Федеральное государственное автономное образовательное учреждение
высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет бизнеса и менеджмента

Домашняя работа

По дисциплине «Временные ряды»

Выполнила:

Студентка группы МЭВ-181

Алеева Вера Александровна

Москва 2019

Вариант 40

Исходный ряд:

21.3333,22.6667,27.8294,30.4124,32.9520,14.2640,28.8415,27.8419,53.2665,19.1
559,16.7668,55.6615,35.6406,30.3929,50.3441,40.2123,28.7693,33.9378,58.5428,
44.5764,49.3091,50.4247,41.9638,62.7497,56.3852,41.9907,61.1028,58.6780,73.2
081,66.2684,54.4457,67.6482,56.9542,64.9572,69.7775,61.5698,84.9554,57.4548,
58.0847,88.6711,75.2904,79.9606,70.9618,83.2560,85.7562,76.4391,77.6164,85.2
613,104.0068,74.5042,86.8909,94.7717,103.2406,100.7451,97.0038,78.8327,96.9
834,102.6197,98.0549,83.8813,111.9083,118.8757,94.5341,105.2204,115.1087,11
9.6549,104.8529,107.0710,129.1234,109.7078,106.9881,139.7480,101.7890,103.0
492,130.6339,114.8879,126.8285,126.5739,125.0049,143.9181,111.3871,139.348
3,131.1971,139.8066,134.0294,131.5055,138.3898,130.1479,142.5471,151.6323,1
43.3908,134.7999,151.3457,147.6873,143.3519,151.3348,143.1829,160.5847,147.
4983,163.1111,154.0063,155.9284,150.6340,185.3708,156.5745,161.2122,160.30
83,164.8771,149.9788,186.9586,174.4658,157.6656,166.4182,185.9162,175.4570,
180.4779,183.1741,179.0722,176.4521,182.3102,180.6805,192.9606,166.6624,19
6.1941,204.3302,166.0524,197.4435,203.4331,182.9194,199.0849,195.7082,192.6
015,182.5659,198.6555,192.8288,200.6871,203.5176,204.4451,199.6403,193.458
9,216.7572,202.2302,214.3964,220.0236,206.6291,198.8249,207.9889,231.3487,2
10.9570,215.7860,217.7013,217.5601,235.6657,211.7818,237.4417,227.7584,235.
0806,218.3528,230.5214,246.4117,227.7168,210.6495,264.0180,237.5281,251.98
05,235.9781,247.4135,230.6109,243.7570,256.7107,252.7107,233.9085,269.4156,
247.2615,240.1998,246.4124,267.3122,258.3539,256.3502,252.6688,262.7675,25
7.0177,269.8386,262.4560,267.8276,273.2983,258.8461,282.0778,265.0453,261.5
477,273.9821,277.2504,283.9545,268.6449,280.3142,285.1080,282.3552,299.604
9,286.6788,264.3836

Задания:

1. Используя процедуру Доладо-Дженкинса-Сосвилла -Риверо, выяснить, относится ли наблюдаемый ряд к типу TSP.
2. Для ряда типа TSP оценить по МНК детерминированную составляющую ряда.
3. Детрендировать ряд.
4. Провести идентификацию случайной составляющей ряда, выбрав в качестве базовых 2-3 подходящие модели.
5. Оценить параметры выбранных моделей.
6. Используя информационные критерии Акаике и Шварца, выбрать адекватную модель.
7. Провести диагностику остатков.
8. Построить прогноз на один шаг.

1. Для начала был построен график ряда (рис.1), на который также была нанесена линия тренда:

Рисунок 1

На основе графика можно сделать вывод, что ряд имеет четко-выраженный линейный тренд. Для определения принадлежности ряда к типу TSP или DSP проведем процедуру Доладо-Дженкинса-Сосвилла-Риверо, начиная с расширенного теста Дики-Фуллера. На данном шаге проверяется гипотеза о наличии единичного корня в модели

$$(1)\ x_t = \mu + \beta t + \alpha_1 x_{t-1} + \cdots + \alpha_n x_{t-n} + \varepsilon_t\ :$$

Преобразуем модель (1) к виду:

$$(2)\ \Delta x_t = x_t - x_{t-1} = \mu + \beta t + \gamma x_{t-1} + \theta_1 \Delta x_{t-1} \ldots + \theta_{n-1} \Delta x_{t-n+1} + \varepsilon_t,$$

где $\gamma = \sum_{i=1}^{n} \alpha_i - 1, \quad \theta_i = -(\alpha_{i+1} + \cdots + \alpha_n).$

В данном случае проверяются гипотезы $H_0: \gamma = 0 \qquad H_A: \gamma < 0$

Для того, чтобы определить, сколько лагов включить в модель, рассчитаем оценки модели с максимальным числом включаемых лагов $\sqrt[3]{n}$, где n – это количество наблюдений, в данном случае равное 200.

Оценки модели с константой, трендом и 6 запаздываниями.

	coef	std err	t	P>|t|
const	39.8818	5.289	7.539	0.000
t	2.6108	0.407	6.414	0.000
L	-0.4679	0.076	-6.360	0.000
L^2	-0.4260	0.081	-5.249	0.000
L^3	-0.0961	0.087	-1.097	0.274
L^4	0.0093	0.087	0.105	0.916
L^5	-0.0014	0.082	-0.017	0.987
L^6	-0.0112	0.074	-0.151	0.880

По значению p-value видно, что значимыми оказываются только первые два лага, следовательно, далее будем использовать модель

(*) $\Delta x_t = \mu + \beta t + \gamma x_{t-1} + \theta_1 \Delta x_{t-1}$.

Найдем расчетное значение ADF-теста с помощью библиотеки statsmodels и встроенной функции adfuller в Python. На вход данной функции список, состоящий из значений временного ряда, указывается тип модели (с константой и трендом, с константой, без константы и без тренда), а также максимальное число лагов, включаемых в модель. В данном случае была указана модель с константой и трендом, максимальное число лагов 2. В результате были получены следующие оценки:

```
Расчетное значение ADF-теста:  -11.780478271034044
p-value:  8.84431704233448e-19
```

Полученное значение статистики $\frac{\hat{\gamma}}{s.e.\hat{\gamma}}$ = -11.78 сравниваем с критическим

```
1%: -3.4645146202692527
5%: -2.8765564361715534
10%: -2.5747745328940375
```

Как видно, расчетное значение теста меньше, чем критическое на всех разумных уровнях значимости, а значит H_0 отвергается, ряд относится к типу TSP и верна модель (*).

2. Оценим детерминированную составляющую ряда с помощью метода наименьших квадратов. Получаем следующие результаты:

```
Коэффициент при t: 1.3231866037650946
Независимый коэффициент 21.35046132160801
Коэффициент детерминации R^2: 0.9838882305069662
```

Таким образом, получаем следующее уравнение тренда:

$$Y_t = 21{,}35 + 1{,}32 \cdot t \quad (3)$$

3. Детрендируем ряд, вычитая из него значения, полученные с помощью уравнения 1. В результате получаем ряд, представленный на графике (рис.2):

Рисунок 2.

4. Для идентификации случайной составляющей ряда, оценим значения автокорреляционной и частной автокорреляционной функции (рис.3)

```
Оценки АКФ:1.0 0.30526821 -0.24707727  0.14134189  0.06739749 -0.07858409
  0.00101671  0.08155622  0.03042871 -0.20562257  0.20416172 -0.05488077
 -0.01463683  0.03931343  0.06915498 -0.05136239 -0.02189363  0.04048514
  0.05741984 -0.19091342]
Оценки ЧАКФ:1.00000000e+00 -3.06802221e-01 -3.79413921e-01 -1.01270295e-01
 -1.88060538e-03 -1.84995495e-02  7.17560067e-04  7.22620256e-02
  1.19763206e-01 -1.45736161e-01  1.28533114e-01 -6.54746152e-02
  6.57860687e-02  4.19940481e-02  1.32527427e-01  6.08480617e-02
  3.89862017e-02  3.88801408e-02  4.89638487e-02 -1.43415866e-01]
Уровень значимости alpha=0.05
```

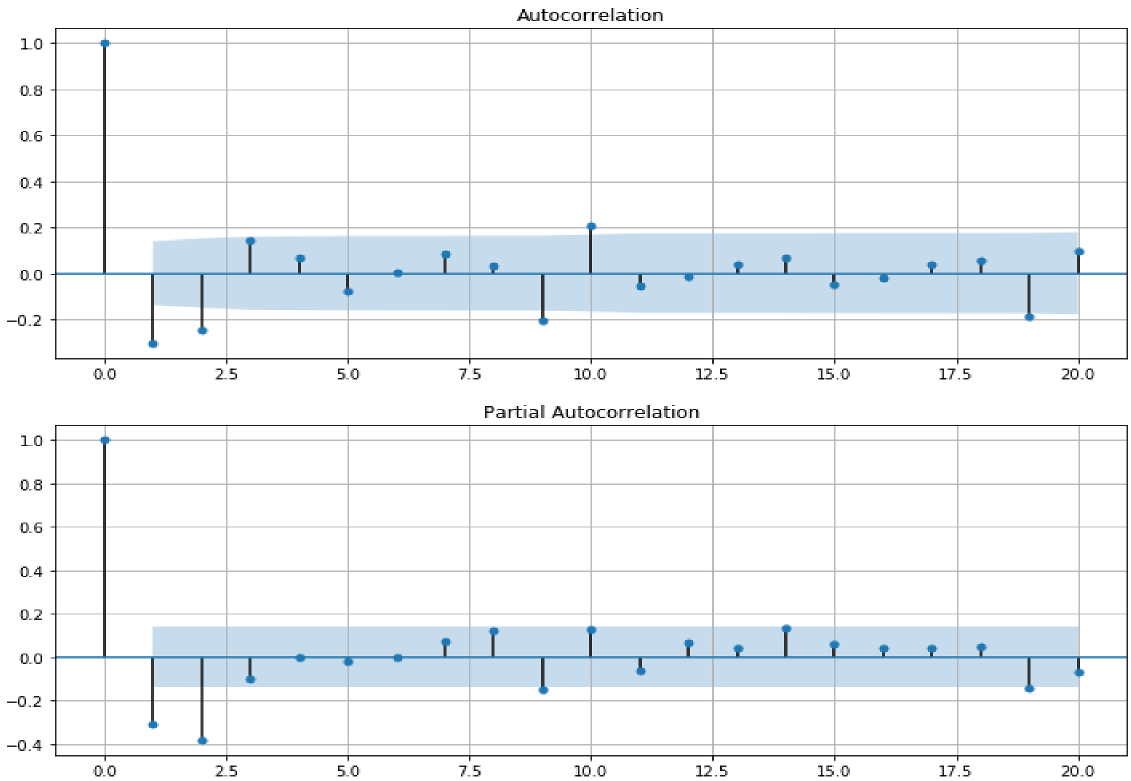

Рисунок 3.

Видно, что оценки АКФ экспоненциально убывают и после второго лага попадают в доверительный интервал. На 9,10 и 19 лаге оценки автокорреляции отличны от 0, что можно считать случайностью на уровне значимости 5%. По графику ЧАКФ видно, что оценки первых двух лагов значительно отличаются от нуля, а начиная с третьего попадают в доверительный интервал. Таким образом следует предположить, что данный ряд представляет собой AR(2)-процесс.

5.Оценим модель AR(2):

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  200
Model:                     ARMA(2, 0)   Log Likelihood                -714.076
Date:                Tue, 26 Mar 2019   AIC                           1434.152
Time:                        14:09:31   BIC                           1444.047

==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y       -0.4313      0.066     -6.529      0.000      -0.561      -0.302
ar.L2.y       -0.3842      0.066     -5.844      0.000      -0.513      -0.255
```

Также возьмем еще две модели в качестве базовых.

AR(1):

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  200
Model:                     ARMA(1, 0)   Log Likelihood                -729.781
Date:                Tue, 26 Mar 2019   AIC                           1463.562
Time:                        14:12:08   BIC                           1470.159

==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y       -0.3113      0.068     -4.587      0.000      -0.444      -0.178
```

AR(3):

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  200
Model:                     ARMA(3, 0)   Log Likelihood                -713.085
Date:                Tue, 26 Mar 2019   AIC                           1434.170
Time:                        14:13:08   BIC                           1447.364

==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y       -0.4684      0.071     -6.616      0.000      -0.607      -0.330
ar.L2.y       -0.4271      0.072     -5.922      0.000      -0.568      -0.286
ar.L3.y       -0.1001      0.071     -1.411      0.160      -0.239       0.039
```

4. Для выбора модели необходимо сравнить их показатели, а именно критерии Акайке и Шварца. Информационный критерий AIC для модели ARMA(p,q), где p и q соответствующие порядки модели, выглядит следующим образом:

$$AIC(p,q) = -2\ln L + 2k,$$

L – максимизированное значение функции правдоподобия модели, $k = p + q + 1$. Критерий Шварца имеет вид:

$$BIC(p,q) = -2\ln L + \ln n \cdot k,$$

n – число наблюдений. Соответственно, чем меньше значение критериев, тем лучше модель описывает процесс. Сравним информационные критерии Акайке и Шварца для выбранных моделей в Таблице 1

Таблица 1

Сравнение информационных критериев моделей

	AIC	BIC
AR(1)	1465,56	1475,45
AR(2)	1434,15	1444,05
AR(3)	1436,14	1452,63

Видно, что самые низке показатели имеет модель AR(2), следовательно выбираем ее для моделирования процесса.

5. Проведем диагностику остатков модели. График остатков представлен на рисунке 4:

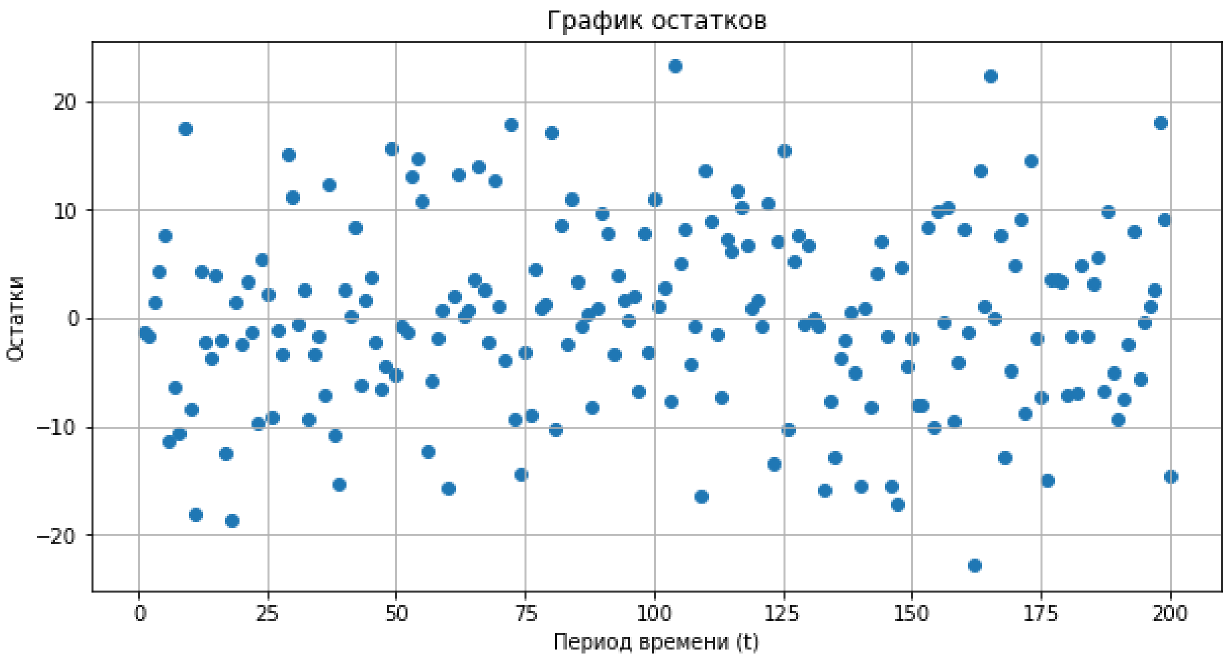

Рисунок 4.

Так как применение ADF-теста предполагает, что остатки имеют нормальное распределение, проведем тест Харке-Бера. Статистика рассчитывается по формуле:

$$JB = \frac{n}{6}[S^2 + \frac{(K-3)^2}{4}]$$

Где $S = \frac{1}{n}\sum_{i=1}^{n}(\frac{\varepsilon_i - \bar{\varepsilon}}{\hat{\sigma}})^3$ – коэффициент асимметрии, $K = \frac{1}{n}\sum_{i=1}^{n}(\frac{\varepsilon_i - \bar{\varepsilon}}{\hat{\sigma}})^4$ - коэффициент эксцесса, $\hat{\sigma}$ – дисперсия остатков.

$H_0: \varepsilon_i \sim N(0; \sigma^2)$

Тестовая статистика имеет распределение χ^2 с двумя степенями свободы. Нулевая гипотеза о нормальности распределения остатков ряда отвергается на уровне значимости 5%, если $JB > JB_{\text{крит}} = 5{,}9916$

```
K: [2.80432656]
S: [0.04653284]
JB:[0.3912443]
p-value: [0.82232289]
```

$JB < JB_{\text{крит}}$, следовательно H_0 не отвергается и остатки имеют нормальное распределение.

Для проверки остатков модели на коррелированность проведем тест Бокса-Пирса. Тест Бокса-Пирса проверяет гипотезу о совместном равенстве нулю всех автокорреляций временного ряда до порядка m включительно.

$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$ (т.е. остатки не коррелированы)

$$H_A: \sum_{k=1}^{m} \rho_k^2 > 0$$

Статистика рассчитывается по формуле

$$Q_{BP} = n\sum_{k=1}^{m} \hat{\rho}_k{}^2,$$

где n – число наблюдений, $\hat{\rho}_k$ – автокорреляция k-го порядка, m – число проверяемых лагов. Гипотеза адекватности подобранной модели отвергается на уровне значимости 5%, если $Q_{BP} > \chi 2(k - p - q)$ и p-value не превышает 5%.

Вычисляем статистику для m=20:

```
Q-расчетное Box-Pierce: [ 0.31874688  0.44678747  1.43253791  2.01805262
2.20085367  2.76458293
2.84826905  3.57025916  8.54828076 13.24999481 13.58384581 15.21272407
15.93178109 17.6753303  17.71117916 17.8015899  17.90467575 17.99749421
22.66742425 23.92662737]
p-value:[0.5723617  0.79979988 0.6979257  0.73243833 0.82071265 0.83776084
0.89867477 0.89366896 0.47997094 0.21002829 0.25688185 0.23000884
0.25283328 0.22197031 0.27815402 0.33564281 0.3948784  0.45581769
0.25230814 0.24561194]
```

Видно, что p-value для всех лагов больше 5%, а значит H_0 не отвергается и остатки не коррелированы.

6. Построим прогноз на один шаг.

Коэффициенты модели AR(2):

AR(1) = -0,4313 AR(2) = -0,3842

Соответствующие значения детрендированного временного ряда:

$x_{200} = -21{,}60 \quad x_{199} = 2{,}014$

Уравнение тренда:

$Y_{201} = 21{,}35 + 1{,}32 \cdot 201$

Прогноз для детрендированного ряда:

$$-0{,}4313 * (-21{,}60) - 0{,}3842 * 2{,}014 = 8{,}544$$

Добавляем трендовую составляющую и получаем

$$x_{201} = 21{,}35 + 1{,}32 \cdot 201 + 8{,}544 = 295{,}214$$