Originalni naučni rad
UDK: 551.577/518

MODELIRANJE NIZOVA MESEČNIH SUMA PADAVINA METODOM VREMENSKIH SERIJA

Ivana TOŠIĆ[1], Jovan MALIŠIĆ[2], Miroslava UNKAŠEVIĆ[1]
[1]Fizički fakultet, Institut za meteorologiju, Beograd
[2]Matematički fakultet, Beograd

REZIME

Za modeliranje serija mesečnih suma padavina mogu se koristiti razni statistički postupci. U njih spadaju i izbori raznih modela korišćenjem statističke analize vremenskih serija. U slučaju stacionarnih serija često se kao pogodni javljaju modeli pokretnih proseka (MA), autoregresijski (AR) modeli i kombinovani (ARMA) modeli. Izbor modela može se vršiti na osnovu raznih kriterijuma. Jedan od njih je izbor na osnovu ponašanja uzoračkih serijskih i uzoračkih parcijalnih koeficijenata korelacije.

Postupak izbora i ocene modela primenjen je na seriju mesečnih suma padavina osmotrenih u periodu 1888 - 2000. god. na Meteorološkoj opservatoriji u Beogradu.

Ključne reči: meteorološka vremenska serija, ARMA modeli, modeliranje, serijska korelacija, parcijalna korelacija

1. UVOD

U dugačkim vremenskim serijama često se zapaža postojanje jedne determinističke sistematske komponente, kojom se izražava dugoročna tendencija razvoja serije. Ta komponenta je trend. U praksi, ona je često manje ili više zamaskirana postojanjem izvesnih slučajnih fluktuacija. Pored trenda, u vremenskim serijama ponekad se zapaža još jedna deterministička komponenta - sezonska komponenta. Ona je rezultat dejstva faktora koji se periodično, u ciklusima, po sezonama javljaju. Mnoge vremenske serije imaju još jednu periodičnu komponentu, cikličnu komponentu. Postoji još i četvrta komponenta, koja je slučajna komponenta. Sve nabrojane komponente su teorijskog karaktera. Prirodno je smatrati da se u većini slučajeva elementi u seriji dobijaju sumiranjem vrednosti trenda, sezonske, ciklične i slučajne komponente. Svaki član vremenske serije

$$X_1, X_2, X_3, \ldots, X_n$$

može se prikazati u obliku

$$X_t = f_t + s_t + c_t + \varepsilon_t \quad t = 1,2,\ldots,n,$$

gde je f_t - trend, s_t - sezonska, c_t - ciklična, a ε_t - slučajna komponenta. U daljem radu će postupak izbora i provere modela (kao i oznake) biti dat prema Mališić [8].

2. SEZONSKA KOMPONENTA

Sezonski efekti u meteorološkim serijama su skoro uvek prisutni. Zato te efekte treba oceniti i eliminisati iz serije pre same obrade podataka iz te serije. Za sezonsku komponentu s_t ćemo pretpostaviti da je to neslučajna funkcija vremena sa poznatom periodom d. U slučaju serija mesečnih vrednosti je $d = 12$. Smatraće se da je $T = \{1,2,\ldots,n\}$, gde je $n = kd$ i $s_{t+d} = s_t$ za sve $t = 1,2,\ldots,n-d$ kao i $s_1 + s_2 + \ldots + s_d = 0$. Ukoliko se u seriji očekuje postojanje sezonske komponente to se može preliminarno ispitati tako što se nalaze prosečne vrednosti po pojedinim sezonama:

$$\overline{X}_j = \frac{1}{k}\sum_{i=1}^{k} X_{(i-1)d+j}, \quad j = \overline{1,d} \tag{1}$$

i ispituje se da li među njima postoje značajne razlike. U slučaju postojanja takvih razlika može se pristupiti otkrivanju i eliminaciji sezonske komponente.

U praksi postoji više načina za eliminaciju sezonske komponente. Ovde će se koristiti jedna takva metoda.

2.1 Metoda konstantnih trendova

U slučaju malih promena trendova u okviru jedne periode, aproksimativno se uzima da je u okviru jedne periode trend konstantan. Dakle, $m_i = f[(i-1)d+1] = f[(i-1)d+2] = \ldots f(id)$ za $i = \overline{1,k}$. Zato će se trend m_i (nepristrasno) ocenjivati statistikom $\hat{m}_i = \frac{1}{d}\sum_{j=1}^{d} X_{(i-1)d+j}$, $i = 1,2,\ldots k$, a sezonski efekat po pojedinim delovima (sezonama) periode sa

$$\hat{s}_j = \frac{1}{k}\sum_{j=1}^{k}\left[X_{(i-1)d+j} - \hat{m}_j\right],\ j = \overline{1,d}, \tag{2}$$

pri čemu će biti $E(\hat{s}_1 + \hat{s}_2 + \ldots + \hat{s}_d) = 0$. Desezonizovani podaci će tada biti

$$\varepsilon_{(i-1)d+j} = X_{(i-1)d+j} - \hat{m}_i - \hat{s}_j, \text{ za sve } i = \overline{1,k} \text{ i sve } j = \overline{1,d}. \tag{3}$$

3. ARMA MODELI VREMENSKIH SERIJA

Autoregresijski (AR), modeli pokretnih proseka (MA) i kombinovani (ARMA) modeli se koriste u meteorologiji/hidrologiji i postali su veoma popularni. Jednostavni su za korišćenje i obuhvataju širok opseg stohastičkih procesa koji su korisni u prognozi; obezbeđuju statističko predstavljanje podataka u zavisnosti od malog broja parametara. Primeri obuhvataju mesečne padavine [3], [10], mesečne indekse suše [2], [4], mesečne proticaje [5], [12], [9], [7], itd.

U statističkoj analizi konkretnih vremenskih serija čest zadatak je formiranje stohastičkog modela koji može da dobro opisuje dati proces. Najčešće mora da se model formira na osnovu diskretnog skupa posmatranja. Tako u meteorologiji se radi sa modelima za dnevne, desetodnevne, mesečne ili godišnje padavine. Mi ćemo se ovde opredeliti za analizu i modeliranje mesečnih suma padavina. Razlozi za to su višestruki. Prvo, mesečne vrednosti padavina svakako dobro odražavaju strukturu procesa padavina, jer sadrže u sebi i sezonsku i slučajnu komponentu (kao i trend - ukoliko on postoji), dok toga nema u serijama godišnjih padavina (u kojima su osnovna svojstva procesa padavina jako zamaskirana). Sem toga, mesečnih podataka ima 12 puta više nego godišnjih, što je mnogo povoljnije za statističku analizu.

Postupak izbora modela prikazaćemo na primerima modela MA, AR i ARMA tipova, primenjenih na desezonizovane serije podataka (koje ćemo ovde označavati sa X_t, $t \in T$, a ne sa ε_t iz (3)), kod kojih je odstranjen trend.

3.1 MA(q) model

U opštem MA(q) modelu

$$X_t = \xi_t + a_1\xi_{t-1} + \ldots + a_q\xi_{t-q},\ t \in T \tag{4}$$

prvih q koefiicijenata korelacije su različiti od nule, a svi ostali su jednaki nuli. Dakle,

$$\rho_1 \neq 0,\ \rho_2 \neq 0, \ldots, \rho_q \neq 0,\quad \rho_l = 0 \text{ za } l > q. \tag{5}$$

Koeficijenti korelacije $\rho_1, \rho_2, \ldots, \rho_q$ se ocenjuju njihovim uzoračkim vrednostima $r_1, r_2, \ldots, r_q$. Serija $\{\xi_t, t \in T\}$ predstavlja beli šum - niz slučajnih promenljivih sa nultim srednjim vrednostima $(E\xi_t = 0)$, konstantnom disperzijom $(D\xi_t = \sigma^2)$, međusobno nekorelisanih ($\mathrm{cov}(\xi_t, \xi_s) = E\xi_t\xi_s = 0$ za $t \neq s$). To ćemo označavati sa: $\xi_t \perp (0;\sigma^2)$. Koeficijenti parcijalne korelacije Φ_{kk} su svi različite od nule, ali pokazuju tendenciju eksponencijalnog opadanja (po apsolutnoj vrednosti). Za MA(q) seriju nalazimo

$$EX_t = 0,\ DX_t = \sigma^2\sum_{j=0}^{q} a_j^2 \tag{6}$$

$$K_X(t,s) \equiv K(\tau) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|\tau|} a_j a_{j+|\tau|}, & 0 \leq |\tau| \leq q \\ 0, & |\tau| > q \end{cases}. \tag{7}$$

Iz relacija (6)-(7) se vidi da je MA(q) jedna stacionarna serija i da koeficijent korelacije sa korakom τ ima vrednost

$$\rho_\tau = \frac{a_0a_\tau + a_1a_{\tau+1} + \ldots + a_{q-\tau}a_q}{a_0^2 + a_1^2 + \ldots a_q^2},\ \tau = \overline{1,q}. \tag{8}$$

Kada je u pitanju MA(1) model

$$X_t = \xi_t + a\xi_{t-1},\ \xi_t \perp (0;\sigma^2), \tag{9}$$

ocenjuju se dva parametra: a i σ^2. Parametar a se obično ocenjuje iz njegove veze sa koeficijentom korelacije sa korakom 1:

$$\rho_1 = \frac{a}{1+a^2}. \tag{10}$$

Zato imamo ocenu:

$$r_1 = \frac{\hat{a}}{1+\hat{a}^2}. \tag{11}$$

Vrednost $\hat{a}$ se dobija rešavanjem jednačine po $\hat{a}$, pri čemu od dva korena jednačine

$$r_1\hat{a}^2 - \hat{a} + r_1 = 0 \tag{12}$$

se bira onaj iz intervala (-1,1) (da bi proces bio obratni). Ocena parametra σ^2 može se dobiti iz veze

$$K_X(0) = (1+a^2)\sigma_\xi^2 \tag{13}$$

u kojoj treba a zameniti prethodno dobijenom ocenom, a $K_X(0)$ nekom od ocena disperzije.

MA(2) serije su serije definisane sa

$$X_t = \xi_t + a_1\xi_{t-1} + a_2\xi_{t-2},\ \xi_t \perp (0;\sigma^2), \tag{14}$$

gde su a_1 i a_2 realni koeficijenti, čije su vrednosti određene uslovom:

$$r_1 = \frac{a_1(1+a_2)}{1+a_1^2+a_2^2},\ r_2 = \frac{a_2}{1+a_1^2+a_2^2}. \tag{15}$$

To znači da se ocene koefecijenata a_1 i a_2 dobijaju rešavanjem jedne jednačine četvrtog stepena. Ocena parametra σ^2 može se dobiti iz veze

$$K_X(0) = (1+a_1^2+a_2^2)\sigma_\xi^2, \tag{16}$$

ako su koeficijenti a_1 i a_2, kao i disperzija $K_X(0)$ prethodno ocenjeni.

3.2 AR(p) model

Neka je X_t jedna stacionarna AR(p), $p \geq 1$ serija određena sa

$$X_t + b_1X_{t-1} + \ldots + b_pX_{t-p} = \xi_t,\ \xi_t \perp (0;\sigma^2). \tag{17}$$

Niz koeficijenata korelacije ove serije se ponaša kao niz parcijalnih korelacionih funkcija MA(p) serija. Dakle, u tom nizu preovladava tendencija opadanja (po modulu) eksponencijalnom brzinom. S druge strane, niz parcijalnih korelacionih funkcija AR(p) serija ponaša se kao niz koeficijenata korelacija MA(p) serija. To znači da u AR(p) seriji je

$$\Phi_{11} \neq 0, \Phi_{22} \neq 0, \ldots, \Phi_{pp} \neq 0, \Phi_{kk} = 0 \text{ za } k > p \tag{18}$$

Kao početne vrednosti koeficijenata $b_1, b_2, \ldots, b_p$ uzimaju se one vrednosti koje slede iz Džul-Vokerovih jednačina u kojima umesto koeficijenata korelacije $\rho_1, \rho_2, \ldots, \rho_p$ se uzimaju njihove uzoračke ocene $r_1, r_2, \ldots, r_p$.

Poznato je da u stacionarnoj AR(p) seriji važe Džul-Vokerove jednačine:

$$C \cdot \Phi = R \tag{19}$$

gde je

$$C = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-3} \\ \rho_3 & \rho_2 & \rho_1 & 1 & \cdots & \rho_{p-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \rho_{p-4} & \cdots & 1 \end{bmatrix},$$

$$\Phi = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \Phi_4 \\ \vdots \\ \Phi_p \end{bmatrix},\ R = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \vdots \\ \rho_p \end{bmatrix} \tag{20}$$

Dakle, C je matrica koeficijenata korelacije $\rho_k = \rho(X_t, X_{t+k})$, Φ je matrica-kolona AR koeficijenata, a R je matrica-kolona koeficijenata korelacije.

Pretpostavimo da je predložen stacionaran model AR(1). To je model određen sa

$$X_t = bX_{t-1} + \xi_t,\ |b| < 1,\ \xi_t \perp (0;\sigma^2),\ t \in T. \tag{21}$$

Za određivanje dva parametra: b i $\sigma^2 \equiv \sigma_\xi^2$ mogu se koristiti rezultati:

$$b = \rho_1,\ \sigma^2 = (1-b^2)K_X(0). \tag{22}$$

Zato će biti:

$$\hat{b} = r_1,\ \hat{\sigma}^2 = (1-\hat{b}^2)\overline{K_X(0)} \equiv (1-\hat{b}^2)\hat{\sigma}_X^2, \tag{23}$$

gde je $\hat{\sigma}_X^2$ jedna od uzoračkih ocena disperzije serije X_t.

Neka se pretpostavi da je kao model prepoznat AR(2) model, tj. model određen sa:

$$X_t = h_1 X_{t-1} + h_2 X_{t-2} + \xi_t,\ t \in T. \tag{24}$$

Rešavanjem Džul-Vokerovih jednačina se dobija:

$$h_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2},\ h_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}, \tag{25}$$

$$\sigma^2 = K(0) - h_1K(1) - h_2K(2) = \frac{(1-\rho_2^2)-2\rho_1^2(1-\rho_2)}{1-\rho_1^2}K(0) \tag{26}$$

Vrednosti parametara h_1, h_2 i σ^2 se dobijaju kada se u gornjim relacijama umesto ρ_1, ρ_2 i $K(0)$ uzmu njihove uzoračke ocene. Seriji X_t definisanoj u (24) odgovara karakteristična jednačina

$$m^2 + b_1 m + b_2 = 0. \tag{27}$$

Da bi serija X_t bila stacionarna, potrebno je da rešenja (27) β_1 i β_2 budu po modulu manja od 1, tj. da su oba korena karakteristične jednačine u jediničnom krugu. Ako se smatra još da je $\beta_1 \neq \beta_2$, tada može da se piše

$$X_t - (\beta_1+\beta_2)X_{t-1} + \beta_1\beta_2 X_{t-2} = \xi_t, \tag{28}$$

odnosno

$$X_t = \sum_{r=0}^{\infty} c_r \xi_{t-r}, \tag{29}$$

gde je

$$c_r = \beta_1^r + \beta_1^{r-1}\beta_2 + \ldots + \beta_2^r = \frac{\beta_1^{r+1}-\beta_2^{r+1}}{\beta_1-\beta_2},\ r = 0,1,2,\ldots \tag{30}$$

U slučaju $\beta_1 = \beta_2$ biće $c_r = (r+1)\beta_1^r$.

3.3 ARMA(p,q)

Među modelima koji se mogu koristiti vrlo su popularni ARMA modeli: mešavine autoregresijskih modela i modela pokretnih proseka. Njihova široka primena u analizi meteoroloških/hidroloških pojava zasniva se na činjenici da oni u dosta slučajeva dobro opisuju realni proces, a da su, po svojoj strukturi, relativno prosti u odnosu na druge modele.

ARMA(p,q) model za seriju $\{X_t, t \in T\}$ definisan je relacijom

$$\Phi(B)X_t = \Theta(B)\xi_t \tag{31}$$

gde je

$$\Phi(B) = 1 + b_1B + \ldots + b_pB^p,\ b_1, b_2, \ldots, b_p \in R, \tag{32}$$

$$\Theta(B) = 1 + a_1B + \ldots + a_qB^q,\ a_1, a_2, \ldots, a_q \in R, \tag{33}$$

a B je operator translacije unazad određen sa $BX_t = X_{t-1}$, $B^0X_t = X_t$, $B^kX_t = X_{t-k}$.

Kod ovih vrsta modela uobičajena je pretpostavka da se radi o modelima stacionarnih serija ö serija koje imaju konstantne srednje vrednosti, konačne momente drugog reda i gde je korelaciona struktura homogena tokom vremena. Pokazuje se da je model stacionaran ako i samo ako svi koreni karakteristične jednačine $\Phi(z) = 0$ leže van jediničnog kruga $|z| = 1$.

Pored stacionarnosti, bitna pretpostavka je i obratnost (inverzibilnost). Naime, serija je obratna ukoliko može biti predstavljena u vidu $\xi_t = \sum_{j=0}^{\infty} c_j X_{t-j}$. Uslovi za obratnost izražavaju se samo u terminima MA koeficijenata. Odnosno, serija je obratna ako i samo ako svi koreni karakterisitčne jednačine $\Theta(z) = 0$ leže van jediničnog kruga $|z| = 1$. Restrikcija po kojoj se posmatraju samo obratne ARMA serije obezbeđuje da se u izražavanju tekuće vrednosti serije ili njene prognoze javljaju samo prošle vrednosti serije, ali ne i buduće vrednosti (što bi bilo i fizički neopravdano i računski komplikovano).

ARMA procesi ne moraju se obavezno smatrati gausovskim. Ipak, mnogo povoljnije je koristiti takve procese zato što linearne transformacije nad takvim procesima čuvaju gausovsko svojstvo. Zato su takvi procesi i dobro izučeni i često primenjivani. Sem toga, neki kriterijumi za izbor modela vezani su za gausovske procese, a za druge vrste mogu se samo izuzetno koristiti.

Neka je kao model predložen ARMA(1,1) model određen sa:

$$X_t + b_1 X_{t-1} = \xi_t + a_1 \xi_{t-1}, \; \xi_t \perp (0; \sigma^2). \tag{34}$$

Za ovaj model važe relacije:

$$\rho_1 = \frac{(1-a_1 b_1)(a_1 - b_1)}{1 - 2a_1 b_1 + a_1^2}, \; \rho_k = (-b_1)\rho_{k-1}. \tag{35}$$

Zamena koeficijenata korelacije ρ_1 i ρ_2 njihovim uzoračkim ocenama r_1 i r_2 daje mogućnost dobijanja koeficijenata a_1 i b_1 iz predloženog modela. U modelu ARMA(p,q) ima $p+q+1$ nepoznatih parametara i to:

- p koeficijenata u AR delu serije,
- q koeficijenata u MA delu serije i
- disperzija $\sigma^2 \equiv \sigma_\xi^2$ belog šuma.

Zbog tako velikog broja nepoznatih parametara, veliki interes predstavlja izbor adekvatnog modela sa što manjim brojem nepoznatih parametara.

4. IZBOR MODELA ARMA TIPA

Postoji više načina za određivanje reda ARMA modela. Među njima su najpoznatiji Akaikeov informacioni kriterijum (AIC) i Bajesov informacioni kriterijum (BIC). U određivanju reda (preciznije rečeno - u izboru modela) mogu se koristiti vrednosti koeficijenata obične korelacije i koeficijenata parcijalne korelacije i taj postupak će i ovde biti primenjen.

Kod mešovitih procesa ARMA(p,q) pokazuje se:

- da je niz koeficijenata korelacije beskonačan i posle prvih (q - p) koraka ima prigušene eksponencijalne i/ili prigušene sinusne komponente;
- da je niz parcijalnih korelacionih funkcija beskonačan i posle prvih (p - q) koraka u njemu preovladavaju prigušene eksponencijalne i/ili sinusne komponente.

Sve ove činjenice treba imati u vidu pri izboru modela za konkretne serije.

Pri identifikaciji modela neophodno je, makar i grubo, proveriti da li je opravdano smatrati da je ρ_k praktično jednako nuli za korake veće od date veličine. Zato se može koristiti kriterijum Bartleta koji za disperziju uzoračkih koeficijenata korelacije iz stacionarne gausovske serije daje ocenu:

$$D(r_k) \approx \frac{1}{n}\left[1 + 2\sum_{v=1}^{q} \rho_v^2\right], \; k > q, \tag{36}$$

koja se praktično koristi tako što se u njoj nepoznate veličine ρ_k zamenjuju uzoračkim koeficijentima korelacije r_k i onda se ocenjuje standardna greška sa $s(r_k) = \sqrt{D(r_k)}$.

U slučaju rada sa konkretnim vremenskim serijama postavlja se zadatak ocenjivanja koeficijenata korelacije ρ_k i koeficijenata parcijalne korelacije Φ_{kk}.

Koeficijenti ρ_k se obično ocenjuju njihovim uzoračkim vrednostima:

$$r_k = \frac{\frac{1}{n-k}\sum_{i=1}^{n-k}(X_i - A_k)(X_{i+k} - B_k)}{\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(X_i - A_k)^2 \; \frac{1}{n-k}\sum_{i=1}^{n-k}(X_{i+k} - B_k)^2}}, \tag{37}$$

u kojima je

$$A_k = \frac{1}{n-k}\sum_{i=1}^{n-k} X_i, \; B_k = \frac{1}{n-k}\sum_{i=1}^{n-k} X_{i+k}, \tag{38}$$

a onda se koeficijenti parcijalne korelacije određuju iz relacije (20) u kojoj se umesto ρ_k i Φ_{kk} uzimaju njihove uzoračke ocene. U praksi se obično umesto A_k i B_k koriste uzoračke sredine

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k. \tag{39}$$

Ako je red procesa autokorelacije jednak p, tada su uzoračke parcijalne korelacione funkcije $\hat{\Phi}_{k,k}$ za $k \geq p+1$ približno nezavisne međusobno i imaju disperziju

$$D\left(\hat{\Phi}_{kk}\right) \approx \frac{1}{n}, \; k \geq p+1. \tag{40}$$

To daje njihovu standardnu grešku $s\left(\hat{\Phi}_{kk}\right) \approx 1/\sqrt{n}$, $k \geq p+1$, na osnovu koje se i odlučuje da li je opravdano smatrati da je $\hat{\Phi}_{k,k}$ jednako nuli za $k \geq p+1$.

Opisani kriterijum izbora modela se može iskoristiti na sledeći način:

- Na osnovu vrednosti koeficijenata $r_1, r_2, \dots$ i zaključivanja da su $r_i \approx 0$ za sve $i \geq q+1$ odlučujemo se za red q u MA delu modela;
- Na osnovu vrednosti $\hat{\Phi}_{11}, \hat{\Phi}_{22}, \dots$ i zaključka da se za $j \geq p+1$ svi $\hat{\Phi}_{jj} \approx 0$, odlučujemo se za red p u AR delu modela.

Time smo izabrali model ARMA(p, q). (Primetimo da može biti kako $p=0$ tako i $q=0$).

5. PROVERA POGODNOSTI IZABRANOG MODELA

Za seriju ostataka $\hat{\xi}_t = \varepsilon_t$ se nalaze uzorački koeficijenti korelacije

$r_1(\varepsilon), r_2(\varepsilon), \dots, r_s(\varepsilon)$.

Pri datom n se nalaze trake $\pm 1/\sqrt{n}$ i $\pm 2/\sqrt{n}$ i gleda se da li svi $r_k(\varepsilon)$ upadaju u te trake ili ne. Na osnovu tog rezultata se i donosi odluka da prihvatimo ili da odbacimo predloženi model.

Drugačiji način je da na koeficijente $r_1(\varepsilon), r_2(\varepsilon), \dots, r_s(\varepsilon)$ gledamo kao na jednu celinu i formiramo sumu

$$Q = n \sum_{k=1}^{s} r_k^2(\varepsilon). \tag{41}$$

Može se pokazati [1] da ako se radi o Gausovim serijama, onda pri velikim vrednostima s i tačnoj nultoj hipotezi H_0: da ostaci obrazuju Gausov beli šum, važi sledeća asimptotika

$$Q \approx \chi^2_{s-p-q}, \tag{42}$$

gde su p i q parametri u izabranom ARMA(p, q) modelu. Na osnovu izabranog praga značajnosti α i toga da li uzoračka vrednost veličine Q upada u kritičnu oblast za χ^2_{s-p-q} ili ne, mi i donosimo odluku da li je izabran odgovarajući model. Ovim je opisan Boks-Pirsov test, koji se često zove i portmanteau test (portmanto test).

Pitanje koje se nameće je izbor veličine s - broja koeficijenata korelacije ostataka koji učestvuju u portmanteau statistici. Za male vrednosti s može aproksimacija raspodele Q - statistike χ^2 - raspodelom biti loša. Pri velikim vrednostima s se radi i sa koeficijentima korelacije izračunatim pri dosta smanjenom broju sabiraka u brojiocu u odnosu na imenilac, što može da dovede do suviše malih signifikantnih vrednosti statistike Q. Kompromis se nalazi u izboru s reda $\sqrt{n}$ (ali ne veće od $2\sqrt{n}$). Opisan test je zasnovan na činjenici da je $E(r_k) = 0$, $D(r_k) \approx 1/n$, (za velike vrednosti n), te je statisitika Q upravo dobijena onako kako se formiraju χ^2 - statistike. Prednost ovog testa je da se ne zahteva specifikacija alternativne hipoteze. Pokazalo se da je test dobar za velike uzorke. Međutim, čak i pri uzorcima obima $n = 100$ aproksimacija χ^2 - raspodelom nije zadovoljavajuća.

Zato su Ljung i Box [6] predložili da se uzme preciznija aproksimacija za $D(r_k)$, tj., da se umesto statistike Q posmatra statistika

$$Q^* = n(n+2) \sum_{j=1}^{s} \frac{r_j^2(\varepsilon)}{n-j}, \tag{43}$$

koja, takođe, ima približno χ^2_{s-p-q} raspodelu.

6. ANALIZA REZULTATA

6.1 Karakteristike korišćene serije

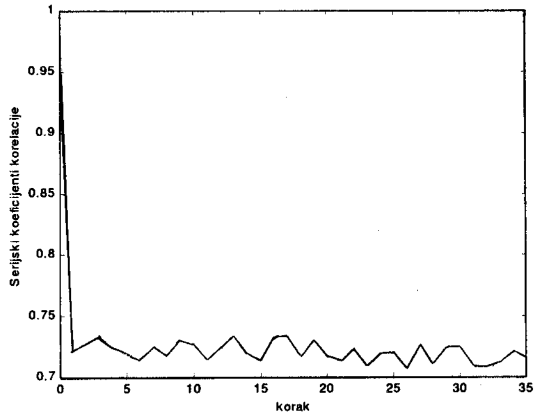
U radu se analizira serija mesečnih suma padavina osmotrenih u periodu od 1888-2000. god. na Meteorološkoj opservatoriji u Beogradu [11]. Za posmatranu vremensku seriju mesečnih suma padavina $\{P_t, t = \overline{1,1356}\}$ dobijene su sledeće karakteristike:

- Srednja vrednost $\overline{P}$ =55.66,
- Standardna devijacija $S = 37.02$,
- Koeficijent asimetrije $C_s = 1.22$,
- min P = 0, max P = 262.5.

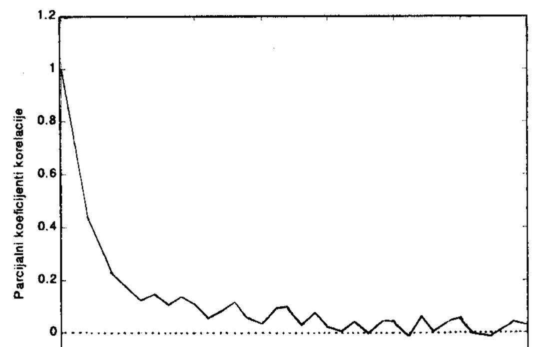
Izvršeno je testiranje empirijske raspodele sa teorijskom normalnom raspodelom (H_0: Serija sledi normalnu raspodelu, H_a: Serija ne sledi normalnu raspodelu).

Na osnovu originalne serije mesečnih suma padavina i podele na 30 klasa, dobijena je statistika hi-kvadrat testa $\chi^2_{exp} = 58.512$. Za prag značajnosti $\alpha = 0.05$ i $\nu = 30 - 2 - 1 = 27$, kritična vrednost hi-kvadrat testa je $\chi^2_{27;0.95} = 40.113$. Kako je $\chi^2_{exp} > \chi^2_{27;0.95}$ zaključuje se da se hipoteza H_0 odbacuje, tj. ne može se smatrati da serija mesečnih suma padavina potiče iz normalne raspodele.

Koeficijenti serijske i parcijalne korelacije za seriju $\{P_t, t = \overline{1,1356}\}$ prikazani su na slikama 1 i 2.



Slika 1: Serijski koeficijenti korelacije za originalnu seriju mesečnih suma padavina $\{P_t\}$



Slika 2: Parcijalni koeficijenti korelacije za originalnu seriju mesečnih suma padavina $\{P_t\}$

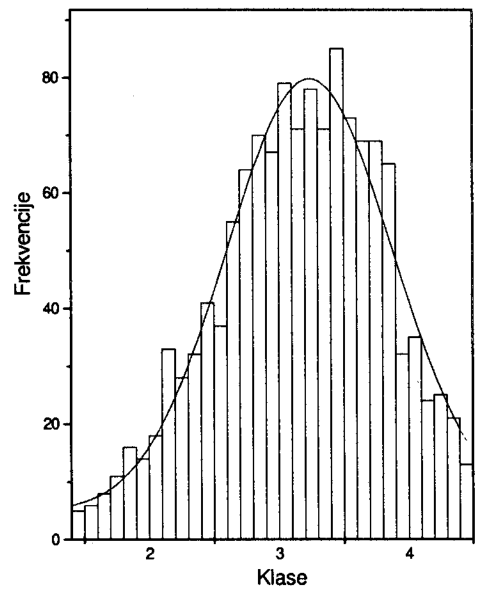
6.2 Izbor transformacije $Y = X^a$

Kako se originalni podaci ne pokoravaju normalnoj raspodeli, pristupilo se traženju transformacije oblika $Y = X^a$ za koju će serija $\{Y_t, t = \overline{1,1356}\}$ imati normalnu raspodelu. Proverom po χ^2 - testu, za razne vrednosti a, pokazalo se da je moguće uzeti $Y = X^{0.3}$. Za seriju $\{Y_t, t = \overline{1,1356}\}$, $\{Y_t = P_t^{0.3}, t = \overline{1,1356}\}$ dobijaju se sledeće karakteristike:

- Srednja vrednost $\overline{Y} = 3.172$,
- Standardna devijacija $S_{(Y)} = 0.704$,
- Koeficijent asimetrije $C_{s(Y)} = -0.312$,
- min Y = 0, max Y = 5.318.

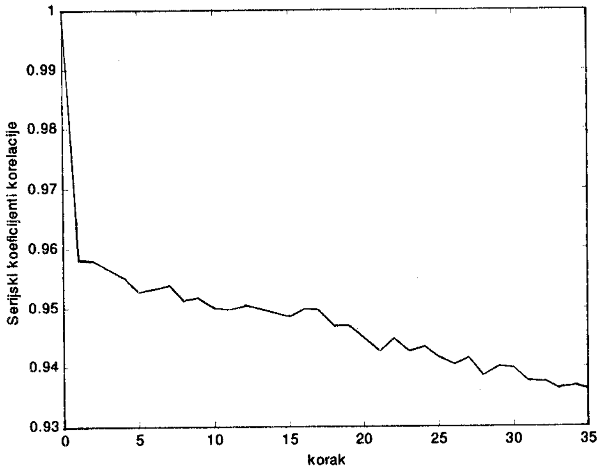
Na osnovu transformisane serije i podele na 31 klasu dobijena je statistika $\chi^2_{exp} = 38.432$. Kako je za prag značajnosti $\alpha = 0.05$ i $\nu = 28$, $\chi^2_{28;0.95} = 41.337$, nulta hipoteza se prihvata. Dakle, serija $\{Y_t, t = \overline{1,1356}\}$, $Y_t = X_t^{0.3}$ potiče iz normalne raspodele.

Empirijske i teorijske frekvencije su prikazane na slici 3.

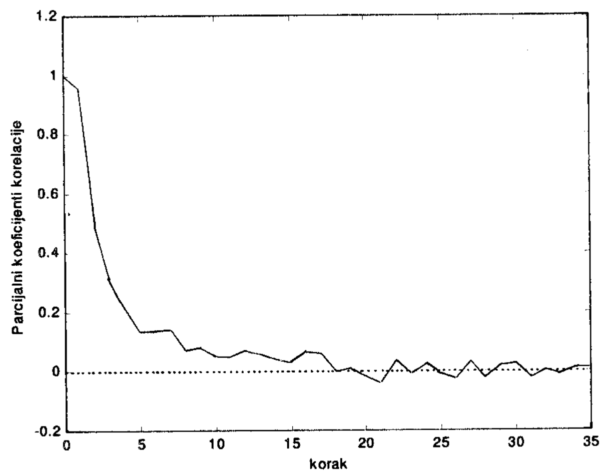


Slika 3: Grafici empirijskih i teorijskih frekvencija za seriju $\{Y_t = P_t^{0.3}, t = \overline{1,1356}\}$

Koeficijenti serijske i parcijalne korelacije za seriju $\{P_t, t = \overline{1,1356}\}$ prikazani su na slikama 4 i 5.



Slika 4: Serijski koeficijenti korelacije za seriju transformisanih mesečnih suma padavina $\{P_t\}$



Slika 5: Parcijalni koeficijenti korelacije za seriju transformisanih mesečnih suma padavina $\{P_t\}$

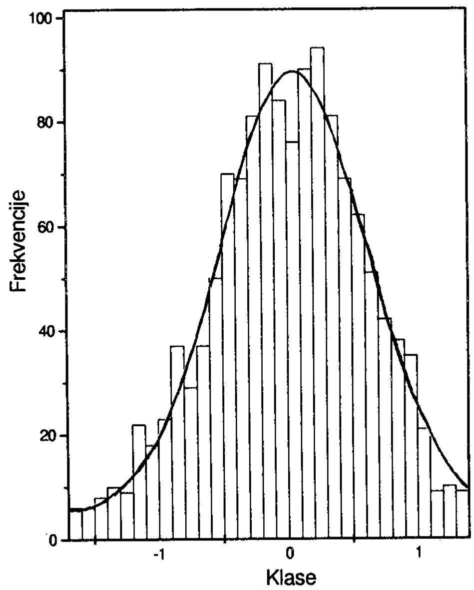
6.3 Desezonizovana serija

Desezonizacija serije $\{Y_t = P_t^{0.3}, t = \overline{1,1356}\}$ je izvršena odstranjivanjem sezonske komponente po metodi konstantnih trendova. Numeričke karakteristike desezonizovane serije $\{Z_t, t = \overline{1,1356}\}$ po metodi konstantnih trendova su:

- Srednja vrednost $\overline{Z} = -1.0152\text{e}-017 = 0$,
- Standardna devijacija $S_{(Z)} = 0.633$,
- Koeficijent asimetrije $C_{s(Z)} = -0.333$,
- $\min Z = -2.9668$, $\max Z = 2.0557$,
- Koeficijenti serijske korelacije: $r_1 = 0.0053$, $r_2 = 0.0371$, $r_3 = 0.0035$, $r_4 = -0.0003$, $r_5 = -0.0278$, $r_6 = -0.0047$, $r_7 = 0.0282$, $r_8 = -0.0014$, $r_9 = 0.0162$, $r_{10} = -0.0287$,
- Koeficijenti parcijalne korelacije: $\alpha_1 = 0.0053$, $\alpha_2 = 0.0371$, $\alpha_3 = 0.0031$, $\alpha_4 = -0.0018$, $\alpha_5 = -0.0281$, $\alpha_6 = -0.0044$, $\alpha_7 = 0.0305$, $\alpha_8 = -0.0011$, $\alpha_9 = 0.0141$, $\alpha_{10} = -0.0300$.

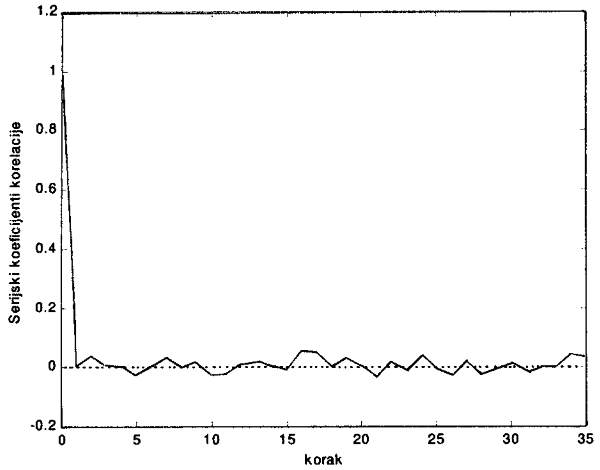
Na osnovu desezonizovane serije $\{Z_t, t = \overline{1,1356}\}$ transformisanih mesečnih suma padavina i podele na 31 klasu dobijena je χ^2 statistika $\chi^2_{\exp} = 30.377$. Kako je kritična vrednost za $\alpha = 0.05$ i $\nu = 28$, sledi da je $\chi^2_{\exp} < \chi^2_{28;0.95}$ pa se hipoteza H_0 prihvata. Dakle, serija $\{Z_t, t = \overline{1,1356}\}$ je jedna Gausovska serija.

Empirijske i teorijske frekvencije su prikazane na sl. 6.

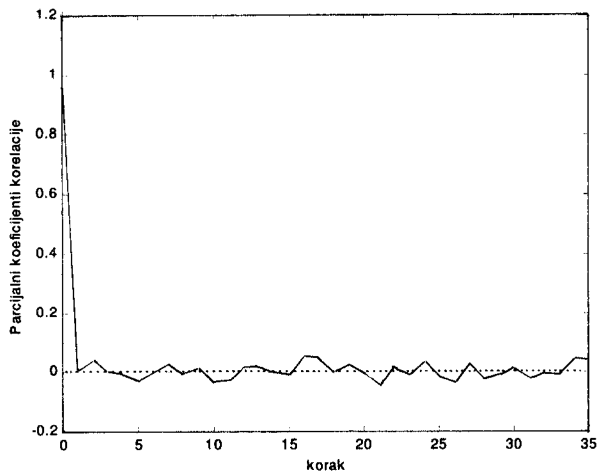


Slika 6: Grafici empirijskih i teorijskih frekvencija za seriju $\{Z_t\}$

Koeficijenti serijske i parcijalne korelacije za seriju $\{Z_t, t = \overline{1,1356}\}$ prikazani su na slikama 7 i 8.



Slika 7: Serijski koeficijenti korelacije za desezonizovanu seriju $\{Z_t\}$



Slika 8: Parcijalni koeficijenti korelacije za desezonizovanu seriju $\{Z_t\}$

6.4 Slučaj izbora MA(1) modela

Za desezonizovanu seriju $\{Z_t, t = \overline{1,1356}\}$ i uzorački koeficijent korelacije r_1 biramo vrednost parametra a iz intervala (-1,1) da bi proces bio obratni. Na osnovu relacije (12) dobijamo $a = 0.0053$. Za seriju ostataka $\{\xi_t, t = \overline{1,1356}\}$ dobijene su sledeće numeričke karakteristike:

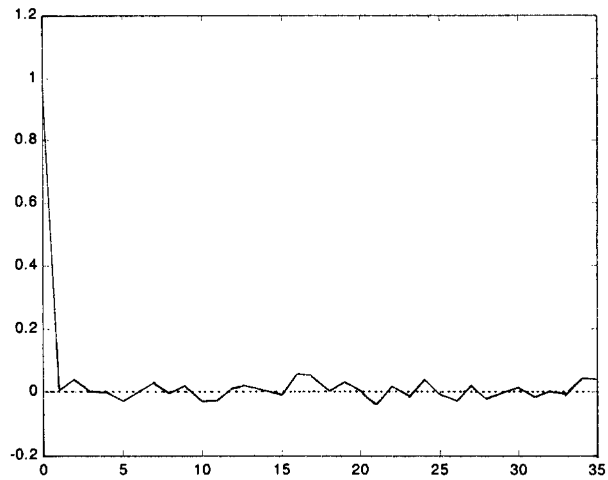
- Srednja vrednost $\overline{\xi} = 1.3602\text{e-}006 = 0$,
- Standardna devijacija $S_{(\xi)} = 0.6326$,
- Koeficijent asimetrije $C_{s(\xi)} = -0.333$,
- $\min \xi = -2.9698$, $\max \xi = 2.0586$,
- Koeficijenti serijske korelacije: $r_1 = -0.0002$, $r_2 = 0.0371$, $r_3 = 0.0033$, $r_4 = -0.0002$, $r_5 = -0.0278$, $r_6 = -0.0047$, $r_7 = 0.0283$, $r_8 = -0.0016$, $r_9 = 0.0163$, $r_{10} = -0.0287$,
- Koeficijenti parcijalne korelacije: $\alpha_1 = -0.0002$, $\alpha_2 = 0.0371$, $\alpha_3 = 0.0033$, $\alpha_4 = -0.0016$, $\alpha_5 = -0.0281$, $\alpha_6 = -0.0047$, $\alpha_7 = 0.0305$, $\alpha_8 = -0.0010$, $\alpha_9 = 0.0142$, $\alpha_{10} = -0.0298$.

Kako se po kriterijumu Bartleta svi koeficijenti serijske i parcijalne korelacije mogu smatrati zanemarljivo malim, zaključuje se da dobijena serija $\{\xi_t, t = \overline{1,1356}\}$ predstavlja beli šum i da se serija $\{Z_t, t = \overline{1,1356}\}$ može predstaviti MA(1) modelom:

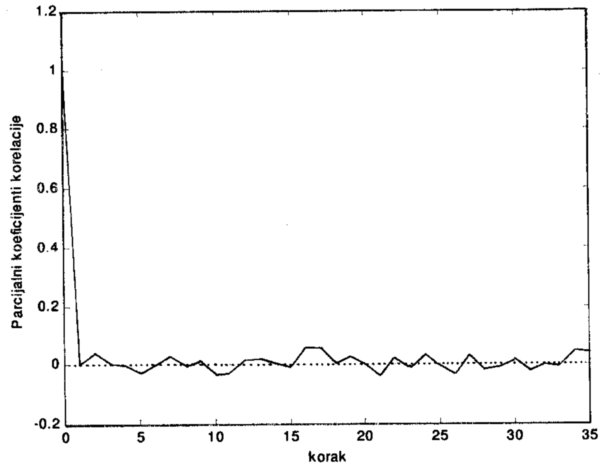
$$Z_t = \xi_t + 0.0053\xi_{t-1}.$$

Takođe, na osnovu Ljung i Boksovog testa za $s = \sqrt{n} = 36$, dobija se vrednost statistike $Q^* = 29.4$. Za prag značajnosti $\alpha = 0.05$, sledi da je $Q^* < \chi^2_{35}$, tj., izabran je odgovarajući model.

Koeficijenti serijske i parcijalne korelacije za seriju $\{\xi_t, t = \overline{1,1356}\}$ prikazani su na slikama 9 i 10.



Slika 9: Serijski koeficijenti korelacije za seriju ostataka $\{\xi_t\}$ dobijenu modelom MA(1)



Slika 10: Parcijalni koeficijenti korelacije za seriju ostataka $\{\xi_t\}$ dobijenu modelom MA(1)

6.5 Slučaj izbora MA(2) modela

Pretpostavimo da biramo model MA(2) oblika. Zamenom podataka iz $\{Z_t\}$ serije dobijamo kompleksne vrednosti koeficijenata a_1 i a_2, što pokazuje da je model neodgovarajući.

6.6 Slučaj izbora AR(1) modela

Parametar b ocenjujemo iz relacije (23), što daje vrednost $b = 0.0053$. Za seriju ostataka $\{\xi_t, t = \overline{1,1356}\}$ dobijene su sledeće numeričke karakteristike:

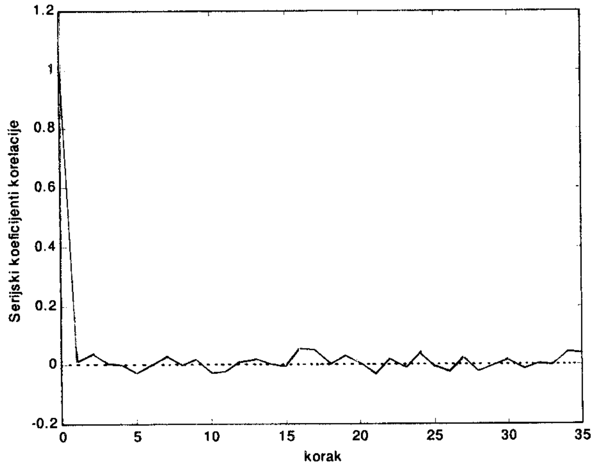
- Srednja vrednost $\bar{\xi} = -1.3868\text{e}-006 = 0$,
- Standardna devijacija $S_{(\xi)} = 0.6326$,
- Koeficijent asimetrije $C_{s(\xi)} = -0.3335$,
- $\min \xi = -2.9638$, $\max \xi = 2.0528$,
- Koeficijenti serijske korelacije: $r_1 = 0.0108$, $r_2 = 0.0371$, $r_3 = 0.0037$, $r_4 = -0.0005$, $r_5 = -0.0278$, $r_6 = -0.0047$, $r_7 = 0.0282$, $r_8 = -0.0012$, $r_9 = 0.0160$, $r_{10} = -0.0287$,
- Koeficijenti parcijalne korelacije: $\alpha_1 = 0.0108$, $\alpha_2 = 0.0371$, $\alpha_3 = 0.0029$, $\alpha_4 = -0.0019$, $\alpha_5 = -0.0281$, $\alpha_6 = -0.0041$, $\alpha_7 = 0.0305$, $\alpha_8 = -0.0012$, $\alpha_9 = 0.0139$, $\alpha_{10} = -0.0302$.

Prema kriterijumu Bartleta svi koeficijenti serijske i parcijalne korelacije mogu se smatrati (po modulu) zanemarljivo malim, zaključuje se da dobijena serija $\{\xi_t, t = \overline{1,1356}\}$ predstavlja beli šum i da se serija $\{Z_t, t = \overline{1,1356}\}$ može predstaviti modelom AR(1) oblika:

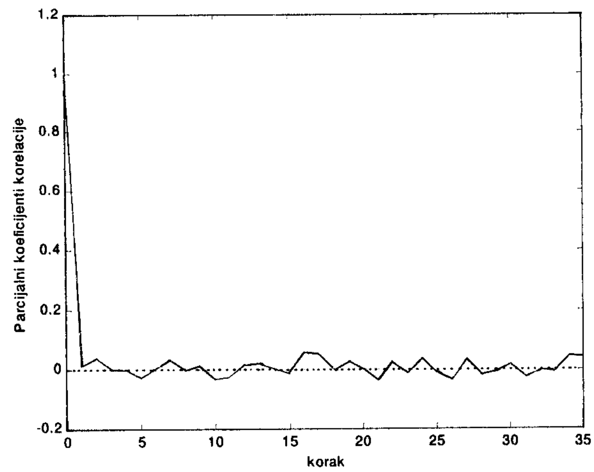
$$Z_t = 0.0053 Z_{t-1} + \xi_t .$$

Na osnovu Ljung i Boksovog testa za $s = 36$, dobija se vrednost statistike $Q^* = 29.8$. Kako je za prag značajnosti $\alpha = 0.05$, $Q^* < \chi^2_{35}$, sledi da je izabran odgovarajući model.

Koeficijenti serijske i parcijalne korelacije za seriju $\{\xi_t, t = \overline{1,1356}\}$ prikazani su na slikama 11 i 12.



Slika 11: Serijski koeficijenti korelacije za seriju ostataka $\{\xi_t\}$ dobijenu modelom AR(1)



Slika 12: Parcijalni koeficijenti korelacije za seriju ostataka $\{\xi_t\}$ dobijenu modelom AR(1)

6.7 Slučaj izbora AR(2) modela

Pretpostavimo da je kao model prepoznat AR(2) model.

Na osnovu relacije (25) dobijamo vrednosti parametara $b_1 = -0.00512$ i $b_2 = -0.03715$.

Za seriju $\{\xi_t, t = \overline{1,1356}\}$ dobijene su sledeće numeričke karakteristike:

- Srednja vrednost $\bar{\xi} = 3.6750\text{e-}005 = 0$,
- Standardna devijacija $S_{(\xi)} = 0.6322$,
- Koeficijent asimetrije $C_{s(\xi)} = -0.3294$,
- $\min \xi = -2.9289$, $\max \xi = 2.0760$,
- Koeficijenti serijske korelacije: $r_1 = -0.0001$, $r_2 = 0.0001$, $r_3 = 0.0041$, $r_4 = -0.0014$, $r_5 = -0.0290$, $r_6 = -0.0046$, $r_7 = 0.0288$, $r_8 = -0.0003$, $r_9 = 0.0162$, $r_{10} = -0.0290$,
- Koeficijenti parcijalne korelacije: $\alpha_1 = -0.0001$, $\alpha_2 = 0.0001$, $\alpha_3 = 0.0041$, $\alpha_4 = -0.0014$, $\alpha_5 = -0.0290$, $\alpha_6 = -0.0047$, $\alpha_7 = 0.0289$, $\alpha_8 = -0.0001$, $\alpha_9 = 0.0163$, $\alpha_{10} = -0.0303$.

Svi koeficijenti serijske i parcijalne korelacije mogu se prema kriterijumu Bartleta smatrati zanemarljivo malim, pa zaključujemo da dobijena serija $\{\xi_t, t = \overline{1,1356}\}$ predstavlja beli šum i da se serija $\{Z_t, t = \overline{1,1356}\}$ može predstaviti modelom AR(2) oblika:

$$X_t = 0.00512 X_{t-1} + 0.03715 X_{t-2} + \xi_t .$$

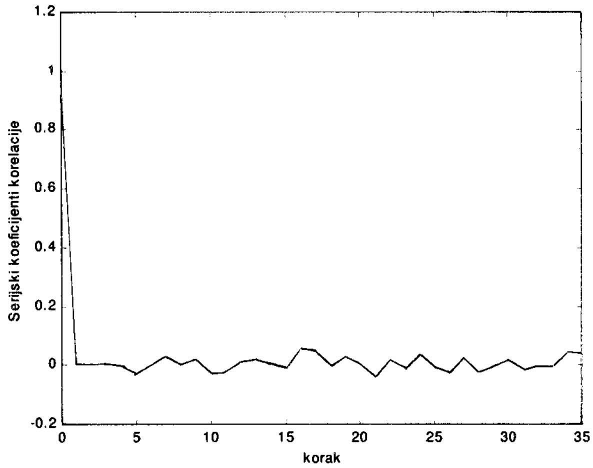
Na osnovu Ljung i Boksovog testa za $s = 36$, dobija se vrednost statistike $Q^* = 28.2$. Kako je za prag značajnosti $\alpha = 0.05$, $Q^* < \chi^2_{34}$, izabran je odgovarajući model.

Koeficijenti serijske i parcijalne korelacije za seriju $\{\xi_t, t = \overline{1,1356}\}$ prikazani su na slikama 13 i 14.

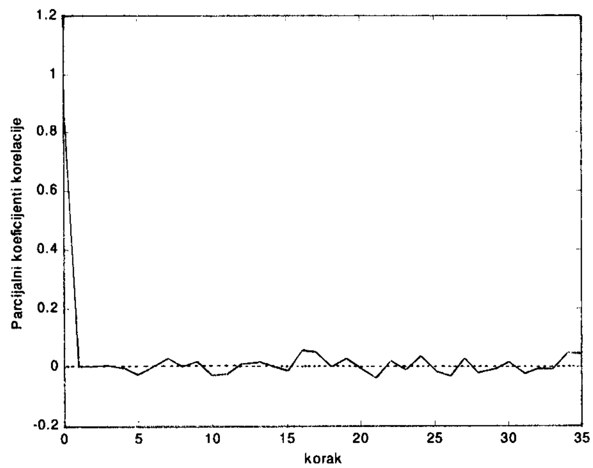
6.8 Slučaj izbora ARMA(1,1) modela

Neka je kao model predložen ARMA(1,1).

Za poznatu seriju $\{Z_t, t = \overline{1,1356}\}$ i uzoračke koeficijente korelacije r_1 i r_2 dobijamo vrednosti koeficijenata ARMA(1,1) koji ne zadovoljavaju uslov za stacionarnost nizova da je $|b_1| < 1$. Dakle, ARMA(1,1) model ne možemo smatrati adekvatnim modelom za mesečne sume padavina osmotrene u Beogradu.



Slika 13: Serijski koeficijenti korelacije za seriju ostataka $\{\xi_t\}$ dobijenu modelom AR(2)



Slika 14: Parcijalni koeficijenti korelacije za seriju ostataka $\{\xi_t\}$ dobijenu modelom AR(2)

ZAKLJUČAK

Predstavljene su tehnike korišćenja ARMA procesa za modeliranje meteoroloških vremenskih serija. Ove tehnike dovode do izbora ARMA procesa koji najviše odgovaraju podacima. Detaljno je razmatrana meteorološka primena ARMA procesa za modeliranje mesečnih suma padavina u Beogradu. Prezentovana je implementacija postupka koja identifikuje odgovarajući izbor ARMA procesa. Pokazano je da AR procesi nižeg reda adekvatno fituju takve vremenske serije i da se opštiji ARMA procesi ne moraju razmatrati.

LITERATURA

[1] Box, G. E. P. and Jenkins, G. M.: Time series analysis: Forecasting and control, Holden-Day, 575 pp, 1976.

[2] Davis, J. M. and Rappoport, P. N.: The use of time series analysis techniques in forecasting meteorological drought, Mon. Wea. Rev., Vol. 102, pp. 176-180, 1974.

[3] Delleur, J. and Kavvas, M. L.: Stochastic models for monthly rainfall forecasting and synthetic generation, J. Appl. Meteor., Vol. 17, pp. 1528-1536, 1978.

[4] Katz, R. W. and Skaggs, R. H: On the use of autoregressive-moving average processes to model meteorological time series, Mon. Wea. Rev., Vol. 109, pp. 479-484, 1981.

[5] Ledolter, J.: The analysis of multivariate time series applied to problems in hydrology, J. Hydrology, Vol. 36, pp. 327-352, 1977.

[6] Ljung, G. M. and Box, G. E. P.: The likelihood function of stationary ARMA models, Biometrika, Vol. 66, pp. 265-270, 1979.

[7] Lungu, E. M. and Sefe, F. T. K.: Stohastic analysis of monthly streamflows, J. Hydrology, Vol. 126, pp. 171-182, 1991.

[8] Mališić, J.: Vremenske serije, Matematički fakultet, Beograd, 304 str, 2002.

[9] Mališić, J., Jevremović, V. i Dmitrović, I.: ARMA modeliranje hidroloških serija, Vodoprivreda, Vol. 30, pp. 203-212, 1998.

[10] Matyasovszky, I.: Further results of the analysis of Central England temperature data, Theor. Appl. Climatol., Vol. 39, pp. 126-136, 1988.

[11] RHMZ Srbije: Rezultati osmatranja Meteoroloske opservatorije u Beogradu u periodu 1887-1986, Republički hidrometeorološki zavod Srbije, 131 str, 1987.

[12] Strikanthan, R., McMahon, T. A. and Irish, J. L.: Time series analysis of annual flows of Australian streams, J. Hydrology, Vol. 66, pp. 213-226, 1983.

MODELING OF MONTHLY PRECIPITATION SUMS BY THE TIME SERIES METHOD

by

MSc Ivana TOŠIĆ[1], Prof. Jovan MALIŠIĆ[2], Prof. Miroslava UNKAŠEVIĆ[1]
[1]Institute of Meteorology, School of Physics, Belgrade
[2]Faculty of Mathematics, Belgrade

Summary

The paper presents the use of ARMA processes to model meteorological time series. The technique used increases the likelihood of choosing the most appropriate ARMA process to model such data. The implementation of a procedure that identifies the choice of the appropriate ARMA process was demonstrated for monthly precipitation sums observed at the Observatory-Belgrade during the period 1888-2000. It was shown that low-order AR processes adequately fit such time series and that more general ARMA processes need not be considered.

Key words: meteorological time series, ARMA models, modeling, serial correlation, partial correlation

Redigovano 04.11.2003.