

b) Testovanie linearity pre model SETAR

Najprv sa odhadnú parametre lineárneho modelu (s absolútnym členom) aj modelu SETAR. Reziduálny rozptyl pri nulovej hypotéze (t. j. lineárnom modele) označíme $\hat{\sigma}_{\varepsilon_L}^2$, pri alternatívnej hypotéze (t. j. modele SETAR) $\hat{\sigma}_{\varepsilon_S}^2(\hat{c})$.

Testovacia štatistika (F - test):

$$LR_{L,S}(\hat{c}) = n \left(\frac{\hat{\sigma}_{\varepsilon_L}^2 - \hat{\sigma}_{\varepsilon_S}^2(\hat{c})}{\hat{\sigma}_{\varepsilon_S}^2(\hat{c})} \right)$$

Ak označíme $p = \max(p_1, p_2)$, pre pevne danú prahovú hodnotu c má štatistika $LR_{L,S}$ asymptotické rozdelenie $\chi^2(p + 1)$. Problém je v tom, že štatistiku $LR_{L,S}$ konštruujeme pre viaceré hodnoty c a testovacie kritérium je maximum z týchto rôznych závislých hodnôt s asymptotickým rozdelením χ^2 . Rozdelenie pravdepodobnosti pre testovacie kritérium je teda väčšie ako $\chi^2(p + 1)$.

$$LR_{L,S} = \sup_{c \in C} LR_{L,S}(c) = \sup_{c \in C} n \left(\frac{\hat{\sigma}_{\varepsilon_L}^2 - \hat{\sigma}_{\varepsilon_S}^2(c)}{\hat{\sigma}_{\varepsilon_S}^2(c)} \right)$$

Hansen v článku Inference in TAR models (1997) odvodil asymptotické rozdelenie pre testovaciu štatistiku $LR_{L,S}$ za nasledovných predpokladov:

1. ε_t je biely šum, $\varepsilon_t \approx N(0, \sigma_\varepsilon^2)$, $E(\varepsilon_t^k) < \infty$

2. Pre $i = 1, 2$ platí: $\sum_{j=1}^p |\varphi_{i,j}| < 1$

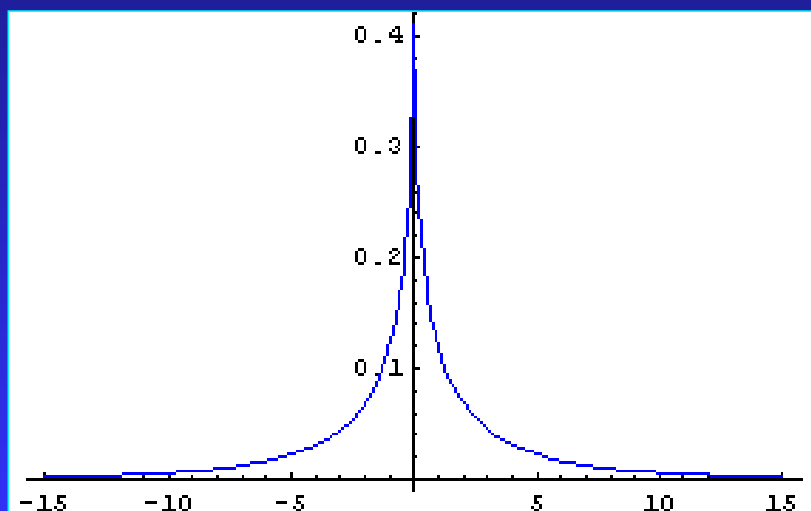
3. Platí buď $(\varphi_{1,0} - \varphi_{2,0}) + (\varphi_{1,d} - \varphi_{2,d}) c \neq 0$ alebo $\varphi_{1,j} \neq \varphi_{2,j}$ pre $j \neq 0, d$

Pri splnení týchto predpokladov má testovacia štatistika približne distribučnú funkciu :

$$P(T \leq x) = \begin{cases} 1 + \sqrt{\frac{x}{2\pi}} e^{-x/8} + \frac{3}{2} e^{x\Phi} \left(-\frac{3\sqrt{x}}{2} \right) - \left(\frac{x+5}{2} \right) \Phi \left(-\frac{\sqrt{x}}{2} \right) & x \geq 0 \\ -\sqrt{\frac{-x}{2\pi}} e^{x/8} - \frac{3}{2} e^{-x\Phi} \left(-\frac{3\sqrt{-x}}{2} \right) + \left(\frac{-x+5}{2} \right) \Phi \left(-\frac{\sqrt{-x}}{2} \right) & x < 0 \end{cases}$$

kde $\Phi(x)$ je distribučná funkcia rozdelenia $N(0, 1)$.

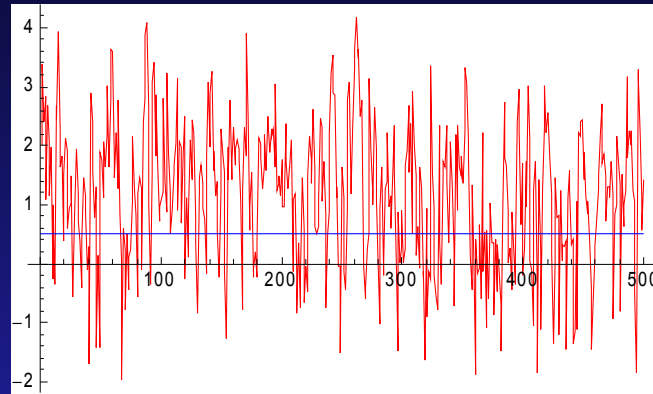
Graf hustoty pravdepodobnosti:



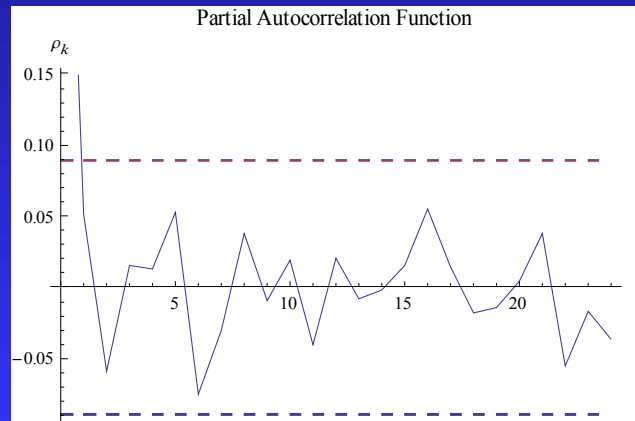
Príklad 1

(2-režimový SETAR, predpoklady Hansenovej vety sú splnené)

Vygenerujeme časový rad dĺžky $n = 500$ pomocou procesu SETAR pre $c = 0.5$;
 $\phi_{1,0} = \phi_{2,0} = 0$; $\phi_{1,1} = -0.8$; $\phi_{2,1} = 0.8$; $X_0 = 2$.



Budeme testovať linearitu tohto časového radu. Pomocou PAF odhadneme vhodný rád pre lineárny AR(p) model:



Vidíme, že vhodný rád je $p = 2$. Potvrdíme si to:

$p = 1$	AIC= 0.112193	BIC= 0.120623
$p = 2$	AIC= 0.11351	BIC= 0.130369
$p = 3$	AIC= 0.11407	BIC= 0.139358
$p = 4$	AIC= 0.117849	BIC= 0.151565
$p = 5$	AIC= 0.121664	BIC= 0.16381

Určíme $c_{\min} = -0.605193$ a $c_{\max} = 2.85402$. Na tejto množine budeme pre $p = 1, 2$ odhadovať parametre minimalizáciou reziduálneho rozptylu a informačných kritérií AIC a BIC.

$c = 0.467162$	$d = 1$	$p = 1$	BIC = 78.8866	AIC = 89.1424
$c = 0.467162$	$d = 1$	$p = 2$	BIC = 79.2374	AIC = 90.121

Menšie hodnoty AIC a BIC sú pre $p = 1$. Autoregresné parametre pre jednotlivé režimy sú:

$$\phi_1 = \{0.16291, -0.64351\}$$
$$\phi_2 = \{0.38273, 0.61582\}$$

Teda $c = 0.467162$, $d = 1$, autoregresné parametre pre jednotlivé režimy sú:

$$\phi_1 = \{0.16291, -0.64351\}$$

$$\phi_2 = \{0.38273, 0.61582\}$$

2. Pre $i = 1, 2$ platí: $\sum_{j=1}^p |\phi_{i,j}| < 1$

$$\sum_{j=1}^1 |\phi_{1,j}| = 0.64351 < 1$$

$$\sum_{j=1}^1 |\phi_{2,j}| = 0.61582 < 1$$

3. Platí buď $(\phi_{1,0} - \phi_{2,0}) + (\phi_{1,d} - \phi_{2,d}) c \neq 0$ alebo $\phi_{1,j} \neq \phi_{2,j}$ pre $j \neq 0, d$

$$= -0.808136$$

Pretože podmienky sú splnené, môžeme použiť na testovanie linearity Hansenovu distribučnú funkciu.

Reziduálny rozptyl pre lineárny AR(1) model je 1.11284, pre SETAR model 0.973114. Potom

$$LR_{L,S}(\hat{c}) = 500 \left(\frac{1.11284 - 0.973114}{0.973114} \right) = 71.7924$$

P-value pre túto hodnotu je $< 10^{-6}$, teda nulovú hypotézu zamietame $\Rightarrow$ lineárny model nestačí; je potrebné pridať ďalší (druhý) režim.

V systéme MATHEMATICA definujeme funkciu:

$$LR[s_1_, s_2_] := n * (s_1 - s_2) / s_2$$

Distribučná funkcia aj vzorec na výpočet P-value je v definičnom súbore

LR_test.def

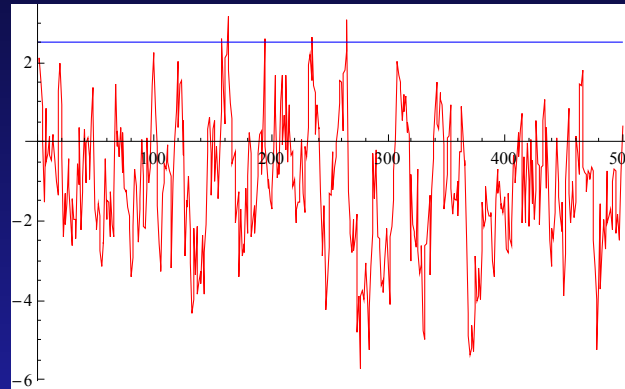
P-value pre danú hodnotu testovacej štatistiky počítame príkazom

pV[LR[s₁, s₂]]

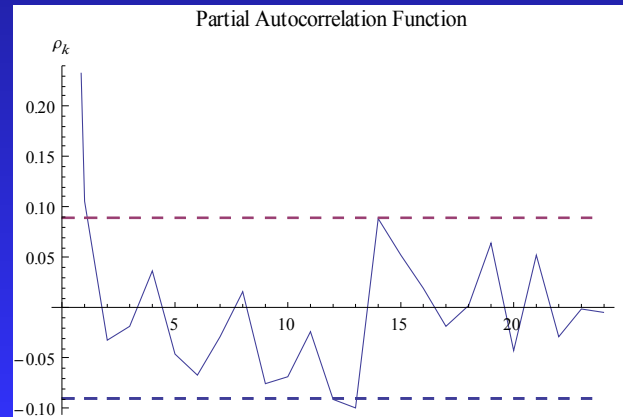
Príklad 2

(1-režimový model, predpoklady Hansenovej vety sú splnené)

Vygenerujeme časový rad dĺžky $n = 500$ pomocou procesu SETAR pre $c = 2.5$;
 $\phi_{1,0} = -0.2$; $\phi_{2,0} = 0$; $\phi_{1,1} = 0.8$; $\phi_{2,1} = 0$; $X_0 = 2$.



Pomocou PAF odhadneme vhodný rád pre lineárny AR(p) model:



Vhodný rád je $p = 2$. Potvrdíme si to:

$p = 1$	AIC= 0.0957753	BIC= 0.104204
$p = 2$	AIC= 0.088500	BIC= 0.105359
$p = 3$	AIC= 0.0915117	BIC= 0.116799
$p = 4$	AIC= 0.0951994	BIC= 0.128916
$p = 5$	AIC= 0.0978899	BIC= 0.140036

Určíme $c_{\min} = -3.51508$ a $c_{\max} = 1.40381$. Na tejto množine budeme pre $p = 1, 2, 3$ odhadovať parametre minimalizáciou reziduálneho rozptylu a informačných kritérií AIC a BIC.

$c = 1.2562$	$d = 1$	$p = 1$	BIC = 129.866	AIC = 118.099
$c = 0.5676$	$d = 1$	$p = 2$	BIC = 150.216	AIC = 131.181
$c = 0.5676$	$d = 1$	$p = 3$	BIC = 159.26	AIC = 133.881

Najmenšie hodnoty AIC a BIC sú pre $p = 1$. Autoregresné parametre pre jednotlivé režimy sú:

$$\phi_1 = \{-0.27192, 0.766574\}$$

$$\phi_2 = \{1.07869, 0.061195\}$$

Teda $c = 1.25624$, $d = 1$, autoregresné parametre pre jednotlivé režimy sú:

$$\phi_1 = \{-0.27192, 0.766574\}$$

$$\phi_2 = \{1.07869, 0.061195\}$$

Pre $i = 1, 2$:

$$\sum_{j=1}^1 |\phi_{1,j}| = 0.766574 < 1$$

$$\sum_{j=1}^1 |\phi_{2,j}| = 0.061195 < 1$$

$$\begin{aligned} (\phi_{1,0} - \phi_{2,0}) + (\phi_{1,d} - \phi_{2,d}) c &= (-0.27192 - 1.07869) + (0.766574 - 0.061195) * 1.25624 = \\ &= -0.464485 \end{aligned}$$

Pretože podmienky sú splnené, môžeme použiť na testovanie linearity Hansenovu distribučnú funkciu.

Reziduálny rozptyl pre lineárny AR(1) model je 1.07374, pre SETAR model 1.0646. Potom

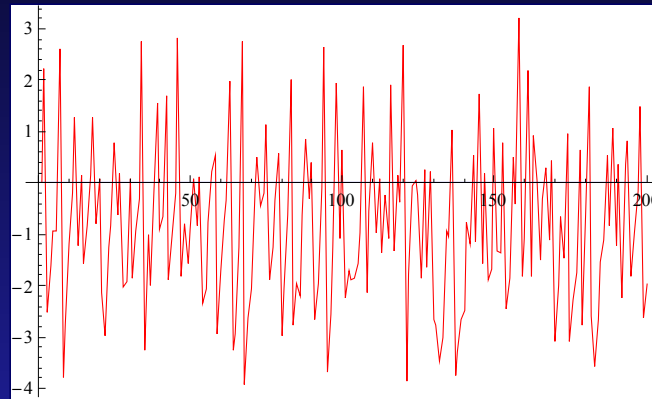
$$LR_{L,S}(\hat{c}) = 500 \left(\frac{1.07374 - 1.0646}{1.0646} \right) = 4.29365$$

P-value pre túto hodnotu je 0.110723, teda nulovú hypotézu na hladine významnosti $\alpha = 0.05$ nezamietame $\Rightarrow$ lineárny model je vhodný a nie je potrebné pridať ďalší (druhý) režim.

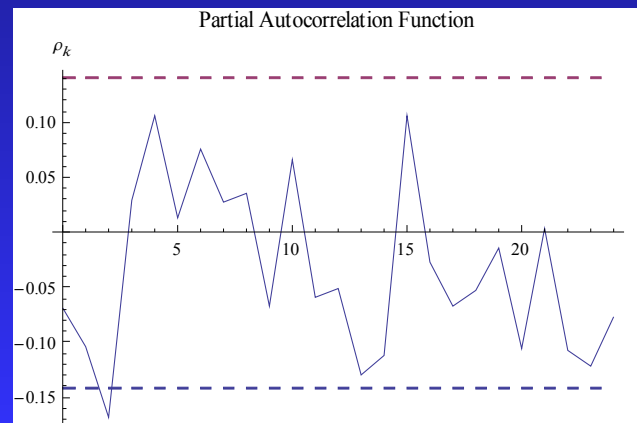
Príklad 3

(2-režimový SETAR, nie sú splnené predpoklady Hansenovej vety)

Vygenerujeme časový rad dĺžky $n = 200$ pomocou procesu SETAR pre $c = 0$;
 $\phi_{1,0} = 1.5$; $\phi_{2,0} = -1.2$; $\phi_{1,1} = 1.2$; $\phi_{2,1} = -0.6$; $X_0 = 0$.



Pomocou PAF odhadneme vhodný rád pre lineárny AR(p) model:



Vidíme, že vhodný rád je $p = 3$. Potvrdíme si to:

$p = 1$	AIC= 0.875102	BIC= 0.891593
$p = 2$	AIC= 0.87407	BIC= 0.907053
$p = 3$	AIC= 0.85558	BIC= 0.905055
$p = 4$	AIC= 0.864732	BIC= 0.930699
$p = 5$	AIC= 0.863548	BIC= 0.946006
$p = 6$	AIC= 0.873365	BIC= 0.972315

Určíme $c_{\min} = -2.9811$ a $c_{\max} = 1.7480$. Na tejto množine budeme pre $p = 1, 2, 3$ odhadovať parametre minimalizáciou reziduálneho rozptylu a informačných kritérií AIC a BIC.

$c = 0.0455426$	$d = 1$	$p = 1$	BIC = 144.904	AIC = 134.865
$c = -0.0963315$	$d = 1$	$p = 2$	BIC = 150.153	AIC = 135.012
$c = -0.0963315$	$d = 1$	$p = 3$	BIC = 157.283	AIC = 137.095

Najmenšie hodnoty AIC a BIC sú pre $p = 1$. Autoregresné parametre pre jednotlivé režimy sú:

$$\phi_1 = \{ 1.33936, \quad 1.09468 \}$$
$$\phi_2 = \{ -1.15862, \quad -0.59035 \}$$

Teda $c = 0.045542$, $d = 1$, autoregresné parametre pre jednotlivé režimy sú:

$$\phi_1 = \{ 1.33936, \quad 1.09468 \}$$

$$\phi_2 = \{-1.15862, \quad -0.59035\}$$

Pre $i = 1, 2$:

$$\sum_{j=1}^1 |\phi_{1,j}| = 1.09468 > 1 \quad \text{neplatí!}$$

$$\sum_{j=1}^1 |\phi_{2,j}| = 0.59035 < 1$$

$$(\phi_{1,0} - \phi_{2,0}) + (\phi_{1,d} - \phi_{2,d}) c = (1.33936 + 1.15862) + (1.09468 + 0.59035) \cdot 0.045542 =$$

$$= 2.57472$$

Pretože jedna z podmienok nie je splnená, nemali by sme použiť na testovanie linearity Hansenovu distribučnú funkciu, ale určiť kritické hodnoty testovacej štatistiky simuláciou.

Reziduálny rozptyl pre lineárny AR(1) model je 1.15185, pre SETAR model 1.11135. Potom

$$LR_{L,S}(\hat{c}) = 200 \left(\frac{1.15185 - 1.11135}{1.11135} \right) = 7.28753$$

P-value pre túto hodnotu je 0.05457, teda nulovú hypotézu na hladine významnosti $\alpha = 0.05$ nezamietame $\Rightarrow$ lineárny model je vhodný a nie je potrebné pridať ďalší (druhý) režim. **Tento záver nie je správny!**

($N_1 = 140$, $N_2 = 60$)

c) Diagnostická kontrola pre model SETAR

Ako prvé treba **otestovať** zostávajúcu **nelinearitu**. V prípade 2-režimového SETAR modelu je H_0 : vhodný je 2-režimový model oproti alternatívnej hypotéze H_1 : vhodný je 3-režimový model.

Ako testovacia štatistika sa opäť používa LR-štatistika

$$LR_{s_2, s_3} = n \left(\frac{\hat{\sigma}_{\varepsilon s_2}^2 - \hat{\sigma}_{\varepsilon s_3}^2}{\hat{\sigma}_{\varepsilon s_3}^2} \right)$$

s rovnakým asymptotickým rozdelením.

Testovanie autokorelácie reziduí si ukážeme neskôr.

Bodová predpoveď pre model SETAR

Naivná jednokroková predpoveď:

$$X_{T+1} = F(X_T, \theta)$$

Bootstrap metóda (resp. Monte Carlo metóda)

$$\begin{aligned} X_{T+1} &= \frac{1}{k} \sum_{i=1}^k F(X_T + \hat{\varepsilon}_i, \theta) = \\ &= \frac{1}{k} \sum_{i=1}^k \begin{cases} \varphi_{1,0} + \varphi_{1,1}(X_T + \hat{\varepsilon}_T) + \varphi_{1,2}X_{T-1} + \dots + \varphi_{1,p_1}X_{T+1-p_1} & X_{T+1-d} \leq c \\ \varphi_{2,0} + \varphi_{2,1}(X_T + \hat{\varepsilon}_T) + \varphi_{2,2}X_{T-1} + \dots + \varphi_{2,p_2}X_{T+1-p_2} & X_{T+1-d} > c \end{cases} \end{aligned}$$

kde k je nejaké veľké prirodzené číslo a $\hat{\varepsilon}_i$ sú reziduá z odhadnutého modelu (resp. hodnota vygenerovaná z predpokladaného rozdelenia pravdepodobnosti náhodnej premennej ε_T).

5 jednokrokových predpovedí pre 2-režimový SETAR z príkladu 1

Časový rad dĺžky $n = 500$, pre ktorý sme odhadli $c = 0.467162$, $d = 1$, $p_1 = p_2 = 1$, autoregresné parametre pre jednotlivé režimy:

$$\phi_1 = \{0.16291, -0.64351\}$$

$$\phi_2 = \{0.38273, 0.61582\}$$

Skeleton:

$$X_t = \begin{cases} 0.16291 - 0.64351 * X_{t-1} & X_{t-1} \leq 0.467162 \\ 0.38273 + 0.61582 * X_{t-1} & X_{t-1} > 0.467162 \end{cases}$$

Posledná hodnota X_{500} časového radu = 2.16535.

$$X_{501} = 0.38273 + 0.61582 * X_{500} = 1.72622$$

$$X_{502} = 0.38273 + 0.61582 * X_{501} = 1.45579$$

$$X_{503} = 0.38273 + 0.61582 * X_{502} = 1.28925$$

$$X_{504} = 0.38273 + 0.61582 * X_{503} = 1.18669$$

$$X_{505} = 0.38273 + 0.61582 * X_{504} = 1.12353$$

Testovanie linearity oproti nelinearite typu SETAR:

- Nahrať súbor LR_test.def a vlastný definičný súbor
- Pre 3 najlepšie modely (s najmenším BIC):
 1. vypočítať hodnotu autoregresných parametrov pre jednotlivé režimy
 2. pomocou informačnej matice vypočítať štandardné chyby odhadov autoregresných parametrov
 3. vypočítať počet dát v prvom a druhom režime
 4. vypočítať rezíduá a ich rozptyl $\hat{\sigma}_{\varepsilon_S}^2$
 5. vytvoriť model a zobrazíť ho s originálnymi dátami (testovacou vzorkou)
 6. odhadnúť parametre lineárneho AR(p) modelu s absolútnym členom (pre príslušné p) a vypočítať rozptyl jeho rezíduí $\hat{\sigma}_{\varepsilon_L}^2$
 7. vypočítať P-value LR testu