



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁRSKA PRÁCA

Július Rábek

Sezónne exponenciálne vyrovnávanie

Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: prof. RNDr. Tomáš Cipra, DrSc.

Študijný program: Matematika

Študijný obor: Finanční matematika

Praha 2018

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Sb., autorského zákona v platnom znení, najmä skutočnosť, že Univerzita Karlova má právo na uzatvorenie licenčnej zmluvy o použití tejto práce ako školského diela podľa §60 odst. 1 autorského zákona.

V dňa

Podpis autora

Na tomto mieste by som sa chcel poďakovať prof. RNDr. Tomášovi Ciprovi, DrSc. za jeho ochotu, venovaný čas a všetky cenné rady a pripomienky pri vzniku tejto bakalárskej práce. Veľká vďaka patrí aj mojim rodičom a sestre za ich neustálu podporu.

Názov práce: Sezónne exponenciálne vyrovňovanie

Autor: Július Rábek

Katedra: Katedra pravdepodobnosti a matematické statistiky

Vedúci bakalárskej práce: prof. RNDr. Tomáš Cipra, DrSc., Katedra pravdepodobnosti a matematické statistiky

Abstrakt: Táto práca sa zaoberá problematikou modelovania časových radov s výskytom sezónnej zložky. Na začiatku je popísaný princíp základných metód exponenciálneho vyrovňovania: jednoduché a dvojité exponenciálne vyrovňovanie, Holtova metóda, ktoré sú aplikovateľné na rady bez sezónnosti. Pre sezónne časové rady je najviac vhodná Holt-Wintersova metóda. Uvedená je v oboch jej variantoch a jej použitie závisí od charakteru sezónnej zložky. V práci je ďalej prezentované sezónne stavové modelovanie ako štatistický rámec pre metódy exponenciálneho vyrovňovania, pričom diskutované sú aj niektoré problémy spojené s praktickou realizáciou týchto techník spolu s návodom na ich riešenie. Na záver je uvedená aplikácia Holt-Wintersovej metódy na dvoch reálnych časových radoch vykazujúcich sezónnosť.

Kľúčové slová: časové rady, exponenciálne vyrovňovanie, sezónnosť

Title: Seasonal exponential smoothing

Author: Július Rábek

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Tomáš Cipra, DrSc., Department of Probability and Mathematical Statistics

Abstract: This thesis deals with the issues of time series modeling, where seasonal component is present. Principles of basic seasonal exponential smoothing methods: simple and double exponential smoothing, Holt's method, which are applicable on time series without seasonality, are described in the beginning. For seasonal time series, Holt-Winters exponential smoothing is the most suitable method. This method is introduced in both of its versions and the usage of either version depends on the characteristics of the seasonal component. Furthermore, state space modeling is presented as a statistical framework for exponential smoothing methods, joined with a discussion of some selected problems related with practical implementation of these techniques together with suggestions of their solution. Finally, Holt-Winters method on two real data time series with seasonality is presented.

Keywords: time series, exponential smoothing, seasonality

Obsah

Úvod	2
1 Exponenciálne vyrovňovanie časových radov	3
1.1 Dekompozícia časových radov	3
1.2 Princíp exponenciálneho vyrovňovania	4
1.3 Jednoduché exponenciálne vyrovňovanie	5
1.4 Dvojité exponenciálne vyrovňovanie	6
1.5 Holtova metóda	8
1.6 Holt-Wintersova metóda	8
1.6.1 Aditívna Holt-Wintersova metóda	9
1.6.2 Multiplikatívna Holt-Wintersova metóda	9
2 Stavové modely exponenciálneho vyrovňovania	10
2.1 Princíp stavových modelov exponenciálneho vyrovňovania	10
2.2 Sezónne stavové modely	11
2.2.1 Modely s aditívnou Sz_t	12
2.2.2 Modely s multiplikatívnou Sz_t	13
3 Praktická realizácia	14
3.1 Problémy spojené s praktickou realizáciou	14
3.1.1 Výber voliteľných parametrov	14
4 Aplikácia na reálnych dátach	16
4.1 Softwareová realizácia	16
Záver	25
Zoznam použitej literatúry	26
Zoznam obrázkov	27
Zoznam tabuliek	28

Úvod

V praxi sa neraz stretávame so situáciami, kedy sa na základe daných skutočností a historických udalostí snažíme konštruovať určité predpovede do budúcnosti. Či už ide o údaje demografické, finančné alebo ekonomické, tieto získané dáta často vykazujú konkrétne zákonitosti, na základe ktorých ich môžeme správne analyzovať a následne predpovedať ich budúci vývoj. Jedným zo základných vzorov, ktorý sa v sledovaných časových radoch často vyskytuje, je sezónnosť.

Pod týmto pojmom rozumieme tendenciu periodickej zmeny v časovom rade, ktorá sa pravidelne opakuje v istom časovom období známej dĺžky (napr. mesačné výkyvy počas kalendárneho roka, denné zmeny behom týždňa apod.). Tieto zmeny sú spôsobené najmä striedaním ročných období a s nimi súvisiacimi ľudskými zvykmi, ktoré sa následne častokrát významne odrazia v ekonomickej aktivite. Pri konštrukcii predpovedí je dôležité predpokladať, že sa sezónnosť v časovom rade vyskytuje naozaj pravidelne, aby bolo jeho modelovanie pomocou opakujúcich sa sezónnych faktorov oprávnené. Ako ilustratívny príklad si môžeme uviesť časový rad množstva mesačne predaného elektronického tovaru počas roka, ktorý pravdepodobne dosiahne svoj vrchol v novembri, prípadne v decembri. Táto charakteristika má vysoký predpoklad, že sa bude rok čo rok opakovať, hoci aj s menšími zmenami v dosiahnutom maxime.

V tejto práci sa budeme bližšie zaoberať práve zmienenou sezónnosťou a jej elimináciou z časových radov, aby sme v nich mohli lepšie popísať dlhodobé zákonitosti a konštruovať presnejšie predpovede do budúcnosti. Konkrétne sa zameriame na metódy exponenciálneho vyrovnávania, ktorých základy položili Holt a Brown-Meyer v druhej polovici minulého storočia.

V prvej časti práce zavedieme základné pojmy a bližšie popíšeme dekompozíciu časových radov. Ďalej sa v nej budeme zaoberať exponenciálnym vyrovnávaním, vysvetlíme jeho princíp a základné metódy. Vhodné stavové modely, ktoré sa pri ňom využívajú sú zhrnuté v kapitole druhej. Problémy spojené s praktickou realizáciou exponenciálneho vyrovnávania, akými sú voľby vyrovnávacích konštánt či počiatočných hodnôt, diskutujeme v tretej kapitole. Hlavným zdrojom pre tieto časti práce sú knihy autorov Cipra [2013] a Hyndman et al. [2008]. Posledná, štvrtá kapitola sa následne sústreďuje na praktickú aplikáciu vhodnej metódy exponenciálneho vyrovnávania na reálnych dátach vykazujúcich sezónnosť spolu s jeho softwareovou realizáciou.

1. Exponenciálne vyrovňovanie časových radov

1.1 Dekompozícia časových radov

Na úvod zavedieme niekoľko základných pojmov, ktoré budeme neskôr v tejto práci využívať. Budeme sa držať textu Cipra [2013], odkiaľ sme čerpali nasledujúce poznatky.

Definícia 1. *Nech sú hodnoty určitej veličiny pozorované v určitom časovom intervale s určitou frekvenciou záznamu (tj. veľkosťou intervalu medzi jednotlivými pozorovaniami). Potom hovoríme o dátach v tvare časového radu.*

Definícia 2. *Postupnosť $\{\varepsilon_t\}$ nazveme bielym šumom, ak sú náhodné veličiny ε_t navzájom nekorelované, majú nulovú strednú hodnotu $E(\varepsilon_t) = 0$ a kladný konečný konštantný rozptyl $\text{var}(\varepsilon_t) = \sigma^2 > 0$. Pridaním predpokladu normality náhodných veličín ε_t dostávame normálny biely šum.*

Definícia 3. *Nech y_t je skutočná hodnota časového radu a \hat{y}_t jeho vyrovnaná hodnota či predpoveď. Potom*

$$MSE = \sum_{t=1}^n \frac{1}{n} (y_t - \hat{y}_t)^2 \quad (1.1)$$

definujeme ako strednú štvorcovú chybu (mean squared error),

$$MAE = \sum_{t=1}^n \frac{1}{n} |y_t - \hat{y}_t|^2 \quad (1.2)$$

ako strednú absolútnu chybu (mean absolute error) a

$$SSE = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (1.3)$$

ako súčet štvorcových chýb (sum of squared errors).

Časové rady, predovšetkým ekonomického charakteru, môžeme rozložiť do niekoľkých špecifických zložiek:

- trend (dlhodobý vývoj časového radu) Tr_t ;
- sezónna zložka (každoročne sa opakujúce sezónne zmeny) Sz_t ;
- reziduálna (náhodná) zložka E_t (obvykle biely šum podľa definície 2).

Takáto dekompozícia časových radov patrí medzi základné metódy štatistickej analýzy a prináša so sebou hneď niekoľko výhod oproti pôvodnej, nerozloženej verzii časového radu. Jej hlavným účelom je analýza radu po jednotlivých zložkách a následné jednoduchšie skúmanie jeho pravidelného správania. Trendové a

sezónne zložky sú vnímané ako deterministické funkcie času, zatiaľ čo reziduálna zložka je náhodnou funkciou času (tj. náhodný proces). Je zrejmé, že časový rad nemusí vo svojej dekompozícii obsahovať všetky zmienené zložky a niektoré z nich môžu úplne chýbať.

Rozlišujeme dva základné prístupy k dekompozícii časových radov: *aditívny* a *multiplikatívny*. Aditívna dekompozícia má tvar

$$y_t = Tr_t + Sz_t + E_t$$

a meria jednotlivé zložky v rovnakých jednotkách, v akých sa uvádza príslušný časový rad y_t . Vyjadruje kvantitatívny nárast časového radu v príslušnej sezóne. Naopak multiplikatívna dekompozícia, keď je časový rad v tvare

$$y_t = Tr_t \cdot Sz_t \cdot E_t,$$

vyjadruje percentuálny nárast časového radu v príslušnej sezóne. V tomto prípade je jedine trendová zložka uvažovaná v jednotkách radu y_t a zvyšné zložky sú vyjadrené v relatívnych hodnotách vzhľadom k trendu, tým pádom sú bezrozmernými veličinami. Pre multiplikatívnu dekompozíciu je charakteristická priama úmernosť medzi sezónnymi výkyvmi a monotónnosťou trendu, čiže sezónne výkyvy sa zväčšujú (resp. znižujú) s rastúcim (resp. klesajúcim) trendom.

1.2 Princíp exponenciálneho vyrovnávania

Exponenciálne vyrovnávanie sa radí k jedným z adaptívnych prístupov k trendovej zložke časových radov, ktorý sa v praxi veľmi často využíva. Analyzovaný časový rad je očistený od sezónnej zložky, teda má tvar

$$y_t = Tr_t + E_t. \tag{1.4}$$

Rad je tvorený aditívnou reziduálnou zložkou (E_t) s vlastnosťami bieleho šumu a trendom (Tr_t), ktorý sa snažíme odhadnúť a zároveň predpovedať jeho budúci vývoj. V prípade exponenciálneho vyrovnávania vyrovnaný rad \hat{y}_t minimalizuje výraz typu

$$(y_t - \hat{y}_t)^2 + (y_{t-1} - \hat{y}_{t-1})^2\beta + (y_{t-2} - \hat{y}_{t-2})^2\beta^2 + \dots, \tag{1.5}$$

kde $0 < \beta < 1$ je dopredu zvolená *diskontná konštanta*.

V súčte (1.5) dochádza k diskontovaniu váh. To znamená, že exponenciálne vyrovnávanie priraduje exponenciálne nižšie váhy starším pozorovaniam. Inými slovami, nedávne pozorovania dostávajú relatívne vyššiu váhu v predpovedi ako dáta, ktoré sú viac vzdialené do minulosti. Práve od exponenciálneho znižovania váh je odvodený aj názov *exponenciálne vyrovnávanie*.

V nasledujúcej časti práce sa budeme venovať základným metódam exponenciálneho vyrovnávania, ktoré sa používajú v závislosti od toho, na aký časový rad bude vyrovnávanie aplikované. Aj naďalej predpokladajme, že pracujeme s časovým radom v tvare (1.4) očisteným od sezónnej zložky Sz_t .

1.3 Jednoduché exponenciálne vyrovnávanie

Jednoduché exponenciálne vyrovnávanie sa používa pre rady, ktorých trendová zložka je v krátkych úsekoch rady rovnaká, teda ju môžeme považovať za lokálne konštantnú

$$Tr_t = \beta_0. \quad (1.6)$$

Cieľom bude nájsť odhad parametru β_0 . Keďže exponenciálne vyrovnávanie sa radí medzi metódy s adaptívnym prístupom k vyrovnávanému radu, odhad tohto parametru bude závislý od časového okamihu, v ktorom sa uskutočnil. Ak si odhad parametru β_0 v čase t označíme ako $b_0(t)$, bude tento odhad reprezentovať nielen vyrovnanú hodnotu \hat{y}_t uvažovaného radu, ale zároveň aj odhadnutú hodnotu trendu v čase t . Jeho veľkosť bude závisieť od všetkých dostupných pozorovaných hodnôt y_1, \dots, y_n . Odhad $b_0(t)$ vzhľadom k (1.5) potom získame minimalizáciou cez β_0 výrazu

$$\sum_{i=0}^{\infty} (y_{t-i} - \beta_0)^2 \beta^i, \quad (1.7)$$

kde $0 < \beta < 1$ je dopredu zvolená diskontná konštanta.

Napriek tomu, že v praktických situáciách poznáme vždy len konečný počet y_1, \dots, y_n , vzhľadom k asymptotickým zjednodušeniam sa uvedený výraz (1.7) používa v tvare nekonečného súčtu. Ak (1.7) zderivujeme podľa β_0 a položíme túto deriváciu rovno nule, dostaneme vzťah pre odhad $b_0(t)$ ako

$$\hat{y}_t = (1 - \beta) \sum_{i=0}^{\infty} \beta^i y_{t-i}. \quad (1.8)$$

Je zrejmé, že výraz (1.8) môžeme interpretovať ako vážený priemer tohto radu do času t s exponenciálne klesajúcimi váhami

$$1 - \beta, (1 - \beta)\beta, (1 - \beta)\beta^2, \dots$$

Tento spôsob výpočtu sa však v praxi nepoužíva, ale (1.8) sa prevádza na lepšie aplikovateľný rekurentný tvar

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}, \quad (1.9)$$

kde $\alpha = 1 - \beta$ ($0 < \alpha < 1$) sa nazýva *vyrovnávací konštanta*.

Na rekurentnom vzorci (1.9) môžeme jasne vidieť výhody exponenciálneho vyrovnávania:

- jednoduchosť výpočtov,
- úspornosť vzhľadom k objemu skladovaných dát: stačí uchovať len predchádzajúcu vyrovnávanú hodnotu namiesto všetkých minulých hodnôt,
- možnosť určovať adaptívnosť metódy: voľbou vyrovnávacej konštanty α môžeme meniť vyrovnávaciu schopnosť metódy.

V prípade, že chceme metódu jednoduchého exponenciálneho vyrovnávania použiť na konštrukciu predpovedí, kladieme

$$\hat{y}_{t+\tau}(t) = \hat{y}_t$$

vzhľadom k (1.6) pre ľubovoľné $\tau > 0$.

Existuje ešte tretí tvar vzorca pre vyrovnávanú hodnotu jednoduchého exponenciálneho vyrovnávania, ktorý je založený na vhodne nastavenej chybe predpovede z predchádzajúceho obdobia. Tento tvar sa niekedy nazýva aj „chybový“:

$$\hat{y}_t = \hat{y}_{t-1} + \alpha(y_t - \hat{y}_{t-1}) = \hat{y}_{t-1} + \alpha(y_t - \hat{y}_t(t-1)) = \hat{y}_{t-1} + \alpha \cdot e_t.$$

Pozn.1.2.1 V rámci exponenciálneho vyrovnávania môžeme okrem bodových predpovedí konštruovať aj tzv. *predpovedné intervaly*, a to za predpokladu normality reziduálneho komponentu E_t uvažovaného časového radu. V prípade, že tento predpoklad vynecháme, budú predpovedné intervaly vytvárané aproximatívne. $(1-p)$.100-percentný predpovedný interval sa doporučuje zostrojovať v tvare

$$\left(\hat{y}_{t+\tau}(n) - u_{1-p/2} \cdot d_\tau \cdot \text{MAE}, \hat{y}_{t+\tau}(n) + u_{1-p/2} \cdot d_\tau \cdot \text{MAE} \right),$$

kde pre ľubovoľné $\tau > 0$ je

- $u_{1-p/2}$ ($1-p/2$)-kvantil normovaného normálneho rozdelenia,
- d_τ konštanta slúžiaca k prevodu MSE (1.1) na MAE (1.2),
- MAE stredná absolútna chyba (1.2).

1.4 Dvojité exponenciálne vyrovnávanie

U radov, ktorých trend môžeme považovať za lokálne lineárny, tj. je tvaru

$$Tr_{t-i} = \beta_0 + \beta_1 \cdot (-i), \quad (1.10)$$

sa používa metóda dvojitého exponenciálneho vyrovnávania, niekedy označovaná aj ako *Brownova metóda*. Pri vytváraní odhadov parametrov β_0 a β_1 v čase t , ktoré analogicky označíme ako $b_0(t)$ a $b_1(t)$, sa aplikuje podobný postup ako pri odhadovaní trendu v jednoduchom exponenciálnom vyrovnávaní. Budeme minimalizovať výraz

$$\sum_{i=0}^{\infty} [y_{t-i} - (\beta_0 + \beta_1(-i))]^2 \beta^i, \quad (1.11)$$

kde $0 < \beta < 1$ je opäť dopredu zvolená diskontná konštanta. V prípade, že sumu (1.11) parciálne zderivujeme podľa parametrov β_0 a β_1 a tieto derivácie položíme rovné nule, dostávame sústavu normálnych rovníc

$$\begin{aligned} \sum_{i=0}^{\infty} \beta^i y_{t-i} - \beta_0 \sum_{i=0}^{\infty} \beta^i + \beta_1 \sum_{i=0}^{\infty} i \beta^i &= 0, \\ \sum_{i=0}^{\infty} i \beta^i y_{t-i} - \beta_0 \sum_{i=0}^{\infty} i \beta^i + \beta_1 \sum_{i=0}^{\infty} i^2 \beta^i &= 0. \end{aligned}$$

Použitím nasledujúcich vzorcov na súčet nekonečných radov

$$\begin{aligned}\sum_{i=0}^{\infty} \beta^i &= \frac{1}{1-\beta} \\ \sum_{i=0}^{\infty} i\beta^i &= \frac{\beta}{(1-\beta)^2} \\ \sum_{i=0}^{\infty} i^2\beta^i &= \frac{\beta(1+\beta)}{(1-\beta)^3}\end{aligned}$$

vieme túto sústavu normálnych rovníc zjednodušiť na tvar

$$\begin{aligned}\beta_0 - \frac{\beta}{1-\beta} \cdot \beta_1 &= (1-\beta) \sum_{i=0}^{\infty} \beta^i y_{t-i}, \\ \beta \cdot \beta_0 - \frac{\beta(1+\beta)}{1-\beta} \cdot \beta_1 &= (1-\beta)^2 \sum_{i=0}^{\infty} i\beta^i y_{t-i}.\end{aligned}\tag{1.12}$$

Kvôli zjednodušeniu značenia sa zavádzajú tzv. *vyrovnávacie štatistiky*. Vzhľadom k (1.8) sa zavádza *jednoduchá vyrovnávacia štatistika*

$$S_t = (1-\beta) \sum_{i=0}^{\infty} \beta^i y_{t-i}\tag{1.13}$$

a obdobne sa vytvára *dvojitá vyrovnávacia štatistika*

$$S_t^{[2]} = (1-\beta) \sum_{i=0}^{\infty} \beta^i S_{t-i}.\tag{1.14}$$

Pomocou práve zavedených vyrovnávacích štatistík vieme sústavu (1.12) prepísať do tvaru

$$\begin{aligned}\beta_0 - \frac{\beta}{1-\beta} \cdot \beta_1 &= S_t, \\ \beta \cdot \beta_0 - \frac{\beta(1+\beta)}{1-\beta} \cdot \beta_1 &= S_t^{[2]} - (1-\beta)S_t,\end{aligned}$$

z ktorého dostávame hľadané odhady $b_0(t)$ a $b_1(t)$ ako

$$\begin{aligned}b_0(t) &= 2S_t - S_t^{[2]}, \\ b_1(t) &= \frac{1-\beta}{\beta}(S_t - S_t^{[2]}).\end{aligned}\tag{1.15}$$

Následná predpoveď hodnoty $y_{t+\tau}$ uskutočnená v čase t má tvar

$$\hat{y}_{t+\tau}(t) = b_0(t) + b_1(t) \cdot \tau = \left(2 + \frac{\alpha \cdot \tau}{1-\alpha}\right) S_t - \left(1 + \frac{\alpha \cdot \tau}{1-\alpha}\right) S_t^{[2]}.\tag{1.16}$$

Špeciálne pre $\tau = 0$ sa vyrovnaná hodnota uvažovaného časového radu určí ako

$$\hat{y}_t = 2S_t - S_t^{[2]}$$

Výrazy (1.13) a (1.14) vieme v analógii so vzťahom (1.9) vyjadriť aj rekurentne, kde α je opäť rovná $1 - \beta$

$$\begin{aligned} S_t &= \alpha y_t + (1 - \alpha)S_{t-1}, \\ S_t^{[2]} &= \alpha S_t + (1 - \alpha)S_{t-1}^{[2]}. \end{aligned} \quad (1.17)$$

Kvôli jednoduchosti sa pri výpočte dvojice vyrovnávacích štatistík S_t a $S_t^{[2]}$, rovnako ako pri jednoduchom exponenciálnom vyrovnávaní, používajú práve rekurentné vzťahy (1.17).

Pozn.1.3.1 Všeobecne môžeme uvažovať exponenciálne vyrovnávanie až do rádu r , avšak z praktického hľadiska je tretí rád najvyšším rádom, ktorý má rozumné použitie. Táto metóda sa nazýva *trojité exponenciálne vyrovnávanie* a vznikla ako prirodzené rozšírenie jednoduchého a dvojitého exponenciálneho vyrovnávania. Využitie má u radov s kvadratickým či konvexným trendom.

1.5 Holtova metóda

Tretou zo základných metód exponenciálneho vyrovnávania je tzv. *Holtova metóda*, ktorá je zovšeobecnením dvojitého exponenciálneho vyrovnávania. Využíva sa u dát, ktoré vykazujú lineárny trend a nevyskytuje sa v nich sezónnosť. Holtova metóda využíva dve vyrovnávacie konštanty: α a β . Zatiaľ čo α ($0 < \alpha$) je určená na vyrovnávanie úrovne L_t , β ($0 < \beta < 1$) slúži na vyrovnávanie smernice T_t . Holtova metóda je daná nasledujúcim rekurentným vzťahom:

$$\begin{aligned} \hat{y}_t &= L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1}), \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}. \end{aligned}$$

V prípade konštruovania predpovedí budúcich hodnôt sa používa tvar

$$\hat{y}_{t+\tau}(t) = L_t + T_t \cdot \tau,$$

pre ľubovoľné $\tau > 0$. Eliminácia úrovne L_t daného radu v čase t sa podľa uvedenej rekurentného vzťahu získa konvexnou kombináciou aktuálnej hodnoty y_t v čase t a odhadom tejto hodnoty z predchádzajúceho času $t - 1$.

Medzi Holtovou metódou a dvojitým exponenciálnym vyrovnávaním existuje jednoduchý prevod, keďže dvojité exponenciálne vyrovnávanie s vyrovnávacou konštantou α je vlastne špeciálnym prípadom Holtovej metódy s vyrovnávacími konštantami α_{Holt} a β_{Holt} v tvare

$$\begin{aligned} \alpha_{Holt} &= \alpha(2 - \alpha) \\ \beta_{Holt} &= \frac{\alpha}{2 - \alpha}. \end{aligned}$$

1.6 Holt-Wintersova metóda

Holtova metóda z predchádzajúceho odstavca sa pridaním tretej vyrovnávacej konštanty môže rozšíriť tak, aby bola pri vyrovnávaní okrem lokálneho trendu

zohľadnená aj sezónnosť v časovom rade. Tento spôsob eliminácie trendovej a sezónnej zložky časového radu sa nazýva *Holt-Wintersova metóda*.

Existuje v dvoch variantoch: aditívnej a multiplikatívnej, a obe využívajú až tri vyrovnávacie konštanty: α , β a γ . Konštanty α a β spĺňajú rovnaký účel ako v Holtovej metóde, t.j. α slúži na vyrovnávanie úrovne L_t , β na vyrovnávanie smernice T_t . Konštanta γ ($0 < \gamma < 1$) sa pridáva kvôli vyrovnávaniu sezónnej zložky S_t daného radu so sezónou dĺžky s , t.j. počtom pozorovaní v jednej sezóne.

1.6.1 Aditívna Holt-Wintersova metóda

Vzťahy pre úroveň, smernicu trendu a sezónnu zložku sú dané nasledujúcimi rekurentnými vzorcami:

$$\begin{aligned}L_t &= \alpha(y_t - Sz_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\Sz_t &= \gamma(y_t - L_t) + (1 - \gamma)Sz_{t-s}.\end{aligned}$$

Vyrovnávaná hodnota je potom rovná súčtu úrovne a sezónnej zložky v čase t :

$$\hat{y}_t = L_t + Sz_t. \quad (1.18)$$

U predpovedí pre budúce obdobia je potrebné rozlíšiť o akú predpovednú sezónu sa jedná. Predpovedné sezóny značne vzdialené od aktuálneho času t sa odvodzujú analogicky, avšak môžeme u nich očakávať, že predpovede už nebudú veľmi spoľahlivé.

$$\hat{y}_{t+\tau}(t) = \begin{cases} L_t + T_t \cdot \tau + Sz_{t+\tau-s}, & \text{pre } \tau = 1, \dots, s, \\ L_t + T_t \cdot \tau + Sz_{t+\tau-2s}, & \text{pre } \tau = s + 1, \dots, 2s. \end{cases}$$

1.6.2 Multiplikatívna Holt-Wintersova metóda

Rekurentná formula multiplikatívnej Holt-Wintersovej metódy je analogická jej aditívnej verzii, avšak s tým rozdielom, že príslušné súčty a rozdiely sa nahradia súčinnými a podielmi:

$$\begin{aligned}L_t &= \alpha(y_t / Sz_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\Sz_t &= \gamma(y_t / L_t) + (1 - \gamma)Sz_{t-s},\end{aligned}$$

Podobne sa pristupuje k predpovediam:

$$\begin{aligned}\hat{y}_t &= L_t \cdot Sz_t, \\ \hat{y}_{t+\tau}(t) &= \begin{cases} (L_t + T_t \cdot \tau) \cdot Sz_{t+\tau-s}, & \text{pre } \tau = 1, \dots, s, \\ (L_t + T_t \cdot \tau) \cdot Sz_{t+\tau-2s}, & \text{pre } \tau = s + 1, \dots, 2s. \end{cases}\end{aligned}$$

2. Stavové modely exponenciálneho vyrovnávania

V tejto kapitole sa zameriame na stavové modely, ktoré sa viažu s jednotlivými metódami exponenciálneho vyrovnávania. Bližšie vysvetlíme zásady stavových modelov exponenciálneho vyrovnávania a následne popíšeme stavové modely, ktoré berú do úvahy sezónnosť. Budeme sa opierať o poznatky z textu Hyndman et al. [2008], z ktorého sme čerpali.

2.1 Princíp stavových modelov exponenciálneho vyrovnávania

Koncepcia stavových modelov je veľmi výhodná pri skúmaní exponenciálneho vyrovnávania, keďže komponenty časového radu sú už explicitne zmienené v používanom modeli. Pre každú metódu existujú dva modely: model s aditívnou reziduálnou zložkou a model s multiplikatívnou reziduálnou zložkou. Tu však musíme byť opatrný a uvedomiť si dôležitý rozdiel medzi *metódou* a *modelom*. Napriek tomu, že stochastické stavové modely generujú rovnaké bodové predpovede ako metódy exponenciálneho vyrovnávania, modely navyše poskytujú aj návod na výpočet intervalových predpovedí s danou hladinou spoľahlivosti a umožňujú použitie mnohých ďalších štatistických nástrojov.

V úvode sme sa zmienili o dekompozícii časových radov do troch častí, viď odstavec 1.1. Na základe tohto rozkladu môžeme zaviesť značenie, ktoré bude charakterizovať individuálne modely a taktiež napomôže pri zapamätaní si poradia, v ktorom sú komponenty definované. Trojica **ETS(-,-,-)** označuje samotné zložky časového radu v nasledujúcom poradí: reziduálna zložka, trendová a sezónna zložka (*error, trend, seasonality*). Podľa tohto značenia vieme o časovom rade povedať, z akých komponentov sa skladá a v akom sú tvare. Napríklad model multiplikatívnej Holt-Wintersovej metódy by sme zapísali v tvare ETS(A,A,M).

V momente, kedy je stavový model špecifikovaný, môžeme skúmať pravdepodobnostné rozdelenie budúcich hodnôt časového radu a určiť niektoré z jeho vlastností. Pre nás najdôležitejšou bude podmienená stredná hodnota budúcich pozorovaní za predpokladu znalosti dát z predošlých období:

$$\hat{y}_{t+\tau} = E(y_{t+\tau} | \mathbf{x}_t), \quad (2.1)$$

kde \mathbf{x}_t je stavový vektor, ktorý obsahuje informácie o úrovni, trende a sezónnosti časového radu v čase t .

Zavedenie stavových modelov metód exponenciálneho vyrovnávania

Obecný stavový model je definovaný stavovým vektorom

$$\mathbf{x}_t = (L_t, T_t, Sz_t, Sz_{t-1}, \dots, Sz_{t-s+1})^T \quad (2.2)$$

a dvojicou stavových rovníc v tvare

$$y_t = w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})E_t, \quad (2.3)$$

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})E_t, \quad (2.4)$$

kde náhodná zložka E_t má vlastnosti bieleho šumu (definícia 2) s rozptylom σ^2 a $w(\mathbf{x}_{t-1}) = \hat{y}_t$.

Pridaním ďalších predpokladov môžeme tento univerzálny model konkretizovať a prepísať do tvaru, s ktorým sa v praxi stretávame častejšie. V modeli s aditívnou reziduálnou zložkou nadobúda $r(\mathbf{x}_{t-1})$ hodnotu 1, odkiaľ dostávame predpis časového radu y_t v tvare $y_t = \hat{y}_t + E_t$. Na druhej strane model s multiplikatívnou reziduálnou zložkou má $r(\mathbf{x}_{t-1}) = \hat{y}_t$ a po zjednodušení sa $y_t = \hat{y}_t(1 + E_t)$.

V oboch prípadoch vieme nájsť určité výhody a nevýhody, podľa ktorých sa pri konkrétnych dátach rozhodujeme, aký model na ne aplikovať. Napríklad modely s multiplikatívnou reziduálnou zložkou sú vhodné pre striktne kladné dáta, pretože ak dáta obsahujú nuly alebo záporné hodnoty, výsledky spravidla nie sú stabilné. V tomto prípade sa používajú aditívne modely.

Už na začiatku tejto sekcie sme načrtli významnú vlastnosť stavových modelov, ktorou je schopnosť určovať predpovedné intervaly. V prípade lineárnych a aj mnohých nelineárnych modelov, v ktorých sa predpovede riadia normálnym rozdelením, ich môžeme určiť deriváciou podmieneného rozptylu $var(y_{t+\tau}|\mathbf{x}_t)$. Existuje však aj priamočiarejší postup, ktorý síce môže byť časovo náročnejší, ale je použiteľný vo všetkých modeloch. Tým je simulácia veľkého počtu scenárov budúceho vývoja časového radu na základe odhadnutého modelu a následné odhadnutie rozdelenia nasimulovaných dát.

2.2 Sezónne stavové modely

V predchádzajúcej časti práce sme ukázali, ako vyzerá a ako je definovaný obecný stavový model. Ako však napovedá názov tejto práce, zameriavame sa hlavne na modely pre časové rady vykazujúce sezónnosť.

Takýchto modelov je celkovo dvanásť a môžeme ich rozdeliť do dvoch kategórií:

1. modely s aditívnou sezónnou zložkou Sz_t (**celkovo 6 modelov**)
2. modely s multiplikatívnou sezónnou zložkou Sz_t (**celkovo 6 modelov**)

Podľa tejto kategorizácie teraz zostrojíme prehľad všetkých sezónnych stavových modelov, ktoré zároveň popíšeme príslušnými stavovými rovnicami. Pri ich zavedení budeme používať nasledovné značenie:

L_t predstavuje úroveň časového radu v čase t , T_t smernicu v čase t , Sz_t sezónnu zložku v čase t a s počet sezón v jednom roku. Pre všetky varianty zavedieme tri vyrovnávacie konštanty: α pre L_t , β pre T_t a γ pre Sz_t , patriace do intervalu $(0,1)$.

2.2.1 Modely s aditívnou Sz_t

U dát, pri ktorých môžeme hovoriť o nejakom signifikantnom absolútnom náraste hodnôt v určitom období počas roka sa využívajú modely s aditívnym prístupom k sezónnemu komponentu časového radu. Sú vhodné pre dáta s nemeniacou sa sezónnosťou, od ktorej môžeme časový rad očistiť jej odčítaním. Príkladom odpovedajúceho časového radu je spotreba energií v domácnostiach počas roka, kedy náklady v zimných mesiacoch pravidelne narastajú o istú čiastku prevyšujúcu celoročný priemer. Pri predpovedaní by sme tým pádom zimným mesiacom pridávali dodatočné hodnoty, práve pre zohľadnenie sezónnych výkyvov.

ETS(-,-,A) modely		
	ETS(A,N,A)	ETS(M,N,A)
Úroveň L_t	$L_{t-1} + \alpha E_t$	$L_{t-1} + \alpha(L_{t-1} + Sz_{t-s})E_t$
Smernica T_t	—	—
Sezónna zložka Sz_t	$Sz_{t-s} + \gamma E_t$	$Sz_{t-s} + \gamma(L_{t-1} + Sz_{t-s})E_t$
Vyrovnaná hodnota \hat{y}_t	$L_{t-1} + Sz_{t-s}$	$L_{t-1} + Sz_{t-s}$
	ETS(A,M,A)	ETS(M,M,A)
Úroveň L_t	$L_{t-1}T_{t-1} + \alpha E_t$	$L_{t-1}T_{t-1} + \alpha(L_{t-1}T_{t-1} + Sz_{t-s})E_t$
Smernica T_t	$T_{t-1} + \beta E_t / L_{t-1}$	$T_{t-1} + \beta(L_{t-1}T_{t-1} + Sz_{t-s})E_t / L_{t-1}$
Sezónna zložka Sz_t	$Sz_{t-s} + \gamma E_t$	$Sz_{t-s} + \gamma(L_{t-1}T_{t-1} + Sz_{t-s})E_t$
Vyrovnaná hodnota \hat{y}_t	$L_{t-1}T_{t-1} + Sz_{t-s}$	$L_{t-1}T_{t-1} + Sz_{t-s}$
	ETS(A,A,A)	ETS(M,A,A)
Úroveň L_t	$L_{t-1} + T_{t-1} + \alpha E_t$	$L_{t-1} + T_{t-1} + \alpha(L_{t-1} + T_{t-1} + Sz_{t-s})E_t$
Smernica T_t	$T_{t-1} + \beta E_t$	$T_{t-1} + \beta(L_{t-1} + T_{t-1} + Sz_{t-s})E_t$
Sezónna zložka Sz_t	$Sz_{t-s} + \gamma E_t$	$Sz_{t-s} + \gamma E_t$
Vyrovnaná hodnota \hat{y}_t	$L_{t-1} + T_{t-1} + Sz_{t-s}$	$L_{t-1} + T_{t-1} + Sz_{t-s}$

Tabuľka 2.1: Modely s aditívnou sezónnou zložkou

2.2.2 Modely s multiplikatívnou Sz_t

V prípade rozboru údajov, akými sú napríklad počty ľudí využívajúcich leteckú dopravu alebo tržby z maloobchodného predaja, hovoríme častejšie o percentuálnych zmenách, než o posunoch o nejakú absolútnu hodnotu. Jedná sa o multiplikatívny prístup k sezónnosti časového radu a je použiteľný aj vtedy, keď sezónne kolísanie závisí od dňa v týždni (napr. zaťaženie mestskej hromadnej dopravy) alebo od dennej doby (napr. spotreba elektriny). Takýto prístup je vhodný u časových radov, v ktorých sa sezónna zložka mení v závislosti od úrovne tohoto radu, a keďže tá je vyjadrená v relatívnej hodnote, rad sezónne očistíme jeho vydelením.

<i>ETS(-,-,M) modely</i>		
	ETS(A,N,M)	ETS(M,N,M)
Úroveň L_t	$L_{t-1} + \alpha E_t / Sz_{t-s}$	$L_{t-1}(1 + \alpha E_t)$
Smernica T_t	—	—
Sezónna zložka Sz_t	$Sz_{t-s} + \gamma E_t / L_{t-1}$	$Sz_{t-s}(1 + \gamma E_t)$
Vyrovnaná hodnota \hat{y}_t	$L_{t-1}Sz_{t-s}$	$L_{t-1}Sz_{t-s}$
	ETS(A,M,M)	ETS(M,M,M)
Úroveň L_t	$L_{t-1}T_{t-1} + \alpha E_t / Sz_{t-s}$	$L_{t-1}T_{t-1}(1 + \alpha E_t)$
Smernica T_t	$T_{t-1} + \beta E_t / (Sz_{t-s}L_{t-1})$	$T_{t-1}(1 + \beta E_t)$
Sezónna zložka Sz_t	$Sz_{t-s} + \gamma E_t / (L_{t-1}T_{t-1})$	$Sz_{t-s}(1 + \gamma E_t)$
Vyrovnaná hodnota \hat{y}_t	$L_{t-1}T_{t-1}Sz_{t-s}$	$L_{t-1}T_{t-1}Sz_{t-s}$
	ETS(A,A,M)	ETS(M,A,M)
Úroveň L_t	$L_{t-1} + T_{t-1} + \alpha E_t / Sz_{t-s}$	$(L_{t-1} + T_{t-1})(1 + \alpha E_t)$
Smernica T_t	$T_{t-1} + \beta E_t / Sz_{t-s}$	$T_{t-1} + \beta(L_{t-1} + T_{t-1})E_t$
Sezónna zložka Sz_t	$Sz_{t-s} + \gamma E_t / (L_{t-1} + T_{t-1})$	$Sz_{t-s}(1 + \gamma E_t)$
Vyrovnaná hodnota \hat{y}_t	$(L_{t-1} + T_{t-1})Sz_{t-s}$	$(L_{t-1} + T_{t-1})Sz_{t-s}$

Tabuľka 2.2: Modely s multiplikatívnou sezónnou zložkou

3. Praktická realizácia

3.1 Problémy spojené s praktickou realizáciou

V tejto časti práce sa sústreďujeme na hlavné problémy spojené s aplikáciou jednotlivých metód exponenciálneho vyrovnávania na rôzne časové rady. Je zrejmé, že jednotlivé metódy nemusia byť vhodné pre všetky časové rady a ich výber bude závisieť práve od typu týchto radov. Ich úspešnosť však nezávisí len od ich primeraného použitia, ale aj od individuálneho určenia voliteľných parametrov.

3.1.1 Výber voliteľných parametrov

Ako už bolo uvedené v prvej kapitole, v praxi sa exponenciálne vyrovnávanie najčastejšie používa v tvare rekurentnej formule. Úspešnosť jednotlivých metód exponenciálneho vyrovnávania je teda z veľkej časti podmienená práve voľbou správnych parametrov v ich rekurentnej realizácii: počiatocnej hodnoty \hat{y}_0 a vyrovnávacích konštánt. Špeciálne výber vyrovnávacích konštánt hrá veľmi dôležitú rolu, pretože ovplyvňujú úroveň vyrovnania dát.

Ak by sme zvolili $\alpha = 1$, nedosiahli by sme žiadne vyrovnanie dát a výsledkom by bol pôvodný časový rad. Naopak, v prípade voľby druhej hraničnej hodnoty $\alpha = 0$ by sme získali úplne vyrovnaný časový rad jednoznačne určený iba počiatočnými podmienkami. V nasledujúcich riadkoch sa preto na základe odporúčaní textu Cipra [2013] bližšie pozrieme na spôsoby adekvátneho určenia týchto parametrov, a to postupne pre každú metódu zvlášť.

a) Jednoduché exponenciálne vyrovnávanie

Za počiatočnú hodnotu \hat{y}_0 v (1.9) sa pri vyrovnávaní časových radov metódou jednoduchého exponenciálneho vyrovnávania najčastejšie volí aritmetický priemer prvých niekoľko hodnôt radu, prípadne priamo prvá pozorovaná hodnota y_1 (potom aj $\hat{y}_1 = y_1$).

Pri voľbe vyrovnávacej konštanty α sa zvyčajne obmedzujeme na interval $0 < \alpha \leq 0,3$ a môžeme použiť jeden zo zaužívaných spôsobov

- fixná voľba, najčastejšie $\alpha = 0,1$ alebo $\alpha = 0,2$;
- preskúmanie niekoľkých hodnôt α a z nich sa potom vyberie tá hodnota, ktorá v danom rade poskytuje najlepšie predpovede v zmysle minimálnej SSE (1.3).

b) Dvojité exponenciálne vyrovnávanie

Tentokrát je nutné hľadať viac počiatočných hodnôt než v predchádzajúcom prípade. Potrebujeme zvoliť počiatocné hodnoty jednoduchej a dvojitej vyrovnávacej štatistiky S_0 a $S_0^{[2]}$, ktoré sa obvykle určia zo vzťahov (1.15) pri $t = 0$. Získame ich preložením nejakého krátkeho počiatočného úseku časového radu

priamkou a použitím metódy najmenších štvorcov odhadneme parametre $b_0(0)$ a $b_1(0)$. Následne získame vzťahy pre S_0 a $S_0^{[2]}$ ako

$$S_0 = b_0(0) - \frac{1-\alpha}{\alpha}b_1(0), \quad S_0^{[2]} = b_0(0) - \frac{2(1-\alpha)}{\alpha}b_1(0). \quad (3.1)$$

Vyrovňavaciu konštantu α volíme analogicky s metódou jednoduchého exponenciálneho vyrovnávania.

c) Holtova metóda

Na inicializáciu Holtovej metódy nám stačí voliť $L_0 = y_1$ a $T_0 = y_2 - y_1$. Pre voľbu vyrovnávacej konštanty α platia rovnaké zásady ako v a) a b).

d) Holt-Wintersova metóda

Realizácia rekurentných formúl Holt-Wintersovej metódy vyžaduje voľbu počiatkových hodnôt $L_0, T_0, Sz_{-s+1}, Sz_{-s+2}, \dots, Sz_0$ a vyrovnávacích konštánt α, β a γ . Pre voľbu vyrovnávacích konštánt platia rovnaké zásady v aditívnej aj multiplikatívnej verzii tejto metódy. Môžeme použiť

- fixnú voľbu, zvyčajne sa doporučuje $\alpha = \gamma = 0,4$ a $\beta = 0,1$;
- výber zo siete hodnôt konštánt α, β a γ podľa predpovedí s najmenšou SSE (1.3), postupuje sa podobne ako v a) a b).

Pri určení počiatkových hodnôt sa však už postupy v jednotlivých variantoch líšia. V aditívnom prístupe odhadneme naše hodnoty vďaka odhadom b_0, b_1 a a_2, \dots, a_s regresného modelu s lineárnym trendom

$$y_t = \beta_0 + \beta_1 t + \alpha_2 x_{t2} + \dots + \alpha_s x_{ts} + E_t$$

ako

$$L_0 = b_0, \quad T_0 = b_1, \quad Sz_{-s+1} = 0, \quad Sz_{-s+2} = a_2, \quad \dots \quad Sz_0 = a_s.$$

V multiplikatívnej verzii regresný prístup k sezónnosti už použiť nemôžeme, preto príslušné počiatkové hodnoty určíme pomocou nasledujúcich vzorcov

$$L_0 = \bar{y}_1 - \frac{s+1}{2}T_0, \quad T_0 = \frac{\bar{y}_m - \bar{y}_1}{(m-1)s}, \quad Sz_{j-s} = \frac{1}{m} \sum_{i=0}^{m-1} \frac{y_{j+s,i}}{\bar{y}_{i+1} - \left(\frac{s+1}{2} - j\right)T_0},$$

kde $j = 1, \dots, s$, \bar{y}_i je aritmetický priemer pozorovaní cez i -tu sezónu dĺžky s a m je celkový počet týchto sezón.

4. Aplikácia na reálnych dátach

V záverečnej kapitole predvedieme praktické použitie definovaných metód exponenciálneho vyrovnávania na reálnych dátach. Pre každú metódu samozrejme existujú vhodné časové rady, pre ktoré je daná metóda optimálna, pričom zvyšné metódy by už nemuseli byť vo vyrovnávaní a predpovedaní také úspešné. Cieľom našej práce však nie je ilustrovať každú z metód individuálne za ideálnych podmienok, ale pozorovať výsledky exponenciálneho vyrovnávania na dátach vykazujúcich sezónnosť. Použijeme preto dva sezónne časové rady z rôznych prostredí a s rôznym počtom sezón s za rok, na ktoré aplikujeme vždy jednu z variant Holt-Wintersovej metódy podľa toho, ktorá bude pre dané dáta vhodnejšia.

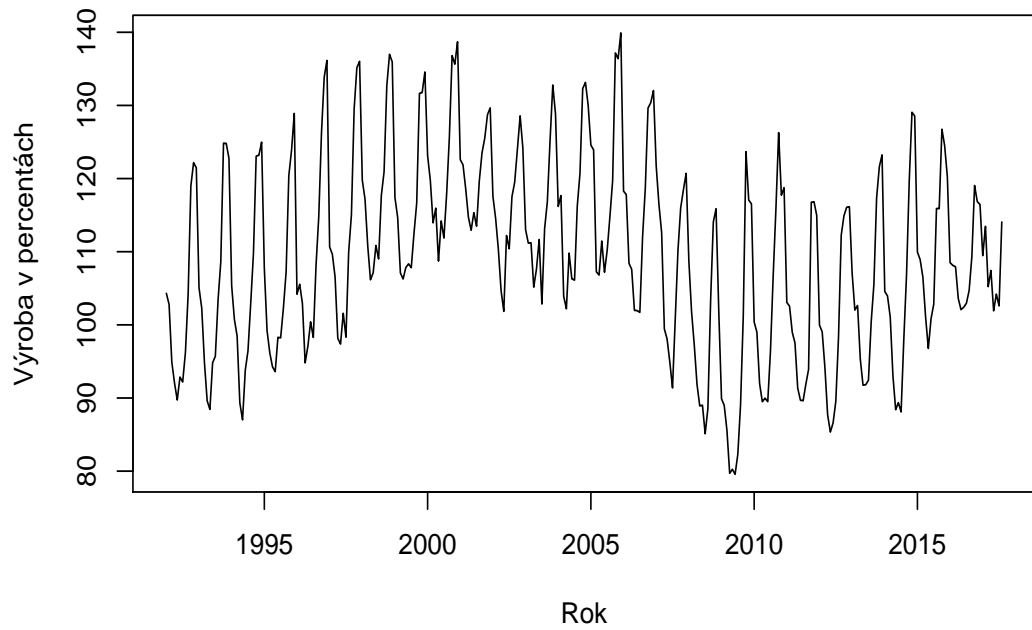
4.1 Softwareová realizácia

K výpočtom budeme používať štatistický software R rozšírený balíčkom *forecast*, v ktorom sú už všetky využívané modely a metódy implementované. Využívať budeme predovšetkým funkcie `HoltWinters` a `forecast`, ktoré vracajú najvhodnejšie odhadnuté výsledky jednotlivých verzií Holt-Wintersovej metódy pomocou postupov popísaných v odstavci 1.6. Požadovaná verzia sa volí použitím argumentu `seasonal` funkcie `HoltWinters`, kedy si môžeme vybrať z možností *additive* a *multiplicative*. Obe funkcie pracujú s objektami triedy `TS`, ktoré získame použitím rovnomennej funkcie na numerický vektor so zadefinovaným začiatkom, koncom a frekvenciou časového radu pomocou argumentov `start`, `end`, `frequency` (napr. prvé pozorovanie mesačného časového radu z januára 2018 by sme získali voľbou `start=c(2018,1)` a `frequency=12`). V momente, kedy bude analyzovaný časový rad vyrovnaný v podobe výstupu funkcie `HoltWinters`, aplikujeme naň funkciu `forecast`, ktorá vracia bodové predpovede pre zvolený časový interval spolu s intervalovými predpoveďami so zvolenými hladinami spoľahlivosti.

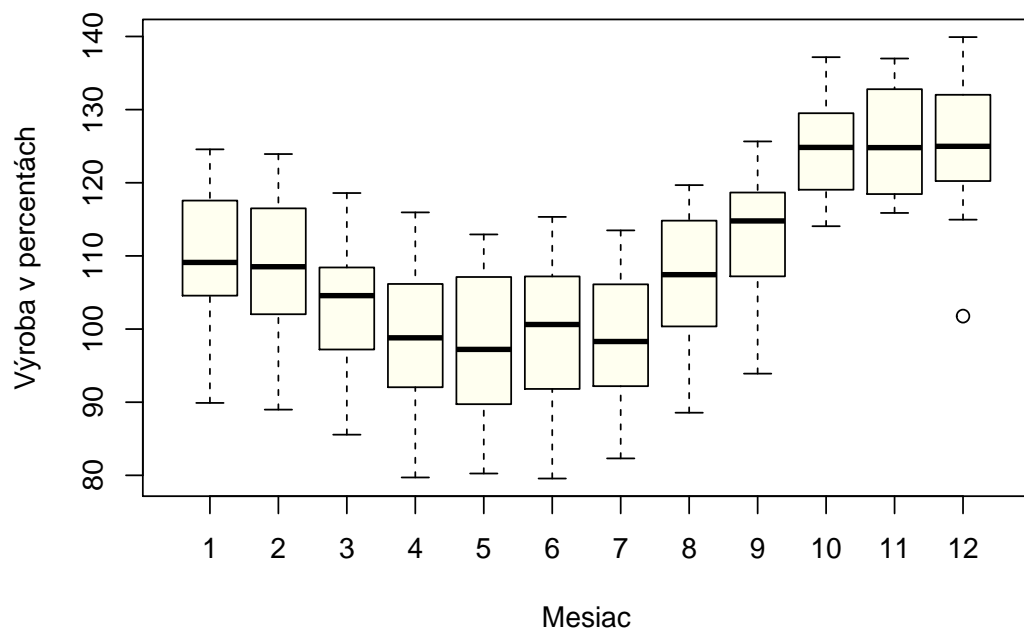
Príklad 1: Mesačná produkcia cukrovínok v USA

Ako prvý príklad uvidíme zahraničný časový rad dĺžky $n = 308$, ktorý je dostupný na portáli <https://research.stlouisfed.org/>. Máme k dispozícii mesačné údaje ($s = 12$), od januára 1992 do augusta 2017, o produkcii cukrovínok v USA v percentuálnom porovnaní s celkovou produkciou z roku 2012. Z obrázku 4.1 môžeme usúdiť, že časový rad obsahuje sezónnu zložku, ktorá indikuje závislosť medzi množstvom mesačne vyrobených cukrovínok a obdobím roka. Keďže sezónne výkyvy zostávajú so vzrastajúcim trendom približne na rovnakej úrovni, vhodnou metódou na tieto dáta sa javí aditívna Holt-Wintersova metóda.

Na boxplote 4.2 môžeme tieto sezónne výkyvy pozorovať na zvýšenej produkcii v období medzi mesiacmi október až december, ktorá sa odzrkadľuje vyššími strednými hodnotami a menšími rozptylmi výroby v tomto období v porovnaní so zvyškom roka. Jej dôvodom môže byť zvýšený dopyt po cukrovinkách pred tradičnými americkými sviatkami, akými sú Halloween či Deň Vďakyvdania a počas vianočného obdobia.

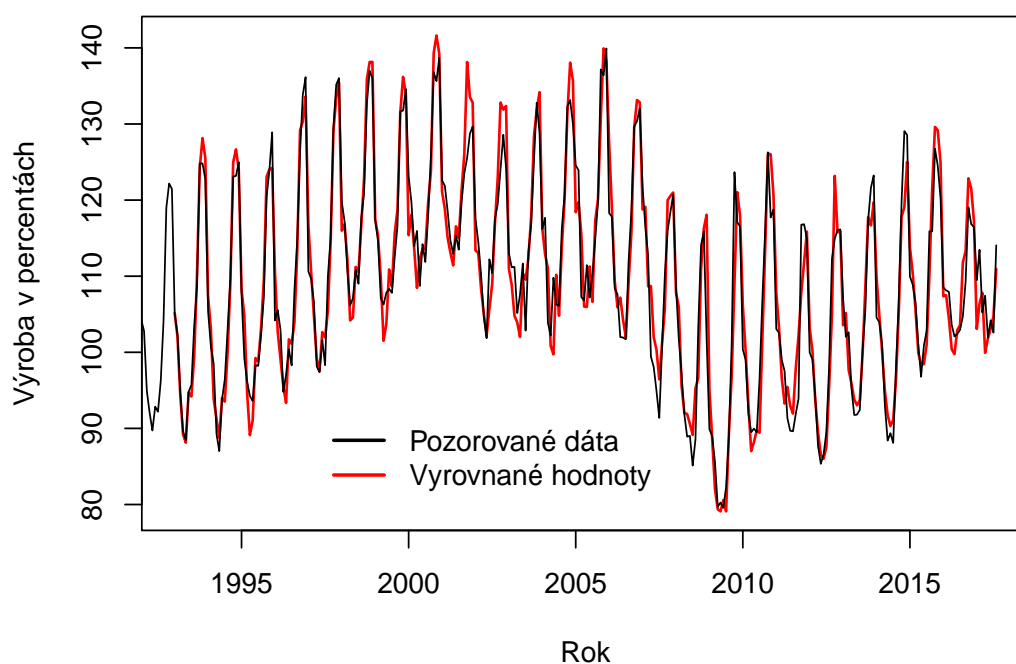


Obr. 4.1: Mesačná výroba cukroví v USA v rokoch 1992-2017



Obr. 4.2: Boxplot produkcie cukroví v USA v rokoch 1992-2017 podľa mesiacov

Na základe vyššie uvedených skutočností sme sa rozhodli použiť aditívnu Holt-Wintersovu metódu. Po použití funkcie `HoltWinters` s voľbou argumentu `seasonal = additive` dostávame vypočítané hodnoty \hat{y}_t , L_t , T_t a Sz_t pre $t = 1, \dots, 296$, ktoré nájdeme v súbore `candy.hw.csv`. Ako môžeme vidieť na obrázku 4.3, použitá metóda bola značne úspešná v zachytení vrcholov jednotlivých sezónnych výkyvov, keďže vyrovnaný rad sa do veľkej miery zhoduje s pozorovanými hodnotami.



Obr. 4.3: Mesačná výroba cukroví v USA v rokoch 1992-2017 a vyrovnanie pomocou aditívnej Holt-Wintersovej metódy

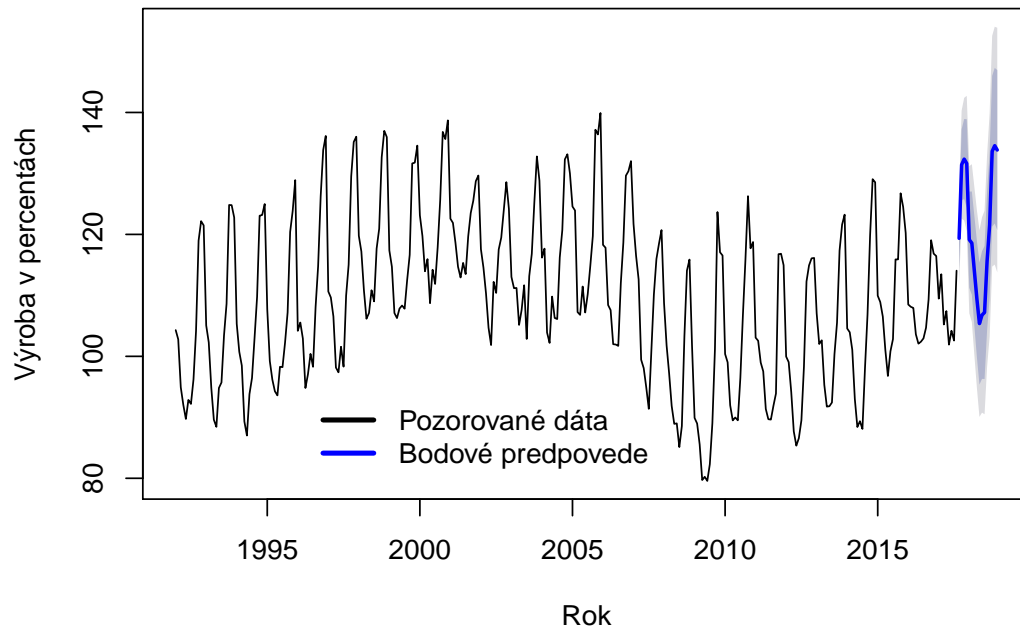
Počiatkové hodnoty $L_0, T_0, Sz_{-s+1}, \dots, Sz_0$, potrebné na realizáciu tejto metódy, odvodzuje funkcia `HoltWinters` pomocou jednoduchšej lineárnej regresie, postupu, ktorý sme priblížili v časti odstavca 3.1.1 venujúcej sa Holt-Wintersovej metóde. Tieto odhady sú obsiahnuté v súbore `candy.est.csv`.

Vyrovnávacie konštanty α, β a γ sú volené zo siete hodnôt podľa predpovedí s najmenšou SSE (1.3) ako

$$\alpha = 0.6258817, \quad \beta = 0, \quad \gamma = 0.4367585.$$

V porovnaní s pevnou voľbou vyrovnávacích parametrov $\alpha = \gamma = 0,4$ a $\beta = 0,1$ sme dostali lepšie výsledky, keďže sme SSE znížili približne o hodnotu 866.2.

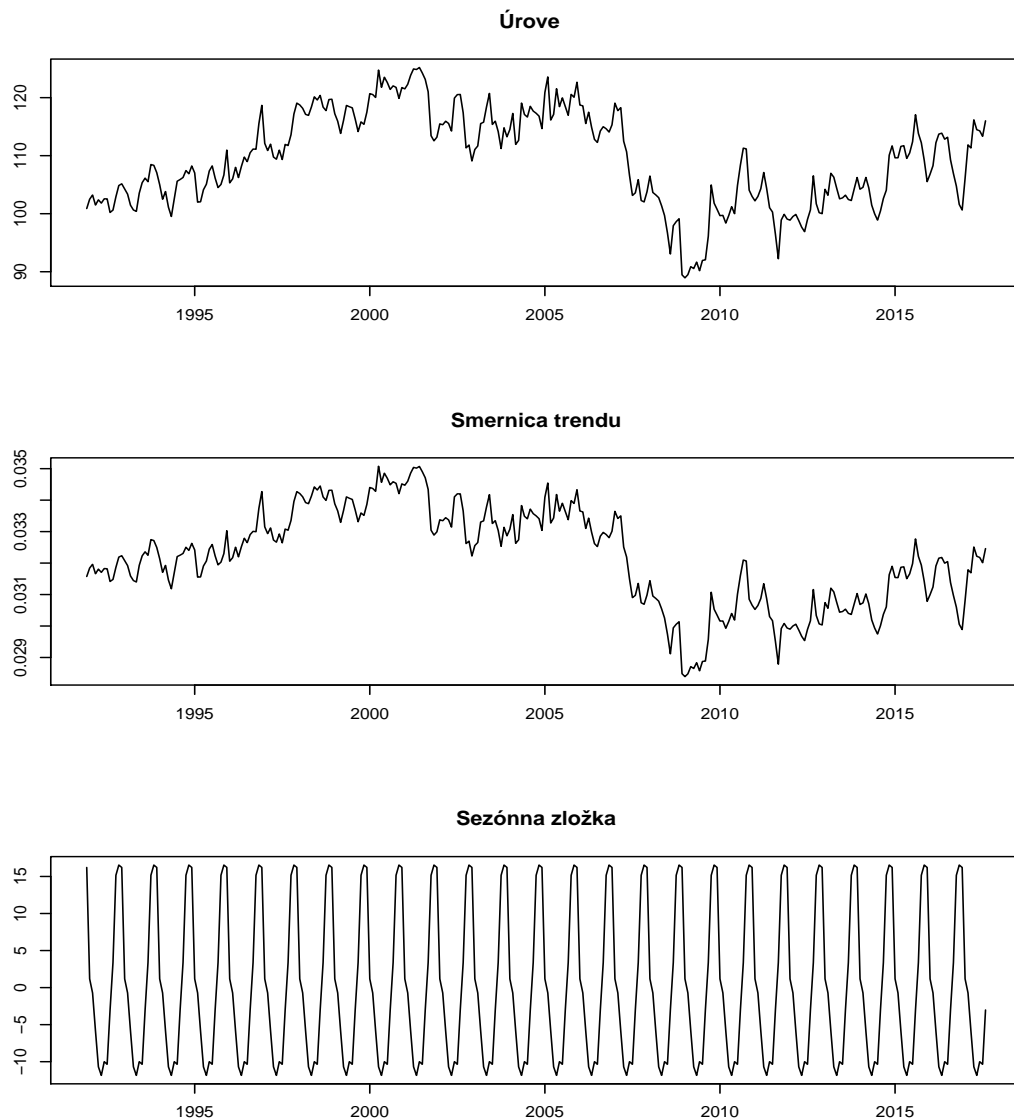
Obrázok 4.4 obsahuje bodové predpovede spolu s 80% a 95% predpovednými intervalmi pre uvažovaný rad do konca roka 2018, obdobia dĺžky 16 mesiacov od posledného pozorovaného údaj. Číselné hodnoty týchto predpovedí sú vypísané v súbore `candy.forecast.csv`.



Obr. 4.4: Bodové predpovede spolu s 80% a 95% intervalovými predpoveďami (znázornené tmavosivou a sivou farbou) získanými aditívnou Holt-Wintersovou metódou pre mesačnú výrobu cukroví v USA

Dekompozíciu nášho časového radu na jednotlivé zložky: úroveň, smernicu trendu a sezónnosť, pomocou aditívnej Holt-Wintersovej metódy znázorňuje obrázok 4.5.

Všetky súbory zmienené v tomto príklade je možné nájsť na priloženom CD.

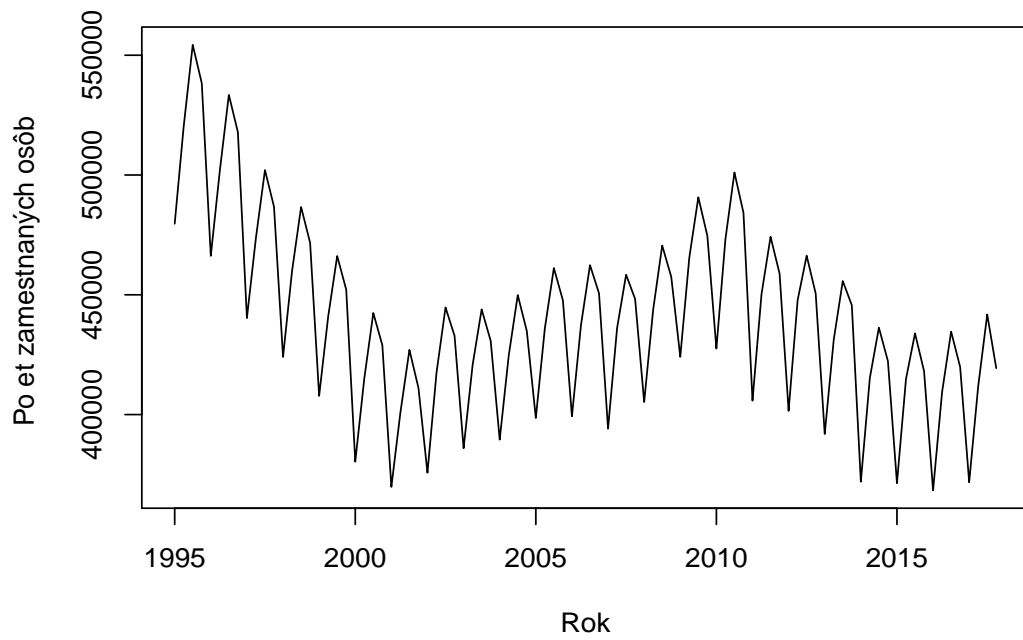


Obr. 4.5: Rozklad uvažovaného časového radu na jednotlivé zložky pomocou aditívnej Holt-Wintersovej metódy

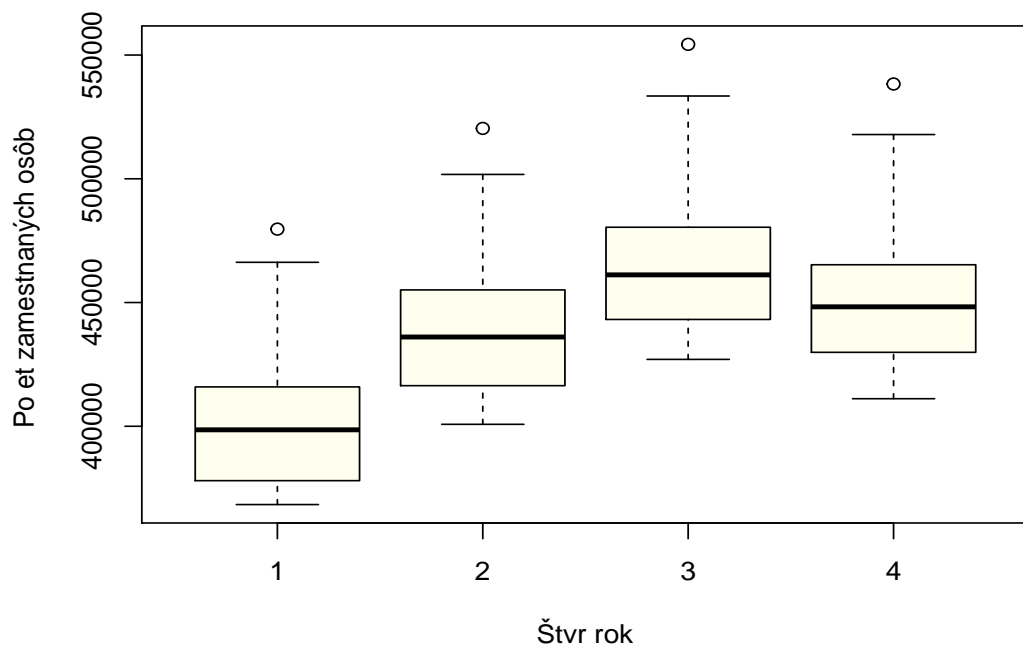
Príklad 2: Štvrťročná zamestnanosť v stavebníctve v ČR

Druhý analyzovaný časový rad tvorí celkový počet osôb zamestnaných v stavebníctve v Českej republike v rokoch 1995-2017. Máme k dispozícii štvrťročné údaje ($s = 4$) za 23 rokov (počet pozorovaní $n = 92$) pochádzajúce z Českého statistického úradu.

Z priebehu radu (obrázok 4.6) je evidentné, že rad obsahuje sezónnu zložku, ktorá sa zväčšuje, respektíve znižuje, s meniacou sa monotónnosťou trendu. Preto sú dáta vhodné pre použitie multiplikatívnej Holt-Wintersovej metódy. Sezónnosť je podložená aj obrázkom 4.7, na ktorom môžeme vidieť zreteľné rozdiely v množstve zamestnaných osôb medzi jednotlivými štvrťrokmi.

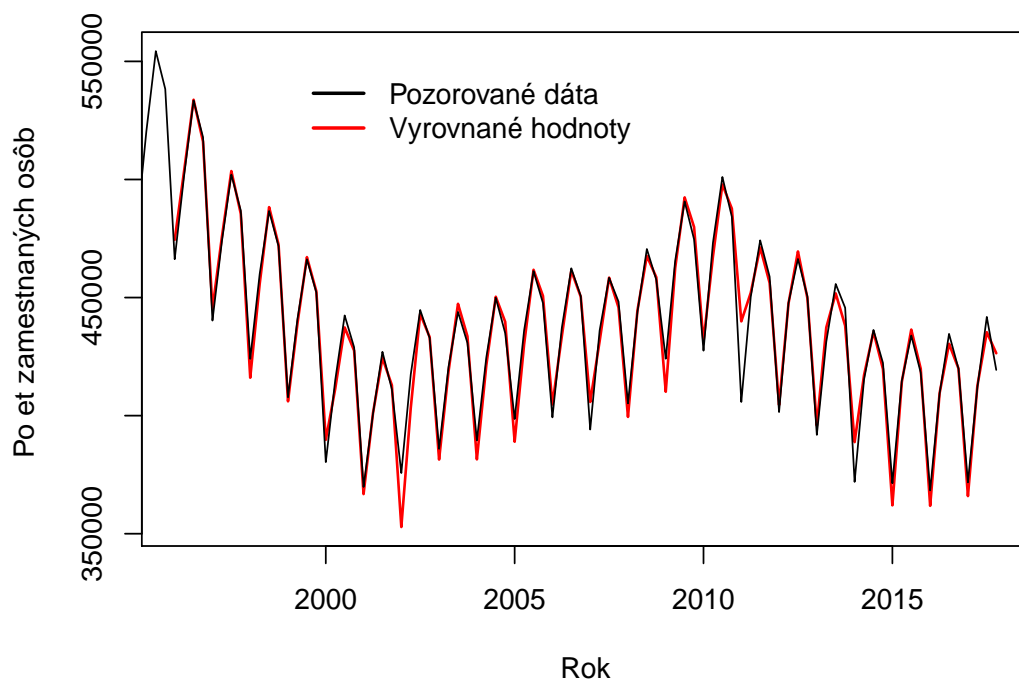


Obr. 4.6: Celková štvrtročná zamestnanosť v stavebníctve v ČR v rokoch 1995-2017



Obr. 4.7: Boxplot celkovej zamestnanosti v stavebníctve v ČR v rokoch 1995-2017 podľa štvrtrokov

Jedným z dôvodov tejto nezrovnalosti môže byť napríklad znížená intenzita stavebných prác v zimných mesiacoch a z nej plynúca nižšia zamestnanosť v sektore v tomto období.



Obr. 4.8: Celková štvrtročná zamestnanosť v stavebníctve v ČR v rokoch 1995-2017 a vyrovnanie pomocou multiplikatívnej Holt-Wintersovej metódy

Po použití funkcie `HoltWinters` s voľbou argumentu `seasonal = multiplicative` dostávame vypočítané hodnoty \hat{y}_t , L_t , T_t a Sz_t pre $t = 1, \dots, 88$, ktoré môžeme nájsť v súbore `zam.hw.csv`. Vyrovnané dáta pomocou multiplikatívnej Holt-Wintersovej metódy sú znázornené na obrázku 4.8. Môžeme si všimnúť, že vyrovnaný rad pomerne dobre kopíruje skutočné dáta a konštatovať úspech použitej metódy na náš časový rad.

Počiatkové hodnoty $L_0, T_0, Sz_{-s+1}, \dots, Sz_0$ odvodila funkcia `HoltWinters` nasledovne:

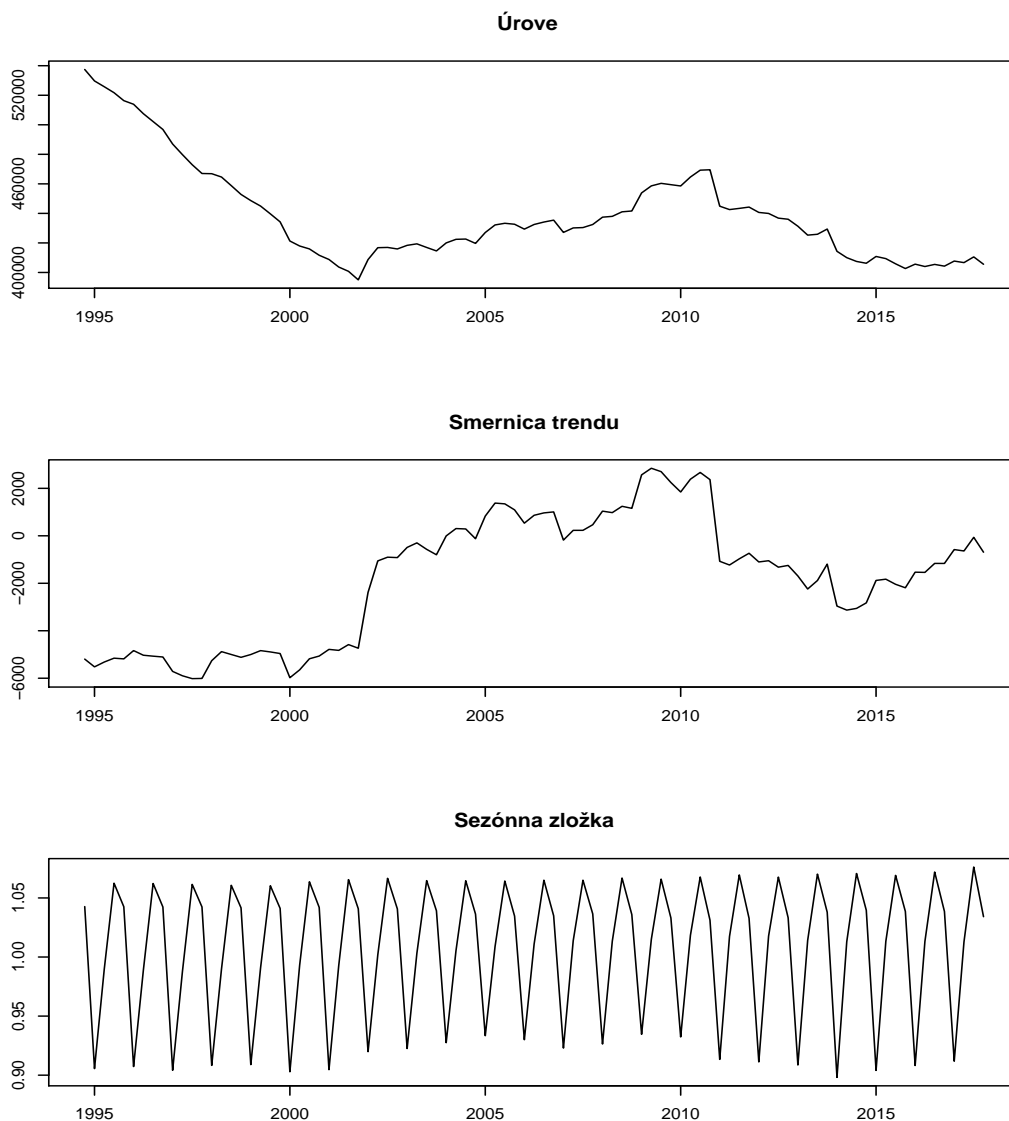
$$L_0 = 406225.4, T_0 = -675.9699$$

$$Sz_{-3} = 0.910094, Sz_{-2} = 1.010982, Sz_{-1} = 1.074214, Sz_0 = 1.032430$$

Výber vyrovnávacích konštánt prebieha rovnako ako v prvom príklade, teda s cieľom dostať vyrovnané hodnoty s čo najmenšou SSE, a tie sme získali pri nasledujúcej voľbe parametrov α, β, γ :

$$\alpha = 0.7343332, \quad \beta = 0.1314865, \quad \gamma = 0.9801001.$$

V porovnaní s pevnou voľbou $\alpha = \gamma = 0,4$ a $\beta = 0,1$ sme aj tentokrát dosiahli presnejšie výsledky s výrazne menšou SSE.

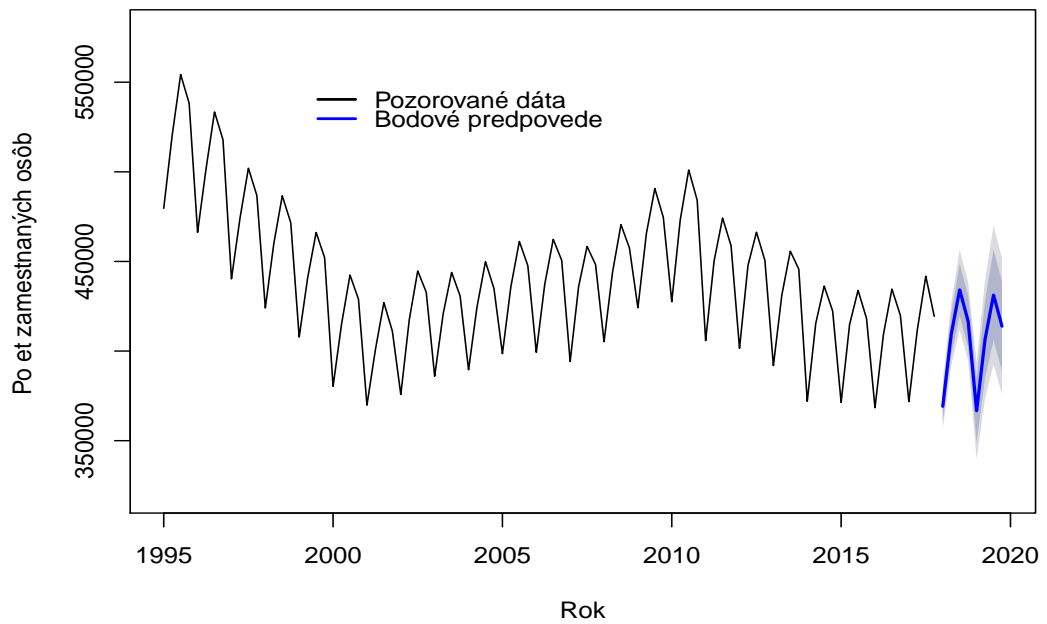


Obr. 4.9: Rozklad uvažovaného časového radu na jednotlivé komponenty pomocou multiplikatívnej Holt-Wintersovej metódy

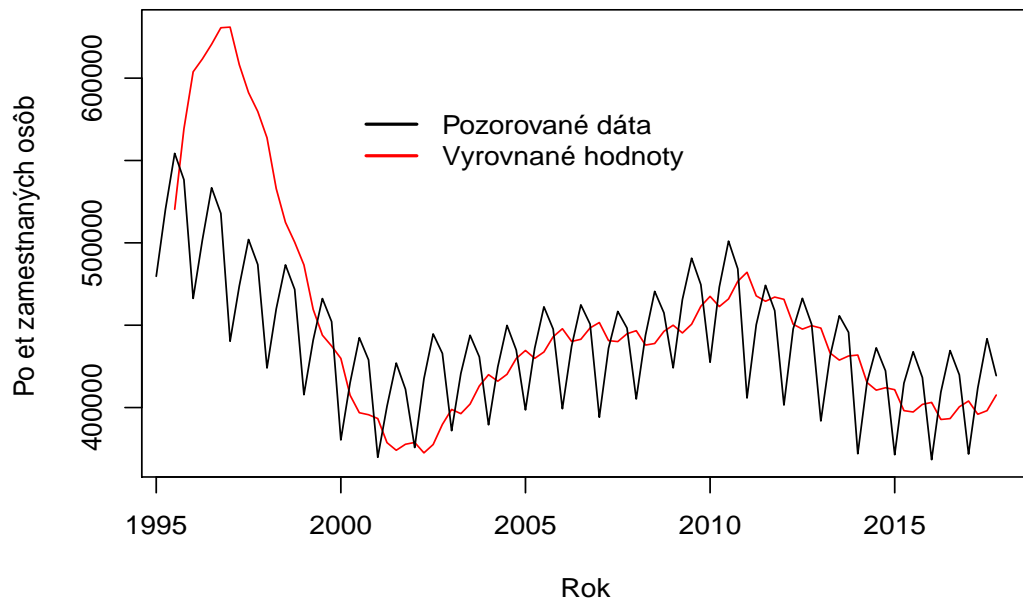
Jednotlivé komponenty radu môžeme vidieť na obrázku 4.9. Obrázok 4.10 zobrazuje bodové predpovede spolu s 80% a 95% predpovednými intervalmi pre uvažovaný rad do konca roka 2019, ktorých číselné hodnoty sú obsiahnuté v súbore `zam.forecast.csv`.

Pre porovnanie na záver uvádzame aj vyrovnanie uvažovaného časového radu Holtovou metódou (1.5). Holtovu metódu sme implementovali pomocou funkcie `HoltWinters` voľbou $\alpha = \beta = 0,2$ a $\gamma = false$ a počiatočných hodnôt $L_0 = y_1$ a $T_0 = y_2 - y_1$. Na obrázku 4.11 vidíme, že vyrovnaný časový rad sa do veľkej miery odlišuje od pozorovaných hodnôt. Tento výsledok by nás však nemal prekvapiť, keďže Holtova metóda neberie do úvahy sezónnu zložku, a preto nie je vhodná práve pre časové rady vykazujúce sezónnosť.

Všetky súbory, ktoré sme v tomto príklade spomenuli, je možné nájsť na priloženom CD.



Obr. 4.10: Bodové predpovede spolu s 80% a 95% intervalovými predpoveďami (znázornené tmavosivou a sivou farbou) získanými multiplikatívnou Holt-Wintersovou metódou pre celkovú štvrťročnú zamestnanosť v stavebníctve v ČR



Obr. 4.11: Celková štvrťročná zamestnanosť v stavebníctve v ČR v rokoch 1995-2017 a vyrovnanie pomocou Holtovej metódy

Záver

V tejto práci sme sa zaoberali metódami exponenciálneho vyrovnávania, ktoré patria medzi často používané dekompozičné techniky pre modelovanie časových radov. Sústredili sme sa predovšetkým na možnosti využitia týchto metód na časových radoch, u ktorých môžeme pozorovať výskyt sezónnosti, s cieľom lepšie popísať dlhodobé zákonitosti v daných radoch a konštruovať presnejšie predpovede do budúcnosti.

V prvej časti práce sme zaviedli nevyhnutné pojmy slúžiace k lepšiemu porozumeniu skúmaného problému a vysvetlili sme princíp dekompozície časových radov na jednotlivé zložky. Následne sme individuálne popísali metódy jednoduchého a dvojitého exponenciálneho vyrovnávania a Holtovu metódu, ktoré sú vhodné na dáta bez sezónnosti a ich voľba závisí od typu trendu. Na sezónne časové rady sa aplikuje Holt-Wintersova metóda, ktorú sme predstavili v dvoch variantách pre aditívnu aj multiplikatívnu sezónnosť.

V druhej kapitole sme upozornili na nutnosť uvedomenia si hlavného rozdielu medzi vyrovnávacou metódou a stochastickým modelom. Ten navyše prináša aj použitie ďalších štatistických nástrojov na hlbšie skúmanie časového radu. Ďalej je v nej obsiahnutý prehľad sezónnych stavových modelov rozdelených podľa typu sezónnosti.

V tretej kapitole sme diskutovali niektoré problémy spojené s praktickou realizáciou metód exponenciálneho vyrovnávania, pričom sme tiež uviedli v praxi doporučované riešenia týchto problémov.

V poslednej časti práce sme popísali implementáciu Holt-Wintersovej metódy v softwari R a ukázali jej aplikáciu na dvoch sezónnych časových radoch, pričom jej variantu sme volili na základe charakteru sezónnosti. V oboch príkladoch sme bližšie skúmali zmienenu sezónnosť so snahou interpretovať jej výskyt v daných radoch. Na našich výsledkoch podložených aj grafickou realizáciou môžeme pozorovať, že Holt-Wintersova metóda bola úspešná v zachytení sezónnych výkyvov v oboch prípadoch a na záver konštatovať, že daná metóda patrí medzi efektívne dekompozičné nástroje.

Zoznam použitej literatúry

T. Cipra. *Finanční Ekonometrie*. 2., upravené vydání. Ekopress, s.r.o, Praha, 2013. ISBN 978-80-86929-93-4.

Český statistický úřad. Hrubý domácí produkt - Časové řady ukazatelů čtvrtletních účtů. https://www.czso.cz/csu/czso/hdp_cr, 2017.

Federal Reserve Bank of St. Louis. Industrial production: Nondurable goods: Sugar and confectionery product. <https://fred.stlouisfed.org/series/IPG3113N>, 2017.

R. Hyndman, A.B. Koehler, J.K. Ord, and R.D. Snyder. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Series in Statistics. Springer-Verlag Berlin Heidelberg, Berlin, 2008. ISBN 978-3-540-71916-8.

Rob Hyndman, Christoph Bergmeir, Gabriel Caceres, Leanne Chhay, Mitchell O'Hara-Wild, Fotios Petropoulos, Slava Razbash, Earo Wang, and Farah Yasmeen. *forecast: Forecasting functions for time series and linear models*, 2018. URL <http://pkg.robjhyndman.com/forecast>. R package version 8.3.

Zoznam obrázkov

4.1	Mesačná výroba cukrovínok v USA v rokoch 1992-2017	17
4.2	Boxplot produkcie cukrovínok v USA v rokoch 1992-2017 podľa mesiacov	17
4.3	Mesačná výroba cukrovínok v USA v rokoch 1992-2017 a vyrovnanie pomocou aditívnej Holt-Wintersovej metódy	18
4.4	Bodové predpovede spolu s 80% a 95% intervalovými predpoveďami (znázornené tmavosivou a sivou farbou) získanými aditívnou Holt-Wintersovou metódou pre mesačnú výrobu cukrovínok v USA	19
4.5	Rozklad uvažovaného časového radu na jednotlivé zložky pomocou aditívnej Holt-Wintersovej metódy	20
4.6	Celková štvrtročná zamestnanosť v stavebníctve v ČR v rokoch 1995-2017	21
4.7	Boxplot celkovej zamestnanosti v stavebníctve v ČR v rokoch 1995-2017 podľa štvrtrokov	21
4.8	Celková štvrtročná zamestnanosť v stavebníctve v ČR v rokoch 1995-2017 a vyrovnanie pomocou multiplikatívnej Holt-Wintersovej metódy	22
4.9	Rozklad uvažovaného časového radu na jednotlivé komponenty pomocou multiplikatívnej Holt-Wintersovej metódy	23
4.10	Bodové predpovede spolu s 80% a 95% intervalovými predpoveďami (znázornené tmavosivou a sivou farbou) získanými multiplikatívnou Holt-Wintersovou metódou pre celkovú štvrtročnú zamestnanosť v stavebníctve v ČR	24
4.11	Celková štvrtročná zamestnanosť v stavebníctve v ČR v rokoch 1995-2017 a vyrovnanie pomocou Holtovej metódy	24

Zoznam tabuliek

2.1	Modely s aditívnou sezónnou zložkou	12
2.2	Modely s multiplikatívnou sezónnou zložkou	13