UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I INFORMATIKU

Anja Mihailović

Analiza vremenskih serija

-Master rad-

mentor:
dr. Zagorka Lozanov- Crvenković

Novi Sad, 2011

Sadržaj

Uvod

Analiza vremenskih serija je statističkia disciplina koja beleži najdinamičnii razvoj poslednjih decenija. Vremenska serija je uređeni niz opservacija, gde se uređivanje vrši s obzirom na vreme u jednakim vremenskim intervalima. Kako je proces donošenja odluka često povezan sa predviđanjem budućih vrednosti promenljivih koje zavise od vremena, vremenske serije i njihova analiza predstavljaju pogodno sredstvo. Naime u ovom kontekstu predviđanje podrazumeva analizu istorijskih podataka i ekstarpolaciju istih u budućnosti, uz upotrebu odgovarajućeg matematičkog modela. Da bi se omogućio stohastički karakter budućih vrednosti, pogodno je da se pretpostavi da je vrednost vremenske serije u trenutku *t* realizacija slučajne promenljive X_t. U tom kontekstu vremenska serija je realizacija familije slučajnih promenljivih koja se zove stohastički proces, pa je u prvom poglavlju dat osvrt na bitne teorijske pojmove vezane za ovu oblast.

U analizi vremenskih serija od interesa je posebna klasa stohastičkih procesa, a to su lineari stacionarni procesi. U drugom poglavlju se uvodi pojam tri grupe linearnih stacionarnih procesa koji se koriste za modeliranje vremenskih serija, a to su autoregresivni procesi, procesi pokretnih sredina i *ARMA* procesi, kao i predviđanje budućih vrednosti vremenskih serija pomoću ovih modela.

Kako se u realnom svetu dešava da su pretpostavke o stacionarnosti narušene, u trećem poglavlju će se izučavati nestacionarni linerni procesi među kojima je i autoregresivni integrisani proces pokretnih sredina (*ARIMA*). Takođe će biti objašnjena opšta strategija izgradnje *ARIMA* modela ili takozvano *Box-Jenkins*-onovo modeliranje koje obuhvata identifikaciju, ocenjivanje i proveru adekvatnosti izabranog modela vremenske serije.

U četvtom poglavlju će se analizirati vektorski modeli serija čija svrha je opisivanje međusobnih uticaja između dve ili više serija. Polazna ideja ovih modela je da na vremensku seriju, osim sopstvenih prošlih vrednosti, utiču i neke druge vremenske serije. Na samom kraju rada, uz korišćenje teorijskih pojmova obrađenih u prethodnim poglavljima, će kroz dva primera biti objašnjen način izbora odgovarajućeg modela u praksi.

1. Stohastički procesi i vremenske serije

1.1. Stohastički procesi

Definicija 1.1.1 Ω je skup svih ishoda jednog eksperimenta. Elementi ovog skupa se označavaju sa $w_i, i = 1,2, \dots$ i nazivaju se elementarni događaji.

*Definicija 1.1.2 (*Aksioma σ −polja *)* Podskup $\mathcal{F}$ partitivnog skupa $\mathcal{P}(\Omega)$ je σ −polje (σ −algebra) nad Ω ako važe uslovi:

- $\Omega \in \mathcal{F}$
- $\mathrm{A} \in \mathcal{F} \Longrightarrow \overline{\mathrm{A}} \in \mathcal{F}$
- $\{\mathrm{A_i}\}_{\mathrm{i}\in\mathbb{N}} \subseteq \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} \mathrm{A_i} \in \mathcal{F}$.

*Definicija 1.1.3 (*Aksioma verovatnoće) Neka je Ω je skup svih elementarnih događaja i $\mathcal{F}$ σ −polje nad Ω. Funkcija $P : \mathcal{F} \longrightarrow [0,1]$ se zove verovatnoća na prostoru $(\Omega, \mathcal{F})$ ako zadovoljava uslove:

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

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{1.1.1}$$

Prostor verovatnoća je uređena trojka $(\Omega, \mathcal{F}, P)$.

Definicija 1.1.4 Neka je $(\Omega, \mathcal{F}, P)$ prostor verovatnoća. Preslikavanje $X : \Omega \longrightarrow \mathbb{R}$ se zove slučajna promenljiva ako $\forall B \in \mathcal{B}$ važi da je

$$X^{-1}(B) \in \mathcal{F} \tag{1.1.2}$$

gde je $\mathcal{B}$ Borelova σ −algebra. Za X kažemo da je $\mathcal{F}$ −merljivo.

Definicija 1.1.5 Familija slučajnih promenljivih $\{X_t(w) : t \in T, w \in \Omega\}$ definisanih nad istim prostorom verovatnoće $(\Omega, \mathcal{F}, P)$ se zove se zove stohastički proces sa indeksnim skupom T.

Promenljiva w se često izostavlja u zapisu, pa se umesto toga stohastički proces označava sa $\{X_t : t \in T\}$. Očigledno je da proces zavisi od dve promenljive t i w . Za fiksirano $t_0 \in T$ stohastički proces je jedna slučajna promenljiva. Za fiksirano $w_0 \in \Omega$ $X_t(\cdot)$ je funkcija definisana na skupu T koja se naziva realizacija ili trajektorija stohastičkog procesa. Za fiksirane $t_0 \in T$ i $w_0 \in \Omega$ X_t je realan broj ili jedna realizacija stohatičkog procesa.

Definicija 1.1.6 Neka je ζ skup $\left((t_1, t_2, \dots, t_n) \in T^n, t_1 < t_2 < \cdots < t_n, n = 1,2\right)$. Tada su konačno dimenzionalne funkcije raspodele stohastičkog procesa funkcije $\{F_t(\cdot), t \in \zeta\}$ definisane sa

$$F_t(x) = P\{X_{t_1} \leq x_1,\ X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n\}, \quad (x_1, x_2, \dots, x_n) \in \mathbb{R}^n. \tag{1.1.3}$$

Sistem funkcija raspodele $\{F_t(\cdot), t \in \zeta\}$ zadovoljava dva uslova:

- uslov simetrije - ako je $\{i_1, \dots, i_n\}$ jedna permutacija brojeva od 1 do n tada važi
$$F_{t_{i_1}, \dots, t_{i_n}}\left(x_{i_1, \dots,} x_{i_n}\right) = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$$
- uslov kompatibilnosti – ako je $m < n$ za proizvoljne $t_{m+1}, \dots, t_n \in T^n$ važi
$$F_{t_1, \dots, t_m, t_{m+1} \dots t_n}(x_1, \dots, x_m, \infty, \dots, \infty) = F_{t_1, \dots, t_m}(x_1, \dots, x_m).$$

Teorema 1.1.1 (Kolmogorova) Za svaku familiju funkcija raspodele koje zadovoljavaju uslove simetrije i kompatibilnosti postoji prostor vreovatnoća $(\Omega, \mathcal{F}, P)$ i stohastički proces $\{X_t: t \in T\}$ definisan na njemu koji ima date raspodele kao svoje konačno – dimenzionalne raspodele.

Neka je $\{X_t: t \in T\}$ stohastički proces. Tada je:

- srednja vrednost stohastičkog procesa
$$E(X_t) = \mu_t, \qquad t \in T,$$
- varijansa stohastičkog procesa
$$Var(X_t) = \sigma_t^2, \qquad t \in T,$$
- kovarijansa stohastičkog procesa
$$\gamma_x(r,s) = Cov(X_r, X_s) = E\left(\left(X_r - E(X_r)\right)\left(X_s - E(X_s)\right)\right), \qquad r, s \in T,$$
- korelacija stohastičkog porcesa
$$\rho_X(r,s) = \frac{Cov(X_r, X_s)}{\sqrt{Var(X_r)Var(X_s)}}, \qquad r, s \in T.$$

Definicija 1.1.7 Stohastički proces je strogo stacionaran ako su njegove konačno dimenzionalne funkcije raspodele invarijantne u odnosu na t, odnosno ako za t_i, $t_i + t \in T$, *i=1,2,...*

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+t, \dots, t_n+t}(x_1, \dots, x_n). \tag{1.1.4}$$

Definicija 1.1.8 Stohastički proces $\{X_t, t \in T\}$ je slabo stacionaran ako su ispunjeni sledeći uslovi:

- $E(X_t) = \mu = const.\ \forall t \in T$
- $E(X_t^2) < \infty, \forall t \in T$
- $\gamma_x(r,s) = \gamma_x(r+t, s+t), \forall r, s, t \in T.$

Očigledno je da uvek iz stroge stacionarnosti sledi slaba stacionarnost stohastičkog procesa.

Definicija 1.1.9 Stohastički proces je Gausovski ako sve njegove konačno dimenzionalne funkcije raspodele imaju višedimenzionalnu normalnu raspodelu.

Ako je $\{ X_t, t \in T\}$ slabo stacionaran Gausovski proces onda slučajni vektori $(X_{t_1}, \dots, X_{t_n})$, $(X_{t_1+h}, \dots, X_{t_n+h})$, $\forall h, t_1, t_2, \dots \in \mathbb{Z}$ i $\forall n \in \mathbb{N}$ imaju istu očekivanu vrednost i autokovarijansnu matricu, pa je onda $\{ X_t, t \in T\}$ strogo stacionaran proces.

Dakle jedino u slučaju kada je stohastički proces Gausovski slaba stacionarnost implicira strogu, a obrnuto uvek važi.

1.2. Kovarijansna i korelaciona funkcija stacionarnog stohastičkog procesa

Definicija 1.2.1 Kovarijansna funkcija stohastičkog procesa $\{ X_t, t \in T\}$ je definisana sa

$$\gamma_k = Cov(X_t, X_{t+k}) = E((X_t - E(X_t))(X_{t+k} - E(X_{t+k}))), \quad t\epsilon\, T,\, k\epsilon\, \mathbb{Z} \tag{1.2.1}$$

a korelaciona funkcija

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)Var(X_{t+k})}}, \quad t\epsilon\, T,\, k\epsilon\, \mathbb{Z}. \tag{1.2.2}$$

Sa

$$\Gamma_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} \tag{1.2.3}$$

je definisana kovarijansna matrica reda n, a sa

$$\mathrm{R}_n = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix} \tag{1.2.4}$$

korelaciona matrica reda n. Očigledno je da važi $\Gamma_n = \sigma^2 \mathrm{R}_n$.

Teorema 1.2.1 Ako su γ_K i ρ_K $k\epsilon\, \mathbb{Z}$ autokovarijansna i autokorelaciona funkcija slabo stacionarnog stohastičkog procesa tada važi:

1. $\gamma_0 = Var(X_t) \geq 0,\ \rho_0 = 1$
2. $\gamma_K = \gamma_{-K}$, $\rho_K = \rho_{-K}$
3. $|\gamma_K| \leq \gamma_0$, $|\rho_K| \leq 1$.

Dokaz.

1. $\gamma_0 = E((X_t - E(X_t))(X_{t+0} - E(X_{t+0})) = E(X_t - E(X_t))^2 = Var(X_t) \geq 0$
$$\rho_0 = \frac{Cov(X_t, X_{t+0})}{\sqrt{Var(X_t)Var(X_{t+0})}} = \frac{Var(X_t)}{Var(X_t)} = 1$$
2. Iz uslova stacionarnosti sledi da je

$$\gamma_K = Cov(X_t, X_{t+k}) = Cov(X_{t-k}, X_{t+k-k}) = Cov(X_{t-k},, X_t) = \gamma_{-K}$$

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)Var(X_{t+k})}} = \frac{Cov(X_{t-k}, X_{t+k-k})}{\sqrt{Var(X_{t-k})Var(X_{t+k-k})}} = \frac{Cov(X_{t-k}, X_t)}{\sqrt{Var(X_{t-k})Var(X_t)}} = \rho_{-k}$$
3. Na osnovu *Koši -Švarcove* nejednakosti važi
$$|\gamma_K| = |Cov(X_t, X_{t+k})| \leq Var(X_t)^{\frac{1}{2}} Var(X_{t+k})^{\frac{1}{2}} \leq \gamma_0$$
$$|\rho_K| = \frac{|\gamma_K|}{\gamma_0} \leq \frac{\gamma_0}{\gamma_0} = 1$$

Definicija 1.2.2 Funkcija $k : \mathbb{Z} \longrightarrow \mathbb{R}$ je nenegativno definitna ako

$$\sum_{i,j=1}^{n} a_i\, k(t_i - t_j) a_j \geq 0 \tag{1.2.5}$$
$\forall n \in \mathbb{N}$, $\forall a = (a_1, \dots, a_n) \in \mathbb{R}^n$ i $\forall t = (t_1, \dots, t_n)$.

Teorema 1.2.2 Ako je $\{X_t, t \in T\}$ slabo stacionaran proces onda je njegova kovarijansna funkcija nenegativno definitna.

Dokaz. Neka je $\{X_t, t \in T\}$ slabo stacionaran stohastički proces i γ_k njegova autokovarijansna funkcija, $a = (a_1, \dots, a_n)$, i $Z_t = \left(X_{t_1} - E(X_{t_1}), \dots, X_{t_n} - E(X_{t_n}) \right)$ tada

$$0 \leq Var(a^T Z_t) = a^T E(Z_t Z_t^T) = a^T \Gamma_n a = \sum_{i,j=1}^{n} a_i\, \gamma(t_i - t_j) a_j ☺$$

1.3. Konvergencija u verovatnoći i konvergencija u raspodeli

Definicija 1.3.1 Niz slučajnih $X_1, X_2, \dots$ promenljivih konvergira u verovatnoći ka slučajnoj promenljivoj X ako $\forall \varepsilon > 0$

$$P\{| X_n - X| \geq \varepsilon\} \longrightarrow 0, n \longrightarrow \infty. \tag{1.3.1}$$

Definicija 1.3.2 Niz slučajnih $X_1, X_2, \ldots$ promenljivih konvergira u raspodeli ka slučajnoj promenljivoj X kada $n \longrightarrow \infty$ ako niz funkcija raspodele $F_{X_1}(x), F_{X_2}(x), \ldots$ konvergira ka funkciji raspodele $F_X(x)$ $\forall x \in \mathbb{R} \cup \{-\infty, \infty\}$ za koje je $F_X(x)$ neprekidna funkcija.

Konvergencija u verovatnoći se označava sa

$$X_n \xrightarrow{p} X, \tag{1.3.2}$$

a konvergencija u raspodeli

$$X_n \xrightarrow{r} X. \tag{1.3.3}$$

Teorema 1.3.1 Ako niz slučajnih promenljivih konvergira u verovatnoći ka slučajnoj promenljivoj X onda taj niz konvergira i u raspodeli ka slučajnoj promenljivoj X. Obrnuto važi samo u slučaju kada je $X = const.$

Lema 1.3.1 Ako su X_n i $X_n \xrightarrow{r} X$ i $Y_n \xrightarrow{p} c = const.$ tada

$$X_n + Y_n \xrightarrow{r} X + c \tag{1.3.4}$$

$$X_n\, Y_n \xrightarrow{r} cX. \tag{1.3.5}$$

Lema 1.3.2 $X_n \xrightarrow{r} X$ i $Y_n \xrightarrow{p} 0$ tada

$$X_n Y_n \xrightarrow{p} 0. \tag{1.3.6}$$

Teorema 1.3.2 (Zakon velikih brojeva Hinčina) Ako nezavisne slučajne promenljive $X_1, X_2, \ldots$ imaju istu raspodelu i konačno očekivanje $E(X_i) = a, i = 1,2, \ldots$ onda važi

$$\frac{1}{n}\sum_{j=1}^{n} X_i \xrightarrow{p} a. \tag{1.3.7}$$

Teorema 1.3.3 (Centralna granična teorema) Ako nezavisne slučajne promenljive $X_1, X_2, \ldots$ imaju istu raspodelu, $E(X_i) = \mu$ i $Var(X_i) = \sigma^2, i = 1,2, \ldots$ onda za

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

važi

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{r} N(0, \sigma^2). \tag{1.3.8}$$

Napomena. Zakona velikih brojeva Hinčina i centralna granična teorema važe i za niz slučajnih vektora kao i za njihove treasformacije ako je preslikavanje neprekidno.

1.4. Beli šum

Proces $\{\varepsilon_t, t \in T\}$ je proces belog šuma ako su $\varepsilon_1, \varepsilon_{2,\dots}$ nekorelisane slučajne promenljive koje imaju jednaku raspodelu, gde su $E(\varepsilon_i) = 0$ i $Var(\varepsilon_i) = \sigma^2, i = 1,2, \dots$

Tada je kovarijansna funkcija procesa

$$\gamma_i = \begin{cases} \sigma^2 & i = 0 \\ 0 & i \neq 0. \end{cases} \tag{1.4.1}$$

Da bi se beli šum definisao kao stohastički proces u klasičnom smislu potrebno je da se definiše uopšteni stohatički proces. Neka je *K* prostor svih beskonačno diferencijabilnih funkcija $\varphi(t), t \in \mathbb{R}$, koje su identički jednake nule van konačnog intervala. Svaka neprekidna linearna funkcionela Φ definisana na *K* se naziva distribucija ili uopštena funkcija.

Uopšten slučajan stohastički proces je slučajna uopštena funkcija u sledećem smislu: svakom $\varphi \in K$ je dodeljena slučajna promenljiva $\Phi(\varphi)$ tako da važe dva uslova:

- Funkcionela Φ je lienarna na K sa verovatnioćom 1, odnosno za proizvoljne φ i ψ iz K i proizvoljno α i β važi da je

$$\Phi(\alpha\varphi + \beta\psi) = \alpha\Phi(\varphi) + \beta\Phi(\psi) \tag{1.4.2}$$

sa verovatnoćom 1.

- $\Phi(\varphi)$ je neprekidna u sledećem smislu: $\varphi_{k_j} \longrightarrow \varphi_k$ u prostoru K, k=1,2,...,n implicira konveregenciju vektora raspodele $(\Phi(\varphi_{1_j}), \dots, (\Phi(\varphi_{n_j}))$ ka raspodeli $(\Phi(\varphi_1), \dots, (\Phi(\varphi_n))$ u smislu konvregencije u raspodeli.

Jedna od bitnih karakteristika uopštenog stohastičkog procesa je ta što njegov izvod uvek postoji i što je isto uopšten stohastički proces. Izvod uopštenog stohastičkog procesa definiše se sa

$$\dot{\Phi}(\varphi) = -\Phi(\dot{\varphi}). \tag{1.4.3}$$

Klasa uopštenih procesa je šira od klase klasičnih, pa svaki klasični proces može da se posmatra kao uopšteni.

Vinerov proces W_t je stohastički proces sa nezavisnim, stacionarnim i $N(0, t-s)$ raspodeljenim priraštajima $W_t - W_s$, početnom vrednošću $W_0 = 0$ i skoro sigurno neprekidnim trajektorijama. Ukoliko se Vinerov proces posmatra kao uopšteni proces, može da se pokaže da je kovarijansna funkcija izvoda Vinerovog procesa Dirakova delta distribucija $\delta(t)$, a to je upravo kovarijansna funkcija belog šuma. Dakle, u smislu poklapanja kovarijansi beli šum je izvod Braunovog kretanja kada se Braunovo kretanje posmatra kao uopšteni stohastički proces.

Nadalje pod procesom belog šuma se podrazumeva familija slučajnih promenljivih, koje su nekorelisane, sa konačnom varijansom i očekivanjem nula.

1.5. Vremenske serije

Definicija 1.5.1 Vremenska serija je jedna realizacija stohastičkog procesa.

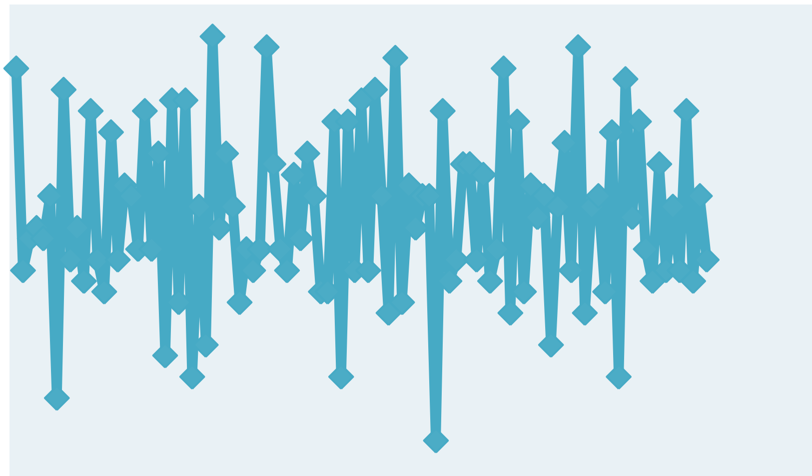
Neka je data jedna realizacija stohastičkog procesa $(x_1, \dots, x_n)$, odnosno vremenska serija. Da bi se odredile neke od osnovnih karakteristika stohastičkog procesa, kao što su srednja vrednost, varijansa i kovarijansa, potrebno je više od jedne njegove realizacije. Čak i u slučaju kada je dostupno beskonačno mnogo vrednosti te iste realizacije postavlja se pitanje je da li je ona dovoljna da bi se proces objasnio u celini.

Zahtev da je vremenska serija ergodična znači da uzorački momenti koji su izračunati na osnovu jedne realizacije srednje kvadratno konvergiraju ka odgovarajućim momentima populacije kada $n \longrightarrow \infty$. U analizi vremenskih serija je od najvećeg značaja da momenti prvog i drugog reda za jednu realizaciju procesa konvergiraju ka odgovarajućim prvim i drugim momentima populacije stohastičkog procesa kada se obim uzorka povećava, odnosno da je ispunjeno

$$lim_{n\to\infty} E\left[(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu)^2\right] = 0, \tag{1.5.1}$$

$$lim_{n\to\infty} E\left[\left(\frac{1}{n}\sum_{i=1}^{n} (x_t - \mu)^2 - \sigma^2\right)^2\right] = 0. \tag{1.5.2}$$

Najbitnija statistička osobina vremenskih serija jeste njihova stacionarnost. Za stacionarne vremenske serije mogu tačno da se odrede njihovi prvi i drugi momenti kao i autokovarijansna funkcija, što omogućava lakše i preciznije predviđanje budućih vrednosti vremenskih serija. Na slici 1 i slici 2 su dati grafici stacionarne i nestacionarne vremenske serije respektivno.



Slika 1 Stacionarna vremenska serija

Slika2 Nestacionarna vremenska serija

Trend vremenske serije opisuje njeno osnovno ponašanje u dužem vremenskom periodu, ili preciznije u čitavom vremenskom periodu za koji je vremenska serija poznata. Trend vremenske serije može da bude deterministički i stohastički. U slučaju kada je trend detreministički postoji matematička funkcija vremena kojom trend može da se opiše. U zavisnosti od izbora funkcije postoje različiti modeli trenda vremenske serije: linearni, parabolični, eksponencijalni, logistički... Ukoliko je trend vremenske serije stohastički tada se on menja pod uticajem slučajnih faktora, odnosno ponašanje vremenske serije ne može da se predvidi i ima stohastički karakter.

Za ocenu teorijskih pojmova se koriste sledeće uzoračke ocene:

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t \tag{1.5.3}$$

$$\widehat{Var}\,(X_t) = \frac{1}{n}\sum_{i=1}^{n}(X_t - \bar{X})^2 \tag{1.5.4}$$

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=k+1}^{n}(X_t - \bar{X})(\,X_t - \bar{X}) \tag{1.5.5}$$

$$\hat{\rho}_K = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k+1}^{n}(X_t - \bar{X})(\,X_{t-k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}. \tag{1.5.6}$$

Za uzoračke ocena $\hat{\gamma}_k$ i $\hat{\rho}_K$ je zadovoljen uslov da je matrica

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{n-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{n-1} & \hat{\gamma}_{n-2} & \cdots & \hat{\gamma}_0 \end{bmatrix} \tag{1.5.7}$$

nenegativno definitna.

Ako se $\hat{\Gamma}_n$ predstavi u obliku

$$\hat{\Gamma}_n = \frac{1}{n} N N^T \tag{1.5.8}$$

gde je N matrica $n \times 2n$ sa elementima

$$N = \begin{bmatrix} 0 & 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n \\ 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & Y_1 & Y_2 & \cdots & Y_n & 0 & \dots & 0 \end{bmatrix} \tag{1.5.9}$$

$$Y_i = X_i - \bar{X}\ ,\ i = 1,2,\dots,n$$

tada za proizvoljan vektor $a \in \mathbb{R}^n$ važi

$$a^T\widehat{\Gamma_n}a = \frac{1}{n}(a^TN)(a^TN)^T \geq 0. \qquad (1.5.10)$$

Kako postoji mnogo parova (X_t, X_{t-k}) *Box* i *Jenkins* su predložili da se u praksi ne izračunava više od $n/4$ autokorelacionih koeficijenata gde je n obim uzorka.

2. Linearni procesi i linearne vremenske serije

2.1. Linearni procesi

U statističkom modeliranju jedan od najbitnijih zadataka jeste pronalaženje odgovarajuće funkcionalne veze između određenih ulaznih i izlaznih informacija. Jedna od najčešćih pretpostavki je da je ta veza linearna. U analizi vremenskih serija linearni filter je operator koji transformiše vremensku seriju $\{X_t\}_{t\in\mathbb{Z}}$ u vremensku seriju $\{Y_t\}_{t\in\mathbb{Z}}$

$$Y_t = L(X_t) = \sum_{j=-\infty}^{\infty} \omega_j X_{t-j} \tag{2.1.1}$$

gde su koeficijenti ω_j vremenski invarijantni. Definicija linearnog filtera implicira da vrednost vremenske serije $\{Y_t\}_{t\in\mathbb{Z}}$ u trenutku *t* zavisi od sopstvenih prošlih i budućih vrednosti. Kako su u praksi na raspolaganju istorijski podaci, uvodi se pretpostavka da je $j \geq 0$, pa se dolazi do definicije linearnog filtera

$$Y_t = L(X_t) = \sum_{j=0}^{\infty} \omega_j X_{t-j} \tag{2.1.2}$$

gde vrednost $\{Y_t\}_{t\in\mathbb{Z}}$ u trenutku *t* zavisi samo od sopstvenih prošlih vrednosti.

Ukoliko suma koeficienata apsolutno konvergira

$$\sum_{j=-\infty}^{\infty} |\omega_j| < \infty \tag{2.1.3}$$

kaže se da su stabilni.

Definicija 2.1.1 Proces $\{X_t, t \in T\}$ je linearan ako može da se predstavi u obliku

$$X_t = \mu_t + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \tag{2.1.4}$$

gde su

- μ_t deterministička komponenta
- $\{\varepsilon_t, t \in T\}$ proces belog šuma
- θ_1, θ_2 ... nepoznati parametri.

Osnovna odlika determinističke komponente je ta da ona može da se aproksimira sa nekom matematičkom funkcijom, što u situaciji prognoziranja vremenske serije znači da će izabrani tip funkcije važiti i u budućem periodu.

Suma

$$\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \tag{2.1.5}$$

predstavlja stohastičku komponentu procesa, i ona opisuje dejstvo slučajnih faktora u modelu. Osnovna pretpostavka je da stohastička komponenta može da se modelira pomoću procesa belog šuma.

Očekivanje i varijansa lineranog procesa su

$$E(X_t) = E(\mu_t) + \sum_{j=0}^{\infty} \theta_j E(\varepsilon_{t-j}) = \mu_t + 0 = \mu_t \tag{2.1.6}$$

$$\begin{aligned} Var(X_t) &= Var(\mu_t) + \sum_{j=0}^{\infty} \theta_j^2 Var(\varepsilon_{t-j}) = 0 + \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \\ &= \sigma^2 \sum_{j=0}^{\infty} \theta_j^2. \end{aligned} \tag{2.1.7}$$

Momenti prvog i drugog reda su konačni u slučaju kada je

$$\mu_t \equiv \mu = const. \tag{2.1.8}$$

$$\sum_{j=0}^{\infty} \theta_j^2 < \infty. \tag{2.1.9}$$

Tada su kovarijansna i korelaciona funkcija

$$\begin{aligned}\gamma_k = Cov\,(X_t, X_{t-k}) = E\big((X_t-\mu)(X_{t-k}-\mu)\big) \\ = E\big((\varepsilon_t+\theta_1\varepsilon_{t-1}+..+\theta_k\varepsilon_{t-k}+\theta_{k+1}\varepsilon_{t-k-1} \\ +\cdots)(\varepsilon_{t-k}+\theta_1\varepsilon_{t-k-1}+\cdots)\big) \\ = \sigma^2(\theta_k+\theta_1\theta_{k+1}+\theta_2\theta_{k+2}+\cdots) \\ = \sigma^2\sum_{j=0}^{\infty}\theta_{j+k}\,\theta_j\end{aligned} \tag{2.1.10}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sigma^2\sum_{j=0}^{\infty}\theta_{j+k}\,\theta_j}{\sum_{j=0}^{\infty}\theta_j^2}. \tag{2.1.11}$$

Ako su ispunjeni uslovi definisani sa 2.1.8 i 2.1.9 tada kovarijansna funkcija zavisi samo od k pa je linearan proces slabo stacionaran. Ukoliko se dodatno pretpostavi da je beli šum Gausovski tada je funkcija raspodele stohastičkog procesa potpuno određena sa prvim i drugim momentima procesa i proces je strogo stacionaran.

Uslov 2.1.9 može da se proveri ispitivanjem apsolutne konvergenicje nepoznatih parametara jer važi

$$\sum_{j=0}^{\infty}\theta_j^2 \le \sum_{j=0}^{\infty}|\theta_j| \tag{2.1.12}$$

pa uslovi slabe stacionarnosti postaju

$$\mu_t \equiv \mu = const. \tag{2.1.13}$$

$$\theta_j, j \ge 0 \text{ su stabilni.} \tag{2.1.14}$$

U slučaju kada važi 2.1.8 onda linearni proces ekvivalentno može da se zapiše u obliku

$$\begin{gathered}X_t - \mu = X_t^* = \sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j} \\ E(X_t^*) = 0.\end{gathered} \tag{2.1.15}$$

1938. godine *Wold* je dokazao teoremu koja može da se smatra egzistencijalnom. On je smatrao da je klasa linearnih procesa dovoljno opšta da obuhvati sve slabo stacionarne vremenske serije.

Teorema 2.1.1 (Woldova dekompozicija) Proces $\{ X_t, t \in T\}$ može da se predstavi kao zbir

$$X_t = V_t + \sum_{j=1}^{\infty} \chi_t \, \varepsilon_{t-j} \tag{2.1.16}$$

gde su

- V_t deterministička komponenta
- $\{ \varepsilon_t, t \in T\}$ je proces belog šuma
- $\sum_{j=0}^{\infty} \chi_j^2 < \infty, \chi_0 = 1$
- $E(\varepsilon_t, V_t) = 0$.

Definicija 2.1.2 Linearan proces $\{ X_t, t \in T\}$ za koji važi da je $E(X_t) = \mu, \forall t \in T$ je invertibilan ako postoje $\varphi_1, \varphi_2, \ldots$ takvi da je

$$\begin{gathered} \varepsilon_t = \sum_{j=0}^{\infty} \varphi_j (X_{t-j} - \mu) = \sum_{j=0}^{\infty} \varphi_j X_{t-j}^* \\ \sum_{j=0}^{\infty} |\varphi_j| < \infty. \end{gathered} \tag{2.1.17}$$

Definicija 2.1.3 Linearna vremenska serija je jedna realizacija linearnog stohastičkog procesa.

Kako u do sada definisanim modelima linearnih vremenskih serija figuriše beskonačan broj parametara, cilj je da se njihov broj ograniči na konačno mnogo. Postoje tri grupe linearnih procesa koje imaju konačan broj parametara:

- autoregresivni proces reda *p* (*AR (p)*)
- proces pokretnih sredina reda *q* (*MA (q))*
- proces *ARMA(p,q)* je kombinacija *AR(p)* i *MA(q)* procesa.

2.2. Operatori u analizi vremenskih serija

U literaturi koja je vezana za analizu vremenskih serija često se koristi L (*lag*) operator. Neka je dat proizvoljan proces $\{ Z_t, t \in T\}$. L operator transformiše proces iz sadašnjeg vremenskog trenutka u isti proces u prethodnom vremenskom trenutki

$$LZ_t = Z_{t-1}. \tag{2.2.1}$$

Koristeći 2.2.1 za proces $\{ Z_t, t \in T\}$ važi

$$\begin{gathered} Z_{t-2} = LZ_{t-1} = LLZ_t = L^2 Z_t \\ Z_{t-3} = LZ_{t-2} = LL^2 Z_t = L^3 Z_t \\ \vdots \\ Z_{t-m} = LZ_{t-(m-1)} = LL^{m-1} Z_t = L^m Z_t. \end{gathered} \tag{2.2.2}$$

Linearan proces definisan sa formulom 2.1.15 pomoću L operatora može da se predstavi u obliku

$$X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \theta_j L^j \varepsilon_t = \theta(L)\varepsilon_t. \tag{2.2.3}$$

Ako je proces i invertibilan tada se koristeći L operator dobija

$$\varepsilon_t = \sum_{j=0}^{\infty} \varphi_j X_{t-j} = \sum_{j=0}^{\infty} \varphi_j L^j X_t = \varphi(L) X_t. \tag{2.2.4}$$

Još jedan od operatora koji se često koristi jeste diferencni operator koji se definiše sa

$$\nabla X_t = X_t - X_{t-1} = X_t - LX_t = (1-L)X_t. \tag{2.2.5}$$

Diferencni operator drugog reda je

$$\begin{aligned} \nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} = X_t - 2LX_t + L^2 X_t \\ &= (1 - 2L + L^2)X_t = (1-L)^2 X_t, \end{aligned} \tag{2.2.6}$$

a diferencni operator reda *d* je

$$\nabla^d X_t = \nabla(\nabla^{d-1})X_t = \cdots = (1-L)^d X_t. \tag{2.2.7}$$

2.3. Autoregresivni proces AR(p)

Definicija 2.3.1 $\{X_t, t \in T\}$ je autoregresivni proces reda *p* ako može da se predstavi u obliku

$$X_t = \mu + \sum_{j=1}^{p} \varphi_j X_{t-j} + \varepsilon_t \tag{2.3.1}$$

gde su

- μ konstanta
- $\{\varepsilon_t, t \in T\}$ proces belog šuma
- $\varphi_j, j = 1, \dots, p$ nepoznati parametri.

Kako je μ konstanta na osnovu 2.1.15 u zapisu *AR (p)* ona može da se izostavi

$$X_t = \sum_{j=1}^{p} \varphi_j X_{t-j} + \varepsilon_t. \tag{2.3.2}$$

Koristeći *L* operator *AR(p)* proces može da se predstavi

$$\varphi(L) X_t = \varepsilon_t \tag{2.3.3}$$

gde je $\varphi(L) = 1 - \varphi_1 L - \cdots - \varphi_p L^p$ karakteristični polinom *AR(p)* procesa.

Jednačina $\varphi(z) = 0$, $z \in \mathbb{C}$ karakteristična jednačina *AR (p)* procesa. Rešenja karakteristične jednačine su karakteristični koreni *AR (p)* procesa.

Neka je dat *AR (1)* proces

$$(1 - \varphi L) X_t = \varepsilon_t \tag{2.3.4}$$

$$X_t = (1 - \varphi L)^{-1} \varepsilon_t. \tag{2.3.5}$$

Za $|\varphi| < 1$ je

$$(1 - \varphi L)^{-1} = \frac{1}{1 - \varphi L} = \sum_{i=1}^{\infty} (\varphi L)^i. \tag{2.3.6}$$

S druge strane karakteristična jednačina *AR (1)* procesa je

$$1-\varphi z=0, \tag{2.3.7}$$

a njen karakteristični koren

$$z=\frac{1}{\varphi}=\varphi^{-1}. \tag{2.3.8}$$

Kako je $|z|=\frac{1}{|\varphi|}$ u slučaju kada je a $|\varphi|<1$ onda je $|z|>1$.

Dakle 2.3.6 važi u slučaju kada je

$$|\varphi|<1 \Leftrightarrow |z|>1. \tag{2.3.9}$$

Lema 2.3.1 (Fundamentalna teorema algebre) Ako je $\alpha(x)$ polinom *p*-tog reda onda postoji faktorizacija

$$\varphi(x)=(1-{x_1}^{-1}x)(1-{x_2}^{-1}x)\ldots(1-{x_p}^{-1}x) \tag{2.3.10}$$

gde su $x_1,x_2,\ldots,x_p$ nule polinoma $\alpha(x)$.

Uslovi stacionarnosti autoregresivnog procesa su definisani sledećom teoremom.

Teorema 2.3.1 Neka su $z_1,z_2,\ldots,z_p$ karakteristični koreni autoregresivni proces reda p. Ako $\forall z_i\in\mathbb{C}$ važi $|z_i|>1, i=1,2,\ldots,p$ onda je AR(p) proces slabo stacionaran.

Dokaz . Predstavićemo AR(p) proces u obliku

$$X_t=\frac{1}{\varphi(L)}\varepsilon_t=(\varphi(L))^{-1}\varepsilon_t \tag{2.3.11}$$

Prema lemi 2.3.1 važi

$$\varphi(z)=(1-{z_1}^{-1}z)(1-{z_2}^{-1}z)\ldots(1-{z_p}^{-1}z) \tag{2.3.12}$$

pa je

$$\begin{aligned}\big(\varphi(z)\big)^{-1}=\frac{1}{\varphi(z)}&=\frac{1}{(1-{z_1}^{-1}z)(1-{z_2}^{-1}z)\ldots\big(1-{z_p}^{-1}z\big)}\\&=\frac{c_1}{(1-{z_1}^{-1}z)}+\cdots+\frac{c_p}{(1-{z_p}^{-1}z)}\end{aligned} \tag{2.3.13}$$

$$X_t = \frac{c_1 \varepsilon_t}{(1 - {z_1}^{-1} z)} + \cdots + \frac{c_p \varepsilon_t}{(1 - {z_p}^{-1} z)}. \tag{2.3.14}$$

Iz 2.3.6 sledi

$$\begin{aligned} \frac{c_j \varepsilon_t}{\left(1 - {z_j}^{-1} z\right)} &= c_j \varepsilon_t \left(1 + \sum_{i=0}^{\infty} (({z_j}^{-1} z)^i \right) \\ &= c_j \left(\varepsilon_t + \sum_{i=1}^{\infty} ({z_j}^{-i} \varepsilon_{t-i} \right), j = 1,2, \dots, p \Rightarrow \end{aligned} \tag{2.3.15}$$

$$X_t = \sum_{j=1}^{p} c_j \sum_{i=0}^{\infty} ({z_i}^{-1} z)^i \, \varepsilon_t = \sum_{i=0}^{\infty} \sum_{j=1}^{p} (c_j {z_j}^{-i}) \, \varepsilon_{t-i} \quad \Longrightarrow \tag{2.3.16}$$

$$X_t = \sum_{i=0}^{\infty} \alpha_i \, \varepsilon_{t-i} \, . \tag{2.3.17}$$

Prema uslovima slabe stacionarnosti linearnih procesa potrebno je da red

$$\sum_{i=0}^{\infty} {\alpha_i}^2 = \sum_{i=0}^{\infty} \left(\sum_{j=1}^{p} (c_j {z_j}^{-i}) \right)^2 \tag{2.3.18}$$

konvergira.

Kako je

$$\lim_{i \to \infty} {\alpha_i}^2 = \lim_{i \to \infty} (c_1 {z_1}^{-i} + \cdots + c_p {z_p}^{-i})^2 = \begin{cases} 0 & |z_i| > 1 \\ \infty & |z_i| \leq 1 \end{cases} \tag{2.3.19}$$

$\Longrightarrow$ za $|z_i| > 1, \; i = 1,2, \dots, p$ red konvergira $\Rightarrow$ *AR (p)* proces je stacionaran. ☺

Primedba 1. Na osnovu 2.3.8 i 2.3.9 *AR (p)* proces je slabo stacionaran ako je $|\varphi_i| < 1, \forall i = 1,2, \dots, p$.

Primedba 2. Ako je *AR(p)* slabo stacionaran onda na osnovu 2.3.17 može da se predstavi kao linearna kombinacija procesa belog šuma sa beskonačnim brojem parametara. Tada se kaže da autoregresivni proces reda *p* ima *MA(*∞) reperezentaciju.

AR (p) proces uvek invertibilan jer je

$$\sum_{j=0}^{\infty}|\varphi_j| = \sum_{j=0}^{p}|\varphi_j| < \infty. \tag{2.3.20}$$

U slučaju kada je *AR(p)* slabo stacionaran, mogu da se odrede $E(X_t)$, $Var\ (X_t)$, γ_k i ρ_k .

$$E(X_t) = E(X_{t-1}) = \cdots = E(X_{t-p}) \tag{2.3.21}$$

$$E(X_t) = \sum_{j=1}^{p}\varphi_j E(X_t) \Rightarrow (1-\sum_{j=1}^{p}\varphi_j)E(X_t) = 1 \Rightarrow$$
$$E(X_t) = \frac{1}{1-\sum_{j=1}^{p}\varphi_j} \tag{2.3.22}$$

$$\begin{aligned}\gamma_k = Cov(X_t, X_{t-k}) = Cov\left(\sum_{i=0}^{p}\varphi_i X_{t-i}, X_{t-k}\right) &= \\ = \sum_{i=0}^{p}\varphi_i Cov(X_{t-i}, X_{t-k}) + Cov(\varepsilon_t, X_{t-k}) & \\ = \sum_{i=0}^{p}\varphi_i\gamma_{k-i} + \begin{cases}\sigma^2 & k=0\\ 0 & k>0\end{cases} &\end{aligned} \tag{2.3.23}$$

$k = 0$

$$\begin{aligned} Var\ (X_t) = \gamma_0 = \sigma^2 + \sum_{i=0}^{p}\varphi_i\gamma_{-i} = \sigma^2 + \sum_{i=0}^{p}\varphi_i\gamma_i \\ \Rightarrow \gamma_0 - \sum_{i=0}^{p}\varphi_i\gamma_i = \sigma^2 \\ \Rightarrow \gamma_0\left(1-\sum_{i=0}^{p}\varphi_i\rho_i\right) = \sigma^2 \\ \Rightarrow Var\ (X_t) = \frac{\sigma^2}{1-\varphi_1\rho_1 - \cdots - \varphi_p\rho_p}\end{aligned} \tag{2.3.24}$$

$k > 0$

$$\begin{aligned} Var\,(X_t) = \gamma_0 = \sigma^2 + \sum_{i=0}^{p} \varphi_i \gamma_{-i} = \sigma^2 + \sum_{i=0}^{p} \varphi_i \gamma_i \\ \Rightarrow \gamma_0 - \sum_{i=0}^{p} \varphi_i \gamma_i = \sigma^2 \\ \Rightarrow \gamma_0 \left(1 - \sum_{i=0}^{p} \varphi_i \rho_i \right) = \sigma^2 \\ \Rightarrow Var\,(X_t) = \frac{\sigma^2}{1 - \varphi_1 \rho_1 - \cdots - \varphi_p \rho_p} \end{aligned} \tag{2.3.25}$$

$$\begin{aligned} \gamma_k = Cov(X_t, X_{t-k}) = \sum_{i=1}^{p} \varphi_i \gamma_{k-i} \\ = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \cdots + \varphi_p \gamma_{k-p} \end{aligned} \tag{2.3.26}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p}. \tag{2.3.27}$$

Ukoliko se sistemi jednačina koji određuju kovarijansnu i korelacionu autoregresivnog procesa zapišu u matričnom obliku dobija se

$$\boldsymbol{\gamma} = \Gamma_p \boldsymbol{\varphi}$$

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix} \; \Gamma_p = \begin{bmatrix} \gamma_0 & \dots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \dots & \gamma_0 \end{bmatrix} \; \boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{bmatrix} \tag{2.3.28}$$

$$\boldsymbol{\rho} = R_p \boldsymbol{\varphi} \tag{2.3.29}$$

$$\boldsymbol{\rho} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_p \end{bmatrix} \; R_p = \begin{bmatrix} 1 & \dots & \rho_{p-1} \\ \vdots & \ddots & \vdots \\ \rho_{p-1} & \dots & 1 \end{bmatrix} \; \boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{bmatrix}$$

Sistem definisan sa 2.3.29 se zove *Yule-Walkers*-ov sistem jednačina, a

$$\widehat{\boldsymbol{\varphi}} = \widehat{R}_p^{-1} \widehat{\boldsymbol{\rho}}_p \tag{2.3.30}$$

Yule-Walkers-ova ocena za nepoznate parametre u modelu. Takođe može da se oceni i nepoznata varijansa procesa belog šuma

$$\sigma^2 = \gamma_0 \big(1 - \widehat{\boldsymbol{\rho}}_p^T \widehat{R}_p^{-1} \widehat{\boldsymbol{\rho}}_p \big). \tag{2.3.31}$$

Autokorelaciona funkcija *AR (p)* zadovoljava linearnu diferencnu jednačinu pa je njeno opšte rešenje

$$\rho_k = A_1(z_1^{-1})^k + \cdots + A_p(z_p^{-1})^k \tag{2.3.32}$$

i tada mogu da se razlikuju dva slučaja:

1) z_i je realan broj, tada će $A_i(z_i^{-1})^k \to 0$ nuli kada $k \to \infty$.
2) z_i i z_j su konjugovano-kopleksni brojevi tada se kaže da autokorelaciona funkcija prati prigušeni sinusni talas, odnosno talas čija se amplituda smanjuje u toku vremena.

Ponašanje korelacione funkcije zavisi od karakterističnih korena, ali kako se u oba slučaja ona približava ka nuli njeno ponašanje može da se okarakteriše kao lagano odumiranje ka nuli.

Da bi se ocenili nepoznati parametri pomoću metode najmanjih kvadrata na osnovu uzorka obima n, mora da se izostavi prvih *p* elemenata uzorka. Kako je *n* uglavnom veliko, a *p* uglavnom malo, broj izbačenih elementa uzorka će biti jako mali. Ako se modifikovani autoregresivni proces predstavi u matričnom obliku

$$Y = X\varphi + \varepsilon$$

$$Y = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{bmatrix}, X = \begin{bmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_n & X_{n-1} & \dots & X_{n-p} \end{bmatrix} \varepsilon = \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.3.33}$$

onda je ocena dobijena metodom najmanjih kvadrata

$$\varphi^* = (X^T X)^{-1} X^T Y. \tag{2.3.34}$$

Teorema 2.3.2 Ako je Ako je φ^* ocena dobijena metodom najmanjih kvadrata onda je ona konzistentna.

Dokaz.

$$\varphi^* = (X^T X)^{-1} X^T Y = \left(\sum_{i=p+1}^{n} X_i X_i^T \right)^{-1} \left(\sum_{i=p+1}^{n} X_i Y_i \right) \tag{2.3.35}$$

$$\varphi^* - \varphi = \left(\frac{1}{n-p} \sum_{i=p+1}^{n} X_i X_i^T \right)^{-1} \left(\frac{1}{n-p} \sum_{i=p+1}^{n} X_i \varepsilon_i \right). \tag{2.3.36}$$

Kada je $\{ X_t, t \in T\}$ stacionaran proces tada na osnovu zakona velikih brojeva Hinčina

$$\frac{1}{n-p}\sum_{i=p+1}^{n} X_i \xrightarrow{p} E(X_i). \tag{2.3.37}$$

Kako za svaku transformaciju slučajnih vektora ukoliko je preslikavanje neprekidno takođe važi zakon velikih brojeva Hinčina dobija se

$$\frac{1}{n-p}\sum_{i=p+1}^{n} X_i X_i^T \xrightarrow{p} E(X_i X_i^T) \tag{2.3.38}$$

$$\frac{1}{n-p}\sum_{i=p+1}^{n} X_i \,\varepsilon_i \xrightarrow{p} E(X_i\varepsilon_i) = 0 \tag{2.3.39}$$

tada na osnovu 2.3.36 i leme 1.3.2

$$\varphi^* - \varphi \xrightarrow{p} 0$$

$\Rightarrow \varphi^*$ konzistentna ocena parametra φ.

Teorema 2.3.3 Ako je φ^* ocena dobijena metodom najmanjih kvadrata onda je

$$\sqrt{n}(\ \varphi^* - \ \varphi) \colon AN(0, \sigma^2\Gamma_p^{-1}) \tag{2.3.40}$$

Dokaz.

$$\begin{aligned}\sqrt{n}(\ \varphi^* - \ \varphi) &= n(X^T X)^{-1}\left(\frac{1}{\sqrt{n}}X^T\varepsilon\right)\\ &= \left(\frac{1}{n}\sum_{i=p+1}^{n} X_i X_i^T\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=p+1}^{n} X_i\,\varepsilon_i\right)\end{aligned} \tag{2.3.41}$$

Neka je $U_i = X_i\varepsilon_i$ tada je

$$E(U_i) = X_i E(\varepsilon_i) = 0 \tag{2.3.42}$$
$$E(U_i U_i^T) = E(X_i X_i^T \varepsilon_i^2) = E(X_i X_i^T)E(\varepsilon_i^2) = \sigma^2\Gamma_p. \tag{2.3.43}$$

Na osnovu centralne granične teoreme se dobija

$$\frac{1}{\sqrt{n}}\sum_{i=p+1}^{n} X_i\,\varepsilon_i \xrightarrow{r} N\left(0,\sigma^2\Gamma_p\right) \tag{2.3.44}$$

pa je zbog 2.3.38 , i 1.3.6

$$\sqrt{n}(\ \varphi^* -\ \varphi) = \left(\frac{1}{n}\sum_{i=p+1}^{n} X_i\,X_i^T\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=p+1}^{n} X_i\,\varepsilon_i\right) \xrightarrow{r} \Gamma_p^{-1}N\left(0,\sigma^2\Gamma_p\right) = N\left(0,\sigma^2\Gamma_p^{-1}\right)$$

$$\sqrt{n}(\ \varphi^* -\ \varphi) : AN\left(0,\sigma^2\Gamma_p^{-1}\right). \tag{2.3.45}$$

Može da se pokaže da ocena dobijena metodom najmanjih kvadrata i *Yule-Walkers*-ova ocena imaju ista asimptotska svojstva

$$\sqrt{n}(\ \hat{\varphi} -\ \varphi) : AN\left(0,\sigma^2\Gamma_p^{-1}\right).$$

2.4. Parcijalna korelacija

U statistici je čest slučaj da je korelacija između dve promenljive ustvari rezultat njihove koreliranosti sa trećom promenljivom u modelu. Zbog toga se uvodi pojam parcijalne korelacije koja je predstavlja korelaciju između dve promenljive uz eliminisan uticaj drugih promenljivih iz modela. Po definiciji parcijalni koeficijent korelacije između X_t i X_{t-k}, u oznaci φ_{kk} je *k*-ti regresioni koeficijent u autoregresivnom procesu reda *p*=*k*

$$X_t = \varphi_{k1}X_{t-1} + \varphi_{k2}X_{t-2} + \cdots + \varphi_{kk}X_{t-k} + \varepsilon_t. \tag{2.4.1}$$

Ako se model definisan formulom 2.4.1 predstavimi u obliku *Yule-Walkers*-ovg sistema tada je

$$\begin{bmatrix}\rho_1\\ \vdots \\ \rho_k\end{bmatrix} = \begin{bmatrix}1 & \dots & \rho_{k-1}\\ \vdots & \ddots & \vdots\\ \rho_{k-1} & \dots & 1\end{bmatrix}\begin{bmatrix}\varphi_{k1}\\ \vdots \\ \varphi_{kk}\end{bmatrix}. \tag{2.4.2}$$

Pomoću Kramerovog pravila može da se odredi $k-$ ti parcijalni korelacioni koeficijent

$$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \dots & \rho_1 \\ \vdots & \ddots & \vdots \\ \rho_{k-1} & \dots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \dots & \rho_{k-1} \\ \vdots & \ddots & \vdots \\ \rho_{k-1} & \dots & 1 \end{vmatrix}}. \tag{2.4.3}$$

Da bi se skratio glomazan postupak rešavanja , *Durbin* je predložio efikasan metod za dobijanje rešenja sistema, a time i za dobijanje uzoračke parcijalne korelacione funkcije. *Darbin*-ova rekurzivna formula je

$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k} \varphi_{k-1}\rho_{k-j}}{1 - \sum_{j=1}^{k} \varphi_{k-1}\rho_{k-j}}$$

$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-j}. \tag{2.4.4}$$

Ako je φ_{kk} koeficijent parcijalne autokorelacije onda važi

$$\sqrt{n}\varphi_{kk} \sim AN(0,1) \tag{2.4.5}$$

pa može da se odredi o intreval poverenja na nivou α za parcijalne korelacione koeficijente

$$\alpha = P\{-c \le \sqrt{n}\varphi_{kk} \le c\}$$

$$\alpha = P\left\{-\frac{c}{\sqrt{n}} \le \varphi_{kk} \le \frac{c}{\sqrt{n}}\right\} \tag{2.4.5}$$

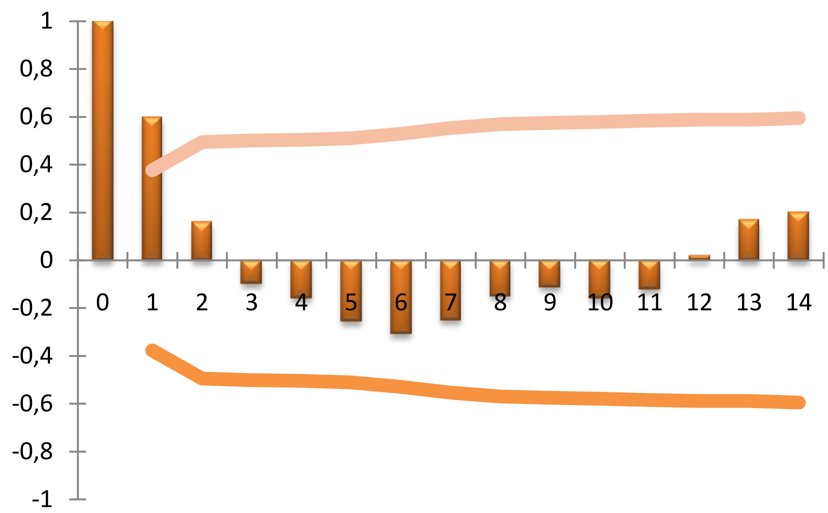
$$(-\frac{c}{\sqrt{n}}, \frac{c}{\sqrt{n}})$$

gde je c kvantil reda $\frac{1+\alpha}{2}$ $N(0,1)$ raspodele.

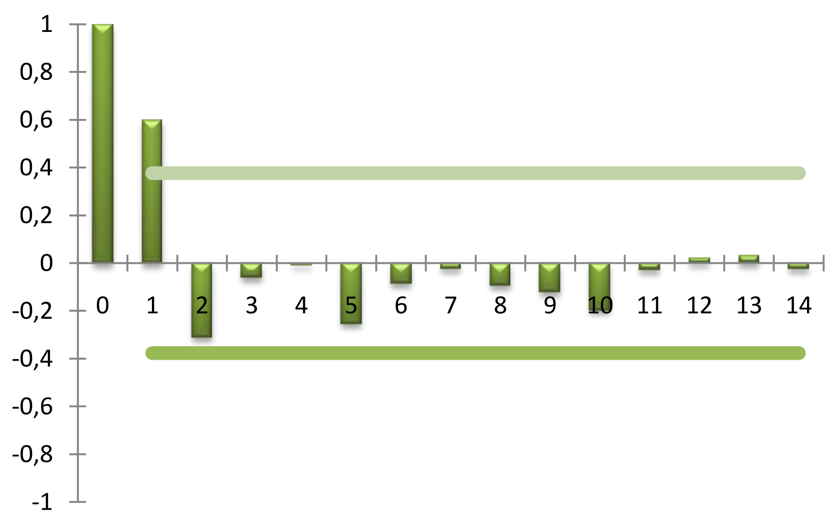
Kako je najčešće $\alpha = 95\%$ intreval poverenja za φ_{kk} je

$$\left(-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right). \tag{2.4.6}$$

Primer 1 Korelaciona i parcijalna korelaziona funkcija *AR (1)* procesa



Slika 3 Korelaciona funkcija AR (1) procesa



Slika 4 Parcijalna korelaciona funkcija AR(1) procesa

2.5. Proces pokretnih sredina MA (q)

Definicija 2.5.1 $\{X_t, t \in T\}$ je proces pokretnih sredina reda q ako je

$$X_t = \mu + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{2.5.1}$$

gde su

- $\mu = E(X_t)$
- $\{\varepsilon_t, t \in T\}$ proces belog šuma
- $\theta_1, \dots, \theta_q$ su nepoznati prametri.

Proces pokretnih sredina reda q je linearni proces sa konačnim očekivanjem i konačnim brojem parametara pa je zbog toga

$$E(X_t) = \mu \tag{2.5.2}$$

$$Var(X_t) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \tag{2.5.3}$$

$$\gamma_k = \begin{cases} 0 & k > q \\ \sigma^2 \sum_{j=0}^{q-k} \theta_j\, \theta_{j+k} & 1 \le k \le q \end{cases} \tag{2.5.4}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 0 & k > q \\ \dfrac{\sum_{j=0}^{q-k} \theta_j\, \theta_{j+k}}{\sum_{j=0}^{q} \theta_j^2} & 1 \le k \le q. \end{cases} \tag{2.5.5}$$

Kako važi

$$\begin{gathered} E(X_t) < \infty \\ Var(X_t < \infty) \\ \gamma_k = f(k) \end{gathered} \tag{2.5.6}$$

proces pokretnih sredina reda q je uvek slabo stacionaran.

Proces pokretnih sredina može da se predstavi pomoću L operatora

$$X_t = \mu + (1 - \theta_1 L - \cdots - \theta_1 L^q)\varepsilon_t = \mu + \theta(L)\varepsilon_t \tag{2.5.7}$$

gde je $\theta(L)$ je karakteristični polinom , a $\theta(z) = 0, z \in \mathbb{C}$ je karakteristična jednačina *MA (q)* procesa.

Ako je proces invertibilan to znači da ε_t može da se izrazi preko beskonačne sume X_t odnosno

$$\begin{gathered} \varepsilon_t = \theta(L)^{-1}(X_t - \mu) = \theta(L)^{-1}X_t - \theta(L)^{-1}\mu \\ = \left(1 - \sum_{i=1}^{\infty} \pi_i L^i\right) X_t - \delta \Longrightarrow \\ \varepsilon_t + \delta = \left(1 - \sum_{i=1}^{\infty} \pi_i L^i\right) X_t \end{gathered} \tag{2.5.8}$$

gde je

$$\sum_{i=1}^{\infty} |\pi_i| < \infty. \tag{2.5.9}$$

Uslovi pod kojima je proces pokretnih sredina invertibilan su definisani sledećom teoremom.

Teorema 2.5.1 Proces pokretnih sredina reda q je invertibilan ako su svi koreni njegove karakteristične jednačine po modulu veći od jedinice ili ekvivalentno ako su svi parametri $\theta_i, i = 1,2,..q$ po modulu manji od jedinice.

Dokaz . Kako je

$$X_t = \theta(L)\varepsilon_t \tag{2.5.10}$$

onda je

$$\varepsilon_t = \theta(L)^{-1}X_t \tag{2.5.11}$$

pa dokaz da karakteristični koreni moraju po modulu da budu veći od jedinice ide isto kao i kod dokaza za uslove stacionarnosti kod autoregresivnih modela. ☺

Ukoliko je proces invretibilan tada se kaže da on ima AR(∞) reprezentaciju.

Neka su dati *AR(p)* i *MA(q)* proces

$$X_t = \theta(L)\varepsilon_t \tag{2.5.12}$$

$$\varepsilon_t = \varphi(L)X_t. \tag{2.5.13}$$

Tada je

$$X_t = \theta(L)\varepsilon_t = \theta(L)\varphi(L)X_t \tag{2.5.14}$$

$$\theta(L)\varphi(L) = 1 \tag{2.5.15}$$

$$\theta(L) = \varphi(L)^{-1} \tag{2.5.16}$$

što daje vezu između koeficijenata procesa.

Kako invertibilan proces pokretnih sredina ima AR(∞) reprezentaciju tada njegova parcijalna korelaciona funkcija nikada nije jednaka nuli i pošto zadovoljava *Yule-Walkers*-ov sistem jednačina za beskonačno mnogo parametara lagano odumire ka nuli.

Sumirajući dobijene rezultate može se reći da stacionarnom autoregreseivnom procesu konačnog reda odgovara proces pokretnih sredina beskonačnog reda, i invertibilnom procesu pokretnih sredina konačnog reda odgovara autoregersivni proces beskonačnog reda. Uzajamna veza između koeficijenata ova dva modela našla je odraza u odnosu njihovih korelacionih i parcijalnih korelacionih koeficijenata. Korelaciona funkcija *AR (p)* procesa i parcijalna korelaciona funkcija *MA (q)* procesa lagano odumiru ka 0, a parcijalna korelaciona funkcija *AR (p)* i korelaciona funkcija *MA (q)* su jednake 0 kada je k veće od reda procesa.

2.6. ARMA (p,q) proces

Definicija 2.6.1 Proces$\{X_t, t \in T\}$ je *ARMA(p,q)* proces ukoliko može da se predstavi

$$X_t = \sum_{j=1}^{p} \varphi_j X_{t-j} + \varepsilon_t - \sum_{i=1}^{q} \theta_i\, \varepsilon_{t-i} \tag{2.6.1}$$

ili u ekvivalentnoj formi

$$\varphi(L)X_t = \theta(L)\varepsilon_t \tag{2.6.2}$$

gde su

- $\varphi(L) = (1 - \varphi_1 L - \cdots \varphi_p L^p)$ *AR* karakteristični polinom *ARMA(p,q)* procesa,
- $\theta(L) = (1 - \varphi_1 L - \cdots \varphi_p L^p)$ *MA* karakteristični polinom *ARMA(p,q)* procesa.

ARMA(p,q) može da se predstavi u obliku

$$\underbrace{X_t - \sum_{j=1}^{p} \varphi_j X_{t-j}}_{AR\ komponenta} = \underbrace{\varepsilon_t - \sum_{i=1}^{q} \theta_i\, \varepsilon_{t-i}}_{MA\ komponenta}. \tag{2.6.3}$$

Pošto je proces pokretnih sredina konačnog reda uvek stacionaran i autoregresivni proces konačnog reda uvek invertibilan onda važi sledeće:

- *ARMA(p,q)* proces je stacionaran ako *AR* komponenta ispunjava uslove stacionarnosti $\Leftrightarrow\ \forall z \in \mathbb{C}$ za koje je $\varphi(z) = 0$ važi $|z| > 1$,
- *ARMA(p,q)* proces je invertibilan ako *MA* komponenta ispunjava uslove invertibilnosti $\Leftrightarrow\ \forall z \in \mathbb{C}$ za koje je $\theta(z) = 0$ važi $|z| > 1$.

Ukoliko je *ARMA(p,q)* proces stacionaran, onda može da se predstavi kao proces pokretnih sredina sa beskonačno mnogo parametara, a ukoliko je invertibilan može da se predstavi kao autoregresivni proces sa beskonačno mnogo parametara.

$$X_t = \sum_{j=1}^{p} \varphi_j X_{t-j} + \varepsilon_t - \sum_{i=1}^{q} \theta_i\, \varepsilon_{t-i} \Big/ X_{t-k}\ /\ E \Rightarrow \tag{2.6.4}$$

$$\begin{aligned}\gamma_k = \varphi_1 \gamma_{k-1} + \cdots + \varphi_p \gamma_{k-p} + E(\varepsilon_t, X_{t-k}) - \theta_1 E(\varepsilon_{t-1}, X_{t-k}) \\ - \cdots \theta_q E\big(\varepsilon_{t-q}, X_{t-k}\big).\end{aligned} \tag{2.6.5}$$

Kako je $E(\varepsilon_{t-i}, X_{t-k}) = 0$ za $k > i$ onda je

$$\gamma_k = \varphi_1 \gamma_{k-1} + \cdots + \varphi_p \gamma_{k-p} \ \text{ za } k \geq q+1 \tag{2.6.6}$$

$$\rho_k = \varphi_1 \rho_{k-1} + \cdots + \varphi_p \rho_{k-p} \ \text{ za } k \geq q+1. \tag{2.6.7}$$

pa autokorelaciona funkcija *ARMA(p,q)* lagano odumire ka nuli.

Kao i u slučaju procesa pokretnih sredina, invertibilan *ARMA(p,q)* proces ima *AR(∞)* reperezentaciju pa njegova parcijalna korelaciona funkcija lagano odumire ka nuli.

2.7. Predviđanje pomoću ARMA(p,q) modela

Jedan od najvažnijih ciljeva u analizi jeste određivanje budućeg toka posmatrane vremenske serije. Najčešće korišćen kriterijum za određivanje prognoze vremenske serije je prognoziranje sa minimalnom srednje kvadratnom greškom . Neka su:

- X_{t+h} buduća vrednost vremenske serije za vremenski interval h
- $\hat{X}_{t+h}$ prognozirana vrednost vremenske serije za vremenski interval h.

Tada je greška prognoziranja

$$e_t(h) = X_{t+h} - \hat{X}_{t+h}, \tag{2.7.1}$$

a srednje kvadratna greška prognoziranja

$$E(X_{t+h} - \hat{X}_{t+h})^2. \tag{2.7.2}$$

Ideja je da se na osnovu poznatih vrednosti vremenske serije formira $\hat{X}_{t+h}$ za koje srednje kvadratna greška prognoziranja ima minimalnu vrednost.

Da bi se izvela prognoza vremenske serije polazi se od najopštijeg invertibilnog i stacionarnog *ARMA* modela

$$\varphi(L)X_t = \theta(L)\varepsilon_t. \tag{2.7.3}$$

Zbog invrertibilnosti proces može da se predstavi kao proces pokretnih sredina beskonačnog reda

$$\begin{gathered} X_t = \psi(L)\varepsilon_t \\ \psi(L) = \frac{\theta(L)}{\varphi(L)} \\ X_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2}+\ldots. \end{gathered} \tag{2.7.4}$$

Kada se zameni $t = n + h$ u 2.7.4 dobija se buduća vrednost vremenske serije za h perioda unapred

$$X_{n+h} = \sum_{j=0}^{\infty} \psi_j\, \varepsilon_{n+h-j}. \tag{2.7.5}$$

U periodu *t=n* se raspolaže sa informacijama $X_n, X_{n-1}, \ldots$ i potrebno je da se formira prognoza buduće vrednosti vremenkse serije za h perioda unapred u obliku linearne kombinacije raspoloživih podataka vremenske serije, odnosno

$$\hat{X}_{n+h} = \psi_h^* \varepsilon_n + \psi_{h+1}^* \varepsilon_{n-1} + \psi_{h+2}^* \varepsilon_{n-2} + \cdots \tag{2.7.6}$$

gde se koeficijenti ψ_j^* određuju minimiziranjem srednje kvadratne greške prognoze.

$$X_{n+h} - \hat{X}_{n+h} = \left(\sum_{j=0}^{h-1} \psi_j \, \varepsilon_{n+h-j} - \sum_{j=0}^{\infty} \psi_{h+j} \varepsilon_{n+j} \right) - \sum_{j=0}^{\infty} \psi_{h+j}^* \varepsilon_{n+j} \tag{2.7.7}$$

$$E(X_{t+h} - \hat{X}_n)^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j + \sigma^2 \sum_{j=0}^{\infty} (\psi_{h+j} - \psi_{h+j}^*)^2 \tag{2.7.8}$$

$$E(X_{n+h} - \hat{X}_{n+h})^2 \longrightarrow min \text{ kada je } \psi_{h+j} = \psi_{h+j}^*. \tag{2.7.9}$$

Kada se uslov 2.7.9 zameni u 2.7.6 dobija se prognoza buduće vrednosti vremenske serije sa minimalnom srednje kvadratnom greškom

$$\hat{X}_{n+h} = \psi_h \varepsilon_n + \psi_{h+1} \varepsilon_{n-1} + \psi_{h+2} \varepsilon_{n-2} + \cdots \tag{2.7.10}$$

Kako je

$$E\left(\varepsilon_{n+h-j} \middle| X_n, X_{n-1}, \dots\right) = \begin{cases} \varepsilon_{n+h-j} & j > h \\ 0 & j \leq h \end{cases} \tag{2.7.11}$$

onda je

$$\begin{aligned} E(X_{n+h} | X_n, X_{n-1}, \dots) &= \sum_{j=0}^{\infty} \psi_j \, E(\varepsilon_{n+h-j}) \\ &= \psi_h \varepsilon_n + \psi_{h+1} \varepsilon_{n-1} + \psi_{h+2} \varepsilon_{n-2} + \cdots \end{aligned} \tag{2.7.12}$$

Dakle prognoza sa minimalnom srednje kvadratnom greškom je data njenom uslovnom očekivanom vrednošću

$$\hat{X}_{n+h} = E(X_{n+h} | X_n, X_{n-1}, \dots). \tag{2.7.13}$$

Greška prognoze je

$$e_n(h) = X_{n+h} - \hat{X}_{n+h} = \sum_{j=0}^{h-1} \psi_j \, \varepsilon_{n+h-j} \tag{2.7.14}$$

pa su

$$E(e_n) = 0 \tag{2.7.14}$$

$$Var(e_n) = Var\left(\sum_{j=0}^{h-1} \psi_j \, \varepsilon_{t+h-j}\right) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{2.7.15}$$

Varijansa greške prognoziranja povećava sa porastom *h*, tako da za što bolju prognozu biramo što manje *h*.

Izračunavanje prognoze

Ako je

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \ldots - \theta_q \varepsilon_{t-1} \tag{2.7.15}$$

ARMA(p,q),onda je buduća vrednost za *h* perioda unapred

$$\begin{aligned} X_{n+h} = \varphi_1 X_{n+h-1} + \cdots + \varphi_p X_{n+h-p} + \varepsilon_{n+h} - \theta_1 \varepsilon_{n+h-1} \ldots \\ - \theta_q \varepsilon_{n+h-q}, \end{aligned} \tag{2.7.16}$$

a njena uslovna očekivana vrednost

$$\begin{aligned} X_{n+h} = \varphi_1 X_{n+h-1} + \cdots + \varphi_p X_{n+h-p} + \varepsilon_{n+h} - \theta_1 \varepsilon_{n+h-1} \ldots \\ - \theta_q \varepsilon_{t-1} \,/\, E(\cdot\, |X_n, X_{n-1}, \ldots) \end{aligned} \tag{2.7.17}$$

$$\begin{aligned} \hat{X}_{n+h} = \varphi_1 E(X_{n+h-1}|X_n, X_{n-1}, \ldots) + \cdots \\ + \varphi_p E\left(X_{n+h-p}\middle|X_n, X_{n-1}, \ldots\right) \\ + E(\varepsilon_{n+h}|X_n, X_{n-1}, \ldots) - \theta_1 E(\varepsilon_{n+h-1}|X_n, X_{n-1}, \ldots) \\ - \cdots - \theta_q E\left(\varepsilon_{n+h-q}\middle|X_n, X_{n-1}, \ldots\right) \end{aligned} \tag{2.7.18}$$

gde je

$$E\left(X_{n+h-j}\middle|X_n, X_{n-1}, \ldots\right) = \begin{cases} X_{n+h-j} & j \geq h \\ \hat{X}_{n+h-j} & j < h. \end{cases} \tag{2.7.19}$$

Primer 2 Izračunavanje prognoze za *AR (1)* proces.

$$X_{n+h} - \mu = \varphi_1 (X_{n+h-1} - \mu) + \varepsilon_{n+h} \tag{2.7.20}$$

h=1

$$\hat{X}_{n+1} = \mu + \varphi_1 (X_n - \mu) \tag{2.7.21}$$

h=2

$$\hat{X}_{n+2} = \mu + \varphi_1\left(\hat{X}_{n+1} - \mu\right) \tag{2.7.22}$$
$$= \mu + \varphi_1(\mu + \varphi_1(X_n - \mu) - \mu)$$
$$= \mu + \varphi_1^2(X_n - \mu)$$

$$\vdots$$

$$\hat{X}_{n+h} = \mu + \varphi_1^h(X_n - \mu). \tag{2.7.23}$$

3. Nestacionarne vremenske serije

3.1. Tipovi nestacionarnosti vremenskih serija

Klasa nestacionarnih procesa obuhvata dve grupe:

- procesi koji imaju eksplozivan tok
- procesi koji transformacijama mogu da se svedu na stacionarne linearne procese.

Kod nestacionarnih procesa koji imaju eksplozivan tok varijansa neograničeno raste u toku vremena pa ova grupa nestacionarnih linearnih procesa nema praktičnog značaja za analizu vremenskih serija. Druga grupa obuhvata sve one procese kod kojih se stacionarnost postiže odgovarajućim transformacijama. Da bi se definisale transformacije potrebno je da se utvrdi uzrok nestacionarnosti procesa. Ako je proces

$$Y_t = D_t + Z_t \tag{3.1.1}$$

- D_t deterministička komponenta
- Z_t stohastička komponenta

mogućnosti koje postoje su sledeće:

1. $D_t = f(t)$ i Z_t je stacionarna vremenska serija
2. $D_t = const.$ i Z_t je nestacionarna vremenska serija
3. $D_t = f(t)$ i Z_t je nestacionarna vremenska serija.

Slučaj 1. Vremenska serija stacionarna oko determinističke komponente koja može da se predstavi nekom matematičkom funkcjom od vremena. Najčešće je to linearna funkcija tako da se nadalje, bez umanjenja opštosti, smatra da je $D_t = \beta_1 + \beta_2 t$.

Ukoliko se oduzme linearni trend sa obe strane jednakosti

$$Y_t = \beta_1 + \beta_2 t + Z_t \; / -(\beta_1 + \beta_2 t) \tag{3.1.2}$$

dobija se

$$Y_t^* = Z_t \tag{3.1.3}$$

koja je stacionarna vremenska serija. Procesi kod kojih se stacionarnost postiže uklanjanjem determinističke komponente se nazivaju trend stacionarni procesi.

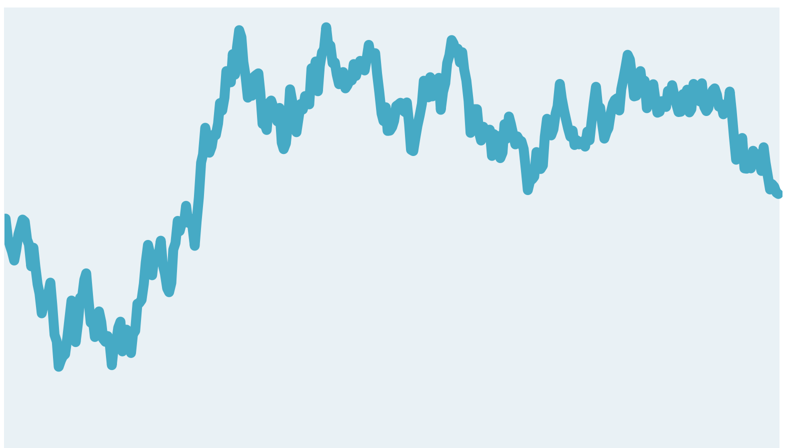
Slučaj 2. Kako je deterministička komponenta konstanta ono što uzrokuje nestacionarnosti procesa je nestacionarnost stohastičke komponente. Da bi se bolje ilustrovao postupak transformacije potrebno je da se definiše proces koji se zove slučajan hod .

Definicija 3.1.1 $\{X_t, t \in T\}$ je proces slučajnog hoda ako važi

$$X_t = X_{t-1} + \varepsilon_t \tag{3.1.4}$$

gde je $\{\varepsilon_t, t \in T\}$ proces belog šuma.

Naziv slučajan hod potiče od karakterističnog ponašanja koje podseća na putanju pijanog čoveka koji slučajno krivuda prilikom kretanja. Vremensku seriju koja prati slučajan hod karakterišu periodi rastućeg i opadajućeg trenda, gde se trend se menja iznenada i promena je nepredvidiva.



Slika 5 Grafik slučajnog hoda

Za proces slučajnog hoda važi

$$X_{t-1} = X_{t-2} + \varepsilon_{t-1} \tag{3.1.5}$$

$$\vdots$$

$$X_1 = X_0 + \varepsilon_1. \tag{3.1.6}$$

$$X_t = X_0 + \sum_{i=1}^{t} \varepsilon_s \tag{3.1.7}$$

$$Var(X_t) = Var\left(X_0 + \sum_{i=1}^{t} \varepsilon_s\right) = t\sigma^2. \tag{3.1.8}$$

Pošto varijansa raste u toku vremena slučajan hod je nestacionaran proces. Sada može da se opiše postupak kojim se postiže stacionarnost.

Ukoliko se sa obe starne jednakosti oduzme X_{t-1}

$$X_t = X_{t-1} + \varepsilon_t \ / -X_{t-1} \tag{3.1.9}$$

$$\nabla X_t = \varepsilon_t \tag{3.1.10}$$

dobija se stacionaran proces. Operacija se zove diferenciranje, a operator ∇ je diferencni operator.

Procesi kod kojih se stacionarnost postiže diferenciranjem se nazivaju diferencno stacionarni procesi. Nekada je potrebno više puta diferencirati proces da bi se postigla stacionarnost.

Definicija 3.1.2 Stohastički proces $\{ X_t, t \in T\}$ je integrisan proces reda d ($I(d)$) ako može da bude transformisan u stacionaran stohastički proces diferenciranjem *d* puta.

Definicija 3.1.3 ARIMA (p,d,q) je proces je definisan sa

$$\varphi(L)\nabla^d X_t = \theta(L)\varepsilon_t \Leftrightarrow \varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \tag{3.1.11}$$

gde su

- ∇^d diferencni opertor reda *d*
- $\varphi(L)$ karakteristični polinom *AR(p)* procesa
- $\theta(L)$ karakteristični polinom *MA(q)* procesa
- $\{\varepsilon_t, t \in T\}$ proces belog šuma

Ukoliko je $\{ X_t, t \in T\}$ *ARIMA (p,d,q)* to znači da je diferenciranjem *d* puta dobijen stacionaran *ARMA(p,q)* proces. To omogućava da se na *ARIMA (p,d,q)* primeni teorija stacionarnih procesa za koju je definisana opšta klasa *ARMA* modela.

Definicija 3.1.4 $\{ X_t, t \in T\}$ je proces slučajnog hoda sa konstantom ako je

$$X_t = c + X_{t-1} + \varepsilon_t \tag{3.1.12}$$

gde je $\{\varepsilon_t, t \in T\}$ proces belog šuma i c konstanta.

Slučajan hod sa konstantom takođe spada u grupu diferencno stacionarnih procesa, jer se njegovim diferenciranjem dobija stacionarna vremenska serija

$$\nabla X_t = c + \varepsilon_t. \tag{3.1.13}$$

Slučajan hod i slučajan hod su prema definiciji 3.1.3 *ARIMA(0,1,0)* procesi.

Slučaj 3. Koristeći prethodna dva slučaja, stacionarnost postižemo tako što prvo uklonimo determinističku komponentu, a zatim primenimo diferenciranje.

3.2. Testovi jediničnih korena

Formalan statistički postupak za testiranje prisustva i tipa nestacionarnosti u modelima linearnih vremenskih serija jeste grupa testova koja se zove testovi jediničnih korena. U prikazu postupka testiranja koristiće se autoregresivni modeli. Ovaj izbor nije restriktivan jer stacionarnost *ARMA* modela zavisi od *AR* komponente.

Prvi model koji se razmatra je *AR (1)*

$$X_t = \phi X_{t-1} + \varepsilon_t. \tag{3.2.1}$$

AR (1) proces je stacionaran ako je $\phi < 1$, a u slučaju kada je $\phi = 1$ tada je 3.2.1 proces slučajnog hoda. Hipoteze koje se postavljaju su sledeće:

$$\begin{gathered} H_0\colon \phi = 1 \\ H_1\colon \phi < 1 \end{gathered} \tag{3.2.2}$$

t statistika

$$\frac{\hat{\phi} - \phi}{SE(\hat{\phi})} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})} \tag{3.2.3}$$

koja se koristi za tetsiranje statističke značajnosti koeficijenta u ovom slučaju nema standardnu Studentovu raspodelu, čak i u asimptotskom slučaju, već je njena raspodela asimetrična ulevo. Ova statistika se zove *Dickey-Fuller* – ova statistika, (*DF* statistika) i označava se sa τ.

Model 3.2.1. ekvivalentno može da se zapiše u obliku

$$\begin{gathered} X_t = \phi X_{t-1} + \varepsilon_t / -X_{t-1} \\ \nabla X_t = (\phi - 1) X_{t-1} + \varepsilon_t \\ \nabla X_t = \alpha X_{t-1} + \varepsilon_t \end{gathered} \tag{3.2.4}$$

$$H_0\colon \alpha = 0$$

$$H_1\colon \alpha < 0,$$

a *DF* statistika

$$\tau = \frac{\hat{\alpha} - 0}{SE(\hat{\alpha})} = \frac{\hat{\alpha}}{SE(\hat{\alpha})}. \quad (3.2.5)$$

Kako proces može da bude i trend stacionaran potrebno je da se model modifikuje 3.2.4 dodavanjem konstante

$$\nabla X_t = \beta_0 + \alpha X_{t-1} + \varepsilon_t \quad (3.2.6)$$

$$H_0: \alpha = 0$$

$$H_1: \alpha < 0.$$

Nultom hipotezom iskazana pretpostavka da vremenska serija prati proces slučajnog hoda sa konstantom protiv altrenativne da je proces stacionaran. Uvođenjem konstante u model menja se prvobitni raspored *DF* statistike.

Naredna modifikacija modela se sastoji u uvođenju lineranog trenda

$$\nabla X_t = \beta_0 + \beta_1 t + (\alpha - 1)X_{t-1} + \varepsilon_t \quad (3.2.7)$$

$$H_0: \alpha = 0$$

$$H_1: \alpha < 0.$$

U ovom slučaju nulta hipoteza je da je linearan proces trend stacionaran, a alternativna da je diferencno stacionaran. I u ovom slučaju se menja prvobitni raspored *DF* statistike.

Za autoregresivne procese višeg reda se koriste modifikacije već opisanih modela. Prema predloženim korekcijama modelima se dodaju članovi $\nabla X_{t-i}, i = 1,2, \dots, p$. Ovime je omogućeno definisanje proširenog *Dickey-Fuller*-ovog testa. Odgovarajuća test statistika se označava sa *ADF (p)*, a model koji se koristi prilikom testiranja prisustva jediničnog korena je

$$\nabla X_t = \beta_0 + \beta_1 t + (\alpha - 1)X_{t-1} + \sum_{i=1}^{p} \lambda_i \nabla X_{t-i} + \varepsilon_t. \quad (3.2.8)$$

Ukoliko se odbaci nulta hipoteza o prisustvu jediničnog korena tada postoji mogućnost da je vremenska serija trend stacionarna.

Kritične vrednosti za sve navedene modifikacije su date u tabeli1. Kritična vrednost se izračunava po formuli

$$\beta_\infty + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} \tag{3.2.9}$$

N	**Model**	**Nivo značajnosti**	β_∞	**Greška ocene**	β_1	β_2
1	Bez konstante	1%	-2.5658	0.0023	-1.960	-10.04
		5%	-1.9393	0.0008	-0.398	0.00
		10%	-1.6156	0.0007	-0.181	0.00
1	Bez trenda	1%	-3.4335	0.0024	-5.999	-29.25
		5%	-2.8621	0.0011	-2.738	-8.36
		10%	-2.5671	0.0009	-1.438	-4.48
1	Sa trendom	1%	-3.9638	0.0019	-8.353	-47.44
		5%	-3.4126	0.0012	-4.039	-17.83
		10%	-3.1279	0.0009	-2.418	-7.58
2	Bez trenda	1%	-3.9001	0.0022	-10.534	-30.03
		5%	-3.3377	0.0012	-5.967	-8.98
		10%	-3.0462	0.0009	-4.069	-5.73
2	Sa trendom	1%	-4.3266	0.0022	-15.531	-34.03
		5%	-3.7809	0.0013	-9.421	-15.06
		10%	-3.4959	0.0009	-7.203	-4.01
3	Bez trenda	1%	-4.2981	0.0023	-13.79	-46.37
		5%	-3.7429	0.0012	-8.352	-13.41
		10%	-3.4518	0.0010	-6.241	-2.79
3	Sa trendom	1%	-4.6676	0.0022	-18.492	-49.35
		5%	-4.1193	0.0011	-12.024	-13.13
		10%	-3.8344	0.0009	-9.188	-4.85
4	Bez trenda	1%	-4.6493	0.0023	-17.188	-59.20
		5%	-4.1000	0.0012	-10.745	-21.57
		10%	-3.8110	0.0009	-8.317	-5.19
4	Sa trendom	1%	-49695	0.0021	-22.504	-50.22
		5%	-4.4294	0.0012	-14.501	-19.54
		10%	-4.1474	0.0010	-11.165	-9.88
5	Bez trenda	1%	-4.9587	0.0026	-22.140	-37.29
		5%	-4.4185	0.0013	-13.641	-21.16
		10%	-4.1327	0.0009	-10.638	-5.48
5	Sa trendom	1%	-5.2497	0.0024	-26.606	-49.56
		5%	-4.7154	0.0013	-17.432	-16.50
		10%	-4.4345	0.0010	-13.654	-5.77
6	Bez trenda	1%	-5.2400	0.0029	-26.278	-41.65
		5%	-4.7048	0.0018	-17.120	-11.17
		10%	-4.4242	0.0010	-13.347	0.00
6	Sa trendom	1%	-5.5127	0.0033	-30.735	-52.50
		5%	-4.9767	0.0017	-20.883	-9.05
		10%	-4.6999	0.0011	-16.445	0.00

Tabela 1

3.3. Stabilizacija varijanse

Neka je data vremenska serija čija varijansa je vremenski zavisna

$$Var(X_t) = \sigma_t^2 \tag{3.3.1}$$

i neka se varijansa nestacionarnog procesa menja sa promenom njene sredine odnosno

$$Var(X_t) = ch(\mu_t) \tag{3.3.2}$$

gde je c je konsatntna, a μ_t je očekivana vrednost.

Za stabilizaciju varijanse, potrebno je da se odredi funkcija f tako da je varijansa transformisane serije konstantna.

$$f(X_t) \approx f(\mu_t) + f'(\mu_t)(X_t - \mu_t) \tag{3.3.3}$$

$$Var(f(X_t)) \approx c(f'(\mu_t)^2 h(\mu_t) \tag{3.3.4}$$

$$f'(\mu_t) = \frac{1}{\sqrt{h(\mu_t)}} \tag{3.3.5}$$

$$f(\mu_t) = \int \frac{1}{\sqrt{h(\mu_t)}} d\mu_t \tag{3.3.6}$$

Primer 3 Neka je $Var(X_t) = c^2{\mu_t}^2$, dakle $h(\mu_t) = {\mu_t}^2$ pa je

$$f(X_t) = \int \frac{1}{\sqrt{{\mu_t}^2}} d\mu_t = \ln(\mu_t). \tag{3.3.7}$$

Ukoliko je varijansa promenljive proporcionalna njenoj očekivanoj vrenosti pogodna transformacija jeste logaritamska.

U cilju stabilizovanja varijanse koristi se takozvana *Box-Cox*-ova transformacija koja je data izrazom

$$X_t^{(\lambda)} = \begin{cases} \dfrac{X_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(X_t) & \lambda = 0. \end{cases} \tag{3.3.8}$$

Koeficijent λ se ocenjuje korišćenjem metode najmanjih kvadrata, odnosno minimiziranjem sume kvadrata reziduala transformisane serije

$$S(\lambda) = \sum_{i=1}^{n} (X_t^{(\lambda)} - \mu^{(\lambda)})^2 \rightarrow min. \tag{3.3.9}$$

Prilikom korišćenja *Box-Cox*-ove transformacije treba voditi računa o sledećem :

- u slučaju kada je $\lambda = 0$ transformacija je definisana samo za pozitivne vrednosti vremenskih serija. U praksi se problem otklanja tako što se dodaje konstanta svim opažanjima jer se tada ne menja korelaciona struktura vremenske serije.
- transformaciju u cilju stabilizovanja varijanse treba sprovesti pre ma koje druge transformacije.

3.4. Izračunavanje prognoze za ARIMA(p,d,q) procese

Neka nam je *ARIMA(p,d,q)* proces

$$\begin{gathered} \varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \Leftrightarrow \\ \alpha(L)X_t = \theta(L)\varepsilon_t \Leftrightarrow \\ \left(1 - \alpha_1 L - \cdots - \alpha_{p+d} L^{p+d}\right) X_t = (1 - \theta_1 L - \cdots - \theta_q L^q)\varepsilon_t \end{gathered} \tag{3.4.1}$$

$t = n + h \Longrightarrow$

$$\begin{aligned} X_{n+h} = \alpha_1 X_{n+h-1} + \cdots + \alpha_{p+d} X_{n+h-(p+d)} + \varepsilon_{n+h} - \theta_1 \varepsilon_{n+h-1} \\ - \cdots - \theta_q \varepsilon_{n+h-q} \ \ / \ E(\cdot\,|X_n, X_{n-1}, \dots) \end{aligned} \tag{3.4.2}$$

$$\begin{aligned} \hat{X}_{n+h} = \alpha_1 E(X_{n+h-1}|X_n, X_{n-1}, \dots) + \cdots \\ + \alpha_{p+d} E\left(X_{n+h-(p+d)}\middle|X_n, X_{n-1}, \dots\right) \\ + E(\varepsilon_{n+h}|X_n, X_{n-1}, \dots) - \theta_1 E(\varepsilon_{n+h-1}|X_n, X_{n-1}, \dots) \\ - \cdots - \theta_q E\left(\varepsilon_{n+h-q}\middle|X_n, X_{n-1}, \dots\right) \end{aligned} \tag{3.4.3}$$

$$E(X_{n+h-j}|X_n, X_{n-1}, \dots) = \begin{cases} X_{n+h-j} & j \geq h \\ \hat{X}_{n+h-j} & j < h \end{cases}$$

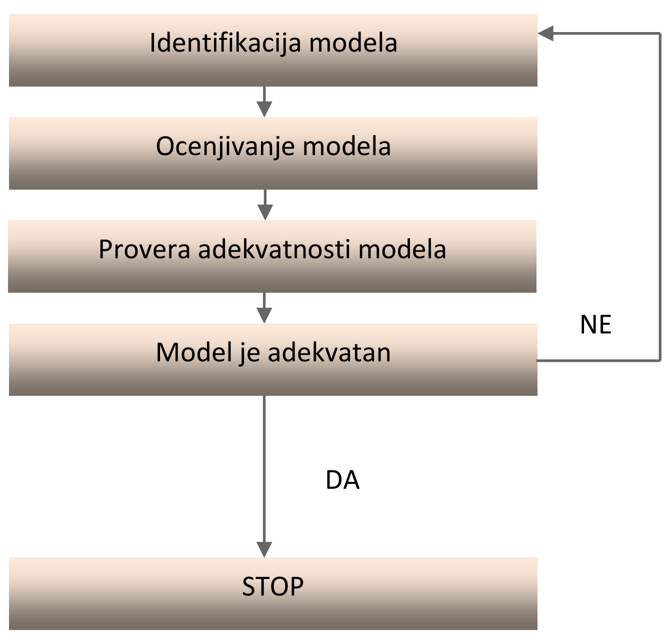
$$E(\varepsilon_{n+h-j}|X_n, X_{n-1}, \dots) = \begin{cases} \varepsilon_{n+h-j} & j > h \\ 0 & j \leq h \end{cases}$$

3.5. Izgradnja ARIMA modela

Opštu strategiju modeliranja *ARIMA* procesa koncipirali su *Box* i *Jenkins* koja je i po njima dobila naziv *Box -Jenkins*-ova metodologija. Ona se satatoji iz 3 etape:

- identifikacija modela
- ocenjivaje modela
- provera adekvatnosti modela.

Koristeći šemu mogu da se prikažu da prikažemo etape u modeliranju:



Ključna etapa *Box-Jenkinsonove* metodologije jeste etapa identifikacije modela. Ona obuhvata postupak utvrđivanja reda diferenciranja, izbor odgovarajućeg modela i utvrđivanje reda procesa. Model koji se izabere u ovoj etapi *Box-Jenkinsonovog* modeliranja se naziva privremeni model.

On prolazi kroz sve ostale etape da bi se na kraju donela konačna odluka o njegovom prihvatanju ili odbacivanju .

Stacionarnost vremenske serije se ispituje sa sa grafika vremenskih serija (time series plot). Ukoliko vremenska serija nije stacionarna onda se, u zavisnosti od tipa nestacionarnosti ona svodi na stacionarnu odgovarajućim transformacijama.

Zatim se na osnovu teorijskih osobina korelacione i parcijalne korelacione funkcije , koje su date u tabeli 2, određuje linearni model i njegov red.

	AR (p)	*MA (q)*	*ARMA (p,q)*
Autokorealciona funkcija	Lagano odumire ka 0	Jednaka 0 kada je red procesa veći od q	Lagano odumire ka 0
Parcijalna autokorelaciona funkcija	Jednaka 0 kada je red procesa veći od *p*	Lagano odumire ka 0	Lagano odumire ka 0

Tabela 2

Osim donošenja zaključka na osnovu korelacione i parcijalne korelacione funkcije, postoji mogućnost da se za određivanje reda modela koriste određeni kriterijumi. Javila se potreba da se u fazi izbora reda *ARMA* procesa *Box-Jenkins*-onov postupak automotizuje korišćenjem statistički utemeljenog skupa pravila. U tom smislu predloženi su brojni kriterijumi za izbor modela, a najčešće korišćena su sledeća dva:

$$AIC\ (p,q) = ln\sigma_{p,q}^2 + \frac{p+q}{n}\ln\,(n) \tag{3.5.1}$$

$$FPE(p,q) = \frac{n+p+q}{n-p-q}\sigma_{p,q}^2. \tag{3.5.2}$$

Vrednosti p i q se biraju tako da minimiziraju $AIC\ (p,q)$ i $FPE(p,q)$.

Primer 4

U tabeli 3 su date vrednosti *AIC* i *FPE* kriterijuma za isti skup podataka i različite modele.

<table>
<tr><td colspan="5">AIC kriterijum</td><td colspan="5">FPE kriterijum (x10^6)</td></tr>
<tr><td rowspan="2">Red AR procesa</td><td colspan="4">Red MA procesa</td><td rowspan="2">Red AR procesa</td><td colspan="4">Red MA polinoma</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>0</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>0</td><td>17.09</td><td>17.12</td><td>17.15</td><td>17.13</td><td>0</td><td>26.51</td><td>27.22</td><td>28.02</td><td>27.39</td></tr>
<tr><td>1</td><td>17.14</td><td>16.86</td><td>16.91</td><td>17.02</td><td>1</td><td>27.91</td><td>20.96</td><td>22.14</td><td>24.58</td></tr>
<tr><td>2</td><td>17.19</td><td>17.08</td><td>17.16</td><td>17.17</td><td>2</td><td>29.32</td><td>26.23</td><td>28.25</td><td>28.58</td></tr>
<tr><td>3</td><td>17.21</td><td>17.16</td><td>17.14</td><td>17.16</td><td>3</td><td>29.39</td><td>28.43</td><td>27.69</td><td>28.34</td></tr>
</table>

Tabela 3

Vrednosti *AIC* i *FPE* kriterijuma su minimalne u slučaju kada je red AR procesa 1 i red MA procesa 1 pa je odgovarajući model izabere *ARMA (1,1)* .

Metod maksimalne verodostojnosti

Za ocenjivanje nepoznatih parametara u modelima vremenskih serija najčešće se koristi metod maksimalne verodostojnosoti. Na onovu uzorka $(x_1, \dots, x_n)$ formira funkcija maksimalne verodostojnosti

$$L(x_1, \dots, x_n; \boldsymbol{\theta}) = f(x_1, \boldsymbol{\theta}) \dots f(x_n, \boldsymbol{\theta}) \tag{3.5.3}$$

gde je $f(\cdot, \boldsymbol{\theta})$ funkcija gustine apsolutno-neprekidne slučajne promenljive i zatim se određuje njen maksimum. Kako L i lnL dostižu maksimum za iste vrednosti, za ocenjivanje nepoznatih parametara će se koristiti maksimum funkcije

$$lnL(x_1, \dots, x_n; \boldsymbol{\theta}). \tag{3.5.4}$$

Da bi mogla da se formira funkcija maksimalne verodostojnosti uvodi se dodatna pretpostavka

$$\varepsilon_i \sim N(0, \sigma^2), i = 1,2, \dots. \tag{3.5.5}$$

Metod makimalne verodostojnosti za autoregersivne procese.

Neka je dat autoregresivni proces reda *p*

$$\begin{gathered} X_t = \mu + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t \\ \varepsilon_t \sim N(0, \sigma^2), \qquad t = 1,2, \dots, n \\ \boldsymbol{\varphi} = \left(c, \varphi_1, \dots \varphi_p, \sigma^2\right). \end{gathered} \tag{3.5.6}$$

Prvo se određuje višedimenzionalna funkcija raspodele za $\left(X_1, \dots, X_p\right)$. Pošto je proces belog šuma Gausovksi onda

$$\left(X_1, \dots, X_p\right) \sim N(c, \sigma^2 V_p) \tag{3.5.7}$$

gde su

$$c = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix} \tag{3.5.8}$$

V_p varijansno-kovarijansna matrica za prvih p promenljivih (3.5.9)

pa je gunkcija gustine

$$f_{(X_1,\dots X_p,)}\left(x_1, \dots x_p\right) = (2\pi)^{-\frac{p}{2}} \frac{1}{\sqrt{\sigma^2 \det\left(V_p\right)}} exp\left\{-\frac{1}{2\sigma^2}(X_p - c)^T V_p^{-1}(X_p - c)\right\}. \tag{3.5.10}$$

Kako je X_t autoregresivni proces onda važi

$$\begin{gathered} X_{p+1} = \mu + \varphi_1 X_p + \dots + \varphi_p X_1 + \varepsilon_p \\ \vdots \\ X_n = \mu + \varphi_1 X_n + \dots + \varphi_p X_{n-p} + \varepsilon_n \end{gathered} \tag{3.5.11}$$

$$X_i \sim N\left(\mu + \varphi_1 X_{i-1} + \dots + \varphi_p X_{i-p}, \sigma^2\right), i = p+1, \dots, n \tag{3.5.12}$$

$$f\left(X_i \middle| X_{i-1}, \dots, X_{i-p}; \boldsymbol{\varphi}\right) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(X_i - \mu - \varphi_1 X_{i-1} - \dots - \varphi_p X_{i-p})^2\right\}. \tag{3.5.13}$$

Funkcija maksimalne verodostojnosti na osnovu uzorka $(x_1, \dots, x_n)$ je

$$L(x_1, \dots, x_n; \boldsymbol{\varphi}) = f\left(x_1, \dots, x_p; \boldsymbol{\theta}\right) \prod_{i=p+1}^{n} f\left(x_i \middle| x_{i-1}, \dots, x_{i-p}; \boldsymbol{\varphi}\right) \tag{3.5.14}$$

$$\ln L(x_1,\dots,x_n;\boldsymbol{\varphi}) = \sum_{t=p+1}^{n} \ln f\left(x_t \middle| x_{i-1},\dots,x_{i-p};\boldsymbol{\theta}\right) + \ln f\left(x_1,\dots,x_p;\boldsymbol{\varphi}\right). \tag{3.5.15}$$

Postoje dve vrste ocena koje se dobijaju metodom maksimalne verodostojnosti:

- Ocene koje se dobijaju maksimizacijom funkcije maksimalne verodostojnosti

$$\widehat{\boldsymbol{\varphi}}_{mle} = max_{\boldsymbol{\varphi}} \ln L(x_1,\dots,x_n;\boldsymbol{\varphi}) \tag{3.5.16}$$

- Ocene koje se dobijaju maksimizacijom uslovne funkcije

$$\widehat{\boldsymbol{\varphi}}_{cmle} = max_{\boldsymbol{\varphi}} \sum_{t=p+1}^{n} \ln f\left(x_t \middle| x_{i-1},\dots,x_{i-p};\boldsymbol{\varphi}\right). \tag{3.5.17}$$

U opštem slučaju funkcija maksimalne verodostojnosti je nelinearna pa se za ocenjivanje koriste metode numeričke analize. Kako $\widehat{\boldsymbol{\varphi}}_{mle}$ i $\widehat{\boldsymbol{\varphi}}_{cmle}$ imaju ista asimptotska svojstva, najčešće se za ocenu nepoznatih parametara koristi uslovna funkcija maksimalne verodostojnosti

$$\begin{aligned}&\sum_{t=p+1}^{n} \ln f\left(x_t \middle| x_{i-1},\dots,x_{i-p};\boldsymbol{\theta}\right)\\ &= -\frac{n-p}{2}\ln(2\pi) - \frac{n-p}{2}\ln(\sigma^2)\\ &- \sum_{i=p+1}^{n} \frac{(x_i-\mu-\varphi_1 x_{i-1}-\dots-\varphi_p x_{i-p})^2}{2\sigma^2}\end{aligned} \tag{3.5.18}$$

$$\frac{\partial lnL}{\partial \mu} = -2\sum_{i=p+1}^{n}(x_i-\mu-\varphi_1 x_{i-1}-\dots-\varphi_p x_{i-p})^2 = 0 \tag{3.5.19}$$

$$\frac{\partial lnL}{\partial \varphi_j} = -2\sum_{i=p+1}^{n}(x_i-\mu-\varphi_1 x_{i-1}-\dots-\varphi_p x_{i-p})^2 x_{i-j} = 0. \tag{3.5.20}$$

Rešenje sistem 3.5.17 i 3.5.18 je

$$\hat{\varphi} = (X^T X)^{-1} X^T Z \tag{3.5.21}$$

$$\hat{\varphi} = \begin{bmatrix} \mu \\ \varphi_1 \\ \vdots \\ \varphi_p \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_p & x_{p-1} & \dots & x_1 \\ 1 & x_{p+1} & x_p & \dots & x_2 \\ 1 & x_{p+2} & x_{p+1} & \dots & x_3 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_{n-1} & \dots & x_{n-p} \end{bmatrix} Z = \begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix} \tag{3.5.22}$$

$$\hat{\sigma}^2 = \sum_{i=p+1}^{n} \frac{(x_i - \mu - \varphi_1 x_{i-1} - \dots - \varphi_p x_{i-p})^2}{n-p}. \tag{3.5.23}$$

Ocene dobijene maksimiziranjem uslovne funkcije maksimalne verodostojnosti su iste kao i ocene parametara dobijene metodom najmnajih kvadrata.

Funkcija maksimalne verodostojnosti za proces pokretnih sredina

Za određivanje uslovne funkcije maksimalne verodostojnosti procesa pokretnih sredina, pored pretpostavke o normalnoj raspodeli, potrebno je da se uvede dodatna pretpostavka

$$\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \dots = 0. \tag{3.5.24}$$

Ako su za *MA(q)* proces

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots \theta_q \varepsilon_{t-q} \tag{3.5.25}$$

u trenutku *t* poznate vrednosti $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ i tada važi

$$X_t \,\middle|\, \varepsilon_{t-1}, \dots, \varepsilon_{t-q} \sim N(\mu - \theta_1 \varepsilon_{t-1} - \cdots \theta_q \varepsilon_{t-q}, \sigma^2) \tag{3.5.26}$$

pa je uslovna funkcija gustine

$$\varphi(X_t; \theta) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\left(X_t - \mu + \theta_1 \varepsilon_{t-1} \dots + \theta_q \varepsilon_{t-q}\right)^2\right). \tag{3.5.27}$$

Zbog uslova 3.5.22 za $X_t = x_1$ uslovna funkcija gustine postaje

$$\varphi(x_1; \theta) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x_1 - \mu)^2\right) \tag{3.5.28}$$

i može da se odredi ε_1

$$\varepsilon_1 = x_1 - \mu. \tag{3.5.29}$$

Nastavljajući dalje proceduru

$$\begin{gathered}\varphi(x_2;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x_2 - \mu + \theta_1\varepsilon_1)^2\right) \\ \varepsilon_2 = x_2 - \mu + \theta_1\varepsilon_1\end{gathered} \tag{3.5.30}$$

$$\vdots$$

$$\begin{gathered}\varphi(x_n;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x_n - \mu + \theta_1\varepsilon_{n-1} \dots + \theta_q\varepsilon_{n-q})^2\right) \\ \varepsilon_n = x_n - \mu + \theta_1\varepsilon_{n-1} \dots + \theta_q\varepsilon_{n-q}\end{gathered} \tag{3.5.31}$$

$$L(x_1, \dots, x_n;\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu + \theta_1\varepsilon_{i-1} \dots + \theta_q\varepsilon_{i-q})^2\right) \tag{3.5.32}$$

ili ekvivalentno

$$L(x_1, \dots, x_n;\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2\right) \tag{3.5.33}$$

$$lnL = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} + \sum_{i=1}^{n}(x_i - \mu + \theta_1\varepsilon_{n-1} \dots + \theta_q\varepsilon_{n-q})^2 \tag{3.5.34}$$

$$lnL = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} + \sum_{i=1}^{n}\varepsilon_i^2. \tag{3.5.35}$$

Funkcija maksimalne verodostojnosti za *ARMA (p,q)*

Ako je

$$\varepsilon_p = \varepsilon_{p-1} = \cdots = \varepsilon_{p-q+1} = 0 \tag{3.5.36}$$

tada je uslovna funkcija maksimalne verodostojnosti za *ARMA (p,q)*

$$L(x_1, \dots, x_n; \theta) = (2\pi\sigma^2)^{-(n-p)/_2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \varepsilon_i^2\right) \tag{3.5.37}$$

$$lnL = -\frac{n-p}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} + \sum_{i=p+1}^{n} \varepsilon_i^2. \tag{3.5.38}$$

Provera adekvatnosti modela

Pod proverom adekvatnosti podrazumevaju se metode koje se koriste za otkrivanje nedostataka modela. U osnovi ove etape su metode koje služe za proveru statističke značajnosti koeficijenata i ispunjenosti pretpostavke da reziduali ocenjenog modela predstavljaju proces belog šuma. Ukoliko se pokaže da je model manjkav moguće je njegovo poboljšanje pa se nastavlja proces izgradnje *ARIMA* modela za datu vremensku seriju. U suprotnom adekvatan model se može koristiti za prognoziranje budućih vrednosti vremenske serije. Nulta hipoteza je da vremnska serija prati proces belog šuma, a alternativna da ne prati.

Klasičnu test statistiku su predložili *Box* i *Pierce*

$$Q_{BP} = n\sum_{k=1}^{m} \rho_k^2. \tag{3.5.39}$$

Ukoliko je nulta hipoteza tačna odnosno ukoliko je vremenska serija beli šum tada

$$Q_{BP} \sim \chi^2_{m-p-q}$$

Modifikovana Q_{BP} statistika koju su predložili *Ljung* i *Box*

$$Q_{BP} = n(n+2)\sum_{k=1}^{m} \frac{\rho_k^2}{n-k} \tag{3.5.40}$$

$$Q_{BP} \sim \chi^2_{m-p-q}$$

se u praksi češće koristi testiranje hipoteza

4. Modeli vektorskih vremenskih serija

4.1. Vektorski linearni procesi

Neka je $\boldsymbol{X}_t = (X_{1,t}, \dots, X_{m,t})$ $t = 0, \pm 1, \pm 2, \dots$. m-dimenzionalni vektorski stohastički proces. Ako je $E(X_{i,t}) = \mu_i$, $i = 1, \dots, m$ tada je

$$E(\boldsymbol{X}_t) = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} \tag{4.1.1}$$

$$Cov(\boldsymbol{X}_t, \boldsymbol{X}_{t-k}) = E\big((\boldsymbol{X}_{t-k} - \boldsymbol{\mu})(\boldsymbol{X}_t - \boldsymbol{\mu})^T\big) \tag{4.1.2}$$

$$= \begin{bmatrix} \gamma_{11}(k) & \gamma_{12}(k) & \dots & \gamma_{1m}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) & \dots & \gamma_{2m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(k) & \gamma_{m2}(k) & \dots & \gamma_{mm}(k) \end{bmatrix} = \boldsymbol{\Gamma}_k$$

gde je

$$\gamma_{ij}(k) = E\Big(\big(X_{i,t-k} - \mu_i\big)\big(X_{j,t} - \mu_j\big)\Big) = E\Big(\big(X_{i,t} - \mu_i\big)\big(X_{j,t+k} - \mu_j\big)\Big) \tag{4.1.3}$$
$$i, j = 1, \dots, m \quad k \in \mathbb{Z}$$

i

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} \gamma_{11}(0) & \gamma_{12}(0) & \dots & \gamma_{1m}(0) \\ \gamma_{21}(0) & \gamma_{22}(0) & \dots & \gamma_{2m}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(0) & \gamma_{m2}(0) & \dots & \gamma_{mm}(0) \end{bmatrix}. \tag{4.1.4}$$

Na osnovu $\boldsymbol{\Gamma}_k$ i $\boldsymbol{\Gamma}_0$ se definiše funkciju korelacione matrice vektorskog stohastičkog procesa

$$\rho(k) = D^{-1/2}\boldsymbol{\Gamma}_k D^{-1/2}$$
$$D = diag\Big(Var\big(X_{1,t}\big), \dots, Var\big(X_{m,t}\big)\Big). \tag{4.1.5}$$

Definicija 4.1.4 Vektorski *m*-dimenzionalni proces belog šuma je definisan sa

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{m,t} \end{bmatrix} \tag{4.1.6}$$

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0} \tag{4.1.7}$$

$$\boldsymbol{\Gamma}_k = E(\boldsymbol{\varepsilon}_{t-k}, \boldsymbol{\varepsilon}_t^T) = \begin{cases} \Sigma_{mm} & k = 0 \\ 0 & k \neq 0 \end{cases}. \tag{4.1.8}$$

Definicija 4.1.2 Linearni vektorski stohastički proces je definisan sa

$$\boldsymbol{X}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Theta}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots = \sum_{i=1}^{\infty} \boldsymbol{\Theta}_i \boldsymbol{\varepsilon}_{t-i} \tag{4.1.9}$$

gde su

- $\boldsymbol{\varepsilon}_{t-i}, i = 0,1,2, \dots$ vektorski *m*-dimenzionalni procesi belog šuma.
- $\boldsymbol{\Theta}_i = \begin{bmatrix} \theta_{11}^i & \cdots & \theta_{mk}^i \\ \vdots & \ddots & \vdots \\ \theta_{m1}^i & \dots & \theta_{mk}^i \end{bmatrix}, i = 1,2, \dots \boldsymbol{\Theta}_0 = I$ matrice koeficijenata.

Koristeći *L* operator vektorski linearni proces može da se predstavi

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \sum_{\boldsymbol{i=0}}^{\infty} \boldsymbol{\Theta}_i L^i = \boldsymbol{\Theta}(L) \boldsymbol{\varepsilon}_t. \tag{4.1.10}$$

Ukoliko je linearni proces invertibilan onda je

$$\begin{aligned} \boldsymbol{X}_t - \boldsymbol{\mu} &= \sum_{\boldsymbol{i=1}}^{\infty} \boldsymbol{\Psi}_i (\boldsymbol{X}_{t-i} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t \\ &= \sum_{\boldsymbol{i=1}}^{\infty} \boldsymbol{\Psi}_i L^i (\boldsymbol{X}_{t-i} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t \Longrightarrow \\ & \boldsymbol{\Psi}(L)(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t. \end{aligned} \tag{4.1.11}$$

Postoje tri grupe konačnih linearnih modela vektorskih vremenskih serija:

1) vektorski autoregersivni proces

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \boldsymbol{\Psi}_1(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \cdots + \boldsymbol{\Psi}_p\big(\boldsymbol{X}_{t-p} - \boldsymbol{\mu}\big) + \boldsymbol{\varepsilon}_t \tag{4.1.12}$$

$$\boldsymbol{\Psi}(L)(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Psi}(L) = \mathbf{I} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{p}} \boldsymbol{\Psi}_i \, L^i$$

2) vektorski proces pokretnih sredina

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} - \cdots - \boldsymbol{\Theta}_q \boldsymbol{\varepsilon}_{t-q} \tag{4.1.13}$$

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Theta}(L) = \mathbf{I} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{q}} \boldsymbol{\Theta}_i \, L^i$$

3) vektorski *ARIMA* proces

$$(\boldsymbol{X}_t - \boldsymbol{\mu}) - \boldsymbol{\Psi}_1(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) - \cdots - \boldsymbol{\Psi}_p(\boldsymbol{X}_{t-p} - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} - \cdots - \boldsymbol{\Theta}_q \boldsymbol{\varepsilon}_{t-q} \tag{4.1.14}$$

$$\boldsymbol{\Psi}(L)(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t.$$

Ukoliko su koreni karakteristične jednačine

$$\det\big(\boldsymbol{I} - \boldsymbol{\Psi}_1 z - \cdots - \boldsymbol{\Psi}_p z^p\big) = 0 \tag{4.1.15}$$

po modulu veći od jedinice, tada je *VAR(p)* proces slabo stacionaran i ima MA reperezentaciju

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \boldsymbol{\Psi}(L)^{-1}\boldsymbol{\varepsilon}_t = \boldsymbol{B}(L)\boldsymbol{\varepsilon}_t = \left(\boldsymbol{I} + \sum_{i=1}^{\infty} \boldsymbol{A}_i L^i\right)\boldsymbol{\varepsilon}_t. \tag{4.1.16}$$

Ukoliko su koreni karakteristične jednačine

$$\det\left(\boldsymbol{I} - \boldsymbol{\Theta}_1 z - \cdots - \boldsymbol{\Theta}_q z^q\right) = 0 \tag{4.1.17}$$

po modulu veći od jedinice, tada je vektorski proces pokretnih sredina invertibilan i ima *VAR* (∞) reprezentaciju

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Theta}(L)^{-1}(\boldsymbol{X}_t - \boldsymbol{\mu}) = \left(\boldsymbol{I} + \sum_{i=1}^{\infty} \boldsymbol{B}_i L^i\right)(\boldsymbol{X}_t - \boldsymbol{\mu}). \tag{4.1.18}$$

Vektorski *ARIMA* proces je slabo stacionaran ukoliko *VAR* komponenta ispunjava uslove stacionarnosti, a invertibilan ukoliko komponente vektorskog procesa pokretnih sredina ispunjava uslove invertibilnosti.

4.2. Granger-ov uticaj za dve vremenske serije

Neka su date dve vremenske serije X i Y . Pitanje koje se posatvlja jeste u kojoj meri vremenska serija Y može da pomogne u predviđanju budućih vrednosti vremenske serije X. Ukoliko ne postoji uticaj Y na X_{t+h} onda se kaže da ne postoji *Granger*-ov uticaj između X i Y. Formalno rečeno to znači da je srednje kvadratna greška budućih vrednosti zasnovana na $(X_t, X_{t-1}, \dots)$ jednaka srednje kvadratnoj greški zasnovanoj na $(X_t, X_{t-1}, \dots, Y_t, Y_{t-1}, \dots)$

$$E(X_{t+h} - E(X_{t+h}|X_t, X_{t-1}, \dots))^2 = E(X_{t+h} - E(X_{t+h}|X_t, X_{t-1}, \dots, Y_t, Y_{t-1}, \dots))^2. \tag{4.2.1}$$

Y ima *Granger*-ov uticaj na X ako je

$$E(X_{t+h} - E(X_{t+h}|X_t, X_{t-1}, \dots))^2 < E(X_{t+h} - E(X_{t+h}|X_t, X_{t-1}, \dots, Y_t, Y_{t-1}, \dots))^2. \tag{4.2.2}$$

U slučaju dvodimenzionalnog *VAR(p)* procesa

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varphi_{11}^{(1)} & \varphi_{12}^{(1)} \\ \varphi_{21}^{(1)} & \varphi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \varphi_{11}^{(p)} & \varphi_{12}^{(p)} \\ \varphi_{21}^{(p)} & \varphi_{22}^{(p)} \end{bmatrix} \begin{bmatrix} X_{t-p} \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{4.2.3}$$

$$E(X_{t+1}|X_t, X_{t-1}, \dots, Y_t, Y_{t-1}, \dots) = \mu_1 + \sum_{i=1}^{p} \varphi_{11}^{(i)} X_{t-p} + \sum_{i=1}^{p} \varphi_{12}^{(i)} X_{t-p} \tag{4.2.4}$$

ako je $\varphi_{12}^{(i)} \equiv 0, i = 1,2,\dots,p$ tada

$$E(X_{t+1}|X_t, X_{t-1}, \dots, Y_t, Y_{t-1}, \dots) = \mu_1 + \sum_{i=0}^{p-1} \varphi_{11}^{(i)} X_{t-i} \tag{4.2.5}$$

pa vrednosti $(Y_t, Y_{t-1}, \dots, Y_{t-p})$ nemaju uticaja na X_{t+1}

$$E(X_{t+1}|X_t, X_{t-1}, \dots, Y_t, Y_{t-1}, \dots) = \sum_{i=0}^{p-1} \varphi_{11}^{(i)} X_{t-p}. \tag{4.2.6}$$

Da bi se formalnim statističkim putem ispitao *Granger*-ov uticaj može da koristi model

$$X_t = c + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + \varepsilon_t \tag{4.2.7}$$

i da se testiramo hipoteza o značajnosti grupe koeficijenata

$$H_0: \beta_1 = \dots . = \beta_p = 0 \tag{4.2.8}$$

ili hipotezu o značajnosti pojedinačnog koeficijenta.

$$H_0: \beta_i = 0, i = 1,2, \dots p. \tag{4.2.9}$$

5. Primer izgradnje dva ARIMA modela

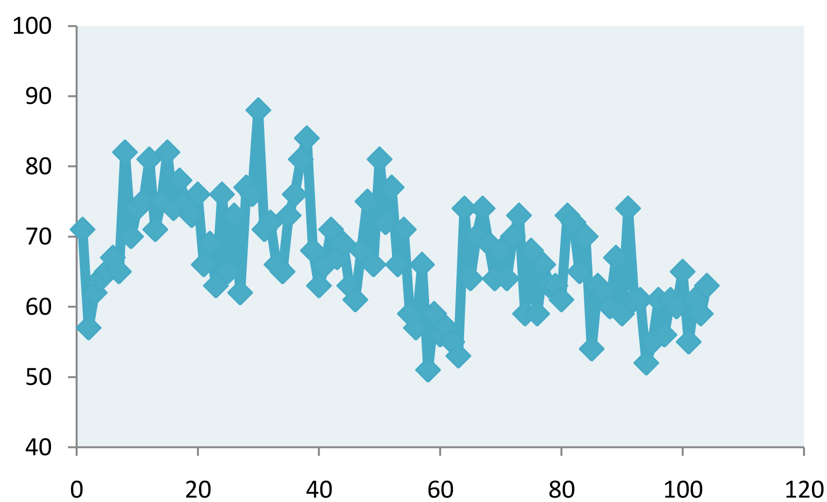
Koristeći *Box-Jenkins*-onovu metodologiju kroz dva primera će biti objašnjen izbor odgovarajućeg modela vremenske serije u praksi, koristeći statističke pakete *Xlstat* i *Eviews*.

t	*Broj kreditnih zahteva*	*t*	*Broj kreditnih zahteva*	*t*	*Broj kreditnih zahteva*	*t*	*Broj kreditnih zahteva*
1	71	27	62	53	66	79	63
2	57	28	77	54	71	80	61
3	62	29	76	55	59	81	73
4	64	30	88	56	57	82	72
5	65	31	71	57	66	83	65
6	67	32	72	58	51	84	70
7	65	33	66	59	59	85	54
8	82	34	65	60	56	86	63
9	70	35	73	61	57	87	62
10	74	36	76	62	55	88	60
11	75	37	81	63	53	89	67
12	81	38	84	64	74	90	59
13	71	39	68	65	64	91	74
14	75	40	63	66	70	92	61
15	82	41	66	67	74	93	61
16	74	42	71	68	69	94	52
17	78	43	67	69	64	95	55
18	75	44	69	70	68	96	61
19	73	45	63	71	64	97	56
20	76	46	61	72	70	98	61
21	66	47	68	73	73	99	60
22	69	48	75	74	59	100	65
23	63	49	66	75	68	101	55
24	76	50	81	76	59	102	61
25	65	51	72	77	66	103	59
26	73	52	77	78	63	104	63

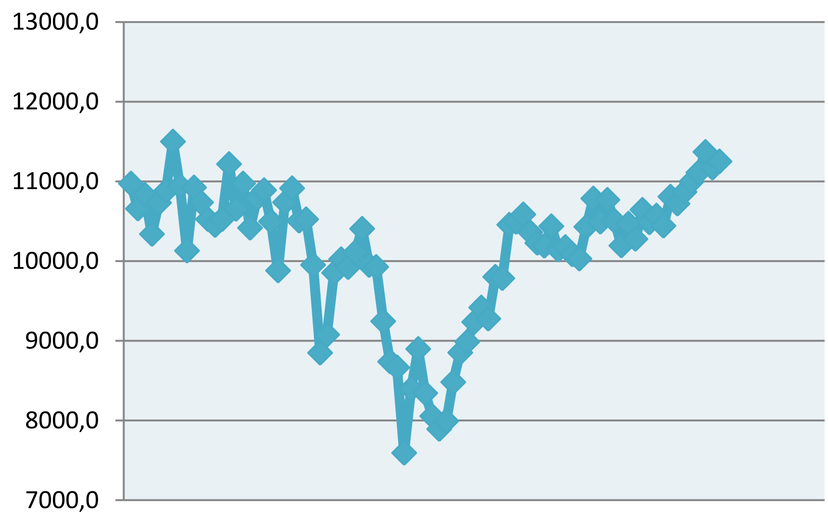
Tabela 4

t	*Dow Jones indeks*	*t*	*Dow Jones indeks*	*t*	*Dow Jones indeks*	*t*	*Dow Jones indeks*	*t*	*Dow Jones indeks*
1	10970.8	18	10414.5	35	9946.22	52	9275.06	69	10766.2
2	10655.2	19	10788	36	9925.25	53	9801.12	70	10503.8
3	10829.3	20	10887.4	37	9243.26	54	9782.46	71	10192.5
4	10337	21	10495.3	38	8736.59	55	10453.9	72	10467.5
5	10729.9	22	9878.78	39	8663.5	56	10488.1	73	10275
6	10877.8	23	10735	40	7591.93	57	10583.9	74	10640.9
7	11497.1	24	10911.9	41	8397.03	58	10357.7	75	10481.6
8	10940.5	25	10502.4	42	8896.09	59	10225.6	76	10568.7
9	10128.3	26	10522.8	43	8341.63	60	10188.5	77	10440.1
10	10921.9	27	9949.75	44	8053.81	61	10435.5	78	10805.9
11	10733.9	28	8847.56	45	7891.08	62	10139.7	79	10717.5
12	10522.3	29	9075.14	46	7992.13	63	10173.9	80	10864.9
13	10447.9	30	9851.56	47	8480.09	64	10080.3	81	10993.4
14	10522	31	10021.6	48	8850.26	65	10027.5	82	11109.3
15	11215.1	32	9920	49	8985.44	66	10428	83	11367.1
16	10650.9	33	10106.1	50	9233.8	67	10783	84	11168.3
17	10971.1	34	10403.9	51	9415.82	68	10489.9	85	11247.9

Tabela 5

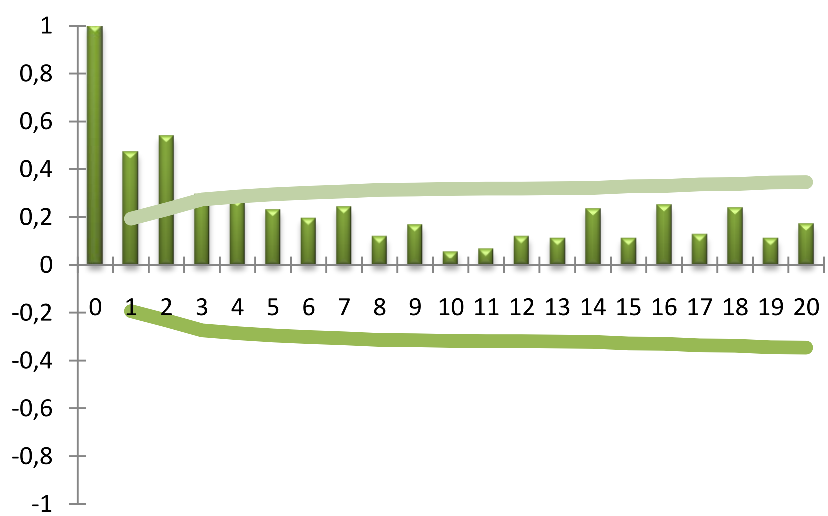


Grafik 1 Grafik broja kreditnih zahteva

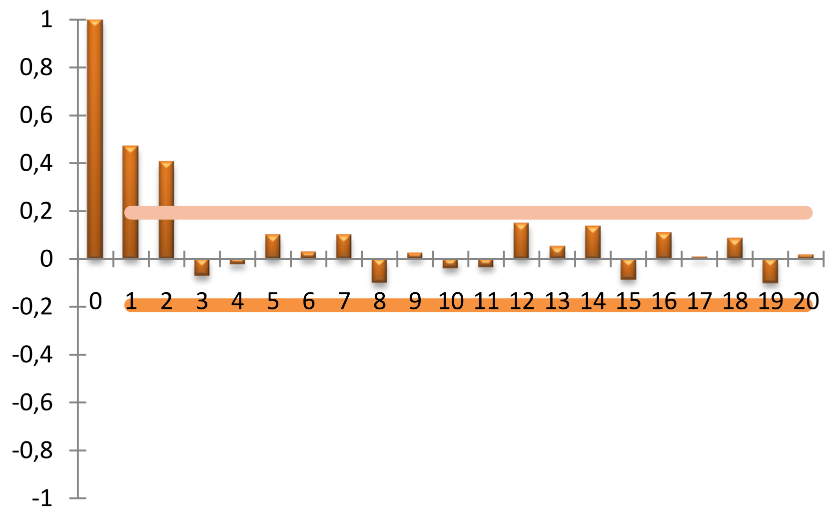


Grafik 2 Grafik Dow Jones-ovog indeksa

Za broj kreditnih zahteva može da se kaže da je stacionarna vremenska serija dok je *Dow Jones*-onov indeks nestacionarna vremenska serija i očigledno sadrži stohastički trend. Za broj kreditnih zahteva mogu odmah da se izučavaju korelaciona i parcijalna korelaciona funkcija dok je podatke za *Dow Jones*-onov indeks prvo potrebno diferencirati, a zatim primeniti teoriju stacionarnih procesa.



Slika 6 Korelaciona funkcija broja kreditnih zahteva



Slika 7 Parcijalna korelaciona funkcija broja kreditnih zahteva

Sa slike 6 i slike 7 može da se zaključi sledeće:

1) korelacioni koeficijenti lagano odumiru kao nuli za $k > 2$
2) parcijalni korelacioni koeficijenti su vrlo bliski nuli za $k > 2$

Na osnovu uporedne analize korelacione i parcijalne korelacione funkcije dobija se da datom skupu podataka odgovara stacionaran autoregresivan proces reda 2, *AR(2)*.

Pomoću *Eviews*-a je su metodom najmnjih kvadrata ocenjeni nepoznati parametri u modelu

$$X_t = 20.73 + 0.28X_{t-1} + 0.41X_{t-2} + \varepsilon_t$$

i

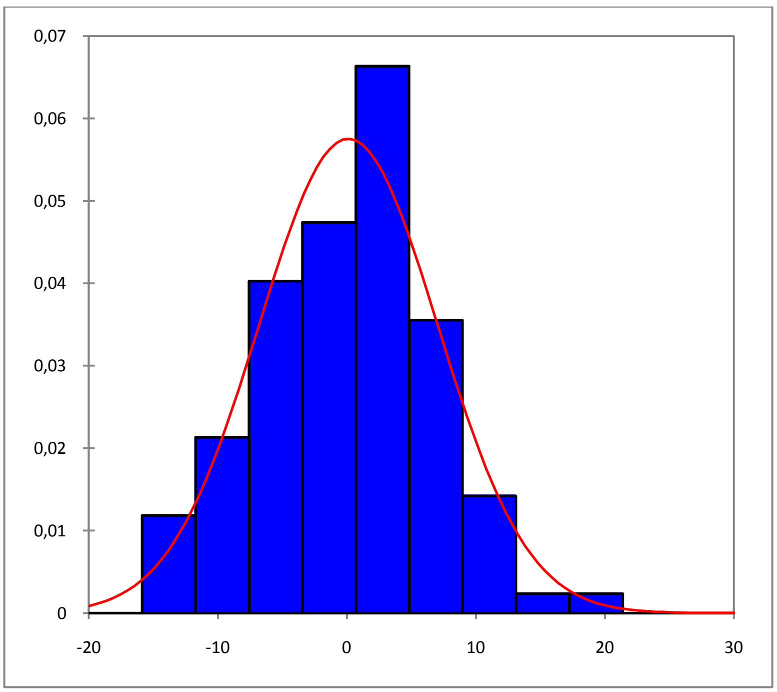
$$\hat{\varepsilon}_t = X_t - 20.73 - 0.28X_{t-1} - 0.41X_{t-2}.$$

Ostalo je još da se proveri da li reziduali u modelu *AR(2)* prate Gausovski proces belog šuma. Vrednosti *Ljung-Box* ove statistike i p- vrednosti se u praksi najčešće izračunavaju za *m=20*, ali je poželjno da budu izračunate za još nekoliko *m*. U Eviews-u su izračunati Q_{BP} i p-vrednosti za *m=10, m=15 i m=20.*

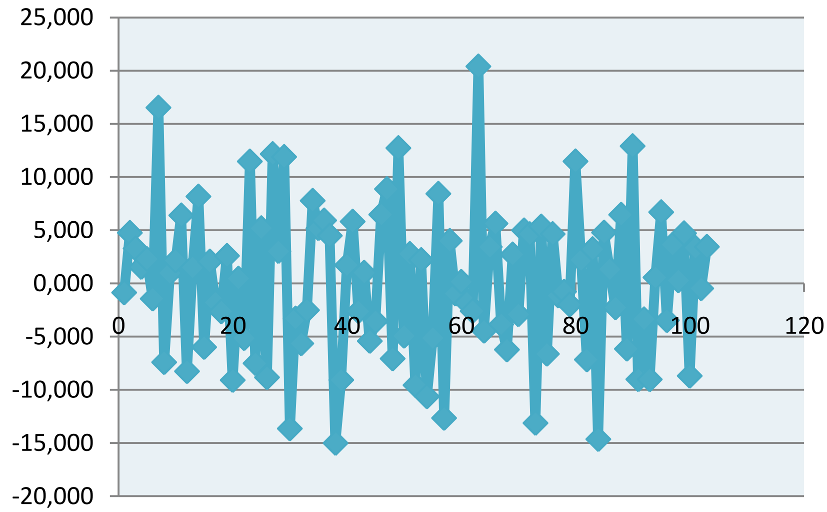
m	*Ljung-Box-ova Q statistika*	*p-vrednost*
10	5.53	0.853
15	10.074	0.815
20	15.55	0.744

p-vrednostu ukazuju da ne treba odbaciti nultu hipotezu, odnosno reziduali su nekorelirani .

U *Ewievs*-u je korišćen *Jarque-Bera* test za testiranje nulte hipoteze da su reziduali normalno raspoređeni protiv alternativne da nisu. Vrednost *JB* statistike je 2.22, a p-vrednost 0.33 što ukazuje da ne treba odbaciti nultu hipotezu na nivou poverenja 5% i 1%.



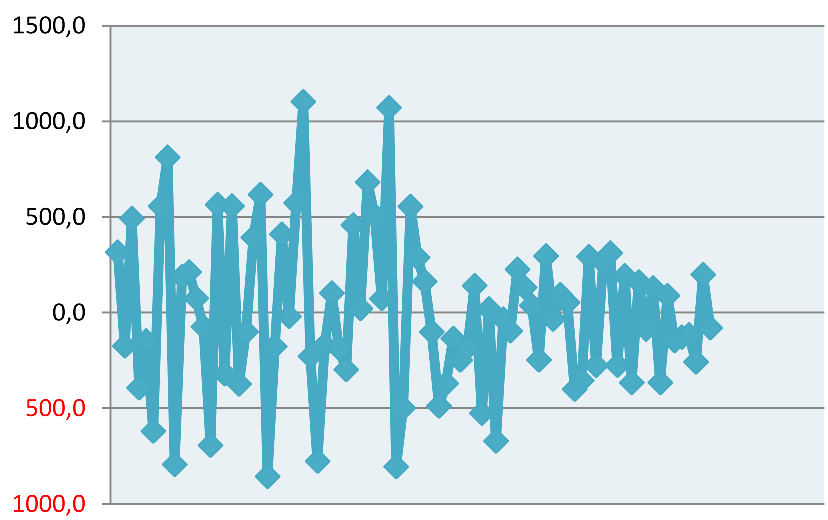
Slika 8 Histogram reziudala



Grafik 3 Grafik reziduala

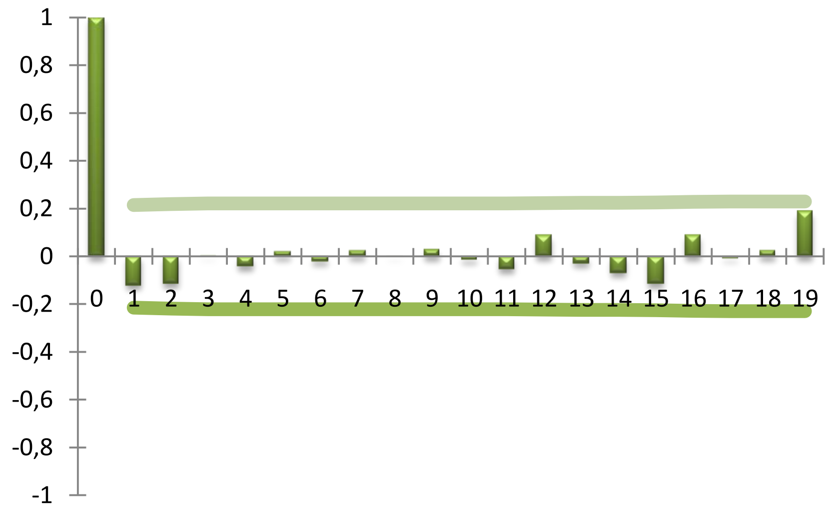
Na grafiku 3 može da se uoči da je varijansa reziduala konstantna. Sumirajući dobijene rezultate može da se zaključi da je *AR (2)* odgovarajući model za broj kreditnih zahteva.

Nakon diferenciranja podataka za *Dow Jones*-onov indeks dobija se stacionarna vremenska serija.

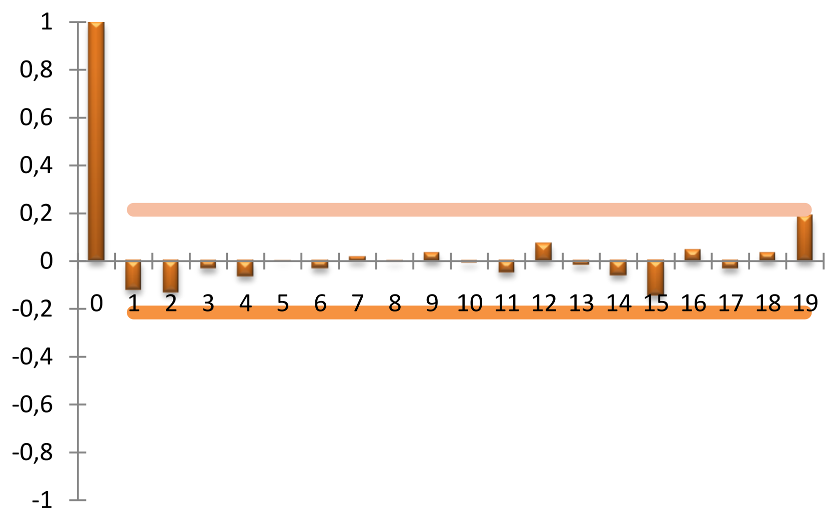


Grafik 4 Grafik prve diference

Sada mogu da se analiziraju korelaciona i parcijalna korelaciona funkcija diferencirane vremenske serije.



Slika 9 Korelciona funkcija prve diference



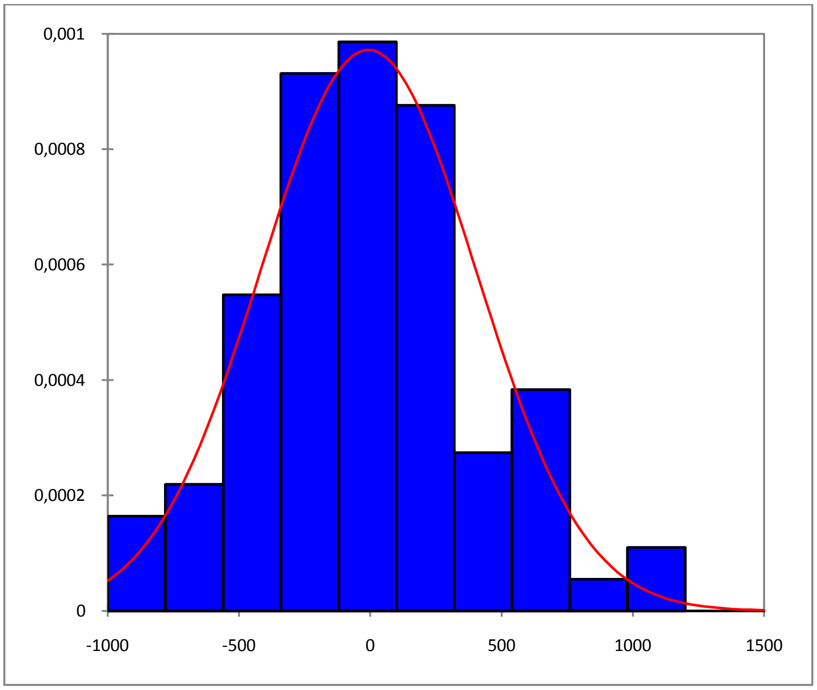
Slika 10 Grafik parcijalne korelacione funkcije

Diferenciranje, korelaciona i parcijalna korelaciona funkcija ukazuju na to da *Dow Jones*-onov indeks prati proces slučajnog hoda (*ARIMA(0,1,0)*)

$$(1-L)X_t = \varepsilon_t \Leftrightarrow \nabla X_t = \varepsilon_t.$$

Vrednost *Ljung-Box* statistike za *m=12* iznosi 3.764 a p-vrednost 0.987 stoga se na nivou poverenja 5% i 1% nulta hipoteza ne odbacije, odnosno reziduali su nekorelirani..

Vrednost *JB* statistike iznosi 1.060, a p–vrednost je 0.589 pa se nulta hipoteza ne odbacuje na nivou poverenja 5% i 1%. Kako za reziduale važi da su nekorelirani i imaju normalnu raspodelu sa konstantnom varijansom, može da se zaključi da oni prate proces belog šuma. Dakle model *ARIMA(0,1,0)* je odgovarajuć za *Dow Jones*-onov indeks.



Slika 11 Histogram reziduala

Zaključak

Modeli lineranih vremenskih serija predstavljaju vrlo moćno sredstvo za predviđanje i donošenje odluka u različitim oblastima-ekonomiji, poljoprivredi, industriji, medicini... U praksi se pokazalo da veliki broj vremenskih pojava može da se modelira pomoću linearnih procesa. Rasprostranjenost i atraktivnost ovih modela je posledica njihove same strukture koja je lako razumljiva. Za razliku od regresionih modela koji razmatraju vezu između dve ili više različitih pojava, modeli vremenskih serija ispituju uticaj istorijskih vrednosti jedne pojave na njenu sadašnju i buduću vrednost. Ovakav pristup omogućava proučavanje ponašanja date pojave u vremenu, i daje dobre rezulate posebno ako je dostupan veliki broj istorijskih podataka.

Naravno kao i svi modeli, ni ovi nisu savršen prikaz stvarnog stanja, ali uprkos tome oni i dalje predstavljaju brz i lak put koji sa dosta preciznosti daje prognozu budućih vrednosti. Na samom kraju je potrebno da se napomene da modeli opisani u ovom radu spaduju u grupu klasičnih modela stoga njihovo poznavanje doprinosi razumevanju moderne teorije analize vremenskih serija.

Literatura

1. J. D. Hamilton- Time Series Analysis, Princeton University Press, New Jersey, 1994.
2. G. Kirchgässner, J. Wolters- Introduction to Modern Time Series Analysis, Springer, 2007.
3. R. S. Tsay- Analysis of Financial Time series, 3rd edition, John Wiley and Sons, New York, 2010.
4. D. C. Montgomery, C. L. Jennings, M. Kulachi- Introduction to Time Series Analysis and Forecasting , John Wiley and Sons, 2008.

Kratka biografija

Rođena sam 01.07.1985. u Novom Sadu. Završila sam osnovnu školu “Jovan Popović ” i gimnaziju “Jovan Jovanović Zmaj”, smer prirodno-matematički, u Novom Sadu.

Prirodno-matematički fakultet u Novom Sadu, odsek za matematiku, smer dipl. matematičar, matematika-finansija, upisala sam 2004. godine.

Master studije na Departmanu za matematiku i informatuku na Prirodno-matematičkom fakultetu u Novom Sadu sam upisala 2010 godine. Poslednji ispit na master studijama sam položila u septembru 2011.

UNIVERZITET U NOVOM SADU
PRIRODNO - MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR
Identifikacioni broj:
IBR
Tip dokumentacije: Monografska dokumentacija
TD
Tip zapisa: Tekstualni štampani materijal
TZ
Vrsta rada: Master rad
VR
Autor: Anja Mihailović
AU
Mentor: Dr Zagorka Lozanov- Crvenković
MN
Naslov rada: Analiza vremenskih serija
MR
Jezik publikacije: Srpski (latinica)
JP
Jezik izvoda: s/e
JI
Zemlja publikovanja: Srbija
ZP
Uže geografsko područje: Vojvodina
UGP
Godina: 2011.
GO
Izdavač: Autorski reprint
IZ
Mesto i adresa: Novi Sad, Departman za matematiku i informatiku, PMF, Trg Dositeja Obradovića 4
MA
Fizički opis rada: (5, 66, 0, 5, 11, 4, 0)
FO
Naučna oblast: Matematika
NO
Naučna disciplina: Statistika
ND
Ključne reči: stohastički proces, linearne vremenske serije, nestacionarne vremenske serije
PO
UDK:

Čuva se: U biblioteci Departmana za matematiku i informatiku
ČU
Važna napomena:
VN
Izvod:
IZ
U ovom radu su razmatrani modeli linearnih vremenskih serija, nestacionarne vremenske serije kao i vektorski modeli vremenskih serija.
Datum prihvatanja teme od strane NN veća: 31.05.2011.
DP
Datum odbrane: 25.10.2011.
DO
Članovi komisije:
KO

Predsednik: : Dr Ljiljana Gajić, redovni profesor Prirodno-matematičkog fakulteta u Novom Sadu

Član: Dr Zagorka Lozanov-Crvenković, redovni profesor Prirodno-matematičkog fakulteta u Novom Sadu

Član : Dr Danijela Rajter-Ćirić, vandredni profesor Prirodno-matematičkog fakulteta u Novom Sadu

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 thesis
CC
Author: Anja Mihailović
AU
Mentor: Dr Zagorka -Lozanov Crvenković
MN
Title: Time series analysis
TI
Language of text: Serbian
LT
Language of abstract: English
LA
Country of publication: Serbia
CP
Locality of publication: Vojvodina
LP
Publication year: 2011.
PY
Publisher: Author's reprint
PU
Publ. place: Novi Sad, Faculty of Natural Sciences and Mathematics, Trg Dositeja Obradovića 4
PP
Physical description: (5, 66, 0, 5, 11, 4, 0)
PD
Scientific field: Mathematics
SF
Scientific discipline: Statistics
SD
Key words: stochastic process, linear time series models, nonstationary time series
SKW
UC:

Holding data: In library of Department of Mathematics
HD
Note:
N
Abstract:
AB
This thesis discusses models of linear time series, nonstationary time series and vector time series models.
Accepted by the Scientific Board on: 31.05.2011.
ASB
Defended: 25.10.2011.
DE
Thesis defend board:
DB

President: Dr Ljiljana Gajić, Full Professor, Faculty of Natural Sciences and Mathematics, University of Novi Sad

Member: Dr Zagorka Lozanov-Crvenković, Full Professor, Faculty of Natural Sciences and Mathematics, University of Novi Sad

Member: Dr Danijela Rajter-Ćirić , Associate Professor, Faculty of Natural Sciences and Mathematics, University of Novi Sad