Moving average modely (kĺzavé priemery)

Beáta Stehlíková

Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

Motivácia

Príklad: Ceny kakaa

Ben Vogelvang: *Econometrics. Theory and Applications with EViews.*, Pearson Education Limited, 2005.

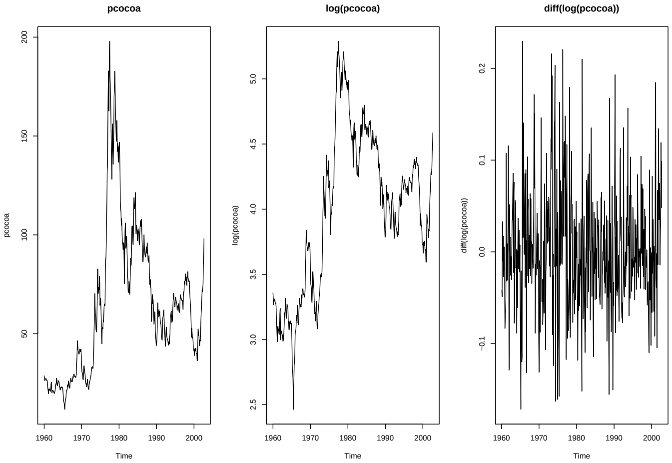
Chapter 14.7. - The Box-Jenkins Approach in Practice

- Mesačné dáta, január 1960 - september 2002
- pcocoa - cena kakka, zlogaritmujeme a kvôli stacionarite budeme pracovať s diferenciami ⇒ modelujeme teda percentuálnu zmenu ceny

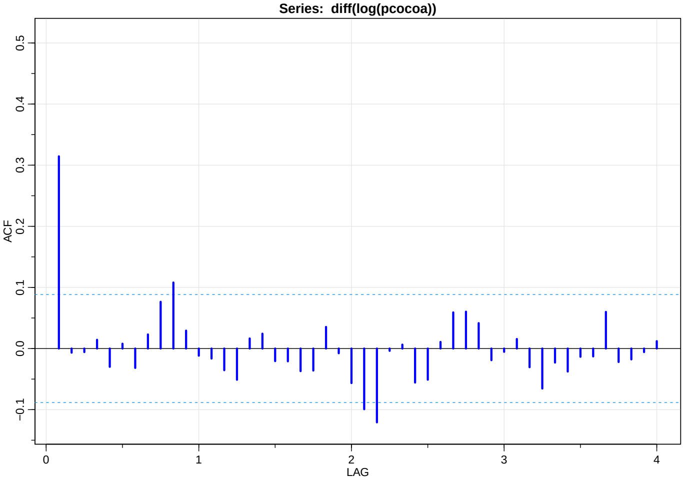
$$
\begin{aligned}
\Delta \ln(p_t) &= \ln(p_t) - \ln(p_{t-1}) = \ln\left(\frac{p_t}{p_{t-1}}\right) \\
&= \ln\left(1 + \frac{p_t - p_{t-1}}{p_{t-1}}\right) \approx \frac{p_t - p_{t-1}}{p_{t-1}}
\end{aligned}
$$

(lebo $\ln(1+x) \approx x$ pre x malé)

- Priebeh dát: pôvodné, zlogaritmované, diferencie logaritmov



- Modelujeme diferencie logaritmov
- Zobrazíme si pre ne odhad ACF



- Prvá výrazne nenulová, ostatné skoro nulové
- Nevyzerá to ako typický priebeh pre AR proces:
  - tam klesala ACF postupne
  - priebeh tohto typu mala PACF

Príklad z prvej prednášky

- Nech u_t je biely šum, definujme

$$x_t = u_t + u_{t-1}$$

- Vypočítali sme:

$$\mathbb{E}(x_t) = 0, \mathbb{D}(x_t) = 2\sigma^2$$

$$\mathrm{Cov}(x_t, x_{t+k}) = \begin{cases} \sigma^2 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \ldots \end{cases}$$

$$\mathrm{Cor}(x_t, x_{t+k}) = \begin{cases} 1/2 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \ldots \end{cases}$$

- ACF je nulová pre $k = 2, 3, \ldots$ - presne tá vlastnosť, ktorú potrebujeme

Zovšeobecnenie príkladu a plán prednášky

- Príklad zovšeobecníme tak, aby
  - sa zachovala nulovosť korelácií okrem prvej
  - prvá korelácia mohla nadobúdať aj iné hodnoty ako $1/2$ $\Rightarrow$ MA(1) procesy
- Procesy, ktoré majú prvých q nenulových a ostatné nulové $\Rightarrow$ MA(q) procesy

MA(1) procesy

Definícia, stacionarita, momenty

- Nech u_t je biely šum, potom

$$x_t = \mu + u_t - \beta u_{t-1}$$

  sa nazýva moving average proces prvého rádu - MA(1)
- Woldova reprezentácia: $x_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}$, pre MA(1) proces: $\psi_0 = 1, \psi_1 = -\beta, \psi_j = 0$ pre $j = 2, 3, \dots$ → **vždy stacionárny**
- Momenty a ACF:

$$\mathbb{E}(x_t) = \mu, \mathbb{D}(x_t) = (1 + \beta^2)\sigma^2$$

$$\mathrm{Cov}(x_t, x_{t+k}) = \begin{cases} -\beta\sigma^2 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \dots \end{cases}$$

$$\mathrm{Cor}(x_t, x_{t+k}) = \begin{cases} -\frac{\beta}{1+\beta^2} & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \dots \end{cases}$$

PACF

- Pripomeňme si PACF:

$$\Phi_{kk} = \frac{\det\begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & \rho(k) \end{pmatrix}}{\det\begin{pmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{pmatrix}}$$

- Pre MA(1) proces dosadzujeme $\rho(k) = 0$ pre $k = 2, 3, \dots$

$$\begin{aligned}
\Phi_{11} &= \rho(1) \\
\Phi_{22} &= \frac{\det\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{pmatrix}}{\det\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix}} = \frac{\det\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 0 \end{pmatrix}}{\det\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix}} = \frac{-\rho(1)^2}{1-\rho(1)^2} \\
\Phi_{33} &= \frac{\det\begin{pmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{pmatrix}}{\det\begin{pmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{pmatrix}} = \frac{\det\begin{pmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & 0 \\ 0 & \rho(1) & 0 \end{pmatrix}}{\det\begin{pmatrix} 1 & \rho(1) & 0 \\ \rho(1) & 1 & \rho(1) \\ 0 & \rho(1) & 1 \end{pmatrix}} \\
&= \frac{\rho(1)^3}{1-2\rho(1)^2}
\end{aligned}$$

Cvičenie: Odvoďte ďalší člen PACF:

$$\Phi_{44} = \frac{-\rho(1)^4}{(1-\rho(1)^2)^2 - \rho(1)^2}$$

Príklad 1: výpočet ACF a PACF v R-ku, $x_t = u_t + 0.7u_{t-1}$

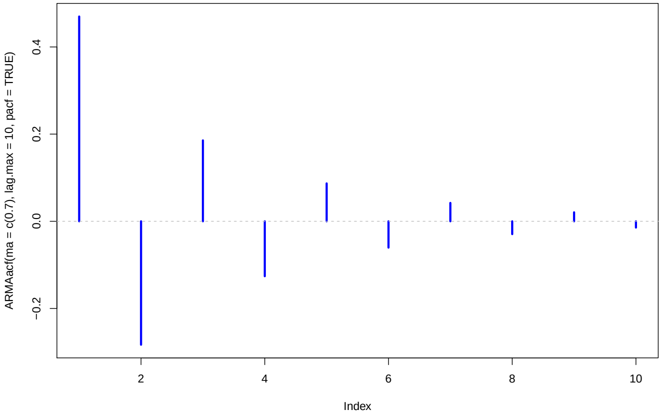
```
ARMAacf(ma = c(0.7), lag.max = 10)
```

```
##         0         1         2         3         4
## 1.0000000 0.4697987 0.0000000 0.0000000 0.0000000 0.0000
##         8         9        10
## 0.0000000 0.0000000 0.0000000
```

```
ARMAacf(ma = c(0.7), lag.max = 10, pacf = TRUE)
```

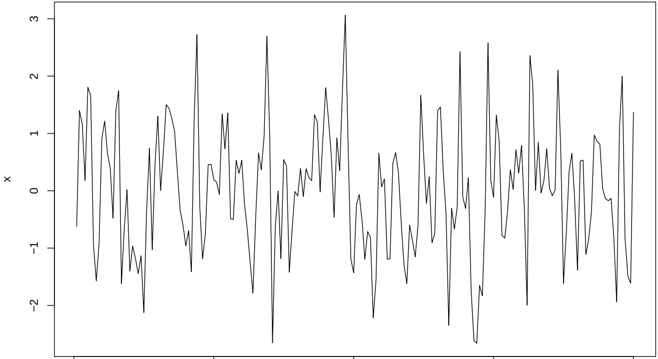
```
##  [1]  0.46979866 -0.28322062  0.18563127 -0.12601048  0.
##  [7]  0.04214074 -0.02944844  0.02059677 -0.01441187
```

```r
plot(ARMAacf(ma = c(0.7), lag.max = 10, pacf = TRUE),
     type = "h", col = "blue", lwd = 3)
abline(h = 0, col = "grey", lty = 2)
```



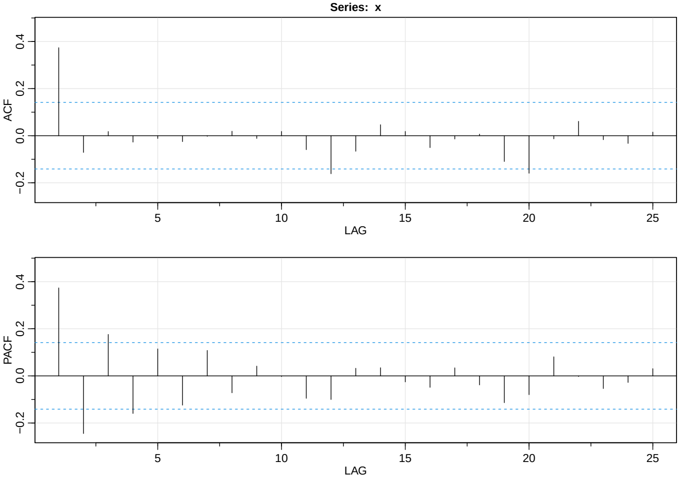
Príklad 2: simulované dáta, $x_t = u_t + 0.7u_{t-1}$

```
set.seed(123)
x <- arima.sim(model = list(ma = c(0.7)), n = 200)
plot(x)
```



Odhad ACF a PACF:

```
acf2(x) # naraz ACF (bez lagu 0) a PACF, balik astsa
```



Príklad 3: procesy s rovnakou ACF

- Nech u_t je biely šum s rozdelením $N(0,4)$, definujme

$$x_t = u_t + \frac{1}{2} u_{t-1}$$

Potom: $\mathbb{E}(x_t) = 0$, $\mathbb{D}(x_t) = (1 + (1/2)^2) \times 4 = 5$

$$\mathrm{Cor}(x_t, x_{t+k}) = \begin{cases} \frac{1/2}{1+1/4} = 2/5 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \ldots \end{cases}$$

- Nech u_t je biely šum s rozdelením $N(0,1)$, definujme

$$x_t = u_t + 2u_{t-1}$$

Potom: $\mathbb{E}(x_t) = 0$, $\mathbb{D}(x_t) = (1 + 2^2) \times 1 = 5$

$$\mathrm{Cor}(x_t, x_{t+k}) = \begin{cases} \frac{2}{1+4} = 2/5 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \ldots \end{cases}$$

MA(1) procesy s danou ACF, invertovateľnosť

MA(1) procesy s danou ACF

- Zovšeobecníme príklad 3 (pre dva procesy vyšla rovnaká ACF)
- Majme MA(1) proces, teda ACF tvaru

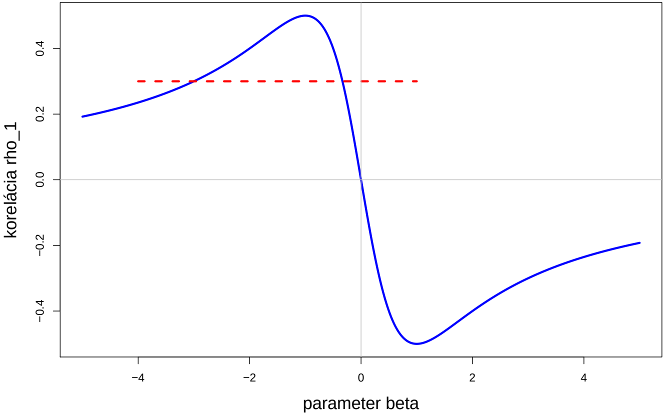
$$\mathrm{Cor}(x_t, x_{t+k}) = \begin{cases} -\frac{\beta}{1+\beta^2} & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \ldots \end{cases}$$

- Predpokladjme teraz, že máme danú hodnotu $\rho_1 = \rho(1)$ a chceme z nej spätne určiť koeficient β:

$$\rho_1 = -\frac{\beta}{1+\beta^2} \Rightarrow \beta = ?$$

- Máme teda rovnicu

$$\rho_1 = -\frac{\beta}{1+\beta^2} \Rightarrow \beta^2 + \frac{1}{\rho_1}\beta + 1 = 0$$



- Rovnica $\beta^2 + \frac{1}{\rho_1}\beta + 1 = 0$ má pre $|\rho_1| < 1/2$ dve riešenia β_1, β_2, ktoré spĺňajú $\beta_1\beta_2 = 1$
- Procesy

$$x_t = \mu + u_t - \beta u_{t-1}, x_t = \mu + u_t - \frac{1}{\beta} u_{t-1}$$

  majú rovnakú ACF
- Ak chceme jednoznačnú parametrizáciu, potrebujeme dodať ďalšiu podmienku

Invertovateľnosť

- Budeme sa snažiť zapísať proces v tvare AR(∞):

$$x_t = \tilde{\mu} + u_t + \psi_1 x_{t-1} + \psi_2 x_{t-2} + \psi_3 x_{t-3} + \dots$$

  Ak sa to dá spraviť, proces sa nazýva invertovateľný

- Pre MA(1) proces:

$$\begin{aligned} x_t &= \mu + (1 - \beta L) u_t \\ (1 - \beta L)^{-1} x_t &= (1 - \beta L)^{-1} \mu + u_t \end{aligned}$$

  inverzia $(1 - \beta L)^{-1}$ existuje pre $|\beta| < 1$, vtedy

$$(1 + \beta L + \beta^2 L^2 + \dots) x_t = \frac{\mu}{1 - \beta} + u_t$$

$$x_t + \beta x_{t-1} + \beta^2 x_{t-2} + \dots = \frac{\mu}{1 - \beta} + u_t$$

$$x_t = \frac{\mu}{1 - \beta} + u_t - \beta x_{t-1} - \beta^2 x_{t-2} - \dots$$

- Dostali sme teda podmienku invertovateľnosti MA(1) procesu: $|\beta| < 1$
- Iný zápis tejto podmienky:
  - máme proces $x_t = \mu + (1 - \beta L)u_t$
  - koreň polynómu $1 - \beta L$ je $1/\beta$
  - podmienka invertovateľnosti teda hovorí, že koreň polynómu $1 - \beta L$ musí byť v absolútnej hodnote väčší ako 1, pri zakreslení do komplexnej roviny mimo jednotkového kruhu
- Táto podmienka sa ešte objaví v súvislosti s predikciami.

Reálne dáta: ceny kakaa zo začiatku slajdov

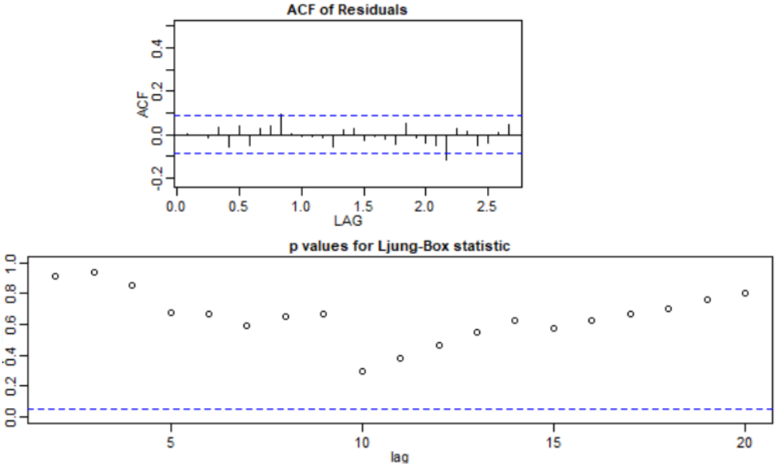
- Odhadneme MA(1) model pre diferencie logaritmov cien (teda percentuálne zmeny cien):

```
##         Estimate     SE t.value p.value
## ma1       0.3520 0.0402  8.7585  0.0000
## xmean     0.0024 0.0037  0.6501  0.5159
```

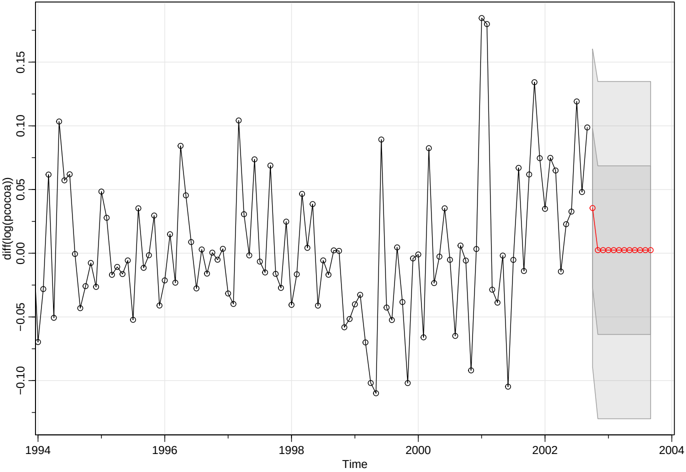
- Zapíšeme model pre premennú $x_t = \Delta \ln(pcocoa_t)$:

$$x_t = 0.0024 + u_t + 0.3520 u_{t-1}$$

- Rezíduá sú v poriadku - model je dobrý



- Predikcie z R-ka na nasledujúci rok.
- Všimnime si, že nie sú konštantné. Prečo nie sú konštantné predikcie z modelu $x_t = \mu + u_t - \beta u_{t-1}$ → k tomu sa vrátime.



MA(q) procesy

Definícia, momenty, ACF, PACF, invertovateľnosť

- Nech u_t je biely šum, potom

$$x_t = \mu + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots - \beta_q u_{t-q}$$

  sa nazýva moving average proces rádu q - MA(q)
- Woldova reprezentácia: $x_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}$, pre MA(q) proces: $\psi_0 = 1, \psi_1 = -\beta_1, \ldots, \psi_q = -\beta_q \psi_j = 0$ pre $j > q$ $\rightarrow$ **vždy je stacionárny**
- Momenty a ACF:

$$\mathbb{E}(x_t) = \mu, \mathbb{D}(x_t) = (1 + \beta_1^2 + \ldots \beta_q^2)\sigma^2$$

$$\mathrm{Cov}(x_t, x_{t+k}) = 0 \text{ pre } k = q+1, q+2, \ldots$$

$$\mathrm{Cor}(x_t, x_{t+k}) = 0 \text{ pre } k = q+1, q+2, \ldots$$

- Výpočet prvých q autokorelácií:

$$\begin{aligned} \mathrm{Cov}(x_t, x_{t+k}) &= \mathbb{E}[(u_t - \beta_1 u_{t-1} - \cdots - \beta_q u_{t-q}) \times \\ &\quad (u_{t+k} - -\beta_1 u_{t+k-1} - \cdots - \beta_q u_{t+k-q})] \end{aligned}$$

- Postupne dostaneme

$$\begin{aligned} k = 1 \quad &\Rightarrow \quad \gamma(1) = (-\beta_1 + \beta_1\beta_2 + \cdots + \beta_{q-1}\beta_q)\sigma^2 \\ k = 2 \quad &\Rightarrow \quad \gamma(2) = (-\beta_2 + \beta_1\beta_3 + \cdots + \beta_{q-2}\beta_q)\sigma^2 \\ &\quad \ldots \\ k = q \quad &\Rightarrow \quad \gamma(q) = (-\beta_q)\sigma^2 \end{aligned}$$

- Pre konkrétny model je praktickejšie počítať to priamo, namiesto dosadzovania do týchto vzťahov (tie sú užitočné kvôli postupu)
- ACF: autokovariancie vydelíme disperziou
- PACF: ACF dosadzujeme do všeobecného vzorca

Cvičenie

Uvažujme proces $x_t = 10 + u_t + 0.5u_{t-1} - 0.2u_{t-2} + 0.1u_{t-3}$

- Ukážte, že je invertovateľný.
- Vypočítajte jeho ACF
- Vypočítajte prvé tri hodnoty jeho PACF

Na kontrolu:

```
ARMAacf(ma=c(0.5,-0.2,0.1), lag.max = 4)[-1] # vynechame a
```

```
##          1           2          3          4
##  0.29230769 -0.11538462  0.07692308  0.00000000
```

```
ARMAacf(ma=c(0.5,-0.2,0.1), lag.max = 3, pacf = TRUE)
```

```
## [1]  0.2923077 -0.2195911  0.2093677
```

Reálne dáta: výmenné kurzy

- Dáta: výmenné kurzy voči euru, zo stránky Európskej centrálnej banky
- Z týchto dát zoberieme hodnoty pre Nový Zéland (novozélandský dolár, NZD) z rokov 2021 a 2022.

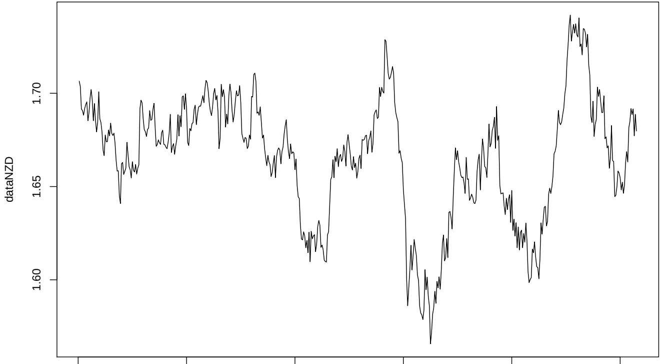
https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html

```r
dataVsetky <- read.csv("data/eurofxref-hist.csv")
head(dataVsetky)
```

```
##          Date    USD    JPY    BGN CYP    CZK    DKK EEK
## 1 2023-09-05 1.0731 158.20 1.9558 N/A 24.161 7.4533 N/A
## 2 2023-09-04 1.0801 158.11 1.9558 N/A 24.108 7.4527 N/A
## 3 2023-09-01 1.0844 157.47 1.9558 N/A 24.118 7.4528 N/A
## 4 2023-08-31 1.0868 158.49 1.9558 N/A 24.072 7.4523 N/A
## 5 2023-08-30 1.0886 159.15 1.9558 N/A 24.107 7.4531 N/A
## 6 2023-08-29 1.0803 158.93 1.9558 N/A 24.163 7.4529 N/A
```

- Vektor s týmito hodnotami `dataNZD` je pripravený vo *workfile* `MAprednaska.Rdata`

```
load("MAprednaska.Rdata")
plot(dataNZD, type = "l") # x-ova os len cislo pozorovania
```



Opačné poradie riadkov v *data frame*:

- Často sa stretneme s tým, že najnovšie pozorovania sú na začiatku a najstaršie na konci:

```
head(dataVsetky$Date)
```

```
## [1] "2023-09-05" "2023-09-04" "2023-09-01" "2023-08-31"
## [6] "2023-08-29"
```

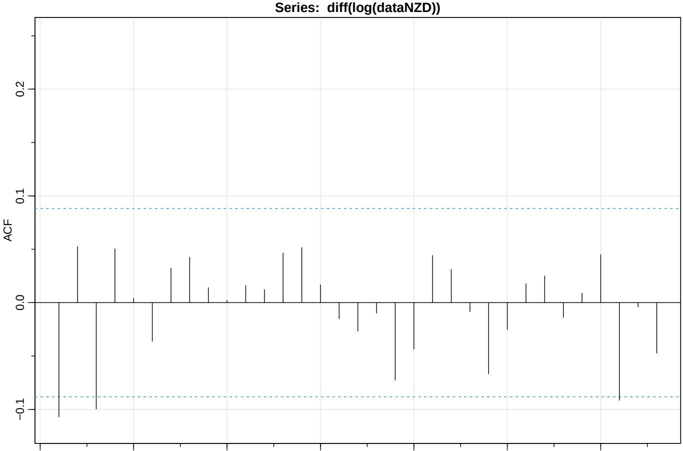
- Jednoduchý postup, ako ich otočiť:

```
dataVsetkyUsporiadane <- dataVsetky[nrow(dataVsetky):1, ]
head(dataVsetkyUsporiadane$Date)
```

```
## [1] "1999-01-04" "1999-01-05" "1999-01-06" "1999-01-07"
## [6] "1999-01-11"
```

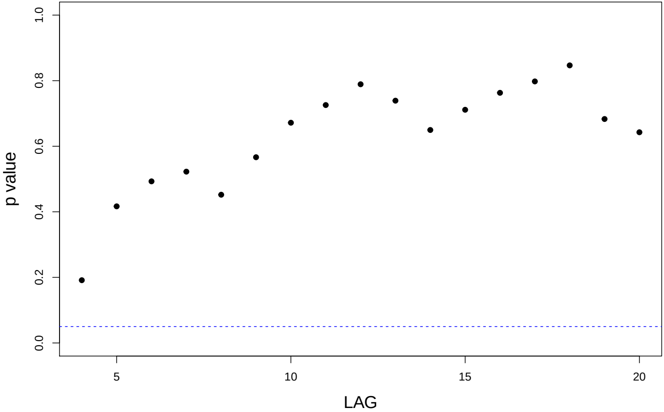
- Budeme pracovať znovu s diferenciami logaritmov, teda modelujeme percentuálne zmeny

```
acf1(diff(log(dataNZD)))
```



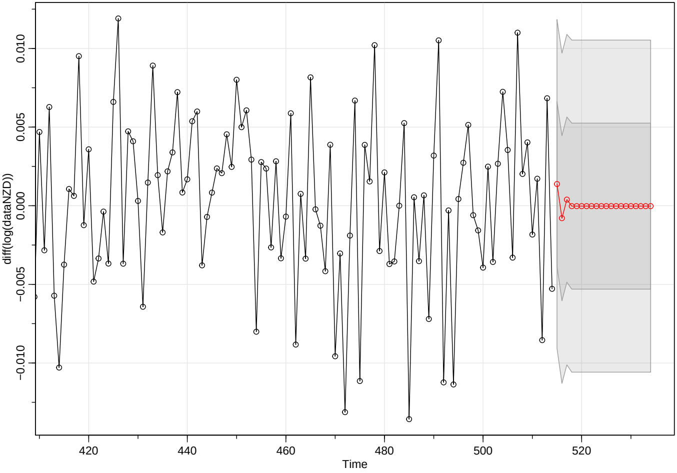
- Odhadneme MA(3) a pozrieme si Ljung-Boxov test pre rezíduá:

```
ma3 <- sarima(diff(log(dataNZD)), 0, 0, 3, details = FALSE)
```



- Predikcie - znovu si všimnime, že nie sú konštantné

```
sarima.for(diff(log(dataNZD)), 20, 0, 0, 3)
```



Nepovinné dodatky

Načítanie dát zo zip archívu

- Dáta sa zo stránky centrálnej banky stiahnu ako zip archív, v ktorom je csv súbor:

```
unzip("data/eurofxref-hist.zip", list = TRUE)
```

```
##                   Name  Length                Date
## 1 eurofxref-hist.csv 1713117 2023-09-05 15:55:00
```

- So zip archívom sa dá v R pracovať aj priamo:

```
# rovnaky data frame ako dataVsetky
dataVsetky2 <- read.csv(unzip("data/eurofxref-hist.zip"))
```

Dátumy v R

```
class(dataVsetky$Date)
```

```
## [1] "character"
```

```
datum <- "2022-01-01"
class(datum)
```

```
## [1] "character"
```

```
datum <- as.Date(datum, format = "%Y-%m-%d")
class(datum)
```

```
## [1] "Date"
```

- Na ukážku (jazyk sa dá zmeniť pomocou `Sys.setlocale`):

```
format(datum, format = "%b %y")
```

```
## [1] "Jan 22"
```

```
format(datum, format = "%B %Y")
```

```
## [1] "January 2022"
```

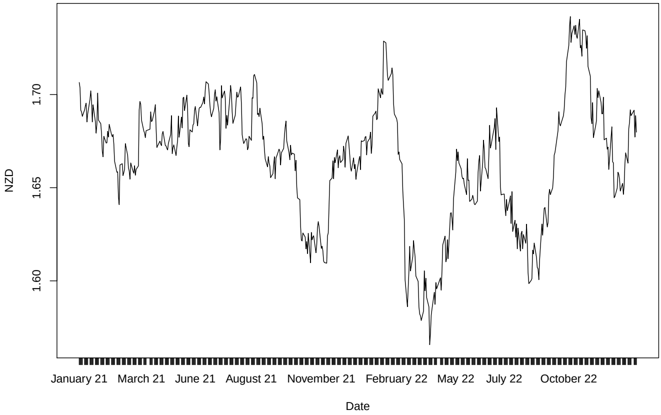
- Existujú aj iné možnosti práce s dátumami a časovými radmi, napr. tu sa nedá použiť typ `ts` - roky nemajú rovnaký počet dní, nedá sa zadať `frequency`
- Balíky: `zoo` (Z's Ordered Observations) `xts` (eXtensible Time Series)

Takto môžeme z *data framu* s dátumami spraviť krajší graf:

```
load("MApred naskaSDATUMOM.Rdata")
str(dataNZDsdatumom)
```

```
## 'data.frame':    515 obs. of  2 variables:
##  $ Date: Date, format: "2021-01-04" "2021-01-05" ...
##  $ NZD : num  1.71 1.7 1.69 1.69 1.69 ...
```

```
plot(NZD ~ Date, dataNZDsdatumom, xaxt = "n", type = "l")
axis(1, dataNZDsdatumom$Date,
     format(dataNZDsdatumom$Date, "%B %y"))
```



Hľadanie riadkov s dátumom v danom roku

- Užitočná je tu funkcia `grep`

```
# v ktorych retazcoch sa nachadza "ok"
grep(pattern = "ok", x = c("pondelok", "streda", "piatok"))
```

```
## [1] 1 3
```

- Vyvorme si dátumy (ďalšia užitočná funkcia je `outer`):

```
datumy <- outer(month.abb, 2020:2023, paste)
head(datumy)
```

```
##      [,1]       [,2]       [,3]       [,4]
## [1,] "Jan 2020" "Jan 2021" "Jan 2022" "Jan 2023"
## [2,] "Feb 2020" "Feb 2021" "Feb 2022" "Feb 2023"
## [3,] "Mar 2020" "Mar 2021" "Mar 2022" "Mar 2023"
```

```
head(c(datumy))
```

```
## [1] "Jan 2020" "Feb 2020" "Mar 2020" "Apr 2020" "May 202
```

```
datumy <- c(datumy)
head(datumy, 30)
```

```
##  [1] "Jan 2020" "Feb 2020" "Mar 2020" "Apr 2020" "May 20
##  [7] "Jul 2020" "Aug 2020" "Sep 2020" "Oct 2020" "Nov 20
## [13] "Jan 2021" "Feb 2021" "Mar 2021" "Apr 2021" "May 20
## [19] "Jul 2021" "Aug 2021" "Sep 2021" "Oct 2021" "Nov 20
## [25] "Jan 2022" "Feb 2022" "Mar 2022" "Apr 2022" "May 20
```

```
grep(pattern = "Jan", x = datumy)
```

```
## [1]  1 13 25 37
```

```r
grep(pattern = as.character("2020"), x = datumy)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12
```

```r
# dva roky -> vystup bude list (vektor pre kazdy rok)
# -> preto `lapply`
lapply(2021:2022,
       function(rok) grep(pattern = as.character(rok),
                          x = datumy))
```

```
## [[1]]
##  [1] 13 14 15 16 17 18 19 20 21 22 23 24
##
## [[2]]
##  [1] 25 26 27 28 29 30 31 32 33 34 35 36
```

```r
# spravime vektor
vyberList <- lapply(2021:2022,
                    function(rok) grep(pattern = as.charact
                                       x = datumy))
vyberVektor <- unlist(vyberList)
vyberVektor
```

```
##  [1] 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
```

```r
head(vyberVektor)
```

```
## [1] 13 14 15 16 17 18
```

```r
tail(vyberVektor)
```

```
## [1] 31 32 33 34 35 36
```

- Takto sa z dát `dataVsetky` dajú vybrať hodnoty zo zvoleného obdobia
- Zmena zvoleného obdobia je len prepísanie hľadaného vektora - môže to byť parameter funkcie, čím sa z výberu dát podľa štátu a obdobia stane jedno zavolanie funkcie