



Univerzitet u Novom Sadu
Prirodno-matematički fakultet
Departman za matematiku i informatiku



Milena Vajagić

**Ocena parametara u autoregresivnim modelima -
stohastički gradijent protiv gradijenta sa
rastućim uzorkom**

-Master rad-

Mentor:

dr Nataša Krklec Jerinkić

Novi Sad, 2020

Sadržaj

| | | |
|----------|--|-----------|
| 1 | Analiza vremenskih serija | 5 |
| 1.1 | Osnovni pojmovi teorije verovatnoće | 5 |
| 1.2 | Pojam stohastičkog procesa, definicija i tipovi vremenskih serija | 9 |
| 1.3 | Pojam autokorelacije i beli šum | 13 |
| 1.4 | Autoregresivni modeli (AR) | 14 |
| 1.4.1 | Određivanje reda p | 17 |
| 1.4.2 | Provera adekvatnosti modela | 18 |
| 1.4.3 | Test normalnosti | 18 |
| 1.4.4 | Test jediničnih korena | 18 |
| 1.4.5 | Formiranje prediktivnih vrednosti AR modela | 19 |
| 1.4.6 | Greške predviđanja | 20 |
| 2 | Ocenjivanje parametara | 21 |
| 2.1 | Problem najmanjih kvadrata | 21 |
| 2.2 | Optimizacija | 23 |
| 2.3 | Gradijentni metod | 26 |
| 2.3.1 | Stohastički gradijentni metod | 28 |
| 2.3.2 | Gradijentni metod sa rastućim uzorkom | 30 |
| 3 | Istraživanje | 32 |
| 3.1 | Grafički prikaz i osnovna analiza | 32 |
| 3.2 | Osnovne statističke osobine | 35 |
| 3.3 | Ocena parametara modela metodom stohastičkog gradijenta (SGD) i gradijenta sa rastućim uzorkom (IGD) | 38 |
| 3.4 | Formiranje prediktivnih vrednosti $AR(3)$ modela | 42 |
| 3.5 | Formiranje prediktivnih vrednosti $AR(5)$ modela | 45 |
| 3.6 | Formiranje prediktivnih vrednosti $AR(8)$ modela | 48 |
| 3.7 | Komparativna analiza rezultata predikcija | 50 |
| 3.8 | Grafički prikaz relativne greške i analiza | 61 |
| 4 | Zaključak | 64 |
| | Literatura | 68 |
| | Biografija | 69 |

Predgovor

Pod vremenskom serijom podrazumeva se uređeni niz opservacija. Uređivanje se najčešće vrši u odnosu na vreme u jednakim vremenskim intervalima. Serije koje će biti razmatrane u ovom radu će imati ovakvu karakterizaciju. Na samom početku, ukažimo na jednu od suštinskih razlika između klasične statističke analize i statističke analize vremenskih serija. Dok su kod klasične statističke analize elementi slučajnog uzorka međusobno nezavisni, kod analize vremenskih serija opservacije u uzorku nisu među sobom nezavisne samim tim što se pri analizi uzima u obzir njihov vremenski poredak. Upravo ova međusobna zavisnost opservacija se koristi u cilju formiranja modela vremenske serije, a zatim kad je poznat model koji serija prati koristimo ga da na osnovu prošlih, prognoziramo buduće opservacije.

Vremenske serije se sreću u različitim sferama ljudskog delovanja. Recimo, u demografiji je to stopa prirodnog priraštaja, u ekonomiji fluktuacije deviznog kursa, mesečno kretanje industrijske proizvodnje, uvoza, izvoza i cena. U medicini se za pacijenta između ostalog beleži njegov elektrokardiogram (EKG) u svakom vremenskom trenutku pregleda. U meteorologiji se svakog sata registruje brzina vetra, prati se temperatura na dnevnom nivou kao i prosečne mesečne ili godišnje padavine na nekoj teritoriji. U radu će pored glavnog cilja - ocene parametara za dobijanje predikcija, predstavljanja i opisa metoda za ocenu parametara, prediktivnih modela, kao i praktičnog dela takode biti reči i o analizi vremenskih serija, slučajnim i linearnim procesima, koji su neophodni radi uvođenja u problematiku formiranja predikcija. Analiza vremenskih serija po pravilu obuhvata i rešavanje problema optimizacije. Različiti metodi optimizacije daju različite rezultate po pitanju kvaliteta aproksimacije rešenja, ali i po pitanju vremena potrebnog za dobijanje aproksimacija ili nekog drugog vida troškova.

Nakon teorijskog dela i predstavljanja metoda za ocenu parametara u vremenskim serijama uslediće i praktičan deo rada, prognoziranje budućih vrednosti vremenske serije sa prethodno ocenjenim parametrima. U radu će se koristiti simulirane serije koje prate AR model, kod kojih će se posmatrati ponašanje usled promene reda modela i disperzije šuma. Posmatraćemo slučaj u kom se vrši rekalkibracija, tj. parametri modela se iznova ocenjuju sa dotokom novih podataka kako bi se dobile što bolje predikcije. Ovaj rad ima za cilj da predstavi i uporedi dva modela za ocenu parametara: stohastički i inkrementalni, kao dve modifikacije gradijentnog metoda. Analiziraće se koji od metoda daje bolje rezultate u vidu preciznijih predikcija. Rezultati testova, grafici i modeli u radu dobijeni su korišćenjem računarskog softvera *Matlab*.

Veliku zahvalnost dugujem svojoj mentorki, dr Nataši Krklec Jerinkić, za savete, posvećenost, razumevanje, kao i za preneto znanje tokom studija. Zahvaljujem se na izdvojenom vremenu i korisnim savetima bez kojih rad ne bi imao sadašnji oblik.

Takode, zahvalila bih se članovima komisije, prof. dr Nataši Krejić i prof. dr Sanji Rapajić, na prenetom znanju tokom studiranja i izdvojenom vremenu prilikom izrade master rada.

*Posebnu zahvalnost dugujem svojoj porodici, sestri Milici, majci Bosiljki i ocu
Mladenu na bezgraničnom strpljenju i podršci koju su mi pružali tokom
celokupnog školovanja.*

1 Analiza vremenskih serija

Analiza vremenskih serija predstavlja jednu od statističkih disciplina koja beleži najdinamičniji razvoj poslednjih decenija. Našla je primenu u mnogim oblastima ekonomije, finansija, prirodnih i društvenih nauka kao i u raznim svakodnevnim situacijama kojih zapravo nismo ni svesni (npr. u praćenju prometa pazara neke radnje, vremenskoj prognozi, itd) tj. u svakoj oblasti koja uključuje merenje neke pojave kroz vreme. Glavni ciljevi u analizi vremenskih serija su da se pomoću određenih statističkih testova izvedu zaključci o osnovnim karakteristikama vremenske serije, objasni njeno ponašanje i da se na osnovu istorijskih podataka prognozira buduća vrednost neke pojave. Korišćenjem različitih modela analize vremenskih serija opisuje se izučavana pojava, daje objašnjenje zašto i kako je do nje došlo, zatim predvideti njeno kretanje u narednom periodu i naposljetku, manipulacijom veličina koje uslovljavaju posmatranu pojavu, istu držati pod kontrolom. Na osnovu prethodnih vrednosti serije ocenjujemo model, koji zatim koristimo za formiranje prognoze budućih vrednosti serije. Na kraju, upoređujući stvarne i prediktivne vrednosti formiramo grešku predviđanja. One će ujedno predstavljati i osnovni kriterijum kada se stigne do zaključivanja koji je model bolje ispratio stvarne vrednosti date serije.

U ovom poglavlju biće predstavljeni statistički testovi, kao i autoregresivni modeli koji su korišćeni pri analizi vremenskih serija i formiranju simuliranih serija i njihovih predikcija. Pre predstavljanja modela definišu se osnovni pojmovi iz teorije verovatnoće i stohastičke analize koji su neophodni za dalje razumevanje.

1.1 Osnovni pojmovi teorije verovatnoće

Osnovni model u teoriji verovatnoće jeste eksperiment (pojava, opit) kod koga ostvarivanje određenih uslova ne dovodi do jednoznačnog rezultata. Skup svih (logički) mogućih ishoda nekog eksperimenta označava se sa Ω , a elementi tog skupa nazivaju se elementarni događaji i označavaju se sa ω .

Definicija 1 *Slučajan događaj A (ili samo događaj A) je podskup skupa elementarnih događaja Ω . On se sastoji od onih elementarnih događaja ω koji imaju svojstvo kojim se događaj A definiše.*

Aksioma 1 (Aksioma σ - polja): *Podskup \mathcal{F} partitivnog skupa $\mathcal{P}(\Omega)$ je σ - algebra nad Ω ako važe uslovi:*

1. $\Omega \in \mathcal{F}$
2. ako $A \in \mathcal{F}$, onda i $\bar{A} \in \mathcal{F}$, gde je \bar{A} komplement skupa A
3. ako niz $A_1, A_2, \dots \in \mathcal{F}$ onda i $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

gde Ω predstavlja skup svih mogućih ishoda nekog eksperimenta.

Definicija 2 *Borelovo σ -polje $\mathcal{B} = \mathcal{B}(\mathbb{R})$ je σ -polje definisano nad skupom realnih brojeva. Formira se pomoću familije poluotvorenih intervala $[a, b)$, $a, b \in \mathbb{R}$ i sadrži sve skupove koji se dobijaju kao konačne ili prebrojive unije ili preseki te familije, kao i skupove koji se dobijaju uzimanjem komplementa.*

Može se pokazati da $\mathcal{B}(\mathbb{R})$ sadrži sve otvorene, zatvorene, poluotvorene intervale i unije tih skupova.

Aksioma 2 (Aksioma verovatnoće) Neka je Ω skup elementarnih događaja i \mathcal{F} σ -polje nad Ω . Funkcija $P: \mathcal{F} \rightarrow [0, 1]$ se zove verovatnoća na prostoru (Ω, \mathcal{F}) ako zadovoljava sledeće uslove:

1. $P(\Omega) = 1$,
2. Ako $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{F}$, $A_i \cap A_j = \emptyset$, $i \neq j$, $i, j = 1, 2, \dots$, onda

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Prostor verovatnoća je uređena trojka (Ω, \mathcal{F}, P) , gde je Ω skup svih elementarnih događaja, \mathcal{F} je σ -polje nad Ω , a P je verovatnoća na (Ω, \mathcal{F}) .

Definicija 3 Preslikavanje $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$, gde je $\mathbf{X} = (X_1, \dots, X_n)$ je n -dimenzionalna slučajna promenljiva na prostoru verovatnoća (Ω, \mathcal{F}, P) ako $\mathbf{X}^{-1}(S) \in \mathcal{F}$ za svako $S \in \mathcal{B}_n$, gde je $\mathcal{B}_n = \mathcal{B}(\mathbb{R}^n)$ Borelovo σ -polje. Ekvivalentno, kažemo da je \mathbf{X} \mathcal{F} -merljivo.

Definicija 4 Funkcija $F_{\mathbf{X}}(x) = F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n)$, za $(x_1, \dots, x_n) \in \mathbb{R}^n$, zove se zajednička funkcija raspodele slučajnih promenljivih (X_1, \dots, X_n) .

Definicija 5 Slučajna promenljiva $\mathbf{X} = (X_1, \dots, X_n)$ je diskretna (diskretnog tipa) ako postoji prebrojiv skup tačaka u \mathbb{R}^n :

$$R_{\mathbf{X}} = \{(x_1^{(k_1)}, \dots, x_n^{(k_n)}); k_1, \dots, k_n = 1, 2, \dots\}$$

takav da je $P\{X \in \overline{R_{\mathbf{X}}}\} = 0$. Skup $R_{\mathbf{X}}$ je skup svih mogućih vrednosti za \mathbf{X} .

Definicija 6 Slučajna promenljiva $\mathbf{X} = (X_1, \dots, X_n)$ je n -dimenzionalna slučajna promenljiva apsolutno neprekidnog tipa ako postoji integrabilna funkcija $f_{\mathbf{X}}(x_1, \dots, x_n) \geq 0$, $-\infty < x_1, \dots, x_n < \infty$, takva da je, za svaki Borelov skup $S \in \mathcal{B}_n$,

$$P\{(X_1, \dots, X_n) \in S\} = \int \cdots \int_S f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Funkcija $f_{\mathbf{X}}(x)$ zove se gustina raspodele verovatnoća (ili samo gustina raspodele) slučajne promenljive \mathbf{X} .

Specijalno, ako izaberemo $S = \{(u_1, \dots, u_n); u_k < x_k, k = 1, \dots, n\}$ dobijamo

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(u_1, \dots, u_n) du_1 \dots du_n$$

za svako $(x_1, \dots, x_n) \in \mathbb{R}^n$

Definisani pojmovi su u nastavku dati za jednodimenzionalne slučajne promenljive ($n = 1$).

Definicija 7 *Kvantil reda p je ona vrednost x_p za koju važi da je $F_X(x_p) = P(X < x_p) = p$, $p \in (0, 1)$, gde je $F_X(x) = P(X < x)$ predstavlja funkciju raspodele za jednodimenzionalnu slučajnu promenljivu.*

Definicija 8 *Očekivanje ili srednja vrednost apsolutno neprekidne slučajne promenljive X se definiše kao*

$$E(X) = \int_{-\infty}^{\infty} x f_x dx$$

i ono postoji ako integral $\int_{-\infty}^{\infty} |x| f_x dx < \infty$ apsolutno konvergira.

Definicija 9 *Momenat reda k , $k \in \mathbb{N}$, slučajne promenljive X je $E(X^k)$. Centralni momenat reda k , $k \in \mathbb{N}$, slučajne promenljive X je*

$$E((X - E(X))^k).$$

Dakle, odavde sledi da je matematičko očekivanje momenat reda 1.

Definicija 10 *Centralni momenat reda 2 apsolutno neprekidne slučajne promenljive X se zove disperzija ili varijansa slučajne promenljive X , i označava se sa $D(X)$ ili $\sigma^2(X)$.*

Za izračunavanje disperzije koriste se sledeći izrazi

$$D(X) = E((X - E(X))^2), \text{ ili}$$

$$D(X) = E(X^2) - E^2(X).$$

Definicija 11 *Standardna devijacija (standardno odstupanje, volatilnost) apsolutno neprekidne slučajne promenljive X definiše se kao*

$$\sigma(X) = \sqrt{D(X)}.$$

Dakle, to je brojna karakteristika koja predstavlja meru odstupanja od srednje vrednosti. Zavisnost između slučajnih promenljivih X i Y opisuje se pojmovima *kovarijansa* i *koficijent korelacije*.

Definicija 12 *Kovarijansa slučajne promenljive (X, Y) je*

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

Definicija 13 *Koficijent korelacije slučajne promenljive (X, Y) je*

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sqrt{D(X)D(Y)}} = \frac{\text{cov}(X, Y)}{\sqrt{D(X)D(Y)}}$$

Slučajne promenljive X i Y su nekorelisane ako je $\rho_{XY} = 0$, pozitivno korelisane za $\rho_{XY} > 0$ i negativno korelisane za $\rho_{XY} < 0$.

Neka slučajna promenljiva X ima normalnu raspodelu $\mathcal{N}(\mu, \sigma^2)$, gde je $\mu \in \mathbb{R}$, a $\sigma > 0$. U slučaju kada su parametri normalne raspodele $\mu = 0$ i $\sigma^2 = 1$ dobija se $\mathcal{N}(0, 1)$ raspodela koja se naziva *standardizovana normalna raspodela*. To je jedna od najčešće korišćenih raspodela u teoriji verovatnoće i matematičke statistike. Funkcija gustine i funkcija raspodele slučajne promenljive sa normalnom $\mathcal{N}(\mu, \sigma^2)$ raspodelom predstavljene su jednačinama

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, x \in \mathbb{R}.$$

Na osnovu ovih formula, može se pokazati da je μ očekivanje, a σ^2 disperzija slučajne promenljive X .

Definicija 14 Neka je X slučajna promenljiva sa očekivanjem $E(X)$ i disperzijom $D(X)$. Standardizovana (normalizovana) slučajna promenljiva X^* je

$$X^* = \frac{X-E(X)}{\sqrt{D(X)}} = \frac{X-E(X)}{\sigma(X)}$$

Teorema 1 Neka je X^* standardizovana slučajna promenljiva. Tada je

$$E(X^*) = 0 \text{ i } D(X^*) = 1.$$

Definicija 15 Momenat reda tri standardizovane slučajne promenljive se naziva koeficijent asimetričnosti i definiše se sa

$$S_X = E\left[\frac{(X-E(X))^3}{\sigma_x^3}\right]$$

Ukoliko je $S_X = 0$ raspodela je simetrična, za $S_X > 0$ raspodela je asimetrična na desno, a za $S_X < 0$ raspodela je asimetrična na levo.

Definicija 16 Momenat reda četiri standardizovane slučajne promenljive naziva se koeficijent spljoštenosti i definiše se sa

$$K_X = E\left[\frac{(X-E(X))^4}{\sigma_x^4}\right]$$

Za $K = 3$ raspodela ima normalnu spljoštenost, za $K > 3$ raspodela je više izdužena u odnosu na normalnu raspodelu i za $K < 3$ raspodela je više spljoštena u odnosu na normalnu raspodelu.

Neka je (Ω, \mathcal{F}, P) prostor verovatnoća i neka su događaji $A, B \in \mathcal{F}$, takvi da je $P(B) > 0$. Verovatnoća realizacije događaja A pod uslovima koji dovode do realizacije događaja B , u oznaci $P(A|B)$, naziva se *uslovna verovatnoća* i računa se na sledeći način

$$P(A|B) = \frac{P(AB)}{P(B)}, P(B) > 0.$$

$E(X|B)$ predstavlja očekivanu vrednost slučajne promenljive X , pod uslovom da se realizovao događaj B i prati sledeću formulu

$$E(X|B) = \frac{1}{P(B)} \int_B X dP.$$

Analogno, očekivanu vrednost slučajne promenljive X , pod uslovom da je poznata vrednost slučajne promenljive Y nazivamo *uslovno očekivanje*, u oznaci $E(X|Y)$.

Definicija 17 Neka je Y prosta slučajna promenljiva i neka je $\{y_1, \dots, y_n\}$ skup vrednosti promenljive Y . Tada je uslovno očekivanje slučajne promenljive X pod uslovom da je poznata vrednost slučajne promenljive Y dato sa

$$E(X|Y) = \begin{cases} E(X|A_1), & \omega \in A_1 \\ \vdots \\ E(X|A_n), & \omega \in A_n \end{cases}$$

ili

$$E(X/Y) : \begin{pmatrix} E(X|A_1) & E(X|A_2) & \cdots & E(X|A_n) \\ P(A_1) & P(A_2) & \cdots & P(A_n) \end{pmatrix}$$

gde je $A_i = \{\omega \in \Omega | Y(\omega) = y_i\}$ i $Y = y_i$ ako $\omega \in A_i$ za $i = 1, \dots, n$.

Teorema 2 Neka je X proizvoljna slučajna promenljiva na (Ω, \mathcal{F}, P) i Y prosta slučajna promenljiva, tada važi:

- $E(X/Y)$ je slučajna promenljiva,
- $E(X/Y)$ je \mathcal{F} -merljivo (jer je $E(X/Y)$ funkcija od Y),
- $\int_A X dP = \int_A E(X/Y) dP$ za sve $A \in \mathcal{F}(Y)$

Teorema 3 Osobine uslovnog očekivanja:

- Ako je X ν -merljivo, tada je $E(X|\nu) = X$ skoro sigurno,
- Ako $a, b \in \mathbb{R}$, tada je $E(aX + bY|\nu) = aE(X|\nu) + bE(Y|\nu)$,
- Ako je X nezavisno od ν , tada $E(X|\nu) = E(X)$ skoro sigurno,
- Ako je X ν -merljivo i ako je XY integrabilna, tada $E(XY|\nu) = XE(Y|\nu)$.

1.2 Pojam stohastičkog procesa, definicija i tipovi vremenskih serija

Pretpostavimo da se u svakom trenutku t vremenskog intervala I posmatra neka karakteristika X , fizičkog sistema koja je slučajnog karaktera. Dakle, za fiksirano $t \in I$ $X(t)$ je slučajna promenljiva. Ovo implicira da na skup svih slučajnih promenljivih $X(t)$, kada t "prolazi" kroz čitav I možemo da gledamo kao na slučajnu funkciju vremena. Na taj način dolazi se do pojma *stohastičkog (slučajnog) procesa*.

Definicija 18 Stohastički (slučajni) proces $\{X(t), t \in I\}$ je familija slučajnih promenljivih definisanih na istom prostoru verovatnoća (Ω, \mathcal{F}, P) , gde je I parametarski skup.

Kako je stohastički proces $X(\omega, t)$ funkcija dva parametra - t i ω , za fiksirani vremenski trenutak $t_0 \in I$ dobija se jedna slučajna promenljiva, a za fiksirano $\omega_0 \in \Omega$ se dobija realna funkcija vremena koja se zove trajektorija (realizacija) stohastičkog procesa.

Definicija 19 *Vremenska serija $\{r_t\}_{t \in N}$ predstavlja niz opservacija (determinističkih ili stohastičkih) uređenih u odnosu na vreme, a to uređenje se obično ostvaruje u jednakim vremenskim intervalima.*

Vremenska serija se može posmatrati kao jedna realizacija stohastičkog procesa ili kao uzorak iz kolekcije svih mogućih realizacija stohastičkog procesa. Cilj analize vremenskih serija je da se na osnovu konkretne vremenske serije izvedu zaključci o karakteristikama celog procesa.

Neka je $\{x_t\}_{t \in T}$ vremenska serija, odnosno slučajni uzorak iz kolekcije na konačnom vremenskom intervalu dužine T . Tada je

- Uzoračka sredina

$$\hat{\mu}_x = \frac{1}{T} \sum_{t=1}^T x_t$$

- Uzoračka disperzija

$$\hat{\sigma}_x^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \hat{\mu}_x)^2$$

- Standardna devijacija uzorka

$$\hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2}$$

- Uzorački koeficijent korelacije

$$\hat{\rho}_{x,y} = \frac{\sum_{t=1}^T (x_t - \hat{\mu}_x)(y_t - \hat{\mu}_y)}{\sum_{t=1}^T (x_t - \hat{\mu}_x)^2 \sum_{t=1}^T (y_t - \hat{\mu}_y)^2}$$

- Koeficijent asimetrije uzorka

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3} \sum_{t=1}^T (x_t - \hat{\mu}_x)^3$$

Koeficijent asimetrije meri simetričnost raspodele u odnosu na očekivanu vrednost tj. daje informaciju da li je većina vrednosti u uzorku manja ili veća od očekivane vrednosti. Ukoliko je manja koeficijent ima negativan predznak, a ukoliko je većina vrednosti iz uzorka veća od očekivane koeficijent ima pozitivan predznak. \hat{S}_x asimptotski prati $\mathcal{N}(0, \frac{6}{T})$ raspodelu. Vrednost test statistike za uzorački koeficijent asimetrije dobija se iz izraza

$$t = \frac{\hat{S}_x}{\sqrt{\frac{6}{T}}}.$$

- Koeficijent spljoštenosti uzorka

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4} \sum_{t=1}^T (x_t - \hat{\mu}_x)^4$$

Ovaj koeficijent meri "debljinu repova" raspodele, odnosno koliko je neka raspodela spljoštena u odnosu na normalnu. Ako je $\hat{K}_x < 3$ to znači da je raspodela više spljoštena u odnosu na normalnu. Za $\hat{K}_x = 3$ raspodela je normalno spljoštena i za $\hat{K}_x > 3$ raspodela je više izdužena u odnosu na normalnu tj. ima deblje repove. $\hat{K}_x - 3$ asimptotski prati $\mathcal{N}(0, \frac{24}{T})$ raspodelu. Vrednost test statistike za uzorački koeficijent spljoštenosti dat je izrazom

$$t = \frac{\hat{K}_x - 3}{\sqrt{\frac{24}{T}}}.$$

Kod vremenskih serija stacionarnost igra jednu od veoma bitnih uloga. Ukoliko se svojstva vremenske serije ne menjaju tokom vremena, onda je data serija *stacionarna*. U suprotnom, ona je *nestacionarna*. U zavisnosti od toga da li je vremenska serija stacionarna ili ne, biramo različite statističke metode za analizu, i pravimo podelu na *stacionarne* i *nestacionarne*.

Definicija 20 Vremenska serija $\{r_t\}_{t \in \mathbb{N}}$ je **striktno stacionarna** ako $(r_{t_1}, \dots, r_{t_k})$ i $(r_{t_1+t}, \dots, r_{t_k+t})$ imaju istu raspodelu za svako t , pri čemu $t_1, \dots, t_k \in \mathbb{N}$.

Odnosno, vremenska serija je striktno stacionarna ako je raspodela $(r_{t_1}, \dots, r_{t_k})$ invarijantna u odnosu na vreme. S obzirom da je ovo dosta jak uslov koji je teško proveriti, uvodi se pojam slabe stacionarnosti.

Definicija 21 Vremenska serija $\{r_t\}_{t \in \mathbb{N}}$ je **slabo stacionarna** ako su $E(r_t)$ i $Cov(r_t, r_{t-l})$ nezavisne od vremenskog trenutka t za svako $l \in \mathbb{Z}$.

Dakle, vremenska serija je slabo stacionarna ako se njena statistička svojstva ne menjaju tokom vremena, odnosno ako važi:

- $E(r_t) = \mu$ (očekivanje je konstanta)
- $Cov(r_t, r_{t-l}) = \gamma_l$ (kovarijansa zavisi samo od l)

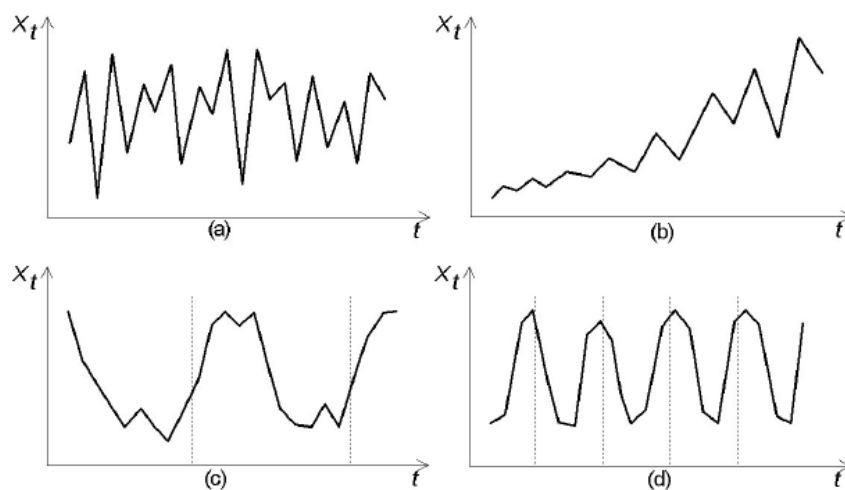
$Cov(r_t, r_{t-l}) = \gamma_l$ se naziva kovarijansa kašnjenja l i važi da je $\gamma_0 = D(r_t)$ i $\gamma_l = \gamma_{-l}$.

Implicitna pretpostavka u slaboj stacionarnosti je da su prva dva momenta konačna, odnosno $E(r_t) < \infty$ i $E(r_t^2) < \infty$.

Striktne ne implicira slabu stacionarnost u opštem slučaju. Samo ako se pretpostavi da postoje odgovarajući momenti ($E(r_t) < \infty$, $E(r_t^2) < \infty$). Takođe, iz slabe stacionarnosti ne sledi stroga stacionarnost. Međutim, uz dodatne uslove, to se može tvrditi.

Nekoliko karakterističnih tipova vremenskih serija prikazano je na Slici 1.2 gde je sa X_t označena opservacija vremenske serije u trenutku t . Prvi tip vremenske serije na Slici 1.2 (a) karakterišu slučajna odstupanja oko konstantnog nivoa. Za takve vremenske serije kažemo da su stacionarne u sredini (nivo serije ne menja se tokom vremena). Na Slici 1.2 (b) prikazana je serija sa trendom, sa slučajnim fluktuacijama oko tog uzlaznog trenda. Ova serija predstavlja primer nestacionarne serije u sredini (nivo serije povećava se tokom vremena). Uporedo sa porastom nivoa ove vremenske serije povećava se i njena varijansa, pa je serija nestacionarna i u varijansi. Na naredne dve slike prikazane su vremenske

serije kod kojih se varijacije u nivou serije periodično ponavljaju. Ako je period ponavljanja sličnog toka serije duži od godinu dana, tada govorimo o cikličnom karakteru te serije (Slika 1.2 (c)).



Slika 1.1: Tipovi vremenskih serija: (a) konstantan proces; (b) proces sa trendom; (c) serija sa cikličnim varijacijama; (d) serija sa sezonskim varijacijama[4]

U daljem tekstu, ukoliko nije naglašeno drugačije, pod pojmom stacionarnosti podrazumevamo slabu stacionarnost.

1.3 Pojam autokorelacije i beli šum

Već je ranije definisan pojam korelacije, koji predstavlja međusobnu povezanost između različitih pojava predstavljenih pomoću dve varijable. Povezanost znači da je vrednost jedne varijable moguće predvideti sa određenom verovatnoćom na osnovu saznanja o vrednosti druge varijable. Pošto je reč o zavisnosti elemenata istog niza, ovom pojmu kao i pojmu kovarijanse, dodaje se prefiks "auto", tj. koriste se nazivi autokorelacija i autokovarijanse. *Autokorelacija* predstavlja korelaciju između sadašnjih i prošlih vrednosti iste vremenske serije. Korelacija između vrednosti r_t i r_{t-l} se naziva autokorelacija kašnjenja l .

Definicija 22 *Autokorelacija kašnjenja l (ACF) slabo stacionarne serije $\{r_t\}_{t \in \mathbb{N}}$ je*

$$\rho_l = \frac{\text{Cov}(r_t, r_{t-l})}{\sqrt{\text{Var}(r_t)\text{Var}(r_{t-l})}} = \frac{\text{Cov}(r_t, r_{t-l})}{\text{Var}(r_t)} = \frac{\gamma_l}{\gamma_0},$$

gde zbog slabe stacionarnosti važi da je $\text{Var}(r_t) = \text{Var}(r_{t-l})$.

ACF zavisi samo od l i iz definicije se zaključuje da je $\rho_0 = 1$, $\rho_l = \rho_{-l}$ i $-1 \leq \rho_l \leq 1$.

Definicija 23 *Uzoračka autokorelacija kašnjenja l (ACF) slabo stacionarne vremenske serije $\{r_t | t = 1, \dots, T\}$ je*

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq l \leq T-1 \quad (1.1)$$

gde je \bar{r} uzoračka srednja vrednost vremenske serije.

Testiranje autokorelacije za određenu vremensku seriju može se vršiti na dva načina. Prvi je pojedinačno testiranje za svako $l \in [0, T-1]$ pomoću t -testa koji govori da li postoji korelacija između r_t i r_{t-l} . Testira se nulta hipoteza $H_0(\rho_l = 0)$, tj. da r_t i r_{t-l} nisu korelisane, protiv alternativne hipoteze da je $H_1(\rho_l \neq 0)$. Prilikom testiranja koristi se sledeća test statistika

$$t\text{-ratio} = \frac{\hat{\rho}_l}{\sqrt{(1 + 2 \sum_{i=0}^{l-1} \hat{\rho}_i^2)/T}} \quad (1.2)$$

Ako je $|t\text{-ratio}| > |z_{\alpha/2}|$ hipoteza H_0 se odbacuje na nivou poverenja α , tj. postoji korelacija između r_t i r_{t-l} , a $z_{\alpha/2}$ je kvantil standardizovane normalne raspodele. U suštini, izračuna se verovatnoća p za test statistiku $t\text{-ratio}$ i ukoliko je p manje od α nulta hipoteza se odbacuje. Drugi način testiranja pomoću Ljung-Box statistike omogućava proveru da li postoji autokorelacija između više članova niza odjednom. Dakle, testira se hipoteza $H_0(\rho_1 = \dots = \rho_m = 0)$ protiv alternativne $H_1(\rho_i \neq 0)$ za neko $i \in \{1, \dots, m\}$. Ljung-Box statistika je

$$Q(m) = T(T+2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{T-l} \quad (1.3)$$

koja ima asimptotsku χ^2 raspodelu sa m stepeni slobode. Ako je $Q(m) > \chi_{m,\alpha}^2$, tj. verovatnoća p Ljung-Box test statistike manja nego nivo poverenja α , odbacuje se nulta hipoteza H_0 .

Definicija 24 Vremenska serija $\{r_t\}$ se naziva **beli šum** ako je $\{r_t\}$ niz nezavisnih slučajnih promenljivih sa istom raspodelom i sa konačnim očekivanjem i disperzijom. Gausov beli šum je beli šum sa raspodelom $\mathcal{N}(0, \delta^2)$.

Za beli šum, autokorelaciona funkcija je 0.

Teorema 4 Neka $\{r_t\}$ prati $AR(1)$ model, tj. $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, gde je a_t beli šum. Ako je $\{r_t\}$ slabo stacionarna vremenska serija, tada je $|\phi_1| < 1$.

Na osnovu Teoreme 4 razlikuju se sledeća tri slučaja:

1. $|\phi_1| < 1$, serija je slabo stacionarna,
2. $|\phi_1| > 1$, serija eksplodira,
3. $|\phi_1| = 1$, serija ima jedinični koren i nestacionarna je.

Od značaja je posmatrati ponašanje γ_l i ρ_l pri povećanju koraka l . Za stacionaran proces i $|\phi_1| < 1$ ACF opada eksponencijalno po apsolutnoj vrednosti. Pritom, znak ρ_l zavisi od znaka ϕ_1 . Ako je ϕ_1 pozitivan, tj. $0 < \phi_1 < 1$, onda su sve autokorelacije pozitivne, a ako je $-1 < \phi_1 < 0$, znak autokorelacije se naizmenično menja, počevši sa negativnom vrednošću.

1.4 Autoregresivni modeli (AR)

U analizi vremenskih serija AR modeli predstavljaju alat pomoću kog se mogu predvideti buduće vrednosti neke vremenske serije. U ovom poglavlju ćemo približnije objasniti kako izgleda autoregresivni model. Prikazaćemo neke njegove značajne osobine i videti kako se koristi za formiranje prediktivnih vrednosti. Pre uvođenja formalnih definicija modela, uvode se još neki bitni pojmovi:

Definicija 25 Vremenska serija $\{r_t\}$, $t = 1, \dots, T$ je linearna ako se može prikazati u obliku

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

gde je μ aritmetička sredina niza $\{r_t\}$, $\psi_0 = 1$, a $\{a_t\}$, $t = 1, \dots, T$ predstavlja niz nezavisnih, jednako raspodeljenih slučajnih promenljivih sa očekivanjem nula i dobro definisanom disperzijom, tj. $\{a_t\}$ je beli šum.

$\{a_t\}$ predstavlja nove informacije u trenutku t koje se nazivaju inovacije ili šokovi. Dinamiku u nizu $\{r_t\}$ određuju koeficijenti ψ_i . Ako je $\{r_t\}$ slabo stacionarna serija, očekivanje i disperzija su

$$E(r_t) = \mu,$$

$$D(r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2$$

gde je σ_a^2 disperzija šuma. Činjenica da je $D(r_t) < \infty$ implicira da $\{\psi_i^2\}$ mora da konvergira, odnosno $\psi_i^2 \rightarrow 0$ kada $i \rightarrow \infty$. To ustvari znači da uticaj šoka a_{t-i} na r_t opada kako i raste.

Autoregresivni proces je regresivni proces u kome je zavisna promenljiva predstavljena članom vremenske serije u trenutku t , dok skup nezavisnih promenljivih čine članovi iste vremenske serije u trenucima $t-1, t-2, \dots, t-p$.

Autoregresivni model reda p ili $AR(p)$ definiše se na sledeći način.

Definicija 26 *Proces $\{r_t\}_{t \in T}$ je autoregresivni proces reda p , $AR(p)$, ako se može predstaviti u obliku*

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots + \phi_p r_{t-p} + a_t$$

gde $p \in \mathbb{N}_0$ predstavlja red modela i govori koliko daleko se ide u prošlost, a_t je beli šum sa očekivanjem nula i disperzijom σ_a^2 , a ϕ_i su koeficijenti modela (determinističke veličine).

Ovo je regresioni model, gde je zavisna promenljiva predstavljena članom vremenske serije u trenutku t , a objašnjavajuće promenljive su članovi iste vremenske serije samo u trenucima $t-1, \dots, t-p$.

$AR(2)$ proces je prvobitno koristio Jil (G. U. Yule), kako bi opisao fenomen broja Sunčevih pega i ponašanje matematičkog klatna. Iz tog razloga se ponekad ovaj proces zove Jilov proces. Prvo će biti razmotreni autoregresivni procesi prvog i drugog reda, kao najjednostavniji. Nakon ispitivanja njihovih osobina, posmatra se opšti autoregresivni proces konačnog reda.

Autoregresivni model prvog reda

Autoregresivni model prvog reda ili $AR(1)$ dat je na sledeći način

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t$$

gde je, kao što je već rečeno a_t beli šum sa očekivanjem nula i disperzijom σ_a^2 , a ϕ_1 autoregresivni koeficijent. Potreban i dovoljan uslov za slabu stacionarnost $AR(1)$ modela navode se u vidu sledeće dve teoreme.

Teorema 5 *Neka $\{r_t\}$ prati $AR(1)$ model, tj. $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, gde je a_t beli šum i važi $E(r_t) = 0$ i $D(r_t) = \sigma_a^2$. Ako je $\{r_t\}$ slabo stacionarna vremenska serija, onda važi $|\phi_1| < 1$.*

Što znači da uticaj prethodnih vrednosti nije velik, kontrolisan je na neki način.

Teorema 6 *Neka $\{r_t\}$ prati $AR(1)$ model, tj. $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, gde je a_t beli šum i važi $E(r_t) = 0$ i $D(r_t) = \sigma_a^2$ i pretpostavimo da je $E(r_t)$ uniformno ograničeno za svako t . Tada je $\{r_t\}$ slabo stacionarna vremenska serija ako je $|\phi_1| < 1$.*

Očekivanje $AR(1)$ modela je dato sa

$$\mu = E(r_t) = \frac{\phi_0}{1 - \phi_1}.$$

Disperzija modela $AR(1)$ predstavljena je formulom

$$\sigma_a^2 = D(r_t) = \frac{\sigma_a^2}{1-\phi_1^2}.$$

Ono što mora da važi da bi očekivanje modela postojalo je da je $\phi_1 \neq 1$. Očekivanje modela je nula ako i samo ako je $\phi_0 = 0$. Disperzija treba da bude ograničena i nenegativna, tako da važi $\phi_1^2 < 1$. Slaba stacionarnost AR(1) modela implicira da $\phi_1 \in (-1, 1)$.

Autokovarijansna i autokorelaciona funkcija kašnjenja l su date sa

$$\gamma = \begin{cases} \phi_1 \gamma_1 + \sigma_a^2, & l = 0 \\ \phi_1 \gamma_{l-1}, & l > 0 \end{cases}$$

$$\rho_l = \phi_1^l,$$

respektivno. Autokorelaciona funkcija kašnjenja slabo stacionarnog modela AR(1) eksponencijalno opada sa stopom ϕ_1 .

Autoregresivni model reda 2

Autoregresivni model drugog reda $AR(2)$, dat je sledećom jednačinom

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t,$$

gde je, a_t beli šum sa očekivanjem nula i disperzijom σ_a^2 , a ϕ_1 i ϕ_2 autoregresivni koeficijenti.

Očekivanje modela $AR(2)$ dato je sa

$$\mu = E(r_t) = \frac{\phi_0}{1-\phi_1-\phi_2}.$$

za koje mora da važi $\phi_1 + \phi_2 \neq 1$.

Disperzija $AR(2)$ modela data je sa

$$\sigma^2 = D(r_t) = \frac{\sigma_a^2 + 2\phi_1\phi_2\gamma_1}{1-\phi_1^2-\phi_2^2}.$$

Kako disperzija treba da bude nenegativna i ograničena treba da važi $\phi_1^2 + \phi_2^2 < 1$.

Autokovarijansna i autokorelaciona funkcija kašnjenja l date su sledećim formulama

$$\gamma = \begin{cases} \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_a^2, & l = 0, \\ \phi_1 \gamma_{l-1} + \phi_2 \gamma_{l-2}, & l > 0, \end{cases}$$

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2}. \quad (1.4)$$

Jednačina (1.4) ukazuje da autokorelaciona funkcija $AR(2)$ modela zadovoljava diferencnu jednačinu drugog reda

$$(1 - \phi_1 B - \phi_2 B^2) \rho_l = 0,$$

gde B predstavlja operator unazad (*engl. back-shift*) za koji važi $B\rho_l = \rho_{l-1}$. Navedena diferencna jednačina određuje osobine autokorelacione funkcije i ponašanje prediktivnih vrednosti.

Odgovarajuća polinomna jednačina za datu diferencnu jednačinu je

$$1 - \phi_1 x - \phi_2 x^2 = 0.$$

Rešenje ove jednačine dato je sa

$$x_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

Inverzna vrednost rešenja $x_{1,2}$ predstavlja karakteristične korene $AR(2)$ modela u oznaci $\omega_{1,2}$. Uslov slabe stacionarnosti postaje $|\omega_{1,2}| < 1$.

Autoregresivni model reda p

$AR(p)$ model je uopštenje modela $AR(1)$ i $AR(2)$. Sve osobine se izvode na sličan način, jedino što dodatno komplikuje račun jeste veći broj koeficijenata. Očekivanje je dato sa

$$\mu = E(r_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p}.$$

Uvođenjem operatora unazad za $AR(p)$ model dobijamo diferencnu jednačinu reda p

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \rho_t = 0,$$

Analogno, kao u $AR(2)$ modelu imamo polinomnu jednačinu

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0.$$

Karakteristični koreni modela $AR(p)$ su inverzne vrednosti rešenja ove polinomne jednačine i uslov slabe stacionarnosti je $|\omega_i| < 1$, gde je $i \in \{1, 2, \dots, p\}$.

1.4.1 Određivanje reda p

Iako se u radu koriste fiksirane vrednosti za red modela, ukratko će biti predstavljeno kako se on određuje u slučaju da se kreće od serije za koju tek treba da se odredi koji model prati.

Često se dešava da je korelacija između dve promenljive rezultat njihove korelisanosti sa trećom promenljivom. Iz tih razloga uvodi se parcijalna korelaciona funkcija (PACF) koja predstavlja korelaciju između dve promenljive uz eliminisan uticaj drugih promenljivih iz modela.

$$\begin{aligned} AR(1): r_t &= \phi_{01} + \phi_{11} r_{t-1} + e_{1t} \\ AR(2): r_t &= \phi_{02} + \phi_{12} r_{t-1} + \phi_{22} r_{t-2} + e_{2t} \\ &\dots \end{aligned}$$

$PACF(p) = \hat{\phi}_{pp}$ predstavlja dodatni doprinos uključivanja r_{t-p} na $AR(p-1)$ model. Ako je $\hat{\phi}_{pp} \approx 0$, tada nema puno smisla uključivati nove parametre u model. Vrednosti parcijalnih korelacionih koeficijenata su jednaki nuli za $k > p$ gde je p red modela.

1.4.2 Provera adekvatnosti modela

Ukoliko su ocenjeni parametri $\hat{\phi} = (\hat{\phi}_p, \dots, \hat{\phi}_0)$ dobijeni na primer metodom najmanjih kvadrata, model je **adekvatan** ako dovodi do toga da se reziduali ponašaju kao beli šum. Odnosno, da se

$$r_t - \hat{r}_t = \hat{a}_t$$

ponaša kao beli šum, gde je $\hat{r}_t = \hat{\phi}_0 + \dots + \hat{\phi}_p \hat{r}_{t-p}$. Adekvatnost modela se ispituje pomoću Ljung-Box statistike gde se testira nulta hipoteza da ne postoji autokorelacija među rezidualima.

1.4.3 Test normalnosti

Test koji se najčešće koristi za ispitivanje normalnosti serije je Jarque-Bera, čija je test statistika:

$$JB = \frac{T}{6} (\hat{S}^2 + \frac{(\hat{K}-3)^2}{4})$$

i ona asimptotski prati χ^2 raspodelu sa dva stepena slobode. Testira se nulta hipoteza

$$H_0 : JB = 0$$

nasuprot

$$H_1 : JB > 0$$

Ukoliko je vrednost test statistike veća od kvantila χ^2 raspodele sa dva stepena slobode, ili ukoliko je registrovana p vrednost manja od zadatog nivoa značajnosti - nulta hipoteza se odbacuje. U suprotnom je prihvatamo. Takođe, ovaj test ispituje da li ranije definisani koeficijenti asimetrije i spljoštenosti odgovaraju normalnoj raspodeli. Nulta hipoteza koja se testira je da ovi koeficijenti odgovaraju normalnoj raspodeli, odnosno da su koeficijenti asimetrije i spljoštenosti jednaki nula i tri, redom.

1.4.4 Test jediničnih korena

Kako je jedna od osnovnih karakteristika vremenskih serija tendencija stalnog rasta ili opadanja, formalni postupak za testiranje prisustva i tipa nestacionarnosti u modelima linearnih vremenskih serija čini grupa testova koji se zovu testovi jediničnog korena. Definisan je veći broj testova jediničnog korena, ali u ovom radu stacionarnost će se ispitivati pomoću Dickey-Fuller testa jediničnih korena (DF).

Radi jednostavnosti, ovaj test će biti opisan samo za AR(1) model

$$r_t = \phi_1 r_{t-1} + a_t$$

gde je a_t beli šum. Na osnovu ranije definisanog *back-shift* operatora i AR polinoma, sledi da je AR(1) model slabo stacionaran ako je $|\phi_1| < 1$. Ako je $|\phi_1| \geq 1$ serija r_t je nestacionarna i varijansa teži u beskonačno kako vreme ide u beskonačno.

DF test testira nultu hipotezu $H_0 : |\phi_1| = 1$ protiv alternativne $H_1 : |\phi_1| < 1$. Ako se u AR(1) modelu sa obe strane oduzme r_{t-1} dobija se

$$r_t - r_{t-1} = (\phi_1 - 1)r_{t-1} + a_t$$

odnosno

$$\Delta r_t = (\phi_1 - 1)r_{t-1} + a_t$$

gde je $\alpha = \phi_1 - 1$ pa nulta i alternativna hipoteza mogu da se zapišu kao $H_0 : \alpha = 0$ protiv alternativne $H_1 : \alpha < 0$. Ako je $\phi_1 = 1$ dobija se da je

$$\Delta r_t = a_t$$

a kako je a_t beli šum sledi da je Δr_t stacionarna vremenska serija. Na taj način se vremenska serija diferenciranjem svodi na stacionarnu vremensku seriju. Može se desiti da je nestacionarne vremenske serije potrebno diferencirati više od jednog puta da bi postale stacionarne.

1.4.5 Formiranje prediktivnih vrednosti AR modela

Predviđanje, tj. određivanje budućeg toka posmatrane vremenske serije je jako bitan korak u procesu analize jedne vremenske serije. Pretpostavimo da se nalazimo u trenutku t i zanima nas vrednost serije u trenutku $t+1$. Posmatramo AR(p) model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots + \phi_p r_{t-p} + a_t$$

Predikcija za jedan korak unapred, odnosno predviđena vrednost za trenutak $t+1$ na osnovu AR(p) modela dobija se na sledeći način

$$\hat{r}_t(1) = E(r_{t+1}|\mathcal{F}_t) = E(\phi_0 + \phi_1 r_t + \phi_2 r_{t-1} + \dots + \phi_p r_{t+1-p} + a_{t+1}|\mathcal{F}_t),$$

$$\hat{r}_t(1) = \phi_0 + \phi_1 r_t + \phi_2 r_{t-1} + \dots + \phi_p r_{t+1-p} + a_t, \quad (1.5)$$

gde je \mathcal{F}_t poznata istorija do trenutka t . U jednačini (1.5) koristili smo osobine uslovnog očekivanja $E(a_{t+1}|\mathcal{F}_t) = E(a_{t+1}) = 0$. Greška predviđanja za jedan korak unapred je

$$e_t(1) = r_{t+1} - \hat{r}_t(1) = a_{t+1}.$$

Ako nam je poznata istorija do trenutka t , stvarna vrednost serije u trenutku $t+1$ i formirana prediktivna vrednost za taj trenutak razlikuju se samo za vrednost šuma u trenutku $t+1$.

Pretpostavimo da se nalazimo u trenutku t i zanima nas vrednost serije u trenutku $t+2$. Dvokoračna predikcija na osnovu modela AR(p) dobija se na sledeći način

$$\hat{r}_t(2) = E(r_{t+2}|\mathcal{F}_t) = E(\phi_0 + \phi_1 r_{t+1} + \phi_2 r_t + \dots + \phi_p r_{t+2-p} + a_{t+2}|\mathcal{F}_t),$$

$$\hat{r}_t(2) = E(r_{t+2}|\mathcal{F}_t) = \phi_0 + \phi_1 \hat{r}_t(1) + \phi_2 r_t + \dots + \phi_p r_{t+2-p}. \quad (1.6)$$

Greška predviđanja za dva koraka unapred je

$$e_t(2) = r_{t+2} - \hat{r}_t(2) = \phi_1 a_{t+1} + a_{t+2}.$$

Uopštenje formule za formiranje prediktivnih vrednosti za l koraka unapred, ukoliko znamo istoriju do trenutka t prati formulu

$$\hat{r}_t(l) = E(r_{t+l}|\mathcal{F}_t) = \phi_0 + \sum_{i=1}^p \phi_i \hat{r}_t(l-i),$$

i važi $\hat{r}_t(i) = r_{t+i}$, ako je $i \leq 0$. Ovakav način predviđanja se naziva dinamičko predviđanje i zasniva se na određivanju predikcija za periode nakon prvog perioda u uzorku, tako što se koriste prethodne predviđene vrednosti. Nasuprot dinamičkom, je statički način predviđanja, koji se zasniva na određivanju predikcija za periode nakon prvog perioda u uzorku, tako što se koriste prethodne stvarne vrednosti. U slučaju jednokoračnih predikcija vrednosti dinamičkih i statičkih predikcija se poklapaju.

Statičko predviđanje podrazumeva da ako formiramo predikcije za period $t+l$, poznata nam je istorija do trenutka $t+l-1$. Tako da statičko predviđanje prati sledeću formulu

$$\hat{r}_t(l) = E(r_{t+l}|\mathcal{F}_{t+l-1}) = \phi_0 + \sum_{i=1}^p \phi_i r_t(t+l-i).$$

U svakom koraku, prediktivne vrednosti dobijene statičkom metodom će se od stvarnih razlikovati samo za vrednost šuma u tom koraku.

Statičko predviđanje ćemo koristiti uvek kada imamo poznate vrednosti za period za koji predviđamo. Tačnije, to je predviđanje "od danas do sutra". Kod dinamičkog načina predviđanja greška se povećava kako se povećava broj perioda za koje tražimo prediktivne vrednosti, jer se koriste vrednosti prethodnih predikcija, koje se razlikuju od stvarnih vrednosti. Tada dolazi do "gomilanja" greške. Dinamički metod se koristi kada na primer treba da se vidi šta će se desiti sledeće godine, na osnovu podataka iz prethodnih nekoliko godina.

1.4.6 Greške predviđanja

Greške predviđanja predstavljaju pokazatelj koliko je model pomoću kog su formirane predikcije dobar. Što su greške predviđanja bliže nuli, model je bolji. Kako se model pravi na 80% uzorka, a za preostalih 20% formiramo prediktivne vrednosti, upoređujući stvarne i prediktivne vrednosti iz uzorka dolazi se do *grešaka predviđanja*.

Neka T predstavlja veličinu uzorka, odnosno serije. Period za koji se prave predikcije označimo sa $T+1, \dots, T+h$, a stvarnu i prediktivnu vrednost za trenutak t sa r_t i \hat{r}_t respektivno.

Standardna devijacija greške predviđanja ($RMSE$) je kvadratni koren prosečne sume kvadrata odstupanja i prati formulu

$$RMSE = \sqrt{\frac{1}{h} \sum_{t=T+1}^{T+h} (\hat{r}_t - r_t)^2} \quad (1.7)$$

Prosečna apsolutna greška (*MAE*) predstavlja prosečnu sumu apsolutnih odstupanja i predstavljena je sledećom jednakosću

$$MAE = \frac{1}{h} \sum_{t=T+1}^{T+h} |\hat{r}_t - r_t| \quad (1.8)$$

Prosečno apsolutno procentualno odstupanje (*MAPE*) prati jednakost

$$MAPE = \frac{100}{h} \sum_{t=T+1}^{T+h} \left| \frac{\hat{r}_t - r_t}{r_t} \right| \quad (1.9)$$

Nisu sve vrste prethodno navedenih grešaka međusobno uporedive. Prve dve greške *RMSE* i *MAE* zavise od skale u kojoj su podaci izraženi. Trebalo bi da se koriste za upoređivanje prediktivne moći iste vremenske serije kroz različite modele. Preostali pokazatelj je uporediv i za modele ocenjene na različitim vremenskim serijama. Za *MAPE* ne postoji gornje ograničenje i što je vrednost bliža nuli, model je bolji.

2 Ocenjivanje parametara

Postoje razni načini i mehanizmi za ocenjivanje parametara:

- Metoda momenta (distribucija populacije ocenjuje se odgovarajućim momentom uzorka)
- Metoda najmanjih kvadrata (minimizira se kvadrat razlike odstupanja stvarnih od predviđenih vrednosti)
- Metoda maksimalne verodostojnosti (metoda izbora jedne vrednosti parametara modela kao ocene parametara, ali tako da funkcija verodostojnosti ima što je moguće veću vrednost)
- Metoda najboljeg linearnog nepristrasnog ocenjivača (ocenjivač je linearan i ima minimalnu varijansu u odnosu na ocenjivače dobijene drugim metodama), itd.

U radu se od prethodno navedenih, između ostalog, koristi i metod najmanjih kvadrata. S obzirom da se sve bazira na simuliranim vremenskim serijama, red modela će unapred biti određen, što znači da se odmah može preći na ocenjivanje parametara modela. Za početak, ocena koeficijenata vrši se metodom najmanjih kvadrata. Ova metoda je najčešće korišćena jer u slučaju odsustva autokorelacije i u slučaju heteroskedastičnosti daje nepristrasne i konzistentne ocene. Nakon što se ovom metodom dobiju početni koeficijenti, na red dolazi primena druga dva metoda, a to su: *Stohastički gradijentni spust (SGD)* i *Gradijentni metod sa rastućim uzorkom (IGD)*.

2.1 Problem najmanjih kvadrata

Suština problema najmanjih kvadrata je minimiziranje sume kvadrata odstupanja ocenjene vrednosti od stvarne, tačne. Drugim rečima, izmereni

rezultati nisu egzaktni, već sadrže neku grešku; te greške mogu nastati kao posledice: nesavršenosti mernih instrumenata, nesavršenosti uslova eksperimenata, ljudskog faktora itd. Ukoliko bi to predstavili grafički, pogodno je da dobijeni grafik što manje odstupa od izmerenih rezultata, tj. da predstavlja dobru aproksimaciju. Postavlja se pitanje, kako definisati najbolju aproksimaciju. U praksi se najčešće minimizuje zbir kvadrata odstupanja, što se još naziva i l_2 aproksimacija ili aproksimacija najmanjih kvadrata koja se i koristi u radu.

Za $AR(p)$ model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + a_t \quad (2.1)$$

gde su $\{r_t\}_{t=1,\dots,T}$ stvarne vrednosti vremenske serije, uslovno očekivanje koje se koristi za dobijanje predikcija dato je sa

$$\hat{r}_t = E(r_t | r_{t-1}) = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} = h_t(\vec{\phi}).$$

Greška u trenutku t data je sa

$$(h_t(\phi) - r_t)^2$$

odnosno, ako se formalizuje izraz funkcije h čiji oblik je poznat onda se nalaženje minimuma svodi na minimum izraza

$$\sum_{t=p+1}^T (h_t(\phi) - r_t)^2.$$

Može se primetiti da je izraz kvadriran da bi se izbegla negativna vrednost, s obzirom da apsolutna vrednost ne može da se iskoristi jer nije diferencijabilna. Upravo zato je metod najmanjih kvadrata i najkorišćeniji.

Dakle, minimizirana funkcija ima sledeći oblik

$$\min_{\phi} f(\phi) = \frac{1}{2} \|F(\phi) - y\|^2, \quad (2.2)$$

gde je $F(\phi) = \begin{bmatrix} h_{p+1}(\phi) \\ \vdots \\ h_T(\phi) \end{bmatrix}$ a $y = \begin{bmatrix} r_{p+1} \\ \vdots \\ r_T \end{bmatrix}$. Jednu polovinu uvodimo samo iz računskih razloga.

U gore navedenoj jednačini (2.1) $\phi_0, \phi_1, \dots, \phi_p$ predstavljaju parametre modela (težine), koje treba odabrati tako da "udaljenost" $h_t(\phi)$ i r_t bude minimalna. To znači da je potrebno odabrati funkciju greške koja govori koliko model funkcija (sa trenutnim vrednostima parametara) odskače od stvarnih vrednosti y .

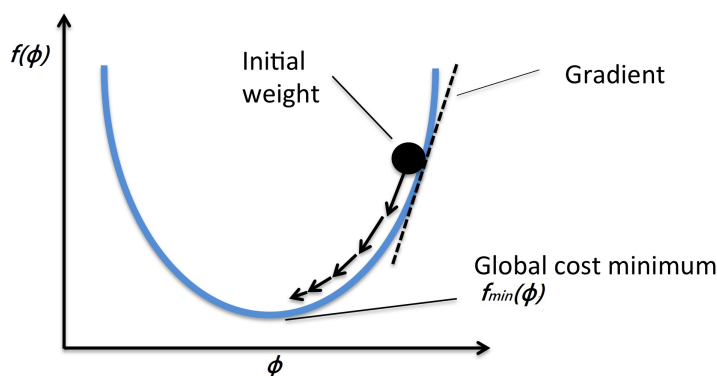
U praksi se za optimizaciju modela najčešće koristi **gradijentni metod**. To je iterativni metod traženja minimuma funkcije koji se zasniva na korišćenju gradijenta - generalizacije izvoda na veći broj promenljivih. Gradijent je vektor čiji su elementi parcijalni izvodi funkcije, pa je tako

$$f(\phi) = \left(\frac{\partial f(\phi)}{\partial \phi_0}, \frac{\partial f(\phi)}{\partial \phi_1}, \dots, \frac{\partial f(\phi)}{\partial \phi_p} \right)^T$$

gradijent funkcije greške. U linearnoj regresiji se najčešće kao funkcija greške koristi prosek kvadrata odstupanja $h(\phi)$ od y (engl. *Mean Squared Error - MSE*):

$$f(\phi) = \frac{1}{2m} \sum_{i=1}^m (h(\phi_i) - y_i)^2 \quad (2.3)$$

gde je m veličina uzorka, a koji takođe može biti zapisan i u obliku (2.2). Minimum ove funkcije moguće je pronaći analitički, što bi podrazumevalo računanje inverznih matrica a to može biti preskupa operacija kada je broj korišćenih podataka veliki. Na Slici 2.1 data je ilustracija gradijentnog metoda koji će detaljno biti opisan u jednom od narednih poglavlja.



Slika 2.1: Ilustracija gradijentnog metoda [8]

2.2 Optimizacija

S obzirom da analiza vremenskih serija podrazumeva i rešavanje problema optimizacije, pre nego što se uvede pojam gradijentnog metoda biće reči o optimizaciji uopšteno. Matematička optimizacija se bavi metodama pronalaženja minimuma i maksimuma funkcija. Opšti problem optimizacije je obično oblika:

$$\min_{x \in \mathcal{D}} f(x)$$

pri čemu se funkcija f naziva *funkcija cilja*, a skup \mathcal{D} *dopustivi skup*. Tačka iz dopustivog skupa koja zadovoljava sva ograničenja, naziva se *dopustivo rešenje*. Potrebno je među svim dopustivim rešenjima naći ono za koje je vrednost funkcije cilja najmanja. Ova formulacija obuhvata i pronalaženje maksimuma, pošto se pronalaženje maksimuma funkcije f može svesti na pronalaženje minimuma funkcije $-f$. Zato će u nastavku biti reči isključivo o metodama pronalaženja minimuma, tj. *minimizacije*. Problemi optimizacije su uglavnom vrlo zahtevni i obično se rešavaju numerički, pomoću relevantnih iterativnih postupaka. To

znači da se njima generiše niz aproksimacija - polazeći od početne i generišući svaku narednu aproksimaciju koristeći prethodnu. Sve to s ciljem da se dobije niz tačaka koje konvergiraju ka rešenju problema optimizacije.

Problemi optimizacije se mogu podeliti na probleme bez ograničenja i na probleme s ograničenjima. Opšta formulacija problema bez ograničenja je

$$\min_{x \in \mathbb{R}^n} f(x) \quad (2.4)$$

gde je funkcija $f : \mathbb{R}^n \rightarrow \mathbb{R}$ funkcija cilja. Kako je problem koji je predmet istraživanja ovog rada problem optimizacije bez ograničenja, u ovom poglavlju fokusiraćemo se na probleme tog tipa. U slučaju problema s ograničenjima, promenljiva x pripada nekom podskupu od \mathbb{R}^n .

Pod rešavanjem problema (2.4) podrazumeva se utvrđivanje tačke minimuma x^* u kojoj funkcija f ima najmanju vrednost na skupu \mathbb{R}^n . Velika većina metoda za rešavanje problema minimizacije bez ograničenja su konstruisani da konvergiraju u tačke koje zadovoljavaju određene uslove optimalnosti. Potrebni i dovoljni uslovi za rešenje problema (2.4) su veoma važni ne samo u teorijskom smislu, nego i u praksi. Kad se pronade tačka koja zadovoljava potrebne uslove, dovoljni uslovi nam pomažu da vidimo da li je taj kandidat pravi minimizator. U zavisnosti od reda izvoda, razlikuju se potreban uslov prvog i drugog reda, dok su dovoljni uslovi drugog reda.[9]

Za glatke funkcije postoje uslovi koji određuju lokalni minimum. Potreban uslov prvog reda je definisan za neprekidno diferencijabilne funkcije cilja.

Teorema 7 *Ako je x^* tačka lokalnog minimuma problema (2.4) sa neprekidno diferencijabilnom funkcijom f u nekoj okolini tačke x^* , tada je $\nabla f(x^*) = 0$.*

Tačka u kojoj je gradijent funkcije jednak nuli je stacionarna tačka funkcije. Dakle, potreban uslov da funkcija ima minimum u tački x^* je da je x^* njena stacionarna tačka.

Potreban uslov drugog reda, kao i dovoljan uslov drugog reda, dati su za dva puta neprekidno diferencijabilne funkcije.

Teorema 8 *Ako je x^* tačka lokalnog minimuma problema (2.4) sa neprekidno diferencijabilnim gradijentom ∇f u nekoj okolini tačke x^* , tada je $\nabla f(x^*) = 0$ i Hesijan $\nabla^2 f(x^*)$ je pozitivno semidefinitan.*

Teorema 9 *Ako je Hesijan $\nabla^2 f$ neprekidan u nekoj okolini tačke x^* , $\nabla f(x^*) = 0$ i $\nabla^2 f(x^*)$ je pozitivno definitan, tada je x^* tačka strogog lokalnog minimuma problema (2.4).*

Dakle, za tačku koja nema pozitivno semidefinitan Hesijan znamo da ne može biti tačka minimuma. Dok za stacionarnu tačku u kojoj je Hesijan pozitivno definitan znamo da je tačka lokalnog minimuma. Kada je funkcija cilja konveksna, tada imamo informacije i o globalnom minimumu.

Teorema 10 *Ako je funkcija cilja konveksna, tada je njena tačka lokalnog minimuma i tačka globalnog minimuma.*

Kada je funkcija cilja konveksna i diferencijabilna, tada je svaka stacionarna tačka ujedno i tačka globalnog minimuma.

Ako je funkcija cilja strogo konveksna, tada je njena tačka globalnog minimuma jedinstvena.

Metodi rešavanja - Linijsko pretraživanje

S obzirom da su metodi za rešavanje problema optimizacije iterativni, odabere se početna tačka x_0 i generiše se niz iterativnih tačaka $\{x_k\}_{k \in \mathbb{N}}$ tako da se naredna tačka određuje na osnovu prethodne. Niz $\{x_k\}$ treba da bude generisan tako da teži ka rešenju x^* problema (2.4). Stoga se izbor naredne iterativne tačke vrši sa ciljem da se dobije manja vrednost funkcije cilja. Tako se u k -toj iteraciji bira naredna iterativna tačka x_{k+1} za koju važi da je $f(x_{k+1}) < f(x_k)$. To se čini odabirom pravca d_k i dužine koraka α_k s kojima se iz x_k dolazi u x_{k+1} , što znači da je

$$x_{k+1} = x_k + \alpha_k d_k.$$

Postoje dva osnovna pristupa određivanja pravca i dužine koraka: postupak oblasti poverenja i linijsko pretraživanje. Mi ćemo se ukratko osvrnuti samo na postupak linijskog pretraživanja.

Postupkom linijskog pretraživanja se u svakoj iteraciji prvo odredi pravac pretraživanja d_k , a potom dužina koraka α_k tako da se umanju vrednost funkcije cilja. Uzimajući za pravac pretraživanja opadajući pravac - pravac d_k sa osobinom da je

$$d_k \nabla^T f(x_k) < 0,$$

obebeđuje se da se duž pravca d_k može smanjiti vrednost funkcije f . Neki metodi koriste proizvoljan opadajući pravac pretraživanja, dok ga drugi konkretno definišu. Mi ćemo se u ovom delu više posvetiti tome kako se vrši izbor dužine koraka u linijskom pretraživanju.

Za izbegavanje dugačkih pravaca po kojima dobijamo jako malo smanjenje vrednosti funkcije cilja, pri linijskom pretraživanju zahtevamo da α_k zadovoljava

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \eta \alpha_k \nabla^T f(x_k) d_k \quad (2.5)$$

gde je $\eta \in (0, 1)$ konstanta. Ovaj uslov zapravo zahteva da je opadanje vrednosti funkcije proporcionalno dužini koraka. Primetimo da, kako je d_k opadajući pravac, važi

$$\eta \alpha_k \nabla^T f(x_k) d_k < 0$$

i zbog toga zahtevana nejednakost zapravo pokazuje da tražimo nešto više od običnog smanjenja vrednosti funkcije. Ovaj uslov se naziva *uslov dovoljnog opadanja*, a poznat je i pod nazivom *Armijo uslov*.

2.3 Gradijentni metod

Najjednostavnija i najpoznatija metoda optimizacije prvog reda za diferencijabilne funkcije je **gradijentni metod** (eng. *gradient descent*). Ako bi se gradijentni metod pokušao opisati intuitivno to bi izgledalo ovako: Zamislite da ste na planini i treba da se spustite do ravne površine sa povezom na očima bez ičije pomoći. Jedina pomoć koju imate je gedžet koji vam govori na kojoj ste nadmorskoj visini (tj. visini od nivoa mora). Šta bi bio vaš pristup? Počeli biste da se spuštate u proizvoljnom pravcu i zatim biste proverili visinu na kojoj se nalazite. Ukoliko bi gedžet pokazao veću visinu nego što je bila prvobitna, to bi bio znak da ste krenuli u pogrešnom pravcu. Promenite pravac i ponovite proces. Ponavljanjem postupka na opisan način, korak po korak uspešno stižete na odredište.

Gradijentni metod je najčešći algoritam optimizacije. To je algoritam optimizacije prvog reda što znači da uzima u obzir samo prvi izvod prilikom izvođenja tj. ažuriranja parametara. Ova metoda, kao i većina metoda optimizacije, zasniva se na postepenom, iterativnom približavanju rešenju problema. Gradijent ukazuje na pravac najbržeg uspona. Stoga, negativna vrednost gradijenta ukazuje na pravac najbržeg pada. Osnovna ideja gradijentnog spusta je da se, polazeći od neke nasumice odabrane tačke, nizom koraka u pravcu gradijenta dođe vrlo blizu rešenju. Ako je polazna tačka x_0 , svaka naredna se dobija primenom pravila

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \quad (2.6)$$

Bitno je napomenuti da se početne vrednosti parametara unapred pretpostavljaju a zatim se pomoću ove metode pronalaze nove i približava se željenoj vrednosti. α_k je pozitivan broj koji predstavlja dužinu koraka, tj. brzinu učenja (eng. *learning rate*) - ona određuje koliko će se veliki skokovi niz gradijent praviti. U vezi sa ovim pristupom, postavlja se više pitanja. Prvo je kako se bira dužina koraka α_k koji se preduzima u pravcu suprotnom gradijentu. Jedan jednostavan izbor je korišćenje konstantne vrednosti koraka $\alpha_k = \alpha$, za svako k .

U tom slučaju, izraz (2.6) dobija oblik

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \quad (2.7)$$

Gradijentni metod može da konvergira sa ovakvom, fiksnom vrednošću koraka α . Pretpostavimo da je funkcija cilja dva puta neprekidno diferencijabilna i njen gradijent Lipšic neprekidan, tj. za sve $x, y \in \mathbb{R}^n$ važi sledeće

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad (2.8)$$

Primetimo, pod pretpostavkom da je $f \in C^2(\mathbb{R}^n)$, uslov (2.8) zapravo znači da su maksimalni karakteristični koreni matrice Hesijana ustvari ograničeni od gore vrednošću L . Dakle, može se dokazati konvergencija konveksnih funkcija. Konvergencija je osigurana samo ako je veličina koraka dovoljno mala. Štaviše, teorema koja sledi ukazuje da je stopa konvergencije R-sublinearna za $f(x_k)$ [9].

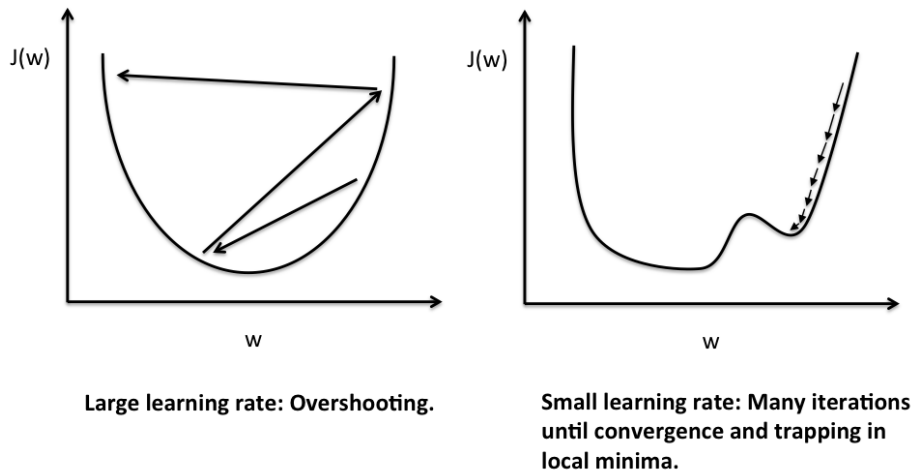
Teorema 11 [9] *Pretpostavimo da je $f \in C^2(\mathbb{R}^n)$ konveksna i da važi (2.8). Tada, ako je $\alpha < \frac{1}{L}$, gradijentni metod (2.7) sa fiksnom vrednošću koraka zadovoljava sledeću nejednakost*

$$f(x_k) - f(x^*) \leq \frac{\|x_0 - x^*\|^2}{2\alpha k}.$$

Takođe, često se α_k bira tako da zadovoljava uslov dovoljnog pada, tj. *Armijo* uslov dat formulom (2.5).

Ono što je sigurno, to je da dužina koraka α mora da bude pažljivo odabrana. Na Slici 2.3 je ilustrovano kakav uticaj ima prevelika a kakav premala dužina koraka α .

- Kada je α premalo kretanje je veoma sporo tj. ima veliki broj koraka. U zavisnosti od oblika funkcije greške, moguće je i zaglavljivanje u lokalnim minimumima.
- Kada je α preveliko, optimizacija može da preskoči minimum ili čak i da divergira.



Slika 2.2: Ilustracija uticaja prevelike i premale dužine koraka α [7]

Kako se spust približava minimumu, vrednost parcijalnih izvoda se sama smanjuje tako da automatski dolazi do konvergencije. Vrednost izvoda u minimumu je jednaka nuli, tako da se parametri x više ne menjaju ako dostignu optimalne vrednosti. Ukoliko promena vrednosti parametara u jednom koraku postane manja od neke predefinisane vrednosti, to je znak da se staje sa iteracijama. Praktično nikad se ne dešava da se dođe baš u rešenje, tj. da gradijent bude nula u praksi. Postoji samo asimptotski rezultat, kad k teži beskonačnosti što u praksi daje približno rešenje.

To je ujedno i odgovor na drugo pitanje koje se nameće - Kada se staje sa izračunavanjem? Broj koraka potrebnih za konvergenciju zavisi od konkretnog zadatka koji se rešava i može znatno da varira. Ono što može predstavljati automatsku proveru konvergencije jeste provera - da li se greška smanjuje u svakom koraku. Za dovoljno malo α greška treba da se smanjuje pri svakom koraku.

Kriterijuma zaustavljanja koji se koriste u praksi ima više. Najčešći su zaustavljanje nakon unapred zadatog broja iteracija, nakon što razlika između susednih koraka $\|x_{k+1} - x_k\|$ postane manja od unapred zadate vrednosti ε , nakon što razlika između vrednosti funkcije u susednim koracima $|f(x_{k+1}) - f(x_k)|$ postane manja od ε ili nakon što ova razlika u odnosu na polaznu vrednost funkcije $|f(x_{k+1}) - f(x_k)|/|f(x_0)|$ postane manja od ε . Takođe, još jedan od dosta korišćenih kriterijuma je taj da je norma gradijenta manja od ε .

Ako je dužina koraka dobro odabrana, gradijentni metod će uvek konvergirati ka globalnom minimumu. U našem slučaju, funkcija greške definisana izrazom (2.3) je kvadratna, konveksna funkcija, pa je njena tačka lokalnog minimuma ujedno i tačka globalnog minimuma.

U sumi, prednosti metode najbržeg pada su njena jednostavnost i široki uslovi primenljivosti, a mane su spora konvergencija, to što je izabrani pravac samo lokalno optimalan, što dodatno usporava konvergenciju cik-cak kretanjem i to što se u mnogim slučajevima za izračunavanje tog optimalnog pravca troši puno vremena. U slučaju kada imamo veliki broj podataka (Big Data problemi), klasičan gradijentni metod može da bude veoma skup. Zbog toga se vrše modifikacije pomenutog metoda i jedna od njih je **Stohastički gradijentni metod** (eng. *SGD - Stochastic Gradient Descent*).

2.3.1 Stohastički gradijentni metod

Jedna, u mnogome primenjivana, modifikacija gradijentnog metoda je **stohastički gradijentni metod** (eng. *Stochastic Gradient Descent*). Stohastički gradijentni metod se primenjuje u modelima sa velikim količinama podataka. Modifikacija se sastoji u tome da je umesto gradijenta dovoljno koristiti neki slučajni vektor čije je očekivanje kolinearno sa gradijentom i istog je smera. Ovakva modifikacija ima smisla pre svega kada se funkcija koja se optimizuje može predstaviti kao prosek drugih jednostavnijih funkcija:

$$f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (2.9)$$

Ovo je tipičan slučaj gde se minimizuje funkcija greške koja je zbir grešaka na pojedinačnim instancama. U našem slučaju, pojedinačne funkcije $f_i(x)$ su oblika

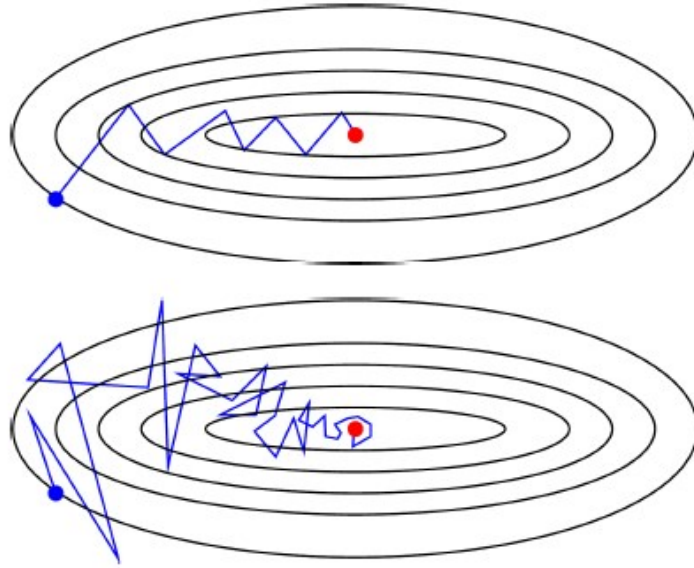
$$f_i(x) = \frac{1}{2}(x^T a_i - r_i)^2$$

gde je x vektor koeficijenata modela oblika $(\phi_0, \phi_1, \dots, \phi_p)^T$ kojih ima onoliko koliki je red modela, a $a_i = (1, r_{i-1}, \dots, r_{i-p})^T$ i r_i su vrednosti očitane iz vremenske serije. Tada je pravilo izračunavanja novog koraka moguće zameniti sledećim pravilom:

$$x_{k+1} = x_k - \alpha_k \nabla f_i(x_k) \quad (2.10)$$

Obično se i bira tako da bude jednako $(k \bmod N) + 1$, odnosno tako da se u svakom koraku koristi naredna funkcija f_i dok se ne dodje do poslednje, a onda se ponovo nastavlja od prve. Ovaj pristup predstavlja jeftinu aproksimaciju gradijenta. Ipak, ona može biti prilično neprecizna, kao što se može videti

sa Slike 2.3.1. Zbog svoje stohastičke prirode put ka globalnom minimumu nije direktan kao kod gradijentnog metoda, ali može ići "cik-cak". Stoga se kao kompromis često, umesto samo jedne od funkcija f_i , koristi prosek nekog podskupa ovih funkcija (eng. *minibatch*). Često se koristi i varijanta kad se na slučajan način (pomoću uniformne raspodele) izabere f_i (ili podskup funkcija).



Slika 2.3: Ponašanje gradijentnog metoda i stohastičkog gradijentnog metoda [11]

Prilikom obrade velikog broja podataka, vreme jedne iteracije gradijentnog metoda, koji u svakoj iteraciji koristi sve podatke, je drastično veće nego u slučaju stohastičkog gradijenta, koji u svakoj iteraciji koristi samo po jednu instancu iz skupa podataka.

Što se tiče brzine učenja, tj. koraka α , za dobijanje dužine koraka u ovom metodu najčešće se koristi izraz $\alpha_k = \frac{1}{k}$ gde k predstavlja broj koraka i uzima vrednosti $k=1, 2, \dots, N$. To je ujedno i jedan od mogućih izbora niza dužina koraka koji zadovoljava Robins-Monroove uslove

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty$$

Intuitivno, smisao prvog uslova je da su koraci dovoljno veliki da se može dostići rešenje problema. Smisao drugog uslova je da su koraci dovoljno mali da niz tačaka x_k konvergira rešenju, umesto da osciluje.

U odnosu na gradijentni metod, prednosti stohastičkog gradijenta su mnogostruke. Gradijent, koji inače može biti skup za izračunavanje, jeftino se aproksimira. Gradijentni metod (GD) mora da prođe kroz ceo skup za obučavanje da bi napravio jedan korak, dok stohastički (SGD) može da krene sa minimizacijom funkcije greške na osnovu samo jednog podatka. SGD se često približi minimumu znatno brže nego grupni. S obzirom da stohastički gradijent teorijski konvergira samo u stohastičkom smislu, u praksi često imamo oscilovanje oko rešenja a ne konvergenciju u klasičnom smislu. Kao što je i prikazano na Slici 2.3.1, on stigne u okolinu rešenja a to kolika je ta okolina zavisi između ostalog od nivoa varijanse. Zbog svoje efikasnosti, SGD se uvek preferira kod velikih skupova podataka jer ovaj metod daje zadovoljavajuće aproksimacije rešenja sa prihvatljivim troškovima.

2.3.2 Gradijentni metod sa rastućim uzorkom

Često se u literaturi nailazi na probleme sa matematičkim očekivanjem. U praksi ponekad nije moguće napraviti analitički oblik tog očekivanja pa se stoga koriste aproksimacije. SAA (engl. *Sample Average Approximation*) metod je veoma rasprostranjen, a njegova osnovna ideja je da se očekivana vrednost aproksimira uzoračkom očekivanom vrednošću.

Za dobru aproksimaciju matematičkog očekivanja, često je potreban veliki uzorak. To u mnogim problemima povlači za sobom veliki broj izračunavanja vrednosti funkcije. U problemima u kojima je evaluacija funkcije skupa, odnosno predstavlja dominantan trošak, direktno rešavanje SAA problema determinističkim postupcima je skupo, te se stoga koriste metodi sa promenljivom veličinom uzorka, takozvani VSS (engl. *Variable Sample Size*) metodi. U VSS metodu se u procesu optimizacije umesto jednog velikog uzorka koriste i uzorci manje veličine. Globalna ideja VSS metoda je da se računa s malim uzorcima dok je iterativna tačka daleko od rešenja, a da se koriste veliki uzorci kada se iterativna tačka približi rešenju. Time što se dopušta računanje i s manjim uzorcima, dolazi se do smanjenja troškova.

Suština postupka povećavanja uzorka jeste da se rešava niz aproksimacija s određenom tačnošću i da se naredni problem rešava s uzorkom čija je veličina veća u odnosu na prethodni. Za početnu tačku novog problema uzima se aproksimacija iz poslednje iteracije prethodnog problema. Algoritmi inkrementalnog gradijenta nude jeftine aproksimacije uzorkovanjem podskupa izraza koji se nalaze pod sumom. Ove metode inicijalno mogu postići veliki napredak, ali često sporo stižu do rešenja. Nasuprot tome, metode punog gradijenta postižu stabilnu konvergenciju ali je svaka iteracija skupa s obzirom da se računa gradijent od celog skupa podataka.

Jasno je da ako je broj merenja N iz formule (2.9) jako velik ili ako su pojedinačne f_i složene funkcije, tada izračunavanje $f(x)$ i $\nabla f(x)$ može biti skupo. S obzirom da se u skupu podataka mogu ponavljati ista ili slična merenja, kompletna evaluacija funkcije $f(x)$ ili $\nabla f(x)$ može biti nepotrebna da bi se napredovalo u rešavanju problema minimizacije funkcije date formulom (2.9). Upravo to daje motivaciju za uvođenje metode inkrementalnog gradijenta, u kojoj svaka iteracija procenjuje gradijent u odnosu na sve dostupne podatke do datog momenta.

Ukoliko bi se u svakom koraku koristila samo jedna instanca, metoda inkrementalnog gradijenta bi bila N puta brža od metode punog gradijenta, jer su iteracije nezavisne od N . Dakle, u vremenu koje je potrebno da se napravi jedna iteracija punog gradijenta, metoda inkrementalnog gradijenta može postići N iteracija, što često rezultira brzim napretkom u početku. Međutim, broj iteracija potrebnih za postizanje istog nivoa tačnosti može biti mnogo veći.[12] Motivisan hibridnim varijantama algoritma kao iz rada [12], naš inkrementalni metod će biti u smislu rastućeg uzorka. S obzirom da stohastički gradijent ima samo stohastičku konvergenciju, a klasičan gradijent je skup, uzima se nešto između, a to je baš rastući uzorak. U nastavku je predstavljen metod rastućeg stohastičkog gradijenta.

Još jedna od modifikacija gradijentnog metoda je ***gradijentni metod sa rastućim uzorkom***. U poređenju sa prethodnom metodom, ovaj metod je kompleksniji iz razloga što pri računanju svake nove iteracije koristi sve prethodne f_i funkcije, tj. uzima sve podatke dostupne do datog momenta. Za početak, predstavlja se formula za izračunavanje koeficijenata $k+1$ iteracije

$$x_{k+1} = x_k - \tilde{\alpha}_k \nabla f_{N_k}(x_k) \quad (2.11)$$

gde je $N_k < N_{k+1}$, tj. N je rastuće. Kod metoda sa rastućim uzorkom funkcija $f_{N_k}(x_k)$ se dobija kao prosečna vrednost svih prethodnih f_i - ova.

$$f_{N_k}(x) = \frac{1}{N_k} \sum_{i=0}^{N_k} f_i(x) \quad (2.12)$$

za $k = 0, 1, 2, \dots$. Osnovna razlika je u tome što stohastički gradijent u svakoj iteraciji uzima samo deo uzorka formiran na osnovu novopridošlih podataka, dok rastući, inkrementalni gradijent zahteva računanje svih gradijenata u novoj tački u svakoj iteraciji i formira po pravilu kvalitetniju, ali i skuplju aproksimaciju rešenja.

Što se tiče variranja veličine uzorka, ono se javlja u više vidova ali ćemo se mi u radu fokusirati na kumulativni uzorak. To znači da svaki put dodajemo nove podatke na postojeći uzorak, ažuriramo ga i samim tim i povećavamo. Takođe, izbor uzorka može biti i takav da su oni međusobno nezavisni jedan od drugog, tj. da imaju prazan presek. [13]

3 Istraživanje

U ovom poglavlju će biti prikazani i analizirani rezultati istraživanja. Pored mnogih programskih paketa, za simulaciju i proveru AR modela korišćen je MATLAB koji zahvaljujući nizu ugrađenih funkcija dosta olakšava rad.

MATLAB je softverski paket namenjen za rešavanje matematičkih problema, analizu i grafički prikaz podataka. On u sebi integriše numeričku i matricnu analizu, dok sve probleme rešava numerički. Sam naziv potiče od engleskih reči *MATrix LABoratory* ("laboratorija za matrice"), koje označavaju njegova osnovna svojstva, a to su rad sa matricama i primena kod obrade mernih rezultata u laboratoriji. Pored matematičkih, koristi se i za rešavanje mnogih inženjerskih i istraživačkih problema.

Osnovu za analizu predstavljale su simulirane serije sa različitim vrednostima disperzije - δ^2 , i reda serije - p . Prilikom formiranja predikcija, svaka od serija podeljena je u odnosu 80% prema 20%, gde je 80% podataka svake serije korišćeno za predviđanje preostalih 20% vrednosti. U nastavku će biti opisano koliko dobro je fitovan model, tj. kakve su predviđene vrednosti u odnosu na stvarne. Za početak, simulirane serije su formirane tako da prate AR model reda p , kada p uzima vrednosti 3, 5 i 8 respektivno, za različite vrednosti δ^2 . Cilj je da se uporede stvarne i predviđene vrednosti, kada su koeficijenti modela dobijeni metodom stohastičkog gradijenta i rastućeg stohastičkog gradijenta. Videće se koji od ova dva gradijentna metoda daje bolje ocene parametara modela, za dobijanje što validnijih predikcija. Kriterijumi koji će se koristiti za ocenu greške su:

- *Standardna devijacija greške predviđanja (1.7)*
- *Prosečna apsolutna greška (1.8)*
- *Prosečna relativna greška (1.9)*

Pored navedenih, posmatraće se i procenat relativne greške, gde kriterijume predstavljaju proizvoljno definisane vrednosti. Da bi se videlo koliko su precizno predikcije ispratile stvarne vrednosti, analiziraće se koliki deo test perioda ima relativnu grešku manju od 1, 3 i 5%.

Prilikom formiranja modela koristiće se metod *in sample - out of sample*, odnosno podaci se dele tako što se deo njih koristi za ocenu koeficijenata pojedinačnih modela, a zatim se pomoću ovih koeficijenata formiraju predikcije i vrši evaluacija dobijenih rezultata preostalim delom podataka. *In sample* metoda predviđanja ima dve prednosti: uvek imate potrebne podatke, i možete proveriti tačnost svoje prognoze upoređujući je sa onim što se zapravo dogodilo.

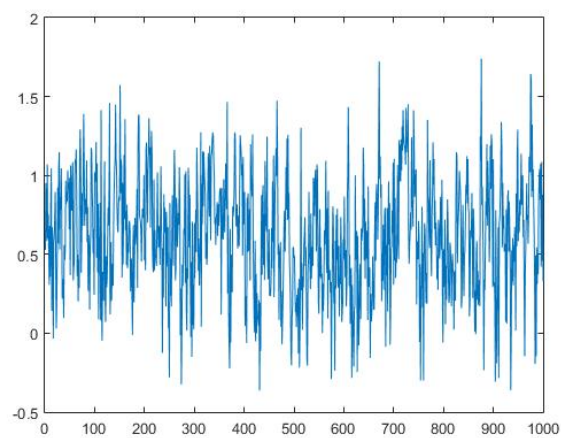
3.1 Grafički prikaz i osnovna analiza

Kao što je već ranije pomenuto, rad se bazira na simuliranim serijama. Analiza će se za početak vršiti na tri simulirane serije koje su napravljene tako da prate $AR(3)$, $AR(5)$ i $AR(8)$ model sa varijansama $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$ i $\sigma_a^2 = 10$, respektivno. To je ono što će se ispitati za početak, dok će se posle ispitati i kako različite kombinacije reda p i varijanse utiču na kretanje simuliranih serija, performanse metoda i njihovih predikcija. Na grafiku je prikazano 1000 simuliranih vrednosti za $AR(3)$, $AR(5)$ i $AR(8)$ procese, čiji su parametri random izabrane vrednosti između 0 i 1 predstavljeni u tabeli ispod:

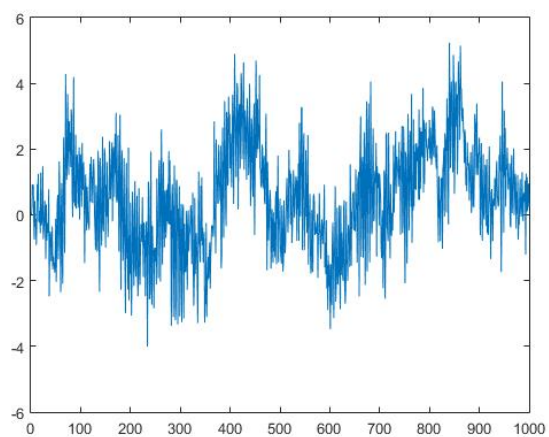
| Model | ϕ_0 | ϕ_1 | ϕ_2 | ϕ_3 | ϕ_4 | ϕ_5 | ϕ_6 | ϕ_7 | ϕ_8 |
|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| AR(3) | 0.215 | 0.342 | 0.283 | 0.008 | - | - | - | - | - |
| AR(5) | 0.026 | 0.109 | 0.002 | 0.007 | 0.617 | 0.182 | - | - | - |
| AR(8) | 0.435 | 0.091 | 0.032 | 0.127 | 0.098 | 0.055 | 0.213 | 0.028 | 0.075 |

Za procese belog šuma $\{a_t\}$ su uzete nezavisne slučajne promenljive sa normalnom raspodelom $\mathcal{N}(\theta, 0.1)$, $\mathcal{N}(\theta, 1)$ i $\mathcal{N}(\theta, 10)$. Slike 3.1, 3.2 i 3.3 daju vizuelnu predstavu kretanja ove tri vremenske serije. Modeli koji se koriste u radu baziraju se na pretpostavkama o stacionarnosti vremenskih serija. Zbog toga se pre same ocene modela mora ispitati stacionarnost. U praksi, slaba stacionarnost je neophodna da bi se mogle odrediti predikcije.

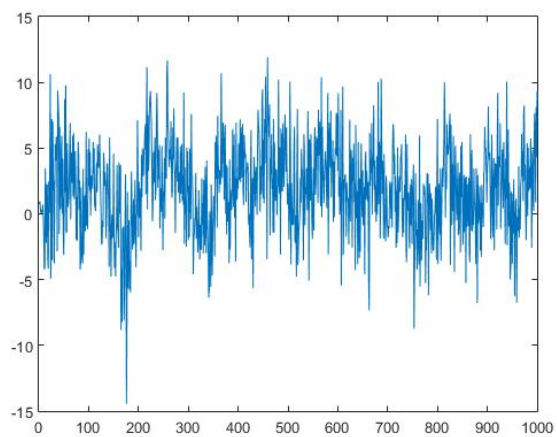
Posmatrajući grafički prikaz na Slikama 3.1, 3.2 i 3.3 uočava se da procesi imaju testerast izgled i da vrednosti osciluju bez opadajućih ili rastućih trendova. To ukazuje na stacionarnost, što se takođe može potvrditi Dickey-Fuller-ovim (DF) testom jediničnih korena, tj. testom stacionarnosti. Dobijena p-vrednost korišćenjem ovog testa za sve tri serije je 0.001 što je manje od 0.05 što znači da se na nivou značajnosti od 5% nulta hipoteza odbacuje pa se zaključuje da su vremenske serije stacionarne. S obzirom da su serije stacionarne, spremne su za dalji korak a to je ocena parametara, a zatim i formiranje predikcija.



Slika 3.1: Kretanje simulirane serije koja prati AR(3) model, gde je $\sigma_a^2 = 0.1$



Slika 3.2: Kretanje simulirane serije koja prati AR(5) model, gde je $\sigma_a^2 = 1$



Slika 3.3: Kretanje simulirane serije koja prati AR(8) model, gde je $\sigma_a^2 = 10$

3.2 Osnovne statističke osobine

U analizi bilo koje vrste podataka početni korak podrazumeva ispitivanje statističkih osobina vremenske serije. Time se bavi deskriptivna statistika, koja sadrži metode i procedure za prezentovanje i sumiranje podataka. Njena svrha je da pomoću dobijenih vrednosti opiše značenje podataka koji se nalaze unutar vremenske serije. Deskriptivna statistika je obično prvi korak u analizi i prethodi statističkom zaključivanju i predviđanju. Osnovne mere koje se koriste za opisivanje podataka su očekivanje, standardna devijacija, minimalna i maksimalna vrednost kao i koeficijenti spljoštenosti i asimetrije. U Tabeli 3.1 su prikazane vrednosti deskriptivnih statistika samo nekih simuliranih vremenskih serija, tj. onih koje prate AR(3) model sa varijansom $\sigma_a^2 = 0.1$, AR(5) model gde je $\sigma_a^2 = 1$ i AR(8) model sa varijansom $\sigma_a^2 = 10$. Za ostale kombinacije varijansi i reda modela, poređenje će se vršiti samo na osnovu grešaka predviđanja.

| | AR(3) | AR(5) | AR(8) |
|------------------------|-----------|-----------|------------|
| Očekivana vrednost | 0.617176 | 0.504789 | 1.858706 |
| Standardna devijacija | 0.371562 | 1.609090 | 3.529850 |
| Min | -0.362892 | -4.005162 | -14.445572 |
| Max | 1.740125 | 5.236287 | 11.911196 |
| Koeficijent asimetrije | -0.027346 | 0.073963 | -0.105924 |
| Koeficijent ekscesa | 2.734708 | 2.777071 | 3.230332 |
| Jarque-Bera test | 3.057126 | 2.982490 | 4.080547 |
| p vrednost | 0.209792 | 0.216804 | 0.116779 |

Tabela 3.1: Deskriptivna statistika simuliranih serija: AR(3), $\sigma_a^2 = 0.1$; AR(5), $\sigma_a^2 = 1$; AR(8), $\sigma_a^2 = 10$

Očekivane vrednosti simuliranih vremenskih serija koje prate AR(3) i AR(5) model su oko 0.5 što je i za očekivati s obzirom na vrednosti sa grafika. Za AR(8) seriju ona je najveća i približno 2, dok je za AR(5) najmanja.

Standardna devijacija nam govori koliko u proseku elementi skupa odstupaju od očekivane vrednosti tog istog skupa. Standardne devijacije su za prvu seriju bliže nuli, za drugu preko 1.5 dok je za treću veća vrednost. Najveće odstupanje od očekivane vrednosti ima serija koja prati AR(8) model.

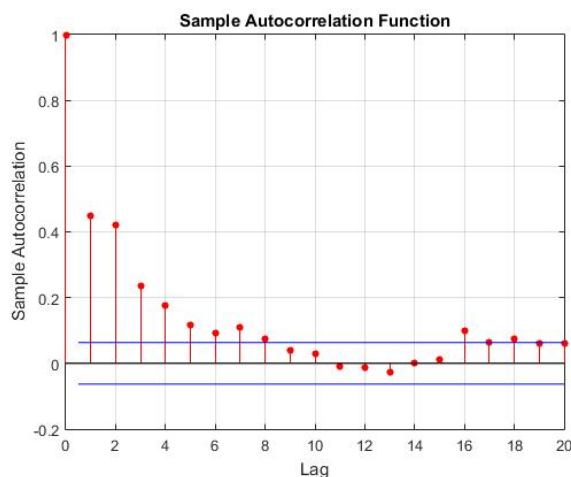
Koeficijent asimetričnosti nam daje informaciju da li je većina vrednosti u uzorku manja ili veća od očekivane vrednosti. Ukoliko je koeficijent asimetričnosti manji od nule to znači da je većina vrednosti u uzorku manja od očekivane vrednosti, i obrnuto. Ako je koeficijent veći od nule to znači da je većina vrednosti u uzorku veća od očekivane vrednosti. U datoj tabeli su za serije koje prate AR(3) i AR(8) model vrednosti koeficijenta asimetrije manje od nule, što znači da je većina podataka iz serija manja od njihovih očekivanih vrednosti. Za seriju koja prati AR(5) model je obrnuta situacija. Koeficijent je veći od nule što znači da je većina podataka iz serije veća od njene očekivane vrednosti.

Koeficijent ekscesa tj. spljoštenosti nam govori koliko je neka raspodela spljoštena u odnosu na normalnu raspodelu. Ako je koeficijent spljoštenosti veći od tri to znači da raspodela ima deblje repove što ukazuje na postojanje aut-lajera, ako je jednak tri to znači da je raspodela normalno spljoštena i ako je manji od tri raspodela ima tanje repove. U datoj tabeli vrednosti koeficijenata

spljoštenosti za AR(3) i AR(5) serije su ispod tri, dok je za seriju koja prati AR(8) model ta vrednost malo iznad 3. Ovo ukazuje na postojanje outlier-a, što znači da postoje vrednosti koje su udaljene od ostalih.

Jarque-Bera (JB) test normalnosti testira da li koeficijenti asimetrije i spljoštenosti odgovaraju normalnoj raspodeli. Nulta hipoteza koja se testira glasi da podaci imaju koeficijent asimetričnosti jednak nuli, a koeficijent spljoštenosti jednak tri, odnosno da ovi koeficijenti odgovaraju normalnoj raspodeli. Na osnovu registrovanih p vrednosti u Tabeli 3.1 zaključuje se da se za sve serije nulte hipoteze prihvataju, tj serije prate normalnu raspodelu. Nulta hipoteza JB testa koja se testira glasi, da podaci prate standardnu normalnu raspodelu, nasuprot alternativne da podaci nisu normalno raspodeljeni. Ako su p vrednosti veće od 0.05 nulta hipoteza se prihvata na nivou poverenja od 95%.

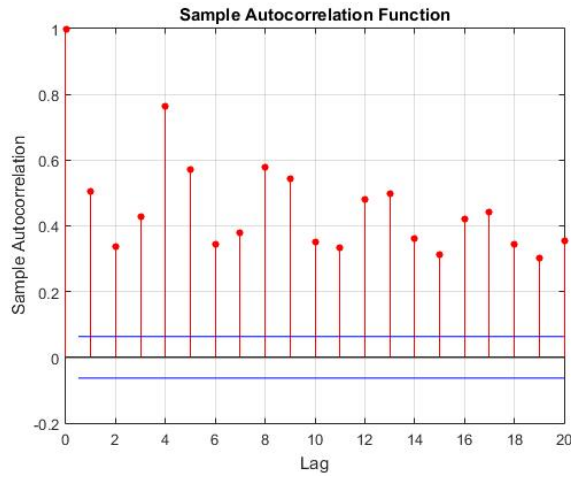
Ono što je još interesantno da se posmatra jeste autokorelaciona povezanost vrednosti unutar serija koja se može očitati sa sledećih grafika:



Slika 3.4: Autokorelaciona funkcija AR(3) serije

Autokorelacija je matematička reprezentacija stepena sličnosti između date vremenske serije i njenih prethodnih vrednosti u više vremenskih intervala. Prilikom izračunavanja autokorelacije mogu se dobiti vrednosti u opsegu od -1 do 1. Autokorelacija koja ima vrednost 1 predstavlja savršenu pozitivnu korelaciju, dok vrednost -1 ukazuje na savršeno negativnu korelaciju. Vidimo da vrednosti koeficijenata kod serije koja prati AR(3) model opadaju što se više ide u prošlost. Nakon četvrtog koraka, korelaciona zavisnost opada i koeficijenti nisu statistički značajni. Kod serija koje prate AR(5) i AR(8) model, situacija je drugačija. Tu imamo vrednosti koje ukazuju na značajniju korelacionu povezanost. Vrednovanje korelacije se može odrediti korišćenjem Ljung - Box testa kod koga se za nultu hipotezu koristi pretpostavka da nema autokorelacije, naspram alternativne da ona postoji. Vrednost test statistike dobija se ranije predstavljenom formulom 1.3.

U Tabeli ispod, date su autokorelacione funkcije za serije koje prate AR(3), AR(5) i AR(8) modele, redom. One samo potvrđuju stanje sa grafikama.



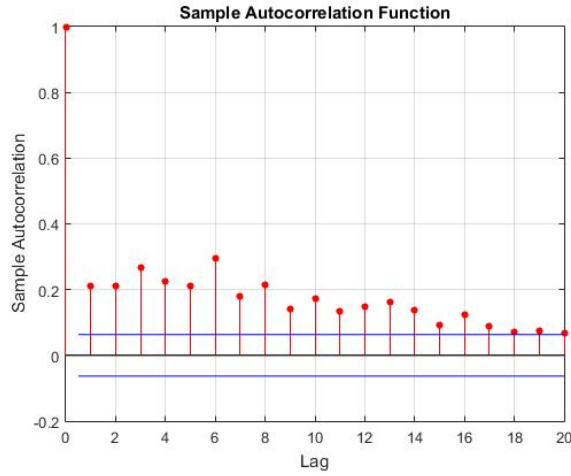
Slika 3.5: Autokorelaciona funkcija AR(5) serije

| AR(3) | ACF | t | hip |
|--------|-------|--------|-----|
| Lag 1 | 0.449 | 8.207 | 1 |
| Lag 4 | 0.177 | 2.852 | 1 |
| Lag 5 | 0.118 | 1.879 | 0 |
| Lag 10 | 0.030 | 0.4751 | 0 |

| AR(5) | ACF | t | hip |
|--------|-------|--------|-----|
| Lag 1 | 0.506 | 9.239 | 1 |
| Lag 4 | 0.763 | 11.910 | 1 |
| Lag 5 | 0.569 | 7.8475 | 1 |
| Lag 10 | 0.349 | 3.979 | 1 |

| AR(8) | ACF | t | hip |
|--------|-------|-------|-----|
| Lag 1 | 0.212 | 3.877 | 1 |
| Lag 4 | 0.225 | 3.902 | 1 |
| Lag 5 | 0.212 | 3.634 | 1 |
| Lag 10 | 0.171 | 2.743 | 1 |

U prvoj koloni je predstavljeno koliko koraka se ide unazad, tj. koliko daleko se gleda u prošlost. Druga kolona daje vrednosti za ACF na osnovu formule (1.1), tj. ρ_l za svako l , treća kolona predstavlja vrednost test statistike t -ratio koji smo računali na osnovu formule (1.2), a poslednja kolona govori da ako se hipoteza H_0 odbacuje onda pišemo jedan, u suprotnom ako se H_0 prihvata onda pišemo nula. U datim matricama kao i na graficima, uočava se da što se više ide u prošlost vrednosti autokorelacije se smanjuju. Pa se tako za prvu, AR(3) seriju, za više od četiri koraka unazad korelacija gubi, dok za preostale dve serije ona postoji sve do 20 koraka unazad iako njene vrednosti nisu visoke.



Slika 3.6: Autokorelaciona funkcija AR(8) serije

3.3 Ocena parametara modela metodom stohastičkog gradijenta (SGD) i gradijenta sa rastućim uzorkom (IGD)

Za ocenu parametara modela koriste se dve konceptualno različite, prethodno spomenute varijante gradijentnog metoda:

1. *Stohastički gradijentni metod* i 2. *Gradijentni metod sa rastućim uzorkom*. S obzirom da se ova dva metoda porede u slučaju kad imamo rekalkulaciju, parametri modela se iznova ocenjuju sa dotokom novih podataka kako bi se dobile što bolje predikcije. Za ocenjivanje parametara modela neće se koristiti cela vremenska serija nego samo prvih 80% podataka, što je u ovom slučaju prvih 800 vrednosti vremenske serije. Početno x_T , gde T predstavlja veličinu uzorka, dobija se primenom metode najmanjih kvadrata na prethodnih 80% podataka. Na osnovu njega, dobija se naredni vektor koeficijenata, x_{T+1} od kog kreću dalje iteracije. Dakle, početna tačka za računanje novih koeficijenata je x_T koje će biti isto za pojedinačne vremenske serije u oba metoda. Zatim će se sa dobijenim koeficijentima praviti predikcije, koje se kasnije porede sa stvarnim vrednostima.

AR(3)

Krenućemo sa vremenskom serijom koja prati $AR(3)$ model. Za početak će se oceniti parametri modela uz pomoć klasičnog pristupa, odnosno koristiće se metod najmanjih kvadrata. Dakle, ocenjeni su parametri AR(3) modela koji je dat formulom:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + a_t$$

U Tabeli 3.2 prikazani su koeficijenti ocenjeni metodom najmanjih kvadrata i njihove standardne greške, što je uobičajena tehnika kod klasičnog pristupa ocene parametara.

Posmatrajući poslednju kolonu ove tabele uočava se da su svi koeficijenti

| Parametar | Ocena | Standardna greška | Test statistika |
|-----------|-----------|-------------------|-----------------|
| ϕ_0 | 0.248819 | 0.034937 | 7.121851 |
| ϕ_1 | 0.323329 | 0.035498 | 9.108137 |
| ϕ_2 | 0.291568 | 0.034940 | 8.344735 |
| ϕ_3 | -0.018927 | 0.02590 | 0.730696 |

Tabela 3.2: Ocena parametara AR(3) modela sa normalnom $\mathcal{N}(0,0.1)$ raspodelom

osim ϕ_3 statistički značajni. Prema tome, naš model se može opisati formulom:

$$r_t = 0.248819 + 0.323329r_{t-1} + 0.291568r_{t-2} + a_t$$

Sada, kada je poznata početna tačka prelazi se na računanje novih koeficijenata, za početak metodom stohastičkog gradijenta.

1. Stohastički gradijentni metod

Da bi se dobili koeficijenti AR(3) modela SGD metodom, tj. metodom *Stohastičkog gradijenta* primenjuje se njegov opšti oblik dat formulom (2.10)

$$x_{T+1} = x_T - \alpha_T \nabla f_i(x_T)$$

Funkcija f_i u ovom metodu predstavlja na neki način novi podatak. Kao što je navedeno u poglavlju 2.3.1 to je kvadratna funkcija oblika

$$f_i(x) = \frac{1}{2}(x^T a_i - r_i)^2$$

gde je x vektor koeficijenata modela oblika $(\phi_0, \phi_1, \dots, \phi_p)^T$ kojih ima onoliko koliki je red modela (u ovom slučaju tri, plus slobodan član ϕ_0), a $a_i = (1, r_{i-1}, \dots, r_{i-p})^T$ i r_i su vrednosti očitane iz vremenske serije. Suština je u tome da kad imamo x_T (gde je T u našem slučaju prvih 800 podataka) napravimo predikciju, a zatim na kraju dana saznamo novi podatak koji postaje novi član uzorka. Na taj način se iz iteracije u iteraciju f_i ažurira, pa samim tim i koeficijenti metoda zajedno sa njom.

S obzirom da je x_{800} trenutak od kog se kreće sa računanjem, novi koeficijent u trenutku $T+1$ kada T uzima vrednost, $T = 800$ biće

$$x_{801} = x_{800} - \alpha_1 \nabla f_{801}(x_{800})$$

gde je x vektor koeficijenata modela oblika $(\phi_0, \phi_1, \phi_2, \phi_3)^T$. Koeficijenti metoda stohastičkog gradijenta α_k se kao što smo ranije naveli, u našem slučaju računaju po formuli $\alpha_k = \frac{1}{k}$. Tako sve do $T+k$ momenta, kada k uzima vrednosti $k = 1, \dots, 200$. Najbitnije je da se vrednosti koeficijenata ažuriraju iz iteracije u iteraciju, pa se tako na kraju stiže i do poslednje iteracije za x_{1000}

$$x_{1000} = x_{999} - \alpha_{200} \nabla f_{1000}(x_{999})$$

Vrednosti prvih nekoliko vektora koeficijenata, dobijenih opisanim iteracijama su:

$$\begin{aligned}
x_{801} &= [0.141931 \quad 0.266438 \quad 0.297845 \quad -0.040401], \\
x_{802} &= [0.139158 \quad 0.265625 \quad 0.296369 \quad -0.040237], \\
&\vdots \\
x_{1000} &= [0.283727 \quad 0.292395 \quad 0.284343 \quad -0.012296]
\end{aligned}$$

Kasnije će biti i grafički predstavljene vrednosti koeficijenata da bi se videlo njihovo kretanje tokom vremena, kao i da bi se uporedilo sa vrednostima dobijenim primenom drugog metoda čiji je postupak opisan u nastavku.

2. Gradijentni metod sa rastućim uzorkom

Primenom drugog metoda za ocenu parametara, datog formulom (2.11) u poglavlju 2.3.2,

$$x_{T+1} = x_T - \tilde{\alpha}_T \nabla f_{N_T}(x_T),$$

vrednost koeficijenata x_{T+1} u trenutku $T = 800$ ima malo drugačiji oblik u odnosu na prvi metod ocene parametara

$$x_{801} = x_{800} - \tilde{\alpha}_{800} \nabla f_{N_{801}}(x_{800})$$

Osnovna razlika je u načinu računanja dužine koraka α i gradijenta funkcije $f_i(x)$. Prvi metod je nešto jednostavniji, jer za dobijanje dužine koraka α uzima samo vrednost razlomka $\alpha_k = \frac{1}{k}$ i gradijent funkcije se računa samo u jednoj tački. Dok je drugi metod zahtevniji, jer je potrebno da se računaju svi gradijenti u novoj tački u svakoj iteraciji. Po uzoru na formulu (2.12) iz dela o inkrementalnom metodu, sada samo uvodimo parametar T i vodimo računa da nam brojač krene od $i = 2$ zbog AR modela

$$f_{N_{T+k}}(x) = \frac{1}{T+k} \sum_{i=2}^{T+k} f_i(x),$$

a pojedinačne funkcije f_i su istog oblika kao u prethodnom metodu, tj. $f_i(x) = \frac{1}{2}(x^T a_i - r_i)^2$. Za dobijanje dužine koraka $\tilde{\alpha}$ korišćena je sledeća formula

$$\tilde{\alpha}_k = \frac{1}{2L_k},$$

gde je

$$L_k = \frac{1}{N_k} \sum_{i=1}^{N_k} l_i,$$

a l_i maksimalan karakteristični koren hesijana funkcije f_i , tj. $l_i = \lambda_{max}(\nabla^2 f_i(x))$. Dakle, za dobijanje dužine koraka α potrebno je izračunati Hesijan funkcije $f_i(x)$ a za gradijent koristiti sve prethodne vrednosti niza, pomnožene vrednošću novodobijenog koeficijenta.

Po uslovu iz Teoreme 11 za fiksnu dužinu koraka α , važi da je $\alpha < \frac{1}{L}$

pri čemu je L Lipšicova konstanta za gradijent. U našem slučaju, mi svaki dan imamo neki novi uzorak koji u tom trenutku predstavlja sve podatke kojima raspolazemo. To znači da svaki dan imamo novi pun gradijent. Jedan od uslova za konvergenciju je da je dužina koraka manja od $\frac{1}{L}$, pri čemu je L kao što smo već rekli - Lipšicova konstanta za gradijent. S obzirom da mi svaki dan imamo novi pun gradijent, onda se osim njega ažurira i dužina koraka $\tilde{\alpha}$. Danas nam je dužina koraka $\tilde{\alpha}_k = \frac{1}{2L_k}$ pri čemu L_k predstavlja Lipšicovu konstantu za ceo uzorak koji imamo. Sutradan dobijamo novi podatak što znači da se cela funkcija menja, a samim tim menja se i gradijent. Takođe ažuriramo i dužinu koraka tj. uzimamo novu ocenu za Lipšicovu konstantu. Motiv za izbor baš ovakve dužine koraka je upravo iz teoreme o fiksnom α , gde je uslov da je ono manje od $\frac{1}{L}$ a naš izbor $\tilde{\alpha}_k = \frac{1}{2L_k}$ ga zadovoljava.

Ostalo je još samo da napomenemo kako računamo Lipšicovu konstantu za gradijent. Već smo rekli da je ona maksimalan karakteristični koren hesijana funkcije f_i . S obzirom da je naša funkcija f_i kvadratna i "lepog" oblika, znamo da izračunamo maksimalan karakteristični koren hesijana. Za dobijanje $\nabla^2 f_i(x)$ koristimo izraz $\nabla^2 f_i(x) = a_i a_i^T$ gde su $a_i = (1, r_{t-1}, \dots, r_{t-p})^T$ podaci očitani iz naše vremenske serije.

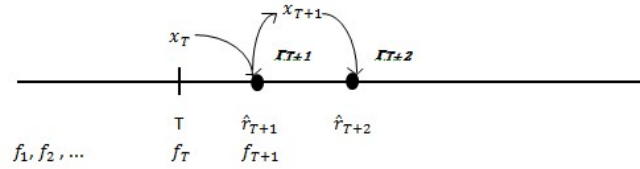
Radi poređenja sa metodom stohastičkog gradijenta, navešćemo nekoliko prvih iteracija koeficijenata

$$\begin{aligned} x_{801} &= [0.248793 \quad 0.323315 \quad 0.291569 \quad -0.018932], \\ x_{802} &= [0.248743 \quad 0.323297 \quad 0.291559 \quad -0.018930], \\ &\vdots \\ x_{1000} &= [0.254574 \quad 0.335715 \quad 0.278388 \quad -0.026889] \end{aligned}$$

da bi bilo lakše uočiti razlike u vrednostima.

Za kraj primene ova dva metoda na seriju koja prati $AR(3)$ model, još jednom ćemo se osvrnuti na to šta tačno predstavlja funkcija f_i . Kao što je prethodno navedeno, suština postupka povećavanja uzorka je da se rešava niz aproksimacija s određenom tačnošću i da se svaki naredni problem rešava s uzorkom čija je veličina veća u odnosu na prethodni. Upravo tako je i formulisana funkcija koja se dobija kao prosečna vrednost svih prethodnih f_i - ova. Sam f_i računa se na isti način kao kod stohastičkog gradijenta, dakle imamo x_T pomoću kog se prave predikcije, na kraju dana dobijemo novi podatak čijim dodavanjem se formira novi uzorak. Osnovna razlika je što inkrementalni gradijent uzima sve raspoložive podatke do i -tog momenta, dok stohastički gradijent uzima samo deo uzorka, odnosno računa gradijent samo u jednoj tački.

Grafički, algoritmi za dobijanje novih koeficijenata i način na koji se računaju predikcije bi se najlakše predstavili na sledeći način, Slika 3.7, gde se lako vidi da koeficijenti modela iz vektora x_T dobijeni metodom najmanjih kvadrata služe za dobijanje predikcija \hat{r}_{T+1} u narednom koraku, čija se vrednost na kraju poredi sa stvarnom r_{T+1} . Postupak se dalje ponavlja onoliko puta koliko koraka se ide unapred. U našem slučaju to je za $k = 1, 2, \dots, 200$.



Slika 3.7: Skica algoritma za dobijanje predikcija pomoću novih koeficijenata

Nakon što se ocene parametri modela, prelazi se na formiranje predikcija.

3.4 Formiranje prediktivnih vrednosti $AR(3)$ modela

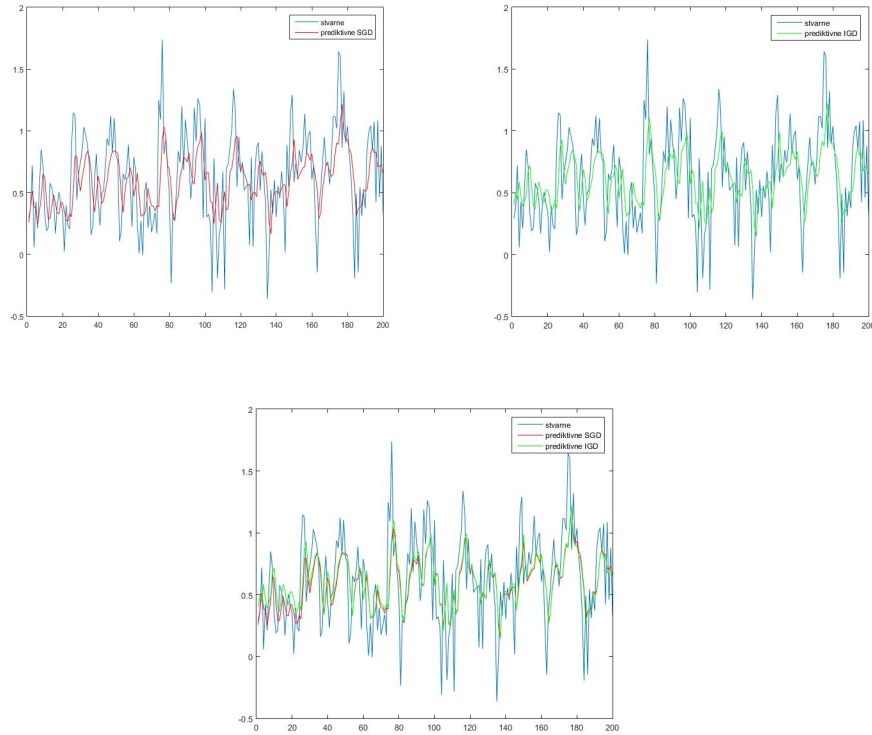
Kada se znaju koeficijenti modela, i to kako on izgleda u svakoj iteraciji može se preći na predikcije za narednih k koraka unapred kad k uzima vrednosti $k = 1, \dots, 200$. Kao što je već navedeno u nekom od prethodnih poglavlja, kao pristup se koristi podela vremenske serije na dva perioda. Uglavnom se deli tako da prvi period sadrži 80% a drugi period preostalih 20% podataka. Prvi potperiod serije koristi se za ocenu modela, dok se drugi potperiod naziva test period i za taj period se formiraju predikcije.

Vremenska serija koja prati $AR(3)$ model, deli se na dva potperioda. S obzirom da je veličina serije $T = 1000$, prvi potperiod koji sadrži 80% podataka je ustvari prvih 800 vrednosti date serije. Test period sadrži preostale podatke, tj. poslednjih 200 vrednosti vremenske serije koju posmatramo. Pomoću *SGD* i *IGD* metoda ocenjeni su parametri modela na osnovu podataka iz prvog potperioda vremenske serije. Nakon toga, formirane su predikcije za test period.

Za početak, predstavljene su predikcije vremenske serije kod koje su parametri modela ocenjeni SGD metodom. Na Slici 3.8 grafički su predstavljene stvarne i prediktivne vrednosti serije dobijene statičkom metodom. Na prvom grafiku, plavom bojom predstavljene su stvarne vrednosti serije, dok su crvenom bojom označene prediktivne vrednosti $AR(3)$ modela dobijene pomoću koeficijenata stohastičkog gradijenta. Sa grafika se vidi da prediktivne vrednosti prate kretanje serije ali da odstupanje od stvarnih svakako postoji.

Drugi metod ocene parametara modela - *IGD*, dao je slične rezultate. Na grafiku desno, predstavljene su stvarne i prediktivne vrednosti serije dobijene opet statičkom metodom. Plavom bojom su isto kao u prethodnom primeru, predstavljene stvarne vrednosti serije, dok su zelenom bojom označene prediktivne vrednosti $AR(3)$ modela dobijene pomoću koeficijenata rastućeg stohastičkog gradijenta. Na poslednjem grafiku Slike 3.8, predstavljene su predikcije dobijene pomoću oba metoda za ocenu parametara zajedno sa stvarnim vrednostima serije.

Na sva tri grafika, x-osa je vremenska osa i na njoj su predstavljeni momenti koji se mogu posmatrati kao sati ili dani za koje su formirane predikcije a y-osa predstavlja vrednosti vremenske serije. Uočava se da predviđene vrednosti ne osciluju mnogo i približno su jednake za oba metoda. One se za prvi metod kreću u intervalu od 0.1634 do 1.2162, dok se za metod rastućeg stohastičkog



Slika 3.8: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(3) model sa varijansom 0.1

gradijenta (IGD) kreću od 0.1433 do 1.2213.

Da bi se izveli neki precizniji zaključci o validnosti predikcija nije dovoljan samo grafički prikaz, nego je potrebno da se izračunaju i greške predviđanja. U Tabeli 3.3 nalaze se greške predviđanja za oba kriterijuma. Standardna devijacija greške predviđanja (RMSE) date formulom (1.7) i prosečna suma apsolutne greške (MAE) date formulom (1.8) trebaju da budu blisko nuli, što je ovde i slučaj za oba metoda. Prosečno apsolutno procentualno odstupanje (MAPE) od stvarnih vrednosti serije, računa se pomoću formule (1.9) i za seriju koja prati AR(3) model iznosi 73.37% za prvi metod, odnosno 80.51% za drugi metod. To je pokazatelj da iako predikcije nisu baš dobre, bolje su prediktivne vrednosti izračunate pomoću koeficijenata dobijenih metodom stohastičkog gradijenta.

| | SG | IG |
|------|----------|----------|
| RMSE | 0.317379 | 0.324089 |
| MAE | 0.248022 | 0.255593 |
| MAPE | 73.37077 | 80.50851 |

Tabela 3.3: Greške predviđanja AR(3) modela sa varijansom $\sigma_a^2 = 0.1$

Kako bi se bolje sagledale razlike između stvarnih i prediktivnih vrednosti u Tabeli 3.4 predstavljen je procenat predikcija sa apsolutnom relativnom greškom manjom od 1%, 3% i 5%.

| Relativna greška | SG | IG |
|------------------|-------|----|
| < 1% | 1.5% | 1% |
| < 3% | 7% | 6% |
| < 5% | 11.5% | 9% |

Tabela 3.4: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 3$ sa varijansom $\sigma_a^2 = 0.1$

Iz tabele se vidi da samo 7% predikcija iz prvog modela, odnosno 6% predikcija iz drugog modela imaju apsolutno relativno odstupanje manje od 3%, dok 11.5% odnosno 9% predikcija ima grešku manju od 5% što potvrđuje prethodnu analizu, da su predikcije sa koeficijentima SGD metoda bolje ispratile stvarne vrednosti iako su ovi procenti dosta niski.

AR(5)

Sledeća serija je ona koja prati $AR(5)$ model, sa varijansom $\sigma_a^2 = 1$. Šest parametara modela koji predstavljaju početnu tačku za oba metoda ocene su

$$x_{800} = \begin{bmatrix} 0.032253 \\ 0.094698 \\ -0.021318 \\ 0.026679 \\ 0.634975 \\ 0.177822 \end{bmatrix}.$$

Dobijeni su kao i u prethodnoj seriji, metodom najmanjih kvadrata i to su ponovo zajednički koeficijenti za oba metoda ocene parametara. U Tabeli 3.5 prikazane su pored koeficijenata njihove standardne greške, što je uobičajena tehnika kod klasičnog pristupa ocene parametara.

| Parametar | Ocena | Standardna greška | Test statistika |
|-----------|-----------|-------------------|-----------------|
| ϕ_0 | 0.032253 | 0.042860 | 0.752527 |
| ϕ_1 | 0.094698 | 0.033529 | 2.824359 |
| ϕ_2 | -0.021318 | 0.03351 | 0.636150 |
| ϕ_3 | 0.026679 | 0.033524 | 0.795816 |
| ϕ_4 | 0.634975 | 0.042857 | 14.81601 |
| ϕ_5 | 0.177822 | 0.045947 | 3.870164 |

Tabela 3.5: Ocena parametara AR(5) modela sa normalnom $\mathcal{N}(0,1)$ raspodelom

Posmatrajući poslednju kolonu ove tabele uočava se da su koeficijenti ϕ_1 , ϕ_4 i ϕ_5 statistički značajni dok preostala tri nisu. Prema tome, naš model se

može opisati formulom:

$$r_t = 0.094698r_{t-1} + 0.634975r_{t-4} + 0.177822r_{t-5} + a_t$$

Ostali koeficijenti su za oba metoda, i metod *Stohastičkog gradijentnog spusta* i za *Gradijentni metod sa rastućim uzorkom* dobijeni primenom formula (2.10) i (2.11), respektivno. U nastavku će se videti do kakvih predviđenih vrednosti ti koeficijenti vode.

3.5 Formiranje prediktivnih vrednosti $AR(5)$ modela

Kao i u prethodnom primeru i ovde se koristi najjednostavniji način da se predvide buduće vrednosti $AR(5)$ serije, odnosno buduće vrednosti simulirane serije koja prati $AR(5)$ model će se predviđati samo u odnosu na njene prethodne vrednosti. Metodologija formiranja predikcija je ista kao i u prethodnom primeru sa $AR(3)$ serijom. Dakle, serija se ponovo deli na dva potperioda. Prvi potperiod sadrži 80% podataka i služi za ocenu modela, dok drugi potperiod čini preostalih 20% i on predstavlja test period tj. za taj period se formiraju predikcije. Simulirana vremenska serija koja prati $AR(5)$ model je iste veličine kao i $AR(3)$ serija, tj. $T = 1000$. Pomoću *SGD* i *IGD* metoda ocenjeni su parametri modela na osnovu podataka iz prvog potperioda, što znači da se može preći na formiranje predikcija.

Na prvom grafiku Slike 3.9 prikazane su predikcije vremenske serije kada su njeni parametri ocenjeni metodom *Stohastičkog gradijentnog spusta* (*SGD*). Plava boja predstavlja stvarne vrednosti simulirane serije, dok crvena boja predstavlja prediktivne vrednosti modela. Primećuje se da su prediktivne vrednosti dosta visoke i u početku beleže velike skokove. To je do toga što su vrednosti gradijenta velike, a isto tako i korak α je na početku najveći pa vrednosti koeficijenata dosta variraju. Tek nešto kasnije pri većim vrednostima k , α se smanjuje međutim, vrednosti koeficijenata ostaju visoke.

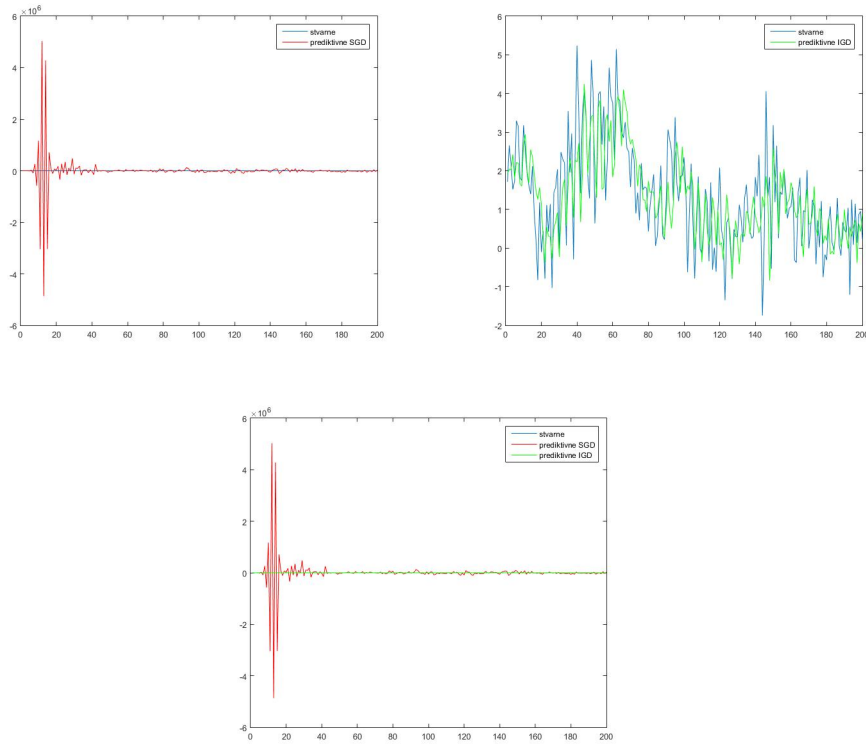
Primenom drugog metoda za ocenu parametara, dobijene su dosta drugačije vrednosti koje vode ka prediktivnim vrednostima, prikazanim na desnom grafiku Slike 3.9. Na tom grafiku se jasno vidi da su predviđene vrednosti bolje ispratile stvarne nego što je to slučaj kod prvog metoda. Takođe, koeficijenti modela dobijeni *IGD* metodom su mnogo manji nego oni ocenjeni prvom metodom. Iz koraka u korak, nema nekih velikih razlika među njima tj. menjaju se dosta sporije. Kako bi predstavili koja je razlika u pitanju, mogu se uporediti vrednosti koeficijenata u istim iteracijama. Tako je recimo za $k = 4$ vrednost koeficijenata dobijenih *SGD* metodom

$$x_{804} = [117.879 \quad 241.765 \quad 336.259 \quad 175.328 \quad 360.442 \quad 230.436]$$

a metodom rastućeg stohastičkog gradijenta, *IGD*

$$x_{804} = [0.032299 \quad 0.094663 \quad -0.021267 \quad 0.026703 \quad 0.634978 \quad 0.177851].$$

Stoga je i jasno iz kog razloga se prediktivne vrednosti toliko razlikuju međusobno. Kako bi se bolje sagledale razlike, opet se oslanjamo na greške predviđanja kako bi se izveli precizniji zaključci. U Tabeli 3.6 vidi se da je prosečno apsolutno procentualno odstupanje (*MAPE*) (1.9) od stvarnih vrednosti manje kod drugog metoda i iznosi oko 55%. Kod predikcija izračunatih pomoću



Slika 3.9: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(5) model sa varijansom 1

koeficijenata dobijenih metodom stohastičkog gradijenta vrednost odstupanja je jako velika iz razloga što su predviđene vrednosti mnogostruko veće od stvarnih. Standardna devijacija greške predviđanja (RMSE) (1.7) i prosečna suma apsolutne greške (MAE) (1.8) trebaju da budu blisko nuli, što je jedino zadovoljeno kod predikcija izračunatih pomoću koeficijenata rastućeg stohastičkog gradijenta.

| | SG | IG |
|------|------------|----------|
| RMSE | 6.6786e+05 | 0.940480 |
| MAE | 1.5730e+05 | 0.712573 |
| MAPE | 1.0484e+07 | 55.17253 |

Tabela 3.6: Greške predviđanja AR(5) modela sa varijansom $\sigma_a^2 = 1$

Dakle, već je jasno da će procenat predikcija čije su relativne greške manje od zadatih granica u Tabeli 3.7 biti nula za predikcije dobijene koeficijentima SGD metoda. Vrednosti za predikcije dobijene koeficijentima IGD metoda su nešto bolje ali takođe niske kao i u slučaju serije koja prati AR(3) model. Bez

| Relativna greška | SG | IG |
|------------------|----|------|
| < 1% | 0% | 2% |
| < 3% | 0% | 3.5% |
| < 5% | 0% | 6.5% |

Tabela 3.7: Procenat predikcija čije su relativne greške manje od zadatih granica

obzira na niske vrednosti iz druge kolone Tabele 3.7, ti rezultati samo potvrđuju prethodnu analizu. Predikcije sa koeficijentima IGD metoda na primeru serije koja prati AR(5) model, su bolje ispratile stvarne vrednosti. To je drugačiji zaključak nego kod AR(3) serije, gde su predikcije sa koeficijentima SGD metoda bolje ispratile stvarne vrednosti.

AR(8)

Nakon analize serija koje prate $AR(3)$ i $AR(5)$ model, preostaje još analiza vremenske serije koja prati $AR(8)$ model, sa varijansom $\sigma_a^2 = 10$. Isto se kreće od klasične ocene parametara metodom najmanjih kvadrata, koji daje vektor koeficijenata

$$x_{800} = \begin{bmatrix} 0.5274 \\ 0.0768 \\ 0.0734 \\ 0.1187 \\ 0.0869 \\ 0.0466 \\ 0.2065 \\ 0.0515 \\ 0.0607 \end{bmatrix}.$$

Dobijeni koeficijenti su zajednički za oba metoda i predstavljaju polaznu tačku za iterativne postupke date formulama (2.10) i (2.11). Pre nego predemo na predikcije, osvrnućemo se na standardne greške i test statistike ovih parametara, kao što je to bio slučaj u prethodne dve serije.

U Tabeli 3.8 na osnovu rezultata test statistike iz poslednje kolone, vidi se da su svi koeficijenti osim ϕ_5 , ϕ_7 i ϕ_8 statistički značajni.

Prema tome, model se može opisati formulom

$$r_t = 0.527 + 0.076r_{t-1} + 0.073r_{t-2} + 0.118r_{t-3} + 0.086r_{t-4} + 0.206r_{t-6} + a_t$$

Koeficijenti u narednim iteracijama dobijaju se primenom već pomenutih formula, tako da se u nastavku bavimo predikcijama ove vremenske serije.

| Parametar | Ocena | Standardna greška | Test statistika |
|-----------|----------|-------------------|-----------------|
| ϕ_0 | 0.527439 | 0.036783 | 14.338902 |
| ϕ_1 | 0.076755 | 0.036837 | 2.083632 |
| ϕ_2 | 0.073427 | 0.036327 | 2.021255 |
| ϕ_3 | 0.118651 | 0.036564 | 3.244947 |
| ϕ_4 | 0.086896 | 0.036547 | 2.377671 |
| ϕ_5 | 0.046648 | 0.036265 | 1.286325 |
| ϕ_6 | 0.206494 | 0.036732 | 5.621602 |
| ϕ_7 | 0.051465 | 0.036653 | 1.404113 |
| ϕ_8 | 0.060742 | 0.160587 | 0.378250 |

Tabela 3.8: Ocena parametara AR(8) modela sa normalnom $\mathcal{N}(0,10)$ raspodelom

3.6 Formiranje prediktivnih vrednosti $AR(8)$ modela

Prateći potpuno iste korake kao sa prethodne dve simulirane serije, ponovo se vremenska serija deli na dva potperioda i određuju predikcije za test period. Na prvom grafiku Slike 3.10 prikazane su predikcije vremenske serije kada su njeni parametri ocenjeni SGD metodom. Ono što se odmah uočava to je da su predviđene vrednosti za test period visoke, s tim da se najveći skokovi dešavaju u prvoj polovini pomenutog perioda. Ako se pogledaju koeficijenti dobijeni prvom metodom, već u trećoj i četvrtoj iteraciji oni imaju jako niske tj. visoke vrednosti. Recimo, za $k = 3$ vektor koeficijenata je

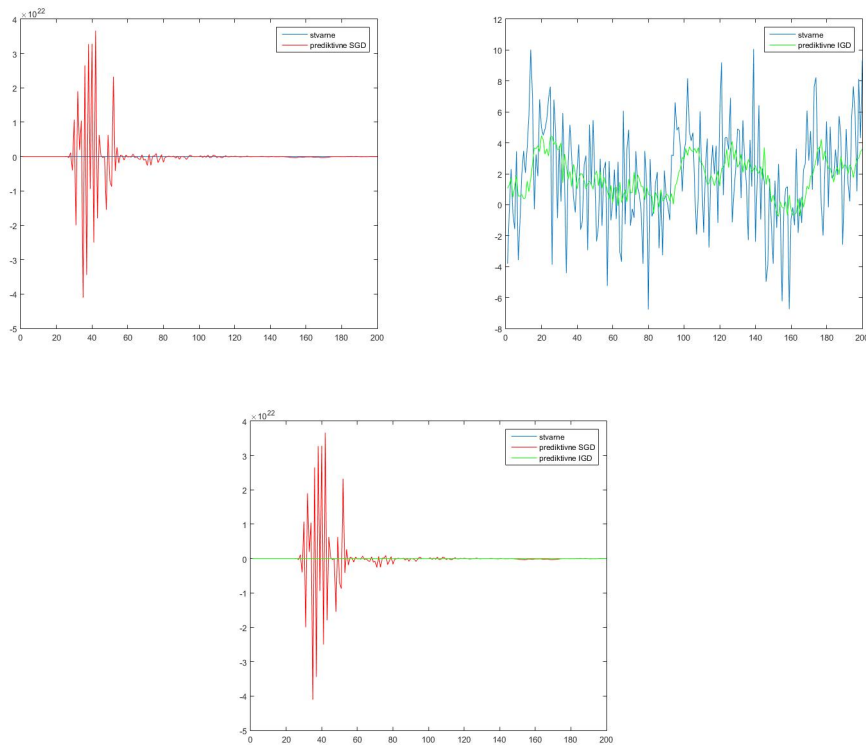
$$x_{803} = \begin{bmatrix} -398.610 \\ 259.6230 \\ 961.9726 \\ -2.6224e + 03 \\ -812.772 \\ 1.1064e + 03 \\ -815.472 \\ -671.782 \\ -47.232 \end{bmatrix}$$

pa je stoga i vrednost predikcije u toj kao i u ostalim tačkama visoka. Međutim, u istoj iteraciji, $k = 3$, vrednost koeficijenata dobijenih IGD metodom je

$$x_{803} = \begin{bmatrix} 0.527361 \\ 0.076342 \\ 0.073133 \\ 0.118881 \\ 0.086832 \\ 0.046494 \\ 0.206476 \\ 0.051484 \\ 0.060880 \end{bmatrix}$$

što je približnije nekim realnim vrednostima, pa se stoga očekuju i validnije predikcije.

Kao što se pokazalo za prethodne dve simulirane serije, tako i u ovoj, drugi metod ocene parametara (IGD) daje predviđene vrednosti bliže očeki-



Slika 3.10: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(8) model sa varijansom 10

vanoj. Deskriptivna statistika prikazana u Tabeli 3.1 pokazuje da je ta vrednost za AR(8) seriju oko 1.85. To se u nekim tačkama sa drugog grafika Slike 3.10 poklapa, međutim postoje poprilična odstupanja.

U Tabeli 3.9 nalaze se greške predviđanja modela. Standardna devijacija greške predviđanja (RMSE) (1.7) i prosečna suma apsolutne greške (MAE) (1.8) trebaju da budu blisko nuli, što ovde i nije slučaj. Vrednosti u prvoj koloni, kod SGD metoda su jako velike dok su kod IGD metoda male, ali opet iznad nule. Prosečno apsolutno procentualno odstupanje (MAPE) (1.9) od stvarnih vrednosti je 44% za predikcije dobijene pomoću koeficijenata izračunatih metodom rastućeg stohastičkog gradijenta. Za predikcije dobijene korišćenjem parametara izračunatih prvom metodom, metodom stohastičkog gradijenta, ta vrednost je mnogo veća. Odstupanje je toliko da se predikcije i ne mogu smatrati validnim.

Kako bi se bolje sagledale razlike između stvarnih i prediktivnih vrednosti, u Tabeli 3.10 predstavljen je procenat predikcija sa apsolutnom relativnom greškom manjom od 1%, 3% i 5%. Iz tabele se vidi da samo 1% predikcija dobijenih koeficijentima IGD metode ima apsolutno relativno odstupanje manje od 3%, dok kod predikcija sa SGD metodom takvih vrednosti uopšte nema. Slična

| | SG | IG |
|------|------------|----------|
| RMSE | 7.1267e+21 | 3.155492 |
| MAE | 2.1681e+21 | 2.562015 |
| MAPE | 7.7259e+22 | 44.7043 |

Tabela 3.9: Greške predviđanja AR(8) modela sa varijansom $\sigma_a^2 = 10$

je situacija i kod apsolutnog relativnog odstupanja od 5%. Samo 2% predikcija dobijenih drugom metodom ima odstupanje manje od 5%, dok je kod prve metode ta vrednost opet 0% kao i kod serije koja prati AR(5) model.

| Relativna greška | SG | IG |
|------------------|----|----|
| < 1% | 0% | 0% |
| < 3% | 0% | 1% |
| < 5% | 0% | 2% |

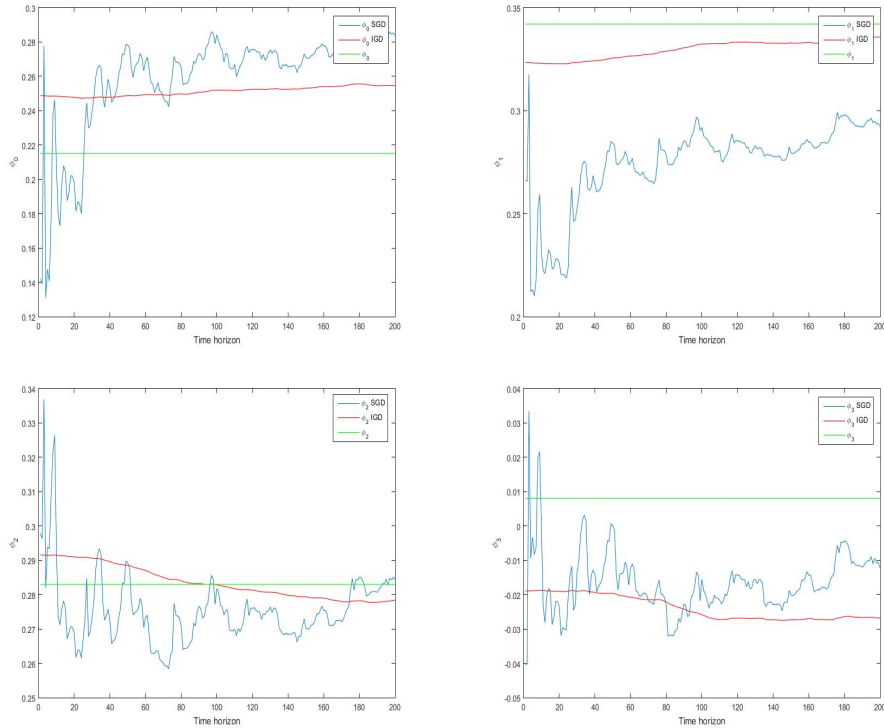
Tabela 3.10: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 8$ sa varijansom $\sigma_a^2 = 10$

Predikcije sa koeficijentima IGD metoda na primeru serije koja prati AR(8) model su bolje ispratile stvarne vrednosti. To je isti zaključak kao kod vremenske serije koja prati AR(5) model. Međutim, taj procenat odstupanja od stvarnih vrednosti iako je manji nego kod predikcija dobijenih koeficijentima SGD metoda, je dosta visok kod obe serije. Jedino su se kod AR(3) modela bolje pokazale predikcije sa koeficijentima dobijenim metodom stohastičkog gradijentnog spusta.

3.7 Komparativna analiza rezultata predikcija

Kako bi se bolje sagledale varijacije koeficijenata i predikcije dobijene pomoću njih, u nastavku je predstavljena analiza kretanja koeficijenata za AR(3) seriju sa varijansom $\sigma_a^2 = 0.1$, AR(5) sa varijansom $\sigma_a^2 = 1$ i AR(8) gde je $\sigma_a^2 = 10$. Nakon toga prelazimo na ispitivanje različitih varijansi na već pomenutim modelima i na to kako različite kombinacije reda p i varijanse utiču na kretanje simuliranih serija, performanse metoda i njihovih predikcija.

Na Slici 3.11 je predstavljeno kretanje ocenjenih koeficijenata AR(3) modela pomoću dva različita metoda - metod stohastičkog gradijenta i metod rastućeg stohastičkog gradijenta. S obzirom da se prilikom računanja dobija matrica koeficijenata, gde prva kolona predstavlja koeficijent ϕ_0 za $k = 1, \dots, 200$, druga kolona ϕ_1 , i tako sve do ϕ_p za isti broj koraka k , na graficima je ustvari predstavljeno poređenje po kolonama dve različite matrice koje predstavljaju dva metoda ocene parametara. Kao što je naznačeno u legendi, plavom bojom je predstavljeno kretanje koeficijenata dobijenih SGD metodom, crvenom bojom su predstavljeni oni dobijeni metodom rastućeg stohastičkog gradijenta (IGD), dok su zelenom bojom predstavljene početne vrednosti koeficijenata koje smo dobili metodom slučajnog izbora.

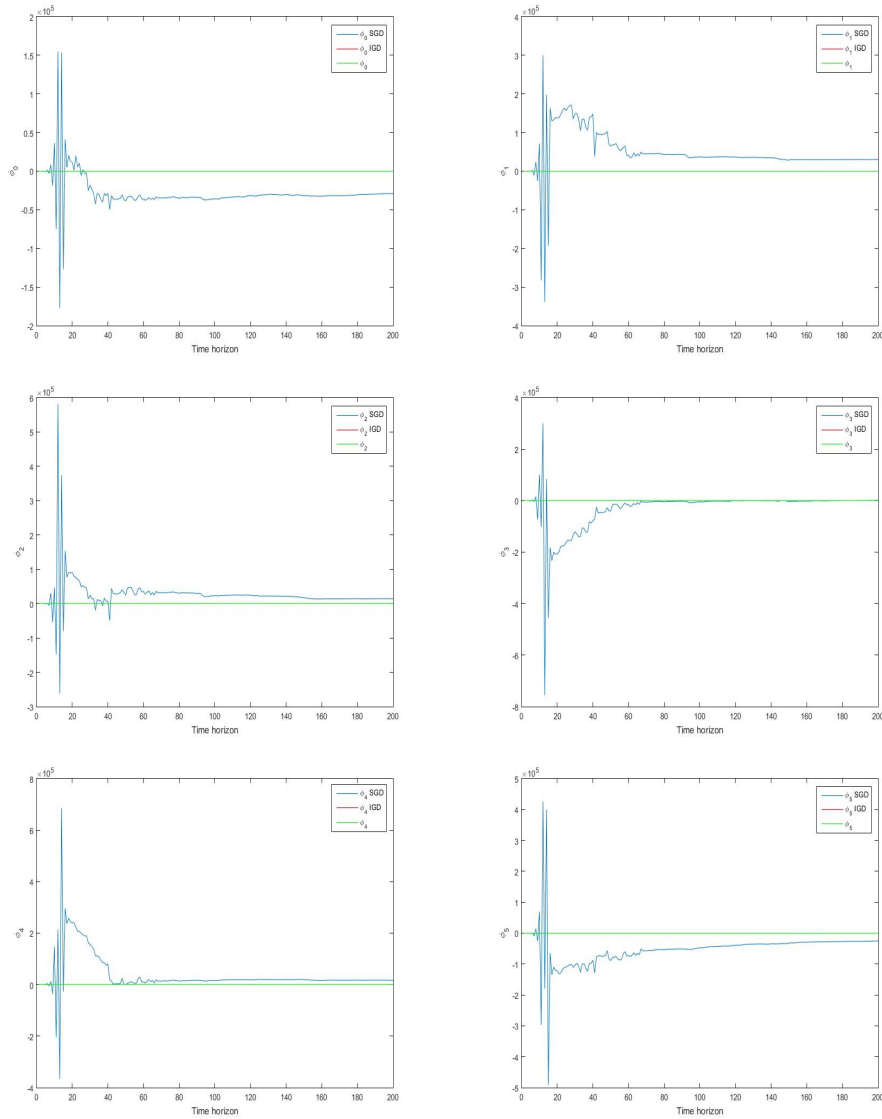


Slika 3.11: Poređenje koeficijenata dva metoda za ocenu parametara kod simulirane serije koja prati AR(3) model

Ono što se prvo uočava to je da kod koeficijenata dobijenih metodom rastućeg stohastičkog gradijenta nema velikih skokova kao kod onih dobijenih prvom metodom. Iz koraka u korak koeficijenti se ažuriraju, ali sa malim promenama u vrednostima. Tako je vrednost za ϕ_0 oko 0.25, za ϕ_1 ona se kreće od 0.32 do 0.33 i slično za ϕ_2 oko 0.28. Samo ϕ_3 ima negativnu vrednost, koja se kreće malo ispod nule za svih $k = 1, \dots, 200$ perioda.

Kod serije koja prati AR(5) model, situacija je slična što se tiče metoda za ocenu parametara. Ponovo se visoke vrednosti i veliki skokovi dešavaju prilikom ocene parametara metodom stohastičkog gradijenta, i to na samom početku, dok su one dobijene IGD metodom dosta niže i sa manjim oscilacijama tokom svih $k = 1, \dots, 200$ perioda.

Bez obzira što su oscilacije u vrednostima koeficijenata u obe vremenske serije slične (kod SGD metoda su vrednosti veće dok kod IGD osciluju oko iste) rezultati o predikcijama se razlikuju. Kao što smo spomenuli u prethodnom poglavlju, kod AR(3) serije su stvarne vrednosti bolje ispraćene predikcijama dobijenim pomoću koeficijenata SGD metoda dok je kod serije koja prati AR(5) model suprotna situacija. Stvarne vrednosti su bolje ispraćene predikcijama dobijenim pomoću koeficijenata rastućeg stohastičkog gradijenta. Ako bismo i koeficijente AR(8) modela predstavili grafički, dobili bismo sličan rezultat kao u AR(5) seriji.



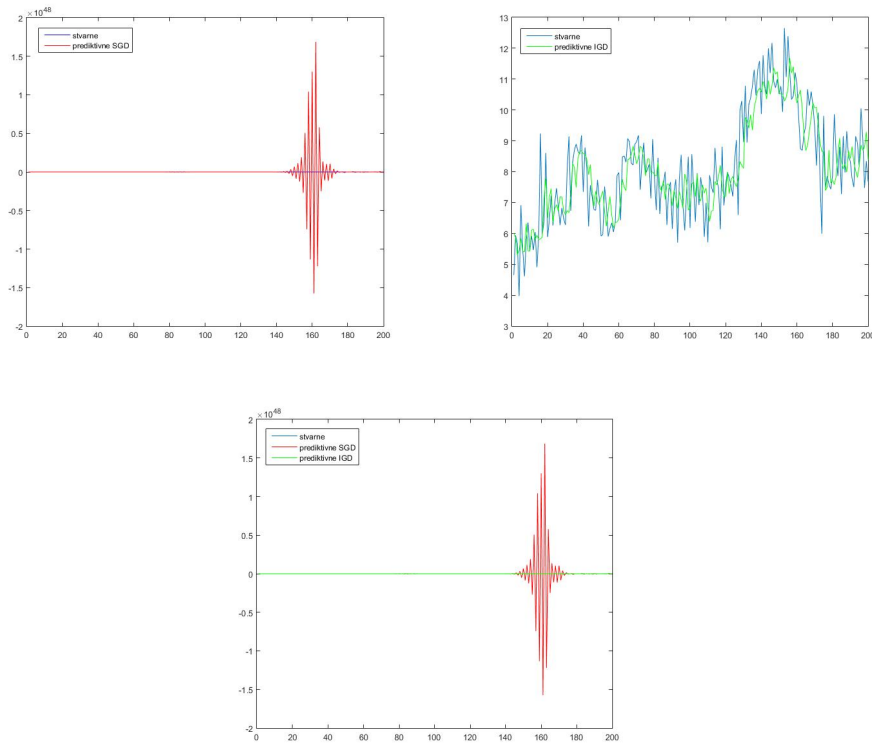
Slika 3.12: Poređenje dva metoda za ocenu parametara kod simulirane serije koja prati AR(5) model

Sve ovo sa koeficijentima je analizirano za prve tri serije, odnosno za simuliranu seriju koja prati AR(3) model sa varijansom $\sigma_a^2 = 0.1$, seriju AR(5) sa varijansom $\sigma_a^2 = 1$ i AR(8) sa varijansom $\sigma_a^2 = 10$. Da bismo dobili još bolju analizu, zanimljivo bi bilo ispitati različite varijanse na prethodnim modelima.

Dakle, potrebno je da izvedemo zaključak o tome kako različite kombinacije reda p i varijanse utiču na kretanje simuliranih serija, performansi metoda i njihovih predikcija. Ovde se nećemo baviti detaljnom analizom statističkih osobina, već samo ispitati ono što je neophodno da bi uopšte imalo smisla tražiti predikcije - a to je stacionarnost. Na slikama koje slede, lako se uočava da sve

serije imaju testerast izgled i da vrednosti osciluju bez rastućih ili opadajućih trendova. To ukazuje na stacionarnost, što je i dodatno potvrđeno Dickey-Fuller-ovim (DF) testom jediničnih korena, ali nećemo navoditi p -vrednosti za svaku seriju ponaosob. S obzirom da je slaba stacionarnost prisutna, može se krenuti sa daljom analizom.

Za početak će biti predstavljene varijacije na AR(3) modelu. Analizi kad serija prati AR model, reda $p = 3$ sa varijansom 0.1, biće pridružena analiza sa varijansama $\sigma^2 = 1$ i $\sigma^2 = 10$. Kako se povećanje varijanse odrazilo na izračunavanje predikcija grafički je predstavljeno na Slikama 3.13 i 3.14.



Slika 3.13: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(3) model sa varijansom 1

Kao što je i naznačeno u legendi, plava boja predstavlja stvarne vrednosti, crvena prediktivne vrednosti dobijene koeficijentima SGD metoda, dok zelena boja predstavlja predikcije dobijene koeficijentima IGD metoda. Ako se ponovo osvrnemo na rezultat koji je dala serija sa varijansom 0.1, uočava se da je sada rezultat drugačiji. Ovog puta, stvarne vrednosti su bolje ispraćene predikcijama dobijenim uz pomoć koeficijenata IGD metoda. Kod predikcija dobijenih SGD metodom sada su приметni skokovi, što nije bio slučaj kod serije sa varijansom 0.1. Ponovo se oslanjamo i na greške predviđanja, koje su predstavljene u Tabeli 3.11. Standardna devijacija greške predviđanja (RMSE) (1.7) i prosečna apso-

lutna greška (MAE) (1.8) trebaju da budu blizu nule, što ovde i jeste slučaj kod metode rastućeg stohastičkog gradijenta. Prosečno apsolutno procentualno odstupanje (MAPE) (1.9) od stvarnih vrednosti je 9.48% kod predikcija dobijenih isto pomoću koeficijenata rastućeg stohastičkog gradijenta. To je dosta dobar rezultat, dok su predikcije dobijene korišćenjem parametara prvog metoda mnogo veće od stvarnih, pa se one ne mogu smatrati validnim.

| | SG | IG |
|------|------------|----------|
| RMSE | 2.4820e+47 | 0.942081 |
| MAE | 5.9255e+46 | 0.745764 |
| MAPE | 5.9785e+47 | 9.480557 |

Tabela 3.11: Greške predviđanja AR(3) modela sa varijansom $\sigma_a^2 = 1$

Kako bi se bolje sagledala valjanost dobijenih predikcija, u Tabeli 3.12 prikazani su procenti predikcija koji su imali apsolutnu relativnu grešku manju od 1%, 3% i 5%. S obzirom na grafik, moglo se i pretpostaviti da će vrednosti kod prvog metoda biti nula, dok se metod rastućeg stohastičkog gradijenta pokazao najbolje do sad. Čak 33.5% predikcija dobijenih ovom metodom ima apsolutno relativno odstupanje manje od 5%. Što znači da je trećina stvarnih vrednosti test perioda vremenske serije dobro ispraćena predikcijama.

| Relativna greška | SG | IG |
|------------------|----|-------|
| < 1% | 0% | 6.5% |
| < 3% | 0% | 20.5% |
| < 5% | 0% | 33.5% |

Tabela 3.12: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 3$ sa varijansom $\sigma_a^2 = 1$

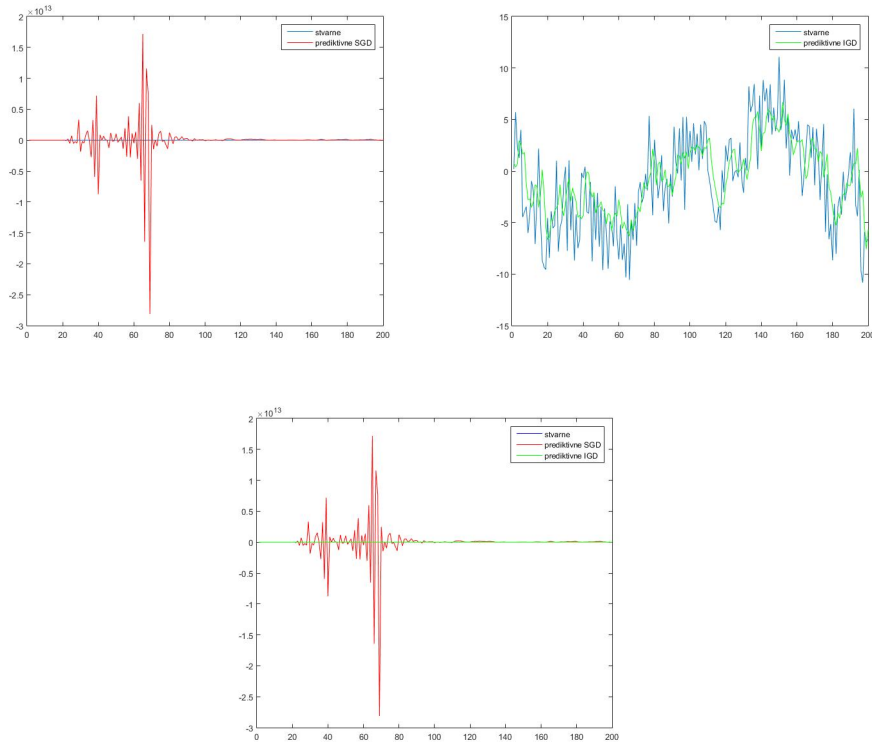
Prateći iste korake, analizira se i poslednja u nizu serija AR(3) modela čija je varijansa $\sigma_a^2 = 10$. Na Slici 3.14 je predstavljeno kretanje stvarnih i prediktivnih vrednosti, prvo sa pojedinačnim predikcijama da bi se lakše uočile razlike, a onda sve tri vrednosti zajedno.

Iako se može naslutiti rezultat koji je metod bolji, greške predviđanja date u Tabeli 3.13 će dati precizne brojeke za poređenje sa drugim vremenskim serijama. Prosečna apsolutna greška (MAE) (1.8) kod predikcija dobijenih pomoću koeficijenata IGD metoda je 2.63, što znači da se u proseku stvarna od prediktivne vrednosti razlikuje za 2.63, što je dosta malo odstupanje. Prosečno apsolutno procentualno odstupanje (MAPE) (1.9) od stvarnih vrednosti je približno 52%.

| | SG | IG |
|------|-------------|----------|
| RMSE | 3.08360e+12 | 3.244902 |
| MAE | 8.8650e+11 | 2.629424 |
| MAPE | 4.0349e+13 | 52.12005 |

Tabela 3.13: Greške predviđanja AR(3) modela sa varijansom $\sigma_a^2 = 10$

U Tabeli 3.14 su i za ovu vremensku seriju prikazani procenti predikcija koji imaju apsolutnu relativnu grešku manju od 1%, 3% i 5%. U slučaju



Slika 3.14: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(3) model sa varijansom 10

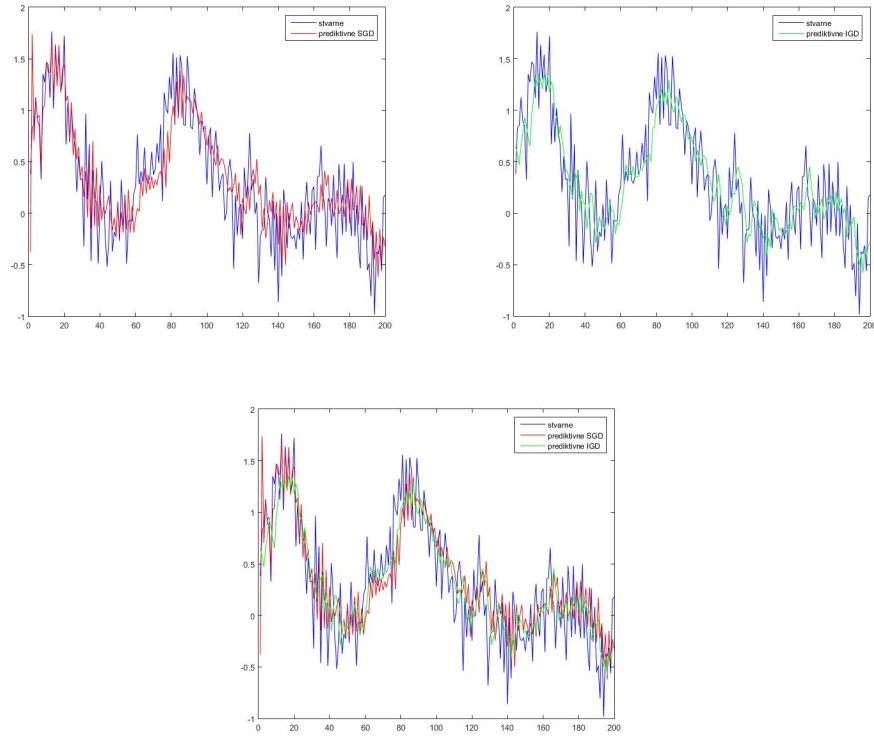
ove vremenske serije te vrednosti su jako male, iako odstupanja nisu velika kod predikcija dobijenih pomoću koeficijenata IGD metoda što je prikazano u prethodnoj tabeli. Razlog tome je što, bez obzira što ta odstupanja nemaju velike vrednosti ona ipak postoje u delu vremenske serije koji se koristi kao test period.

| Relativna greška | SG | IG |
|------------------|----|----|
| < 1% | 0% | 0% |
| < 3% | 0% | 1% |
| < 5% | 0% | 2% |

Tabela 3.14: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 3$ sa varijansom $\sigma_a^2 = 10$

U nastavku se po istom principu prelazi na analizu vremenske serije koja prati AR model reda $p = 5$. Gledajući grafički prikaz serije koja prati AR(5) model sa varijansom $\sigma_a^2 = 0.1$, vidi se da po prvi put kombinacija reda i varijanse jedne simulirane serije rezultira predikcijama koje su iako dobijene različitim metodama, slično ispratile kretanje stvarnih vrednosti. Reklo bi se da

su najpreciznije do sad, ali računski deo će pomoći da se vidi da li je to zaista tako.



Slika 3.15: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(5) model sa varijansom 0.1

Na osnovu Tabele 3.15 vidi se da su vrednosti u obe kolone slične, tj. da predikcije dobijene pomoću koeficijenata oba metoda deluju podjednako dobro. Koeficijenti stohastičkog gradijenta dali su za nijansu preciznije predikcije. Svega 13.2% iznosi apsolutno procentualno odstupanje (MAPE) (1.9) od stvarnih vrednosti.

| | SG | IG |
|------|-----------|-----------|
| RMSE | 0.330729 | 0.337792 |
| MAE | 0.264885 | 0.274188 |
| MAPE | 13.212817 | 13.873350 |

Tabela 3.15: Greške predviđanja AR(5) modela sa varijansom $\sigma_a^2 = 0.1$

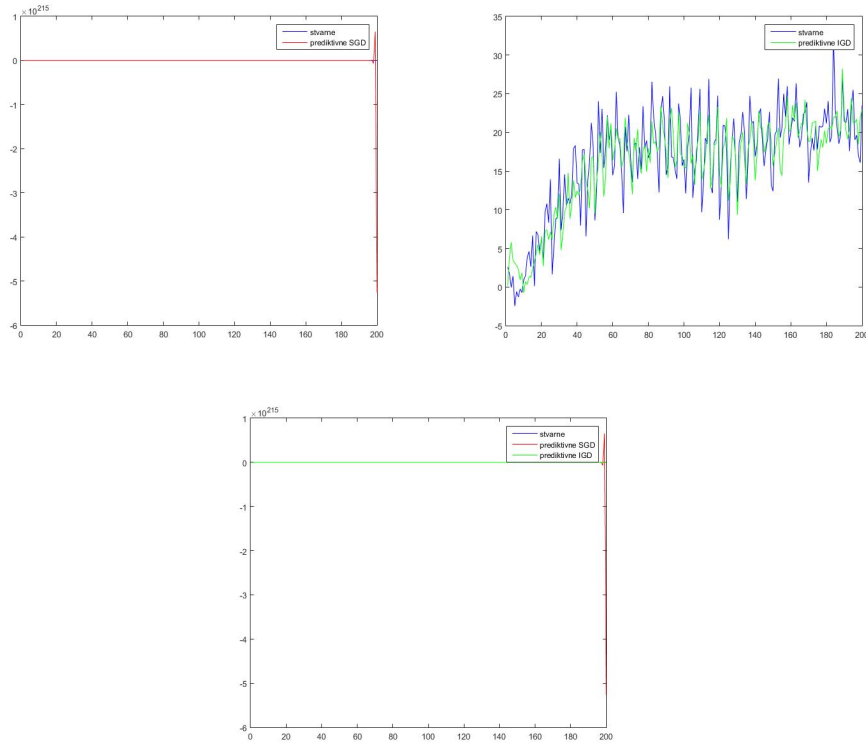
U Tabeli 3.16 predstavljen je procenat predikcija sa apsolutnom relativnom greškom manjom od 1%, 3% i 5%. Rezultati potvrđuju brojke iz prethodne tabele. Procenti se, poredeći oba metoda, malo razlikuju i kao što je već navedeno metod stohastičkog gradijenta je dao bolje predikcije. Iako se

na grafiku činilo da su ovo možda najbolji rezultati do sad, odstupanja su iako mala, prisutna u velikoj meri.

| Relativna greška | SG | IG |
|------------------|------|------|
| < 1% | 1.5% | 0.5% |
| < 3% | 2% | 1% |
| < 5% | 5% | 4% |

Tabela 3.16: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 5$ sa varijansom $\sigma_a^2 = 0.1$

Kako bi analiza za vremensku seriju koja prati AR(5) model bila kompletna, preostaje još da se vidi kako se ponašaju predikcije kada je varijansa najveća, odnosno $\sigma_a^2 = 10$. Na Slici 3.16 predstavljene su stvarne i prediktivne vrednosti dobijene pomoću oba gradijentna metoda.



Slika 3.16: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(5) model sa varijansom 10

Sem grafičkog prikaza određene su i vrednosti grešaka predviđanja, prikazane u Tabeli 3.17 koje će kao i do sad pomoći da se prediktivne vrednosti bolje izanaliziraju. Podsećanja radi, vrednost MAPE (1.9) izražava se u procentima i u ovom slučaju govori da je prosečno relativno odstupanje od stvarnih vred-

nosti 27.03% kod predikcija dobijenih koeficijentima IGD metoda. Vrednosti standardnog i apsolutnog odstupanja od stvarnih vrednosti nisu bliska nuli ali su niska kod drugog metoda, dok su za prvi metod sve vrednosti jako velike.

| | SG | IG |
|------|-------------|----------|
| RMSE | ∞ | 3.225143 |
| MAE | 2.9966e+213 | 2.584442 |
| MAPE | 1.3448e+214 | 27.02902 |

Tabela 3.17: Greške predviđanja AR(5) modela sa varijansom $\sigma_a^2 = 10$

Kada su u pitanju procenti predikcija koji imaju apsolutnu relativnu grešku manju od 1%, 3% i 5%, Tabela 3.18 daje jasan prikaz. Predikcije dobijene pomoću koeficijenata stohastičkog gradijenta nisu ispratile stvarne vrednosti uopšte, dok su predikcije sa koeficijentima iz stohastičkog gradijenta sa rastućim uzorkom u nekoj meri ispratile stvarne vrednosti. Skoro petina vrednosti iz test perioda koji se posmatra ima relativnu grešku manju od 5%.

| Relativna greška | SG | IG |
|------------------|----|-------|
| < 1% | 0% | 2% |
| < 3% | 0% | 10% |
| < 5% | 0% | 18.5% |

Tabela 3.18: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 5$ sa varijansom $\sigma_a^2 = 10$

Za kraj je preostala da se analizira još serija koja prati AR(8) model, sa varijansama $\sigma_a^2 = 0.1$ i $\sigma_a^2 = 1$. Kao i kod većine modela do sad, na Slici 3.17 se vidi da su i ovde parametri ocenjeni gradijentnom metodom sa rastućim uzorkom dali bolje predikcije nego oni iz stohastičkog gradijenta.

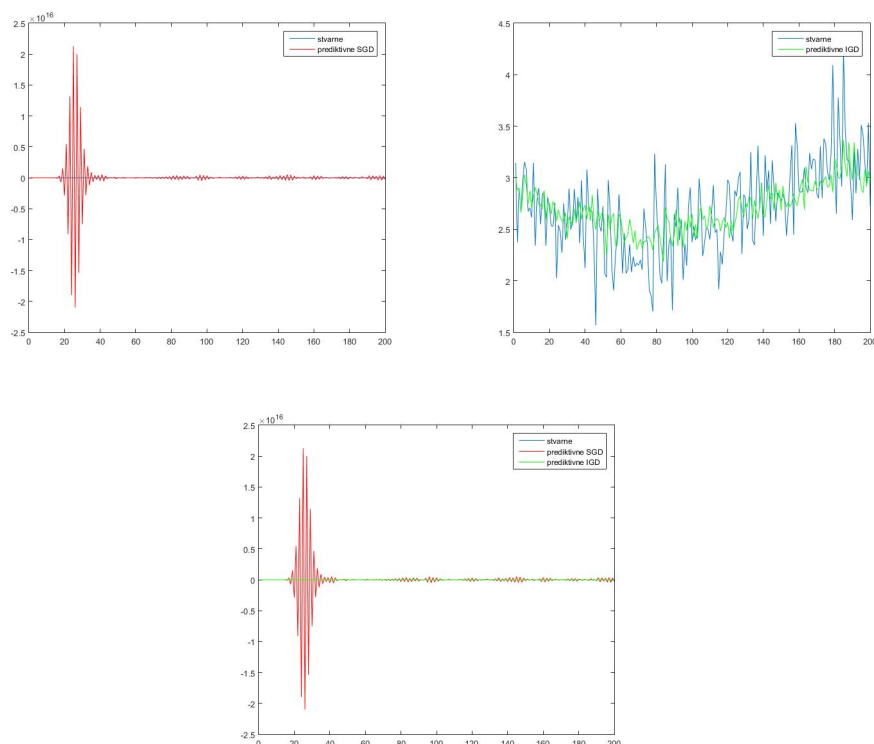
Detaljna analiza u Tabeli 3.19 govori da su standardna devijacija greške predviđanja (RMSE) (1.7) i vrednost apsolutnih odstupanja (MAE) (1.8) bliske nuli za IGD metod. Vrednost (MAPE) (1.9) govori da je relativno odstupanje od stvarnih vrednosti prosečno za 9.87%.

| | SG | IG |
|------|------------|----------|
| RMSE | 3.4765e+15 | 0.326132 |
| MAE | 9.5033e+14 | 0.255162 |
| MAPE | 3.8154e+16 | 9.871197 |

Tabela 3.19: Greške predviđanja AR(8) modela sa varijansom $\sigma_a^2 = 0.1$

Iz Tabele 3.20 se vidi da čak 33% predikcija dobijenih koeficijentima IGD metoda ima apsolutno relativno odstupanje manje od 5%, dok 20.5% ima odstupanje manje od 3% od stvarne vrednosti. Slična situacija kod relativne greške manje od 5% bila je kod vremenske serije koja prati AR(5) model ali sa varijansom $\sigma_a^2 = 1$, gde je isto 33% prediktivnih vrednosti imalo manju relativnu grešku od navedene.

Pre sumiranja svih rezultata i donošenja zaključaka ostalo je još da se osvrnemo na simuliranu seriju koja prati AR(8) model, sa varijansom $\sigma_a^2 = 1$



Slika 3.17: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(8) model sa varijansom 0.1

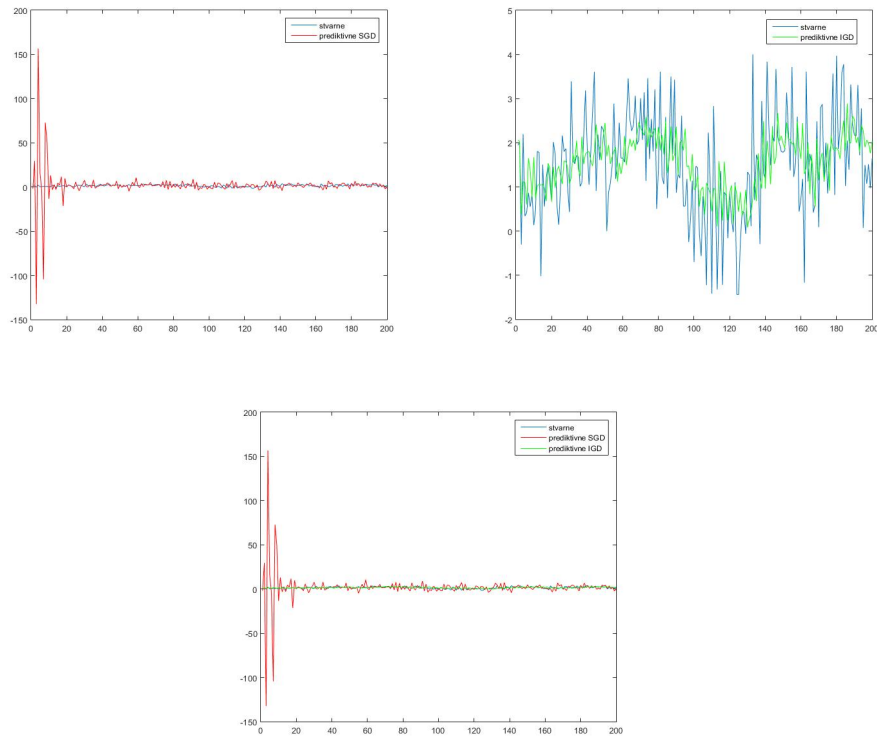
| Relativna greška | SG | IG |
|------------------|----|-------|
| < 1% | 0% | 6.5% |
| < 3% | 0% | 20.5% |
| < 5% | 0% | 33% |

Tabela 3.20: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 8$ sa varijansom $\sigma_a^2 = 0.1$

- Slika 3.18. Ono što se odmah uočava, to je da vrednosti predikcija dobijenih SGD metodom nisu mnogostruko veće od stvarnih.

U Tabeli 3.21 vrednost MAPE (1.9) izražena u procentima govori da je kod oba metoda prosečno relativno odstupanje od stvarnih vrednosti dosta veliko. Kod predikcija dobijenih koeficijentima IGD metoda ta vrednost je veća nego kod onih dobijenih SGD metodom. Dok je pak vrednost preostala dva pokazatelja manja kod predikcija dobijenih koeficijentima SGD metoda.

Procenti predikcija koje imaju apsolutno relativno odstupanje manje od zadatih granica prikazano je u Tabeli 3.22. Na osnovu ovoga reklo bi se da su prediktivne vrednosti dobijene koeficijentima IGD metoda za nijansu bolje ispratile stvarne, ali je taj procenat svakako mali. S obzirom na veličinu test



Slika 3.18: Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(8) model sa varijansom 1

| | SG | IG |
|------|------------|------------|
| RMSE | 17.866242 | 1.0023300 |
| MAE | 5.3021950 | 0.7874282 |
| MAPE | 2.6523e+02 | 6.0809e+02 |

Tabela 3.21: Greške predviđanja AR(8) modela sa varijansom $\sigma_a^2 = 1$

perioda za koji se prave predikcije, samo za 15 prediktivnih vrednosti (od 200) je relativna greška manja od 5%.

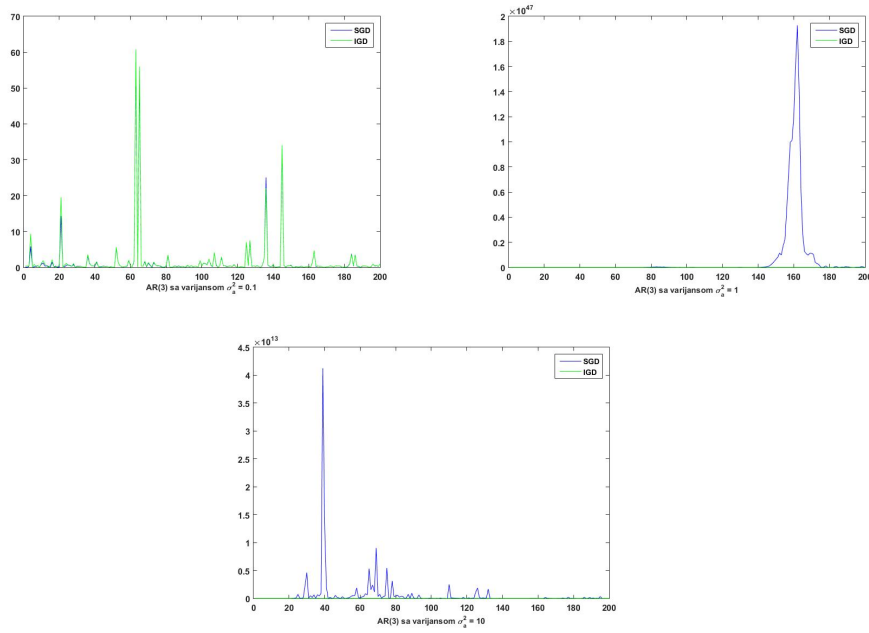
| Relativna greška | SG | IG |
|------------------|------|------|
| < 1% | 1.5% | 0.5% |
| < 3% | 2% | 4.5% |
| < 5% | 3% | 7.5% |

Tabela 3.22: Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 8$ sa varijansom $\sigma_a^2 = 1$

3.8 Grafički prikaz relativne greške i analiza

Ukoliko bi posmatrali nešto jednostavnije kao kriterijum kada se dođe do zaključivanja koji je model bolje ispratio stvarne vrednosti, to bi bila relativna greška. Razmotriće se kakve su stvarne u odnosu na prediktivne vrednosti kod svih predstavljenih serija, počevši od simulirane serije reda $p = 3$.

Na Slici 3.19 grafički su predstavljene greške predikcija za oba metoda ocene parametara, pa je tako plavom bojom predstavljen metod stohastičkog gradijenta, a zelenom metod rastućeg stohastičkog gradijenta. Vrednosti varijanse su s leva na desno - $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$ i $\sigma_a^2 = 10$ u kombinaciji sa redom $p = 3$.

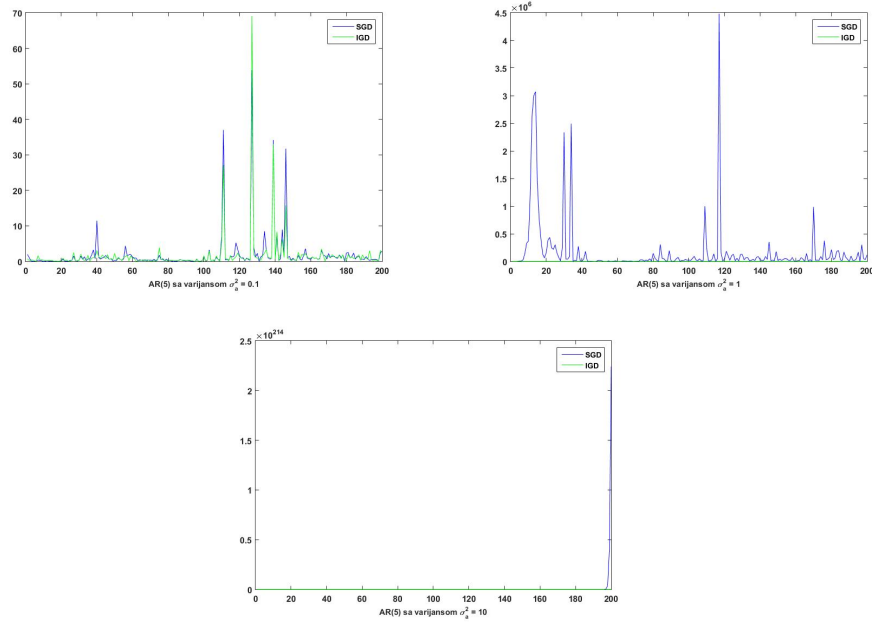


Slika 3.19: Relativne greške kod simulirane serije koja prati AR(3) model, gde je $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$, $\sigma_a^2 = 10$ respektivno; SGD vs. IGD

Ono što se odmah uočava to je da su greške odstupanja najmanje za seriju sa najmanjom varijansom. Primećuje se i da se na prvom grafiku vrednosti za oba metoda međusobno prate, nema drastičnih razlika. Uzimajući u obzir i prethodnu analizu sa greškama predviđanja, ovo je samo potvrda da je SGD metod dao bolje predikcije u ovom slučaju. Na drugom i trećem grafiku se vidi jasna razlika ova dva metoda. Parametri dobijeni SGD metodom vode ka predikcijama koje dosta odstupaju od stvarnih vrednosti. One su mnogostruko veće. Stoga se u ovom slučaju, za ocenu parametara, bolje pokazao gradijentni metod sa rastućim uzorkom.

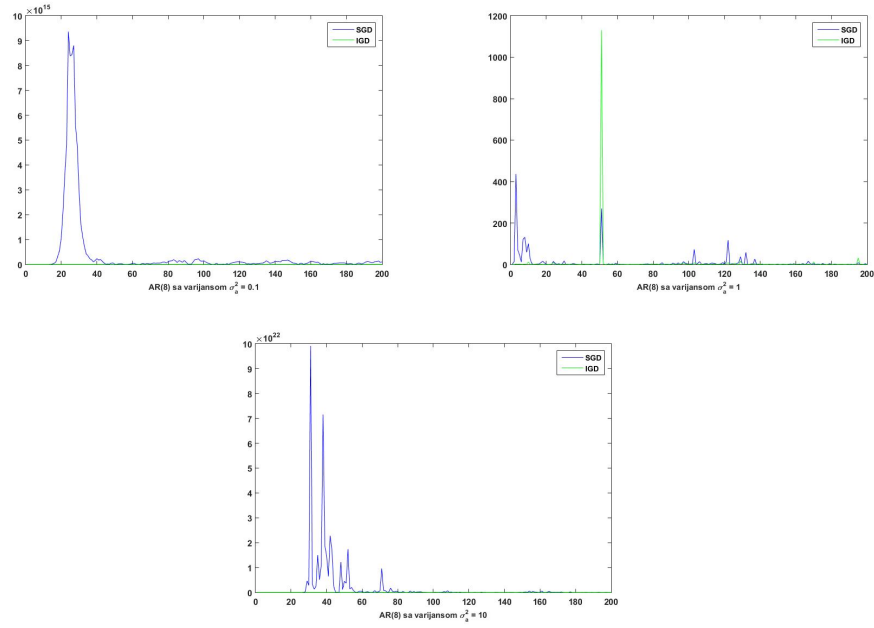
Kod simulirane serije koja prati AR(5) model, vidi se da je najidealnija situacija kada varijansa ima najmanju vrednost, Slika 3.20. To su ujedno i najpreciznije predikcije od svih posmatranih, što je već predstavljeno Slikom 3.15. Ono što se ovde možda jasnije vidi kad je u pitanju serija reda $p = 5$ sa

varijansom $\sigma_a^2 = 0.1$, to je da kod predikcija sa IGD metodom ima manji broj skokova ali su oni veći. Slično kao i kod AR(3) serije, i ovde se sa povećanjem varijanse povećavaju i greške predviđanja. Parametri dobijeni SGD metodom vode ka predikcijama koje u mnogome odstupaju od stvarnih vrednosti. Toliko da se one i ne bi mogle uzeti kao validne.



Slika 3.20: Relativne greške kod simulirane serije koja prati AR(5) model, gde je $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$, $\sigma_a^2 = 10$ respektivno; SGD vs. IGD

Kada se dođe do serije sa najvećim redom, što je u našem slučaju $p = 8$ - situacija se malo razlikuje. Na Slici 3.21 se vidi da su relativne greške predikcija najmanje u slučaju kad je varijansa $\sigma_a^2 = 1$, dok su za preostale dve vrednosti one jako velike kod SGD metoda.



Slika 3.21: Relativne greške kod simulirane serije koja prati AR(8) model, gde je $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$, $\sigma_a^2 = 10$ respektivno; SGD vs. IGD

Posmatrajući u globalu, nakon ove analize može se zaključiti da je metod rastućeg stohastičkog gradijenta dao parametre koji su doveli do boljih predikcija u svim kombinacijama vrednosti reda p i varijanse σ_a^2 .

4 Zaključak

Analizom vremenskih serija donosilac odluke dobija informacije o kretanju pojave, kao i mogućnost predviđanja uz određenu grešku. Na bazi dobijenih informacija o kretanju pojedinih statističkih parametara donosi se najpouzdanija odluka. Analiza vremenskih serija po pravilu obuhvata i rešavanje problema optimizacije. Različiti metodi optimizacije daju različite rezultate po pitanju kvaliteta aproksimacije rešenja.

Osnovni cilj istraživanja je bilo upoređivanje ocene parametara dobijenih korišćenjem dve konceptualno različite varijante gradijentnog metoda: *stohastički* i *inkrementalni*. Istraživanje je sprovedeno na simuliranim podacima uz pomoć softverskog paketa MATLAB. Posmatrano je ukupno devet vremenskih serija koje prate autoregresivni model, različitog reda i varijanse u cilju upoređivanja koji od dva gradijentna metoda daje bolje ocene parametara modela, i kako promene reda i varijanse utiču na performanse metoda i njihovih predikcija. Za prve tri serije je urađena kompletna statistička analiza, deskriptivna statistika, ocena parametara metodom najmanjih kvadrata, predikcije i ocena parametara pomoću metode stohastičkog gradijenta i gradijenta sa rastućim uzorkom. Kasnije se analiza preostalih serija sve više bazira na analizi samih grešaka predikcija.

Posmatrajući sve simulirane serije i njihove predikcije, stohastički gradijent sa rastućim uzorkom je formirao kvalitetniju aproksimaciju rešenja. Samim tim što je njegova aproksimacija kvalitetnija, ona je i skuplja. On uzima sve raspoložive podatke i računa pun gradijent, dok metod običnog stohastičkog gradijenta koristi deo uzorka, formiranog na osnovu novopridošlih podataka. Na osnovu sprovedenog istraživanja, posmatrajući serije i njihove kombinacije reda i varijanse, došli smo do zaključka da metod stohastičkog gradijenta radi bolje samo kad su varijansa i red modela jako mali. Ukoliko je varijansa velika, ili imamo veliki red modela a malu varijansu - model eksplodira.

To za koji model se treba odlučiti diktira i veličina uzorka i koja je frekvencija kojom stižu novi podaci. U našem slučaju podaci pristižu na dnevnom nivou, ali ako bi to bilo na primer svake sekunde, onda metod rastućeg stohastičkog gradijenta ne bi bilo moguće koristiti jer bi trebalo mnogo vremena da bi se izračunao pun gradijent. U tom slučaju kad je frekvencija dotoka novih informacija velika, stohastički gradijentni metod bi sigurno bio bolja opcija.

Predlog za naredno istraživanje bi bio da se napravi varijanta metoda koji nije ni stohastički ni inkrementalni, nego *minibatch* koji sam definiše veličinu uzorka u svakoj iteraciji.

Pregled slika

| | | |
|------|---|----|
| 1.1 | Tipovi vremenskih serija: (a) konstantan proces; (b) proces sa trendom; (c) serija sa cikličnim varijacijama; (d) serija sa sezonskim varijacijama[4] | 12 |
| 2.1 | Ilustracija gradijentnog metoda [8] | 23 |
| 2.2 | Ilustracija uticaja prevelike i premale dužine koraka α [7] | 27 |
| 2.3 | Ponašanje gradijentnog metoda i stohastičkog gradijentnog metoda [11] | 29 |
| 3.1 | Kretanje simulirane serije koja prati AR(3) model, gde je $\sigma_a^2 = 0.1$ | 34 |
| 3.2 | Kretanje simulirane serije koja prati AR(5) model, gde je $\sigma_a^2 = 1$ | 34 |
| 3.3 | Kretanje simulirane serije koja prati AR(8) model, gde je $\sigma_a^2 = 10$ | 34 |
| 3.4 | Autokorelaciona funkcija AR(3) serije | 36 |
| 3.5 | Autokorelaciona funkcija AR(5) serije | 37 |
| 3.6 | Autokorelaciona funkcija AR(8) serije | 38 |
| 3.7 | Skica algoritma za dobijanje predikcija pomoću novih koeficijenata | 42 |
| 3.8 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(3) model sa varijansom 0.1 | 43 |
| 3.9 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(5) model sa varijansom 1 | 46 |
| 3.10 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(8) model sa varijansom 10 | 49 |
| 3.11 | Poređenje koeficijenata dva metoda za ocenu parametara kod simulirane serije koja prati AR(3) model | 51 |
| 3.12 | Poređenje dva metoda za ocenu parametara kod simulirane serije koja prati AR(5) model | 52 |
| 3.13 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(3) model sa varijansom 1 | 53 |
| 3.14 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(3) model sa varijansom 10 | 55 |
| 3.15 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(5) model sa varijansom 0.1 | 56 |
| 3.16 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(5) model sa varijansom 10 | 57 |
| 3.17 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(8) model sa varijansom 0.1 | 59 |
| 3.18 | Stvarne i prediktivne vrednosti serije dobijene metodom običnog i rastućeg stohastičkog gradijenta kod simulirane serije koja prati AR(8) model sa varijansom 1 | 60 |
| 3.19 | Relativne greške kod simulirane serije koja prati AR(3) model, gde je $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$, $\sigma_a^2 = 10$ respektivno; SGD vs. IGD | 61 |

| | | |
|------|---|----|
| 3.20 | Relativne greške kod simulirane serije koja prati AR(5) model, gde je $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$, $\sigma_a^2 = 10$ respektivno; SGD vs. IGD . . . | 62 |
| 3.21 | Relativne greške kod simulirane serije koja prati AR(8) model, gde je $\sigma_a^2 = 0.1$, $\sigma_a^2 = 1$, $\sigma_a^2 = 10$ respektivno; SGD vs. IGD . . . | 63 |

Pregled tabela

| | | |
|------|---|----|
| 3.1 | Deskriptivna statistika simuliranih serija: AR(3), $\sigma_a^2 = 0.1$; AR(5), $\sigma_a^2 = 1$; AR(8), $\sigma_a^2 = 10$ | 35 |
| 3.2 | Ocena parametara AR(3) modela sa normalnom $\mathcal{N}(0, 0.1)$ raspodelom | 39 |
| 3.3 | Greške predviđanja AR(3) modela sa varijansom $\sigma_a^2 = 0.1$ | 43 |
| 3.4 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 3$ sa varijansom $\sigma_a^2 = 0.1$ | 44 |
| 3.5 | Ocena parametara AR(5) modela sa normalnom $\mathcal{N}(0, 1)$ raspodelom | 44 |
| 3.6 | Greške predviđanja AR(5) modela sa varijansom $\sigma_a^2 = 1$ | 46 |
| 3.7 | Procenat predikcija čije su relativne greške manje od zadatih granica | 47 |
| 3.8 | Ocena parametara AR(8) modela sa normalnom $\mathcal{N}(0, 10)$ raspodelom | 48 |
| 3.9 | Greške predviđanja AR(8) modela sa varijansom $\sigma_a^2 = 10$ | 50 |
| 3.10 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 8$ sa varijansom $\sigma_a^2 = 10$ | 50 |
| 3.11 | Greške predviđanja AR(3) modela sa varijansom $\sigma_a^2 = 1$ | 54 |
| 3.12 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 3$ sa varijansom $\sigma_a^2 = 1$ | 54 |
| 3.13 | Greške predviđanja AR(3) modela sa varijansom $\sigma_a^2 = 10$ | 54 |
| 3.14 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 3$ sa varijansom $\sigma_a^2 = 10$ | 55 |
| 3.15 | Greške predviđanja AR(5) modela sa varijansom $\sigma_a^2 = 0.1$ | 56 |
| 3.16 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 5$ sa varijansom $\sigma_a^2 = 0.1$ | 57 |
| 3.17 | Greške predviđanja AR(5) modela sa varijansom $\sigma_a^2 = 10$ | 58 |
| 3.18 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 5$ sa varijansom $\sigma_a^2 = 10$ | 58 |
| 3.19 | Greške predviđanja AR(8) modela sa varijansom $\sigma_a^2 = 0.1$ | 58 |
| 3.20 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 8$ sa varijansom $\sigma_a^2 = 0.1$ | 59 |
| 3.21 | Greške predviđanja AR(8) modela sa varijansom $\sigma_a^2 = 1$ | 60 |
| 3.22 | Procenat predikcija čije su relativne greške manje od zadatih granica za seriju reda $p = 8$ sa varijansom $\sigma_a^2 = 1$ | 60 |

Literatura

- [1] Tsay, R. S., Analysis of Financial Time Series, Second edition, New York, A John Wiley & Sons, Inc, 2005
- [2] Nocedal, J., Wright S. J., Numerical Optimization, Springer, 1999.
- [3] Spall, J. C., Introduction to Stochastic Search and Optimization, Wiley-Interscience Series in Discrete Mathematics, New Jersey, 2003.
- [4] Kovačić, J. Zlatko, Analiza vremenskih serija, Beograd, 1995.
- [5] Rajter-Ćirić D., Verovatnoća, Prirodno-matematički fakultet, UNS, 2009.
- [6] Lozanov-Crvenković Z., Statistika, Prirodno-matematički fakultet, UNS, 2012.
- [7] [https : //sebastianraschka.com/Articles/2015_singlelayer_neurons.html](https://sebastianraschka.com/Articles/2015_singlelayer_neurons.html)
- [8] <https://sebastianraschka.com/faq/docs/closed-form-vs-gd.html>
- [9] Friedlander A., Krejić N., Krklec Jerinkić N., Lectures on Fundamentals of Numerical Optimization, Novi Sad, 2019.
- [10] <https://stackoverflow.com/questions/21064030/gradient-descent-in-linear-regression>
- [11] Zečević A., Nikolić M., Mašinsko učenje, Beograd, 2019.
- [12] Friedlander P.M., Schmidt M., Hybrid Deterministic-Stochastic Methods for Data fitting, 2013.
- [13] Rožnjik A., Optimizacija problema sa stohastičkim ograničenjima tipa jednakosti - kazneni metodi sa promenljivom veličinom uzorka, Doktorska disertacija, Novi Sad: Prirodno-matematički fakultet, 2018.
- [14] Herceg Đ., Krejić N., Numerička analiza, UNS Prirodno-matematički fakultet, 1997.

Biografija



Milena Vajagić rođena je 26. februara 1992. godine u Novom Sadu. Osnovnu školu "Miroslav Antić" završila je u Futogu 2007. godine kao nosilac "Vukove diplome". Iste godine upisala je opšti smer gimnazije "Svetozar Marković" u Novom Sadu. Gimnaziju završava 2011. godine, gde je tokom sve četiri godine bila odličan učenik. Nakon završetka gimnazije, iste godine upisuje Prirodno-matematički fakultet u Novom Sadu, smer Primenjena matematika, modul Matematika finansija. Osnovne studije uspešno završava 2015. godine. Iste godine upisuje se na master akademske studije Primenjene matematike, modul Matematika finansija, na istom fakultetu. Položila je sve ispite predviđene nastavnim planom i programom master studija u junu 2018. godine. Od juna 2017. godine radi na poziciji test konsultanta u firmi "Sixsentix" u Novom Sadu.

UNIVERZITET U NOVOM SADU
PRIRODNO - MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJA INFORMACIJA

Redni broj:

RBR

Identifikacioni broj:

IBR

Tip dokumentacije: Monografska dokumentacija

TD

Tip zapisa: Tekstualni štampani materijal

TZ

Vrsta rada: Master rad

VR

Autor: Milena Vajagić

AU

Mentor: dr Nataša Krklec Jerinkić

ME

Naslov rada: Ocena parametara u autoregresivnim modelima - stohastički
gradijent protiv gradijenta sa rastućim uzorkom

NR

Jezik publikacije: Srpski (latinica)

JP

Jezik izvoda: s / en

JI

Zemlja publikovanja: Republika Srbija

ZP

Uže geografsko područje: Vojvodina

UGP

Godina: 2020

GO

Izdavač: Autorski reprint

IZ

Mesto i adresa: Novi Sad, Trg D. Obradovića 4

MA

Fizički opis rada: (3/68/14/22/25/0/0)(broj poglavlja/broj strana/broj literarnih citata/broj tabela/broj slika/broj grafika/broj priloga)

FO:

Naučna oblast: Matematika

NO

Naučna disciplina: Numerička matematika

ND

Ključne reči: Gradijent, gradijentni metod, ocena parametara, Stohasticki gradijent, Inkrementalni gradijent, AR model, vremenske serije

PO, UDK

Čuva se: U biblioteci Departmana za matematiku i informatiku, Prirodno-

matematički fakultet, Univerzitet u Novom Sadu

ČU

Važna napomena:

VN

Izvod: Cilj rada je da se za ocenu parametara u autoregresivnim modelima uporede dve konceptualno različite varijante gradijentnog metoda: stohastički gradijentni metod i gradijentni metod sa rastućim uzorkom. Podatke će predstavljati simulirane serije AR modela različitog reda p , sa različitim vrednostima varijanse. Uporediće se koji od dva gradijentna metoda daje bolje ocene parametara modela, za dobijanje što validnijih predikcija. Izvešće se zaključak o tome kako različite kombinacije reda p i varijanse utiču na kretanje simuliranih serija, performanse metoda i njihovih predikcija.

IZ

Datum prihvatanja teme od strane NN veća: 04.08.2020.

DP

Datum odbrane:

DO

Članovi komisije:

ČK

Predsednik: dr Nataša Krejić, redovni profesor Prirodno-matematičkog fakulteta u Novom Sadu

Član: dr Sanja Rapajić, redovni profesor Prirodno-matematičkog fakulteta u Novom Sadu

Mentor: dr Nataša Krklec Jerinkić, vanredni profesor Prirodno-matematičkog fakulteta u Novom Sadu

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCE
KEY WORDS DOCUMENTATION

Accession number:

ANO

Identification number:

INO

Document type: Monograph type

DT

Type of record: Printed text

TR

Contents Code: Master's thesis

CC

Author: Milena Vajagić

AU

Mentor: dr Nataša Krklec Jerinkić

MN

Title: Parameter estimation in Autoregressive Models - Stochastic Gradient vs Incremental Gradient

TI

Language of text: Serbian (Latin)

LT

Language of abstract: s / en

LA

Country of publication: Republic of Serbia

CP

Locality of publication: Vojvodina

LP

Publication year: 2020

PY

Publisher: Author's reprint

PU

Publication place: Novi Sad, Trg D. Obradovića 4

PP

Physical description: (3/68/14/22/25/0/0)(chapters/ pages/ quotations/ tables/ pictures/ graphics/ enclosures)

PD

Scientific field: Mathematics

SF

Scientific discipline: Numerical Mathematics

SD

Subject/Key words: Gradient Descent, parameter estimation, Stochastic Gradient Descent, Incremental Gradient Descent, Autoregressive Models, Time series

SKW Holding data: The Library of the Department of Mathematics and In-

formatics, Faculty of Science and Mathematics, University of Novi Sad

HD

Note:

N

Abstract: This thesis describes parameter estimation in autoregressive models using two different gradient methods: Stochastic Gradient Descent and Incremental Gradient Descent. Research is based on simulated data of time series following AR model, with different variations of order p and variance. More precisely, we will compare which of these two gradient methods gives better estimation of parameters which leads to more valid prediction values. At the end, we will conclude how different combinations of AR model's order p and variance, influence quality of predictions.

AB

Accepted by the Scientific Board on: 04.08.2020.

ASB Defended:

DE

Thesis defend board:

DB President: dr Nataša Krejić, full professor at Faculty of Science in Novi Sad

Member: dr Sanja Rapajić, full professor at Faculty of Science in Novi Sad

Mentor: dr Nataša Krklec Jerinkić, associate professor at Faculty of Science in Novi Sad