

APLIKÁCIE ČASOVÝCH RADOV

MODELY S PREMENLIVÝMI REŽIMAMI

Hodnotenie programu na cvičení

- 1 bod** funkčný program, obsahujúci všetky predpísané body; študent ho vie adaptovať aj na iné dáta
- 0,5 bodu** program, v ktorom chýbajú niektoré predpísané body (menej ako polovica), ale študent ho vie adaptovať aj na iné dáta
- 0 bodov** program, ktorý je odpísaný od kolegov (študent mu nerozumie a nevie ho adaptovať na iné dáta)
- 0,5 bodu** program, v ktorom chýba viac ako polovica predpísaných bodov
- 1 bod** študent nemá žiadny program

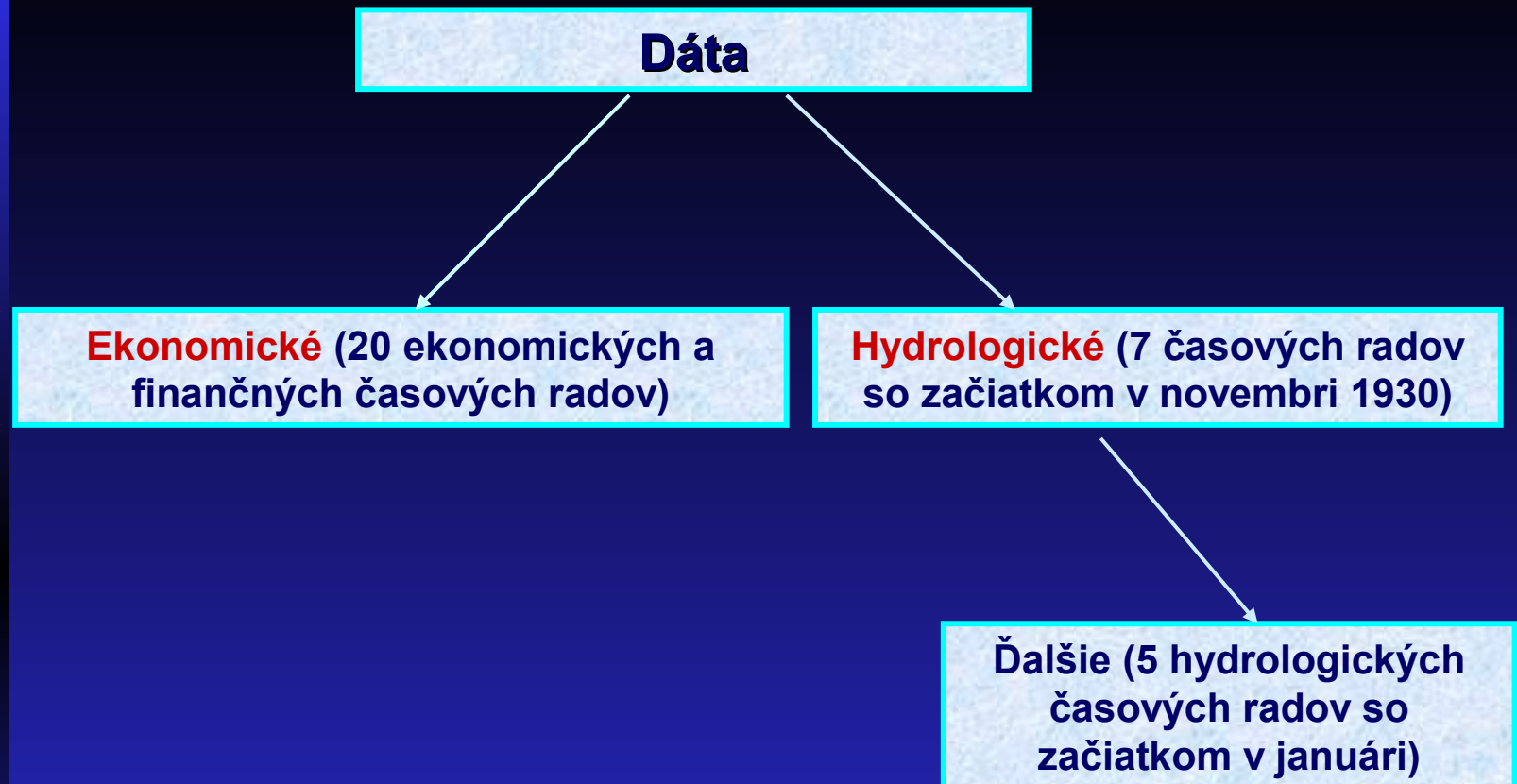
Zápočet a skúška

V polovici semestra bude písomka z teórie (maximálny počet bodov je 20). Tieto body spolu z bodmi z projektov (maximálne 10 bodov) sa pripočítajú k bodom na skúške. Na zápočet je potrebných minimálne 60% (t. j. 18 bodov).

Skúška má dve časti: teoretickú (písomka - maximálne 35 bodov) a praktickú (projekt - maximálne 35 bodov).

Projekt

Pre zvolené dáta (iné, aké ste spracovali na cvičení) odhadnúť parametre vhodných modelov zo všetkých troch preberaných tried (SETAR, LSTAR, ESTAR). Určiť najlepší model z hľadiska popisného a z hľadiska predikcií na 5 rokov.



Ekonomické dáta pochádzajú zo stránok Národnej Banky Slovensko (www.nbs.sk) a Európskej centrálnej banky (www.ecb.int), ktorá preberá niektoré údaje priamo z Eurostatu (www.ec.europa.eu/eurostat)

Popis ekonomických dát

NEZAMESTNANOSŤ SR– mesačné dáta evidovanej miery nezamestnanosti (január 1999 – marec 2009) – označíme SVKNEZAM

NEZAMESTNANOSŤ OČISTENÁ SR – mesačné dáta evidovanej miery nezamestnanosti SR, sezónne očistené (január 1998 – február 2008) – označíme UNEMPLOY očist

INFLÁCIA SR– medziročné dáta na mesačnej báze (január 1999 – marec 2009) - SVKINFLACIA

VÝMENNÝ KURZ SKK/EUR – priemerný výmenný kurz NBS na mesačnej báze (roky 1999 - 2008)

VÝMENNÝ KURZ HUF/EUR - priemerný výmenný kurz NBS na mesačnej báze (roky 1999 - 2008)

VÝMENNÝ KURZ PLN/EUR - priemerný výmenný kurz NBS na mesačnej báze (roky 1999 - 2008)

VÝMENNÝ KURZ CZK/EUR - priemerný výmenný kurz NBS na mesačnej báze (roky 1999 - 2008)

ZAMESTNANOSŤ SR – štvrťročné hodnoty celkovej priemernej domácej zamestnanosti SR (v mil osôb, 1Q 1994 – 4Q 2008)

DEFICIT (ALEBO ZISK) USA – štvrťročné percentuálne hodnoty deficitu (resp. zisku) vo všetkých sektoroch USA, nie sú sezónne očistené (1Q 1960 – 3Q 2008) - označíme DoS USA

Popis ekonomických dát - pokračovanie

HDP MAĎARSKO - štvrťročné údaje hrubého domáceho produktu Maďarska v trhových cenách (v mld HUF), sezónne očistené (1Q 1995 – 4Q 2008)

HDP SR – štvrťročné údaje hrubého domáceho produktu SR v trhových cenách (v mld SKK), sezónne očistené (1Q 1993 – 4Q 2008)

HRUBÁ PRIDANÁ HODNOTA POĽNOHOSPODÁRSTVA SR (gross value-added) – štvrťročné dáta hrubej pridanej hodnoty poľnohospodárstva, rybolovu, lesníctva a lovu, v základných cenách (v mld €, 1Q 1993 - 4Q 2008) – označíme GVA agri

HRUBÁ PRIDANÁ HODNOTA STAVEBNÍCTVA SR – štvrťročné dáta hrubej pridanej hodnoty v stavebníctve, v základných cenách (v mld €, 1Q 1993 - 4Q 2008) - označíme GVA constr

HRUBÁ PRIDANÁ HODNOTA FINANČNÉHO SEKTORU SR – štvrťročné dáta hrubej pridanej hodnoty vo finančnom sektore, v základných cenách (v mld €, 1Q 1993 - 4Q 2008) -označíme GVA fin

HRUBÁ PRIDANÁ HODNOTA PRIEMYSLU SR – štvrťročné dáta hrubej pridanej hodnoty celkového priemyslu, zahrňujúce energie, v základných cenách (v mld €, 1Q 1993 - 4Q 2008) -označíme GVA industry

HRUBÁ PRIDANÁ HODNOTA OSTATNÉ SR – štvrťročné dáta hrubej pridanej hodnoty pre ostatné služby, v základných cenách (v mld €, 1Q 1993 - 4Q 2008) -označíme GVA other

Popis ekonomických dát - pokračovanie

INVESTIČNÉ MAJETKY EU16 (CAPITAL GOODS) – mesačné hodnoty obchodu EÚ extra 16 (v mld €), sezónne očistené (január 1999 – december 2008)

POČET BANKOVIEK V OBEHU V HODNOTE 100 SK – mesačné počty čistého obehu bankoviek nominálnej hodnoty 100 SK, v miliardách ks (január 2005 – december 2007) - označíme NofB100

POČET MINCÍ V OBEHU V HODNOTE 10 SK – mesačné počty čistého obehu mincí nominálnej hodnoty 10 SK, v miliardách ks (január 2005 – december 2007) - označíme NofB10

INDEX SPOTREBITEĽSKÝCH CIEN PRE DOPRAVU SR (ICP) – mesačné hodnoty indexu spotrebiteľských cien pre dopravu SR, nie sú sezónne očistené (január 1995 – február 2009) – označíme TRANSPORT

LITERATÚRA

Arlt, J. – Arltová, M. (2003) *Finanční časové řady – Vlastnosti, metody modelování, příklady a aplikace*, GRADA Publ.

Franses, P. H., Van Dijk, D. (2000) *Non – linear time series models in empirical finance*. Cambridge Univ. Press

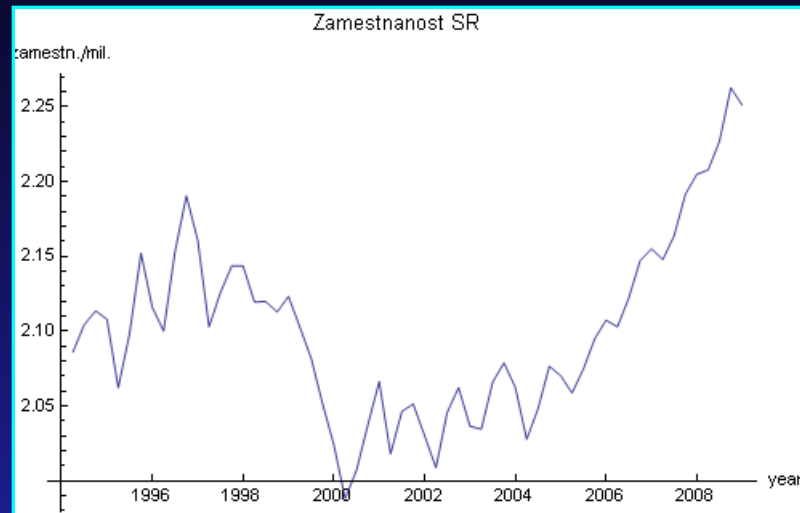
Hamilton, J.D. (1994): *Time series analysis*. Princeton University Press, 1994.

Hamilton, J.D. (1996): *Specification testing in Markov-switching time series models*. Journal of Econometrics, 70 (1996), 127-157.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988), Testing linearity in univariate time series models. *Scandinavian Journal of Statistics* 16, 161 - 175.

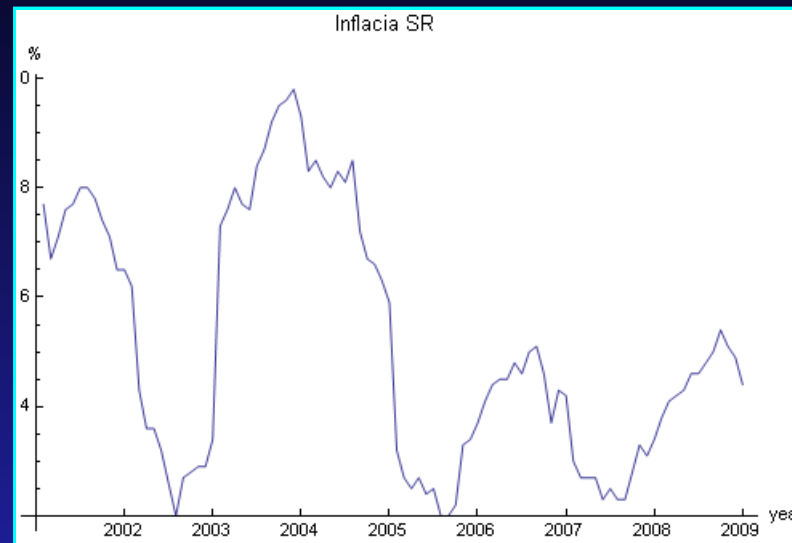
Teräsvirta, T. (1994): Specification, estimation, and evaluation of smooth transition models. *Journal of American Statistical Association* 89, 208 – 218.

Ekonomické časové rady



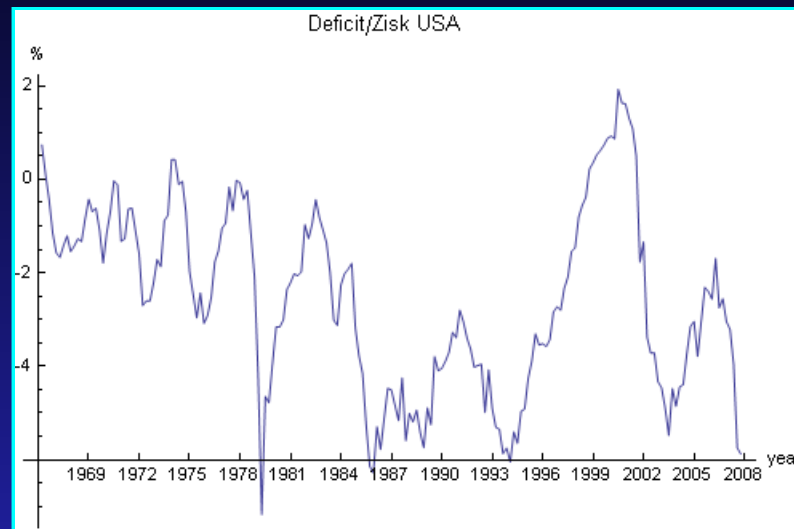
ZAMESTNANOSŤ SR – štvrťročné hodnoty celkovej priemernej domácej zamestnanosti SR (v mil osôb, 1Q 1994 – 4Q 2008)

Ekonomické časové rady



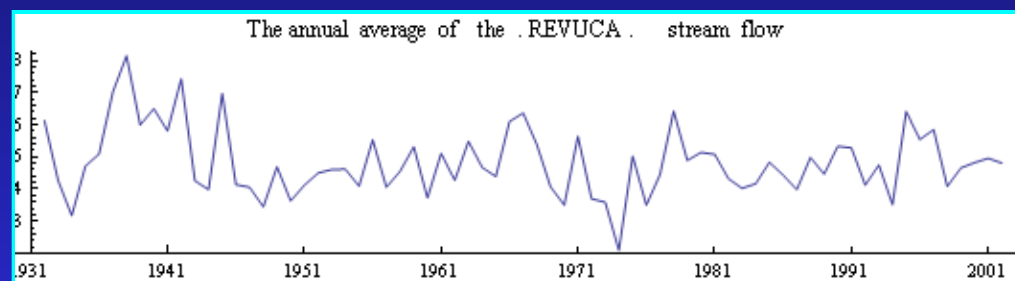
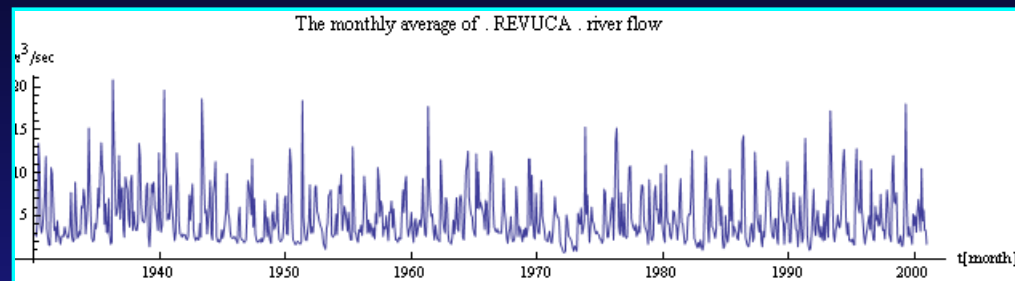
INFLÁCIA SR– medziročné dáta na mesačnej báze (január 2001 – marec 2009)

Ekonomické časové rady

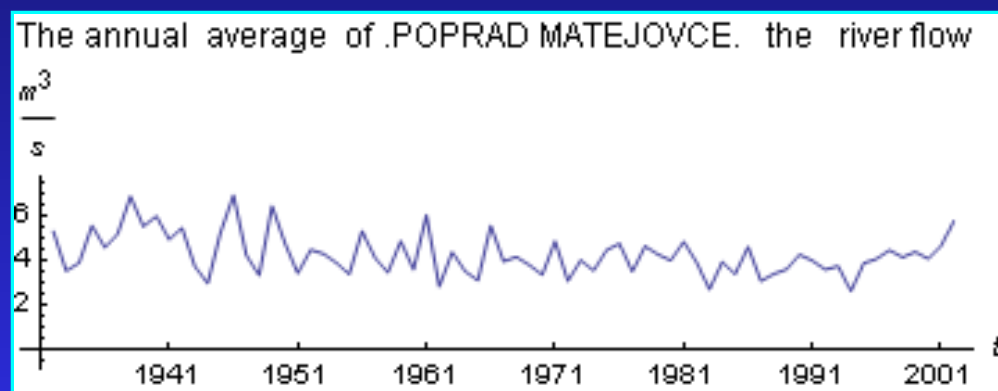
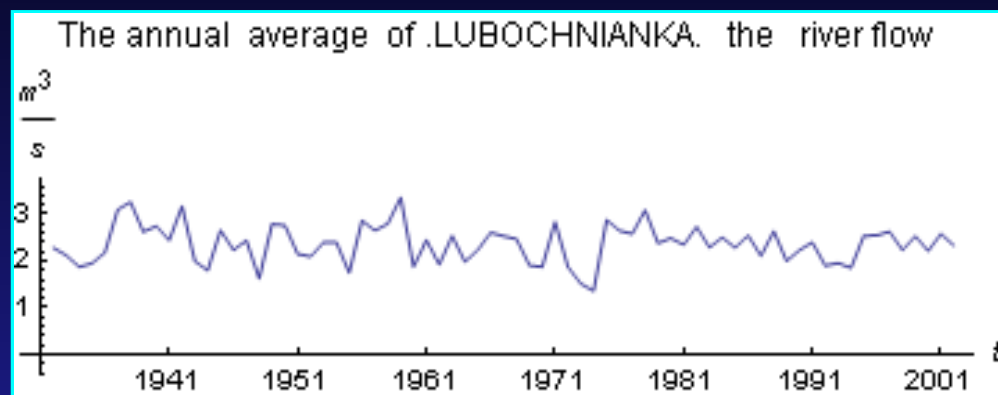


DEFICIT (ALEBO ZISK) USA – štvrťročné percentuálne hodnoty deficitu (resp. zisku) vo všetkých sektoroch USA, nie sú sezónne očistené (1Q 1960 – 3Q 2008)

Hydrologické časové rady



Hydrologické časové rady



Modely s premenlivým režimom – viacrežimové modely

1. Modely s režimami určenými pozorovateľnými premennými

V týchto modeloch sa predpokladá, že režimy, ktoré sa objavili v minulosti a súčasnosti sú známe, resp. je ich možné identifikovať štatistickými technikami.

premennou s_t , nadobudajúcou len hodnoty 1, 2, ..., m, ktorú reprezentuje m-stavový Markovov proces prvého rádu s prechodovou maticou P. Jednotlivé režimy teda nie je možné identifikovať presne, ale len s určitou pravdepodobnosťou.

ε_t – proces bieleno sumu, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma_\varepsilon^2$

Skeleton modelu je deterministická časť modelu. **Ekvilibrium** modelu je v bode X^* , ak X^* je pevný bod skeletonu, t.j. bod, pre ktorý platí $F(X^*, \theta) = X^*$. **Ekvilibrium** sa nazýva **stabilné**, ak časový rad konverguje k X^* . Časový rad tvorený nelineárnym procesom môže mať jedno aj viac stabilných, resp. nestabilných ekviliórií, ale nemusí mať žiadne ekvilibrium.

Modely TAR (Threshold Autoregressive)

V modeloch TAR sa vhodný režim v čase t určuje podľa vzťahu pozorovateľnej indikačnej premennej q_t k prahovej (*threshold*) hodnote, označovanej ako c . Indikačná premenná q_t môže byť hodnota iného časového radu, modelujúceho ukazovateľ, ovplyvňujúci hodnoty sledovaného časového radu, resp. hodnota sledovaného časového radu posunutá v čase o d časových intervalov, teda $q_t = x_{t-d}$ pre určité prirodzené číslo d .

Napr. 2-režimový TAR model s režimami $AR(p_1)$ a $AR(p_2)$ má tvar:

$$X_t = Y'_{t,1} \Phi_1 I(q_t \leq c) + Y'_{t,2} \Phi_2 I(q_t > c) + \varepsilon_t$$

kde $Y_{t,i} = (1, X_{t-1}, \dots, X_{t-p_i})'$ a $\Phi_i = (\phi_{i,0}, \phi_{i,1}, \dots, \phi_{i,p_i})'$, $i = 1, 2$

$I[A]$ je indikačná funkcia s hodnotami $I[A] = 1$, ak udalosť A nastane a $I[A] = 0$, ak nenastane

$\{\varepsilon_t\}$ je proces bieleho šumu s nulovou strednou hodnotou a rozptylom σ_ε^2 .

Nakreslite 2-režimový TAR proces (dĺžky $n = 500$) s AR(1) v obidvoch režimoch pre rôzne hodnoty autoregresných parametrov a parametra c (pre $q_t = X_{t-1}$, s počiatočnou podmienkou $X_0 = 2$). Vypočítajte počet dát v jednotlivých režimoch.

Model má tvar:

$$X_t = \begin{cases} \varphi_{1,0} + \varphi_{1,1}X_{t-1} + \varepsilon_t & \text{ak } X_{t-1} \leq c \\ \varphi_{2,0} + \varphi_{2,1}X_{t-1} + \varepsilon_t & \text{ak } X_{t-1} > c \end{cases}$$

resp. pomocou indikačnej funkcie

$$\begin{aligned} X_t &= Y'_{t,1} \Phi_1 I(\mathbf{X}_{t-1} \leq \mathbf{c}) + Y'_{t,2} \Phi_2 I(\mathbf{X}_{t-1} > \mathbf{c}) + \varepsilon_t = \\ &= Y'_{t,1} \Phi_1 [1 - I(\mathbf{X}_{t-1} > \mathbf{c})] + Y'_{t,2} \Phi_2 I(\mathbf{X}_{t-1} > \mathbf{c}) + \varepsilon_t \end{aligned}$$

kde $Y_{t,i} = (1, X_{t-1})'$ a $\Phi_i = (\varphi_{i,0}, \varphi_{i,1})'$, $i = 1, 2$

- a) $c = 0.5$; $\varphi_{1,0} = \varphi_{2,0} = 0$; $\varphi_{1,1} = -0.8$; $\varphi_{2,1} = 0.8$;
- b) $c = 2$; $\varphi_{1,0} = \varphi_{2,0} = 0$; $\varphi_{1,1} = -0.8$; $\varphi_{2,1} = 0.8$;
- c) $c = 2$; $\varphi_{1,0} = 1.2$; $\varphi_{2,0} = -0.5$; $\varphi_{1,1} = -0.8$; $\varphi_{2,1} = 0.8$;
- d) $c = 1$; $\varphi_{1,0} = -1.2$; $\varphi_{2,0} = 0.5$; $\varphi_{1,1} = -0.8$; $\varphi_{2,1} = 0.8$;

Modelovací cyklus viacrežimových modelov

Pri tvorbe nelineárnych modelov sa vo všeobecnosti prijíma zásada “od špeciálneho ku všeobecnému”. Vhodné je začínať jednoduchšími lineárnymi modelmi a potom, keď nie sú splnené podmienky pre ich použitie, prejsť k modelom všeobecnejším a tiež komplexnejším, akými sú modely nelineárne. Doporučujú sa nasledovné kroky:

1. Pre daný časový rad určiť vhodný lineárny model typu AR(p)
2. Otestovať nulovú hypotézu linearity modelu oproti alternatívnej hypotéze nelinearity modelu
3. Odhadnúť parametre zvoleného nelineárneho modelu
4. Overiť vhodnosť modelu diagnostickými testami
5. Ak je potrebné, modifikovať model.
6. Použiť model na popisné účely a predikciu.

TESTOVANIE LINEARITY

Prvým krokom pri konštrukcii viacrežimových modelov je štatistické porovnanie „kvality“ lineárneho modelu a modelu s viacerými režimami. Testujeme hypotézu, že vhodný je lineárny model oproti alternatívnej hypotéze, že model má premenlivé režimy.

V prípade dvojrežimového modelu teda testujeme:

$H_0: \Phi_1 = \Phi_2$ $H_1: \varphi_{1,i} \neq \varphi_{2,i}$ pre minimálne jedno $i \in \{1, \dots, \max(p_1, p_2)\}$

Problém štatistických testov v prípade viacrežimových modelov je, že v prípade testovanej nulovej hypotézy H_0 sa identifikuje menej parametrov než v prípade alternatívnej hypotézy H_1 . Tieto parametre sa nazývajú **prebytočné (nuisance)**. Nelineárne modely totiž obsahujú parametre, ktoré sa v lineárnych modeloch nevyskytujú. Dôsledkom toho je, že testovacie štatistiky majú neštandardné rozdelenia pravdepodobnosti, ktoré sa často dajú vyjadriť len pomocou simulačných metód.

BODOVÉ PREDPOVEDE

Predpokladajme nelineárny model so skeleonom $F(X_{t-1}, \theta)$, t. j. model v tvare

$$X_t = F(X_{t-1}, \theta) + \varepsilon_t$$

kde F je nelineárna funkcia, θ je vektor parametrov a $\{\varepsilon_t\}$ je proces bieleho šumu s nulovou strednou hodnotou a rozptylom σ_ε^2 .

Optimálna bodová predpoveď s horizontom h zostrojená v čase T je podmienená stredná hodnota

$$X_T(h) = E(X_{T+h} | \Omega_T)$$

kde $\Omega_T = \{X_T, X_{T-1}, \dots\}$.

Základný problém je, že na rozdiel od lineárnych modelov v prípade nelineárnych modelov neplatí (vo všeobecnosti) rovnosť:

$$E(F(X_{T+h}, \theta) | \Omega_T) = F(E(X_{T+h} | \Omega_T), \theta)$$

t.j. lineárny operátor podmienenej strednej hodnoty nie je zameniteľný s nelineárnym operátorom F . Dôsledkom je, že v prípade nelineárnych modelov vo všeobecnosti neexistuje jednoduchý rekurzívny vzťah medzi predpoveďami s rôznymi horizontami.

1. Uzavretá forma predpovedi (Closed - Form Forecast)

$$X_T(h) = \int_{-\infty}^{\infty} F(X_{T+h-1}, \theta) f(X_{T+h-1} | \Omega_T) dX_{T+h-1}$$

kde $f(X_{T+h-1} | \Omega_T)$ je podmienená hustota pravdepodobnosti.

2. Metóda Monte Carlo

$$X_T(h) = \frac{1}{k} \sum_{i=1}^k F(X_T(h-1) + \varepsilon_i^v, \theta)$$

kde k je nejaké veľké prirodzené číslo a ε_i^v je vybraná hodnota z predpokladaného rozdelenia pravdepodobnosti náhodnej premennej ε_{T+h-1} .

3. Bootstrap metóda

$$X_T(h) = \frac{1}{k} \sum_{i=1}^k F(X_T(h-1) + \hat{\varepsilon}_i, \theta)$$

kde k je nejaké veľké prirodzené číslo a $\hat{\varepsilon}_i$ sú reziduá z odhadnutého modelu.

Predpovedné chyby

Aritmetický priemer druhých mocnín odchýliek (mean squared error) MSE:

$$\text{MSE} = \frac{\sum_{T=1}^P (\mathbf{x}_T - \mathbf{F}_T)^2}{P}$$

Druhá odmocnina z aritmetického priemeru druhých mocnín odchýliek (root mean squared error) RMSE:

$$\text{RMSE} = \sqrt{\frac{\sum_{T=1}^P (\mathbf{x}_T - \mathbf{F}_T)^2}{P}}$$

Priemer absolútnych odchýliek (mean absolute error) MAE :

$$\text{MAE} = \frac{\sum_{T=1}^P |\mathbf{x}_T - \mathbf{F}_T|}{P}$$

Osnova programu na prvé cvičenie

Dáta:

Z adresára „data“ si vyberte ekonomický alebo hydrologický časový rad, s ktorým budete cez semester pracovať.

Úvodné operácie:

- ✓ Nahratie balíkov na analýzu časových radov;
- ✓ Nastavenie pracovného adresára;
- ✓ Nastavenie podmienok pre ListPlot;
- ✓ Načítanie dát (v prípade potreby odstránenie prvých dvoch údajov).

Základná analýza časového radu:

- ✓ Pri hydrologických časových radoch vypočítať časový rad ročných priemerov;
- ✓ Vykresliť časový rad s popisom osí;
- ✓ Vypočítať strednú hodnotu, smerodajnú odchýlku a rozptyl.

Osnova programu na prvé cvičenie - pokračovanie

Príprava dát na modelovanie a prognózu:

- ✓ Vytvoriť skúšobnú vzorku (5 posledných dát) a testovaciu vzorku (ostávajúce dáta);
- ✓ Vykresliť skúšobnú aj testovaciu vzorku s popisom osí;
- ✓ Uložiť skúšobnú aj testovaciu vzorku spolu s ich grafmi do definičného súboru;
- ✓ Vypočítať strednú hodnotu, smerodajnú odchýlku a rozptyl testovacej vzorky;
- ✓ Vypočítať ACF a PACF pre testovaciu vzorku;
- ✓ Na základe PACF určiť pomocou `LevinsonDurbinEstimate` a informačných kritérií AIC, BIC vhodný rád p pre model $AR(p)$.