

## M0130 – 7. PRAKTIKUM : M0130pr07 (Modely lokálního postupného trendu)

### A. Klouzavé průměry

Klouzavé průměry aproximují neznámý trend **polynomem stupně  $m$**  uvnitř **symetrického vyhlazovacího okénka**  $[t - s, t + s]$  a pro náhodný proces  $\{Y_t, t \in \mathbb{Z}\}$  uvažují model:

$$Y_{t+\tau} = \sum_{j=0}^m \beta_j(t) \tau^j + \varepsilon_{t+\tau} \quad \tau \in [-s, s]; \quad E\varepsilon_{t+\tau} = 0; \quad D\varepsilon_{t+\tau} = \sigma^2; \quad C(\varepsilon_{t+\tau}, \varepsilon_{t+\tau'}) = 0; \quad \tau \neq \tau'.$$

$$\text{Označme: } \mathbf{X} = \begin{pmatrix} 1 & -s & \dots & (-s)^m \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0^m \\ \vdots & \vdots & \dots & \vdots \\ 1 & s & \dots & s^m \end{pmatrix} \Rightarrow \mathbf{H} = \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\text{proj. matice}} = \begin{pmatrix} \mathbf{h}'_1 \\ \vdots \\ \mathbf{h}'_{s+1} \\ \vdots \\ \mathbf{h}'_{2s+1} \end{pmatrix}$$

Rozdělme časovou řadu na tři segmenty:

$$\mathbf{Y}_{(F)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_s \\ \vdots \\ Y_{2s+1} \end{pmatrix}, \quad \mathbf{Y}_{(t)} = \begin{pmatrix} Y_{t-s} \\ \vdots \\ Y_t \\ \vdots \\ Y_{t+s} \end{pmatrix} \quad \text{a} \quad \mathbf{Y}_{(L)} = \begin{pmatrix} Y_{n-2s} \\ \vdots \\ Y_{n-s} \\ \vdots \\ Y_n \end{pmatrix}.$$

$$\hat{Y}_1 = \mathbf{h}'_1 \mathbf{Y}_{(F)}$$

$\vdots$

$$\hat{Y}_s = \mathbf{h}'_s \mathbf{Y}_{(F)}$$

Pak  $\hat{Y}_t = \mathbf{h}'_{s+1} \mathbf{Y}_{(t)}$  pro  $t = s+1, \dots, n-s$ , přičemž prvky projekční matice  $\mathbf{H}$  se nazývají váhy.

$$\hat{Y}_{n-s+1} = \mathbf{h}'_{s+2} \mathbf{Y}_{(L)}$$

$\vdots$

$$\hat{Y}_n = \mathbf{h}'_{2s+1} \mathbf{Y}_{(L)}$$

Většina statistických programů pracuje pouze s prostředním segmentem a první a poslední část časové řady nechává nevyhlazenou.

Uvažujme například  $m = 2, s = 2, r = 2s + 1 = 5$ , pak

$$\text{matice plánu } \mathbf{X} = \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \mathbf{x}'_3 \\ \mathbf{x}'_4 \\ \mathbf{x}'_5 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}, \quad \text{informační matice } \mathbf{X}'\mathbf{X} = \begin{pmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{pmatrix}$$

a projekční matice

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}'_1 \\ \mathbf{h}'_2 \\ \mathbf{h}'_3 \\ \mathbf{h}'_4 \\ \mathbf{h}'_5 \end{pmatrix} = \begin{pmatrix} \frac{31}{35} & \frac{9}{35} & \frac{-3}{35} & \frac{-5}{35} & \frac{3}{35} \\ \frac{9}{35} & \frac{13}{35} & \frac{12}{35} & \frac{6}{35} & \frac{-5}{35} \\ \frac{-3}{35} & \frac{12}{35} & \frac{17}{35} & \frac{12}{35} & \frac{-3}{35} \\ \frac{-5}{35} & \frac{6}{35} & \frac{12}{35} & \frac{13}{35} & \frac{9}{35} \\ \frac{3}{35} & \frac{-5}{35} & \frac{-3}{35} & \frac{9}{35} & \frac{3}{35} \end{pmatrix} \quad \begin{array}{l} \text{váhy pro první bod (asymetrické váhy)} \\ \text{váhy pro druhý bod (asymetrické váhy)} \\ \text{váhy pro "středové" body (symetrické váhy)} \\ \text{váhy pro předposlední bod (asymetrické váhy)} \\ \text{váhy pro poslední bod (asymetrické váhy)} \end{array}$$

Uvažujme ještě  $m = 3, s = 2, r = 2s + 1 = 5$ , pak matice plánu

$$\mathbf{X} = \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \\ x'_4 \\ x'_5 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix}, \text{ informační matice } \mathbf{X}'\mathbf{X} = \begin{pmatrix} 5 & 0 & 10 & 0 \\ 0 & 10 & 0 & 34 \\ 10 & 0 & 34 & 0 \\ 0 & 34 & 0 & 130 \end{pmatrix}$$

a projekční matice

$$\mathbf{H} = \begin{pmatrix} h'_1 \\ h'_2 \\ h'_3 \\ h'_4 \\ h'_5 \end{pmatrix} = \begin{pmatrix} \frac{69}{70} & \frac{2}{35} & \frac{-3}{35} & \frac{2}{35} & \frac{-1}{70} \\ \frac{2}{35} & \frac{27}{35} & \frac{12}{35} & \frac{-8}{35} & \frac{2}{35} \\ \frac{-3}{35} & \frac{12}{35} & \frac{17}{35} & \frac{12}{35} & \frac{-3}{35} \\ \frac{2}{35} & \frac{-8}{35} & \frac{12}{35} & \frac{27}{35} & \frac{2}{35} \\ \frac{-1}{70} & \frac{2}{35} & \frac{-3}{35} & \frac{2}{35} & \frac{69}{70} \end{pmatrix} \begin{array}{l} \text{váhy pro první bod (asymetrické váhy)} \\ \text{váhy pro druhý bod (asymetrické váhy)} \\ \text{váhy pro "středové" body (symetrické váhy)} \\ \text{váhy pro předposlední bod (asymetrické váhy)} \\ \text{váhy pro poslední bod (asymetrické váhy)} \end{array}$$

Vidíme, že

- Součet vah v jednom řádku je roven jedné (jde o prvky projekční matice s jednotkovou normou).
- Středové váhy jsou symetrické kolem střední hodnoty ( $\mathbf{H}$  je symetrická).
- Je-li  $m$  sudé, pak "středové" váhy řádu  $m$  a  $m + 1$  pro stejnou délku  $r = 2s + 1$  jsou totožné (součty s lichými mocninami jsou nulové).

### Jednoduché klouzavé průměry

Pro  $m = 0, r = 2s + 1 \Rightarrow \mathbf{X} = (1, \dots, 1)'$ ,  $\mathbf{X}'\mathbf{X} = 2s + 1$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \frac{1}{2s+1}\mathbf{X}\mathbf{X}' = \frac{1}{2s+1} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$$

Pro  $m = 1, r = 2s + 1 \Rightarrow \mathbf{X}' = \begin{pmatrix} 1 & \dots & 0 & \dots & 1 \\ -s & \dots & 0 & \dots & s \end{pmatrix}$ ,  $\mathbf{X}'\mathbf{X} = \begin{pmatrix} 2s+1 & 0 \\ 0 & \frac{s(s+1)(2s+1)}{3} \end{pmatrix}$ ,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{pmatrix} h'_1 \\ \vdots \\ h'_{s+1} \\ \vdots \\ h'_{2s+1} \end{pmatrix} = \begin{pmatrix} h'_1 \\ \vdots \\ \frac{1}{2s+1} \mathbf{1}'_{2s+1} \\ \vdots \\ h'_{2s+1} \end{pmatrix}$$

### Centrované klouzavé průměry

Využívají se v případě, kdy chceme k vyhlazení použít okénko **sudé délky**, například u ročních či čtvrtletních časových řad, kdy potřebujeme časovou řadu očistit od sezónních vlivů.

1. krok •  $p_1 = \frac{1}{2s}(Y_{t-s} + \dots + Y_{t+s-1})$  a  $p_2 = \frac{1}{2s}(Y_{t-s+1} + \dots + Y_{t+s})$
2. krok •  $p = \frac{1}{2}(p_1 + p_2) = \frac{1}{4s}(Y_{t-s} + 2Y_{t-s+1} \dots + 2Y_{t+s-1} + Y_{t+s})$

PŘÍKLAD 1

Pro použití klouzavých průměrů si zvolíme měsíční časovou řadu s počty nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002. Nejprve načteme popis časové řady, který je v datovém souboru umístěn na prvních dvou řádcích. Následně načteme samotnou časovou řadu pomocí příkazu `scan`, volbou `skip3` zajistíme přeskočení prvních dvou řádků s textem a jednoho volného řádku.

```
> fileDat <- paste(data.library, "MontlyUnemplWomen.dat", sep = "")
> con <- file(fileDat)
> (popis <- readLines(con, n = 2))
```

```
[1] "Monthly unemployed young women between ages 16 and 19"
[2] "in the U.S. from January 1961 to August 2002"
```

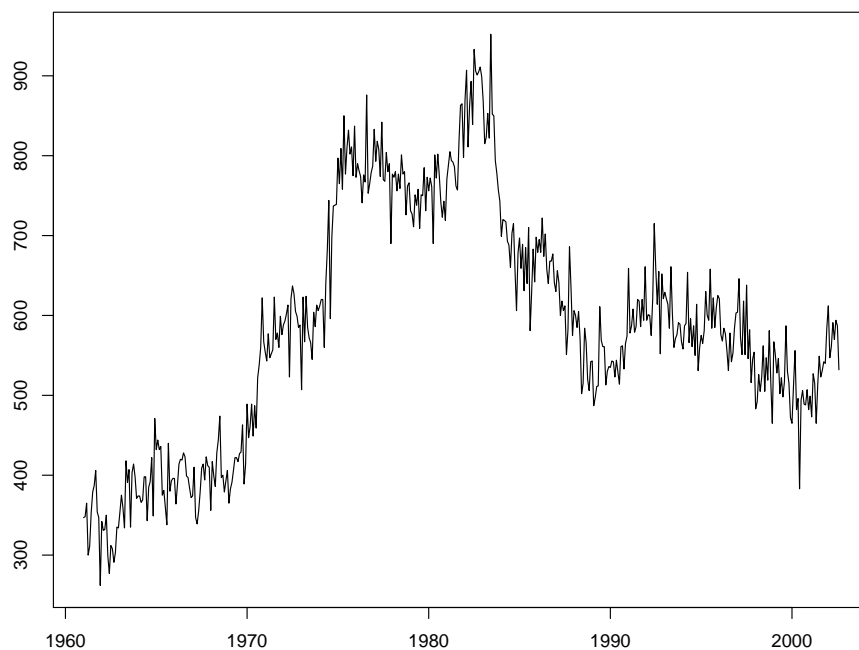
```
> close(con)
> UnemplWomen <- scan(fileDat, skip = 3)
```

Z načtených dat vytvoříme časovou řadu a vykreslíme ji.

```
> UnemplWomenTS <- ts(UnemplWomen, start = 1961, frequency = 12)
> str(UnemplWomenTS)
```

```
Time-Series [1:500] from 1961 to 2003: 347 348 365 300 311 353 379 387 406 354 ...
```

```
> par(mar = c(2, 2, 0, 0) + 0.5)
> plot(UnemplWomenTS)
```



Obrázek 1: Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002

Klouzavé průměry se nejčastěji používají k odfiltrování sezónních vlivů. K tomu bychom však potřebovaly, aby klouzavé průměry měly šířku 12, což není liché číslo. Proto použijeme jednoduché centrované klouzavé průměry.

V prostředí R máme k dispozici příkaz `filter()`, ve kterém musíme zadat váhy (tj. prostřední řádek  $\mathbf{h}'_{s+1}$  projekční matice  $\mathbf{H}$ ). Proto nejprve vytvoříme funkci `CenterFilter()`, která spočítá váhy pro centrované klouzavé průměry pro sudé  $k$ .

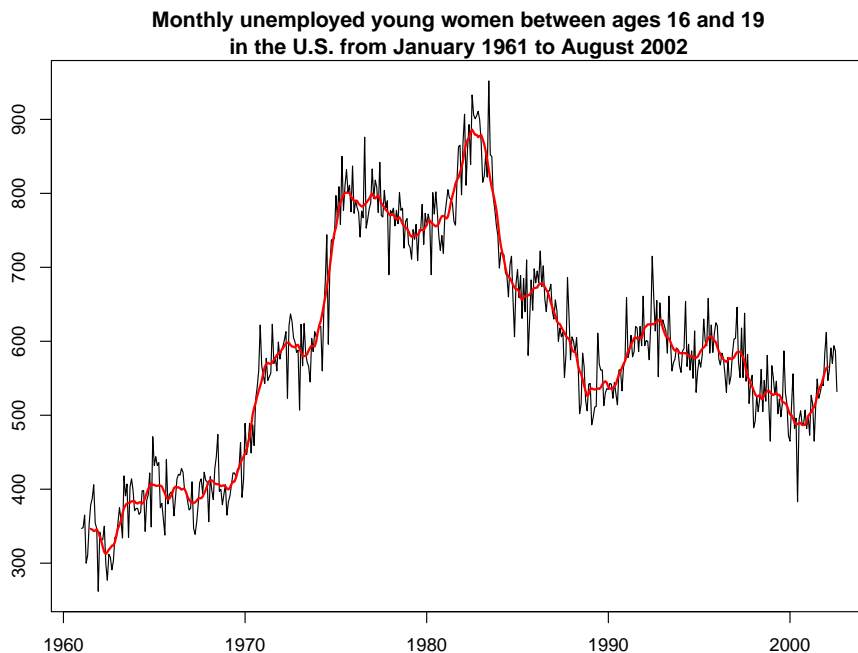
```
> CenterFilter <- function(k) {
  if (k%2 != 0)
    stop("k must be a odd number")
  centerFilter <- c(1/(2 * k), rep(1/k, k - 1), 1/(2 * k))
}
```

A nyní již můžeme provést vyhlazení (*smoothing*) časové řady pomocí centrovaných klouzavých průměrů a výsledky graficky znázornit.

```
> x <- UnemplWomenTS
> k <- 12
> xSmooth <- filter(x, CenterFilter(k))
> str(xSmooth)
```

```
Time-Series [1:500] from 1961 to 2003: NA NA NA NA NA ...
```

```
> par(mar = c(2, 2, 2, 0) + 0.5)
> plot(x)
> lines(xSmooth, col = "red", lwd = 2)
> txt <- paste(popis[1], popis[2], sep = "\n")
> title(main = txt)
```



Obrázek 2: Použití centrovaných klouzavých průměrů pro měsíční data *Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002*

## B. Exponenciální vyrovnávání

Na rozdíl od klouzavých průměrů vychází z polynomiální lokální **vážené** metody nejmenších čtverců, kde váhy jednotlivých čtverců se směrem do minulosti exponenciálně snižují – odtud název metody.

Uvažujeme-li vyhlazovací okno pouze směrem do minulosti, pak pro každé  $t, \tau = 0, 1, \dots$  dostaneme regresní model tvaru

$$Y_{t-\tau} = \sum_{j=0}^m (-\tau)^j \beta_j(t) + \varepsilon_{t-\tau}, \quad E\varepsilon_{t-\tau} = 0; \quad E\varepsilon_q \varepsilon_s = 0, \quad q \neq s; \quad D\varepsilon_{t-\tau} = \alpha^{-\tau} \sigma^2; \quad \alpha \in (0, 1).$$

tj. matice vah je rovna

$$\mathbf{W} = \text{diag}\{w_1, \dots, w_n, \dots\} = \text{diag}\{\alpha^0, \alpha^1, \dots, \alpha^\tau \dots\}.$$

Odhad parametrů  $\beta$  metodou **nejmenších vážených čtverců** (neboť rozptyly nejsou konstantní) je dán vzorcem:

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$$

kde

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \begin{pmatrix} \sum_{\tau=0}^{\infty} \alpha^\tau & \sum_{\tau=0}^{\infty} (-\tau)^1 \alpha^\tau & \dots & \sum_{\tau=0}^{\infty} (-\tau)^m \alpha^\tau \\ \sum_{\tau=0}^{\infty} (-\tau)^1 \alpha^\tau & \sum_{\tau=0}^{\infty} (-\tau)^2 \alpha^\tau & \dots & \sum_{\tau=0}^{\infty} (-\tau)^{m+1} \alpha^\tau \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{\tau=0}^{\infty} (-\tau)^m \alpha^\tau & \sum_{\tau=0}^{\infty} (-\tau)^{m+1} \alpha^\tau & \dots & \sum_{\tau=0}^{\infty} (-\tau)^{2m} \alpha^\tau \end{pmatrix}, \quad \mathbf{X}'\mathbf{W}\mathbf{Y} = \begin{pmatrix} \sum_{\tau=0}^{\infty} \alpha^\tau Y_{t-\tau} \\ \sum_{\tau=0}^{\infty} (-\tau)^1 \alpha^\tau Y_{t-\tau} \\ \vdots \\ \sum_{\tau=0}^{\infty} (-\tau)^m \alpha^\tau Y_{t-\tau} \end{pmatrix}.$$

Tento přístup založený na vážené polynomiální metodě nejmenších čtverců se nazývá *Brownův přístup*.

### ZNAČENÍ:

Pro dobrou srozumitelnost zavedeme následující značení. Necht'  $\{Y_t, t \in \mathbb{Z}\}$  je náhodná posloupnost, její realizace v časových okamžicích  $t_1, t_2, \dots, t_n$  označme  $y_1, y_2, \dots, y_n$ . Symbolem  $\hat{y}_{t|k}$  označme odhad hodnoty  $Y_t$  v čase  $t$  na základě hodnot do časového okamžiku  $k$  včetně.

- Jestliže  $k < t$ , pak  $\hat{y}_{t|k}$  nazýváme **predikcí**,
- pokud  $k = t$ ,  $\hat{y}_{t|t}$  nazýváme **filtrací**
- a je-li  $k = n > t$ , pak  $\hat{y}_{t|n}$  nazýváme **vyrovnáním** (*smoothing*).

### Jednoduché exponenciální vyrovnávání

Exponenciální vyrovnávání pro  $m = 0$  se nazývá jednoduché exponenciální vyrovnávání.

Použijeme-li označení  $\hat{\beta}_0(t) = b_0(t)$  a uvážíme-li, že pro  $\alpha \in (0, 1)$  je  $\sum_{\tau=0}^{\infty} \alpha^\tau = \frac{1}{1-\alpha}$ , dostaneme

$$b_0(t) \sum_{\tau=0}^{\infty} \alpha^\tau = \sum_{\tau=0}^{\infty} \alpha^\tau Y_{t-\tau} \quad \Rightarrow \quad b_0(t) = \hat{Y}_t = (1 - \alpha) \sum_{\tau=0}^{\infty} \alpha^\tau Y_{t-\tau}$$

Abychom získali rekurentní vztah, upravujeme

$$\begin{aligned}\hat{Y}_t &= (1 - \alpha) \sum_{\tau=0}^{\infty} \alpha^\tau Y_{t-\tau} = (1 - \alpha) Y_t + (1 - \alpha) \sum_{\tau=1}^{\infty} \alpha^\tau Y_{t-\tau} = \left| \begin{array}{l} \text{subst.} \\ k = \tau - 1 \end{array} \right| \\ &= (1 - \alpha) Y_t + (1 - \alpha) \sum_{k=0}^{\infty} \alpha^{k+1} Y_{t-1-k} \\ &= (1 - \alpha) Y_t + \alpha (1 - \alpha) \underbrace{\sum_{k=0}^{\infty} \alpha^k Y_{t-1-k}}_{\hat{Y}_{t-1}} = (1 - \alpha) Y_t + \alpha \hat{Y}_{t-1}\end{aligned}$$

Protože predikce o  $\tau$  ( $\tau > 0$ ) kroků dopředu pro jednoduché exponenciální vyrovnávání je rovna  $\hat{Y}_{t+\tau|t} = \hat{Y}_t = b_0(t)$ , můžeme předchozí rekurentní vztah přepsat pro realizace a dále upravovat

$$\begin{aligned}\hat{y}_{t+1|t} &= (1 - \alpha) y_t + \alpha \hat{y}_{t|t-1} \\ &= (1 - \alpha) y_t + \alpha \hat{y}_{t|t-1} + \hat{y}_{t|t-1} - \hat{y}_{t|t-1} \\ &= \hat{y}_{t|t-1} + (1 - \alpha) \underbrace{(y_t - \hat{y}_{t|t-1})}_{\text{chyba predikce } \hat{\varepsilon}_{t|t-1}}\end{aligned}$$

a o rekurentním vzorci s chybou predikce  $\hat{\varepsilon}_{t|t-1}$  se říká, že je ve formě korekce chyby předpovědi (*error correction form*).

### Ad hoc přístupy Holta a Winterse

Pokud chceme na základě pozorování  $y_1, \dots, y_t$  sestavit předpověď budoucí hodnoty  $y_{t+1}$  v čase  $t + 1$ , označme ji  $\hat{y}_{t+1|t}$ , pak nejjednodušším odhadem může být obyčejný průměr. Tato předpověď je vhodná, pokud hodnoty časové řady náhodně kolísají kolem střední hodnoty, která se v čase nemění. Jako rozumnější se však jeví použít pro predikci budoucí hodnoty ve větší míře pozorování, která jsou časově nejbližší. Pak se nabízejí vážené průměry

$$\hat{y}_{t+1|t} = \sum_{j=0}^{t-1} w_{j,t} y_{t-j}, \quad (1)$$

kde součet vah je roven jedné, tj.  $\sum_{j=0}^n w_{j,t} = 1$ .

Exponenciální vyrovnávání je založeno na myšlence použití vah, které do minulosti klesají exponenciálně.

S využitím vztahu

$$\sum_{j=0}^{t-1} \alpha^j = \frac{1 - \alpha^t}{1 - \alpha}, \quad \text{pro } \alpha \in (0, 1), \quad (2)$$

chceme-li, aby součet vah, které exponenciálně klesají, byl roven jedné, položíme

$$w_{j,t} = \frac{1 - \alpha}{1 - \alpha^t} \alpha^j. \quad (3)$$

Protože pro  $t \rightarrow \infty$  konvergují váhy  $w_{j,t} \rightarrow w_j = (1 - \alpha) \alpha^j$ , můžeme uvažovat jednokrokovou předpověď ze všech minulých pozorování ve tvaru

$$\hat{y}_{t+1|t} = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j y_{t-j}, \quad \text{pro } \alpha \in (0, 1). \quad (4)$$

Analogicky jako u Brownova přístupu odvodíme rekurentní vztahy

$$\begin{aligned}\widehat{y}_{t+1|t} &= (1 - \alpha)y_t + (1 - \alpha) \sum_{j=1}^{\infty} \alpha^j y_{t-j} \\ &= (1 - \alpha)y_t + \alpha(1 - \alpha) \sum_{k=0}^{\infty} \alpha^k y_{t-1-k} \\ &= (1 - \alpha)y_t + \alpha \widehat{y}_{t|t-1}\end{aligned}$$

Obdobně získáme i tvar využívající korekci chyby předpovědi

$$\begin{aligned}\widehat{y}_{t+1|t} &= (1 - \alpha)y_t + \alpha \widehat{y}_{t|t-1} + \widehat{y}_{t|t-1} - \widehat{y}_{t|t-1} \\ &= \widehat{y}_{t|t-1} + (1 - \alpha)(y_t - \widehat{y}_{t|t-1}) \\ &= \widehat{y}_{t|t-1} + (1 - \alpha)\widehat{\varepsilon}_{t|t-1}\end{aligned}$$

Na tomto ad-hoc přístupu se nám podařilo ukázat, že se v podstatě jedná o **jednoduché exponenciální vyrovnávání**, které přepokládá model

$$Y_t = \beta_0(t) + \varepsilon_t$$

s lokální hladinou  $\beta_0(t)$ .

Použijeme-li značení obvyklá pro tento přístup, kdy váhy mají tvar

$$w_j = \beta(1 - \beta)^j, \quad (5)$$

tj.  $\alpha = 1 - \beta$ , místo  $\widehat{\beta_0(t)}$ , píšeme  $L_t$  (level). Odvozené vztahy v novém značení:

$$\widehat{y}_{t+1|t} = \beta y_t + (1 - \beta)\widehat{y}_{t|t-1} = \widehat{y}_{t|t-1} + \beta \widehat{\varepsilon}_{t|t-1} \quad (6)$$

$$L_{t+1} = L_t + \beta \widehat{\varepsilon}_{t+1|t} \quad (7)$$

### Holtovo exponenciální vyrovnávání

Oproti jednoduchému exponenciálnímu vyrovnávání Holtova metoda předpokládá lokálně lineární trend, jehož koeficienty  $\beta_0(t)$  i  $\beta_1(t)$  se v čase mění. Hodnota časové řady v okamžiku  $t$  je určena jednak její úrovní  $\beta_0(t)$ , jednak směrnici  $\beta_1(t)$ . V Holtově metodě se úroveň v čase  $t$  značí symbolem  $L_t$  (zkratka pro *level*) a směrnice jako  $T_t$  (zkratka pro *trend*).

Úroveň  $L_t$  je zároveň vyrovnanou hodnotou realizace  $y_t$  v okamžiku  $t$ . Směrnice lokálně lineárního trendu  $T_t$  (někdy se mluvíme krátce o trendu) vyjadřuje očekávanou změnu úrovně časové řady při jednotkové časové změně. Pokud chceme pomocí Holtovy metody přepovídat hodnotu časové řady o  $h > 0$  jednotek dopředu, položíme

$$\widehat{y}_{t+h|t} = L_t + T_t h. \quad (8)$$

Takže, je-li  $h = 1$ , dostaneme jednokrokovou předpověď jako

$$\widehat{y}_{t+1|t} = L_t + T_t. \quad (9)$$

Protože by přibližně mělo platit, že realizace  $y_{t+1} \approx L_{t+1}$ , pak se jeví vhodné získat  $L_{t+1}$ , jako konvexní lineární kombinaci hodnot  $(L_t + T_t)$  a  $y_{t+1}$ . V Holtově metodologii bývá zvykem místo  $\alpha \in (0, 1)$  používat  $\beta = 1 - \alpha$ , takže konvexní lineární kombinace bude mít tvar

$$L_{t+1} = (1 - \beta)(L_t + T_t) + \beta y_{t+1} . \quad (10)$$

Hodnota  $\beta$  se nazývá *vyrovnávací konstanta pro úroveň řady*.

Analogickou úvahu použijeme i pro směrnici trendu  $T_t$ . Z předpokladu, že řada má lokálně lineární trend vyplývá, že by přibližně mělo platit

$$T_{t+1} \approx T_t,$$

ale zároveň má také smysl očekávat, že směrnice trendu je přibližně rozdílem sousedních úrovní, tj.

$$T_{t+1} \approx L_{t+1} - L_t .$$

Novou hodnotu směrnice  $T_{t+1}$  budeme uvažovat jako konvexní lineární kombinaci

$$T_{t+1} = (1 - \gamma)T_t + \gamma(L_{t+1} - L_t), \quad \text{kde } \gamma \in (0, 1) \quad (11)$$

$\gamma$  je tzv. *vyrovnávací konstanta pro lineární růst (pro směrnici)*.

Na závěr odstavce ještě ukážeme přepsání předchozích rekurentních vztahů do chybového tvaru.

$$\begin{aligned} L_{t+1} &= (1 - \beta)(L_t + T_t) + \beta y_{t+1} = (1 - \beta)(L_t + T_t) + \beta y_{t+1} + \beta \widehat{y}_{t+1|t} - \beta \widehat{y}_{t+1|t} \\ &= \underbrace{\beta (y_{t+1} - \widehat{y}_{t+1|t})}_{\widehat{\varepsilon}_{t+1|t}} + (1 - \beta)(L_t + T_t) + \underbrace{\beta \widehat{y}_{t+1|t}}_{L_t + T_t} \\ &= \beta \widehat{\varepsilon}_{t+1|t} + L_t + T_t \end{aligned}$$

$$\begin{aligned} T_{t+1} &= (1 - \gamma)T_t + \gamma(L_{t+1} - L_t) = T_t - \gamma T_t + \gamma \underbrace{L_{t+1}}_{L_t + T_t + \beta \widehat{\varepsilon}_{t+1|t}} - \gamma L_t \\ &= T_t - \gamma T_t + \gamma(L_t + T_t + \beta \widehat{\varepsilon}_{t+1|t}) - \gamma L_t \\ &= T_t + \gamma \beta \widehat{\varepsilon}_{t+1|t} . \end{aligned}$$

### Holtovo-Wintersovo exponenciální vyrovnávání

V případě, kdy časová řada má sezonní charakter, nevystačíme se žádnou z předchozích metod. Rozšíření Holtovy metody na sezónní časové řady je známo jako **Holtova-Wintersova metoda**. Autorem je Holtův student Peter R. Winters.

Holtova-Wintersova metoda je založena na třech vyrovnávacích konstantách. Jedna je pro hladinu, druhá pro trend a třetí pro sezónnost. Dle charakteru dat využívá aditivní nebo multiplikatívni notaci.



Uvažujme časovou řadu s lokálně lineárním trendem a sezónností s periodou  $p \geq 2$ . Stejně jako u Holtovy metody označme symbolem  $L_t$  úroveň v čase  $t$ , symbolem  $T_t$  směrnici lokálně lineárního trendu a symbolem  $S_t$  sezónní výkyv čase  $t$ . Součet úrovně  $L_t$  s hodnotou sezónního výkyvu  $S_t$  představuje v okamžiku  $t$  vyrovnanou hodnotu realizace  $y_t$ . Předpověď hodnoty časové řady o  $h > 0$  jednotek dopředu je pak dána vztahem

$$\hat{y}_{t+h|t} = L_t + S_{t-p+h} + T_t h, \quad (12)$$

takže v případě jednokrokové predikce platí

$$\hat{y}_{t+1|t} = L_t + S_{t+1-p} + T_t \quad (13)$$

Protože by mělo přibližně platit

$$y_{t+1} \approx L_{t+1} + S_{t+1-p}$$

a

$$L_{t+1} \approx L_t + T_t,$$

má smysl získat úroveň  $L_{t+1}$  jako konvexní lineární kombinaci hodnot  $(L_t + S_t)$  a  $(y_{t+1} - S_{t+1-p})$ , tj.

$$L_{t+1} = (1 - \beta)(L_t + T_t) + \beta(y_{t+1} - S_{t+1-p}). \quad (14)$$

Protože řada má lokálně lineární trend, mělo by přibližně platit

$$T_{t+1} \approx T_t,$$

ale zároveň lze směrnici lokálně lineárního trendu vyjádřit pomocí rozdílu sousedních hladin

$$T_{t+1} \approx L_{t+1} - L_t.$$

Oba předchozí vztahy využijeme při konstrukci směrnice lokálně lineárního trendu díky konvexní lineární kombinaci

$$T_{t+1} = (1 - \gamma)T_t + \gamma(L_{t+1} - L_t),$$

kde  $\gamma \in (0, 1)$  se nazývá *vyrovnávací konstanta pro směrnici trendu*. Pro sezónní výkyvy musí platit vztah

$$S_{t+1} \approx S_{t+1-p},$$

a také

$$S_{t+1} \approx y_{t+1} - L_{t+1}$$

Tedy označíme-li symbolem  $\delta \in (0, 1)$  *vyrovnávací konstantu pro sezónní výkyvy*, pak

$$S_{t+1} = (1 - \delta)S_{t+1-p} + \delta(y_{t+1} - L_{t+1})$$

Na závěr odstavce odvodíme rekurentní vztahy v chybové formě. Tedy upravujeme

$$\begin{aligned}
 L_{t+1} &= (1 - \beta)(L_t + T_t)\beta(y_{t+1} - S_{t+1-p}) \\
 &= (1 - \beta)(L_t + T_t) + \beta(y_{t+1} - S_{t+1-p}) + \beta\hat{y}_{t+1|t} - \beta\hat{y}_{t+1|t} \\
 &= \beta(y_{t+1} - \hat{y}_{t+1|t}) + L_t + T_t - \beta L_t + -\beta T_t - \beta S_{t+1-p} + \beta \underbrace{\hat{y}_{t+1|t}}_{L_t + S_{t+1-p} + T_t} \\
 &= L_t + T_t + \beta\hat{\varepsilon}_{t+1|t}
 \end{aligned}$$

$$\begin{aligned}
 T_{t+1} &= (1 - \gamma)T_t + \gamma(L_{t+1} - L_t) = T_t - \gamma T_t + \gamma \underbrace{L_{t+1}}_{L_t + T_t + \beta\hat{\varepsilon}_{t+1|t}} - L_t \\
 &= T_t - \gamma T_t + \gamma(L_t + T_t + \beta\hat{\varepsilon}_{t+1|t} - \gamma L_t) \\
 &= T_t + \gamma\beta\hat{\varepsilon}_{t+1|t}
 \end{aligned}$$

$$\begin{aligned}
 S_{t+1} &= (1 - \delta)S_{t+1-p} + \delta(y_{t+1} - L_{t+1}) \\
 &= S_{t+1-p} - \delta(S_{t+1-p} - y_{t+1}) - \delta(L_t + T_t + \beta\hat{\varepsilon}_{t+1|t}) \\
 &= S_{t+1-p} + \delta y_{t+1} - \delta(L_t + T_t + S_{t+1-p}) - \delta\beta\hat{\varepsilon}_{t+1|t} \\
 &= S_{t+1-p} + \delta(1 - \beta)\hat{\varepsilon}_{t+1|t}.
 \end{aligned}$$

## Exponenciální vyrovnávání v prostředí R

V prostředí R je pro exponenciální vyhlazování v balíčku `stats` k dispozici funkce `HoltWinters()`, která například v případě aditivního modelu uvažuje rekurentní vztahy

úroveň ( <i>level</i> )	• $L_t = \alpha(y_t - S_{t-p}) + (1 - \alpha)(L_{t-1} + T_{t-1})$
lineární růst ( <i>growth</i> )	• $T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$
sezónní výkyvy ( <i>seasonal</i> )	• $S_t = \gamma(y_t - L_{t-1} - T_{t-1}) + (1 - \gamma)S_{t-p}$
předpověď ( <i>forecast</i> )	• $\hat{y}_{t+h t} = L_t + T_t h + S_{t-p+h_p^+}$

kde  $h_p^+ = [(h - 1) \bmod p] + 1$ .

Počáteční stavy  $L_0, T_0, S_{1-p}, \dots, S_0$  a tzv. vyrovnávací konstanty  $\alpha, \beta, \gamma$  jsou odhadnuty na základě dat. Na podrobný popis funkce se podívejte pomocí příkazu `?HoltWinters`.

Mnohem komplexnější možnosti nabízí funkce `ets()` z balíčku `forecast`, která navíc dovoluje tzv. tlumící faktor.

Tak například, označíme-li

$$\mu_t = \hat{y}_t,$$

pak model s **tlumeným lineárním trendem**, kdy

$$\mu_t = \hat{y}_t = L_{t-1} - T_{t-1}$$

a který je ve formě korekce chyby predikce (*error correction form*), bude definován rekurzivními vztahy

$$\begin{aligned}y_t &= L_{t-1} + \phi T_{t-1} + \varepsilon_t \\L_t &= L_{t-1} + \phi T_{t-1} + \alpha \varepsilon_t \\T_t &= \phi T_{t-1} + \beta(L_t - L_{t-1} - \phi T_{t-1}) = \phi T_{t-1} + \alpha \beta \varepsilon_t\end{aligned}$$

Optimální model je vybrán na základě tzv. AIC kritéria

$$AIC = -2 \log(\text{Likelihood}) + 2m,$$

kde  $m$  je počet parametrů.

Při výstupu procedura `ets()` používá následující notaci

Trendová komponenta	Sezónní komponenta		
	$N$ (None)	$A$ (Additive)	$M$ (Multiplikative)
$N$ (None)	$N, N$	$N, A$	$N, M$
$A$ (Additive)	$A, N$	$A, A$	$A, M$
$A_d$ (Additive damped)	$A_d, N$	$A_d, A$	$A_d, M$
$M$ (Multiplikative)	$M, N$	$M, A$	$M, M$
$M_d$ (Multiplikative damped)	$M_d, N$	$M_d, A$	$M_d, M$

Označení optimálního modelu tvoří trojici  $ETS(E, T, S)$ , kde

$E$	<i>error</i>	možné hodnoty	$A, M$
$T$	<i>trend</i>		$N, A, A_d, M, M_d$
$S$	<i>seasonal</i>		$N, A, M$

### PŘÍKLAD 1 (pokračování)

Na načtená data vyzkoušíme Holtův–Wintersenův model se všemi komponentami a výsledky exponenciálního vyrovnávání vypíšeme.

```
> x <- UnemplWomenTS
> model <- HoltWinters(x)
> summary(model)
```

```
      Length Class  Mode
fitted    1952  mts   numeric
x          500   ts   numeric
alpha      1  -none- numeric
beta       1  -none- numeric
gamma      1  -none- numeric
coefficients 14  -none- numeric
seasonal   1  -none- character
SSE        1  -none- numeric
call      2  -none- call
```

```
> coefficients(model)
```

```

      a          b          s1          s2          s3          s4
559.3306093  0.3683431 11.8684563  5.4251826  3.3297314  2.1699366
      s5          s6          s7          s8          s9          s10
12.7939957  2.0405548 11.2096531  6.7428758 -1.0626908  0.7839631
      s11         s12
13.3458782  0.1804263

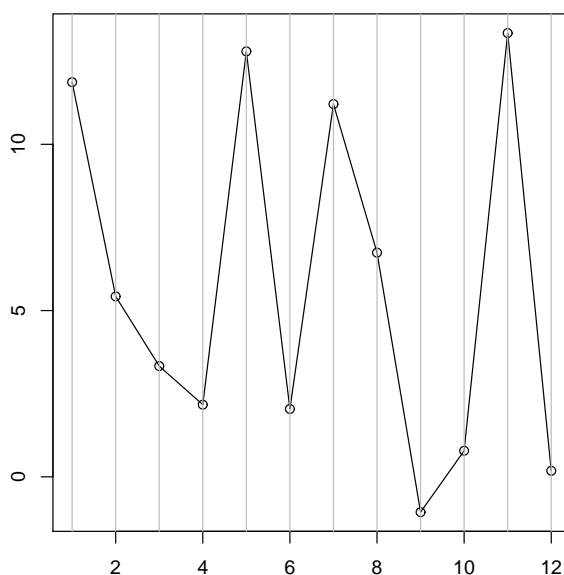
```

Hodnoty sezónních složek vykreslíme do grafu.

```

> par(mar = c(2, 2, 0, 0) + 0.5)
> plot(1:12, coefficients(model)[3:14], type = "o")
> for (k in 1:12) abline(v = k, col = "gray", lty = 1)

```



Obrázek 3: Hodnoty sezónních složek z Holtova–Wintersonova exponenciálního vyrovnání pomocí funkce `HoltWinters()` pro časovou řadu: *Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002*

Na základě tohoto modelu provedeme predikci na dva roky dopředu. Výsledky Holtova–Wintersonova exponenciálního vyrovnání i predikci vykreslíme.

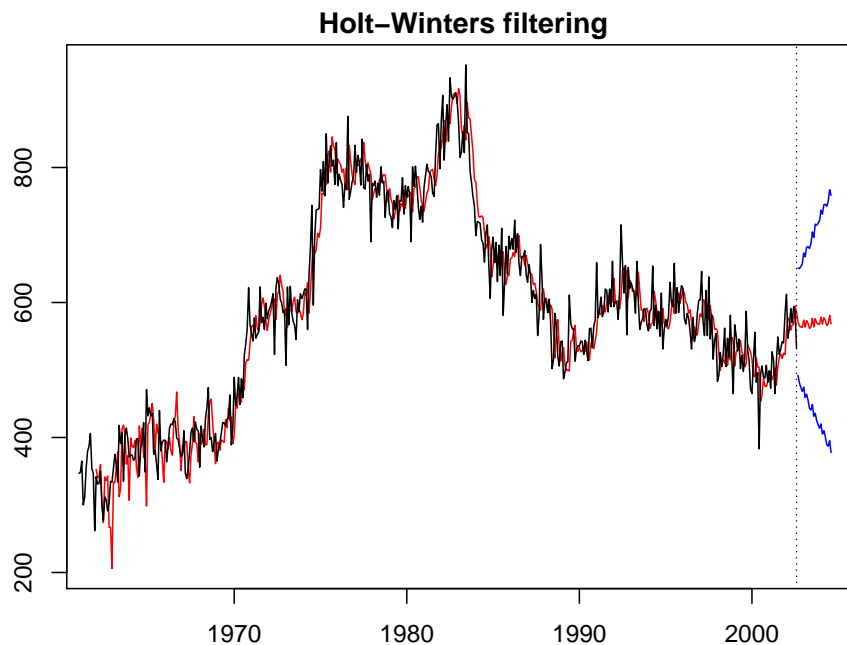
```

> pred <- predict(model, n.ahead = 24, prediction.interval = T)
> summary(pred)

```

fit	upr	lwr
Min. :561.6	Min. :650.1	Min. :377.8
1st Qu.:565.1	1st Qu.:678.4	1st Qu.:408.8
Median :568.3	Median :709.5	Median :433.7
Mean :569.7	Mean :705.5	Mean :433.9
3rd Qu.:573.6	3rd Qu.:738.2	3rd Qu.:460.5
Max. :581.1	Max. :767.1	Max. :492.0

```
> par(mar = c(2, 2, 1, 0) + 0.5)
> plot(model, predicted.values = pred)
```



Obrázek 4: Holtovo–Wintersonovo exponenciální vyrovnávání pomocí funkce `HoltWinters()` pro časovou řadu: *Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002*

Podíváme se, jak dopadnou výsledky v případě použití funkce `ets()` z balíčku `forecast`.

```
> library(forecast)
> x <- UnemplWomenTS
> modelETS <- ets(x)
> summary(modelETS)
```

ETS(A,N,N)

Call:  
ets(y = x)

Smoothing parameters:  
alpha = 0.3982

Initial states:  
l = 346.2374

sigma: 36.3603

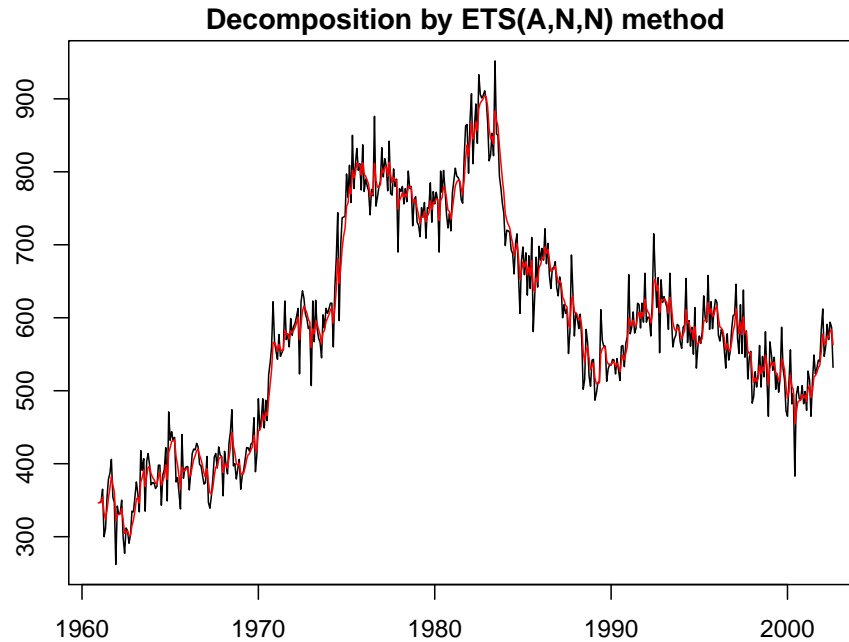
AIC	AICc	BIC
6704.782	6704.806	6713.211

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE
1.0886768	36.3603155	28.0809481	-0.1103519	5.0476089	0.8542579

Na základě AIC kritéria funkce `ets()` vybrala pro naše data jednoduché exponenciální vyrovnávání (značení  $ETS(A, N, N)$ ). Výsledky jednoduchého exponenciálního vyhlazení opět vykreslíme.

```
> par(mar = c(2, 2, 1, 0) + 0.5)
> plot(modelETS, plot.type = "single", col = 1:3, ylab = "")
```



Obrázek 5: Holtovo–Wintersonovo exponenciální vyrovnávání pomocí funkce `ets()` pro časovou řadu: *Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002*

```
> predETS <- forecast(modelETS)
> summary(predETS)
```

Forecast method: ETS(A,N,N)

Model Information:  
ETS(A,N,N)

Call:  
`ets(y = x)`

Smoothing parameters:  
alpha = 0.3982

Initial states:  
l = 346.2374

sigma: 36.3603

AIC	AICc	BIC
6704.782	6704.806	6713.211

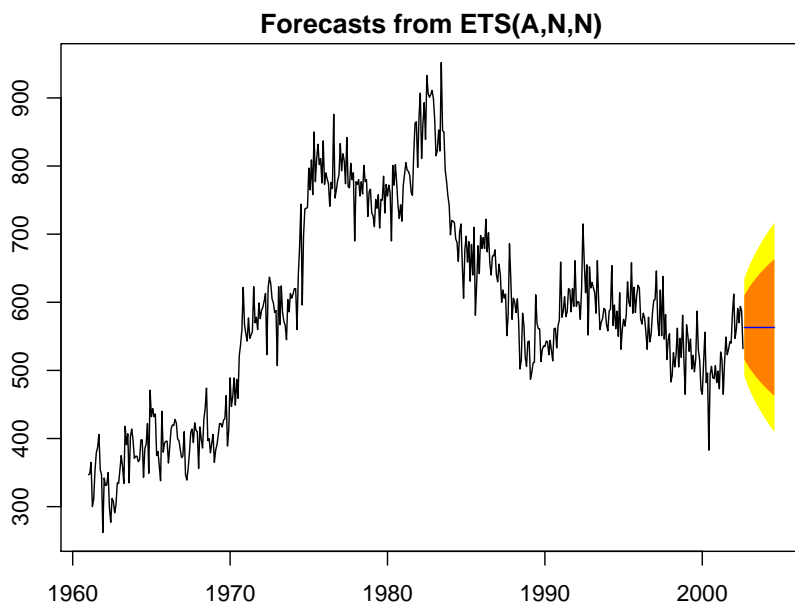
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE
1.0886768	36.3603155	28.0809481	-0.1103519	5.0476089	0.8542579

Forecasts:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Sep 2002	562.9935	516.3959	609.5911	491.7286	634.2584
Oct 2002	562.9935	512.8374	613.1496	486.2864	639.7006
Nov 2002	562.9935	509.5152	616.4718	481.2055	644.7815
Dec 2002	562.9935	506.3876	619.5994	476.4223	649.5647
Jan 2003	562.9935	503.4240	622.5630	471.8898	654.0972
Feb 2003	562.9935	500.6010	625.3860	467.5725	658.4145
Mar 2003	562.9935	497.9004	628.0866	463.4421	662.5449
Apr 2003	562.9935	495.3074	630.6796	459.4765	666.5105
May 2003	562.9935	492.8101	633.1769	455.6573	670.3297
Jun 2003	562.9935	490.3987	635.5883	451.9693	674.0177
Jul 2003	562.9935	488.0649	637.9221	448.4000	677.5870
Aug 2003	562.9935	485.8016	640.1854	444.9386	681.0484
Sep 2003	562.9935	483.6027	642.3843	441.5758	684.4112
Oct 2003	562.9935	481.4632	644.5238	438.3037	687.6833
Nov 2003	562.9935	479.3784	646.6086	435.1152	690.8718
Dec 2003	562.9935	477.3443	648.6427	432.0044	693.9826
Jan 2004	562.9935	475.3574	650.6296	428.9657	697.0213
Feb 2004	562.9935	473.4146	652.5724	425.9944	699.9926
Mar 2004	562.9935	471.5131	654.4739	423.0862	702.9008
Apr 2004	562.9935	469.6502	656.3368	420.2373	705.7497
May 2004	562.9935	467.8238	658.1632	417.4441	708.5429
Jun 2004	562.9935	466.0319	659.9551	414.7035	711.2835
Jul 2004	562.9935	464.2724	661.7146	412.0126	713.9744
Aug 2004	562.9935	462.5438	663.4433	409.3689	716.6181

```
> par(mar = c(2, 2, 1, 0) + 0.5)
> plot(predETS)
```



Obrázek 6: Predikce pomocí funkce `forecast()` pro časovou řadu: Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002

**C. Úkol:**

Na časové řady, které jste si našli v úvodním praktiku, aplikujte klouzavé průměry a exponenciální vyhlazování.