

Универзитет у Београду
Математички факултет



МАСТЕР РАД

Модели временских серија са променом режима

Студент:
Маријана Миленковић 1080/2018

Ментор:
Јелена Јоцковић

Београд, септембар 2020.

Мојој породици, која је безрезервно уз мене

Садржај

1.	Увод	1
2.	Стационарна линеарна временска серија	3
2.1.	Линеарни модел временске серије.....	3
2.2.	Ауторегресиони модел	4
2.3.	Предвиђања ауторегресионих модела	6
3.	Режими одређени помоћу опажене променљиве	8
3.1.	Репрезентација модела са променом режима	8
3.2.	Модели вишег реда	16
3.3.	Идентификација реда кашњења.....	16
3.4.	Више режима.....	21
3.5.	Стационарност	23
4.	Оцењивање параметара модела.....	26
4.1.	Оцењивање параметара <i>SETAR</i> модела.....	26
4.1.2.	Интервали поверења	27
4.1.3.	Избор променљиве прага.....	28
4.2.	Оцењивање параметара <i>STAR</i> модела	29
4.2.1.	Оцењивање параметра глаткости γ	30
4.3.	Адитивни аутлајери и робусне оцене	30
5.	Тестирање нелинеарности промене режима.....	31
5.1.	Тестирање <i>SETAR</i> – типа нелинеарности.....	32
5.2.	Тестирање <i>STAR</i> - типа нелинеарности.....	33
5.3.	Хетероскедастичност и тестирање нелинеарности	35
6.	Дијагностичка провера	36
6.1.	Дијагностички тестови за <i>SETAR</i> и <i>STAR</i> моделе	37
7.	Предвиђања.....	39
7.1.	Тачкаста предвиђања	39
7.2.	Интервална предвиђања	41
8.	Евалуација предвиђања	43
8.1.	Метрике за евалуацију предвиђања	43
8.2.	DM статистика.....	44
8.3.	Узроци лоших предвиђања	46
9.	Вишедимензиони модели са променом режима	48

10.	Режими одређени помоћу неопажене променљиве	49
11.	Примери	51
11.1.	Пример 1.	51
11.2.	Пример 2.	69
12.	Закључак	87
	Референце.....	88

1. Увод

Иако су линеарни модели временских серија били у првом плану академског и примењеног истраживања, често се откривало да они остављају необјашњеним одређене аспекте економских и финансијских података. Познато је да економски и финансијски системи пролазе кроз различите структурне промене, али и промене које су последица утицаја психолошких, емоционалних, културних и друштвених фактора. Због тога је разумно претпоставити да је за објашњење података у различито време потребно више различитих модела временских серија.

Да би се моделирало нелинеарно понашање неких економских и финансијских временских серија, чини се природним да се омогући разматрање у различитим државама света или режимима и да се омогући да се динамика разликује у различитим режимима. У овом раду биће приказани модели који претпостављају да је динамичко понашање временске серије у сваком од режима описано ауторегресионим *AR* моделом. Овакви модели називају се *TAR*¹ моделима, односно ауторегресионим моделима са прагом, или ауторегресионим моделима са променом режима. Тонг је први 1978. године предложио ову класу нелинеарних модела временских серија, које је детаљно описао 1990. године у својој књизи². Једноставни *AR* модели су веома популарни, па проширивањем *AR* модела како би се омогућило нелинеарно понашање, добијају се нелинеарни модели који се лако могу разумети и интерпретирати. *TAR* модели су с једне стране једноставни, док су са друге стране довољно богати да опишу сложену нелинеарну динамику. На пример, може се показати да они могу имати граничне циклусе чиме се омогућава моделовање периодичних временских серија. Такође, њима се могу моделовати различите асиметрије и појаве скокова које не могу бити ухваћене линеарним моделом. Упркос једноставности обрасца *TAR* модела, постоји много параметара које треба оценити и променљивих које се могу одабрати приликом израде самог модела. Иако је ово ометало његову рану употребу, почетком 2000-их година постигнут је велики напредак у погледу спецификације и процене *TAR* модела.

Када је променљива која утиче на промену прага једна од претходних вредности саме временске серије, онда се говори о *SETAR*³ моделу. На примеру ће бити показано колико је класа ових модела богата и колики је утицај слободних чланова. Ако су само слободни чланови различити у различитим режимима, *SETAR* модели се могу користити за хватање промене нивоа временске серије, под условом да су дисперзије шума различитих режима различите. Такође се могу користити и за хватање адитивних аутлајера⁴ посматране временске серије. Ове моделе детаљније су описали Тсаи⁵ 1989. године и Хансен⁶ 1997. године, који имају различит приступ оцењивања непознатих параметара, али и тестирања *SETAR* нелинеарности. Иако је процедура коју је предложио Тсаи једноставна за извођење, она се не може у потпуности аутоматизовати, посебно за одабир вредности прага/прагова. Зато се користи Хансенова процедура чија је предност у томе

¹ Threshold Autoregressive models

² Tong, H. (1990). Non-Linear Time Series: A Dynamical System Approach. Oxford University Press.

³ Self - Exciting Threshold Autoregressive model

⁴ additive outliers

⁵ Tsay, R.S. (1989). "Testing and Modeling Threshold Autoregressive Processes", Journal of the American Statistical Association

⁶ Hansen, B.E. (1997). "Inference in TAR Models", Studies in Nonlinear Dynamics and Econometrics

што се вредности прагова могу оценити заједно са осталим параметрима модела и у томе што се могу добити и валидни интервали поверења оцене прагова.

Код *SETAR* модела промена режима се дешава када променљива која утиче на промену режима пређе одређени праг. Иако такав модел може открити многе нелинеарне карактеристике које се обично посматрају у економским и финансијским временским серијама, понекад није интуитивно наметати да прелаз из једног режима у други буде нагли, односно прекидни. У неким случајевима је оправдано претпоставити да се прелаз између режима одвија постепено и глатко. Да би се то омогућило, уводе се *STAR*⁷ модели. О приступу моделовања *STAR* модела писао је Терасвирта⁸ 1994. године, док су ван Дијк, Терасвирта и Франсес⁹ 2002. године приказали до тада развијене *STAR* моделе. У зависности од тога која се функција узима како би се омогућила постепена промена прага, дефинишу се различити *STAR* модели. У овом раду пажња ће бити посвећена *STAR* моделима код којих је та функција логистичка. На тај начин се добијају такозвани *LSTAR*¹⁰ модели.

Међутим, с обзиром на то да је могућност лажно доброг уклапања временске серије у нелинеарни модел веома велика, потребно је извршити тестирање (не)линеарности. Постоје тестови којима се може тестирати линеарност против нелинеарности и који имају велику моћ против широког спектра нелинеарних модела временских серија. Они су погодни када нема претходних сазнања о врсти нелинеарности коју временска серија има и када је довољно утврдити да ли постоји (не)линеарност. У пракси, понекад економска или финансијска теорија, или чак стилизоване емпиријске чињенице¹¹, тј. чињенице које су толико конзистентне да се прихватају као истините, могу предложити специфичан облик нелинеарности посматране временске серије. У оваквим ситуацијама се преферира тестирање одређеног облика нелинеарности, или конкретније *SETAR*-, *STAR*- и *LSTAR*- нелинеарности.

Поменути нелинеарни модели са променом режима чине прву класу модела код којих је променљива која утиче на промену режима опажена¹² и они ће бити детаљно описани у овом раду. Биће приказано како се оцењују њихови параметри, који је критеријум одабира модела на скупу за обучавање, како се тестира хипотеза о нелинеарности, како се пореде предвиђања различитих модела са променом режима на посматраном скупу за тестирање. Такође ће се упоредити предвиђања ових модела са онима које даје линеарни *AR* модел.

Другу класу модела са променом режима чине модели код којих је та променљива неопажена¹³. Они неће бити детаљно описани у овом раду, али ће бити дефинисан један од најпознатијих модела ове класе, да би се истакао њен значај, али и створила шира слика о моделима са променом режима уопште. [1], [18]

⁷ Smooth Transition Autoregressive models

⁸ Teräsvirta, T. (1994). "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models", Journal of the American Statistical Association

⁹ van Dijk, D., T. Teräsvirta and P.H. Franses (2002). "Smooth Transition Autoregressive Models — A Survey of Recent Developments", Econometric Reviews

¹⁰ Logistic Smooth Transition Autoregressive model

¹¹ Stylized empirical facts

¹² Observable

¹³ Unobservable

2. Стационарна линеарна временска серија

2.1. Линеарни модел временске серије

Низ опсервација $(y_t, t \in T)$ променљиве у посматраних у различитим временским тренуцима назива се **временска серија**. Скуп T који се посматра је дискретан и такав да су интервали између забележених временских тренутака једнаки. Низ посматраних опсервација почиње у неком тренутку $t = 1$ и завршава се у неком тренутку $t = n$, па се временска серија може записати као (y_1, y_2, \dots, y_n) . Често се претпоставља да постоје неке претходне опсервације $y_0, y_{-1}, y_{-2}, \dots$, или неке будуће опсервације y_{n+1}, y_{n+2}, \dots , па се узорак (y_1, y_2, \dots, y_n) може видети као коначан подниз бесконачног низа $\{y_t\}_{t=-\infty}^{\infty}$, тј.

$$\{y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots, y_n, y_{n+1}, y_{n+2}, \dots\}.$$

Обично је временска серија $\{y_t\}_{t=-\infty}^{\infty}$ задата описивањем t -ог елемента. На пример, ако је временска серија **Гаусов процес белог шума**, онда се записује као :

$$y_t = \varepsilon_t, \quad (2.1)$$

при чему важи да је:

$$E[\varepsilon_t] = 0, \quad (2.2)$$

$$E[\varepsilon_t^2] = \sigma^2, \quad (2.3)$$

$$E[\varepsilon_t \varepsilon_s] = 0, \forall s \neq t \in \mathbb{N}, \quad (2.4)$$

односно, случајне величине ε_t међусобно су независне и имају заједничку нормалну расподелу са очекивањем 0 и дисперзијом σ^2 .

Линеарни модел временске серије y_t може бити посматран као сума детерминистичког и стохастичког процеса, или једноставније, сума онога што се може предвидети и онога што се не може предвидети, користећи знање о историји посматране временске серије. **Историја** или скуп опсервација који је на располагању у тренутку $t - 1$ означава се са Ω_{t-1} , па се линеарни модел записује као

$$y_t = E[y_t | \Omega_{t-1}] + v_t, \quad (2.5)$$

при чему $E[\cdot | \cdot]$ представља оператор условног очекивања, а v_t непредвидиви део и важи да је $E[v_t | \Omega_{t-1}] = 0$. Претпоставља се и да је v_t горе наведени Гаусов бели шум, па се (2.5) може записати као:

$$y_t = E[y_t | \Omega_{t-1}] + \varepsilon_t. \quad (2.6)$$

За случајни процес $\{y_t, t \in T\}$ каже се да је **слабо стационаран** ако важе следећи услови:

$$E[y_t] = \mu < +\infty \text{ за } \forall t \in T, \quad (2.7)$$

$$\text{cov}(y_t, y_{t-k}) = \gamma_k < +\infty \text{ за } \forall t \in T \text{ и } k = 0, 1, 2, \dots. \quad (2.8)$$

2.2. Ауторегресиони модел

Ауторегресиони модел првог реда, у ознаци $AR(1)$ ¹⁴ записује се као:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t, \quad (2.9)$$

где се претпоставља да је слободан члан φ_0 различит од нуле како би се добила временска серија са очекивањем различитим од нула. Одређивањем очекивања обе стране једнакости (2.9), добија се:

$$E[y_t] = \varphi_0 + \varphi_1 E[y_{t-1}] + E[\varepsilon_t]. \quad (2.10)$$

Ако је y_t слабо стационарна временска серија која задовољава (2.7) и (2.8) и ε_t Гаусов бели шум, добија се да је очекивање μ једнако:

$$\mu = \frac{\varphi_0}{1 - \varphi_1}. \quad (2.11)$$

Запис (2.11) има смисла само када је $|\varphi_1| < 1$, што ће се детаљније показати. Користећи рекурзију, (2.9) се може представити на следећи начин:

$$\begin{aligned} y_t &= \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t \\ &= \varphi_0 + \varphi_1(\varphi_0 + \varphi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \varphi_0 + \varphi_1 \varphi_0 + \varphi_1^2 y_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \varphi_0 + \varphi_1 \varphi_0 + \varphi_1^2 (\varphi_0 + \varphi_1 y_{t-3} + \varepsilon_{t-2}) + \varphi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \varphi_0 + \varphi_1 \varphi_0 + \varphi_1^2 \varphi_0 + \varphi_1^3 y_{t-3} + \varphi_1^2 \varepsilon_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t = \dots \\ &= \varphi_1^t y_0 + \sum_{i=0}^{t-1} \varphi_1^i \varphi_0 + \sum_{i=0}^{t-1} \varphi_1^i \varepsilon_{t-i}. \end{aligned} \quad (2.12)$$

Ако се на ову једнакост примени очекивање, добија се да је:

$$E[y_t] = \varphi_1^t y_0 + \varphi_0 \sum_{i=0}^{t-1} \varphi_1^i. \quad (2.13)$$

Када је $|\varphi_1| < 1$, онда се геометријски ред може записати као:

$$\sum_{i=0}^{t-1} \varphi_1^i = \frac{1 - \varphi_1^t}{1 - \varphi_1} < +\infty \text{ за } \forall t \geq 0, \quad (2.14)$$

при чему важи да $\varphi_1^t \rightarrow 0$ када $t \rightarrow +\infty$, па се добија да је:

$$E[y_t] \approx \frac{\varphi_0}{1 - \varphi_1} \text{ за довољно велико } t. \quad (2.15)$$

¹⁴ Autoregressive model

Са друге стране, када је $|\varphi_1| \geq 1$ посматрани геометријски ред не конвергира за велико t , па горе наведено не важи. На пример, када је $\varphi_1 = 1$, добија се да је $E[y_t] = y_0 + \varphi_0 t$ што није константно.

За слабо стационарну серију y_t **аутоковаријациони коефицијент са кашњењем k** , или **коваријација реда k** означава се са $\gamma_k = cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$, за $k \geq 0$, а аутоковаријациона функција представља низ ових коефицијената, $\gamma_0, \gamma_1, \dots$. Да би се одредили аутоковаријациони коефицијенти $AR(1)$ модела, (2.9) се записује на следећи начин користећи (2.11):

$$(y_t - \mu) = \varphi_1(y_{t-1} - \mu) + \varepsilon_t. \quad (2.16)$$

Квадрирањем обе стране једнакости и узимањем очекивања, добија се да је:

$$E[(y_t - \mu)^2] = \varphi_1^2 E[(y_{t-1} - \mu)^2] + E[\varepsilon_t^2] + 2\varphi_1 E[(y_{t-1} - \mu) \varepsilon_t]. \quad (2.17)$$

Посматрањем (2.12) за y_{t-1} види се да је y_{t-1} функција шокова $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$, тј. да y_{t-1} и ε_t нису корелисане. Комбинујући ово са (2.2) и (2.3) види се да је последње очекивање десне стране једнакости једнако нула. Узимајући у обзир да је $|\varphi_1| < 1$, из (2.17) се добија да је:

$$\gamma_0 = \frac{\sigma^2}{1 - \varphi_1^2}. \quad (2.18)$$

Користећи (2.16) аутоковаријација првог реда $AR(1)$ модела може се видети на следећи начин:

$$\begin{aligned} \gamma_1 &= E[(y_t - \mu)(y_{t-1} - \mu)] \\ &= \varphi_1 E[(y_{t-1} - \mu)(y_{t-1} - \mu)] + E[(y_{t-1} - \mu) \varepsilon_t] \\ &= \varphi_1 \gamma_0. \end{aligned} \quad (2.19)$$

За $AR(1)$ модел аутоковаријација k -тог реда може се записати као:

$$\gamma_k = \varphi_1 E[(y_t - \mu)(y_{t-k} - \mu)] = \varphi_1 \gamma_{k-1}, \quad \text{за } k \geq 1. \quad (2.20)$$

Ауторегресиони модел p -ог реда, у ознаци **$AR(p)$** јесте линеарна комбинација претходних p вредности временске серије која се записује као:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t. \quad (2.21)$$

Да би се искористили резултати добијени за $AR(1)$ модел, приказује се његова **карактеристична једначина**:

$$1 - \varphi_1 z = 0, \quad (2.22)$$

а затим и карактеристична једначина $AR(p)$ модела:

$$1 - \varphi_1 z - \dots - \varphi_p z^p = 0. \quad (2.23)$$

Решење, или **корен карактеристичне једначине** (2.22) је $z = \varphi_1^{-1}$. Дакле, услов да је $|\varphi_1|$ мање од 1 временске серије y_t генерисане $AR(1)$ моделом, да би она била слабо стационарна, еквивалентан је услову да је корен карактеристичне једначине (2.22) већи од 1.

Услов слабе стационарности временске серије генерисане $AR(p)$ моделом тада је једноставно да су сва p решења једначине (2.23) већа од 1. Познато је да се (2.23) може записати као:

$$(1 - \alpha_1 z)(1 - \alpha_2 z) \cdots (1 - \alpha_p z) = 0, \quad (2.24)$$

што показује да је услов стационарности еквивалентан томе да су све $\alpha_i, i = \overline{1, p}$ унутар јединичног круга. Ако постоји бар једно k тако је да $|\alpha_k| = 1$, а сви остали α_i такви да је $|\alpha_i| < 1, z = 1$ је решење једначине (2.23). Тада се за временску серију генерисану $AR(p)$ моделом каже да има **јединични корен**. Број јединичних корена указује на то колико пута треба диференцирати временску серију да би она постала стационарна. Ако је $\alpha_k = -1$, реч је о такозваном сезонском диференцирању.

Тестирање стационарности временске серије је веома битно, јер када је она стационарна, њено понашање током времена се може предвидети. Постоји више тестова за тестирање стационарности временске серије а један од њих је **ADF¹⁵ тест** који тестира нулту хипотезу о постојању јединичног корена. Алтернативна хипотеза зависи од верзије теста који се користи, али се обично узима да је у питању стационарна временска серија.

Параметри $AR(p)$ модела датог са (2.21) могу се оценити помоћу обичне **методе најмањих квадрата**, у ознаци **OLS¹⁶**. Може се показати¹⁷ да оцене параметара добијене методом **OLS** јесу конзистентне и асимптотски нормалне, и да се стандардна t -статистика може користити за истраживање значаја ауторегресионих параметара ϕ_1, \dots, ϕ_p . Оцена средње вредности временске серије μ , може се добити као:

$$\hat{\mu} = \frac{\widehat{\varphi}_0}{1 - \widehat{\varphi}_1 - \widehat{\varphi}_2 - \cdots - \widehat{\varphi}_p}. \quad (2.25)$$

Користећи оцене параметара може се добити низ оцењених резидуала $\{\hat{\varepsilon}_t, t \in T\}$.

2.3. Предвиђања ауторегресионих модела

Нека је $\hat{y}_{t+h|t}$ **предвиђена вредност временске серије y_t за h корака унапред**, а y_{t+h} њена реализована вредност у тренутку $t + h$. Придружена **грешка предвиђања¹⁸** означава се са:

$$\xi_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}. \quad (2.26)$$

Као што се параметри модела временске серије одабирају тако да минимизују дисперзију резидуала ε_t , у предвиђању се често сматра пожељним одабрати предвиђање $\hat{y}_{t+h|t}$ које минимизује квадратну грешку предвиђања¹⁹ **SPE**:

$$SPE(h) = E[\xi_{t+h|t}^2] = E[(y_{t+h} - \hat{y}_{t+h|t})^2]. \quad (2.27)$$

¹⁵ Augmented Dickey – Fuller test

¹⁶ Ordinary Least Squares

¹⁷ Box, G.E.P. and G.M. Jenkins, 1970. Time Series Analysis; Forecasting and Control, San Francisco: Holden-Day

¹⁸ forecast error

¹⁹ Squared Prediction Error

Показује се да је предвиђање које минимизује (2.27) условно очекивање за y_{t+h} у тренутку t , тј.

$$\hat{y}_{t+h|t} = E[y_{t+h}|\Omega_t], \quad (2.28)$$

где је са Ω_t означена историја временске серије у тренутку t .

Ово ће детаљније бити показано на примеру $AR(2)$ модела:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t. \quad (2.29)$$

Предвиђање један корак унапред у тренутку t дато је са:

$$\hat{y}_{t+1|t} = \varphi_1 y_t + \varphi_2 y_{t-1}, \quad (2.30)$$

под претпоставком да се знају оцене параметара модела. Предвиђање два корака унапред у тренутку t дато је са:

$$\begin{aligned} \hat{y}_{t+2|t} &= E[y_{t+2}|\Omega_t] = \varphi_1 E[y_{t+1}|\Omega_t] + \varphi_2 y_t \\ &= \varphi_1 \hat{y}_{t+1|t} + \varphi_2 y_t = \varphi_1(\varphi_1 y_t + \varphi_2 y_{t-1}) + \varphi_2 y_t. \end{aligned} \quad (2.31)$$

Из (2.31) може се видети да постоји **рекурзивна веза између предвиђања** један корак унапред и предвиђања два корака унапред у тренутку t . Лако се може показати да важи:

$$\hat{y}_{t+h|t} = \varphi_1 \hat{y}_{t+h-1|t} + \varphi_2 \hat{y}_{t+h-2|t}, \quad (2.32)$$

при чему је $\hat{y}_{t+i|t} = y_{t+i}$, за $i \leq 0$. Рекурзивна веза се може искористити за добијање предвиђања више корака унапред. [1], [19]

3. Режими одређени помоћу опажене променљиве

3.1. Репрезентација модела са променом режима

Најистакнутији модел ове класе модела, који претпоставља да режим који се догоди у тренутку t може бити одређен помоћу опажене променљиве²⁰ q_t , која се назива **променљивом прага**²¹, јесте **ауторегресиони модел са прагом**²², или **ауторегресиони модел са променом режима**, **TAR модел**. Претпоставља се да је режим одређен помоћу вредности q_t у односу на **вредност прага**, која се означава са c .

Дефиниција: Ауторегресиони модел са два режима првог реда, у ознаци **TAR(2, 1)** модел, при чему први број означава број режима, а други ред AR модела сваког режима, са променљивом прага q_t и вредношћу c која утиче на промену прага, записује се на следећи начин:

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_{t-1} + \varepsilon_t, & \text{за } q_t \leq c \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + \varepsilon_t, & \text{за } q_t > c \end{cases} \quad (3.1)$$

Претпоставља се да је $\{\varepsilon_t, t \in T\}$ низ једнакораспоређених независних случајних величина (белог шума) условљен историјом временске серије $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \dots, y_{1-(p-1)}, y_{1-p}\}$, при чему p означава ред $AR(p)$ модела. Такође је $E[\varepsilon_t | \Omega_{t-1}] = 0$ и $E[\varepsilon_t^2 | \Omega_{t-1}] = \sigma^2$, односно постоји претпоставка о **хомоскедастичности** случајних величина ε_t .

Када се за променљиву прага q_t узме нека од претходних вредности саме временске серије, тј. $q_t = y_{t-d}$ за $d > 0$, добија се специјални случај. Како је сада режим одређен самом временском серијом резултујући модел назива се **SETAR**²³ модел.

Дефиниција: **SETAR(2, 1) модел** са два режима при чему је сваки од њих описан ауторегресионим $AR(1)$ моделом, са претходном вредношћу посматране временске серије y_{t-1} као променљивом прага и вредношћу прага c , дат је на следећи начин:

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} \leq c \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} > c \end{cases} \quad (3.2)$$

при чему за случајне величине ε_t важи горе наведено.

Алтернативни запис **SETAR** модела (3.2) је:

$$y_t = (\varphi_{0,1} + \varphi_{1,1} y_{t-1})(1 - I[y_{t-1} > c]) + (\varphi_{0,2} + \varphi_{1,2} y_{t-1})I[y_{t-1} > c] + \varepsilon_t, \quad (3.3)$$

где је $I[A]$ индикатор функција при чему је $I[A] = 1$ ако се деси догађај A и $I[A] = 0$ у супротном.

Наведени **SETAR** модел претпоставља да су границе између два режима дате специфичном вредношћу променљиве прага y_{t-1} . Иако овакав модел може ухватити многе нелинеарне карактеристике, понекад је у супротности са интуицијом захтевати да прелаз између режима буде

²⁰ observable variable

²¹ threshold variable

²² Threshold Autoregressive model

²³ Self - Exciting Threshold Autoregressive model

нагли или прекидни. Некада је оправдано претпоставити да се прелаз одвија постепено и глатко. Постепени прелаз између различитих режима може се добити заменом индикаторске функције $I[y_{t-1} > c]$ у (3.3) непрекидном функцијом $G(y_{t-1}; \gamma, c)$ која се назива **функцијом преласка**²⁴. У овом случају она зависи од променљиве прага y_{t-1} која се глатко мења од 0 ка 1 како y_{t-1} расте. Параметар γ назива се **параметар глаткости** јер одређује глаткоћу²⁵ промене вредности функције $G(y_{t-1}; \gamma, c)$, и на тај начин, прелаз између једног режима у други. Параметар c може бити интерпретиран као праг између два режима који одговарају вредностима $G(y_{t-1}; \gamma, c) = 0$ и $G(y_{t-1}; \gamma, c) = 1$ како y_{t-1} расте.

У *SETAR* моделу, опсервације y_t генеришу се или из првог режима када је $y_{t-1} \leq c$ или из другог режима када је $y_{t-1} > c$. Ако се функција бинарног индикатора замени функцијом глатког преласка $G(y_{t-1}; \gamma, c)$, модел постаје **глатко прелазни ауторегресиони модел**²⁶, *STAR* модел.

Дефиниција: *STAR(2, 1)* модел са два режима при чему је сваки од њих описан ауторегресионим *AR(1)* моделом и одређен функцијом преласка $G(y_{t-1}; \gamma, c)$ са параметром глаткости γ и вредношћу прага c , дат је на следећи начин:

$$y_t = (\varphi_{0,1} + \varphi_{1,1} y_{t-1})(1 - G(y_{t-1}; \gamma, c)) + (\varphi_{0,2} + \varphi_{1,2} y_{t-1})G(y_{t-1}; \gamma, c) + \varepsilon_t. \quad (3.4)$$

Сада се опсервације y_t глатко пребацују између два режима у смислу да динамику временске серије y_t могу одредити оба режима, при чему један режим има више утицаја у једним тренуцима, а други режим више утицаја у другим тренуцима. Друга интерпретација је да *STAR* модели заправо омогућавају „континуум“ режима, тј. описују временску серију чије се суседне опсервације не разликују много међусобно, али су у крајности веома различите, и повезане различитим вредностима функције преласка $G(y_{t-1}; \gamma, c)$. Најчешћи избор функције преласка $G(y_{t-1}; \gamma, c)$ јесте логистичка функција

$$G(y_{t-1}; \gamma, c) = \frac{1}{1 + e^{-\gamma(y_{t-1} - c)}}. \quad (3.5)$$

а резултујући модел назива се *LSTAR*²⁷ модел, при чему се лако види да је $G(c; \gamma, c) = 0.5$.

Дефиниција: *LSTAR(2, 1)* модел са два режима описаним ауторегресионим *AR(1)* моделом, записује се на исти начин као и *STAR(2, 1)* модел (3.4) при чему се наглашава да је функција преласка $G(y_{t-1}; \gamma, c)$ заправо логистичка функција дата са (3.5).

Такође, за функцију преласка може се узети и експоненцијална функција:

$$G(y_{t-1}; \gamma, c) = 1 - e^{-\gamma(y_{t-1} - c)^2}, \quad \gamma > 0, \quad (3.6)$$

и на тај начин добити *ESTAR*²⁸ модел.

²⁴ transition function

²⁵ smoothness

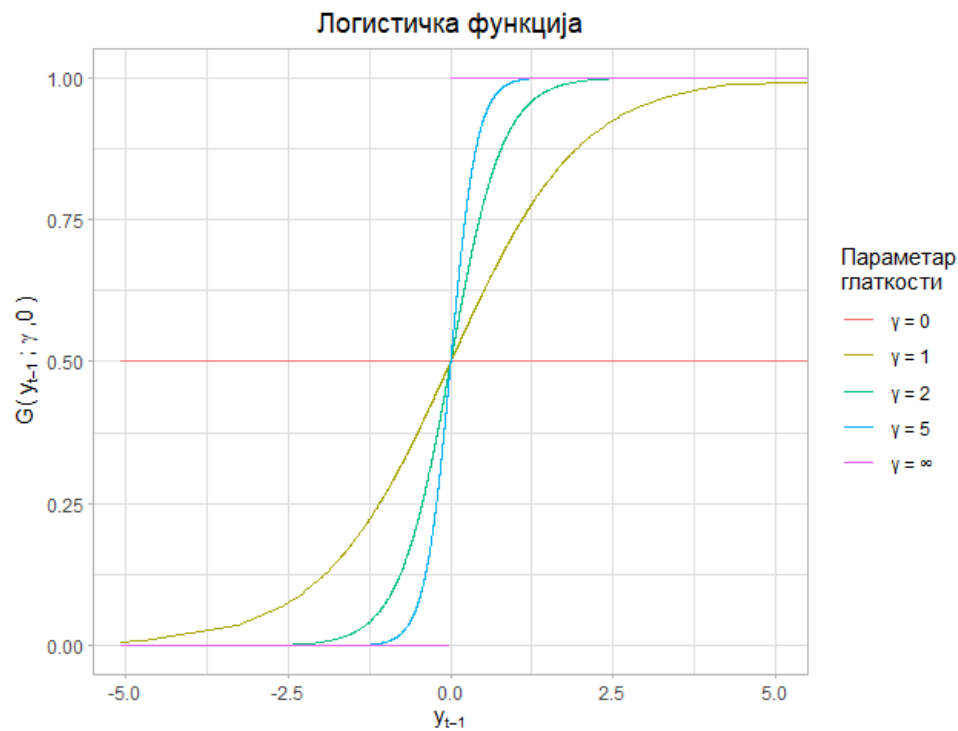
²⁶ Smooth Transition Autoregressive model

²⁷ Logistic Smooth Transition Autoregressive model

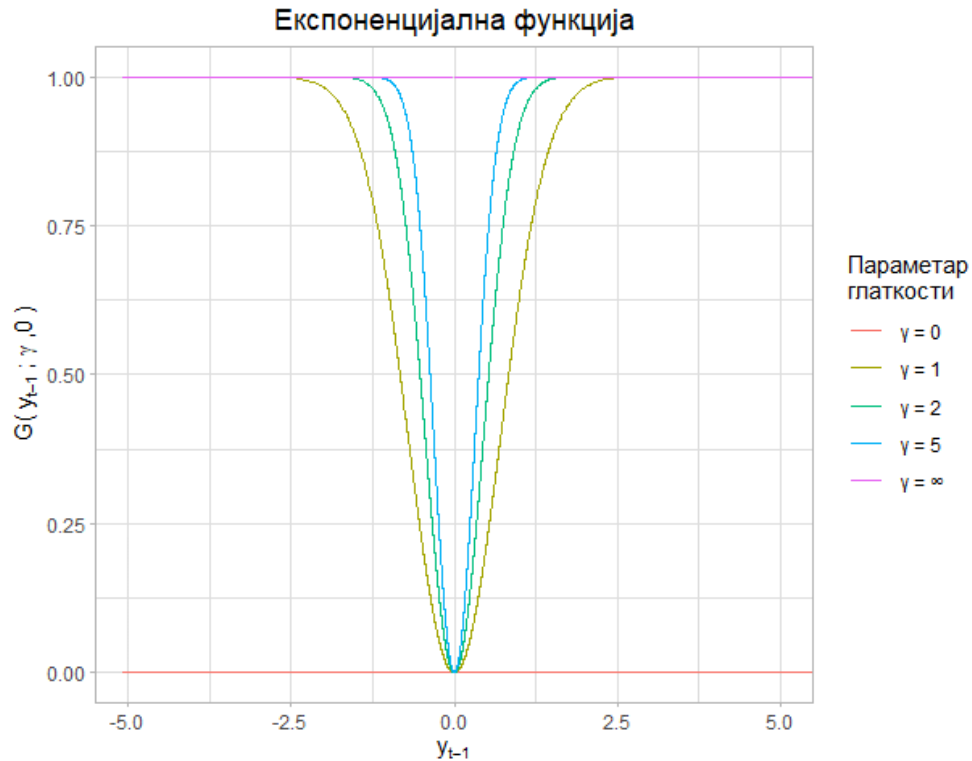
²⁸ Exponential Smooth Transition Autoregressive model

Дефиниција: *ESTAR(2, 1) модел* са два режима описаним ауторегресионим $AR(1)$ моделом, записује се на исти начин као и $STAR(2,1)$ модел (3.4) при чему се наглашава да је функција преласка $G(y_{t-1}; \gamma, c)$ експоненцијална функција дата са (3.6).

Иако су *LSTAR* и *ESTAR* модели слични, они заправо омогућавају различите врсте прелаза између режима. Да би се ово показало, на сликама 3.1 и 3.2 могу се видети логистичка и експоненцијална функција преласка за различите вредности параметра глаткости $\gamma = \{0, 1, 2, 5, \infty\}$ и за вредност прага $c = 0$, респективно.



Слика 3.1. Логистичка функција за различите вредности параметра глаткости γ примењена на бази података која ће касније бити детаљно описана.



Слика 3.2. Експоненцијална функција за различите вредности параметра глаткости γ примењена на бази података која ће касније бити детаљно описана.

Следећа својства се могу лако уочити:

1) Ако је γ мало, обе функције преласка прелазе између 0 и 1 врло глатко и споро, а што је γ веће, оне брже прелазе између 0 и 1.

2) Када $\gamma \rightarrow 0$, логистичка функција је једнака константи 0.5, експоненцијална функција једнака је 0 и оба модела постају линеарна.

3) Када $\gamma \rightarrow \infty$, логистичка функција се приближава индикаторској функцији $I[y_{t-1} > c]$ и *LSTAR* модел се своди на *TAR* модел. Односно, *SETAR* модел (3.3) се може произвољно добро апроксимирати *LSTAR* моделом са функцијом преласка (3.5).

Међутим, када $\gamma \rightarrow \infty$ експоненцијална функција тежи индикаторској функцији $I[y_{t-1} = c]$, па дефинисани *ESTAR* модел не представља специјалан случај *TAR* модела, јер и када $\gamma \rightarrow 0$ и када $\gamma \rightarrow \infty$, он постаје линеаран.

4) Логистичка функција је монотона и *LSTAR* модел се пребацује између два режима постепено, у зависности од тога колико је променљива прага y_{t-1} мања или већа од прага $c = 0$. Експоненцијална функција је симетрична и *ESTAR* модел глатко прелази између два режима у зависности од тога колико је променљива прага y_{t-1} удаљена од прага $c = 0$. За *LSTAR* модел битни су и удаљеност y_{t-1} од $c = 0$ и његов знак; за *ESTAR* модел важна је само удаљеност y_{t-1} од $c = 0$, али не и знак.

Ако се за функцију преласка $G(\cdot)$ узме квадратна логистичка функција

$$G(y_{t-1}; \gamma, c_1, c_2) = \frac{1}{1 + e^{-\gamma(y_{t-1} - c_1)(y_{t-1} - c_2)}}, c_1 \leq c_2, \gamma > 0, \quad (3.7)$$

лако се може видети се да кад $\gamma \rightarrow 0$, модел постаје линеаран, док је за $\gamma \rightarrow \infty$, функција (3.7) једнака 1, ако је променљива прага мања од c_1 , и једнака 0 ако је променљива прага већа од c_2 . Дакле, *STAR* модел са квадратном логистичком функцијом преласка јесте специјалан случај *TAR* модела, и то трорежимног модела, о чему ће касније бити речи. Међутим, у овом раду ће детаљно бити описан само *STAR* моделом са логистичком функцијом преласка, односно *LSTAR* модел.

Прилично велики број динамичких образаца може бити генерисан једноставним моделом као што је *SETAR* модел (3.2). Како би се показало колико је ова класа модела богата, симулираће се 4 различита *SETAR* модела првог реда, при чему ће параметри $\varphi_{1,1}$ и $\varphi_{1,2}$ бити фиксирани, док ће слободни чланови $\varphi_{0,1}$ и $\varphi_{0,2}$ узимати различите вредности.

Пре него што се дискутује о својствима симулираних временских серија које ће бити и приказане, увешће се неки термини нелинеарних модела временских серија. Функцијом

$$F(y_{t-1}) = (\varphi_{0,1} + \varphi_{1,1} y_{t-1})(1 - I[y_{t-1} > c]) + (\varphi_{0,2} + \varphi_{1,2} y_{t-1})I[y_{t-1} > c] \quad (3.8)$$

представљен је детерминистички и предвидиви део модела, познатији као **скелет модела**. Заправо $F(y_{t-1})$ јесте условно очекивање за y_{t-1} у тренутку $t - 1$. За **модел** се каже да је у **равнотежном стању** на y^* , ако је y^* **фиксирана тачка скелета**, тј. ако је $y^* = F(y^*)$. За равнотежно стање се каже да је **стабилно равнотежно стање** ако временска серија y_t конвергира ка y^* кад $t \rightarrow \infty$, при чему је шум ε_t искључен, што једноставно значи да су сви $\varepsilon_t = 0$.

Како код стационарне линеарне временске серије све вредности конвергирају ка њеној очекиваној вредности, она увек има јединствено и стабилно равнотежно стање y^* које је једнако баш том очекивању. Као што ће се видети и на симулираним моделима, нелинеарна временска серија може имати једно (стабилно или нестабилно) равнотежно стање, вишеструка равнотежна стања или да их уопште нема. Осим тога, ако је равнотежно стање јединствено и стабилно, оно не мора бити једнако њеном очекивању. Стабилно равнотежно стање се такође назива и **тачком привлачења**²⁹ јер је у недостатку шума временска серија привучена стабилним равнотежним стањем. Како нелинеарне временске серије могу имати различит број стабилних равнотежних стања, то значи да могу имати и различит број тачака привлачења. Осим тога, стабилно равнотежно стање није једини могући облик тачке привлачења нелинеарне временске серије, што ће се видети на примеру. Слободни чланови ће бити генерисани тако да демонстрирају различита понашања ових модела. Посматраће се узорак обима $n = 200$ опсервација, са $\varphi_{1,1} = -0.5$, $\varphi_{1,2} = 0.5$, $c = 0$ и $\varepsilon_t \sim \mathcal{N}(0, 0.25^2)$ за $t = \overline{1, n}$. Почетна вредност серије је $y_0 \equiv 0$. На графицима слике 3.3 зеленом линијом је означена вредност прага $c = 0$.

Следи приказ кода који се користи за симулирање наведених модела.

²⁹ Attractor


```

# funkcija za simuliranje SETAR modela sa fiksnim parametrima f11 i f22
# za prosledjene vrednosti slobodnih parametara f01 i f02

model <- function(f01,f02, ime_modela){
  n <- 200          # Broj opservacija
  f11 <- -0.5
  f12 <- 0.5
  B <- matrix(c(f01,f11,f02,f12), ncol=2) # parametri modela
  lag <- 1          # broj kasnjenja u svakom od rezima - AR(1) model u svakom od rezima
  nthresh <- 1      # broj pragova
  Thresh <- 0       # vrednost praga (c=0)
  starting <- 0     # pocetna vrednost simulirane serije
  sd <- 0.25        # standardno odstupanje
  thDelay = 0       # da bismo obezbedili da promenljiva  $Y_{t-1}$  bude promenljiva praga
  m1 <- setar.sim( type="simul", B=B , n=n, lag=lag , nthresh = nthresh,
                  Thresh = Thresh, starting = starting, sd=sd )$serie

  plot(m1, type= 'l', xlab="", ylab = expression(italic(y[t])))
  abline(h=0, col="green")

  cat("Model ", ime_modela,
      "\nSrednja vrednost      :", mean(m1),
      "\nBroj opservacija < 0:", sum(m1<0))
  return(m1)
}

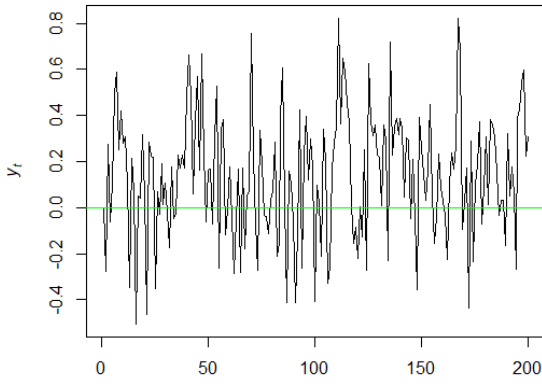
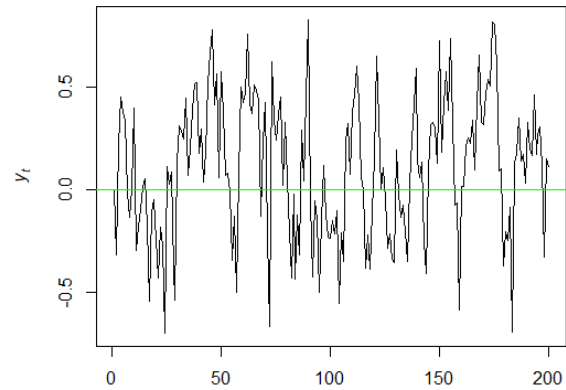
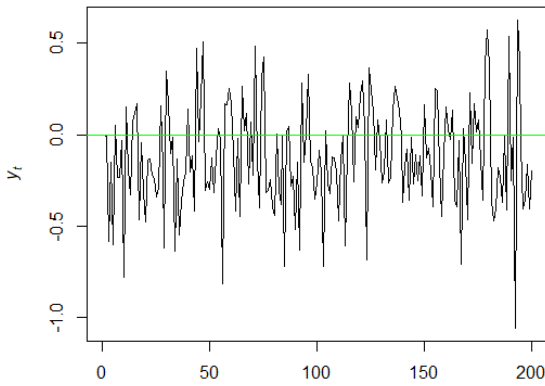
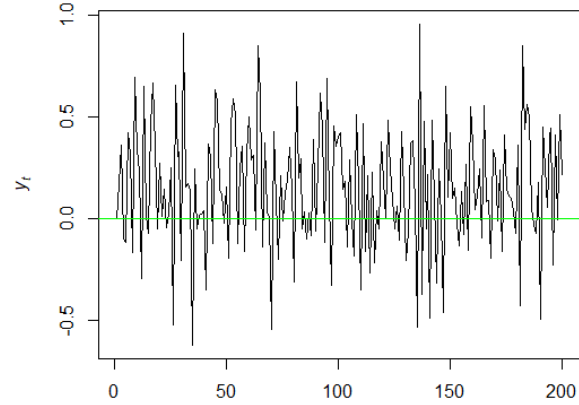
> model_1<-model(0,0,"a")
Model a)
Srednja vrednost      : 0.1387278
Broj opservacija < 0: 57

> model_2<-model(-0.3,0.1,"6")
Model 6)
Srednja vrednost      : 0.1071968
Broj opservacija < 0: 70

> model_3<-model(-0.3,-0.1,"B")
Model B)
Srednja vrednost      : -0.1212466
Broj opservacija < 0: 135

> model_4<-model(0.3,-0.1,"r")
Model r)
Srednja vrednost      : 0.1522672
Broj opservacija < 0: 63

```

Слика 3.3 а) $\varphi_{0,1} = 0, \varphi_{0,2} = 0$ Слика 3.3 б) $\varphi_{0,1} = -0.3, \varphi_{0,2} = 0.1$ Слика 3.3 в) $\varphi_{0,1} = -0.3, \varphi_{0,2} = -0.1$ Слика 3.3 г) $\varphi_{0,1} = 0.3, \varphi_{0,2} = -0.1$

Сада ће детаљно бити образложен сваки од симулираних модела при чему се за рачунање очекивања сваког од режима користи (2.11), односно $\mu = \frac{\varphi_0}{1-\varphi_1}$.

На слици 3.3 (а) приказан је модел

$$y_t = \begin{cases} -0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} \leq 0 \\ 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} > 0 \end{cases}.$$

Средња вредност $AR(1)$ модела у оба режима једнака је нули. Решавањем система $y = -0.5y$ и $y = 0.5y$ добија се да је $y = 0$, односно да је $y^* = 0$ равнотежно стање које је јединствено и стабилно. Међутим, из приложених резултата се види се да је очекивана вредност временске серије различита од нуле. Ово се може разумети ако се примети да серија има тенденцију да напусти нижи режим прилично брзо зато што је $\varphi_{1,1}$ негативно. Заправо, ако је $\varepsilon_t = 0$, серија се одмах враћа горњем режиму јер је $E[y_t | y_{t-1}] = \varphi_{1,1}y_{t-1} > 0$ ако је $y_{t-1} < 0$. Како је $\varphi_{1,2}$ позитивно, очекује се

да серија остане у горњем режиму, када су сви $\varepsilon_t = 0$, иако ће бити повучена према прагу $c = 0$ зато што је $\varphi_{1,2} < 1$. Дакле, у просеку серија ће бити позитивна, што потврђују и резултати.

На слици 3.3 (б) приказан је модел

$$y_t = \begin{cases} -0.3 - 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} \leq 0 \\ 0.1 + 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} > 0 \end{cases}.$$

Решавањем система $y = -0.3 - 0.5y$ и $y = 0.1 + 0.5y$, добија се да је $y^{*1} = -0.2$, а из друге једначине да је $y^{*2} = 0.2$. Дакле, ово је пример модела са два равнотежна стања, при чему су оба стабилна. Интуитиван начин да се види ово је ако се примети да су слободни чланови такви да је очекивана вредност AR модела у нижем режиму $y_{t-1} \leq 0$ унутар нижег режима, као и да је средња вредност AR модела у горњем режиму $y_{t-1} > 0$ унутар горњег режима. Због тога у одсуству спољашњих шокова³⁰, односно шума, временска серија нема тенденцију преласка у други режим, већ ће конвергирати ка очекиваној вредности AR модела унутар режима у ком је. Другим речима, области привлачења за y^{*1} (y^{*2}) су негативне (позитивне) вредности временске серије. На примеру се види да је очекивана вредност симулиране серије позитивна.

На слици 3.3 (в) приказан је модел

$$y_t = \begin{cases} -0.3 - 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} \leq 0 \\ -0.1 + 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} > 0 \end{cases}.$$

Решавањем система $y = -0.3 - 0.5y$ и $y = -0.1 + 0.5y$, добија се да је $y^* = -0.2$ јединствено стабилно равнотежно стање. Види се да је очекивана вредност AR модела у горњем режиму негативна. Дакле, ако серија почиње у горњем режиму, привлачи је нижи режим. Како је очекивана вредност AR модела у нижем режиму такође негативна, та очекивана вредност је тачка привлачења модела.

На слици 3.3 (г) приказан је модел

$$y_t = \begin{cases} 0.3 - 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} \leq 0 \\ -0.1 + 0.5 y_{t-1} + \varepsilon_t, & \text{за } y_{t-1} > 0 \end{cases}.$$

Систем једначина $y = 0.3 - 0.5y$ и $y = -0.1 + 0.5y$ нема решења јер је решење прве једначине $y = 0.2$, што не задовољава услов да је мање или једнако нули. Решење друге једначине је $y = -0.2$, што није веће од нуле. Дакле, овај модел нема ниједно равнотежно стање. Такође, очекиване вредности AR модела у оба режима су у оном другом режиму. Ово указује на то да временска серија нема тачку у којој „се може одморити“. Ако је у горњем режиму, повлачи се према доњем, и обрнуто. Ипак, модел има тачку привлачења. Заправо, модел има тзв. **гранични циклус**. Гранични циклус са k периода дефинише се као скуп тачака y^{*1}, \dots, y^{*k} , где је $y^{*j} = F(y^{*(j-1)})$ за $j = \overline{2, k}$ и $y^{*1} = F(y^{*k})$. Или, једноставније, ако временска серија почиње у некој од тачака y^{*j} , $j = \overline{1, k}$, и ако нема шума, она би се кретала између тих k тачака y^{*1}, \dots, y^{*k} . За овај модел се може показати, као што је то наведено у књизи [1], да је $y^{*1} = 0.06667$, $y^{*2} = -0.06667$ и $y^{*3} = 0.33333$, при чему је се добија да је $F(y^{*1}) = y^{*2}$ рачунајући у горњем режиму, $F(y^{*2}) = y^{*3}$ рачунајући у доњем режиму $F(y^{*3}) = y^{*1}$ рачунајући у горњем режиму. Као што је наведено, може се показати да је гранични циклус област привлачења, тј. да временска серија конвергира ка њему ако се шум искључи.

³⁰ exogenous shocks

Последњи пример илуструје да нелинеарни модели могу садржати унутрашњу³¹ динамику, што значи да и у одсуству шума временска серија може имати колебања. Ово је у супротности са линеарним временским серијама код којих колебања у потпуности узрокују спољашњи шумови ε_t .

Дакле, симулирана четири модела разликују се само у вредностима слободних чланова оба режима. Чињеница да четири модела ипак генеришу временске серије са прилично различитим понашањима показује колико важну улогу имају слободни чланови нелинеарних модела временских серија.

3.2. Модели вишег реда

Иако *SETAR* и *STAR* модели са *AR*(1) моделом у оба режима могу да генеришу велики број различитих образаца, у пракси је често потребно дозволити већи ред *AR* модела у различитим режимима. На пример, у случају два режима, ред *AR* модела може бити постављен на p_1 у нижем режиму и на p_2 у вишем режиму.

Дефиниција: *SETAR*(2, p_1 , p_2) модел са два режима и редовима *AR* модела постављеним на p_1 и p_2 , са променљивом прага y_{t-1} и прагом c , дат је као:

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_{t-1} + \dots + \varphi_{p_1,1} y_{t-p_1} + \varepsilon_t, & \text{за } y_{t-1} \leq c, \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + \dots + \varphi_{p_2,1} y_{t-p_2} + \varepsilon_t, & \text{за } y_{t-1} > c. \end{cases} \quad (3.9)$$

Дефиниција: *STAR*(2, p_1 , p_2) модел са два режима и редовима *AR* модела постављеним на p_1 и p_2 , са променљивом прага y_{t-1} и прагом c , дат је као:

$$y_t = (\varphi_{0,1} + \varphi_{1,1} y_{t-1} + \dots + \varphi_{p_1,1} y_{t-p_1})(1 - G(y_{t-1}; \gamma, c)) + (\varphi_{0,2} + \varphi_{1,2} y_{t-1} + \dots + \varphi_{p_2,1} y_{t-p_2})G(y_{t-1}; \gamma, c) + \varepsilon_t. \quad (3.10)$$

Код модела вишег реда може бити значајно разматрати променљиву y_{t-d} са $d > 1$, као променљиву прага. За такве моделе је теже аналитички утврдити постојање равнотежног стања, тачке привлачења и/или граничних циклуса. Као опште користан начин истраживања својстава скелета модела вишег реда може се користити детерминистичка симулација. Значи да се за дате почетне вредности y_0, \dots, y_{1-p} , са $p = \max(p_1, p_2)$, могу се израчунати вредности које узимају y_1, y_2, \dots ако се постави да је $\varepsilon_t = 0$ за $t \geq 1$. Радећи ово за више различитих почетних вредности, могу се донети неки закључци о карактеристикама скелета модела. [1]

3.3. Идентификација реда кашњења

Важно питање односи се на одређивање одговарајућих редова кашњења p_1 и p_2 у општим моделима са два режима (3.9) и (3.10).

³¹ endogenous

Један од приступа који се обично примењује, нарочито у случају *STAR* модела, јесте да се крене од одређивања реда линеарног $AR(p)$ модела за временску серију y_t и да се претпостави да је тај ред p одговарајући ред за оба режима нелинеарног модела. Ред p је заснован на (парцијалним) аутокорелацијама временске серије или **информационим критеријумима** као што су AIC^{32} или BIC^{33} . Ови критеријуми упоређују прилагођеност модела скупу за обучавање, која се мери преосталом дисперзијом, у односу на број оцењених параметара модела. Постоји више начина записивања ових критеријума који су засновани на методу максималне веродостојности. Да би се боље разумели, AIC ће бити записан на следећи начин:

$$AIC(k) = n \ln \hat{\sigma}^2 + 2k, \quad (3.11)$$

где је $\hat{\sigma}^2$ оцењена дисперзија, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_t^2$, $\hat{\varepsilon}_t$ резидуали модела, n величина узорка а k број оцењених параметара $AR(p)$ модела. Овим критеријумом се постиже варирање између тачности модела и броја оцењених параметара. Јасно, ако је ред p $AR(p)$ модела већи, потребно је оценити више параметара. Стога AIC не награђује само тачност предвиђања, већ „кажњава“ и додавање нових параметара. Главна предност овог критеријума је што спречава преприлагођавање модела. Бајесов информациони критеријум, BIC , може се приказати као:

$$BIC(k) = n \ln \hat{\sigma}^2 + k \ln n, \quad (3.12)$$

где се $\hat{\sigma}^2$ и k дефинишу као и код AIC -а. Како је $\ln n > 2$ за $n > 8$, овај критеријум додатне параметре „кажњава“ много јаче него AIC критеријум, па је ред модела добијен на основу BIC -а вероватно мањи од реда модела који је добијен на основу AIC -а. [2], [3]

Међутим, ова процедура се не препоручује, јер се може лако догодити да ред кашњења који се добија на овај начин није одговарајући. Може се показати да није тешко конструисати нелинеарне моделе временских серија за које су све аутокорелације једнаке нули. На основу тога би се могло закључити да је посматрана нелинеарна временска серија бели шум, што није тачно. Такође, чак и релативно једноставни нелинеарни модели временских серија могу довести до прилично компликованих структура аутокорелације, који могу бити ухваћени само помоћу линеарног $AR(p)$ модела са веома великим редом кашњења p . С друге стране, AIC и BIC могу преценити ред линеарног $AR(p)$ модела.

Алтернативни приступ је одабир реда кашњења p_1 и p_2 у (3.9) и (3.10) директно на основу информационих критеријума. Показано је да је овај приступ доследан, у смислу да ће бити изабран одговарајући ред модела са вероватноћом 1 асимптотски. Међутим, очигледан недостатак овог приступа је да *SETAR* и *STAR* модел морају бити оцењени за све могуће комбинације p_1 и p_2 . Као што је поменуто, у случају коришћења BIC -а, побољшање уклапањем у сложенији модел мора бити велико да надокнади казну насталу због укључивања додатних параметара у модел. Ово може бити проблем у моделовању нелинеарних модела временских серија. Такође се може десити да дође до прецењивања реда модела са променом режима.

У примени модела са променом режима дешава се да је велика већина опсервација у једном од режима. У том случају, побољшање у поређењу са линеарним моделом је вероватно веома мало и недовољно да би се нелинеарни модел одабрао информационим критеријумом. Међутим, треба узети у обзир чињеницу да параметри у додатном режиму, односно у додатним

³² Akaike's information criterion

³³ Bayesian information criterion

режимима, јесу потребни за неки, иако мали, број опсервација. То се може постићи тиме што се не „кажњава“ укључивање додатних параметара за читаву величину узорка, већ се „кажњава“ само за онај број опсервација за које су додатни параметри потребни. Зато је дефинисан **алтернативни AIC** за *SETAR* модел са 2 режима као сума *AIC* -а *AR* модела у оба режима, тј.

$$AIC(p_1, p_2) = n_1 \ln \hat{\sigma}_1^2 + n_2 \ln \hat{\sigma}_2^2 + 2(p_1 + 1) + 2(p_2 + 1), \quad (3.13)$$

где је $n_j, j = 1, 2$, број опсервација j -тог режима а $\hat{\sigma}_j^2, j = 1, 2$, дисперзија резидуала j -тог режима. Иако ε_t може имати исту дисперзију током оба режима, оцењене $\hat{\sigma}_1^2$ и $\hat{\sigma}_2^2$ се могу разликовати.

Алтернативни BIC за *SETAR* модел се може аналогно дефинисати као

$$BIC(p_1, p_2) = n_1 \ln \hat{\sigma}_1^2 + n_2 \ln \hat{\sigma}_2^2 + (p_1 + 1) \ln n_1 + (p_2 + 1) \ln n_2, \quad (3.14)$$

Из записа алтернативног *BIC*-а (3.14) још је очигледније колики је значај броја опсервација у сваком од режима када се рачуна информациони критеријум.

Као што се очекује, за дате горње границе p_1^* и p_2^* , за p_1 и p_2 респективно, најадекватнији редови *AR* модела у оба режима јесу оне вредности за које је информациони критеријум минималан. [1]

Узимајући све ово у обзир, алтернативни *AIC* ће бити примењен на горе симулираним временским серијама како би се показала његова предност у односу на стандардни *AIC*.

Свака симулирана временска серија, чији је ред *AR(p)* модела у сваком од режима једнак 1, уклопиће се у *SETAR* модел коришћењем функције *setar* за генерисање модела за различите вредности параметра p , при чему $p \in \{1, 2, 3, 4, 5, 6\}$. То ће се обезбедити навођењем параметра m функције *setar*. Затим ће се израчунати вредност стандардног *AIC*-а за сваки од тих модела, наћи минимална вредност *AIC*-а као и одговарајућа вредност параметра p . Циљ је показати да се коришћењем стандардне формуле за рачунање *AIC*-а може преценити ред *AR(p)* модела у сваком од режима.

Затим ће се применити функција *selectSETAR* која користи алтернативни *AIC* (3.13), „*pooled – AIC*“, приликом одабира најбољег модела. Њој је довољно проследити само једну, највећу могућу вредност за p , односно m . Навођењем *same.lags = TRUE* обезбеђује се да ред нижег и вишег режима сваког модела буде једнак. Функција ће аутоматски обезбедити да у сваком од режима буде 15% опсервација, што је препоручени минимални проценат опсервација у сваком од режима. Након тога ће се упоредити добијени резултати.

```
aic_function<-function(model, ime_modela){
  # po default-u je:
  # thDelay = 0, tj. promenljiva praga je  $Y_{t-1}$ ,
  # d = 1 obezbedjuje posmatranje serije  $Y_t = f_{00} + f_{01} Y_{t-1}$  u nizem rezimu
  # i analogno u visem rezimu
  # include = "const" obezbedjuje postojanje slobodnih parametara modela

  m_s1<- setar(model, m=1, mL=1, mH=1) # mL je red AR modela u nizem rezimu
  m_s2<- setar(model, m=2, mL=2, mH=2) # mH je red AR modela u visem rezimu
  m_s3<- setar(model, m=3, mL=3, mH=3)
  m_s4<- setar(model, m=4, mL=4, mH=4)
  m_s5<- setar(model, m=5, mL=5, mH=5)
  m_s6<- setar(model, m=6, mL=6, mH=6)

  m_s1_aic <- AIC(m_s1) # funkcija AIC prima fitted model
  m_s2_aic <- AIC(m_s2)
  m_s3_aic <- AIC(m_s3)
  m_s4_aic <- AIC(m_s4)
  m_s5_aic <- AIC(m_s5)
  m_s6_aic <- AIC(m_s6)

  p <- which.min(c(m_s1_aic, m_s2_aic, m_s3_aic, m_s4_aic, m_s5_aic, m_s6_aic))
  min_aic <- min(c(m_s1_aic, m_s2_aic, m_s3_aic, m_s4_aic, m_s5_aic, m_s6_aic))
  cat("Model ", ime_modela,
      "\nNajmanja vrednost AIC-a je ", min_aic, ", za p = ", p[1])
}
```

```
> model_1<-model(0,0,"a")
Model a)
Srednja vrednost : 0.1387278
Broj opservacija < 0: 57
> aic_function(model_1,"a")
Model a)
Najmanja vrednost AIC-a je -550.1731 , za p = 6
> selectSETAR( model_1, m=6, same.lags = TRUE, plot = FALSE )
Using maximum autoregressive order for low regime: mL = 6
Using maximum autoregressive order for high regime: mH = 6
Searching on 134 possible threshold values within regimes with sufficient ( 15% ) number of observations
Searching on 804 combinations of thresholds ( 134 ), thDelay ( 1 ) and m ( 6 )
Results of the grid search for 1 threshold
  thDelay m      th pooled-AIC
1      0 1 -0.11686398    19.13111
2      0 1 -0.11565939    19.92284
3      0 1 -0.11233233    20.04875
4      0 1 -0.09556265    20.19025
5      0 3  0.04411230    21.43958
6      0 3  0.05818588    21.51633
7      0 3 -0.11686398    21.73180
8      0 1 -0.12786351    21.82211
9      0 1 -0.06164096    22.03656
10     0 3  0.05186303    22.13408
```

```

> model_2<-model(-0.3,0.1,"6")
Model 6)
Srednja vrednost : 0.1071968
Broj opservacija < 0: 70
> aic_function(model_2,"6")
Model 6)
Najmanja vrednost AIC-a je -558.4544 , za p = 1
> selectSETAR( model_2, m=6, same.lags = TRUE, plot = FALSE )
Using maximum autoregressive order for low regime: mL = 6
Using maximum autoregressive order for high regime: mH = 6
Searching on 134 possible threshold values within regimes with sufficient ( 15% ) number of observations
Searching on 804 combinations of thresholds ( 134 ), thDelay ( 1 ) and m ( 6 )
Results of the grid search for 1 threshold

```

	thDelay	m	th	pooled-AIC
1	0	1	0.0002690014	11.73722
2	0	1	-0.0097510908	15.98005
3	0	1	0.0000890798	16.13592
4	0	2	0.0002690014	16.58051
5	0	3	0.0002690014	17.11020
6	0	2	-0.0097510908	17.27039
7	0	4	0.0002690014	18.48774
8	0	3	-0.0097510908	18.52994
9	0	4	-0.0097510908	18.61054
10	0	5	0.0002690014	18.84460

```

> model_3<-model(-0.3,-0.1,"b")
Model b)
Srednja vrednost : -0.1212466
Broj opservacija < 0: 135
> aic_function(model_3,"b")
Model b)
Najmanja vrednost AIC-a je -548.8935 , za p = 1
> selectSETAR( model_3, m=6, same.lags = TRUE, plot = FALSE )
Using maximum autoregressive order for low regime: mL = 6
Using maximum autoregressive order for high regime: mH = 6
Searching on 134 possible threshold values within regimes with sufficient ( 15% ) number of observations
Searching on 804 combinations of thresholds ( 134 ), thDelay ( 1 ) and m ( 6 )
Results of the grid search for 1 threshold

```

	thDelay	m	th	pooled-AIC
1	0	1	-0.004773613	21.37940
2	0	1	0.004273031	22.87239
3	0	2	0.004273031	23.38658
4	0	2	-0.004773613	24.23958
5	0	1	-0.006269671	25.91015
6	0	3	0.004273031	26.76360
7	0	1	0.008811641	27.55443
8	0	3	-0.004773613	27.87025
9	0	2	0.008811641	28.34555
10	0	1	-0.008785523	28.69993


```

> model_4<-model(0.3,-0.1,"r")
Model r)
Srednja vrednost      : 0.1522672
Broj opservacija < 0: 63
> aic_function(model_4,"r")
Model r)
Najmanja vrednost AIC-a je -557.5244 , za p = 5
> selectSETAR( model_4, m=6, same.lags = TRUE, plot = FALSE )
Using maximum autoregressive order for low regime: mL = 6
Using maximum autoregressive order for high regime: mH = 6
Searching on 134 possible threshold values within regimes with sufficient ( 15% ) number of observations
Searching on 804 combinations of thresholds ( 134 ), thDelay ( 1 ) and m ( 6 )
Results of the grid search for 1 threshold
  thDelay m      th pooled-AIC
1      0 1 -0.001288481    14.88896
2      0 5 -0.001288481    15.64619
3      0 3 -0.001288481    15.91021
4      0 1 -0.048662371    16.19747
5      0 5 -0.048662371    16.52319
6      0 6 -0.001288481    16.75133
7      0 5 -0.053034119    17.10533
8      0 5 -0.051432564    17.14312
9      0 2 -0.001288481    17.17686
10     0 1 -0.046951330    17.28801

```

Као што се може видети, вредности за p добијене позивом функције *AIC* могу бити веће од оних које су добијене коришћењем функције *selectSETAR*, која користи алтернативни *AIC*, која ће као најбољи ред модела сваког од режима узети баш 1.

3.4. Више режима

Некада је од интереса дозволити више од два режима. На примеру *SETAR* и *STAR* модела биће показано да се то може омогућити на једноставан начин. При томе се могу разликовати два случаја у зависности од тога да ли је режим описан једном променљивом или комбинацијом више њих.

Прво ће бити приказани модели са више режима код којих је режим одређен једном променљивом.

Дефиниција: *SETAR* модел са m режима описаних *AR*(1) моделом, у ознаци ***SETAR*($m, 1$)**, аналоган једнакости (3.3), добија се дефинисањем скупа $m + 1$ прагова c_0, c_1, \dots, c_m , за $-\infty = c_0 < c_1 < \dots < c_m = \infty$, $m \geq 1$ као:

$$y_t = \varphi_{0,j} + \varphi_{1,j} y_{t-1} + \varepsilon_t, \text{ ако је } c_{j-1} < y_{t-1} \leq c_j, j = \overline{1, m}. \quad (3.15)$$

За *STAR* модел процедура је слична. Најпре се расписује једнакост (3.4) како би се показало да се *STAR* модел може представити и на следећи начин:

$$\begin{aligned} y_t &= (\varphi_{0,1} + \varphi_{1,1} y_{t-1})(1 - G) + (\varphi_{0,2} + \varphi_{1,2} y_{t-1})G + \varepsilon_t \\ &= \varphi_{0,1} + \varphi_{1,1} y_{t-1} + (\varphi_{0,2} - \varphi_{0,1})G + (\varphi_{1,2} - \varphi_{1,1})G y_{t-1} + \varepsilon_t. \end{aligned} \quad (3.16)$$

Нека је $x_t = (1, y_{t-1})'$, при чему је $'$ ознака за транспоновање, $\varphi_j = (\varphi_{0,j}, \varphi_{1,j})'$ за $j = 1, 2$ а G функција преласка $G(y_{t-1}; \gamma, c)$. Тада је:

$$y_t = \varphi_1' x_t + (\varphi_2 - \varphi_1)' x_t G(y_{t-1}; \gamma, c) + \varepsilon_t. \quad (3.17)$$

Користећи подскуп c_1, \dots, c_{m-1} прагова дефинисаних изнад за *SETAR* модел и додатни скуп параметара глаткости $\gamma_1, \dots, \gamma_{m-1}$, интуитивно је јасно како се *STAR* модел са m режима може дефинисати.

Дефиниција: *STAR* модел са m режима описаних *AR*(1) моделом, у ознаци ***STAR*($m, 1$)**, са параметрима глаткости $\gamma_1, \dots, \gamma_{m-1}$ и праговима c_0, c_1, \dots, c_m , за које важи $-\infty = c_0 < c_1 < \dots < c_m = \infty$, $m \geq 1$ дат је као:

$$y_t = \varphi_1' x_t + (\varphi_2 - \varphi_1)' x_t G_1(y_{t-1}) + (\varphi_3 - \varphi_2)' x_t G_2(y_{t-1}) + \dots + (\varphi_m - \varphi_{m-1})' x_t G_{m-1}(y_{t-1}) + \varepsilon_t, \quad (3.18)$$

где су $G_j(y_{t-1}) \equiv G_j(y_{t-1}; \gamma_j, c_j)$, $j = \overline{1, m-1}$, функције преласка са параметрима глаткости γ_j и праговима c_j .

Ако је функција преласка модела (3.18) једнака логистичкој функцији (3.5) са одговарајућим параметрима, добија се ***LSTAR*($m, 1$)** модел.

Као пример модела у ком је режим одређен помоћу више од једне променљиве, претпоставиће се да понашање временске серије y_t не зависи само од вредности y_{t-1} у односу на неки праг c_1 , већ и од вредности y_{t-2} у односу на неки праг c_2 . На овај начин добија се *SETAR* модел са четири режима, који се назива **угњежђен *TAR* модел**, односно ***NeTAR*³⁴** модел.

Дефиниција: **Угњежђени *TAR* модел**, односно *SETAR* модел са четири режима који су дати *AR*(1) моделом и одређени помоћу две променљиве прага y_{t-1} и y_{t-2} , са одговарајућим вредностима прагова, c_1 и c_2 , респективно, дефинише се на следећи начин:

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,2} y_{t-1} + \varepsilon_t, & \text{ако је } y_{t-1} \leq c_1 \text{ и } y_{t-2} \leq c_2 \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + \varepsilon_t, & \text{ако је } y_{t-1} \leq c_1 \text{ и } y_{t-2} > c_2 \\ \varphi_{0,3} + \varphi_{1,3} y_{t-1} + \varepsilon_t, & \text{ако је } y_{t-1} > c_1 \text{ и } y_{t-2} \leq c_2 \\ \varphi_{0,4} + \varphi_{1,4} y_{t-1} + \varepsilon_t, & \text{ако је } y_{t-1} > c_1 \text{ и } y_{t-2} > c_2 \end{cases} \quad (3.19)$$

Модел (3.19) се назива угњежђени *TAR* модел јер се временска серија може описати дворежимним *SETAR* моделом са режимима дефинисаним помоћу y_{t-1} , а унутар сваког од тих режима је дворежимни *SETAR* модел са режимима дефинисаним помоћу y_{t-2} , или обрнуто. Да би се ово разумело, представиће се на следећи начин:

Ако је $y_{t-1} \leq c_1$, онда је	$\varphi_{0,1} + \varphi_{1,2} y_{t-1} + \varepsilon_t$	за $y_{t-2} \leq c_2$,
	$\varphi_{0,2} + \varphi_{1,2} y_{t-1} + \varepsilon_t$	за $y_{t-2} > c_2$,
Ако је $y_{t-1} > c_1$, онда је	$\varphi_{0,3} + \varphi_{1,3} y_{t-1} + \varepsilon_t$	за $y_{t-2} \leq c_2$,
	$\varphi_{0,4} + \varphi_{1,4} y_{t-1} + \varepsilon_t$	за $y_{t-2} > c_2$.

На основу овога је очигледно како се одговарајући *STAR* модел са више режима може записати.

³⁴ *Nested TAR*

Дефиниција: *STAR* модел са четири режима који су дати $AR(1)$ моделом и одређени помоћу две променљиве прага y_{t-1} и y_{t-2} , са одговарајућим функцијама преласка $G_j(y_{t-j})$, $j = \overline{1,2}$, које зависе од параметара глаткости γ_j и прагова c_j , дефинише се на следећи начин:

$$y_t = [\varphi'_1 x_t (1 - G_2(y_{t-2})) + \varphi'_2 x_t G_2(y_{t-2})][1 - G_1(y_{t-1})] + [\varphi'_3 x_t (1 - G_2(y_{t-2})) + \varphi'_4 x_t G_2(y_{t-2})]G_1(y_{t-1}) + \varepsilon_t. \quad (3.20)$$

Запис модела (3.20) можда јасније илуструје интерпретацију угњеждених модела. [1]

3.5. Стационарност

Да би се приказало оно што се зна о условима стационарности *SETAR* и *STAR* модела, записаће се најпре општи облик *TAR* модела и услови стационарности који важе за њега.

Дефиниција: *TAR* модел са m режима описаних $AR(p)$ моделом и променљивом прага q_t , у ознаци $TAR(m, p)$, дат је на следећи начин:

$$y_t = \sum_{k=1}^m [\varphi_1^{(k)} y_{t-1} + \dots + \varphi_p^{(k)} y_{t-p} + \varepsilon_t] I(q_t \in \mathcal{R}_k), \quad (3.21)$$

при чему је $\mathcal{R}_k = (r_{k-1}, r_k]$ и важи да је $\bigcup_{k=1}^m \mathcal{R}_k = \mathbb{R}$ за $-\infty = r_0 < r_1 < \dots < r_{m-1} < r_m = \infty$. Функција I је индикаторска функција. За бели шум ε_t важе исте претпоставке као и до сада. Ако важи да је променљива прага q_t независна од y_t и ε_t , онда се за модел (3.21) може рећи да припада класи ауторегресионих модела са случајним коефицијентима који су детаљно описани у књизи Николса и Квина³⁵. Међутим, теоријски резултати ове класе модела не могу се директно применити на модел (3.21) јер структура *TAR* модела не испуњава претпоставку о независности коефицијената која постоји у наведеној књизи. Зато ће се *TAR* модел (3.21) са два режима, за $m = 2$, записати на следећи начин:

$$y_t = \sum_{i=1}^p (\varphi_i + \psi_i^{(1)} I(q_t \in \mathcal{R}_1)) y_{t-i} + \varepsilon_t, \quad (3.22)$$

где је $\varphi_i^{(1)} = \varphi_i + \psi_i^{(1)}$, за $i = \overline{1, p}$. Ово се може јасније видети ако се распише и добије:

$$y_t = \begin{cases} \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t, & q_t \notin \mathcal{R}_1 \\ (\varphi_1 + \psi_1^{(1)}) y_{t-1} + (\varphi_2 + \psi_2^{(1)}) y_{t-2} + \dots + (\varphi_p + \psi_p^{(1)}) y_{t-p} + \varepsilon_t, & q_t \in \mathcal{R}_1. \end{cases}$$

Векторска репрезентација модела (3.22) дата је са:

$$\mathbf{y}_t = (\mathbf{\Phi} + \mathbf{\Psi}^{(1)} I(q_t \in \mathcal{R}_1)) \mathbf{y}_{t-1} + \varepsilon_t, \quad (3.23)$$

где је $\mathbf{y}_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'$, $\varepsilon_t = (\varepsilon_t, \mathbf{0}_{[1 \times (p-1)]})'$. Матрице $\mathbf{\Phi}$ и $\mathbf{\Psi}^{(1)}$ дате су на следећи начин:

³⁵ Nicholls, D.F., Quinn, B.G. *Random Coefficient Autoregressive Models: An Introduction* 1982.

$$\Phi = \begin{bmatrix} \varphi_1 & \varphi_2 & \dots & \varphi_p \\ I_{[(p-1) \times (p-1)]} & \mathbf{0}_{[(p-1) \times 1]} \end{bmatrix} \text{ и } \Psi^{(1)} = \begin{bmatrix} \psi_1^{(1)} & \psi_2^{(1)} & \dots & \psi_p^{(1)} \\ I_{[(p-1) \times (p-1)]} & \mathbf{0}_{[(p-1) \times 1]} \end{bmatrix},$$

при чему је I идентичка матрица и $\mathbf{0}$ нула вектор. Различите карактеристике стохастичких процеса могу се видети посматрањем његових момената. За очекивање модела (3.23) добија се $E[y_t] = (\Phi + \Psi^{(1)}\lambda)E[y_{t-1}]$, где је $\lambda = E[I(q_t \in \mathcal{R}_1)]$. Очекивање је коначно ако су испуњени одговарајући услови за коефицијенте матрица у оба режима. Може се показати да ако оба режима имају матрицу коефицијената са доминантним сопственим вредностима мањим од 1, тада (3.23) неће експлодирати. Ово утиче на стационарност модела (3.23), али указује и на разлике које постоје између TAR модела и $SETAR$ модела. Тачније, последњи резултати се морају прилагодити ако се посматра $SETAR$ модел код ког је $q_t = y_{t-d}$, $d > 0$, јер тада не важи услов о независности променљиве прага и временске серије коју посматрамо.

Стационарност $SETAR$ модела је истраживана на различите начине. Резултати се углавном односе на $SETAR$ моделе за које је $p = 1$, али се они могу уопштити и за $p \geq 1$. [4]

Услови стационарности $SETAR$ модела са два режима описаних $AR(p)$ моделом дати су са:

$$\begin{aligned} \max_k \sum_{i=1}^p |\varphi_i^{(k)}| &< 1 \text{ и} \\ \sum_{i=1}^p \max_k |\varphi_i^{(k)}| &< 1, \end{aligned} \quad (3.24)$$

при чему је са $k \in \{1,2\}$ означен режим који се посматра.

Међутим, показује се да стационарност $SETAR$ модела са $AR(1)$ моделом у оба режима има мање рестриктивне услове, па је тако $SETAR$ модел (3.2) стационаран ако и само ако је неки од следећих услова испуњен :

- (1) $\varphi_{1,1} < 1$, $\varphi_{1,2} < 1$, $\varphi_{1,1}\varphi_{1,2} < 1$;
 - (2) $\varphi_{1,1} = 1$, $\varphi_{1,2} < 1$, $\varphi_{0,1} > 0$;
 - (3) $\varphi_{1,1} < 1$, $\varphi_{1,2} = 1$, $\varphi_{0,2} < 0$;
 - (4) $\varphi_{1,1} = 1$, $\varphi_{1,2} = 1$, $\varphi_{0,2} < 0 < \varphi_{0,1}$;
 - (5) $\varphi_{1,1}\varphi_{1,2} = 1$, $\varphi_{1,1} < 0$, $\varphi_{0,2} + \varphi_{1,2}\varphi_{0,1} > 0$.
- (3.25)

Треба нагласити да (1) дозвољава да један од параметара AR модела буде мањи од -1. Услови (2), (3) и (4) показују на то да AR модел у једном, или чак оба режима, може имати јединични корен. Тада се каже да је временска серија **локално нестационарна**. Услови на слободне чланове $\varphi_{0,1}$ и $\varphi_{0,2}$ су такви да временска серија има склоност ка враћању стационарном режиму, па је стога она глобално стационарна.

Што се тиче стационарности $STAR$ модела, коришћењем Монте Карло симулације пронађена је расподела **нелинеарног ADF теста**, тј. $NLADF^{36}$ теста. Он за разлику од ADF теста, нема асимптотски стандардну нормалну расподелу. Међутим, показало се да је $NLADF$ тест много бољи од на стандардног линеарног ADF теста под алтернативом $ESTAR$ нестационарности.

Груба процена стационарности нелинеарних модела временских серија уопште, јесте провера да ли је скелет модела стабилан. Интуитивно, ако је скелет такав да временска серија има

³⁶ Nonlinear ADF test

склоност да експлодира за неке почетне вредности, серија је нестационарна. Ово се лако може утврдити симулацијом. [\[1\]](#)

4. Оцењивање параметара модела

Да би се тестирала нулта хипотеза о линеарности против алтернативне о конкретној нелинеарности, потребно је оценили параметре нелинеарних модела, односно модела са променом режима. Због једноставности, претпоставиће се да је ред AR модела у сваком од режима једнак, тј. да је $p_1 = p_2 = p$. Такође, приказаће се само опште методе оцењивања параметара $SETAR$ и $STAR$ модела.

4.1. Оцењивање параметара $SETAR$ модела

Параметри од интереса дворежимног $SETAR$ модела (3.9) су φ_{ij} , $i = \overline{0, p}$, $j = 1, 2$, c и σ^2 . Они могу бити оцењени **условном методом најмањих квадрата**. Под претпоставком да су случајне величине ε_t нормално расподељене, оцена која се добија еквивалентна је оцени која се добија методом максималне веродостојности. Да би се показало зашто је метод најмањих квадрата одговарајући, (3.9) се записује као:

$$y_t = (\varphi_{0,1} + \varphi_{1,1} y_{t-1} + \dots + \varphi_{p,1} y_{t-p}) I[y_{t-1} \leq c] + (\varphi_{0,2} + \varphi_{1,2} y_{t-1} + \dots + \varphi_{p,2} y_{t-p}) I[y_{t-1} > c] + \varepsilon_t, \quad (4.1)$$

или, једноставније, као:

$$y_t = \varphi'_1 x_t I[y_{t-1} \leq c] + \varphi'_2 x_t I[y_{t-1} > c] + \varepsilon_t, \quad (4.2)$$

где су $\varphi_j = (\varphi_{0,j}, \varphi_{1,j}, \dots, \varphi_{p,j})'$, $j = 1, 2$ и $x_t = (1, y_{t-1}, \dots, y_{t-p})'$. Када је праг c фиксиран, модел постаје линеаран по осталим параметрима, па се оцена за $\varphi = (\varphi'_1, \varphi'_2)'$ лако добија методом најмањих квадрата као:

$$\hat{\varphi}(c) = \left(\sum_{t=1}^n x_t(c) x_t(c)' \right)^{-1} \left(\sum_{t=1}^n x_t(c) y_t \right), \quad (4.3)$$

где је $x_t(c) = (x'_t I[y_{t-1} \leq c], x'_t I[y_{t-1} > c])'$. Запис $\hat{\varphi}(c)$ се користи да би се нагласило да оцена за φ зависи од вредности прага c . Одговарајући резидуали означени су као $\hat{\varepsilon}_t = y_t - \hat{\varphi}(c)' x_t(c)$ са придруженом дисперзијом $\hat{\sigma}^2(c) = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2(c)$. Оцена за c изведена методом најмањих квадрата може се добити минимизацијом дисперзије резидуала, тј.

$$\hat{c} = \underset{c \in C}{\operatorname{argmin}} \hat{\sigma}^2(c), \quad (4.4)$$

при чему C означава скуп свих дозвољених вредности прага c . Коначна оцена ауторегресионих параметара дата је са $\hat{\varphi} = \hat{\varphi}(\hat{c})$, док је дисперзија резидуала оцењена са $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{c})$. Скуп C дозвољених вредности прага c у (4.4) треба бити такав да сваки режим садржи довољно опсервација да би оцењени ауторегресиони параметри били поуздани. Најчешћи избор скупа C је такав да захтева да сваки режим садржи најмање део π_0 укупног броја опсервација, тј.

$$C = \{c \mid y_{(\lfloor \pi_0(n-1) \rfloor)} \leq c \leq y_{(\lceil (1-\pi_0)(n-1) \rceil)}\}, \quad (4.5)$$

где $y_{(0)}, \dots, y_{(n-1)}$ означава статистике поретка променљиве прага y_{t-1} , $y_{(0)} \leq \dots \leq y_{(n-1)}$, а $[\cdot]$ цео део. Препоручен избор за π_0 је $\pi_0 = 0.15$.

Проблем минимизације израза (4.4) може се решити директном претрагом. Може се показати да се вредност $\hat{\sigma}^2(c)$ не мења када c варира између две уређене статистике променљиве прага y_{t-1} , јер тада опсервације не прелазе из једног режима у други. Зато је довољно израчунати дисперзију резидуала $\hat{\sigma}^2(c)$ само за вредности прага које су једнаке уређеним статистикама променљиве прага y_{t-1} , тј. за $c = y_{(i)}$ за свако i такво да $y_{(i)} \in C$. Показано је да је оцена прага, \hat{c} , добијена методом најмањих квадрата, конзистента када је ред брзине конвергенције једнак n ,³⁷ али и асимптотски зависна од оцена осталих параметара. Касније је изведена и гранична расподела за \hat{c} која зависи само од скалираних параметара; не и од параметара као што су, на пример, параметри φ .³⁸ Оцене ауторегресионих параметара су конзистентне када је ред брзине конвергенција једнак \sqrt{n} , и асимптотски су нормалне.

4.1.2. Интервали поверења

Асимптотска расподела оцене прага c доступна је у затвореној форми³⁹, па се може користити за конструкцију интервала поверења. Међутим, то захтева оцењивање скалираних параметара расподеле. Зато се препоручује алтернативни приступ који је заснован на инверзији теста количника веродостојности, LR ,⁴⁰ да би се тестирала хипотеза о једнакости прага некој специфичној вредности c_0 , дата са

$$LR(c_0) = n \frac{\hat{\sigma}^2(c_0) - \hat{\sigma}^2(\hat{c})}{\hat{\sigma}^2(\hat{c})}. \quad (4.6)$$

Лако се види да је $LR(\hat{c}) = 0$. **Интервал поверења** дат нивоом значајности α , прага c , **дат је скупом** \hat{C}_α који се састоји од вредности c за које се нулта хипотеза не одбацује са прагом значајности α , тј.

$$\hat{C}_\alpha = \{c : LR(c) \leq z(\alpha)\}, \quad (4.7)$$

где је $z(\alpha) = 100\alpha$ перцентил⁴¹ асимптотске расподел LR статистике. Оцене ауторегресионих параметара φ_1 и φ_2 имају асимптотски нормалну расподелу, па се може конструисати асимптотски 95% интервал поверења за, на пример, $\varphi_{1,2}$ као $(\hat{\varphi}_{1,2} - 1.96\hat{\sigma}_{\varphi_{1,2}}, \hat{\varphi}_{1,2} + 1.96\hat{\sigma}_{\varphi_{1,2}})$ где је $\hat{\sigma}_{\varphi_{1,2}}$ оцењена стандардна грешка за $\varphi_{1,2}$. Међутим, показано је да интервали поверења добијени на овакав начин нису одговарајући.⁴² Алтернативна процедура је налажења 95% интервала поверења за φ_1 и φ_2 за сваку вредност c скупа \hat{C}_α , а као коначни интервал поверења узима се унија тих

³⁷ Chan K.S., *Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model*, The Annals of Statistics, 21, 520-533. 1993.

³⁸ Hansen, B.E., *Inference in TAR models*, Studies in Nonlinear Dynamics and Econometrics 2, 1-14, Quarterly Journal April 1997, Volume 2, Number 1 The MIT Press

³⁹ closed - form

⁴⁰ Likelihood Ratio

⁴¹ percentile

⁴² Hansen, B.E., *Inference in TAR models*, Studies in Nonlinear Dynamics and Econometrics 2, 1-14, Quarterly Journal April 1997, Volume 2, Number 1 The MIT Press

интервала. На основу симулација добијено је да је $\alpha = 0.8$ разуман ниво поузданости за скуп \hat{C}_α у овом случају.

Поменути два приступа су различита па су другачији и резултати који се добијају. У књизи [1] је показано да се код једне временске серије интервал поверења прага c базиран на асимптотски нормалној расподели и интервал поверења $\hat{C}_{0.8}$ поклапају са веома малим одступањима, док код друге временске серије то не важи. Код ње је интервал поверења базиран на асимптотски нормалној расподели ужи од интервала поверења $\hat{C}_{0.8}$, који је у овом случају дат као унија подинтервала који немају заједничких елемената. Разлог за то може бити непрецизно оцењен праг c .

4.1.3. Избор променљиве прага

До сада се претпостављало да је променљива прага q_t , која дефинише који се режим јавља у било ком тренутку времена, позната и једнака y_{t-1} . У пракси је она непозната и веома је важно одредити је. Када је реч о *SETAR* моделима, довољно је посматрати претходне вредности саме временске серије, y_{t-d} , за позитиван цео број d , као кандидата променљиве прага.

Да би се добила оцена параметра d претраживање по мрежи⁴³ (4.4) може се допунити претрагом по d тако да минимизациони проблем постане једнак:

$$(\hat{c}, \hat{d}) = \underset{c \in C, d \in D}{\operatorname{argmin}} \hat{\sigma}^2(c, d), \quad (4.8)$$

где је $D = \{1, \dots, d^*\}$ скуп потенцијалних вредности параметра d . Запис $\hat{\sigma}^2(c, d)$ се користи да би се нагласило да оцена дисперзије резидуала сада зависи од два параметра, c и d . Како је простор параметара D дискретан, оцена \hat{d} добијена методом најмањих квадрата је веома доследна и може се сматрати да је d познато приликом рачунања интервала поверења осталих параметара.

Може се приметити да постоји петља у делу процедуре дискутоване до сада. На пример, да би се одредили одговарајући редови *AR* модела са два режима у *AIC*-у дефинисаном са (3.13), за променљиву прага се претпостављало да је позната, док се за одређивање променљиве прага користећи претрагу по мрежи, као што је горе наведено, претпоставља да је ред *AR* модела познат. Један од начина да се прекине ова петља је да се потрага одговарајуће променљиве прага укључи у минимизацију информационог критеријума. Прецизније, вршила би се минимизација *AIC*-а преко p_1, p_2 и d , иако се на овај начин повећава рачунска захтевност јер је $p_1^* \cdot p_2^* \cdot d^*$ различитих модела које треба оценити. [1]

⁴³ grid search

4.2. Оцењивање параметара *STAR* модела

Оцењивање параметара *STAR* модела (3.10) са два режима који су описани $AR(p)$ моделом, релативно је једноставна примена **нелинеарне методе најмањих квадрата**⁴⁴ *NLS*, тј. параметри $\theta = (\varphi'_1, \varphi'_2, \gamma, c)'$ могу се оценити помоћу:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} Q_n(\theta) = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^n [y_t - F(x_t; \theta)]^2, \quad (4.9)$$

где је $F(x_t; \theta)$ скелет модела дат са $F(x_t; \theta) = \varphi'_1 x_t (1 - G(y_{t-1}; \gamma, c)) + \varphi'_2 x_t G(y_{t-1}; \gamma, c)$. Под претпоставком да су грешке ε_t нормално расподељене, метода *NLS* је еквивалентна методи максималне веродостојности. Када услов о нормалној расподељености грешака ε_t није испуњен, овај метод се назива квази метод максималне веродостојности⁴⁵. Под одређеним условима регуларности $E[\varepsilon_t^2] < \infty$, оцена добијена методом *NLS* је конзистентна и асимптотски нормална, тј. важи да је:

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, C), \quad (4.10)$$

где θ_0 означава праве вредности параметара θ , а C асимптотску матрицу коваријације параметара θ која може бити конзистентно оцењена неком од уобичајених нелинеарних оптимизационих метода.

Веома је битан избор почетних вредности јер добре почетне вредности могу смањити оптерећеност оптимизационог алгоритма. За фиксиране вредности параметара функције преласка, γ и c , *STAR* модел је линеаран у ауторегресионим параметрима φ_1 и φ_2 , што је слично *SETAR* моделу. Дакле, за дате γ и c оцена за $\varphi = (\varphi'_1, \varphi'_2)'$ може се добити помоћу обичне методе најмањих квадрата као:

$$\hat{\varphi}(\gamma, c) = \left(\sum_{t=1}^n x_t(\gamma, c) x_t(\gamma, c)' \right)^{-1} \left(\sum_{t=1}^n x_t(\gamma, c) y_t \right), \quad (4.11)$$

где је $x_t(\gamma, c) = (x'_t (1 - G(y_{t-1}; \gamma, c)), x'_t G(y_{t-1}; \gamma, c))'$, а запис $\hat{\varphi}(\gamma, c)$ означава да оцена за φ зависи од γ и c . Одговарајући резидуали могу се добити као $\hat{\varepsilon}_t = y_t - \hat{\varphi}(\gamma, c)' x_t(\gamma, c)$ са придруженом дисперзијом $\hat{\sigma}^2(\gamma, c) = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2(\gamma, c)$. Разумне почетне вредности нелинеарног оптимизационог алгоритма могу се добити из дводимензионалне претрага по γ и c тако што се одаберу они оцењени параметри који дају најмању вредност дисперзије резидуала $\hat{\sigma}^2(\gamma, c)$.

⁴⁴ Nonlinear Least Squares

⁴⁵ Quasi - Maximum Likelihood

4.2.1. Оцењивање параметра глаткости γ

Добијање прецизне оцене параметра глаткости γ може бити веома тешко, јер чак и разумне оцене параметра γ могу имати веома велике стандардне грешке. Показује се, ипак, да се за велике промене вредности γ облик логистичке функције (3.5) мења само мало, што ублажава проблем. Као што је очекивано, за добијање што боље оцене за γ потребно је много опсервација из непосредне околине прага c . Међутим, и оцена прага c је подложна сличним проблемима, али и вредност параметра c , која статистички није различита од нуле, не утиче на укупан квалитет модела. Ипак, у пракси није увек могуће обезбедити велики број опсервација па оцена за γ остаје прилично непрецизна и често се чини статистички безначајном, али то не треба тумачити као доказ против присуства *STAR*-типа нелинеарности. [1] , [15]

4.3. Адитивни аутлајери и робусне оцене

Нестандардне опсервације се могу често јавити у пракси, па је битно одредити да ли оне припадају посматраној временској серији или се могу посматрати као аутлајери. Веома је важно проучити присуство оваквих опсервација као и утицај који имају на моделовање временских серија, њихова предвиђања и статистичке анализе уопште.

Постоји више врста нестандартних опсервација. Један пример су адитивни аутлајери који се могу посматрати као опсервације које су из посматране временске серије, али увећане или умањене за неку вредност. Та вредност може бити не-нула због грешке у мерењу. Тада би се, на пример, могло закључити да, иако то можда није тачно, повраћају на берзи узимају неочекивано велике апсолутне вредности, што утиче на закључке целокупног економског система. Такође, овакве опсервације могу утицати на генерисање модела са променом режима јер се може закључити да је потребан додатни режим како би оне биле ухваћене.

Присуство нестандартних опсервација утиче и на поузданост оцењених параметара модела. Зато је некад потребно користи робусне оцене које се труде да опонашају стандардне методе оцењивања параметара, за које често постоје различите претпоставке које морају бити испуњене. Једна од претпоставки је нормална расподељеност резидуала која не мора бити испуњена баш због присуства нестандартних опсервација. Међутим, чак и ако су претпоставке испуњене само приближно, робусне оцене ће и даље бити довољно добре, имаће малу пристрасност, а за довољно велики обим узорка биће и асимптотски непристрасне.

На основу овога се може закључити колики је значај робусних оцена модела са променом режима, иако у овом раду неће бити више речи о њима. [5] , [6]

5. Тестирање нелинеарности промене режима

С обзиром на то да су нелинеарни модели веома флексибилни, могућност лажног доброг уклапања у било који скуп података временских серија веома је велика. Због тога се препоручује да се изврши **тест линеарности против нелинеарности** пре него што се генерише сложен нелинеарни модел. Постоје различити портмантео тестови⁴⁶, који су обично засновани на резидуалима и немају специфичну алтернативну хипотезу, па у случају да је линеарност одбијена, не зна се који нелинеарни модел треба узети у обзир. Зато се јавила потреба за развијањем тестова линеарности против конкретне нелинеарности, као и различити приступи који се у том случају користе. [7]

Када је реч о моделима временских серија са променом режима, можда је најважније одредити да ли додатни режим, у односу на један режим описан линеарним *AR* моделом, објашњава динамично понашање посматране временске серије. Један од начина да се ово утврди јесте упоређивање одговарајућег модела са променом режима и линеарног модела добијених на скупу за обучавање. Линеарни модел се узима као нулта хипотеза, а модел са променом режима као алтернативна. У случају дворежимног модела, нулта хипотеза H_0 може бити представљена као једнакост ауторегресионих параметара оба режима, против алтернативне хипотезе H_1 о њиховој различитости. Прецизније,

$$\begin{aligned} H_0 : \varphi_1 &= \varphi_2 \\ H_1 : \varphi_{i,1} &\neq \varphi_{i,2} \end{aligned} \quad (5.1)$$

за најмање једно $i \in \{0, \dots, p\}$.

Статистички тестови који као алтернативу имају неки од до сада наведених модела са променом режима, имају тзв. проблем **неидентификованих**⁴⁷ параметара под нултом хипотезом. Прецизније, нелинеарни модели садрже одређене параметре који нису ограничени нултом хипотезом, нити су присутни у линеарном моделу. У *SETAR* и *STAR* моделу праг c је неидентификован параметар, док је у *STAR* моделу параметар γ такође неидентификован. Због присуства оваквих параметара, конвенционална статистичка теорија не може се применити за добијање асимптотске расподеле тест статистике, јер тест статистике имају нестандартну расподелу, за коју често није доступан аналитички израз. То подразумева да се критичне вредности морају одредити помоћу симулираних метода.

Иако су методе за оцену модела са променом режима доступне, тек након оцењивања *SETAR* модела, може се проценити релевантност додатног режима. За *STAR* модел, са друге стране, доступна је *LM* тест статистика⁴⁸ која избегава оцену модела под алтернативном хипотезом. [1]

⁴⁶ portmanteau tests

⁴⁷ nuisance

⁴⁸ Lagrange Multiplier test

5.1. Тестирање *SETAR* – типа нелинеарности

Проблем неидентификованих параметара код тестирања линеарности насупрот алтернативи *SETAR* модела решен је коришћењем оцена *SETAR* модела за дефинисање теста коефицијента веродостојности⁴⁹ када су грешке независне и нормално расподељене. Тест заснован на оценама добијених методом најмањих квадрата⁵⁰, који тестира ограничења која су дата нултом хипотезом, дат је са :

$$F(\hat{c}) = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2(\hat{c})}{\hat{\sigma}^2(\hat{c})}. \quad (5.2)$$

Оцена $\tilde{\sigma}^2$ је оцена дисперзије резидуала под нултом хипотезом о линеарности (*SETAR*(1) модел) дата са $\tilde{\sigma}^2 = \sum_{t=1}^n \tilde{\varepsilon}_t^2$, за $\tilde{\varepsilon}_t = y_t - \hat{\phi}' x_t$. Оцена $\hat{\sigma}^2(\hat{c})$ дисперзије резидуала *SETAR*(2) модела зависи од оцене \hat{c} .

Лако се може видети да је овај тест једнак конвенционалном *F*-тесту, односно Валдовом тесту⁵¹, као и *LM*-тесту. Статистика (5.2) је монотона трансформација параметра $\hat{\sigma}^2(\hat{c})$ јер $F(\hat{c})$ расте када $\hat{\sigma}^2(\hat{c})$ расте, и обрнуто. Како \hat{c} минимизује дисперзију резидуала у скупу C дозвољених вредности за праг, $F(\hat{c})$ је еквивалентна супремуму на скупу C тест статистика $F(c)$. На тај начин добија се тзв. *sup* – *LR* тест:

$$F(\hat{c}) = \sup_{c \in C} F(c), \quad (5.3)$$

где је:

$$F(c) = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2(c)}{\hat{\sigma}^2(c)}, \quad (5.4)$$

а $\hat{\sigma}^2(c) = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2(c)$ дисперзија резидуала. Велике вредности за $F(\hat{c})$ воде ка одбацивању нулте хипотезе, али да би се имплементирао овај тест мора се знати расподела под нултом хипотезом. Статистика $F(c)$ може бити израчуната као $n R^2$, где је R^2 коефицијент детерминације вештачке регресије на $x_t I(y_{t-1} \leq c)$ и $x_t I(y_{t-1} \geq c)$. Дакле, $F(c)$ има асимптотски χ^2 расподелу са $p + 1$ степени слободе, па је тест статистика (5.3) супремум броја зависних статистика од којих свака има χ^2 расподелу. Ово показује да расподела за $F(\hat{c})$ није стандардна. Како је тешко одредити тачан облик зависности између различитих статистика $F(c)$, критичне вредности се најлакше одређују симулацијом. Међутим, испоставља се да ако вредност статистике (5.2) није статистички значајна у поређењу са χ^2 расподелом, онда она није статистички значајна ни када се пореди са тачном асимптотском расподелом. [1]

Овај тест се може уопштити за тестирање *SETAR*(*i*) нелинеарности против *SETAR*(*k*) нелинеарности, који је имплементиран у програмском језику *R*, па се (5.2) записује на следећи начин:

$$F_{ik} = n \frac{S_i - S_k}{S_k}, \quad (5.5)$$

⁴⁹ Likelihood ratio test

⁵⁰ Least-squares test

⁵¹ Waldov test

где је, као што се очекује, S_i дисперзија резидуала $SETAR(i)$ модела, а S_k дисперзија резидуала $SETAR(k)$ модела, за $i < k$. Узимајући ово у обзир, запис F_{12} указује на тестирање линеарности против $SETAR(2)$ модела, F_{13} на тестирање линеарности против $SETAR(3)$ модела. Такође, за тестирање $SETAR(2)$ нелинеарности против $SETAR(3)$ нелинеарности користи се F_{23} , што омогућава тестирање преостале нелинеарности, о чему ће бити више речи у поглављу 6.1..

Приказани тест има велику моћ, али је потребно увести рестрикције за параметре глаткости γ_j тако да у сваком од режима буде бар 10% укупног броја опсервација. [8]

5.2. Тестирање $STAR$ - типа нелинеарности

Код тестирања линеарности против $STAR$ нелинеарности могуће је на више начина решити проблем неидентификованих параметара. Тачније, постоји више од једног ограничења које се може користити.

Поред једнакости AR параметара у оба режима, $H_0 : \varphi_1 = \varphi_2$, нулта хипотеза се може изразити и као $H'_0 : \gamma = 0$. Ако је $\gamma = 0$, логистичка функција (3.5) једнака је 0.5 за све вредности y_{t-1} и $STAR$ модел се своди на линеарни AR модел са параметрима $\frac{1}{2}(\varphi_1 + \varphi_2)$. Међутим, без обзира на то која се формулација хипотезе користи, модел садржи неидентификоване параметре. Када се користи H_0 за карактерисање нулте хипотезе о линеарности, параметри γ и c у функцији преласка су неидентификовани параметри. Када се користи хипотеза H'_0 , праг c и параметри φ_1 и φ_2 су неидентификовани параметри, при чему се лако види да у случају коришћења хипотезе H'_0 , параметри φ_1 и φ_2 могу узети било које вредности све док је њихов просек исти.

Приступ који се користи у овом случају за решавање проблема неидентификованих параметара разликује се од приступа који се користи код $SETAR$ модела. Испоставља се да је у случају тестирања линеарности против $STAR$ нелинеарности, могуће користити LM - тест статистику која има асимптотски χ^2 расподелу. Главна предност могућности коришћења ове статистике је та што није неопходно оцењивати модел под алтернативном хипотезом.

Да би се показало зашто је овај приступ добар, разматра се $STAR$ модел дат у (3.10), али се записује као:

$$y_t = \frac{1}{2}(\varphi_1 + \varphi_2)' x_t + (\varphi_2 - \varphi_1)' x_t G^*(y_{t-1}; \gamma, c) + \varepsilon_t, \quad (5.6)$$

где је $G^*(y_{t-1}; \gamma, c) = G(y_{t-1}; \gamma, c) - \frac{1}{2}$. Под нултом хипотезом $H'_0 : \gamma = 0$, вредност функције $G^*(y_{t-1}; \gamma, c)$ једнака је нула, па се предлаже апроксимација функције $G^*(y_{t-1}; \gamma, c)$ Тејлоровом апроксимацијом првог реда око $\gamma = 0$, тј.

$$T_1(y_{t-1}; \gamma, c) \approx G^*(y_{t-1}; 0, c) + \gamma \frac{\partial G^*(y_{t-1}; \gamma, c)}{\partial \gamma} \Big|_{\gamma=0} = \frac{1}{4}\gamma(y_{t-1} - c), \quad (5.7)$$

где се користи да је $G^*(y_{t-1}; 0, c) = 0$. Након замене $G^*(\cdot)$ са $T_1(\cdot)$ у (5.6) добија се помоћни регресиони модел

$$y_t = \beta_{0,0} + \beta'_0 \tilde{x}_t + \beta'_1 \tilde{x}_t y_{t-1} + \eta_t, \quad (5.8)$$

где је $\tilde{x}_t = (y_{t-1}, \dots, y_{t-p})'$, а $\beta_j = (\beta_{1,j}, \dots, \beta_{p,j})'$, $j = 0, 1$, функције оригиналних параметара *STAR* модела. Може се показати да су они дати са:

$$\beta_{0,0} = \frac{1}{2}(\varphi_{0,1} + \varphi_{0,2}) - \frac{1}{4}\gamma c(\varphi_{0,2} - \varphi_{0,1}) \quad (5.9)$$

$$\beta_{1,0} = \frac{1}{2}(\varphi_{1,1} + \varphi_{1,2}) - \frac{1}{4}\gamma c[(\varphi_{0,2} - \varphi_{0,1}) - (\varphi_{0,2} - \varphi_{0,1})] \quad (5.10)$$

$$\beta_{i,0} = \frac{1}{2}(\varphi_{i,1} + \varphi_{i,2}) - \frac{1}{4}\gamma c(\varphi_{i,2} - \varphi_{i,1}), \quad i = \overline{2, p} \quad (5.11)$$

$$\beta_{i,1} = \frac{1}{4}\gamma c(\varphi_{i,2} - \varphi_{i,1}), \quad i = \overline{1, p} \quad (5.12)$$

Горње једнакости показују да рестрикција $\gamma = 0$ подразумева да је $\beta_{i,1} = 0$ за $i = \overline{1, p}$. Дакле, тестирање нулте хипотезе $H'_0: \gamma = 0$ у (5.6) еквивалентно је тестирању нулте хипотезе $H''_0: \beta_1 = 0$ у (5.8). Оваква нулта хипотеза може бити тестирана стандардном тест статистиком на директан начин. Под нултом хипотезом о линеарности, тест статистика асимптотски има χ^2 расподелу са $p + 1$ степени слободе. Како статистика не тестира оригиналну нулту хипотезу $H'_0: \gamma = 0$, обично се назива *LM* – тест статистиком.

Међутим, горе наведена тест статистика нема моћ у ситуацијама када се само слободни чланови разликују у различитим режимима, тј. када је $\varphi_{0,1} \neq \varphi_{0,2}$ али $\varphi_{i,1} = \varphi_{i,2}$ за $i = \overline{1, p}$. Ово се види директно из (5.12) који показује да ниједан од параметара $\beta_{i,1}$, $i = \overline{1, p}$, не зависи од $\varphi_{0,2}$ и/или $\varphi_{0,1}$. Да би се ово решило,⁵² предложено је замењивање функције преласка $G^*(y_{t-1}; \gamma, c)$ Тејлоровом апроксимацијом трећег реда око $\gamma = 0$, тј.

$$\begin{aligned} T_3(y_{t-1}; \gamma, c) &\approx G^*(y_{t-1}; 0, c) + \gamma \frac{\partial G^*(y_{t-1}; \gamma, c)}{\partial \gamma} \Big|_{\gamma=0} + \frac{\gamma^2}{2!} \frac{\partial^2 G^*(y_{t-1}; \gamma, c)}{\partial \gamma^2} \Big|_{\gamma=0} \\ &+ \frac{\gamma^3}{3!} \frac{\partial^3 G^*(y_{t-1}; \gamma, c)}{\partial \gamma^3} \Big|_{\gamma=0} = \frac{1}{4}\gamma(y_{t-1} - c) + \frac{1}{48}\gamma^3(y_{t-1} - c), \end{aligned} \quad (5.13)$$

где се користи чињеница да је други извод функције $G^*(y_{t-1}; \gamma, c)$ са рестрикцијом $\gamma = 0$ оцењен са нула. Коришћењем ове апроксимације добија се помоћни модел:

$$y_t = \beta_{0,0} + \beta'_0 \tilde{x}_t + \beta'_1 \tilde{x}_t y_{t-1} + \beta'_2 \tilde{x}_t y_{t-1}^2 + \beta'_3 \tilde{x}_t y_{t-1}^3 + \eta_t, \quad (5.14)$$

где су $\beta_{0,0}$ и $\beta_j, j = 1, 2, 3$, поново функције параметара $\varphi_1, \varphi_2, \gamma$ и c . Посматрањем тачних односа између параметара модела (5.14) и оригиналних параметара, показује се да нулта хипотеза $H'_0: \gamma = 0$ сада одговара хипотези $H''_0: \beta_1 = \beta_2 = \beta_3 = 0$, што може бити тестирано стандардним *LM* - тестом. Под нултом хипотезом о линеарности тест статистика има асимптотски χ^2 расподелу.

[1]

⁵² Luukkonen, R., P. Saikkonen and T. Terasvirta, 1988. Testing linearity against smooth transition autoregressive models, *Biometrika* 75, 491–9

5.3. Хетероскедастичност и тестирање нелинеарности

До сада је претпоставка била да се ради са хомоскедастичним временским серијама, односно да се сви шумови могу описати $\mathcal{N}(0, \sigma^2)$ расподелом, при чему је дисперзија константа током времена. Када се дисперзија мења током времена, говори се о **хетероскедастичности**. Међутим, занемарена хетероскедастичност може довести до лажног одбацавања нулте хипотезе о линеарности. Зато су развијени тестови који могу бити коришћени и у присуству хетероскедастичности, и то без потребе одређивања тачног облика хетероскедастичности, који је често непознат, али у овом раду неће бити више речи о томе. [8]

6. Дијагностичка провера

Када је реч о **евалуацији модела** прво што се испитује јесу резидуали, као што је то уобичајено у моделовању линеарних временских серија. За резидуале ε_t се често претпоставља да су међу собом независни и једнакорасподељени из $\mathcal{N}(0, \sigma^2)$ расподеле. Тестирање нормалности резидуала је веома битно јер у случају да тај услов није испуњен, добијени резултати често нису поуздани, а самим тим ни закључци донети на основу изабраног модела.

Испоставља се, ипак, да нису све тест статистике које су развијене у контексту линеарних *AR* модела применљиве и на резидуале нелинеарних модела. Међутим, за тестирање нормалне расподељености резидуала, како линеарних, тако и нелинеарних временских серија, може се користити статистика коју су добили Бера и Жарк⁵³, а која је дата са:

$$JB = n \left(\frac{\widehat{SK}_{\varepsilon}^2}{6} + \frac{(\widehat{K}_{\varepsilon} - 3)^2}{24} \right), \quad (6.1)$$

при чему је $\widehat{K}_{\varepsilon}$ ⁵⁴ **мера спљоштености**, $\widehat{SK}_{\varepsilon}$ ⁵⁵ **мера асиметрије** на реализованом узорку, а ε наглашава да се ради о серији оцењених резидуала. Постоје различити начини дефинисања поменутих мера, али ће бити представљен само један од њих.

Нека је оцењени моменат резидуала дат као:

$$\widehat{m}_j = \frac{1}{n} \sum_{i=1}^n \varepsilon_t^j. \quad (6.2)$$

Тада се **мера спљоштености** може дефинисати као:

$$\widehat{K}_{\varepsilon} = \frac{\widehat{m}_4}{\widehat{m}_2^2}, \quad (6.3)$$

а **мера асиметрије** као:

$$\widehat{SK}_{\varepsilon} = \frac{\widehat{m}_3}{\sqrt{\widehat{m}_2^3}}. \quad (6.4)$$

Мера спљоштености је мера која описује облик репова расподеле у односу на облик целе расподеле. Расподела може бити бесконачна у пику, тј. у максимуму, са малом вредношћу мере спљоштености, али и савршено равна са бесконачном мером спљоштености. Вредности мере спљоштености се упоређују са вредношћу мере спљоштености стандардне нормалне расподеле код које је она једнака три. Што се тиче мере асиметрије, све симетричне расподеле, укључујући и нормалну расподелу, имају меру асиметрије једнаку нули.

⁵³ Bera and Jarque

⁵⁴ Kurtosis

⁵⁵ Skewness

Под нултом хипотезом о нормалности резидуала, стандардизована мера спљоштености

$\sqrt{\frac{n}{24}}(\hat{K}_{\varepsilon} - 3)$ и стандардизована мера асиметрије $\sqrt{\frac{n}{6}}S\hat{K}_{\varepsilon}$ јесу независне и имају асимптотски нормалну расподелу $\mathcal{N}(0, 1)$, па тест статистика (6.1) има асимптотски χ^2 расподелу са два степена слободе. Одбијање нулте хипотезе о нормалности може указивати на то да постоје аутлајери, да није испуњена претпоставка о хомоскедастичкости резидуала, и/или да временска серија треба бити описана неким другим моделом. [1], [13], [16]

6.1. Дијагностички тестови за *SETAR* и *STAR* моделе

Осим тестирања резидуала, постоји више метода за евалуацију модела са променом режима, али ће овом раду пажња бити посвећена **тестирању преостале нелинеарности**.

Питање које се поставља када се користе нелинеарни модели временских серија јесте да ли предложени модел обухвата све нелинеарне карактеристике временске серије која се посматра. Један од начина да се ово испита јесте да се примени тест преостале нелинеарности оцењеног модела. Природни приступ је навођење присуства додатног режима као алтернативне хипотезе преостале нелинеарности. На пример, можда се чини потребним да се тестира нулта хипотеза о адекватном моделу са два режима против алтернативне хипотезе да је неопходан и трећи режим. Испоставља се да је само за *STAR* модел доступан *LM*-тест који омогућава тестирање ове хипотезе без потребе за оценом сложенијег модела. За *SETAR* модел, за тестирање преостале нелинеарности неопходно је оценити и модел са више режима. Заправо, ово је аналогно тестирању нелинеарности поменуто у поглављу 5..

За *SETAR* моделе се у суштини може применити методологија описана у поглављу 5.1. на сваком од подузорака дефинисаних оцењеним прагом \hat{c} . Тачније, може се тестирати линеарност против алтернативе дворежимног *SETAR* модела на подузorcима за које је $y_{t-1} \leq \hat{c}$ и $y_{t-1} > \hat{c}$ користећи тест статистику (5.3). Као што је напоменуто, за рачунање тест статистике потребно је оценити дворежимни модел под алтернативном хипотезом. Зато се чини да тамо где статистика указује на присуство додатног режима, оцено трорежимног модела су лако доступне комбиновањем оригиналних резултата оцењеног дворежимног *SETAR* модела са онима за дворежимни модел на подузорку за који је линеарност одбијена. Међутим, у случају да је трорежимни модел заиста бољи избор од дворежимног модела, може се показати да иако је оцена другог прага, \hat{c}_2 , конзистентна, оцена првог прага $\hat{c}_1 \equiv \hat{c}$, сада не мора бити. Да би се добила конзистентна оцена и првог прага, неопходно је извести тзв. „корак поделе“ у ком је дворежимни *SETAR* модел оцењен на подузорку дефинисаном са $y_{t-1} \leq \hat{c}_2$, ако је $\hat{c}_1 < \hat{c}_2$ и на подузорку дефинисаном са $y_{t-1} > \hat{c}_2$, ако је $\hat{c}_1 > \hat{c}_2$.

За **тестирање дворежимног *STAR* модела против алтернативе о трорежимном *STAR* моделу** развијена је *LM* тест статистика⁵⁶ која се може записати као:

$$y_t = \varphi_1' x_t + (\varphi_2 - \varphi_1)' x_t G_1(y_{t-1}; \gamma_1, c_1) + (\varphi_3 - \varphi_2)' x_t G_2(y_{t-1}; \gamma_2, c_2) + \varepsilon_t, \quad (6.5)$$

⁵⁶ Eitrheim and Terasvirta, *Testing for the Adequacy of Smooth Transition Autoregressive Models*, Journal of Econometrics 1996.

где је $x_t = (1, \tilde{x}_t)'$, $\tilde{x}_t = (y_{t-1}, \dots, y_{t-p})'$, G_1 и G_2 су дате са (3.5) и претпоставља се да је $c_1 < c_2$ без губитка општости. Нулта хипотеза дворежимног модела може бити дата са $H_0 : \gamma_2 = 0$. Овај начин тестирања има сличне проблеме као и случају испитивања нулте хипотезе о линеарности потив алтернативне о дворежимном *STAR* моделу приказаном у поглављу 5.2.. Решење је исто као и тамо. Функција преласка $G_2(y_{t-1}; \gamma_2, c_2)$ замени се Тејлоровом апроксимацијом трећег реда око тачке $\gamma_2 = 0$. У случају апроксимације трећег реда, резултујући модел дат је са:

$$y_t = \beta_0' \tilde{x}_t + (\varphi_2 - \varphi_1)' x_t G_1(y_{t-1}; \gamma_1, c_1) + \beta_1' \tilde{x}_t y_{t-1} + \beta_2' \tilde{x}_t y_{t-1}^2 + \beta_3' \tilde{x}_t y_{t-1}^3 + \eta_t, \quad (6.6)$$

где су $\beta_j, j = 0, 1, 2, 3$, функције параметара $\varphi_1, \varphi_2, \gamma_2$ и c_2 . Нулта хипотеза $H_0 : \gamma_2 = 0$ у (6.5) прелази у $H_0' : \beta_1 = \beta_2 = \beta_3 = 0$ у (6.6). Тест статистика може бити рачуната као nR^2 из резидуала помоћне регресије добијених из оцењеног модела под нултом хипотезом. Резултујућа тест статистика има асимптотску χ^2 расподелу.

У горе наведеном, претпостављено је да је додатни режим одређен истом променљивом (у овом случају то је y_{t-1}) као и оригинална два режима. Као што је дискутовано раније, може се разматрати и ситуација где су режими одређени помоћу неколико променљивих, на пример y_{t-1} и y_{t-2} . За *STAR* модел, нулта хипотеза дворежимног модела може бити тестирана против алтернативне хипотезе о четворорежимном моделу (3.20) тестирањем $H_0 : \gamma_2 = 0$. *LM* тест статистика која се користи у овом случају слична је *LM* типу статистике за тестирање против дискутоване трорежимне алтернативе.⁵⁷ Почевши од модела (3.20), друга функција преласка $G_2(y_{t-1}; \gamma_2, c_2)$ замењена је Тејлоровом апроксимацијом трећег реда за добијање помоћне регресије:

$$\begin{aligned} y_t = & \beta_0' \tilde{x}_t + (\varphi_2 - \varphi_1)' x_t G_1(y_{t-1}; \gamma_1, c_1) \\ & + \beta_1' \tilde{x}_t y_{t-2} + \beta_2' \tilde{x}_t y_{t-2}^2 + \beta_3' \tilde{x}_t y_{t-2}^3 \\ & + \beta_4' x_t G_1(y_{t-1}; \gamma_1, c_1) y_{t-2} + \beta_5' x_t G_1(y_{t-1}; \gamma_1, c_1) y_{t-1}^2 \\ & + \beta_6' x_t G_1(y_{t-1}; \gamma_1, c_1) y_{t-1}^3 + \eta_t. \end{aligned} \quad (6.7)$$

Нулта хипотеза $H_0 : \gamma_2 = 0$ у (3.19) сада постаје $H_0' : \beta_j = 0, j = \overline{1,6}$, која може бити тестирана на потпуно исти начин као и раније. [1]

⁵⁷ D. van Dijk and P.H. Franses, *Modeling multiple regimes in the business cycle*, Macroeconomic Dynamics 3, Printed in the United States of America, 1999.

7. Предвиђања

Нелинеарни модели временских серија могу бити разматрани у различите сврхе. Некада је довољно добити адекватан опис динамичких образаца који постоје у посматраној временској серији. Међутим, врло често је додатни циљ употреба модела за предвиђање будућих вредности временске серије. У том контексту, предвиђање ван узорка, тачније на скупу за тестирање, сматра се начином евалуације оцењеног модела са променом режима. Конкретно, упоређивањем предвиђања нелинеарних модела са онима из референтног линеарног модела може се утврдити значај додате карактеристике нелинеарног модела.

Постоји неколико начина за добијање тачкастих и интервалих предвиђања нелинеарних модела.

7.1. Тачкаста предвиђања

Нека је временска серија y_t описана нелинеарним ауторегресионим моделом првог реда, тј. нека је:

$$y_t = F(y_{t-1}; \theta) + \varepsilon_t, \quad (7.1)$$

за неку нелинеарну функцију $F(y_{t-1}; \theta)$. Када се користи метод најмањих квадрата, оптимално тачкасто предвиђање будуће вредности временске серије дато је њеним условним очекивањем. Прецизније, оптимално **предвиђање h корака унапред** за вредност временске серије y_{t+h} у тренутку t , у ознаци $\hat{y}_{t+h|t}$, дато је са:

$$\hat{y}_{t+h|t} = E[y_{t+h} | \Omega_t], \quad (7.2)$$

где Ω_t означава историју временске серије до тренутка t укључујући и опсервацију у тренутку t . Користећи (7.1) и чињеницу да $E[\varepsilon_{t+1} | \Omega_t] = 0$, оптимално предвиђање један корак унапред добија се као:

$$\hat{y}_{t+1|t} = E[y_{t+1} | \Omega_t] = F(y_t; \theta), \quad (7.3)$$

што је еквивалентно предвиђању један корак унапред када је функција $F(y_{t-1}; \theta)$ линеарна. Када је период предвиђања већи од један, ствари се компликују. На пример, за оптимално предвиђање два корака унапред, на основу (7.2) и (7.1), добија се:

$$\hat{y}_{t+2|t} = E[y_{t+2} | \Omega_t] = E[F(y_{t+1}; \theta) | \Omega_t]. \quad (7.4)$$

где користимо чињеницу да се линеарни оператор очекивања E не може заменити нелинеарним оператором F , тј. важи $E[F(\cdot)] \neq F(E[\cdot])$. Дакле,

$$E[F(y_{t+1}; \theta) | \Omega_t] \neq F(E[y_{t+1} | \Omega_t]; \theta) = F(\hat{y}_{t+1|t}; \theta). \quad (7.5)$$

Ипак, веза између предвиђања један корак унапред и предвиђања два корака унапред дата је са:

$$\hat{y}_{t+2|t} = E[F(F(y_t; \theta) + \varepsilon_{t+1}; \theta) | \Omega_t] = E[F(\hat{y}_{t+1|t} + \varepsilon_{t+1}; \theta) | \Omega_t]. \quad (7.6)$$

Ово показује да рекурентна веза између предвиђања са различитим периодима, код нелинеарних модела не постоји. Током година, развијено је неколико метода за добијање што адекватнијег предвиђања више корака унапред. Неке од њих ће бити приказане.

Прво, може се покушати са добијањем условног очекивања (7.6) директним путем на следећи начин:

$$\hat{y}_{t+2|t}^{(c)} = \int_{-\infty}^{+\infty} F(\hat{y}_{t+1|t} + \varepsilon; \theta) f(\varepsilon) d\varepsilon, \quad (7.7)$$

где f означава густину серије резидуала ε_t . Ово предвиђање се назива **предвиђање затвореног облика**, па отуд и ознака c^{58} .

Другачији начин израчунавања интеграла добија се из (7.4) као:

$$\hat{y}_{t+2|t}^{(c)} = \int_{-\infty}^{+\infty} F(y_{t+1}; \theta) g(y_{t+1} | \Omega_t) dy_{t+1} = \int_{-\infty}^{+\infty} E[y_{t+2} | y_{t+1}] g(y_{t+1} | \Omega_t) dy_{t+1}, \quad (7.8)$$

где $g(y_{t+1} | \Omega_t)$ означава расподелу за y_{t+1} под условом Ω_t . Када се узме у обзир да је $y_{t+1} = F(y_t; \theta) + \varepsilon_{t+1}$, ова условна расподела је заправо једнака расподели $f(\cdot)$ шума ε_{t+1} са очекивањем једнаким $F(y_t; \theta)$, тј. важи да је $g(y_{t+1} | \Omega_t) = f(y_{t+1} - F(y_t; \theta))$. Како аналитички израз интеграла (7.7) или (7.8) у општем случају не постоји, потребно га је апроксимирати користећи нумеричке методе. Међутим, додатно компликује чињеница да расподела за ε_t никад није позната са сигурношћу, иако се уобичајено претпоставља да ε_t има нормалну расподелу.

Предвиђање затвореног облика је заморно за рачунање предвиђања више од два периода. Лако се може видети и зашто. Користећи релацију Колмогоров – Чепмена⁵⁹, расподела $g(y_{t+1} | \Omega_t)$ се може записати као:

$$g(y_{t+h} | \Omega_t) = \int_{-\infty}^{+\infty} g(y_{t+h} | y_{t+h-1}) g(y_{t+h-1} | \Omega_t) dy_{t+h-1}, \quad (7.9)$$

где је $g(y_{t+h} | y_{t+h-1})$ условна расподела за y_{t+h} при услову y_{t+h-1} . Узимањем условног очекивања са обе стране једнакости (7.9) добија се:

$$E[y_{t+h} | \Omega_t] = \int_{-\infty}^{+\infty} E[y_{t+h} | y_{t+h-1}] g(y_{t+h-1} | \Omega_t) dy_{t+h-1}, \quad (7.10)$$

што се може посматрати као генерализација израза (7.8). Да би се оценио овај интеграл, потребна је условна расподела $g(y_{t+h-1} | \Omega_t)$. У принципу, ова расподела се може добити рекурзивно из (7.9), што је рачунски захтевно, па је алтернатива претпоставити да грешка предвиђања $h - 1$ корак унапред $\varepsilon_{t+h-1|t} = y_{t+h-1} - \hat{y}_{t+h-1|t}$ има нормалну расподелу са очекивањем једнаким нула и дисперзијом σ_{h-1}^2 . Тада је $g(y_{t+h-1} | \Omega_t)$ нормално расподељена са очекивањем једнаким предвиђању $h - 1$ корак унапред, $\hat{y}_{t+h-1|t}$, и дисперзијом σ_{h-1}^2 . Ова тзв. **метода предвиђања нормалне грешке**⁶⁰, у ознаци **NFE**, развијена је за уопштене ауторегресионе нелинеарне моделе, (који су уско повезани са **STAR** моделом), као и за **SETAR** моделе.

За дворежимни **SETAR** модел (3.2), **NFE** за h корака унапред може се добити из рекурзије:

⁵⁸ closed

⁵⁹ Kolmogorov-Chapman

⁶⁰ normal forecast error

$$\begin{aligned}
\hat{y}_{t+h|t}^{(NFE)} = & \Phi(z_{t+h-1|t})(\varphi_{0,1} + \varphi_{1,1}\hat{y}_{t+h-1|t}) \\
& + \Phi(-z_{t+h-1|t})(\varphi_{0,2} + \varphi_{1,2}\hat{y}_{t+h-1|t}) \\
& + \Phi(z_{t+h-1|t})(\varphi_{1,2} - \varphi_{1,1})\sigma_{h-1},
\end{aligned} \tag{7.11}$$

где су $\Phi(\cdot)$ и $\phi(\cdot)$ стандардна нормална расподела и њена густина, респективно, а σ_{h-1}^2 дисперзија предвиђања $h - 1$ корак унапред грешке $\varepsilon_{t+h-1|t}$ и важи да је $z_{t+h-1|t} = (c - \hat{y}_{t+h-1|t})/\sigma_{h-1}$. Лако се може видети да је (7.11) у суштини пондерисани просек оптималних предвиђања оба режима, са тежинама које су једнаке вероватноћи да је модел у одређеном режиму у тренутку $t + h - 1$ под условом нормалности, с тим што постоји додатни фактор корекције.

Алтернативни приступ за рачунање предвиђања више корака унапред је коришћење Монте Карло методе или *bootstrap* методе за апроксимацију условног очекивања (7.4). **Предвиђање два корака унапред добијено методом Монте Карла** дато је са:

$$\hat{y}_{t+2|t}^{(MK)} = \frac{1}{k} \sum_{i=1}^k F(\hat{y}_{t+1|t} + \varepsilon_i; \theta), \tag{7.12}$$

где је k неки велики број и ε_i су извучене из претпостављене расподеле за ε_{t+1} . **Предвиђање добијено *bootstrap* методом** је веома слично. Једина разлика је у томе што се користе резидуали оцењеног модела, $\hat{\varepsilon}_t, t = \overline{1, n}$, тј. важи:

$$\hat{y}_{t+2|t}^{(B)} = \frac{1}{k} \sum_{i=1}^k F(\hat{y}_{t+1|t} + \hat{\varepsilon}_i; \theta). \tag{7.13}$$

Предност *bootstrap* методе у односу на Монте Карло методу је што нису потребне претпоставке о расподели за ε_{t+1} . [1]

Овај алтернативни приступ је имплементиран и у програмском језику R , при чему се претпоставља да је испуњен услов о хомоскедастичности резидуала.

7.2. Интервална предвиђања

Осим тачкастих предвиђања, од интереса могу бити и **интервали поверења добијених тачкастих предвиђања**. За предвиђања добијена из линеарних модела, уобичајена област поузданости предвиђања јесте интервал симетричан у односу на тачкасто предвиђање. То је засновано на чињеници да је условна расподела $g(y_{t+h} | \Omega_t)$ линеарне временске серије нормална са очекивањем $\hat{y}_{t+h|t}$. За нелинеарне моделе временских серија то није случај. Заправо, условна расподела може бити асиметрична и да чак има више различитих модела, па симетрични интервал поверења не мора бити најадекватнија област поузданости предвиђања. Показано је да постоје три методе за конструкцију $100(1 - \alpha)\%$ области предвиђања:

1. Интервал симетричан око средње вредности

$$S_\alpha = (\hat{y}_{t+h|t} - \omega, \hat{y}_{t+h|t} + \omega), \text{ где је } \omega \text{ такво да важи } P(y_{t+h} \in S_\alpha | \Omega_t) = 1 - \alpha.$$

2. Интервал између $\frac{\alpha}{2}$ и $1 - \frac{\alpha}{2}$ квантила расподеле предвиђања, у ознаци $q_{\frac{\alpha}{2}}$ и $q_{1-\frac{\alpha}{2}}$,
респективно $Q_{\alpha} = (q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}})$
3. Регион највеће густине (The highest-density region - HDR)

$$HDR_{\alpha} = \{ y_{t+h} \mid g(y_{t+h} \mid \Omega_t) \geq q_{\alpha} \}, \quad (7.15)$$
где је q_{α} такво да је $P(y_{t+h} \in HDR_{\alpha} \mid \Omega_t) = 1 - \alpha$.

За симетричну и унимодалну расподелу, сва три региона су једнака. Код асиметричне расподеле то не важи и су они различити међу собом. Показује се да је HDR_{α} најприроднији избор, јер је пре свега најмањи од свих могућих $100(1 - \alpha)\%$ интервала предвиђања и свака тачка унутар њега има условну густину $g(y_{t+h} \mid \Omega_t)$ велику бар онолико колико има и свака тачка изван региона. Региони HDR_{α} се лако могу израчунати коришћењем Монте Карло методе или *bootstrap* методе за рачунање тачкастих предвиђања $\hat{y}_{t+h|t}$.⁶¹

У књизи [1] је на симулираном *SETAR*(1) моделу (3.1) са параметрима истим као код модела (2), али смањеном дисперзијом резидуала, (што не мења скелет модела), показано да једино HDR_{α} за предвиђања два корака унапред не мора бити континуалан и да може бити асиметричан.

⁶¹ Hyndman, R.J., *Highest-density forecast regions for nonlinear and nonnormal time series*, Journal of Forecasting, 1995.

8. Евалуација предвиђања

До сада се модел бирао на основу тога какве резултате даје на скупу за обучавање. Тачније, као најбољи модел проглашен је онај који има најмању вредност AIC -а, или неког другог информационог критеријума. Ови критеријуми избора модела су добре мере које пореде резидуале на скупу за обучавање, мерењем дисперзије резидуала у односу на број оцењених параметара. Међутим, чињеница да одређени модел добро описује особине посматране временске серије на скупу за обучавање, не гарантује да тај модел даје добре резултате и на предвиђањима на скупу за тестирање. Дакле, други приступ је процена тачности предвиђања предложених модела.

8.1. Метрике за евалуацију предвиђања

У овом делу ће се навести неке од стандардних метрика које се користе за евалуацију модела. Због једноставности, грешка предвиђања у тренутку t означава се са $\xi_t = y_t - \hat{y}_t$, где је y_t реализована вредност посматране временске серије у тренутку t , а \hat{y}_t предвиђена вредност те временске серије у тренутку t , без обзира на то на који начин је предвиђање извршено.

Мере које се најчешће користе су **средње квадратна грешка** MSE^{62} , **корен средње квадратне грешке** $RMSE^{63}$, **средња апсолутна грешка** MAE^{64} , **средња релативна грешка изражена у процентима** $MAPE^{65}$. За сваку од ових мера мања вредност указује на већу прецизност предвиђања, а саме мере се дефинишу на следећи начин:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (8.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (8.2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (8.3)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (8.4)$$

Употреба апсолутних или квадратних вредности спречава да се негативне и позитивне грешке међусобно поништавају. Значај MSE је очигледан, али је $RMSE$ кориснији, пошто је изражена на истој скали на којој је и посматрана временска серија, што олакшава интерпретацију добијених вредности. Међутим, MSE , $RMSE$ и MAE зависе од скале, па се могу користити само за поређење различитих предвиђања у оквиру исте временске серије.

Због јаког утицаја аутлајера на ове грешке, користи се и средња релативна грешка изражена у процентима, $MAPE$. Она је независна од скале, па се често користи за поређење предвиђања

⁶² Mean Square Error

⁶³ Root Mean Square Error

⁶⁴ Mean Absolute Error

⁶⁵ Mean Absolute Percentage Error

између различитих временских серија. Међутим, мана мере засноване на процентним грешкама је то што је она бесконачна или недефинисана ако у временској серији постоје нулте вредности. Она има још један недостатак: теже „кажњава“ позитивне грешке него негативне. Зато се уводи симетрична средња релативна грешка изражена у процентима $sMAPE$ ⁶⁶, чија формула није сада од интереса.

Још једна мера која се предлаже је скалирање грешака методом наивног предвиђања⁶⁷. Користећи наивни метод, генеришу се предвиђања за један корак унапред за сваку опсервацију скупа за обучавање. Прецизније, дефинисана скалирана грешка је дата је као:

$$q_t = \frac{\xi_t}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}. \quad (8.5)$$

Грешка q_t је независна од скале на којој су подаци. Овако скалирана грешка је мања од један ако је предвиђена грешка мања од упросечене грешке добијене наивном методом, и супротно, она је већа од један ако је предвиђена грешка већа од грешке добијене упросечавањем. На тај начин се добија **скалирана апсолутна средња грешка $MASE$** ⁶⁸ дефинисана као:

$$MASE = \frac{1}{n} \sum_{i=1}^n |q_i|. \quad (8.6)$$

Једина околност под којом би $MASE$ била бесконачна или недефинисана је када су све опсервације скупа за обучавање једнаке међу собом, што се ретко дешава. Главна предност $MASE$ у односу на MAE је та што је она широко применљива. Коришћењем мере MAE претпоставља да је серија стабилна током времена, што не мора увек бити испуњено. Насупрот томе, $MASE$ је погодна чак и када серија има тренд на пример. $MASE$ се може користити за упоређивање различитих метода предвиђања на једној серији и, с обзиром на то да је независна од скале, за упоређивање тачности предвиђања између серија. Упоређивањем различитих вредности за $MASE$ добијених на основу више временских серија, може се идентификовати које је серије најтеже предвидети. Типичне вредности за $MASE$ код предвиђања један корак унапред мање су од један, јер је обично могуће добити предвиђања тачније од наивне методе. За предвиђања више корака често су веће од један, јер је теже предвиђати када се хоризонт предвиђања повећава. Ипак, сматра се да је ова мера једна од најбољих мера за евалуацију предвиђања. [2] , [9]

8.2. DM статистика

Наведене метрике упоређују грешке предвиђања добијених из различитих модела. Међутим, занемарује се врло важно питање, а то је да ли постоји квалитативна разлика између предвиђања оцењених модела, тј. да ли су разлике у предвиђањима статистички значајне. Зато се јавља потреба за тестирањем ових разлика. Оригинални тестови који се уводе у ову сврху намећу

⁶⁶ Symmetric Mean Absolute Percentage Error

⁶⁷ naïve forecast method

⁶⁸ Mean Absolute Scaled Error

одређене услове **функцији губитка**⁶⁹, која мери одступање предвиђених и реализованих вредности посматране временске серије, и грешкама предвиђања, које је често веома тешко постићи. Функција губитка мора бити квадратна, а грешке предвиђања морају бити Гаусове, са очекивањем нула. Не сме постојати корелација између грешке предвиђања у тренутку t и грешке предвиђања у претходним тренуцима. Такође, не сме постојати корелација између конкурентних грешака предвиђања у тренутку t . Да би се превазишла ова ограничења, предложени су тестови који тестирају нулту хипотезу да нема разлике у тачности два конкурентна предвиђања, при чему се омогућава одређена флексибилност функције губитка, као и грешака предвиђања. [10]

Нека су $\{\hat{y}_{it}\}_{t=1}^h$ и $\{\hat{y}_{jt}\}_{t=1}^h$ два различита предвиђања за реализоване вредности посматране временске серије $\{y_t\}_{t=1}^h$ и одговарајуће грешке предвиђања $\{\varepsilon_{it}\}_{t=1}^h$ и $\{\varepsilon_{jt}\}_{t=1}^h$, где је h број предвиђених вредности. Идеја је проценити функцију очекиваног губитка, која се означава са g . Циљ је тестирати нулту хипотезу

$$H_0 : E[g(\varepsilon_{it})] = E[g(\varepsilon_{jt})], \quad (8.7)$$

или, ако се са d_t означи разлика функција губитака различитих предвиђања, $d_t = g(\varepsilon_{it}) - g(\varepsilon_{jt})$, тада се нулта хипотеза може записати као:

$$H_0 : E[d_t] = 0. \quad (8.8)$$

Другим речима, тестира се нулта хипотеза да је очекивана разлика функција губитака једнака нула. Да би се тестирала ова хипотеза, претпоставља се да d_t , за $t = \overline{1, h}$, има асимптотску расподелу:

$$\sqrt{h} (\bar{d} - \mu) \rightarrow \mathcal{N}(0, 2\pi f_d(0)), \quad (8.9)$$

где је $\bar{d} = \frac{1}{h} \sum_{t=1}^h d_t$, а $f_d(0)$ спектрална густина⁷⁰ на фреквенцији нула. Када је скуп за обучавање довољно велики, \bar{d} има расподелу $\mathcal{N}(\mu, 2\pi f_d(0))$, па се за тестирање нулте хипотезе о једнакости грешака предвиђања користи **DM**⁷¹ **тест статистика** дефинисана као:

$$DM = \frac{\bar{d}}{\sqrt{\text{var}(\bar{d})}} = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{h}}} \sim \mathcal{N}(0,1). \quad (8.10)$$

Лако се може видети да је гранична дисперзија за \bar{d} , $\text{var}(\bar{d})$, дата на приказан начин. Важи да је:

$$\text{var}(\bar{d}) = \frac{1}{h^2} \sum_{s=1}^h \sum_{t=1}^h \gamma_d(t-s), \quad (8.11)$$

где је са $\gamma_d(t-s)$ означена одговарајућа коваријација. Нека је $\tau = t-s$. Тада се претходна сума може једноставније записати као:

⁶⁹ function loss

⁷⁰ https://en.wikipedia.org/wiki/Spectral_density

⁷¹ Diebold and Mariano

$$\text{var}(\bar{d}) = \frac{1}{h^2} \sum_{\tau=-(h-1)}^{h-1} (h - |\tau|) \gamma_d(\tau) = \frac{1}{h} \sum_{\tau=-(h-1)}^{h-1} \left(1 - \frac{|\tau|}{h}\right) \gamma_d(\tau). \quad (8.12)$$

Ако се позове на то да је спектрална функција густине за d дефинисана као:

$$f_d(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau) e^{-i\omega\tau}, \quad (8.13)$$

онда за $\omega = 0$ важи да је:

$$f_d(0) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau). \quad (8.14)$$

Користећи (8.14), за $h \rightarrow \infty$ у (8.12) директно се добија да је:

$$\text{var}(\bar{d}) = \frac{1}{h} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau) = \frac{2\pi f_d(0)}{h}. \quad (8.15)$$

Међутим, симулирани експерименти показују да нормална расподела може бити лоша апроксимација DM тест статистике под нултом хипотезом на мањим узорцима. Зато се предлага побољшана DM тест статистика, у ознаци **HLN – DM**⁷²:

$$\text{HLN} - DM = \sqrt{\frac{n + 1 - 2h + h(h-1)/n}{n}} DM, \quad (8.16)$$

која има Студентову t -расподелу са $(n-1)$ степени слободе, иако је Студентова t -расподела оправдана само у случају да су d_t независне и нормално расподељене. Тест статистика (8.16) је имплементирана у програмском језику *R*. [17]

8.3. Узроци лоших предвиђања

Приказане су различите метрике за евалуацију предвиђања, међутим често се дешава да, иако нелинеарни модели описују одређене карактеристике временске серије боље од линеарних, резултати које дају предвиђања линеарних модела не морају бити лошија од резултата предвиђања нелинеарних модела. Биће наведени и неки од разлога зашто је то тако.

На пример, нелинеарност може бити лажна, у смислу да друге карактеристике временске серије, као што су хетероскедастичност, структурни ломови или аутлајери, указују на присуство

⁷² Harvey, Leybourne, and Newbold 1997.

нелинеарности. Иако се може успешно оценити нелинеарни модел за такву серију, мало је вероватно да ће резултати добијени приликом предвиђања бити добри.

Још један узрок лоших резултата предвиђања нелинеарних модела је да се нелинеарност не показује током периода предвиђања. У случају модела са променом режима може бити да се само један режим догодио током целог периода предвиђања. Потенцијално решење проблема непостојања нелинеарности током периода предвиђања, јесте извршити симулирани експеримент у коме се оцењени модел са променом режима користи за генерисање вештачких временских серија, а затим се изврше предвиђања ван узорка на свакој од тих серија. У овако контролисаном окружењу, сигурно је да су у сваком од режима укључена предвиђања.

Избор критеријума евалуације је веома битан. Иако су традиционални критеријуми применљиви на предвиђања нелинеарних модела, они нису увек најадекватнији. Може се добити да је нека мера линеарног модела мања од мере модела са променом режима, ако постоји могућност да се погрешно предвиди у којем ће режиму бити временска серија. Као што је приметио Тонг⁷³ „Колико добро можемо предвидети, зависи од тога где смо.“⁷⁴ У случају модела са променом режима, може се десити да је моћ предвиђања временске серије веома различита у различитим режимима. Стога би се могло евалуирати предвиђање за сваки режим посебно да би се испитало да ли је нелинеарни модел користан за добијање предвиђања у одређеном режиму.

[1]

У поглављу 11. ће на примерима бити примењене стандардне метрике и $HLN - DM$ статистика да би се донео што адекватнији закључак о квалитету изабраних модела.

⁷³ Tong, H. "A personal overview of non-linear time series analysis from a chaos perspective" Scandinavian Journal of Statistics, 22, 399-445, (1995)

⁷⁴ Tong H : "How well we can forecast depends on where we are."

9. Вишедимензиони модели са променом режима

До сада је пажња била посвећена једнодимензионим моделима временских серија у којима је временска серија y_t описана у терминима својих претходних вредности, $y_{t-1}, y_{t-2}, \dots, y_{t-p}$. Некада је корисно моделирати неколико временских серија заједно да би се искористиле могуће везе које могу постојати између њих. Показано је да се, на пример, повраћај и волатилност различитих берзанских индекса крећу заједно током времена. Такође, вишедимензиони модели се могу користити за описивање неколико карактеристика одређене финансијске подлоге.⁷⁵ У овом раду се неће испитивати вишедимензиони модели са променом режима, већ ће се као пример приказати само вишедимензиони *SETAR* модел који се добија уопштавањем једнодимензионог модела. Посматра се s -димензиона временска серија $\mathbf{Y}_t = (y_{1t}, y_{2t}, \dots, y_{st})'$. **S -димензиони *SETAR* модел са k режима**, у ознаци $SETAR(d; p_1, p_2, \dots, p_k)$ дат је као:

$$\mathbf{Y}_t = \begin{cases} \mathbf{c}_0^{(1)} + \sum_{j=1}^{p_1} \Phi_j^{(1)} \mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t^{(1)}, & z_{t-d} \leq r_1 \\ \mathbf{c}_0^{(2)} + \sum_{j=1}^{p_2} \Phi_j^{(2)} \mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t^{(2)}, & r_1 < z_{t-d} \leq r_2 \\ \dots & \dots \\ \mathbf{c}_0^{(k)} + \sum_{j=1}^{p_k} \Phi_j^{(k)} \mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t^{(k)}, & r_{k-1} < z_{t-d} \end{cases} \quad (9.1)$$

где су $\mathbf{c}_0^{(i)}$ ($s \times 1$)-димензиони вектори слободних чланова, $\Phi_j^{(i)}$ ($s \times s$)- димензионе матрице параметара модела за $i = \overline{1, k}$. За шумове $\boldsymbol{\varepsilon}_t^{(i)}$ важи да је $\boldsymbol{\varepsilon}_t^{(i)} = \Sigma_i^{1/2} \mathbf{a}_t$ где $\Sigma_i^{1/2}$ означава симетричне позитивно дефинитне матрице, а $\{\mathbf{a}_t\}$ низ некорелисаних случајних вектора са очекивањем $\mathbf{0}$ и матрицом коваријације \mathbf{I} која је ($s \times s$)- димензиона и идентичка. За променљиву прага z_{t-d} претпоставља се да је стационарна и да зависи од претходних вредности посматране временске серије \mathbf{Y}_{t-d} . На пример, може се ставити да је $z_{t-d} = \omega' \mathbf{Y}_{t-d}$, где је ω ($s \times 1$)-димензиони вектор. Када је $\omega = (1, 0, \dots, 0)'$ променљива прага z_{t-d} је једноставно $z_{t-d} = y_{1,t-d}$. Када је $\omega = (\frac{1}{s}, \frac{1}{s}, \dots, \frac{1}{s})'$ променљива прага z_{t-d} је просек свих елемената временске серије \mathbf{Y}_{t-d} . [12]

⁷⁵ Financial asset, <https://www.investopedia.com/terms/f/financialasset.asp>

10. Режими одређени помоћу неопажене променљиве

Друга класа модела са променом режима претпоставља да режим који се догоди у тренутку t не може бити опажен. Тачније, он је одређен неопаженим процесом који ће бити означен са s_t . Када постоје само два режима, за s_t се једноставно може претпоставити да узима вредности 1 и 2, па је модел са $AR(1)$ моделом у оба режима дат као:

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_{t-1} + \varepsilon_t, & \text{ако је } s_t = 1, \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + \varepsilon_t, & \text{ако је } s_t = 2, \end{cases} \quad (10.1)$$

или, користећи једноставнији запис,

$$y_t = \varphi_{0,s_t} + \varphi_{1,s_t} y_{t-1} + \varepsilon_t. \quad (10.2)$$

Да би се употпунио модел, својства процеса s_t морају бити прецизирана. Најпопуларнији модел ове класе, јесте **Марковљев модел пребацивања**⁷⁶, у ознаци **MSW**, у ком се претпоставља да је s_t Марковљев процес првог реда.⁷⁷ То подразумева да режим у тренутку t зависи само од режима у тренутку $t - 1$, односно да он не зависи од историје посматране временске серије. Дакле, модел је употпуњен дефинисањем **вероватноћа преласка** из једног стања у друго на следећи начин:

$$\begin{aligned} P(s_t = 1 | s_{t-1} = 1) &= p_{11} \\ P(s_t = 2 | s_{t-1} = 1) &= p_{12} \\ P(s_t = 1 | s_{t-1} = 2) &= p_{21} \\ P(s_t = 2 | s_{t-1} = 2) &= p_{22}. \end{aligned} \quad (10.3)$$

Као што се може закључити, p_{ij} представља вероватноћу са којом Марковљев процес прелази из стања i у тренутку $t - 1$ у стање j у тренутку t , или, другим речима, вероватноћу да је режим i у тренутку $t - 1$ праћен режимом j у тренутку t . Да би p_{ij} биле одговарајуће вероватноће, оне морају бити ненегативне и мора важити да је $p_{11} + p_{12} = 1$ и $p_{21} + p_{22} = 1$. Од интереса су и безусловне вероватноће са којима је процес у једном од режима, тј. $P(s_t = i)$ за $i = 1, 2$. Користећи теорију ергодичности⁷⁸ Марковљевог процеса, директно се показује да су за **MSW** модел са два стања, безусловне вероватноће дате са:

$$\begin{aligned} P(s_t = 1) &= \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \\ P(s_t = 2) &= \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \end{aligned} \quad (10.4)$$

⁷⁶ Markov switching model

⁷⁷ Hamilton, J.D., 1989. *A new approach to the economic analysis of nonstationary time series subject to changes in regime*, *Econometrica* 57, 357–84

⁷⁸ ergodicity

Ако се дозволи да неопажени Марковљев процес s_t узме неку од $m > 2$ различитих вредности, добија се *MSW* модел са m режима, при чему свака од тих вредности одређује један режим. Тачније, модел постаје:

$$y_t = \varphi_{0,j} + \varphi_{1,j} y_{t-1} + \varepsilon_t, \quad \text{ако је } s_t = j, \quad (10.5)$$

за $j = \overline{1, m}$, са вероватноћама преласка

$$p_{ij} = P(s_t = j | s_{t-1} = i), \quad i, j = \overline{1, m}, \quad (10.6)$$

које задовољавају услов $p_{ij} \geq 0$ за $i, j = \overline{1, m}$ и $\sum_{j=1}^m p_{ij} = 1$ за $\forall i = \overline{1, m}$.

Параметри *MSW* модела могу бити оцењени користећи метод максималне веродостојности. Међутим, због чињенице да Марков процес s_t није опажен, проблем оцењивања је није стандардан. Циљ процедуре оцењивања није само добијање оцена параметара у ауторегресионим моделима у различитим режимима и вероватноће прелаза из једног режима у други, већ и добијање оцена стања која се догађају у свакој тачки узорка, или, прецизније, вероватноће са којима се јавља свако стање у сваком тренутку времена. Итерацијом кроз развијену процедуру могу се добити оценое жељених параметара, при чему се показује да свака итерација ове процедуре повећава вредност функције веродостојности, што гарантује да су крајње оценое максимално вероватне.

Занимљива особина *MSW* модела је релативна лакоћа са којом се може добити аналитички израз предвиђања више корака унапред, јер се предвиђање будуће вредности временске серије h корака унапред, y_{t+h} , може разложити на предвиђање за y_{t+h} под условом режима који ће се догодити у тренутку $t + h$, s_{t+h} , и предвиђања вероватноћа са којима ће се сваки од ових режима десити у тренутку $t + h$. [1]

Више о моделима ове класе може се наћи и у књизи [19].

11. Примери

Резултати који ће бити приказани добијени су коришћењем *Rstudio*-а, интегрисаног развојног окружења програмског језика *R* и његове верзије 3.6.2 (2019-12-12).

11.1. Пример 1.

База која се посматра може се наћи на приказаном линку [14]. Она садржи вредности девизног курса холандског цеха према америчком долару. Узорак се односи на дневне податке у периоду од 2. јануара 1980. године до 31. децембра 1997. године. Често се уместо дневних података посматрају недељни, па ће се у ту сврху користити свака седма опсервација почев од прве.

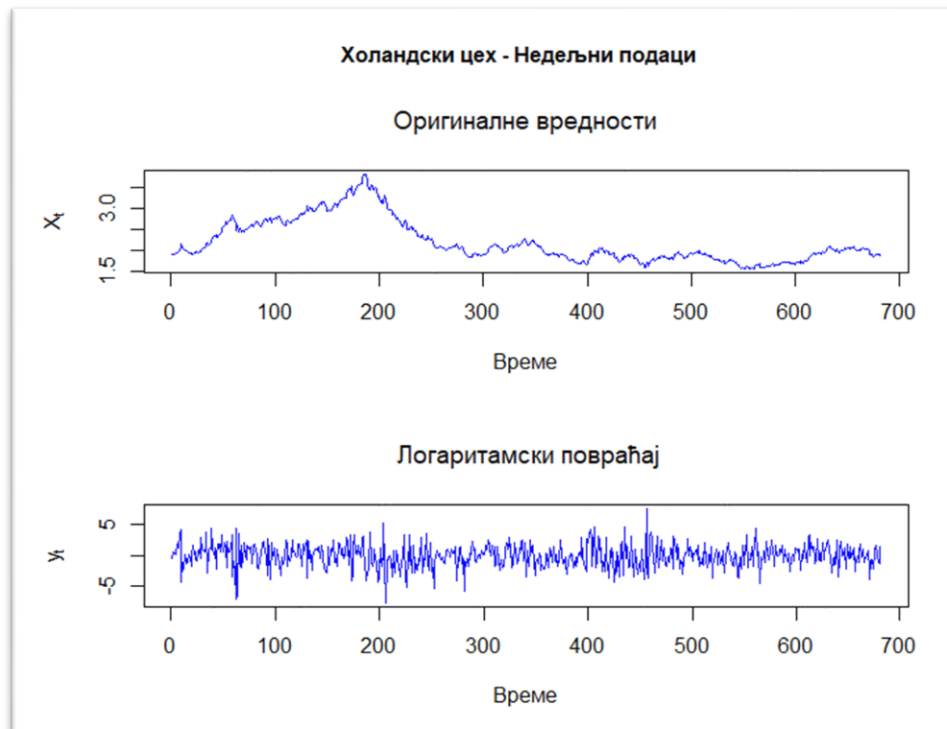
Најпре ће бити представљене неке статистике посматране временске серије. На слици 11.1 приказани су дневни подаци оригиналне временске серије, а на слици 11.2 њени недељни подаци. У горњем делу слика приказана је оригинална временска серија X_t , док је у доњем делу приказан њен логаритамски повраћај дат у процентима који се означава са y_t . Он се у финансијама често посматра и дат је као

$$y_t = 100 \cdot (\ln X_t - \ln X_{t-1}), \quad (11.1)$$

при чему је \ln ознака за природни логаритам, логаритам са основом e .



Слика 11.1. Вредности оригиналне серије X_t и логаритамски повраћај y_t холанског цеха према америчком долару- дневни подаци



Слика 11.2. Вредности оригиналне серије X_t и логаритамски повраћај y_t холанског цеха према америчком долару- недељни подаци

Позивом функције *summary*, *kurtosis*, *skewness* и *var* добијају се следеће вредности за дневни и недељни повраћај, приказане у табели 11.1.

Табела 1.11 - Дневни повраћај						
<i>min</i>	<i>Median</i>	<i>mean</i>	<i>max</i>	<i>Kurtosis</i>	<i>skewness</i>	<i>Var</i>
-3.9846	0.0117	-0.0003	3.1883	4.9719	-0.1429	0.4644
Недељни повраћај						
<i>min</i>	<i>Median</i>	<i>mean</i>	<i>max</i>	<i>Kurtosis</i>	<i>skewness</i>	<i>Var</i>
-7.7629	0.0241	-0.0012	7.7031	4.0429	-0.1902	3.3886

Као што смо је наведено, мера спљоштености (*kurtosis*) је мера која описује облик репова расподеле у односу на облик целе расподеле. Познато је да је мере спљоштености стандардне нормалне расподеле једнака три. Из табеле 11.1 се види да је мера спљоштености за дневне повраћаје једнака 4.9719, а за недељне 4.0429. У оба случаја она је већа од три, што указује на то да су репови ових временских серија дебљи од репова нормалне расподеле. Другим речима, велике вредности временске серије, тачније велике апсолутне вредности приноса, дешавају се много чешће него код нормално расподељене случајне величине.

Мера асиметрије (*skewness*) стандардне нормалне расподеле једнака је нули. Из табеле 11.1 види се да је за дневне повраћаје она једнака -0.1429, а за недељне -0.1902, што значи да је леви реп

расподеле дебљи од десног репа, односно да се велики негативни приноси јављају чешће него позитивни.

Вредности мере спљоштености и мере асиметрије се пореде са вредностима истих статистика стандардне нормалне расподеле зато што је устаљена претпоставка да логаритамска вредност финансијске подлоге⁷⁹ може бити описана случајним лутањем. Другим речима, када се користе линеарни модели, претпоставља се да важи

$$E[\ln X_{t+1} | \Omega_t] = \ln X_t + \mu, \quad (11.2)$$

односно да очекивана вредност логаритма сутрашње вредности девизног курса, под условом да су дате све релевантне информације до данас, укључујући и информације које се односе на данашњи дан, у ознаци Ω_t , треба да буде једнака логаритму садашње вредности девизног курса увећаног за детерминистичку вредност μ .

Као што се може видети, карактеристике временске серије y_t , било за дневне, било за недељне податке, разликују се од особина које има нормална расподела, што указује на то да линеарни модел можда није најадекватнији у овом случају.

Како би се показало да је нелинеарни модел са променом режима бољи избор од линеарног, користиће се *SETAR* модел.

Укупан број опсервација базе које садржи недељне податке је 677. Овај скуп се дели на два скупа, скуп за обучавање, који ће садржати већину опсервација, 666, и скуп за тестирање са 11 опсервација који ће се користити за предвиђања ван узорка.

Прво се одређује променљива прага и ред кашњења *SETAR* модела користећи алтернативни *AIC* (3.13). Осим променљивих y_{t-d} , $d > 0$, такође се разматра и мера волатилности као кандидат за променљиву прага, јер важи да што је волатилност финансијске временске серије већа, то је она ризичнија. За меру волатилности се често узима стандардна девијација, али ће се у овом примеру, као што је препоручено у [1], променљива $v_{t,j}$ дефинисати као просек апсолутних повраћаја у последњих j недеља, тј.

$$v_{t,j} = \frac{1}{j} \sum_{i=0}^{j-1} |y_{t-i}|. \quad (11.3)$$

Разматрају се $v_{t-1,j}$ са $j = \overline{1,4}$ као потенцијалне променљиве прага, или прецизније:

$$\begin{aligned} v_{t-1,1} &= |y_{t-1}|, \\ v_{t-1,2} &= \frac{1}{2} (|y_{t-1}| + |y_{t-2}|), \\ v_{t-1,3} &= \frac{1}{3} (|y_{t-1}| + |y_{t-2}| + |y_{t-3}|) \text{ и} \\ v_{t-1,4} &= \frac{1}{4} (|y_{t-1}| + |y_{t-2}| + |y_{t-3}| + |y_{t-4}|). \end{aligned}$$

Променљиве y_{t-d} разматрају се за $d = \overline{1,4}$. Такође, посматрају се само модели у којима је ред *AR* модела у оба режима једнак и не већи од 5.

Да би се разумео код, потребно је представити на који начин је записан *SETAR* модел пакета "*tsDyn*" програмског језика *R*. *SETAR* модел временске серије y_t са два режима записује се као:

⁷⁹ asset

$$y_{t+s} = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_t + \varphi_{2,1} y_{t-d} + \dots + \varphi_{mL,1} y_{t-(mL-1)d} + \varepsilon_{t+s}, & \text{за } z_t \leq th, \\ \varphi_{0,2} + \varphi_{1,2} y_t + \varphi_{2,2} y_{t-d} + \dots + \varphi_{mH,2} y_{t-(mH-1)d} + \varepsilon_{t+s}, & \text{за } z_t > th, \end{cases} \quad (11.4)$$

при чему је s број корака предвиђања унапред, d време кашњења, и важи да је $s = d$, при чему је подразумевано да је $d = 1$. Параметар mL означава ред AR модела нижег реда, параметар mH ред AR модела вишег реда, z_t је променљива прага, а th вредност која утиче на промену режима.

За налажење најбољег модела на скупу за обучавање користи се функција ***selectSETAR*** која врши аутоматски избор хипер-параметара $SETAR$ модела. Поред горе наведених параметара који се односе на сам запис $SETAR$ модела, ова функција има и параметар *nthresh* који представља број вредности које утичу на промену режима. Његова подразумевана вредност је *nthresh* = 1, односно подразумева се да је реч о дворежимном $SETAR$ моделу. Навођењем *same.lags* = *TRUE* обезбеђује се да ред AR модела у свим режимима буде исти. Параметар *trim* обезбеђује минимални проценат опсервација посматране временске серије у сваком од режима. Подразумевана вредност је *trim* = 0.15. Параметар *criterion* је критеријум избора најбољег модела. Подразумевано је *criterion* = "pooled – AIC", односно алтернативни *AIC* (3.13). Параметри s и d су фиксирани. Прослеђивањем параметра m , који означава максимални ред AR модела у сваком од режима, и жељене променљиве прага, функција *selectSETAR* извршиће претрагу свих могућих комбинација одређених хипер-параметара и на основу критеријума *criterion* приказаће десет најбољих модела. На тај начин се израчунава број кашњења у сваком режиму, mL и mH , и вредност која утиче на промену режима, th .

Омогућено је да се променљива z_t изабере на неки од следећих начина:

- (1) $z_t = y_{t-thDelay \cdot d}$;
одакле се види улога параметра *thDelay*, с тим што он мора бити строго мањи од параметра m .
- (2) $z_t = y_t \cdot mTh[1] + y_{t-d} \cdot mTh[2] + \dots + y_{t-(m-1)d} \cdot mTh[m]$,
што се обезбеђује навођењем параметра *mTh* који представља вектор коефицијената уз одговарајуће заостале променљиве временске серије. Из самог записа се види да дужина вектора *mTh* мора бити једнака m .
- (3) $z_t = thVar_t$,
при чему параметар *thVar* обезбеђује да се као променљива преласка користи нека спољна променљива.

Запис $SETAR$ модела (11.4) за $d = s = 1$ дат је са:

$$y_{t+1} = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_t + \varphi_{2,1} y_{t-1} + \dots + \varphi_{mL,1} y_{t-(mL-1)} + \varepsilon_{t+s}, & \text{за } z_t \leq th, \\ \varphi_{0,2} + \varphi_{1,2} y_t + \varphi_{2,2} y_{t-1} + \dots + \varphi_{mH,2} y_{t-(mH-1)} + \varepsilon_{t+s}, & \text{за } z_t > th, \end{cases} \quad (11.5)$$

при чему је $z_t = y_{t-thDelay}$. Да би се ово усагласило са начином на који се записује $SETAR$ модел у теоријском делу и да би се користиле жељене променљиве прага, *thDelay* треба узети вредности 0, 1, 2 и 3. Да би се мера волатилности узела као променљива прага, користи се параметар *thVar*, што ће бити приказано у самом коду. Међутим, иако су ове променљиве прага дефинисане помоћу претходних вредности саме временске серије, оне неће бити третиране на тај начин, већ ће бити посматране као било које друге спољне променљиве. Због овога, резултати који се добијају можда

неће бити онакви какви се очекују, па се уводи још један кандидат за променљиву прага, просек повраћаја у претходних j недеља, у ознаци $p_{t,j}$ који се дефинише као:

$$p_{t,j} = \frac{1}{j} \sum_{i=0}^{j-1} y_{t-i}. \quad (11.6)$$

Разматра се $p_{t-1,j}$, са $j = \overline{2,4}$, тачније:

$$p_{t-1,2} = \frac{1}{2}(y_{t-1} + y_{t-2}),$$

$$p_{t-1,3} = \frac{1}{3}(y_{t-1} + y_{t-2} + y_{t-3}) \text{ и}$$

$$p_{t-1,4} = \frac{1}{4}(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}).$$

Да би се ово обезбедило, користи се $mTh = c\left(\frac{1}{2}, \frac{1}{2}\right)$ за $m = 2$, $mTh = c\left(\frac{1}{2}, \frac{1}{2}, 0\right)$ да би ред AR модела могао да буде $m = 3$, $mTh = c\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right)$ за $m = 4$, $mTh = c\left(\frac{1}{2}, \frac{1}{2}, 0, 0, 0\right)$ за $m = 5$ и аналогно за остале променљиве прага.

Вредности које враћа функција *selectSETAR* јесу вредности параметара m , вредност која утиче на промену прага th као и *pooled – AIC* у том случају. Коришћењем $\$best$ може се видети који параметри представљају најбољи избор. Међутим, она ће, на пример, за прослеђено $m = 5$ приказати десет најмањих вредности за *pooled – AIC*, али се често дешава да свих десет вредности буду баш за $m = 5$, иако није велика разлика између вредности *pooled – AIC*-а за неко $m < 5$ и оних за $m = 5$. Зато се позива функција *selectSETAR* за више различитих вредности прослеђеног максималног реда параметра $m \in \{2, 3, 4, 5\}$ и на основу тога одабира најадекватнији модел. Модел треба пре свега бити практичан и повећањем реда AR модела у сваком од режима не мора значити да се добио бољи модел. Често је модел мањег реда сасвим прихватљив.

При позиву функције *selectSETAR* мора се водити рачуна о ограничењима која постоје када је ред AR модела у питању.

Да би се генерисао конкретан *SETAR* модел, користи се функција *setar* којој се прослеђују вредности параметра m , односно mL и mH , и одговарајућа променљива прага. О њој ће касније бити више речи. Поново се наглашава да параметар m мора бити већи од параметра $thDelay$, као и да дужина вектора mTh мора бити једнака вредности параметра m .

Имајући ово у виду, резултати добијени коришћењем функције *selectSETAR* наводе се у табели 11.2. Променљиве $v_{t-1,j}$, $j = \overline{1,4}$ дефинишу се коришћењем функције *lag* која за прослеђени вектор $x = (x_1, x_2, x_3, \dots, x_n)$ враћа вектор $lag(x) = (NA, x_1, x_2, \dots, x_{n-1})$. Недостајуће вредности се морају избацити па се мора водити рачуна о томе да сви скупови буду једнаке дужине. Неки од позива функције *selectSETAR* приказују се ради илустрације.

```

selectSETAR(x_train, m=4, same.lags=TRUE, thDelay = 1, plot=FALSE)

selectSETAR(x_train, m=2, same.lags=TRUE, thvar = thvar1, plot=FALSE)
# sa thvar1,..., thvar4 oznacene su promenljive praga koje se odnose na volatilnost

selectSETAR(x_train, m=3, same.lags=TRUE, mTh = c(1/2,1/2,0), plot=FALSE)

```

Посебно се приказује излаз функције *selectSETAR* за прослеђену променљиву прага *thVar4*, односно $v_{t-1,4}$ јер се у [1] она користи као изабрана променљива прага на основу резултата добијених у програмском језику *GAUSS*. На основу онога што се добија коришћењем програмског језика *R*, она се не показује као најбоља променљива прага. Разлог за то је вероватно тај што се параметром *thVar* прослеђене променљиве прага виде као спољне променљиве.

```

> (m5 <- selectSETAR(x_train, m=5, same.lags=TRUE, thvar = thvar4, plot=FALSE))
Using maximum autoregressive order for low regime: mL = 5
Using maximum autoregressive order for high regime: mH = 5
Using only first 661 elements of thvar
Searching on 461 possible threshold values within regimes with sufficient ( 15% )
number of observations
Searching on 2305 combinations of thresholds ( 461 ), thDelay ( 1 ) and m ( 5 )
Results of the grid search for 1 threshold
  thDelay m      th pooled-AIC
1      0 5 2.009699 2690.380
2      0 5 1.940651 2690.833
3      0 5 1.941809 2690.833
4      0 5 1.944276 2690.833
5      0 5 2.011733 2690.849
6      0 5 2.023653 2690.849
7      0 5 2.032228 2690.849
8      0 5 1.907572 2690.856
9      0 5 1.908807 2690.856
10     0 5 1.931682 2691.172
> m5$bests
  thDelay      th pooled-AIC      m
0.000000 2.009699 2690.379968 5.000000

```

У табели 11.2 приказане су вредности *pooled – AIC*-а за различите вредности реда модела и различите променљиве прага.

Табела 11.2 – вредности за <i>pooled – AIC</i>				
Променљива прага	<i>p</i>			
	2	3	4	5
y_{t-1}	2698.38458	2697.84984	2695.46423	2688.81856
y_{t-2}	2696.30957		2690.29619	2682.55211
y_{t-3}				2694.41812
y_{t-4}				2690.26542
$v_{t-1,1}$	2697.25934		2694.67507	2690.37996
$v_{t-1,2}$	2699.56699		2697.00611	2690.37996
$v_{t-1,3}$	2700.58419	2700.35681	2697.29050	2690.37996
$v_{t-1,4}$	2700.62104	2700.35681	2697.29050	2690.37997
$p_{t-1,2}$	2698.385	2697.84984	2695.46423	2688.81856
$p_{t-1,3}$		2698.36460	2695.99770	2689.23701
$p_{t-1,4}$			2696.62960	2689.23701

Издвојене вредности представљају најмање вредности *pooled – AIC*-а за модел реда 2 и модел реда 5, иако се оне суштински много не разликују.

Занимљиво је да се када функцији *selectSETAR* проследи скуп за обучавање који садржи само првих 70% опсервација, добијају се мање вредности *pooled – AIC*-а. Повећањем скупа за обучавање очекује се да добијени модел буде што бољи. Овде то није случај, што може указивати на то да у другом делу базе постоје аутлајери који значајно утичу на избор модела, о чему ће касније бити речи. Резултати су приказани у табели 11.3, иако ће се у наставку користити већи скуп за обучавање.

Табела 11.3 - вредности за <i>pooled – AIC</i>				
Променљива прага	<i>p</i>			
	2	3	4	5
y_{t-1}	1971.66027	1972.052	1969.84066	1962.58805
y_{t-2}	1975.41754	1975.533	1970.49492	1963.00665
y_{t-3}				1973.51315
y_{t-4}				1966.009198
$v_{t-1,1}$	1975.83342		1974.42967	1967.59643
$v_{t-1,2}$	1979.73129	1979.55565	1977.13542	1968.94404
$v_{t-1,3}$	1979.61002	1979.23642	1977.13542	1968.94404
$v_{t-1,4}$	1979.61002	1979.23642	1977.13542	1968.94404
$p_{t-1,2}$	1972.52058	1972.926	1970.74638	1963.31156
$p_{t-1,3}$		1972.052	1969.841	1962.58805
$p_{t-1,4}$			1969.84066	1962.58805

На основу резултата Табеле 11.2, али и Табеле 11.3, види се да није велика разлика између вредности *pooled* – *AIC*-а за различит реда модела и прослеђене променљиве прага. Зато ће се генерисати неки од модела реда 2. Применом функције **summary**, између осталог, може се видети вредност и значајност оцењених параметара. Како је излаз ове функције обимнији, приказује се у потпуности само за *SETAR* модел реда 2 са променљивом прага *thDelay* = 1, односно y_{t-2} .

```
> mod_1 <- setar(x_train, mL=2, mH=2, thDelay=1)
> summary(mod_1)
```

Non linear autoregressive model

SETAR model (2 regimes)

Coefficients:

Low regime:

const.L	phiL.1	phiL.2
0.04916926	0.04093929	0.11524166

High regime:

const.H	phiH.1	phiH.2
1.5745943	0.0738488	-0.5585644

Threshold:

-variable: $z(t) = + (0) x(t) + (1)x(t-1)$

-value: 1.675

Proportion of points in low regime: 83.43% high regime: 16.57%

Residuals:

Min	1Q	Median	3Q	Max
-6.90565001	-1.12537606	0.00067234	1.13799762	8.10978191

Fit:

residuals variance = 3.326, AIC = 814, MAPE = 144.9%

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
const.L	0.049169	0.082676	0.5947	0.552233
phiL.1	0.040939	0.042950	0.9532	0.340844
phiL.2	0.115242	0.052761	2.1842	0.029297 *
const.H	1.574594	0.557464	2.8246	0.004877 **
phiH.1	0.073849	0.089694	0.8233	0.410612
phiH.2	-0.558564	0.193936	-2.8801	0.004104 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
variable: $z(t) = + (0) x(t) + (1) x(t-1)$

value: 1.675

```
> |
```

Приказане су оцењене вредности коефицијената у оба режима, променљива прага која се користи, вредност која утиче на промену режима, $th = 1.675$. У нижем режиму је 83.43% опсервација, док је у вишем свега 16.57% опсервација. Могу се видети и неке од статистика резидуала оцењеног модела. Вредност за AIC која је приказана односи се на стандардан AIC , што није од превеликог значаја. Дате су и стандардне грешке оцењених коефицијената модела, вредност t -статистике и p вредност на основу које се може видети колико је сваки од коефицијената значајан. Што се тиче нижег режима, значајан је коефицијент уз y_{t-2} . Што се тиче вишег режима, значајан је слободан члан и коефицијент уз y_{t-2} , при чему су коефицијенти вишег режима значајнији. Два коефицијента вишег режима су значајна, иако он има свега $\sim 16\%$ опсервација, док је значајан само један коефицијент нижег режима са $\sim 84\%$ опсервација.

Сада ће се претходни позив функције допунити навођењем `include = "none"` чиме се омогућава да генерисани модел буде без слободних чланова.

```
> mod_1_0 <- setar(x_train, mL=2, mH=2, thDelay=1, include="none")
> summary(mod_1_0)
```

Proportion of points in low regime: 15.06% high regime: 84.94%

Residuals:

	Min	1Q	Median	3Q	Max
	-7.52435	-1.08093	0.12927	1.19835	8.24409

Fit:

residuals variance = 3.356, AIC = 816, MAPE = 124.1%

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
phiL.1	-0.057364	0.094163	-0.7154	0.47461
phiL.2	0.141839	0.059032	2.4028	0.01655 *
phiH.1	0.038893	0.042409	2.0961	0.03645 *
phiH.2	-0.022708	0.051158	-0.4439	0.65728

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: $z(t) = + (0) x(t) + (1) x(t-1)$

value: -1.746

Занимљиво је да се само избацивањем слободних чланова добија да је $th = -1.746$. У нижем режиму је свега 15.06% опсервација, што једва прелази границу за вредност параметра *trim*. У вишем режиму је 84.94% опсервација. Ово је у супротности са резултатима који се добијају када у се модел укључе слободни чланови. Што се значајности коефицијената тиче, они су мање значајни од коефицијената претходно приказаног модела.

Поново се позива функција *selectSETAR* да би се видела вредност за th која се препоручује. Та вредност ће бити прослеђена функцији *setar*.

```

> s <- selectSETAR(x_train, m=2, same.lags=TRUE, thDelay=1, plot=FALSE)
Using maximum autoregressive order for low regime: mL = 2
Using maximum autoregressive order for high regime: mH = 2
Searching on 462 possible threshold values within regimes with sufficient ( 15% )
number of observations
Searching on 924 combinations of thresholds ( 462 ), thDelay ( 1 ) and m ( 2 )
> th<- s$bests[['th']]
> mod_1_2<- setar(x_train, mL=2, mH=2, thDelay=1, th=th)
> summary(mod_1_2)

Proportion of points in low regime: 17.02%          High regime: 82.98%

Residuals:
      Min       1Q   Median       3Q      Max
-7.679328 -1.155714  0.042471  1.184554  8.319466

Fit:
residuals variance = 3.34,  AIC = 815, MAPE = 126.3%

Coefficient(s):

      Estimate Std. Error t value Pr(>|t|)
const.L  1.1429552   0.4651978   2.4569 0.014270 *
phiL.1   -0.0182545   0.0940997  -0.1940 0.846243
phiL.2    0.4524907   0.1547560   2.9239 0.003575 **
const.H   0.0168882   0.0845786   0.1997 0.841797
phiH.1    0.0873407   0.0425262   2.0538 0.040387 *
phiH.2    0.0025825   0.0560457   0.0461 0.963262
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: z(t) = + (0) x(t) + (1) x(t-1)

value: -1.648 (fixed)

```

У нижем режиму је 17.02% опсервација, док је у вишем 82.98% опсервација. Вредност која утиче на промену режима једнака је $th = -1.648$, што је слично вредности за th коју даје модел без слободних чланова. Два значајна коефицијента су у нижем режиму који има свега $\sim 17\%$ опсервација, док је само један значајан коефицијент вишег режима са $\sim 83\%$ опсервација.

Како је код трећег модела, *mod_1_2*, у нижем режиму 17.02% опсервација, док је код осталих модела проценат опсервација у режиму са мањим број опсервација мањи, и узимајући у обзир значајност коефицијената, која је код првог и трећег модела слична, може се рећи да је модел *mod_1_2* најбољи од приказаних.

Сада ће се приказати неки од позива функције *setar*, значајни коефицијенти сваког генерисаног модела, број опсервација у сваком од режима, као и вредност која утиче на промену режима. Звезде представљају значајност коефицијената која је дата на следећи начин:

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```

mod_2 <- setar(x_train, mL=2, mH=2, thDelay=0)
summary(mod_2)
# const.L **
# phiL.1 .
# phiH.2 .
# nizi rezim: 30.12%   visi rezim: 69.88%
# th = -0.8885

mod_3 <- setar(x_train, mL=2, mH=2, thvar=thvar1)
summary(mod_3)
# phiL.1 *
# phiH.2 ***
# nizi rezim: 62.5%   visi rezim: 37.5%
# th = 1.538

mod_4 <- setar(x_train, mL=2, mH=2, thvar=thvar2)
summary(mod_4)
# phiL.1 ***
# nizi rezim: 63.7%   visi rezim: 36.3%
# th = 1.584

mod_5 <- setar(x_train, mL=2, mH=2, thvar=thvar3)
summary(mod_5)
# phiL.1 ***
# nizi rezim: 75.15%   visi rezim: 24.85%
# th = 1.799

mod_6 <- setar(x_train, mL=2, mH=2, thvar=thvar4)
summary(mod_6)
# phiL.1 **
# nizi rezim: 46.69%   visi rezim: 53.31%
# th = 1.253

mod_7<- setar(x_train, mL=2, mH=2, mTh=c(1/2,1/2))
summary(mod_7)
# const.L *
# phiL.1 *
# nizi rezim: 16.87%   visi rezim: 83.13%
# th = -1.242

```

Види се да модел *mod_3* са променљивом прага *thVar1* у нижем режиму има ~62% опсервација, а у вишем скоро ~38% опсервација. Иако је мањи број значајних коефицијената, чини се да је *mod_3* бољи од модела *mod_1_2*.

Наравно, треба тестирати линеарност против *SETAR* - нелинеарности. У ту сврху користи се функција *setarTest* која користи *bootstrap* методу и резидуале оцењеног модела под нултом хипотезом уз претпоставку да су они нормално расподељени и хомоскедастични.

Функцији *setarTest* је као променљиву прага могуће проследити само вредност за *thDelay*. Спецификацијом параметра *test* могу се тестирати различите хипотезе. Навођењем:

- (1) *test* = "1vs" тестира се нулта хипотеза о линеарности против *SETAR* модела са два и три режима, тј. обухватају се два теста *test* = "1vs2" и *test* = "1vs3",

- (2) $test = "2vs3"$ тестира се нулта хипотеза *SETAR* модела са два режима против алтернативе *SETAR* модела са три режима.

Подразумевана вредност аргумента *trim* у овој функцији је $trim = 0.10$.

```
> test1 <- setarTest(x_train, m=2, thDelay=1, test='1vs', trim=0.15)

1 T: Trim not respected: 0.1490964 0.8509036 from th: -1.720304
1 T: Trim not respected: 0.8509036 0.1490964 from th: 1.667837
1 T: Trim not respected: 0.1490964 0.8509036 from th: -1.892177
> summary(test1)
Test of linearity against setar(2) and setar(3)

      Test Pval
1vs2 12.38178 0.2
1vs3 20.83080 0.2

Critical values:
      0.9      0.95      0.975      0.99
1vs2 12.57223 13.19625 13.50826 13.69547
1vs3 23.84265 24.93584 25.48244 25.81040

SSR of original series:
      SSR
AR      2256.459
SETAR(2) 2215.152
SETAR(3) 2187.823

Threshold of original series:
      th1      th2
SETAR(2) 1.674705      NA
SETAR(3) -1.606795 1.674705

Number of bootstrap replications: 10
Asymptotic bound: 2196.456
```

Навођењем да је $trim = 0.15$ тестирање хипотезе иде у корист линеарном моделу. Такође, може се видети за које вредности прага није испуњен услов дат параметром *trim*.

```

> test2 <- setarTest(x_train, m=2, thDelay=1, test='1vs')

1 T: Trim not respected: 0.09939759 0.9006024 from th: -2.129808
1 T: Trim not respected: 0.9006024 0.09939759 from th: 2.143965
> summary(test2)
Test of linearity against setar(2) and setar(3)

      Test Pval
1vs2 15.33313 0.0
1vs3 24.80742 0.1

Critical values:
      0.9      0.95      0.975      0.99
1vs2 13.85038 13.94464 13.99177 14.02005
1vs3 24.04052 25.39874 26.07786 26.48533

SSR of original series:
      SSR
AR      2256.459
SETAR(2) 2205.529
SETAR(3) 2175.192

Threshold of original series:
      th1      th2
SETAR(2) 2.237313      NA
SETAR(3) -1.606795 2.237313

Number of bootstrap replications: 10
Asymptotic bound: 2720.008

```

Сада је p вредност теста "1vs2" мала, што значи да се прихвата хипотезу о *SETAR* моделу са два режима.

```

> test3 <- setarTest(x_train, m=2, thDelay=1, test = '2vs3')

1 T: Trim not respected: 0.09939759 0.9006024 from th: -2.259981
1 T: Trim not respected: 0.9006024 0.09939759 from th: 2.196448
> summary(test3)
Test of setar(2) against setar(3)

      Test Pval
2vs3 9.260453 0.5

Critical values:
      0.9      0.95      0.975      0.99
2vs3 12.41994 14.82831 16.03249 16.755

SSR of original series:
      SSR
AR      2256.459
SETAR(2) 2205.529
SETAR(3) 2175.192

Threshold of original series:
      th1      th2
SETAR(2) 2.237313      NA
SETAR(3) -1.606795 2.237313

Number of bootstrap replications: 10

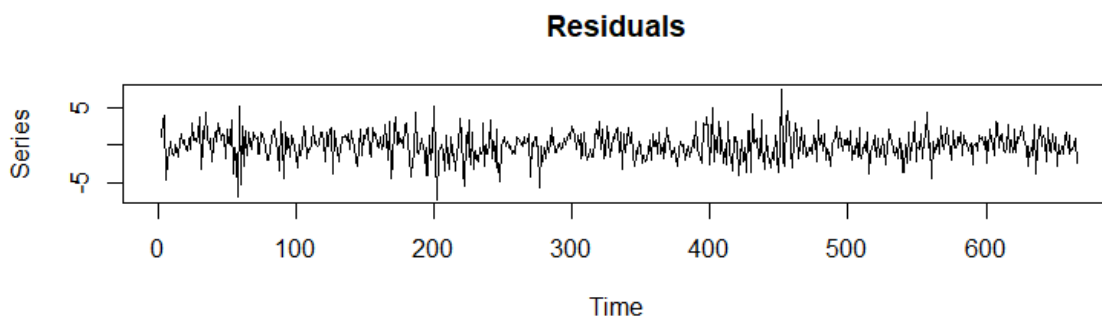
```

Овде је p вредност тест статистике једнака 0.5, што значи да се не одбацује нулта хипотеза о *SETAR* моделу са два режима.

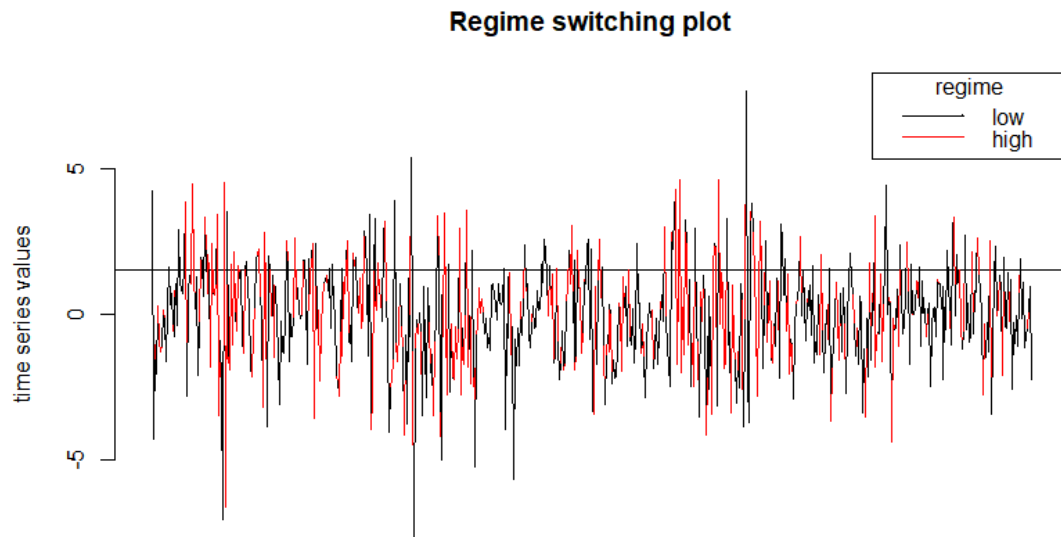
Сада ће се испитати резидуали два изабрана модела, који ће се због практичности преименовати. Функција **regime** враћа индикаторску променљиву чија је вредност 1 ако се у том тренутку модел налази у нижем режиму, и вредност 2 ако се у том тренутку модел налази у вишем режиму. Користећи функцију *sum* може се добити колико је опсервација у ком режиму, иако се то може израчунати и на основу процента опсервација у сваком од режима.

```
> model_setar_1 <- mod_1_2
> model_setar_2 <- mod_3
> # prvi model
> sum(regime(model_setar_1, initval=FALSE)==1)
[1] 113
> sum(regime(model_setar_1, initval=FALSE)==2)
[1] 551
> # drugi model
> sum(regime(model_setar_2, initval=FALSE)==1)
[1] 415
> sum(regime(model_setar_2, initval=FALSE)==2)
[1] 249
```

Једноставним позивом *plot(model_setar_2)* добија се неколико графика. Неки од њих представљени су на слици 11.3.



Слика 11.3 а) График резидуала посматраног модела



Слика 11.3 б) Приказ вредности опсервација временске серије са назнаком ком режиму свака опсервација припада ; хоризонтална линија представља вредност која утиче на промену режима $th = 1.538$

На графицима слике 11.3 може се видети како изгледају резидуали модела као и вредност прага која је коришћена. Остаје да се тестира нормалност резидуала.

```
> e_setar_2 <- residuals(model_setar_2)
> head(e_setar_2)
Time Series:
Start = 1
End = 6
Frequency = 1
[1]      NA      NA  1.175428  3.375547  3.981973 -4.568966
> e_setar_2 <- e_setar_2[!is.na(e_setar_2)]
> jarque.bera.test(e_setar_2)

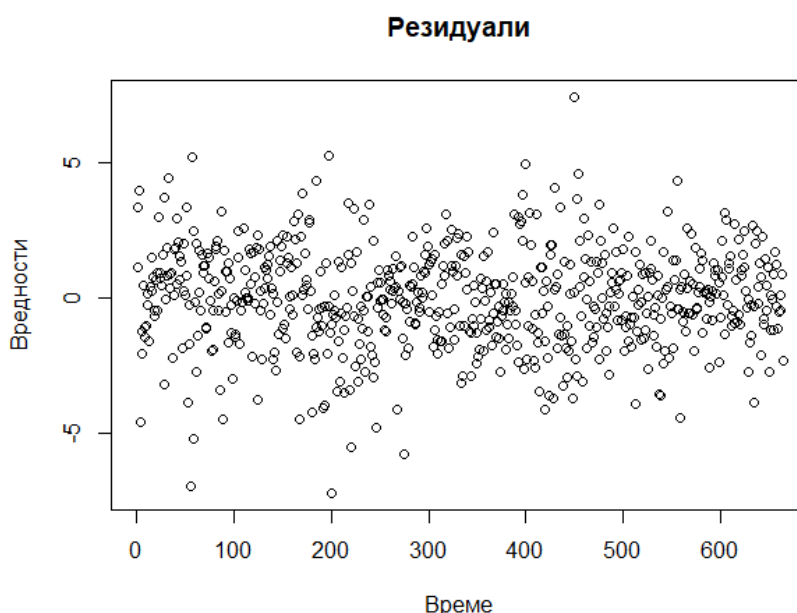
Jarque Bera Test

data:  e_setar_2
X-squared = 28.003, df = 2, p-value = 8.303e-07
```

Види се да је p вредност тест статистике јако мала, што значи да се одбацује нулта хипотеза о нормалности резидуала. Ово може бити последица хетероскедастичности или присуства аутлајера, што се може наслутити на основу вредности *pooled* — *AIC*-а мањег скупа за обучавање. Проблем постајања аутлајера могао би се решити коришћењем робусних оцена, што превазилази оквире овог рада. Хетероскедастичност такође не сме бити занемарена и она може бити индикатор погрешне спецификације модела.

Занемарена хетероскедастичност може довести до лажног одбацивања нулте хипотезе о линеарности, па би се на основу овога могло посумњати да је адекватнији линеарни модел, јер резултати добијени применом функције *setarTest* нису меродавни. Међутим, на основу резултата приказаних у књизи [1] и статистика саме временске серије, јасно се види да серију боље описује нелинеарни модел, иако је врло вероватно потребно користи модел заснован на робусним оценама.

На слици 11.10 биће приказан график резидуала на основу кога би се могло нешто више закључити о њиховој хомоскедастичности/ хетероскедастичности.



Слика 11.10 Резидуали модела *model_setar_2*

На основу приказаног могло би се рећи да су резидуали хомоскедастични. Међутим, ово се не може сматрати чврстим доказом, па ће се покушати са генерисањем модела реда 5 са y_{t-2} као променљивом прага, јер се на основу вредности *pooled* – *AIC*-а ово чини као добар избор. Следи и излаз функције *summary*.

```

mod <- setar(x_train, m=5, thDelay = 1)
summary(mod)
Proportion of points in low regime: 15.43%      High regime: 84.57%

Residuals:
      Min       1Q   Median       3Q      Max
-7.739742 -1.078279  0.028977  1.166606  8.070454

Fit:
residuals variance = 3.242,  AIC = 809, MAPE = 145.1%

Coefficient(s):

      Estimate Std. Error t value Pr(>|t|)
const.L  0.7924405   0.5018760   1.5790  0.11483
phiL.1   -0.0297125   0.0951700  -0.3122  0.75498
phiL.2    0.3695573   0.1613569   2.2903  0.02232 *
phiL.3    0.0279843   0.0875175   0.3198  0.74926
phiL.4    0.1860179   0.0799465   2.3268  0.02028 *
phiL.5    0.2082725   0.0929893   2.2397  0.02544 *
const.H   0.0438953   0.0820487   0.5350  0.59284
phiH.1    0.0739477   0.0422372   1.7508  0.08045 .
phiH.2   -0.0242700   0.0545691  -0.4448  0.65664
phiH.3   -0.0075344   0.0429111  -0.1756  0.86068
phiH.4   -0.0756970   0.0439321  -1.7230  0.08535 .
phiH.5    0.0292385   0.0421820   0.6932  0.48846
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
variable: Z(t) = + (0) x(t) + (1) x(t-1)+ (0) x(t-2)+ (0) x(t-3)+ (0) x(t-4)

value: -1.719

```

Види се да коефицијенти вишег режима немају велику значајност, па се већ сада може одустати од овог модела.

На основу резултата теста о тестирању (не)линеарности, показује се да нема потребе генерисати *SETAR* модел са три режима, јер је тест био у корист *SETAR* модела са два режима. Међутим, вреди покушати са генерисањем трорежимног модела, јер се показује⁸⁰ да се повраћај девизних курсева најбоље описује помоћу модела са три режима и променљивом прага y_{t-1} , при чему је серија белог шума у спољним режимима (где y_{t-1} има велику апсолутну вредност па је самим тим волатилност велика), а средњи режим се може описати помоћу стационарног $AR(p)$ модела у (где y_{t-1} има мање апсолутне вредности па је и волатилност умерена).

Генерисаће се *SETAR* модел са три режима код кога је ред модела првог и трећег режима једнак 1 (јер функција *setar* не омогућава серију белог шума у неком од режима), за различите вредности реда модела средњег режима. Који год ред (≤ 5) *AR* модела средњег режима да је прослеђен функцији *setar*, добиће се да су коефицијенти средњег режима безначајни. Због практичности, приказује се само један од тих модела генерисан на следећи начин:

⁸⁰ Krager and Kugler 1993.,

```
> mod <- setar(x_train, mL=1, mM=5, mH=1, nthresh=2)
Warning message:
Possible unit root in the medium regime. Roots are: 0.7328
1.6792 2.3676 3.0623 2.3676
```

Добија се упозорење од могућности постојање јединичних корена у средњем режиму, тј. добијени модел можда није стационаран. Свакако ће се приказати излаз функције *summary*, иако теоријски није представљен модел код ког је ред модела међу режимима различит.

```
Proportion of points in low regime: 43.27%      Middle regime: 16.79%      High regime: 39.94%

Residuals:
      Min       1Q   Median       3Q      Max
-7.440590 -1.118947  0.036384  1.131084  6.783742

Fit:
residuals variance = 3.244,  AIC = 808,  MAPE = 153%

Coefficient(s):

      Estimate Std. Error t value Pr(>|t|)
const.L -4.7813e-01  1.8178e-01  -2.6303  0.008732 **
phi.L.1 -1.1883e-01  8.9727e-02  -1.3244  0.185830
const.M  2.4392e-01  1.9216e-01   1.2694  0.204760
phi.M.1 -1.4829e+00  8.6794e-01  -1.7085  0.088019 .
phi.M.2 -1.0464e-01  1.1417e-01  -0.9165  0.359735
phi.M.3 -1.5576e-05  1.0203e-01  -0.0002  0.999878
phi.M.4 -7.0813e-02  1.0796e-01  -0.6559  0.512119
phi.M.5  4.7345e-02  9.0428e-02   0.5236  0.600761
const.H  6.3017e-01  2.1612e-01   2.9159  0.003668 **
phi.H.1 -2.2593e-01  1.0720e-01  -2.1075  0.035456 *
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: Z(t) = + (1) x(t) + (0) x(t-1)+ (0) x(t-2)+ (0) x(t-3)+ (0) x(t-4)

value: -0.2622  0.4456
```

На основу процента опсервација које припадају сваком од режима, може се рећи да генерисани модел има смисла, или бар донекле може потврдити оно што је показано за повраћај девизних курсева; да се најбоље описују трорежимним моделом, иако *JB* тест статистика примењена на резидуале овог модела говори да они нису нормално расподељени, а самим тим ни модел није одговарајући.

Како се не добија модел који има нормално расподељене резидуале, што је основа за добијање добрих предвиђања, нема смисла оцењивати модел на скупу за тестирање. Међутим, модели који су изабрани као најбољи за анализу, добијени су на основу претпоставке о нормалној расподељености хомоскедастичних резидуала. Такође, оцене параметара генерисаних модела добијене су под истом претпоставком.

Посматрана база садржи реалне податке са берзе, па је очекивано да се применом основних метода не добије одговарајући модел. Међутим, коришћењем робусних оцена и тестова за

тестирање хетероскедастичности могао би се добити модел који добро описује посматрану временску серију, што превазилази оквире овог рада.

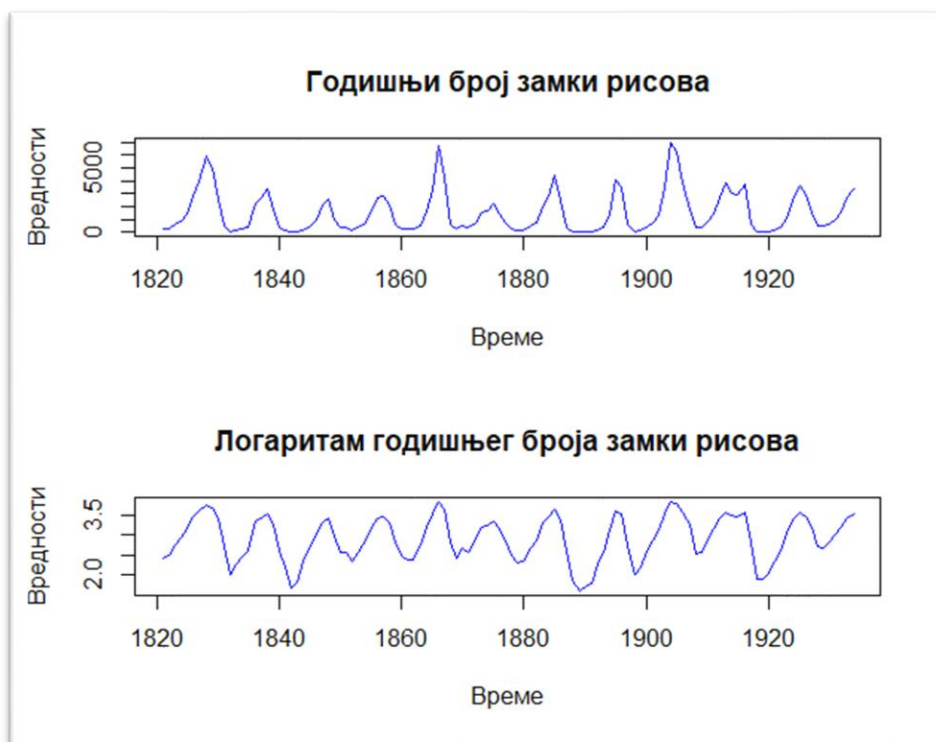
Ипак, представљен је основни начин размишљања који се користи при одређивању *SETAR* модела, описане су неке од функција које су развијене у програмском језику *R* пакета „*tsDyn*“, као и неки од проблема који се могу јавити.

11.2. Пример 2.

У овом примеру је на бази *lynx*, која садржи годишњи број замки за рисове у Канади у периоду од 1821. до 1934. године, приказан цео поступак описан у теоријском делу. Једноставним позивом `?lynx` могу се видети доступне информације жељене базе која је садржана у пакету „*datasets*“ програмског језика *R*.

Укупан број опсервација је 114. Скуп је мали, али довољно репрезентативан. Он се дели на скуп за обучавање, који садржи 104 опсервација, и скуп за тестирање са преосталих 10 опсервација. Користе се логаритмоване вредности посматране временске серије јер се показује да ова трансформација често стабилизује дисперзију временске серије. При томе је лакше радити са мањим вредностима које се добијају логаритмовањем.

На слици 11.4 је приказана оригинална и логаритмована временска серија.



Слика 11.4 У горњем делу приказан је годишњи број замки рисова у Канади у периоду од 1821. до 1934. године; У доњем делу слике приказан је логаритам ових вредности

Прво се одређује најбољи *SETAR* модел који се може добити на скупу за обучавање. Како су функције које се користе за одређивање *SETAR* модела описане у претходном примеру, наводи се само основно. Функцијом *selectSETAR* на основу *pooled – AIC*-а налази се најбољи модел, затим се тестира хипотезу о линеарности коришћењем функције *setarTest*. Након тога се генерише модел функцијом *setar*.

```
m1<- selectSETAR(x.train, m=5, same.lags = TRUE, thDelay=0, plot=F)
m2<- selectSETAR(x.train, m=5, same.lags = TRUE, thDelay=1, plot=F)
m3<- selectSETAR(x.train, m=2, same.lags = TRUE, thDelay=1, plot=F)
```

```
> m1$bests
  thDelay      th pooled-AIC      m
0.000000  2.557507 -17.101809  2.000000
> m2$bests
  thDelay      th pooled-AIC      m
1.000000  3.310056 -20.300859  3.000000
> m3$bests
  thDelay      th pooled-AIC      m
1.000000  3.310056 -18.572979  2.000000
```

На основу приказаног одабира се *SETAR* модел реда 2 и променљивом прага $thDelay = 1$, односно променљивом прага y_{t-2} .

Резултати тестова за тестирање хипотезе о (не)линеарности приказани су у наставку.

```

> (test1<- setarTest(x.train, m = 2, thDelay = 1, check = F, test = "1vs"))

1 T: Trim not respected: 0.09803922 0.9019608 from th: 2.326073
1 T: Trim not respected: 0.09803922 0.9019608 from th: 2.35155Test of linearity
against setar(2) and setar(3)

      Test Pval
1vs2 30.29186    0
1vs3 39.98759    0
Warning messages:
1: Possible unit root in the high regime. Roots are: 1 1
2: Possible unit root in the high regime. Roots are: 1 1

> (test2<- setarTest(x.train, m = 2, thDelay = 1, check = F, test = "2vs3"))

1 T: Trim not respected: 0.09803922 0.9019608 from th: 2.332414
1 T: Trim not respected: 0.9019608 0.09803922 from th: 3.588778
1 T: Trim not respected: 0.09803922 0.9019608 from th: 2.204201Test of setar(2)
against setar(3)

      Test Pval
2vs3 7.475626 0.6
Warning messages:
1: Possible unit root in the high regime. Roots are: 1 1
2: Possible unit root in the high regime. Roots are: 1 1
3: In if (positionSecond <= length(storeMinus)) newThresh <- gammaMinus[positionSecond] else newThresh <- gammaPlus[positionSecond - 1] :
the condition has length > 1 and only the first element will be used

```

Као што се може видети, иако се прихвата хипотеза о *SETAR* моделу са два режима, јавља се упозорење о могућности постојања јединичног корена у вишем режиму. Позивом *m2\$bests* добија се да је најбољи ред модела једнак 3 када је променљива прага *thDelay* = 1. Позивом функције *selectSETAR* за максимални ред модела 2, добија се вредност *pooled* – *AIC*-а која је мања од оне која се добија позивом функције са променљивом преласка *thDelay* = 0. Међутим, упозорења која се добијају могу бити индикатор тога да је процена била лоша и да је ипак модел реда 2 са променљивом преласка *thDelay* = 0 бољи. То се најпре проверава тестирањем хипотезе о (не)линеарности.

```

> (test3<- setarTest(x.train, m = 2, thDelay = 0, trim = 0.15, check = F, test = "1vs"))
Test of linearity against setar(2) and setar(3)

      Test Pval
1vs2 28.93258    0
1vs3 37.45568    0
> (test4<- setarTest(x.train, m = 2, thDelay = 0, trim=0.15, check = F, test = "2vs3"))
Test of setar(2) against setar(3)

      Test Pval
2vs3 6.639726 0.3

```

Нема никаквих упозорења о постојању јединичних корена иако се уместо подразумеване вредности за *trim* у датој функцији (*trim* = 0.10) наводи да је *trim* = 0.15. На основу *p* вредности тестова види се да су и *SETAR* модел са два режима и *SETAR* модел са три режима бољи избор од линеарног модела, али да је *SETAR* модела са два режима бољи избор од *SETAR* модела са три

режима. Зато се одлучује за *SETAR* модел са два режима код кога је ред *AR* модела у сваком од режима једнак два, а променљива прага $thDelay = 0$, односно y_{t-1} .

Сада се генерише описани *SETAR* модел са вредношћу прага која је изабрана функцијом *selectSETAR*.

```
> m_select <- selectSETAR(x.train, m = 2, same.lags = TRUE, thDelay = 0, plot = F)
Using maximum autoregressive order for low regime: mL = 2
Using maximum autoregressive order for high regime: mH = 2
Searching on 66 possible threshold values within regimes with sufficient ( 15% ) number
of observations
Searching on 132 combinations of thresholds ( 66 ), thDelay ( 1 ) and m ( 2 )
> (m_select$th)
      th
2.557507
> model_setar<- setar(x.train, m = 2, thDelay = 0, th = m_select$th)
> summary(model_setar)
```

```
Proportion of points in low regime: 30.39%      High regime: 69.61%

Residuals:
      Min       1Q   Median       3Q      Max
-0.594904 -0.123141  0.039197  0.133222  0.493207

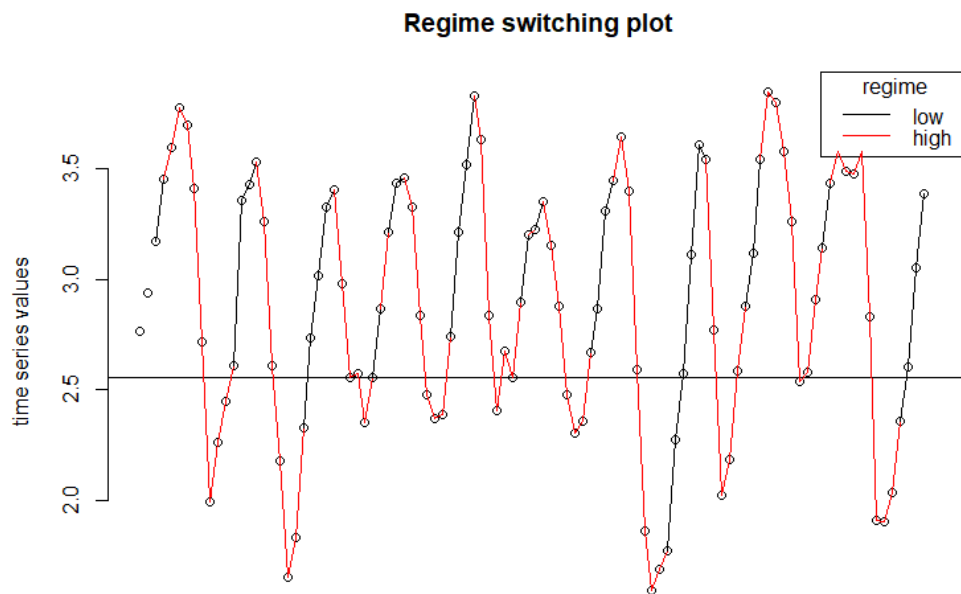
Fit:
residuals variance = 0.04209,  AIC = -317, MAPE = 6.118%

Coefficient(s):
              Estimate Std. Error t value Pr(>|t|)
const.L    0.405943    0.294594   1.3780  0.171350
phiL.1     1.245677    0.151636   8.2149 8.806e-13 ***
phiL.2    -0.333929    0.101346  -3.2949 0.001371 **
const.H     1.123878    0.218980   5.1323 1.450e-06 ***
phiH.1     1.571763    0.091381  17.2001 < 2.2e-16 ***
phiH.2    -0.966309    0.074140 -13.0336 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: z(t) = + (1) x(t) + (0) x(t-1)

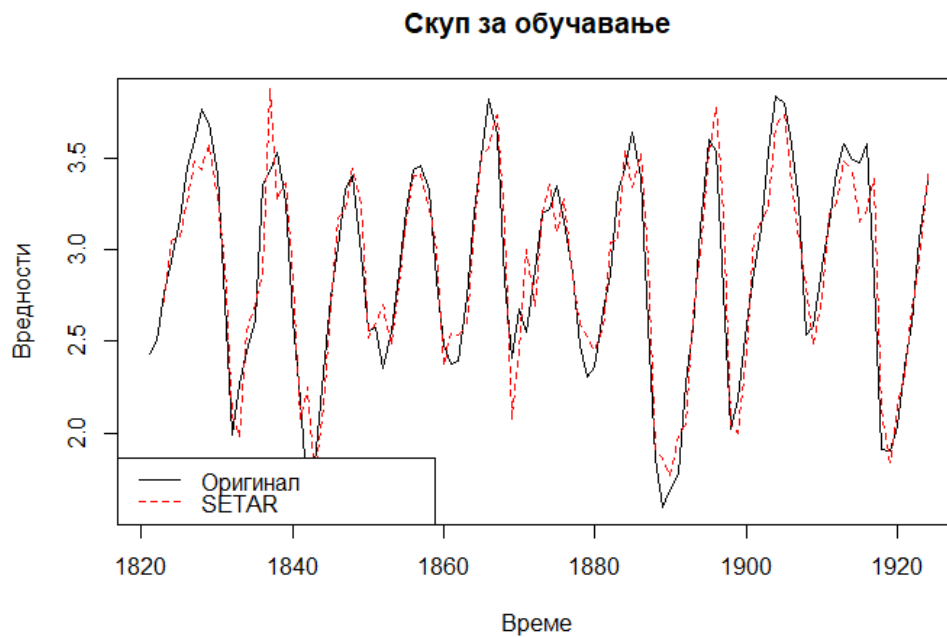
value: 2.558 (fixed)
```

На основу процента опсервација у сваком од режима, види се да је модел одговарајући. У примеру 1 дисперзија резидуала била је ~ 3 , док је у овом случају дисперзија резидуала ~ 0.04 , што је одличан показатељ квалитета модела. Такође, коефицијенти модела су веома значајни. На слици 11.5 може се видети ком режиму припада свака од опсервација посматране временске серије. Због мањег броја опсервација које су у скупу за обучавање, график који се добија позивом *plot(model_setar)* прегледнији је од оног који је приказан у примеру 1.



Слика 11.5 Приказ вредности опсервација временске серије са назнаком ком режиму свака опсервација припада ; хоризонтална линија представља вредност која утиче на промену режима $th = 2.5575$

На слици 11.6 приказане су вредности посматране временске серије и вредности добијене генерисаним *SETAR* моделом. Види се да се модел добро уклопио у податке скупа за обучавање.



Слика 11.6 Приказ оригиналних вредности временске серије и вредности добијене генерисаним *SETAR* моделом

Испод су приказани резултати теста за тестирање нормалности резидуала.

```
> e_setar <- residuals(model_setar)
> e_setar <- e_setar[!is.na(e_setar)]
> jarque.bera.test(e_setar)

      Jarque Bera Test

data:  e_setar
X-squared = 3.4719, df = 2, p-value = 0.1762
```

На основу p вредности теста види се да су резидуали нормално расподељени, па има смисла предвиђати на скупу за тестирање. Најпре треба пробати са одређивањем најбољег *STAR* и *LSTAR* модела и у случају да и ова два модела имају добре перформансе на скупу за обучавање, извршава се предвиђање на скупу за тестирање и упоређују се резултати сва три модела са онима које даје *AR* модел реда 2.

За налажење *STAR* модела постоји имплементирана само једна функција *star* која је описана у наставку. Овом функцијом аутоматски се тестира линеарност насупрот *STAR* нелинеарности, најпре за *STAR* модел са два режима, при чему се може видети и p вредност тест статистике. У случају да је одбијена хипотеза о линеарности, ова функција генерише *STAR* модел са два режима користећи мрежу претраге и исписује најбољу вредност параметра глаткости γ и вредност која утиче на промену прага, γ ознаци *th*. Затим се тестира да ли је потребно додавати трећи режим исписујући p вредност тест статистике. Могу се видети и коефицијенти модела као и остали параметри. Приликом процене параметра глаткости γ и вредности која утиче на промену прага, *th*, као и за рачунања стандардних грешака, може доћи до проблема са конвергенцијом. У таквој ситуацији корисна је опција *maxit* којом се задаје максимални број итерација методе оптимизације. Аргумент *starting.control* омогућава да се ограничи број вредности које утичу на вредност прага *th*, чија је подразумевана вредност $nTh = 200$. Може се утицати и на број вредности за γ у мрежи претраге. Подразумевана вредност је $nGamma = 40$. Минималан број опсервација у сваком од режима обезбеђује се параметром *trim* чија је подразумевана вредност $trim = 0.10$, тј. 10%. Могућа вредност за *th* у том случају налази се између 0.1 и 0.9 квантила. Подразумевани интервал у коме ће се наћи оцена за γ је (1,40). То се може променити навођењем аргумента *gammaInt*.

Неки од излаза функције *star* биће приказани како би било јасније зашто се одлучује за конкретан модел.

```

> control <- list(maxit = 3000)
> controls <- list(gammaInt = c(1,500), nGamma = 100, trim = 0.15)
> (star0 <- star(x.train, m = 2, thDelay = 1, control = control, starting.control = controls))
Testing linearity... p-value = 0.0009537218
The series is nonlinear. Incremental building procedure:
Building a 2 regime STAR.
Performing grid search for starting values...
Starting values fixed: gamma = 8.615385 , th = 3.363624 ; SSE = 4.312901
Optimization algorithm converged
Optimized values fixed for regime 2 : gamma = 8.615499 , th = 3.36229 ; SSE = 4.312895

Testing for addition of regime 3.
Estimating gradient matrix...
Done. Computing the test statistic...
Done. Regime 3 is NOT accepted (p-value = 0.9867115 ).

Finished building a MRSTAR with 2 regimes

Non linear autoregressive model

Multiple regime STAR model

Regime 1 :
  Linear parameters: 0.454201, 1.2502866, -0.3544742

Regime 2 :
  Linear parameters: -1.6984602, 0.4417287, -0.1036671
  Non-linear parameters:
  8.6154991, 3.3622897

```

Види се да је p вредност јако мала и једнака ~ 0.000954 . Серија је нелинеарна и гради се *STAR* модел реда 2. Оптимизациони алгоритам конвергира. Могу се видети вредности оцењених параметара модела; $\gamma = 8.615$, $th = 3.362$ и $SSE = 4.313$. Тестира се да ли је потребно додати трећи режим и на основу p вредности, која је ~ 1 , прихвата се нулта хипотеза о дворежимном *STAR* моделу. Функција *summary* примењена на генерисаном *STAR* моделу не пружа нове информације, па њен излаз неће бити приказан.

За остале позиве функција, због практичности, приказан је само део излаза.

```

> (star1 <- star(x.train, m = 2, thDelay = 0, control = control, starting.control = controls))
Testing linearity... p-value = 0.004572888
The series is nonlinear. Incremental building procedure:
Building a 2 regime STAR.
Performing grid search for starting values...
Starting values fixed: gamma = 100 , th = 2.57116 ; SSE = 4.418747
Grid search selected lower/upper bound gamma (was: 1 100 ]).
                                Might try to widen bound with arg: 'starting.control'
=list(gammaInt=c(1,200))'
Optimization algorithm converged
Optimized values fixed for regime 2 : gamma = 100.0001 , th = 2.568856 ; SSE = 4.417719

Testing for addition of regime 3.
Estimating gradient matrix...
Done. Computing the test statistic...
Done. Regime 3 is NOT accepted (p-value = 0.9633968 ).

Finished building a MRSTAR with 2 regimes

```

На основу излаза функције, види се да серија одговара *STAR* моделу са два режима са p вредности једнакој 0.00457. Ова p вредност је већа од p вредности модела *star0*. Вредност параметра γ сада је много већа и износи $\gamma = 100$. Вредност параметра th је мања и износи $th = 2.569$. Вредност за *SSE* је за нијансу већа, што би значило да је *star0* бољи избор од модела *star1*.

У следећем кораку се испитује квалитет *STAR* модела већег реда и променљивом прага $thDelay = 1$, односно y_{t-2} .

```
> (star2 <- star(x.train, m = 3, thDelay = 1, control = control, starting.control = controls))
Testing linearity... p-value = 0.0003822478
The series is nonlinear. Incremental building procedure:
Building a 2 regime STAR.
Performing grid search for starting values...
Starting values fixed: gamma = 6.076923 , th = 3.540261 ; SSE = 3.983377
Convergence problem code 1. You might want to increase maximum number of iterations by setting
'control=list(maxit=1000)'
Optimized values fixed for regime 2 : gamma = 5.359141 , th = 3.746337 ; SSE = 3.97443

Testing for addition of regime 3.
Estimating gradient matrix...
Done. Computing the test statistic...
Done. Regime 3 is NOT accepted (p-value = 0.7044974 ).

Finished building a MRSTAR with 2 regimes
```

За *STAR* модел реда 3 мања је вредност за *SSE* ~4. Такође, p вредност је најмања од приказаних и износи 0.00038. Ово је добар показатељ, али је ред овог модела за 1 већи у сваком од режима.

Сада ће се приказати разлике које се могу добити само променом променљиве која утиче на промену прага.

```
> (star3 <- star(x.train, m = 4, thDelay = 1, control = control, starting.control = controls))
Testing linearity... p-value = 0.003944386
The series is nonlinear. Incremental building procedure:
Building a 2 regime STAR.
Performing grid search for starting values...
Starting values fixed: gamma = 6.076923 , th = 3.579166 ; SSE = 3.836354
Optimization algorithm converged
Optimized values fixed for regime 2 : gamma = 4.775792 , th = 3.936964 ; SSE = 3.814912

> (star4 <- star(x.train, m = 4, thDelay = 0, control = control, starting.control = controls))
Testing linearity... p-value = 0.08782714
The series is linear. Using linear model instead.

Non linear autoregressive model

AR model
Coefficients:
      const      phi.1      phi.2      phi.3      phi.4
1.4545374  1.2573117 -0.6945811  0.1570410 -0.2234118
```

Дакле, за исти ред модела и исте почетне вредности, али различите променљиве прага, добија се да је у једном случају серију боље описати *STAR* моделом, а у другом случају да је адекватнији линеарни модел истог реда. На овом примеру види се значај променљиве прага на изградњу модела са променом режима.

Надаље се користити модел *star_0* који је преименован у *model_star*.

Резултат теста за тестирање нормалности резидуала приказан је испод.

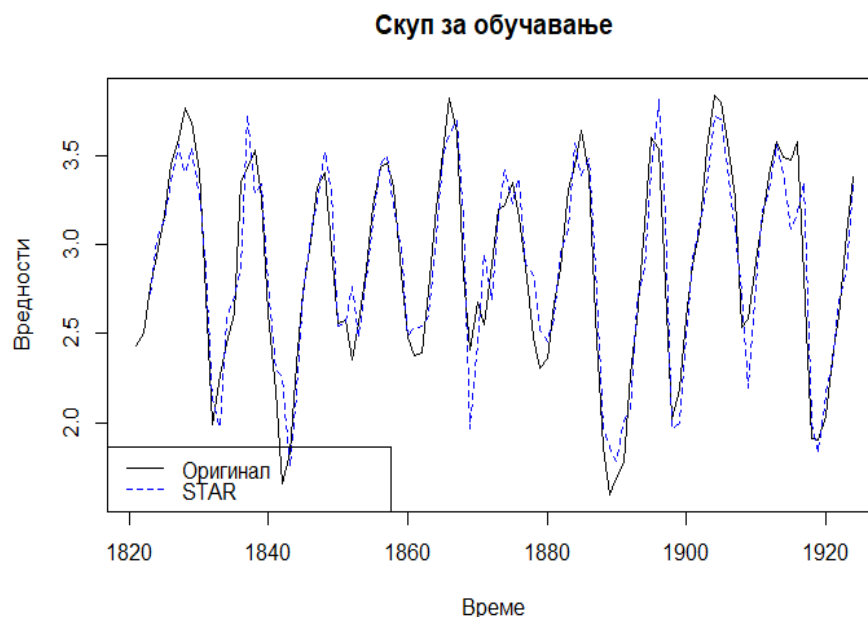
```
> e_star <- residuals(model_star)
> e_star <- e_star[!is.na(e_star)]
> jarque.bera.test(e_star)

Jarque Bera Test

data:  e_star
X-squared = 1.0265, df = 2, p-value = 0.5985
```

На основу p вредности теста види се да су резидуали нормално расподељени.

На слици 11.7 приказане су вредности посматране временске серије и вредности добијене генерисаним *STAR* моделом.



Слика 11.7 Приказ оригиналних вредности временске серије и вредности добијене генерисаним *STAR* моделом

На основу приказне слике, види се да се *STAR* модел добро уклопио у податке скупа за обучавање.

Остало је одредити *LSTAR* модел. Наглашава се да је у програмском језику *R* могуће генерисати само *LSTAR* модел са два режима, иако у овом примеру то не представља ограничење. За разлику од *STAR* модела, постоји функција *selectLSTAR* која најбоље моделе одређује на основу стандардног *AIC*-а, па се она не користи, већ се „ручно“ одређује најбољи могући модел. Процедура је иста као и код претходних модела. Функција *summary* је овде од значаја. Пошто је њен излаз обимнији, у потпуности се приказује само за први модел. За остале моделе наводи се само део који је од већег интереса.

```

> lstar1 <- lstar(x.train, m = 2, thDelay = 1, control = control, starting.control = controls)
Using maximum autoregressive order for low regime: mL = 2
Using maximum autoregressive order for high regime: mH = 2
Performing grid search for starting values...
Starting values fixed: gamma = 11.08081 , th = 3.339366 ; SSE = 4.315477
Optimization algorithm converged
Optimized values fixed for regime 2 : gamma = 9.291303 , th = 3.353325 ; SSE = 4.313023
> summary(lstar1)

Non linear autoregressive model

LSTAR model
Coefficients:
Low regime:
  const.L   phiL.1   phiL.2
  0.467266  1.249901 -0.360018

High regime:
  const.H   phiH.1   phiH.2
 -1.4792186  0.4315208 -0.1481375

Smoothing parameter: gamma = 9.291

Threshold
Variable: Z(t) = + (0) X(t) + (1) X(t-1)

value: 3.353

Residuals:
      Min       1Q   Median       3Q      Max
-0.596351 -0.123669  0.019384  0.125127  0.507841

Fit:
residuals variance = 0.04147,  AIC = -315, MAPE = 6.006%

Coefficient(s):
      Estimate Std. Error t value Pr(>|z|)
const.L  0.467266   0.232684   2.0082  0.04463 *
phiL.1   1.249901   0.072744  17.1822 < 2e-16 ***
phiL.2  -0.360018   0.115814   -3.1086  0.00188 **
const.H -1.479219   3.261176   -0.4536  0.65013
phiH.1   0.431521   0.211212   2.0431  0.04105 *
phiH.2  -0.148138   0.755343   -0.1961  0.84452
gamma    9.291303   8.483198   1.0953  0.27340
th       3.353325   0.135842  24.6855 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Non-linearity test of full-order LSTAR model against full-order AR model
F = 10.53 ; p-value = 7.2676e-05

Threshold
Variable: Z(t) = + (0) X(t) + (1) X(t-1)

```

Функција *summary* даје оцене коефицијента оба режима, параметра γ који је једнак $\gamma = 9.291$, параметра th који је једнак $th = 3.353$. Оцена за SSE једнака је $SSE = 4.313$. Може се видети и променљива прага која је коришћена, $thDelay = 1$. Коефицијенти нижег режима су значајни, док је само један коефицијент вишег режима значајан. Параметар th је такође значајан, док је параметар γ безначајан, што не изненађује. На основу приказане p вредности тест статистике, која је веома мала, закључује се да је *LSTAR* модел бољи избор од линеарног *AR* модела.

Приказани модел има слободне чланове и он је бољи од модела који се добија искључивањем слободних чланова, па се из тог разлога приказују следећи резултати.

```
> lstar2 <- lstar(x.train, m = 2, thDelay = 0, control = control, starting.control = controls)
Using maximum autoregressive order for low regime: mL = 2
Using maximum autoregressive order for high regime: mH = 2
Performing grid search for starting values...
Starting values fixed: gamma = 500 , th = 2.562035 ; SSE = 4.378012
Grid search selected lower/upper bound gamma (was: 1 500 ]).
                                Might try to widen bound with arg: 'starting.control=
list(gammaInt=c(1,200))'
Optimization algorithm converged
Optimized values fixed for regime 2 : gamma = 500 , th = 2.563439 ; SSE = 4.377729
> summary(lstar2)
```

Coefficient(s):					
	Estimate	Std. Error	t value	Pr(> z)	
const.L	0.401630	0.289653	1.3866	0.1655670	
phiL.1	1.246160	0.147640	8.4405	< 2.2e-16	***
phiL.2	-0.332369	0.099592	-3.3373	0.0008459	***
const.H	0.722996	0.359675	2.0101	0.0444167	*
phiH.1	0.325423	0.172148	1.8904	0.0587092	.
phiH.2	-0.633981	0.122951	-5.1564	2.518e-07	***
gamma	500.000002	1331.927074	0.3754	0.7073660	
th	2.563439	0.021522	119.1051	< 2.2e-16	***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Non-linearity test of full-order LSTAR model against full-order AR model
F = 7.3549 ; p-value = 0.0010617

Threshold
Variable: z(t) = + (1) x(t) + (0) x(t-1)
```

Као што се може видети, сви параметри, осим параметра γ , јесу значајни, што је и очекивано, иако он сад има већу вредност, $\gamma = 500$. Мала је p вредност теста, али је већа од претходно приказане p вредности. Вредност за SSE је за нијансу већа од вредности SSE претходног модела.

```

> lstar3 <- lstar(x.train, m = 1, include = "none", control = control, starting.control = controls)
Using maximum autoregressive order for low regime: mL = 1
Using maximum autoregressive order for high regime: mH = 1
Using default threshold variable: thdelay=0
Performing grid search for starting values...
Starting values fixed: gamma = 500 , th = 3.239588 ; SSE = 12.10419
Grid search selected lower/upper bound gamma (was: 1 500 ]).
                                Might try to widen bound with arg: 'starting.control=list
(gammaInt=c(1,200))'
Optimization algorithm converged
Optimized values fixed for regime 2 : gamma = 500 , th = 3.236998 ; SSE = 12.10419
> summary(lstar3)
Coefficient(s):
              Estimate Std. Error t value Pr(>|z|)
phiL.1  1.0395e+00  1.5857e-02  65.5548 < 2.2e-16 ***
phiH.1 -9.3274e-02  2.2996e-02  -4.0561 4.991e-05 ***
gamma   5.0000e+02  1.0285e+04   0.0486  0.9612
th      3.2370e+00  4.0463e-01   7.9999 1.332e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Non-linearity test of full-order LSTAR model against full-order AR model
F = 0.64902 ; p-value = 0.42237

Threshold
Variable: z(t) = + (1) x(t)

```

Прво што се примети јесте вредност за SSE која је једнака $SSE = 12.104$, што се значајно разликује од оцењених вредности за SSE претходно приказаних модела. Коефицијенти су значајни, али је p вредност већа и износи 0.4224, што иде у корист линеарном моделу.

На основу приказаног, бира се модел *lstar2*, преименован у *model_lstar* и тестира нормалност његових резидуала.

```

> model_lstar <- lstar2
> e_lstar <- residuals(model_lstar)
> e_lstar <- e_lstar[!is.na(e_lstar)]
> jarque.bera.test(e_lstar)

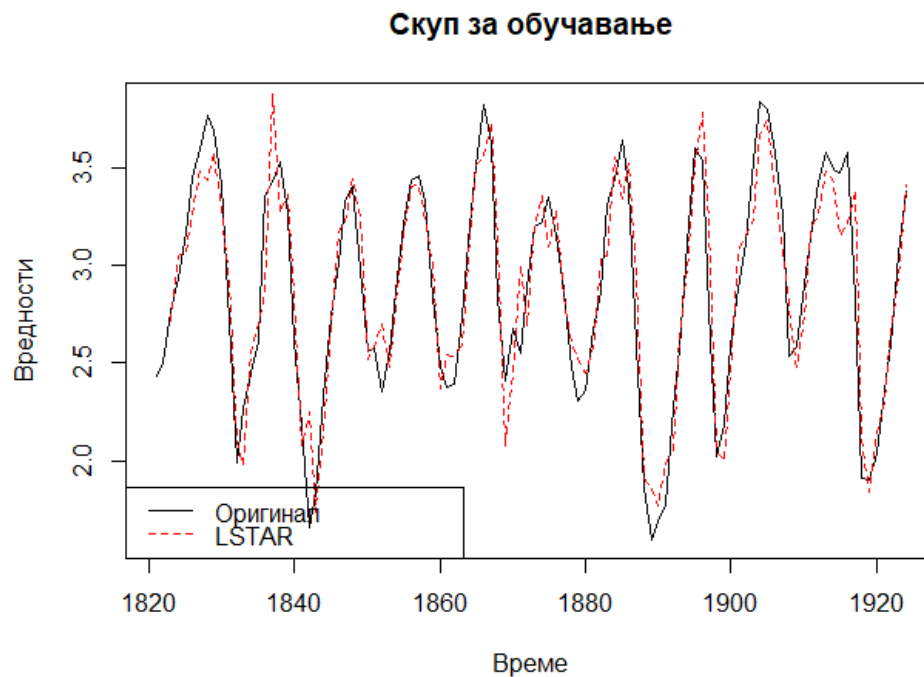
      Jarque Bera Test

data:  e_lstar
X-squared = 3.5112, df = 2, p-value = 0.1728

```

Види се да је p вредност теста већа од 0.1 па се прихвата хипотезу о нормалности резидуала.

На слици 11.8 приказане су вредности посматране временске серије и вредности добијене генерисаним *LSTAR* моделом.



Слика 11.8 Приказ оригиналних вредности временске серије и вредности добијене генерисаним *LSTAR* моделом

На основу приказаног, види се да се и *LSTAR* модел добро уклопио у податке скупа за обучавање.

Сада се генерише *AR* модел реда 2 како би се добијена предвиђања нелинеарних модела упоредила са предвиђањима линеарног модела.

```
> model_ar <- Arima(x.train, order=c(2,0,0))
```

Приказан модел се, као и претходни модели, добро уклапа у скуп за обучавање и има нормално расподељене резидуале.

У наставку су приказани графици оригиналних вредности и предвиђених вредности сваког модела. Користе се две методе за добијање предвиђених вредности нелинеарних модела; *bootstrap* метода и метода Монте Карла, које се бирају навођењем параметра *type*. Број репликација за обе методе дефинише се параметром *n.boot*. За добијање интервала поузданости користи се параметар *ci*.

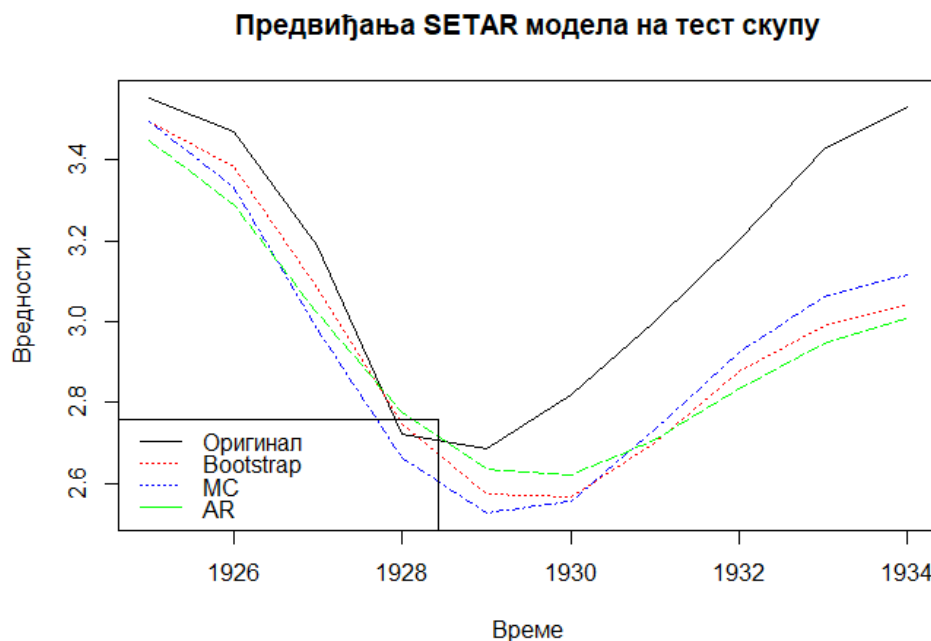
```

# SETAR
pred_setar_boot <- predict(model_setar, n.ahead=10, type = "bootstrap",
                           n.boot=200, ci=0.95)
pred_setar_MC   <- predict(model_setar, n.ahead=10, type = "MC",
                           n.boot=200, ci=0.95)
# STAR
pred_star_boot  <- predict(model_star, n.ahead=10, type = "bootstrap",
                           n.boot=200, ci=0.95)
pred_star_MC    <- predict(model_star, n.ahead=10, type = "MC",
                           n.boot=200, ci=0.95)
# LSTAR
pred_lstar_boot <- predict(model_lstar, n.ahead=10, type = "bootstrap",
                           n.boot=200, ci=0.95)
pred_lstar_MC   <- predict(model_lstar, n.ahead=10, type = "MC",
                           n.boot=200, ci=0.95)
# AR
pred_AR <- predict(AR, n.ahead = 10)

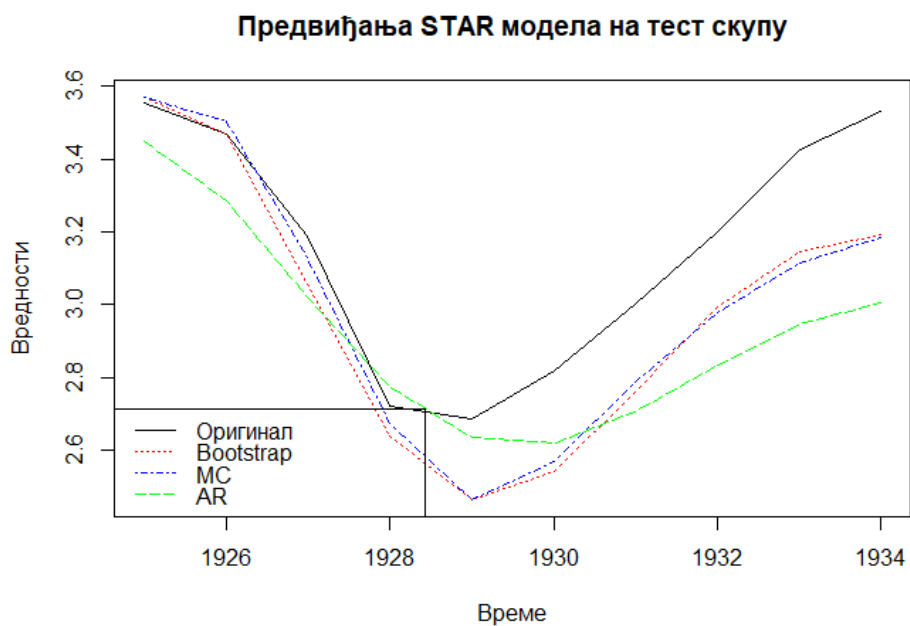
```

Функција *predict* враћа предвиђене вредности добијене на основу прослеђене методе и стандардне грешке за добијање 95% интервала поверења.

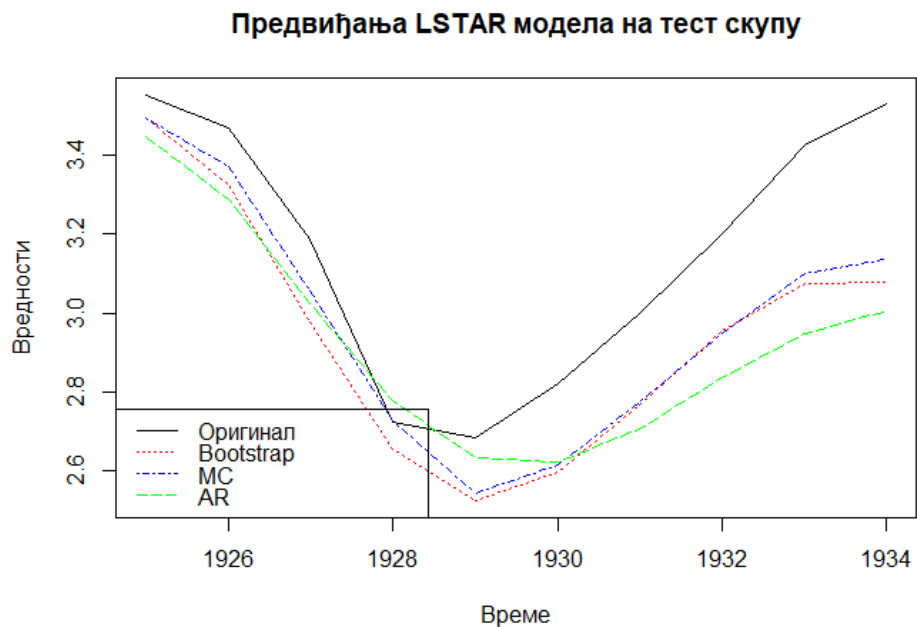
На слици 11.9 приказани су графици предвиђања нелинеарних модела и на сваком од њих предвиђања линеарног модела.



Слика 11.9 а) Предвиђања SETAR модела добијена *bootstrap* методом и методом Монте Карла и предвиђања линеарног AR модела



Слика 11.9 б) Предвиђања STAR модела добијена *bootstrap* методом и методом Монте Карла и предвиђања линеарног AR модела



Слика 11.9 в) Предвиђања LSTAR модела добијена *bootstrap* методом и методом Монте Карла и предвиђања линеарног AR модела

Са слике 11.9 види се да су предвиђања линеарног модела слична предвиђањима нелинеарних модела када је мањи корак предвиђања у питању. Међутим, за предвиђање више корака унапред чини се да је нелинеарни модел ипак бољи.

Да би овако нешто било оправдано тврдити, примениће се метрике тачности које су сачуване у бази *baza_acc*.

```
> baza_acc
```

	ME	RMSE	MAE	MPE	MAPE	MASE
setar_boot	0.2396691	0.2641229	0.2396691	7.648940	7.648940	1.2380414
setar_MC	0.1873211	0.2255138	0.1885991	5.833947	5.880875	1.1335462
star_boot	0.1407981	0.1630480	0.1443831	4.612119	4.713016	0.6809603
star_MC	0.1899603	0.2296284	0.1935453	6.053665	6.154561	1.0352665
lstar_boot	0.2113573	0.2407356	0.2113573	6.625825	6.625825	1.2246400
lstar_MC	0.1969781	0.2496397	0.2068407	6.079491	6.441624	1.3369473
AR	0.1969781	0.2496397	0.2068407	6.079491	6.441624	1.7966732

Користећи функцију *which* на једноставан начин може се видети која су предвиђања најбоља.

```
> ( ME_min <- which(baza_acc[,1] == min(baza_acc[,1], na.rm = TRUE)) )
[1] 3
> ( RMSE_min<- which(baza_acc[,2] == min(baza_acc[,2], na.rm = TRUE)) )
[1] 3
> ( MAE_min <- which(baza_acc[,3] == min(baza_acc[,3], na.rm = TRUE)) )
[1] 3
> ( MPE_min <- which(baza_acc[,4] == min(baza_acc[,4], na.rm = TRUE)) )
[1] 3
> ( MAPE_min<- which(baza_acc[,5] == min(baza_acc[,5], na.rm = TRUE)) )
[1] 3
> ( MASE_min<- which(baza_acc[,6] == min(baza_acc[,6], na.rm = TRUE)) )
[1] 3
```

Предвиђања *STAR* модела добијена методом *bootstrap* најмање одступају од оригиналних вредности скупа за тестирање приказане временске серије.

Сада ће бити приказани резултати добијени *DM* тест статистиком. Нулта хипотеза је да две методе имају исту тачност предвиђања. Алтернативна хипотеза се може формулисати на више начина. Навођењем *alternative = "less"* она гласи „Прва метода је тачнија од друге“. Да би се применила *DM* тест статистика, потребно је одредити грешке предвиђања. Начин на који се оне добијају приказан је за једну методу предвиђања. За остале методе предвиђања оне се могу добити аналогно.

```
> error_setar_boot <- as.vector(pred_setar_boot$pred- x.test)
```

У наставку су приказани резултати *dm.test*-а. Због начина дефинисања алтернативне хипотезе, *p* вредности мање од 0.05 говоре да је прва метода боља од друге. У супротном се прихвата нулта хипотеза да оне имају исту тачност предвиђања.


```
> dm.test(error_setar_boot, error_AR, alternative = "less")

Diebold-Mariano Test

data: error_setar_booterror_AR
DM = -0.69775, Forecast horizon = 1, Loss function power = 2, p-value =
0.2515
alternative hypothesis: less
```

Због p вредности која је једнака 0.2515 прихватамо хипотезу да ове методе имају исту тачност. Сада се наводи *alternative = "greater"* за дефинише алтернативне хипотезе која гласи „Друга метода је тачнија од прве“.

```
> dm.test(error_setar_boot, error_AR, alternative = "greater")

Diebold-Mariano Test

data: error_setar_booterror_AR
DM = -0.69775, Forecast horizon = 1, Loss function power = 2, p-value =
0.7485
alternative hypothesis: greater
```

Како је p вредност велика, прихвата се хипотеза да тестиране две методе имају исту тачност предвиђања.

Због једноставности у наставку ће бити приказани само позиви *dm.test*-ова, одговарајуће p вредности тест статистике са назнаком која је метода боља.

```
#SETAR
dm.test(error_setar_boot, error_AR, alternative = "less") # 0.2515 # iste
dm.test(error_setar_MC , error_AR, alternative = "less") # 0.02532 # setar_MC

#STAR
dm.test(error_star_boot , error_AR, alternative = "less") # 0.03821 # star_boot
dm.test(error_star_MC , error_AR, alternative = "less") # 0.03382 # star_MC

#LSTAR
dm.test(error_lstar_boot, error_AR, alternative = "less") # 0.06204 # iste
dm.test(error_lstar_MC , error_AR, alternative = "less") # 0.02679 # lstar_MC
```

Види се да предвиђања које даје *setar_boot* нису ништа тачнија од предвиђања *AR* модела. Исто важи и за *lstar_boot*.

```
# boot i MC metoda
dm.test(error_setar_boot, error_setar_MC, alternative = "less") # 0.9825 # iste
dm.test(error_star_boot , error_star_MC , alternative = "less") # 0.04834 # star_boot
dm.test(error_lstar_boot, error_lstar_MC, alternative = "less") # 0.275 # iste
```

Поређењем предвиђања добијена *bootstrap* методом и методом Монте Карла, види се да се она разликују једино код *STAR* модела, при чему су предвиђања добијена *bootstrap* методом боља.

```
# SETAR i STAR/LSTAR
dm.test(error_setar_MC, error_star_boot, alternative = "less") # 0.9357 # iste
dm.test(error_setar_MC, error_star_MC, alternative = "less") # 0.3428 # iste
dm.test(error_setar_MC, error_lstar_MC, alternative = "less") # 0.03337 # setar_MC

# STAR i LSTAR
dm.test(error_star_boot, error_lstar_MC, alternative = "less") # 0.0509 # iste
dm.test(error_star_MC, error_lstar_MC, alternative = "less") # 0.1021 # iste
```

Сада се показује да су предвиђања *SETAR* модела добијена *bootstrap* методом боља од предвиђања *LSTAR* модела добијена истом методом.

Применом *dm.test*-а за различите комбинације метода предвиђања добија се да већина метода предвиђања модела са променом режима јесте тачнија од предвиђања које даје линеарни *AR* модел. Такође, предвиђања нелинеарних модела се међу собом малом разликују. С друге стране, све метрике тачности иду у корист *STAR* модела са предвиђањима добијених *bootstrap* методом. У случају да је потребан модел са променом режима са глатким прелазом између режима, био би изабран овај модел. Ако је потребан модел са дискретном променом режима, добијени *SETAR* модел даће довољно добре резултате.

12. Закључак

У овом раду су детаљно описани модели са променом режима код којих је променљива која утиче на промену режима опажена. Највећа пажња посвећена је *SETAR* и *STAR* моделима. Представљено је како се оцењују параметри ових модела, на који начин се тестира да ли је изабрани модел са променом режима бољи избор од референтног линеарног модела, који се информациони критеријуми користе приликом избора модела на скупу за обучавање. Приказана је и тест статистика којом се може тестирати нормалност резидуала као и начин на који се тестира преостала нелинеарност као вид евалуације модела.

Описано је како се добијају тачкаста предвиђања на скупу за тестирање и како се могу добити интервали поверења тих предвиђања. Представљене су метрике које се користе за евалуацију предвиђања.

Наведено је које претпоставке морају бити испуњене, шта може бити узрок лоших предвиђања и колики је утицај аутлајера.

На примерима је приказано колико може бити захтевно радити са моделима са променом режима када се они користе за опис реалних података, јер у том случају многе претпоставке не морају бити испуњене. За рад са реалним подацима потребно је укључити робусне оцене и тестове за тестирање хетероскедастичности резидуала. Било би корисно обезбедити методе које могу да предвиде у ком ће се режиму наћи модел на скупу за тестирање и на основу тога дати оцену предвиђања. Такође, Марковљев модел пребацивања се показује као значајан у пракси, па би корисно било проучити детаљније његове особине и начин на који га је могуће имплементирати. Међутим, да би се модели са променом режима користили у пракси, потребно је укључити, на пример, софтверски програм *GAUSS* који има велики број имплементираних функција за рад са овим моделима, за чије је разумевање пружена добра теоријска подлога.

Референце

- [1] Philip Hans Franses and Dick van Dijk. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press 2003.
- [2] Shouyi Wang and Wanpracha Art Chaovalitwongse. *Evaluating and Comparing Forecasting Models*. 2011 . DOI: 10.1002/9780470400531.eorms0307
- [3] Adnan M. Awad. *Properties of the Akaike information criterion*. Department of Mathematics, University of Jordan, Amman, Jordan 1995.
- [4] Francesco Giordano, Marcella Niglio and Cosimo Damiano Vitale. *On the stationarity of the Threshold Autoregressive process: the two regimes case*. Events and meetings of Italian statistical society, 2012.
- [5] Thompson, B. *The influence of aberrant values on the statistics related to a selection program*. Theoret. Appl. Genetics 52, 201–207 (1978).
- [6] https://en.wikipedia.org/wiki/Robust_statistics
- [7] Zacharias Psaradakis, Marián Vávra. *Portmanteau Tests for Linearity of Stationary Time Series*. Working paper National Bank of Slovakia 2016.
- [8] Bruce E. Hansen , University of Wisconsin. *Testing for linearity*. Blackwell Publishers Ltd. 1999, 108 Cowley Rd., Oxford OX4 1JF, UK and 350 Main St., Malden, MA 02148, USA
- [9] Rob J. Hyndman Monash University (Australia), *Another look at forecast-accuracy metrics for intermittent demand* , 2014
- [10] Francis X. Diebold and Roberto S. Mariano, University of Pennsylvania, *Comparing predictive accuracy I: an asymptotic test*, Article in Journal of Business and Economic Statistics 2002.
- [11] <https://cran.r-project.org/web/packages/tsDyn/tsDyn.pdf>
- [12] Wai-Sum Chan, Albert C. S. Wong and Howell Tong, *Some Nonlinear Threshold Autoregressive Time Series Models for Actuarial Use*, Article in North American Actuarial Journal, 2004.
- [13] <https://www.investopedia.com/>
- [14] https://drive.google.com/drive/folders/1zMYloesmpP5SGPh_G3wY_9vP2xQIC21iM?usp=sharing
- [15] Gleb P. Severov editor, *International Finance and Monetary Policy* , Nova Science Publisher, inc. New York, 2006.
- [16] Panagiotis Mantalos , *Three Different Measures of Sample Skewness and Kurtosis and their Effects on the Jarque-Bera Test for Normality*, Jönköping International Business School Jönköping University JIBS Working Papers No. 2010-9
- [17] Richard Luger , *Exact Tests of Equal Forecast Accuracy with an Application to the Term Structure of Interest Rates*, Monetary and Financial Analysis Department Bank of Canada Ottawa, Ontario, Canada K1A 0G9, 2004.

[18] Eric Zivot and Jiahui Wang , *Modeling financial time series with S-PLUS®* Second Edition, Springer Science + Business Media, Inc., New York, USA 2006.

[19] James D. Hamilton. *Time Series Analysis*. Princeton University Press, New Yersey 1994.

[20] <http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/vignettes/tsDyn/tsDyn.pdf>