

a) Testovanie linearity

Testujeme hypotézu, že vhodný je lineárny model oproti alternatívnej hypotéze nelinearity typu STAR.

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 aj v prípade STAR 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)**. Modely STAR totiž obsahujú parametre, ktoré sa v lineárnych modeloch nevyskytujú (prahová hodnota c a „vyhladzovací“ parameter γ).

Riešenie tohto problému navrhli ako prví Luukkonen, Saikkonen a Teräsvirta (1988). My budeme používať modifikáciu odvodenú autormi Franses, Van Dijk (2000) pre LSTAR.

Označme $p = \max(p_1, p_2)$. Model STAR má tvar:

$$X_t = \Phi_1' Y_t (1 - G(q_t, \gamma, c)) + \Phi_2' Y_t G(q_t, \gamma, c) + \varepsilon_t$$

kde

$$Y_t = (1, X_{t-1}, \dots, X_{t-p})' \text{ a } \Phi_i = (\phi_{i,0}, \phi_{i,1}, \dots, \phi_{i,p})', i = 1, 2$$

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

V prípade **LSTAR** ho môžeme prepísať do tvaru:

$$X_t = \frac{1}{2}(\Phi_1 + \Phi_2)' Y_t + (\Phi_2 - \Phi_1)' Y_t G^*(q_t, \gamma, c) + \varepsilon_t$$

kde $G^*(q_t, \gamma, c) = G(q_t, \gamma, c) - 0.5$.

Funkciu $G^*(q_t, \gamma, c)$ môžeme aproximovať Taylorovým polynómom 3. stupňa v okolí bodu $\gamma = 0$.

Po úpravách dostaneme regresný model

$$X_t = \alpha_0 + \beta'_0 \tilde{\mathbf{Y}}_t + \beta'_1 \tilde{\mathbf{Y}}_t q_t + \beta'_2 \tilde{\mathbf{Y}}_t q_t^2 + \beta'_3 \tilde{\mathbf{Y}}_t q_t^3 + \hat{e}_t$$

kde

$$\tilde{\mathbf{Y}}_t = (X_{t-1}, \dots, X_{t-p})',$$

$$\beta_j = (\beta_{j,1}, \dots, \beta_{j,p})', j = 0, 1, 2, 3,$$

sú funkcie parametrov $\Phi_1, \Phi_2, \gamma, c$.

Nulová hypotéza $H_0: \gamma = 0$ (t. j. $G^*(q_t, 0, c) = 0$) je ekvivalentná hypotéze

$$H_0: \beta_1 = \beta_2 = \beta_3 = \mathbf{0}.$$

V prípade **ESTAR** je priamo $G(q_t, 0, c) = 0$. Prepíšeme ho do tvaru:

$$X_t = \Phi_1' Y_t (1 - G(q_t, \gamma, c))' + \Phi_2' Y_t G(q_t, \gamma, c) + \varepsilon_t$$

Funkciu $G(q_t, \gamma, c)$ budeme aproximovať Taylorovým polynómom 1. stupňa v okolí bodu $\gamma = 0$.

Dostaneme regresný model

$$X_t = \alpha_0 + \beta_0' \tilde{Y}_t + \beta_1' \tilde{Y}_t q_t + \beta_2' \tilde{Y}_t q_t^2 + \hat{e}_t$$

kde

$$\tilde{Y}_t = (X_{t-1}, \dots, X_{t-p})'$$

$$\beta_j = (\beta_{j,1}, \dots, \beta_{j,p})', j = 0, 1, 2 \text{ sú funkcie parametrov } \Phi_1, \Phi_2, \gamma, c.$$

Nulová hypotéza $H_0: \gamma = 0$ (t. j. $G(q_t, 0, c) = 0$) je ekvivalentná hypotéze

$$H_0: \beta_1 = \beta_2 = 0.$$

Postup pri testovaní hypotézy linearity:

1. Na základe lineárneho modelu, v ktorom je na ľavej strane X_t a na pravej strane $Y_t = (1, X_{t-1}, \dots, X_{t-p})'$ sa vypočítajú rezíduá e_t a reziduálny súčet štvorcov SSR_0 .
2. Z regresného modelu, v ktorom sú na ľavej strane rezíduá e_t z lineárneho modelu a na pravej strane $1, \tilde{Y}_{t,q}^j$ ($j = 1, 2, 3$ pri LSTAR, resp. $j = 1, 2$ pri ESTAR) sa vypočítajú rezíduá, reziduálny súčet štvorcov SSR_1 a index determinácie $R^2 = 1 - SSR_1/SSR_0$.
3. Testovacia štatistika $LM = n R^2$ má za predpokladu platnosti nulovej hypotézy pri LSTAR asymptoticky rozdelenie $\chi^2(3p)$ a pri ESTAR asymptoticky rozdelenie $\chi^2(2p)$.

Tento test sa používa aj na výber vhodnej prahovej premennej $q_t = X_{t-d}$. Test sa urobí pre viacerých vhodných kandidátov a vyberie sa prahová premenná s minimálnou P-value.

Voľba vhodnej prechodovej funkcie (výber LSTAR alebo ESTAR):

Pre **LSTAR** je regresný model

$$X_t = \alpha_0 + \beta'_0 \tilde{Y}_t + \beta'_1 \tilde{Y}_t q_t + \beta'_2 \tilde{Y}_t q_t^2 + \beta'_3 \tilde{Y}_t q_t^3 + \hat{e}_t$$

Pre **ESTAR** je regresný model

$$X_t = \alpha_0 + \beta'_0 \tilde{Y}_t + \beta'_1 \tilde{Y}_t q_t + \beta'_2 \tilde{Y}_t q_t^2 + \hat{e}_t$$

Nulová hypotéza $H_0: \beta_3 = 0$.

Zamietnutie H_0 znamená zamietnutie modelu z triedy ESTAR.

Lineárna regresia

Lineárny regresný model:

$$\mathbf{X} = \mathbf{A} \cdot \boldsymbol{\theta} + \mathbf{e}$$

kde $\mathbf{X}$ je $(n \times 1)$ vektor známych hodnôt, $\mathbf{A}$ je matica regresných koeficientov typu $n \times m$ ($n > m$), $\boldsymbol{\theta}$ je $(m \times 1)$ vektor neznámych parametrov, $\mathbf{e}$ je $(n \times 1)$ vektor rezíduí (nezávislé náhodné premenné s nulovými strednými hodnotami a rozptylmi σ_i^2).

Parametre $\boldsymbol{\theta}$ odhadneme metódou najmenších štvorcov, t.j. riešime rovnice

$$\sum_{i=1}^n \sigma_i^{-2} (\mathbf{X}_i - \mathbf{A}_i \hat{\boldsymbol{\theta}})^2 = \min_{\boldsymbol{\theta} \in \mathbb{R}^m} \sum_{i=1}^n \sigma_i^{-2} (\mathbf{X}_i - \mathbf{A}_i \boldsymbol{\theta})^2$$

Označme $\boldsymbol{\Sigma}$ diagonálnu maticu

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

Vektor $\hat{\theta}$ dostaneme riešením normálnych rovníc:

$$\mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{A} \cdot \hat{\theta} = \mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{X}$$

ktoré majú vždy aspoň jedno riešenie. Matica $\mathbf{M} = \mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{A}$ sa nazýva **informačná matica**, pričom $h(\mathbf{M}) = h(\mathbf{A})$.

Ak $h(\mathbf{M}) = h(\mathbf{A}) = m$, normálne rovnice majú jediné riešenie

$$\hat{\theta} = \mathbf{M}^{-1} \cdot \mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{X}$$

Kovariančná matica odhadu $\hat{\theta}$ je matica $\mathbf{M}^{-1}$.

Osnova programu na piate cvičenie

Výpočet predpovedí pomocou modelu SETAR:

Pre 3 najlepšie modely (podľa výsledkov testu ostávajúcej nelinearity 2- alebo 3- režimové)

- vypočítať jednokrokové predpovede na 5 časových jednotiek (naivnou metódou aj metódou Monte Carlo, resp. bootstrap metódou)
- zobrazit' ich spolu so skúšobnou vzorkou
- vypočítať chyby predpovedí.

Zhodnotenie modelu SETAR:

Vybrať najlepší SETAR model

- a) na popisné účely (podľa smerodajnej odchýlky rezíduí, ale aj počtu dát v jednotlivých režimoch a polohe parametra c)
- b) na prognózovanie (podľa predpovedných chýb).