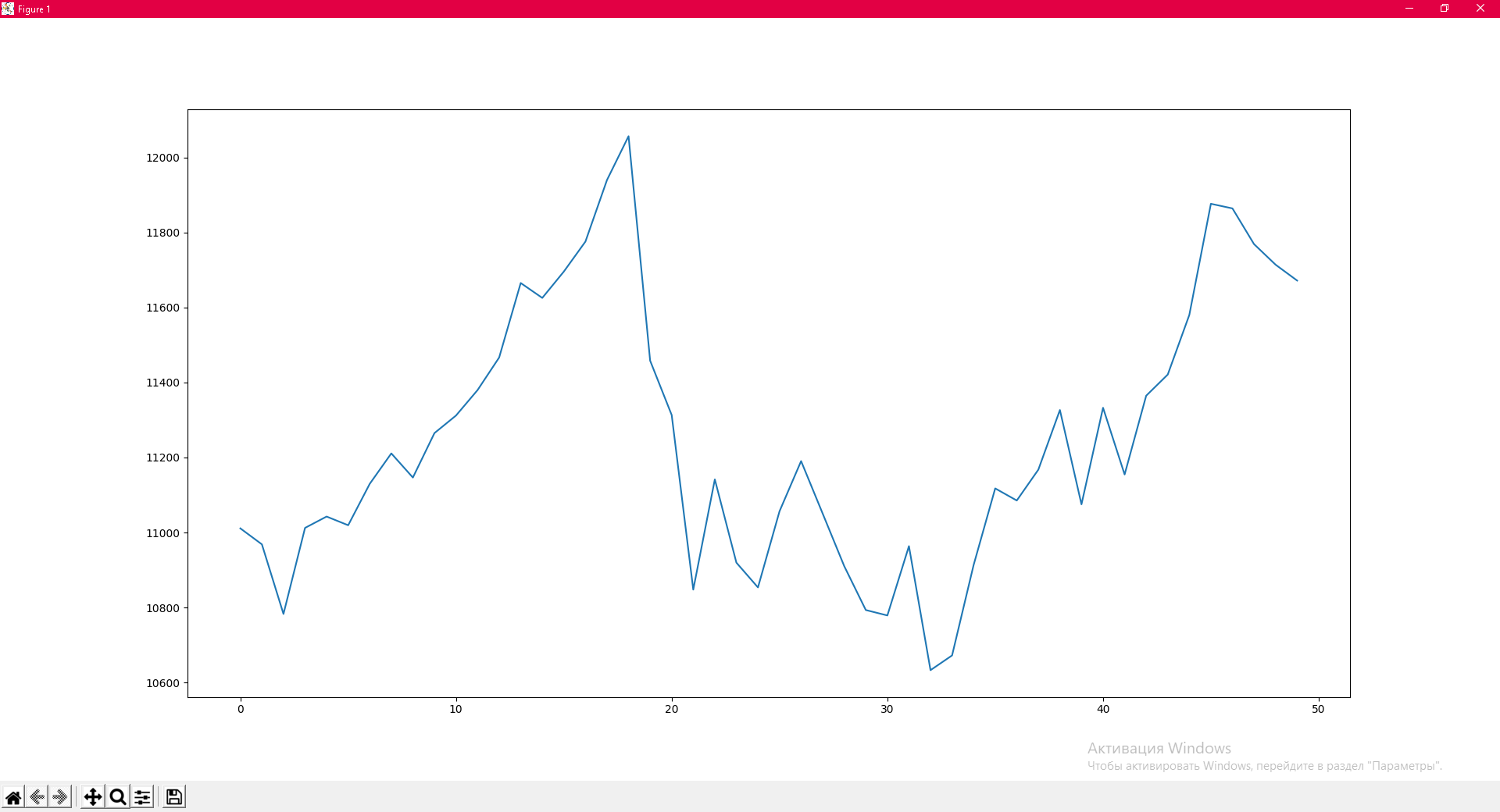
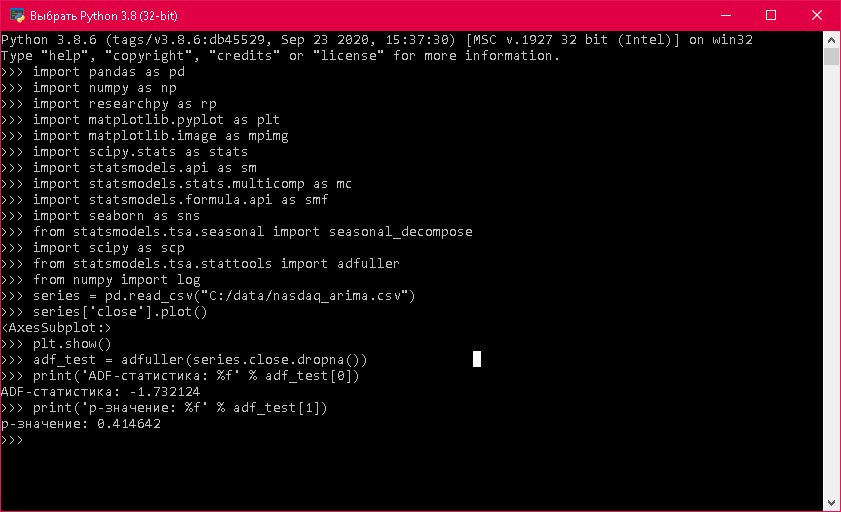
**Анализ временных рядов и прогнозирование**

1. Установим стационарность ряда.



Фондовый индекс сам по себе представляет нестационарный показатель, для которых применяется модель для разностей соседних уровней ряда: разности уже представляют собой стационарный ряд.

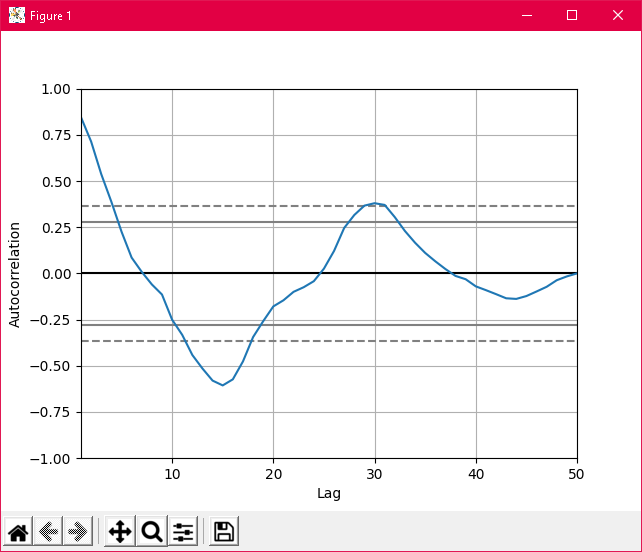
В статистике для проверки ряда динамики на стационарность применяется т.н. Дополненный тест Дики-Фуллера (ADF-тест). #Нулевая гипотеза в этом тесте "H0: ряд динамики нестационарный".



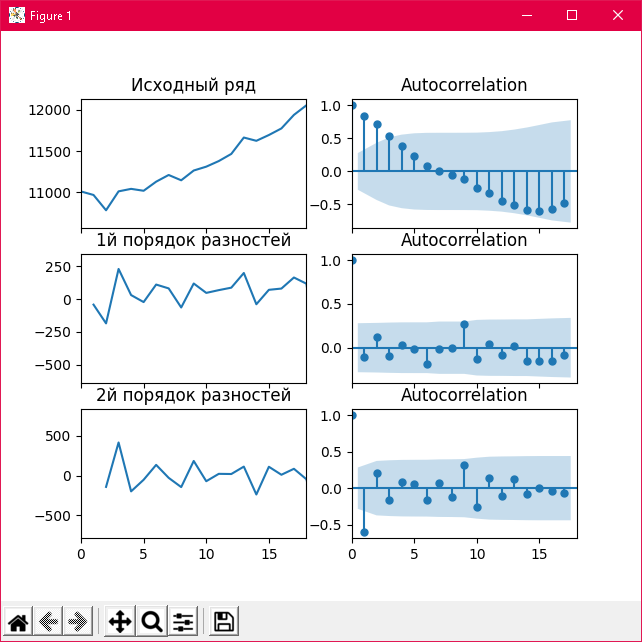
P-значение >0,05 соответственно H0 принимается, т.е. ряд динамики нестационарный.

Теперь мы хотим выяснить, нужны ли нам разности и какого порядка.

# строим график автокорреляции:

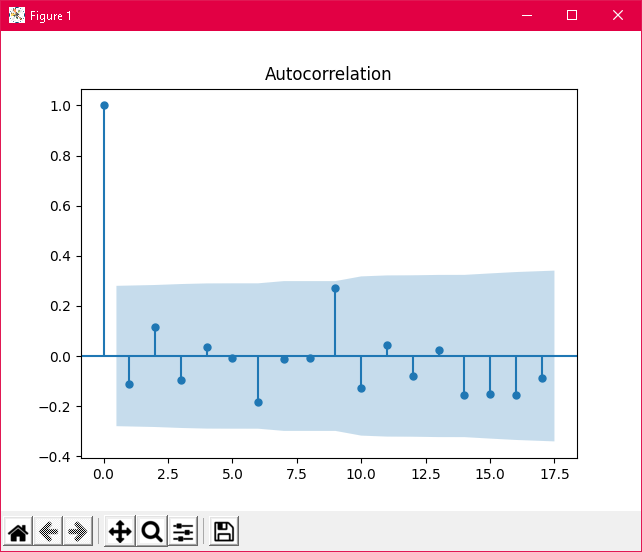


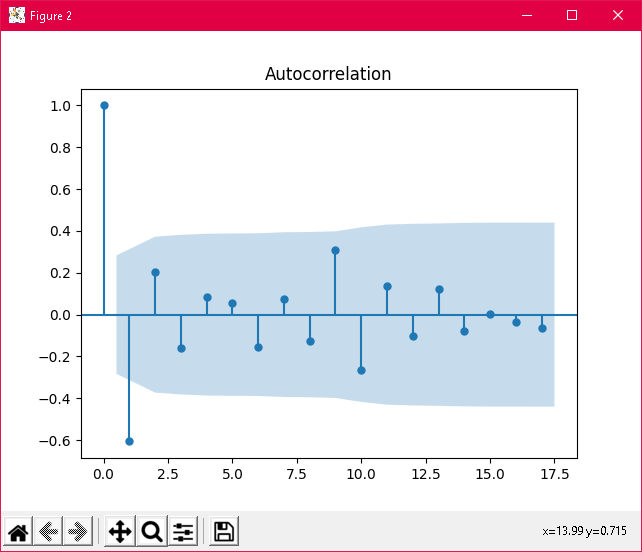
Автокорреляция ряда положительная для примерно 7-8 предыдущих наблюдений (т.е. с лагами от 1 до 7-8), значима для лага 3-4, и выше 0.5 для лага 3 и менее.



Для вышеуказанного ряда временной ряд достигает стационарности с двумя порядками разностей.

Построим разности на одном графическом поле:





Порядок дифференцирования определяется следующим образом:

0 - модель стационарная, включает долгосрочную компоненту (ненулевое долгосрочное среднее).

Модель с 1-м порядком дифференцирования должна иметь константу и имеет постоянный средний тренд (случайное блуждание в одном направлении).

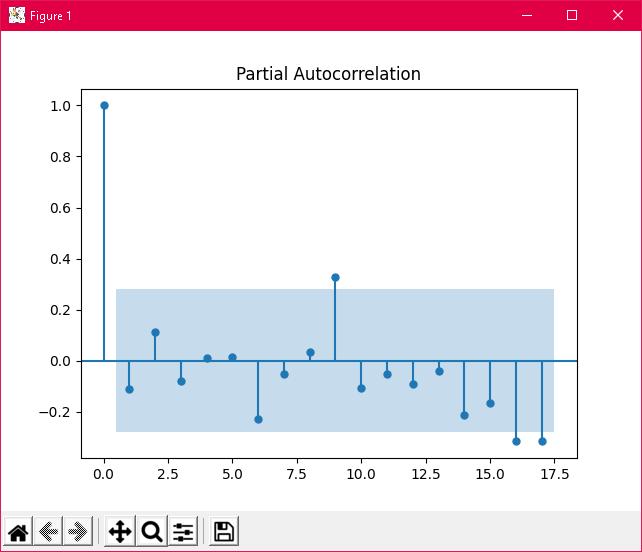
Модель со 2-м порядком дифференцирования предполагает, что тренд меняется со временем (т.е., например, случайный тренд)

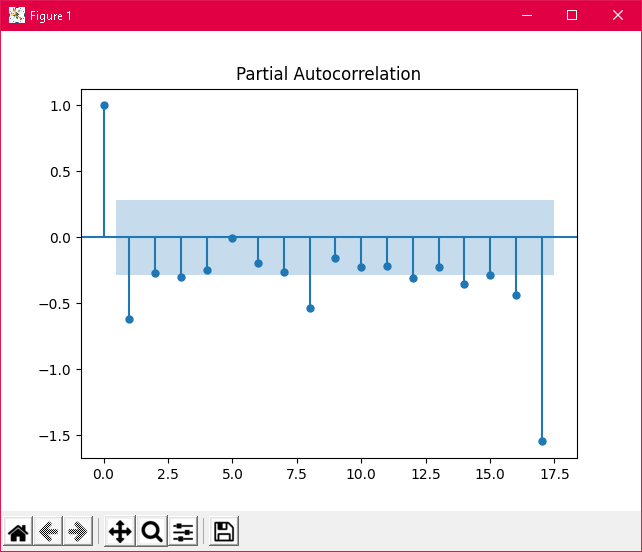
Для слишком высокого порядка разности модели (2) возникают признаки чрезмерного порядка, например, смена знака остатков от одного наблюдения к следующему.

На графике автокорреляции регрессионных остатков при этом виден отрицательный пик примерно в 0,5 на лаге 1; также повышается стандартное отклонение модели. Возьмем одну несезонную разность. Потом при подгонке модели может обнаружиться, что добавление еще одного порядка разности снимает некоторые проблемы.

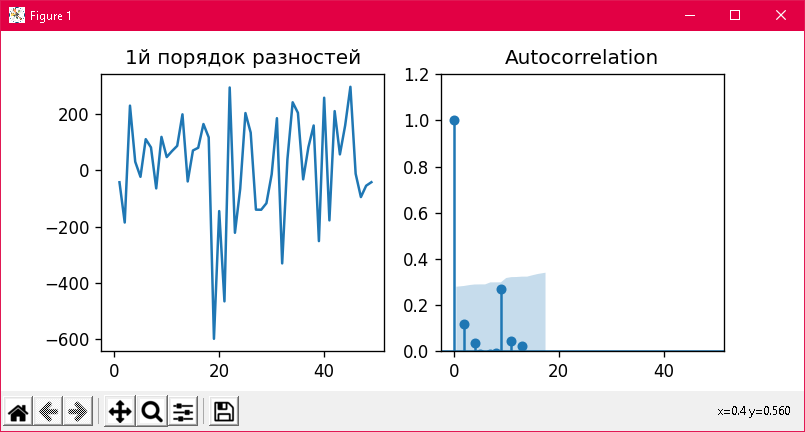
График автокорреляции порядка дифференцирования 1. Этот график также проясняет, сколько параметров авторегрессии взять в модель.

#Консервативно - берем 5. Менее консервативно - 9-11. Очень консервативно - начинаем с 1.





ACF - автокорреляция самого ряда. Этот график проясняет, сколько параметров скользящей средней целесообразно включить в модель.Строим график



По-видимому, также 5-7. Что, если в модели чрезмерный (или недостаточный) порядок разности?

#- если текущий порядок немного слишком большой, добавить один параметр МА.

#- если текущий порядок немного недостаточен, добавить один параметр AR.

В моделях ARMA/ARIMA порядок модели указывается в терминах параметров p,d,q, где p - количество параметров авторегрессии (определяемое частной автокорреляцией ряда с его регрессионными остатками), d - порядок разностей, q - количество параметров скользящей средней, определяемое автокорреляцией самого ряда).

Мы можем применить также следующую методику построения таких моделей: сначала строится модель с порядком 1,х,1 (вне зависимости от того, каков порядок разностей), затем число параметров авторегрессии повышается до наилучшей подгонки под "естественное" движение ряда, и затем \_строго по одному\_ повышается количество параметров скользящего среднего для "доводки" модели.

При этом в какой-то момент возникнет ситуация, когда добавление в модель еще одного параметра авторегрессии не дает видимого изменения качества модели на графике. Следует руководствоваться общей рекомендацией, что модель с меньшим количеством параметров лучше, при приблизительно одинаковом результате.

Для большинства процессов количество параметров авторегрессии и скользящего среднего приблизительно одинаковое. Следует также помнить, что чистые процессы авторегрессии (вида p,x,0) , как и чистые процессы скользящего среднего (вида 0,x,q) встречаются крайне редко.

Например, построим модель Арима с порядком 1,1,2

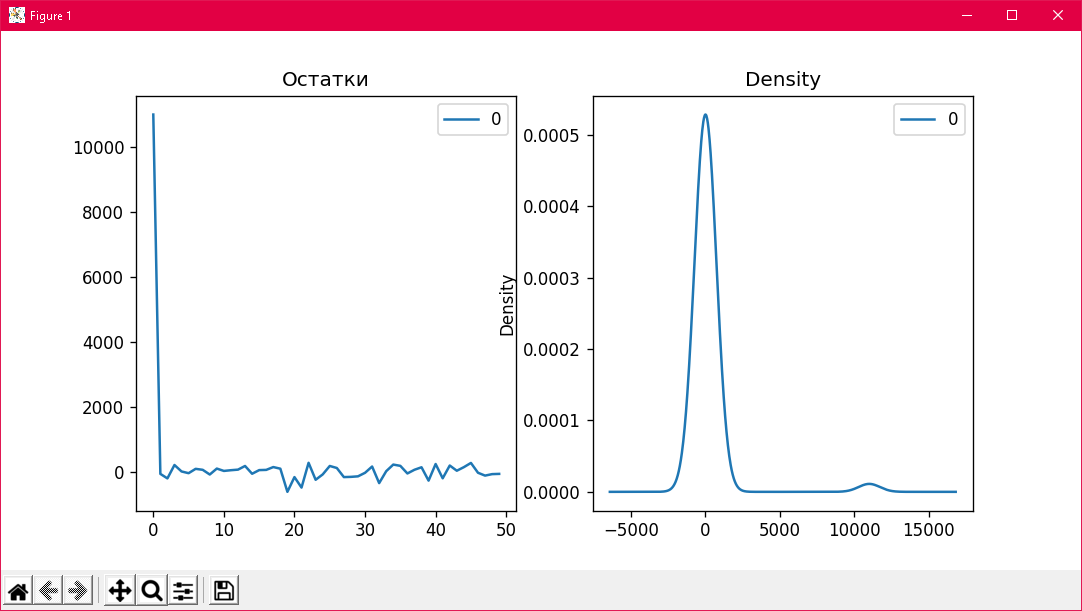
# порядок = число параметров AR, порядок разностей, число параметров MA.



Коэффициенты AR и оба МА незначимы. Попробуем перестроить модель без МА2 или взять больше коэффициентов.



Построим график регрессионных остатков, чтобы убедиться, что в них отсутствуют закономерности (т.е. смотрим, что средняя и дисперсия постоянна):



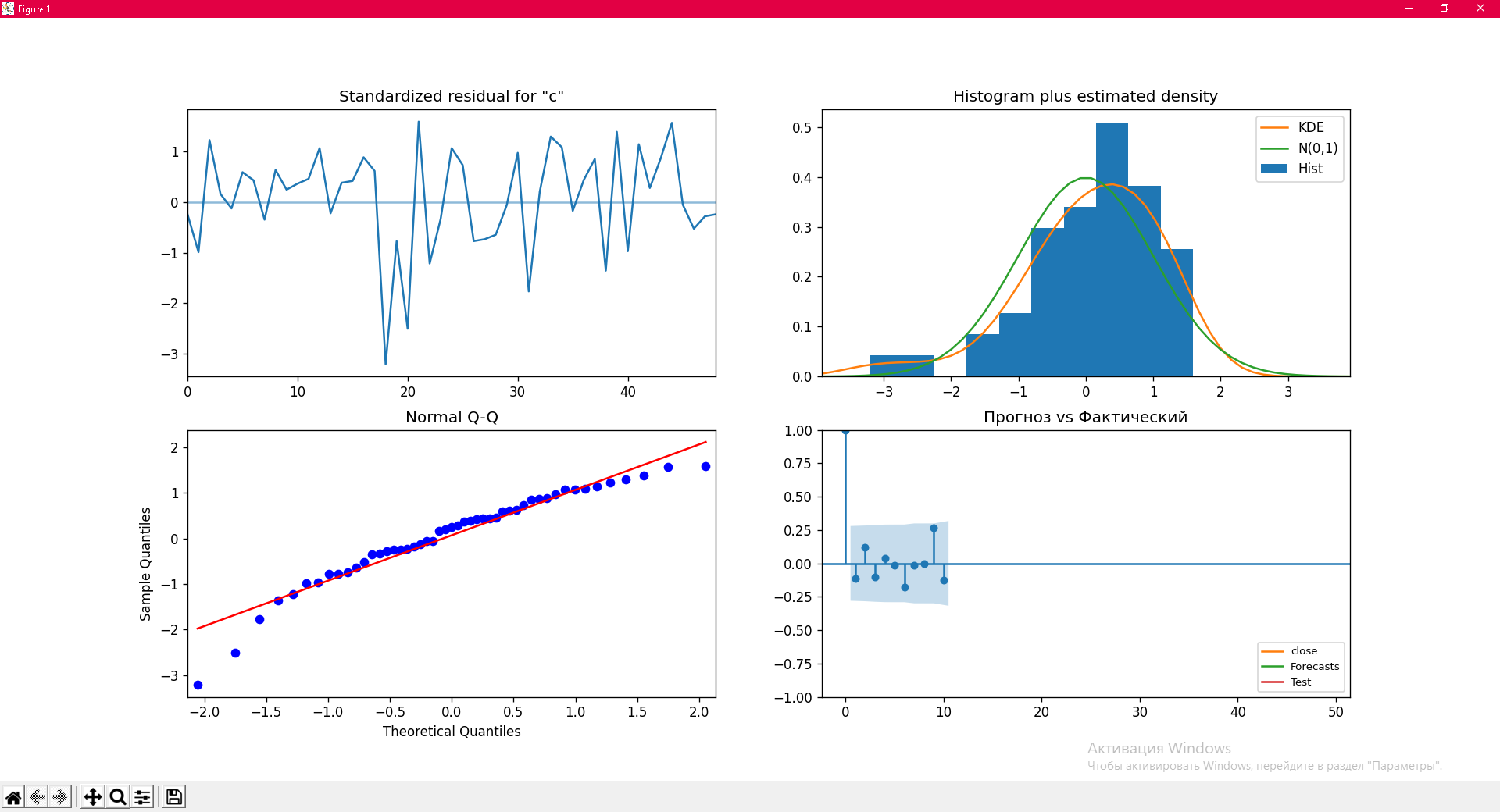
Также

#Левый верхний график - остаточные ошибки колеблются около нулевого среднего и имеют однородную дисперсию.

#Правый верхний график – нормальное распределение.

#Левый нижний график - на квантиль-квантильном графике есть незначительные отклонения. В целом картина хорошая.

#Правый нижний график - Любая автокорреляция будет означать, что существует некоторая закономерность в остаточных ошибках, которая не объясняется в модели.  В целом, похоже, он подходит.

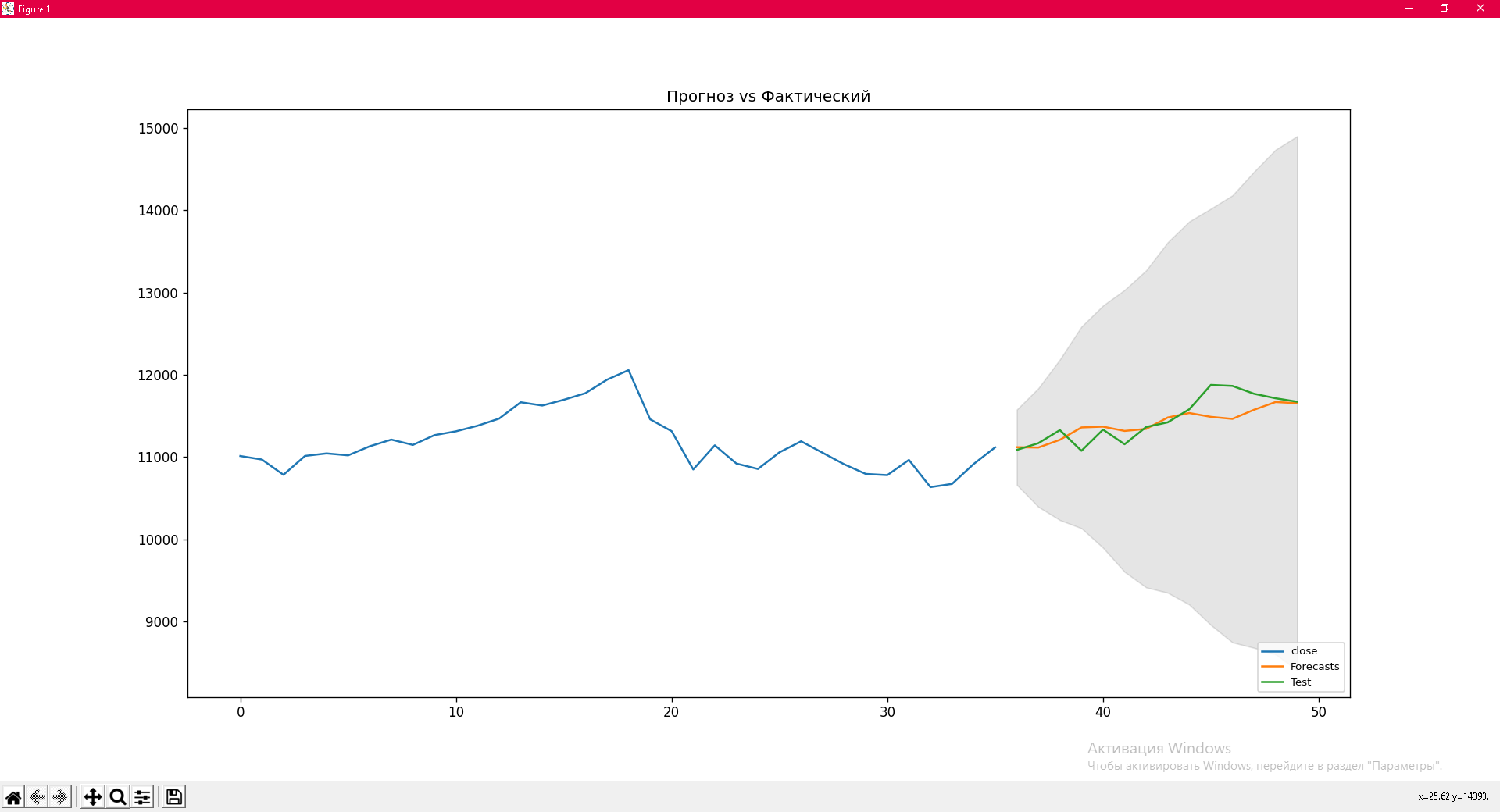


Для дальнейшей оценки модели проведем кросс-валидацию, которая предполагает сокращение ряда динамики на N шагов с последнего значения, а затем построение прогноза на N шагов вперед, с проверкой соответствия фактическим известным данным.

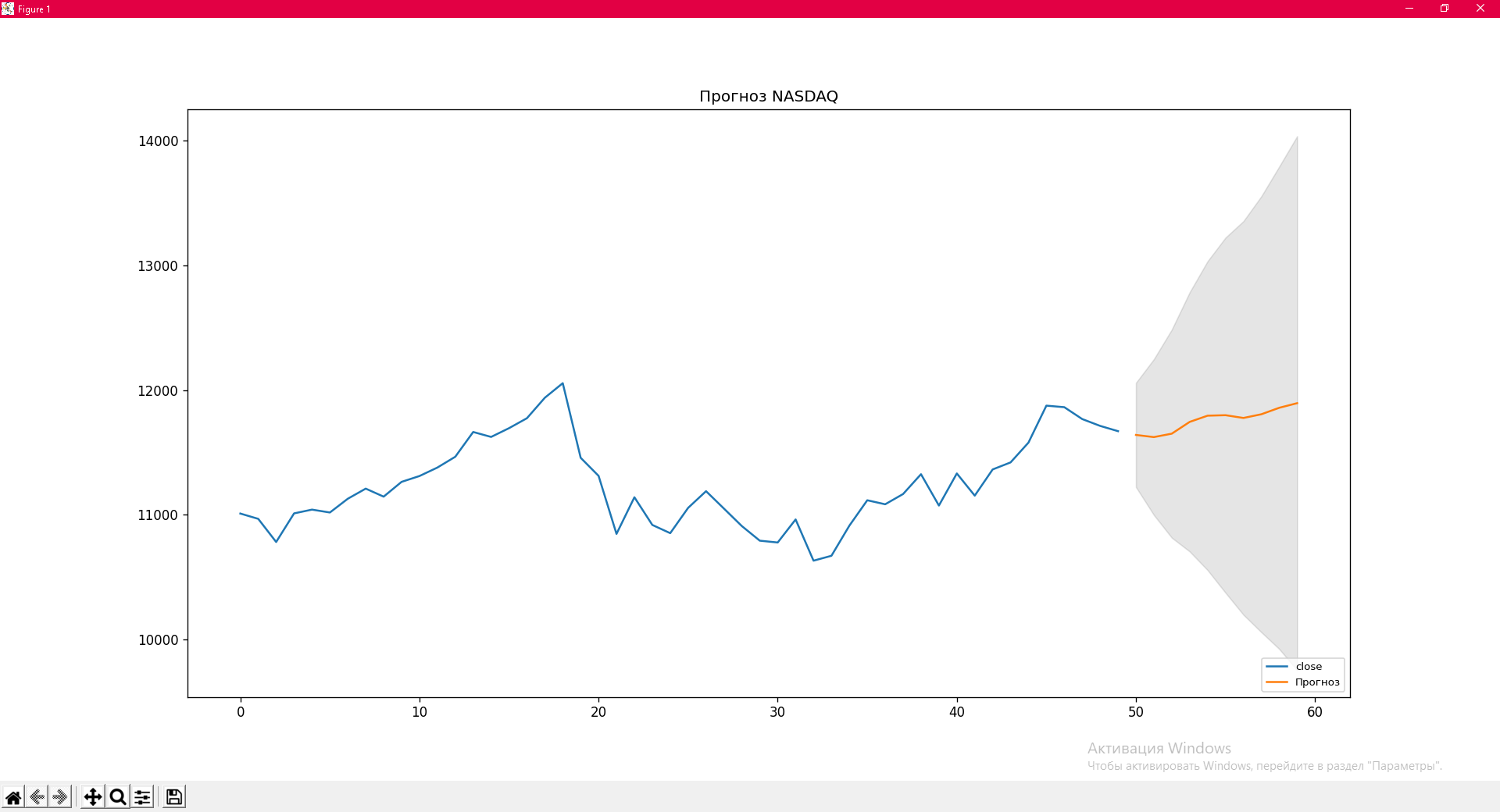
Прогноз находится ниже фактических значений, хотя модель верно "определяет" дальнейшее направление в целом. То есть добавление некоей константы, или дополнение тренда, к нашему прогнозу, должно привести к его улучшению.

Оставляем порядок разности 2, и последовательно поднимем p до 6 и затем q до 4, и пронаблюдаем, какая модель дает наиболее низкий критерий информативности Акаике (AIC) и дает наилучшее приближение прогнозных значений к фактическим.

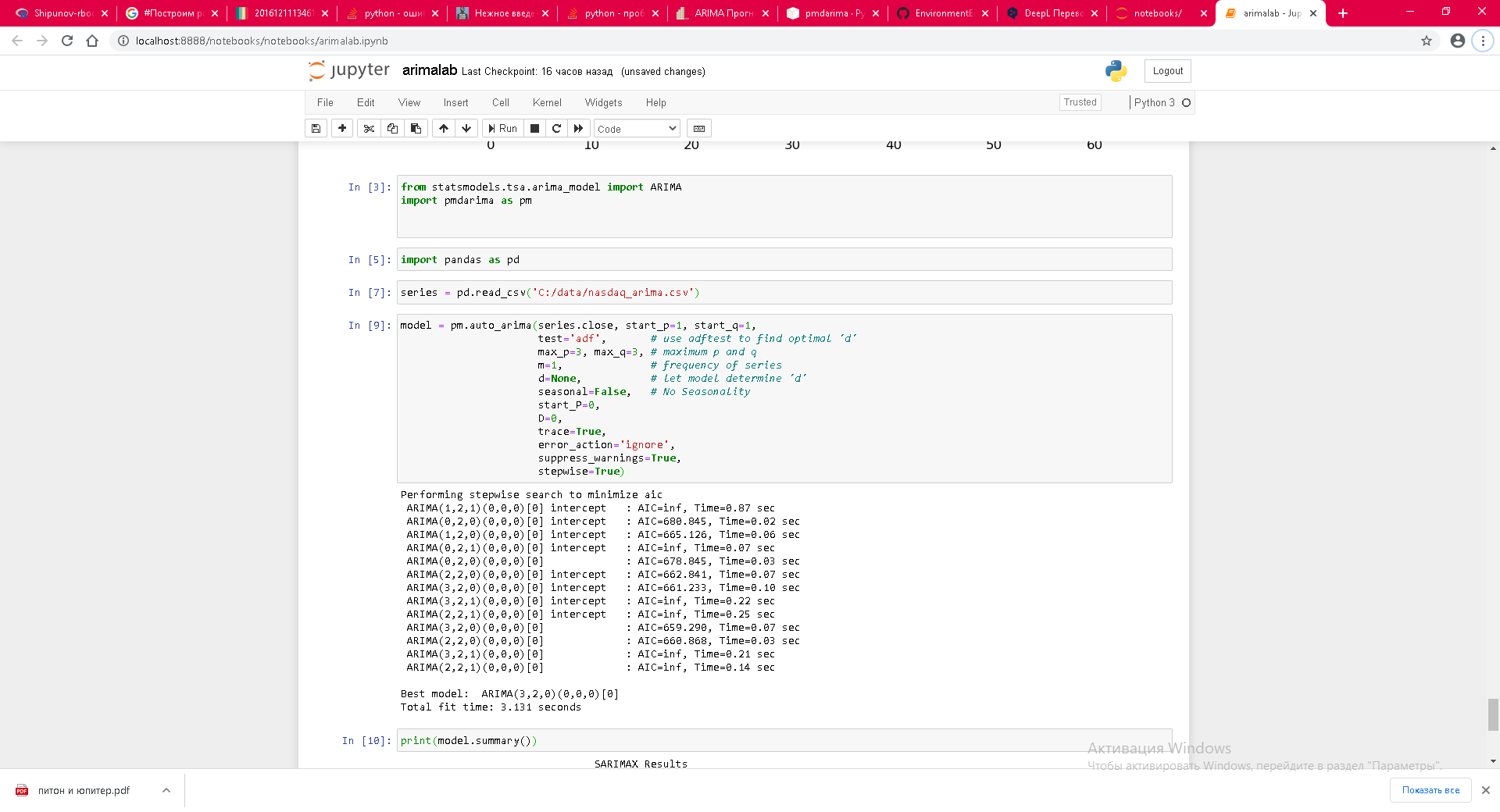
В то же время необходимо отслеживать p-значения для коэффициентов AR и MA в модели, которые должны быть как можно ближе к 0.

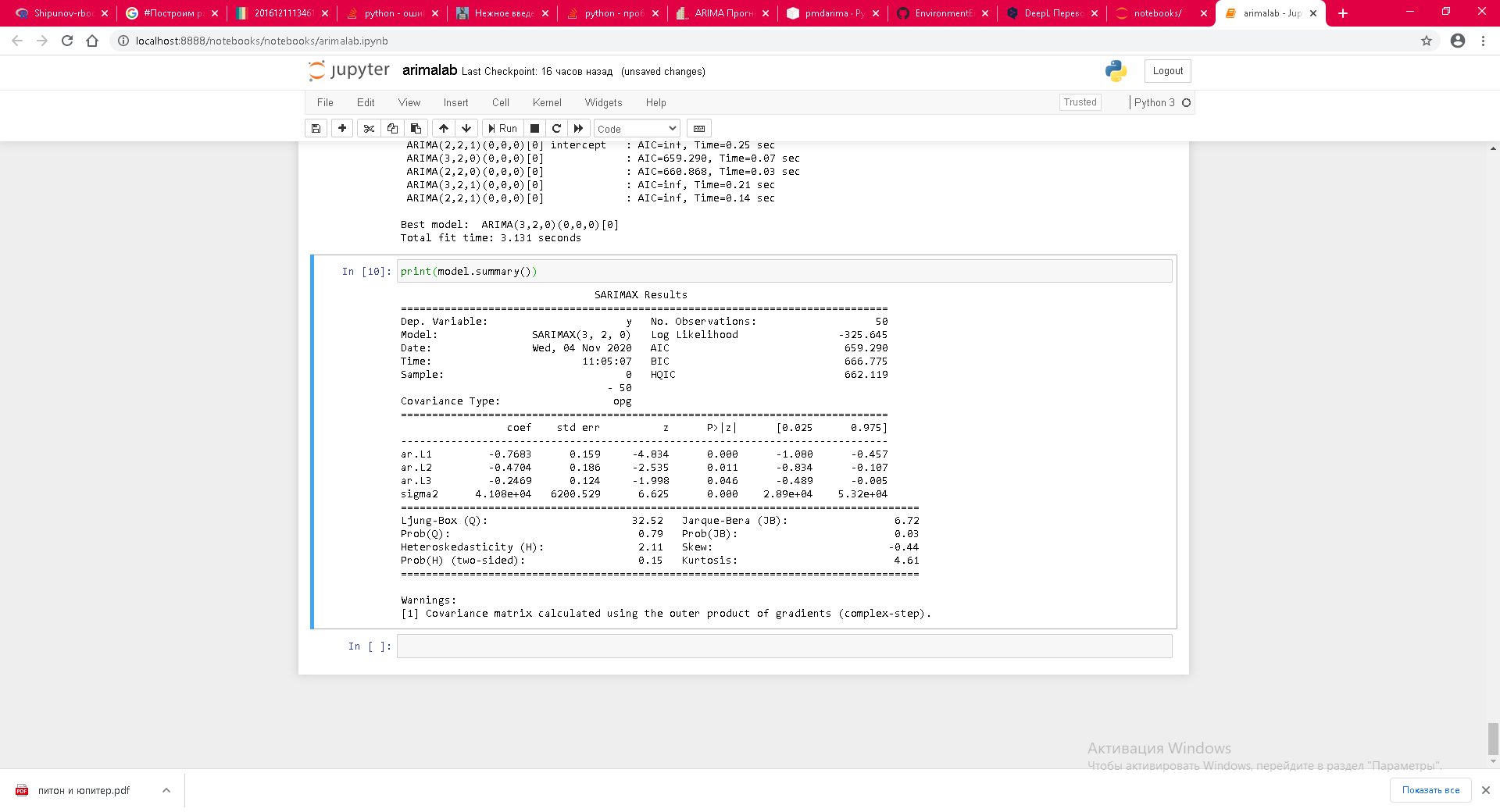


Для проверки модели целесообразно также менять количество дней, на которые строится кросс-валидационный прогноз. Теперь строим модель с наилучшим порядком для прогнозирования фактических данных



Автоматическое построение модели:





P-значения X членов меньше <0,05, и это хорошо.

Строим график и ряд для прогноза.

