

Министерство науки и высшего образования
Российской Федерации
Московский Физико-технический институт
(Государственный Университет)

Магистерская диссертация
**"Разработка алгоритмов для поиска арбитражных сделок на
рынке ценных бумаг"**

Студент
Козуб Валентина Романовна

Научный руководитель
Горбатилов Андрей Анатольевич

Москва, 2022

Содержание

Используемые сокращения	3
1 Введение	4
2 Анализ рынка роботизированной торговли	5
2.1 Общее описание рынка	5
2.2 Развитие рынка	7
2.3 Основные рынки для торговли	9
2.4 Основные технологии	10
2.5 Основные производители и продукты - конкуренты	14
2.6 Тенденции развития рынка	17
2.7 Описание целевой аудитории – потребителей продукта проекта	19
2.8 Резюме	20
3 Статистический арбитраж	21
3.1 Метрические стратегии	24
3.2 Коинтеграционные стратегии	26
3.3 Стратегии на основе стохастического моделирования	28
3.4 Стратегии на основе машинного обучения	30
3.5 Статистический арбитраж на Российских рынках	32
3.6 Резюме	34
4 Технологическое описание алгоритма	35
4.1 Логическое описание алгоритма	35
4.2 Корреляция	37
4.3 Метод наименьших квадратов	38
4.4 Нормировка	40
4.5 Поиск оптимальной точки	41
4.6 Данные для исследования и параметры алгоритма	42
4.7 Оценка возможных потерь	44
4.8 Детали работы программы	46
4.9 Результат работы алгоритма	47
4.10 Резюме	50
5 Вывод	51
Список источников	53

Используемые сокращения

1. МНК - метод наименьших квадратов
2. SSD - sum of squared distances (Сумма квадратов расстояний)
3. CRSP - Center in research in security prices (центр исследований цен на финансовые инструменты)
4. HFT - High-frequency trading (высокочастотная торговля)
5. DMA - Direct market access (прямой доступ к рынку)
6. ETF - Exchange Traded Funds (биржевой фонд)
7. FOREX - Foreign Exchange (валютный обмен)
8. н.с.в - независимые случайные величины
9. ОУ процесс - процесс Орнштейна-Ухлебека
10. DBN - Deep Brief Network

1 Введение

В связи с распространением компьютерных и сетевых технологий появилась возможность совершать сделки на бирже, используя автоматические торговые алгоритмы. В основном такими системами пользуются крупные финансовые организации: банки, фонды, биржи. Сейчас объем операций, производимых в автоматическом режиме по данным МосБиржи, в среднем составляет более 60% и может достигать более 80% в моменте.

Особенностью рынка торговых роботов является нежелание производителей делиться своими наработками. Такие роботы извлекают пользу из несовершенства рынка, при этом объемы операций ограничены ликвидностью и, как следствие, ограничена максимальная потенциальная прибыль робота. Это ведет к тому, что компании не предоставляют доступ к своим алгоритмам, а эффективность тех немногочисленных систем, которые находятся в открытом доступе, сомнительна.

Кроме того, самыми популярными стратегиями для торговых роботов являются анализ тренда и временных рядов. Во многом их распространенность объясняется возможностью применения к широкому спектру финансовых инструментов, однако они не учитывают экономические показатели и не балансируют портфель вкладами в различные активы, и, как следствие, отличаются высоким риском.

Целью данной работы является создание алгоритма, обладающего большей надежностью, чем импульсная торговля и анализ тренда. Арбитражные стратегии считаются менее рискованными в отличие от остальных. Поэтому в данной работе был разработан алгоритм для торгового робота на основе статического арбитража. Алгоритм извлекает прибыль, анализируя изменение цен двух коррелирующих активов и выявляя момент, когда один из активов недооценен. Для определения оптимальной точки входа используется прогнозирование спреда между активами на основе метода наименьших квадратов. В работе тестирование алгоритма проводится на парах акций, так как акции являются самым доступным и распространенным финансовым инструментом.

Помимо этого, в работе также проведен детальный анализ рынка роботизированной торговли, произведена оценка перспективности данного способа трейдинга. Изучена информация об открытых системах роботизированной торговли, персонализированы наиболее репрезентативные (цитируемые) публикации по теме статистического арбитража. Предложенный алгоритм торговли программно реализован, качество его работы оценено применительно к Московской и Санкт-Петербургской биржам на актуальном срезе данных за 2021 год. Произведена оценка рисков и экономической эффективности разработанного алгоритма.

2 Анализ рынка роботизированной торговли

2.1 Общее описание рынка

В настоящее время алгоритмическая торговля достаточно популярна.

Существует два распространенных типа роботов: полностью автоматизированные системы и полуавтоматические, то есть рекомендательные роботы или советники.

Полностью автоматизированная торговля - это форма, при которой компьютерная программа используется для выполнения определенного набора инструкций или правил, которые включают покупку или продажу актива в отношении меняющихся рыночных данных. Пользователям просто нужно настроить технические параметры, а компьютерная программа сделает все остальное.

В полуавтоматической системе робот генерирует сигналы к покупке или продаже соответствующего актива, при этом не может совершать сделку самостоятельно. Пользователь должен вручную выполнить сделку на основе предложения робота.

Преимущества роботов. Первое преимущество использования робота заключается в том, что он устраняет эмоциональный аспект торговли. Робот будет выполнять сделку на основе алгоритмов, которые не подвержены чувствам и эмоциям. Он проведет сделку независимо от незначительных отклонений, которые могли бы сбить с толку даже самых профессиональных трейдеров. Этот инструмент поможет трейдеру избежать паники при принятии торговых решений. Например, он не закроет сделку, опасаясь, что торговая сессия откроется немного выше или ниже, чем обычно. Автоматизированное программное обеспечение поможет трейдеру выполнять стратегии, основанные на заранее определенных правилах.

Еще одним большим, но очевидным преимуществом использования робота является то, что он остается полностью функциональным 24/7. Полностью автоматизированный робот не спит и продолжает торговать как днем, так и ночью. Используя такое программное обеспечение, не нужно постоянно следить за рынком.

Кроме этого, робот может производить большое количество операций за короткое время, гораздо быстрее человека и без остановок. Высокочастотные роботы способны совершать несколько операций в минуту.

И конечно, робот способен обрабатывать гораздо больший объем информации о рынке, чем человек. Алгоритм может ежеминутно сканировать десятки активов, проводить вычисления и выдавать результат.

Еще одним преимуществом является возможность тестирования различных торговых идей. Такое тестирование на исторических данных можно выполнить на различных активах.

Недостатки использования роботов. Роботам требуется круглосуточное подключение к Интернету для автоматического совершения сделок. Кроме того, сервер, на котором размещен робот, должен быть надежным и обеспечивать максимальное время безотказной работы. Задержка или сбой на сервере могут привести к тому, что робот будет совершать неправильные сделки.

Торговый робот это высокотехнологичный инструмент, принцип работы которого является коммерческой тайной и не разглашается производителем. Кроме того на такие инструменты не предоставляются гарантии. Поэтому заранее объективно оценить качество робота крайне сложно.

Подводя итог, роботы предлагают трейдерам возможность действовать в соответствии с рыночными возможностями 24 часа в сутки. Робот выполняет сделку в наиболее подходящее время, что приводит к максимальной прибыли. Он также заботится об эмоциональных аспектах при торговле на рынке, тем самым сводя к минимуму вероятность совершения неудачных сделок.

Однако недостатком использования автоматической торговли является то, что качество алгоритмов и их доходность не гарантируется.

В целом можно сказать, что роботы ускоряют и упрощают процесс биржевой торговли, увеличивают ликвидность, и самое главное, позволяют зарабатывать практически каждому участнику трейдинга.

2.2 Развитие рынка

Алгоритмическая торговля начала активно развиваться в конце 20 века. Сейчас крупнейшие брокерские и финансовые компании заключают сделки в основном автоматически.

Доля роботизированной торговли в разных странах составляет от 40 до 80%.



*По данным аналитического агентства Rangea Strategic Intelligence [1]

По разным подсчетам доля торговых роботов в России составляет от 50% до 80% и ежегодно данная цифра увеличивается. Это достаточно высокий показатель, по которому можно сказать, что роботизированная торговля не является новшеством. Роботы на данный момент широко используются и составляют значительную долю рынка.



*По данным МосБиржи [2]

На графике представлена динамика объема торгов на МосБирже. За последние три года объем торгов вырос почти в три за (по сравнению с 2019 годом), за 2021 на 25,5%.

Такой резкий рост объясняется в первую очередь алгоритмической торговлей, так как роботы совершают большое количество сделок и тем самым увеличивают объем торгов.

В первом квартале 2022 года объем составил 9 295, что на 30% превышает показатели аналогичного квартала 2021 года. Это свидетельствует о продолжении активного роста объемов торгов и увеличении числа роботов.

2.3 Основные рынки для торговли

Валютный рынок (FOREX). Это глобальный «децентрализованный» или «внебиржевой» рынок для торговли валютами. Рынок FOREX определяет обменные курсы для каждой валюты в мире. Это один из наиболее активно торгуемых рынков, и из-за его часто непредсказуемого и изменчивого характера алгоритмическая торговля может быть очень выгодной для трейдеров.

Фондовые рынки. Фондовый рынок – это все рынки и биржи, на которых осуществляется регулярная деятельность по покупке и продаже публично торгуемых акций компаний. Будучи одним из самых популярных финансовых рынков, трейдеры часто используют алгоритмы, основанные на их торговых стратегиях, что повышает точность и скорость их сделок.

Биржевые фонды (ETF). Это тип инвестиционного фонда, акции которого торгуются на фондовых биржах. Они состоят из различных отдельных акций и часто являются отличным способом для инвесторов получить доступ к широкому спектру облигаций и акций. Алгоритмическая торговля используется для торговли ETF, однако большинство ETF используются в качестве долгосрочных инвестиций, что делает их менее популярными для алгоритмической торговли, чем фондовый рынок или FOREX.

Облигации. Облигация — это платеж, сделанный инвестором заемщику (компании или правительству). Алгоритмическая торговля на рынке облигаций не редкость, но сопряжена с определенными трудностями. Каждая облигация имеет разные характеристики, такие как процентные ставки, сроки погашения и условия. Алгоритмы должны учитывать доходность к погашению, доходность к отзыву, фактическую доходность, кредитный рейтинг и многое другое, чтобы отличить хорошую облигацию от плохой. Это требует сложных и хорошо отработанных алгоритмов.

Криптовалюты. Криптовалюта — это цифровая/виртуальная валюта, транзакции в которой осуществляются через децентрализованную систему, такую как Биткойн или Эфириум. Криптовалютный рынок чрезвычайно волатилен и очень непредсказуем. Из-за огромной волатильности этого рынка алгоритмическая торговля может быть чрезвычайно прибыльной, но также может привести к большим потерям, если алгоритм не совершенен.

Другие. Рынки для алгоритмической торговли бесконечны, включая товары, такие как золото или нефть, кредитно-дефолтные свопы, процентные свопы и рынок деривативов. Приложения и способы использования алгоритмического трейдинга бывают самых разных форм и размеров, каждый из которых имеет свои преимущества и риски.

2.4 Основные технологии

Существует большое количество разных стратегий, самые популярные:

Статистическая. Эта стратегия основана на анализе исторических данных временных рядов, которые в этом контексте будут относиться к цене предложения (покупки-продажи) актива.

Эти точки данных, часто являющиеся ценой закрытия самой высокой или самой низкой ценой покупки-продажи, записываются через регулярные промежутки времени в течение определенного периода времени. Анализируя движение выбранных точек данных, можно определить тенденции и возможности на основе этих данных.

Затем эти исторические тенденции можно использовать для сравнения с текущими рыночными данными, чтобы можно было определить прибыльные торговые возможности.

Импульсная торговля. Помимо простого изучения движения цен на активы, существует стратегия, которая изучает импульс тренда. Она основывается на использовании технических торговых индикаторов, чтобы дать сигналы о том, когда тренд изменится. Примерами часто используемых индикаторов являются скользящие средние и стохастический осциллятор. Скользящее среднее - это множество точек, каждая из которых равна средней цене актива за определенный промежуток времени. Стохастический осциллятор - показатель, который отражает динамику цены по сравнению с предыдущим значением.

Скользящие средние помогают определить долгосрочный тренд. Вместо того, чтобы вручную отображать изменения цен, алгоритм можно запрограммировать на отслеживание ценовых движений в виде 20-дневных или 50-дневных средних значений. Чем короче дневная средняя, тем ближе она будет к самой последней цене актива.

Стохастические осцилляторы часто используются как признаки перекупленности или перепроданности. Они могут сигнализировать о приближении восходящего или нисходящего тренда из-за перекупленности или перепроданности актива.

Прибыль таких алгоритмов зависит от качества предсказанного поведения котировок, в случае нестандартных ситуаций на рынке или технического сбоя алгоритм может достаточно быстро проиграть все деньги.

Следование за трендом. Будучи одной из самых простых стратегий, эта стратегия алготрейдинга включает в себя следование рыночным тенденциям. Когда тренд растущий, алгоритм занимает длинную позицию. Когда он ниспадающий, алгоритм, скорее всего, закроет короткую позицию.

Для определения длительности действия тренда такие алгоритмы обычно используют исторические данные.

Данная стратегия очень похожа на импульсную торговлю, использует те же индикаторы. Их отличие заключается во временных промежутках. Импульсная торговля заключается в поиске резких выбросов, которые не длятся долго. Соответственно импульсная торговля - это краткосрочные сделки. Трендовая торговля заключается в поиске общего направления движения котировок и удержании сделки до конца этого тренда.

Новостная стратегия. События внутри страны или в другой стране могут оказать влияние на мировые валюты и фондовые рынки. Например, политические волнения, пандемии, выборы, инфляция, война и так далее.

Торговая стратегия, основанная на новостях, запрограммирована реагировать на новости. Система предназначена для отслеживания новостных лент, а затем для генерации торговых сигналов на основе этих событий в режиме реального времени.

Однако события, попадающие в основные новостные агентства, как правило, устаревают, поскольку преимущество заключается в наличии информации или знаний, позволяющих действовать раньше конкурентов.

Кроме этого, есть новости, которые приходят не от новостных агентств и журналистов, а от известных личностей в социальных сетях и других источниках. Например, последнее время на рынок сильно оказывают влияние высказывания Илона Маска в твиттере. Алгоритмы пока не способны собирать информация из всевозможных источников.

Поэтому данная стратегия не является популярной, трейдеры предпочитают игнорировать новостной фон и реагировать только на движения котировок.

Арбитраж — это одновременная покупка и продажа одного и того же актива на разных рынках с целью получения прибыли от крошечных различий в заявленной цене актива. Он использует краткосрочные колебания цен на идентичные или похожие финансовые инструменты на разных рынках или в разных формах. Арбитраж существует в результате неэффективности рынка, и он одновременно использует эту неэффективность и устраняет ее.

Раньше арбитраж был более прибыльным. Теперь, из-за большого количества торговых роботов, ценовые аномалии не остаются надолго. Таким образом, арбитражная стратегия должна осуществляться путем торговли очень большими объемами, чтобы получить достаточно существенную прибыль. А так же необходимо высокотехнологичное оборудование с быстрым доступом к бирже.

Статистический арбитраж - инвестиционная стратегия, которая основывается на поиске зависимых активов и использовании этой зависимости для получения прибыли. Является подвидом арбитража. Отличие заключается в том, что при простом ар-

битраже используются идентичные активы на разных площадках либо с разным сроком действия, а в статистическом арбитраже участвуют разные зависимые активы. Поэтому в статистическом арбитраже больше возможностей для сделок.

Маркет-мэйкинг. Данная стратегия заключается в одновременном поддержании заявок на покупку и продажу. Независимо от общего тренда, цена актива совершает небольшие колебания. Если правильно подобрать цены покупки и продажи, то на этих колебаниях можно зарабатывать, что и делают маркет-мэйкеры.

Часто маркет-мэйкинг подразумевает под собой доступ не только заявкам на покупку и продажу, но также к тейк-профиту (сигнал брокеру на продажу в случае когда цена поднимется до определенного уровня, используется для того, чтобы не упустить прибыль) и стоп-лоссу (сигнал брокеру на продажу в случае когда цена опустится до определенного уровня, используется для того, чтобы ограничить убытки). Это позволяет маркет-мэйкерам прогнозировать дальнейшее поведение цены.

Стоит отметить, что данная стратегия положительно влияет на рынок, потому что, во-первых, обеспечивает ликвидность, во-вторых, удерживает цену актива от сильных колебаний. Часто биржи поддерживают маркет-мэйкеров снижением комиссий или даже денежным вознаграждением.

Алгоритмы на основе ИИ. Относительно новая форма алготрейдинга использует машинное обучение и искусственный интеллект (ИИ). Последнее время данное направление активно развивается.

Преимущество заключается в том, что, анализируя огромные объемы данных, алгоритм самостоятельно извлекает из них опыт, это позволяет находить сложные нестандартные зависимости, которые человек не способен заметить.

Кроме стратегических особенностей, роботы отличаются используемыми инструментами, далее будут приведены примеры нескольких из них:

Высоочастотная торговля (HFT). Одной из подкатегорий алготрейдинга является высоочастотная торговля (HFT). HFT предполагает использование алгоритмов для совершения до сотен тысяч сделок за доли секунды.

Учитывая, что алгоритмы автоматизированы, они могут работать с такой частотой и скоростью, с которой не может работать ни один трейдер-человек.

Ручные трейдеры будут подвержены ограничениям своих физических и умственных способностей, в то время как алгоритм может отслеживать движение цен и работать 24 часа в сутки 7 дней в неделю без простоев.

Следовательно, алгоритмы могут определять и фиксировать лучшие цены за миллисекунды, например, при торговле парами с самыми высокими спредами между ценами покупки и продажи в любое время дня. Это максимизирует потенциальную прибыль

для трейдеров с большей легкостью.

Но для работы такого робота необходимо мощное оборудование, хорошее качество соединения и быстрый доступ к данным биржи. Из-за этого высокочастотный трейдинг требует крупных капиталовложений на техническое обеспечение и использовать его могут только крупные компании.

Еще одним негативным моментом является риск. Данный инструмент может как быстро увеличивать доходность, так и быстро совершать сделки в минус в случае сбоя работы. Так в 2012 одной Американской компании удалось потерять свыше 400 млн. долларов за час.

Использование прямого доступа к рынку (DMA). Эта алгоритмическая торговая стратегия предполагает использование прямого доступа к рынку для размещения сделок. Прямой доступ к памяти означает доступ к сложной технологической инфраструктуре, которая часто принадлежит фирмам-продавцам.

Эта инфраструктура подключается к нескольким торговым платформам и содержит данные о стакане бирж финансового рынка.

Обычно сделки на бирже совершаются через брокера или маркет-мэйкера, DMA предполагает выставление заявок напрямую без посредников. Это позволяет значительно сократить время совершения сделки. Например, для обычного клиента время доступа к бирже составляет 50 - 250 миллисекунд, а при подключении с DMA 250 - 500 микросекунд, то есть выигрыш во времени увеличивается примерно в 1000 раз.

Конечно, для обычного алгоритма, который совершает только порядка 10 сделок в день, данное преимущество не является существенным, но для спекуляций с валютами и высокочастотной торговли DMA играет существенную роль.

2.5 Основные производители и продукты - конкуренты

Основными пользователями роботов являются фонды, банки и крупные финансовые организации, которые оперируют большими капиталами. Данные организации используют не готовых роботов, а создают собственных с нужным уровнем риска и прибыльностью или пользуются для этих целей услугами других компаний. В данный момент крупнейшими производителями торговых роботов являются следующие:

1. 63 Moons Technologies Limited [3] является мировым лидером в предоставлении технологических предприятий, инноваций, платформ и решений для создания цифровых рынков и торговых площадок, которые обеспечивают ценообразование и эффективность транзакций в отраслевых сегментах. Данная компания создает роботов на основе высокочастотной торговли, потребителям предлагается 14 типов стратегий, а также настраиваемый риск. То есть компания предоставляет инструменты, из которых предлагается собрать собственного робота. Компания также предлагает оборудование.

2. Virtu Financial [4] также является крупнейшей мировой организацией по предоставлению финансовых решений. Компания предлагает настраиваемые решения с торговыми технологиями, которые позволяют повысить доходность и снизить риск. Virtu Financial предоставляет сервисы аналитики с интерактивными инструментами. Производитель делает акцент на высокое качество предоставляемых решений и высокий уровень профессионализма экспертов, к которым всегда можно обратиться за помощью.

3. MetaQuotes [5] - компания, которая предоставляет не отдельных роботов, а целую платформу MetaTrader Market. Платформа предоставляет собственный язык программирования MQL5 для написания торговых роботов. Также реализованы все возможности для обеспечения безопасной реализации сделок купли-продажи роботов и аренды. Кроме этого реализован интерфейс, который отображает графики, различные технические индикаторы и возможности для тестирования роботов перед покупкой на исторических данных. На платформе есть всевозможные современные инструменты для торговых роботов: нейросети, советники, усредняющие системы. На данный момент платформа насчитывает более 10000 приложений и является крупнейшим в мире брокерским сообществом.

4. Algo Trader AG [6] - платформа для торговли криптовалютами. Торговая система основана на технологии блокчейн. Первым преимуществом компании является доступ к различным биржам со всего мира, что позволяет географически диверсифицировать капитал. Вторым преимуществом производитель считает минимальный операционный риск и риск исполнения заявок при использовании системы.

Провести детальный анализ готовых продуктов достаточно сложно, потому что

основными пользователями роботов являются фонды, банки и крупные финансовые организации, которые используют собственных роботов и не стремятся ими делиться. В такой отрасли, как алгоритмическая торговля, большое количество пользователей не выгодно производителю, так как ресурсы роботов ограничены ликвидностью рынка. Услуги компании, которые предлагают инструменты и инфраструктуру для создания торговых роботов, достаточно дорогостоящие и требуют значительно стартового капитала. В свободном доступе для частных инвесторов, как правило, есть либо роботы не очень высокого качества, либо роботы в режиме тестирования. Ниже представлен список самых популярных готовых решений:

1. **Learn2Trade [7]** - полуавтоматический робот на рынке FOREX. Платформа предоставляет порядка 3 сигналов в день для совершения какой-либо сделки.

Потенциальная доходность: 76% успешных прогнозов, информации о приросте капитала нет

Комиссия: 40 фунтов (около 2760 руб.) в месяц

Минимальный взнос: -

2. **Pionex [8]** - полностью автоматизированный робот на рынке криптовалюты.

Потенциальная доходность: непредсказуема, компания не предоставляет никакие конкретные характеристики

Комиссия: -

Минимальный взнос: от 20 долларов (около 1200 руб.) в месяц

3. **1000pip Climber System [9]** - полуавтоматическая платформа для торговли на FOREX. Есть три режима генерации сигналов: раз в 15 минут, раз в час и раз в 4 часа.

Потенциальная доходность: до 60-70%

Комиссия: 97 долларов (около 5820 руб.) в год

Минимальный взнос: -

4. **Forex Fury [10]** - советник, то есть полуавтоматический робот на FOREX рынке.

Потенциальная доходность: до 100%, зависит от уровня риска

Комиссия: разовая оплата от 300 долларов (около 18000 руб.)

Минимальный взнос: -

5. **Algo Signals [11]** - система для обеспечения торговых сигналов, то есть полуавтоматический робот с неограниченным выбором активов.

Потенциальная доходность: зависит от торговой площадки и уровня риска, численные показатели компания не предоставляет

Комиссия: -

Минимальный взнос: от 1500 долларов (около 90000 руб.)

6. **BinBot Pro [12]** - автоматический робот для торговли бинарными опционами

Потенциальная доходность: до 90%

Комиссия: -

Минимальный взнос: от 250 долларов (около 15000 руб.)

7. **EA Builder [13]** - робот советник с неограниченным выбором активов.

Потенциальная доходность: зависит от актива

Комиссия: разовая выплата от 97 долларов (около 6000 руб.)

Минимальный взнос: -

Все рассмотренные роботы используют разновидности стратегий импульсной торговли и следования тренду.

Таким образом, среди доступных готовых решений наиболее популярными являются роботы, торгующие на рынках криптовалюты и FOREX. Возможно, из-за слишком большого разнообразия акций и не такой высокой волатильности, по-сравнению с криптовалютой, фондовый рынок не так популярен.

Не каждая компания сообщает примерный уровень доходности, так как это зависит от настроек алгоритма, и ни одна не предоставляет гарантии доходности. По информации, которую удалось найти, максимальная доходность роботов составляет от 60 до 100% при торговле на крипторынке и FOREX.

2.6 Тенденции развития рынка

Алгоритмическая торговля пока не реализовала весь свой потенциал и продолжает активно развиваться. В предыдущих главах были рассмотрены несколько направлений, но самыми перспективными являются следующие:

Рост ИИ и машинного обучения– Алгоритмическая торговля нуждается в постоянном обновлении и изменении алгоритмов и стратегий в соответствии с торговой инфраструктурой. Ситуация на рынках меняется трейдеры вынуждены дополнять свои алгоритмы различными условиями и особенностями. В результате модели становятся более сложными, громоздкими, с большим количеством параметров, в которых сложно разобраться частному инвестору. Алгоритмы на основе ИИ могут сами выявлять закономерности, разработчику не обязательно детально прописывать модели и параметры. ИИ в автоматическом режиме подбирает нужные значения, что упрощает использование таких роботов. Доверие к ИИ и машинному обучению среди трейдеров и клиентов, использующих алгоритмическую торговлю, будет расти. В будущем мы можем увидеть мир полной автоматизации. Рост интеллекта машин может произвести революцию в трейдинге, развивая новые технологии, такие как квантовые вычисления, технология блокчейн и облачные вычисления. Машинное обучение в сочетании с перечисленными ниже возможными технологиями может открыть совершенно другое будущее для трейдинга и алгоритмической торговли: – Использование нанотехнологий в трейдинге – Микрочипы, позволяющие выполнять сделки за 74 наносекунды – Технология микроволновой передачи может позволить передавать данные со скоростью света

Улучшение качества прогнозирования. Заглядывая в будущее, мы увидим, что основные алгоритмы станут более интеллектуальными и реагирующими на живые рыночные условия. Алгоритмы, основанные на искусственном интеллекте/машинном обучении, будут прогнозировать микроволатильность и рыночные события. Это естественным образом повысит эффективность алгоритмической торговли за счет использования предсказательных механизмов и технологий, способных предвидеть изменения в реальном времени.

Таким образом, алгоритмы предоставят лучшие пред-торговые рекомендации относительно производительности всех доступных алгоритмов. Это означает, что трейдеры смогут использовать совершать сделки эффективнее.

Распространение по активам. Еще одна основная тенденция, которая наблюдалась в 2020/2021 гг., заключается в том, что алгоритмы используются во все большем количестве классов активов. С 2016 года использование алгоритмов в торговле фьючерсами увеличилось на 40% в годовом исчислении. JP Morgan также зафиксировал, что в 2016 на валютном рынке было зафиксировано менее 35% объема сделок с исполь-

зованием алгоритмов. В 2020 году этот показатель составлял почти 60%. В торговле на валютном рынке алгоритмы позволяют лучше выполнять и принимать решения и минимизировать торговые издержки. Ожидается, что алгоритмическая торговля будет постоянно расширяться за счет большего количества классов активов, поскольку доверие к ней растет, а выгоды от результатов становятся все более очевидными.

2.7 Описание целевой аудитории – потребителей продукта проекта

Алгоритмическая высокочастотная торговля требует большого количества высококлассного оборудования и прямого доступа к рынку. Этот вид в основном используется институциональными инвесторами и крупными брокерскими конторами для сокращения расходов, связанных с торговлей. Согласно исследованиям, такая алгоритмическая торговля особенно выгодна для больших размеров сделок, которые могут составлять до 10% от общего объема торгов.

Еще одними распространенными пользователями являются маркет-мейкеры и биржи. Они используют алгоритмы для автоматизации сделок покупки/продажи.

Также алгоритмами пользуются небольшие компании и частные инвесторы. как уже упоминалось ранее этому сегменту потребителей подобрать качественного работника сложно, так как не известны детали работы и нет гарантий. Данная работа ориентирована именно на данный сегмент потребителей.

2.8 Резюме

Таким образом, в целом из анализа рынка можно сделать следующие выводы:

1. Алгоритмическая торговля занимает значительную долю рынка и продолжает развиваться дальше. Роботизированная торговля позволяет проводить в автоматическом режиме анализ активов, находить возможности для сделок там, где человек не способен их заметить. Полностью автоматизированные роботы способны работать 24/7 без вмешательства человека. А также значительным плюсом является отсутствие эмоционального аспекта.

2. На рынке представлено большое количества различных стратегий. Среди доступных роботов самыми популярными являются импульсная торговля и следование за трендом. Данные направления отличаются возможностью торговать абсолютно любым активом. В рамках данной работы реализован принципиально другой алгоритм на основе статистического арбитража.

3. Есть несколько площадок для алгоритмической торговли: валюта, фондовый рынок, криптовалюта, фонды, облигации... В данной работе алгоритм будет тестироваться на акциях Российского инструмента, так как данный инструмент является самым легкодоступным.

4. Алгоритмической торговлей пользуются не только крупные финансовые организации, но и частные инвесторы. Но при отсутствии значительного капитала выбрать торгового робота достаточно сложно. Представленный в данной работе алгоритм ориентирован на частного инвестора или компанию с небольшим капиталом, так как не требует больших затрат на инфраструктуру, то есть не использует высокочастотную торговлю и прямое подключение к рынку.

5. На рынке есть готовые решения в том числе и для частного инвестора с небольшим капиталом, но детальная информация о работе алгоритмов является коммерческой тайной и качество таких роботов крайне сложно проверить. К тому же производитель не предоставляет гарантий. В данной работе предлагаемый алгоритм детально описан.

6. По информации, которую удалось найти, максимальная доходность роботов составляет до 60 - 100% при торговле на крипторынке и FOREX. При этом производители не указывают средний размер доходности. Поэтому сложно судить о справедливости таких показателей.

3 Статистический арбитраж

Алгоритмы, позволяющие эффективно вкладывать капитал в различные биржевые инструменты представляют большой интерес для научного сообщества. Помимо непосредственной практической пользы, заключающейся в максимизации доходности имеющегося капитала, многие работы затрагивают в том числе вопросы, связанные с выявлением факторов, которые влияют на цену активов, а также пытаются анализировать причины, по которым цена актива отклоняется от справедливой стоимости.

В то время, как информация об устройстве большинства коммерческих разработок держится в тайне, научные публикации подробно описываются все детали предлагаемых в них стратегий, а также данные на которых производилась оценка качества, что позволяет составить более четкое представление о современных методах алгоритмической торговли и возможностях их применения.

Существует много различных подходов к созданию торговых алгоритмов. Их описание и анализ приводился в разделе 2.4. В данном же разделе будут рассматриваться только стратегии, основанные на концепции статистического арбитража.

Термином *арбитраж* в экономике и финансах обозначается группа сделок на один и тот же актив, производимая в один момент времени на разных рынках с целью безрискового извлечения прибыли. При арбитраже происходит извлечение прибыли за счет эксплуатации рыночной неэффективности (различие цен на один и тот же товар в разных местах), которая может возникать, например, из-за асимметрии информации или различиях в рыночной ликвидности между площадками. Статистический арбитраж, хоть и входит в класс арбитражных стратегий, существенно выделяется тем, что сделки ведутся в рамках одного рынка. При этом рассматривается не один актив, а пара статистически взаимосвязанных активов (в некоторых случаях может одновременно рассматриваться большее число активов). В сообществе трейдеров статистический арбитраж на паре активов также хорошо известен под названием парная торговля или парный трейдинг.

Определение пар взаимосвязанных активов это отдельная задача, требующая решения. Часто хорошим вариантом оказывается использование привилегированных и обыкновенных акции одной компании, или пары акции и производного инструмента (например, фьючерсный контракт на данную акцию). Возможно также использование акций компаний из одного сектора экономики, однако для этого необходимо более тщательный анализ исторических данных.

Цель статистического арбитража состоит в выявлении связи между ценами активов и отслеживании моментов, когда данная связь нарушается. Связь обычно представля-

ется в виде некоторого уравнения. Нарушение связи является индикатором неэффективности рынка: либо один из активов недооценен, либо другой переоценен. Предполагается, что такие нарушения являются кратковременными, поэтому по недооцененной акции открывается длинная позиция (в то время как по переоцененной может быть открыта короткая позиция). После восстановления связи между ценами открытые позиции закрываются. На практике, однако, зависимость никогда не выполняется идеально, в следствии чего приходится дополнительно определять уровень допустимого отклонения от уравнения связи. Кроме того, сам характер связи может меняться с течением времени, что также необходимо учитывать, например, за счет определения максимальной степени отклонения цен от связи, или ограничения максимального времени удержания открытых позиций.

К основным достоинствам парной торговли можно отнести сравнительно низкие риски: в случае хорошего отбора пар вероятность сохранения взаимосвязи между ценами инструментов очень высока. Кроме того, зависимость обычно сохраняется даже в случае больших глобальных потрясений. Другим достоинством является широкие возможности по диверсификации портфеля: пары можно выбирать практически в любых секторах экономики и использовать широкий набор инструментов (начиная с акций и заканчивая опционами). Однако, имеются и недостатки. Для хорошей работы стратегий как правило требуется значительный капитал, который нужен для хорошей балансировки открытых коротких и длинных позиций. Для той же цели желательно работать с высоколиквидными активами. Помимо этого, открытие сразу двух позиций вызывает дополнительные комиссионные издержки. Стоит заметить, что все эти проблемы возникают в случае, когда стратегия предполагает открытие коротких позиций. При отказе от этого условия недостатки утрачивают свою значимость, но одновременно возрастает риск.

Есть несколько теорий, объясняющих успешность применения статистического арбитража. Первая теория предполагает наличие инертности инвесторов в случае обнаружения новой информации (например, выхода аналитического отчета) [14]. В этот момент инвесторы уделяют сначала больше внимания более популярными компаниями и компаниям с больше капитализацией, что приводит к небольшой задержке в изменениях цен на коррелирующие активы. Согласно другой теории доходность таких стратегий является премией за предоставление ликвидности в случае создания давления на рынки со стороны неинформированных инвесторов [15].

Для оценки эффективности торговых алгоритмов как правило используют такие показатели как средняя доходность стратегии (3.1) и коэффициент Шарпа (3.2) [16], которые могут быть дневными, месячным или годовыми. На практике, очевидно, опре-

делить точные значения этих величин нельзя, поэтому их оценивают как среднее значение по всем проведенным торговым экспериментам.

$$\mathbb{E}\{I\} = \mathbb{E}\left\{\frac{C_f - C_s}{C_s}\right\}, \quad (3.1)$$

где C_s - начальный капитал, C_f - капитал в конце торгового периода.

$$S = \frac{\mathbb{E}\{I\} - \mathbb{E}\{R_w\}}{\sigma(I - R_w)} \quad (3.2)$$

где R_w - безрисковая ставка, σ - среднеквадратичное отклонение.

Используя в качестве опорной точки обзорную работу Краусса [17] стратегии на основе статистического арбитража были разделены на 4 класса: метрические стратегии, коинтеграционные стратегии, стратегии на основе стохастического моделирования и стратегии на основе машинного обучения. Данное разделение удобно тем, что алгоритмы, находящиеся внутри одного подкласса, опираются на единый концептуальный и математический базис, что позволяет значительно упростить изложение.

3.1 Метрические стратегии

В данном классе, отбор пар активов происходит на основе специально заданной метрики $\rho(P_1, P_2)$, где $P_1 = \{p_{1,1}, \dots, p_{1,T}\}$ и $P_2 = \{p_{2,1}, \dots, p_{2,T}\}$ - временные ряды цен на активы (где второй индекс указывает на временные отсчеты, интервалы между которыми могут быть различными и зависят от специфики конкретного алгоритма). Самой известной в рамках данного направления является работа Гатева, Гетцмана и Рувенхорста [18], в которой используется среднеквадратичное расхождение нормированных цен активов:

$$\rho_{SSD}(\overline{P}_1, \overline{P}_2) = \frac{1}{T} \sum_{t=1}^T (\bar{p}_{1,t} - \bar{p}_{2,t})^2, \quad (3.3)$$

где $\bar{p}_{k,j} = p_{k,j}/p_{k,1}$, $k = 1, 2$.

При рассмотрении набора из N активов, расстояние ρ_{SSD} вычисляется на некотором историческом промежутке, для каждой из $N(N-1)/2$ пар, после чего отбираются 20 пар с наименьшим значением метрики. Для каждой отобранной пары по историческим данным определяется среднее значение спреда цен и его стандартное отклонение. Позиции открываются в случае, если отклонение спреда от среднего значения превышает 2 стандартных отклонения, а закрываются в случае возвращения спреда к среднему значению (либо по окончании торгового периода или в случае делистинга компании). В соответствии с первыми буквами фамилий авторов, данного подход получил название GGR.

В оригинальной работе [18] тестирование алгоритма проводилось на пуле наиболее ликвидных акций из базы данных CRSP за период с 1962 по 2002 год и показала среднегодовую доходность в 11%. Впоследствии, работоспособность данного алгоритма была проверена в работе Ду и Фаффа [19], где тот же набор активов изучался на временном интервале с 1962 по 2009 год. Была получена среднегодовая доходность в уже 7%. В статье так же показывается, что имеется достаточно большое количество пар (порядка 32%), спред между которыми не возвращается к среднему историческому значению.

Наиболее привлекательными, с точки зрения статистического арбитража, являются пары активов, которые:

1. Имеют устойчивое среднее значение спреда, что свидетельствует о наличии стабильной взаимосвязи (корреляции) между ценами активов;
2. Распределение спреда обладает большим среднеквадратичным отклонением. Это важно, так как именно в моменты отклонения спреда от среднего значения, создаются условия для открытия позиций и извлечения прибыли.

Использование ρ_{SSD} позволяет находить пары, хорошо удовлетворяющие первому усло-

вию, однако с удовлетворением второго условия возникают трудности. Действительно, из формул 3.4 и 3.5 следует, низкое значение ρ_{SSD} автоматически влечет за собой низкое среднеквадратичное отклонение.

$$\sigma^2(P_1 - P_2) = \frac{1}{T} \sum_{t=1}^T (p_{1,t} - p_{2,t})^2 - \left(\frac{1}{T} \sum_{t=1}^T (p_{1,t} - p_{2,t}) \right)^2 \quad (3.4)$$

$$\rho_{SSD}(P_1, P_2) = \sigma^2(P_1 - P_2) + \left(\frac{1}{T} \sum_{t=1}^T (p_{1,t} - p_{2,t}) \right)^2 \quad (3.5)$$

В качестве альтернативы ρ_{SSD} , в работе [20] предложено использовать корреляцию Пирсона, а также рассматривать вместо отклонений спреда специальную, эвристически сконструированную величину:

$$d_y = \beta(r_{1,t} - r_f) - (r_{2,t} - r_f) \quad (3.6)$$

где β - регрессионный коэффициент, r_f - безрисковая ставка, $r_{k,t}$ - доходность инструмента к моменту времени t , определяемая формулой $(p_{k,t} - p_{k,1} + D_{k,t})/p_{k,t}$, $D_{k,t}$ - сумма дивидендов (или купонов), выплаченная к моменту времени t .

Используя тот же набор данных, что и в [18] была получена среднемесячная доходность 1.7%, однако, стоит отметить, что как следует из формулы 3.7, корреляция Пирсона $Corr_P$ не лишена тех же проблем, что и ρ_{SSD} . Действительно, при больших значениях корреляции следует ожидать слабое отклонение разности доходностей от среднего значения.

$$\sigma^2(R_1 - R_2) = \sigma^2(R_1) + \sigma^2(R_2) - 2 \cdot \text{Corr}_P(R_1, R_2)\sigma(R_1)\sigma(R_2) \quad (3.7)$$

3.2 Коинтеграционные стратегии

В коинтеграционных подходах, по сравнению с метрическими добавляется дополнительная проверка отобранных пар активов на коинтеграцию. Группа временных рядов называется коинтегрированной, если существует их линейная комбинация, которая образует стационарный временной ряд, то есть такой временной ряд, среднее значение которого не меняется со временем (ряд не имеет тренда). Пары, признанные коинтегрированными, используются для торгов, остальные пары отбрасываются. Один из наиболее популярных коинтеграционных подходов представлен в [21]. Для определения коинтеграции в работе использовался использовался двухступенчатый тест Энгла и Грейнджер, а точки открытия и закрытия позиций определялись на основе пороговых значений спреда. Несмотря на то, что представленный метод не сопровождался эмпирической проверкой, он, тем не менее, лег в основу многих последующих решений.

Вариант данного подхода был реализован Радом в 2015 году [22]. В нем предлагается двухэтапный отбор пар: сначала отбирают пары с наименьшим значением ρ_{SSD} , а уже затем из них выбирают 20 пар с самым высоким показателем коинтеграции (Энгл-Грэнджер тест). В работе так же используется другой способ определения спреда:

$$\varepsilon_t = \gamma \cdot p_{1,t} + p_{2,t}, \quad (3.8)$$

где γ - коэффициент линейной комбинации, при котором достигается стационарность.

Если ряды коинтегрированы, то спред ε_t является стационарным рядом и поэтому периодически должен возвращаться к своему среднему значению. Эмпирическая проверка алгоритма производилась на исторических данных по американским акциям, собираемых CRSP за период с 1962 по 2014 год. Эксперименты показали среднемесячную доходность в 0.83%, которая сравнима с методом GGR, где среднемесячная доходность составила 0.88%. Худший результат вероятнее всего связан с двухступенчатым подходом к выбору пар, т.к. как первый этап смещает фокус в сторону пар с более слабыми отклонениями от среднего.

В 2006 году Лин совместно с коллегами [23] опираясь на уравнение 3.8 предложил удобную формулу, с помощью которой можно оценить выгоду открытия позиций по коинтегрированной паре:

$$\begin{aligned} \text{Profit}_{tc} = & -\frac{n}{|\gamma|}(p_{2,tc} - p_{2,to}) + \frac{n}{|\gamma|}(p_{1,to} - p_{1,tc}) = \\ & -\frac{n}{|\gamma|}((\varepsilon_{tc} - p_{1,tc}) - (\varepsilon_{to} - p_{1,to})) + \frac{n}{|\gamma|}(p_{1,to} - p_{1,tc}) = \frac{n(\varepsilon_{to} - \varepsilon_{tc})}{|\gamma|}, \end{aligned} \quad (3.9)$$

В данном случае предполагается, что цена второго актива ниже цены первого, и мы открываем по нему длинную позицию на n акций. Для балансировки также открывается короткая позиция на первый актив на $n/|\gamma|$ акций. При этом t_o и t_c - цены, соответственно, открытия и закрытия позиций. Уравнение позволяет определить пороговое значение отклонения среда. Например, такой порог, открытие позиций ниже которого точно не покроет издержки, и следовательно такая торговля будет бессмысленной.

В 2013 году Калдейра и Моура опубликовали работу [24], где коинтеграционный подход применялся к 50 наиболее ликвидным акциям бразильского фондового индекса IBovespa. Тесты Энгла–Грейнджера и Йохансена на уровне значимости 5% для данных длиной в год, перед началом периода торгов, использовались для проверки взаимосвязей между всеми возможными парами активов. 20 лучших пар отбирались для последующего четырехмесячного торгового периода. Для торговли использовалась модификация правила из алгоритма GGR, основанная на стандартном отклонении. Авторы показали, что даже при учете транзакционной комиссии в 0.5% алгоритм показывает среднегодовую доходность порядка 16.4%, а индекс Шарпа имеет значение 1.34.

3.3 Стратегии на основе стохастического моделирования

Данный класс стратегий предполагает наличие некоторой стохастической модели, которая описывает процесс изменения спреда активов. Стохастическая модель, помимо детерминированных параметров, содержит в себе случайные величины и, следовательно, не позволяет получать точные предсказания, а только распределение их вероятностей. Детерминированные параметры модели уточняются исходя из исторических данных. В ходе торгового периода сделки совершаются в случае отклонения спреда от предсказания модели.

Одной из наиболее известных работ в классе является [25]. Элиот, совместно с коллегами рассматривает в качестве модели стохастический процесс вида:

$$y_k = x_k + D \cdot \omega_k, \quad k = \overline{1, \dots, T} \quad (3.10)$$

где $Y = \{y_1, \dots, y_T\}$ - наблюдаемый процесс, отвечающий значению спреда пары активов; D - некоторая положительная константа, определяемая имперически; $\{\omega_1, \dots, \omega_T\}$ - набор н.с.в. с распределением $\mathcal{N}(0, 1)$; $X = \{x_1, \dots, x_T\}$ - скрытый стохастический процесс, поведение которого описывается уравнением 3.11:

$$x_{k+1} - x_k = (a - bx_k) \cdot \tau + \sigma \sqrt{\tau} \xi_{k+1}, \quad (3.11)$$

где a, b, σ - некоторые положительные действительные параметры, τ - квант времени, $\{\xi_1, \dots, \xi_T\}$ - набор н.с.в. с распределением $\mathcal{N}(0, 1)$;

Выражение 3.11 описывает случайный процесс обладающий свойством возврата к среднему значению, при этом $\lim_{k \rightarrow \infty} \mathbb{E}\{x_k\} \rightarrow a/b$. С практической точки зрения намного более удобной является запись вида:

$$x_{k+1} = A + Bx_k + C\xi_{k+1}, \quad (3.12)$$

$$A = a\tau, \quad B = 1 - b\tau, \quad C = \sigma\sqrt{\tau} \quad (3.13)$$

Здесь A, B и C - являются параметрами, которые, совместно с D определяются по историческим данным. В работе Элиота для этого используется ЕМ-алгоритм, основанный на методе максимального правдоподобия, который реализует итерационный процесс оптимизации параметров, где каждая итерация состоит из двух шагов: (I) по приближенным параметрам с помощью выражений 3.10, 3.12 оцениваются значения скрытой переменной $X_H = \{x_{h_1}, \dots, x_{h_l}\}$ и правдоподобие исторических значений $Y_H = \{y_{h_1}, \dots, y_{h_l}\}$; (II) значения параметров оптимизируются, чтобы максимизировать

правдоподобие Y_H при фиксированном X_H . Кроме того, в ходе самих торгов, для более точной оценки значения спреда применяется фильтр Калмана (фильтр, позволяющий рекурсивно уточнять очередное значение ряда данных используя предыдущие зашумленные значения). Позиции открываются при условии, что отклонение оценки спреда y_k от теоретического среднего значения выше некоторого порога.

Недостатком описанного подхода является то, что он применим только для пар активов с примерно одинаковыми доходностями (это в основном случаи с двойным или кросс-листингом), однако, как было показано в дальнейшем, его можно обобщить на другие пары активов [26]. Другим спорным моментом является использование Гауссовского шума, который с одной стороны, упрощает некоторые формулы, но с другой, может не соответствовать реальности.

Кумминс и Бука использовали для моделирования спреда стохастический процесс Орнштейна-Ухлебека (ОУ) с возвратом к среднему [27] вида:

$$dy_t = -a \cdot y_t dt + b \cdot dW_t, \quad (3.14)$$

где y_t - спред между активами в момент времени t , W_t - процесс Винера, a и b - параметры.

Тестирование алгоритма проводилось на рынке сырых и рафинированных нефтяных продуктов. Всего использовалось 861 отобранных в ручную пары. Торговый период составлял от 9 до 55 дней. Рассматривался большой пул торговых стратегий, лучшие 20 из которых показали среднюю дневную доходность в диапазоне 0.07-0.55%.

3.4 Стратегии на основе машинного обучения

Машинное обучение (МО) позволяет строить алгоритмы для торговли без априорных предположений о том, как изменяются цены активов. Так, в работе 2009 года Джованни, Костас и Теодорас представили торговый фреймворк, основанный на адаптивном методе наименьших квадратов [28]. Рассматривается редкий для статистического арбитража в целом, но популярных в случае использования МО подход: вместо пары активов рассматривается группа активов $\{X_1, \dots, X_n\}$, где $X_i = \{x_{i,1}, \dots, x_{i,T}\}$ которые авторы называют информационными потоками. Оценка справедливой стоимости целевого актива $Y = \{y_1, \dots, y_T\}$ в момент времени t производится по формуле $y_t = \sum_{i=1}^n x_{i,t} \beta_{t,i}$, где $B_t = \{\beta_{t,1}, \dots, \beta_{t,n}\}$ - вектор адаптивных коэффициентов, который итеративно определяется исходя из минимизации выражения 3.15:

$$\min_{B_{t+1}} \left\{ \sum_{t=1}^T (y_t - X_t B_t^T) + \frac{(1 - \delta)}{\delta} \sum_{t=1}^T \|B_{t+1} - B_t\|_2^2 \right\} \quad (3.15)$$

где $X_t = \{X_{1,t}, \dots, X_{n,t}\}$, $\delta \in (0, 1)$ - подбираемый гиперпараметр, определяющий “скорость изменения” коэффициентов.

В работе предлагается способ рекурсивного обновления коэффициентов B_t , что позволяет снизить вычислительную сложность алгоритма, также показана связь метода с фильтром Калмана, из чего следует устойчивость получаемых оценок в случае зашумленности данных. Торговая стратегия состоит в оценке стоимости активов $X_t B_t^T$ и последующей сверке их с наблюдаемой ценой y_t . По результатам сравнения принимается решение о покупке или продаже актива.

Авторы апробировали метод на торговле фьючерсными контрактами индекса S&P500, а в роли информационных потоков выступали прологорифмированные доходности по 432 акциям индекса. Использовался временной промежуток с 1997 по 2005 год. Торговля осуществлялась один раз в день. Алгоритм показал среднегодовую доходность в 13.1%, а индекс Шарпа составил 0.804.

В 2021 году Карта, совместно с коллегами, представил сложный алгоритм, основанный на ансамблевом машинном обучении [29]. Алгоритм состоит из 3 основных блоков: предсказание цен активов, ранжирование предсказаний по надежности и определение торговых сигналов на основе предсказаний. На первом этапе для предсказаний используется метод опорных векторов, случайный лес, алгоритм прогнозирования временных рядов ARIMA и градиентный бустинг. Полученные предсказания ансамблируются посредством усреднения. На втором шаге производится ранжирование предсказаний с

использованием величины MDA:

$$MDA(i, T, d) = \frac{1}{T} \sum_{t=d-1}^{d-T-1} [sign(o_t^i) = sign(y_t^i)] \quad (3.16)$$

где d - текущий момент времени, T - кол-во предыдущих дней, которые принимаются в расчет, $[.]$ - индикаторная функция (1 - если выражение истинно и 0, если ложно), i - номер актива, o_t^i - предсказание ансамбля для i -го актива в момент времени t , y_t^i - изменение цены i -го актива в день t .

Среди акций, для которых значения MDA выше порогового, каждый день отбираются 5 с наибольшим ожидаемым ростом для открытия длинных позиций и 5 с наименьшим ожиданием роста для открытия коротких позиций. Под ожиданием понимается значение вероятности, возвращаемая ансамблевым классификатором на предыдущем шаге.

Тестирование алгоритма проводилось на акциях S&P500 в период с марта 2007 года по январь 2016 года. Среднегодовой доход с учетом транзакционных издержек в 0.4% составил более 22%, при этом среднеквадратичное отклонение доходности было значительно ниже, по сравнению с алгоритмами 5-DAY и Buy&Hold.

В исследовании [30] изучается возможность использования таких методов машинного обучения как случайный лес и нейронные сети для предсказания цен активов на основе большого количества индикаторов, которые, включают историю цен на целевой актив включает, а также цены на другие рыночные инструменты (например, на нефть марки WTI, цены на акции в индексе S&P100 или S&P300), информацию о торговом дне недели, месяце, индексы VIX, ICB классификатор и некоторые другие показатели. Всего авторы насчитывают порядка 592 величины по которым строится прогноз. Прогноз осуществляется с помощью классификатора, который определяет вероятность того, что актив будет расти быстрее рынка в среднем. Каждый день открывается 10 длинных позиций для акций с наилучшими прогнозами и 10 коротких позиций для акций с наихудшими прогнозами. В качестве классификаторов использовались алгоритмы случайного леса, нейронная сеть DBN и линейная регрессия с регуляризацией Elastic Net. В качестве тренировочных и торговых использовались данные с 1990 по 2015 год по акциям, которые за этот промежуток времени хотя бы раз появлялись в SP 900. Имеющиеся данные были разделены на 45 пересекающихся периодов по 2.5 года. В рамках каждого периода первые два года использовались для обучения, а оставшиеся полгода для торговли. Среди многочисленных экспериментов авторам удалось при некоторых условиях добиться среднегодовой доходности более чем в 35%. Однако, стоит заметить, что разброс в зависимости от используемых данных весьма значителен и

при добавлении новой информации некоторые алгоритмы показывали отрицательную среднюю доходность.

3.5 Статистический арбитраж на Российских рынках

К сожалению, работ, посвященных российскому рынку чрезвычайно мало, причем их эмпирическая составляющая весьма скромная. Так, в работе Володина и Коченкова [31] изучается возможность применения парной торговли на российском рынке на примере фьючерсных контрактов на обыкновенные и привилегированные акции ОАО «Сбербанк России». Для торговли используется простое пороговое правило по отношению к спреду. Значение порога является гиперпараметром и подбирается на тестовом участке исторических данных. Балансировка коротких и длинных позиций происходит по заранее определенному жесткому правилу. Авторами было проведено два эксперимента длительностью порядка трех месяцев: один охватывает временной интервал с декабря 2012 по марта 2013 года, а другой с марта 2013 года по июнь 2013 год. Эксперименты показали годовую доходность 691% и 1546% годовых соответственно. Представленные результаты, скорее всего, являются чрезмерно оптимистичными. Торговые периоды слишком короткие, а все рассматриваемые активы относятся к одной компании, что не позволяет надежно оценить среднюю эффективность предлагаемого метода.

Схожая работа была опубликована Липатниковым и др. в 2017 году [32]. В статье также рассматривается торговля фьючерсными контрактами, ввиду сравнительно более низких комиссий на при работе с данным инструментом. Рассматриваются две пары контрактов:

1. контракты на акции компаний «Северсталь» и «Норильский никель»;
2. контракты на акции компании «Транснефть» и «Татнефть».

Возможность торговли представленными парами изначально оценивается с помощью коинтеграционного подхода. Сначала рассчитывается уравнение линейной регрессии между логарифмированными ценами контрактов (3.17).

$$\ln(p_{1,t}) = \mu + \ln(p_2) + \xi_t \quad (3.17)$$

где μ - коэффициент регрессии.

Затем ряд, состоящий из невязок $\xi = \{\xi_1, \dots, \xi_T\}$, проверяется на стационарность. Торговля происходит на основе симметричного порогового правила по отношению к

индикатору на основе разнице в относительных изменениях цен (3.18).

$$Tind = \frac{p_{1,t}}{p_{1,t-1}} - \frac{p_{2,t}}{p_{2,t-1}} \quad (3.18)$$

Тестирование проводилось по данным в период с января 2014 года по март 2017 года. Квант времени равнялся 1 часу. Доходность составила 96.6% для первой пары и 107.0% для второй пары. В работе также отмечается устойчивый рост капитала даже в условиях геополитических потрясений 2014-2015 годов, что свидетельствует о возможности данного типа методов противостоять рискам, вызванным политической повесткой. Недостатком данной работой также является малое количество рассматриваемых пар.

3.6 Резюме

В научной литературе имеется большое разнообразие методов парной торговли. Каждый из подходов имеет достоинства и недостатки, нельзя однозначно выявить наиболее эффективный метод. Сложность сравнение усугубляется разным набором используемых активов, исторических периодов и выбираемой скоростью торгов. Тем не менее, как показывает опыт исследований Российского рынка: даже сравнительно простые подходы, основанные на пороговых правилах по отношению к спреду могут приносить большую доходность на некоторых парах инструментов. Исходя из этого, кажется, по крайней мере в ближайшей перспективе, нецелесообразным использование сложных в технической реализации методов, основанных на ансамблевом машинном обучении и нейронных сетях. Они требуют большого количества данных для обучения, долгой и тонкой настройки гиперпараметров [30], а принимаемые ими решения сложно интерпретировать. Поэтому для разработки метода торговли было решено использовать хорошо интерпретируемую концепцию линейной регрессии с динамически обновляемыми коэффициентами, которые в тоже время, позволят алгоритму быть достаточно гибким.

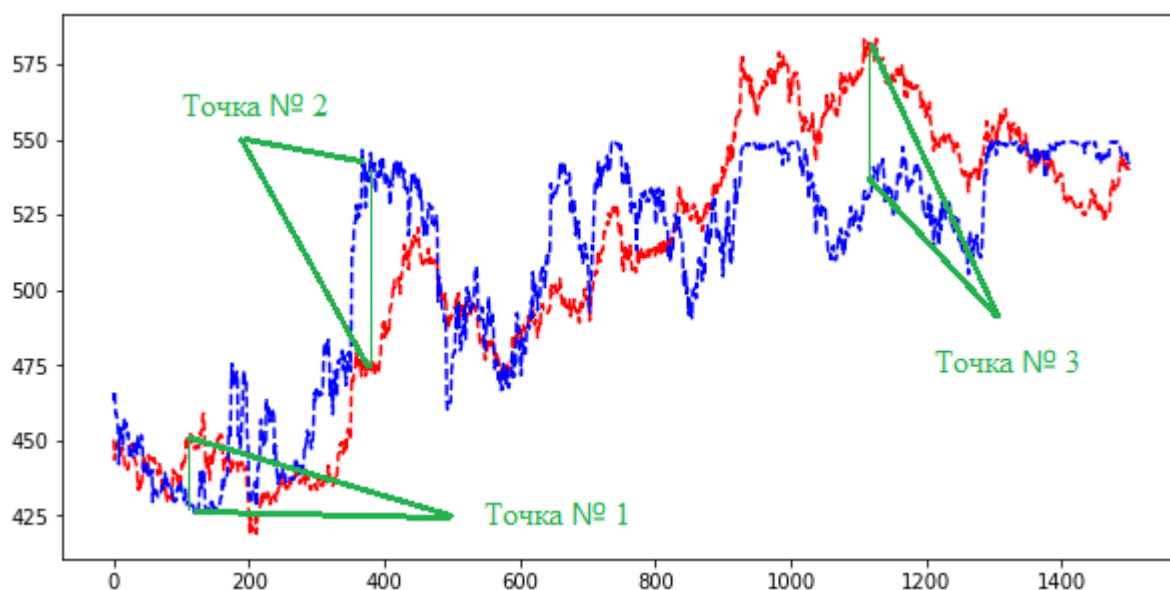
4 Технологическое описание алгоритма

4.1 Логическое описание алгоритма

Данная работа посвящена рассмотрению статистического арбитража на паре акций с использованием коинтеграционного подхода. Для работы будут отбираться активы, линейная комбинация которых с некоторыми поправками является постоянной величиной, то есть активы имеющие общий тренд.

В качестве активов для тестирования используются различные акции, торгующиеся на Московской и Петербургской бирже. Так как акции являются самым доступным инструментом.

Для оценки доходности метрических и коинтеграционных стратегий авторы использовали подход, когда при наличии спреда по одному активу открывается длинная позиция, по другому коротка, закрытие сделки происходит в случае возвращения спреда к среднему значению. И доходность таких подходов составляет от 7 до 16% в зависимости от работы. При этом в периоды, когда не найдена возможность для сделки, капитал простаивает. Поэтому в данной работе будет рассмотрен другой подход: без открытия коротких позиций. Например, два актива:



Алгоритм действует следующим образом: в Точке №1 открывается длинная позиция по синему активу, в Точке №2 закрывается позиция по синему активу и открывается длинная позиция по красному, в Точке №3 соответственно закрывается по красному и открывается по синему. Таким образом, к доходности, которая получается в результате арбитража, прибавляется доходность самих активов. А также такая логика позволяет избежать дополнительных трат на комиссию от открытия коротких позиций. Но сто-

ит отметить негативный момент, в случае падения рынка, доходность алгоритма тоже будет падать. Это дополнительный риск.

Вторым отличием данной работы от рассмотренных выше алгоритмов является оценка порогового значения спреда. Которое будет динамически рассчитываться на каждом шаге, при этом без использования технически сложных нейронных сетей и машинного обучения.

А также тестирование алгоритма будет проводиться на активах российского рынка.

4.2 Корреляция

В статистике корреляция — это любая статистическая связь между двумя случайными величинами или двумерными данными. Хотя в самом широком смысле «корреляция» может обозначать любой тип ассоциации, в статистике она обычно относится к показателю линейной связи пары активов.

Корреляции полезны, потому что они могут указывать на взаимосвязь, которую можно использовать для прогнозирования. Корреляция — это мера того, как две или более переменных связаны друг с другом. Существует несколько коэффициентов корреляции, часто обозначаемых ρ или же r . Наиболее распространенным из них является коэффициент корреляции Пирсона, который чувствителен только к линейной зависимости между двумя переменными (которая может присутствовать, даже когда одна переменная является нелинейной функцией другой). Другие коэффициенты корреляции, такие как ранговая корреляция Спирмена, были разработаны, чтобы быть более чувствительными к нелинейным отношениям.

В данной работе отбираются активы с линейной зависимостью, поэтому используется корреляция Пирсона.

Коэффициент корреляции Пирсона — это ковариация двух переменных, деленная на произведение их стандартных отклонений.

$$\rho_{x,y} = \frac{\text{cov}(x, y)}{(\sigma_x \sigma_y)}$$

или

$$\rho_{x,y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Корреляция принимает значения от +1 до -1. Абсолютное значение, равное ровно 1, означает, что линейное уравнение точно описывает взаимосвязь между X и Y . При этом +1 означает, что независимые величины имеют общий тренд, то есть при увеличении одной вторая тоже увеличивается. А -1 свидетельствует об обратных трендах, при увеличении одной вторая величина уменьшается. Корреляция равна 0, когда величины не связаны между собой.

4.3 Метод наименьших квадратов

В данной работе используется метод наименьших квадратов (МНК), поэтому сразу рассмотрим его теоретическое описание.

МНК является одним из классических методов регрессионного анализа и используется для аппроксимации решений переопределенных систем линейных уравнений (в которых неизвестных меньше чем самих уравнений), путем минимизации суммы квадратов остатков (под остатком понимается разность между величиной наблюдения и предсказаниями модели).

В данной работе МНК используется для нахождения полиномиальной аппроксимирующей функции. Предполагается, что есть два ряда значений: $x = \{x_1, x_2, \dots, x_n\}$, $y = \{y_1, y_2, \dots, y_n\}$. Мы хотим найти такую функцию $f_m(x_k) = \sum_{j=0}^m a_j x_k^j$, чтобы минимизировать разницу:

$$\sum_{i=0}^m |f_m(x_i) - y_i|^2 \quad (4.19)$$

Дальнейшие выкладки удобнее производить в векторном виде. Тогда уравнение 4.19 приобретает вид:

$$\min_a \{\|Xa - y\|_2\} = \min_a \{(Xa - y)^T (Xa - y)\} \quad (4.20)$$

где $\bar{a} = [a_1, a_2, \dots, a_m]^T$ - вектор-столбец подбираемых параметров, X - транспонированная матрица Вандермонда для ряда x :

$$X = \begin{bmatrix} x_1^0 & x_1^1 & \dots & x_1^m \\ x_2^0 & x_2^1 & \dots & x_2^m \\ \vdots & \vdots & \ddots & \vdots \\ x_n^0 & x_n^1 & \dots & x_n^m \end{bmatrix}$$

В силу выпуклости функции 4.20, решение можно найти аналитически, с помощью необходимого условия минимума:

$$\begin{aligned} \text{grad}_a((Xa - y)^T (Xa - y)) &= \text{grad}(a^T X^T Xa - a^T X^T y - y^T Xa - y^T y) = \\ 2X^T Xa - 2X^T y &= 0 \Leftrightarrow X^T Xa = X^T y \end{aligned}$$

Матрица X обычно не является квадратной, и, следовательно, для нее не существует обратной матрицы. Однако, если все точки в x различны, а $n \geq m$ (что обычно выполняется на практике), то матрица X является полноранговой, и следовательно,

существует $(X^T X)^{-1}$. При этом решение задачи может быть записано в виде:

$$\hat{a} = (X^T X)^{-1} X^T y \quad (4.21)$$

Ценным свойством метода является несмещенность и оптимальной получаемых оценок в случае шума. Если истинное связь между x и y может быть описана уравнением $y = Xa + \varepsilon$, где $\varepsilon = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ - вектор н.с.в. с нулевым математическим ожиданием и дисперсией σ^2 , то:

$$\begin{aligned} \mathbb{E}\{\hat{a}\} &= \mathbb{E}\{(X^T X)^{-1} X^T y\} = \mathbb{E}\{(X^T X)^{-1} X^T (Xa + \varepsilon)\} = \\ &= \mathbb{E}\{(X^T X)^{-1} (X^T X)a + (X^T X)^{-1} X^T \varepsilon\} = \mathbb{E}\{a\} + \mathbb{E}\{(X^T X)^{-1} X^T \varepsilon\} = \mathbb{E}\{a\} = a \end{aligned} \quad (4.22)$$

где последнее равенство верно в силу нулевого математического ожидания ε и линейности преобразования $(X^T X)^{-1} X^T$.

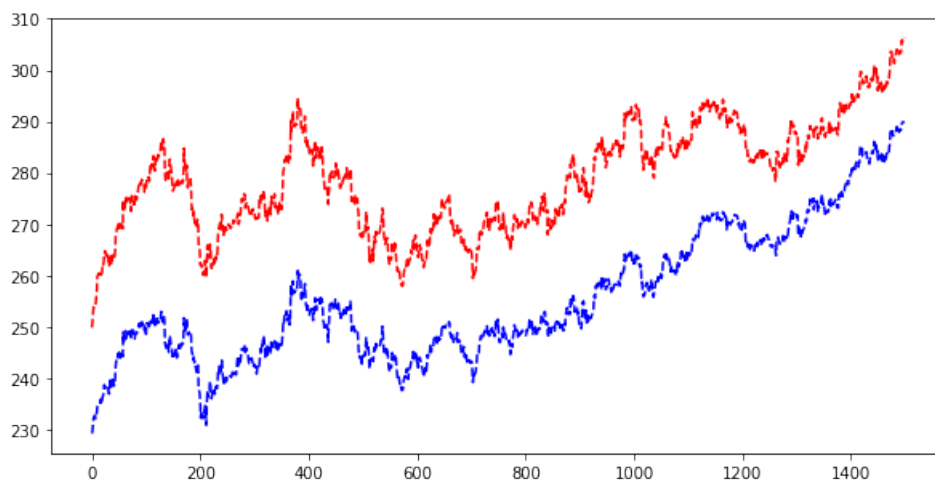
Уравнение 4.22 показывает несмещенность оценки. Для дисперсии имеем:

$$\begin{aligned} \mathbb{D}\{\hat{a}\} &= \mathbb{E}\{(\hat{a} - \mathbb{E}\{a\})(\hat{a} - \mathbb{E}\{a\})^T\} = \mathbb{E}\{(\hat{a} - a)(\hat{a} - a)^T\} = \\ &= \mathbb{E}\{(a - (X^T X)^{-1} X^T \varepsilon + a)(a - (X^T X)^{-1} X^T \varepsilon + a)^T\} = \mathbb{E}\{(X^T X)^{-1} X^T \varepsilon \varepsilon^T X (X^T X)^{-T}\} \\ &= (\text{в силу предположений относительно } \varepsilon \text{ имеем, что } \varepsilon \varepsilon^T = \sigma^2 I) = \\ &= \sigma^2 \mathbb{E}\{(X^T X)^{-1} X^T I X (X^T X)^{-1}\} = \sigma^2 \mathbb{E}\{(X^T X)^{-1}\} = \sigma^2 (X^T X)^{-1} \end{aligned} \quad (4.23)$$

В силу теоремы Гаусса-Маркова оценка, представленная в 4.21 является оптимальной, т.е. дисперсия, полученная в 4.23 является наименьшей среди всех возможных несмещенных оценок [33].

4.4 Нормировка

В основном активы имеют разный номинал. Например, котировки привилегированной и обычной акции Сбербанка выглядят следующим образом:

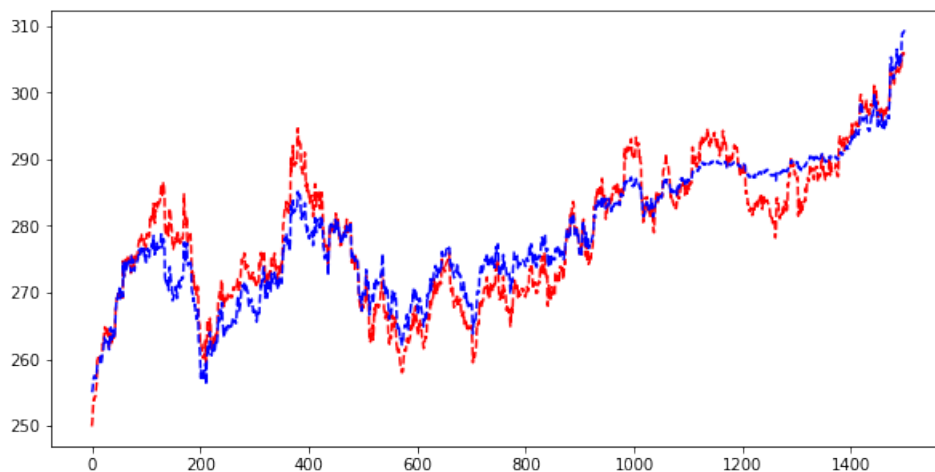


Перед оценкой спреда нужно привести котировки к общему виду. В данной работе отбираются активы с линейной зависимостью, поэтому актив y может быть выражен через актив x как:

$$y = ax + b$$

где a и b коэффициенты линейной функции, которые можно найти, используя рассмотренный выше МНК 1 степени.

Результат нормировки на примере привилегированной и обычной акции Сбера:



Таким образом, для нормировки используется МНК первой степени.

4.5 Поиск оптимальной точки

Колебания спреда не поддаются точному описанию, в зависимости от ситуации они разного размера. Для максимизации прибыли сделку нужно совершать в момент максимального расхождения. Но на практике неизвестно в какую сторону пойдут акции и предсказать этот момент крайне сложно. В связи с этим встает задача поиска оптимальной точки для совершения сделки, зная только статистическое распределение котировок до настоящего момента.

В качестве метода аппроксимации был использован МНК 5 степени.

При решении задачи поиска оптимальной точки входа за x принимается множество значений спреда, за y ряд натуральных чисел. Используется функция нормировки 5 порядка. Оценка спреда происходит на данных за 10 дней с шагом 1 час, за данный промежуток обычно наблюдается не более 5 перегибов, данное значение было получено на практике. Поэтому пятого порядка достаточно.

Пример аппроксимации спреда:



Оптимальной точкой для совершения сделки является перегиб аппроксимирующей функции.

4.6 Данные для исследования и параметры алгоритма

В данной работе используются различные датасеты Московской и Санкт-Петербургской биржи. Такие как:

1. Топ МосБиржи (50шт.)
2. Все акции Мосбиржи (628 шт.)
3. Все акции Санкт-Петербургской биржи (2800 шт.)

Для выгрузки данных был создан специальный скрипт, который в автоматическом режиме генерирует ссылки для скачивания котировок акций с сайта Finam [34]. Скрипт принимает на вход следующие параметры:

Рынок	МосБиржа акции	конец_диапазона_(день)	01
интервал_измерения	1 час	конец_диапазона_(месяц)	05
формат_файла	.csv	конец_диапазона_(год)	2022
формат_даты	ггггммдд	начало_диапазона_(день)	01
формат_времени	ччммсс	начало_диапазона_(месяц)	01
московское_время	да	начало_диапазона_(год)	2020
Разделитель_полей	запятая (,)		
Разделитель_разрядов	нет		
Формат_записи_в_файл	TICKER, PER, DATE, TIME, OPEN, HIGH, LOW		
Добавить_заголовок_файла	да		
Заполнять_периоды_без_сделок	да		

куда положить результат? /content/drive/MyDrive/

Скачать

Загружено 7/648

Итоговые данные выгружаются в csv таблицы, со следующими атрибутами:

<TICKER>	<PER>	<DATE>	<TIME>	<OPEN>	<HIGH>	<LOW>	<CLOSE>	<VOL>
GAZP	60	20200103	110000	259.1	260.27	258.52	259.36	3897800
GAZP	60	20200103	120000	259.36	261.86	259.04	261.75	8670090
GAZP	60	20200103	130000	261.8	261.9	260.91	261.5	3347900
GAZP	60	20200103	140000	261.54	261.82	260.94	261.18	2201500
GAZP	60	20200103	150000	261.2	261.57	259.43	259.69	3036390

где, TICKER - название компании, PER - интервал измерения в секундах, DATE - дата, TIME - время, OPEN - цена открытия, HIGH - максимальная цена за интервал, LOW - минимальная цена за интервал, CLOSE - цена закрытия, VOL - объем торгов за интервал.

Параметры данных, используемых для тестирования алгоритма:

1. Все данные выгружаются с шагом в 1 час. В данной работе рассматривается именно такой шаг, так как для использования более маленького шага повышаются риски ликвидности, а использование более длинного промежутка приведет к сокращению количества сделок за год. Кроме этого, такой промежуток позволяет использовать алгоритм для полуавтоматического робота, так как период между сделками достаточно большой и можно их совершать в ручном режиме.

2. В данной работе в основном используется временной промежуток с 01.10.2020 по 01.11.2021. На данных 2022 году тестирование проводилось отдельно, так как из-за нестабильной политической обстановки на рынке наблюдается резкое падение котировок всех акций. Такие ситуации наблюдаются редко и непригодны для оценки среднестатистической работы алгоритма.

Параметры алгоритма:

1. Корреляция. В данной работе отбираются активы с корреляцией от 0,9 до 1. Данные значения были получены практическим путем.
2. Временную промежуток для отбора активов и нормировки. Выбран период для расчета корреляции и нормировки размером в один месяц. Из данных, которые программа получает на вход, первый месяц считается известным и используется для предварительных действий. На практике было установлено, что данный период является достаточным, его увеличение слабо влияет на результат но увеличивает вычислительную сложность алгоритма.
3. Временной промежуток для аппроксимации спреда. Для данной задачи используется 10 дней. Спред быстро меняется, если использовать месяц, как в случае с нормировкой, аппроксимирующая функция получается либо сглаженной и слабо отражает реальную картину, либо, если увеличить степень МНК до 7 и выше, начинает накапливаться вычислительная ошибка, которая искажает результат.
4. Используемый метод. Для сравнения в данной работе проведено тестирование алгоритма на основе среднего значения. То есть сделка совершается, когда спред превышает своё среднее значение за последний месяц.
5. Начальный капитал. Все тесты проведены для значения первоначального капитала равного 100 000 руб.

Стоит отметить, что нормировка активов и аппроксимация спреда на каждом шаге рассчитываются заново.

4.7 Оценка возможных потерь

Так как в данной работе рассматривается арбитражная стратегия, когда один из пары коррелирующих активов находится в портфеле, будем рассматривать риски для этого вида арбитража.

Ликвидность. В данном алгоритме при недостаточной ликвидности не реализуется возможность арбитража, но отрицательного эффекта не будет. Например, в случае использования коротких позиций или фьючерсов недостаточная ликвидность приводит к дополнительным транзакционным издержкам, при торговле акциями сделка просто не совершится.

Риск падения акций. Так как при рассматриваемом алгоритме в портфеле находится одна из акций, алгоритм чувствителен к падению рынка и требует ручного вмешательства в случае кризисных ситуаций. В стандартных ситуациях объемы рынка растут в долгосрочной перспективе. От локальных падений в алгоритм добавлено условие, оно заключается в том, что сделки не совершаются в минус, то есть акции не продаются дешевле, чем были куплены.

Транзакционные издержки. Все брокеры берут комиссии за сделки, в данной работе учитывается комиссия в размере - 0,05 процента от сделки. В алгоритм вставлено условие на совершение сделки только в случае, если спред превышает транзакционные издержки. Кроме этого, есть риск отмены сделки по какой-либо причине. В таком случае, как и с риском ликвидности, не реализуется возможность арбитража, но негативного эффекта не будет.

Модельный риск. Возникает в результате неточностей в модели предлагаемого алгоритма. Можно выделить два случая:

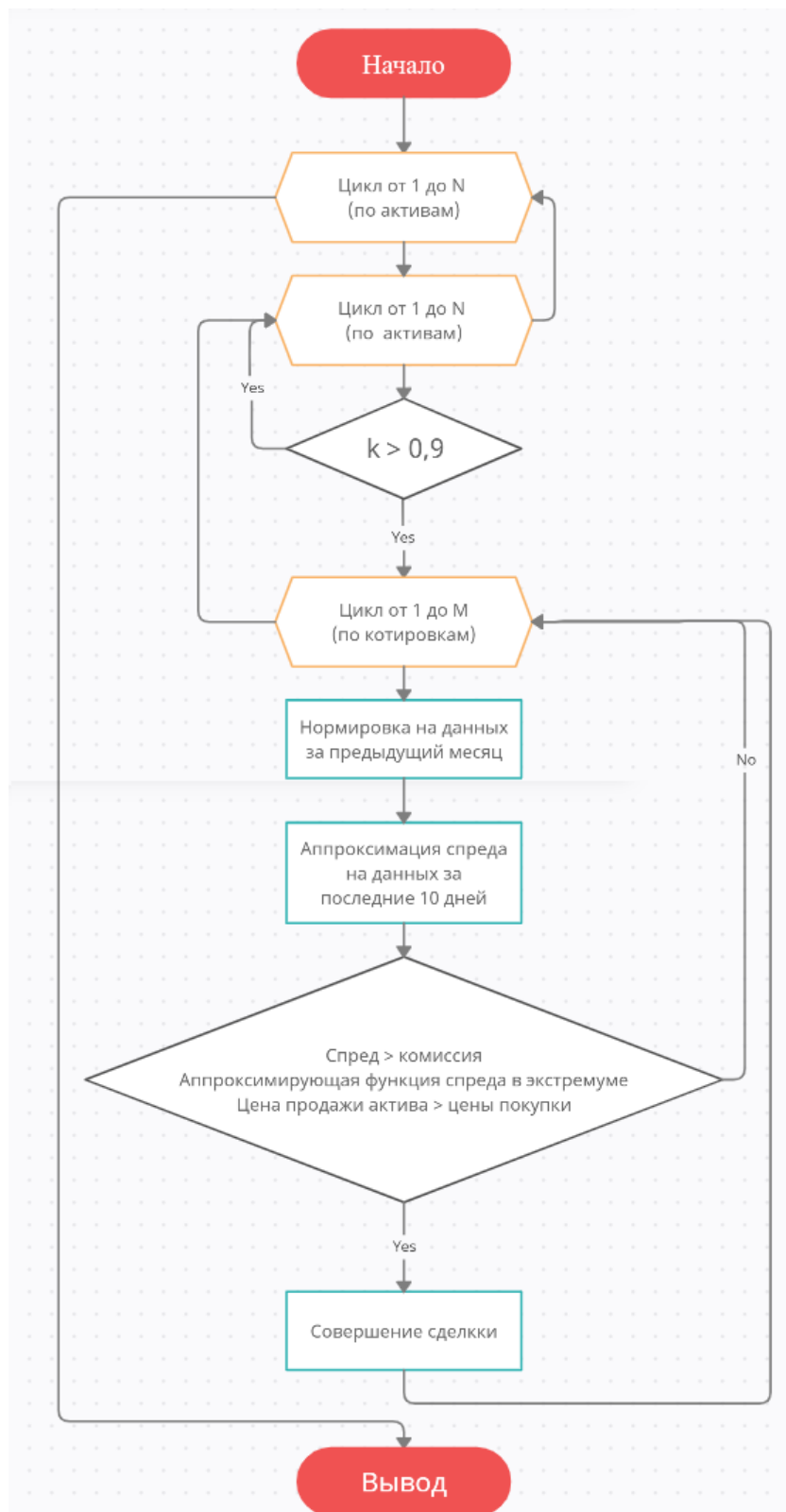
1. Неверно найдена точка входа. Аппроксимация сглаживает колебания и в случае смены тренда спреда, аппроксимирующая функция среагирует не сразу, а на 2-3 шаге. Из-за этого есть риск того, что сделка будет совершена не оптимально. Но при этом в программе заложен фильтр на то, чтобы размер спреда превышал транзакционные издержки и цена продажи не была ниже цены покупки в прошлый раз, то есть сделки не совершаются в минус. Поэтому при не оптимальном выборе точки входа, прибыль получается не максимальной, но отрицательного эффекта нет.
2. Некорректно отобраны активы. Алгоритм отбирает активы с общим трендом однократно на данных за месяц, если после этого тренды активов изменятся, есть риск получить отрицательную прибыль. Например, в течении месяца два актива А и В показывали схожую растущую динамику, в результате были отобраны в порт-

фель и была открыта длинная позиция по активу А. С этого момента котировки компании А сменили тренд на ниспадающий и по прошествии года её котировки упали на 30%, а котировки компании В выросли на 20%. Алгоритм не нашел возможности для повторной сделки, т.к. компания А на всем промежутке была дешевле В, как результат доходность -30%. Такая ситуация возможна, но в целом фондовый рынок по статике растет (по оценке BCS EXPRESS среднегодовая доходность российского рынка с 2000 по 2020 составляет 17%), при диверсификации убыток покроется прибылью остальных пар.

Операционный и системные риски. Успешность работы алгоритма зависит от стабильной работы инфраструктуры: брокера, биржи и остальных участников. Сбои системы такие как глобальный финансовый кризис, заморозки активов, потери информации о транзакциях... принесут убытки

4.8 Детали работы программы

Учитывая все вышеперечисленное, в целом алгоритм выглядит следующим образом:



4.9 Результат работы алгоритма

На графике представлен пример работы алгоритма для одной пары. Зеленым отмечены места сделок. На рисунке видно, что сделки совершаются достаточно оптимально в нужной последовательности.



В численном выражении результаты работы представлены ниже.

Для промежутка 01.10.2020 - 01.11.2021 с выбором порогового спреда на основе МНК:

Датасет	СД,%	СДД,%	ИШ	Мед,%	min,%	max,%	<0,%	<10%	N
Топ МосБиржи (50)	38,07	20,03	0,86	23,73	-0,16	230,08	5	47	142
Акции МосБиржи (628)	28,6	21,2	0,69	4,8	- 2,03	1287	664	2377	6040
Акции СПБ (1988)	20,01	15,07	0,63	-0,09	-5,06	1710	150776	360500	850135

На промежутке 01.10.2020 - 01.11.2021 с выбором порогового спреда на основе среднего значения:

Датасет	СД,%	СДД,%	ИШ	Мед,%	min,%	max,%	<0,%	<10%	N
Топ МосБиржи (50)	36,01	19,97	0,58	16,47	-5,1	290	22	56	142
Акции МосБиржи (628)	24,7	17,3	0,38	2,66	-7,23	1070	2721	3479	5897
Акции СПБ (1988)	18,05	13,7	0,27	-9,19	0,1	1311	434987	482341	834920

где СД - средняя доходность арбитража, СДД - средняя добавочная доходность от арбитража, ИШ - индекс Шарпа, Мед. - медианное значение, min - минимальное

значение, \max - максимальное значение, $<0, \%$ - количество акций, у которых доходность $<0, \%$, $<10, \%$ - количество акций, у которых доходность $<10, \%$, N - количество отобранных пар.

Для сравнения работы алгоритма с использованием МНК проводился эксперимент, где пороговое значение спреда вычислялось как среднее значение. По результатам видно, что алгоритм с МНК эффективнее примерно на 10%, если смотреть по доходности. Кроме этого, индекс Шарпа алгоритма на основе среднего гораздо ниже, что говорит о более сильной волатильности. Далее все численные значения будут приведены для МНК.

Нормальным значением индекса Шарпа для инвестиционного портфеля является 1, тогда волатильность компенсируется доходностью. Но индекс Шарпа не отличает колебания вверх и вниз. В данной работе при использовании МНК значение коэффициента колеблется от 0,63 до 0,86. Эти значения ниже нормы, при этом процент пар, доходность которых выше 10% составляет 66% для первого датасета (Топ МосБиржи), 60% для всех акций МосБиржи, 57% всех акций СПБ. Процент пар с доходностью ниже 0 составляет соответственно 3%, 10%, 17%. И минимальное значение доходности не ниже -5,06%. Это значит, что есть достаточно высокий риск получить значение доходности ниже среднего, но риск потери капитала низкий, так как доля активов с отрицательной доходностью невелика.

В рассмотренной выше работе Калдейра и Моура индекс Шарпа для 50 наиболее ликвидных акций бразильского индекса получился равным 1,34. В данной работе для 50 ликвидных акций России значение составило 0,86. Стоит также отметить, что для расчета индекса Шарпа в качестве безрисковой ставки использовалась доходность ОФЗ за 2021 год, равная 6,4%. В зарубежных исследованиях для этих целей обычно используют ставку по долговым обязательствам США, которая составляет порядка 1,5%. Среднегодовая доходность предложенного алгоритма более чем на 20% выше, чем в работе Калдейра и Моура. Все вышесказанное доказывает предположение о том, что отказ от использования коротких позиций увеличивает доходность алгоритма за счет доходности самих акций, но в следствии чувствительности к падению котировок данная стратегия показывает более высокую волатильность и риск не получить среднее значение доходности.

Готовые продукты, представленные на рынке, предлагают доходность до 100% годовых. Не очень понятно, как интерпретировать данный результат, так как это не среднее значение. В рассматриваемой работе обнаружены пары доходность, которых составляет более 1000%, что гораздо больше. Но такой результат нельзя считать достоверным, так как это выбросы. Среднее значение доходность анализируемого алгоритма

составляет 20-35% годовых.

Теперь проанализируем работу алгоритма в кризисный период. Для промежутка 01.11.2021 - 01.06.2022 с выбором порогового спреда на основе МНК:

Датасет	СД,%	СДД,%	ИШ	Мед,%	min,%	max,%	<0,%	<10%	N
Топ МосБиржи (50)	3,77	39,17	-0,06	1,84	-0,1	98,25	24	85	103
Акции МосБиржи (628)	2,87	38,67	-0,31	-0,09	-0,18	124,0	2356	2620	2894
Акции СПБ (1988)	-3,08	35,81	-24,5	-0,1	-0,1	34,3	300327	301876	262347

На промежутке 01.11.2021 - 01.06.2022 с выбором порогового спреда на основе среднего значения:

Датасет	СД,%	СДД,%	ИШ	Мед,%	min,%	max,%	<0,%	<10%	N
Топ МосБиржи (50)	0,9	36,33	-1,15	1,09	-5,1	30,5	23	91	103
Акции МосБиржи (628)	-1,6	34,2	-0,19	0,09	-7,11	192,0	2250	2572	2894
Акции СПБ (1988)	-7,04	35,77	-13,89	-0,09	-15,1	39,7	259052	262141	262347

За данный промежуток времени российский рынок упал на 37%. Доходность алгоритма для наиболее ликвидных акций остается положительной. Это доказывает низкий риск потери капитала.

Индекс Шарпа отрицательный, потому что в качестве безрисковой ставки по-прежнему используются ОФЗ. Хотя в такой кризисный период, ОФЗ считать безрисковым активом некорректно, так как рейтинг страны понизили до преддефолтного. Но за неимением альтернатив используются российские облигации.

4.10 Резюме

В рамках данной работы был реализован алгоритм на основе статистического арбитража на парах акций. Для отбора линейно зависимых активов использует коэффициент корреляции Пирсона.

Отличительной особенностью данной работы являются, во-первый, использование только длинных позиций, за счет чего в портфеле всегда находится один из активов. Преимущество такого подхода заключается в увеличении доходности алгоритма за счет роста самих активов. Недостатком является чувствительность к падению рынка. Второй особенностью данной работы является тестирование алгоритма в Российских реалиях. Для работы использовались все акции торгующие на Московской и Санкт-Петербургской бирже, а также для расчета индекса Шарпа в качестве безрисковой доходности принималась ставка ОФЗ. Третьим отличием данного алгоритма от уже описанных является нормировка и поиск оптимальной точки входа с использованием метода наименьших квадратов.

К преимуществам алгоритма относится доходность в размере 20-35%. Количество пар, доходность которых превышает 10% составляет порядка 60%. Минимальные значения доходности не ниже -6% при стабильном рынке. Максимальные значения доходности составляют сотни процентов. Алгоритм хорошо себя показал при тестировании на кризисном периоде. Положительная доходность при общем падении рынка более чем на 35% - свидетельствует стабильной работе алгоритма и низком риске потери капитала.

Лучшее значение индекса Шарпа составляет 0,86, что говорит о повышенной волатильности и риске не получить значения средней доходности. Это относится к негативным сторонам алгоритма.

5 Вывод

В рамках данной работы был проведен анализ рынка автоматизированных торговых систем. В результате него было выяснено, что алгоритмическая торговля широко распространилась и занимает значительную долю в объеме торгов на биржах по всему миру. Крупные финансовые организации используют высокочастотные алгоритмы и прямой доступ к биржам, что позволяет проводить большое количество операций за короткий срок и извлекать из них прибыль. Для использования таких технологий необходимо дорогостоящее оборудование и крупные первоначальные вложения. Для частного инвестора с небольшим капиталом на рынке представлены готовые роботы. Производители заявляют о доходности таких решений до 100%, но не предоставляют гарантии. Кроме того, открытые роботы как правило используют стратегии импульсной торговли и анализа тренда. Такие роботы в различных нестандартных ситуациях могут достаточно быстро “проиграть” капитал за счет коротких позиций.

Задачей данной работы являлось создание алгоритма, который приносил бы прибыль без использования высокочастотной торговли, а так же обладал бы большей надежностью, чем импульсные и трендовые решения.

Арбитражные стратегии считаются менее рискованными в отличие от остальных. Поэтому в данной работе реализован алгоритм статистического арбитража на паре акций. Был проведен анализ научных работ по данной теме. Разработанный алгоритм относится к классу коинтеграционных стратегий со следующими отличиями:

1. Используются только длинные позиции
2. Для определения порогового спреда используются аппроксимирующая функция, найденная методом наименьших квадратов
3. Тестирование алгоритма проводится на российском рынке акций

Использование только длинных позиций увеличивает доходность, так как к прибыли получаемой от арбитража прибавляется доходность самих активов. В то же время алгоритм становится чувствительным к падению рынка и повышается волатильность. Это было показано на практике. По сравнению с коинтеграционным алгоритмом Калдейра и Моура, рассматриваемый алгоритм имеет доходность на 20% больше, при этом индекс Шарпа значительно ниже (на 0,48). Но количество пар, доходность которых ниже 10%, для датасета ликвидных акций составляет менее 10%, что говорит о низком риске потери капитала.

В результате, по проведенным тестам можно выделить следующие конкурентные преимущества алгоритма:

1. Доходность порядка 20-35% годовых;
2. Низкий риск потери капитала;
3. Низкий порог входа, то есть не обязательно иметь большой капитал.

И недостатки:

1. Индекс Шарпа меньше 1, значит есть риск получить доходность ниже средней;
2. Ограниченный набор активов, так как отбираются только пары с коррелирующими ценами.

Таким образом, алгоритм обладает экономической эффективностью и может использоваться в качестве основы для торгового робота.

Список источников

- [1] Algorithmic Trading. — 2022. — Access mode: <https://www.pangea-si.com/algo-trading-overview/> (online; accessed: 2022-06-12).
- [2] Статистика объемов торгов Московская Биржа. — 2022. — Режим доступа: <https://www.moex.com/ru/ir/interactive-analysis.aspx> (дата обращения: 2022-06-12).
- [3] 63 Moons Technologies Limited. — 2022. — Access mode: <https://brokeragetechnologiesolutions.63moons.com/products/trader-workstation-pt> (online; accessed: 2022-06-12).
- [4] Virtu Financial. — 2022. — Access mode: <https://www.virtu.com/solutions/> (online; accessed: 2022-06-12).
- [5] MetaTrader Market. — 2022. — Режим доступа: <https://www.metaquotes.net/ru/services/market> (дата обращения: 2022-06-12).
- [6] Algo Trader AG. — 2022. — Access mode: <https://www.algotrader.com/> (online; accessed: 2022-06-12).
- [7] Learn2Trade. — 2022. — Access mode: <https://learn2.trade/forex-signals> (online; accessed: 2022-06-12).
- [8] Pionex. — 2022. — Access mode: <https://www.pionex.com/offers/#/w/grid/SybH8BhsSX> (online; accessed: 2022-06-12).
- [9] 1000 pip Climber System. — 2022. — Access mode: <https://www.1000pipclimbersystem.com/> (online; accessed: 2022-06-12).
- [10] Forex Fury. — 2022. — Access mode: <https://www.forexfury.com/> (online; accessed: 2022-06-12).
- [11] Algo Signals. — 2022. — Access mode: <https://algo-signals.com/ru/> (online; accessed: 2022-06-12).
- [12] BinBot Pro. — 2022. — Access mode: <https://www.binbotpro.com/> (online; accessed: 2022-06-12).
- [13] EA Builder. — 2022. — Access mode: <https://www.eabuilder.com/> (online; accessed: 2022-06-12).

- [14] Papadakis George, Wysocki Peter. Pairs trading and accounting information // Boston university and mit working paper. — 2007.
- [15] Andrade Sandro, Di Pietro Vadim, Seasholes M. Understanding the profitability of pairs trading // Unpublished working paper, UC Berkeley, Northwestern University. — 2005.
- [16] Sharpe William F. The sharpe ratio // *Streetwise—the Best of the Journal of Portfolio Management*. — 1998. — P. 169–185.
- [17] Krauss Christopher. Statistical arbitrage pairs trading strategies: review and outlook. — 2017. — Vol. 31, no. 2. — P. 513–545. — Access mode: <https://onlinelibrary.wiley.com/doi/10.1111/joes.12153> (online; accessed: 2022-06-08).
- [18] Gatev Evan, Goetzmann William, Rouwenhorst K. Pairs Trading: Performance of a Relative-Value Arbitrage Rule // *Review of Financial Studies*. — 2006. — Vol. 19, no. 3. — P. 797–827. — Access mode: <https://EconPapers.repec.org/RePEc:oup:rfinst:v:19:y:2006:i:3:p:797-827>.
- [19] Do Binh, Faff Robert. Does Simple Pairs Trading Still Work? // *Financial Analysts Journal*. — 2010. — Vol. 66, no. 4. — P. 83–95. — <https://doi.org/10.2469/faj.v66.n4.1>.
- [20] Empirical investigation of an equity pairs trading strategy / Chen Huafeng, Chen Shaojun, Chen Zhuo, and Li Feng // *Management Science*. — 2019. — Vol. 65, no. 1. — P. 370–389.
- [21] Vidyamurthy Ganapathy. *Pairs Trading: Quantitative Methods and Analysis*. Wiley Finance. — Nashville, TN : John Wiley & Sons, 2004. — Aug.
- [22] Rad Hossein, Low Rand Kwong Yew, Faff Robert. The profitability of pairs trading strategies: distance, cointegration and copula methods // *Quantitative Finance*. — 2016. — Vol. 16, no. 10. — P. 1541–1558. — <https://doi.org/10.1080/14697688.2016.1164337>.
- [23] Lin Yan-Xia, McCrae Michael, Gulati Chandra. Loss protection in pairs trading through minimum profit bounds: A cointegration approach // *Journal of Applied Mathematics and Decision Sciences*. — 2006. — Vol. 2006, no. 1. — P. 2090–3359.
- [24] Caldeira João Frois, Moura Guilherme Valle. Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy // *Brazilian Review of Finance*. — 2013. — Vol. 11, no. 1. — P. 49–80. — Access mode: <https://ideas.repec.org/a/brf/journl/v11y2013i1p49-80.html>.

- [25] Elliott Robert J., * John Van Der Hoek, Malcolm William P. Pairs trading // *Quantitative Finance*. — 2005. — Vol. 5, no. 3. — P. 271–276. — <https://doi.org/10.1080/14697680500149370>.
- [26] Avellaneda Marco, Lee Jeong-Hyun. Statistical arbitrage in the US equities market // *Quantitative Finance*. — 2010. — Vol. 10, no. 7. — P. 761–782. — <https://doi.org/10.1080/14697680903124632>.
- [27] Cummins Mark, Bucca Andrea. Quantitative Spread Trading on Crude Oil and Refined Products Markets // *Quantitative Finance*. — 2011. — 08. — Vol. 12.
- [28] Montana Giovanni, Triantafyllopoulos Kostas, Tsagaris Theodoros. Flexible least squares for temporal data mining and statistical arbitrage. — 2009. — Vol. 36, no. 2. — P. 2819–2830. — Access mode: <https://linkinghub.elsevier.com/retrieve/pii/S0957417408000614> (online; accessed: 2022-06-08).
- [29] Ensembling and Dynamic Asset Selection for Risk-Controlled Statistical Arbitrage / Carta Salvatore M., Consoli Sergio, Podda Alessandro Sebastian, Recupero Diego Riformiato, and Stanciu Maria Madalina // *IEEE Access*. — 2021. — Vol. 9. — P. 29942–29959.
- [30] Huck Nicolas. Large data sets and machine learning: Applications to statistical arbitrage. — 2019. — Vol. 278, no. 1. — P. 330–342. — Access mode: <https://linkinghub.elsevier.com/retrieve/pii/S0377221719303339> (online; accessed: 2022-06-08).
- [31] Володин Сергей Николаевич, Коченков Илья Алексеевич. Статистический арбитраж на российском фондовом рынке // *Аудит и финансовый анализ*. — 2013. — № 6. — С. 237–244.
- [32] Эффективность парного статистического арбитража на российском фондовом рынке / Липатников ВС, Ломджария СГ, Мазуровский ПА и Маркова ТС // *Банковское дело*. — 2017. — № 5. — С. 47–51.
- [33] Давнис В.В., Тинякова В.И. Учебное пособие для слушателей магистерских программ. — 2009. — УДК 338.27:51.
- [34] Финам.ru. — 2022. — Режим доступа: <https://www.finam.ru/profile/moex-akcii/gazprom/export/> (дата обращения: 2022-06-12).