APLIKÁCIA EXPONENCIÁLNEHO VYROVNÁVANIA ČASOVÝCH RADOV

PhDr. Eva Ostertagová, PhD.
Technická univerzita v Košiciach
Fakuta elektrotechniky a informatiky
Katedra matematiky a teoretickej informatiky
Němcovej 32, 042 00 Košice
eva.ostertagova@tuke.sk

Abstract

The aim of article is an application of exponential smoothing method for time series analysis by means of MATLAB software.

Key words

Time series analysis, trend component, exponential smoothing, prediction, root mean square error.

ÚVOD

Pri modelovaní reálnych časových radov sa často stáva, že v priebehu analyzovaného obdobia sa hodnoty štrukturálnych parametrov modelu v čase menia, prípadne dochádza aj k zmenám analytického tvaru modelu. Uvedené dôvody boli podnetom ku konštrukcii tzv. adaptívnych modelov, ktoré sa označujú aj ako modely s premenlivými parametrami.

Z metodologického hľadiska sú adaptívne modely veľmi podobné klasickým trendovým modelom. Obe spomenuté triedy modelov časových radov sú totiž orientované len na dôkladný opis priebehu analyzovanej premennej v čase a vôbec nie na objasnenie kauzálneho (príčinného) mechanizmu vývoja tejto premennej prostredníctvom skúmania dynamiky iných premenných. Adaptívne modely sa od klasických trendových modelov v zásade líšia tým, že nepredpokladajú stabilitu analytického tvaru trendovej funkcie ani jej štrukturálnych parametrov v čase a nepredpokladajú ani spojitosť trendovej funkcie. Na základe novo získaných údajov sú tieto modely plynule korigované a aktualizované a môžu tak pracovať aj s takými trendovými zložkami, ktorých charakter sa v čase výrazne a nepravidelne mení.

1. Analýza trendovej zložky časových radov

Hlavnou úlohou analýzy časových radov je vystihnutie základnej tendencie ich vývoja, teda stanovenie ich trendu. Trend sa určuje metódami, ktoré sa súhrnne nazývajú vyrovnávanie, resp. vyhladzovanie časových radov, t.j. nahradenie časového radu empirických hodnôt $y_1, y_2, \ldots, y_n$ radom hodnôt bez periodického a náhodného kolísania.

Budeme uvažovať neperiodický časový rad $y_i = T_i + \varepsilon_i$, $i = 1, 2, \ldots, n$. Pri jeho spracovaní sa snažíme nájsť jeho trendovú zložku T_i očistenú od náhodných vplyvov ε_i. Vyrovnané hodnoty označíme vo všeobecnosti $\hat{y}_i$. Podobne môžeme vyrovnať aj periodický časový rad, keď jeho hodnoty očistíme od periodickej a náhodnej zložky. Vyrovnávanie časových radov sa realizuje najčastejšie grafickou metódou, analytickou metódou a adaptívnymi metódami.

Grafické vyrovnávanie časových radov spočíva v tom, že sa lomená čiara v spojnicovom grafe skusmo „od ruky“ preloží vyrovnávajúcou čiarou, t.j. lineárnou alebo nelineárnou funkciou tak, aby sústredenie bodov okolo tejto funkcie bolo čo najtesnejšie. Tento spôsob je pomerne rýchly, je to však len odhad, ktorý je veľmi často subjektívny.

Analytické vyrovnávanie časových radov spočíva vo vystihnutí priebehu časového radu analytickou funkciou (regresia). Je to najpoužívanejší spôsob opisu vývoja ukazovateľov časových radov, ktorý umožňuje aj ich extrapoláciu, t.j. jednoduché odhady budúceho vývoja časových radov.

Vyrovnávanie časových radov adaptívnymi metódami spočíva v nahradení pôvodných empirických hodnôt časového radu radom aritmetických priemerov (prostých, resp. vážených), ktoré sa vypočítajú z väčšieho či menšieho počtu po sebe idúcich empirických hodnôt časového radu. Adaptívne metódy nepredpokladajú nemennosť parametrov modelu ani stabilitu analytického tvaru ani stabilitu štrukturálnych parametrov v priebehu sledovaného obdobia. Taktiež nepožadujú ani spojitosť trendovej funkcie. K adaptívnym metódam patrí predovšetkým *metóda kĺzavých priemerov* a *exponenciálne vyrovnávanie*.

2. Exponenciálne vyrovnávanie časových radov

Metóda je odvodená od metódy kĺzavých priemerov, pričom odstraňuje jej najväčšiu nevýhodu, a to chýbajúce hodnoty na koncoch.

Exponenciálne vyrovnávanie sa zakladá na myšlienke, že pri konštrukcii extrapolačnej prognózy sa najvyššia váha pridelí poslednej hodnote časového radu, pretože táto obsahuje najviac informácií o možných zmenách vo vývoji sledovaného ukazovateľa a očakáva sa aj ich čiastočný prenos do budúcnosti. Odhad trendu je získavaný vo forme lineárnej kombinácie všetkých predchádzajúcich pozorovaní, pričom váhy skorších pozorovaní exponenciálne klesajú.

Existuje a navrhnutých bolo viacero metód exponenciálneho vyrovnávania, z ktorých sa najčastejšie používajú *Brownove, Holtove* a *Wintersove modely*.

V rámci techniky *Brownovho exponenciálneho vyrovnávania* je možné rozlíšiť tri základné varianty:

- *Jednoduché exponenciálne vyrovnávanie*, ktoré sa využíva vtedy, ak v určitých rovnako dlhých časových úsekoch je trend časového radu *konštantný*, ale pritom sa pomaly meniaci v čase.
- *Dvojité exponenciálne vyrovnávanie* – využíva sa vtedy, ak časový rad možno na rovnako dlhých časových úsekoch vyrovnať *lineárnou funkciou*.
- *Trojité exponenciálne vyrovnávanie* – trend je v krátkych úsekoch časového radu modelovaný *kvadratickou funkciou*.

Princíp techniky exponenciálneho vyrovnávania vysvetlíme na modeli jednoduchého exponenciálneho vyrovnávania. Aj keď je model konštantného trendu zdanlivo zriedkavý, používa sa pre časové rady so značne nerovnomerným vývojom. Nerovnomernosť vývoja môže byť pritom zapríčinená sezónnosťou alebo silným vplyvom náhodných zložiek.

Brownov model jednoduchého exponenciálneho vyrovnávania vychádza z predstavy, že každé pozorovanie analyzovaného časového radu je možné vyjadriť v tvare súčtu konštantnej deterministickej zložky a náhodnej (reziduálnej) zložky. Teda v *i*-tom období je hodnota ukazovateľa časového radu daná súčtom $y_i = T_i + \varepsilon_i$, $i = 1, 2, \ldots, n$; kde $T_i = a_0$, $a_0 \in R$. O konštantnej zložke sa predpokladá, že v každom úseku časového radu zostáva relatívne stabilná, ale v čase môže prechádzať pomalými zmenami. Úlohou je teda nájsť odhad parametra a_0, ktorý sa v tomto prípade rovná vyrovnanej hodnote $\hat{y}$.

Prognostické využitie tejto metódy je možné zapísať jednoducho takto:

nová predpoveď = stará predpoveď + určitá časť z chyby predpovede.

Jednoduchý vzorec, ktorý vyjadruje tento princíp je: nová predpoveď = stará predpoveď + $\alpha \cdot$ (najneskoršie pozorovanie – stará predpoveď), kde α je vyrovnávacia konštanta.

Odhad trendu, t.j. vyrovnanú hodnotu časového radu, je možné vypočítať podľa nasledujúceho rekurentného vzorca

$$\hat{y}_2 = y_1,\ \hat{y}_i = \hat{y}_{i-1} + \alpha \cdot (y_{i-1} - \hat{y}_{i-1}),$$

pre $i = 3, \ldots, n$; kde $0 < \alpha < 1$ je zvolená konštanta, tzv. *vyrovnávacia konštanta*, y_{i-1} – empirická hodnota ukazovateľa v (*i*–1)-tom období, $\hat{y}_{i-1}$ – vyrovnaná hodnota ukazovateľa v (*i*–1)-tom období, ktorá bola prognózovaná na toto obdobie; $\hat{y}_i$ – nová vyrovnaná hodnota ukazovateľa, ktorá je prognózou na *i*-té obdobie. Hodnota $\hat{y}_1$ nie je definovaná.

Po úprave dostaneme rekurentný vzorec v tvare $\hat{y}_i = \alpha \cdot y_{i-1} + (1-\alpha) \cdot \hat{y}_{i-1}$.

Určenie optimálnej hodnoty vyrovnávacej konštanty α je zásadným problémom aplikácie uvedeného modelu. Jej hodnota je závislá od charakteru zmien analyzovaného ukazovateľa. Ak počítame s možnosťou rýchlych a nepravidelných zmien trendu v čase, je logické pripisovať väčšiu váhu najnovším pozorovaniam a menšiu váhu odhadu trendu v predchádzajúcom časovom období. V tomto prípade budeme teda voliť α blízke 1, čo znamená rýchle tlmenie vplyvu minulých pozorovaní. Ak naopak predpokladáme, že bude pokračovať doterajší vývoj, t.j. že k zmenám trendu bude dochádzať pozvoľne, je vhodné voliť α blízke 0, čo znamená relatívne pomalé tlmenie vplyvu minulých pozorovaní.

Je potrebné poznamenať, že uvedené odporúčania sú len orientačné. V reálnych aplikáciách exponenciálneho vyrovnávania sa hodnota vyrovnávacej konštanty α určuje spravidla experimentálne pomocou „metódy skúšok a omylov“. Táto metóda spočíva v postupnom skúšaní rôznych hodnôt α, napr. $0,1;\ 0,2;\ \ldots;\ 0,9$ (prípadne aj s jemnejším krokom). Potom za najvhodnejšiu hodnotu vyrovnávacej konštanty α vyberáme tú, ktorá minimalizuje priemernú kvadratickú chybu odhadu *MSE*, resp. jej druhú odmocninu *RMSE*. Pre *RMSE* platí vzťah $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$, kde $e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n$; nazývame *rezíduá časového radu*. Sú to odhady nesystematickej zložky skúmaného časového radu.

3. Konkrétny príklad s použitím MATLABu

Máme k dispozícii údaje ŠÚSR o ročných priemerných hodnotách koncentrácie prízemného ozónu (y_i) v μg·m^{-3}, ktoré boli namerané v Košiciach (Ďumbierska) v rokoch 1996 – 2009: 55, 43, 40, 34, 60, 58, 64, 68, 60, 67, 49, 57, 56, 81. Úlohou je odhadnúť priemernú koncentráciu ozónu v Košiciach v roku 2010.

Keďže daný časový rad vykazuje fluktuácie, je teda otázkou, do akej miery sú splnené predpoklady modelu jednoduchého exponenciálneho vyrovnávania. Hodnoty časového radu aj napriek istým odchýlkam sa pohybujú od začiatku do konca v približne rovnakom intervale, teda je možné usudzovať, že trendová zložka by mohla byť konštantná.

Použijeme uvedený rekurentný vzorec, pričom vyskúšame rôzne hodnoty vyrovnávacej konštanty α. Pre ilustráciu však uvedieme výsledky len pre tri rôzne hodnoty vyrovnávacej konštanty α aj s vypočítanými hodnotami *RMSE*.

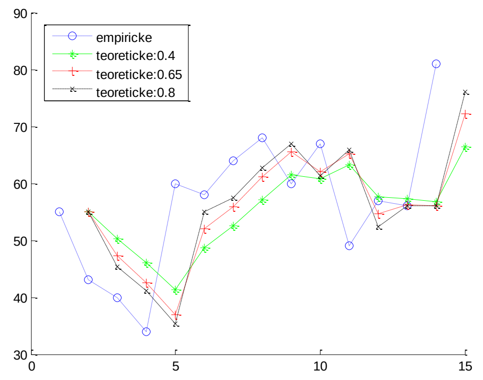
Realizácia úlohy v programe MATLAB:

```
z=[55,43,40,34,60,58,64,68,60,67,49,57,56,81];
n=length(z);a=[0.4 0.65 0.8];
yi1=[];
yi1(1)=NaN;
yi1(2)=z(1);
for i=3:15
    yi1(i)=a(1)*z(i-1)+(1-a(1))*yi1(i-1);
end
yi2=[];
yi2(1)=NaN;
yi2(2)=z(1);
for i=3:15
    yi2(i)=a(2)*z(i-1)+(1-a(2))*yi2(i-1);
end
yi3=[];
yi3(1)=NaN;
yi3(2)=z(1);
for i=3:15
    yi3(i)=a(3)*z(i-1)+(1-a(3))*yi3(i-1);
end
RMSE1=...
sqrt((1/(n-1)*sum((z(2:end)-yi1(2:14)).^2)));
RMSE2=...
sqrt((1/(n-1)*sum((z(2:end)-yi2(2:14)).^2)));
RMSE3=...
sqrt((1/(n-1)*sum((z(2:end)-yi3(2:14)).^2)));
RM=[NaN,NaN,RMSE1,RMSE2,RMSE3];
i=1:15;
fprintf('\ti\t\tYi\t\t\tY^i(0.4) Y^i(0.65) Y^i(0.8)\n')
t=[i;z,NaN;yi1;yi2;yi3]';
t=[t;RM];
disp(t)
hold on
plot(i(1:14),z,'o:')
plot(i(2:15),yi1(2:15),'*g-')
plot(i(2:15),yi2(2:15),'+r-.')
plot(i(2:15),yi3(2:15),'xk--')
legend('empiricke','teoreticke:0.4',...
'teoreticke:0.65','teoreticke:0.8')
hold off
```

Výsledky riešenia úlohy pre $\alpha = 0,4;\ 0,65;\ 0,8$ sú uvedené v tab.1 a na obr.1.

Tab.1 Tabuľka výsledkov z MATLABu

i	Yi	Y^i(0.4)	Y^i(0.65)	Y^i(0.8)
1.0000	55.0000	NaN	NaN	NaN
2.0000	43.0000	55.0000	55.0000	55.0000
3.0000	40.0000	50.2000	47.2000	45.4000
4.0000	34.0000	46.1200	42.5200	41.0800
5.0000	60.0000	41.2720	36.9820	35.4160
6.0000	58.0000	48.7632	51.9437	55.0832
7.0000	64.0000	52.4579	55.8803	57.4166
8.0000	68.0000	57.0748	61.1581	62.6833
9.0000	60.0000	61.4449	65.6053	66.9367
10.0000	67.0000	60.8669	61.9619	61.3873
11.0000	49.0000	63.3201	65.2367	65.8775
12.0000	57.0000	57.5921	54.6828	52.3755
13.0000	56.0000	57.3553	56.1890	56.0751
14.0000	81.0000	56.8132	56.0661	56.0150
15.0000	NaN	66.4879	**72.2732**	76.0030
NaN	NaN	12.1387	*12.0742*	12.1462



Obr. 1 Grafické znázornenie empirických hodnôt a teoretických hodnôt pre $\alpha = 0,4;\ 0,65;\ 0,8$

Z posledného riadku tabuľky 1 je zrejmé, že najmenšia hodnota *RMSE* bola dosiahnutá pre hodnotu $\alpha = 0,65$; a to $RMSE = 12,0742$. Teda najpresnejšou predpoveďou priemernej koncentrácie ozónu v Košiciach v roku 2010 je hodnota 72,2732 $\mu g \cdot m^{-3}$ (viď posledný riadok tabuľky 1).

ZÁVER

Jednoduché exponenciálne vyrovnávanie nachádza svoje uplatnenie najmä pre krátkodobé predpovede. Metóda je vhodná pre pozorovania, ktorých hodnoty kolíšu okolo priemeru, t.j. nie je pri nich zjavný trend. Výhodou metódy je aj jej výpočtová nenáročnosť a nízke náklady.

Literatúra

[1] ARLT, J.: Moderní metody modelování časových řad. Grada Publishing, Praha, 1999.

[2] CYHELSKÝ, L., KAŇOKOVÁ, J., NOVÁK, I.: Základy teorie statistiky pro ekonomy. SNTL/Alfa, Praha, 1979.

[3] HANČLOVÁ, J., TVRDÝ, L.: Úvod do analýzy časových řad. VŠB-TU, Ostrava, 2003.

[4] CHAJDIAK, J., RUBLÍKOVÁ, E., GUDÁBA, M.: Štatistické metódy v praxi. Statis, Bratislava, 1994.

[5] KŘIVÝ, I.: Analýza časových řad. Ostravská univerzita, 2006.

[6] LUCEY, T.: Quantitative Techniques. DP Publications Ltd., London, 1992.

[7] OSTERTAGOVÁ, E.: Aplikovaná štatistika. Elfa, Košice, 2011.

[8] PIRČ, V., OSTERTAGOVÁ, E.: *Matematika s MATLABom*. Košice, 2007.

[9] SVATOŠOVÁ, L., KÁBA, B.: Statistické metódy 2. PEF ČZU, Praha, 2008.

Príspevok bol spracovaný v rámci riešenia grantovej úlohy KEGA 3/7353/09.