Федеральное государственное образовательное бюджетное

учреждение высшего профессионального образования

«**Финансовый университет**

**при Правительстве Российской Федерации»**

**(Финансовый университет)**

**Департамент анализа данных, принятия решений и финансовых технологий**

**Курсовая работа**

**на тему:**

**«Моделирование экстремальных событий на финансовом рынке»**

**Выполнила:**

студентка группы ПМ 3-4

Чиркова Елена

**Научный руководитель:**

к.н., доцент   
Иванюк Вера Алексеевна

**Москва 2020**

**Содержание**

**1. Введение**

**2. Основная часть.**

**2.1. Теоретическая часть**

**2.1.1. План работы**

**2.1.2. Описание моделей**

**2.2. Практическая часть**

**2.2.1. Тесты на выбор модели**

**2.2.2. Оценивание моделей**

**2.2.3. Проверка состоятельности**

**2.2.4. Оценка шоков**

**2.3. Прогнозирование.**

**2.3.1. Описание процесса**

**2.3.2. Прогнозирование**

**4. Заключение**

**5. Источники**

**6. Приложения**

1. Введение.

Рынок – вне всякого сомнения нестабильный и неидеальный механизм, находящийся в постоянном движения в ответ на внешние обстоятельства. Множество процессов, происходящих на финансовом рынке, в ом числе его цикличность, уже изучено человечеством. Выявленные взаимосвязи используются во множестве различных направлений деятельности во всем мире. Однако ввиду не изученности всех факторов, на рынок влияющих, невозможно предсказать движения рынка со 100% уверенностью. Притом, иногда непредвиденные скачки рынка приводят к таким экстремальным ситуациям, когда целые государства получают убытки в триллионы долларов, что в принципе могло бы поставить под сомнение использование данного механизма ввиду его рискованности. Однако в наш век, нельзя и пренебрегать привлекательностью финансового рынка не представляется возможным из-за того, насколько глубоко он проник во все мировые структуры. Разумный выход – глубокое изучение упомянутых выше экстремальных событий (шоков) с целью подготовки к ним и даже возможного использования.

Глобальный кризис 2008-2009 гг., последствия кризиса, сбой обеспечения равновесия на макроэкономическом уровне, неспособность мировых держав «быстро исправить ситуацию» – все это явилось главным показателем того, насколько данный феномен не изучен и не предсказуем. И выявило необходимость поиска новых концепций для описания закономерности и природы финансовых шоков.

Таким образом, изучение скрытых закономерностей в возникновении шоков рынка является важной и тяжело осуществимой задачей, привлекающей внимание исследовательских сообществ в области математики, экономики и data science. Поиск путей прогнозирования шоков рынка – трудная, но осуществимая задача. И ее решение способно обнаружить точки неэффективности рынка с целью дать сигнал в случае экстремальной ситуации, а также вывести на новый уровень мировые финансовые системы.

Предсказуемы ли шоки на коротких таймфреймах на финансовом рынке? На этот вопрос я хотела бы ответить в своей работе. Используя пятиминутный таймфрейм для индекса РТС, смоделировать его «движение» и по возможности предсказать его дальнейшее изменение и вероятность шока. Для этого сначала необходимо выбрать модель, которую будем использовать для обучающей выборки по индексу. Далее с помощью нейронных сетей прогнозировать изменение.

2.1. Теоретическая часть

2.1.1. План работы.

Мы собираемся использовать для оценки выгруженные значения индекса РТС. Это данные по определению являются временным рядом, значит, для моделирования будем использовать соответствующие типы моделей.

Выбор модели для описания ряда – одна из главных задач работы. Ошибка в выборе модели может повлечь за собой неудовлетворительные результаты, несостоятельность спецификации модели и, самое главное, неверную интерпретацию результатов.

Для описания временных финансовых рядов используются  авторегрессионые модели, интегральные модели и модели скользящего среднего. Чтобы понять какую модель выбрать, необходимо провести ряд тестов:

- тест на автокорреляцию ряда (на белый шум) – теста Бокса-Льюнг (Q-тест)

- тест на стационарность ряда - тест Дики-Фуллера ( DF/ADF-тест)

- тест на стационарность ряда (и на стационарность вокруг тренда) – тест KPSS

Цель анализа временного ряда – определить от каких компонент зависит последующее значение ряда и определить модель для каждой из этих компонент. Например, модель ARMA (Autoregressive Moving Average – модель авторегрессии - скользящего среднего) линейно совмещает авторегрессионый характер ряда доходностей и скользящую среднюю остатков.

Стоит так же обратить внимание на модели типа ARCH (Autoregressive Conditional Heteroskedasticity – Авторегрессинная условная гетероскедастичность), широко используемая для моделирования финансовых рядов. Данная модель занимается объяснением кластеризации волатильности (подразумевает зависимость условной дисперсии от прошлых значений ряда). Боллерслев Т.П. в своей работе (1986 г) выводит обобщение ARCH-модели: Generalized ARCH – GARCH, которая в свою очередь, помимо особенностей основной модели, учитывает так же зависимость условной дисперсии и от прошлых значений дисперсии. В связке ARMA(p,q)-GARCH(r,s) метод GARCH(r,s) дает возможность отразить динамику волатильности метода ARMA(p,q).

Будем выбирать из моделей ARMA(ARIMA), ARMA-GARCH подходящую модель для выбранных данных.

Шоки на финансовом рынке определим как непредвиденные движения рыночной доходности, которая, в свою очередь, может быть определена с помощью модели ценообразования.

Определение может быть записано в следующем виде:

Допустим, модель доходности может быть оценена с помощью функции f:

где вектор состоит из факторов, доступных в момент времени t – 1, и распределение ошибки в момент t.

Тогда, шок на финансовом рынке в момент t определяется как:

Далее определим параметры модели p,q с помощью ACF и PACF, оценим модель ARMA и проведем тесты:

- тест на порядок временного ряда - тест Филлипса-Перрона (PP-тест), чтобы проверить, насколько верно выбраны параметры, с использованием ACF, PACF

- тест на автокорреляцию остатков ряда (на белый шум) – теста Бокса-Льюнг (Q-тест) для модели ARMA.

- тест на гетероскедастичность случайных ошибок – тест Уайта. При наличии гетероскедастичности оценим модель GARCH для условных дисперсий ряда.

После оценки и проверок сравним значения следующих тестов для моделей ARMA (ARMA-GARCH):

-тест на проверку того, что выбранные данные соответствует [предполагаемой модели](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9) – критерий согласия Колмогорова.

После расчета шоков (знаков шоков) с помощью выбранной модели, исследуем их прогнозируемость, применяя метод машинного обучения, задачей которого, в сущности, будет оптимизация функции g():

, (3)

где Ω - факторы, связанные с нелинейными составляющими, такими как технические показатели, показатели внутридневной изменчивости и др.,  
а также минимизация ошибки прогнозирования D:

Схематичное описание всего процесса работы в Приложении 4.

2.1.2. Описание моделей

ARMA(p,q):

Данная модель - модель авторегрессии AR(p), обобщенная с моделью скользящего среднего MA(q). p и q – целые числа, которые задают порядок модели.

Для доходности в период времени t, общий вид модели ARMA(p,q):

Где с – константа, – последовательность независимых величин с нулевым средним, которые одинаково распределены. – коэффициенты скользящего среднего ( и авторегрессионные коэффициенты .

GARCH(r,s):

Это модель, часто применяемая для оценки финансовых рядов ввиду того, что она подразумевает зависимость условной дисперсии текущего периода от предыдущих ее значений и от предыдущих значений ряда. Эта модель применяется еще также для кластеризации волатильности.

Для ряда доходностей в период времени t, общий вид модели условной дисперсии GARCH(r,s) имеет вид:

Здесь сумма всех коэффициентов должна быть меньше единицы. Иначе – интегрированный IGARCH.

Безусловная дисперсия GARCH(r,s):

Описание данных.

Данные взяты по индексу РТС в качестве показателя изменения российского финансового рынка в целом. Для проверки реакции на нисходящий тренд взяты данные с 21.01.2020 по 05.03.2020.

ARMA(p,q)-GARCH(r,s):

Для доходности в период времени t, общий вид модели ARMA(p,q)-GARCH(r,s) представляется следующим образом:

; (4)

,

Здесь модель (4) иллюстрирует ARMA компонент, а модели (5) и (6) – GARCH компонент.  
 – авторегрессионный ряд доходностей, для определения которого используются данные по индексу за последние p периодов.

- значения скользящей средней за последние q периодов.

– коэффициенты модели.

с, – константы.

Автокорреляция отражается в компоненте GARCH, то есть в (5), (6).

– шоки модели ARMA-GARCH.

Вообще говоря, порядки моделей (p, q, s, m) определяются уровнем значимости, но в чаще всего принимаются равными единице. Такая модель использовалась также Jacob F. (2015 г) в работе, посвященной высокочастотным данным. Принимая порядки равными единице, получаем модель оценки шока рынка:

Модель GARCH не подходит для нашего эксперимента, так как не подразумевает построение модели регрессии самого временного ряда.

2.2. Практическая часть

2.2.1. Тесты на выбор модели.

1 – Тест Бокса-Льюнг (Q-тест)

Данный тест предназначен для выявления автокорреляции временных рядов. Даже при ненормальности распределения тест продолжает быть состоятельным.

Выдвигаются гипотезы:  
 – данные выборки являются белым шумом (являются случайными).

– данные выборки не являются случайными.

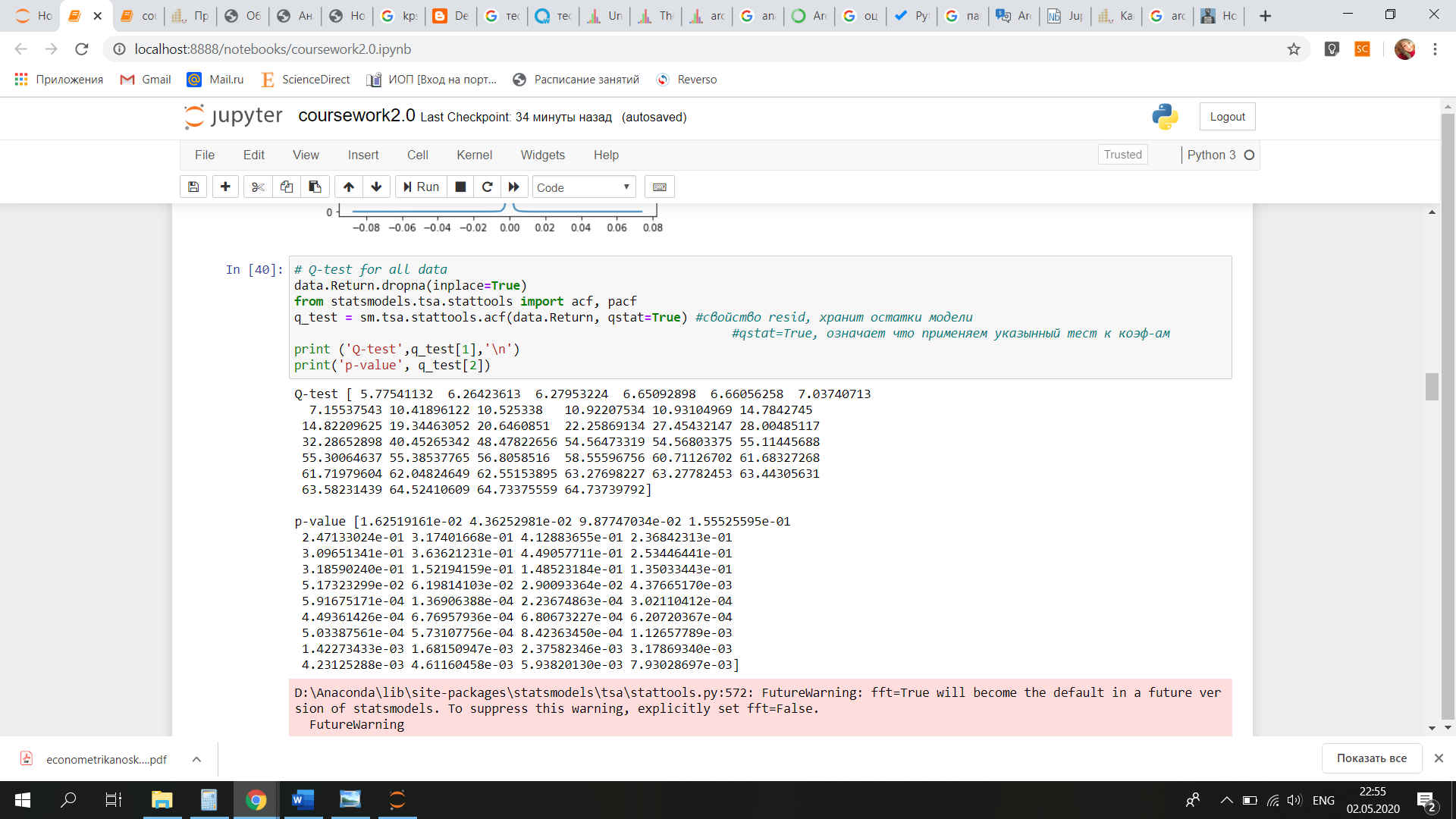
Статистика для теста:

где n – число наблюдений в выборке, – автокорреляция k-ого порядка, m - число проверяемых лагов.

Если

где – квантили хи-квадрат распределения с m степенями свободы, то нулевая гипотеза отвергается, и соответственно признается автокорреляция во временном ряду m-ого порядка.

Результат теста для имеющейся выборки:



Нулевая гипотеза отвергается, следовательно данные выборки не являются белым шумом, и имеет место автокорреляция.

2 – Тест Дики-Фуллера (DF-тест)

Тест для проверки ряда на стационарность, на наличие, так называемых, единичных корней (unit root).

Ряд имеет единичный корень (порядок интеграции один), если это разности , образуют стационарный ряд.

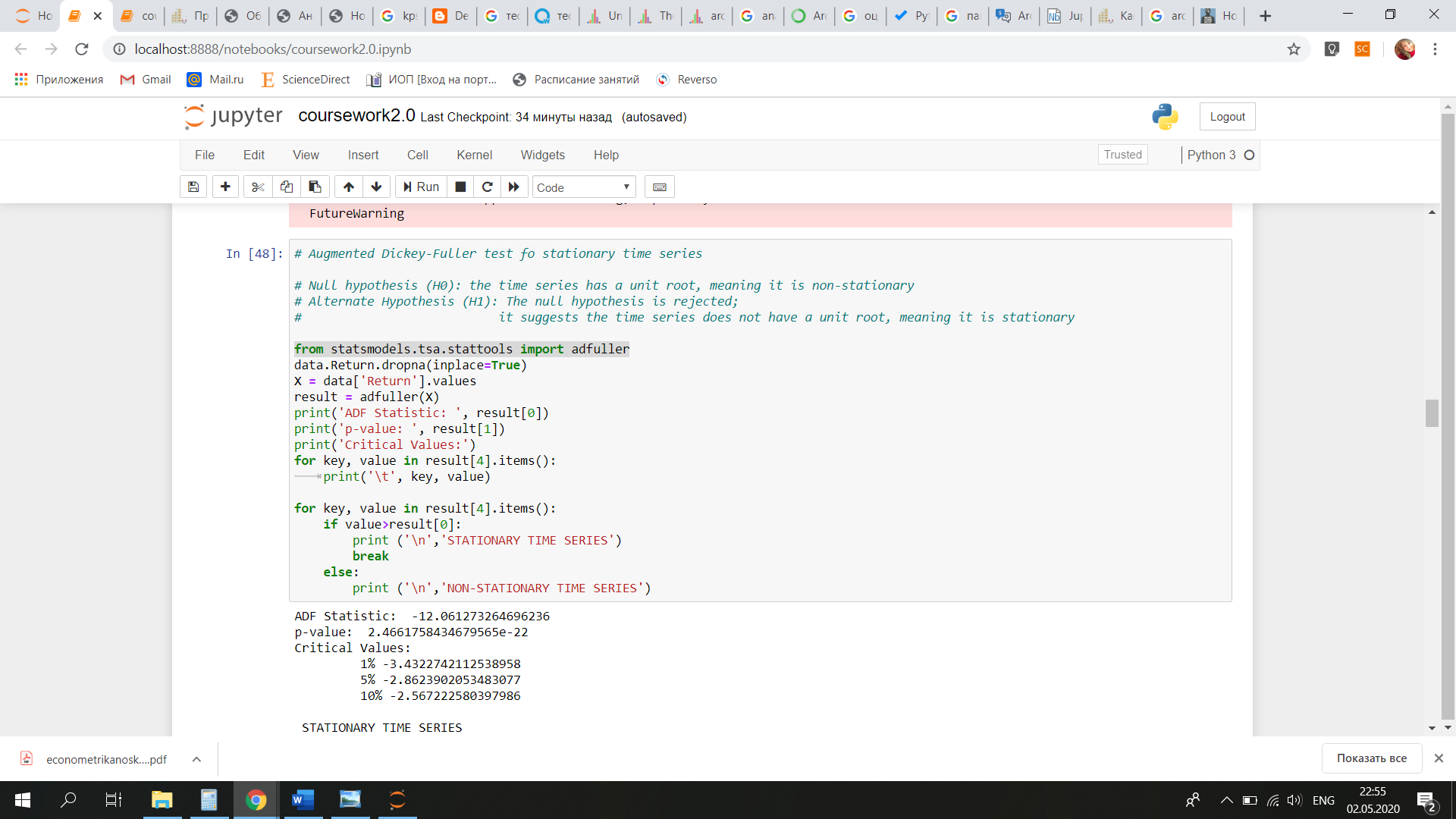
Статистика для теста:

Где - оператор разности первого порядка

Выдвигаются гипотезы:  
 – a = 1, в регрессии ряда присутствует единичный корень, ряд не стационарный.

– , ряд стационарный.

Значение |a|>1 для финансово-экономических рядов не свойственно, так как при этом ряд является «взрывным», что ввиду инертности этой среды очень мало вероятно.  
Результаты теста для имеющейся выборки:



Нулевая гипотеза отвергается, ряд является стационарным.

3 - Тест KPSS

Тест является дополнением к тесту Дики-Фуллера и предполагает возможность стационарности вокруг некоторого тренда

Выдвигаются гипотезы:  
 – ряд стационарен вокруг некоторого тренда

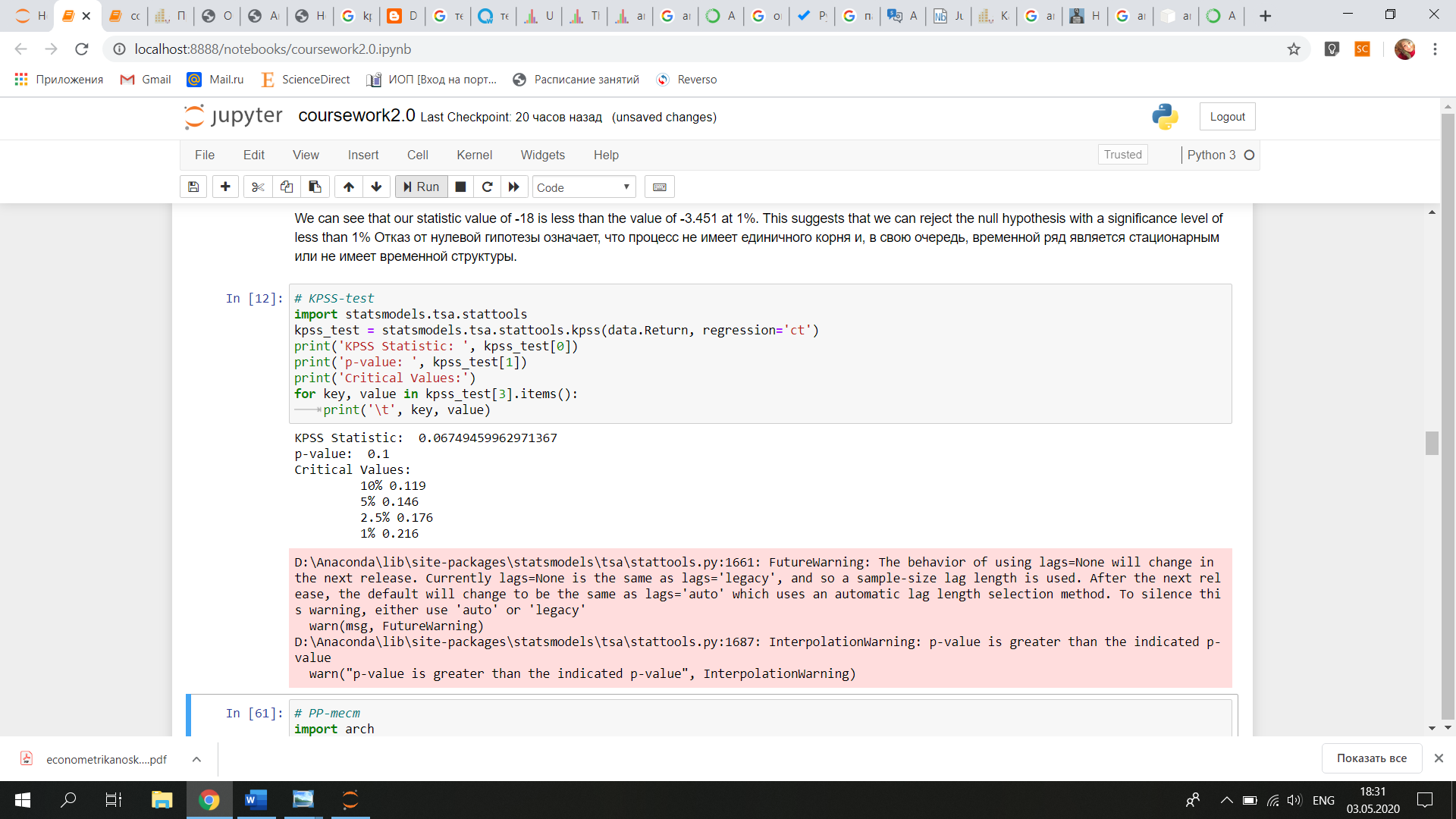
– существует единичный корень

Статистика для теста:

Где – детерминированный ряд, - случайные блуждания, – стационарная ошибка.

Нулевой гипотезе соответствует значение дисперсии , альтернатива - .

Результат теста для выбранных данных:



Нулевая гипотеза отвергается, нельзя утверждать о стационарности вокруг некоторого тренда.

2.2.2. Оценивание моделей

Начнем с визуального анализа. График 1. иллюстрирует изменение входных данных во времени.

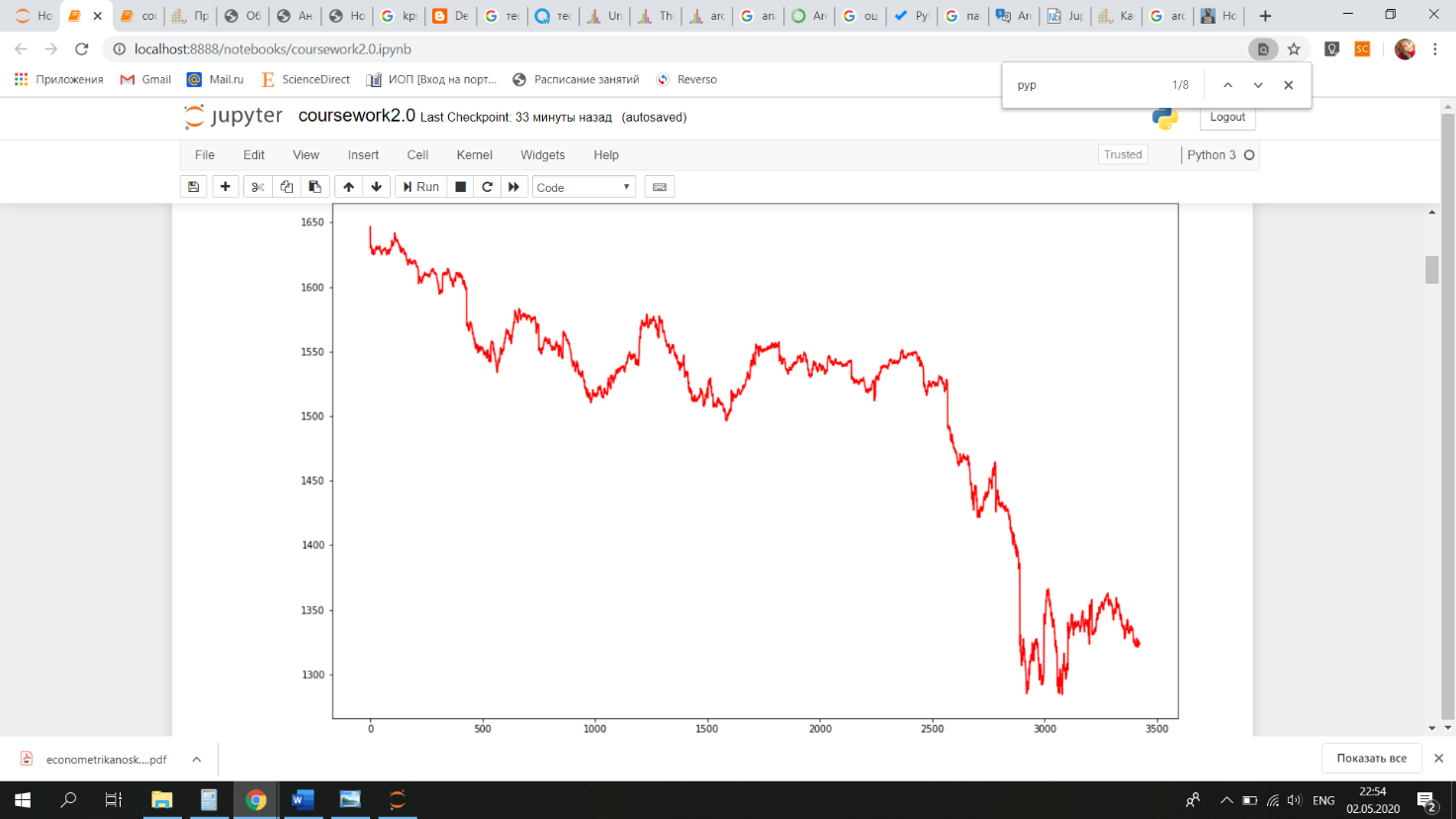


График 1. Изменение индекса РТС.

Заметим нисходящий тренд, однако нас интересуют доходности. Их можно увидеть на Графике 2.

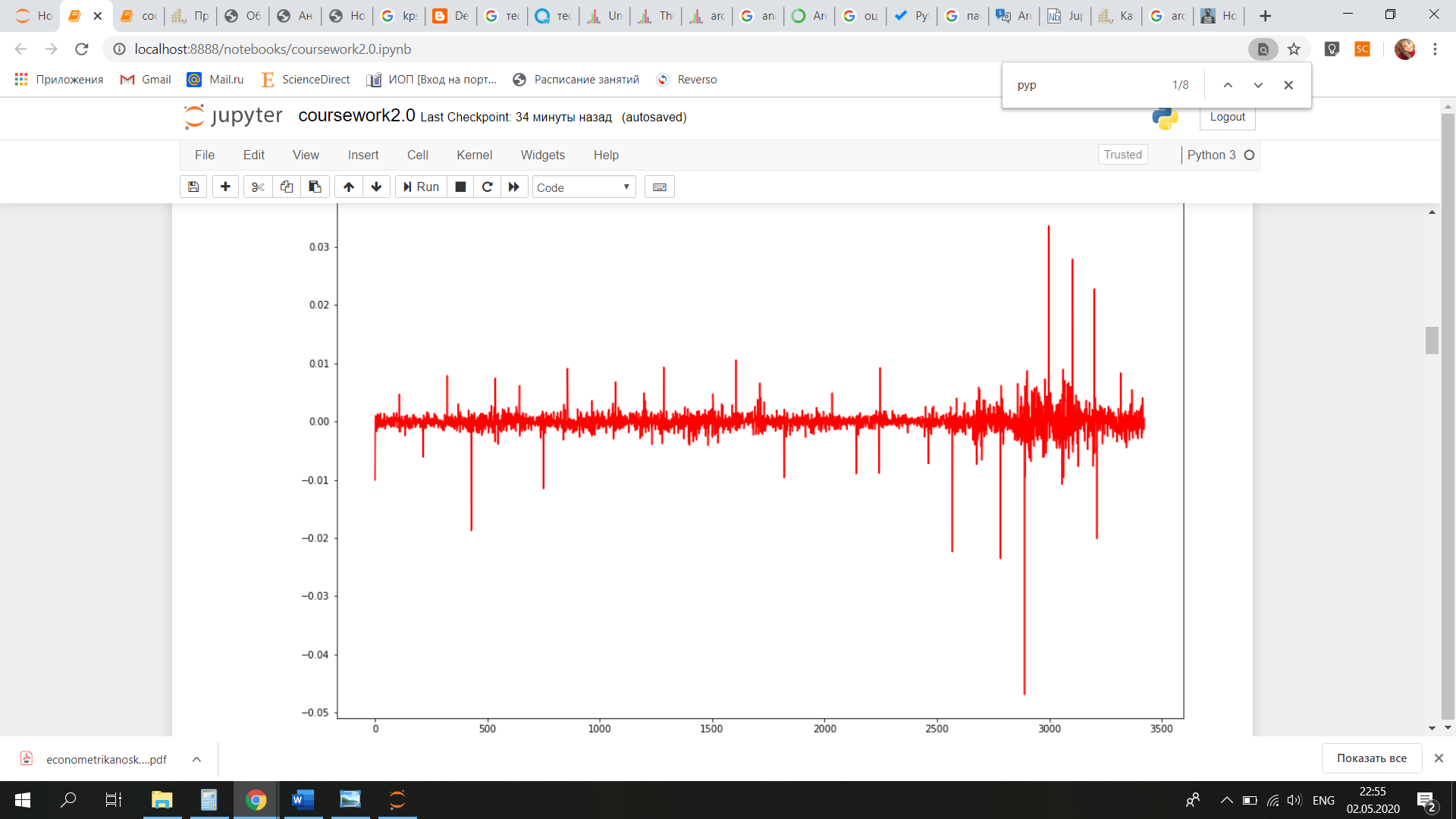


График 2. Доходности индекса РТС.

По данным графика доходностей можем заметить отсутствие тренда и стационарность. Эти наблюдения подтверждают проведенные тесты.

Результаты тестов дают нам основания для того, чтобы использовать в качестве модели для прогнозирования модель AR(p). Целесообразность использования модели ARMA(p,q) проверим, оценив ее и проведя тест Бокса-Льюнга для остатков модели.

Чтобы оценить модель ARMA(p,q) необходимо выяснить ее порядки p,q. Для этого воспользуемся ACF и PACF – автокорреляционная функция и частичная автокорреляционная функция.

Для lag=1:

|  |  |
| --- | --- |
|  |  |

Для lag=10:

|  |  |
| --- | --- |
|  |  |

Получаем p = 1 , q = 1

Оценим модель ARMA(1,1):



Проверим правильность выбранных параметров p,q с помощью теста Филлипса-Перрона.

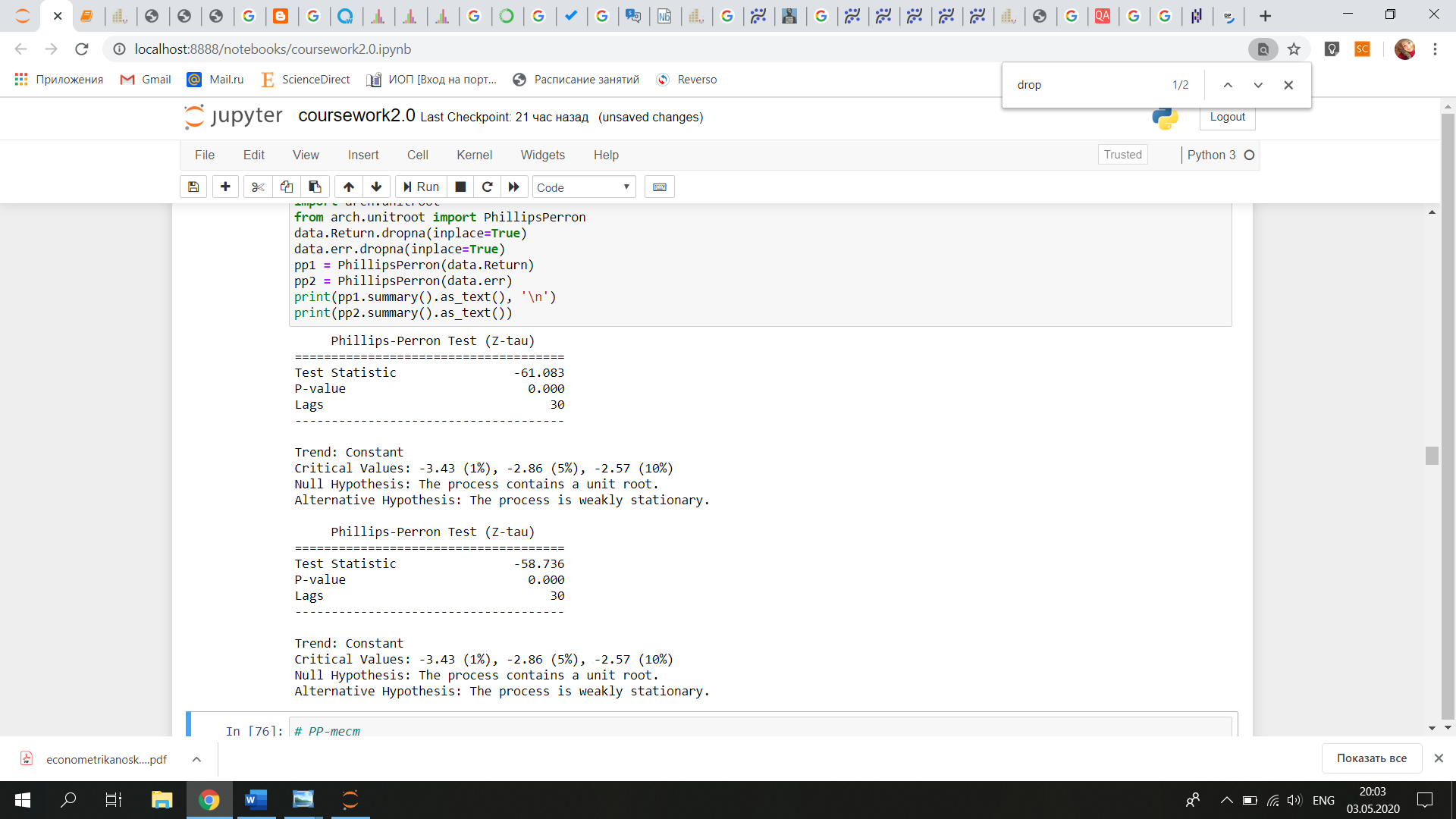
Тест Филлипса-Перрона (PP-тест):

Данный тест проверяет гипотезу о том, что автокорреляция в данных может иметь более высокий порядок.

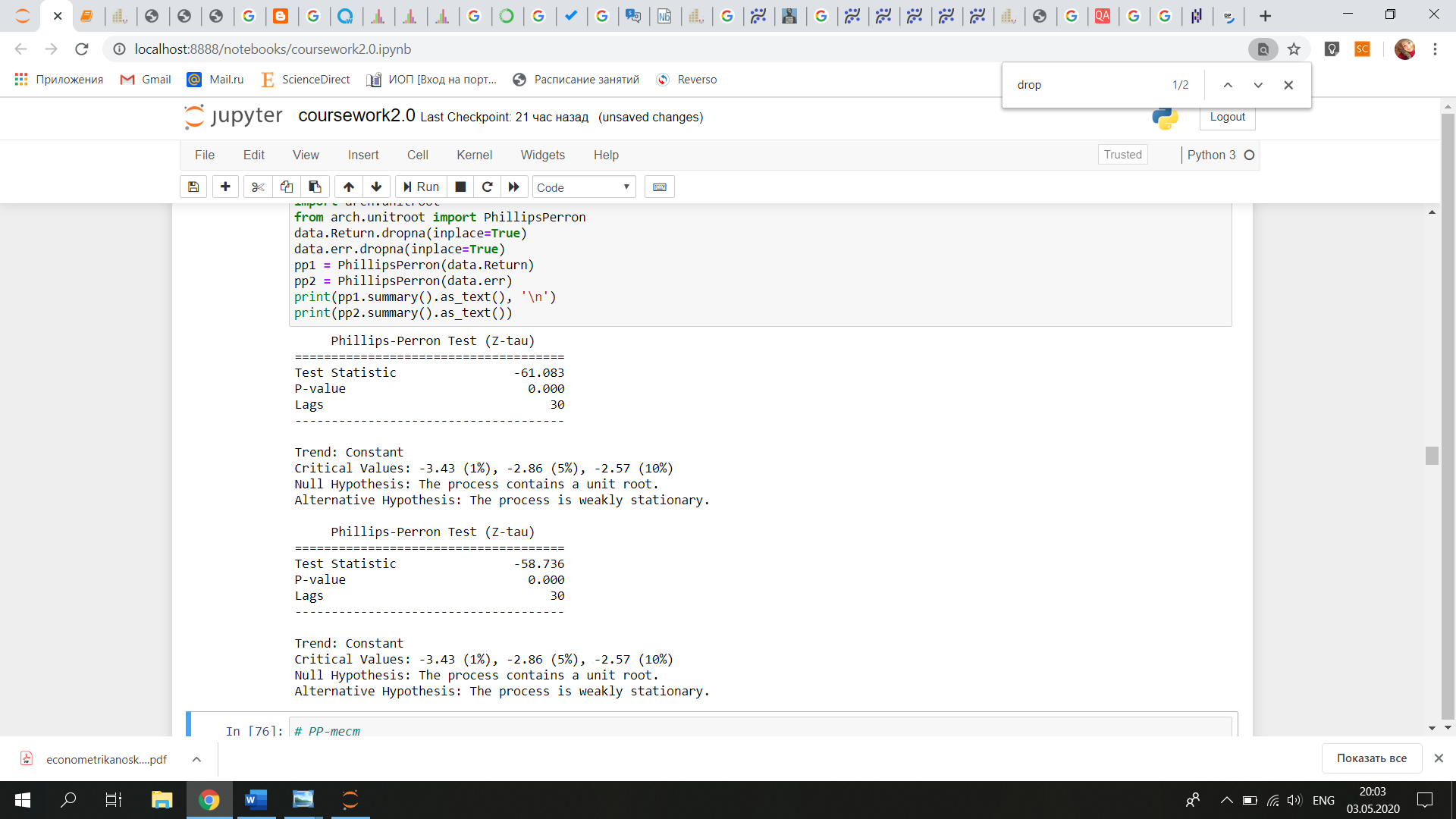
Выдвигаются гипотезы:  
 – существует единичный корень

– не существует единичного корня, автокорреляция первого порядка

Результат теста для p (применяется к ряду доходностей):



Результат теста для q (применяется к ошибкам модели):



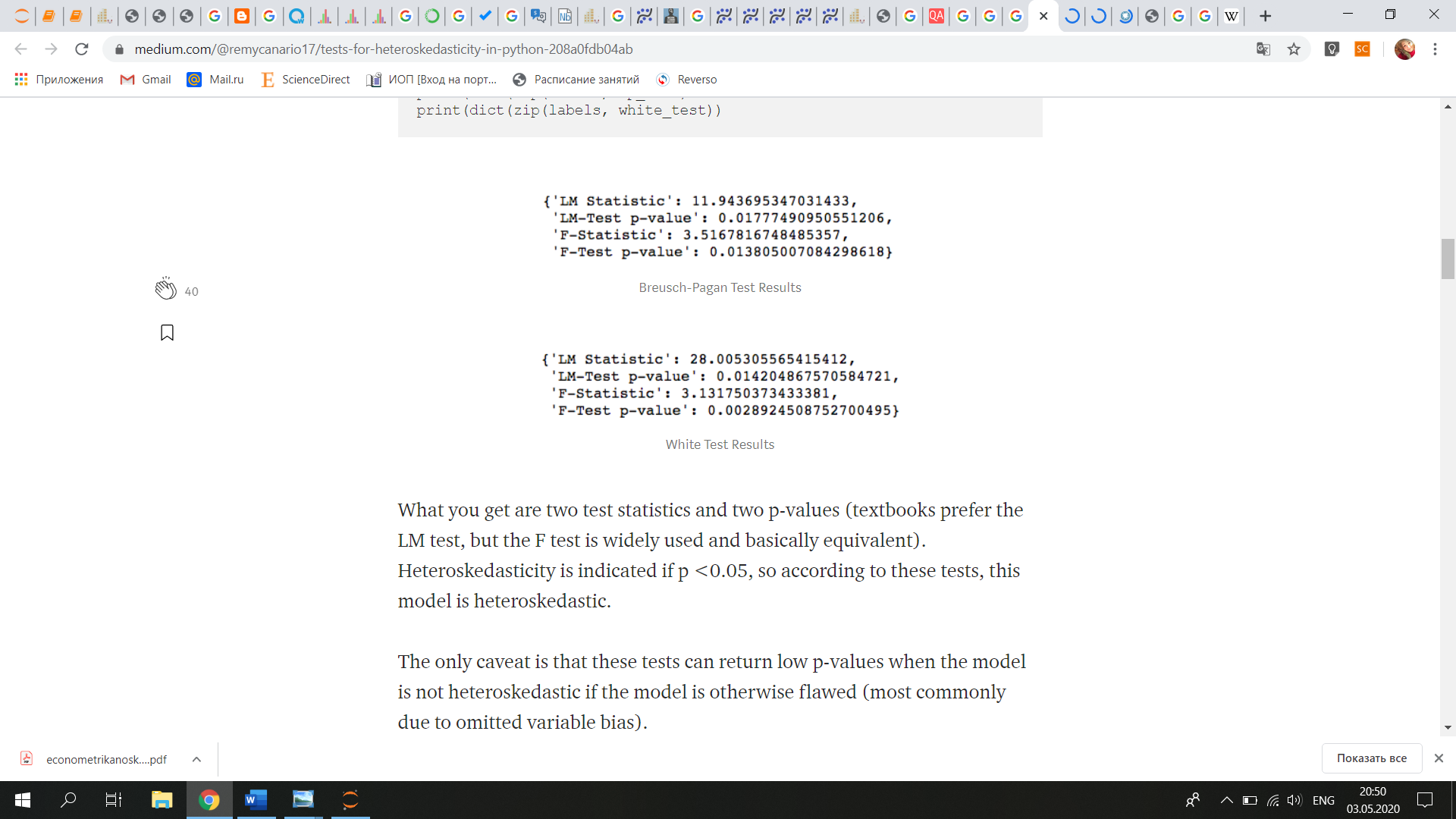
Для порядка (1,1), получаем результат: нулевая гипотеза отвергается: порядок модели, равный p=1, q=1, подобран верно.

Проверим необходимость включать в нашу модель компонент GARCH.  
Для этого проведем тест на гетероскедастичность остатков модели, тест Уайта.

Тест Уайта:

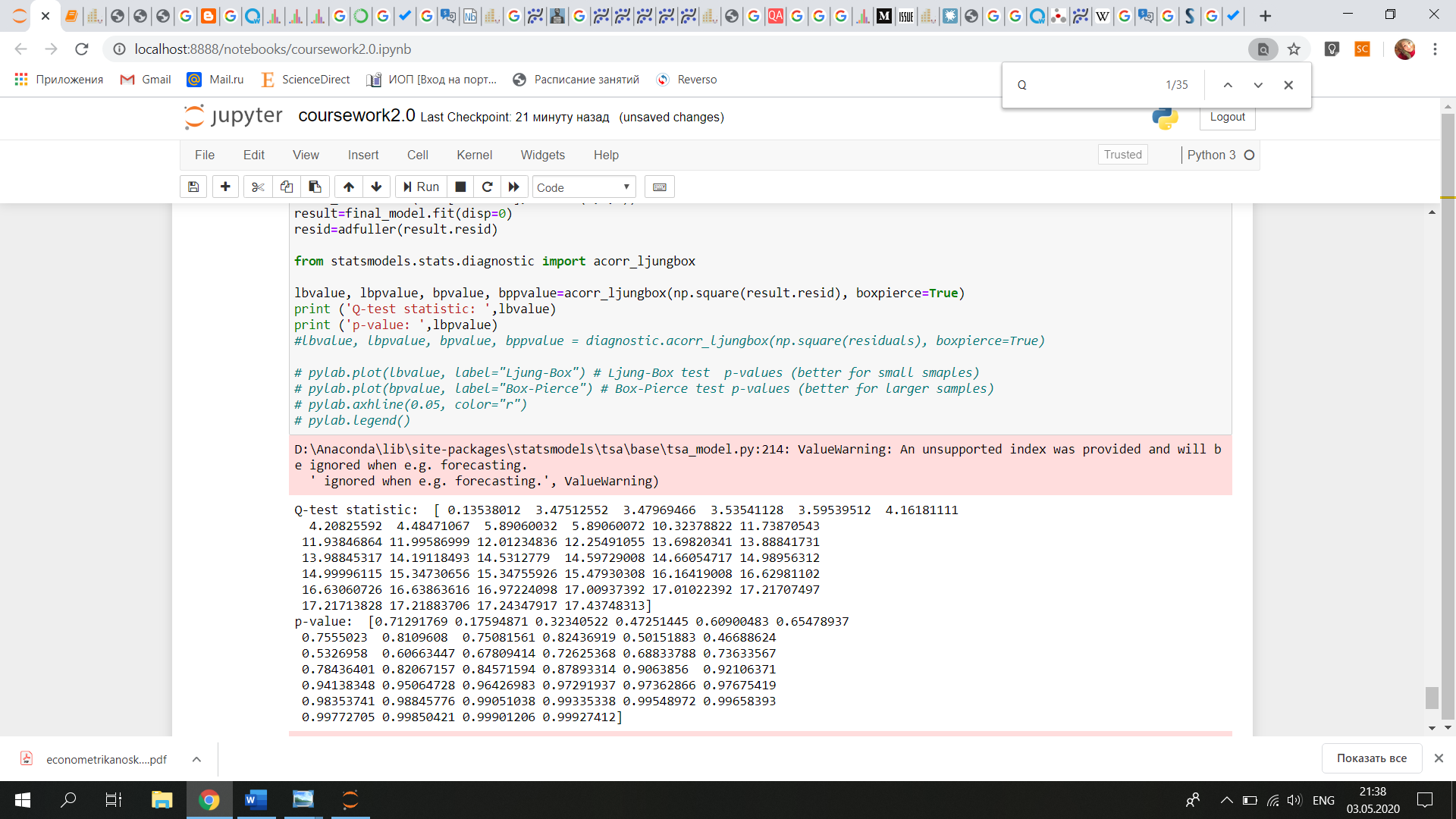
Данный тест не налагает ограничений на структуру гетероскедастичнгости. Он использует остатки регрессии и строит вспомогательную регрессию квадратов остатков (+ константа, регрессоры исходной модели и их произведения попарно). Используется LM-статистика.

Статистика теста:



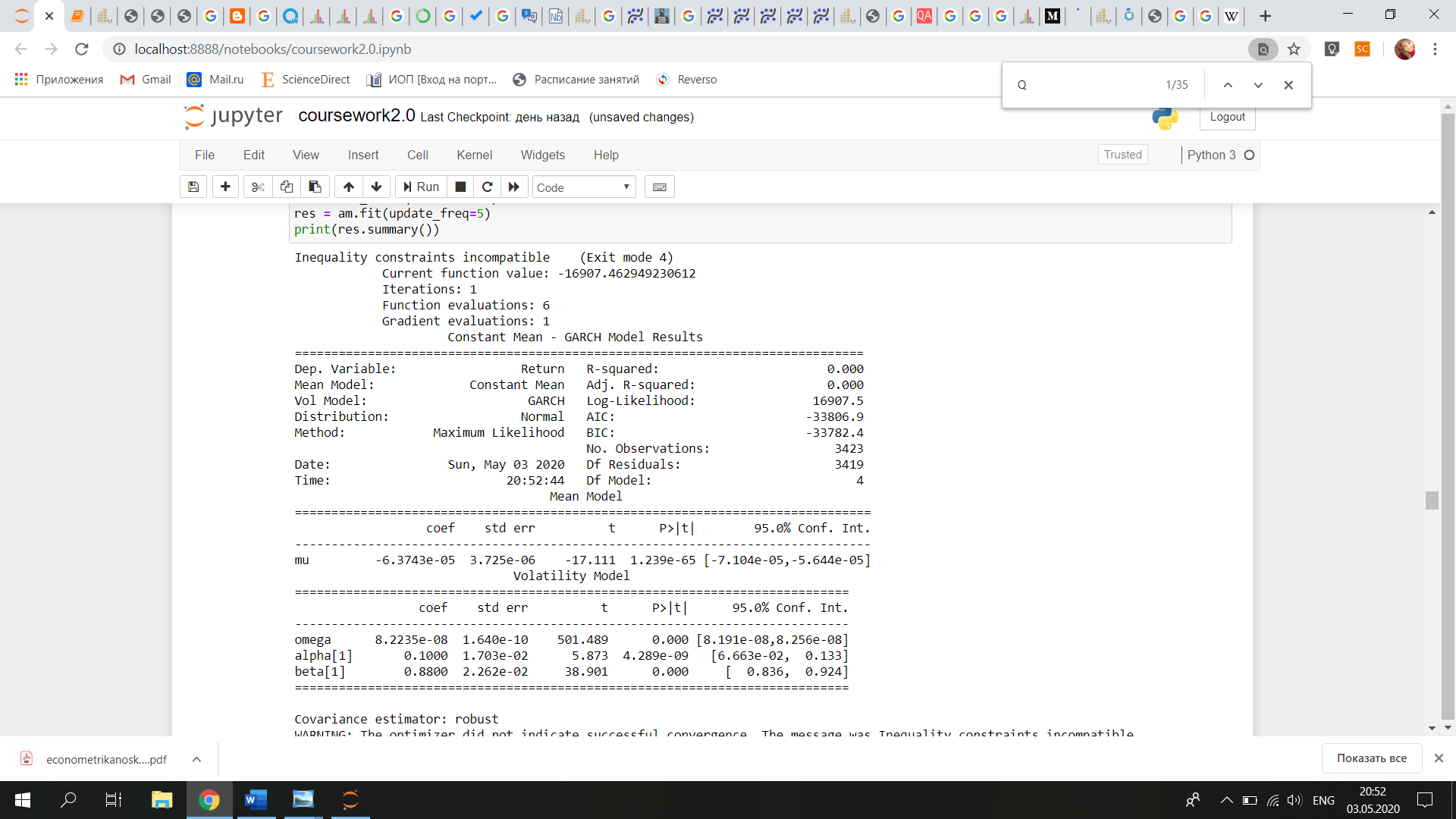
По результатам теста Уайта заметим, что – остатки действительно показывают гетероскедастичность.

Тест Бокса-Льюнга для остатков (на белый шум):



Нулевая гипотеза отвергается, следовательно ошибки модели не являются белым шумом, и имеет место автокорреляция.

Оценим модель GARCH()

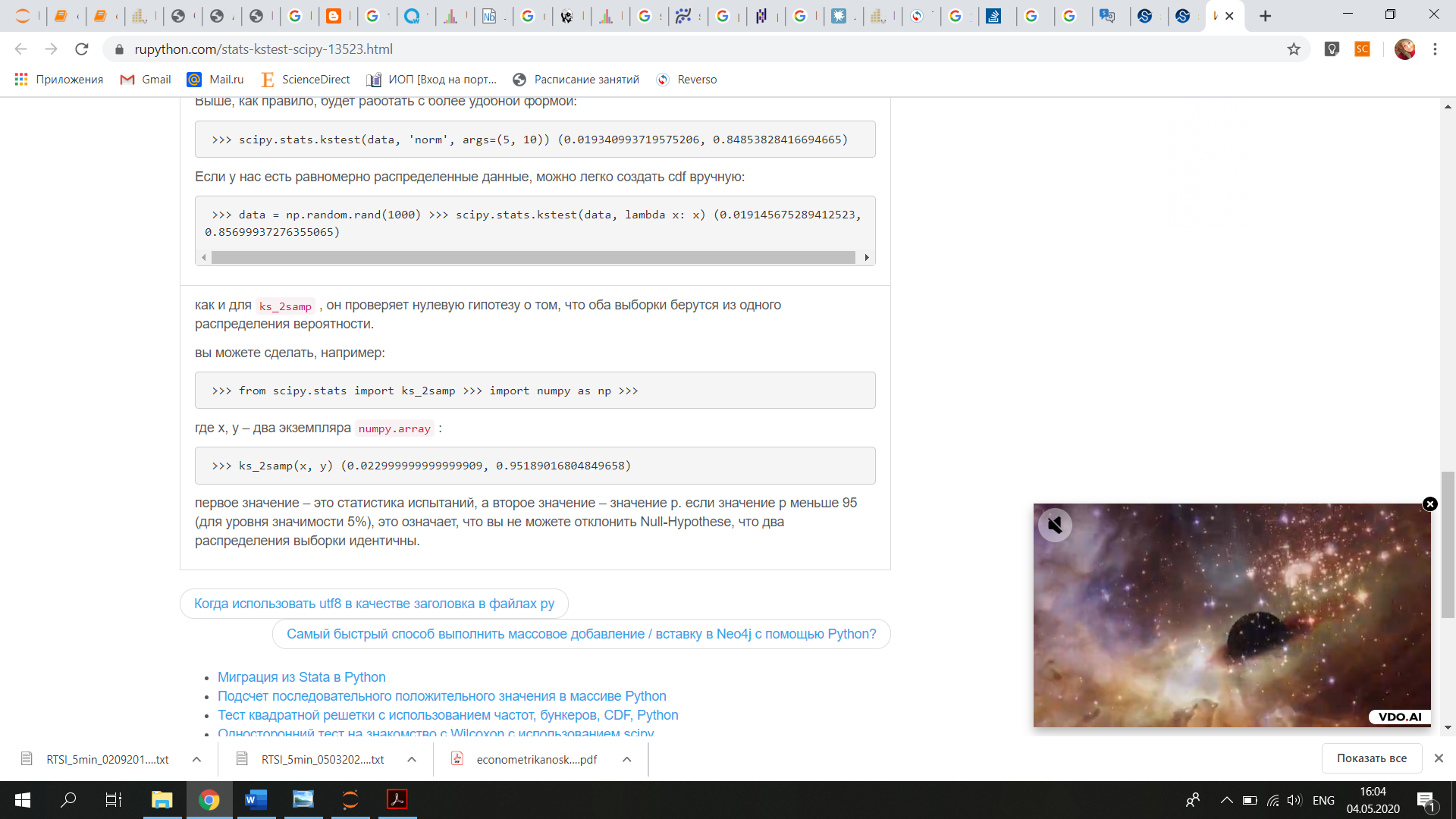


2.2.3. Проверка состоятельности.

Проведем тест на проверку того, что выбранные данные соответствует [предполагаемой модели](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9) – критерий согласия Колмогорова.

Данный тест вычисляет меру отклонения данного распределения от некоторого полностью известного распределения.

Нулевая гипотеза: тестируемое распределение принадлежит известному распределению. Она отвергается, если статистика теста превышает квантиль распределения заданного уровня значимости.  
Статистика теста:



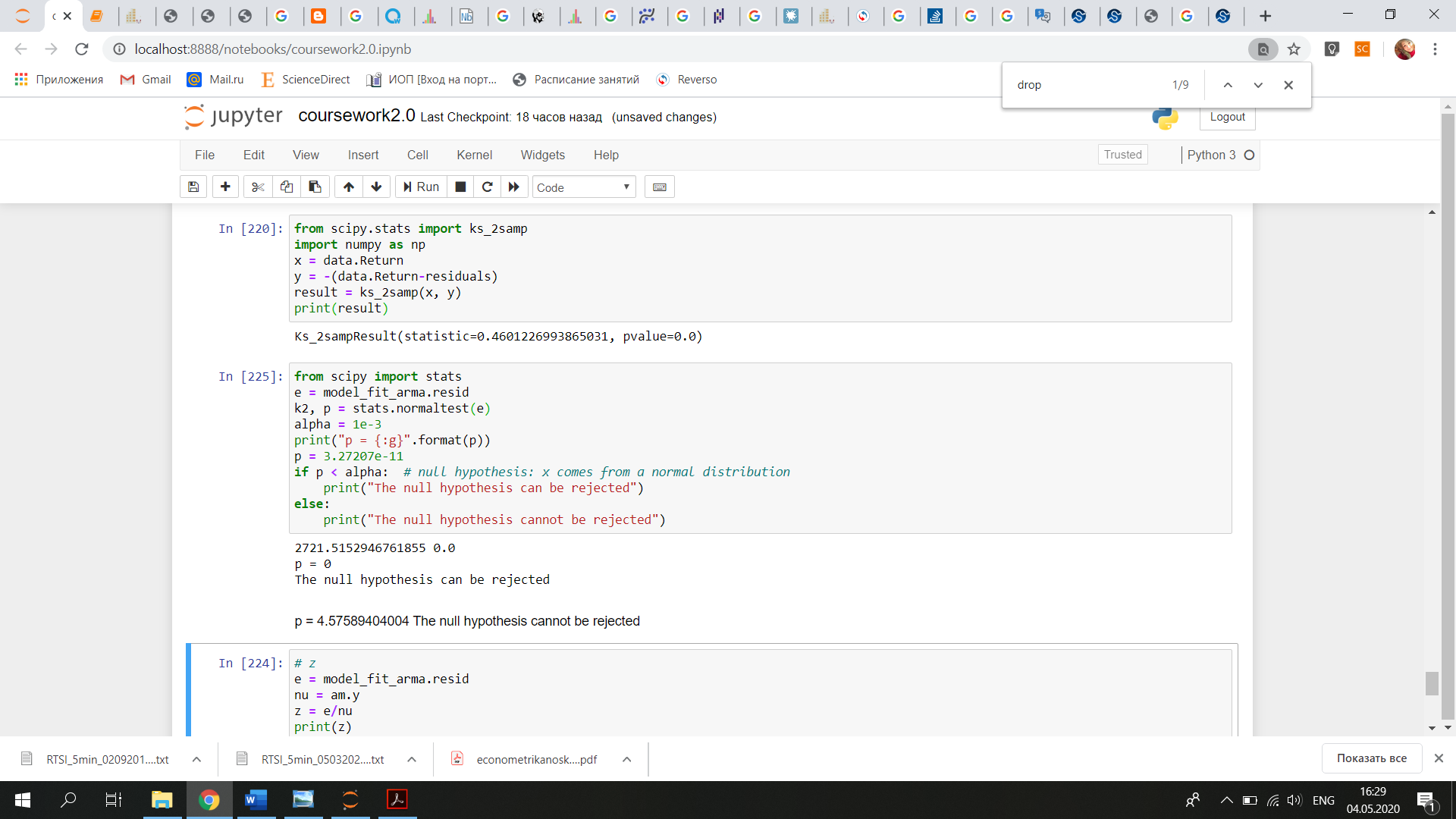
Нулевая гипотеза не отвергается.

2.2.4. Оценка шоков.

В пункте 2.1.1. План работы мы определили значение шока на финансовом рынке в момент t следующим образом:

При этом – ошибка модели, нормально распределена.  
Проверим ошибки модели на нормальность:

Оценим с помощью stats.normaltest(x). Нулевая гипотеза: x – принадлежит нормальному распределению.



Оценим шоки:

|  |  |
| --- | --- |
| Оцененное значение : | Знаки ошибок\* sign(: |
|  |  |

\* - Если , sign(  
Если , sign(

Если , sign(

2.3. Прогнозирование.

2.3.1. Описание процесса

Осуществим прогноз с помощью рекуррентных нейронных сетей типа LSTM (Long Short-Term Memory). Данный тип сетей отличается тем, что учитывает долгосрочные последовательности для короткосрочных лаговых таймфреймов.

Используется библиотека Keras. Серия (batch) в Keras – фиксированное количество строк из тренировочной выборки, определяющее сколько зависимостей будет «проигрываться» до переопределения весов сети.

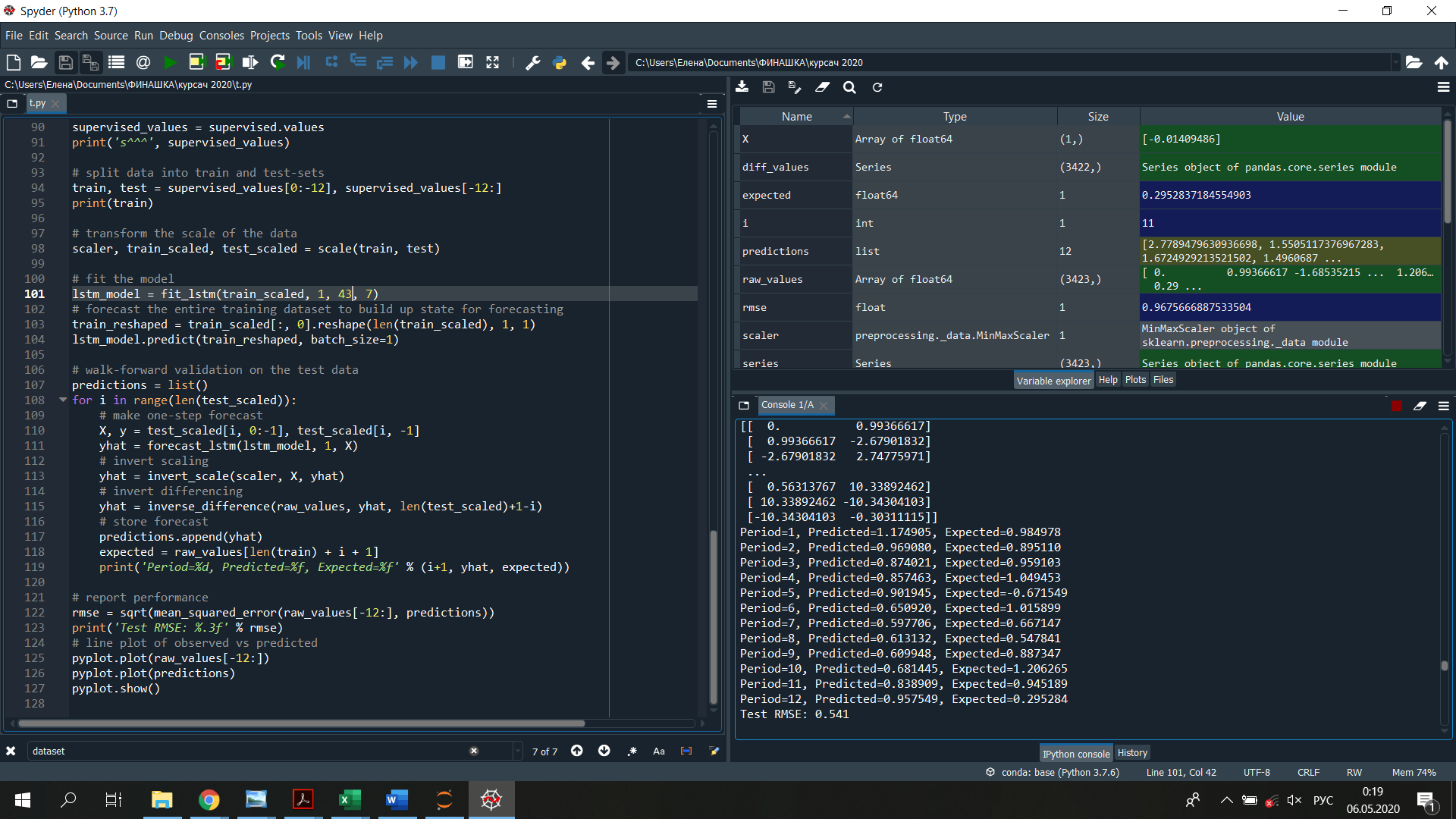
Алгоритм таков:

1. Загрузка данных
2. Преобразование данных, чтобы их можно было применить к LSTM моделям. Шкалирование от -1 до 1
3. Моделирование LSTM сети по тренировочной выборке.
4. Рассчитать статистику для тестирующей выборки
5. Просчитать прогноз

Тестирующая выборка шкалируется по смоделированной тренировочной выборке, показывает минимальны и максимальные значения тестирующей выборки и в то же время не влияет на модель.

2.3.2. Прогнозирование

Посмотрим, как прогнозирование работает на 12 пятисекундных таймфреймов вперед. Сравним спрогнозированные значения и наблюдаемые. Важно обратить внимание на показатель RMSE, который показывает среднее расстроение между прогнозируемым и наблюдаемым. Чем он меньше, тем лучше сработал прогноз.



4. Заключение

В данной работе я собиралась составить модель доходности индекса РТС. И далее оценить шоки и спрогнозировать их с помощью методов машинного обучения, нейронных сетей. После проведенной работы, можно утверждать, что шоки на финансовом рынке – процесс, который можно прогнозировать. Модель ARMA-GARCH может быть успешно использована для моделирования вышеуказанных процессов. В дальнейшем прогнозируемые значения, полученный алгоритм можно применить в создании роботов для трейдинга и для создания сигналов об изменении направления рынка в целом.

5. Источники

1. Fuller, R.J. (1998). Behavioral finance and sources of alpha. Journal of Pension Pan Investing. 2(3), 291-293

2. Jacob F. (2015). Risk estimation on high frequency financial data: Empirical analysis of the DAX 30. Springer.

3. В.А. Цветков. (2012). Циклы и кризисы: теоретико-методологический аспект.

4. В.А. Бывшев. Эконометрика: учебное пособие/ В.А. Бывшев. – М.: Финансы и статистика, 2008. – 480 с.: ил.

5. Kyungwon K. (2013). Modeling financial crisis period: A volatility perspective of Credit Default Swap market. Physica A 392. 4977-4988

6. Приложения

Приложение 1. - Файл «RTSI\_5min\_21012020\_05032020» со значениями обучающей выборки для нисходящего тренда (21.01.2020 – 05.03.2020)

Приложение 2. - Файл «RTSI\_5min\_05032020\_01042020» со значениями тестирующей выборки для нисходящего тренда (05.03.2020 – 01.04.2020)

Приложение 3. - Файл «coursework2.ipynb» с кодом программы на языке Python

Приложение 4.

