ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет Высшая Школа Бизнеса

Департамент Бизнес-информатики

Проект по предмету «Анализ и прогнозирование рыночных рисков»

Образовательная программа Бизнес-информатика

Студент

Сторожок Мария Константиновна,

Группа БИИН1908

Проверяющий:

Попов Виктор Юрьевич

Москва, 2022 г.

**Оглавление**

[**Проект по предмету "Анализ и прогнозирование рыночных рисков"** 3](#_Toc98199235)

[**Что такое S&P/TSX Composite Index?** 3](#_Toc98199236)

[**Основные характеристики временного ряда доходности** 4](#_Toc98199237)

[**Волатильность доходности** 6](#_Toc98199238)

[**Кластеры волатильности** 8](#_Toc98199239)

[**Ненормальность и жирные хвосты** 10](#_Toc98199240)

[**Идентификация толстых хвостов** 14](#_Toc98199241)

[**Статистические тесты для определения толстых хвостов** 15](#_Toc98199242)

[**Графические методы анализа жирных хвостов** 16](#_Toc98199243)

[**Нелинейная зависимость** 19](#_Toc98199244)

[**Выборочное доказательство нелинейной зависимости** 20](#_Toc98199245)

[**Корреляции превышения** 23](#_Toc98199246)

[**Копулы** 24](#_Toc98199247)

[**Прогнозирование временных рядов** 33](#_Toc98199248)

[**Проверка стационарности** 34](#_Toc98199249)

[**ARMA модель** 34](#_Toc98199250)

[**ARCH модель** 40](#_Toc98199251)

[**GARCH модель** 43](#_Toc98199252)

[**Сравнение моделей** 45](#_Toc98199253)

[**Заключение** 48](#_Toc98199254)

[**Список литературы** 49](#_Toc98199255)

**Проект по предмету "Анализ и прогнозирование рыночных рисков"**

В данной работы мы поговорим о крупнейшем канадском индексе S&P/TSX Composite (^GSPTSE), который публикуется с 1979 года.

Мы будем использовать данные за период с 1979 года по 2021 год.

## **Что такое S&P/TSX Composite Index?**

S&P/TSX Composite Index — это взвешенный по капитализации индекс акций, который отслеживает результаты крупнейших и наиболее известных компаний, котирующихся на основной фондовой бирже Канады, Фондовой бирже Торонто (TSX). Это эквивалент индекса S&P500 в Соединенных Штатах, и поэтому канадские инвесторы внимательно следят за ним.

По состоянию на 10 августа 2021 г. главные компоненты по рыночной капитализации в индексе S&P/TSX включают следующие акции (точный порядок может меняться изо дня в день):

Изображение выглядит как стол

Автоматически созданное описание

## **Основные характеристики временного ряда доходности**

Временной ряд цен закрытия индекса не является информативным ввиду инфляции. По той же причине он не может быть стационарным, а значит, с ним будет неудобно работать и многие модели, рассмотренные в данном проекте, вовсе не могут быть применены к нему.

Поэтому мы будем работать с рядом доходностей индекса, который не зависит от времени, то есть является стационарным. Будем рассматривать простые доходности, которые вычисляются по формуле ниже:

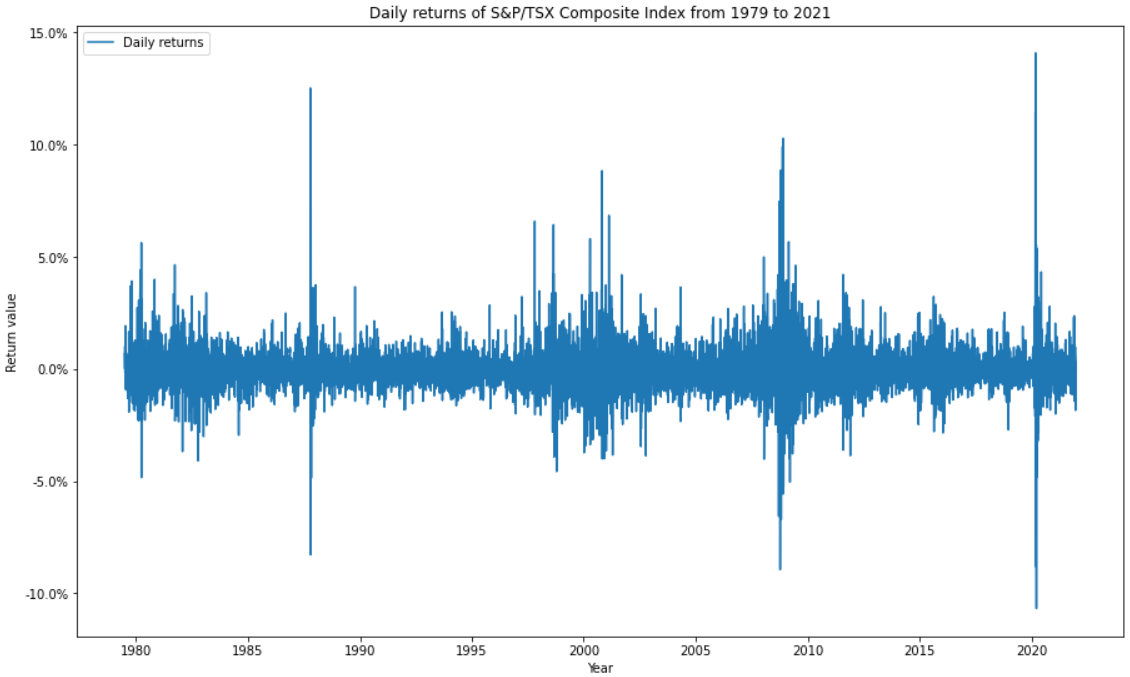
Изображение выглядит как текст, часы

Автоматически созданное описание,

где Pt - цена в момент t.

Для начала посмотрим на доходности S&P/TSX Composite Index за период с 1979 года по 2021 год.

Данные возьмем с сайта finance.yahoo.com.



На графике выше наблюдается высокая волатильность в 1988 году, 2000-х годах, 2008 и 2020 годах, среди прочих.

Это может быть связано со следующими событиями:

* Канадская экономика испытывала слабость с начала 1980 г. до конца 1983 г., с низкими годовыми темпами роста реального ВВП на уровне 2,1% и 2,6% в 1980 и 1983 гг., соответственно, и резким падением реального ВВП на 3,2% в 1982 г., а в начале 1980-х годов у Канады было два отдельных экономических спада.
* В 1988 г. Канада вступила в «большую семерку». Также с 13 по 28 февраля 1988 года в Канаде проходили XV зимние Олимпийские игры, ставшие первыми в истории этой страны.
* 22 октября во время экономический кризис 2008 года луни, или канадский доллар, упал ниже 80 долларов США впервые с середины 2005 года.
* Во II квартале 2020 года экономика Канады сократилась на рекордные 11,5% относительно предыдущего квартала и вошла в рецессию из-за коронавируса. В годовом измерении ВВП Канады в апреле-июне рухнул на 38,7%, и это тоже рекорд.

Длительные периоды высокой волатильности обычно связаны с большой неопределенностью в реальной экономике.

Посмотрим на подборку сводных статистических данных по ежедневной доходности S&P/TSX Composite Index с 1979 по 2021 гг.

Изображение выглядит как стол

Автоматически созданное описание

Среднее значение дневной доходности не просто мало, но даже отрицательно, и составляет -0.019%, дневная волатильность составляет около 0.98%.

Самая низкая дневная доходность в -10.7%, ровно как и самая высокая в 14%, соответствует экономическим последствиям короновируса в 2020 году.

Доходности имеют небольшую положительную асимметрию и, что более важно, довольно высокий эксцесс.

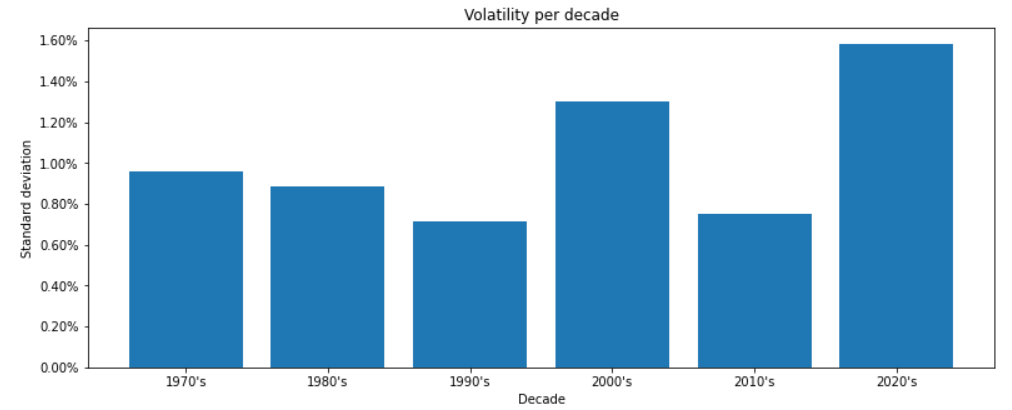
Наконец, дневная доходность имеет автокорреляцию около 3%, в то время как квадрат доходности, являющийся показателем волатильности, имеет автокорреляцию почти 24%, что дает очень убедительные доказательства предсказуемости волатильности и кластеров волатильности.

В таблице также приведены результаты теста на нормальность Жака-Бера (Jarque–Bera (JB) test) и теста на наличие автокорреляции (до 20 лагов) Льюинга-Бокса (Ljung–Box (LB) test).

## **Волатильность доходности**

Волатильность является наиболее распространенной мерой неопределенности рынка (хотя и довольно грубой).

Рассчитаем волатильность в подпериодах данных по ежедневной доходности S&P/TSX Composite Index.

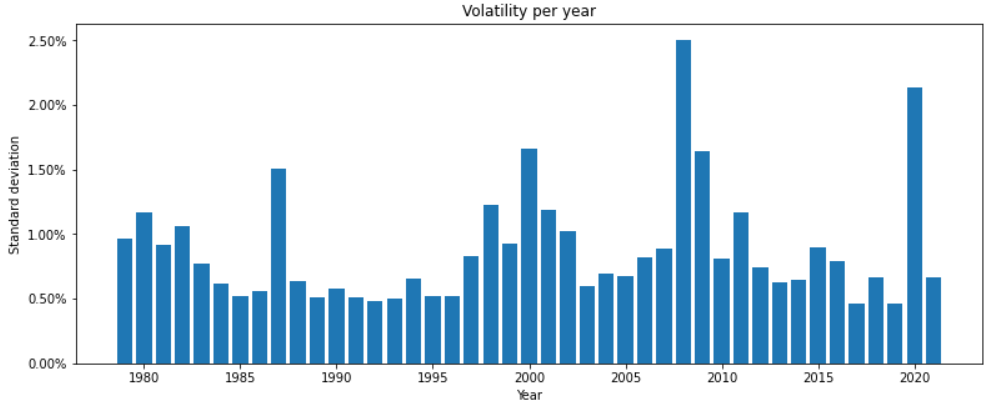


На рисунка выше показана волатильность по десятилетиям с 1979 по 2020 года, и можно увидеть явные свидетельства циклических закономерностей волатильности от одного десятилетия к другому.

Волатильность была самой низкой в 1990-х и 2010-х годах и самой высокой во время эпидемиологической ситуации, связанной с коронавирусной инфекцией, в 2020-х годах.

Обратите внимание, что значения 2020-х годов содержат данные только за два года (2020 и 2021).

Посмотрим теперь на волатильность доходности рассматриваемого индекса в разрезе лет.



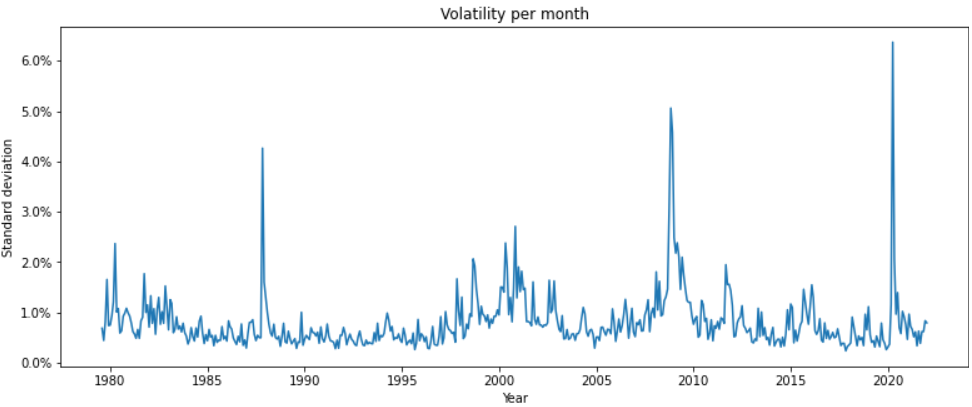
Самыми нестабильными являются 2008 и 2020 года, т.е. периоды экономического кризиса и распространения коронавирусной инфекции, соответственно.

Самые спокойные года — 2017 и 2019, как раз перед эпидемией, вспыхнувшей в 2020 году; 1991–1993 года тоже достаточно спокойные.

Однако тот факт, что волатильность была низкой в эти периоды, не означает, что риск на финансовых рынках в те годы был низким, поскольку волатильность может быть низкой, когда хвосты толстые.

Другими словами, переменная с низкой волатильностью может иметь гораздо более экстремальные результаты, чем другая переменная с более высокой волатильностью. Вот почему волатильность является вводящей в заблуждение мерой риска.

Наконец, взглянем на среднюю дневную волатильность в разрезе месяцев в период с 1979 по 2021 года.



И снова видно, что волатильность имела тенденцию к снижению и была очень низкой во втором квартале 1990-х и 2017-2019 годах. Ситуация меняется в результате кризиса 2008 года, а также событий 2020 года.

Взятые вместе, цифры дают убедительные доказательства того, что существуют как долгосрочные циклы волатильности, охватывающие десятилетия, так и короткие циклы, охватывающие недели или месяцы.

В этом случае мы наблюдаем циклы внутри циклов внутри циклов.

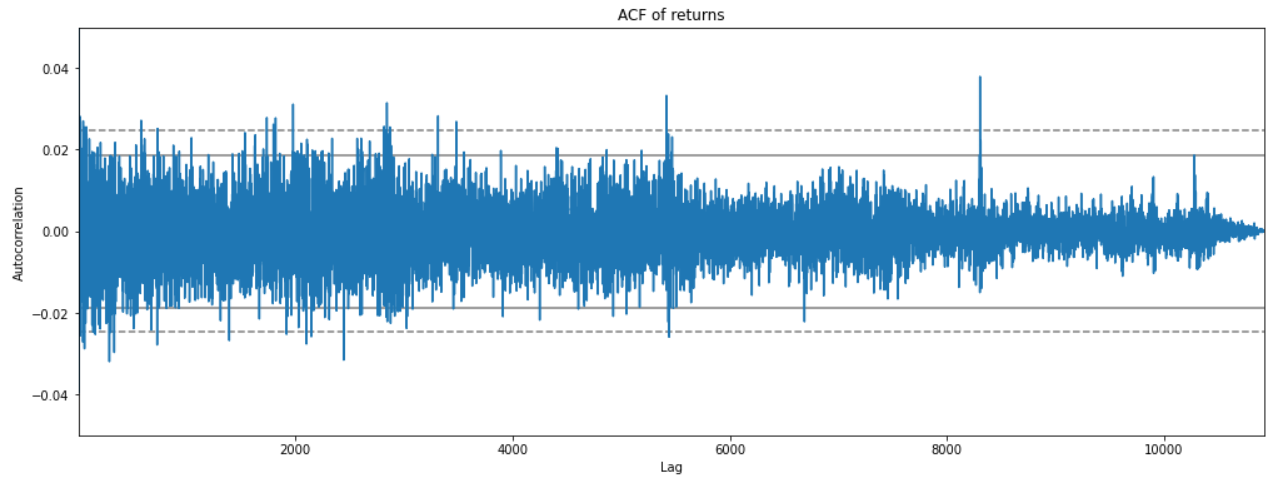
## **Кластеры волатильности**

Смотря на графики волатильности выше, становится очевидным, что она меняется со временем.

Кроме того, учитывая очевидные циклы, волатильность частично предсказуема.

Эти явления известны как кластеры волатильности.

Посмотрим на функцию автокорреляции (autocorrelation function (ACF)), которая измеряет, как доходность за один день коррелирует с доходностью в предыдущие дни. Если такие корреляции статистически значимы, у нас есть веские доказательства предсказуемости.

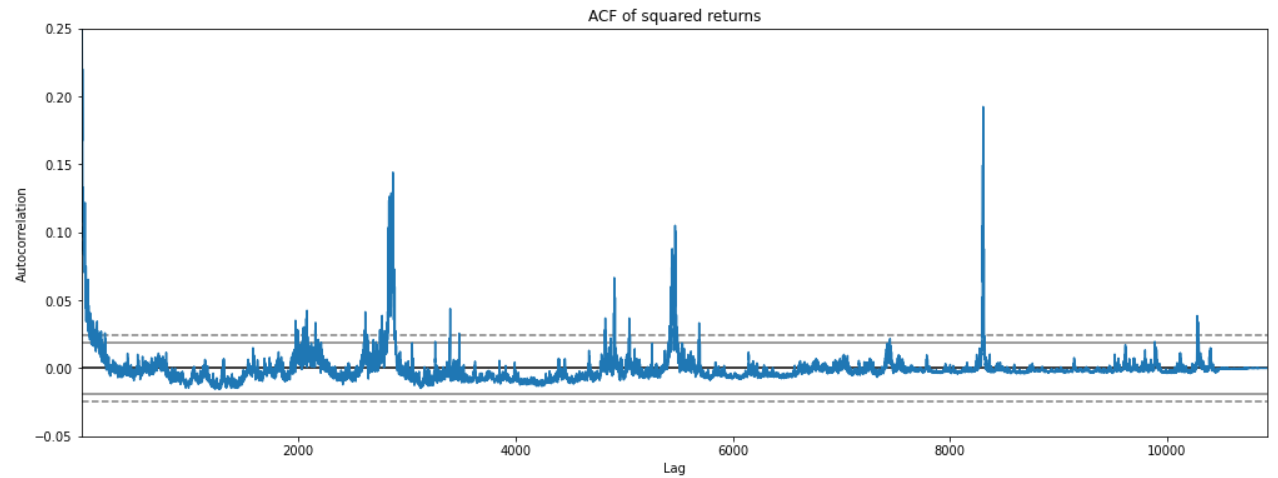


Итак, мы построили график автокорреляции дневных доходностей S&P/TSX Composite Index в рассматриваемый период.

Горизонтальные сплошная и пунктирная линии на графике соответствуют 95% и 99% доверительным интервалам.

Видна низкая автокорреляция, но есть выбросы, что удовлетворяет гипотезе, согласно которой в спокойные периоды автокорреляция низкая, а когда происходят катастрофические события, автокорреляция начинает возрастать.

Сравним полученную ACF с ACF квадратов доходностей на графике ниже (с теми же доверительными интервалами), где она значима даже при длительных лагах, предоставляя убедительные доказательства предсказуемости волатильности.



Мы можем проверить совместную значимость коэффициентов автокорреляции для нескольких лагов, используя тест Льюинга-Бокса (Ljung–Box (LB) test).

Тест Льюинга-Бокса — это статистический тест, который проверяет наличие автокорреляции во временном ряду.

В этом тесте проверяются следующие гипотезы:

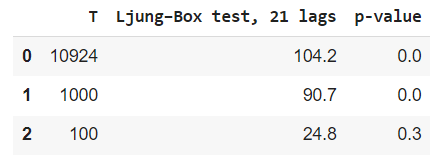
H0: остатки распределяются независимо.

HA: остатки не распределяются независимо, они демонстрируют корреляцию.

В идеале мы хотели бы не отвергать нулевую гипотезу. То есть мы хотели бы, чтобы p-значение (p-value) теста было больше 0,05, потому что это означало бы, что остатки для нашей модели временных рядов независимы, что часто является предположением, которое мы делаем при создании модели.

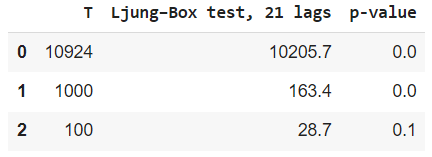
Проведем тест, используя 21 лаг дневной доходности S&P/TSX Composite Index.

Тест выполняется с использованием полного размера выборки, а также последних 2500 и 100 наблюдений. Результаты приведены в таблице ниже:



Тестовая статистика теста для всей выборки составляет 104.2, а p-значение теста составляет 0.0, что намного меньше 0.05. Таким образом, мы еще раз отвергаем нулевую гипотезу теста и заключаем, что остатки не являются независимыми.

Повторим ту же процедуру для квадратов доходностей индекса:



Причина сосредоточения внимания на квадратах доходностей заключается в том, что они являются косвенными показателями волатильности.

p-значение (p-value) для наименьшего размера выборки квадратов доходностей намного ниже, чем соответствующее значение для выборки доходностей.

Таким образом, таблицы выше еще раз показывают, что прогнозировать волатильность легче, чем доходность.

## **Ненормальность и жирные хвосты**

Под ненормальностью мы будем понимать отсутствие нормального распределения.

Часто при анализе финансовых рисков предполагается, что доходности индекса распределяются нормально.

Посмотрим на ежедневные доходности S&P/TSX Composite Index и их вероятности при условии нормальности.

Изображение выглядит как стол

Автоматически созданное описание

Из приведенной в самом начале статистики видно, что самое большое однодневное падение индекса составило 11% (10.7%), в то время как самое большое поднятие индекса равнялось 14% (14.1%).

Если бы ряд доходностей S&P/TSX Composite Index действительно был нормально распределен, вероятности этих однодневных краха и пика были бы 1.6∗10-28 и 1.7∗10-45, соответственно, согласно таблице выше.

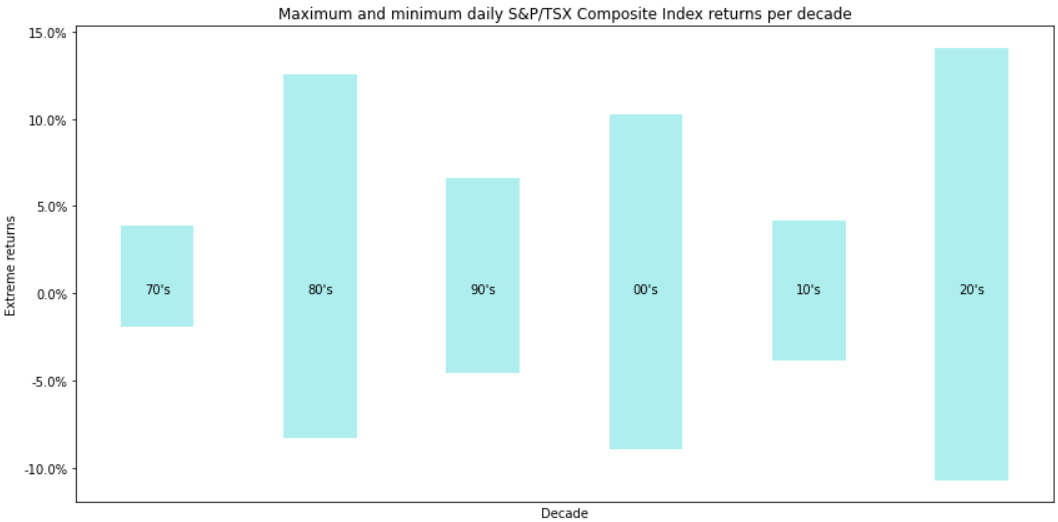
Другими словами, такое крушение должно происходить раз в 10-28 лет, а подобное поднятие - раз в 10-45 лет (с учетом выходных и праздников).

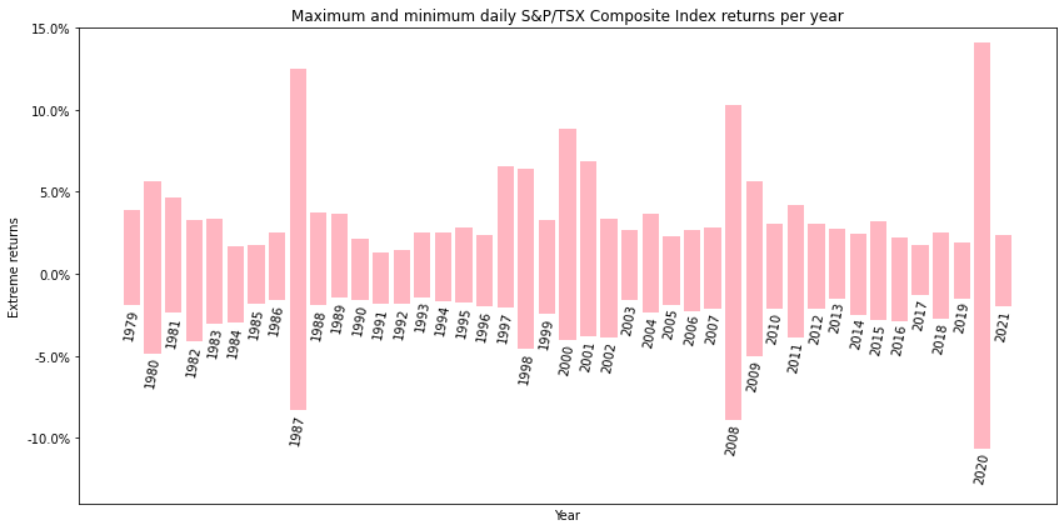
Чтобы поместить это в контекст, ученые обычно предполагают, что Земле около 10-7 лет, а Вселенной 10-13 лет.

Таким образом, предполагая нормальность, мы утверждаем, что крах 2020 года происходит лишь в каждой второй вселенной, тем временем, достижение доходности в 14%, как это произошло в том же 2020 году, случается только в одной из каждых 3 вселенных.

Тем не менее, можно утверждать, что поднятие 2020 года, как и крах, было аномалией, поэтому допущение нормальности для всех остальных дней было бы относительно безобидным.

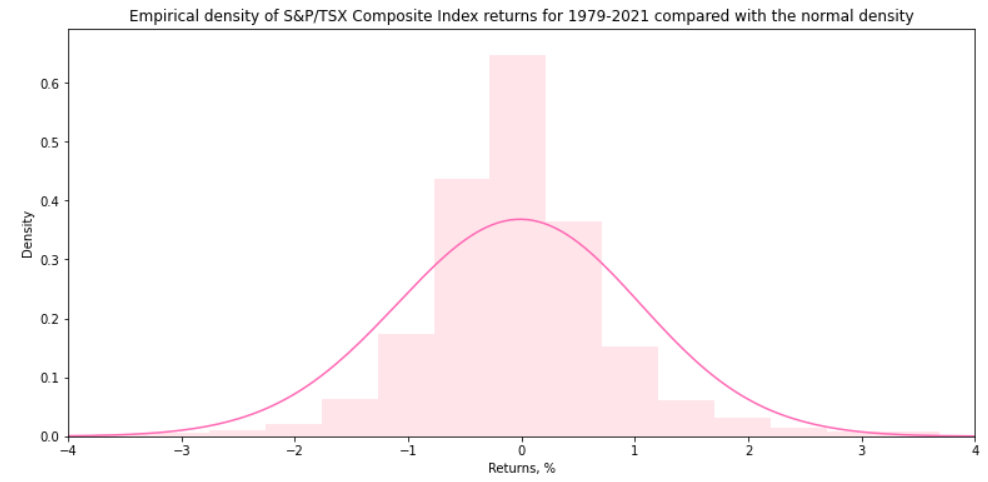
Поэтому взглянем на самые экстремальные значения дневной доходности индекса в разрезе отдельных десятилетий и годов, соответственно.





Из приведенных выше графиков видно, что экстремумов гораздо больше, чем предсказывает таблица, соответствующая перцентилям нормального распределения.

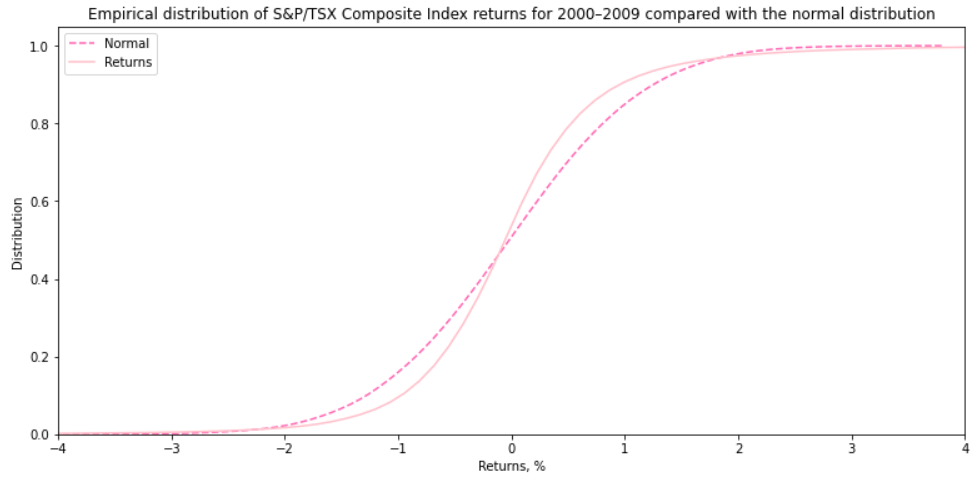
Воспользуемся альтернативным способом анализа распределения доходностей S&P/TSX Composite Index:



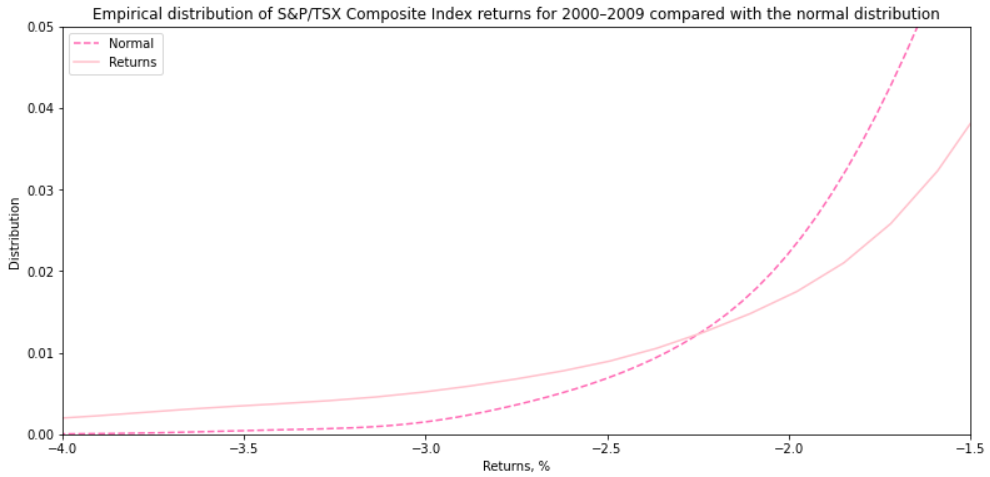
На графике выше построена гистограмма доходности (в %), на которую наложено нормальное распределение с теми же средним значением и дисперсией.

Для удобства хвосты обрезаны на уровне 4%.

Посмотрим также на нормальное распределение и эмпирическое распределение доходности индекса:



Увеличим левый хвост распределения:



Итого, мы можем сделать следующие выводы:

1. Пик распределения доходностей индекса намного выше, чем для нормального распределения.
2. Стороны распределения доходностей индекса ниже, чем у нормального распределения.
3. Хвосты распределения доходностей индекса намного толще, чем у нормального распределения.

Другими словами, существует больше дней, когда на рынке происходит очень мало, чем предсказывается нормальным распределением, и больше дней, когда рыночные цены значительно меняются.

## **Идентификация толстых хвостов**

Как мы только что убедились, доходности индекса демонстрируют то, что известно как толстые хвосты.

Как уже было замечено, говорят, что случайная величина имеет толстые хвосты, если она показывает более экстремальные результаты, чем нормально распределенная случайная величина с теми же средним значением и дисперсией.

Это означает, что рынок имеет больше относительно больших и малых результатов, чем можно было бы ожидать при нормальном распределении, и, наоборот, меньше доходов промежуточной величины.

В частности, вероятность больших результатов намного выше, чем можно было бы предсказать для нормально распределенной величины.

### **Статистические тесты для определения толстых хвостов**

Коэффициент асимметрии — величина, характеризующая асимметрию распределения вероятностей случайной величины.

Эксцесс — степень остроконечности распределения по отношению к хвостам.

Высокий эксцесс обычно означает, что большая часть дисперсии обусловлена ​​нечастыми экстремальными отклонениями, чем предсказывается нормой, и является сильным, но не идеальным признаком того, что ряд возвратов имеет толстые хвосты.

Эксцесс свыше 3 считается избыточным. Таким образом, быстрый, но, конечно, не самый надежный тест на толстые хвосты состоит в том, чтобы увидеть, превышает ли эксцесс 3.

Вспомним, что значение эксцесса в таблице со всей основной статистикой равно 21 (21.4), что является довольно веским доказательством против нормальности. При этом асимметрия составляет 1.4, в то время как третий момент нормального распределения всегда равен 0, по чему также можно судить о ненормальности распределения временного ряда.

Если говорить о более "официальных" тестах в этой категории, распространенным является тест Жака-Бера (Jarque–Bera (JB) test), значение которого мы также считали вначале.

Этот тест проверяет ошибки наблюдений (в нашем случае, дневные доходности индекса) на нормальность посредством сверки их третьего момента (асимметрия) и четвёртого момента (эксцесс) с моментами нормального распределения, у которого S=0, K=3.

В тесте Жака—Бера проверяется нулевая гипотеза H0: S=0, K=0 против гипотезы H1: S≠0, K≠0, где S — коэффициент асимметрии (Skewness), K — коэффициент эксцесса (Kurtosis).

Чем ближе распределение ошибок к нормальному, тем меньше статистика Жака—Бера отличается от нуля. При достаточно большом значении статистики p-value будет мало, и тогда будет основание отвергнуть нулевую гипотезу (статистика попала в «хвост» распределения).

Для S&P/TSX Composite Index p-value принимает значение 0, что еще раз подтверждает ненормальность распределения доходностей индекса.

Другим распространенным тестом на нормальность является тест Колмогорова-Смирнова (Kolmogorov–Smirnov test (KS)), который основан на оценке минимального расстояния, сравнивающей выборку с эталонным распределением вероятностей (например, с нормальным распределением).

Преимущество теста KS заключается в том, что он не делает никаких предположений о распределении данных, за исключением непрерывности обеих функций распределения (т. е. с технической точки зрения, он непараметричен и не зависит от распределения).

Иногда утверждают, что тест KS более мощный, чем тест JB, потому что он рассматривает все распределение.

Тест KS чувствителен к различиям как в расположении, так и в форме кумулятивной функции распределения, и на практике требуется относительно большое количество наблюдений, чтобы отклонить нулевую гипотезу.

Однако в большинстве случаев тесты KS и JB совпадают.

Проведем такой тест для индекса и получим: 0.0 %.

Таким образом, в нашем случае значения тестов действительно совпадают.

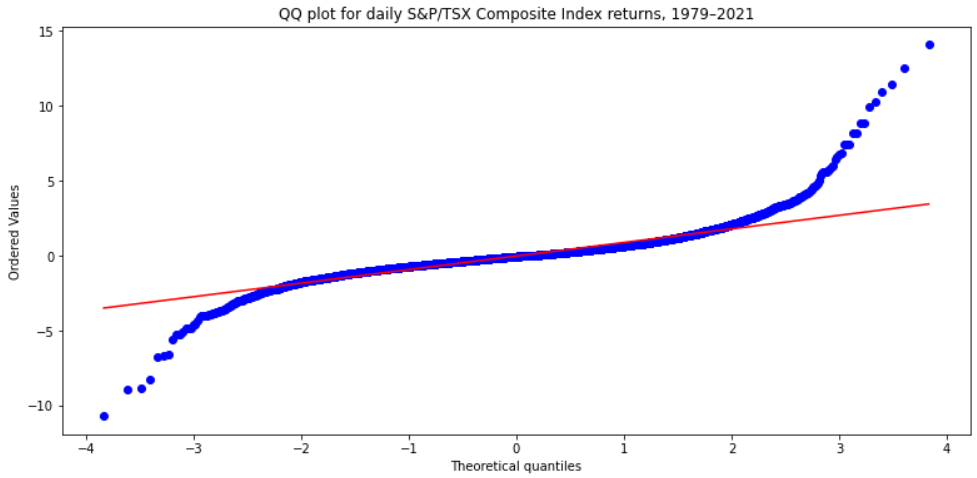
### **Графические методы анализа жирных хвостов**

Возможно, наиболее часто используемым графическим методом для анализа хвостов распределений является QQ plot (quantile–quantile plot).

График квантиль-квантиль используют для оценки того, имеет ли набор наблюдений определенное распределение или два набора данных имеют одинаковое распределение.

QQ plot сравнивает квантили выборочных данных с квантилями эталонного распределения, таким образом, при его использовании оценивается, имеет ли набор выборочных наблюдений то же распределение.

Посмотрим на QQ plot для S&P/TSX Composite Index по сравнению с нормальным распределением:



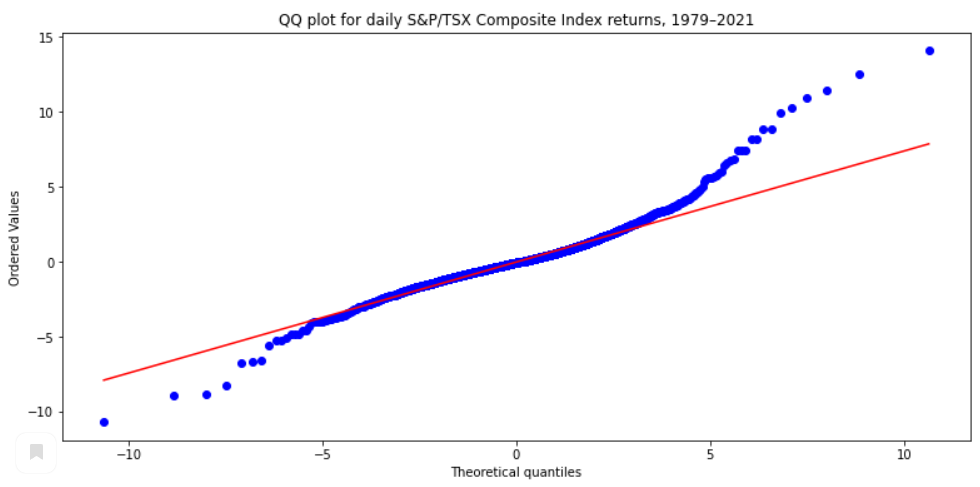
Прямая линия — это нормальный прогноз.

Из QQ plot'а выше видно, что многие наблюдения, кажется, отклоняются от нормальности, как в минус, так и в плюс, поскольку график QQ имеет четкую S-образную форму.

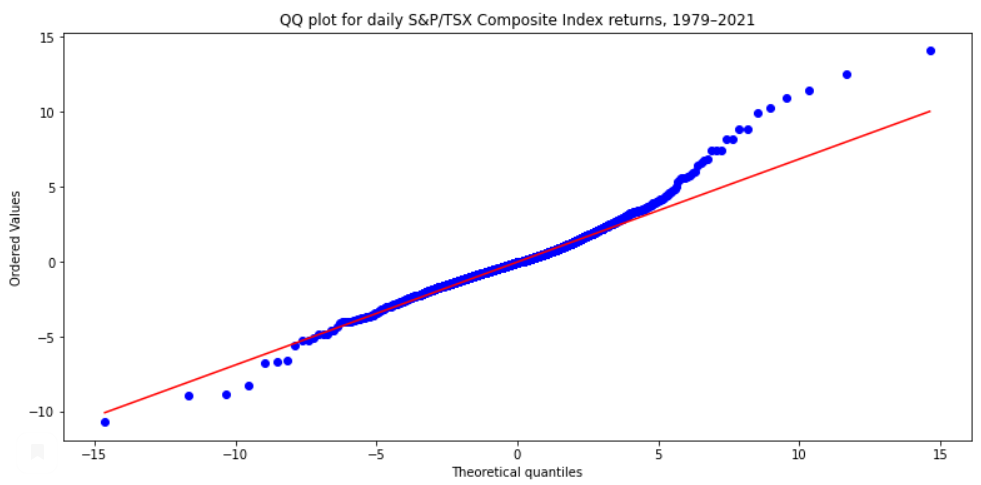
Хвосты у доходностей кажутся более толстыми, чем у нормального распределения. Мы можем получить некоторое представление об упитанности хвоста, сравнивая данные по индексу с распределением с толстыми хвостами.

Например, распределение Стьюдента-t имеет толстые хвосты, где степени свободы указывают, насколько толстыми на самом деле являются эти хвосты.

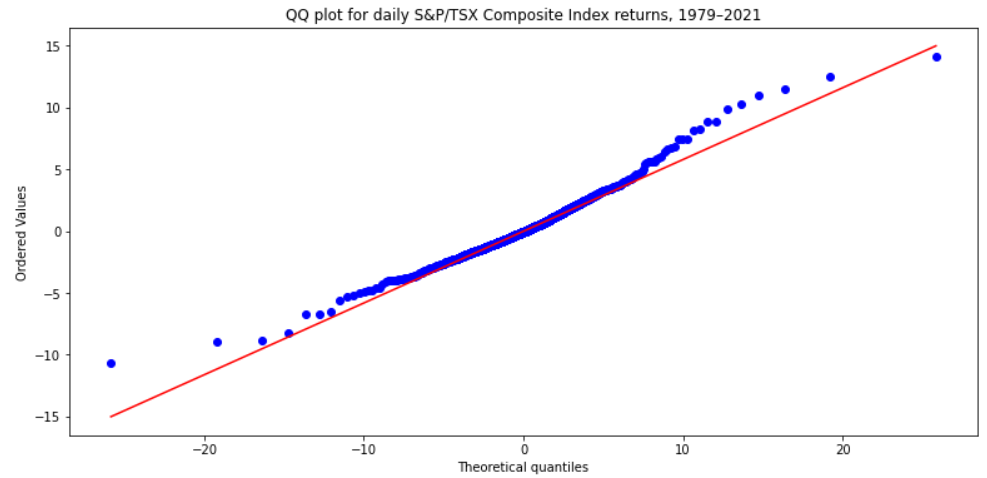
На графике ниже распределение Стьюдента с 5 степенями свободы, t(5), выбрано в качестве эталонного распределения.



Доходности явно кажутся высокими по сравнению со Student-5 как с нижней, так и с верхней стороны. Посмотрим на соответствующие графики для t(4) и t(3):



Кажется, что данные становятся ближе, нижняя сторона примерно соответствует t(4), в то время как верхняя сторона все еще довольно жирная.



Однако, из графика для t(3) мы видим, что данные больше соответствуют Student-3, как с верхней, так и с нижней стороны.

Вывод состоит в том, что доходность S&P/TSX Composite Index имеет хвосты, приблизительно равные t(3), где верхний хвост кажется немного толще нижнего.

Альтернативным графическим методом обнаружения толстых хвостов является график последовательных моментов, в основе которого лежит теория экстремальных значений.

Здесь толщина хвоста распределения измеряется индексом хвоста, обозначенным *l*. Чем ниже этот индекс, тем толще хвосты.

В частном случае распределения Стьюдента-t хвостовой индекс соответствует степеням свободы.

Таким образом, этот простой графический метод проверяет толщину хвоста с использованием выборочных моментов данных. Для этого вычисляется m-й центральный момент, который определяется выражением:

Изображение выглядит как текст, часы, датчик

Автоматически созданное описание

Этот интеграл не имеет конечного решения для всех m и всех распределений. В частности, если распределение имеет толстый хвост, мы можем вычислить моменты только для *m < l*.

Из этого следует, что если число моментов, которые мы можем вычислить, конечно (т. е. мы не можем вычислить моменты для всех *m > 0*), данные должны иметь толстые хвосты.

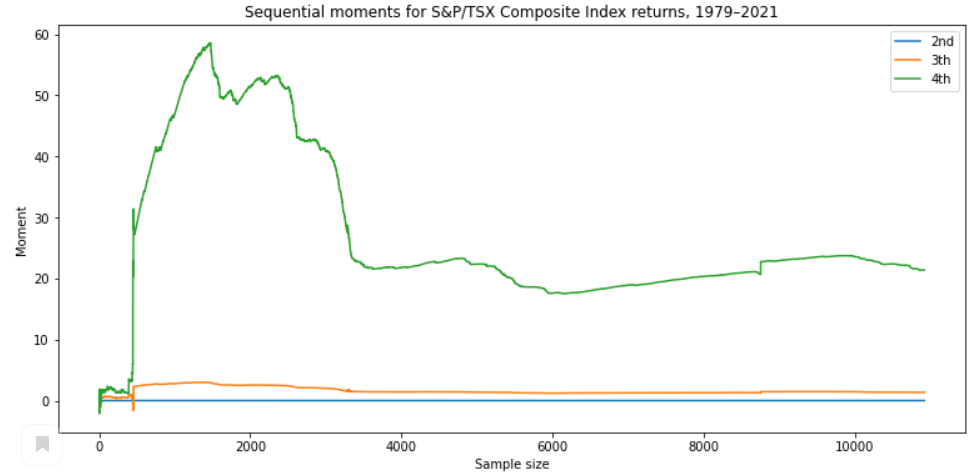
В случае нормального распределения имеем *l = ∞*, а значит, можем вычислить все моменты.

Таким образом, мы можем измерить толщину хвостов, графически отображая моменты набора данных по мере добавления все большего количества наблюдений:

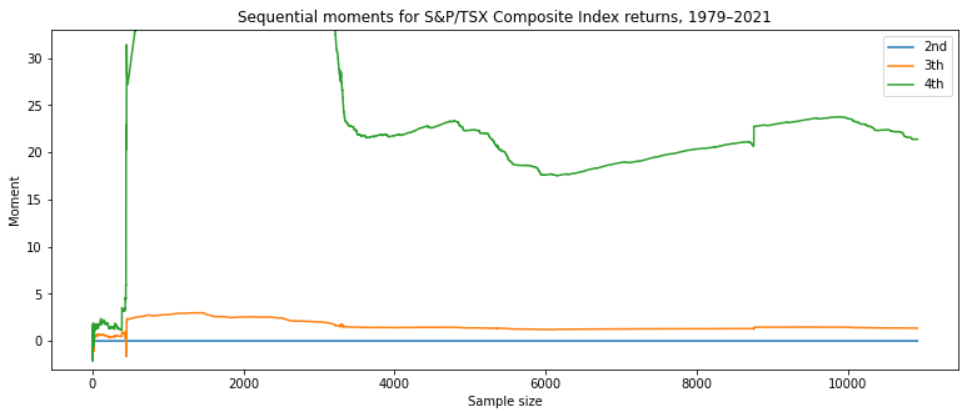
Изображение выглядит как текст, часы

Автоматически созданное описание

Посчитаем 2, 3 и 4 моменты для наших данных и построим соответствующий график:



Немного приблизим график:



Как и ожидалось, второй момент сходится, а вот четвертый - нет, что указывает на то, что хвостовой индекс доходности важнейшего канадского индекса находится между 2 и 4.

## **Нелинейная зависимость**

Последним стилизованным фактом финансовой доходности является нелинейная зависимость - наблюдение, что зависимость между различными рядами доходности изменяется в зависимости от рыночных условий.

Например, большую часть времени цены на активы меняются относительно независимо друг от друга, но в кризис все они падают вместе.

На практике совместные экстремальные результаты более вероятны, чем это предсказывает многомерная нормальность и линейные корреляции.

Большинство статистических моделей предполагают, что взаимосвязь между различными доходами является линейной, однако недавние исследования показали, что предположение о линейной зависимости обычно не выполняется для доходности активов, где корреляция обычно ниже на бычьих рынках, чем на медвежьих.

Кроме того, если бы финансовые данные были совместно распределены нормально, корреляции уменьшались бы для экстремальных событий, тогда как эмпирически мы видим, что корреляции имеют тенденцию увеличиваться до единицы во время кризиса, как показано ниже.

Чтобы уловить такие явления, модели нелинейной зависимости позволяют изменять структуру зависимости в зависимости от рыночных условий.

При этом линейные корреляции завышают зависимость в некризисные периоды и занижают корреляции во время кризисов.

## **Выборочное доказательство нелинейной зависимости**

Мы проиллюстрируем в данном разделе нелинейную зависимость с помощью долгосрочных корреляций между доходностями трех банков, Bank of Montreal (BMO), Royal Bank of Canada (RY) и The Toronto-Dominion Bank (TD), и одной нефинансовой фирмы Alimentation Couche-Tard Inc (ATD), специализирующейся на розничной торговле, все они являются компонентами S&P/TSX Composite Index, за период времени с 6 сентября 1996 г. по 10 декабря 2021 г.

Итак, взглянем на корреляции (и доверительные интервалы) ежедневных доходностей для Bank of Montreal (BMO), Royal Bank of Canada (RY), The Toronto-Dominion Bank (TD), а также Alimentation Couche-Tard Inc (ATD) за период с 6 сентября 1996 г. по 23 марта 2020 г.:

Изображение выглядит как текст

Автоматически созданное описание

Как можно заметить, финансовые показатели относительно сильно коррелируют, в то время как компания, занимающаяся продуктовой розницей, коррелирует с ними совсем незначительно — около 11% (можно сказать, что и не коррелирует вовсе), и это логично.

Посмотрим, что изменится для временного промежутка с 23 марта 2020 г. по 1 апреля 2020 г., во время пика экономической нестабильности, связанной со вспышкой вируса в Канаде:

Изображение выглядит как текст

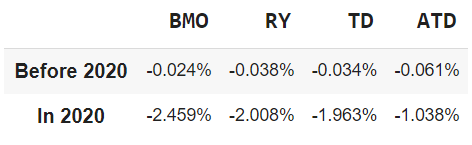
Автоматически созданное описание

Мы видим, что корреляция активов резко и значительно возросла. Например, если судить по таблице выше, корреляция между доходностями Royal Bank of Canada (RY) и The Toronto-Dominion Bank (TD) увеличилась с 79% до 100%, с доверительным интервалом (98%, 100%), что важно иметь в виду.

Такая высокая корреляция указывает на то, что акции двух компаний движутся не просто почти синхронно, а буквально шаг в шаг.

Существенный рост коэффициентов корреляции затронул также Alimentation Couche-Tard Inc (ATD), чья деятельность никак не связана с банковским сектором. Корреляции этой компании с финансовыми организациями выросли в несколько раз.

Кроме того, в таблице ниже показано, как период экономической нестабильности повлиял на фактические цены (доходности) акций:



Средняя ежедневная доходность снизилась примерно на 2% для банков и на 1%, что тоже не мало, для продуктовой сети.

Это эмпирический факт, что очень высокие корреляции обычно связаны с очень отрицательной доходностью.

Повторим те же действия для кризисного 2008 для большей наглядности, рассмотрим для этого доходности активов с 6 сентября 1996 г. по 10 октября 2008 г.:

Изображение выглядит как текст

Автоматически созданное описание

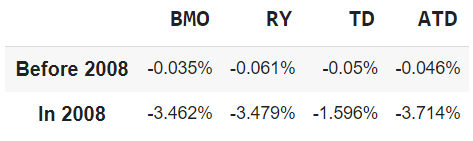
Видно, что доходности банковских акций изначально коррелируют между собой, между тем, корреляции нефинансовой организации с ними, можно сказать, равны нулю.

Посмотрим, что изменилось с 10 октября 2008 г. по 17 октября 2008 г.:

Изображение выглядит как текст

Автоматически созданное описание

Как и в случае с рыночной ситуацией, обусловленной обостренной эпидемиологической ситуацией в стране в 2020 г., кризис 2008 г. ощутимо повлиял на корреляции акций разных фирм.



И в этот раз, кризисная ситуация в значительной мере отразилась на средних ежедневных доходностях активов как финансовых, так и нефинансовых организаций.

## **Корреляции превышения**

Один из методов документирования наличия нелинейной зависимости заключается в использовании корреляций превышения.

Рассмотрим доходности двух акций X и Y, которые были стандартизированы (среднее значение равно нулю, а дисперсия равно единице).

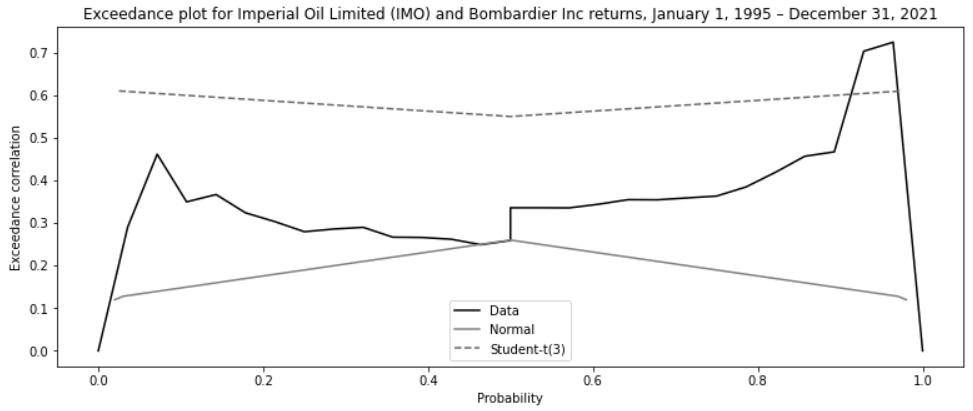
Корреляции превышения показывают, что корреляции двух акций зависят от превышения некоторого порога, то есть:

Изображение выглядит как текст

Автоматически созданное описание

где QX(p) и QY(p) — p-е квантили X и Y, соответственно, с учетом предположения о распределении.

Форма графика корреляции превышения зависит от распределения данных.



На приведенном выше графике показаны эмпирические корреляции превышения для ежедневных доходностей Imperial Oil Limited (IMO) и Bombardier Inc (BBD), входящих в S&P/TSX Composite Index, за 27 лет, наложенные на корреляции превышения для двумерного нормального распределения и двумерного распределения Стьюдента-t(3) с тем же коэффициентом корреляции.

Если говорить о нормальном распределении и распределении Стьюента с 3 степенями свободы, графики нелинейные по p, но симметричные.

Корреляции превышения уменьшаются для нормального распределения, когда мы приближаемся к крайним квантилям, в то время как они увеличиваются для Стьюдента-t(3).

В случае с активами IMO и BBD, корреляции превышения демонстрируют существенную асимметрию.

Доходности акций совсем не коррелируют в самых крайних положениях, по мере перемещения в центр распределения они резко становятся сильно коррелированными, после чего их корреляции постепенно уменьшаются.

## **Копулы**

Корреляционный анализ, который мы провели выше, помогает определить характер нелинейной зависимости (НЛЗ).

Однако есть более формальные методы моделирования НЛЗ, один из которых заключается в использовании копул.

Копулы предоставляют средства для создания многомерного распределения с рядом типов зависимости. Рассмотрим технологию использования копул ниже.

Совместное распределение нескольких случайных величин состоит из информации о каждой переменной в отдельности, а также информации о том, как различные случайные величины связаны друг с другом.

Предположим, что X и Y — две случайные величины, представляющие доходности двух разных акций,

Изображение выглядит как текст

Автоматически созданное описание

Вместе совместное распределение и маргинальные распределения представлены совместной плотностью:



Идея копулного подхода заключается в том, что мы отдельно фокусируемся на маргинальных распределениях (F, G) и функции, которая объединяет их в совместное распределение, H.

Эта функция и есть копула. Другими словами, копула извлекает информацию о структуре зависимости из совместного распределения.

Первым шагом в применении копул является преобразование X и Y в случайные величины, равномерно распределенные между нулем и единицей, что удаляет индивидуальную информацию из двумерной плотности h.

Получим равномерные величины из X и Y следующим образом (согласно теореме 1.1. Фишера Р. Э. (1925), Statistical Methods for Research Workers, Oliver & Boyd, London States):

Изображение выглядит как текст

Автоматически созданное описание

Далее мы приступим к моделированию структуры зависимости между полученными равномерными случайными величинами с помощью копулы.

Копула — это распределение вероятностей на единичном кубе, для которого каждое маргинальное распределение равномерно на интервале [0;1].

Копула содержит всю информацию о зависимости в исходной двумерной плотности h, но не содержит никакой индивидуальной информации.

Согласно Sklar, A. (1959), ‘‘Fonctions de reґpartition a` n dimensions et leurs marges,’’ Publ. Inst. Statis. Univ. Paris, 8, 229–231., при всех вышеперечисленных условиях существует единственная копула C такая, что:



Можно также разложить совместную плотность, используя плотности, на:

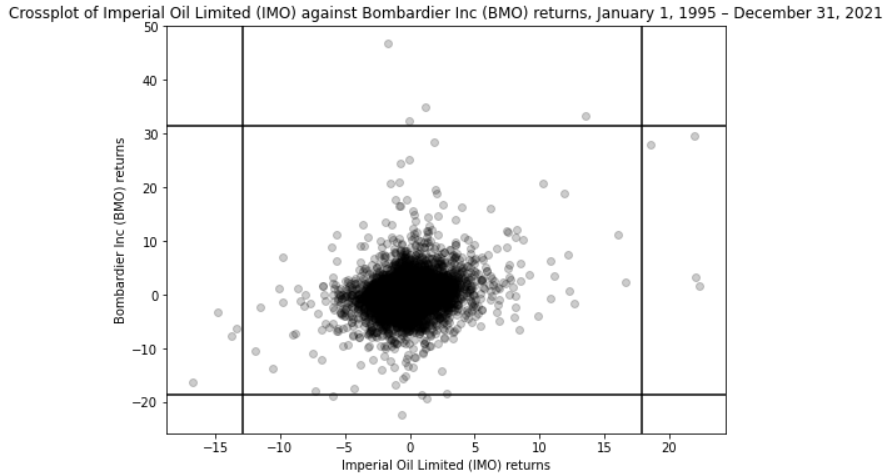


Итак, мы отделяем распределения отдельных активов от распределения, которое связывает их вместе, и копула предоставляет информацию о том, как активы ведут себя совместно.

Таким образом, мы можем построить совместное распределение из любых двух маргинальных распределений и любой копулы и извлечь подразумеваемую копулу и маргинальные распределения из любого совместного распределения.

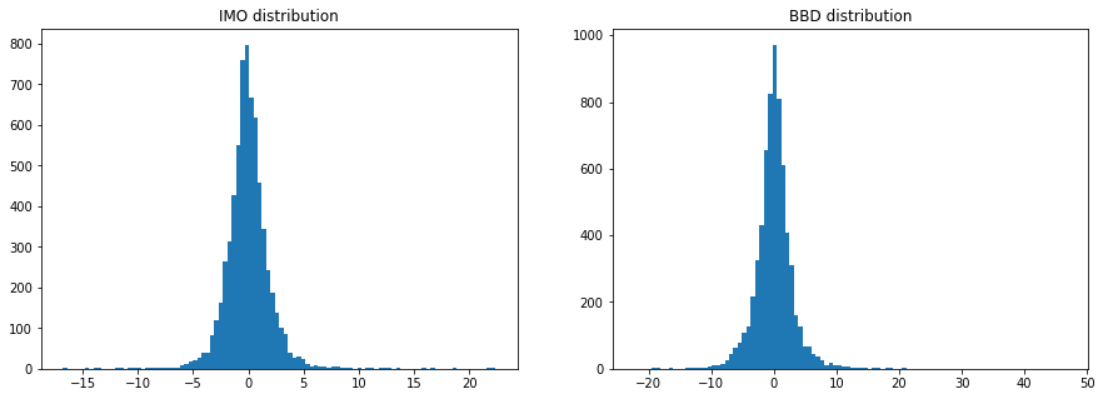
Попробуем применить копулы к тем же данным акций IMO и BBD.

Для начала взглянем на кросс-график доходностей акций вместе с их квантилями 0.05% и 99.95%.



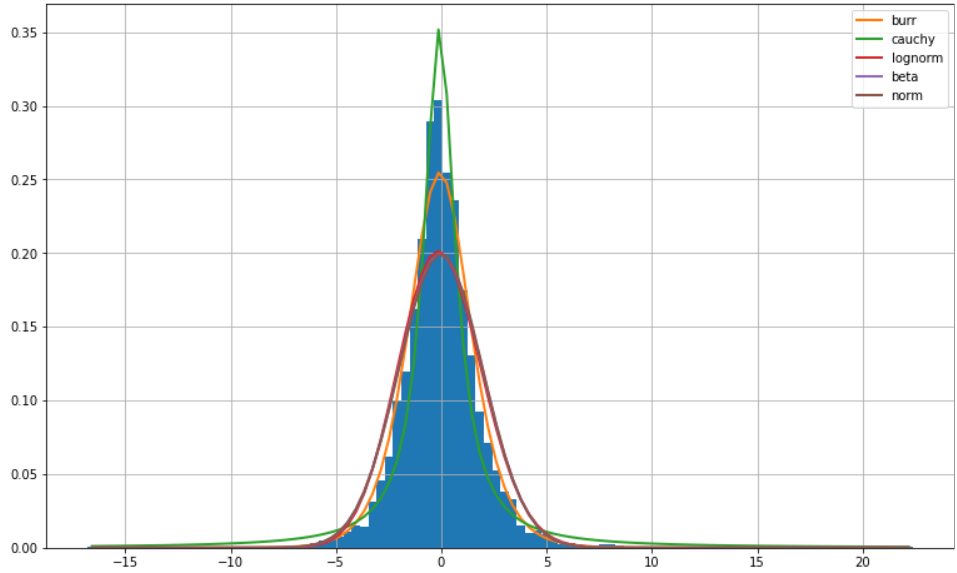
Подавляющее большинство точек данных, кажется, сосредоточены вокруг нуля, но есть несколько выбросов для обеих акций, которые лежат за пределами квантилей их выборок. Нормально распределенные данные не имели бы таких общих экстремумов.

Посмотрим теперь на распределения доходностей обеих акций по отдельности и попробуем оценить на глаз, какие это могут быть распределения:



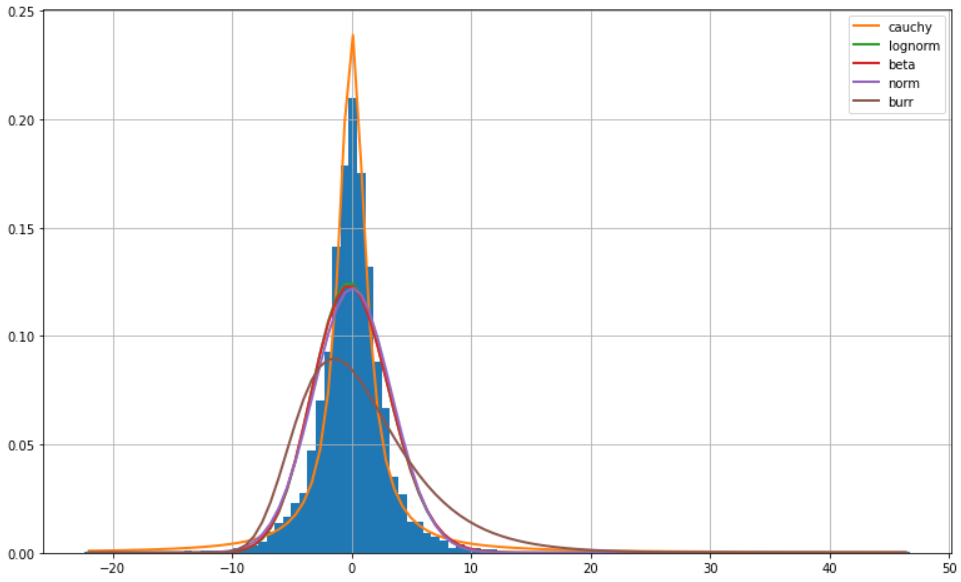
Далее воспользуемся встроенной питоновской функцией Fitter(), которая по подаваемому ей на вход списку названий распределений определяет, насколько вероятно, что данные подчиняются каждому из перечисленных распределений.

Сначала применим данную функцию к данным акции IMO:



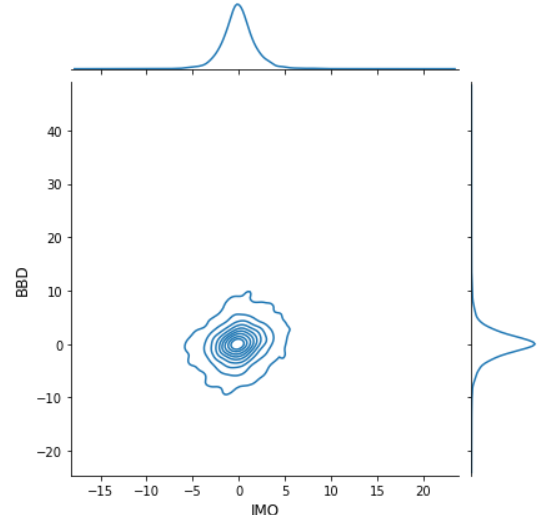
Итак, мы видим, что данные по ежедневным доходностям акции IMO, вероятнее всего, подчиняются распределению Заусенцев с параметрами x = 35.12657387945953, c = 37.16212848306275, k = 0.878834202573739, λ = −35.018063931325656.

Сделаем то же самое для ряда доходностей BBD:



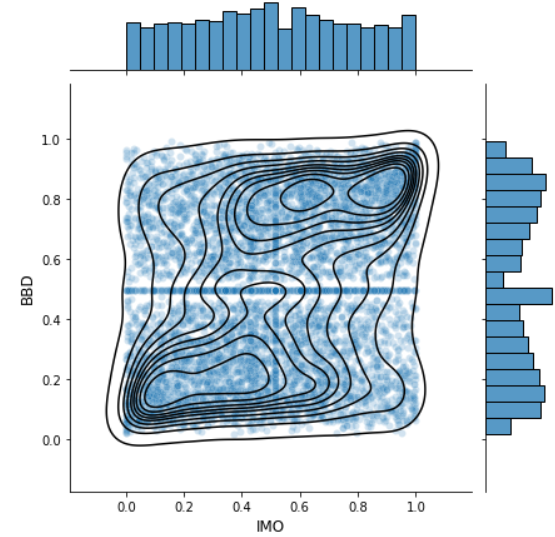
Ежедневные доходности акции BBD, в свою очередь, вероятнее всего подчиняются распределению Коши с параметрами x0 = 0.017072134826404222 и γ = 1.3268151095776355.

Итак, теперь взглянем на маргинальные и совместное распределения вместе:



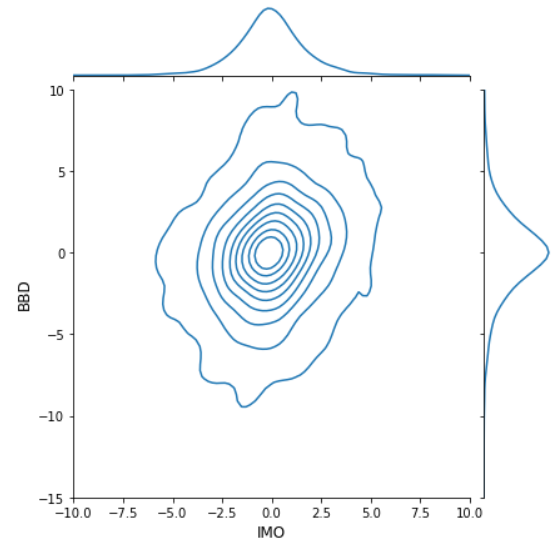
Мы знаем, или, хотя бы, примерно догадываемся, по каким законам распределены доходности наших акций, а значит, можем преобразовать их в равномерно распределенные случайные величины.

Сделав это, получим следующий график:

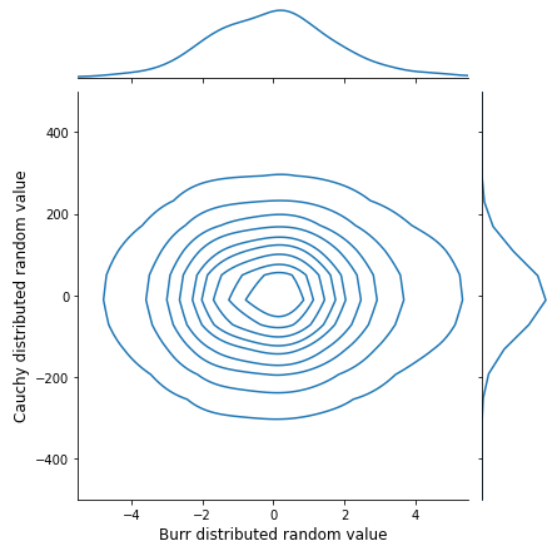


Совместный график, полученный выше, и есть копула.

Теперь преобразуем маргинальные распределения обратно, чтобы увидеть, что такое преобразование действительно обратимо и сохраняет в себе информацию о структуре связи между маргинальными распределениями, не меняя их:



Для большей наглядности сравним полученное распределение с совместным распределением без корреляций, т.е. совместным распределений двух случайных величин, сгенерированных независимо друг от друга из тех же распределений:



Таким образом, можно подвести итог, что копулы действительно работают.

Также мы можем не делать никаких предположений о распределении данных, т.е. не использовать fitter.Fitter(), а попробовать применить к данным некоторые распространенные копулы, реализованные, например, в библиотеке copulae, и посмотреть на результат.

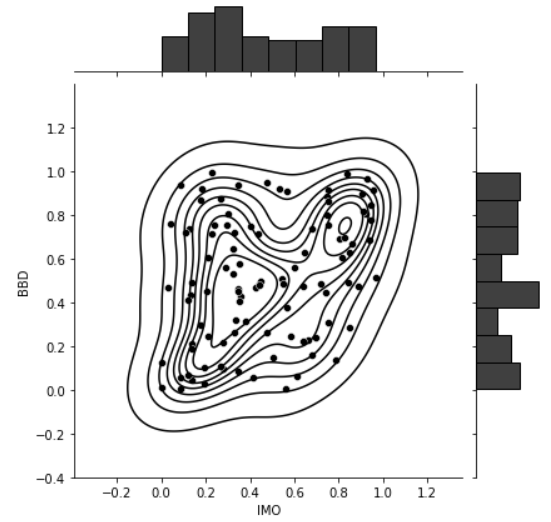
Начнем с самой распространенной копулы, Гауссовской.

Обучив такую копулу на наших данных, получим для нее следующий оцениваемый параметр, который является коэффициентом корреляции для Гауссовской копулы:

Изображение выглядит как текст

Автоматически созданное описание

Попробуем применить обученную копулу. Для этого сгенерируем с ее помощью 2 равномерно распределенные случайные величины, которые должны иметь ту же зависимость двумерной плотности, что и изначальные, и посмотрим на их совместное распределение (по сути, визуализируем копулу):



На всякий случай проверим, что полученные при помощи копулы случайные величины распределены равномерно на отрезке от 0 до 1:

Изображение выглядит как стол

Автоматически созданное описание

И снова видим, что копула действительно работает.

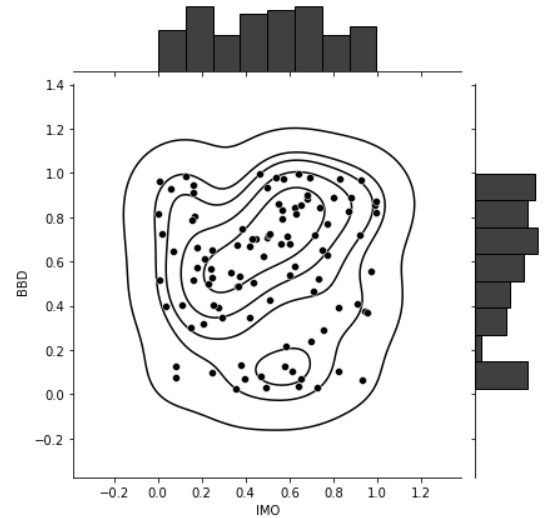
Повторим все то же самое для копулы Стьюдента.

Аналогично, подгоним выбранную копулу под данные и вновь посмотрим на оцениваемые параметры. На этот раз, это коэффициент корреляции, а также степень свободы t:

Изображение выглядит как текст

Автоматически созданное описание

Визуализируем полученную копулу:



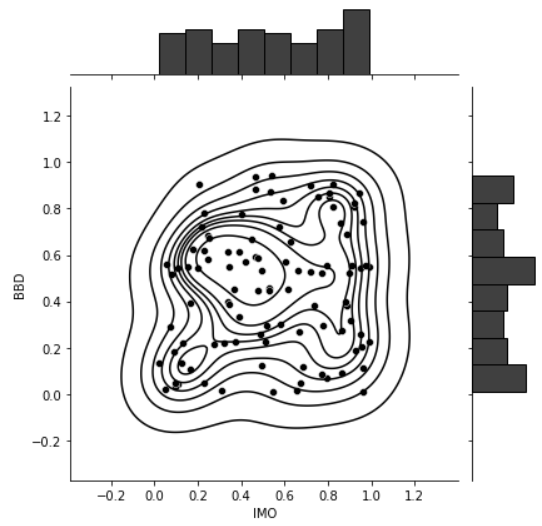
Теперь используем для данных копулу Клейтона.

Соответствующий оцениваемый параметр:

Изображение выглядит как текст

Автоматически созданное описание

Полученная копула:



Таким образом, мы попробовали применить к нашим данным некоторые типы копул, но неясно, дает ли копула, которая имеет хорошее соответствие, хорошее соответствие для распределения данных.

Итак, мы увидели, что доходность наших финансовых активов не представляют собой независимые одинаково распределенные случайные величины (independent and identically distributed random variables, IID), как предполагают многие приложения в финансах.

Хорошо известно, что финансовые доходы подчиняются сложной и постоянно меняющейся функции распределения вероятностей, и мы можем надеяться статистически смоделировать лишь очень небольшую часть распределения доходов в любой момент времени.

Мы попытались эмпирически выявить такие стилизованные факты о рядах доходностей, как кластеры волатильности, тяжелые хвосты и нелинейная зависимость, с использованием ряда статистических методов.

Теперь мы можем попробовать применить к рядам доходностей статистические методы прогнозирования.

## **Прогнозирование временных рядов**

Ниже мы рассмотрим некоторые самые распространенные модели, используемые для прогнозирования временных рядов, и применим их к ряду дневных доходностей S&P/TSX Composite Index.

## **Проверка стационарности**

Чаще всего ряд доходностей (ряд первых разностей) стационарен, вследствие чего нет необходимости обосновывать применимость тех или иных методов. Посмотрим, является ли ряд дневных доходностей S&P/TSX Composite Index стационарным, для этого применим к нему расширенный тест Дикки-Фуллера (аugmented Dickey-Fuller Testing), который является наиболее распространенным тестом на единичный корень (unit root testing).

Нулевая гипотеза, в этом случае, состоит в том, что ряд содержит единичный корень, а альтернативная - в том, что ряд является стационарным.

Изображение выглядит как текст

Автоматически созданное описание

Так как уровень значимости значительно меньше 0.05 и даже 0.01, у нас есть основание отвергнуть нулевую гипотезу, а значит, мы принимаем альтернативную гипотезу, которая гласит, что временной ряд стационарен.

Теперь, еще раз убедившись в стационарности наших данных, перейдем к построению моделей.

## **ARMA модель**

Модель авторегрессии — скользящего среднего (autoregressive moving-average model, ARMA) - это совмещение более простых моделей AR (авторегрессионной модели, или autoregressive model) и MA (модели скользящего среднего, или moving average model).

Модель AR(L2) пытается объяснить эффекты импульса, то есть взаимосвязи с данными за прошлые периоды, часто наблюдаемые на торговых рынках (эффекты участников рынка):

Изображение выглядит как часы, антенна, датчик

Автоматически созданное описание

Модель MA(L1) пытается отразить шоковые эффекты, наблюдаемые в терминах белого шума. Эти шоковые эффекты можно рассматривать как неожиданные события, влияющие на процесс наблюдения, такие как внезапные заработки, войны, кризисы и прочее:

Изображение выглядит как текст, часы, антенна

Автоматически созданное описание

В обеих формулах выше случайные шоки, или остатки, ϵt являются независимыми одинаково распределёнными случайными величинами, причем ϵt ∼ N(0, σ2).

Модель ARMA(L2, L1) пытается охватить оба этих аспекта при моделировании финансовых временных рядов. Но она не принимает во внимание кластеризацию волатильности, ключевой эмпирический феномен многих финансовых временных рядов:

Изображение выглядит как текст, антенна

Автоматически созданное описание

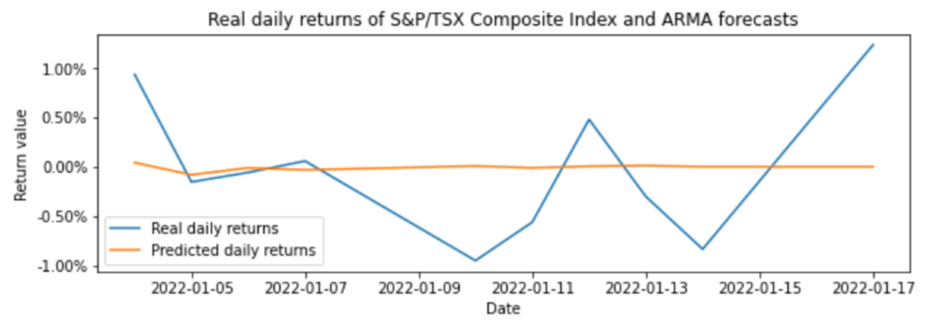
Итак, приступим к обучению первой модели.

Изображение выглядит как стол

Автоматически созданное описание

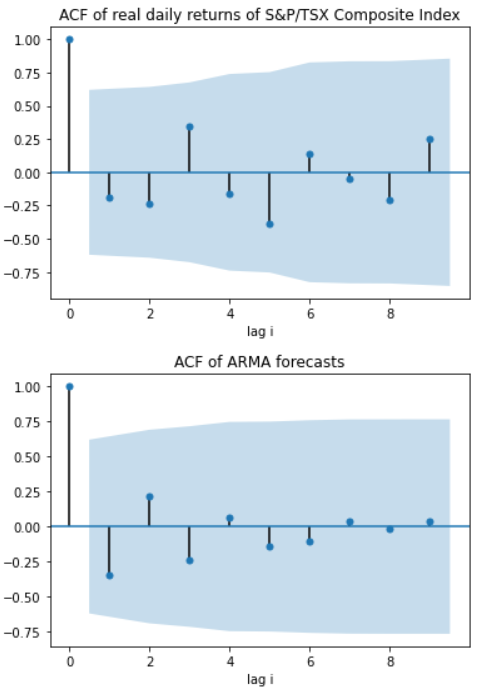
Мы задали модель, после чего обучили ее на имеющихся данных за прошлые периоды, используя метод максимального правдоподобия (задали соответствующий параметр при обучении). В таблице выше приведены полученные по результатам обучения модели ARMA(2, 2) коэффициенты, все они по модулю не превышают 1.

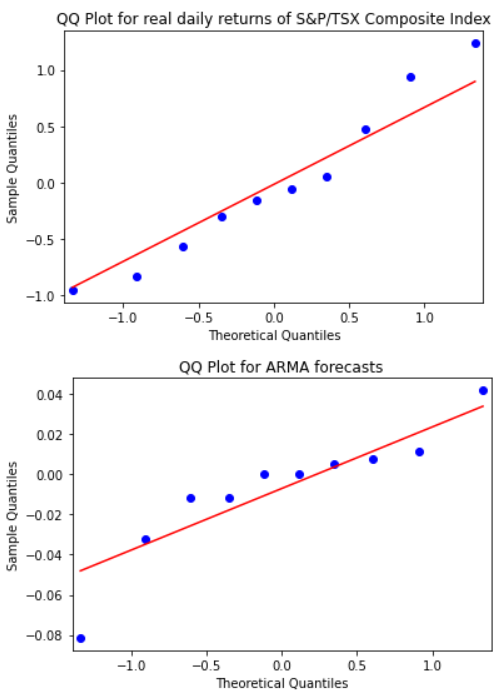
Сравним полученные прогнозы ежедневной доходности индекса для первых дней 2022 года (для 10 дней) с реальными значениями.

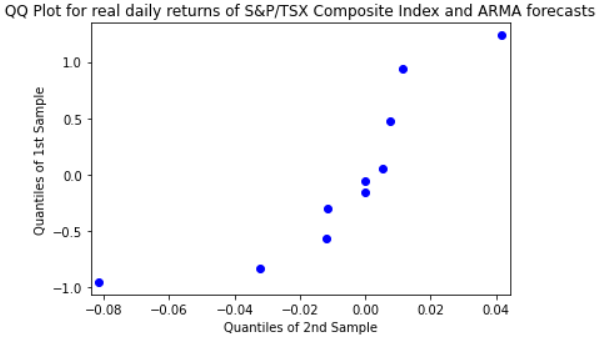


На графике четко видно, что предсказательная способность обученной модели оставляет желать лучшего, а точнее, она очень и очень плохая.

Попробуем взглянуть на автокорреляционные функции (и их доверительные интервалы) и графики квантиль-квантиль для реальных и спрогнозированных данных, соответственно.







Полученная модель уловила некоторые (совсем чуть-чуть) взаимосвязи данных во времени, что видно по автокорреляционной функции, однако предсказанные ею значения для первых 10 дней 2022 года не просто значительно отличаются от реальных, но и имеют другое распределение, что видно, в том числе, из приведенных выше QQ Plots.

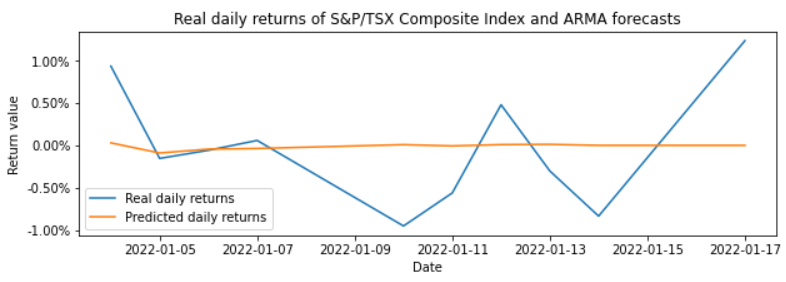
Попробуем подобрать значения L1 и L2, а не задавать их самостоятельно, чтобы улучшить результаты модели. Для этого будем перебирать значения для L1 и L2 от 1 до 5 и оценивать получаемые результаты модели, обученной на каждой паре значений, по критерию AIC (Akaike Information Criterion), который автоматически подсчитывается для обученной модели.

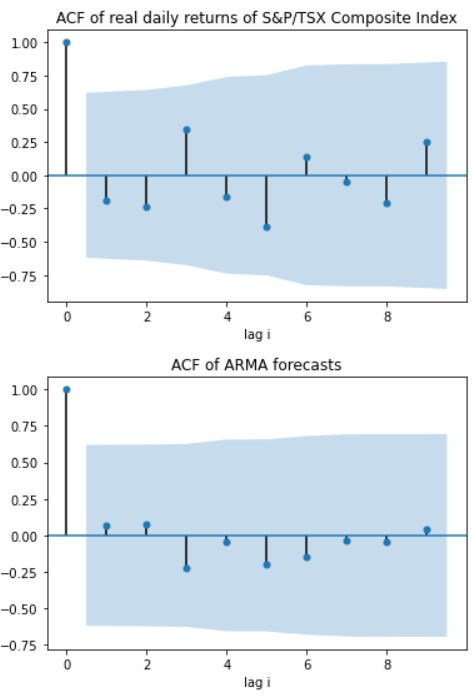
Обучив модели со всеми парами значений для L1 и L2, мы получили лучшее соотношение: L1 = 4 и L2 = 4.

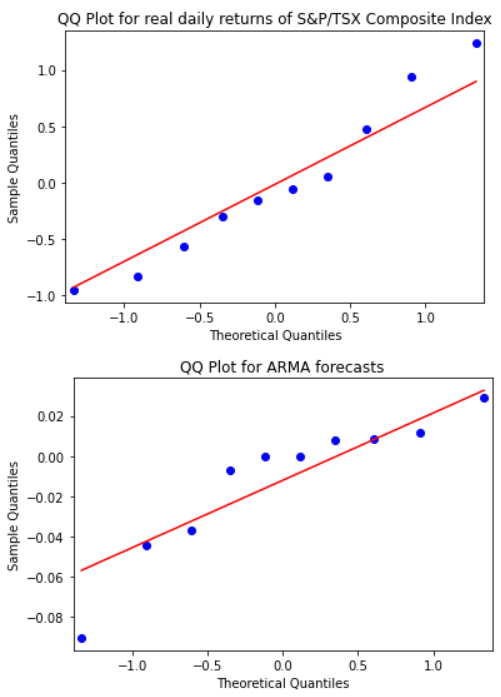
Изображение выглядит как стол

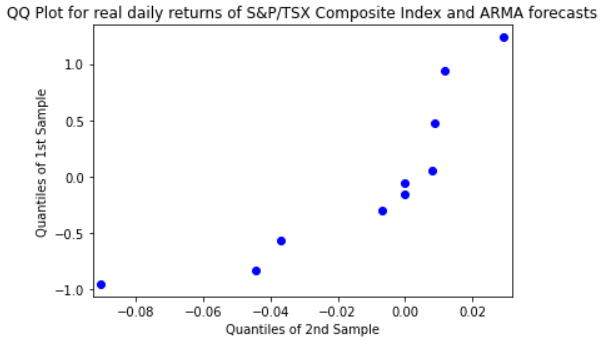
Автоматически созданное описание

Посмотрим на графики, аналогичные приведенным выше, для лучшей модели.









Качество предсказаний по-прежнему страдает, хотя и можно заметить чуть больше совпадений корреляционных функций для реальных и предсказанных значений.

Применим к остаткам лучшей модели уже опробованный нами ранее тест Льюнга-Бокса, чтобы наверняка определить, было ли достигнуто хорошее соответствие для L1 = 4 и L2 = 4. Если p-значение теста будет больше требуемой значимости, можно сделать вывод, что остатки независимы и являются белым шумом.

Проведя тест, мы получили значение p-value, равное 0.00014850656311690398.

Как можно было ожидать, p-значение значительно меньше 0.05, что говорит нам о том, что остатки модели ARMA(4, 4), обученной на исторических доходностях S&P/TSX Composite Index, не являются дискретным белым шумом, а значит, в остатках присутствует дополнительная автокорреляция, которая не объясняется подобранной моделью ARMA(4, 4). Это видно и из графика остатков, приведенного выше, мы видим области явной условной волатильности (гетероскедастичности), которые модель не уловила (возможно, по этому все же не стоит судить, так как мы решили предсказать данные лишь на 10 дней вперед).

Попробуем обучить другую модель и посмотреть на ее результаты.

## **ARCH модель**

Модель авторегрессионной условной гетероскедастичности (AutoRegressive Conditional Heteroscedasticity, ARCH) использует в качестве шоков волатильности наблюдаемые значения доходности, или остатки, и выглядит следующим образом:

Изображение выглядит как текст, часы

Автоматически созданное описание

где Zt - это белый шум, а L1 - количество лагов.

Важно отметить, что как условное, так и безусловное математическое ожидание такого процесса равно нулю.

Децентрировав временной ряд, так же перебором подберем наиболее оптимальное значение для L1 в диапазоне от 1 до 5 и построим наилучшую модель:

Изображение выглядит как стол

Автоматически созданное описание

Проверим, что выполняются следующие важные ограничения, накладываемые на параметры ARCH(L1) модели:

* 

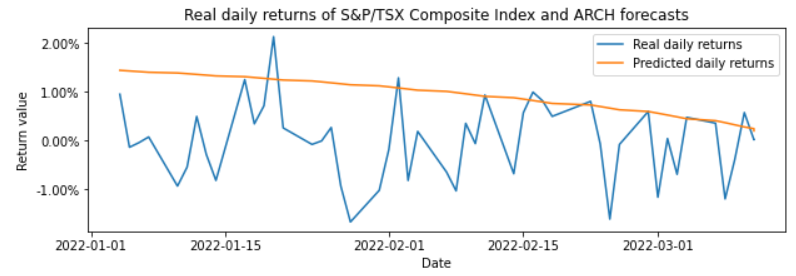
что обеспечивает положительность прогнозов волатильности

* 

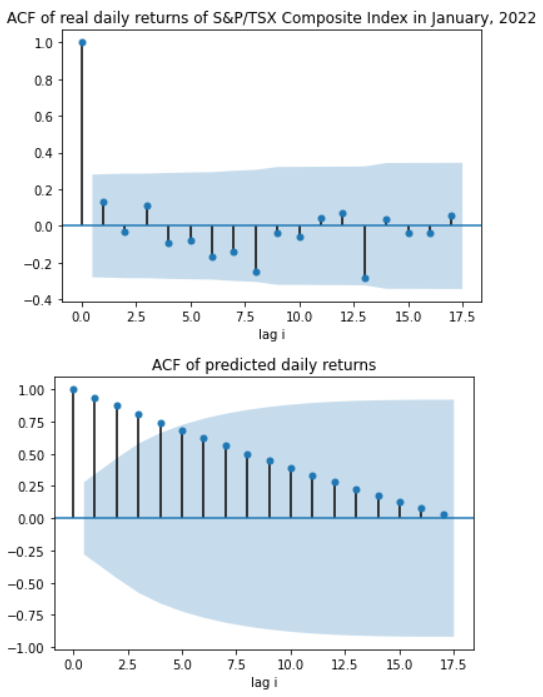
что обеспечивает стационарность ковариации, чтобы определить безусловную волатильность.

Также мы можем вспомнить о правиле, которое гласит, чем выше альфы, тем толще хвосты. Все соответствующие коэффициенты нашей обученной модели лежат в интервале от 0.02 до 0.3, из чего мы можем сделать предположение, что распределение данных, которые будет прогнозировать полученная модель, будет иметь не сильно толстые хвосты.

И снова посмотрим, как предсказания модели соотносятся с реальными данными:



Построим автокорреляционные функции:



## **GARCH модель**

ARCH-модель предполагает зависимость условной дисперсии только от квадратов прошлых значений временного ряда, однако, можно обобщить данную модель, предположив, что условная дисперсия зависит также от квадратов прошлых значений самой условной дисперсии, таким образом, мы сможем учесть влияние исторической доходности. Такая модель называется обобщённым ARCH (Generalized ARCH — GARCH(L1, L2)) и описывается следующим образом:

Изображение выглядит как текст, антенна

Автоматически созданное описание

Перебрав значения для L1 и L2 от 1 до 5, получили для наших данных лучшую модель с соответствующими характеристиками:

Изображение выглядит как стол

Автоматически созданное описание

Как и в модели ARCH(L1), на параметры модели GARCH(L1, L2) накладываются два типа ограничений, которые, как можно видеть из таблицы выше, выполняются для реализованной модели:

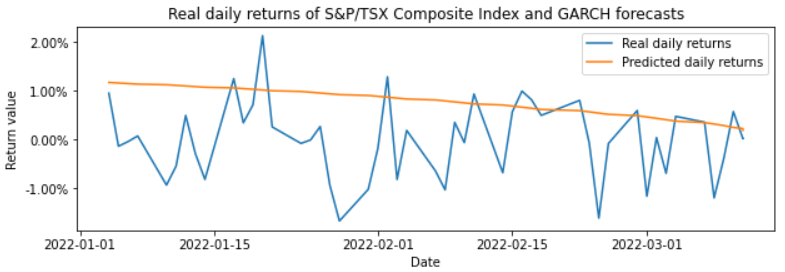
1. Для обеспечения положительных прогнозов волатильности:



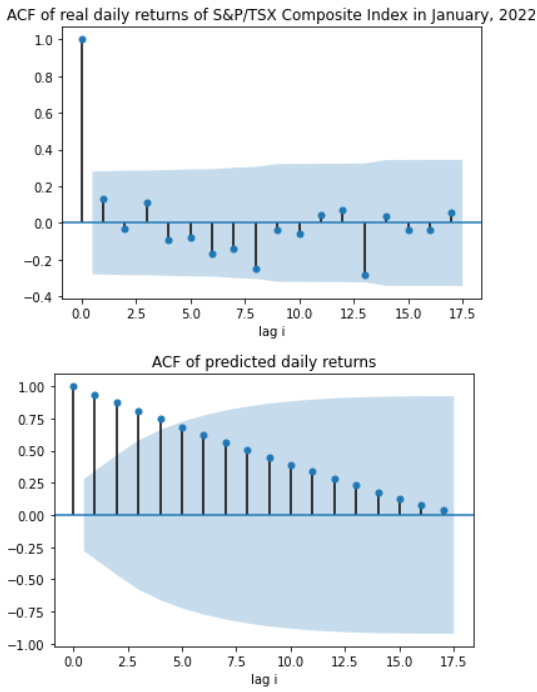
1. Для обеспечения стационарности ковариации:



Посмотрим на предсказания модели:



И вновь на ACF:



## **Сравнение моделей**

Мы построили некоторые модели, теперь попробуем продиагностировать их вместе. Но перед этим построим еще 2 модели типа GARCH(L1, L2), на этот раз уже будем использовать распределение Стьюдента-t.

Параметры всех таблиц приведены в таблице ниже:

Изображение выглядит как стол

Автоматически созданное описание

Из таблицы выше можно увидеть, что сумма коэффициентов для большинства обученных моделей близка к 1:



Как мы знаем, при таких больших значениях коэффициентов появляются проблемы с безусловной волатильностью.

Посмотрим на результаты тестов отношения правдоподобия (likelihood ratio tests) оценочных моделей в таблице ниже.

Изображение выглядит как текст

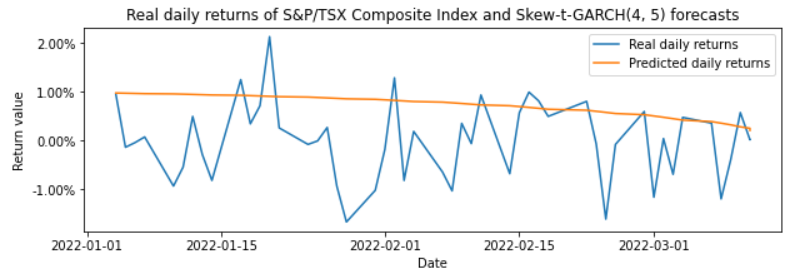
Автоматически созданное описание

Из представленной таблицы видно, что ARCH(5) значительно лучше ARCH(1), что неудивительно, поскольку мы ожидаем, что зависимость от волатильности продлится много дней.

При добавлении запаздывающей волатильности в модели GARCH(4, 5) соответствие между моделями резко улучшаться, и если сравнивать GARCH(4, 5) с GARCH(1, 1), они оказываются гораздо более схожими (на это, конечно же, также влияют сами данные, под которые подгоняются модели).

Использование Стьюдента-t в качестве условного распределения также значительно улучшает обычную версию модели GARCH, в то время как асимметричное распределение Стьюдента-t оказывается еще лучше, чем симметричное.

Наконец, посмотрим на соотношение предсказаний лучшей модели с реальными данными:

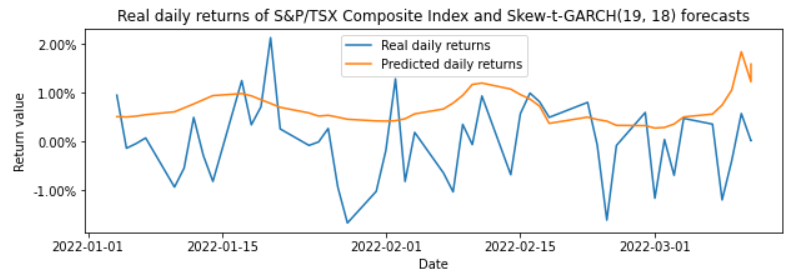


На графике видно, что значения предсказываются уже гораздо лучше, однако все равно далеко не идеально. Для лучшего понимания, добились ли мы успеха, можно посмотреть, для какого число дней был предсказан верный знак доходности (то есть будет она отрицательной или положительной), потому что саму доходность, которая является дробным числом с большим количеством знаков после запятой и содержит в себе нелинейные закономерности, как мы уже выяснили раньше, предсказывать все же непросто, из чего можно сделать вывод, что, возможно, нам и не стоит требовать от модели предсказывать доходности с точностью до какого-то определенного знака после запятой.

Вычисление такой метрики покажет, что модель предсказывает с точностью до знака 51% значений доходностей, что является неплохим результатом.

Попробуем увеличить для лучшей модели число лагов L1 и L2 и обучим ее на данных только за 2021 год.

Значение максимального правдоподобия выросло с -1002.941648 до -231.51, а график с предсказаниями принял следующий вид:



Можно увидеть, что модель не стала предсказывать доходности значительно лучше, однако ее график все же приблизился к истинному, а это значит, что мы добились еще лучшего результата.

## **Заключение**

Итак, мы проанализировали ряд доходностей канадского индекса S&P/TSX и посмотрели на предсказания будущих значений этого ряда, для получения которых обучили несколько статистических моделей.

Таким образом, мы достигли поставленных целей, подробно изучив выбранный нами финансовый ряд и применив к нему знания, полученные во время прохождения курса "Анализ и прогнозирование финансовых рисков".

## **Список литературы**

* 1. R. A. Fisher (1925), Statistical Methods for Research Workers, Oliver & Boyd, London States
  2. Sklar, A. (1959), ‘‘Fonctions de reґpartition a` n dimensions et leurs marges,’’ Publ. Inst. Statis. Univ. Paris, 8, 229–231.