

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

Отчет о программном проекте

на тему: _____ Рекомендательные системы – анализ портфеля ценных бумаг и/или криптовалют _____

(промежуточный, этап 1)

Выполнил:

Студент группы БПМИ 221	 _____	Г.А. Бельковец И.О.Фамилия
	Подпись	

Студент группы БПМИ 221	 _____	П.В. Перов И.О.Фамилия
	Подпись	

Студент группы БПМИ 221	 _____	А.А. Хасанов И.О.Фамилия
	Подпись	

Студент группы БПМИ 221	 _____	К.И. Миннибаев И.О.Фамилия
	Подпись	

11.01.2025

Дата

Принял:


Руководитель проекта	Илья Викторович Мунерман Имя, Отчество, Фамилия
----------------------	----------------------------------------------------

доцент, к.э.н.

Должность, ученое звание

Интерфакс

Место работы (Компания или подразделение НИУ ВШЭ)

Дата проверки	?? ?? 2025	?	
		Оценка (по 10-ти балльной шкале)	Подпись

Москва 2025

Содержание

1	Введение	3
2	Актуальность	4
3	Цели, задачи, план работы	5
4	Описание функциональных и нефункциональных требований к программному проекту	7
5	Коннектор к бирже (Перов П.В.)	8
6	Оптимизация портфеля с теоретической точки зрения (Бельковец Г.А.)	9
6.1	Доходность, риск, ковариация активов	9
6.2	Портфельная теория Марковица	10
6.3	Портфельная теория Блэка-Литтермана	11
6.4	Применение на практике	12
7	Бэктестинг и оценка качества портфелей (Хасанов А.А.)	15
7.1	Что такое бэктэстинг и зачем он нужен?	15
7.2	С какими проблемами сталкиваются при бэктестинге?	15
7.3	Построение бэктестера и эксперименты	16
8	Извлечение сигналов и прогнозирование цен акций (Миннибаев К.И., Перов П.В.)	18
8.1	Введение в анализ сигналов	18
8.2	Парсинг сигналов	18
8.3	Исследование влияния новостей	19

1 Введение

Оптимизация инвестиционного портфеля — это задача, которая корнями уходит в древние времена, когда люди впервые столкнулись с необходимостью управлять ресурсами и столкнулись с задачей баланса двух противоположных желаний: максимизации прибыли и минимизации рисков.

1. В древние времена эта задача в основном решалась с помощью интуиции. Ключевой метод — диверсификация — использовался в Вавилоне и древнем Египте торговцами, которые распределяли свой товар на несколько кораблей, чтобы минимизировать убытки в случае потери одного из них.
2. В средневековые управление торговыми рисками вышло на новый уровень. Возрождение торговли и рост банковского дела в Италии привели к использованию более сложных методов оценки рисков. Купцы Флоренции, Генуи и Венеции разработали системы страхования грузов, а также начали использовать финансовые контракты для снижения рисков.
3. В новое время задача стала обретать научный подход. В 17 веке работы Паскаля и Ферма в области теории вероятностей заложили основы для количественной оценки рисков. Эти идеи позже были применены в финансах для анализа доходностей и вероятностей убытков.
4. В 18 веке в Лондоне начали появляться компании, предлагающие страхование рисков, связанных с морскими перевозками. Эти компании активно использовали статистику для оценки вероятности убытков.
5. В 19 веке математики и экономисты впервые использовали математический анализ для моделирования цен акций, заложив основы для дальнейших финансовых исследований.

Теперь обратимся к современным реалиям. В современном мире финансов задача оптимизации инвестиционного портфеля занимает центральное место в управлении активами. Она заключается в нахождении баланса между доходностью и риском для достижения финансовых целей инвестора, учитывая его риск профиль и ожидаемую доходность. Однако изменчивость финансовых рынков и растущая сложность данных требуют новых подходов к решению данной проблемы.

Традиционные методы, такие как модель Марковица и модель Блэка-Литтермана, предлагают устойчивую основу для построения портфеля, однако их эффективность ограничивается в условиях высокой волатильности и при анализе нетрадиционных данных. Современные технологии, включая методы машинного обучения и обработки естественного языка, открывают новые возможности для учета широкого спектра рыночных сигналов, что делает управление портфелем более гибким и адаптивным.

Рекомендательные системы, изначально разработанные для персонализации контента в таких областях, как электронная коммерция и социальные сети, находят применение и в управлении инвестиционными портфелями. Использование подобных систем позволяет не только учитывать разнообразные источники данных, но и адаптироваться к изменяющимся рыночным условиям в режиме реального времени.

Целью настоящего исследования является изучение методов оптимизации инвестиционного портфеля с акцентом на использование новых рыночных сигналов и современных моделей машинного обучения. Работа сосредоточена на анализе комбинации традиционных и неочевидных данных, таких как новости, квартальные отчёты и судебные решения, с целью повышения эффективности управления портфелем.

2 Актуальность

В настоящее время финансовые технологии сделали брокерские услуги крайне доступными. Любой желающий может установить приложение любимого банка к себе на телефон и начать инвестировать, торговать на бирже используя всё разнообразие финансовых инструментов.

Современные технологии и скорость компьютеров нашего времени развивается очень быстро. В наши дни принятие решений о покупке или продаже того или иного актива принимаются за наносекунды в фондах высокочастотной торговли с помощью новейшего оборудования и сложных алгоритмов.

Рынки капитала становятся всё более динамичными и сложными по вышеописанным причинам. В этих условиях инвесторы сталкиваются с рядом вызовов, включая необходимость быстрой обработки больших объемов данных и своевременной адаптации стратегий к изменениям в рыночной среде.

Применение методов машинного обучения и анализа больших данных в управлении инвестиционными портфелями позволяет решать эти вызовы. Ключевая актуальность заключается в возможности использовать ранее недоступные источники данных, такие как текстовая информация из новостей и отчётов, а также прогнозировать влияние таких факторов на волатильность и доходность активов.

Кроме того, интеграция рекомендательных систем в процесс принятия инвестиционных решений позволяет учитывать персонализированные предпочтения инвесторов и адаптировать стратегии под их цели. Это особенно важно в условиях растущей конкуренции на финансовых рынках, где успешное управление портфелем требует инновационных решений.

Таким образом, исследование, посвященное разработке методов оптимизации портфеля с учетом новых рыночных сигналов и применением современных технологий, имеет не только теоретическое, но и практическое значение.

3 Цели, задачи, план работы

Цель нашей работы – исследовать четко постановку задачи оптимизации инвестиционного портфеля и существующие решения этой задачи, исследовать новые рыночные сигналы и их влияние на оптимизацию инвестиционного портфеля, а также изучить методы построения моделей для принятия инвестиционных решений. Дополнительно мы ставим перед собой задачи разработки инструментов для анализа рыночных данных и оценки моделей в условиях реального времени.

В качестве первого шага мы сосредотачиваем внимание на традиционных финансовых рынках, так как о них опубликовано больше исследований, и для анализа доступны более очевидные и проверенные сигналы (например, судебные решения, новости, квартальные отчетности и т.д.). Это также позволяет использовать преимущественно API Московской биржи для работы с данными.

В рамках исследования будут решаться следующие задачи:

1. Постановка задачи оптимизации инвестиционного портфеля, исследование существующих подходов к решению, включая использование классических методов, таких как модель Марковица (Mean-Variance Optimization), модель Блэка-Литтермана и более современных подходов, таких как модели классического машинного обучения (линейные регрессии, градиентные бустинги и решающие деревья) и модели глубинного обучения (нейронные сети и их разновидности, например, рекуррентные).
2. Поиск и анализ новых рыночных сигналов, включая комбинацию традиционных и неочевидных источников данных, таких как данные о судебных решениях и различные финансовые показатели фундаментального анализа. Мы также планируем использовать Natural Language Processing (NLP) для анализа текстов и новостей.
3. Реализация различных подходов к оптимизации портфеля, включая вышеописанные классические подходы и более современные методы машинного обучения с использованием внешних факторов как признаков для обучения.
4. Изучение взаимосвязи между рыночными сигналами и волатильностью портфеля с целью дальнейшей динамической адаптации портфеля в ответ на изменения в рыночной среде (ребалансировка весов финансовых активов в портфеле).
5. Оценка устойчивости предложенных моделей к изменяющимся рыночным условиям с помощью исторических данных.

Для достижения целей работы разработан следующий план:

1. Изучить литературу на тему решения задачи оптимизации инвестиционного портфеля, включая методы классического портфельного анализа, а также современные методы, использующие машинное обучение для предсказания рыночных движений.
2. Создать коннектор для взаимодействия с биржей через API, с возможностью получения данных о ценах активов, объемах торгов и новостных потоках.
3. Реализовать систему бэктестинга для оценки эффективности предложенных методов, используя исторические данные с биржи. Мы будем использовать подходы, такие как кросс-валидация для моделей машинного обучения, а также методы подбора параметров (например, Grid Search и Random Search) для поиска оптимальных гиперпараметров.
4. Разработать методику выявления наиболее значимых сигналов и их ранжирования, исследуя качество моделей при использовании разного подмножества признаков.
5. Исследовать влияние различных сигналов на прогнозы и оптимизацию портфеля, изучая поведение моделей в различные моменты времени на исторических данных.
6. Оценить производительность моделей при изменении гиперпараметров, добавлении шума в данные и изменении рыночных режимов, используя написанные модули для бэктестинга.
7. Внедрить систему оценки рисков на основе предложенных сигналов, используя классические метрики (Value at Risk, Conditional Value at Risk, и т.д.)

При успешном выполнении вышеописанных задач, в работе также будут рассмотрены дополнительные подходы, такие как:

1. Создание инструментов визуализации и анализа полученных результатов для интерпретации моделей.
2. Разработка методов автоматического поиска и адаптации гиперпараметров для улучшения производительности моделей.

Мы планируем построить гибкий и масштабируемый подход, который позволит интегрировать результаты работы в реальные системы управления портфелями.

4 Описание функциональных и нефункциональных требований к программному проекту

В конечном итоге ожидается реализация различных алгоритмов оптимизации инвестиционного портфеля с учётом новых изученных сигналов. Работа будет происходить на языке Python, а также для удобства экспериментов будет использоваться Jupyter Notebook. Совместная работа будет происходить на Gitlab, в качестве менеджера зависимостей в проекте - *poetry*.

5 Коннектор к бирже (Перов П.В.)

Создадим базовый интерфейс *ExchangesInterface*, от которого впоследствии будем наследовать все подключения к биржам:

```
1 class ExchangesInterface(object):
2     apiKey: str | None
3     __metaclass__ = ABCMeta
4
5     @property
6     def exchangeName(self):
7         return self._exchangeName
8
9     @exchangeName.setter
10    def exchangeName(self, value):
11        self._exchangeName = value
12
13    @abstractmethod
14    def __init__(self, apiKey: Optional[str]):
15        self.apiKey = apiKey
```

Для получения исторических данных Московской Биржи будем использовать библиотеку **MoexAlgo**. Реализуем методы получения свеч и разных статистик по тикеру за указанный период:

```
1 class MOEX(ExchangesInterface):
2     def __init__(self, apiKey: Optional[str] = None):
3         session.TOKEN = apiKey
4         super().__init__(apiKey)
5
6     def error_handler(method):
7         def wrapper(self, *args, **kwargs):
8             try:
9                 return method(self, *args, **kwargs)
10            except Exception as e:
11                error_message = f"Error occurred: {e}"
12                raise MOEXAlgoPackException(error_message)
13
14        return wrapper
15
16    @error_handler
17    def get_candles(
18        self, ticker: str, start: str, end: str, period: MOEXTimePeriods
19    ) -> pd.DataFrame:
20        stock = Ticker(ticker)
21        candles_data = stock.candles(
22            start=start, end=end, period=period.value, use_dataframe=True
23        )
24        return candles_data
25
26    @error_handler
27    def get_tradestats(self, ticker: str, start: str, end: str) -> pd.DataFrame:
28        stock = Ticker(ticker)
29        tradestats_data = stock.tradestats(start=start, end=end, use_dataframe=True)
30        return tradestats_data
31
32    @error_handler
33    def get_orderstats(self, ticker: str, start: str, end: str) -> pd.DataFrame:
34        stock = Ticker(ticker)
35        orderstats_data = stock.orderstats(start=start, end=end, use_dataframe=True)
36        return orderstats_data
37
38    @error_handler
39    def get_obstats(self, ticker: str, start: str, end: str) -> pd.DataFrame:
40        stock = Ticker(ticker)
41        obstats_data = stock.obstats(start=start, end=end, use_dataframe=True)
42        return obstats_data
43
44    @error_handler
45    def get_tickers(self) -> list[str]:
46        stocks = Market("stocks")
47        return stocks.tickers()["ticker"].tolist()
```


6 Оптимизация портфеля с теоретической точки зрения (Бельковец Г.А.)

Давайте введем теоретический минимум, чтобы понять, что вообще такое оптимальный портфель, и как его оптимизировать. Далее перейдем к основным современным теориям оптимизации портфеля и рассмотрим их на практике.

6.1 Доходность, риск, ковариация активов

1. Доходность актива

Доходность актива — это показатель, который отражает, насколько выгодно было вложение в данный актив за определенный период времени. Доходность может быть рассчитана как для отдельного актива, так и для портфеля в целом. Формула доходности актива:

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

где: r_t - непосредственно доходность за определенный период, P_t - цена актива на конец периода, P_{t-1} - цена актива на начало периода. Для упрощения мы не рассматриваем возможные выплаты дивидендов за этот период и при надобности добавим их позже.

Средняя доходность актива считается как просто усреднение доходности за несколько периодов и обозначается буквой μ

2. Риск актива

Чтобы посчитать дисперсию актива, надо воспользоваться стандартной формулой дисперсии случайной величины из теории вероятности:

$$\sigma^2 = \frac{\sum_{t=1}^m (r_t - \mu)^2}{m - 1} \quad (2)$$

где: σ^2 - дисперсия рассматриваемого актива, r_t - доходность актива за определенный период, μ - средняя доходность, m - количество рассматриваемых периодов

Однако инвесторами чаще рассматривается стандартное отклонение как мера риска актива:

$$\sigma = \sqrt{\sigma^2} \quad (3)$$

В целом очевидно, что нет особо разницы что использовать, так как и для дисперсии и для стандартного отклонения выполнен тот факт, что чем эта мера больше, тем более рисковый актив (и наоборот).

3. Ковариация активов

Ковариация - показатель между парой активов, который измеряет степень совместного изменения доходностей двух активов. Он показывает, насколько доходности двух активов склонны двигаться в одном направлении (положительная ковариация) или в противоположных направлениях (отрицательная ковариация). Если же ковариация близка к нулю, то скорее всего между активами отсутствует какая-либо зависимость. Формула ковариации выглядит следующим образом:

$$\sigma_{ij} = \frac{1}{m} \sum_{t=1}^m (r_t^i - \mu_i)(r_t^j - \mu_j) \quad (4)$$

где: σ_{ij} - ковариация между парой активов i, j , r_t^i - доходность актива с индексом i за период времени t , μ_i - средняя доходность актива i

Когда мы работаем с задачей оптимизации портфеля, то говорим чаще всего о каком-либо множестве из n активов, для которого удобно ввести матрицу ковариаций:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix} \quad (5)$$

6.2 Портфельная теория Марковица

До публикации Гарри Марковицем своей портфельной теории [1] инвесторы сильно фокусировались на оценке ожидаемой доходности и рисков одного конкретного актива, пытаясь извлечь максимальную выгоду только из него. Однако Марковиц произвел в каком-то роде революцию, формализовав свой подход, в основе которого лежит тот факт, что акции и любые активы можно представить как случайные величины со своим матожиданием и стандартным отклонением, и что комбинация активов с разными рисками и доходностями вместе может дать гораздо больше выгоды, чем по отдельности.

Будем работать в терминах, описанных выше: имеем n различных активов, их доходности $\mu_1, \mu_2 \dots$ (или просто все вместе в одной матрице - μ), а также матрицу их ковариаций Σ . Для нас портфель - это просто распределение весов по разным активам, то есть вектор из n элементов, который мы назовем w .

Определим как $\mu_P = w^T \mu$ - доходность нашего портфеля, по сути просто средневзвешенные доходности всех активов которые мы взяли. Также определим $\sigma_P = \sqrt{w^T \Sigma w}$ - риск нашего портфеля. Обе эти формулы абсолютно понятным образом возникают из матожидания и стандартного отклонения суммы случайных величин.

Далее мы готовы к формулировке задачи оптимизации. Чаще всего она встречается в 2 вариациях:

1. Максимизация доходности при заданном уровне риска

$$\sum_{i=1}^n w_i = 1 \quad (6)$$

$$\sigma_P = \sqrt{w^T \Sigma w} \leq C \quad (7)$$

$$\mu_P = w^T \mu \rightarrow \max \quad (8)$$

2. Минимизация риска при заданном уровне доходности

$$\sum_{i=1}^n w_i = 1 \quad (9)$$

$$\mu_P = w^T \mu = R \quad (10)$$

$$\sigma_P = \sqrt{w^T \Sigma w} \rightarrow \min \quad (11)$$

Эти задачи неразрывно связаны друг с другом через кривую эффективных портфелей ("Efficient frontier") - если на двумерной плоскости изобразить точками (доходность, риск) все возможные портфели, то на плоскости образуется кривая наиболее эффективных портфелей (с лучшим риском при заданной доходности).

Классическая задача Марковица — это задача квадратичной оптимизации, так как целевая функция (риск портфеля) является квадратичной, а ограничения — линейными. Эта задача имеет множество решений (численные методы, методы квадратичной оптимизации), в нашей работе я не буду углубляться в них.

Также отмечу, что помимо двух классических формулировок (описанных выше) существует множество вариаций, которые часто используются на практике:

1. Наличие безрискового актива

Безрисковый актив (которым на практике может являться банковский вклад или облигации федерального займа) меняет условия оптимизационной задачи - в наше множество просто добавится актив со стандартным отклонением 0, который также очевидно некоррелирован ни с какими другими активами. Решение задачи немного поменяется, но не станет более сложным.

2. Ограничение на веса

Потенциально инвестор может захотеть явно контролировать степень диверсификации нашего портфеля с помощью ограничения на веса, например, установить $w_i \leq 0.2$, то есть на каждый из активов нельзя будет выделить более 20% капитала. Эта задача также несложно решается, так как описанные ограничения также являются линейными функциями, и сильно концепция решения не поменяется

6.3 Портфельная теория Блэка-Литтермана

Модель Блэка-Литтермана — это модель построения оптимального портфеля, используемая для создания сбалансированных инвестиционных портфелей с учетом не только исторической статистики активов, но и субъективных взглядов инвесторов на будущую цену активов. Она была разработана Фишером Блэком и Робертом Литтерманом в 1990-х годах как расширение классической модели Марковица. Основная проблема в модели Марковица заключена в ожидаемых доходностях. Если рассчитывать их на исторических данных, то точность будем плачевно, это показывают многочисленные эксперименты на практике.

1. В модели Блэка-Литтермана отправной точкой является предположение, что текущие рыночные цены отражают равновесные ожидания всех участников рынка. Это означает, что рыночные веса активов в рыночном портфеле (например, в индексе Московской Биржи для российского рынка, или в индексе S&P500 для американского) считаются оптимальными.

Формула рыночного равновесия в модели Блэка-Литтермана выглядит следующим образом:

$$\mu_{eq} = \lambda \Sigma w_{market} \quad (12)$$

где: μ_{eq} — вектор ожидаемых доходностей в состоянии равновесия, λ — коэффициент неприятия риска инвестора, Σ — ковариационная матрица доходностей активов, w_{market} — веса активов в рыночном портфеле

2. Следующий шаг - выражение субъективных взглядов инвестора. Инвестор может выразить свои собственные взгляды на ожидаемую доходность активов. Они могут быть как абсолютными (например, актив вырастет на $x\%$), так и относительными (например, актив 1 обойдет по доходности актив 2 на $y\%$). Вообще таким образом можно будет выразить любой взгляд на любую линейную комбинацию из доходностей активов.

Взгляды формализуются в виде линейных уравнений:

$$P\mu = Q + \varepsilon \quad (13)$$

где: P - матрица, которая связывает активы с взглядами, Q - вектор значений, отражающих ожидания инвестора, ε - вектор ошибок, который учитывает неопределенность взглядов. Например, пусть мы работаем с 3 активами, и утверждаем, что 1 актив вырастет в цене на 5%, а 2 актив обгонит 3 актив на 10%. Тогда выражение примет следующий вид:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.10 \end{pmatrix} + \varepsilon \quad (14)$$

3. Объединение рыночного равновесия и взглядов происходит с помощью байесовского подхода. Результирующие ожидаемые доходности μ_{BL} вычисляются как взвешенное среднее между равновесными ожиданиями и взглядами инвестора:

$$\mu_{BL} = [(\tau \Sigma)^{-1} + P^T \Omega P]^{-1} [(\tau \Sigma)^{-1} \mu_{eq} + P^T \Omega^{-1} Q] \quad (15)$$

где: τ - скаляр, регулирующий вес, который придается априорным ожиданиям, Σ - ковариационная матрица доходностей активов, P - матрица взглядов, Ω - ковариационная матрица неопределенности взглядов, Q - вектор ожидаемых значений для взглядов

6.4 Применение на практике

Для наглядности рассмотрим все вышеописанные факты на реальных данных, взятых с Московской Биржи (с помощью библиотеки из прошлой главы).

Для первоначальных тестов мною были отобраны 3 сектора российской экономики и по 3 больших компании в каждом секторе, в качестве первоначальной выборки. Главным критерием была высокая ликвидность акций на бирже (то есть это должны быть известные компании, акции которых можно в любой момент купить или продать у брокера), а также многолетняя история торгов на бирже для более точных расчетов (от 6 лет). По итогу выбор пал на следующие компании:

1. Металлургия: Северсталь, НЛМК, ММК
2. Нефть: Лукойл, Роснефть, Татнефть
3. Финансовые технологии: Сбербанк, ВТБ, Московская биржа

График изменения цены акций этих компаний вы можете наблюдать на рисунке 1. Отмечу, что на графиках ниже будут рассмотрены только данные по ценам до 2023 года. Это сделано для того, чтобы проводить дальнейшие тестирования на данных 2024 года.

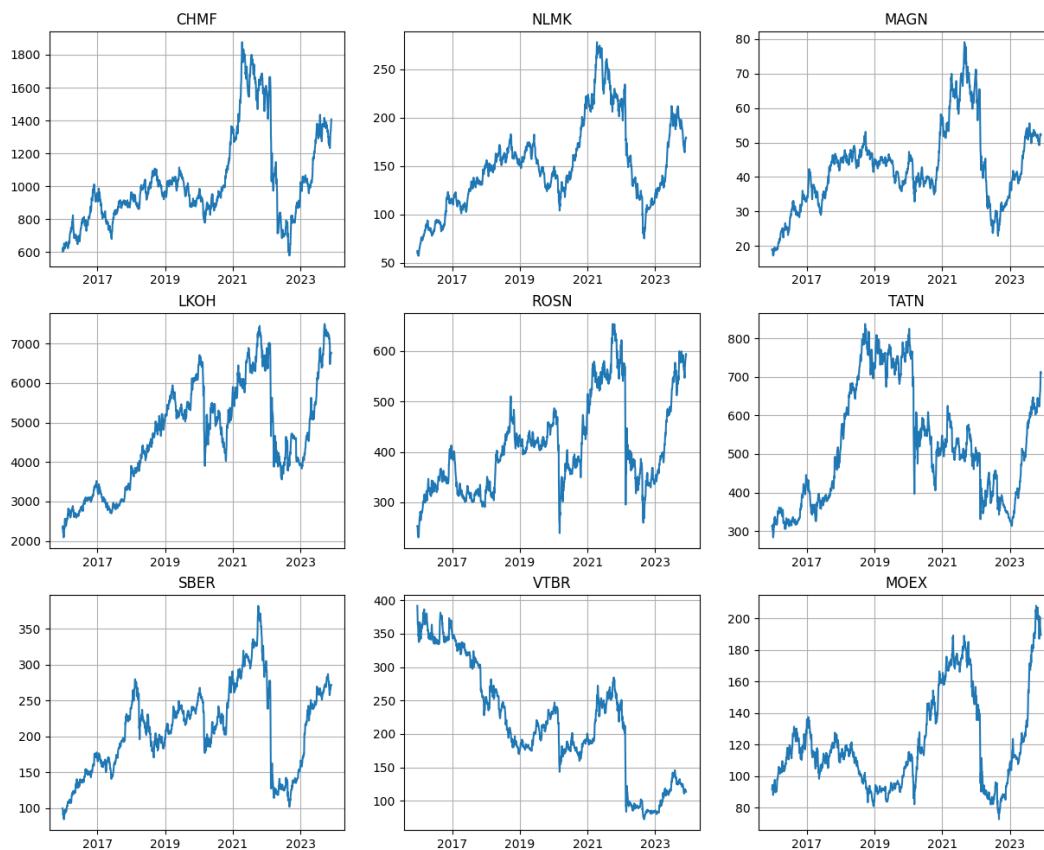


Рис. 1: График изменения цен 9 выбранных компаний

Посчитаем теперь все описанные выше термины на практике. Для расчета буду использоваться известную библиотеку PyPortfolioOpt, в которой реализованы все нужные нам вычисления.

Сложив все цены акций в один датафрейм `prices`, посчитаем среднюю доходность и их ковариационную матрицу с помощью соответствующих функций из упомянутой выше библиотеки:

```
1 from pypfopt import expected_returns, risk_models
2
3 mu = expected_returns.mean_historical_return(prices)
4 sigma = risk_models.sample_cov(prices)
```

Получились следующие результаты: если говорить про среднюю годовую доходность μ , то имеем такие числа:

Тикер на московской бирже	Средняя годовая доходность
CHMF	0.114
NLMK	0.144
MAGN	0.137
LKOH	0.144
ROSN	0.114
TATN	0.111
SBER	0.136
VTBR	-0.143
MOEX	0.095

В фаворитах по доходности Лукойл и НЛМК, в худшую же сторону явно выделяется ВТБ. Далее посмотрим на отношение между активами. Выше мы посчитали матрицу ковариаций, но визуально проще анализировать будет немного другую величину - корреляцию, которая по сути является просто отнормированной ковариацией в отрезок $[-1, 1]$. Благо функционал PyPortfolioOpt позволяет несложно визуализировать требуемую матрицу:

```
1 from pypfopt import plotting
2 plotting.plot_covariance(sigma, plot_correlation=True)
```

Полученные результаты можно увидеть на рисунке 2.

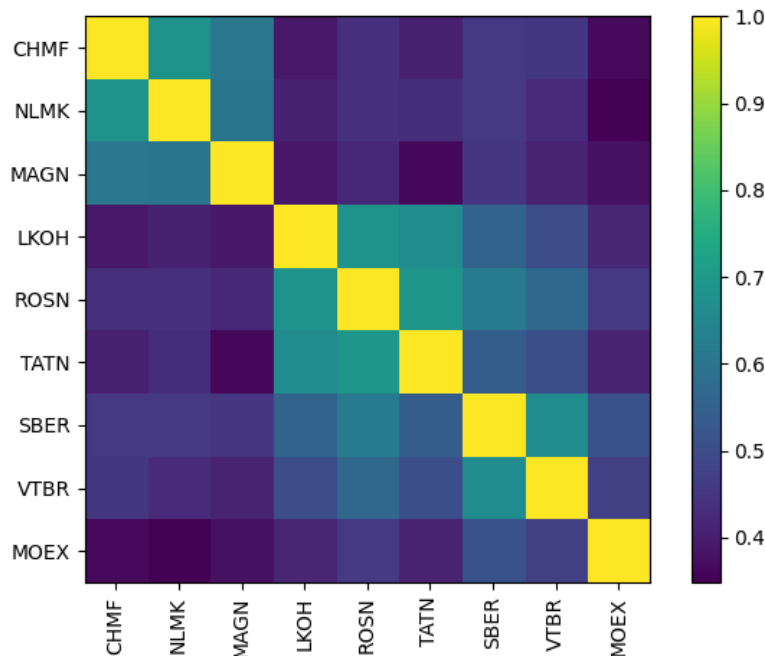


Рис. 2: Матрица корреляций между выбранными акциями

Отмечу некоторые интересные наблюдения, видимые по матрице корреляций. Самое бросающееся в глаза - четкое разделение компаний на сектора. Голубой квадрат 3 на 3 сверху слева - это компании из сектора металлургии, попарные корреляции между ними находятся на уровне 0.7, что показывает высокую степень линейной зависимости между доходностями этих активов. По центру квадрат 3 на 3 - компании из сектора нефти, между ними аналогично высокая степень корреляции. Справа же снизу цвета более блеклые, но все равно видно что корреляция между Сбером и ВТБ высокая, выделяется немного Московская биржа - это, конечно, тоже финансовая компания, но все-таки не банк, как первые две.

Перейдем теперь непосредственно к построению портфеля, так как нужные нам вещи мы уже посчитали выше. Все эффективные портфели строятся в PyPortfolioOpt через функционал класса EfficientFrontier.

```
1 from pypfopt import EfficientFrontier
2 ef = EfficientFrontier(mu, sigma)
3 w = ef.min_volatility()
4 res = ef.portfolio_performance(verbose=True)
```

Выше построили портфель с минимальным возможным риском с помощью функции `min_volatility`. PyPortfolioOpt оценивает потенциальные параметры нашего портфеля следующим образом:

1. Expected annual return: 11.2%
2. Annual volatility: 22.6%
3. Sharpe Ratio: 0.50

Важно отметить, что это не реальные метрики качества нашего портфеля, а лишь их оценка по историческим данным. По ним не стоит пока слишком серьезно рассуждать о качестве портфеля, это лучше делать с помощью бэктестинга на данных о следующих годах, о чем будет рассказано в следующих главах.

Полученные веса портфеля изображены на рисунке 3.

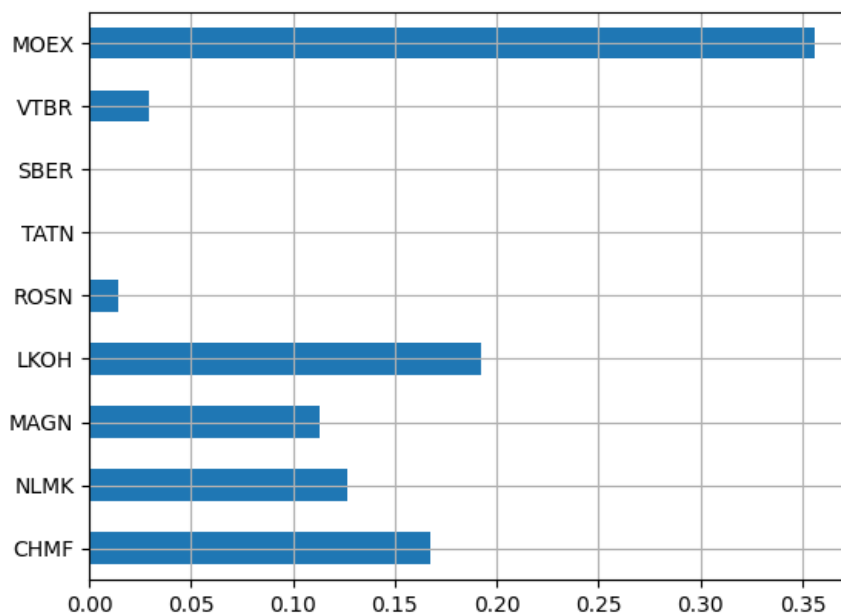


Рис. 3: Веса портфеля, обладающего минимальной волатильностью

7 Бэктестинг и оценка качества портфелей (Хасанов А.А.)

7.1 Что такое бэктэстинг и зачем он нужен?

Бэктестинг - это метод, который позволяет трейдеру сделать оценку того, насколько эффективной могла бы быть его торговая стратегия в прошлом. Этот подход широко применяется в алгоритмической торговле, так как при ее использовании нам важно иметь представление о том, как та или иная стратегия поведет себя на реальных данных. Затем на основе проведенного исследования посмотреть на различные ключевые показатели торговой стратегии: прибыль/убыток, технические индикаторы и сравнительный анализ по отношению к другим известным стратегиям.

Процесс бэктестинга можно разделить на несколько логических этапов:

1. Сбор данных

Для начала нужно получить качественные исторические данные: цены, объемы торгов, спреда и другие показатели. Чем точнее данные, тем достовернее будут результаты. Обычно сами биржи предоставляют удобные инструменты для доступа к информации о торгах, совершаемых у них. Мы же, например, как это было упомянуто выше, будем брать информацию с московской биржи.

2. Разработка стратегии

На этом этапе создаются правила и алгоритмы, которые определяют, когда открывать и закрывать сделки.

3. Тестирование

Стратегия применяется к историческим данным, чтобы проверить, как она сработала бы в прошлом.

4. Анализ

После тестирования анализируются результаты: прибыль, убытки, риски и другие параметры. Это помогает понять, стоит ли использовать стратегию в реальной торговле.

7.2 С какими проблемами сталкиваются при бэктестинге?

С первого взгляда процесс может показаться довольно простым, однако, как это часто бывает, без трудностей не обойтись:

1. Неполнота данных

Исторические данные могут быть неполными или содержать ошибки. Например, отсутствие информации о спредах или комиссиях может исказить результаты. Например, в ходе одного из экспериментов, который будет описан дальше, мы не имели данных о некоторых тикерах (в таком случае в таблицу вставляют специальное значение NaN).

2. Игнорирование реальных условий

К таким, например, относятся задержки в исполнении ордеров, целочисленные объемы позиций, изменение состава индекса и др. Здесь, кстати, стоит сделать важное замечание, что для упрощения бэктестинга считают, что объемы сделок не столь значительный, чтобы повлиять на цены в будущем. Мы тоже будем пользоваться этим предположением.

3. Вычислительные сложности

Сейчас этот пункт особенно актуален, т.к. бурно развивается сфера ML, с помощью которого люди пытаются составлять прогнозы временных рядов.

4. Устойчивость (Robustness)

Стоит учитывать разные параметры. Так, например, хочется знать как наши результаты будут зависеть от того, запустим ли мы стратегию в понедельник или во вторник

5. Переобучение

Здесь мы подразумеваем то, что важно следить за тем, чтобы при построении стратегии мы не использовали данные из будущего.

7.3 Построение бэктестера и эксперименты

Для бэктестинга наших портфелей будем использовать фреймворк `bt`. Конечно, существуют и другие как, например, `Backtrader`, `PyAlgoTrade` или `Zipline`. Однако, в `bt` есть вся необходимая для нас логика и кроме того он более удобен для портфельных оптимизаций.

Давайте немного расскажем о том, как фреймворк работает и какие сущности использует.

Глобально здесь используется только две сущности: стратегии (наследуются от класса `bt.core.Algo`) и активы (наследуются от класса `bt.core.SecurityBase`). Данные предоставляются в виде `pandas.DataFrame` табличек. Все устроено так, чтобы в дальнейшем можно было максимально переиспользовать код.

Важной особенностью является древовидная структура:

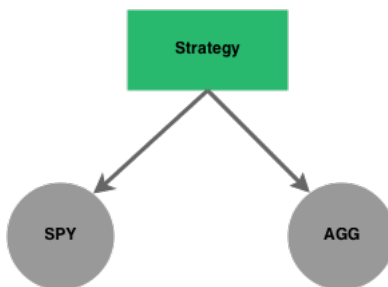


Рис. 4: Наша стратегия и два актива, которыми она оперирует (наш портфель)

Сверху можем видеть простейшую модель нашей стратегии. А теперь давайте представим, что у нас есть три актива (`SPY`, `EEM`, `AGG` на рис. 5) и при наступлении каких-то условий мы бы хотели по особенному обрабатывать первые два. Тогда мы просто создаем новую стратегию и объявляем ее ребенком первой и теперь при обработке первых двух тикеров наша стратегия после материнской стратегии будем применять дочернюю стратегию.

Отметим, что данная ситуация вполне часто встречается на практике, например, когда какое-то время мы пользовались стратегией 1, а затем в ходе исследований мы поняли, что будет оптимальнее будет применять стратегию 2 для каких-то элементов нашего портфеля. Так вот благодаря данной структуре нам не нужно будет переписывать весь код. Мы допишем логику для стратегии 2, объявим ее ребенком изначальной стратегии.

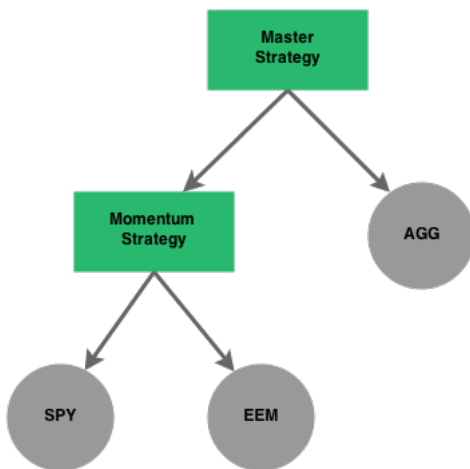


Рис. 5: Новая стратегия, которая теперь применяет Momentum Strategy для первых двух активов

Теперь давайте реализуем простейшую стратегию "Buy and Hold" (только в нашем случае, конечно, могут быть отрицательные веса и тогда формально мы будем открывать не длинную, а короткую позицию).


```

1 def MakeStatic(name, prices, weights):
2     """
3     Creates a static portfolio with given weights.
4
5     Args:
6     * name (str): a name of strategy
7     * prices (pd.DataFrame): data for backtest
8     * weights (list): weights for balancing the portfolio
9
10    Return:
11    * backtest (bt.Backtest): a backtest instance object
12
13    """
14
15    weights = pd.Series(weights, index = prices.columns)
16    weighSpecifiedAlgo = bt.algos.WeighSpecified(**weights)
17
18    strat = bt.Strategy(name,
19        [
20            bt.algos.RunOnce(),
21            weighSpecifiedAlgo,
22            bt.algos.Rebalance()
23        ]
24    )
25
26    backtest = bt.Backtest(
27        strat,
28        prices,
29        integer_positions=False
30    )
31
32    return backtest

```

Теперь давайте протестируем данную стратегию на данных за 2024 год на портфеле составленном в предыдущей главе (топ-3 компании из топ-3 сфер в РФ) и сравним ее, например, с лидерами каждой из трех отраслей и MOEX (т.к. данный тикер все же немного выделяется среди остальных).

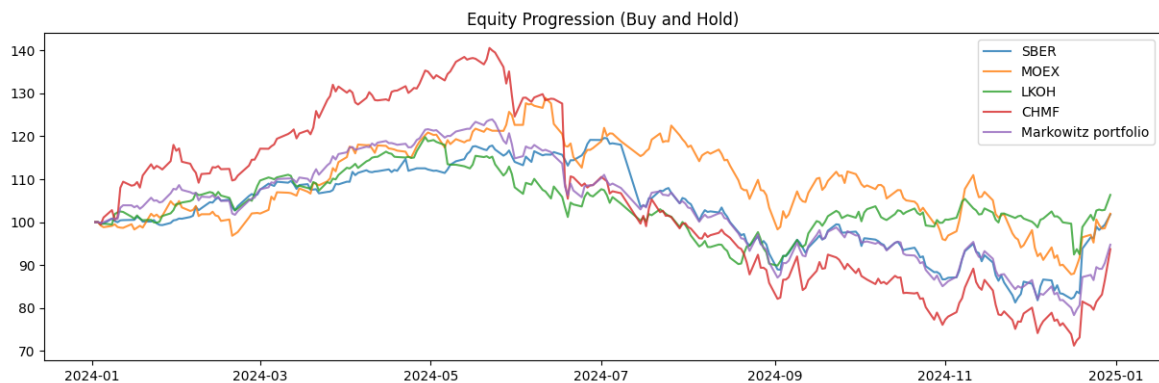


Рис. 6: Прогресс портфелей из стартовой позиции в 100 условных единиц

На рис. 6 видим, что на дистанции в 1 год портфель не показал себя хорошо, несмотря на то что . Выгоднее было бы купить, скажем, акций "Лукойл".

8 Извлечение сигналов и прогнозирование цен акций (Миннибаев К.И., Перов П.В.)

8.1 Введение в анализ сигналов

В современном финансовом мире цена акций формируется под влиянием множества факторов, среди которых ключевую роль играют новости и различные сигналы, поступающие с рынка. Инвесторы, аналитики и трейдеры активно используют информационные потоки для принятия решений, что делает анализ влияния новостей на стоимость активов крайне актуальной задачей.

Одной из основных причин, по которой изучение данного явления имеет практическую ценность, является высокая скорость распространения информации. Новости о макроэкономических показателях, корпоративные отчёты, заявления руководителей компаний и даже социальные медиа могут вызывать значительные колебания на фондовом рынке. Эффективность рынка предполагает, что цена актива мгновенно учитывает всю доступную информацию, однако на практике наблюдаются задержки и дисбалансы, создающие возможности для предсказания движений цены.

8.2 Парсинг сигналов

В качестве источников сигналов для начала возьмём известные новостные каналы в телеграмм (РБК, РИА Новости и т.д.). Для дальнейшего анализа мы провели парсинг всех постов с 2016 по 2025 год.

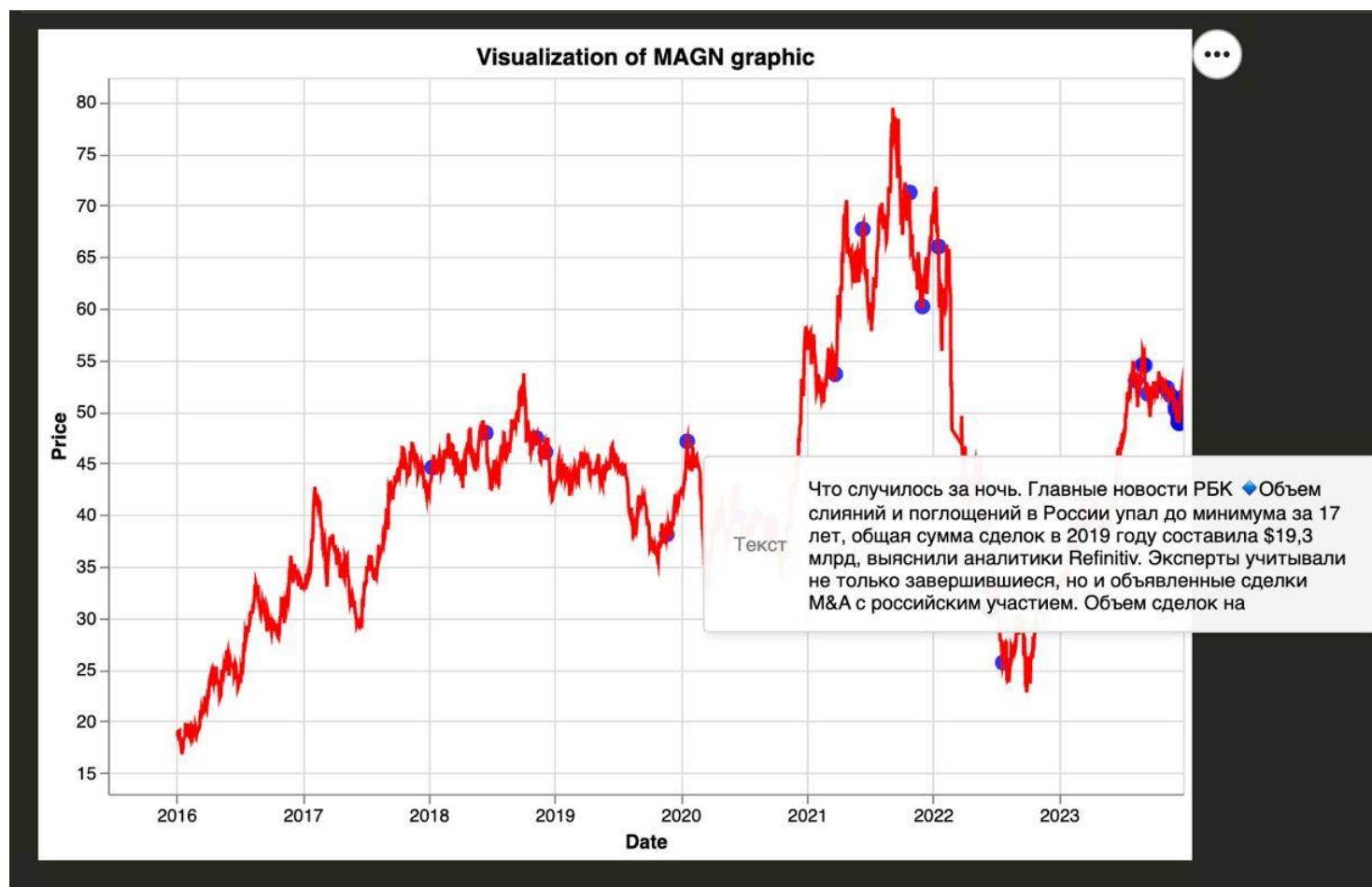


Рис. 7: График цены акции "Магнит" с нанесёнными на график новостями

Для поиска новостей с упоминанием компании мы использовали регулярные выражения и лемматизацию текста новости.

8.3 Исследование влияния новостей

Мы хотим выявить зависимость между новостями и графиком цены акций, для этого соберём датасет с новостями и изменением цены в процентах.

Список литературы

- [1] H.M.Markowitz. *Portfolio Selection*. 1952.