

Как прогнозировать дефолты банков: эволюция методов, моделей и факторов риска

© 2026 г. М.А. Щепелева, М.И. Столбов

М.А. Щепелева,

к.э.н, доцент, НИУ «Высшая школа экономики», Москва; e-mail: mshchepelova@hse.ru, mschepelova@yandex.ru

М.И. Столбов,

д.э.н, профессор, заведующий кафедрой прикладной экономики, МГИМО МИД России, Москва; e-mail: stolbov.m@my.mgimo.ru

Поступила в редакцию 29.04.2025

Исследование выполнено при поддержке Российского научного фонда (проект № 23-18-00756).

Аннотация. Прогнозирование банковских дефолтов представляет собой важную задачу для всей экономики. Раннее выявление проблемных банков помогает предотвратить надвигающиеся банкротства либо минимизировать связанные с ними потери. В статье обсуждается современное состояние инструментальных методов и данных, используемых для этой цели. Последовательно рассматриваются теоретические предпосылки, эволюция методологических подходов, применяемых для прогнозирования банковских дефолтов, особенности работы с данными, а также перечни предикторов, которые включают в модели раннего оповещения. Делается вывод, что в литературе до сих пор наблюдаются существенные противоречия в отношении как методов, так и переменных, которые следует использовать в прогнозных моделях. Методы машинного обучения демонстрируют лучшую по сравнению с традиционными статистическими моделями способность выявлять нелинейные зависимости и работать с большими выборками. Их преимущества часто нивелируются при вневыборочной оценке. Другие ограничения подобных методов связаны с риском переобучения и сложностью интерпретации результатов. Списки потенциальных предикторов банковских дефолтов также разнятся от страны к стране. Чаще всего в прогнозных моделях используют данные банковских балансов и финансовые коэффициенты. Однако есть исследования, которые показывают для отдельных стран повышение точности прогнозов при включении рыночных, макроэкономических и нефинансовых показателей. Перспективы дальнейших исследований в этой области заключаются в поиске оптимального сочетания параметрических и непараметрических подходов, исследовании потенциала нефинансовых показателей в качестве факторов банковских банкротств, а также в проведении исследований на больших выборках, включающих как развитые, так и развивающиеся страны.

Ключевые слова: дефолт банка, машинное обучение, системы раннего оповещения, факторы банковских дефолтов.

Классификация JEL: G17, G21.

УДК: 336.71.

Для цитирования: **Щепелева М.А., Столбов М.И.** (2026). Как прогнозировать дефолты банков: эволюция методов, моделей и факторов риска // *Экономика и математические методы*. Т. 62. № 1. С. 63–77. DOI: 10.7868/S3034617726010057

1. ВВЕДЕНИЕ

Прогнозирование дефолтов банков представляет собой важную задачу как с точки зрения отдельных инвесторов, так и с точки зрения регулятора. Дефолт одного банка, даже небольшого размера, связан с денежными убытками для отдельных агентов, но он также может вызвать цепочку неплатежей и кризис доверия, что может угрожать стабильности банковской системы в целом. Повышенная вероятность возникновения заражения отличает банки от нефинансового сектора. Проблемы в банковской системе могут также распространяться на реальный сектор и в конечном итоге негативно повлиять на экономический рост. Таким образом, серьезные последствия, которые могут повлечь дефолты банков, требуют тщательного изучения.

На сегодняшний день накоплен значительный опыт прогнозирования дефолтов банков. Чаще всего для решения этой задачи используются такие статистические методы, как дискриминантный анализ, а также логит- или пробит-модели для панельных данных. Сравнительно новый подход предполагает применение методов машинного обучения, решающих задачу классификации: случайного леса (random forest), метода опорных векторов (support vector machines), бинарных деревьев (binary decision trees), а также более сложных техник — ансамблей алгоритмов и нейронных сетей.

Несмотря на то что методы машинного обучения получили широкое распространение и имеют высокую предсказательную силу, в современных условиях дефолты банков по-прежнему случаются и вызывают значительный резонанс. Последний пример — крах Silicon Valley Bank (SVB) в США — 16-го по размеру активов банка США на конец 2022 г. После SVB обанкротился еще один банк — Signature Bank of New York (SBNY), который по системе CAMELS имел удовлетворительную оценку, присвоенную банку за два дня до банкротства (Citterio, 2024). Эти два события показывают, что, очевидно, до сих пор остаются неучтенные исследователями факторы, которые могут влиять на вероятность банкротства.

Цель данной статьи — систематизировать накопленный массив исследований прогнозирования дефолтов банков, уделив особое внимание применению методов машинного обучения для решения этой задачи.

Работа структурирована следующим образом: в разд. 2 рассматриваются причины дефолтов банков; далее, в разд. 3, технические нюансы работы с данными, которые используются при прогнозировании дефолтов банков. В разд. 4 обсуждается эволюция методологических подходов для прогнозирования банковских дефолтов. В разд. 5 приведены выводы на основании итогов исследования.

2. ОСНОВНЫЕ ПРИЧИНЫ БАНКОВСКИХ ДЕФОЛТОВ

Д. Мартин (Martin, 1977) подразумевал под банкротством банка ситуацию, когда его чистое богатство становится отрицательным, в результате чего банк не может продолжать свою деятельность без серьезных убытков. В большинстве исследований, которые будут рассмотрены ниже, используется близкое определение: под дефолтом понимается невозможность исполнения банком своих обязательств, что требует вмешательства регулятора¹.

Есть три гипотезы, которые объясняют, с чем могут быть связаны дефолты банков (Fungacova, Turk, Weill, 2021).

1. Гипотеза слабых фундаментальных показателей. Согласно этой гипотезе, ухудшение основных показателей деятельности компании тесно связано с вероятностью банкротства. Одним из примеров системы опережающих индикаторов банкротства банка, которая была разработана на основе этой гипотезы, является система CAMELS. Она оценивает показатели деятельности банка в соответствии с пятью компонентами: достаточность капитала (*Capital*), качество активов (*Assets*), качество менеджмента (*Management*), прибыльность (*Earnings*), состояние ликвидности (*L*), платежеспособность (*Solvency*). CAMELS считается одной из самых аккуратных систем для оценки финансового положения банка. Основная проблема — результаты CAMELS доступны с отставанием.

Гипотеза слабых фундаментальных показателей логично подводит к тому, что для предотвращения дефолта необходимо развивать микропруденциальное регулирование.

2. Гипотеза нехватки ликвидности. Ликвидность является одним из элементов системы CAMELS, но во многих исследованиях именно проблемы в этой компоненте выделяют как главную причину дефолтов банков. Основная причина банкротства, согласно этой теории, связана с самой природой банковской деятельности, а именно — с финансированием неликвидных длинных активов короткими ликвидными обязательствами (Diamond, Dybvig, 1983; Kashyap, Rajan, Stein, 2002). Если происходит внешний шок, а также имеется «ограничение на последовательное обслуживание», вкладчики, которые оказываются первыми в очереди, предпочитают изъять свои

¹ В данной работе рассматриваются банкротства, которые происходят в результате нарушения банками выплат по обязательным платежам либо в результате нарушения обязательных нормативов ликвидности и достаточности капитала. Случаи отзыва лицензии в связи с мошенническими действиями не рассматриваются. Это — отдельное направление исследований, в которых используются другие данные: о сделках с наличными расчетами, электронных переводах и платежах, банковских счетах, инвестиционных сделках через посредников и др. (Jullum et al., 2020; Lokanan, 2024).

средства. А по мере того как способность банка рассчитываться со всеми вкладчиками снижается, любые шоки ликвидности становятся все более и более чувствительными. Таким образом, изначально банк является платежеспособным, но из-за ухудшающихся проблем в ликвидности он приближается к банкротству.

Гипотеза дефицита ликвидности свидетельствует в пользу создания более доверительных отношений между банками и другими игроками, а также формирования специальных институтов, снижающих врожденную уязвимость банков, — систем страхования вкладов, возможность получения финансовой помощи от кредитора последней инстанции, спасение банка государством.

В (Fungacova, Turk, Weill, 2021) авторы предлагают третью гипотезу, которая утверждает, что повышенный риск банкротства может быть связан с созданием чрезмерной ликвидности.

Представленные выше гипотезы не свидетельствуют о том, что является причиной дефолта, они скорее описывают симптомы — слабые финансовые коэффициенты либо неблагоприятную ситуацию с ликвидностью. Опираясь на них, можно подобрать информативные показатели, которые будут использоваться в моделях раннего оповещения для предотвращения банкротства.

Какие же факторы могут вести к слабым фундаментальным показателям либо к неблагоприятной ликвидности?

К внешним факторам относится состояние экономики в целом в данной стране. При прочих равных риски оказаться банкротом в стране с высокими темпами роста ВВП и стабильной инфляцией ниже, чем в стране с низкими темпами роста ВВП и высокой нестабильной инфляцией (Flannery, 1998; Jagtiani, Lemieux, 2001; Karminsky, Khromova, 2016). В исследованиях в отношении России аналогичный вывод показан в (Карминский, 2015; Пересецкий, 2007). При этом структура банковской системы также может влиять на показатели деятельности банков. Уровень конкуренции влияет на прибыльность и устойчивость банков. Прямая или обратная зависимость между конкуренцией и вероятностью дефолтов банков зависит чаще всего от других показателей, в частности от проводимой политики, качества институтов, структуры собственности банка, макроэкономических условий (Diallo, 2015; Goetz, 2018).

Также проводимая политика в некоторых случаях может сделать отдельные банки уязвимыми: снижение объема международных резервов, политика высоких процентных ставок, увеличение налоговой нагрузки (Giordana, Schumacher, 2017; Erer E., Erer D., 2024).

Тем не менее, когда происходит внешний шок, через дефолт проходят далеко не все банки. Станет ли банк банкротом или нет в случае реализации внешних шоков в большей степени зависит от его внутренней устойчивости. К внутренним факторам, ведущим к слабым фундаментальным показателям и неблагоприятной ликвидности, можно отнести неэффективный менеджмент, неверно проводимую бизнес-стратегию и слабые процедуры внутреннего контроля.

Авторы (DeYoung, Torna, 2013) связывают ухудшение финансовых показателей с неверной стратегией банка, а именно с расширением таких нетрадиционных банковских операций, как инвестиционный банкинг, секьюритизация, венчурные инвестиции. Авторы (Cole, White, 2012), рассматривая традиционный подход к оценке вероятности дефолта банков на основе CAMELS, показывают, что *наиболее значимым предиктором* является объем инвестиций банка в коммерческую недвижимость

Неэффективный менеджмент также серьезно влияет на результаты деятельности банка, и это всегда отражается в показателях баланса и отчета о прибылях и убытках. Менеджмент отвечает за стратегию, которой следует банк, за его устойчивость и прибыльность. Агрессивная стратегия, нацеленная на быстрый рост кредитного портфеля, в сочетании с недостаточностью залога, наличие большой доли волатильных обязательств и неликвидных активов быстро ухудшает баланс банка, особенно если ухудшается внешняя среда (Serrano-Cinca et al., 2014; Köhler, 2015). Также менеджмент отвечает за устойчивость банка, обеспечивает следование кредитной политике, соответствие основным правилам деятельности банков и надзор над деятельностью ключевых департаментов. Недостаточно развитые процедуры внутреннего контроля также могут являться причиной банкротства, особенно в развивающихся странах и странах с низким уровнем дохода (Podriega, Weill, 2008).

3. ОСОБЕННОСТИ РАБОТЫ С ДАННЫМИ ДЛЯ ПРОГНОЗИРОВАНИЯ БАНКОВСКИХ ДЕФОЛТОВ

В этом разделе рассматриваются основные технические моменты, с которыми сталкиваются исследователи при работе с данными для прогнозирования банковских дефолтов.

Прогнозирование банкротства банка представляет собой задачу классификации: на основе анализируемого набора данных, где банки имеют разметку на «стабильные» и «обанкротившиеся», необходимо предсказать, в какой класс попадут банки с заданными характеристиками.

Первая особенность работы с подобными данными связана с разметкой зависимой переменной, т.е. с определением факта, обанкротился банк в отчетном периоде или нет. Такая разметка напрямую связана с тем, что *конкретные исследователи подразумевают* под банкротством, или дефолтом, банка.

Авторы (Valcaen, Ooghe, 2006) выделили три основных подхода к определению банкротства банка. Первый подход предполагает формальное определение банкротства, т.е. на основании соответствующего решения регулятора (юридическое определение банкротства). Вторым подходом базируется на концепции финансового стресса, т.е. метка «банкрот» присваивается тем банкам, которые находятся в уязвимом финансовом положении, а это устанавливается на основе субъективных критериев исследователей. Наконец, третий подход устанавливает связь банкротства банка с определенными действиями и процедурами, — такими как понижение рейтинга банка, объявление дефолта по кредитному портфелю, проведение процедуры рекапитализации банка, наличие финансовой поддержки со стороны государства, пересмотр ковенант² и др.

Несмотря на то что первое определение является объективным, его использование связано со сложностями. Во-первых, всегда есть некий временной лаг между моментом, когда банк уже фактически прекратил свою деятельность, и моментом, когда выносится соответствующее решение регулятора. Во-вторых, в некоторых случаях банки, как и другие компании, могут объявлять себя банкротами по стратегическим причинам, например для избавления от долгов или в результате непредвиденных событий, природных катастроф, смены правительства и т.д. Кроме того, банкротство банка не обязательно заканчивается ликвидацией банка, возможны сделки слияния/поглощения. На самом деле официальное признание банка банкротом — нечастое явление. Если ориентироваться только на юридическое определение, можно сильно занижить долю слабых банков, возможно, стоящих на грани банкротства, в исследуемом наборе данных (Citterio, 2024).

Недостаток второго подхода связан с существенной долей субъективности исследователей, в результате чего мы имеем большое число работ, результаты которых несопоставимы между собой. Авторы (Switzer, Wang, 2013a) для выделения финансово уязвимых банков используют данные о кредитно-дефолтных свопах, а в (Cipollini, Fiordelisi, 2012) ориентируются на показатель стоимости компании для акционеров. Авторы (Poghosyan, Ćihak, 2011) выделяют слабые банки на основе текстового анализа, т.е. путем поиска банков, в отношении которых наблюдаются наиболее высокие баллы по негативному сентименту в новостях и прессе. В (Halteh, Kumar, Gepp, 2018) используется показатель z-score Альтмана, а в (Zaki, Bah, Rao, 2011) — процентное изменение в доходности акций банка, среднегодовую доходность на капитал, а также среднегодовую чистую процентную маржу.

Третий подход предполагает широкий взгляд на банкротство банка и отнесение к этому явлению разных альтернативных действий и процедур, что также не способствует сопоставимости проведенных исследований.

Вторая особенность связана с тем, как исследователи работают с проблемой несбалансированных данных. Она часто встречается и в других экономических задачах, где используются модели бинарного выбора, в частности при прогнозировании дефолтов по кредитам, прогнозировании банкротств компаний реального сектора, финансовых кризисов и пр. Если специально не работать с дисбалансом классов, прогнозы любых моделей будут смещены в сторону мажоритарного класса.

² Ковенант (covenant) — договорное обязательство в праве и финансах. В финансовой сфере — условия в кредитном договоре, обязывающие заемщика соблюдать финансовые или нефинансовые обязательства (например, поддерживать определенный уровень прибыли) в обмен на более выгодные условия займа; нарушение ведет к штрафам или досрочному возврату долга.

Применение моделей машинного обучения для прогнозирования дефолтов предлагает новые и относительно эффективные способы работы с дисбалансом классов: первый — это ресэмплинг³, который включает андерсэмплинг, или удаление лишних мажоритарных наблюдений, и оверсэмплинг, или дублирование наблюдений миноритарного класса; второй класс — добавление специальных штрафов в функцию потерь.

Согласно (Naixiang et al., 2017), ресэмплинг является более популярным способом работы с дисбалансом классов, чем изменение функции потерь; при этом оверсэмплинг используется в два раза чаще, чем андерсэмплинг.

Третья особенность, которая может приводить к несопоставимости результатов, — установление определенного порога для ошибок первого и второго рода.

Порог служит для трансформации полученной в результате оценки модели вероятности в бинарный прогноз. По умолчанию для моделей дискретного выбора, если вероятность превышает 0,5, то предсказание равно 1, что означает наступление дефолта. В силу того что при прогнозировании дефолтов ошибки первого рода имеют большую стоимость для общества — классификация слабого банка в качестве финансово устойчивого имеет гораздо более сильные негативные последствия, чем признание хорошего банка банкротом; часто в исследованиях порог берется ниже 0,5.

Наличие разных пороговых значений, которые выбираются исследователями субъективно, не позволяет напрямую сравнивать получаемые ими результаты. Это объясняет, почему предсказательная сила различных методов разнится от исследования к исследованию.

Четвертая проблема, связанная с данными, — игнорирование панельной структуры данных при прогнозировании дефолтов с помощью методов машинного обучения. В большинстве работ тот или иной метод машинного обучения применяется к панели данных, но рассматривает каждое наблюдение «время—объект» как уникальное значение, игнорируя, что анализируется один и тот же объект в течение некоторого промежутка времени.

4. ОБЗОР МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ ОЦЕНКИ ВЕРОЯТНОСТИ ДЕФОЛТА БАНКОВ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

В этом разделе мы рассмотрим эволюцию методологических подходов для исследования банковских дефолтов. Все представленные ниже работы решали две задачи — предиктивную и дескриптивную — вспомогательную задачу, которая заключалась в составлении портрета обанкротившегося банка.

В целом методы, которые использовали исследователи, можно разделить на две большие группы: параметрические (все разновидности дискриминантного анализа, логит- и пробит-регрессии, модель пропорциональных рисков, наивный Байес) и непараметрические (нейронные сети, метод опорных векторов, деревья решений, k -ближайших соседей, различные нечеткие алгоритмы).

Как параметрические, так и непараметрические методы направлены на поиск закономерностей в данных. Отличия между ними заключаются в следующем. Параметрические методы предполагают, что данные порождены определенным распределением, в то время как непараметрические рассматривают данные как результат некоторого неизвестного процесса.

Исследование возможностей прогнозирования банковских дефолтов началось в 1966 г., когда была опубликована работа (Beaver, 1966). В данной работе автор анализировал *динамику отдельных финансовых показателей*, в частности: леввериджа, доходности на активы, коэффициентов ликвидности. Автор пришел к выводу, что ухудшение этих показателей происходит тогда, когда банк близок к банкротству.

Затем исследователи стали рассматривать возможность *адаптировать модели предсказания банкротства, применяемые к компаниям нефинансового сектора, для банковской отрасли*. В (Altman, 1968) впервые был представлен подход оценки риска банкротства для нефинансовых публичных компаний. На основе пяти коэффициентов: отношения рабочего капитала — к совокупным активам, нераспределенной прибыли — к совокупным активам, EBIT — к совокупным активам, рыночной

³ Ресэмплинг (resampling) — процесс создания нового набора данных (например, изображения, звука, статистики) на основе существующего, путем изменения его размера или характеристик, например увеличения/уменьшения числа пикселей, изменения частоты дискретизации или генерации новых выборок из имеющихся данных.

стоимости — к балансовой (для совокупного долга), объема продаж — к совокупным активам — он вывел показатель z -score⁴. Впоследствии z -score часто использовался как мера расстояния до дефолта. Статья (Sinkey, 1975) стала первой, где эта модель была применена для прогнозирования дефолтов банков в США в 1969–1972 гг. Точность ее прогнозов составила 72%. С этого момента дискриминантный анализ получил широкое распространение.

В 1970 г. авторы (Meuer, Pifer, 1970) использовали МНК-регрессию для оценки риска банкротства банка. В качестве предикторов в их модели выступали различные показатели, рассчитанные на основе бухгалтерской отчетности.

В 1977 г. для прогнозирования банковских дефолтов Д. Мартин (Martin, 1977) использовал *логит-регрессию*. Он сравнил результаты логит-регрессии с многомерным дискриминантным анализом на выборке из американских банков и показал, что обе модели прогнозируют на достаточно хорошем уровне.

Впоследствии в некоторых работах также использовалась *пробит-регрессия* (Bovenzi, Marino, McFadden, 1983; Chiaramonte, Croci, Poli, 2015). Исследования показали, что по точности прогнозов пробит- и логит-модели очень близки.

Традиционная логит-модель, обладая преимуществом выделения наиболее значимых предикторов банкротства банка, тем не менее, не может предсказать, *когда точно* произойдет банкротство, или сколько времени еще просуществует банк. Для исследования периода времени, в течение которого произойдет дефолт, авторы (Lane, Looney, Wansley, 1986) использовали *модель пропорциональных рисков*, применив ее к выборке из американских банков за период с 1979 по 1984 г. (Whalen, 1991), а также (Braga et al., 2006) использовали ту же модель. Модель дает относительно точные прогнозы, но ее недостаток заключается в использовании предпосылки о том, что все банки в итоге окажутся в состоянии дефолта, что не соответствует реальному положению дел.

С 1990-х годов активно используются *нейронные сети*. Одним из первых исследований с использованием нейронных сетей была работа (Tam, 1991), в которой были изучены обанкротившиеся банки в период 1985–1987 г. Автор пришел к выводу, что нейронные сети обладают более высоким прогностическим потенциалом, чем дискриминантный анализ. Авторы (Salchenberger, Cinar, Lash, 1992) также сравнивали эффективность нейронной сети с логит-моделью. Однако в данной работе существенного превосходства нейронной сети по части аккуратности прогнозов обнаружено не было. Тем не менее, исследователи пришли к выводу, что нейронная сеть лучше подходит для моделирования нелинейных взаимосвязей в банковском секторе.

С середины 2000-х годов широкое распространение получили *методы машинного обучения*, в частности метод опорных векторов, классификационные деревья, случайный лес. Такая популярность была связана с ростом вычислительных мощностей и лучшей способностью обрабатывать большие данные.

В работах (Mompalmer, Carmona, Climent, 2016; Malikkidou, Bräuning, Scalone, 2019) в качестве основного метода были выбраны *классификационные деревья*. В первой работе они использовались для выявления наиболее важных предикторов банковских дефолтов для 155 банков стран Еврозоны. Таковыми стали величина активов и отношение кредитного портфеля к привлеченным депозитам. В работе (Malikkidou, Bräuning, Scalone, 2019) авторы сравнивали предсказательный потенциал деревьев решений с логистической регрессией для выборки, включавшей порядка 3000 средних и малых европейских банков. Авторы показывают, что обе модели дают высокую точность прогнозов, как внутри-, так и вневыборочно, но логит-модель гораздо более чувствительна к количеству пропущенных значений в наборе данных.

В работе (Gogas, Papadimitriou, Agrapetidou, 2018) анализировалась выборка более чем из 1400 банков с помощью *метода опорных векторов*. Наиболее важными предикторами банковских дефолтов были признаны норматив достаточности капитала и отношение совокупных процентных расходов к совокупным процентным доходам.

⁴ Z-score (z -оценка) — статистический показатель, который показывает, на сколько стандартных отклонений данное значение отличается от среднего значения в группе. Он стандартизирует данные, позволяя сравнивать значения из разных распределений, и рассчитывается по формуле $z = (x - \mu) / \sigma$, где x — ваше значение, μ — среднее, σ — стандартное отклонение. Положительный Z-score означает, что значение выше среднего, отрицательный — ниже, а значение около 0 — близко к среднему; обычно значения от -2 до $+2$ считаются нормальными.

Стоит отметить, что большая часть работ, где используются методы машинного обучения для прогнозирования банковских дефолтов, выполнена на данных развитых стран, среди которых особенно выделяются США и страны ЕС. Это обусловлено большей доступностью массивов банковских микроданных.

Среди стран с формирующимся рынком также наблюдается выраженная асимметрия: фактически все известные нам публикации в авторитетных международных журналах выполнены на данных банков Индии и Турции. Большая часть таких исследований показывает, что наибольшим предсказательным потенциалом для развивающихся рынков является *метод опорных векторов*.

В последние годы появился пласт работ, где авторы применяют различные алгоритмы машинного обучения и сравнивают их эффективность.

В (Tanaka, Kinkyo, Namori, 2016) авторы использовали случайный лес, классификационные деревья и логистическую регрессию. Авторы строили систему индикаторов раннего оповещения для заблаговременного выявления банков в странах ОЭСР, находящихся в преддефолтном состоянии. Авторы показали, что система, построенная с помощью случайного леса, обеспечивает наибольшую точность.

В (Le, Viviani, 2018) были применены следующие методы: нейронная сеть, метод K-ближайших соседей (K-NN), метод опорных векторов, дискриминантный анализ и логистическая регрессия. Была проанализирована выборка из более 3000 американских банков. Наибольшую предсказательную силу продемонстрировали модели, основанные на *нейронной сети и методе K-NN*, однако разрыв с логистической регрессией оказался не таким уж большим.

В (Tuan, Lin, Teng, 2023) авторы использовали метод K-NN, дерево решений, логистическую регрессию в отношении выборки почти из 11 тысяч американских банков. Было показано, что простейшие методы машинного обучения в обычных условиях не превосходят предсказательной силы логистической регрессии. Их преимущество как при внутри-, так и вневыборочной оценке проявляется только после предварительного снижения размерности данных посредством метода главных компонент. Наилучшим методом при этом оказался метод K-NN.

В настоящее время большой популярностью также пользуются *ансамблевые модели* (бустинг, бэггинг, стэкинг⁵). Ансамблевые методы объединяют базовые модели для достижения меньшей дисперсии, большей точности и устойчивости модели, чем это может оказаться при использовании каждого классификатора в отдельности (Citterio, 2024).

Авторы (Climent, Momparler, Carmona, 2019) рассмотрели выборку, состоящую из 156 американских банков. Наиболее перспективным методом (всего было рассмотрено три метода — экстремальный градиентный бустинг, логистическая регрессия и случайный лес) для оценки вероятности дефолта был признан *экстремальный градиентный бустинг*.

Авторы (Kolari, López-Iturriaga, Sanz, 2019) использовали адаптивный бустинг. Выборка состояла из 273 банков стран ЕС. Адаптивный бустинг позволяет с точностью порядка 90% воспроизвести результат стресс-тестирования европейских банков. Также высокая эффективность адаптивного бустинга, но применительно к индийским банкам, была продемонстрирована в работе (Shrivastava, Jeyanthi, Singh, 2020).

Ансамблевые модели направлены на повышение точности прогнозов, однако их использование сопряжено с рядом сложностей.

Во-первых, эффективность ансамблевых методов во многом зависит от разнообразия входящих в них моделей. Модели, допускающие различные ошибки, могут дополнять друг друга, и это приводит к повышению точности прогноза. Однако создание такого разнообразия требует тщательного подбора различных методов обучения и гиперпараметров.

Второй проблемой становятся большие вычислительные затраты из-за необходимости обучения нескольких моделей.

Третья проблема — слабая интерпретируемость. В целом это — общая проблема для всех моделей машинного обучения, однако при использовании ансамблей она усугубляется.

⁵ В бэггинге используются однородные модели, они строятся параллельно. Для расчета усредненного ответа используется бутстреп. В бустинге используются однородные модели, которые строятся последовательно, с учетом ошибок на предыдущем шаге. Стэкинг и блендинг используют разнородные модели и метамоделли.

В-четвертых, хотя ансамблевые методы и направлены на снижение переобучения путем усреднения ошибок отдельных моделей, эту цель не всегда удается достигнуть. Если отдельные модели в ансамбле чрезмерно подогнаны под обучающие данные, ансамбль может унаследовать и даже усилить эту проблему.

Наконец, ансамблевые методы хорошо работают на больших массивах данных. Однако получение и обработка больших наборов данных может быть затруднена, особенно в специализированных областях, где данные могут быть дефицитными, конфиденциальными или дорогостоящими. Синтетические методы дополнения данных являются потенциальными решениями, но их нужно использовать осторожно, чтобы не внести неточностей.

Таким образом, в данном разделе мы показали, что методология прогнозирования банковских дефолтов существенно эволюционировала: от простого анализа финансовых коэффициентов в 1930-е годы до современных ансамблевых алгоритмов машинного обучения. На первом этапе (1930–1960 гг.) исследователи преимущественно сравнивали показатели обанкротившихся и стабильных банков. Затем на смену пришли многомерный дискриминантный анализ (1960–1970 гг.) и логит-регрессия (1970–2000 гг.), которые позволили в большей степени формализовать взаимосвязи между переменными. С 1990-х годов началась эра нейронных сетей, а с 2000-х годов — активное внедрение таких методов машинного обучения, как случайный лес, классификационные деревья, метод опорных векторов, различные варианты бустинга и ансамблевые методы, что стало ответом на широкое использование больших данных в финансовой отрасли. Методы машинного обучения в настоящее время по-прежнему активно применяются, однако нельзя сказать, что частота их использования увеличивается экспоненциально: счет обстоятельных исследований на основе данных методов идет пока на десятки, даже не на сотни.

Сравнение методологических подходов выявило двоякую природу методов машинного обучения. С одной стороны, эти методы демонстрируют лучшую по сравнению с традиционными статистическими моделями способность выявлять нелинейные зависимости, сложные паттерны и работать с большими выборками. С другой, их преимущества часто нивелируются при вневыборочной оценке: так, авторы (Jing, Fang, 2018) обнаружили, что логит-регрессия превосходит методы машинного обучения на новых данных. Авторы (Beutel, List, Schweinitz, 2019; Tuan, Lin, Teng, 2023) зафиксировали те же результаты. Ключевые ограничения методов машинного обучения связаны с риском переобучения и сложностью интерпретации результатов. Это объясняет, почему логит-регрессия сохраняет конкурентоспособность, — ее прозрачность и устойчивость делают метод «золотым стандартом» для прогнозирования банковских дефолтов.

Среди самих методов машинного обучения также нельзя выделить единственный наиболее эффективный. По предсказательной силе, как правило, лидируют нейронные сети, но сопоставимое качество прогнозов также демонстрируют градиентный бустинг и случайный лес. Ключевым трендом сегодня является использование ансамблевых методов.

Итак, методы машинного обучения стали мощным инструментом в прогнозировании банковских дефолтов, но их применение требует учета трех аспектов. Во-первых, региональной специфика данных: существующие модели, обученные на банках ЕС и США, могут быть не релевантны для развивающихся рынков. Во-вторых, для таких моделей обязательная проверка на вневыборочных данных для минимизации переобучения. В-третьих, необходимо соблюдать баланс между точностью и интерпретируемостью: в силу того что такие модели и так дают высокоточные прогнозы, больше внимания стоит уделять отбору предикторов, основываясь на теоретических знаниях. Лица, принимающие решения, должны совершать прозрачные и обоснованные действия, которые необходимо предпринять исходя из прогнозов моделей. Наконец, модели машинного обучения также несовершенны в условиях изменяющейся среды: при структурных изменениях в системе *найденные ранее закономерности перестают работать* и прогнозы дефолтов банков также оказываются несостоятельными.

5. ОТБОР ПРЕДИКТОРОВ ДЛЯ ПРОГНОЗИРОВАНИЯ БАНКРОТСТВ БАНКОВ

Выбор объясняющих переменных имеет большое значение при построении моделей раннего прогнозирования банкротств в банковском секторе.

Традиционно в моделях раннего оповещения о банковских дефолтах используют следующие переменные (Citterio, 2024): показатели из бухгалтерской отчетности; финансовые коэффициенты;

рыночные показатели (рыночная капитализация, доходность акций, кредитные рейтинги и др.); макроэкономические показатели (ВВП, инфляция, государственный долг и пр.); индивидуальные характеристики отдельных банков и всего банковского сектора (расположение, качество управления, модель видения бизнеса и т.д.); показатели, характеризующие политику регулирования банковской отрасли.

Начиная с работ (Altman, 1968; Beaver, 1966), многие исследователи пытались оценить вероятность дефолта с использованием бухгалтерской информации. В качестве значимых предикторов банковских дефолтов чаще всего в исследованиях были идентифицированы совокупная величина активов, доля просроченной задолженности в совокупном кредитном портфеле, а также отношение выданных кредитов к привлеченным депозитам (для банков России одним из первых исследований с использованием балансовых данных была статья (Пересецкий, 2007)).

Несмотря на то что бухгалтерские показатели легко доступны из публичной отчетности банка, их использование в моделях прогнозирования дефолтов связано с некоторыми сложностями. Во-первых, информация из бухгалтерского баланса «обращена в прошлое», т.е. характеризует состояние банка в прошлом и может не подходить для предсказания будущего дефолта. В (Hillegeist et al., 2004) подчеркивается, что баланс формируется по принципу «непрерывности деятельности», что само по себе предполагает, что компания не обанкротится в будущем. В итоге вероятность банкротства можно предсказать только при существенном превышении пороговых значений балансовыми показателями.

Кроме того, некоторая свобода действий со стороны бухгалтера может быть использована для манипулирования данными отчетности и предоставления неточной информации о стоимости и ликвидности активов компании.

В работе (Meuer, Pifer, 1970) продемонстрировано, что слабые финансовые коэффициенты могут являться предикторами будущего дефолта банка. В (Zhao, Sinha, Ge, 2009) показано, что использование коэффициентов вместо сырых балансовых показателей существенно повышает точность прогнозов.

После того как власти США стали использовать систему CAMELS для оценки состояния банков, многие исследователи стали включать в системы опережающих индикаторов финансовые коэффициенты из этой системы. Несмотря на большую популярность системы CAMELS⁶, некоторые исследования показывают, что предсказательной силы индикаторов из этого набора может быть недостаточно. Так, например, в (Canbas, Sabuk, Kilic, 2005) продемонстрировано, что CAMELS работает не лучшим образом на выборке турецких банков.

Относительно использования рыночных данных в моделях прогнозирования банковских дефолтов имеются смешанные результаты. Так, в (Curgu, Elmer, Fissel, 2007) авторы демонстрируют, что внедрение таких показателей в наборы данных бухгалтерских показателей может улучшить предиктивную способность моделей. В (Poghosyan, Ćihak, 2011) авторы подтверждают этот вывод. Тем не менее, в (Distinguin, Rous, Tarazi, 2016) авторы показывают, что предсказательной силы рыночных цен, аномальной избыточной доходности и стандартных отклонений ежедневных доходностей недостаточно, если банки полагаются на систему страхования вкладов. В (Avkiran, Cai, 2012) авторы продемонстрировали, что ни аннуализированные доходности акций⁷, ни рыночная капитализация не улучшают способности различать здоровые банки и банки с финансовыми трудностями.

Авторы (Bongini, Laeven, Majnoni, 2002) использовали для прогнозирования дефолтов в Западной Азии в 1996–1997 гг. бухгалтерские показатели, цены на фондовом рынке и кредитные рейтинги. Согласно их расчетам, цены на фондовом рынке обладают меньшей предиктивной способностью по сравнению с бухгалтерскими показателями, но реагируют на изменяющиеся финансовые условия быстрее, чем рейтинги кредитных агентств, что отражает способность рыночных показателей быстрее впитывать новую информацию по сравнению с бухгалтерскими данными.

⁶ Система CAMELS представляет собой методику оценки надежности и финансового состояния банков, где каждая буква обозначает некий критерий: capital (капитал), assets quality (качество активов), management (управление), earnings (доход), liquidity (ликвидность) и sensitivity to market risk (чувствительность к рыночным рискам). Используется банковскими регуляторами для всестороннего анализа банковской деятельности, помогая оценить их устойчивость и выявить потенциальные проблемы.

⁷ Чтобы правильно учитывать волатильность доходностей и их влияние на результат, используют среднюю геометрическую, или аннуализированную, доходность (compound average growth rate). Она рассчитывается как корень степени n из произведения доходностей за n периодов.

Однако необходимо учитывать, что рыночная информация, как правило, может быть полезной для прогнозирования лишь на ограниченном горизонте времени, а ее информативность зависит от того, считаем ли мы справедливым предположение об эффективности рынков. В силу того что финансовые рынки часто подвержены влиянию асимметрии информации и иррационального поведения, рыночные показатели могут служить лишь дополнением к другим переменным, включаемым в модели раннего оповещения.

Некоторые авторы подчеркивают возможность включения в модели прогнозирования дефолтов, помимо вышеперечисленного, таких дополнительных переменных, как макроэкономические и нефинансовые показатели.

Исследования показывают, что ухудшающаяся макроэкономическая ситуация может повышать вероятность банкротства банков. Несмотря на некоторые неопределенные результаты относительно информативности макроэкономических показателей для прогнозирования банковских дефолтов, большинство исследований приходят к выводу, что дефолты чаще происходили в странах с более слабым ростом ВВП и более высокой инфляцией (Flannery, 1998; Jagtiani, Lemieux, 2001).

В литературе также исследована взаимосвязь между несколькими специфическими для страны показателями банковского сектора и вероятностью банковских дефолтов. К основным предикторам банковских дефолтов в этой группе переменных относят уровень концентрации, характер регулирования и наличие/отсутствие государственного вмешательства.

Конкуренция в банковском секторе обычно оценивается с помощью индекса Лернера⁸ и индекса Херфиндаля–Хиршмана⁹. Литература представляет два противоположных взгляда на взаимосвязь между концентрацией в банковском секторе и устойчивостью отдельных банков: концепция «концентрации–хрупкости», согласно которой банки, расположенные в более концентрированном банковском секторе, оказываются более уязвимыми, и концепция «концентрации–стабильности», которая предполагает отрицательную взаимосвязь между рыночной концентрацией и вероятностью дефолта.

Также в исследованиях отмечается влияние банковского регулирования и государственного вмешательства на вероятность дефолта банков. В статье (Arena, 2008) проанализированы банковские кризисы в Восточной Азии и Латинской Америке и доказано, что регулирующая и институциональная среда должны учитываться при анализе банкротства.

В статье (Davis, Karim, 2008) оценивается влияние финансовой либерализации на стабильность банков на выборке из 105 банков за период с 1979 по 2003 г. Результаты показывают, что в странах, где наблюдалась либерализация финансового сектора, усиление конкуренции может подорвать внутреннюю устойчивость отдельных банков. Без эффективного надзора и регулирования банки могут отказаться от разумной оценки кредитного риска в стремлении привлечь заемщиков. Такая банковская практика усиливается наличием государственных защитных сетей.

В целом литература показывает, что для повышения точности прогнозов банковских дефолтов необходимо использовать разнообразные переменные. Макроэкономические переменные помогают понять, когда проблемы наиболее вероятны, в то время как специфические для банка переменные помогают выявить банки, наиболее подверженные негативным факторам. Структура банковского сектора играет значимую роль в объяснении того, какие страны, вероятно, столкнутся с проблемами при данных макроэкономических и банковских условиях.

Относительно недавний тренд — включение в состав предикторов банковских дефолтов различных показателей, относящихся к ESG-концепции. Наибольшее внимание при этом уделяется роли корпоративного управления. В (Pathan, 2009) показано, что небольшие по размеру советы директоров и советы директоров, которые не находятся под контролем CEO, были ассоциированы с повышенными рисками американских банков в период с 1997–2004 гг.

В статье (Switzer, Wang, 2013b) исследуется та же взаимосвязь между корпоративным управлением и риском дефолта банка, используя CDS-спреды на выборке из 228 банков США за период с 2001 по 2010 г. Авторы продемонстрировали, что коммерческие банки с более многочисленными и более независимыми советами директоров характеризуются более низким кредитным риском.

⁸ Индекс Лернера представляет собой степень, в которой рыночная власть позволяет компаниям устанавливать цену выше предельных издержек.

⁹ Индекс Херфиндаля–Хиршмана измеряет концентрацию банковской системы как сумму квадратов долей рынка каждого банка, конкурирующего в стране в данный период.

На выборке из 341 банка США за 2007–2010 гг. в работе (Berger, Imbierowicz, Rauch, 2016) показано, что структура собственности играет первостепенную роль в объяснении вероятности краха. Чем выше доля акций у линейных сотрудников и менеджеров среднего звена, тем более выражена связь с риском дефолта. В статье (Switzer, Tu, Wang, 2018) авторы анализируют эту взаимосвязь для выборки финансовых компаний в 28 странах за пределами Северной Америки. Результаты показывают, что институциональная структура собственности и независимость совета негативно связаны с вероятностью дефолта, в то время как большая доля сотрудников, владеющих акциями компании, выполнение председателем совета директоров функций CEO и размер совета директоров повышают вероятность дефолта. Интересно, что статья подчеркивает большее влияние рассмотренных показателей на азиатские компании, чем на европейские.

Есть также ряд работ, в которых изучается влияние гендерного состава совета директоров на устойчивость банков (Gulamhussen, Santa, 2015; Cardillo, Onali, Torluccio 2021). Делается вывод о снижении индивидуальных рисков банков и улучшении финансовых показателей при включении женщин в совет директоров.

Исследований, в которых рассматривается взаимосвязь между дефолтами банков и оставшимися двумя компонентами ESG, т.е. экологическими и социальными факторами, гораздо меньше. Так, в (Gangi et al., 2019) показано, что риск дефолта банков снижается при приверженности банка своей повестке по охране окружающей среды. Аналогичные выводы продемонстрированы и в (Galletta, Mazzù, 2023). В статье (Neitzert, Petras, 2022) рассматривается взаимосвязь между всеми тремя компонентами ESG-счета (количественная оценка деятельности компании по трем направлениям: экология (environmental), социальная ответственность (social) и корпоративное управление (governance)) американских банков и уровнем рисков, принимаемых банками в 2002–2018 гг. Авторы приходят к выводу, что наиболее сильным снижающим влиянием на принимаемые риски (как на риск дефолта, так и на риск портфеля активов) обладает компонента, связанная с защитой окружающей среды. В отношении системы социальной ответственности и корпоративного управления выводы оказываются менее однозначными — авторы подтверждают статистически значимое влияние этих компонентов на снижение риска дефолта, но не выявляют его для риска портфеля активов.

В статье (Citterio, King, 2023) продемонстрировано, что включение ESG-показателей в модели раннего оповещения банковских дефолтов (авторы используют широкую альтернативных методов, среди которых есть как методы машинного обучения, так и статистические методы) существенно снижает ошибки классификации.

ЗАКЛЮЧЕНИЕ

Цель представленного систематического обзора литературы — рассмотреть основные тренды в моделировании банковских дефолтов. Мы рассматриваем всю историю развития подходов к построению моделей банковских дефолтов в 1960–2024 гг. Данный обзор сфокусирован на том, какие данные и методы используются для прогнозирования дефолтов банков, а также систематизирует сложности, с которыми сталкиваются исследователи при обработке данных.

Наш вклад в систематический обзор литературы о банкротствах банков заключается в расширении периода анализа, круга рассматриваемых исследовательских проблем, в частности связанных с внедрением новых методов обработки данных, выявлением их сильных и слабых сторон.

По результатам обзора мы приходим к следующим выводам.

Несмотря на относительно долгую историю исследования вопроса банковских дефолтов, в литературе остается много противоречий. Во-первых, они касаются метода прогнозирования банкротства. Начиная с 1960-х годов, эволюция используемых методов прошла путь от простого сравнения финансовых показателей здоровых и проблемных банков до активного использования методов машинного обучения. Основные преимущества моделей машинного обучения сводятся к возможности учитывать нелинейные взаимосвязи и добиваться довольно высокой точности прогнозов внутри выборки. Недостатки связаны со сложностями в интерпретации результатов. Кроме того, такие модели могут быть чувствительны к выбросам и подвержены переобучению.

Альтернатива методам машинного обучения — традиционные эконометрические модели, которые помогают выявлять значимые факторы, связанные с вероятностью банковских дефолтов, но они не применяются в целях прогнозирования.

Наиболее популярным методологическим подходом в настоящее время является использование ансамблевых алгоритмов и применение серии конкурирующих алгоритмов для последующего сравнения их эффективности.

Вероятно, дальнейшее развитие методологии будет сопряжено с нахождением оптимального сочетания параметрических и непараметрических методов в различных комбинациях для обеспечения лучшего понимания (и предсказания) дефолтов в банковском секторе.

Сохраняются противоречия относительно наиболее информативных предикторов дефолтов, так как для разных стран перечни подобных индикаторов заметно разнятся. Мы делаем вывод, что традиционный набор, в который входят финансовые, макроэкономические и рыночные показатели, целесообразно дополнить ESG-показателями и другими качественными и институциональными индикаторами.

Таким образом, перечисленные методологические сложности и вызовы делают изучение вопроса прогнозирования банковских дефолтов по-прежнему актуальным.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Карминский А.М.** (2015). *Кредитные рейтинги и их моделирование*. М.: Издательский дом «Высшей школы экономики». [**Karminsky A.M.** (2015). *Credit ratings and their modeling*. Moscow: Publishing House of the “Higher School of Economics” (in Russian).]
- Пересецкий А.А.** (2007). Методы оценки вероятности дефолта банков // *Экономика и математические методы*. Т. 43. № 3. С. 37–62. [**Peresetsky A.A.** (2007). Methods for assessing the probability of bank default. *Economics and Mathematical Methods*, 43 (3), 37–62 (in Russian).]
- Altman E.** (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23, 4, 589–609.
- Arena M.** (2008). Bank failures and bank fundamentals: A comparative analysis of Latin America and East Asia during the nineties using bank-level data. *Journal of Banking & Finance*, 32, 2, 299–310.
- Avkiran N., Cai L.** (2012). Predicting bank financial distress prior to crises. In: “*New Zealand Finance Colloquium*”. Albany: Massey University, 8–10.
- Balcaen S., Ooghe H.** (2006). 35 Years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38, 1, 63–93.
- Beaver W.** (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. DOI: 10.2307/2490171
- Berger A., Imbierowicz B., Rauch C.** (2016). The roles of corporate governance in bank failures during the recent financial crisis. *Journal of Money, Credit and Banking*, 48, 4, 729–770.
- Beutel J., List S., Schweinitz G. von** (2019). Does machine learning help us predict banking crises? *Journal of Financial Stability*, 45, 100693.
- Bongini P., Laeven L., Majnoni G.** (2002). How good is the market at assessing bank fragility? A horse race between different indicators. *Ratings, Rating Agencies and the Global Financial System*, 159–176.
- Bovenzi J., Marino J., McFadden F.** (1983). Commercial bank failure prediction models. *Economic Review*, 68, 1, 14–26.
- Braga M.J., Fully Bressan V., Colosimo E., Bressan A.** (2006). Investigating the solvency of Brazilian credit unions using a proportional hazard model. *Annals of Public and Cooperative Economics*, 77, 1, 83–106.
- Canbas S., Cabuk A., Kilic S.B.** (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166, 2, 528–546.
- Cardillo G., Onali E., Torluccio G.** (2021). Does gender diversity on banks’ boards matter? Evidence from Public Bailouts. *Journal of Corporate Finance*, 71, 101560.
- Chiaramonte L., Croci E., Poli F.** (2015). Should we trust the z-score? Evidence from the European banking industry. *Global Finance Journal*, 28, 111–131.
- Cipollini A., Fiordelisi F.** (2012). Economic value, competition and financial distress in the European banking system. *Journal of Banking and Finance*, 36 (11), 3101–3109.
- Citterio A.** (2024). Bank failure prediction models: Review and outlook. *Socio-Economic Planning Sciences*, 101818. DOI:10.1016/j.seps.2024.101818

- Citterio A., King T.** (2023). The role of environmental, social, and governance (ESG) in predicting bank financial distress. *Finance Research Letters*, 51, 103411.
- Climent F., Momparler A., Carmona P.** (2019). Anticipating bank distress in the Eurozone: An extreme gradient boosting approach. *Journal of Business Research*, 101, 885–896.
- Cole R., White L.** (2012). Déjà vu all over again: The causes of U.S. commercial bank failures this time around. *Journal of Financial Services Research*, 42, 5–29.
- Curry T., Elmer P., Fissel G.** (2007). Equity market data, bank failures and market efficiency. *Journal of Economics and Business*, 59, 6, 536–559.
- Davis E., Karim D.** (2008). Comparing early warning systems for banking crises. *Journal of Financial Stability*, 4, 2, 89–120.
- DeYoung R., Torna G.** (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, 22 (3), 397–421.
- Diallo B.** (2015). Bank competition and crises revisited: New results. *Economics Letters*, 129, 81–86.
- Diamond D., Dybvig P.** (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91, 401–419.
- Distinguin I., Rous P., Tarazi A.** (2006). Market discipline and the use of stock market data to predict bank financial distress. *Journal of Financial Services Research*, 30, 2, 151–176.
- Erer E., Erer D.** (2024). The domino effect of Silicon Valley Bank's bankruptcy and the role of Fed's monetary policy. *Borsa Istanbul Review*, 24 (3), 573–591.
- Flannery M.** (1998). Using market information in prudential bank supervision: A review of the US empirical evidence. *Journal of Money, Credit and Banking*, August, 273–305.
- Fungacova Z., Turk R., Weill L.** (2021). High liquidity creation and bank failures. *Journal of Financial Stability*, 57, 100937.
- Galletta S., Mazzù S.** (2023). ESG controversies and bank risk taking. *Business Strategy and the Environment*, 32 (1), 274–288.
- Gangi F., Meles A., D'Angelo E., Daniele L.** (2019). Sustainable development and corporate governance in the financial system: Are environmentally friendly banks less risky. *Corporate Social Responsibility and Environmental Management*, 26 (3), 529–547.
- Giordana G., Schumacher I.** (2017). An empirical study on the impact of Basel III standards on banks' default risk: The case of Luxembourg. *Journal of Risk and Financial Management*, 10 (2), 8.
- Goetz M.** (2018). Competition and bank stability. *Journal of Financial Intermediation*, 35, 57–69.
- Gogas P., Papadimitriou T., Agrapetidou A.** (2018). Forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting*, 34, 3, 440–455.
- Gulamhussen M., Santa S.** (2015). Female directors in bank boardrooms and their influence on performance and risk-taking. *Global Finance Journal*, 28, 10–23.
- Haixiang G., Yijing L., Shang J., Mingyun G., Yuanue H., Bing G.** (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73, 220–239.
- Halteh K., Kumar K., Gepp A.** (2018). Financial distress prediction of Islamic banks using tree-based stochastic techniques. *Managerial Finance*, 44, 6, 759–773.
- Hillegeist S., Keating E., Cram D., Lundstedt K.** (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9, 5–34.
- Jagtiani J., Lemieux C.** (2001). Market discipline prior to bank failure. *Journal of Economics and Business*, 53, 2–3, 313–324.
- Jing Z., Fang Y.** (2018). Predicting US bank failures: A comparison of logit and data mining models. *Journal of Forecasting*, 37, 2, 235–256.
- Jullum M., Løland A., Huseby R.B., Ånonsen G., Lorentzen J.** (2020). Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control*, 23 (1), 173–186.
- Karminsky A., Khromova E.** (2016). Modelling banks' credit ratings of international agencies. *Eurasian Economic Review*, 6, 341–363.
- Kashyap A., Rajan R., Stein J.** (2002). Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking. *Journal of Finance*, 57 (1), 33–73.
- Köhler M.** (2015). Which banks are more risky? The impact of business models on bank stability. *Journal of Financial Stability*, 16, 195–212.

- Kolari J., López-Iturriaga F., Sanz I.** (2019). Predicting European bank stress tests: Survival of the fittest. *Global Finance Journal*, 39, 44–57.
- Lane W., Looney S., Wansley J.** (1986). An application of the cox proportional hazards model to bank failure. *Journal of Banking Finance*, 10, 4, 511–531.
- Le H., Viviani J.** (2018). Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research in International Business and Finance*, 44, 16–25.
- Lokanan M.** (2024). Predicting money laundering using machine learning and artificial neural networks algorithms in banks. *Journal of Applied Security Research*, 19 (1), 20–44.
- Malikidou D., Bräuning M., Scalone S.** (2019). A new approach to early warning systems for smaller European banks. *ECB Working Paper no. 2348*, December.
- Martin D.** (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, 1, 3, 249–276.
- Meyer P., Pifer H.** (1970). Prediction of bank failures. *The Journal of Finance*, 25, 4, 853–868.
- Mompalmer A., Carmona P., Climent F.** (2016). Banking failure prediction: A boosting classification tree approach. *Spanish Journal of Finance and Accounting*, 45, 1, 63–91.
- Neitzert F., Petras M.** (2022). Corporate social responsibility and bank risk. *Journal of Business Economics*, 92 (3), 397–428.
- Pathan S.** (2009). Strong boards, CEO power and bank risk-taking. *Journal of Banking & Finance*, 33 (7), 1340–1350.
- Podpiera J., Weill L.** (2008). Bad luck or bad management? Emerging banking market experience. *Journal of Financial Stability*, 4 (2), 135–148.
- Poghosyan T., Čihak M.** (2011). Determinants of bank distress in Europe: Evidence from a new data set. *Journal of Financial Services Research*, 40, 163–184.
- Salchenberger L., Cinar E., Lash N.** (1992). Neural networks: A new tool for predicting thrift Failures. *Decision Sciences*, 23, 4, 899–916.
- Serrano-Cinca C., Fuertes-Callén Y., Gutiérrez-Nieto B., Cuellar-Fernández B.** (2014). Path modelling to bankruptcy: Causes and symptoms of the banking crisis. *Applied Economics*, 46 (31), 3798–3811.
- Shrivastava S., Jeyanthi P., Singh S.** (2020). Failure prediction of Indian banks using SMOTE, lasso regression, bagging and boosting. *Cogent Economics & Finance*, 8, 1, 1729569.
- Sinkey J. Jr.** (1975). A multivariate statistical analysis of the characteristics of problem banks. *The Journal of Finance*, 30, 1, 21–36.
- Switzer L., Wang J.** (2013a). Default risk and corporate governance in financial vs. non-financial firms. *Risk and Decision Analysis*, 4, 4, 243–253.
- Switzer L., Wang J.** (2013b). Default risk estimation, bank credit risk, and corporate governance. *Financial Markets, Institutions and Instruments*, 22 (2), 91–112.
- Switzer L., Tu Q., Wang J.** (2018). Corporate governance and default risk in financial firms over the post-financial crisis period: International evidence. *Journal of International Financial Markets, Institutions and Money*, 52, 196–210.
- Tam K.** (1991). Neural network models and the prediction of bank bankruptcy. *Omega*, 19, 5, 429–445.
- Tanaka K., Kinkyo T., Hamori S.** (2016). Random forests-based early warning system for bank failures. *Economics Letters*, 148, 118–121.
- Tuan L.Q., Lin C.Y., Teng H.W.** (2023). Machine learning methods for predicting failures of US commercial bank. *Applied Economics Letters*, Mar., 1–7. DOI: 10.1080/13504851.2023.2186353
- Whalen G.** (1991). A proportional hazards model of bank failure: An examination of its usefulness as an early warning Tool. *Economic Review*, 27, 1, 21–31.
- Zaki E., Bah R., Rao A.** (2011). Assessing probabilities of financial distress of banks in UAE. *International Journal of Managerial Finance*, 7, 3, 304–320.
- Zhao H., Sinha A., Ge W.** (2009). Effects of feature construction on classification performance: An empirical study in bank failure prediction. *Expert Systems with Applications*, 36, 2, 2633–2644.

Forecasting bank defaults: Evolution of methods, models and risk factors

© 2026 М.А. Shchepeleva, М.И. Stolbov

M.A. Shchepeleva,

Cand. Sc. (Economics), Associate Professor, National Research University Higher School of Economics, Moscow, Russia; e-mail: mshchepeleva@hse.ru, mschepeleva@yandex.ru.

M.I. Stolbov,

Doct. Sc. (Economics), Chair of the Department of Applied Economics, Professor at the Department of Economics and Banking Business, Moscow State Institute of International Relations (MGIMO University), Moscow, Russia; e-mail: stolbov.m@my.mgimo.ru

Received 29.04.2025

The research was funded by the Russian Science Foundation (project No. 23-18-00756).

Abstract. Predicting bank defaults is an important task for the entire economy. Early identification of troubled banks helps to prevent impending bank failures or minimize the losses associated with them. The paper discusses the state of the art of instrumental methods and data used for this purpose. The theoretical background, the evolution of methodological approaches used to predict bank defaults, the specifics of data handling, and the lists of predictors that are included in early warning models are successively reviewed. We conclude that there is still considerable controversy in the literature regarding both the methods and the variables to be used in predictive models. Machine learning methods show a better ability than traditional statistical models to detect non-linear dependencies and to handle large samples. Their advantages are often offset by out-of-sample estimation. Other limitations of such methods are the risk of overfitting and the difficulty in interpreting the results. The lists of potential predictors of bank defaults also vary from country to country. Most commonly, predictive models use bank balance sheet data and financial ratios. However, there are studies that show that forecast accuracy improves when market, macroeconomic and non-financial indicators are included for special countries. Prospects for further research in this area include finding an optimal combination of parametric and non-parametric approaches, investigating the potential of non-financial indicators as factors in bank failures, and research on large samples including both developed and developing countries.

Keywords: bank default, machine learning, early warning systems, bank default risk factors.

JEL Classification: G17, G21.

UDC: 336.71.

For reference: **Shchepeleva M.A., Stolbov M.I.** (2026). Forecasting bank defaults: Evolution of methods, models and risk factors. *Economics and Mathematical Methods*, 62, 1, 63–77. DOI: 10.7868/S3034617726010057 (in Russian).