

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

имени М.В.ЛОМОНОСОВА

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

«Расширения многофакторных моделей доходности для российского
фондового рынка»

«Extensions of multifactor return models for Russian stock market»

Выполнил студент

Э-401:

Петраков Сергей Денисович

Научный руководитель:

Ассистент Байбаков Владислав Игоревич

Москва, 2021

Аннотация

В данной работе проводится оценка ожидаемой доходности российских компаний на основе многофакторных моделей Фамы-Френча, Кархарта и расширений над ними, включающими учёт долговой нагрузки компаний, специфические для России факторы, а также базовые детерминанты доходности и их лаги. Полученные результаты подтверждают значимость как классических факторов, так и добавленных, и свидетельствуют о целесообразности расширений для большинства крупных российских компаний. Дополнительное сравнение эффективности классических подходов к оценке параметров моделей и подходов, основанных на методах машинного обучения, показало, что в среднем, они сопоставимы по качеству, однако с некоторым преимуществом последних. Результаты устойчивы к различным спецификациям моделей.

Кроме того, методы машинного обучения, учитывающие нелинейные зависимости, позволяют оценить степень важности отдельных факторов при прогнозе ожидаемой доходности. Большую количественную оценку значимости получает премия за рыночный риск, что согласуется с экономической теорией и интуицией. Стоит также отметить, что выводы, сделанные в работе, могут помочь в решении «factor zoo» проблемы.

JEL-коды: C10, C12, C13, C45, C49, C50, G12, G19

Оглавление

Введение	5
Глава 1. Теоретические основы ценообразования активов на развивающихся рынках.....	12
1.1. Факторные модели ожидаемой доходности акции	12
1.1.1. Модель CAPM	13
1.1.2. Модель APT	16
1.1.3. Трёхфакторная модель Фамы-Френча	18
1.1.4. Четырёхфакторная модель Кархарта.....	20
1.2. Специфические особенности развивающихся рынков.....	21
1.3. Выводы по главе.....	23
Глава 2. Многофакторные линейные модели доходности акции	24
2.1. Краткая характеристика основных используемых данных.....	24
2.2.1. Безрисковая ставка	26
2.2.2. Премия за рыночный риск	27
2.2.3. Факторы SMB, HML, MOM	27
2.3. Факторы, существенные для российского фондового рынка	30
2.3.1. Цена на нефть.....	30
2.3.2. Курс рубля к доллару США	31
2.4. Дополнительные факторы	31
2.5. Тестирование классических и расширенных моделей.	32
2.6. Выводы по главе.....	36
Глава 3. Многофакторные модели доходности акции на основе методов машинного обучения	37
3.1. Теоретические основы алгоритмов машинного обучения	37
3.1.1. Метод k ближайших соседей. kNN (k Nearest Neighbors)	37
3.1.2. Lasso, Ridge (гребневая регрессия), Elastic Net	38
3.1.3. Решающее дерево (DT), Случайный лес (RF)	39
3.1.4. Метод опорных векторов (SVM)	43
3.1.5. Градиентный бустинг (Gradient Boosting, GB).....	44
3.1.6. Нейронная сеть (Neural Network, NN).....	45
3.2. Применение алгоритмов ML в моделировании доходности акции	45
3.3. Сравнение качества моделей ML и линейных расширений классических моделей.....	49

3.4. Оценка важности факторов с помощью случайного леса	53
3.5. Проверка устойчивости полученных результатов на алгоритмах градиентного бустинга.....	55
3.6. Отбор факторов с помощью LASSO	56
3.7. Пример практического использования полученных расширений моделей доходности акции на базе торговой рекомендательной стратегии	58
3.8. Выводы по главе.....	65
Заключение.....	67
Библиографический список.....	69
Приложения.	74

Введение

Актуальность темы исследования. Ценообразование активов является одной из важнейших областей экономической теории, а её **актуальность** подтверждается как большим числом академических работ, так и спросом на прикладные модели со стороны финансовой индустрии. Среди ключевых областей применения моделей ценообразования активов можно выделить, например, разработку и реализацию торговых стратегий и алгоритмов, управление инвестиционным портфелем, определение ставки дисконтирования при оценке бизнеса и прочие. От точности моделей, определяющих справедливые стоимости и ожидаемые доходности активов, зачастую будет зависеть финансовый результат сделок.

Несмотря на популярность тематики, посвященной ценообразованию финансовых активов, в академических исследованиях, в современных моделях оценки ожидаемой доходности имеет место ряд проблем, ограничивающих применимость моделей на практике. Одной из самых известных является широкий набор доступных для выбора факторов и отсутствие ответа на вопрос о том, что в итоге имеет смысл использовать в модели в качестве детерминант доходности, также известная как «factor zoo» проблема. Данная проблема становится особенно острой в наше время из-за стремительного увеличения объёмов доступной для анализа информации. Ограничением для выбора набора исследуемых факторов является, в том числе, потребность в вычислительной мощности, поскольку обработать все возможные варианты наборов детерминант доходности на данный момент не представляется возможным¹. Однако проблема заключается не только в вычислительных мощностях. Зачастую трудности возникают и при относительно небольших размерах изучаемых данных из-за отсутствия единого критерия выбора значимых факторов доходности.

Кроме того, общепринятой практикой при изучении детерминант доходности является использование линейных или линеаризованных моделей. При этом взаимосвязь между различными показателями может быть нелинейной и поддаваться линеаризации с существенной погрешностью. Таким образом, возникает дилемма при выборе между

¹ Однако стоит заметить, что проблемы вычислительных мощностей в скором будущем могут быть решены за счёт внедрения квантовых компьютеров, которые позволят вывести вычисления на новый уровень. Иными словами, в такой ситуации перебрать все варианты комбинаций параметров в модели и выбрать лучший станет уже возможным.

использованием более сложных и точных моделей, с одной стороны, и использованием простых линеаризованных моделей, с другой стороны. Нельзя не отметить, что при оценке ожидаемой доходности акций на практике зачастую ориентируются на самую простую модель ценообразования капитальных активов (CAPM) или ее модификации. Она проста в построении и интерпретации, но за её простотой может скрываться довольно существенная потеря в качестве и точности оценки.

Данная работа предлагает подходы к решению проблем путём оценки ожидаемой доходности российских компаний на основе многофакторных моделей Фамы-Френча, Кархарта и расширений над ними, включающими учёт долговой нагрузки компаний, специфические для России факторы, а также базовые детерминанты доходности и их лаги. Получаемые результаты помогают определить значимость классических и дополнительных факторов, ответить на вопрос о целесообразности подобных расширений моделей. Более того, дополнительный анализ по сравнению эффективности классических подходов к оценке параметров моделей и подходов, основанных на методах машинного обучения (ML), позволяет сопоставить качество результатов, получаемых разными методами.

Кроме того, методы машинного обучения, учитывающие нелинейные зависимости, дают возможность численно оценить степень важности отдельных факторов при прогнозе ожидаемой доходности [Simonian J. et al 2019]. Большую количественную оценку значимости получает премия за рыночный риск, что согласуется с экономической теорией и интуицией. Стоит также отметить, что выводы, сделанные в работе, могут помочь в решении «factor zoo» проблемы, поднимаемой во многих исследованиях [Harvey C. R., Liu Y. 2015, Feng G. et al 2017, 2020], основной идеей которой является сложность оценки релевантности тех или иных факторов для многофакторной модели. Результат достигается за счёт использования оценок важности факторов, полученных при настройке модели.

Учитывая все вышесказанное, тематика, выбранная для исследования, представляется весьма актуальной, а результаты исследования могут быть полезны как с теоретической, так и прикладной точек зрения.

Степень научной разработанности проблемы определяется большим количеством исследований в рамках данной области. Развитие теории факторных моделей доходности акции началось с модели CAPM, в которой единственным фактором, определяющим равновесную ожидаемую доходность, является рыночная премия за риск [Sharpe W. F., 1964; Lintner, 1965; Mossin J., 1966]. Модель была протестирована множеством исследователей как на развитых, так и на развивающихся рынках [Estrada J., 2002]. Много работ было посвящено различным аспектам модели, начиная от

корректировки её предпосылок [Black, Jensen, Scholes, 1972] и заканчивая рассмотрением отдельных феноменов на фондовом рынке, таких как эффект дня [Gibbons M. 1981].

Несколько позднее была разработана модель арбитражного ценообразования (APT). Она была предложена экономистом Стивеном Россом [Ross S., 1976] как развитие идей CAPM с более слабыми предпосылками. Модель тестировалась, в основном, в целях идентификации значимых детерминант доходности, что является самым слабым местом модели [Roll R. 1980; Chen N., Ingersoll Jr J. E., 1983; Chen N. F., Roll R., Ross S., 1986]. Полученные результаты говорили в пользу важности инфляции, ВВП и премии к доходности в корпоративных облигациях.

Трёхфакторная модель доходности Фамы-Френча является развитием моделей CAPM и APT [Fama E. F., French K. R., 1993]. Её основной идеей является включение двух дополнительных детерминант равновесной доходности: величины капитализации и соотношения балансовой и рыночной стоимостей компании. Модель тестировалась в ряде работ [Fama E. F., French K. R., 2012; Liu J., Stambaugh R. F., Yuan Y., 2019].

В 1997 Марк Кархарт предложил добавить в многофакторную модель Фамы-Френча дополнительный фактор – моментум (MOM) [Carhart, 1997]. К тому времени накопилось достаточно оснований к тому, что считать данный фактор достаточно важным относительно ожидаемой доходности акции [Jegadeesh N., Titman S., 1993; Hendricks D., Patel J., Zeckhauser R., 1993], в связи с чем и возникла идея использовать его как дополнительный фактор для трехфакторной модели Фамы-Френча.

В настоящее время в различных исследовательских областях всё больше и больше применяются продвинутые методы анализа данных – методы машинного обучения. При этом как задачу оценки ожидаемой доходности акции, так и вопрос о выборе факторов для модели и оценки их важности можно решить методами машинного обучения. Существует множество исследований, в которых рассматриваются приложения методов машинного обучения к теории ценообразования активов [Patel J. 2015; Bryzgalova S., et al 2019; Diallo B., 2019].

Кроме того, есть работы, где демонстрируются способы оценить количественную степень важности факторов для анализа ожидаемой доходности акции, а также оценить многофакторные модели, используя ML [Simonian J. et al, 2019; Gu S. et al, 2018, 2021]. Исследователи используют ML методы для таких задач как анализ ожидаемой доходности компании и динамики стоимости акции, анализ нелинейности зависимостей финансовых данных, выявления статистического арбитража и других смежных с этой темой вопросов.

Несмотря на широкую изученность проблематики, открытыми остаются различные вопросы: каковы возможности по улучшению результатов моделей для различных рынков, за счёт чего это можно реализовать, какие детерминанты доходности, способствующие улучшению качества, можно выделить дополнительно, будет ли их влияние устойчивым. С каждым годом появляется множество исследований с использованием ML инструментария в большом количестве вопросов, связанных с финансовой тематикой. Практика показывает, что они способны значительно улучшать качество результатов, что объясняет их активное использование. Шаг в направлении решения озвученных проблем делается в данной работе.

Целью работы является разработка модификации многофакторной модели доходности акций Фамы-Френча-Кархарта для развивающихся рынков с использованием методов машинного обучения.

Для достижение поставленной цели необходимо выполнить ряд **исследовательских задач**:

- 1) Выделить преимущества и ограничения современных многофакторных моделей доходности.
- 2) Выделить специфику развивающихся фондовых рынков с точки зрения ограничений применения на них многофакторных моделей доходности.
- 3) Оценить классические многофакторные модели Фамы-Френча и Кархарта на российском фондовом рынке.
- 4) Предложить расширение модели Фамы-Френча-Кархарта, улучшающее характеристики модели при применении на развивающихся рынках.
- 5) Оценить разработанную модификацию модели Фамы-Френча на российском фондовом рынке.
- 6) Путем сравнительного анализа показать более высокую эффективность разработанной модели по сравнению с классическими моделями при применении на российском фондовом рынке.
- 7) Сравнить качество расширенной модели и прочих моделей, построенных на основе методов машинного обучения.
- 8) Оценить степень важности факторов с точки зрения моделей, основанных на методах машинного обучения и позволяющих улавливать нелинейные взаимосвязи в данных.

В качестве **объекта исследования** в данной работе выступает ожидаемая доходность акции на российском фондовом рынке.

Предметом исследования являются детерминанты ожидаемой доходности акции на российском фондовом рынке.

В качестве элементов **научной новизны** в работе можно выделить следующее:

- 1) Построены расширения многофакторной модели доходности Фамы-Френча-Кархарта для акций российских компаний на свежих данных по российскому фондовому рынку с учетом классических детерминант доходности и их лагов, долговой нагрузке компаний, лагах зависимой переменной, специфических для России факторов и применением методов машинного обучения. Проведено сопоставление эффективности различных моделей на основе различных критериев
- 2) Вычислены количественные оценки важности факторов расширенных моделей ожидаемой доходности акций российских компаний с помощью методов машинного обучения.

Исследование обладает **практической значимостью**. С прикладной точки зрения, результаты данной работы могут применяться в таких областях как разработка торговых стратегий и алгоритмов, а также определение требуемой доходности собственного капитала компании при оценке бизнеса. Выводы, сделанные в исследовании, могут представлять интерес в трейдинге, так как количественное понимание того, насколько тот или иной фактор важен в предсказании доходности акции может упростить выбор индикаторов для отслеживания, на которых может базироваться торговая стратегия. Кроме того, результаты исследования могут быть интересны портфельным управляющим, так как количественные оценки важности детерминант доходности могут использоваться не только для определения уточненных оценок доходности портфеля, но и, в том числе, для целей диверсификации.

Теоретическая значимость работы заключается в применении продвинутого ML-инструментария для анализа фундаментальных факторов доходности. Данное исследование является попыткой решения проблем множества потенциальных детерминант доходности («factor zoo») и нелинейных взаимосвязей в данных, а его результаты могут использоваться при дальнейшем изучении анализируемой проблематики с помощью продвинутых методов машинного обучения.

Основными **методами исследования** являются:

- сравнительный анализ,
- синтез,
- индукция,

- дедукция,
- математические методы:
 - регрессионный анализ,
 - методы машинного обучения (случайный лес, метод опорных векторов, Lasso и Ridge регрессии, градиентный бустинг над лесами (XGBoost, LightGBM, Catboost), нейронная сеть (Многослойный персептрон)).

Информационную базу исследования составляют работы зарубежных и отечественных исследователей, посвященные многофакторным моделям доходности, а также данные, полученные из следующих информационных систем:

- Bloomberg (данные по ценам акций компаний, отношения рыночных цен к балансовой стоимости акций, текущей рыночной капитализации компании, отношении долга к рыночной капитализации)
- Сайт Московской биржи (данные о кривой бескупонной доходности и курсе валютной пары доллар/рубль²)
- Сайт независимого агентства по управлению энергетической информацией³ (данные о динамике нефтяных цен).

Структура и логика исследования устроена классическим образом. Работа состоит из следующих частей: введение, три главы, заключение, список использованной литературы и приложения.

Основной текст выпускной квалификационной работы состоит из 68 страниц. В тексте присутствуют 15 таблиц, 13 рисунков, 31 формул, 26 приложений.

В первой главе проводится анализ классических многофакторных моделей доходности и исследований на эту тему на разных рынках, что помогает выделить преимущества и ограничения данных моделей. Кроме того, анализируется специфика развивающихся рынков, которая может оказывать влияние на эффективность применения моделей. Таким образом, первая глава посвящена решению первых двух исследовательских задач.

Во второй главе производится описание собранных и рассчитанных данных, а также построение классических многофакторных моделей и их расширений на основе

² <https://www.moex.com/>

³ <https://www.eia.gov/>

стандартного инструментария. Далее проводится анализ сопутствующих аргументов по их использованию. Соответственно, во второй главе внимание уделено решению задач 3 – 6.

В третьей главе рассматриваются теоретические основы используемых алгоритмов машинного обучения, после чего обучаются модели и сравниваются их результаты с расширениями классических моделей. Производится оценка важности факторов. Выполняются проверки на устойчивость результатов. Это позволяет решить последние две исследовательские задачи.

Все программные коды по расчётам и визуализации данных на языке Python, а также сами используемые данные находятся в открытом доступе⁴.

⁴ <https://github.com/SergeyPetrakov/Diploma>

Глава 1. Теоретические основы ценообразования активов на развивающихся рынках

В этой главе проводится анализ базовых моделей ожидаемой доходности акции, и результатов их тестирований в уже существующих исследованиях на развитых и развивающихся рынках. Отдельное внимание уделяется специфике развивающихся рынков, которая может оказывать влияние на эффективность применения моделей.

1.1. Факторные модели ожидаемой доходности акции

Основа для многофакторных теорий доходности была заложена Г.Марковицем, который продемонстрировал положительную зависимость между доходностью акции и риском, и последующими исследованиями двух составляющих риска: систематическую [Treynor Jack L., 1961; Sharpe W. F., 1964; Lintner, 1965; Mossin J., 1966] и идиосинкратическую [Merton R. C., 1973, 1987; Levy, 1978]. Первая группа учёных, которая отрицала роль идиосинкратической составляющей, стала основоположниками ядра факторных моделей – модели CAPM. Затем теория продолжила своё развитие, беря во внимание уже множество других факторов кроме систематического риска, то есть не считая его единственным влияющим на доходность фактором.

Основной идеей факторных моделей ожидаемой доходности является принцип того, что ожидаемая доходность акции поддается влиянию некоторого набора факторов. В финансовой теории такие взаимосвязи называются факторными моделями, основной задачей которых является выявление этих факторов, а также определены чувствительности динамики доходности целевого актива к их динамике.

Общий вид однофакторной модели:

$$E[r_j] = r_f + \beta E[x_j] \quad (1)$$

Где $E[r_j]$ – это ожидаемая доходность ценной бумаги j , $E[*]$ – это оператор математического ожидания, β – чувствительность ожидаемой доходности к изменению ожидаемого значения фактора x_j , r_f – безрисковая ставка.

Для многофакторных моделей доходности акции общая формула имеет следующий вид:

$$E[r_j] = r_f + \beta_1 E[x_{1j}] + \beta_2 E[x_{2j}] + \dots + \beta_k E[x_{kj}] \quad (2)$$

Основными многофакторными моделями ожидаемой доходности акций являются:

- модели в рамках теории арбитражного ценообразования (АРТ),
- трехфакторная модель Фамы-Френча,
- четырехфакторная модель Фамы-Френча-Кархарта,
- пятифакторная модель Фамы-Френча.

1.1.1. Модель CAPM

CAPM (Модель ценообразования капитальных активов) является основополагающей финансовой моделью ценообразования активов, которую в период 1960-ых годов одновременно и независимо друг от друга предложил ряд выдающихся учёных: Джек Тейнор [Treynor, 1961], Уильям Шарп [Sharpe W. F., 1964], Джон Линтнер [Lintner, 1965] и Ян Моссин [Mossin J., 1966] на основе предпосылок заложенной ранее Гарри Марковицем [Markowitz, 1952] фундаментальной теории оптимального выбора активов, основанной на соотношении доходности и риска.

Предпосылки модели CAPM могут быть разделены на 2 основные группы:

1. Предпосылки особенностей поведения инвесторов.
 - 1.1. Гомогенность ожиданий инвесторов.
 - 1.2. Одинаковый горизонт планирования инвестиций всех участников рынка.
 - 1.3. Избегание инвесторами риска (risk-averse), на основе идей Марковица инвесторы стремятся оптимизировать комбинацию риска и доходности.
2. Предпосылки рыночной структуры.
 - 2.1. Отсутствие транзакционных издержек.
 - 2.2. Отсутствие налогов.
 - 2.3. Публичная доступность всей информации.
 - 2.4. Доступность коротких продаж.
 - 2.5. Возможность для участников рынка брать и давать в долг под одинаковую безрисковую процентную ставку⁵.
 - 2.6. Делимость активов.

Общая формула модели имеет следующий вид

$$r_e = r_f + \beta(r_m - r_f) \quad (3)$$

⁵ В оригинальном варианте модели CAPM Линтнер и Шарп постулировали предпосылку существования единой безрисковой ставки, однако она была опровергнута в статье Блек, Йенсен, Шоулз [Black, Jensen, Scholes, 1972] и заменена на возможность неограниченного количества коротких продаж.

Где r_f – доходность безрисковых активов,

β – коэффициент чувствительности оцениваемой бумаги к изменениям рыночной доходности, r_e – интересующая ожидаемая доходность акции, r_m – ожидаемая средняя норма прибыли рыночного портфеля

Таким образом, в модели CAPM единственным фактором, определяющим ожидаемую доходность акции в равновесии, является рыночная премия за риск.

Критика модели, в основном, строилась на нереалистичности ее предпосылок. При этом тестированию модели посвящен широкий пласт академических исследований, некоторые из которых приведены в Таблице 1.

Год	Автор	Рынок	Итог
1998	R Garcia, E Ghysels	Развивающиеся рынки	Значимая роль структурных изменений на развивающихся рынках. Необходимость проведения соответствующих тестов.
2001	Estrada J.	Развивающиеся рынки	Важность учёта специфических факторов для развивающихся рынков.
2008	Bruner R. F. et al.	Развитый и развивающийся	Важность выбора рыночного портфеля на развитом рынке выше, чем на развивающемся.
2010	L.E. Pereiro	Развитый рынок (США), развивающийся рынок (Китай)	Проблема в эквивалентности показателя бета для компаний из одинаковых отраслей развитых и развивающихся рынков.
2012	Dranev Y., Fomkina S.	Россия 2003 - 2012	Важность учёта особенностей развивающихся рынков капитала: низкой ликвидности, недостатка информации и специфических рисков.

Таблица 1. Тестирование модели CAPM на разных рынках.

Источник: составлено автором на основе обзора научных статей.

Приведенные в Таблице 1 исследования демонстрируют ряд существенных различий между развитыми и развивающимися рынками. Они оказываются важны даже при тестировании базовой модели CAPM.

Так, учёные отдельное внимание уделяют специфическим особенностям развивающихся рынков, выражающихся в наличии для них специфических факторов.

Множество эмпирических исследований посвящены исследованию различных аномалий на фондовом рынке, объяснение которых выходит за рамки модели CAPM (см. Таблицу 2).

Год	Авторы	Рынок	Итог
1977 1983	Basu S.	Обыкновенные акции NYSE, развитый рынок	Акции фирм с высоким значением прибыли/цена (Е/Р) приносят в среднем более высокую доходность, чем при низком Е/Р при поправке на риск. Эффект размера нивелируется при контроле различий в риске и соотношениях Е/Р.
1981	Banz	Обыкновенные акции NYSE	Фирмы небольшого размера в среднем имели более высокую доходность с поправкой на риск, чем крупные фирмы.
1981	Gibbons M.	Акции S&P 500	Наличие эффекта дня недели на доходность ценной бумаги - устойчиво более низкая средняя доходность в понедельник.
1983	Stoll H. R., Whaley R. E.	NYSE, AMEX, развитый рынок	Общая рыночная стоимость обыкновенных акций и их цена обратно пропорциональны доходности с поправкой на риск. Транзакционные издержки частично объясняют отклонение от нормы.
1983	Reinganum M. R.	Обыкновенные акции NYSE	Первые дни января оказывают положительное влияние на доходность фирм с малой капитализацией.
1988	Bhandari L.C.	Обыкновенные акции NYSE	Ожидаемая доходность акций положительно связана с соотношением долга к собственному капиталу.

Таблица 2. Тестирование аномалий модели CAPM.

Источник: составлено автором на основе анализа научных статей.

На основе представленных в Таблице 2 итогов можно сказать, что есть множество аномалий, которые невозможно объяснить с помощью модели CAPM: календарный эффект (в месячном и дневном варианте), эффект капитализации компаний, эффект положительной связи величины отношения долга к собственному капиталу компании, эффект стоимости компании (высокое значение Е/Р).

Невозможность контроля вышеперечисленных факторов является недостатком этой модели, что ведёт к недостаточно точным результатам. Однако при этом стоит заметить, что все эти недостатки являются следствием простоты модели, что является её самым большим преимуществом, по причине чего она зачастую используется при анализе

ожидаемой доходности. Тем не менее, исследователи стали предлагать способы повысить точность модели и объяснить аномалии CAPM.

Это привело к тому, что на базе модели CAPM были предложены многофакторные модели, такие как АРТ, модель Фамы-Френча, Кархарта.

1.1.2. Модель АРТ

Модель арбитражного ценообразования (АРТ) предложил Стивен Росс [Ross, 1976] как некоторую альтернативу господствующей на тот момент CAPM. Эта модель опиралась на несколько основополагающих предпосылок. Во-первых, отсутствие возможностей добиться арбитражной сделки в долгосрочной перспективе. Во-вторых, доходность ценной бумаги может быть описана несколькими факторами и представлена в виде линейной функции от них. В-третьих, на рынке присутствует возможность диверсифицировать риск за счёт добавления в портфель некоторого количества активов. Последняя предпосылка постулирует тот факт, что несистематический риск портфеля нивелируется большим количеством активов в портфеле.

У АРТ по сравнению с CAPM есть как преимущества, так и недостатки. Причём во многом сложности с АРТ носят скорее технический характер, заключающиеся в трудоёмкости анализа и подготовки таких данных по макроэкономическим условиям, рыночной нише, деятельности конкурентов компании. И самый большой недостаток, конечно же, связан с тем, что модель не представляет некоторого эталонного набора факторов - она лишь говорит о его существовании, что повышает риски столкнуться с проблемой пропуска важной переменной.

При этом у моделей в рамках АРТ есть и существенные преимущества, среди которых не такие сильные предпосылки как у модели CAPM: из-за присутствия в модели не одного, а нескольких факторов, осуществляется контроль на большее количество влияющих переменных, приближая ситуацию к реальным условиям функционирования компании.

Общий вид модели можно представить следующим уравнением:

$$E[r_j] = r_f + \beta_1 E[x_{1j}] + \beta_2 E[x_{2j}] + \dots + \beta_n E[x_{nj}] \quad (4)$$

Где r_f – ставка безрисковой доходности, x_k – k – ый фактор модели,

β_k – чувствительность k фактора

В академических исследованиях множество работ посвящено тестированию модели АРТ и определению ключевых детерминант ожидаемой доходности акции. Продemonстрируем несколько примеров таких работ в Таблице 3.

Год	Авторы	Рынок	Итог
1986	Cho D. C., Eun C. S., Senbet L. W.	349 акций 11 развитых стран за 1973 - 1983 гг.	Отсутствие удовлетворительных результатов при тестировании АРТ на международном уровне. Вопрос локальных результатов открыт.
2003	Altay E. et al.	Германия январь 1988-июнь 2002, Турция январь 1993-июнь 2002	Для Германии удалось выделить 4 фактора, для Турции - 3. Для Германии неожиданные изменения ставки процента и уровня инфляции существенно влияют на доходность активов, тогда как для Турции не было выделено ни одного макроэкономического фактора, имевшего существенное влияние.
2010	Febrian E., Herwany A.	Джакартская фондовая биржа, 1992 - 2007, развивающийся рынок	Бета - не единственный фактор, который может объяснить избыточную доходность портфеля. АРТ может подтвердить избыточную доходность портфеля в периоды наблюдения, в течение которых средние значения избыточной доходности оказываются постоянно отрицательными.
2015	Kisman Z., Restiyanita S. M.	Индонезийская фондовая биржа, 2008 - 2010, развивающийся рынок	Преимущество АРТ перед САРМ в прогнозировании доходности акций.

Таблица 3. Результаты тестирования модели АРТ.

Источник: составлено автором на основе анализа научных статей.

Среди классических факторов риска можно привести валютный курс, динамику ВВП, отраслевой и общий индексы промышленного производства, значения котировок государственных ценных бумаг и др.

Одной из главных практических проблем является неопределённость набора факторов, которые следует использовать при построении модели на практике. Более того, если для одного региона важными являются одни факторы, то совсем необязательно, что они будут иметь существенную значимость при другом тестировании, как можно обнаружить из Таблицы 3. Эта техническая проблема является основной сложностью данной модели.

1.1.3. Трёхфакторная модель Фамы-Френча

Модель была предложена Кеннетом Френчем и Юджином Фамой в 1993 году [Fama E. F., French K. R., 1993]. Она являлась следствием расширения модели ценообразования капитальных активов – с учётом модели межвременного ценообразования ICAPM Мертона [Merton R. C., 1973] и модели АРТ С.Росса. Авторы взяли за основу регрессионный подход и на данных с нескольких американских бирж (NASDAQ, AMEX, NYSE) за 1963 – 1990 годы выявили, что факторы рыночной капитализации компании и отношения рыночной стоимости к балансовой в большей степени, чем рыночная премия за риск, объясняли динамику ожидаемой доходности акций. В связи с этим Фама и Френч предложили добавить эти факторы в модель.

$$E(r_{it}) = r_f + \beta_i(E(r_{m,t}) - r_f) + s_iSMB_t + h_iHML_t \quad (5)$$

Где, r_{it} – это доходность i компании в момент времени t ,

r_f – безрисковая ставка доходности,

$r_{m,t}$ – рыночный портфель,

SMB_t – фактор, основанный на размере компании,

HML_t – фактор, основанный на стоимости компании,

Фактор размера (SMB)

Данный фактор представляет разность двух слагаемых: средней арифметической доходности небольших по капитализации компаний и больших по капитализации компаний. Также называется показателем размера.

$$SMB = \frac{(r(S,L) + r(S,M) + r(S,H))}{3} - \frac{(r(B,L) + r(B,M) + r(B,H))}{3} \quad (6)$$

$r(S,L)$ – соответствует доходности компаний, которые попали в группу S (небольшие по капитализации) и в группу L (компании, показатель В/М которых самый низкий, то есть компании с самым большим потенциалом роста цены акции или самые дешёвые акции).

Аналогично для всех других доходностей $r(*,*)$, где на месте * располагаются два множества, рассчитываемая доходность которых является средневзвешенной по капитализации доходностью из компаний, соответствующих пересечению множеств. Одно соответствует размеру компании и соответственно принимает одно из двух значений (S, B), второе соответствует стоимости компании или отношению В/М (Book to Market ratio) и принимает одно из трёх значений (L, M, H). В оригинальной версии работы Фамы-Френча для размера процентилем, который разделяет большие и маленькие компании, принято

брать 50%. Для показателя стоимости берутся персентиля соответствующие значениям 30% и 70%.

Фактор стоимости (HML)

$$HML = \frac{r(S, L) + r(B, L)}{2} - \frac{r(S, H) + r(B, H)}{2} \quad (7)$$

Доходности вычисляются тем же способом, как описано для фактора размера.

То есть, HML – это показатель разности между двумя портфелями: доходности акций с низким и высоким показателем В/М (дешёвыми и дорогими акциями или большим и небольшим потенциалом роста).

Год	Авторы	Рынок	Итог
1993	Fama E. F., French K. R.	Американский рынок, (NASDAQ, AMEX, NYSE) 1963 – 1990	Факторы стоимости и размера в большей степени, чем фактор бета, объясняют динамику доходности активов.
2000	Davis J. L., Fama E. F., French K. R.	Американский рынок	Явная положительная связь между средней доходностью и балансовой стоимостью капитала в период с 1929 по 1963.
2001	Daniel K., Titman S., Wei K. C. J.	Японский рынок	Доходность японских акций еще более тесно связана с их соотношением балансовой и рыночной стоимости, чем американские. Отвергается модель Фамы-Френча.
2007	Lee E., Liu W., Strong N.	Британский рынок	Дорогие акции малых по капитализации компаний приносят более высокую доходность независимо от ковариации их доходности.
2012	Fama E. F., French K. R.	Европейский рынок	На локальном уровне были получены удовлетворительные результаты при использовании фактора размера и стоимости.
2019	Liu J., Stambaugh R. F., Yuan Y., 2019	Китайский рынок	Расширение над моделью Фамы-Френча объясняет большинство известных китайских аномалий, включая аномалии прибыльности и волатильности, тогда как просто модель Фама-Френча работает плохо.

Таблица 4. Результаты тестирования 3-факторной модели Фамы-Френча.

Источник: составлено автором на основе анализа научных статей.

Итоги проделанных исследований говорят в пользу важности факторов стоимости и размера при оценке ожидаемой доходности акции на множестве рынков. Учёные получают

одинаковые направления взаимосвязи между фактором размера, стоимости компании и её доходностью (См. Таблицу 4), наблюдается устойчивость наблюдаемых зависимостей.

Тем не менее, ряд исследователей находит слабые места у модели Фамы-Френча при её практической реализации и делает шаги по расширению данной модели с целью получения более качественных результатов.

1.1.4. Четырёхфакторная модель Кархарта

В 1997 году Марк Кархарт [Carhart, 1997] предложил рассмотреть 4 факторную модель, добавив в 3-факторную модель Фамы Френча фактор моментума, являющийся одной из прокси переменных скорости изменения цен акций на рынке.

Показатель моментума по определению означает, что будущая динамика зависит от предшествующей динамики движения цены актива или доходности. Если рассматривать динамику цены акции как случайный процесс, то моментум теряет смысл, однако довольно часто на реальных данных можно заметить присутствие влияния данного фактора. Объяснения этого феномена можно найти в теории поведенческой экономики, утверждающая, что агенты могут поддаваться поведенческим аномалиям и верить в сохранение текущей тенденции. Более того, данные эффекты наблюдаются и среди профессиональных участников рынка.

Одним из наиболее известных исследований моментума является работа Джегадиш и Титман [Jegadeesh N., Titman S., 2001], результатом которой является иллюстрация факта превышения примерно на 1% доходности компаний, которые ранее показывали рост, над теми, у которых за предыдущий период такого роста не наблюдалось.

Здесь стоит отметить, что некоторой альтернативой позиции моментум, является средняя реверсия, то есть гипотеза о том, что цена имеет некоторое среднее значение или тренд, вокруг которого происходят ценовые колебания. То есть, отклонившись от него в сторону повышения или понижения, цена будет иметь тенденцию вернуться, тогда как при моментуме она должна была бы сохранить тенденцию. На выборе между данными тенденциями строится множество алгоритмических стратегий торговли на рынке ценных бумаг.

Фактор моментум может быть рассчитан по следующей формуле:

$$MOM_t = r(high\ returns_{t-1}) - r(low\ returns_{t-1}) \quad (8)$$

По построению в модели используется средневзвешенная по капитализации разность между 50% самых доходных за предыдущий период ценных бумаг и 50% наименее доходных за предыдущий период ценных бумаг.

2008	Bello Z. Y.	Американский рынок, 1990 - 2001	Выявлены преимущества 4 факторной модели над прочими классическими моделями
2013	Cakici N., Fabozzi F. J., Tan S.	18 развивающихся рынков с января 1990 по декабрь 2011, более 800 акций из таких стран, как Китай, Индия, Таиланд, Южная Корея, Индонезия, Филиппины, Малайзия и Тайвань	Высокая значимость моментума и стоимости во всех развивающихся регионах, кроме Восточной Европы (значимость отсутствует у моментума)

Таблица 5. Результаты тестирования 4-факторной модели Кархарта на развитом и развивающемся рынке.

Источник: составлено автором на основе анализа научных статей.

Преимуществом модели Кархарта является учёт дополнительных рыночных аномалий, которые не могут учесть классические модели Фамы-Френча и CAPM. Исследователи показывают в своих работах значимость данного фактора на различных выборках данных на развитых и развивающихся рынках, а также демонстрируют преимущества этой модели над прочими классическими моделями в сравнительной статике. Это даёт повод утверждать, что при анализе доходности как минимум в ряде случаев фактор моментума оказывает значимое влияние, поэтому его недоучёт может спровоцировать проблему пропущенной переменной.

Однако вместе с преимуществами данная модель несёт и недостатки, так, например, при практическом применении для реализации необходимо рассчитывать дополнительный фактор, что сравнительно более трудоёмко, чем использование простой CAPM или остановиться на анализе, используя модель Фамы-Френча.

1.2. Специфические особенности развивающихся рынков

Развивающиеся фондовые рынки характеризуются рядом важных особенностей, отличающих их от развитых рынков и не согласующихся с предпосылками классических моделей ожидаемой доходности [Dragotă V., Mitrică E., 2004; Rojas-Suarez, 2014; Spulbar and Birau, 2018; Spulbar et al., 2020]:

- незрелость институциональной системы,
- более низкая степень эффективности рынков,
- более высокий уровень транзакционных издержек,

- низкая ликвидность,
- асимметрия информации,
- более низкий уровень профессионализма участников рынка и, как следствие, более сильное влияние поведенческих факторов на принятие решений инвесторами,
- сильная подверженность внешним шокам.

Год	Автор	Рынок	Итог
2005	Dey	Развитые и развивающиеся	Концепция риска отличается между развитыми и развивающимися рынками.
2007	Blitz, Vliet van	Развитые и развивающиеся	Большая альфа, получаемая за счёт корректировок доходности на риск, не может быть объяснена классическими факторами (рыночной премией за риск, размером, стоимостью и моментумом)
2011	Baker, Bradley, Wurgler		
2014	Frazzini, Pedersen		
2013	Cakici, Fabozzi, Tan	Развивающиеся рынки (18 стран за 1990 - 2011 года)	Фактор стоимости положительно влияет на доходность.
2014	Pungulescu	Развивающиеся	Процент доходности может быть объяснён фактором размера.
2015	Tianchuan L., Pritamani M.	Развивающиеся	Доходности на развивающихся рынках определяются фактором размера.
2016	Zaremba A.	Развитые и развивающиеся	Эффект размера сильнее проявляется для развитых рынков.

Таблица 6. Исследования особенностей развивающегося рынка.

Источник: составлено автором на основе анализа научных статей.

В этом разделе собраны исследования, которые содержат некоторые различия между развитыми и развивающимися рынками (См. таблицу 6). Исследователи показывают существенные особенности развивающихся рынков. Они важны при проведении исследований данных регионов, так как оказывают на результаты значимое влияние. В таблице 6 приводится ряд наблюдений, характеризующих природу развивающихся рынков. Авторы отмечают важность разницы концепций риска для развитых и развивающихся рынков, показывают существенность фактора размера при прогнозе доходности.

Таким образом, если применять классические модели для развивающихся рынков без учёта их специфических характеристик, то можно столкнуться с некорректными

результатами, поскольку не были учтены важные особенности рынка. Из этого следует, что при анализе развивающегося рынка исследователь заведомо сталкивается с некоторыми ограничениями, которые необходимо корректно учесть в анализе.

1.3. Выводы по главе

В данной главе был проведен анализ классических факторных моделей доходности акции. Преимуществами многофакторных моделей являются их относительная простота, интерпретируемость и гибкость. Тем не менее, классические линейные модели объясняют малую долю кроссекционной дисперсии доходности, что является их существенным недостатком. Более того, при работе с данными по развивающимся рынкам наблюдаются специфические явления, которые требуют дополнительного внимания. Можно сделать вывод, что у расширения моделей являются перспективными с точки зрения более точного моделирования ожидаемой доходности акций, в особенности на развивающихся рынках.

Глава 2. Многофакторные линейные модели доходности акции

В данной главе производится выделение потенциальных детерминант доходности, на базе которых далее строятся и оцениваются как использоваться при построении классические модели доходности (САРМ, трёхфакторная модель Фамы-Френча, 4-факторная модель Кархарта), так и их расширения. (расширения модели Кархарта с добавлением лагов, автолагов и специфических для России факторов – нефти и курса доллара). Во всех моделях учитывается долговая нагрузка компании по формуле Хамады. В рамках самой главы все модели будут представлены на основе данных по Сбербанку, другие примеры доступны в приложениях или по ссылке⁶, где располагаются данные и коды реализации моделей.

2.1. Краткая характеристика основных используемых данных

Для исследования были отобраны данные по 48 обыкновенным акциям российских компаний, торгующихся на ММВБ, за период с 13.08.2010 по 04.06.2020. В качестве индекса, описывающего рыночную динамику, использовался IMOEX Index за тот же период времени.

Изначально для каждой акции из терминала Bloomberg были выгружены такие показатели как цена последней сделки на момент окончания торгового дня (Last Price), отношение рыночной цены к балансовой стоимости акции (Price to Book Ratio), текущая рыночная капитализация компании (Current Market Capitalization of a Share Class), отношение долга к рыночной капитализации (Debt To Market Cap Ratio). На базе первичных данных были рассчитаны классические факторы трёхфакторной модели Фамы-Френча.

Используемый набор акций был выбран, так как он в наилучшей степени соответствовал требованиям исследования. Во-первых, необходимо иметь историю торгов достаточно сопоставимую по времени с большинством выбранных крупных компаний. На российском рынке есть примеры крупных компаний, имеющих достаточно мало данных в силу позднего выхода на биржу, например, Mail.ru. Во-вторых, по некоторым компаниям отсутствуют доступные данные о показателе отношения рыночной цены к балансовой стоимости акции, что автоматически делает невозможным корректный учёт её в портфелях для формирования классических факторов. Как следствие, после предварительного анализа

⁶ <https://github.com/SergeyPetrakov/Diploma>

был отобран такой набор акций, который включает довольно большой промежуток времени и при этом содержит достаточное для анализа количество компаний.

В таблице 7 представлены тикеры отобранных для анализа акций, разбитых по соответствующим отраслям. Пояснения относительно тикеров компаний можно найти в приложении (См. Приложение 22).

Отрасли	Тикеры акций
Нефтегазовый сектор	ROSN, LKOH, SNGS, NVTK, GAZP, TATN, BANE
Электроэнергетика	HYDR, FEES, IRAO, UPRO, RSTI, MSNG, MSRS, OGKB
Металлургия и сектор добычи	NLMK, ALRS, GMKN, CHMF, PLZL, MAGN, TRMK, MTLR, VSMO, RASP
Химическое производство	PHOR, URKA, AKRN, NKNC
Финансы и банки	SBER, MOEX, VTBR, AFKS, BSPB, VZRZ
Машиностроение	SVAV, KMAZ
Потребительские товары и торговля	MGNT, LNTA, MVID, GCHE
Строительство	PIKK, LSRG, MSTT
Телекоммуникация	MTSS, RTKM
Транспорт	AFLT, NMTP

Таблица 7. Тикеры используемых в качестве исходных данных компаний по отраслям экономики.

Источник: составлено автором на основе данных Мосбиржи.

2.2. Классические детерминанты ожидаемой доходности акции

2.2.1. Безрисковая ставка

В качестве прокси переменной безрисковой ставки можно использовать ряд доступных альтернатив. Так, например, можно использовать ставку RUONIA, которая при использовании ежедневных данных позволяет соблюсти требование о равенстве горизонтов инвестирования. Однако операции по данной ставке доступны лишь институциональным инвесторам.

Ввиду данной проблемы, в качестве безрисковой ставки доходности использовалась месячная ставка из кривой бескупонной доходности Московской биржи (G-Curve, КБД)⁷ в непрерывном начислении процентов, рассчитанная для каждого торгового дня (См. Рисунок 1)⁸. КБД Московской биржи имеет преимущество перед другими аналогами оценки безрисковой ставки процента. Одним из них является большая стабильность над ставкой межбанковского рынка кредитования. Она, в целом, лучше отражает экономическую ситуацию в стране.



Рис. 1. Динамика месячной бескупонной доходности Московской биржи за период с 13.08.2010 по 04.06.2020.

Источник: составлено автором на основе данных с Мосбиржи.

⁷ Стоит отметить, что в 2018 году методология расчёта КБД Московской биржи несколько изменилась, тем не менее, существенного влияния на результаты исследования это оказать не должно.

⁸ Непрерывность ставки была рассчитана по принципу натурального логарифма отношения цен актива за два соседних торговых дня.

Ставка RUONIA склонна к большей волатильности в периоды, когда нужна ликвидность, что является её минусом. Кроме того, инвестирование в ОФЗ, на базе котировок которых строится КБД, доступно частным инвесторам.

Проведённый анализ показал, что расчёт доходности с 30-дневным лагом позволяет избавиться от ежедневного шума в данных, достичь их большей стабильности и способствует получению лучших результатов.

2.2.2. Премия за рыночный риск

В качестве данных для расчёта рыночной премии за риск была использована рассчитанная для каждого дня разность месячной доходности индекса московской биржи (IMOEX) и месячной безрисковой ставкой процента, описанной выше (См. Рисунок 2).

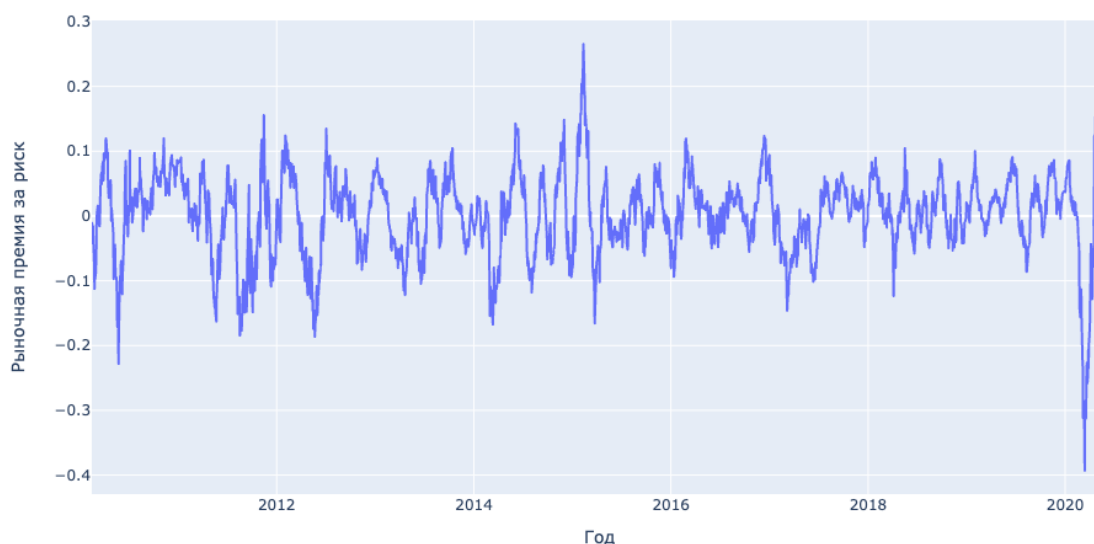


Рис. 2. Динамика месячной премии за рыночный риск за период с 13.08.2010 по 04.06.2020.

Источник: составлено автором на основе данных из терминала Bloomberg.

2.2.3. Факторы SMB, HML, MOM

Расчёты факторов SMB, HML, MOM из классической модели Фамы-Френча-Кархарта были произведены на основе данных Московской фондовой биржи и системы Bloomberg по 48 компаниям (см. Таблицу 7). Необходимые вычисления были произведены в соответствии с методологией в классическом варианте, предложенном в работе Фамы и Френча. Все доходности компаний при расчёте факторов входили в расчёт с соответствующими им весами — рыночными капитализациями компаний.

Согласно теории для вычисления классических факторов, Фама и Френч формировали шесть портфелей ценных бумаг в зависимости от размера компании по капитализации и её стоимости (соотношения Е/Р). Описательная статистика для доходностей этих портфелей приведена в Таблице 8.

	<i>S/L</i>	<i>S/M</i>	<i>S/H</i>	<i>B/L</i>	<i>B/M</i>	<i>B/H</i>
<i>Среднее</i>	0,018	0,004	-0,006	0,02	0,011	0,002
<i>Ст. откл.</i>	0,083	0,08	0,1	0,059	0,078	0,077
<i>Минимальное</i>	-0,34	-0,339	-0,5	-0,266	-0,439	-0,49
<i>25%</i>	-0,024	-0,032	-0,052	-0,016	-0,03	-0,044
<i>50%</i>	0,021	0,009	0,005	0,016	0,019	0,007
<i>75%</i>	0,064	0,047	0,053	0,054	0,06	0,049
<i>Максимальное</i>	0,334	0,361	0,308	0,303	0,304	0,294

Таблица 8. Описательная статистика по доходностям построенных 6 портфелей для расчёта факторов из трёхфакторной модели Фамы-Френча.

Источник: составлено автором на основе данных из терминала Bloomberg.

Анализируя полученные значения описательных статистик, можно заметить относительную устойчивость доходностей, что демонстрируют показатели квантилей, также стоит отметить невысокое значение стандартного отклонения, что свидетельствует об относительной стабильности доходностей портфелей.

Иллюстрация рассчитанных факторов классических моделей доходности Фамы-Френча-Кархарта приведены в приложениях (см. Приложение 23). На основе доходностей портфелей из таблицы 8 формируются значения классических факторов.

	<i>SMB</i>	<i>HML</i>	<i>MOM</i>
<i>Среднее</i>	-0,004	0,019	0,115
<i>Ст. откл.</i>	0,046	0,059	0,036
<i>Минимальное</i>	-0,140	-0,165	0,045
<i>25%</i>	-0,036	-0,019	0,091
<i>50%</i>	-0,004	0,011	0,109
<i>75%</i>	0,028	0,051	0,132
<i>Максимальное</i>	0,135	0,439	0,344

Таблица 9. *Описательная статистика по классическим факторам из модели Фамы-Френча-Кархарта.*

Источник: составлено автором на основе данных из терминала Bloomberg.

Невысокие значения дисперсий для факторов могут положительно сказаться на качестве статистик в линейных моделях (см. Таблицу 9). Это даёт надежду на получение значимых коэффициентов при построении классических моделей, используя линейные регрессии. Также, дополнительный анализ показал, что проблема мультиколлинеарности в модели не наблюдается, так как значения VIF для каждого фактора не превышают 10 (См. Таблицу 10). Таким образом, проблемы, связанные с большими и нестабильными ошибками коэффициентов, являющиеся следствием мультиколлинеарности, не мешают анализу.

Фактор	VIF
<i>SMB</i>	1,019
<i>HML</i>	1,142
<i>MOM</i>	1,158

Таблица 10. *Иллюстрация теста на мультиколлинеарность факторов классической модели Фамы-Френча-Кархарта.*

Источник: составлено автором на основе данных из терминала Bloomberg.

Для дополнительного понимания зависимостей между классическими факторами в приложениях представлена корреляционная матрица (См. Приложение 24). Она демонстрирует слабую корреляцию между классическими факторами.

2.3. Факторы, существенные для российского фондового рынка

Среди дополнительных факторов, потенциально способных влиять на ожидаемые доходности акций на российском рынке, можно выделить динамику курса доллара и цен на нефть. И цены на нефть, и курс рубля к доллару США использовался при анализе в других работах, посвящённых изучению детерминант доходности на российском фондовом рынке, например, в исследовании Горяева [Goriaev A., 2004].

2.3.1. Цена на нефть

Для российского рынка характерна повышенная зависимость от цен на нефть. Динамика этих цен оказывает существенное влияние на величину доходов государственного бюджета РФ. Также нефтедобычей в России занимается ряд компаний, которые входят в число крупнейших по капитализации в России (ROSN, LKOH, SNGS, TATN). Таким образом, цены на нефть имеет смысл добавить в анализ как дополнительный объясняющий, специфический для российского рынка фактор. При этом, чтобы работать со стационарными временными рядами, имеет смысл сразу перейти к первым разностям, которые стационарны⁹.

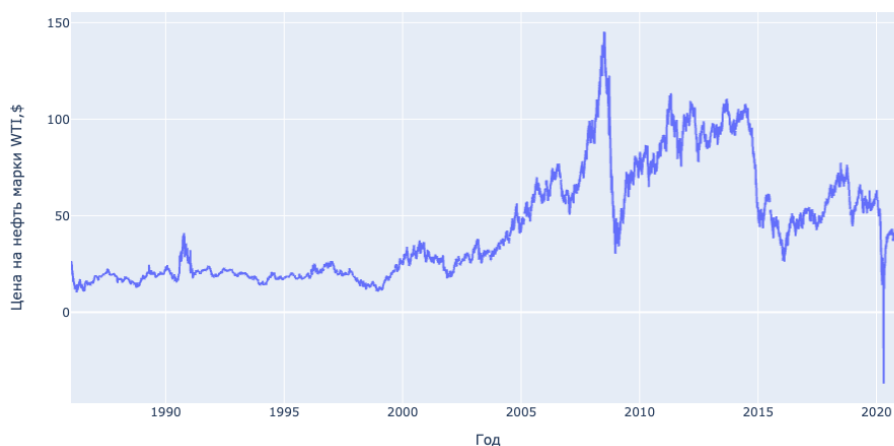


Рис. 3. Динамика стоимости нефти марки WTI.

Источник: составлено автором на основе данных с сайта *eia.gov*.

⁹ Этот вывод был получен после проведения расширенного теста Дики-Фуллера

2.3.2. Курс рубля к доллару США

Также для российской экономики существенным является влияние на фондовый рынок динамики валютного курса рубля к доллару США, так как значительная часть сделок осуществляется в долларовом выражении, многие доходы и расходы зависят от валютной составляющей, что не может не влиять на доходности компаний, которые используют долларовую валюту в своих операциях. Как следствие, валютный курс имеет смысл включить в анализ в качестве потенциальной дополнительной детерминанты ожидаемой доходности акций.

Дополнительно была проведена проверка на стационарность по расширенному тесту Дики-Фуллера. Его результаты показали стационарность первой разности для временного ряда курса рубль/доллар, с которыми далее и была построена работа.

2.4. Дополнительные факторы

При построении многофакторных моделей оценки ожидаемой доходности акции при расчёте детерминант доходности использовалась дополнительная информация о величине коэффициента отношения долга компании к её рыночной капитализации (D/E). Роберт Хамада [Hamada R. S. 1972] учёл тот факт, что наличие долга у компании сказывается на её доходности, и добавил корректирующий множитель в формулу оценки ожидаемой доходности компании. Однако при внедрении многофакторных моделей в анализ появился вопрос относительно использования корректирующего коэффициента, содержащего отношение долга к рыночной капитализации, для факторов модели Фамы-Френча-Кархарта. В работе [Lally M. (2004)] демонстрируется корректировка факторов коэффициента бета, перед рыночной премией за риск, а также коэффициентов при SMB и HML по формуле Р. Хамады, учитывающей долг компании. Приведём ниже эти формулы.

$$b = b_u \left(1 + \frac{D}{E} (1 - T_c) \right) \quad (8)$$

$$s = s_u \left(1 + \frac{D}{E} (1 - T_c) \right) \quad (9)$$

$$h = h_u \left(1 + \frac{D}{E} (1 - T_c) \right) \quad (10)$$

Где b – это коэффициент бета перед премией за рыночный риск,
индекс u означает показатель, соответствующий компании без долга,

D - рыночная стоимость долга,

E - рыночная стоимость капитала,

T_c - ставка корпоративного налога,

s – коэффициент чувствительности перед фактором SMB,

h – коэффициент чувствительности перед фактором HML.

В рамках практической реализации факт наличия долга будет учтён именно этим образом. Так как модели строятся в терминах ожиданий, то предполагается, что агенты рациональны и не делают систематических ошибок, тогда фактическая доходность - прокси ожидаемой, поэтому надо использовать отношение D/E на начало периода, поэтому при расчётах был сделан лаг в 1 месяц (30 дней) для соотношения долга к рыночной капитализации в множителе. Показатель налогов на прибыль соответствует ставке 20% согласно законодательству РФ.

Авторегрессионная составляющая нередко позволяет достичь более качественных результатов, поэтому в модель были включены лаги зависимой переменной и регрессоров. После нескольких итераций анализа значимости коэффициентов были получены результаты, иллюстрирующие, что в среднем значимость пропадает после трёхмесячного лага в связи с чем был выбран именно этот порог для факторов и зависимой переменной.

2.5. Тестирование классических и расширенных моделей.

Данный параграф посвящен оценке и тестированию следующих линейных моделей ожидаемой доходности акций на российском фондовом рынке:

- CAPM
- Трёхфакторная модель Фамы-Френча
- Четырёхфакторная модель Кархарта
- Расширение четырёхфакторной модели Кархарта с помощью добавления лагов классических факторов SMB, HML, MOM
- Расширение четырёхфакторной модели Кархарта с помощью добавления лагов классических факторов SMB, HML, MOM и автолагов доходности анализируемой акции
- Расширение четырёхфакторной модели Кархарта с помощью добавления лагов классических факторов SMB, HML, MOM, автолагов доходности анализируемой акции и специфических для России факторов – изменение курса рубль/доллар и цен на нефть

При переходе от CAPM к многофакторной модели доходности Фамы-Френча-Кархарта наблюдается небольшое улучшение в метриках качества моделей, что подтверждается теорией, рассмотренной в первой главе, заметное улучшение в качестве

появляется при расширении модели Кархарта с помощью лагов и автолагов. Отметим, что все улучшения в качестве с точки зрения исправленного R^2 монотонны.

В рамках этой части в качестве примера проведём рассуждения относительно одной из компаний – Сбер (см. Таблицу 11), для других российских компаний ход рассуждений будет такой же. Так, например, для акций Сбера исправленный R^2 в 4-факторной модели Кархарта принимает значение 0.409, тогда как в модели CAPM он составляет 0.403¹⁰. При переходе к дальнейшим расширениям качество модели, оценённое по исправленному R^2 значительно выросло. Во-вторых, в приложении представлены F-статистики значимости уравнений в целом, все они значимы при любом разумном уровне значимости. В-третьих, можно обратиться к статистической значимости каждого из коэффициентов. Мы видим, что в большинстве случаев практически все коэффициенты значимы при любом разумном уровне значимости¹¹. В-четвёртых, значения среднеквадратической ошибки моделей становятся меньше при переходе к расширенным моделям, что является хорошим сигналом улучшения качества оценивания. В-пятых, дополнительный анализ показал, что наблюдается улучшение информационного критерия Шварца, который учитывает в себе значение функции правдоподобия модели и штрафует за увеличение сложности модели в виде роста числа регрессоров.

Также стоит отметить, что в линейных моделях использовались робастные в условиях гетероскедастичности стандартные ошибки коэффициентов модели (НС3).

В целях предупреждения ложной регрессии в модели были проведены соответствующие тесты на стационарность зависимой переменной и регрессоров. Расширенный тест Дики-Фуллера подтвердил стационарность временных рядов SMB, HML, MOM, Премии за рыночный риск при любом разумном уровне значимости. Ввиду полученных результатов можно не беспокоиться о ложной регрессии в модели (См. Приложение 25).

¹⁰ Имеет место сравнивать между собой модели именно по скорректированному коэффициенту R^2 ($\text{adj } R^2$), так как если рассматривать просто R^2 , то при увеличении количества регрессоров этот показатель точно не уменьшится в силу его построения.

¹¹ Полагаем под разумными уровнями значимости стандартные значения в 1%, 5% и 10%

	CAPM	Фама- Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.444***	0.421***	0.422***	0.427***	0.435***	0.434***
доходность Сбера лаг месяц					0.053***	0.068***
доходность Сбера лаг 2 месяца					0.080***	0.129***
доходность Сбера лаг 3 месяца					0.085***	0.058***
SMB		0.081***	0.080***	0.054**	0.045**	0.063***
HML		-0.021**	-0.019	0.000	0.008	0.054***
SMB лаг месяц				0.055***	0.052***	0.045***
HML лаг месяц				-0.133***	-0.128***	0.055***
SMB лаг 2 месяца				0.068***	0.064***	0.047***
HML лаг 2 месяца				-0.021	-0.009	0.031**
SMB лаг 3 месяца				0.029*	0.007	0.049***
HML лаг 3 месяца				0.003	0.022*	0.025*
MOM			-0.011	0.081*	0.065	0.061*
MOM лаг месяц				-0.450***	-0.473***	-0.117**
MOM лаг 2 месяца				-0.030	-0.023	-0.199***
MOM лаг 3 месяца				0.429***	0.434***	0.272***
Изменение курса доллара за месяц						-0.761***
Изменение цены на нефть за месяц						0.015
R ²	0.403	0.410	0.410	0.545	0.555	0.680
Adjusted R ²	0.403	0.409	0.409	0.542	0.551	0.677
MSE	0.006	0.006	0.006	0.005	0.004	0.003

Таблица 11. Пример регрессий для Сбербанка.

Источник: расчеты автора

При анализе использовалась корректировка для коэффициентов при факторах рыночной премии за риск, SMB, HML по формуле Р. Хамады, учитывающей долговую

нагрузку компаний. Количество наблюдений в регрессиях соответствует 1979 наблюдений.

Для некоторых других компаний результаты классических моделей и их расширений приведены в приложениях, также в приложении доступны ссылки на данные и код, которые можно использовать, чтобы получить результаты регрессий для любой из 48 анализируемых компаний.

Однако стоит отметить и некоторые недостатки данной части исследования. При анализе приложений или при самостоятельном построении регрессий на представленных данных можно заметить, что ряд коэффициентов недостаточно устойчив к разным спецификациям модели. Они могут приобретать значимость при добавлении новых регрессоров или менять знак. Таких примеров не очень много, но они есть. В ответ на это можно сказать, что они, как правило, очень близки к нулю по абсолютному значению, в связи с чем возможность смены знака повышается. Кроме того, можно сказать, что это характерно для ситуации, когда для классических факторов добавляются лаги, что несколько меняет значение коэффициента. Но тем не менее, оценка коэффициента при ключевом показателе – премии за рыночный риск – остаётся устойчивой к спецификациям модели, как и многие другие коэффициенты.

Для некоторого общего понимания, насколько произошло улучшение модели проиллюстрируем динамику R_{adj}^2 и MSE в зависимости от модели в среднем для всех 48 компаний для каждой модели (см. Рисунок 4).

В качестве дополнительной проверки устойчивости результатов стоит учесть влияние шока в связи с Covid-2019. В рамках анализа использовалась булевская переменная, которая принимала значения 0 до даты 21-02-2020 и 1 после неё, считая начало шока моментом начала паники на мировых рынках. Примеры результатов расчета с учетом фактора пандемии представлены в Приложениях 9-11. Можно сказать, что в среднем влияние пандемии оказало значимое влияние на доходности акций, хотя есть и исключения. При этом наблюдается устойчивость полученных ранее значимых коэффициентов.

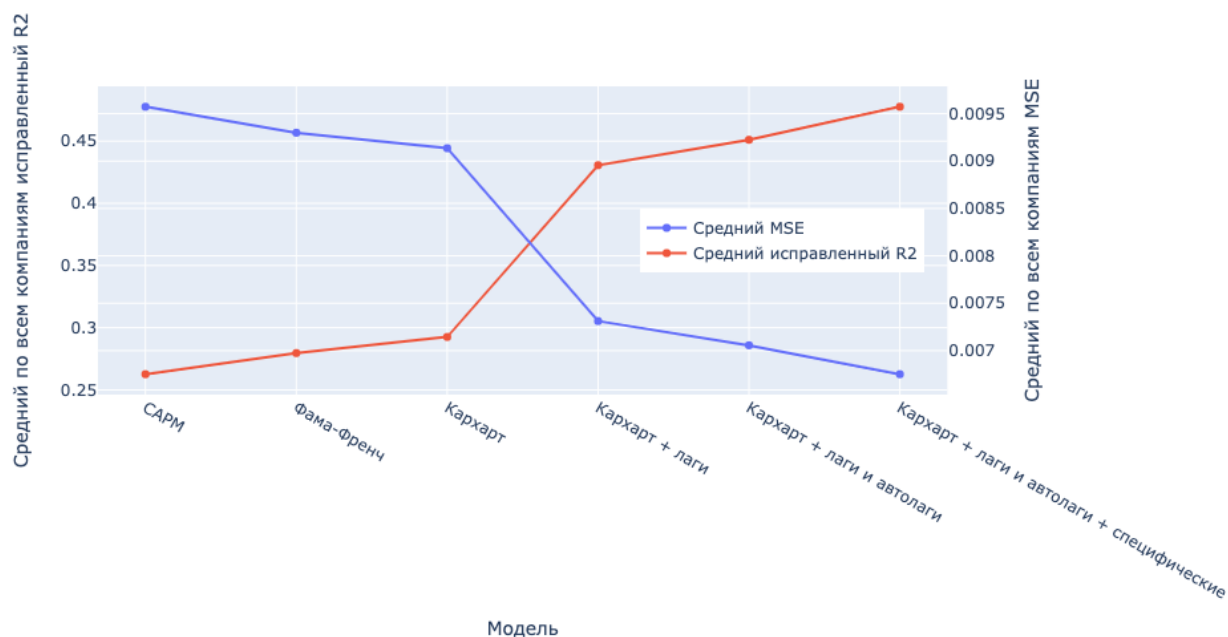


Рис. 4. Усреднённые по всем компаниям R_{adj}^2 и MSE для классических моделей и их расширений.

Источник: составлено автором основе данных из терминала Bloomberg.

2.6. Выводы по главе

С использованием данных по 48 российским компаниям были построены и оценены как классические факторные модели доходности, так и их модификации, учитывающие дополнительные факторы: классические CAPM, модель Фам-Френча, модель Кархарта и расширения над ними с использованием лагов классических факторов, лагов зависимой переменной и добавлением специфических для Российской Федерации факторов – цены на нефть и курса доллара. Классические факторы строились с учётом более устойчивых месячных доходностей, которые взвешивались по рыночной капитализации компаний. Дополнительно во всех моделях была учтена долговая нагрузка компаний по формуле Р. Хамады.

Результаты оценки регрессий демонстрируют значительное улучшение качества моделей. В среднем по российским компаниям наблюдается заметный рост R_{adj}^2 и снижение среднеквадратической ошибки модели при расширении классических моделей, таким образом расширения имеют смысл даже при использовании линейных моделей.

Глава 3. Многофакторные модели доходности акции на основе методов машинного обучения

В данной главе проводится оценка влияния анализируемых факторов на доходность акций с использованием методов машинного обучения. Первый параграф главы посвящен краткому обзору используемых в работе ML-методов. Во втором параграфе продемонстрируем результаты работы моделей на обучающей и тестовой выборке, сравним их между собой и с классическими многофакторными моделями линейной регрессии из предыдущей главы. Из обученных моделей на основе методов машинного обучения получим количественные оценки важности факторов, сравним их между собой на основании разных алгоритмов, а также на основе различных отраслей экономики, а также проведём анализ устойчивости полученных результатов. Дополнительно в качестве примера практического использования полученных результатов продемонстрируем построение торговой рекомендательной стратегии.

3.1. Теоретические основы алгоритмов машинного обучения

3.1.1. Метод k ближайших соседей. kNN (k Nearest Neighbors)

kNN представляет собой метрический алгоритм, базовой идеей которого является сравнение расстояний. Алгоритм рассматривает наиболее похожие объекты выборки с точки зрения дистанции друг от друга и предполагает для близких объектов схожие результаты. Основным параметр метода – это число таких соседей, с которыми сравнивается объект, за него отвечает гиперпараметр k . Если k соответствует единице, то алгоритм ищет расстояние до единственного ближайшего соседа для каждого объекта и сравнивает с его характеристиками для прогноза будущих результатов. Если число k соответствует пяти, то выбирается среднее из 5 ближайших соседей и сравнивается уже с этим значением и т.д. Стандартной проблемой данного метода является проклятие размерности, что выражается в значительном увеличении требуемых мощностей для анализа массивных многомерных данных. Этот алгоритм используется относительно нечасто в задачах регрессии, так как часто правило уступает по качеству другим, более продвинутым, алгоритмам, но всегда бывают свои исключения, поэтому он тоже будет рассмотрен.

С идейной точки зрения регрессия на основе алгоритма kNN целесообразна в том смысле, что она позволяет улавливать похожие значения доходности для похожих факторов на основе расстояний между ними. Алгоритм в такой постановке вполне способен продемонстрировать хорошие результаты по предсказанию значений ожидаемой доходности акции. Так, например, по большому расстоянию между объектами внутри

факторов стоимости или размера алгоритм может выделять разные значения доходности, более соответствующие этой группе.

3.1.2. Lasso, Ridge (гребневая регрессия), Elastic Net

Алгоритм регрессии Lasso (least absolute shrinkage and selection operator) это линейная регрессия с дополнительным условием регуляризации для коэффициентов модели, то есть с учётом штрафа за большое значение весов коэффициентов. Отличие Lasso от Ridge регрессии в том, что в первом случае штраф в регрессии учитывается как сумма модулей значений весов, тогда как во втором случае сумма квадратов весов коэффициентов. Этот алгоритм регрессии за счёт своей архитектуры справляется с проблемой переобучения¹². У Lasso и у Ridge регрессии есть свои преимущества и недостатки. Для того, чтобы выбрать некоторый компромисс между ними используется Эластичная сеть – алгоритм регрессии, который учитывает с некоторым весом α регуляризацию из Lasso и с весом $(1 - \alpha)$ регуляризацию из Ridge регрессии, где α – это настраиваемый гиперпараметр модели. Эти модели активно используются в эконометрике, в частности они помогают бороться с мультиколлинеарностью.

Отличие от обычной линейной регрессии заключается в слагаемом, которое отвечает за регуляризацию.

Для гребневой регрессии функция потерь и соответствующая задача её минимизации представима в виде:

$$\min_{w_j} \left(\sum_{i=1}^N \left(y_i - \sum_{j=0}^k w_j x_{ij} \right)^2 + \lambda \sum_{j=0}^k w_j^2 \right) \quad (11)$$

Где N – это размер выборки,

k – количество коэффициентов в модели, включая константу,

w_j – j -ый коэффициент в регрессии

λ – константа регуляризации, являющаяся гиперпараметром модели.

Для регрессии LASSO функция потерь и соответствующая задача её минимизации представима в виде:

¹² Переобучение модели (алгоритма) – это термин из области машинного обучения, который означает настройку алгоритма под конкретные зависимости обучающей выборки, а не под действительные зависимости генеральной совокупности.

$$\min_{w_j} \left(\sum_{i=1}^N \left(y_i - \sum_{j=0}^k w_j x_{ij} \right)^2 + \lambda \sum_{j=0}^k |w_j| \right) \quad (12)$$

Данная группа методов подходит для подобного исследования, поскольку базируется на линейных моделях, являющихся классическими, при этом добавляя над ними надстройку в виде регуляризации, что помогает улучшать качество модели.

3.1.3. Решающее дерево (DT), Случайный лес (RF)

Решающее дерево – это пример логического алгоритма, который в самой простой постановке может быть записан через условие «Если А, то а, иначе b». Иначе говоря, это граф, представляющий из себя бинарное дерево, где каждой внутренней вершине присвоена некоторая функция, которая каждому листу (листовой вершине R) даёт прогноз $y_R \in Y$, где Y – это множество значений целевой переменной.

Основной принцип работы решающего дерева сводится к выбору фактора, по которому будет произведено разбиение, и определения величины порога разбиения. Данная процедура происходит по принципу максимизации критерия расщепления.

$$Q(R, \theta) = H(R) - \frac{|R_l|}{|R|} * H(R_l) - \frac{|R_r|}{|R|} * H(R_r) \quad (13)$$

Где $H(R)$ – это мера неоднородности для листа R, в нашем случае для регрессии принимает вид среднеквадратичной ошибки (MSE), индексы l и r отвечают за левый и правый узел дерева, Q – критерий расщепления, θ – фактор, и его значение, по которому происходит разбиение на левый и правый узел дерева.

$$H(R) = \frac{1}{|R|} * \sum_{(x_i, y_i) \in R} (y_i - \hat{y}_i)^2 \quad (14)$$

R – это конкретная вершина дерева, $|R|$ – мощность множества R.

y_i – истинное значение целевой переменной, x_i – факторы,

\hat{y}_i – оценённое значение целевой переменной

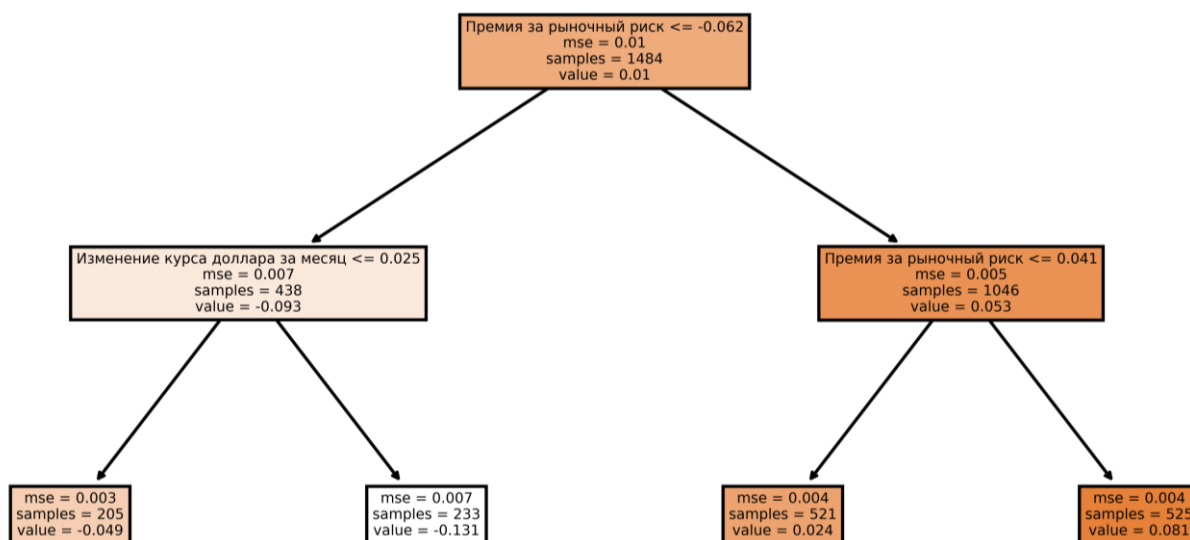


Рис. 5. Пример решающего дерева глубины 2, построенного для сбербанка.

Источник: расчеты автора.

Построение дерева осуществляется за счёт максимизации критерия расщепления (см. формулу 13). Случайный лес является примером построения ансамбля алгоритмов способом бэггинга (Bootstrap Aggregating), и схематично алгоритм построения случайного леса может быть представлен в виде следующих шагов:

- 1) Выбирается бутстрап-подвыборка¹³ наблюдений.
- 2) На этой подвыборке выстраивается решающее дерево на некотором случайном наборе факторов модели (регрессоров), разбиение происходит каждый раз оптимальным с точки зрения максимизации критерия расщепления образом, разбиение на каждом узле проводится на случайной подвыборке факторов.
- 3) Строится заранее заданное количество деревьев по принципу из шагов 1,2 и их результаты усредняются.

Для данного исследования алгоритм случайного леса подходит в силу того, что он по построению позволяет улавливать нелинейные зависимости в данных. Его архитектура позволяет по факторам формировать представления об ожидаемой доходности и при оценке модели параллельно вычислять важность всех факторов при обучении модели, что может

¹³ Случайная выборка наперед заданного объёма с возвращениями.

помочь в решении «factor zoo» проблемы. Также достоинством метода является простота в понимании и интерпретации.

У случайного леса есть преимущество перед решающим деревом. Он гораздо более устойчив к выбросам, так как каждый раз, во-первых, выбирается лишь часть исследуемых наблюдений и, во-вторых, лишь часть регрессоров. Такая модель гораздо менее склонна к переобучению. Важно правильно подбирать гиперпараметры модели так, чтобы модель была сбалансирована в смысле дилеммы между смещением и разбросом. Важность факторов в алгоритме решающего дерева, а также в алгоритмах бэггинга и бустинга над деревьями может определяться несколькими способами, перечислим наиболее популярные из них:

- 1) По количеству разбиений по данному фактору. Чем их больше, тем важнее фактор.
- 2) По среднему показателю критерия расщепления. Это соответствует среднему приросту в информационной определённости, при разбиении по данному фактору.
- 3) По местоположению в дереве, то есть «высоте» фактора. Смысл в том, что по построению дерева понятно, что прежде всего разбиение проходит по тем, факторам, которые вносят в данные больше определённости, таким образом, они располагаются выше по дереву.
- 4) Алгоритм Боруа. Этот алгоритм по очереди перемешивает значения единственного фактора и строит на полученной выборке случайный лес, сравнивая его результативность с результативностью работы случайного леса на первоначальной выборке. Этот процесс происходит с каждым фактором. В конечном итоге важность фактора тем больше, чем более упало качество работы модели при его перемешивании.

В пакетах, отвечающих за градиентный бустинг над деревьями (XGBoost, LightGBM, Catboost), есть методы, которые позволяют определить значимость каждого фактора модели. В рамках этих методов есть встроенные настраиваемые параметры, по которым производится расчёт важности факторов (ряд из них описан выше под пунктами 1-4). В данной работе в качестве основы будет использоваться информационный выигрыш (пункт 2). Данный подход является наиболее теоретически оправданным, так как в отличие от прочих способов этот способ более устойчив к результатам и напрямую связан с оптимизируемой метрикой – критерием расщепления.

В рамках данной работы используется атрибут `feature_importance_` из библиотеки `sklearn` для расчёта важности факторов¹⁴. В данном пакете важность фактора реализована как нормализованное совокупное увеличение критерия расщепления (уменьшение средней меры однородности) на основе конкретного фактора.

Расчёт важности факторов, реализованный в библиотеке `sklearn Python`, который будет использоваться в рамках изучения важности факторов, состоит из нескольких частей:

1) Расчёт важности узла j :

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \quad (15)$$

Где ni_j – Важность узла j ,

w_j – взвешенное число объектов выборки узла j ,

C_j – мера неоднородности листа j (MSE для листа j , так как работаем с задачей регрессии),

$left(j)$ – дочерний узел, соответствующий левой части разбиения узла j на две части,

$right(j)$ – дочерний узел, соответствующий правой части разбиения узла j на две части.

2) Расчёт важности фактора j :

$$fi_i = \frac{\sum_{j:\text{узел } j \text{ разбивается по фактору } i} ni_j}{\sum_{k \in \text{все узлы}} ni_j} \quad (16)$$

Где fi_i – это важность фактора i , ni_j – важность узла j .

3) Нормировка важности фактора.

$$normfi_i = \frac{fi_i}{\sum_{j \in \text{все факторы}} fi_j} \quad (17)$$

$normfi_{ij}$ – нормированная важность фактора i в дереве j .

4) Расчёт важности фактора по всем деревьям, то есть по всему случайному лесу

$$RFfi_i = \frac{\sum_{j \in \text{все деревья}} normfi_{ij}}{T} \quad (18)$$

Где T – это количество деревьев в случайном лесу,

$RFfi_i$ – важность фактора i , посчитанного по всему случайному лесу.

¹⁴ [https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html?highlight=feature%20importance#sklearn.ensemble.RandomForestRegressor.feature_importances_)

[learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html?highlight=feature%20importance#sklearn.ensemble.RandomForestRegressor.feature_importances_](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html?highlight=feature%20importance#sklearn.ensemble.RandomForestRegressor.feature_importances_)

Получаемая по данному методу важность фактора – это число от нуля до единицы. Чем оно больше, тем больше значимость. Для простоты интерпретации можно рассуждать в терминах процентной шкалы от 0% (совсем не важен относительно других факторов) до 100% (важность прочих сопоставима с нулём относительно большой важности этого фактора). Также стоит отметить, что ввиду нормировки по построению все важности при суммировании дают единицу.

3.1.4. Метод опорных векторов (SVM)¹⁵

Метод опорных векторов представляет собой линейный метод, базирующийся на выводе теоремы Каруша-Куна-Такера, который может быть применён для решения задачи регрессии.

Постановка задачи регрессии на основе метода опорных векторов.

$$\begin{cases} \min_w \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \right) \\ |y_i - \langle w, x_i \rangle - b| \leq \xi_i \\ \xi_i \geq 0 \end{cases} \quad (19)$$

Где, w – это веса при регрессорах – коэффициенты модели,

C – константа регуляризации,

b – свободный член,

ξ_i – случайная ошибка модели

Достоинством данного метода является возможность перейти в спрямляющее пространство с помощью ядерной функции.

$$\varphi: X \rightarrow H, k(x, z) = \langle \varphi(x), \varphi(z) \rangle \quad (20)$$

Где H – спрямляющее пространство, $\langle \varphi(x), \varphi(z) \rangle$ – скалярное произведение функций, k – ядро, в котором данные могут быть описаны линейной функцией.

Обычно используется Гауссовское ядро, иногда полиномиальное, остальные – реже.

Метод опорных векторов является одним из классических алгоритмов для решения задач регрессии, которые используются в машинном обучении. Важным преимуществом, почему этот алгоритм может показывать хорошие результаты в приложении к многофакторному анализу ожидаемой доходности – это его способность перейти в спрямляющее пространство за счёт ядерных функций и найти нелинейные зависимости в данных, предложив решение проблемы нелинейностей.

¹⁵ SVM – Support vector machine

3.1.5. Градиентный бустинг (Gradient Boosting, GB)

Градиентный бустинг – это последовательное итеративное обучение простых алгоритмов с некоторыми весами.

Финальный вид модели:

$$a_N(x) = \sum_{j=0}^N \gamma_j b_j(x) \quad (21)$$

Где $a_N(x)$ – результирующий алгоритм,

γ_j – веса базовых алгоритмов,

$b_j(x)$ – простые алгоритмы (деревья).

$$b(x) = \sum_{t=1}^T w_t [x \in R_t] \quad (22)$$

Где, w_t – значение, которое ставится в соответствие листу с индексом t ,

R_t – лист дерева, который имеет индекс t

Основной принцип градиентного бустинга – это итеративный подбор алгоритмов, которые обеспечивают последовательно лучшее качество с точки зрения наперёд заданной функции потерь.

Очень схематично процесс обучения градиентного бустинга можно изобразить в три шага:

- 1) Вычисление антиградиентов функции потерь в точках композиции предыдущего шага:

$$s_i^{(N)} = -\frac{\partial}{\partial z} \mathcal{L}(y, z)|_{z=a_{N-1}(x_i)} \quad (23)$$

Где, \mathcal{L} – функция потерь, $\frac{\partial}{\partial z}$ – частная производная по z , $|_{z=a_{N-1}(x_i)}$ – в точке z

принимавшей значение $a_{N-1}(x_i)$, $s_i^{(N)}$ – это ошибка на N итерации,

a_{N-1} – алгоритм на $N-1$ итерации

- 2) Настройка очередного дерева:

$$b_N(x) = \underset{b \in \text{Множество решающих деревьев}}{\operatorname{argmin}} \sum_{i=1}^n \left(b(x_i) - s_i^{(N)} \right)^2 \quad (24)$$

- 3) На этом шаге происходит обновление констант в листьях дерева в соответствии с функцией потерь

$$w_{Nt}^*(x) = \operatorname{argmin}_{w \in R} \sum_{i: (x_i, y_i) \in R_t} \mathcal{L}(y_i, a_{N-1}(x_i) + w) \quad (25)$$

Градиентный бустинг – алгоритм, который занимает доминирующее положение в анализе табличных данных, с большой долей вероятности лучшая модель является градиентным бустингом над деревьями.

В рамках данного исследования алгоритмы градиентного бустинга уместны по всем причинам, которые были описаны для алгоритма случайного леса, поскольку они имеют схожую природу (за исключением простоты и интерпретации). Но вместе с тем, данный алгоритм позволяет последовательно настраивать деревья, учитывая ошибки предыдущих алгоритмов, на что не способен решающий лес. При решении задач регрессии градиентный бустинг как правило показывает себя очень хорошо, что даёт повод ориентироваться на результаты такой модели, результат которой довольно не просто достичь.

3.1.6. Нейронная сеть (Neural Network, NN)

Нейронная сеть – это ациклический ориентированный граф, в котором вершины – это вводные данные или нейроны, а рёбра – это какие-то преобразования (См. Приложение 26). Основным достоинством нейронной сети является способность улавливать сложные нелинейные зависимости.

Как правило нейронная сеть – это лучшая модель при работе с рядами данных, в которых есть взаимосвязь между собой, например временной ряд, поэтому в качестве одной из альтернатив она применяется для оценки влияния анализируемых факторов на доходности акций российских компаний в данном исследовании. В рамках данной работы для анализа из класса нейронных сетей был взят довольно простой представитель – многослойный персептрон.

3.2. Применение алгоритмов ML в моделировании доходности акции

В отличие от однофакторной модели CAPM, многофакторная модель позволяет более гибко улавливать воздействие прочих факторов на доходность ценной бумаги, делает возможным анализ линейных чувствительностей. Однако данные модели предполагают линейность зависимостей, это является недостатком. В общем случае предположение о линейности модели является неверным без специального анализа. Говоря о реальном финансовом мире, правильнее утверждать наличие нелинейных зависимостей. Вследствие данного факта открывается поле для их анализа, отправной точкой которого можно сделать набор факторов из многофакторной модели Фамы-Френча-Кархарта в силу серьёзных фундаментальных аргументов в пользу их использования.

В исследовании Симонян [Simonian J. et al, 2019] берется за основу данная идея и проводится тестирование на классических факторах более сложных моделей квантильной регрессии и случайного леса. Такой подход позволил точнее описать поведение данных в хвосте распределения доходностей, что позволяет лучше справляться с проблемой толстых хвостов эмпирического распределения доходностей.

Многие исследователи применяют современные методы машинного обучения для моделирования сложных зависимостей и в попытке наилучшим образом предсказать динамику стоимости акции. Одним из таких примеров, является работа Патель [Patel, 2015], где автор демонстрирует сравнительную эффективность предсказаний нейронной сети, метода опорных векторов, случайного леса и наивного Байесовского классификатора. Диалло [Diallo, 2019] применил регрессию на основе метода опорных векторов и байесовскую оптимизацию на основе факторов 3-х и 5-и факторной модели Фамы-Френча для анализа доходностей различных компаний.

Брызгалова и соавторы [Bryzgalova et al., 2019] подтвердили важность учёта нелинейности при оценке многофакторных моделей доходности, в качестве инструментария авторы использовали случайный лес. Такие продвинутое модели как глубокие нейронные сети, градиентный бустинг и случайный лес показывают хорошие результаты при выстраивании стратегий статистического арбитража на базе анализа доходностей компаний, входящих в индекс S&P 500 [Krauss et al., 2017].

Данные факты говорят в пользу того, что продвинутое статистические и эконометрические модели и методы машинного обучения активно применяются многими исследователями и часто показывают достойные внимания результаты. Они способны решить классическую проблему учёта нелинейных связей в финансовых данных или, по крайней мере, продвинуться в её решении. Более того, некоторые исследователи подвергают критике классические линейные модели [Bartholdy J., Peare P, 2005], демонстрируя, что объясняющая способность CAPM и многофакторных моделей доходности составляет лишь 3 – 5%. Использование более продвинутое подходов является возможным ответом на данную критику ввиду вышеупомянутых возможностей этих моделей.

Важное место в исследовании детерминант доходности акций занимают исследования, посвященные анализу объясняющей силы факторов. Так, в работе Коннор [Connor G., 1995] подобный анализ произведен на американских данных на основе линейной модели через сравнение отношений дисперсий в модели с присутствием фактора и при его отсутствии. Автор разбивает факторы на три группы: макроэкономические,

статистические и фундаментальные. Данная работа показывает наличие интереса к важности отдельных факторов, однако подобный подход базируется на линейных моделях, что не учитывает нелинейные зависимости, вследствие чего может давать несколько отличные оценки реального влияния отдельных факторов от действительных. Также в рамках данной работы важным замечанием является оценка данных показателей на американском рынке, в то время как особый интерес представляет изучение важности факторов на российском фондовом рынке с учётом возможных нелинейных взаимосвязей между факторами и доходностями.

Ниже приведены конкретные примеры методов, которые применялись разными исследователями в своих работах (см. Таблицу 12). для отбора факторов, анализа их важности, предсказательного анализа будущей динамики цены акции.

Методы анализа	Статьи
Метод опорных векторов (SVM)	Patel J. et al. 2015, Chen Y., Hao Y. 2017, Abe M., Nakayama H. 2018, Diallo B., Bagudu A., Zhang Q. 2019
Lasso, Ridge, Elastic net	Feng G., Giglio S., Xiu D. 2017, Messmer M., Audrino F., 2017; Freyberger J. et al, 2020;
Решающее дерево (DT), Случайный лес (RF)	Patel J. et al. 2015, Krauss C. et al 2017, Abe M., Nakayama H. 2018, Bryzgalova S., Pelger M., Zhu J. 2019, Simonian J. et al, 2019
К ближайших соседей (kNN)	Chen Y., Hao Y. 2017
Градиентный бустинг (XGBoost, LightGBM, Catboost)	Krauss C. et al 2017, Gu S., Kelly B., Xiu D., 2018
Нейронная сеть (NN)	Patel J. et al. 2015, Di Persio L., Honchar O. 2016, Krauss C. et al 2017, Abe M., Nakayama H. 2018 ¹⁶ , Gu S., Kelly B., Xiu D. 2018, 2021, Montavon G. et al. 2019

Таблица 12. Применение методов машинного обучения в современных исследованиях в области отбора факторов, выяснения их важности и предиктивной аналитики.

Источник: составлено автором на основе анализа научных статей

¹⁶ В данной работе автор делает вывод о преимуществе алгоритма нейронной сети над базовыми алгоритмами случайного леса и метода опорных векторов, которые формируют некоторый основной результат, от которого отталкивается исследователь.

Одной из важных причин, почему стоит рассматривать несколько алгоритмов машинного обучения для решения задач – это проблема отсутствия единого наиболее оптимального алгоритма. Если бы такой алгоритм существовал, то он вытеснил бы все прочие, однако в действительности применяется широкое разнообразие методов. Теоретическая почва данной проблемы заложена в работе Волперта и Макреди [Wolpert D. H., Macready W. G., 1997], где была сформулирована теорема, известная под названием «No free lunch theorem», что можно перевести, исходя из контекста, как теорема об отсутствии единственного наиболее эффективного алгоритма. Поэтому при анализе следует использовать несколько различных алгоритмов, чтобы получить возможность сравнить их результаты в конкретной задаче и выбрать наиболее оптимальный.

Особое место занимает работа Симонян [Simonian J. et al, 2019], поскольку в данной статье автор применяет метод значимости факторов на основе алгоритма случайного леса (См. таблицу 13). В рамках данного исследования будет использоваться данный метод оценки важности факторов модели, а также более продвинутые способы учета нелинейных зависимостей при оценке значимости факторов, такие как градиентный бустинг.

Индустрия	фактор			
	Rm-Rf ¹⁷	SMB	HML	MOM
Потребительские товары роскоши ¹⁸	0,84	0,04	0,05	0,07
Потребительские товары ¹⁹	0,5	0,2	0,15	0,15
Энергетика	0,52	0,13	0,18	0,17
Финансовый сектор	0,72	0,07	0,16	0,05
Здравоохранение	0,53	0,19	0,16	0,13
Промышленные предприятия	0,84	0,05	0,07	0,04
Информационные технологии	0,73	0,05	0,15	0,08
Сырьевой сектор	0,69	0,07	0,11	0,13
Услуги сектора телекоммуникаций	0,49	0,18	0,17	0,16
Коммунальные услуги	0,37	0,25	0,2	0,18

¹⁷ Рыночная премия за риск

¹⁸ В оригинале сектор называется Consumer Discretionary

¹⁹ В оригинале сектор называется Consumer Staples

Таблица 13. Сравнительная важность факторов на основе случайного леса в модели Фамы-Френча-Кархарта на рынке США (Январь 1991 - Август 2018).

Источник: Simonian J. et al, 2019

В работе Гу и соавторов [Gu et al., 2018] получены несколько другие оценки важности за счёт алгоритмов случайного леса и многослойного персептрона. Это можно объяснить во многом тем, что авторы использовали большое количество параметров, среди которых не было HML и SMB. Авторы наблюдают важность только у фактора моментум с лагом в 1 месяц (более 10%, что является наибольшим из полученных значений), если говорить о классических факторах. Также исследователи получили результат относительной неважности рыночной премии за риск. Данные различия можно объяснить спецификой рынков, а также несколько иной методологией расчёта важности факторов. В этой работе авторы вычисляют важность фактора на основе изменения доли объяснённой дисперсии с помощью модели.

3.3. Сравнение качества моделей ML и линейных расширений классических моделей

При построении моделей использовалась метрика среднеквадратичной ошибки в качестве меры эффективности. Это классическая мера для задач линейной регрессии, которая очень часто встречается в качестве способа измерения качества результатов работы модели. Для соответствия канонам и традициям машинного обучения в данном исследовании используется именно эта метрика.

Общая формула среднеквадратичной ошибки (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (27)$$

Где y_i – это зависимая переменная модели, \hat{y}_i – предсказание модели.

В качестве моделей для поставленной задачи были использованы метод опорных векторов, случайный лес, линейная регрессия (регрессия, построенная по методу наименьших квадратов; приведена здесь для сравнения результатов), метод К ближайших соседей, линейная регрессия с регуляризацией типа L1, линейная регрессия с регуляризацией типа L2, алгоритмы градиентного бустинга над деревьями и нейронная сеть.

После обучения перечисленных моделей были получены результаты²⁰, которые обобщены на Рисунке 6. Обобщение заключалось в усреднении метрик MSE для всех компаний по каждому алгоритму для обучающей и тестовой выборки соответственно. Можно сказать, что на тестовой части выборки лучше себя показали алгоритмы LASSO и Ridge. Заметим, что в целом результаты оказались сопоставимы по разным моделям, как по линейным, так и по нелинейным.

На Рисунке 7 представлено обобщение полученных результатов как проекция на каждую компанию, то есть в этом случае усреднение проходило по алгоритмам и высчитывался средний результат для каждой компании.

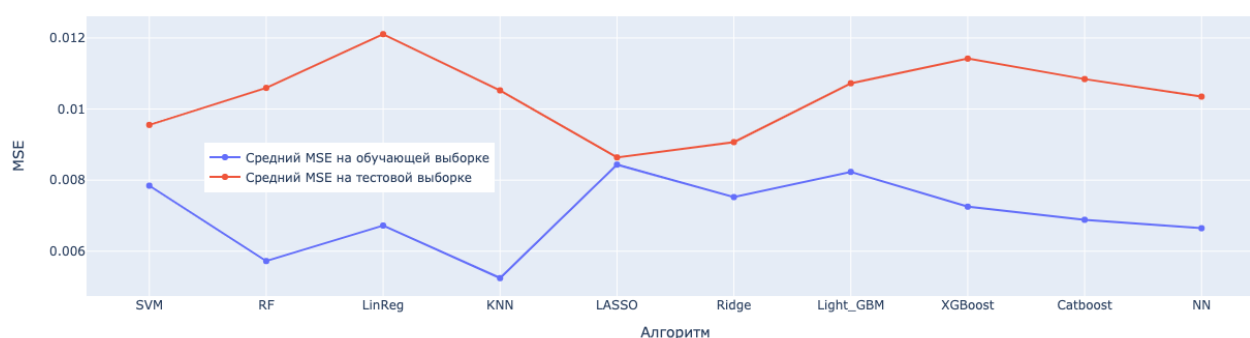


Рис. 6. Средняя MSE для всех алгоритмов на тестовой и обучающей выборках.

Источник: расчеты автора

²⁰ Все коды и данные расположены в открытом доступе <https://github.com/SergeyPetrakov/Diploma>

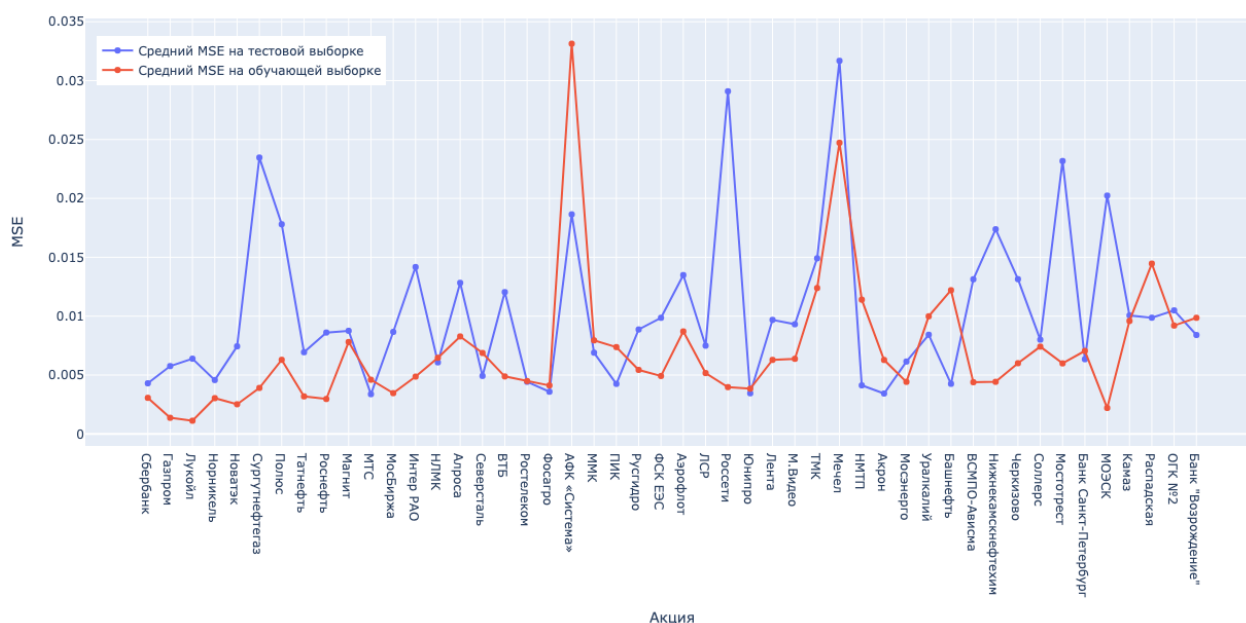


Рис. 7. Средняя результативность алгоритмов по MSE для каждой акции на тестовой и обучающей выборке.

Источник: составлено автором на основе данных из терминала Bloomberg.

Следует отметить, что наблюдается гетерогенность в полученных результатах для разных компаний. Есть компании, для которых показатель среднеквадратической ошибки мал как на обучающей, так и на тестовой выборках, например, Сбербанк, Норильский Никель или МТС. Есть примеры, когда относительно высоко значение MSE и для тестовой, и для обучающей выборки, например, АФК Система и Мечел являются яркими представителями такого случая. Помимо этого, встречаются компании, для которых значение на тестовой выборке на относительно значимую величину превосходят MSE на обучении.

Превышение качества работы модели на обучающей выборке над качеством на тестовой выборке в разумных пределах – закономерное явление. Однако, одной из важнейших проблем при обучении моделей является их переобучение, то есть ситуация, когда результативность на тестовой выборке заметно хуже, чем на обучении. Для предотвращения этого явления был проделан процесс настройки гиперпараметров²¹

²¹ Гиперпараметр – это параметр, который не обучается моделью в процессе её обучения, он задаётся экзогенно извне. Вопрос оптимальности настройки этих параметров непрост. Существует множество разных подходов к их настройке.

модели. Для этого использовалась процедура случайного поиска (Random search) среди наперёд заданного распределения параметров моделей. Эта процедура оценивает модель на случайной выборке параметров из заданных распределений с учётом кросс-валидации²². Из полученного множества результатов выбираются те значения гиперпараметров, которые показали себя лучшим образом. Весь этот процесс автоматизирован в библиотеке Python Sklearn²³. В результате, в процессе оптимизации гиперпараметров моделей был получен оптимальный с данной точки зрения набор для каждого ML алгоритма для каждой компании.

Архитектура²⁴ нейронной сети подбиралась отдельно от настройки гиперпараметров остальных моделей машинного обучения в силу особенностей метода. Одним из принципов построения было правило геометрической пирамиды из книги Мастера [Masters, 1993], также в качестве количества нейронов в скрытых слоях нейронной сети использовались числа, являющиеся степенями двойки [Gu et al., 2018, 2020]. После оптимизации её параметров путём перебора некоторого числа приемлемых вариантов была использована полносвязная нейронная сеть, с 7 скрытыми слоями, в которых было соответственно 4, 8, 32, 64, 32, 16, 1 нейрона, функции активации – везде гиперболический тангенс, функция потерь – MSE, оптимизатор – Adam, который является одним из самых эффективных алгоритмов оптимизации при обучении нейронных сетей, сочетающем в себе идеи и RMSProp и оптимизатора импульса.

Формула гиперболического тангенса, выраженная через экспоненту (e):

$$th(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (28)$$

Где x – это линейное преобразование над данными из нейронов предыдущего слоя.

²² Кросс-валидация – это проверка работы модели, используя несколько итераций с различными подвыборками данных.

²³ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

²⁴ Архитектура нейронной сети – это конструкция, которая задаёт её форму, сколько у неё слоёв, их связность, количество нейронов в каждом слое, функцию активации, оптимизатор.

3.4. Оценка важности факторов с помощью случайного леса

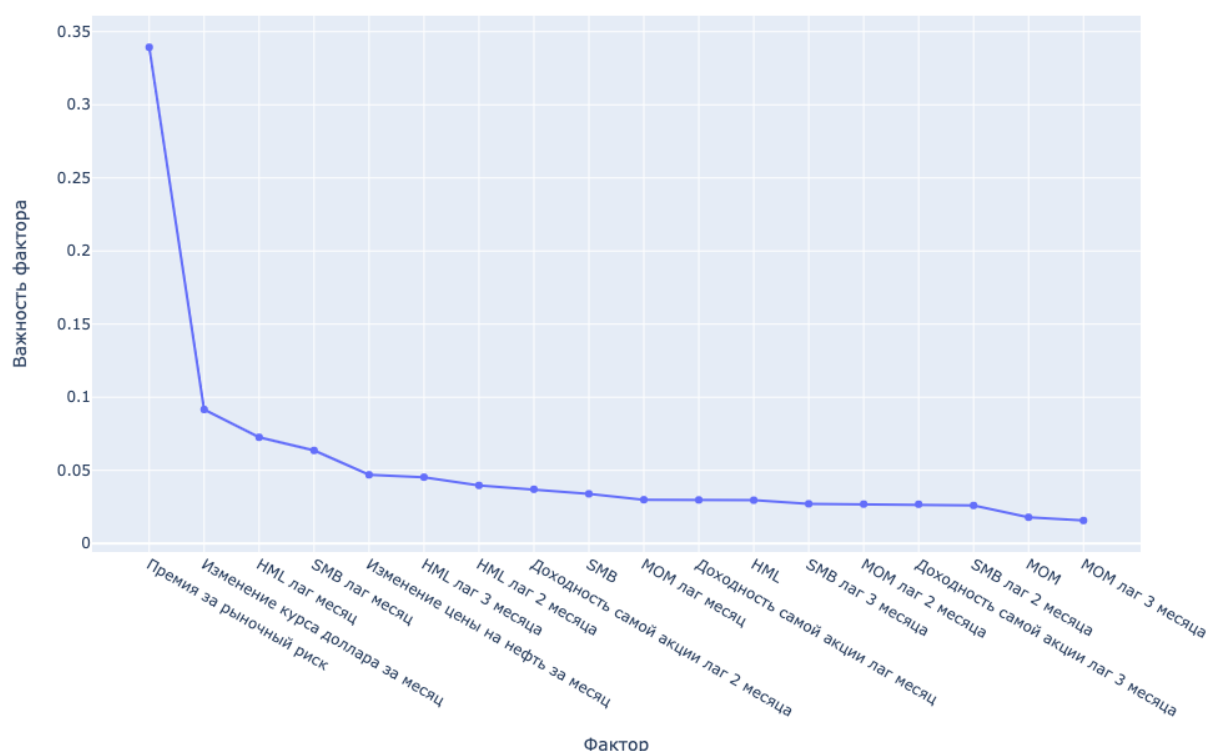


Рис. 8. Важность факторов для всего рынка, полученная при обучении случайного леса.

Источник: составлено автором на основе данных из терминала Bloomberg.

Анализ показал, что в среднем по всем компаниям большую часть важности при обучении случайный лес возлагает именно на премию за рыночный риск. Согласно полученным результатам, треть информации извлекается именно из неё (см. Рисунок 8). Стоит также отметить, что относительно прочих факторов заметно выделяются классические факторы из модели Фамы-Френча взятые с лагом в 1 месяц, а также специфические для России факторы – цены на нефть и курс рубля к доллару США. Их важность при построении модели была оценена в 5-10%. Можно сделать явный вывод в пользу того, что доминирующее значение для модели имеет премия за рыночный риск, однако степень важности меняется от 17 до 70% в зависимости от рассматриваемой отрасли (см. Рисунок 9).

Для отдельных отраслей есть примеры факторов, оказывающих значимое воздействие на ожидаемую доходность с точки зрения случайного леса, например, для химической отрасли особенное значение имеет динамика курса доллара за последний месяц.

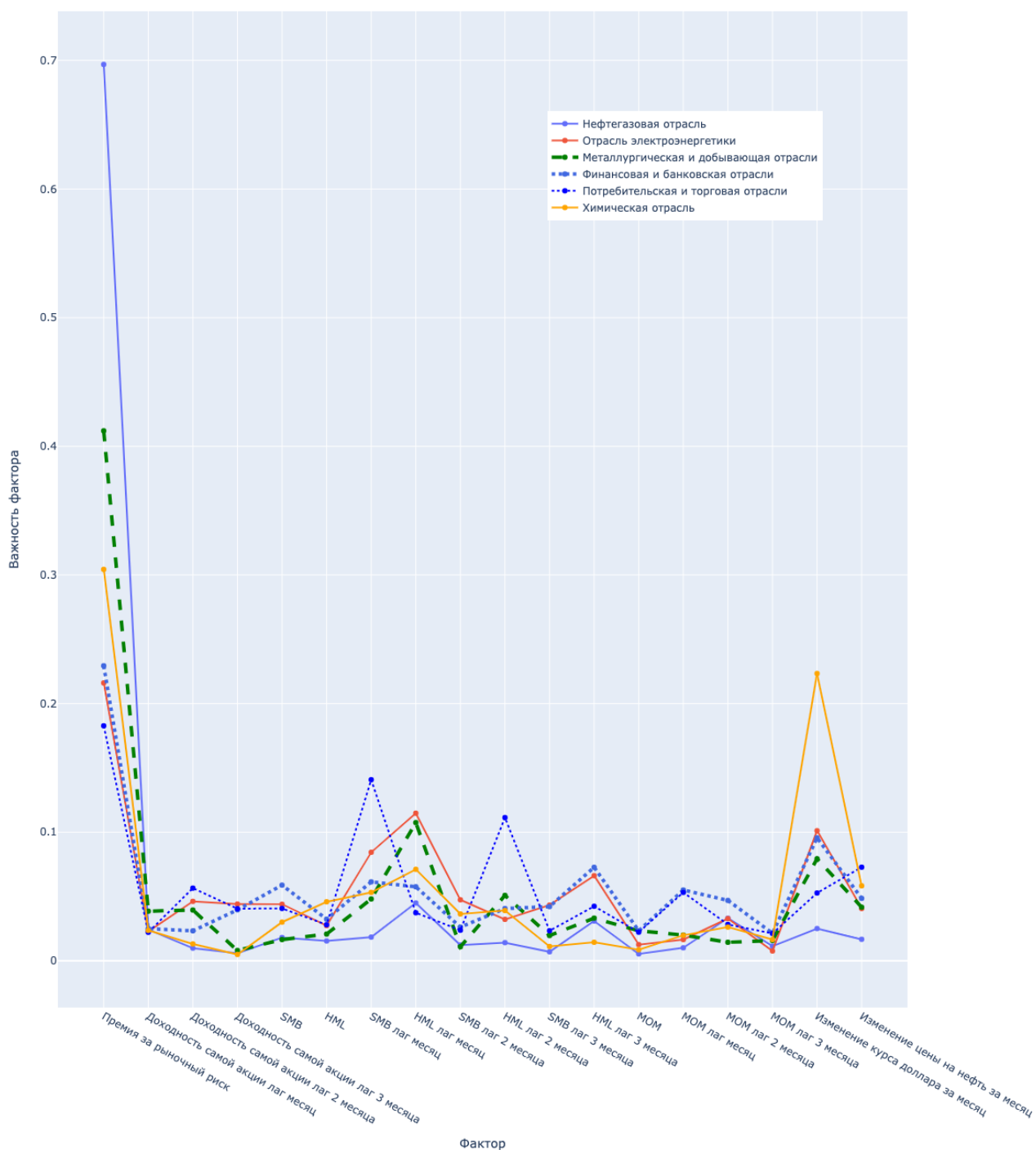


Рис. 9. Важность факторов, полученная при обучении случайного леса, для отраслей экономики, где есть хотя бы 4 компании в выборке. Источник: расчеты автора

В целом тенденция согласуется с базовой интуицией о большой значимости рыночной премии за риск и меньшей значимостью прочих факторов. Полученные результаты согласуются с показателями, поученными Симонян и соавторами [Simonian J.

et al, 2019]. Наблюдается разнотечение с результатами, полученными Гу и соавторами [Gu et al., 2018, 2021] о важности параметра импульса с месячным лагом - на российском рынке данных фактор оказывается не так важен.

3.5. Проверка устойчивости полученных результатов на алгоритмах градиентного бустинга

Для анализа устойчивости результатов проведём вычисления важности факторов с точки зрения алгоритмов градиентного бустинга, реализованного в вариантах XGBoost, LightGBM, CatBoost и сопоставим с важностью, полученной с помощью случайного леса.

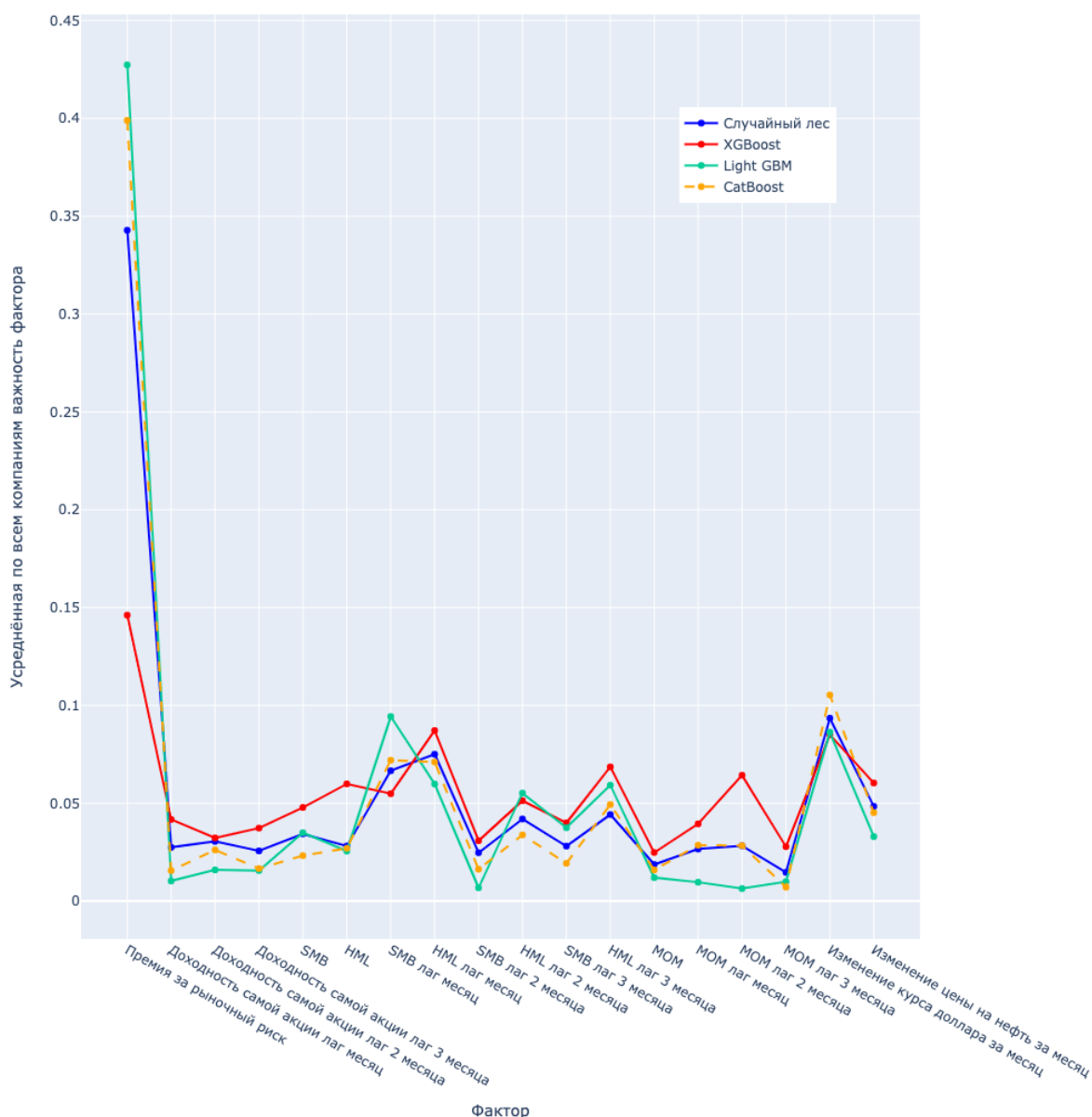


Рис. 10. Проверка устойчивости важности факторов на разных алгоритмах.

Источник: составлено автором на основе данных из терминала Bloomberg.

Везде использовался принцип максимизации среднего выигрыша от расщеплений в процессе построении деревьев²⁵. Как и прежде для каждой пары модель-акция используется набор гиперпараметров, которые оптимизируются при построении прогнозов для тестовой выборки.

На Рисунке 10 видно, что результаты устойчивы к разным типам моделей градиентного бустинга.

3.6. Отбор факторов с помощью LASSO

Полезным побочным результатом, который можно получить при использовании алгоритма LASSO является способ отбора факторов через L1 регуляризацию. Дело в том, что регуляризация такого типа использует Манхэттенскую норму (см. Рисунок 11), что при условии выпуклых целевых функций позволяет обнулить веса (коэффициенты) при многих переменных. В ряде работ [Yuan M., Lin Y., 2006; Huang J. et al, 2010] можно найти примеры такого анализа. В финансовом мире эта техника также нашла применение, например, в некоторых работах [Freyberger J. et al, 2020; Messmer M., Audrino F., 2017] производится отбор факторов для модели с помощью LASSO.

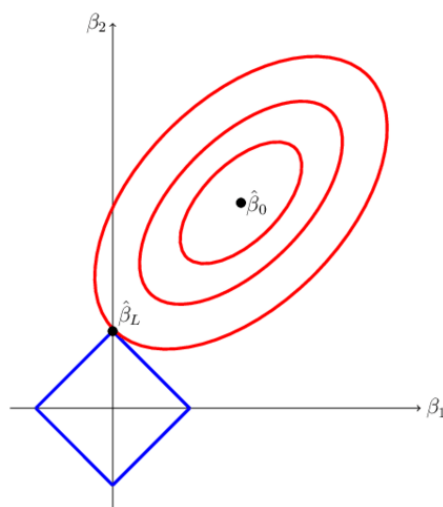


Рис. 11. Иллюстрация принципа обнуления весов в случае Манхэттенской метрики.

Источник: https://www.researchgate.net/figure/Behaviour-of-1-and-2-penalty-in-comparison-Red-lines-represent-RSS-contour-lines-and-the_fig1_330439412

²⁵ Максимизация среднего выигрыша от расщепления реализуется через `importance_type = 'gain'` во всех упомянутых моделях.

В исследовании Мессмера и Аудрино [Messmer and Audrino, 2017] авторы использовали как LASSO, так и расширение этой модели. Их расчёты показали, что веса при классических факторах из модели Фамы-Френча статистически не обнуляются, то есть признаются важными. Дополнительный анализ показывает, что при проведении МНК регрессии есть случаи статистической незначимости, однако, LASSO всё равно оставляет веса ненулевыми. Отмечается значительная важность фактора моментум.

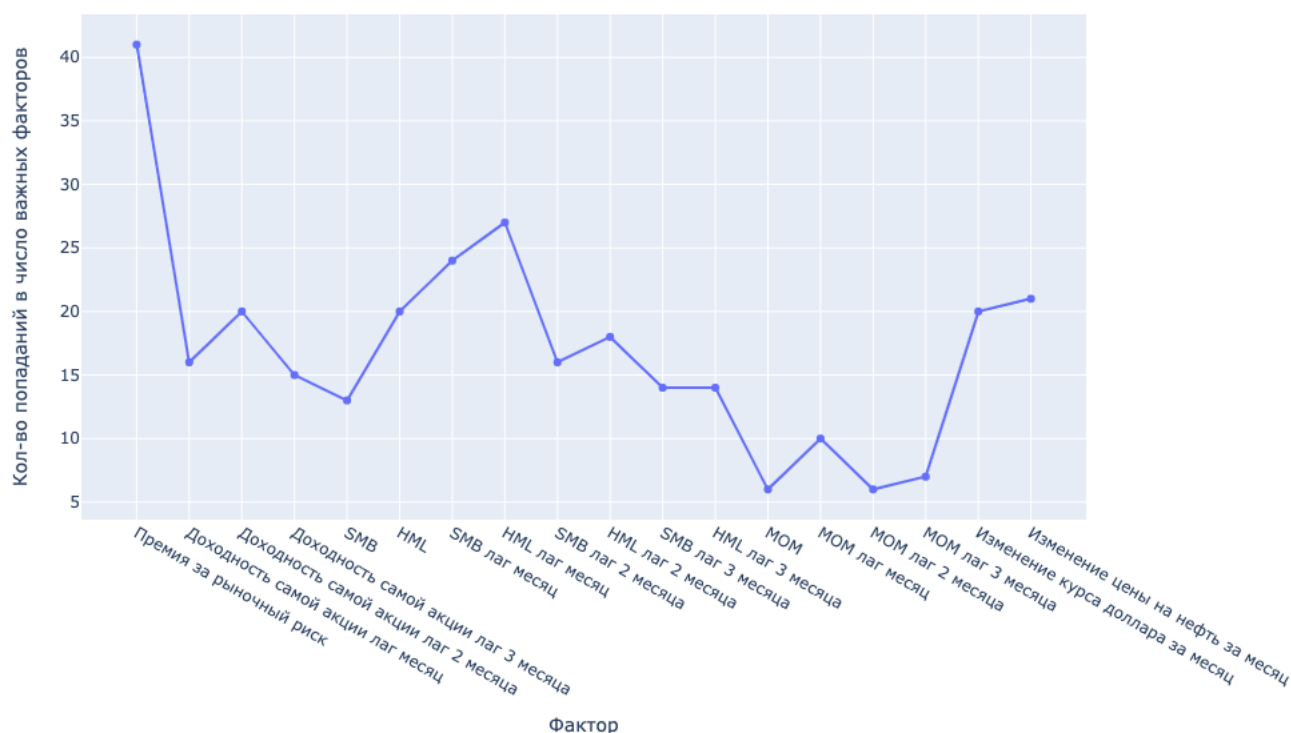


Рис. 12. Столько раз LASSO включила конкретный фактор в модель²⁶.

Источник: расчеты автора.

Чаще всего включается в модель премия за рыночный риск (см. Рисунок 12), что согласуется с большой важностью, полученной при применении случайного леса. Можно заметить, что характер динамики важности похож на рисунки 9–10, где демонстрировалась важность, поученная альтернативными методами. Это является дополнительным аргументом в пользу устойчивости результатов.

²⁶ Гиперпараметры LASSO совпадают с теми, которые использовались при обучении и оценках ранее в работе. В качестве параметров отбора использовались значения по умолчанию, предлагаемые в пакете Skleran (https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectFromModel.html)

Рисунок 13 является свидетельством наличия ряда компаний, где данные довольно плохо подвергаются факторному анализу и поэтому все факторы отвергаются алгоритмом, что видно на рисунке. Поэтому, в среднем, если в модель включён хотя бы один фактор по LASSO, то там есть премия за рыночный риск.

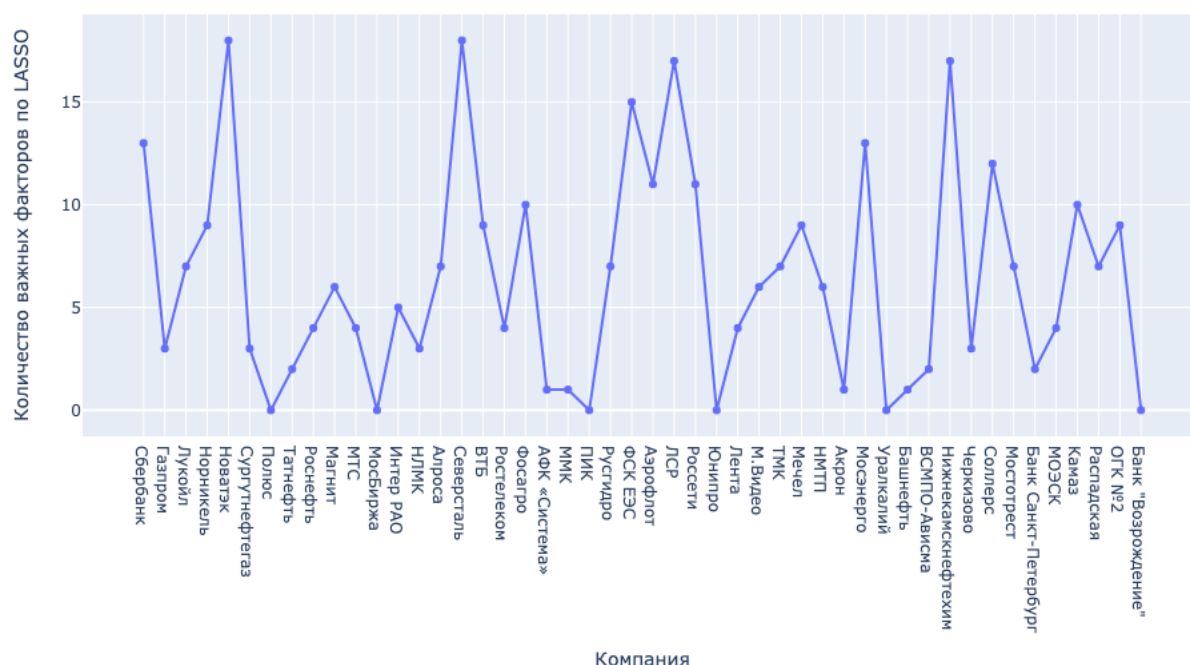


Рис. 13. Количество важных факторов для каждой компании, которые выделила LASSO.

Источник: расчеты автора.

3.7. Пример практического использования полученных расширений моделей доходности акции на базе торговой рекомендательной стратегии

В данном параграфе демонстрируется, как можно использовать расширения моделей доходности на основе машинного обучения и классических моделей для построения торговых рекомендаций, что может являться основой для торговой стратегии.

Базой для включения акции в рекомендацию по продаже, удержанию или покупке является факт её переоценки или недооценки рынком.

Ниже описывается алгоритм, на основе которого осуществляются рекомендации:

- 1) Модель²⁷ настраивается на тренировочной выборке.

²⁷ На основе алгоритмов ML или классических моделей и их расширений.

- 2) Для последней даты из тренировочной выборки²⁸ вычисляется значение модельной доходности.
- 2.1) Если модельное значение доходности превышает фактическую более чем на 1%, тогда следует рекомендовать данную акцию к покупке, так как она недооценена.
- 2.2) Если фактическое значение доходности превышает модельное более чем на 1%, тогда следует рекомендовать данную акцию к короткой продаже²⁹, так как она переоценена.
- 3) 2.3) Если не выполнены условия пунктов 3 и 4, тогда следует рекомендовать держать данную акцию.
- 4) Формируется портфель из акций, рекомендованных к покупке, удержанию и продаже. При этом предполагается использование модификации стратегии 1/N, т.е. все акции, получившие рекомендацию покупать или продавать, участвуют в стратегии с равными весами. Это позволяет несколько упростить анализ, используя усреднения доходностей.

$$r_t^{model} = \frac{P_t^{model} - P_{t-1}^{fact}}{P_{t-1}^{fact}} \quad (29)$$

$$r_t^{fact} = \frac{P_t^{fact} - P_{t-1}^{fact}}{P_{t-1}^{fact}} \quad (30)$$

Где индекс model означает модельный показатель, fact – фактический показатель, P_t – цена акции в момент времени t.

Основная идея рекомендательной стратегии:

$$r_t^{model} > r_t^{fact} \Leftrightarrow P_t^{model} > P_t^{fact} \quad (31)$$

Откуда следует недооценка акции.

Проверка результатов стратегии осуществлялась на 4 горизонтах инвестирования: 1 день, 10 дней, 30 дней, 60 дней после осуществления рекомендации, чтобы можно было посмотреть на динамику результатов прогноза во времени. Сравнивались два ключевых

²⁸ Выборка данных до интересующего временного промежутка, на которых обучается модель, чтобы сделать прогноз

²⁹ Продажа, если бумага есть в наличии и короткая продажа, если её нет в наличии.

показателя – абсолютная доходность и превышение над рыночной доходностью в соответствующей длинной или короткой позиции.

Тестирование стратегии проводилось с разными пороговыми процентными величинами, отражёнными в пунктах 3-5 стратегии (0,5%, 1%, 1,5%). Результаты были относительно устойчивыми, в представленных далее результатах использовалось значение 1%.

Проверка работы рекомендаций моделей осуществлялась на разных состояниях рынка:

- 1) Бычий рынок (см. Приложение 12) соответствует, например, временному промежутку с 2017-06-13 по 2017-09-05.
- 2) Медвежий рынок (см. Приложение 13) соответствует, например, временному промежутку с 2017-01-30 по 2017-05-23.
- 3) Рынок без явно выраженного бычьего или медвежьего тренда (далее - нейтральный рынок) (см. Приложение 14) соответствует, например, временному промежутку с 2018-04-24 по 2018-08-07.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации SARМ	Продавать	Кол-во акций:	16	Кол-во акций:	3	Кол-во акций:	1	Кол-во акций:	20
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,318%	1	-0,263%	1	-0,444%	1	0,193%
		10	1,657%	10	0,453%	10	1,385%	10	1,463%
		30	2,407%	30	1,638%	30	13,903%	30	2,866%
	60	2,8960%	60	-1,520%	60	19,134%	60	3,045%	
	Держать	Кол-во акций:	0	Кол-во акций:	2	Кол-во акций:	0	Кол-во акций:	2
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1	0,152%	1		1	0,152%
		10		10	1,205%	10		10	1,205%
		30		30	6,468%	30		30	6,468%
	60		60	-4,725%	60		60	-4,725%	
	Покупать	Кол-во акций:	3	Кол-во акций:	1	Кол-во акций:	20	Кол-во акций:	24
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,774%	1	0,833%	1	-0,280%	1	-0,295%
		10	0,298%	10	3,539%	10	1,507%	10	1,441%
		30	-0,529%	30	-3,081%	30	-0,210%	30	-0,370%
	60	-6,989%	60	-3,497%	60	-2,487%	60	-3,091%	
Итого		Кол-во акций:	19	Кол-во акций:	6	Кол-во акций:	21	Кол-во акций:	46
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,146%	1	0,058%	1	-0,288%	1	-0,064%
		10	1,443%	10	1,218%	10	1,501%	10	1,440%
		30	1,944%	30	2,461%	30	0,462%	30	1,335%
60	1,335%	60	-2,918%	60	-1,457%	60	-0,494%		

Таблица 14. Сравнение результативности рекомендаций, данных с помощью Lasso и SARМ на нейтральном рынке. В таблице представлены абсолютные доходности³⁰.

Источник: расчеты автора.

Основными моделями, с которыми проводилось сравнение результатов являлись классические SARМ и 3-факторная модель Фамы-Френча. С ними сопоставлялись различные модели машинного обучения, в качестве представителя здесь продемонстрируем результаты анализа самой точной модели при сравнении на тестовой выборке— Lasso.

³⁰ Большинство аналогичных таблиц результатов для медвежьего и бычьего рынков, а также примеры сравнения с 3-факторной моделью Фамы-Френча, представлены в приложениях 15-20.

При интерпретации результатов для рекомендации «Продавать» следует умножать доходности на -1, чтобы получать значение выигрыша стратегии, поскольку короткая позиция выгодна при уменьшении доходности акции. Так, например, CAPM при 60 дневном горизонте инвестирования принесла бы убыток в 3,045%.

Для стратегий «Держать» и «Покупать» направление изменения доходностей совпадает с выигрышем от стратегии.

Нейтральный рынок.

Можно сказать, что на нейтральном рынке при среднесрочном горизонте (10-30 дней) при опоре на рекомендации Lasso, абсолютная доходность в среднем получается сопоставимой с рекомендациями, основанным на CAPM, однако Lasso предлагает более диверсифицированный портфель для стратегии «Держать» и показывает меньшую просадку для рекомендаций по продаже.

Плохие результаты могут обуславливаться скоростью изменения факта недооценки или переоценки. На данных заметно, что, в целом, спустя 2 месяца рекомендации теряют свою релевантность так как наблюдается отрицательная доходность, однако закономерности с другими ситуациями, которые представлены в приложениях (см. Приложения 15-21) найти сложно. Данный вопрос требует дополнительного изучения.

Lasso демонстрирует на 10-60 дневной перспективе для стратегии «Продавать» результаты лучше рыночных, по сравнению с рекомендациями CAPM (см. Таблицу 15). Для стратегии «Держать» Lasso предлагает рекомендацию, несколько уступающую по доходности CAPM, однако более диверсифицированный набор акций.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации SARМ	Продавать	Кол-во акций:	16	Кол-во акций:	3	Кол-во акций:	1	Кол-во акций:	20
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,330%	1	-0,251%	1	-0,432%	1	0,205%
		10	-0,482%	10	-1,686%	10	-0,755%	10	-0,676%
		30	-0,453%	30	-1,223%	30	11,043%	30	0,006%
		60	0,3602%	60	-4,056%	60	16,598%	60	0,510%
	Держать	Кол-во акций:	0	Кол-во акций:	2	Кол-во акций:	0	Кол-во акций:	2
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1	0,164%	1		1	0,164%
		10		10	-0,935%	10		10	-0,935%
		30		30	3,608%	30		30	3,608%
		60		60	-7,260%	60		60	-7,260%
	Покупать	Кол-во акций:	3	Кол-во акций:	1	Кол-во акций:	20	Кол-во акций:	24
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,762%	1	0,845%	1	-0,268%	1	-0,283%
		10	-1,842%	10	1,399%	10	-0,632%	10	-0,699%
		30	-3,389%	30	-5,941%	30	-3,070%	30	-3,230%
		60	-9,524%	60	-6,033%	60	-5,022%	60	-5,627%
Итого		Кол-во акций:	19	Кол-во акций:	6	Кол-во акций:	21	Кол-во акций:	46
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,158%	1	0,070%	1	-0,276%	1	-0,052%
		10	-0,697%	10	-0,921%	10	-0,638%	10	-0,699%
		30	-0,917%	30	-0,399%	30	-2,398%	30	-1,525%
		60	-1,201%	60	-5,454%	60	-3,993%	60	-3,030%

Таблица 15. Сравнение результативности рекомендаций, данных с помощью Lasso и SARМ на нейтральном рынке. В таблице представлены доходности, которые превышают рыночную доходность.

Источник: расчеты автора.

При сравнении с моделью Фамой-Френча, можно сказать, что при рекомендации на продажу абсолютная просадка по Lasso меньше (см. Приложение 21). Также, не в первый раз Lasso предлагает более диверсифицированный портфель для удержания акций, при том, что абсолютные значения доходности довольно приемлемы. В дополнение к этому, рекомендации на покупку для нейтрального рынка Lasso даёт лучше.

При анализе сравнительной динамики рынок демонстрирует себя лучше, чем «короткий портфель» на основе Lasso, на более чем 10 дневном горизонте инвестирования

включительно. Однако «короткий портфель», в свою очередь, показывают себя лучше, чем модель Фамы-Френча с рекомендацией продажи для 30-60 дневного горизонта инвестирования (см. Приложение 20). Рекомендации Lasso на покупку уступают рынку, но не так, как рекомендации модели Фамы-Френча, тогда как результаты стратегии «Держать» получились сопоставимы.

Бычий рынок

При анализе абсолютных результатов наблюдается положительная доходность для «коротких» рекомендаций на малом горизонте инвестирования (1-10 дней) и выгодность «длинных» рекомендаций для большого горизонта инвестиций (30-60 дней) как в случае Lasso, так и в случае модели Фамы-Френча. При этом результаты сопоставимы как по доходности, так и по диверсификации для моделей Lasso и Фамы-Френча (см. Приложение 17). Для акций, рекомендованных моделями к удержанию, портфели, сформированные по рекомендациям обеих моделей, показывают хорошую доходность на 60-дневном горизонте, однако довольно низкую диверсификацию. Наблюдается небольшое преимущество SARМ.

При анализе результатов в сравнении с рыночной доходностью обнаруживается стабильное преимущество у портфеля, сформированного с помощью Lasso, над рынком. При этом портфель, в целом, хорошо диверсифицирован. Результаты примерно сопоставимы с моделью SARМ, но на коротком горизонте в 1-10 дней результативность Lasso несколько выше (см. Приложение 18).

В сравнении Lasso и модели Фамы-Френча наблюдаются похожие результаты по рекомендации «покупать», с небольшим преимуществом модели Фамы-Френча на 30-60-дневном горизонте, однако на коротком горизонте просадка Lasso меньше. Обе модели дают приемлемую доходность при рекомендации «Держать» на 2-месячном горизонте.

Медвежий рынок

При анализе результатов на медвежьем рынке, можно сказать, что все модели показали себя приемлемо только при рекомендации на продажу (см. Приложение 15), т.е. когда угадывается общий рыночный тренд. Абсолютная доходность была довольно высока от «короткой» рекомендации, хотя и уступила короткой позиции в рыночном портфеле.

Неудачную рекомендацию модели Lasso по направлению «Держать» можно объяснить отсутствием диверсификации, так как она состояла из одной акции. Достичь выигрыша по сравнению с рыночной доходностью получилось при рекомендации «покупать» (см. Приложение 16).

Результаты стратегии и перспективы применения ML методов.

Таким образом, значительного улучшения качества рекомендаций с помощью усложнения модели с использованием алгоритмов машинного обучения в рамках построенной стратегии получить не удалось, хотя в ряде случаев преимущества продвинутой модели показывают.

Анализ при различной динамике рынка показал, что при прочих равных на медвежьем рынке стоит реализовывать рекомендации на продажу, тогда как на бычьем неплохие результаты показывают реализованные рекомендации к покупке и удержанию, и относительно рынка и абсолютные доходности приемлемы. Модели не очень хорошо выбирают к рекомендациям акции, динамика которых будет идти против тренда.

Направление по улучшению результатов моделей на основе машинного обучения безусловно есть, так как данные рекомендации составлены на примере одного алгоритма машинного обучения, тогда как можно использовать группы алгоритмов (ансамблирование), усреднение их прогнозов, настройку алгоритмов под конкретные рекомендации, использование ещё более продвинутых моделей. Более того, потенциал использования методов только растёт ввиду повышающихся возможностей к оптимизации параметров модели, например, за счёт байесовской оптимизации гиперпараметров. Кроме того, использовалась наиболее простая стратегия, представляющая собой модификацию стратегии $1/N$. При этом, возможно, использование более продвинутых стратегий позволило бы добиться определенного преимущества над рынком. В любом случае, данный вопрос, относящийся к эффективности рынка, требует дальнейшего изучения.

Как основа, созданная стратегия на основе алгоритма Lasso сопоставима с результатами на основе классических моделей, а в ряде случаев несколько лучше. У продвинутых методов есть большой потенциал в применении к задачам оценки ожидаемой доходности.

3.8. Выводы по главе

В этой главе была проанализирована эффективность работы различных алгоритмов машинного обучения и проведено сравнение их результаты с линейными моделями. На тестовой выборке алгоритмы машинного обучения немного, но систематически обыгрывают расширения классических многофакторных моделей. Это даёт основания считать рациональным применение этих методов при прогнозе доходности акции.

Важно отметить, каким образом исследователю помогают методы машинного обучения. Дело в том, что методы машинного обучения в отличие от линейных моделей

могут улавливать более сложные, нелинейные взаимосвязи в исследуемых данных. Стоит отметить, что разрыв в качестве небольшой, поэтому можно говорить о примерной сопоставимости результатов, однако всё равно результаты показывают, что линейная регрессия проигрывает другим алгоритмам.

Таким образом, важность, которую оценивают алгоритмы машинного обучения, строится на заведомо более выигрышных моделях, чем классические многофакторные модели. Более того, оцениваемая важность устойчива к различным спецификациям моделей и алгоритмов.

Как было продемонстрировано на практическом примере, у продвинутых методов анализа ожидаемой доходности акции, основанных на алгоритмах машинного обучения, есть большой потенциал в коммерческом применении, хотя их использование требует дальнейшего анализа

Заключение

В данном исследовании проведена оценка ожидаемой доходности акций российских компаний на основе многофакторных моделей Фамы-Френча, Кархарта и расширений над ними, включающими учёт долговой нагрузки компаний, специфические для России факторы, а также базовые детерминанты доходности и их лаги. Полученные в исследовании результаты подтверждают значимость как классических факторов, так и добавленных, и свидетельствуют о целесообразности расширений моделей для большинства крупных российских компаний. При добавлении дополнительных факторов, в среднем, наблюдается значительное улучшение метрик качества моделей линейной регрессии, таких как R_{adj}^2 и среднеквадратическая ошибка модели. Так, у CAPM в среднем по всем компаниям качество оценки R_{adj}^2 составляет 0.26, а MSE 0,0096, а для классической трёхфакторной модели доходности Фамы-Френча 0,27 и 0,0093 соответственно, тогда как для расширенной модели аналогичные значения составляют 0,48 и 0,0068. Улучшение моделей при добавлении большего количества факторов происходит монотонно.

Дополнительное сравнение эффективности классических подходов к оценке параметров моделей и подходов, основанных на методах машинного обучения, показало, что, в среднем, они сопоставимы по качеству, хотя линейная регрессия показывает себя немного хуже на тестовой выборке. Использовалось большое количество разных алгоритмов машинного обучения для возможности многостороннего сопоставления результатов. Они оказались устойчивыми к различным спецификациям моделей. Это было продемонстрировано как на моделях линейной регрессии при добавлении новых факторов в классические модели, так и в моделях машинного обучения с использованием процедуры кросс-валидации.

Кроме того, методы машинного обучения, учитывающие нелинейные зависимости в данных, позволили количественно оценить степень важности отдельных факторов при прогнозе ожидаемой доходности. Большую количественную оценку важности получает премия за рыночный риск, что согласуется с экономической теорией и интуицией. В среднем на неё приходится примерно треть (35%) всей важности, согласно результатам, полученным при обучении случайного леса.

Важность несколько по-разному оценивается разными алгоритмами, тем не менее сохраняется общая тенденция в динамике оценки факта важности разными моделями, то есть случайный лес и разные алгоритмы градиентного бустинга похожим образом учитывают важность факторов модели. Степень важности ключевого фактора в зависимости от алгоритма менялась от 15% до 43%. Причём случайный лес и алгоритмы

градиентного бустинга (Light GBM и CatBoost) показали среднюю важность рыночной премии за риск более 34%. Кроме того, схожая динамика важности факторов наблюдалась и при отборе факторов методом LASSO, которая не использует решающие деревья при обучении, что говорит об устойчивости результатов к разным подходам к оценке важности. Модели уделяли ключевое внимание премии за рыночный риск, отмечали повышенную важность месячных лагов классических факторов Фамы-Френча и специфических для России факторов, таких как нефть и курс доллара.

Дополнительно было продемонстрировано, как можно использовать полученные результаты для создания торговой рекомендательной стратегии. Этот практический пример подчёркивает вариативность использования продвинутого инструментария и потенциала к его дальнейшему применению в коммерческих целях.

Стоит также отметить, что выводы, сделанные в работе, могут помочь в решении «factor zoo» проблемы. Это можно осуществить путём сортировки факторов по вычисленной с помощью алгоритмов машинного обучения важности и выбора только тех факторов, которые соответствуют требуемому уровню относительной важности, который можно определить заранее.

Для того, чтобы достичь новых успехов в сфере оценки ожидаемой доходности стоит применять всё более продвинутые методы, в частности инструментарий машинного обучения, сравнивать получаемую результативность с классическими многофакторными моделями, находить их ограничения и стараться решить их. Необходимо тестировать получаемые результаты на разных рынках, для проверки устойчивости моделей.

Библиографический список.

- Abe M., Nakayama H. Deep learning for forecasting stock returns in the cross-section //Pacific-Asia conference on knowledge discovery and data mining. – Springer, Cham, 2018. – С. 273-284.
- Altay E. et al. The effect of macroeconomic factors on asset returns: A comparative analysis of the German and the Turkish stock markets in an APT framework. – Universitäts-und Landesbibliothek Sachsen-Anhalt, 2003.
- Baker M., Bradley B., Wurgler J. Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly // Financial Analysts Journal. 2011. Т. 67. No 1.
- Banz R. W. The relationship between return and market value of common stocks //Journal of financial economics. – 1981. – Т. 9. – №. 1. – С. 3-18.
- Bartholdy J., Peare P. Estimation of expected return: CAPM vs. Fama and French //International Review of Financial Analysis. – 2005. – Т. 14. – №. 4. – С. 407-427.
- Basu S. The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence //Journal of financial economics. – 1983. – Т. 12. – №. 1. – С. 129-156.
- Basu S. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis //The journal of Finance. – 1977. – Т. 32. – №. 3. – С. 663-682.
- Bello Z. Y. A statistical comparison of the CAPM to the Fama-French Three Factor Model and the Carhart's Model //Global Journal of Finance and Banking Issues. – 2008. – Т. 2. – №. 2.
- Bhandari L. C. Debt/equity ratio and expected common stock returns: Empirical evidence //The journal of finance. – 1988. – Т. 43. – №. 2. – С. 507-528.
- Blitz D., Vliet P. van. The Volatility Effect: Lower Risk Without Lower Return // Journal of Portfolio Management. 2007. С. 102–113.
- Bruner R. F. et al. Market integration in developed and emerging markets: Evidence from the CAPM //Emerging Markets Review. – 2008. – Т. 9. – №. 2. – С. 89-103.
- Bryzgalova S., Pelger M., Zhu J. Forest through the trees: Building cross-sections of stock returns //Available at SSRN 3493458. – 2019.
- Cakici N., Fabozzi F. J., Tan S. Size, value, and momentum in emerging market stock returns //Emerging Markets Review. – 2013. – Т. 16. – С. 46-65.
- Carhart M. M. On persistence in mutual fund performance //The Journal of finance. – 1997. – Т. 52. – №. 1. – С. 57-82.

- Chen N., Ingersoll Jr J. E. Exact pricing in linear factor models with finitely many assets: A note //The Journal of Finance. – 1983. – T. 38. – №. 3. – C. 985-988.
- Chen N. F., Roll R., Ross S. A. Economic forces and the stock market //Journal of business. – 1986. – C. 383-403.
- Chen Y., Hao Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction //Expert Systems with Applications. – 2017. – T. 80. – C. 340-355.
- Cho D. C., Eun C. S., Senbet L. W. International arbitrage pricing theory: An empirical investigation //The Journal of Finance. – 1986. – T. 41. – №. 2. – C. 313-329.
- Connor G. The three types of factor models: A comparison of their explanatory power //Financial Analysts Journal. – 1995. – T. 51. – №. 3. – C. 42-46.
- Daniel K., Titman S., Wei K. C. J. Explaining the cross-section of stock returns in Japan: Factors or characteristics? //The Journal of Finance. – 2001. – T. 56. – №. 2. – C. 743-766.
- Davis J. L., Fama E. F., French K. R. Characteristics, covariances, and average returns: 1929 to 1997 //The Journal of Finance. – 2000. – T. 55. – №. 1. – C. 389-406.
- Dey M.K. Turnover and return in global stock markets // Emerging Markets Review. 2005. T. 6. C. 45–67.
- Diallo B., Bagudu A., Zhang Q. A Machine Learning Approach to the Fama-French Three- and Five-Factor Models //Available at SSRN 3440840. – 2019.
- Di Persio L., Honchar O. Artificial neural networks architectures for stock price prediction: Comparisons and applications //International journal of circuits, systems and signal processing. – 2016. – T. 10. – №. 2016. – C. 403-413.
- Dragotă V., Mitrică E. Emergent capital markets' efficiency: The case of Romania //European Journal of Operational Research. – 2004. – T. 155. – №. 2. – C. 353-360.
- Dranev Y., Fomkina S. An asymmetric approach to the cost of equity estimation: empirical evidence from Russia //Higher School of Economics Research Paper No. WP BPR. – 2012. – T. 12.
- Estrada J. The cost of equity in emerging markets: A downside risk approach (II) //Available at SSRN 249579. – 2001.
- Estrada J. Systematic risk in emerging markets: the D-CAPM //Emerging Markets Review. – 2002. – T. 3. – №. 4. – C. 365-379.
- Fama E. F., French K. R. Common risk factors in the returns on stocks and bonds //Journal of financial economics. – 1993. – T. 33. – №. 1. – C. 3-56.
- Fama E. F., French K. R. Size, value, and momentum in international stock returns //Journal of financial economics. – 2012. – T. 105. – №. 3. – C. 457-472.

- Febrian E., Herwany A. CAPM and APT validation test before, during, and after financial crisis in emerging market: evidence from Indonesia //The International Journal of Business & Finance Research. – 2010. – T. 10. – №. 1.
- Feng G., Giglio S., Xiu D. Taming the factor zoo //Chicago Booth research paper. – 2017. – №. 17-04.
- Feng G., Giglio S., Xiu D. Taming the factor zoo: A test of new factors //The Journal of Finance. – 2020. – T. 75. – №. 3. – C. 1327-1370.
- Frazzini A., Pedersen L.H. Betting Against Beta // Journal of Financial Economics. 2014. T. 111. No 1. C. 1–25.
- Freyberger J., Neuhierl A., Weber M. Dissecting characteristics nonparametrically //The Review of Financial Studies. – 2020. – T. 33. – №. 5. – C. 2326-2377.
- Garcia R., Ghysels E. Structural change and asset pricing in emerging markets //Journal of International Money and Finance. – 1998. – T. 17. – №. 3. – C. 455-473.
- Gibbons M. R., Hess P. Day of the week effects and asset returns //Journal of business. – 1981. – C. 579-596.
- Goriaev A. Risk factors in the Russian stock market //Emerging Market Review. – 2004. – T. 2. – №. 1. – C. 67-89.
- Gu S., Kelly B., Xiu D. Empirical asset pricing via machine learning. – National bureau of economic research, 2018. – №. w25398.
- Gu S., Kelly B., Xiu D. Autoencoder asset pricing models //Journal of Econometrics. – 2021. – T. 222. – №. 1. – C. 429-450.
- Hamada R. S. The effect of the firm's capital structure on the systematic risk of common stocks //The journal of finance. – 1972. – T. 27. – №. 2. – C. 435-452.
- Harvey C. R., Liu Y. Lucky factors //Available at SSRN. – 2015. – T. 2528780. – C. 980.
- Hendricks D., Patel J., Zeckhauser R. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988 //The Journal of finance. – 1993. – T. 48. – №. 1. – C. 93-130.
- Huang J., Horowitz J. L., Wei F. Variable selection in nonparametric additive models //Annals of statistics. – 2010. – T. 38. – №. 4. – C. 2282.
- Jegadeesh N., Titman S. Profitability of momentum strategies: An evaluation of alternative explanations //The Journal of finance. – 2001. – T. 56. – №. 2. – C. 699-720.
- Jegadeesh N., Titman S. Returns to buying winners and selling losers: Implications for stock market efficiency //The Journal of finance. – 1993. – T. 48. – №. 1. – C. 65-91.
- Jensen M. C., Black F., Scholes M. S. The capital asset pricing model: Some empirical tests. – 1972.

- Kisman Z., Restiyanita S. M. The Validity of Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT) in Predicting the Return of Stocks in Indonesia Stock Exchange //American Journal of Economics, Finance and Management. – 2015. – T. 1. – №. 3. – C. 184-189.
- Krauss C., Do X. A., Huck N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500 //European Journal of Operational Research. – 2017. – T. 259. – №. 2. – C. 689-702.
- Lally M. The Fama-French model, leverage, and the Modigliani-Miller propositions //Journal of Financial Research. – 2004. – T. 27. – №. 3. – C. 341-349.
- Lee E., Liu W., Strong N. UK evidence on the characteristics versus covariance debate //European Financial Management. – 2007. – T. 13. – №. 4. – C. 742-756.
- Levy H. Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio //The American Economic Review. – 1978. – T. 68. – №. 4. – C. 643-658.
- Lintner J. Security prices, risk, and maximal gains from diversification //The journal of finance. – 1965. – T. 20. – №. 4. – C. 587-615.
- Liu J., Stambaugh R. F., Yuan Y. Size and value in China //Journal of Financial Economics. – 2019. – T. 134. – №. 1. – C. 48-69.
- Markowitz H. Portfolio Selection // The journal of finance. – 1952. – T. 7. – №. 1. – C. 77-91.
- Merton R. C. An intertemporal capital asset pricing model //Econometrica: Journal of the Econometric Society. – 1973. – C. 867-887.
- Merton R. C. et al. A simple model of capital market equilibrium with incomplete information. – 1987.
- Messmer M., Audrino F. The (adaptive) lasso in the zoo-firm characteristic selection in the cross-section of expected returns //Available at SSRN 2930436. – 2017.
- Montavon G. et al. Layer-wise relevance propagation: an overview //Explainable AI: interpreting, explaining and visualizing deep learning. – 2019. – C. 193-209.
- Mossin J. Equilibrium in a capital asset market //Econometrica: Journal of the econometric society. – 1966. – C. 768-783.
- Patel J. et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques //Expert systems with applications. – 2015. – T. 42. – №. 1. – C. 259-268.
- Pereiro L. E. The beta dilemma in emerging markets //Journal of Applied Corporate Finance. – 2010. – T. 22. – №. 4. – C. 110-122.
- Pungulescu C. Market Size Effects and Financial Integration // 2014.

Reinganum M. R. The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects //Journal of financial economics. – 1983. – T. 12. – №. 1. – C. 89-104.

Rojas-Suarez L. Towards strong and stable capital markets in emerging market economies //BIS Paper. – 2014. – №. 75c.

Roll R., Ross S. A. An empirical investigation of the arbitrage pricing theory //The journal of finance. – 1980. – T. 35. – №. 5. – C. 1073-1103.

Ross S. The arbitrage theory of capital asset pricing // Journal of Economic Theory. 1976a. T. 31. C. 341–60.

Sharpe W. F. Capital asset prices: A theory of market equilibrium under conditions of risk //The journal of finance. – 1964. – T. 19. – №. 3. – C. 425-442.

Simonian J. et al. A Machine Learning Approach to Risk Factors: A Case Study Using the Fama–French–Carhart Model //The Journal of Financial Data Science. – 2019. – T. 1. – №. 1. – C. 32-44.

Spulbar C., Birau R. Testing weak-form efficiency and long-term causality of the RIPH emerging capital markets //International Journal of Business Quantitative Economics and Applied Management Research. – 2018. – T. 5. – №. 2. – C. 1-19.

Spulbar C., Trivedi J., Birau R. Investigating abnormal volatility transmission patterns between emerging and developed stock markets: A case study //Journal of Business Economics and Management. – 2020. – T. 21. – №. 6. – C. 1561-1592.

Stoll H. R., Whaley R. E. Transaction costs and the small firm effect //Journal of Financial Economics. – 1983. – T. 12. – №. 1. – C. 57-79.

Tianchuan L., Pritamani M. Country Size and Country Momentum Effects in Emerging and Frontier Markets // Journal of Investing. 2015. T. 24. C. 102–8.

Timothy Masters. Practical neural network recipes in C++. – Morgan Kaufmann, 1993.

Treynor J. L. Market value, time, and risk //Time, and Risk (August 8, 1961). – 1961.

Wolpert D. H., Macready W. G. No free lunch theorems for optimization //IEEE transactions on evolutionary computation. – 1997. – T. 1. – №. 1. – C. 67-82.

Yuan M., Lin Y. Model selection and estimation in regression with grouped variables //Journal of the Royal Statistical Society: Series B (Statistical Methodology). – 2006. – T. 68. – №. 1. – C. 49-67.

Zaremba A. Risk-Based Explanation for the Country-Level Size and Value Effects // Finance Research Letters. 2016. T. 18. C. 226033.

Приложения

Во всех регрессиях предполагается наличие 1979 наблюдений, а также Во всех моделях учтено наличие долга у компаний по формуле Р.Хамады для рыночной премии за риск, SMB, HML и соответствующих лагов. Все расчеты проведены автором. Пояснение:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Приложение 1. Результаты регрессий для Сбербанка (Sber)

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.444***	0.421***	0.422***	0.427***	0.435***	0.434***
доходность Sber лаг месяц					0.053***	0.068***
доходность Sber лаг 2 месяца					0.080***	0.129***
доходность Sber лаг 3 месяца					0.085***	0.058***
SMB		0.081***	0.080***	0.054**	0.045**	0.063***
HML		-0.021**	-0.019	0.000	0.008	0.054***
SMB лаг месяц				0.055***	0.052***	0.045***
HML лаг месяц				-0.133***	-0.128***	0.055***
SMB лаг 2 месяца				0.068***	0.064***	0.047***
HML лаг 2 месяца				-0.021	-0.009	0.031**
SMB лаг 3 месяца				0.029*	0.007	0.049***
HML лаг 3 месяца				0.003	0.022*	0.025*
MOM			-0.011	0.081*	0.065	0.061*
MOM лаг месяц				-0.450***	-0.473***	-0.117**
MOM лаг 2 месяца				-0.030	-0.023	-0.199***
MOM лаг 3 месяца				0.429***	0.434***	0.272***
Изменение курса доллара за месяц						-0.761***
Изменение цены на нефть за месяц						0.015
R ²	0.403	0.410	0.410	0.545	0.555	0.680
Adjusted R ²	0.403	0.409	0.409	0.542	0.551	0.677
MSE	0.006	0.006	0.006	0.005	0.004	0.003
F Statistic	220.504***	80.489***	62.265***	103.034***	90.173***	185.670***

Приложение 2. Результаты регрессий для Газпрома (Gazp)

	CAPM	Фама- Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.744***	0.755***	0.768***	0.690***	0.688***	0.701***
Доходность Gazp лаг месяц					0.010	0.010
Доходность Gazp лаг 2 месяца					-0.070***	-0.080***
Доходность Gazp лаг 3 месяца					-0.016	-0.021*
SMB		-0.073***	-0.103***	-0.087***	-0.078***	-0.079***
HML		-0.048***	-0.013	0.015	0.015	0.005
SMB лаг месяц				-0.189***	-0.191***	-0.193***
HML лаг месяц				-0.207***	-0.211***	-0.243***
SMB лаг 2 месяца				-0.006	0.008	0.002
HML лаг 2 месяца				-0.068***	-0.071***	-0.077***
SMB лаг 3 месяца				-0.012	-0.043***	-0.043**
HML лаг 3 месяца				0.067***	0.041***	0.036***
MOM			-0.104***	-0.137***	-0.146***	-0.148***
MOM лаг месяц				0.237***	0.244***	0.195***
MOM лаг 2 месяца				-0.052	-0.048	-0.039
MOM лаг 3 месяца				-0.110***	-0.106***	-0.068**
Изменение курса доллара за месяц						0.059**
Изменение цены на нефть за месяц						-0.025*
R ²	0.618	0.624	0.641	0.725	0.728	0.732
Adjusted R ²	0.618	0.623	0.641	0.723	0.726	0.729
MSE	0.003	0.003	0.003	0.002	0.002	0.002
F Statistic	2026.203***	775.529***	669.507***	240.885***	198.346***	188.065***

Приложение 3. Результаты регрессий для Лукойла (Lkoh)

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.900***	0.906***	0.901***	0.898***	0.901***	0.852***
Доходность Lkoh лаг месяц					0.029**	0.016
Доходность Lkoh лаг 2 месяца					0.005	0.030**
Доходность Lkoh лаг 3 месяца					-0.027*	-0.013
SMB		-0.008	0.005	0.056***	0.044**	0.057***
HML		0.062***	0.045***	0.017	0.027*	0.031*
SMB лаг месяц				-0.195***	-0.201***	-0.183***
HML лаг месяц				0.073***	0.075***	0.041**
SMB лаг 2 месяца				-0.085***	-0.070***	-0.052**
HML лаг 2 месяца				0.077***	0.077***	0.063***
SMB лаг 3 месяца				0.009	0.019	-0.001
HML лаг 3 месяца				-0.017	-0.022*	-0.002
MOM			0.037***	0.041	0.024	0.010
MOM лаг месяц				0.028	0.025	-0.016
MOM лаг 2 месяца				-0.121***	-0.102***	-0.056*
MOM лаг 3 месяца				0.064**	0.066**	0.071***
Изменение курса доллара за месяц						0.191***
Изменение цены на нефть за месяц						0.081***
R ²	0.677	0.680	0.683	0.716	0.717	0.734
Adjusted R ²	0.677	0.680	0.683	0.714	0.714	0.731
MSE	0.002	0.002	0.002	0.002	0.002	0.002
F Statistic	1940.700***	678.098***	578.531***	256.144***	216.783***	218.357***

Приложение 4. Результаты регрессий для Норильского Никеля (Gmkn)

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.753***	0.753***	0.743***	0.859***	0.854***	0.873***
Доходность Gmkn лаг месяц					-0.052***	-0.049***
Доходность Gmkn лаг 2 месяца					0.140***	0.157***
Доходность Gmkn лаг 3 месяца					0.001	-0.003
SMB		0.011	0.037	0.090***	0.083***	0.079***
HML		0.032*	-0.003	-0.062***	-0.071***	-0.102***
SMB лаг месяц				-0.010	0.015	0.011
HML лаг месяц				0.310***	0.316***	0.195***
SMB лаг 2 месяца				-0.088***	-0.161***	-0.180***
HML лаг 2 месяца				-0.077***	-0.051**	-0.075***
SMB лаг 3 месяца				-0.079***	-0.034	-0.024
HML лаг 3 месяца				-0.109***	-0.132***	-0.129***
MOM			0.073***	0.095***	0.120***	0.111***
MOM лаг месяц				0.087**	0.047	-0.101**
MOM лаг 2 месяца				-0.001	-0.043	0.005
MOM лаг 3 месяца				-0.132***	-0.088**	0.010
Изменение курса доллара за месяц						0.241***
Изменение цены на нефть за месяц						-0.025**
R ²	0.427	0.428	0.438	0.521	0.541	0.569
Adjusted R ²	0.427	0.427	0.437	0.518	0.538	0.565
MSE	0.004	0.004	0.004	0.003	0.003	0.003
F Statistic	1076.255***	361.965***	308.039***	134.228***	123.214***	136.299***

Приложение 5. Результаты регрессий для компании Новатэк (Nvtk)

	CAPM	Фама- Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.803***	0.845***	0.835***	0.843***	0.850***	0.808***
доходность Nvtk лаг месяц					-0.005	-0.036**
доходность Nvtk лаг 2 месяца					-0.034*	-0.032*
доходность Nvtk лаг 3 месяца					0.037**	0.051***
SMB		-0.239***	-0.209***	-0.223***	-0.221***	-0.215***
HML		-0.062**	-0.102***	-0.139***	-0.139***	-0.137***
SMB лаг месяц				-0.366***	-0.367***	-0.349***
HML лаг месяц				0.201***	0.207***	0.120***
SMB лаг 2 месяца				0.102***	0.109***	0.116***
HML лаг 2 месяца				-0.053***	-0.068***	-0.094***
SMB лаг 3 месяца				-0.027	-0.051	-0.076**
HML лаг 3 месяца				-0.103***	-0.095***	-0.074***
MOM			0.078***	-0.173***	-0.177***	-0.196***
MOM лаг месяц				-0.055	-0.042	-0.110**
MOM лаг 2 месяца				0.224***	0.223***	0.275***
MOM лаг 3 месяца				0.098***	0.090**	0.124***
Изменение курса доллара за месяц						0.240***
Изменение цены на нефть за месяц						0.064***
R ²	0.413	0.432	0.444	0.534	0.536	0.552
Adjusted R ²	0.413	0.431	0.443	0.530	0.532	0.548
MSE	0.004	0.004	0.004	0.003	0.003	0.003
F Statistic	1009.723***	365.485***	300.801***	140.286***	118.230***	101.167***

Приложение 6. Результаты регрессий для Мосэнерго (Msng)

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.532***	0.500***	0.503***	0.544***	0.543***	0.520***
доходность Msng лаг месяц					-0.110***	-0.129***
доходность Msng лаг 2 месяца					-0.014	-0.027
доходность Msng лаг 3 месяца					-0.065***	-0.086***
SMB		0.096**	0.094**	-0.033	-0.001	0.022
HML		-0.166***	-0.158***	-0.079***	-0.083***	-0.053***
SMB лаг месяц				0.518***	0.563***	0.576***
HML лаг месяц				-0.439***	-0.426***	-0.292***
SMB лаг 2 месяца				0.218***	0.255***	0.268***
HML лаг 2 месяца				-0.075***	-0.139***	-0.119***
SMB лаг 3 месяца				0.229***	0.274***	0.300***
HML лаг 3 месяца				0.059***	0.035	0.019
MOM			-0.030	0.310***	0.319***	0.317***
MOM лаг месяц				-0.107**	-0.060	0.134**
MOM лаг 2 месяца				-0.252***	-0.300***	-0.376***
MOM лаг 3 месяца				0.062	0.055	-0.047
Изменение курса доллара за месяц						-0.341***
Изменение цены на нефть за месяц						0.037***
R ²	0.191	0.218	0.219	0.580	0.589	0.613
Adjusted R ²	0.190	0.217	0.218	0.577	0.586	0.610
MSE	0.011	0.011	0.011	0.006	0.006	0.005
F Statistic	182.107***	77.219***	61.767***	118.831***	104.354***	116.795***

Приложение 7. Результаты регрессий для Башнефти (Bane)

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.766***	0.730***	0.734***	0.702***	0.779***	0.784***
доходность Bane лаг месяц					0.101***	0.087***
доходность Bane лаг 2 месяца					0.155***	0.141***
доходность Bane лаг 3 месяца					-0.099***	-0.113***
SMB		0.135***	0.131***	0.116**	0.028	0.032
HML		-0.333***	-0.324***	-0.278***	-0.243***	-0.229***
SMB лаг месяц				0.061	0.040	0.038
HML лаг месяц				-0.135***	-0.011	0.064*
SMB лаг 2 месяца				0.153***	0.135***	0.130***
HML лаг 2 месяца				0.051**	0.073***	0.084***
SMB лаг 3 месяца				0.045	0.105**	0.128***
HML лаг 3 месяца				0.157***	0.166***	0.145***
MOM			-0.021	0.075	0.127**	0.132**
MOM лаг месяц				-0.134**	-0.204***	-0.122*
MOM лаг 2 месяца				-0.178***	-0.206***	-0.262***
MOM лаг 3 месяца				0.190***	0.207***	0.182***
Изменение курса доллара за месяц						-0.238***
Изменение цены на нефть за месяц						-0.033**
R ²	0.264	0.324	0.325	0.362	0.392	0.399
Adjusted R ²	0.264	0.323	0.323	0.358	0.386	0.393
MSE	0.011	0.010	0.010	0.009	0.009	0.009
F Statistic	661.305***	245.951***	196.052***	101.859***	77.695***	77.212***

Для компании Башнефть было доступно только 1805 наблюдений.

Приложение 8. Результаты регрессий для Соллерса (Svav)

	CAPM	Фама- Френч	Кархарт	Кархарт + лаги	Кархарт + лаги и автолаги	Кархарт + лаги и автолаги + специфические
Премия за рыночный риск	0.483***	0.488***	0.495***	0.502***	0.527***	0.532***
доходность Svav лаг месяц					0.111***	0.054**
доходность Svav лаг 2 месяца					0.095***	0.085***
доходность Svav лаг 3 месяца					0.025	-0.006
SMB		-0.093**	-0.123***	-0.191***	-0.198***	-0.190***
HML		-0.185***	-0.147***	-0.127***	-0.167***	-0.112***
SMB лаг месяц				0.447***	0.465***	0.466***
HML лаг месяц				0.057**	0.061**	0.282***
SMB лаг 2 месяца				0.100***	0.058*	0.070**
HML лаг 2 месяца				-0.083***	-0.071***	-0.028
SMB лаг 3 месяца				-0.003	-0.044	-0.002
HML лаг 3 месяца				0.075***	0.076***	0.047*
MOM			-0.101***	0.027	0.092*	0.106**
MOM лаг месяц				-0.570***	-0.578***	-0.264***
MOM лаг 2 месяца				0.174***	0.188***	-0.001
MOM лаг 3 месяца				0.267***	0.226***	0.090
Изменение курса доллара за месяц						-0.693***
Изменение цены на нефть за месяц						-0.045***
R ²	0.185	0.210	0.220	0.370	0.387	0.461
Adjusted R ²	0.184	0.209	0.219	0.366	0.382	0.456
MSE	0.011	0.011	0.010	0.008	0.008	0.007
F Statistic	340.016***	151.194***	133.101***	122.441***	105.879***	101.174***

Приложение 9. Пример регрессии с учётом Covid 2019 для Сбера

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги и автолаги + специфические	Кархарт + лаги и автолаги + специфические + Covid 2019
Премия за рыночный риск	0.444***	0.421***	0.422***	0.434***	0.429***
доходность Sber лаг месяц				0.068***	0.056***
доходность Sber лаг 2 месяца				0.129***	0.122***
доходность Sber лаг 3 месяца				0.058***	0.059***
SMB		0.081***	0.080***	0.063***	0.061***
HML		-0.021**	-0.019	0.054***	0.050***
SMB лаг месяц				0.045***	0.048***
HML лаг месяц				0.055***	0.052***
SMB лаг 2 месяца				0.047***	0.051***
HML лаг 2 месяца				0.031**	0.029**
SMB лаг 3 месяца				0.049***	0.051***
HML лаг 3 месяца				0.025*	0.022*
MOM			-0.011	0.061*	0.060
MOM лаг месяц				-0.117**	-0.105**
MOM лаг 2 месяца				-0.199***	-0.180***
MOM лаг 3 месяца				0.272***	0.253***
Изменение курса доллара за месяц				-0.761***	-0.763***
Изменение цены на нефть за месяц				0.015	0.006
Covid 2019					-0.024***
R ²	0.403	0.410	0.410	0.680	0.681
Adjusted R ²	0.403	0.409	0.409	0.677	0.678
MSE	0.011	0.011	0.010	0.007	0.007

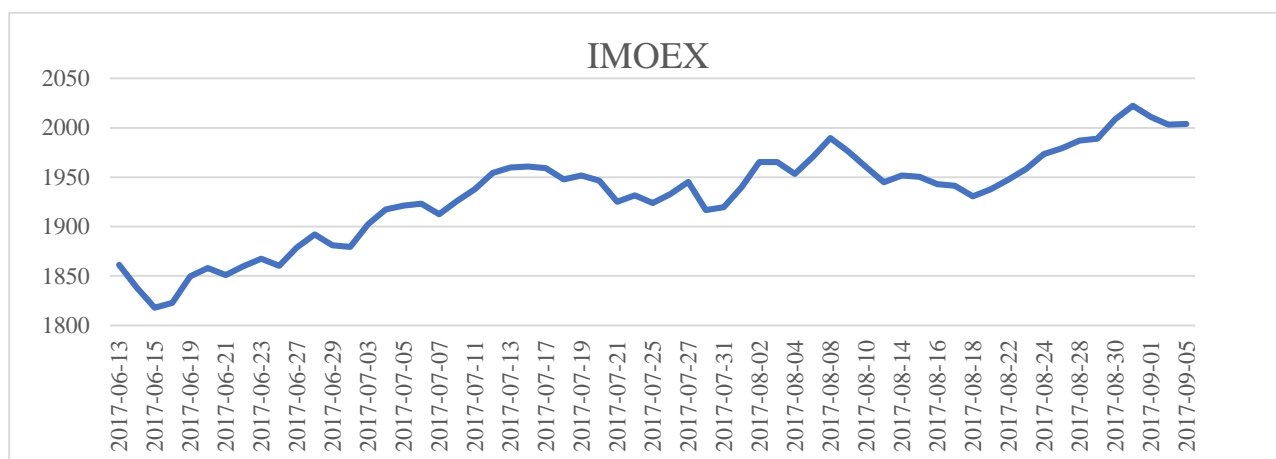
Приложение 10. Пример регрессии с учётом Covid 2019 для компании Аэрофлот

	CAPM	Фама-Френч	Кархарт	Кархарт + лаги и автолаги + специфические	Кархарт + лаги и автолаги + специфические + Covid 2019
Премия за рыночный риск	0.298***	0.267***	0.270***	0.325***	0.305***
Доходность Aflt лаг месяц				0.119***	0.084***
Доходность Aflt лаг 2 месяца				0.117***	0.102***
Доходность Aflt лаг 3 месяца				0.047**	0.046**
SMB		0.134***	0.130***	0.069***	0.074***
HML		-0.002	0.004	0.043***	0.037***
SMB лаг месяц				0.243***	0.249***
HML лаг месяц				0.009	0.014
SMB лаг 2 месяца				-0.026	-0.007
HML лаг 2 месяца				0.043***	0.054***
SMB лаг 3 месяца				0.039**	0.065***
HML лаг 3 месяца				0.019	0.010
MOM			-0.052**	-0.100*	-0.115**
MOM лаг месяц				0.039	0.075
MOM лаг 2 месяца				-0.211***	-0.152***
MOM лаг 3 месяца				0.239***	0.186***
Изменение курса доллара за месяц				-0.408***	-0.449***
Изменение цены на нефть за месяц				0.010	-0.016
Covid 2019					-0.091***
R2	0.256	0.276	0.278	0.451	0.461
Adjusted R2	0.256	0.274	0.276	0.446	0.456
MSE	0,013	0,013	0,013	0.010	0.009

Приложение 11. Пример регрессии с учётом Covid 2019 для компании Мосэнерго

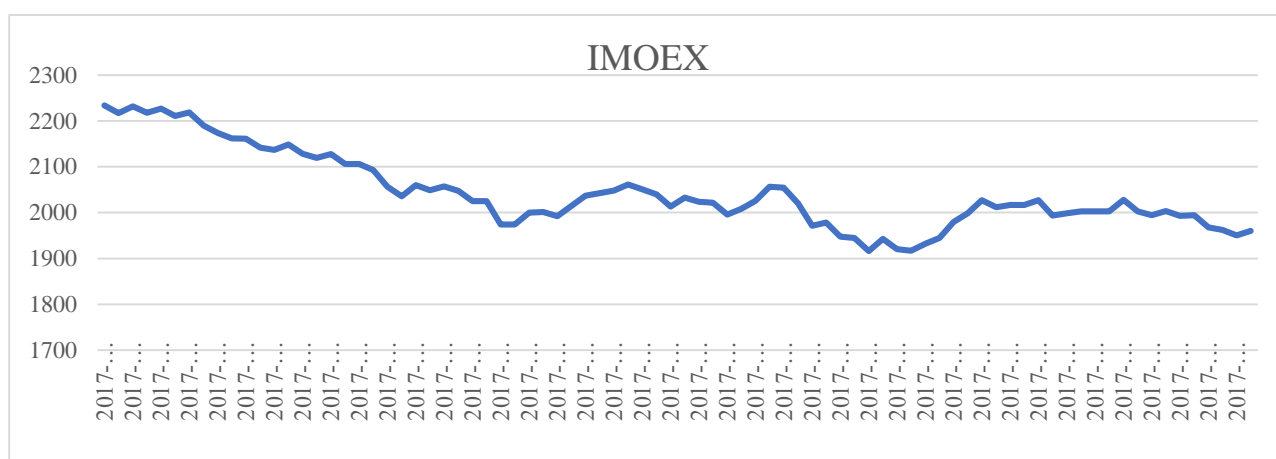
	CAPM	Фама-Френч	Кархарт	Кархарт + лаги и автолаги + специфические	Кархарт + лаги и автолаги + специфические + Covid 2019
Премия за рыночный риск	0.532***	0.500***	0.503***	0.520***	0.517***
доходность Msnг лаг месяц				-0.129***	-0.132***
доходность Msnг лаг 2 месяца				-0.027	-0.028
доходность Msnг лаг 3 месяца				-0.086***	-0.086***
SMB		0.096**	0.094**	0.022	0.021
HML		-0.166***	-0.158***	-0.053***	-0.055***
SMB лаг месяц				0.576***	0.577***
HML лаг месяц				-0.292***	-0.292***
SMB лаг 2 месяца				0.268***	0.272***
HML лаг 2 месяца				-0.119***	-0.120***
SMB лаг 3 месяца				0.300***	0.303***
HML лаг 3 месяца				0.019	0.017
MOM			-0.030	0.317***	0.315***
MOM лаг месяц				0.134**	0.138**
MOM лаг 2 месяца				-0.376***	-0.367***
MOM лаг 3 месяца				-0.047	-0.055
Изменение курса доллара за месяц				-0.341***	-0.343***
Изменение цены на нефть за месяц				0.037***	0.034**
Covid 2019					-0.010
R2	0.191	0.218	0.219	0.613	0.614
Adjusted R2	0.190	0.217	0.218	0.610	0.610
MSE	0.011	0.011	0.011	0.005	0.005

Приложение 12. Иллюстрация бычьего рынка



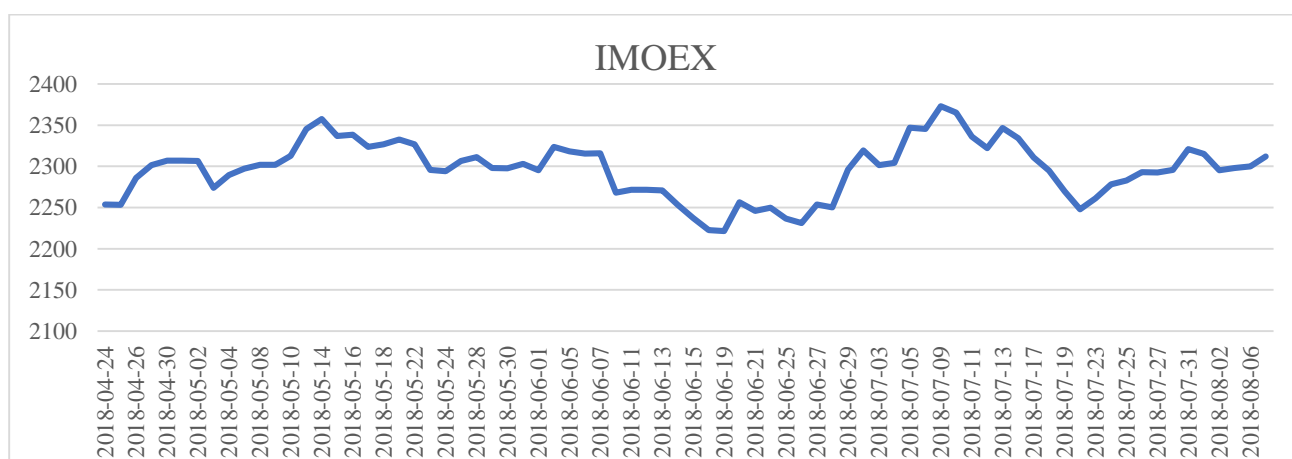
Источник: построено автором на основе данных МосБиржи

Приложение 13. Иллюстрация медвежьего рынка



Источник: построено автором на основе данных МосБиржи

Приложение 14. Иллюстрация нейтрального рынка



Источник: построено автором на основе данных МосБиржи

Приложение 15. Сравнение результативности рекомендаций, данных с помощью Lasso и 3-факторной модели Фамы-Френча на медвежьем рынке

В таблице представлены абсолютные доходности.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации Fama-French	Продавать	Кол-во акций:	22	Кол-во акций:	1	Кол-во акций:	4	Кол-во акций:	27
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,032%	1	-1,493%	1	0,170%	1	-0,056%
		10	-0,525%	10	-4,096%	10	-2,769%	10	-0,990%
		30	-10,901%	30	-15,875%	30	-10,632%	30	-11,045%
		60	-7,611%	60	-18,768%	60	-12,765%	60	-8,787%
	Держать	Кол-во акций:	2	Кол-во акций:	0	Кол-во акций:	3	Кол-во акций:	5
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,396%	1		1	0,461%	1	0,119%
		10	-2,525%	10		10	-0,909%	10	-1,555%
		30	-5,694%	30		30	-5,614%	30	-5,646%
		60	-13,014%	60		60	-5,512%	60	-8,513%
	Покупать	Кол-во акций:	1	Кол-во акций:	0	Кол-во акций:	10	Кол-во акций:	11
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-1,682%	1		1	-0,225%	1	-0,358%
		10	-2,109%	10		10	-1,710%	10	-1,746%
		30	-6,599%	30		30	-9,943%	30	-9,639%
		60	-10,766%	60		60	-11,030%	60	-11,006%
Итого		Кол-во акций:	25	Кол-во акций:	1	Кол-во акций:	17	Кол-во акций:	43
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,127%	1	-1,493%	1	-0,011%	1	-0,113%
		10	-0,748%	10	-4,096%	10	-1,818%	10	-1,249%
		30	-10,312%	30	-15,875%	30	-9,341%	30	-10,058%
		60	-8,169%	60	-18,768%	60	-10,464%	60	-9,323%

Источник: расчеты автора

Приложение 16. Сравнение результативности рекомендаций, данных с помощью Lasso и 3-факторной модели Фамы-Френча на медвежьем рынке

В таблице представлены доходности, которые превышают рыночную доходность.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации Fama-French	Продавать	Кол-во акций:	22	Кол-во акций:	1	Кол-во акций:	4	Кол-во акций:	27
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,718%	1	-0,743%	1	0,920%	1	0,694%
		10	2,727%	10	-0,844%	10	0,482%	10	2,262%
		30	-0,427%	30	-5,401%	30	-0,159%	30	-0,571%
		60	3,784%	60	-7,374%	60	-1,371%	60	2,607%
	Держать	Кол-во акций:	2	Кол-во акций:	0	Кол-во акций:	3	Кол-во акций:	5
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,354%	1		1	1,211%	1	0,868%
		10	0,727%	10		10	2,343%	10	1,697%
		30	4,780%	30		30	4,860%	30	4,828%
		60	-1,620%	60		60	5,882%	60	2,882%
	Покупать	Кол-во акций:	1	Кол-во акций:	0	Кол-во акций:	10	Кол-во акций:	11
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,932%	1		1	0,525%	1	0,392%
		10	1,143%	10		10	1,542%	10	1,506%
		30	3,874%	30		30	0,531%	30	0,835%
		60	0,628%	60		60	0,365%	60	0,389%
Итого		Кол-во акций:	25	Кол-во акций:	1	Кол-во акций:	17	Кол-во акций:	43
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,623%	1	-0,743%	1	0,739%	1	0,637%
		10	2,504%	10	-0,844%	10	1,434%	10	2,003%
		30	0,162%	30	-5,401%	30	1,132%	30	0,416%
		60	3,225%	60	-7,374%	60	0,930%	60	2,072%

Источник: расчеты автора

Приложение 17. Сравнение результативности рекомендаций, данных с помощью Lasso и 3-факторной модели Фамы-Френча на бычьем рынке

В таблице представлены абсолютные доходности.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации Fama-French	Продавать	Кол-во акций:	15	Кол-во акций:	1	Кол-во акций:	1	Кол-во акций:	17
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,115%	1	-1,658%	1	0,886%	1	-0,147%
		10	-1,035%	10	-4,251%	10	0,707%	10	-1,121%
		30	2,823%	30	2,631%	30	3,627%	30	2,859%
		60	3,706%	60	4,022%	60	3,489%	60	3,712%
	Держать	Кол-во акций:	0	Кол-во акций:	3	Кол-во акций:	2	Кол-во акций:	5
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1	-0,791%	1	0,456%	1	-0,292%
		10		10	-1,715%	10	2,966%	10	0,157%
		30		30	-1,127%	30	0,560%	30	-0,452%
		60		60	12,763%	60	1,515%	60	8,264%
	Покупать	Кол-во акций:	0	Кол-во акций:	0	Кол-во акций:	21	Кол-во акций:	21
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1		1	-0,707%	1	-0,707%
		10		10		10	-0,543%	10	-0,543%
		30		30		30	3,153%	30	3,153%
		60		60		60	6,989%	60	6,989%
Итого		Кол-во акций:	15	Кол-во акций:	4	Кол-во акций:	24	Кол-во акций:	43
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,115%	1	-1,008%	1	-0,544%	1	-0,438%
		10	-1,035%	10	-2,349%	10	-0,199%	10	-0,690%
		30	2,823%	30	-0,187%	30	2,957%	30	2,618%
		60	3,706%	60	10,578%	60	6,387%	60	5,842%

Источник: расчеты автора

Приложение 18. Сравнение результативности рекомендаций, данных с помощью Lasso и 3-факторной модели Фамы-Френча на бычьем рынке

В таблице представлены доходности, которые превышают рыночную доходность.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации Fama-French	Продавать	Кол-во акций:	15	Кол-во акций:	1	Кол-во акций:	1	Кол-во акций:	17
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	1,057%	1	-0,487%	1	2,057%	1	1,025%
		10	0,184%	10	-3,032%	10	1,925%	10	0,097%
		30	0,253%	30	0,060%	30	1,057%	30	0,289%
		60	-2,664%	60	-2,348%	60	-2,882%	60	-2,659%
	Держать	Кол-во акций:	0	Кол-во акций:	3	Кол-во акций:	2	Кол-во акций:	5
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1	0,381%	1	1,628%	1	0,880%
		10		10	-0,497%	10	4,185%	10	1,376%
		30		30	-3,697%	30	-2,010%	30	-3,022%
		60		60	6,393%	60	-4,856%	60	1,893%
	Покупать	Кол-во акций:	0	Кол-во акций:	0	Кол-во акций:	21	Кол-во акций:	21
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1		1	0,464%	1	0,464%
		10		10		10	0,675%	10	0,675%
		30		30		30	0,583%	30	0,583%
		60		60		60	0,619%	60	0,619%
Итого		Кол-во акций:	15	Кол-во акций:	4	Кол-во акций:	24	Кол-во акций:	43
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	1,057%	1	0,164%	1	0,628%	1	0,734%
		10	0,184%	10	-1,131%	10	1,020%	10	0,528%
		30	0,253%	30	-2,758%	30	0,387%	30	0,047%
		60	-2,664%	60	4,207%	60	0,017%	60	-0,529%

Источник: расчеты автора

Приложение 19. Сравнение результативности рекомендаций, данных с помощью Lasso и модели SARМ на бычьем рынке

В таблице представлены абсолютные доходности.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации SARМ	Продавать	Кол-во акций:	15	Кол-во акций:	1	Кол-во акций:	2	Кол-во акций:	18
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,115%	1	-1,658%	1	1,540%	1	-0,017%
		10	-1,035%	10	-4,251%	10	-0,619%	10	-1,167%
		30	2,823%	30	2,631%	30	-0,099%	30	2,488%
		60	3,706%	60	4,022%	60	2,333%	60	3,571%
	Держать	Кол-во акций:	0	Кол-во акций:	3	Кол-во акций:	1	Кол-во акций:	4
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1	-0,791%	1	-1,282%	1	-0,914%
		10		10	-1,715%	10	7,875%	10	0,682%
		30		30	-1,127%	30	4,945%	30	0,391%
		60		60	12,763%	60	1,852%	60	10,035%
	Покупать	Кол-во акций:	0	Кол-во акций:	0	Кол-во акций:	21	Кол-во акций:	21
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1		1		1	-0,707%	1	-0,707%
		10		10		10	-0,543%	10	-0,543%
		30		30		30	3,153%	30	3,153%
		60		60		60	6,989%	60	6,989%
Итого		Кол-во акций:	15	Кол-во акций:	4	Кол-во акций:	24	Кол-во акций:	43
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-0,115%	1	-1,008%	1	-0,544%	1	-0,438%
		10	-1,035%	10	-2,349%	10	-0,199%	10	-0,690%
		30	2,823%	30	-0,187%	30	2,957%	30	2,618%
		60	3,706%	60	10,578%	60	6,387%	60	5,842%

Источник: расчеты автора

Приложение 20. Сравнение результативности рекомендаций, данных с помощью Lasso и 3-факторной моделью Фамы-Френча на нейтральном рынке

В таблице представлены доходности, которые превышают рыночную доходность.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации Fama-French	Продавать	Кол-во акций:	16	Кол-во акций:	3	Кол-во акций:	2	Кол-во акций:	21
		Доходности (дней):				Доходности (дней):			
		1	0,330%	1	-0,251%	1	-0,440%	1	0,174%
		10	-0,482%	10	-1,686%	10	-0,986%	10	-0,702%
		30	-0,453%	30	-1,223%	30	1,281%	30	-0,398%
		60	0,3602%	60	-4,056%	60	-1,310%	60	-0,430%
	Держать	Кол-во акций:	1	Кол-во акций:	2	Кол-во акций:	1	Кол-во акций:	4
		Доходности (дней):				Доходности (дней):			
		1	0,093%	1	0,164%	1	0,558%	1	0,245%
		10	-1,265%	10	-0,935%	10	-2,546%	10	-1,420%
		30	-7,299%	30	3,608%	30	4,297%	30	1,054%
		60	-10,552%	60	-7,260%	60	8,188%	60	-4,221%
	Покупать	Кол-во акций:	2	Кол-во акций:	1	Кол-во акций:	18	Кол-во акций:	21
		Доходности (дней):				Доходности (дней):			
		1	-1,189%	1	0,845%	1	-0,304%	1	-0,334%
		10	-2,130%	10	1,399%	10	-0,494%	10	-0,559%
		30	-1,434%	30	-5,941%	30	-3,179%	30	-3,144%
		60	-9,011%	60	-6,033%	60	-4,968%	60	-5,403%
Итого		Кол-во акций:	19	Кол-во акций:	6	Кол-во акций:	21	Кол-во акций:	46
		Доходности (дней):				Доходности (дней):			
		1	0,158%	1	0,070%	1	-0,276%	1	-0,052%
		10	-0,697%	10	-0,921%	10	-0,638%	10	-0,699%
		30	-0,917%	30	-0,399%	30	-2,398%	30	-1,525%
		60	-1,201%	60	-5,454%	60	-3,993%	60	-3,030%

Источник: расчеты автора

Приложение 21. Сравнение результативности рекомендаций, данных с помощью Lasso и 3-факторной моделью Фамы-Френча на нейтральном рынке

В таблице представлены абсолютные доходности.

		Рекомендации Lasso						Итого	
		Продавать		Держать		Покупать			
Рекомендации Fama-French	Продавать	Кол-во акций:	16	Кол-во акций:	3	Кол-во акций:	2	Кол-во акций:	21
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,318%	1	-0,263%	1	-0,452%	1	0,162%
		10	1,657%	10	0,453%	10	1,153%	10	1,437%
		30	2,407%	30	1,638%	30	4,141%	30	2,462%
	60	2,8960%	60	-1,520%	60	1,226%	60	2,106%	
	Держать	Кол-во акций:	1	Кол-во акций:	2	Кол-во акций:	1	Кол-во акций:	4
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	0,081%	1	0,152%	1	0,546%	1	0,233%
		10	0,874%	10	1,205%	10	-0,406%	10	0,720%
		30	-4,438%	30	6,468%	30	7,157%	30	3,914%
	60	-8,016%	60	-4,725%	60	10,723%	60	-1,686%	
	Покупать	Кол-во акций:	2	Кол-во акций:	1	Кол-во акций:	18	Кол-во акций:	21
		Доходности (дней):		Доходности (дней):		Доходности (дней):		Доходности (дней):	
		1	-1,201%	1	0,833%	1	-0,316%	1	-0,346%
		10	0,010%	10	3,539%	10	1,646%	10	1,580%
		30	1,426%	30	-3,081%	30	-0,319%	30	-0,284%
	60	-6,475%	60	-3,497%	60	-2,432%	60	-2,868%	
	Итого		Кол-во акций:	19	Кол-во акций:	6	Кол-во акций:	21	Кол-во акций:
Доходности (дней):			Доходности (дней):		Доходности (дней):		Доходности (дней):		
1			0,146%	1	0,058%	1	-0,288%	1	-0,064%
10			1,443%	10	1,218%	10	1,501%	10	1,440%
30			1,944%	30	2,461%	30	0,462%	30	1,335%
60	1,335%	60	-2,918%	60	-1,457%	60	-0,494%		

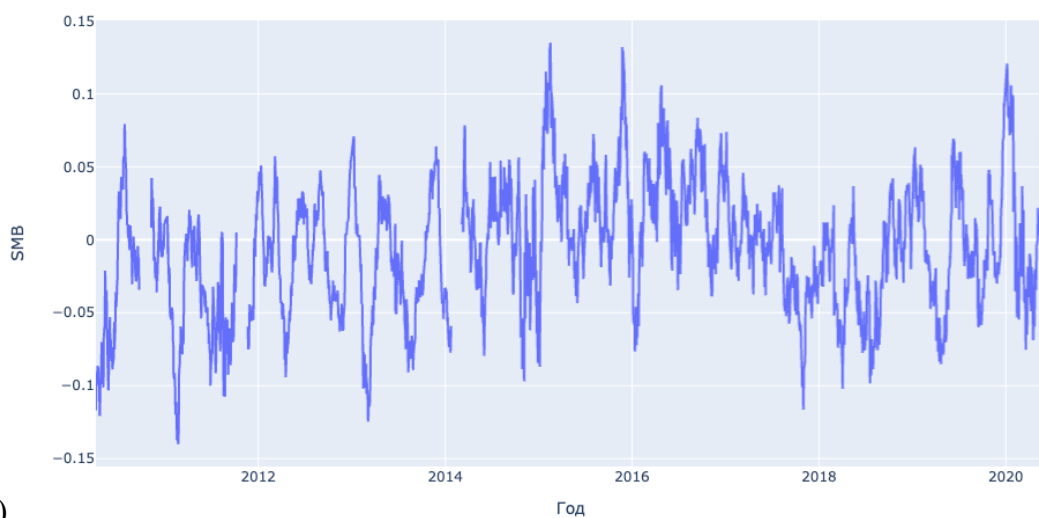
Источник: расчеты автора

Приложение 22. Соответствие тикеров и компаний, используемых при построении моделей

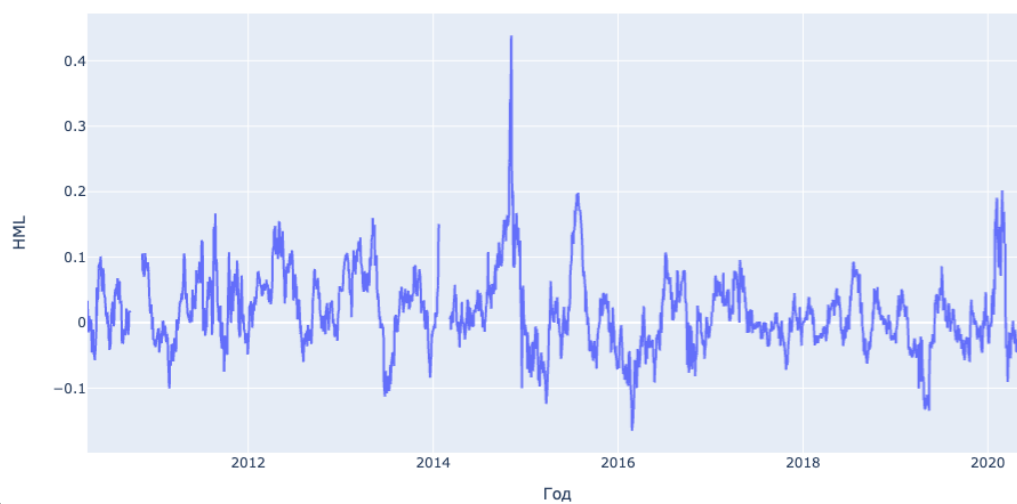
№	Тикер	Компания	№	Тикер	Компания
0	SBER	Сбербанк	24	AFLT	Аэрофлот
1	GAZP	Газпром	25	LSRG	ЛСР
2	LKOH	Лукойл	26	RSTI	Россети
3	GMKN	Норникель	27	UPRO	Юнипро
4	NVTK	Новатэк	28	LNTA	Лента
5	SNGS	Сургутнефтегаз	29	MVID	М.Видео
6	PLZL	Полюс	30	TRMK	ТМК
7	TATN	Татнефть	31	MTLR	Мечел
8	ROSN	Роснефть	32	NMTP	НМТП
9	MGNT	Магнит	33	AKRN	Акрон
10	MTSS	МТС	34	MSNG	Мосэнерго
11	MOEX	МосБиржа	35	URKA	Уралкалий
12	IRAO	Интер РАО	36	BANE	Башнефть
13	NLMK	НЛМК	37	VSMO	ВСМПО-Ависма
14	ALRS	Алроса	38	NKNC	Нижнекамскнефтехим
15	CHMF	Северсталь	39	GCHE	Черкизово
16	VTBR	ВТБ	40	SVAV	Соллерс
17	RTKM	Ростелеком	41	MSTT	Мостотрест
18	PHOR	Фосагро	42	BSPB	Банк Санкт-Петербург
19	AFKS	АФК «Система»	43	MSRS	МОЭСК
20	MAGN	ММК	44	KMAZ	Камаз
21	PIKK	ПИК	45	RASP	Распадская
22	HYDR	Русгидро	46	OGKB	ОГК №2
23	FEES	ФСК ЕЭС	47	VZRZ	Банк "Возрождение"

Источник: составлено автором на основе данных МосБиржи

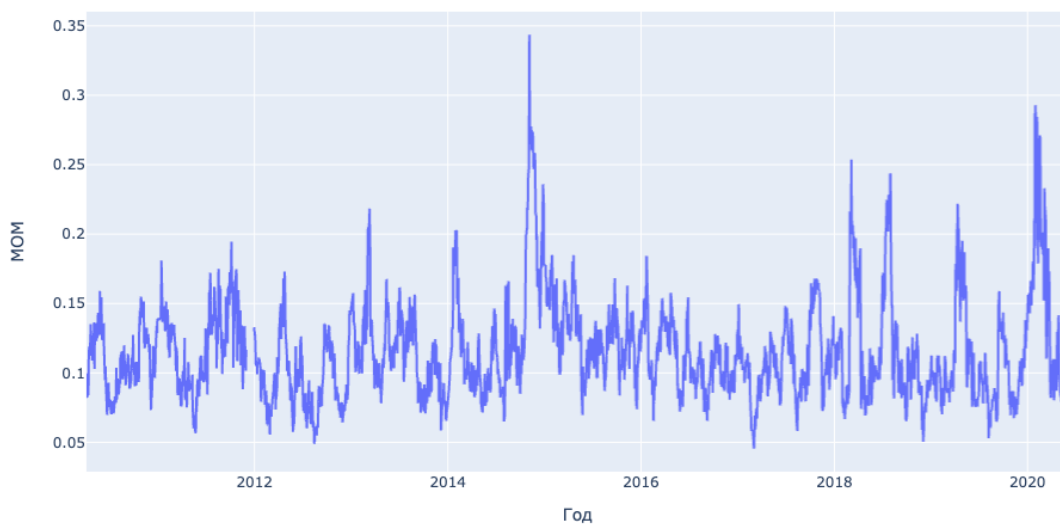
Приложение 23. Динамика классических факторов модели Фамы-Френча-Кархарта за период с 13.08.2010 по 04.06.2020



а)



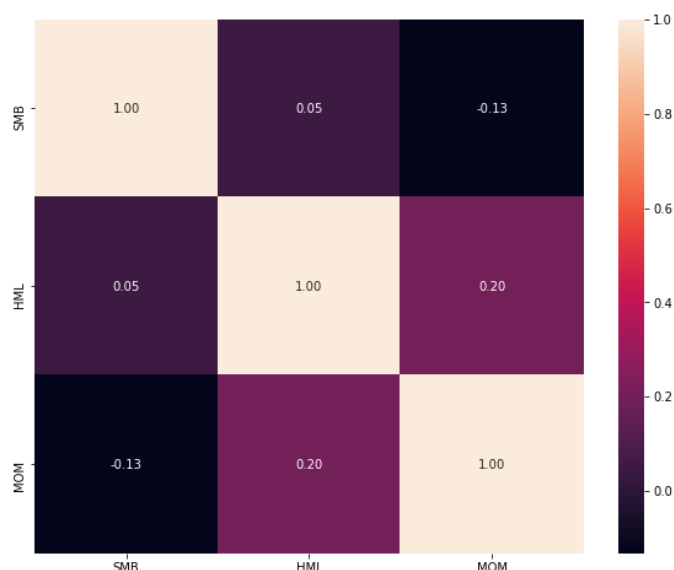
б)



в)

Источник: составлено автором

Приложение 24. Корреляционная матрица классических факторов модели Фамы-Френча-Кархарта



Приложение 25. Результаты проверки данных на стационарность с помощью расширенного теста Дики-Фуллера при построении классических линейных моделей доходности акции, а также их расширений для Сбера

P-value при проведении ADF теста составило $1.4711353642e-20$, поэтому премия за рыночный риск для Сбера стационарна,

P-value при проведении ADF теста составило $9.167462503430895e-14$, поэтому доходность акций Сбера стационарна,

P-value при проведении ADF теста составило $2.689501089068641e-12$, поэтому первые разности (месячные) для курса рубль/доллар стационарны,

P-value при проведении ADF теста составило $2.2919574273281916e-13$, поэтому фактор SMB стационарен,

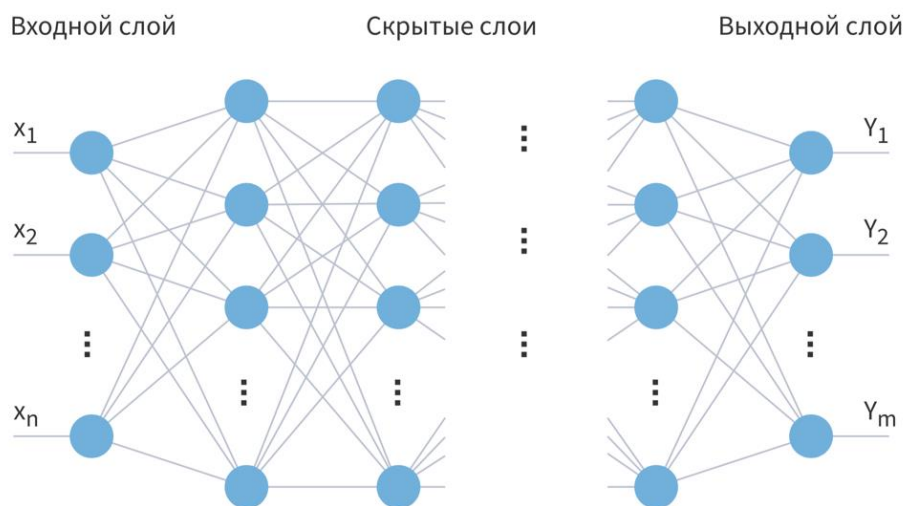
P-value при проведении ADF теста составило $8.155779676005055e-12$, поэтому фактор HML стационарен,

P-value при проведении ADF теста составило $5.575353483573121e-12$, поэтому фактор MOM стационарен,

P-value при проведении ADF теста составило $9.127276708193843e-12$, поэтому первые разности (месячные) для цены на нефть стационарны

Приложение 26. Иллюстрация архитектуры базовой нейронной сети – многослойного персептрона

Это нейронная сеть с n входными слоями, m выходными слоями и каким-то количеством скрытых слоёв.



Источник: <https://wiki.loginom.ru/articles/neural-network.html>