

M5201 – 8. CVIČENÍ:  
**Procesy diferenčně stacionární – ARIMA modely. Jejich generování a následné odhady**

## 1 Procesy nestacionární ve střední hodnotě

U procesů nestacionárních ve střední hodnotě je třeba odlišovat dva různé pojmy:

- deterministický trend
- stochastický trend

Připomeňme pojem stochastického trendu a uveďme jeho souvislost s ARIMA procesy.

### 1.1 Stochastický trend

U ARMA procesů jsme jako podmínku stacionarity požadovali, aby všechny kořeny polynomu

$$\Phi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p$$

ležely vně jednotkové kružnice, tj. aby proces byl kauzální. Pokud ale nějaký kořen leží

- **na jednotkové kružnici**, mluvíme o procesu **nestacionárním se stochastickým trendem**,
- **uvnitř jednotkové kružnice**, mluvíme o procesu **nestacionárním explozivního typu**.

První případ, tedy nestacionární proces se stochastickým trendem, lze převést na stacionární proces pomocí **diferencování**. Pomocí *operátoru zpětného chodu*  $B$ , který je definován vztahem

$$BY_t = Y_{t-1}$$

můžeme zavést **diferenční operátor** a vyjádřit 1. diferenci, 2. diferenci, ...  $d$ -tou diferenci:

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} = (1 - B)Y_t \\ \Delta^2 Y_t &= \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - B)^2 Y_t \\ &\vdots \\ \Delta^d Y_t &= (1 - B)^d Y_t \end{aligned}$$

Nestacionární proces se stochastickým trendem nazýváme **integrovaný smíšený model** a značíme jej

$$ARIMA(p, d, q).$$

Formální zápis pomocí operátoru zpětného chodu je

$$ARIMA(p, d, q) : \Phi(B)(1 - B)^d Y_t = \Theta(B)\varepsilon_t.$$

Provedeme-li  $d$ -krát diferencování a položíme-li

$$W_t = (1 - B)^d Y_t,$$

pak  $W_t$  je stacionární  $ARMA(p, q)$  proces.

Některé náhodné procesy, se kterými jsme se již setkali, jsou speciálními případy  $ARIMA(p, d, q)$  procesu. Přehled je uveden v následující tabulce.

p	d	q	Zkratka	Název	Formální zápis
0			$IMA(d, q)$	Integrovaný proces klouzavých součtů	$(1 - B)^d Y_t = \Theta(B) \varepsilon_t$
0	0		$MA(q)$	Proces klouzavých součtů	$Y_t = \Theta(B) \varepsilon_t$
		0	$ARI(p, d)$	Integrovaný autoregresní proces	$\Phi(B)(1 - B)^d Y_t = \varepsilon_t$
	0	0	$AR(p)$	Autoregresní proces	$\Phi(B) Y_t = \varepsilon_t$
0		0	$I(d)$	Integrovaný proces	$(1 - B)^d Y_t = \varepsilon_t$
0	1	0	$I(1)$	Náhodná procházka	$(1 - B) Y_t = \varepsilon_t$

Operátor

$$\nu(B) = \Phi(B)(1 - B)^d$$

se někdy nazývá **zobecněný autoregresní operátor**.

Pokud  $\nu(B)$  chápeme jako polynom proměnné  $B$ , pak vzhledem ke kauzalitě procesu  $\Phi(B)W_t = \Theta(B)\varepsilon_t$  má  $\nu(B)$  **právě  $p$  kořenů ležících vně jednotkového kruhu a  $d$  kořenů rovných 1**.

V praxi se nejprve  $d$ -násobným diferencováním časové řady získá stacionární řada  $W_t$  a po ní se pak vybuduje proces  $ARMA(p, q)$ . Pokud jsme původně měli  $n$  pozorování  $Y_1, Y_2, \dots, Y_n$ , po diferencování zůstane  $n - d$  pozorování  $W_{d+1}, \dots, W_n$ .

#### PŘÍKLAD 1 - ARI PROCES

Zaměříme se nejprve na konkrétní ARI proces řádu (1,1,0) (tzn. po jednom diferencování dostaneme AR(1) proces). Zvolme konkrétní hodnotu  $\varphi = 0.9$ , tj. proces lze zapsat ve tvaru

$$(1 - B)(1 - 0.9B)Y_t = \varepsilon_t.$$

AR(1) proces s tímto koeficientem by měl být kauzální, protože u AR(1) procesů je podmínka pro kauzalitu splněna, jestliže  $|\varphi| < 1$ . Můžeme ale ověřit i obecnou podmínku kauzality pro AR proces, tj. zda kořen polynomu  $1 - \varphi z = 0$  leží vně jednotkové kružnice.

$$1 - 0.9z = 0$$

$$z = \frac{1}{0.9} = \frac{10}{9} > 1$$

Tento výpočet můžeme provést také pomocí R:

```
> ARphi <- c(0.9)
> ARroots <- polyroot(c(1, -ARphi))
> Mod(ARroots)
```

[1] 1.111111

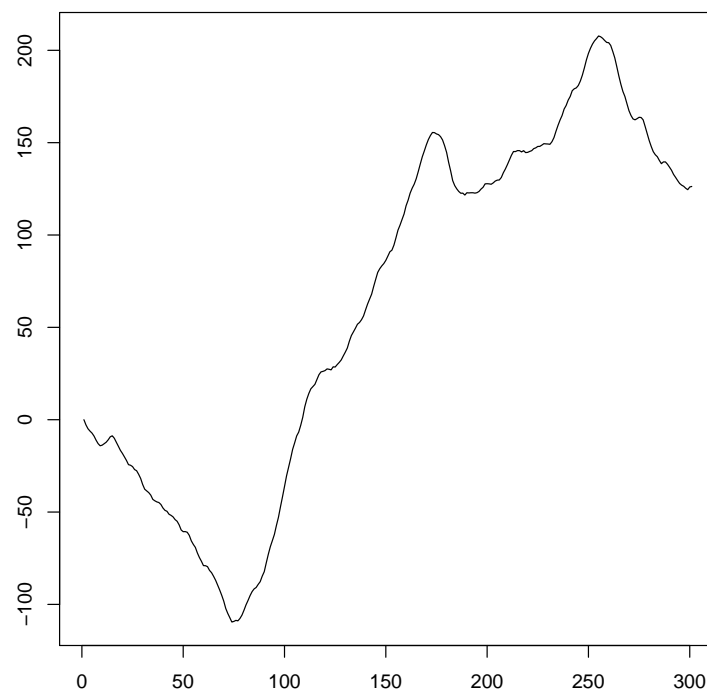
Z předchozích cvičení víme, že kauzální proces je stacionární. ARI proces se od AR procesu liší v tom, že polynom na levé straně rovnice popisující AR proces je navíc vynásoben výrazem  $(1 - B)$ , a proto má polynom na levé straně rovnice navíc ještě jeden kořen rovný 1.

ARI proces můžeme postupně rozepsat následujícím způsobem:

$$\begin{aligned}(1 - B)(1 - 0.9B)Y_t &= \varepsilon_t \\(1 - 1.9B + 0.9B^2)Y_t &= \varepsilon_t \\Y_t - 1.9Y_{t-1} + 0.9Y_{t-2} &= \varepsilon_t \\Y_t &= 1.9Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t\end{aligned}$$

Nyní vygenerujeme konkrétní realizaci ARI procesu.

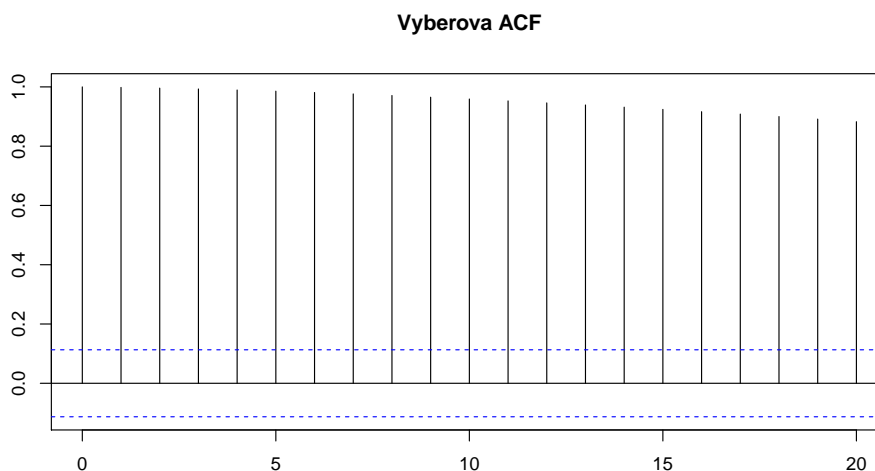
```
> ARphi <- c(0.9)
> ari.sim <- arima.sim(model = list(order = c(1, 1, 0),
  ar = ARphi), n = 300)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> plot(ari.sim)
```



Obrázek 1: Simulace ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

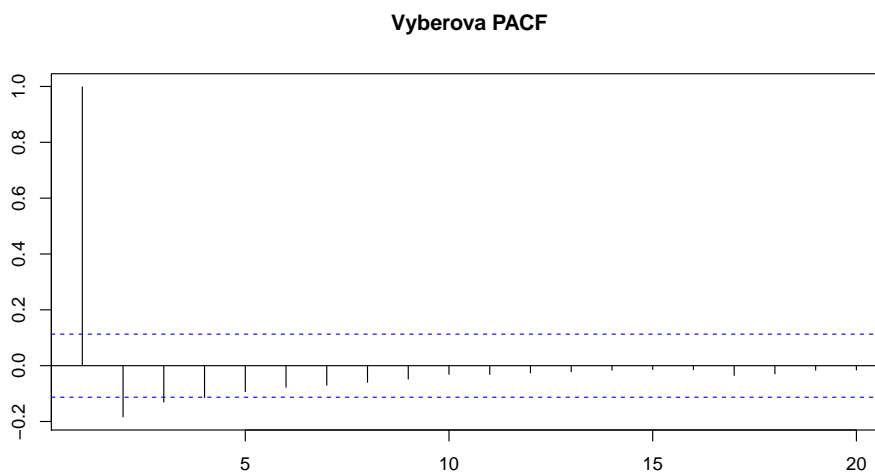
Prohlédneme si autokorelační a parciální autokorelační funkci simulovaného ARI procesu.

```
> vARIacf <- acf(ari.sim, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vARIacf, main = "Vyberova ACF")
```



Obrázek 2: ACF pro simulovanou řadu ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

```
> vARIpacf <- pacf(ari.sim, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vARIpacf, main = "Vyberova PACF")
```



Obrázek 3: ACF pro simulovanou řadu ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

Vidíme, že autokorelační funkce klesá velmi pomalu, což odpovídá procesům s jednotkovým kořenem.

Parametry ARI procesu odhadneme pomocí funkce `arima`, kde jako parametr `order` zadáme řád ARIMA procesu `c(1,1,0)`.

```
> ari.fit <- arima(ari.sim, order = c(1, 1, 0))
> print(ari.fit)
```

Call:

```
arima(x = ari.sim, order = c(1, 1, 0))
```

Coefficients:

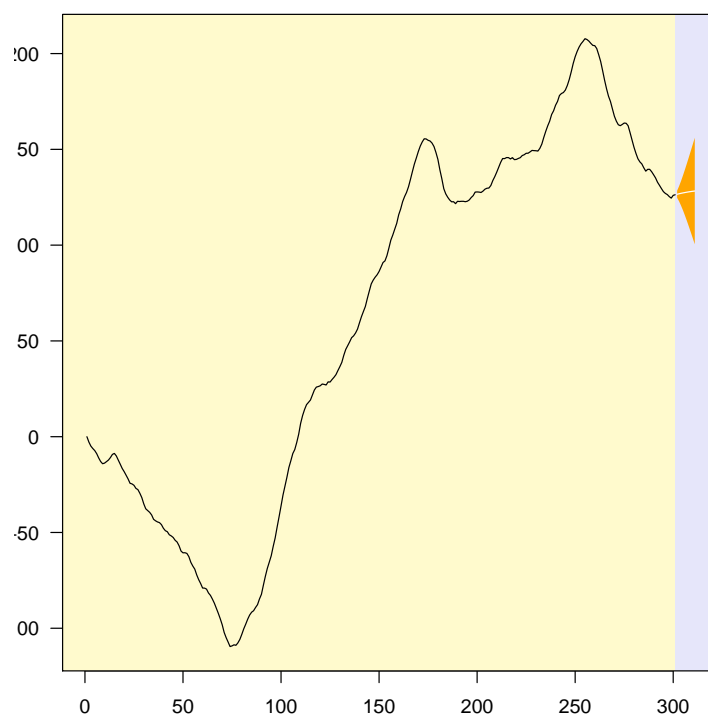
```
ar1
```

```
0.9219
s.e. 0.0218
```

```
sigma^2 estimated as 0.8711: log likelihood = -405.93, aic = 815.86
```

Nyní, když máme odhadnuty parametry ARI procesu, využijeme jejich znalosti k predikování budoucích hodnot procesu. K tomu použijeme funkci `PlotPredictARIMA`, jejíž zdrojový kód musíme nejprve načíst ze souboru `FunkceM5201.R`.

```
> fileFun <- paste(data.library, "FunkceM5201.R", sep = "")
> source(fileFun)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> PlotPredictARIMA(ari.sim, ari.fit, n.ahead = 10)
```

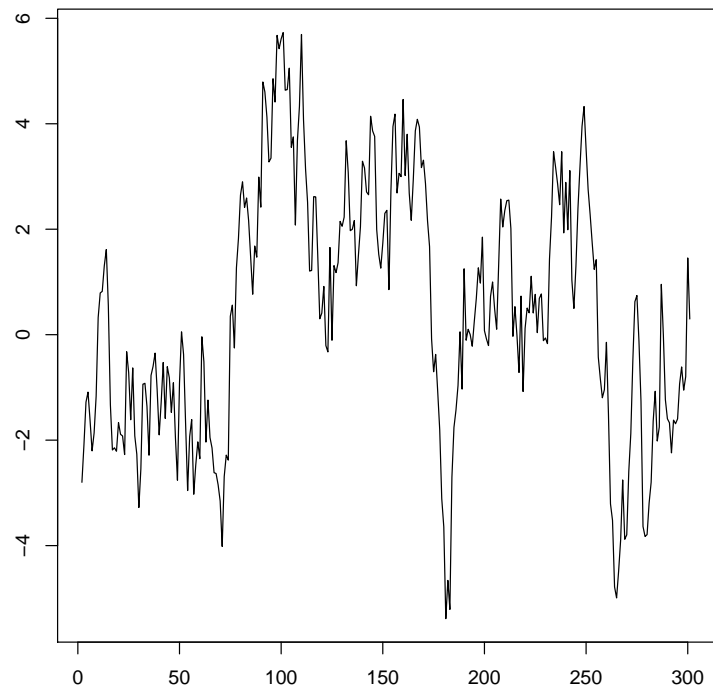


Obrázek 4: Predikované hodnoty pro simulovanou řadu ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

Pokud bychom neznali přesný řád odhadovaného ARI procesu, museli provést diferencování tolikrát, dokud bychom nezískali stacionární řadu, a poté odhadnout parametry ARMA procesu pro diferencovanou řadu.

Tento postup nyní vyzkoušíme. Nejprve spočítáme první diference simulované řady.

```
> dif.ari <- diff(ari.sim)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> plot(dif.ari)
```

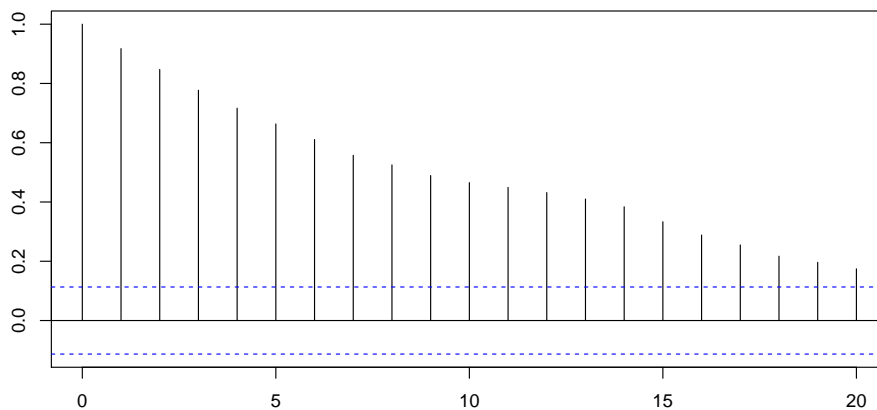


Obrázek 5: 1. diference simulované řady ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

Prohlédneme si autokorelační a parciální autokorelační funkci diferencované řady.

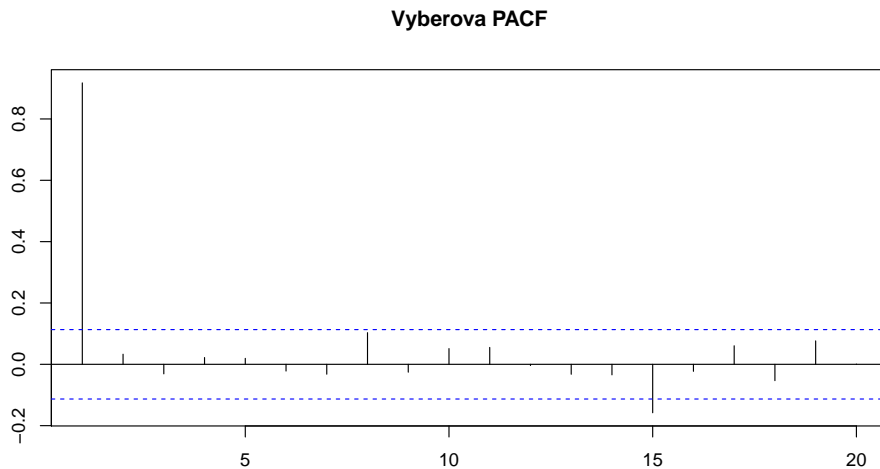
```
> difARIacf <- acf(dif.ari, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(difARIacf, main = "Vyberova ACF")
```

**Vyberova ACF**



Obrázek 6: ACF pro simulovanou řadu ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

```
> vARIpacf <- pacf(dif.ari, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vARIpacf, main = "Vyberova PACF")
```



Obrázek 7: PACF pro simulovanou řadu ARI procesu  $(1 - B)(1 - 0.9B)Y_t = \varepsilon_t$

Vidíme, že autokorelační funkce klesá exponenciálně k nule, zatímco parciální autokorelační funkce je významně nenulová pouze v bodě  $k = 1$ , což by odpovídalo  $AR(1)$  procesu. Proto odhadneme parametry takového modelu pro diferencovanou řadu.

```
> dif.ari.fit <- arima(dif.ari, order = c(1, 0, 0))
> print(dif.ari.fit)
```

Call:

```
arima(x = dif.ari, order = c(1, 0, 0))
```

Coefficients:

	ar1	intercept
	0.9201	0.3020
s.e.	0.0222	0.6509

```
sigma^2 estimated as 0.8706: log likelihood = -405.83, aic = 817.65
```

Z výsledku vidíme, že jsme opět dostali podobný odhad parametru  $\varphi$ .

### PŘÍKLAD 2 - IMA PROCES

Nyní se podívejme na konkrétní případ IMA procesu řádu  $(0,1,2)$  (tzn. proces, u něž po jednom diferencování dostaneme  $MA(2)$  proces). Zvolíme konkrétní hodnoty parametrů  $\theta_1 = 0.75$  a  $\theta_2 = 0.2$ , tj. proces lze zapsat ve tvaru

$$\begin{aligned}(1 - B)Y_t &= \varepsilon_t + 0.75\varepsilon_{t-1} + 0.2\varepsilon_{t-2} \\ Y_t - Y_{t-1} &= \varepsilon_t + 0.75\varepsilon_{t-1} + 0.2\varepsilon_{t-2} \\ Y_t &= Y_{t-1} + \varepsilon_t + 0.75\varepsilon_{t-1} + 0.2\varepsilon_{t-2}\end{aligned}$$

Ještě předtím, než vygenerujeme realizaci tohoto procesu, bylo by vhodné ověřit, že polynom definující  $MA$  část procesu má kořeny ležící vně jednotkové kružnice a že  $MA$  proces, který bychom získali diferencováním, je tím pádem invertibilní.

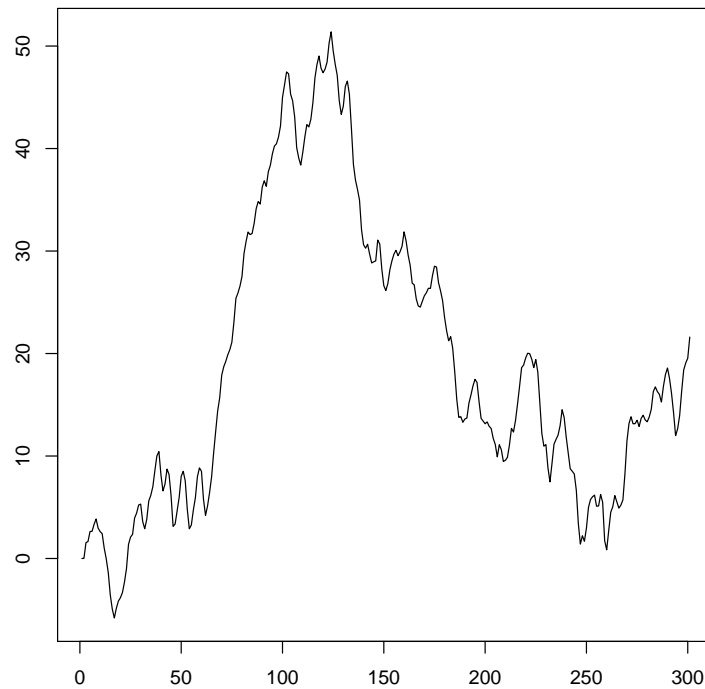
```
> Mtheta <- c(0.75, 0.2)
> Mod(polyroot(c(1, Mtheta)))
```

[1] 2.236068 2.236068

Absolutní hodnota obou kořenů je větší než 1, tudíž leží vně jednotkové kružnice a podařilo se nám ověřit invertibilitu.

Analogicky jako v předchozím příkladu vygenerujeme 300 hodnot  $IMA(1,2)$  procesu a takto získanou řadu vykreslíme do grafu.

```
> ima.sim <- arima.sim(model = list(order = c(0, 1, 2),
  ma = MAteta), n = 300)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> plot(ima.sim)
```

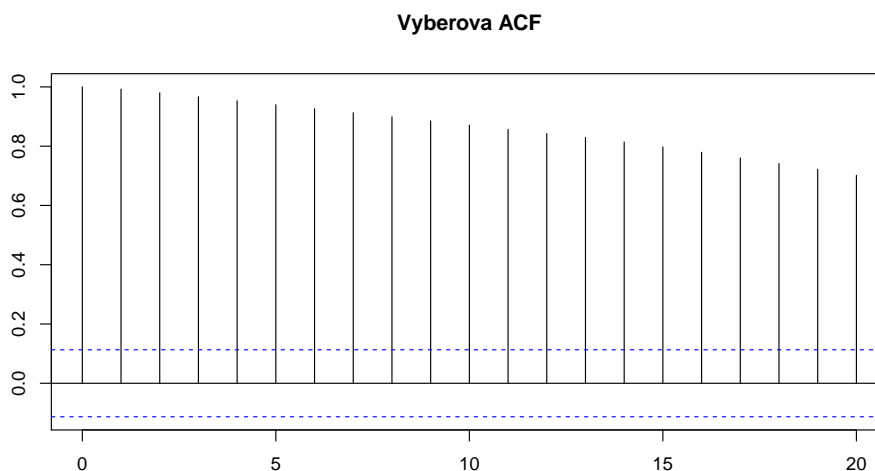


Obrázek 8: Simulovaná řada IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

Opět si můžeme prohlédnout grafy výběrové autokorelační a parciální autokorelační funkce.

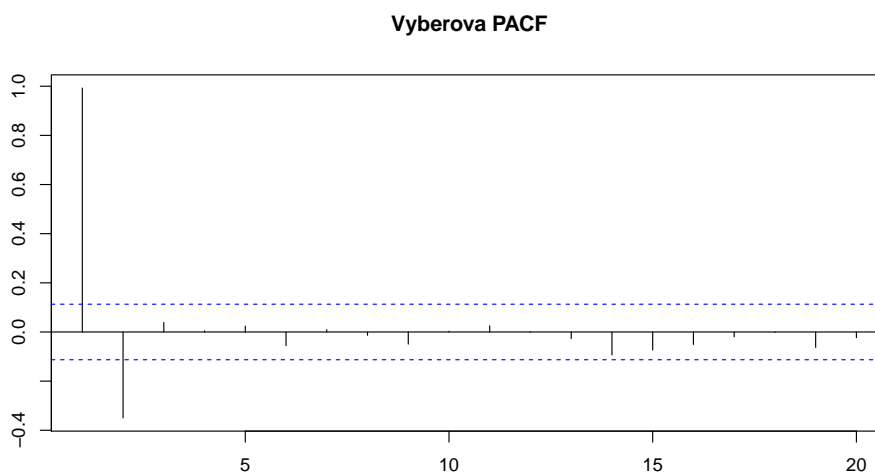
```
> vIMAacf <- acf(ima.sim, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vIMAacf, main = "Vyberova ACF")
```





Obrázek 9: ACF simulované řady IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

```
> vIMApacf <- pacf(ima.sim, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vIMApacf, main = "Vyberova PACF")
```



Obrázek 10: PACF simulované řady IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

Opět vidíme, že autokorelační funkce klesá k nule velmi pomalu, což obvykle indikuje přítomnost jednotkového kořenu.

Podíváme se, jak by dopadl odhad parametrů *IMA* procesu, a na základě odhadnutého modelu budeme predikovat 10 budoucích hodnot procesu.

```
> ima.fit <- arima(ima.sim, order = c(0, 1, 2))
> print(ima.fit)
```

Call:

```
arima(x = ima.sim, order = c(0, 1, 2))
```

Coefficients:

```

      ma1    ma2
0.7251 0.2331
s.e. 0.0534 0.0590

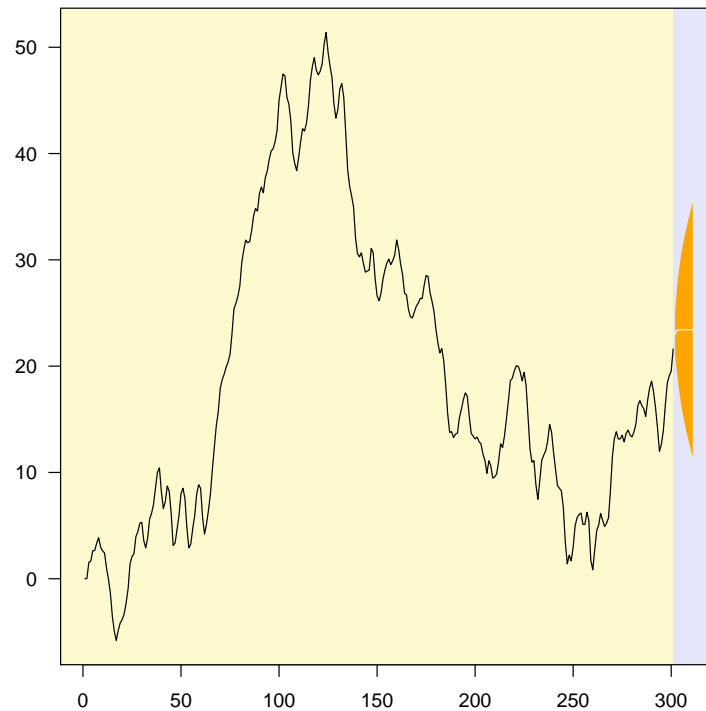
```

```
sigma^2 estimated as 1.078: log likelihood = -437.19, aic = 880.37
```

```

> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> PlotPredictARIMA(ima.sim, ima.fit, n.ahead = 10)

```



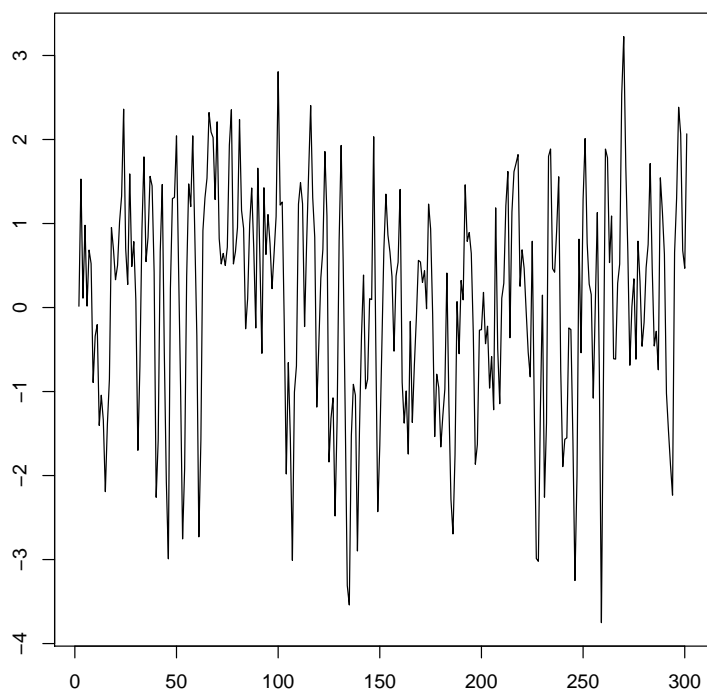
Obrázek 11: Predikce pro simulovanou řadu IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

Opět se přesvědčíme, že diferencování získáme z původního  $IMA(1, 2)$  procesu  $MA(2)$  proces. Nejprve si prohlédneme graf diferencované časové řady.

```

> dif.ima <- diff(ima.sim)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> plot(dif.ima)

```

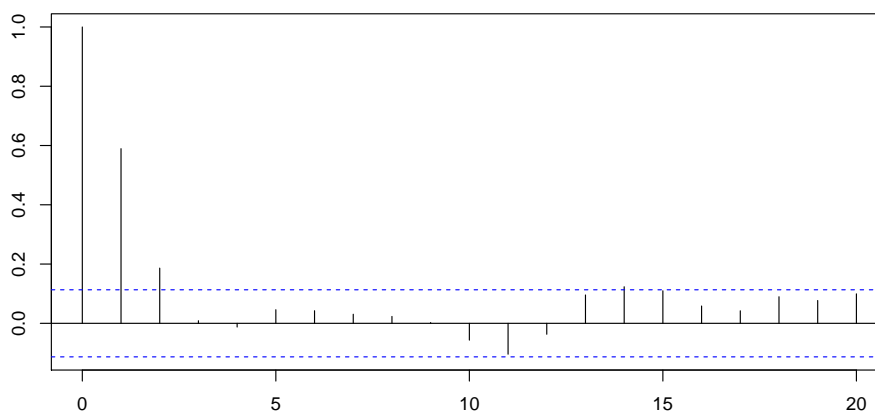


Obrázek 12: 1. diference simulované řady IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

Dále se jako obvykle podíváme na výběrovou autokorelační a parciální autokorelační funkci

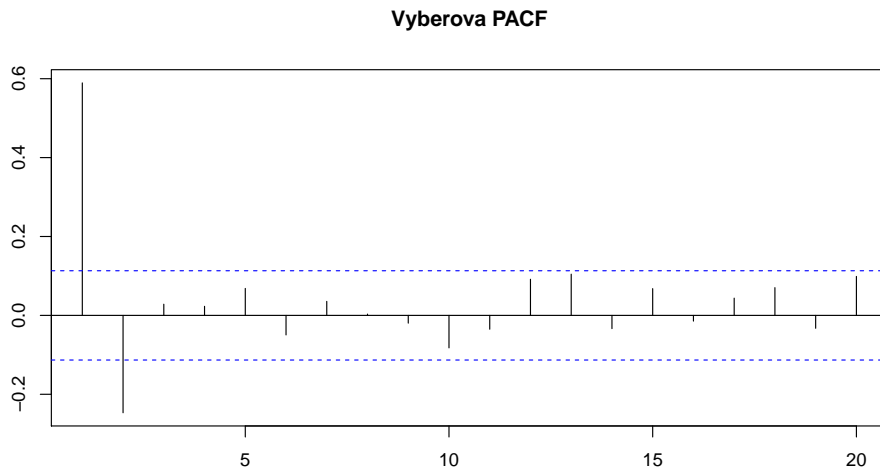
```
> difIMAacf <- acf(dif.ima, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(difIMAacf, main = "Vyberova ACF")
```

**Vyberova ACF**



Obrázek 13: ACF diferencovaného IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

```
> vIMApacf <- pacf(dif.ima, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vIMApacf, main = "Vyberova PACF")
```



Obrázek 14: PACF diferencovaného IMA procesu  $(1 - B)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

Z vlastností  $MA(2)$  procesu vyplývá, že autokorelační funkce by měla být významně nenulová pouze v bodech  $k = 0, 1, 2$  a parciální autokorelační funkce by měla exponenciálně klesat k nule. Ani jeden z předchozích dvou grafů není v rozporu s očekáváním.

Na závěr ještě ověříme, že když odhadneme  $MA(2)$  model pro diferencovanou řadu, dostaneme podobné odhady parametrů  $\theta_1$  a  $\theta_2$ , jako když jsme odhadovali parametry  $IMA(1, 2)$  modelu pro původní nediferencovaná data.

```
> dif.ima.fit <- arima(dif.ima, order = c(0, 0, 2))
> print(dif.ima.fit)
```

Call:

```
arima(x = dif.ima, order = c(0, 0, 2))
```

Coefficients:

	ma1	ma2	intercept
	0.7243	0.2324	0.0756
s.e.	0.0534	0.0591	0.1170

sigma<sup>2</sup> estimated as 1.076: log likelihood = -436.98, aic = 881.96

### PŘÍKLAD 3 – ARIMA PROCES

Ještě se podíváme na konkrétní případ ARIMA procesu řádu (2,1,2) (tzn. proces, u něž po jednom diferencování dostaneme ARMA(2,2) proces). Zvolíme konkrétní hodnoty parametrů  $\varphi_1 = 0.9$ ,  $\varphi_2 = -0.25$ ,  $\theta_1 = 0.75$  a  $\theta_2 = 0.2$ , tj. proces lze zapsat ve tvaru

$$\begin{aligned} (1 - B)(1 - 0.9B + 0.25B^2)Y_t &= (1 + 0.75B + 0.2B^2)\varepsilon_t \\ (1 - 0.9B + 0.25B^2 - B + 0.9B^2 - 0.25B^3)Y_t &= (1 + 0.75B + 0.2B^2)\varepsilon_t \\ (1 - 1.9B + 1.15B^2 - 0.25B^3)Y_t &= (1 + 0.75B + 0.2B^2)\varepsilon_t \\ Y_t - 1.9Y_{t-1} + 1.15Y_{t-2} - 0.25Y_{t-3} &= \varepsilon_t + 0.75\varepsilon_{t-1} + 0.2\varepsilon_{t-2} \\ Y_t &= 1.9Y_{t-1} - 1.15Y_{t-2} + 0.25Y_{t-3} + \varepsilon_t + 0.75\varepsilon_{t-1} + 0.2\varepsilon_{t-2} \end{aligned}$$

Nejprve ověříme kauzalitu a invertibilitu AR a MA části procesu, tj. zjistíme, zda kořeny polynomu  $\Phi(z)$  a  $\Theta(z)$  leží vně jednotkové kružnice.

```
> ARphi <- c(0.9, -0.25)
> Mod(polyroot(c(1, -ARphi)))
```

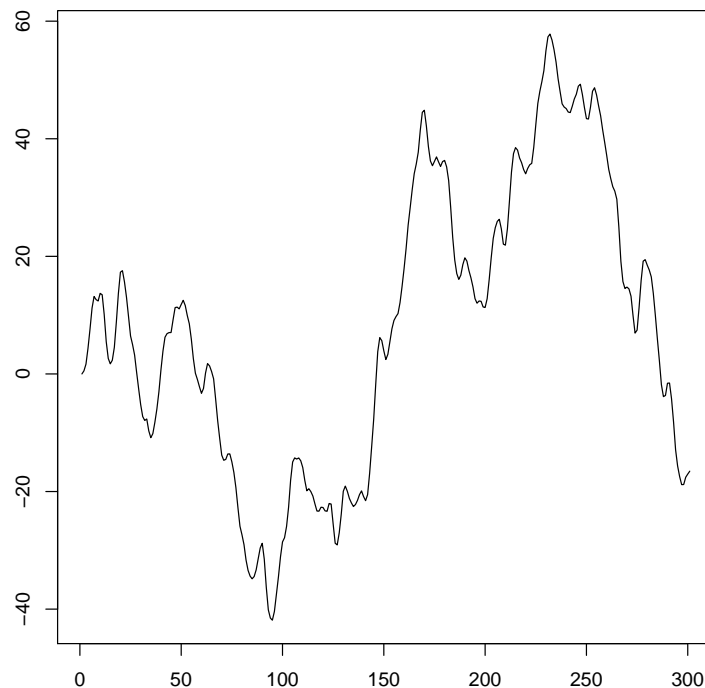
```
[1] 2 2
```

```
> Mtheta <- c(0.75, 0.2)
> Mod(polyroot(c(1, Mtheta)))
```

```
[1] 2.236068 2.236068
```

Všechny kořeny leží vně jednotkové kružnice, takže se nám podařilo ověřit kauzalitu resp. invertibilitu AR a MA části ARIMA procesu. Teď už můžeme vygenerovat 300 hodnot procesu a získanou časovou řadu znázornit graficky.

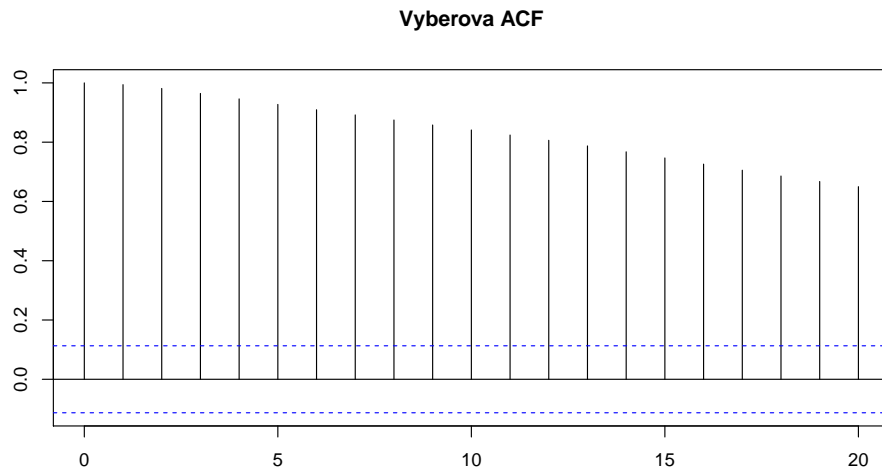
```
> arimaSim <- arima.sim(model = list(order = c(2, 1, 2),
  ar = ARphi, ma = Mtheta), n = 300)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> plot(arimaSim)
```



Obrázek 15: Simulovaný ARIMA proces  $(1 - B)(1 - 0.9B + 0.25B^2)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

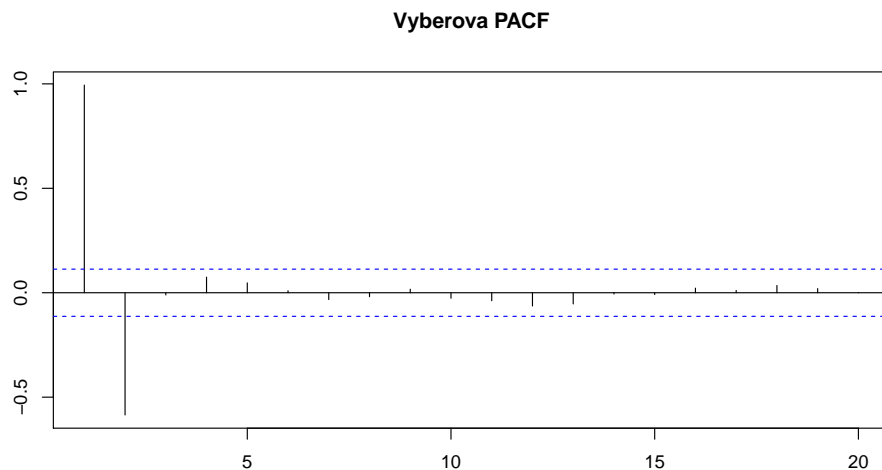
Opět si prohlédneme výběrovou autokorelační a parciální autokorelační funkci.

```
> VARIMAacf <- acf(arimaSim, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(VARIMAacf, main = "Vyberova ACF")
```



Obrázek 16: ACF simulovaného ARIMA procesu  $(1-B)(1-0.9B+0.25B^2)Y_t = (1+0.75B+0.2B^2)\varepsilon_t$

```
> VARIMApacf <- pacf(arimaSim, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(VARIMApacf, main = "Vyberova PACF")
```



Obrázek 17: PACF simulovaného ARIMA procesu  $(1-B)(1-0.9B+0.25B^2)Y_t = (1+0.75B+0.2B^2)\varepsilon_t$

Stejně jako v případě předchozích procesů s jednotkovým kořenem vidíme, že autokorelační funkce klesá k nule velmi pozvolna.

Pomocí funkce `arima` opět odhadneme parametry  $ARIMA(2, 1, 2)$  procesu, abychom je mohli porovnat s teoretickými hodnotami zadanými při simulaci časové řady. Potom na základě odhadnutého modelu provedeme predikci 10 budoucích hodnot procesu.

```
> arima.fit <- arima(arimaSim, order = c(2, 1, 2))
> print(arima.fit)
```

Call:

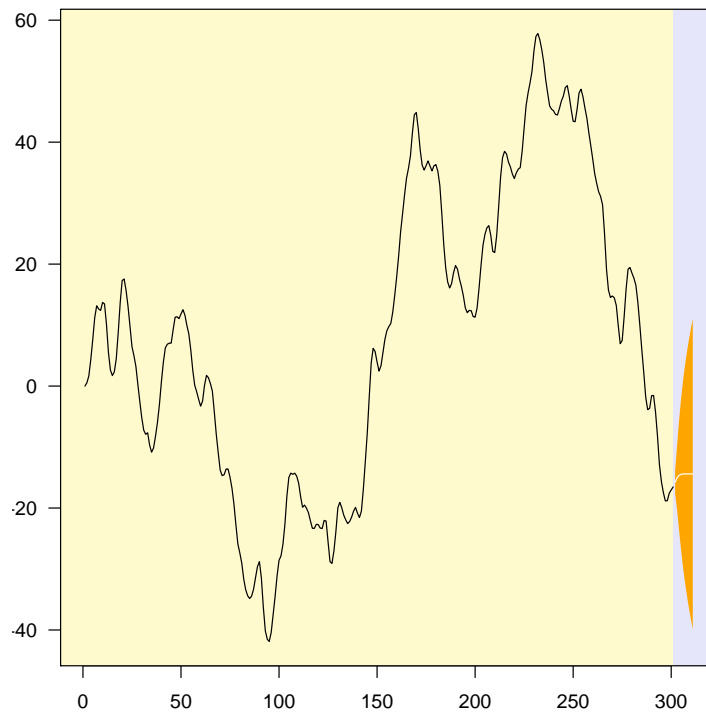
```
arima(x = arimaSim, order = c(2, 1, 2))
```

Coefficients:

	ar1	ar2	ma1	ma2
	0.7284	-0.1648	0.8108	0.2638
s.e.	0.2767	0.2095	0.2814	0.1973

sigma^2 estimated as 0.9598: log likelihood = -420.72, aic = 851.44

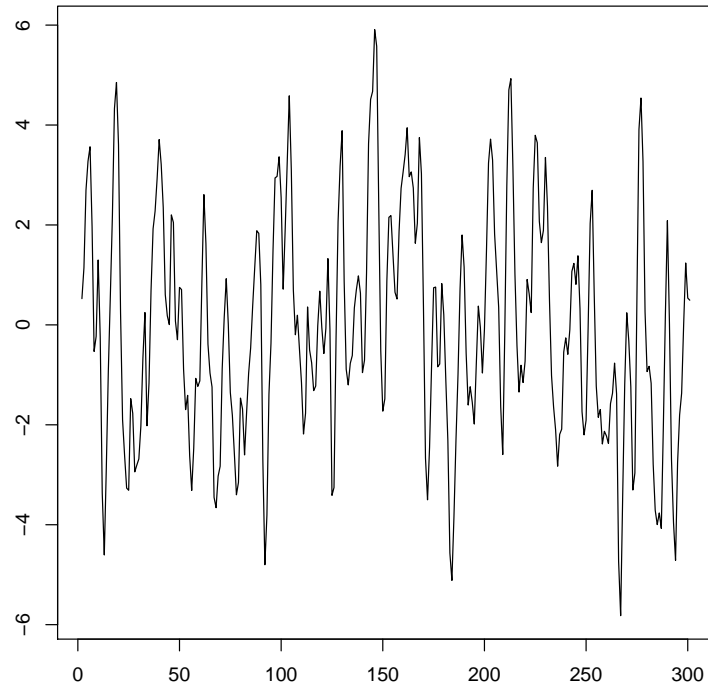
```
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> PlotPredictARIMA(arimaSim, arima.fit, n.ahead = 10)
```



Obrázek 18: Predikované hodnoty ARIMA procesu  $(1 - B)(1 - 0.9B + 0.25B^2)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

V závěrečné části příkladu ověříme, že diferencováním  $ARIMA(2, 1, 2)$  procesu dostaneme  $ARMA(2, 2)$  proces s týmiž parametry. Nejprve proto spočítáme 1. difference simulované časové řady a vykreslíme je do grafu.

```
> dif.arima <- diff(arimaSim)
> par(mfrow = c(1, 1), mar = c(2, 2, 1, 0) + 0.05)
> plot(dif.arima)
```

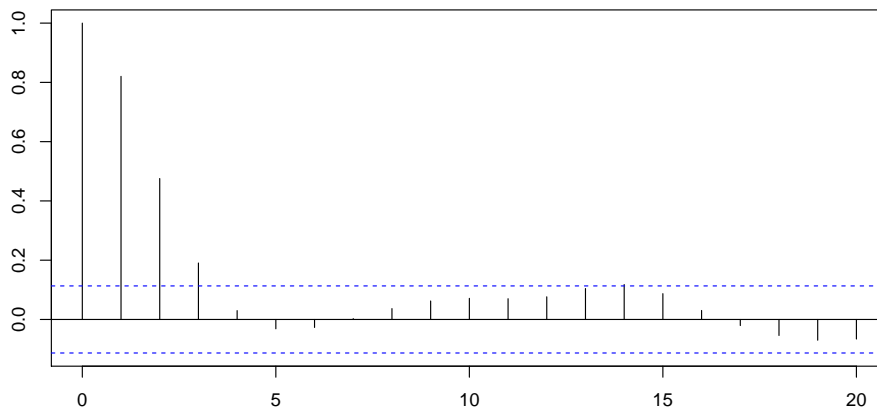


Obrázek 19: 1. diference simulovaného ARIMA procesu  $(1 - B)(1 - 0.9B + 0.25B^2)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$

Podíváme se na výběrovou autokorelační a parciální autokorelační funkci diferencované řady. Jestliže se jedná o ARMA proces, jak předpokládáme, měly by obě funkce exponenciálně klesat k nule.

```
> difARIMAacf <- acf(dif.arima, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(difARIMAacf, main = "Vyberova ACF")
```

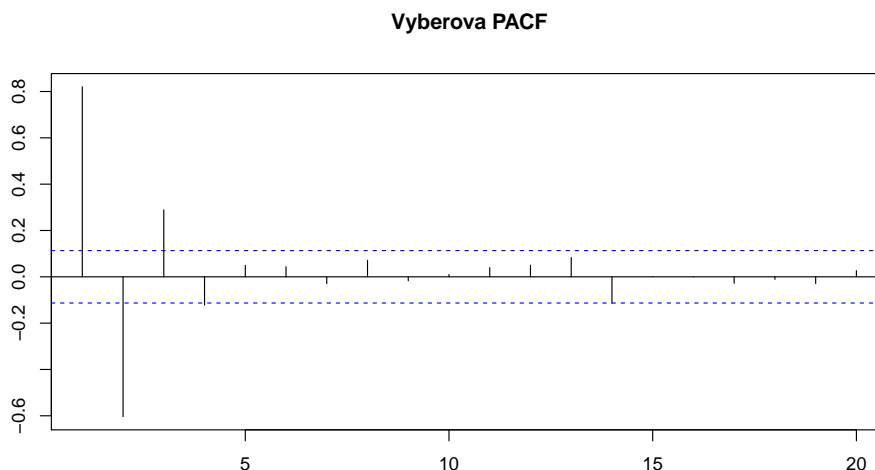
**Vyberova ACF**



Obrázek 20: ACF simulovaného ARIMA procesu  $(1 - B)(1 - 0.9B + 0.25B^2)Y_t = (1 + 0.75B + 0.2B^2)\varepsilon_t$



```
> vARIMApacf <- pacf(dif.arima, lag.max = 20, plot = FALSE)
> par(mfrow = c(1, 1), mar = c(2, 2, 3, 0) + 0.05)
> plot(vARIMApacf, main = "Vyberova PACF")
```



Obrázek 21: PACF simulovaného ARIMA procesu  $(1-B)(1-0.9B+0.25B^2)Y_t = (1+0.75B+0.2B^2)\varepsilon_t$

Vidíme, že průběhy obou funkcí odpovídají ARMA procesu. Na samotný závěr odhadneme parametry  $ARMA(2, 2)$  procesu pro diferencovanou řadu, abychom se přesvědčili, že při odhadu parametrů pro původní časovou řadu (ARIMA proces) před diferencováním i při odhadu parametrů ARMA procesu pro časovou řadu po diferencování dostaneme podobné výsledky, které odpovídají teoretickým hodnotám použitým při simulaci procesu.

```
> dif.arima.fit <- arima(dif.arima, order = c(2, 0, 2))
> print(dif.arima.fit)
```

Call:

```
arima(x = dif.arima, order = c(2, 0, 2))
```

Coefficients:

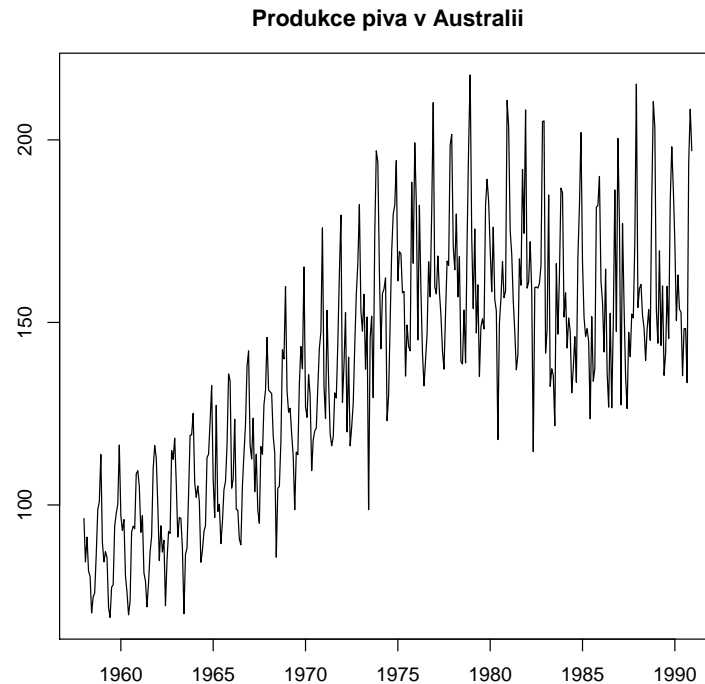
	ar1	ar2	ma1	ma2	intercept
	0.7292	-0.1656	0.8098	0.2632	-0.0433
s.e.	0.2768	0.2096	0.2815	0.1974	0.2673

sigma<sup>2</sup> estimated as 0.9597: log likelihood = -420.71, aic = 853.41

#### PŘÍKLAD 4 - PRODUKCE PIVA V AUSTRÁLII

Nyní vyzkoušíme aplikovat ARIMA modely na reálná data. K tomu použijeme měsíční údaje o množství piva vyrobeného v Austrálii, které byly zjišťovány od roku 1958. Data jsou uložena v souboru `beer.txt`. Data načteme pomocí funkce `read.table` jako obvykle a pak z nich vytvoříme časovou řadu, kterou vykreslíme do grafu.

```
> fileDat <- paste(data.library, "beer.txt", sep = "")
> beer <- read.table(fileDat, header = TRUE)
> beerTS <- ts(beer, start = 1958, frequency = 12)
> par(mar = c(2, 2, 3, 0) + 0.05)
> plot(beerTS, main = "Produkce piva v Australii")
```



Obrázek 22: Měsíční údaje o množství vyrobeného piva v Austrálii (1958-1991)

V časové řadě lze vysledovat rostoucí trend ve výrobě piva ve zkoumaném období. Z toho důvodu by mohl být vhodným modelem jednoduchý integrovaný proces klouzavých součtů  $IMA(1, 1)$  (tj.  $ARIMA(0, 1, 1)$ ), protože umožňuje modelovat tento trend tak, že přenáší pozorování v čase  $t-1$  do pozorování v čase  $t$  a přidává k němu náhodnou odchylku.  $IMA(1, 1)$  proces bývá pro podobná data často vhodný z toho důvodu, že představuje lineární trend s přidáním bílým šumem.

Parametry IMA procesu odhadneme pomocí funkce `arima`.

```
> ima.beer <- arima(beerTS, order = c(0, 1, 1))
> ima.beer
```

Call:

```
arima(x = beerTS, order = c(0, 1, 1))
```

Coefficients:

```
      ma1
      -0.3334
s.e.    0.0558
```

```
sigma^2 estimated as 360.4: log likelihood = -1723.27, aic = 3450.53
```

Odhadnutý model má podobu

$$Y_t = Y_{t-1} + \varepsilon_t - 0.333\varepsilon_{t-1}.$$

Odhadnutý model můžeme použít k predikci budoucího vývoje. K tomu v R použijeme jako obvykle funkci `predict()` s parametrem `n.ahead` udávajícím počet predikovaných hodnot.

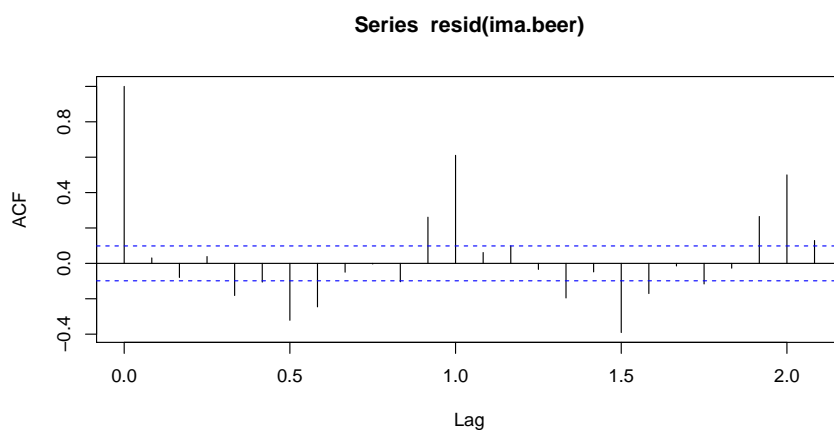
Pokud by nás například na konci roku 1991 zajímalo, jaké je predikované množství vyrobeného piva za celý následující rok 1992, a to dohromady za všech dvanáct měsíců, spočítáme predikci 12 budoucích hodnot a sečteme je.

```
> beer1992 <- predict(ima.beer, n.ahead = 12)
> sum(beer1992$pred)
```

```
[1] 2365.412
```

Poté, co jsme odhadli model pro časovou řadu, bychom měli správně ověřit vhodnost tohoto modelu. Diagnostickým grafem může být například graf výběrové autokorelační funkce pro rezidua.

```
> acf(resid(ima.beer))
```



Obrázek 23: ACF reziduí IMA(1,1) modelu pro data *Měsíční údaje o množství vyrobeného piva v Austrálii (1958-1991)*

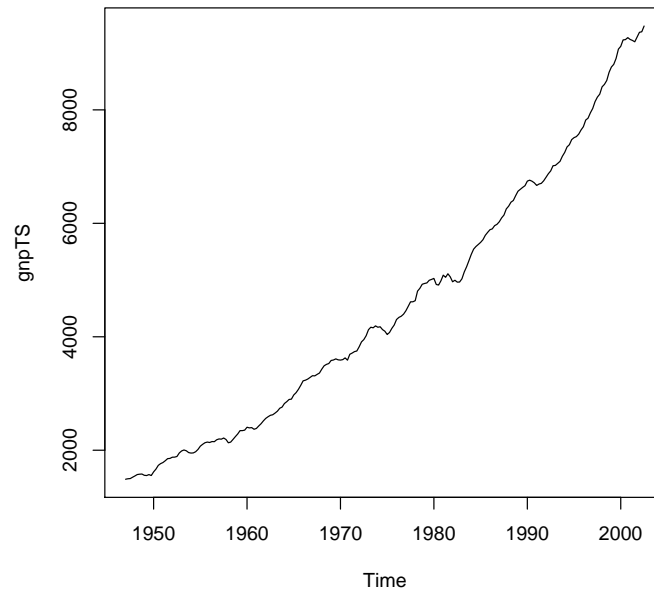
Graf autokorelační funkce neklesá exponenciálně k nule, ale má vrcholy s periodou 12, což naznačuje, že do modelu by bylo třeba zahrnout ještě sezónní složku (k tomu jsou vhodné například SARIMA modely).

#### PŘÍKLAD 5 - ANALÝZA AMERICKÉHO HNP

V posledním příkladu použijeme ARIMA modely k analýze hrubého národního produktu USA. Data uložená v souboru `gnp96.dat` jsou čtvrtletní údaje o velikosti HNP USA od 1. čtvrtletí roku 1947 do 3. čtvrtletí roku 2002. Pozorování je celkem 223. Konkrétně se jedná o hodnoty reálného HNP v bilionech dolarů, v cenách odpovídajících roku 1996, údaje jsou sezónně očištěné.

Data musíme nejdřív načíst. Soubor je bez hlavičky a obsahuje dva sloupce, v prvním sloupci je uvedeno období a ve druhém odpovídající hodnota HNP. Data načteme funkcí `read.table`, vytvoříme z nich časovou řadu a vykreslíme je do grafu.

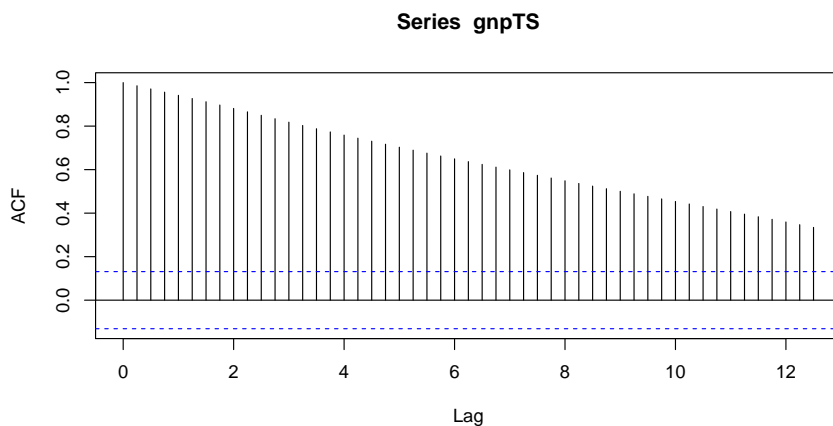
```
> fileDat <- paste(data.library, "gnp96.dat", sep = "")
> gnp96 <- read.table(fileDat)
> gnpTS <- ts(gnp96[, 2], start = 1947, frequency = 4)
> plot(gnpTS)
```



Obrázek 24: HNP USA v letech 1947-2002

Na první pohled je zřejmé, že v datech je obsažen výrazný trend, který ale přehluší většinu dalších charakteristik časové řady. Z obrázku například není vůbec patrné, že s časem roste také variabilita dat. Podívejme se na autokorelační funkci.

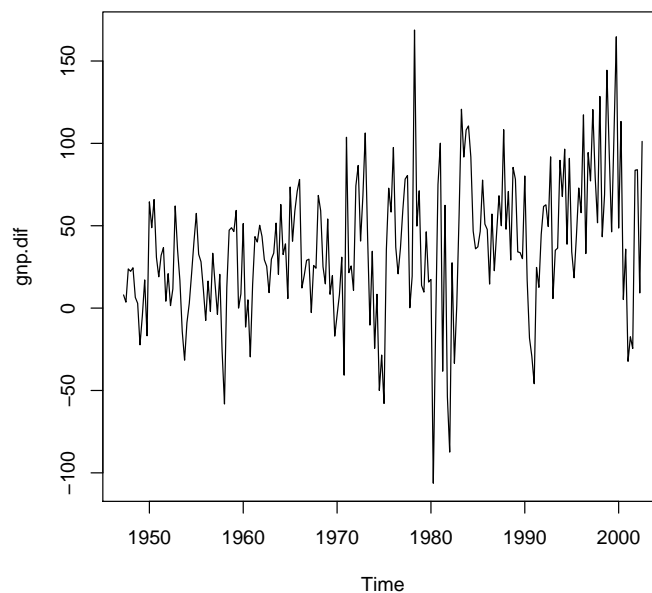
```
> acf(gnpTS, lag.max = 50)
```



Obrázek 25: Autokorelační funkce HNP USA v letech 1947-2002

Pokusme se odstranit trend pomocí diferencování.

```
> gnp.dif <- diff(gnpTS)
> plot(gnp.dif)
```



Obrázek 26: 1. diference HNP USA v letech 1947-2002

Z tohoto obrázku je již patrné, že variabilita druhé poloviny dat je vyšší. Navíc se zdá, že ani diferencováním se nám nepodařilo z řady úplně eliminovat trend.

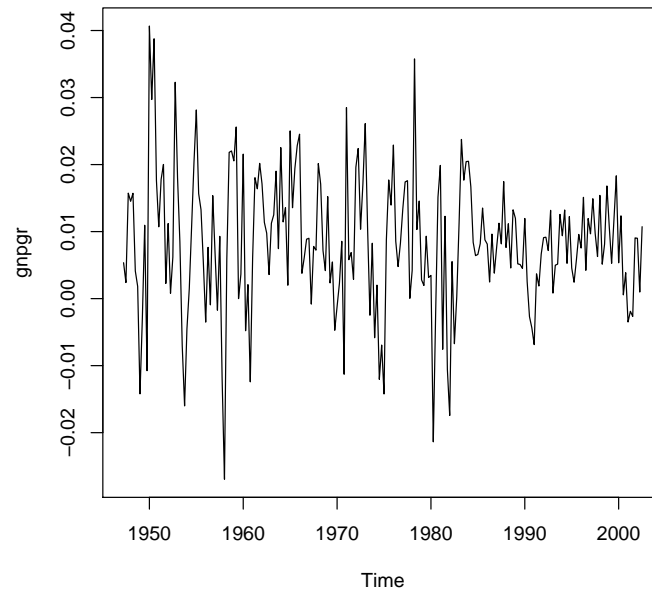
Proto zvolíme poněkud odlišný postup, který je v ekonometrii poměrně obvyklý. Zkoumanou časovou řadu, HNP USA, zlogaritmujeme a pro zlogaritmovaná data spočítáme 1. diference. Budeme tedy pracovat s časovou řadou

$$Z_t = \Delta \ln Y_t.$$

Takto vzniklá časová řada má zároveň dobrou interpretaci, protože jednotlivé údaje vyjadřují tempo růstu HNP vztahující se k danému období.

Podívejme se na obrázek nově vytvořené časové řady.

```
> gnpgr <- diff(log(gnpTS))
> plot(gnpgr)
```

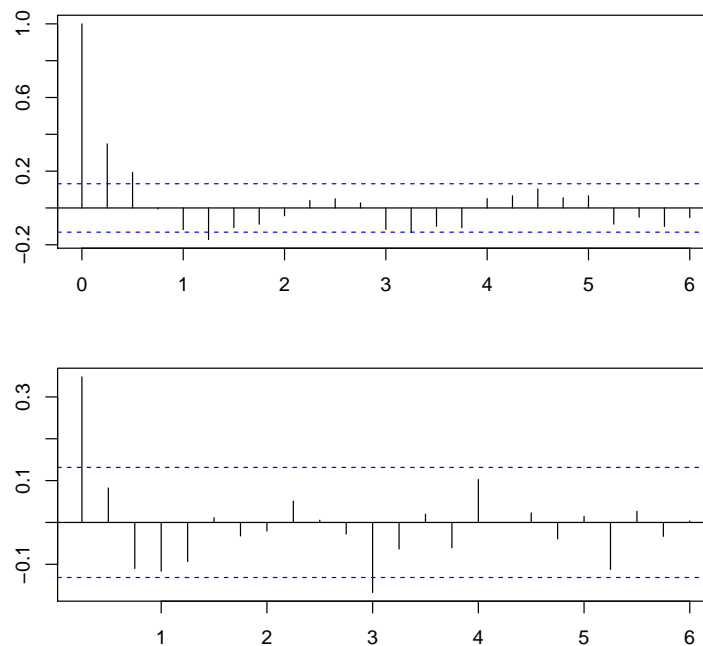


Obrázek 27: Tempa růstu HNP USA v letech 1947-2002

Časová řada se od pohledu jeví jako stacionární. Tomu by měla odpovídat také autokorelační a parciální autokorelační funkce, kdy by obě měly poměrně rychle klesat k nule.

```
> par(mfrow = c(2, 1), mar = c(2, 2, 3, 0) + 0.05)
> acf(gnpgr, lag.max = 24)
> pacf(gnpgr, lag.max = 24, main = "")
```

#### Series gnpgr



Obrázek 28: ACF a PACF pro tempa růstu HNP USA v letech 1947-2002

Z výše uvedených grafů ACF a PACF bychom mohli usoudit, že ACF se pro  $k > 2$  významně neliší od nuly, zatímco PACF klesá exponenciálně k nule. To by nás vedlo k názoru, že tempa růstu HNP se dají modelovat pomocí  $MA(2)$  procesu, a tudíž zlogaritmované hodnoty HNP můžeme modelovat jako  $IMA(1, 2)$  proces.

Grafy ACF a PACF ale můžeme zároveň interpretovat tak, že ACF klesá exponenciálně k nule a PACF se neliší významně od nuly pro  $k > 1$ . To by vedlo k modelování temp růstu pomocí  $AR(1)$  procesu a zlogaritmovaného HNP pomocí  $ARIMA(1, 1, 0)$  procesu.

Postupně odhadneme oba dva modely.

```
> ma.gnpgr <- arima(gnpgr, order = c(0, 0, 2))
> ma.gnpgr
```

```
Call:
arima(x = gnpgr, order = c(0, 0, 2))
```

```
Coefficients:
      ma1      ma2  intercept
 0.3028 0.2035    0.0083
s.e. 0.0654 0.0644    0.0010
```

```
sigma^2 estimated as 8.919e-05:  log likelihood = 719.96,  aic = -1431.93
```

```
> ar.gnpgr <- arima(gnpgr, order = c(1, 0, 0))
> ar.gnpgr
```

```
Call:
arima(x = gnpgr, order = c(1, 0, 0))
```

```
Coefficients:
      ar1  intercept
 0.3467    0.0083
s.e. 0.0627    0.0010
```

```
sigma^2 estimated as 9.03e-05:  log likelihood = 718.61,  aic = -1431.22
```

Odhadnuté modely jsou:

$$MA(2) : Z_t = 0.0083 + \varepsilon_t + 0.3028\varepsilon_{t-1} + 0.2035\varepsilon_{t-2}$$

$$AR(1) : Z_t = 0.0083 + 0.3467Z_t + \varepsilon_t$$

Funkce `arima` automaticky odhaduje ARIMA model se střední hodnotou, což se projeví v odhadnuté konstantě označené `intercept`. V tomto konkrétním případě je takový postup zcela na místě, protože průměrný růst HNP USA je kladný. Pokud bychom konstantu do modelu nezahrnuli, znamenalo by to, že předpokládáme průměrný růst HNP rovný 0.

Odhadli jsme dva modely. Pomocí analýzy reziduí, kterou se budeme podrobně zabývat v jednom z pozdějších cvičení, můžeme ověřit, zda jsou oba modely vyhovující. Nyní však

předpokládejme, že oba modely dobře popisují data. Naskytá se otázka, který z modelů je lepší, vhodnější?

Ve skutečnosti jsou oba modely velmi podobné. Dá se to snadno ověřit, pokud vyjádříme odhadnutý  $AR(1)$  proces pomocí ekvivalentní  $MA$  reprezentace (což je, jak víme, u kauzálních  $AR$  procesů možné). Budeme při tom uvažovat odhadnutý  $AR(1)$  model, v němž budeme prozatím ignorovat odhadnutou konstantu, a vyjádříme ho díky vlastnosti kauzality jako proces  $MA(\infty)$ .

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

Prvních deset koeficientů  $\psi_j$   $MA$  reprezentace kauzálního  $AR(1)$  procesu zjistíme příkazem `ARMAtoMA()`.

```
> ARMAtoMA(ar = 0.35, ma = 0, 10)
```

```
[1] 3.500000e-01 1.225000e-01 4.287500e-02 1.500625e-02 5.252187e-03
[6] 1.838266e-03 6.433930e-04 2.251875e-04 7.881564e-05 2.758547e-05
```

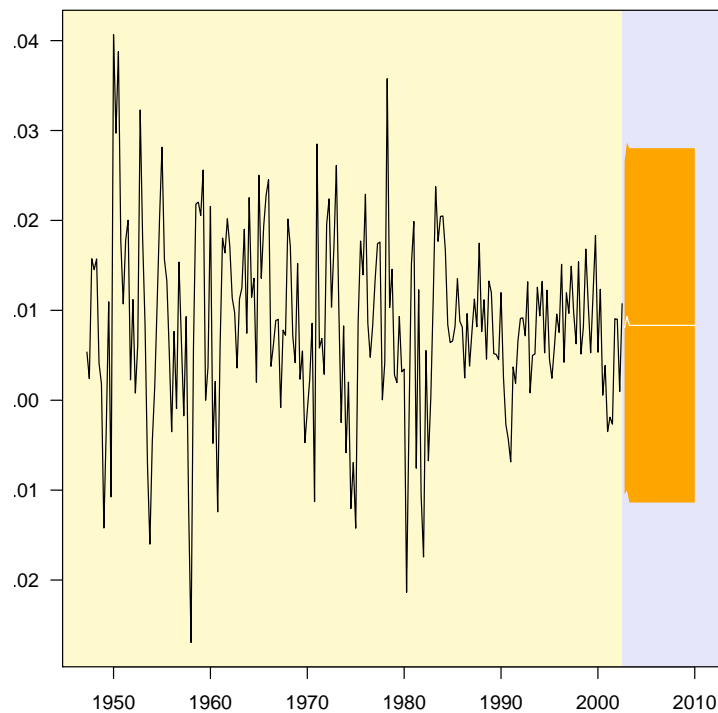
Odtud vidíme, že

$$Y_t \approx \varepsilon_t + 0.35\varepsilon_{t-1} + 0.12\varepsilon_{t-2},$$

což je velmi blízké odhadnutému  $MA(2)$  procesu.

Na závěr na základě odhadnutého  $MA(2)$  modelu spočítáme predikci pro tempa růstu amerického HNP do roku 2010.

```
> PlotPredictARIMA(gnpgr, ma.gnpgr, n.ahead = 30)
```



Obrázek 29: Predikce tempa růstu amerického HNP



## 2 Úkol:

1. Simulujte 300 hodnot  $IMA(2, 2)$  procesu s hodnotami parametrů  $\theta_1 = 0.2$ ,  $\theta_2 = -0.75$ 
  - (a) Vykreslete simulovanou časovou řadu.
  - (b) Vykreslete ACF a PACF.
  - (c) Odhadněte parametry  $IMA(2, 2)$  procesu ze simulovaných dat.
  - (d) Vykreslete predikci pro 10 budoucích hodnot.
  - (e) Spočítejte 1. diference simulované řady a prohlédněte si autokorelační funkci.
  - (f) Spočítejte 2. diference a opět se podívejte na autokorelační funkci.
2. Časová řada v datovém souboru `robot.dat` udává  $x$ -ové souřadnice odpovídající pozicím, v nichž se nacházel průmyslový robot po ukončení série operací. Data udávají odchylky od cílové pozice.
  - (a) Data načtěte a vykreslete.
  - (b) Prohlédněte si graf 1. diferencí.
  - (c) Pro původní (nediferencovaná) data odhadněte parametry modelu  $IMA(1,1)$  a vykreslete graf s predikcí pro 10 budoucích hodnot.