

УНИВЕРЗИТЕТ У БЕОГРАДУ

МАТЕМАТИЧКИ ФАКУЛТЕТ



Ана Јовичић

# Примена бутстреп метода у анализи временских серија

мастер рад

Београд, 2023

## **Ментор:**

др Бојана Милошевић, ванредни професор  
Универзитет у Београду, Математички факултет

## **Чланови комисије:**

др Марко Обрадовић, доцент  
Универзитет у Београду, Математички факултет

др Марија Цупарић, доцент  
Универзитет у Београду, Математички факултет

## Наслов мастер рада:

Примена бутстреп метода у анализи временских серија

## Резиме:

Развој рачунарске технике средином XX века допринео је процвату многих научних дисциплина, с обзиром на то да су створени услови за решавање великог броја проблема који су захтевали интензивна израчунавања. Међу наукама које почињу нагло да се развијају наша се и математичка статистика. Нагли раст рачунарске моћи омогућио је лакша и бржа испитивања великих скупова података сложене структуре, али и развој нових статистичких техника који обезбеђују добре резултате за узорке малог обима.

У овом раду биће описан бутстреп метод, један од савремених алата статистичке анализе и примена ове методе у анализи временских серија. У другом поглављу дат је опис тзв. стандардног бутстрепа, тј. бутстреп методе за независне скупове података, као и примена ове методе за оцену средње вредности, дисперзије и интервала поверења. У трећем поглављу дата је основна анализа временских серија и модела временских серија. Четврто поглавље описује бутстреп методе за временске серије, као што су блок и решеткасти бутстреп и наводи теореме којима је показана постојаност ових метода. Такође, у овом поглављу упоређене су перформансе ових метода за оцену средње вредности, параметара модела и аутокорелационих коефицијената. Поглавље пет садржи примену бутстреп метода са скуповима реалних података.

## Кључне речи:

бутстреп метод, реузорковање, временске серије, зависност података

## Захвалница

Израдом овог рада, након две године паузе, завршавам своје студије на Математичком факултету. Писање овог рада вратило ми је интересовање за статистику и изучавање њене примене у разним областима. Иако су испред мене велики задаци у даљем раду, сматрам да је оно што сам до сада постигла на студијама успех, па бих се захвалила свима који су том успеху допринели.

Највећу захвалност дугујем својој менторки, проф. др Бојани Милошевић, која ми је била велика подршка током читаве израде мастер рада. Њено несебично залагање и бројне сугестије значајно су допринели квалитету овог рада. Такође, захвална сам јој и на подршци током основних студија, њено знање, ангажовање и љубав према математици дали су ми мотивацију за даљи рад.

Захвалила бих се и осталим члановима комисије на издвојеном времену за читање овог рада и корисним сугестијама.

На крају, захвалила бих се својој породици и пријатељима на стрпљењу и подршци током студија.

У Београду,  
Августа 2023. године,  
Ана Јовичић.

# Садржај

<b>1</b>	<b>Увод</b>	<b>1</b>
<b>2</b>	<b>Стандардни бутстреј</b>	<b>2</b>
2.1	Параметарски бутстреј . . . . .	3
2.2	Непараметарски бутстреј . . . . .	3
2.3	Примена бутстреј методе за оцену средине и дисперзије . . . . .	4
2.4	Примена бутстреј методе за оцену интервала поверења . . . . .	5
<b>3</b>	<b>Временске серије</b>	<b>7</b>
3.1	Дефиниција, стационарност и ергодичност временских серија . . . . .	7
3.2	Компоненте временске серије . . . . .	9
3.3	Декомпозиција временских серија . . . . .	10
3.4	Аутоковаријациона и аутокорељациона функција . . . . .	12
3.5	Модели стационарних временских серија . . . . .	14
3.5.1	Линеарни процес . . . . .	14
3.5.2	Ауторегресиони модели . . . . .	16
3.5.3	Модели покретних просека . . . . .	20
3.5.4	Ауторегресиони модели покретних просека . . . . .	25
3.5.5	$\alpha$ - мешајући процес . . . . .	28
<b>4</b>	<b>Бутстреј методе за временске серије</b>	<b>29</b>
4.1	Блок бутстреј . . . . .	29
4.1.1	Блок бутстреј без преклапања . . . . .	30
4.1.2	Покретни блок бутстреј . . . . .	30
4.1.3	Кружни блок бутстреј . . . . .	33
4.1.4	Стационарни бутстреј . . . . .	34
4.1.5	Симулација метода . . . . .	35
4.2	Решеткасти бутстреј . . . . .	49
<b>5</b>	<b>Примери са скуповима реалних података</b>	<b>53</b>
5.1	Пример 1: Оцена стандардне девијације грешака прогнозе . . . . .	54
5.2	Пример 2: Побољшање тачности прогнозе . . . . .	59
<b>6</b>	<b>Закључак</b>	<b>68</b>
<b>7</b>	<b>Додатак 1</b>	<b>70</b>
<b>8</b>	<b>Додатак 2</b>	<b>73</b>



## 1 Увод

Принцип бутстреп методологије увео је амерички статистичар Бредли Ефрон <sup>1</sup> 1979. године у раду "*Bootstrap methods: Another look at the Jackknife*" [2], који је објављен у часопису "*The Annals of Statistics*". Назив бутстреп потиче од енглеског израза "*to pull yourself up by your own bootstraps*", што у пренесеном смислу значи, постићи успех без ослањања на помоћ споља.

Бутстреп је једноставан, али моћан статистички метод, који је Ефрон представио први пут 1977. године, међутим и након објаве рада 1979. године, научници га нису често користили због недовољно развијених рачунарских система, који тада нису били у стању да изводе сложене и дуготрајне процесе, као и због лаичког објашњења саме методе и недовољно добрих доказа о асимптотским својствима која се јављају у бутстрепу. Неколико година касније, са појавом темељних доказа, бутстреп постаје битан алат у анализи података и регуларна метода реузорковања података.

**Реузорковање података (енгл. resampling)** подразумева генерисање већег скупа података на основу датог оригиналног скупа података. Реализовани узорак из почетне популације посматрамо као нову популацију из које правимо велики број нових узорака (реузорака). Добијени реузорци користе се за апроксимацију расподеле полазног узорака, као и за оцену параметара. Предност овакве апроксимације је што не треба правити никакве претпоставке о расподели на основном узорку, довољан услов је да је основни узорак популације репрезентативан, јер уколико у њему постоји грешка она ће се у даљем раду акумулирати и добићемо погрешне резултате.

Овакав принцип је посебно погодан у случајевима када је оригиналан узорак малих димензија, па самим тим асимптотска теорија (Закон великих бројева и Централно гранична теорема) не дају поуздане податке апроксимације расподеле.

---

<sup>1</sup>Bradley Efron, 1938.

## 2 Стандардни бутстреп

Бутстреп метода је метода реузорковања која омогућава једноставну оцену расподеле статистике од интереса без претпоставки о типу расподеле датог узорка и не ослања се на асимптотске резултате. Упознајмо се са принципом функционисања бутстреп методе.

Нека је  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  прост случајан узорак популације са функцијом расподеле  $F(\theta)$  и нека је  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  реализован узорак генерисан функцијом  $F$ . Основни циљ бутстреп методе јесте оцена расподеле статистике којом оцењујемо параметре расподеле.

Нека је  $Z = f(\mathbf{X})$  статистика узорка  $\mathbf{X}$  којом оцењујемо параметар  $\theta = \theta(F)$  расподеле  $F$  на основу простог случајног узорка  $\mathbf{X}$ , коју ћемо означити са  $\hat{\theta} = \hat{\theta}(\mathbf{X})$ . Задатак је одредити функцију расподеле  $H$  статистике  $Z$ .

У зависности да ли је расподела  $F$  позната разликујемо два случаја:

1. **параметарски бутстреп** - позната је функција расподеле  $F$ , али у њој фигурише непознати параметар  $\theta$ ,
2. **непараметарски бутстреп** - не постоје никакве претпоставке о функцији расподеле  $F$ .

У оба случаја из датог реализованог узорка  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  генеришемо  $R$  реузорака  $\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_R^*$ <sup>2</sup> који су истог обима као и полазни узорак. Добијене реузорке можемо представити на следећи начин:

$$\begin{aligned} \mathbf{x}_1^* &= (x_{11}^*, x_{21}^*, \dots, x_{n1}^*), \\ \mathbf{x}_2^* &= (x_{12}^*, x_{22}^*, \dots, x_{n2}^*), \\ &\vdots \\ \mathbf{x}_R^* &= (x_{1R}^*, x_{2R}^*, \dots, x_{nR}^*). \end{aligned}$$

На основу добијених бутстреп реузорака  $\mathbf{x}^* = (\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_R^*)$  рачунамо вредности статистике  $Z$ , тј.

$$z_i = f(\mathbf{x}_i), \mathbf{z} = (z_1, z_2, \dots, z_R).$$

На основу добијених вредности  $\mathbf{z} = (z_1, z_2, \dots, z_R)$  генеришемо емпиријску функцију расподеле, којом апроксимирамо праву расподелу статистике  $Z$ :

$$H(z|\mathbf{Z} = \mathbf{z}) = \frac{1}{R} \sum_{i=1}^R I\{z_i \leq z\}, z \in \mathbb{R}.$$

---

<sup>2\*</sup> означава бутстреп реализације



## 2.1 Параметарски бутстреп

У случају параметарског бутстрепа претпостављамо да је познат облик функције расподеле  $F$  и да та функција зависи од непознатог параметра  $\theta$ . Као што је већ поменуто, потребно је наћи функцију расподеле неке оцене  $\hat{\theta}$  параметра  $\theta$ . Поступак је следећи:

1. на основу датог реализованог узорка  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  рачунамо вредност оцене  $\hat{\theta}$  параметра  $\theta$ ,
2. функцију  $F_\theta$  апроксимирамо функцијом  $F_{\hat{\theta}}$ ,
3. генеришемо из  $F_{\hat{\theta}}$   $R$  независних узорака  $\mathbf{x}_i^* = (x_{1i}^*, x_{2i}^*, \dots, x_{ni}^*)$ ,  $i = 1 \dots R$ , које називамо бутстреп узорцима,
4. за сваки бутстреп узорак рачунамо вредност оцене  $\hat{\theta}_i^* = \hat{\theta}(\mathbf{x}_i^*)$ , добијене вредности зовемо бутстреп оцене посматране статистике,
5. одређујемо емпиријску функцију расподеле на основу бутстреп оцена добијених у кораку 4.

## 2.2 Непараметарски бутстреп

У непараметарском бутстрепу полазимо од претпоставке да немамо никакве информације о функцији расподеле  $F$ . Поступак је следећи:

1. на основу датог реализованог узорка  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  рачунамо емпиријску функцију расподеле

$$F_n(x|\mathbf{X} = \mathbf{x}) = \frac{1}{n} \sum_{i=1}^n I\{x_i \leq x\}, x \in \mathbb{R},$$

2. на основу добијене емпиријске функције генеришемо нових  $R$  узорака, што се заправо своди на генерисање нових узорака

$$\mathbf{x}_i^* = (x_{1i}^*, x_{2i}^*, \dots, x_{ni}^*), i = 1 \dots R$$

путем извлачења са враћањем из основног узорка, који сада посматрамо као нову популацију, и ово је основна разлика између ове две врсте бутстрепа. Приметимо да би укупан број могућих реузорака у овом случају био  $n^n$ , дакле, непараметарски бутстреп би идејно посматрао свих  $n^n$  могућих реузорака, а овај поступак извлачења  $R$  реузорака заправо представља Монте Карло апроксимацију непараметарског бутстрепа.

3. за сваки бутстреп узорак рачунамо вредност оцене

$$\hat{\theta}_i^* = \hat{\theta}(\mathbf{x}_i), i = 1 \dots R,$$

4. одређујемо емпиријску функцију расподеле на основу бутстреп оцена добијених у кораку 3.

### 2.3 Примена бутстреп методе за оцену средине и дисперзије

На основу статистичке теорије познато је да за оцену средине обележја  $X$  на основу простог случајног узорка  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  користимо узорачку средину

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

а за оцену дисперзије користимо узорачку дисперзију

$$\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

или уколико желимо непристрасну оцену дисперзије користи се поправљена узорачка дисперзија

$$\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Овим се намеће да се бутстреп оцена средње вредности оцене  $\hat{\theta}$  параметра  $\theta$  добија рачунањем средине колекције бутстреп оцена  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_R^*)$ ,

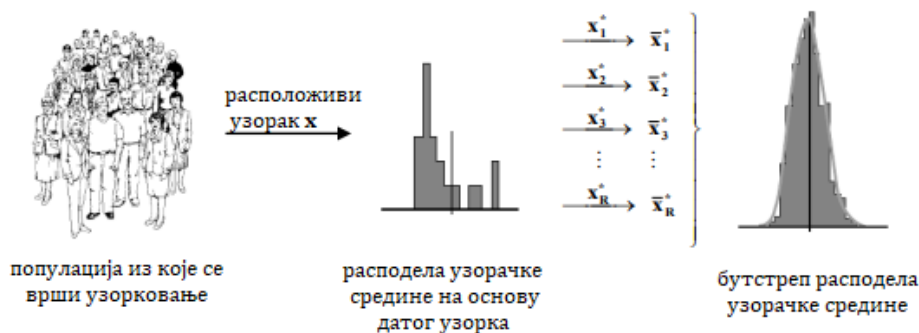
$$\bar{\theta}^* = \frac{1}{R} \sum_{i=1}^R \hat{\theta}_i^*.$$

Аналогно, бутстреп оцена дисперзије оцене  $\hat{\theta}$  параметра  $\theta$  је

$$\hat{\sigma}^{2*} = \frac{1}{R-1} \sum_{i=1}^R (\hat{\theta}_i^* - \bar{\theta}^*)^2.$$

Илустроваћемо ово на примеру нормалне расподеле  $N(m, \sigma^2)$ . Применом бутстреп методе треба одредити оцену средине и варијансе оцене  $\hat{\theta}$  непознатог параметра  $\theta = m$ , тј. треба оценити расподелу статистике  $\bar{X}_n$  којом се оцењује параметар  $m$  (в. Слику 1). Разликујемо два случаја:

- Уколико нам је познато да је основни узорак из нормалне расподеле, користимо параметарски бутстреп. Генеришемо  $R$  реузорака из нормалне расподеле са оцењеним параметрима  $m$  и  $\sigma^2$  реализованим вредностима статистика  $\bar{x}_n$  и  $\tilde{s}_n^2$  које добијамо на основу датог узорка. За сваки од реузорака рачуна се реализована вредност узорачке средине чиме се добија  $R$  бутстреп оцена статистике  $\bar{X}_n: \bar{x}_1^*, \bar{x}_2^*, \dots, \bar{x}_R^*$ . На основу добијених бутстреп оцена можемо да одредимо емпиријску функцију расподеле, као и оцене средње вредности и дисперзије статистике  $\bar{X}_n$ .
- Уколико расподела основног узорка није позната користимо непараметарски бутстреп. Основна разлика је што реузорке генеришемо као случајне узорке са понављањем из популације коју чине вредности основног узорка. Поступак је даље исти као код параметарског бутстреп.



Слика 1: Илустрација бутстреп за оцену расподеле параметра  $m$

## 2.4 Примена бутстреп методе за оцену интервала поверења

Постоји више бутстреп начина за конструкцију интервала поверења, као што су Ефронов перцентилни бутстреп, Ефронов перцентилни бутстреп са корекцијом пристрасности, перцентилни т-бутстреп и нормални т-бутстреп. Овде ћемо описати *Ефронов перцентилни бутстреп метод*, а о осталим методама може се више наћи у раду [3].

Нека је дат прост случајан узорак  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  обележја  $X$  са расподелом  $F$  у којој фигурише непознати параметар  $\theta$ . Потребно је конструисати интервал поверења за параметар  $\theta$ .

Примењујемо стандардну бутстреп процедуру, тј. на основу реализованог узорка  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  рачунамо вредност оцене  $\hat{\theta}$ , а затим генеришемо бутстреп оцене  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_R^*$  из расподеле  $F_{\hat{\theta}}$ .

Ефрон је предложио да уколико желимо да направимо нпр. 90% интервал поверења, изаберемо 90% вредности бутстреп оцена  $\hat{\theta}_i^*, i = 1, \dots, R$  који се налазе у средини варијационог низа ових оцена, а да одбацимо 5% најмањих и 5% највећих вредности.

Ако је  $H^*$  функција расподеле бутстреп оцене  $\hat{\theta}^* = \theta(\mathbf{X})$ , горњу границу интервала поверења налазимо решавањем једначине  $H^*(x) = (1 + \alpha)/2$  по  $x$ . Добија се

$$\hat{\theta}^g(\alpha) = (H^*)^{-1}((1 + \alpha)/2).$$

Када је  $R$  велико, функција расподеле  $H^*$  може се апроксимирати емпријском функцијом расподеле за  $\hat{\theta} = (\hat{\theta}_1^*, \dots, \hat{\theta}_R^*)$

$$H_R(\rho|\hat{\theta}) = \frac{1}{R} \sum_{i=1}^R I(\hat{\theta}_i^* \leq \rho), \rho \in \mathbb{R}.$$

Одавде важи:

$$(H^*)^{-1}((1 + \alpha)/2) \rightarrow H_R^{-1}((1 + \alpha)/2|\hat{\theta}), R \rightarrow \infty.$$

У пракси ову вредност добијамо одређивањем броја

$$g = \frac{1 + \alpha}{2} R.$$

За горњу границу интервала поверења бирамо  $g$ -ту вредност  $\hat{\theta}_{(g)}^*$  варијационог низа

$$\hat{\theta}_{(1)}^* \leq \hat{\theta}_{(2)}^* \leq \dots \leq \hat{\theta}_{(R)}^*.$$

Уколико  $g$  није цео број, узимамо први цео број  $n$  мањи од  $g$ , а затим интерполацијом добијамо горњу границу која је облика:

$$\hat{\theta}_{(g)}^* = \hat{\theta}_{(n)}^* + (g - n)(\hat{\theta}_{(n+1)}^* - \hat{\theta}_{(n)}^*).$$

Аналогно рачунамо и доњу границу интервала поверења

$$\hat{\theta}^d(\alpha) = (H^*)^{-1}((1 - \alpha)/2) \rightarrow H_R^{-1}((1 - \alpha)/2|\hat{\theta}), R \rightarrow \infty.$$

У пракси рачунамо број  $d$  тако да важи:

$$d = \frac{1 - \alpha}{2} R,$$

а онда статистику поретка  $\hat{\theta}_{(d)}^*$  проглашавамо доњом границом. Уколико  $d$  није цео број онда већ описаним поступком интерполације налазимо доњу границу.

Добијени интервал поверења  $I = (\hat{\theta}_{(d)}^*, \hat{\theta}_{(g)}^*)$  одређује  $100\alpha\%$  Ефронов перцентилни интервал поверења за параметар  $\theta$ .

### 3 Временске серије

Анализа временских серија представља једну од најпопуларнијих статистичких дисциплина данашњице. Закључивање у анализи временских серија битно се разликује од уобичајног начина статистичког закључивања. Основни појам код статистичког закључивања је *прост случајан узорак*, под којим посматрамо скуп од  $n$  независних једнакорасподељених случајних променљивих, док се код временских серија посматра скуп случајних променљивих за које се претпоставља да су међусобно зависне, а веома често и корелисане.

Циљеви анализе временских серија су:

**Описивање временске серије** даје основне информације о временској серији помоћу графичког приказа и сумарних статистика те временске серије. Описивањем могу се добити основне информације о временској серији, којима можемо утврдити суштинску карактеристику посматране појаве и тиме избећи сложеније статистичке методе анализе временске серије.

**Објашњење временске серије** представља избор одговарајућег математичког модела који описује кретање временске серије на задовољавајући начин.

**Предвиђање временске серије** подразумева да се изабрани модел користи за прогнозирање будућег стања временске серије на основу прошлих опсервација.

Временске серије имају велику примену у пракси, користе се у економији, статистици, финансијама; њима се представља природни прираштај, флукуација девизног курса, временска прогноза месечно кретање индустријске производње, увоза, извоза и цена...

#### 3.1 Дефиниција, стационарност и ергодичност временских серија

Како бисмо дефинисали временску серију неопходно је прво дефинисати случајни (стохастички) процес.

**Дефиниција 3.1** *Случајни процес је фамилија случајних променљивих  $\{X_t(\omega) : t \in T, \omega \in \Omega\}$  дефинисаних над истим простором вероватноћа  $(\Omega, \mathcal{F}, \mathcal{P})$ , где је  $T$  индекси скуп.*

Индексни skup  $T$  може бити дефинисан на више начина. Уколико је  $T = \mathbb{R}$  или  $T \subseteq \mathbb{R}$  добија се случајни процес са непрекидним временом  $\{X_t\}_{t \in \mathbb{R}}$ , а уколико је skup  $T$  skup целих или природних бројева добија се случајни процес са дискретним временом,  $\{X_t\}_{t \in \mathbb{Z}}$ .

**Дефиниција 3.2** *Временска серија  $\{X_t, t \in T\}$  је реалан случајни процес, при чему случајне величине  $X_t$  и  $X_s$ ,  $t \neq s$  не морају бити некорелисане.*

Можемо рећи да временске серије представљају уређени низ података, при чему се уређење остварује у односу на време и најчешће у једнаким временским интервалима.

Уколико се својства временске серије у току времена не мењају онда је дата серија стационарна, у супротном она је нестационарна. Стационарност је веома битна у анализи временских серија, у зависности од стационарности временске серије бирају се различите статистичке методе за анализу. Постоје две основне дефиниције стационарности:

**Дефиниција 3.3** *(Строга стационарност)*

*Временска серија  $X_t$  је строго стационарна ако за било која два природна броја  $n$  и  $k$  и било коју  $n$ -торку природних бројева  $(t_1, \dots, t_n)$  случајни низови  $(X_{t_1}, \dots, X_{t_n})$  и  $(X_{t_1+k}, \dots, X_{t_n+k})$  имају исту функцију расподеле.*

**Дефиниција 3.4** *(Слаба стационарност)*

*За временску серију  $X_t$  кажемо да је слабо стационарна уколико важе следећи услови:*

1.  $E(X_t) = \mu = \text{const.}, \forall t \in Z,$
2.  $E(X_t^2) < \infty, \forall t \in Z,$
3.  $\text{cov}(X_t, X_{t-k}) = E(X_t - \mu)E(X_{t-k} - \mu) = \gamma(k), \forall t, k \in Z,$  где је  $\gamma(\cdot)$  аутоковаријациона функција.

Ако је временска серија строго стационарна онда је она и слабо стационарна једино ако има коначну дисперзију, обрнуто не важи.

**Дефиниција 3.5** *Нека је  $X_t$  слабо стационарна временска серија чије је очекивање  $EX_t = \mu$ . Таква временска серија је ергодична у односу на средњу вредност ако важи услов*

$$\overline{X} \xrightarrow{P} \mu, T \rightarrow \infty,$$

где је  $\bar{X} = \frac{1}{T} \sum_{i=1}^T X_i$ .

Дакле, временска серија је ергодична у односу на средњу вредност ако аритметичка средина датог скупа конвергира у вероватноћи ка стварној средњој вредности временске серије како се повећава обим узорка.

### 3.2 Компоненте временске серије

Компоненте временских серија су тренд, сезонска, циклична и случајна компонента. Временска серија не мора садржати све четири наведене компоненте, али мора садржати случајну компоненту.

**Тренд** изражава дугорочну тенденцију развоја серије. Може бити растући или опадајући у зависности од тога да ли вредности серије током времена систематски расту или опадају. Такође, може бити детерминистички или стохастички, у зависности од тога да ли се промене временске серије током времена могу предвидети или не.

**Сезонска компонента** представља правилност у кретању временске серије која се појављује у оквиру једне календарске године. Присуство сезонске компоненте утиче на то да постоји већи степен корелације између истих месеци различитих година, него између различитих месеци током једне године.

**Циклична компонента** представља правилност у кретању временске серије која се појављује са периодом различитим од једне године. У пракси, сматра се да тренд укључује цикличну компоненту.

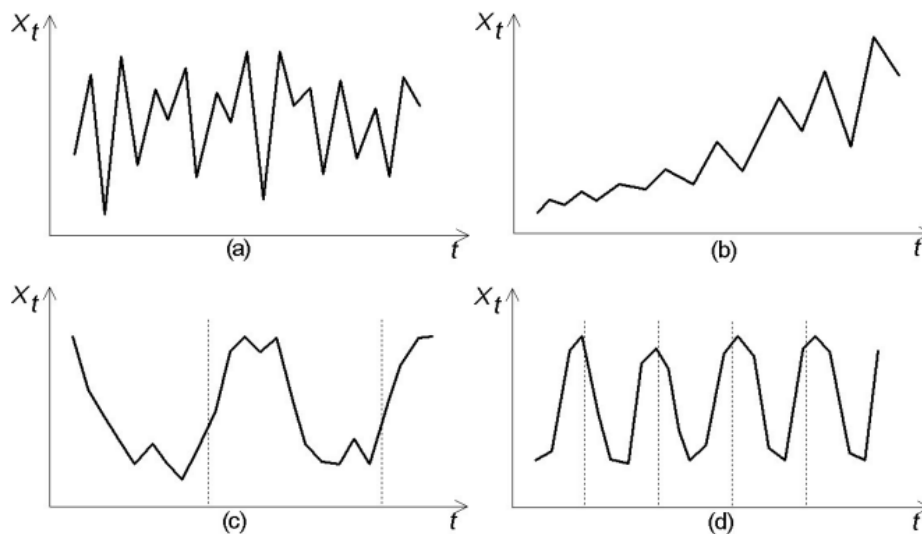
**Случајна компонента** представља промене у кретању временске серије које су случајног карактера, она је стационарни процес познатији као *бели шум*.

**Дефиниција 3.6** (*Бели шум*) Процес  $\{\varepsilon_t, \forall t \in Z\}$  назива се *бели шум* уколико важе следећи услови:

1.  $E(\varepsilon_t) = 0, \forall t \in Z,$
2.  $E(\varepsilon_t^2) = \sigma^2 = const, \forall t \in Z,$
3.  $\gamma(t, t - k) = E(\varepsilon_t \varepsilon_{t-k}) = 0,$  где је  $\gamma(\cdot)$  аутоковаријациона функција.

Уколико се овим условима дода услов да су чланови низа независне случајне променљиве са нормалном функцијом расподеле, онда се такав процес назива *Гаусов бели шум*.

На Слици 2 приказано је неколико карактеристичних типова временских серија, где  $X_t$  означава опсервацију временске серије у тренутку  $t$ . Први тип временске серије (a) је стационарна временска серија. Серија (b) је нестационарна временска серија са израженим трендом, серија (c) је нестационарна временска серија са израженим цикличним варијацијама и серија (d) је са израженим сезонским варијацијама. Слика је преузета из литературе [4].



Слика 2: Графички приказ неких временских серија

### 3.3 Декомпозиција временских серија

#### Адитивни, мултипликативни и мешовити модел

Нека је  $X_t$  опсервација временске серије у тренутку  $t$ ,  $f_t$  тренд,  $S_t$  сезонска компонента,  $C_t$  циклична компонента и  $e_t$  случајна компонента те временске серије у тренутку  $t$ .

**Адитивни модел** подразумева да се опсервације серије представљају као збир све четити компоненте:

$$X_t = f_t + S_t + C_t + e_t.$$

Овај модел претпоставља да се компоненте мењају независно једна од друге, нпр. повећање тренда неће узроковати повећање сезонске компоненте. Овај модел користи се када је сезонска варијација релативно стабилна током времена, тј. ако су сезонски ефекти исти из године у годину.



**Мултипликативни модел** подразумева да се опсервације серије представљају као производ све четити компоненте временске серије:

$$X_t = f_t \cdot S_t \cdot C_t \cdot e_t.$$

Овај модел претпоставља да су све компоненте међусобно зависне. Овај модел се користи уколико се повећањем (смањивањем) појаве пропорционално повећава (смањује) утицај сезонске компоненте током времена. Мултипликативни модел се може записати у адитивној форми коришћењем логаритма:

$$\log X_t = \log f_t + \log S_t + \log C_t + \log e_t.$$

Мултипликативни модели или лог-адитивни модели се не могу користити када оригинална временска серија садржи нуле или негативне вредности. У овом случају може се користити мешовити модел.

**Мешовити модел** даје могућност да опсервације серије буду резултат различитих комбинација компоненти, нпр.

$$X_t = f_t \cdot S_t \cdot C_t + e_t \text{ или } X_t = f_t \cdot C_t \cdot (S_t + e_t - 1).$$

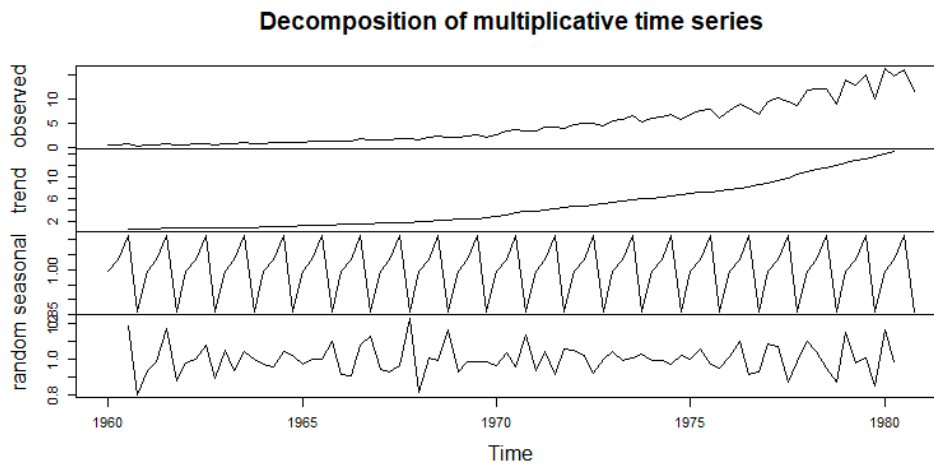
Овај модел комбинује елементе адитивних и мултипликативних веза и претпоставља да су сезонска и случајна компонента зависне од тренда, али независне међусобно.

### Метод декомпозиције

Метод декомпозиције представља разлагање временске серије на њене компоненте, а такође даје могућност да се поједине компоненте елиминишу из временске серије, што је посебно погодно за сезонске временске серије, где се елиминацијом сезонске компоненте, тренд може видети јасније. Да би декомпозиција временске серије била успешно изведена, пре свега је потребно одредити који модел одговара датој серији: адитивни, мултипликативни или мешовити. Декомпозиција временске серије подразумева: оцену тренда, уклањање тренда из временске серије, оцену сезонских фактора и њихова нормализација, одређивање случајне компоненте која треба да представља стационарне резидуале.

Декомпозицију података показаћемо на скупу података JohnsonJohnson који се налази у бази података **R** програма. Овај скуп података представља тромесечну зараду (у доларима) по деоницама компаније Johnson & Johnson у периоду од 1960. до 1980. године. Декомпозиција временске серије у **R** програму врши се применом функције `decompose()`. За декомпозицију ове временске серије коришћен је мултипликативни модел. На

Слици 3 можемо видети графички приказ оригиналног скупа података, као и графички приказ сваке од компоненти. Приметимо да овај скуп података има јасно изражен тренд и сезонску компоненту.



Слика 3: Декомпозиција података JohnsonJohnson

### 3.4 Аутоковаријациона и аутокорељациона функција

Битна особина временских серија јесте корелација између случајних величина. За испитивање корелације користе се аутокорељациона и аутоковаријациона функција.

**Дефиниција 3.7** (Аутоковаријациони коефицијент)

Нека је  $\{X_t, t \in Z\}$  слабо стационарна временска серија, чија је средња вредност  $\mu$ . Аутоковаријациони коефицијент  $\gamma_k$  за временску серију  $X_t$  дефинише се на следећи начин:

$$\gamma_k = \text{cov}(X_t, X_{t-k}) = E(X_t - \mu)E(X_{t-k} - \mu), k = 1, 2, \dots$$

**Дефиниција 3.8** (Аутоковаријациона функција)

Аутоковаријациона функција представља низ аутоковаријационих коефицијената  $\gamma_1, \gamma_2, \dots$

**Дефиниција 3.9** (Аутокорељациони коефицијент)

Аутокорељациони коефицијент  $\rho_k$  представља коефицијент корелације између  $X_t$  и  $X_{t-k}$  и дефинисан је са:

$$\rho_k = \frac{\text{cov}(X_t, X_{t-k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-k})}} = \frac{\text{cov}(X_t, X_{t-k})}{\text{Var}(X_t)} = \frac{\gamma_k}{\gamma_0}, k = 1, 2, \dots$$

**Дефиниција 3.10** (Аутокорелациона функција)

Скуп аутокорелационих коефицијената  $\rho_1, \rho_2, \dots$  који су уређени у односу на време назива се аутокорелациона функција.

**Дефиниција 3.11** Графички приказ аутокорелационе функције назива се **корелограм**.

**Особине аутоковаријационих и аутокорелационих коефицијената**

1. за  $k = 0$  добија се:  $\gamma_0 = \text{Var}(X_t)$  и  $\rho_0 = 1$ ,
2. важи симетричност:  $\gamma_k = \gamma_{-k}$  и  $\rho_k = \rho_{-k}$ ,
3.  $|\gamma_k| \leq \gamma_0$  и  $\rho_k \leq 1$ ,
4. Аутоковаријациона ( $\mathbf{\Gamma}_n$ ) и аутокорелациона ( $\mathbf{P}_n$ ) матрица су позитивно семидефинитне, при чему је:

$$\mathbf{\Gamma}_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix} = \sigma^2 \mathbf{P}_n$$

**Оцена аутоковаријационих и аутокорелационих коефицијената**

Нека је низ  $X_1, X_2, \dots, X_T$  слабо стационарна временска серија обима  $T$ . За оцену средње вредности  $\mu$  користи се узорачка средина  $\bar{X}$ :

$$\bar{X} = \frac{1}{T} \sum_{i=1}^T X_i.$$

**Дефиниција 3.12** Оцена аутоковаријационих коефицијената дефинише се на два начина:

1.  $\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})$ ,
2.  $\hat{\gamma}_k = \frac{1}{T-k} \sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})$ .

**Дефиниција 3.13** Оцена аутокорелационих коефицијената дефинише се са:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}.$$

**Дефиниција 3.14** *Графички приказ оцена аутокорелационих коефицијената назива се **узорачки корелограм** и означава се са **ACF** (енгл. autocorrelation function).*

Узорачки корелограм служи за препознавање модела стационарних временских серија и помоћу њега се може проценити да ли временска серија има тренд и сезонску компоненту.

## 3.5 Модели стационарних временских серија

### 3.5.1 Линеарни процес

**Дефиниција 3.15** *Временска серија  $X_t$  је линеарни процес уколико се она може записати на следећи начин*

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i},$$

где је  $\mu$  очекивање временске серије  $X_t$ ,  $\psi_0 = 1$ ,  $\psi_1, \psi_2, \dots$  су произвољни параметри помоћу којих је описана динамичка структура временске серије и  $\{\varepsilon_t\}$  је низ независних случајних променљивих са истом расподелом.

Процес  $\{\varepsilon_t\}$  је најчешће бели шум.

Особине линеарног процеса:

1. Очекивање:

$$E(X_t) = E(X_t) + \sum_{i=0}^{\infty} \psi_i E(\varepsilon_{t-i}) = \mu + 0 = \mu.$$

2. Дисперзија:

$$\text{Var}(X_t) = \text{Var}(\mu) + \sum_{i=0}^{\infty} \psi_i^2 \text{Var}(\varepsilon_{t-i}) = 0 + \sigma^2 \sum_{i=0}^{\infty} \psi_i^2 = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2.$$

3. Аутоковаријациона функција:

$$\begin{aligned} \gamma_k &= \text{cov}(X_t, X_{t-k}) = E(X_t - \mu)(X_{t-k} - \mu) \\ &= E((\varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_k \varepsilon_{t-k} + \psi_{k+1} \varepsilon_{t-k-1} + \dots)(\varepsilon_{t-k} + \psi_1 \varepsilon_{t-k-1} + \dots)) \\ &= \sigma^2 (\psi_k + \psi_1 \psi_{k+1} + \psi_2 \psi_{k+2} + \dots) \\ &= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}. \end{aligned}$$

4. Аутокорелациона функција:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

Услов стационарности линеарног процеса дефинишемо преко коефицијената  $\psi_i$ . У општем случају ових коефицијената има бесконачно много, дакле за стационарност потребно је показати да је  $\gamma_k$  коначно за свако  $k$ . Без умањења општости претпоставићемо да је  $\mu = 0$ , како је:

$$|\gamma_k| = |E(X_t X_{t-k})| \leq [Var(X_t) Var(X_{t-k})]^{1/2} = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2,$$

значи да је услов стационарности задовољен ако је  $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ .

**Теорема 3.1** (Волдова теорема разлагања)

Стационарни процес  $\{X_t, t \in T\}$  може се представити на следећи начин

$$X_t = V_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i},$$

где су

- $V_t$  детерминистичка компонента,
- $\{\varepsilon_t, t \in T\}$  процес белог шума,
- $\sum_{i=0}^{\infty} \psi_i^2 < \infty, \psi_0 = 1$ ,
- $E(V_t, \varepsilon_t) = 0$ .

Дакле, према датој Волдовој теореме разлагања сваки стационарни процес се може представити као збир два међусобно некорелисана процеса, од којих је један детерминистички, а други недетерминистички. Детерминистичка компонента представља константу (средњу вредност) и/или функцију линеарног тренда, и она се може предвидети на основу информација из прошлости, док се недетерминистичка компонента најчешће представља у облику линеарног процеса и она се не може предвидети на основу података из прошлости.

На основу линеарног процеса могу се извести три класе процеса. Погодним избором коефицијената  $\psi$  добијамо моделе са коначним бројем параметара и њима описујемо слабо стационарне временске серије. Ти модели су:

1. Ауторегресиони модел(енгл. Autoregressive model )
2. Модел покретних просека(енгл. Moving average model)
3. Ауторегресиони модел покретних просека(енгл. Autoregressive moving average model)

### 3.5.2 Ауторегресиони модели

**Дефиниција 3.16** Ауторегресиони модел реда  $p$  у ознаци  $AR(p)$  дефинисан је на следећи начин:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

где су  $\phi_1, \phi_2, \dots, \phi_p$  ауторегресиони параметри, а  $\varepsilon_t$  је процес белог шума.

$AR(p)$  модел може се записати и у облику

$$\Phi(L)X_t = \varepsilon_t,$$

где је  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 + \dots + \phi_p L^p$  и  $L$  је оператор померања  $L(X_t) = X_{t-1}$ .

Услов стационарности  $AR(p)$  модела одређен је са  $|q_i^{-1}| > 1$ , што је еквивалентно услову  $|q_i| < 1$  за  $i = 1, \dots, p$ , где су  $q_i^{-1}$  корени карактеристичне једначине  $\Phi(L) = 0$ .

#### 3.5.2.1 Ауторегресиони модел првог реда $AR(1)$

Ауторегресиони модел првог реда  $AR(1)$  представља се на следећи начин:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t.$$

Рекурзивном заменом добија се:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \varepsilon_t \\ &= \phi_1 [\phi_1 X_{t-2} + \varepsilon_{t-1}] + \varepsilon_t \\ &= \phi_1^2 [\phi_1 X_{t-3} + \varepsilon_{t-2}] + \varepsilon_t + \phi_1 \varepsilon_{t-1} \\ &= \dots = \\ &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots \end{aligned}$$

Дакле, ауторегресиони модел првог реда је специјалан случај линеарног процеса при услову  $\psi_j = \phi_1^j, j = 1, 2, 3, \dots$ . Из овога следи и да је

$$E(X_t) = 0 \text{ и } Var(X_t) = \sigma^2(1 + \phi_1^2 + \phi_1^4 + \phi_1^6 + \dots) = \frac{1}{1-\phi_1^2}.$$

Дисперзија временске серије је коначна ако је  $|\phi_1| < 1$ , а то је и услов стационарности. Слично се може показати и за моделе вишег реда.

### 3.5.2.2 Аутоковаријациона и аутокорелациона функција AR(1) и AR(2) модела

Као што је раније дефинисано за аутоковаријациони коефицијент важи  $\gamma_k = cov(X_t, X_{t-k}) = E(X_t - \mu)(X_{t-k} - \mu)$ .

Како је очекивање једнако нули добија се  $\gamma_k = E(X_t X_{t-k})$ .

- Како за AR(1) модел важи:

$$X_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots,$$

следи да је:

$$X_{t-k} = \varepsilon_{t-k} + \phi_1 \varepsilon_{t-k-1} + \phi_1^2 \varepsilon_{t-k-2} + \phi_1^3 \varepsilon_{t-k-3} + \dots$$

Одавде добијамо облик аутоковаријационог коефицијента на растојању  $k$ :

$$\begin{aligned} \gamma_k &= E(X_t X_{t-k}) \\ &= E(\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots)(\varepsilon_{t-k} + \phi_1 \varepsilon_{t-k-1} + \phi_1^2 \varepsilon_{t-k-2} + \dots) \\ &= \sigma^2(\phi_1^k + \phi_1 \phi_1^{k+1} + \phi_1^2 \phi_1^{k+2} + \dots) \\ &= \sigma^2 \phi_1^k (1 + \phi_1^2 + \phi_1^4 + \dots) \\ &= \frac{\sigma^2 \phi_1^k}{1 - \phi_1^2}. \end{aligned}$$

Аутокорелациони коефицијент на растојању  $k$  одређен је са:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\frac{\sigma^2 \phi_1^k}{1 - \phi_1^2}}{\frac{\sigma^2}{1 - \phi_1^2}} = \phi_1^k.$$

- AR(2) модел дат је са:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t.$$

Аутоковаријационе и аутокорелационе коефицијенте добићемо тако што дату једначину модела помножимо са  $X_{t-k}$  и затим одредимо очекивање добијеног израза:

$$E(X_t X_{t-k}) = \phi_1 E(X_{t-1} X_{t-k}) + \phi_2 E(X_{t-2} X_{t-k}) + E(\varepsilon_t X_{t-k}),$$

одакле се добија

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + E(\varepsilon_t X_{t-k}).$$

Како је

$$\begin{aligned} E(\varepsilon_t X_{t-k}) &= E(\varepsilon_t (\varepsilon_{t-k} + \phi_1 \varepsilon_{t-k-1} + (\phi_1^2 + \phi_2) \varepsilon_{t-k-2} + \dots)) \\ &= \begin{cases} E(\varepsilon_t^2) = \sigma^2 & , k = 0 \\ 0 & , k \neq 0, \end{cases} \end{aligned}$$

за аутоковаријационе коефицијенте AR(2) модела важи:

$$\gamma_k = \begin{cases} \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \sigma^2 & , k = 0 \\ \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} & , k > 0. \end{cases}$$

Уколико израз  $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$  поделимо са  $\gamma_0$  добијамо аутокорелационе коефицијенте:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, k \geq 1.$$

Заменом  $k = 1$  и  $k = 2$  добија се:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 = \frac{\phi_1}{1 - \phi_2}; \rho_2 = \phi_1 \rho_1 + \phi_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2,$$

преостале аутокорелационе коефицијенте добијамо коришћењем дате рекурзивне формуле.

### 3.5.2.3 Оцена параметара AR(p) модела

При одређивању аутоковаријационе функције AR(p) модела користимо исти поступак као у случају AR(2) модела. Може се показати да се аутоковаријациони и аутокорелациони коефицијенти AR(p) модела могу изразити на следећи начин:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}, k > 0,$$



$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, k > 0.$$

Првих  $p$  једначина аутоковаријационих коефицијената гласе:

$$\begin{aligned} \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \dots + \phi_p \gamma_{p-1} \\ \gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 + \dots + \phi_p \gamma_{p-2} \\ &\dots \\ \gamma_p &= \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \dots + \phi_p \gamma_0 \end{aligned}$$

и називамо их Јул-Вокерове <sup>3</sup> једначине. На основу ових једначина можемо успоставити релацију између коефицијената  $AR(p)$  модела и аутоковаријационих коефицијената. У матричној форми Јул-Вокерове једначине дате су изразом:

$$\boldsymbol{\gamma} = \boldsymbol{\Gamma} \boldsymbol{\phi},$$

где су  $\boldsymbol{\gamma}' = [\gamma_1, \gamma_1, \dots, \gamma_p]$ ,  $\boldsymbol{\phi}' = [\phi_1, \phi_1, \dots, \phi_p]$  и  $\boldsymbol{\Gamma}$  аутоковаријациона матрица. Одавде можемо изразити параметре модела:

$$\boldsymbol{\phi} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}.$$

Оценом аутоковаријационих коефицијената добија се и оцена параметара модела, која се назива *Јул-Вокерова оцена параметара модела*:

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}^{-1} \hat{\boldsymbol{\gamma}}.$$

### 3.5.2.4 Парцијална аутокорелациона функција

Како бисмо лакше одредили ред  $AR(p)$  модела уводимо нови појам тзв. парцијалну аутокорелациону функцију (енгл. partial autocorrelation function). Као што је раније дефинисано, аутокорелациони коефицијент на растојању  $k$  мери степен корелације између величина  $X_t$  и  $X_{t-k}$ . С обзиром на то да  $\rho_k$  може да буде под утицајем корелисаности  $X_t$  и  $X_{t-k}$  са члановима временске серије на растојањима између тренутака  $t$  и  $t - k$ , елиминацијом утицаја тих чланова добија се парцијални аутокорелациони коефицијент.

Јул-Вокерове једначине могу се записати и помоћу аутокорелационе матрице, тј. важи:

$$\boldsymbol{\rho} = \mathbf{P} \boldsymbol{\phi},$$

где је  $\boldsymbol{\rho}' = [\rho_1, \rho_2, \dots, \rho_p]$ ,  $\boldsymbol{\phi}' = [\phi_1, \phi_2, \dots, \phi_p]$  и  $\mathbf{P}$  је аутокорелациона матрица.

---

<sup>3</sup>Yule-Walker, у литератури се може наћи и превод Џул-Вокерове једначине

Ако у ову једначину уврстимо  $p = k$  и  $\phi_i = \phi_{ii}$ , онда можемо доћи до вредности парцијалних аутокорељационих коефицијената помоћу Крамеровог поступка примењеног на тако добијени систем једначина. Добија се следећи облик парцијалног аутокорељационог коефицијента у општем случају:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}.$$

Парцијална аутокорељациона функција има исте особине као аутокорељациона функција, а интерпретација парцијалних аутокорељационих коефицијената назива се *парцијални корелограм*.

Сада можемо одредити парцијалне аутокорељационе коефицијенте AR процеса:

AR(1) процес:  $\phi_{11} = \rho_1 = \phi_1, \phi_{kk} = 0, k > 1,$

AR(2) процес:  $\phi_{11} = \rho_1 = \frac{\phi_1}{1-\phi_2}, \phi_{22} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \phi_2, \phi_{kk} = 0, k > 2,$

AR( $p$ ) процес:  $\phi_{11} = \rho_1, \phi_{11} \neq 0, \dots, \phi_{pp} \neq 0, \phi_{kk} = 0, k > p.$

Закључујемо да су парцијални аутокорељациони коефицијенти  $\phi_{kk}$  једнаки нули када је  $k$  веће од реда AR модела и последњи парцијални аутокорељациони коефицијент AR( $p$ ) модела једнак је последњем ауто-регресионом коефицијенту  $\phi_p$ , што је битно својство за идентификацију AR модела временске серије.

### 3.5.3 Модели покретних просека

**Дефиниција 3.17** *Модел покретних просека реда  $q$  у ознаци MA( $q$ ) дефинише се на следећи начин:*

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q},$$

где су  $\theta_1, \theta_2, \dots, \theta_q$  параметри модела, а  $\varepsilon_t$  је процес белог шума.

MA( $q$ ) модел је специјалан случај линеарног процеса при услову

$$\psi_j = -\theta_j \text{ за } j \leq q \text{ и } \psi_j = 0 \text{ за } j > q.$$

Линеарни процес се може означити и као МА модел бесконачног реда  $MA(\infty)$ . Приметимо да за МА модел важи да је

$$E(X_t) = 0, Var(X_t) = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2) < \infty,$$

па закључујемо да је коначни процес покретних просека увек стационаран.

$MA(q)$  модел може се записати и у облику:

$$X_t = \Theta(L)\varepsilon_t,$$

где је  $\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$  и  $L$  је оператор померања  $L(X_t) = X_{t-1}$ .

Битна особина  $MA(q)$  модела је инвертибилност. Инвертибилност показује везу између форми  $AR$  и  $MA$  модела. Може се показати да стационарном  $AR(p)$  процесу коначног реда одговара  $MA$  процес бесконачног реда, а да инвертибилном  $MA(q)$  процесу коначног реда одговара  $AR$  процес бесконачног реда.

Услов инвертибилности  $MA(q)$  модела одређен је са  $|q_i^{-1}| > 1$ , што је еквивалентно услову  $|q_i| < 1$  за  $i = 1, \dots, q$ , где су  $q_i^{-1}$  корени карактеристичне једначине  $\Theta(L) = 0$ .

$AR$  модел облика  $\Phi(L)X_t = \varepsilon_t$  може се записати у облику  $MA$  модела

$$X_t = \frac{1}{\Phi(L)}\varepsilon_t = \Theta(L)\varepsilon_t.$$

На основу добијене релације  $\Phi(L)\Theta(L) = 1$  може се извести веза између коефицијената  $AR$  и  $MA$  модела.[4]

### 3.5.3.1 Аутоковаријациона и аутокорелациона функција $MA(1)$ и $MA(2)$ модела:

- $MA(1)$  модел одређен је са:  $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ .  
Дисперзија овог модела је  $Var(X_t) = \sigma^2(1 + \theta_1^2)$ .  
 $MA(1)$  процес може се записати као линеарни процес за  $\psi_1 = -\theta_1$  и  $\psi_k = 0$  за  $k > 1$ . Одавде лако можемо добити аутоковаријационе коефицијенте  $MA(1)$  процеса:

$$\gamma_1 = -\sigma^2\theta_1 \text{ и } \gamma_k = 0 \text{ за } k > 1.$$

Аутокорелациони коефицијенти одређени су са:

$$\rho_1 = -\frac{\theta_1}{1+\theta_1^2} \text{ и } \rho_k = 0 \text{ за } k > 1.$$

- МА(2) модел одређен је са:  $X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$ . Дисперзија овог модела је  $Var(X_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2)$ . МА(2) процес може се записати као линеарни процес за  $\psi_1 = -\theta_1$ ,  $\psi_2 = -\theta_2$  и  $\psi_k = 0$  за  $k > 2$ . Одавде лако можемо добити аутоковаријационе коефицијенте МА(2) процеса:

$$\gamma_1 = \sigma^2(-\theta_1 + \theta_1\theta_2), \gamma_2 = -\sigma^2\theta_2 \text{ и } \gamma_k = 0 \text{ за } k > 2.$$

Аутокорелациони коефицијенти одређени су са:

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \text{ и } \rho_k = 0 \text{ за } k > 2.$$

### 3.5.3.2 Парцијална аутокорелациона функција МА(1) модела

С обзиром на то да МА модел карактерише особина инвертибилности, тј. МА(1) моделу одговара АР модел бесконачног реда, можемо да закључимо да код МА модела парцијални аутокорелациони коефицијенти опадају са растом кашњења.

За произвољно  $k$  облик парцијалне аутокорелационе функције МА(1) модела дат је са:

$$\phi_{kk} = -\frac{\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}, k = 1, 2, \dots$$

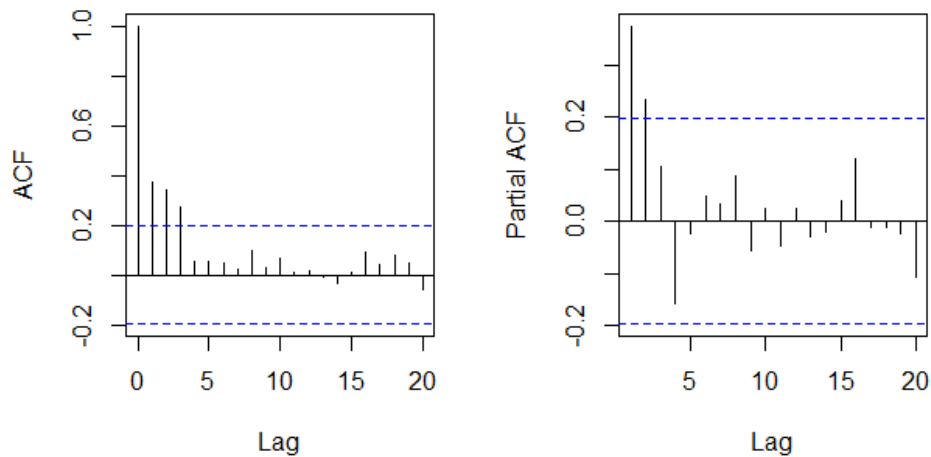
### 3.5.3.3 Избор модела

У пракси ред  $p$  и  $q$  АР/МА модела су непознати и треба их емпиријски одредити. Један од начина одређивања реда модела јесте помоћу аутокорелационе и парцијалне аутокорелационе функције, а за упоређивање модела користи се АИС критеријум (енгл. Akaike information criterion), који је заснован на теорији максималне веродостојности.

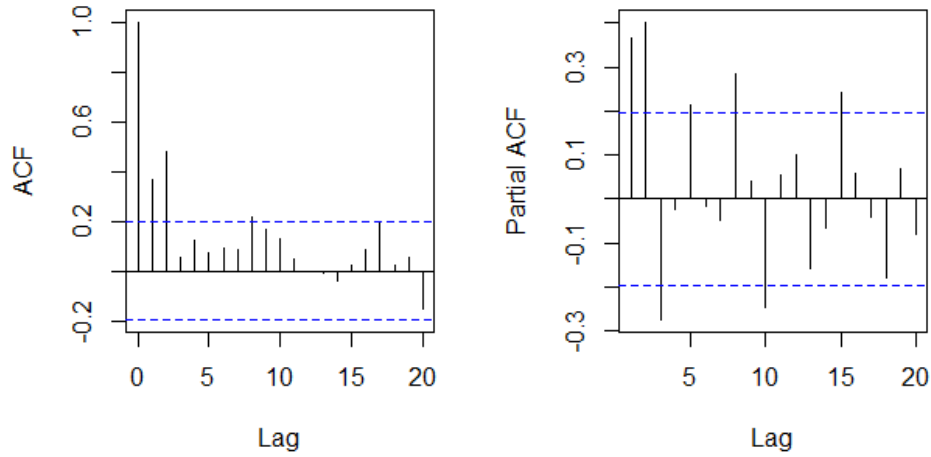
На основу већ дате анализе аутокорелационе и парцијалне аутокорелационе функције АР и МА модела, можемо закључити да аутокорелациона функција експоненцијално опада након кашњења  $p$  код АР( $p$ ) модела и прекида се након кашњења  $q$  код МА( $q$ ) модела, док

за пацијалну аутокорељациону функцију важи обрнуто, она се прекида након кашњења  $p$  код  $AR(p)$  модела и експоненцијално опада након кашњења  $q$  код  $MA(q)$  модела. Сходно томе можемо закључити да је за одређивање реда  $AR(p)$  модела кориснија парцијална аутокорељациона функција, а за одређивање реда  $MA(q)$  модела кориснија је аутокорељациона функција.

На Слици 4 дат је пример узорачког корелограма и парцијалног корелограма за  $AR(2)$  модел:  $X_t = 0.3X_{t-1} + 0.2X_{t-2} + \varepsilon_t$ , а на Слици 5 дат је пример узорачког корелограма и парцијалног корелограма за  $MA(2)$  модел:  $X_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$ .



Слика 4: ACF и PACF за  $AR(2)$  модел



Слика 5: ACF i PACF за MA(2) модел

Даље ћемо описати AIC критеријум за избор модела.

Нека је  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  прост случајан узорак са заједничком густином  $f(\mathbf{x}; \theta)$ . Тада је функција веродостојности  $L(\theta)$  дефинисана са:

$$L(\theta) = f(\mathbf{x}; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

односно функција веродостојности представља заједничку густину посматрану као функцију параметра  $\theta$ . Оцена параметра  $\theta$  је вредност која максимизује функцију веродостојности, тј. важи:

$$\hat{\theta} = \arg \max_{\theta} L(\theta).$$

Да би се лакше одредио овај максимум прелази се на тражење максимума логаритма функције веродостојности:

$$\ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta),$$

одакле је оцена параметра  $\theta$  дефинисана са:

$$\hat{\theta} = \arg \max_{\theta} \ln L(\theta).$$

Код временских серија вредности узорка нису независне величине, тако да заједничка функција густине неће бити производ маргиналних густина, већ заједничку функцију густине дефинишемо преко условне функције густине:

$$f_{X|Y}(x; \theta) = \frac{f_{X,Y}(x, y; \theta)}{f_Y(y; \theta)},$$

одакле је заједничка функција густине одређена са:

$$f_{X,Y}(x, y; \theta) = f_Y(y; \theta)f_{X|Y}(x; \theta).$$

Применом горе наведене условне густине добијамо заједничку функцију густине за временску серију  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ :

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta)f(x_2|x_1; \theta) \dots f(x_n|x_1, x_2, \dots, x_{n-1}; \theta),$$

што се може записати и на следећи начин:

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \prod_{i=1}^n f(x_i|x_1, \dots, x_{i-1}; \theta),$$

где је  $f(x_1; \theta)$  маргинална функција густине прве вредности, а  $f(x_i|x_1, \dots, x_{i-1}; \theta)$  условна функција густине за  $x_i$  у односу на претходне вредности.

Као што је већ поменуто, функција веродостојности користи се за оцену параметара модела, али може се искористити и за упоређивање два модела. Информациони критеријум за избор модела рачуна се на основу оцењене вредности функције веродостојности и броја параметара модела.

Нека је  $\hat{\theta}$  вредност параметра која максимализује вредност функције веродостојности и нека је  $\hat{L}$  тај максимум. Нека је  $k$  димензија параметра  $\theta$ , односно број оцењених параметара посматраног модела. Тада

$$AIC = 2k - 2 \ln \hat{L}$$

даје вредност на основу које бирамо модел. Међу посматраним моделима бира се онај модел за који је АИС вредност најмања.

### 3.5.4 Ауторегресиони модели покретних просека

**Дефиниција 3.18** Ауторегресиони модел покретних просека у ознаци  $ARMA(p, q)$  дефинише се на следећи начин:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

где је где је  $p$  ред ауторегресионе компоненте и  $q$  ред компоненте покретних просека, а је  $\{\varepsilon_t\}$  бели шум.

ARMA( $p, q$ ) може се записати и помоћу оператора кашњења:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)X_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)\varepsilon_t,$$

односно у сажетом облику

$$\Phi(L)X_t = \Theta(L)\varepsilon_t.$$

Специјални случајеви ARMA( $p, q$ ) модела:

$$\begin{aligned} ARMA(p, 0) &= AR(p) \\ ARMA(0, q) &= MA(q) \end{aligned}$$

### 3.5.4.1 Аутоковаријациона и аутокорелациона функција ARMA модела

Да бисмо одредили аутоковаријациону и аутокорелациону функцију изразимо ARMA( $p, q$ ) модел у облику

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}.$$

Множењем ове једначине са  $X_{t-k}$  и одређивањем очекивања тог израза добијамо:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + E(X_{t-k} \varepsilon_t) - \theta_1 E(X_{t-k} \varepsilon_{t-1}) - \dots - \theta_q E(X_{t-k} \varepsilon_{t-q}).$$

Како је  $E(X_{t-k} \varepsilon_{t-i}) = 0$  за  $k > i$ , добијамо аутоковаријационе коефицијенте:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}, \text{ за } k \geq q + 1,$$

одакле добијамо и аутокорелационе коефицијенте:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}.$$

### 3.5.4.2 Ауторегресиони модели покретних просека за интегрисане временске серије

У раду са ARMA моделима често се јавља проблем стационарности, који се решава увођењем ARIMA модела који нестационарну временску серију прво своди на стационарну, диференцирањем, а затим се серија моделира ARMA моделом.



Диференцирање временске серије представља операцију одузимања узастопних опсервација

$$\Delta X_t = X_t - X_{t-1}.$$

За временску серију кажемо да је интегрисана временска серија реда  $d$  уколико може бити трансформисана у стационарну временску серију диференцирањем  $d$  пута.

**Дефиниција 3.19** Ауторегресиони модели покретних просека за интегрисане временске серије у ознаци  $ARIMA(p, d, q)$ , где је  $p$  ред ауторегресионе компоненте,  $d$  ниво интегрисаности временске серије и  $q$  ред компоненте покретних просека дефинисан је формулом:

$$\underbrace{(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)}_{\Phi(L)} (1 - L)^d X_t = \underbrace{(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)}_{\Theta(L)} \varepsilon_t.$$

Процес  $\varepsilon_t$  је процес белог шума,  $L$  је оператор померања  $L(X_t) = X_{t-1}$ , док полиноми  $\Phi(L)$  и  $\Theta(L)$  описују редом ауторегресиону компоненту и компоненту покретних просека стационарне временске серије  $(1-L)^d X_t$  и претпоставља се да не садрже заједничке факторе.

Специјални случајеви  $ARIMA(p, d, q)$  модела:

$$\begin{aligned} ARIMA(p, 0, 0) &= AR(p) \\ ARIMA(0, 0, q) &= MA(q) \\ ARIMA(p, 0, q) &= ARMA(p, q) \\ ARIMA(0, 0, 0) &= \text{бели шум.} \end{aligned}$$

### 3.5.4.3 Сезонски ARIMA модели

Временске серије које садрже сезонску компоненту називају се сезонске временске серије. Приликом моделирања сезонских временских серија користи се оператор сезонског диференцирања:

$$\Delta_s X_t = X_t - X_{t-s} = (1 - L^s) X_t.$$

Природа сезонске компоненте може бити детерминистичка и стохастичка, у зависности од тога да ли се сезонске варијације мењају током година или не. Стохастичка сезонска компонента описује се посебном врстом ARIMA модела, који се називају сезонски ARIMA модели. Уводи

се мултипликативни сезонски ARIMA модел  $ARIMA(p, d, q) \times (P, D, Q)_s$  дефинисан са:

$$(1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps})(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L^s)^D (1 - L)^d X_t = (1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_Q L^{Qs})(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t,$$

при чему је  $s$  период сезоне,  $\Phi_1, \dots, \Phi_P$  су параметри сезонске регресионе компоненте реда  $P$ ,  $\Theta_1, \dots, \Theta_Q$  су параметри сезонске компоненте покретних просека реда  $Q$ ,  $D$  ниво сезонског диференцирања, а остали параметри су као код ARIMA модела.

### 3.5.5 $\alpha$ - мешајући процес

**Дефиниција 3.20** Нека је  $(\Omega, \mathcal{F}, P)$  простор вероватноћа и нека је  $\mathcal{F}_i^j$ ,  $-\infty \leq i \leq j \leq \infty$  генерисано  $\sigma$ -поље над променљивим  $X_t$ , за  $i \leq t \leq j$  и  $t \in Z$ . Коefицијент зависности за било који позитиван број  $n$  дат је са

$$\alpha(n) = \sup_{A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+n}^\infty, m \in Z} |P(A)P(B) - P(A \cap B)|.$$

За стационарни процес  $\{X_t\}$  каже се да је  $\alpha$ -мешајући ако важи

$$\alpha(n) \rightarrow 0, \text{ када } n \rightarrow \infty.$$

На основу дате дефиниције закључујемо да  $\alpha$ -мешајући процес подразумева да се зависност између величина  $X_{t_1}$  и  $X_{t_2}$  приближава нули, када  $|t_1 - t_2|$  расте, тј. подразумева се асимптотска независност. ARMA( $p, q$ ) јесте  $\alpha$ -мешајући процес [5].

## 4 Бутстреп методе за временске серије

Већ описани стандардни бутстреп подразумева да су добијени подаци независни и једнако расподељени, међутим због брзог развоја статистике и њене употребе у економији, финансијама и другим областима јавља се потреба за све већом употребом зависних серија података, а најчешће су ти подаци представљени временским серијама. Употреба стандарног бутстрепа у анализи временских серија би нарушила зависност међу датим подацима, сходно томе дошло је до модификације стандарног бутстрепа како би се ова метода прилагодила зависним подацима. Најпопуларније бутстреп методе које се користе у раду са временским серијама су блок бутстреп (енгл. block bootstrap) и мрежаста бутстреп (енгл. sieve bootstrap).

### 4.1 Блок бутстреп

Прву бутстреп методу за временске серије увео је Карлштајн<sup>4</sup> 1986. године у раду [6], овај метод познат је као блок бутстреп метод без преклапања (енгл. non-overlapping block bootstrap). Он подразумева да су подаци подељени у блокове без преклапања, чиме се одржава зависност података унутар блока. Међутим, поделом података на блокове фиксне дужине може доћи до одбацивања последњих неколико опсервација, што може утицати негативно на тачност закључивања. Како би се надоместио овај проблем 1989. године уведен је покретни блок бутстреп (енгл. moving block bootstrap) [7]. У овој методи сваки блок почиње другом опсервацијом из претходног блока, што омогућава да сви подаци буду укључени у разматрање, неке опсервације ће бити укључене више пута, а прва и неколико последњих опсервација ће бити укључено само једном. Током 1994. године уводи се кружни бутстреп (енгл. circular bootstrap) [9], који све опсервације укључује у блокове више пута, уколико је дужина временске серије  $n$  биће толико и блокова. Све ове три методе блок бутстрепа подразумевају фиксну дужину блока. Друга врста блок бутстрепа је стационарни бутстреп (енгл. stationary bootstrap) који дели дати скуп података на блокове чије дужине прате геометријску расподелу са параметром  $p$ . У даљем раду описаћемо детаљније ове методе.

---

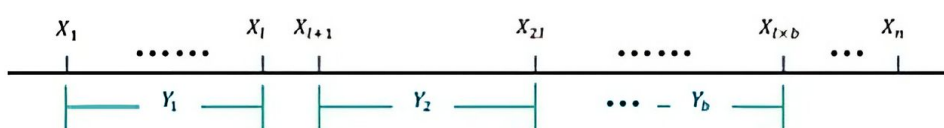
<sup>4</sup>Edward Carlstein, 1934.

### 4.1.1 Блок бутстреп без преклапања

Основна идеја ове методе је поделити дати скуп података дужине  $n$  на  $b$  независних блокова  $Y_i$  дужине  $l$ ,  $1 < l < n$  где је

$$Y_i = \{X_{(i-1)l+1}, \dots, X_{il}\}, i = 1, \dots, b.$$

Подела серије на блокове илустрована је на Слици 6. Случајним избором са понављањем бира се  $b$  блокова  $Y_1^*, \dots, Y_b^*$ , њиховим спајањем редоследом којим су изабрани добија се бутстреп временска серија  $X_1^*, \dots, X_m^*$ , где је  $m = lb$  дужина те серије.



Слика 6: Илустрација блок бутстрепа без преклапања

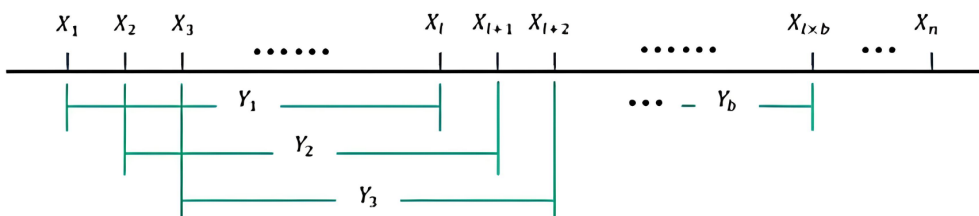
### 4.1.2 Покретни блок бутстреп

Покретни блок бутстреп, за разлику од блок бутстрепа без преклапања, дозвољава преклапање блокова, а тиме омогућава да се генерише већи број блокова.

Претпоставимо да је дата временска серија слабо стационарна. Под условом слабе стационарности може се претпоставити да се зависност података одржава унутар блока, а да се блокови могу сматрати независним. Након одабира дужине блока  $l$ , добија се  $N = n - l + 1$  независних блокова

$$Y_i = \{X_i, X_{i+1}, \dots, X_{i+l-1}\}, 1 \leq i \leq N.$$

Подела серије на блокове илустрована је на Слици 7. Од тако добијених  $N$  блокова, случајним избором са понављањем бирамо  $b = \lceil n/l \rceil$  блокова,  $Y_1^*, \dots, Y_b^*$ , и њиховим спајањем добијамо нову временску серију  $X_1^*, \dots, X_m^*$ , где је  $m = lb$ .



Слика 7: Илустрација покретног блок бутстрепа

**Теорема 4.1** *Претпоставимо да је низ  $\{X_t\}$ ,  $t = 1, \dots, n$ ,  $m$ -зависна стационарна временска серија тако да је  $EX_1 = \mu$  и  $E|X_1|^{4+\delta} < \infty$  за неко позитивно  $\delta$ . Ако  $\frac{l}{n} \rightarrow 0$ ,  $n \rightarrow \infty$  тада*

$$\sup_x |P^*\{\sqrt{m}(\bar{X}_m^* - E^*(\bar{X}_m^*)) \leq x\} - P\{\sqrt{n}(\bar{X}_n - \mu) \leq x\}| \xrightarrow{P} 0,$$

где је  $P^*$  бутстреп оцена вероватноће,  $E^*$  је бутстреп оцена средње вредности и  $\bar{X}_m^* = \frac{1}{m} \sum_{i=1}^m X_i^*$ . Ако се услов  $\frac{l}{n} \rightarrow 0$  може заменити са  $\frac{l}{\sqrt{n}} \rightarrow 0$  када  $n \rightarrow \infty$ , тада се  $E^*(\bar{X}_m^*)$  може заменити са  $\bar{X}_n$ .

У датој теорему  $m$ -зависност дефинисана је на следећи начин: за низ  $\{X_t\}_{t \in \mathbb{Z}}$  кажемо да је  $m$ -зависан за  $m \in \mathbb{N}_0$  ако за свако  $t$  важи да је низ  $\{X_s\}_{s \leq t}$  независан са  $\{X_s\}_{s \geq t+m+1}$ . Дата теорема показује постојаност методе за оцену средње вредности. Доказ теореме може се пронаћи у раду [7].

С обзиром на то да дата теорема показује постојаност покретног бутстрепа, како би се показала постојаност блок бутстрепа без преклапања довољно је показати да је разлика у перформансама између ове две методе занемарљива.

Нека је  $\{X_n\}_{n \in \mathbb{Z}}$  стационарна серија. Означимо са  $\chi_n = \{X_1, X_2, \dots, X_n\}$  низ од  $n$  опсервација дате серије.

Означимо са  $\bar{X}_m^{(1)}$  оцену средње вредности добијену методом бутстрепа без преклапања, тада је:

$$\bar{X}_m^{*(1)} = \frac{1}{m} \sum_{i=1}^m X_i^{*(1)}.$$

Приметимо да узорковани блокови  $Y_1^{*(1)}, \dots, Y_b^{*(1)}$  представљају независне и једнако расподељене  $l$ - димензионе случајне векторе са вероватноћом избора сваког блока  $\frac{1}{b}$ , тј. важи:

$$P\{(X_1^{*(1)}, \dots, X_l^{*(1)}) = (X_{(i-1)l+1}, \dots, X_{il}) | \chi_n\} = \frac{1}{b}, 1 \leq i \leq b.$$

Одавде може се извести очекивање оцене  $\overline{X}_m^{*(1)}$ , тј. важи:

$$\begin{aligned} E(\overline{X}_m^{*(1)} | \chi_n) &= E\left(\frac{1}{l} \sum_{i=1}^l X_i^{*(1)} | \chi_n\right) \\ &= \frac{1}{b} \sum_{j=1}^b \left(\frac{1}{l} \sum_{i=1}^l X_{(j-1)l+i}\right) \\ &= \frac{1}{bl} \left(\sum_{i=1}^n X_i - \sum_{j=bl+1}^n X_j\right). \end{aligned}$$

Означимо са  $\overline{X}_m^{*(2)}$  оцену средње вредности добијену методом покретног бутстрепа, тада је:

$$\overline{X}_m^{*(2)} = \frac{1}{m} \sum_{i=1}^m X_i^{*(2)}.$$

Као и код блок бутстрепа без преклапања, и овде важи да узорковани блокови  $Y_1^{*(2)}, \dots, Y_b^{*(2)}$  представљају независне и једнако расподеле  $l$ -димензионе случајне векторе са вероватноћом избора сваког блока  $\frac{1}{N}$ :

$$P\{(X_1^{*(2)}, \dots, X_l^{*(2)}) = (X_i, \dots, X_{i+l-1}) | \chi_n\} = \frac{1}{N}, 1 \leq i \leq N.$$

Одавде може се извести очекивање оцене  $\overline{X}_m^{*(2)}$ , тј. важи:

$$\begin{aligned} E(\overline{X}_m^{*(2)} | \chi_n) &= E\left(\frac{1}{l} \sum_{i=1}^l X_i^{*(2)} | \chi_n\right) \\ &= \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{l} \sum_{i=1}^l X_{j+i-1}\right) \\ &= \frac{1}{N} \left(\sum_{i=1}^n X_i - \frac{1}{l} \sum_{j=1}^{l-1} (l-j)(X_j + X_{n-j+1})\right). \end{aligned}$$

Приметимо да бутстреп оцене средње вредности имају различита очекивања у зависности од методе. У раду [8], под претпоставком да процес  $\{X_n\}_{n \in \mathbb{Z}}$  задовољава услове  $\alpha$ -мешајућег процеса, изведен је закључак  $E\{E(\overline{X}_m^{*(2)} | \chi_n) - E(\overline{X}_m^{*(1)} | \chi_n)\}^2 = O(\frac{l}{n^2})$ , што показује да за велике обиме датог узорка постоји занемарљива разлика између ове две методе, а тиме је показана постојаност блок бутстрепа без преклапања.

### 4.1.3 Кружни блок бутстреп

Кружни блок бутстреп уводи се како би се искористили сви подаци дате временске серије већи број пута. Нпр. у покретном блок бутстрепу прва вредност појавиће се само у првом блоку, док ће се пета вредност појавити у пет блокова, уколико је број блокова већи или једнак од пет, док у блок бутстрепу без преклапања последњих неколико елемената се неће укључити у блокове, уколико дужина блока није делилац дужине дате временске серије.

Овај метод, такође дозвољава преклапање блокова, али дати низ података треба продужити, како би се направили додатни блокови. Продужење дате временске серије дефинисано је са:

$$X_i \equiv X_j \text{ за } i > n, \text{ где је } j = i \pmod{n} \text{ и } X_0 = X_n$$

и конструише се  $n$  блокова.

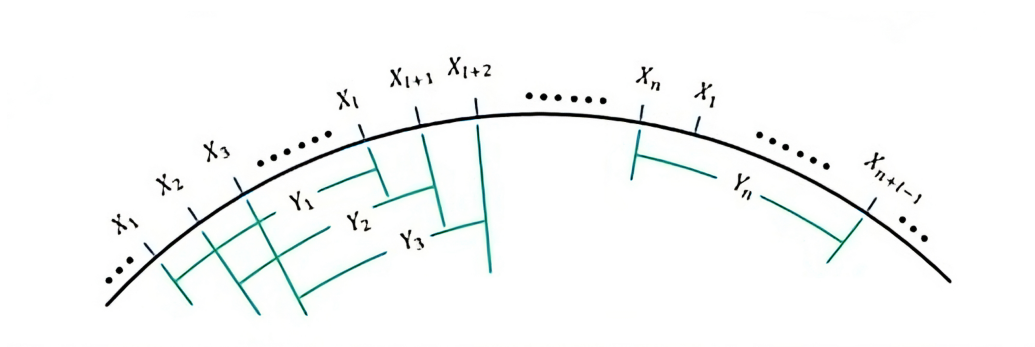
Нека је дата серија  $X_1, \dots, X_n$ , цео низ опсервација које учествују у конструисању блокова одређен је са

$$X_1, \dots, X_n, X_1, \dots, X_{n+l-1},$$

док је сваки блок дефинисан са

$$Y_i = \{X_i, X_{i+1}, \dots, X_{i+l-1}\},$$

где је  $l$  дужина блока. Подела серије на блокове илустрована је на Слици 8. Излачењем са враћањем бира се  $b$  блокова. Спајањем блокова добија се нова серија  $X_1^*, \dots, X_m^*$ , где је  $m = lb$ .



Слика 8: Илустрација кружног блок бутстреп

Следећа теорема, доказана у раду [9], показује постојаност и асимптотску тачност кружног блок бутстреп за  $\alpha$ -мешајућег процесе.

**Теорема 4.2** Нека је  $\{X_t\}$ ,  $t = 1, \dots, n$  слабо стационарна временска серија, уз претпоставку да важи услов  $\alpha$ -мешајућег процеса ( $\alpha(h) \rightarrow 0$  када  $h \rightarrow 0$ ). Претпоставимо да је  $E|X_t|^{6+\delta} < \infty$ , за неко позитивно  $\delta$  и  $\sum_{h=1}^{\infty} n^2(\alpha(h))^{\frac{\delta}{6+\delta}} < \infty$ . Додатно важи за  $n \rightarrow \infty$ ,  $\frac{m}{n} \rightarrow 1$ , где је  $m = lb$  и за  $l \rightarrow \infty$ ,  $\frac{b}{n} \rightarrow 0$ . Тада је  $\sigma_n^2 \equiv \text{Var}(\sqrt{n}\bar{X}_n)$  ограничено са  $\sigma_\infty^2 < \infty$  и  $\text{Var}^*(\sqrt{m}\bar{X}_m^*) \xrightarrow{P} \sigma_\infty^2$ , где је  $\text{Var}^*$  бутстреп оцена дисперзије статистике  $\sqrt{m}\bar{X}_m^*$ , и још важи

$$\sup_x |P^*\{\sqrt{m}(\bar{X}_m^* - \bar{X}_n) \leq x\} - P\{\sqrt{n}(\bar{X}_n - \mu) \leq x\}| \xrightarrow{P} 0.$$

#### 4.1.4 Стационарни бутстреп

Стационарни блок бутстреп представили су Политис и Романо<sup>5</sup> 1994. године [10]. Овај метод користи се за слабо стационарне временске серије и генерише реузорке временских серија као стационарне временске серије, тј. ова метода тежи ка томе да се очува стационарност оригиналне временске серије у реузорцима. За разлику од претходно описаних метода које подразумевају да је дужина блока фиксна, овде је блок произвољне дужине, а дужина блока се може представити неком расподелом. Обично се за дужину блока бира геометријска расподела. Претпоставимо да је  $\{X_t\}_{t \in \mathbb{Z}}$  дата слабо стационарна временска серија. Дефинишимо блокове са

$$B_{i,l} = \{X_i, X_{i+1}, \dots, X_{i+l-1}\},$$

где је  $l$  дужина блока. С обзиром на то да насумично бирамо почетну тачку сваког блока, а дужина блока има геометријску расподелу, избор опсервација за неке блокове може надмашити дужину дате временске серије. Како би се ово избегло користи се исти принцип као у кружном бутстрепу, где се временска серија продужује на исти начин:

$$X_i \equiv X_j \text{ за } i > n, \text{ где је } j = i \pmod{n} \text{ и } X_0 = X_n.$$

Дужине блокова  $l_i$  представљају низ независних и једнако расподељених вредности са геометријском расподелом

$$P\{l_i = u\} = (1-p)^{u-1}p,$$

за  $u = 1, 2, \dots$  и за дату вредност  $p$ . Индекс почетне опсервације сваког блока  $i_i$  бира се независно од  $X_i$  и  $l_i$  случајним избором из скупа

<sup>5</sup>Dimitris N. Politis и Joseph P. Romano



$\{1, \dots, n\}$ , тј. низ  $\{i_i\}$  је низ независних једнако расподељених вредности са дискретном униформном расподелом на скупу  $\{1, \dots, n\}$ .

Дакле, реузорковани низ блокова је

$$\{B_{i_1, l_1}, B_{i_2, l_2}, \dots\} = \{X_{i_1}, X_{i_1+1}, \dots, X_{i_1+l_1-1}, X_{i_2}, X_{i_2+1}, \dots, X_{i_2+l_2-1}, \dots\}.$$

Следећа теорема, доказана у раду [10], показује постојаност и асимптотску тачност стационарног блок бутстрепа за  $\alpha$ -мешајуће процесе.

**Теорема 4.3** *Нека је  $X_t$ ,  $t = 1, 2, \dots$  строго стационарни процес са аутоковаријационом функцијом  $\gamma(\cdot)$  која задовољава услов*

$$\gamma(0) + \sum_r (r\gamma(r)) < \infty.$$

*Претпоставимо да важи*

$$\sum_{u,v,w} |k_4(u, v, w)| = K < \infty,$$

где је  $k_4(u, v, w)$  четврти заједнички кумулант расподеле за  $(X_i, X_{i+u}, X_{i+v}, X_{i+u+v+w})$  [12]. Претпоставимо да важи  $E|X|^{2+\delta} < \infty$  и  $\sum_h [\alpha(h)]^{\frac{\delta}{2+\delta}} < \infty$ , за неко позитивно  $\delta$ . Тада је гранична дисперзија

$$\sigma_\infty^2 = \text{Var}(X_1) + \sum_{i=1}^{\infty} \text{cov}(X_1, X_{1+i})$$

коначна. Ако је  $\sigma_\infty > 0$ , тада важи

$$\sup_x |P\{\sqrt{n}(\bar{X}_n - \mu) \leq x\} - \Phi(x/\sigma_\infty)| \rightarrow 0,$$

где је  $\Phi$  стандардизована нормална расподела.

Уколико још важи да  $p \rightarrow 0$  и  $np \rightarrow \infty$  тада је бутстреп расподела блиска правој расподели, тј. важи:

$$\sup_x |P^*\{\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x\} - P\{\sqrt{n}(\bar{X}_n - \mu) \leq x\}| \xrightarrow{P} 0.$$

#### 4.1.5 Симулација метода

У даљем раду спровешћемо истраживање како бисмо видели перформансе описаних модела блок бутстрепа за AR(2) модел:  $x_t = 0.2x_{t-1} + 0.4x_{t-2} + \varepsilon_t$ , где је  $\varepsilon_t$  је процес белог шума са  $N(0, 1)$  расподелом. Искористићемо ове методе да одредимо оцену средње

вредности дате серије, аутокорелационих коефицијената и параметара модела и њихове интервале поверења, као и пристрасност и стандардну девијацију добијених оцена. Резултати неких симулација за МА(2) модел:  $x_t = 0.2\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + \varepsilon_t$  дати су у поглављу Додатак 2.

Симулација метода подразумева:

1. генеришемо оригиналну временску серију AR(2) модела функцијом `arima.sim` програмског језика **R**,
2. генеришемо  $R = 1000$  бутстреп временских серија једном од метода на следећи начин:
  - дату временску серију поделимо на одређен број блокова у зависности од методе,
  - случајним избором са понављањем бирамо  $b$  блокова,
  - на основу изабраних блокова формирамо бутстреп временску серију,
3. за сваку бутстреп серију рачунамо оцену посматраног параметра и добијамо  $R$  вредности оцене параметра:  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_R^*$ ,
4. рачунамо средњу вредност добијених оцена, што представља бутстреп оцену посматраног параметра:  $\bar{\theta}^* = \frac{1}{R} \sum_{i=1}^n \hat{\theta}_i^*$ ,
5. рачунамо 95% Ефронов перцентилни интервал поверења, описан у одељку 2.4,
6. рачунамо стандардну девијацију добијених оцена параметра на следећи начин:

$$sd(\hat{\theta}) = \sqrt{\frac{1}{R-1} \sum_{i=1}^n (\hat{\theta}_i^* - \bar{\theta}^*)^2},$$

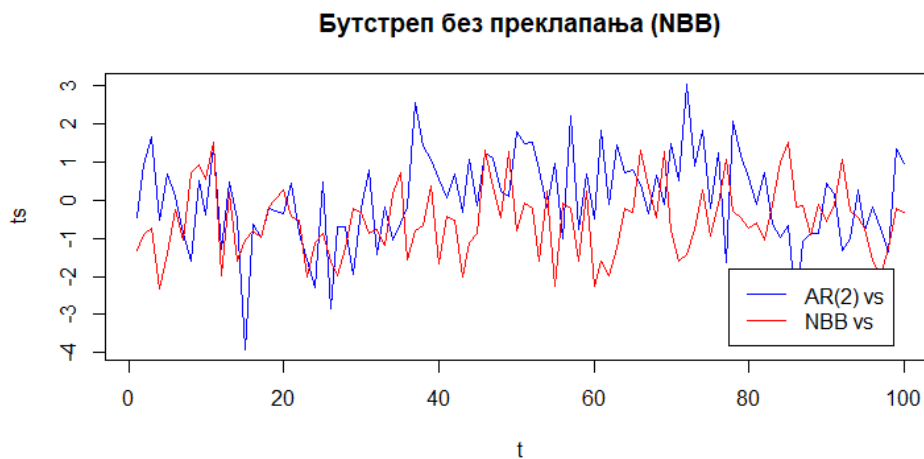
7. рачунамо пристрасност бутстреп оцене параметра на следећи начин:

$$Bias(\hat{\theta}) = \bar{\theta}^* - \hat{\theta},$$

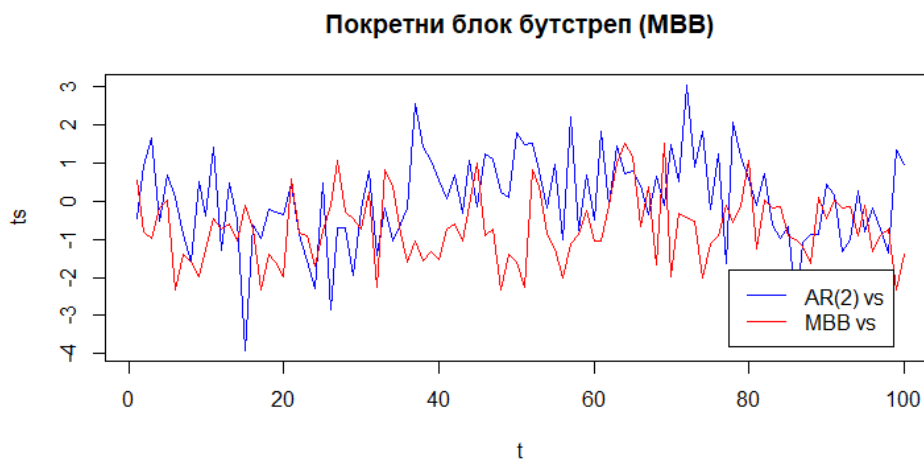
где је  $\hat{\theta}$  оцена параметра добијена на основу оригиналне временске серије.

На основу описаних блок бутстреп метода генерисане су функције за сваку методу у програмском језику **R**, у случају када је обим оригиналне временске серије  $n$  дељив са дужином блока  $l$ . Имплементација

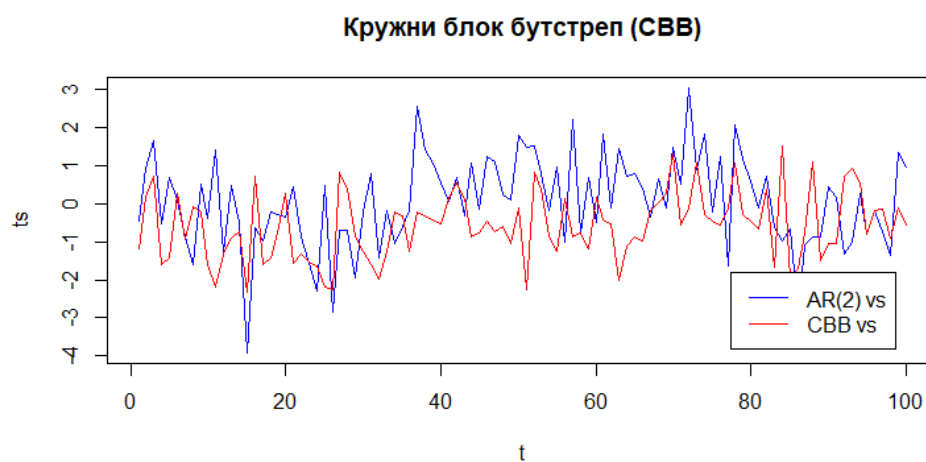
функција налази се у поглављу Додатак 1. Свака функција враћа вектор који представља једну бутстреп временску серију. Како бисмо упоредили оригиналну временску серију и бутстреп серију, представићемо графички серију датог AR(2) модела дужине  $n = 100$  и једну бутстреп серију добијену блок бутстрепом за дужину блока  $l = 5$ , за сваку од метода (в. Сliku 9, Сliku 10, Сliku 11, Сliku 12).



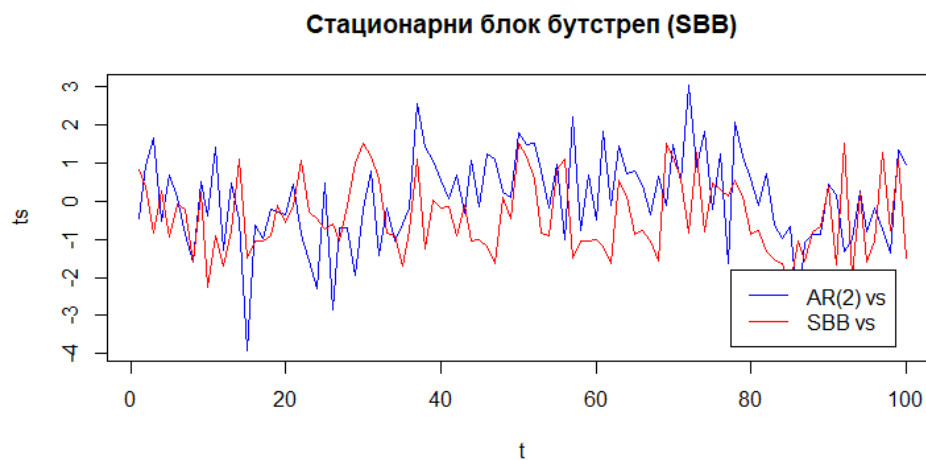
Слика 9: Графички приказ оригиналне временске серије и бутстреп временске серије добијене бутстрепом без преклапања



Слика 10: Графички приказ оригиналне временске серије и једне бутстреп временске серије добијене покретним блок бутстрепом



Слика 11: Графички приказ оригиналне временске серије и једне бутстреп временске серије добијене кружним блок бутстрепом



Слика 12: Графички приказ оригиналне временске серије и једне бутстреп временске серије добијене стационарним блок бутстрепом

Табела 1 садржи средње вредности и стандардне девијације приказане оригиналне временске серије и бутстреп серија.

	средња вредност	стандардна девијација
оригинална серија	0.04	1.19
NBB серија	-0.54	0.92
MBB серија	-0.71	0.90
CBV серија	-0.58	0.89
SBB серија	-0.41	0.98

Табела 1: Средња вредност и стандардна девијација оригиналне серије и бутстреп серија

#### 4.1.5.1 Оцена средње вредности модела блок бутстреп методом

Упоредићемо бутстреп интервале поверења средње вредности на основу просечне покривености интервала поверења, тј. одређујемо просечан број интервала поверења који садрже праву вредност средине серија, кроз 500 симулација. Током 500 симулација генеришемо 1000 бутстреп серија на основу којих одредимо 95% интервал поверења, и за сваки интервал проверимо да ли средња вредност оригиналне серије, тј. нула припада добијеном интервалу, а добијени број интервала делимо са 500 и тако добијамо просечну покривеност интервала поверења. Разматраћемо све четири методе блок бутстрепа за дужину узорка  $n = 512$  и различите дужине блока  $l = 4, l = 8, l = 16, l = 32$  (у случају стационарног блок бутстрепа приказана је очекивана дужина блока, што заправо значи да се  $r$  вредност геометријске расподеле рачуна као  $1/l$ ). Добијени резултати дати су табелом Табела 2.

модел	$l = 4$	$l = 8$	$l = 16$	$l = 32$
NBB	0.806	0.854	0.908	0.882
MBB	0.786	0.860	0.9	0.88
CBV	0.796	0.884	0.922	0.90
SBB	0.83	0.874	0.912	0.902

Табела 2: Просечна покривеност интервала поверења за средњу вредност узорка за  $n = 512$  и  $l = 4, 8, 16, 32$

Упоредићемо и како величина узорка утиче на просечну покривеност интервала поверења. Дужина блока је фиксирана  $l = 8$ , а разматраћемо различите обиме оригиналне серије  $n = 64, 200, 800$ . Добијени резултати приказани су табелом Табела 3.

$l = 8$	$n = 64$	$n = 200$	$n = 800$
NBB	0.83	0.826	0.874
MBB	0.8	0.862	0.888
CBV	0.818	0.846	0.878
SBB	0.794	0.824	0.892

Табела 3: Просечна покривеност интервала поверења за средњу вредност узорка за  $l = 8$  и  $n = 64, 200, 800$

У наредној табели приказане су бутстреп оцене средње вредности, пристрасност и стандардна девијација за обим узорка  $n = 512$  и дужине блока  $l = 4, 8, 16, 32$ .

		$l = 4$	$l = 8$	$l = 16$	$l = 32$
NBB	$\bar{x}^*$	0.205	0.204	0.203	0.202
	$bias(\bar{x}^*)$	-0.0003	-0.0018	-0.0027	-0.0034
	$sd(\bar{x}^*)$	0.0669	0.0883	0.0908	0.0942
MBB	$\bar{x}^*$	0.206	0.209	0.215	0.198
	$bias(\bar{x}^*)$	0.0007	0.0034	0.0095	-0.0072
	$sd(\bar{x}^*)$	0.0683	0.0816	0.0889	0.0981
CBV	$\bar{x}^*$	0.205	0.204	0.208	0.204
	$bias(\bar{x}^*)$	-0.0005	-0.0016	0.0025	-0.0019
	$sd(\bar{x}^*)$	0.0707	0.0791	0.0867	0.0977
SBB	$\bar{x}^*$	0.202	0.201	0.204	0.201
	$bias(\bar{x}^*)$	-0.0035	-0.0042	-0.0018	-0.0047
	$sd(\bar{x}^*)$	0.0735	0.0857	0.0959	0.0995

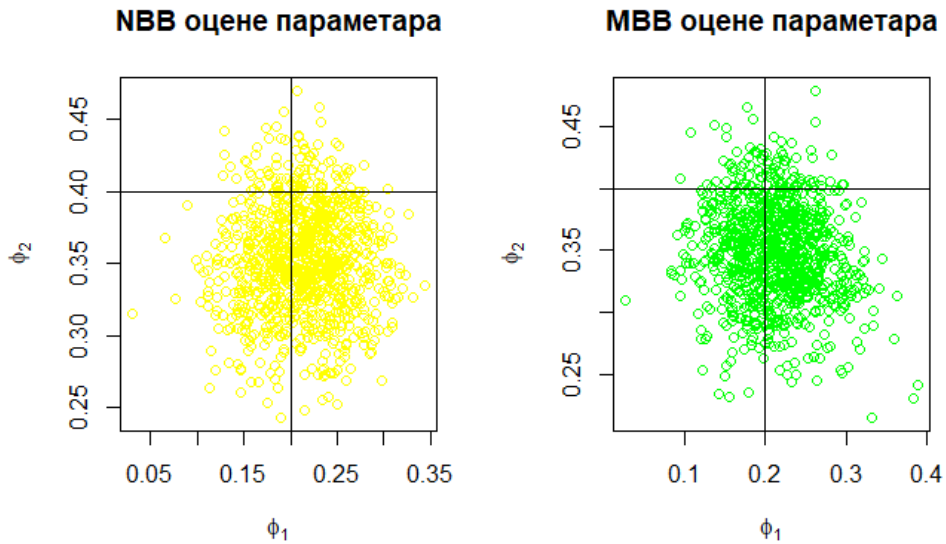
Табела 4: Бутстреп оцене средње вредности

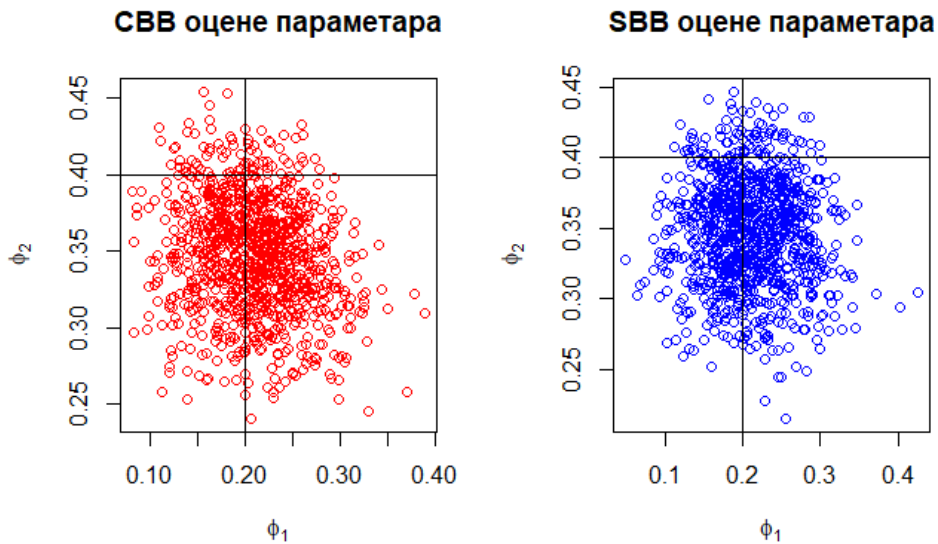
На основу добијаних резултата можемо закључити:

- ✓ за дужину блока  $l = 16$  добија се највећа просечна покривеност интервала,
- ✓ просечна покривеност интервала поверења битно зависи од дужине узорка и дужине блока,
- ✓ пристрасност варира у зависности од методе и дужине блока, јавља се и потцењивање и прецењивање средње вредности у односу на средњу вредност оригиналне временске серије,
- ✓ стандардна девијација расте са повећањем дужине блока,
- ✓ највећу стандардну девијацију постиже стационарни блок бутстреп.

#### 4.1.5.2 Оцена параметара модела блок бутстреп методом

У овом делу упоредићемо описане блок бутстреп методе за оцену параметара модела. Под претпоставком да је ред модела два, уклопићемо бутстреп серије у одговарајући AR(2) модел функцијом `arima` програмског језика **R**. Посматраћемо серију обима  $n = 512$  и за дужину блока узимамо  $l = 16$ . На слици испод представљене су добијене оцене параметара бутстреп серија за све блок бутстреп методе. Црне линије на слици означавају тачне вредности параметара.





Слика 13: Оцене параметара бутстреп временских серија

Резиме резултата у смислу бутстреп оцене параметара, интервала поверења, пристрасности и стандардне девијације дати су наредном табелом.

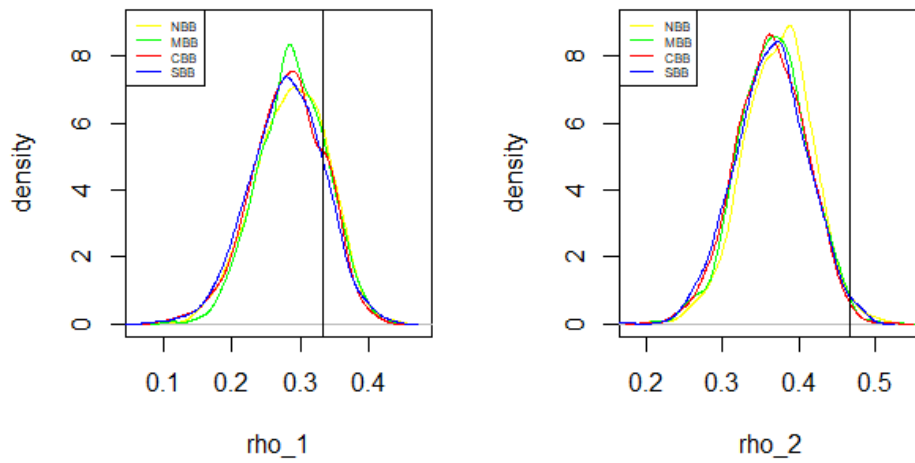
Модел	NBV	MBV	CBV	SBB
$\hat{\phi}_1^*$	0.198 (0.1299,0.2672)	0.202 (0.1249,0.2767)	0.193 (0.1200,0.2667)	0.193 (0.1221,0.2692)
$\hat{\phi}_2^*$	0.314 (0.2322,0.3920)	0.304 (0.2109,0.3889)	0.307 (0.2217,0.3886)	0.310 (0.2288,0.3881)
$bias(\hat{\phi}_1^*)$	0.0023	0.0058	-0.0026	-0.0031
$bias(\hat{\phi}_2^*)$	-0.0488	-0.0584	-0.0555	-0.0527
$sd(\hat{\phi}_1^*)$	0.0350	0.0378	0.0366	0.0373
$sd(\hat{\phi}_2^*)$	0.0405	0.0438	0.0428	0.0408

Табела 5: Бутстреп оцене параметара модела

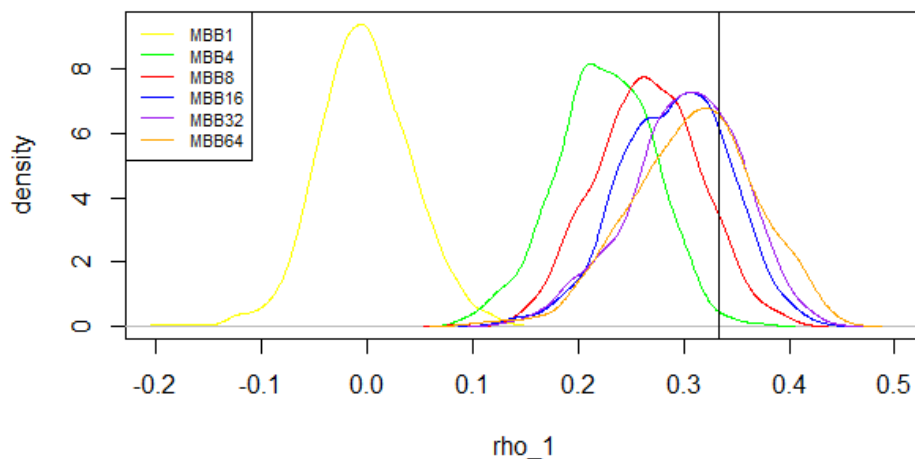
Да бисмо уочили истовремено ефекат процене оба параметра можемо погледати коефицијенте колерације. Теоријске аутокорелационе коефицијенте рачунамо по већ датим формулама:  $\rho_1 = \frac{\phi_1}{1-\phi_2} = 0.333$ ,  $\rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2 = 0.467$ . Из процене параметара за сваку од 1000 бутстреп серија можемо извести процену аутокорелационих коефицијената  $\rho_1$  и  $\rho_2$



на основу датих једначина. Добијене вредности представљене су наредним графиком, где црна вертикална линија означава теоријску вредност аутокорелационих коефицијената.



Слика 14: Оцена коефицијената корелације са кашњењем 1 и 2

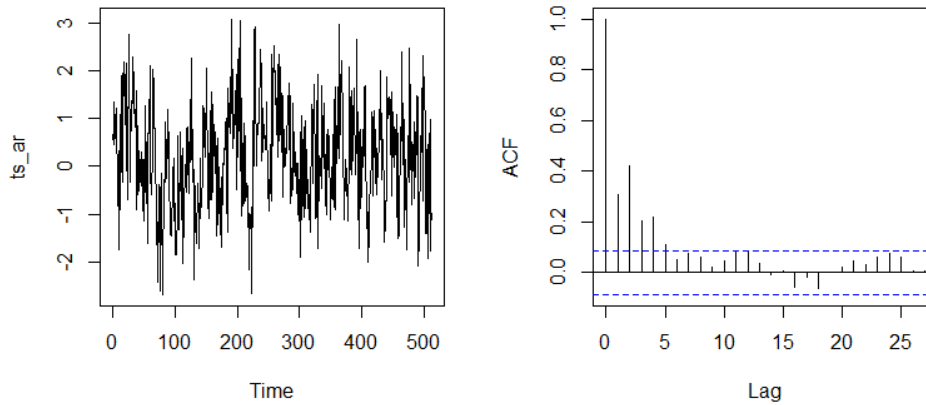


Слика 15: Оцена коефицијента корелације са кашњењем 1 за различите дужине блока

#### 4.1.5.3 Оцена аутокорелационих коефицијената модела блок бутстреп методом

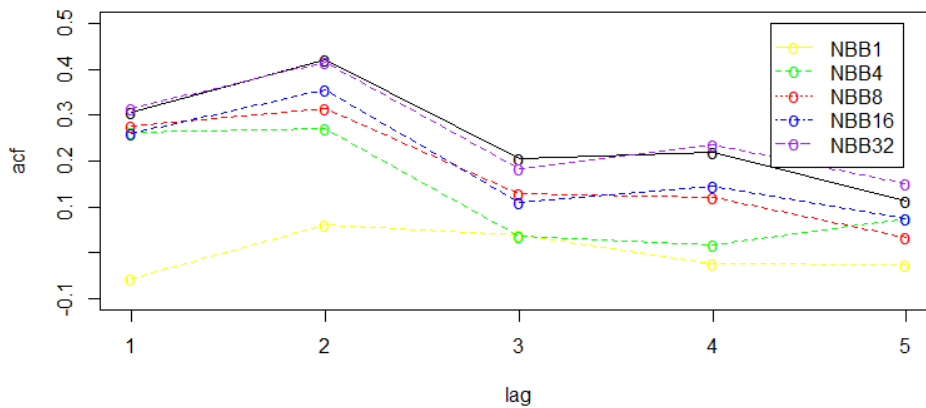
Како бисмо увидели шта се дешава са корелацијом код бутстреп серија упоредићемо аутокорелационе коефицијенте бутстреп серија и оригиналне серије. Оцене аутокорелационих коефицијената добићемо функцијом `acf` програмског језика **R**.

Погледајмо прво узорачки корелограм серије AR модела дужине  $n = 512$  (в. Слика 16).

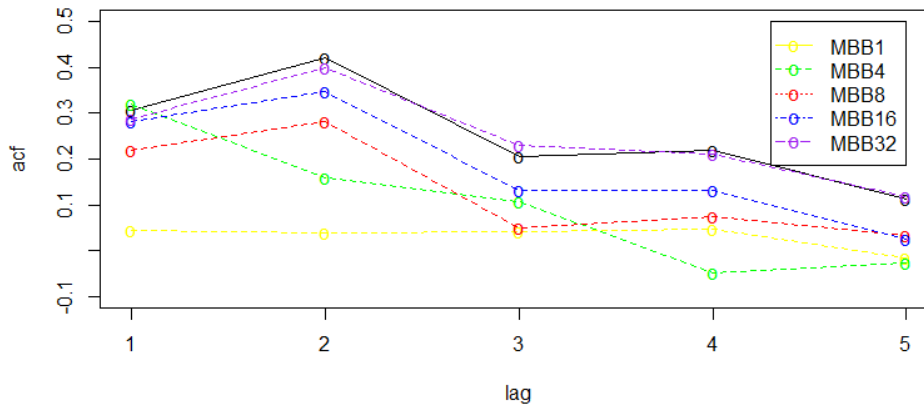


Слика 16: Графички приказ AR серије и њен узорачки корелограм

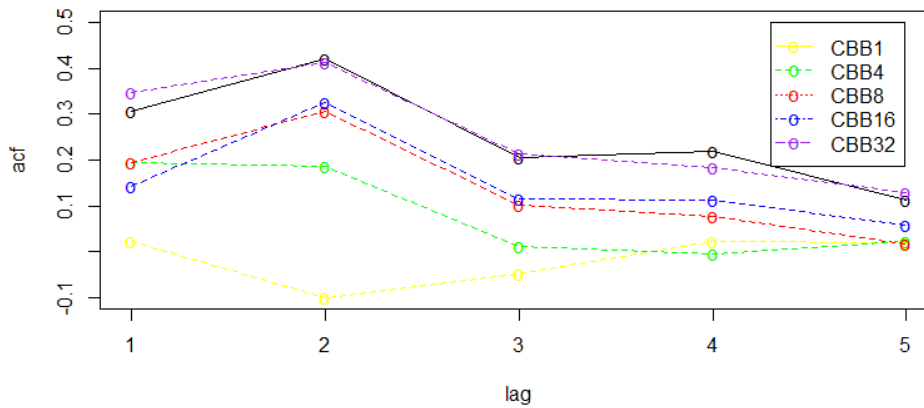
На основу датог узорачког корелограма можемо да приметимо да су аутокорелациони коефицијнти са кашњењем већим од пет безначајни, тако да ћемо упоређивати првих пет коефицијената корелације. Даље генеришемо по једну бутстреп серију за различите дужине блокова и упоредићемо оцене њихових аутокорелационих коефицијената на основу датих графика, за сваку од метода појединачно.



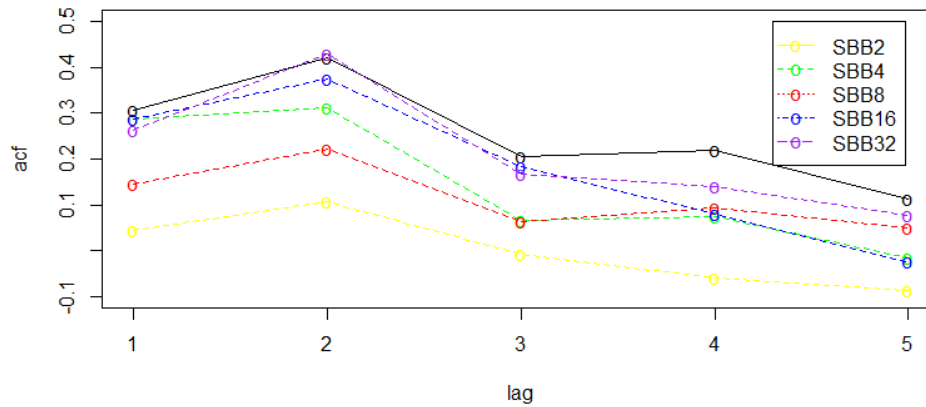
Слика 17: Вредност првих пет аутокорелационих коефицијената ( $\text{lag} = 1, 2, 3, 4, 5$ ) за бутстреп временске серије блок бутстрепа без преклапања за дужине блока  $l = 1, 4, 8, 16, 32$ .



Слика 18: Вредност првих пет аутокорелационих коефицијената ( $\text{lag} = 1, 2, 3, 4, 5$ ) за бутстреп временске серије покретног блок бутстрепа за дужине блока  $l = 1, 4, 8, 16, 32$ .

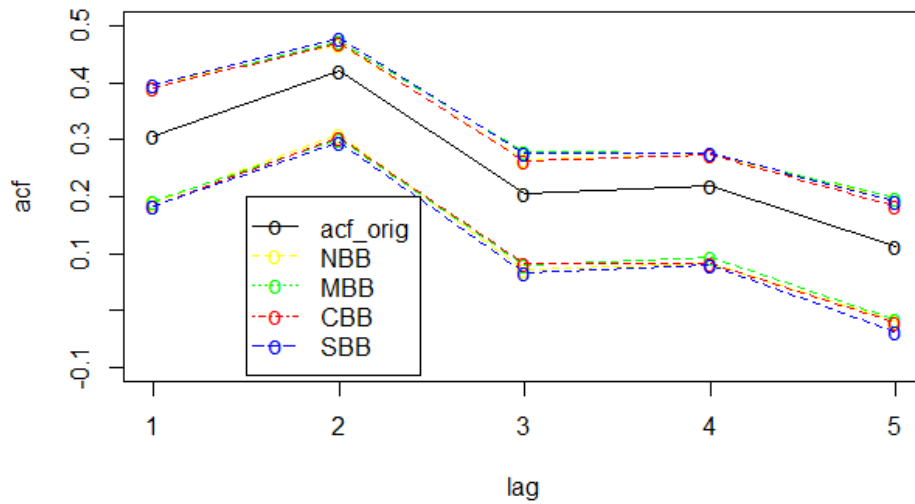


Слика 19: Вредност првих пет аутокорелационих коефицијената ( $\text{lag} = 1, 2, 3, 4, 5$ ) за бутстреп временске серије кружног блок бутстрепа за дужине блока  $l = 1, 4, 8, 16, 32$ .



Слика 20: Вредност првих пет аутокорелационих коефицијената ( $\text{lag} = 1, 2, 3, 4, 5$ ) за бутстреп временске серије стационарног блок бутстрепа за очекиване дужине блока  $l = 2, 4, 8, 16, 32$ .

Као што можемо приметити на основу датих графика, корелација се код бутстреп узорака смањила, али и видимо да се повећава са повећањем дужине блока, можемо претпоставити да је дужина блока најбитнији фактор тачности бутстреп оцена аутокорелационих коефицијената. Приметимо и да је корелација временских серија са дужином блока један близу нуле, а то показује зашто стандардни бутстреп није погодно користити за временске серије. За даљи рад узимамо да је дужина блока 32. Погледајмо прво бутстреп интервале поверења (в. Слика 21).



Слика 21: Бутстреп интервали поверења аутокорељационих коефицијената са кашњењем 1,2,3,4,5.

Резиме резултата у смислу бутстреп оцене аутокорељационих коефицијената, пристрасности и стандардне девијације дати су наредном табелом.

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
NBB	$\hat{\rho}_k^*$	0.293	0.392	0.173	0.186	0.092
	$bias(\hat{\rho}_k^*)$	-0.014	-0.030	-0.033	-0.033	-0.021
	$sd(\hat{\rho}_k^*)$	0.051	0.039	0.049	0.049	0.055
MBB	$\hat{\rho}_k^*$	0.299	0.395	0.184	0.190	0.095
	$bias(\hat{\rho}_k^*)$	-0.007	-0.027	-0.022	-0.030	-0.017
	$sd(\hat{\rho}_k^*)$	0.052	0.044	0.051	0.047	0.054
CBB	$\hat{\rho}_k^*$	0.289	0.389	0.178	0.180	0.085
	$bias(\hat{\rho}_k^*)$	-0.018	-0.032	-0.029	-0.039	-0.027
	$sd(\hat{\rho}_k^*)$	0.052	0.043	0.048	0.049	0.052
SBB	$\hat{\rho}_k^*$	0.288	0.387	0.177	0.180	0.088
	$bias(\hat{\rho}_k^*)$	-0.019	-0.035	-0.029	-0.039	-0.025
	$sd(\hat{\rho}_k^*)$	0.054	0.047	0.052	0.050	0.060

Табела 6: Бутстреп оцене аутокорељационих коефицијената

## 4.2 Решетки бутстреп

Поред блок бутстрепа постоји и решетки бутстреп (други називи су мрежасте или ситасти) бутстреп (енгл. AR-sieve bootstrap), који не дели податке на блокове, већ користи модел који је одређен на основу оригиналног скупа података и помоћу добијеног модела генерише нове податке који се користе за процену параметара.

Нека је  $\{X_1, X_2, \dots, X_n\}$  оригинална временска серија. Решетки бутстреп у основи захтева апроксимацију модела  $AR(\hat{p})$ :

$$X_t - \mu = \sum_{i=1}^{\hat{p}} \phi_i (X_{t-i} - \mu) + \varepsilon_t,$$

где је  $\mu = E(X_t)$  и  $\varepsilon_t$  процес белог шума. Апроксимација модела захтева оцену параметра  $p$ . Параметар  $p$  најчешће се оцењује помоћу АИС критеријума.

Други корак у апроксимацији моделе јесте оцена параметара модела. Нека је  $\hat{\phi}_{\hat{p}} = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_{\hat{p}})$  Јул-Вокерова оцена параметара. Тада је

$$\hat{\phi}_{\hat{p}} = \hat{\Gamma}_{\hat{p}}^{-1} \hat{\gamma}_{\hat{p}},$$

где је  $\hat{\Gamma}_{\hat{p}} = [\hat{\gamma}(i-j)]_{i,j=1}^{\hat{p}}$ ,  $\hat{\gamma}_{\hat{p}} = (\hat{\gamma}(1), \hat{\gamma}(2), \dots, \hat{\gamma}(\hat{p}))'$  и  $\hat{\gamma}$  је оцена аутоковаријационе функције дефинисана са

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=k+1}^n (X_t - \bar{X}_n)(X_{t-k} - \bar{X}_n), 0 \leq k \leq \hat{p}.$$

Како би се одржале особине оригиналног скупа података даље треба реузорковати резидуале. Нека је

$$\hat{\varepsilon}_t = X_t - \sum_{i=1}^{\hat{p}} \hat{\phi}_i X_{t-i}, t = \hat{p} + 1, \dots, n,$$

оцена резидуала добијена на основу одређеног модела, тада можемо да дефинишемо емпиријску функцију расподеле резидуала са:

$$F_{\hat{\varepsilon}_t}(x) = \frac{1}{n - \hat{p}} \sum_{t=\hat{p}+1}^n I\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}_t \leq x\},$$

где је

$$\bar{\hat{\varepsilon}}_t = \frac{1}{n - \hat{p}} \sum_{t=\hat{p}+1}^n \hat{\varepsilon}_t.$$

На основу добијене емпиријске функције расподеле вршимо реузорковање резидуала  $\varepsilon_t^*$ .

Последњи корак јесте конструисати реузорке временске серије:

$$X_t^* = \bar{X}_n + \sum_{i=1}^{\hat{p}} \hat{\phi}_i(X_{t-1}^* - \bar{X}_n) + \varepsilon_t^*.$$

У пракси се често за почетне опсервације реузорковане серије узима да су оне једнаке нули или  $\bar{X}_n$ , након генерисања потпуне серије на основу процењеног модела почетне вредности се бришу.

Да бисмо показали постојаност ове методе размотрићемо пажљивије процес  $\{X_t\}_{t \in \mathbb{Z}}$  и дати прецизне претпоставке о овом процесу.

Претпоставимо да је процес  $\{X_t\}_{t \in \mathbb{Z}}$  стационарни процес са очекивањем  $EX_t = \mu$ . Према Волдовој теореме разлагања, стохастички процес  $\{X_t\}_{t \in \mathbb{Z}}$  може се записати као  $MA(\infty)$  процес

$$X_t - \mu = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i},$$

где је  $\psi_0 = 1$ ,  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  низ некорелисаних променљивих и  $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ .

Под додатним условом инвертибилности, процес  $\{X_t\}$  може се записати као  $AR(\infty)$  модел

$$\sum_{i=0}^{\infty} \phi_i (X_{t-i} - \mu) = \varepsilon_t,$$

где је  $\psi_0 = 1$  и  $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ .

Запишимо

$$\Phi(z) = \sum_{i=0}^{\infty} \phi_i z^i, \quad \phi_0 = 1, \quad z \in \mathbb{C},$$

$$\Psi(z) = \sum_{i=0}^{\infty} \psi_i z^i, \quad \psi_0 = 1, \quad z \in \mathbb{C}.$$

На основу дате репрезентације модели  $MA(\infty)$  и  $AR(\infty)$  могу се, редом записати на следећи начин

$$\Phi(B)(X - \mu) = \varepsilon, \quad X - \mu = \Psi(B)\varepsilon,$$

где је  $B$  оператор померања,  $B(X_t) = X_{t-1}$ ,  $x \in \mathbb{R}^{\mathbb{Z}}$ .



Да бисмо увели додатне претпоставке модела означимо са  $\mathcal{F}_t = \sigma(\{\varepsilon_s, s \leq t\})$ ,  $\sigma$  - поље генерисано над низом  $\{\varepsilon_s\}_{s=-\infty}^t$ .

Да бисмо показали постојаност методе уводимо следеће претпоставке:

**Претпоставка 1:**  $X_t - \mu = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ ,  $\psi_0 = 1$ ,  $t \in \mathbb{Z}$ , где је  $\{\varepsilon_t\}$  ергодичан стационаран процес,  $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ ,  $E[\varepsilon_t^2 | \mathcal{F}_{t-1}] = \sigma^2 < \infty$ , и  $E|\varepsilon_t|^s < \infty$ ,  $s \geq 4$ .

**Претпоставка 1':**  $X_t - \mu = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ ,  $\psi_0 = 1$ ,  $t \in \mathbb{Z}$ , где је  $\{\varepsilon_t\}$  низ независних једнако расподелених случајних променљивих,  $E[\varepsilon_t] = 0$  и  $E|\varepsilon_t|^s < \infty$ ,  $s \geq 4$ .

**Претпоставка 2:** Постоји позитиван број  $c > 0$  тако да је  $\Psi(z) > c$  или  $\Psi(z) < -c$  за  $|z| \leq 1$ ,  $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$  за неко  $r \in \mathbb{N}$ .

**Претпоставка 3:** Нека је  $\hat{p} = \hat{p}(n) \rightarrow \infty$ ,  $\hat{p}(n) = o(n)$  када  $n \rightarrow \infty$  и нека оцене  $\hat{\phi}_{\hat{p}} = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_{\hat{p}})'$  задовољавају емпиријске Јул-Вокерове једначине

$$\hat{\Gamma}_{\hat{p}} \hat{\phi}_{\hat{p}} = \hat{\gamma}_{\hat{p}},$$

где је  $\hat{\Gamma}_{\hat{p}} = [\hat{\gamma}(i-j)]_{i,j=1}^{\hat{p}}$ ,  $\hat{\gamma}_{\hat{p}} = (\hat{\gamma}(1), \hat{\gamma}(2), \dots, \hat{\gamma}(\hat{p}))'$  и  $\hat{\gamma}(k) = \frac{1}{n} \sum_{t=k+1}^n (X_t - \bar{X}_n)(X_{t-k} - \bar{X}_n)$ ,  $0 \leq k \leq \hat{p}$ .

Следећа теорема, показује постојаност решеткастог бутстрепа за средњу вредност, чак и у случају када важи само претпоставка 1, тј. када не важи независнот резидуала. Доказ теореме налази се у раду [11].

**Теорема 4.4** Нека важи претпоставка 1 за  $s = 4$ , претпоставка 2 за  $r = 1$  и претпоставка 3 за  $p(n) = o((n/\log(n))^{1/4})$ , тада важи:

1.  $Var^*(n^{-1/2} \sum_{t=1}^n X_t^*) - Var(n^{-1/2} \sum_{t=1}^n X_t) = o_{\hat{p}}(1)$ ,  $n \rightarrow \infty$
2. Додатно ако важи  $n^{-1/2} \sum_{t=1}^n (X_t - \mu) \rightarrow^d N(0, \sum_{h=-\infty}^{\infty} \gamma(h))$ , тада за  $n \rightarrow \infty$  важи

$$\sup_x |P^* \{n^{-1/2} \sum_{t=1}^n (X_t^* - \bar{X}) \leq x\} - P \{n^{-1/2} \sum_{t=1}^n (X_t - \mu) \leq x\}| = o_{\hat{p}}(1).$$

Како бисмо упоредили перформансе метода блок бутстрепа и решеткастог бутстрепа приказаћемо вредности оцена средње вредности, параметара модела и аутокорелационих коефицијената добијене решеткастим

бутстрепом. Бутстреп серије добијене су функцијом `tsboot` програмског језика **R**.

```
tsboot(tseries, statistic, R, l = NULL, sim = "model",
      endcorr = TRUE, n.sim = NROW(tseries), orig.t = TRUE,
      ran.gen, ran.args = NULL, norm = TRUE, ...,
      parallel = c("no", "multicore", "snow"),
      ncpus = getOption("boot.ncpus", 1L), cl = NULL)
```

Функција `tsboot` је функција пакета `boot` која генерише бутстреп временске серије на основу модела који је одређен аргументом `sim` ('fixed' означава покретни блок бутстреп, 'geom' означава стационарни блок бутстреп, 'model' означава решеткасти бутстреп) и враћа вредности на основу прослеђене функције аргументу `statistic`, `R` одређује број бутстреп серија, `n.sim` одређује дужину бутстреп серија, `l` дужину блока, аргумент `ran.gen` користи се код решеткастог бутстрепа за генерисање нових серија, `ran.args` се користи за прослеђивање додатних аргумената функцији `ran.gen`.

Резултати су приказани Табелом 7.

	бутстреп оцена	пристрасност	стандардна девијација
$\bar{x}$	0.061 (-0.193,0.315)	0.0045	0.1276
$\phi_1$	0.223 (0.146,0.302)	0.0240	0.0393
$\phi_2$	0.436 (0.356,0.507)	0.0165	0.0388
$\rho_1$	0.400 (0.270,0.517)	0.0593	0.0635
$\rho_2$	0.526 (0.435,0.613)	0.0395	0.0441
$\rho_3$	0.290 (0.166,0.413)	0.1099	0.0632
$\rho_4$	0.292 (0.166,0.395)	0.0882	0.0584
$\rho_5$	0.1903 (0.0633,0.3070)	0.0571	0.0648

Табела 7: Вредности оцена добијених решеткастим бутстрепом

## 5 Примери са скуповима реалних података

Као што је већ поменуто раније, бутстреп метода је данас јако заступљена и има широку примену. Већ смо видели како се користе бутстреп методе за оцене параметара модела, сада ћемо се бавити применом бутстреп метода у предвиђању. Главни циљ анализе временских серија управо је предвиђање, а бутстреп методе могу бити корисне у побољшању тачности и одређивању интервала поверења прогноза. Да бисмо уопште могли да говоримо о прогнози треба прво моделирати дату серију, тј. уклопити податке у одговарајући модел на основу ког вршимо прогнозирање. Моделирање временских серија спроводи се у неколико корака:

1. визуелизација серије:

представљамо графички дату серију и на основу графика можемо да закључимо да ли је серија стационарна, да ли има тренд и сезонску компоненту, као и шта се дешава са дисперзијом током времена, а од великог значаја је и декомпозиција серије и приказ њених одвојених компоненти,

2. стационаризација серије:

уколико дата серија није стационарна постоји неколико метода трансформације којима се постиже стационарност

- Бокс-Кокс трансформација користи се у циљу стабилизације дисперзије и дата је изразом

$$X_t^{(\lambda)} = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(X_t) & \lambda = 0, \end{cases}$$

коэффициент  $\lambda$  оцењује се минимизацијом суме квадрата резидуала трансформисане серије,

- диференцирањем серије уклањамо тренд и сезонску компоненту,

3. ACF/PACF графикон:

на основу ACF и PACF графикона одређујемо ред модела,

4. оценити модел:

одређујемо оцену параметара одабраног модела,

5. провера адекватности модела: резидуали одабраног модела морају бити независни и нормално расподељени.

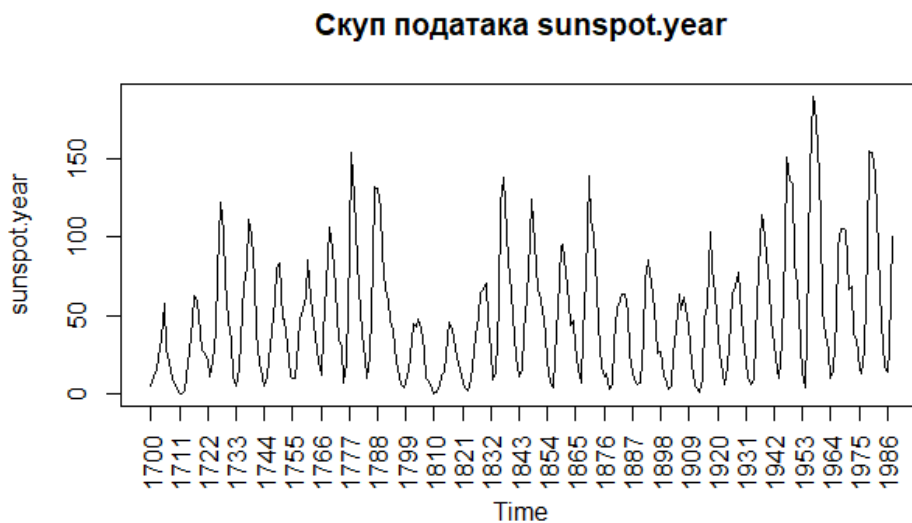
## 5.1 Пример 1: Оцена стандардне девијације грешака прогнозе

Размотримо примену решеткистог бутстрепа у оцини стандардне девијације грешака прогноза које се користе за добијање интервала поверења прогноза код стационарних временских серија, за које важи да је интервал поверења прогнозе дат са

$$\hat{X}_n(h) \pm Z_{1-\alpha/2}sd(e_n(h)),$$

где је  $Z_{1-\alpha/2}$  квантил стандардизоване нормалне расподеле,  $e_n(h) = \hat{X}_n(h) - X_{n+h}$  грешка прогнозе за  $h$  периода унапред.

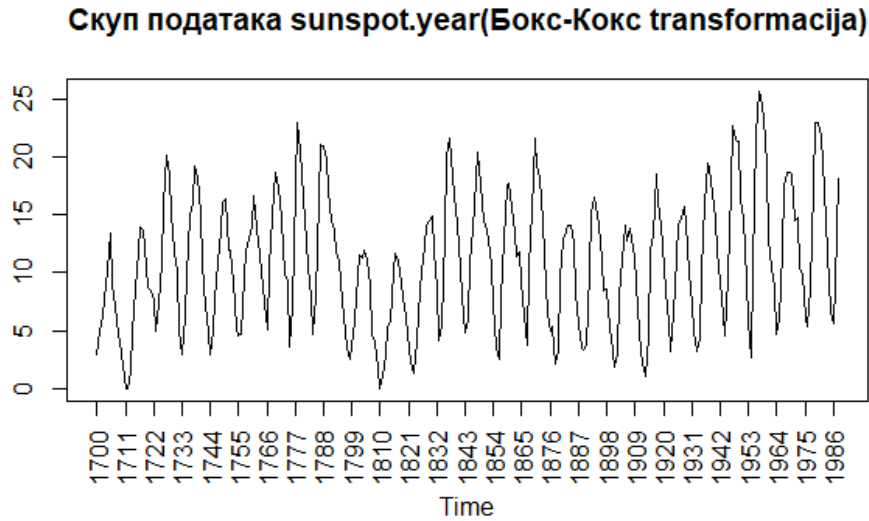
Посматраћемо временску серију *sunspot.year* која је имплементирана у пакету *datasets* програмској језика **R**. Серија показује годишњи број сунчевих пега од 1700. до 1988. године (в. Слика 22).



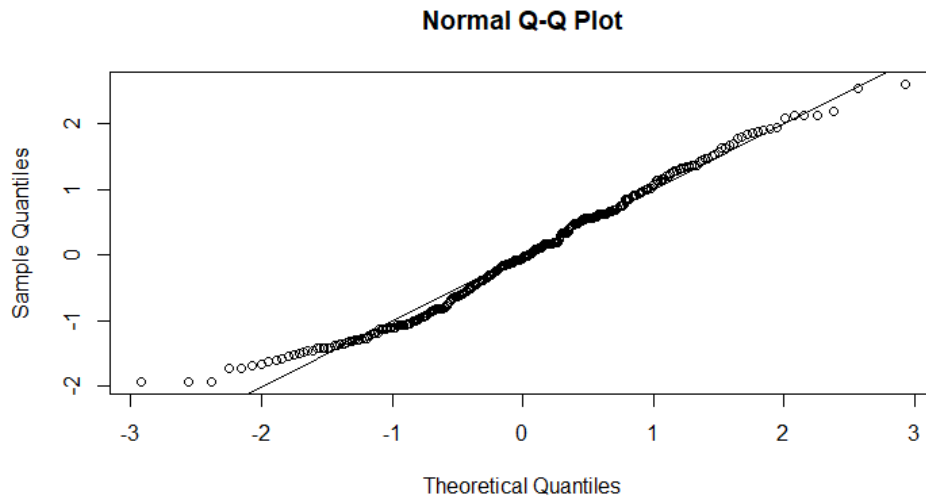
Слика 22: Графички приказ временске серије *sunspot.year*

На основу графика серије можемо закључити да серија има нестабилну дисперзију и циклус од 11 година. Како бисмо стабилизовали дисперзију и добили стационарну серију применићемо Бокс-Кокс трансформацију. Како је за Бокс-Кокс трансформацију потребно да вредности серије буду строго позитивне, а посматрана серија има вредности једнаке нули, увећаћемо све вредности серије за један, а затим применити Бокс-Кокс

трансформацију за параметар  $\lambda = 0.5$ . Погледајмо график тако добијене серије.



Слика 23: Графички приказ временске серије sunspot.year са Бокс-Кокс трансформацијом



Слика 24: QQ plot трансформисане серије sunspot.year

Стационарност добијене серије проверићемо помоћу Дики-Фулеровог

теста <sup>6</sup>, који је у програму **R** имплементиран функцијом `adf.test`. Тестира се нулта хипотеза  $H_0$ : Временска серија је нестационарна, против алтернативне  $H_1$ : Временска серија је стационарна. Ако је р-вредност теста мања од нивоа значајности 0.05 одбацујемо нулту хипотезу и закључујем да је временска серија стационарна.

```
> adf.test(sun_bx)
```

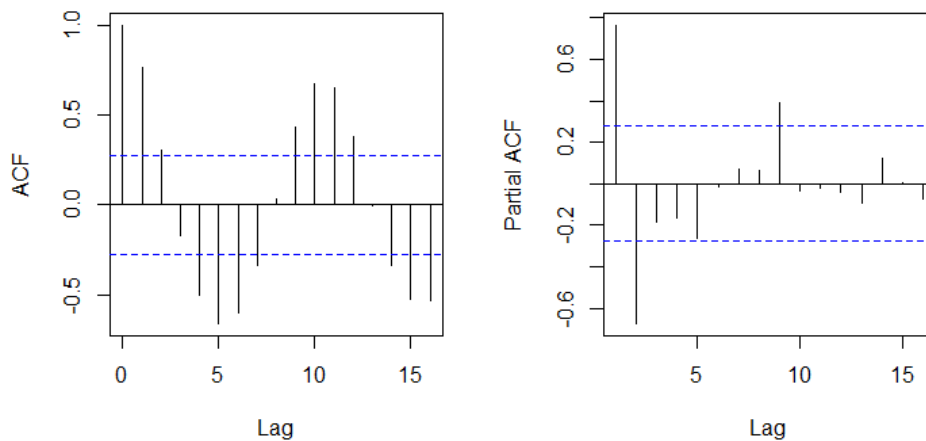
Augmented Dickey-Fuller Test

```
data: sun_bx
Dickey-Fuller = -4.6028, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Warning message:

```
In adf.test(sun_bx) : p-value smaller than printed p-value
```

Дисперзија трансформисане серије је знатно стабилнија, на основу QQ plot-а и adf теста можемо серију посматрати као стационарну. За даљи рад поделићемо серију на тренинг и тест скуп. За тренинг скуп узећемо серију од 1930. до 1979. године, а за тест скуп од 1980. до 1989. године. Погледајмо корелограме за податке тренинг скупа.



Слика 25: ACF и PACF тренинг скупа

---

<sup>6</sup>Dickey-Fuller test

Узорачки корелограм прати образац синусне функције због циклуса који се јавља у серији, а код парцијалног корелограма видимо значајне вредности за кашњења 1 и 2, али и касније, са кашњењем 9. Помоћу функције `auto.arima` одредићемо најбољи AR модел, који ћемо искористити за предвиђање. Функција `auto.arima` враћа ред модела и оцењене коефицијенте, добија се да је најбољи модел реда 11.

Поступак за одређивање стандардне девијације грешака прогнозе заснива се на генерисању нове серије дужине 59 применом решеткастог бутстрепа, затим се оцењује модел за тако добијену серију и врши се прогноза за првих 50 опсервација, даље се рачуна грешка прогнозе, тј. разлика добијених прогноза и одговарајућих вредности бутстреп серије. На овај начин добија се 2000 грешака прогнозе за сваки корак и на основу њих рачуна се стандардна девијација. За даљи рад и примену решеткастог бутстрепа одредићемо Јул-Вокерове оцене параметара AR(11) модела помоћу `ar` функције, а за решеткасти бутстреп користимо функцију `tsboot`.

У наставку је приказан код којим су дефинисане неопходне функције за оцену стандардне девијације грешака прогнозе.

```
sun_fun = function(ts) {
  # за проследјену бутстреп серију фитује се модел
  ar.fit = ar(window(ts, 1, 50), aic = FALSE, order.max = p)
  # функција враћа грешке предвиђања за 9 корака унапред
  c(predict(ar.fit, newdata = window(ts, 1, 50),
    n.ahead = 9, se.fit = FALSE)-window(ts,51,59))
}
# функција за генерисање бутстреп серија
sun_sim = function(tseries, n.sim, ran.args) {
  rg1 = function(n, res){
    # центрирамо резидуале
    sample(res - mean(res), n, replace = TRUE)
  }
  ts.orig = ran.args$ts
  ts.mod = ran.args$model
  # генерисемо нову серију помоћу функције arima.sim
  mean(ts.orig) + ts(arima.sim(model = ts.mod, n = n.sim,
    rand.gen = rg1, res = as.vector(ran.args$res)))
}
# ред и коефицијенти добијеног модела
```

```
sun_model = list(order = c(sun_ar$order, 0, 0), ar = sun_ar$ar)
# centriramo rezidualne pocetnog modela
sun_res = c(scale(sun_ar$resid[!is.na(sun_ar$resid)], scale = FALSE))
# sieve bootstrap
sun_boot = tsboot(ts(c(window(sun_bx, 1930))), sun_fun, R = 2000,
sim = "model", n.sim = 59, ran.gen = sun_sim, ran.args = list(res =
sun_res, ts = window(sun_bx, 1930), model = sun_model))
# racunamo standardnu devijaciju gresaka
sd_boot = apply(sun_boot$t, 2, sd)
```

Добијене вредности приказане су табелом:

Стандардна девијација грешака прогнозе									
Модел	2.094	3.176	3.601	3.605	3.608	3.658	3.725	3.765	3.813
Бутстреп	2.386	3.478	3.797	3.707	3.647	3.658	3.755	3.806	3.836

Табела 8: Вредности стандардне девијације грешака прогнозе

Интервали поверења прогноза су често веома уски и имају мању покривеност, јер се не узимају сви извори случајности приликом предвиђања. Постоје најмање четири извора случајности у предвиђању коришћењем модела временских серија:

- случајне грешке,
- оцене параметара,
- избор модела,
- грешке прогнозе.

Приликом конструисања интервала поверења узима се у обзир само први извор случајности. Бутстреп методе решавају овај проблем јер се генерише велики број бутстреп серија, где ће за сваку серију избор модела и оцена параметара бити другачија, чиме се узима у обзир случајност избора модела и оцене параметара. Применом бутстреп методе добијају се шири интервали поверења, а и већи проценат покривености. [18]

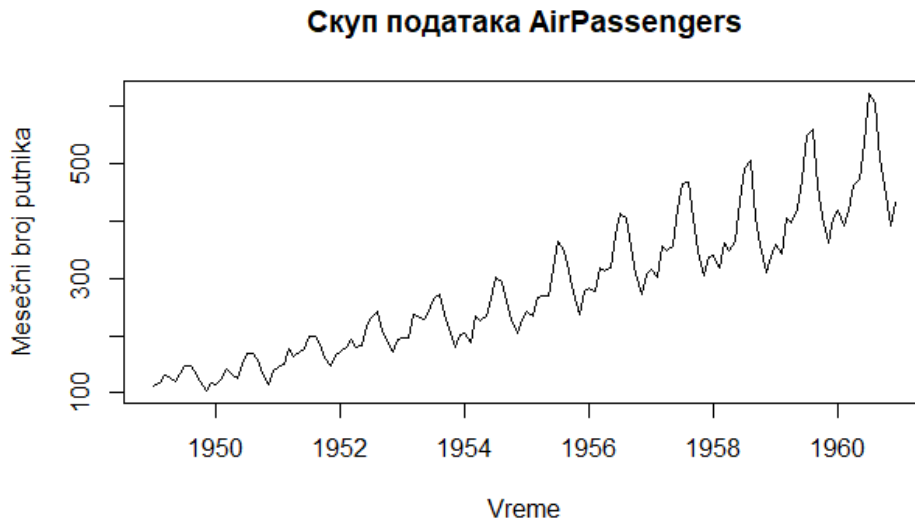
Приметимо да су вредности стандардне девијације грешака прогноза добијене помоћу решеткастог бутстрепа веће од стандардне девијације грешака добијених на основу одређеног AR модела, због чега би и интервали поверења били шири (в. Табела 8).



## 5.2 Пример 2: Побољшање тачности прогнозе

Представићемо примену бутстреп метода у побољшању тачности предвиђања нестационарних временских серија, код којих је честа појава корелисаност података случајне компоненте. Овај приступ данас је нарочито заступљен у машинском учењу, а основна разлика од приступа којим се врши оцена параметара модела јесте да се овде бутстреп методе примењују над случајном компонентом дате временске серије. Заправо, приступ се огледа у томе да се изврши декомпозиција дате временске серије, примени се бутстреп метода на случајну компоненту и добијеним реузорцима додаје се тренд и сезонска компонента, чиме се добија бутстреп серија којом се врши предвиђање. Уколико серија има нестабилну дисперзију примени се Бокс-Кокс трансформација пре декомпозиције, и након предвиђања на основу бутстреп серија примени се инверзна Бокс-Кокс трансформација [19]. Описани поступак применићемо на временску серију *AirPassengers* програмској језика **R**. Прво ћемо моделирати серију и одредити прогнозе на основу добијеног модела, а потом и помоћу бутстреп методе.

База *AirPassengers* представља временску серију о месечном броју путника интернационалне авиокомпаније у периоду од 1949. до 1960. године. На Слици 26 дат је графички приказ ове серије.

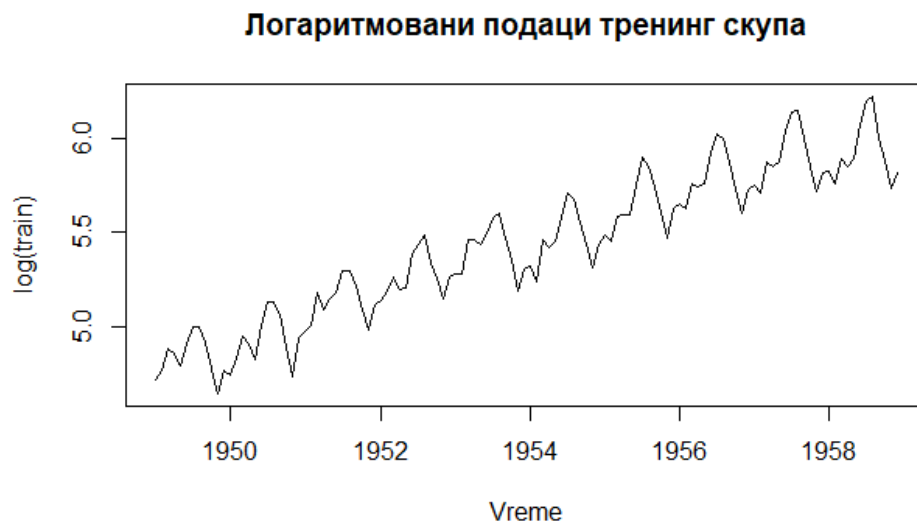


Слика 26: Графички приказ временске серије *AirPassengers*

Можемо видети да ова серија има јасно изражен тренд, сезонски образац и варијасна није стабилна.

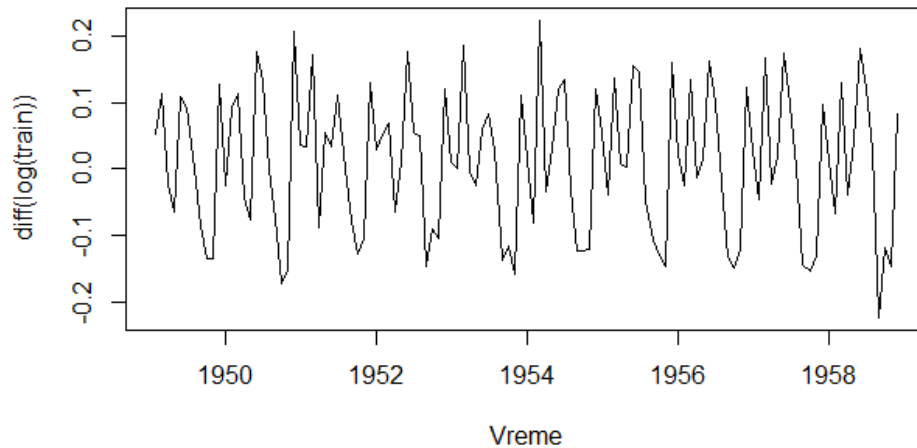
Дату серију поделићемо на скуп за тренирање и тестирање. На тренинг скупу извршићемо моделирање и предвиђање, а тест скуп служи како бисмо упоредили добијене резултате. Скуп за тренирање садржаће податке од 1949. до краја 1958. године, укупно 120 вредности, а тест скуп податке за последње две године. Дакле, прогнозу вршимо за две године, тј. 24 месеци.

Због нестабилне дисперзије извршићемо прво логаритмовање временске серије, а потом и диференцирање због тренда и сезонске компоненте. За диференцирање серије користимо функцију `diff` програма **R**. Графички приказ тако добијених података дат је сликама Слика 27 и Слика 28.



Слика 27: Графички приказ логаритмованог тренинг скупа

### Диференцирани и логаритмовани подаци тренинг скупа



Слика 28: Графички приказ диференцираног и логаритмованог тренинг скупа

Стационарност серије након диференцирања и логаритмовања можемо проверити функцијом `adf.test`.

```
> adf.test(train_diff)
```

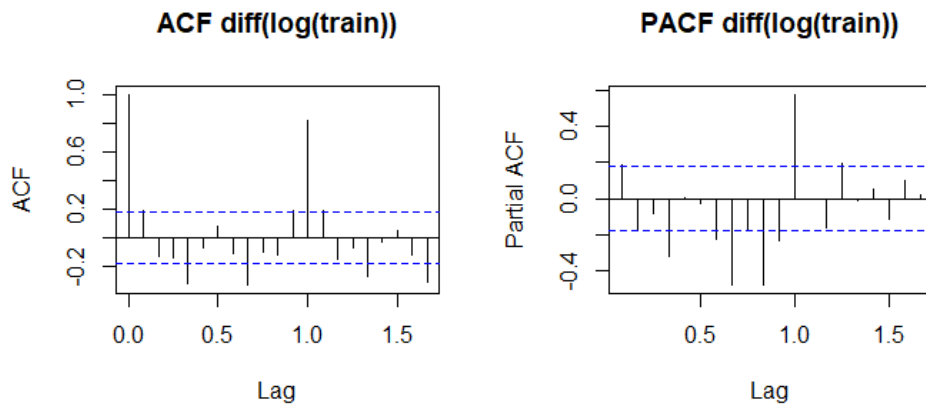
Augmented Dickey-Fuller Test

```
data: train_diff
Dickey-Fuller = -6.3698, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Warning message:

```
In adf.test(train_diff) : p-value smaller than printed p-value
```

На основу добијене  $p$ -вредности можемо закључити да је добијена серија стационарна. Погледајмо корелограме ове серије.



Слика 29: ACF и PACF диференцираног и логаритмованог тренинг скупа

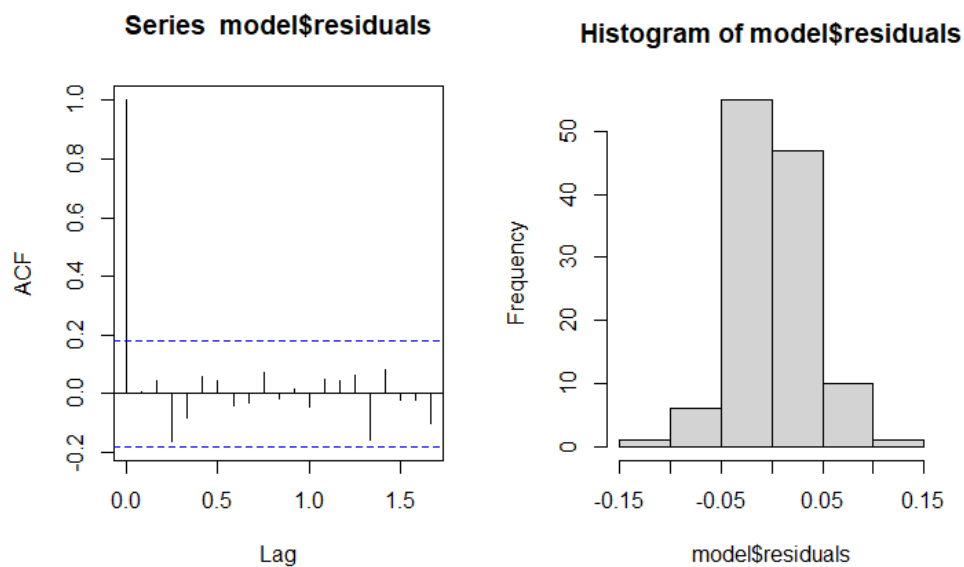
На основу корелограма серије можемо да закључимо да постоји одређена сезоналност, тако да је вероватно потребно извршити и сезонско диференцирање, такође, графици указују да треба прилагодити и МА модел. Како бисмо добили најбољи модел искористићемо функцију `auto.arima`.

```
> model = auto.arima(train)
> model
Series: train
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1      sma1
      -0.3424  -0.5405
s.e.      0.1009   0.0877

sigma^2 = 0.001432: loglikelihood = 197.51
AIC=-389.02  AICc=-388.78  BIC=-381
```

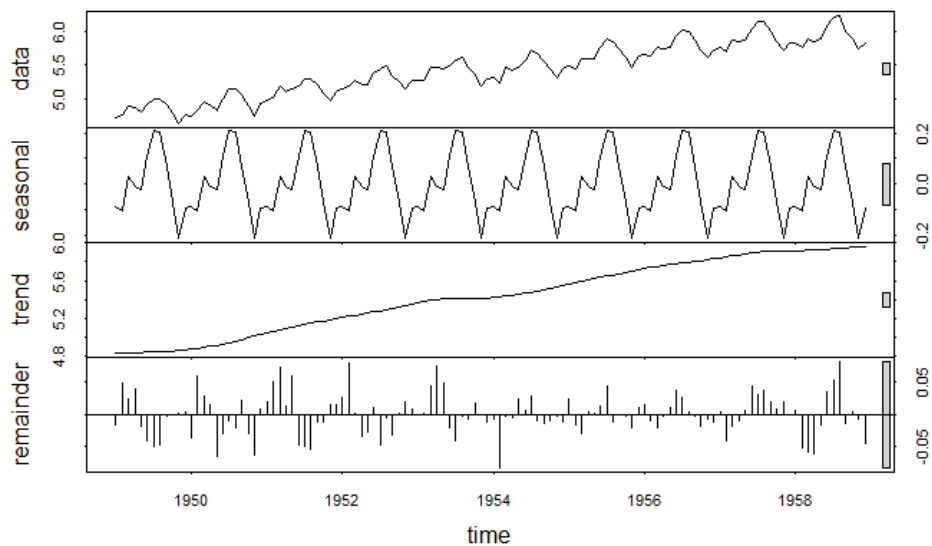
Функцијом `auto.arima` добили смо сезонски  $ARIMA(0,1,1)(0,1,1)$  модел. Оценићемо модел за наше податке функцијом `arima` и проверавамо адекватност модела на основу узорачког корелограма и хистограма за резидуале (в. Слика 30).



Слика 30: Провера адекватности модела на основу резидуала

Сада можемо да извршимо предвиђање на основу добијеног модела функцијом `predict` за 24 корака унапред.

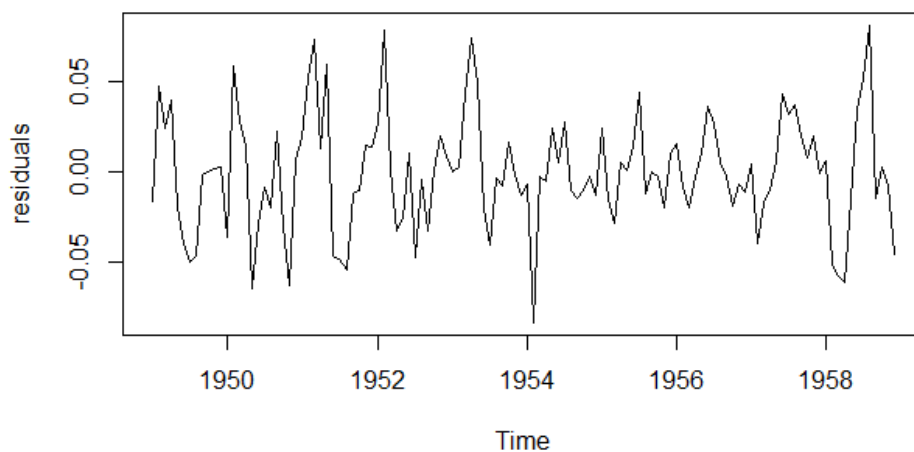
Имплементираћемо даље примену покретног блок бутстрепа за предвиђање. Први корак јесте декомпозиција логаритмованих података тренинг скупа коју постижемо функцијом `stl`.



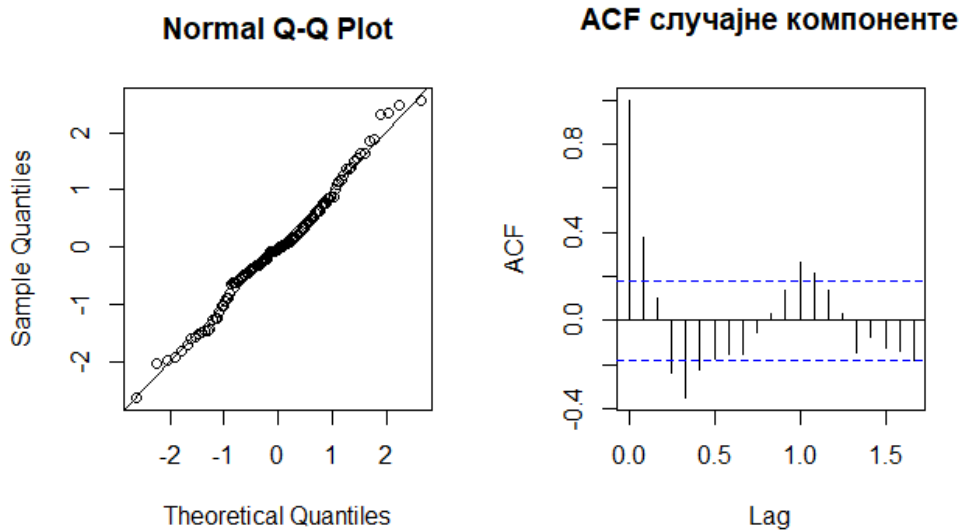
Слика 31: Декомпозиција тренинг скупа

Можемо уочити јасан тренд и сезоналност. Погледајмо случајну компоненту детаљније.

### Случајна компонента временске серије



Слика 32: Случајна компонента тренинг скупа



Слика 33: QQ plot и ACF случајне компоненте

```
> adf.test(rez)
```

Augmented Dickey-Fuller Test

```
data: rez
Dickey-Fuller = -5.975, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Warning message:

```
In adf.test(rez) : p-value smaller than printed p-value
```

Можемо закључити да је случајна компонента стационарна серија са израженом корелацијом. Управо постојање корелације код случајне компоненте захтева употребу неке блок бутстреп методе за реузорковање. Вршимо реузорковање случајне компоненте применом покретног блок бутстрепа функцијом `tsboot`.

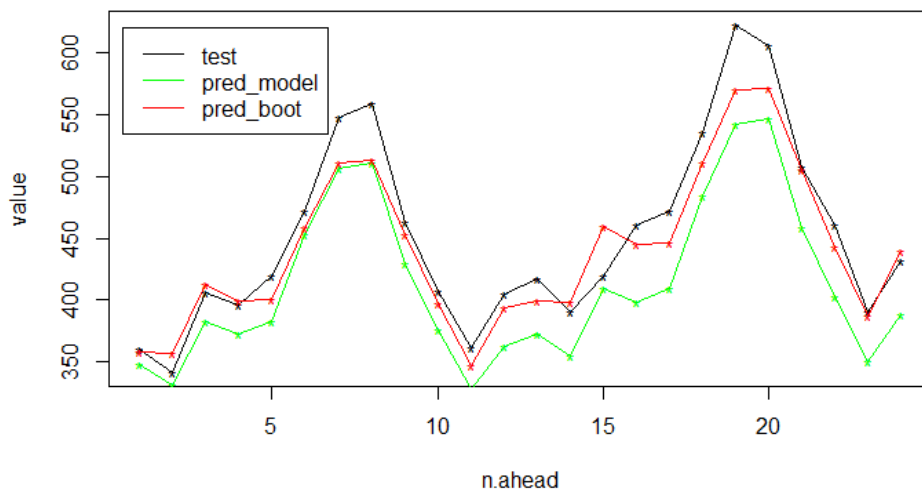
```
fun = function(ts){
  return(ts)
}
rez_boot = tsboot(rez, statistic = fun, R = 2000, l = 24,
sim = "fixed", n.sim = n, orig.t = F)
```

Функцији `tsboot` прослеђујемо случајну компоненту `rez`, генерисану функцију `fun` која враћа добијене реузорке, постављамо број реузорка на `R=2000`, дужину блока `l=24`, јер су подаци дати месечно, `sim =`

"fixed" означава да се ради о покретном блок бутстрепу, `n.sim = n` означава да је дужина блока једнака дужини тренинг скупа, а `orig.t = F` искључује податке о оригиналној серији случајне компоненте. Као резултат добијамо 2000 реузорака случајне компоненте. Сваком реузорку се додаје тренд и сезонска компонента, оцени се  $ARIMA(0,1,1)(0,1,1)$  модел за бутстреп серију и врши се предвиђање на основу добијеног модела за 24 корака, а коначна прогноза добија се инверзним логаритмовањем. На овај начин имамо по 2000 вредности за сваки корак, а бутстреп оцена прогноза за сваки корак добија се као њихова средња вредност.

```
for(i in 1:2000){
  ts_boot = rez_boot$t[i,]+t+s # butstrep us
  m = arima(ts_boot, c(0, 1, 1),
  seasonal = list(order = c(0, 1, 1), period = 12))
  p = predict(m, n.ahead = 24)
  pred_boot[i,] = 2.718~p$pred
}
p = rep(0,24)
p = apply(pred_boot, 2, mean)
```

Добијени резултати приказани су наредним графиком:



Слика 34: Прогноза серије за 1959. и 1960. годину

Приметимо да је примена бутстреп методе дала значајна побољшања



предвиђања, а на то указују и вредности RMSE (енгл. root mean square error) која износи 43.43 за предвиђање на основу модела и 23.06 за предвиђање помоћу бутстреп методе.

## 6 Закључак

У овом раду упознали смо се са бутстреп методама које су корисне у анализи стационарних временских серија, као и са њиховом применом. Описане су четири методе блок бутстрепа: блок бутстреп без преклапања, покретни, кружни и стационарни блок бутстреп. Свака од ових метода има сличан механизам који се заснива на подели оригиналне временске серије на блокове, а случајним избором блокова са понављањем генерише се бутстреп серија. Поред блок бутстрепа описан је и решеткасти бутстреп који се заснива на оцени AR модела и реузорковањем резидуала омогућава генерисање бутстреп серија. Бутстреп серије користе се за оцену посматраних параметара и конструисање интервала поверења. Такође, дате су и теореме које показују постојаност наведених метода за оцену средње вредности временских серија. У литератури [8] (Поглавље 4) могу се наћи и теореме и докази које показују постојаност бутстреп метода за оцене параметара које се могу представити као глатка функција средње вредности, као што су оцене аутоковаријационих и аутокорелационих коефицијената и Јул-Вокер оцене параметара AR модела, али и такозваних генерализованих M-оцена као што су оцене параметара ARMA модела методом максималне веродостојности. Симулација метода у четвртом поглављу даје нумеричке резултате који описују перформансе ових модела за оцену средње вредности, параметара серије и аутокорелационих коефицијената. Овде смо показали да дужина блока, као и обим оригиналне серије утичу на квалитет оцена. Стандардна девијација оцена се повећава са повећањем дужине блока, а стационарни блок бутстреп даје највећу стандардну девијацију због додатне случајности која се уноси насумичним избором дужине блока. Такође, закључујемо да описане бутстреп методе одржавају основну зависност међу подацима временских серија, али и да зависност података значајно зависи од дужине блока. Зависност података је знатно смањена у бутстреп временским серијама за неке дужине блока, а у случају стандардног блок бутстрепа, за који је дужина блока један, губи се зависност података. Приметимо још да се у случају AR(2) модела који је коришћен у симулацијама добијају добри резултати за оцену параметра  $\phi_1$ , али знатно лошији резултати, у смислу потцењивања оцена, за параметар модела  $\phi_2$ . Слични резултати могу се наћи у раду [15], а у овом раду могу се наћи и резултати симулација за друге моделе. Свакако постоји потреба за додатним истраживањем о перформансама ових метода и квалитету оцена параметара за различите моделе временских серија. Теоријска поређења описаних метода могу се наћи у радовима

[16] и [17].

У поглављу пет описане су неке од примена бутстреп метода у побољшању предвиђања и интервала поверења прогноза. Приметимо да је у Примеру 2 овог поглавља описана примена блок бутстрепа за предвиђање у случају нестационарних серија, реузорковањем серија резидуала. Свакако, овде постоји простор за даље истраживање о применама бутстреп метода за анализу нестационарних серија.

На крају, можемо закључити да избор бутстреп методе није лак задатак и да га треба обавити са опрезом. Величина узорка и структура зависности имају значајну улогу у одабиру бутстреп методе.

*"Ако мислите да статистика нема ништа да каже о ономе што радите или како да то радите боље, онда или грешите или је време да нађете интересантнији посао": Стивен Сен<sup>7</sup>.*

---

<sup>7</sup>Stephen Senn - енглески статистичар

## 7 Додатак 1

```
# ne preklapajuci blok butstrep
NBB = function(ts, l){

  n = length(ts) # duzina vs
  b = floor(n/l) # br. blokova

  Y = matrix(0, nrow = b, ncol=1) # matrica blokova
  for(i in 1:b)
    Y[i,] = ts[((i-1)*l+1):(i*l)] # jedna vrsta je jedan blok

  # biramo b blokova slucajnim izborom sa ponavljanjem
  index = sample(c(1:b), size = b, replace = T)

  X = c() # butstrep vs
  for(i in 1:b)
    X = c(X,Y[index[i],])

  return(X)
}

# pokretni blok butstrep
MBB = function(ts, l){

  n = length(ts) # duzina vs
  N = n-l+1 # br. blokova

  Y = matrix(0, nrow = N, ncol=1) # matrica blokova
  for(i in 1:N)
    Y[i,] = ts[i:(i+l-1)]

  b = floor(n/l) # biramo b blokova
  index = sample(c(1:N), size = b, replace = T)

  X = c() # butstrep vs
  for(i in 1:b)
    X = c(X,Y[index[i],])

  return(X)
}
```

```
# kruzni blok butstrep
CBB = function(ts, l){

  n = length(ts)
  ts1 = array(0, dim = n+1-1) # proizena us

  for(i in 1:n)
    ts1[i]=ts[i]
  for(i in (n+1):(n+1-1)){
    if(i %% n == 0)
      j = n
    else
      j = i %% n

    ts1[i]=ts[j]
  }

  # broj blokova je jednak n
  Y = matrix(0, nrow = n, ncol=1) # matrica blokova
  for(i in 1:n)
    Y[i,] = ts1[i:(i+1-1)]

  b = floor(n/l) # biramo b blokova
  index = sample(c(1:n), size = b, replace = T)

  X = c() # butstrep us
  for(i in 1:b)
    X = c(X,Y[index[i],])

  return(X)
}

# stacionarni blok butstrep
SBB = function(ts, l){

  n = length(ts)

  p = 1/l # p vrednost geometrijske raspodele biramo na osnovu toga da
# se ocekivanje raspodele poklapa sa datom duzinom bloka

  X = c() # nova us
```

```
n1=0 # duzina nove us

while(n1<=n){

  l1=rgeom(1,p) # generisemo duzine blokova iz geom. raspodele sa parametrom p

  if((l1!=0) && (l1<n)){
    # generisemo indekse pocetnih opservacija blokova iz uniformne raspodele
    i1 = round(runif(1,1,n))

    # generisemo blokove i smestamo ih u novu us
    for(i in i1:(i1+l1-1)){

      if(i %% n == 0)
        j = n
      else
        j = i %% n

      X=c(X,ts[j])

    }

    n1=n1+l1

  }
}

X = X[1:n]
return(X)
}
```

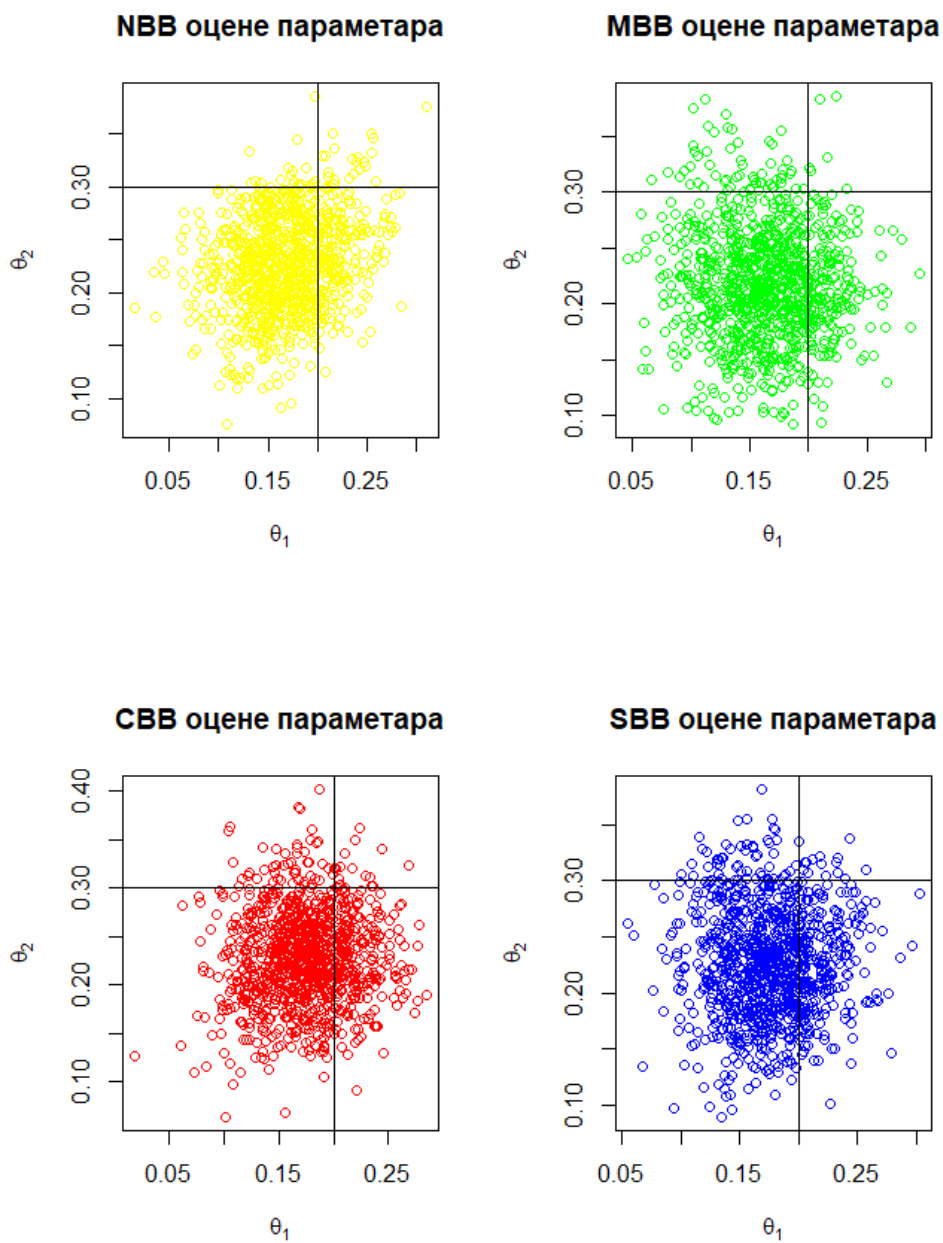
## 8 Додатак 2

модел	$l = 4$	$l = 8$	$l = 16$	$l = 32$
NBB	0.920	0.914	0.934	0.938
MBB	0.920	0.922	0.934	0.952
CBV	0.920	0.920	0.938	0.952
SBB	0.918	0.924	0.930	0.946

Табела 9: Просечна покривеност интервала поверења за средњу вредност серије МА(2) модела за  $n = 512$  и  $l = 4, 8, 16, 32$

		$l = 4$	$l = 8$	$l = 16$	$l = 32$
NBB	$\bar{x}^*$	0.0170	0.0140	0.0180	0.0130
	$bias(\bar{x}^*)$	0.0025	0.0001	0.0036	-0.0006
	$sd(\bar{x}^*)$	0.0593	0.0596	0.0666	0.0602
MBB	$\bar{x}^*$	0.0160	0.0110	0.0090	0.0010
	$bias(\bar{x}^*)$	0.0024	-0.0027	-0.0053	-0.0126
	$sd(\bar{x}^*)$	0.0609	0.0636	0.0657	0.0662
CBV	$\bar{x}^*$	0.0160	0.0100	0.0140	0.0130
	$bias(\bar{x}^*)$	0.0018	-0.0038	-0.0005	-0.0010
	$sd(\bar{x}^*)$	0.0594	0.0656	0.0635	0.0622
SBB	$\bar{x}^*$	0.0160	0.0120	0.012	0.0170
	$bias(\bar{x}^*)$	0.0025	-0.0016	-0.002	0.0032
	$sd(\bar{x}^*)$	0.0604	0.0645	0.061	0.0621

Табела 10: Бутстреп оцене средње вредности серије МА(2) модела

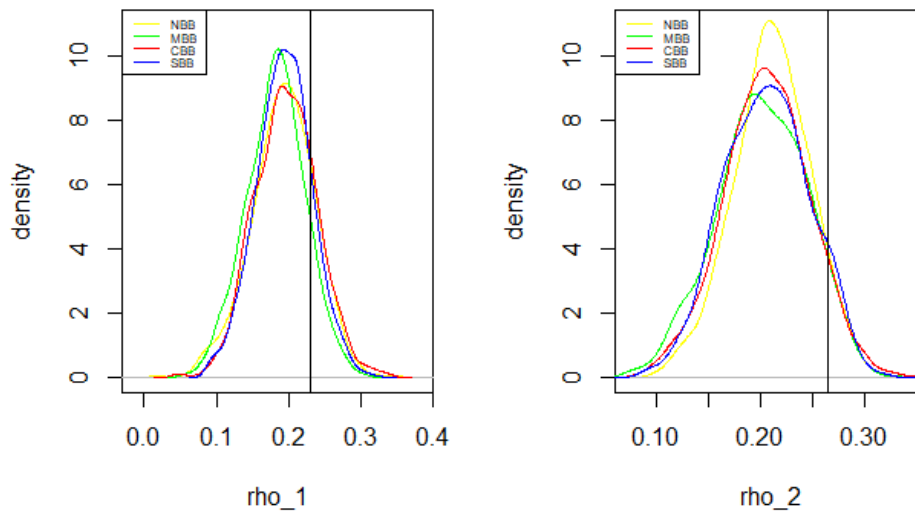


Слика 35: Оцене параметара бутстреп временских серија  $MA(2)$  модела за  $n = 512$  и  $l = 32$

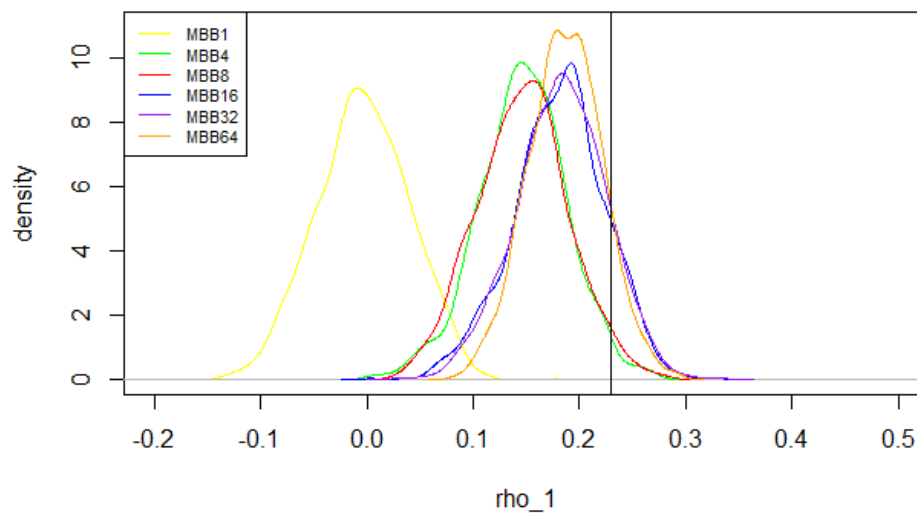


Модел	NBB	MBB	CBB	SBB
$\hat{\theta}_1^*$	0.1710 (0.0887,0.2566)	0.1640 (0.0852,0.2457)	0.1740 (0.0976,0.2593)	0.1740 (0.1066,0.2447)
$\hat{\theta}_2^*$	0.2280 (0.1408,0.3129)	0.2200 (0.1250,0.3191)	0.2270 (0.1343,0.3310)	0.2260 (0.1267,0.3172)
$bias(\hat{\theta}_1^*)$	0.0138	0.0074	0.0172	0.0172
$bias(\hat{\theta}_2^*)$	0.0189	0.0105	0.0176	0.0165
$sd(\hat{\theta}_1^*)$	0.0431	0.0395	0.0398	0.0354
$sd(\hat{\theta}_2^*)$	0.0442	0.0496	0.0526	0.0487

Табела 11: Бутстреп оцене параметара МА(2) модела за  $n = 512$  и  $l = 32$



Слика 36: Оцена коефицијената корелације са кашњењем 1 и 2 за МА(2) модел



Слика 37: Оцена коефицијента корелације са кашњењем 1 за различите дужине блока за MA(2) модел

## Литература

- [1] Bradley Efron, Robert Tibshirani. *An Introduction to the Bootstrap*, 1993.
- [2] Bradley Efron. *Bootstrap methods: Another look at the Jackknife*. The Annals of Statistics, 1979, Vol.7, No.1, 1-26
- [3] [http://elibrary.matf.bg.ac.rs/bitstream/handle/123456789/4242/Boskovic\\_Milica.pdf?sequence=1](http://elibrary.matf.bg.ac.rs/bitstream/handle/123456789/4242/Boskovic_Milica.pdf?sequence=1)
- [4] Златко Ј. Ковачић. *Анализа временских серија*. Економски факултет, Београд 1995.
- [5] Abdelkader Mokkadem. *Mixing properties of ARMA processes*. Stochastic Processes and Their Applications, 29, 309–315, 1988.
- [6] Edward Carlstein. *The use of subseries methods for estimating the variance of a general statistic from a stationary time series*. The Annals of Statistics, 14:1171–9, 1986.
- [7] Regina Y. Liu and Kesar Singh. *Moving blocks jackknife and bootstrap capture*. In Raoul Lepage and Lynne Bilard, editors, Exploring the limits of bootstrap, pages 225–248. Wiley-Interscience, 1992.
- [8] Soumendra N. Lahiri. *Resampling Methods for Dependent Data*. Springer, 2003.
- [9] Dimitris N. Politis and Joseph P. Romano. *A circular block resampling procedure for stationary data*. In Raoul Lepage and Lynne Bilard, editors, Exploring the limits of bootstrap, page 263270. Wiley-Interscience, 1992.
- [10] Dimitris N. Politis and Joseph P. Romano. *The stationary bootstrap*. Journal of the American Statistical Association, 89(428): 1303–1313, 1994.
- [11] Peter Bühlmann. *Sieve bootstrap for time series*. Bernoulli, 3(2):123–148, 1997.
- [12] <http://www.stat.uchicago.edu/~pmcc/courses/stat306/2013/cumulants.pdf>
- [13] Jonathan D. Cryer, Kung-Sik Chan. *Time Series Analysis With Applications in R*

- [14] Michael R. Chernick, Robert A. LaBudde. *An Introduction to Bootstrap Methods with Applications to R*
- [15] Fanny Bergström. *Bootstrap Methods in Time Series Analysis*. Stockholm University, 2018.
- [16] Boris Radovanović and Aleksandra Marcikić. *A comparison of four different block bootstrap methods*. Croatian Operational Research Review CRORR 5(2014), 189–202
- [17] Soumendra N. Lahiri. *Theoretical comparasions of block bootstrap methods*. The Annals of Statistics 1999, Vol. 27, No. 1, 386–404
- [18] <https://otexts.com/fpp2/bootstrap.html?fbclid=IwAR2MxNiobHxjV1lrw3dxtsXevk4J2hCeureXiPU0JJ42hMaAXgyy6ytBEws>
- [19] Christoph Bergmeir, Rob J Hyndman, José M Benitez. *Bagging Exponential Smoothing Methods using STL Decomposition and Box-Cox Transformation*.

## Биографија аутора

Ана Јовичић рођена је 01.07.1996. у Параћину. Завршила је Гимназију у Параћину 2015. године и носилац је Вукове дипломе. Исте године уписала је Математички факултет Универзитета у Београду, смер Статистика, актуарска и финансијска математика. Основне студије завршила је 2019. године са просечном оценом 8.5 и исте године уписује мастер студије на истом смеру. У септембру 2020. године положила је све испите и тиме стекла услов за одбрану мастер рада. Прво радно искуство стиче 2020. године као професор математике у Војној гимназији, а након тога, 2021. године, добија позицију професор-руководилац актива Средње стручне војне школе "1300 каплара", где и данас ради. Интересовање за математику гаји од основне школе, а током студија развила је интересовање за област статистике, обраду података и машинско учење.