Cvičenie 5

Almonovej model

Jeden z problémov pri predikcii časových radov lineárnym modelom je silná lineárna závislosť medzi jednotlivými hodnotami posunutými v čase $x_t, x_{t-1}, x_{t-2,}$ .... Aby sa táto závislosť odstránila, pôvodné dáta sa transformujú nahradením parametrov β_m. Pri polynominálnom modeli Almonovej sa *M* + 1 parametrov $\beta_0, \beta_1, \beta_2, \dots, \beta_M$ nahradí za menší počet *K*+1 parametrov $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_K$ dosadením polynómu:

$$\beta_j = \alpha_0 + \alpha_1.j + \alpha_2.j^2 + \alpha_3.j^3 + \cdots + \alpha_K.j^K$$

Napr. majme lineárny model s časovým oknom o veľkosti *M*=3:

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3}$$

Tento model chceme transformovať polynómom stupňa *K* = 2, tzn. parametre $\beta_0, \beta_1, \beta_2, \beta_3$ nahradíme za $\alpha_0, \alpha_1, \alpha_2$, kde pre transformované parametre platí:

$$\beta_j = \alpha_0 + \alpha_1.j + \alpha_2.j^2$$

Po dosadení dostaneme:

$$\begin{aligned}
\beta_0 &= \alpha_0 \\
\beta_1 &= \alpha_0 + \alpha_1 + \alpha_2 \\
\beta_2 &= \alpha_0 + \alpha_1 2 + \alpha_2 2^2 = \alpha_0 + \alpha_1 2 + \alpha_2 4 \\
\beta_3 &= \alpha_0 + \alpha_1 3 + \alpha_2 3^2 = \alpha_0 + \alpha_1 3 + \alpha_2 9
\end{aligned} \tag{1}$$

Ak transformované hodnoty $\beta_0, \beta_1, \beta_2, \beta_3$ dosadíme do pôvodného modelu a zlúčime členy pre jednotlivé parametre $\alpha_0, \alpha_1, \alpha_2$ dostaneme transformovaný model:

$$y_t = \alpha_0(x_t + x_{t-1} + x_{t-2} + x_{t-3}) + \alpha_1(x_{t-1} + 2x_{t-2} + 3x_{t-3}) + \alpha_2(x_{t-1} + 4x_{t-2} + 9x_{t-3})$$

Členy pri transformovaných parametroch $\alpha_0, \alpha_1, \alpha_2$ si označíme ako nové vstupné atribúty z_0, z_1, z_2, pre ktoré platí:

$$\begin{aligned}
z_0 &= x_t + x_{t-1} + x_{t-2} + x_{t-3} \\
z_1 &= x_{t-1} + 2x_{t-2} + 3x_{t-3} \\
z_2 &= x_{t-1} + 4x_{t-2} + 9x_{t-3}
\end{aligned}$$

Po dosadení transformovaných dát dostaneme lineárny model:

$$y_t = \alpha_0 z_0 + \alpha_1 z_1 + \alpha_2 z_2$$

z ktorého môžeme priamo vypočítať transformované parametre $\alpha_0, \alpha_1, \alpha_2$ pomocou funkcie lm. Vypočítané parametre $\alpha_0, \alpha_1, \alpha_2$ potom môžeme spätne dosadiť do (1) a vypočítame pôvodné parametre $\beta_0, \beta_1, \beta_2, \beta_3$.

Načítame si dáta USMacroG a vytvoríme model pre predikovanie dpi podľa atribútu consumption s časovým oknom M=3 a veľkosťou polynómu K=2. Najprv skonvertujeme časové rady na číselné vektory.

```
data("USMacroG", package="AER")
y = as.numeric(USMacroG[,"dpi"])
x = as.numeric(USMacroG[,"consumption"])
```

Vytvoríme si vektory pre posunuté atribúty x_t, x_{t-1}, x_{t-2}, x_{t-3} a y_t:

```
N = length(x)
x_t3 = x[1:(N-3)]
x_t2 = x[2:(N-2)]
x_t1 = x[3:(N-1)]
x_t = x[4:N]
y_t = y[4:N]
```

Vypočítame si transformované atribúty z_0, z_1, z_2:

```
z0 = x_t + x_t1  + x_t2 + x_t3
z1 = x_t1 + (2*x_t2) + (3*x_t3)
z2 = x_t1 + (4*x_t2) + (9*x_t3)
```

Pomocou funkcie lm si vypočítame parametre $\alpha_0, \alpha_1, \alpha_2$. Pre model bez posunu je potrebné vo výraze lm zadať medzi závislé parametre 0.

```
td = data.frame(y_t, z0, z1, z2)
td_lm = lm(y_t  ~ 0 + z0 + z1 + z2, td)
alfa = as.numeric(coefficients(td_lm))
```

Z parametrov $\alpha_0, \alpha_1, \alpha_2$ spätne vypočítame $\beta_0, \beta_1, \beta_2, \beta_3$:

```
beta0 = alfa[1]
beta1 = alfa[1] + alfa[2] + alfa[3]
beta2 = alfa[1] + 2*alfa[2] + 4*alfa[3]
beta3 = alfa[1] + 3*alfa[2] + 9*alfa[3]
```

Koyckov model

Ak by sme uvažovali model s nekonečným časovým oknom:

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots$$

Tak by sa priamo nedal vypočítať bez dodatočných obmedzení (mali by sme nekonečne veľa parametrov β_k). Pri Koyckovom modeli nahradíme parametre β_k podľa vzťahu $\beta_k = b\lambda^k$, pričom aby sa model dal spočítať pre nekonečnú postupnosť $\beta_1, \beta_2, \beta_3$, … pre b musí platiť $0 < |\lambda| < 1$ aby bola geometrická postupnosť $b + b\lambda + b\lambda^2 + b\lambda^3 + \cdots$ klesajúca. Transformovaný model je potom možné zapísať v tvare:

$$y_t = a + bx_t + \lambda y_{t-1}$$

Tzn. namiesto výpočtu nekonečného počtu parametrov sa z dát vypočítajú iba parametre a, b a λ. Pre výpočet Koyckovho modelu je možné priamo použiť funkciu dynlm, napr. pre predikciu "dpi" podľa "consumption":

```
library("dynlm")
model = dynlm(dpi ~ consumption + L(dpi,1), USMacroG)
```

Úlohy na cvičení

1. Vytvorte Almonovej model s časovým oknom *M*=5 a polynómom stupňa *K*=3 pre predikovanie časového radu "dpi" na základe radu "consumption". Vypočítajte predikciu modelu na trénovacích dátach a zobrazte skutočnú hodnotu "dpi" a predikovanú hodnotu na spoločnom grafe.
2. Vypočítajte korelačnú maticu medzi atribútmi x_t, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5} z predchádzajúceho príkladu a vyhodnoťte, ktoré dvojice sú silno korelované.
3. Vytvorte štandardný autoregresný model pre predikovanie časového radu "dpi" na základe "consumption" s časovým oknom *M*=5:

   $$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \beta_4 x_{t-4} + \beta_5 x_{t-5}$$

   Porovnajte presnosť modelu s Almonovej modelom. Porovnajte vypočítané koeficienty podľa štandardného modelu a po Almonovej transformácii.
4. Vytvorte Koyckov model pre predikciu časového radu "gdp" od hodnoty "invest". Vypočítajte predikciu modelu na trénovacích dátach a zobrazte ich na spoločnom grafe s priebehom "gdp".