

SARIMA modely: modelovanie sezónnych časových radov

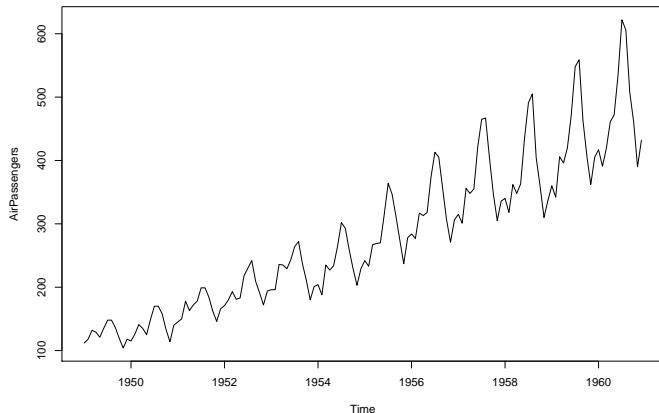
Beáta Stehlíková

2-EFM-102 Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

Obsah

- ▶ Modelovanie sezónnych dát: kvartálny HDP (sezónne neočistený), mesačný prietok riek, mesačný počet návštevníkov v turistických oblastiach



- ▶ Niekoľko teoretických príkladov so sezónnymi členmi: aký priebeh ACF a PACF môžeme pri nich očakávať
- ▶ Sezónne diferencovanie
- ▶ SARIMA modely

Príklady ARMA procesov so sezónnymi členmi, ich ACF PACF

Príklad 1: $x_t = \alpha x_{t-12} + u_t$

- ▶ Stacionarita: $(1 - \alpha L^{12})x_t = u_t \Rightarrow \alpha < 1$
- ▶ Prenásobíme x_{t-s} a spravíme strednú hodnotu \rightarrow rovnice pre autokovariancie:

$$\gamma(s) = \alpha \gamma(s - 12), \quad (s > 0)$$

$$\gamma(0) = \alpha \gamma(12) + \sigma^2$$

- ▶ Riešenie:

$$\text{disperzia: } \gamma(0) = \frac{\sigma^2}{1 - \alpha^2}$$

$$\gamma(12k) = \alpha^k \gamma(0), \quad \text{ostatné } \gamma(s) = 0$$

- ▶ ACF:

$$\rho(12k) = \alpha^k, \quad \text{ostatné } \rho(s) = 0$$

Proces $x_t = \alpha x_{t-12} + u_t$ v R-ku

Máme vlastne AR(12) proces s mnohými nulovými koeficientami:

$$x_t = \alpha x_{t-12} + u_t$$

$$x_t = 0x_{t-1} + \dots + 0x_{t-11} + \alpha x_{t-12} + u_t$$

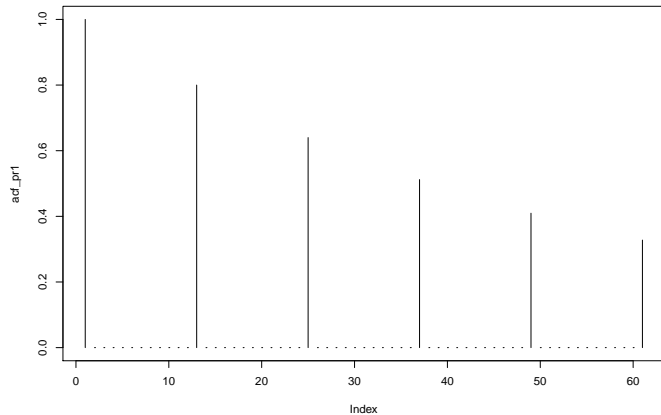
```
# ACF
```

```
acf_pr1 <- ARMAacf(ar = c(rep(0, 11), 0.8), lag.max = 60)  
plot(acf_pr1, type = "h")
```

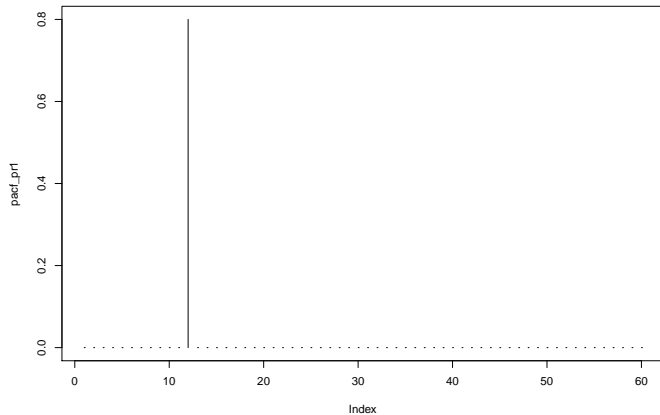
```
# PACF
```

```
pacf_pr1 <- ARMAacf(ar = c(rep(0, 11), 0.8), lag.max = 60,  
                    pacf = TRUE)  
plot(pacf_pr1, type = "h")
```

Autokorelačná funkcia:



Parciálna autokorelačná funkcia:



Príklad 2: $x_t = 0.8 x_{t-12} - 0.3 x_{t-24} + u_t$

- ▶ Stacionarita:

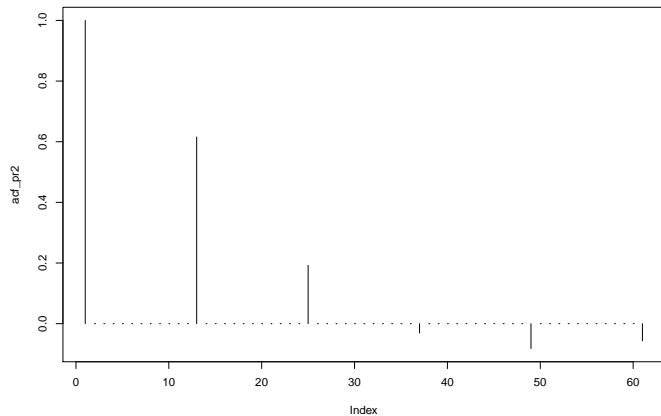
$$(1 - 0.8L^{12} + 0.3L^{24})x_t = u_t$$

substitúcia $L^{12} = y \rightarrow$ overujeme korene polynómu
 $1 - 0.8y + 0.3y^2$ (z každej hodnoty y dostaneme 12 koreňov L ,
 pre ktoré $|L|^{12} = |y|$)

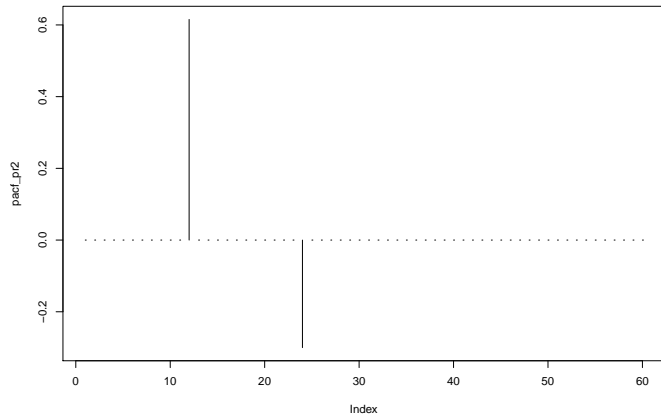
- ▶ Koeficienty do výpočtu ACF a PACF

```
ar_koef <- c(rep(0, 11), 0.8, rep(0, 11), -0.3)
acf_pr2 <- ARMAacf(ar = ar_koef, lag.max = 60)
pacf_pr2 <- ARMAacf(ar = ar_koef, lag.max = 60, pacf = TRUE)
```

Autokorelačná funkcia:



Parciálna autokorelačná funkcia:



Príklad 3: $(1 - 0.5L)(1 + 0.8L^{12})x_t = u_t$

- ▶ V SARIMA modeloch sa budú klasické a sezónne polynómy **násobiť**
- ▶ Máme teraz:

$$(1 - 0.5L)(1 + 0.8L^{12}) = 1 - 0.5L + 0.8L^{12} - 0.4L^{13},$$

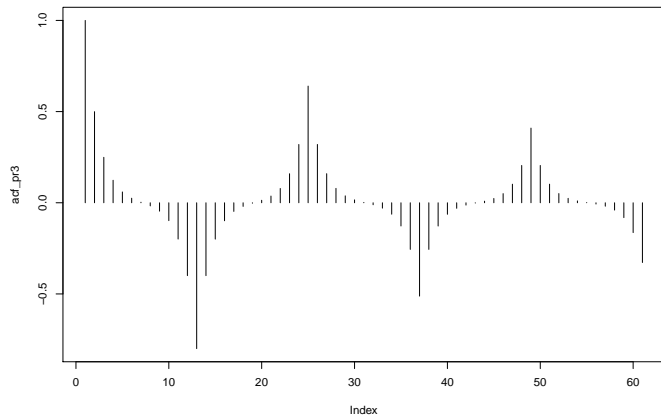
proces teda je

$$x_t = 0.5x_{t-1} - 0.8x_{t-12} + 0.4x_{t-13} + u_t$$

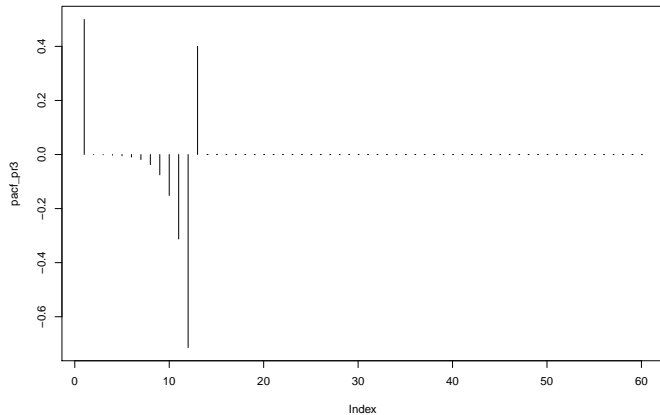
- ▶ Koeficienty do výpočtu ACF a PACF

```
ar_koef <- c(0.5, rep(0, 10), -0.8, 0.4)
acf_pr3 <- ARMAacf(ar = ar_koef, lag.max = 60)
pacf_pr3 <- ARMAacf(ar = ar_koef, lag.max = 60, pacf = TRUE)
```

Autokorelačná funkcia:



Parciálna autokorelačná funkcia:

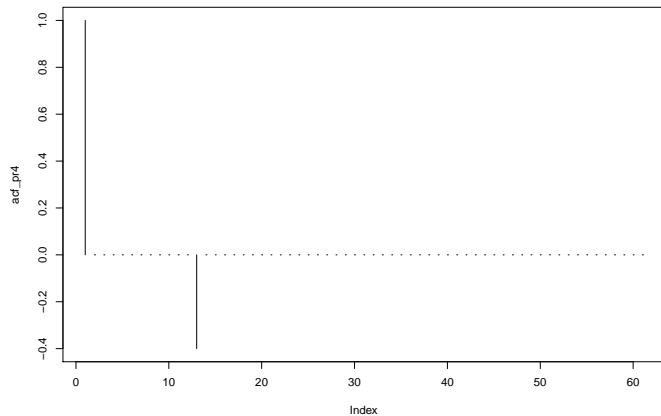


Príklad 4: $x_t = u_t - 0.5u_{t-12}$

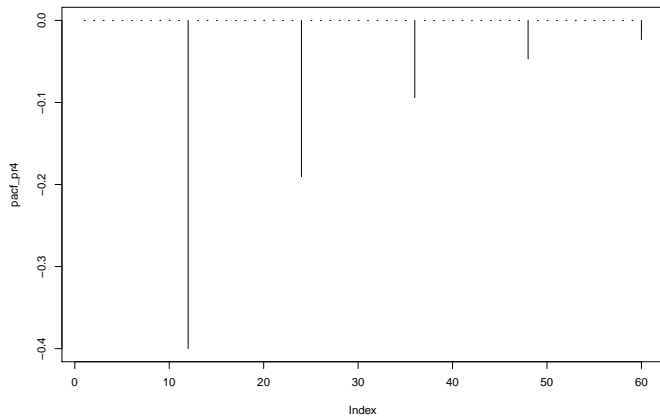
- ▶ Teoretická ACF: jediná nenulová hodnota je pre lag 12
- ▶ Koeficienty do výpočtu ACF a PACF

```
ma_koef <- c(rep(0, 11), -0.5)
acf_pr4 <- ARMAacf(ma = ma_koef, lag.max = 60)
pacf_pr4 <- ARMAacf(ma = ma_koef, lag.max = 60, pacf = TRUE)
```


Autokorelačná funkcia:



Parciálna autokorelačná funkcia:



Príklad 5: $x_t = 0.8x_{t-12} + u_t - 0.5u_{t-1}$

Cvičenie. Odvodte ACF pre všeobecný model tohto tvaru:

Example 3.47 A Mixed Seasonal Model

Consider an $\text{ARMA}(0, 1) \times (1, 0)_{12}$ model

$$x_t = \Phi x_{t-12} + w_t + \theta w_{t-1},$$

where $|\Phi| < 1$ and $|\theta| < 1$. Then, because x_{t-12} , w_t , and w_{t-1} are uncorrelated, and x_t is stationary, $\gamma(0) = \Phi^2 \gamma(0) + \sigma_w^2 + \theta^2 \sigma_w^2$, or

$$\gamma(0) = \frac{1 + \theta^2}{1 - \Phi^2} \sigma_w^2.$$

In addition, multiplying the model by x_{t-h} , $h > 0$, and taking expectations, we have $\gamma(1) = \Phi \gamma(11) + \theta \sigma_w^2$, and $\gamma(h) = \Phi \gamma(h - 12)$, for $h \geq 2$. Thus, the ACF for this model is

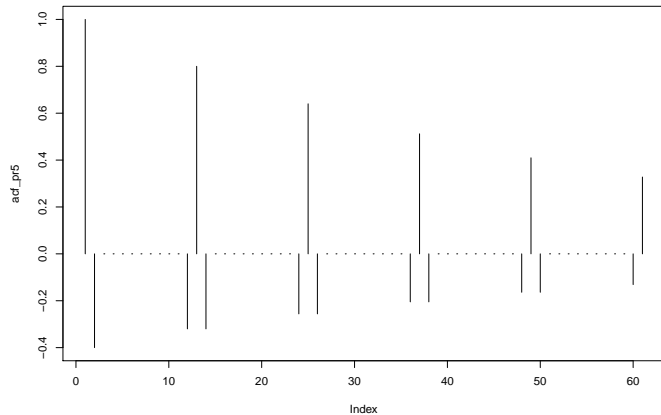
$$\begin{aligned} \rho(12h) &= \Phi^h \quad h = 1, 2, \dots \\ \rho(12h - 1) &= \rho(12h + 1) = \frac{\theta}{1 + \theta^2} \Phi^h \quad h = 0, 1, 2, \dots, \\ \rho(h) &= 0, \quad \text{otherwise.} \end{aligned}$$

```
#  $x_t = 0.8 x_{t-12} + u_t - 0.5 u_{t-1}$ 
```

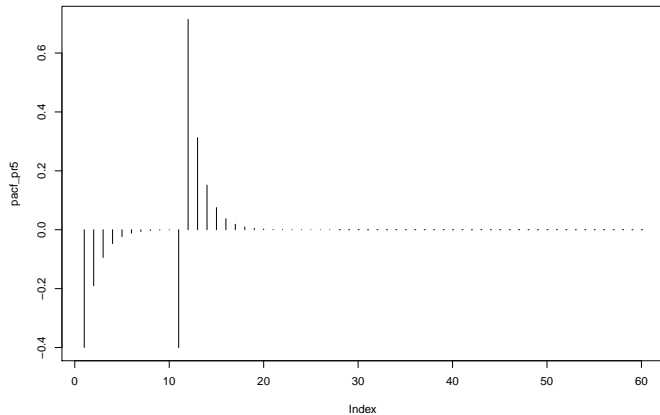
```
acf_pr5 <- ARMAacf(ar = c(rep(0, 11), 0.8),  
                  ma = c(-0.5),  
                  lag.max = 60)
```

```
pacf_pr5 <- ARMAacf(ar = c(rep(0, 11), 0.8),  
                   ma = c(-0.5),  
                   lag.max = 60,  
                   pacf = TRUE)
```

Autokorelačná funkcia:



Parciálna autokorelačná funkcia:



Sezónne diferencovanie

Vzorový príklad

- ▶ Predstavme si model pre časový rad x_t :

$$x_t = S_t + u_t$$

so sezónnou zložkou S_t a bielym šumom u_t

- ▶ Sezónnu zložku modelujeme ako

$$S_t = S_{t-12} + w_t,$$

kde w je biely šum nezávislý od u

- ▶ Potom náš proces x je

$$\begin{aligned} x_t &= S_t + u_t = (S_{t-12} + w_t) + u_t \\ &= (x_{t-12} - u_{t-12} + w_t) + u_t \\ (1 - L^{12})x_t &= w_t + u_t - u_{t-12} \end{aligned}$$

- ▶ Náš proces teda nie je stacionárny kvôli jednotkovému koreňu (prakticky to vidíme aj na ACF, že klesá príliš pomaly)
- ▶ V takomto prípade zoberieme **sezónne diferencie** $x_t - x_{t-12}$
- ▶ V R-ku `diff(x, lag = 12)`

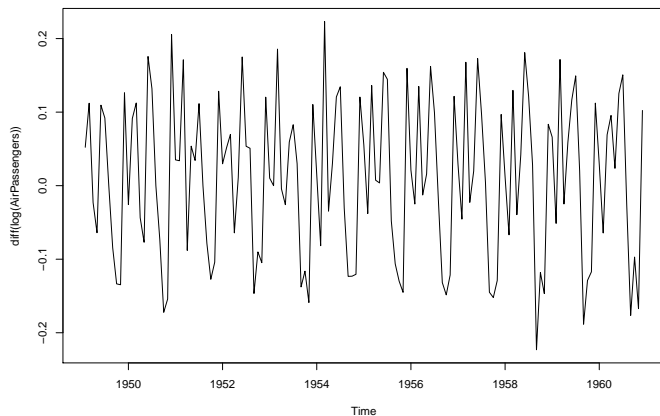
Príklad sezónneho diferencovania

- ▶ Zoberieme dáta o počte cestujúcich aerolinkami od Boxa a Jenkinsa
- ▶ Zlogaritmujeme, aby sme odstránili rastúcosť disperzie
- ▶ Zdiferencujeme kvôli trendu

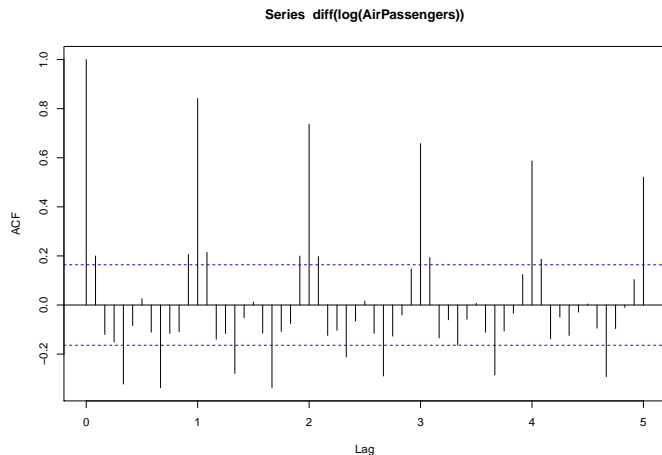
Pozrieme sa pre tieto diferencie na:

- ▶ priebeh dát
- ▶ výberovú ACF
- ▶ ADF test

```
plot(diff(log(AirPassengers)))
```



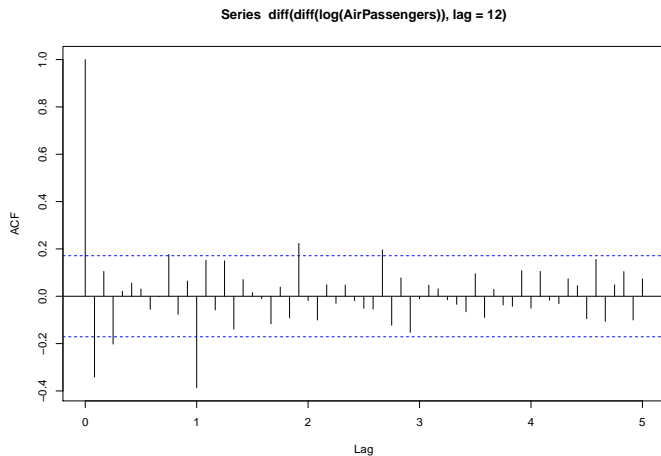
```
acf(diff(log(AirPassengers)), lag.max = 60)
```



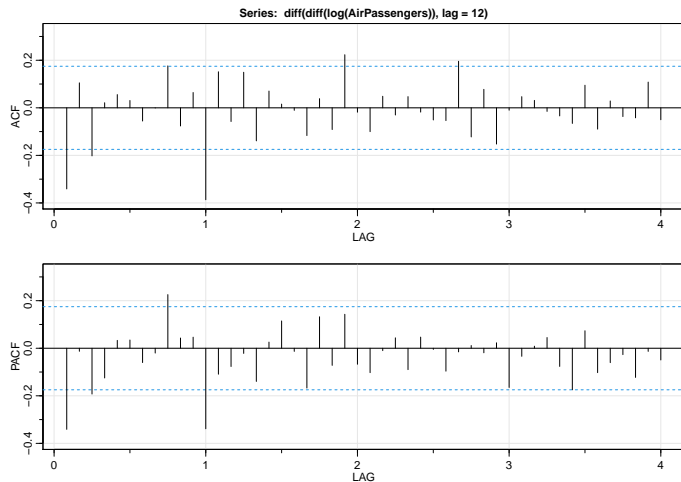
```
library(urca)
# kritická hodnota je -1.95
ur.df(diff(log(AirPassengers)),
      type = "none", lags = 12, selectlags = "BIC")

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration
## #####
##
## The value of the test statistic is: -1.2228
```

```
acf(diff(diff(log(AirPassengers)), lag = 12), lag.max = 60)
```



```
library(astsa)
acf2(diff(diff(log(AirPassengers)), lag = 12))
```



SARIMA modely

- ▶ Dáta môžu byť klasicky a/alebo sezónne diferencované
- ▶ AR členy
 - ▶ klasické, napr. $(1 - \alpha_1 L)x_t, (1 - \alpha_1 L - \alpha_2 L^2)x_t, \dots$
 - ▶ sezónne, napr. $(1 - \phi_1 L^{12})x_t, (1 - \phi_1 L^{12} - \phi_2 L^{24})x_t, \dots$
 - ▶ vynásobia sa
- ▶ MA členy
 - ▶ klasické, napr. $(1 - \beta_1 L)u_t, (1 - \beta_1 L - \beta_2 L^2)u_t, \dots$
 - ▶ sezónne, napr. $(1 - \theta_1 L^{12})u_t, (1 - \theta_1 L^{12} - \theta_2 L^{24})u_t, \dots$
 - ▶ vynásobia sa

Príklady

Príklad 1. Model pre mesačné dáta:

$$(1 - \alpha_1 L - \alpha_2 L^2)(1 - \psi L^{12})y_t = (1 - \beta_1 L)(1 - \theta_1 L^{12})u_t,$$

kde y_t sú diferencie pôvodných dát x_t :

$$y_t = (1 - L)x_t$$

Príklad 2. Model pre kvartálne dáta:

$$(1 - \alpha_1 L)(1 - \psi L^4)y_t = (1 - \beta_1 L)(1 - \theta_1 L^4 - \theta_2 L^8)u_t,$$

kde y_t vzniknú z pôvodných dát x_t diferencovaním a následným sezónnym diferencovaním:

$$y_t = (1 - L^4)(1 - L)x_t$$

Terminológia SARIMA modelov

- ▶ Pripomeňme si $ARIMA(p, d, q)$ modely:
 - ▶ p = počet AR členov
 - ▶ d = koľkokrát dáta diferencujeme
 - ▶ q = počet MA členov
- ▶ $SARIMA(p, d, q) \times (P, D, Q)_s$ modely majú navyše
 - ▶ P = počet sezónnych AR členov (vo výstupe z R-ka `sar1`, `sar2`, ...)
 - ▶ D = koľkokrát dáta sezónne diferencujeme
 - ▶ Q = počet sezónnych MA členov (vo výstupe z R-ka `sma1`, `sma2`, ...)
 - ▶ s = perióda dát
- ▶ $SARIMA(p, d, q) \times (P, D, Q)_s$ model v R-ku:

```
sarima(data, p, d, q, P, D, Q, s) #model  
sarima.for(data, N, p, d, q, P, D, Q, s) # predikcie
```

Príklad 1. Model pre mesačné dáta:

$$(1 - \alpha_1 L - \alpha_2 L^2)(1 - \psi L^{12})y_t = (1 - \beta_1 L)(1 - \theta_1 L^{12})u_t,$$

kde y_t sú diferencie pôvodných dát x_t :

$$y_t = (1 - L)x_t$$

Je to

$$SARIMA(2, 1, 1) \times (1, 0, 1)_{12}$$

Príklad 2. Model pre kvartálne dáta:

$$(1 - \alpha_1 L)(1 - \psi L^4)y_t = (1 - \beta_1 L)(1 - \theta_1 L^4 - \theta_2 L^8)u_t,$$

kde y_t vzniknú z pôvodných dát x_t diferencovaním a následným sezónnym diferencovaním:

$$y_t = (1 - L^4)(1 - L)x_t$$

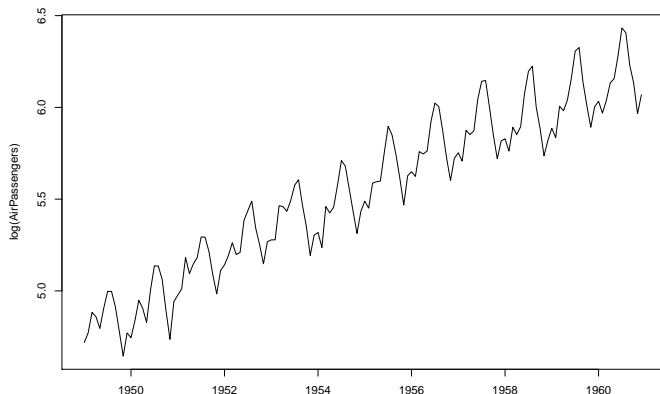
Je to

$$SARIMA(1, 1, 1) \times (1, 1, 2)_4$$

SARIMA modely pre reálne dáta

Príklad 1: počet cestujúcich

Vo videu dokončíme hľadanie SARIMA modelu pre $\log(\text{AirPassengers})$ (zatiaľ vieme, že dáta budeme diferencovať klasicky aj sezónne):



Príklad 2: ceny kurčiat (z balíka astsa)

Potom nájdeme model pre mesačné ceny kurčiat:

```
plot(chicken)
```

