Практическое занятие 11. **ARFIMA-модели.**

Задача 1. Запишите разложение $(1\text{-}L)^{0.3}$.

Замечание. Используйте формулу биномиального разложения (ряд Маклорена)

$$(1+x)^\alpha = 1 + \sum_{n=1}^{\infty} \binom{\alpha}{n} x^n, \text{где } \binom{\alpha}{n} = \prod_{k=1}^{n} \frac{\alpha - k + 1}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}$$

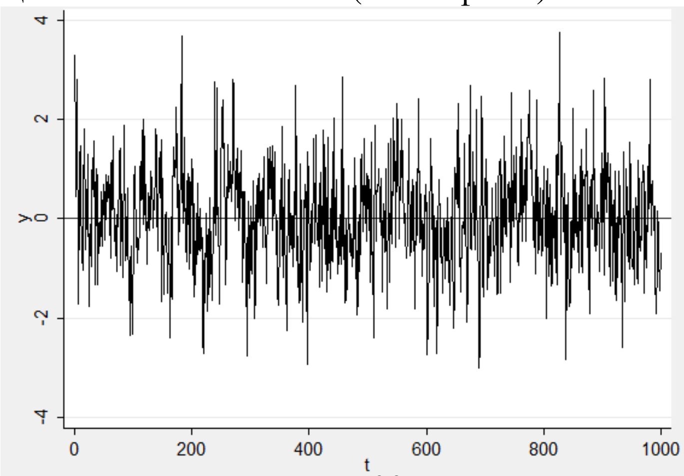
https://ru.wikipedia.org/wiki/%D0%A0%D1%8F%D0%B4_%D0%A2%D0%B5%D0%B9%D0%BB%D0%BE%D1%80%D0%B0

Запишите разложение $(1\text{-}L)^{0.1} y_{t} = \varepsilon_t$.

Задача 2. Запишите модели в общем виде

ARFIMA(0, 0.3, 0)
ARFIMA(1, 0.45, 1)
ARFIMA(2, 1.3, 1)

Задача 3. (Stata) Генерирование процесса ARFIMA. Как сгенерировать случайный процесс ARFIMA с d=0.3? (см do-файл).



Сгенерируйте процесс $(1\text{-}L)^{0.3} y_t$ в Stata. Проанализируйте поведение графика процесса и ACF/PACF.

Задача 4. (Excel). Показатель Херста. R/S анализ.

Рассчитайте экспоненту Херста и определите параметр **d** .

Файл: RS.xls

Замечание. Для выявления длинной памяти Б. Мандельброт предложил использовать R/S статистику (rescaled range), придуманную Х. Хёрстом (1951). R/S статистика определяется размахом частичных сумм отклонений ряда от его среднего, деленного на его стандартное отклонение.

- Рассмотрим ряд $y_1, y_2, \dots y_T$
- Делим ряд на несколько интервалов n=T, n=T/2, n=T/4, n=T/8 и т.д.
- Рассчитываем и сравниваем частичные суммы

$$\left(\tfrac{R}{S}\right)_t = \tfrac{1}{\hat{\sigma}_t}\left(\max_{1<j<T} \sum_{j=1}^{t} (y_j - \bar{y}) - \min_{1<j<T} \sum_{j=1}^{t} (y_j - \bar{y}) \right)$$

$$\hat{\sigma}^2 = \tfrac{1}{T} \sum_{i=1}^{T} (y_i - \bar{y})^2, \bar{y} = \tfrac{1}{T} \sum_{i=1}^{T} y_i$$

Hurst H. E. Long term Storage Capacity of Reservoirs // Transactions of the American Society of Civil Engineers. 1951. N 116. P. 770–799.

Mandelbrot B. Statistical Methodology for Non-Periodic Cycles:From the Covariance to R/S Analysis // Annals of Economic and Social Measurement. 1972. N 1. P. 259–290;

Величина d связана с экспонентой Херста H из уравнения **$\ln(R/S)_n = \ln c + H \ln n + u$** равенством **H = d + 0.5**.

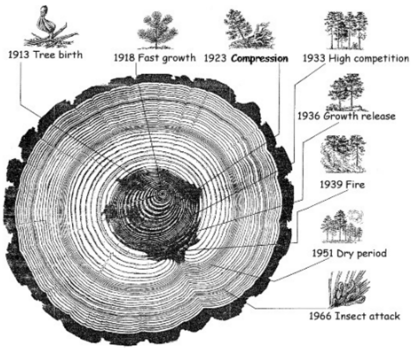
Задача 5. (Stata открыть в Gretl)**. Анализ процесса с длинной памятью.**
Данные: Mount Campito tree ring data.

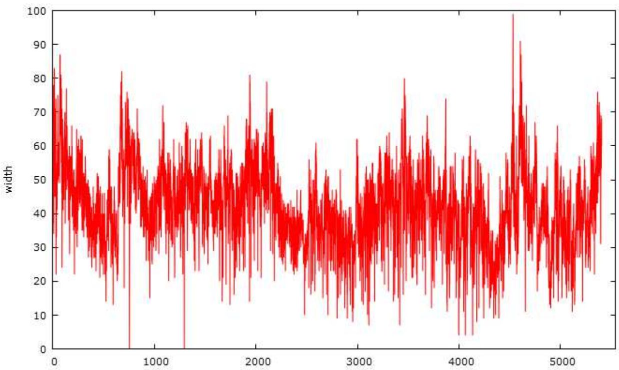
Baillie (1996) discusses a time series of measurements of the widths of the annual rings of a Mount Campito Bristlecone pine. The series contains measurements on rings formed in the tree from 3436 BC to 1969 AD. Essentially, larger widths were good years for the tree and narrower widths were harsh years.

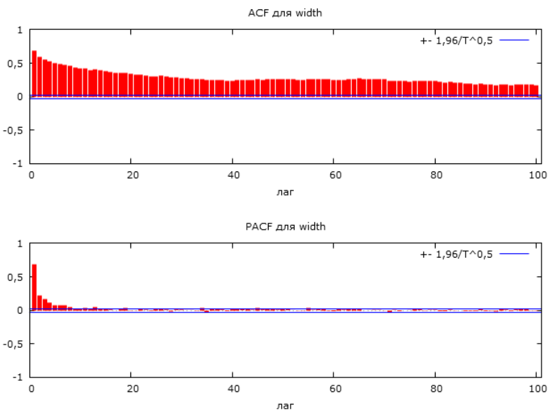
use http://www.stata-press.com/data/r13/campito
*Campito Mnt. tree ring data from 3435BC to 1969AD

Файл: **campito.gdt.** Откройте данные в Gretl.

1. **Поведение ACF.** Постройте график ряда width и его ACF/PACF. Как ведет себя ACF? Убывание с медленной гиперболической скоростью?







2. **Стационарность процессов.** Протестируйте процесс на наличие стохастического тренда (используйте три вида тестов ADF, PP, KPSS). Стационарен/нестационарен процесс согласно критериям?

Замечание. Проблема обнаружения «длинной памяти». Широко используемые тесты Дики–Фуллера и Филипса–Перрона на наличие единичных корней обладают малой мощностью и плохо отличают I(1) процессы от I(d) процессов с d<1. Тест Квятковского–Филипса–Шмидта–Шина (KPSS), имеющий нулевую гипотезу о стационарности, состоятелен при стационарных I(d) процессах с $|d| < 0.5$, но требует большого количества наблюдений (не менее 1000).(Подкорытова)

Свойства процесса:

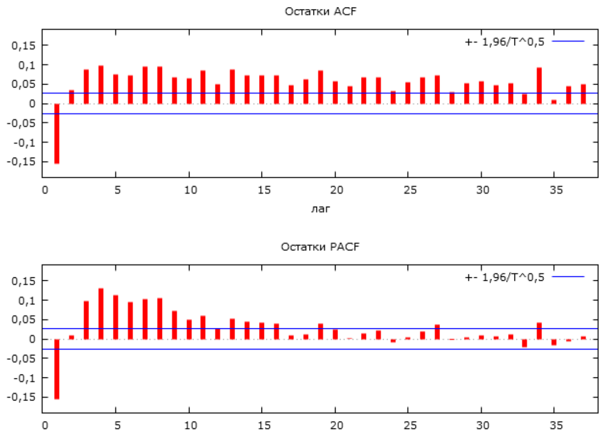
$d < 0{,}5$ процесс стационарен,

$d > -0{,}5$ процесс обратим.

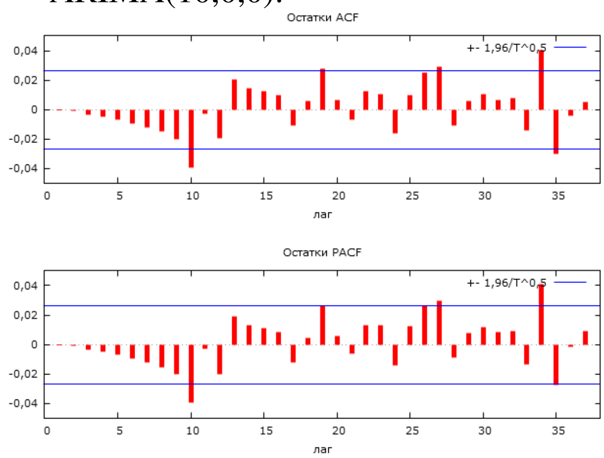
Условия $-0{,}5 < d < 0{,}5$ всегда можно добиться, применив необходимое количество обычных дифференцировании.

3. **ARIMA и их недостаток.** Для ряда оцените различные варианты моделей:
ARIMA(1,0,0), ARIMA(1,1,0),
ARIMA(5,0,0), ARIMA(5,1,0),
ARIMA(10,0,0), ARIMA(10,1,0),
в чем недостаток этих моделей? Как ведут себя ACF/PACF остатков? Удалось ли подобрать адекватную модель для прогнозирования? Почему?

ARIMA(1,0,0):



ARIMA(10,0,0):



Замечание. Большое число параметров ARMA моделей высокого порядка, которое потребуется оценить, сделают модель низкого качества согласно критериям Акаике и Шварца. А дробно-интегрируемые процессы «схватывают» эту структуру одним параметром d. (Подкорытова)

4. **Показатель Херста.**
4.1. Оцените показатель Херста (в Gretl) и вычислите порядок дробной разности d=H-0.5.

```
Rescaled range figures for width
(логарифмы имеют основание 2)

Размер      RS(avg)  log(Размер)     log(RS)
  5405       914,52       12,400      9,8369
  2702       338,88       11,400      8,4046
  1351       271,14       10,400      8,0829
   675       161,24       9,3987      7,3331
   337       75,620       8,3966      6,2407
   168       35,913       7,3923      5,1664
    84       20,582       6,3923      4,3633
    42       11,158       5,3923      3,4800
    21       6,1480       4,3923      2,6201
    10       3,3786       3,3219      1,7564

Результаты регрессии (n = 10)

                     коэфф.    Ст. ошибка
   Константа        -1,2273      0,18262
  Угл. коэф.        0,88175     0,021741

Оценка модели экспонента Хёрста = 0,881745
```

4.2. Оцените модели ARFIMA(1,0,0), ARFIMA(2,0,1), в Gretl, взяв соответствующую дробную разность. Запишите полученные модели (в явном виде и с использованием лагового оператора). Проанализируйте остатки.
4.3. Сравните результаты оценивания ARFIMA-моделей с параметром **d=0,44,** оцененным в Stata. (ARFIMA(0,0,2), ARFIMA(2,0,2)). Проанализируйте остатки.

Задача 5. Подход Ло. Проведите тест Ло в Gretl. Сделайте вывод о наличии долгосрочной памяти.

Замечание. Недостатки R/S-статистики - чувствительность к краткосрочным зависимостям и гетероскедастичности. Э. Ло (1988, 1991) модифицировал статистику, преобразовал σ. Ло (1988, 1991) вывел таблицу критических значений - интервалы, при попадании в которые статистики, нулевая гипотеза не отвергается.

Lo's (1991) modified R/S test

Null Hypothesis: short-range dependence

Alternative hypothesis: long-range dependence

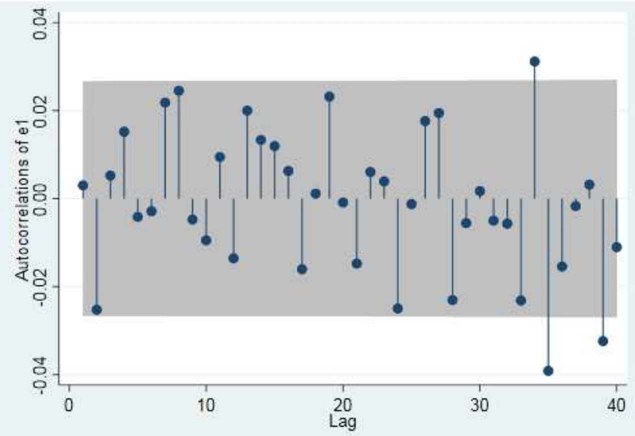
```
? ModRS_test(width, 10, 2)
-------------------------------------------------
      Lo's(1991) Modified R/S (Hurst) Test
-------------------------------------------------
Null hypothesis: short-range memory
Alternative hypothesis: long-range memory
-------------------------------------------------
Note: this function uses the built-in function
lrvar to compute the long-run variance (lrvar)
-------------------------------------------------
Lo's modified R/S statistic: 0,16179
-------------------------------------------------
Critical values for 5 percent: 0,809 and 1,862
H0 rejected at 5 percent
-------------------------------------------------
```

Задача 6. Оценивание дробноинтегрированного процесса в Stata.
Данные: Mount Campito tree ring data.

Файл: **campito.dta.** Откройте данные в Stata и оцените ARFIMA-модель. Проанализируйте остатки. Сравните с результатами, полученными в Gretl.

Оценивание ARFIMA (Stata 13)	
ssc install lomodrs lomodrs width arfima width predict e1, residual ac e1 wntestq e1 pac e1	arfima y, nonconstant arfima y, ar(1) arfima y, ar(1) ma(1)

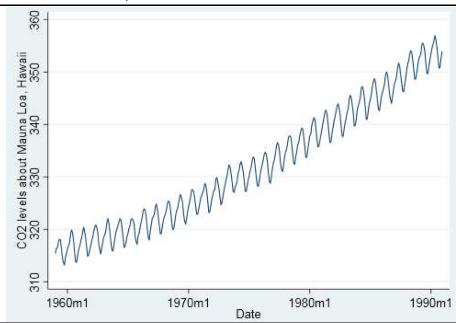
```
. lomodrs width

Lo Modified R/S test for width

Critical values for H0: width is not long-range dependent

90%: [ 0.861, 1.747 ]
95%: [ 0.809, 1.862 ]
99%: [ 0.721, 2.098 ]

Test statistic:     3.03  (38 lags via Andrews criterion)  N = 5367
ARFIMA regression

Sample: -3435 - 1969                         Number of obs     =      5,405
                                             Wald chi2(1)      =    1864.44
Log likelihood = -18907.279                  Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OIM
       width |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
width        |
       _cons |   44.01432   9.174317     4.80   0.000     26.03299    61.99565
-------------+----------------------------------------------------------------
ARFIMA       |
           d |   .4468887   .0103496    43.18   0.000     .4266038    .4671737
-------------+----------------------------------------------------------------
     /sigma2 |   63.92927   1.229754    51.99   0.000       61.519    66.33955
------------------------------------------------------------------------------
```



Задача 7. (самостоятельно)

Данные: log of the monthly levels of carbon dioxide above Mauna Loa, Hawaii.

Мау́на-Ло́а (англ. Mauna Loa — «длинная гора») — действующий щитовой вулкан высотой 4169 метров на острове Гавайи.



Файл: **mloa.dta.**

Откройте данные в Stata. Исследуйте свойства длинной памяти для показателя log. В случае необходимости возьмите сезонную разность. (S12.log)

Оцените ARFIMA-модель.

Сравните модели:

arima S12.log, ar(1) ma(2)

arfima S12.log, ar(1) ma(2)

Проанализируйте остатки.

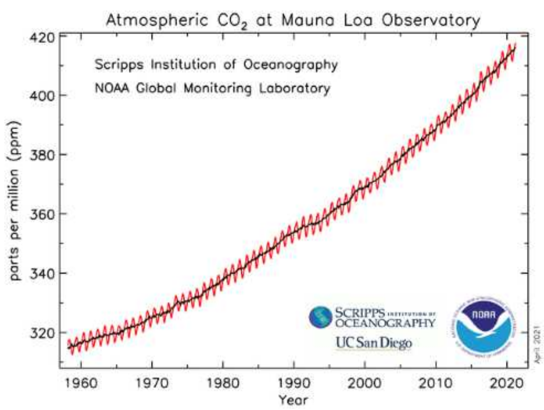
Домашнее задание (ТДЗ) 11. ARFIMA

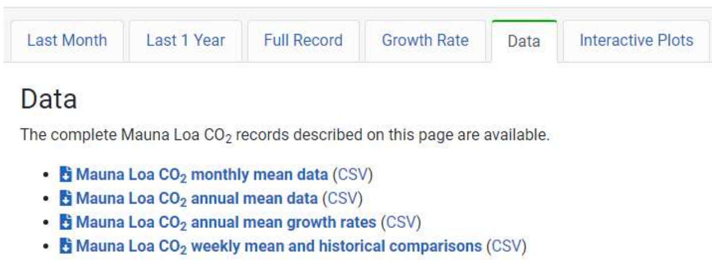
По данным Mauna Loa Observatory (Hawaii) выберите один показатель выброс CO2 (опишите какой показатель был взят для анализа, за какой период).

Global Monitoring Laboratory https://www.esrl.noaa.gov/gmd/ccgg/trends/

Данные: https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html

!Можно взять свои данные





Задание. Требуется подобрать соответствующую модель ARFIMA для описания динамики *выбранного показателя*, оценить ее параметры и построить прогноз на основании полученной модели.

1. Опишите какой показатель был взят для анализа, за какой период, постройте график и сделайте предположение о стационарности ряда. В случае необходимости возьмите несезонную и/или сезонную разность. (S12.log или DS12.log)
2. Проведите тесты единичного корня (ADF, PP, KPSS) и их модификации. Сравните результаты и сделайте вывод по результатам тестирования.
3. Проверьте гипотезу о наличии долгосрочной памяти.
4. Оцените ARFIMA-модель. Запишите модели в математической форме через *лаговый* оператор. Опишите статистические свойства модели (значимость коэффициентов, ошибка модели, информационные критерии).
5. Проверьте адекватность моделей на основе анализа остатков (автокорреляция, нормальность)).
6. Постройте прогноз на 12 шагов вперед (точечная и интервальная оценка) по наилучшей модели. В отчете приведите графики

(наблюдаемые+предсказанные значения). Сделайте вывод как изменится показатель.

7. Запишите разложение $(1\text{-}L)^{-0.4}y_{t=}\ \varepsilon_t$.