ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
Факультет компьютерных наук

Каюмов Руслан Асхатович

Reinforcing Learning для ребалансировки инвестиционного портфеля

Выпускная квалификационная работа [отчет #4]
по направлению подготовки 01.04.02 Прикладная математика и информатика
образовательная программа магистратуры «Машинное обучение и высоконагруженные системы»

Научный руководитель:
Максимовская Анастасия Максимовна

Рецензент:
_____

Москва 2023

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

[об актуальности и значения исследования - пока в черновых тезисах]

Развить и подтвердить источниками следующие тезисы:

- Существенный рост числа частных инвесторов в мире и России, привести график, фактически речь про бум частного инвестирования …
- Не забыть упомянуть и институциональных инвесторов …
- Развитие сервисов для частных инвесторов. …
- В задаче инвестирования, которая становится все более массовой, важным является не только грамотно составить инвестиционный портфель, но и осуществлять его регулярную оптимизацию с учетом изменений как рынков, так и стратегии. Этой цели отвечает задача ребалансировки портфеля …
- Развитие методов машинного обучения ставит вопрос о возможности использования новых подходов для повышения эффективности …

… … …

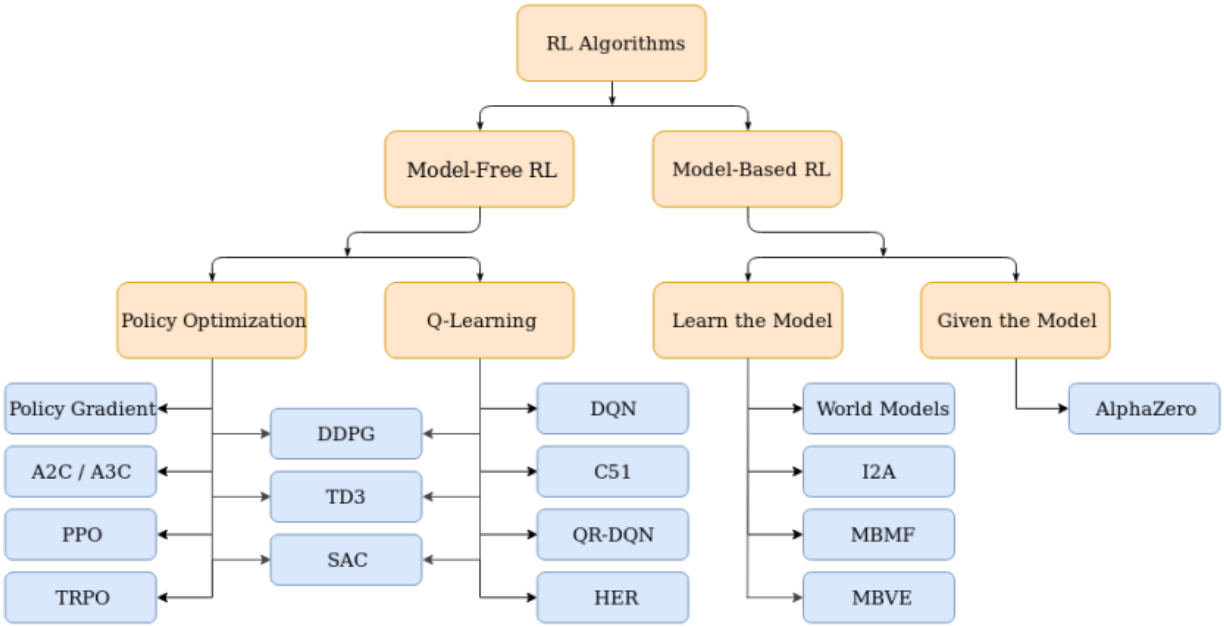
ГЛАВА 1. ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ

1.1 Существующие исследования

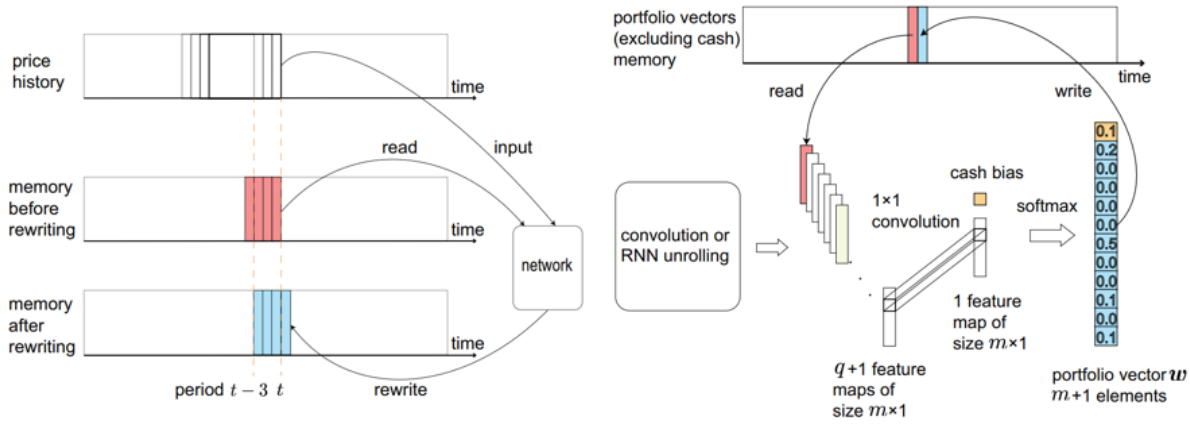
Ребалансировка инвестиционного портфеля представляет собой действия по изменению структуры инвестиций для восстановления баланса риска и доходности инвестиционного портфеля. Методы формирования портфеля подходят, как правило, и для ребалансировки. Но ребалансировка проводится регулярно при обновлении исторических данных по активам, при изменении уровня приемлемого риска или доходности, добавлении новых потенциальных активов и др.

В свою очередь, методы DRL (Deep Reinforcement Learning) для управления портфелем начали активно развиваться в последние 10 лет. В прошлом веке управление портфелем строилось традиционно на основе идей Марковица (Harry Markowitz), которые легли в основу современной портфельной теории (Modern Portfolio Theory, MPT) [1] Модель Марковица продолжала развиваться в последующие десятилетия. Например, в 90-е гг. появилась модель Модель Блэка-Литтермана от сотрудников Goldman Sachs Group [2].

С бурным развитием методов ML и ростом производительности ПК начало появляться все больше публикаций по применению методов машинного обучения и, в частности, RL к трейдингу и управлению портфелем [12]. Складывается определенная таксономия RL-методов [13]



Наконец, в последние 5 лет развиваются фреймворки с имплементированными моделями DRL [20]. Во второй версии своей статьи «A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem» авторы предлагают новые решения в виде независимых оценщиков (IEEI), векторной памяти портфеля (PVM) и OSBL.

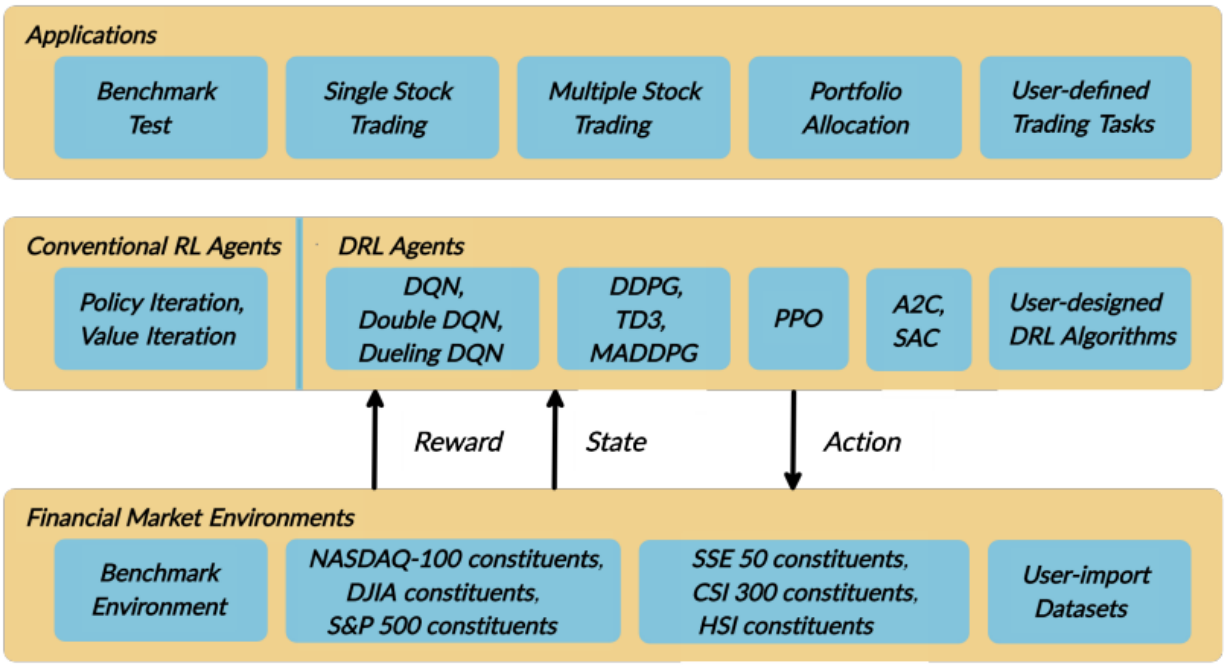


(a) Mini-Batch Viewpoint (b) Network Viewpoint

Одними из популярных агентов для реализации задач управления становятся A2C (Advantage Actor Critic), DDPG (Deep Deterministic Policy Gradient) и PPO (Proximal Policy Optimization). Так DDPG используется для решения задач, когда действия необходимо выбирать в непрерывном

пространстве; примером такой задачи является и распределение средств среди активов инвестиционного портфеля.

Многие современные модели реализованы в развивающемся современном open-source фреймворке FinRL от A4Finance Foundation [7], который позволяет



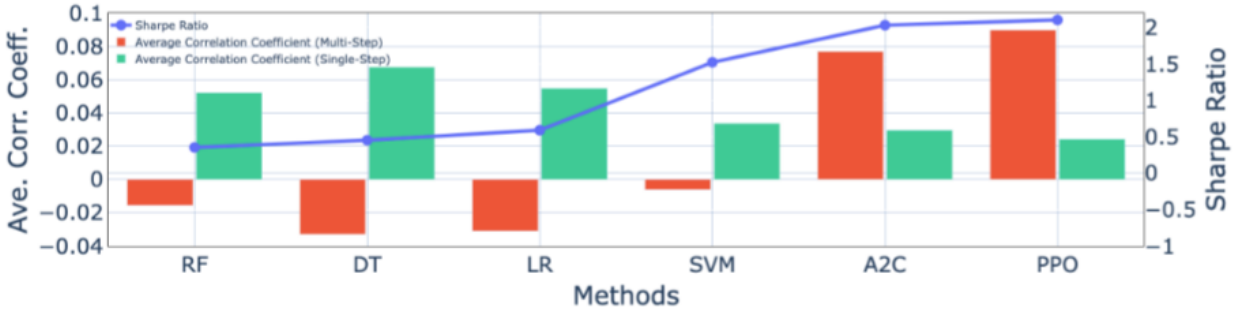
сравнивать работу моделей между собой и с бенчмарками, а также подключать собственные разработки. В проект имплементирован не только DDPG, но и расширения в виде TD3 или мультиагентного MADDPG. В целом, фреймворк поддерживает три источника моделей, среди которых Stable Baselines 3 – форк от OpenAI Baselines, переписанный на PyTorch и обладающий большим покрытием тестами [21].

	SB3	OAI Baselines	PFRL	RLlib	Tianshou	Acme	Tensorforce
Backend	PyTorch	TF	PyTorch	PyTorch/TF	PyTorch	Jax/TF	TF
User Guide / Tutorials	✓/ ✓	✗/ –	–/ ✓	✓/ ✓	–/ ✓	–/ ✓	✓/ –
API Documentation	✓	✗	✓	✓	✓	✗	✓
Benchmark	✓	✓	✓	✓	–	–	–
Pretrained models	✓	✗	✓	✗	✗	✗	✗
Test Coverage	95%	49%	?	?	94%	74%	81%
Type Checking	✓	✗	✗	✓	✓	✓	✗
Issue / PR Template	✓	✗	✗	✓	✓	✗	✗
Last Commit (age)	< 1 week	> 6 months	< 1 month	< 1 week	< 1 month	< 1 week	< 1 month
Approved PRs (6 mo.)	75	0	13	222	85	5	7

Исследования по DDPG [18] показывают, что с помощью данного агента, состоящего из двух нейронных структур (для Actor и Critic), могут

быть реализовано большинство торговых алгоритмов. Авторам удалось превзойти бенчмарк.

В той же работе авторы подчеркивает, что DRL-модели не должны рассматриваться, как «черный ящик», так как для них существует полная теория, позволяющая настраивать и контролировать работу модели. В публикации «Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach» [9], представленной на ICAIF'21, предлагается эмпирический подход к объяснению стратегий агентов DRL применительно к задаче управления портфелем. Подход сравнивается с другими ML-методами, в том числе с Random Forest. При этом изучается сила прогнозирования в двух случаях: одно- и многоэтапному (Multi-Step) предсказание. Авторы приходят к выводу, что DRL (A2C, PPO) обладает в контексте проводимого эксперимента большей способностью к многоэтапному прогнозированию, чем обычные методы машинного обучения.

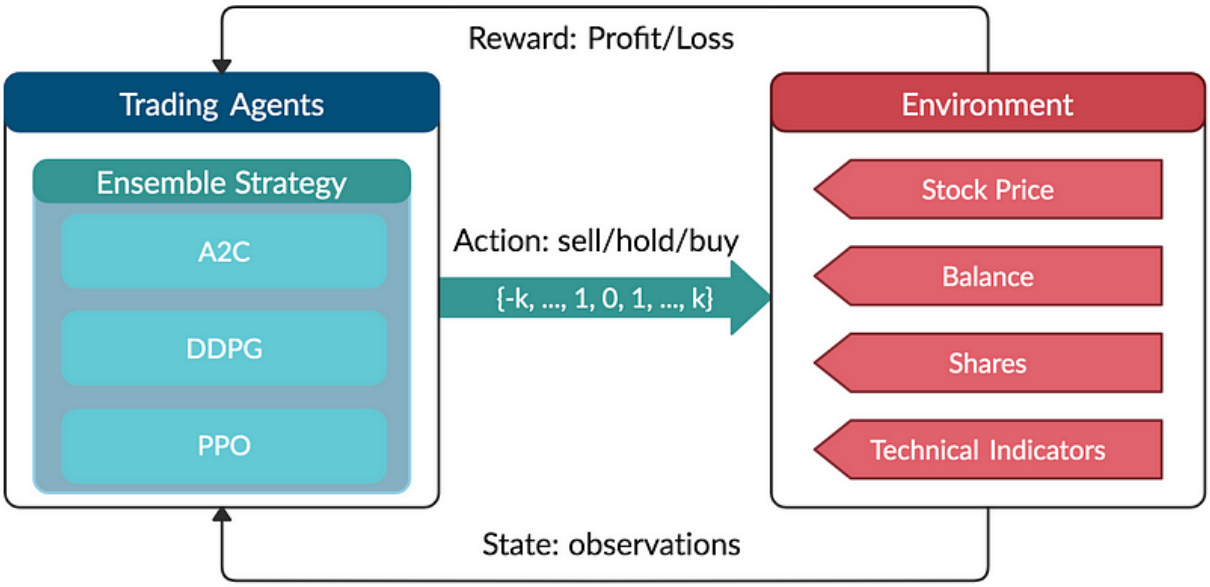


Некоторые работы сравнивают между собой DRL-агенты [15]. Вышедшая совсем недавно в 2023 году публикация [16] сравнивает PPO с подходами actor-critic и actor-only. PPO в исследовании хорошо защищает от падения рынка, но на растущем рынке получает доход ниже двух упомянутых подходов. Вводится алгоритм Reward Clipping, который сохраняет преимущества трех моделей.

Вместе с тем нейросети применяются параллельно для прогнозирования стоимости активов. В первую очередь, RNN и LSTM, адаптированные для временных рядов. Так в недавней работе «Dynamic portfolio rebalancing through reinforcement learning» [19] автор показывает, что RL-агент с использованием

LSTM показал улучшил результаты примерно на 30-90%. Другим важным моментом в работе является использование частичной перебалансировки вместо полной.

Некоторые авторы исследуют ансамблевые методы. Разработчикам FinRL в работе «Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy» [8], которая была опубликована на ICAIF'20, удалось совместить A2C, PPO, DDPG и показать, что она превосходит отдельные модели.



В упомянутых выше работах часто используются в качестве дополнительной информации о рынке (среде, если в терминах RL) технические индикаторы. Стандартные MA, BB, MACD, RSI и др.

В публикации «MAPS: Multi-agent Reinforcement Learning-based Portfolio Management System» [11] авторы демонстрируют мультиагентную RL-систему MAPS, где каждый агент является независимым инвестором, создающим свой собственный портфель. Каждый агент должен максимизировать свой доход при максимально возможном разнообразии активов. По итогам эксперимента на 12-летних данных американского рынка MAPS превосходит большинство базовых методов, такие как MOM, MR, MLP, CNN, DARNN.

Важным моментом является выбор метрики, как для оценки эффективности управления, так и в качестве параметра оптимизации. Рассматриваются такие показатели, как кумулятивная доходность, максимальная просадка (MDD), коэффициент Шарпа, Сортино, Кальмара. Практически во всех работах основными метриками являются доходность и Шарп. В свежей статье «A Comparative Study on the Sharpe Ratio, Sortino Ratio, and Calmar Ratio in Portfolio Optimization» [10] автор напрямую сравнивает разные метрики для целей оптимизации портфеля и приходит к выводу, что именно портфели, ориентированные на максимальный коэффициент Шарпа, приносят оптимальную прибыль. Автор подчеркивает превосходство коэффициента Шарпа в качестве параметра оптимизации для выбранных условий.

Помимо метрик и базовых моделей часто сравнивают эффективность с бенчмарком, в роли которого выступают в большинстве случаев фондовые индексы по американскому рынку. Чаще всего DJI (Dow Jones Index).

1.2 Теоретические аспекты по ребалансировке портфеля

Эконометрические методы долгое время оставались основными при составлении и оптимизации портфеля. Результаты работы наиболее распространенного из методов могут выступать в качестве бейзлайна для оценки эффективности предлагаемой модели.

1.2.1 Современная портфельная теория

Идеи Марковица (Harry Markowitz) лежат в основе современной портфельной теории (Modern Portfolio Theory, MPT) [1]. Метод основан на анализе ожидаемых средних значений и вариаций случайных величин. Активы подбираются таким образом, чтобы доходность была максимальной при заданном уровне риска.

С математической точки зрения метод сводится к задаче квадратической оптимизации при линейных ограничениях. Эта задача хорошо изучена и представлена большим числом эффективных алгоритмов.

Основой практической реализации модели является построение границы эффективности (Efficient Frontier), которая представляет собой набор портфелей, дающих максимальную ожидаемую доходность

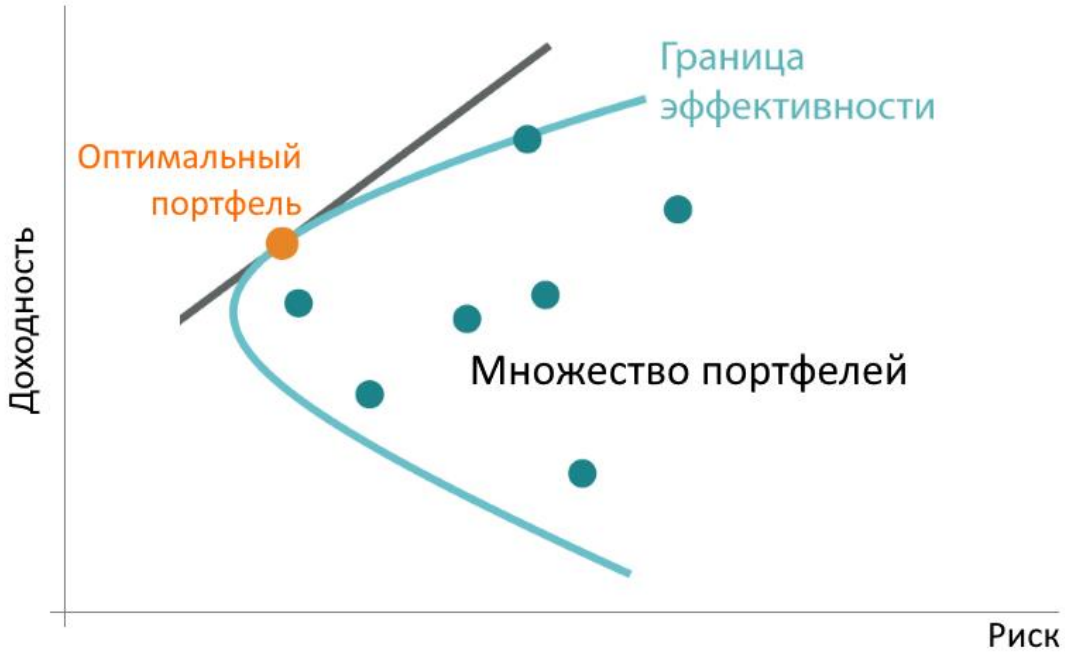


Рис 1. Иллюстрация границы эффективности в модели Марковица для выбора оптимального портфеля

1.2.2 Другие не RL-методы (черновой раздел)

Модель Блэка-Литтермана от сотрудников Goldman Sachs Group [2], которая призвана решить проблемы, возникающие при практическом применении модели Марковица.

[Монте-Карло] Метод имитационного моделирования для подбора оптимального соотношения весов.

CART - Classification and Regression Tree, или дерево решений. Непараметрический метод. Со ссылкой на [5]

[модели сокращения и иерархического сокращения проверить]

… … …

1.3 RL и DRL-методы (черновой, теория)

1.3.1 Теория, терминология

[Q-learning и прочее]

… …

1.3.2 Агенты A2C, PPO, DDPG

[теория, архитектура, отличия]

О развитии DDPG [перепроверить, добавить источники]

DDPG – это базовый алгоритм, оптимизированный для работы с непрерывными средами, разработан в 2015 году. С тех пор было создано множество его вариантов, расширений и улучшений. Некоторые из них:

- Twin Delayed DDPG (TD3) — расширение DDPG, которое включает в себя два критика и добавляет задержку в обновление весов критика. Это может привести к более стабильному обучению и повышению производительности.
- Soft Actor-Critic (SAC) —использует энтропийную регуляризацию для повышения исследования, более эффективного использования буфера воспроизведения и улучшения стабильности обучения.

- Distributed Distributional Deterministic Policy Gradients (D4PG) — использует распределенное обучение и дискретизированные оценки Q-функции для решения задач в непрерывных пространствах действий.
- Addressing Function Approximation Error in Actor-Critic Methods (ADA) — использует методы для уменьшения ошибок в функции приближения ценности Q.
- Soft Q-Learning (SQL) — использует энтропийную регуляризацию для повышения исследования, более эффективного использования буфера воспроизведения, улучшения стабильности обучения и улучшения быстродействия.
- Multi-Agent Deep Deterministic Policy Gradients (MADDPG) — используется для решения задач многих агентов, где несколько агентов действуют в среде, взаимодействуя друг с другом.

Общая тенденция в развитии DDPG — это добавление более продвинутых функций, таких как регуляризация, дискретизация, и более мощных слоев нейронных сетей для обучения больших наборов данных.

… … …

ГЛАВА 2. АНАЛИЗ И СПЕЦИФИКА ВХОДНЫХ ДАННЫХ

2.1 Особенности и структура входных данных

Используются финансовые данные, особенностями которых является:

- хронологическое представление (временные ряды),
- разнотипные активы с разными источниками в зависимости от страны и площадки размещения,
- история финансовых активов обычно представлена достаточным объемом данных (за 10 и более лет) с минимальным количеством ошибок.

Так как цели инвестиционного управления не являются краткосрочными, то достаточным шагом изменения стоимости актива является день или неделя. Стандартным таймфремом хранения рыночных данных является D1 – ежедневные данные. При необходимости их можно переконвертировать в другие таймфреймы (недельные и месячные)

При этом история по каждому активу должна включать в себя два и более лет, что позволит учитывать сезонные циклы и глобальные тренды.

Формат хранения рыночных данных обычно включает в себя следующие элементы:

- Дата и время начала периода изменения стоимости актива.
- Open (O) – цена открытия периода.
- High (H) – максимум, который был достигнут ценой за период.
- Low (L) – минимум, который достигнут ценой.
- Close (C) – цена закрытия.
- Volume (Vol) – объем торгов за период.

Наиболее важными из указанных элементов являются дата и цена закрытия. Цены максимума и минимума могут быть использованы для определения просадок по активу.

2.2 Классы инвестиционных активов и их первичный отбор

Классы инвестиционных активов, которые могут быть включены в портфель или задействованы в работе модели:

- Акции российских компаний, торгующиеся на Московской бирже (MOEX).
- Акции иностранных компаний, торгующиеся на американских биржах NYSE, NASDAQ и др.
- Основные валютные инструменты: EUR/USD, GBP/USD, USD/JPY
- Драгоценные металлы – золото.
- Фьючерсы и CFD на основные сырьевые товары и сельхозтовары: Нефть, Алюминий, Пшеница и др.
- Фьючерсы на биржевые индексы
- Основные криптовалютные инструменты: биткоин BTC/USD, эфир ETH/USD

На данном этапе можно сделать следующие предположения и заключения:

1. Учитывая присутствие активов, стоимость которых выражена в национальной валюте, на стадии разработки можно ограничиться международными активами, выраженными в USD.
2. Некоторые активы могут предусматривать выплату дивидендов, которые являются важной частью структуры инвестиционного

дохода. В данной работе будет учитываться только изменение стоимости самого актива.

3. В каждом из классов можно выделить наиболее популярные активы для составления портфеля на этапе разработки, всего 25:
   - EUR, JPY
   - GOLD
   - AAPL (Apple), WMT (Walmart), TSLA (Tesla), MSFT(Microsoft), AMZN (Amazon), GOOG (Google), JPM (JP Morgan Chase), LMT (Lockheed Martin), PFE (Pfizer), XOM (Exxon Mobil), V (Visa), PEP (PepsiCo), MCD (McDonald's), DIS (The Walt Disney), NFLX (Netflix), IBM, META
   - Cruide Oil, Natural Gas, Wheat
   - BTC (Bitcoin), ETH (Ethereum)
4. Историю для обучения модели можно взять за 13 лет – с начала 2010 года до конца 2022 года. Такой период позволит включить ряд провалов на рынке и учесть глобальные сезонные циклы.
5. Для MVP и разведочного анализа предлагается отобрать данные для 7 активов: EUR, GOLD, BTC, AAPL, XOM, V, CL.

2.3 Источники данных

Архивы финансовых данных могут предоставляться самими торговыми площадками (биржами), брокерами, агрегирующими сервисами.

Популярными универсальными источниками рыночных данных являются сервисы Google Finance и Yahoo Finance с публичным API. При этом доступ к Google Finance был ограничен для российских пользователей с 2022 года, а в Yahoo Finance с мая 2022 года отсутствуют котировки российских акций.

Чтобы избежать смещения акцента на решение инфраструктурных проблем использования разных источников планируется разрабатывать и тестировать сервис с использованием исторических данных, полученных через Yahoo Finance [3]. Архитектурно можно предусмотреть подключение коннекторов к другим источникам.

Временное размещение загруженных осуществляется в облаке по ссылке - https://drive.google.com/drive/folders/1JargvYx5-qjBfl95U6X4MhdTi-ccL5y2

2.4 Разведочный анализ данных

Как было условлено в предыдущих пунктах, для разведочного анализа используем 7 финансовых активов из 5 классов с историей данных за 3 года с 2020 по 2022 г.

В финансовых данных практически не наблюдаются пропуски и ошибки. Пропуски в истории некоторых активов могут быть обусловлены неторговыми днями в праздники и выходные. В зависимости от класса актива такие неторговые периоды могут быть разными. Например, биржевые инструменты торгуются по графикам бирж с соблюдением праздников, выходных и внутридневных неторговых часов. Тогда как внебиржевые котировки валют могут изменять всегда, кроме выходных, а котировки криптовалют изменяются и в выходные.

Для поставленной задачи формирования и оптимизации инвестиционного портфеля ввиду долгосрочности горизонта планирования решено перейти с дневного таймфрейма на недельный, где будет фактически нивелирована разница в графике формирования котировок между разными классами активов.

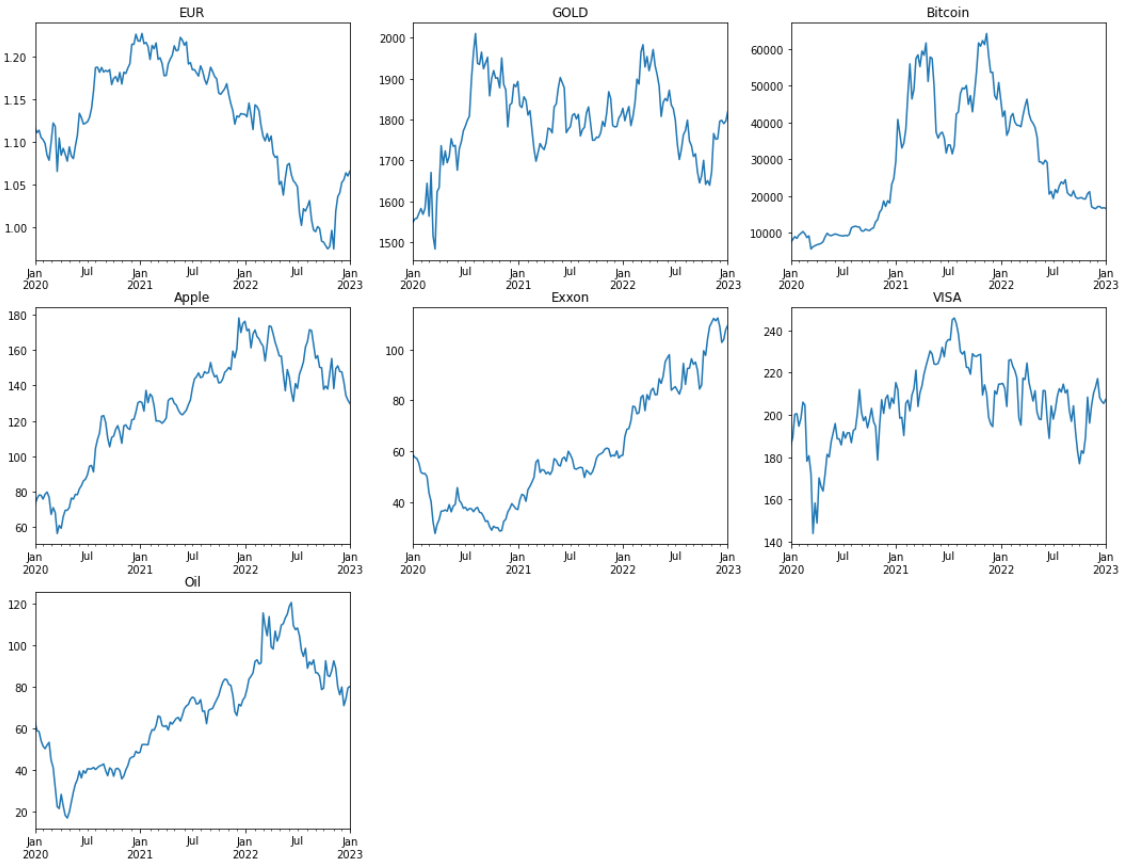


Рис 2. График изменения стоимости активов с 2020 по 2022 год.

Посмотрим на изменение стоимости активов по годам.

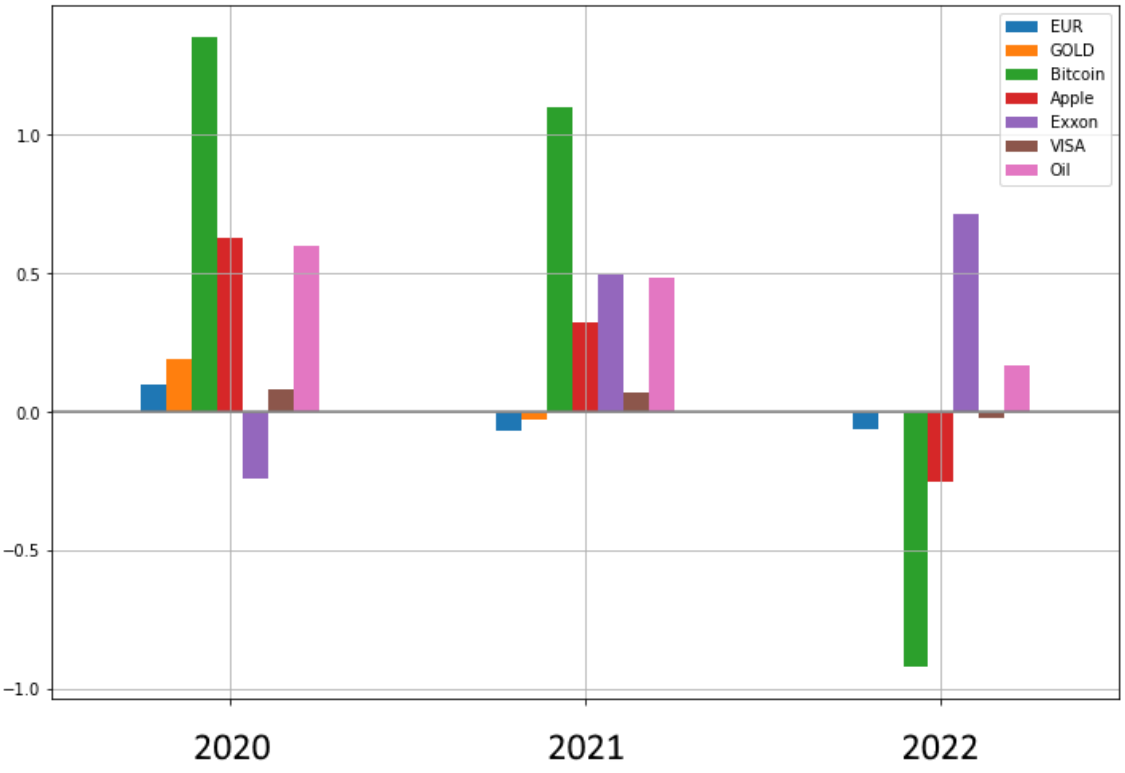


Рис 3. Сравнение изменения стоимости активов в долях по годам.

Хорошо видно, что периоды роста и падения стоимости активов различаются. Так в 2022 году лидером падения стал биткоин, а лидером роста Exxon Mobile, хотя годом ранее оба актива показали рост. При этом Exxon Mobile оказался единственным из рассматриваемых активов, который показал снижение в 2020 году.

В свою очередь евро и золото показывают очень слабое изменение сравнительно с другими инструментами на протяжении всего рассматриваемого отрезка. Очевидно, что такие инструменты вносят слабый вклад в результат портфеля при немарижнальной торговле. Но модель должна уметь работать с такими инструментами (например, самостоятельно исключать их), поэтому они будут учитываться в работе.

Попробуем оценить корреляцию между активами. Для этого с помощью тепловой карты сравним помесячное относительное изменение стоимости активов.

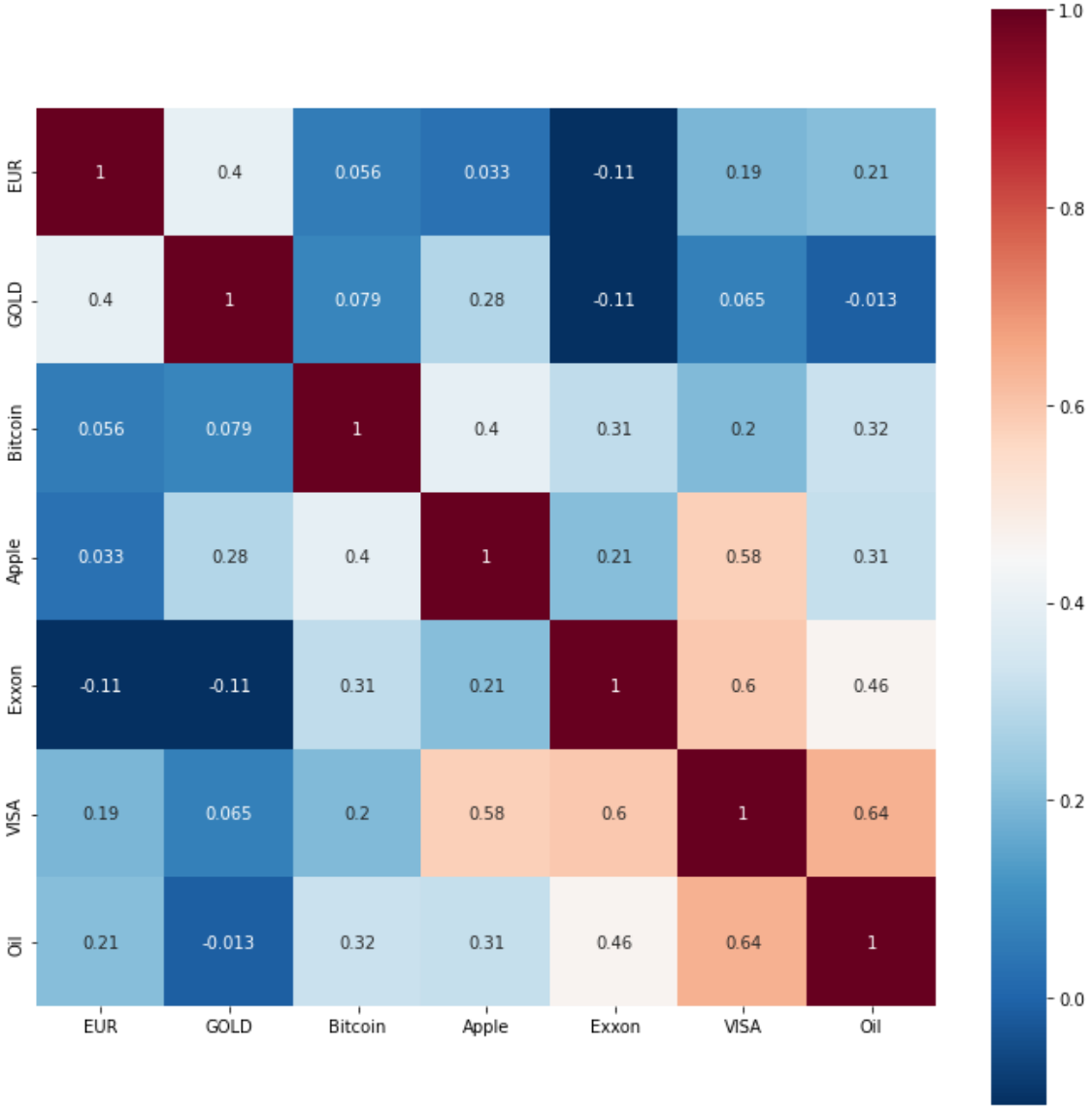


Рис 4. Корреляция помесячного относительного изменения стоимости активов.

Наблюдается заметная корреляция между некоторыми активами. Например, вполне очевидная положительная корреляция между стоимостью сырой нефти и акциями Exxon Mobile, что соответствует и более ранним наблюдениям [4]. Можно предположить, что в портфеле не должны преобладать коррелированные инструменты, так как это повышает риски просадки всего портфеля.

Посмотрим на распределение относительных недельных изменений стоимости активов на одном графике.

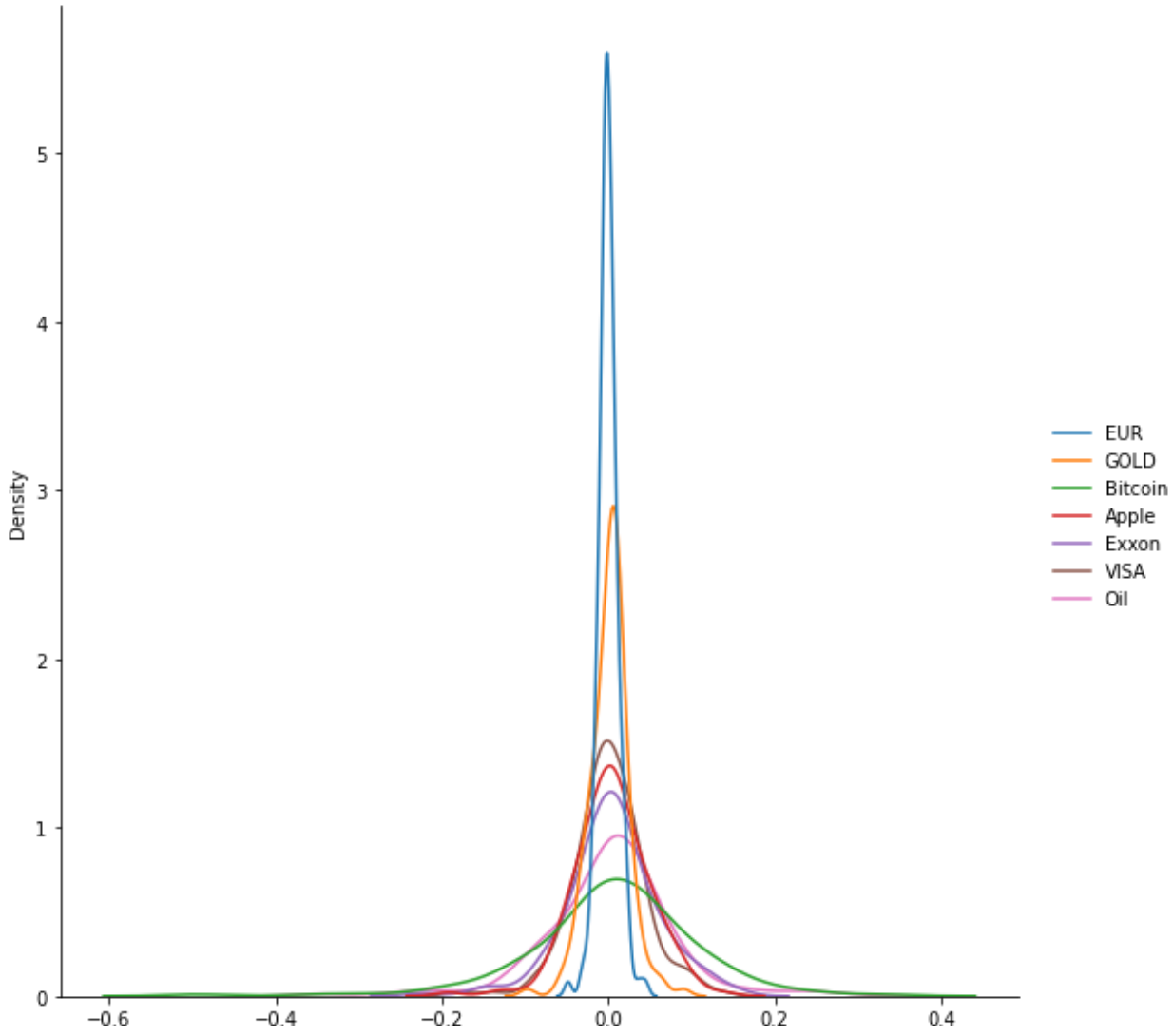


Рис 5. Распределение относительных недельных изменений стоимости активов.

Распределение близко к нормальному. Большая часть изменений происходит в околонулевой области с хвостами в положительную и отрицательную стороны. При этом можно отметить, что у разных активов может сильно отличаться ширина распределения. Так у EUR большое число недельных изменений сосредоточено ближе к нулю. Волатильность актива гораздо ниже, чем, например, у Биткоина.

Отметим выбросы, предварительно отнормировав шкалу стоимости активов.

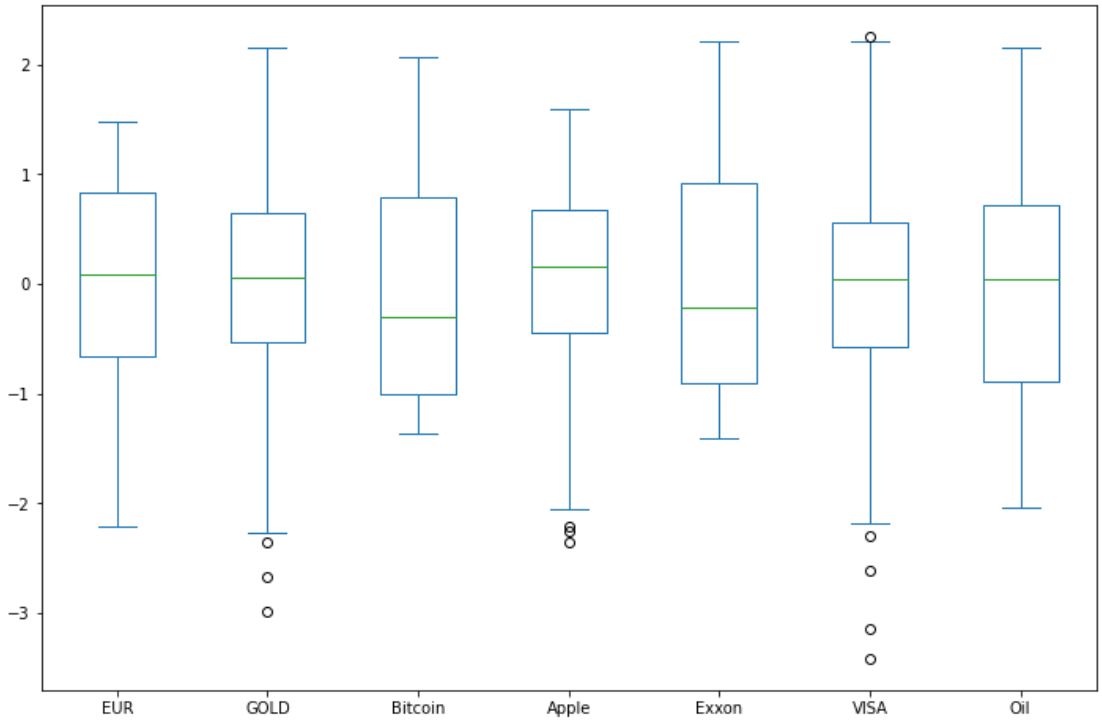


Рис 6. Boxplots по отнормированной стоимости активов.

Тут можно заметить выбросы для одних инструментов и их отсутствие для других, как и отличие в форме распределения. Это свидетельствует о разном характере волатильности стоимости активов.

Так как имеем дело с временными рядами, то посмотрим на графики автокорреляции.

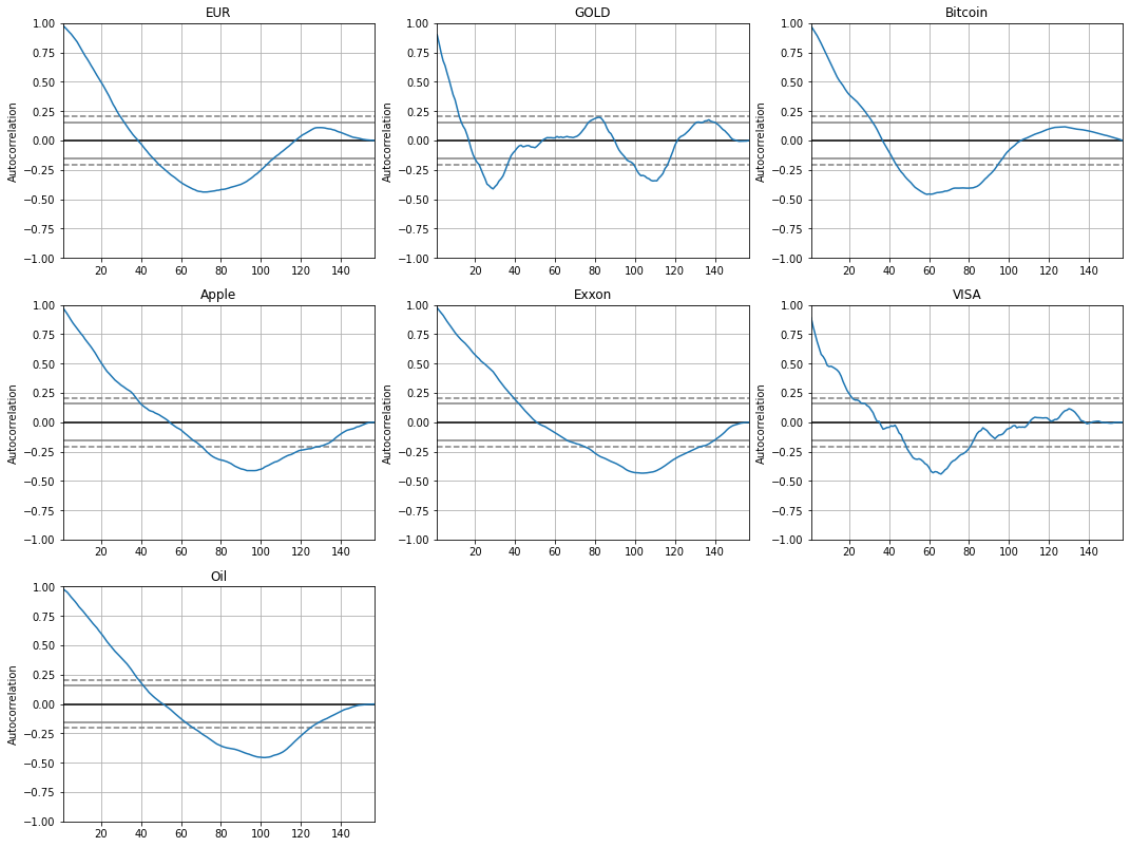


Рис 7. Графики автокорреляции.

Ожидаемо финансовые ряды не являются стационарными. Можно выделить компоненты тренда и сезонности. Для более детального анализа компонент проведем STL декомпозицию с периодом 12, или квартальный период для недельного таймфрейма.

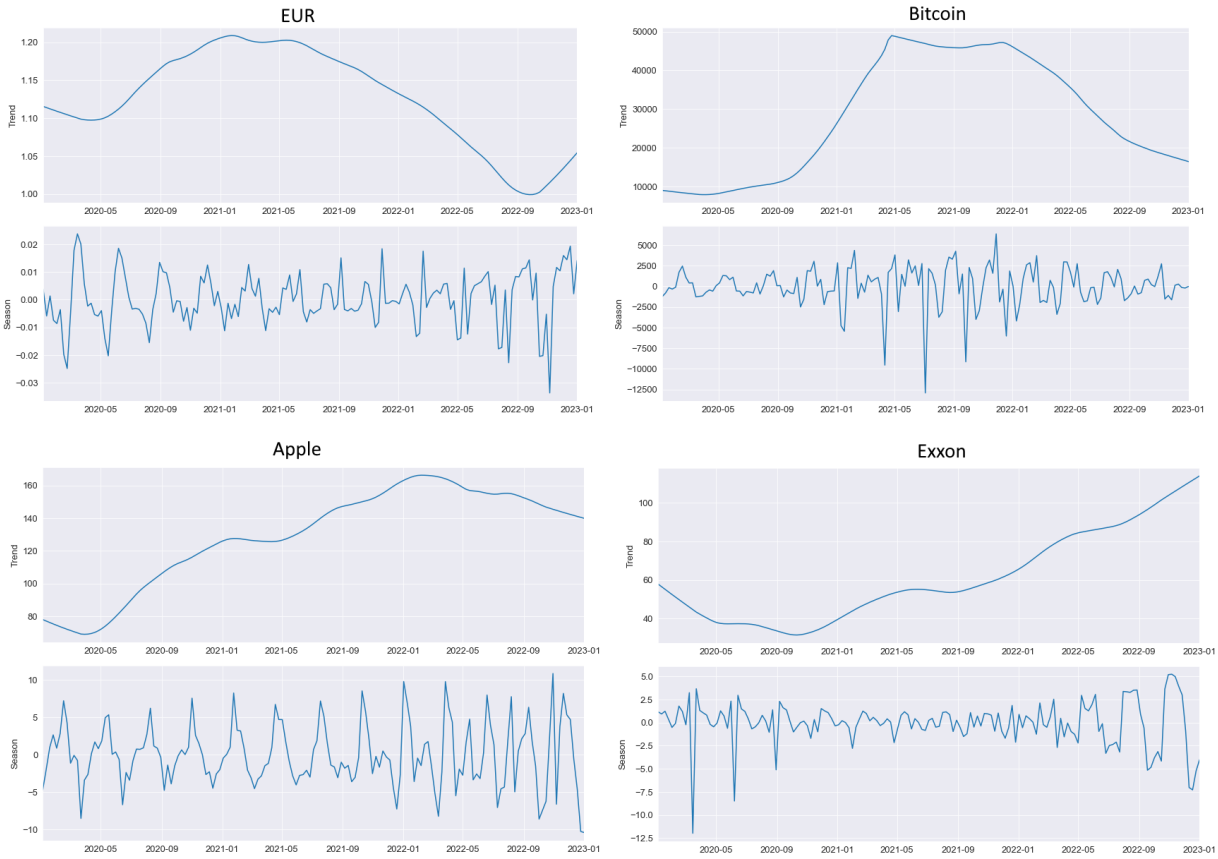


Рис 8. Трендовые и сезонные компоненты на графиках STL-декомпозиции.

При разложении хорошо выражены и тренд, и сезонность. Так можно выделить сезонные зависимости для Apple и тренд для Exxon Mobile.

Выводы:

- Активы отличаются разным характером волатильности, могут обладать сильными и слабыми трендовыми и сезонными компонентам. Активы с положительным выраженным трендом целесообразнее использовать в портфеле.
- Существуют как периоды долгосрочного роста актива, так и долгосрочного снижения стоимости.
- Среднесрочно актив может расти на фоне общей картины снижения.

- Могут наблюдаться пики снижения, что повышает риски для портфеля. Оптимально было бы учитывать форс-мажорные изменения на рынке.
- Между инструментами может наблюдаться корреляция. Модель должна уметь составлять портфель из некоррелированных активов для диверсификации рисков. В то же время частичная отрицательная корреляция может позволить хеджировать риски по некоторым активам.

[Связанный файл проекта - https://github.com/KayumovRu/RL-invest-optimization/blob/master/notebooks/download_and_EDA.ipynb]

ГЛАВА 3. BASELINE МОДЕЛЬ

3.1 Практическая реализация модели Марковица

Расчеты проведены для MVP-данных: 7 активов из 5 классов с трехлетней историей торгов. При этом период для обучения составляет 2 года, тестирование – 1 год.

Сгенерировано 10 000 портфелей для последующего отбора, рассчитаны границы эффективности, сформулированные в классической теории Марковица.

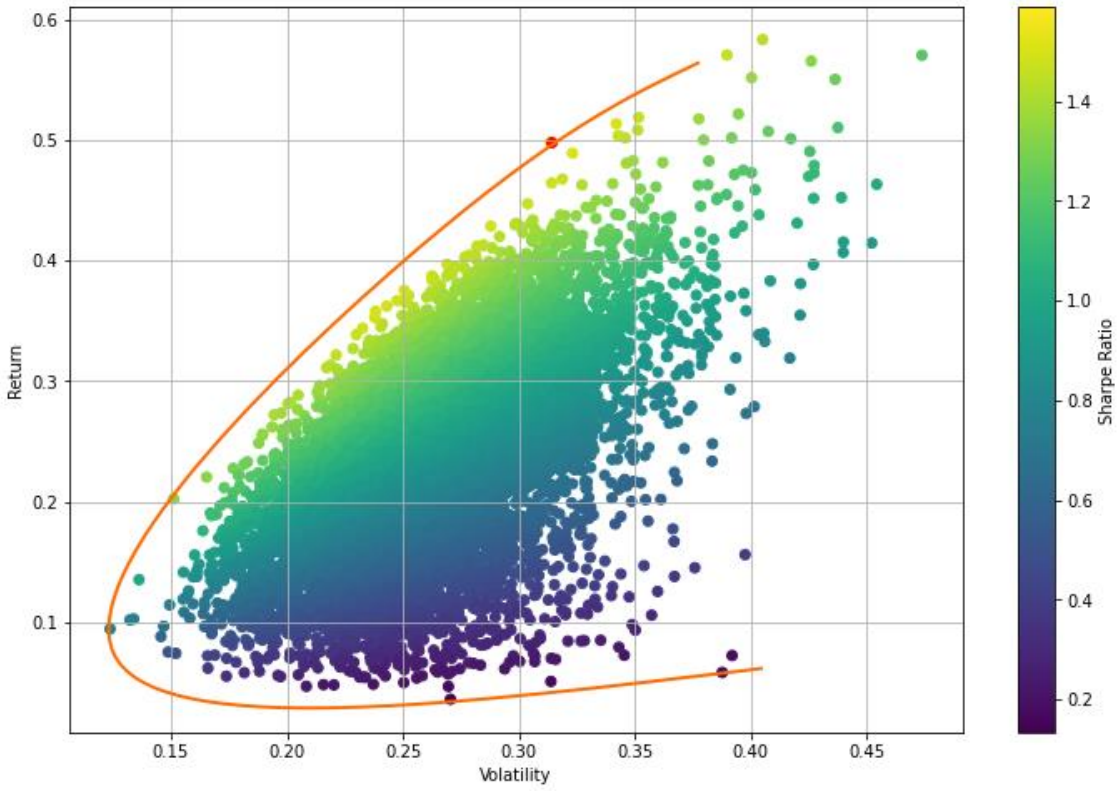


Рис 8. Визуализация границ эффективности (Efficient Frontier)

Отобран оптимальный с точки зрения модели портфель с максимальным коэффициентом Шарпа. На обучающих данных прибыль за 2 года составила около +90%, или 45% годовых

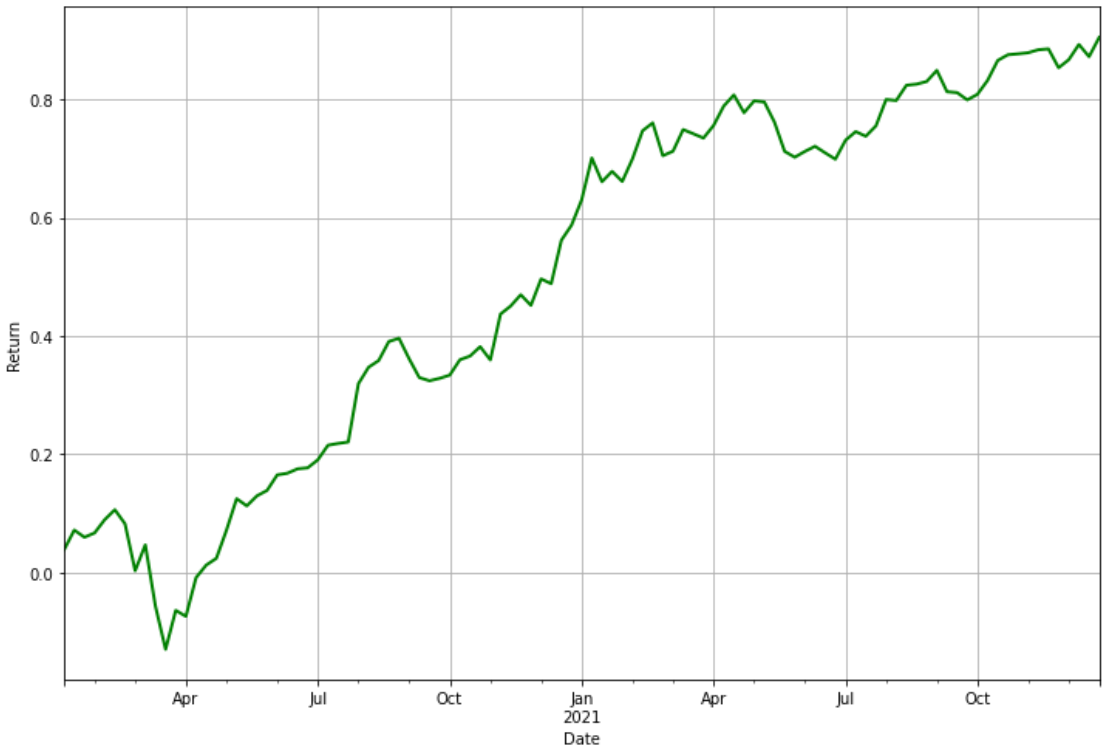


Рис 9. Портфель по модели Марковица на обучающих данных.

Характерно, что при проверке на бэктесте положительная картина не сохранилась. Портфель показал на тесте отрицательную доходность -27% годовых.

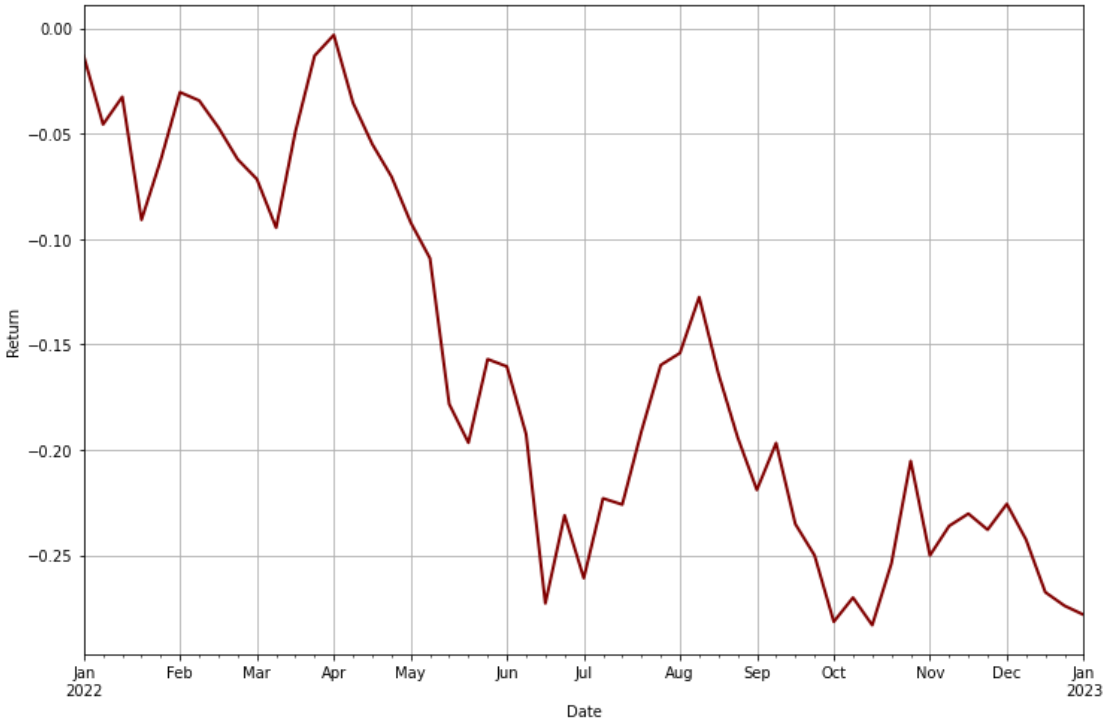


Рис 10. Портфель показал убыток при проверке на бэктесте.

Этот ожидаемый результат демонстрирует, что рассчитанные на прошлой истории веса быстро устаревают, модель обладает плохой робастностью. Но регулярная ребалансировка позволит с некоторой периодичностью пересчитывать доли активов в портфеле и таким образом актуализировать модель.

3.2 Добавление модуля периодической ребалансировки

Было решено поддержать скользящее окно ребалансировки и периодичность ее проведения.

Опытным путем подобрано окно скольжения 180 дней с шагом ребалансировки каждые 90 дней. В том числе это означает, что во время каждой ребалансировки новые данные составляют половину от всей обучающей истории.

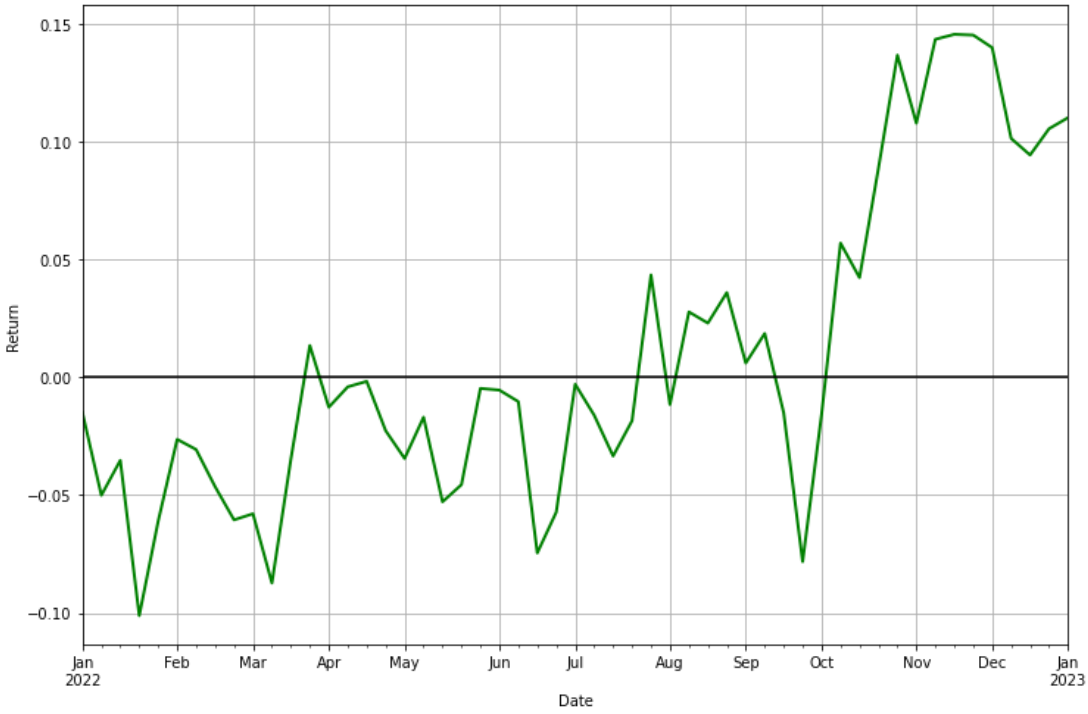


Рис 11. Портфель показал прибыль на бэктесте при добавлении модуля периодической ребалансировки.

Таким образом, всего за 2022 год доли в портфеле пересматривались 4 раза, что позволило выйти по итогам года на прибыль в +11%. Полученный результат может свидетельствовать о высокой роли периодической ребалансировки для улучшения работы модели.

На следующем графике можно отследить изменение доли активов в портфеле после ребалансировок.

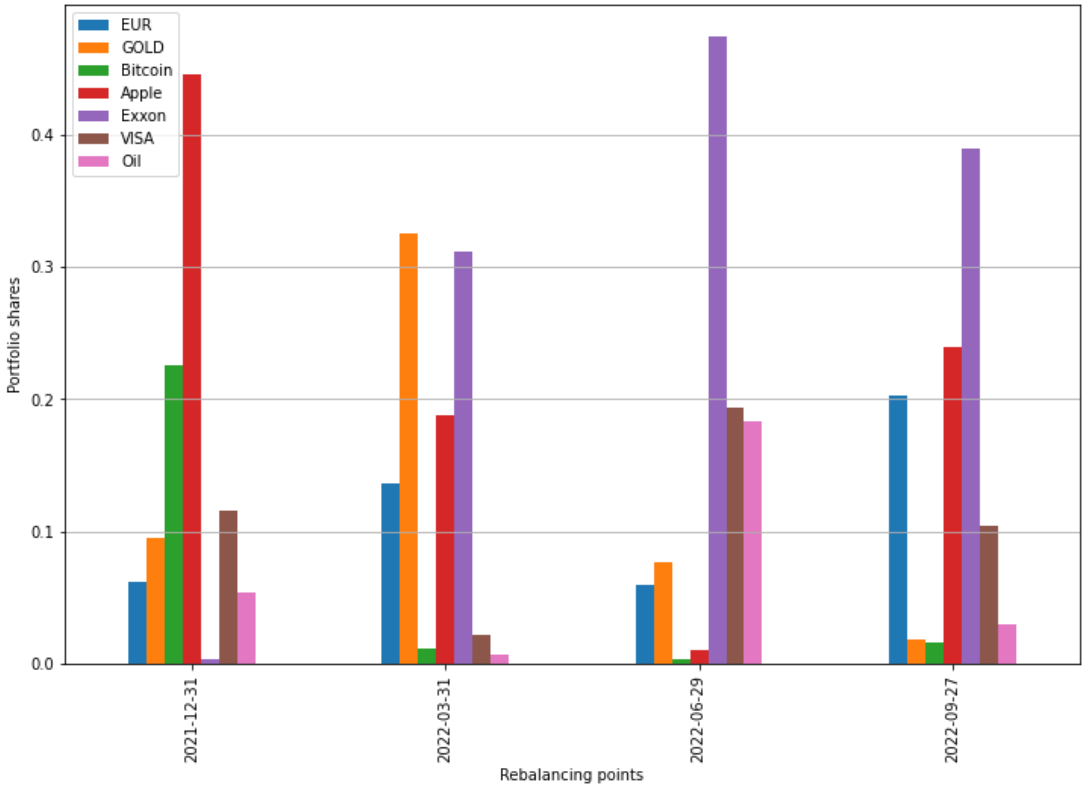


Рис 12. Изменение доли активов в портфеле после ребалансировок.

К примеру, доля Биткоина после ребалансировки в марте была сокращена до минимума. И действительно, после роста в 2020-2021 гг. биткоин падал или незначительно рос в 2022 году. Ребалансировка позволила своевременно исключить его из портфеля. Тогда как доля акций нефтяной компании Exxon Mobil выросла до лидера портфеля.

[Связанный файл проекта – https://github.com/KayumovRu/RL-invest-optimization/blob/master/notebooks/Markowitz_rebalance.ipynb]

ГЛАВА 4. DRL-МОДЕЛЬ

4.1 Эксперимент MVP – условия и схема валидации

Цель на данном этапе - имплементировать и сравнить DRL-агенты между собой, с базовой моделью и с бенчмарком. Для выбора оптимальной модели, которая будет реализована в конечном сервисе.

- Используются расширенные данные по 7 активам из 5 классов - EUR, GOLD, BTC, AAPL, XOM, V, CL - за период в 8 лет с 2015 по 2022 гг.
- Таймфрейм – минимальный временный промежуток, на которые разделены данные - одна неделя. Это связано с удобством использования недельных данных для активов из разных рынков. При этом такого деления достаточно для реагирования на рыночные изменения.
- Обучение производится на периоде в 5 лет: 2015 – 2019 гг.
- Тестирование (бэктест) проводим на периоде 2020 – 2022.
- В качестве бенчмарка используется индекс DJI за тот же период.
- В качестве метрики сравнения эффективности моделей: кумулятивная прибыль и коэффициент Шарпа.
- Реализуемые DRL-агенты: A2C, PPO, DDPG.
- Добавлены индикаторы в качестве дополнительных признаков: MACD, Bollinger Bands (ub и lb), RSI, CCI, DX.

4.2 Проведение эксперимента

Собрана среда и имплементированы DRL-агенты A2C, PPO, DDPG.

Число итераций по каждой модели составило от 50 000 до 80 000. Отмечено, что при обучении DDPG результат перестал изменяться уже после примерно 15 тыс. шагов.

Агенты обучены на данных до 2019 года и запущены на тестовых данных с 2020 года для бэктеста.

Кроме того, оказалось, что скользящее окно ребалансировки, которое использовалось для модели Марковица на стадии создания базовой модели, не помогло при увеличении тестовых данных до трех лет, а только ухудшило результат базовой модели. Поэтому принято решение на данном эатпе строить модель Марковица без скользящего окна ребалансировки.

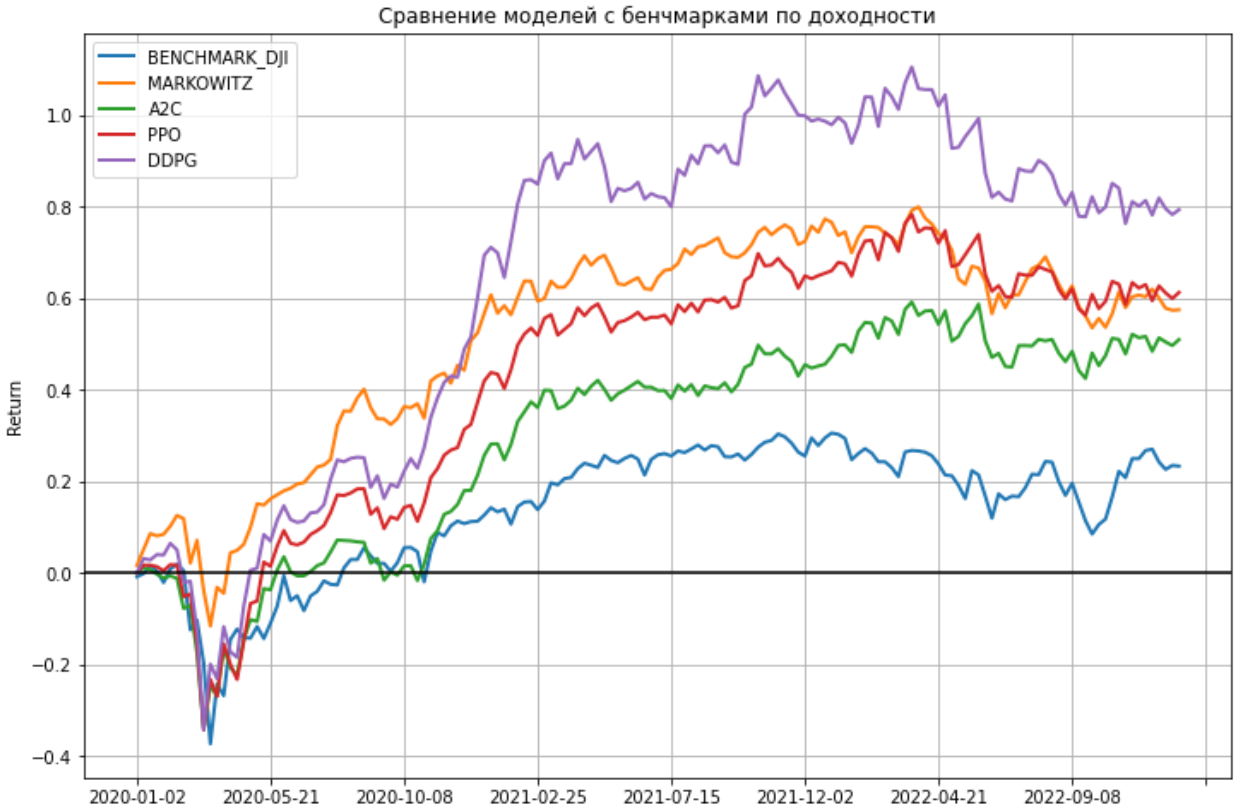
[Связанный файл проекта – FinRL_02_tests.ipynb]

4.3 Результаты эксперимента

По кумулятивной доходности на бэктестах с 2020 по 2022 гг лучший результат показала DRL-модель на основе DDPG +79%.

	DJI (бенчмарк)	Марковиц (базовая модель)	A2C	PPO	DDPG
Доходность за все время	+23.3%	+57.5%	+51.0%	+61.3%	**+79.4%**
Доходность среднегодовая	+7.7%	+19.1%	+17.0%	+20.4%	**+26.4%**
Коэффициент Шарпа	1.67	**2.45**	2.15	2.16	2.04

Результаты бэктеста представлены в таблице и на графике ниже.



Бенчмарк DJI продемонстрировал наихудший результат по обоим показателям. То есть в данном эксперименте любая модель превосходит простую покупку фондового индекса.

По доходности A2C уступил Марковицу, PPO незначительно превзошел его, а DDPG показал существенно лучший результат.

В то же время по коэффициенту Шарпа модель Марковица оказалась лучше, чем любые другие, включая DDPG.

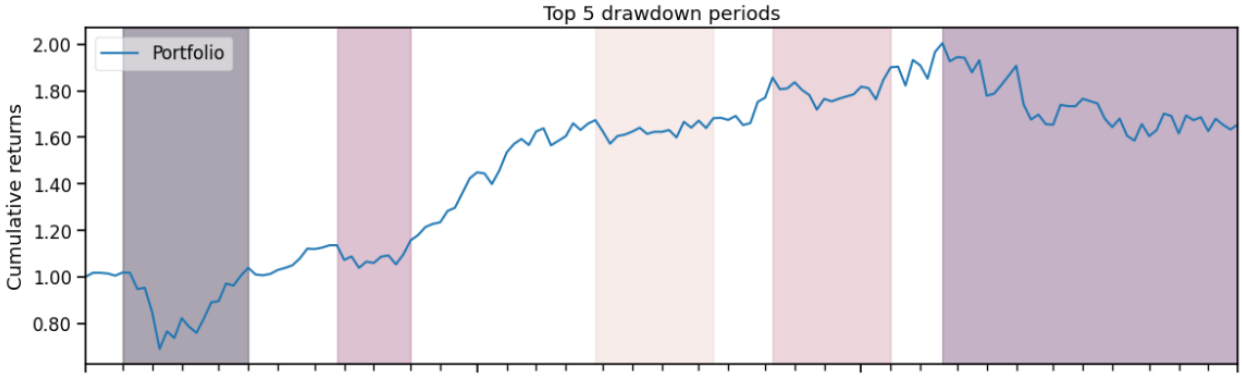
Это связано прежде всего с просадкой в начале 2020 года. Модель Марковица, ориентированная на максимальный Шарп, смогла лучше остальных справиться с просадкой.

На графиках, сформированных с помощью автоматических отчетов PyPortfolioOpt [6] для модели DDPG, можно обнаружить, что:

- Наибольшая просадка была получена за один месяц 2020 года.
- По итогам 2022 году был получен убыток – это общая проблема всех моделей. Интересно, что применение скользящего окна ребалансировки к Марковицу помогает избежать убытка в 2022,

но приводит к значительному падению прибыли в предыдущие годы.

5 наиболее больших периодов просадки на DDPG – самый темный участок в начале 2020 года соответствует наибольшей просадке.



[Связанные файлы проекта

- FinRL_03_mvp_results.ipynb
- test_results]

4.4 Выводы по эксперименту и дальнейший план

Наиболее перспективной моделью из проверенных является DDPG. Но в то же время она проигрывает по Шарпу. Меры, которые можно предпринять для улучшения результатов на основном этапе:

- более тонкая настройка агента и изменение архитектуры,
- проверка на большем числе активов (23 вместо 7 в MVP),
- увеличение истории (начало обучения с 2009 года вместо 2015),
- увеличение числа признаков «сенсоров» агента за счет добавления большего числа индикаторов,
- ансамблирование нескольких DRL-моделей [8],
- применение постепенности перебалансировки [19].

4.5 Финальный эксперимент

Увеличено количество активов до 23, размер истории с 8 лет до 14 лет. Классы активов – 18 акций и 5 сырьевых фьючерсов.

В качестве базовой модели выбрана равнодолевая модель (Equal Weight), которая равномерно распределяет доли по всем активам. Эта модель показывает положительные результаты и является хорошим бенчмарком.

Архитектура слоев нейронных сетей Actor и Critic в модели DDPG

Слой	**Размерность**
Conv2d	(1, 32, (1,3))
Conv2d	(32, 32, (1, 3))
Linear	(24*1*32, 64)
Linear	(64, 64)
Linear	(64, 24)

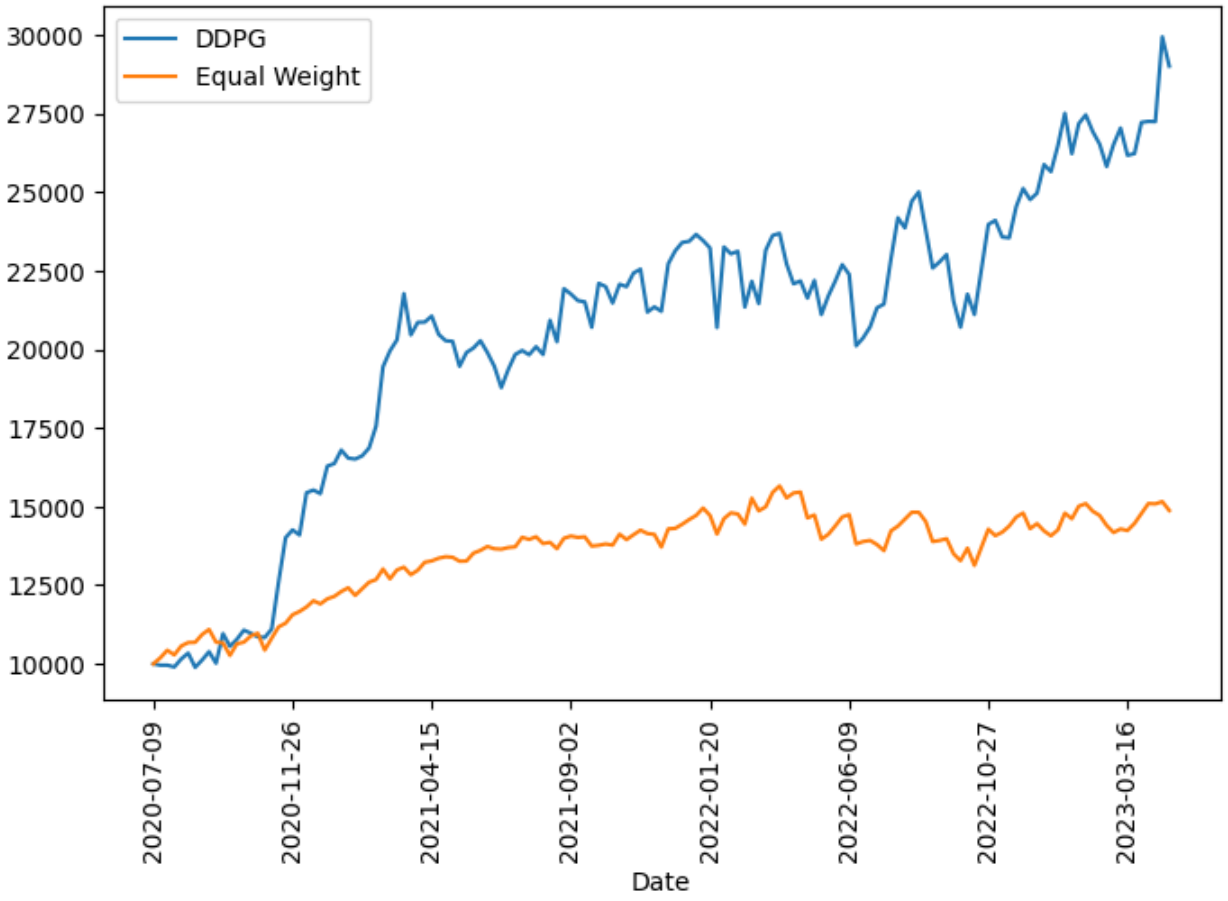
Обучение проведено на периоде с 2009.01.01 – 2020.06.17. Тестирование: 2020.06.18 – 2023.04.28

Получены результаты на тестовом периоде [добавить для периода обучения]

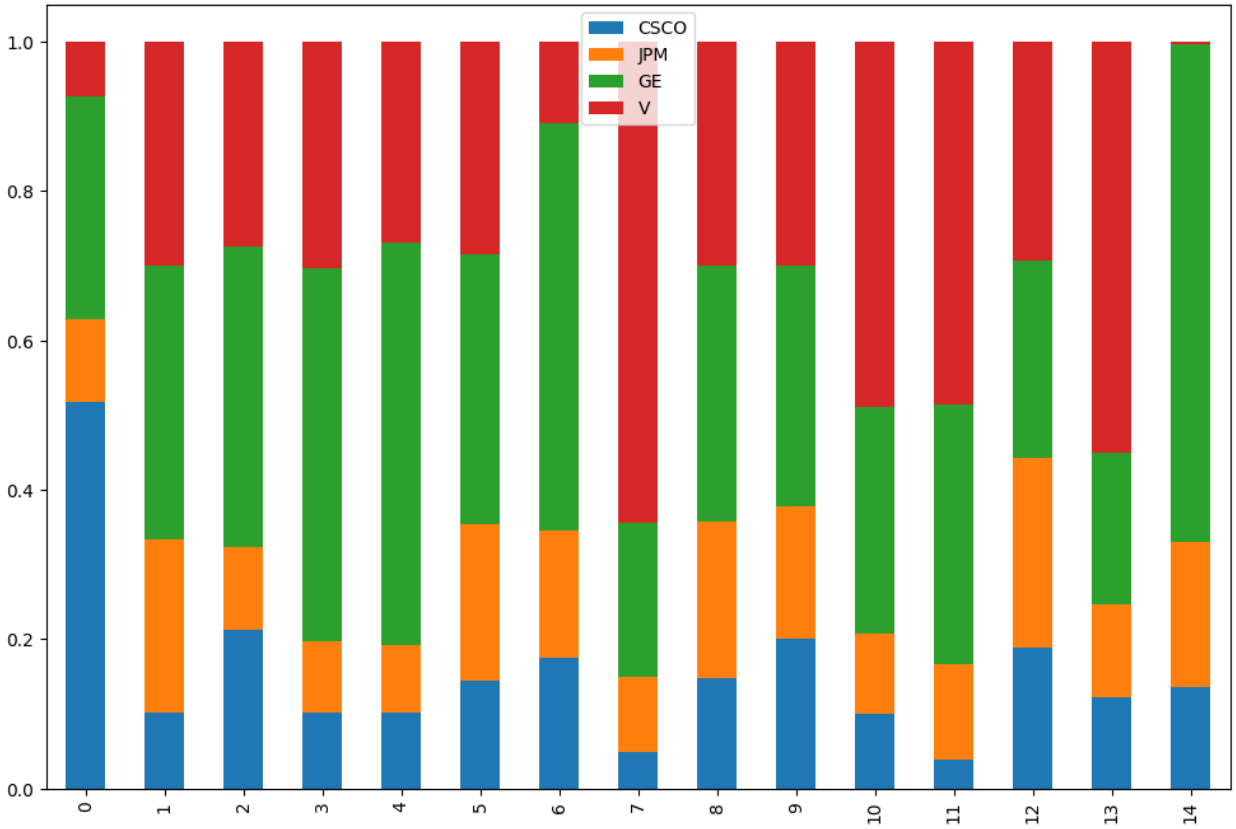
	Максимальная просадка	Коэффициент Шарпа	Прибыль
Equal Weight (benchmark)	**0.16**	2.23	+48%
DDPG	0.17	**3.22**	**+189%**

При почти одинаковой просадке прибыль существенна выросла, что привело и к значительному росту коэффициента Шарпа.

График сравнения эффективности DDPG с равнодолевой моделью:



При этом изменение распределения активов портфеля во время тестирования выглядит следующим образом:


CSCO
JPM
GE
V
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14

ГЛАВА 5. РАЗРАБОТКА КЛИЕНТ-СЕРВЕРНОГО ПРИЛОЖЕНИЯ

5.1 Архитектура проекта

5.1.2 Стек

- PostgresSQL – БД для хранения данных,
- Streamlit – фронтенд сервиса для визуализации и формирования пользовательского интерфейса,
- FastAPI – бэкенд сервиса.

[добавить больше информации]

5.1.3 Расчет ресурсов

Планируется отобрать 23 актива. Данные по неделям за 14 лет – с 2009 по 2022 г. Исторические данные будут храниться в одной таблице с единой колонкой даты и ценами закрытия по каждому инструменту.

[добавить больше информации]

5.2 Реализация клиент-серверного приложения

Интерфейс приложения поддерживает выбор моделей, размер депозита, размера окна исторических данных, выбор соотношения обучающей и тестовой выборок, выбор используемых для обучения признаков и активов.

На основе выбранных настроек обучаются и применяются модели (включая DDPG) для расчета оптимального распределения активов в портфеле.

Интерфейс клиентской части:

[добавить больше информации]

ЗАКЛЮЧЕНИЕ

[Подведем итоги проведенного выпускного квалификационного исследования и охарактеризуем кратко его основные результаты.]

… … …

СПИСОК ИСТОЧНИКОВ

1. Markowitz, H.M.: “Portfolio Selection”. The Journal of Finance. 7 (1), March 1952: pp. 77–91.
2. Black F., Litterman R.: “Global Portfolio Optimization”. Financial Analysts Journal, September 1992, pp. 28–43
3. Михаил Шардин.: “Все финансовые рынки мира в API Yahoo Finance”, Habr.ru, Июнь 2020, URL: https://habr.com/ru/post/505674/
4. Alex Rosenberg.: “The 5 stocks that are most correlated to oil”. Yahoo Finance, December 2015, URL: https://finance.yahoo.com/news/want-bet-oil-bounce-204900158.html
5. Buyanova E., Sarkisov A.: “Constructing of Optimal Portfolio on Russian Stock Market Using Nonparametric Method – Classification and Regression Tree”. Corporate Finance Journal, 1 (37), 2016: pp. 46 – 58.
6. Martin, R. A.: “PyPortfolioOpt: portfolio optimization in Python”. Journal of Open-Source Software, 6(61), 2021. URL: https://doi.org/10.21105/joss.03066
7. Xiao-Yang Liu, Hongyang Yang, Qian Chen, Runjia Zhang, Liuqing Yang, Bowen Xiao, Christina Dan Wang.: “FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance”. In NeurIPS 2020 Deep RL Workshop, 2020. URL: https://arxiv.org/abs/2011.09607
8. Hongyang Yang, Xiao-Yang Liu, Shan Zhong, Anwar Walid.: “Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy”. In ICAIF ’20, 2020, NY. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3690996
9. Mao Guan, Xiao-Yang Liu.: «Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach». In ICAIF ’21, 2021. URL: https://arxiv.org/abs/2111.03995

10. Jaydip Sen.: «A Comparative Study on the Sharpe Ratio, Sortino Ratio, and Calmar Ratio in Portfolio Optimization», Dec 2022. URL: https://www.researchgate.net/publication/366517929_A_Comparative_Study_on_the_Sharpe_Ratio_Sortino_Ratio_and_Calmar_Ratio_in_Portfolio_Optimization
11. Jinho Lee, Raehyun Kim, Seok-Won Yi, Jaewoo Kang.: «MAPS: Multi-agent Reinforcement Learning-based Portfolio Management System», 2020. URL: https://arxiv.org/abs/2007.05402
12. John Moody, Matthew Saffell.: «Learning to Trade via Direct Reinforcement», IEEE, 2001. URL: https://bi.snu.ac.kr/SEMINAR/Joint2k1/ojm5.pdf
13. Nitin Kanwar.: «Deep Reinforcement Learning-based Portfolio Management», The University of Texas at Arlington, 2019. URL: https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/28108/KANWAR-THESIS-2019.pdf
14. Zhengyao Jiang, Jinjun Liang.: «Cryptocurrency Portfolio Management with Deep Reinforcement Learning», Xi'an Jiaotong-Liverpool University, 2017. URL: https://arxiv.org/abs/1612.01277
15. Ricard Durall.: «Asset Allocation: From Markowitz to Deep Reinforcement Learning», Open University of Catalonia, Jul 2022. URL: https://arxiv.org/abs/2208.07158
16. Jiwon Kim, Moon-Ju Kang, KangHun Lee, HyungJun Moon, Bo-Kwan Jeon.: «Deep Reinforcement Learning for Asset Allocation: Reward Clipping», SK Inc. (SK C&C), Jan 2023. URL: https://arxiv.org/abs/2301.05300
17. Pengqian Yu, Joon Sern Lee, Ilya Kulyatin, Zekun Shi, Sakyasingha Dasgupta.: «Model-based Deep Reinforcement Learning for Dynamic Portfolio Optimization», Neuri PTE LTD, Jan 2019. URL: https://arxiv.org/abs/1901.08740

18. Gang Huang, Xiaohua Zhou, Qingyang Song.: «Deep reinforcement learning for portfolio management», Chongqing University, Dec 2020. URL: https://arxiv.org/abs/2012.13773
19. Qing Yang Eddy Lim, Qi Cao, Chai Quek.: «Dynamic portfolio rebalancing through reinforcement learning», Neural Computing and Applications Journal 34, pages 7125–7139 (2022), 2022. URL: https://link.springer.com/article/10.1007/s00521-021-06853-3
20. Zhengyao Jiang, Dixing Xu, Jinjun Liang.: «A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem», Xi'an Jiaotong-Liverpool University, 2017. URL: https://arxiv.org/abs/1706.10059
21. Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto4, Maximilian Ernestus, Noah Dormann: «Stable-Baselines3: Reliable Reinforcement Learning Implementations», Journal of Machine Learning Research 22 (2021) 1-8, 2021. URL: https://jmlr.org/papers/volume22/20-1364/20-1364.pdf