

## LASSO, RIDGE И ELASTICNET В МОДЕЛЯХ ФОРМИРОВАНИЯ СПРЕДА ДОХОДНОСТИ РОССИЙСКИХ КОРПОРАТИВНЫХ ОБЛИГАЦИЙ

Карамян Анаит Араратовна

аспирант,

МГУ имени М.В. Ломоносова, экономический факультет

(г. Москва, Россия)

### Аннотация

*Исследование посвящено анализу моделей формирования спреда доходности российских корпоративных облигаций нефинансового сектора, что имеет важное значение для понимания факторов, определяющих разницу в доходности финансовых инструментов. Цель работы заключается в построении моделей формирования спреда методами Lasso, Ridge и ElasticNet и сделать вывод по полученным результатам о применимости данных методов, которые ранее не использовались в отечественной литературе для решения подобных задач. Методология исследования основана на сравнительном анализе двух временных выборок: данных за период с 01.01.2014 г. по 23.02.2022 г. и за 24.02.2022 г. по 31.12.2023 г., что позволяет оценить устойчивость моделей в разных рыночных условиях. Результаты показали, что, несмотря на частичную способность методов отбирать экономически значимые факторы, их применение для российского рынка корпоративных облигаций нефинансового сектора ограничено. Это связано с наличием нелогичных зависимостей и сложностями в интерпретации коэффициентов. В дальнейших исследованиях целесообразно рассмотреть альтернативные подходы, поскольку разработка точной модели позволит участникам рынка эффективнее управлять рисками и оптимизировать доходность в соответствии с индивидуальным риск-аппетитом.*

**Ключевые слова:** Lasso, Ridge, ElasticNet, спред доходности, корпоративные облигации.

**JEL коды:** G12.

**Для цитирования:** Карамян А.А. Lasso, Ridge и ElasticNet в моделях формирования спреда доходности российских корпоративных облигаций // Научные исследования экономического факультета. Электронный журнал. 2025. Том 17. Выпуск 4. С. 99-122. DOI: 10.38050/2078-3809-2025-17-4-99-122.

## Введение

Облигации как альтернатива банковскому кредиту считаются надежным инструментом для инвестирования капитала и важным механизмом для привлечения денежных средств для эмитентов. Понимание особенностей российского облигационного рынка может позволить инвесторам правильно управлять своими рисками. Современные реалии в мировой и отечественной экономике, нестабильность на финансовых рынках, которые провоцируют многообразие факторов, влияющих на оценку риска и доходности инструментов долгового рынка, а также возросший интерес инвесторов и эмитентов обуславливают необходимость изучения особенностей российского долгового рынка.

Тема формирования спредов доходности корпоративных облигаций достаточно хорошо изучена в зарубежной литературе с использованием широкого инструментария анализа данных, в том числе рекуррентных нейронных сетей (Gu et al., 2020). Однако развивающиеся рынки остаются малоизученными. Проблема исследования развивающихся рынков связана с их недостаточной ликвидностью, низкой информационной транспарентностью, небольшими объемами, недостатком данных для полноценного исследования. В связи с этим, сложно получить интерпретируемые результаты и оценки анализа. В отечественной литературе исследования проводились с применением простых методов регрессионного анализа, более сложные методы регрессии не рассматривались (Ерофеева, 2021). Многие зарубежные исследователи в своих работах сравнивают результаты простых и сложных методов и указывают на малую точность результатов простых методов, например, таких, которые используются на данный момент в российской литературе. Это и представляет актуальность данного исследования в применении новых методов в модели факторов формирования спреда на российском рынке облигаций. Так, исследования о формировании доходности корпоративных облигаций развивающихся стран, например, Китая (Guo, Wang, 2019), позволяют обратить внимание на новые значимые факторы, благодаря применению более сложных методов.

Слабые, но значимые взаимосвязи факторов представляют собой статистически достоверные, хотя и не выраженные по величине, зависимости между предикторами и целевой переменной, которые при совокупном рассмотрении вносят существенный вклад в предсказательную способность модели. В контексте моделирования спредов доходности на развивающихся рынках, включая российский, такие взаимосвязи часто остаются незамеченными при использовании традиционных методов вследствие их невысокой объясняющей способности в отдельности и сильного влияния рыночных шумов. Регуляризационные методы (Lasso, Ridge, ElasticNet) позволяют выявлять и учитывать эти зависимости за счет сжатия коэффициентов и отбора информативных предикторов, что особенно важно в условиях низкой ликвидности и сложной структуры взаимосвязей на долговом рынке. Регуляризованные методы – это методы, которые добавляют штрафные члены к функции потерь для уменьшения переобучения и улучшения обобщающей способности модели, что особенно полезно в финансах для работы с мультиколлинеарностью и отбора значимых признаков. Однако их применение ограничено в случаях, когда интерпретируемость коэффициентов критически важна, а также при наличии нелинейных зависимостей, требующих более сложных методов.

Большинство российских исследователей применяли статистические методы анализа, однако за последние годы методы машинного обучения начали приобретать все большую популярность за счет своей эффективности относительно традиционных статистических методов. Так, в зарубежной литературе, в том числе китайской, методы машинного обучения уже

обладают широким применением. Многие зарубежные исследователи сравнивают Lasso, Ridge и ElasticNet с другими методами, такими как дерево решений, PCA (Principal component analysis), LSTM (Long Short-Term Memory) и др. При этом результаты у исследователей получались разные, на некоторых рынках какие-то из методов Lasso, Ridge и ElasticNet показывали действительно высокий уровень точности, однако на других ни один из методов не демонстрировал положительных результатов. В связи с этим основная цель данного исследования – построить модели формирования спреда доходности для российского рынка облигаций методами Lasso, Ridge и ElasticNet и сделать вывод по полученным результатам о применимости данных методов.

Таким образом, в условиях повышенной волатильности и структурных изменений на финансовых рынках исследование формирования спредов доходности рублевых корпоративных облигаций нефинансового сектора приобретает особую актуальность. В отличие от существующих отечественных работ, опирающихся на простые регрессионные методы, данное исследование предлагает применение более сложных методов (Lasso, Ridge, ElasticNet), получивших широкое распространение в зарубежной практике, но пока не применяемых в российской литературе. Это позволит не только выявить слабые, но значимые взаимосвязи факторов, но и повысить точность моделирования спреда с учетом специфики развивающегося рынка, характеризующегося низкой ликвидностью и ограниченной транспарентностью. Полученные результаты могут быть полезны как профессиональным участникам рынка (финансовым аналитикам, управляющим активами) для оценки рисков и принятия инвестиционных решений, так и эмитентам, использующим облигации в качестве альтернативы банковскому кредитованию.

### **1.1. Анализ применения методов Lasso, Ridge, ElasticNet в моделях формирования спредов доходности на зарубежных рынках**

С. Анатольев и А. Микущева в своей статье рассматривают проблему оценки премии за риск в моделях линейного ценообразования активов (Anatolyev, Mikusheva, 2021). Авторы рассматривают довольно большую выборку. Особое внимание уделялось факторам, обладающим слабой силой в моделях. Исследование слабых факторов обусловлено тем, что их влияние может быть неизменным с течением времени, а значит, данные факторы могут учитываться как перманентные. Авторы предлагают метод оценки премии за риск с учетом слабых факторов. Также они выводят асимптотическое распределение предложенной оценки по множеству слабых факторов, вычисляют стандартные ошибки и проверяют эффективность метода при моделировании с помощью различных инструментов машинного обучения, в том числе с помощью методов Lasso и Ridge. Однако большую часть внимания авторы уделяют рынку акций, чем облигаций.

С. Гиглио, Д. Ксию и Д. Жанг тестируют модели, основанные на SPCA (Supervised principal component analysis), PCA, Ridge, Lasso и на методе частичных наименьших квадратах (PLS, Partial least squares) для исследования слабых факторов в модели регрессии ценообразования активов (Giglio, Xiu, Zhang, 2021). В результате показывается, что PCA (и некоторые другие его вариации), Ridge и PLS несовместимы при наличии слабых факторов, а подход Lasso совместим с оценкой SDF (Stochastic Discount Factor), а также с оценкой премии за риск. Последние модели продемонстрировали положительные результаты.

С. Гу, Б. Келли и Д. Ксию проводили сравнительный анализ методов машинного обучения для решения канонической задачи эмпирического ценообразования активов - измерения премий за риск активов (Gu, Kelly, Xiu, 2020). Авторы в процессе эмпирического анализа изучают несколько методов, в том числе Lasso, Ridge, ElasticNet и иные методы машинного обучения. Все методы основаны на одном и том же наборе доминирующих прогностических факторов. По результатам исследования авторам удалось получить интерпретируемые результаты по методам Lasso и ElasticNet.

Т. Бали и его соавторы проводили сравнительный анализ методов машинного обучения для прогнозирования доходности корпоративных облигаций США (Bali et al., 2020). Прогнозируемые факторы включают в себя большой набор из 43 характеристик облигаций, таких как размер выпуска, кредитный рейтинг, срок до погашения и проч. Выводы, которые удалось сделать авторам с помощью эмпирической оценки, состоят в том, что если модель состоит из множества коррелирующих между собой факторов, линейная модель без регуляризации работает хуже, так как применение штрафных санкций и уменьшение весов факторов существенно улучшают производительность линейной модели. Авторы также отмечают низкую точность простых регрессий при решении подобных задач.

Наряду с другими исследователями авторы (He et al., 2021) отмечают, что методы машинного обучения действительно открывают новые перспективы для анализа рынка корпоративных облигаций относительно простых методов. Эмпирические результаты свидетельствуют, что линейные модели демонстрируют статистически значимую прогнозную способность, тогда как нелинейные методы обеспечивают более устойчивую и превосходящую производительность на различных подвыборках облигаций и временных интервалах. Ключевое преимущество регуляризованных линейных регрессий (Lasso, Ridge, ElasticNet) заключается в их способности сохранять исходные предикторы без преобразований – в отличие от методов снижения размерности, таких как PCA и PLS, – что позволяет сохранить экономическую интерпретируемость модели. Однако сравнительный анализ выявляет относительную нестабильность результатов линейных методов: в частности, для краткосрочных облигаций наблюдаются значительные колебания прогнозной точности во времени, а в условиях финансового кризиса 2008 г. модели Lasso демонстрируют снижение эффективности для высоконадежных облигаций (рейтинг AAA), выдавая статистически незначимые или отрицательные коэффициенты. Тем не менее, в стандартных рыночных условиях Elastic Net и Lasso сохраняют свои конкурентные преимущества, уступая по устойчивости лишь наиболее совершенным нелинейным методам (Random Forest), которые лучше адаптируются к структурным изменениям и гетерогенности данных.

Методы Lasso, Ridge и ElasticNet активно применяются в зарубежных исследованиях различных научных направлений, демонстрируя свою эффективность в задачах отбора признаков и регуляризации. Однако в российской практике их использование для анализа факторов спреда практически не встречается, несмотря на их потенциал для построения более точных моделей. Зарубежный опыт применения этих методов неоднороден: в одних рыночных условиях они показывают высокую точность, в других – незначительно превосходят простые подходы. В связи с этим представляет научный и практический интерес проверить, насколько данные методы применимы к российским рыночным данным и способны ли они улучшить понимание механизмов формирования спреда российских корпоративных рублевых облигаций нефинансового сектора.

## 1.2. Описание выборки

Для исследования зависимости спреда доходности от ряда детерминант используются ежедневные данные о доходности российских рублевых корпоративных облигаций, торгуемых на Московской бирже. В целях данного исследования будут рассмотрены два периода: с 01.01.2014 г. по 23.02.2022 г. и с 24.02.2022 г. по 31.12.2023 г. Выборка делится по дате размещения облигации, и сделано это для демонстрации специфических факторов, характерных для российского долгового рынка капитала после начала СВО (специальной военной операции) в 2022 г.

В выборку не включены облигации финансового сектора (банков, лизинговых и страховых компаний), поскольку их бизнес-модели и структура капитала существенно отличаются от нефинансовых корпораций. Финансовые институты подвержены уникальным регуляторным рискам и особым требованиям к капиталу, что может исказить результаты исследования, направленного на анализ фундаментальных факторов кредитного риска. Их включение потребовало бы отдельной методологии учета системных рисков и нормативных особенностей, что выходит за рамки данного исследования.

Облигации с плавающей купонной ставкой были исключены из-за принципиальных различий в механизмах ценообразования. Поскольку их доходность привязана к макроэкономическим показателям (ключевой ставке, инфляции и т. д.), это создает дополнительную волатильность, не связанную с кредитными характеристиками эмитента. Такие инструменты требуют особых подходов к оценке, а их включение осложнило бы выделение «чистого» эффекта исследуемых фундаментальных факторов на спреды доходности.

Эмиссионные выпуски, по которым отсутствует открытая финансовая отчетность, не рассматривались в исследовании, так как это лишает возможности анализировать ключевые показатели финансового состояния компаний. Без доступа к балансам, отчетам о прибылях и убытках невозможно достоверно оценить кредитное качество эмитентов и рассчитать необходимые факторные переменные, что ставит под сомнение надежность любых выводов по таким облигациям.

Также из выборки исключены облигации с неполными данными по исследуемым факторам, поскольку это нарушает принцип сопоставимости наблюдений. Отсутствие исторических данных по кредитным рейтингам, финансовым коэффициентам или другим ключевым переменным делает невозможным их включение в единую модель.

Итого, в выборке присутствуют следующие типы облигаций: облигации с амортизацией, биржевые, коммерческие, зеленые, отзывные облигации, для квалифицированных инвесторов. Широкая выборка позволяет охватить практически весь рынок корпоративных облигаций нефинансового сектора.

## 1.3. Компоненты модели формирования спреда

Зависимой переменной в целях настоящего исследования является G-спред. Он рассчитывается индивидуально для каждой облигации на основе данных Cbonds, что обеспечивает учет специфических характеристик каждого выпуска<sup>1</sup>.

---

<sup>1</sup> Облигационный калькулятор Cbonds (Cbonds.ru): URL: <https://cbonds.ru/calc/> (дата обращения: 01.04.2024).

Формула G-спред для выпусков в рублях одинакова для всех выпусков и представляет собой разницу между текущей доходностью облигации и значением на G-кривой по облигациям федерального займа с такой же дюрацией. Дюрация Маколея – это средневзвешенное время до возврата инвестиции в облигации, учитывающее размер купонных выплат, срок до погашения и доходность по облигации. Этот показатель помогает инвесторам оценить чувствительность цены облигации к изменениям процентных ставок. Спред характеризует меру риска, который инвесторы готовы принять на себя при покупке конкретной облигации. Положительный спред означает, что доходность облигации выше, чем доходность ОФЗ. Обычно это свидетельствует о том, что рынок считает данную облигацию менее надежной и более рискованной, поэтому требуется дополнительное вознаграждение за инвестирование в нее. Отрицательный спред доходности указывает на то, что рынок считает данную облигацию более надежной и менее рискованной, и поэтому инвесторы готовы получить меньшую доходность за ее приобретение.

Отбор независимых переменных осуществлялся на основании их теоретической значимости, подтвержденной работами других исследователей в области долгового финансирования, а также с учетом специфики российского рынка корпоративных облигаций. Т.М. Ерофеева исследовала выборку на промежутке времени с 2010 г. по 2019 г., которая включает в себя период времени до 2014 г. и после (Ерофеева, 2019). Отдельное внимание уделяется влиянию наличию кредитного рейтинга у эмитента. Автор утверждает, что исследования, проведенные на выборке с низкой долей присвоенных рейтингов, являются менее качественными. Итоговая модель формирования спреда доходности корпоративных облигаций включает в себя такие факторы, как курс золота в долларах, темп роста ВВП, рейтинг, срок до погашения, инфляция, размер компании и объем эмиссии. В данной модели приводится много макроэкономических факторов.

В другом ее исследовании, касающемся функциональной оценки взаимосвязи между спредом доходности и спредом дефолта, доказывалось наличие параметрической связи между этими показателями (Ерофеева, 2020). Автором делается предположение, что остальные факторы, кроме риска дефолта, уже включены в расчет спреда. Поэтому модель состоит из одной переменной, значимость которой и доказывается.

Д.В. Жидкова и Д.С. Шароватов исследовали изменения, которые произошли на российском долговом рынке с учетом кризиса 2022 г. (Жидкова, Шароватов, 2023). Так, авторы уделяют внимание макроэкономическим показателям, таким как динамика ключевой ставки Банка России, инфляции, ИПЦ. Однако авторы не проводят эмпирического анализа, статья носит информационный характер с приведением статистических данных по рынку без эконометрического или математического моделирования.

Особенность исследования Е.В. Королевой, М.Л. Копейкина и С. Поповой заключается в оценке влияния макроэкономических факторов на изменение доходности корпоративных облигаций 20 развитых стран (Королева, Копейкин, Попова, 2023). В выборку попали облигации за временной промежуток с 2016 г. по 2020 г. В целях регрессионного анализа были сформированы две модели: одна модель включала изменение цен на нефть и остальные факторы, кроме изменения цен на золото, другая модель включала изменение цен на золото и остальные факторы, кроме изменения цен на нефть. Данное деление связано с единичной корреляцией между этими факторами. По результатам регрессионного анализа двух моделей, выяснилось, что вторая модель уступает первой в части описания изменения доходности, что свидетельствует о существенной значимости изменения цен на нефть на долговом рынке 20 развитых

стран. При этом остальные макроэкономические индикаторы (инфляция, ВВП, обменный курс) являются незначимыми.

Дж. Кабаллеро и А. Фернандез проводили аналогичное исследование, однако для 18 развивающихся стран (в том числе и для России) на большей выборке данных (с 2000 г. по 2016 г.) (Caballero, Fernandez, 2019). Авторы отмечали наличие связей между долговым и фондовым рынками. Работа выделяется на фоне остальных применением в модели цен на пшеницу, цен на сталь, так как большинство исследуемых стран являются одними из основных мировых экспортеров зерна/металлов.

Б. Келли, Д. Палхарес и С. Пруитт предлагают условную факторную модель доходности корпоративных облигаций США с пятью факторами и изменяющейся во времени факторной нагрузкой (Kelly, Palhares, Pruitt, 2021). Авторы сделали три основных вывода по результатам исследования. Во-первых, факторная модель описывает риски и доходность корпоративных облигаций, значительно превосходя ранее предложенные в литературе модели на рынке облигаций США. Так, в модель входят такие факторы, как кредитный рейтинг, темп роста ВВП, срок до погашения, ключевая ставка ФРС, объем эмиссии. Во-вторых, модель рекомендует систематический портфель инвестиций в облигации с высоким коэффициентом Шарпа, что свидетельствует о том, что в данной модели высокая премия за кредитный риск. В-третьих, выводится более тесная интеграция между долговым и фондовым рынками.

П. Дхони, Х. Сри и М. Ахмад проводили исследование на рынке облигаций, зарегистрированных на бирже Индонезии (Dhony, Sri, Ahmad, 2019). Статья описывает факторную зависимость спредов облигаций в строительной отрасли как одной из самых капитализируемых и ключевых отраслей в стране. Так, в качестве характеристики облигаций используется срок их погашения, который положительно влияет на доходность облигаций. Макроэкономическими значимыми переменными являются инфляция и ставка по безрисковым активам, которые оказывают негативное влияние на доходность облигаций. Цена на золото в качестве замещающей инвестиции в облигации положительно влияет на их доходность. Обменный курс и мировые цены на нефть не влияют на доходность облигаций.

Б. Гуо, С. Ванг исследовали макроэкономические факторы, определяющие структуру доходности казначейских облигаций Китая по срокам, кредитные спреды по корпоративным китайским облигациям и спреды ликвидности по ним без арбитража (Guo, Wang, 2019). Рассматриваются четыре экономических фактора – денежно-кредитные условия, инфляция, реальный объем производства и волатильность финансового рынка. В докризисный период шоки волатильности снижали доходность казначейских облигаций и увеличивали как кредитные спреды, так и спреды ликвидности по всем рейтинговым классам, а кредитные спреды увеличивались по мере ужесточения денежно-кредитных условий, но влияние инфляции и реального объема производства было незначительным. В периоды стресса волатильность финансового рынка оказывает аналогичное воздействие, и влияние инфляции и реального объема производства также становится значительным. Показано, что игнорирование компонента ликвидности в спредах доходности компаний приводит к неточной оценке влияния экономических факторов на корпоративные кредитные спреды. Доказательства поведения «бегства к ликвидности», которое усиливается в трудные времена, проливают свет на отрицательную корреляцию между безрисковой ставкой и спредами корпоративной доходности, а также на положительную корреляцию между кредитными спредами и спредами ликвидности.

С. Чанг, М. Фанг особое внимание уделяли анализу макроэкономических факторов кредитных спредов китайских корпоративных облигаций нефинансовых предприятий с фиксированной процентной ставкой (Chang, Fang, 2020). Так, индекс потребительских цен, ставка по казначейским облигациям и темп роста ВВП оказывают значительное влияние на формирование спреда китайских облигаций нефинансового сектора.

Еще одно исследование на развивающемся рынке китайских облигаций проводили Д. Хуанг, К. Чанг и У. Тиан (Huang, Chang, Tian, 2019). В статье основное внимание также уделяется макроэкономическим факторам, влияющим на спред доходности корпоративных облигаций. Они обнаружили, что цены на промышленные товары (ИПЦ) и индекс цен на корпоративные товары (CGPI) являются значимыми в модели. С ростом цен на промышленную продукцию растут производственные издержки, что приводит к повышению риска дефолта, поэтому инвесторы требуют более высокую премию за риск. Обменный курс играет важную роль в моделировании спреда. Когда китайская валюта укрепляется по отношению к доллару США, экспорт увеличивается, а уровень корпоративного риска снижается, что приводит к снижению спреда доходности корпоративных облигаций. Однако влияние обменного курса юаня ниже, чем у CGPI. Авторы дополнительно в качестве гипотезы рассмотрели влияние индекса деловой активности и обнаружили, что она отрицательно коррелирует со спредом.

В рамках настоящего исследования был сформирован набор из 28 различных объясняющих переменных, который представлен в табл. 1 в Приложении. Особенностью данных переменных является наличие мультиколлинеарности (VIF, variance inflation factor) между ними. Именно эта структурная особенность данных требует применения специализированных аналитических подходов, позволяющих корректно идентифицировать вклад отдельных факторов при наличии устойчивых корреляционных зависимостей между ними. Выбор в пользу конкретных 28 переменных обусловлен частотой их анализа в российской и зарубежной литературе, а также с учетом наличия данных по ним в открытом доступе.

Помимо общепринятых факторов, таких как темп роста ВВП, валютный курс, ключевая ставка, депозитные и кредитные ставки, в модель также включается фактор (дамми-переменная) вхождения эмитента в SDN-лист. SDN-лист представляет собой список из людей и организаций, на которых действуют ограничения (санкции) со стороны Министерства финансов США. Санкции предусматривают визовые ограничения и блокирование собственности и счетов физических и юридических лиц. Всего около 72 эмитента из рассматриваемых в данном исследовании находятся в SDN-листе по состоянию на конец 1 квартала 2024 г.

Присутствие в эмиссионных документах облигации информации о наличии поручителей или гарантов дает инвесторам дополнительный знак надежности облигации, так как если финансовое состояние самого эмитента ухудшится, обязательство возьмет на себя другая компания.

Кредитный рейтинг эмитента – независимая комплексная оценка кредитного качества эмитента, присвоенная рейтинговым агентством. Основанный на всесторонней оценке финансового состояния компании, результатов ее деятельности, структуры капитала, учитывающий все риски бизнеса эмитента рейтинг позволяет более обоснованно судить об уровне надежности эмитента и рисках обеспечения его обязательств. По мере развития отечественного долгового рынка и рейтинговой индустрии в России в последние годы стало возможным формировать достаточную по количеству выборку эмитентов, имеющих рейтинг. Как правило, высокий



рейтинг позволяет компаниям предлагать более низкую процентную ставку, так как им не нужно компенсировать кредитный риск. Именно поэтому спред по облигациям с низким рейтингом будет значительно шире. В целях данного исследования будут использованы рейтинги отечественных кредитных рейтинговых агентств. Показатели рейтингов переводятся в числовые значения по шкале от 1 до 10, например, облигации (или эмитенту облигации) с самым высоким рейтингом будет присвоено «10».

Вхождение в группу компаний подразумевает поддержку со стороны материнской структуры, которая может проявляться через покупку высокорисковых (но низкодоходных) облигаций дочерней компании, предоставлении беспроцентного займа для погашения облигаций и прочих видов финансовой поддержки. В период кризиса многие компании оказываются в ситуации финансовой нестабильности, показатели финансовой отчетности ухудшаются, что подразумевает актуальность поддержки со стороны группы.

Факт дефолта в кредитной истории эмитента служит важным сигналом для инвесторов, существенно повышающим воспринимаемый уровень кредитного риска. Дефолт является фактическим нарушением обязательств по выплатам (например, пропущенным купонным платежом по облигациям). Дефолт может быть временным техническим нарушением, которое впоследствии может устраниться, в отличие от банкротства.

Банкротство представляет собой юридически признанную несостоятельность субъекта, при которой он неспособен в полном объеме исполнять свои финансовые обязательства, что влечет за собой судебную процедуру реструктуризации долга или ликвидации активов под контролем арбитражного управления. Банкротство означает окончательную утрату платежеспособности и влечет за собой системные последствия для всех заинтересованных сторон.

Коэффициент долг/активы отражает степень защищенности обязательств компании ее активами, демонстрируя потенциальную способность предприятия обслуживать долговые обязательства даже в неблагоприятных рыночных условиях. Суть коэффициента звучит следующим образом: «Сколько рублей активов требуется эмитенту для покрытия 1 рубля долга?».

Учет величины долга компании при моделировании спреда доходности является необходимым, поскольку уровень заемных средств непосредственно влияет на кредитный риск эмитента. Компании с высокой уровнем долговых средств демонстрируют более широкие спреды из-за повышенной вероятности дефолта, что отражает требуемую инвесторами премию за риск. Логарифмирование данной величины позволит избежать высокого разброса данных.

Эмпирические исследования свидетельствуют о наличии устойчивой обратной зависимости между объемом выручки и величиной спреда, что объясняется снижением кредитного риска у компаний с более значительными и стабильными доходами. Особую значимость данный фактор приобретает в условиях макроэкономической нестабильности, когда инвесторы проявляют повышенную чувствительность к фундаментальным показателям финансового состояния эмитентов. Величина выручки также будет логарифмирована в моделях.

Объем эмиссии отражает совокупную стоимость выпущенных компанией долговых инструментов. Более крупные эмиссии, как правило, характеризуются повышенной ликвидностью за счет большего числа обращающихся бумаг и широкого круга инвесторов, что способствует сужению спреда доходности. Однако чрезмерный объем заимствований может сигнали-

зировать о повышенной долговой нагрузке, особенно если рост эмиссии не подкрепляется соответствующим увеличением денежных потоков компании. Объем эмиссии будет логарифмирован в моделях.

Многими исследованиями доказана прямая зависимость риска дефолта от срока, оставшегося до погашения облигации из-за высокой неопределенности относительно будущего эмитента. Соответственно, инвесторы требуют более высокую доходность по более длинным облигационным выпускам.

Большинство российских облигаций выпускаются с наличием встроенного опциона *put* (для инвестора) или опциона *call* (для эмитента), что подразумевает, что инвестор или эмитент имеют право погасить облигацию по номиналу в день оферты. С другой стороны, после дня оферты эмитент имеет право изменить купонные выплаты, и если эти выплаты оказываются ниже текущего уровня, то скорее всего большинство инвесторов реализуют опцион. Это накладывает дополнительные риски, которые и являются одной из составляющих спреда доходности. На развитых рынках, где лучше защищены права инвесторов, более распространены облигации со встроенными опционами *call*, а на рынках развивающихся стран, напротив, – облигации со встроенными опционами *put*. Если в условиях выпуска облигации предусмотрен опцион, то доходность считается не к дате погашения, а к дате оферты. В целях настоящего исследования опционы разделены отдельно на *put* и *call*, которые являются дамми-переменными.

## **2.1. Описание методов Lasso, Ridge, ElasticNet**

Методы Lasso, Ridge и ElasticNet являются разновидностями линейной регрессии, специально адаптированными для данных, которые демонстрируют сильную мультиколлинеарность. Они автоматизируют части выбора модели, такие как выбор переменных или исключение параметров.

В этих методах заложен процесс сжатия коэффициентов, когда значения данных приближаются к центральной точке (например, к среднему значению). Процесс сжатия добавляет регрессионным методам несколько преимуществ: более точные и стабильные оценки истинных параметров и сглаживание пространственных флуктуаций. В этом процессе они автоматически помогают устранить или исказить сильно коррелированные и избыточные функции в методе с низкой дисперсией. В процессе работы данных методов заложена регуляризация, которая представляет собой «штрафование» переменных через увеличение или уменьшение их весов ( $\alpha$  на языке Python). Слишком большое значение  $\alpha$  может привести к недообученной модели, когда все регрессоры становятся незначимыми или слабо значимыми. Слишком малое значение  $\alpha$  может спровоцировать появление переобученной модели, в которой будет присутствовать большое количество регрессоров с большими весами. Выражения «большая  $\alpha$ » и «малая  $\alpha$ » являются относительными, так как для каждой модели будут индивидуальные пороговые значения.

Lasso использует регуляризацию L1, которая взвешивает ошибки по их абсолютному значению. Применение только такого типа регуляризации способно обнулять регрессоры с большими весами. Lasso не всегда может быть информативна в практических задачах, так как данный метод может обнулять значимые с точки зрения логики факторы. В Ridge используется

регуляризация L2, которая более «строгая», так как учитывает квадрат весов переменных, вместо модуля по сравнению с регуляризацией L1 и не обнуляет переменные с большими весами, т. е. чем больше вес независимой переменной, тем ближе ее значение будет к нулю. В Ridge веса факторов размываются, что затрудняет интерпретацию данной метода. ElasticNet совмещает применение обоих методов регуляризации в одной модели. В таблице 2 приведены основные характеристики методов.

Таблица 2

### Характеристики методов

Метод	Моделирование	Регуляризация
Метод Lasso-регрессии	<ol style="list-style-type: none"> <li>1. Моделирование на пространственных данных</li> <li>2. Помогает бороться с переобученными моделями</li> <li>3. Используется регуляризация L1</li> <li>4. Большой «штраф» приводит к недообученной модели, а небольшой – к переобученной</li> </ol>	$  y - X\omega  _2^2 + \lambda   \omega  _1 \rightarrow \min,$ <p>где</p> $  y - X\omega  _2^2 - \text{MSE (loss-функция)},$ $\lambda   \omega  _1 - \text{регуляризация L1},$ $  \omega  _1 = \sum_{i=1}^n  \omega_i $
Метод Ridge-регрессии	<ol style="list-style-type: none"> <li>1. Моделирование на пространственных данных</li> <li>2. Помогает бороться с переобученными моделями</li> <li>3. Используется регуляризация L2</li> <li>4. Большой «штраф» приводит к недообученной модели, а небольшой – к переобученной</li> </ol>	$  y - X\omega  _2^2 + \lambda   \omega  _2^2 \rightarrow \min,$ <p>где</p> $  y - X\omega  _2^2 - \text{MSE (loss-функция)},$ $\lambda   \omega  _2^2 - \text{регуляризация L2},$ $  \omega  _2^2 = \sum_{i=1}^n \omega_i^2$
Метод ElasticNet	<ol style="list-style-type: none"> <li>1. Моделирование на пространственных данных</li> <li>2. Помогает бороться с переобученными моделями</li> <li>3. Модель совмещает в себе регуляризации L1 и L2</li> </ol>	$  y - X\omega  _2^2 + \lambda   \omega  _1 + \lambda   \omega  _2^2 \rightarrow \min,$ <p>где</p> $  y - X\omega  _2^2 - \text{MSE (loss-функция)},$ $\lambda   \omega  _1 - \text{регуляризация L1},$ $\lambda   \omega  _2^2 - \text{регуляризация L2}$

Источник: составлено автором.

С учетом набора из 28 регрессоров, которые имеют высокую корреляционную связь между собой, применение регуляризации становится необходимым. В целях настоящего исследования моделирование с применением вышеописанных методов производится автором на языке Python в Jupyter Notebook<sup>2</sup>. Применение именно данного языка программирования обусловлено тем, что в Excel и в других подобных программных обеспечениях нет подходящего функционала. Модели Lasso, Ridge и ElasticNet можно построить только с помощью языков программирования. Язык Python выделяется своей простотой. Так, моделирование происходит с помощью классов LassoCV (для метода Lasso), RidgeCV (для метода Ridge) и ElasticNetCV (для метода ElasticNet), которые находятся в составе модуля sklearn. linear\_model библиотеки

<sup>2</sup> Код см. в Приложении.

Scikit-learn (sklearn) языка программирования Python. Важной составляющей является кросс-валидация (Cross-validation, CV), которая производится с помощью класса RepeatedKFold модуля sklearn.model\_selection библиотеки Scikit-learn. Кросс-валидация необходима для объективной оценки обобщающей способности моделей, минимизации переобучения и обеспечения устойчивости результатов на новых данных путем последовательного тестирования на различных подмножествах исходной выборки. Библиотека Numpy будет использоваться для создания массивов подбора наилучшего коэффициента регуляризации. Библиотека Pandas помогает работать с таблицами. Jupyter Notebook представляет собой интерактивный блокнот с кодом и текстом, есть также его онлайн версия – Google Colab (Google Colaboratory), которая представляет собой облачную платформу от Google, где можно писать код на языке Python прямо из браузера.

## **2.2. Результаты применения методов Lasso, Ridge, ElasticNet**

В результате проведенного исследования было получено шесть моделей с учетом временного разделения выборки. Все коэффициенты округлены до сотых долей. Модели, построенные на подвыборке за период с 01.01.2014 г. по 23.02.2022 г., маркированы индексом «\_1», а модели на данных с 24.02.2022 г. по 31.12.2023 г. – индексом «\_2». В таблице 3 в Приложении представлены результаты по этим моделям.

Анализ коэффициентов различных моделей регрессии с регуляризацией позволяет сделать ряд существенных выводов о значимости рассматриваемых факторов. Применение моделей с L1-регуляризацией (столбцы model\_Lasso\_coefs\_1 и model\_Lasso\_coefs\_2) позволило выявить как преимущества метода Lasso, так и его содержательные ограничения. Ключевым достоинством выступила способность модели к автоматическому отбору признаков, что проявилось в обнулении коэффициентов для менее значимых переменных и позволило выделить устойчивый набор факторов, влияющих на спред доходности. В первой модели наибольший вклад вносят три ключевых показателя: срок до погашения ценной бумаги с положительным влиянием  $-0,51$ , что указывает на обратную зависимость, объем эмиссии с положительным влиянием  $-0,18$ , риск дефолта с отрицательным влиянием  $0,15$  и наличие опциона на продажу с коэффициентом  $0,20$ . Вторая модель демонстрирует несколько иную структуру значимых переменных. Здесь наибольший вес имеют риск дефолта ( $0,27$ ), опцион на продажу ( $0,30$ ) и кредитный рейтинг эмитента ( $0,10$ ). Опираясь на полученные результаты по данному методу, можно сделать вывод о том, что после 24.02.2022 г. модель формирования спреда лишь частично модифицировалась, появились рейтинг эмитента и величина долга с отрицательным влиянием, стали незначимыми величина выручки и срок до погашения, по остальным факторам веса немного изменились без изменения знаков. Появление рейтинга с положительным знаком является довольно нелогичным, потому что при увеличении рейтинга эмитента или облигации, кредитные риски наоборот снижаются, что и приводит к малым премиям за риск, т. е. к уменьшению спреда, а не к его увеличению. Данная ситуация демонстрирует принципиальное ограничение Lasso, когда метод, будучи эффективным инструментом сокращения размерности, не гарантирует содержательной корректности отобранных переменных. Это может быть связано как с наличием скрытых зависимостей в данных (например, нелинейных эффектов или взаимодействий между переменными), так и с изменением структурных параметров рынка в период кризиса, когда традиционные соотношения между показателями временно нарушаются (например, когда у эмитентов с высоким кредитным рейтингом выпускаются облигации с вы-

сокой доходностью). Таким образом, несмотря на ценность Lasso для первичного отбора признаков, интерпретация его результатов требует дополнительной экономической верификации, особенно в условиях структурных шоков, когда стандартные теоретические ожидания могут терять силу.

При анализе моделей с L2-регуляризацией (столбцы `model_Ridge_coefs_1` и `model_Ridge_coefs_2`) наблюдается принципиально иная картина. В отличие от предыдущего случая данный метод не обнуляет коэффициенты, а лишь уменьшает их абсолютные значения, сохраняя все переменные в модели. В первой модели выделяются несколько особенно значимых факторов: отношение долга к активам с сильным отрицательным коэффициентом  $-3,57$ , что подчеркивает важность финансовой устойчивости эмитента, риск дефолта (1,55), цена на пшеницу (1,82) и доходность государственных облигаций (1,12). Вторая модель демонстрирует более сложную структуру влияния факторов. Наибольшие по абсолютной величине коэффициенты наблюдаются у доходности государственных облигаций (2,56), процентной ставки по кредитам ( $-2,51$ ) и ИПЦ (1,20). Примечательно появление значимого положительного влияния кредитных ставок, что может отражать чувствительность зависимой переменной к условиям заимствования в экономике. Сравнительный анализ двух исследуемых периодов позволяет выявить существенные изменения в направлении влияния ключевых факторов на формирование спреда доходности после начала специальной военной операции. Как показывают результаты моделирования, следующие параметры изменили характер воздействия с положительного на отрицательный: темп роста ВВП, обменный курс, ключевая ставка, цена на нефть, уровень государственного участия, торговый баланс, кредитный рейтинг эмитента, принадлежность к корпоративной группе и наличие встроенного опциона *put*. Данная трансформация коэффициентов, аналогично результатам, полученным при применении метода Lasso, не находит убедительного экономического обоснования и противоречит теоретическим ожиданиям. В условиях высокой мультиколлинеарности, характерной для финансовых данных, Ridge распределяет влияние между коррелированными переменными, искусственно сохраняя их в модели, но при этом искажая экономический смысл оценок коэффициентов, что особенно проблематично при анализе структурных сдвигов, когда истинные зависимости между факторами претерпевают существенные изменения. Кроме того, отсутствие обнуления коэффициентов делает динамику их изменений во времени экономически необоснованной, как продемонстрировано в исследовании, что противоречит теоретическим ожиданиям и снижает аналитическую ценность модели.

Особый интерес представляет модель, сочетающая свойства обоих типов регуляризации (столбец `model_ElasticNet_coefs_1`, столбец `model_ElasticNet_coefs_2`). Результаты показывают, что в данном случае первая модель ведет себя аналогично методу с L1-регуляризацией (благодаря кросс-валидации), обнуляя большинство коэффициентов. Сохранились лишь несколько значимых фактора: ИПЦ (0,01), риск дефолта (0,47), банкротство (0,18), объем эмиссии ( $-0,18$ ), размер компании ( $-0,03$ ), срок до погашения ( $-0,54$ ) и объем заемных средств (0,01). Во второй модели наибольшее влияние на зависимую переменную оказывает срок до погашения ценной бумаги с коэффициентом  $-0,47$ . Вторым по значимости фактором выступает наличие встроенного опциона на продажу с отрицательным влиянием 0,33. Положительное влияние оказывают такие параметры, как размер компании ( $-0,09$ ), объем эмиссии ( $-0,12$ ) и кредитный рейтинг эмитента ( $-0,11$ ). Оценивая изменения в моделях после СВО, можно выделить обнуление влияния ИПЦ, что нелогично, так как именно после СВО наблюдается высокий

уровень инфляции в России, с которым борется Банк России путем установления высокой ключевой ставки (Годовой отчет..., 2024). Также обнулились риск дефолта и банкротство, хотя после 24.02.2022 г. наблюдались на рынке частые дефолты под влиянием ввода санкций (Сбои трансграничных платежей..., 2024). Отсутствие ИПЦ и ожидаемой реакции на учащение дефолтов указывают на системное ограничение ElasticNet в условиях структурных шоков – метод, эффективно балансирующий между отбором признаков и устойчивостью к мультиколлинеарности в стабильных условиях, может давать содержательно проблематичные результаты при резких изменениях рыночных взаимосвязей. Таким образом, несмотря на теоретические преимущества гибридного подхода, практическое применение ElasticNet в финансовых исследованиях требует дополнительных проверок экономической состоятельности отобранных переменных, особенно при анализе кризисных периодов.

Проведенный анализ шести регрессионных моделей с различными типами регуляризации позволяет сделать вывод об их неприменимости и лишь частичном соответствии экономической теории. Арифметический знак во всех моделях для срока до погашения, демонстрирующего стабильно положительное влияние, не соответствует логике, так как при увеличении срока до погашения увеличиваются риски, которые в свою очередь отражаются на увеличении спреда. Знак при риске дефолта во всех моделях соответствует теоретическим ожиданиям, так как наличие дефолта по выплате купонов по облигациям увеличивает спред. Наличие встроенного опциона *put*, рост величины долга, ИПЦ также увеличивают спред доходности, как и показывается во всех моделях. Проблему интерпретации вызывает поведение кредитного рейтинга в Lasso-моделях, где наблюдается либо его исключение из модели, либо парадоксальный положительный знак коэффициента, противоречащий базовой аксиоме кредитного риска. Также присутствует неоднозначная оценка влияния объема эмиссии. Ridge-регрессии демонстрируют более сбалансированный подход, сохраняя значимость как микроэкономических параметров эмитентов, так и макроэкономических индикаторов, хотя тоже содержат некоторые противоречия в знаках коэффициентов для отдельных переменных. Модели ElasticNet занимают промежуточное положение, сочетая черты обоих методов, но проявляют высокую чувствительность к спецификации, что выражается в существенных различиях между двумя вариантами модели. Присутствуют также разногласия по поводу влияния рейтинга, причем во второй модели ElasticNet знак при рейтинге указан правильно с точки зрения логики, хотя и по некоторым другим переменным результат не поддается интерпретации.

Полученные результаты указывают на существенные методологические ограничения существующих подходов и необходимость разработки новых методик моделирования кредитных спредов, способных последовательно интегрировать микро- и макроэкономические детерминанты без внутренних логических противоречий.

## Заключение

Таким образом, в данном исследовании были построены 6 моделей с применением методов Lasso, Ridge и ElasticNet для периодов с 01.01.2014 г. по 23.02.2022 г. и с 24.02.2022 г. по 31.12.2023 г.. Присутствие некоторых различий в части интерпретации и нелогичность результатов позволяют сделать вывод о том, что данные методы не применимы для российского рынка корпоративных облигаций нефинансового сектора, они способны лишь частично подбирать экономически целесообразные параметры. В связи с этим в качестве направлений для будущих исследований предполагается дальнейший поиск метода, наиболее подходящего для

российского рынка облигаций, так как построение и применение точной модели позволит всем участникам облигационного рынка эффективно управлять своими рисками, получая требуемую доходность с учетом риск-аппетита каждого участника.

## Список литературы

Ерофеева Т.М. Исследование факторов и построение модели прогнозирования спреда доходности корпоративных облигаций на российском рынке // Финансы и бизнес. 2019. № 4. Т. 15. С. 81–10.

Ерофеева Т.М. Оценка функциональной взаимосвязи между спредом доходности и спредом дефолта // Экономический журнал ВШЭ. 2020. № 1. С. 28–52.

Ерофеева Т.М. Прогнозирование спреда доходности на российском долговом рынке // Вест. Моск. ун-та. Сер. 6. Экономика. 2021. № 6. С. 1–6.

Жидкова Д.В., Шароватов Д.С. Анализ рынка и оценка инвестиционной привлекательности отечественных корпоративных облигаций // Инновации и инвестиции. 2023. № 1. С. 3–10.

Королева Е.В., Копейкин М.Л., Попова С. Оценка влияния макросреды на доходность корпоративных облигаций в развитых странах // Вестник Академии знаний. 2023. № 2 (55). С. 14–19.

Anatolyev S., Mikusheva A. Factor models with many assets: strong factors, weak factors, and the two-pass procedure // Journal of Econometrics. 2021. No. 229 (1). P. 40–51.

Bali T.G., Goyal A., Huang D., Jiang F., Wen Q. Predicting Corporate Bond Returns: Merton Meets Machine Learning // Social Science Research Network (SSRN). 2020. No. 20–110. P. 82–86.

Caballero J., Fernandez A. On corporate borrowing, credit spreads and economic activity in emerging economies: an empirical investigation // Journal of International Economics. 2019. No. 118. P. 160–178.

Chang C.L., Fang M. Macro-Factors of Credit Spreads in Corporate Bonds // Social Science Research Network (SSRN). 2020. No. 1. P. 163–171.

Dhony P., Sri H., Ahmad M.T.N. Effect of Bond Characteristics and Macroeconomic Factors on the Return of Corporate Bonds in the Sector of Property, Real Estate and Building Construction // RJOAS. 2019. No. 3 (87). P. 100–105.

Giglio S., Dacheng X., Zhang D. Test Assets and Weak Factors // Social Science Research Network (SSRN). 2021. No. 2021–79. P. 68–72.

Gu S., Kelly B., Xiu D. Empirical asset pricing via machine learning // The Review of Financial Studies. 2020. No. 33 (5). P. 2223–2273.

Guo B., Wang S. Macro Factors in Corporate Bond Credit and Liquidity Spreads // Social Science Research Network (SSRN). 2019. No. 1. P. 20–50.

He X., Feng G., Wang J., Wu C. Predicting Individual Corporate Bond Returns // Journal of Banking and Finance. 2024. No. 171. P. 71–83.

Huang J., Chang K., Tian Y. Analysis of Macroeconomic Factors Affecting the Corporate Bond Yield Spread // Engineering Letters. 2019. Vol. 27. No. 1. P. 120–130.

Kelly B., Palhares D., Pruitt S. Modeling Corporate Bond Returns // Journal of Finance. 2021. No. 78. P. 12–35.

Годовой отчет Банка России (Банк России): URL: [https://cbr.ru/about\\_br/publ/god/](https://cbr.ru/about_br/publ/god/) (дата обращения: 01.04.2024).

Официальный сайт Федеральной службы государственной статистики (Росстат): URL: <https://rosstat.gov.ru/> (дата обращения: 01.04.2024).

Облигационный калькулятор Cbonds (Cbonds.ru): URL: <https://cbonds.ru/calc/> (дата обращения: 01.04.2024).

Поиск облигаций и Карты рынка (Cbonds.ru): URL: <https://cbonds.ru/bonds>. (обновлено 01.03.2024; процитировано 01.04.2024).

Сбои трансграничных платежей назвали причиной дефолтов на долговом рынке (РБК): URL: <https://www.rbc.ru/quote/news/article/66cc67ee9a7947dbbc0c318f?from=copy> (дата обращения 01.09.2024).



## Приложение

Таблица 1

### Описание переменных

Группа показателей	№	Показатель/ наименование в модели	Описание	VIF	Источник	Влияние на спред
Макроэкономические и монетарные факторы	1	Темп роста ВВП/ «GDP»	Логарифм роста ВВП за квартал, предшествующий размещению	3042	Росстат	Обратное (положительное)
	2	Валютный курс/ «RUB/USD»	На дату размещения	3994	Банк России	Прямое (отрицательное)
	3	Ключевая ставка/ «Key rate»		3747		
	4	Депозитная ставка/ «Deposit rate»		4832		
	5	Кредитная ставка/ «Credit rate»		4044		
	6	Индекс первичного жилья/ «Real estate»	Месяц, предшествующий дате размещения	4912	Cbonds	
	7	Индекс потребительских цен (ИПЦ)/ «CPI»		2487	Банк России	
Рыночные и товарные цены	8	Цена на нефть/ «Oil»	На дату размещения	3090	Cbonds	
	9	Цена на золото/ «Gold»		2789		
	10	Цена на сталь/ «Steel»		3532		
	11	Цена на пшеницу/ «Wheat»		2168		
Государственные и внешнеэкономические факторы	12	Участие государства в капитале компании/ «Government»	Дамми-переменная: 1, если >50% + 1 голос; 0.5, если 50%; 0.25, если 25% + 1 голос; 0, если <25%	1834	Спарк-Интерфакс	Обратное (положительное)
	13	Торговый баланс/ «Trade_balance»	Значение за квартал, предшествующего размещению	1733	Банк России	
	14	Наличие эмитента в SDN-листе/ «SDN»	Дамми-переменная: 0 – эмитент не входит в санкционный список, 1 – эмитент входит в санкционный список	823	Спарк-Интерфакс	Прямое (отрицательное)
	15	Ставка по облигациям федерального займа (ОФЗ)/ «Rus treasures»	На дату размещения	3481	Банк России	
Кредитные и юридические риски	16	Наличие поручительств или гарантий/ «Guarantees»	Дамми-переменная: 0 – не имеются по-	730	Cbonds	Обратное (положительное)

Группа показателей	№	Показатель/ наименование в модели	Описание	VIF	Источник	Влияние на спред
			ручительства/гарантии, 1 – имеются поручительства/гарантии		Спарк-Интерфакс	Прямое (отрицательное)
	17	Рейтинг эмитента/эмиссии/ «Rating»	Числовое значение в зависимости от рейтинга (от 1 до 10)	1834		
	18	Вхождение в группу компаний/ «Group»	0 – не входит в группу компаний, 1 – входит в группу компаний	1011		
	19	Риск дефолта/ «Default»	Дамми-переменная: 0 – нет дефолтов, 1 – есть.	1504		
	20	Банкротство/ «Bankruptcy»	Дамми-переменная: 0 – нет банкротства, 1 – есть.	1703		
Финансовые показатели эмитентов	21	Долг/Активы/ «Debt/Asset»	Если из квартальной отчетности, то квартал, предшествующий дате размещения, если годовая отчетность, то год, предшествующий дате размещения	1307	Спарк-Интерфакс	Обратное (положительное)
	22	Величина долга/ «Debt»	Логарифм величины (если из квартальной отчетности, то квартал, предшествующий дате размещения, если годовая отчетность, то год, предшествующий дате размещения)	1476		
	23	Величина выручки/ «Sales»		1534		
	24	Размер компании/ «Assets»		1789		
	25	Объем эмиссии/ «Emissions»		1555		
Технические параметры облигации	26	Срок до погашения/ «Maturity»	В годах	1894	Cbonds	Прямое (отрицательное)
	27	Встроенный опцион call/ «Call»	0 – нет встроенного опциона call; 1 – есть встроенный опцион call	1636	Cbonds	
	28	Встроенный опцион put/ «Put»	0 – нет встроенного опциона put; 1 – есть встроенный опцион put	1862	Cbonds	

Источник: составлено автором.

Пошаговый код на языке Python для построения моделей:

1. Импортируются необходимые библиотеки и модули:

```
import pandas as pd
import sklearn
import numpy as np
from sklearn.model_selection import RepeatedKFold
from sklearn.linear_model import LassoCV
from sklearn.linear_model import RidgeCV
from sklearn.linear_model import ElasticNetCV
```

2. Загрузка Excel-файлов с данными по 28 независимым переменным и зависимой переменной в формате 2 таблиц по периодам с 01.01.2014 по 23.02.2022 и с 24.02.2022 по 31.12.2023 (DataFrame, df). df1 – таблица с облигациями за период с 01.01.2014 по 23.02.2022, X1 – набор факторов для этой таблицы, Y1 – спреды облигаций по этому периоду. df2 – таблица с облигациями за период с 24.02.2022 по 31.12.2023, X2 – набор факторов для этой таблицы, Y2 – спреды облигаций по этому периоду.

```
df1 = pd.read_excel («...»)
```

```
df2 = pd.read_excel («...»)
```

```
X1 = df1[['GDP', 'RUB/USD', 'Key_rate', 'Deposit_rate', 'Credit_rate', 'Real_estate',
'CPI', 'Oil', 'Gold', 'Steel', 'Wheat', 'Government', 'Trade_balance', 'SDN', 'Rus_treasures',
'Guarantees', 'Rating', 'Group', 'Default', 'Bankruptcy', 'Debt/Asset', 'Debt', 'Sales', 'Emissions',
'Maturity', 'Call', 'Put']]
```

```
Y1 = df1[['G-spread']]
```

```
X2 = df2[['GDP', 'RUB/USD', 'Key_rate', 'Deposit_rate', 'Credit_rate', 'Real_estate',
'CPI', 'Oil', 'Gold', 'Steel', 'Wheat', 'Government', 'Trade_balance', 'SDN', 'Rus_treasures',
'Guarantees', 'Rating', 'Group', 'Default', 'Bankruptcy', 'Debt/Asset', 'Debt', 'Sales', 'Emissions',
'Maturity', 'Call', 'Put']]
```

```
Y2 = df2[['G-spread']]
```

3. Следующим шагом приводятся параметры кросс-валидации, можно применить часто используемые: n\_splits = 10 разбивает данные на 10 частей или по-другому folds (по сути каждая 10-я часть по очереди становится тестовой выборкой, а остальные – обучающей), n\_repeats = 3 повторяет процесс 3 раза для повышения надежности результатов, перемешивая данные с фиксированным random\_state = 1 (для воспроизводимости). Итого получается 30 проверок (10 folds \* 3 повтора), что снижает зависимость случайного разбиения. Такая кросс-валидация признана устойчивой для моделей с нестабильными данными. Код кросс-валидации следующий:

```
cv = RepeatedKFold(n_splits = 10, n_repeats = 3, random_state = 1)
```

4. Далее строим модель Lasso с автоматическим подбором силы регуляризации (alphas): для каждого  $\alpha$  проводится кросс-валидация (по стратегии cv) и выбирается оптимальное  $\alpha$ , дающее наилучший результат. Для этого значение  $\alpha$  заключается в массив значений, где в скобках сперва указывается начальное значение, которые мы хотим рассмотреть, затем – конечное (оно не включается в подбор) и третье значение – шаг между значениями. То есть для приведенного ниже кода будут тестироваться  $\alpha$  из следующего массива: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. Затем с помощью функции fit библиотеки sklearn происходит обучение модели по заданным X и Y. Код выглядит следующим образом:

```
model = LassoCV(alphas= np.arange (0.1, 1, 0.1), cv = cv)
```

```
model1_Lasso = model.fit(X1, Y1)
```

```
model2_Lasso = model.fit(X2, Y2)
```

5. Последним шагом выводятся коэффициенты в модели через атрибут coef\_:

```
model1_Lasso_coefs = model_lasso.coef_1
```

```
model2_Lasso_coefs = model_lasso.coef_2
```

6. Модель ElasticNet строится аналогичным образом, как Lasso, код представлен ниже:

```
model = ElasticNetCV(alphas= np.arange (0.1, 1, 0.1), cv = cv)
```

```
model1_ElasticNet = model.fit(X1, Y1)
```

```
model2_ElasticNet = model.fit(X2, Y2)
```

```
model1_ElasticNet_coefs = model_ElasticNet.coef_1
```

```
model1_ElasticNet_coefs = model_ElasticNet.coef_2
```

7. В модели RidgeCV принципиальное значение имеет выбор метрики оценки (scoring), определяющей критерий оптимизации коэффициента регуляризации  $\alpha$  в процессе кросс-валидации. Использование метрики 'neg\_mean\_absolute\_error' обусловлено особенностью реализации scikit-learn, где процедура подбора параметров всегда ориентирована на максимизацию целевого показателя: отрицательное значение средней абсолютной ошибки (MAE) позволяет достичь минимизации фактической ошибки прогноза. Алгоритм последовательно тестирует значения  $\alpha$  из заданного диапазона, на каждом шаге осуществляя обучение на тренировочных подвыборках и валидацию на тестовых, окончательно выбирая параметр  $\alpha$ , обеспечивающий минимальное значение MAE. В отличие от принятого по умолчанию коэффициента детерминации ( $R^2$ ), MAE демонстрирует повышенную робастность к наличию выбросов в данных, поскольку использует абсолютные отклонения вместо квадратичных, что особенно значимо при работе с зашумленными или аномальными наблюдениями. Код для построения Ridge выглядит следующим образом:

```
model = RidgeCV(alphas = np.arange(0.1, 1, 0.1), cv = cv, scoring = 'neg_mean_absolute_error')
```

```
model1_Ridge = model.fit(X1, Y1)
```

```
model2_Ridge = model.fit(X2, Y2)
```

model1\_Ridge\_coefs = model\_Ridge.coef\_1

model2\_Ridge\_coefs = model\_Ridge.coef\_2

Таблица 3

Результаты применения методов Lasso, Ridge и ElasticNet

	model_ Lasso_coefs_1	model_ Lasso_coefs_2	model_ Ridge_coefs_1	model_ Ridge_coefs_2	model_ ElasticNet_ coefs_1	model_ ElasticNet_ coefs_2
Темп роста ВВП	0,00	0,00	-1,00	0,04	0,00	0,03
Валютный курс	0,00	0,00	-0,41	0,79	0,00	0,00
Ключевая ставка	0,00	0,00	-1,01	1,07	0,00	0,00
Депозитная ставка	0,00	0,00	1,26	0,61	0,00	0,00
Кредитная ставка	0,00	0,00	-0,30	-2,51	0,00	0,00
Индекс первичного жилья	0,00	0,00	-0,25	-0,07	0,00	0,00
ИПЦ	0,01	0,02	0,00	1,20	0,01	0,00
Цена на нефть	0,00	0,00	-0,88	0,80	0,00	0,00
Цена на золото	0,00	0,00	0,72	0,70	0,00	0,00
Цена на сталь	0,00	0,00	0,00	0,28	0,00	0,00
Цена на пшеницу	0,00	0,00	1,82	-0,29	0,00	0,00
Участие государства	0,00	0,00	0,19	-0,06	0,00	0,00
Торговый баланс	0,00	0,00	-0,09	0,46	0,00	0,00
Наличие эмитента в SDN-листе	0,00	0,00	0,00	0,16	0,00	0,00
ОФЗ	0,00	0,00	1,12	2,56	0,00	0,00
Поручительства/гарантии	0,00	0,00	-0,18	-0,05	0,00	0,00
Рейтинг эмитента	0,00	0,10	0,21	-0,65	0,00	-0,11

	<b>model_ Lasso_coefs_1</b>	<b>model_ Lasso_coefs_2</b>	<b>model_ Ridge_coefs_1</b>	<b>model_ Ridge_coefs_2</b>	<b>model_ ElasticNet_ coefs_1</b>	<b>model_ ElasticNet_ coefs_2</b>
Вхождение в группу компаний	0,00	0,00	<b>0,14</b>	<b>-0,21</b>	0,00	0,00
Риск де- фолта	<b>0,15</b>	<b>0,27</b>	<b>1,55</b>	0,00	<b>0,47</b>	0,00
Банкрот- ство	0,00	0,00	<b>0,79</b>	0,00	<b>0,18</b>	0,00
Долг/Ак- тивы	0,00	0,00	<b>-3,57</b>	<b>-0,15</b>	0,00	0,00
Величина долга	0,00	<b>0,01</b>	<b>0,25</b>	<b>0,06</b>	<b>0,01</b>	0,00
Величина выручки	<b>0,01</b>	0,00	0,00	<b>-0,01</b>	0,00	<b>-0,01</b>
Размер компании	<b>-0,02</b>	<b>-0,01</b>	<b>-0,25</b>	<b>-0,18</b>	<b>-0,03</b>	<b>-0,09</b>
Объем эмиссии	<b>-0,18</b>	<b>-0,10</b>	<b>-0,18</b>	<b>-0,06</b>	<b>-0,18</b>	<b>-0,12</b>
Срок до погашения	<b>-0,51</b>	0,00	<b>-0,68</b>	<b>-0,83</b>	<b>-0,54</b>	<b>-0,47</b>
Встроен- ный оп- цион call	0,00	0,00	<b>1,19</b>	<b>-0,16</b>	0,00	0,00
Встроен- ный оп- цион put	<b>0,20</b>	<b>0,30</b>	<b>1,06</b>	<b>0,75</b>	0,00	<b>0,33</b>

Источник: составлено автором.

## LASSO, RIDGE AND ELASTICNET IN MODELS OF THE FORMATION OF THE YIELD SPREAD OF RUSSIAN CORPORATE BONDS

**Anait A. Karamian**

*Postgraduate student,*

*Lomonosov Moscow State University, Faculty of Economics*

*(Moscow, Russia)*

### Abstract

*The article is devoted to the analysis of models for the formation of the yield spread of Russian corporate bonds in the non-financial sector, which is important for understanding the factors determining the difference in the profitability of financial instruments. The purpose of the work is to build models of spread formation using Lasso, Ridge and ElasticNet methods and to draw a conclusion based on the results obtained about the applicability of these methods, which have not previously been used in the Russian literature to solve such problems. The research methodology is based on a comparative analysis of two samples: data for the period from 01.01.2014 to 23.02.2022 and for 24.02.2022 to 31.12.2023, which makes it possible to assess the stability of models in different market conditions. The results showed that, despite the partial ability of the methods to select economically significant factors, their application for the Russian corporate bond market in the non-financial sector is limited. This is due to the presence of illogical dependencies and difficulties in interpreting the coefficients. In further research, it is advisable to consider alternative approaches, since the development of an accurate model will allow market participants to manage risks more effectively and optimize returns in accordance with individual risk appetite.*

**Keywords:** Lasso, Ridge, ElasticNet, spread yield, corporate bonds.

**JEL:** G12.

**For citation:** Karamian, A.A. (2025) Lasso, Ridge and ElasticNet in Models of the Formation of the Yield Spread of Russian Corporate Bonds. Scientific Research of Faculty of Economics. Electronic Journal, vol. 17, no. 4, pp. 99-122. DOI: 10.38050/2078-3809-2025-17-4-99-122.

### References

Erofeeva T.M. Issledovanie faktorov i postroenie modeli prognozirovaniya spreda dokhodnosti korporativnykh obligatsiy na rossiyskom rynke. *Finansy i biznes*. 2019. № 4. T. 15. S. 81–10. (In Russ.).

Erofeeva T.M. Otsenka funktsional'noy vzaimosvyazi mezhdu spredom dokhodnosti i spredom defolta. *Ekonomicheskiiy zhurnal VShE*. 2020. № 1. S. 28–52. (In Russ.).

Erofeeva T.M. Prognozirovanie spreda dokhodnosti na rossiyskom dolgovom rynke. Vest. Mosk. un-ta. Ser. 6. Ekonomika. 2021. № 6. S. 1–6. (In Russ.).

Zhidkova D.V., Sharovатов D.S. Analiz rynka i otsenka investitsionnoy privlekatel'nosti otechestvennykh korporativnykh obligatsiy. Innovatsii i investitsii. 2023. № 1. S. 3–10. (In Russ.).

Koroleva E.V., Kopeykin M.L., Popova S. Otsenka vliyaniya makrosredy na dokhodnost' korporativnykh obligatsiy v razvitykh stranakh. Vestnik Akademii znaniy. 2023. № 2 (55). S. 14–19. (In Russ.).

Anatolyev S., Mikusheva A. Factor models with many assets: strong factors, weak factors, and the two-pass procedure. Journal of Econometrics. 2021. No. 229 (1). P. 40–51.

Bali T.G., Goyal A., Huang D., Jiang F., Wen Q. Predicting Corporate Bond Returns: Merton Meets Machine Learning. Social Science Research Network (SSRN). 2020. No. 20–110. P. 82–86.

Caballero J., Fernandez A. On corporate borrowing, credit spreads and economic activity in emerging economies: an empirical investigation. Journal of International Economics. 2019. No. 118. P. 160–178.

Chang C.L., Fang M. Macro-Factors of Credit Spreads in Corporate Bonds. Social Science Research Network (SSRN). 2020. No. 1. P. 163–171.

Dhony P., Sri H., Ahmad M.T.N. Effect of Bond Characteristics and Macroeconomic Factors on the Return of Corporate Bonds in the Sector of Property, Real Estate and Building Construction. RJOAS. 2019. No. 3 (87). P. 100–105.

Giglio S., Dacheng X., Zhang D. Test Assets and Weak Factors. Social Science Research Network (SSRN). 2021. No. 2021–79. P. 68–72.

Gu S., Kelly B., Xiu D. Empirical asset pricing via machine learning. The Review of Financial Studies. 2020. No. 33 (5). P. 2223–2273.

Guo B., Wang S. Macro Factors in Corporate Bond Credit and Liquidity Spreads. Social Science Research Network (SSRN). 2019. No. 1. P. 20–50.

He X., Feng G., Wang J., Wu C. Predicting Individual Corporate Bond Returns. Journal of Banking and Finance. 2024. No. 171. P. 71–83.

Huang J., Chang K., Tian Y. Analysis of Macroeconomic Factors Affecting the Corporate Bond Yield Spread. Engineering Letters. 2019. Vol. 27. No. 1. P. 120–130.

Kelly B., Palhares D., Pruitt S. Modeling Corporate Bond Returns. Journal of Finance. 2021. No. 78. P. 12–35.

Godovoy otchet Banka Rossii (Bank Rossii): Available at: [https://cbr.ru/about\\_br/publ/god/](https://cbr.ru/about_br/publ/god/) (accessed: 01.04.2024). (In Russ.).

Ofitsial'nyy sayt Federal'noy sluzhby gosudarstvennoy statistiki (Rosstat): Available at: <https://rosstat.gov.ru/> (accessed: 01.04.2024). (In Russ.).

Obligatsionnyy kal'kulyator Cbonds (Cbonds.ru): Available at: <https://cbonds.ru/calc/> (accessed: 01.04.2024). (In Russ.).

Poisk obligatsiy i Karty rynka (Cbonds.ru): Available at: <https://cbonds.ru/bonds>. (accessed: 01.09.2024). (In Russ.).

Sboi transgranichnykh platezhey nazvali prichinoy defoltov na dolgovom rynke (RBK): Available at: <https://www.rbc.ru/quote/news/article/66cc67ee9a7947dbbc0c318f?from=copy> (accessed: 01.09.2024). (In Russ.).