

# Lab5

## Завантаження пакетів

```
if (!require("pacman")) install.packages("pacman")

## Loading required package: pacman

pacman::p_load(pacman, ggplot2,
library(moments)
library(agricolae)
library(corrplot)
library(tidyverse)
library(corr)
library(model4you)
library(metrics)
library(readxl)
library(MASS)
library(stats)
library(dplyr)
library(ggplot2)
library(plotly)
library(quantmod)
library(zoo)
library(forecast)
library(astsa)
```

## Завантаження даних

```
# CSV
AstraZeneca.daily <- read.csv(
  "Users/Victoria/Documents/HistoricalPrices2016-2021AstraZeneca.csv", stringsAsFactors = FALSE)
```

## 1. Опис даних

Цей набір даних містить ціни на акції фармацевтичної компанії AstraZeneca з 15.11.2016 по 12.11.2021. Джерело: <https://www.wsj.com/market-data/quotes/AZN/historical-prices>

Розглянемо перші рядки датасету:

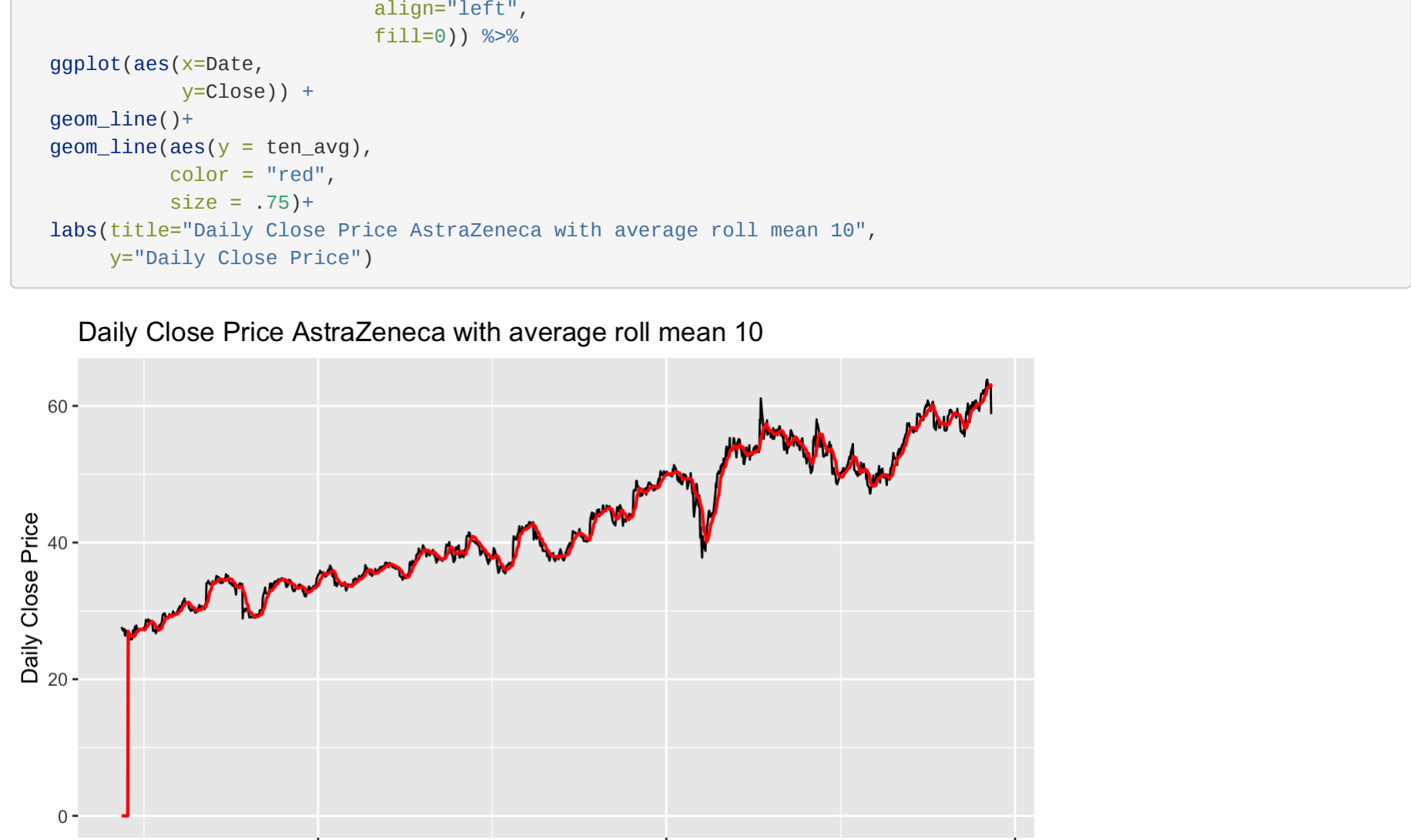
```
head(AstraZeneca.daily)

##      Date      Open      High      Low      Close      Volume
## 1 11/12/21 60.34 60.530 58.465 58.80 14155790
## 2 11/11/21 63.04 63.290 62.850 62.92 2452039
## 3 11/10/21 63.27 63.450 62.970 63.14 2519074
## 4 11/09/21 62.94 62.990 62.580 62.70 3618217
## 5 11/08/21 63.66 63.870 62.970 62.99 4557085
## 6 11/05/21 62.86 63.135 62.230 63.09 4868487
```

## 2. Зображення ряду.

```
AstraZeneca.daily$Date <- as.Date(AstraZeneca.daily$Date, format = "%m/%d/%y")

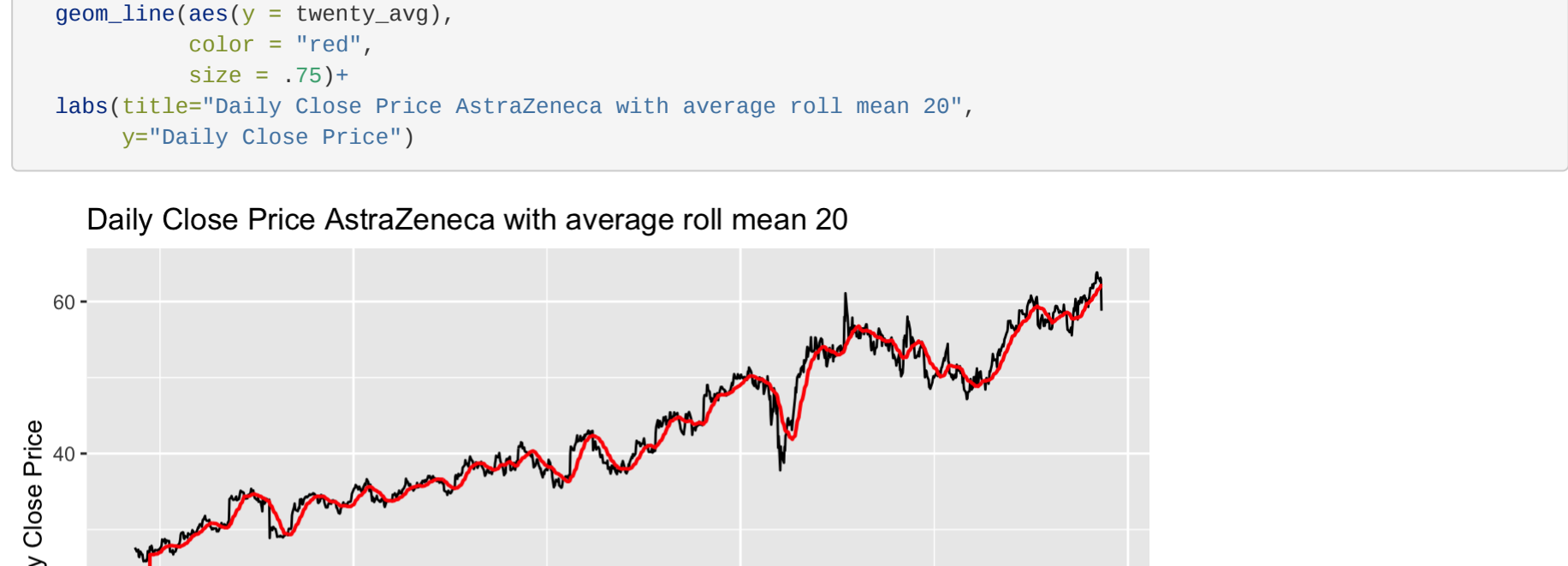
ggplot(aes(x=Date, y=Close)) +
  geom_line() +
  labs(title="Daily Close Price AstraZeneca",
        y="Daily Close Price")
```



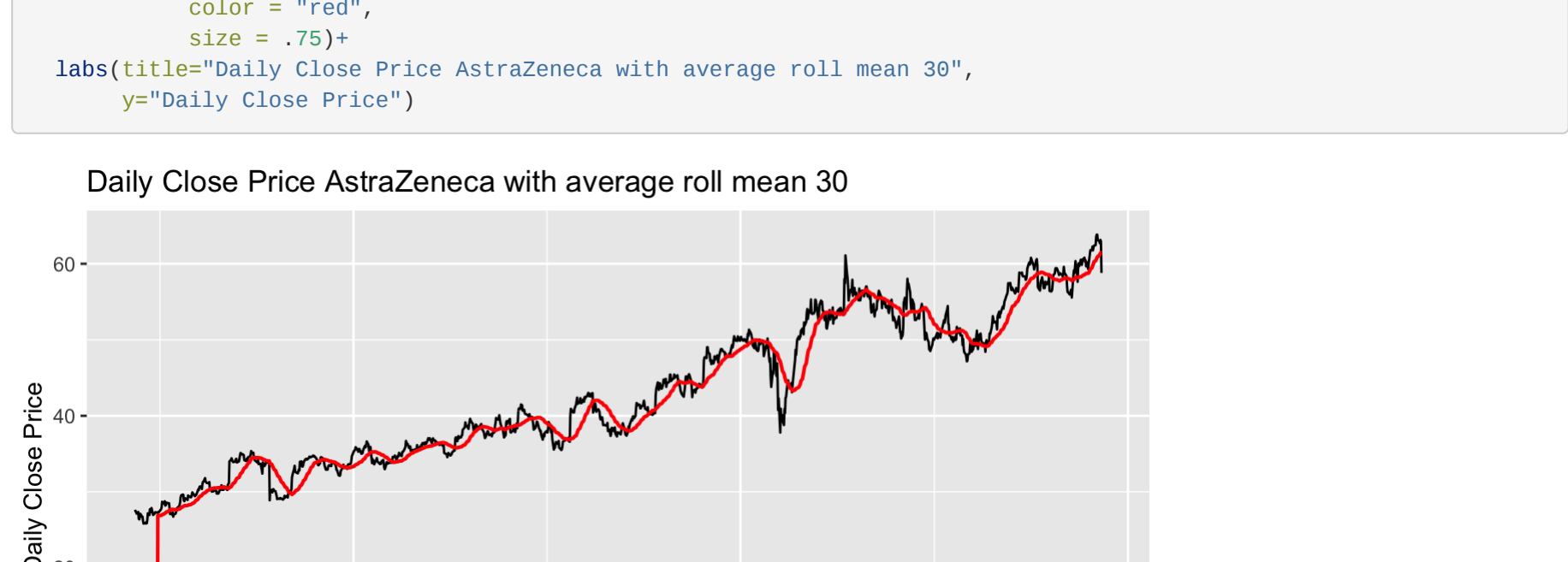
## 3. Проведемо згладжування ряду методом рухомого середнього з різним кроком.

Візьмемо значення порядку 10, 20, 30.

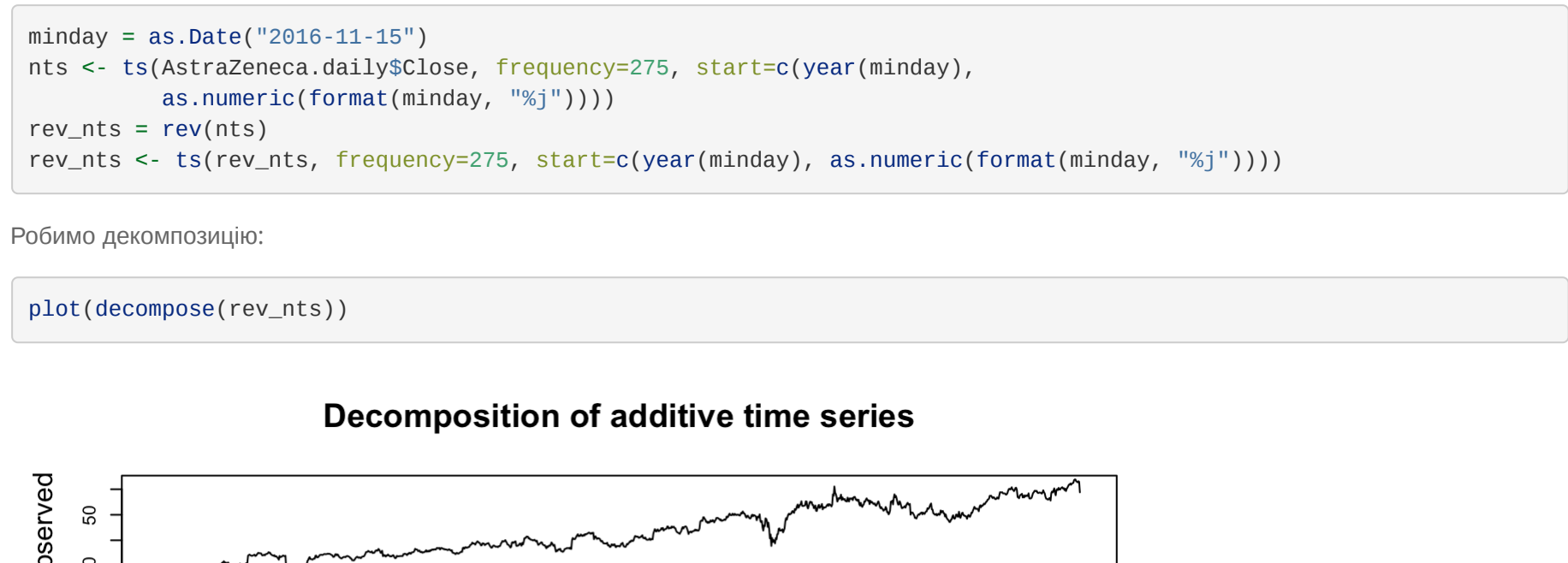
```
AstraZeneca.daily %>%
  mutate(ten_avg= rollmean(Close, 10,
    align="left",
    fill=0)) %>%
  ggplot(aes(x=Date, y=Close)) +
  geom_line() +
  geom_line(aes(y = ten_avg),
    color = "red",
    size = .75) +
  labs(title="Daily Close Price AstraZeneca with average roll mean 10",
        y="Daily Close Price")
```



```
AstraZeneca.daily %>%
  mutate(twenty_avg= rollmean(Close, 20,
    align="left",
    fill=0)) %>%
  ggplot(aes(x=Date, y=Close)) +
  geom_line() +
  geom_line(aes(y = twenty_avg),
    color = "red",
    size = .75) +
  labs(title="Daily Close Price AstraZeneca with average roll mean 20",
        y="Daily Close Price")
```



```
AstraZeneca.daily %>%
  mutate(thirty_avg= rollmean(Close, 30,
    align="left",
    fill=0)) %>%
  ggplot(aes(x=Date, y=Close)) +
  geom_line() +
  geom_line(aes(y = thirty_avg),
    color = "red",
    size = .75) +
  labs(title="Daily Close Price AstraZeneca with average roll mean 30",
        y="Daily Close Price")
```



## 4. Декомпозиція

Зробимо розбиття часового ряду на систематичну, періодичну та хаотичну складові.

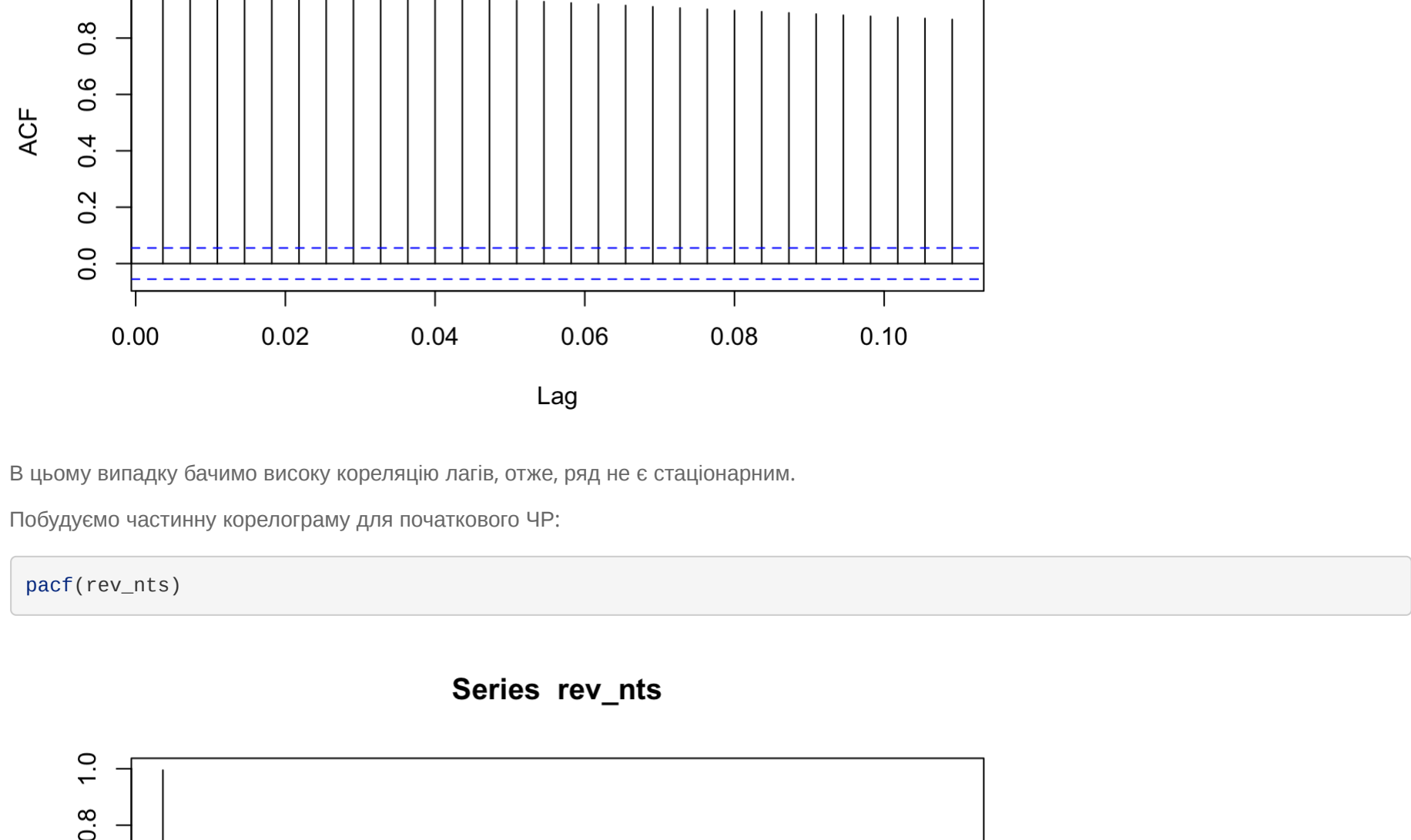
```
#library(zoo)
#A <- zoo(AstraZeneca.daily$Close, seq(from = as.Date("2016-11-15"), to = as.Date("2021-11-12"), by = 1))
```

Перетворимо датасет у часовий ряд:

```
m1day = as.Date("2016-11-15")
nts <- ts(AstraZeneca.daily$Close, frequency=275, start=c(year(m1day),
  as.numeric(format(m1day, "%j"))))
rev_nts = rev(nts)
rev_nts <- ts(rev_nts, frequency=275, start=c(year(m1day), as.numeric(format(m1day, "%j"))))
```

Робимо декомпозицію:

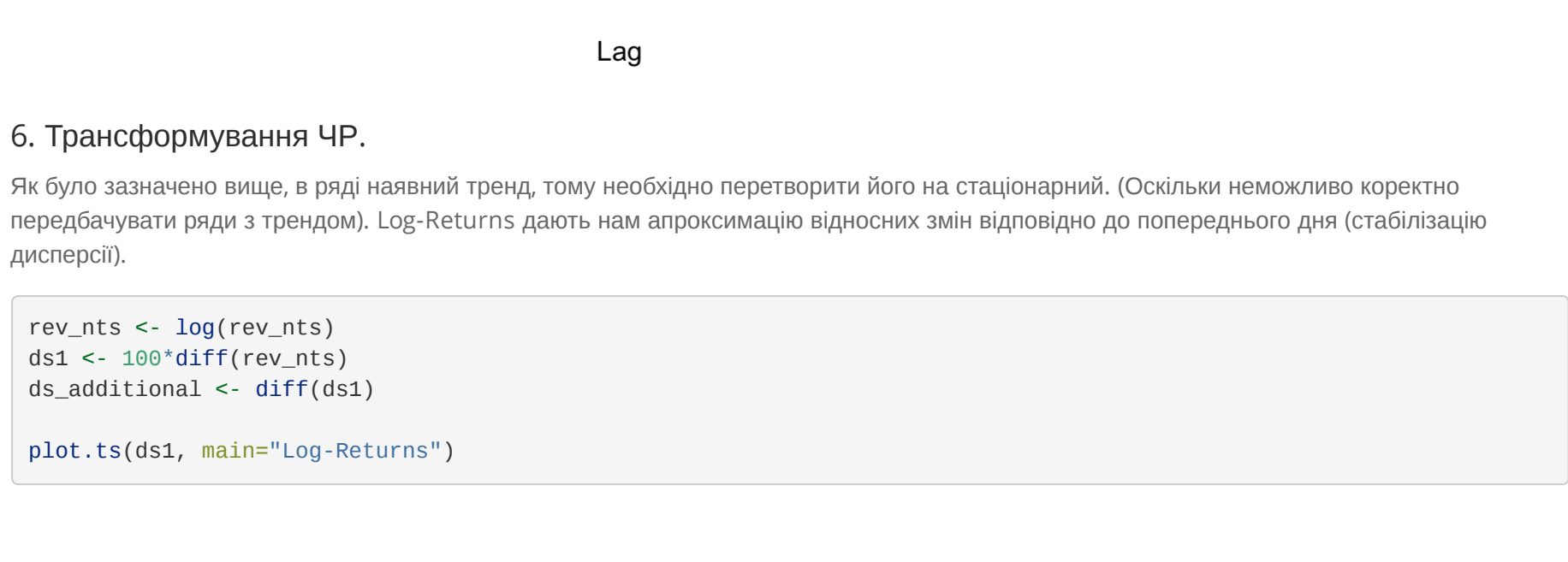
```
plot(decompose(rev_nts))
```



## 5. Корелограма та частинна корелограма

Побудуємо корелограму на початковому ЧР:

```
acf(rev_nts)
```



В цьому випадку бачимо високу кореляцію лагів, отже, ряд не є стаціонарним.

Побудуємо частинну корелограму для початкового ЧР:

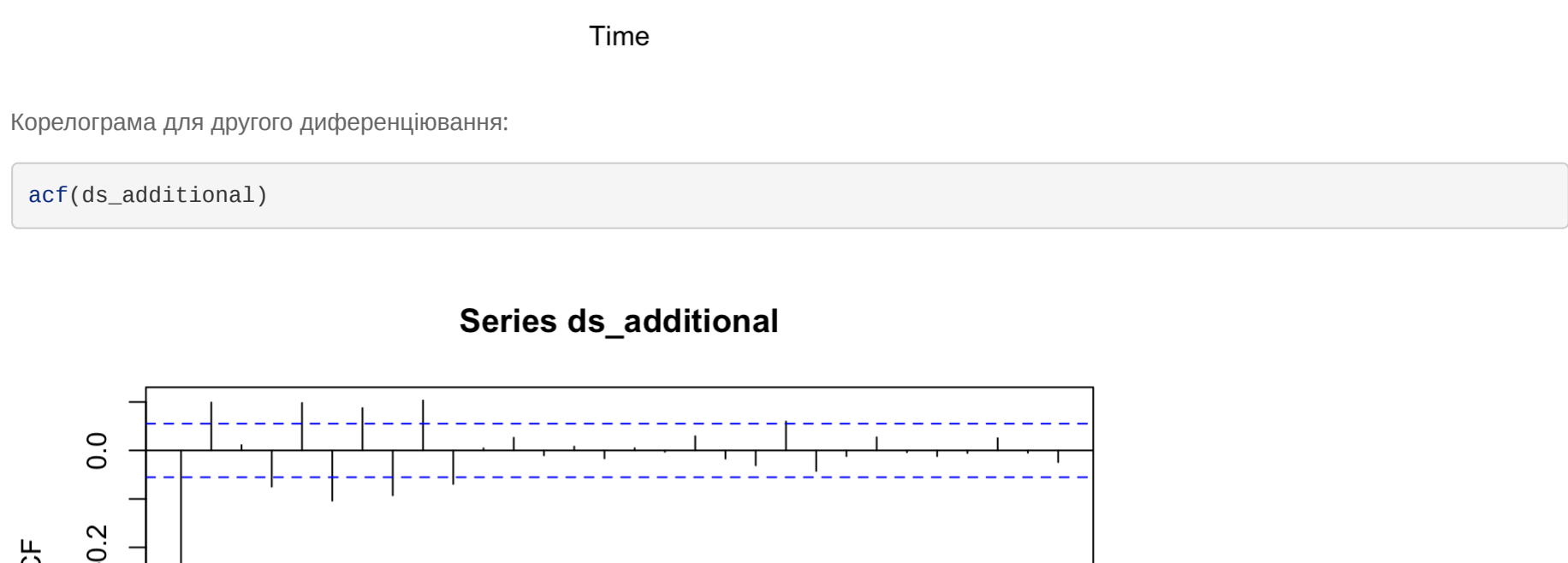
```
pacf(rev_nts)
```



## 6. Трансформування ЧР.

Як було зазначено вище, в ряді наявний тренд, тому необхідно перетворити його на стаціонарний. (Оскільки неможливо коректно передбачувати ряди з трендом). Log-Returns дають нам апроксимацію відносних змін відносно до попереднього дня (стабілізацію дисперсії).

```
rev_nts <- log(rev_nts)
ds1 <- diff(rev_nts)
ds_additional <- diff(ds1)
plot.ts(ds1, main="Log>Returns")
```



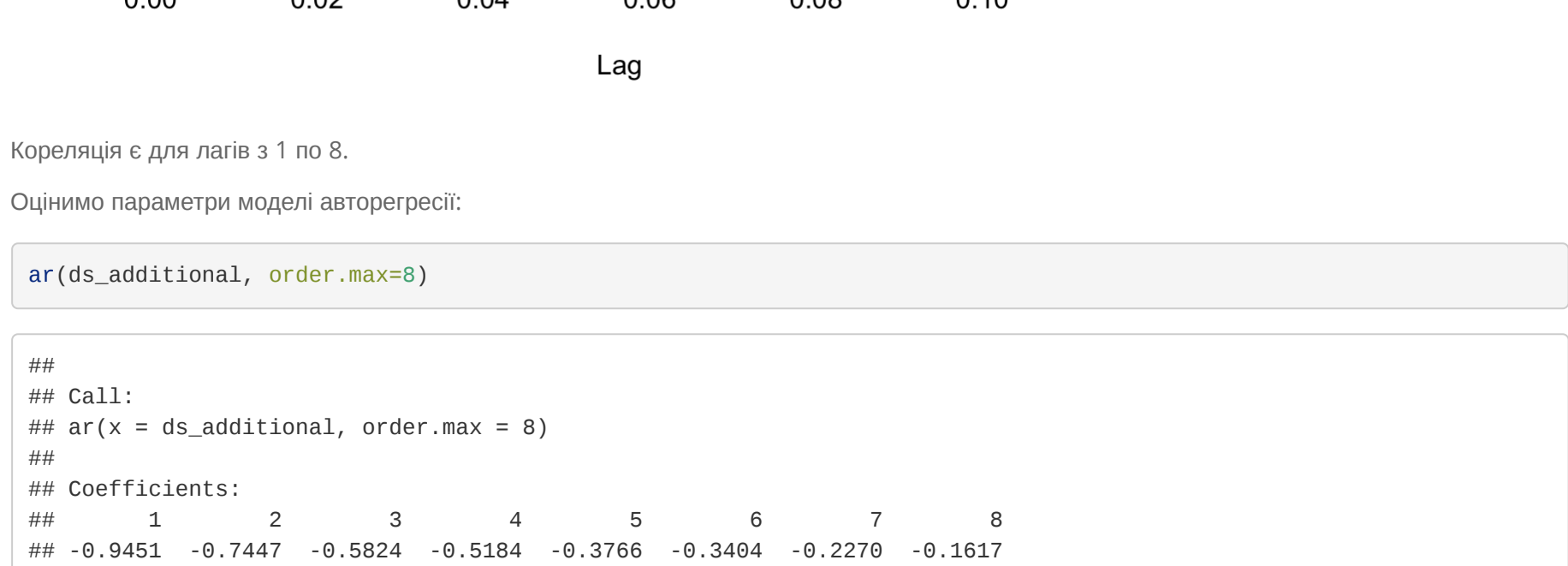
Ще раз застосуємо різнищевий оператор:

```
plot.ts(ds_additional, main="Additional Trend Removal Step")
```



Корелограма для другого диференціювання:

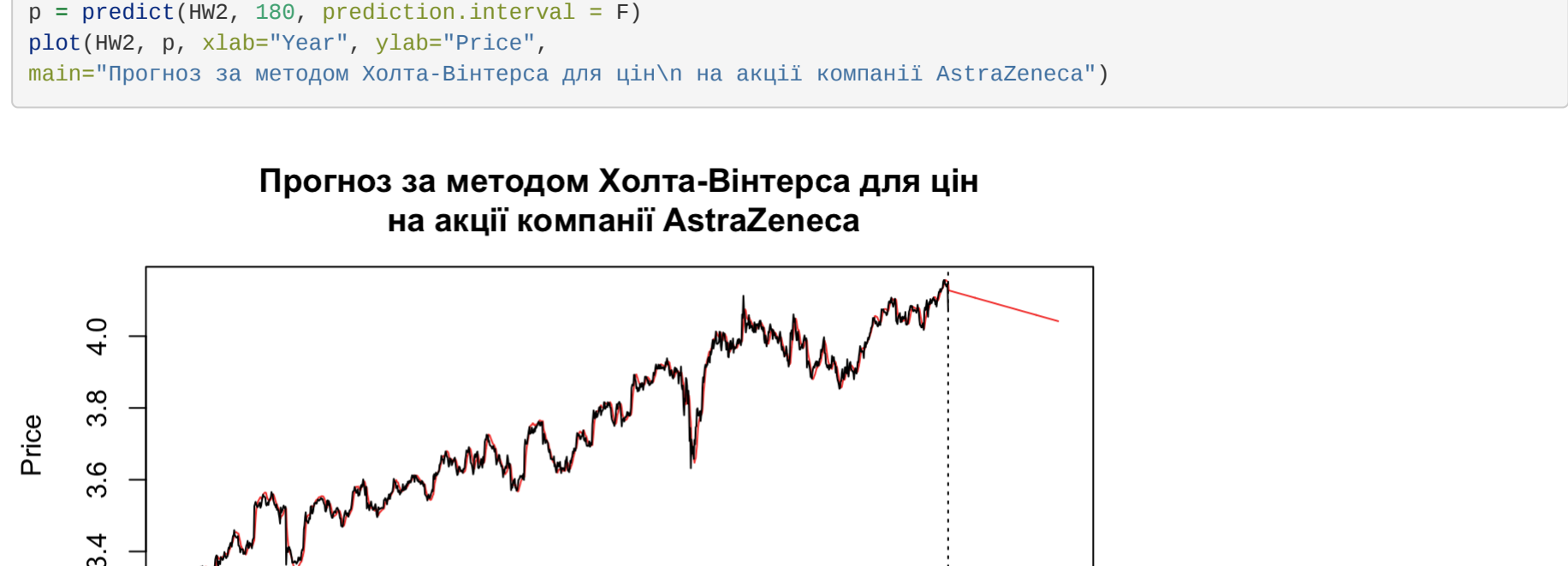
```
acf(ds_additional)
```



Повітка кореляція для лагів 1, 2, 4, 5, 6, 7, 8, 9, 10.

Частинна корелограма:

```
pacf(ds_additional)
```



Кореляція є для лагів з 1 по 8.

Оцінюмо параметри моделі авторегресії:

```
ar(ds_additional, order.max=8)
```

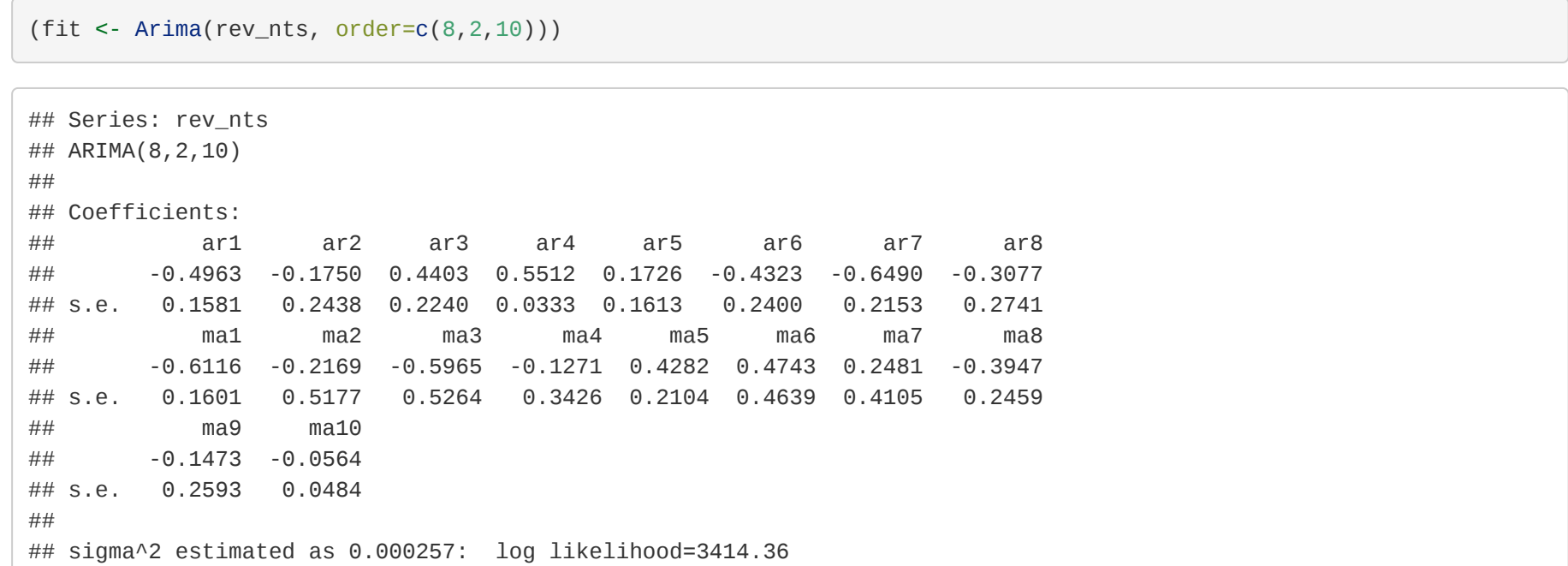
```
##
## Call:
## ar(x = ds_additional, order.max = 8)
##
## Coefficients:
##      1      2      3      4      5      6      7      8
## -0.9451 -0.7447 -0.5824 -0.5184 -0.3766 -0.3404 -0.2270 -0.1617
##
## Order selected 8 sigma^2 estimated as 2.975
```

## 7. Побудова прогнозу

Оскільки у нас є тренд та немає сезонних компонент, використовуємо подвійне експоненційне згладжування (метод Холта).

Метод експоненційного згладжування (для короткострокових прогнозів):

```
Hw2 <- HoltWinters(rev_nts, alpha=0.3, beta=0.1, gamma=FALSE)
p = predict(Hw2, 180, prediction.interval = F)
plot(Hw2, p, xlab="Year", ylab="Price",
  main="Прогноз за методом Холта-Вінтерса для цін на акції компанії AstraZeneca")
```



Побудуємо методом з групи ARIMA:

Оцінюємо якість моделі за допомогою AIC - інформаційний критерій Akaike:

```
AIC(arima(ds_additional, order = c(8,0,10)))
```

```
## Warning in stats::arima(x = x, order = order, seasonal = seasonal, xreg =
## xreg, : possible convergence problem: optim gave code = 1
```

```
## [1] 4787.181
```

```
AIC(arima(ds_additional, order = c(1,0,1)))
```

```
## [1] 4779.83
```

Оскільки значення функції AIC на наборі (8,0,10) менше, ніж на наборі (1,0,1), використаємо для прогнозу ARIMA перші параметри.

Метод ARIMA

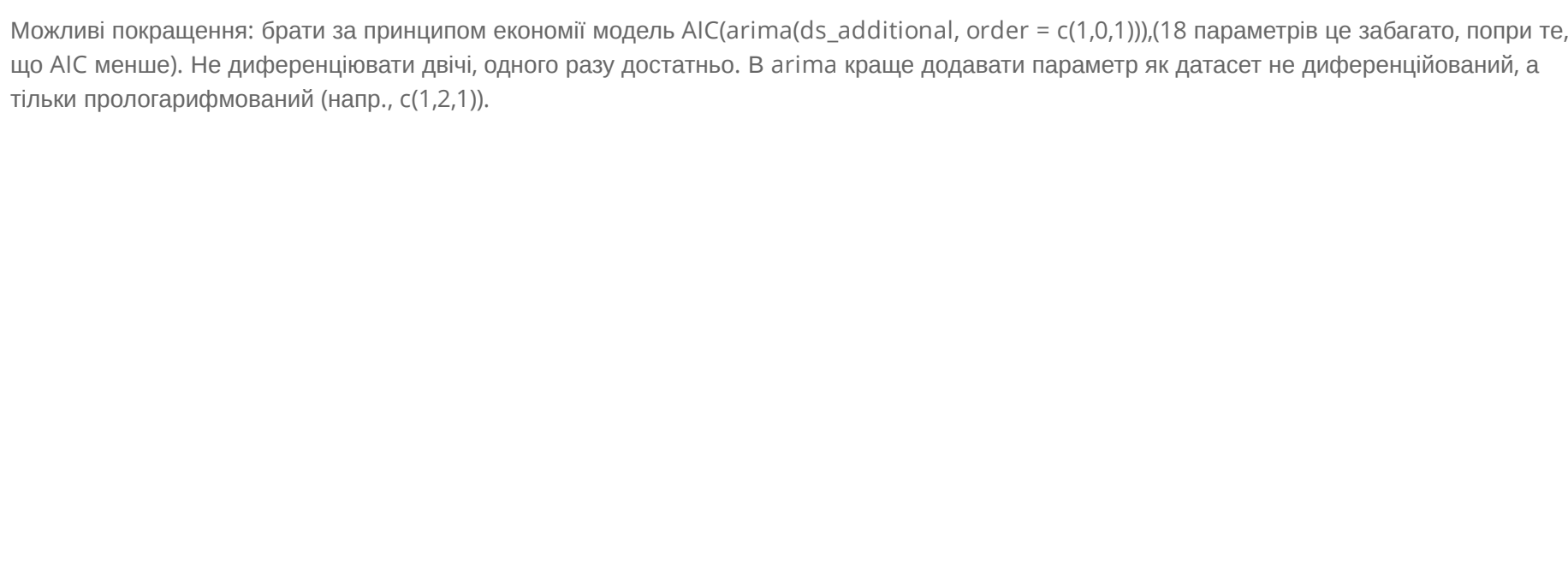
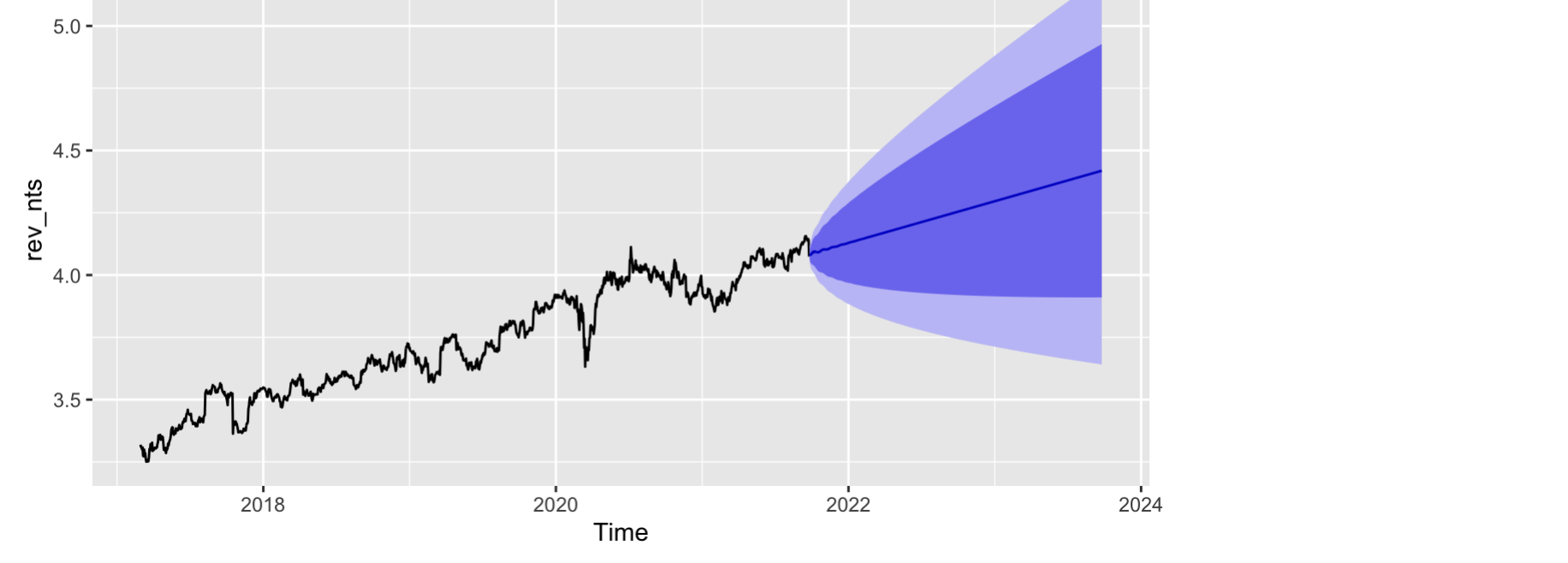
```
(fit <- Arima(rev_nts, order=c(8,2,10)))

## Series: rev_nts
## ARIMA(8,2,10)
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
##    -0.4963 -0.1750  0.4403  0.5512  0.1726 -0.4323 -0.6490 -0.3077
## s.e.   0.1581  0.2430  0.2240  0.0333  0.1613  0.2400  0.2153  0.2741
##      ma1      ma2      ma3      ma4      ma5      ma6      ma7
##    -0.6116 -0.2169 -0.5965 -0.1271  0.4282  0.4743  0.2481 -0.3947
## s.e.   0.1601  0.5177  0.5264  0.3426  0.2104  0.4639  0.4105  0.2459
##      me9      me10
##    -0.1473 -0.0564
## s.e.   0.2593  0.0484
##
## sigma^2 estimated as 0.068257: log likelihood=3414.36
## AIC=-6790.73 AICc=-6790.11 BIC=-6693.15
```

## 8. Побудова корелограм залишків

Залишки моделі схожі на білий шум: хаотично змінюються навколо середнього значення 0, зі сталою дисперсією, вони некорельовані (корелограма), та приблизно нормально розподілені (Q-Q plot). Тест Льюнга-Бокса має високі p-значення для всіх лагів, отже, маємо підстави прийняти основну гіпотезу про односторонню рівність нулю всіх автокореляцій показов моделі (для всіх лагів одночасно).

```
checkresiduals(fit)
```

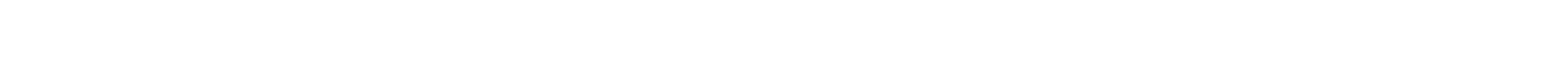


```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(8,2,10)
## Q* = 222.07, df = 234, p-value = 0.7819
## Model df: 18. Total lags used: 252
```

Залишки моделі схожі на білий шум: хаотично змінюються навколо середнього значення 0, зі сталою дисперсією, вони некорельовані (корелограма), та приблизно нормально розподілені. P-value більше за 0.05, отже, приймаємо гіпотезу про незалежність усіх залишків.

Прогноз:

```
autoplot(forecast(fit))
```



Можливо порівняння: брати за принципом економічної моделі AIC(arima(ds\_additional, order = c(1,0,1)))(18 параметрів це забагато, попри те, що AIC менше). Не диференціювати дані, одного разу достатньо. В агіта краще додавати параметр як датасет не диференційований, а тільки прогнозований (напр., c(1,2,1)).