

Разработка модели прогнозирования вероятности дефолта по субъектам Российской Федерации и муниципалитетам

Бубаня Ксения Костовна
РЭУ им. Г.В. Плеханова
студент

Аннотация

В статье посвящена исследованию и описанию проекта по разработке модели оценки вероятности дефолта клиентов из сегмента административно-территориальных образований местного и регионального значения, включающих в себя Субъекты РФ, для использования в скоринговых системах Банка и иных кредитных процессах средствами Python. Материалы и методы: официальные открытые источники: базы данных рейтинговых агентств, сайт Казначейства России, каталог публикаций «Регионы России. Социально-экономические показатели», сайт Налоговой службы Российской Федерации, Долговая Книга субъектов Российской Федерации, системный анализ, машинное обучение, а также метод множественной линейной регрессии.

Ключевые слова: прогнозирование, машинное обучение, PD, дефолт, субъекты РФ, макроэкономика, прогнозирование данных

Development of a model for predicting the probability of default for the subjects of the Russian Federation and municipalities

Bubanya Kseniya Kostovna
Plekhanov Russian University of Economic
student

Abstract

The article is devoted to the study and description of the project on the development of a model for estimating the probability of default of customers from the segment of local and regional administrative and territorial entities, including the Subjects of the Russian Federation, for use in the scoring systems of the Bank and other credit processes using Python. Materials and methods: official open sources: databases of rating agencies, site of the Treasury of Russia, catalog of publications «Regions of Russia. Socio-economic indicators», the site of the Tax Service of the Russian Federation, the Debt Book of the Subjects of the Russian Federation, system analysis, machine learning, and the multiple linear regression method.

Keywords: forecasting, machine learning, PD, default, subjects of the Russian Federation, macroeconomics, data forecasting

Введение

Консервативная политика управления рисками является одним из ключевых элементов бизнес-моделей современных коммерческих банков. Исходя из этого, в данной исследовательской работе описан процесс разработки модели оценки вероятности дефолта на уровне региональной аналитики.

Целью работы является создание модели оценки вероятности дефолта клиентов из сегмента административно-территориальных образований местного и регионального значения, включающих в себя Субъекты РФ, для использования в скоринговых системах Банка и иных кредитных процессах [2].

Актуальность исследования заключается в том, что ключевым драйвером снижения расходов на риски в коммерческих банках является совершенствование системы управления рисками за счёт внедрения продвинутой аналитики в процессы риск-менеджмента, мониторинга риска в режиме реального времени и поддержания высокого качества портфеля [4].

Контрагенты, по которым строится модель, определяются как административно-территориальные образования местного и регионального значения, и включают в себя Субъекты Российской Федерации и муниципалитеты.

Определение вероятности дефолта

Модели оценки вероятности дефолта (PD, probability of default) играют важную роль в риск-менеджменте коммерческих банков, поскольку позволяют оценить кредитоспособность контрагентов и разработать свод единых бизнес-правил работы с клиентскими сегментами [12]. В рамках данной работы под дефолтом целесообразно понимать определения дефолтов рейтинговых агентств, описанные в таблице (табл. 1). Методология Базель II предполагает использование оценки кредитоспособности кредитного портфеля с использованием внутренних рейтинговых моделей для оценки кредитного риска, что требует разработки отдельных моделей для клиентов различных групп [27].

Таблица 1 – Определения дефолтов рейтинговых агентств

Определение дефолта по Moody's	Определение дефолта по S&P	Определение дефолта по Fitch
Пропущенная или отсроченная выплата процентов и / или основной суммы	Невыплата неустойки одного или нескольких своих финансовых обязательств (оцененных или не оцененных), если S&P не считает, что такие платежи будут производиться в течение пяти рабочих дней, независимо от любого	Неспособность должника своевременно выплачивать основную сумму и / или проценты по договорным условиям любого финансового обязательства

	льготного периода	
Банкротство, администрация, юридическое правонарушение или другие правовые блоки (возможно, регулирующие органы) для своевременной выплаты процентов и / или принцепала	Подача эмитентом заявки на банкротство или принятие аналогичных мер, которые ставят под угрозу платежи по финансовому обязательству. Заемщик находится под контролем надзора из-за его финансового состояния	Подача заявки на банкротство, администрирование, конкурсное производство, ликвидация или иное прекращение или прекращение деятельности должника.
Эмитент предлагает держателям долговых обязательств новую ценную бумагу или пакет ценных бумаг, размер которых ограничен уменьшенным финансовым обязательством (таким как привилегированные или обыкновенные акции, или задолженность с более низким купоном или номинальной стоимостью, более низкий стаж или более длительный срок погашения); и обмен имеет эффект, позволяющий эмитенту избежать банкротства или невыполнения платежа.	Завершение предложения о бедственном обмене, в результате которого одно или несколько финансовых обязательств либо выкупаются на сумму денежных средств, либо заменяются другими инструментами, имеющими общую стоимость, которая меньше номинальной.	Обмен проблемным долгом (DDE) обязательства, когда кредиторам предлагались ценные бумаги с уменьшенными структурными или экономическими условиями по сравнению с существующим обязательством.

PD (Probability of Default) – это вероятность наступления дефолта в течение годового горизонта прогнозирования. Данная компонента является важной характеристикой кредитоспособности контрагента [28]. Модели оценки вероятности дефолта играют важную роль в системах риск-менеджмента коммерческих банков. Риск-менеджмент дает инструменты, которые позволяют руководителю выявить, оценить и управлять нежеланными ситуациями [7]. Целью стандартов риск-менеджмента является установление контекста как мероприятия в начале общего процесса управления рисками. Под установлением контекста можно понимать фиксирование целей организации, условия, при которых организация пытается достичь своих целей. Международные стандарты управления рисками могут быть взяты за основу при разработке внутрикорпоративных стандартов риск-менеджмента [29].

Принятым в рассматриваемой организации стандартом риск-менеджмента является Базель II, выпущенный Базельским комитетом по банковскому надзору. Базель II (Basel II) — соглашение об определении достаточности банковского капитала для обеспечения эффективной и стабильной деятельности банковских учреждений и их регулирования. Оно

было принято Базельским комитетом по банковскому надзору в 2004 г. на замену предыдущему соглашению (Базель I), введенному в 1988 г. Целью принятия Базеля II является содействие адекватной капитализации банков, совершенствованию систем управления рисками, и, таким образом, усилению стабильности банковской системы в целом. Для этого регулятивный капитал банка должен соответствовать его реальным рискам, покрывать все основные риски и определять объем операций, который может осуществлять банк в соответствии с уровнем своего капитала [3].

Базель II ориентирован преимущественно на банки, которые являются активными участниками международного рынка и предусматривает распространение надзора на банковские группы и холдинговые компании, то есть касается не только активов банков, но и страховых и других организаций [13]. С целью повышения точности оценки рисков, связанных с различными категориями активов, и их чувствительности к качеству этих активов, Базельский комитет предоставляет возможность использовать внешние кредитные рейтинги специализированных финансовых агентств, среди которых наиболее известными являются Standard and Poor's, Moody's Service, Fitch IBCA [1].

При построении моделей PD для Субъектов Российской Федерации рассматривались различные способы построения модели [26]. Перечень и описание трёх возможных подходов для построения модели отражены ниже (табл. 2).

Таблица 2 – Возможные подходы к построению PD модели

	1. Статистический анализ клиентов в дефолте и не в дефолте	2. Использование внешних рейтингов (shadowrating)	3. Экспертное ранжирование
Методология	Анализ факторов, предсказывающих дефолт, на основе статистического сравнения клиентов в дефолте и не в дефолте	Построение статистической модели, аппроксимирующей внешние рейтинги, и использование этой модели для клиентов без рейтинга	Выбор факторов и их весов путем построения модели на основе экспертного ранжирования клиентов
Необходимые данные	Наличие достаточного для обеспечения статистической значимости количества данных по дефолтам	Подход применяется в случае отсутствия данных о дефолтах заемщиков и наличии достоверной информации о внешних рейтингах	Подход применяется в случае отсутствия данных для количественного анализа
Преимущества подхода	Высокое качество модели при наличии достаточного количества данных по дефолтам.	Высокое качество модели при наличии доступных данных о внешних рейтингах. Использование	Нет требований к данным при разработке

	Использование допустимо с точки зрения регулятора	допустимо с точки зрения регулятора	
Недостатки подхода	Необходимо значимое количество дефолтов в портфеле банка	Высокая зависимость от качества внешних рейтингов	Высокая зависимость от качества экспертного суждения. Использование ограничено с точки зрения регулятора

1. Статистический анализ клиентов в дефолте и не в дефолте

Построение модели на основе статистической информации о дефолтах контрагентов Банка, являющихся Субъектами РФ, невозможно ввиду отсутствия достаточного количества данных по дефолтам во внутренних системах Банка. Это является особенностью прогнозирования вероятности дефолта по низкодефолтному портфелю: фактически, ни одного дефолта не было, поскольку речь идёт о государственных территориальных образованиях, однако государственные бюджетные предприятия, являющиеся контрагентами Банка и относящиеся к территории рассматриваемых Субъектов РФ, могут как просрочить выплату долга, так и уйти в дефолт, и прогнозируемая в данной выпускной квалификационной работе вероятность дефолта служит как вспомогательный фактор для оценки кредитоспособности государственных учреждений, являющихся клиентами Банка [5].

2. Использование внешних рейтингов (shadowrating approach)

При использовании подхода на основе внешних рейтингов (shadowrating approach), выбор факторов осуществляется на основе их способности предсказывать PD, соответствующие внешним рейтингам. Процесс разработки модели подразумевает сбор информации о доступных внешних рейтингах, сопоставление внутренних PD Банка внешним рейтингам и проведение количественного анализа на основе полученных значений PD [5].

Внешние рейтинги имеют 74 субъекта РФ. Данное количество наблюдений является достаточным для проведения статистического анализа на базе внешних рейтингов. Данный подход к разработке модели PD не противоречит требованиям Базель II, а также требованиям ЦБ РФ и регуляторным требованиям. В мировой практике подход «внешних рейтингов» является наиболее распространенным методом для построения моделей PD для сходных субсуверенных сегментов.

При применении подхода «внешних» рейтингов в качестве показателя, характеризующего PD, используются рейтинги, присвоенные Субъектам РФ международными рейтинговыми агентствами. В качестве внешних рейтингов использовались следующие рейтинги [17, 18, 19]:

- Moody's Credit View Foreign Currency LT Issuer Rating
- Fitch Issuer Default Rate Foreign Currency LT

- Standard & Poor's LT Issuer Rating

Рейтинги данных рейтинговых агентств признаются ЦБ РФ и являются наиболее распространенными при оценке уровня риска контрагента. Данные виды рейтингов уникальны для каждого заемщика в соответствии с методиками рейтинговых агентств. Рейтинги постоянно проходят мониторинг рейтинговых агентств, то есть актуальны не только на дату публикации, но и в течение всего срока присвоения рейтингов внешними рейтинговыми агентствами.

При построении выборки не отдавалось предпочтения оценке одного определенного международного рейтингового агентства, так как оценка рисков Субъектов РФ в трёх агентствах схожа. При этом в случае, если один субъект одновременно получал оценки нескольких рейтинговых агентств, в выборке учитывалась средняя оценка.

Подход построения модели на базе внешних рейтингов подразумевает использование определения дефолта внешних рейтинговых агентств. Определения дефолта Moody's, S&P и Fitch практически идентичны, поэтому возможно использование рейтингов нескольких рейтинговых агентств при моделировании [6].

3. Экспертное ранжирование

Применение подхода экспертного ранжирования возможно, если использование других подходов невозможно или нецелесообразно. В этом случае заёмщику присваивается рейтинг на основе экспертной оценки, и выбор факторов и их весов осуществляется на основе данного экспертного ранжирования. Однако в случае с Субъектами РФ существуют достоверные данные о присвоении им внешних рейтингов. На основе этой информации было принято решение о нецелесообразности использования данного подхода при разработке модели [5].

Описание источников данных

В качестве исходных данных для базовой модели используются характеристики контрагента, на основе которых определяется балл контрагента. Основными исходными данными для определения балла служат наблюдаемые характеристики контрагента. Факторы, включенные в модель, делятся на следующие типы:

1. Бюджетные факторы и факторы рентабельности – количественные факторы, расчет которых основан на данных бюджета Субъекта РФ и количественных микроэкономических параметрах. К данному типу факторов относятся:

- финансовые соотношения, например, уровень государственного долга – соотношение объема государственного долга и объема общих доходов бюджета за вычетом безвозмездных поступлений;
- финансовые показатели в абсолютном выражении, например, величина собственных доходов контрагента.

2. Макроэкономические факторы - количественные факторы, расчет которых основан на макроэкономических параметрах, отражают демографические и экономические факторы [16]. К данному типу факторов относятся:

- финансовые соотношения, например, доходы на душу населения;
- финансовые показатели в абсолютном выражении, например, численность населения.

Вышеописанные факторы анализируются с помощью набора критериев. С помощью многофакторного анализа из факторов, прошедших первичный отбор, формируется итоговый список факторов. Также на данном этапе определяется вес каждого фактора в модели.

Модель основывается на исторических финансовых данных из бюджетной отчетности субъектов РФ и на исторических макроэкономических данных [14, 15, 16, 23]. Источники информации, использованные для построения модели, перечислены ниже (табл. 3). Таким образом, была учтена вся существенная финансовая и качественная информация. Все необходимые данные были собраны из открытых источников [24].

Таблица 3 – Перечень источников информации для построения выборки

Необходимые данные	Источник информации	Использованные временные периоды
Данные о внешних рейтингах Субъектов РФ	<ul style="list-style-type: none"> • Базы данных рейтинговых агентств Moody's, S&P • Сайт истории кредитных рейтингов Cbonds (http://www.cbonds.ru) 	01/01/2008 – 01/06/2018
Финансовые данные из бюджетов Субъектов РФ	<ul style="list-style-type: none"> • Официальный сайт Казначейства России (http://www.roskazna.ru/ispolnenie-byudzhetrov/byudzhety-subektov-rf-i-mestnye-byudzhety/) • Долговая книга субъекта РФ - Объем и структура государственного долга субъектов Российской Федерации и долга муниципальных образований (https://www.minfin.ru/ru/performance/public_debt/subdbt/#) • Данные по формам статистической налоговой отчетности, отчет о задолженности по налогам и сборам, пеням и налоговым санкциям в бюджетную систему Российской Федерации, форма №4-НМ, раздел 1010 (https://www.nalog.ru/rn77/related_activities/statistics_and_analytics/forms/) 	01/01/2008 – 01/01/2018
Макроэкономические	<ul style="list-style-type: none"> • Каталог публикаций «Регионы России. 	01/01/2008 –

данные	Социально-экономические показатели» Федеральной службы государственной статистики (http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156) • Единая межведомственная информационно – статистическая система (ЕМИСС) (https://fedstat.ru/)	01/01/2018
--------	--	------------

Для разработки модели был использован период 01.01.2008 - 01.01.2018. Начало данного периода в 2008 г. обусловлено наличием бюджетных данных на сайте Казначейства России, начиная с этого времени [20]. Стоит отметить, что до 2011 года отчётная информация велась в формате html. Окончание периода определяется годовой отчётностью по настоящее время.

Данный временной период соответствует регуляторным требованиям, так как включает в себя период экономического спада и длится 10 лет, охватывая полный цикл. Классическое определение экономического спада гласит, что экономическая ситуация может называться рецессией, если два квартала подряд наблюдается отрицательный рост ВВП по сравнению с предыдущим кварталом [8]. Обоснование границ периода экономического спада представлено в приложении (рис. 1).

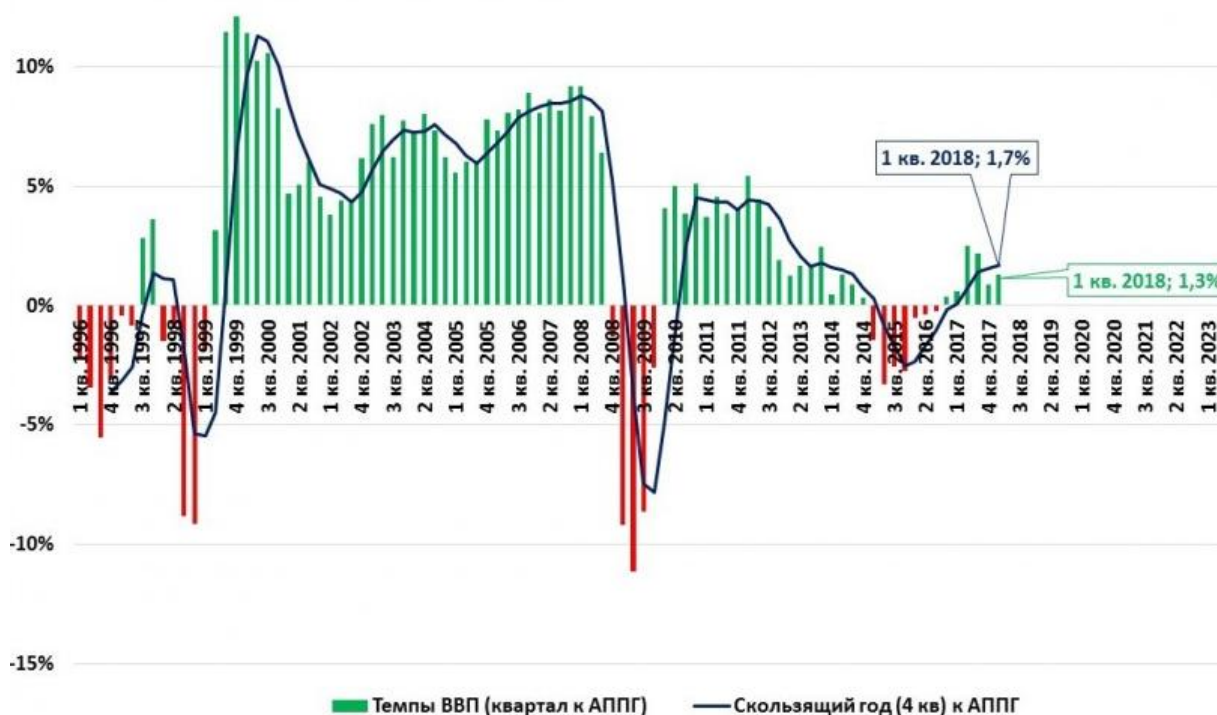


Рисунок 1 – Динамика ВВП РФ с 1996 года по текущий период

Как видно из графика (рис. 1), 4 квартал 2008 года – второй квартал с отрицательным приростом ВВП, следовательно, данный квартал является кварталом начала экономического спада. Также, 3 квартал 2009 года – первый квартал, в котором наблюдается положительный прирост ВВП,

следовательно, данный квартал является периодом окончания рецессии. Исходя из роста ВВП, периодом экономического спада является 4 квартал 2008 года – 3 квартал 2009 года.

В список изначальных факторов, использованных при разработке модели, вошли факторы из таблицы (табл. 4). Источником информации для расчёта бюджетных факторов и план-фактного анализа является Отчёт об исполнении консолидированного бюджета Субъекта Российской Федерации и бюджета территориального государственного внебюджетного фонда. Столбец «Исполнено: бюджет субъекта РФ» использован для фактических показателей и столбец «Утверждено: бюджет субъекта РФ» для плановых показателей.

Из обучающей выборки исключены город Севастополь и республика Крым, поскольку они фигурируют в отчётности с 2015 года, что не соответствует наблюдаемому периоду, а также имеют различные обозначения региона в разные годы [20]. Кроме того, отсутствуют данные внешних рейтинговых агентств по данным субъектам РФ. Также из выборки исключён город Байконур как самостоятельный субъект РФ, поскольку по нему доступна лишь бюджетная статистика. В налоговую отчётность и долговые книги, как и в рейтинги внешних агентств, Байконур не входит.

Использование внешних рейтингов позволяет ориентироваться на изменения рейтинга, поскольку каждый пересмотр рейтинга основан на анализе актуальной бюджетной отчетности, макроэкономических параметров и качественных факторов, используемых рейтинговыми агентствами [25]. Рейтинговые агентства производят постоянный мониторинг присвоенных рейтингов, поэтому стабильный рейтинг также основан на актуальной информации [24].

Общая выборка делится на выборку для разработки модели, в которую входят наблюдения за период 01.01.2008 - 01.01.2017 с имеющимися оценками рейтинговых агентств, на выборку для тестирования, в которую входят наблюдения за 2017 год, имеющие оценку рейтинговых агентств, и выборку для прогнозирования, в которую входят данные за период 01.01.2011 - 01.01.2018 без оценок внешних рейтинговых агентств. Для тестирования выбран период, не участвующий в разработке модели.

Таблица 4 – Перечень используемых показателей

Название	Описание
F1	Уровень государственного долга по итогам отчетного периода
F2	Уровень прямого государственного долга по итогам отчетного периода
F3	Динамика государственного долга
F4	Способность к погашению долга за счет собственных средств
F5	Уровень обслуживания долга
F6	Уровень условных обязательств
F7	Уровень долга без учета бюджетных кредитов
F8	Уровень задолженности по налогам в бюджет

F9	Объем собственных доходов бюджета
F10	Объем общих доходов бюджета
F11	Уровень дефицита бюджета
F12	Уровень безвозмездных поступлений
F13	Исполнение бюджета по доходам
F14	Исполнение бюджета по расходам
F15	Динамика доходов бюджета
F16	Уровень выплат социального характера
F17	Уровень налоговых доходов бюджета
F18	Динамика собственных доходов бюджета
F19	Уровень исполнения собственных доходов бюджета
F20	Динамика налоговых доходов бюджета в отчетном году
F21	Исполнение плана по налоговым доходам бюджета в отчетном году
F22	ВРП на душу населения
F23	Индекс физического объёма ВРП
F24	Доля инвестиций в ВРП
F25	Доля прибыльных предприятий
F26	Доля совокупной прибыли в ВРП
F27	Доходы на душу населения
F28	Диверсификация налоговой базы
F29	Уровень бюджетной обеспеченности региона
F30	Уровень безработицы
F31	Уровень инфляции
F32	Численность населения
F33	Уровень урбанизации
F34	Возрастная структура
F35	Рост населения
F36	Рост уровня безработицы
F37	Рост ВРП на душу населения

Построение модели

Этапы процесса построения финальной модели состоят из формирования списка факторов для машинного обучения с помощью многофакторного анализа и выбора итогового перечня факторов для построения наилучшей модели [10].

Многофакторный анализ включает в себя следующие шаги [21]:

- Трансформация и нормализация значений факторов;
- Обработка пропущенных значений;
- Анализ корреляции между факторами;
- Проведение регрессионного анализа всевозможных комбинаций факторов;
- Обработка полученных данных и выбор лучшей модели исходя из определенных критериев.

Перед началом моделирования был проведён анализ каждого из тридцати семи факторов. Проверялись наличие пропусков, его монотонность,

интуитивность для понимания бизнес-партнёрами и экономическая логика: увеличивается ли вероятность дефолта с увеличением значения или уменьшается.

Характеристики, которыми должны обладать факторы для попадания в обучающую выборку с целью дальнейшего проведения многофакторного анализа [22]:

- Высокая доступность данных
Для того чтобы фактор попал в обучающую выборку, он должен быть доступен для большинства наблюдений в выборке. Так как данные собирались из официальных источников, количество отсутствующих значений в факторах не превышало 20%.
- Сбалансированное распределение значений фактора
Факторы, которые необходимо включить в обучающую выборку, должны иметь сбалансированное распределение значений. В противном случае высокая концентрация значений фактора изначально предполагает низкую дифференцирующую способность фактора.

Предварительный анализ качества данных показал, что перед моделированием данные должны подвергнуться предобработке. Предобработка данных включает в себя трансформацию, нормализацию и обработку пропущенных значений.

В отношении факторов последовательно применяются процедуры логистической трансформации для снижения влияния выбросов и нормализации для приведения всех факторов к единой балльной шкале [24]. В условиях ограниченности данных, учет нелинейных зависимостей для факторов с использованием Weight of Evidence (WoE) преобразования осуществлен не был с целью исключения дополнительного переобучения моделей.

Логистическая трансформация факторов осуществляется по формуле 1:

$$X_{transf} = \frac{1}{1 + \exp\{-Slope \times (X - Midpoint)\}} \quad (1)$$

где X_{transf} – трансформированное значение фактора;

X – значение фактора;

$Slope$ – коэффициент трансформации для фактора X ;

$Midpoint$ – значение, рассчитываемое по формуле 2;

$X_{5\%}$ – значение 5%-ного перцентиля фактора X ;

$X_{95\%}$ – значение 95%-ного перцентиля фактора X .

Формула для нахождения значения $Midpoint$:

$$\frac{X_{5\%} + X_{95\%}}{2} \quad (2)$$

Значения коэффициента трансформации $Slope$ находятся из следующего условия нормировки (3) или (4), которые в данном случае являются равнозначными:

$$\frac{1}{1+\exp\{-Slope \times (X_{5\%} - Midpoint)\}} = 0,05, \quad (3)$$

$$\frac{1}{1+\exp\{-Slope \times (X_{95\%} - Midpoint)\}} = 0,95, \quad (4)$$

где $X_{5\%}$, $X_{95\%}$, $Midpoint$ были определены после формулы 1.

Значения коэффициентов $Slope$ зависят от коэффициента экономической логики: если значение коэффициента у фактора равно «-1», то при вычислении трансформированного значения фактора полученное значение будет вычитаться из 1. Таким образом, после проведения трансформации более высокому трансформированному значению каждого показателя будет соответствовать более низкая вероятность дефолта Субъекта РФ с учетом заложенной в пункте 1 третьей главы экономической логики изменения вероятности дефолта с изменением значения факторов.

В отношении трансформированных значений факторов производится нормализация их значений формуле 5. После осуществления нормализации все факторы приводятся к единой сопоставимой шкале с нулевым средним значением и стандартным отклонением, равным 1.

$$X_{norm} = \left(\frac{X_{transf} - mean(X_{transf})}{std(X_{transf})} \right), \quad (5)$$

где X_{norm} – нормализованное значение фактора;

X_{transf} – трансформированное значение фактора;

$mean(X_{transf})$ – среднее значение трансформированного фактора;

$std(X_{transf})$ – стандартное отклонение трансформированного фактора.

При проведении многофакторного анализа появляется необходимость обработки пропущенных значений. Исключение всех наблюдений хотя бы с одним пропущенным значением приведет к потере значительной доли наблюдений. Исходя из небольшого количества наблюдений в выборках, было принято решение полностью не исключать наблюдения с пропущенными значениями.

Два типа пропущенных значений включают в себя следующие типы:

1. Случайно пропущенные значения. Это значения, пропуски в которых не имеют определенной тенденции. Данные пропуски обусловлены тем, что внешние рейтинговые агентства не обязаны давать оценку Субъектам РФ каждый год.

2. Значения, отсутствующие для определенного сегмента Субъектов РФ. Это значения, пропуски в которых имеют определенную причину.

Часть пропущенных значений, требующих заполнения в ручном режиме, восполнялась в Excel, иные же факторы восполнялись при помощи встроенных функций библиотек sklearn непосредственно в Python [11].

Для оценки дискриминационных способностей отдельных факторов и моделей используется коэффициент Somers'D. Коэффициент Somers'D показывает ранговую корреляцию (взаимосвязь) между факторами (или скоринговыми баллами для многофакторных моделей) и целевой переменной [9]. Для вычисления значений показателя Somers'D для каждого фактора был

написан программный модуль функции “Somers_D. По результатам анализа дискриминационной способности показателей были исключены те факторы, чья предсказательная способность была меньше или равна 10%.

После проведения корректировок пропущенных значений в факторах, проводился анализ корреляции между объясняющими факторами. В результате анализа корреляций между парами факторов были сформированы группы факторов, объединённые как по экономическому смыслу, так и по силе корреляции членов группы друг с другом. Разбиение на группы проводилось членами экспертной группы, при этом в ходе разбиения были исключены некоторые макроэкономические факторы, имевшие сильную непонятную интуитивно корреляцию с бюджетными показателями и факторы, срок публикации которых составляет 28 месяцев с момента завершения отчётного периода, поскольку модель, построенная на таких факторах, будет некорректно вести себя в начале и конце экономических циклов (табл. 5).

Таблица 5 – Группы корреляций

	Название	SomersD	Группа	Название группы
F1	Уровень государственного долга по итогам отчетного периода	46%	1	Debt
F2	Уровень прямого государственного долга по итогам отчетного периода	41%	1	Debt
F4	Способность к погашению долга за счет собственных средств	13%	1	Debt
F5	Уровень обслуживания долга	51%	1	Debt
F7	Уровень долга без учета бюджетных кредитов	42%	1	Debt
F8	Уровень задолженности по налогам в бюджет	16%	1	Debt
F9	Объем собственных доходов бюджета	42%	2	Size
F10	Объем общих доходов бюджета	44%	2	Size
F11	Уровень дефицита бюджета	22%	3	Budget
F12	Уровень безвозмездных поступлений	27%	4	Subsidies
F13	Исполнение бюджета по доходам	30%	3	Budget
F16	Уровень выплат социального характера	31%	4	Subsidies
F17	Уровень налоговых доходов бюджета	24%	4	Subsidies
F19	Уровень исполнения собственных доходов бюджета	28%	3	Budget
F21	Исполнение плана по налоговым доходам бюджета в отчетном году	22%	3	Budget
F25	Доля прибыльных предприятий	20%	8	Macro4
F27	Доходы на душу населения	32%	5	Macro1
F29	Уровень бюджетной обеспеченности региона	17%	6	Macro2
F30	Уровень безработицы	16%	7	Macro3

Чтобы сформировать обучающую выборку, необходимо взять лишь один фактор из каждой группы, или исключить группу целиком. Этот перебор различных вариантов моделей необходим ввиду того, что некоторые факторы могут оказать мультипликативный эффект на общую дискриминационную способность модели. Для выбора наилучшего списка обучающих факторов была сформирована таблица со всеми возможными комбинациями факторов из различных групп. В каждой группе также присутствует единичный вектор, чтобы была возможность полностью исключить целую группу факторов. Всего получилось 6720 различных комбинаций факторов в модели.

Проведение регрессионного анализа всевозможных комбинаций факторов

Целью многофакторного анализа являлось построение зависимости между PD, значения которой были получены на основе соотнесения рейтингов внешних рейтинговых агентств внутренним PD Банка по полученной мастер-шкале, и линейной комбинации трансформированных и нормализованных значений факторов, которые прошли этап корреляционного анализа.

Многофакторный регрессионный анализ позволяет оценить оптимальную комбинацию весов факторов-регрессоров, которая наилучшим образом предсказывает PD, полученную на основе внешних рейтингов.

Линейная регрессия имеет вид:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + \varepsilon, \quad (6)$$

где y является зависимой переменной,

x_i – вектор независимых объясняющих переменных

(трансформированных и нормализованных значений факторов),

β_i – вектор коэффициентов при каждой из объясняющих переменных,

α – константа,

ε – случайная ошибка.

Ввиду того, что целью многофакторной регрессионной модели является предсказание PD, необходимо провести логарифмическое преобразование значений вероятности дефолта. Таким образом, финальная регрессионная модель, которая использовалась для анализа, имеет вид:

$$\ln\left(\frac{PD}{1-PD}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + \varepsilon. \quad (7)$$

Данная спецификация регрессионной модели оценивалась методом наименьших квадратов.

На каждом множестве факторов обучающей выборки из таблицы комбинаций факторов была построена модель множественной регрессии и оценена её дискриминационная способность. При построении моделей автоматически формировался массив, содержащий сведения о вошедших в модель параметрах, коэффициенте R^2 , значении Somers' D, коэффициенты линейной регрессии и нормированные веса факторов в модели, удовлетворяющий следующим условиям:

1. Все коэффициенты линейной регрессии имеют неположительные значения, чтобы избежать некорректного построения модели, при которой более высокому значению фактора будет соответствовать более высокое значение вероятности дефолта.

2. Значение коэффициента Somers' D должно быть выше 60%.

Из массива пошагово формируется список моделей, содержащий список наилучших моделей. Алгоритм отбора следующий:

1. В списке находится наибольшее значение коэффициента Somers' D, умножается на 100 для перевода в проценты и округляется в меньшую сторону. Из списка удаляются те значения, чей коэффициент Somers' D меньше округлённого максимального на 2%, чтобы сформировать новый список из моделей с наиболее высокой дискриминационной способностью.

2. Из полученного списка удаляются те модели, чей коэффициент R^2 меньше 0,49 и те модели, где нормированный вес хотя бы одного из факторов составляет менее 5% (но не равен 0). Этот шаг необходим для устранения тех моделей, где в список вошли факторы, имеющие слишком слабое влияние на прогнозную переменную. Проверка R^2 является информационной проверкой, поскольку чаще всего вошедшие в финальный список модели являются достаточно качественными.

3. На финальном списке моделей проводится прогнозирование вероятности дефолта по тестовой выборке. Результаты теста оцениваются по значению коэффициента Somers' D. Из списка исключаются те модели, чей Somers' D меньше, чем максимальный, округлённый в меньшую сторону (рис. 2).

	Factors	SomersD	R2	Coefficients	Weight
0	[F5, F9, tmp, tmp, F27, tmp, F30, tmp]	0.651786	0.512211	[-0.6971019445149939, -0.3659438210141498, -5....	[0.471, 0.247, 0.0, 0.0, 0.151, 0.0, 0.13, 0.0]
1	[F5, F10, tmp, F12, F27, F29, F30, tmp]	0.651568	0.543550	[-0.6774305545560203, -0.21801716411765337, -5....	[0.416, 0.134, 0.0, 0.069, 0.12, 0.165, 0.094, ...
2	[F5, F10, tmp, F12, F27, tmp, F30, tmp]	0.652906	0.512898	[-0.7058673515377406, -0.2830295627000365, -1....	[0.463, 0.186, 0.0, 0.083, 0.143, 0.0, 0.124, ...
3	[F5, F10, tmp, F16, F27, tmp, F30, tmp]	0.651008	0.515015	[-0.6671131472402261, -0.3047605544506669, 0.0....	[0.442, 0.202, 0.0, 0.091, 0.152, 0.0, 0.113, ...

Рисунок 2 – Финальный список моделей, принятых к разработке

Выводы

Данный алгоритм построения модели определения вероятности дефолта методом множественной линейной регрессии позволяет сформировать сразу несколько моделей, удовлетворяющих требованиям к разрабатываемой модели, что обеспечивает гибкость в принятии проектной группой финального решения, поскольку позволяет руководствоваться мнением экспертов при выборе финального варианта, а значит, обеспечить большую прозрачность для бизнес-партнёров.

Применение рассмотренных подходов, включающих использование методики, модели и программно-технических средств реализации существенно упрощает процесс регионального прогнозирования вероятности

дефолта и позволяет улучшить качество управленческих решений в банковской отрасли для успешной реализации запланированной стратегии развития.

Библиографический список

1. Банк России. Письмо от 06.02.2012 № 14-Т «О рекомендациях Базельского комитета по банковскому надзору «Принципы совершенствования корпоративного управления» : Справочно-правовая система «Консультант плюс». –URL: http://www.consultant.ru/document/cons_doc_LAW_126196 (дата обращения: 04.05.2018)
2. Банк России. Письмо от 29.12.2012 No 192-Т «о Методических рекомендациях по реализации подхода к расчету кредитного риска на основе внутренних рейтингов банков» // справочно-правовая система «консультант плюс» URL: http://www.consultant.ru/document/cons_doc_LAW_140871/ (дата обращения: 10.06.2018).
3. Анализ математических моделей Базель II / Алескеров, Ф.Т. [и др.]. М. ФИЗМАТЛИТ, 2015. 286 с.
4. Кабушкин, С.Н. Управление банковским кредитным риском: учебное пособие С.Н. Кабушкин. Минск: Новое знание, 2017. 336 с.
5. Карминский А.М., Пересецкий А.А., Петров А.Е. Рейтинги в экономике: методология и практика: Монография / Под ред. А.М. Карминского. М.: Финансы и статистика, 2015.
6. Карминский, А.М. Моделирование вероятности дефолта российских банков: расширенные возможности / А.М. Карминский, А.В. Костров // Журнал Новой экономической ассоциации. 2013. №1, т. 17. С. 64-86.
7. Катасонов, В.Ю. Проектное финансирование, управление риском, страхование / В.Ю. Катасонов, Д.С. Морозов. М.: Анкил, 2014. 272 с.
8. Моделирование вероятности дефолта корпоративных заемщиков / А.А. Жевага и др. // Управление финансовыми рисками. 2016. №1. т. 45. С. 12-26.
9. Моргунов, А.В. Использование сводных макроэкономических индикаторов для калибровки внутренних рейтинговых моделей в банках / А.А. Жевага, А.В. Моргунов // Деньги и Кредит. 2015. № 8. С. 39-46
10. Пугачев, В.С. Теория вероятностей и математическая статистика: Учеб. Пособие. М.: ФИЗМАТЛИТ, 2002. 496 с.
11. Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. М.: ДМК Пресс, 2017. 418 с.
12. Рогов М.А. Риск-менеджмент. М.: Финансы и статистика, 2015. – 120 с.
13. Симановский А.Ю. Базельские принципы эффективного банковского надзора, издание второе // Деньги и кредит. 2016. № 1. С. 20-30.
14. Бугакова Н.С., Гельвановский М.И., Глисин Ф.Ф., Гохберг Л.М. – Каталог публикаций «Регионы России. Социально-экономические показатели – 2017, статистический сборник». // Федеральная служба государственной

- статистики. – 1995-2018. URL: http://www.gks.ru/free_doc/doc_2017/region/reg-pok17.pdf (дата обращения: 05.01.2018).
15. Данные по формам статистической налоговой отчетности / Федеральная Налоговая Служба России. – 1995-2018. URL: https://www.nalog.ru/rn77/related_activities/statistics_and_analytics/forms/ (дата обращения: 19.04.2018).
16. Единая межведомственная информационно-статистическая система: Официальный сайт / Федеральная служба государственной статистики. – 1995-2018. URL: <https://www.fedstat.ru/> (дата обращения: 20.04.2018).
17. Информационный ресурс Fitch [Электронный ресурс]: Официальный сайт Fitch. URL: www.fitchratings.com (дата обращения: 05.05.2018).
18. Информационный ресурс Moody's // Официальный сайт Moody's. URL: www.moody.com (дата обращения: 05.05.2018).
19. Информационный ресурс S&P [Электронный ресурс]: Официальный сайт S&P. – Режим доступа: www.standardandpoors.com (дата обращения: 05.05.2018).
20. Консолидированные бюджеты субъектов Российской Федерации и бюджетов территориальных государственных внебюджетных фондов / Официальный сайт Казначейства России. – 1995-2018. URL: <http://roskazna.ru/> (дата обращения: 14.04.2018).
21. Моргунов А.В. Методы оценки кредитных рисков инвестиционных проектов. URL: <https://www.hse.ru/sci/diss/201793626> (дата обращения: 08.04.2018).
22. Моргунов А.В. Моделирование вероятности дефолта инвестиционных проектов // Корпоративные финансы. 2016. №1. Т.37. С.30-45. URL: <https://cfjournal.hse.ru/article/view/1424/2589> (дата обращения: 15.06.2018).
23. Объем и структура государственного долга субъектов Российской Федерации и долга муниципальных образований [Электронный ресурс] / Министерство Финансов России. – 1995-2018. URL: https://www.minfin.ru/ru/performance/public_debt/subdbt/ (дата обращения: 18.04.2018).
24. Рейтинги субъектов РФ / Информационное агентство Cbonds URL: <http://cbonds.ru/ratings/> (дата обращения: 25.04.2018).
25. Федеральная служба государственной статистики: Официальный сайт / Федеральная служба государственной статистики. – 1995-2018. URL: <http://www.gks.ru> (дата обращения: 15.04.2018).
26. Тотьмянина К. М. Обзор моделей вероятности дефолта // Сборник Управление финансовыми рисками 25.01.2015, 15 с. Режим доступа: <https://www.hse.ru/data/2011/11/14/1270200082/%D0%BE%D0%B1%D0%B7%D0%BE%D1%80%20%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D0%B5%D0%B9%20%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B8%20%D0%B4%D0%B5%D1%84%D0%BE%D0%BB%D1%82%D0%B0.PDF> (дата обращения: 25.05.2018).

-
27. Basel Committee on Banking Supervision. Basel III: A global regulatory framework for more resilient banks and banking system [Электронный ресурс] / Bank for International Settlements – 2014 – Режим доступа: <http://www.bis.org/publ/bcbs189.pdf> (дата обращения: 10.06.2018).
28. Benninga, S. Financial Modelling. [Text] / S. Benninga. – 3 ed. – The MIT Press, 2008. 1168 p.