

ARMA modely
část 1: autoregresné modely (AR)

Beáta Stehlíková
Časové rady, FMFI UK

ARMA modely

- Terminológia:
 - ◇ AR - autoregresný model
 - ◇ MA - moving average, kĺzavé priemery
 - ◇ ARMA - ich kombinácia
- Najskôr: autoregresný proces prvého rádu - AR(1)
 - ◇ definícia
 - ◇ stacionarita, ohraničenie na parametre
 - ◇ výpočet momentov a ACF
 - ◇ simulované dáta
 - ◇ praktický príklad s reálnymi dátami
- Potom:
 - ◇ autoregresné procesy vyšších rádov
 - ◇ ako určiť vhodný rád procesu pre dané dáta

I.

Autoregresný proces prvého rádu - AR(1)

AR(1) - definícia

- AR(1) proces:

$$x_t = \delta + \alpha x_{t-1} + u_t,$$

kde δ a α sú konštanty a $\{u_t\}$ je biely šum

- Nech pre $t = t_0$ je daná hodnota x_{t_0} :

$$x_{t_0+1} = \delta + \alpha x_{t_0} + u_{t_0+1},$$

$$x_{t_0+2} = \delta + \alpha x_{t_0+1} + u_{t_0+2} =$$

$$\delta(1 + \alpha) + \alpha^2 x_{t_0} + (\alpha u_{t_0+1} + u_{t_0+2}),$$

$$x_{t_0+3} = \dots$$

vo všeobecnosti:

$$(1) \quad x_{t_0+\tau} = \alpha^\tau x_{t_0} + \frac{1 - \alpha^\tau}{1 - \alpha} \delta + \sum_{j=0}^{\tau-1} \alpha^j u_{t_0+\tau-j}$$

AR(1) - stacionarita

- Zo vzťahu (1):

$$x_t = \alpha^{t-t_0} x_{t_0} + \frac{1 - \alpha^{t-t_0}}{1 - \alpha} \delta + \sum_{j=0}^{t-t_0-1} \alpha^j u_{t-j}$$

- **Deterministické začiatkové podmienky:** hodnota procesu v čase t_0 je $x_0 \rightarrow$ proces nie je stacionárny
- **Náhodné začiatkové podmienky:**
 - ◇ Proces je generovaný pre $t \in \mathbb{R} \rightarrow$ hodnota x_{t_0} je náhodná.
 - ◇ Ak $-1 < \alpha < 1$, tak pre $t_0 \rightarrow -\infty$ dostaneme
(2)
$$x_t = \frac{1}{1 - \alpha} \delta + \sum_{j=0}^{\infty} \alpha^j u_{t-j}$$
 - ◇ Woldova reprezentácia: $\psi_j = \alpha^j$ pre $|\alpha| < 1 \rightarrow$ proces je slabo stacionárny.

AR(1) - momenty

- Pripomeňme si explicitné vyjadrenie procesu (2):

$$x_t = \frac{\delta}{1 - \alpha} + \sum_{j=0}^{\infty} \alpha^j u_{t-j}$$

- Stredná hodnota:

$$\begin{aligned} E[x_t] &= E \left[\frac{\delta}{1 - \alpha} + \sum_{j=0}^{\infty} \alpha^j u_{t-j} \right] \\ &= \frac{\delta}{1 - \alpha} + \sum_{j=0}^{\infty} \alpha^j E[u_{t-j}] = \frac{\delta}{1 - \alpha} \end{aligned}$$

- ◇ $E[x_t] = 0$ práve vtedy, keď $\delta = 0$
- ◇ vo všeobecnosti $E[x_t] \neq \delta$, ale majú rovnaké znamienko (lebo $|\alpha| < 1$)

AR(1) - momenty

- Variancia:

$$\begin{aligned} \text{Var}[x_t] &= \text{Var} \left[\frac{\delta}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j u_{t-j} \right] \\ &= \sum_{j=0}^{\infty} \text{Var}[\alpha^j u_{t-j}] = \sum_{j=0}^{\infty} \alpha^{2j} \text{Var}[u_{t-j}] \\ &= \sigma^2 \sum_{j=0}^{\infty} \alpha^{2j} = \sigma^2 \frac{1}{1-\alpha^2} \end{aligned}$$

kde

- ◇ sme využili, že disperzia súčtu nekorelovaných premenných je súčet disperzií
- ◇ σ^2 je variancia bieleho šumu $\{u_j\}$

AR(1) - momenty

- **Autokovariancie** (využijeme, že $Cov[u_k, u_l] = \sigma^2$ pre $k = l$ a $Cov[u_k, u_l] = 0$ pre $k \neq l$):

$$\begin{aligned} Cov[x_t, x_{t-s}] &= E \left[\left(\sum_{i=0}^{\infty} \alpha^i u_{t-i} \right) \left(\sum_{j=0}^{\infty} \alpha^j u_{t-s-j} \right) \right] \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \alpha^{i+j} E[u_{t-i} u_{t-s-j}] \\ &= \sigma^2 \sum_{j=0}^{\infty} \alpha^{s+2j} = \alpha^s \frac{\sigma^2}{1 - \alpha^2} \end{aligned}$$

- **Autokorelácie :**

$$Cor[x_t, x_{t-s}] = \frac{Cor[x_t, x_{t-s}]}{Var[x_t]Var[x_{t-s}]} = \alpha^s$$

Príklad - simulované dáta

- AR(1) proces

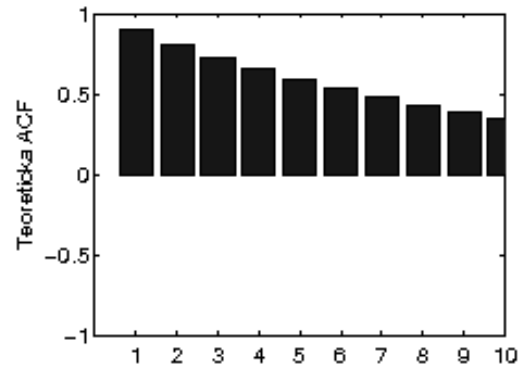
$$x_t = \delta + \alpha x_{t-1} + u_t,$$

kde biely šum u_t má normálne rozdelenie, $\delta = 0$,
 $\sigma^2 = 1$

- Postupne zoberieme $\alpha = \{0.9, 0.6, -0.9\}$
- Zobrazíme:
 - ◇ teoretickú autokorelačnú funkciu
 - ◇ realizáciu procesu
 - ◇ výberovú autokorelačnú funkciu odhadnutú z dát

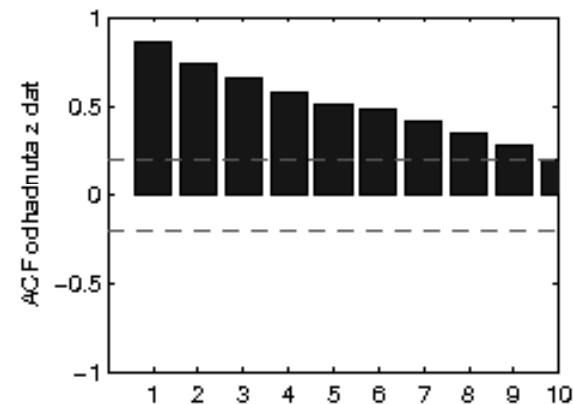
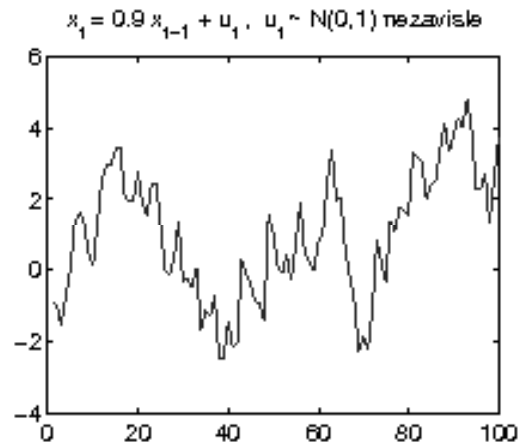
Príklad - simulované dáta, $\alpha = 0.9$

- Teoretická autokorelačná funkcia:



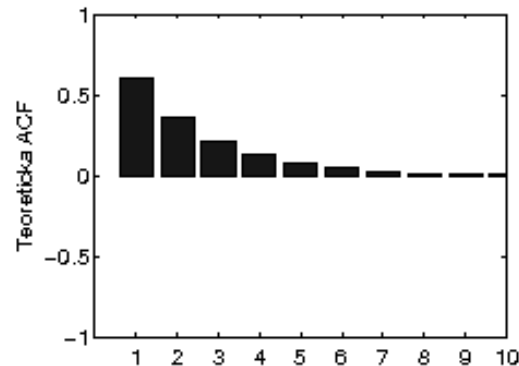
Kontrolná otázka: Čomu sa rovnajú jej hodnoty?

- Vygenerovaný proces a jeho výberová ACF:



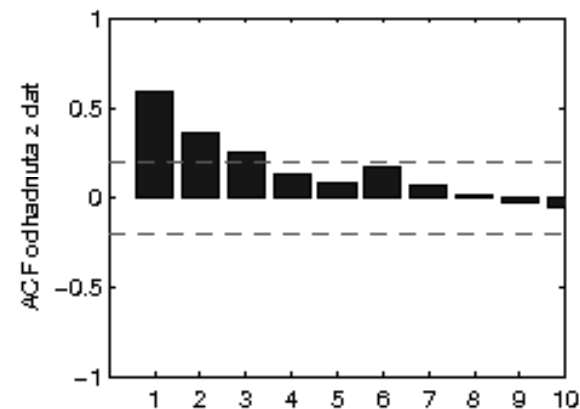
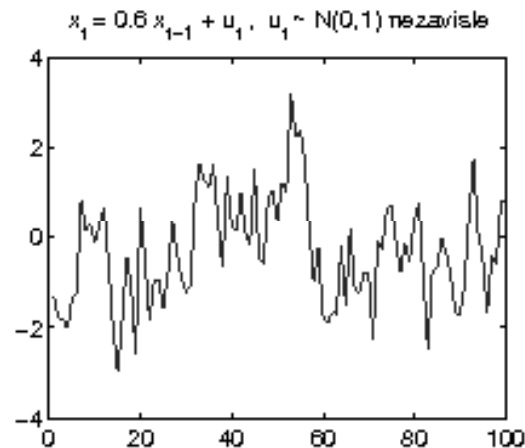
Príklad - simulované dáta, $\alpha = 0.6$

- Teoretická autokorelačná funkcia:



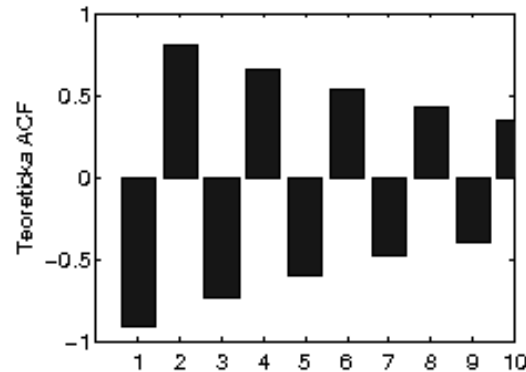
Kontrolná otázka: Čomu sa rovnajú jej hodnoty?

- Vygenerovaný proces a jeho výberová ACF:



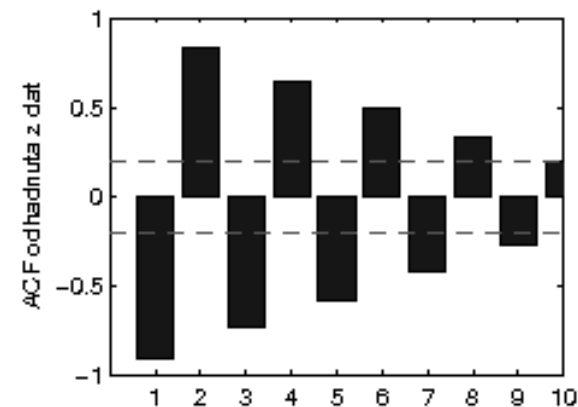
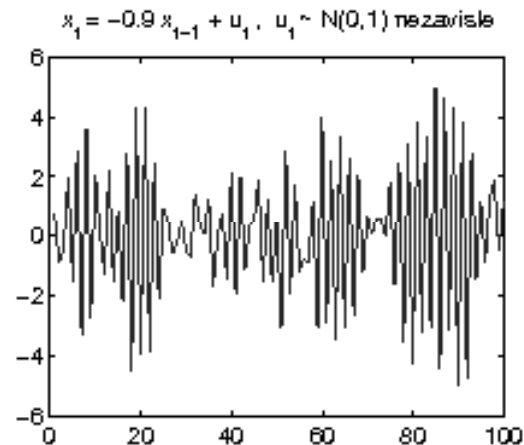
Príklad - simulované dáta, $\alpha = -0.9$

- Teoretická autokorelačná funkcia:



Kontrolná otázka: Čomu sa rovnajú jej hodnoty?

- Vygenerovaný proces a jeho výberová ACF:



Príklad - reálne dáta

- G. Kirchgässner: **Causality Testing of the Popularity Function: An Empirical Investigation for the Federal Republic of Germany, 1971-1982**, Public Choice 45 (1985), p. 155-173.
- [Kirchgässner, Wolters], example 2.2
- Nemecko, január 1971 - apríl 1982
- CDU_t = volebné preferencie CDU/CSU



a) Popularity of the CDU/CSU, 1971 – 1982

Príklad - reálne dáta

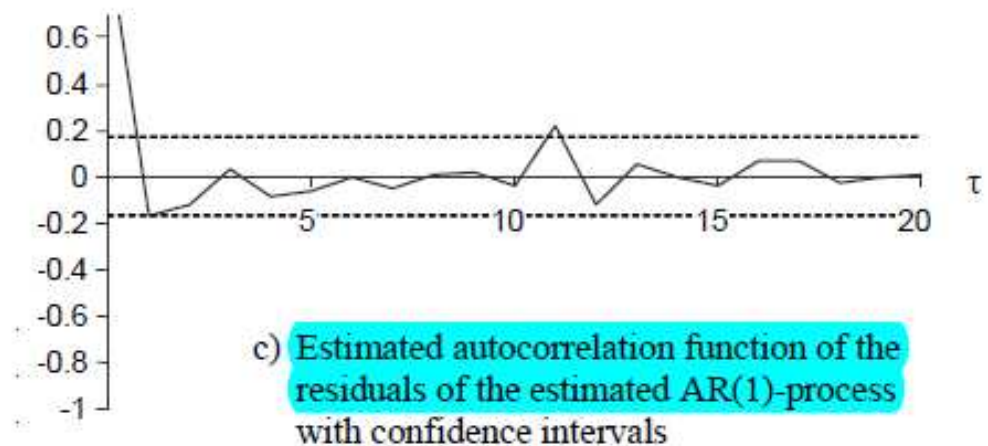
- Odhadnutý AR(1) model:

$$\text{CDU}_t = 8.053 + 0.834 \text{CDU}_{t-1} + \hat{u}_t$$

(3.43) (17.10)

$$\bar{R}^2 = 0.683, \text{ SE} = 1.586, \text{ Q}(11) = 12.516 \text{ (p} = 0.326\text{)}.$$

The estimated t values are given in parentheses, SE denotes the standard error of the residuals. The autocorrelogram, which is also given in *Figure 2.4*, does not indicate any higher-order process. Moreover, given the high p-value, the Ljung-Box Q statistic with 12 correlation coefficients (i.e. with 11 degrees of freedom) gives no reason to reject this model.



Príklad - reálne dáta

- Je odhadnutý model stacionárny? Z čoho to vyplýva?
- Rezíduá modelu by mali byť bielym šumom:
 - ◇ Na grafe sú pri autokoreláciách zostrojené intervaly. Na čo slúžia? Vypočítajte pomocou známych údajov ich hranice.
 - ◇ V texte sa spomínajú autokorelácie rezíduí a Ljung-Boxova Q štatistika - aké hypotézy sa testujú (a prečo), akým spôsobom a s akými závermi?
- Čomu sa rovná stredná hodnota premennej CDU_t ?

Predikcie

- Proces je stacionárny → má konštantnú strednú hodnotu
- Má však zmysel počítat' aj podmienenú strednú hodnotu
- V predchádzajúcom príklade:
 - ◇ Máme stacionárny proces ako model pre volebné preferencie
 - ◇ Našli sme nepodmienenú strednú hodnotu procesu - je konštantná
 - ◇ Podmienená stredná hodnota - napr.: *Aká je očakávaná hodnota preferencií nasledujúcim mesiaci, ak dnešné preferencie sú 40 percent? Čo ak sú dnešné preferencie 55 percent?* - odpovede budú rôzne

Predikcie v AR(1) modeli

- Intuitívne (presnejšie pri zložitejších modeloch, kde postup nebude zrejmý)
- Pre $x_t := CDU_t$ máme model

$$x_t = 8.053 + 0.834x_{t-1} + u_t$$

- Biely šum u_t nahradíme jeho strednou hodnotou - nulou
- Za x_{t-1} dosadíme
 - ◇ skutočnú hodnotu x_{t-1} , ak ju máme k dispozícii
 - ◇ predikciu hodnoty x_{t-1} , ak sa ešte nerealizovala

Predikcie v AR(1) modeli

- Praktický výpočet:

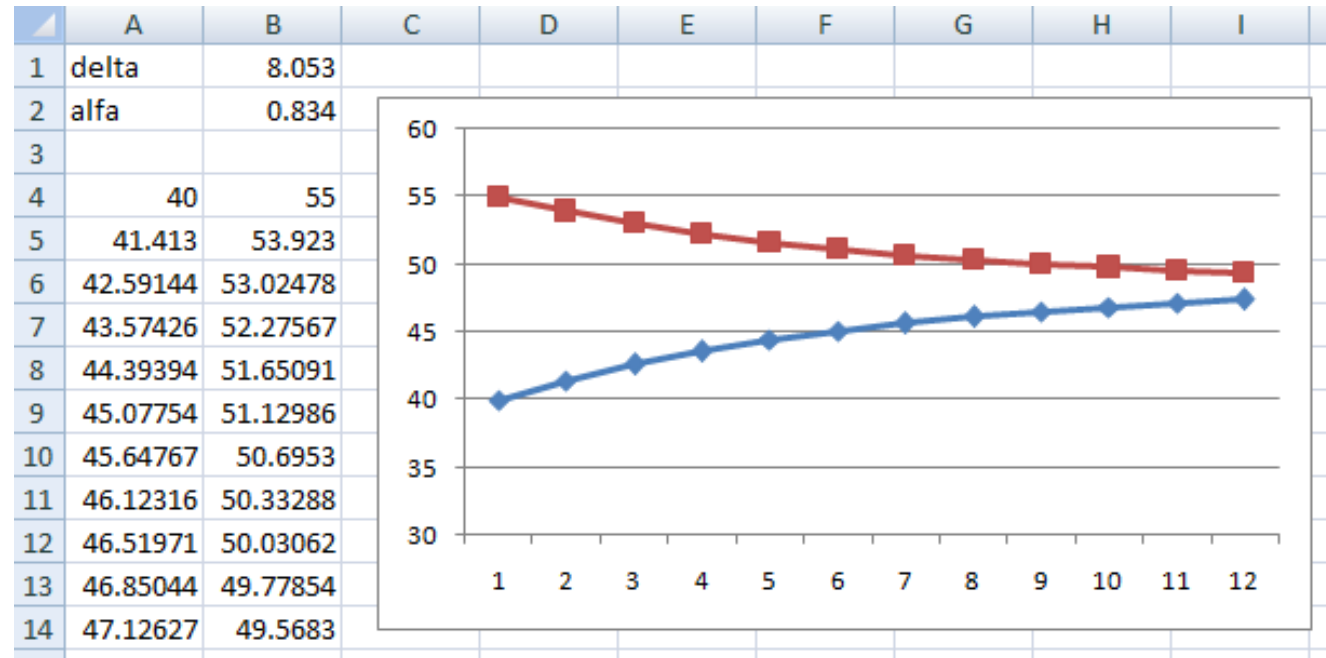
	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	
5	=B\$1+B\$2*A4	
6		

	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	
5	41.413	
6	=B\$1+B\$2*A5	
7		

	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	
5	41.413	
6	42.59144	
7	43.57426	
8	44.39394	
9	45.07754	
10	45.64767	
11	46.12316	
12	46.51971	
13	46.85044	
14	47.12627	
15	=B\$1+B\$2*A14	
16		

Predikcie v AR(1) modeli

- Predikcie pre dve rôzne začiatočné hodnoty:



- K akej spoločnej hodnote tieto predikcie konvergujú?

Motivácia k zložitejším modelom

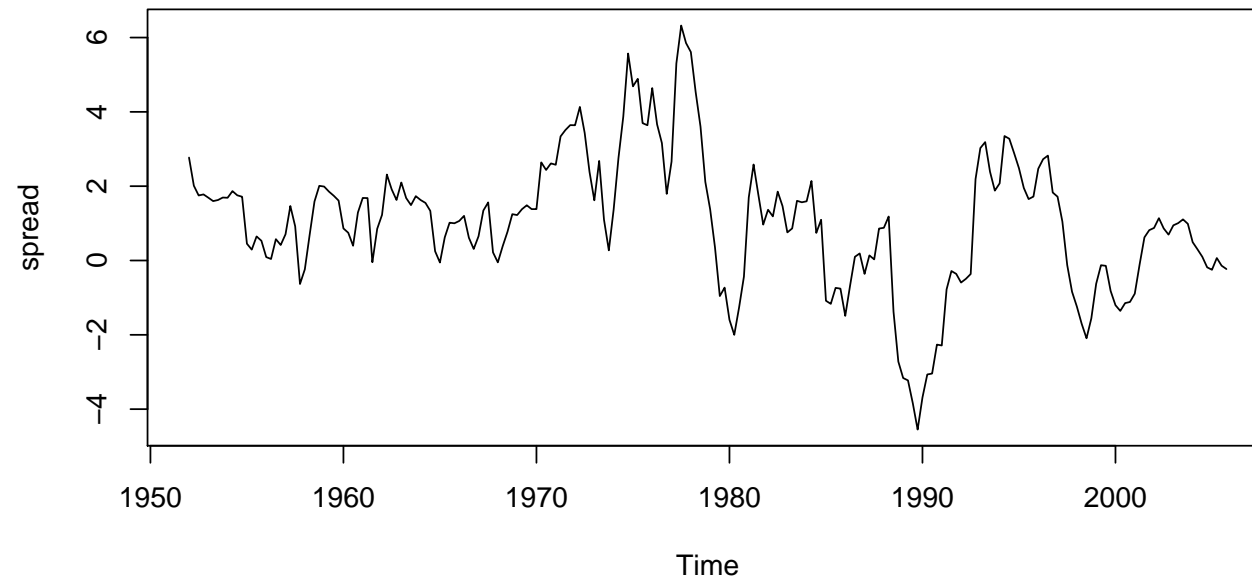
Mills, Markellos: **The Econometric Modelling of Financial Time Series**. Cambridge University Press, 2008

Dáta: <http://www.lboro.ac.uk/departments/ec/cup/data.html>

- Štvrt' ročné dáta, 1952Q1 - 2005Q4
- Premenné:
 - ◇ krátkodobá úroková miera (3 mesiace)
 - ◇ dlhodobá úroková miera (20 rokov)
- Budeme modelovať rozdiel dlhodobej a krátkodobej úrokovej miery

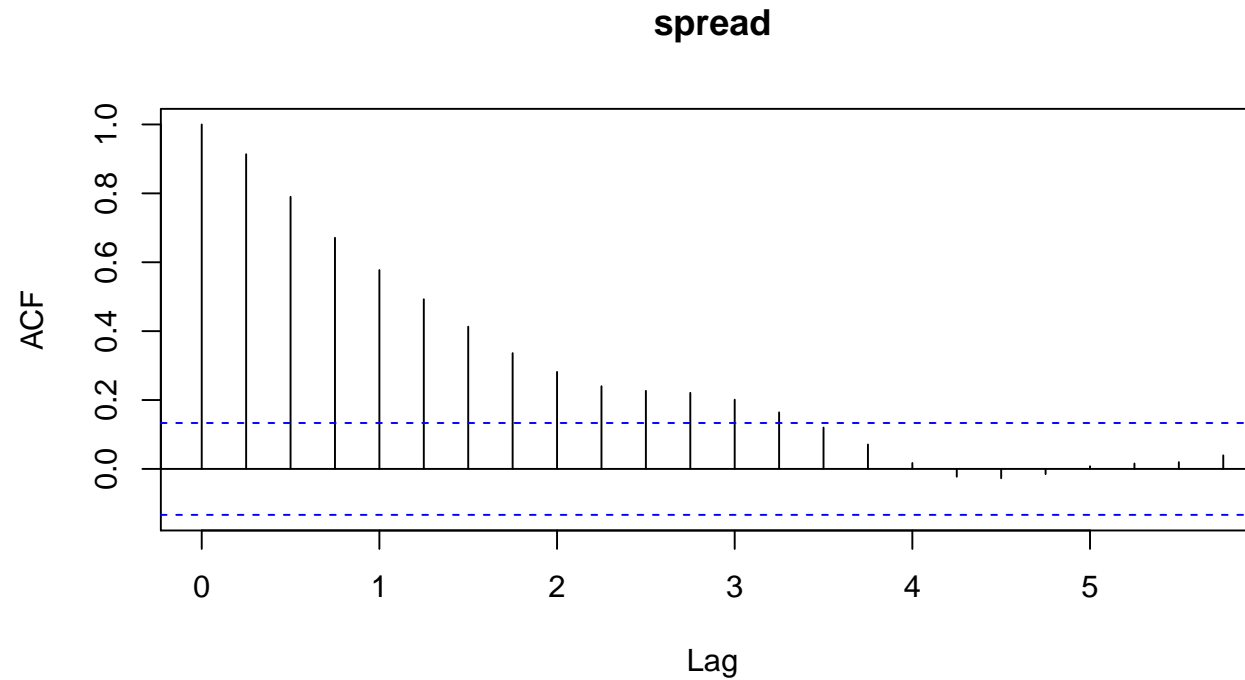
Motivácia k zložitejším modelom

- Priebek modelovanej premennej:



Motivácia k zložitejším modelom

- Odhadnutá autokoreačná funkcia:



Motivácia k zložitejším modelom

- V R-ku budeme na odhadovanie modelov používať balík **astsa: Applied Statistical Time Series Analysis**
- Odhadneme AR(1) model:

```
sarima(spread,1,0,0,details="FALSE")
```

- Aby bol proces stacionárny, autoregresný koeficient musí byť v absolútnej hodnote menší ako 1.
- **AR** v názve **SARIMA** sa vzťahuje na autoregresné členy
- SARIMA označuje všeobecnejšie modely, budeme sa nimi postupne zaoberať neskôr

Motivácia k zložitejším modelom

```
> sarima(spread,1,0,0,details="FALSE")
```

```
$fit
```

```
Call:
```

```
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,  
  Q), period = S), xreg = xmean, include.mean = FALSE, optim.control = list(t$  
  REPORT = 1, reltol = tol))
```

```
Coefficients:
```

```
      ar1    xmean  
0.9156  1.0473  
s.e.  0.0266  0.5491
```

```
sigma^2 estimated as 0.5106:  log likelihood = -234.8,  aic = 475.61
```

```
$AIC
```

```
[1] 0.3463184
```

```
$AICc
```

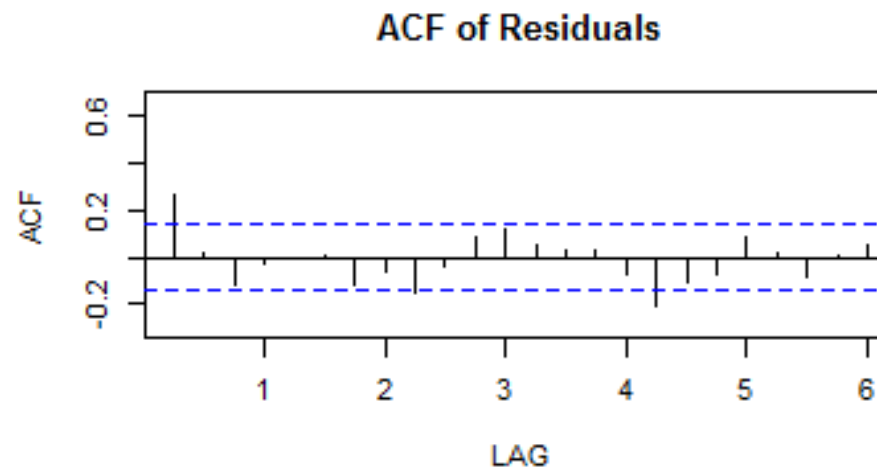
```
[1] 0.3561018
```

```
$BIC
```

```
[1] -0.622429
```


Motivácia k zložitejším modelom

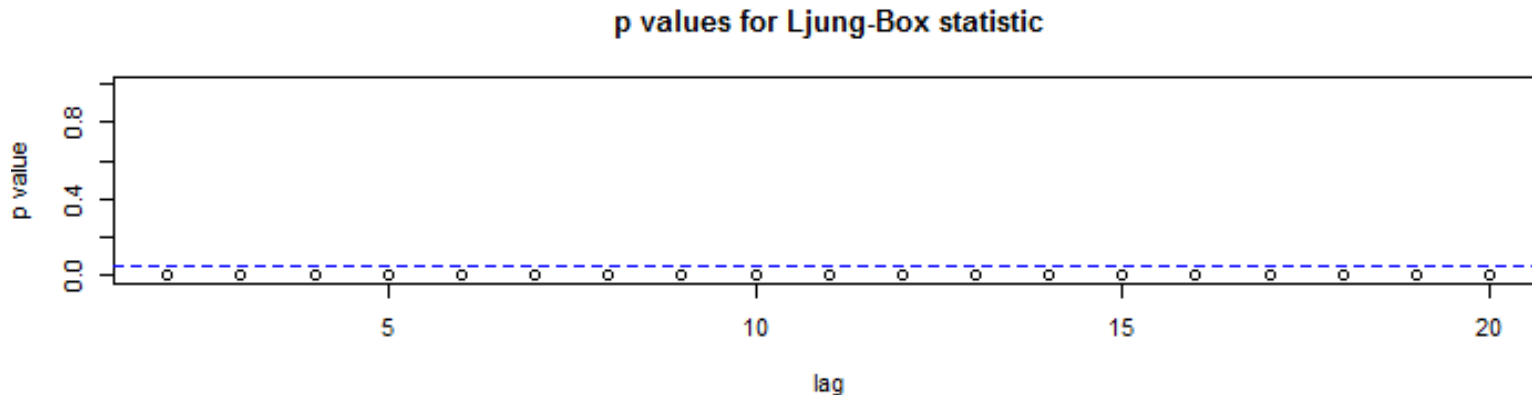
- Kontrola rezíduí - autokorelačná funkcia:



- Na zopakovanie:
 - ◇ Akú hypotézu testujeme?
 - ◇ Na čo slúžia vyznačené intervaly a ako sa konštruujú?
 - ◇ Aký záver z tohto grafu vypláva?

Motivácia k zložitejším modelom

- Kontrola rezíduí - P hodnoty Ljung-Boxovej štatistiky:



- Ide o rezíduá z AR(1) modelu, počet stupňov voľnosti sa znižuje o 1.
- Na zopakovanie:
 - ◇ Akú hypotézu testujeme? Aký záver vypláva z grafu?
 - ◇ Ako sa počíta testovacia štatistika a aké je jej pravdepodobnostné rozdelenie za platnosti nulovej hypotézy?

II.

Autoregresný proces druhého rádu - AR(2)

Predch. príklad - modelovanie spreadu

- Zistili sme, že AR(1) proces

$$x_t = \delta + \alpha x_{t-1} + u_t,$$

nevyhovuje (rezíduá odhadnutého modelu nie sú biely šum).

- Skúsme okrem závislosti od x_{t-1} pridať aj závislosť od x_{t-2} :

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$$

- Takýto proces sa nazýva **autoregresný proces druhého rádu**
- Analogicky **autoregresný proces p -teho rádu**:

$$x_t = \delta + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + u_t$$

- Najskôr sa budeme zaoberať AR(2) procesom

AR(2) - definícia

- AR(2) proces:

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$$

- Aj bez u_t komplikovanejšie ako AR(1) - korene charakteristického polynómu
- Skúsime iný postup ako dosadzovanie
- Pomocou operátora posunu:

$$(1 - \alpha_1 L - \alpha_2 L^2)x_t = \delta + u_t$$

$$\alpha(L)x_t = \delta + u_t$$

- Woldova reprezentácia a stacionarita:

$$x_t = \alpha^{-1}(L)\delta + \alpha^{-1}(L)u_t$$

→ potrebujeme inverzný operátor $\alpha^{-1}(L)$

AR(2) - definícia

- Inverzný operátor $\alpha^{-1}(L)$; nájdeme metódou neurčitých koeficientov:

$$\alpha^{-1}(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots$$

pričom

$$(3) \quad 1 = (1 - \alpha_1 L - \alpha_2 L^2)(\psi_0 + \psi_1 L + \psi_2 L^2 + \dots)$$

- Porovnáme koeficienty pri L^j na oboch stranách (3):

$$\psi_j - \alpha_1 \psi_{j-1} - \alpha_2 \psi_{j-2} = 0,$$

$$\psi_0 = 1, \quad \psi_1 = \alpha_1$$

AR(2) - stacionarita

- **Podmienka stacionarity:** Kvôli splneniu podmienky $\sum \psi_j^2 < \infty$ musia byť korene charakteristickej rovnice

$$\lambda^2 - \alpha_1 \lambda - \alpha_2 = 0$$

v absolútnej hodnote menšie ako 1

- Inak povedané: **korene rovnice**

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 = 0$$

musia byť v absolútnej hodnote väčšie ako 1, t. j. **mimo jednotkového kruhu**

- To isté vyšlo predtým pre AR(1): koreň $\alpha(L) = 0$ mimo jednotkového kruhu

Príklad - modelovanie spreadu

Odhadnutý AR(2) model:

```
> sarima(spread, 2, 0, 0, details="FALSE")
```

```
$fit
```

```
Call:
```

```
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), xreg = xmean, include.mean = FALSE, optim.control = list(t$REPORT = 1, reltol = tol))
```

```
Coefficients:
```

	ar1	ar2	xmean
	1.1809	-0.2886	1.0449
s.e.	0.0650	0.0651	0.4212

```
sigma^2 estimated as 0.4677: log likelihood = -225.42, aic = 458.84
```

```
$AIC
```

```
[1] 0.2678181
```

```
$AICc
```

```
[1] 0.277955
```

```
$BIC
```

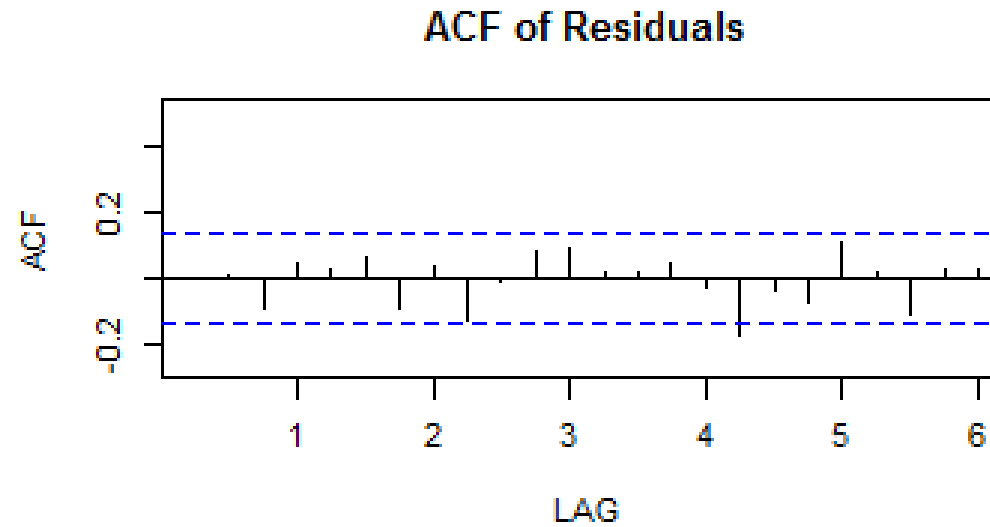
```
[1] -0.6853031
```


Príklad - modelovanie spreadu

- Ukážte, že získaný proces je stacionárny.
- Čo testujeme pri rezíduách - aké hypotézy a pomocou akých štatistík?
- Čo - na základe výstupov na nasledujúcich slajdoch - dostávame v tomto prípade?

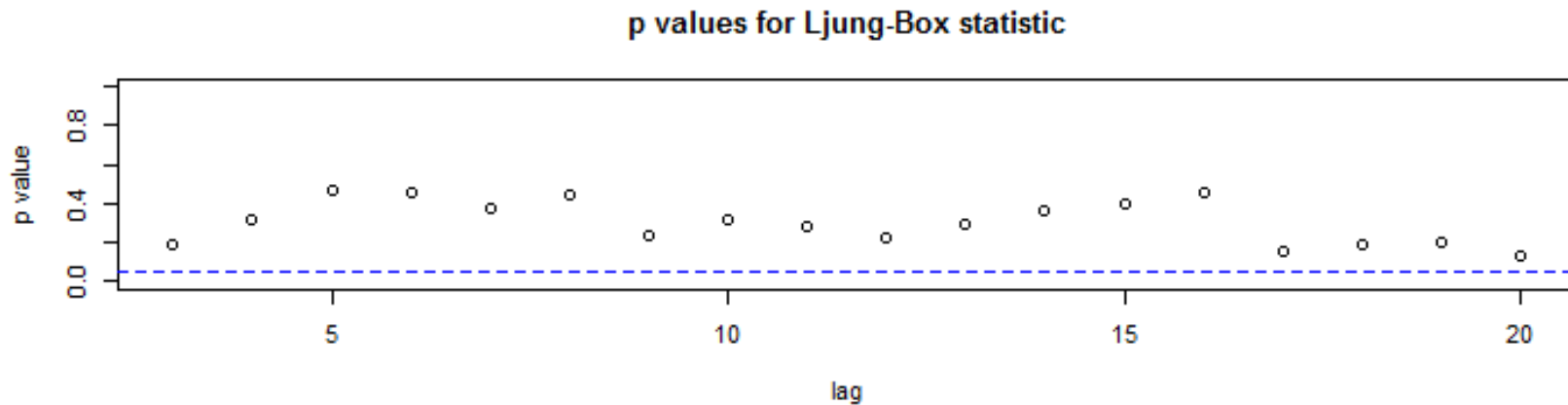
Príklad - modelovanie spreadu

- ACF rezíduí:



Príklad - modelovanie spreadu

- P hodnoty Ljung-Boxovej štatistiky



Pre rezíduá z AR(p) modelu sa počet stupňov voľnosti znižuje o p .

AR(2) - momenty

- Slabo stacionárny AR(2) proces:

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$$

- Stredná hodnota:

- ◇ označme $\mu = E[x_i]$; potom

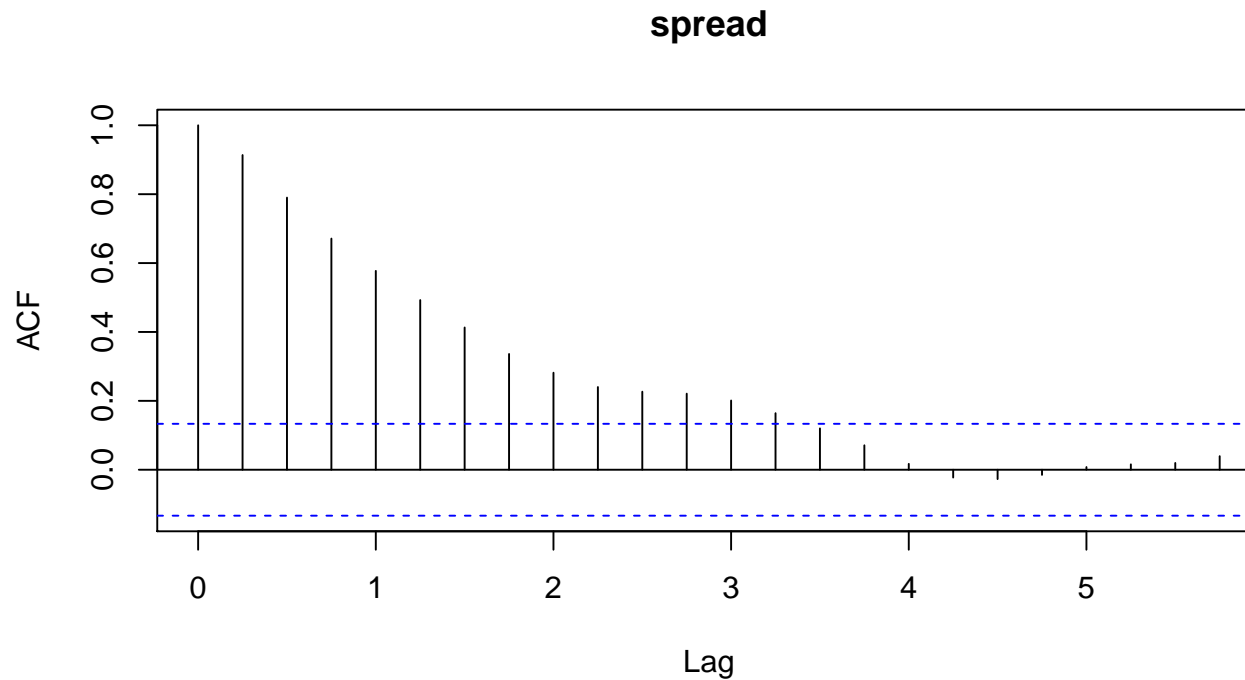
$$\mu = \delta + \alpha_1 \mu + \alpha_2 \mu,$$

$$\mu = \frac{\delta}{1 - \alpha_1 - \alpha_2}$$

- ◇ Podobne ako pre AR(1): stredná hodnota sa nerovná parametru δ (okrem prípadu $\delta = 0$), ale majú rovnaké znamienko

AR(2) - momenty

- Autokovariancie AR(2) procesu - motivácia :
 - ◇ pre pripomenutie - výberová ACF pre spread:



AR(2) - momenty

- Autokovariancie AR(2) procesu - motivácia :
 - ◇ výberová ACF pre spread sa podobala na AR(1) proces
 - ◇ napriek tomu AR(1) proces nebol dobrým modelom, ale AR(2) áno
 - ◇ aký priebeh má ACF pre AR(2) proces?
 - ◇ môže mať podobný priebeh ako AR(1)? (zdá sa že áno)
 - ◇ môže mať aj "úplne iný" priebeh? (t.j. "toto určite nie je AR(1), ale AR(2) to môže byť")

AR(2) - momenty

- Autokovariancie - výpočet : znovu môžeme predpokladať nulovú strednú hodnotu, t. j.

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t \quad / \quad \times x_{t-s}, E[.]$$

$$E[x_{t-s}x_t] = \alpha_1 E[x_{t-s}x_{t-1}] + \alpha_2 E[x_{t-s}x_{t-2}] + E[x_{t-s}u_t]$$

- Pre $s = 0, 1, 2$ dostaneme:

$$\gamma(0) = \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \sigma^2$$

$$\gamma(1) = \alpha_1 \gamma(0) + \alpha_2 \gamma(1)$$

$$\gamma(2) = \alpha_1 \gamma(1) + \alpha_2 \gamma(0)$$

- sústava rovníc $\rightarrow \gamma(0) = Var[x_t], \gamma(1), \gamma(2)$

- Pre $s \geq 2$ - diferenčná rovnica:

$$(4) \quad \gamma(s) - \alpha_1 \gamma(s-1) - \alpha_2 \gamma(s-2) = 0,$$

začiatkové podmienky z predchádzajúceho bodu

AR(2) - momenty

- **Autokorelácie:** diferenčnú rovnicu (4) a jej začiatočné podmienky vydelíme $\gamma(0)$:

$$\rho(s) - \alpha_1\rho(s-1) - \alpha_2\rho(s-2) = 0$$

$$\rho(0) = 1, \rho(1) = \frac{\alpha_1}{1 - \alpha_2}$$

AR(2) - ACF - príklad 1

- Spread modelovaný AR(2) procesom:

```
Coefficients:
          ar1      ar2      xmean
          1.1809 -0.2886  1.0449
s.e.      0.0650   0.0651  0.4212
```

- Diferenčná rovnica pre autokorelácie:

$$\rho(s) - 1.1809\rho(s - 1) + 0.2886\rho(s - 2) = 0$$

začiatkové podmienky: $\rho(0) = 1$, $\rho(1) = \frac{1.1809}{1+0.2886}$

AR(2) - priebeh ACF

- ACF je riešením diferenčnej rovnice

$$\rho(s) - \alpha_1 \rho(s-1) - \alpha_2 \rho(s-2) = 0$$

⇒ priebeh závisí od **koreňov** charakteristickej rovnice

$$\lambda^2 - \alpha_1 \lambda - \alpha_2 = 0$$

- λ_1, λ_2 - **reálne** (a rôzne): ACF má tvar

$$\rho(s) = c_1 \lambda_1^s + c_2 \lambda_2^s$$

Zo stacionarity: $|\lambda_{1,2}| < 1$

- λ_1, λ_2 - **komplexné**: ACF je tlmená kombinácia sínusu a kosínusu

$$\rho(s) = r^s (c_1 \cos(ks) + c_2 \sin(ks))$$

Zo stacionarity: $r < 1$

AR(2) - ACF - príklad

- Proces: $x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$

- ◇ korelácie splňajú diferenčnú rovnicu

$$\rho(t) - 1.4\rho(t-1) + 0.85\rho(t-2) = 0$$

- ◇ jej všeobecné riešenie

$$\rho(t) = 0.922^t (c_1 \cos(0.709t) + c_2 \sin(0.709t))$$

- ◇ c_1, c_2 zo začiatkových podmienok $\rho(0), \rho(1)$

- ◇ $\cos(kt), \sin(kt) \rightarrow$ perióda $\frac{2\pi}{k}$

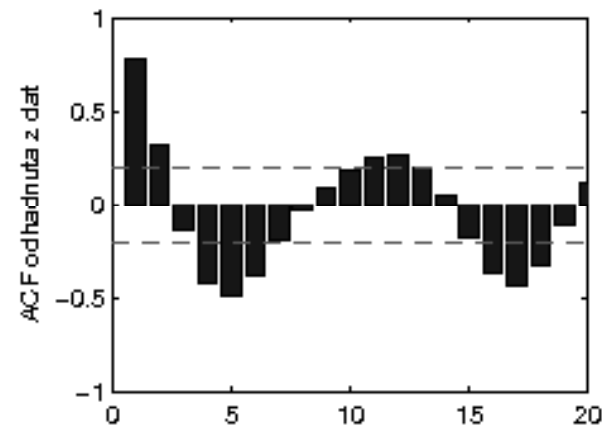
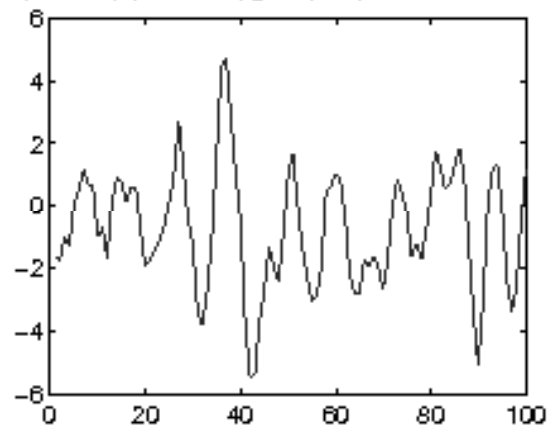
v našom prípade $\frac{2\pi}{k} = \frac{2\pi}{0.709} = 8.862 \approx 9$

\Rightarrow v dátach generovaných týmto procesom sa dá očakávať takáto perióda

AR(2) - ACF - príklad

- Na obrázku:
 - ◇ realizácia procesu $x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$
 - ◇ výberová ACF

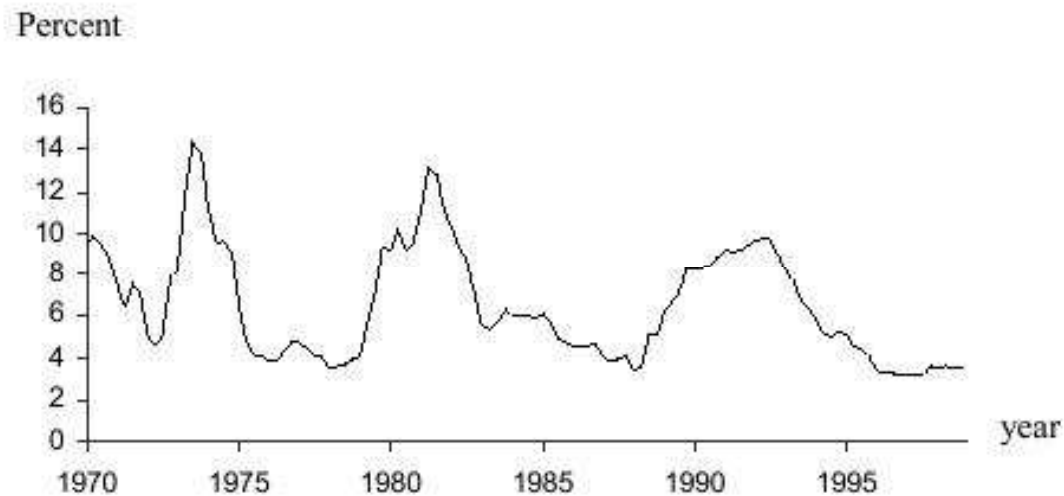
$$x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t, \quad u_t \sim N(0,1) \text{ nezávisle}$$



AR(2) - reálne dáta

[Kirchgässner, Wolters], example 2.6

- 3-mesačná úroková miera, Nemecko, 1970q1-1998q4



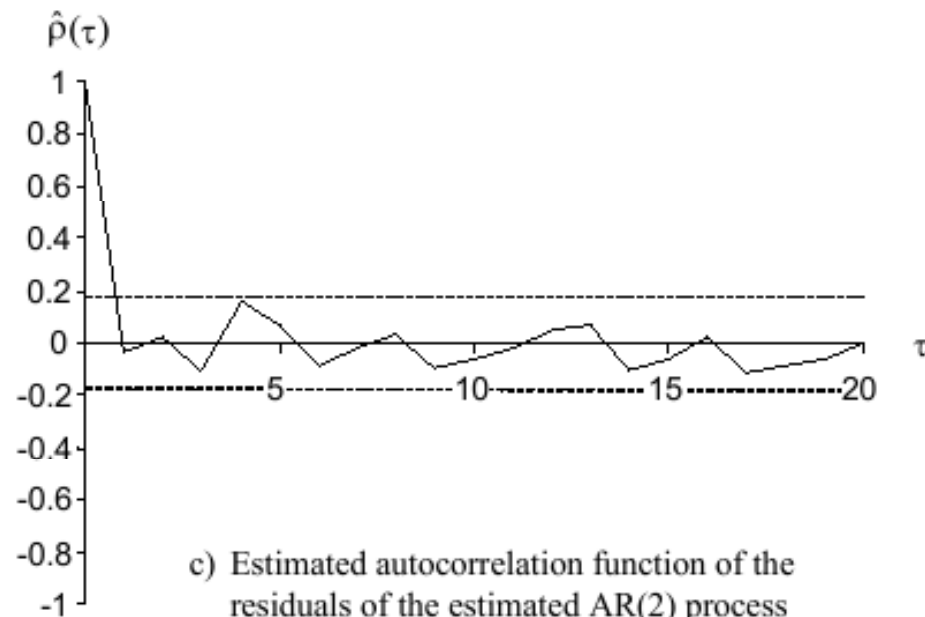
AR(2) - reálne dáta

- Odhadnutý AR(2) model:

$$\text{GSR}_t = 0.577 + 1.407 \text{GSR}_{t-1} - 0.498 \text{GSR}_{t-2} + \hat{u}_t,$$

(2.82) (17.49) (-6.16)

$$\bar{R}^2 = 0.910, \text{ SE} = 0.812, Q(6) = 6.431 \text{ (p} = 0.377\text{)}$$



c) Estimated autocorrelation function of the residuals of the estimated AR(2) process with confidence intervals

AR(2) - reálne dáta

- Otázky k odhadnutému modelu:
 - ◇ Je stacionárny?
 - ◇ Analyzujte rezíduá - autokorelogram, Q-štatistika (aký je počet stupňov voľnosti?).
 - ◇ Aká je stredná hodnota odhadnutého procesu?
 - ◇ Aký je priebeh jeho autokorelačnej funkcie?
 - ◇ Vysvetlite nasledovné tvrdenie z knihy (str.49) a vypočítajte uvedené hodnoty:
"The two roots of the process are $0.70 \pm 0.06i$, i.e. they indicate cycles ... the frequency $f = 0.079$ corresponds to a period of 79.7 quarters and therefore of nearly 20 years."

III.

Autoregresný proces p -teho rádu - $AR(p)$

$AR(p)$ - úvod

- Videli sme $AR(1)$ a $AR(2)$ proces, ich ACF môže byť podobná - ako ich rozlíšiť?
- Analogicky sa dá definovať $AR(p)$ proces - ako vyzerá jeho ACF?
- Ako určiť správny rád?
- $AR(p)$ proces - ukážeme:
 - ◇ **stacionarita**: korene mimo jednotkového kruhu
 - ◇ **ACF**: daná diferenčnou rovnicou p -teho rádu
 - ◇ **prvých p autokorelácií** (začiatkové podm. diferenčnej rovnice): zo sústavy lineárnych rovníc; užitočný postup, ešte ho budeme potrebovať

AR(p) proces - stacionarita

- AR(p) proces:

$$(5) \quad x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + u_t,$$

t. j. $\alpha(L)x_t = \delta + u_t$, kde $\alpha(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$

- Woldova reprezentácia a stacionarita:

$$x_t = \alpha(L)^{-1}(\delta + u_t),$$

inverzný operátor $\alpha(L)^{-1}$ hľadáme v tvare

$$\alpha(L)^{-1} = 1 + \psi_1 L + \psi_2 L^2 + \dots$$

- Pre koeficienty ψ_j dostaneme diferenčnú rovnicu

$$\psi_k - \alpha_1 \psi_{k-1} - \dots - \alpha_p \psi_{k-p} = 0$$

\Rightarrow kvôli konv. $\sum \psi_j^2$ musia byť korene charakt. rovnice

$\lambda^k - \alpha_1 \lambda^{k-1} - \dots - \alpha_p = 0$ vnútri jednotkového kruhu,

t. j. korene $\alpha(L) = 0$ musia byť mimo jednotkového kruhu

AR(p) proces - momenty

- **Stredná hodnota:**

označíme $\mu = E[x_t]$ a spravíme strednú hodnotu z ľavej aj pravej strany (5):

$$\mu = \delta + \alpha_1\mu + \dots + \alpha_p\mu \Rightarrow \mu = \frac{\delta}{1 - \alpha_1 - \dots - \alpha_p}$$

aj teraz: stredná hodnota má rovnaké znamienko ako parameter δ

- **Variancia, autokovariancie** - nech $\delta = 0$

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + u_t \quad / \quad \times x_{t-s}, E[.]$$
$$\gamma(s) = \alpha_1 \gamma(s-1) + \dots + \alpha_p \gamma(s-p) + E[u_t x_{t-s}]$$

AR(p) proces - momenty

- Variancia, autokovariancie - pokračovanie:
 - ◇ $s = 0, 1, \dots, p \rightarrow$ sústava $p + 1$ rovníc s neznámymi $\gamma(0), \gamma(1), \dots, \gamma(p)$:

$$\gamma(0) = \alpha_1\gamma(1) + \alpha_2\gamma(2) + \dots + \alpha_p\gamma(p) + \sigma^2$$

$$\gamma(1) = \alpha_1\gamma(0) + \alpha_2\gamma(1) + \dots + \alpha_p\gamma(p-1)$$

...

$$\gamma(p) = \alpha_1\gamma(p-1) + \alpha_2\gamma(p-2) + \dots + \alpha_p\gamma(0)$$

(6)

- ◇ ostatné autokovariancie z diferenčnej rovnice

$$(7) \quad \gamma(t) - \alpha_1\gamma(t-1) - \dots - \alpha_p\gamma(t-p) = 0$$

AR(p) proces - momenty

- **ACF** :

- ◇ diferenčná rovnica pre autokorelácie - rovnicu (7) vydelíme disperziou $\gamma(0)$:

$$\rho(t) - \alpha_1\rho(t-1) - \dots - \alpha_p\rho(t-p) = 0$$

- ◇ začiatočné podmienky - posledných p rovníc zo sústavy (6) vydelíme $\gamma(0)$:

$$\rho(1) = \alpha_1 + \alpha_2\rho(1) + \dots + \alpha_p\rho(p-1)$$

$$\rho(2) = \alpha_1\rho(1) + \alpha_2 + \dots + \alpha_p\rho(p-2)$$

...

$$\rho(p) = \alpha_1\rho(p-1) + \alpha_2\rho(p-2) + \dots + \alpha_p$$

(8)

- nazývajú sa **Yule-Wolkerove rovnice**

AR(p) proces - ACF - príklad 1

- Výpočet ACF v softvéri R:
 - ◇ funkcia **ARMAacf** z balíka **stats**
 - ◇ počítali sme ACF procesu

$$x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$$

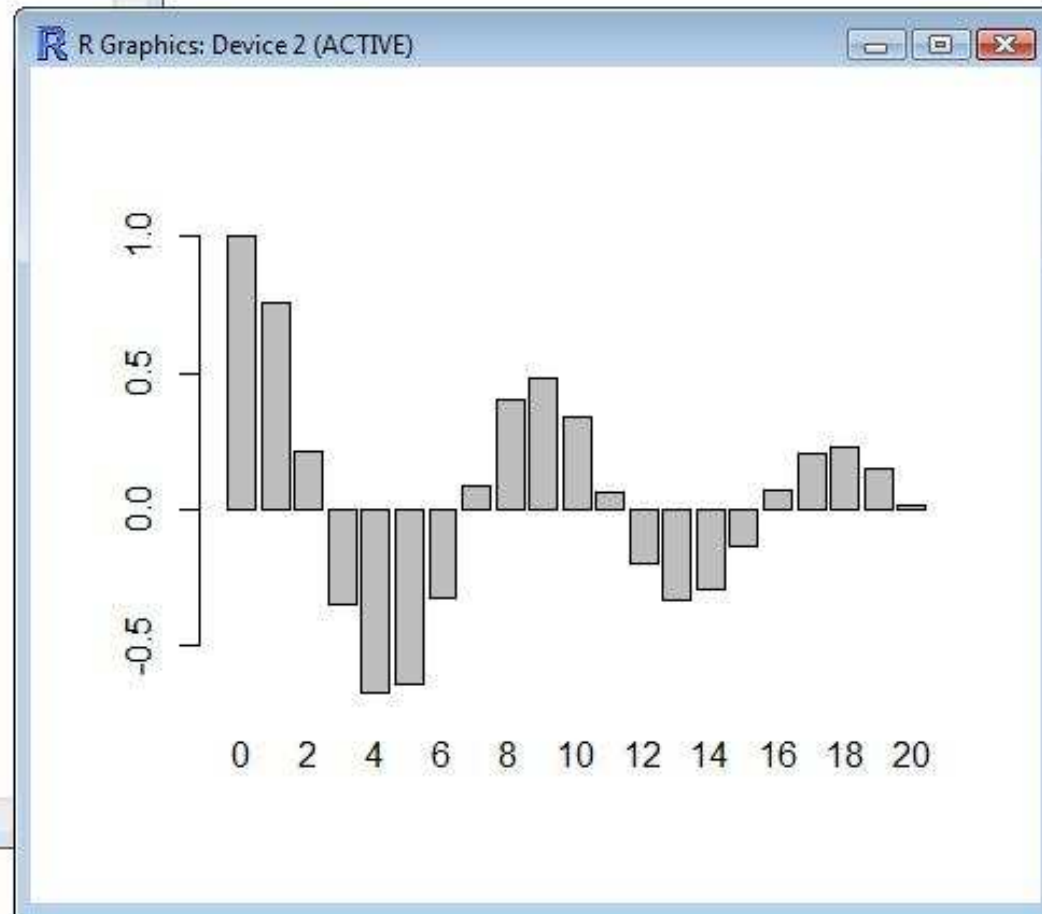
- ◇ teraz v softvéri R:

ARMAacf(ar=c(1.4,-0.85), lag.max=20)

- Alternatíva: funkcia **armaTrueacf** z balíka **fArma**

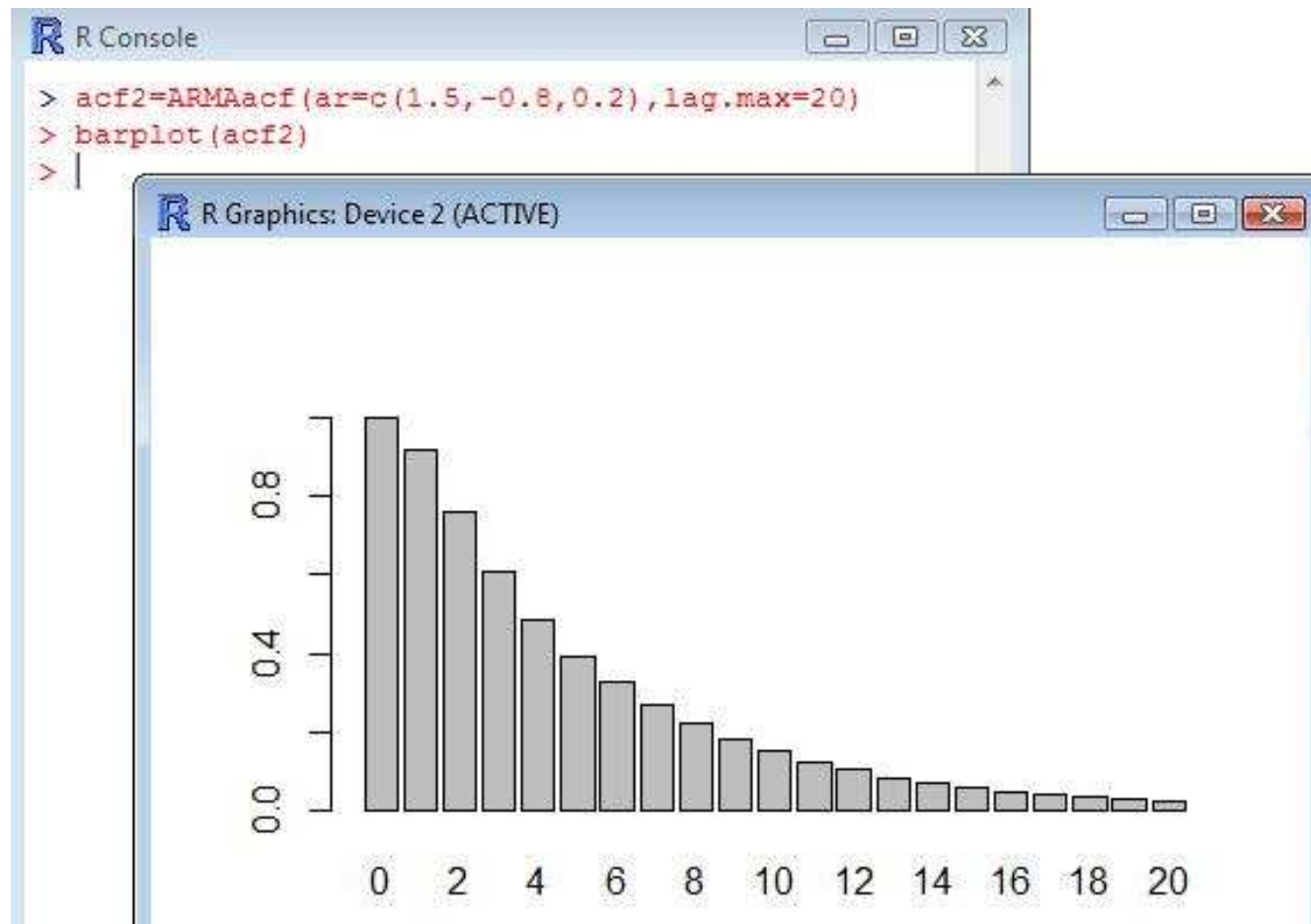
AR(p) proces - ACF - príklad 1

```
R Console
> acf1=ARMAacf(ar=c(1.4,-0.85),lag.max=20)
> acf1
      0      1      2
1.0000000 0.75675676 0.20945946
      3      4      5
-0.3500000 -0.66804054 -0.63775676
      6      7      8
-0.32502500 0.08705824 0.39815279
      9     10     11
 0.48341440 0.33835029 0.06278816
     12     13     14
-0.19969432 -0.33294198 -0.29637861
     15     16     17
-0.13192936 0.06722071 0.20624895
     18     19     20
 0.23161093 0.14894369 0.01165188
> barplot(acf1)
> |
```



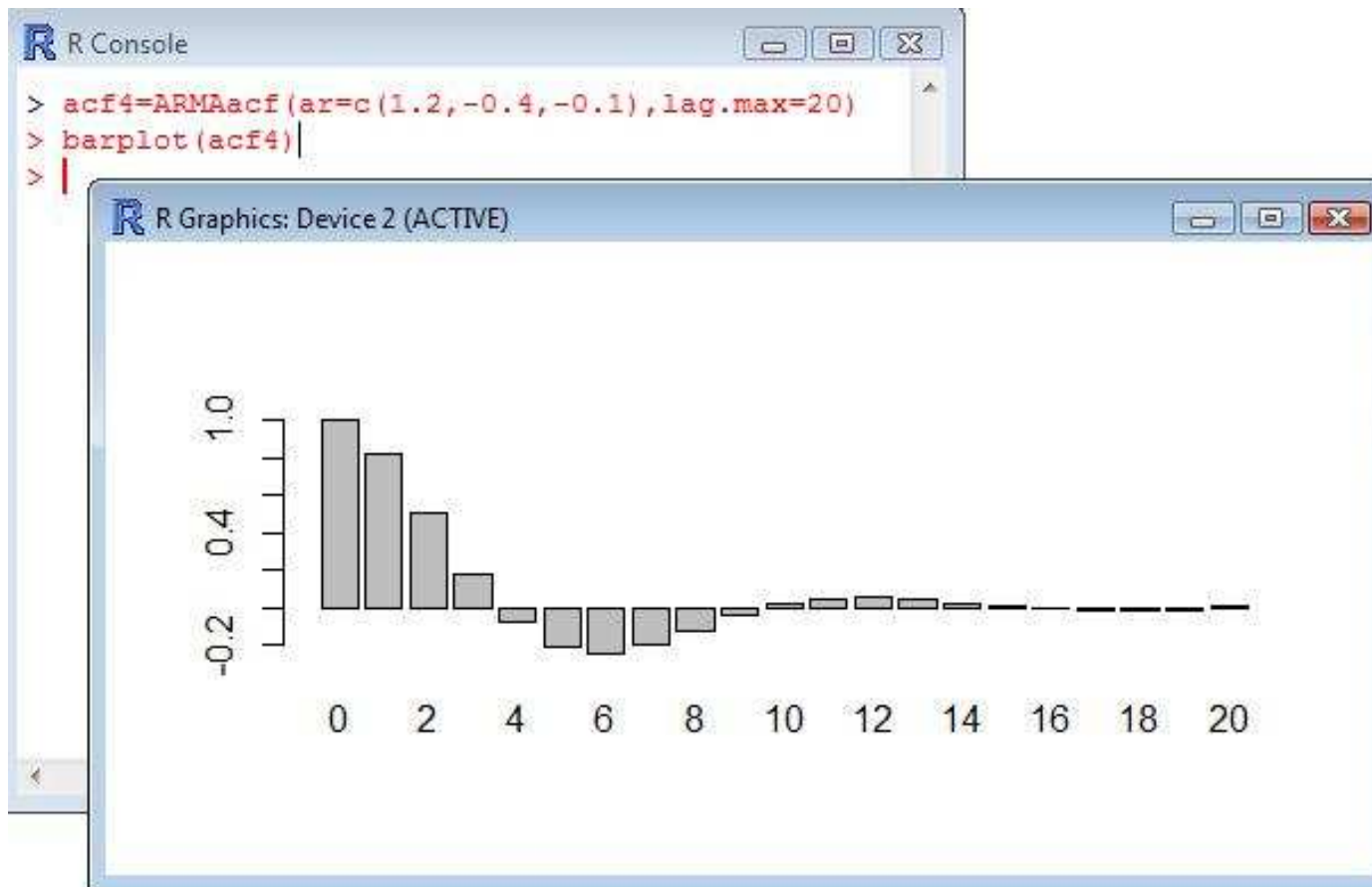
AR(p) proces - ACF - príklad 2

- AR(3) proces $x_t = 1.5 x_{t-1} - 0.8 x_{t-2} + 0.2 x_{t-3} + u_t$



AR(p) proces - ACF - príklad 3

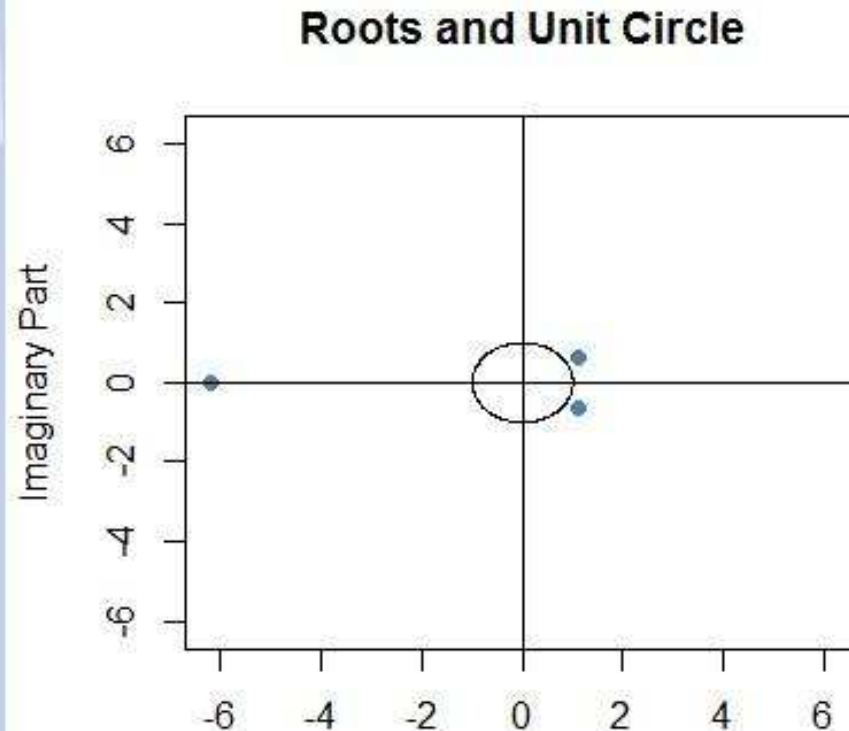
- AR(3) proces $x_t = 1.2 x_{t-1} - 0.4 x_{t-2} - 0.1 x_{t-3} + u_t$
- Dajú sa očakávať komplexné korene.



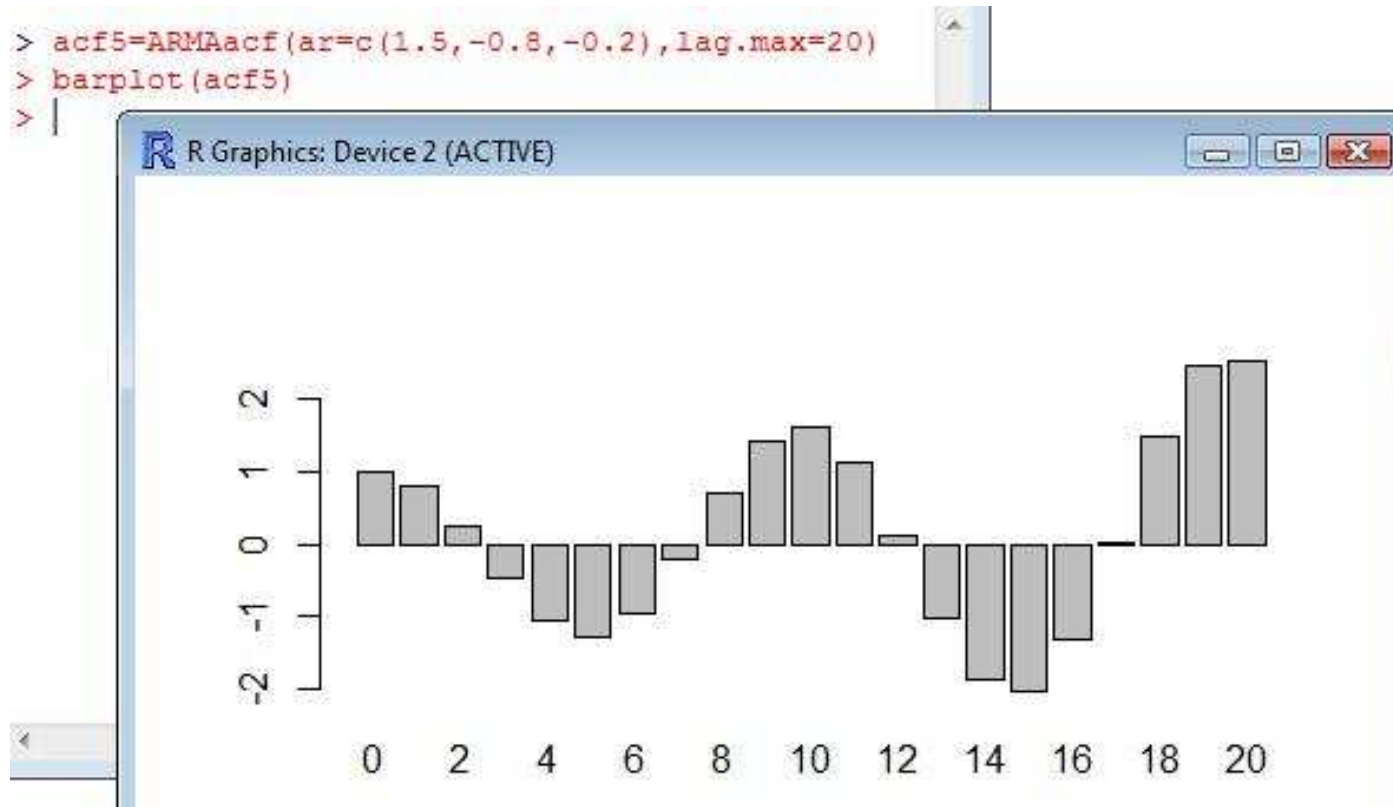
AR(p) proces - ACF - príklad 3

- Výpočet koreňov v R:
 - ◇ funkcia **armaRoots** z balíka **fArma**
 - ◇ vráti hodnoty koreňov - musia teda byť mimo jednotkového kruhu
- CVIČENIE: napíšte polynóm, ktorého korene počítame

```
> armaRoots(c(1.2, -0.4, -0.1))
      re      im  dist
1  1.0984  0.6381 1.2703
2  1.0984 -0.6381 1.2703
3 -6.1969  0.0000 6.1969
v |
```

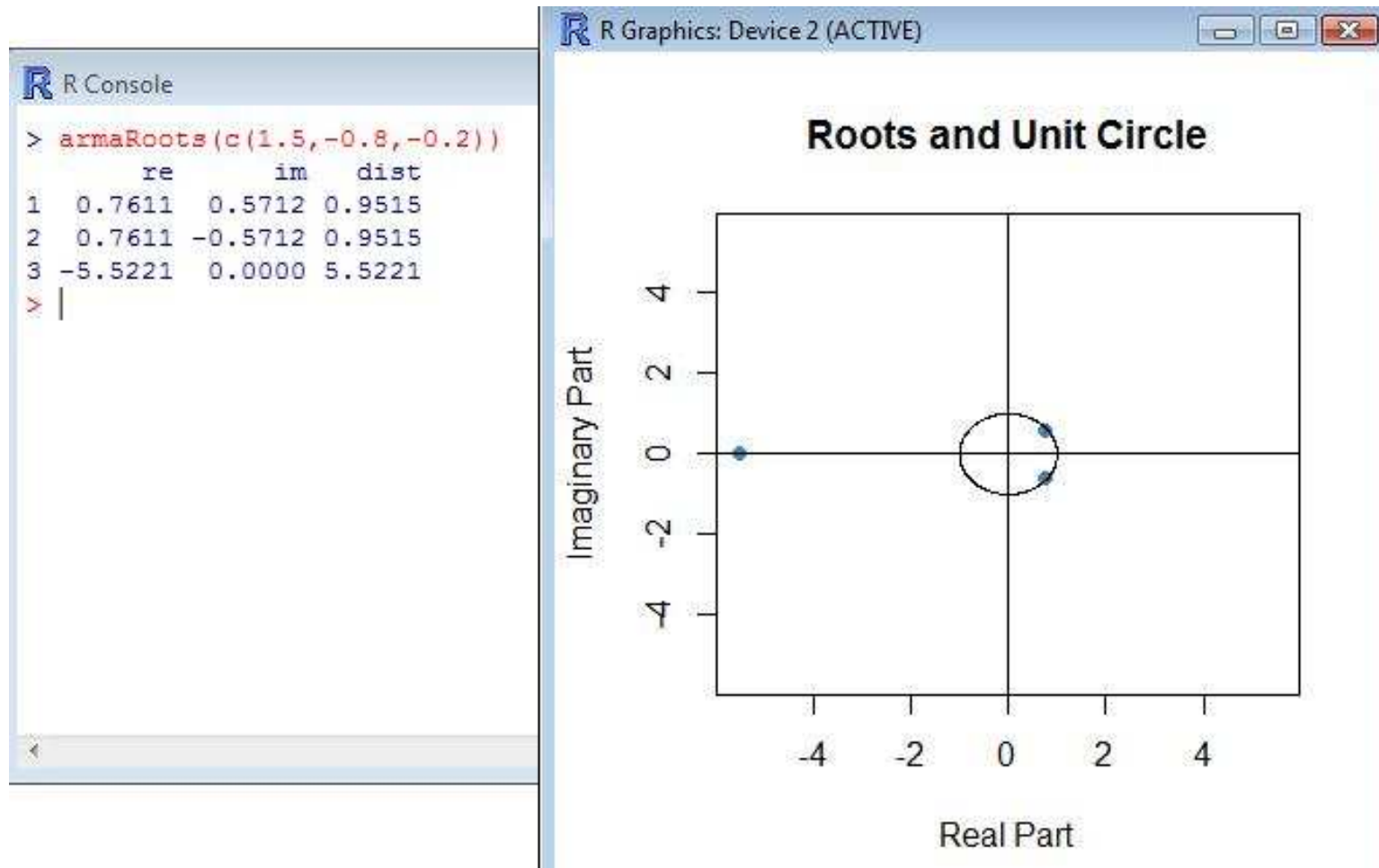


AR(p) proces - ACF - príklad 4



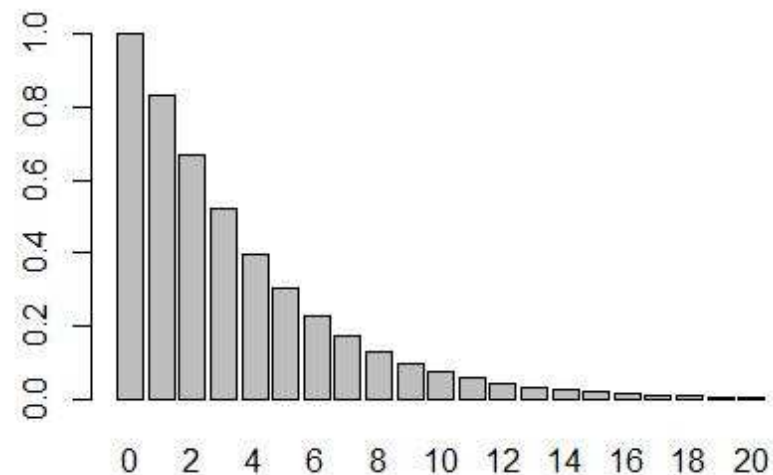
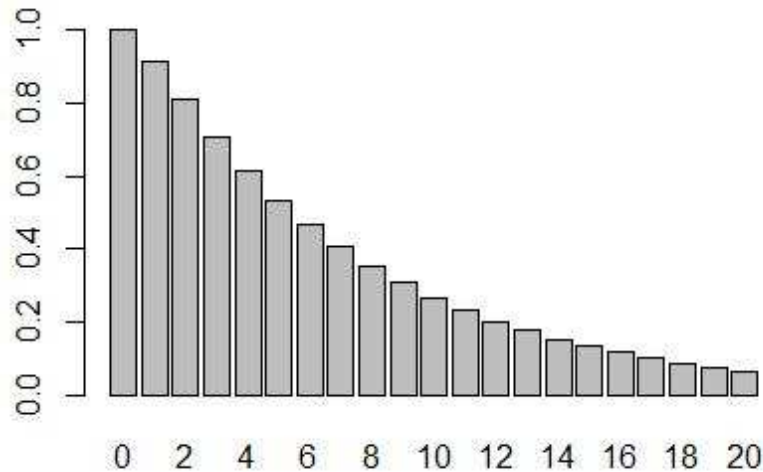
- Ako je to možné?
 - ◇ absolútna hodnota ACF väčšia ako 1
 - ◇ maximálna absolútna hodnota sa ešte zväčšuje

$AR(p)$ proces - ACF - príklad 4



- Proces nie je stacionárny → predchádzajúci výpočet nemá zmysel

$AR(p)$ proces - ACF - príklad 5



- ACF pre dva procesy: jeden je $AR(2)$ a druhý $AR(3)$
- **Nevieme ich takto na prvý pohľad rozlíšiť**
- Pri práci s reálnymi dátami sa navyše pridáva náhodnosť, nemáme presnú ACF, ale odhadujeme ju

IV.

Parciálna autokorelačná funkcia - určovanie rádu AR procesu

PACF - motivácia

- Uvažujme nejaký náhodný proces x_t s nulovou strednou hodnotou a modelujme jeho hodnotu pomocou prechádzajúcich k hodnôt:

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_k x_{t-k} + u_t$$

- Označme koeficienty Φ_{ki} , kde k je počet použitých starších hodnôt procesu x a i je koeficient pri x_{t-i}
- Teda:

$$x_t = \Phi_{11} x_{t-1} + u_t$$

$$x_t = \Phi_{21} x_{t-1} + \Phi_{22} x_{t-2} + u_t$$

$$x_t = \Phi_{31} x_{t-1} + \Phi_{32} x_{t-2} + \Phi_{33} x_{t-3} + u_t$$

...

$$x_t = \Phi_{k1} x_{t-1} + \Phi_{k2} x_{t-2} + \Phi_{k3} x_{t-3} + \dots + \Phi_{kk} x_{t-k} + u_t$$

- Ak x je AR(p) proces, tak $\Phi_{kk} = 0$ pre $k > p$.

PACF - definícia a výpočet

- Koeficient Φ_{kk} sa nazýva **parciálna autokorelácia rádu k**
- Postupnosť týchto koeficientov vytvára **parciálnu autokorelačnú funkciu (PACF)**
- Výpočet: vyjdeme zo vzt'ahu

$$x_t = \Phi_{k1}x_{t-1} + \Phi_{k2}x_{t-2} + \Phi_{k3}x_{t-3} + \dots + \Phi_{kk}x_{t-k} + u_t$$

a rovnako ako pri odvodení Yule-Wolkerovych rovníc dostaneme

$$\rho(1) = \Phi_{k1} + \Phi_{k2} \rho(1) + \dots + \Phi_{kk} \rho(k-1)$$

$$\rho(2) = \Phi_{k1} \rho(1) + \Phi_{k2} + \dots + \Phi_{kk} \rho(k-2)$$

...

$$\rho(k) = \Phi_{k1} \rho(k-1) + \Phi_{k2} \rho(k-2) + \dots + \Phi_{kk}$$

PACF - definícia a výpočet

- Maticový zápis:

$$\begin{bmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \Phi_{k1} \\ \Phi_{k2} \\ \dots \\ \Phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(k) \end{bmatrix}$$

- Zaujímá nás iba Φ_{kk} , použijeme Cramerovo pravidlo:

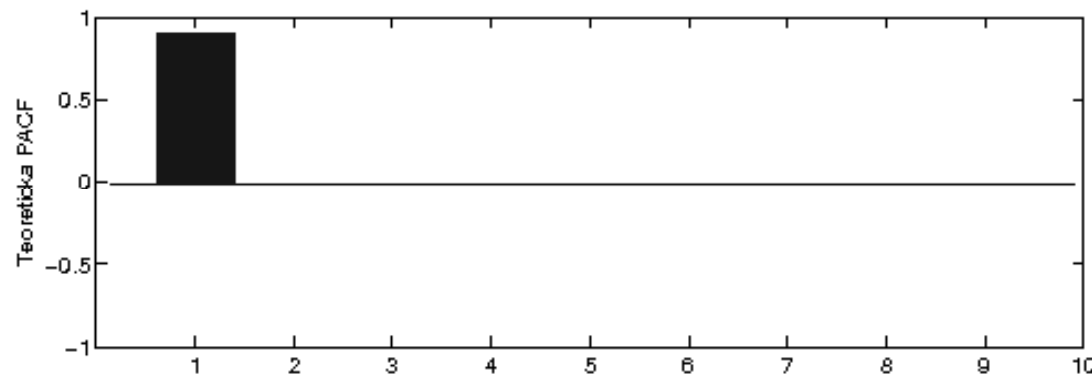
$$(9) \quad \Phi_{kk} = \frac{\det \begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & \rho(k) \end{pmatrix}}{\det \begin{pmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ & & \dots & \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{pmatrix}}$$

PACF - príklad: AR(1) proces

- Postupne počítame:

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

- Z def. PACF vyplýva, že aj nasledujúce $\Phi_{kk} = 0$
- Pre $\alpha = 0.9$:

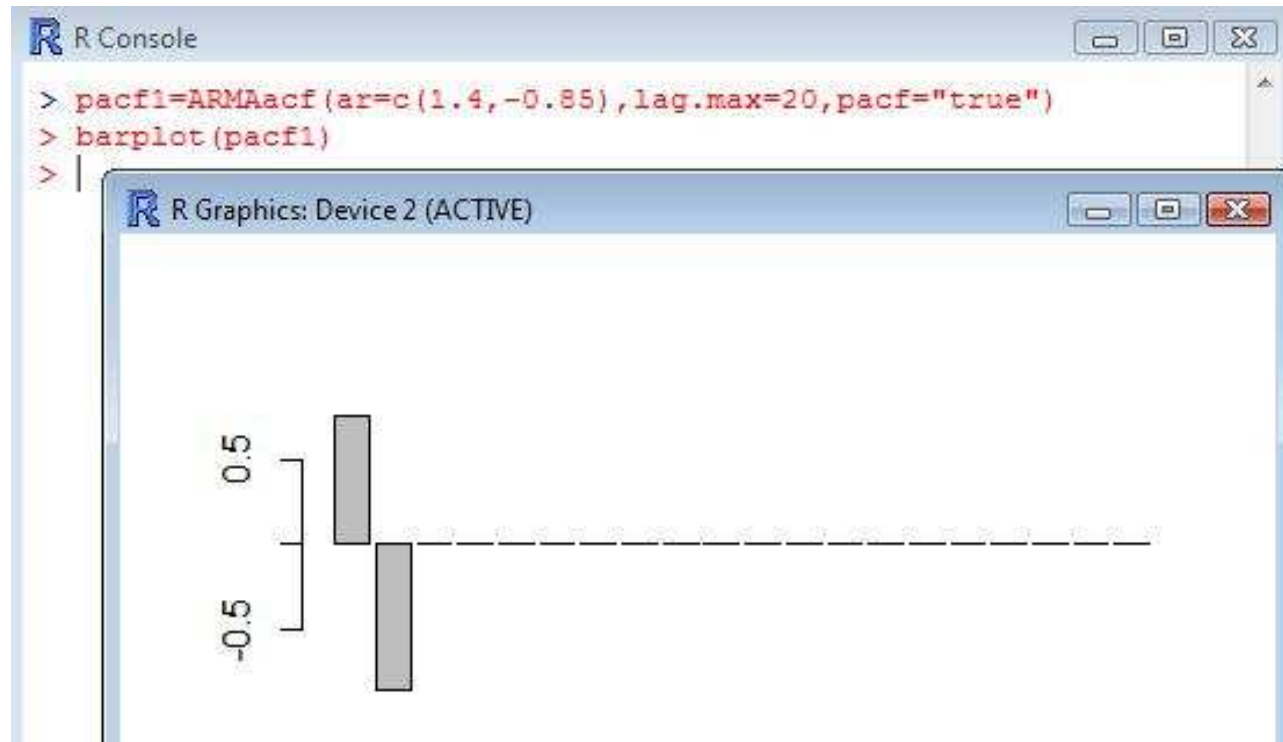


PACF - príklad 1

- Výpočet PACF v softvéri R - znovu funkcia **ARMAacf** z balíka **stats**
- Pre proces $x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$ sme počítali ACF, teraz PACF:

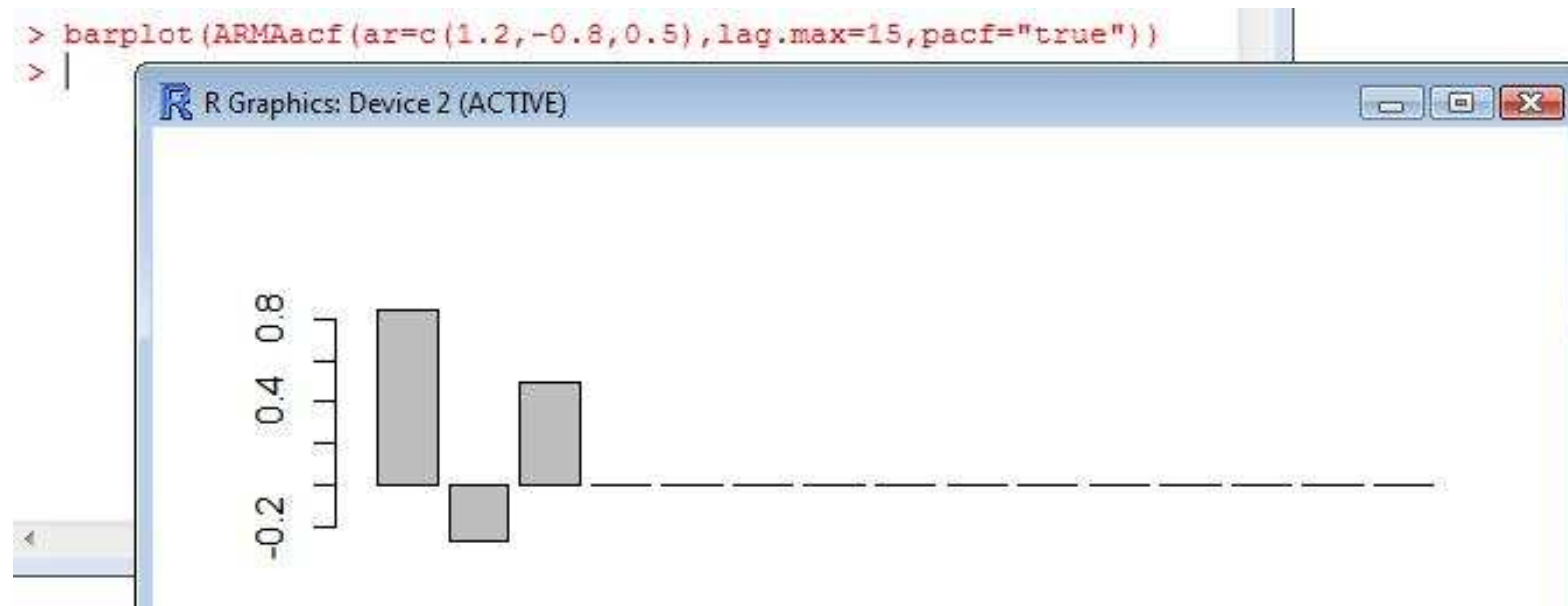
ARMAacf(ar=c(1.4,-0.85), lax.max=20, pacf="true")

PACF - príklad 1



PACF - príklad 2

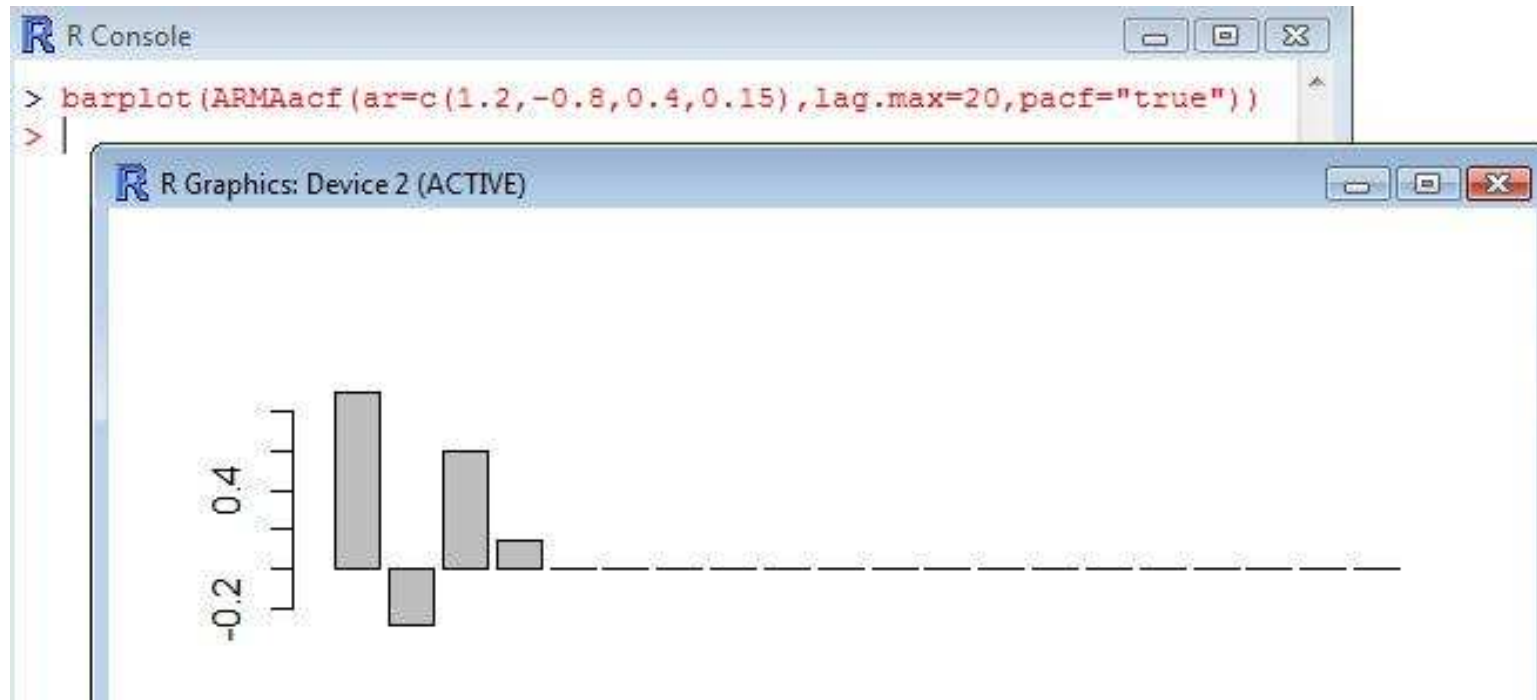
- AR(3) proces $x_t = 1.2 x_{t-1} - 0.8 x_{t-2} + 0.5 x_{t-3} + u_t$



PACF - príklad 3

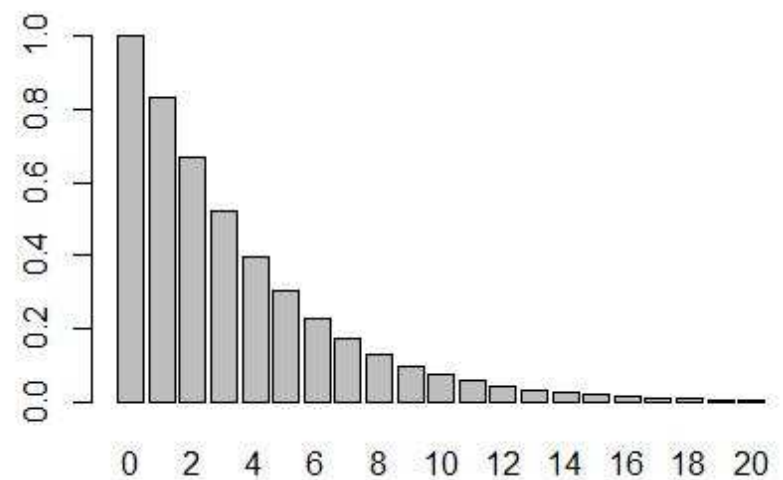
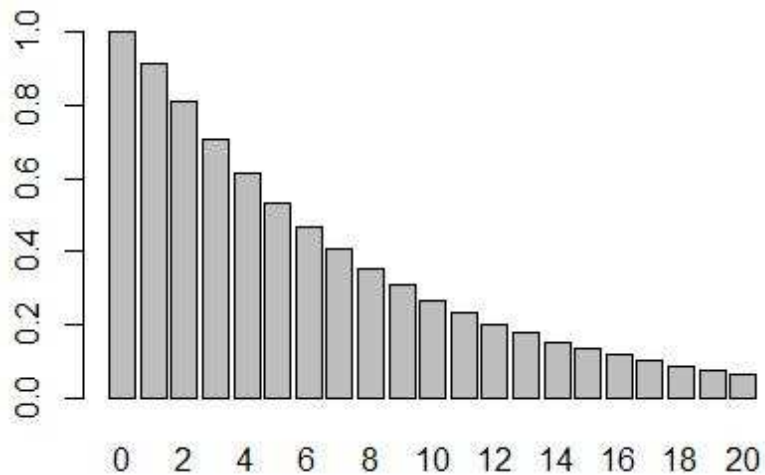
- AR(4) proces

$$x_t = 1.2 x_{t-1} - 0.8 x_{t-2} + 0.4 x_{t-3} + 0.15 x_{t-4} + u_t$$



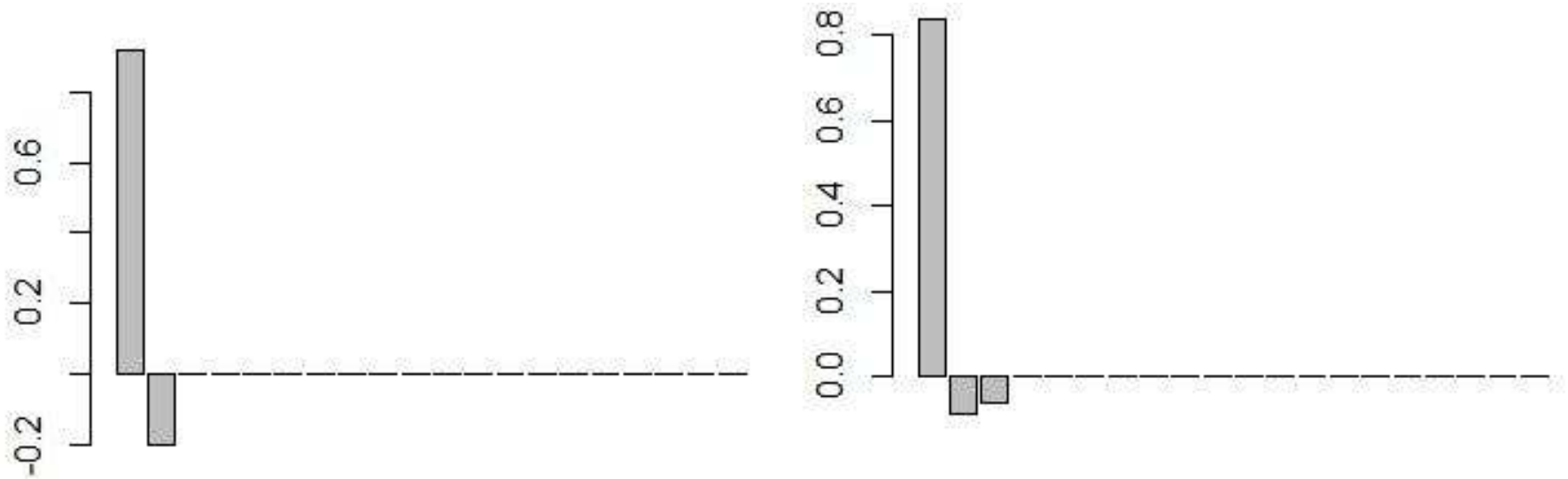
PACF - príklad 4

- Pripomeňme si:
ACF pre dva procesy, jeden je **AR(2)** a druhý **AR(3)**,
nevieme ich takto rozlíšiť



PACF - príklad 4

- PACF týchto procesov:



- Tu je jasné, že vľavo je AR(2) a vpravo je AR(3) proces

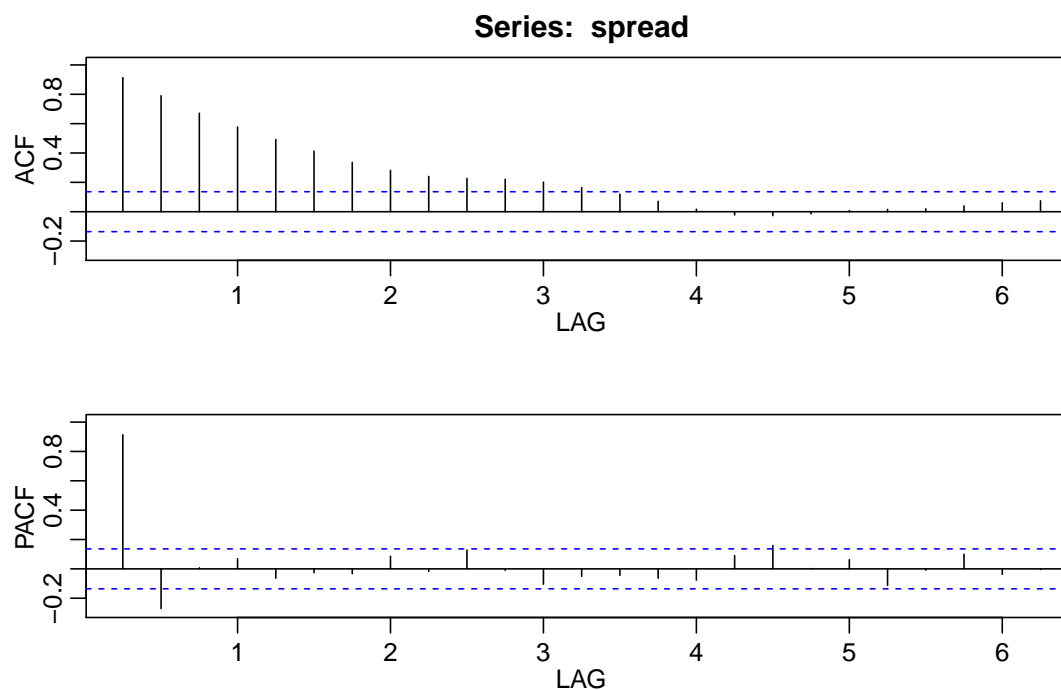
PACF - odhadovanie z dát

- Za teoretické autokorelácie vo vzt'ahu (15) dosadíme ich konzistentné odhady \rightarrow dostaneme konzistentný odhad $\hat{\Phi}_{kk}$
- Pre AR(p) proces je $\Phi_{kk} = 0$ pre $k > p$, pre tieto k asymptoticky platí

$$\text{Var}[\hat{\Phi}_{kk}] \approx \frac{1}{T}$$

Odhadovanie PACF - príklad 1

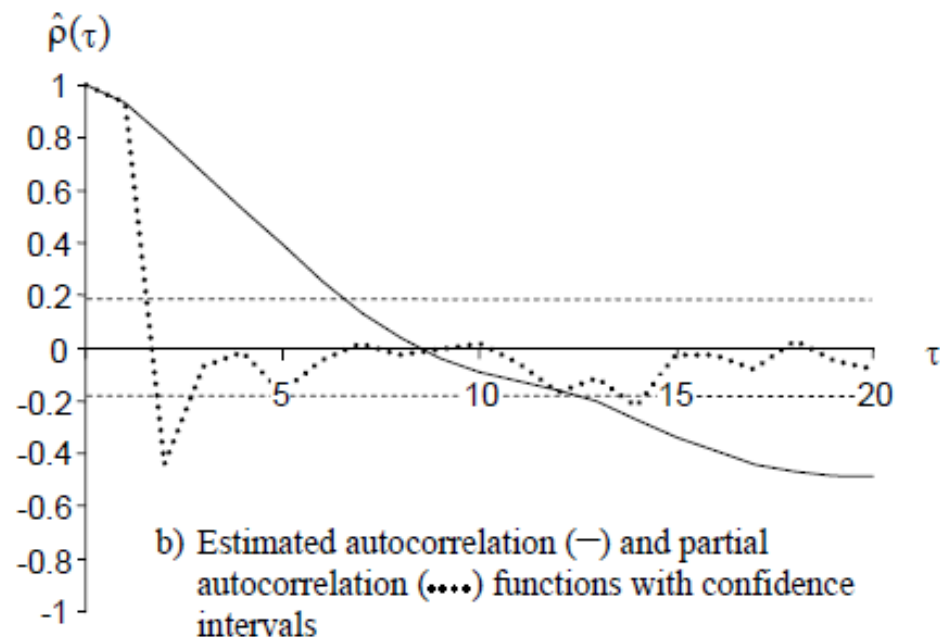
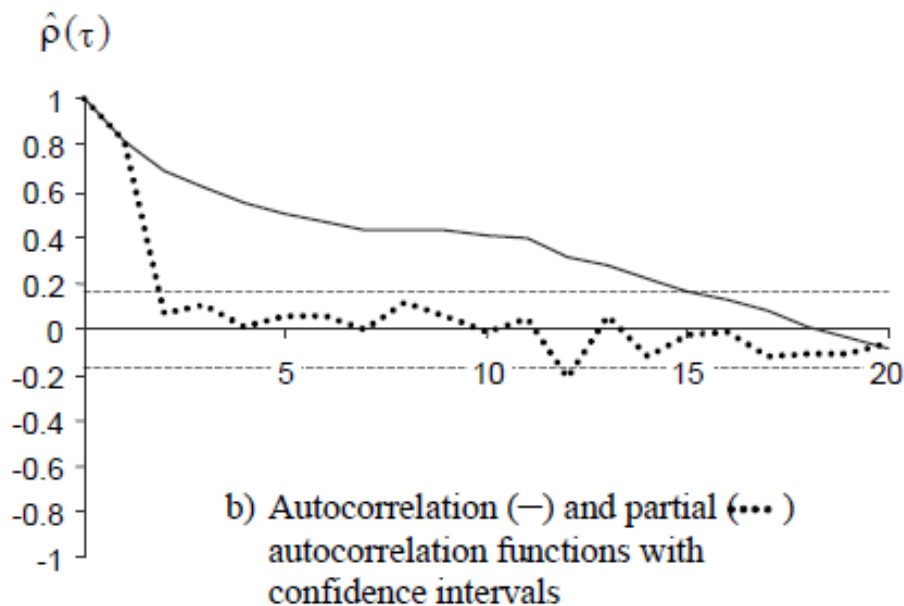
- Modelovali sme spread; príkazom `acf2(spread)` dostaneme ACF aj PACF:



- Vidíme, že treba odhadovať AR(2) proces (to sme aj spravili)

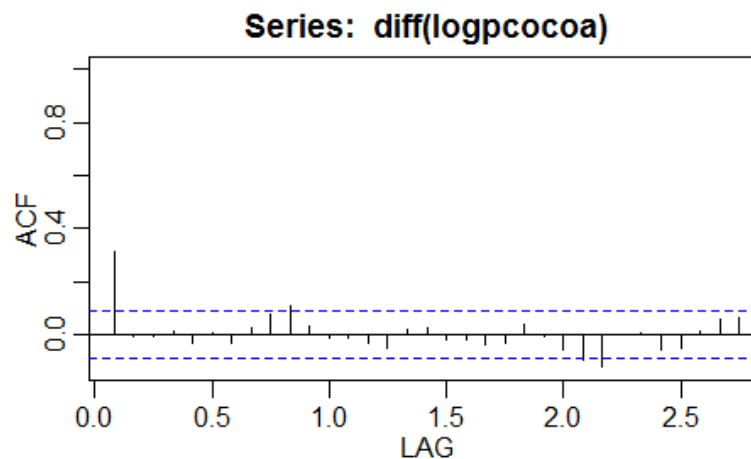
Odhadovanie PACF - príklad 2

- Z predchádzajúcich príkladov s reálnymi dátami:
 - ◇ volebné preferencie (vľavo) - AR(1)
 - ◇ úrokové miery (vpravo) - AR(2)



Na ďalších prednáškach

- Pripomenutie: AR(p) proces - ACF rýchlo klesá (monotónne alebo oscilujúco), PACF sa po p hodnotách rovná nule
- Dáta: **pcocoa** - ceny kakaa; ACF pre diferencie logaritmov:



- Tu je to v podstate naopak → **nebude to AR proces**
- Na ďalších prednáškach: modely pre takéto dáta