Семинар 3. Анализ неслучайной составляющей временного ряда

Достовалов М.Ю.

08.02.2024

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

Установим необходимые пакеты и подгрузим библиотеки

# install.packages("DescTools")  
# install.packages("TTR")  
  
library(stats)  
library(TTR)  
library(DescTools)  
library(haven)

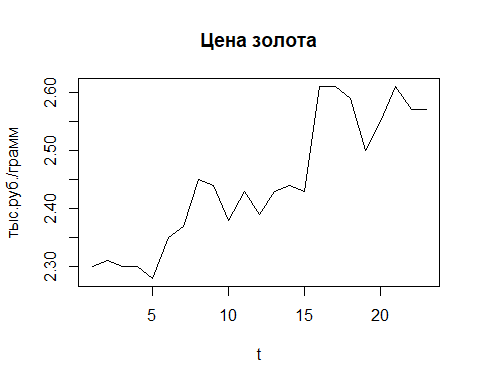
Задание 1. Изучается динамика цен на золото. Файл gold.dta.

Загрузим данные из файла

#file.choose()  
gold=read\_dta("gold.dta") # укажите свой путь, где лежит файл

Построим график временного ряда

plot.ts(gold$gold,  
 main="Цена золота",  
 ylab="тыс.руб./грамм",   
 xlab="t")



Посмотрим на характеристики ВР. Найдем медиану.

summary(gold$gold)

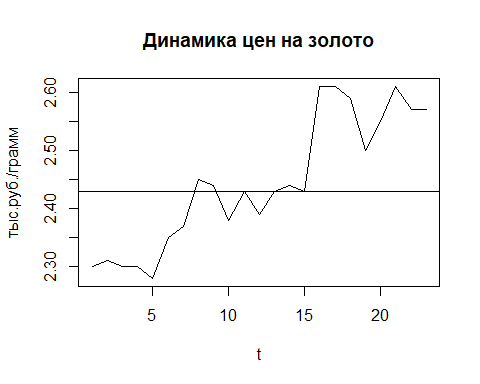
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.280 2.360 2.430 2.444 2.560 2.610

paste('Медиана: Me = ', round(median(gold$gold), 2))

## [1] "Медиана: Me = 2.43"

Добавим на график линию медианы ВР

plot.ts(gold$gold,  
 main="Динамика цен на золото",  
 ylab="тыс.руб./грамм",   
 xlab="t")  
abline(h=median(gold$gold))



Строим серии для критерия серий

gold$runs=ifelse(gold$gold<= 2.43, 0, 1)  
gold$runs

## [1] 0 0 0 0 0 0 0 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1

Приближенное правило: При отсутствии трендовой составляющей число серий должно быть достаточно велико, серии не должны быть слишком длинными.

trunc(1/2\*(23+2-1.96\*sqrt(23-1)))

## [1] 7

trunc(1.43\*log(23+1))

## [1] 4

Число серий должно быть r>7 и длина серий должна быть tau<4. Для нашего примера мы получили r=6, tau=11. Какой вывод можно сделать?

Проведем тест на случайность ВР (Н0: временной ряд имеет случайные наблюдения, тренд отсутствует)

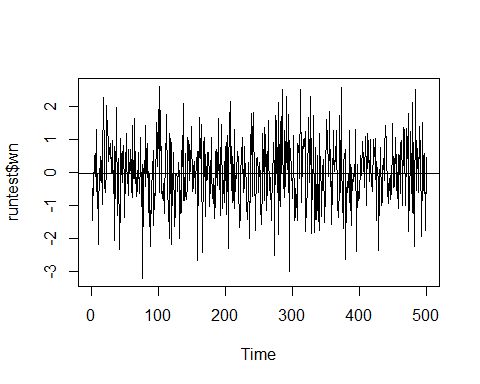
RunsTest(gold$gold)

##   
## Runs Test for Randomness  
##   
## data: gold$gold  
## runs = 6, m = 12, n = 11, p-value = 0.008617  
## alternative hypothesis: true number of runs is not equal the expected number  
## sample estimates:  
## median(x)   
## 2.43

Задание 2. Критерий случайности для БШ и тренда

Построим график БШ и проверим процесс на случайность.

runtest=read\_dta("runtest.dta")  
plot.ts(runtest$wn)  
abline(h=median(runtest$wn))



summary(runtest$wn)

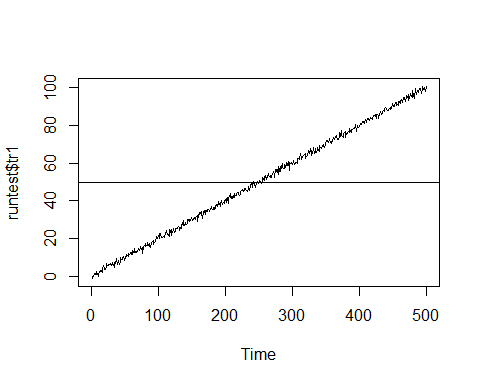
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -3.217909 -0.623221 -0.029982 0.005489 0.695767 2.606215

RunsTest(runtest$wn)

##   
## Runs Test for Randomness  
##   
## data: runtest$wn  
## z = -0.22383, runs = 248, m = 250, n = 250, p-value = 0.8229  
## alternative hypothesis: true number of runs is not equal the expected number  
## sample estimates:  
## median(x)   
## -0.02998218

Построим график тренда и проверим процесс на случайность.

plot.ts(runtest$tr1)  
abline(h=median(runtest$tr1))



summary(runtest$tr1)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.976 25.203 49.994 50.205 75.321 100.622

RunsTest(runtest$tr1)

##   
## Runs Test for Randomness  
##   
## data: runtest$tr1  
## z = -21.712, runs = 8, m = 250, n = 250, p-value < 2.2e-16  
## alternative hypothesis: true number of runs is not equal the expected number  
## sample estimates:  
## median(x)   
## 49.99431

1. Выделение трендовой составляющей. Скользящие средние.

SMA -Simple moving average.

EMA - Exponential moving average.

WMA - Weighted moving average.

Задание 3. Выделить трендовую компоненту в ценах на золото.

Найдем 3-членную скользящую среднюю.

sma3 = SMA(gold$gold,n=3)  
sma3

## [1] NA NA 2.303333 2.303333 2.293333 2.310000 2.333333 2.390000  
## [9] 2.420000 2.423333 2.416667 2.400000 2.416667 2.420000 2.433333 2.493333  
## [17] 2.550000 2.603333 2.566667 2.546667 2.553333 2.576667 2.583333

Найдем 5-членную скользящую среднюю.

sma5 = SMA(gold$gold,n=5)  
sma5

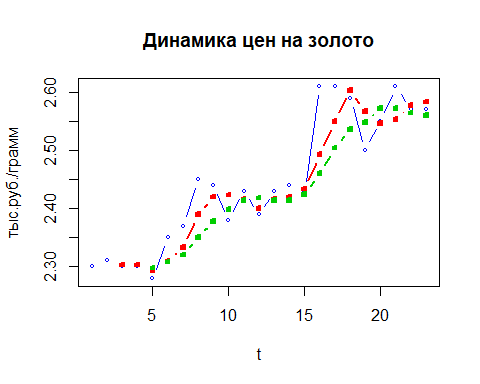
## [1] NA NA NA NA 2.298 2.308 2.320 2.350 2.378 2.398 2.414 2.418  
## [13] 2.414 2.414 2.424 2.460 2.504 2.536 2.548 2.572 2.572 2.564 2.560

Создадим новую таблицу со сглаженными значениями ВР

data\_ma= data.frame(gold$gold, sma3, sma5)  
View(data\_ma)

Построим графики исходного и сглаженных временных рядов

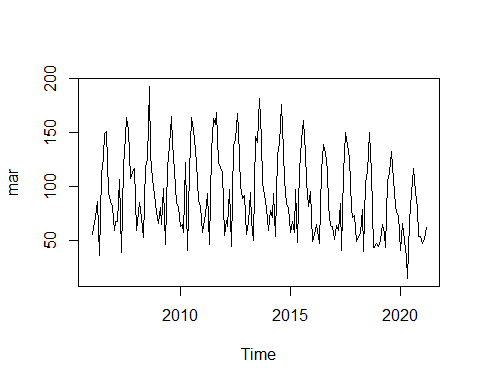
plot(gold$gold, col="blue", main="Динамика цен на золото", ylab="тыс.руб./грамм",   
 xlab="t", type="b", pch=21, cex=0.5)  
lines(sma3, col="red", lwd=2, type = "b", pch = 12, cex=0.5)  
lines(sma5, col="green3", lwd=2, type = "b", pch = 12, cex=0.5)



Задание 4. Для показателя числа браков постройте простую СС с длиной интервала сглаживания k=12 (файл marit.dta)

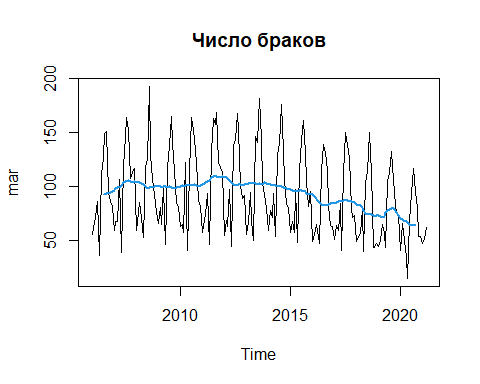
Загрузим данные из файла и построим график ВР

marit=read\_dta("marit.dta")  
mar <- ts(marit$mar, start=c(2006,1), frequency=12)   
plot(mar)



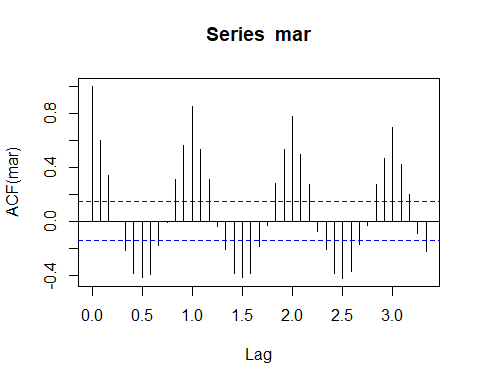
Рассчитаем 12-членную скользящую среднюю и построим графики

wgts=c( 0.5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5)/12  
wgts1 = c(.5, rep(1,11), .5)/12 # rep(1,11) повторить "1" 11 раз  
yma=stats::filter(mar, sides=2, filter=wgts) #sides=2 включать наблюдения до и после точки сглаживания   
plot(mar, main = "Число браков")  
lines(yma,lwd=2, col=4)

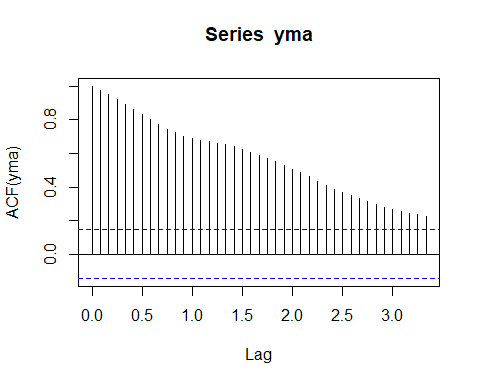


ACF/PACF числа браков

acf(mar,40, ylab="ACF(mar)")



acf(yma, 40, na.action = na.pass, ylab="ACF(yma)")



Проведем тест на случайность (отсутствие тренда)

RunsTest(mar)

##   
## Runs Test for Randomness  
##   
## data: mar  
## z = -5.7816, runs = 53, m = 92, n = 91, p-value = 7.399e-09  
## alternative hypothesis: true number of runs is not equal the expected number  
## sample estimates:  
## median(x)   
## 84

RunsTest(yma, na.rm = TRUE) # na.rm = TRUE -удаляем пропуски при выполнении теста

##   
## Runs Test for Randomness  
##   
## data: yma  
## z = -12.118, runs = 7, m = 86, n = 85, p-value < 2.2e-16  
## alternative hypothesis: true number of runs is not equal the expected number  
## sample estimates:  
## median(x)   
## 99.04167

Математические функции в R: <http://www.itmathrepetitor.ru/matematicheskie-funkcii-v-r/>