Семинар 4. Анализ трендовой компоненты временного ряда. Тест Чоу

Достовалов М.Ю.

08.02.2024

1. Анализ и моделирование трендовой компоненты. Типы кривых роста.

Установим необходимые пакеты и подгрузим библиотеки

# install.packages("forecast")  
# install.packages("dplyr")  
# install.packages("fBasics")  
# install.packages("broom")  
# install.packages("strucchange") # пакет для анализа структурных сдвигов  
  
library(strucchange)  
library(fBasics)  
library(ggplot2)  
library(dplyr)  
library(forecast)  
library(haven)

Задание 1. Постройте графики распространенности анемии среди детей. Опишите существующие тренды. Какие кривые роста (функции от времени) подходят для описания такого типа трендов? Файл hunger.dta.

Загрузим данные из файла

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

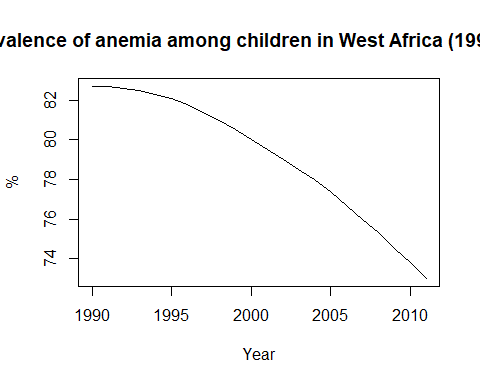
Создадим ts объект для отражения динамики анемии в Западной Африке

West <- ts(hunger$West, start=1990)  
West

## Time Series:  
## Start = 1990   
## End = 2011   
## Frequency = 1   
## [1] 82.7 82.7 82.6 82.5 82.3 82.1 81.8 81.4 81.0 80.5 80.0 79.5 79.0 78.5 78.0  
## [16] 77.4 76.7 76.0 75.3 74.5 73.8 73.0  
## attr(,"format.stata")  
## [1] %8.0g

Построим график динамики детской анемии в Западной Африке

plot.ts(West,  
 main="Prevalence of anemia among children in West Africa (1990-2011)",  
 ylab="%",   
 xlab="Year")



Оценим кривые роста и проверим модели на адекватность

1.Построим линейный тренд и оценим адекватность модели

hunger<-hunger%>% mutate(t = 1:22)  
hunger<-hunger%>% mutate(t2=t^2)  
model1=lm(West~t, data=hunger)  
summary(model1)

##   
## Call:  
## lm(formula = West ~ t, data = hunger)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.4344 -0.5112 0.2500 0.5557 0.6632   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 84.60909 0.29989 282.13 < 2e-16 \*\*\*  
## t -0.47470 0.02283 -20.79 5.15e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6795 on 20 degrees of freedom  
## Multiple R-squared: 0.9558, Adjusted R-squared: 0.9536   
## F-statistic: 432.2 on 1 and 20 DF, p-value: 5.149e-15

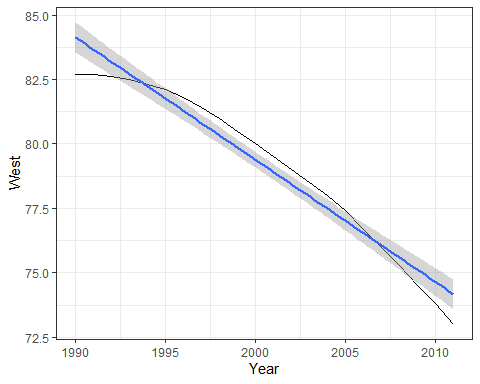
anova(model1)

## Analysis of Variance Table  
##   
## Response: West  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 199.542 199.542 432.22 5.149e-15 \*\*\*  
## Residuals 20 9.233 0.462   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Добавим на график линию тренда

ggplot(hunger, aes(x=Year, y=West)) +  
 geom\_line()+  
 stat\_smooth(method = "lm")+  
 theme\_bw()

## `geom\_smooth()` using formula = 'y ~ x'



Посчитаем информационные критерии Акайке(AIC) и Шварца(BIC)

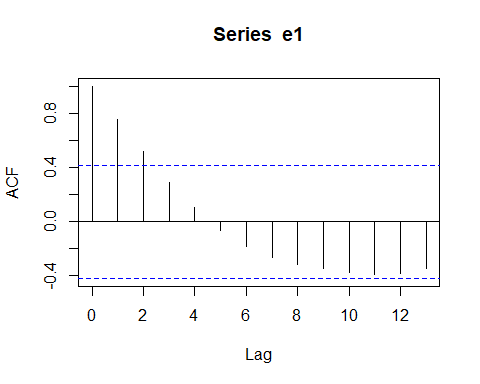
CV(model1) #Cross-validation statistic

## CV AIC AICc BIC AdjR2   
## 0.5473806 -13.1007951 -11.7674618 -9.8276678 0.9535622

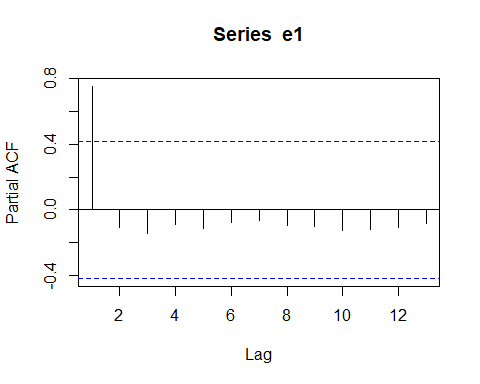
Проведем анализ остатков для модели линейного тренда

Проверим автокорреляцию остатков (ACF, PACF, Тест Льюинга-Бокса)

west1=fitted(model1)  
e1=residuals(model1)  
acf(e1)



pacf(e1)



Box.test(residuals(model1),lag=10, type="Lj") # Тест Льюинга-Бокса: Н0: ρ1=…= ρk=0 («автокорреляция до лага k отсутствует»)

##   
## Box-Ljung test  
##   
## data: residuals(model1)  
## X-squared = 42.723, df = 10, p-value = 5.573e-06

Проведем тесты на нормальность остатков (Н0: наблюдения имеют нормальное распределение)

jarqueberaTest(model1$resid) # Тест Жака-Бера

##   
## Title:  
## Jarque - Bera Normalality Test  
##   
## Test Results:  
## STATISTIC:  
## X-squared: 2.6477  
## P VALUE:  
## Asymptotic p Value: 0.2661

shapiro.test(model1$resid) # Шапиро-Уилка

##   
## Shapiro-Wilk normality test  
##   
## data: model1$resid  
## W = 0.86445, p-value = 0.006181

2.Построим параболический тренд и оценим адекватность модели

model2=lm(West~t+t2, data=hunger)  
summary(model2)

##   
## Call:  
## lm(formula = West ~ t + t2, data = hunger)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.182907 -0.060056 -0.008271 0.070677 0.160346   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 82.9642838 0.0676522 1226.336 < 2e-16 \*\*\*  
## t -0.0635019 0.0135503 -4.686 0.000161 \*\*\*  
## t2 -0.0178783 0.0005721 -31.251 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0963 on 19 degrees of freedom  
## Multiple R-squared: 0.9992, Adjusted R-squared: 0.9991   
## F-statistic: 1.125e+04 on 2 and 19 DF, p-value: < 2.2e-16

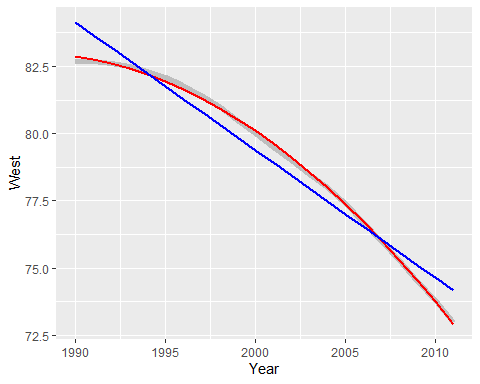
anova(model2)

## Analysis of Variance Table  
##   
## Response: West  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 199.542 199.542 21516.05 < 2.2e-16 \*\*\*  
## t2 1 9.057 9.057 976.61 < 2.2e-16 \*\*\*  
## Residuals 19 0.176 0.009   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Добавим на график линии тренда

west2=fitted(model2)  
ggplot(hunger) +  
 geom\_line(aes(Year, West), colour = 'grey', size = 2) +  
 geom\_line(aes(Year, west2), colour = 'red', size = 1)+  
 geom\_line(aes(Year, west1), colour = 'blue', size = 1)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



Посчитаем информационные критерии Акайке(AIC) и Шварца(BIC)

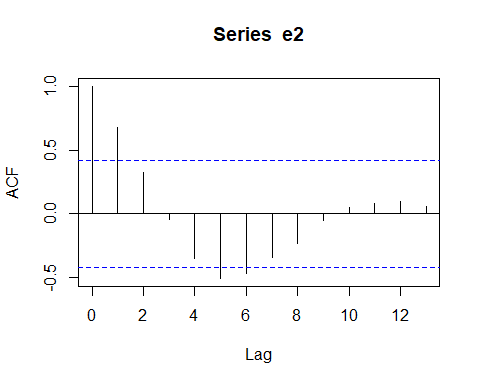
CV(model2) #Cross-validation statistic

## CV AIC AICc BIC AdjR2   
## 0.01195227 -98.19697691 -95.84403573 -93.83280709 0.99906715

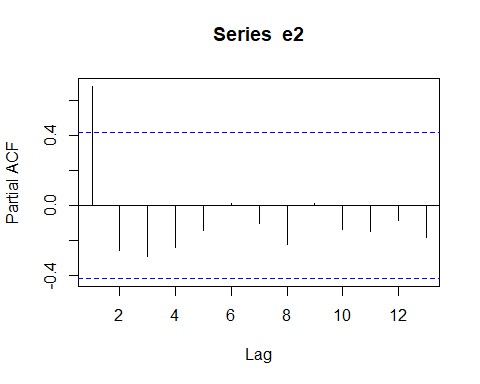
Проведем анализ остатков для модели линейного тренда

Проверим автокорреляцию остатков (ACF, PACF, Тест Льюинга-Бокса)

e2=residuals(model2)  
acf(e2)



pacf(e2)



Box.test(residuals(model2),lag=10, type="Lj") # Тест Льюинга-Бокса: Н0: ρ1=…= ρk=0 («автокорреляция до лага k отсутствует»)

##   
## Box-Ljung test  
##   
## data: residuals(model2)  
## X-squared = 39.589, df = 10, p-value = 2.002e-05

Проведем тесты на нормальность остатков (Н0: наблюдения имеют нормальное распределение)

jarqueberaTest(model2$resid) # Тест Жака-Бера

##   
## Title:  
## Jarque - Bera Normalality Test  
##   
## Test Results:  
## STATISTIC:  
## X-squared: 0.4428  
## P VALUE:  
## Asymptotic p Value: 0.8014

shapiro.test(model2$resid) # Шапиро-Уилка

##   
## Shapiro-Wilk normality test  
##   
## data: model2$resid  
## W = 0.98274, p-value = 0.9532

1. Тест Чоу на структурный сдвиг.

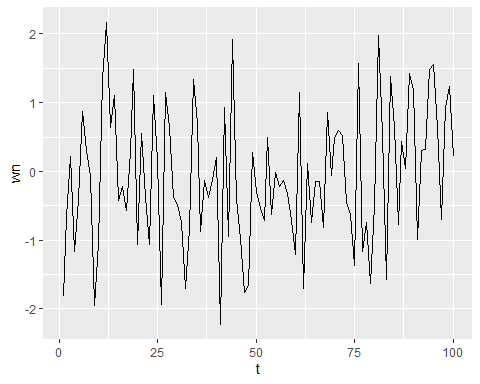
Задание 2. Для ВР со структурными изменениями из файла Chow.dta: tr, tr1, tr2, tr3, (предварительно изучив данные и определив точку структурного сдвига) используйте тест Чоу. Сделайте вывод.

Загружаем данные (chow.dta).

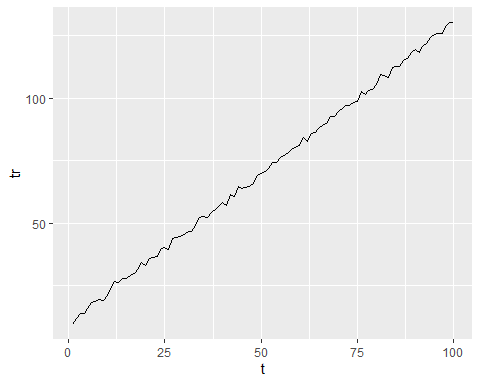
chow=read\_dta("chow.dta")

Графики ВР

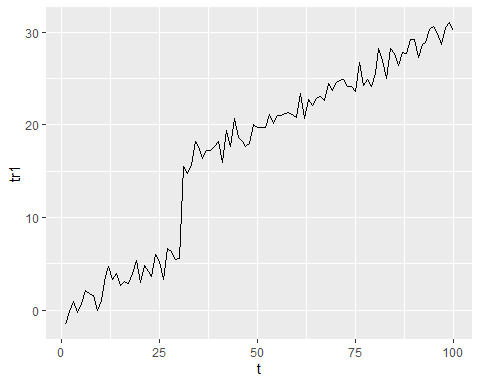
ggplot(chow, aes(x=t, y=wn)) + geom\_line()



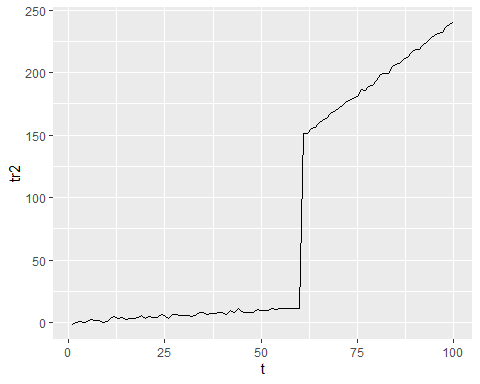
ggplot(chow, aes(x=t, y=tr)) + geom\_line()



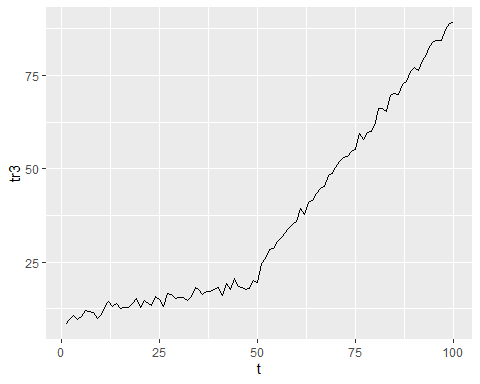
ggplot(chow, aes(x=t, y=tr1)) + geom\_line()



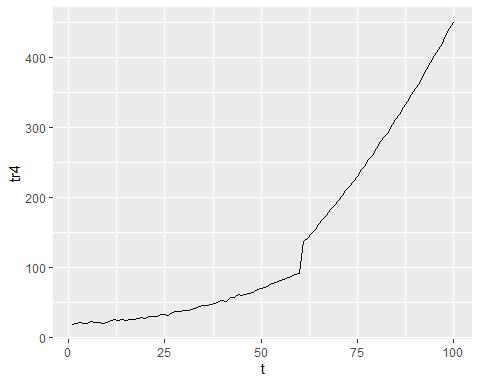
ggplot(chow, aes(x=t, y=tr2)) + geom\_line()



ggplot(chow, aes(x=t, y=tr3)) + geom\_line()



ggplot(chow, aes(x=t, y=tr4)) + geom\_line()



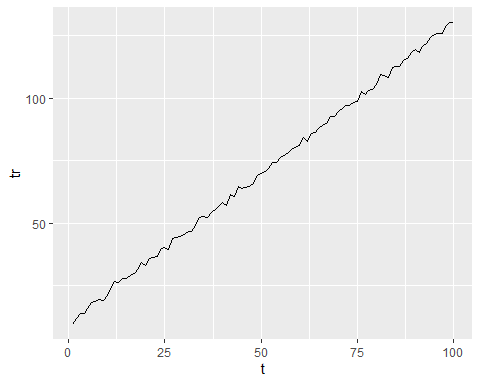
Проведем тест Чоу на структурный сдвиг для БШ (H0: структурного сдвига нет)

sctest(chow$wn~chow$t, type="Chow")

##   
## Chow test  
##   
## data: chow$wn ~ chow$t  
## F = 2.5029, p-value = 0.08718

График tr

ggplot(chow, aes(x=t, y=tr)) + geom\_line()



Проведем тест Чоу на структурный сдвиг для tr в точке t=20

sctest(chow$tr~chow$t, type="Chow", point=20) # Тест Чоу: point=20 - точка сдвига

##   
## Chow test  
##   
## data: chow$tr ~ chow$t  
## F = 2.1149, p-value = 0.1262

fs.tr <- Fstats(chow$tr~chow$t)  
plot(fs.tr)

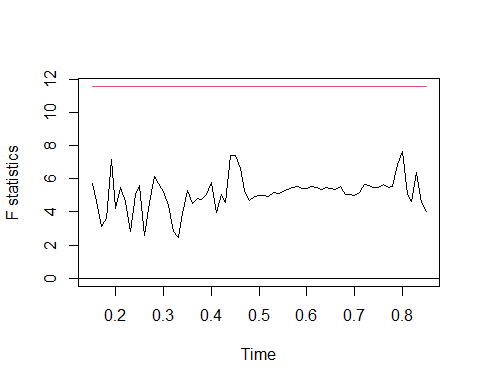
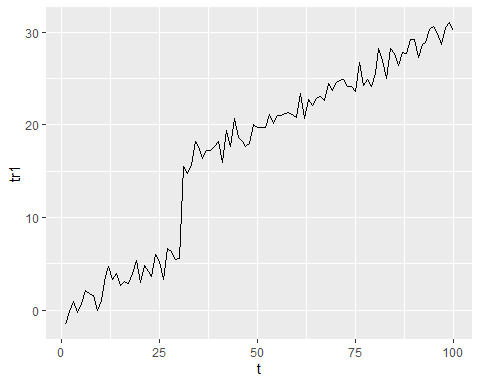
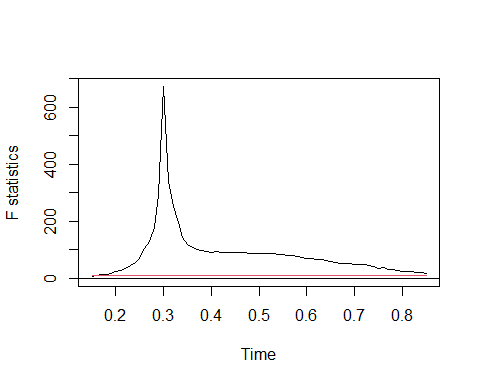


График tr1

tr1 <- ts(chow$tr1, start=1)  
t <- 1:100  
ggplot(chow, aes(x=t, y=tr1)) + geom\_line()



fs.tr1 <- Fstats(chow$tr1~chow$t)  
plot(fs.tr1)



Рассчитаем точку сдвига на основе теста Бай-Перрона

breakpoints(fs.tr1)

##   
## Optimal 2-segment partition:   
##   
## Call:  
## breakpoints.Fstats(obj = fs.tr1)  
##   
## Breakpoints at observation number:  
## 30   
##   
## Corresponding to breakdates:  
## 0.29

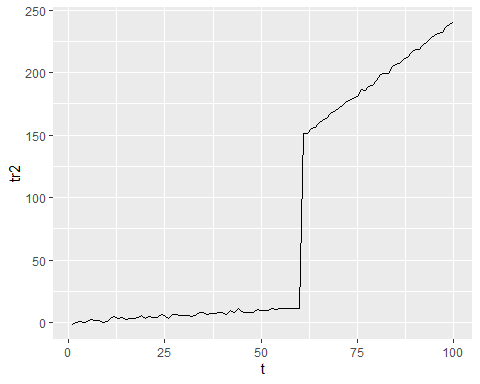
Проведем тест Чоу на структурный сдвиг для tr1 в точке t=30

sctest(chow$tr1~chow$t, type="Chow", point=30) # Тест Чоу: point=30 - точка сдвига

##   
## Chow test  
##   
## data: chow$tr1 ~ chow$t  
## F = 336.57, p-value < 2.2e-16

График tr2

ggplot(chow, aes(x=t, y=tr2)) + geom\_line()



Проведем анализ структурных сдвигов для tr2

bp.tr2 <- breakpoints(chow$tr2~chow$t)  
breakpoints(bp.tr2)

##   
## Optimal 2-segment partition:   
##   
## Call:  
## breakpoints.breakpointsfull(obj = bp.tr2)  
##   
## Breakpoints at observation number:  
## 60   
##   
## Corresponding to breakdates:  
## 0.6

sctest(chow$tr2~chow$t, type="Chow", point=60)

##   
## Chow test  
##   
## data: chow$tr2 ~ chow$t  
## F = 94746, p-value < 2.2e-16

fs.tr2 <- Fstats(chow$tr2~chow$t)  
plot(fs.tr2)

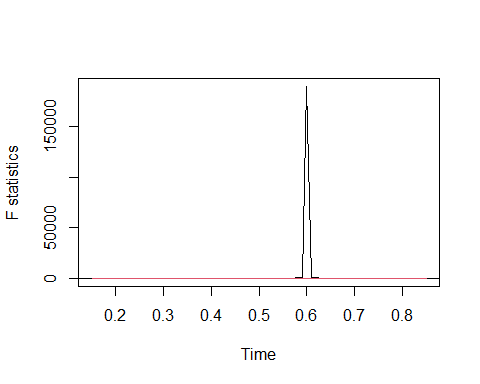
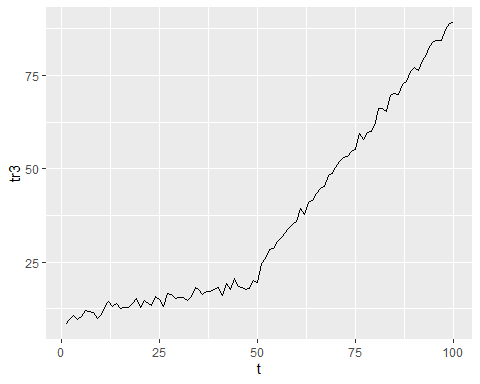


График tr3

ggplot(chow, aes(x=t, y=tr3)) + geom\_line()



Проведем анализ структурных сдвигов для tr3

bp.tr3 <- breakpoints(chow$tr3~chow$t)  
breakpoints(bp.tr3)

##   
## Optimal 2-segment partition:   
##   
## Call:  
## breakpoints.breakpointsfull(obj = bp.tr3)  
##   
## Breakpoints at observation number:  
## 50   
##   
## Corresponding to breakdates:  
## 0.5

sctest(chow$tr3~chow$t, type="Chow", point=50)

##   
## Chow test  
##   
## data: chow$tr3 ~ chow$t  
## F = 3531.9, p-value < 2.2e-16

fs.tr3 <- Fstats(chow$tr3~chow$t)  
plot(fs.tr3)

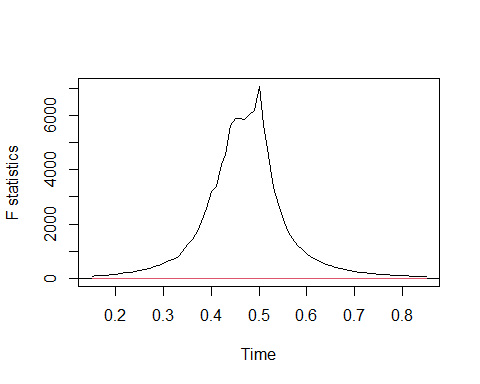
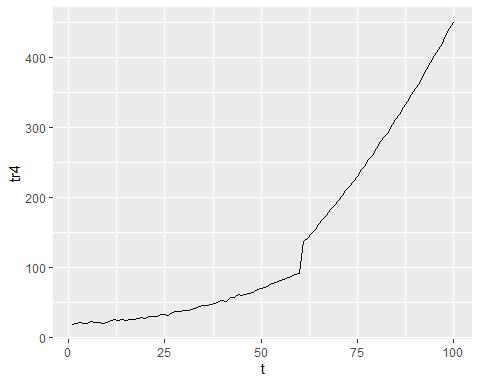


График tr4

ggplot(chow, aes(x=t, y=tr4)) + geom\_line()



Проведем анализ структурных сдвигов для tr4

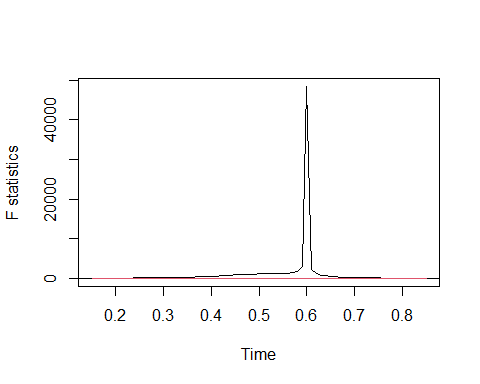
chow<-chow%>% mutate(t2=t^2)  
bp.tr4 <- breakpoints(chow$tr4~chow$t2)  
breakpoints(bp.tr4)

##   
## Optimal 2-segment partition:   
##   
## Call:  
## breakpoints.breakpointsfull(obj = bp.tr4)  
##   
## Breakpoints at observation number:  
## 60   
##   
## Corresponding to breakdates:  
## 0.6

sctest(chow$tr4~chow$t2, type="Chow", point=60)

##   
## Chow test  
##   
## data: chow$tr4 ~ chow$t2  
## F = 24211, p-value < 2.2e-16

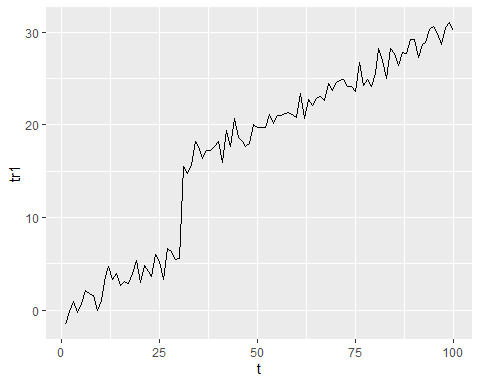
fs.tr4 <- Fstats(chow$tr4~chow$t2)  
plot(fs.tr4)



Моделирование структурных сдвигов

График tr1

ggplot(chow, aes(x=t, y=tr1)) + geom\_line()



Построим и проведем анализ модели без фиктивной переменной

model1=lm(tr1~t, data=chow) # без фиктивной переменной  
summary(model1)

##   
## Call:  
## lm(formula = tr1 ~ t, data = chow)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.6052 -2.2047 -0.4836 1.9914 6.5371   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.387816 0.551749 0.703 0.484   
## t 0.329994 0.009485 34.789 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.738 on 98 degrees of freedom  
## Multiple R-squared: 0.9251, Adjusted R-squared: 0.9243   
## F-statistic: 1210 on 1 and 98 DF, p-value: < 2.2e-16

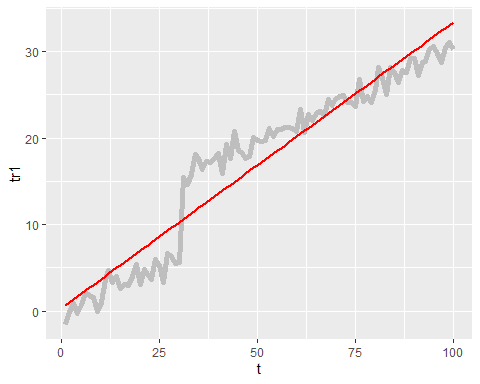
anova(model1)

## Analysis of Variance Table  
##   
## Response: tr1  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 9073.8 9073.8 1210.3 < 2.2e-16 \*\*\*  
## Residuals 98 734.7 7.5   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

y1=fitted(model1)  
e1=residuals(model1)

Построим график линии тренда

ggplot(chow) +  
 geom\_line(aes(t, tr1), colour = 'grey', size = 2) +  
 geom\_line(aes(t, y1), colour = 'red', size = 1)



Создадим фиктивную переменную для моделирования “скачка” (изменение констант)

chow$t30 <- ifelse(chow$t<= 30, 1, 0)

Построим и проведем анализ модели с фиктивной переменной

model2=lm(tr1~t+t30, data=chow) # с фиктивной переменной  
summary(model2)

##   
## Call:  
## lm(formula = tr1 ~ t + t30, data = chow)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.04552 -0.63565 -0.01412 0.60336 2.33713   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.906650 0.380903 23.38 <2e-16 \*\*\*  
## t 0.215351 0.005538 38.89 <2e-16 \*\*\*  
## t30 -9.097784 0.348846 -26.08 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9723 on 97 degrees of freedom  
## Multiple R-squared: 0.9907, Adjusted R-squared: 0.9905   
## F-statistic: 5139 on 2 and 97 DF, p-value: < 2.2e-16

anova(model2)

## Analysis of Variance Table  
##   
## Response: tr1  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 9073.8 9073.8 9597.81 < 2.2e-16 \*\*\*  
## t30 1 643.0 643.0 680.15 < 2.2e-16 \*\*\*  
## Residuals 97 91.7 0.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

y2=fitted(model2)  
e2=residuals(model2)

Построим график линии тренда с учетом сдвига

ggplot(chow) +  
 geom\_line(aes(t, tr1), colour = 'grey', size = 2) +  
 geom\_line(aes(t, y2), colour = 'red', size = 1)

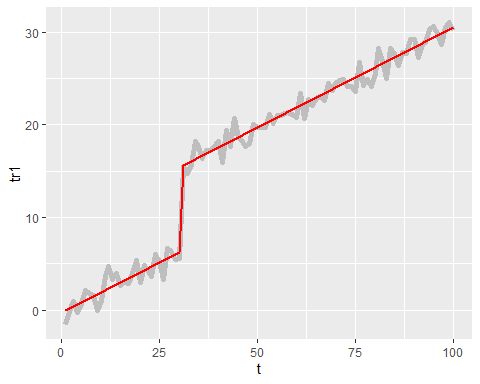
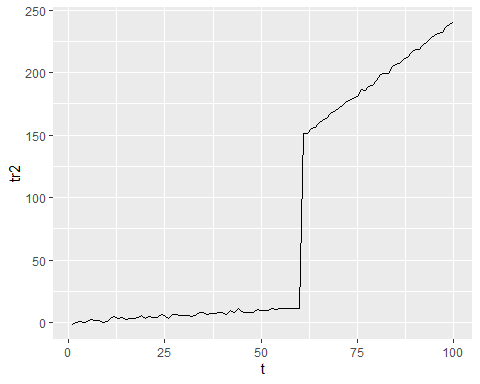


График tr2

ggplot(chow, aes(x=t, y=tr2)) + geom\_line()



Создадим фиктивную переменную для моделирования “скачка” и изменения угла наклона

chow$t60 <- ifelse(chow$t<= 60, 1, 0)  
chow<-chow%>% mutate(tt60=t60\*t)

Построим и проведем анализ модели с учетом “скачка” и изменения угла наклона

model3=lm(tr2~t+tt60, data=chow)  
summary(model3)

##   
## Call:  
## lm(formula = tr2 ~ t + tt60, data = chow)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.6196 -0.7300 -0.0719 0.7393 3.4598   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.514789 0.306287 1.681 0.096 .   
## t 2.417641 0.004401 549.289 <2e-16 \*\*\*  
## tt60 -2.235144 0.006327 -353.258 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.203 on 97 degrees of freedom  
## Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998   
## F-statistic: 3.072e+05 on 2 and 97 DF, p-value: < 2.2e-16

anova(model3)

## Analysis of Variance Table  
##   
## Response: tr2  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 708862 708862 489581 < 2.2e-16 \*\*\*  
## tt60 1 180685 180685 124791 < 2.2e-16 \*\*\*  
## Residuals 97 140 1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

y3=fitted(model3)  
e3=residuals(model3)

Построим график линии тренда с учетом структурных сдвигов

ggplot(chow) +  
 geom\_line(aes(t, tr2), colour = 'grey', size = 2) +  
 geom\_line(aes(t, y3), colour = 'red', size = 1)

