Dugoročna zavisnost

Grgić, Lucijana

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Department of Mathematics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:126:480768

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**



Repository / Repozitorij:

Repository of School of Applied Mathematics and Computer Science







Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Lucijana Grgić

Dugoročna zavisnost

Diplomski rad

Osijek, 2018.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Lucijana Grgić

Dugoročna zavisnost

Diplomski rad

Mentor: doc. dr. sc. Danijel Grahovac

Osijek, 2018.

Sadržaj

Uvod **1**

1 Stacionarni procesi s dugoročnim pamćenjem **2**

1.1 Definicije i osnovni rezultati 2

1.2 Prirasti sebi–sličnih procesa 6

2 FARIMA model **11**

2.1 Definicija i osnovni rezultati 12

2.2 Prediktor FARIMA modela 15

3 Statistički procjenitelji dugoročne zavisnosti **17**

3.1 R/S statistika 17

3.2 Korelogram 19

3.3 Metoda varijance 22

3.4 Periodogram 23

3.5 Whittleov aproksimativni ML procjenitelj 25

4 Modeliranje stvarnih podataka FARIMA modelom **26**

4.1 Minimalni vodostaj rijeke Nil 26

4.1.1 Opis podataka 26

4.1.2 Identifikacija i dijagnostika modela 27

4.1.3 Predviđanje modela 32

4.2 Globalna temperatura sjeverne polutke 34

4.2.1 Opis podataka 34

4.2.2 Identifikacija i dijagnostika modela 35

Literatura **40**

Sažetak **41**

Summary **42**

Životopis **43**

Uvod

Jedna od važnijih primjena statistike je modeliranje prirodnih pojava i procesa u raznim područjima poput hidrologije, geografije, ekonomije, klimatologije i agronomije. Kako se u tim područjima radi o procesima koje se ne može kontrolirati i koje pokreću kompliciranі fizički mehanizmi i zakoni, zavisnost među promatranjima u različitim vremenskim trenutcima je nešto što očekujemo. Znanstvenici u području primjenjene statistike primijetili su kako zavisnost među promatranjima, osim što se pojavljuje, može trajati duži period, odnosno prošlost u nekim procesima može imati dugoročan utjecaj na budućnost. Da je pojava takozvanog dugoročnog pamćenja bila poznata puno prije nego su odgovarajući stohastički modeli razvijeni, pokazuje poznati hidrolog Hurst (1951) koji je primijetio ovo svojstvo istraživajući kako regulirati tok rijeke Nil. U Knjizi Postanka 41, 29–30, čitamo: *"Dolazi, evo, sedam godina velikog obilja svoj zemlji egipatskoj. A poslije njih nastat će sedam gladnih godina, kada će se zaboraviti sve obilje u zemlji egipatskoj."* Podatke vodostaja rijeke Nil iz vremena prije Krista nemamo, no dostupni podaci iz 622. godine pa do 1281. godine poslije Krista pokazuju ponašanje koje bi moglo biti dobro "objašnjenje" za sedam dobrih i sedam loših godina iz Knjige Postanka. Naime, postoje dugi periodi gdje je maksimalan vodostaj rijeke Nil bio visok, te dugi periodi kada je isti bio vrlo nizak. Kako je navedeni citat iz Knjige Postanka bio interpretacija faraonova sna o sedam lijepih i sedam ružnih krava koju je izrekao sluga Josip, francuski matematičar Mandelbrot (1977) je, osvrćući se na Bibliju, ovo ponašanje prozvao *Josipov efekt*.

U ovom diplomskom radu upoznat ćemo stacionarne procese s dugoročnim pamćenjem ili dugoročnom zavisnošću koji dobro opisuju procese kod kojih je prošlost dugoročno povezana s budućnosti. Definicije i osnovna svojstva takvih procesa uvodimo u prvom poglavlju kao i primjer značajne klase procesa koji su stacionarni procesi s dugoročnim pamćenjem. U drugom poglavlju donosimo definiciju i rezultate o drugoj značajnoj klasi procesa s dugoročnom zavisnošću, tj. o frakcionalnim ARIMA procesima. U trećem poglavlju upoznajemo statističke procjenitelje dugoročne zavisnosti koji će nam u četvrtom poglavlju poslužiti za uočavanje dugoročne zavisnosti kod vremenskog niza o godišnjem minimalnom vodostaju rijeke Nil i vremenskog niza o globalnoj temperaturi sjeverne polutke Zemlje koje ćemo modelirati frakcionalnim ARIMA procesom.

1 Stacionarni procesi s dugoročnim pamćenjem

U analizi vremenskih nizova, svrha nam je slučajnim procesom $\{X_t, t \in T\}$ modelirati stohastički mehanizam koji dovodi do realizacija koje čine vremenski niz. Jedna od važnijih karakteristika vremenskog niza je zavisnost čiju strukturu promatramo u funkciji autokorelacija procesa, tj.

$$\rho(t,s) = \mathrm{Corr}(X_t, X_s) = \frac{\gamma(t,s)}{\sqrt{\gamma(t,t)\gamma(s,s)}},\ t,s \in T,$$

gdje je $\gamma(t,s) = \mathrm{Cov}(X_t, X_s)$ funkcija autokovarijanci procesa. Za stacionarne procese funkcija autokovarijanci ovisi samo o vremenskom pomaku, tj. funkcija je jedne varijable zbog čega za takve procese pišemo $\gamma(h) = \gamma(0,h) = \mathrm{Cov}(X_t, X_{t+h})$. Tada je funkcija autokorelacija stacionarnog procesa jednaka

$$\rho(h) = \rho(0,h) = \frac{\gamma(h)}{\gamma(0)}, \quad h \in \mathbb{Z}.$$

1.1 Definicije i osnovni rezultati

U ovom ćemo poglavlju predstaviti slučajne procese koji su stacionarni, no čija zavisnost među trenutcima opada neuobičajeno sporo kako se ti trenutci udaljavaju. Funkcija autokorelacija takvog procesa opada u nulu toliko sporo da nije apsolutno sumabilna, odnosno vrijedi:

$$\sum_{h=-\infty}^{\infty} |\rho(h)| = \infty.$$

Sljedeća definicija definira takve procese (vidi [2, str.42]).

Definicija 1.1 (Stacionaran proces s dugoročnim pamćenjem) *Neka je $\{X_t\}$ stacionaran proces za koji postoje realni brojevi $\alpha \in (0,1)$ i $c_\rho > 0$ takvi da vrijedi*

$$\lim_{h\to\infty} \frac{\rho(h)}{c_\rho h^{-\alpha}} = 1. \tag{1.1}$$

Tada $\{X_t\}$ zovemo ***stacionaran proces s dugoročnim pamćenjem ili dugoročnom zavisnošću****.*

U svrhu interpretacije (1.1) definiramo asimptotsku jednakost funkcija iz [10, str.4].

Definicija 1.2 (Asimptotska jednakost) *Za kompleksne funkcije f i g kažemo da su* ***asimptotski jednake*** *kada $z \to z_0$ ako je*

$$\lim_{z\to z_0} \frac{f(z)}{g(z)} = 1,$$

i pišemo $f \sim g$, kada $z \to z_0$.

Prema tome, stacionaran proces $\{X_t\}$ je stacionaran proces s dugoročnim pamćenjem ukoliko za $h \to \infty$ vrijedi da je

$$\rho(h) \sim c_\rho h^{-\alpha},$$

za neke realne brojeve $c_\rho > 0$ i $\alpha \in (0, 1)$. Primijetimo da u slučaju $\alpha > 1$, vrijedi da je

$$\sum_{h=-\infty}^{\infty} |\rho(h)| < \infty.$$

Stacionarne procese koji imaju apsolutno sumabilnu autokorelacijsku funkciju nazivat ćemo stacionarnim procesima s kratkoročnim pamćenjem ili kratkoročnom zavisnošću.

Ekvivalentna analizi stacionarnih procesa u vremenskoj domeni je analiza stacionarnih procesa u frekvencijskoj domeni. Dok je u vremenskoj domeni analiza procesa temeljena na funkciji autokovarijanci procesa, u frekvencijskoj domeni ona je temeljena na spektralnoj reprezentaciji procesa, točnije, spektralnoj dekompoziciji procesa koja je analogon Fourierove reprezentacije determinističkih funkcija.
Pretpostavimo da je $\{X_t\}$ stacionaran proces s autokorelacijskom funkcijom γ za koju vrijedi $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. **Spektralna gustoća** procesa $\{X_t\}$ je funkcija $f : \mathbb{R} \to \mathbb{C}$ definirana s

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h). \tag{1.2}$$

Kako je $|e^{-ih\lambda}| = |\cos(h\lambda) - i \sin(h\lambda)| = \sqrt{\cos^2(h\lambda) + \sin^2(h\lambda)} = 1$, zbog sumabilnosti funkcije γ funkcija f definirana s (1.2) je dobro definirana. Funkcije sin i cos su periodične funkcije s periodom 2π pa je stoga i f periodična s periodom 2π i dovoljno ju je promatrati za $\lambda \in (-\pi, \pi]$. Svojstva tako definirane funkcije iskazana su u sljedećoj propoziciji.

Propozicija 1.1 (Svojstva spektralne gustoće)

(i) $f(\lambda) = f(-\lambda)$,

(ii) $f(\lambda) \geq 0$, $\forall \lambda \in (-\pi, \pi]$,

(iii) $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$.

Dokaz. Vidi [4, str.112]. □

Primijetimo da u slučaju apsolutno sumabilne funkcije autokorelacija procesa spektralna gustoća procesa uvijek postoji. No, tema ovog rada su procesi čija funkcija autokorelacija ne zadovoljava tu pretpostavku. U takvom slučaju spektralna gustoća procesa postoji ako su zadovoljene dvije pretpostavke iskazane u sljedećoj definiciji.

Definicija 1.3 (Spektralna gustoća) *Funkcija f je* ***spektralna gustoća*** *stacionarnog procesa $\{X_t\}$ s funkcijom autokovarijanci $\gamma(\cdot)$ ako je:*

(i) $f(\lambda) \geq 0$, $\forall \lambda \in (0, \pi]$,

(ii) $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$, $\forall h \in \mathbb{Z}$.

Takva funkcija f je jedinstvena do na jednakost skoro svuda, pa je poznavati funkciju autokovarijanci procesa ekvivalentno poznavanju spektralne gustoće procesa ukoliko ona postoji (vidi [4, Poglavlje 4.1]). Stoga, definiciju stacionarnog procesa s dugoročnim pamćenjem možemo iskazati preko spektralne gustoće na sljedeći način.

Definicija 1.4 (Stacionaran proces s dugoročnim pamćenjem) *Neka je $\{X_t\}$ stacionaran proces za koji postoji spektralna gustoća f i realni brojevi $\beta \in (0, 1)$ i $c_f > 0$ takvi da vrijedi*

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{c_f |\lambda|^{-\beta}} = 1. \tag{1.3}$$

Tada $\{X_t\}$ zovemo ***stacionaran proces s dugoročnim pamćenjem ili dugoročnom zavisnošću.***

Gledajući stvari u frekvencijskoj domeni, ova definicija govori kako je stacionaran proces $\{X_t\}$ čija spektralna gustoća postoji i ima pol u točki 0 u iznosu $c_f|\lambda|^{-\beta}$ za $c_f > 0$ i $0 < \beta < 1$, proces s dugoročnim pamćenjem.

Definicija 1.1 i Definicija 1.4 su ekvivalentne u sljedećem smislu:

Teorem 1.1 *Neka je $\{X_t\}$ stacionaran proces s funkcijom autokovarijanci γ koja je monotona na nekom intervalu (h_0, ∞), te neka postoji spektralna gustoća f toga procesa.*

(i) Ako (1.1) vrijedi za $0 < \alpha = 2 - 2H < 1$, onda spektralna gustoća f zadovoljava

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{c_f(H)|\lambda|^{1-2H}} = 1, \tag{1.4}$$

gdje je

$$c_f = \sigma^2 \pi^{-1} c_\rho \Gamma(2H - 1) \sin(\pi - \pi H)$$

i $\sigma^2 = \mathrm{Var}(X_t)$.

(ii) Ako (1.3) vrijedi za $0 < \beta = 2H - 1 < 1$, onda funkcija autokovarijanci γ zadovoljava

$$\lim_{h \to \infty} \frac{\rho(h)}{c_\rho h^{2H-2}} = 1, \tag{1.5}$$

gdje je

$$c_\rho = \frac{c_\gamma}{\sigma^2}$$

i

$$c_\gamma = 2c_f\Gamma(2-2H)\sin(\pi H - \frac{1}{2}\pi).$$

Ako u pretpostavkama Teorema 1.1 umjesto asimptotske jednakosti funkcija u jednadžbama (1.1) i (1.3) stoji jednakost, tvrdnje Teorema 1.1 su klasični rezultati o Fourierovim nizovima koje možemo naći u [13]. No u slučajevima asimptotske jednakosti, kao što je to u Teoremu 1.1, tvrdnje ne vrijede uvijek. Tako [2, Teorem 2.1] nije istinit općenito o čemu govore rezultati članka [6]. Iz tog razloga, u Teoremu 1.1 pretpostavljamo postojanje spektralne gustoće i monotonost funkcije autokovarijanci na nekom intervalu.

Umjesto parametara α i β, u nastavku ćemo koristiti parametar H koji je definiran na način kao što je to učinjeno u Teoremu 1.1. U terminima parametra H, dugoročna zavisnost se pojavljuje za $\frac{1}{2} < H < 1$.

Bitno je primijetiti kako su definicije stacionarnih procesa s dugoročnim pamćenjem (1.1) i (1.3) asimptotske definicije. Definicija 1.1 opisuje granično ponašanje funkcije autokorelacija procesa kada korak h teži u beskonačnost, no ne govori kako funkcija autokorelacija procesa izgleda za određeni konačni korak h. Štoviše, vrijednost autokorelacijske funkcije procesa može biti proizvoljno mala za svaki korak h pojedinačno, a da je njezino opadanje u nulu neuobičajeno sporo.

Još jednu karakteristiku stacionarnih procesa s dugoročnim pamćenjem donosi sljedeći teorem iz [1].

Teorem 1.2 *Neka je $\{X_t\}$ stacionaran proces s dugoročnim pamćenjem. Tada vrijedi*

$$\lim_{n\to\infty} \frac{\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)}{c_\gamma n^{2H}} = \frac{1}{H(2H-1)}. \tag{1.6}$$

Kako je $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$, lijeva strana jednakosti (1.6) je jednaka

$$\lim_{n\to\infty} \frac{\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)}{c_\gamma n^{2H}} = \lim_{n\to\infty} \frac{\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)}{c_\gamma n^{2H-2}} = \lim_{n\to\infty} \frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{c_\gamma n^{-(2-2H)}} = \lim_{n\to\infty} \frac{\mathrm{Var}\left(\overline{X}_n\right)}{c_\gamma n^{-(2-2H)}}.$$

Zbog definiranja $0 < \alpha = 2 - 2H < 1$ u Teoremu 1.1, slijedi da nam Teorem 1.2 govori kako je varijanca aritmetičke sredine stacionarnog procesa s dugoročnim pamćenjem

$\{X_t\}$ asimptotski proporcionalna $n^{-\alpha}$, za neki $0 < \alpha < 1$.
Za stacionarne procese s kratkoročnim pamćenjem varijanca aritmetičke sredine ima sljedeće svojstvo (vidi [4]).

Propozicija 1.2 *Ako je $\{X_t\}$ stacionaran proces s funkcijom autokovarijanci γ za koju vrijedi $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, tada*

$$\frac{1}{n}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \longrightarrow \sum_{|h|<\infty} \gamma(h), \quad n \to \infty.$$

Drugim riječima, stopa opadanja varijance aritmetičke sredine stacionarnog procesa s kratkoročnim pamćenjem bit će n^{-1}, dok će kod procesa s dugoročnim pamćenjem ista biti sporija, točnije, jednaka $n^{-\alpha}$, za neki $0 < \alpha < 1$.

Nakon što smo uveli definiciju stacionarnih procesa s dugoročnim pamćenjem i naveli neka njegova kvantitativna svojstva, u sljedećem potpoglavlju ćemo prikazati značajnu klasu modela za koje vrijedi (1.1), tj. klasu modela koji su stacionarni procesi s dugoročnim pamćenjem.

1.2 Prirasti sebi–sličnih procesa

Osnovna ideja sebi–sličnosti potječe iz 17. stoljeća. U deterministčkom pogledu, geometrijski oblik je sebi–sličan ako možemo uočiti istu geometrijsku strukturu bez obzira iz koje udaljenosti gledamo geometrijski oblik. U stohastičkom pogledu, sebi–sličnost procesa se definira u terminima distribucije procesa na sljedeći način.

Definicija 1.5 (Sebi–sličan proces) *Slučajan proces $\{Y_t, t \geq 0\}$ je H–**sebi–sličan** ili **sebi–sličan s parametrom sebi–sličnosti** H za neki $H > 0$ ako za njegove konačnodimenzionalne distribucije vrijedi*

$$\left(c^H Y_{t_1}, c^H Y_{t_2}, \ldots, c^H Y_{t_n}\right) \stackrel{d}{=} \left(Y_{ct_1}, Y_{ct_2}, \ldots, Y_{ct_n}\right),$$

za svaki $c > 0$ i proizvoljne vremenske trenutke $0 < t_1 < t_2 < \cdots < t_n$.

Sada pretpostavimo da je $\{Y_t, t \geq 0\}$, $Y_0 = 0$ *g.s.*, sebi–sličan proces s parametrom sebi–sličnosti H. Nadalje, neka su njegovi prirasti stacionarni, tj. distribucija od $Y_{t+h} - Y_t$ ne ovisi o vremenskom trenutku t već samo o duljini intervala h, te neka je $\mathrm{E}(Y_t) = 0$ radi jednostavnosti zapisa.
Označimo sa $\sigma^2 := \mathrm{Var}(Y_t - Y_{t-1}) = \mathrm{E}[(Y_t - Y_{t-1})^2] = \mathrm{E}[Y_1^2]$. Tada je za $s < t$ vrijedi

$$\mathrm{E}[(Y_t - Y_s)^2] = \mathrm{E}[(Y_{t-s} - Y_0)^2] = \mathrm{E}[\left((t-s)^H Y_1\right)^2] = (t-s)^{2H}\sigma^2 \qquad (1.7)$$

zbog sebi–sličnosti procesa $\{Y_t\}$ i stacionarnosti njegovih prirasta.
S druge strane imamo

$$\mathrm{E}[(Y_t - Y_s)^2] = \mathrm{E}[Y_t^2] + \mathrm{E}[Y_s^2] - 2\mathrm{E}[Y_t Y_s] = \sigma^2 t^{2H} + \sigma^2 s^{2H} - 2\gamma_Y(t,s),$$

gdje je $\gamma_Y = \mathrm{Cov}(Y_t, Y_s)$. Zbog toga slijedi

$$\gamma_Y(t,s) = \frac{1}{2}\sigma^2 \left[t^{2H} - (t-s)^{2H} + s^{2H}\right]. \tag{1.8}$$

Neka je $\{X_i, i \in \mathbb{N}\}$, za $X_i = Y_i - Y_{i-1}$, proces jednokoračnih prirasta procesa $\{Y_t\}$ u diskretnom vremenu. Primijetimo da je tada σ^2 varijanca procesa $\{X_i\}$. Kako je $\{X_i\}$ stacionaran proces, funkcija autokovarijanci za korak $h > 0$ je jednaka

$$\begin{aligned}
\gamma(h) &= \mathrm{Cov}(X_i, X_{i+h}) = \mathrm{Cov}(X_1, X_{h+1}) = \mathrm{E}(X_1 X_{h+1}) \\
&= \mathrm{E}\left[X_1 X_{h+1} + X_1(X_2 + \cdots + X_h) - X_1(X_2 + \cdots + X_h)\right] \\
&= \frac{1}{2}\mathrm{E}\left[2X_1 \sum_{j=2}^{h+1} X_j - 2X_1 \sum_{j=2}^{h} X_j\right] \\
&= \frac{1}{2}\mathrm{E}\left[X_1^2 + 2X_1 \sum_{j=2}^{h+1} X_j + \left(\sum_{j=2}^{h+1} X_j\right)^2 - \left(\sum_{j=2}^{h+1} X_j\right)^2\right. \\
&\quad \left. - X_1^2 - 2X_1 \sum_{j=2}^{h} X_j - \left(\sum_{j=2}^{h} X_j\right)^2 + \left(\sum_{j=2}^{h} X_j\right)^2\right] \\
&= \frac{1}{2}\mathrm{E}\left[\left(\sum_{j=1}^{h+1} X_j\right)^2 - \left(\sum_{j=2}^{h+1} X_j\right)^2 - \left(\sum_{j=1}^{h} X_j\right)^2 + \left(\sum_{j=2}^{h} X_j\right)^2\right] \\
&= \frac{1}{2}\mathrm{E}[(Y_{h+1} - Y_0)^2] - \mathrm{E}[(Y_h - Y_0)^2] - \mathrm{E}[(Y_h - Y_0)^2] + \mathrm{E}[(Y_{h-1} - Y_0)^2].
\end{aligned} \tag{1.9}$$

Koristeći (1.7) dobivamo formulu

$$\gamma(h) = \frac{1}{2}\sigma^2[(h+1)^{2H} - 2h^{2H} + (h-1)^{2H}], \tag{1.10}$$

za $h \geq 0$ i $\gamma(h) = \gamma(-h)$ za $h < 0$. Odnosno, autokorelacije procesa $\{X_i\}$ su dane s

$$\rho(h) = \frac{1}{2}[(h+1)^{2H} - 2h^{2H} + (h-1)^{2H}], \tag{1.11}$$

za $h \geq 0$ i $\rho(h) = \rho(-h)$ za $h < 0$.

Kako bismo uočili asimptotsko ponašanje ove autokorelacijske funkcije, primijetimo da je $\rho(h) = \frac{1}{2}h^{2H}g(h^{-1})$ gdje je $g(x) = (1+x)^{2H} - 2 + (1-x)^{2H}$. Taylorov razvoj funkcije g u okolini točke x_0 do trećeg člana je

$$g(x) = g(x_0) + \left[2H(1+x_0)^{2H-1} - 2H(1-x)^{2H-1}\right](x-x_0) + \frac{1}{2}\left[2H(2H-1)(1+x_0)^{2H-2} - 2H(2H-1)(1-x_0)^{2H-2}\right](x-x_0)^2.$$

Za $0 < H < 1$ i $H \neq \frac{1}{2}$, prvi član u Taylorovom razvoju funkcije g u okolini točke 0 koji je različit od nula je jednak $2H(2H-1)x^2$. Stoga, kada $h \to \infty$, $\rho(h)$ je asimptotski jednaka $H(2H-1)h^{2H-2}$, tj. vrijedi

$$\frac{\rho(h)}{H(2H-1)h^{2H-2}} \to 1 \tag{1.12}$$

kada $h \to \infty$. Time smo dokazali sljedeće.

Propozicija 1.3 *Neka je $\{Y_t, t \geq 0\}$ sebi–sličan proces s parametrom sebi–sličnosti H, stacionarnim prirastima i konačnom varijancom $\mathrm{Var}(Y_t) < \infty$ za svaki $t \geq 0$. Tada za autokorelacijsku funkciju procesa $\{X_i, i \in \mathbb{N}\}$, gdje je $X_i = Y_i - Y_{i-1}$, vrijedi*

$$\lim_{h\to\infty} \frac{\rho(h)}{H(2H-1)h^{2H-2}} = 1.$$

U slučaju kada je $\frac{1}{2} < H < 1$, autokorelacijska funkcija procesa $\{X_i\}$ iz Propozicije 1.3 nije sumabilna. Tada je po Definiciji 1.1 proces $\{X_i\}$ stacionaran proces s dugoročnim pamćenjem.

Primjer 1.1 (**Frakcionalno Brownovo gibanje i frakcionalni Gaussovski šum**) *Neka je $\{Y_t, t \geq 0\}$, $Y_0 = 0$ g.s., sebi–sličan proces sa stacionarnim prirastima. Neka je $\mathrm{E}[X_i] = \mathrm{E}[Y_i - Y_{i-1}] = 0$, za svaki $i = 1, 2, \ldots$. Tada je funkcija autokovarijanci procesa $\{X_i\}$ dana s (1.10). Nadalje, pretpostavimo da je $\{X_t\}$ Gaussovski proces. Tada je distribucija procesa potpuno određena očekivanjem i funkcijom autokovarijanci. Stoga, za svaku vrijednost $H \in (0,1)$ imamo točno jedan Gaussovski proces $\{X_i\}$ koji je stacionaran proces prirasta sebi–sličnog procesa $\{Y_t\}$. Tako definiran proces $\{X_t\}$ se naziva **frakcionalni Gaussovski šum**. Pripadajući sebi–sličan proces $\{Y_t\}$ se naziva frakcionalno Brownovo gibanje koje ćemo definirati na sljedeći način.*

Definicija 1.6 (Frakcionalno Brownovo gibanje) *Neka je* $B^H = \{B_t^H, t \geq 0\}$ *slučajan proces u neprekidnom vremenu za koji vrijedi sljedeće:*

(i) B^H *je Gaussovski proces,*

(ii) $B_0^H = 0$ *g.s.,*

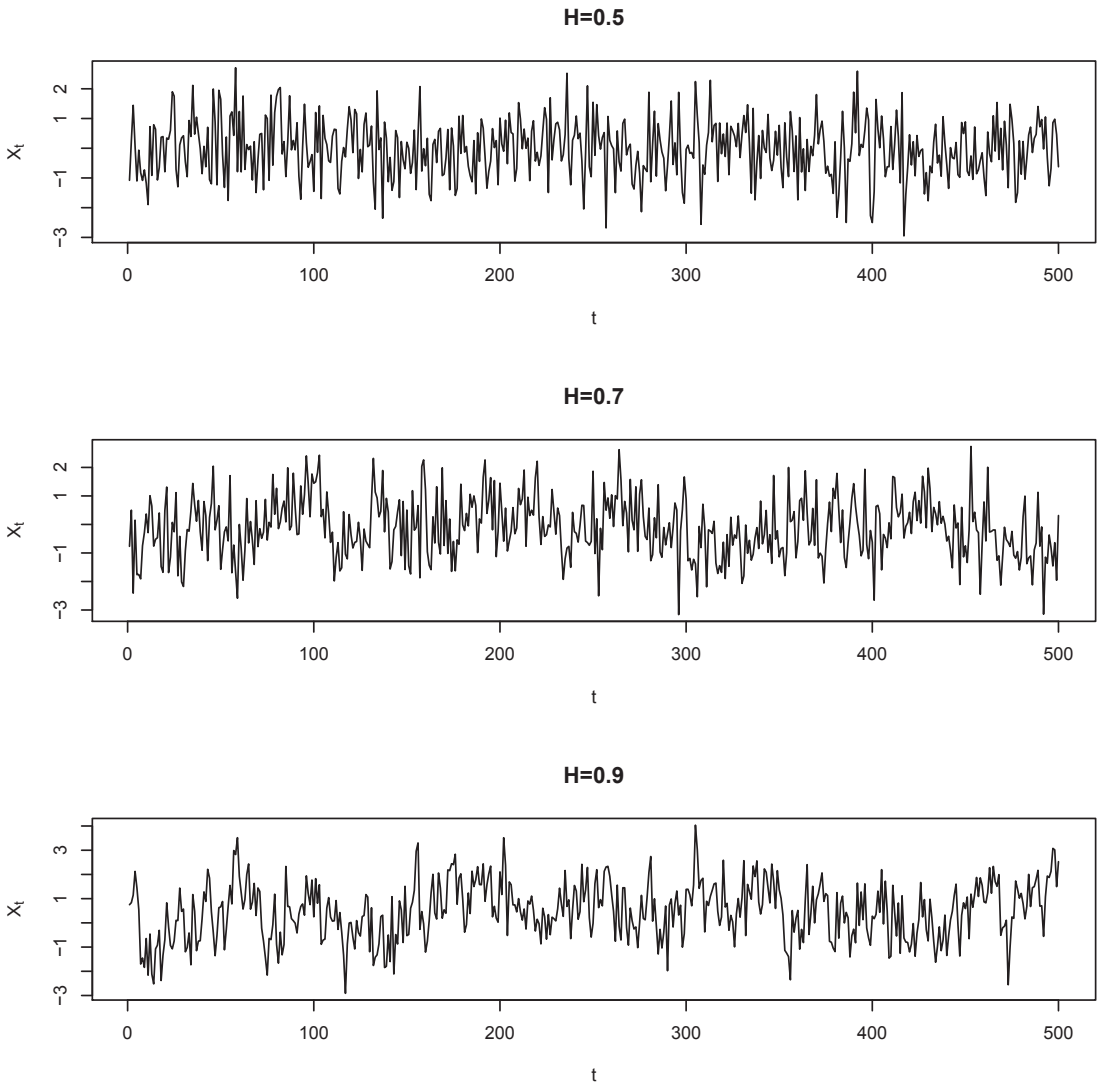
(iii) B^H *ima stacionarne priraste,*

(iv) B^H *je sebi–sličan proces s parametrom sebi-sličnosti* H*,*

(v) $\mathrm{Cov}(B_t^H, B_s^H) = \frac{1}{2}\sigma^2 \left[t^{2H} - (t-s)^{2H} + s^{2H}\right]$.

Tada se B^H *naziva* ***frakcionalno Brownovo gibanje****.*

Za $H = \frac{1}{2}$*, funkcija autokovarijanci (1.10) je jednaka nula za* $h \neq 0$*, pa je za takav* H *proces* $\{X_t\}$ *Gaussovski nezavisni jednako distribuirani (NJD) šum, a* B^H *standardno Brownovo gibanje. Za* $\frac{1}{2} < H < 1$*, frakcionalni Gaussovski šum* $\{X_t\}$ *je stacionaran proces s dugoročnim pamćenjem.*
Na Slici 1.1 prikazana je simulirana trajektorija Gaussovskog NJD šuma ($H = 0.5$*), te frakcionalnih Gaussovskih šumova za* $H = 0.7$ *i* $H = 0.9$*. Dugoročna zavisnost se očituje u pojavljivanju dugih perioda visokih vrijednosti i dugih perioda niskih vrijednosti procesa. Ako pogledamo kraći vremenski period, mogli bismo pomisliti kako se pojavljuje sezonalnost ili lokalni trend, no gledajući cijeli niz, uviđamo kako nema postojanog ciklusa ni trenda već se pojavljuju ciklusi skoro svih frekvencija u različitim razmacima. Primjećujemo kako je dugoročna zavisnost jača što je parametar* H *bliže* 1*.*



Slika 1.1: *Simulirana trajektorija frakcionalnog Gaussovskog šuma za* $H = 0.5$, $H = 0.7$ *i* $H = 0.9$.

2 FARIMA model

Druga klasa modela za koju u određenim slučajevima vrijedi (1.1) su frakcionalni ARIMA procesi. Oni su prirodno proširenje klasičnih ARIMA procesa koje ćemo u ovom poglavlju predstaviti kako bismo pomoću njih mogli modelirati vremenske nizove koji nas zanimaju. Na početku ćemo uvesti definicije nekih procesa za lakše razumijevanje definicije i rezultata vezanih za frakcionalne ARIMA procese.

Definicija 2.1 (Bijeli šum) *Slučajan proces* $\{Z_t, t \in \mathbb{Z}\}$ *za koji vrijedi:*

- $\mathrm{E}Z_t^2 < \infty$, $\forall t \in \mathbb{Z}$,
- $\mathrm{E}Z_t = 0$, $\forall t \in \mathbb{Z}$,
- $\mathrm{Var}Z_t = \sigma^2$, $\forall t \in \mathbb{Z}$,
- $\mathrm{Cov}(Z_t, Z_s) = 0$, *za* $t \neq s$,

naziva se ***bijeli šum****. Kraće pišemo:* $\{Z_t\} \sim WN(0, \sigma^2)$.

Ako je $\{Z_t\}$, dodatno, niz nezavisnih jednako distribuiranih slučajnih varijabli, $\{Z_t\}$ nazivamo nezavisni jednako distribuirani (NJD) šum.

Definicija 2.2 (ARMA proces) *Slučajan proces* $\{X_t, t \in \mathbb{Z}\}$ *je* ***autoregresivni pomični prosjek reda*** (p, q), $p, q \geq 0$, *ako je stacionaran i ako je*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, t \in \mathbb{Z}, \tag{2.1}$$

gdje je $\{Z_t\} \sim WN(0, \sigma^2), \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \in \mathbb{R}, \phi_p, \theta_q \neq 0$ *i polinomi*

$$\begin{aligned} \phi(z) =& 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \\ \theta(z) =& 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q, \quad z \in \mathbb{C}, \end{aligned} \tag{2.2}$$

nemaju zajedničke nultočke. Kraće pišemo: $\{X_t\} \sim ARMA(p, q)$.

Koristeći operator pomaka unazad B definiran s $BX_t = X_{t-1}$, (2.1) možemo zapisati kao $\phi(B)X_t = \theta(B)Z_t$.

Teorem 2.1 *Stacionarno rješenje jednadžbe (2.1) postoji i jedinstveno je ako i samo ako* $\phi(z) \neq 0$, $\forall z \in \mathbb{C}$ *takav da je* $|z| = 1$.

Dokaz. Vidi [3, Poglavlje 3.]. □

Definicija 2.3 (Kauzalan proces) *Za $ARMA(p,q)$ proces $\{X_t\}$ kažemo da je* ***kauzalan*** *ako postoji niz $\{\psi_j\}$, $\sum_{j=0}^{\infty}|\psi_j|<\infty$, takav da je*

$$X_t=\sum_{j=0}^{\infty}\psi_j Z_{t-j},\quad t\in\mathbb{Z}.$$

Definicija 2.4 (Invertibilan proces) *Za $ARMA(p,q)$ proces $\{X_t\}$ kažemo da je* ***invertibilan*** *ako postoji niz $\{\pi_j\}$, $\sum_{j=0}^{\infty}|\pi_j|<\infty$, takav da je*

$$Z_t=\sum_{j=0}^{\infty}\pi_j X_{t-j},\quad t\in\mathbb{Z}.$$

Definicija 2.5 (ARIMA proces) *Neka je $d\in\mathbb{N}_0$. Slučajan proces $\{X_t\}$ za koji vrijedi*

$$\phi(B)(1-B)^d X_t=\theta(B)Z_t,\quad Z_t\sim WN(0,\sigma^2),\tag{2.3}$$

je ***integrirani ARMA proces reda*** *(p,d,q) ako je $(1-B)^dX_t$ kauzalan $ARMA(p,q)$ proces. Kraće pišemo: $\{X_t\}\sim ARIMA(p,d,q)$.*

2.1 Definicija i osnovni rezultati

Uočimo da možemo pisati

$$(1-B)^d=\sum_{k=0}^{d}\binom{d}{k}(-1)^k B^k\tag{2.4}$$

s binomnim koeficijentima

$$\binom{d}{k}=\frac{d!}{k!(d-k)!}=\frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

Za negativne cijele brojeve gama funkcija nije definirana pa su binomni koeficijenti jednaki nula ako je $k>d$, za d cijeli broj. Stoga (2.4) možemo zamijeniti s

$$(1-B)^d=\sum_{k=0}^{\infty}\binom{d}{k}(-1)^k B^k.$$

Kako je gama funkcija definirana za sve realne brojeve, jednadžba (2.3) može biti generalizirana ako dopustimo da d bude realan broj. Tada imamo:

Definicija 2.6 (FARIMA proces) *Neka je $\{X_t, t\in\mathbb{Z}\}$ stacionaran proces takav da vrijedi*

$$\phi(B)(1-B)^d X_t=\theta(B)Z_t,\quad Z_t\sim WN(0,\sigma^2),\tag{2.5}$$

za $-\frac{1}{2}<d<\frac{1}{2}$ i polinome ϕ, θ stupnjeva p,q redom. Tada $\{X_t\}$ zovemo ***frakcionalni ARIMA proces reda*** *(p,d,q). Kraće pišemo: $\{X_t\}\sim FARIMA(p,d,q)$.*

Proces koji zadovoljava (2.5) postoji što pokazuje sljedeći teorem iz [3, str.525].

Teorem 2.2 *Pretpostavimo da je* $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ *i da polinomi* ϕ *i* θ *nemaju zajedničkih nultočaka. Ako je* $\phi(z) \neq 0$ *za* $|z| = 1$*, tada postoji jedinstveno stacionarno rješenje jednadžbe (2.5) dano s*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j (1-B)^{-d} Z_{t-j}, \tag{2.6}$$

gdje je $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$.

Za $|d| \geq \frac{1}{2}$ proces nije stacionaran, no odgovarajućim diferenciranjem dobivamo stacionarni proces koji zadovoljava (2.5) s $0 < |d| < \frac{1}{2}$. Ako je kod frakcionalnog ARIMA(p, d, q) procesa $p = q = 0$, tada imamo proces koji nazivamo frakcionalni bijeli šum o čemu govori sljedeća definicija.

Definicija 2.7 (Frakcionalni bijeli šum) *Neka je* $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$*. Stacionaran proces* $\{W_t\}$ *za koji vrijedi* $\mathrm{E}[W_t] = 0$ *i*

$$(1-B)^d W_t = Z_t, \quad Z_t \sim WN(0, \sigma^2)$$

zovemo ***frakcionalni ARIMA proces reda*** $(0, d, 0)$ *ili* ***frakcionalni bijeli šum****.*

Dokaz da frakcionalni bijeli šum postoji može se naći u [3]. Također, u [3] se može naći sljedeća propozicija koja opisuje strukturu funkcije autokovarijanci frakcionalnog bijelog šuma.

Propozicija 2.1 *Neka je* $\{W_t\}$ *frakcionalni bijeli šum. Tada je njegova funkcija autokovarijanci definirana s*

$$\gamma_W(h) = \sigma^2 \frac{(-1)^h \Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-h-d)}. \tag{2.7}$$

Strukturu funkcije autokovarijanci frakcionalnog ARIMA(p, d, q) procesa dobit ćemo koristeći Propoziciju 2.1.
Neka je $\{X_t\}$ invertibilan proces koji zadovoljava (2.5). Tada jednadžbu (2.5) možemo zapisati kao

$$\phi(B) X_t = \theta(B)(1-B)^{-d} Z_t. \tag{2.8}$$

Ako označimo da je $W_t = (1-B)^{-d} Z_t$, onda vrijedi

$$(1-B)^d W_t = Z_t,$$

pa je po Definiciji 2.7 proces $\{W_t\}$ frakcionalni bijeli šum. Dakle, možemo reći da je proces $\{X_t\}$ dobiven djelovanjem linearnog filtera

$$\beta(B) = \sum_{j=0}^{\infty} \beta_j B^j = \phi(B)\theta^{-1}(B)$$

na frakcionalni ARIMA$(0, d, 0)$ proces. Kako bismo dobili funkciju autokovarijanci procesa $\{X_t\}$, prvo odredimo koeficijente β_j tako da izjednačimo odgovarajuće potencije od $\phi(B)\theta^{-1}(B)$ s onima od $\beta(B)$. Tada dobijemo da je funkcija γ, zadana s

$$\gamma(h) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \beta_j\beta_k\gamma_W(h+j-k),$$

funkcija autokovarijanci procesa $\{X_t\}$ koji zadovoljava (2.8).

Propozicija 2.2 *Neka je $\{X_t\}$ kauzalan proces koji zadovoljava jednadžbu (2.5). Spektralna gustoća procesa $\{X_t\}$ je dana s*

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} \cdot \frac{\sigma^2|\theta(e^{-i\lambda})|^2}{2\pi|\phi(e^{-i\lambda})|^2}. \tag{2.9}$$

Dokaz. Neka je $\{X_t\}$ kauzalan proces koji zadovoljava jednadžbu (2.5). Tada (2.5) možemo zapisati kao

$$(1-B)^d X_t = Y_t, \tag{2.10}$$

gdje je $\{Y_t\}$ ARMA(p, q) proces definiran s

$$\phi(B)Y_t = \theta(B)Z_t.$$

Spektralna gustoća ARMA procesa $\{Y_t\}$ (vidi [3, str.123]) je dana s

$$f_{ARMA}(\lambda) = \frac{\sigma^2|\theta(e^{-i\lambda})|^2}{2\pi|\phi(e^{-i\lambda})|^2}.$$

Ako je proces $\{X_t\}$ dobiven djelovanjem linearnog filtera $\sum a(j)Y_{t-j}$ na proces $\{Y_t\}$ sa spektralnom gustoćom f_Y, onda je spektralna gustoća procesa $\{X_t\}$ jednaka $|A(e^{-i\lambda})|^2 f_Y(\lambda)$ gdje je $A(e^{-i\lambda}) = \sum a(j)e^{-ij\lambda}$ (vidi [11, str.266]). Zbog tog rezultata i jednadžbe (2.5), spektralna gustoća procesa $\{X_t\}$ je

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_{ARMA}(\lambda) = |1 - e^{-i\lambda}|^{-2d} \cdot \frac{\sigma^2|\theta(e^{-i\lambda})|^2}{2\pi|\phi(e^{-i\lambda})|^2}. \tag{2.11}$$

□

Frakcionalni ARIMA(p, d, q) procesi su stacionarni procesi s dugoročnim pamćenjem za određeni raspon parametra d što nam govori Propozicija 2.3.

Propozicija 2.3 *Ako je* $0 < d < \frac{1}{2}$, *frakcionalni ARIMA*(p, d, q) *proces je stacionaran proces s dugoročnim pamćenjem.*

Dokaz. Spektralna gustoća frakcionalnog ARIMA(p, d, q) procesa $\{X_t\}$ je zadana s

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} \cdot \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2}.$$

Modul kompleksnog broja $z = x+yi$ zadan je s $|z| = \sqrt{x^2 + y^2}$. Zbog trigonometrijskih identiteta vrijedi da je $\sin^2 \lambda = \frac{1}{2}(1 - \cos(2\lambda))$, pa stoga imamo

$$\begin{aligned} |1 - e^{-i\lambda}| &= |1 - \cos\lambda + i \sin\lambda| = \sqrt{(1 - \cos\lambda)^2 + (\sin\lambda)^2} \\ &= \sqrt{1 - 2\cos\lambda + \cos^2\lambda + \sin^2\lambda} \\ &= \sqrt{2 - 2\cos\lambda} = \sqrt{4\sin^2\frac{\lambda}{2}} \\ &= 2\sin\frac{\lambda}{2}, \end{aligned}$$

za $\lambda \in (0, \pi]$. Pošto je $\lim_{\lambda \to 0} \lambda^{-1} \left(2\sin\left(\frac{1}{2}\lambda\right)\right) = 1$, slijedi da je spektralna gustoća procesa $\{X_t\}$ oko nule dana s

$$f(\lambda) \sim \frac{\sigma^2 |\theta(1)|^2}{2\pi |\phi(1)|^2} |\lambda|^{-2d} = f_{ARMA}(0) |\lambda|^{-2d}.$$

Stoga, za $d > 0$ spektralna gustoća ima pol u nuli u iznosu pozitivna konstanta puta $|\lambda|^{-2d}$. Po Definiciji 1.4, za $0 < d < \frac{1}{2}$, frakcionalni ARIMA(p, d, q) proces je stacionaran proces s dugoročnim pamćenjem. □

2.2 Prediktor FARIMA modela

U slučaju kauzalnog i invertibilnog frakcionalnog ARIMA(p, d, q) procesa $\{X_t\}$, postoje $\{\psi_j\}$ i $\{\pi_j\}$ takvi da vrijedi:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \tag{2.12}$$

i

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \tag{2.13}$$

gdje je $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)\phi^{-1}(z)(1-z)^{-d}$ i $\sum_{j=0}^{\infty} \pi_j z^j = \phi(z)\theta^{-1}(z)(1-z)^{d}$. Koeficijente ψ_j i π_j dobijemo izjednačavajući odgovarajuće potencije iz prikazanih reprezentacija. U slučaju frakcionalnog bijelog šuma, imamo eksplicitne formule za koeficijente koje se mogu pronaći u [7]. Ovi rezultati su korisni za predviđanje budućih realizacija procesa na osnovu beskonačne prošlosti. Definirat ćemo takav prediktor kojim ćemo se poslužiti u Poglavlju 4 u svrhu predviđanja stvarnih podataka.
Najbolji linearni prediktor za X_{n+h}, $h \geq 1$, na osnovu beskonačne duge prošlosti je

$$\widetilde{\sqcap}_n X_{n+h} = \sqcap_{\overline{sp}\{1, X_n, X_{n-1}, \dots\}} X_{n+h}. \tag{2.14}$$

Za njega vrijede sljedeća svojstva:

(i) $\widetilde{\sqcap}_n U = U$, za $U = \sum_{j=-\infty}^{n} \chi_j X_j$,

(ii) $\widetilde{\sqcap}_n Z = \mathrm{E}Z$ ako je $\mathrm{Cov}(Z, X_j) = 0, \forall j \leq n$,

(iii) $\widetilde{\sqcap}_n$ je linearan.

Ako na (2.12) i (2.13) primijenimo linearni prediktor (2.14) s navedenim svojstvima, imamo

$$\widetilde{\sqcap}_n X_{n+h} = -\sum_{j=1}^{\infty} \pi_j \widetilde{\sqcap}_n X_{n+h-j} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j}. \tag{2.15}$$

Srednje kvadratna greška predikcije dana je s

$$\widetilde{v_h} = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2.$$

Kako $X_0, X_{-1}, \dots$ ne znamo, umjesto predikcije na osnovu beskonačno duge prošlosti koristit ćemo odrezanu predikciju, odnosno u (2.15) imat ćemo sumu za $1 \leq j \leq n + h - 1$. U [2, Poglavlje 8.7] je pokazano kako su predikcije kod modela dugoročne zavisnosti preciznije ukoliko se prikladno koristi struktura zavisnosti. Uzimajući u obzir svu raspoloživu prošlost, pouzdani intervali za predikcije se značajno sužavaju. Također, predikcije kod FARIMA modela su postojanije i sporije opadaju u procijenjenu očekivanu vrijednost nego što to čine predikcije ARMA modela (vidi [3, str.534]).

3 Statistički procjenitelji dugoročne zavisnosti

U ovom ćemo poglavlju predstaviti nekoliko jednostavnih metoda za identificiranje dugoročne zavisnosti u vremenskom nizu. Započet ćemo s R/S statistikom koja je bila prvi pokazatelj dugoročne zavisnosti. Preostale tri metode uočavanja dugoročne zavisnosti se oslanjaju na Definicije 1.1 i 1.4 te na svojstvo iz Teorema 1.6.

3.1 R/S statistika

Najpoznatija metoda uočavanja dugoročne zavisnosti je R/S statistika koju je predložio Hurst (1951) u području hidrologije. Naime, promatrajući tok rijeke Nil, Hurst je istraživao na koji se način može regulirati njezin tok. Htio je izračunati idealan kapacitet korita za vremensko razdoblje između trenutaka t i $t+k$.
Radi jednostavnosti, pretpostavit ćemo da je vrijeme diskretno i da nema nepredviđenih gubitaka vode zbog curenja, isparavanja i slično. Idealnim kapacitetom se smatra sljedeće: izljev je uniforman, u vremenskom trenutku $t+k$ korito je puno jednako kao što je bilo u vremenskom trenutku t i korito ni u jednom trenutku nije prepunjeno.
Neka je X_i priljev u vremenskom trenutku i i $Y_j = \sum_{i=1}^{j} X_i$ cjelokupan priljev do vremenskog trenutka j. Tada je, prema [2, str.33], idealan kapacitet dan s

$$R(t,k) = \max_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}\left(Y_{t+k} - Y_t\right)\right] - \min_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}\left(Y_{t+k} - Y_t\right)\right]. \quad (3.1)$$

$R(t,k)$ se naziva prilagođen raspon. U svrhu istraživanja svojstava koja su neovisna o skali, $R(t,k)$ je standardiziran s

$$S(t,k) = \sqrt{\frac{1}{k}\sum_{i=t+1}^{t+k}(X_i - \overline{X}_{t,k})^2}, \quad (3.2)$$

gdje je $\overline{X}_{t,k} = \frac{1}{k}\sum_{i=t+1}^{t+k} X_i$. Omjer

$$R/S = \frac{R(t,k)}{S(t,k)} \quad (3.3)$$

se naziva reskalirani prilagođen raspon ili R/S statistika (vidi [2, str.33]). Hurst je nacrtao graf logaritma R/S–a u odnosu na nekoliko vrijednosti logaritma od k i uočio da su za velike k vrijednosti logaritma R/S–a nakupljene oko pravca s koeficijentom smjera većim od $\frac{1}{2}$. To znači da je za velike k

$$\log[R/S] \approx a + H\log k, \text{ za } H > \frac{1}{2}. \quad (3.4)$$

Hurstov zaključak da se za rijeku Nil R/S statistika ponaša na način da vrijedi (3.4) naziva se *Hurstov efekt*. Sličan fenomen pojavljuje se i na drugim podacima vezanim za geofiziku, klimatologiju i hidrologiju.
Označimo s $Q = Q(t,k) = \frac{R(t,k)}{S(t,k)}$ R/S statistiku. Kako bismo uočili dugoročnu zavisnost i procijenili parametar H, crtat ćemo graf od $\log Q$ u odnosu na $\log k$ i nazivat ćemo ga R/S graf u nastavku. Tako ćemo za svaki k imati $n-k$ vrijednosti od Q. Navest ćemo dva teorema iz [9] koja će nam dati temelje za zaključivanje na osnovu ovog grafa. Za razumijevanje pretpostavki teorema navodimo sljedeću definiciju iz [5].

Definicija 3.1 (Ergodičan proces) *Neka je $\{X_t\}$ stacionaran proces, $\mu = \mathrm{E}[X_t]$, $\forall t$ i $\overline{X_T} = \frac{1}{2T}\int_{-T}^{T} X_t dt$ njegov vremenski prosjek. Kažemo da je proces $\{X_t\}$* ***ergodičan*** *ako je*

$$\lim_{T\to\infty} \overline{X_T} = \mu.$$

Teorem 3.1 *Neka je $\{X_t\}$ stacionaran proces za koji vrijedi da je $\{X_t^2\}$ ergodičan i $\left\{n^{-\frac{1}{2}}\sum_{s=1}^{\lfloor nt\rfloor} X_s\right\}$ konvergira slabo prema standardnom Brownovom gibanju kada n teži u beskonačnost. Tada za $k\to\infty$, vrijedi*

$$\frac{1}{\sqrt{k}}Q \xrightarrow{d} \xi,$$

gdje je ξ nedegenerirana slučajna varijabla.

Jednostavnije rečeno, Teorem 3.1 nam govori da kada god centralni granični teorem vrijedi, statistika $k^{-\frac{1}{2}}Q$ konvergira prema dobro definiranoj slučajnoj varijabli. Za statističku primjenu to znači da će za velike k vrijediti $\log Q \approx \frac{1}{2}\log k + \log \xi$. Odnosno, u R/S grafu će za velike k vrijednosti $\log Q$ biti nakupljene oko pravca s koeficijentom smjera $\frac{1}{2}$. Sljedeći teorem iz [9] opisuje ponašanje statistike Q za procese s dugoročnim pamćenjem.

Teorem 3.2 *Neka je $\{X_t\}$ stacionaran proces za koji vrijedi da je $\{X_t^2\}$ ergodičan i $\{n^{-H}\sum_{s=1}^{\lfloor nt\rfloor} X_s\}$ konvergira slabo prema frakcionalnom Brownovom gibanju kada n teži u beskonačnost. Tada za $k\to\infty$, vrijedi*

$$\frac{1}{k^H}Q \xrightarrow{d} \xi,$$

gdje je ξ nedegenerirana slučajna varijabla.

Može se pokazati da, primjerice, Gaussovski stacionaran proces s dugoročnim pamćenjem zadovoljava uvjet konvergencije iz Teorema 3.2. U tom slučaju, očekujemo da će se logaritmi vrijednosti Q statistike u R/S grafu nakupljati oko pravca s koeficijentom

smjera $H > \frac{1}{2}$ za dovoljno velike k.
Lijepo svojstvo R/S statistike je da njezino asimptotsko ponašanje ne pogađaju distribucije teških repova. Odnosno, ako je $\{X_t\}$ NJD niz slučajnih varijabli s teškim repom, R/S statistika svejedno ima asimptotski koeficijent smjera u iznosu od $\frac{1}{2}$. No ukoliko imamo odstupanja od stacionarnosti, R/S statistika nas može odvesti u krivi zaključak o čemu govori Teorem 4.4 iz [2].
Osvrnuvši se na sve ovo, metodu uočavanja dugoročne zavisnosti preko R/S statistike možemo sažeti na korake:

1. Izračunaj Q za sve moguće vrijednosti od k i t.

2. Prikaži graf vrijednosti $\log Q$ u odnosu na $\log k$.

3. Nacrtaj pravac $y = a + b \log k$. Koeficijente a i b procijeni metodom najmanjih kvadrata. Tada $\widehat{b}$ predstavlja procjenu parametra H, tj. $\widehat{H}$.

Ono što predstavlja problem u ovoj metodi je to što distribucija R/S statistike ovisi o distribuciji podataka. Zbog toga ju je teško odrediti pa se postavlja pitanje je li metoda najmanjih kvadrata primjerena u ovom slučaju? Nadalje, raspolažemo konačnim i vrlo često manjim skupom podataka. Kada odrediti od kojeg k počinje ono "granično ponašanje"? Koliko je pouzdana procjena parametra H u tom slučaju? No unatoč ovim nedostatcima, R/S graf je koristan i može nam dati prvu procjenu o dugoročnoj zavisnosti.

3.2 Korelogram

Standardna metoda u analizi vremenskih nizova je korelogram. Definirajmo prvo uzoračku funkciju autokovarijanci i uzoračku autokorelacijsku funkciju.

Definicija 3.2 *Za niz podataka $x_1, x_2, \ldots, x_n$ uzoračka funkcija autokovarijanci je definirana s*

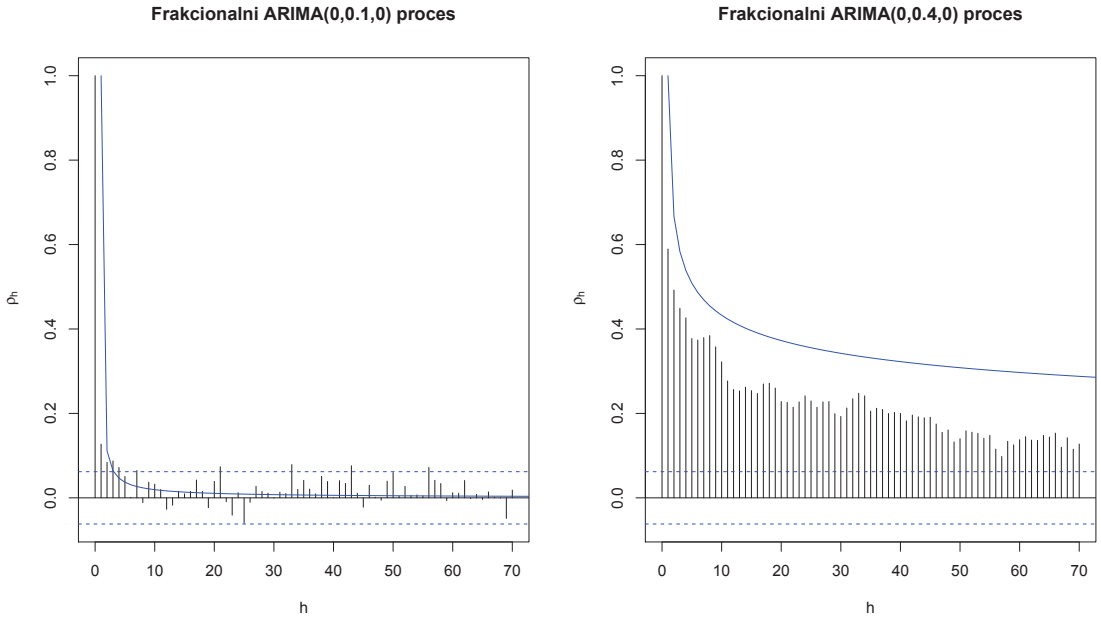
$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \overline{x}_n)(x_j - \overline{x}_n), \quad za \;\; 0 \leq h < n,$$

i $\widehat{\gamma}(h) = \widehat{\gamma}(-h)$, za $-n < h < 0$, gdje je $\overline{x}_n = \frac{1}{n} \sum_{j=1}^{n} x_j$.
Uzoračka autokorelacijska funkcija je zadana s

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)},$$

za $-n < h < n$.

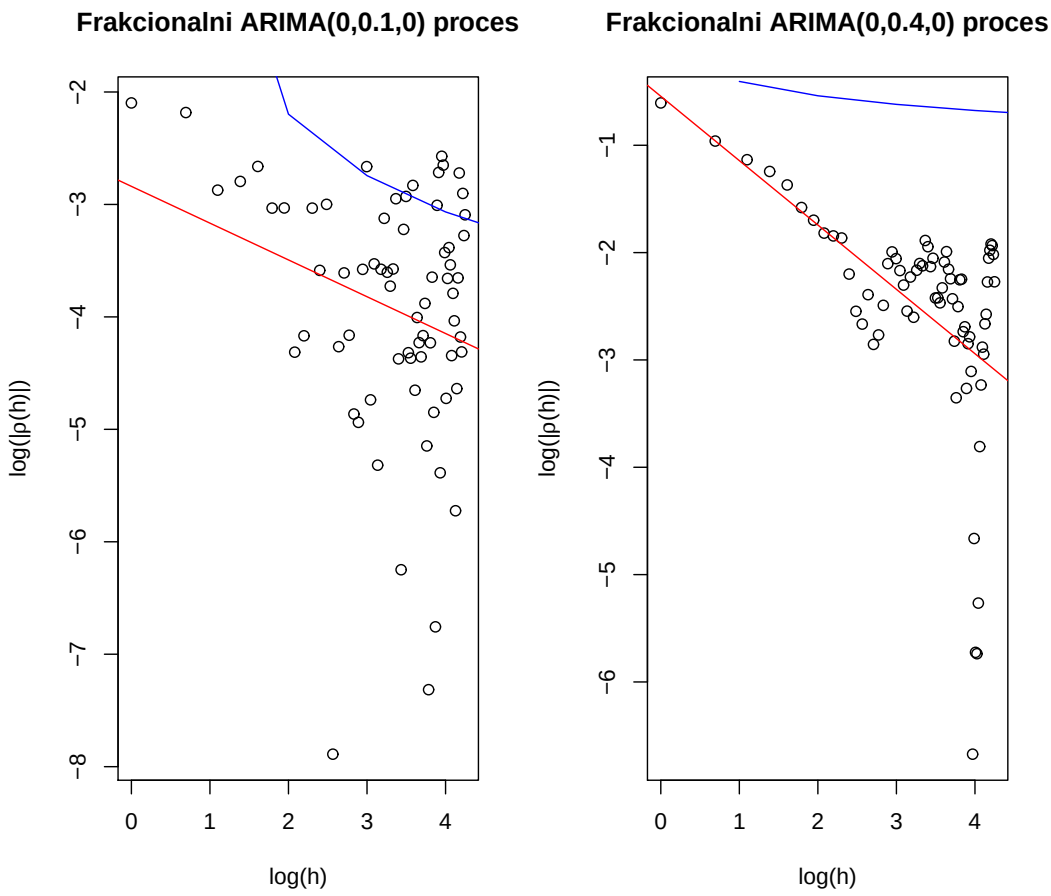
Graf parova $(h, \widehat{\rho}(h))$ za $h \geq 0$ naziva se korelogram. U korelogramu se nalaze dvije horizontalne linije na vrijednostima $\pm 1.96/\sqrt{n}$. Ukoliko uzoračke korelacije prelaze te linije, one se smatraju statistički značajno različite od nula na razini značajnosti 0.05. Ovo jednostavno pravilo opravdava asimptotska distribucija i 95% pouzdani interval od $\widehat{\rho}(h)$ u slučaju uzorka iz NJD niza.
Dugoročnu zavisnost karakterizira neuobičajeno sporo opadanje autokorelacijske funkcije u nulu pošto je ona proporcionalna s k^{2H-2} za neki $\frac{1}{2} < H < 1$. U Poglavlju 2 smo napomenuli kako struktura takve autokorelacijske funkcije može značiti da su korelacije niske, ali opadanje u nulu neuobičajeno sporo. Takvu asimptotsku karakteristiku je teško uočiti čak i kada imamo veliki uzorak.



Slika 3.1: *Korelogram simuliranog frakcionalnog ARIMA*$(0, d, 0)$ *procesa za* $d = 0.1$ *i* $d = 0.4$ *i njihova teorijska autokorelacijska funkcija (plava linija).*

Na Slici 3.1 možemo primijetiti kako je kod frakcionalnog ARIMA$(0, 0.1, 0)$ procesa većina vrijednosti uzoračke autokorelacijske funkcije ispod granica pouzdanog intervala, odnosno nije statistički značajno različita od nule. Za $H = d + \frac{1}{2}$ bliže 1, dugoročna zavisnost je uočljivija što pokazuje korelogram frakcionalnog ARIMA$(0, 0.4, 0)$ procesa. Prikladniji prikaz korelograma bi bio graf s logaritmiranim osima, tj. graf $\log|\rho(h)|$ u odnosu na $\log h$. Ako je $\rho(h) \sim c_\rho h^{2H-2}$ za $h \to \infty$, onda će za velike h vrijediti $\log|\rho(h)| \approx \log c_\rho + (2H-2)\log h$. Za procese s kratkoročnom zavisnosti log–log korelogram će divergirati u minus beskonačno, dok će se za procese s dugoročnom

zavisnosti vrijednosti log–log korelograma na velikim koracima gomilati oko pravca s negativnim koeficijentom smjera približno jednakim $2H - 2$. Linearnom regresijom $\log|\rho(h)|$ s obzirom na $\log h$ dobit ćemo pravac $y = a + b \log h$ te će procjena parametra H biti $\widehat{H} = \frac{\widehat{b}+2}{2}$.



Slika 3.2: *Log–log korelogram simuliranog frakcionalnog* $ARIMA(0, d, 0)$ *procesa za* $d = 0.1$ *i* $d = 0.4$*, njihova teorijska autokorelacijska funkcija (plava linija) i linearna regresija podataka (crvena linija).*

Na Slici 3.2 su prikazani log–log korelogrami promatranih FARIMA procesa. U slučaju FARIMA$(0, d, 0)$ procesa za $d = 0.1$ i $d = 0.4$, dobijemo procjene za b redom $\widehat{b} = -0.3272$ i $\widehat{b} = -0.5996$, odnosno procjene za parametar H redom $\widehat{H} = 0.8364$ i $\widehat{H} = 0.7002$. Kako je kod prvog procesa $H = 0.6$, a kod drugog $H = 0.9$, vidimo da su procjene poprilično grube. Linearna regresija koristi metodu najmanjih kvadrata za procjenu parametara stoga dolazimo do istih pitanja i poteškoća kao kod R/S metode. No, možemo primijetiti kako smo u oba slučaja dobili $\frac{1}{2} < \widehat{H} < 1$ što znači kako nam

ova metoda može poslužiti za uočavanje dugoročne zavisnosti.

3.3 Metoda varijance

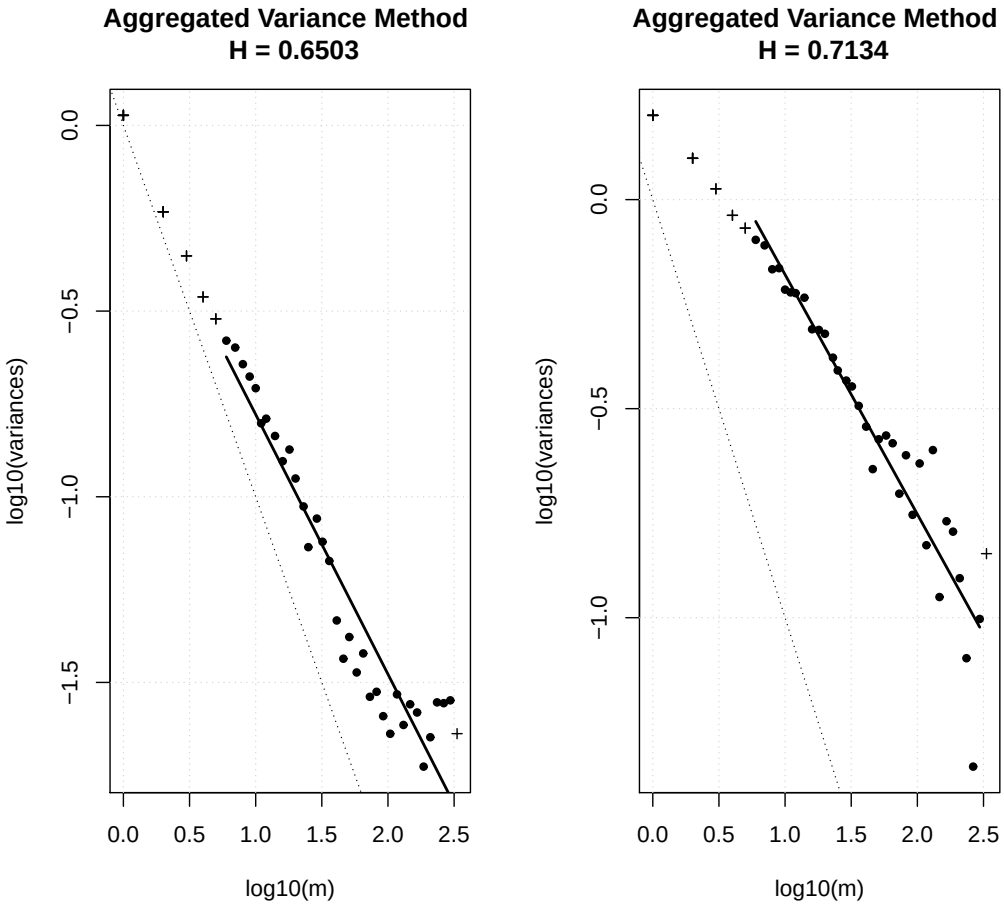
U Teoremu 1.2 prikazano je još jedno svojstvo stacionarnog procesa s dugoročnim pamćenjem koje govori kako je stopa opadanja varijance aritmetičke sredine procesa $\{X_t\}$ sporija od n^{-1}. Odnosno, ona je jednaka $n^{-\alpha}$ za neki $0 < \alpha < 1$. Ovo će nam poslužiti kao još jedan alat za uočavanje dugoročne zavisnosti nekog vremenskog niza. Metoda za procjenu parametra H koristeći ovo svojstvo varijance aritmetičke sredine procesa može se definirati na sljedeći način.

1. Neka je h prirodan broj. Za različite brojeve h u rangu $2 \leq h \leq n/2$ i m_h podnizova duljine h, izračunaj uzoračke aritmetičke sredine $\overline{X}_1(h), \overline{X}_2(h), \ldots, \overline{X}_{m_h}(h)$ i sveukupnu aritmetičku sredinu
$$\overline{X}(h) = \frac{1}{m_h} \sum_{j=1}^{m_h} \overline{X}_j(h).$$

2. Za svaki h, izračunaj uzoračku varijancu uzoračkih aritmetičkih sredina $\overline{X}_j(h)$ $(j = 1, \ldots, m_h)$:
$$s^2(h) = \frac{1}{m_h - 1} \sum_{h=1}^{m_h} (\overline{X}_j(h) - \overline{X}(h))^2.$$

3. Nacrtaj graf $\log s^2(h)$ u odnosu na $\log h$.

Za velike brojeve h, točke grafa bi se trebale gomilati oko pravca negativnog koeficijenta smjera u iznosu od $2H - 2$. Za uočavanje dugoročne zavisnosti, poželjno je nacrtati pravac s koeficijentom smjera -1 i promotriti odstupanje podataka od tog pravca. Parametar H se procjenjuje na isti način kao i u prethodne dvije metode.
Slika 3.3 je dobivena funkcijom `aggvarFit` iz paketa `fArima` u R-u koja predstavlja graf izračunatih vrijednosti ove metode. Procjena parametra H kod FARIMA$(0, 0.1, 0)$ je $\widehat{H} = 0.6503$, a kod FARIMA$(0, 0.4, 0)$ $\widehat{H} = 0.7134$, dok su stvarne vrijednosti parametra H redom $H = 0.6$ i $H = 0.9$. Primjećujemo da u drugom slučaju metoda varijance poprilično podcjenjuje stvarni parametar H procesa, no kao uočavanje dugoročne zavisnosti i prva "gruba" procjena parametra H metoda varijance može poslužiti.



Slika 3.3: *Graf metode varijance za simulirani frakcionalni* $ARIMA(0, d, 0)$ *proces za* $d = 0.1$ *i* $d = 0.4$*, pravac s koeficijentom smjera* -1 *(iscrtkana linija) i linearna regresija podataka (puna linija).*

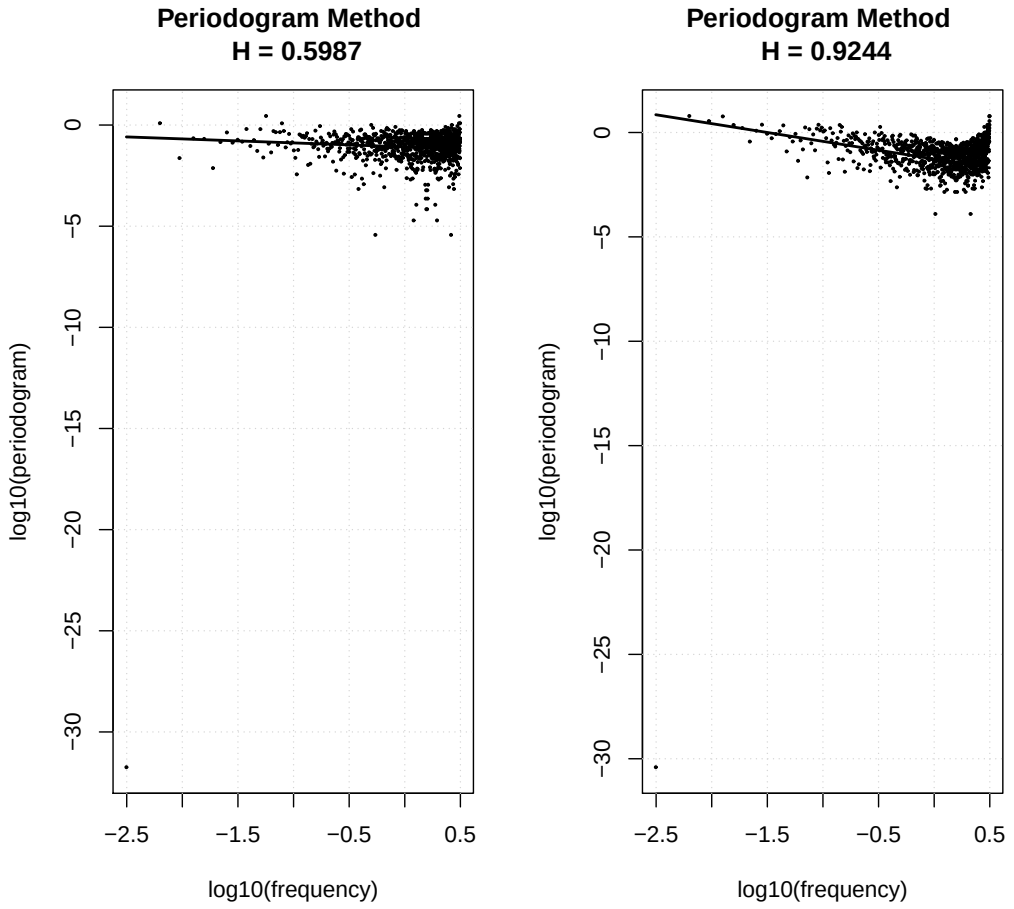
3.4 Periodogram

U Definiciji 1.4 procese s dugoročnim pamćenjem nazivamo stacionarne procese čija spektralna gustoća ima pol u nuli u iznosu proporcionalnom članu $|\lambda|^{1-2H}$, za $0 < 2H - 1 < 1$. Procjena spektralne gustoće nekog vremenskog niza omogućit će nam da gledajući podatke oko nule pokušamo zaključiti radi li se o realizacijama stacionarnog procesa s dugoročnim pamćenjem ili ne. Procjena spektralne gustoće ili uzoračka spektralna gustoća za podatke $x_1, x_2, \ldots, x_n$ je periodogram definiran kao

$$I(\lambda) = \frac{1}{n} \left| \sum_{t=1}^{n} x_t e^{-2\pi i \lambda t} \right|,$$

za $\lambda \in \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Kao i u prethodnim metodama, prikazat ćemo graf od $\log(I(\lambda))$ u odnosu na $\log(\lambda)$ gdje ćemo podacima pridružiti pravac linearne regresije $\log(I(\lambda))$ na $\log(\lambda)$ koristeći metodu najmanjih kvadrata. Koeficijent smjera tog pravca bit će procjena od $1 - 2H$.



Slika 3.4: *Logaritmirani periodogram u odnosu na logaritmirane frekvencije simuliranog frakcionalnog ARIMA*$(0, d, 0)$ *proces za* $d = 0.1$ *i* $d = 0.4$ *i linearna regresija podataka (puna linija).*

Funkcija `perFit`, također iz paketa `fArima`, za simulirani FARIMA$(0, 0.1, 0)$ i FARIMA $(0, 0.4, 0)$ dala je kao rezultat Sliku 3.4 i procjene $\widehat{H} = 0.5986559$ i $\widehat{H} = 0.9243703$ redom. Za procjenu parametara pravca linearne regresije koriste se podaci za najnižih 10% frekvencija s obzirom da se, teorijski, spektralna gustoća procesa s dugoročnim pamćenjem ponaša kao $c_f|\lambda|^{1-2H}$ samo oko nule. Više detalja o ovoj metodi i pozadinskoj teoriji može se naći u [2].

3.5 Whittleov aproksimativni ML procjenitelj

Procjenitelji metodom maksimalne vjerodostojnosti ili ML procjenitelji su efikasni i asimptotski normalni, stoga ćemo za vrijednost parametra H u modeliranju koristiti procjenu parametra H tom metodom. Metoda maksimalne vjerodostojnosti se oslanja na distribuciju procesa zbog čega ovdje pretpostavljamo da imamo Gaussovski proces. Kako je izračun funkcije vjerodostojnosti za frakcionalne ARIMA procese jako spor i zahtjevan, koristi se Whittleova aproksimacija funkcije vjerodostojnosti, tj. L_W definirana kao

$$-2\ln(L_W) = n\ln(2\pi) + 2n\ln\sigma + \frac{1}{\sigma^2}\sum_j \frac{I_n(\omega_j)}{g(\omega_j)} + \sum_j \ln g(\omega_j),$$

gdje je I_n periodogram, $(\sigma^2/(2\pi))g$ spektralna gustoća modela, dok $\sum_j$ označava sumu po svim Fourierovim frekvencijama $\omega_j = 2\pi j/n \in (-\pi, \pi]$ koje su različite od nule. Više o ovoj metodi i izvodu Whittleovog aproksimativnog ML procjenitelja može se pronaći u [2, Poglavlje 5, Poglavlje 6].

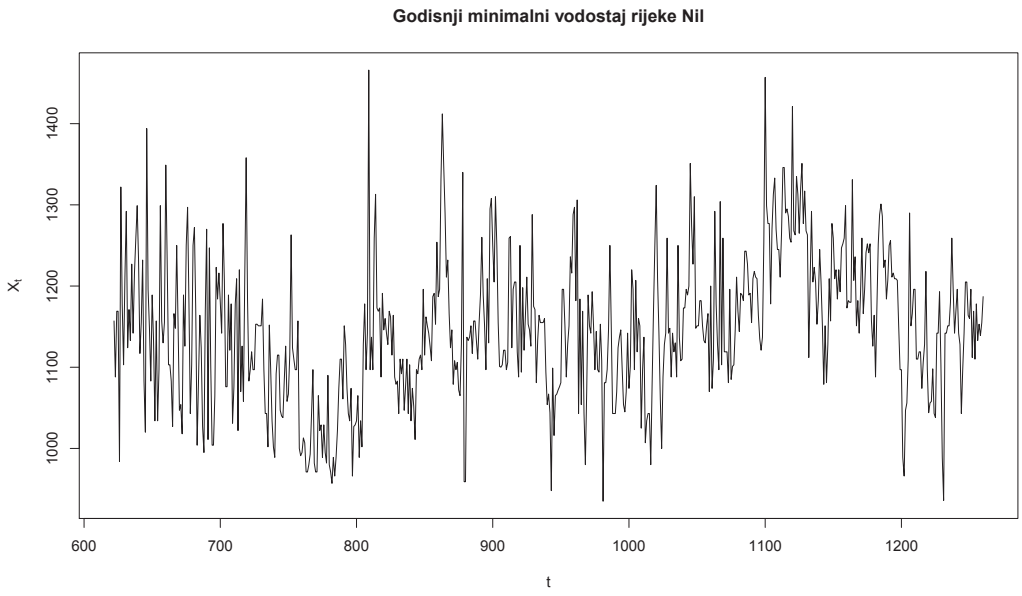
Funkcija `WhittleEst` iz paketa `longmemo` izračunava Whittleov aproksimativni ML procjenitelj za parametar H kako je to opisano u [2, Poglavlje 6.1].

4 Modeliranje stvarnih podataka FARIMA modelom

4.1 Minimalni vodostaj rijeke Nil

4.1.1 Opis podataka

U ovom potpoglavlju modelirat ćemo vremenski niz podataka o vodostaju rijeke Nil koji smo naveli kao primjer podataka gdje se uočavaju karakteristike stacionarnih procesa s dugoročnim pamćenjem. U prethodnim poglavljima smo mogli primijetiti kako je ovaj skup podataka od povijesne važnosti. Naime, analiza ovog skupa podataka i sličnih dovela je do otkrića takozvanog Hurstovog efekta koje je, nekoliko godina kasnije, Mandelbrota i njegove suradnike potaknulo na uvođenje stohastičkih modela s dugoročnom zavisnošću.
Vremenski niz $\{x_t\}$ koji ćemo promatrati se sastoji od 663 godišnja podatka o minimalnom vodostaju rijeke Nil u razdoblju od 622. do 1281. godine koji su mjereni na mjerilu na otoku Roda pokraj Kaira. Originalni podaci su dani u knjizi Jana Berana [2].



Slika 4.1: *Godišnji minimalni vodostaj rijeke Nil u razdoblju od 622. do 1260. godine.*

Na Slici 4.1 je prikazan minimalni vodostaj rijeke Nil kroz vrijeme u razdoblju od 622. do 1260. godine. Ovaj vremenski niz modelirat ćemo FARIMA modelom i tako napraviti predviđanje. Zadnjih 21 podatak ostavit ćemo u svrhu usporedbe s

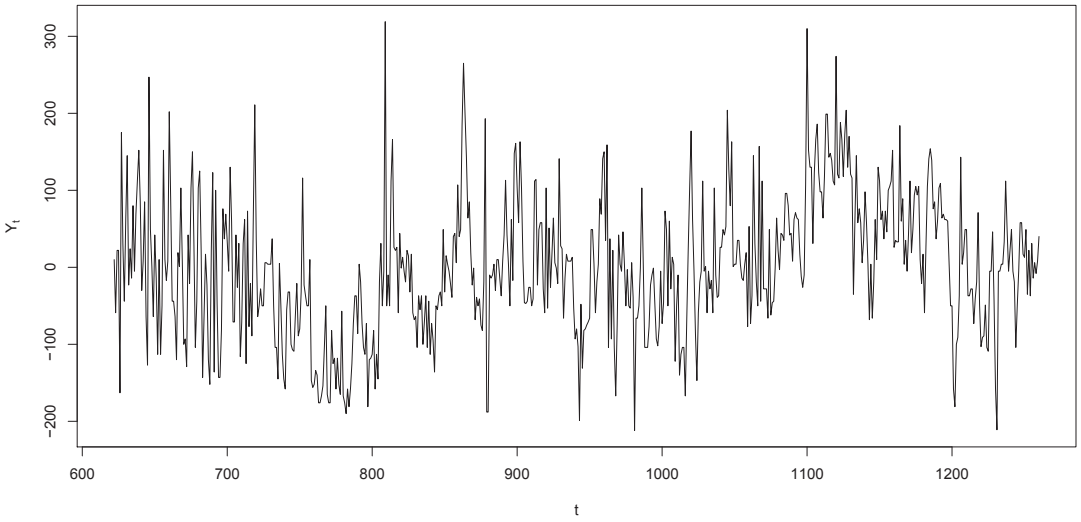
predviđenim vrijedostima našeg modela. Numeričke karakteristike vremenskog niza $\{x_t\}$ prikazane su u Tablici 1.

Minimum	Donji kvartil	Medijan	Aritmetička sredina	Gornji kvartil	Maksimum
935	1089	1146	1147	1200	1466

Tablica 1: *Numeričke karakteristike niza podataka o godišnjem minimalnom vodostaju rijeke Nil 622.–1260. godine.*

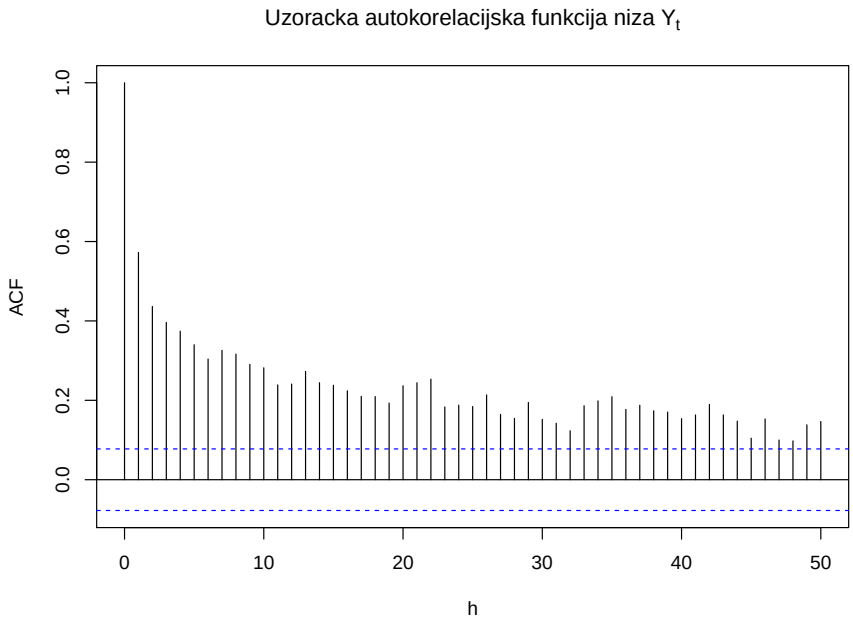
4.1.2 Identifikacija i dijagnostika modela

Kako procjena očekivanog godišnjeg minimalnog vodostaja rijeke Nil u promatranom razdoblju iznosi 1147, FARIMA modelom $\{Y_t\}$ ćemo modelirati vremenski niz $\{y_t\}$ gdje je $y_t = x_t - 1147$.



Slika 4.2: *Vremenski niz $\{y_t\}$ u odnosu na vrijeme t.*

Promatrajući Sliku 4.2 možemo reći kako je niz $\{y_t\}$ stacionaran niz u kojem se izmjenjuju dugi periodi visokih i niskih vrijednosti. Također, uočavamo puno ciklusa različitih frekvencija za što ne možemo reći da je posljedica nekakve sezonalnosti. Stacionarnost niza možemo primijetiti i gledajući uzoračku autokorelacijsku funkciju niza $\{y_t\}$ prikazanu na Slici 4.3.

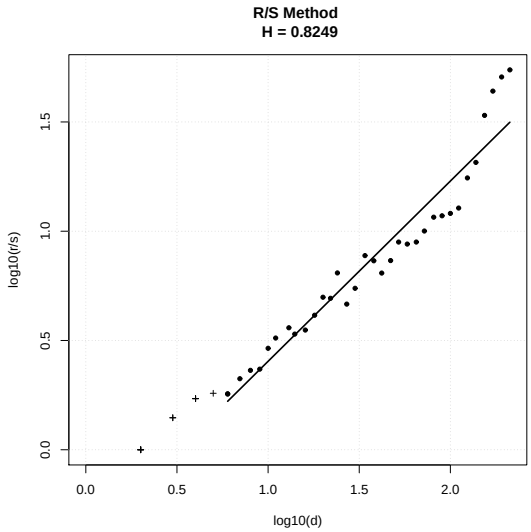


Slika 4.3: *Uzoračka autokorelacijska funkcija vremenskog niza* $\{y_t\}$.

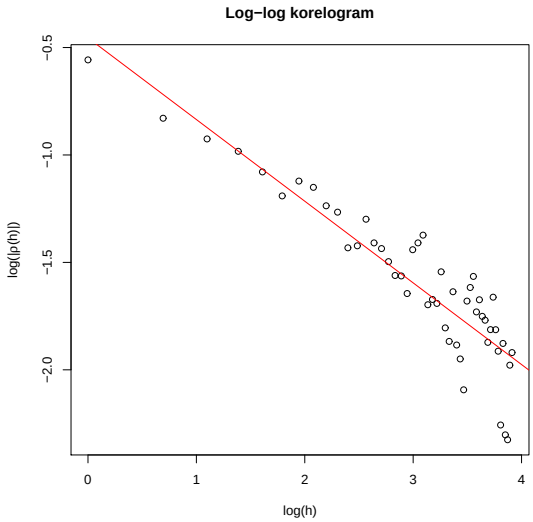
Sporo opadanje uzoračke autokorelacijske funkcije u nulu koje je za ovaj niz podataka očito, sugerira kako bismo $\{y_t\}$ mogli modelirati frakcionalnim ARIMA(p, d, q) procesom. Kako bismo potvrdili sumnju da je niz $\{y_t\}$ realizacija stacionarnog procesa s dugoročnom zavisnosti, proveli smo metode za statističku procjenu dugoročne zavisnosti koje su opisane u Poglavlju 3. Na Slici 4.4 su prikazani grafovi tih metoda za vremenski niz $\{y_t\}$.

Graf R/S statistike prikazan na Slici 4.4a pokazuje kako se vrijednosti R/S statistike gomilaju oko pravca s koeficijentom smjera većim od $\frac{1}{2}$. Procjena koeficijenta smjera pravca koji aproksimira podatke R/S statistike je 0.8248789 što je ujedno i procjena za parametar dugoročne zavisnosti H po ovoj metodi. Korelogram vremenskog niza $\{y_t\}$ u logaritmiranim osima dan je na Slici 4.4b. Procjena koeficijenta smjera pravca linearne regresije podataka $\log|\widehat{\rho}(h)|$ u odnosu na $\log h$ je -0.3801, što znači da je po ovoj metodi $\widehat{H} = 0.80995$. Metoda varijance, čiji je graf prikazan na Slici 4.4c, također pokazuje kako vremenski niz $\{y_t\}$ upućuje na proces s dugoročnim pamćenjem. Procjena parametra H metodom varijance za promatrani niz je $\widehat{H} = 0.8392439$. Log–log periodogram (Slika 4.4d) niza $\{y_t\}$ prikazuje logaritmirane vrijednosti periodograma za logaritme niskih frekvencija, tj. frekvencija oko nule. Aproksimirajući te podatke pravcem, dobivamo procjenu parametra H koja je jednaka 0.8721554.

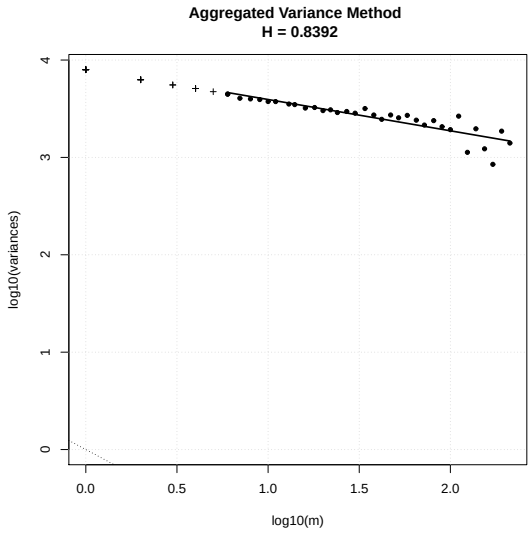
Svaka od ovih metoda dala je procjenu parametra H u intervalu $(0.8, 0.9)$, stoga ćemo



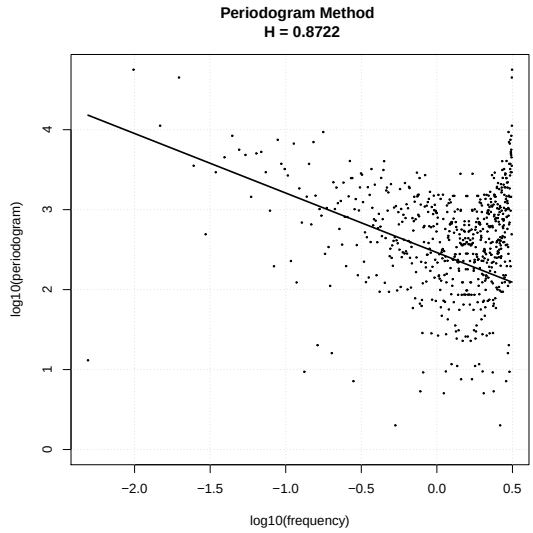
(a) *Graf R/S statistike.*



(b) *Log–log korelogram.*



(c) *Graf metode varijance.*

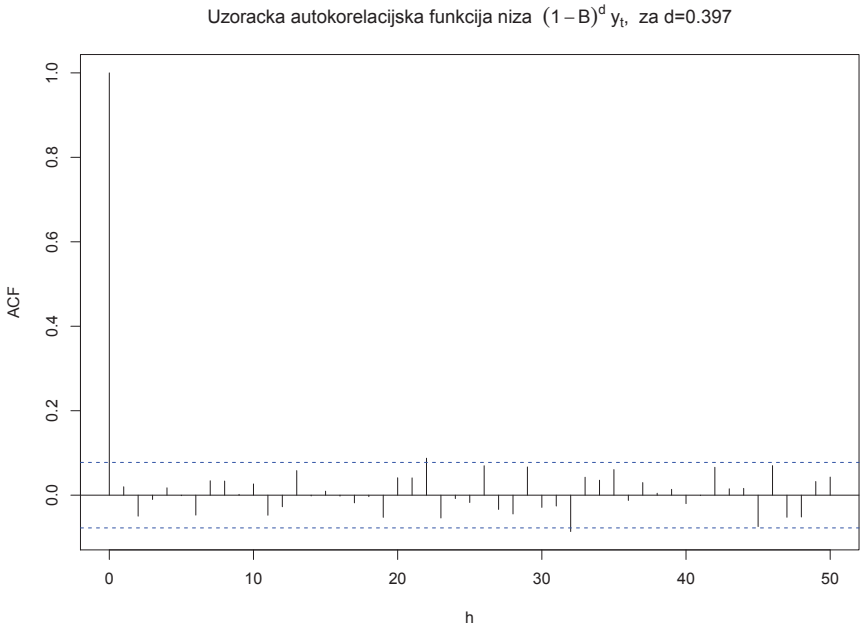


(d) *Log–log periodogram.*

Slika 4.4: *Statistički procjenitelji za parametar dugoročne zavisnosti H promatranog vremenskog niza $\{y_t\}$.*

vremenski niz podataka $\{y_t\}$ modelirati frakcionalnim ARIMA(p, d, q) modelom gdje je $d = H - \frac{1}{2}$. Procjena parametra H dobivena pomoću Whittleovog aproksimativnog ML procjenitelja je $\widehat{H} = 0.89656833$, što znači da ćemo za parametar d u FARIMA(p, d, q) modelu uzeti $d = \widehat{H} - \frac{1}{2} = 0.397$.

Nakon što smo funkcijom `diffseries` iz paketa `fracdiff` frakcionalno diferencirali niz $\{y_t\}$ na koraku $d = 0.397$, dobivamo niz $\{(1-B)^{0.397}y_t\}$ gdje se $(1-B)^{0.397}$ računa kao (2.4). Uzoračka autokorelacijska funkcija tog niza prikazana je na Slici 4.5.

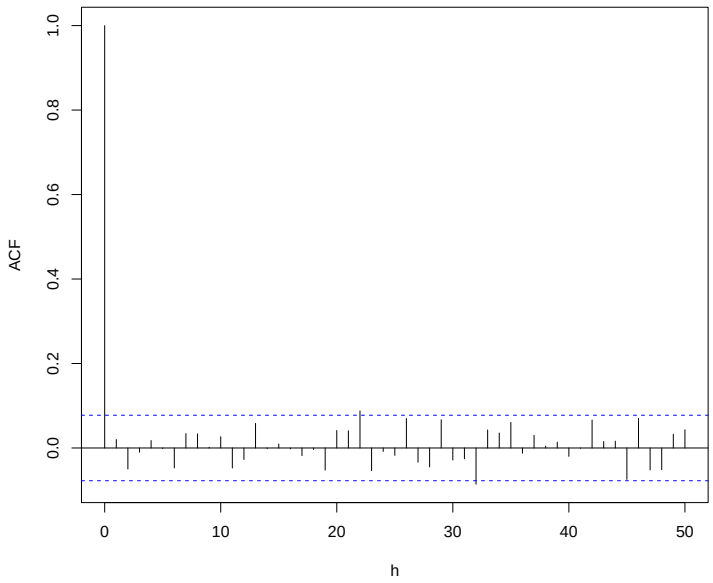


Slika 4.5: *Uzoračka autokorelacijska funkcija frakcionalno diferenciranog vremenskog niza* $\{(1-B)^{0.397}y_t\}$.

Primjećujemo kako niti jedna korelacija na koraku h, $h \geq 1$ nije statistički značajno različita od nule. Zbog toga će prvi model za niz $\{y_t\}$ biti frakcionalni bijeli šum ili frakcionalni ARIMA$(0, d, 0)$ proces za $d = 0.397$. Nakon procjene parametara funkcijom `arfima` iz paketa `arfima`, dobijemo model:

$$(1-B)^{0.397}Y_t = Z_t, \quad Z_t \sim WN(0, 4963.68). \tag{4.1}$$

Ispitujući nekoreliranost reziduala ovog modela Ljung–Box testom (vidi [8]) za korake $h = 1, \ldots, 20$ zaključujemo kako nema razloga sumnjati u nekoreliranost reziduala ovoga modela što možemo uočiti i iz grafa procijenjene autokorelacijske funkcije tih reziduala (Slika 4.6).
U želji da provjerimo postoji li još koji prikladan FARIMA model za niz $\{y_t\}$, potražili smo najbolji ARMA(p, q) model za niz $\{(1-B)^{0.397}y_t\}$ na osnovu informacijskog kriterija AIC, statističke značajnosti koeficijenata i nekoreliranosti reziduala modela. Tako



Slika 4.6: *Procijenjena autokorelacijska funkcija reziduala modela (4.1).*

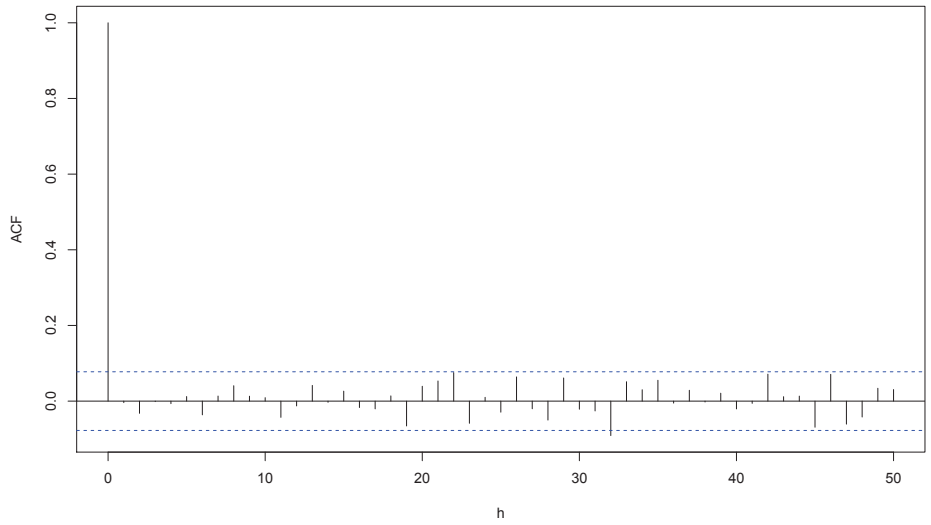
dobiven FARIMA model za $\{y_t\}$ je FARIMA$(2, 0.397, 2)$ model definiran s

$$\begin{gathered}(1-B)^{0.397}(1+0.8884B+0.9492B^2)Y_t = (1-0.9156B-0.9601B^2)Z_t, \\ Z_t \sim WN(0, 4959.23).\end{gathered} \tag{4.2}$$

Procijenjena autokorelacijska funkcija reziduala modela (4.2) dana je na Slici 4.7. Kod oba modela, (4.1) i (4.2), Shapiro–Wilk testom normalnosti (vidi [12]) odbačena je hipoteza o normalnosti reziduala. Akaike informacijski kriterij ili AIC za FARIMA (p, d, q) model je definiran s

$$AIC(p, d, q) = -2 \ln L_W + 2(p + q + 2),$$

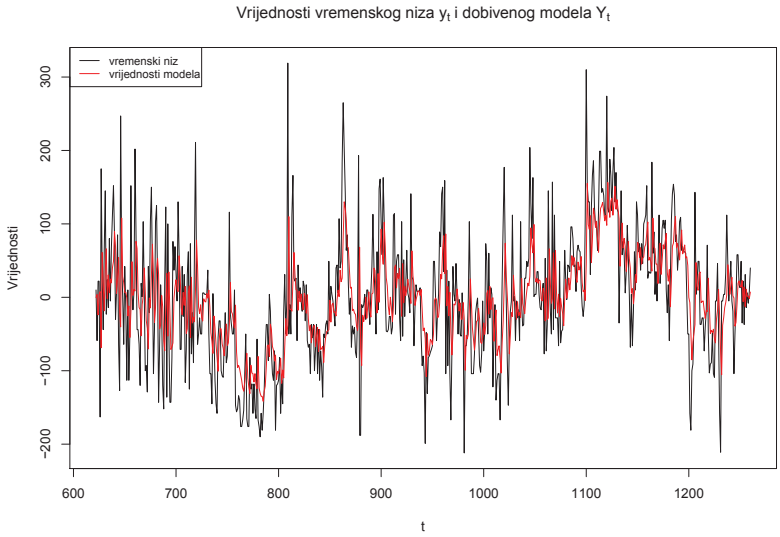
gdje je L_W Whittleov aproksimativni ML procjenitelj. Vrijednost AIC-a kod modela (4.1) iznosi 5444.61, dok za model (4.2) iznosi 5448.203. Procjena varijance je nešto manja u modelu (4.2) nego u modelu (4.1), pa uspoređujući ova dva modela, možemo reći da su podjednako dobri.



Slika 4.7: *Procijenjena autokorelacijska funkcija reziduala modela (4.2).*

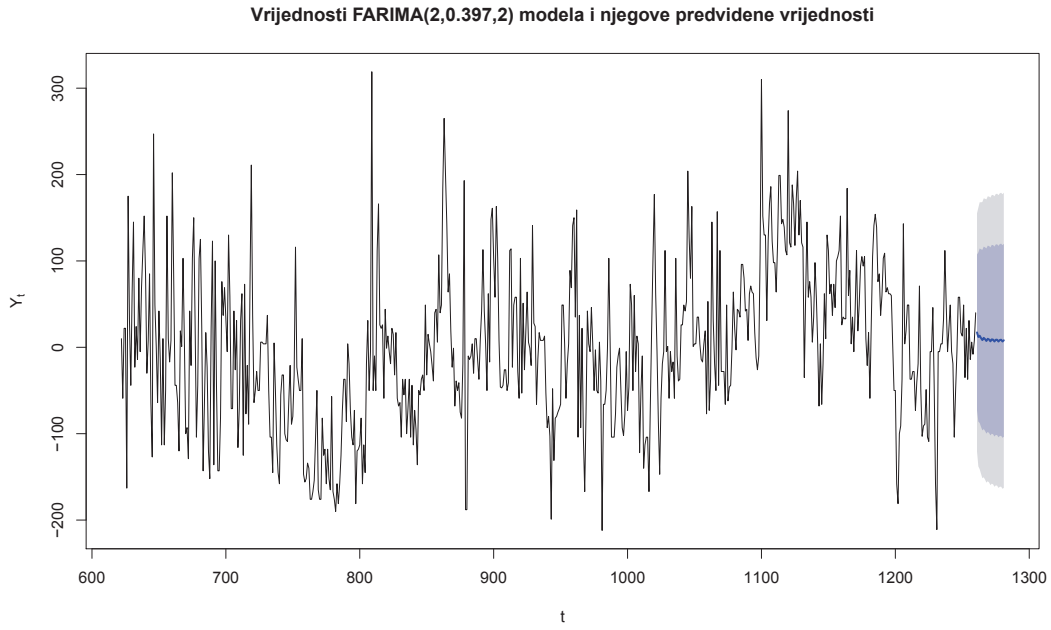
4.1.3 Predviđanje modela

U ovom potpoglavlju ćemo prikazati predviđanje dobivenog FARIMA$(2, 0.397, 2)$ modela (4.2). Najprije promotrimo vrijednosti vremenskog niza $\{y_t\}$ i vrijednosti modela (4.2) za taj vremenski niz koje su prikazane na Slici 4.8.



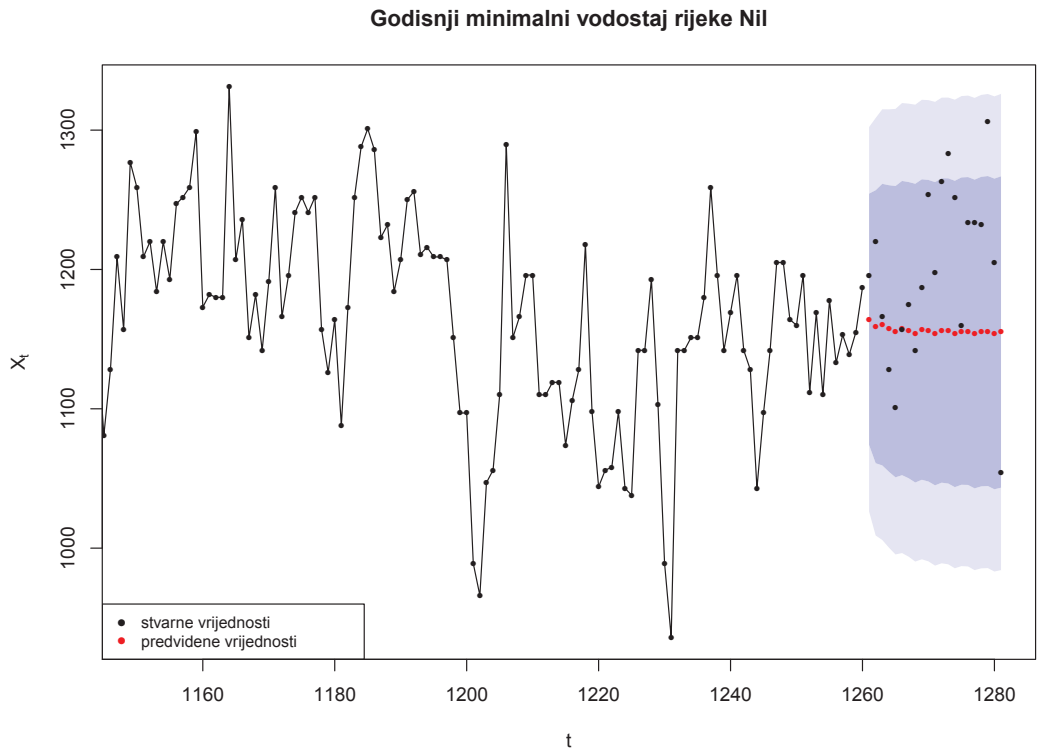
Slika 4.8: *Vrijednosti vremenskog niza* $\{y_t\}$ *i vrijednosti FARIMA*$(2, 0.397, 2)$ *modela za taj niz.*

Možemo primijetiti kako model (4.2) dobro prati trendove vremenskog niza, no one ekstremno visoke i niske vrijednosti ne dostiže.
Da bismo dobili predviđene vrijednosti modela, koristili smo funkciju `forecast` iz paketa `forecast`. Na Slici 4.9 prikazane su vrijednosti modela (4.2) i njegove predviđene vrijednosti za narednu 21 godinu koje su računate na način kao što je to opisano u Poglavlju 4.4. Plava područja oko predviđenih vrijednosti predstavljaju njihove 80% i 95% pouzdane intervale.



Slika 4.9: *Vrijednosti* $FARIMA(2, 0.397, 2)$ *modela i njegova 21 predviđena vrijednost.*

Usporedili smo stvarne vrijednosti i predviđene vrijednosti modela (4.2) za razdoblje od 1260. do 1281. godine (Slika 4.10). Vidimo da se stvarne vrijednosti godišnjeg minimalnog vodostaja rijeke Nil nalaze unutar granica 95% pouzdanog intervala za predviđene vrijednosti modela.

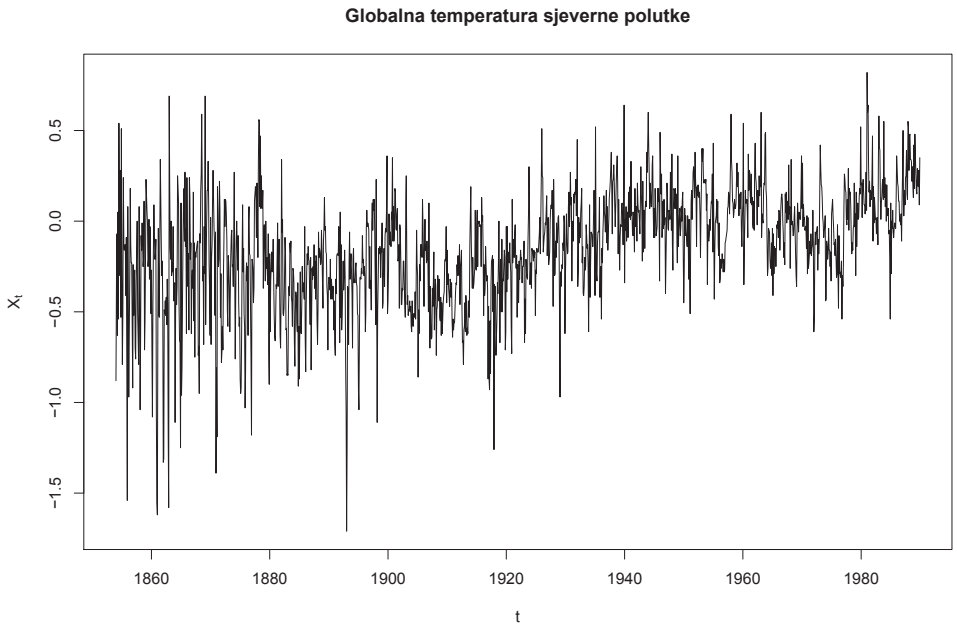


Slika 4.10: *Stvarne i predviđene vrijednosti godišnjeg minimalnog vodostaja rijeke Nil za razdoblje 1260.–1281. godine.*

4.2 Globalna temperatura sjeverne polutke

4.2.1 Opis podataka

Sljedeći vremenski niz koji ćemo modelirati je niz podataka o globalnoj temperaturi sjeverne polutke Zemlje u razdoblju od 1854. do 1990. godine. Radi se o mjesečnim podacima iz baze podataka čuvane na Odsjeku za klimatska istraživanja na Sveučilištu Istočne Anglie, Norwich u Engleskoj. Promatrani vremenski niz $\{x_t\}$ se sastoji od 1632 podatka o promjenama mjesečne globalne temperature u odnosu na referentno razdoblje koje čine prosječne mjesečne temperature u razdoblju 1950.–1979. godine mjerene u Celzijevim stupnjevima. Ovaj skup podataka se također može pronaći u [2]. Na Slici 4.11 prikazane su vrijednosti globalne temperature u razdoblju od 1854. do 1990. godine. Na osnovu tih podataka napravit ćemo model za vremenski niz $\{x_t\}$.



Slika 4.11: *Globalna temperatura sjeverne polutke u razdoblju od 1854. do 1990. godine.*

Numeričke karakteristike vremenskog niza $\{x_t\}$ prikazane su u Tablici 2.

Minimum	Donji kvartil	Medijan	Aritmetička sredina	Gornji kvartil	Maksimum
-1.71	-0.33	-0.12	-0.1495	0.07	0.82

Tablica 2: *Numeričke karakteristike niza podataka o globalnoj temperaturi sjeverne polutke 1854.–1990. godine.*

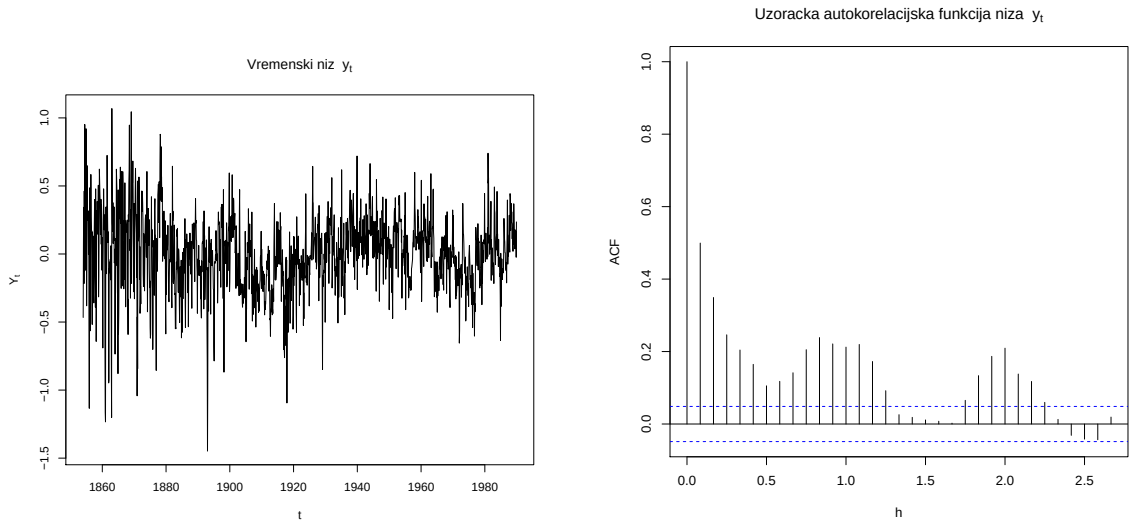
4.2.2 Identifikacija i dijagnostika modela

Ako promotrimo Sliku 4.11, možemo uočiti kako se pojavljuje rastući trend u vremenskom nizu $\{x_t\}$. Taj rastući trend bi mogao biti odraz globalnog zatopljenja prisutnog tijekom zadnjih 100 godina. Zbog toga ima smisla promatrati model s linearnim trendom

$$X_t = \beta_0 + \beta_1 t + Y_t,$$

gdje je $\{Y_t\}$ stacionaran proces. Procjenu parametara trenda napravit ćemo generaliziranom metodom najmanjih kvadrata koja pretpostavlja da vrijedi $\{Y_t\} \sim \text{AR}(1)$. Dobivene procjene su $\widehat{\beta}_0 = -7.612508$ i $\widehat{\beta}_1 = 0.003883$. Pouzdani intervali ovih procjena sugeriraju kako su parametri statistički značajno različiti od nule što ukazuje na

opravdanost rastućeg trenda vremenskog niza $\{x_t\}$.
Oduzimanjem procijenjenog linearnog trenda od promatranog niza $\{x_t\}$, dobijemo niz $\{y_t\}$, gdje je $y_t = x_t - \widehat{\beta}_0 - \widehat{\beta}_1 t$. Vremenski niz $\{y_t\}$ i njegova uzoračka autokorelacijska funkcija prikazani su na Slici 4.12.

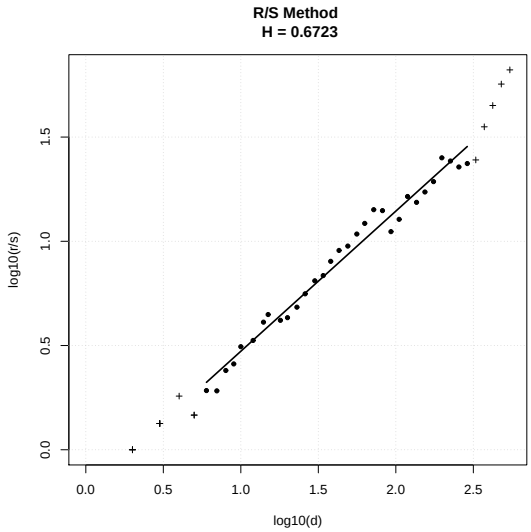


(a) *Vremenski niz* $\{y_t\}$.

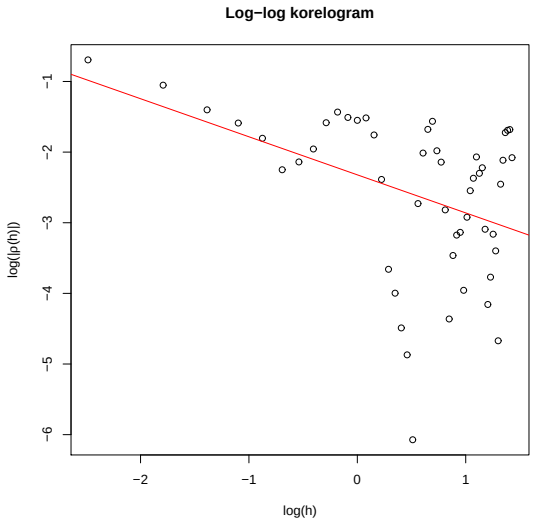
(b) *Uzoračka autokorelacijska funkcija vremenskog niza* $\{y_t\}$.

Slika 4.12: *Vremenski niz* $\{y_t\}$ *i njegova uzoračka autokorelacijska funkcija.*

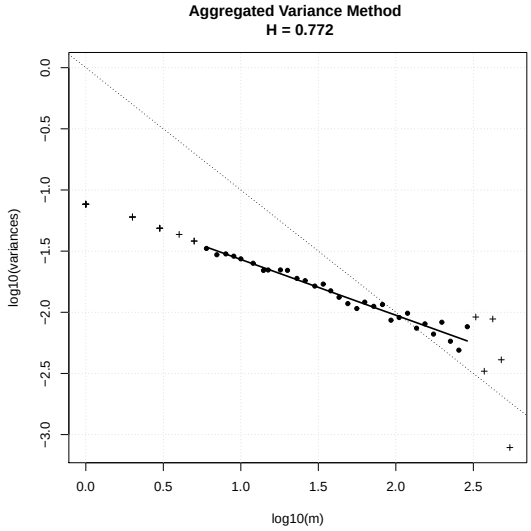
Gledajući Sliku 4.12a, možemo primjetiti duge periode viših i nižih vrijednosti vremenskog niza $\{y_t\}$ te cikluse različitih frekvencija. Iako uzoračka autokorelacijska funkcija niza $\{y_t\}$ upućuje na nekakvu sezonalnost (Slika 4.12b), uočavamo sporo opadanje autokorelacijske funkcije u nulu i naslućujemo da bismo ovaj niz mogli modelirati frakcionalnim ARIMA(p, d, q) modelom. Dugoročnu zavisnost vremenskog niza $\{y_t\}$ ispitat ćemo statističkim metodama iz Poglavlja 3.
Vrijednosti R/S statistike ovog niza gomilaju se oko pravca čiji je procijenjeni koeficijent smjera jednak 0.6723202 (Slika 4.13a) što ukazuje na dugoročnu zavisnost. Log-log korelogram vremenskog niza $\{y_t\}$ prikazan je na Slici 4.13b. Procjena koeficijenta smjera pravca koji aproksimira vrijednosti od $\log|\widehat{\rho}(h)|$ u odnosu na $\log h$ je -0.5389, pa je procjena parametra H po ovoj metodi $\widehat{H} = 0.73055$. Metoda varijance procjenjuje parametar dugoročne zavisnosti niza $\{y_t\}$ na $\widehat{H} = 0.7720168$ što vidimo na Slici 4.13c. Periodogram ovog niza daje procjenu $\widehat{H} = 0.8176876$. Sve ove metode koje smo naveli kao koristan alat za uočavanje dugoročne zavisnosti, pokazuju kako dugoročna zavisnost ovog niza postoji. Svaka od njih je parametar dugoročne zavisnosti



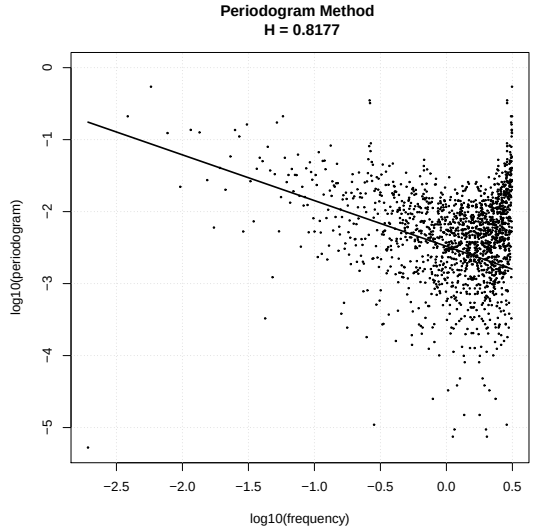
(a) *Graf R/S statistike.*



(b) *Log–log korelogram.*



(c) *Graf metode varijance.*

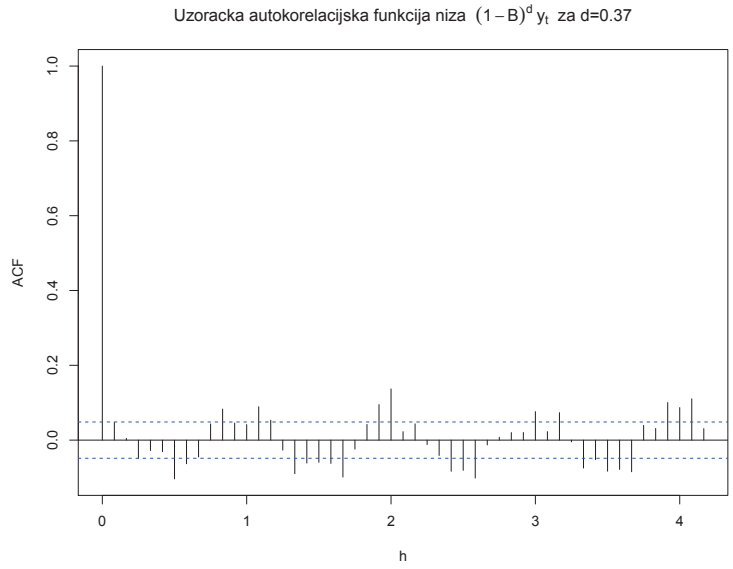


(d) *Log–log periodogram.*

Slika 4.13: *Statistički procjenitelji dugoročne zavisnosti promatranog vremenskog niza* $\{y_t\}$.

H procijenila na vrijednost iz intervala $(0.65, 0.85)$. Zbog toga odlučujemo niz $\{y_t\}$ modelirati frakcionalnim ARIMA(p, d, q) procesom za $d = H - \frac{1}{2}$.
Whittleov aproksimativni ML procjenitelj daje procjenu parametra $\widehat{H} = 0.87009947$, pa ćemo u našem modelu staviti da je $d = 0.37$. Frakcionalnim diferenciranjem niza $\{y_t\}$ za $d = 0.37$ dobijemo niz $\{(1 - B)^{0.37} y_t\}$. Njegova procijenjena autokorelacijska

funkcija prikazana je na Slici 4.14.

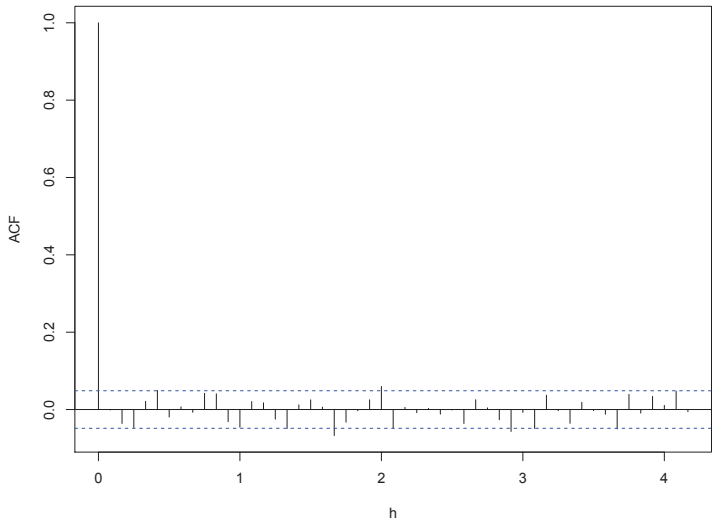


Slika 4.14: *Procijenjena autokorelacijska funkcija frakcionalno diferenciranog vremenskog niza* $\{(1-B)^{0.37}y_t\}$.

Ostatak korelacija koje su statistički značajno različite od nula želimo modelirati ARMA (p, q) procesom. Ispitujući nekoreliranost reziduala ARMA modela za različite redove modela p i q, nismo naišli na model za koji Ljung–Box test ne odbacuje hipotezu o nekoreliranosti reziduala na koracima $h = p + q, \ldots, 20$. Uzimajući u obzir druge vrijednosti parametra $d = H - \frac{1}{2}$, za $H \in (0.65, 0.85)$, također nismo naišli na model koji zadovoljava pretpostavku o nekoreliranosti reziduala. Odlučili smo se niz $\{(1-B)^{0.37}y_t\}$ modelirati ARMA$(2, 3)$ procesom jer je on model s najmanje parametara među modelima čiji su reziduali vizualno najbliže bijelom šumu.
Tako dobiveni FARIMA$(2, 0.37, 3)$ model je definiran s

$$(1-B)^{0.37}(1-1.7319B+0.9984B^2)Y_t = (1-1.7468B+1.0251B^2-0.0242B^3)Z_t,$$
$$Z_t \sim WN(0, 0.05191). \tag{4.3}$$

Procijenjena autokorelacijska funkcija reziduala modela (4.3) prikazana je na Slici 4.15.



Slika 4.15: *Procijenjena autokorelacijska funkcija reziduala modela (4.3).*

Tipično svojstvo stacionarnih procesa s dugoročnim pamćenjem je pojavljivanje lokalnih trendova i ciklusa koji nestaju nakon nekog perioda. Iz tog razloga je teško razlučiti je li neki vremenski niz realizacija stacionarnog procesa s dugoročnim pamćenjem ili nestacionarnog procesa. Ako bismo promatrali jedan vremenski niz podataka o temperaturi, mogli bismo pomisliti da je trend posljedica dugoročne zavisnosti. No veliki broj vremenskih nizova podataka o temperaturi upućuju na povećanje temperature u zadnjih 100 godina. Također, istraživanja u području klimatologije potkrjepljuju postojanje globalnog zatopljenja teorijskim modelima. S obzirom na te rezultate i ostale vremenske nizove podataka o temperaturi, možemo reći da u vremenskom nizu podataka $\{x_t\}$ zaista postoji rastući trend.

Literatura

[1] J. Beran, *A test of location for data with slowly decaying serial correlations*, Biometrika, **76**(1989), 261–269.

[2] J. Beran, *Statistics for Long–Memory Processes*, Chapman & Hall, New York, 1994.

[3] P. J. Brockwell, R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.

[4] P. J. Brockwell, R. A. Davis, *Introduction to Time Series and Forecasting*, Springer-Verlag, New York, 2002.

[5] R. H. Chitale, *Probability and Queueing Theory*, Technical Publications, Pune, 2008.

[6] J. A. Gubner, *Theorems and fallacies in the theory of long–range–dependent processes*, IEEE Transactions on Information Theory, **51**(2005), 1234–1239.

[7] J. R. M. Hosking, *Fractional differencing*, Biometrika, **68**(1981), 165–176.

[8] G. M. Ljung, G. E. P. Box, *On a measure of lack of fit in time series models*, Biometrika, **65**(1978), 297–303.

[9] B. B. Mandelbrot, *Limit theorems of the self–normalized range for weakly and strongly dependent processes*, Z. Wahr. verw. Geb., **31**(1975), 271–285.

[10] J. D. Murray, *Asymptotic Analysis*, Springer-Verlag, New York, 1984.

[11] M. B. Priestley, *Spectral Analysis of Time Series*, Academic Press, London, 1981.

[12] P. Royston, *Remark AS R94: A remark on Algorithm AS 181: The W test for normality*, Applied Statistics, **44**(1995), 547–551.

[13] A. Zygmund, *Trigonometric Series*, Cambridge Univ. Press, Cambridge, U.K., 1959, vol. I.

Sažetak

U ovom diplomskom radu predstavljeni su procesi čija autokorelacijska funkcija nije apsolutno sumabilna. Takvi procesi se nazivaju stacionarni procesi s dugoročnim pamćenjem i svoju primjenu nalaze u različitim područjima poput klimatologije i hidrologije. Navedene su definicije i osnovni rezultati stacionarnih procesa s dugoročnim pamćenjem te dvije klase takvih procesa. Jedna od njih su frakcionalni ARIMA procesi koji su prirodno proširenje klasičnih ARIMA procesa. Prikazane su strukture autokorelacijske funkcije i spektralne gustoće takvih procesa. Zatim su uvedeni statistički procjenitelji koji služe za uočavanje dugoročne zavisnosti u nekom vremenskom nizu. To su R/S statistika, log–log korelogram, graf varijance, periodogram te Whittleov aproksimativni ML procjenitelj. Uvedeni koncept dugoročne zavisnosti primijenjen je na dva vremenska niza: godišnji minimalni vodostaj rijeke Nil i globalnu temperaturu sjeverne polutke Zemlje.

Ključne riječi: autokorelacijska funkcija, stacionarni procesi s dugoročnim pamćenjem, frakcionalni ARIMA procesi, R/S statistika, dugoročna zavisnost.

Summary

In this graduate work, the processes of which the autocorrelation function is not absolutely summable are presented. Such processes are called stationary processes with long memory and its application are found in different areas such as climate and hydrology. Definitions and basic results of stationary processes with long memory and two classes of such processes are presented. One of them are fractional ARIMA processes that are the natural expansion of the classical ARIMA processes. The structures of autocorrelation function and spectral density of such processes are presented. Then, statistical estimators are introduced to identify long–range dependence in some time series. These are R/S statistic, log–log correlogram, variance plot, periodogram and Whittle's approximate ML estimator. The concept of long–range dependence has been introduced in two time series: yearly minimum water level of the Nile River and the global temperature of the northern hemisphere of the Earth.

Keywords: autocorrelation function, stationary processes with long memory, fractional ARIMA processes, R/S statistics, long–range dependence.

Životopis

Rođena sam 13. listopada 1993. godine u Osijeku u Hrvatskoj. Pohađala sam Osnovnu školu "Dobriša Cesarić" od 2000. do 2008. godine. Nakon toga sam upisala III. gimnaziju Osijek. Godine 2012. završila sam srednju školu i upisala Preddiplomski studij matematike na Odjelu za matematiku u Osijeku. Akademski stupanj *prvostupnice matematike* stekla sam 2015. godine napisavši završni rad *Nelder-Meadova metoda: lokalna metoda direktne bezuvjetne optimizacije* pod mentorstvom prof. dr. sc. Kristiana Sabe. Iste godine sam upisala Diplomski studij matematike, smjer Financijska matematika i statistika na Odjelu za matematiku u Osijeku. Tijekom diplomskog studija odradila sam stručnu praksu u tvrtci Farmeron d.o.o. gdje sam, uz pomoć mentora, primijenila znanje statistike u mliječnoj industriji.