

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ СЕТЕЙ ЧАСТНЫХ  
КОРРЕЛЯЦИЙ, ПОСТРОЕННЫХ ДЛЯ РОССИЙСКОГО  
ФОНДОВОГО РЫНКА**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студентки 2 курса 248 группы

направления 09.04.03 — Прикладная информатика

механико-математического факультета

Чекмаревой Альфии Жавдятовны

Научный руководитель

д. ф.-м. н., доцент.

\_\_\_\_\_

С. П. Сидоров

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2022

## ВВЕДЕНИЕ

**Актуальность темы.** Современная теория портфельного инвестирования характеризуется большим количеством возможных моделей формирования портфелей инвестиций, что оказывает на инвестора дополнительное давление в виду широкого выбора инструментов. Стоит также отметить, что вся совокупность инвестиционных моделей и стратегий не всегда может быть адаптирована к реалиям российского фондового рынка. Данный аспект сильнее усугубляет проблему выбора конкретного инструмента формирования портфеля ценных бумаг, так как необходимо детально изучить ограничения каждой модели и её реализацию в условиях деятельности российских компаний.

Обращая внимание на непрекращающееся усложнение механизмов инвестирования, развития фондового рынка, совершенствования финансовых и информационных технологий трейдинга, а также ростом неопределенности, что является неотъемлемой частью инвестирования в ценные бумаги, делающее операцию по формированию инвестиционных портфелей достаточно объёмной и время затратной, представляется актуальной задачей автоматизации и оптимизации построения инвестиционных портфелей.

**Целью магистерской работы** – решение проблемы автоматизации построения инвестиционных портфелей.

**Объект исследования** – российский фондовый рынок.

**Предмет исследования** – инвестиционные портфели, сформированные на основе 120 российских компаний.

Для достижения поставленной цели в работе необходимо решить следующие **задачи**:

Основные задачи:

- проанализировать научно-методическую литературу по теме магистерской работы;
- подготовка данных;
- построение финансовых сетей на основе частных корреляций и корреляций Пирсона;
- разработка алгоритма построения тривиальных и диверсифицирован-

ных инвестиционных портфелей;

- используя разработанный алгоритм проанализировать поведение полученных портфелей за рассматриваемый период времени.

**Практическая значимость** исследования состоит в разработке нового алгоритма для построения финансовых сетей и инвестиционных портфелей.

**Структура и содержание магистерской работы.** Магистерская работа состоит из введения, пяти разделов, списка использованных источников и приложения.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** описана краткая информация по изучаемой проблеме, приведены цели и задачи дипломной работы. В первых трех разделах рассказывается о основной теории портфельного инвестирования, теории графов, корреляционных сетях. В четвертом разделе, по материалам работы, написанной С. П. Сидоровым, В. А. Балашом, А. Р. Файзлиевым, Д.В. Мельничуком и А. Григорьевым, были описаны методы построения корреляционных финансовых сетей и приведен их анализ для нестабильного фондового рынка. Пятый, шестой и седьмой разделы посвящены практической части дипломной работы, а именно данным и методам их обработки, подробному описанию методов разработки алгоритма построения инвестиционных портфелей и полученным результатам.

**В первом разделе** рассказывается о основах теории портфельного инвестирования.

Традиционно при решении задачи формирования оптимального портфеля инвестор сначала фокусирует внимание на общем состоянии экономики, затем на определенных отраслях и, в конце концов, выбирает финансовые активы для инвестирования.

В целом формирование оптимального инвестиционного портфеля реализуется в виде последовательности следующих действий:

1. выбор целей и методов управления портфелем;
2. анализ рынка в целом, а также характеристик ценных бумаг, доступных инвестору;

3. выбор методик измерения рисков и оценки рисков;
4. формирование оптимального портфеля в соответствии с принятыми целями;
5. управление портфелем, т. е. его реструктуризация в случае, если характеристики портфеля не отвечают поставленным целям;
6. оценка эффективности портфеля и пересмотр принятой стратегии.

В теории портфельного инвестирования исходят из того, что значения доходности отдельной ценной бумаги портфеля являются случайными величинами, распределенными по нормальному (Гауссовскому) закону.

Чтобы определить распределение вероятностей случайной величины необходимо знать, какие фактические значения принимает данная величина, и какова вероятность  $i$  каждого подобного результата. При этом инвестора интересует доходность инвестиций в конце инвестиционного, холдингового периода, то есть будущие значения  $r_i$ , которые в начальный момент инвестирования неизвестны.

Под ожидаемой доходностью портфеля понимается средневзвешенное значение ожидаемых значений доходности ценных бумаг, входящих в портфель. При этом "вес" каждой ценной бумаги определяется относительным количеством денег, направленных инвестором на покупку этой ценной бумаги.

Ожидаемая доходность рассчитывается по формуле:

$$E(R_p) = \sum_{i=1} w_i E(R_i), \quad (0.1)$$

где  $E(R_p)$  — ожидаемая доходность портфеля;  $E(R_i)$  — ожидаемая доходность ценной бумаги  $i$ , за период времени;  $w_i$  — доля  $i$ -ой ценной бумаги в портфеле. С нахождением коэффициента данного выше инвестор может предполагать доходность как отдельной ценной бумаги, так и портфеля в целом.

При определении риска портфеля следует учитывать, что дисперсию портфеля нельзя найти как средневзвешенную величин дисперсий входящих в портфель ценных бумаг. Это объясняется тем, что дисперсия портфеля зависит не только от дисперсий входящих в портфель ценных бумаг, но также

и от взаимосвязи доходностей ценных бумаг портфеля друг с другом. Иными словами, риск портфеля объясняется не только индивидуальным риском каждой отдельно взятой ценной бумаги портфеля, но и тем, что существует риск воздействия изменений наблюдаемых ежегодных величин доходности одной акции на изменения доходности других акций, включаемых в инвестиционный портфель.

Дисперсия инвестиционного портфеля используется как мера риска и получается по формуле:

$$Var(R_p) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j Cov(R_i, R_j), \quad (0.2)$$

где  $Var(R_p)$  - дисперсия портфеля;  $w_i$  - доля  $i$ -ой ценной бумаги в портфеле;  $Cov(R_i, R_j)$  - ковариация ценной бумаги  $i$  и  $j$

Если портфель состоит из более чем двух ценных бумаг, то для любого заданного уровня доходности существует бесконечное число портфелей, или, иными словами, можно сформулировать бесконечное количество портфелей, имеющих одну и ту же доходность.

Тогда задача инвестора сводится к следующему: из всего бесконечного набора портфелей с ожидаемой доходностью  $E(r_p)$  необходимо найти такой, который обеспечивал бы минимальный уровень риска. Иными словами, можно задачу инвестора свести к следующему: необходимо найти минимальное значение дисперсии портфеля.

Базовой моделью для формирования портфеля является модель Марковица. Модель Марковица – математическая модель формирования оптимального портфеля ценных бумаг при определенных условиях на основе теоретико-вероятностной формализации понятия доходности и риска. Используя формулы 1.1 и 1.2 можно сформулировать прямую задачу по модели Марковица следующим образом:

$$\begin{cases} \sum_{i=1}^n w_i E(R_i) \rightarrow \max; \\ \sum_{i=1}^n w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i^2 w_j^2 \sigma_i \sigma_j \rho_{ij} \leq \sigma_{req}, w_i \geq 0; \\ \sum w_i = 1. \end{cases} \quad (0.3)$$

Решением задачи Марковица является множество неулучшаемых порт-

фелей, то есть таких, в которых ожидаемая доходность не может быть увеличена без увеличения риска и, наоборот – риск не может быть уменьшен без уменьшения ожидаемой доходности.

**Во втором разделе** описываются основные понятия теории графов.

**Третий раздел** посвящен корреляционным сетям и методам их построения по материалам работы, написанной С. П. Сидоровым, В. А. Балашом, А. Р. Файзлиевым, Д.В. Мельничуком и А. Григорьевым, были описаны методы построения корреляционных финансовых сетей и приведен их анализ для нестабильного фондового рынка.

Для расчета корреляционной матрицы ( $C$ ) для акций необходимо использовать временной ряд цен (Adj Close)  $r_i(t)$  для каждой компании  $i$  в один и тот же период времени  $t$ . Затем, чтобы сгладить колебания, мы используем доходность журнала  $Y_i(t)$  компании  $i$  за период времени  $[t-\Delta t, t]$ , определяемый формулой

$$Y_i(t) = \ln r_i(t) - \ln r_i(t - \Delta t), \quad (0.4)$$

, где  $\Delta t = 1$  для ежедневной цены.

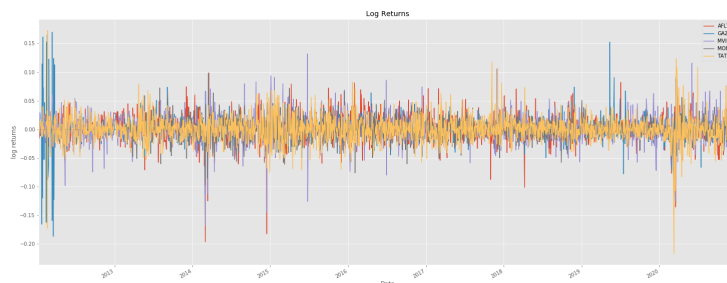


Рисунок 0.1 – График логарифмической доходности

Затем корреляционная матрица Пирсона ( $C$ ) вычисляется для каждой пары компаний  $i$  и  $j$  следующим образом.

$$C_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \quad (0.5)$$

где  $\sigma$  ковариационная матрица.

Матрица частичной корреляции ( $P$ ) была подсчитана следующим об-

разом

$$P_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}}, \quad (0.6)$$

где  $\Theta$  инверсия ковариационной матрицы или матрицы точности.

И на основе полученной матрицы корреляций рассчитывается матрица расстояний по формуле:

$$d_{ij} = \sqrt{(1 - P_{ij})} \quad (0.7)$$

**В четвертом разделе** описаны методы построения инвестиционных портфелей.

В качестве исходных данных были взяты дневные скорректированные цены закрытия 120 российских компаний из 10 секторов (технологические и коммуникационные компании объединены в один сектор). Период торгов на Московской бирже взят с 10.01.2012 по 30.12.2021. Таким образом, торгуется всего 2249 дней. Пропущенные торговые дни были заполнены значениями предыдущих торговых дней, или, если данные отсутствуют с самого начала, с первого дня торговли акциями. Всего таких дней не превышает 15 % от общего количества.

В данной работе были использованы ежедневные данные из Индекса Московской Биржи (IMOEX.ME). Индекс Московской Биржи - взвешенный по рыночной капитализации индекс российского фондового рынка, включающий наиболее ликвидные ценные бумаги.

	ABRD	AFKS	AFLT	AKRN	ALRS	AMEZ	APTK	AQUA	ARSA	BANEP
<b>Date</b>										
2012-01-10	16.161558	16.238918	34.958534	581.853760	15.103805	7.987	41.079899	75.877228	1.401672	967.262085
2012-01-11	16.161558	15.949866	34.889580	580.996155	14.941051	8.150	40.901501	76.770020	1.369355	963.897583
2012-01-12	16.161558	15.851783	35.151600	580.138550	15.232313	8.101	41.266800	74.100616	1.351736	963.056335
2012-01-13	16.161558	16.396761	34.606876	570.233765	15.083955	8.100	39.882099	74.082718	1.450695	971.383179
2012-01-16	16.161558	16.547461	34.296597	571.949036	14.959910	8.194	39.542301	74.216690	1.392177	1000.906006
...	...	...	...	...	...	...	...	...	...	...

Рисунок 0.2 – Пример исследуемых данных

На основе полученных и обработанных данных были построены финансовые сети, по которым были построено несколько видов инвестиционных портфелей. И с помощью алгоритма кластеризации и коэффициента Шарпа были выбраны портфели на каждом из рассмотренных периодов времени.

Как уже говорилось выше, данные рассматриваются за 10 лет. Для построения и анализа поведения инвестиционных портфелей был использован метод "скользящего" окна. С учетом анализа доходностей и подсчитанных значений корреляции в течение 250 торговых дней были сформированы инвестиционные портфели: тривиальный и диверсифицированный, затем на последующих 30 днях были посчитаны их математическое ожидание, как средняя доходность и дисперсия, как мера риска портфеля. После чего был осуществлен сдвиг на месяц и по аналогии построены новые портфели за такой же период времени. И также для сравнения подобный анализ полученных портфелей был проведен на том же промежутке времени, что и был сформирован.

По результату работы данного алгоритма были посчитаны средняя доходность и дисперсия каждого из полученных портфелей и проанализировано их поведение за весь рассматриваемый период.

Для сравнения и оценки эффективности диверсифицированных портфелей был построен и рассмотрен тривиальный портфель, включающий себя все акции всех рассматриваемых компаний, каждая ценная  $i$  - бумага имела равную долю, и общая сумма была равна 1.

Для оценки эффективности тривиального портфеля его числовые характеристики были посчитаны по формулам 1 и 2, как в период внутри выборки, то есть здесь рассматриваются доходности за 250 дней, так и вне выборки (30 дней).

Для построения диверсифицированного портфеля были использованы корреляционные сети по методу Пирсона и частной корреляции, а также полученные на их основе матрицы расстояний.

Для определения ценных бумаг, входящих в такой портфель на каждом из рассмотренном периоде был использован иерархического алгоритма кластеризации.

Иерархические алгоритмы работают либо «агломеративно», склеивая на каждом шаге два наиболее близких кластера, начиная с тривиального разбиения на одноэлементные кластеры, либо «дивизимно» – разбивая на каждом шаге какой-либо кластер, начиная с универсального кластера, состоящего из всех объектов.



Основная идея агломеративных алгоритмов заключается в том, что результаты алгоритма агломерации зависят от формулы пересчета расстояний до вновь построенного кластера. Общая формула Ланса и Вильямса относится к случаю, когда связи (расстояния) до объединенного кластера определяются, исходя из связей (расстояний) с объединившимися частями. Хорошо известный пример – метод ближайшего (дальнейшего) соседа, когда связь с объединенного кластера рассчитывается как максимум (минимум) связей с объединившимися кластерами. Эти два метода приводят к кластерам, которые в какой-то мере аналогичны графотеоретическим концепциям компоненты связности (ближайший сосед) и клики (дальнейший сосед).

В работе использовался агломеративный иерархический алгоритм, реализованный на Python с помощью библиотеки `scikit-learn`.

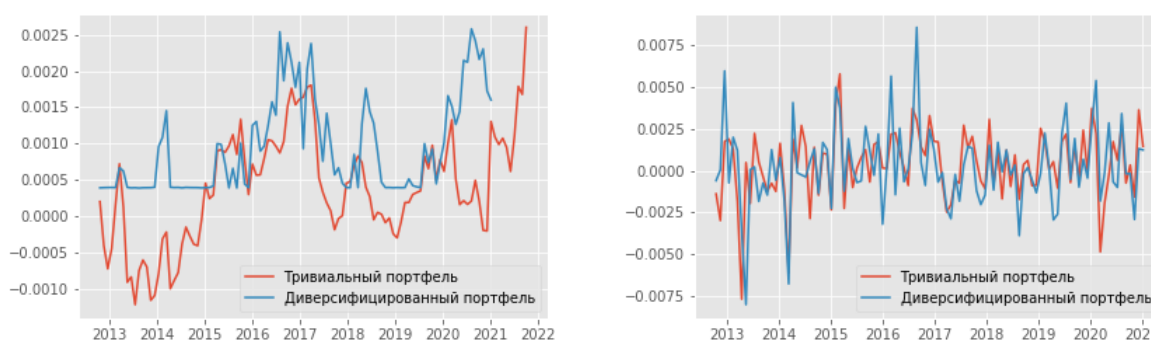
Принцип работы агломеративного алгоритма можно описать, как выполнение работы "снизу вверх". То есть, начиная с множества небольших кластеров объединяем их вместе, чтобы создать более крупные кластеры. Другими словами, каждую точку данных обозначается как один кластер, затем берется два ближайших кластера и объединяются в один. И повторяем предыдущий шаг пока не образуется кластер с нужным числом элементов. Существует различные виды связи между элементами в зависимости от способа подсчета расстояния. В данном случае используется одиночная связь, то есть кластеризуются элементы с наименьшим расстоянием и затем из каждого полученного кластера, выбирается та ценная бумага у которой наибольший коэффициент Шарпа. Таким образом формируется диверсифицированный портфель на каждом из рассмотренных этапов. После чего вычисляется доля каждой  $i$  - ценной бумаги методом оптимизации по Марковицу. Начальными условиями является, что сумма долей всех ценных бумаг должна быть равна единице и средняя доходность диверсифицированного портфеля больше, доходности тривиального. Метод реализован также на Python с использованием библиотеки `scikit-learn` и метода `optimize.minimize`.

Таким образом, в результате работы алгоритма был получен диверсифицированный портфель с оптимальными долями каждой ценной бумаги, далее считались его числовые характеристики как внутри выборки, так и вне для более четкой оценки эффективности.

**В пятом разделе** приведены результаты работы алгоритма.

После разработки алгоритма построения тривиального и диверсифицированного портфеля он был использован для анализа и сравнения поведения портфелей за рассматриваемый период времени. Основными оценками эффективности являлись математическое ожидания, как средняя доходность портфелей и дисперсия, как мера риска.

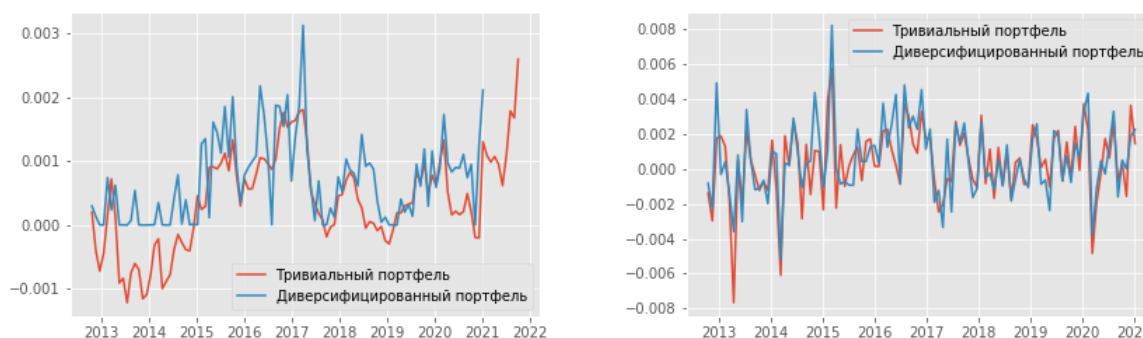
Ниже представлен график изменения средней доходности тривиального портфеля внутри и вне выборки, рассчитанной на основе частных корреляций и корреляционных значений, полученных методом Пирсона.



(a) Ожидаемая доходность портфелей – метод Пирсона (IN-SAMPLE)

(b) Ожидаемая доходность портфелей – метод Пирсона (OUT-OF-SAMPLE)

Рисунок 0.3 – Ожидаемая доходность портфелей по методу Пирсона



(a) Ожидаемая доходность портфелей – частная корреляция (IN-SAMPLE)

(b) Ожидаемая доходность портфелей – частная корреляция (OUT-OF-SAMPLE)

Рисунок 0.4 – Ожидаемая доходность портфелей по методу частных корреляций

И также были посчитаны дисперсии тривиального и дивесифиционных

портфелей, как мера риска. График результатов на каждой из выборок ниже приведены на рисунках.



(a) Дисперсия портфелей – метод Пирсона (IN-SAMPLE)



(b) Дисперсия портфелей – метод Пирсона (OUT-OF-SAMPLE)

Рисунок 0.5 – Дисперсии портфелей по методу Пирсона

После получения числовых характеристик каждого портфеля в течении всего рассматриваемого периода были посчитаны их средние значения также внутри и вне выборки по двум видам корреляционных сетей, построенных по методу Пирсона и методу частной корреляции. Результаты приведены ниже в таблице.

	Trivial_Portfolio	Diver_Potfolio
Dispersion	0.000070	0.000092
Average_Return	0.000393	0.000841

(a) Данные числовых характеристик портфелей, метод Пирсона – in-sample

	Trivial_Portfolio	Diver_Potfolio
Dispersion	0.000068	0.000107
Average_Return	0.000354	0.000186

(b) Данные числовых характеристик портфелей, метод Пирсона – out-of-sample

	Trivial_Portfolio	Diver_Potfolio
Dispersion	0.000070	0.000100
Average_Return	0.000395	0.000825

(c) Данные числовых характеристик портфелей, метод частных корреляций – in-sample

	Trivial_Portfolio	Diver_Potfolio
Dispersion	0.000068	0.000107
Average_Return	0.000354	0.000604

(d) Данные числовых характеристик портфелей, метод частных корреляций – out-of-sample

Рисунок 0.6 – Данные числовых характеристик тривиальных и диверсифицированных портфелей

В этом исследовании используется Python 3, библиотека pandas для обработки данных и библиотека yfinance для загрузка исторических данных торгового тикера, NumPy и SciPy для общих сценариев, sklearn для реализации методов оптимизации, matplotlib для построения графиков.

**В заключении** приведены результаты магистерской работы.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. Разработан алгоритм построения рыночных сетей на основе частных корреляций и по методу Пирсона.
2. На основе построенных финансовых сетей был разработан алгоритм построения инвестиционных портфелей.
3. Были посчитаны числовые характеристики сетей и проанализировано их поведение за весь рассматриваемый период.