



UNIVERZITA KOMENSKÉHO V BRATISLAVE  
Fakulta matematiky, fyziky a informatiky  
Katedra aplikovanej matematiky a štatistiky

# Nelineárne autoregresné modely

Diplomová práca

Bc. Erika Erdélyi

9.1.9 Aplikovaná matematika  
Ekonomická a finančná matematika

Školiteľ:  
Doc. RNDr. Ján Pekár, PhD.

BRATISLAVA 2010

## Čestné prehlásenie

Prehlasujem, že som diplomovú prácu vypracovala samostatne s využitím svojich poznatkov a s použitím uvedenej literatúry.

V Bratislave 26. apríla 2010

.....  
Erika Erdélyi

## PodĎakovanie

Touto cestou by som sa chcela poĎakovať môjmu diplomovému vedúcemu

Doc. RNDr. Jánovi Pekárovi, PhD.,

za ochotu, pomoc pri zaobstaraní potrebnej literatúry a za všetok svoj čas a námahu, ktoré vynaložil, aby ma viedol pri písaní tejto práce.

Samostatné poĎakovanie patrí aj

Ing. Tomášovi Bacigálovi, PhD.,

ktorého program vo výpočtovom systéme Mathematica 5.2 som využila pri aplikácii na reálnych dátach.

Ďakujem.

## Abstrakt

Práca je venovaná najznámejším nelineárnym modelom, ide o SETAR, STAR modely a Markovove modely prepínania režimov. Cieľom práce je popísať tieto modely a aplikovať získané teoretické poznatky o týchto modeloch aj v praxi. Špeciálne sa budeme zaoberať modelmi SETAR a STAR, ktoré využijeme aj pri modelovaní našich reálnych dát.

**Kľúčové slová:** modelovanie časových radov, nelineárne modely, SETAR modely, STAR modely, MSW modely

## Abstract

This thesis deals with the theory of the most popular nonlinear models: SETAR, STAR and Markov switching models. Our objective is to apply theoretical results in real data analysis, especially models SETAR and STAR.

**Keywords:** modelling time series, nonlinear models, SETAR models, STAR models, MSW models

# Obsah

Úvod	1
<b>1 Časový rad HDP Švédska</b>	<b>2</b>
1.1 Použitý časový rad	2
1.2 Ekonomická história Švédska od 1980 až po krízu 90-tych rokov	2
1.3 Švédska finančná kríza 90-tych rokov	3
1.4 Ekonomika v Švédsku po „veľkej“ kríze	4
1.5 Vplyv svetovej finaçnej krízy na Švédsko	4
<b>2 Teória</b>	<b>6</b>
2.1 Základné pojmy	6
2.1.1 Stacionarita	6
2.1.2 Momenty náhodných procesov	7
2.1.3 Biely šum	8
2.1.4 Operátor posunutia	9
2.1.5 Waldova reprezentácia	9
2.2 Dekompozícia časového radu	10
2.3 Testovanie hypotéz a rozhodovacie kritériá	11
<b>3 Boxove-Jenkinsove modely</b>	<b>15</b>
3.1 Procesy $AR(p)$	15
3.2 Procesy $MA(q)$	16
3.3 Procesy $ARMA(p, q)$	17
3.4 Integrované procesy $ARIMA(p, q)$	17
3.5 Sezónne procesy $SARIMA(p, d, q)(P, D, Q)_s$	17
<b>4 Nelineárne modely</b>	<b>19</b>
4.1 Samobudiacie prahové autoregresné modely	
SETAR( $p; d, r$ )	19
4.1.1 Preusporiadanie autoregresie a prediktívne reziduá	20
4.1.2 Identifikácia parametra omeškania $d$	21
4.1.3 Určenie hodnôt prahov $r$	22
4.2 Autoregresné modely hladkého prechodu STAR	22
4.2.1 Špecifikácia modelu STAR	23
4.2.2 AR model	24
4.2.3 Test linearity a voľba vhodnej funkcie prechodu	24

4.2.4	Odhad parametrov . . . . .	27
4.3	Markovove modely prepínania režimov . . . . .	27
4.3.1	Základný popis a vyžitie . . . . .	27
4.3.2	Tvorba MSW modelov . . . . .	28
4.3.3	Testovanie linearity oproti nelinearite typu MSW . . . . .	29
4.3.4	Odhady parametrov modelu MSW . . . . .	29
<b>5</b>	<b>Aplikácia teórie v praxi</b>	<b>31</b>
5.1	Modely SETAR . . . . .	32
5.1.1	Lineárne a nelineárne modelovanie časového radu HDP <sub>LTS</sub> . . . . .	32
5.1.2	Identifikovanie nelineárneho modelu časového radu HDP <sub>LTS</sub> pre $d = 3$ . . . . .	34
5.1.3	Model SETAR(5; 3, -4500) . . . . .	34
5.1.4	Model SETAR(5; 3, 4000) . . . . .	34
5.1.5	Model SETAR(5, 3; (-4500, 4000)) . . . . .	35
5.1.6	Identifikovanie nelineárneho modelu časového radu HDP <sub>LTS</sub> pre $d = 1$ . . . . .	35
5.1.7	Model SETAR(5; 1, -7000) . . . . .	35
5.1.8	Model SETAR(5; 1, 3000) . . . . .	36
5.1.9	Model SETAR(5, 1; (-7000, 3000)) . . . . .	36
5.2	Modely STAR . . . . .	36
5.2.1	Hľadanie najvhodnejšieho STAR modelu . . . . .	37
5.2.2	Najlepšie odhadnuté LSTAR modely . . . . .	37
	<b>Záver</b>	<b>39</b>
	<b>Literatúra</b>	<b>39</b>
	<b>Appendix</b>	<b>42</b>

# Úvod

V poslednom období sa problematika časových radov stala veľmi populárnou. Snaha spoznávať javy a zákonitosti do hĺbky a vedieť skúmané premenné predikovať, našla svoje opodstatnenie azda v najrôznejších oblastiach vedy. Dáta vo forme časových radov vznikajú vo fyzikálnych (napr. seizmický záznam v geofyzike), technických (napr. priebeh výstupného signálu určitého elektrotechnického zariadenia), biologických (napr. sledovanie znečistenia ovzdušia v ekológii) ale aj vo spoločenských vedách (napr. zmeny počtu a zloženia obyvateľstva sledované v demografii).

V súčasnej dobe, v oblasti praktického využitia časových radov, orientované najmä na ekonómiu, sa používajú rôzne metódy vyrovnania časových radov a eliminovania ich význačných zložiek, Boxova-Jenkinsova metodológia a spektrálna analýza. Cieľom analýzy časových radov je konštrukcia vhodného modelu. To predovšetkým umožní porozumieť mechanizmu, na základe čoho sú „generované“ sledované údaje. Môžeme si predstaviť, že znalosť modelu zodpovedá znalosti algoritmu, podľa ktorého počítač generuje dáta, pričom do algoritmu sú zapojené generátory náhodných čísel, čo zabezpečí náhodnosť celého procesu. Ďalej, znalosť modelu nám umožňuje predpovedať aj budúci vývoj systému.

V ideálnom prípade pozorovaný jav opísaný časovým radom vykazuje lineárny charakter, ale nie je tomu vždy tak, preto boli vytvorené aj nelineárne modely pre časové rady. Aj v tejto práci sa budeme zameriavať hlavne na tieto modely, presnejšie povedané na modely s rôznymi režimami, kde jednotlivé režimy sa určujú podľa istej pozorovateľnej, tzv. prahovej premennej, alebo prepnutie medzi režimami je náhodné, určené podľa Markovovho reťazca.

Práca sa skladá z úvodu, piatich kapitol a záveru. V prvej kapitole opíšeme modelovaný časový rad a oboznámime sa s ekonomickou históriou krajiny. Informácie o švédskej ekonomike sme získali z elektronických zdrojov [16], [17], [18], [19]. V druhej kapitole zadefinujeme jednotlivé pojmy, ktoré sa vyskytujú v teórii analýzy a modelovania časových radov. V tretej kapitole sa oboznámime s Boxovými-Jenkinsovými modelmi. K druhej a tretej kapitole sme čerpali teóriu hlavne z kníh Cipru [3] a Pekára [10]. Vo štvrtej kapitole opíšeme jednotlivé nelineárne modely. Potrebnú teóriu sme získali z článkov Chana, Wonga, Tonga [7], Luukkonena, Saikkonena, Teräsvirtu [9] a Tsaya [14], z knihy Pekára [10] a z diplomovej práce Petričkovej [11]. Piata kapitola je ťažiskom práce, tu budú aplikované v praxi vo štvrtej kapitole opísané SETAR a STAR modely. V závere zhrnieme výsledok našej snahy nájsť čo najvhodnejší model pre vybraný časový rad.

# Kapitola 1

## Časový rad HDP Švédska

Ekonomický rast každej krajiny vyjadrený rastom hrubého domáceho produktu (HDP) závisí od vicerých faktorov, ako sú štrukturálne zmeny v ekonomike krajiny, globálna recesia, politická nestabilita, prírodné katastrofy ale aj práve prebiehajúce fázy hospodárskeho cyklu. Zo štatistického hľadiska však môžeme štvrťročné údaje o HDP považovať za kombináciu štyroch, navzájom oddelených štatistických procesov, t.j. dlhodobého trendu, hospodárskeho cyklu, sezónnych prejavov a krátkodobých šokov, ktoré sa dajú od seba oddeliť pomocou štandardných štatistických metód.

### 1.1 Použitý časový rad

V tejto práci sa budeme venovať modelovaniu časového radu HDP Švédska. HDP máme dané vo forme kvartálnych dát, od začiatku roku 1980 až do konca roku 2008, čo znamená stošestnásť údajov.

### 1.2 Ekonomická história Švédska od 1980 až po krízu 90-tych rokov

Do polovice 19. storočia bolo Švédsko rozvojovou krajinou s hladujúcim obyvateľstvom. Avšak počas niekoľko desaťročí v priebehu 19. storočia poskytla skupina politikov hlásiacich sa ku klasickému liberalizmu Švédsku slobodu vierovyznania, prejavu, pohybu a taktiež ekonomickú slobodu, takže ľudia mohli začať podnikáť a slobodne obchodovať na trhu. Voľný obchod umožnil Švédom špecializovať sa na to v čom boli dobrí, teda na drevárske a železiarske odvetvia, a kupovať za tieto veci produkty potravinárskeho a strojárkeho priemyslu, ktoré Švédi tak lacne vyrábať nedokázali. Výsledkom bol ekonomický rast a industrializácia, ktoré umožnili vzrast blahobytu a investície do školstva a zdravotníctva. Medzi rokmi 1860-1910 vzrástla mzda v priemysle o 170%, teda omnoho viac ako v neskoršom období. Švédsko v tej dobe nebolo sociálnym štátom, malo skôr bližšie k minimálnemu štátu. Až do 1. svetovej vojny nebola spotreba verejného sektoru vyššia ako 6% HDP. Sociálni demokrati, ktorí sa chopili moci v roku 1932, pokračovali v liberálnom prístupe k podnikaniu a v politike voľného obchodu. Napriek tomu že vládne investície pomaly rástli, v roku 1950 bol



verejný sektor menší ako vo väčšine krajín, zhruba 25% HDP, približne ako v USA a Švajčiarsku. Ekonomike tiež prospievalo, že sa Švédsko nezúčastnilo oboch svetových vojen. Švédske podniky predávali obom stranám, priemysel nebol poškodený a mladí Švédi nezomierali v zákopoch. Medzi rokmi 1870-1970 bol švédsky rast po Japonsku najväčší na svete. V roku 1970 bolo Švédsko štvrtým najbohatším členom OECD, po USA, Luxembursku a Švajčiarsku. Lenže potom začal narastať sociálny štát „blahobytu“, umožňujúci politikom redistribuovať bohatstvo, ktoré bolo vytvorené jednotlivcami a trhom. Ekonomika pokračovala v raste; s ohľadom na počiatočné podmienky, dobrý priemysel, vzdelanie a intenzívne pracujúcich ľudí by tento vývoj mohla prekaziť snáď len plánovaná ekonomika. Rast bol ale pomalší ako v ostatných krajinách. Štát blahobytu ľahko spotreboval trhom vytvorené bohatstvo a skomplikoval vytvorenie nového. Od roku 1950 nevytvoril súkromný sektor (v čistom vyjadrení) ani jediné pracovné miesto, zatiaľ čo verejný sektor sa rozrástol najmenej o milión zamestnancov. Počas 70. rokov sa verejný sektor stával väčším a neproduktívnym a trh práce bol regulovaný. Od roku 1976 do roku 1982 vzrástla verejná spotreba z 50% na 65%. V rovnakom období musela byť päťkrát devalvovaná mena, celkom o 45%. Priemerné tempo rastu kleslo o polovicu na 2% v 70-tych rokoch a klesalo naďalej v 80-tych rokoch. V 90-tych rokoch došlo k veľkej kríze.

### 1.3 Švédska finančná kríza 90-tych rokov

Kríza bola vyvolaná monetárnym uvoľnením a následným vytvorením bubliny. Monetárne uvoľnenie alebo deregulácia tvorby úverov začala v roku 1985, kedy bolo zrušené kvantitatívne obmedzenie pre banky, ktoré ich obmedzovalo v tom, koľko úverov mohli poskytnúť. V tejto dobe bola švédska koruna zafixovaná ku košu mien. Finančná liberalizácia ovplyvnila správanie veriteľov ako aj dlžníkov, nakoľko sa z regulovaného úverového trhu, ktorý sa riadil hlavne politickými preferenciami, stal trh voľný. Na trhu tak začala obrovská úverová expanzia, ktorá zasahovala všetky oblasti ekonomiky, no najviac sektor domov, komerčných nehnuteľností a akciový trh. Boom podporovala aj vláda, nakoľko daňový systém favorizoval požičiavanie a reálna úroková miera tak bola záporná. Takýto vývoj zvyšoval cenu aktív, boom ekonomiky, rastúcu zamestnanosť, rastúcu spotrebu a znižujúcu sa mieru úspor, ktorá sa nakoniec stala tiež zápornou. Kvôli fixnému menovému kurzu nemohol byť tento boom zastavený zvyšovaním úrokových sadzieb (ináč by bolo možné realizovať úrokovú arbitráž). Možnosťou tak ostávala fiškálna politika (fiškálna kontrakcia), no tá, aj napriek tomu, že štátny rozpočet dosahoval vysoké prebytky, nebola dostatočne tvrdá aby boom zastavila. Tento vývoj bol ukončený viacerými domácimi aj medzinárodnými udalosťami v rokoch 1989-90. Po zjednotení Nemecka bola monetárna politika Bundesbanky utesňovaná, čo zvyšovalo reálne úrokové sadzby. Kvôli tomuto vývoju sa začal vytvárať tlak aj na švédske úrokové sadzby, nakoľko koruna bola naviazaná na ECU virtuálnu európsku menu. Ďalším šokom pre úrokové miery bolo ich zvyšovanie kvôli ochrane meny voči špekulačným útokom na menu z obdobia rokov 1989-92. Z hľadiska domácej ekonomiky ovplyvnila prasknutie úverovej bubliny daňová reforma zavedená v rokoch 1990-91. Tá výrazne zredukovala výšku splátok hypoték, ktoré si mohli domácnosti odpočítať z daní a znížila aj hraničné daňové sa-

dzby. Požičiavanie tak bolo menej atraktívne. Okrem toho výšku úrokových sadziieb ovplyvnil aj výrazný pokles inflácie. Reálna úroková sadzba tak vzrástla z úrovne -5% na úroveň +5%. Na finančný trh mal tento vývoj negatívny dopad a v celej ekonomike sa objavila deflácia aktív. Svoju situáciu sa snažili subjekty v ekonomike zlepšiť predajom aktív, no tí pokles cien nehnuteľností ako aj ostatných aktív sa len urýchlili. Domácnosti zas začali oveľa viac sporiť, čím sa ekonomický rast ocitol v kontrakcii. Na jeseň roku 1992 sa dostala mena pod veľký špekulatívny útok, kvôli ktorým bojovala v septembri banka zvýšením dennej úrokovej sadzby (overnight) až na 500%, aby udržala naviazanosť na ECU. Jej obrana parity bola veľmi vytrvalá, no v novembri Riksbank nakoniec situácii podľahla a koruna prešla do režimu floatingu (voľne plávanie meny). Nakoľko krajina získala zrušením naviazanosti a depreciáciou meny konkurenčnú výhodu pri exporte, označuje sa tento krok za štartér obnovy ekonomického rastu. Počas krízy totiž ekonomický rast poklesol kumulatívne o 5% a celková zamestnanosť o 10%. Celková finančná a ekonomická kríza vážne ohrozila fungovanie hlavných bánk v krajine.

Švédskej ekonomike sa nakoniec podarilo opäť naštartovať svoj rast. Kríza ale nebola pre Švédsko dostatočným poučením. Počas posledného úverového šialenstva totiž švédske banky rozširovali svoje aktivity po nordickom a baltickom regióne a stali sa jedným z najväčších veriteľov v tejto oblasti. Po tom, čo sa prevalila aktuálna finančná kríza, musela vláda opäť pristúpiť ku garancii pôžičiek bánk v objeme 158 mld. EUR a vytvorila stabilizačný fond v objeme 1,4 mld. EUR určený na prípadné kapitálové investície do bánk. Tri zo štyroch najväčších švédskych bánk už v reakcii na krízu museli získavať kapitál zo súkromných zdrojov.

## 1.4 Ekonomika v Švédsku po „velkej“ kríze

Obdobie od roku 1994 je charakteristické dynamickým ekonomickým rastom, ktorý patrí k najväčším v EU. Táto skutočnosť prispela k výraznému poklesu miery nezamestnanosti. Miera inflácie sa stabilizovala na úrovni zodpovedajúcej ostatným členským krajinám EU. Napriek tomu, že rast ekonomiky spomalila v prvej polovici roku 2001, dlhodobé predpovede rastu zostali priaznivé. Hlavným motorom hospodárskeho rastu sa ukázal export. V období 1995-2003 ale ceny exportu boli znížené o 4%, kým import ceny sa zvýšili o 11%, čo spôsobilo nepriaznivý vývoj obchodu. Vládny rozpočet sa zlepšil oproti stavu roku 1993, z hladiny deficitu 12% HDP k prebytku 8% HDP v 2001. Nový, prísny rozpočtový proces určený parlamentom a zmena ústavy značne zlepšili dôveryhodnosť politiky. Kým v roku 1994 úroveň hrubého štátneho dlhu bol 78% HDP, v roku 2000 sa táto úroveň dostala pod kľúčovú hodnotu 60% HDP, a v roku 2007 už bol len na úrovni 37% HDP. Čaká sa, že na rok 2010 sa to zníži až na úroveň 20% HDP.

## 1.5 Vplyv svetovej finančnej krízy na Švédsko

Švédsko zvládlo finančnú krízu najlepšie z pätnástich najbohatších krajín EU, na rebríčku ekonomickej prosperity organizácie OECD sa umiestnilo na jedenástom mieste.

Švédska ekonomika klesla v roku 2009 o 5%. Ide o najväčší pokles od doby pred druhou svetovou vojnou. Súčasne vzrástla nezamestnanosť na 8% a podľa očakávaní by sa to malo naďalej zvyšovať. Jedným z dôvodov švédskeho úspechu je to, že ako predtým, tak aj súčasnej vláde sa podarilo dosiahnuť rozpočtový prebytok, a to umožnilo stimulovať ekonomiku. Podľa odhadu ministra financií v roku 2010 sa očakáva rast 3%, a na ďalší rok 3.6%. Nezamestnanosť v roku 2010 by mala dosiahnuť hodnotu 9.5% a v roku 2011 by malo klesnúť na 8.9%.

# Kapitola 2

## Teória

### 2.1 Základné pojmy

V tejto kapitole zavedieme základné pojmy, s ktorými Boxova-Jenkinsova metodológia pracuje.

**Definícia 1.** Rad náhodných premenných  $\{X_t, t \in T\}$ , usporiadaných v čase, kde parameter  $t$  voláme čas, voláme *náhodný proces*  $X$ . Ak  $T$  je interval, tak náhodný proces voláme *spojitý*, ak  $T$  je podmnožinou množiny prirodzených čísiel, tak náhodný proces voláme *diskrétny*. Reálnu funkciu  $x = x(t), t \in T$ , kde  $x(t)$  je realizáciou  $X(t)$  v čase  $t$  voláme *realizáciou náhodného procesu*  $X$ .

Na začiatku skúmaní teórie o časových radoch by nás zaujímala otázka, čo vlastne je *časový rad*? Z matematického hľadiska je to usporiadaný súbor dát  $(x_1, x_2, \dots, x_n)$ , pozorovania náhodnej premennej  $X$ . Dolný index pozorovania  $x_t$  reprezentuje čas  $t = 1, 2, \dots, n$ , kde  $n$  predstavuje dĺžku daného časového radu. Je pochopiteľné uvažovať, že pozorovania  $x_t$  majú stochastický charakter, a preto  $(x_1, x_2, \dots, x_n)$  budeme považovať za jednu realizáciu stochastického procesu  $(X_1, X_2, \dots, X_n)$ .

#### 2.1.1 Stacionarita

Špeciálnym prípadom náhodných procesov sú stacionárne náhodné procesy. Všeobecne povedané, stacionarita časového radu znamená, že chovanie tohto radu je v istom zmysle ustálené. Rozlišujeme slabú a ostrú stacionaritu, ich definície sú nasledujúce:

**Definícia 2.** Náhodný proces  $X$  nazývame *slabo (kovariančne) stacionárny* ak prvý a druhý moment procesu existujú a sú konečné, majú konštantné hodnoty v čase a kovariančná funkcia závisí len od časovej vzdialenosti náhodných premenných, t.j.

$$E(X_t) = \mu \quad \text{pre všetky } t \in T, \quad (2.1a)$$

$$V(X_t) = \sigma \quad \text{pre všetky } t \in T, \quad (2.1b)$$

$$\text{cov}(X_t, X_s) = \text{cov}(X_{t+k}, X_{s+k}) \quad \text{pre všetky } k \in Z, s \neq t. \quad (2.1c)$$

Silná stacionarita zase hovorí, že pravdepodobnostné správanie daného stochastického procesu je časovo invariantné.

**Definícia 3.** Náhodný proces  $X$  nazývame *ostro stacionárny*, ak pre akékoľvek reálne číslo  $k$  platí

$$F(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = F(x_{t_1+k}, x_{t_2+k}, \dots, x_{t_n+k}), \quad (2.2)$$

kde  $F(\cdot)$  je združená distribučná funkcia.

### 2.1.2 Momenty náhodných procesov

Každý náhodný proces môžeme opísať pomocou nasledujúcich charakteristík:

- stredná hodnota

$$\mu_t = E(X_t) \quad (2.3)$$

- rozptyl (druhý centrálny moment)

$$\sigma_t^2 = V(X_t) = E(X_t - \mu_t)^2 \quad (2.4)$$

- šikmosť (tretí centrálny moment)

$$\mu_t^{(3)} = E(X_t - \mu_t)^3 \quad (2.5)$$

- strmosť (štvrtý centrálny moment)

$$\mu_t^{(4)} = E(X_t - \mu_t)^4 \quad (2.6)$$

- kovariančná funkcia

$$\gamma_{t,t-k} = \text{cov}(X_t, X_{t-k}) = E(X_t - \mu_t)(X_{t-k} - \mu_{t-k}) \quad (2.7)$$

- korelačná funkcia v bode  $k$  je definovaná vzťahom

$$\rho_{t,t-k} = \frac{\gamma_{t,t-k}}{\sqrt{\sigma_t^2} \sqrt{\sigma_{t-k}^2}} \quad (2.8)$$

Výberové momenty náhodných procesov zase môžeme charakterizovať pomocou nasledovných charakteristík:

- výberový priemer

$$\bar{x}_t = \frac{1}{N} \sum_{t=1}^N x_t \quad (2.9)$$

kde  $N$  je počet hodnôt časového radu, je odhadom strednej hodnoty  $\mu$ ;

- výberový rozptyl

$$\hat{s}^2 = \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2 \quad (2.10)$$

je odhadom rozptylu procesu  $\sigma^2$ ;

- výberová šikmosť

$$\hat{S} = \frac{1}{N} \sum_{t=1}^N \left( \frac{x_t - \bar{x}}{\hat{s}} \right)^3 \quad (2.11)$$

je odhadom tretieho centrálneho momentu procesu  $\mu^{(3)}$ ;

- výberová strmosť

$$\hat{K} = \frac{1}{N} \sum_{t=1}^N \left( \frac{x_t - \bar{x}}{\hat{s}} \right)^4 \quad (2.12)$$

je odhadom štvrtého centrálneho momentu procesu  $\mu^{(4)}$ ;

- výberová autokorelačná funkcia (ACF) s posunutím  $k$

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (2.13)$$

kde  $k = 1, 2, \dots, N - 1$ , je odhadom kovariančnej funkcie  $\rho_k$  a je daná výberovou autokoreláciou a výberovým rozptylom. Chovanie autokorelačnej funkcie je v rámci Boxovej-Jenkinsovej metodológie dôležitým ukazovateľom, lebo predpovedá, aký typ modelu je vhodné pre daný časový rad použiť.

- výberová parciálna autokorelačná funkcia (PACF)  $r_{kk}$  je definovaná rekurentným vzťahom

$$\begin{aligned} r_{11} &= r_1, \\ r_{kk} &= \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \quad \text{pre } k > 1, \end{aligned} \quad (2.14)$$

kde

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad \text{pre } j = 1, 2, \dots, k-1. \quad (2.15)$$

### 2.1.3 Biely šum

Základným prvkom pre vybudovanie nami použitej teórie sú nezávislé a rovnako rozdelené, tzv. i.i.d. (independent identically distributed) stochastické procesy. 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 často príliš silné, a preto pre ďalšie účely vystačíme s bielym šumom, ktorý definujeme nasledovne:

**Definícia 4.** Náhodný proces  $E$ , v ktorom  $E_t$  sú nekorelované náhodné premenné s rovnakým rozdelením pravdepodobnosti sa nazýva *biely šum*, ak platí, že

$$E(E_t) = \mu \quad \text{obvykle } \mu = 0, \quad (2.16a)$$

$$V(E_t) = \sigma^2, \quad (2.16b)$$

$$\text{cov}(E_t E_s) = 0 \quad \forall s \neq t. \quad (2.16c)$$

Ak  $E_t$  majú normálne rozdelenie  $N(0,1)$ , biely šum voláme *gaussovský*.

**Poznámka 1.** *Biely šum je stacionárny, centrovany náhodný proces.*

### 2.1.4 Operátor posunutia

Aby sa dali časové rady jednoduchšie a zjednotenejšie zapísať, zadefinujeme si operátor posunutia.

**Definícia 5.** Predpokladajme, že  $X$  je náhodný proces. Potom operátor posunutia  $B$  je daný vzťahom

$$Bx_t = x_{t-1}, \quad (2.17)$$

a operátor spätnej diferencie vzťahom

$$\Delta x_t = (1 - B)x_t = x_t - x_{t-1}. \quad (2.18)$$

**Poznámka 2.** *Oba operátory môžeme aplikovať viackrát za sebou, takže dostávame operátor posunutia rádu  $s$*

$$B^s x_t = x_{t-s} \quad \forall s \in \mathbb{N}, \quad (2.19)$$

a operátor spätnej diferencie rádu  $s$

$$\Delta^s x_t = (1 - B)^s x_t \quad \forall s \in \mathbb{N}. \quad (2.20)$$

### 2.1.5 Waldova reprezentácia

**Definícia 6.** Náhodný proces, ktorého stredná hodnota sa rovná nule, t.j.  $\mu = 0$ , nazývame *centrovany*.

**Veta 1.** (*Waldova reprezentácia*) *Každý stacionárny centrovany proces  $x_t$  ktorý neobsahuje deterministickú zložku, môžeme vyjadriť ako lineárnu kombináciu radu nekorelovaných rovnako rozdelených náhodných premenných, ktorá sa označuje ako Waldova reprezentácia a matematicky sa dá zapísať ako*

$$x_t = \sum_{i=0}^{\infty} \theta_i E_{t-i}, \quad \theta_0 = 1, \quad \sum_{i=0}^{\infty} \theta_i^2 < \infty, \quad (2.21)$$

kde  $E(E_t) = 0$ ,  $E(E_t E_{t-i}) = 0 \quad \forall i \neq 0$  a  $E(E_t^2) = \sigma^2$ .

## 2.2 Dekompozícia časového radu

Niektoré časové rady, hlavne z ekonomickej praxe, môžeme rozložiť na niekoľko zložiek, ktorými sú

- dlhodobý trend  $T_t$
- cyklická zložka  $C_t$
- sezónna zložka  $S_t$
- reziduálna zložka  $I_t$

Dekompozícia časového radu je motivovaná tým, že niekedy v jednotlivých zložkách ľahšie odhalíme pravidelné správanie sa časového radu, ako v prípade nerozloženého časového radu. Aby sme mali o dekompozícii istú predstavu, charakterizujeme jej jednotlivé zložky.

*Trend* odráža dlhodobé zmeny v priemernom správaní sa časového radu (napr. dlhodobý rast alebo dlhodobý pokles). Môžeme si predstaviť, že trendová zložka vzniká v dôsledku systematického pôsobenia síl v tom istom smere. Poznáme viac typov trendov, ako napr. konštantný, lineárny, kvadratický, exponenciálny a logaritmický trend. Typ najvhodnejšej matematickej krivky pre túto uvažovanú zložku určíme na základe predbežného rozboru časového radu, najčastejšie nám pomôže grafické znázornenie uvažovaného časového radu, alebo určíme na základe predpokladaných vlastností trendovej zložky, ktoré vyplývajú z ekonomickej teórie.

*Sezónna zložka* popisuje periodické zmeny v časovom rade, ktoré sa odohrávajú behom jedného kalendárneho roku a každý rok sa opakujú. Pre skúmanie sezónnej zložky sú vhodné najmä mesačné alebo kvartálne dáta. Rozbor eliminovanej sezónnej zložky môže podstatne rozšíriť naše znalosti o zákonitostiach chovania učitého ekonomického javu a prispieť ku konštrukcii dokonalejšej predpovede v uvažovanom časovom rade (napr. rozbor sezónnych zmien v počte cestujúcich leteckou spoločnosťou). Dôležitým cieľom je sezónne očistenie časového radu, najlepšie je, ak sa nám podarí eliminovať túto zložku celkom z časového radu, alebo aspoň potlačiť jej vplyv na minimum.

*Cyklická zložka* je najspornejšou zložkou časového radu. Niektorí autori sa vyhýbajú nazvať túto zložku cyklickou, hovoria radšej o fluktuáciách okolo trendu, kde sa strieda fáza rastu a poklesu. Dĺžka jednotlivých cyklov je obvykle premenlivá, a takisto môže byť premenlivá aj intenzita jednotlivých fáz cyklickeho priebehu. Pre západnú ekonomiku je typickým predstaviteľom tejto zložky tzv. „business cycle“, ktorý charakterizuje rast a pokles ekonomickej aktivity, jeho dĺžka je v rozmedzí 5-7 rokov.

*Reziduálna zložka* zostáva v časovom rade po odstránení trendu, sezónnej a cyklickej zložky. Je tvorená náhodnými fluktuáciami v priebehu časového radu, ktoré



nemajú systematický charakter. Reziduálna zložka takisto pokrýva aj chyby v meraní údajov a aj chýb, ktorých sa dopustíme pri analýze časového radu. Väčšinou sa predpokladá, že reziduálna zložka je biely šum s normálnym rozdelením.

Skúmaný časový rad na základe predošlého výkladu môžeme predstaviť ako trend, na ktorý sú „nabalené“ periodické zložky (sezónna a cyklická zložka) a biely šum. Tvar rozkladu môže byť dvojakeho typu. Aditívna dekompozícia má tvar

$$Y_t = T_t + C_t + S_t + It. \quad (2.22)$$

Pri aditívnom rozklade sú jednotlivé zložky uvažované vo svojich skutočných absolútnych hodnotách, a sú merané v jednotkách skúmaného časového radu  $Y_t$ .

Multiplikatívna dekompozícia má tvar

$$Y_t = T_t \cdot C_t \cdot S_t \cdot It. \quad (2.23)$$

Pri multiplikatívnom rozklade je väčšinou len tendová zložka uvažovaná v absolútnej hodnote, meraná v jednotkách  $Y_t$ . Ostatné zložky sú uvažované v relatívnych hodnotách oproti trendu, a sú bezrozmernými veličinami.

## 2.3 Testovanie hypotéz a rozhodovacie kritériá

**Akaikeho informačné kritérium (AIC)** sa používa pri výbere z alternatívnych modelov, pričom sa preferuje jeho čo najnižšia hodnota. Počíta sa podľa vzorca

$$AIC = -\frac{2l}{N} + \frac{2k}{N}, \quad (2.24)$$

pričom

- $l$  je hodnota logaritmickej funkcie maximálnej vierohodnosti,
- $k$  je počet odhadovaných parametrov,
- $N$  je počet meraní.

**Chowov test zlomu** Test sa používa na testovanie štrukturálnej zmeny (napr. vojna, prírodná katastrofa) v časovom rade. Testuje sa tu konkrétny bod zlomu v dátach, teda test je vhodný na overenie hypotézy o zmene chodu ekonomiky v prípade ekonomických časových radov. Nulová hypotéza testu je, že nami rozdelené dáta vyhovujú tomu istému modelu, teda odhady parametrov modelov sú totožné. Pri zamietnutí nulovej hypotézy sa prijíma alternatívna hypotéza, že v časovom rade nastal štrukturálny zlom. Testovacia štatistika je

$$F = \frac{(SSR_R - SSR_U)/k}{SSR_U/(N - k)} \sim F_{k, n-2k}, \quad (2.25)$$

kde

- $SSR_U$  je suma štvorcov reziduí z modelu bez reštrikcie,
- $SSR_R$  je suma štvorcov reziduí z modelu s reštrikciou,
- $N$  je počet meraní,
- $k$  je počet odhadovaných parametrov.

**Jarqueho-Berov test** Tento test slúži na testovanie, či rad má normálne rozdelenie. Budeme ho používať na testovanie normality reziduí. Testovacia štatistika meria rozdiely medzi šikmosťou a strmosťou meraného radu a normálneho rozdelenia. Štatistika sa počíta podľa vzorca

$$JB = \frac{N - k}{6} (\hat{S}^2 + \frac{1}{4} (\hat{K} - 3)^2), \quad (2.26)$$

kde

- $\hat{S}$  je odhad tretieho centrálného momentu daný výberovou šikmosťou (2.11),
- $\hat{K}$  je odhad štvrtého centrálného momentu daný výberovou strmosťou (2.12),
- $N$  je počet meraní,
- $k$  je počet odhadovaných parametrov.

Nulová hypotéza  $H_0$  pri tomto teste je, že rad má normálne rozdelenie proti alternatívnej hypotéze  $H_A$ , ktorá predpokladá opak  $H_0$ .

JB štatistika má  $\chi^2$  rozdelenie s dvoma stupňami voľnosti,  $p$ -hodnota testu zase udáva pravdepodobnosť, že štatistika v absolútnej hodnote prekročí hodnotu danú nulovou hypotézou. Ak testom získaná  $p$ -hodnota bude nižšia, než  $\alpha = 0.05$ , tak normalitu uvažovaného radu zamietame.

**Ljungova-Boxova Q-štatistika**  $Q$ -štatistika s posunutím  $p$  je testovacia štatistika nulovej hypotézy, že v rade nie je autokorelácia až po rád  $p$ , t.j.  $H_0$ : v rade je autokorelácia až po rád  $p$  proti alternatíve  $H_A$ : opak  $H_0$ . Testovacia štatistika má tvar

$$Q_{LB} = N(N + 2) \sum_{j=1}^p \frac{r_j^2}{N - j}, \quad (2.27)$$

kde

- $r_j$  je koeficient autokorelácie s posunutím  $j$ ,
- $N$  je počet meraní.

$Q$  má asymptoticky  $\chi^2$  rozdelenie s počtom stupňov voľnosti rovným počtu autokorelácií.

**Miery presnosti vyrovnávaní** Pri odhade parametrov modelu budeme zisťujú tieto chyby:

- Koeficient determinácie

$$RS = R^2 = 1 - \frac{\sum_{t=1}^N (x_t - \hat{x}_t)^2}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (2.28)$$

vyjadruje mieru zhody meraných a modelovaných údajov.

- Upravený koeficient determinácie

$$ARS = \bar{R}^2 = 1 - (1 - R^2) \frac{N - 1}{N - k} \quad (2.29)$$

kde

- $R^2$  je koeficient determinácie,
- $N$  je počet meraní,
- $k$  je počet odhadovaných parametrov,

vyjadruje mieru zhody meraných a modelovaných údajov, penalizujúci zaradenie regresorov, ktoré neprispievajú k vysvetľovaniu modelu.

- Štandardná chyba regresie

$$SER = \hat{s} = \sqrt{\frac{\sum_{t=1}^N (x_t - \hat{x}_t)^2}{N - k}}, \quad (2.30)$$

kde

- $N$  je počet meraní,
- $k$  je počet odhadovaných parametrov,

- Súčet štvorcov reziduí

$$SSR = \sum_{t=1}^N (x_t - \hat{x}_t)^2 \quad (2.31)$$

je kvantitatívnym vyjadrením presnosti vyrovnania.

**Rozšírený Dickeyov-Fullerov test jednotkového koreňa** Používame na testovanie, či skúmaný rad je stacionárny, a v prípade, že je stacionárny, používame na určenie rádu integračného člena. Test kontroluje korelácie vyšších rádo pri dávajúc členy posunutých spätných diferencií na pravej strane regresie

$$\Delta x_t = \mu + \gamma x_{t-1} + \delta_1 \Delta x_{t-1} + \delta_2 \Delta x_{t-2} + \dots + \delta_{p-1} \Delta x_{t-p+1} + \varepsilon_t. \quad (2.32)$$

Test spočíva v testovaní nulovej hypotézy  $H_0 : \gamma = 0$  proti alternatívnej hypotéze  $H_A : \gamma < 0$ . Štatistiku získanú z tohoto testu porovnávame s MacKinnovými kritickými hodnotami, a v prípade, že testovacia štatistika je menšia než MacKinnova kritická hodnota, hypotézu zamietame.

**Test významnosti ACF a PACF** Ak ACF leží vnútri intervalu  $(-\frac{2}{\sqrt{N}}, \frac{2}{\sqrt{N}})$ , kde  $N$  je počet meraní, tak nie je na hladine významnosti  $\alpha = 0.05$  významne rozdielna od nuly. To isté platí aj pre PACF.

# Kapitola 3

## Boxove-Jenkinsove modely

Ako úvod k štúdiu tejto metodológie zhrnieme jej hlavné výhody a nevýhody. Začneme s pozitívnymi vlastnosťami:

1. Boxove-Jenkinsove modely rýchlo reagujú na zmeny v charaktere modelovaného procesu. Sú to stochastické modely, ktoré dokážu stochasticky modelovať aj trend a sezónnosť. Preto sú často schopné popísať aj také časové rady, na ktoré klasická analýza časových radov nestačí.
2. Existuje postup, ako zostaviť pre daný časový rad model tohoto typu. Tento postup môžeme rozdeliť do troch základných krokov: identifikácia, odhad parametrov a overovanie modelu.

Hlavnými nedostatkami modelu sú:

1. Pre vybudovanie spoľahlivého modelu podľa tejto metodológie je potrebné mať časový rad dĺžky aspoň 50 pozorovaní, takže tento model je vhodný predovšetkým na modelovanie mesačných alebo kvartálnych dát.
2. Praktická aplikácia Boxovo-Jenkinsovej metodológie je často nákladná a časovo náročná. Ale našťastie pre túto problematiku už existuje veľa dobrých programov, ktoré modelovanie časového radu značne uľahšia.
3. Nezanedbateľnou nevýhodou týchto modelov je, že často stratíme jednoduchú interpretáciu modelovaného javu.

V ďalšej časti kapitoly uvedieme definície základných prvkov tejto metodológie, ide o AR a MA procesy. Na základe týchto „stavebných prvkov“ už vieme opísať rôzne varianty Boxovo-Jenkinsových modelov, akými sú modely ARMA, ARIMA, SARIMA.

### 3.1 Procesy AR( $p$ )

Autoregresný centrováný proces  $p$ -teho rádu AR( $p$ ) môžeme vyjadriť vzťahom:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + E_t = \sum_{i=1}^p \phi_i X_{t-i} + E_t, \quad (3.1)$$

čo môžeme zapísať aj v operátorovom tvare

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = E_t, \quad (3.2)$$

pričom B je operátor posunu

$$B^i X_t = X_{t-i}. \quad (3.3)$$

Ak proces  $X_t$  nie je centrovateľný, tak AR( $p$ ) má tvar

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + E_t = c + \sum_{i=1}^p \phi_i X_{t-i} + E_t, \quad (3.4)$$

kde konštanta  $c$  je rovná

$$c = \mu - \phi_1 \mu - \phi_2 \mu - \dots - \phi_p \mu. \quad (3.5)$$

**Veta 2.** Náhodný proces  $X$  daný vzťahom (3.1) je stacionárny práve vtedy, ak všetky korene polynomickej rovnice

$$\phi_p(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0 \quad (3.6)$$

ležia mimo jednotkového kruhu.

## 3.2 Procesy MA( $q$ )

Proces klzavých priemerov  $q$ -teho rádu MA( $q$ ) môžeme vyjadriť vzťahom:

$$X_t = E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2} + \dots + \theta_q E_{t-q} = E_t + \sum_{j=1}^q \theta_j E_{t-j}, \quad (3.7)$$

čo môžeme zapísať aj v operátorovom tvare

$$X_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) E_t, \quad (3.8)$$

pričom B je operátor posunu.

Proces MA( $q$ ) je stacionárny a platí

$$\mu_t = 0, \quad \sigma^2 = \sigma_E^2 \left( 1 + \sum_{j=1}^q \theta_j^2 \right).$$

**Definícia 7.** Náhodný proces MA( $q$ ), ktorý sa dá zapísať v tvare konečného AR procesu, voláme invertovateľný.

**Veta 3.** Náhodný proces MA( $q$ ), definovaný vzťahom (3.7) je invertovateľný práve vtedy, ak všetky korene polynomickej rovnice

$$\theta_q(x) = 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q = 0 \quad (3.9)$$

ležia mimo jednotkového kruhu.

### 3.3 Procesy ARMA( $p, q$ )

ARMA( $p, q$ ) procesy sú kombináciou AR( $p$ ) a MA( $q$ ) procesov, a dajú sa vyjadriť v tvare

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + E_t + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q} \\ &= \sum_{i=1}^p \phi_i X_{t-i} + E_t + \sum_{j=1}^q \theta_j E_{t-j}, \end{aligned} \quad (3.10)$$

alebo pomocou operátorov

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) E_t. \quad (3.11)$$

**Veta 4.** *Náhodný proces ARMA( $p, q$ ), definovaný vzťahom (3.10) je stacionárny, práve vtedy ak všetky korene polynomickej rovnice (3.6) ležia mimo jednotkového kruhu a je invertovateľný práve vtedy, ak všetky korene polynomickej rovnice (3.9) ležia mimo jednotkového kruhu.*

### 3.4 Integrované procesy ARIMA( $p, q$ )

Ak máme časový rad, ktorý nie je centrováný, prípadne jeho trend je polynomický, na modelovanie používame autoregresné integrované modely kľzavých priemerov ARIMA( $p, d, q$ ), ktoré sú založené na  $d$  násobnom diferencovaní uvažovaného časového radu. Diferencovaný proces označme  $W_t$ . Potom máme

$$W_t = \nabla^d X_t = (1 - B)^d X_t \quad (3.12)$$

kde  $\nabla = (1 - B)$  je operátor spätnej diferencie. Teda ARIMA( $p, d, q$ ) môžeme vyjdiť v tvare

$$\begin{aligned} \phi_p(B) W_t &= \theta_q(B) E_t, \\ \phi_p(B) (1 - B)^d X_t &= \theta_q(B) E_t, \end{aligned} \quad (3.13)$$

kde

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \theta_q(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q. \end{aligned}$$

Proces ARIMA( $p, d, q$ ) nie je stacionárny, lebo na ľavej strane vzorca (3.13) je  $d$  koreňov na jednotkovej kružnici.

### 3.5 Sezónne procesy SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$

V prípade, že časový rad vykazuje sezónnosť, t.j. závislosť medzi zodpovedajúcimi veličinami v jednotlivých sezónach, teda medzi  $\dots, X_{t-2s}, X_{t-s}, X_t, X_{t+s}, X_{t+2s} \dots$ , kde  $s$  je dĺžka periódy, použijeme sezónny autoregresný integrovaný proces kľzavých priemerov SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$ , kde

- $p$  je rád procesu AR,
- $d$  je rád diferencie vnútri periódy,
- $q$  je rád procesu MA,
- $P$  je rád sezónneho AR procesu,
- $D$  je rád sezónnej diferencie,
- $Q$  je rád sezónneho MA procesu,
- $s$  je dĺžka sezónnej periódy.

Proces môžeme vyjadriť v tvare

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)E_t, \quad (3.14)$$

kde

$$\Phi_P(B) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}), \quad (3.15)$$

$$\Theta_Q(B) = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}), \quad (3.16)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad (3.17)$$

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q. \quad (3.18)$$

Sezónnosť v časových radoch môžeme identifikovať pomocou periodogramov alebo pomocou autokorelačných funkcií.



# Kapitola 4

## Nelineárne modely

V tejto kapitole sa budeme zaoberať tromi typmi modelov, ktoré sa najčastejšie používajú v nelineárnom modelovaní časových radov. Zakladajú sa na myšlienke, že uvažovaný proces prebieha v rôznych režimoch, pričom v každom z nich časový rad sa dá modelovať rôznymi lineárnymi autoregresnými modelmi.

Najjednoduchším modelom z nelineárnych modelov sú samobudiacie prahové autoregresné modely, v ktorých indikátorom typu režimu je niektorá z endogénnych premenných modelu, zvyčajne je to  $d$ -te omeškanie procesu, a samotné režimy určia isté vybrané hodnoty pre tieto premenné, tzv. prahy. Pri tejto metóde azda najťažšou úlohou je určiť hodnotu omeškania a hodnoty prahov.

V mnohých prípadoch však prudká zmena pri prechode medzi režimami nezodpovedá modelovanému javu. Tento nedostatok sa snaží odstrániť autoregresný model hladkého prechodu, kde sa do jednotlivých režimov vstupuje pomocou váhových funkcií, najčastejšie ide o logistickú a exponenciálnu funkciu. Aj v prípade tohoto modelu, indikátorom režimu je niektorá endogénna premenná, najčastejšie niektoré omeškanie samotného procesu, a takisto sa musia určiť aj hodnoty prahov.

Ako posledný typ, uvádzme Markovov model prepínania režimov. Tento model sa líši od predošlých v tom, že zmenu režimu neriadi nejaká pozorovateľná veličina, ale prepnutie medzi režimami je náhodná, určená podľa Markovovho reťazca.

### 4.1 Samobudiacie prahové autoregresné modely

#### SETAR( $p; d, r$ )

Prahové autoregresné modely, tzv. TAR modely sú základom viacrežimových modelov. Navrhol ich Tong v roku 1978. Tieto modely predpokladajú, že režim, ktorý nastane v čase  $t$ , môže byť určený nejakou pozorovateľnou premennou  $u_t$ . Hodnoty tejto tzv. prahovej (threshold) premennej sú porovnávané s tzv. prahovou hodnotou  $r$ . Vo všeobecnom tvare model môžeme zapísať ako

$$X_t = \sum_{j(t)=1}^k \left( \phi_0^{j(t)} + \sum_{i=1}^p \phi_i^{j(t)} X_{t-i} + E_t^{j(t)} \right), \quad (4.1)$$

pričom  $j(t) = 1$  ak  $g(u_t) < r_1$ ,  $j(t) = 2$  ak  $r_1 \leq g(u_t) < r_2$ , ...,  $j(t) = k$  ak  $r_{k-1} < g(u_t)$  a  $E_t^{j(t)}$  je biely šum. Premennú  $u_t$  nazývame prahovou premennou, parametre  $r_t$  nazývame prahmi a hodnoty  $j(t)$  voláme režimami. Proces daný vzťahom (4.1) rozkladá časovú os na  $k$  režimov, kde jednotlivé režimy sú určené iným lineárnym autoregresným modelom. Proces  $X_t$  je nelineárny, ak sú v ňom definované aspoň dva rôzne režimy s rôznymi lineárnymi procesmi.

V prípade, že za premennú  $u_t$  budeme uvažovať oneskorené hodnoty analyzovaného časového radu  $X_{t-d}$ , kde  $d > 0$  je číslo udávajúce oneskorenie, tak dostávame samobudiaci prahový autoregresný model SETAR( $p; d, r$ ) (Self-Exciting Threshold Autoregressive Model), kde  $p$  je maximálny rád autoregresie,  $d$  je hodnota omeškania a  $r$  je vektor prahov. Tento model môžeme zapísať nasledujúcim spôsobom

$$X_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^p \phi_i^{(1)} X_{t-i} + E_t^{(1)} & \text{ak } X_{t-d} < r_1 \\ \phi_0^{(2)} + \sum_{i=1}^p \phi_i^{(2)} X_{t-i} + E_t^{(2)} & \text{ak } r_1 \leq X_{t-d} < r_2 \\ \vdots & \vdots \\ \phi_0^{(k)} + \sum_{i=1}^p \phi_i^{(k)} X_{t-i} + E_t^{(k)} & \text{ak } r_{k-1} \leq X_{t-d} \end{cases} \quad (4.2)$$

kde  $k$  je počet režimov v modeli. Hodnota prahov spĺňa  $-\infty < r_1 < r_2 < \dots < r_k = \infty$ . Inovácie v  $i$ -tom režime  $E_t^{(i)}$  sú i.i.d. s nulovou strednou hodnotou a varianciou  $\sigma_i^2 < \infty$  pre  $i = 1, 2, \dots, k$ . Procedúra identifikácie SETAR modelu je nasledujúca

- Voľba lineárneho AR procesu rádu  $p$  a množiny  $S$  možných hodnôt parametra omeškania  $d$ .
- Pomocou preusporiadanej bázevej základne procesu AR( $p$ ) a možných hodnôt omeškania  $d \in S$  otestovať prahovú nelinearitu štatistikou  $\hat{F}(p, d)$ . Ak nelinearita sa preukáže, tak určiť parameter omeškania  $d_p$ .
- Pre dané  $p$  a  $d_p$  určiť hodnoty prahov  $r$  z bodového grafu.
- Vylepšenie rádu AR procesu a hodnôt prahov v každom režime pomocou lineárnych autoregresných metód.

Napriek tomu, že určenie prahových premenných a hodnôt samotných prahov sú obtiažne, modely SETAR sa stali pomerne populárne pri modelovaní rôznych nelineárnych ekonomických a finančných časových radov.

#### 4.1.1 Preusporiadanie autoregresie a prediktívne reziduá

Uvažujme lineárny autoregresný proces AR( $p$ ), identifikovaný v uvažovanom rade  $\{X_1, X_2, \dots, X_n\}$ . Nech  $X_{t-d}$  je prahová premenná, ktorá nadobúda hodnoty  $\{X_h, \dots, X_{n-d}\}$ , kde  $h$  je definované vzťahom  $h = \max\{1, p+1-d\}$ . Nech  $\pi_i$  je časový index  $i$ -teho najmenšieho pozorovania z množiny možných hodnôt  $\{X_h, \dots, X_{n-d}\}$ . Potom v skonštruovaní preusporiadaneého autoregresného procesu podľa prahovej pre-

menšej  $X_{t-d}$  nám pomáha *rolling order autoregression* (ROA), ktorá má tvar

$$\begin{pmatrix} X_{\pi_1+d} \\ X_{\pi_2+d} \\ \vdots \\ X_{\pi_i+d} \end{pmatrix} = \begin{pmatrix} 1 & X_{\pi_1+d-1} & \cdots & X_{\pi_1} & \cdots & X_{\pi_1+d-p} \\ 1 & X_{\pi_2+d-1} & \cdots & X_{\pi_2} & \cdots & X_{\pi_2+d-p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & X_{\pi_i+d-1} & \cdots & X_{\pi_i} & \cdots & X_{\pi_i+d-p} \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} E_{\pi_1+d} \\ E_{\pi_2+d} \\ \vdots \\ E_{\pi_p+d} \end{pmatrix} \quad (4.3)$$

kde  $i = b, b+1, \dots, n-p$ , a  $b$  je počiatkový počet pozorovaní, ktorý berieme z časového radu. Tsay [14] vo svojej práci navrhuje ako počiatkovú hodnotu  $b \approx n/10 + p$ . Pre každé  $i$  vieme počítať prediktívne reziduá preusporiadanej autoregresie o jeden krok dopredu  $\hat{E}_{\pi_{i+1}+d}$ . Tieto reziduá získame tak, že postupne urobíme regresiu vždy z niekoľko prvých pozorovaní pre dané  $i$ . Získané reziduá použijeme k ďalšej regresii danej vzťahom

$$\hat{E}_{\pi_i+d} = \omega_0 + \sum_{\nu=1}^p \omega_{\nu} X_{\pi_i+d-\nu} + \delta_{\pi_i+d} \quad (4.4)$$

pre  $i = b+1, \dots, n-d-h+1$ . Už nám stačí vypočítať testovaciu štatistiku

$$\hat{F}(p, d) = \frac{(\sum_{t=b+1}^{n-d-h+1} \hat{E}_t^2 - \sum_{t=b+1}^{n-d-h+1} \hat{\delta}_t^2) / (p+1)}{(\sum_{t=b+1}^{n-d-h+1} \hat{\delta}_t^2) / (n-d-b-p-h)} \quad (4.5)$$

kde  $\hat{E}_t \equiv \hat{E}_{\pi_t+d}$  a  $\hat{\delta}_t \equiv \hat{\delta}_{\pi_t+d}$ , aby sme otestovali prahovú nelinearitu modelu.

**Veta 5.** (Tsay (1989)) *Predpokladajme, že  $X_t$  je lineárny stacionárny AR proces rádu  $p$ . Teda  $X_t$  spĺňa (4.2) pre  $k = 1$ . Potom pre veľké  $n$  štatistika  $\hat{F}(p, d)$  definovaná vzťahom (4.5) má asymptoticky Fisherovo  $F$  rozdelenie s  $(p+1)$  a  $(n-d-b-p-h)$  stupňami voľnosti.*

### 4.1.2 Identifikácia parametra omeškania $d$

Pri SETAR modeloch jedným z najťažších úloh je určenie parametra omeškania  $d$ , ktorý zohráva kľúčovú úlohu v nelineárnom charaktere modelu. Tong a Lim [13] odporúčajú na určenie hodnoty parametra  $d$  Akaikeho informačné kritérium až po identifikovaní ostatných parametrov modelu. Tsay [14] navrhol odlišnú procedúru, v istom zmysle heuristickú metódu voľby parametra  $d$ , založenú na analýze reálnych dát, kde parameter  $d$  sa špecifikuje ako prvý. Pre daný AR proces rádu  $p$ , najvhodnejší parameter omeškania  $d$  si vybereme podľa štatistiky danej vzťahom (4.5), a vyberáme podľa toho, kde pre možné hodnoty parametra  $d \in S = \{1, 2, \dots, p\}$  táto štatistika nadobúda svoje maximum, t.j.

$$\hat{F}(p, d_p) = \max_{\nu \in S} \{\hat{F}(p, \nu)\}, \quad (4.6)$$

$$d_p = \arg \max_{\nu \in S} \{\hat{F}(p, \nu)\}, \quad (4.7)$$

kde index  $p$  pri parametri  $d$  vyjadruje, že parameter závisí na ráde autoregresie.

### 4.1.3 Určenie hodnôt prahov $r$

Špecifikácia hodnôt prahov vyžaduje špeciálny prístup. Nech  $r_i$  je skutočná hodnota prahu vyhovujúca nerovnosti  $X_{\pi_i} < r_i < X_{\pi_{i+1}}$ . Potom každé reálne číslo z intervalu  $[X_{\pi_i}, X_{\pi_{i+1}}]$  môže byť dobrým odhadom hodnoty prahu, lebo určuje rovnaký SETAR model. Vo všeobecnosti, v literatúre nájdeme rôzne spôsoby intervalových a bodových odhadov hodnôt prahov. V práci použijeme metódy, ktoré pri bodových odhadoch používajú výberové empirické percentily. Tieto metódy lokalizujú prahy pomocou analýzy bodového grafu rôznych štatistík verzus špecifikovaná prahová premenná. Nie sú to formalizované metódy, avšak poskytujú informácie o možnej lokalizácii prahov. Štandardne sa používa metóda, ktorú navrhli Tong a Lim [13], ktorá analyzuje bodový graf prediktívnych reziduí získaných v rekurzívnych lineárnych regresiach pri teste nelinearity versus príslušnú prahovú premennú modelu  $X_{t-d_p}$ . Tsay [14] ako porovnávacie štatistiky berie t-štatistiky jednotlivých autoregresných koeficientov získaných pri rekurzívnych regresiach. Vychádza z úvahy, že priebeh hodnôt prediktívnych reziduí aj t-štatistík sa vychyluje z pôvodného smeru, ak hodnota prahovej premennej dosahuje prah. Graf prediktívnych reziduí dokáže lokalizovať prahy, a grafy t-štatistík okrem toho identifikujú aj významnosť jednotlivých autoregresných koeficientov.

## 4.2 Autoregresné modely hladkého prechodu STAR

Najväčším nedostatkom samobudiacich prahových autoregresných modelov SETAR je ostrý zlom modelu pri prechode prahom. Tento nedostatok sa snaží odstrániť autoregresný model hladkého prechodu STAR (Smooth Transition Autoregressive Model), o ktorého popularitu sa zaslúži najmä Granger a Teräsvirta [4]. V tomto modeli uvažujeme dva režimy s váženým prístupom do režimov, ktorý závisí na realizácii nejakej premennej  $u_t$ . Vážený prístup do režimov nám zabezpečuje spojitá prechodová funkcia  $G(u_t; \gamma, c)$ , ktorá vyjadruje váhu druhého režimu, a jej hodnoty sú medzi 0 a 1. Funkciu  $u_t$  nazývame prechodová premenná. Prechodovou premennou môže byť buď posunutá endogénna premenná, napríklad  $u_t = X_{t-d}$  pre  $d > 0$ , exogénna premenná alebo aj funkcia niekoľkých posunutých endogénnych a exogénnych premenných. Prechodová funkcia obsahuje dva parametre. Prvý z nich,  $r$ , je prah a druhý  $\gamma$ , je parameter hladkosti.

Model môžeme vyjadriť v tvare

$$X_t = \left( \phi_0 + \sum_{i=1}^{p_1} \phi_i X_{t-i} \right) (1 - G(u_t; \gamma, r)) + \left( \varphi_0 + \sum_{i=1}^{p_2} \varphi_i X_{t-i} \right) G(u_t; \gamma, r) + E_t, \quad (4.8)$$

kde  $E_t$  je gaussovský biely šum.

Alebo vo vektorovom tvare

$$\mathbf{X}_t = \boldsymbol{\phi}' \mathbf{X}_t (1 - G(u_t; \gamma, r)) + \boldsymbol{\varphi}' \mathbf{X}_t G(u_t; \gamma, r) + E_t, \quad (4.9)$$

kde  $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_p)'$ ,  $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \dots, \varphi_p)'$  a  $\mathbf{X}_t = (1, x_{t-1}, \dots, x_{t-p})'$ . Model (4.8) vykazuje prítomnosť dvoch režimov združených s extrémnymi hodnotami prechodovej funkcie  $G(u_t; \gamma, r) = 0$  a  $G(u_t; \gamma, r) = 1$ , pričom prechod z jedného režimu

do druhého je postupný. Tento model má aj inú interpretáciu, a to ako model s nekonečne veľa režimami, každý spojený s jednou hodnotou prechodovej funkcie. Rovnako, ako aj v modeloch typu SETAR, aj v modeloch typu STAR je režim v čase  $t$  pozorovateľný, a môže byť určený pomocou prechodovej funkcie.

V praxi sa najčastejšie používajú dva typy prechodovej funkcie, logistická funkcia a exponenciálna funkcia. Logistická prechodová funkcia, ktorá má tvar

$$G(u_t; \gamma, r) = \frac{1}{1 + \exp(-\gamma(u_t - r))} \quad (4.10)$$

nám zabezpečuje rôzne správanie modelu v takých hodnotách prechodovej premennej  $u_t$ , ktoré neprekračujú hodnotu prahu  $r$ , čo znamená, že model sa mení monotónne vzhľadom na prechodovú premennú. Model s logistickou prechodovou funkciou budeme označovať LSTAR. Pre extrémne hodnoty parametra hladkosti model LSTAR prechádza do dvojrežimového modelu SETAR ak  $\gamma \rightarrow 0$ , a do lineárneho AR modelu ak  $\gamma \rightarrow \infty$ .

Exponenciálna prechodová funkcia má tvar

$$G(u_t; \gamma, r) = 1 - \exp(-\gamma(u_t - r)^2) \quad (4.11)$$

Zabezpečuje nám rôzne správanie modelu v malých a veľkých odchýlkach hodnoty prechodovej premennej  $u_t$  od hodnoty prahu  $r$ , čo znamená, že model sa mení „symetricky“ k hodnote prahu. Model s exponenciálnou prechodovou funkciou budeme označovať ESTAR. Pre obe extrémne hodnoty parametra hladkosti  $\gamma \rightarrow 0$  a  $\gamma \rightarrow \infty$  model ESTAR prechádza do lineárneho AR modelu.

### 4.2.1 Špecifikácia modelu STAR

Ako je zrejmé z (4.8) a (4.12), lineárny AR( $p$ ) model je vnorený do modelu STAR. Preto je prirodzené predpokladať, že prvým krokom pri špecifikácii modelu je testovanie linearity modelu proti špecifikácii STAR. Ak nulová hypotéza linearity modelu je akceptovaná, tak daný časový rad môže byť opísaný pomocou lineárneho AR modelu. Ak nulová hypotéza je zamietnutá, má zmysel zaoberať sa špecifikáciou nelineárneho modelu. Postupnosť krokov pri špecifikácii STAR modelu definoval Teräsvirta [12] takto

- Špecifikácia východzieho AR modelu.
- Testovanie linearity pre rôzne hodnoty parametra omeškania  $d$ , a v prípade zamietnutia nulovej hypotézy linearity, špecifikovať  $d$ .
- Voľba typu modelu medzi LSTAR a ESTAR.
- Odhad parametrov modelu.

### 4.2.2 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  $p$  sa odporúča vybrať si niekoľko alternatívnych modelov a vhodný model určiť na základe informačných kritérií.

### 4.2.3 Test linearity a voľba vhodnej funkcie prechodu

Uvažujme autoregresný model hladkého prechodu v tvare (4.8) s prechodovou premennou  $X_{t-d}$ ,  $d > 0$ . Luukkonen, Saikkonen a Teräsvirta [9] navrhli niekoľko testov, ktoré môžu byť použité v celej takej triede STAR modelov s neznámym  $d$ , v ktorej prechodová funkcia  $G(u_t; \gamma, r)$  spĺňa nasledujúce dve podmienky

- Prechodová funkcia  $G(u; \cdot, \cdot)$  je vo svojej prvej premennej nepárna, rýdzo rastúca a pre nejaké  $a > 0$  a  $s \geq 0$  existuje na otvorenom intervale  $(-a, a)$  nenulová derivácia rádu  $2s + 1$ .
- $G(0; \cdot, \cdot) = 0$  a  $[\partial^k G(u; \cdot, \cdot) / \partial u^k]_{u=0} \neq 0$  pre všetky také nepárne  $k$ , kde platí  $1 \leq k \leq 2s + 1$ .

Navrhnuté testy testujú, či sa model (4.8) redukuje na lineárny v tvare

$$X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} + E_t. \quad (4.12)$$

Za platnosti nulovej hypotézy, keď sa model redukuje na lineárny tvar, nedajú sa určiť parametre modelu. Uvedení autori tento problém odstránili nahradením prechodovej funkcie  $G(u; \cdot, \cdot)$  aproximáciou, v tvare Taylorovho polynómu. Linearita je potom testovaná pomocou štatistiky Lagrangeovho multiplikátora so štandardným asymptotickým  $\chi^2$  rozdelením pri platnosti nulovej hypotézy.

Z testov si všimnime test tretieho rádu. Ak sú hore uvedené dve podmienky splnené pre  $s = 1$ , tak je možné funkciu  $G(u; \cdot, \cdot)$  aproximovať pomocou Taylorovho polynómu tretieho rádu

$$T_3(u) = g_1 u + g_3 u^3, \quad (4.13)$$

kde

$$g_1 = [\partial G(u; \cdot, \cdot) / \partial u]_{u=0}, \quad g_3 = \frac{1}{6} [\partial^3 G(u; \cdot, \cdot) / \partial u^3]_{u=0}. \quad (4.14)$$

Dosadením  $u_t = X_{t-d}$  a nahradením funkcie  $G$  funkciou  $T_3$  dostaneme pomocný model tvaru

$$\begin{aligned} X_t = & \beta_0 + \sum_{i=1}^p \beta_i X_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \varphi_{ij} X_{t-i} X_{t-j} + \sum_{i=1}^p \sum_{j=1}^p \zeta_{ij} X_{t-i} X_{t-j}^2 \\ & + \sum_{i=1}^p \sum_{j=1}^p \xi_{ij} X_{t-i} X_{t-j}^3 + E_t. \end{aligned} \quad (4.15)$$

Budeme testovať hypotézu

$H_0^{(3)}$  :  $\varphi_{ij} = 0$  pre  $i = 1, \dots, p, j = i, \dots, p, \zeta_{ij} = 0, \xi_{ij} = 0$  pre všetky  $i, j = 1, \dots, p$ ,  
proti alternatívnej hypotéze

$H_A^{(3)}$  : aspoň jeden z koeficientov  $\varphi_{ij}, \zeta_{ij}, \xi_{ij}$  je nenulový.

### Procedúra (1)

1. Zbehnúť regresiu  $X_t$  na  $\{1, X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$ , vypočítať reziduá  $\hat{E}_t$  a súčet ich štvorcov, t.j.

$$SSR_0 = \sum_{t=n-p+1}^n \hat{E}_t^2. \quad (4.16)$$

2. V prípade, že  $d$  nie je známe, zbehnúť regresiu  $\hat{E}_t$  na  $\{1, X_{t-i}, i = 1, 2, \dots, p, X_{t-i}X_{t-j}, i = 1, 2, \dots, p, j = i, \dots, p, X_{t-i}X_{t-j}^k, i = 1, 2, \dots, p, j = 1, 2, \dots, p, k = 2, 3\}$ .

V prípade, že  $d$  je známe, zbehnúť regresiu  $\hat{E}_t$  na  $\{1, X_{t-i}, i = 1, 2, \dots, p, X_{t-i}X_{t-d}, i = 1, 2, \dots, p, X_{t-i}X_{t-d}^k, i = 1, 2, \dots, p, k = 2, 3\}$ , a vypočítať reziduály  $\hat{\eta}_t$  a súčet ich štvorcov

$$SSR_3 = \sum_{t=n-p+1}^n \hat{\eta}_t^2. \quad (4.17)$$

3. Vypočítať štatistiku

$$S_3 = (n - p) \frac{SSR_0 - SSR_3}{SSR_0}. \quad (4.18)$$

**Veta 6.** (Luukkonen (1988)) Nech  $X_t$  je stacionárny, invertovateľný autoregresný proces rádu  $p$  s reprezentáciou

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + E_t \quad (4.19)$$

kde  $E_t$  sú nezávislé náhodné premenné z rovnakého rozdelenia s nulovou strednou hodnotou, rozptylom  $\sigma_E^2$  a konečným štvrtým momentom. Potom za platnosti hypotézy  $H_0^{(3)}$  má  $S_3$  asymptoticky  $\chi^2$  rozdelenie so stupňami voľnosti  $\frac{1}{2}p(p+1) + 2p^2$  ak nepoznáme  $d$ , a  $3p$  ak poznáme  $d$ .

Test daný procedúrou (1) je súčasne aj testom linearite modelu proti nelinearite typu LSTAR. Saikkonen a Luukkonen [9] navrhli test, kde ako alternatívnu hypotézu k lineárnemu AR modelu postavíme model typu ESTAR. Podľa ich návrhu pomocný model nadobúda tvar

$$X_t = \beta_0 + \sum_{i=1}^p \beta_i X_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \varphi_{ij} X_{t-i} X_{t-j} + \sum_{i=1}^p \sum_{j=1}^p \zeta_{ij} X_{t-i} X_{t-j}^2 + E_t. \quad (4.20)$$

Nulová hypotéza má tvar

$$H_0^{(2)} : \varphi_{ij} = 0 \text{ pre } i = 1, \dots, p, j = i, \dots, p, \zeta_{ij} = 0 \text{ pre všetky } i, j = 1, \dots, p,$$

proti alternatívnej hypotéze

$$H_A^{(2)} : \text{aspoň jeden z koeficientov } \varphi_{ij}, \zeta_{ij} \text{ je nenulový.}$$

### Procedúra (2)

1. Zbehnúť regresiu  $X_t$  na  $\{1, X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$ , vypočítať reziduá  $\hat{E}_t$  a súčet ich štvorcov, t.j.

$$SSR_0 = \sum_{t=n-p+1}^n \hat{E}_t^2. \quad (4.21)$$

2. V prípade, že  $d$  nie je známe, zbehnúť regresiu  $\hat{E}_t$  na  $\{1, X_{t-i}, i = 1, 2, \dots, p, X_{t-i}X_{t-j}, i = 1, 2, \dots, p, j = i, \dots, p, X_{t-i}X_{t-j}^2, i = 1, 2, \dots, p, j = 1, 2, \dots, p\}$ .

V prípade, že  $d$  je známe, zbehnúť regresiu  $\hat{E}_t$  na  $\{1, X_{t-i}, i = 1, 2, \dots, p, X_{t-i}X_{t-d}, i = 1, 2, \dots, p, X_{t-i}X_{t-d}^2, i = 1, 2, \dots, p\}$ , a vypočítať reziduály  $\hat{\eta}_t$  a súčet ich štvorcov

$$SSR_2 = \sum_{t=n-p+1}^n \hat{\eta}_t^2. \quad (4.22)$$

3. Vypočítať štatistiku

$$S_2 = (n-p) \frac{SSR_0 - SSR_2}{SSR_0}. \quad (4.23)$$

**Veta 7.** (Luukkonen (1988)) *Nech  $X_t$  splňa predpoklady z vety 6. Potom za platnosti hypotézy  $H_0^{(2)}$  má  $S_2$  asymptoticky  $\chi^2$  rozdelenie so stupňami voľnosti  $\frac{1}{2}p(p+1) + p^2$  ak nepoznáme  $d$ , a  $2p$  ak poznáme  $d$ .*

V roku 1999 Escibano a Jordá došli na to, že aproximácia Taylorovým polynómom prvého rádu dostatočne nevystihuje charakteristickú vlastnosť exponenciálnej prechodovej funkcie. Preto navrhli aproximáciu prechodovej funkcie Taylorovým polynómom druhého rádu, ktorá vedie k pomocnej rovnici

$$\begin{aligned} X_t = & \beta_0 + \sum_{i=1}^p \beta_i X_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \varphi_{ij} X_{t-i} X_{t-j} + \sum_{i=1}^p \sum_{j=1}^p \zeta_{ij} X_{t-i} X_{t-j}^2 \\ & + \sum_{i=1}^p \sum_{j=1}^p \xi_{ij} X_{t-i} X_{t-j}^3 + \sum_{i=1}^p \sum_{j=1}^p \psi_{ij} X_{t-i} X_{t-j}^4 + E_t. \end{aligned} \quad (4.24)$$

Budeme testovať hypotézu

$$H_0^{(4)} : \varphi_{ij} = 0 \text{ pre } i = 1, \dots, p, j = i, \dots, p, \zeta_{ij} = 0, \xi_{ij} = 0, \psi_{ij} = 0 \text{ pre všetky } i, j = 1, \dots, p,$$

proti alternatívnej hypotéze

$$H_A^{(3)} : \text{aspoň jeden z koeficientov } \varphi_{ij}, \zeta_{ij}, \xi_{ij}, \psi_{ij} \text{ je nenulový.}$$



**Procedúra (3)**

1. Zbehnúť regresiu  $X_t$  na  $\{1, X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$ , vypočítať reziduá  $\hat{E}_t$  a súčet ich štvorcov, t.j.

$$SSR_0 = \sum_{t=n-p+1}^n \hat{E}_t^2. \quad (4.25)$$

2. V prípade, že  $d$  nie je známe, zbehnúť regresiu  $\hat{E}_t$  na  $\{1, X_{t-i}, i = 1, 2, \dots, p, X_{t-i}X_{t-j}, i = 1, 2, \dots, p, j = i, \dots, p, X_{t-i}X_{t-j}^k, i = 1, 2, \dots, p, j = 1, 2, \dots, p, k = 2, 3, 4\}$ .

V prípade, že  $d$  je známe, zbehnúť regresiu  $\hat{E}_t$  na

$\{1, X_{t-i}, i = 1, 2, \dots, p, X_{t-i}X_{t-d}, i = 1, 2, \dots, p, X_{t-i}X_{t-d}^k, i = 1, 2, \dots, p, k = 2, 3, 4\}$ , a vypočítať reziduály  $\hat{\eta}_t$  a súčet ich štvorcov

$$SSR_4 = \sum_{t=n-p+1}^n \hat{\eta}_t^2. \quad (4.26)$$

3. Vypočítať štatistiku

$$S_4 = (n-p) \frac{SSR_0 - SSR_4}{SSR_0}. \quad (4.27)$$

**Veta 8.** (Luukkonen (1988)) *Nech  $X_t$  spĺňa predpoklady z vety 6. Potom za platnosti hypotézy  $H_0^{(4)}$  má  $S_4$  asymptoticky  $\chi^2$  rozdelenie so stupňami voľnosti  $\frac{1}{2}p(p+1) + 3p^2$  ak nepoznáme  $d$ , a  $4p$  ak poznáme  $d$ .*

**4.2.4 Odhad parametrov**

V prípade, ak linearita bola zamietnutá, použijeme na modelovanie daného časového radu STAR model. Uvažujme dvojrežimový model v tvare (4.9). Na určenie parametrov  $\hat{\theta} = (\psi', \varphi', \gamma, r)$  môžeme použiť nelineárnu metódu najmenších štvorcov

$$\hat{\theta} = \arg \min_{\theta} \sum_{t=0}^N (x_t - F(\mathbf{X}_t; \theta))^2, \quad (4.28)$$

kde

$$F(\mathbf{X}_t, \theta) = \phi' \mathbf{X}_t (1 - G(u_t; \gamma, r)) + \varphi' \mathbf{X}_t G(u_t; \gamma, r) + E_t, \quad (4.29)$$

$\phi = (\phi_0, \phi_1, \dots, \phi_p)'$ ,  $\varphi = (\varphi_0, \varphi_1, \dots, \varphi_p)'$  a  $\mathbf{X}_t = (1, x_{t-1}, \dots, x_{t-p})'$ .

**4.3 Markovove modely prepínania režimov****4.3.1 Základný popis a vyžitie**

Tieto modely neposkytujú len zmysluplnú cestu na modelovanie časových radov podmienených zmenami v režime, ale pomocou vhodných filtrujúcich a vyhladzujúcich algoritmov vytvárajú aj optimálne závery o nepozorovateľných stavoch v každom bode v rámci vzorky. Modely s premenlivými režimami sú stále značne využívané na modelovanie agregovaných makroekonomických časových radov, kde ukazujú radikálne

odlišné správanie počas období exenzie a recesie. Tieto modely boli úspešne použité na fitovanie množstva ekonomických a finančných časových radov, napríklad úrokovej miery, výmenných kurzov.

Táto trieda modelov vychádza z predpokladu, že daný režim je určený nejakým nepozorovateľným stochastickým procesom, ktorý môžeme označiť ako  $\{S_t\}$ . Z toho vyplýva, že jednotlivé režimy nie je možné identifikovať presne, ale len s určitou pravdepodobnosťou. V prípade dvoch režimov proces  $\{S_t\}$  nadobúda len hodnoty 1 a 2. Ak vychádzame z modelu AR(1), potom model s režimami určenými nepozorovateľnými veličinami má tvar

$$X_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} X_{t-1} + E_t^{(1)} & \text{ak } S_t = 1 \\ \phi_0^{(2)} + \phi_1^{(2)} X_{t-1} + E_t^{(2)} & \text{ak } S_t = 2, \end{cases} \quad (4.30)$$

alebo zjednodušene

$$X_t = \phi_0^{(S_t)} + \phi_1^{(S_t)} X_{t-1} + E_t^{(S_t)} \quad \text{pre } S_t = 1, 2. \quad (4.31)$$

Aby bol model (4.30) a (4.31) kompletný, musí byť špecifikovaný proces  $\{S_t\}$ . Hamilton [5] prišiel s predstavou, že tento proces je Markovov proces prvého rádu, alebo MSW model (Markov-Switching Model). Predpokladá sa tu, že režim  $S_t$  závisí len na režime  $S_{t-1}$ . Model (4.30) a (4.31) je kompletný, ak definujeme pravdepodobnosti prechodu z jedného stavu do druhého, t.j.

$$P(S_t = 1 \mid S_{t-1} = 1) = p_{11},$$

$$P(S_t = 2 \mid S_{t-1} = 1) = p_{12},$$

$$P(S_t = 1 \mid S_{t-1} = 2) = p_{21},$$

$$P(S_t = 2 \mid S_{t-1} = 2) = p_{22}.$$

Pravdepodobnosť prechodu zo stavu  $i$  v čase  $t-1$  do stavu  $j$  v čase  $t$  v Markovovom reťazci, resp. pravdepodobnosť prechodu zo stavu  $i$  v čase  $t-1$  do režimu  $j$  v čase  $t$  v modeli časového radu je rovná  $p_{ij}$ . Platia rovnice  $p_{11} + p_{12} = 1$  a  $p_{21} + p_{22} = 1$ . Nepodmienené pravdepodobnosti, že model je v čase  $t$  v danom režime, odvodil Hamilton ako vlastný vektor zodpovedajúci vlastnému číslu  $\lambda = 1$  matice

$$\begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}$$

a týmito pravdepodobnosťami sú

$$P(S_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \quad (4.32)$$

$$P(S_t = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \quad (4.33)$$

### 4.3.2 Tvorba MSW modelov

Pri modelovaní vždy je vhodné začínať jednoduchšími lineárnymi modelmi a potom, keď nie sú splnené podmienky pre ich použitie, prejsť k modelom všeobecnejším a tiež komplexnejším, akými sú modely nelineárne. Odporúčajú sa nasledovné kroky:

- Pre daný časový rad určiť vhodný lineárny model typu AR( $p$ ).
- Otestovať nulovú hypotézu linearity modelu oproti alternatívnej hypotéze nelinearity modelu.
- Odhadnúť parametre zvoleného nelineárneho modelu.
- Overiť vhodnosť modelu.
- Ak je potrebné, model treba vylepšiť.

### 4.3.3 Testovanie linearity oproti nelinearite typu MSW

Pre testovanie linearity oproti alternatíve nelinearity typu MSW sa používa test vierohodnostným pomerom, v ktorom sa testuje nulová hypotéza linearity, t.j.

$$H_0 : \Phi_1 = \Phi_2, \quad (4.34)$$

$$H_1 : \phi_{i,1} \neq \phi_{i,2} \text{ pre } i \in \{0, 1, \dots, p\}, \quad p = \max(p_1, p_2). \quad (4.35)$$

Testovacie kritérium je  $LR = L_{MSW} - L_{AR}$ , kde  $L_{MSW}$  a  $L_{AR}$  sú hodnoty logaritmov vierohodnostných funkcií zodpovedajúcich modelom MSW a AR. Za predpokladu nulovej hypotézy pravdepodobnosti prechodu  $p_{11}$  a  $p_{22}$  sú prebytočnými parametrami. Hansen [6] ukázal, že táto testovacia štatistika linearity nemá štandardné pravdepodobnostné rozdelenie. Toto rozdelenie dokonca ani nemôže byť vyjadrané analyticky, ale jedine simulačne.

### 4.3.4 Odhady parametrov modelu MSW

Pri MSW modeloch potrebujeme odhadnúť

- počet stavov  $N$
- rozdelenie pravdepodobnosti prechodu z jedného stavu do druhého  
 $p_{ij} = P(q_t = j \mid q_{t-1} = i) \quad 1 \leq i, j \leq N$
- počiatkové stavové pravdepodobnosti  
 $\pi = \pi_i$   
 $\pi_i = P(q_1 = i) \quad i = 1, 2, \dots, N$
- autoregresné parametre a rozptyly reziduálnej zložky v jednotlivých režimoch, t.j.  $\Phi_i = (\phi_0^{(i)}, \phi_1^{(i)}, \dots, \phi_{p_i}^{(i)}, \sigma_i^2) \quad i = 1, 2, \dots, N$
- odhad postupnosti stavov pre daný časový rad, t.j.  $q_1, q_2, \dots, q_n$ . Tu budeme odhadovať pravdepodobnosť, s ktorou nastane stav  $j$  v čase  $t$ . Nakoniec sa v každom čase  $t$  vyberie stav s najväčšou pravdepodobnosťou.

Parametre MSW modelu môžeme odhadnúť pomocou metódy maximálnej vierohodnosti. Vzhľadom na to, že Markovovský proces  $q_t$  je nepozorovateľný, problém odhadu je neštandardný. Cieľom odhadovacej procedúry je získať odhady autoregresných parametrov, rozptylov reziduálnej zložky v rôznych režimoch, odhady pravdepodobností

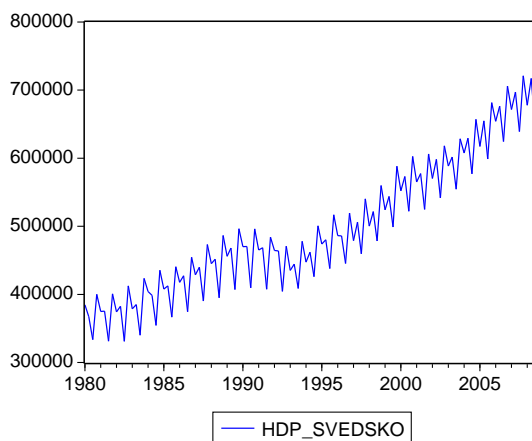
prechodu z jedného režimu do druhého a odhady pravdepodobností, s ktorými nastane istý stav v jednotlivých časoch. Podrobnosti sa dajú nájsť v diplomovej práci Petričkovej [11].

# Kapitola 5

## Aplikácia teórie v praxi

V tejto kapitole sa budeme snažiť nájsť vhodné modely pre štvrťročné dáta HDP Švédska. Hlavnou úlohou a takisto ťažiskom práce je aplikovať teóriu nelineárneho modelovania. Konkrétne budeme hľadať SETAR a STAR modely pre náš časový rad. V týchto modeloch jedným z parametrov je  $p$ , tzv. maximálny rád autoregresie. Preto pred hľadaním nelineárnych modelov, najprv musíme nájsť vhodné lineárne  $AR(p)$  modely. Z tohto dôvodu začneme kapitolu s analýzou dát a ich lineárnym modelovaním. Pri modelovaní sme mali na pomoci softvéry EViews 5, Matlab 6.5 a Mathematicu 5.2.

Dáta máme z rokov 1980-2008, graficky znázornené ich vidíme na obrázku 5.1.

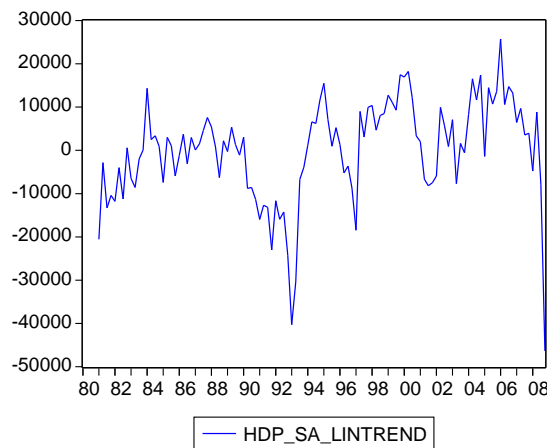


Obr. 5.1: Kvartálne dáta HDP Švédska dané v miliónoch domácej meny

Ako aj z obrázka vidieť, v dátach máme sezónnosť, a keďže ide o štvrťročné dáta, tak sme zobrali sezónnu zložku  $s$  dĺžky 4. Sezónnosť sme odstránili „ručne“, pomocou sezónnych diferencií.

V ďalšom kroku ku vytvoreniu správneho modelu sme odstránili trend z daného časového radu. Jednoznačne sa nedá určiť ani jeden konkrétny typ trendu, ktorý by perfektne kopíroval naše dáta, mi sme napriek tomu skúšali lineárny, kvadratický a hp trend. Najviac sa nám osvedčil model, kde sme odstránili z pôvodného časového radu lineárny trend. Detrendovaný a sezónne očistený časový rad budeme označovať  $HDP_{LTS}$ .

Po úpravách časového radu HDP Švédska sme dostali časový rad, ktorý je graficky znázornený na obrázku 5.2.



Obr. 5.2: Časový rad HDP po odstránení lineárneho trendu a po sezónnom očistení

Ďalej sme otestovali prítomnosť jednotkového koreňa v skúmanom časovom rade, t.j., či daný časový rad je stacionárny. Na testovanie stacionarity sme použili Dickey-Fullerov test. Nulovou hypotézou pri tomto teste je prítomnosť jednotkového koreňa, ale túto hypotézu sme zamietli s p-hodnotou 0.0003. Teda vychádzame zo stacionárneho časového radu a môžeme začať s jeho lineárnym modelovaním.

## 5.1 Modely SETAR

### 5.1.1 Lineárne a nelineárne modelovanie časového radu $HDP_{LTS}$

V prípade časového radu  $HDP_{LTS}$  najlepšie  $AR(p)$  modely, ktoré sme našli sú  $AR(5)$ ,  $AR(8)$ ,  $AR(11)$ . Pri ďalšej analýze dát, presnejšie, pri hľadaní vhodného nelineárneho modelu, v prípade  $AR(8)$  a  $AR(11)$  sa preukázalo, že tieto časové rady nie sú vhodné na nelineárne modelovanie, lebo prahová nelinearita sa v ich prípade zamietla. Otestovali sme nekorelovanosť reziduí a druhých mocnín reziduí, a korelovanosť dát sme zamietli. Normalitu reziduí sme otestovali empirickými testami normality (Lilliefors, Cramer-von Mises, Watson), a nulovú hypotézu normality nezamietli. Uvádžame tabuľky odhadov koeficientov pre nájdený modely, a takisto aj korelogramy reziduí a druhých mocnín reziduí na obrázkach 5.3, 5.4, 5.5.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.777532	0.086896	8.947856	0.0000
AR(4)	-0.218040	0.118872	-1.834246	0.0695
AR(5)	0.205940	0.108275	1.901998	0.0599
R-squared	0.492348	Mean dependent var	561.9121	
Adjusted R-squared	0.482586	S.D. dependent var	11596.82	
S.E. of regression	8341.769	Akaike info criterion	20.92358	
Sum squared resid	7.24E+09	Schwarz criterion	20.99851	
Log likelihood	-1116.411	Durbin-Watson stat	1.886497	

Obr. 5.3: Odhady koeficientov pre model AR(5)

Correlogram of Residuals						
Date: 04/25/10 Time: 16:40						
Sample: 1982Q2 2008Q4						
Included observations: 107						
Q-statistic probabilities adjusted for 3 ARMA term(s)						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1 -0.147	-0.147	2.3894		
		2 0.109	0.090	3.7195		
		3 0.063	0.093	4.1598		
		4 -0.069	-0.060	4.7016	0.030	
		5 0.148	0.118	7.2043	0.027	
		6 -0.017	0.028	7.2390	0.065	
		7 0.001	-0.019	7.2391	0.124	
		8 -0.086	-0.117	8.1041	0.151	
		9 0.026	0.017	8.1825	0.225	
		10 0.064	0.082	8.6795	0.276	
		11 -0.096	-0.079	9.7928	0.280	
		12 -0.043	-0.100	10.018	0.349	
		13 -0.008	0.018	10.026	0.438	
		14 -0.078	-0.049	10.794	0.461	
		15 0.091	0.051	11.842	0.458	
		16 0.011	0.054	11.857	0.539	
		17 -0.020	0.013	11.910	0.614	
		18 0.023	0.007	11.981	0.680	
		19 0.041	0.047	12.200	0.730	
		20 0.002	-0.014	12.201	0.788	
		21 0.073	0.066	12.917	0.796	
		22 -0.076	-0.076	13.702	0.801	
		23 0.032	0.008	13.847	0.838	
		24 0.014	0.021	13.874	0.875	
		25 -0.061	-0.078	14.396	0.887	

Obr. 5.4: Korelogram reziduí pre model AR(5) časového radu  $HDP_{LTS}$ 

Correlogram of Residuals Squared						
Date: 04/25/10 Time: 16:36						
Sample: 1982Q2 2008Q4						
Included observations: 107						
Q-statistic probabilities adjusted for 3 ARMA term(s)						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1 0.046	0.046	0.2352		
		2 0.013	0.011	0.2541		
		3 0.014	0.013	0.2772		
		4 0.012	0.010	0.2932	0.588	
		5 0.012	0.010	0.3091	0.857	
		6 -0.005	-0.006	0.3117	0.958	
		7 0.011	0.011	0.3248	0.988	
		8 0.004	0.003	0.3266	0.997	
		9 -0.011	-0.011	0.3403	0.999	
		10 0.016	0.017	0.3708	1.000	
		11 0.013	0.012	0.3927	1.000	
		12 -0.004	-0.005	0.3944	1.000	
		13 -0.014	-0.014	0.4187	1.000	
		14 -0.005	-0.004	0.4217	1.000	
		15 0.020	0.021	0.4747	1.000	
		16 0.008	0.007	0.4838	1.000	
		17 0.058	0.057	0.9143	1.000	
		18 0.019	0.013	0.9621	1.000	
		19 -0.019	-0.022	1.0093	1.000	
		20 0.007	0.006	1.0155	1.000	
		21 -0.009	-0.011	1.0272	1.000	
		22 0.002	0.001	1.0277	1.000	
		23 -0.026	-0.025	1.1189	1.000	
		24 -0.007	-0.004	1.1249	1.000	
		25 -0.001	-0.001	1.1250	1.000	

Obr. 5.5: Korelogram druhých mocnín reziduí pre model AR(5) časového radu  $HDP_{LTS}$ 

Odhadnutý model aj s koeficientami je nasledujúci

$$x_t = \phi_1 x_{t-1} + \phi_4 x_{t-4} + \phi_5 x_{t-5} + E_t,$$

$$x_t = 0.7775x_{t-1} - 0.2180x_{t-4} + 0.2059x_{t-5} + E_t.$$

Po nájdení vhodného „východiskového“ lineárneho modelu, snažili sme sa nájsť vhodný nelineárny model pre dané dáta. Pomocou testovacej štatistiky 4.5 sme zistili akú hodnotu máme zobrať za parameter omeškania. Preukázalo sa, že najlepšie je zobrať hodnotu  $d = 3$ , ale hypotéza prahovej nelinearity sa nezamietla ani v prípade  $d = 1$ .

### 5.1.2 Identifikovanie nelineárneho modelu časového radu HDP<sub>LTS</sub> pre $d = 3$

Parameter omeškania najprv sme zobrali  $d = 3$ , a neskôr budeme analyzovať aj prípad  $d = 1$ . Ako poslednú vec, potrebnú na určenie modelov jednotlivých režimov, sú hodnoty prahov, ktoré sme odhadli zo scatterplotov, a dostali sme hodnoty  $r_1 = -4500$  a  $r_2 = 4000$ . Pomocou týchto informácií o parametroch SETAR modelu sme sa zaoberali modelovaním dát dĺžky 32, 45 a 30. Pred modelovaním sme otestovali Chowov testom zlomu, či naozaj je preukázateľný zlom v preusporiadaných dátach, a prijali sme alternatívnu hypotézu, t.j., že zlom v dátach je preukázateľný. Zlom s vysokými p-hodnotami sa preukázal v prípade, keď sme testovali prahy zvlášť, a takisto aj v prípade, keď sme zobrali oba prahy a testom sme testovali dva zlomy naraz v dátach. To znamená, že v ďalšej časti kapitoly sa budeme zaoberať týmito tromi možnosťami, a všetky tri prípady vyšetříme.

#### 5.1.3 Model SETAR(5; 3, -4500)

Prah  $r=-4500$  nám rozdelil dáta na dve sady dát dĺžky 32 a 75. Na zistenie najlepšej kombinácie autoregresíí modelov určujúcich správanie sa modelu v jednotlivých režimoch sme použili AIC a BIC informačné kritériá. Tieto kritériá pri možnej najmenšej hodnote určujú akú kombináciu rádu autoregresie máme brať v jednotlivých režimoch. Tu sa najviac osvedčil model s  $p_1 = 1$ ,  $p_2 = 3$  rádmí autoregresie v jednotlivých režimoch, s hodnotami informačných kritérií AIC= 1934.13 a BIC= 1946.33. Suma štvorcov reziduí pre odhadnutý model je SSR=  $7.3658 * 10^9$ . Výsledný model aj s odhadnutými koeficientami je nasledujúci

$$x_t = \begin{cases} -1326.1511 + 0.8099x_{t-1} + E_t & \text{ak } x_{t-3} < -4500, \\ 771.7286 + 0.3381x_{t-1} + 0.4001x_{t-2} - 0.0320x_{t-3} + E_t & \text{ak } x_{t-3} \geq -4500. \end{cases} \quad (5.1)$$

Namodelované a pôvodné dáta sú porovnané na obrázku A.6.

#### 5.1.4 Model SETAR(5; 3, 4000)

Pomocou hodnoty prahu  $r=4000$  sa rad vhodný na určenie modelov v jednotlivých režimoch rozdelil na jednotlivé úseky dĺžky 67 a 30. Pomocou AIC a BIC informačných kritérií sme určili, aké rády autoregresie sú najvhodnejšie na modelovanie „rozsekaného“ radu. Dostali sme  $p_1 = 2$  a  $p_2 = 3$ . Informačné kritériá modelu vyšli AIC= 1805.66 a BIC= 1817.88. Suma štvorcov reziduí pre odhadnutý model má hodnotu SSR=  $1.0342 * 10^{10}$ . Odhad modelu je nasledovný

$$x_t = \begin{cases} -685.0370 + 0.7214x_{t-1} + 0.0059x_{t-2} + E_t & \text{ak } x_{t-3} < 4000, \\ 334.2376 + 0.3731x_{t-1} + 0.5442x_{t-2} - 0.1293x_{t-3} + E_t & \text{ak } x_{t-3} \geq 4000. \end{cases} \quad (5.2)$$

Modelované a pôvodné dáta sú znázonené na obrázku A.7.



### 5.1.5 Model SETAR(5, 3; (-4500, 4000))

Pomocou hodnôt prahov  $r_1=-4500$  a  $r_2=4000$  sme rozdelili na nelineárne modelovanie vhodný rad na jednotlivé úseky dĺžky 32, 45 a 30. Najlepšiu kombináciu rádu autoregresných členov určených na modelovanie jednotlivých režimov sme našli ako  $p_1 = 1$ ,  $p_2 = 2$  a  $p_3 = 3$ . Informačné kritériá pre tieto parametre modelu nám vyšli AIC=1933.60 a BIC=1947.56. Suma štvorcov reziduí odhadnutého modelu je SSR=  $7.2911 * 10^9$ . Výsledný odhadnutý model má tvar

$$x_t = \begin{cases} -1326.1511 + 0.8099x_{t-1} + E_t & \text{ak } x_{t-3} < -4500, \\ 635.6774 + 0.2610x_{t-1} + 0.3787x_{t-2} + E_t & \text{ak } -4500 \leq x_{t-3} < 4000, \\ 630.3101 + 0.4547x_{t-1} + 0.3833x_{t-2} & \\ -0.0690x_{t-3} + E_t & \text{ak } x_{t-3} \geq 4000. \end{cases} \quad (5.3)$$

Modelované a pôvodné dáta porovnávame na obrázku A.8.

### 5.1.6 Identifikovanie nelineárneho modelu časového radu HDP<sub>LTS</sub> pre $d = 1$

V druhom prípade sme za parameter omeškania zobrali  $d = 1$ . Pri poznaní maximálneho rádu autoregresie modelu a hodnoty parametera omeškania nám zostalo za úlohu určiť len hodnoty prahov, ktoré sme odhadli zo scatterplotov, a dostali sme hodnoty  $r_1 = -7000$  a  $r_2 = 3000$ . Pomocou týchto informácií o parametroch SETAR modelu sme sa zaoberali modelovaním dát dĺžky 23, 25 a 49. Pred modelovaním sme otestovali Chowov testom zlomu, či naozaj je preukázateľný zlom v preusporiadaných dátach, a prijali sme alternatívnu hypotézu tohoto testu, t.j., že zlom v dátach je preukázateľný. Zlom sa preukázal v prípade, keď sme testovali prahy zvlášť, a takisto aj v prípade, keď sme zobrali oba prahy a testom sme testovali dva zlomy naraz v dátach. V ďalšej časti kapitoly vyšetríme všetky tri prípady.

### 5.1.7 Model SETAR(5; 1, -7000)

Prah  $r=-4500$  nám rozdelil na nelineárne modelovanie vhodné dáta na dve sady dát dĺžky 23 a 84. Z hľadiska hodnôt informačných kritérií sa najviac osvedčil model s  $p_1 = 3$ ,  $p_2 = 1$  rádmi autoregresie v jednotlivých režimoch, s hodnotami informačných kritérií AIC= 1922.30 a BIC= 1931.71. Suma štvorcov reziduí pre odhadnutý model vyšiel SSR=  $7.0059 * 10^9$ . Výsledný model aj s odhadnutými koeficientami je nasledujúci

$$x_t = \begin{cases} -4451.4726 + 0.5488x_{t-1} - 0.6012x_{t-2} + 0.7799x_{t-3} + E_t & \text{ak } x_{t-1} < -7000, \\ 1270.0407 + 0.5957x_{t-1} + E_t & \text{ak } x_{t-1} \geq -7000. \end{cases} \quad (5.4)$$

Namodelované a pôvodné dáta sú graficky znázornené na obrázku A.9.

### 5.1.8 Model SETAR(5; 1, 3000)

Pomocou hodnoty prahu  $r=3000$  sa rad vhodný na určenie modelov v jednotlivých režimoch rozdelil na jednotlivé úseky dĺžky 58 a 49. Pomocou AIC a BIC informačných kritérií sme určili, aké rády autoregresie sú najvhodnejšie na modelovanie „rozsekaného“ radu. Dostali sme  $p_1 = 1$  a  $p_2 = 3$ . Suma štvorcov reziduí pre odhadnutý model je  $SSR = 8.7944 * 10^9$ . Nami špecifikované parametre určujú model v tvare

$$x_t = \begin{cases} 134.9210 + 0.7696x_{t-1} + E_t & \text{ak } x_{t-1} < 3000, \\ -89.2339 + 0.5577x_{t-1} + 0.3317x_{t-2} - 0.0918x_{t-3} + E_t & \text{ak } x_{t-1} \geq 3000. \end{cases} \quad (5.5)$$

Modelované a pôvodné dáta porovnávame na obrázku A.10.

### 5.1.9 Model SETAR(5, 1; (-7000, 3000))

Pomocou hodnôt prahov  $r_1=-7000$  a  $r_2=3000$  sme rozdelili na nelineárne modelovanie vhodný rad na jednotlivé úseky dĺžky 23, 25 a 49. Najlepšiu kombináciu rádu autoregresných členov určených pre modelovanie jednotlivých režimov sme našli ako  $p_1 = 5$ ,  $p_2 = 1$  a  $p_3 = 3$ . Informačné kritériá pre tieto parametre nám vyšli  $AIC=1958.67$  a  $BIC=1976.16$ . Suma štvorcov reziduí pre odhadnutý model vyšiel  $SSR = 8.3665 * 10^9$ .

$$x_t = \begin{cases} -5787.4024 + 0.5584x_{t-1} - 0.6306x_{t-2} \\ + 0.5366x_{t-3} - 0.0398x_{t-4} + 0.2618x_{t-5} + E_t & \text{ak } x_{t-1} < -7000, \\ 861.5781 + 0.2691x_{t-1} + E_t & \text{ak } -7000 \leq x_{t-1} < 3000, \\ -89.2339 + 0.5577x_{t-1} + 0.3317x_{t-2} \\ - 0.0918x_{t-3} + E_t & \text{ak } x_{t-1} \geq 3000. \end{cases} \quad (5.6)$$

Modelované a pôvodné dáta porovnávame na obrázku A.11.

## 5.2 Modely STAR

V nasledujúcej časti práce sa budeme zaoberať nájdením najvhodnejších nelineárnych STAR modelov pre naše dáta. Na to budeme používať zdrojové kódy Ing. Tomáša Bacigála, PhD. [1]. Samotné STAR modely sme hľadali pre sezónne očistený a detrendovaný časový rad HDP Švédska, ako v prípade SETAR modelov. Zaoberali sme sa teda modelovaním časového radu  $HDP_{LTS}$ . Pri identifikovaní najvhodnejších STAR modelov sme brali vstupné údaje  $p=5$ ,  $d=3$  a  $p=5$ ,  $d=1$ . Pri Tsay teste prahovej nelinearity sme prijali alternatívnu hypotézu testu, t.j., že dáta je vhodnejšie nelineárne modelovať. Ale keď sme použili LM test na otestovanie STAR modelu ako alternatívnu hypotézu oproti lineárneho modelu, tak test nám zamietol hypotézu o vhodnejšom použití STAR modelu pre naše dáta oproti lineárnemu modelovaniu. Napriek tomu odhadneme model pre vyššie určené parametre. Program nám umožnil odhadovať 2-režimové modely s logistickou prechodovou funkciou, teda budeme odhadovať LSTAR modely. Výsledné modely budeme porovnávať podľa AIC, BIC kritérií.

### 5.2.1 Hľadanie najvhodnejšieho STAR modelu

Hlavná funkcia na odhad STAR modelov v použitom programe ako vstup si žiadala parametre  $p$ ,  $d$ ,  $r$ ,  $\gamma$ . Vhodnosť zvolených parametrov  $p$  a  $d$  sme overili pomocou testu prahovej nelinearity. V prípade parametra  $r$  program nám ponúkol funkciu `fThresholdRange`, kde sme si mohli zadať, koľko najvýznamnejších prahov si žiadame, a funkcia nám vrátila interval možných hodnôt. Konkrétne pre naše dáta, v prípade dvadsiatich prahov sme dostali výstup s hodnotou minimálneho prahu  $r_{min} = -11236.1$ , maximálneho prahu  $r_{max} = 10689.9$  a krokom 1096.3. Pri testovaní sme si zvolili množinu hodnôt pre parameter hladkosti ako  $\gamma = \{0.25, 0.5, 2, 5\}$ . Funkciu prechodu sme si zvolili logistickú. Hľadali sme najlepšie modely z hľadiska AIC, BIC informačných kritérií. Po zvolení parametrov stačilo dosadiť do funkcie `fConditionalRegimeSwitching`, a tá nám odhadla parametre modelu. V prípade, ak sme odhadli model s rôznymi autoregresnými stupňami v jednotlivých režimoch, dostali sme menšie hodnoty informačných kritérií.

### 5.2.2 Najlepšie odhadnuté LSTAR modely

Modely, ktoré nám najlepšie fitovali pôvodné dáta sa našli pre parametre  $p = 5$ ,  $d = 3$ . Ide o tri modely. Prvý najlepší odhadnutý model má hodnoty parametrov  $d = 3$ ,  $r = -3562$  a  $\gamma = 0.8$ . Výsledný model s odhadnutými koeficientami má tvar

$$\begin{aligned} x_t = & (-1229.88 + 0.7408x_{t-1})(1 - G(.)) & (5.7) \\ & + (145.039 + 0.4730x_{t-1} + 0.4905x_{t-2} + 0.0076x_{t-3} \\ & - 0.3816x_{t-4} + 0.2018x_{t-5})G(.) + E_t, \end{aligned}$$

kde

$$G(.) = \frac{1}{1 + \exp(-0.8(x_{t-3} + 3562))}.$$

SSR modelu vyšiel  $1.194902 * 10^{10}$ . Pôvodné a namodelované dáta sú znázornené na obrázku A.12. Tento model je jediný, kde sme odhad hľadali pre množinu možných prahov. V ďalších dvoch prípadoch sme si už zvolili len jednu hodnotu pre prah, konkrétne sme tam otestovali pri SETAR modeloch nájdené prahy  $r_1 = -4500$  a  $r_2 = 4000$ . Najlepšie sa ukázali modely pre parametre  $d_1 = 3$ ,  $r_1 = -4500$ ,  $\gamma_1 = 0.25$  a  $d_2 = 3$ ,  $r_2 = 4000$ ,  $\gamma_2 = 0.8$ . Odhadnutý druhý model má tvar

$$\begin{aligned} x_t = & (-1326.15 + 0.8099x_{t-1})(1 - G(.)) & (5.8) \\ & + (946.3 + 0.3756x_{t-1} + 0.5061x_{t-2} + 0.0073x_{t-3} - 0.3783x_{t-4} \\ & + 0.1656x_{t-5})G(.) + E_t, \end{aligned}$$

kde

$$G(.) = \frac{1}{1 + \exp(-0.25(x_{t-3} + 4500))}.$$

Pre SSR modelu sme dostali hodnotu  $1.1545 * 10^{10}$ . Namodelované a pôvodné dáta sú porovnané na obrázku A.13. Nami špecifikovaný tretí model je nasledujúci

$$\begin{aligned} x_t &= (-695.093 + 0.7252x_{t-1})(1 - G(\cdot)) \\ &+ (848.881 + 0.4946x_{t-1} + 0.4287x_{t-2} - 0.0582x_{t-3} - 0.4603x_{t-4} \\ &+ 0.3539x_{t-5})G(\cdot) + E_t, \end{aligned} \quad (5.9)$$

kde

$$G(\cdot) = \frac{1}{1 + \exp(-0.8(x_{t-3} - 4000))}.$$

Suma štvorcov reziduí pre model má hodnotu  $9.4868 * 10^9$ . Namodelované a pôvodné dáta pre tretí najlepší model sú znázornené na obrázku A.14.

# Záver

V práci sme sa venovali triede viacrežimových modelov STAR, SETAR a MSW, ale v praktickej časti sme porovnávali len odhady získaných zo SETAR a LSTAR modelov.

Modely z teoretického hľadiska sme popísali vo štvrtjej kapitole, a v piatej kapitole sme sa zaoberali už len aplikáciou nelineárneho modelovania na reálnych dátach HDP Švédska. Piatu kapitolu by sme mohli rozdeliť na dve hlavné časti, v jednej z nich sme hľadali odhady parametrov SETAR modelu, a v druhej časti sme sa zaoberali odhadovaním LSTAR modelov. Pri hľadaní vhodného SETAR aj LSTAR modelu sme použili časový rad  $HDP_{LTS}$ , ktorý sme získali z časového radu HDP Švédska odstránením trendovej a cyklickej zložky.

V triede SETAR modelov sa nám podarilo odhadnúť šesť modelov a v prípade LSTAR modelov ide len o tri odhadnuté modely. Pri odhadovaní modelov, zvolíť najlepší rád autoregresie modelov, ktoré určujú správanie sa dát v jednotlivých režimoch, nám pomáhali informačné kritériá AIC a BIC, pričom sme uprednostnili ich nižšiu hodnotu. Takisto sme brali do úvahy aj sumy štvorcov reziduí pre jednotlivé modely. V prílohe na obrázkach sú porovnané pôvodné a namodelované údaje. Pre LSTAR modely sme nedostali až také dobré zhody, čo potvrdzujú aj väčšie hodnoty sumy štvorcov reziduí v prípade týchto modelov. Ale dá sa to zdôvodniť tým, že LM test nelinearity testujúci lineárny vs. nelineárny model typu STAR pre konkrétne  $p$  a  $d$  parametre zamietol hypotézu o vhodnom modelovaní časového radu  $HDP_{LTS}$  nelineárnym STAR modelom. Medzi hodnotami informačných kritérií pre nájdené modely nie sú rádovo veľké rozdiely, preto pri výbere najvhodnejšieho nelineárneho modelu pre náš odhadovaný časový rad porovnáваме sumy štvorcov reziduí pre dané modely. Z tohoto dôvodu ako najlepší model sme si zvolili SETAR(5; 1, -7000) so  $SSR = 7.0059 * 10^9$ , ktorý má tvar

$$x_t = \begin{cases} -4451.4726 + 0.5488x_{t-1} - 0.6012x_{t-2} + 0.7799x_{t-3} + E_t & \text{ak } x_{t-1} < -7000, \\ 1270.0407 + 0.5957x_{t-1} + E_t & \text{ak } x_{t-1} \geq -7000. \end{cases}$$

Modelované a pôvodné dáta máme porovnané na obrázku A.9 v prílohe. Fitované hodnoty v prípade tohoto modelu kopírujú najkrajšie modelované dáta zo všetkých uvedených modelov.

Ďalšou možnosťou rozšírenia tejto práce pre hľadanie lepších modelov môžu byť porovnané predikcie na základe odhadnutých modelov, čím by sa dali nájsť ešte lepšie modely pre skúmaný časový rad.

# Literatúra

- [1] BACIGAL, T. 2007. *Moderné metódy modelovania časových radov s aplikáciou v geodézii* : dizertačná práca PhD štúdia. Bratislava: Slovenská Technická Univerzita, 2007.  
Program v programovacom prostredí Mathematica:  
[www.math.sk/bacigal/homepage/research.html](http://www.math.sk/bacigal/homepage/research.html)
- [2] BOX, G. E. P., JENKINS, G. M. 1976. *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day, 1976.
- [3] CIPRA, T. 1986. *Analýza časových rad s aplikacemi v ekonomii*. Praha: SNTL/ALFA, 1986.
- [4] GRANGER, C. W. J., TERÄSVIRTA, T. 1993. *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press, 1993.
- [5] HAMILTON, J. D. 1994. *Time Series Analysis*. Princeton University Press, 1994.
- [6] HANSEN, B. E. 1994. The likelihood ratio test under nonstandard conditions: Testing the Markow-Switching model of GNP. *Journal of Applied Econometrics*. 1994, 7, 61-82.
- [7] CHAN, W.-S., C.S.WONG, A., TONG, H. 2004. Some Nonlinear Threshold Autoregressive Time Series Model for Actuarial Use. *North American Actuarial Journal*. 2004, 8, 37-61.
- [8] HANDLOVIČ, M. 2007. *Informačné kritériá pre STAR modely* : diplomová práca. Bratislava: Fakulta matematiky, fyziky a informatiky UK, 2007.
- [9] LUUKKONEN, R., SAIKKONEN, P., TERÄSVIRTA, T. 1988. Testing Linearity Against Smooth Transition autoregressive Models. *Biometrika*. 1988, 75, 491-499.
- [10] PEKÁR, J. 2004. *Autorgresné modely hrubého domáceho produktu Slovenska*. Bratislava: Fakulta matematiky, fyziky a informatiky UK, 2004.
- [11] PETRIČKOVÁ, A. 2009. *Modely časových radov s premenlivými režimami* : diplomová práca. Bratislava: Fakulta matematiky, fyziky a informatiky UK, 2009.
- [12] TERÄSVIRTA, T. 1994. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*. 1994, 89, 208-218.

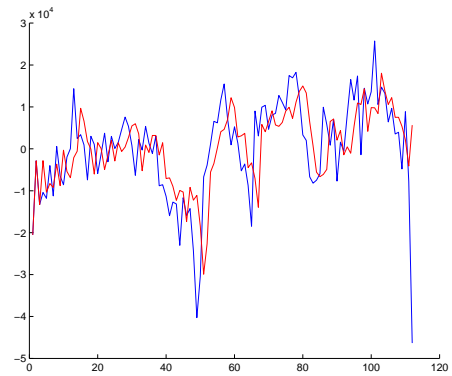
- 
- [13] TONG, H., LIM, K. S. 1980. Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society*. 1980, SER. B, 42, 245-292.
- [14] TSAY, R. S. 1989. Testing and Modeling Threshold Autoregressive Processes. *Journal of the American Statistical Association*. 1989, 84, 231-240.
- [15] [www.epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/](http://www.epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/)
- [16] [www.wikipedia.org](http://www.wikipedia.org)
- [17] [wikipedia.infostar.cz](http://wikipedia.infostar.cz)
- [18] [www.cikomodity.com/trhovy-komentar](http://www.cikomodity.com/trhovy-komentar)
- [19] [library.muni.cz/EU/html/181tsve.htm](http://library.muni.cz/EU/html/181tsve.htm)

# Appendix

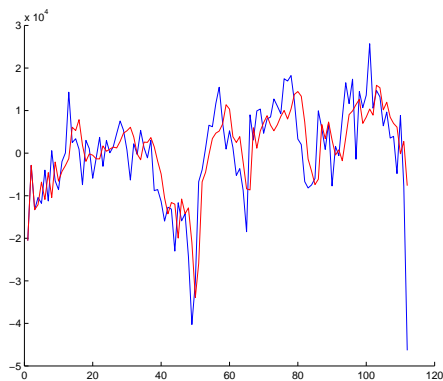




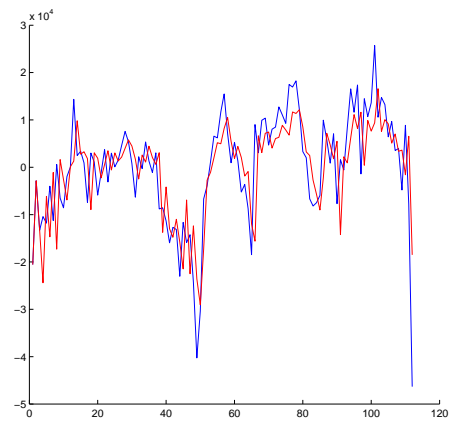
Obr. A.6: Porovnanie pôvodých a modelovaných dát odhadnutých modelom SETAR(5; 3, -4500)



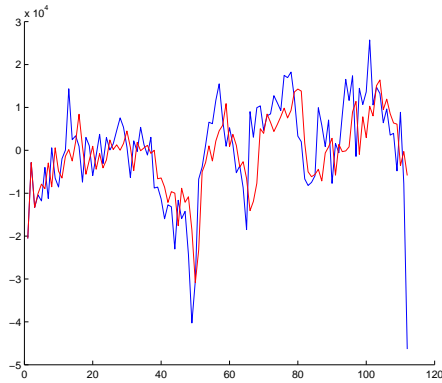
Obr. A.7: Porovnanie pôvodých a modelovaných dát odhadnutých modelom SETAR(5; 3; 4000)



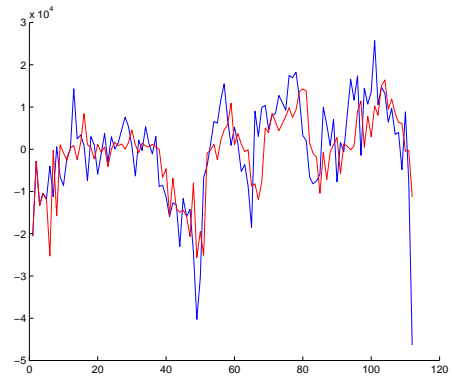
Obr. A.8: Porovnanie pôvodých a modelovaných dát odhadnutých modelom SETAR(5; 3, (-4500, 4000))



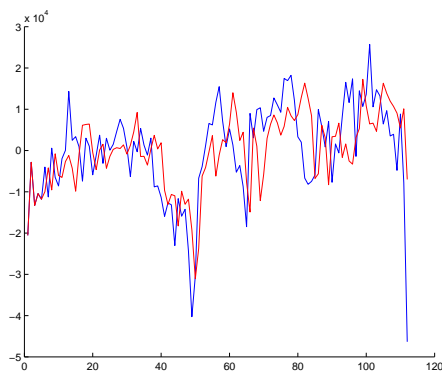
Obr. A.9: Porovnanie pôvodých a modelovaných dát odhadnutých modelom SETAR(5; 1, -7000)



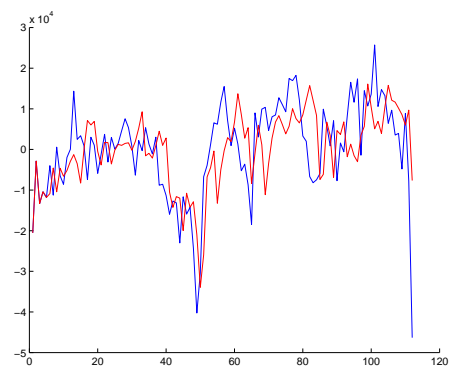
Obr. A.10: Porovnanie pôvodých a modelovaných dát odhadnutých modelom SETAR(5; 1, 3000)



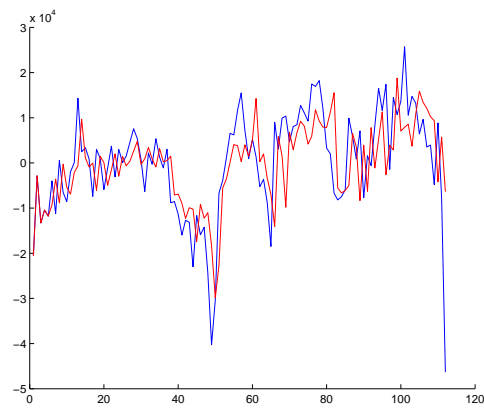
Obr. A.11: Porovnanie pôvodých a modelovaných dát odhadnutých modelom SETAR(5, 1; (-7000, 3000))



Obr. A.12: Porovnanie pôvodých a modelovaných dát odhadnutých z modelu LSTAR s  $G(x_{t-3}; 0.8, -3562)$



Obr. A.13: Porovnanie pôvodých a modelovaných dát odhadnutých z modelu LSTAR s  $G(x_{t-3}; 0.25, -4500)$



Obr. A.14: Porovnanie pôvodných a modelovaných dát odhadnutých z modelu LSTAR s  $G(x_{t-3}; 0.8, 4000)$