

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ ДНР  
**ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ**

**ФИЗИКО-ТЕХНИЧЕСКИЙ ФАКУЛЬТЕТ**

**ЛАБОРАТОРНАЯ РАБОТА №3**

По теме: «Анализ временных рядов средствами языка R»

По предмету: «Машинное обучение»

Выполнил:

студент 2 курса

группы ИВТ-3

Дворников Г.Н.

Проверил:

Савенков И.Н.

Донецк  
2021

**Лабораторная работа № 3**

**Тема**: анализ временных рядов в пакете R.

**Цель:** изучить основные возможности языка R для исследования временных рядов, изучить особенности анализа временных рядов, структур данных и имеющийся статистический инструментарий.

**Ход** **работы**

Листинг программы:

from pandas import read\_csv, DataFrame

import statsmodels.api as sm

from statsmodels.iolib.table import SimpleTable

from statsmodels.tsa.arima.model import ARIMA

from sklearn.metrics import r2\_score

import ml\_metrics as metrics

import numpy as np

import matplotlib.pyplot as plt

dataset = read\_csv('tovar\_moving.csv',';', index\_col=['date'], parse\_dates=['date'], dayfirst=True)

dataset.head(15)

otg = dataset.qty

otg.head()

otg.plot(figsize=(12,6))

otg = otg.resample('W').mean()

otg.plot(figsize=(12,6))

itog = otg.describe()

otg.hist()

itog

print ('V = %f' % (itog['std']/itog['mean']))

row = [u'JB', u'p-value', u'skew', u'kurtosis']

jb\_test = sm.stats.stattools.jarque\_bera(otg)

a = np.vstack([jb\_test])

itog = SimpleTable(a, row)

print (itog)

test = sm.tsa.adfuller(otg)

print ('adf: ', test[0])

print ('p-value: ', test[1])

print ('Critical values: ', test[4])

if test[0]> test[4]['5%']:

print ('есть единичные корни, ряд не стационарен')

else:

print ('единичных корней нет, ряд стационарен')

otg1diff = otg.diff(periods=1).dropna()

test = sm.tsa.adfuller(otg1diff)

print ('adf: ', test[0])

print ('p-value: ', test[1])

print ('Critical values: ', test[4])

if test[0]> test[4]['5%']:

print ('есть единичные корни, ряд не стационарен')

else:

print ('единичных корней нет, ряд стационарен')

m = otg1diff.index[len(otg1diff.index)//2+1]

r1 = sm.stats.DescrStatsW(otg1diff[m:])

r2 = sm.stats.DescrStatsW(otg1diff[:m])

print ('p-value: ', sm.stats.CompareMeans(r1,r2).ttest\_ind()[1])

otg1diff.plot(figsize=(12,6))

fig = plt.figure(figsize=(12,8))

ax1 = fig.add\_subplot(211)

fig = sm.graphics.tsa.plot\_acf(otg1diff.values.squeeze(), lags=25, ax=ax1)

ax2 = fig.add\_subplot(212)

fig = sm.graphics.tsa.plot\_pacf(otg1diff, lags=25, ax=ax2)

src\_data\_model = otg[:'2013-05-26']

model = ARIMA(src\_data\_model, order=(1,1,1), freq='W').fit()

print(model.summary())

q\_test = sm.tsa.stattools.acf(model.resid, qstat=True)

print (DataFrame({'Q-stat':q\_test[1], 'p-value':q\_test[2]}))

pred = model.predict('2013-05-26','2014-12-31', typ='levels')

trn = otg['2013-05-26':]

r2 = r2\_score(trn, pred[1:34])

print ('R^2: %1.2f' % r2)

metrics.rmse(trn,pred[1:34])

metrics.mae(trn,pred[1:34])

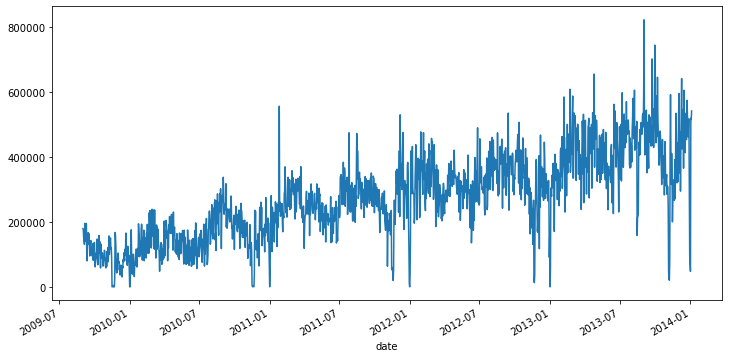
otg.plot(figsize=(12,6))

pred.plot(style='r--')

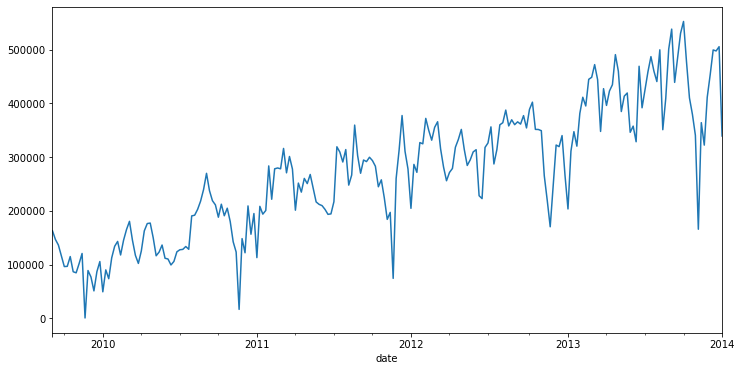
Описание набора данных:

|  |  |  |
| --- | --- | --- |
| **№** | **Обозначение** | **Описание** |
| 1 | data | Число, месяц, год поставки |
| 2 | qty | Количество товаров принято |

График нашего ряда:



Убираем выбросы:



Линия тренда:

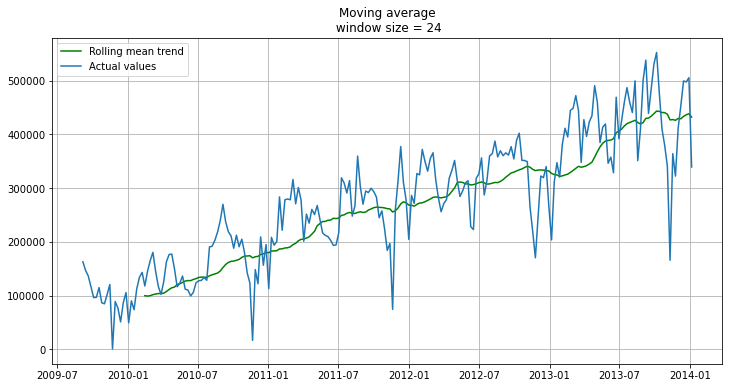
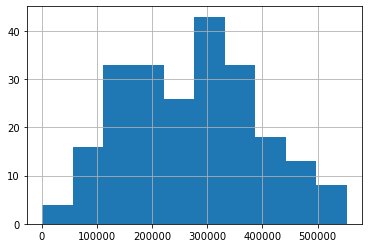


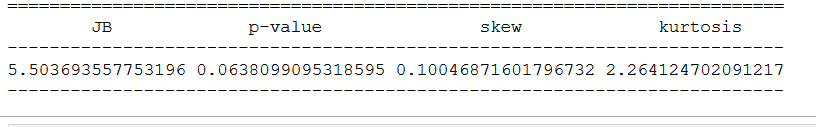
График однородности ряда:



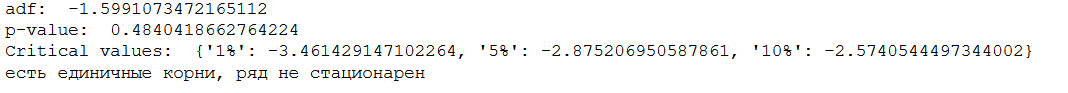
Коэффициент вариации:



Тест Харки – Бера:



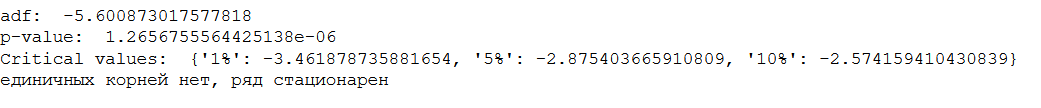
Обобщенный тест Дикки-Фуллера:



Определение порядка ряда



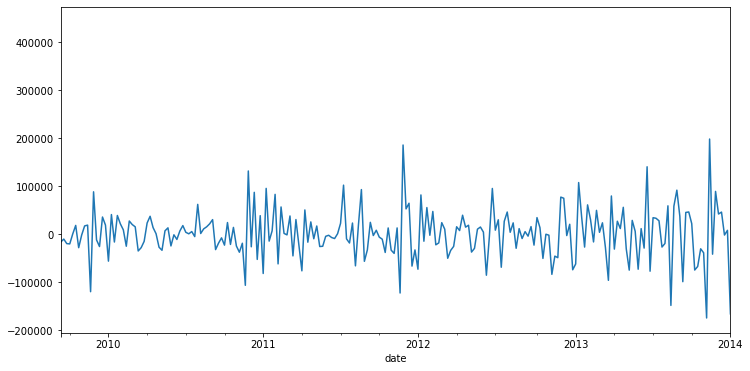
Проверим получившийся ряд на стационарность:



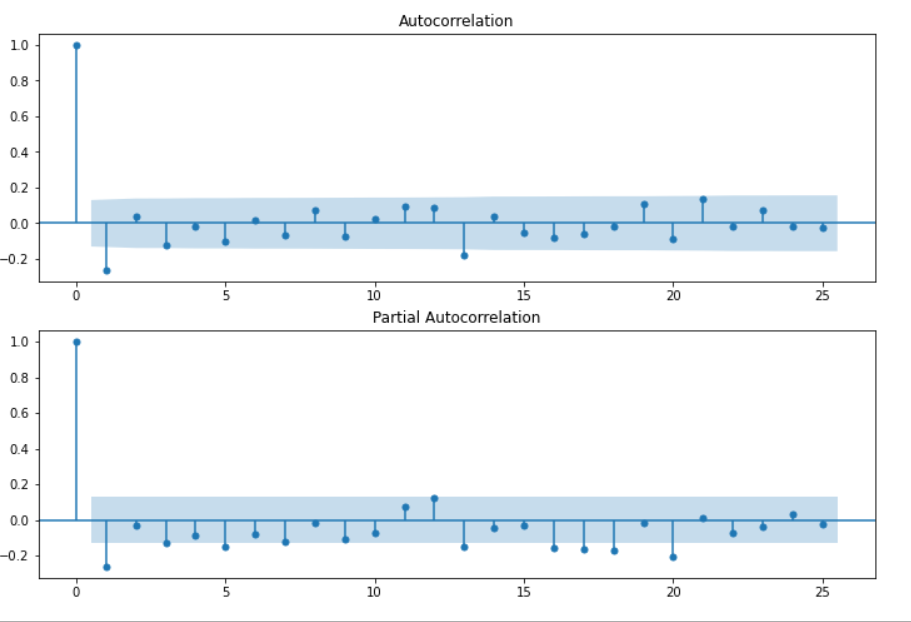
Среднее значение мат ожидания на разных интервалах:



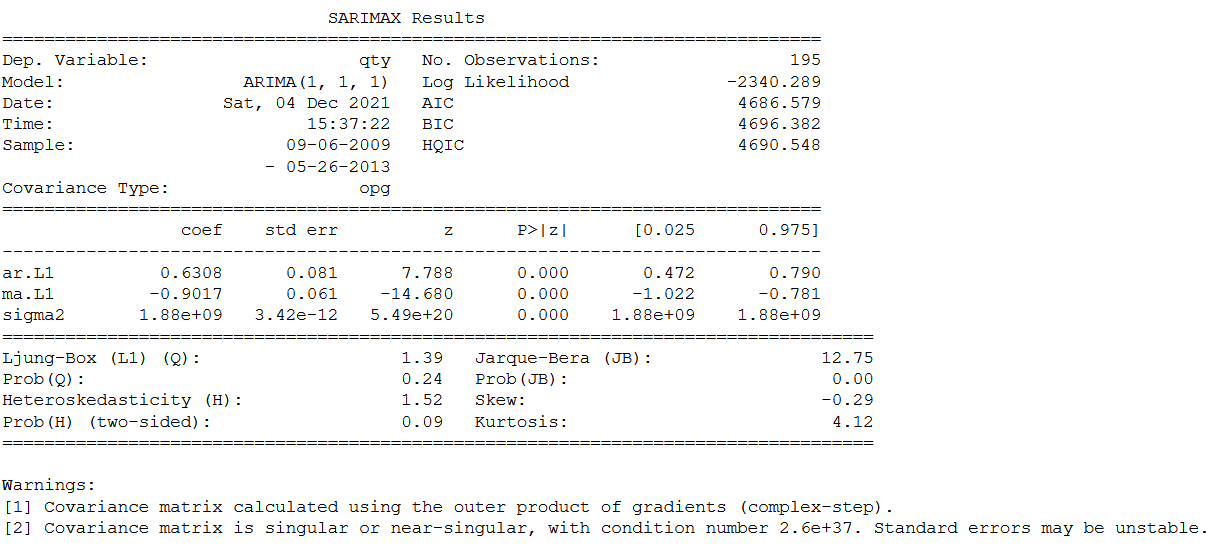
Исчезновение тренда в обработанном ряде:



Графики автокорреляции и частичной автокорреляции:



ARIMA модель:



Коэффициент детерминации:



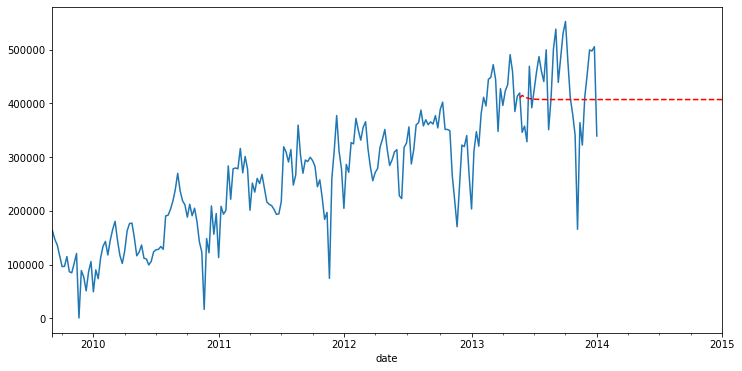
Среднеквадратичное отклонение модели:



Средняя абсолютная ошибкапрогноза:



Прогноз на графике:



**Вывод**: Как видно из графика модель не очень хорошо строит прогнозы. Это связано с выбросами в исходных данных, которые убраны по всей видимости не до конца.