**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ** **ВЫСШЕГО ОБРАЗОВАНИЯ**

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

**Факультет экономических наук**

Зайцева Екатерина Сергеевна

***«Выявление детерминант доходности Европейского рынка высокодоходных облигаций с использованием машинного обучения»***

Выпускная квалификационная работа – МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

по направлению подготовки 38.04.08 Финансы и кредит

Образовательная программа «Финансовые рынки и финансовые институты»

Руководитель

доцент, к.э.н.

Столяров Андрей Иванович

Соруководитель

Тлехугов Николай Владимирович

Москва 2023

**Содержание**

[**ВВЕДЕНИЕ** 3](#_Toc134253632)

[**ГЛАВА 1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.** 6](#_Toc134253633)

[1.1. Причины и процесс формирования рынка высокодоходных облигаций в Европе. 6](#_Toc134253634)

[1.2. Характеристика и структура Европейского рынка ВДО. 11](#_Toc134253635)

[1.3. Критерии отнесения облигации к высокодоходной. 18](#_Toc134253636)

[1.4. Обзор исследуемой литературы. 20](#_Toc134253637)

[1.4.1. Эмпирические исследования эффективности финансовых ценных бумаг с использованием числовых и текстовых данных. 26](#_Toc134253638)

[**ГЛАВА 2. МЕТОДОЛОГИЯ.** 32](#_Toc134253639)

[2.1. Описание переменных 32](#_Toc134253640)

[2.2. Описание выборки 41](#_Toc134253641)

[2.3. Сбор данных 42](#_Toc134253642)

[2.4. Описательные данные о выпусках облигаций 44](#_Toc134253643)

[2.5. Корреляционный анализ детерминант 48](#_Toc134253644)

[**ГЛАВА 3. РЕЗУЛЬТАТЫ ВЫЯВЛЕНИЯ ДЕТЕРМИНАНТ ДОХОДНОСТИ ЕВРОПЕЙСКОГО РЫНКА ВЫСОКОДОХОДНЫХ ОБЛИГАЦИЙ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ** 51](#_Toc134253645)

[3.1. Подготовка к настройке модели 52](#_Toc134253646)

[3.2. Настройка и выбор наиболее эффективной модели 54](#_Toc134253647)

[3.3. Прогнозирование купонной ставки ВДО 58](#_Toc134253648)

[3.4. Обсуждение результатов прогноза купонной ставки ВДО с использованием ML 59](#_Toc134253649)

[3.5. Значение исследования для научных кругов и участников рынка ВДО 61](#_Toc134253650)

[3.6. Допущения и ограничения модели 64](#_Toc134253651)

[3.7. Предложения для дальнейших исследований 65](#_Toc134253652)

[**ЗАКЛЮЧЕНИЕ** 68](#_Toc134253653)

[**СПИСОК ЛИТЕРАТУРЫ** 71](#_Toc134253654)

[**ПРИЛОЖЕНИЯ** 74](#_Toc134253655)

[Приложение 1. – Стрим в Orange. 74](#_Toc134253656)

[Приложение 2. – Назначение неопределенных переменных в Orange. 74](#_Toc134253657)

[Приложение 3. – Виджет Rank в Orange. 75](#_Toc134253658)

[Приложение 4. – Перечень моделей, используемых в Orange. 75](#_Toc134253659)

[Приложение 5. – Выгрузка модели для прогноза Orange. 76](#_Toc134253660)

[Приложение 6. – Схема для предсказания купонной ставки в Orange. 76](#_Toc134253661)

[Приложение 7. – Результат предсказания купонной ставки в Orange. 77](#_Toc134253662)

# **ВВЕДЕНИЕ**

Существование уже сформированного, достаточно зрелого рынка высокодоходных облигаций, то есть корпоративных облигаций с рейтингом ниже инвестиционного уровня, представляется положительным фактором для экономического развития любого государства или региона. Это позволяет корпорациям привлекать средства быстрее и на более гибких условиях, чем это было бы возможно в противном случае. Существование этого направления корпоративного финансирования, дополняющего банковское и акционерное финансирование, особенно выгодно для стран с большим количество мелких и средних фирм.

Рынок корпоративных облигаций, деноминированных в евро, который все еще отстает от своего американского аналога как по абсолютной величине, так и по отношению к размеру экономики, за последние два десятилетия претерпел существенные изменения, которые включают не только рост его объема в обращении, но и также существенные изменения в его разложении. Индексы спредов корпоративных облигаций, которые традиционно используются для отражения динамики рынка посредством единой статистики, могут дать вводящую в заблуждение картину. Имеющиеся в настоящее время индексы для корпоративных облигаций, деноминированных в евро, различают только разные рейтинги выпусков, таким образом, они маскируют неоднородную динамику спредов корпоративных облигаций как в отдельных странах, так и в разных отраслевых сегментах рынка. В то же время ни один страновой или отраслевой индекс корпоративных облигаций еврозоны не учитывает существенных изменений в декомпозиции рейтинга, произошедших за последнее десятилетие. Кроме того, большинство индексов корпоративных облигаций, номинированных в евро, включают неевропейские облигации, номинированные в евро, а также облигации со встроенными специальными характеристиками; они также смешивают связи с различными погашение и купоны.

Выпуск облигаций с высоким кредитным риском становится все более важным источником финансирования для компаний, ведущих свою деятельность в Европе, после того, как законодательные ограничения были наложены на банки после большого финансового кризиса 2008 года. Как и в случае с акциями, эти облигации торгуются публично, и аналитики пытаются предсказать, как они будут функционировать, чтобы сделать прибыльные инвестиции. Но в отличие от акций, рынок высокодоходных корпоративных облигаций все еще относительно молодой и нишевый рынок, доступный только для институциональных инвесторов. Это означает, что качество и количество доступных данных, а также ресурсов для их обработки меньше. Без полной информации рынки капитала не могут оптимально распределять ресурсы, что делает их прозрачными и свободными.

Как правило, аналитики проводят тщательный финансовый анализ рассматриваемой компании, используя данные, полученные из финансовых отчетов и отчетов о прибылях и убытках. Но с недавним развитием в области обработки естественного языка и машинного обучения разрабатываются новые способы извлечения, анализа и прогнозирования информации, связанной с финансовыми ценными бумагами, что делает эту область интересной для исследований.

В своей дипломной работе я попытаюсь распространить анализ числовых данных для прогнозирования купонной ставки по высокодоходной облигации на основе заданных параметров.

**Цель работы:** проанализировать влияние различных детерминант на доходность ВДО, построить качественные модели, которые благодаря машинному обучению смогут с большой точностью предсказать купонную ставку по высокодоходной облигации на основе заданных параметров.

**Предмет работы**: взаимосвязь между характеристиками эмитента, финансовыми показателями компании и купонной ставкой высокодоходной облигации Европейского рынка при размещении.

**Объект работы**: купонная ставка по высокодоходной облигации

**В соответствии с поставленной целью были определены следующие задачи:**

1. Определить источники для написания работы.
2. Изучить исследования различных ученых, посвящённые детерминантам доходности европейского рынка высокодоходных облигаций.
3. Проанализировать уже имеющиеся модели оценки доходности и купонной ставки высокодоходных облигаций эмитентов.
4. Выявить детерминанты, влияющие на купонную ставку ВДО Европы.
5. Построить модели, которые позволят оценить влияние выбранных характеристик на купонную ставку различных выпусков ВДО европейского рынка, а также помогут с большой точностью предсказать купонную ставку.
6. Подвести выводы по работе.

**Методы исследования**:

* описательный;
* аналитический;
* сравнительный;
* математический.

Практическая значимость данного исследования заключается в том, что результаты могут быть полезными для инвесторов во время проведения анализа и обозначения перспектив инвестирования в те или иные ВДО.

# **ГЛАВА 1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.**

# **Причины и процесс формирования рынка высокодоходных облигаций в Европе.**

Европейский рынок высокодоходных облигаций определяется как рынок облигаций, выпущенных компаниями в европейской стране с кредитным рейтингом BBB- или ниже (S&P) или Baa- или ниже (Moody's) (Fridson, 2018).

Европейский высокодоходный рынок по-прежнему является относительно молодым и быстрорастущим рынком по сравнению с другими европейскими финансовыми рынками, такими как рынок облигаций инвестиционного уровня или фондовый рынок.

Первая волна высокой доходности (что имеется в виду?) в Европе началась в конце 1990-х годов, когда европейские инвесторы попытались повторить успех финансирования развития телекоммуникационного и медиа-сектора в США путем выпуска высокодоходных облигаций (Stone Harbour, 2015). В 1998 году рынок составлял всего 4,9 миллиарда евро на основе 35 выпусков (Stone Harbour, 2015). Эта фаза европейского высокодоходного рынка закончилась неудачно вскоре после начала века, поскольку в телекоммуникации стали чрезмерно инвестироваться заемными средствами, а многие проекты не соответствовали требованиям по ежегодному купону (Stone Harbour, 2015).

В начале 1990-х годов некоторые авторы утверждали, что развитие рынка высокодоходных облигаций в Европе, вероятно, произойдет после введения в 1993 году Единого рынка финансовых услуг (Molyneux, 1990). Однако, несмотря на все сопутствующие факторы, европейский рынок высокодоходных облигаций, начал расти действительно существенно в середине 1997 к началу третьего этапа EMU. Такое развитие событий, по-видимому, связано с ожиданиями действительно интегрированного финансового рынка, которые, в свою очередь, были вызваны введением евро (MarquÈs et al., 2000, ECB, 2000 и 2001).

Вторая волна роста (?европейских выпусков высокодоходных облигаций пришлась на середину 2000-х годов. Этот раунд роста также имитировал рынок США и был вызван мощной волной выкупа заемных средств (также известных как инвестиции в частные инвестиции или LBO), когда были выпущены большие суммы долга, чтобы сделать зарегистрированные на бирже компании частными с высоким балансом заемных средств. Европейский рынок высокодоходных облигаций вырос в период с марта 2003 г. по март 2007 г. с 53 млрд евро до 84 млрд евро (Stone Harbour, 2015).

После кратковременного замедления на пике финансового кризиса в 2008 году рост европейского рынка высокодоходных облигаций резко возрос. Финансовые проблемы европейских банков и новая нормативно-правовая база, которая требовала от банков уделять больше внимания подверженности рискам и устанавливать новые требования к капиталу, ограничивали кредитный аппетит банков и, следовательно, способность компаний финансировать свои инвестиции через банки. Это побудило многие компании обратиться к рынкам капитала за новым финансированием и рефинансированием, что вызвало резкий рост числа новых выпусков европейских облигаций с высокой доходностью (Stone Harbour, 2015). Еще одним источником роста европейского высокодоходного рынка в посткризисный период стало появление падших ангелов. Выпуски облигаций, первоначально выпущенные с инвестиционным рейтингом, были понижены до уровня высокодоходных. С 2009 года примерно 27% роста европейского высокодоходного рынка приходится на долю падших ангелов (Stone Harbor, 2015). К 2015 году европейский рынок высокодоходных облигаций увеличился более чем в четыре раза и достиг 387 млрд евро (AFME, 2020). С тех пор рост продолжился, хотя и более медленными темпами. По состоянию на 31.12.2019 европейский рынок высокодоходных облигаций оценивался в 508 млрд евро в обращении. На рисунке 1 показано развитие европейского рынка высокодоходных облигаций.

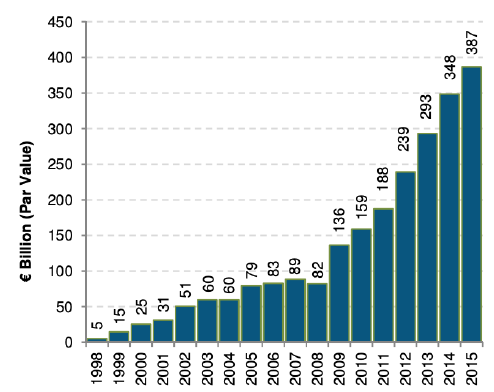


Рисунок 1. – Развитие Европейского рынка ВДО.

Источник: составлено автором на основе данных Development of the European high yield market, AFME, 2020.

Европейский рынок корпоративных облигаций постепенно стал альтернативой традиционным банковским кредитам (Aussenegg et al., 2015). За период с января 2012 года по первый квартал 2016 года объем банковских кредитов уменьшился на 471 млрд евро. За тот же период европейские нефинансовые компании выпустили корпоративные облигации чистой стоимостью 344 млрд евро (BlackRock, 2016). Таким образом, корпоративные облигации сыграли ключевую роль в финансировании корпоративного долга в Европе.

Дальнейший сильный рост наблюдался в 2015 году, когда европейский рынок высокодоходных облигаций достиг 500 млрд евро. Европейский рынок, занимающий менее одной шестой части рынка США, соответствовал примерно 30% рынка США в 2015 году.

Европейский рынок ВДО стал больше, глубже и ликвиднее после мирового финансового кризиса, увеличившись с примерно 110 миллиардов долларов США в 2008 году до более чем 508 евро сегодня — совокупный годовой темп роста составляет около 16% (по сравнению с менее чем 5% для США с высокой доходностью). Всплеск финансовой активности, связанный с Covid-19, увеличил непогашенную задолженность европейских высокодоходных облигаций еще на 21% в прошлом году.

Ключевым структурным изменением стала миграция «падших ангелов» из инвестиционного класса в высокодоходный в прошлом году, что изменило кредитный профиль рынка, поскольку доля выпусков с рейтингом BB выросла с 67% до 71%.

Рисунок 2. – Структура Европейского рынка ВДО по регионам.

Источник: составлено автором на основе данных The Euro High Yield Credit Chartbook.

Германия, Великобритания и Франция входят в тройку лидеров по объемам выпуска высокодоходных облигаций в регионе EMEA и составляют 43% от общего объема за двенадцать месяцев до июня 2015 года. В то время как растущее число малых и средних эмитентов из Италии (что составляет 10% рынка) и Испания (8%) стремились получить рейтинги и вышли на рынок облигаций в 2013 и 2014 годах, эта тенденция замедлилась в первой половине 2015 года из-за улучшения условий кредитования со стороны банков. Местным банкам (исторически прочно закрепившимся на своих рынках до финансового кризиса 2009 г.) стало доступно увеличение ликвидности благодаря операциям ЕЦБ по целевому долгосрочному рефинансированию (TLTRO) и количественному смягчению, направленным на поощрение корпоративных инвестиций в поддержку Рост ВВП и занятость.

В 2022 году объем выпуска высокодоходных ценных бумаг в Европе составил всего 15,4 млрд евро в 1 кв. Об этом говорится в последнем ежеквартальном отчете European High Yield Insight. Война в Украине еще больше ограничила аппетит инвесторов и эмитентов, усугубив влияние довоенной корректировки рынка на рост кривых доходности и ожидания более высоких ставок с 4 кв. 2021 г. В марте 2022 года было выпущено высокодоходных облигаций всего на 2,8 миллиарда евро.

Средняя доходность европейских высокодоходных облигаций категории «BB» выросла до 3,45% в марте 2022 года после падения до рекордно низкого уровня в 1,88% в августе 2021 года по сравнению с пандемическим максимумом в 4,4% в апреле 2020 года. Доходность категории «B» составила 5,57% в конце марта. Доходность в 1 квартале 2022 года и доходность CCC составила 9,18% по сравнению с рекордно низким уровнем в 6,46% в сентябре 2021 года.

Рисунок 3. - Объем Европейского рынка ВДО, млрд евро.

Источник: составлено автором на основе данных High Yield Market Statistics.

Хотя лучшие макроэкономические перспективы могут привлечь на рынок эмитентов из более цикличных секторов, краткосрочные изменения в общей структуре отрасли маловероятны, что может по-прежнему затруднять диверсификацию инвестиционных инструментов.

# **Характеристика и структура Европейского рынка ВДО.**

Несмотря на высокие темпы роста европейского рынка высокодоходных облигаций, его доля на общем европейском рынке облигаций остается относительно низкой (AFME, 2020). Отчасти это связано с тем, что выпуски облигаций с высокой доходностью в основном производятся частными фирмами, а публично торгуемые компании предпочитают облигации или акции инвестиционного уровня (S&P, 2020).

Рост рынка изменил структуру европейского высокодоходного рынка и значительно диверсифицировал отрасль: от почти только телекоммуникационных и медиа-компаний до наличия практически всех основных отраслей. На Рисунке 4 показан европейский рынок высокодоходных облигаций с разбивкой по отраслям.

Рисунок 4. – Структура Европейского рынка ВДО по отраслям.

Источник: составлено автором на основе данных European Corporate HY bonds outstanding by sector, AFME, 2020.

Несмотря на то, что индустрия связи по-прежнему широко представлена, крупнейшей отраслью, безусловно, является финансовая индустрия, также сектора Материалы и Потребительский скриннинг )что это такое?, они занимают значительную часть европейского рынка высокодоходных облигаций (AFME, 2020).

Примерно 85% рынка европейских корпоративных облигаций состоит из облигаций инвестиционного класса с рейтингом BBB или лучше, в то время как оставшиеся примерно 15% рынка составляют высокодоходные облигации.

Рисунок 5. – Объем Европейского рынка ВДО по рейтингу, млрд евро.

Источник: составлено автором на основе данных European corporate HY bonds outstanding by rating, : Q4 2022 (AFME, 2022)

На рисунке 5 показана разбивка европейского высокодоходного рынка по кредитному рейтингу. Подавляющее большинство европейского рынка высокодоходных облигаций имеют рейтинг BB, BB+, BB и BB-, что составляет более двух третей рынка.

В конце 2021 года средняя продолжительность европейского рынка высокодоходных облигаций составляла около 3,5 лет по сравнению с более чем 5 годами для евро облигаций инвестиционного уровня.

Рисунок 6. - Структура Европейского рынка ВДО по сроку погашения.

Источник: составлено автором на основе данных The Euro High Yield Credit Chartbook.

На рисунке 6 показана разбивка европейского высокодоходного рынка по сроку погашения. Наиболее распространенный срок погашения составляет 5-7 лет, но значительная часть рынка также имеет срок погашения 8-10 лет. Лишь очень небольшая часть рынка имеет срок погашения выше или ниже этого уровня (AFME, 2020). Это прямо противоречит тому что дано в предыдущем абзаце

Рисунок 7. – Использование поступлений Европейского рынка ВДО.

Источник: составлено автором на основе данных*:* European HY bonds Use of Proceeds (AFME, 2022).

На рисунке 7 показано использование средств, привлеченных при выпуске европейских высокодоходных облигаций. За последние 10 лет вырученные средства в основном использовались на общекорпоративные цели или общие капитальные вложения и развитие компании-эмитента. Значительная часть вырученных средств также направляется на рефинансирование долга. Это указывает на то, что значительная часть европейского высокодоходного рынка сохраняет высокий уровень долга в структуре капитала на более постоянной и стратегической основе. Еще одним важным направлением использования вырученных средств являются приобретения. Это более волатильное использование и соответствует общему уровню активности в сфере слияний и поглощений в Европе (AFME, 2022). Выкупы с привлечением заемных средств снова приобрели некоторую популярность (Gottfried, 2018), но по-прежнему составляют довольно небольшую часть от общего объема использования выручки.

Профиль риска европейских высокодоходных облигаций сильно отличается от европейского рынка облигаций инвестиционного класса. В то время как профиль риска облигаций инвестиционного класса в основном связан с риском изменения процентных ставок и риском ликвидности, и они обладают очень небольшим кредитным риском от потенциальных дефолтов, со средней ставкой дефолта 0,03% (S&P, 2017), высокодоходные облигации несут значительный кредитный риск, средний риск дефолта составляет 3,1% (S&P, 2019).

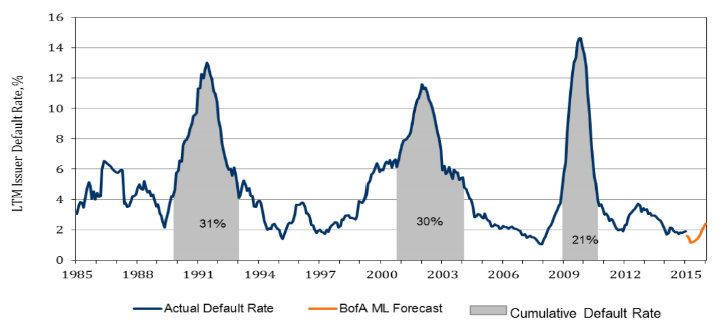


Рисунок 8. – Годовые и совокупные дефолты по высокодоходным облигациям Европейского рынка.

Источник: составлено автором на основе данных*:* European HY bonds default rates (BofA Merril Lynch, 2016).

На рисунке 8 показаны исторические годовые и совокупные дефолты по высокодоходным облигациям. На рисунке показано, насколько волатильна ставка дефолта по европейским высокодоходным облигациям и насколько высокодоходные облигации чувствительны к экономическому циклу, а значительная часть совокупных дефолтов накапливается во время экономических кризисов, таких как кризис доткомов в начале 90-х годов 2001 года и финансовый кризис 2009 года. Уровень дефолтов достиг своего пика в 2009 году после финансового кризиса, тогда уровень дефолтов составил более 14%.

Согласно данным Moody's, уровень дефолта по европейскому спекулятивному классу (высокой доходности) в октябре 2021 года составил 4,3% по сравнению с 9,1% в США. Здесь играют роль структурные факторы. Например, в США выше доля долга с рейтингом B; она имеет более высокий вес по сравнению с нефтью и газом (по которым в 2019 году был нанесен серьезный удар),(непонятно как связана доля с рейтингом и нефть) и объявить себя банкротом в США, как правило, легче, чем в Европе. Значительные фискальные и монетарные интервенции, которые наблюдаются в Европе в последнее время, вероятно, помогут смягчить последствия пандемии Covid-19 и некоторых политических аспектов. Вероятнее всего, мы увидим новые дефолты в 2023 году, но я считаю, что уровень дефолтов, достигший пика в 2020 году, так или иначе продолжит снижаться. (та к будет расти или снижаться)?

Повышение базовых ставок и расширение спредов в 1 кв. 2022 г. отражали влияние на отношение к риску продолжительного роста цен на входные данные для европейских эмитентов и вероятность ужесточения денежно-кредитной политики. Война на Украине вызывает среднесрочные опасения по поводу степени разрушения спроса на операционную среду европейских корпораций. Непосредственное влияние на европейские дефолты в основном связано с «падшими ангелами» Газпрома и РЖД, которые подняли уровень дефолта по корпоративным облигациям в регионе EMEA за последние 12 месяцев до 2,8%. Уровень дефолта в развитых странах Европы оставался исторически низким на уровне 0,5% за последние 12 месяцев.

Список Fitch «Облигации, вызывающие обеспокоенность рынка» вырос до 1,5% в марте с 0,6% в феврале. Несмотря на недавнюю волатильность на вторичном рынке, список остается существенно ниже пика апреля 2020 года, связанного с пандемией, в 6,7%. Активное рефинансирование с более низкими купонами и увеличенными сроками погашения в 2021 году вместе с продолжающимся восстановлением экономики после пандемии ограничивает существенное краткосрочное ухудшение кредитных перспектив, за исключением тех эмитентов, которые непосредственно подвержены санкциям Украины и России.

На рисунке 9 показана доходность ключевых индексов, используемых в отрасли для оценки эффективности акций, корпоративных облигаций инвестиционного уровня и высокодоходных корпоративных облигаций.

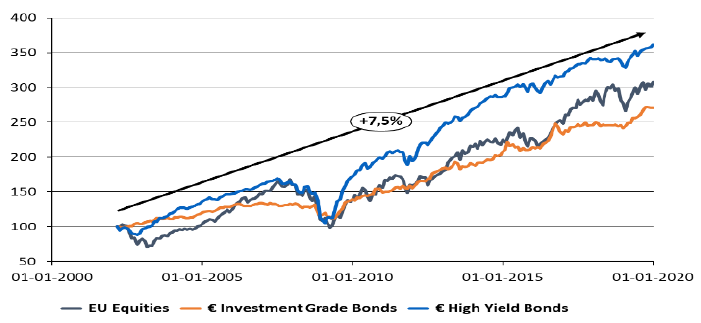


Рисунок 9. – Доходность ключевых индексов, %.

Источник: составлено автором на основе данных High Yield Market Statistics.

Хотя европейский рынок высокодоходных облигаций является часто упускаемым из виду классом активов, который не привлекает почти такого же внимания, как фондовый рынок или денежный рынок, исторически он показал чрезвычайно хорошие результаты. Европейский рынок высокодоходных облигаций за последние 20 лет превзошел как европейский фондовый рынок, так и рынок облигаций инвестиционного класса с совокупным годовым темпом роста (CAGR) ~7,5%. Это превосходство в значительной степени материализовалось за последние 10 лет, в период после финансового кризиса 2008 года (Alfawise, 2020; Bloomberg, 2020; ICE Dataservices, 2020). Эта высокая эффективность делает европейские высокодоходные облигации как класс активов интересным для исследования.

Я считаю, что Европейский рынок ВДО недостаточно изучен по сравнению с эквивалентами инвестиционного уровня. Многие эмитенты с высокой доходностью являются частными компаниями и не котируются на публичной фондовой бирже. Это увеличивает вероятность того, что ценные бумаги будут оценены неправильно, и означает, что можно выявить возможности посредством глубокого анализа.

# **Критерии отнесения облигации к высокодоходной.**

В этом разделе излагаются теоретические перспективы, имеющие отношение к пониманию того, как функционируют высокодоходные облигации, и вводится финансовая теория, используемая при ценообразовании этих ценных бумаг.

Облигация - это тип долгового инструмента, при котором эмитент облигации получает прямой приток денежных средств при выпуске и в соответствии с контрактом обязан выплатить кредитору/инвестору в течение срока действия облигации плюс дополнительные процентные платежи. Облигации отличаются от других типов долговых инструментов, таких как банковские кредиты, тем, что все облигации имеют некоторые стандартизированные договорные характеристики, что делает их гораздо более доступными для торговли на финансовых рынках (Fabozzi, 2013).

Во всем мире насчитываются сотни тысяч выпусков облигаций (Fabozzi, 2013). Большинство облигаций выпускаются с девятисимвольной идентификационной меткой, называемой CUSIP (Комитет по единообразным процедурам идентификации ценных бумаг). Этот тег CUSIP позволяет точно идентифицировать конкретные облигации при торговле на финансовых рынках, особенно когда у одного эмитента может быть много разных облигаций в обращении. Альтернативной и используемой во всем мире формой идентификации тега CUSIP является номер ISIN (Международный идентификационный номер ценных бумаг), который аналогичным образом присваивает код, однозначно идентифицирующий конкретную проблему безопасности. На протяжении всей этой статьи номера ISIN будут использоваться для отслеживания конкретных облигаций и базовых данных, относящихся к конкретной облигации.

Для большинства крупных выпусков кредитный риск облигаций оценивается тремя основными рейтинговыми агентствами: Standard & Poor Corporation (S&P), Moody's Investor Service (Moody's) и Fitch Investor service (Fitch). Эти рейтинговые агентства предоставляют относительную оценку кредитного риска эмитента и его способности своевременно погасить облигацию в течение срока погашения облигации. Основываясь на этой оценке, они присваивают буквенную оценку каждой облигации, которую они покрывают. В таблице ниже представлена сводная шкала рейтингов, используемая S&P и Moody's. S&P дополнительно модифицирует буквенные рейтинги на + и –, а Moody's - на цифры 1,2,3, чтобы достичь более высокой детализации рейтингов:

Таблица 1. – Международные рейтинги облигаций.

|  |  |  |  |
| --- | --- | --- | --- |
| **Кредитный риск** | **Moody’s** | **Standard & Poor’s** | **Fitch** |
| **Инвестционный уровень** |  |  |  |
| Высокий уровень | Aaa | AAA | AAA |
| Высокий уровень (очень уверенный) | Aa | AA | AA |
| Выше среднего (уверенный) | A | A | A |
| Средний уровень | Baa | BBB | BBB |
| **Ниже инвестиционного уровня (ВДО)** |  |  |  |
| Ниже среднего | Ba | BB | BB |
| Низкий уровень (спекулятивный) | B | B | B |
| Плохой уровень | Caa | CCC | CCC |
| Наиболее спекулятивные | Ca | CC | CC |
| Возможно банкротство | C | C | C |
| Дефолт | C | D | D |

Источник: составлено автором на основе данных Moody’s, Standard & Poor’s, Fitch.

Тем не менее, некоторое время назад несколько исследователей обнаружили, что кредитные спреды могут сильно различаться в пределах одних и тех же рейтингов (Фридсон и др., 2016; Джон и др., 2010). Например, Фридсон и др. (2016) обнаружили, что спред субординированного долга и старшего долга одного и того же рейтинга торгуются с очень разными спредами. Джон и др. (2010) также обнаружили большие различия в распределении различных аналогичных рейтинговых кредитов. Кроме того, поскольку рейтинги общедоступны для всех, они не являются отличным источником информации для управляющих фондами, позволяющим получить конкурентное преимущество и превзойти рынок. Для исследователей отсутствие прозрачности в точных расчетах, составляющих окончательный кредит, делает их неидеальными в качестве модели для объяснения распространения кредитного риска, и поэтому несколько исследователей пытались создать модели, которые могут объяснить распространение кредитного риска (Фишер, 1959; Альтман, 1968; Мертон, 1974; Фридсон & Гарман, 1998; Лонгстафф и др., 2005; Фридсон и др., 2016). Многие из них использовали структурные модели и анализ коэффициентов, уделяя особое внимание финансовому риску из вышеуказанной структуры. Хотя точная используемая модель и соотношение варьируются от исследователя к исследователю, следует применять переменные, охватывающие следующие пять областей: ликвидность, прибыльность, кредитное плечо, платежеспособность и активность (Altman, 2000).

С учетом особенностей, обычно ассоциируемых с высокодоходными облигациями, и теоретических основ, мы переходим к обзору литературы связанной с темой Европейского рынка высокодоходных облигаций. Этот анализ соответствующей литературы служит двум целям: он дополняет теоретические определения, изложенные в предыдущем разделе, контекстуализируя их с большей степенью детализации, и представляет эмпирическую сторону моего исследования, исследуя, как предыдущие ученые оценивали и прогнозировали купонные ставки во европейским ВДО.

# **Обзор исследуемой литературы.**

В данном разделе я проанализировала эмпирические исследования спредов и эффективности корпоративных облигаций, а также исследования, в которых уже использовалось машинное обучение для определения эффективности ценных бумаг.

Одно из самых ранних эмпирических исследований рынка корпоративных облигаций было проведено в 1959 году Лоуренсом Фишером. Фишер (1959) исследовал детерминанты премий за риск, обнаруженных в корпоративных облигациях. В 1968 году Альтман разработал систему определения риска дефолта, основного компонента премии за риск, с помощью аналитической методики ratio analysis с разработкой его Z-балла (Altman, 1968). Ценообразование облигаций с акцентом на рискованный характер актива получило дальнейшее развитие благодаря развитию структурного подхода к ценообразованию рискованного долга, разработанного Мертоном (1974). С тех пор во многих исследованиях пытались использовать структурный подход с акцентом на кредитный риск при изучении ценообразования облигаций (Black & Cox, 1976; Sundaresan et al, 1993; Shimko et al, 1993; Nielsen et al, 1993; Longstaff & Schwartz, 1995; Андерсон и Сундарезан, 1996; Джарроу и Тернбулл, 1995; Ландо, 1998; Коллин-Даффресн и Голдштейн, 2001; Кэмпбелл и Тасклер, 2003; Бутера и Фафф, 2006; Эрикссон и др., 2009). Однако такие исследования, как Huang & Huang (2003), Huang & Huang (2012), Longstaff и др. (2005), Ericsson & Renault (2006) и Dick-Nielsen и др. (2012), документально подтверждают, что структурный подход и используемые ими модели недооценивают риск доходности, подразумевая, что разброс доходности корпоративных облигаций содержит другие премии, помимо премии за кредитный риск. Этот разрыв между наблюдаемыми спредами и спредом, необходимым для компенсации ожидаемых убытков от дефолта, был назван "загадкой кредитного риска" (Amato & Remolona, 2003). Вслед за выводом о том, что структурные модели и исследования фокусируются только на премии за риск, необходимой для компенсации ожидаемых убытков от дефолта, появилась новая литература, посвященная значению ликвидности для объяснения спредов доходности, наблюдаемых на рынках корпоративных облигаций. Как следствие, большинство современных исследований либо фокусируются на объяснении кредитного риска, либо премии за ликвидность, одновременно контролируя другие.

Как уже упоминалось, Фишер (1959) разработал самые ранние эмпирические оценки распределения кредитного риска по корпоративным облигациям. Фишер исследовал рынок облигаций США и разработал следующие четыре гипотезы: 1) средняя премия за риск по облигациям фирмы зависит, во-первых, от риска дефолта и, во-вторых, от их товарности; 2) риск дефолта может быть оценен по трем переменным: ‘коэффициент вариации чистого дохода фирмы за последние девять лет’ (стабильность прибыльности), ‘продолжительность деятельности фирмы, не вынуждая ее кредиторов нести убытки" (последовательная управленческая деятельность) и ‘отношение рыночной стоимости собственного капитала фирмы к номинальной стоимости долга фирмы’ (структура капитала / левередж); 3) ‘рыночность’ фирмы может быть оценена с использованием одной переменной: рыночной стоимости всех публично торгуемых облигаций, находящихся в обращении у фирмы (ранняя альтернатива ликвидности); 4) логарифм средней премии за риск может быть оценен с помощью линейной функции логарифмов четырех только что перечисленных переменных. Фишер проверил это на выборке из 366 наблюдений за облигациями промышленных компаний за 5-летний промежуток времени и смог объяснить 74 процента изменений премий за риск.

Другие исследования основывались на анализе коэффициентов в попытках предсказать доходность корпоративных облигаций. Хурана и Раман (2003) исследуют актуальность долгосрочных фундаментальных показателей для риска дефолта, поскольку они настраивают регрессионную модель, используя широкий спектр фундаментальных показателей эффективности, включая Z-балл Альтмана, для прогнозирования сроков погашения новых выпусков корпоративных облигаций. Было обнаружено, что как совокупный фундаментальный балл, так и отдельные фундаментальные показатели обеспечивают дополнительную объяснительную силу при оценке новых облигаций как показателей ожидаемой будущей прибыли и платежеспособности, и что они обладают значительной объяснительной силой, превышающей опубликованную классификацию рейтингов облигаций (Kuhrana & Raman, 2003).

Многие из вышеупомянутых исследований проводятся на выборке облигаций инвестиционного класса или на выборке корпоративных облигаций в целом, что подразумевает, что большая часть выборки - облигации инвестиционного класса. Как уже упоминалось, более высокий профиль риска высокодоходных облигаций означает, что информация о будущих результатах деятельности компании и оценки вероятности дефолта более важны для спредов. Это в сочетании с относительно меньшим размером рынка и более низкой ликвидностью по сравнению с облигациями инвестиционного класса (AFME, 2020) означает, что литература, специально посвященная рынку высокодоходных облигаций, актуальна для углубления понимания влияния этих различий. Однако литература по высокодоходным облигациям гораздо более скудна, чем относительно богатая литература по корпоративным облигациям в целом. Особенно ограничена литература по европейскому рынку высокодоходных облигаций из-за его относительно молодого возраста и нехватки доступных данных, поскольку многие европейские эмитенты HY являются частными компаниями.

Из-за высокого профиля риска облигаций не инвестиционного класса предложения облигаций HY исторически обозначались как "облигации с историей", для которых трудно установить количественные, объективные критерии оценки для определения цены или спреда. Аналитикам пришлось бы узнать о природе этих компаний, используя более глубокий специальный анализ (Fridson & Garman, 1998). Фридсону и Гарману (1998) удается объяснить 56% различий в премиях за риск по высокодоходным облигациям США с помощью поддающихся количественной оценке факторов, таких как рейтинги, структура сроков и спреды на вторичном рынке. Профиль высокого риска также делает прогнозирование вероятности дефолта еще более ценным, и в нескольких исследованиях, таких как Huffman & Ward (1996), применялись различные структурные подходы и многомерный анализ для прогнозирования ставок дефолта по высокодоходным облигациям. Хаффман и Уорд (1996) смогли правильно предсказать 73,3 процента облигаций с дефолтом и 68,6 процента облигаций без дефолта. Применение машинного обучения для прогнозирования доходности облигаций не является полностью неизученным направлением. Эшби и Кумар (1996) попытались применить раннюю стадию нейросетевых моделей с использованием одного скрытого слоя, достигнув 89% точности классификации 56 записей.

Фридсон (1990) обнаружил, что первоначальная цена обладает объяснительной силой в качестве предиктора последующей доходности облигаций HY. Он обнаружил, что высокодоходные облигации, которые в конечном итоге сталкиваются с серьезными кредитными проблемами, как правило, не выходят на рынок как выпуск среднего качества. Вместо этого распространение, с которым они выходят на рынок в рамках первоначальных предложений, помечает их как потенциальных проблемных кандидатов. Как следствие, простое стремление к максимальной доходности в пределах класса рейтинга представляется сомнительной инвестиционной стратегией, и требуются превосходные аналитические навыки (Фридсон, 1990). В статье 1996 года Фридсон и Гао (1996) исследуют систематическое занижение цен при первоначальном размещении высокодоходных облигаций. Установлено, что первичные выпуски обеспечивают более высокую доходность с поправкой на риск в период, последующий за выпуском, и модель способна объяснить 64% различий в первичный и вторичный спред доходности, используя прокси для предложения, спроса и ликвидности на высокодоходном рынке.

В более поздних исследованиях Джентри и др. (2010) исследуют статистические свойства спреда кредитного риска для высокодоходных облигаций и анализируют влияние набора переменных, которые, как ожидается, окажут влияние на спред кредитного риска. Они изучают группу из 9 переменных с помощью многомерного анализа и приходят к выводу, что наибольшее влияние оказывает ряд переменных риска дефолта. Подобно выводам Кэмпбелла и Тасклера (2003) и Эрикссона и др. (2009) о взаимосвязи между спредами по облигациям и волатильностью базовых акций на общих рынках корпоративных облигаций, Ву и Чжан (2014) исследуют взаимосвязь между доходностью рынка высокодоходных облигаций и фондового рынка. Исследование показало, что акции обеспечивают высокую доходность и что эта взаимосвязь между опережением и запаздыванием наиболее сильна на медвежьих рынках, поскольку тенденция к снижению на фондовом рынке подразумевает высокую вероятность использования акций с короткой позицией, вложенных в высокодоходные облигации при погашении, аналогично модели, разработанной Мертоном (1974).

Ли и др. (2014) исследуют эффективность высокодоходных облигаций по сравнению с инвестиционным классом. Они показывают, что, когда предполагается нормальное распределение, высокодоходные облигации достигают более высокой ожидаемой доходности и меньшего стандартного отклонения. Но также и то, что как высокодоходные облигации, так и облигации инвестиционного класса имеют "жирные хвосты", что означает, что при исследовании профилей риска высокодоходных облигаций следует применять распределения, допускающие "хвосты" или асимметрию.

Не все высокодоходные долговые обязательства с одинаковым кредитным рейтингом имеют одинаковый профиль риска или требуют одинакового распределения кредитного риска. Фридсон и др. (2016) показывают, что спреды доходности по корпоративным облигациям с рейтингом В были больше по облигациям с длительным сроком погашения, чем по субординированным облигациям. Джон и др. (2010) также обнаружили, что предполагаемые менее рискованные облигации с длительным сроком погашения имели более широкие спреды, чем субординированные выпуски с аналогичным рейтингом. Это указывает на то, что можно ожидать, что более низкая вероятность дефолта по субординированному выпуску с рейтингом B4 с лихвой компенсирует увеличение убытков в случае дефолта эмитента с высшим рейтингом B1, что приведет к снижению общего ожидаемого убытка по субординированному выпуску (Fridson et al, 2016).

Подводя итог, можно сказать, что литература по корпоративным облигациям в целом и высокодоходным облигациям в частности подчеркивает важность правильной оценки кредитного риска при попытке объяснить общие спреды доходности. Поскольку высокодоходные облигации являются типом актива с наименьшим количеством доступной информации и в то же время в наибольшей степени подвержены кредитному риску, увеличение объема данных, используемых для прогнозирования эффективности, путем включения качественных параметров из проспектов облигаций кажется привлекательным предлогом. Это также хорошо согласуется с идеей облигаций HY как «исторических облигаций», нуждающихся в дополнительном кредитном анализе (Fridson & Garman, 1998), и позволяет лучше прогнозировать последующие показатели по мере увеличения объема информации при выпуске (Fridson, 1990).

# **Эмпирические исследования эффективности финансовых ценных бумаг с использованием числовых и текстовых данных.**

Цель этого раздела не состоит в том, чтобы представить теорию и литературу, лежащие в основе моделей обработки данных и машинного обучения, которые будут использоваться для анализа в моей дипломной работе, поскольку это будет рассмотрено в последующем разделе методологии. Вместо этого цель этого раздела состоит в том, чтобы изложить существующую литературу по теме применения обработки естественного языка и числовых данных для анализа эффективности финансовых ценных бумаг, и, таким образом, изложить существующие структуры передовой практики, а также предоставить дополнительные доказательства обоснованности использование числовых и текстовых данных в качестве входных данных для более количественных моделей финансовых рынков.

Область использования текстового и числового анализа относительно молода из-за юного возраста технологии, которая позволяет это сделать. Но недавние достижения в области НЛП и технологии машинного обучения позволили исследователям проводить более глубокий анализ с использованием текста и цифр. Хотя не существует каких-либо широко известных или признанных исследовательских работ, использующих методы НЛП для анализа данных по рынку высокодоходных облигаций, ряд исследователей пытались использовать анализ НЛП для объяснения тенденций на фондовых рынках, что мы дальше и рассмотрим.

Одним из первых исследований использования анализа НЛП на фондовых рынках является Тетлок (2007). Тетлок (2007) пытается проанализировать взаимосвязь между контентом, публикуемым СМИ, и последующими движениями на фондовом рынке с помощью количественного анализа ежедневных котировок акций по индексу S&P 500 и анализа НЛП популярной колонки Wall Street Journal. Тетлок (2007) использует Гарвардский психосоциальный словарь IV-4, чтобы разделить каждое слово столбца на различные категории, такие как ‘положительный’ или ‘отрицательный’. Тетлок (2007) затем продолжает использовать количество этих категорий в качестве показателя настроения или тона в средствах массовой информации и обнаруживает, что высокий пессимизм СМИ предсказывает понижательное давление на рыночные цены с последующим возвратом к фундаментальным показателям, а необычно низкий пессимизм предсказывает высокий объем торгов на рынке. В Tetlock et al (2008) единица анализа перенесена с уровня рынка на уровень отдельной компании, исследуя, можно ли использовать простой количественный показатель языка для прогнозирования бухгалтерской прибыли отдельных фирм и доходности акций. В ходе анализа они использовали отрицательную оценку слов в качестве входных данных для своей количественной модели, основанной на психосоциальном словаре Гарварда IV-4. Установлено, что доля негативных слов в новостях, посвященных конкретной фирме, предсказывает низкие доходы фирмы (Tetlock et al, 2008).

В статье 2011 года Loughran & McDonald обнаружили, что списки слов, основанные на общих словарях, таких как Harvard IV-4 психологический словарь, часто неправильно классифицируют слова в финансовых текстах. В их выборке три четверти всех слов, классифицированных гарвардским словарем как "негативные", обычно не считаются негативными в финансовом контексте. Исследователи приступают к разработке собственного списка слов, используя метод набора слов и взвешивая важность слов с использованием матрицы TF-IDF, которую затем можно использовать в качестве входных данных для обучения классификатора с помощью методов машинного обучения. Они объединяют этот список слов другими списками слов, чтобы проанализировать взаимосвязь между тоном в годовых отчетах компании и объем по результатам торгов, волатильность, доходность и риск неожиданного снижения прибыли по базовым акциям и компании. Джегадеш и Ву (2013) разработали свою собственную схему взвешивания в качестве альтернативы TF-IDF и предложили подход, который присваивает определенный вес каждому слову на основе реакции рынка на документы, содержащие эти слова. Это ручной подход, который очень похож на то, как современные алгоритмы машинного обучения присваивают веса с помощью грубой силы вычислительной мощности (Géron, 2017).

В контексте анализа НЛП первичное публичное размещение акций компании имеет некоторое сходство с размещением высокодоходных облигаций, поскольку, как и при размещении облигаций, вся информация об IPO и компании-эмитенте представлена в длинном проспекте. Несколько исследователей использовали лингвистический и текстологический анализ проспектов IPO, чтобы пролить новый свет на эффективность IPO. Loughran & McDonald (2011) используют анализ тона в заявках по форме S-1 (первой в SEC 6 при IPO) и обнаруживают, что IPO с высоким уровнем неопределенности текста в их заявках по форме S-1 имеют более высокую доходность в первый день, пересмотр абсолютной цены предложения и последующую волатильность.

Бартов (2011) использует НЛП для анализа взаимосвязи между содержанием раздела о рисках проспектов IPO и будущими доходами и прогнозом аналитиков. Они используют три различных подхода для количественной оценки отрицательной стороны текстовой информации о рисках, все с использованием списка слов, составленного на основе чтения многочисленных проспектов. Они обнаружили, что текстовая информация о риске снижения прибыли в разделе "Риски" проспектов IPO коррелирует с будущей прибылью и ошибкой прогноза аналитиков, но не с самим прогнозом аналитиков, что также доказывает, что текст раздела "Риски" может обеспечить объяснительную силу, превосходящую ту, которая доступна в прогнозах аналитиков (Бартов, 2011). Фишер и др. (2015) пытаются объяснить доходность IPO в первый день с помощью анализа НЛП и Past-of-Speech (PoS) раздела рисков проспектов IPO. Они обнаружили, что списки слов с настроениями значительно коррелируют с результатами первого дня, но также обнаружили, что слова, которые недопредставлены в таких списках, обладают еще большей объяснительной силой. Кроме того, анализ PoS показывает, что существительные и прилагательные обладают наибольшей объяснительной силой (Fisher et al, 2015). В более поздней статье Yan et al (2019) показывают, что негативный или неопределенный тон в проспектах снижает долгосрочную доходность акций, анализируя взаимосвязь между тоном IPO и первоначальной доходностью IPO на китайском рынке.

Hanley & Hoberg (2010) использует текстовое содержание проспектов IPO в качестве показателя того, сколько времени андеррайтер потратил на проведение предрыночной проверки. Если содержание очень похоже на общее содержание (близко к среднему значению), большая часть текста является общим текстом, который был взят из предыдущих выпусков, что свидетельствует о нехватке времени, затраченного на должную осмотрительность. Более уникальные тексты IPO сигнализируют о больших усилиях андеррайтера на премаркете. Deokar & Tao (2015) используют анализ NLP в проспектах IPO на уровне предложений и разрабатывают структуру FOCUS-IE для анализа финансовых текстовых документов (функционально-ориентированный, контекстно-зависимый, систематическое извлечение информации). Они объединяют текстовое настроение, полученное из фреймворка FOCUS-IE, с несколькими современными методами машинного обучения и обнаруживают, что ценообразование может быть лучше предсказано с использованием текста функции наряду с функциями количественного IPO, чем только с функциями количественного IPO. Они достигают наилучшего результата с помощью модели дерева решений, но также значительно превосходят количественную модель только с помощью модели искусственной нейронной сети (Deokar & Tao, 2015). Эти выводы основаны на статье Deokar et al (2018). В этой статье они разрабатывают экстрактор прогнозных утверждений и классификатор, использующий архитектуру глубокого обучения, которая, как оказалось, превосходит любую предыдущую модель. Прогнозные заявления используются для анализа изменений цен до IPO и доходности в первый день после IPO, обнаруживая, что функции FLS более предсказуемы для предварительного IPO по сравнению с прогнозом оценки после IPO (Deokar et al, 2018).

Анализ НЛП также был применен к другому контенту проспектов IPO для объяснения доходности акций. Несколько исследователей использовали текстовые данные Twitter в качестве показателя общественного настроения для прогнозирования движений на фондовом рынке (Patel, 2016; Fabozzi et al, 2016). Патель (2016) использует классификатор настроений для извлечения настроений из данных Twitter для изучения взаимосвязи между настроениями в Twitter и движениями фондового рынка. Он использует три метода машинного обучения: наивный Байес, дерево решений и машины опорных векторов и обнаруживает, что модель может успешно предсказывать движение акций на основе настроений в Твиттере. Фабоцци и др. (2016) проводят аналогичное исследование, но используют матрицу текстового контента TF-IDF, а не настроения, чтобы попытаться предсказать движение акций.

Из обзора литературы ясно, что использование машинного обучения является актуальным и может быть эффективным при попытке предсказать эффективность финансовых инструментов. Но поскольку ранее не проводился анализ высокодоходных облигаций, моё исследование попытается внести свой вклад в литературу, продемонстрировав актуальность использования машинного обучения для этого класса активов. Кроме того, поскольку данная область находится в быстром и постоянном развитии, можно будет применять модели, более сложные, чем те, которые использовались ранними исследователями, тем самым расширяя понимание того, как различные факторы влияют на купонную ставку и последующую эффективность облигаций.

# **ГЛАВА 2. МЕТОДОЛОГИЯ.**

Основываясь на представленной теории, а также на обзоре эмпирической литературы, цель этой главы - объяснить теоретические основы, лежащие в основе нашей эмпирической модели. В этом разделе будут рассмотрены теоретические соображения при выборе независимых и зависимых переменных для оценки купонных ставок высокодоходных облигаций. Будет представлен краткий обзор лучших теоретических предикторов, а также то, как они могут быть опосредованы измеряемыми переменными.

Выбор переменных можно разделить на две группы: выбор зависимой переменной, переменной, которую модель стремится объяснить и / или предсказать; и выбор независимых переменных, которые должны иметь теоретический эффект или корреляцию с зависимой переменной. Выбор зависимой переменной в этой статье касается выбора правильного показателя эффективности облигаций. Выбор независимых переменных делится на две общие категории: 1) учет переменных, связанных с фирмой-эмитентом, целью которого является отражение текущих и ожидаемых будущих результатов деятельности фирмы-эмитента; 2) другие переменные, не относящиеся к финансам и бухгалтерии фирмы.

# **2.1. Описание переменных**

*Выбор зависимой переменной (таргет)*

Первым шагом к построению эмпирической модели является выбор переменной для измерения. Эта переменная будет функционировать как метка для входных данных и является переменной, подлежащей объяснению или прогнозированию. Поэтому она должна соответствовать исследовательскому вопросу, на который я и пытаюсь ответить в своей работе. В своей работе я буду прогнозировать купонную ставку по высокодоходным облигациям европейского рынка.

В качестве таргетируемой переменной в моей работе выступает купонная ставка бумаги при размещении.

**Купон**– это процентная ставка по облигации, которую должен заплатить эмитент облигации. Купонная ставка (CR) = Периодический доход, выплачиваемый раз в полгода и измеряемый в процентах. Ожидается, что облигация с более высоким купонным платежом будет платить более высокий налог. Чтобы сделать связь привлекательной, инвестору предлагается более высокая доходность. Это говорит о положительном отношении между доходностью облигации и ставкой купона.

Купон или купонный платеж представляет собой годовую процентную ставку, выплачиваемую по облигации, выраженную в процентах от номинальной стоимости и выплачиваемую с даты выпуска до погашения. Купоны обычно упоминаются с точки зрения ставки купона (сумма купонов, выплачиваемых в течение года, деленная на номинальную стоимость рассматриваемой облигации).

*Выбор объясняющих переменных*

В этом разделе будет описан наш выбор объясняющих переменных для таргетируемой переменной (купона).

1. **Страна эмитента.**

Экономические и правовые условия могут варьироваться в зависимости от страны. Как и в случае с отраслью, две компании с одинаковыми финансовыми показателями, но расположенные в разных странах, могут иметь разные профили риска. Таким образом, страна может повлиять на распространение облигаций, поскольку инвестор предпочел бы облигации страны с сильной экономикой и правовой базой, а не страны с более проблемной экономикой и правовой базой, если в остальном две облигации идентичны. Например, Lie & Nielsen (2015) изучали премию по облигациям, выпущенным в Южной Европе, по сравнению с Северной Европой.

1. **Отраслевая группа.**

Риск компаний в разных отраслях может сильно различаться. Например, поставщики коммунальных услуг имеют очень надежные и стабильные денежные потоки, в то время как будущие денежные потоки компаний, занимающихся модой, будут гораздо более неопределенными. Таким образом, две компании с одинаковыми финансовыми показателями, но где одна является поставщиком коммунальных услуг, а другая - модной компанией, будут иметь очень разные профили риска. Как следствие, облигация, выпущенная модной компанией, потребует гораздо более высокого спреда. В результате классификация отрасли или вида деятельности станет потенциальной объясняющей переменной для прогнозирования спредов по облигациям.

1. **Эмитент.**

На Европейском рынке ВДО в заданных нами параметрах обращается не так много различных эмитентов, поэтому для нас эмитент будет играть важную роль при дальнейшем прогнозировании купонной ставки.Над объяснить в чем различия эмитентов – количественно и качественно

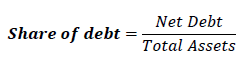
1. **Структура капитала/запас собственного капитала (Share of Debt).**

Одним из первых ключевых показателей для анализа является структура капитала компании – какую долю от полной стоимости компании составляет долг. Капитал, который акционеры предоставляют компании, в структуре капитала подчинен любой форме долга. Это означает, что акционеры понесут убытки до того, как какая-либо стоимость будет потеряна для держателей долгов. В результате, в целом, компания должна быть в состоянии избежать дефолта до тех пор, пока стоимость остается в транше акционерного капитала компании, поскольку компания может использовать этот капитал для обслуживания долга, а в случае технического дефолта ставка возмещения для держателей облигаций должна составлять 100% если какая-либо ценность останется для акционеров. По этой причине уровень собственного капитала по отношению к общей стоимости компании часто называют запасом собственного капитала, поскольку он функционирует как запас собственного капитала, который необходимо будет уничтожить, прежде чем какая-либо ценность будет потеряна для держателей облигаций. В идеале следовало бы использовать рыночную стоимость компании и собственный капитал компании:

Запас собственного капитала:



Однако для частных компаний такие данные недоступны. Вместо рыночной стоимости компании и собственного капитала в качестве другого показателя можно использовать балансовую стоимость компании, определяемую балансовой стоимостью совокупных активов из бухгалтерского баланса:



Более высокий запас собственного капитала и меньшая доля долга теоретически приведут как к снижению риска дефолта, так и к меньшим убыткам в случае дефолта и, как ожидается, будут коррелировать со спредом, по которому торгуется облигация.

1. **Кредитное плечо (Leverage).**

Одним из наиболее широко применяемых показателей является уровень долга компании по отношению к ее прибыльности, известный как кредитное плечо (Fridson, 2018). Удерживая уровень долга относительно прибыли компании, это дает представление о том, насколько устойчивым является уровень долга, или, другими словами, каков риск дефолта по долгу. Наиболее широко используемым показателем кредитного плеча является соотношение долг/ EBITDA:



Поскольку этот показатель доступен как для частных, так и для публичных компаний (при условии, что у вас есть доступ к финансовым данным частных компаний). Кроме того, показатель EBITDA используется в качестве показателя рентабельности по трем причинам: во-первых, это показатель прибыли, который наиболее тесно связан с фактическими денежными средствами, полученными от прибыли, поскольку он рассчитывается до начисления амортизации. Во-вторых, это доход, доступный для обслуживания долга, поскольку налоги сначала рассчитываются после обслуживания долга. В-третьих, он оценивает полученные денежные средства, устраняя при этом шум, такой как разовые расходы, списания и т.д. Уровень кредитного плеча может изменяться двумя способами: либо компании увеличивают или уменьшают свой уровень долга до желаемого уровня левереджа, либо показатель EBITDA улучшается или ухудшается при стабильном уровне долга. Устойчивый уровень левереджа сильно варьируется от отрасли к отрасли, в зависимости от того, насколько надежны или стабильны доходы в отрасли. Однако при прочих равных условиях более высокое кредитное плечо привело бы к увеличению вероятности дефолта. Показатели левереджа в виде отношения чистого долга к EBITDA или в виде доли в полной структуре капитала используются во многих исследовательских работах о спредах кредитного риска по облигациям, таких как Fisher (1959), Blume и др. (1998), Collin-Duffresne и др. (2001), Ericsson и др. (2009), Kovner & Wei (2012), Ashby & Kumar (1996), которые все сочли термины кредитного плеча важным показателем для объяснения спредов кредитного риска по облигациям.

1. **ROA.**

Термин рентабельность активов (ROA) относится к финансовому коэффициенту, который показывает, насколько прибыльна компания по отношению к ее общим активам. Корпоративное управление, аналитики и инвесторы могут использовать ROA, чтобы определить, насколько эффективно компания использует свои активы для получения прибыли.

**Коэффициент рентабельности = чистая прибыль/совокупные активы \* 100**

Метрика обычно выражается в процентах с использованием чистой прибыли компании и ее средних активов. Более высокая рентабельность инвестиций означает, что компания более эффективна и продуктивна в управлении своим балансом для получения прибыли, в то время как более низкая рентабельность инвестиций указывает на то, что есть возможности для улучшения.

1. **Рейтинг эмитента.**

Это значение каждого рейтинга согласно рейтингу, выпущенному рейтинговым агенством. Рейтинг облигации представляет собой рейтинг от ВВ+ и ниже до ССС. Облигации с низкими рейтингами указывают на более высокий риск дефолта и, соответственно, ожидается, что от этого доходность облигации будет выше.

Как уже объяснялось, кредитные рейтинги содержат большой объем информации о кредитном риске конкретного выпуска. Риск включения этого параметра в модель заключается в риске полной мультиколлинеарности между рейтингами и некоторыми другими метриками, поскольку они могут быть ключевыми показателями для присвоения самого рейтинга. Это необходимо будет проверить, прежде чем включать кредитные рейтинги в какую-либо модель. В отсутствие идеальной мультиколлинеарности рейтинги являются хорошим показателем для включения. Кроме того, рейтинг будет служить ориентиром для любой модели, основанной на данных, поскольку он должен будет превзойти модель, которая использует только кредитные рейтинги в качестве входных данных, чтобы добавить новую объяснительную силу.

1. **Годовая инфляция страны эмитента.**

**Инфляция** — это данные, выгруженные из Unctad Stat по состоянию на 31 января в течение исследовательского периода с 2003 по 2022 год.

Инфляция отражает в целом экономическую ситуацию в стране, что конечно же сказывается на купонных ставках по ВДО.

1. **Месяц выпуска ВДО.**

Проведя сравнительный анализ, нами была обнаружена интересная особенность. Большинство бумаг с доходностью больше 8% как правило выпускались ближе к концу года, либо в самом начале года. Также наибольший выпуск облигаций, доходностью около 5% приходился на середину и конец года. Исходя из этого было принято брать данный фактор в модель для более точного предсказания купонной ставки.

1. **Дюрация.**

Срок владения облигацией, измеряемый в годах; это значение Доли к Текущей стоимости, деленное на полученный Денежный поток, умноженное на год, в котором он был выплачен, формулируется следующим образом:

Изображение выглядит как текст

Автоматически созданное описание

Ожидается, что более длительный срок погашения укажет на более высокий риск процентной ставки и, как таковой, прогнозируется предложит более высокую доходность.

1. **Периодичность выплаты купонов.**

Исходя из анализа собранных данных было обнаружено, что наибольшее количество облигаций с купонной ставкой не более 5% имеют только одну выплату купона, в тоже время, те облигации, которые предусматривают выплату дважды имеют самые высокие ставки. Поэтому данный фактор было решено оставить в модели.

1. **Объем размещения.**

Данный показатель мы брали с Cbonds и нами была обнаружена следующая особенность. В бумагах, чей объем не превышал 500 млн шт, средняя купонная ставка не была выше, чем 6%. Напротив, в бумагах с объемом более 500 млн шт купон чаще достигал значения от 6% до 9%, поэтому данный показатель было также решено оставить как объясняющий.

1. **Номинал / мин. торговый лот.**

Минимальный торговый лот — это минимальный объем ценных бумаг по номиналу, доступный к продаже.

В ситуации с нашими данными, если торговый лот был равен 100 тыс единиц, то распределение процентной ставки купона было относительно равномерным, но в ситуациях с торговым лотом меньше 100 тыс единиц ставка купона росла и начиналась не с 0% как в предыдущем случае, а как правило с 2%.

На рынке еврооблигаций зачастую минимальный торговый лот значительно превышает 1000 и обычно составляет 100 000 или 200 000 (чаще долларов США или евро), что зачастую ограничивает круг потенциальных инвесторов и делает еврооблигации менее ликвидными инструментами. Однако среди этой практики также встречаются исключения с меньшими минимальными торговыми лотами. Например: Турция, 11.875% 15jan2030, USD (US900123AL40). Минимальный торговый лот по этой бумаге составляет 1000 долларов США. Есть и пример с еще более маленьким минимальным лотом — Россия, 7.5% 31mar2030, USD (XS0114288789), по которой минимальный торговый лот составляет лишь 1 доллар США. Для повышения спроса на еврооблигации в последние несколько лет брокеры начали дробить большие минимальные лоты и предлагать их своим клиентам.

1. **Тип облигации.**

Еврооблигации выпускаются в валюте, не являющейся национальной для страны эмитента, тогда как облигация представляет собой облигацию с номиналом в валюте страны эмитента, выпущенную компанией за рубежом. Другими словами, облигация позволяет эмитенту облигаций привлекать средства от зарубежных инвесторов в «родной» валюте, в то время как еврооблигации выпускаются в валюте, отличной от местной валюты эмитента. В облигации заемщику не нужно беспокоиться об обесценивании местной валюты, и она не связана с валютным риском.

Могут быть различия, связанные с валютными рисками

1. **Условия досрочного выкупа.**

Как правило облигации имеющие условие досрочного выкупа, имеют более низкие ставки, чем бумаги, не имеющие такого условия.

1. **Цена первичного размещения, %**

Бумаги с ценой ниже первичного размещения на 1-2% имеют более высокую процентную ставку по сравнению с бумагами, имеющими ставку 100% от первичного размещения.

1. **ВВП, млрд долл**

Экономические тенденции являются критически важными факторами, определяющими производительность рынка облигаций. Как и другие виды инвестиций, облигации привязаны к состоянию экономики, потому что предприятия и правительства, которые их выпускают, существуют внутри этой экономики.

Экономический рост происходит, когда страна увеличивает темпы экономического производства, известного как «валовой внутренний продукт» (ВВП). ВВП является наиболее часто используемым показателем эффективности экономики. Положительное изменение ВВП означает экономический рост; отрицательное изменение означает усадку.

Более сильная экономика снижает доходность казначейских облигаций и облигаций, но гораздо более вероятно, что это будет положительным фактором для высокодоходных облигаций, где кредитоспособность эмитента является серьезной проблемой.

1. **Индекс потребительских цен, ежегодный рост %**

Индекс потребительских цен (ИПЦ) измеряет ежемесячное изменение цен, уплачиваемых потребителями в той или иной стране. Бюро трудовой статистики рассчитывает ИПЦ как средневзвешенное значение цен на корзину товаров и услуг, представляющих совокупные потребительские расходы в стране.

ИПЦ является одним из самых популярных показателей инфляции и дефляции. В отчете об ИПЦ используется другая методология обследования, ценовые выборки и веса индекса, чем в индексе цен производителей (ИЦП), который измеряет изменения в ценах, полученных производителями товаров и услуг.

1. **Легкость ведения бизнеса, место в рейтинге.**

Страны ранжируются по легкости ведения бизнеса от 1 до 190. Высокий рейтинг легкости ведения бизнеса означает, что нормативно-правовая среда более благоприятна для создания и работы местной фирмы. Рейтинги определяются путем сортировки совокупности баллов по 10 темам, каждая из которых состоит из нескольких показателей, придающих каждой теме равный вес.

1. **Индекс коррупции.**

Индекс коррупции или индекс восприятия коррупции (Corruption Perception Index) – ежегодный составной индекс, измеряющий уровень восприятия коррупции в государственном секторе различных стран. Индекс готовится Transparency International на основе данных опросов, проведенных среди экспертов и в деловых кругах. На основе этой информации страны мира ранжируются по шкале от 0 до 100 баллов. Ноль обозначает самый высокий уровень восприятия коррупции, а сто – наименьший. Индекс коррупции рассчитывается и публикуется ежегодно с 1995 года.

# **2.2. Описание выборки**

В исследовании будут использоваться два источника данных, в одном источнике будут присутствовать данные в целом по выпуску и эмитенту, в другом источнике сосредоточены данные об общей экономической ситуации в стране.

В следующих параграфах будут описаны источники используемых данных, а также процесс сбора данных, прежде чем представить описательную статистику и обсудить любые шаги, предпринятые для очистки и обработки данных надлежащим образом для ввода в анализ. Чтобы должным образом охватить различные стратегии, используемые для обработки данных, в отдельных разделах будет дано подробное описание данных по облигациям, бухгалтерских данных компаний, лежащих в их основе, и, наконец, данных об экономическом состоянии той или иной страны.

Исследование будет проведено с использованием разных источников данных, которые представляют собой совокупность информации о высокодоходных облигациях, выпущенных в Европе. Один из них - Cbonds и включает в себя в основном данные, касающиеся непосредственно выпуска и немного затрагивающий общее экономическое положение той или иной страны, в то время как другой – это данные с сайта Unctad Stat, которые включают в себя только данные относительно экономического положения той или иной страны.

Моя выборка состоит из облигаций публично торгуемых компаний и частных компаний в том числе. Несомненно, сбор данных был более легкодоступен для публично торгуемых фирм, были доступны более детализированные данные. Примером этого являются данные бухгалтерского учета, которые все публично торгуемые фирмы обязаны представлять на ежеквартальной основе. Для частных фирм это не всегда так, что означает, что некоторые компании, включенные в мой набор данных, могут быть исключены ввиду отсутствия необходимой мне информации.

Хотя процесс сбора данных был полностью ручным, вся обработка данных и последующий анализ выполнялись с использованием системы визуального программирования для отображения данных, машинного обучения и интеллектуального датамайнинга - Orange. На протяжении всего документа будут делаться ссылки на соответствующие системные шаги в программе, а полученные выводы программы будут доступны в качестве приложения. Ниже приводится краткое описание процесса сбора данных для моей выборки.

# **2.3. Сбор данных**

Первый набор данных собран в Cbonds и доступен через корпоративный доступ. Cbonds хранит финансовые и бухгалтерские данные по всем публичным эмитентам. Основа набора данных была создана путем фильтрации всех активных и погашенных корпоративных облигаций, выпущенных в Восточной или Западной Европе, которые публично торгуются по состоянию на 23 февраля 2023 года в EUR, являются купонными с фиксированной ставкой, а также рейтинг эмитента S&P должен быть в диапазоне от CCC до BB+. Всего в выборку попали 530 облигаций с датой выпуска с 2003 по 2023 год.

Затем применяются фильтры в соответствии с целью исследования. Во-первых, сразу из выборки были исключены бумаги, по которым отсутствуют бухгалтерские или финансовые данные как в Cbonds, так и в открытом доступе. После данной фильтрации остаются 480 бумаг. Затем применяется фильтр, позволяющий исключить облигации с купонным процентом выше 11%, т.к таких бумаг немного и они существенно искажают результаты модели. Нужно объяснить Это приводит к уменьшению набора данных до 460 облигаций. Далее из модели исключались бумаги, которые имеют хоть одно пустое значение объясняющей переменной. Это необходимо для того, чтобы в данных, загружаемых в модель, не осталось пустых ячеек. Итого на выходе после всех фильтраций остаётся 446 бумаг.

Для разделения данных на обучающую и тестовую подгруппу мы принимаем отраслевую норму разделения 76/24, с 76% данных, разделенных на обучающий набор и 24% на прогнозный набор. Такой шаг оставляет большую часть данных для оптимизации алгоритмов, но при этом также оставляет существенный набор данных для последующего прогноза. Мы разделяем данные случайным образом, чтобы убедиться, что тестовый набор репрезентативен для общего распределения данных. Для проведения случайного сплита мы использовали тренировочный/тестовый сплит алгоритм из библиотеки Sci-Kit для иPython.

Итак, общую выборку мы делим на тестовую выборку, которую будем вводить в модель для тренировки и на выборку для прогноза. Таким образом тестовая выборка состоит из 339 бумаг и 7 458 уникальных значений. В данной выборке присутствуют 18 стран из 37 отраслей и 96 эмитентов с 11 разными рейтингами. Выборка, которую мы будем использовать для прогноза ставки купона состоит из 107 различных бумаг и 2 354 уникальных значений. В данной выборке присутствуют 14 стран из 20 отраслей и 36 эмитентов с 10 разными рейтингами.

Используя надстройку в Microsoft Excel для Cbonds мы собираем основные данные для объясняющих переменных нашей выборки. Кроме того, активно используются открытые источники и финансовые результаты компаний. Наконец, аналогичным образом собираются данные ВВП, Consumer price index и инфляция, % с сайта Конференции Организации Объединенных Наций по торговле и развитию UnctadStat начиная с 30 декабря 2003 года.

# **2.4. Описательные данные о выпусках облигаций**

Следующий подпункт состоит из графиков, иллюстрирующих описательные данные по моей тестовой и прогнозной выборкам.

Помимо разделения на тестовую и прогнозную выборки, прогнозный набор данных из 107 наблюдений и тестовый набор данных из 339 наблюдений кажутся похожими и сопоставимыми по всем параметрам. Это значит, что мы корректно разграничили две выборки и в дальнейшем наша модель предскажет более точные результаты. Ниже приводится подробное описание и обсуждение всех рассматриваемых параметров.

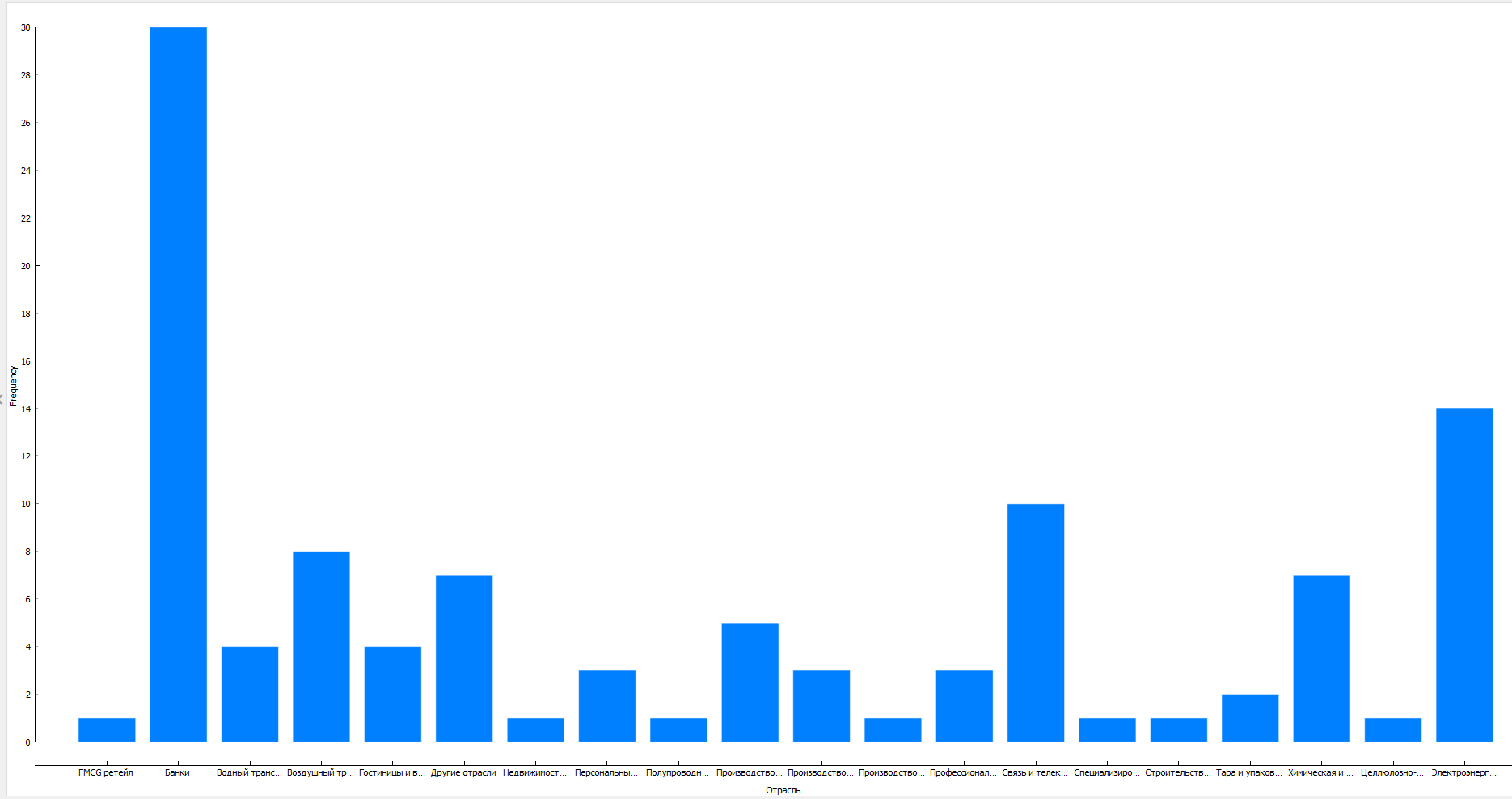
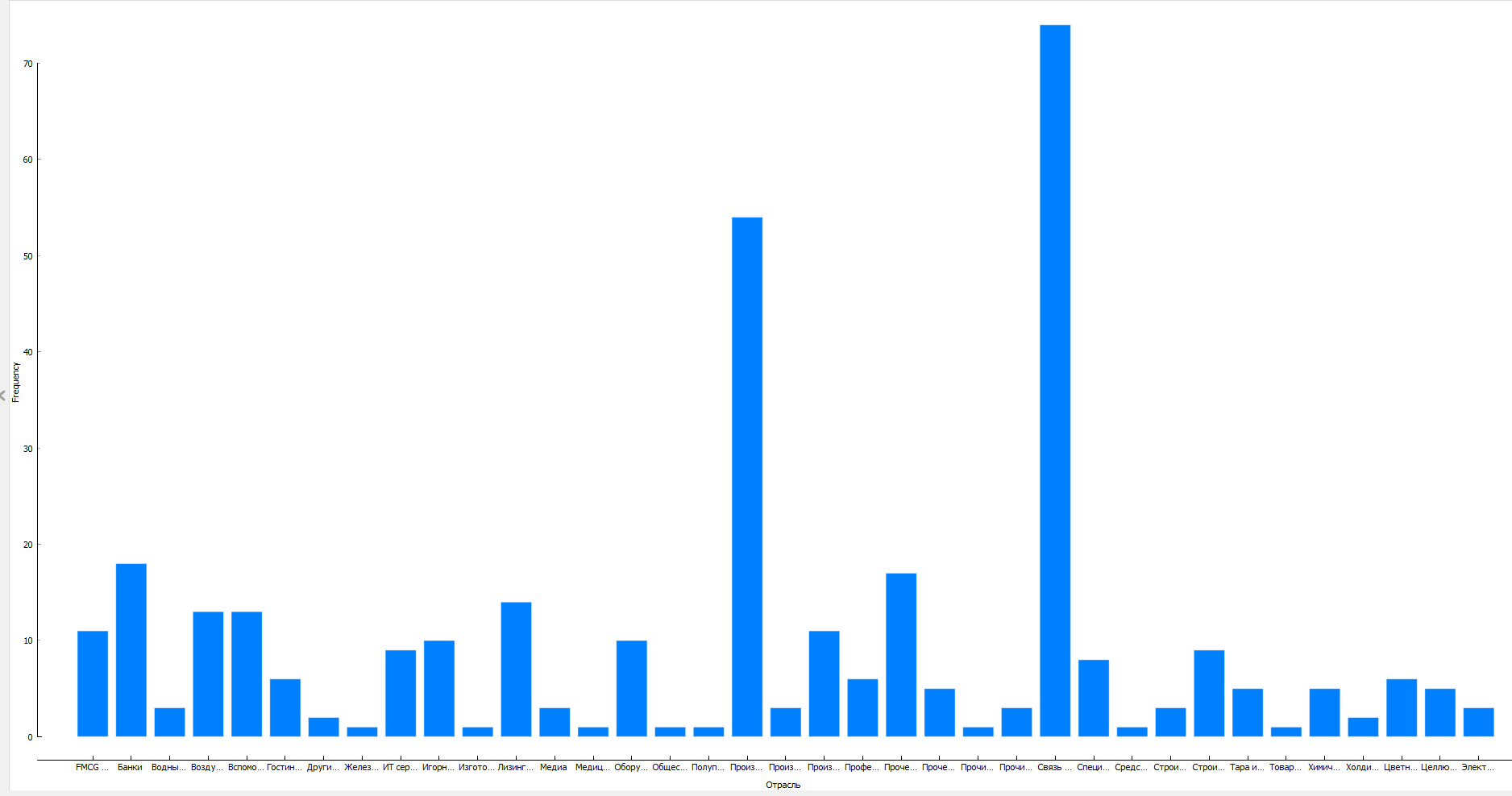


Рисунок 10. – Отраслевое распределение тестовая выборка vs прогнозная.

Cbonds классифицирует отрасль компании с помощью своей системы отраслевой классификации, которая была применена для отраслевой классификации рассматриваемых ценных бумаг. Чтобы иметь достаточное количество точек данных для каждой категории, компании были классифицированы на основе их отраслевой метки, которая является самым высоким доступным уровнем классификации.

Смотря на представленные отрасли, мы можем сделать вывод о том, что в целом распределение равномерное. Исключение составляют сектора «Связь и коммуникация» и «Производство автотранспорта» в тестовой выборке и сектора «Банки» и «Электроэнергетика» в прогнозной выборке. Данные сектора являются преобладающими в своей выборке, но средняя величина купона не превышает 4,5% в обеих выборках, что не может негативно сказаться на результатах нашей модели. Всего в тестовой выборке представлено 37 различных отраслей, в прогнозной выборке – 20 различных отраслей.

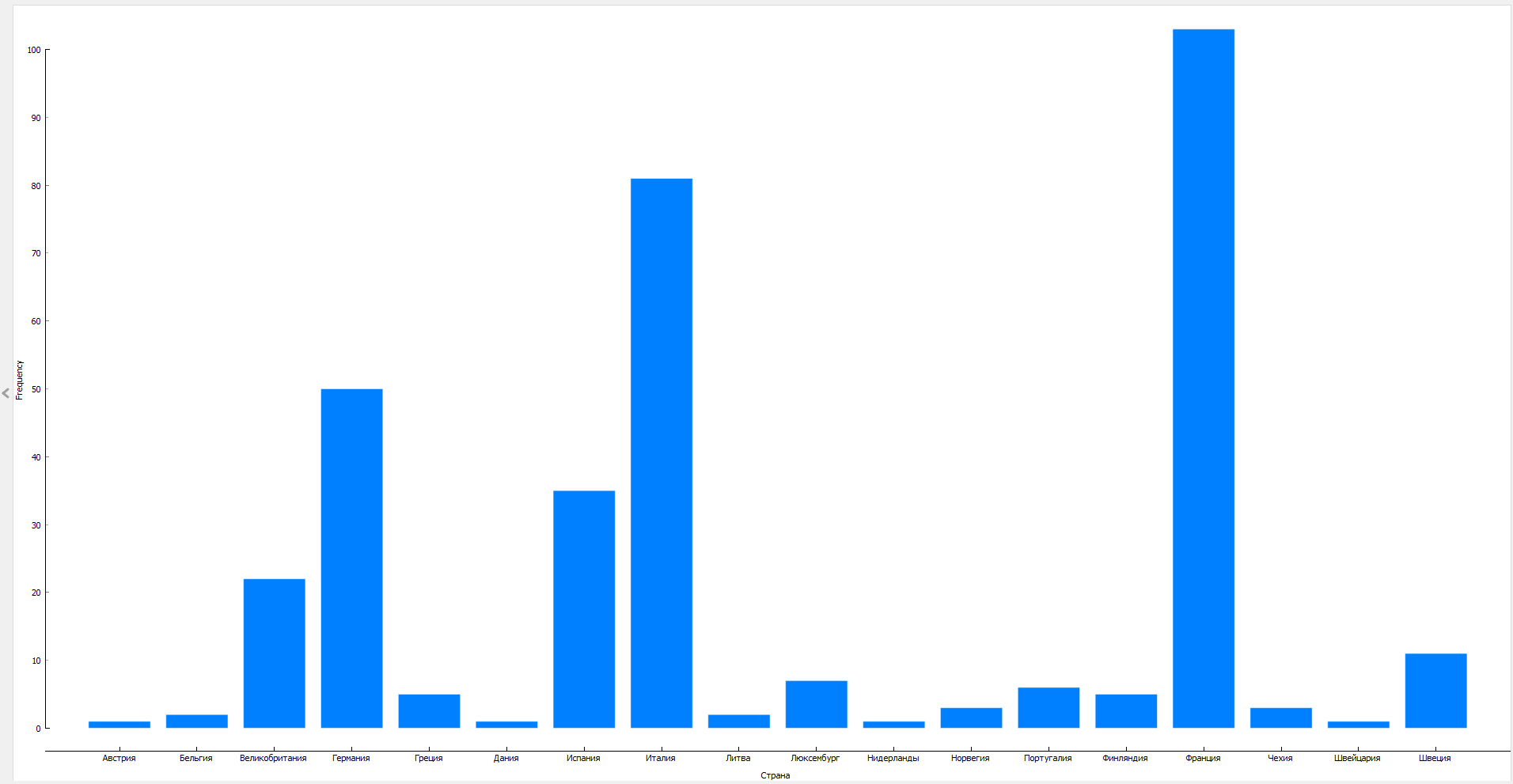
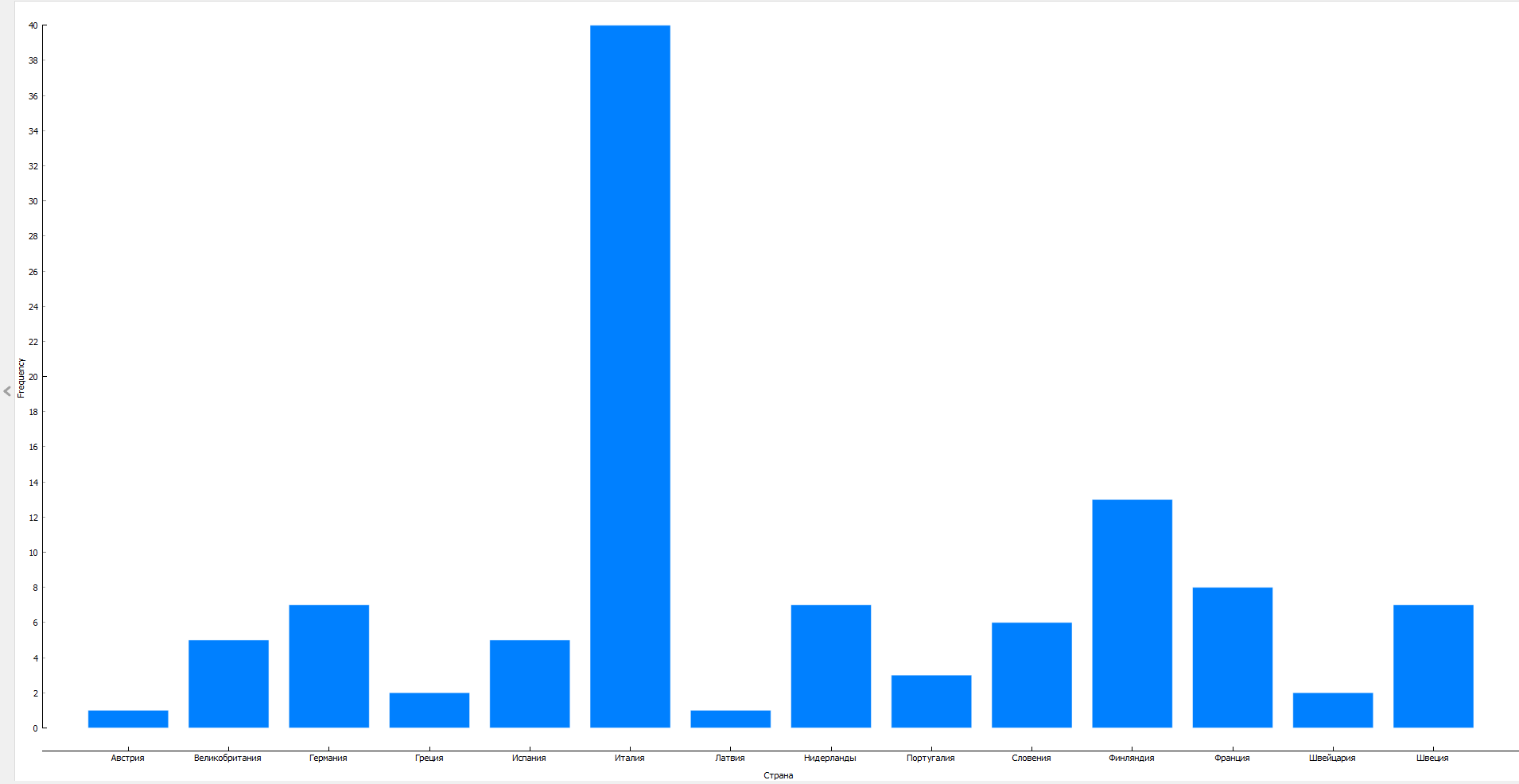
 

Рисунок 11. – Страновое распределение тестовая выборка vs прогнозная.

Фирмы, работающие в Европе, с большей вероятностью будут выпускать долговые обязательства на европейских финансовых рынках, и обзор стран является свидетельством этого. Указанная страна — это страна, в которой создана компания, лежащая в основе ценной бумаги, что объясняет, почему могут быть представлены страны за пределами Европы. Однако как для прогнозного, так и для тестового набора данных основными компонентами набора данных являются европейские страны, в частности Германия, Франция, Италия и Испания.

Одно потенциальное смещение, возникающее при включении всех ценных бумаг, выпущенных в данной области, заключается в том, что ценные бумаги компаний, не являющихся местными в данной области, будут, как правило, касаться крупных транснациональных фирм, поскольку компании, которые работают только на своем местном рынке, вряд ли будут искать финансирование на иностранных финансовых рынках. Примером этого является представительство таких стран, как Швейцария, которая представлена облигациями, выпущенными Dufry, швейцарским туристическим ритейлером, который управляет магазинами беспошлинной торговли, а также магазинами товаров повседневного спроса в аэропортах, на круизных линиях, в морских портах, на вокзалах и в центральных туристических зонах; Норвегия, которая представлена облигациями, выпущенными Adevinta, глобальным специалистом по объявлениям, занимающем лидирующие позиции на ключевых европейских рынках и Великобританией, которая представлена облигациями, выпущенными International Game Technology, мировым лидером в области игр, Playtech — крупнейшим в мире поставщиком программного обеспечения для онлайн-игр и ставок на спорт, а также Rolls-Royce — британский производитель автомобилей класса люкс, который с 2003 года является дочерней компанией BMW AG, являясь эксклюзивным производителем автомобилей марки Rolls-Royce. Другими словами, хотя набор данных содержит компании среднего размера из европейских стран, маловероятно, что этот баланс сохраняется в разных странах.

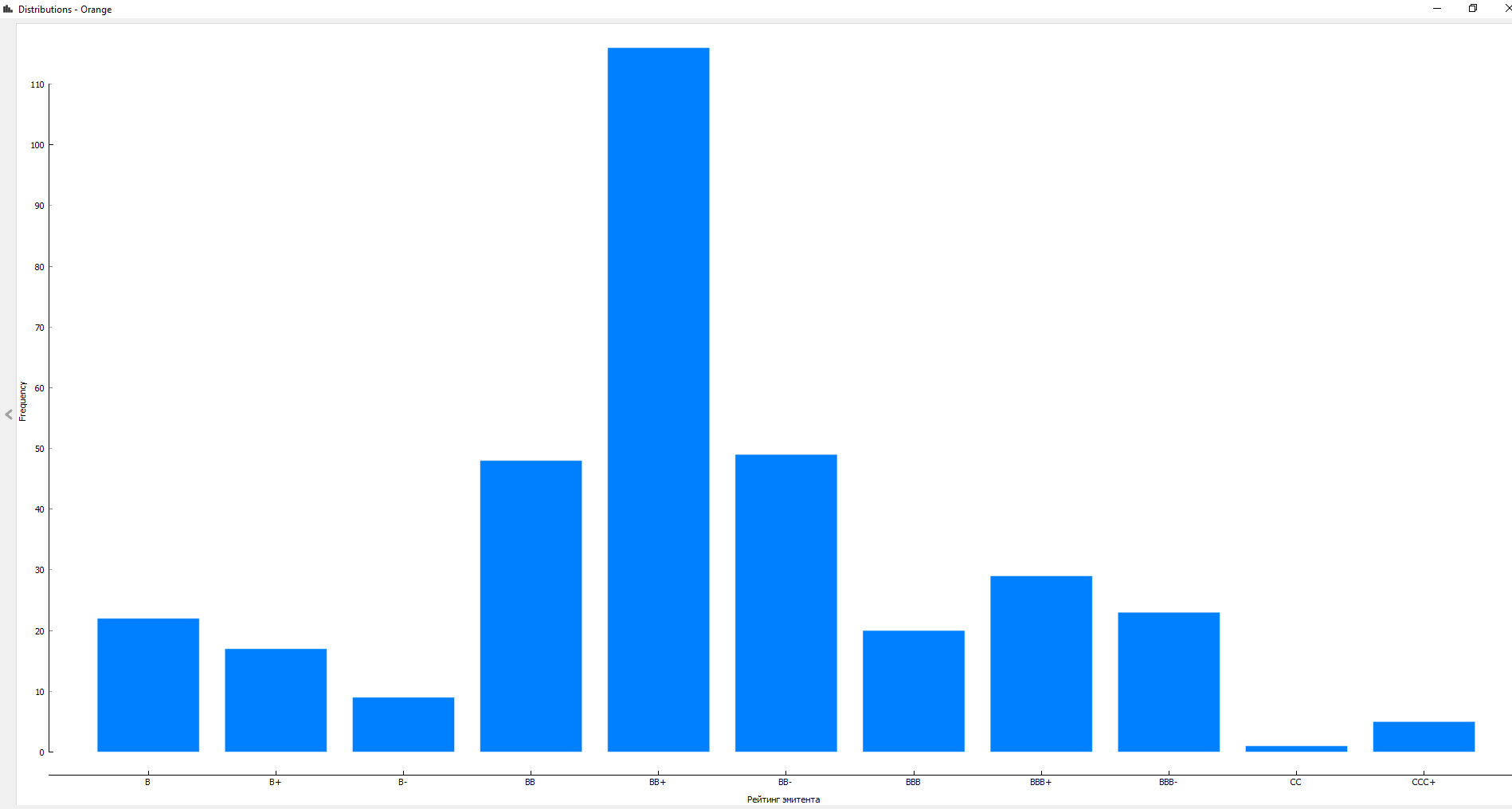
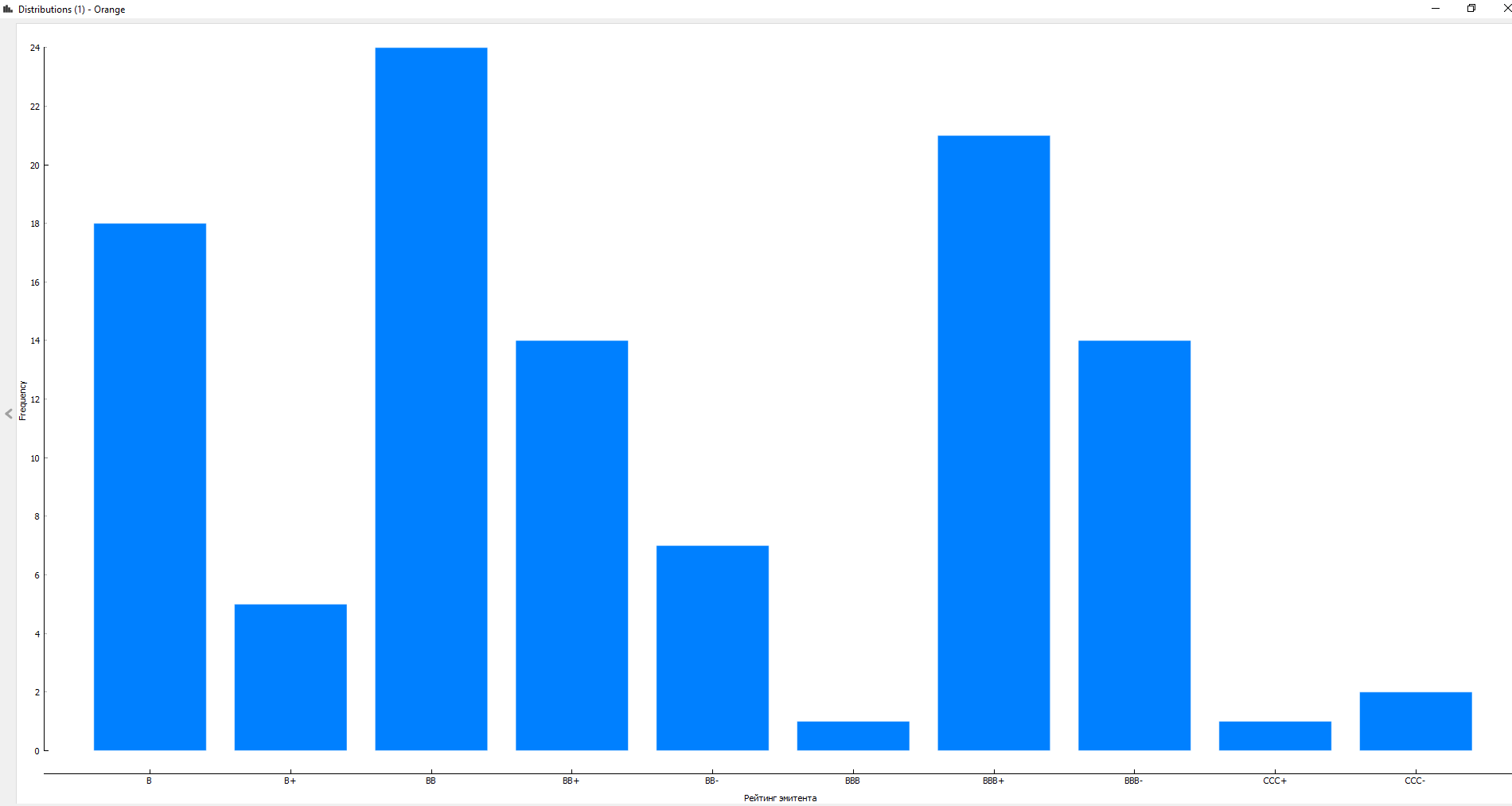
 

Рисунок 12. – Распределение по рейтингу эмитента тестовая выборка vs прогнозная.

Рейтинговые агентства, такие как Standard & Poors, Moodys и Fitch, обычно оценивают ценную бумагу до ее размещения, хотя некоторым частным выпускам рейтинг присваивается необязательно. Так как я использовала в своей выборке не только торгующиеся в данный момент бумаги, но и уже погашенные, соответственно половина рейтингов погашенных бумаг отсутствовала. В связи с этим мне пришлось взять для рассмотрения рейтинг эмитента. Каждый рейтинг компании был получен с сайта Cbonds. Я рассматривала только оценки S&P Rating в диапазоне от CCC до BB+. Высокодоходные облигации без кредитных рейтингов были исключены из нашей выборки. Как и следовало ожидать, оба набора данных состоят в основном из ценных бумаг с рейтингом от BBB- до BB.

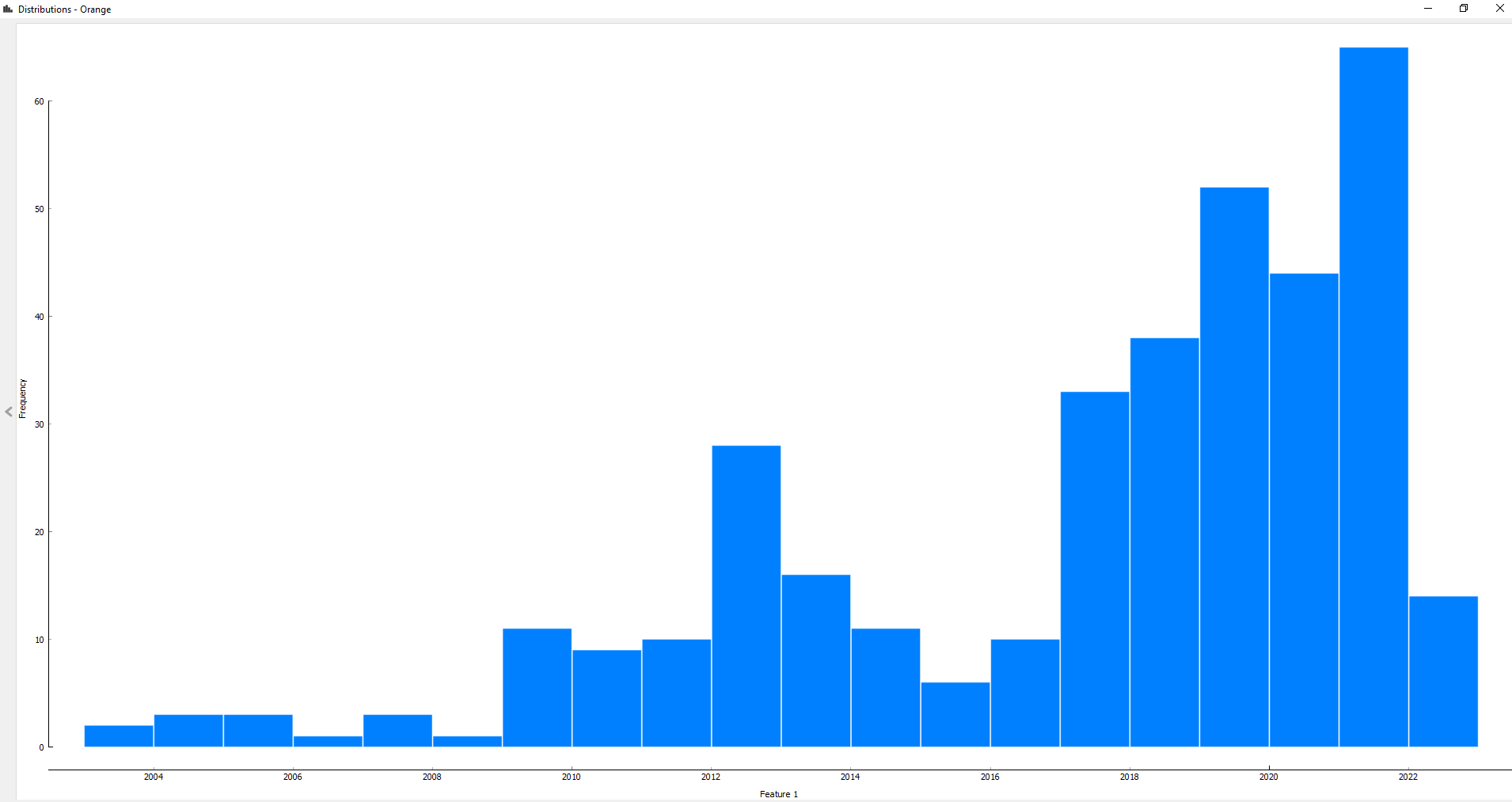
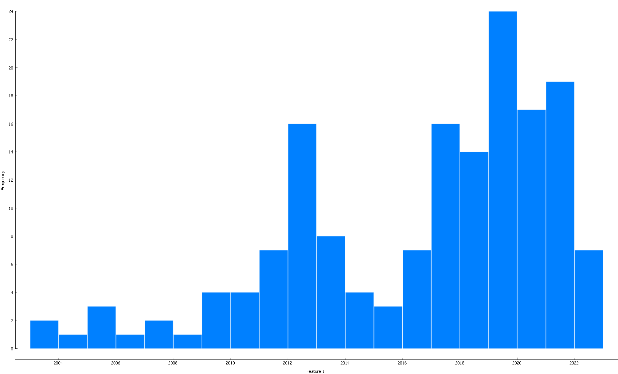
 

Рисунок 13. – Распределение по годам тестовая выборка vs прогнозная.

В моих выборках присутствуют облигации выпущены в период с 2003 по 2023 год, причем наиболее распространенным наблюдением является с 2017 по 2022 год. Прогнозный набор данных несколько больше смещен в сторону более старых выпусков, чем тестовый набор данных, и содержит больше выпусков, датированных ранее в 2012 году.

Определенный процент облигаций будет отозван или объявит дефолт в любой данный период, что означает, что они больше не подлежат обращению на рынке, и по этой причине можно было бы ожидать, учитывая фиксированное количество выпусков в год, что количество облигаций, торгуемых в настоящее время, будет больше в течение нескольких лет ближе к настоящее. Однако выпуск облигаций носит весьма циклический характер (AFME, 2021), и поэтому разница в годах выпуска, наблюдаемая в наборе данных, сама по себе не вызывает тревоги.

Изображение выглядит как диаграмма

Автоматически созданное описание Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 14. – Распределение дюрации тестовая выборка vs прогнозная.

Для обоих наборов данных большинство облигаций обеспечено средним сроком погашения от 4 до 11 лет. Конечно, в двух выборках присутствуют бумаги со сроком погашения более 50 лет, но они составляют менее одного процента от всех бумаг, соответственно не могут повлиять негативно на результаты.

Изображение выглядит как График, диаграмма, снимок экрана, линия

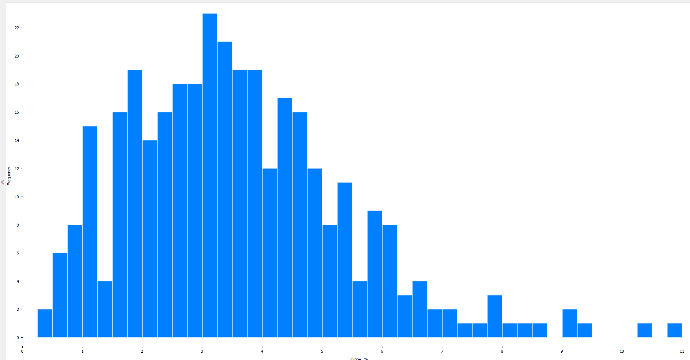
Автоматически созданное описание

Рисунок 15. – Распределение купонной ставки тестовая выборка vs прогнозная.

Наибольшее количество бумаг в тестовой выборке имеют ставку купона в пределах от 1.5 – 5,5%. В прогнозной выборке превалирующий диапазон 1,5 – 4,5%.

Изображение выглядит как диаграмма

Автоматически созданное описаниеИзображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 16. – Распределение объема размещения тестовая выборка vs прогнозная.

Средний размер размещения облигаций в тестовом наборе данных составляет от 300 млн евро до 1 млрд евро против 50 – 600 млн евро для прогнозного набора данных с сопоставимым распределением.

# **2.5. Корреляционный анализ детерминант**

После того, как все объясняющие переменные были ручным образом собраны и рассчитаны, мы проводим корреляционный анализ, чтобы обеспечить идеальную мультиколлинеарность. Если некоторые из объясняющих объектов отражают схожие или зависимые экономические показатели, они дадут очень высокие показатели корреляции, и было бы лучше исключить один из объектов из модели.



Рисунок 17. – Корреляционная матрица тестовой выборки.

На рисунке 17 показана корреляционная матрица вычисленной попарной корреляции между объясняющими переменными. Среди различных объектов корреляция близка к -1 только в одном случае: в паре «Индекс коррупции» и «Место в рейтинге Doing Business». Корреляция достаточно сильная и отрицательная, т.к индекс коррупции имеет обратное значение, т.е чем он выше, тем меньше коррупции в стране. Соответственно, чем выше номинальное значение индекса коррупции, тем быстрее к первому месту в рейтинге Doing Business стремится страна. Тем не менее, не смотря на достаточно высокое значение корреляции между этими двумя переменными, я оставляю в модели обе переменные. Это связано с тем, что индекс коррупции в целом измеряет только уровень восприятия коррупции в государственном секторе той или иной страны. В это же время, Рейтинг Doing Business оценивает экономическое положение страны по десяти критериям. Соответственно, место в рейтинге «Doing Business» включает в себя огромное количество различных факторов, что достаточно полно отражает некоторые аспекты экономической деятельности той или иной страны. Таким образом, обе объясняющие переменные было решено оставить для нашей модели. Также стоит прокомментировать ещё несколько коэффициентов корреляции.

Положительный коэффициент равный 0,56 между годовой инфляцией и ежегодным ростом индекса потребительских цен означает, что более высокий показатель инфляции приводит к росту индекса потребительских цен. Это соответствует тому, что мы ожидали, поскольку теоретически должно быть совпадение между этими двумя показателями, т.к так или иначе ИПЦ является одним из самых популярных показателей инфляции и дефляции. Но т.к корреляционная связь между этими двумя показателями некритичная, то принято решение оставить два показателя в модели.

Далее коэффициент 0,37 попарной корреляции между кредитным плечом и структурой капитала. Это обусловлено тем, что в двух коэффициентах при расчете присутствует чистый долг. Коэффициент корреляции незначительный, значит оставляем обе переменные в модели.

Что касается остальных переменных, то зависимость между ними не превышает 0,3 соответственно в остальных случаях исключаются случаи идеальной мультиколлинеарности.

# **ГЛАВА 3. РЕЗУЛЬТАТЫ ВЫЯВЛЕНИЯ ДЕТЕРМИНАНТ ДОХОДНОСТИ ЕВРОПЕЙСКОГО РЫНКА ВЫСОКОДОХОДНЫХ ОБЛИГАЦИЙ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ**

Собрав и обработав все данные, мы приступаем к фактической настройке нашей модели. В этой главе будет описана используемая модель и проведён прогноз величины купона для прогнозной выборки с использованием программы для машинного обучения Orange.

Orange — это набор инструментов для визуализации данных, машинного обучения и интеллектуального анализа данных с открытым исходным кодом. Он имеет интерфейс визуального программирования для исследовательского качественного анализа данных и интерактивной визуализации данных.

Оранжевые компоненты называются виджетами. Они варьируются от простой визуализации данных, выбора подмножества и предварительной обработки до эмпирической оценки алгоритмов обучения и прогнозного моделирования.

При написании дипломной работы мой выбор пал на эту программу, т.к она имеет достаточно простой интерфейс, но в тоже время помогает делать существенные операции, для которых в обычной ситуации пригодился бы сложный код на языке Python. Но т.к я не владею данным языком программирования в совершенстве и не смогу написать большой и сложный код, я остановила свой выбор на Orange.

Перед выполнением работы передо мной была поставлена четкая цель максимально точно и качественно предсказать купонную ставку для прогнозного набора данных с помощью машинного обучения и программы Orange. В нашем случае интеллектуальный анализ данных и прогнозирование помогут не только потенциальным эмитентам корректно оценить выпускаемую бумагу, но и потенциальным участникам рынка - покупателям. Так эмитенты смогут корректно оценить предполагаемую ставку купона, является ли она реальной, отвечающей текущему состоянию копании – эмитента и общей экономической ситуации в стране эмитента или она должна быть наоборот увеличена или понижена. Также интеллектуальный анализ данных поможет эмитенту объективно выявить факторы, по которым купонная ставка завышена или, наоборот, занижена.

Что касается покупателей, участников рынка, то здесь интеллектуальный анализ данных поможет понять, реальна ли купонная ставка для данной бумаги или она необоснованно завышена.

# **3.1. Подготовка к настройке модели**

После сбора данных мы получили следующие визуальные данные, которые могут помочь нам лучше понять, какие факторы влияют на купонную ставку бумаги и на какие характеристики следует обращать внимание, когда вы хотите приобрести в свой портфель или проанализировать какую-либо бумагу (Приложение 1).

Загрузив документ в систему мне необходимо определить таргет, числовые и текстовые переменные (Приложение 2).

Как только я обозначила системе таргет и типы данных мне нужно было проанализировать какие факторы в наибольшей степени влияют на таргетируемую переменную, а какие не влияют вообще и могут быть исключены из модели для более быстрого и чистого анализа и работы системы. Чтобы понять, какие переменные следует учитывать в модели, я построила в Orange виджет Rank, откуда я могу видеть, какие факторы и насколько сильно влияют на мою целевую переменную (Ставка купона)(Приложение 3).

Виджет «Rank» оценивает переменные в соответствии с их корреляцией с дискретной или числовой целевой переменной на основе применимых внутренних показателей (таких как прирост информации, хи-квадрат и линейная регрессия) и любых подключенных внешних моделей.

В данном виджете использовались следующие методы оценки (классификации):

* прирост информации (ожидаемое количество информации (уменьшение энтропии));
* Gain Ratio (соотношение прироста информации и внутренней информации атрибута, которое уменьшает смещение в сторону многозначных признаков, возникающее при получении информации);
* Джини (неравенство между значениями частотного распределения);
* хи-квадрат (зависимость между функцией и классом, измеренная статистикой хи-квадрат);
* ReliefF (способность атрибута различать классы в похожих экземплярах данных).

Основываясь на перечисленных выше методах оценки переменных мы можем сделать вывод о том, что во всех рассматриваемых моделях топ восемь влияющих на таргетированную переменную факторов составляют: годовая инфляция, эмитент, отрасль, месяц выпуска, страна, рейтинг эмитента, место в рейтинге Doing Business и индекс коррупции. Также метод оценки ReliefF в отличии от других методов выделяет ещё влияние цены первичного размещения и тип облигации.

В целом все выбранные для моей модели факторы показали некоторое влияние на таргетированную переменную, поэтому было решено оставить все объясняющие переменные в модели для более точного дальнейшего предсказания купонных ставок в прогнозной выборке.

После разделения набора данных и определения показателей оценки следующим шагом будет настройка структуры и алгоритмов регрессии машинного обучения.

Регрессия машинного обучения может быть выполнена с помощью множества различных алгоритмов, но, как утверждает теорема Макреди и Вольперта еще в 1997 г., в машинном обучении не бывает бесплатных обедов. Будут неотъемлемые компромиссы для каждой модели и алгоритма, и ни одна модель не является универсально лучшей. Производительность разных алгоритмов машинного обучения часто зависит от размера и структуры базовых данных (Daumé, 2017). Эти структуры часто скрыты, в том смысле, что они не являются структурами, легко воспринимаемыми человеческим разумом. В результате правильный выбор алгоритма и настройка модели будут часто оставаться неясными до тех пор, пока модели не будут протестированы и оценены (Géron, 2017).

# **3.2. Настройка и выбор наиболее эффективной модели**

После того, как мы определились со всеми данными для модели, определили влияние детерминант, мы можем приступать к выбору оптимальной и точной модели для нашей прогнозной выборки.

Итак, когда я, наконец, определилась с переменными для модели и фактической целью модели, я начала думать о том, какие модели я могу использовать в своих исследованиях. В первую очередь хотелось попробовать множество моделей и выбрать из них лучшую. Поэтому я попробовал kNN, SVM, Neural Network, Gradient Boosting, Tree, CN2 Rule Induction, Random Forest, Naïve Bayes, Logistic Regression и AdaBoost (Приложение 4).

В программе Orange я пробую различные модели следующим образом. Как уже говорилось ранее, мы загрузили файл в значок «Файл для модели» и определили тип данных, если это не произошло корректно в автоматическом режиме. В качестве таргетируемой переменной я выбрала Купонную ставку, исключила из выборки Бумага — это столбец с названием бумаг, из дюрацией и купонной ставкой, соответственно система выбирает в этом названии ставку купона и предсказывает результаты практически со 100% точностью, что полностью не отвечает цели и задачам нашего исследования. Поэтому данная переменная исключается из модели.

На следующем этапе все модели проходят через виджет Test and Score. Test and Score проверяет алгоритмы обучения на данных и выдает результаты тестирования алгоритмов классификации. Виджет тестирует алгоритмы обучения. Доступны различные схемы выборки, в том числе с использованием отдельных тестовых данных. Виджет делает две вещи. Во-первых, он показывает таблицу с различными показателями производительности классификатора, такими как точность классификации и площадь под кривой. Во-вторых, он выводит результаты оценки, которые могут использоваться другими виджетами для анализа производительности классификаторов, таких как анализ ROC или матрица путаницы.

Виджет поддерживает различные методы выборки. Для своей работы я выбрала метод «Тестирование на тренировочном наборе данных». Тестирование на тренировочном наборе данных использует весь набор данных для обучения, а затем для тестирования.

Все алгоритмы машинного обучения содержат параметры, которые задаются вручную перед обучением модели начинаются. Чтобы отличить их от параметров, на которые влияют данные, они называются гиперпараметры. Поскольку они могут сильно повлиять на производительность моделей, поэтому необходимо выполнить точную настройку гиперпараметров для каждой модели.

Чтобы оценить результаты конфигураций гиперпараметров мы будем использовать меру R2, поскольку именно она предоставляет информацию о производительности модели, которую лучше всего сравнивать между моделями.

После того, как виджет Test and Score отработал, появляется классификация моделей (Рисунок 18).

Далее я построила регрессионную модель в Orange. Результаты были следующими.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 18. – Результаты регрессии выбранных моделей.

Для оценки моделей мы применяем следующие четыре метрики оценки: Средние абсолютные ошибки (MAE), среднеквадратичные ошибки (RMSE), R-квадрат (R2) и MSE. Каждая из них, по сути, указывает, насколько хорошо модель предсказала набор данных, но каждая из них обеспечивает различные нюансы информации о том, насколько хорошо модель предсказала точки данных (Casella et al, 2017).

MSE измеряет среднее значение квадратов ошибок или отклонений (разница между оценщиком и тем, что оценивается).

RMSE - это квадратный корень из среднего арифметического квадратов набора чисел (мера несовершенства соответствия оценки данным). RMSE похожа на MAE, поскольку также выражает среднюю модель ошибки предсказания. Но вместо того, чтобы просто взять абсолютное значение парных ошибок, RMSE возведёт в квадрат ошибки, прежде чем брать среднее значение. В результате она придает больший вес большой ошибке, и таким образом лучше для обнаружения выбросов и для обнаружения больших ошибок, вызванных моделью (Stock & Watson,2015).

MAE используется для измерения того, насколько близки прогнозы или предсказания к возможным результатам.

R2 интерпретируется как доля дисперсии зависимой переменной, которую можно предсказать по независимой переменной. Проблема с использованием R2 заключается в том факте, что добавление новых объясняющих переменных всегда увеличивает показатель, независимо от того, обладают ли они фактической объясняющей силой. Поскольку в исследовании используется множество переменных, важно оценивать модели с помощью соответствующих показателей. R2 отличается от остальных показателей, поскольку вместо измерения средней ошибки он описывает, насколько хорошо модель отражает лежащие в основе дисперсии зависимой переменной. Если зависимая переменная имеет очень высокую дисперсию, но модель также очень хорошо предсказывает дисперсию, она будет иметь высокий показатель R2, но модель все еще может возвращать высокие значения MAE и RMSE из-за лежащей в её основе высокой дисперсии зависимой переменной. Это делает R2 лучшим показателем для сравнения наборов данных.

Число в таблице показывает вероятность того, что модель, соответствующая строке, имеет более высокий балл, чем модель, соответствующая столбцу. Что означает более высокая оценка, зависит от метрики: более высокая оценка может означать либо лучшую модель (например, CA или AUC), либо наоборот (например, RMSE). Если включена незначительная разница, меньшее число ниже показывает вероятность того, что разница между парой незначительна. Тест основан на байесовской интерпретации t-критерия (краткое введение).

Я пробовал разные настройки как в моделях, так и в Test and Score, но R2 не менялся, а если и менялся, то в пределах 0,001. При повторе или смене метода выборки коэффициенты увеличивались, результаты ухудшились. Таким образом, многие из вышеперечисленных в начале главы моделей не работали с моим набором данных или были слишком низкого качества. Поэтому в итоге я решила сосредоточиться на следующих трех моделях: Neural Network, Gradient Boosting (Gradient Boosting — это метод машинного обучения для задач регрессии и классификации, который создает модель прогнозирования в виде ансамбля слабых моделей прогнозирования, обычно деревьев решений) и Random Forest (Случайный лес — это метод ансамблевого обучения, используемый для классификации, регрессии и других задач). Эти три модели показали следующие результаты.

Основываясь на результатах выше R2, мы можем сделать вывод, что лучшей моделью является Neural Network, за которой следуют Gradient Boosting и Random Forest.

Для дальнейшего прогноза я использовала Neural Network. Алгоритм многослойного персептрона (MLP) с обратным распространением. Виджет Neural Network использует алгоритм многослойного персептрона sklearn, который может изучать нелинейные модели так же, как и линейные.

Нейронная сеть использует предварительную обработку по умолчанию, когда другие препроцессоры не указаны. Она выполняет их в следующем порядке:

* удаляет экземпляры с неизвестными целевыми значениями
* продолжает категориальные переменные (с горячим кодированием)
* удаляет пустые столбцы
* вменяет пропущенные значения со средними значениями
* нормализует данные, центрируя их по среднему значению и масштабируя до стандартного отклонения 1.

# **3.3. Прогнозирование купонной ставки ВДО**

После того, как я попробовала все модели через Test and Score и «натренировала» мою систему, я перешла к тестированию модели на тестовых данных. Для этого я подключила Save Model к модели Neural Network (Приложение 5). Я сохранила модель в формате .pkcls и потом работала с ними во время теста (Приложение 6).

Для корректной работы предсказания я для начала подгрузила прогнозную часть выборки в файловый виджет Test file. Далее я подгрузила ранее выгруженную модель в формате .pkcls в виджет Predictions, который показывает прогнозы моделей по данным. Виджет получает набор данных и один или несколько предикторов (предиктивные модели). Он выводит данные и прогнозы.

Виджет Predictions показывает вероятности и окончательные решения прогностических моделей. Результатом виджета является другой набор данных, к которому добавляются прогнозы в виде новых метаатрибутов. Вы можете выбрать, какие функции вы хотите вывести (исходные данные, прогнозы, вероятности). Результат можно наблюдать в таблице данных. Если прогнозируемые данные включают истинные значения класса, результат прогнозирования также можно наблюдать в матрице путаницы (Приложение 7).

# **3.4. Обсуждение результатов прогноза купонной ставки ВДО с использованием ML**

В конце прошлого пункта мы увидели предсказание купонной ставки прогнозных данных в виджете Predictions в Orange. Я перенесла эти данные в excel и сопоставила их с реальными ставками данных бумаг. В целом по всей выборке среднее отклонение прогнозных значений составило 0,87 единиц. Соответственно учитывая величину средней ставки купона в нашей прогнозной выборке 3,43% можем сделать вывод, что наша модель ошиблась в 25% случаев из 100%, что свидетельствует о том, что мы подобрали правильные объясняющие переменные и грамотно настроили модели. Как-то можно это показать?

Для более детального анализа и дальнейших выводов, я разобью прогнозную выборку на группы по величине купона (Таблица 2).

Таблица 2. – Отклонение прогнозных значений от реальных по секторам.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| %  % | <1 | 1≤x<2 | 2≤x<3 | 3≤x<4 | 4≤x<5 | 5≤x<6 | 6≤x<7 | ≥7 |
| Величина ошибки в выборке | 67 | 45 | 20 | 16 | 12 | 18 | 29 | 21 |

Итак, у меня получилось семь диапазонов купонной ставки. Я сделала данную сегментацию для того, чтобы понимать в каких диапазонах модель делает самые большие отклонения от реальной ставки купона. В Таблице 2 представлены данные, на основе которых можно сделать вывод о том, что самые большие расхождения между прогнозом и реальной величиной купоной ставки наблюдаются в диапазоне от 0 до 2 %. Здесь модель ошибается практически в каждой второй бумаге. Данный результат можно объяснить тем, что в как в изначальной тестовой, так и в прогнозной моделях было очень мало бумаг с подобными ставками купона, соответственно модель просто не смогла корректно научиться на слишком маленьких или на слишком больших ставках, как например в бумагах с величиной ставки от 6 до 7 % (здесь предсказание сработало с точностью 70%. Что касается бумаг с величиной ставки от 2 до 6%, здесь показатели довольно оптимистичные, так в среднем отклонение прогнозных значений от реальных показателей составило не более 17%. Это обусловлено тем, что тестовая выборка в большинстве своём состоит преимущественно из бумаг со схожими купонными ставками. Поэтому можем уверенно сказать, что основную цель исследования мы выполнили и получили отклонения достаточно небольшие для выборки всего 25%. В будущем можно доработать перечень новыми объясняющими переменными для точности модели.

В разрезе стран самые точные предсказания оказались по таким странам как Испания, Италия, Франция и Финляндия. Такой хороший результат для первых трёх стран объясняется тем, что они достаточно часто встречаются в тестовой выборке. Что касается Финляндии, то купонный диапазон представленных в прогнозной выборке бумаг составляет от 1,7 до 4,2%, что, как мы сказали ранее, является достаточно распространённым диапазоном для тестовой выборки, соответственно модель смогла научиться и построить для себя сильные связи, чтобы в дальнейшем как можно точно предсказать купонные ставки на прогнозном диапазоне.

С точки зрения надежности результатов, чувствительность к разделению обучения/тестирования наблюдалась во время изучения. Согласно Daumé (2017), это характерно для небольших наборов данных. Таким образом, тиражирование исследования с большим количеством точек данных было бы эффективным способом минимизировать риск такой чувствительности и повысить общую надежность результатов. В частности, для рассматриваемой доходности сумма данных и качество данных привели к тому, что окончательный набор данных был маленьким. Это ограничивает уверенность в любом выводе, сделанном на основе результатов набора данных.

# **3.5. Значение исследования для научных кругов и участников рынка ВДО**

В этом пункте будет обсуждаться значение результатов, полученных в этой статье, как для литературы, связанной с высокодоходными облигациями, так и для исследователей, включая оценку того, как аналитическая база, представленная в исследовании, может быть полезна для участников рынка ВДО.

Обсуждение будет разделено на три части: краткий обзор всего аналитического процесса, включая то, как это может быть реализовано на практике, раздел о значении моего исследования для научных кругов и раздел о значении исследования для исследователей.

В Приложении 1 показан полный аналитический процесс окончательного стрима в Orange для прогнозирования купонной ставки высокодоходных облигаций. Сначала собираются все данные. Здесь информация о конкретной облигации, а также бухгалтерские данные для фирмы-эмитента собираются из Cbonds, некоторые показатели общего экономического развития страны, как ВВП, рост индекса потребительских цен берутся из Unctad Stat. Далее данные обрабатываются, поэтому они готовы к анализу. Для количественных данных мы конвертируем все к единой валюте EUR и масштабу, удаляются выбросы и т. д. Для текстовых данных текст обрабатывается с помощью шагов предварительной обработки текста. После этого особенности на основе данных вычисляются.

Мы создали три разные модели машинного обучения. Затем мы вычисляем предварительные оценки этих моделей и настраиваем их для получения окончательного набора функций и настройки гиперпараметров для каждой из этих моделей. Когда настройки модели и набор функций были доработаны, модели затем обучаются на тестовых данных и каждая из четырех моделей машинного обучения оцениваются.

Следующим шагом, если эта структура будет дальше использоваться участниками рынка ВДО, будет включение лучшей из окончательных моделей в операционную фазу. Здесь модель может быть использована как инструмент скрининга или как часть инвестиционного решения. На этапе эксплуатации, когда на рынке должна появиться новая бумага, данные по новой эмиссии будут собираться и обрабатываться. Особенности будут вычислены и введены в окончательную модель, которая затем даст практикам прогноз купонной ставки облигации, который можно сравнить с предлагаемым эмитентом купоном. Это сравнение можно использовать в процессе принятия инвестиционных решений или в качестве инструмента проверки, помечая проблемы, в которых реальная купонная ставка сильно отличается от предсказания. Наконец, каждый новый прогноз и результат на этапе эксплуатации должны быть сохранены и добавлены в тестовый набор данных для дальнейшего повышения производительности модели.

Основным вкладом этой модели в общую литературу по купонным ставкам высокодоходных облигаций является разработка полного аналитического конвейера и модели, которые можно использовать для получения высокоточного предсказания купонной ставки ВДО.

Насколько известно авторам, это первая работа по прогнозированию купонных ставок ВДО с помощью модели машинного обучения, созданная в академической сфере высокодоходных облигаций в России. До этого по ВДО с использованием машинного обучения не прогнозировали даже спредов в нашей стране. Таким образом, в перспективе есть возможность прогнозирования других показателей ВДО путем принятия более продвинутых моделей с большими объемами вычислений, которые не полагаются на простые линейные отношения.

Модели машинного обучения могут предсказывать купонные ставки и спреды высокодоходных облигаций с помощью гораздо большего разнообразия входных данных признаков и от более сложных признаков, чем в предыдущих исследованиях, упомянутых в литературном обзоре. Полный аналитический конвейер обеспечивает модель, которую можно легко адаптировать в будущем.

Моё исследование также вносит свой вклад в литературу, создавая модель, которая позволяет исследовать купонную ставку ВДО через определённый перечень параметров, неиспользуемых для прогноза с использованием машинного обучения ранее. Наконец, аналитический конвейер, разработанный для прогнозирования купонов высокодоходных облигаций с помощью набора собранных данных, может быть обобщен для использования в областях, отличных от высокодоходных облигаций. Данное исследование можно легко применять для изучения ценообразования IPO, ценовых движений после финансовых отчетов, таких как годовые или ежеквартальные отчеты, как новости и иные факторы влияют на финансовые рынки и аналогичные исследования, направленные на объединение количественных данных для объяснения ценообразования на финансовых рынках. Ключевой проблемой, которая все еще остается актуальной для эффективного обобщения всего аналитического конвейера, является эффективное получение выбранных данных для обучения прогностических моделей.

# **3.6. Допущения и ограничения модели**

Одним из первых ограничений модели, как и большинства статистических моделей, является то, что она показывает только корреляцию, а не причинность. Несмотря на то, что мы обнаружили, что некоторые переменные имеют высокую объяснительную силу в предсказании купонной ставки ВДО, мы не можем быть уверены, что эти переменные вызывают распространение. Модель настроена с конечным числом функций, чтобы ограничить степени потери свободы, и чтобы размеры набора данных не взрывались. В результате сильная объяснительная сила некоторых переменных может быть просто связана с высокой корреляцией с третьей переменной, не включенной в модель, оказывающей фактическое причинно-следственное влияние на купонную ставку.

Модель по-прежнему может дать важную информацию о том, на какие экономические, облигационные и учетные функции финансовый аналитик, а также менеджеры компаний-эмитентов должны обратить внимание, но это не может подтвердить каких-либо причинно-следственных связей между ними и купонной ставкой высокодоходных облигаций.

Еще одно допущение модели, созданной в этой работе, заключается в том, что текущие и прошлые данные работают как предсказатель будущего. Предполагается, что купонные ставки высокодоходных облигаций отражают ожидания относительно будущего этой облигации и рынка в целом. Таким образом, вся модель построена на предположениях о том, что нынешние облигационные данные бумаги и эмитента, а также исторические данные бухгалтерского учета будут иметь некоторую способность предсказывать будущие события. Это было доказано исторически, т.к. данные бухгалтерского учета часто попадают в определенный диапазон данных бухгалтерского учета за предыдущие годы. Но это предположение может быть верным только в нормальных ситуационных, экономических и политических обстоятельствах, кризис Covid-19 и кризис, связанный с СВО на Украине, доказали это. Кризисы привели к тому, что многие исторические показатели, которые обычно были сильными предикторами будущих результатов, оказались совершенно неуместными предикторами.

Моя модель, специально разработанная для данной дипломной работы, хороша настолько, насколько хороши данные, на которых она построена. Мне пришлось ограничить количество функций и сложность функций из-за ограниченного размера конечного набор данных, так как мы не хотели терять много степеней свободы. Большой набор данных позволит построить более сложный набор функций.

# **3.7. Предложения для дальнейших исследований**

Выводы, сделанные в моей работе в сочетании с предположениями и ограничениями, рассмотренными выше, поднимают несколько интересных и актуальных действий и перспектив для дальнейших исследований.

Во-первых, в дальнейшем мне было бы интересно запустить такой же аналитический конвейер для более расширенной выборки европейских высокодоходных облигаций и сравнить результаты, чтобы убедиться, что результаты также согласуются с очень большим набором данных.

Во-вторых, было бы очень интересно в следующем исследовании применить аналитическую основу для изучения высокодоходных облигаций рынка США. Было бы интересно сравнить результаты прогнозирования купонных ставок высокодоходных облигаций США с выводами этой статьи.

В-третьих, в этой статье делается попытка только предсказать и объяснить обычную купонную ставку высокодоходных облигаций. Однако, нормальные различия и изменения в части купонной ставки ВДО становятся непредсказуемыми во время наступления экстремальных ситуаций, таких как Covid-19 и кризис, связанный с СВО на Украине. Еще один следующий шаг исследования может заключаться в использовании модели, разработанной в моей работе, чтобы попытаться объяснить непредсказуемое движение купонных ставок или спредов отдельных высокодоходных облигаций в условиях экстремального кризиса, такого как пузырь доткомов 2001 г.; 2008 г. финансовый кризис; Covid-19; кризис, связанный СВО на Украине. Это было бы интересно для исследования высокодоходных облигаций, так как многие потери, связанные с высокодоходными облигации в новое время вызваны такими событиями.

Цель моей дипломной работы состояла не только в том, чтобы создать академическую модель с академическими последствиями, но и в том, чтобы создать модель, которую можно использовать на практике.

Первый очевидный вывод для практиков заключается в том, что некоторые экономические показатели страны очень сильно влияют на купонную ставку. Поэтому они должны быть включены в процесс принятия инвестиционных решений. Кроме того, если модель, разработанная в моей работе, была идеально настроена со всеми включенными соответствующими функциями, инвестиционный процесс может быть автоматизирован. Это можно наблюдать в мире акций, где большая часть торговли и инвестиционный процесс теперь осуществляются с помощью алгоритмов (Frasincar et al, 2013). Тем не менее, рынок высокодоходных облигаций ограничен небольшими объемами данных, что являлось самым большим ограничением в моей работе.

Кроме того, модель можно использовать на этапе скрининга инвестиционного процесса. Если прогнозируется, что облигация будет торговаться со значительной купонной ставкой, выше ставки, предлагаемой на рынке, данную бумагу можно пометить как заслуживающую внимания. То же самое относится к облигациям, торгующимся с гораздо более низкими купонными ставками, чем прогнозировалось, их можно будет быстро убрать из рассмотрения, как бумаги, не стоящие внимания аналитика.

Ко всему вышеперечисленному хочется добавить, что модель может служить подтверждением принятого инвестиционного решения. В мире управления активами портфельные менеджеры делают предварительные решения, которые не всегда оказываются хорошими решениями постфактум. Они часто сталкиваются с задачей доказать, что инвестиционное решение все еще является хорошей идеей, учитывая информацию известную в данное время. Модель может быть использована в качестве еще одного аналитического инструмента в подтверждение уже принятого решения.

Наконец, модель можно использовать как некоторую форму отработанной информации. Например, если бы управляющий активами маркировал каждое инвестиционное решение и хранил бы его в надлежащей базе данных вместе с данными, которые были проанализирована командой аналитиков компании. Затем такую модель можно было бы обучать постепенно, со временем превращая её в более мощный инструмент, который будет служить коллективной памятью для всей команды аналитиков, поскольку каждая функция вместе с меткой инвестиционного аналитика будет храниться должным образом.

Чтобы увеличить преимущества модели, разработанной в моей работе, на практике, самая большая проблема, которую придётся решитьпри дальнейших исследованиях — это автоматизация процесса сбора данных. Для данного исследования сбор всех необходимых данных по облигациям, выпускам и экономической ситуации в странах был очень ручной и трудоемкой задачей. Таким образом, чтобы действительно извлечь выгоду из полученной в данной работе модели, набор данных, используемых в качестве входных данных, должен быть автоматизирован полностью.

# **ЗАКЛЮЧЕНИЕ**

Это исследование направлено на оценку влияния различных детерминант таких как Share of Debt, Leverage, ROA, рейтинг эмитента, годовая инфляция, дюрация бумаги, периодичность выплаты купонов, объем размещения, ВВП, ежегодный рост индекса потребительских цен, место в рейтинге Doing Business и индекс коррупции на купонную ставку высокодоходных облигаций европейского рынка с 2003 по 2022 годы и дальнейшее прогнозирование купонных ставок для прогнозной выборки бумаг.

На основе результатов оценки с использованием языка программирования Python очевидно, что годовая инфляция, рейтинг эмитента и место в рейтинге Doing Business оказывают самое существенное влияние на купонную ставку облигаций, в то время как тип облигации, периодичность выплаты купонов и объем размещения не являются сильными детерминантами, влияющими на купонную ставку ВДО европейского рынка. Влияние инфляции несложно объяснить, ведь она порождает интересный момент восприятия риска инвесторами, что помогает объяснить некоторые недавние изменения купонной ставки облигаций. Идея заключается в том, что во время кризиса инвесторы хотят обезопасить свои инвестиции и поэтому отступают к наименее рискованным активам. Данное исследование подразумевает, что компании, выпускающие облигации, должны обеспечить ставку купона, привлекательную для инвесторов, особенно когда экономика испытывает высокие темпы инфляции в качестве противовеса реальным результатам. Результаты этого исследования показывают, что инвесторы будут заинтересованы в инвестировании в облигации с более высокой доходностью, если компания будет вынуждена установить более высокую купонную ставку, особенно в экономических условиях с низким уровнем инфляции.

После изучения основных детерминант, влияющих на купонную ставку облигаций, я создала базу данных из описанных выше детерминант за период с 2003 по 2022 год, провела анализ данных, построила их взаимную корреляцию и вес каждого фактора в модели. Итак, коэффициенты, обычно принимаемыми в качестве определяющих факторов купона по облигациям, а также стандартные переменные характеристик облигаций, представляли в моей модели объясняющие переменные для таргетовой переменной (купонная ставка).

Первоначально мною была выполнена множественная линейная регрессия, сначала только с рейтингом в качестве объясняющей переменной, затем я добавила все остальные детерминанты. Точно так же я настроила три алгоритма машинного обучения для прогнозирования спредов в невидимом наборе тестовых данных. Использовались три модели: Neural Network, Gradient Boosting, Random Forest. По итогу исследования, лучшей моделью оказалась модель Neural Network. Алгоритм работает лучше всего, что указывает на то, что связь между всеми детерминантами носит сложный, нелинейный характер. Итогом работы стало достаточно точное прогнозирование купонной ставки ВДО европейского рынка из определённой заранее прогнозной выборки. Результаты выборки оказались достаточно точными. В среднем по выборке из 107 различных бумаг и 2 354 уникальных значений наша модель неправильно предсказала купонную ставку в 25% случаев из 100%. Средне отклонение купонной ставки составило 0,88% при том, что средняя ставка купона в выборке 3,43%. Самые большие расхождения между прогнозом и реальной величиной купоной ставки наблюдаются в диапазоне реальной купонной ставки от 0 до 2 %. Здесь модель ошибалась практически в каждой второй бумаге. Что касается бумаг с величиной ставки от 2 до 6%, здесь показатели довольно оптимистичные, так в среднем отклонение прогнозных значений от реальных показателей составило не более 17%.

В разрезе стран самые точные предсказания оказались по таким странам как Испания, Италия, Франция и Финляндия. Такой хороший результат для первых трёх стран объясняется тем, что они достаточно часто встречаются в тестовой выборке. Что касается Финляндии, то купонный диапазон представленных в прогнозной выборке бумаг составляет от 1,7 до 4,2%, что, как мы сказали ранее, является достаточно распространённым диапазоном для тестовой выборки, соответственно модель смогла научиться и построить для себя сильные связи, чтобы в дальнейшем как можно точно предсказать купонные ставки на прогнозном диапазоне.

Выводы, сделанные в моей работе, вносят вклад в академические круги, делая вклад в ещё не до конца изученные исследования относительно европейского рынка ВДО и давая дальнейшую почву для развития исследований ВДО с использованием машинного обучения. Кроме того, мои исследования представляют основу для отраслевых практиков, которая может повысить точность принятия инвестиционных решений, а также их дальнейшей оценки.

Исследование имеет ограничения, особенно факторы, используемые при определении купонной ставки облигаций, так что многие факторы не оказывают никакого влияния. Поэтому в дальнейших исследованиях предлагается рассмотреть вопрос о добавлении других переменных в модель, как внутренних, так и внешних факторов, например; ликвидность, прибыльность, корпоративное управление, уровень долга, размер компании и настроения инвесторов.

# **СПИСОК ЛИТЕРАТУРЫ**

1. Милицкова Т. М. Влияние специфических факторов на спреды доходности корпоративных облигаций //Корпоративные финансы. – 2013. – Т. 7. – №. 2.
2. Теплова Т. В., Родина В. А. Высокодоходные облигации: от истории становления рынка в США до российских реалий. М.: ИНФРА-М, 2019.
3. Alexander G. J., Edwards A. K., Ferri M. G. The determinants of trading volume of high-yield corporate bonds //Journal of Financial Markets. – 2000. – Т. 3. – №. 2. – С. 177-204.
4. Andres C., Betzer A., Limbach P. Underwriter reputation and the quality of certification: Evidence from high-yield bonds //Journal of Banking & Finance. – 2014. – Т. 40. – С. 97-115.
5. Anwar, C.J., & Suhendra, I. Monetary Policy Independence and Bond Yield in Developing Countries // Journal of Asian Finance, Economics and Business. – 2019. - № 7. - С. 23-31.
6. Asquith P., Mullins Jr D. W., Wolff E. D. Original issue high yield bonds: Aging analyses of defaults, exchanges, and calls //The Journal of Finance. – 1989. – Т. 44. – №. 4. – С. 923-952.
7. Barnhill Jr T. M., Joutz F. L., Maxwell W. F. Factors affecting the yields on noninvestment grade bond indices: a cointegration analysis //Journal of empirical finance. – 2000. – Т. 7. – №. 1. – С. 57-86.
8. Bhojraj, S., & Sengupta, P. Effect of Corporate Governance on Bond Rating and Yields: The Role of Institutional Investors and Outside Directors // Journal of Business. – 2003. - №76(3). – С. 455-475.
9. Blume M. E., Keim D. B., Patel S. A. Returns and volatility of low‐grade bonds 1977–1989 //The Journal of Finance. – 1991. – Т. 46. – №. 1. – С. 49-74.
10. Chen, L., Lesmond, D. A., & Wei, J. Corporate Yield Spreads and Bond Liquidity // Journal of Finance. – 2007. - № 6(1). – С. 119-149.
11. Cornell B. Liquidity and the pricing of low-grade bonds //Financial Analysts Journal. – 1992. – Т. 48. – №. 1. – С. 63-67.
12. Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. Factors Affecting the Valuation of Corporate Bonds // Journal of Banking and Finance. – 2004. - №28(11). – С. 2747-2767.
13. Fridson M. S., Cherry M. A. Initial Pricing as a Predictor of Subsequent Performance of High-Yield Bonds //Financial Analysts Journal. – 1990. – Т. 46. – №. 4. – С. 61-67.
14. Fridson M. S., Gao Y. Primary versus secondary pricing of high-yield bonds //Financial Analysts Journal. – 1996. – Т. 52. – №. 3. – С. 20-27.
15. Fridson M. S., Garman C. M. Determinants of spreads on new high-yield bonds //Financial Analysts Journal. – 1998. – Т. 54. – №. 2. – С. 28-39.
16. Grammenos C. T., Alizadeh A. H., Papapostolou N. C. Factors affecting the dynamics of yield premia on shipping seasoned high yield bonds //Transportation Research Part E: Logistics and Transportation Review. – 2007. – Т. 43. – №. 5. – С. 549-564.
17. Hammami, Y., & Bahri, M. On the determinants of expected corporate bond returns in Tunisia // Research in International Business and Finance. – 2016. - №38. – С. 224-235.
18. Hamid, A. A., Siagian, A., Razak, A., & Endri, E. Determinants of Bond Rating and its Implications to Corporate Bond Yield // International Journal of Engineering and Advanced Technology. – 2019. - № 9(2). – С. 195-200.
19. Joutz F. L., Maxwell W. F. Modeling the yields on noninvestment grade bond indexes: Credit risk and macroeconomic factors //International Review of Financial Analysis. – 2002. – Т. 11. – №. 3. – С. 345-374.
20. Lace, N., Macerinskiene, I., & Balciunas, A. Determining the EUR/USD Exchange Rate with U.S. and German Government Bond Yields in the Post-Crisis Period // Intellectual Economics. – 2015. - № 9. С. 150-155.
21. Liu, Y., & Jiraporn, P. The Effect of CEO Power on Bond Ratings and Yield. // Journal of Empirical Finance. – 2010. - № 17(4). – С. 744-762.
22. Sarig O., Warga A. Some empirical estimates of the risk structure of interest rates //The Journal of Finance. – 1989. – Т. 44. – №. 5. – С. 1351-1360.
23. Shulman J., Bayless M., Price K. Marketability and default influences on the yield premia of speculative-grade debt //Financial Management. – 1993. – С. 132-141.
24. Van Landschoot, A. Determinants of yield spread dynamics: Euro versus US dollar corporate bonds // Journal of Banking & Finance. – 2008. - № 32(12). – С. 2597-2605.
25. Yieand, C. V., & Chen, N.H. Determinants of Bond Yield // International Journal of Recent Technology and Engineering (IJRTE). – 2019. - № 7(5S). – С. 238-244.
26. Zhang T. W., Wu W. H. The asymmetric predictability of high-yield bonds //The North American Journal of Economics and Finance. – 2014. – Т. 29. – С. 146-155.

**Электронные ресурсы**

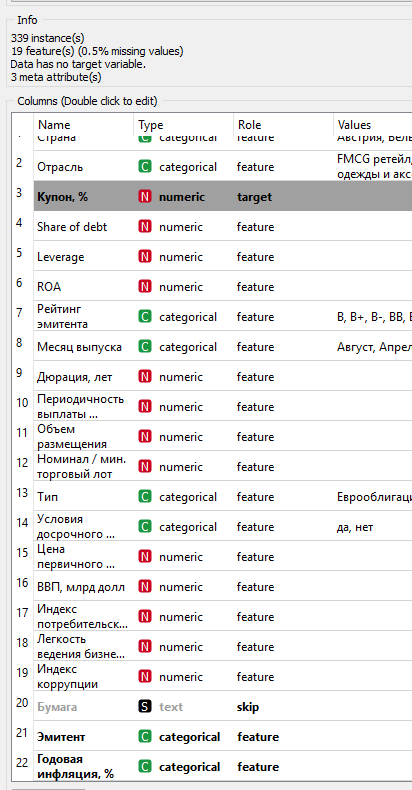
1. Запись в telegram-канале Высокодоходные облигации. – 2018. –URL: <https://t.me/russianjunkbonds/16> (дата обращения: 13.12.2021).
2. Козлов Д. Рынок высокодоходных облигаций. Весна 2021 [вебинар] //YouTube. – 2021. –URL: <https://youtu.be/aQ3HahPww3M> (дата обращения: 15.04.2022).
3. Федеральная служба государственной статистики. –URL:<https://rosstat.gov.ru/folder/210/document/13223?print=1>(дата обращения: 02.03.2021).

# **ПРИЛОЖЕНИЯ**

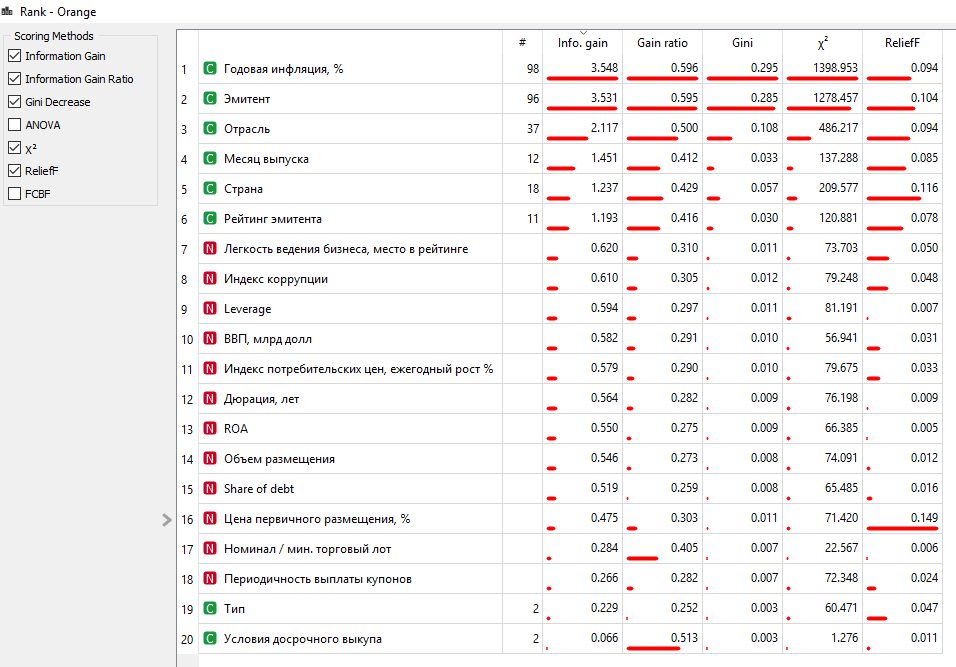
Изображение выглядит как диаграмма, карта

Автоматически созданное описание

# Приложение 1. – Стрим в Orange.



# Приложение 2. – Назначение неопределенных переменных в Orange.



# Приложение 3. – Виджет Rank в Orange.

Изображение выглядит как диаграмма

Автоматически созданное описание

# Приложение 4. – Перечень моделей, используемых в Orange.

Изображение выглядит как текст, диаграмма, карта

Автоматически созданное описание

# Приложение 5. – Выгрузка модели для прогноза Orange.

Изображение выглядит как текст, диаграмма, круг, карта

Автоматически созданное описание

# Приложение 6. – Схема для предсказания купонной ставки в Orange.

Изображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

# Приложение 7. – Результат предсказания купонной ставки в Orange.