Prognózovanie s využitím exaktných metód

Martin Lukáčik – Juraj Pekár

Medzi základné exaktné metódy prognózovania patria

- *extrapolačné metódy*, ktoré využívajú predpoklad, že skúmané javy sa dajú rozložiť na kvantifikovateľné zložky a že ich vzájomné súvislosti sú pre budúcnosť rozhodujúce
- *ekonometrické metódy*, ktoré sa snažia hľadať závislosť medzi skúmaným radom a inými súvisiacimi premennými buď v priamom jednorovnicovom modeli alebo priamo a súčasne aj cez sprostredkovaný vplyv v simultánnom viacrovnicovom modeli.

Extrapolačné metódy sa pri prognózovaní vždy využívajú ako základný rámec, ktorý poskytuje bázické hodnoty pre porovnávanie všetkých následne realizovaných modelov:

1. *dekompozičné metódy* – analyzovaný rad sa chápe ako súhrn základných zložiek: trendu, sezónnej zložky, cyklickej zložky a náhodnej zložky, ktoré sa snažíme kvantifikovať a popísať funkciami časovej premennej (analytickým vyrovnávaním v prípade; pravidelného vývoja) alebo kĺzavými priemermi (v prípade nepravidelného vývoja)
2. *Box-Jenkinsova metodológia* – vysvetľuje vývoj skúmaného časového radu náhodnou zložkou a nekonštruuje modely na základe teórie, ale priamo z empirických dát;
3. *spektrálna analýza* – k časovým radom sa pristupuje pomocou Fourierovej analýzy a tým dekompozície radu v tvare množiny sínusových a kosínusových kriviek s rôznymi amplitúdami a frekvenciami;
4. *viacrozmerná analýza časových radov* – analyzuje súvislosti medzi skúmaným časovým radom a inými radmi o ktorých predpokladáme, že určitým spôsobom ovplyvňujú skúmaný rad – patria sem faktorové modely alebo kombinované modely s vysvetľujúcimi aj časovými premennými.

Meranie presnosti prognóz

Cieľom hodnotenia presnosti prognóz je analýza a minimalizácia ich chýb. Chyby prognózy – teda rozdielu skutočnej a prognózovanej hodnoty – sa posudzujú individuálne pre jednotlivé časové horizonty (podhodnotenie alebo nadhodnotenie), ale pre porovnanie rôznych použitých postupov sa posudzujú skôr súhrnne pomocou rôznych mier.

Nech r_t je chyba prognózy definovaná ako rozdiel napozorovanej reálnej hodnoty y_t a prognózovanej hodnoty $\hat{y}_t$. Teda platí vzťah $r_t = y_t - \hat{y}_t$ Potom najčastejšie požívanými mierami hodnotenia presnosti prognóz sú:

- priemerná štvorcová chyba prognózy (Mean Squared Error)

$$MSE = \frac{1}{n}\sum_{t=1}^{n} r_t^2$$

- priemerná absolútna chyba prognózy (Mean Absolute Error)

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |r_t|$$

- priemerná absolútna relatívna chyba prognózy (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|r_t|}{y_t}$$

- relatívna chyba prognózy nazývaná Theilov koeficient

$$TC = \frac{\sum_{t=1}^{n} r_t^2}{\sum_{t=1}^{n} y_t^2}.$$

Bodová a intervalová prognóza

Všetky spomínané metódy prognózovania priamo poskytujú ako prognózu jednu hodnotu – teda bodový odhad. Samozrejme predpokladať, že práve tento konkrétny bod bude presnou realizáciou budúcnosti by bolo príliš trúfalé. Preto sa ako doplnok bodovej prognózy konštruuje sprievodný interval – intervalová prognóza.

Intervalová prognóza teda spočíva v konštrukcii intervalu, ktorého koncové body sú štatistiky s vopred zvolenou pravdepodobnosťou pokrytia skutočnej hodnoty. Zapisujemeí

$$P\left(q_1 \le Q \le q_2\right) = 1 - \alpha .$$

Krajné hodnoty tohto intervalu q_1 a q_2 sú štatistiky s pravdepodobnosťou P rovnou koeficientu spoľahlivosti (1–α), že skutočná hodnota y sa bude nachádzať v intervale. Interval sa nazýva *100*(1–α) %-ný interval spoľahlivosti*. Obvykle sa vyberá na úrovni 95 %. Číslo α (alfa) sa nazýva *riziko odhadu* (hladina významnosti). Zvyšovanie spoľahlivosti, a tým pádom znižovanie rizika odhadu rozširuje veľkosť intervalu spoľahlivosti.

Pre výber typu intervalu spoľahlivosti existujú tri možnosti:

1. interval je *obojstranný*, teda obmedzený zhora aj zdola a bodový odhad ho rozdeľuje na dve rovnako veľké časti, zodpovedajúce riziku $\alpha/2$;
2. interval je *pravostranný*, teda obmedzený len zhora a vzdialenosť hornej hranice od bodového odhadu zodpovedá celému riziku α;

3. interval je *ľavostranný*, teda obmedzený len zdola a vzdialenosť dolnej hranice od bodového odhadu zodpovedá celému riziku α.

Základom pre konštrukciu intervalu spoľahlivosti je vždy bodový odhad. Ak zvolíme hladinu významnosti α, potom v tabuľkách Studentovho rozdelenia nájdeme $t_{\alpha/2}$ také, že platí

$$P\left\{-t_{\alpha/2} \leq \frac{y-\hat{y}}{s_p} \leq t_{\alpha/2}\right\} = 1-\alpha\ ,$$

kde s_p je štandardná odchýlka prognózy.

Malou úpravou odvodíme vzťah pre *obojstranný interval spoľahlivosti prognózy* (za predpokladu $n > 30$, aby sa dalo použiť normálne rozdelenie), teda

$$P\left(\hat{y} - t_{1-\alpha/2} s_p \leq y \leq \hat{y} + t_{1-\alpha/2} s_p\right) = 1-\alpha\ ,$$

ktorý s pravdepodobnosťou (1–α) bude obsahovať skutočnú budúcu hodnotu y.

Jednotlivé zložky časového radu

Na pozorovaný ukazovateľ, ktorého hodnoty sú usporiadané v časovom rade, vplýva veľké množstvo rôznych činiteľov s rôznou intenzitou. Napriek odlišnej povahe každého činiteľa ich môžeme zhrnúť do troch základných skupín, ktoré predstavujú základné zložky časového radu.

1. *vývojové činitele*, ktoré pôsobia na ukazovateľ stále a dlhodobo, čím určujú hlavný smer jeho vývoja a tým určujú *trend* (T_t) časového radu
2. *periodické činitele*, ktoré pôsobia občas, ale pravidelne sa opakujú, pričom môžu striedavo pôsobiť v smere trendu alebo proti nemu

   Podľa toho ako často sa opakuje ich výskyt, rozoznávame dve zložky časového radu:
   - *sezónnu zložku* (S_t), ktorej periodicita je menšia ako rok, alebo práve rok
   - *cyklickú zložku* (C_t), ktorej periodicita je väčšia ako rok (hospodársky cyklus)
3. *náhodné činitele*, ktoré pôsobia úplne nepravidelne, rôznym smerom a bez možnosti ich predvídania, a preto tvoria *náhodnú zložku* (ε_t)

Časový rad môže obsahovať tieto zložky buď všetky spolu, prípadne nemusí obsahovať jednu alebo obidve periodické zložky. Ak predpokladáme, že medzi nimi existuje aditívny vzťah, teda ak každú hodnotu časového radu Y_t tvorí súčet týchto zložiek, tak platí

$$y_t = T_t + S_t + C_t + \varepsilon_t.$$

Ak predpokladáme, že medzi nimi existuje multiplikatívny vzťah, tak platí

$$y_t = T_t \, . \, S_t \, . \, C_t \, . \, \varepsilon_t.$$

V praktických modeloch sa častejšie využíva aditívny model, pretože logaritmickou transformáciou sa dá multiplikatívny vzťah transformovať na aditívny. Ale pri vyhľadávaní najlepších modelov pre prognózu by sa nemali obísť ani multiplikatívne modely.

Analytický odhad funkcie trendu

Táto metóda je založená na existencii závislosti hodnôt časového radu (závislá premenná) od časového faktora (nezávislá premenná). Potom je možné na vyrovnanie časového radu použiť matematickú funkciu, ktorá najlepšie popisuje jeho priebeh. Výber vhodnej funkcie závisí od grafickej analýzy a rozboru diferencií. K dispozícii sú:

1. *Lineárna funkcia* sa pre opis časového radu využíva vtedy, ak je trend lineárny. Teda priebeh hodnôt osciluje okolo priamky. Takže sa predpokladá nasledovná závislosť hodnôt sledovaného ukazovateľa

   $$y_t = \beta_0 + \beta_1 t \, ,$$

   kde t predstavuje časovú nezávislú premennú (t = 1, 2, ..., n).

   Minimalizáciou sumy štvorcov odchýlok odhadov od skutočných pozorovaní a následnými úpravami dostaneme sústavu normálových rovníc. Riešením tejto sústavy dostaneme vzorce pre odhad parametrov lineárnej funkcie. Podobný postup sa použije aj v prípade odhadu parametrov iných typov funkcií. Niekedy však tento postup vyžaduje použitie vhodného typu transformácie – obvykle zlogaritmovanie.

2. *Kvadratická funkcia* sa pre opis časového radu využíva vtedy, ak je trend charakterizovaný parabolou. Takže sa predpokladá nasledovná závislosť hodnôt sledovaného ukazovateľa

   $$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 \, .$$

3. *Exponenciálna funkcia* sa pre opis časového radu využíva vtedy, ak je trend charakterizovaný exponenciálnou krivkou. Takže sa predpokladá nasledovná závislosť

   $$y_t = \beta_0 \beta_1^t \, .$$

   Rovnica sa najprv zlogaritmuje a následne sa pokračuje ako pri lineárnom trende.

4. *S krivka* je modifikáciou exponenciálneho trendu, ak hodnoty radu majú po logaritmickej transformácii hyperbolický priebeh v čase. Takže sa predpokladá závislosť

   $$y_t = e^{\left(\beta_0 + \beta_1 \frac{1}{t}\right)} \, .$$

5. *Logistický trend* patrí do triedy symetrických S kriviek. Pre opis časového radu využíva vtedy, ak je trend charakterizovaný tromi rôznymi úsekmi. Prvý je charakterizovaný pomalým vzostupom, nasleduje prudký rast a poslednú časť charakterizuje vrcholová stagnácia. Takže sa predpokladá závislosť v nasledovnom tvare
$$y_t = \frac{1}{c + \beta_0 \beta_1^t}.$$
6. *Gompertzova krivka* patrí medzi S krivky, ale je asymetrická. Predpokladá závislosť
$$y_t = c\beta_0^{\beta_1^t}.$$

Metódy adaptívnych modelov

Trendové funkcie sa najčastejšie používajú na popis takých premenných, ktorých vývoj zodpovedá uvedeným funkciám. Ale mnohé premenné v skutočnosti nemajú taký ideálny priebeh, pretože v určitom časovom období sa mení ich vecný charakter vplyvom rôznych činiteľov. Tieto zmeny sa prejavia ako zmeny v sledovaných premenných v určitých úsekoch časového radu. Preto vyrovnať rady pomocou trendových funkcií by znamenalo nerešpektovať vzniknuté zmeny a vypočítať nepresné prognózy. Uvedené problémy je preto nevyhnutné riešiť pomocou adaptívnych modelov, ktorých parametre sa menia v čase.

Najjednoduchšie adaptívne vyrovnávanie je exponenciálne vyrovnávanie. Zakladá sa na myšlienke, že pri konštrukcii extrapolačnej prognózy sa najvyššia váha pridelí poslednej hodnote radu, pretože táto obsahuje najviac informácie o možných zmenách vo vývoji sledovanej premennej a očakáva sa aj ich čiastočný prenos do budúcnosti. Pre ostatné hodnoty platí, že ich váha exponenciálne klesá s ich zväčšujúcou vzdialenosťou do minulosti.

Exponenciálny priemer s_t v čase t možno vypočítať ako súčet exponenciálneho priemeru v čase t-1 a α násobku rozdielu pozorovania v čase t a exponenciálneho priemeru v čase t-1. α sa nazýva *vyrovnávajúcou konštantou* (váhy posledného pozorovania) a je z intervalu $\langle 0,1 \rangle$. Teda
$$s_t = \alpha y_t + (1-\alpha) s_{t-1} = s_{t-1} + \alpha (y_t + s_{t-1})$$
Keďže ide o rekurentný vzťah, po vzájomnom dosadzovaní a zanedbaní člena $(1-\alpha)^n s_0$, kvôli tomu, že $(1-\alpha)^n \to 0$, sa dá exponenciálny priemer vyjadriť ako vážený súčet všetkých členov radu s exponenciálne klesajúcimi váhami
$$s_t = \alpha \sum_{i=0}^{n-1} (1-\alpha)^i \, y_{t-i}.$$

Existuje a navrhnutých bolo viacero metód exponenciálneho vyrovnávania, z ktorých sú najčastejšie používané Brownove modely, Holtove modely a Wintersove modely.

Brownov model jednoduchého exponenciálneho vyrovnávania

Jednoduché exponenciálne vyrovnávanie sa využíva vtedy, ak v určitých rovnakých časových úsekoch je trend časového radu konštantný, ale pritom sa pomaly meniaci v čase. Predpokladá sa model časového radu

$$y_t = T_t + \varepsilon_t,$$

kde t = 1, 2, ..., n a v ktorom trend $T_t = A_{0t}$ = konštanta.

Výsledkom odhadu parametra A_{0t} váženou metódou najmenších štvorcov je, že tento parameter sa rovná exponenciálnemu priemeru s_t v čase t. Potom platí, že nová prognóza v čase t sa určí ako súčet prognózy so začiatkom v čase t-1 a α násobku chyby prognózy, teda

$$\hat{y}_t = s_t = s_{t-1} + \alpha\left(y_t + s_{t-1}\right) = \hat{y}_{t-1} + \alpha\varepsilon_t.$$

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. V takých prípadoch sa vyrovnávajúca konštanta α volí blízka 1 – teda prognózu určuje najmä posledná aktuálna hodnota radu.

Brownov lineárny model dvojitého exponenciálneho vyrovnávania

Tento model sa využíva vtedy, ak časový rad možno na rovnako dlhých časových úsekoch vyrovnať lineárnou funkciou. Predpokladá sa model časového radu

$$y_t = A_{0t} + A_{1t}t + \varepsilon_t,$$

kde t = 1, 2, ..., n a v ktorom je lineárny trend $T_t = A_{0t} + A_{1t}t$.

Výsledkom odhadu parametrov A_{0t} a A_{1t} váženou metódou najmenších štvorcov je, že pre exponenciálny priemer s_t a exponenciálny priemer druhého rádu $s_t^{(2)}$ (čo znamená použitie exponenciálneho vyrovnávania na exponenciálne priemery) platia vzťahy

$$s_t = \alpha y_t + \left(1-\alpha\right)s_{t-1}$$
$$s_t^{(2)} = \alpha s_t + \left(1-\alpha\right)s_{t-1}^{(2)}$$

a prognóza má tvar

$$\hat{y}_{t+i} = \left(2 + \frac{\alpha i}{1-\alpha}\right)s_t - \left(1 + \frac{\alpha i}{1-\alpha}\right)s_t^{(2)}.$$

Ak je najlepšia hodnota α (MAE...) blízka 0, vhodnejšie je použiť jednoduché diferencovanie. Existuje možnosť použiť aj trojité vyrovnávanie, potom ide o Brownov kvadratický model.

Holtov lineárny model exponenciálneho vyrovnávania

Tento model bol vypracovaný preto, lebo v prípade výrazného lineárneho trendu dáva exponenciálne vyrovnávanie skreslené prognózy. Holtov model síce využíva exponenciálne vyrovnávanie, ale používa dve vyrovnávajúce konštanty α_1 a α_2 z intervalu $\langle 0,1 \rangle$. Konštanta α_1 vyrovnáva zmeny v úrovni jednotlivých úsekov časového radu a konštanta α_2 vyrovnáva smernice lineárnych trendov v jednotlivých úsekoch časového radu.

Predpokladá sa lineárny trend v tvare $T_t = A_{0t} + A_{1t}t$. Výsledkom odhadu parametrov A_{0t} a A_{1t} váženou metódou najmenších štvorcov a po zapracovaní jednotlivých vzťahov je, že pre jednotlivé parametre platí

$$A_{0t} = \alpha_1 y_t + (1-\alpha_1)(A_{0t-1} - A_{1t-1})$$
$$A_{1t} = \alpha_2 (A_{0t} - A_{0t-1}) + (1-\alpha_2) A_{1t-1}$$

a prognóza má tvar

$$\hat{y}_{t+i} = A_{0t} + A_{1t}i$$

Ak sú najlepšie hodnoty α_1 a α_2 (z hľadiska porovnávacích mier MAE a iných) blízke 0, tak to znamená, že vyrovnávaný trend je konštantný a vhodnejšie je použitie ARIMA modelov.

Skúmanie cyklických zložiek vo vývoji časového radu

Okrem základných činiteľov, ktoré určujú trend časového radu, je jeho vývoj zvyčajne ovplyvňovaný periodickými činiteľmi, ktoré vyvolávajú striedavo rast a pokles jeho hodnôt. Ak jedna perióda zahŕňa viac rokov, hovoríme o cyklickej zložke, ak sa prejaví do dĺžky maximálne jedného roka, tak ide o sezónnu zložku. Sezónne vplyvy sú vyvolané hlavne prírodnými činiteľmi a popisujeme ich pomocou sezónnych indexov.

Sezónne *indexy* sú pomerné čísla, v ktorých hodnoty časového radu porovnávame voči „očisteným“ hodnotám, ktoré zahŕňajú len trend (hodnoty vypočítané metódou najmenších štvorcov). Teda platí vzťah

$$S_t = \frac{y_t}{y_t^v}, \qquad \text{kde } y_t^v \text{ sú očistené hodnoty v čase } t.$$

Keďže v skúmanej perióde sa môžu prejaviť aj iné ako sezónne vplyvy, je potrebné rátať sezónne indexy za viac periód a nakoniec vypočítať priemerné hodnoty sezónnych indexov.

Cyklickú zložku potom určíme jednoduchým odpočítaním trendovej a sezónnej zložky od pôvodnej hodnoty časového radu, pričom abstrahujeme od náhodnej zložky. Takže platí

$$C_t = Y_t - (T_t + S_t).$$

Cyklické vplyvy sa prejavujú predovšetkým na makroekonomických ukazovateľoch a signalizujú, v akom stave sa nachádza ekonomika.

Keďže existujú modely exponenciálneho vyrovnávania zohľadňujúce aj sezónne výkyvy, uvedieme dva najpoužívanejšie Wintersove modely.

Wintersov model sezónneho exponenciálneho vyrovnávania – aditívny typ

Wintersov model vyrovnávania je metóda exponenciálneho vyrovnávania využívaná pri sezónnych časových radoch. Využíva tri vyrovnávajúce konštanty α, β, γ. Konštanta α sa používa na odhad úrovne radu, konštanta β sa využíva na odhad koeficienta lineárneho trendu a konštanta γ sa využíva na odhad sezónnych indexov. Predpokladá sa model časového radu

$$y_t = A_{0t} + A_{1t} i + S_p(t) + \varepsilon_{t-i},$$

Výsledkom odhadu parametrov sú vyrovnávajúce vzťahy

$$\begin{aligned} L_t &= \alpha\left(y_t - S_{t-p}\right) + (1-\alpha)\left(L_{t-1} + T_{t-1}\right) \\ T_t &= \beta\left(L_t - L_{t-1}\right) + (1-\beta) T_{t-1} \\ S_t &= \gamma\left(y_t - L_t\right) + (1-\gamma) S_{t-p} \end{aligned}$$

a prognóza má tvar

$$\hat{y}_{t+i} = L_t + iT_t + S_{t-p+i}$$

Wintersov model sezónneho exponenciálneho vyrovnávania – multiplikatívny typ

Pri použití tohto modelu musia byť časové rady a všetky prognózy striktne kladné. Predpokladá sa model časového radu v tvare

$$y_t = \left(A_{0t} + A_{1t} i\right) S_p(t) + \varepsilon_{t-i},$$

Výsledkom odhadu parametrov sú vyrovnávajúce vzťahy

$$\begin{aligned} L_t &= \alpha\left(y_t / S_{t-p}\right) + (1-\alpha)\left(L_{t-1} + T_{t-1}\right) \\ T_t &= \beta\left(L_t - L_{t-1}\right) + (1-\beta) T_{t-1} \\ S_t &= \gamma\left(y_t / L_t\right) + (1-\gamma) S_{t-p} \end{aligned}$$

a prognóza má tvar

$$\hat{y}_{t+i} = \left(L_t + iT_t\right) S_{t-p+i}$$

Pri oboch Wintersovych modeloch, ak sú sezónne indexy veľmi blízke 1, tak by mali byť použité modely bez sezónnosti. A ak sú sezónne indexy veľmi blízke 0, tak to znamená prítomnosť deterministického sezónneho faktora.

Okrem metód dekompozície sa pre základný rámec porovnávaných modelov využíva Boxova-Jenkinsova metodológia modelov typu SARIMA.

Boxova-Jenkinsova metodológia

Základným kameňom pre budovanie teórie časových radov je proces *bieleho šumu*. Biely šum je uvažovaný ako množina identicky a nezávisle normálne rozdelených hodnôt.

To implikuje nasledovné základné predpoklady:

1. stredná hodnota bieleho šumu ε_t je pre všetky *t* nulová, $E\left(\varepsilon_t\right) = E\left(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right) = 0$,
2. kovariancia dvoch rôznych členov ε_t a ε_{t-j} je nulová, $E\left(\varepsilon_t, \varepsilon_{t-j}\right) = \operatorname{cov}\left(\varepsilon_t, \varepsilon_{t-j}\right) = 0$,
3. rozptyl bieleho šumu ε_t je pre všetky *t* rovnaký, $\operatorname{var}\left(\varepsilon_t\right) = \operatorname{var}\left(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right) = \sigma_\varepsilon^2$.

Veľká trieda modelov je vytvorená práve lineárnou kombináciou bieleho šumu. Ako príklad môžeme uviesť nasledovné typy, ktorých identifikáciu rozpracovali Box a Jenkins

- autoregresný model 1. rádu, AR(1): $y_t = \phi y_{t-1} + \varepsilon_t$,
- model kĺzavých priemerov 1. rádu, MA(1): $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$,
- autoregresný model rádu p, AR(p): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$,
- model kĺzavých priemerov rádu q, MA(q): $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$,
- všeobecný model ARMA(*p*, *q*): $y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$.

Pre ľahšiu manipuláciu s radmi obsahujúcimi v čase posunuté hodnoty je výhodné zaviesť *operátor* oneskorenia *L*, potom platí $Ly_t = y_{t-1}$, $Ly_{t-1} = y_{t-2} \;\rightarrow\; L^2 y_t = y_{t-2}$. Pomocou polynómu operátora oneskorenia, napríklad $a\left(L\right) y_t = \left(a_0 L^0 + a_1 L^1 + a_2 L^2\right) y_t = a_0 y_t + a_1 y_{t-1} + a_2 y_{t-2}$, sa následne jednoducho dajú vyjadriť modely ARMA:

- *AR* model: $a\left(L\right) y_t = \varepsilon_t$,
- *MA* model: $y_t = b\left(L\right) \varepsilon_t$,
- model *ARMA*: $a\left(L\right) y_t = b\left(L\right) \varepsilon_t$.

Identifikácia konkrétneho modelu sa zakladá na skúmaní priebehu odhadnutej *autokorelačnej funkcie* (ACF) a odhadnutej *parciálnej autokorelačnej funkcie* (PACF). Pri výbere modelu je dôležité zistiť, od ktorého bodu sú hodnoty autokorelačnej a parciálnej autokorelačnej funkcie štatisticky nevýznamné. Pri ACF pomáha zisťovať tento *identifikačný bod* Bartlettova aproximácia a pri PACF Quenouillova aproximácia.

Pri analýze však netreba zabúdať, že sa pracuje iba s odhadmi ACF a PACF a preto závery môžu byť skreslené. Preto sa odporúča preskúšať niekoľko alternatív. Obvykle sa preskúmavajú všetky kombinácie, kde *p* = 0, 1, 2 a q = 0, 1, 2.

Dôležitým pojmom pri modelovaní časových radov je ich *stacionarita*. Stacionarita je definovaná *v striktnom zmysle*, ak požadujeme nemennosť distribučnej funkcie rozdelenia skúmanej náhodnej premennej v čase. Pre praktický výskum postačuje skúmanie stacionarity *v slabom zmysle* ako nemennosť strednej hodnoty, rozptylu a kovariancie.

Pri všetkých doteraz spomínaných ARMA modeloch predpokladáme ich stacionaritu. Riešenie problému nestacionarity sa realizuje pomocou *diferencovania*. A podľa toho koľko násobnú diferenciu je nevyhnutné použiť, aby sa rad stacionarizoval, sa určí *rád integrácie* časového radu. Prvé diferencie sú definované vzťahom $\Delta y_t = y_t - y_{t-1}$, druhé diferencie vzťahom $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$, atď.

Preto sa pri nestacionarite namiesto ARMA modelov používajú ARIMA modely charakterizované okrem parametrov (*p*, *q*) aj parametrom *d* – určujúcim rád integrácie. Potom celý postup je analogický, ale nepracuje sa s pôvodnými úrovňami, ale so zodpovedajúcimi diferenciami.

Pri diferencovaní nemusíme využiť, iba rozdiel susedných členov, ale ak pracujeme zo sezónnymi údajmi, aj rozdiel medzi rovnakými sezónami. Teda pre štvrťročné údaje je sezónna diferncia prvého rádu definovaná vzťahom $\Delta_4 y_t = y_t - y_{t-4}$ a pre mesačné údaje vzťahom $\Delta_{12} y_t = y_t - y_{t-12}$. V takom prípade hovoríme o modeloch typu SARIMA.

Pre odhad parametrov Box-Jenkinsových modelov sa používajú iteračné metódy ako je napríklad Marquardtova nelineárna metóda najmenších štvorcov, ktoré podporuje väčšina štandardných štatistických alebo ekonometrických programových balíkov.

Odhadnutý tvar modelu je priamo použiteľný pre výpočet prognózy. Prognóza sa konštruuje rekurentne, to znamená, že najprv sa vypočíta prognóza o jedno obdobie dopredu, pomocou nej sa vypočíta prognóza o dve obdobia dopredu, atď.

Záver

Všetky uvedené metódy sú v súčasnosti podporované štandardným štatistickým softvérom ako napr. STATGRAPHICS alebo ekonometrickým softvérom ako napr. EVIEWS. Výber preskúmavaného typu modelu respektíve rozhodnutie o úrovni vplývajúcich konštánt modelov, rovnako ako rozhodnutie o výbere modelu pre prognózu závisí od tvorcu prognózy. Programové vybavenie ako výborná podpora tvorby prognóz poskytuje odhadnuté hodnoty parametrov modelov, na ich základe vypočítava bodové a intervalové prognózy a vypočítava sprievodné štatistiky uľahčujúce rozhodnutia.

Zoznam literatúry

[1] BOX, G. E. P. – JENKINS, G. M. – REINSEL, G. C.: *Time Series Analysis. Forecasting and Control*. Prentice – Hall 1994.

[2] GOURIEROUX, CH. – MONFORT, A.: *Time Series and Dynamic Models*. Cambridge University Press 1997.

[3] RUBLÍKOVÁ, E. – HATRÁK, M.: *Prognostická štatistika*. ES VŠE 1985.

[4] HAMILTON, J. D.: *Time Series Analysis*, Princeton: Princeton University Press 1994.

[5] MAKRIDAKIS, S. G. – WHEELWRIGHT, S. C. – HYNDMAN, R. J.: *Forecasting: Methods and Applications*, 3rd ed. John Wiley and Sons 1998.

[6] MILLS, T. C.: *Time Series Techniques for Economists*, Cambridge University Press, 1990.