

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY  
UNIVERZITY KOMENSKÉHO V BRATISLAVE

Katedra aplikovanej matematiky a štatistiky



# Informačné kritéria pre STAR modely

Viacrežimové modelovanie časových radov

Diplomová práca

Bratislava 2007

Martin Handlovič

Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislave  
Katedra aplikovanej matematiky a štatistiky

Matematika

Ekonomická a finančná matematika



# INFORMAČNÉ KRITÉRIA PRE STAR MODELY

## Viacrežimové modelovanie časových radov

Diplomová práca

Diplomant: Martin Handlovič

Vedúci diplomovej práce: Prof. RNDr. Magdaléna Komorníková, CSc.

Bratislava 2007

# Čestné prehlásenie

Čestne prehlasujem, že túto diplomovú prácu som vypracoval samostatne len s použitím uvedenej citovanej literatúry.

V Bratislave, máj 2007

.....

# Pod'akovanie

Touto cestou by som sa chcel poďakovať mojej školiteľke p.prof. RNDr. Magdaléne Komorníkovej, za jej vysoko odbornú spoluprácu na tejto diplomovej práci, za jej nesmiernu ochotu. Poskytla mi množstvo cennej literatúry, z ktorej som čerpal námety. V neposlednom rade jej patrí veľká vďaka za množstvo času a trpezlivosti, ktorý mi pri vzniku tejto práce venovala.

Samostatné poďakovanie patrí aj Ing.Tomášovi Bacigálovi, ktorého program vo výpočtovom systéme Mathematica som využil pri aplikácií na reálnych dátach.

# Abstrakt

Cieľom diplomovej práce je popísať najčastejšie používané moderné techniky na analyzovanie časových radov a následne vyhodnocovanie týchto radov na základe informačných kritérií. Špeciálne sa budeme zaoberať STAR modelmi a informačnými kritériami pre tieto modely. Vhodnosť týchto modelov pre analýzu časových radov podporíme aplikáciou na reálnych dátach.

Kľúčové slová: Modelovanie časových radov, informačné kritéria, STAR modely

# Abstract

The main purpose of this diploma thesis is to demonstrate the modern techniques used to analyse the given time series by a model and compare it with alternative models using the information criteria. We will focus on STAR models and its information criteria. We will prove the need of using these models in applications on real data.

Key words: Model selection, information criteria, STAR models

# Obsah

<b>1</b>	<b>Úvod</b>	<b>8</b>
<b>2</b>	<b>Analýza časového radu</b>	<b>10</b>
2.1	Spoločný úvod . . . . .	10
2.2	Dekompozícia časového radu . . . . .	11
2.2.1	Trendová zložka . . . . .	11
2.2.2	Sezónna zložka . . . . .	12
2.2.3	Cyklická zložka . . . . .	12
2.2.4	Reziduálna zložka . . . . .	12
<b>3</b>	<b>Prehľad modelov</b>	<b>13</b>
3.1	Lineárne modely - jednorozmerné . . . . .	13
3.1.1	MA, AR, ARMA modely . . . . .	14
3.1.2	Stacionarita a invertovateľnosť . . . . .	16
3.1.3	Určenie vhodného modelu . . . . .	17
3.1.4	Odhad parametrov modelu . . . . .	19
3.1.5	Bodové predpovede . . . . .	21
3.1.6	Výhody/Nevýhody lineárnych modelov a príklady použitia . . . . .	21
3.2	Nelineárne modely . . . . .	22
3.2.1	Úvod k viacrežimovým modelom . . . . .	23
3.2.2	TAR modely . . . . .	23
<b>4</b>	<b>STAR modely</b>	<b>25</b>
4.1	Definícia STAR modelov . . . . .	25
4.2	Prechodová funkcia . . . . .	25
4.3	Výstavba STAR modelov . . . . .	27
<b>5</b>	<b>Informačné kritéria</b>	<b>30</b>
5.1	Úvod . . . . .	30
5.1.1	FPE . . . . .	31
5.1.2	Kullback-Leiblerova informácia . . . . .	31
5.2	IC založené na K-L informácii . . . . .	33
5.3	IC založené na penalizácii počtu parametrov . . . . .	34
5.3.1	Bayesovské informačné kritérium - BIC . . . . .	34
5.3.2	DIC, HQIC . . . . .	35

5.4	IC pre STAR modely . . . . .	36
5.4.1	Všeobecný tvar IC pre STAR modely . . . . .	36
5.4.2	Nové informačné kritéria pre STAR modely . . . . .	36
<b>6</b>	<b>Konzistentnosť informačných kritérií</b>	<b>38</b>
6.1	Lineárne modely . . . . .	38
6.2	SETAR modely . . . . .	39
6.3	STAR modely . . . . .	40
6.4	Konzistentnosť IC pre STAR modely . . . . .	42
<b>7</b>	<b>Aplikácia na reálnych dátach</b>	<b>43</b>
7.1	Popis programu . . . . .	44
7.2	HDP . . . . .	45
7.3	Inflácia . . . . .	47
7.4	Nezamestnanosť . . . . .	49
7.5	Výmenný kurz EUR/SKK . . . . .	52
7.6	Výmenný kurz USD/SKK . . . . .	55
<b>8</b>	<b>Odporúčania a záver</b>	<b>58</b>
	<b>Literatúra</b>	<b>59</b>

# 1 Úvod

V poslednom období sa problematika časových radov sa stala veľmi populárna. Snaha spoznávať javy a zákonitosti do hĺbky a vedieť skúmané premenné predikovať (predpovedať nasledujúce hodnoty na základe skúmaného časového radu) našla svoje opodstatnenie nielen v matematike, ale aj vo finančnom, bankovom, či ekonomickom sektore ako takom.

Na to, aby sme mohli predikovať budúce hodnoty, musíme v prvom rade byť schopní danému časovému radu priradiť "vhodný" model. Na základe tohto modelu potom určíme budúce hodnoty.

V súvislosti s vyššie uvedenými skutočnosťami čelíme štyrom zásadným problémom:

- Aké zložité modely použiť?
- Ako určiť, ktorý model je lepší?
- Sú dané metódy, ktoré sa používajú na výber "vhodného" modelu aj matematicky "najvhodnejšie"?
- Ako z reálnych dát určiť vhodný model?

Z matematického hľadiska môžeme zostaviť schému, ktorú odporúča Granger [11]:

- ŠPECIFIKÁCIA MODELU - určenie vhodného modelu
- ODHAD MODELU - výpočet parametrov modelu
- OVERENIE MODELU - určenie kvality modelu a jeho porovnanie s inými modelmi a jeho prípadná modifikácia
- PREDPOVEĎ alebo POPIS - predpovedanie ďalších hodnôt, alebo popísanie daného problému výsledným modelom

V nasledujúcich kapitolách sa pokúsime odpovedať na tieto problémy a položíme všeobecný, ako aj matematický základ zvoleným modelom.

Práca sa skladá z úvodu, šiestich kapitol a záveru.

V druhej kapitole popíšeme základnú analýzu časového radu - rozklad na deterministickú a reziduálnu zložku, tzv. dekompozíciu.



V tretej kapitole sa pokúsime zodpovedať prvý problém. Zameriame sa na triedy modelov. V jednotlivých podkapitolách si prejdeme modely od jednoduchých k zložitejším, od lineárnych modelov až po nelineárne modely. Uvedieme si rozdiely medzi týmito modelmi a pokúsime sa v skratke popísať výhody a nevýhody daných modelov. Budeme sa zaoberať špeciálnou podtriedou jednorozmerných nelineárnych modelov, a to viacrežimovými modelmi.

V štvrtej kapitole sa špeciálne zameriame na STAR modely, ktoré budú predmetom skúmania tejto diplomovej práce, vysvetlíme si ich pôvod, uvedieme základnú charakteristiku a postupne sa prepracujeme cez viaceré typy prechodovej funkcie STAR modelov.

V piatej kapitole sa pokúsime zodpovedať druhý problém, t.j. zameriame sa na definíciu informačných kritérií a uvedieme si prehľad najpoužívanejších kritérií od najelementárnejších po sofistikované kritéria vyžadujúce isté matematické znalosti.

Šiesta kapitola bude rýdzo matematickou kapitolou, kde zhrnieme doterajšie poznatky o vhodnosti výberu modelu pomocou konzistentnosti informačných kritérií a pokúsime sa aplikovať tieto poznatky pre STAR modely.

Siedma kapitola bude obsahovať modelovanie reálnych dát pomocou vybraných modelov za pomoci programu MATHEMATICA. V jednotlivých podkapitolách si bližšie rozoberieme získané výsledky a sformulujeme závery o vhodnosti použitia jednotlivých tried modelov.

V závere sa pokúsime zhrnúť ciele tejto diplomovej práce a načrtneme možné rozšírenia menovaním ďalších oblastí využitia časových radov.

## 2 Analýza časového radu

### 2.1 Spoločný úvod

Položme si na začiatku otázku: Čo vlastne je *časový rad*?

Z matematického hľadiska je to usporiadaný súbor dát  $(y_1, y_2, \dots, y_n)$  pozorovaní premennej  $Y$ . Dolný index pozorovania  $y_t$  reprezentuje čas  $t = 1, 2, \dots, n$ , kde  $n$  predstavuje dĺžku daného časového radu. Je pochopiteľné uvažovať, že pozorovania  $y_t$  majú stochastický charakter, a preto  $(y_1, y_2, \dots, y_n)$  budeme považovať za jednu realizáciu stochastického procesu  $(Y_1, Y_2, \dots, Y_n)$ . Kvôli jednoduchosti budeme v našej práci používať pre náhodnú premennú, ako aj pre jednu jej realizáciu rovnaké označenie  $y_t$ .

Na to, aby sme mohli uvažovať o modeloch pre časové rady, musíme prijať isté predpoklady pre tieto časové rady. Budeme uvažovať len diskrétné časové rady s ekvidistančným časovým krokom [6], t.j. rovnakým časovým krokom, lebo všetky ekonomické ukazovatele sú na dennej, mesačnej, štvrtročnej, či ročnej báze. V prípade, že by sme chceli uvažovať o spojitej premennej, z pochopiteľných dôvodov z nej spravíme diskrétnu verziu použijúc časové rezy.

Pri pohľade na akékoľvek dáta si uvedomujeme, že časový rad je tvorený dvomi typmi premenných:

- Deterministické premenné
- Stochastické premenné.

Z vyššie uvedeného vyplýva, že všeobecne môžeme časový rad napísať v tvare aditívnej dekompozície

$$y_t = D_t + \mathcal{E}_t. \quad (2.1)$$

V praxi sa často stretávame aj s časovým radom v tvare multiplikatívnej dekompozície

$$y_t = D_t \cdot \mathcal{E}_t. \quad (2.2)$$

Poznamenajme, že prechod od multiplikatívneho tvaru k aditívnemu možno previesť logaritmickou transformáciou modelu.

V tejto diplomovej práci v krátkosti zhrnieme dekompozíciu časového radu a vo zvyšku práce budeme uvažovať výhradne časové rady tvorené stochastickým procesom. Podrobnosti o vlastnostiach a určení deterministických premenných možno nájsť hlavne v [6].

## 2.2 Dekompozícia časového radu

Uvažujme časový rad v aditívnom tvare

$$y_t = T_t + S_t + C_t + \mathcal{E}_t, \quad (2.3)$$

kde  $y_t$  je skúmaný časový rad,  $T_t$  je trendová,  $S_t$  je sezónna,  $C_t$  je cyklická a  $\mathcal{E}_t$  je reziduálna zložka.

### 2.2.1 Trendová zložka

Pre jednoduchosť uvažujme o trende ako o deterministickej zložke. Vo všeobecnosti tento predpoklad nemusí byť splnený, trendová zložka môže byť aj stochastická. Tejto problematike sa venuje napr. [9].

Metódy na určenie trendovej zložky časového radu môžeme rozdeliť na tri základné:

- Subjektívne určený trend
- Trend určený matematickou funkciou
- Klízavé priemery.

Prvá z uvedených metód sa prevažne používa na určenie trendu na základe vizuálnej predstavy, kým druhá metóda sa snaží matematickou funkciou opísať trend daného časového radu.

Ako trend určený matematickou funkciou sa najčastejšie používajú polynómy (hlavne konštantný, lineárny a kvadratický trend), logaritmické a exponenciálne funkcie. V prípade, že ide o dáta z finančného, či ekonomického sektora, veľmi často uvažujeme logistickú krivku, prípadne jej modifikácie (viď [5]).

Najadaptabilnejšia z uvedených metód je metóda klízavých priemerov, ktorá je schopná vyrovnávať sa aj s časovými radmi, ktorých trend sa v čase výrazne mení. Nevýhodou tejto metódy je, že ňou strácame množstvo dát, čo sa môže výrazne prejavíť pri obmedzenej dĺžke pozorovanej vzorky.

### 2.2.2 Sezónna zložka

Nie menej významnú úlohu v dekompozícii časového radu hrá sezónna zložka, ktorej dôležitosť stúpa v prípade reálnych dát z ekonómie. Je napríklad prirodzené, že najviac peňazí ľudia míňajú na Vianoce na darčeky a v lete na dovolenku, kým v ostatných mesiacoch je spotreba nižšia. Preto v takomto časovom rade bude sezónna zložka nezanedbateľná. Pozor si treba dávať v prípade použitia kľzavých priemerov (2.2.1), pretože tie môžu deformovať sezónnu zložku ([6]).

Na dekompozíciu sezónnej zložky možno použiť jednoduchú aritmetickú metódu založenú na centrovaní, alebo zložitejšiu regresnú metódu s použitím tzv. dummy premenných. V poslednej dobe veľmi dobre výsledky vykazuje aj metóda založená na goniometrických funkciách  $\sin x$  a  $\cos x$ .

### 2.2.3 Cyklická zložka

Ide o veľmi ťažko identifikovateľnú zložku časového radu. V princípe sa táto zložka prejavuje až pri väčšom počte dát z dlhšieho časového obdobia, ale paradoxne ju možno interpretovať ako stochastickú zložku okolo trendu, preto sa často z dekompozície časového radu úplne vynecháva a chyba sa väčšinou vyrovnáva pri analýze stochastickej zložky.

Na presnejšiu analýzu cyklickej zložky sa využíva tzv. spektrálna analýza, ktorej podrobnosti možno nájsť v [6].

Z hľadiska ekonomiky je táto zložka veľmi dôležitá, keďže pojem "business cycle" sa v posledných rokoch dostáva do popredia.

### 2.2.4 Reziduálna zložka

Reziduálna zložka ostane v časovom rade po odstránení deterministických zložiek, t.j. trendu, sezónnej a cyklickej zložky. Je tvorená náhodnými pohybmi v časovom rade, ktoré nemajú rozpoznateľný systematický charakter. Ako sme už spomenuli vyššie, v tejto práci sa budeme venovať modelovaniu reziduálnej zložky.

## 3 Prehľad modelov

### 3.1 Lineárne modely - jednorozmerné

V tejto podkapitole sa zameriame na lineárne modely, popíšeme ich základné vlastnosti, ich výhody a nevýhody a príklady použitia.

Po dekompozícii časového radu sme z neho odstránili všetky deterministické zložky, t.j. trendovú, cyklickú, sezónnu, prípadne aj iné zložky. Ostal nám teda časový rad v tvare

$$y_t = \mathcal{E}_t, \quad (3.1)$$

čo znamená, že môžeme modelovať stochatickú časť časového radu.

Pre jednoduchosť budeme pod pojmom stacionarita časového radu uvažovať *slabú stacionaritu*, t.j. príslušný časový rad musí mať konštantnú strednú hodnotu a rozptyl, nakoľko silná stacionarita vyžaduje invariantnosť v čase, t.j. aby  $F(y_{t_1}, y_{t_2}, \dots, y_{t_k}) = F(y_{t_1+h}, y_{t_2+h}, \dots, y_{t_k+h})$ , kde  $F$  je združená distribučná funkcia.

**Definícia 3.1.1.** *Stochastický proces  $\{y_t\}$  nazývame (kovariančne) stacionárny, ak*

$$\begin{aligned} E[y_t] &= m \\ D[y_t] &= \sigma^2 \\ \text{cov}(y_t, y_s) &= \text{cov}(y_{t+k}, y_{s+k}) \quad \forall k, \quad s \neq t \end{aligned}$$

*t.j. kovariančná funkcia závisí len od časovej vzdialenosti náhodných premenných.*

Základným stavebným prvkom pre budovanie tejto teórie sú nezávislé a identicky rozdelené stochastické procesy. (Budeme používať anglické označenie i.i.d). Časový rad je i.i.d, ak sú jeho zložky navzájom nezávislé a všetky pochádzajú z toho istého rozdelenia. Z toho vyplýva, že ak uvažujeme pozorovanie v čase  $T$ , tak hodnota v čase  $T+1$  nie je ovplyvnená predchádzajúcimi hodnotami a bude mať opäť to isté rozdelenie.

Tieto podmienky sú ale častokrát príliš silné, a preto si pre ďalšie účely vystačíme s *bielym šumom*. Na začiatok si definujme *biely šum* [11]:

**Definícia 3.1.2.** *Stochastický proces  $\epsilon_t$ , pre  $t = 1, 2, \dots, n$ , ktorý je tvorený nekorelovanými náhodnými premennými s rovnakým rozdelením pravdepodobnosti sa nazýva biely šum, ak platí, že:*

$$\begin{aligned} E[\epsilon_t] &= m && \text{obvykle } m = 0 \\ D[\epsilon_t] &= \sigma^2 \\ \text{cov}[\epsilon_t \epsilon_s] &= 0 && \forall s \neq t, \end{aligned}$$

kde  $E[\cdot]$  je stredná hodnota,  $D[\cdot]$  je disperzia a  $\text{cov}[\cdot]$  kovariancia.

Pod označením  $\epsilon_t$  budeme v celej diplomovej práci uvažovať proces bieleho šumu.

Každý stacionárny proces  $y_t$ , ktorý neobsahuje deterministickú zložku, môžeme vyjadriť ako lineárnu kombináciu radu nekorelovaných rovnako rozdelených náhodných premenných, ktorá sa označuje ako Waldova reprezentácia (alebo tiež lineárny proces)

$$y_t - m = \epsilon_t + \psi_1 \epsilon_{t-1} + \dots = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = 1.$$

### 3.1.1 MA, AR, ARMA modely

Ako sme si povedali, základným stavebným prvkom lineárnych modelov sú i.i.d časové rady, resp. biele šumy.

**Definícia 3.1.3.** *Uvažujme časový rad  $y_t$ , pre  $t = 1, 2, \dots, n$ . V prípade, že tento časový rad môžeme zapísať ako*

$$y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots + \psi_q \epsilon_{t-q} \tag{3.2}$$

*hovoríme, že časový rad je v tvare MA( $q$ ) procese, teda ide o proces kľzavých priemerov rádu  $q$ .*

Analogicky môžeme vytvoriť časový rad aj pomocou predchádzajúcich hodnôt časového radu  $y_t$ , z čoho dostávame definíciu autoregresného AR( $p$ ) procesu rádu  $p$ .

**Definícia 3.1.4.** Uvažujme časový rad  $y_t$ , pre  $t = 1, 2, \dots, n$ . V prípade, že tento časový rad môžeme zapísať ako

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (3.3)$$

hovoríme, že časový rad je v tvare  $AR(p)$  procesu.

V praxi sa často stretávame so zápisom časového radu pomocou tzv. lag operátora (operátora spätného posunu)  $B$ , definovaného ako

$$By_t = y_{t-1}. \quad (3.4)$$

MA, resp. AR proces potom možno prepísať ako

$$y_t = \Psi(B)\epsilon_t \quad (3.5)$$

$$\Phi(B)y_t = \epsilon_t, \quad (3.6)$$

kde

$$\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots + \psi_q B^q = 1 + \sum_{j=1}^q \psi_j B^j \quad (3.7)$$

je MA polynóm a

$$\phi(B) = 1 - \sum_{j=1}^p \phi_j B^j \quad (3.8)$$

je AR polynóm.

Táto forma zápisu sa využíva pre určenie stacionarity a invertovateľnosti časového radu.

Najpoužívanejšou triedou modelov je kombinácia MA a AR modelov.

**Definícia 3.1.5.** Uvažujme časový rad  $y_t$ , pre  $t = 1, 2, \dots, n$ . V prípade, že tento časový rad môžeme zapísať ako

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots + \psi_q \epsilon_{t-q} \quad (3.9)$$

hovoríme, že časový rad je v tvare  $ARMA(p, q)$  procesu.

Analogicky ako pre AR(p), resp. MA(q) procesy, môžeme tento proces prepísať pomocou lag operátora do tvaru

$$\Phi(B)y_t = \Psi(B)\epsilon_t.$$

kde  $\Psi(B)$ , resp.  $\Phi(B)$  definujeme ako (3.7) a (3.8).

### 3.1.2 Stacionarita a invertovateľnosť

Podľa [6] hovoríme, že časový rad  $\{y_t\}$  je invertovateľný, ak sa dá zapísať v tvare

$$\Pi(B)y_t = \epsilon_t, \tag{3.10}$$

kde

$$\Pi(B) = 1 - \sum_{j=1}^{\infty} \Pi_j B^j. \tag{3.11}$$

Zamerajme sa na porovnanie týchto vlastností pre AR(p), resp. MA(q), resp. ARMA(p,q) procesy.

Proces MA(q) je automaticky stacionárny z definície pre akékoľvek jeho parametre. Postačujúcou podmienkou (dôkaz viď [9]) pre invertovateľnosť tohto časového radu je, že

$$\Pi(B) = \Psi^{-1}(B) \quad \text{konverguje pre } -1 \leq B \leq 1.$$

Táto podmienka je ekvivalentná tvrdeniu, že všetky korene polynómu  $\Psi(B)$  ležia mimo jednotkového kruhu.

Z definície AR(p) procesu vyplýva, že tento proces je invertovateľný bez nutnosti klásť ďalšie požiadavky na tento časový rad. Analogicky ako pre MA(q) proces sa dá dokázať, že proces AR(p) je stacionárny, ak všetky korene polynómu  $\Phi(B)$  ležia mimo jednotkového kruhu.

Pre ARMA modely nie je ani jedna z podmienok splnená automaticky. Ale podmienka pre stacionaritu je rovnaká, ako podmienka pre AR(p) procesy a podmienka pre invertovateľnosť je rovnaká, ako je podmienka pre MA(q) procesy.



**Poznámka 3.1.1.** *Invertovateľný ARMA proces môžeme pre dostatočne veľké  $k$  aproximovať modelom  $AR(k)$  (pozri napr. [6]). Táto vlastnosť sa využíva pri odhade parametrov  $ARMA(p,q)$  modelov, tzv. "Long AR metódou".*

### 3.1.3 Určenie vhodného modelu

Ako sme si už povedali predtým, vybrať správnu triedu modelov a následne z nej vybrať ten najvhodnejší model, nie je veľmi jednoduché. V tejto podkapitole si rozoberieme postup ako identifikovať model a ako následne určiť jeho rád. Nebudeme sa zaoberať porovnávaním jednotlivých modelov, keďže to bude náplňou samostatnej kapitoly o informačných kritériách(ďalej IC).

Na identifikáciu modelu si definujeme autokorelačnú a parciálnu autokorelačnú funkciu  $\rho_k$ , resp.  $\rho_{kk}$ , ako aj ich odhady  $r_k$ , resp.  $r_{kk}$  podľa [6].

**Definícia 3.1.6.** *Uvažujme stacionárny časový rad  $\{y_t\}$ . Potom hodnota jeho autokorelačnej funkcie v bode  $k$  je definovaná ako*

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad k = \dots, -1, 0, 1, \dots, \quad (3.12)$$

kde

$$\gamma_k = \text{cov}(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu), \quad k = \dots, 0, 1, \dots \quad (3.13)$$

$$\mu = E(y_t) \quad (3.14)$$

**Dôsledok 3.1.1.** *Z definície 3.1.6 vyplýva, že odhad výberovej autokorelačnej funkcie pre posunutie  $k$  má potom tvar*

$$r_k = \frac{c_k}{c_0}, \quad k = \dots, -1, 0, 1, \dots, \quad (3.15)$$

kde

$$c_k = \sum_{t=1}^{n-k} \frac{(y_t - \bar{y})(y_{t+k} - \bar{y})}{n}, \quad k = 0, 1, \dots, n-1, \quad (3.16)$$

$$\bar{y} = \sum_{t=1}^n \frac{y_t}{n}. \quad (3.17)$$

Definíciu parciálnej autokorelačnej funkcie možno nájsť v [9] a [6], my sa budeme zaoberať s odhadom jej hodnôt, ktoré majú tvar

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \quad \text{pre } k > 1, \quad (3.18)$$

$$r_{11} = r_1, \quad (3.19)$$

kde

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}, \text{ pre } j = 1, 2, \dots, k-1. \quad (3.20)$$

Z definície 3.1.3 MA(q) procesu vyplýva, že

$$r_{kk} \neq 0 \quad \forall k > 0, \quad (3.21)$$

$$r_k = 0 \quad \forall k > q. \quad (3.22)$$

Analogicky z definície 3.1.4 AR(p) procesu vyplýva, že

$$r_{kk} = 0 \quad \forall k > p, \quad (3.23)$$

$$r_k \neq 0 \quad \forall k > 0. \quad (3.24)$$

Najjednoduchšou a zároveň asi najpoužívanejšou metódou je určenie rádu na základe odhadu hodnôt výberovej autokorelačnej funkcie  $r_k$  a parciálnej autokorelačnej funkcie  $r_{kk}$ .

Na určenie rádu MA procesu hľadáme bod  $q_0$  taký, že

$$|r_k| < 2 \sqrt{\frac{1}{n} \left(1 + \sum_{j=1}^{q_0} r_j^2\right)} \quad \forall k > q_0.$$

Na určenie rádu AR procesu hľadáme bod  $p_0$  taký, že

$$|r_{kk}| < 2 \sqrt{\frac{1}{n}} \quad \forall k > p_0.$$

V prípade, že takýto bod  $q_0$ , resp.  $p_0$  existuje, tak potom za vhodných kandidátov na model zvolíme príslušné AR( $p_0$ ), MA( $q_0$ ), resp. ich kombináciu ARMA( $p_0, q_0$ ) a modely navzájom porovnáme. V súčasnosti najrozšírenejšou metódou porovnávaní modelov je porovnanie na základe informačných kritérií, ktoré si bližšie rozoberieme v kapitole 5. Pre naše účely sa zatiaľ obmedzíme na konštatovanie, že informačné kritérium nám charakterizuje vybraný model a jeho minimalizáciou dostávame najlepší možný model z hľadiska schopnosti "fitovať" dáta.

### 3.1.4 Odhad parametrov modelu

Na odhad parametrov sa používa niekoľko metód. Uvedieme si tri:

- Yule-Walkerova metóda
- Long AR metóda
- Hannan-Rissanenova procedúra.

#### Yule-Walkerova metóda

Používa sa na odhad parametrov modelu AR(p) pre dané  $p$ . Využívajúc 3.1.6 dostávame

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} + a_k,$$

kde

$$a_k = \begin{cases} \sigma_\epsilon^2 & \text{ak } k = 0, \\ 0 & \text{inak.} \end{cases} \quad (3.25)$$

Upravou tejto rovnice dostávame sústavu  $(p+1)$  rovníc o  $(p+1)$  neznámych:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \quad \forall k = 1, 2, \dots, p, \quad (3.26)$$

$$\rho_0 = \phi_1 \rho_1 + \phi_2 \rho_2 + \cdots + \phi_p \rho_p + \sigma_\epsilon^2 \quad k = 0. \quad (3.27)$$

Ak namiesto autokorelačnej funkcie  $\rho_k$  použijeme výberovú autokorelačnú funkciu  $r_k$ , riešením tejto sústavy sú príslušné parametre AR(p) modelu  $\phi_1, \phi_2, \dots, \phi_p, \sigma_\epsilon^2$ .

#### Long AR metóda

Používa sa na odhad parametrov invertovateľného modelu ARMA(p,q) pomocou AR(k) modelu (pre dostatočne veľké  $k$ ). Najskôr sa pomocou MNS odhadnú parametre AR(k):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k} + z_t \quad (3.28)$$

Tieto odhady  $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_k$  sa použijú na odhad reziduí  $\hat{z}_t$

$$\hat{z}_t = y_t - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \cdots - \hat{\phi}_k y_{t-k}, \quad (3.29)$$

ktoré sa použijú na odhad parametrov ARMA(p,q) pomocou MNŠ

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \hat{z}_t + \psi_1 \hat{z}_{t-1} + \psi_2 \hat{z}_{t-2} + \dots + \psi_q \hat{z}_{t-q} \quad (3.30)$$

**Hannan-Rissanenova procedúra** Hannan-Rissanenova procedúra je veľmi podobná Long AR metóde. Pozostáva z nasledujúcich krokov:

- Lewinson-Durbinov odhad parametrov modelov AR(1), ... AR( $k_{max}$ ) - upravená YW metóda (kde  $k_{max}$  je rád long AR metódy)
- Výpočet AIC pre všetky modely a výber AR(k) modelu s minimálnym AIC
- Výpočet reziduí vybraného AR(k) modelu
- Odhad parametrov modelov ARMA(p,q) pre  $p \leq \min(k, p_{max})$ , resp.  $q < q_{max}$  pomocou MNŠ
- Výber jedného, alebo niekoľkých najlepších modelov na základe BIC,

kde AIC resp. BIC má tvar

$$AIC = -2\ln(L) + 2K, \quad (3.31)$$

$$BIC = -2\ln(L) + K\ln(n), \quad (3.32)$$

kde  $K$  je počet parametrov,  $L$  je funkcia vierohodnosti (max-likelihood function) a  $n$  je počet pozorovaní. Viac o nich si povieme v kapitole 5.

Na výslednom modeli (alebo modeloch) podrobíme reziduá diagnostickým testom, ako je test nulovosti autokorelačnej funkcie, Portmanteauov test, test normality, znamienkový test, test kritických bodov a mnohé iné (viď [6]). V prípade, že model nie je dostatočne "kvalitný", je nutné celý postup (identifikácia, odhad, diagnostika) opakovať.

### 3.1.5 Bodové predpovede

Ak diagnostika preukáže vhodnosť modelu, okrem popisných účelov môže byť model časového radu použitý na predpovedanie budúcich hodnôt. Označme predpovedanú hodnotu v čase  $t+k$  ako  $\hat{y}_{t+k}$ . Pojmom predikčná chyba (chyba predpovede) budeme označovať výraz

$$e_{t+k} = y_{t+k} - \hat{y}_{t+k}.$$

Poznamenajme, že existuje viacero spôsobov ako odhadovať  $\hat{y}_{t+k}$ . Preto sa ukazuje ako najvhodnejšie odhadovať  $\hat{y}_{t+k}$  pomocou minimalizácie štvorca predikčnej chyby (SPE z anglického squared prediction error)

$$SPE(k) = E[e_{t+k}^2].$$

V praxi sa často používajú aj iné kritéria, ako napríklad RMSE (root mean squared error) a MAE (mean absolute error). Sú založené na priemere  $m$   $k$ -krokových predpovedí urobených na modeli zostavenom na  $n$  pozorovaniach. Podrobnejšie sa o nich píše napr. v ([3], [9]).

$$RMSE = \sqrt{\frac{1}{m} \sum_{j=1}^m (\hat{y}_{n+j|n+j-k} - y_{n+j})^2}, \quad (3.33)$$

$$MAE = \frac{1}{m} \sum_{j=1}^m |\hat{y}_{n+j|n+j-k} - y_{n+j}|, \quad (3.34)$$

kde  $\hat{y}_{n+j|n+j-k}$  znamená  $k$ -krokovú predpoveď v čase  $n+j$ , urobenú na základe informácií do času  $n+j-k$ .

### 3.1.6 Výhody/Nevýhody lineárnych modelov a príklady použitia

Lineárne jednorozmerné modely poskytujú relatívne slušne výsledky a ich veľkou výhodou je mnohokrát veľmi ľahká interpretácia výsledkov, ako aj ich celkové porovnanie s realitou. Spravidla sa ako vhodné modely uvádzajú len modely, kde  $p+q < 4$ , nanajvýš 10, lebo pri príliš vysokom počte vysvetľujúcich premenných stúpa časová náročnosť výpočtu, ako aj neschopnosti interpretovať výsledky. Vyššie rády AR modelov sa používajú napr. v geodézii, v ekonómii na modelovanie výmenných kurzov (kde je často rád modelu

7 a viac) a podobne. Vyššie rády MA modelov signalizujú silnú cyklickú zložku.

Hlavnou nevýhodou týchto modelov je paradoxne ich jednoduchosť. Veľmi zle tieto modely predpovedajú budúce hodnoty, hlavne pokiaľ ide o reálne dáta, ktoré sú častokrát závislé od iných premenných, prípadne sami o sebe majú špecifický pôvod a charakter.

Najlepšie sa tieto modely uplatňujú na jednoduchých dátach, ktoré nevykazujú veľkú stochastickosť a nemajú veľké a nečakané štrukturálne zmeny. Podmienka stacionarity je nevyhnutná. V prípade, že časový rad nie je stacionárny, použitie lineárnych modelov je bezvýznamné a nevhodné. Určite nie sú vhodné napríklad ani na modelovanie výmenných kurzov, ktoré majú premenlivú disperziu (volatilitu). Slúžia skôr ako základ pre pokročilejšie metódy a lepšie poznanie dát, s ktorými sa pracuje.

## 3.2 Nelineárne modely

Je zrejmé, že lineárne modely nie sú vo všeobecnosti vždy vhodné na použitie. Častokrát sa stretávame s časovými radmi, ktoré vykazujú silné nelineárne črty, a preto je použitie klasických lineárnych modelov nedostatočné. Do pozornosti sa preto dostávajú nelineárne modely, reprezentované napr. bilineárnymi modelmi, neurónovými sieťami a viacrežimovými modelmi. Je nutné si klásť prirodzený predpoklad, že dané modely musia byť okrem schopnosti dobre "fitovať" dáta aj interpretovateľné. To sa však ukazuje v prípade neurónových sietí častokrát zložité, ak nie priam nemožné.

Matematicky možno nelineárne modely zapísať v tvare podľa [11]

$$y_t = F(\mathbf{X}_t, \boldsymbol{\theta}) + \epsilon_t \quad (3.35)$$

kde  $\mathbf{X}_t = (1, y_{t-1}, \dots, y_{t-p})$ ,  $\boldsymbol{\theta}$  je vektor parametrov modelu a tzv. skelet  $F(\mathbf{X}_t, \boldsymbol{\theta})$  je nelineárna funkcia, ktorá je minimálne dvakrát spojitely diferencovateľná.

V ďalšej práci sa zameriame na triedu viacrežimových modelov, ktoré sú dobre interpretovateľné a ako sa neskôr v aplikáciach ukáže, sú aj veľmi dobré na modelovanie reálnych dát. Základný prvok týchto modelov je ich

"riadenie" pomocou jednej, alebo viacerých premenných. V prípade, že tieto premenné nie sú pozorovateľné, t.j. ich hodnoty nie sú priamo odvoditeľné, hovoríme, že ide o viacrežimové modely Markovovho typu [10], [21]. V tejto práci sa budeme zaoberať len modelmi s režimami určenými pozorovateľnými premennými.

Zostavíme si podobnú schému ako v kapitole 1. Začínajúc lineárnymi modelmi prejdeme k nelineárnym modelom v prípade, že použitie lineárnych modelov sa ukazuje ako nedostatočné a časový rad vykazuje nelineárny charakter. Zostavíme vhodný model, odhadneme jeho parametre, otestujeme testami náhodnosti a podľa potreby modifikujeme. Nakoniec urobíme na základe modelu popis procesu, alebo predpovedáme budúce hodnoty.

### 3.2.1 Úvod k viacrežimovým modelom

Ako sme si už povedali vyššie, základom viacrežimových nelineárnych modelov sú, ako už názov prezrádza, rôzne režimy (stavy) modelu. Tieto režimy sa menia na základe pozorovateľných premenných (jednej alebo viacerých). Zmena stavu môže nastať v ktoromkoľvek okamihu časového radu. Z pochopiteľných dôvodov je zrejmé, že takto definovaný model má premenlivú strednú hodnotu, rozptyl a autokoreláciu, v závislosti na momentálnom stave. Jednoduchým príkladom viacrežimového modelu môže byť meranie teploty v závislosti od ročného obdobia, teda máme definované správanie v 4 rôznych stavoch.

Vzhľadom na to, že trieda viacrežimových modelov je veľmi široká, v našej práci sa obmedzíme len na viacrežimové modely s pozorovateľnými premennými, kde každý stav môžeme vyjadriť ako AR proces.

### 3.2.2 TAR modely

Threshold Autoregressive procesy, tzv. TAR procesy sú základom viacrežimových modelov. Navrhol ich v 80-tych rokoch Tong (viď [24], [23]). Dynamika modelu je zabezpečená tzv. threshold premennou  $z_t$ , ktorá je porovnávaná s referenčnou hodnotou  $r$ . Podľa [24], dvojrežimový TAR proces s  $AR(p)$  v oboch režimoch má tvar

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \phi_{2,1}y_{t-2} + \cdots + \phi_{p,1}y_{t-p} + \epsilon_t & \text{ak } z_t \leq r \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \phi_{2,2}y_{t-2} + \cdots + \phi_{p,2}y_{t-p} + \epsilon_t & \text{ak } z_t > r. \end{cases} \quad (3.36)$$

V literatúre sa často stretávame s maticovým tvarom

$$y_t = \Phi_1' \mathbf{X}_t (1 - I[z_t > r]) + \Phi_2' \mathbf{X}_t I[z_t > r], \quad (3.37)$$

kde  $\Phi_k = (\phi_{0,k}, \phi_{1,k}, \dots, \phi_{p,k})'$  je vektor neznámych parametrov k-teho procesu,  $\mathbf{X}_t = (1, y_{t-1}, \dots, y_{t-p})'$  a  $I[A]$  je indikačná funkcia nadobúdajúca 1 v prípade nastania udalosti A a 0 inak.

V prípade, že za premennú  $z_t$  budeme uvažovať predchádzajúcu hodnotu  $y_{t-d}$ , pre  $d > 0$ , hovoríme o tzv. **SETAR modeloch**, z anglického Self-Exciting TAR modeloch. Pre ne platí

$$y_t = \Phi_1' \mathbf{X}_t (1 - I[y_{t-d} > r]) + \Phi_2' \mathbf{X}_t I[y_{t-d} > r] + \epsilon_t,$$

teda skelet

$$F(\mathbf{X}, \boldsymbol{\theta}) = \Phi_1' \mathbf{X}_t (1 - I[y_{t-d} > r]) + \Phi_2' \mathbf{X}_t I[y_{t-d} > r]$$

pre  $\boldsymbol{\theta} = (\Phi_1, \Phi_2, r, d)$ .



## 4 STAR modely

Ako sme si už povedali vyššie, predmetom tejto diplomovej práce budú STAR modely. Preto sa pri nich bližšie pristavíme. Definujme si ich, vysvetlíme si pojem prechodovej funkcie a zhrnieme základné vlastnosti týchto modelov.

### 4.1 Definícia STAR modelov

Základom pre STAR modely sú TAR modely. Ich veľkou nevýhodou je, že indikačná funkcia  $I[A]$  je funkcia nespojitá, nadobúdajúca len hodnoty 0 a 1. Tým pádom aj prepínanie medzi jednotlivými AR procesmi je neprirodzené. Jediným východiskom, ako urobiť tento prechod "hladkým", je zameniť indikačnú funkciu hladkou funkciou  $g(z_t)$ , resp.  $g(y)$  v špeciálnom prípade.

Potom dostávame jednoduchý model vychádzajúci z [11], upravený do tvaru podľa ([3], [10])

$$y_t = \Phi_1' \mathbf{X}_t (1 - G(z_t, r, \gamma)) + \Phi_2' \mathbf{X}_t G(z_t, r, \gamma) + \epsilon_t, \quad (4.1)$$

kde  $\Phi_k = (\phi_{0,k}, \phi_{1,k}, \dots, \phi_{p,k})'$  je vektor neznámych parametrov k-teho procesu,  $\mathbf{X}_t = (1, y_{t-1}, \dots, y_{t-p})'$  a  $G(z_t, r, \gamma)$  je tzv. *prechodová funkcia* (hladká funkcia, ktorá s rastúcou hodnotou  $z_t$  prechádza plynule od 0 do 1).

Pre naše účely si tento model ešte preformulujeme do tvaru

$$y_t = \Phi_1' \mathbf{X}_t G(z_t, r, \gamma) + (\Phi_2 - \Phi_1)' \mathbf{X}_t G(z_t, r, \gamma) + \epsilon_t, \quad (4.2)$$

Podobne definujme STAR model s  $m$  režimami a  $m - 1$  prechodovými funkciami ako

$$\begin{aligned} y_t = & \Phi_1' \mathbf{X}_t G(z_t, r, \gamma) + (\Phi_2 - \Phi_1)' \mathbf{X}_t G(z_t, r_1, \gamma_1) + \dots \\ & + (\Phi_m - \Phi_{m-1})' \mathbf{X}_t G(z_t, r_{m-1}, \gamma_{m-1}) + \epsilon_t, \quad r < r_1 < \dots < r_{m-1} \end{aligned} \quad (4.3)$$

### 4.2 Prechodová funkcia

Zamerajme sa bližšie na prechodovú funkciu  $G(z_t, r, \gamma)$ . Z definície STAR modelu vieme, že táto funkcia nadobúda hodnoty medzi 0 a 1. Z 4.1 vyplýva, že v prípade hraničných hodnôt, t.j. pre  $G(z_t, r, \gamma) = 0$ , resp.  $G(z_t, r, \gamma) = 1$ , dostávame príslušný klasický AR model.

Najčastejšie sa za prechodovú funkciu volí logistická funkcia prvého rádu, ktorá sa pre svoje dobré vlastnosti hojne využíva aj vo finančnom a ekonomickom sektore. Prechodová funkcia má potom tvar

$$G(z_t, r, \gamma) = \frac{1}{1 + \exp^{-\gamma(z_t - r)}} \quad \gamma > 0, \quad (4.4)$$

kde parameter  $\gamma$  sa nazýva *parameter vyhladzovania*. Takto definovaný STAR model sa nazýva **LSTAR model** (logistic STAR model).

Poznamenajme, že v prípade  $\gamma = 0$  je hodnota  $G(z_t, r, \gamma) = \frac{1}{2}$ , z čoho dostávame AR model. Analogicky pre  $\gamma \rightarrow \infty$  dostávame TAR model. Parameter  $\gamma \rightarrow 0$  má funkciu vyhladzovania, nakoľko pre veľké hodnoty  $\gamma$  je hodnota výrazu  $1 + \exp^{-\gamma(z_t - r)}$  buď veľmi blízka k 0, alebo veľmi veľká, v závislosti od znamienka výrazu  $(z_t - r)$ , čo indikuje, že prechod medzi jednotlivými režimami je takmer okamžitý a funkciu  $G(z_t, r, \gamma)$  možno stotožniť s indikačnou funkciou  $I[A]$ .

Parameter  $r$  hrá úlohu "threshold" premennej, t.j. prahu medzi dvomi režimami, nakoľko v prípade, že  $z_t = r$ , dostávame  $G(z_t, r, \gamma) = \frac{1}{2}$

**Iné prechodové funkcie** Za prechodovú funkciu si môžeme zvoliť aj iné funkcie. V oblasti ekonómie sa veľmi často používa aj *exponenciálna funkcia*, ktorá má veľmi dobré využitie v prípade absolútnych hodnôt premennej  $z_t$ . Má tvar

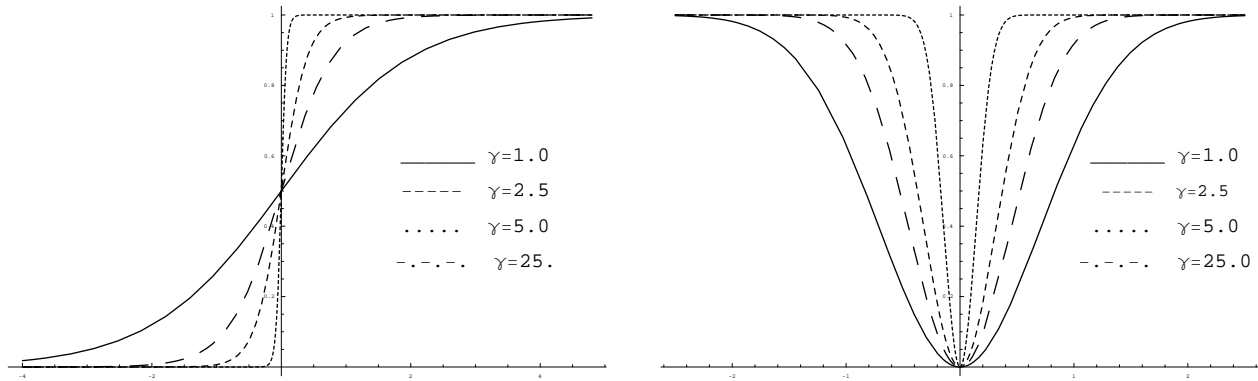
$$G(z_t, r, \gamma) = 1 - \exp^{-\gamma(z_t - r)^2} \quad \gamma > 0. \quad (4.5)$$

Z jej definície vyplýva, že  $G(z_t, r, \gamma) \rightarrow 1$  pre  $z_t \rightarrow \infty$ , ako aj pre  $z_t \rightarrow -\infty$ . V prípade, že  $z_t = r$  je  $G(z_t, r, \gamma) = 0$ .

Hojne využívanou sa stala aj *logistická funkcia* druhého, resp. vyššieho rádu. Tá zodpovedá LSTAR modelu s tromi, resp. viacerými režimami. Prechodová funkcia má potom tvar

$$G(z_t, \mathbf{r}, \gamma) = \frac{1}{1 + \exp^{-\gamma(z_t - r_1)(z_t - r_2)}} \quad r_1 < r_2, \quad \gamma > 0, \quad (4.6)$$

kde  $\mathbf{r} = (r_1, r_2)$ .



Obrázok 1: Porovnanie logistických (vľavo) a exponenciálnych funkcií (vpravo) pre  $r = 0$  a rôzne hodnoty parametra  $\gamma$

V našej ďalšej práci sa budeme zaoberať prevažne LSTAR a ESTAR modelmi, ktoré vďaka svojej prechodovej funkcii poskytujú široké možnosti využitia a je silný predpoklad na ich dobré modelovacie schopnosti na reálnych dátach.

### 4.3 Výstavba STAR modelov

V tejto podkapitole sa budeme venovať výstavbe STAR modelov a načrtneme si odhad parametrov. Postup výstavby STAR modelov sa riadi princípom, "od špeciálneho ku všeobecnému". Preto je vhodné začať s jednoduchšími lineárnymi modelmi a až potom, keď nie sú splnené podmienky pre ich použitie, prejsť k nelineárnym modelom. Postup možno zhrnúť do nasledovnej schémy [11]

- Špecifikácia príslušného AR modelu na základe IC
- Test linearity oproti nelinearite viacrežimových modelov
- Odhad parametrov modelu
- Overenie modelu
- Použitie modelu na popis, alebo predikciu hodnôt

Viac o jednotlivých krokoch možno nájsť v [10], [8].

**AR model** Modelovanie procesov STAR modelmi má nepochybne oproti lineárnym modelom niekoľko výhod. Keďže každý proces je špecifický, je dobré začať modelovať všetky procesy najskôr lineárnymi AR(p) modelmi.

Na stanovenie parametra  $p$  odporúčame vybrať niekoľko alternatívnych modelov a vhodný model určiť na základe AIC, alebo BIC. Bližšie si o nich povieme v ďalšej kapitole.

**Test linearity** Druhou dôležitou časťou je test linearity, ktorý nám preukáže, či je alebo nie je nutné modelovať proces pomocou STAR modelov. Najčastejšie sa linearita testuje pomocou Lagrangeových multiplikátorov, tzv. LM test. Bližší popis možno nájsť aj v [3],[10], [8].

**Odhad parametrov** V prípade, že linearita bola zamietnutá, použijeme na modelovanie procesu STAR modely. Na začiatok začíname s dvojrežimovým STAR modelom v tvare 4.2. Na určenie parametrov  $\hat{\theta} = (\Phi_1', \Phi_2', \gamma, r)$  môžeme použiť nelineárnu metódu najmenších štvorcov

$$\hat{\theta} = \arg \min_{\theta} \sum_{t=0}^n [y_t - F(\mathbf{X}_t; \theta)]^2, \quad (4.7)$$

kde

$$F(\mathbf{X}_t; \theta) = \Phi_1' \mathbf{X}_t (1 - G(z_t, r, \gamma)) + \Phi_2' \mathbf{X}_t G(z_t, r, \gamma) \quad (4.8)$$

**Overenie modelu** Po použití STAR modelu na príslušný proces nám ostane reziduálna zložka, ktorú podrobíme testom autokorelácie, zvyškovej nelinearity (otestujeme dvojrežimový model oproti trojrežimovému), prípadne otestujeme normalitu alebo použijeme iné testy [10], [8]. Tým ukážeme správnosť nášho modelu a môžeme použiť model na popis, resp. predikciu hodnôt

**Predikcia hodnôt** Pre jednoduchosť uvažujme model v zjednodušenej verzii v tvare

$$y_t = F(y_{t-1}, \theta) + \epsilon_t \quad (4.9)$$

Potom  $k$ -kroková predikcia v čase  $t + k$  z informácií v čase  $t$  je

$$\hat{y}_{t+k|t} = E[y_{t+k} | \Omega_t], \quad (4.10)$$

kde  $\Omega_t$  je informačná množina v čase  $t$ , t.j. poznatky o časovom rade do času  $t$ . V prípade, že  $k = 1$  platí, že  $\hat{y}_{t+1|t} = E[y_{t+1} | \Omega_t] = F(y_t, \boldsymbol{\theta})$ . Vo všeobecnosti nie je jednoduché nájsť predpis pre  $k$ -krokovú predikciu. Jednou z možností je

$$\hat{y}_{t+k|t} = \int_{-\infty}^{\infty} F(\hat{y}_{t-(k-1)|t} + \epsilon, \boldsymbol{\theta}) f(\epsilon) d\epsilon, \quad (4.11)$$

kde  $f(\cdot)$  je hustota  $\epsilon_t$ .

Vyrábať takéto zložité integrály je častokrát priam nemožné, a preto sa v praxi používajú hlavne dve metódy - Bootstrap metóda a Monte Carlo metóda.

**Monte Carlo metóda** Pre  $k$ -krokovú predikciu dostávame

$$\hat{y}_{t+k|t} = \frac{1}{m} \sum_{i=0}^m F(y_{t+(k-1)|t} + \epsilon_i, \boldsymbol{\theta}), \quad (4.12)$$

kde  $m$  je dostatočne veľké číslo a  $\epsilon_i$  sú náhodné čísla získané z predpokladaného rozdelenia  $\epsilon_{t+(k-1)}$ .

**Bootstrap metóda** Bootstrap metóda je analogická ako Monte Carlo metóda, jediným rozdielom je, že používame reziduá odhadnuté z modelu. Pre  $k$ -krokovú predikciu teda dostávame

$$\hat{y}_{t+k|t} = \frac{1}{m} \sum_{i=0}^m F(y_{t+(k-1)|t} + \hat{\epsilon}_i, \boldsymbol{\theta}), \quad (4.13)$$

Výhodou Bootstrap metódy oproti Monte Carlo metóde je, že nepotrebujeme žiadny predpoklad na rozdelenie  $\epsilon_{t+(k-1)}$ .

## 5 Informačné kritéria

### 5.1 Úvod

V predchádzajúcich kapitolách sme si povedali niečo o triedach modelov, vysvetlili sme si rozdiely medzi modelmi a spomenuli odhadovanie ich parametrov. Viackrát sme narazili na problém porovnávania modelov medzi sebou. Je lepší AR(2) model, alebo ARMA (2,1), prípadne STAR model?

Odpoveďou na túto otázku a zároveň odpoveďou na druhý problém zo schémy v kapitole 1 sú informačné kritéria. V tejto kapitole si definujeme niektoré základné informačné kritéria, vysvetlíme si rozdiely medzi nimi.

Prírodnou snahou pre výber najvhodnejšieho modelu je minimalizácia rozptylu bieleho šumu  $\sigma_\epsilon^2$ . Nech nami sledovaný časový rad je modelovaný ARMA(p,q) modelom. Potom odhad rozptylu bieleho šumu označíme ako  $\sigma_{p,q}^2$ . Samotná minimalizácia tejto hodnoty častokrát nevedie k správnejmu cieľu, pretože sa preferujú vysokorádové modely. Problémom vysokorádových modelov sme sa už venovali v predchádzajúcich kapitolách. Preto sa za optimálne považuje minimalizácia

$$A_{p,q} = (1 + w_n(p, q))\sigma_{p,q}^2, \quad (5.1)$$

kde  $w_n(p, q)$  je tzv. penalizačná funkcia, ktorá penalizuje vysokorádové modely. Je zrejmé, že funkcia  $w_n(p, q)$  je pre pevné  $n$  rastúca v parametroch  $p$  a  $q$ . Analogicky pre pevné parametre  $p$  a  $q$  penalizačná funkcia konverguje k nule pre  $n \rightarrow \infty$ . Viac o nej možno nájsť v [6].

Je prirodzené predpokladať, že neexistuje model, ktorý by presne vystihol realitu a dokonale by fitoval dáta. Preto sa snažíme vybrať najlepší z modelov. Je zrejmé, že kvalita modelu bude úzko súvisieť s počtom dát. V prípade 6000 pozorovaní dostávame iný model ako pri 60 pozorovaniach. V tejto súvislosti ale narážame na zásadný problém [1]:

**Jednoduchosť / Zložitosť** Čelíme dvom protichodným veciam. Jednoduché modely sú dobre interpretovateľné. Zložité modely naopak dosahujú malú odchýlku, varianciu. Preto treba vždy uvážiť, či zvolíme model s veľkým počtom parametrov a malou varianciou, alebo model s malým počtom parametrov, ale s veľkou varianciou. Tomu problému sa venuje hlavne Foster [7].

V nasledujúcich podkapitolách si prejdeme historický vývin IC, rozdelíme si ich do troch samostatných tried a predstavíme si niekoľko informačných kritérií často používaných, resp. vyvinutých, pre ARMA modely. Tieto informačné kritéria neskôr našli svoj význam aj pre zložitejšie modely.

### 5.1.1 FPE

Historicky prvým kritériom bolo FPE kritérium uvedené Akaikem v roku 1969. Toto kritérium bolo odvodené pre AR(p) modely, neskôr rozšírené pre ARMA(p,q) modely. Má tvar

$$FPE = \sigma^2 \frac{n + p + q}{n - p - q} \quad (5.2)$$

Toto kritérium je však nekonzistentné a bolo neskôr nahradené AIC (Akaike information criterion) a BIC (Bayesian information criterion) používanými dodnes.

IC kritéria vznikli ako prirodzená požiadavka porovnávať kvalitatívne alternatívne modely. Na základe týchto kritérií sa stanoví najvhodnejší model. V zásade môžeme rozdeliť informačné kritéria do troch skupín [2]

- IC založené na MSE, alebo MSPE (vid 3.33)
- IC odvodené od Kullbeck - Leiblerovej informácie
- IC založené na penalizácii počtu parametrov

My sa však budeme venovať prevažne posledným dvom skupinám.

### 5.1.2 Kullback-Leiblerova informácia

Základným pilierom pre IC z druhej skupiny sa stala Kullbeck - Leiblerova informácia, ďalej len K-L informácia, ktorej základ vznikol v roku 1951 [14].

Môžeme ju interpretovať ako "vzdialenosť" medzi "realitou", reprezentovanou hustotou rozdelenia pravdepodobnosti  $f(y)$  a modelom, reprezentovaným pravdepodobnostnou funkciou  $g(y, \theta)$ , kde  $y$  sú dáta a  $\theta$  sú parametre modelu odhadované z dát  $y$ .

Prirodzenou požiadavkou je definovať všeobecne platné kritérium, ktoré nebude mať žiadne obmedzenia a bude s pravdepodobnosťou blízkou 1 vyberať "vhodný" model. Preto môžeme definovať K-L informáciu podľa [18].

$$I(f, g) = \int f(x) \ln \frac{f(x)}{g(x, \theta)} dx \quad (5.3)$$

pre spojité rozdelenie, resp.

$$I(f, g) = \sum_{i=1}^n p_i \ln \frac{p_i}{\phi_i} \quad (5.4)$$

pre diskrétné rozdelenia, kde  $p_i = P(Y_i = y_i)$  je skutočná pravdepodobnosť a  $\phi_i = g(y_i, \theta)$  je pravdepodobnosť daného modelu [2].

Poznamenajme, že v prípade ak  $f \equiv g$  je  $I(f, g) = 0$ , t.j. nie je žiadna informačná strata.

Pretože na výpočet K-L informácie podľa 5.3, resp. 5.4 je potrebné poznať hustotu rozdelenia  $f$  ako aj parametre  $\theta$ , v praxi sa používa minimalizácia odhadu relatívnej hodnoty K-L informácie, ktorú získame nasledovne:

Prepíšeme si

$$\begin{aligned} I(f, g) &= \int f(x) \ln \frac{f(x)}{g(x, \theta)} dx \\ &= \int f(x) \ln f(x) dx - \int f(x) \ln g(x, \theta) dx. \end{aligned}$$

Označme

$$\begin{aligned} E_f[\ln(f(y))] &= \int f(x) \ln f(x) dx \\ E_f[\ln(g(y | \theta))] &= \int f(x) \ln g(x, \theta) dx. \end{aligned}$$

Potom

$$I(f, g) = E_f[\ln(f(y))] - E_f[\ln(g(y | \theta))], \quad (5.5)$$

Prvú strednú hodnotu  $E_f[\ln(f(y))]$  môžeme považovať za konštantu, pretože popisuje stále tú istú realitu a nemení sa v závislosti od použitého modelu. Označíme ju ako neznámu konštantu  $C$ . Následne potom 5.5 dostáva tvar

$$I(f, g) = C - E_f[\ln(g(y | \theta))], \quad (5.6)$$



resp.

$$I(f, g) - C = -E_f[\ln(g(y | \theta))].$$

Na určenie relatívnej vzdialenosti modelu od reality stačí teda odhadnúť len  $E_f[\ln(g(y | \theta))]$ . Treba však poznamenať, že v praxi rozdelenie pravdepodobnosti  $f(y)$  nepoznáme a len ho odhadujeme, rovnako ako parametre modelu  $\theta$ . Podľa toho, akú metódu pritom zvolíme, získame rôzne IC založené na K-L informácií.

## 5.2 IC založené na K-L informácií

Predpokladajme, že neznáme parametre modelu  $\theta$  sú odhadované metódou maximálnej virohodnosti a logaritmus funkcie virohodnosti je

$$\log L(\theta | y) = \ln(g(y, \theta)).$$

Najznámejšie kritérium z tejto skupiny je Akaikeho IC. Akaike navrhol odhadnúť relatívnu K-L informáciu z empirických dát a maximalizácie logaritmu funkcie virohodnosti. Nech  $K$  je počet neznámych parametrov. Potom vhodný odhad relatívnej K-L informácie  $E_f[\ln(g(y | \theta))]$  (alebo presnejšie  $I(f, g) - C$ ) je

$$\ln(L(\hat{\theta} | y)) - K.$$

Akaike potom definoval informačné kritérium (AIC), vynásobením tohto odhadu výrazom  $-2$ .

$$AIC = -2 \ln(L(\hat{\theta} | y)) + 2K, \quad (5.7)$$

je odhad očakávanej relatívnej vzdialenosti medzi modelom a neznámym reálnym mechanizmom, ktorý generuje pozorované dáta.

Pre ARMA(p,q) modely má toto kritérium tvar

$$AIC = \ln(\sigma_{p,q}^2) + 2\frac{p+q}{n}. \quad (5.8)$$

Pre  $m$ -režimové SETAR modely s  $n_j$  prvkami v  $j$ -tom režime  $AR(p_j)$ ,  $j = 1, \dots, m$ ,  $\sum_{j=1}^m n_j = n$  je toto kritérium definované podľa [10]

$$AIC(p_1, \dots, p_m) = \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + 2(p_j + 1)), \quad (5.9)$$

kde  $\hat{\sigma}_j^2, j = 1, \dots, m$  je rozptyl reziduí v  $j$ -tom režime.

Pre STAR modely je problém určiť (vzhľadom ku plynulému prechodu z jedného režimu do druhého) počet prvkov v jednotlivých režimoch. Návrhu AIC pre STAR modely sa budeme venovať v kapitole 5.4.

Akaikeho prístup nie je vždy dostatočný. V prípade, že počet parametrov  $K$  je relatívne veľký oproti  $n$  (ak podiel  $\frac{n}{K} \geq 40$ ), je nutné upraviť AIC do tvaru

$$AIC_c = -2\ln(L) + 2K + \frac{2K(K+1)}{n-K-1}. \quad (5.10)$$

V praxi sa veľmi často stáva aj to, že chceme porovnávať veľa modelov rôznych tried, aby sme vedeli určiť ten najvhodnejší. Práve tu sa ukazuje, že všetky informačné kritéria AIC-typu sú efektívne len pre tzv. "nested" modely, t.j. jeden model je len rozšírením/zúžením toho druhého, v zmysle členov. Podrobnejšie viď [19].

### 5.3 IC založené na penalizácii počtu parametrov

Do tejto skupiny patria hlavne BIC, SIC, HQIC.

Ako už bolo naznačené vyššie, druhý člen v AIC penalizuje modely vzhľadom na počet parametrov. V nasledujúcich podkapitolách si predstavíme IC, ktoré sa od AIC líšia práve v druhom člene, t.j. v penalizačnom člene. Ich snahou je hlavne sa zamerať na konzistenciu (asymptotickú nevychýlenosť, rozptyl idúci k 0) odhadu pre  $K$ . V neposlednom rade je snaha o výber takého modelu, ktorý pre veľkú testovaciu vzorku, t.j.  $n \rightarrow \infty$ , vyberie správny model s pravdepodobnosťou 1.

#### 5.3.1 Bayesovské informačné kritérium - BIC

Hlavnou nevýhodou AIC je jeho asymptotická nevychýlenosť a bolo dokázané [17], že AIC má tendenciu vyberať preparametrizované modely.

Prvým, v súčasnosti najpoužívanejším kritériom, je BIC, odvodené na základe tzv. Bayesovského prístupu, ktorý sa oproti AIC líši iba tým, že penalizuje rád modelu logaritmom počtu pozorovaní. BIC kritérium má potom vo všeobecnosti tvar

$$BIC = -2 \ln(L(\hat{\theta} | y)) + K \ln(n) \quad (5.11)$$

V prípade, že proces je modelovaný ARMA(p,q) má toto kritérium tvar

$$BIC = -2 \ln(\sigma_{k,l}^2) + \frac{p+q}{n} \ln(n). \quad (5.12)$$

Pre  $m$ -režimové SETAR modely s  $n_j$  prvkami v  $j$ -tom režime  $AR(p_j)$ ,  $j = 1, \dots, m$ ,  $\sum_{j=1}^m n_j = n$  je toto kritérium definované analogicky ako 5.9 podľa [10]

$$BIC(p_1, \dots, p_m) = \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + 2(p_j + 1) \ln n_j), \quad (5.13)$$

kde  $\hat{\sigma}_j^2, j = 1, \dots, m$  je rozptyl reziduí v  $j$ -tom režime.

BIC kritérium odstránilo nedostatok AIC a dodnes sa používa hlavne pre svoju jednoduchosť a ľahkú spočítateľnosť. Výhodou tohto kritéria je, že pravdepodobnosť výskytu realizácie časového radu, pre ktorú by BIC neurčilo správne rád modelu, je asymptoticky nulová. Z toho vyplýva, že BIC je silne konzistentný odhad.

### 5.3.2 DIC, HQIC

Ako sa neskôr ukázalo, hodnoty AIC, resp. BIC sú determinované počtom parametrov  $K$ . Tento nedostatok odstránilo zovšeobecnenie, nazývané aj DIC (deviance information criterion) v tvare [15]

$$DIC = p_D + \bar{D}, \quad (5.14)$$

kde  $p_D = \bar{D} - D(\tilde{\theta})$ ,  $\tilde{\theta} = E[\theta | y]$ ,  $D(\tilde{\theta}) = -2 \ln(L)$  a  $\bar{D} = E[D(\theta | y)]$ .

Výhodou tohto kritéria je, že zohľadňuje schopnosť "fitovať" dáta, ako aj rozmery modelu.

Komplexnejším kritériom sa ukázalo HQ (Hannan-Quinn) informačné kritérium, ktoré vykazuje silnú konzistenciu pre konštantu  $c > 2$  pre ARMA modely. HQIC má tvar

$$HQIC = \ln(\sigma_{p,q}^2) + 2c \frac{p+q}{n} \ln(\ln(n)). \quad (5.15)$$

Otázkou zostáva, či je lepšie používať AIC alebo BIC. Odpoveď nie je jednoznačná, keďže obidve IC, podľa Reschenhofera, v princípe odpovedajú na iné otázky.

## 5.4 IC pre STAR modely

Ako už bolo spomenuté vyššie, najväčším problémom pre určenie informačných kritérií pre STAR modely je ich hladký prechod od jedného režimu k druhému. Častokrát sa preto autori vedeckých prác zameriavajú skôr na odhad parametrov a celkovú výstavbu STAR modelov, ako na informačné kritéria. Za všetky spomeňme aspoň [16]. V týchto prácach sa používa AIC, resp. BIC v štandardnom tvare.

Naším cieľom ale bude zamerať sa práve na informačné kritéria. Podľa [20] máme všeobecný tvar informačného kritéria a navrhne niekoľko typov penalizačných členov tohto kritéria, ktoré následne overíme na reálnych dátach a podporíme matematickými vetami.

### 5.4.1 Všeobecný tvar IC pre STAR modely

Uvažujme  $m$ -režimový STAR model podľa definície 4.3 s  $n$  pozorovaniami a  $K$  parametrami. Potom všeobecný tvar informačného kritéria je

$$Q_n(m) = \max_{r_1, \dots, r_m} \ln\left(\frac{\sigma^2}{\sigma^2(r_1, \dots, r_m)}\right) - Km \frac{\lambda_n}{n}, \quad (5.16)$$

kde  $\sigma^2$  je variancia reziduí lineárneho modelu,  $\sigma^2(r_1, \dots, r_m)$  je variancia reziduí STAR modelu s  $m$  threshold premennými a  $\lambda_n$  je neznáma funkcia.

V prípade  $\lambda_n = 2$  dostávame AIC. V prípade  $\lambda_n = \ln(n)$  dostávame BIC.

### 5.4.2 Nové informačné kritéria pre STAR modely

V tejto podkapitole navrhne nové informačné kritéria pomocou neznámej funkcie  $\lambda_n$ .

Z prvom rade modifikujeme všeobecne známe informačné kritéria pre účely STAR modelov.

$$\begin{aligned}
AIC(p_1, \dots, p_m) &= \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + 2(p_j + 1)) \\
BICC(p_1, \dots, p_m) &= \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + (p_j + 1) \ln(n_j)) \\
HQIC(p_1, \dots, p_m) &= \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + 2(p_j + 1) \ln(\ln(n_j))),
\end{aligned}$$

kde podľa [3]  $n_j = \sum_{t=1}^n \Delta G_{j,t}$ ,  $\Delta G_{j,t} = G_{j-1,t} - G_{j,t}$ , pričom  $G_{j,t}$  je prechodová funkcia j-tého režimu. Navyše  $G_{0,t} = 1$  a  $G_{m,t} = 0$ .

Vzhľadom na to, že AIC je nekonzistentný odhad, nebudeme sa pokúšať modifikovať informačné kritéria penalizačným členom v tvare konštanty. Napriek tomu ho budeme v aplikácií používať pre jeho rozšírenosť a popularitu. Zameriame sa hlavne na BIC, keďže BIC je všeobecne uznávané a používané kritérium. Budeme sa snažiť jeho hlavnú zložku, t.j.  $\ln(n_j)$ , nahradiť podobným výrazom. Tento výraz ohraničíme pomocou  $\sqrt{n_j}$  zhora, resp.  $n_j^{\frac{1}{4}}$  zdola. Takto dostávame dve nové informačné kritéria v tvare

$$\begin{aligned}
OIC(p_1, \dots, p_m) &= \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + (p_j + 1) n_j^{\frac{1}{4}}) \\
SIC(p_1, \dots, p_m) &= \sum_{j=1}^m (n_j \ln \hat{\sigma}_j^2 + (p_j + 1) \sqrt{n_j}).
\end{aligned}$$

## 6 Konzistentnosť informačných kritérií

V tejto kapitole si zhrnieme doterajšie poznatky z literatúry o konzistentnosti informačných kritérií pre jednotlivé typy modelov a pokúsime sa načrtnúť možnosti ich rozšírenia na STAR modely.

Uvedieme definície konzistentnosti podľa [22].

**Definícia 6.0.1.** *Odhad  $\hat{C}_{n,k}$  parametra  $C_{n,k}$  na základe dát  $y(n)$  sa nazýva konzistentný (alebo aj slabó konzistentný), ak pravdepodobnosť, že odhad sa rovná skutočnej hodnote  $\lambda_0$ , konverguje k 1 pre  $n$  idúce do nekonečna.*

$$P(\hat{C}_{n,k} = \lambda_0) \rightarrow 1 \quad \text{ak } n \rightarrow \infty \quad (6.1)$$

**Definícia 6.0.2.** *Odhad  $\hat{C}_{n,k}$  sa nazýva silne konzistentný, ak sa odhad rovná skutočnej hodnote  $\lambda_0$  skoro všade pre  $n$  idúce do nekonečna.*

$$\hat{C}_{n,k} = \lambda_0, \text{ alebo skoro všade pre } n \rightarrow \infty \quad (6.2)$$

Odhad parametra  $C_{n,k}$  je iný pre každú triedu modelov, a preto aj nutné a postačujúce podmienky na slabú, resp. silnú konzistenciu sú rôzne pre tieto triedy modelov. Zhrňme si historický vývoj tried modelov a uvedieme si nutné a postačujúce podmienky pre ne.

### 6.1 Lineárne modely

Historicky najstaršia trieda modelov je veľmi dobre zmapovaná z hľadiska konzistentnosti odhadov. Už Hannan sa v roku 1980 zaoberal konzistenciou odhadov. V [12] dokázal silnú konzistentnosť BIC pre ARMA modely, ako aj pre AR modely ako také.

Vo všeobecnosti pre informačné kritérium v tvare  $-L_n(\hat{\theta}) + C_{n,k}$  sú nutné a postačujúce podmienky pre slabú konzistentnosť odhadu

$$C_{n,k} \rightarrow \infty \quad \text{pre } n \rightarrow \infty \quad \text{a} \quad \lim_{n \rightarrow \infty} \frac{C_{n,k}}{n} = 0 \quad (6.3)$$

Sin a White [25] rozšírili tieto poznatky na všeobecné lineárne a nelineárne modely za predpokladu, že skutočný model sa nachádza medzi uvažovanými modelmi.

V oblasti silnej konzistencie dokázali, že odhad parametra je silne konzistentný za predpokladu 6.3 a podmienky, že penalizačný člen ide k nekonečnu rýchlejšie ako výraz  $\ln(\ln(n))$ .

## 6.2 SETAR modely

S potrebou aplikácie viacrežimových modelov vyvstala otázka nutných a postačujúcich podmienok pre nelineárne modely. Z triedy nelineárnych modelov sa vývoj zamerl hlavne na najjednoduchšiu triedu - triedu SETAR modelov.

Nutné a postačujúce podmienky pre SETAR modely definoval Kapetanios [13], ktorý rozšíril prácu svojich predchodcov. Ukazuje sa, že pri uvažovaní niekoľkých dodatočných predpokladov dostávame podobné podmienky ako pre lineárne modely. Uvedieme si výsledky práce Kapetaniosa.

Pre jednoduchosť uvažujme SETAR model s 2 režimami v tvare

$$y_t = \Phi_1 \mathbf{X}_t (1 - I[y_{t-d} > r]) + \Phi_2 \mathbf{X}_t I[y_{t-d} > r] + \epsilon_t, \quad (6.4)$$

**Predpoklad 6.2.1.** *Uvažujme proces  $y_t$  modelovaný SETAR modelom 6.4. Potom*

$$(F(\mathbf{X}, \theta) - y_t) \rightarrow_p 0, \quad (6.5)$$

kde  $F(\mathbf{X}, \theta) = \Phi_1 \mathbf{X}_t (1 - I[y_{t-d} > r]) + \Phi_2 \mathbf{X}_t I[y_{t-d} > r]$ .

**Predpoklad 6.2.2.**  $\epsilon_t$  sú absolútne spojité s rovnomerne spojitou, kladnou hustotou rozdelenia a konečným štvrtým momentom.

**Predpoklad 6.2.3.**  $y_t$  sú stacionárne s konečným štvrtým momentom.

**Predpoklad 6.2.4.** Autoregresná funkcia je nespojitá.

**Predpoklad 6.2.5.** Pre  $p < d$ , odhad parametra  $r$ , t.j.  $\hat{r}$ , konverguje skoro všade k tej istej konštante  $r^*$ .

**Predpoklad 6.2.6.** Pre dané  $r$  a  $p$  platí, že  $|C_{n,k} - (C_{n_1,k_1} + C_{n_2,k_2})| < C$  skoro všade, pre všetky  $p$ , kde  $C$  je kladná konštanta a  $C_{n_j,k_j}$  je penalizačný člen pre  $j$ -ty režim. Poznamenajme, že  $n_j$  je počet dát pre  $j$ -ty režim. Tvrdenie platí aj pre dané  $n$  a  $k$  a rozdielne  $r$ .

Ak sú splnené predpoklady 6.2.1 - 6.2.6, potom platia nasledovné tvrdenia:

**Veta 6.2.1.** *Uvažujme proces  $y_t$  modelovaný SETAR modelom. Potom je odhad dosiahnutý pomocou informačného kritéria s penalizačným členom  $C_{n,k}$  slabo konzistentný práve vtedy, keď platí*

$$\begin{aligned} C_{n,k} &\rightarrow_p \infty \\ \frac{C_{n,k}}{n} &\rightarrow_p 0 \\ C_{n,k_1} - C_{n,k_2} &\rightarrow_p \infty \quad \text{ak } k_1 > k_2. \end{aligned}$$

**Veta 6.2.2.** *Uvažujme proces  $y_t$  modelovaný SETAR modelom. Potom je odhad dosiahnutý pomocou informačného kritéria s penalizačným členom  $C_{n,k}$  silne konzistentný práve vtedy, keď platí*

$$\begin{aligned} \frac{C_{n,k}}{\ln(\ln(n))} &\rightarrow_{s.v.} C \quad k < C \leq \infty \\ \frac{C_{n,k}}{n} &\rightarrow_{s.v.} 0 \\ \frac{C_{n,k_1} - C_{n,k_2}}{\ln(\ln(n))} &\rightarrow_{s.v.} C \quad k < C \leq \infty \quad \text{ak } k_1 > k_2. \end{aligned}$$

Dôkazy týchto viet možno nájsť priamo v [13].

### 6.3 STAR modely

V našej diplomovej práci sa zaoberáme STAR modelmi, a preto sa pokúsime o nutné, príp. o postačujúce podmienky konzistentnosti IC pre STAR modely. Pokiaľ je nám známe, doposiaľ tieto podmienky nikdy neboli sformulované a matematicky dokázané.

Výrazný rozdiel, ktorý komplikuje použitie analógie dôkazov viet pre iné triedy modelov, je plynulý prechod od jedného režimu k druhému, t.j. časový rad sa nechová ako jeden samostatný režim, ale ako "zmes" viacerých režimov, čo zväčšuje nielen počet parametrov, ale i náročnosť dôkazov konzistentnosti odhadov.

Napriek tomuto nedostatku budeme vychádzať z dôkazov pre SETAR modely, keďže STAR modely môžeme považovať pre dostatočne veľké hodnoty parametra  $\gamma$  za SETAR modely. Z tohto dôvodu modifikujeme predpoklady z časti 6.2 a sformulujeme vety, zatiaľ bez dôkazov.



**Predpoklad 6.3.1.** Uvažujme proces  $y_t$  modelovaný STAR modelom s  $m$ -režimami podľa 4.3. Potom

$$(F(\mathbf{X}, \boldsymbol{\theta}) - y_t) \rightarrow_p 0, \quad (6.6)$$

kde

$$\begin{aligned} F(\mathbf{X}, \boldsymbol{\theta}) &= \Phi_1 \mathbf{X}_t G(z_t, r, \gamma) + (\Phi_2 - \Phi_1) \mathbf{X}_t G(z_t, r_1, \gamma_1) + \dots \\ &+ (\Phi_m - \Phi_{m-1}) \mathbf{X}_t G(z_t, r_{m-1}, \gamma_{m-1}), \quad r < r_1 < \dots < r_{m-1}. \end{aligned} \quad (6.7)$$

**Predpoklad 6.3.2.** Pre dané  $r$  a  $p$  platí, že  $|C_{n,k} - \sum_{j=1}^m C_{n_j, k_j}| < C$  skoro všade, pre všetky  $p$ , kde  $C$  je kladná konštanta,  $C_{n_j, k_j}$  je penalizačný člen pre  $j$ -ty režim,  $n_j = \sum_{t=1}^n \Delta G_{j,t}$  a  $G_{0,t} = 1, G_{m,t} = 0$ . Tvrdenie platí aj pre dané  $n$  a  $k$  a rozdielne  $r$ .

Ak platia predpoklady 6.2.2 až 6.2.5 z časti 6.2 a predpoklady 6.3.1 a 6.3.2, potom platia nasledovné tvrdenia:

**Veta 6.3.1.** Uvažujme proces  $y_t$  modelovaný STAR modelom s  $m$ -režimami podľa 4.3. Potom je odhad dosiahnutý pomocou informačného kritéria s penalizačným členom  $C_{n,k}$  slabo konzistentný práve vtedy, keď platí

$$\begin{aligned} C_{n,k} &\rightarrow_p \infty \\ \frac{C_{n,k}}{n} &\rightarrow_p 0 \\ C_{n,k_i} - C_{n,k_j} &\rightarrow_p \infty \\ \text{pre } k_i > k_j \quad i &= 1, \dots, j-1 \quad j = 2, \dots, m. \end{aligned}$$

**Veta 6.3.2.** Uvažujme proces  $y_t$  modelovaný STAR modelom s  $m$ -režimami podľa 4.3. Potom je odhad dosiahnutý pomocou informačného kritéria s penalizačným členom  $C_{n,k}$  silne konzistentný práve vtedy, keď platí

$$\begin{aligned} \frac{C_{n,k}}{\ln(\ln(n))} &\rightarrow_{s.v.} C \quad k < C \leq \infty \\ \frac{C_{n,k}}{n} &\rightarrow_{s.v.} 0 \\ \frac{C_{n,k_i} - C_{n,k_j}}{\ln(\ln(n))} &\rightarrow_{s.v.} C \quad k < C \leq \infty \\ \text{pre } k_i > k_j \quad i &= 1, \dots, j-1 \quad j = 2, \dots, m. \end{aligned}$$

## 6.4 Konzistentnosť IC pre STAR modely

V tejto podkapitole overíme konzistentnosť informačných kritérií z časti 5.4.2.

Je zrejmé, že AIC nebude konzistentné, lebo nespĺňa prvú podmienku. Naopak BIC, resp. HQIC sú slabo konzistentné, a BIC je dokonca aj silne konzistentné kritérium.

Pre nové informačné kritéria definované v časti 5.4.2 možno poľahky ukázať, že majú analogické vlastnosti ako BIC, t.j. sú silne konzistentné. Overenie tohto tvrdenia nechávame (pre jeho jednoduchosť) na čitateľa.

## 7 Aplikácia na reálnych dátach

V tejto kapitole sa zameriame na aplikáciu doteraz spomenutých poznatkov na reálnych dátach. Modelovanie časových radov možno bez újmy na všeobecnosti používať na akékoľvek reálne dáta. V praxi sa stretneme so širokým spektrom dát, napr. od modelovania makroekonomických ukazovateľov až po modelovanie stavov a prietokov slovenských riek.

My sa však v našej práci zameriame na makroekonomické ukazovatele a výmenné kurzy. Po prvé pre ich ľahkú dostupnosť a dôveryhodnosť zdroja. A po druhé kvôli tomu, že schopnosť predpovedať a modelovať tieto ukazovatele sa stalo populárnym a žiadaným aj vďaka integrácií do Európskej únie, ako aj do eurozóny. Momentálne sa sleduje napríklad schopnosť splniť Maastrichtské kritéria pre vstup do eurozóny.

Zameriame sa teda na slovenské dáta, ktoré najviac charakterizujú stav slovenskej ekonomiky. Konkrétne pôjde o

- HPD - štvrtročný hrubý domáci produkt v bežných cenách
- INFLÁCIA - medziročné dáta na mesačnej báze
- NEZAMESTNANOSŤ - mesačné dáta evidovanej miery nezamestnanosti
- VÝMENNÝ KURZ USD/SKK - priemerný výmenný kurz NBS na mesačnej báze
- VÝMENNÝ KURZ EUR/SKK - priemerný výmenný kurz NBS na mesačnej báze

Vyššie uvedené dáta pochádzajú zo stránok Národnej Banky Slovenska ([www.nbs.sk](http://www.nbs.sk)), ktorá niektoré údaje priamo preberá zo Štatistického úradu SR ([www.statistics.sk](http://www.statistics.sk)). Každému sledovanému údaju bude venovaná samostatná kapitola, zhŕňajúca výsledky a ich následnú interpretáciu.

V každej časti zistíme na základe IC, ako aj predikčných chýb, najlepší model z triedy STAR modelov, ktorý zobrazíme spolu s modelovaným časovým radom reziduí po dekompozícii deterministických zložiek a prahovou hodnotou  $r$  (modelovaný časový rad zobrazíme čiernou, model červenou a

prahovú hodnotu modrou farbou). Do ďalšieho obrázku zobrazíme skutočné hodnoty sledovaného časového radu (čiernou farbou) a hodnoty predpovedané na základe modelu (zelenou farbou).

Na modelovanie týchto ukazovateľov použijeme výpočtový systém Mathematica a v ňom program, ktorý napísal Ing. Tomáš Bacigal, interný doktorand na KMaDG SvF STU v Bratislave. Jeho konkrétne časti si zhrnieme v samostatnej podkapitole.

## 7.1 Popis programu

V tejto časti si bližšie popíšeme program Ing. Tomáša Bacigála na analýzu a predikciu časových radov pomocou TAR a STAR modelov.

Samotná aplikácia je vyvinutá pre viacrozmerné (vektorové) modelovanie časových radov, preto pre naše účely budeme používať len časť tejto aplikácie. Aplikáciu možno rozdeliť na niekoľko častí

- Definícia potrebných funkcií a premenných
- Dekompozícia deterministických zložiek časového radu
- Testovanie linearity oproti nelinearite typu STAR a v prípade zamietnutia lineárnosti následné modelovanie reziduálnej zložky STAR modelmi
- Predikcia hodnôt a ich porovnanie so skutočným stavom

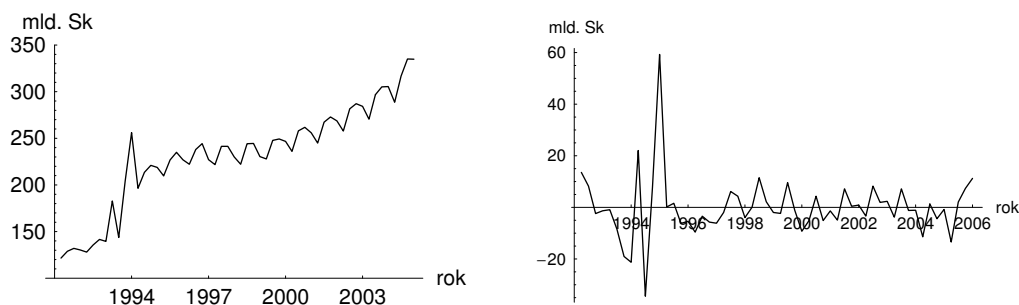
Na analýzu deterministických zložiek použijeme lineárnu regresiu (odstránime ňou trend) a spektrálnu analýzu a funkcie  $\sin(x)$ ,  $\cos(x)$  (na odstránenie sezónnej a cyklickej zložky).

Z časového radu nám ostane len reziduálna zložka. Spravidla už na korelograme vidno nelinearitu, ktorú následne podrobíme LM testu s nulovou hypotézou, že daný časový rad má lineárny charakter. Rád lineárneho modelu určíme na základe informačných kritérií. V prípade zamietnutia linearity modelujeme reziduálnu zložku dvoj-, prípadne viacrežimovým STAR modelom, s logistickou, resp. exponenciálnou funkciou.

Poslednou časťou je predikcia hodnôt. Analýzu časového radu vykonávame len na časti dát. Posledných niekoľko údajov, spravidla rok-dva, použijeme na porovnanie s predikciami danými modelom.

## 7.2 HDP

Ako už bolo povedané vyššie, budeme modelovať HDP na štvrťročnej báze v stálych cenách v období od roku 1993 do roku 2005. Tento model potom použijeme na predpoveď na rok 2006, ktorú porovnáme so skutočnými hodnotami. Zobraziť si dáta do grafu (obrázok 2). Z grafu dobre vidno, že dáta budú mať lineárny rastúci trend a silnú sezónnu a cyklickú zložku. Po ich odstránení dostávame graf reziduí, ktorý budeme modelovať STAR modelmi (obrázok 2).



Obrázok 2: Graf dát použitých na modelovanie s  $\mu = 232,16$  a  $\sigma = 53,81$  (vľavo) a graf reziduí po odstránení deterministických zložiek (vpravo)

Na základe LM3 testu modelujeme reziduálnu zložku STAR modelmi s 2 režimami. Odhady parametrov, ich štandardné chyby, základné vlastnosti modelov, príslušné informačné kritéria a chyby predikcií nájdeme v nasledujúcej tabuľke 1, ako aj v zápise výsledného modelu.

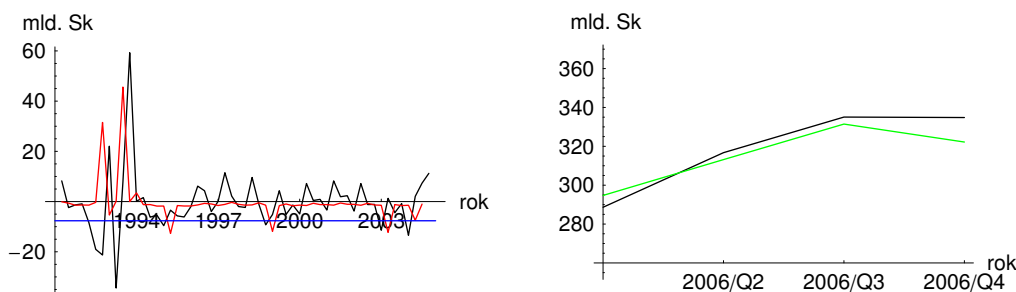
Ako vidno, najlepším modelom (podľa predikčnej chyby MAE, ako aj podľa informačných kritérií) je model (hodnoty štandardných chýb odhadov parametrov budeme písať za parametrom do zátvoriek)

$$y_t = [-42,920_{(0,0151)} - 0,996_{(0,0005)}y_{t-1} - 2,798_{(0,0008)}y_{t-2}](1 - G(\cdot)) + [-1,248_{(0,0027)} + 0,0108_{(0,0002)}y_{t-1} + 0,079_{(0,0002)}y_{t-2}]G(\cdot) + \epsilon_t, \quad (7.1)$$

$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
2	2	-1,97	20	6	88,1	25,95	27,92	27,33	33,19	43,86	7,23	7,63
4	4	-7,64	20	4	53,12	31,54	31,57	32,63	40,31	58,31	3,69	4,03
4	4	-6,02	20	6	50,53	31,82	32,77	33,12	41,33	58,79	8,20	9,47
5	5	-7,64	20	6	48,50	35,35	35,23	36,47	45,59	66,57	7,02	9,27
2	2	-7,64	20	2	72,55	24,40	24,85	25,26	30,13	41,23	3,48	4,24

Tabuľka 1: Tabuľka pre logistickú funkciu

kde  $G(\cdot) = \frac{1}{1 + \exp^{-20(y_t - 2 + 7,64)}}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 3.



Obrázok 3: Graf výsledného modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

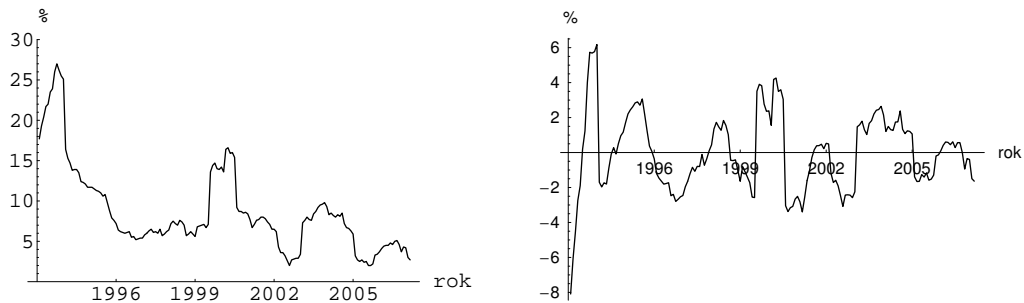
HDP nebudeme modelovať exponenciálnou funkciou, keďže táto metóda je pre dané dáta numericky nestabilná a nedáva dobré výsledky (zlá podmienosť matíc). Vo všeobecnosti ale môžeme povedať, aj vzhľadom na dosiahnuté výsledky (vysoké hodnoty parametra  $\gamma$ ), že HDP bude lepšie modelovať pomocou (SE)TAR modelov.

Model s logistickou funkciou dosiahol minimálnu hodnotu MAE (a druhú najmenšiu pre RMSE) zároveň s minimálnou hodnotou pre každé informačné kritérium, čo potvrdzuje dobré výberové vlastnosti našich informačných kritérií. Ako vidíme z výsledkov,  $d = 2$ , t.j. zmena režimu nastane v závislosti od hodnoty pred pol rokom. Z toho môžeme usúdiť, že hladký prechod nie je schopný dostatočne dobre vysvetliť vývoj HDP. Je zrejmé, že naň vply-

va veľa ďalších exogénnych premenných. Rozptyl reziduí  $\sigma^2$  je pre tento model relatívne veľký, čo je však spôsobené nízkym rádom procesov AR v jednotlivých režimoch.

### 7.3 Inflácia

Ako ďalšiu veličinu budeme modelovať čistú infláciu na mesačnej báze v období od roku 1993 do roku 2005. Tento model potom použijeme na predpoveď inflácie v roku 2006 a prvých dvoch mesiacov roku 2007, ktoré porovnáme so skutočnými hodnotami. Zobraziť si dáta do grafu (obrázok 4). Z grafu dobre vidno, že dáta majú veľmi zložitý trend, a preto ich budeme modelovať kubickým trendom, keďže kvadratický je nedostatočný. Je tu prítomná aj cyklická zložka, ale nie taká silná, ako v prípade HDP. Po ich odstránení dostávame graf reziduí, ktorý budeme modelovať STAR modelmi (obrázok 4).



Obrázok 4: Graf dát použitých na modelovanie s  $\mu = 8,63$  a  $\sigma = 5,34$  (vľavo) a graf reziduí po odstránení deterministických zložiek (vpravo)

Podľa výsledku LM3 testu modelujeme reziduálnu zložku STAR modelmi s 2 režimami. Odhady parametrov, ich štandardné chyby, základné vlastnosti modelov, príslušné informačné kritéria a chyby predikcií nájdeme v nasledujúcich tabuľkách (2, 3), ako aj v zápise výsledného modelu.

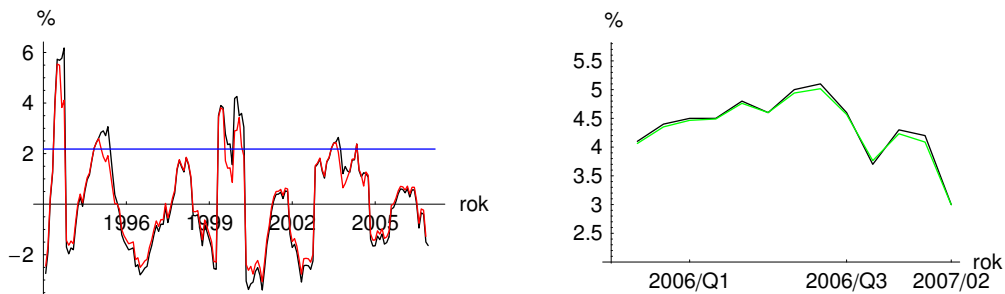
Ako vidno z tabuľky 2, najlepším modelom (podľa predikčných chýb, ako aj podľa informačných kritérií) je model

$$y_t = [0,1362_{(0,0005)} + 0,9425_{(0,0002)}y_{t-1}](1 - G(\cdot)) + [-0,2398_{(0,0013)} + 0,7040_{(0,0001)}y_{t-1}]G(\cdot) + \epsilon_t, \quad (7.2)$$

$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
2	2	2,18	10	4	1,06	18,39	23,16	23,56	31,13	57,44	0,47	0,57
1	1	2,18	10	3	1,14	14,55	17,73	18,00	23,05	40,66	0,43	0,54
1	1	2,18	10	4	1,11	14,48	17,66	17,93	22,98	40,52	0,42	0,53
2	2	2,18	10	3	1,08	18,46	23,27	23,67	31,27	57,71	0,46	0,58
2	2	2,18	10	5	1,03	18,33	23,09	23,48	31,05	57,25	0,52	0,62
2	2	2,18	10	1	1,09	18,44	23,22	23,65	31,23	57,85	0,46	0,58

Tabuľka 2: Tabuľka pre logistickú funkciu

kde  $G(\cdot) = \frac{1}{1 + \exp^{-10(y_t - 4 - 2,18)}}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 5.



Obrázok 5: Graf výsledného LSTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

V tabuľke 3 je najlepším modelom (podľa predikčných chýb, ako aj podľa informačných kritérií) model

$$y_t = [-0,1283_{(0,0010)} + 1,1376_{(0,0006)}y_{t-1}](1 - G(\cdot)) + [-0,0252_{(0,0006)} + 0,7610_{(0,0002)}y_{t-1}]G(\cdot) + \epsilon_t, \quad (7.3)$$

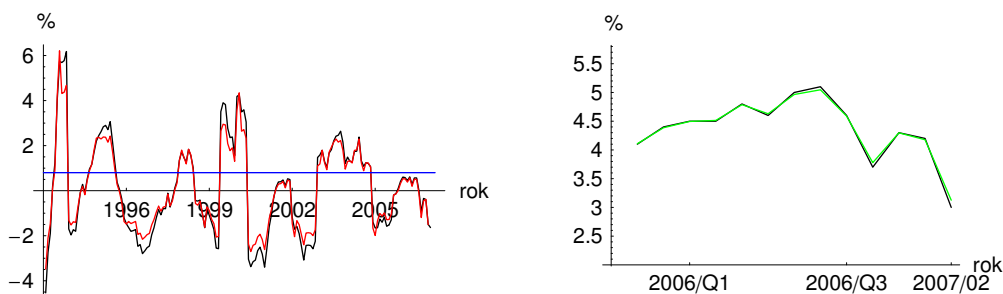
kde  $G(\cdot) = 1 - \exp^{-0,5(y_t - 3 - 0,8)^2}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 6.

Obidva modely dosiahli minimálnu hodnotu MAE a RMSE zároveň s minimálnou hodnotou pre každé informačné kritérium, čo potvrdzuje dobré



$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
1	1	1,03	0,5	4	1,14	15,58	19,38	19,57	25,11	43,69	0,44	0,52
1	1	0,8	0,5	3	1,14	15,50	19,32	19,52	25,02	43,46	0,42	0,51
2	2	0,8	0,5	3	1,09	19,52	25,26	25,54	33,87	61,91	0,44	0,57
1	1	1,03	0,5	5	1,14	15,56	19,36	19,54	25,07	43,58	0,47	0,57
2	2	-0,58	1	1	1,09	19,22	24,68	25,00	33,10	60,75	0,58	0,68

Tabuľka 3: Tabuľka pre exponenciálnu funkciu



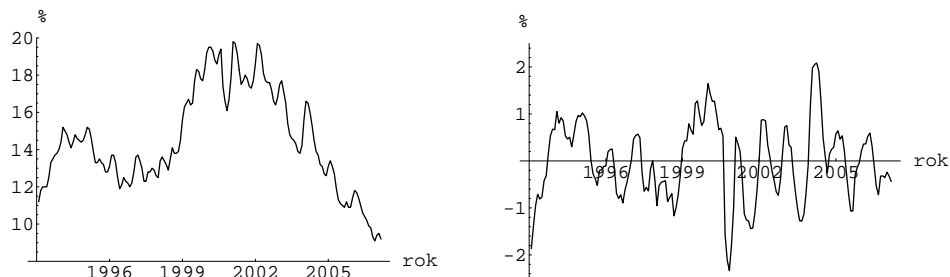
Obrázok 6: Graf výsledného ESTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

výberové vlastnosti našich informačných kritérií. Pri porovnaní IC a predikčných chýb pre STAR model s logistickou a exponenciálnou prechodovou funkciou vidno, že na popis dát je vhodnejší LSTAR model, zatiaľčo na predikciu je mierne lepší ESTAR model. Inflácia je veľmi dobrý príklad STAR modelov, čo potvrdzuje hlavne veľmi silná predikčná schopnosť. Za poznámku stojí aj fakt, že model využíva v jednotlivých režimoch jednoduchý AR model, medzi ktorými prepína na základe hodnoty spred štvrt' roka.

## 7.4 Nezamestnanosť

Poslednou makroekonomickou veličinou, ktorú budeme modelovať, je evidovaná nezamestnanosť meraná na mesačnej báze v období od roku 1993 do roku 2005. Tento model potom použijeme na predpoveď na rok 2006 a prvé dva mesiace roku 2007, ktoré porovnáme so skutočnými hodnotami. Zobrazme si dáta do grafu (obrázok 7). Z grafu dobre vidno, že dáta budú mať

opäť kubický trend (kvôli dvom výrazným "kopcom"). Vzhľadom na očakávanú silnú sezónnu zložku (s periódou 12) a cyklickú zložku budeme tieto dáta modelovať väčším počtom funkcií v tvare  $\sin(kx)$ , resp.  $\cos(kx)$ . Po odstránení deterministických zložiek dostávame graf reziduí, ktorý budeme modelovať STAR modelmi (obrázok 7).



Obrázok 7: Graf dát použitých na modelovanie s  $\mu = 14,49$  a  $\sigma = 2,66$  (vľavo) a graf reziduí po odstránení deterministických zložiek (vpravo)

Podľa výsledkov LM3 testu modelujeme reziduálnu zložku STAR modelmi s 2 režimami. Odhady parametrov, ich štandardné chyby, základné vlastnosti modelov, príslušné informačné kritéria a chyby predikcií nájdeme v nasledujúcich tabuľkách (4, 5), ako aj v zápise výsledného modelu.

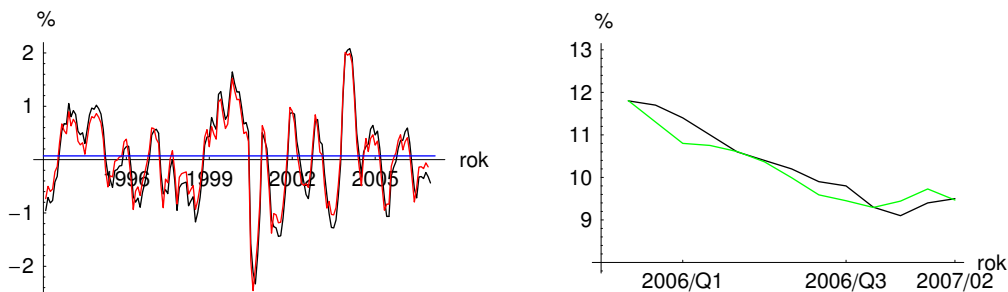
$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
1	1	0,07	20	3	0,11	11,19	15,09	15,28	20,89	39,74	0,32	0,36
1	1	0,07	20	4	0,12	11,26	15,14	15,33	20,93	39,70	0,35	0,41
2	2	0,79	20	4	0,10	13,74	18,57	18,96	26,58	52,97	0,36	0,40

Tabuľka 4: Tabuľka pre logistickú funkciu

Ako vidno v tabuľke 4, najlepším modelom pre logistickú funkciu (podľa predikčných chýb, ako aj podľa informačných kritérií) je model

$$y_t = [0,1647_{(0,0007)} + 0,9374_{(0,0009)}y_{t-1}](1 - G(\cdot)) + [-0,2184_{(0,0008)} + 1,06_{(0,0009)}y_{t-1}]G(\cdot) + \epsilon_t, \quad (7.4)$$

kde  $G(\cdot) = \frac{1}{1 + \exp^{-20(y_t - 3 - 0,07)}}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 8.



Obrázok 8: Graf výsledného LSTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
2	2	-0,81	20	4	0,10	13,28	17,64	18,15	25,42	51,29	0,37	0,41
1	1	0,63	4	4	0,12	11,10	14,87	15,06	20,58	39,10	0,35	0,40
1	1	0,55	5	3	0,12	11,05	14,76	14,95	20,45	39,04	0,31	0,35

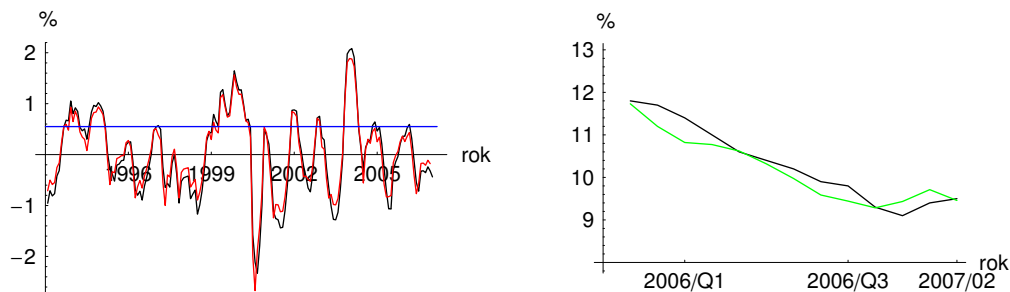
Tabuľka 5: Tabuľka pre exponenciálnu funkciu

Ako vidno v tabuľke 5, najlepším modelom pre exponenciálnu funkciu (podľa predikčných chýb, ako aj podľa informačných kritérií) je model

$$y_t = [-0,2741_{(0,0009)} + 1,1365_{(0,0013)}y_{t-1}](1 - G(.)) + [0,1040_{(0,0006)} + 0,8518_{(0,0007)}y_{t-1}]G(.) + \epsilon_t, \quad (7.5)$$

kde  $G(.) = 1 - \exp^{-5(y_{t-3}-0,55)^2}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 9.

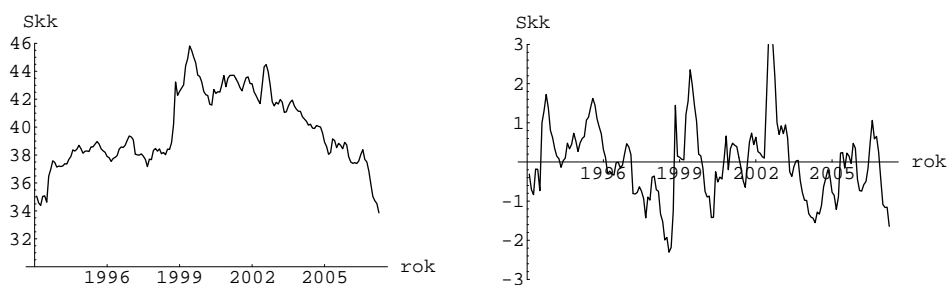
Obidva modely dosiahli minimálnu hodnotu MAE aj RMSE zároveň s minimálnou hodnotou pre každé informačné kritérium, čo potvrdzuje dobré výberové vlastnosti našich informačných kritérií. Napriek dost veľkej odchýlke od reálnych dát sa na modelovanie nezamestnanosti ukazuje ako vhodnejšie použiť exponenciálnu prechodovú funkciu, nakoľko v prípade logistickej funkcie je hodnota parametra  $\gamma$  príliš vysoká a indikuje skôr (SE)/TAR model ako STAR model. Podobne možno konštatovať, že predikčná schopnosť nie je tak silná, ako v prípade inflácie, i keď dodržiava správny trend.



Obrázok 9: Graf výsledného ESTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

## 7.5 Výmenný kurz EUR/SKK

Zamerajme sa aj na výmenné kurzy, konkrétne na modelovanie mesačného výmenného kurzu EUR/SKK v období od roku 1993 do roku 2005. Tento model potom použijeme na predpoveď na rok 2006 a prvé tri mesiace roku 2007, ktoré porovnáme so skutočnými hodnotami. Zobrazme si dáta do grafu (obrázok 10). Z dlhodobého hľadiska vieme, že výmenné kurzy majú silnú cyklickú zložku, preto budeme v regresii klasť dôraz práve na ňu. Po odstránení cyklickej zložky dostávame graf reziduí, ktorý budeme modelovať STAR modelmi (obrázok 10).



Obrázok 10: Graf dát použitých na modelovanie s  $\mu = 38,78$  a  $\sigma = 2,66$  (vľavo) a graf reziduí po odstránení deterministických zložiek (vpravo)

Na základe výstupu z LM3 testu modelujeme reziduálnu zložku STAR modelmi s 2 režimami. Odhady parametrov, ich štandardné chyby, základné vlastnosti modelov, príslušné informačné kritéria a chyby predikcií nájdeme

v nasledujúcich tabuľkách (6, 7), ako aj v zápise výsledného modelu.

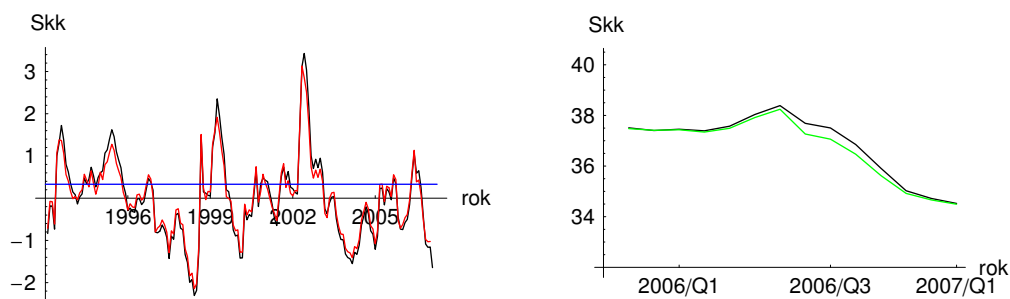
$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
2	2	0,42	20	4	0,21	16,05	21,61	21,91	30,10	57,77	0,64	0,73
2	2	0,42	20	5	0,20	15,97	21,57	21,85	30,08	57,74	0,63	0,71
2	2	0,33	20	3	0,21	16,18	21,86	22,16	30,44	58,44	0,62	0,70
1	1	0,33	20	3	0,22	12,24	16,03	16,23	21,75	40,42	0,58	0,66
2	2	-0,03	20	6	0,20	16,25	22,06	22,33	30,72	58,73	0,60	0,68

Tabuľka 6: Tabuľka pre logistickú funkciu

Pre logistickú prechodovú funkciu je najlepším modelom (podľa predikčných chýb, ako aj podľa informačných kritérií) model

$$y_t = [0,1010_{(0,0006)} + 0,9737_{(0,0006)}y_{t-1}](1 - G(\cdot)) + [-0,1393_{(0,0010)} + 0,8735_{(0,0009)}y_{t-1}]G(\cdot) + \epsilon_t, \quad (7.6)$$

kde  $G(\cdot) = \frac{1}{1 + \exp^{-20(y_t - 3 - 0,33)}}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 11.



Obrázok 11: Graf výsledného LSTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

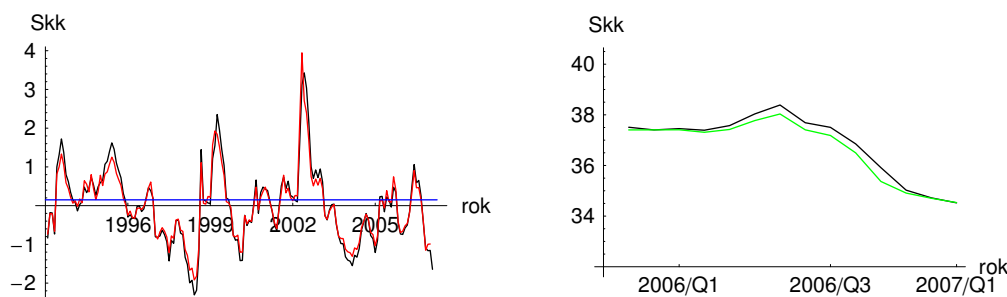
$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
2	2	0,6	20	4	0,22	14,96	19,47	19,95	27,32	53,43	0,63	0,71
2	2	0,24	4	5	0,20	16,13	21,86	22,14	30,46	58,38	0,60	0,68
1	1	0,15	20	3	0,20	11,44	14,81	15,05	20,25	38,18	0,59	0,67
2	2	0,69	20	6	0,20	14,58	18,79	19,33	26,49	52,11	0,60	0,69

Tabuľka 7: Tabuľka pre exponenciálnu funkciu

Pre exponenciálnu prechodovú funkciu je najlepším modelom (podľa predikčných chýb, ako aj podľa informačných kritérií) model

$$y_t = [0,1537_{(0,0012)} + 1,2437_{(0,0013)}y_{t-1}](1 - G(\cdot)) + [-0,0574_{(0,0005)} + 0,8071_{(0,0005)}y_{t-1}]G(\cdot) + \epsilon_t, \quad (7.7)$$

kde  $G(\cdot) = 1 - \exp^{-20(y_{t-3} - 0,15)^2}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 12.



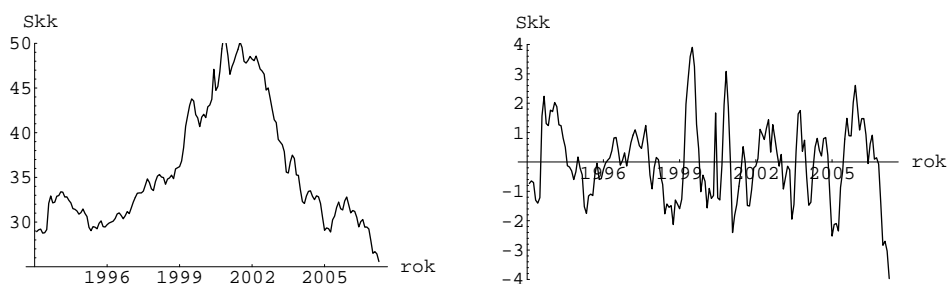
Obrázok 12: Graf výsledného ESTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

Obidva modely dosiahli minimálnu hodnotu MAE aj RMSE zároveň s minimálnou hodnotou pre každé informačné kritérium, čo potvrdzuje dobré výberové vlastnosti našich informačných kritérií. Ako vidno z výsledkov, obidva modely majú vysoké hodnoty parametra  $\gamma$ , t.j. vhodnejšie je modelovať tieto časové rady pomocou (SE)TAR modelov. V prípade exponenciálnej prechodovej funkcie možno z tohto dôvodu akceptovať zložitejší model (s AR(2)

modelmi v jednotlivých režimoch), ktorý vykazuje skoro identické výsledky, ale je to vhodnejší STAR model pre tieto dáta (vzhľadom na  $\gamma = 4$ ).

## 7.6 Výmenný kurz USD/SKK

V súvislosti s modelovaním výmenného kurzu EUR/SKK sa upriamime aj na modelovanie mesačného výmenného kurzu USD/SKK v období od roku 1993 do roku 2005. Tento model potom použijeme na predpoveď na rok 2006 a prvé tri mesiace roku 2007, ktoré porovnáme so skutočnými hodnotami. Zobrazieme si dáta do grafu (obrázok 13). V prípade výmenného kurzu USD/SKK budeme postupovať rovnako ako v prípade EUR/SKK v časti 7.5. Po odstránení cyklickej zložky dostávame graf reziduí, ktorý budeme modelovať STAR modelmi (obrázok 13).



Obrázok 13: Graf dát použitých na modelovanie s  $\mu = 35,76$  a  $\sigma = 6,63$  (vľavo) a graf reziduí po odstránení deterministických zložiek (vpravo)

Podľa výsledkov LM3 testu modelujeme reziduálnu zložku STAR modelmi s 2 režimami. Odhady parametrov, ich štandardné chyby, základné vlastnosti modelov, príslušné informačné kritéria a chyby predikcií, nájdeme v nasledujúcich tabuľkách (8, 9), ako aj v zápise výsledného modelu.

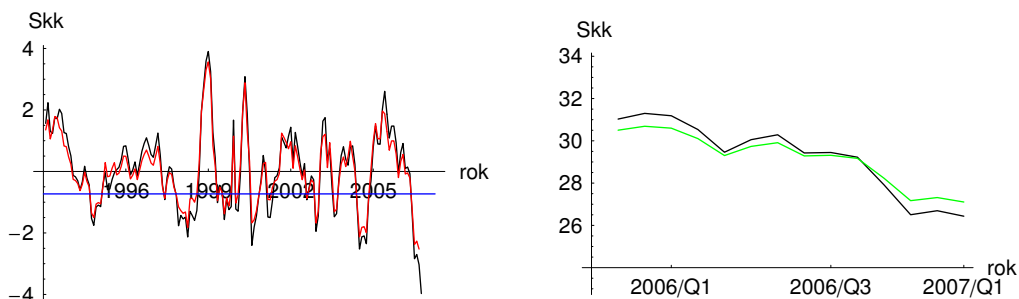
Najlepším modelom (podľa predikčných chýb, ako aj podľa informačných kritérií) pre logistickú funkciu je model

$$y_t = [0,3224_{(0,0009)} + 0,8305_{(0,0005)}y_{t-1}](1 - G(\cdot)) + [-0,1575_{(0,0005)} + 0,7807_{(0,0004)}y_{t-1}]G(\cdot) + \epsilon_t, \quad (7.8)$$

kde  $G(\cdot) = \frac{1}{1 + \exp^{-20(y_t - 7 + 0,73)}}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 14.

$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
1	1	-1,01	20	2	0,58	13,86	17,51	17,63	23,06	41,12	1,10	1,48
2	2	-0,73	20	3	0,56	17,91	23,40	23,71	31,84	59,46	1,02	1,32
2	2	-0,73	20	7	0,46	17,44	22,84	23,12	31,19	58,28	0,93	1,21
1	1	-0,31	20	6	0,61	14,31	18,11	18,29	23,82	42,33	0,09	1,23
1	1	-0,73	20	7	0,56	13,82	17,42	17,60	22,99	41,04	0,91	1,20

Tabuľka 8: Tabuľka pre logistickú funkciu



Obrázok 14: Graf výsledného LSTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

Najlepším modelom pre exponenciálnu funkciu je podľa predikčných chýb model 7.9 s AR(2) v obidvoch režimoch,  $d = 2$ ,  $r = -1,29$  a  $\gamma = 10$ . Tento má však najväčšie hodnoty všetkých IC. Je teda najlepší na predikciu, ale nie je veľmi vhodný na popisné účely (pre ne by bol lepší prvý z modelov v tabuľke 9).

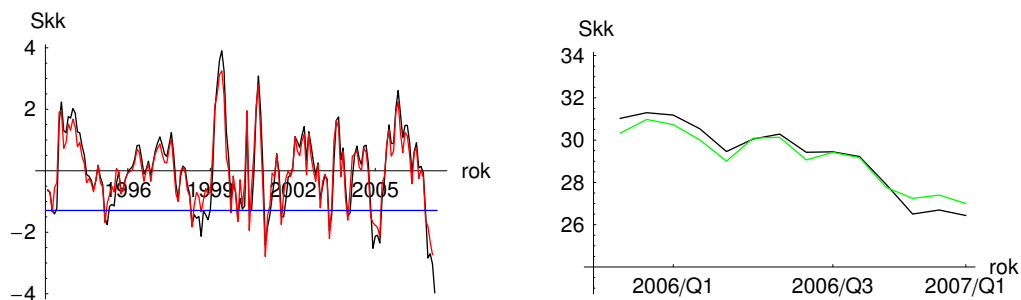
$$y_t = [-0,544_{(0,0132)} + 0,849_{(0,0013)}y_{t-1} - 0,915_{(0,0102)}y_{t-2}](1 - G(\cdot)) + [-0,100_{(0,0005)} + 1,117_{(0,0006)}y_{t-1} - 0,281_{(0,0006)}y_{t-2}]G(\cdot) + \epsilon_t, \quad (7.9)$$

kde  $G(\cdot) = 1 - \exp^{-10(y_{t-2} + 1,29)^2}$ . Výsledný graf modelu a predpovedí možno nájsť na obrázku 15.



$p_1$	$p_2$	$r$	$\gamma$	$d$	$\sigma^2$	AIC	HQIC	OIC	BICC	SIC	MAE	RMSE
1	1	-1,29	20	2	0,57	12,47	15,20	15,62	20,33	37,60	1,03	1,34
2	2	-1,29	20	3	0,56	16,43	20,52	21,14	28,20	54,01	1,02	1,26
2	2	-0,45	20	7	0,47	15,96	19,85	20,49	27,42	52,69	1,03	1,27
2	2	-1,29	10	2	0,53	16,78	21,34	21,82	29,23	55,57	1,01	1,25

Tabuľka 9: Tabuľka pre exponenciálnu funkciu



Obrázok 15: Graf výsledného ESTAR modelu (vľavo) a graf predpovedí tohto modelu porovnaných so skutočnými hodnotami (vpravo)

## 8 Odporúčania a záver

V tejto diplomovej práci sme zhrnuli základné poznatky o analýze časových radov, od dekompozície až po prehľad základných lineárnych jednorozmerných modelov. Na tomto základe sme zhrnuli nové moderné metódy analýzy pomocou nelineárnych viacrežimových modelov, pre ktoré sme definovali všeobecne známe informačné kritéria a k nim sme doplnili niekoľko nových kritérií, ktoré sme v aplikačnej časti úspešne empiricky otestovali.

Ukázalo sa, že STAR modely sú veľmi silný nástroj na modelovanie reziduálnej zložky mnohých reálnych časových radov. V aplikačnej časti bol však pre viaceré časové rady predpoklad hladkého prechodu medzi režimami prílišný a stačili by nám na modelovanie aj (SE)TAR modely. Je ale zrejmé, že STAR modely majú potenciál vyborne modelovať hlavne veľmi zložitú časovú radu, k čomu môžu pomôcť aj nami navrhnuté IC pre STAR modely. Veľký potenciál vidíme aj vo viacrozmernom modelovaní pomocou STAR modelov. Na skúmaných dátach vidno, že na ne vplyva veľké množstvo exogénnych zmien, ktoré sa veľmi ťažko presne popisujú. V prípade výmenných kurzov by sa lepšie výsledky dosiahli pomocou úspešného modelovania stochastických pohybov výmenného kurzu a špekulatívnych obchodov na trhu.

V našej aplikácii sa ukázalo, že so stúpajúcim rádom modelu klesá parameter  $\sigma^2$ , ale zároveň IC penalizujú zvýšenie rádu zväčšením svojej hodnoty. Zachováva sa tým rovnováha medzi jednoduchosťou modelu a presnosťou modelu. Dané výsledky plne potvrdzujú správnosť a vhodnosť používania informačných kritérií pre STAR modely.

V našej práci sme naznačili budovanie teórie o konzistentnosti týchto informačných kritérií a uviedli sme nové tvrdenia, zatiaľ bez dôkazu. Vychádzali sme z dosiaľ zverejnených prác v oblasti SETAR modelov a upravili sme podľa našich potrieb predpoklady a tvrdenia o konzistentnosti informačných kritérií pre SETAR modely.

Táto práca sa ďalej môže rozvíjať po matematickej stránke, t.j. úplný a korektný dôkaz daných tvrdení, ako aj v oblasti skúmania kointegrácie, či zahrnutia agregáčnych operátorov do výpočtov, čím by mohol vzniknúť priestor na lepšie a kvalitnejšie modely.

## Referencie

- [1] Anderson, D.R., Burnham, K.P.: *Multimodel inference: Understanding AIC and BIC in Model Selection*, Colorado State University, 2004.
- [2] Anderson, D.R., Burnham, K.P.: *Understanding information criteria for selection among capture-recapture or ring recovery models*, Colorado State University, 1999.
- [3] Bacigal, T.: *Advanced methods of time series modeling and their application in geodesy*, STU Bratislava, 2006.
- [4] Bacigal, T.: *Smooth Transition and Threshold AutoRegressive Model - Application to artificial time series*, [www.math.sk/bacigal/homepage/research.html](http://www.math.sk/bacigal/homepage/research.html), STU Bratislava, 2007.
- [5] Balcaen, S., Ooghe, H.: *35 years of studies on bussines failure: An overview of the classic statistical methodologies and their related problems*, Vlerick Leuven Ghent Management School, 2004.
- [6] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*, SNTL/ALFA, 1986.
- [7] Foster, R.M.: *Key Concepts in Model Selection: Performance and Generalizability*, University of Wisconsin, Madison, 2000.
- [8] Franses, P.H., van Dijk, D., Teräsvirta, T.: *Smooth Transition Autoregressive Models - A Survey of Recent Developments*, Econometric Institute Research Report EI 2000-23/A (2000).
- [9] Franses, P. H.: *Time series models for business and economic forecasting*, Cambridge University Press, 1998.

- [10] Franses, P. H., Dijk, D.: *Non-linear time series models in empirical finance*, Cambridge University Press, 2000.
- [11] Granger, C.W.J., Teräsvirta, T.: *Modelling Nonlinear Economic Relationship*, Oxford University Press, 1993.
- [12] Hannan, J.E.: *The estimation of the order of an ARMA process*, Annals of Statistics Vol.8,. No.5, pp. 1071-1081, 1980.
- [13] Kapetanios, G.: *Model Selcetion in Threshold Models*, Churchill College, Cambridge, National Institute of Economic and Social Research, London, 1999.
- [14] Kullback, S., Leibler, R.A.: *On information and sufficiency*, Annals of Mathematical Statistics 22:79-86, 1951.
- [15] Lopez, H.F., Salazar, E.: *Bayesian model uncertainty in smooth transition autoregressions*, University of Chicago and Federal University of Rio de Janeiro, 2004.
- [16] Maringer, D., Meyer, M.: *Smooth Transition Autoregressive Models - New Approach to the model selection problem*, University of Essex, 2006.
- [17] Sen, L.K., Shitan, M.: *The performance of AICC as an order selection criterion in ARMA time series model*, Universiti Putra Malaysia, 2000
- [18] Shibata, E.: *Bootstrap estimate of Kullback-Leibler information for model selection*, Keio University, 1997
- [19] Sugiyama, M., Ogawa, H.: *Subspace information criterion for model selection*, Department of computer science, Tokyo Institute of Technology, 2001

- [20] Strikholm, B., Teräsvirta, T.: *Determining the Number of Regimes in Threshold Autoregressive Model Using Smooth Transition Autoregression*,  
Department of Economic Statistics, Stockholm School of Economics, 2002.
- [21] Szokeova, D.: *Analýza vlastností stochastických procesov*,  
STU Bratislava, 2006.
- [22] Talata, Zs.: *Model Selection via Information Criteria*,  
Stochastic Research Group, Hungarian Academy of Sciences,  
Budapest, 2005
- [23] Tong, H.: *On a threshold model. Pattern recognition and Signal processing*,  
Sijthoff & Nordhoff, Amsterdam, 1978, 101-141.
- [24] Tong, H.: *Non-linear time series: A dynamical systems approach*,  
Oxford University Press, Oxford, 1990.
- [25] White, H., Sin, C.: *Information criteria for selecting possibly mis-specified parametric models*,  
Journal of Econometrics, 71, 207–225.