ARIMA-модели

Достовалов М.Ю.

21.03.2024

Установим необходимые пакеты и подгрузим библиотеки

# install.packages('urca')  
# install.packages('aTSA')  
# install.packages('forecast')  
# install.packages('zoo')  
# install.packages('xts')  
# install.packages('pander')  
  
library(aTSA)  
library(urca)  
library(haven)  
library(stats)  
library(tseries)  
library(forecast)  
library(zoo) # временные ряды  
library(xts) # временные ряды  
library(pander) # красивые таблицы  
library(lmtest) # тесты  
library(car) # тесты  
library(ggplot2)

Задание 2. Анализ и моделирование демографических данных (Методология Бокса-Дженкинса) Исходные данные (Данные с сайта <http://www.gks.ru>): 1. Коэффициент суммарной рождаемости в России (1990-2014 г.г.) 2. Ожидаемая продолжительность жизни при рождении в Москве (мужчины, женщины). 3. Ожидаемая продолжительность жизни при рождении в России (мужчины, женщины). Файл: TFR\_LE.dta

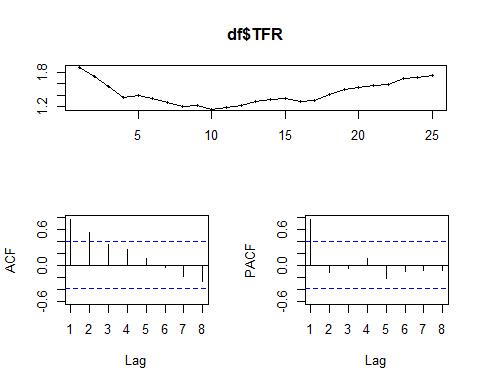
Загрузим данные из файла

#file.choose()  
df <- read\_dta('TFR\_LE.dta') # укажите свой путь, где лежит файл

Идентификация модели

Графический анализ ВР

tsdisplay(df$TFR)



Тест Дики-Фуллера для исходного ряда

summary(ur.df(df$TFR, type = c("trend")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.083281 -0.021354 0.002079 0.027502 0.099587   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.200640 0.081574 2.460 0.023667 \*   
## z.lag.1 -0.243933 0.066435 -3.672 0.001621 \*\*   
## tt 0.010858 0.002481 4.376 0.000325 \*\*\*  
## z.diff.lag -0.095211 0.186355 -0.511 0.615297   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04635 on 19 degrees of freedom  
## Multiple R-squared: 0.6908, Adjusted R-squared: 0.6419   
## F-statistic: 14.15 on 3 and 19 DF, p-value: 4.41e-05  
##   
##   
## Value of test-statistic is: -3.6718 7.3755 10.9256   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.38 -3.60 -3.24  
## phi2 8.21 5.68 4.67  
## phi3 10.61 7.24 5.91

summary(ur.df(df$TFR, type = c("drift")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.101817 -0.046837 -0.001844 0.027758 0.123131   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.13604 0.11080 1.228 0.23377   
## z.lag.1 -0.09347 0.07851 -1.191 0.24775   
## z.diff.lag 0.53713 0.16251 3.305 0.00353 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06402 on 20 degrees of freedom  
## Multiple R-squared: 0.3791, Adjusted R-squared: 0.317   
## F-statistic: 6.106 on 2 and 20 DF, p-value: 0.008514  
##   
##   
## Value of test-statistic is: -1.1906 0.7809   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.75 -3.00 -2.63  
## phi1 7.88 5.18 4.12

Тест KPSS (Квятковского-Филлипса-Шмидта-Шина) для исходного ряда

stationary.test(df$TFR, method = "kpss")

## KPSS Unit Root Test   
## alternative: nonstationary   
##   
## Type 1: no drift no trend   
## lag stat p.value  
## 1 0.731 0.1  
## -----   
## Type 2: with drift no trend   
## lag stat p.value  
## 1 1.04 0.01  
## -----   
## Type 1: with drift and trend   
## lag stat p.value  
## 1 0.0368 0.1  
## -----------   
## Note: p.value = 0.01 means p.value <= 0.01   
## : p.value = 0.10 means p.value >= 0.10

Вывод: После добавления тренда процесс на уровне значимости 5% становится стационарным. Значит можно предположить, что это тренд-стационарный процесс

Тест Дики-Фуллера для первой разности

d\_TFR = diff(df$TFR)  
summary(ur.df(d\_TFR, type = c("drift")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.117934 -0.033402 -0.004596 0.039004 0.108317   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.01102 0.01325 0.832 0.4159   
## z.lag.1 -0.52024 0.18518 -2.809 0.0112 \*  
## z.diff.lag -0.18163 0.19197 -0.946 0.3560   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06167 on 19 degrees of freedom  
## Multiple R-squared: 0.4016, Adjusted R-squared: 0.3386   
## F-statistic: 6.375 on 2 and 19 DF, p-value: 0.007615  
##   
##   
## Value of test-statistic is: -2.8093 4.4448   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.75 -3.00 -2.63  
## phi1 7.88 5.18 4.12

PP-тест Филлипса-Перрона для первой разности

stationary.test(d\_TFR, method = "pp")

## Phillips-Perron Unit Root Test   
## alternative: stationary   
##   
## Type 1: no drift no trend   
## lag Z\_rho p.value  
## 2 -8.67 0.0328  
## -----   
## Type 2: with drift no trend   
## lag Z\_rho p.value  
## 2 -8.28 0.228  
## -----   
## Type 3: with drift and trend   
## lag Z\_rho p.value  
## 2 -16.5 0.0794  
## ---------------   
## Note: p-value = 0.01 means p.value <= 0.01

Вывод: После взятия первой разности процесс не стал стационарным.

Тест Дики-Фуллера для второй разности

d2\_TFR = diff(d\_TFR)  
#summary(ur.df(d2\_TFR, type = c("drift")))  
adf.test(d2\_TFR)

## Warning in adf.test(d2\_TFR): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: d2\_TFR  
## Dickey-Fuller = -5.7106, Lag order = 2, p-value = 0.01  
## alternative hypothesis: stationary

Вывод: После взятия второй разности процес стал стационарным. Значит это разностно-стационарный процесс с порядком интегрируемости I(2)

Оценивание и диагностика моделей

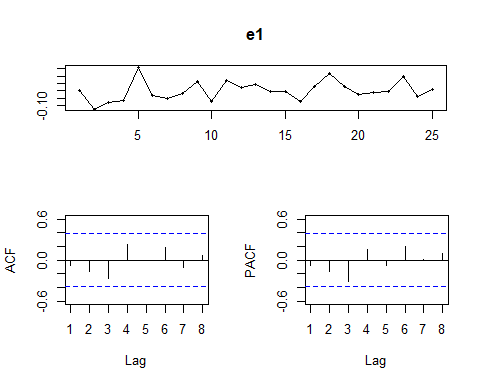
ARIMA(2,1,0)

m1 <- Arima(df$TFR, order=c(2,1,0))  
summary(m1)

## Series: df$TFR   
## ARIMA(2,1,0)   
##   
## Coefficients:  
## ar1 ar2  
## 0.5162 0.1862  
## s.e. 0.1966 0.2187  
##   
## sigma^2 = 0.004651: log likelihood = 31.15  
## AIC=-56.29 AICc=-55.09 BIC=-52.76  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.002027987 0.06397653 0.04915885 0.2017137 3.504562 0.7751724  
## ACF1  
## Training set -0.07193065

Тестирование остатков модели ARIMA(2,1,0)

e1 <- m1$residuals  
tsdisplay(e1)



pander(shapiro.test(e1)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e1

| Test statistic | P value |
| --- | --- |
| 0.9758 | 0.7914 |

pander(Box.test(e1,lag=10, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

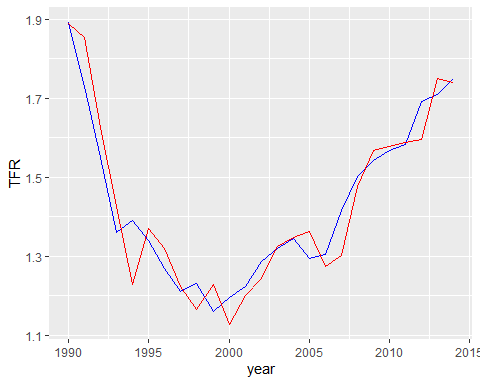
Box-Ljung test: e1

| Test statistic | df | P value |
| --- | --- | --- |
| 7.064 | 10 | 0.7194 |

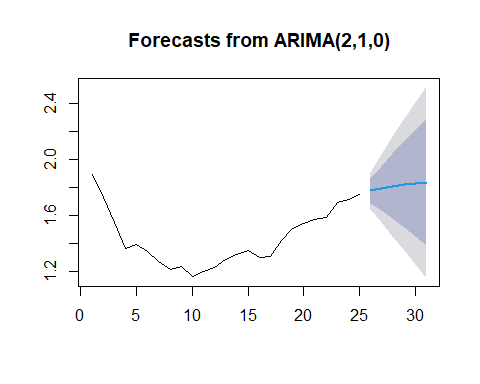
Прогнозирование по модели ARIMA(2,1,0)

pred1=fitted(m1)  
ggplot(data=df, aes(year, TFR))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred1), colour = 'red', size = 0.5)

## 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.



prognoz1 <- forecast(m1, h=6)  
plot(prognoz1)



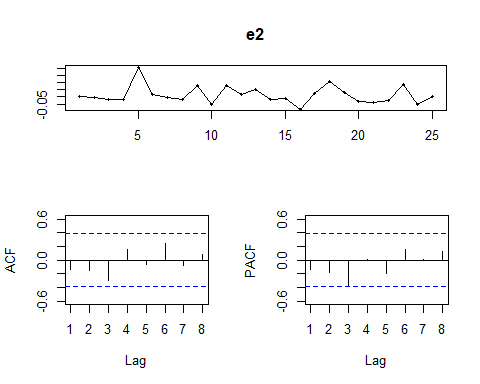
ARIMA(2,2,0)

m2 <- Arima(df$TFR, order=c(2,2,0))  
summary(m2)

## Series: df$TFR   
## ARIMA(2,2,0)   
##   
## Coefficients:  
## ar1 ar2  
## -0.4395 -0.2233  
## s.e. 0.1998 0.2006  
##   
## sigma^2 = 0.004788: log likelihood = 29.72  
## AIC=-53.44 AICc=-52.18 BIC=-50.04  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.01373992 0.06341718 0.04403039 1.029874 3.199885 0.6943032  
## ACF1  
## Training set -0.1347162

Тестирование остатков модели ARIMA(2,2,0)

e2 <- m2$residuals  
tsdisplay(e2)



pander(shapiro.test(e2)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e2

| Test statistic | P value |
| --- | --- |
| 0.8973 | 0.01604 \* |

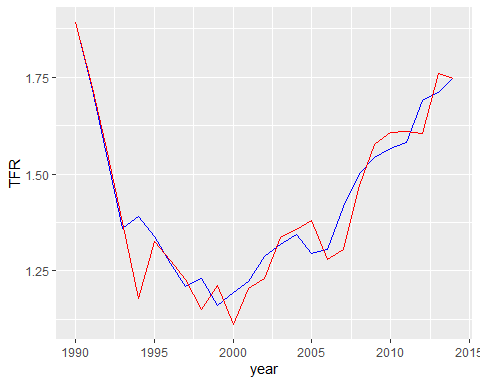
pander(Box.test(e2,lag=10, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e2

| Test statistic | df | P value |
| --- | --- | --- |
| 7.759 | 10 | 0.6524 |

Прогнозирование по модели ARIMA(2,2,0)

pred2=fitted(m2)  
ggplot(data=df, aes(year, TFR))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred2), colour = 'red', size = 0.5)



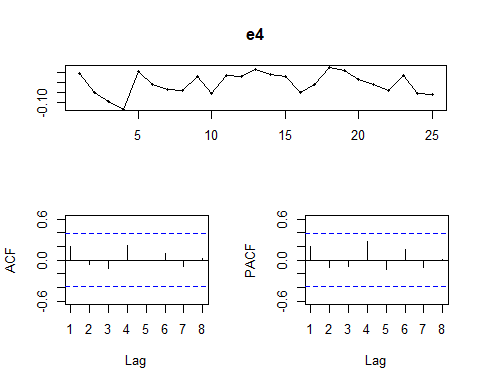
ARIMA(1,0,0)+параболический тренд

t2=df$t^2  
m4 <- Arima(df$TFR, order=c(1,0,0),include.drift=TRUE, xreg=t2)  
summary(m4)

## Series: df$TFR   
## Regression with ARIMA(1,0,0) errors   
##   
## Coefficients:  
## ar1 intercept drift xreg  
## 0.7799 1.9193 -0.1070 0.0041  
## s.e. 0.1439 0.1050 0.0172 0.0006  
##   
## sigma^2 = 0.003372: log likelihood = 37.39  
## AIC=-64.78 AICc=-61.62 BIC=-58.69  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.003488029 0.05322254 0.04629974 -0.2061366 3.262276 0.7300878  
## ACF1  
## Training set 0.1990827

Тестирование остатков модели ARIMA(1,0,0)+параболический тренд

e4 <- m4$residuals  
tsdisplay(e4)



pander(shapiro.test(e4)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e4

| Test statistic | P value |
| --- | --- |
| 0.9451 | 0.1935 |

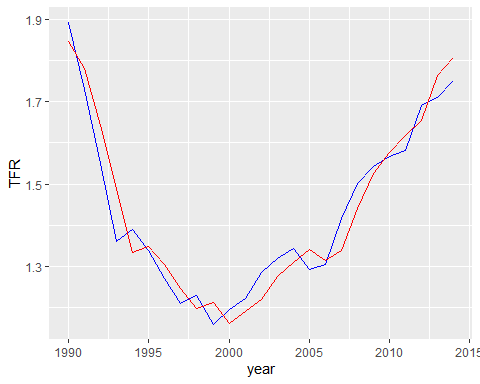
pander(Box.test(e4,lag=10, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e4

| Test statistic | df | P value |
| --- | --- | --- |
| 6.594 | 10 | 0.7631 |

Прогнозирование по модели ARIMA(1,0,0)+параболический тренд

pred4=fitted(m4)  
ggplot(data=df, aes(year, TFR))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred4), colour = 'red', size = 0.5)

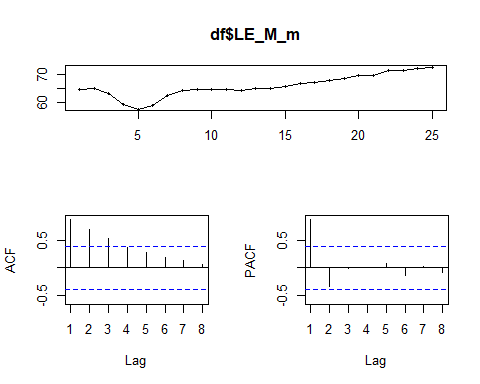


Задание 3. Устранение автокорреляции в остатках. Ожидаемая продолжительность жизни при рождении в Москве (мужчины)

Идентификация модели

Графический анализ ВР

tsdisplay(df$LE\_M\_m)



Тест Дики-Фуллера для исходного ряда

summary(ur.df(df$LE\_M\_m, type = c("trend")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3772 -0.3513 -0.0559 0.2619 1.2682   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 32.85669 4.70165 6.988 1.17e-06 \*\*\*  
## z.lag.1 -0.55948 0.07929 -7.056 1.03e-06 \*\*\*  
## tt 0.31135 0.04504 6.913 1.36e-06 \*\*\*  
## z.diff.lag 0.62824 0.10616 5.918 1.07e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6514 on 19 degrees of freedom  
## Multiple R-squared: 0.8144, Adjusted R-squared: 0.7851   
## F-statistic: 27.8 on 3 and 19 DF, p-value: 3.702e-07  
##   
##   
## Value of test-statistic is: -7.0559 17.8065 26.0391   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.38 -3.60 -3.24  
## phi2 8.21 5.68 4.67  
## phi3 10.61 7.24 5.91

summary(ur.df(df$LE\_M\_m, type = c("drift")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3686 -0.4019 0.1092 0.5272 1.9106   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.32885 4.56869 1.166 0.25718   
## z.lag.1 -0.07903 0.06977 -1.133 0.27071   
## z.diff.lag 0.63327 0.19399 3.264 0.00388 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.19 on 20 degrees of freedom  
## Multiple R-squared: 0.3477, Adjusted R-squared: 0.2825   
## F-statistic: 5.33 on 2 and 20 DF, p-value: 0.01395  
##   
##   
## Value of test-statistic is: -1.1328 0.8424   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.75 -3.00 -2.63  
## phi1 7.88 5.18 4.12

Вывод: После добавления тренда процесс на уровне значимости 5% становится стационарным. Значит можно предположить, что это тренд-стационарный процесс

Тест Дики-Фуллера для первой разности

d\_LE\_M\_m = diff(df$LE\_M\_m)  
summary(ur.df(d\_LE\_M\_m, type = c("drift")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.00934 -0.36527 0.05594 0.51030 1.61655   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3068 0.2201 1.394 0.17953   
## z.lag.1 -0.6891 0.1721 -4.004 0.00076 \*\*\*  
## z.diff.lag 0.5194 0.1821 2.852 0.01020 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9989 on 19 degrees of freedom  
## Multiple R-squared: 0.4752, Adjusted R-squared: 0.42   
## F-statistic: 8.603 on 2 and 19 DF, p-value: 0.002186  
##   
##   
## Value of test-statistic is: -4.0038 8.0933   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.75 -3.00 -2.63  
## phi1 7.88 5.18 4.12

Оценивание и диагностика моделей

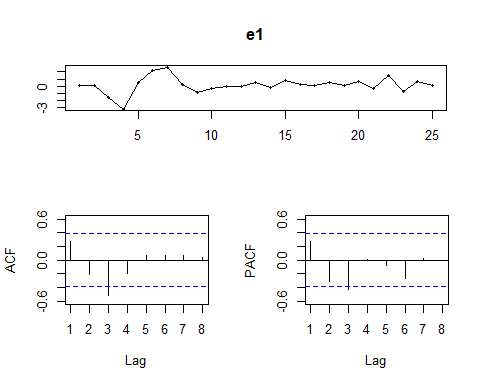
ARIMA(1,1,0)

m1 <- Arima(df$LE\_M\_m, order=c(1,1,0))  
summary(m1)

## Series: df$LE\_M\_m   
## ARIMA(1,1,0)   
##   
## Coefficients:  
## ar1  
## 0.5568  
## s.e. 0.1610  
##   
## sigma^2 = 1.337: log likelihood = -37.22  
## AIC=78.43 AICc=79 BIC=80.79  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 0.1534417 1.10924 0.726981 0.2273566 1.140315 0.7595796 0.2798533

Тестирование остатков модели ARIMA(1,1,0)

e1 <- m1$residuals  
tsdisplay(e1)



pander(shapiro.test(e1)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e1

| Test statistic | P value |
| --- | --- |
| 0.9039 | 0.02239 \* |

pander(Box.test(e1,lag=1, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e1

| Test statistic | df | P value |
| --- | --- | --- |
| 2.203 | 1 | 0.1378 |

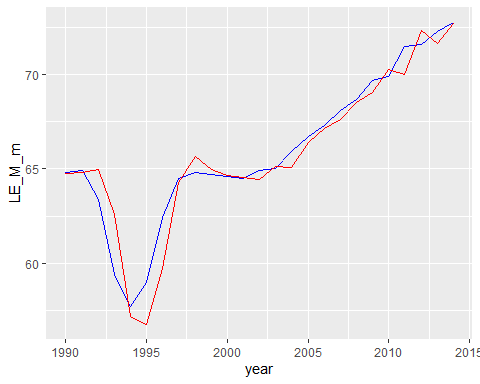
pander(Box.test(e1,lag=3, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e1

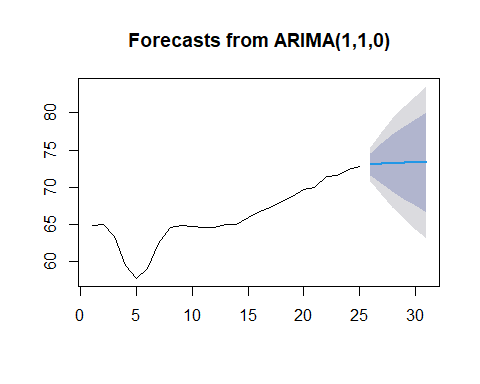
| Test statistic | df | P value |
| --- | --- | --- |
| 11.67 | 3 | 0.008619 \* \* |

Прогнозирование по модели ARIMA(1,1,0)

pred1=fitted(m1)  
ggplot(data=df, aes(year, LE\_M\_m))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred1), colour = 'red', size = 0.5)



prognoz1 <- forecast(m1, h=6)  
plot(prognoz1)



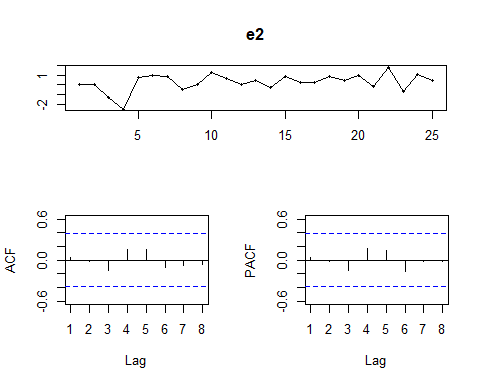
ARIMA(3,1,0)

m2 <- Arima(df$LE\_M\_m, order=c(3,1,0))  
summary(m2)

## Series: df$LE\_M\_m   
## ARIMA(3,1,0)   
##   
## Coefficients:  
## ar1 ar2 ar3  
## 0.7156 -0.2617 -0.2986  
## s.e. 0.1957 0.2430 0.2277  
##   
## sigma^2 = 0.9984: log likelihood = -33.09  
## AIC=74.18 AICc=76.28 BIC=78.89  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.2833425 0.9158043 0.7067372 0.4089244 1.085636 0.7384281  
## ACF1  
## Training set 0.03500647

Тестирование остатков модели ARIMA(3,1,0)

e2 <- m2$residuals  
tsdisplay(e2)



pander(shapiro.test(e2)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e2

| Test statistic | P value |
| --- | --- |
| 0.9114 | 0.03273 \* |

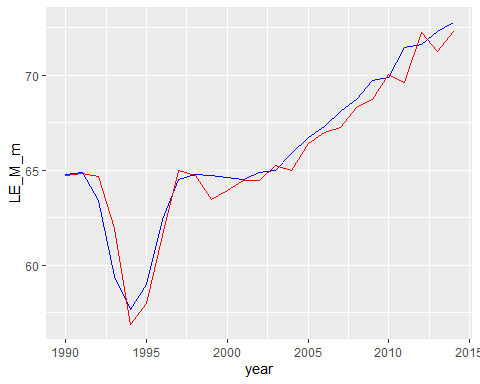
pander(Box.test(e2,lag=10, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e2

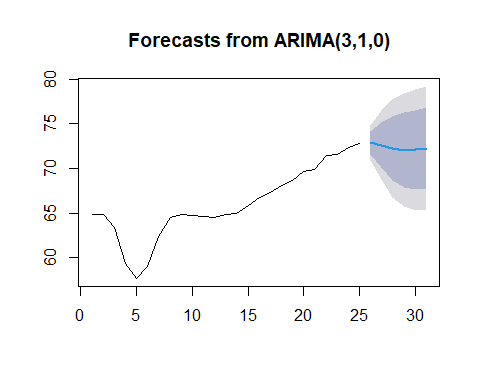
| Test statistic | df | P value |
| --- | --- | --- |
| 3.542 | 10 | 0.9656 |

Прогнозирование по модели ARIMA(3,1,0)

pred2=fitted(m2)  
ggplot(data=df, aes(year, LE\_M\_m))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred2), colour = 'red', size = 0.5)



prognoz2 <- forecast(m2, h=6)  
plot(prognoz2)



Задача 4. Оценивание ARIMA со структурным сдвигом Исходные данные: ОПЖ в США в период с 1900 по 1999 Файл с данными: LE\_USA.gdt Замечание. Необходимо учитывать структурный сдвиг 1918 года, связанный с эпидемией «Испанки».

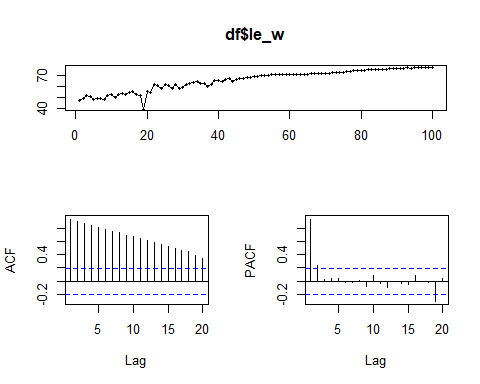
Загрузим данные из файла

#file.choose()  
df <- read\_dta('LE\_USA.dta') # укажите свой путь, где лежит файл

Идентификация модели

Графический анализ ВР

tsdisplay(df$le\_w)



Проведем тест Эндрюса-Зивота с учетом структурного сдвига 1-го типа (на константу)

summary(ur.za(df$le\_w, model = c("intercept")))

##   
## ################################   
## # Zivot-Andrews Unit Root Test #   
## ################################   
##   
##   
## Call:  
## lm(formula = testmat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.5751 -0.4467 0.1699 0.9128 3.1112   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.49948 3.82242 8.241 9.34e-13 \*\*\*  
## y.l1 0.34641 0.07824 4.427 2.55e-05 \*\*\*  
## trend 0.15064 0.02248 6.700 1.47e-09 \*\*\*  
## du 4.85718 0.74427 6.526 3.28e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.883 on 95 degrees of freedom  
## (1 пропущенное наблюдение удалено)  
## Multiple R-squared: 0.9579, Adjusted R-squared: 0.9565   
## F-statistic: 719.7 on 3 and 95 DF, p-value: < 2.2e-16  
##   
##   
## Teststatistic: -8.3532   
## Critical values: 0.01= -5.34 0.05= -4.8 0.1= -4.58   
##   
## Potential break point at position: 19

Проведем тест Эндрюса-Зивота для у4 с учетом структурного сдвига 2-го типа (на наклон)

summary(ur.za(df$le\_w, model = c("trend")))

##   
## ################################   
## # Zivot-Andrews Unit Root Test #   
## ################################   
##   
##   
## Call:  
## lm(formula = testmat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.8519 -0.3809 0.1209 0.4667 5.4691   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 35.93193 4.67667 7.683 1.39e-11 \*\*\*  
## y.l1 0.24044 0.09952 2.416 0.0176 \*   
## trend 0.32722 0.04627 7.071 2.58e-10 \*\*\*  
## dt -0.19529 0.03805 -5.133 1.51e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.006 on 95 degrees of freedom  
## (1 пропущенное наблюдение удалено)  
## Multiple R-squared: 0.9522, Adjusted R-squared: 0.9507   
## F-statistic: 631 on 3 and 95 DF, p-value: < 2.2e-16  
##   
##   
## Teststatistic: -7.6319   
## Critical values: 0.01= -4.93 0.05= -4.42 0.1= -4.11   
##   
## Potential break point at position: 50

Проведем тест Эндрюса-Зивота для у4 с учетом структурного сдвига 3-го типа (на константу и наклон)

summary(ur.za(df$le\_w, model = c("both")))

##   
## ################################   
## # Zivot-Andrews Unit Root Test #   
## ################################   
##   
##   
## Call:  
## lm(formula = testmat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.7355 -0.6316 0.1611 0.9567 3.7409   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.24477 3.77287 8.812 6.18e-14 \*\*\*  
## y.l1 0.35687 0.07610 4.690 9.26e-06 \*\*\*  
## trend -0.06667 0.08663 -0.770 0.4435   
## du 6.53849 0.97117 6.733 1.31e-09 \*\*\*  
## dt 0.21690 0.08367 2.592 0.0111 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.829 on 94 degrees of freedom  
## (1 пропущенное наблюдение удалено)  
## Multiple R-squared: 0.9607, Adjusted R-squared: 0.959   
## F-statistic: 574 on 4 and 94 DF, p-value: < 2.2e-16  
##   
##   
## Teststatistic: -8.4513   
## Critical values: 0.01= -5.57 0.05= -5.08 0.1= -4.82   
##   
## Potential break point at position: 19

Тест Дики-Фуллера для исходного ряда

summary(ur.df(df$le\_w, type = c("trend")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.0237 -0.5568 0.0695 0.8336 7.1450   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.55904 4.60837 3.376 0.00107 \*\*  
## z.lag.1 -0.29633 0.09123 -3.248 0.00161 \*\*  
## tt 0.08509 0.02878 2.957 0.00393 \*\*  
## z.diff.lag -0.28470 0.09913 -2.872 0.00504 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.185 on 94 degrees of freedom  
## Multiple R-squared: 0.2709, Adjusted R-squared: 0.2476   
## F-statistic: 11.64 on 3 and 94 DF, p-value: 1.485e-06  
##   
##   
## Value of test-statistic is: -3.2482 4.8101 5.4713   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.04 -3.45 -3.15  
## phi2 6.50 4.88 4.16  
## phi3 8.73 6.49 5.47

summary(ur.df(df$le\_w, type = c("drift")))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.3253 -0.2210 0.0949 0.5605 9.5394   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.82647 1.70577 1.657 0.101   
## z.lag.1 -0.03673 0.02574 -1.427 0.157   
## z.diff.lag -0.41771 0.09187 -4.547 1.6e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.272 on 95 degrees of freedom  
## Multiple R-squared: 0.2031, Adjusted R-squared: 0.1863   
## F-statistic: 12.1 on 2 and 95 DF, p-value: 2.077e-05  
##   
##   
## Value of test-statistic is: -1.4268 2.6303   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.51 -2.89 -2.58  
## phi1 6.70 4.71 3.86

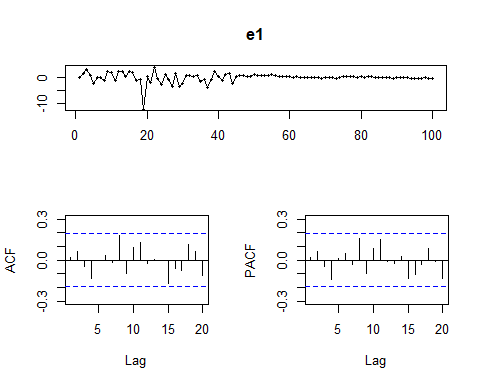
ARIMA(2,0,0)+дамми+тренд

d19 <- c(rep(c(0), times=19), rep(c(1), times=81))  
m1 <- Arima(df$le\_w, order=c(2,0,0), include.drift=TRUE, xreg=d19)  
summary(m1)

## Series: df$le\_w   
## Regression with ARIMA(2,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 intercept drift xreg  
## 0.4395 0.1529 47.5718 0.2215 8.5373  
## s.e. 0.1208 0.0984 0.9885 0.0211 1.6307  
##   
## sigma^2 = 3.308: log likelihood = -199.32  
## AIC=410.64 AICc=411.55 BIC=426.27  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.0008449899 1.772635 0.9820817 -0.1208943 1.74986 0.8149713  
## ACF1  
## Training set 0.01836534

Тестирование остатков модели ARIMA(2,0,0)+дамми+тренд

e1 <- m1$residuals  
tsdisplay(e1)



pander(shapiro.test(e1)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e1

| Test statistic | P value |
| --- | --- |
| 0.7225 | 1.859e-12 \* \* \* |

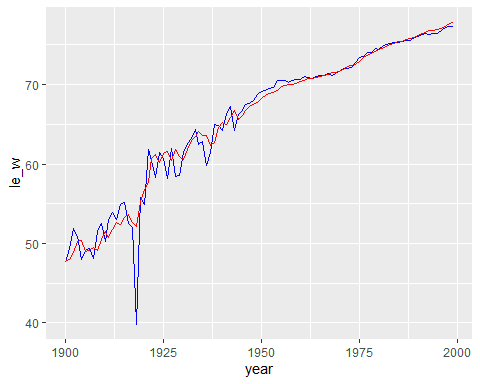
pander(Box.test(e1,lag=10, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e1

| Test statistic | df | P value |
| --- | --- | --- |
| 8.356 | 10 | 0.5941 |

Прогнозирование по модели ARIMA(2,0,0)+дамми+тренд

pred1=fitted(m1)  
ggplot(data=df, aes(year, le\_w))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred1), colour = 'red', size = 0.5)



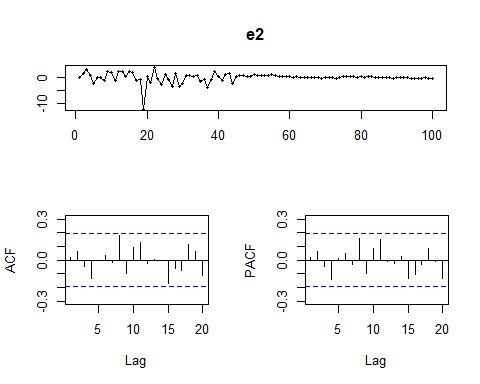
ARIMA(3,0,0)+дамми

m2 <- Arima(df$le\_w, order=c(3,0,0), xreg=d19)  
summary(m2)

## Series: df$le\_w   
## Regression with ARIMA(3,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 intercept xreg  
## 0.7842 0.1367 0.0407 54.6705 14.2431  
## s.e. 0.1295 0.1594 0.1014 4.3299 2.5760  
##   
## sigma^2 = 3.768: log likelihood = -206.86  
## AIC=425.73 AICc=426.63 BIC=441.36  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.1424443 1.891904 1.097428 0.03330078 1.932392 0.9106901  
## ACF1  
## Training set -0.01539607

Тестирование остатков модели ARIMA(3,0,0)+дамми

e2 <- m1$residuals  
tsdisplay(e2)



pander(shapiro.test(e2)) # Тест Шапиро-Уилка на нормальность остатков

Shapiro-Wilk normality test: e2

| Test statistic | P value |
| --- | --- |
| 0.7225 | 1.859e-12 \* \* \* |

pander(Box.test(e2,lag=10, type="Lj")) # Тест Льюинга-Бокса на автокорреляцию

Box-Ljung test: e2

| Test statistic | df | P value |
| --- | --- | --- |
| 8.356 | 10 | 0.5941 |

Прогнозирование по модели ARIMA(3,0,0)+дамми

pred2=fitted(m2)  
ggplot(data=df, aes(year, le\_w))+  
 geom\_line(colour = 'blue')+  
 geom\_line(aes(year, pred2), colour = 'red', size = 0.5)

