

Практическое занятие 3. «Анализ неслучайной составляющей временного ряда»

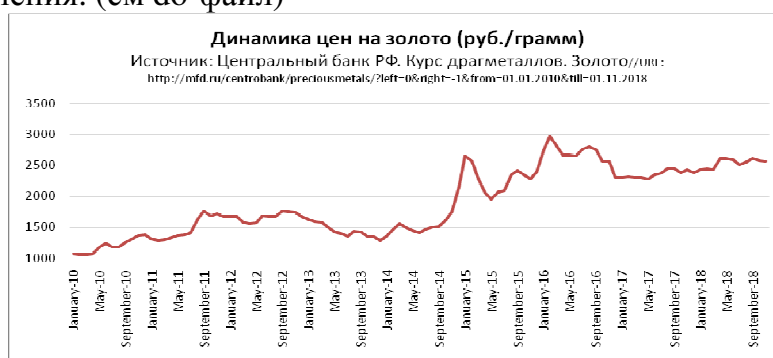
1. Тест по занятию 1-2.
2. Проверка гипотезы о случайности (неизменности среднего значения ВР) (критерий серий, основанный на медиане выборки, runtest)
3. Выделение трендовой составляющей. Скользящие средние.

1. Проверка гипотезы о неизменности среднего значения ВР (критерий серий, основанный на медиане выборки)

1. В чем суть критерия серий, основанного на медиане выборки?

Задание. Изучается динамика цен на золото. **Файл gold.dta.**

Рассмотрите данные за последние два года. Что можно сказать о трендовой составляющей? Чему равна медиана ряда? Проверить гипотезу о неизменности среднего значения. (см do-файл)



t	Стоимость золота (тыс. руб./гр), y_t	Серии
January-17	2,30	
February-17	2,31	
March-17	2,30	
April-17	2,30	
May-17	2,28	
June-17	2,35	
July-17	2,37	
August-17	2,45	
September-17	2,44	
October-17	2,38	
November-17	2,43	
December-17	2,39	
January-18	2,43	
February-18	2,44	
March-18	2,43	
April-18	2,61	
May-18	2,61	
June-18	2,59	
July-18	2,50	
August-18	2,55	
September-18	2,61	
October-18	2,57	

November-18	2,57	
-------------	------	--

Сделайте вывод о наличии трендовой составляющей в изучаемом ВР.

1. 2. Используйте непараметрический критерий случайности в Stata: **runtest (Test for random order)** (Swed and Eisenhart (1943)) для ряда «Стоимость золота»

Подробности см Help Stata.

runtest begins by calculating the number of observations below the threshold, n_0 ; the number of observations above the threshold, n_1 ; the total number of observations, $N = n_0 + n_1$; and the number of runs, r . These statistics are always reported, so the exact tables of critical values in Swed and Eisenhart (1943) may be consulted if necessary.

The expected number of runs under the null is $\mu_r = \frac{2n_0n_1}{N} + 1$

the variance is $\sigma_r^2 = \frac{2n_0n_1(2n_0n_1 - N)}{N^2(N - 1)}$

and the normal approximation test statistic is $\hat{z} = \frac{r - \mu_r}{\sigma_r}$

Swed, F. S., and C. Eisenhart. 1943. Tables for testing randomness of grouping in a sequence of alternatives. *Annals of Mathematical Statistics* 14: 66–87.

Аналог теста есть в Gretl: Инструменты – Непараметрические тесты – Критерий серий

Результат:

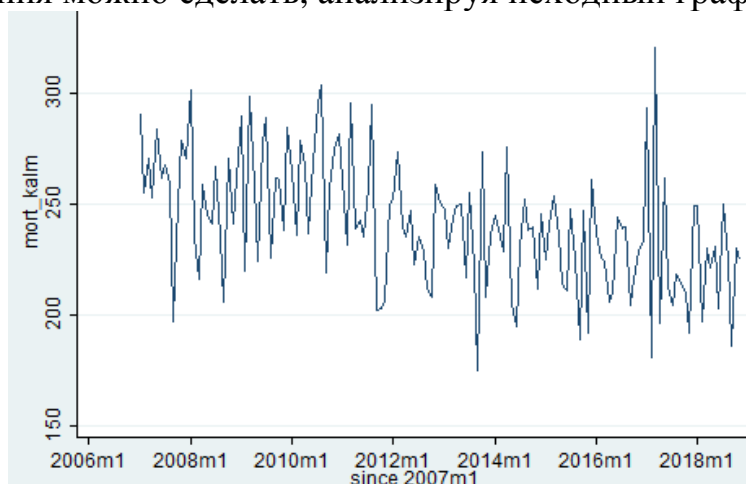
```
. runtest gold
N(gold <= 2.430000066757202) = 12
N(gold > 2.430000066757202) = 11
      obs = 23
      N(runs) = 6
      z = -2.77
      Prob>|z| = .01
```

Сравните с результатами полученными ранее.

1.3. **Файл: runtest.dta** . Используйте критерий случайности (**runtest**) для белого шума, сделайте вывод о неизменности среднего значения ВР.

1.4. **Файл: runtest.dta** . **Сезонность**. Для показателя смертности (число умерших (чел.))в Калмыкии проверьте гипотезу о неизменности среднего значения.

Какие предположения можно сделать, анализируя исходный график ВР?



По данным 2017 г. в 5 регионах-субъектах федерации число умерших увеличилось, наиболее значительно в Республике Калмыкии (на 1,7%).

Щербакова Е.М. Россия: предварительные демографические итоги 2017 года (часть II) //Демоскоп Weekly. 2018. № 761-762. URL: <http://demoscope.ru/weekly/2017/0761/barom01.php>

Сделайте свой вывод, как меняется показатель в период с 2007 года? Как влияет наличие сезонности на результаты проверки гипотезы?

2. Выделение трендовой составляющей. Скользящие средние.

2.1. В чем суть построения простой скользящей средней (СС)? Рассчитайте по формулам значение простой СС по 3,5 точкам (k-длина интервала сглаживания) для марта-апреля 2018

t	Стоимость золота (тыс. руб./гр), y_t	СС (k=3)	СС (k=5)
January-18	2,43		
February-18	2,44		
March-18	2,43		
April-18	2,61		
May-18	2,61		
June-18	2,59		
July-18	2,50		
August-18	2,55		
September-18	2,61		
October-18	2,57		
November-18	2,57		

2.2. Для показателя смертности постройте простую СС с длиной интервала сглаживания $k=12$, используя встроенные процедуры стат пакетов.

tssmooth ma -- Moving average with uniform weights

```
tssmooth ma [type] newvar = exp [if] [in], window(#l[#c[#f]])
[replace]
```

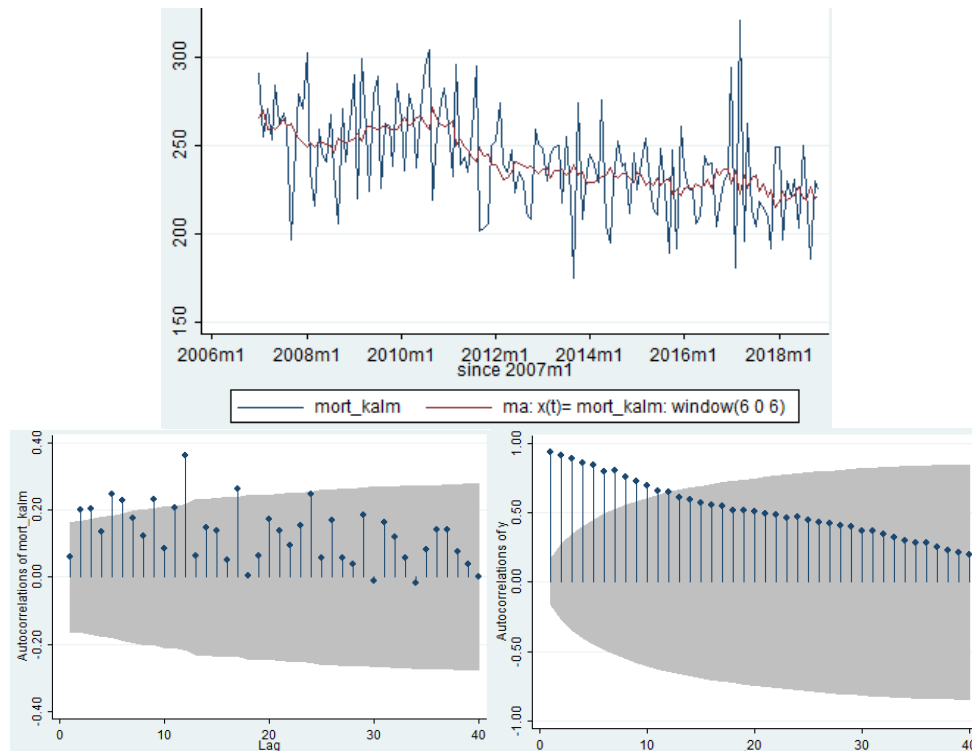
window(#l[#c[#f]]) describes the span of the uniformly weighted moving average.

#l specifies the number of lagged terms to be included, $0 \leq \#l \leq$ one-half the number of observations in the sample.

#c is optional and specifies whether to include the current observation in the filter. A 0 indicates exclusion and 1, inclusion. The current observation is excluded by default.

#f is optional and specifies the number of forward terms to be included, $0 \leq \#f \leq$ one-half the number of observations in the sample.

Постройте совмещенные графики: исходный и сглаженный ряд. Сравните поведение ACF исходного и сглаженного рядов.



2.3. Проведите (**runtest**) для сглаженных данных смертности. Как изменились Ваши выводы?

```
. runtest y
N(y <= 236.5833282470703) = 72
N(y > 236.5833282470703) = 71
obs = 143
N(runs) = 14
z = -9.82
Prob>|z| = 0
```

2.4. Выведите формулы для весовых коэффициентов СС в случае, если локальное поведение сглаженного ВР описывается параболой 2-го порядка, $k=5$. Определите значения для весовых коэффициентов в краевых точках.

Формулы:

Сглаживание в точке t'

$$f(t') = a_0 + a_1 t' + a_2 (t')^2$$

$$t' = t - (m + 1)$$

t	1	2	3	4	5
t'	-2	-1	0	1	2

МНК

$$Q = \sum_{t'=-2}^2 (y_{t'} - a_0 - a_1 t' - a_2 (t')^2)^2 \rightarrow \min$$

$$\frac{\partial Q}{\partial a_0} = 0 \rightarrow -2 \sum_{t'=-2}^2 (y_{t'} - a_0 - a_1 t' - a_2 (t')^2) = 0 \quad (1)$$

$$\frac{\partial Q}{\partial a_1} = 0 \rightarrow -2 \sum_{t'=-2}^2 (y_{t'} - a_0 - a_1 t' - a_2 (t')^2) \cdot t' = 0 \quad (2)$$

$$\frac{\partial Q}{\partial a_2} = 0 \rightarrow -2 \sum_{t'=-2}^2 (y_{t'} - a_0 - a_1 t' - a_2 (t')^2) \cdot (t')^2 = 0 \quad (3)$$

$$(1): \sum_{t'=-2}^2 y_{t'} - \sum_{t'=-2}^2 a_0 - \sum_{t'=-2}^2 a_1 t' - \sum_{t'=-2}^2 a_2 (t')^2 = 0$$

$$\sum_{t'=-2}^2 y_{t'} - \sum_{t'=-2}^2 a_0 - (-2a_1 - a_1 + a_1 + 2a_1) + a_2(4+1+0+1+4) = 0 \rightarrow$$

$$\begin{cases} \sum_{t'=-2}^2 y_{t'} = 5a_0 + 10a_2 & (1) \\ \sum_{t'=-2}^2 y_{t'} t' = 10a_1 & (2) \\ \sum_{t'=-2}^2 y_{t'} (t')^2 = 10a_0 + 34a_2 & (3) \end{cases}$$

$$a_0: (1) * 17 - (3) * 5$$

$$17 \sum_{t'=-2}^2 y_{t'} = 17 \cdot 5a_0 + 17 \cdot 10a_2$$

$$-5 \sum_{t'=-2}^2 y_{t'} (t')^2 = 5 \cdot 10a_0 + 5 \cdot 34a_2$$

$$\text{Выражаем } a_0 = \frac{1}{35} \left(17 \sum_{t'=-2}^2 y_{t'} - 5 \sum_{t'=-2}^2 y_{t'} (t')^2 \right) = \frac{1}{35} (17(y_{t'=-2} + y_{-1} + y_0 + y_1 + y_2) - 5(\dots)) =$$

$$\hat{f}(t'=0) = a_0 + a_1 t' + a_2 (t')^2 = a_0$$

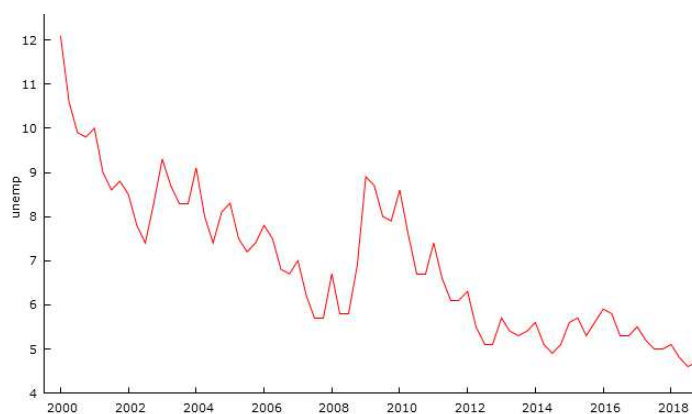
Весовые коэффициенты при сглаживании по полиномам второго и третьего порядка

Длина интервала Сглаживания	Весовые коэффициенты
5	$\frac{1}{35}[-3, +12, +17]$
7	$\frac{1}{21}[-2, +3, +6, +7]$
9	$\frac{1}{231}[-21, +14, +39, +54, +59]$
11	$\frac{1}{429}[-36, +9, +44, +69, +84, +89]$
13	$\frac{1}{143}[-11, 0, +9, +16, +21, +24, +25]$

Домашняя работа (ТДЗ) 3.

Рассмотрите ВР с сезонностью (безработица в России/свои данные). Требования к данным: Ежемесячные/ежеквартальные данные за 3-4 года.

Файл: unemp.dta




1. **Сглаживание.** Выделите трендовую составляющую с помощью простой СС/взвешенной СС, используя встроенные процедуры стат пакетов (Stata/Gretl/любой пакет). Какой интервал сглаживания необходимо взять для ежеквартальных данных? Постройте совмещенные графики: исходный и сглаженный ряд. Опишите основную тенденцию вашего показателя. Сравните поведение ACF исходного и сглаженного рядов (приведите графики ACF).
2. **Критерий серий.** Вручную реализуйте критерий серий, основанный на медиане выборки, для сглаженного ряда. Приведите подробные рассуждения и выводы. Какая нулевая гипотеза проверяется в критерии серий? Можно ли утверждать, что данные содержат тренд?
3. **runttest.** Для ряда используйте критерий случайности (Stata/Gretl/любой пакет): runttest (Stata). Какая нулевая гипотеза проверяется? Приложите скрины результатов тестирования, дайте интерпретацию и сделайте выводы. Сравните результаты с полученными ранее.

Выполненная домашняя работа загружается в LMS в формате pdf.

Срок выполнения – 1 неделя.

Сдача работы в группе по 2 человека (не забывайте указывать авторов).

edit	редактирование данных (открытие редактора данных)
clear	очистить память компьютера
display	Вывод на экран значения переменной или выражения
dis	калькулятор
list	Вывод на экран значений переменных из активного множества данных
Описательные статистики	
list [v1]	вывести значения переменных (v1) на экран, кнопка BREAK  (прервать выполнение команды)
describe [v1]	вывести описание переменных
sum [v1]	расчет дескриптивных статистик для переменной (v1)
Действия над переменными	
gen v2=g(v1)	создать новую переменную v2 как функцию g от v1
drop v1	удалить переменную v1
ren v1 v2	переименовать переменную v1 в v2
egen t=seq()	Создание последовательности целых чисел
Работа с временными рядами	
tsset t	Объявить переменную t переменной времени
tsline y	Построить график временного ряда y
help tsvarlist	Time-series varlists
regress y t	Построить линейную регрессию
predict y1, xb	Сохранить предсказанные значения в y1
predict e1, residuals	Сохранить значения остатков в e1
ac y	построить автокорреляционную функцию для y
pac y	построить частную автокорреляционную функцию для y
runttest y	Критерий случайности для y

tssmooth ma wn3 = wn, window(1 1 1)	Простая СС по 3 точкам
corrgram y	Вычислить значения автокорреляционной функций для y
pergram y	Построение периодограммы для y
sktest e1	тест на нормальность для e1
wntestq e1	Статистика Льюинга-Бокса для e1
estat dwatson	Статистика Дарбина-Уотсона на наличие автокорреляции 1-го порядка
Число ПИ	_pi
dfuller y	Тест Дики-Фуллера для y