оборотные переменные становятся тем значимее, чем ближе отзыв лицензии – оборотные переменные вообще не присутствуют в модели для 12 периодов до отзыва лицензии. На

рис. 4 показаны ROC-кривые по тестовой выборке для каждой из четырех моделей.

Дадим экономическую интерпретацию полученным коэффициентам. Переменная  $Cp$  – объем собственного капитала фирмы, который можно встретить в формах 123 и 134. Этот показатель имеет отрицательное влияние на вероятность отзыва лицензии, что можно объяснить одним: чем больше у банка собственный капитал, тем больше средств он имеет для предотвращения внешних финансовых шоков.



Также этот показатель отражает размер банка, как и переменная  $Act$ , являющаяся прокси-переменной для активов банка. Размер активов банка отрицательно влияет на вероятность отзыва лицензии, что может быть связано с эффектом *too big to fail* или с тем фактом, что крупные кредитные организации обычно менее подвержены финансовым колебаниям.

Чем выше доля денежных средств и чеков в активах банка (переменная  $TA\_W$ ), тем выше вероятность отзыва лицензии, что можно объяснить, например, стремлением обеспечить себя финансовой «подушкой» в условиях приближающейся финансовой нестабильности.

Повышение доли обязательных резервов в активах банка ( $TA\_Or$ ) отражает рост резервов под полученные депозиты. Рост обязательных резервов можно связать с возросшим спросом на депозиты, что, в свою очередь, может быть объяснено ухудшением финансового состояния банка.

Доля расчетных счетов в иностранных организациях в активах ( $TA\_Kf$ ) и доля кредитов по МБК в активах (срок до 1 дня –  $TA\_Lb\_1$ , срок до 7 дней  $TA\_Lb\_7$ ) отрицательно влияют на вероятность отзыва лицензии, так как кредитные организации (как иностранные, так и отечественные) стремятся работать с надежными контрагентами.

Доля долгосрочных кредитов финансовым организациям ( $TA\_La\_m3y$ ), домохозяйствам ( $TA\_Lh\_m3y$ ) и кредитным организациям ( $TA\_Li\_m3y$ ) влияют на финансовую стабильность банков положительно. Этот факт можно объяснить тем, что дол-

Таблица 2

## Результаты оценивания моделей для случаев 12, 6, 3 и 1 периодов

Периоды до отзыва		12	6	3	1
<i>Балансовые переменные</i>					
<i>Cr</i>	Объем собственного капитала фирмы	–0,17			
<i>TA_W</i>	Доля денежных средств и чеков	0,94			
<i>TA_Or</i>	Доля обязательных резервов	9,10	12,92	13,85	13,23
<i>TA_Kf</i>	Доля расчетных счетов в иностранных организациях	–3,30	–4,39	–4,32	–3,79
<i>TA_Lb_1</i>	Доля кредитов по МБК (сроком до 1 дня)	–1,60			
<i>TA_La_1y</i>	Доля кредитов коммерческим организациям (сроком до 1 года)	0,30			
<i>TA_La_m3y</i>	Доля кредитов коммерческим организациям (сроком свыше 3 лет)	–0,90	–1,00	–0,94	
<i>TA_Lh_m3y</i>	Доля кредитов домохозяйствам (сроком свыше 3 лет)	–1,08	–1,01	–1,00	
<i>TA_Li_m3y</i>	Доля кредитов финансовым организациям (сроком свыше 3 лет)	–7,46			
<i>TA_Ca</i>	Доля счета собственного капитала (активная часть)		0,70	0,86	1,66
<i>TA_Ca</i>	Доля кредитов по МБК (сроком до 7 дней)		0,70	0,86	1,66
<i>TA_La_180</i>	Доля кредитов коммерческим организациям (сроком до 180 дней)				–2,16
<i>TA_S</i>	Доля приобретенных ценных бумаг				–0,75
<i>Act</i>	Активы банка				–0,13
<i>Оборотные переменные</i>					
<i>t_Ni</i>	Оборачиваемость по расчетным счетам кредитных организаций		–0,04		
<i>t_Lf_1</i>	Оборачиваемость по кредитам иностранным организациям (сроком до 1 дня)		–0,06	–0,10	
<i>t_Lf_7</i>	Оборачиваемость по кредитам иностранным организациям (сроком до 7 дней)		–0,05		–0,09
<i>t_Lc_1</i>	Оборачиваемость по кредитам центральному банку (сроком до 1 дня)			–0,04	–0,05
<i>t_Lb_1</i>	Оборачиваемость по кредитам банкам по МБК организациям (сроком до 1 дня)			–0,02	–0,05
<i>Макропеременные</i>					
<i>GDP</i>	ВВП	–0,63	–0,27		
<i>Exc</i>	Курс доллара в рублях		0,01	0,02	
<i>Exp</i>	Объем экспорта				–0,06
<i>Дополнительная информация</i>					
<i>n<sub>train</sub></i>		10 804	5 592	2 870	960
<i>AUC<sub>train</sub></i>		0,769	0,796	0,816	0,848
<i>AUC<sub>test</sub></i>		0,772	0,797	0,797	0,850

госрочное кредитование предоставляется только финансово надежным контрагентам, а высокая доля долгосрочных кредитов может говорить о качестве кредитного портфеля банка.

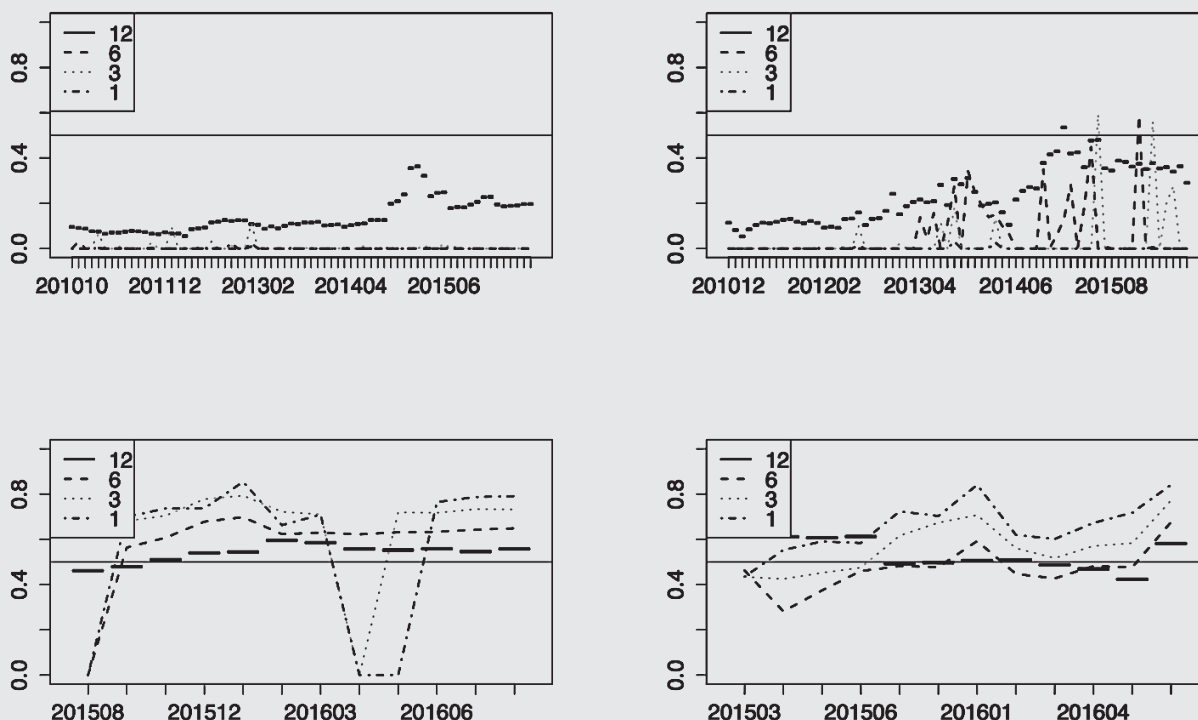
Разная связь кредитов коммерческим организациям сроком до одного года (*TA\_La\_1y*) и до 180 дней (*TA\_La\_180*) с финансовым положением банка объясняется разницей во временных интервалах. Так, переменная *TA\_La\_1y* присутствует в модели с

12 периодами до отзыва лицензии, т. е. обязательства по данным кредитам наступают в течение предсказываемого периода. Таким образом, в рамках этой модели высокая доля кредитов является кредитным риском невозврата. А переменная *TA\_La\_180* присутствует в модели с одним периодом до отзыва и, наоборот, является отложенной прибылью.

Переменная *TA\_Ca* включает в себя отложенные расходы будущих периодов и/или нераспределен-



Рис. 5. Оценка вероятности отзыва лицензии для некоторых банков



ные полученные убытки. Высокая доля отложенных расходов имеет определенно негативное влияние на финансовое состояние банка. Доля приобретенных ценных бумаг ( $TA\_S$ ) может свидетельствовать о диверсифицированности банка, что положительно связано с его финансовой устойчивостью.

Что касается оборотных коэффициентов, то все они имеют негативное влияние на вероятность отзыва лицензии у банка. Это можно объяснить тем, что высокая активность банка отражает его «живучесть» и устойчивое функционирование с контрагентами.

Интерпретация макропеременных достаточно проста. На хорошем макроэкономическом фоне, когда выпуск ( $GDP$ ) растет, курс еще не успел резко вырасти ( $Exs$ ), а поступления от экспорта высоки ( $Exr$ ), банковский сектор также чувствует себя хорошо.

Оцененные модели позволяют достаточно хорошо классифицировать банки по отзыву лицензии. Можно утверждать, что при более системном подходе к отбору модели можно достичь лучших результатов. Так, было эмпирически проверено, что модель с одним периодом до отзыва лицензии значительно улучшается при увеличении количества регрессоров до 20, при этом не переобучаясь. В то же время представляет интерес предсказывающая способность построенных моделей.

На рис. 5 отображена оценка вероятности отзыва лицензии для каждого периода с использованием всех четырех моделей. Названия банков опускаются из-за конфиденциальности (так, верхние рисунки

отражают предсказания модели для двух банков, входящих в топ-100 по размерам активов). В целом представленные кредитные организации демонстрируют финансовую устойчивость, но показатели одного банка (правый рисунок) может сигнализировать о некоторых финансовых проблемах.

Наибольший интерес представляют нижние два рисунка, отражающие предсказания модели для двух кредитных организаций, у которых отзывали лицензию (все четыре линии практически все время превышают значение 0,5 – отзыв лицензии у этих банков можно было предвидеть).

## ВЫВОДЫ

В статье предложен набор моделей вероятности отзыва банковской лицензии регулятором. Процедура построения моделей базируется на LASSO-регуляризации с применением методов оценивания, основанных на кросс-валидации. Такой метод позволил выделить наиболее значимые факторы более чем из 600 переменных, описывающих финансовое положение конкретного банка и общее состояние экономики. Важно заметить, что для оценивания модели использовались только открытые данные, предоставляемые Банком России на регулярной основе.

Использование различных индикативных переменных, учитывающих будущий отзыв лицензии, в качестве объясняемых переменных, в свою очередь, позволило классифицировать эти факторы в зависимости от возможности их использования в качестве опережающих индикаторов. Отдельно

отметим, что модель имеет прямое практическое значение – она может использоваться участниками финансового рынка для оперативного мониторинга

финансового состояния контрагентов и банковской системы в целом.

### Список литературы

1. *Beaver W. H.* Financial ratios as predictors of failure. Journal of accounting research. 1966.
2. *Altman E. I.* Financial ratios, discriminant analysis and the prediction of corporate bankruptcy // The Journal of finance, 1968.
3. *Ohlson J. A.* Financial ratios and the probabilistic prediction of bankruptcy // Journal of accounting research. 1980.
4. *Lo A. W.* Logit versus discriminant analysis: A specification test and application to corporate bankruptcies // Journal of econometrics. 1986.
5. *Kolari J., Glennon D., Shin H., Caputo M.* Predicting large US commercial bank failures // Journal of Economics and Business. 2002.
6. *Бризицкая А. В.* Особенности внедрения стандартов третьего поколения по достаточности капитала в банковскую практику зарубежных стран // Финансовый журнал. 2015. № 4. С. 112–121.
7. *Пересецкий А. А.* Модели причин отзыва лицензий российских банков. Влияние неучтенных факторов // Прикладная эконометрика. 2014. № 30. С. 49–64; 2014. № 4. С. 24.
8. *Карминский А. М., Костров А. В.* Моделирование вероятности дефолта российских банков: расширенные возможности // Журнал Новой экономической ассоциации. 2013.
9. *Louzada F., Ara A., Fernandes G. B.* Classification methods applied to credit scoring: A systematic review and overall comparison. 2016.
10. *Haerdle W. K., Prastyo D. D.* Default Risk Calculation based on Predictor Selection for the Southeast Asian Industry. 2013.
11. *Леонидов А. В., Румянцев Е. А.* Оценка системных рисков межбанковского рынка России на основе сетевой топологии // Журнал Новой экономической ассоциации. 2013. № 3, С. 65–80.
12. *Cont R., Moussa A., others.* Network structure and systemic risk in banking systems. Edson Bastos-e, Network Structure and Systemic Risk in Banking Systems. Dec. 2010.
13. *Беляков И. В.* О бюджетной стоимости кризисов // Финансовый журнал. 2015. № 5.
14. *Пильник Н. П., Поспелов И. Г., Дедова М. С.* Описание потребности в ликвидности со стороны российской банковской системы на основе статистики оборотов // Журнал новой экономической ассоциации. 2014. № 4. С. 87–109.

ИНФОРМБАНК

**Заголовок**

**2000 знаков  
(без заголовка)**