

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

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

Кафедра теории функций и стохастического анализа

СРАВНЕНИЕ ПОРТФЕЛЕЙ ЦЕННЫХ БУМАГ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 412 группы

направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Гука Кирилла Олеговича

Научный руководитель

доцент, к. ф.-м. н.

О. А. Мыльцина

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2022

Введение

Актуальность темы. Над проблемой выбора инвестиционного портфеля работают с середины XX-го века. Под портфелем понимается набор инвестиций в ценные бумаги, обращающиеся на финансовом рынке. В 1952 г. Гарри Марковиц опубликовал фундаментальную работу, которая является основой подхода к инвестициям с точки зрения современной теории формирования портфеля. Допустим, инвестор в настоящий момент времени имеет конкретную сумму денег, которые будут инвестированы на определенный промежуток времени, называемый периодом владения. В конце периода инвестор продает ценные бумаги. В начальный момент инвестор должен принять решение о покупке конкретных ценных бумаг, которые будут находиться в его портфеле до конечного момента. Задача любого инвестора – это выбрать оптимальный портфель из набора возможных, именно в этом и заключается **актуальность** темы данной работы и ее **практическая значимость**. Работы с теорией портфельного инвестирования актуальны и разработки в данной сфере идут до сих пор.

Целью бакалаврской работы является моделирование портфелей ценных бумаг при помощи методов машинного обучения и анализ полученных результатов.

Объект исследования – портфели ценных бумаг.

Предмет исследования данной работы – анализ изменений основных характеристик портфеля ценных бумаг при использовании предварительных предсказаний изменений цены каждого актива перед созданием портфеля.

Для достижения данной цели необходимо решить следующие **задачи**:

1. Изучение теории портфелей и основных экономических понятий;
2. Подбор активов при помощи метода кластеризации;
3. Сделать предсказания цен для каждого актива с помощью методов машинного обучения;
4. Построение пар портфелей ценных бумаг с и без использования предсказаний цен активов;
5. Произвести сравнения полученных результатов.

Основное содержание работы

Бакалаврская состоит из введения, 3 разделов, заключения, списка использованных источников и приложение.

Во **введении** обосновывается актуальность темы работы и ее практическая значимость, формулируется цель работы и ставятся основные задачи.

В **первом** разделе вводятся общие понятия и определения экономики.

Понятие риска, виды рисков. Риск — одно из важнейших понятий, сопутствующих любой активной деятельности человека. Вместе с тем это одно из самых неясных и многозначных понятий. Однако во многих ситуациях суть риска хорошо понимается и воспринимается. Его содержание определяется той конкретной задачей, где данный термин используется.

Общее понятие экономического риска можно сформулировать следующим образом:

риск — это вероятность или угроза потери лицом или организацией части своих ресурсов, части своих доходов, или появления дополнительных расходов.

Назовем некоторые виды рисков, связанные с экономической деятельностью человека.

- *Валютный* — риск, связанный с резкими колебаниями курсов валют;
- *Инвестиционный* — возникает за счет обесценивания инвестиционно-финансового портфеля, состоящего из собственных и приобретенных ценных бумаг;
- *Риск ликвидности* — возникает за счет неожиданного изменения кредитных и депозитных потоков;
- *Производственный* — связан с возможностью невыполнения фирмой своих обязательств перед заказчиком.
- *Риск разорения* — вероятность столь больших потерь, которые не могут быть компенсированы и ведут к разорению конкретного лица или организации.
- *Рыночный* — риски, связанный с возможным обесценением ценных бумаг;

Риски также подразделяются на:

- *динамические* — они связаны с непредвиденными изменениями стоимости основного капитала вследствие управленческих решений, а также рыночных или политических обстоятельств;
- *статические* — такие риски связаны с возможностью потерь реальных активов при нанесении ущерба собственности и потерь доходов из-за недееспособности организации.

Анализ риска конкретного экономического проекта обычно осуществляется в следующей последовательности:

- выявляются объективные и субъективные факторы, влияющие на риск;
- производится анализ данных факторов;
- дается оценка риска с позиций, определяющих финансовую состоятельность проекта;
- устанавливается допустимый уровень риска;
- проводится анализ отдельных операций по выбранному уровню риска;
- разрабатываются мероприятия по снижению риска.

Выделяют два принципиально разных пути, ведущих к снижению риска.

1-й путь состоит в компенсации рисков с помощью так называемых рискованных премий, которые представляют собой различной рода надбавки (к цене, уровню процентной ставки, тарифу и т.д.), выступающие в виде «платы за риск».

2-й путь заключается в управлении риском, которое осуществляется на основе диверсификации, приемов хеджирования, заключения форвардных контрактов, покупки опционов, страхования.

Измерение рисков. Определим количественную оценку риска того или иного действия лица, принимающего решение (ЛПР), или его финансовых операций.

Самое важное нам нужно уметь различать риск и неопределенность. Это легло в основу определения риска, данного Ф. Найтом (1921). Именно он предложил различать риск и неопределенность.

Риск имеет место тогда, когда некоторое действие может привести к нескольким исходам с известным распределением их вероятностей. Если же распределение вероятностей неизвестно, то ситуация рассматривается как неопределенность.

При таком подходе, если ситуация позволяет его использовать, риск измеряется с помощью дисперсии или среднеквадратичного отклонения случайной величины показателя эффективности принимаемого решения от ожидаемого. В пользу выбора этого способа говорит тот факт, что чем меньше разброс (дисперсия) результата решения, тем он более предсказуем, то есть риск меньше. Если разброс равен нулю, то риск полностью отсутствует.

Предположим, что в результате выполнения некоторой финансовой операции возможны n исходов с вероятностями возникновения p_i и доходностями q_i , $i = \overline{1, n}$. Таким образом, доход от финансовой операции является случайной величиной R . Ее *средний ожидаемый доход*, то есть математическое ожидание случайной величины R , есть

$$m_R = p_1q_1 + p_2q_2 + \dots + p_nq_n.$$

Дисперсия случайной величины R (дисперсия данной операции) есть

$$D[R] = M[(R - m_R)].$$

Риском данной операции будем называть среднеквадратичное отклонение случайной величины R

$$\sigma_R = \sqrt{D[R]}.$$

В качестве применения количественного значения риска рассмотрим выбор некоторым ЛПР одного из двух вариантов инвестиций в условиях риска.

Пусть имеются два проекта A и B , в которые ЛПР может вложить средства. Проект A в определенный момент в будущем дает случайную величину прибыли R_A . Предположим, что ее среднее ожидаемое значение (математическое ожидание) равно $m_A = M[R_A]$ с дисперсией $D[R_A] = M[(R_A - m_A)^2]$. Для проекта B числовые характеристики прибыли R_B как случайной величины соответственно равны

$$m_B = M[R_B], D[R_B] = M[(R_B - m_B)^2].$$

Среднеквадратичные отклонения случайных величин R_A и R_B соответ-

ственно равны

$$\sigma_A = \sqrt{D[R_A]}, \sigma_B = \sqrt{D[R_B]}.$$

Если оказалось, что имеет место один из случаев:

1) $m_A \leq m_B, \sigma_A > \sigma_B$;

2) $m_A < m_B, \sigma_A \geq \sigma_B$,

то следует выбрать проект B .

Если же имеют место случаи:

3) $m_A < m_B, \sigma_A < \sigma_B$;

4) $m_A > m_B, \sigma_A > \sigma_B$,

то решение о выборе проекта зависит от отношения ЛПР к риску. В последнем случае проект A характеризуется большим значением ожидаемой доходности, однако он более рискован. В случае 3) для проекта A риск меньше, но и ожидаемая доходность меньше по сравнению с проектом B . Субъективное отношение к риску с математической точки зрения можно учитывать с помощью функции полезности данного ЛПР.

Диверсификация. Диверсификация капитала, то есть его распределение в различные финансовые операции, является самым распространенным способом снижения риска в инвестиционном деле. Приведем в краткой форме математическую основу метода. Пусть $Q_1, Q_2, \dots, Q_n$ — некоторые финансовые операции, требующие одинаковых вложений, случайные величины доходности которых

$\{R_i\}_{i=\overline{1,n}}$ ведут себя независимо. Обозначим через $\{m_i\}_{i=\overline{1,n}}$ и $\{\sigma_i\}_{i=\overline{1,n}}$ соответствующие значения ожидаемой доходности и риска этих операций. Предположим, что у ЛПР есть возможность распределить свой капитал, достаточный для финансирования только одной операции, в равных долях в финансировании всех операций, получая впоследствии и соответствующую часть дохода от каждой операции. Поскольку случайная величина доходности от такого способа вложения есть

$$R = \frac{R_1}{n} + \frac{R_2}{n} + \dots + \frac{R_n}{n},$$

то и ее ожидаемая доходность является средним арифметическим $m = (m_1 + m_2 + \dots + m_n)/n$ от ожидаемых доходностей всех операций. Тогда дисперсия

случайной величины R , учитывая независимость случайных величин R_i и R_j , есть

$$\begin{aligned} D[R] &= M[(R - m)^2] = n^{-2} M \left[\left(\sum_{i=1}^n R_i - \sum_{i=1}^n m_i \right)^2 \right] = \\ &= n^{-2} M \left[\left(\sum_{i=1}^n (R_i - m_i) \right)^2 \right] = n^{-2} \sum_{i=1}^n \sum_{j=1}^n M[(R_i - m_i)(R_j - m_j)] = \\ &= n^{-2} \sum_{i=1}^n M[(R_i - m_i)^2] = n^{-2} \sum_{i=1}^n \sigma_i^2. \end{aligned}$$

Отсюда, если обозначим через σ^* максимальное значение из $\{\sigma_i\}_{i=1, \overline{n}}$, вытекает, что риск операции можно оценить как

$$\sigma = \sqrt{D[R]} \leq \frac{\sigma^*}{\sqrt{n}}$$

. Таким образом, получаем важный вывод: при увеличении числа независимых финансовых операций ожидаемая доходность сконструированной нами операции находится на уровне среднего арифметического значения ожидаемых доходностей всех операций, а ее риск уменьшается, становясь сколь угодно малым числом при достаточно больших n .

Данный вывод называется *эффектом диверсификации* и представляет разумное правило работы на финансовом и других рынках. Этот же эффект воплощен в народной мудрости — «не клади все яйца в одну корзину». Принцип диверсификации гласит, что нужно проводить разнообразные, не связанные друг с другом операции. Тогда доходность окажется усредненной, а риск однозначно уменьшится.

Во **втором** разделе приводится теория по построению портфелей с использованием аппарата теории вероятностей.

В **третьем** разделе забираются основные моменты из программ на языке Python и методы машинного обучения для построения портфелей. Реализуется **численный эксперимент** и проводится построение всех необходимых графиков для каждого найденного портфеля, делаются все необходимые выводы. Данные берутся на времяянке промежутке с 01.01.2015 по 01.01.2020

года, а также проводится проверка результатов с использованием ранее неизвестных для программы данных взятых с 02.01.2020 по 31.12.2021.

Разбиение ценных бумаг на кластеры. При работе с ценными бумагами перед инвесторами возникает вопрос о выборе ограниченного набора из огромного множества активов. Без смысла в подборе нельзя выбирать активы для портфеля, так как повышается риск потерь из-за схожего поведения активов в одинаковой экономической ситуации. Для избежания этого используют диверсификацию для портфеля ценных бумаг.

Для этого будем использовать метод машинного обучения – **k-means clustering**. K-means clustering – это метод машинного обучения без учителя. Он разбивает множество элементов векторного пространства на заранее известное число кластеров **k**. Действие алгоритма таково, что он стремится минимизировать среднеквадратичное отклонение на точках каждого кластера. На каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров.

Для работы с этим алгоритмом необходимо заранее знать количество кластеров **k**. Выбор оптимального **k** можно производить ручным и автоматизированным способом. К первому способу относится "метод локтя" (Elbow method). Метод состоит из построения объясненного варианта как функции количества кластеров и выбора изгиба кривой в качестве количества кластеров для использования.

Ко второму способу относится метод по нахождению коэффициента силуэта (The Silhouette Coefficient). С помощью данного способа можно автоматически найти коэффициент **k**. Нахождение коэффициента силуэта. Коэффициент силуэта рассчитывается с использованием среднего расстояния внутри кластера a и среднего расстояния до ближайшего кластера b для каждой выборки. Формула для нахождения:

$$s_i = (b_i - a_i) / \max(a_i, b_i).$$

Значение коэффициента силуэта может варьироваться от -1 до 1 . Отрицательное значение нежелательно, поскольку оно соответствует случаю, когда a_i – среднее расстояние до точек в кластере, больше, чем b_i – минимальное среднее расстояние до точек в другом кластере. Желательно, чтобы коэффициент силуэта был положительным ($a_i < b_i$), а для этого a_i он был как можно ближе к 0 , и тогда коэффициент принимает максимальное значение 1 , когда $a_i = 0$. При выборе из нескольких коэффициентов выбирается максимально близкий по значению к 1 . Таким образом можно найти наилучшее количество кластеров k .

Для кластеризации 505 активов из индекса *S&P500* получим при методе локтя необходимо брать $k = 5$ или $k = 7$, а на методе силуэта – $k = 7$.

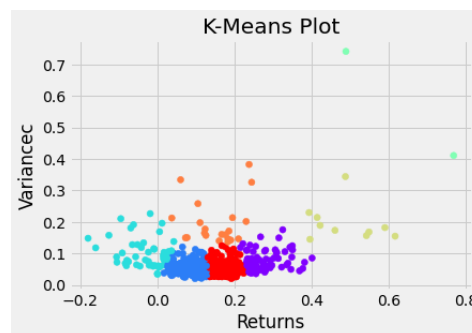


Рисунок 1 – Результаты кластеризации

Из 7 кластеров было выбрано 6 наборов по 7 активов для дальнейшей работы.

Предсказание цены. Прогнозирование временных рядов может быть сложной задачей, поскольку существует множество различных методов, которые вы можете использовать, и множество различных гиперпараметров для каждого метода.

Библиотека Prophet — это библиотека с открытым исходным кодом, предназначенная для прогнозирования одномерных наборов данных временных рядов. Он прост в использовании и предназначен для автоматического поиска хорошего набора гиперпараметров для модели, чтобы делать точные прогнозы для данных с тенденциями и сезонной структурой по умолчанию.

Математическое уравнение, лежащее в основе модели Prophet, опреде-

ляется как:

$$y(t) = g(t) + s(t) + h(t) + e(t),$$

где $g(t)$ представляет тренд, Prophet использует кусочно-линейную модель для прогнозирования тенденций; $s(t)$ представляет периодические изменения (еженедельно, ежемесячно, ежегодно); $h(t)$ представляет влияние праздников и $e(t)$ – член ошибки.

Процедура подбора модели Prophet обычно очень быстра (даже для тысяч наблюдений) и не требует предварительной обработки данных. Также можно работать с отсутствующими данными и выбросами

Prophet требует, чтобы данные были в Pandas DataFrames. Поэтому будем загружать и суммировать данные с помощью Pandas.

Можем загрузить данные непосредственно из URL-адреса, вызвав функцию `read_csv()` Pandas. Чтобы использовать Prophet для прогнозирования, сначала определяется и настраивается объект `Prophet()`, затем он подгоняется к набору данных, вызывая функцию `fit()` и передавая данные.

Объект `Prophet()` принимает аргументы для настройки нужного типа модели, например типа роста, типа сезонности и т. д. По умолчанию модель будет усердно работать, чтобы вычислить почти все автоматически.

Функция `fit()` принимает `DataFrame` данных временного ряда. `DataFrame` должен иметь определенный формат. Первый столбец должен иметь имя `"ds"` и содержать дату и время. Второй столбец должен иметь имя `"y"` и содержать наблюдения.

Это означает, что мы меняем имена столбцов в наборе данных. Также требуется, чтобы первый столбец был преобразован в объекты даты и времени, если они еще не преобразованы (например, это может быть отключено как часть загрузки набора данных с правильными аргументами в `read_csv`). По умолчанию библиотека предоставляет много подробных выходных данных во время процесса подбора.

То есть можно сделать прогноз данных, используемых в качестве входных данных для обучения модели. В идеале модель уже видела данные раньше и могла бы сделать идеальный прогноз. Тем не менее, это не так, поскольку модель пытается обобщить все случаи в данных.

Это называется созданием прогноза в выборке (в выборке обучающего

набора), и просмотр результатов может дать представление о том, насколько хороша модель. То есть насколько хорошо он усвоил обучающие данные.

Прогноз делается путем вызова функции `predict()` и передачи `DataFrame`, который содержит один столбец с именем `"ds"` и строки с датой и временем для всех прогнозируемых интервалов.

Нахождение оптимального портфеля ценных бумаг. Современная теория портфеля (МРТ) используется многими фирмами, управляющими фондами и консультантами по благосостоянию. Нобелевский лауреат Гарри Марковиц предложил эту идею. И теория остается чрезвычайно популярной сегодня из-за ее относительной простоты и производительности. МРТ предполагает, что инвесторы не склонны к риску. То есть предполагается, что инвесторы предпочтут менее рискованный портфель, чем более рискованный с той же доходностью.

Библиотека `PyPortfolioOpt` имеет несколько очень полезных методов для реализации методов оптимизации портфеля. Один из которых находит портфель с максимальным коэффициентом Шарпа. Для этого строится граница эффективности для множества портфелей. Вывод представляет собой массив весов для каждой акции. При работе с данной библиотекой есть возможность работать с "короткими позициями" (**Short positions**).

Результаты моделирования. Объединив все подпрограммы в одну получается таблица с результатами, например, для набора активов (*AMZN, BIIB, AVY, AMD, TFC, CZR, ETSY*):

	Name	Values_predict_weight	Name	Values_real_weight	Values_predict_weight	Name	Values_real_weight	Values_predict_weight
0	predict_price_AMZN	0.201730	real_price_AMZN	0.350640	0.201730	today_price_AMZN	0.350640	0.201730
1	predict_price_BIIB	0.000000	real_price_BIIB	0.000000	0.000000	today_price_BIIB	0.000000	0.000000
2	predict_price_AVY	0.664130	real_price_AVY	0.053380	0.664130	today_price_AVY	0.053380	0.664130
3	predict_price_AMD	0.024770	real_price_AMD	0.137290	0.024770	today_price_AMD	0.137290	0.024770
4	predict_price_TFC	0.000000	real_price_TFC	0.000000	0.000000	today_price_TFC	0.000000	0.000000
5	predict_price_CZR	0.109370	real_price_CZR	0.458690	0.109370	today_price_CZR	0.458690	0.109370
6	predict_price_ETSY	0.000000	real_price_ETSY	0.000000	0.000000	today_price_ETSY	0.000000	0.000000
7	Expected annual return	0.329479	Expected annual return	0.595945	0.308353	Expected annual return	0.310830	0.273450
8	Annual volatility	0.049428	Annual volatility	0.273841	0.192720	Annual volatility	0.411919	0.273662
9	Sharpe Ratio	6.261181	Sharpe Ratio	2.103210	1.600007	Sharpe Ratio	0.754592	0.999223

Рисунок 2 – Результаты моделирования

Также программа создает zip-архив. В нем хранятся графики изменения цен для активов, таблица `Portfolio` с данными по каждому из смоделированных портфелей и таблица `Price` с данными об изменениях цен активов.

Последняя таблица имеет реальные, предсказанные данные цен и разницу между этими значениями.

В **заключение** приведены основные результаты и выводы работы.

Основные результаты

В результате работы было изучено несколько методов машинного обучения и написана программа. С её помощью можно принимать решение при работе с портфельным инвестированием, но главная её особенность заключается в модульности, так как три блока программы можно использовать отдельно. Если же использовать все 3 блока, то получим таблицу с результатами для двух портфелей основанных на одном наборе активов, но имеющие разные веса. Это происходит из-за того, что первый портфель строится на реальных данных, а второй - с использованием предсказанных значений цены для каждого актива на заданном временном промежутке. Программа сохраняет подсчеты в excel-формате и далее с ним можно продолжать работать.

В **приложении** приведен код программы для создания и анализа портфелей.