# Modeliranje koncentracije peludi ambrozije vektorskim autoregresivnim modelom

**Knežević, Marinela**

**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:126:981679

*Rights / Prava:* In copyright / Zaštićeno autorskim pravom.

*Download date / Datum preuzimanja:* **2024-03-29**



*Repository / Repozitorij:*

Repository of School of Applied Mathematics and Computer Science









Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

**Marinela Knežević**

# Modeliranje koncentracije peludi ambrozije vektorskim autoregresivnim modelom

Diplomski rad

Osijek, 2018.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

**Marinela Knežević**

# Modeliranje koncentracije peludi ambrozije vektorskim autoregresivnim modelom

Diplomski rad

Mentor: izv. prof. dr. sc. Nenad Šuvak

Osijek, 2018.

# Sadržaj

**Uvod** **1**

**1 Osnovne analize vremenskih nizova** **3**

1. Općenito o vremenskom nizu i konceptu stacionarnosti . . . . . 3

1.1 Vremenski niz . . . . . 3

1.2 Stacionarni procesi . . . . . 4

1.3 Nestacionarni procesi . . . . . 7

2. Modeli stacionarnih vremenskih nizova . . . . . 9

2.1 Linearan proces . . . . . 9

2.2 Model pomičnih prosjeka . . . . . 10

2.3 Autoregresivni model . . . . . 11

2.4 Autoregresivni pomični prosjek . . . . . 12

3. Modeli nestacionarnih vremenskih nizova . . . . . 13

3.1 Diferenciranje i integrirani autoregresivni pomični prosjek . . . . 13

3.2 Sezonalni autoregresivni pomični prosjek . . . . . 14

3.3 Sezonalni integrirani autoregresivni pomični prosjek . . . . . 14

**2 Višedimenzionalni vremenski nizovi** **15**

1. Višedimenzionalni vremenski nizovi . . . . . 15

2. Vektorski autoregresivni modeli . . . . . 18

2.1 Stabilnost . . . . . 19

2.2 Stacionarnost . . . . . 21
2.3 Grangerova uzročnost . . . . . 28
2.4 Identifikacija modela . . . . . 31
2.5 Dijagnostika modela . . . . . 35
2.6 Predikcija . . . . . 38

**3 Vektorski autoregresivni model na stvarnim podacima 41**

1. Koraci u izgradnji modela . . . . . 41
1.1 Deskriptivna analiza varijabli . . . . . 42
1.2 Analiza uzročnosti . . . . . 47
1.3 Odabir reda modela . . . . . 48
2. Adekvatnost modela . . . . . 49
2.1 Reziduali . . . . . 53
2.2 Stabilnost i stacionarnost . . . . . 55
3. Predikcija modela . . . . . 55

**Literatura 59**

**Sažetak 60**

**Summary 61**

**Životopis 62**

**Dodatak 63**

# Uvod

Alergija je pretjerano izražena reakcija imunološkog sustava na tvari iz okoline koje su obično bezazlene i neopasne. Takve tvari nazivamo alergenima. Uobičajeni alergeni su različite tvari iz okoliša ili hrane, a najčešći su oni koji u tijelo ulaze dišnim sustavom iz zraka. Dodir osjetljivih osoba s takvim tvarima potiče otpuštanje upalnih tvari iz stanica (npr. sluznica vjeđa i dišnog sustava), što uzrokuje simptome alergije: curenje, začepljenost ili svrbež nosa, peckanje ili škakljanje u grlu, suzenje ili svrbež te crvenilo i otekline oko očiju, otežano disanje, iznenadno kihanje, kašljanje, glavobolja i slično.
Pelud je najčešći uzrok alergijskih bolesti dišnog sustava i jedan od najsnažnijih prirodnih alergena. U zraku ga je gotovo nemoguće izbjeći. Kada dođe u dodir s vlažnom sredinom poput sluznice našeg dišnog sustava, pelud otpušta niz kemijskih alergogenih spojeva (njih čak 52, prema [6]). Pojava alergija na pelud se podudara s cvatnjom biljaka koje stvaraju taj pelud, zbog čega je važno poznavati peludni kalendar koji daje podatke o sezoni cvjetanja biljaka koje izazivaju simptome alergije. Peludni kalendar je grafički prikaz peludnog spektra u zraku, a najčešće se izrađuje za razdoblje od godine dana. Izrađuje se zasebno za svako biogeografsko područje zbog različitosti biljnih vrsta i različitosti početaka cvatnje. Jako alergenu pelud ima drveće lijeske, johe i breze, od korovnih biljaka pelin i ambrozija te pripadnici porodice trava. Izrazito osjetljive osobe razviju simptome alergijske reakcije već pri niskoj razini peluda, većina osjetljivih osoba pri umjerenoj, dok pri visokim i vrlo visokim razinama sve osobe osjetljive na pelud razviju simptome alergijske reakcije. Procjenjuje se da u Republici Hrvatskoj 7-10 % stanovništva pati od peludne alergije.
Meteorološki uvjeti poput dnevne temperature, vlažnosti zraka, vjetra i padalina igraju veliku ulogu u koncentraciji peludi u smislu taloženja, prijenosa zrakom, ispiranja uslijed padalina i slično. U ovom diplomskom radu su korišteni podaci o koncentraciji peludi ambrozije u gradu Somboru, Republika Srbija, za razdoblje od siječnja 2008. do srpnja 2017. godine te podaci o meteorološkim uvjetima na mjernoj postaji u Rimskim Šančevima za isto vremensko razdoblje. Na osnovu danih podataka cilj je pronaći model koji dovoljno dobro prati kretanje tih pojava i predviđa njihove buduće vrijednosti, kako bi pojedincima koji pate od alergije olakšali taj period u godini. Primjena ovog ili sličnih modela za predikciju buduće dnevne koncentracije peludi podrazumijeva stalno pristizanje novih mjerenja, tj. novih podataka, i korekciju parametara modela u skladu s njima. U tom kontekstu ovaj se problem može svrstati u kategoriju problema iz popularnog *"big data"* područja.
U radu su najprije objašnjeni osnovni pojmovi analize vremenskih nizova koji će se koristiti u ostatku diplomskog rada. Model koji se pokazao prikladan za ovakav tip

podataka je vektorski autoregresivni model (VAR model) te su opisani rezultati koji su osnova za razumijevanje ovakve vrste modela. Na kraju je napravljena predikcija budućih koncentracija peludi te su dobivene vrijednosti podijeljene u kategorije - niska, umjerena, visoka i vrlo visoka koncentracija. Logičan nastavak ovog smjera modeliranja koncentracije peludi je implementacija VAR modela u domeni strojnog učenja (*engl. machine learning*), no to prelazi granice ovog diplomskog rada.

# Poglavlje 1

# Osnovne analize vremenskih nizova

## 1. Općenito o vremenskom nizu i konceptu stacionarnosti

### 1.1 Vremenski niz

**Definicija 1.1.** (Slučajni proces) ***Slučajni proces*** *je familija* $\{X_t, t \in T\}, T \subseteq \mathbb{R}$ *slučajnih varijabli definiranih na istom vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, P)$.

Dva bitna skupa koja karakteriziraju slučajni proces su:

- skup stanja - skup svih mogućih realizacija slučajnih varijabli $X_t$

- skup indeksa ili parametarski skup $T$
  $T$ može biti diskretan pa govorimo o slučajnom procesu u diskretnom vremenu, odnosno nizu slučajnih varijabli ili neprekidan pa govorimo o slučajnom procesu u neprekidnom vremenu.

**Vremenski niz** je niz podataka $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ prikupljenih u uzastopnim vremenskim trenutcima $t_1, t_2, \ldots, t_n$. Svaki podatak vremenskog niza $x_{t_i}, i = 1, \ldots, n$ smatramo realizacijom jedne slučajne varijable $X_{t_i}, i = 1, \ldots, n$, odnosno, vremenski niz je realizacija slučajnog procesa $\{X_t, t \in T\}$, u trenutcima iz $T_0, T_0 = \{t_1, t_2, \ldots, t_n\}, T_0 \subseteq T$. Slučajni proces $\{X_t, t \in T\}$ zovemo model vremenskog niza. Svrha analize vremenskih nizova je modelirati i razumjeti stohastički mehanizam koji dovodi do realizacije kojom raspolažemo te predviđati buduće vrijednosti vremenskog niza.

## 1.2 Stacionarni procesi

**Definicija 1.2.** (Strogo stacionaran proces) *Slučajni proces* $\{X_t, t \in T\}$ *je* ***strogo stacionaran (stacionaran u užem smislu)*** *ako za svaki* $h > 0$ *vrijedi:*

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}), \forall t_1, \ldots, t_n, n \in \mathbb{N}.$$

Posebno, $X_t \stackrel{d}{=} X_s, \forall s, t \in T$.
Promatrajmo nadalje diskretan T, pri čemu je $T = \mathbb{Z}$.

**Definicija 1.3.** *Neka je* $\{X_t, t \in \mathbb{Z}\}$ *slučajni proces takav da je* $E[X_t^2] < \infty, \forall t \in \mathbb{Z}$.
***Funkcija očekivanja procesa*** $\{X_t, t \in \mathbb{Z}\}$ *je* $\mu \colon \mathbb{Z} \to \mathbb{R}, \mu(t) = E[X_t]$.
***Funkcija autokovarijanci (ACVF) procesa*** $\{X_t, t \in \mathbb{Z}\}$ *je* $\gamma \colon \mathbb{Z}^2 \to \mathbb{R}$, *definirana s*

$$\gamma(t, s) = Cov(X_t, X_s) = E[(X_t - E[X_t])(X_s - E[X_s])] = E[X_t X_s] - E[X_t]E[X_s].$$

***Autokorelacijska funkcija (ACRF) procesa*** $\{X_t, t \in \mathbb{Z}\}$ *je* $\rho \colon \mathbb{Z}^2 \to [-1, 1]$, *definirana s*

$$\rho(t, s) = Corr(X_t, X_s) = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}} = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t)\gamma(s, s)}}.$$

**Definicija 1.4.** (Slabo stacionaran proces) *Slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ *je* ***slabo stacionaran (stacionaran u širem smislu)*** *ako vrijedi:*

*(i)* $E[X_t^2] < \infty, \forall t \in \mathbb{Z}$

*(ii)* $\mu(t) = c, \forall t \in \mathbb{Z}$, *gdje je* $c \in \mathbb{R}$ *konstantno očekivanje*

*(iii)* $\gamma(t, s) = \gamma(t + h, s + h), \forall t, s, h \in \mathbb{Z}$.

Uočimo, ako u (iii) stavimo $s = t$, dobivamo $\gamma(t, t) = \gamma(t+h, t+h)$, odnosno varijanca stacionarnog procesa je konstantna. Stavimo li u (iii) $h = -t$, onda vrijedi $\gamma(t, s) = \gamma(0, s - t), \forall t, s \in \mathbb{Z}$. Dakle, funkcija autokovarijanci ovisi samo o razlici $s - t$ pa je možemo promatrati kao funkciju jedne varijable: $\gamma(h) = \gamma(0, h) = Cov(X_t, X_{t+h})$. Ubuduće, stacionaran proces će podrazumijevati proces stacionaran u širem smislu.

**Propozicija 1.1.** (Svojstva funkcije autokovarijanci stacionarnog procesa) *Za funkciju autokovarijanci* $\gamma$ *stacionarnog procesa* $\{X_t, t \in \mathbb{Z}\}$ *vrijedi:*

*(i)* $\gamma(0) \geq 0$

*(ii)* $|\gamma(h)| \leq \gamma(0), \forall h \in \mathbb{Z}$

*(iii)* $\gamma(h) = \gamma(-h), \forall h \in \mathbb{Z}$.

Analogno za autokorelacijsku funkciju stacionarnog procesa.

**Propozicija 1.2.** (Svojstva autokorelacijske funkcije stacionarnog procesa) *Za autokorelacijsku funkciju $\gamma$ stacionarnog procesa $\{X_t, t \in \mathbb{Z}\}$ vrijedi:*

*(i)* $\rho(0) = 1$

*(ii)* $|\rho(h)| \leq 1, \forall h \in \mathbb{Z}$

*(iii)* $\rho(h) = \rho(-h), \forall h \in \mathbb{Z}$.

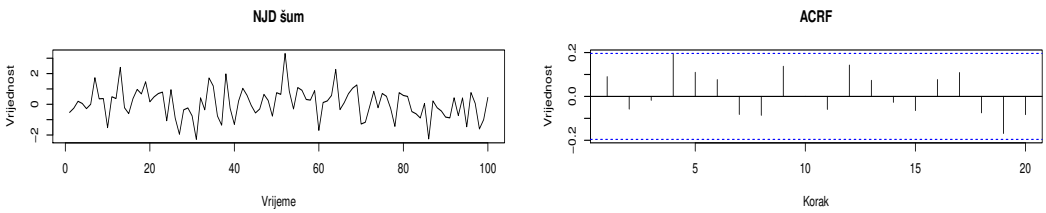
Navedena svojstva proizlaze iz samih definicija funkcija i dokazi se mogu vrlo lako pokazati.
Neki od primjera stacionarnih procesa su niz nezavisnih jednako distribuiranih slučajnih varijabli, bijeli šum itd.

**Definicija 1.5.** (Nezavisni jednako distribuirani šum (NJD šum)) *Neka je $\{X_t, t \in \mathbb{Z}\}$ niz nezavisnih jednako distribuiranih slučajnih varijabli takvih da vrijedi:*

- $E[X_t^2] < \infty, \forall t \in \mathbb{Z}$
- $E[X_t] = 0, \forall t \in \mathbb{Z}$
- $Var(X_t) = \sigma^2, \forall t \in \mathbb{Z}$.

*Tada je $\{X_t\}$* ***nezavisni jednako distribuirani šum****, oznaka: $\{X_t\} \sim IID(0, \sigma^2)$.*

**Primjer 1.1.** (NJD šum) *NJD šum takav da je $X_t \sim \mathcal{N}(0, 1)$.*




(a) Trajektorija (b) Empirijska ACRF

Slika 1.1: NJD šum

**Definicija 1.6.** (Bijeli šum) *Neka je* $\{X_t, t \in \mathbb{Z}\}$ *niz slučajnih varijabli takvih da vrijedi:*

- $E[X_t^2] < \infty, \forall t \in \mathbb{Z}$
- $E[X_t] = 0, \forall t \in \mathbb{Z}$
- $Var(X_t) = \sigma^2, \forall t \in \mathbb{Z}$
- $Cov(X_t, X_s) = 0, \forall t \neq s.$

*Tada je* $\{X_t\}$ ***bijeli šum****, oznaka:* $\{X_t\} \sim \mathcal{WN}(0, \sigma^2)$.

Uočimo da slučajne varijable ne moraju biti jednako distribuirane i nezavisne, bijeli šum je relaksacija NJD šuma.

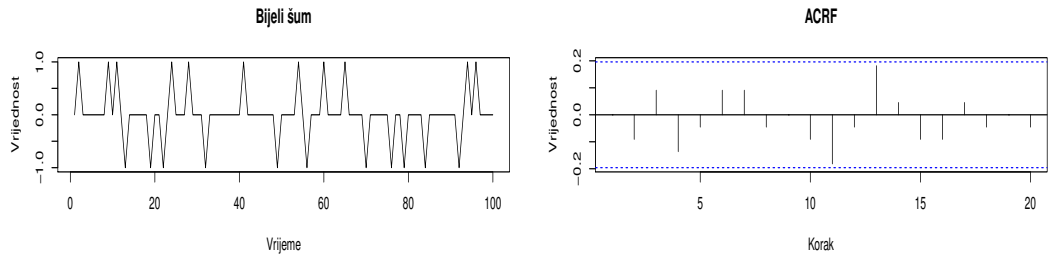
**Primjer 1.2.** (Bijeli šum) *Neka su* $\{Y_t, t \in \mathbb{Z}\}$ *i* $\{Z_t, t \in \mathbb{Z}\}$ *nizovi nezavisnih jednako distribuiranih slučajnih varijabli takvi da je:*

$$Y_t \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

*i*

$$Z_t \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

*te međusobno nezavisni. Proces* $\{X_t, t \in \mathbb{Z}\}$ *definiran s* $X_t = Y_t(1-Y_t)Z_t$ *je bijeli šum.*




(a) Trajektorija (b) Empirijska ACRF

Slika 1.2: Bijeli šum

## 1.3 Nestacionarni procesi

Ukoliko slučajni proces ne zadovoljava neku od pretpostavki iz Definicije 1.4., kažemo da proces nije stacionaran. Nestacionarnost može imati različite oblike, može naprimjer dolaziti od deterministčkog trenda, od promjenjivosti varijance, od egzogenih šokova, iz sezonalnosti itd.

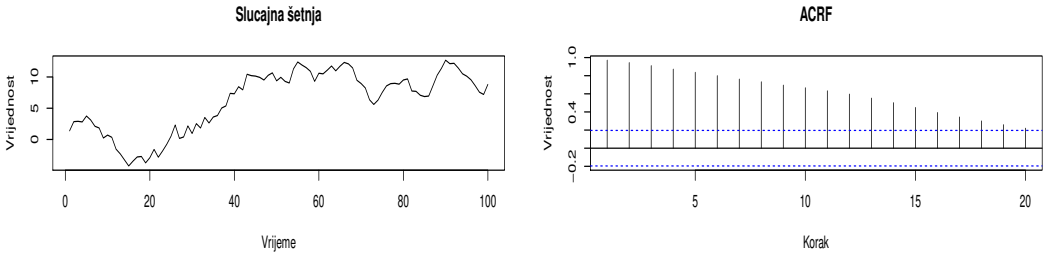
**Definicija 1.7.** (Slučajna šetnja) *Neka je* $\{Y_t\} \sim IID(0, \sigma^2)$. ***Slučajna šetnja*** *je proces* $\{X_t, t \in \mathbb{N}_0\}$ *definiran s:*

$$X_0 = 0, X_t = \sum_{i=1}^{t} Y_i, t \in \mathbb{N}.$$

Uočimo da proces nije stacionaran jer varijanca nije konstantna:

$$Var(Y_t) = Var\left(\sum_{i=1}^{t} Y_i\right) = tVar(Y_1) = t\sigma^2.$$

**Primjer 1.3.** (Slučajna šetnja) *Neka je* $\{Y_t\} \sim IID(0, 1)$ *iz standardne normalne distribucije te* $\{X_t\}$ *pripadajuća slučajna šetnja. Očigledna je promjenjivost varijance slučajne šetnje (Slika 1.3).*



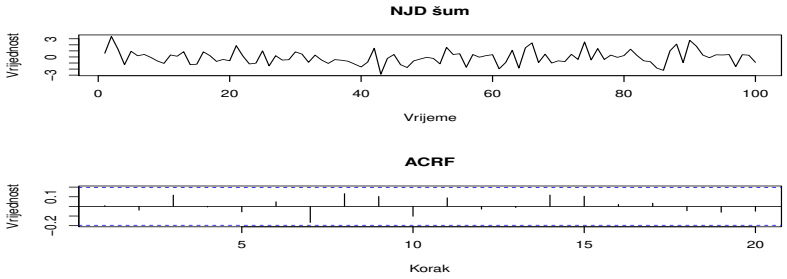

(a) Trajektorija (b) Empirijska ACRF

Slika 1.3: Slučajna šetnja

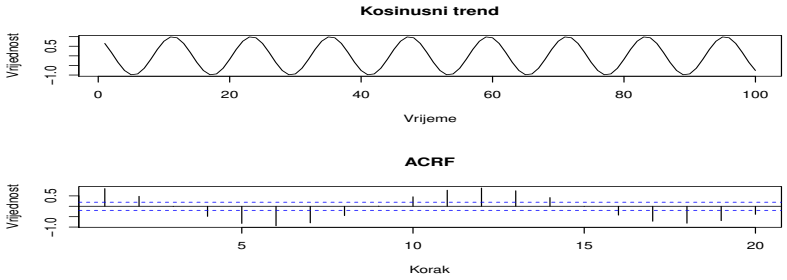
**Definicija 1.8.** (Proces s deterministčkim trendom) *Kažemo da proces* $\{X_t\}$ *ima* ***deterministički trend*** *ako* $\forall t \in \mathbb{Z}$ *vrijedi* $X_t = \mu_t + Y_t$, *gdje je* $\{Y_t\}$ *stacionaran proces takav da je* $E[Y_t] = 0$, *a* $\mu \colon \mathbb{Z} \to \mathbb{R}$ *funkcija koja nije konstantna.*

$\{X_t\}$ nije stacionaran, ali $\{X_t - \mu_t\}$ je. Oblik funkcije $\mu$ obično procijenimo na osnovu podataka, a česti su linearni, polinomijalni, kosinunsni trendovi itd.

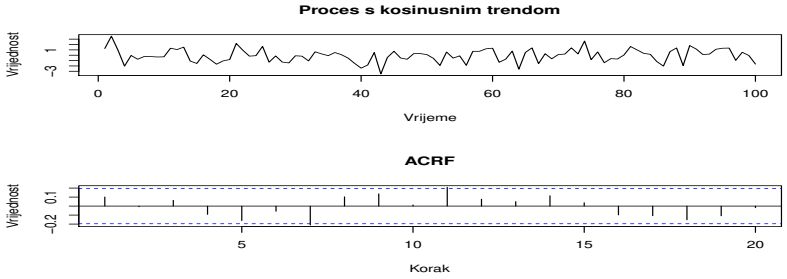
**Primjer 1.4.** (Kosinusni trend) *U slučaju kosinusnog trenda, $\mu$ je općenito oblika: $\mu_t = \beta cos\,(2\pi ft + \Phi)$, gdje je $\beta$ amplituda, $f$ period, i $\Phi$ faza. Primjera radi, neka je $Y_t \sim IID(0,1)$ iz standardne normalne distribucije, $\mu_t = cos(2\pi\frac{1}{12}t + \Phi)$ kosinusni trend tako da je faza slučajno odabran broj iz uniformne distribucije na intervalu (0,1), period je 1/12, a amplituda je jednaka 1. Na taj način dobijemo sljedeći nestacionarni slučajni proces $X_t$.*




(a) NJD šum




(b) Kosinusni trend




(c) Nestacionarni proces s kosinusnim trendom

Slika 1.4: Trajektorije i empirijske ACRF

# 2. Modeli stacionarnih vremenskih nizova

Pretpostavimo da podaci sugeriraju kako se radi o realizaciji stacionarnog procesa pa tražimo stacionaran slučajni proces za model tog vremenskog niza. Pretpostavit ćemo i da traženi model ima očekivanje 0. Opći okvir za proučavanje stacionarnih procesa nam daju linearni procesi.

## 2.1 Linearan proces

**Definicija 1.9.** (Linearan proces) *Kažemo da je* $\{X_t, t \in \mathbb{Z}\}$ ***linearan proces*** *ako se može prikazati kao:*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \forall t \in \mathbb{Z}, \tag{1.1}$$

*gdje je* $\{Z_t\} \sim \mathcal{WN}(0, \sigma^2)$ *i* $(\psi_j)_{j\in\mathbb{Z}}$ *niz konstanti takav da je* $\sum_{j=-\infty}^{\infty} |\psi_j| \leq \infty$.

Odnosno, proces je linearan ako se može zapisati kao linearna kombinacija prošlih, sadašnjih i budućih vrijednosti bijelog šuma.

**Napomena 1.1.**

*(i) Nekad ćemo pretpostaviti da je* $\{Z_t\} \sim IID(0, \sigma^2)$ *i reći da se radi o linearnom procesu uz NJD šum.*

*(ii) Desna strana izraza (1.1) je red slučajnih varijabli i shvaćamo ga kao* $X_t = lim_{n\to\infty} \sum_{j=-n}^{n} \psi_j Z_{t-j}$*, gdje je limes u smislu gotovo sigurno ili* $L^2$ *smislu (vidi [2], Poglavlje 2.2).*

**Definicija 1.10.** (Operator pomaka unazad) *Za proces* $\{X_t, t \in \mathbb{Z}\}$ ***operator pomaka unazad*** *(engl. backward shift) je definiran s:*

$$BX_t = X_{t-1}.$$

Analogno, uzastopnom primjenom $j$ puta, dobivamo

$$B^j X_t = B^{j-1} B X_t = B^{j-1} X_{t-1} = \cdots = X_{t-j}.$$

Linearan proces (1.1) možemo zapisati kao:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j Z_t = \psi(B) Z_t,$$

gdje je

$$\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j.$$

**Definicija 1.11.** (Linearan filter) *Neka je* $\{Y_t, t \in \mathbb{Z}\}$ *slučajni proces i* $\{\beta_j, j \in \mathbb{Z}\}$ *niz koeficijenata. Za slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ *definiran s*

$$X_t = \sum_{j=-\infty}^{\infty} \beta_j Y_{t-j}, t \in \mathbb{Z},$$

*kažemo da je linearan filter procesa* $\{Y_t\}$.

**Propozicija 1.3.** *Neka je* $\{Y_t\}$ *stacionaran proces s očekivanjem 0 i funkcijom autokovarijanci* $\gamma_y$. *Ako je* $\psi(B)$ *linearan filter takav da je* $\sum_{j=-\infty}^{\infty} |\psi_j| \leq \infty$, *onda je proces*

$$X_t = \psi(B)Y_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}, t \in \mathbb{Z}$$

*stacionaran s očekivanjem 0 i vrijedi:*

$$\gamma_x(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_y(h+k-j), h \in \mathbb{Z}.$$

*Specijalno, ako je* $\{X_t\}$ *linearan proces,*

$$\gamma_x(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2, h \in \mathbb{Z}.$$

Dakle, linearan filter stacionarnog proces je opet stacionaran proces. Dokaz vidljiv u [2], Propozicija 2.2.1.

## 2.2 Model pomičnih prosjeka

Prisjetimo se **Woldove dekompozicije** procesa: ako je $\{X_t, t \in \mathbb{Z}\}$ stacionaran nedeterminisitčki proces, onda je on zbroj dva nekorelirana procesa, jednog determinističkog koji se može linearno izraziti iz proizvoljno daleke prošlosti procesa $\{X_t\}$ i jednog nedeterminističog koji je linearna kombinacija sadašnjih i prošlih vrijednosti nekog bijelog šuma. Drugim riječima, $\{X_t\}$ se može prikazati u obliku:

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

gdje je:

(i) $\psi_0 = 1, \sum_{j=0}^{\infty} \psi_j^2 < \infty$

(ii) $\{Z_t\} \sim \mathcal{WN}(0, \sigma^2)$

(iii) $Cov(Z_t, V_s) = 0, \forall s, t \in \mathbb{Z}$

(iv) $Z_t = \Pi_t Z_t, \forall t \in \mathbb{Z}$, a $\Pi_t$ je projekcija na $\{X_s, s \leq t\}$

(v) $V_t = \Pi_s V_t, \forall s, t \in \mathbb{Z}$

(vi) $\{V_t\}$ je deterministički proces.

Vidimo da je prvi član "približno" linearan proces jer ne vrijedi nužno $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Dakle, iz Woldove dekompozicije slijedi da je, uz neke pretpostavke, stacionaran proces i linearan. Više o ovoj dekompoziciji u [3], Poglavlje 5.7.
Promatrajmo sad linearan proces $\{X_t\}$ kao u (1.1). $\{X_t\}$ se može zapisati i kao:

$$X_t = \sum_{j=-\infty}^{-1} \psi_j Z_{t-j} + \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

Uočimo da je prvi član sume zapravo budućnost, a drugi sadašnjost i prošlost bijelog šuma. Za lineran proces za kojeg je $\psi_j = 0, j \leq -1$, kažemo da je **kauzalan** (neovisan o budućnosti). U Woldovoj dekompoziciji je nedeterministički dio kauzalan te stoga možemo promatrati samo linearne procese oblika:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, t \in \mathbb{Z}. \tag{1.2}$$

Umjesto (1.2), možemo promatrati linearne procese za koje je $\psi_j = 0, j > q, q \in \mathbb{N}$. To će biti upravo proces pomičnih prosjeka kojim se opisuju slučajni procesi koji generiraju vremenski niz čija je trenutna vrijednost povezana s greškama trenutnog i prethodnih razdoblja.

**Definicija 1.12.** (Proces pomičnih prosjeka) *Slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ *je* ***proces pomičnih prosjeka reda*** $q \in \mathbb{N}$ *ako je*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, t \in \mathbb{Z},$$

*gdje je* $\{Z_t\} \sim \mathcal{WN}(0, \sigma^2), \theta_1, \ldots, \theta_q \in \mathbb{R}$, $\theta_q \neq 0$. *Oznaka:* $\{X_t\} \sim MA(q)$.

## 2.3 Autoregresivni model

Ideja ovih modela jest sadašnju vrijednost procesa prikazati kao linearnu kombinaciju $p$ prošlih vrijednosti uz šum.

**Definicija 1.13.** (Autoregresivni proces) *Slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ *je* ***autoregresivni proces reda p*** $\in \mathbb{N}$ *ako je stacionaran i ako je*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t, t \in \mathbb{Z}, \tag{1.3}$$

*gdje je* $\{Z_t\} \sim \mathcal{WN}(0, \sigma^2)$ *i* $\phi_1, \phi_2, \ldots, \phi_p \in \mathbb{R}, \phi_p \neq 0$. *Oznaka:* $\{X_t\} \sim AR(p)$.

Izraz (1.3) možemo zapisati kao $\phi(B)X_t = Z_t$, gdje je $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 \ldots \phi_p z^p, z \in \mathbb{C}$. Navedeni polinom se naziva i **karakteristični AR polinom**, a B je operator pomaka unazad.

## 2.4 Autoregresivni pomični prosjek

Ukoliko je sadašnja vrijednost procesa linearna kombinacija $p$ prošlih vrijednosti procesa te sadašnje i $q$ prošlih vrijednosti šuma, govorimo o autoregresivnom pomičnom prosjeku kao kombinaciji prethodna dva procesa.

**Definicija 1.14.** (Autoregresivni pomični prosjek) *Slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ *je* ***autoregresivni pomični prosjek reda (p,q)****,*$p, q \geq 0$ *ako je stacionaran i ako je*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, t \in \mathbb{Z}, \tag{1.4}$$

*gdje je* $\{Z_t\} \sim \mathcal{WN}(0, \sigma^2), \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \in \mathbb{R}, \phi_p, \theta_q \neq 0$ *i polinomi* $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 \ldots \phi_p z^p$ *i* $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$, $z \in \mathbb{C}$ *nemaju zajedničke nultočke. Oznaka:* $\{X_t\} \sim ARMA(p, q)$.

Koristeći operator pomaka unazad, (1.4) možemo zapisati kao $\phi(B)X_t = \theta(B)Z_t$.
Nadalje, neka su $\alpha(B)$ i $\beta(B)$ linearni filteri, odnosno:

$$\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j, \sum_{j=-\infty}^{\infty} |\alpha_j| < \infty$$

i

$$\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j, \sum_{j=-\infty}^{\infty} |\beta_j| < \infty.$$

Kompozicicja ovih operatora je linearan filter s koeficijentima $\psi_j$ koji se odrede iz zahtjeva $\psi(z) = \alpha(z)\beta(z)$ i slijedi

$$\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta j - k.$$

**Teorem 1.1.** *Stacionarno rješenje jednadžbe (1.4) postoji i jedinstveno je ako i samo ako* $\phi(z) \neq 0, \forall z \in \mathbb{C}$ *takav da je* $|z| = 1$.

Odnosno, stacionarno rješenje jednadžbe (1.4) postoji i jedinstveno je ako i samo ako AR polinom nema jedinični korijen.

# 3. Modeli nestacionarnih vremenskih nizova

Sjetimo se da smo spomenuli, ukoliko vremenski niz $x_1, \ldots, x_n$ sugerira da ga ima smisla modelirati nekim stacionarnim procesom, logično ga je pokušati modelirati nekim ARMA procesom. No, u praksi je dosta čest slučaj da podaci ne sugeriraju stacionarnost tako te ih je potrebno transformirati i doći do vremenskog niza čija trajektorija sugerira stacionarnost. Spomenuti su i neki najčešći oblici nestacionarnosti - deterministički trend i sezonalnost. U tu klasu treba dodati i procese sa stohastičkim trendom.

## 3.1 Diferenciranje i integrirani autoregresivni pomični prosjek

**Definicija 1.15.** (Operator diferenciranja) *Za slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ ***operator diferenciranja*** $\boldsymbol{\Delta}$ *je definiran s:* $\Delta X_t = X_t - X_{t-1} = (1-B)X_t$.

Primjenom $d \in \mathbb{N}$ puta operator diferenciranja, dobivamo $\Delta^d X_t = (1-B)^d X_t$.

**Definicija 1.16.** (Integrirani ARMA proces) *Neka je* $d \in \mathbb{N}_0$. *Slučajni proces* $\{X_t\}$ *je* ***integrirani ARMA proces reda (p,d,q)*** *ako je* $Y_t = \Delta^d X_t$ *kauzalan ARMA(p,q) proces. Oznaka:* $\{X_t\} \sim ARIMA(p,d,q)$.

Dakle, diferenciranjem $d$ puta, dobijemo kauzalan ARMA proces. Koristeći standardne oznake, možemo pisati:

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

Uz oznaku $\widetilde{\phi}(B) = \phi(B)(1-B)^d$, dobivamo $\widetilde{\phi}(B)X_t = \theta(B)Z_t$. Dakle, polinom $\widetilde{\phi}(z) = \phi(z)(1-z)^d$ ima jedinični korijen. Kažemo da proces $\{X_t\}$ ima **stohastički trend**.

Pitanje je za podatke vremenskog niza ima li pripadni model jedinični korijen? Treba li diferencirati? Ako treba, koliko puta? U tu svrhu koristimo **prošireni Dickey-Fuller test** (*engl. Augmented Dickey Fuller test*) - test jediničnog korijena (ADF test). Hipoteze su sljedeće:

$\mathcal{H}_0$: postoji jedinični korijen
$\mathcal{H}_1$: ne postoji jedinični korijen.

Uočimo, $\mathcal{H}_0$ povlači da je proces nestacionaran, dok $\mathcal{H}_1$ ne povlači da je proces stacionaran. Proces i dalje može biti nestacionaran iako nema jediničnog korijena (tada je izvor nestacionarnosti nešto drugo). Dakle, ovaj test koristimo kad želimo zaključiti ima li smisla diferencirati vremenski niz.

## 3.2 Sezonalni autoregresivni pomični prosjek

Promatramo modele za podatke vremenskog niza koji pokazuju jasan sezonalni period $s$.

**Definicija 1.17.** (Sezonalni ARMA proces) *Za proces* $\{X_t\}$ *kažemo da je* ***sezonalni ARMA proces reda*** $(\mathbf{p}, \mathbf{q}) \times (\mathbf{P}, \mathbf{Q})_\mathbf{s}$, $p, q, P, Q \in \mathbb{N}_0, s \in \mathbb{N}$, *ako je*

$$\phi(B)\Phi(B^s)X_t = \theta(B)\Theta(B^s)Z_t, t \in \mathbb{Z},$$

*gdje je* $\{Z_t\} \sim \mathcal{WN}(0, \sigma^2)$, *s period, a*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p,$$

$$\Phi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_P z^P,$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q,$$

$$\Theta(z) = 1 + \vartheta_1 z + \vartheta_2 z^2 + \cdots + \vartheta_Q z^Q.$$

*Oznaka:* $\{X_t\} \sim SARMA(p, q) \times (P, Q)_s$.

Uočimo da se radi o ARMA procesima, ali veći broj parametara je 0.

## 3.3 Sezonalni integrirani autoregresivni pomični prosjek

**Definicija 1.18.** (Operator sezonalnog diferenciranja) *Za proces* $\{X_t\}$ ***operator sezonalnog diferenciranja na koraku*** $\mathbf{s} \in \mathbb{N}$ *definiran je s:* $\Delta_s X_t = X_t - X_{t-s} = (1 - B^s)X_t$.

**Definicija 1.19.** (Sezonalni ARIMA proces) *Neka su* $d, D \in \mathbb{N}_0$. *Kažemo da je* $\{X_t\}$ ***sezonalni ARIMA proces reda*** $(\mathbf{p}, \mathbf{d}, \mathbf{q}) \times (\mathbf{P}, \mathbf{D}, \mathbf{Q})_\mathbf{s}$ *ako je* $\Delta^d \Delta_s^D X_t$ *stacionaran* $ARMA(p, q) \times (P, Q)$ *proces. Oznaka:* $\{X_t\} \sim SARIMA(p, d, q) \times (P, D, Q)_s$.

Uz prijašnje oznake, vrijedi: $\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D X_t = \theta(B)\Theta(B^s)Z_t$.

# Poglavlje 2

# Višedimenzionalni vremenski nizovi

U prethodnim poglavljima smo razmatrali kako modelirati vremenski niz podataka nekim slučajnim procesom (modelom vremenskog niza). No, u praksi je prilično čest slučaj da raspolažemo s više vremenskih nizova i želimo ih modelirati istovremeno. Postoje generalizacije spomenutih modela koje svaki vremenski niz shvaćaju kao jednu komponentu višedimenzionalnog (vektorskog) niza.
Američki ekonomist Christopher Albert Sims je zaslužan za primjenu višedimenzionalnih vremenskih nizova i njihovu popularizaciju u makroekonomiji. Zajedno s Thomasom Sargentom je 2011. godine dobio Nobelovu nagradu za ekonomiju koja se dodjeljuje za iznimna postignuća i otkrića u tom području (vidljivo u [9]).

## 1. Višedimenzionalni vremenski nizovi

**Napomena 2.1.** *U daljnjim razmatranjima ćemo radi jednostavnosti izostaviti notaciju podebljanim matematičkim slovima, kao npr.* $\mathbf{X}_t$ *i pisat ćemo* $X_t$*, ali će iz konteksta biti jasno da se radi o vektoru.*

Pretpostavimo da raspolažemo s $m \geq 2$ vremenskih nizova $x_{i1}, x_{i2}, \ldots, x_{in}, i = 1, \ldots, m$. Model za ovakve podatke je višedimenzionalni (vektorski) slučajni proces $\{X_t, t \in \mathbb{Z}\}$.

**Definicija 2.1.** (Vektorski slučajni proces) ***Vektorski slučajni proces*** *je familija* $\{X_t, t \in T\}, T \subseteq \mathbb{R}$ *slučajnih vektora definiranih na istom vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, P)$.

U fiksnom trenutku $t \in \{1, \ldots, n\}$ imamo vektor $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{m,t})^T$. Za fiksni $i \in 1, \ldots, m, \{X_{i,t}, t \in \mathbb{Z}\}$ je slučajni proces. Navedimo sad neka osnovna svojstva i osnovne funkcije za opisivanje ovakvih vremenskih nizova.

**Definicija 2.2.** *Neka je* $\{X_t, t \in \mathbb{Z}\}$ *vektorski slučajni proces.*

- ***Funkcija očekivanja procesa*** $\mu\colon \mathbb{Z} \to \mathbb{R}^m$ *:*

$$\mu(t) = E[X_t] = \begin{pmatrix} E[X_{1,t}] \\ \vdots \\ E[X_{m,t}] \end{pmatrix}$$

- ***Funkcija kovarijanci procesa*** $\Gamma\colon \mathbb{Z}^2 \to \mathbb{R}^{m\times m}$*:*

$$\Gamma(t,s) = Cov(X_t, X_s) = \begin{pmatrix} \gamma_{11}(t,s) & \dots & \gamma_{1m}(t,s) \\ \vdots & \ddots & \vdots \\ \gamma_{m1}(t,s) & \dots & \gamma_{mm}(t,s) \end{pmatrix},$$

*gdje je* $\gamma_{ij}(t,s) = Cov(X_{i,t}, X_{j,s})$.

- ***Autokorelacijska funkcija procesa*** $R\colon \mathbb{Z}^2 \to \mathbb{R}^{m\times m}$*:*

$$R(t,s) = Corr(X_t, X_s) = \begin{pmatrix} \rho_{11}(t,s) & \dots & \rho_{1m}(t,s) \\ \vdots & \ddots & \vdots \\ \rho_{m1}(t,s) & \dots & \rho_{mm}(t,s) \end{pmatrix},$$

*gdje je* $\rho_{ij}(t,s) = \dfrac{Cov(X_{i,t}, X_{j,s})}{\sqrt{Var(X_{i,t})Var(X_{j,s})}} = \dfrac{\gamma_{ij}(t,s)}{\sqrt{\gamma_{ii}(t,t)\gamma_{jj}(s,s)}}$.

**Napomena 2.2.**

- *Uočimo, za i=j u definiciji matrice kovarijanci je* $\gamma_{ii}$ ***funkcija autokovarijanci (ACVF) procesa*** $\{X_{i,t}, t \in \mathbb{Z}\}$*. Za* $i \neq j$ *je* $\gamma_{ij}$ ***funkcija kroskovarijanci (CCVF) procesa*** $\{X_{i,t}\}$ *i* $\{X_{j,t}\}$.

- *Možemo pisati* $\Gamma(t,s) = E[(X_t - \mu(t))(X_s - \mu(s))^T]$*, pri čemu očekivanje matrice uzimamo po elementima.*

- *Uočimo, za i=j u definiciji matrice korelacija je* $\rho_{ii}$ ***autokorelacijska funkcija (ACRF) procesa*** $\{X_{i,t}, t \in \mathbb{Z}\}$*. Za* $i \neq j$ *je* $\rho_{ij}$ ***kroskorelacijska funkcija (CCRF) procesa*** $\{X_{i,t}\}$ *i* $\{X_{j,t}\}$.

- *Matrica kovarijanci ne mora biti simetrična. Vrijedi:* $\gamma_{ij}(t,s) = Cov(X_{i,t}, X_{j,s}) = Cov(X_{j,s}, X_{i,t}) = \gamma_{ji}(s,t)$*. Posljedično isto vrijedi i za matricu korelacija.*

**Primjer 2.1.** (Matrica kovarijanci ne mora biti simetrična) *Neka je* $\{(X_t, Y_t)\}$ *dvodimenzionalni vremenski niz takav da je* $\{X_t\} \sim \mathcal{WN}(0, \sigma^2)$ *i* $Y_t = X_t + 0.8X_{t-1}$, *odnosno* $\{Y_t\} \sim MA(1)$ *proces. Pogledajmo matricu kovarijanci:*

$$\Gamma(t,s) = \begin{pmatrix} \gamma_{X,X}(t,s) & \gamma_{X,Y}(t,s) \\ \gamma_{Y,X}(t,s) & \gamma_{Y,Y}(t,s). \end{pmatrix}$$

*Kako bismo pokazali da nije simetrična, ključni su elementi na sporednoj dijagonali, odnosno* $\gamma_{X,Y}(t,s)$ *i* $\gamma_{Y,X}(t,s)$.

$$\begin{aligned}
\gamma_{X,Y}(t,s) &= Cov(X_t, Y_s) \\
&= E[X_t Y_s] - E[X_t]E[Y_s] \\
&= E[X_t Y_s] \\
&= E[X_t(X_s + 0.8X_{s-1})] \\
&= E[X_t X_s] + 0.8E[X_t X_{s-1}] \\
&= \begin{cases} \sigma^2, & \text{t = s} \\ 0.8\sigma^2, & \text{t = s} - 1 \\ 0, & \textit{inače} \end{cases}
\end{aligned}$$

$$\begin{aligned}
\gamma_{Y,X}(t,s) &= Cov(Y_t, X_s) \\
&= \vdots \\
&= E[X_t X_s] + 0.8E[X_{t-1} X_s] \\
&= \begin{cases} \sigma^2, & \text{t = s} \\ 0.8\sigma^2, & \text{t} - 1 = \text{s} \\ 0, & \textit{inače} \end{cases}
\end{aligned}$$

*Iz nejednakosti* $\gamma_{X,Y}(t,s)$ *i* $\gamma_{X,Y}(t,s)$ *slijedi da matrica kovarijanci nije simetrična.*

**Definicija 2.3.** (Stacionarnost) *Višedimenzionalan proces* $\{X_t, t \in \mathbb{Z}\}$ *je* ***stacionaran*** *ako vrijedi:*

*(i)* $E[X_{i,t}^2] < \infty, \forall t \in \mathbb{Z}, \forall i \in 1, \ldots, m$

*(ii)* $\mu(t) = \mu \in \mathbb{R}^m, \forall t \in \mathbb{Z}$

*(iii)* $\Gamma(t, t+h)$ *ne ovisi o* $t$, $\forall h \in \mathbb{Z}$.

Budući da kod stacionarnih procesa $\Gamma(t, t+h)$ ne ovisi o $t$, a posljedično ni $R(t, t+h)$, ima smisla uvesti oznaku $\Gamma(h) := \Gamma(t, t+h)$ i $R(h) := R(t, t+h)$.

# 2. Vektorski autoregresivni modeli

U Poglavlju 1. smo spomenuli AR procese kao procese kod kojih trenutnu vrijednost modeliramo pomoću prošlih vrijednosti procesa i slučajne greške. Također, spomenuli smo da kod višedimenzionalnih vremenskih nizova u fiksnom trenutku $t = 1, \ldots, n$ imamo vektor podataka $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{m,t})$. Cilj vektorskih autoregresivnih (VAR) modela je trenutnu vrijednost tog vektora modelirati preko prošlih vrijednosti tog vektora i greške, odnosno, svaku komponentu tog višedimenzionalnog vektora modelirati preko prošlih vrijednosti te, ali i ostalih komponenti vektora i greške.
VAR (*engl. vector autoregressive*) modeli su, dakle, dinamički modeli skupine vremenskih nizova, a rezultati dobiveni VAR modelom su osnova za razumijevanje i istraživanje međusobne ovisnosti skupa varijabli. Također se koriste u analizi Grangerove uzročnosti (*engl. Granger causality*) i inovacijskoj analizi (*engl. innovation analysis*). Za potrebe modela koji je rađen u okviru ovog diplomskog rada, definirat ćemo samo pojam Grangerove uzročnosti koji je korišten u izradi modela, a više o inovacijskoj analizi se može vidjeti u [4].
Opći oblik vektorskog autoregresivnog modela definiranog na temelju $m$ varijabli s duljinom koraka $p$ je:

$$X_t = \nu + A_1 X_{t-1} + \cdots + A_p X_{t-p} + Z_t, t \in \mathbb{Z}, \tag{2.1}$$

gdje je $X_t = (X_{1,t}, \ldots, X_{m,t})^T$ $m$-dimenzionalan slučajni vektor, $\nu = (\nu_1, \nu_2, \ldots, \nu_m)^T$ je vektor konstantnih članova, $A_i$ su kvadratne $(m \times m)$ matrice koeficijenata i $Z_t = (Z_{1,t}, \ldots, Z_{m,t})^T$ je $m$-dimenzionalan proces bijelog šuma, odnosno proces za kojeg vrijedi: $E[Z_t] = 0, E[Z_t Z_t^T] = \Sigma_Z$ i $E[Z_t Z_s^T] = 0$ (nul-matrica) $\forall s \neq t$. Za matricu kovarijanci $\Sigma_Z$ se smatra da nije singularna, osim ako nije drugačije naglašeno. Uz ovakav korak $p$, proces (2.1) se naziva **vektorski autoregresivni model reda $p$**. Oznaka: $\{X_t\} \sim VAR(p)$.

## 2.1 Stabilnost

Razmotrimo sada VAR(1) model koji starta u trenutku $t$=1.

$$\begin{aligned} X_1 &= \nu + A_1 X_0 + Z_1 \\ X_2 &= \nu + A_1 X_1 + Z_2 \\ &= \nu + A_1(\nu + A_1 X_0 + Z_1) + Z_2 \\ &= (I_m + A_1)\nu + A_1^2 X_0 + A_1 Z_1 + Z_2 \\ &\vdots \\ X_t &= (I_m + A_1 + \cdots + A_1^{t-1})\nu + A_1^t X_0 + \sum_{i=0}^{t-1} A_1^i Z_{t-i} \\ &\vdots \end{aligned}$$

Dakle, slučajni vektori $X_1, \ldots, X_t$ su jednoznačno određeni s $X_0, Z_1, \ldots, Z_t$.
Ukoliko u zadnjem retku uvedemo oznaku $j = t - 1$, dobivamo:

$$X_t = (I_m + A_1 + \cdots + A_1^j)\nu + A_1^{j+1} X_{t-j-1} + \sum_{i=0}^{j} A_1^i Z_{t-i}.$$

Ako su sve svojstvene vrijednosti matrice $A_1$ modula manjeg od jedan, onda je niz $A_1^i, i = 0, 1, \ldots$ apsolutno sumabilan (vidi [4], Dodatak A, Poglavlje A.9.1). Dakle, prebrojiva suma $\sum_{i=0}^{\infty} A_1^i Z_{t-i}$ postoji u srednje kvadratnom smislu (vidi [4], Dodatak C, Propozicija C.9). Štoviše, $(I_m + A_1 + \cdots + A_1^j)\nu \xrightarrow{j\to\infty} (I_m - A_1)^{-1}\nu$ (vidi [4], Dodatak A, Poglavlje A.9.1). Nadalje, $A_1^{j+1}$ kovergira k nuli kad $j \to \infty$, tako da je u limesu izraz $A_1^{j+1} X_{t-j-1}$ jednak nuli.
Stoga, ako su sve svojstvene vrijednosti matrice $A_1$ modula manjeg od 1, smatramo da je dobro definiran slučajni proces

$$X_t = \mu + \sum_{i=0}^{\infty} A_1^i Z_{t-i}, \forall t \in \mathbb{Z},$$

gdje je $\mu := (I_m - A_1)^{-1}\nu$. Kao što smo rekli, distribucija vektora $X_t$ je jedinstveno određena distribucijom procesa $\{Z_t\}$ i vrijedi:

$$E[X_t] = \mu, \forall t \in \mathbb{Z}$$

i

$$\begin{aligned}\Gamma_X(h) &:= E[(X_t-\mu)(X_{t-h}-\mu)^T] \\ &= \lim_{n\to\infty}\sum_{i=0}^{n}\sum_{j=0}^{n} A_1^i E[Z_{t-i}Z_{t-h-j}^T](A_1^j)^T \\ &= \lim_{n\to\infty}\sum_{i=0}^{n} A_1^{h+i}\Sigma_Z(A_1^i)^T \\ &= \sum_{i=0}^{\infty} A_1^{h+i}\Sigma_Z(A_1^i)^T,\end{aligned}$$

zbog definicije matrice kovarijanci bijelog šuma $\Sigma_Z$.
Kažemo da je VAR(1) proces **stabilan** ako su sve svojstvene vrijednosti matrice $A_1$ modula manjeg od jedan. Taj uvjet je ekvivalentan uvjetu $det(I_m - A_1 z) \neq 0, \forall |z| \leq 1$ (vidi [4], Dodatak A.6, Pravilo 7). Bitno je za naglasiti da se proces (2.1) može definirati i ako uvjet stabilnosti nije zadovoljen, ali mi ćemo se baviti stabilnim procesima u ostatku rada.
Prethodno razmatranje se može lako proširiti na VAR($p$) procese, $p > 1$, budući da se svaki VAR($p$) proces može zapisati u formi VAR(1) procesa. Konkretno, ako imamo VAR($p$) proces kao u (2.1), onda je odgovarajući $mp$-dimenzionalan VAR(1) proces dan s:

$$X_t = \nu + \mathbf{A}X_{t-1} + Z_t, \tag{2.2}$$

gdje je:

$$\begin{aligned}X_t &:= (X_t, X_{t-1}, \ldots, X_{t-p+1})^T, (mp\times 1), \\ \nu &:= (\nu, 0, \ldots, 0)^T, (mp\times 1), \\ \mathbf{A} &:= \begin{pmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_m & 0 & \ldots & 0 & 0 \\ 0 & I_m & & 0 & 0 \\ \vdots & & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & I_m & 0 \end{pmatrix}, (mp\times mp), \\ Z_t &:= (Z_t, 0, \ldots, 0)^T, (mp\times 1)\end{aligned}$$

s dimenzijama navedenim u zagradi. Analogno kao u slučaju VAR(1) procesa, slijedi da je $\{X_t\}$ stabilan ako je

$$det(I_{mp} - Az) = det(I_{mp} - A_1 z - \cdots - A_p z^p) \neq 0, \forall |z| \leq 1.$$

## 2.2 Stacionarnost

### 2.2.1 Stabilnost i stacionarnost

Zapišimo proces $\{X_t\}$ u MA reprezentaciji koristeći operator pomaka unazad:

$$X_t = \nu + (A_1 B + \dots A_p B^p) X_t + Z_t,$$

odnosno uz oznaku $A(B) := I_m - A_1 B - \dots - A_p B^p$,

$$A(B) X_t = \nu + Z_t.$$

Neka je $\Phi(B) := \sum_{i=0}^{\infty} \Phi_i B^i$ operator takav da vrijedi $\Phi(B)A(B) = I_m$. Odnosno, operator $\Phi(B)$ je inverz operatora A(B) pa se nekada i označava s $A(B)^{-1}$. Općenito, kažemo da je operator A(B) invertibilan ako je $det(A(z)) \neq 0, \forall |z| \leq 1$. Ako to vrijedi, koeficijenti matrice $\Phi(B) = A(B)^{-1}$ su apsolutno sumabilni te je proces $\Phi(B)Z_t$ dobro definiran što će nam poslužiti zbog sljedećeg zapisa:

$$\begin{aligned} X_t &= \Phi(B)\nu + \Phi(B)Z_t \\ &= \sum_{i=0}^{\infty} \Phi_i \nu + \sum_{i=0}^{\infty} \Phi_i Z_{t-i}. \end{aligned}$$

Sada je matrica kovarijanci procesa $X_t$ dana s:

$$\begin{aligned} \Gamma_X(h) &= E\left[ (X_t - \nu)\,(X_{t-h} - \nu)^T \right] \\ &= E\left[ \left( \sum_{i=0}^{h-1} \Phi_i Z_{t-i} + \sum_{i=0}^{\infty} \Phi_{h+i} Z_{t-h-i} \right) \left( \sum_{i=0}^{\infty} \Phi_i Z_{t-h-i} \right)^T \right] \\ &= \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_Z \Phi_i^T. \end{aligned}$$

Ovaj rezultat je vrlo važan jer temeljem njega i Propozicije 1.3. slijedi sljedeća propozicija:

**Propozicija 2.1.** *Stabilan VAR(p) proces $\{X_t, t \in \mathbb{Z}\}$ je stacionaran.*

Dakle, stabilnost implicira stacionarnost te je to razlog zašto se u literaturi uvjet stabilnosti često referira kao i uvjet stacionarnosti. Pogledajmo funkcije autokovarijanci i autokorelacija stabilnog, pa time i stacionarnog VAR($p$) procesa.

### 2.2.2 Funkcija autokovarijanci stabilnog VAR($p$) procesa

Pretpostavimo da je $\{X_t\}$ stabilan i stacionaran VAR($p$) proces:

$$X_t = \nu + A_1 X_{t-1} + \cdots + A_p X_{t-p} + Z_t.$$

Ukoliko proces zapišemo u tzv. 'mean-adjusted' formi, dobivamo:

$$X_t - \mu = A_1(X_{t-1} - \mu) + \cdots + A_p(X_{t-p} - \mu) + Z_t,$$

gdje je $\mu = E[X_t]$. Pomnožimo li sve s $(X_t - \mu)^T$ i uzmemo li očekivanje, korištenjem svojstva $\Gamma_X(-h) = \Gamma_X(h)^T$, za $h$=0 dobivamo:

$$\begin{aligned} \Gamma_X(0) &= A_1\Gamma_X(-1) + \cdots + A_p\Gamma_X(-p) + \Sigma_Z \\ &= A_1\Gamma_X(1)^T + \cdots + A_p\Gamma_X(p)^T + \Sigma_Z, \end{aligned}$$

dok za $h > 0$ dobivamo: $\Gamma_X(h) = A_1\Gamma_{\mathbf{X}}(h-1) + \cdots + A_p\Gamma_{\mathbf{X}}(h-p)$. Ove jednadžbe se obično zovu **Yule-Walkerove jednadžbe**.

### 2.2.3 Autokorelacijska funkcija stabilnog VAR($p$) procesa

Uz iste oznake kao kod računanja funkcije autokovarijanci stabilnog VAR($p$) procesa, autokorelacijske funkcije stabilnog VAR($p$) procesa možemo dobiti na sljedeći način: $R_X(h) = D^{-1}\Gamma_X(h)D^{-1}$, gdje je $D$ dijagonalna matrica sa standardnim devijacijama komponenata od $X_t$ na glavnoj dijagonali. Odnosno, dijagonalni elementi matrice $D$ su zapravo korijeni dijagonalnih elemenata matrice $\Gamma_X(0)$.
Uz oznaku $\gamma_{ij}(h) := Cov(X_{i,t}, X_{j,t-h})$ (odnosno $\gamma_{ij}(h)$ je $ij$-ti element od $\Gamma_X(h)$), dijagonalni elementi $\gamma_{11}(0), \ldots, \gamma_{mm}(0)$ matrice $\Gamma_X(0)$ su varijance od $X_{1,t}, \ldots, X_{m,t}$. Dakle,

$$D^{-1} = \begin{pmatrix} 1/\sqrt{\gamma_{11}(0)} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\gamma_{mm}(0)} \end{pmatrix},$$

i koeficijent korelacije između $X_{i,t} i X_{j,t-h}$ je dan s: $\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$.

### 2.2.4 Procjena parametara metodom najmanjih kvadrata

Pretpostavimo da je $\{X_t, t \in \mathbb{Z}\}$ stabilan, pa time i stacionaran, VAR($p$) proces uz oznake kao u (2.1). No, u prethodnim razmatranjima smo pretpostavljali da su koeficijenti $\nu, A_1, \ldots, A_p$ i matrica kovarijanci $\Sigma_Z$ poznati, no zapravo se u praksi oni procjenjuju na osnovi podataka vremenskog niza.

**Napomena 2.3.** *Notacija se u daljnjim razmatranjima ne razlikuje u slučaju slučajnog procesa i vremenskog niza kao realizacije slučajnog procesa. No, iz konteksta će biti jasno značenje.*

Procjena spomenutih parametara se može provoditi na više načina - npr., procjena metodom najmanjih kvadrata, procjena metodom najmanjih kvadrata sa 'mean-adjusted' podacima, Yule-Walker procjena i *maximum likelihood* procjena. No, za potrebe ovog diplomskog rada navodimo samo metodu koja je korištena i u modelu, a to je metoda najmanjih kvadrata. Diskutirat ćemo kako izgleda taj procjenitelj te spomenuti neka njegova svojstva. Detaljnije o preostalim metodama se može vidjeti u [4], Poglavlje 3.

### 2.2.5 Procjena matrice koeficijenata

Pretpostavimo kao i dosad da raspolažemo s $m$ vremenskih nizova duljine $T$, ali i da za svaki vremenski niz imamo poznatih $p$ prethodnih vrijednosti. Uvedimo sljedeće oznake:

$$\begin{aligned} X &:= (X_1, \ldots, X_T), (m \times T) \\ B &:= (\nu, A_1, \ldots, A_p), (m \times (mp+1)) \\ Y_t &:= (1, X_t, \ldots, X_{t-p+1})^T, ((mp+1) \times 1) \\ Y &:= (Y_0, \ldots, Y_{T-1}), ((mp+1) \times T) \\ Z &:= (Z_1, \ldots, Z_T), (m \times T), \end{aligned}$$

dimenzija navedenih u zagradi. Koristeći te oznake, VAR($p$) model (2.1) možemo zapisati kao: $X = BY + Z$, odnosno kao $vec(X) = vec(BY) + vec(Z) = Y^T \otimes I_m vec(B) + vec(Z)$, gdje je $\otimes$ Kroneckerov produkt, a *vec* operator slaganja stupaca (više u [4], Dodatak A).
Procjenitelja metodom najmanjih kvadrata možemo dobiti na sljedeći način: izraz

$$X_t = BY_{t-1} + Z_t$$

pomnožimo s $Y_{t-1}^T$ i uzmemo očekivanje:

$$E[X_t Y_{t-1}^T] = BE[Y_{t-1} Y_{t-1}^T].$$

Procjenimo li $E[X_t Y_{t-1}^T]$ s

$$\frac{1}{T} \sum_{t=1}^{T} X_t Y_{t-1}^T = \frac{1}{T} X Y^T,$$

a $E[Y_{t-1}Y_{t-1}^T]$ s

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}Y_{t-1}^T = \frac{1}{T}YY^T,$$

dobivamo normalne jednadžbe

$$\frac{1}{T}XY^T = B\frac{1}{T}YY^T,$$

iz čega slijedi da je

$$\widehat{B} = XY^T(YY^T)^{-1}.$$

Spomenimo sad neke rezultate koji će nam trebati kasnije.

**Definicija 2.4.** (Standardni bijeli šum) *Bijeli šum $Z_t = (Z_{1,t}, \ldots, Z_{m,t})^T$ se naziva* ***standardni bijeli šum*** *ako su $Z_t$ neprekidni slučajni vektori takvi da vrijedi $E[Z_t] = 0, \Sigma_Z = E[Z_t Z_t^T]$ nije singularna, $Z_t$ i $Z_s$ su nezavisni za $t \neq s$ i za neku konačnu konstantu $c$ vrijedi:*

$$E[|Z_{i,t}Z_{j,t}Z_{k,t}Z_{l,t}|] \leq c, \text{ za } i,j,k,l = 1,\ldots,m \ i \ \forall t.$$

Nadalje, označimo $\Gamma = p - lim\frac{YY^T}{T}, T \to \infty$.

**Lema 2.1.** *Neka je $\{X_t\}$ stabilan, m-dimenzionalan VAR(p) proces kao u (2.1) gdje su reziduali $\{Z_t\}$ standardni bijeli šum. Tada vrijedi:*

*(i)* $\Gamma$ *postoji i nije singularna*

*(ii)* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} vec(Z_t Y_{t-1}^T) = \frac{1}{\sqrt{T}} vec(ZY^T) \xrightarrow{d} \mathcal{N}(0, \Gamma \otimes \Sigma_Z)$, *za* $T \to \infty$.

*Dokaz.* Vidi Fuller(1976, str. 340, tm. 8.2.3). □

**Propozicija 2.2.** (Asimptotska svojstva LS procjenitelja) *Neka je $\{X_t\}$ stabilan, m-dimenzionalan VAR(p) proces kao u (2.1) gdje su reziduali $\{Z_t\}$ standardni bijeli šum i $\widehat{B} = XY^T(YY^T)^{-1}$ LS procjenitelj VAR koeficijenata B uz oznake kao u 2.2.5. Onda vrijedi:*

*(i)* $\widehat{B}$ *je konzistentan procjenitelj za B*

*(ii)* $\sqrt{T}vec(\widehat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_Z)$.

*Dokaz.* (i) Zapišimo $\widehat{B}$ na sljedeći način

$$\begin{aligned}\widehat{B} &= XY^T(YY^T)^{-1} \\ &= (BY+Z)Y^T(YY^T)^{-1} \\ &= B + ZY^T(YY^T)^{-1}\end{aligned}$$

Sada vrijedi

$$p-lim(\widehat{B}-B) = p-lim(\frac{ZY^T}{T})p-lim(\frac{YY^T}{T})^{-1} = 0$$

jer Lema 2.1.(ii) povlači da je $p-lim(\frac{ZY^T}{T}) = 0$.
Iz toga slijedi konzistentnost procjenitelja $\widehat{B}$.
(ii) Označimo $\beta = vec(B^T)$. Analogno je dokazati da

$$\sqrt{T}vec(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1}\otimes\Sigma_Z).$$

Uz oznaku $z = vec(Z)$, vrijedi

$$\begin{aligned}\sqrt{T}vec(\widehat{\beta}-\beta) &= \sqrt{T}((YY^T)^{-1}Y\otimes I_m)z \\ &= \left(\left(\frac{1}{T}YY^T\right)^{-1}\otimes I_m\right)\frac{1}{\sqrt{T}}(Y\otimes I_m)z.\end{aligned}$$

Prema [4], Dodatak C, Propozicija C.2 (4), $\sqrt{T}vec(\widehat{\beta}-\beta)$ ima ista asimptotska svojstva kao

$$(p-lim\left(\frac{1}{T}YY^T\right)^{-1}\otimes I_m)\frac{1}{\sqrt{T}}(Y\otimes I_m)z = (\Gamma^{-1}\otimes I_m)\frac{1}{\sqrt{T}}(Y\otimes I_m)z.$$

Prema Lemi 2.1. je asimptotska distribucija od $\sqrt{T}vec(\widehat{\beta}-\beta)$ normalna s matricom kovarijanci

$$(\Gamma^{-1}\otimes I_m)(\Gamma\otimes\Sigma_Z)(\Gamma^{-1}\otimes I_m) = \Gamma^{-1}\otimes\Sigma_Z.$$

□

Vidimo da su nam za procjenitelja matrice koeficijenata potrebne matrice $\Gamma$ i $\Sigma_Z$. Konzistentan procjenitelj za $\Gamma$ je iz Leme 2.1. upravo $\widehat{\Gamma} = \frac{YY^T}{T}$, dok nam još preostaje naći konzistentnog procjenitelja za matricu kovarijanci bijelog šuma $\Sigma_Z$.

### 2.2.6 Procjena matrice kovarijanci bijelog šuma

Budući da je $\Sigma_Z = E[Z_t Z_t^T]$, procjenitelja možemo naći na sljedeći način:

$$\begin{aligned}
\widetilde{\Sigma}_Z &= 1/T \sum_{t=1}^{T} (\widehat{Z}_t (\widehat{Z}_t)^T) \\
&= \frac{1}{T} \widehat{Z}_t (\widehat{Z}_t)^T \\
&= \frac{1}{T} (X - \widehat{B}Y)(X - \widehat{B}Y)^T \\
&= \frac{1}{T} [(X - XY^T(YY^T)^{-1})Y][(X - XY^T(YY^T)^{-1})Y]^T \\
&= \frac{1}{T} X[I_T - Y^T(YY^T)^{-1})Y][I_T - Y^T(YY^T)^{-1})Y]X^T \\
&= \frac{1}{T} X[I_T - Y^T(YY^T)^{-1})Y]X^T.
\end{aligned}$$

Navedeni se procjenitelj korigira stupnjevima slobode i dobivamo: $\widehat{\Sigma}_Z = \frac{T}{T-mp-1} \widetilde{\Sigma}_Z$. Procjenitelji su asimptotski ekvivalentni i u uvjetima Propozicije 2.2. su konzistentni.

**Propozicija 2.3.** (Asimptotska svojstva procjenitelja matrice kovarijanci WN) *Neka je $\{X_t\}$ stabilan, m-dimenzionalan VAR(p) proces kao u (2.1) s rezidualima koji su standardni bijeli šum i neka je $\overline{B}$ procjenitelj matrice koeficijenata B takav da $\sqrt{T} vec(\overline{B} - B)$ konvergira po distribuciji. Nadalje, koristeći oznake kao u 2.2.5, pretpostavimo da je $\overline{\Sigma}_Z = (X - \overline{B}Y)(X - \overline{B}Y)^T/(T - c)$, gdje je c konstanta. Tada vrijedi:*

$$p - lim \sqrt{T}(\overline{\Sigma}_Z - ZZ^T/T) = 0.$$

*Dokaz.* Vrijedi

$$\begin{aligned}
\frac{1}{T}(X - \overline{B}Y)(X - \overline{B}Y)^T &= (B - \overline{B}) \left( \frac{YY^T}{T} \right) (B - \overline{B})^T \\
&+ (B - \overline{B}) \frac{YZ^T}{T} + \frac{ZY^T}{T} (B - \overline{B})^T + \frac{ZZ^T}{T}.
\end{aligned}$$

Pod uvjetima propozicije je $p - lim(B - \overline{B}) = 0$. Stoga, po Lemi 2.1. je

$$p - lim(B - \overline{B}) YZ^T/\sqrt{T} = 0$$

i

$$p - lim \left( (B - \overline{B}) \frac{YY^T}{T} \sqrt{T} (B - \overline{B})^{-1} \right) = 0.$$

Konačno,

$$p - lim \sqrt{T}[(X - \overline{B}Y)(X - \overline{B}Y)^T/T - ZZ^T/T] = 0.$$

Budući da $T/(T - c) \to 1$, za $T \to \infty$, slijedi tvrdnja propozicije. □

Ova propozicija pokriva oba procjenitelja $\widehat{\Sigma}_Z$ i $\widetilde{\Sigma}_Z$.

**Korolar 2.1.** *Pod uvjetima Propozicije 2.3. vrijedi:*

$$p - lim\widetilde{\Sigma}_Z = p - lim\widetilde{\Sigma}_Z = p - limZZ^T/T = \Sigma_Z.$$

*Dokaz.* Zbog Propozicije 2.3., dovoljno je pokazati da vrijedi $p - limZZ^T/T = \Sigma_Z$. Budući da je

$$E\left[\frac{1}{T}ZZ^T\right] = \frac{1}{T}\sum_{t=1}^{T} E(Z_t Z_t^T) = \Sigma_Z$$

i

$$Var\left(\frac{1}{T}vec(ZZ^T)\right) = \frac{1}{T^2}\sum_{t=1}^{T} Var[vec(Z_t Z_t^T)] \leq \frac{T}{T^2}q \to 0, T \to \infty,$$

gdje je $q$ konstanta koja je gornja granica za $Var[vec(Z_t Z_t^T)]$ i koja postoji jer su četvrti momenti od $Z_t$ omeđeni prema Definiciji 2.4. □

### 2.2.7 Procjena očekivanja

Prisjetimo se *mean-adjusted* forme modela iz 2.2.2.:

$$X_t - \mu = A_1(X_{t-1} - \mu) + \cdots + A_p(X_{t-p} - \mu) + Z_t.$$

Često očekivanje procesa $\mu$ nije poznato te ga procjenjujemo aritmetičkom sredinom uzorka, odnosno:

$$\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t.$$

Zapravo, koristeći *mean-adjusted* formu, $\overline{X}$ možemo zapisati na sljedeći način:

$$\begin{aligned}
\overline{X} &= \mu + A_1[\overline{X} + \frac{1}{T}(X_0 - X_T) - \mu] + \dots \\
&+ A_p[\overline{X} + \frac{1}{T}(X_{-p+1} + \cdots + X_0 - X_{T-p+1} - \cdots - X_T) - \mu] \\
&+ \frac{1}{T}\sum_{t=1}^{T} Z_T.
\end{aligned}$$

Vrijedi i:

$$(I_m - A_1 - \cdots - A_p)(\overline{X} - \mu) = \frac{1}{T}Y_T + \frac{1}{T}\sum_{t}^{T} Z_T, \tag{2.3}$$

gdje je $Y_T = \sum_{i=1}^{p} A_i[\sum_{j=0}^{i-1}(X_{0-j} - X_{T-j})]$.
Budući da je $\{X_t\}$ stabilan, vrijedi: $E[Y_T/\sqrt{T}] = 0$ i $Var(Y_T/\sqrt{T}) \to 0$, kad $T \to \infty$

odnosno $Y_T/\sqrt{T}$ konvergira k nuli u srednje kvadratnom smislu. Stoga, ukoliko (2.3) pomnožimo s $\sqrt{T}$ i primijenimo to svojstvo, dobivamo da $\sqrt{T}(I_m - A_1 - \cdots - A_p)(\overline{X} - \mu)$ ima ista asimptotska distribucijska svojstva kao $\frac{1}{\sqrt{T}} \sum_t Z_T$ (više u [4], Dodatak A, Propozicija C.2). Budući da po centralnom graničnom teoremu vrijedi:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \xrightarrow{d} \mathcal{N}(0, \Sigma_Z), T \to \infty$$

slijedi sljedeći rezultat:

**Propozicija 2.4.** (Asimptotska svojstva aritmetičke sredine prosjeka) *Ako je* $\{X_t\}$ *stabilan VAR(p) proces dan s (2.1) i reziduali* $\{Z_t\}$ *standardni bijeli šum, onda vrijedi:*

$$\sqrt{T}(\overline{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\overline{X}}),$$

*gdje je*

$$\Sigma_{\overline{X}} = (I_m - A_1 - \cdots - A_p)^{-1} \Sigma_Z ((I_m - A_1 - \cdots - A_p)^T)^{-1}.$$

*Specijalno,* $\overline{X} \xrightarrow{d} \mu$.

*Dokaz.* Dokaz slijedi iz zapisa koji prethode Propoziciji 2.4. i Propoziciji C.15 iz [4], Dodatak C. □

Alternativni procjenitelj očekivanja se može dobiti iz LS procjenitelja matrice koeficijenata i sljedećeg je oblika: $\widehat{\mu} = (I_m - \widehat{A}_1 - \cdots - \widehat{A}_p)^{-1} \widehat{\nu}$. Procjenitelj je također konzistentan i asimptotski normalan ([4], Propozicija C.15, Dodatak C).

## 2.3 Grangerova uzročnost

Granger je 1969. godine definirao koncept uzročnosti (kauzalnosti) koji je posljednjih godina postao prilično popularan. Ideja je da uzrok ne može doći nakon učinka (efekta). Da bismo formalizirali ideju, označimo s $\Omega_t$ skup informacija koji sadrži sve relevantne informacije u svemiru do trenutka $t$, uključujući i njega. Neka je $z_t(h|\Omega_t)$ optimalan $h$-prediktor procesa $\{z_t\}$ u odnosu na informacije u $\Omega_t$. $h$-prediktor je prediktor koji radi predviđanje $h$ koraka unaprijed, a optimalnost se promatra u odnosu na minimalnu srednje kvadratnu grešku (MSE). Odgovarajuća MSE predikcije neka bude označena sa $\Sigma_z(h|\Omega_t)$.

Kažemo da proces $\{\mathbf{x_t}\}$ **uzrokuje** $\{\mathbf{z_t}\}$ **u Grangerovom smislu** ako vrijedi:

$$\Sigma_z(h|\Omega_t) < \Sigma_z(h|\Omega_t \setminus \{x_s | s \leq t\}), \text{ za barem jedan h} = 1, 2, \ldots$$

$\Omega_t \setminus \{x_s | s \leq t\}$ je skup koji sadrži sve relevantne informacije iz svemira osim prošlih i sadašnjih vrijednosti procesa $\{x_t\}$.
Pojednostavljeno rečeno, ako $\{z_t\}$ može biti efikasnije predviđen na temelju informacija procesa $\{x_t\}$, onda kažemo da je $\{x_t\}$ Granger-kauzalan za $\{z_t\}$.
Proširimo ovu definiciju na slučaj kad je $\{z_t\}$ $M$-dimenzionalan, a $\{x_t\}$ $N$-dimenzionalan proces. Tada kažemo da $\{x_t\}$ je Granger-kauzalan za $\{z_t\}$ ako vrijedi:

$$\Sigma_z(h|\Omega_t) \leq \Sigma_z(h|\Omega_T \setminus \{x_s | s \leq t\}),$$

odnosno ako je razlika desne i lijeve matrice pozitivno semidefinitna. Budući da je i nul-matrica pozitivno semidefinitna, nužno je zahtijevati da te dvije matrice nisu jednake.
Inače, ako $\{x_t\}$ uzrokuje $\{z_t\}$ i $\{z_t\}$ uzrokuje $\{x_t\}$, kažemo da imamo povratni sustav (*engl. feedback system*).

### 2.3.1 Karakterizacija Grangerove uzročnosti VAR procesa

Podsjetimo se MA reprezentacije $m$-dimenzionalnog VAR procesa:

$$X_t = \mu + \sum_{i=0}^{\infty} (\Phi_i Z_{t-i}) = \mu + \Phi(B) Z_t, \tag{2.4}$$

gdje je $\Phi_0 = I_m$. Pretpostavimo sada da se $\{X_t\}$ sastoji od $M$-dimenzionalnog procesa $\{z_t\}$ i ($m$-$M$)-dimenzionalnog procesa $\{x_t\}$ i da je MA reprezentacija particionirana:

$$X_t = \begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{pmatrix} \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix}. \tag{2.5}$$

Vrijedi:

$$z_t(1|\{X_s | s \leq t\}) = [I_M : 0] X_t(1) = \mu_1 + \sum_{i=1}^{\infty} \Phi_{11,i} Z_{1,t+1-i} + \sum_{i=1}^{\infty} \Phi_{12,i} Z_{2,t+1-i}, \tag{2.6}$$

te je stoga greška predikcije:

$$z_{t+1} - z_t(1|\{X_s | s \leq t\}) = Z_{1,t+1}.$$

$z_t$ također ima MA reprezentaciju kao podproces stacionarnog procesa te vrijedi:

$$z_t = \mu_1 + \sum_{i=1}^{\infty} \Phi_{11,i} Z_{1,t-i} + \sum_{i=1}^{\infty} \Phi_{12,i} Z_{2,t-i} = \mu_1 + \sum_{i=0}^{\infty} F_i v_{t-i}, \tag{2.7}$$

gdje je $F_0 = I_M$.
Sada je 1-prediktor temeljem $\{z_t\}$

$$z_t(1|\{z_s|s \leq t\}) = \mu_1 + \sum_{i=1}^{\infty} F_i v_{t+1-i}, \tag{2.8}$$

a pripadajuća greška predikcije je

$$z_{t+1} - z_t(1|\{z_s|s \leq t\}) = v_{t+1}.$$

Prediktori (2.6) i (2.8) su jednaki ako i samo ako je $v_t = u_{1,t}, \forall t$, a jednakost tih prediktora povlači da $\{z_t\}$ ima i sljedeću MA reprezentaciju:

$$z_t = \mu_1 + \sum_{i=0}^{\infty} F_i u_{q,t-i}. \tag{2.9}$$

Zbog jedinstvenosti MA reprezentacije (2.7) i (2.9) mora biti jednako, a to povlači da je $F_i = \Phi_{11,i}$ i $\Phi_{12,i} = 0, \forall i = 1, 2, \ldots$. Iz toga slijedi propozicija:

**Propozicija 2.5.** (Karakterizacija Grangerove neuzročnosti) *Neka je $\{X_t\}$ VAR(p) proces kao u (2.4) i (2.5) s kanonskim MA operatorom $\Phi(z)$. Tada vrijedi:*

$$z_t(1|\{X_s|s \leq t\} = z_t(1|\{z_s|s \leq t\}) \Leftrightarrow \Phi_{12,i} = 0, \forall i = 1, 2, \ldots.$$

Ova propozicija daje nužan i dovoljan uvjet da $x_t$ nije Granger-kauzalan za $z_t$ i to za sve procese koji imaju MA reprezentaciju. Odnosno, ako imamo proces $\{X_t\}$ kao u (2.5), proces $\{x_t\}$ nije Granger-kauzalan za $\{z_t\}$ ako i samo ako su sve matrice koeficijenata uz $\{x_t\}$ u jednadžbi za $\{z_t\}$ jednake nul-matrici.
U pogledu stacionarnih, stabilnih VAR($p$) modela imamo sljedeće:

$$X_t = \begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} + \begin{pmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} A_{11,p} & A_{12,p} \\ A_{21,p} & A_{22,p} \end{pmatrix} \begin{pmatrix} z_{t-p} \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix}. \tag{2.10}$$

**Korolar 2.2.** *Ako je $X_t$ stabilan VAR(p) proces kao u (2.10) sa nesingularnom matricom kovarijanci bijelog šuma $\Sigma_Z$, onda vrijedi:*

$$z_t(h|\{X_s|s \leq t\}) = z_t(h|\{z_s|s \leq t\}), h = 1, 2, \cdots \Leftrightarrow A_{12,i} = 0, \forall i = 1, \ldots, p.$$

*Alternativno,*

$$x_t(h|\{X_s|s \leq t\}) = x_t(h|\{x_s|s \leq t\}), h = 1, 2, \cdots \Leftrightarrow A_{21,i} = 0, \forall i = 1, \ldots, p.$$

Ovaj korolar implicira da se neuzročnost može uočiti samo iz promatranja VAR reprezentacije procesa. Odnosno, ako imamo VAR proces kao u (2.10), proces $\{x_t\}$ nije Granger-kauzalan za $\{z_t\}$ ako su u jednadžbi za proces $\{z_t\}$ sve matrice koeficijenta uz proces $\{x_t\}$ jednake nul-matrici.

### 2.3.2 Waldov test Grangerove uzročnosti

Ukoliko u procijenjenom VAR($p$) modelu želimo testirati Grangerovu uzročnost, potrebno je testirati ograničenja koeficijenata za što nam je potrebna Waldova statistika. Uz oznake kao u Poglavlju 2.2.5, neka je $\beta := vec(B)$ vektor dimenzija $((m^2p+m)\times 1)$. Testiramo hipoteze $\mathcal{H}_0 : C\beta = c$ i $\mathcal{H}_1 : C\beta \neq c$, gdje je $C$ $(N \times (m^2p + m))$ matrica ranga $N$, a $c$ je $(N \times 1)$ vektor.
Pretpostavimo li da je

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_Z), T \to \infty$$

kao u prethodnom potpoglavlju, dobivamo

$$\sqrt{T}(C\widehat{\beta} - C\beta) \xrightarrow{d} \mathcal{N}(0, C(\Gamma^{-1} \otimes \Sigma_Z)C^T), T \to \infty$$

te je stoga

$$T(C\widehat{\beta} - c)^T[C(\Gamma^{-1} \otimes \Sigma_Z)C^T]^{-1}(C\widehat{\beta} - c) \xrightarrow{d} \chi^2(N), T \to \infty.$$

Ta statistika se zove **Waldova statistika**.
Zamijenimo li $\Gamma$ i $\Sigma_Z$ njihovim procjeniteljima $\widehat{\Gamma} = YY^T/T$ i $\widehat{\Sigma}_Z$ , dobivamo statistiku:

$$\lambda_W = (C\widehat{\beta} - c)^T[C((YY^T)^{-1} \otimes \widehat{\Sigma}_Z)C^T]^{-1}(C\widehat{\beta} - c)$$

koja u uvjetima istinitosti nul-hipoteze ima asimptotsku $\chi^2(N)$ distribuciju.
Moguća je sljedeća korekcija test statistike

$$\lambda_F = \lambda_W/N \approx F(N, mT - m^2p - m) \approx F(N, T - mp - 1).$$

Time smo dobili statistiku koja se može aproksimirati F-statistikom koja ima "lakše" repove od $\chi^2(N)$ distribucije.

## 2.4 Identifikacija modela

U prethodnim poglavljima smo pretpostavljali da imamo poznat red VAR modela $p$. U praksi je red obično nepoznat. Pod identifikacijom modela smatramo određivanje reda modela. Odabirom nepotrebno velikog reda $p$ smanjujemo preciznost predikcije odgovarajućeg procijenjenog VAR modela. Iz dosadašnjih razmatranja možemo zaključiti da ne postoji jedan točan red VAR modela. Primjerice, ako imamo proces:

$$X_t = \nu + A_1X_{t-1} + \cdots + A_pX_{t-p} + Z_t, t \in \mathbb{Z},$$

isto vrijedi i za proces

$$X_t = \nu + A_1 X_{t-1} + \cdots + A_p X_{t-p} + A_{p+1} X_{t-p-1} + Z_t, t \in \mathbb{Z}$$

uz uvjet $A_{p+1} = 0$. Odnosno, ako je $\{X_t\}$ VAR($p$) proces, tada je on u ovom smislu i VAR(*p+1*) proces jer nismo postavili nikakav uvjet koji isključuje jednakost matrice koeficijenata nul-matrici. Stoga je nužno da u ovom trenutku uvedemo uvjet za jedinstveni broj koji će biti red modela. U nastavku ćemo reći da je $\{X_t\}$ VAR($p$) proces ako je $A_p \neq 0$ i $A_i = 0, \forall i > p$, tako da je $p$ najmanji mogući red.

### 2.4.1 LR test statistika

Pretpostavimo da je poznata gornja granica $M$ za mogući red VAR modela. Od različitih pristupa za određivanje reda, promatrat ćemo onaj koji se temelji na *likelihood ratio* statistici. Statistički test temeljen na LR statistici se bazira na usporedbi maksimuma logaritma funkcije vjerodostojnosti (*engl. log-likelihood*) restringiranog i nerestringiranog prostora parametara na temelju kojeg je definirana nul-hipoteza. Konkretno, LR statistika je oblika: $\lambda_{LR} = 2[\ln(l(\widetilde{\delta}) - \ln(l(\widetilde{\delta}_r))]$, gdje je $\widetilde{\delta}$ nerestringirani ML procjenitelj za vektor parametara $\delta$ dobiven maksimiziranjem funkcije vjerodostojnosti nad cijelim prostorom parametara, a $\widetilde{\delta}_r$ restringirani ML procjenitelj koji je dobiven maksimiziranjem funkcije vjerodostojnosti nad prostorom parametara s ispunjenim restrikcijama. $\lambda_{LR}$ ima asimptotski $\chi^2(N)$ distribuciju, gdje je $N$ broj linearnih restrikcija.

Najprije testiramo hipotezu $\mathcal{H}_0 : A_M = 0$. Ako se nul-hipoteza ne odbaci, testiramo dalje hipotezu $\mathcal{H}_0 : A_{M-1} = 0$ i tako sve dok ne dođemo do situacije da odbacimo nul-hipotezu. Pogledajmo shemu:

$$\begin{aligned} \mathcal{H}_0^1 : A_M = 0 \quad &; \quad \mathcal{H}_1^1 : A_M \neq 0 \\ \mathcal{H}_0^2 : A_{M-1} = 0 \quad &; \quad \mathcal{H}_1^2 : A - M - 1 \neq 0 | A_M = 0 \\ \vdots \quad &; \quad \vdots \\ \mathcal{H}_0^M : A_1 = 0 \quad &; \quad \mathcal{H}_1^M : A_1 \neq 0 | A_M = \cdots = A_2 = 0. \end{aligned}$$

Dakle, svaka nova nul-hipoteza se testira uvjetno na to da prethodna nije odbačena. LR statistika za testiranje $i$-te nul-hipoteze je dana s:

$$\lambda_{LR}(i) = T[\ln|\widetilde{\Sigma}_Z(M-i)| - \ln|\widetilde{\Sigma}_Z(M-i+1)|],$$

gdje $\widetilde{\Sigma}_Z$ označava ML procjenitelja $\Sigma_Z(k)$ kad imamo VAR($k$) model na uzorku veličine $T$ i u tom slučaju vrijedi $\lambda_{LR}(i) \sim \chi^2(m^2)$, budući da je u svakoj $\mathcal{H}_0^i$ $m^2$ parametara

postavljeno na nulu.

Bitno je za napomenuti da je razina značajnosti pojedinog testa različita od greške tipa I cijele procedure budući da odbacivanje $\mathcal{H}_0^i$ znači da su odbačene i $\mathcal{H}_0^{i+1}, \ldots, \mathcal{H}_0^M$. Ukoliko s $D_j$ označimo događaj da je $\mathcal{H}_0^j$ odbačena u $j$-tom testu kad je zapravo istinita, vjerojatnost greške tipa I za $i$-ti test u nizu je $\varepsilon_i = P(D_1 \cup D_2 \cup \cdots \cup D_i)$, dok je $\gamma_j = P(D_j)$ razina značajnosti $j$-tog testa. Nadalje, može se pokazati da su za $l \neq j, l, j \leq i$ $\lambda_{LR}(l)$ i $\lambda_{LR}(j)$ asimptotski nezavisne statistike ako su $\mathcal{H}_0^1, \ldots, \mathcal{H}_0^i$ istinite. Stoga su $D_l$ i $D_j$ nezavisni događaji za velike uzorke te vrijedi:

$$\begin{aligned} \varepsilon_i &= P(D_1 \cup D_2 \cup \cdots \cup D_{i-1}) + P(D_i) - P((D_1 \cup D_2 \cup \cdots \cup D_{i-1}) \cap D_i) \\ &= \varepsilon_{i-1} + \gamma_i - \varepsilon_{i-1}\gamma_i \\ &= \varepsilon_{i-1} + \gamma_i(1 - \varepsilon_{i-1}), \end{aligned}$$

$i = 2, 3 \ldots, M$.
Za prvi test vrijedi $\gamma_i = \varepsilon_i$. Induktivno, vrijedi:

$$\varepsilon_i = 1 - (1 - \gamma_1) \ldots (1 - \gamma_i), i = 1, 2, \ldots, M.$$

### 2.4.2 Kriteriji za određivanje reda modela

Iako je izvođenje statističkih testova dobra strategija za detekciju parametara različitih od nule, takav pristup za specifične svrhe nije potpuno zadovoljavajuć. Primjerice, ako se VAR model koristi za predikciju, onda nismo posebno zainteresirani za pronalazak točnog reda modela. Više nas zanima model koji će biti dobar u smislu predikcije. Stoga ćemo u ovom poglavlju navesti kriterije koji se koriste za određivanje reda modela.
Ukoliko je cilj modela predviđanje, onda ima smisla odabrati takav red za kojeg je srednje kvadratna greška (*engl. mean squared error*) predikcije minimalna. Stoga Akaike (1969, 1971) predlaže da red modela odaberemo na temelju srednje kvadratne greške predikcije jedan korak unaprijed (vidljivo u [4], Poglavlje 3):

$$\Sigma_{\widehat{X}}(1) = \frac{T + mr + 1}{t}\Sigma_Z,$$

gdje je $r$ red modela procijenjen na podacima. Da bismo mogli ovo koristit, $\Sigma_Z$ zamijenimo procjenom matrice kovarijanci bijelog šuma $\widehat{\Sigma}_Z$. Budući da je jednostavnije imati skalarni kriterij nego matricu, Akaike predlaže da se koristi LS procjenitelj s korekcijom stupnjeva slobode i uzme determinanta dobivene procjene matrice kovarijanci:

$$\Sigma_Z(r) = \frac{T}{T - mr - 1}\widetilde{\Sigma}_Z(r).$$

Time dobivamo kriterij/mjeru **finalna prediktivna greška** (FPE) definiranu izrazom:

$$FPE(r) = det\left(\frac{T+mr+1}{T}\frac{T}{t-mr-1}\widetilde{\Sigma}_Z(r)\right) = \left(\frac{T+mr+1}{T-mr-1}\right)^m det\left(\widetilde{\Sigma}_Z(r)\right).$$

Na temelju ovog kriterija procjenu reda $\widehat{p}$ dobivamo kao arg $\min_{r=0,1,\ldots,M} FPE(r)$.

Akaike (1973, 1974) izvodi novi kriterij poznat kao **Akaike informacijski kriterij** (AIC) definiran s:

$$AIC(r) = \ln|\widetilde{\Sigma}_Z(r)| + \frac{2rm^2}{T}.$$

Na temelju ovog kriterija procjenu reda $\widehat{p}$ dobivamo kao arg $\min_{r=0,1,\ldots,M} AIC(r)$.

Može se pokazati da vrijedi sljedeća veza između ova dva kriterija:

$$\ln FPE(r) = AIC(r) + 2m/T + \mathcal{O}(T^{-2}).$$

No, ovi procjenitelji nisu konzistentni (vidi [4] Poglavlje 4, Propozicija 4.2 i Korolar 4.2.1). Stoga, spomenimo dva procjenitelja koji su u posljednje vrijeme postali dosta popularni u primjeni, a konzistentni su. To su **Hannan-Quinn kriterij** (HQ) i **Schwarzov kriterij** (SC) definirani na sljedeći način:

$$HQ(r) = \ln|\widetilde{\Sigma}_Z(r)| + \frac{2\ln(\ln T)}{T}rm^2,$$

$$SC(r) = \ln|\widetilde{\Sigma}_Z(r)| + \frac{\ln T}{T}rm^2.$$

Sljedeći rezultat uspoređuje ove kriterije.

**Propozicija 2.6.** (Usporedba AIC, HQ i SC kriterija za male uzorke) *Neka je $x_{-M+1}, \ldots, x_0, x_1, \ldots, x_T$ bilo koji m-dimenzionalan vremenski niz te neka je VAR(r), $r = 0, 1, \ldots, M$ model kojim modeliramo podatke $x_1, \ldots, x_T$. Tada vrijede sljedeće relacije:*

$$\begin{aligned} \widehat{p}(SC) \leq \widehat{p}(AIC) \quad & \forall T \geq 8, \\ \widehat{p}(SC) \leq \widehat{p}(HQ) \quad & \forall T, \\ \widehat{p}(HQ) \leq \widehat{p}(AIC) \quad & \forall T \geq 16. \end{aligned}$$

Dakle, za manje uzorke AIC i FPE imaju bolja svojstva (izabiru dobar red češće) od HQ i SC kriterija, a također i minimiziraju varijancu greške predikcije i općenito daju bolje predikcije.

## 2.5 Dijagnostika modela

Pod dijagnostikom modela smatramo kritički analizirati model i odabrati najbolji među kandidatima. Dijagnostika se temelji na analizi reziduala. Poželjno svojstvo bi bilo da se reziduali ponašaju kao bijeli šum, no sve ovisi o svrsi modela. Ako je, na primjer, predikcija glavna svrha modela, to svojstvo nije toliko bitno ukoliko model dobro predviđa. No, ako je primjerice red modela određen nekom nestatističkom metodom (recimo na osnovu neke teorije), onda bi bilo poželjno imati neke statističke alate koji će analizirati svojstva reziduala te provjeriti ispravnost te odluke. Također, budući da različite procedure i kriteriji provjeravaju različite aspekte modela, poželjno bi bilo koristiti više statističkih alata te između više potencijalnih modela odabrati bolji, a te usporedbe se upravo vrše na osnovi analize reziduala i prediktivne sposobnosti modela.

### 2.5.1 Korelacija reziduala

Pretpostavimo da je $\{Z_t\}$ $m$-dimenzionalan bijeli šum s nesingularnom matricom kovarijanci $\Sigma_Z$. $\{Z_t\}$ može predstavljati na primjer reziduale VAR($p$) procesa. Matricu kovarijanci procesa $\{Z_t\}$ procjenjujemo s:

$$\widehat{\Gamma}_Z(h) := \frac{1}{T}\sum_{t=h+1}^{T} Z_t Z_{t-h}^T.$$

U 2.2.3 smo naveli da korelacijsku matricu možemo dobiti iz matrice kovarijanci na sljedeći način: $R_Z(h) = D^{-1}\Gamma_Z(h)D^{-1}$, gdje je $D$ dijagonalna matrica sa standardnim devijacijama komponenata od $Z_t$ na glavnoj dijagonali, odnosno, dijagonalni elementi matrice $D$ su zapravo korijeni dijagonalnih elemenata matrice $\Gamma_Z(0)$. Stoga, temeljem procjenitelja matrice kovarijanci definiramo i procjenitelja matrice korelacija: $\widehat{R}_Z(h) = \widehat{D}^{-1}\widehat{\Gamma}_Z(h)\widehat{D}^{-1}$.

Testovi koji se najčešće koriste za testiranje hipoteze o nekoreliranosti reziduala su **Portmanteau test** i **Lagrange Multiplier test** poznat kao **Breusch-Godfrey test**. Portmanteau test testira hipotezu o nekoreliranosti reziduala do koraka $h$:

$$\mathcal{H}_0 : \mathbf{R}_h = (R_1, \ldots, R_h) = 0;\ \mathcal{H}_1 : \mathbf{R}_h \neq 0.$$

Test statisika:

$$Q_h := T\sum_{i=1}^{h} tr(\widehat{R}_i^T \widehat{R}_0^{-1} \widehat{R}_i \widehat{R}_0^{-1})$$

ima asimptotski $\chi^2(m^2(h-p))$ distribuciju. Ovi testovi se mogu koristiti za potrebe ispitivanja jesu li reziduali bijeli šum na način da se postavi niz hipoteza kao

u određivanju reda modela u 2.4.1 i, prema Lütkepohlu [4], prigodni su za velike korake $h$ zbog asimptotskih svojstava.
Lagrangeov test je, s druge strane, prigodan za manje korake $h$ jer je ideja da se postavi VAR model za vektor greške:

$$Z_t = D_1 Z_{t-1} + \cdots + D_h Z_{t-h} + V_t,$$

gdje je $V_t$ bijeli šum i on je jednak $Z_t$ ako su reziduali nekorelirani. Dakle, hipoteze su:

$$\mathcal{H}_0 : D_1 = \cdots = D_h = 0 \text{ i } \mathcal{H}_1 : D_j \neq 0 \text{ za barem jedan } j \in 1, \ldots, h.$$

Test statistika $\chi_{LM}(h)$ se izvodi iz pomoćnog regresijskog modela i ima asimptotsku $\chi^2(hm^2)$ distribuciju. Radi jednostavnosti, ovdje ne navodimo spomenuti regresijski model i izvod test statistike, što je dostupno u [4], Poglavlje 4.4.4.

### 2.5.2 Normalnost reziduala

**2.5.2.1 Normalnost bijelog šuma** Svojstvo normalnosti reziduala je poželjno, ali nije i nužan uvjet prihvaćanja modela. Testovi se baziraju na trećem i četvrtom centralnom momentu normalne distribucije poznatima kao koeficijent asimetričnosti (*engl. skewness*) i koeficijent spljoštenosti (*engl. kurtosis*) redom. Prisjetimo se svojstva da ako je $X$ slučajna varijabla sa standardnom normalnom distribucijom, onda vrijedi: $E[X^3] = 0$ i $E[X^4] = 3$.
Neka je $\{Z_t\}$ $m$-dimenzionalan Gaussov bijeli šum, odnosno vrijedi: $Z_t \sim \mathcal{N}(\mu_Z, \Sigma_Z)$ i neka je $P$ matrica takva da vrijedi: $PP^T = \Sigma_Z$. Neka je

$$W_t = (W_{1,t}, \ldots, W_{m,t})^T := P^{-1}(Z_t - \mu_Z) \sim \mathcal{N}(0, I_m),$$

odnosno, komponente od $Z_t$ su nezavisne standardne normalne slučajne varijable pa vrijedi:

$$E[W_{1,t}^3, \ldots, W_{m,t}^3]^T = 0, E[W_{1,t}^4, \ldots, W_{m,t}^4]^T = [3, \ldots, 3]^T = 3I_m. \tag{2.11}$$

Ovaj rezultat će se koristiti za testiranje normalnosti na temelju usporedbe trećeg i četvrtog momenta transformiranog procesa i teorijskih vrijednosti Gaussovog procesa. Pripadajući test je poznat kao **Jarque-Berra test**. Definirajmo:

$$\overline{Z} := \frac{1}{T}\sum_{t=1}^{T}(Z_t), S_Z := \frac{1}{T-1}\sum_{t}^{T}(Z_t - \overline{Z})(Z_t - \overline{Z})^T,$$

i neka je $P_s$ matrica takva da vrijedi $P_s P_s^T = S_Z$ i $p - lim(P_s - P) = 0$. Nadalje, neka je:

$$v_t := (v_{1,t}, \ldots, v_{m,t})^T = P_s^{-1}(Z_t - \overline{Z}), t = 1, \ldots, T,$$
$$b_1 := (b_{1,1}, \ldots, b_{m,1})^T, \text{ gdje je } b_{k,1} = \frac{1}{T}\sum_t (v_{k,t}^3), k = 1, \ldots, m,$$
$$b_2 := (b_{1,2}, \ldots, b_{m,2})^T, \text{ gdje je } b_{k,2} = \frac{1}{T}\sum_t (v_{m,t})^4, k = 1, \ldots, m.$$

Dakle, $b_1$ i $b_2$ su procjenitelji vektora iz (2.11). Za asimptotsku distribuciju koeficijenata asimetričnosti i spljoštenosti vrijedi:

$$\sqrt{T}\begin{pmatrix} b_1 \\ b_2 - 3I_m \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} 6I_m & 0 \\ 0 & 24I_m \end{pmatrix}\right), kad T \to \infty.$$

Odnosno, $b_1$ i $b_2$ su asimptotski nezavisni i normalno distribuirani. Prethodni rezultat implicira sljedeće:

$$\lambda_s := T b_1 b_1^T / 6 \xrightarrow{d} \chi^2(m)$$

i

$$\lambda_k := T(b_2 - 3I_m)^T (b_2 - 3I_m)/24 \xrightarrow{d} \chi^2(m).$$

$\lambda_s$ statistika se koristi za testiranje hipoteza:

$$\mathcal{H}_0 : E[W_{1,t}^3, \ldots, W_{m,t}^3]^T = 0; \mathcal{H}_1 : E[W_{1,t}^3, \ldots, W_{m,t}^3]^T \neq 0,$$

a $\lambda_k$ za testiranje:

$$\mathcal{H}_0 : E[W_{1,t}^4, \ldots, W_{m,t}^4]^T = 3I_m; \mathcal{H}_1 : E[W_{1,t}^4, \ldots, W_{m,t}^4]^T \neq 3I_m.$$

Konačno, statistika $\lambda_{s,k} := \lambda_s + \lambda_k \xrightarrow{d} \chi^2(2m)$ se može koristiti za testiranje obje hipoteze.

**2.5.2.2 Normalnost VAR procesa** Stabilan, stacionaran VAR($p$) proces dan s

$$X_t - \mu = A_1(X_{t-1} - \mu) + \cdots + A_p(X_{t-p} - \mu) + Z_t,$$

je normalno distribuiran ako i samo ako je proces bijelog šuma $\{Z_t\}$ normalno distribuiran. Stoga, normalnost procesa $\{X_t\}$ ispitujemo preko normalnosti $\{Z_t\}$. U primjeni se to ispituje zapravo preko procjene reziduala $\widehat{Z}_t$. Može se pokazati kako procjenitelji koeficijenta spljoštenosti i asimetričnosti i u ovom slučaju imaju asimptotski normalnu distribuciju, odnosno vrijedi:

$$\sqrt{T}\begin{pmatrix} \widehat{b}_1 \\ \widehat{b}_2 - 3I_m \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} 6I_m & 0 \\ 0 & 24I_m \end{pmatrix}\right).$$

(vidi [4], Poglavlje 4.5, Propozicija 4.10).
Iz toga računamo i vrijednosti test statistika:

$$\begin{aligned}\widehat{\lambda}_s &:= T\widehat{b}_1(\widehat{b}_1)^T/6 \xrightarrow{d} \chi^2(m),\\ \widehat{\lambda}_k &:= T(\widehat{b}_2 - 3I_m)^T(\widehat{b}_2 - 3I_m)/24 \xrightarrow{d} \chi^2(m),\\ \widehat{\lambda}_{s,k} &:= \widehat{\lambda}_s + \widehat{\lambda}_k \xrightarrow{d} \chi^2(2m).\end{aligned}$$

## 2.6 Predikcija

Nakon što smo kritički analizirali i usporedili modele te se odlučili za najprikladniji, možemo pomoću njega predviđati buduće vrijednosti varijabli modela. To je zapravo glavna svrha ove analize.
Pretpostavimo da u specifičnom trenutku $t$ želimo predvidjeti buduće vrijednosti varijabli $X_1, \ldots, X_m$. Neka je $\Omega_t$ skup svih dostupnih informacija u trenutku $t$. U slučaju VAR($p$) procesa, $\Omega_t$ sadrži sadašnje i prošle vrijednosti varijabli sustava, odnosno: $\Omega_t = \{X_s | s \leq t\}$, gdje je $X_s = (X_{1,s}, \ldots, X_{ms})^T$. Sjetimo se, prediktor koji radi predviđanje $h$ koraka unaprijed se naziva $h$-prediktor. Cilj je da prediktor minimizira očekivani gubitak. Funkcije gubitka mogu biti definirane na različite načine, ali u kontekstu VAR modela, najrašireniji su prediktori koji minimiziraju srednje kvadratnu grešku (MSE). Granger (1969) i Granger & Newbold (1986) su pokazali da minimalna MSE predikcije zapravo minimizira i široki rang drugih funkcija gubitaka. Osim toga, u slučaju nepristranih prediktora, MSE je varijanca greške predikcije što je korisno u određivanju intervala predikcije.

Neka je $\{X_t\}$ stabilan $m$-dimenzionalan VAR($p$) proces kao u (2.1). Tada je prediktor koji minimizira MSE za period $h$ **uvjetno očekivanje**:

$$E_t[X_{t+h}] := E[X_{t+h}|\Omega_t] = E[X_{t+h}|\{X_s|s \leq t\}]$$

i on minimizira MSE po svim komponentama od $X_t$.
Optimalnost uvjetnog očekivanja implicira i da je

$$E_t[X_{t+h}] = \nu + A_1 E_t[X_{t+h-1}] + \cdots + A_p E_t[X_{t+h-p}] \tag{2.12}$$

optimalan $h$-prediktor VAR($p$) procesa $\{X_t\}$ ako je $\{Z_t\}$ nezavisni bijeli šum, odnosno ako su $Z_t$ i $Z_s$ nezavisni za $s \neq t$ pa je stoga i $E_t[Z_{t+h}] = 0$ za $h > 0$.
Formula (2.12) se može rekurzivno računati za $h$-prediktora počevši s $h$=1:

$$\begin{aligned}E_t[X_{t+1}] &= \nu + A_1 X_t + \cdots + A_p X_{t-p+1}\\ E_t[X_{t+2}] &= \nu + A_1 E_t[X_{t+1}] + A_2 X_t + \cdots + A_p X_{t-p+2}\\ &\vdots\end{aligned}$$

te na taj način dobijemo VAR(1) proces:

$$E_t[X_{t+h}] = (I_m + A_1 + \cdots + A_1^{h-1})\nu + A_1^h X_t.$$

Uvjetno očekivanje ima sljedeća svojstva:

- nepristran je procjenitelj, odnosno vrijedi $E[X_{t+h} - E_t[X_{t+h}]] = 0$,
- ako je $\{Z_t\}$ nezavisni bijeli šum, $MSE[E_t[X_{t+h}]] = MSE[E_t[X_{t+h}|X_t, X_{t-1}, \dots]]$.

Ukoliko $\{Z_t\}$ nije nezavisni bijeli šum, $E_t[Z_{t+h}]$ ne mora biti jednako nuli. Bez tog svojstva možemo pronaći manje ambicioznog prediktora koji minimizira MSE među linearnim funkcijama od $X_t, X_{t-1}, \dots$. Razmotrimo VAR(1) proces:

$$X_t = A_1 X_{t-1} + Z_t.$$

Kao u (2.2) vrijedi:

$$X_{t+h} = A_1^h X_t + \sum_{i=0}^{h-1} A_1^i Z_{t+h-i}.$$

Stoga za prediktora

$$X_t(h) = B_0 X_t + B_1 X_{t-1} + \dots,$$

gdje su $B_i(m \times m)$ matrice koeficijenata, dobijemo grešku predikcije:

$$X_{t+h} - X_t(h) = \sum_{i=0}^{h-1} A_1^i Z_{t+h-i} + (A_1^h - B_0)X_t - \sum_{i=1}^{\infty} B_i X_{t-i}.$$

Koristeći činjenicu da su $Z_{t+j}$ i $X_{t-i}$ za $j > 0$ i $i \geq 0$, dobijemo:

$$\begin{aligned} MSE|X_t(h)| &= E\left[\left(\sum_{i=0}^{h-1} A_1^i Z_{t+h-i}\right)\left(\sum_{i=0}^{h-1} A_1^i Z_{t+h-i}\right)^T\right] \\ &+ E\left[\left[(A_1^h - B_0)X_t - \sum_{i=1}^{\infty} B_i X_{t-i}\right]\left[(A_1^h - B_0)X_t - \sum_{i=1}^{\infty} B_i X_{t-i}\right]^T\right]. \end{aligned}$$

Minimalna vrijednost MSE se postiže za $B_0 = A_1^h$ i $B_i = 0$ pa je optimalni prediktor

$$X_t(h) = A_1^h X_t = A_1 X_t(h-1)$$

i naziva se **prediktor minimalne MSE među linearnim funkcijama**. Pripadna greška predikcije je:

$$\sum_{i=0}^{h-1} A_1^i Z_{t+h-i}.$$

Budući da smo pokazali u (2.2) da se VAR($p$) proces može zapisati u obliku VAR(1) procesa, ovaj rezultat se može poopćiti i na VAR($p$) procese te vrijedi:

$$X_t(h) = \nu + A_1 X_t(h-1) + \cdots + A_p X_t(h-p).$$

Sada kada smo vidjeli kako općenito izgleda predviđanje kod VAR procesa, pogledajmo kako ono izgleda na procijenjenim modelima.

### 2.6.1 Predviđanje s procijenjenim modelima

Ukoliko u formulu za $X_t(h)$ uvrstimo procjene za koeficijente $\nu, A_1, \ldots$ i $A_p$, dobivamo procjenu za $h$-prediktora:

$$\widehat{X}_t(h) = \widehat{\nu} + \widehat{A}_1 \widehat{X}_t(h-1) + \cdots + \widehat{A}_p \widehat{X}_t(h-p),$$

gdje je je $\widehat{X}_t(j) = X_{t+j}$ za $j \leq 0$.
Greška predikcije je:

$$\begin{aligned} X_{t+h} - \widehat{X}_t(h) &= [X_{t+h} - X_t(h)] + [X_t(h) - \widehat{X}_t(h)] \\ &= \sum_{i=0}^{h-1} \Phi_i Z_{t+h-i} + [X_t(h) - \widehat{X}_t(h)], \end{aligned}$$

gdje su $\Phi_i$ matrice koeficijenata MA reprezentacije procesa $\{X_t\}$ (vidljivo u 2.2.1) Vrijedi: $E[X_{t+h} - \widehat{X}_t(h)] = 0$, dakle predviđanje je nepristrano čak i ako su koeficijenti procijenjeni.

# Poglavlje 3

# Vektorski autoregresivni model na stvarnim podacima

Sada kada smo napravili kratki uvod u teoriju koja stoji iza VAR modela, pogledajmo kako je izgledala primjena te vrste modela na stvarne podatke. Podaci koji su korišteni za modeliranje su meteorološki podaci i koncentracija peludi ambrozije u gradu Somboru, Republika Srbija za period od siječnja 2008. godine zaključno s lipnjem 2017. godine. Podaci o koncentraciji peludi su izvorno bili podaci s dvosatnim razmakom, dakle mjereno u 2 h, 4 h, 6 h, itd., odnosno 12 podataka za svaki dan. Cilj ovog modela je pokušati prognozirati dnevnu koncentraciju peludi. Budući da je prepostavka da bi osobe koje pate od alergije trebao zanimati 'najgori' slučaj, modelirali smo maksimalnu dnevnu koncentraciju.
Meteorološki podaci kojima raspolažemo su minimalna, prosječna i maksimalna temperatura, vjetar, padaline te vlažnost zraka i prikupljeni su s mjerne postaje Rimski Šančevi. Za vjetar, padaline i vlažnost zraka raspolažemo dnevnim prosjecima. Teorijski se smatra da je temperatura najznačajniji čimbenik, padaline i vlažnost su bitne zbog taloženja i ispiranja peludi, a vjetar zbog prijenosa zrakom. U ovom radu je analizom Grangerove uzročnosti donešena odluka o izboru prediktora za model.

## 1. Koraci u izgradnji modela

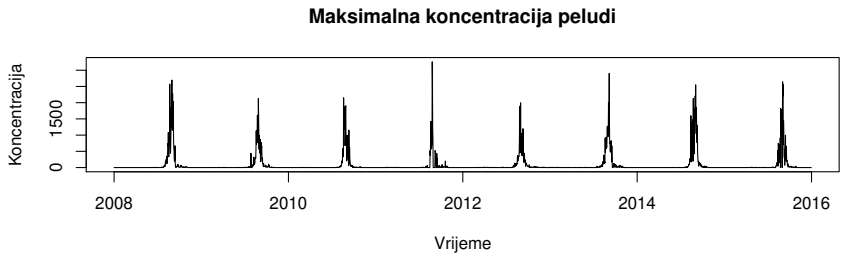
Na početku raspolažemo s 3466 dnevnih vrijednosti svake od navedenih varijabli. Od toga je odvojen dio podataka za modeliranje, a dio za validaciju dobivenog modela. Podaci od početka 2008. do kraja 2015. godine (njih 2920) su ostavljeni za modeliranje, a podaci od početka 2016. do srpnja 2017. (njih 546) su ostavljeni za validaciju. Daljnji postupak je rađen u programskom paketu R, verzija 3.4.4., a korištena je razina značajnosti 0.05 u svim statističkim testovima u nastavku.

## 1.1 Deskriptivna analiza varijabli

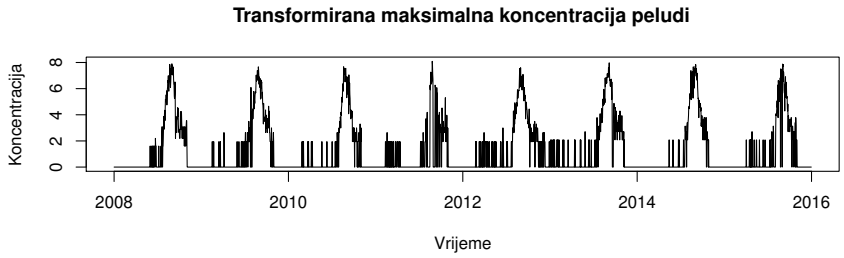
Pogledajmo najprije grafičke prikaze pojedinih varijabli, a potom ćemo analizirati svojstva svih varijabli zajedno.

### 1.1.1 Pelud

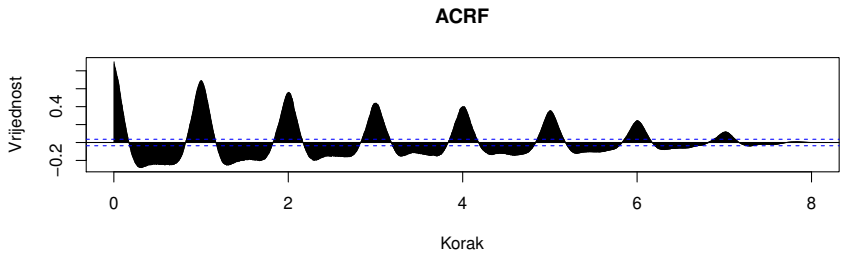
Koncentracija peludi se mjeri brojem zrnaca po kubnom metru zraka. Maksimalna koncentracija peludi izgleda vrlo nestacionarno te smo ju stoga transformirali funkcijom $x \mapsto \ln(1+x)$ i kao takvu ćemo je modelirati.




(a) Originalna koncentracija peludi




(b) Transformirana koncentracija peludi




(c) Empirijska ACRF transformirane koncentracije peludi

Slika 3.1: Trajektorije i empirijska ACRF

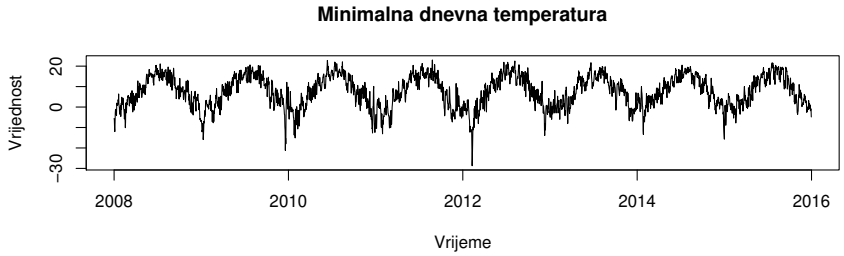
| Pelud | Min. | Donji kvartil | Prosjek | Gornji kvartil | Max. | Varijanca |
|---|---|---|---|---|---|---|
| Vrijednosti | 0 | 0 | 0 | 2.08 | 8.09 | 4.39 |

Tablica 3.1: Numeričke karakteristike transformirane koncentracije peludi

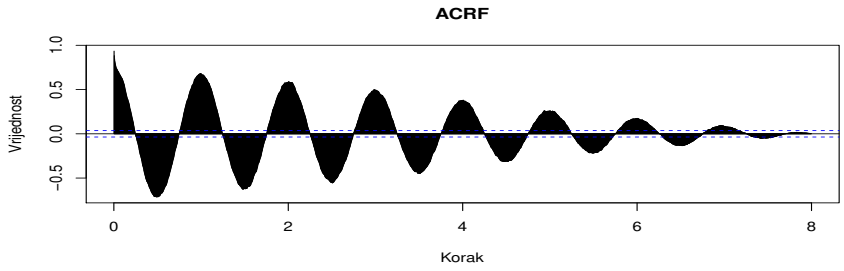
Sezona cvjetanja ambrozije je okvirno od početka lipnja do kraja listopada s maksimalnim koncentracijama obično u kolovozu i rujnu. Bitno je za napomenuti i da u danima kad nije mjerena koncentracija peludi (jer nije bila sezona) imamo nedostajuću vrijednost za koncentraciju, ali je ona iz jasnih razloga postavljena na 0.

### 1.1.2 Temperatura

U podacima je bilo 10-ak dana gdje nismo imali navedenu temperaturu te je ona procjenjena prosjekom temperature od prethodnog i budućeg dana.

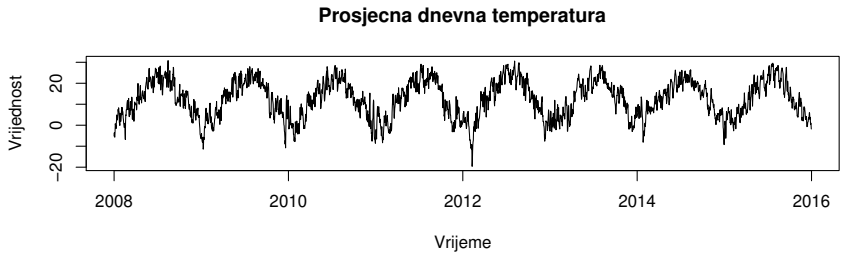



(a) Trajektorija

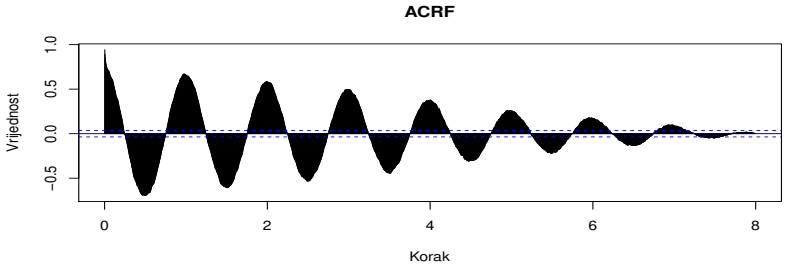



(b) Empirijska ACRF

Slika 3.2: Minimalna temperatura

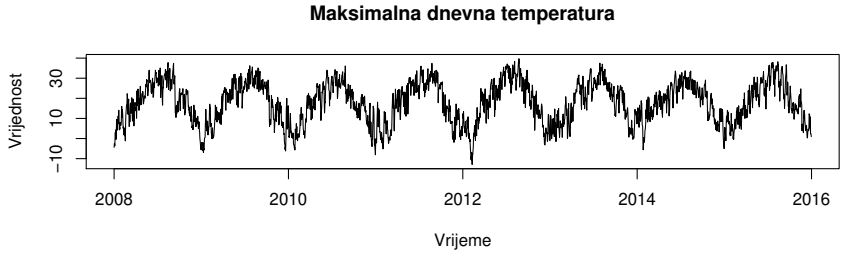



(a) Trajektorija

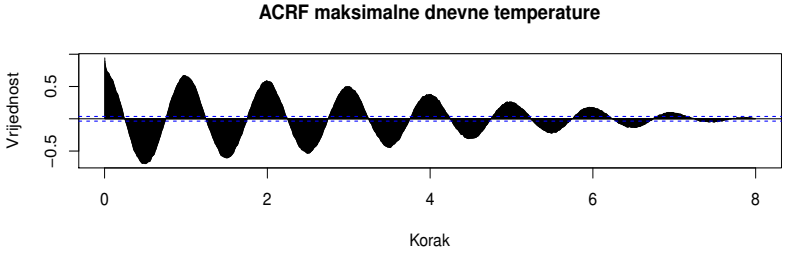



(b) Empirijska ACRF

Slika 3.3: Prosječna temperatura




(a) Trajektorija




(b) Empirijska ACRF

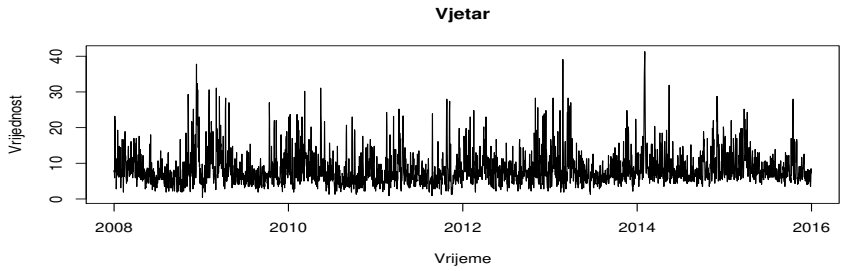
Slika 3.4: Maksimalna temperatura

| Temperatura | Min. | Donji kvartil | Prosjek | Gornji kvartil | Max. | Varijanca |
|---|---|---|---|---|---|---|
| Minimalna | -28.7 | 1.6 | 8 | 13.9 | 23 | 58.8 |
| Prosječna | -19.6 | 5.5 | 13.1 | 19.7 | 30.8 | 78.22 |
| Maksimalna | -12.9 | 9.9 | 18.8 | 26.3 | 39.7 | 105.28 |

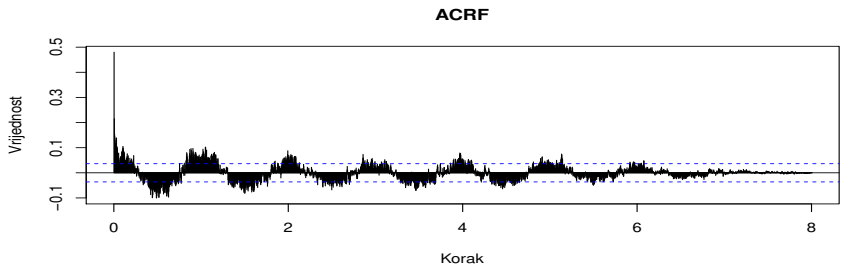
Tablica 3.2: Numeričke karakteristike temperatura

### 1.1.3 Vjetar

Brzina vjetra je mjerena u $m/s$.




(a) Trajektorija




(b) Empirijska ACRF

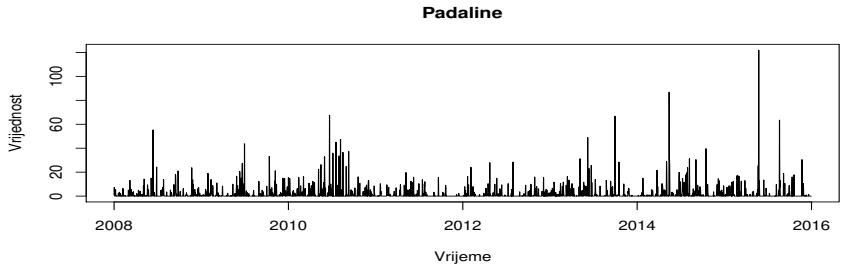
Slika 3.5: Vjetar

| Vjetar | Min. | Donji kvartil | Prosjek | Gornji kvartil | Max. | Varijanca |
|---|---|---|---|---|---|---|
| Vrijednosti | 0.4 | 5.4 | 7.2 | 10.4 | 41.3 | 23.2 |

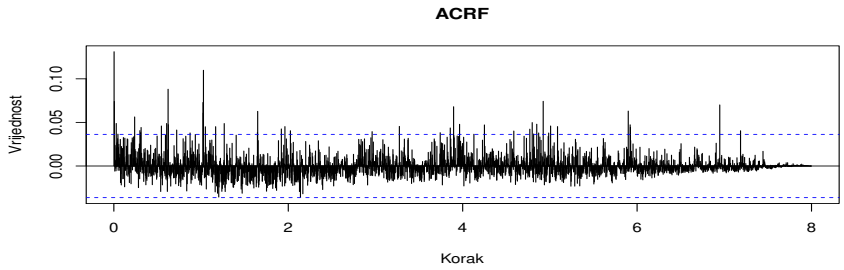
Tablica 3.3: Numeričke karakteristike vjetra

### 1.1.4 Padaline

Količina padalina se mjeri u $mm$ što predstavlja njihovu količinu u $l/m^2$.




(a) Trajektorija




(b) Empirijska ACRF

Slika 3.6: Padaline

| Padaline | Min. | Donji kvartil | Prosjek | Gornji kvartil | Max. | Varijanca |
|---|---|---|---|---|---|---|
| Vrijednosti | 0 | 0 | 0 | 0.7 | 121.9 | 32.9 |

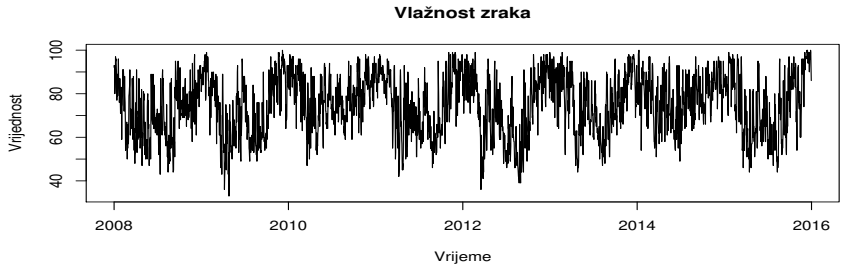
Tablica 3.4: Numeričke karakteristike padalina

### 1.1.5 Vlažnost zraka

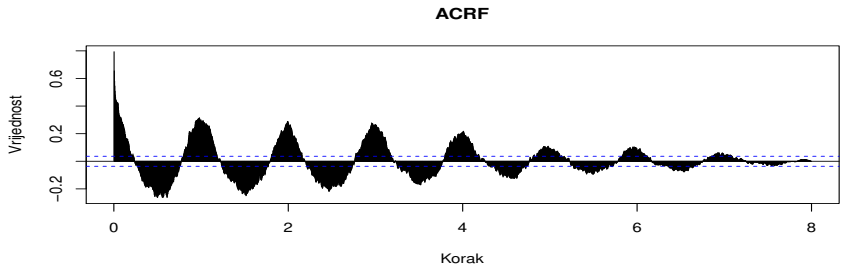
Relativna vlažnost zraka je mjerena u postotcima od 1 do 100%.

| Vlažnost | Min. | Donji kvartil | Prosjek | Gornji kvartil | Max. | Varijanca |
|---|---|---|---|---|---|---|
| Vrijednosti | 33 | 64 | 75.5 | 86 | 100 | 190.1 |

Tablica 3.5: Numeričke karakteristike vlažnosti zraka




(a) Trajektorija




(b) Empirijska ACRF

Slika 3.7: Vlažnost zraka

Uočimo da neke varijable pokazuju jasnu sezonalnost na godišnjoj razini koju bismo trebali obuhvatiti modelom. Prisutna je nestacionarnost, što je vidljivo i iz funkcija autokorelacija uzorka. Korišten je ADF test o postojanju jediničnog korijena (funkcija *adf.test()* iz paketa *'tseries'*) kako bismo donijeli odluku treba li dani vremenski niz diferencirati ili ne. Na razini značajnosti 0.05 je za svaku varijablu odbačena nul-hipoteza u korist alternativne hipoteze te nema potrebe diferencirati dane vremenske nizove. Sjetimo se, to i dalje ne znači da su nizovi stacionarni, nego samo da nemaju jedinični korijen. Dobivene p-vrijednosti ovog testa su navedene u sljedećoj tablici.

| | Min. t. | Prosj. t. | Maks. t. | Vjetar | Padaline | Vlažnost |
|---|---|---|---|---|---|---|
| p-vrijednosti | 0.01 | 0.01989 | 0.01 | 0.01 | 0.01 | 0.01 |

Tablica 3.6: p-vrijednosti dobivene testiranjem jediničnog korijena

## 1.2 Analiza uzročnosti

Pri izradi modela, jedan od najvažnijih koraka je odabir prediktora koje ćemo staviti u model. U tu svrhu, možemo se poslužiti različitim metodama koje će nešto sugerirati. Naravno, ukoliko je moguće, poželjno je isprobati više različitih modela i odlučiti se za

najprikladniji. Upravo tako je i u ovom slučaju odabran model. Naime, jedan od važnih kriterija odabira modela je jednostavnost. Stoga, nakon isprobanih više mogućnosti, odlučili smo se za model koji je bio "jednako dobar" kao ostali, ali je najjednostavniji. Sjetimo se, za vremenski niz $\{X_t\}$ kažemo da uzrokuje $\{Y_t\}$ u smislu Grangera ako se buduće vrijednosti $\{Y_t\}$ mogu predvidjeti s većom preciznošću na osnovu prethodnih vrijednosti $\{X_t\}$ nego bez njih. Iako VAR model radi predikcije svih varijabli sustava, varijabla koja je nama od interesa je transformirana maksimalna koncentracija peludi. Stoga je provedena analiza Grangerove uzročnosti u smislu samo te varijable, odnosno, provjeravali smo koja od ostalih varijabli uzrokuje nju u smislu Grangera. U tu svrhu je korištena funkcija *causality()* iz paketa *'vars'*. Nul-hipoteza jest da vremenski niz $\{X_t\}$ ne uzrokuje $\{Y_t\}$ u smislu Grangera, a alternativna da uzrokuje. Dobivene p-vrijednosti ovog testa su navedene u sljedećoj tablici.

| | Min. t. | Prosj. t. | Maks. t. | Vjetar | Padaline | Vlažnost |
|---|---|---|---|---|---|---|
| p-vrijednosti | 0.1003 | 0.03642 | 0.05851 | 0.6265 | 0.2451 | 0.05849 |

Tablica 3.7: p-vrijednosti dobivene testiranjem Grangerove uzročnosti

Kao što vidimo, pokazala se značajno jedino prosječna temperatura, dok za maksimalnu temperaturu i vlažnost zraka ne odbacujemo nul-hipotezu, ali vidimo da je $p$-vrijednost malo veća od 0.05. S obzirom da je suvišno uvrštavati obje dnevne temperature u model, odlučili smo se za jednu za koju mislimo da nosi najviše informacija (ujedno uočimo i da je p-vrijednost najmanja), a to je prosječna dnevna temperatura. Model koji je napravljen na temelju tog rezultata, a koji sadrži samo dvije varijable - transformiranu koncentraciju peludi i prosječnu temperaturu nije dovoljno dobro pratio kretanje podataka. Potom je u model dodana i vlažnost zraka, što je značajno popravilo model. Konačno, opredijelili smo se za ta tri prediktora.

## 1.3 Odabir reda modela

Nakon što smo se odabrali prediktore koje ćemo staviti u model, potrebno je odrediti red modela, odnosno odabrati broj prethodnih dana koje ćemo koristiti za predviđanje budućih vrijednosti varijabli modela. Spomenuli smo da odluku o tome možemo donijeti korištenjem LR test statistike, ali i optimiziranjem kriterija FPE, AIC, HQ i SC, što je preporučljivo ukoliko je svrha modela predviđanje. U tu svrhu je korištena funkcija *VARselect()* iz paketa *'vars'* koja sugerira red modela prema svakom od tih kriterija. Potrebno je kao argument dati i maksimalan broj koraka, što je stavljeno na 30 dana. Dobiveni rezultati su prikazani u Tablici 3.8. Odlučili smo se za korak 16.

| Kriterij | FPE | AIC | HQ | SC |
|---|---|---|---|---|
| Red | 16 | 16 | 7 | 4 |

Tablica 3.8: Sugerirani red modela prema pojedinom kriteriju

## 2. Adekvatnost modela

Sada kada smo se odlučili koje varijable uzeti i kad smo odabrali red modela, imamo sve potrebno za izgraditi model. U tu svrhu koristimo funkciju *VAR()* iz paketa *'vars'*. U pozivu ove funkcije smo uvrstili i primjećenu godišnju sezonalnost podataka tako što smo kao argument dodali *season = 365* , a to rezultira dodavanjem u model 364 *dummy* varijabli oblika:

$$sd_i(x) = \begin{cases} 1, & \text{x} = \text{i} \\ 0, & \text{x} \neq \text{i} \end{cases}, i = 1, \ldots, 364.$$

Na ovaj način dobijemo model za svaku od varijabli iz modela. Ukoliko je $s$ sezona, u model se dodaje $s-1$ *dummy* varijabli kako matrica regresora ne bila singularna, budući da imamo i vektor konstantnih članova u modelu. U sljedećim tablicama su navedeni koeficijenti uz svakog prediktora i za svaki pojedini model, ali je radi preglednosti izostavljeno 364 koeficijenata uz *dummy* varijable za svaki od modela (isti su navedeni u Dodatku).

| Korak | Prosječna temp. | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| 1 | 0.9726$^\star$ | 0.0021 | -0.0034 |
| 2 | -0.2292$^\star$ | -0.0078 | -0.0789 |
| 3 | 0.0383 | 0.0034 | -0.0203 |
| 4 | 0.0603$^\star$ | 0.0067 | 0.0062 |
| 5 | -0.0410 | -0.0041 | -0.0366 |
| 6 | 0.0568$^\star$ | 0.0077 | 0.0637 |
| 7 | -0.0247 | -0.0018 | 0.0853 |
| 8 | 0.0155 | 0.0097 | -0.0858 |
| 9 | -0.0087 | 0.0030 | -0.0089 |
| 10 | -0.0034 | -0.0010 | 0.01791 |
| 11 | 0.0050 | -0.0003 | -0.0053 |
| 12 | 0.0014 | 0.0107 | 0.0102 |
| 13 | -0.0289 | -0.0110 | -0.0245 |
| 14 | 0.0208 | 0.0021 | -0.0210 |
| 15 | 0.0009 | 0.0010 | -0.0987 |
| 16 | 0.0160 | 0.0051 | 0.1443$^\star$ |

Tablica 3.9: Procijenjeni koeficijenti uz svakog prediktora za prosječnu dnevnu temperaturu za svaki korak $i = 1, \ldots, 16$, $^\star$ - značajni na nivou značajnosti 0.05

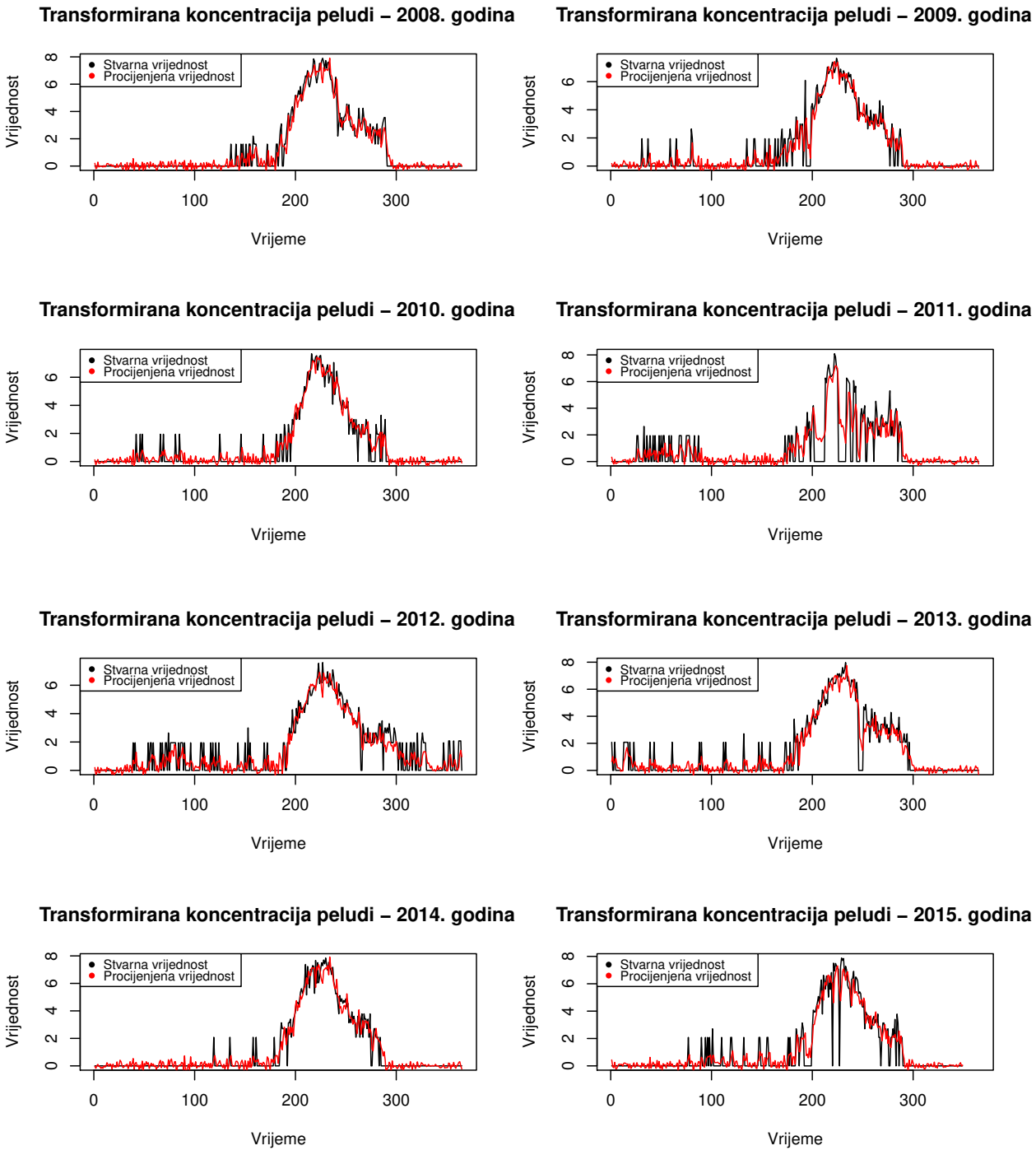
| Korak | Prosječna temp. | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| 1 | 1.673e-01$^\star$ | 6.879e-01$^\star$ | -1.366e-01 |
| 2 | 3.520e-02 | -7.616e-03 | 4.277e-01* |
| 3 | 1.102e-02 | 1.484e-02 | -1.189e-01 |
| 4 | -1.831e-01 | 3.864e-02 | -3.504e-01 |
| 5 | 1.498e-01 | 2.143e-02 | 2.184e-01 |
| 6 | -1.611e-02 | -3.714e-03 | -1.864e-01 |
| 7 | -7.045e-02 | 5.393e-03 | -2.039e-01 |
| 8 | 3.979e-02 | 1.309e-02 | 1.444e-01 |
| 9 | 5.325e-03 | -9.496e-05 | 1.0519e-01 |
| 10 | -8.812e-03 | 3.597e-02 | -8.429e-02 |
| 11 | 6.484e-02 | 2.856e-02 | 6.247e-02 |
| 12 | -2.199e-02 | -1.811e-02 | 9.673e-02 |
| 13 | 1.074e-01 | 2.598e-02 | 6.027e-01$^\star$ |
| 14 | 1.223e-01 | 6.674e-02$^\star$ | 2.723e-02 |
| 15 | -1.135e-01 | -1.885e-02 | 8.226e-02 |
| 16 | -1.389e-03 | 4.294e-02$^\star$ | -3.025e-01 |

Tablica 3.10: Procijenjeni koeficijenti uz svakog prediktora za vlažnost zraka za svaki korak $i = 1, \ldots, 16$, $^\star$ - značajni na nivou značajnosti 0.05

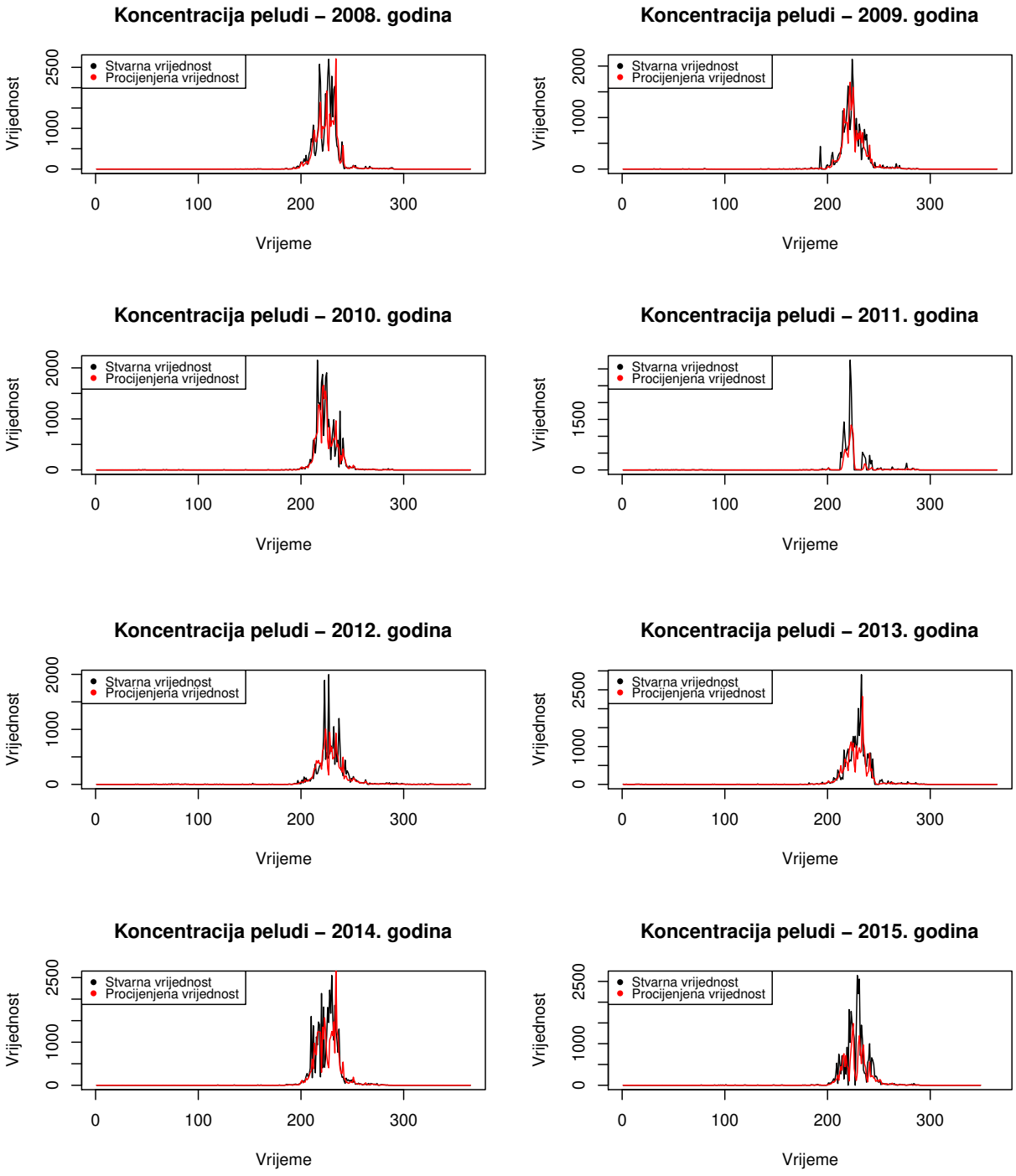
| Korak | Prosječna temp. | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| 1 | 1.018e-02 | -2.472e-04 | 3.672e-01$^\star$ |
| 2 | -1.349e-03 | -9.117e-04 | 1.419e-01$^\star$ |
| 3 | 3.375e-03 | 3.817e-03 | 6.133e-02$^\star$ |
| 4 | 3.842e-03 | -6.578e-04 | 2.866e-02 |
| 5 | -3.039e-03 | 2.220e-03 | -3.661e-06 |
| 6 | -9.050e-03 | 1.478e-03 | 6.545e-02$^\star$ |
| 7 | 1.912e-03 | 8.000e-04 | 7.622e-02$^\star$ |
| 8 | 6.304e-04 | -2.990e-03 | -3.004841e-02 |
| 9 | 5.325e-03 | 4.039e-03 | 1.980e-03 |
| 10 | -1.502e-02 | -6.374e-03$^\star$ | -2.150e-02 |
| 11 | 9.013e-03 | 1.544e-04 | -2.171e-02 |
| 12 | 7.893e-03 | 4.325e-03 | 3.335e-02 |
| 13 | -2.757e-02$^\star$ | -3.807e-03 | -3.755e-02$^\star$ |
| 14 | 2.621e-02$^\star$ | 1.422e-03 | 2.908e-02 |
| 15 | -4.955e-03 | 2.653e-03 | 3.832e-02$^\star$ |
| 16 | -5.783e-03 | -2.19e-03 | 2.867e-02 |

Tablica 3.11: Procijenjeni koeficijenti uz svakog prediktora za transformiranu koncentraciju peludi za svaki korak $i = 1, \ldots, 16$, $^\star$ - značajni na nivou značajnosti 0.05

Pogledajmo usklađenost procijenjenih i pravih vrijednosti za varijablu od interesa - transformiranu koncentraciju peludi, a time i originalnu koncentraciju peludi, ukoliko vratimo transformaciju inverznom funkcijom $x \mapsto (e^x - 1)$.




Slika 3.8: Trajektorije procijenjene i stvarne transformirane koncentracije peludi po godinama




Slika 3.9: Trajektorije procijenjene i stvarne originalne koncentracije peludi po godinama

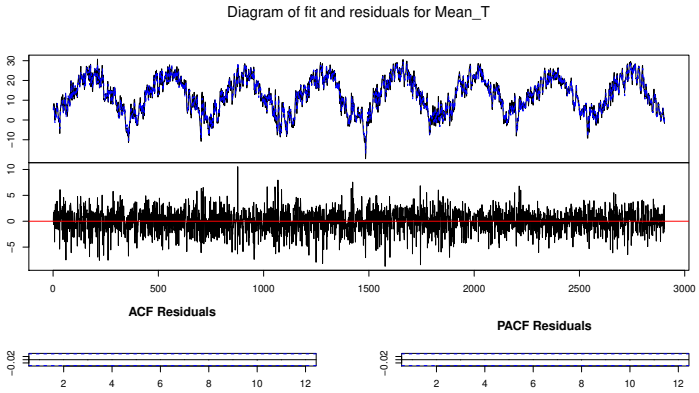
Poželjeno je da je što veći udio ukupne varijabilnosti podataka objašnjen modelom, što je u ovom slučaju u redu budući da je korigirani koeficijent determinacije $\widetilde{R}^2$ u modelu za prosječnu temperaturu 0.977, u modelu za vlažnost zraka 0.9887, a u modelu za transformiranu koncentraciju peludi 0.8969. Spomenuli smo da je poželjno da je model stabilan, da su greške nekorelirane i normalno distribuirane te da je varijanca konstanta. Potrebno je provjeriti ta svojstva. Greške su slučajne varijable, a reziduali, kao procjene greške, podaci na temelju kojih provjeravamo te pretpostavke.

## 2.1 Reziduali

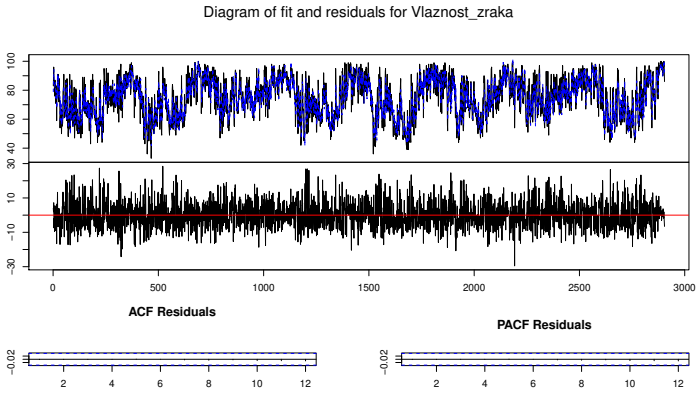
Za reziduale smo rekli da normalnost nije nužno, ali je poželjno svojstvo prihvaćanja modela. Ispitivali smo ju pomoću funkcije *normality.test()* iz paketa *'vars'* koja računa vrijednosti test statistika i pripadne p-vrijednosti multivarijatnog Jargue-Berra testa i testova asimetričnosti i spljoštenosti. Korištene su i funkcije *serial.test()* za Portmanteauov test serijske koreliranosti i funkcija *arch.test()* iz istog paketa za provjeru prisutnosti uvjetne heteroskedastičnosti koju naslućujemo iz grafičkih prikaza (drugi prikaz na svakoj od Slika 3.10). Dobiveno je sljedeće:

| Test | Skewness | Kurtosis | Jarque Berra test | Serial test | Arch test |
|---|---|---|---|---|---|
| p-vrijednost | $<$ 2.2e-16 | $<$ 2.2e-16 | $<$ 2.2e-16 | $<$ 2.2e-16 | $<$ 2.2e-16 |

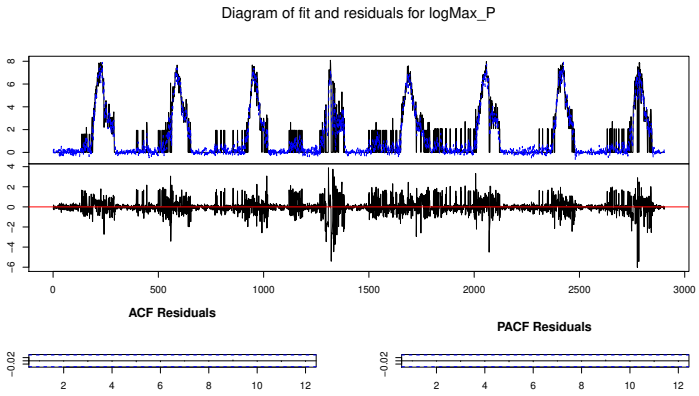
Dakle, odbacujemo sve nul-hipoteze i možemo reći da reziduali ne dolaze iz normalne distribucije te da je prisutna uvjetna heteroskedastičnost i serijska korelacija. Ta svojstva nam govore da se ne trebamo pouzdati u rezultate dobivene modelom u smislu pouzdanih intervala, interpretacije koeficijenata i slično. Ukoliko je svrha modela interpretacija, trebali bismo posegnuti za korekcijom modela kako bismo imali ispunjene pretpostavke na reziduale i kako bismo se mogli pouzdati u dobivene rezultate, no ukoliko je svrha modela predikcija, možemo pogledati koliko model dobro predviđa buduće vrijednosti ovisnih varijabli modela i ukoliko je to zadovoljavajuće, model se može koristiti u te svrhe. Na sljedećim prikazima vidimo prave i temeljem ovog modela procijenjene vrijednosti transformirane maksimalne koncentracije peludi, prosječne dnevne temperature i vlažnosti zraka te pripadajuće reziduale. Vidimo da model dosta dobro prati izmjerene (stvarne) vrijednosti svih tih varijabli.




(a) Prosječna dnevna temperatura




(b) Vlažnost zraka




(c) Transformirana koncentracija peludi

Slika 3.10: Modeli

## 2.2 Stabilnost i stacionarnost

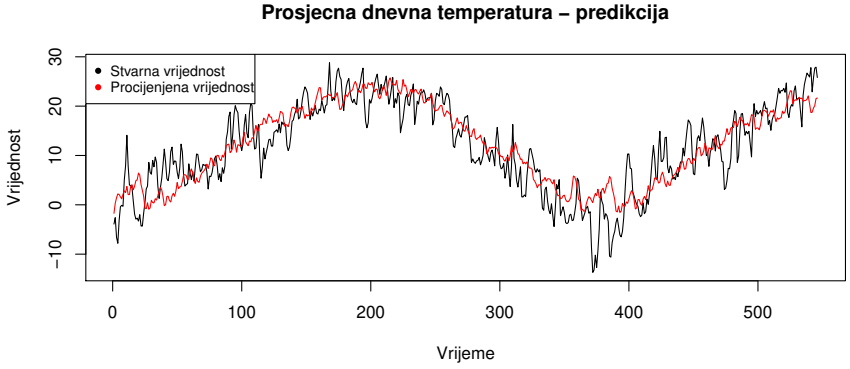
Stabilnost smo provjerili tako što smo pogledali svojstvene vrijednosti matrice koeficijenata i svi su modula manjih od jedan. Svojstvene vrijednosti matrice koeficijenata su prikazane u Tablici 3.12. Kako smo pokazali da stabilnost modela povlači i stacionarnost, ovaj model je stacionaran.

| Svojstvene vrijednosti matrice koeficijenata | | | | | |
|---|---|---|---|---|---|
| 0.9998606 | 0.9357207 | 0.8895893 | 0.8895893 | 0.8700376 | 0.8594369 |
| 0.8594369 | 0.8581790 | 0.8581790 | 0.8568425 | 0.8568425 | 0.8563004 |
| 0.8563004 | 0.8501665 | 0.8501665 | 0.8479065 | 0.8479065 | 0.8472676 |
| 0.8472676 | 0.8452040 | 0.8452040 | 0.8450405 | 0.8450405 | 0.8372544 |
| 0.8372544 | 0.8371385 | 0.8371385 | 0.8314723 | 0.8314723 | 0.8291664 |
| 0.8291664 | 0.8121786 | 0.8121786 | 0.8120374 | 0.8120374 | 0.8034546 |
| 0.8034546 | 0.8026833 | 0.8026833 | 0.8018929 | 0.8018929 | 0.7972833 |
| 0.7972833 | 0.7718810 | 0.7718810 | 0.7242792 | 0.7242792 | 0.2817423 |

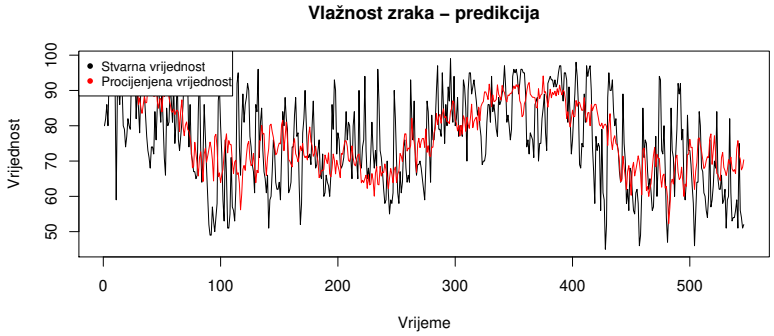
Tablica 3.12: Svojstvene vrijednosti matrice koeficijenata

# 3. Predikcija modela

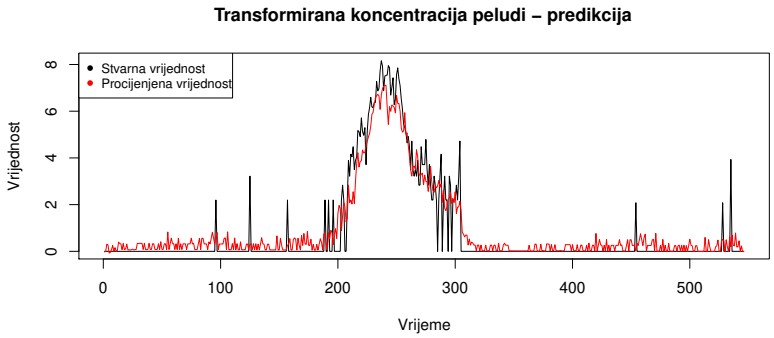
Iako naš model ne ispunjava bitne pretpostavke, pogledajmo ima li barem dobru prediktivnu moć. Koristeći funkciju *forecast()* iz istoimenog paketa smo napravili predikciju za 2016. godinu i prvu polovicu 2017. godine (ukupno 546 dana). Na sljedećim prikazima vidimo predikciju svake varijable modela.




(a) Prosječna dnevna temperatura



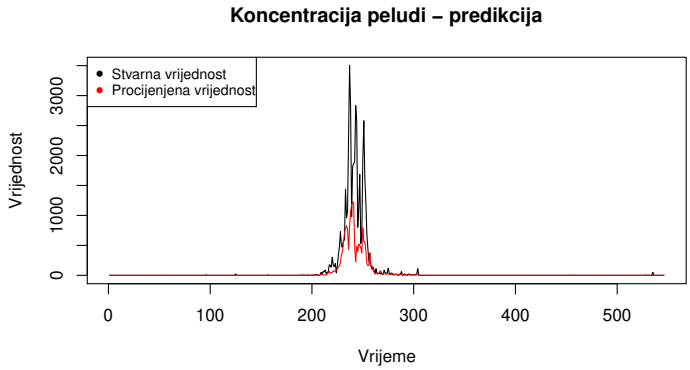

(b) Vlažnost zraka




(c) Transformirana koncentracija peludi

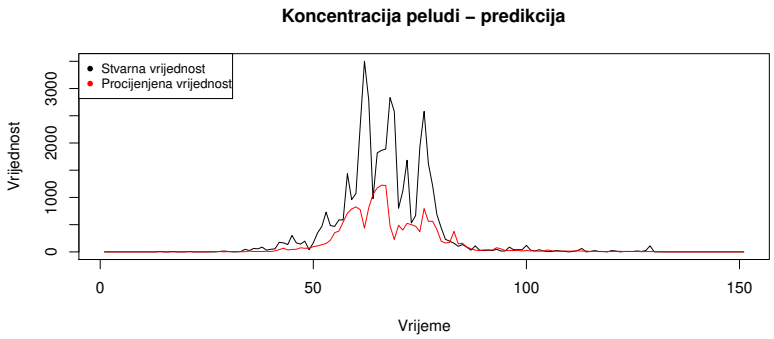
Slika 3.11: Trajektorije predikcije

Što se tiče predikcije originalne maksimalne dnevne koncentracije peludi, dobivene predviđene vrijednosti transformirane koncentracije su vraćene nazad inverznom funkcijom $x \mapsto (e^x - 1)$.




Slika 3.12: Trajektorija predikcije originalne koncentracije peludi

Istaknimo period oko sezone:




Slika 3.13: Trajektorija predikcije originalne koncentracije peludi - sezona

Vidljivo je da model ne predviđa dobro vrhunce koncentracije. No, prosječnoj osobi broj zrnaca peludi ne znači puno nego ju zanima kakva je općenito razina koncentracije peludi u zraku. Stoga je cilj modela predviđanje pripadnosti maksimalne dnevne koncentracije peludi unaprijed definiranim kategorijama. Dobivene vrijednosti su stoga podijeljene u kategorije prema sljedećoj tablici:

| Koncentracija | Niska | Umjerena | Visoka | Vrlo visoka |
|---|---|---|---|---|
| Vrijednosti | $[0, 10]$ | $(10, 50]$ | $(50, 500]$ | $\geq 500$ |

Tablica 3.13: Kategorizacija koncentracije peludi

Navedena kategorizacija je preuzeta s web stranice Nastavnog zavoda za javno zdravstvo "Dr. Andrija Štampar" ([7]), a tumačenje je sljedeće: kod niske koncentracije peludi samo vrlo osjetljive osobe na pelud mogu razviti simptome alergijske reakcije, kod umjerene koncentracije peludi 50% osjetljivih ljudi na pelud razviti će simptome alergijske reakcije, kod visoke koncentracije peludi većina osjetljivih osoba, a kod vrlo visoke koncentracije peludi gotovo sve osjetljive osobe na pelud razvit će simptome alergijske rekacije. U sljedećoj tablici je prikazano koliko je dana pojedine kategorije u stvarnim i predviđenim vrijednostima.

| Koncentracija | Niska | Umjerena | Visoka | Vrlo visoka |
|---|---|---|---|---|
| Predviđene vrijednosti | 467 | 35 | 29 | 15 |
| Stvarne vrijednosti | 463 | 31 | 27 | 25 |

Tablica 3.14: Podjela predviđenih i stvarnih koncentracija peludi u kategorije

Od 546 dana je:

- ispravno procijenjeno 502 (92.49%)
- neispravno procijenjeno 41 (7.51%), a od toga
    - precijenjeno 10 dana (1.83%) (predvidjeli smo koncentraciju više kategorije)
    - podcijenjeno 31 dan (5.68%) (predvidjeli smo koncentraciju niže kategorije).

Za osobe koje pate od alergija bitnije je biti pripravan, pa čak i ako koncentracija ne bude visoka koliko je predviđeno, tako da je onih 1.83% modelom netočno kategoriziranih dana manji "problem" nego 5.68% dana u kojima je modelom predviđena niža koncentracija nego stvarna. U čak 10 dana (1.83%) s vrlo visokom koncentracijom peludi je modelom predviđena visoka koncentracija, no u tom slučaju je vjerojatno većina osoba već i pripravna. Samo su 3 dana (0.54%) u kojima se modelom previđena kategorija koncentracije od stvarne koncentracije razlikovala za dvije kategorije, dok se u ostalim netočno kategoriziranim danima radilo o razlici za jednu kategoriju. Uzevši u obzir to i stopu ispravnih pogodaka od 92.49%, mogli bismo reći da model dobro predviđa buduće vrijednosti te da je njegova svrha kao takva ispunjena. Model je moguće i korigirati ukoliko bi se radila iterativna korekcija greške te potom nova procjena parametara modela.

# Literatura


[1] V. Bahovec, N. Erjavec, *Uvod u ekonometrijsku analizu*, Element, Zagreb, 2009.

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

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

[4] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin, 2005.

[5] Ž. Salinger, *Modeliranje stope nezaposlenosti i inflacije mjerene indeksom potrošačkih cijena u Republici Hrvatskoj*, Sveučilište J.J. Strossmayera u Osijeku, Odjel za matematiku, Diplomski rad, 2014.
URL: `http://www.mathos.unios.hr/~mdjumic/uploads/diplomski/SAL24.pdf`

[6] *Hrvatski zavod za javno zdravstvo*,
URL: `https://www.hzjz.hr/sluzba-zdravstvena-ekologija/alergije-kad-preburno-reagira-imunosni-sustav/`

[7] *Nastavni zavod za javno zdravstvo "Dr. Andrija Štampar"*,
URL: `http://www.stampar.hr/hr/alergijski-semafor-peludna-prognoza-i-peludni-kalendar`

[8] *Wikipedia*,
URL: `https://en.wikipedia.org/wiki/Christopher_A._Sims`

# Sažetak

U ovom radu su predstavljeni osnovni pojmovi, svojstva i rezultati analize vremenskih nizova. Navedeni su modeli za stacionarne i nestacionarne jednodimenzionalne vremenske nizove - procesi pomičnih prosjeka (MA), autoregresivni procesi (AR), autoregresivni (integrirani) procesi pomičnih prosjeka (ARMA i ARIMA) te sezonalni autoregresivni (integrirani) procesi pomičnih prosjeka (SARMA i SARIMA). Pokazani su i neki primjeri takvih procesa. Od modela za višedimenzionalne vremenske nizove je opisan vektorski autoregresivni proces (VAR). Navedene su metode procjene parametara i određivanja reda modela i analizirana je uzročnost u smislu Grangera. Opisana je i dijagnostika modela koja se zasniva na analizi reziduala te je opisana predikcija pomoću odabranog modela.

Na temelju podataka o koncentraciji peludi u Somboru i meteoroloških podataka s mjerne postaje Rimski Šančevi je napravljen vektorski autoregresivni model s ciljem predviđanja maksimalne dnevne koncentracije peludi. Za potrebe modela su odabrane i varijable vlažnost zraka i prosječna dnevna temperatura, a red modela je 16, što znači da za predikciju maksimalne koncentracije peludi pojedinog dana koristimo vrijednosti koncentracije peludi, vlažnosti zraka i prosječne temperature prethodnih 16 dana. Dobivene vrijednosti su podijeljene u sljedeće kategorije - niska, umjerena, visoka i vrlo visoka koncentracija. S obzirom na nezadovoljene pretpostavke o svojstvima reziduala, model se ne može koristiti u svrhu interpretacije, no za predikciju se pokazao točnim u 92.49 % promatranih dana. Navedena je i mogućnost poboljšanja korištenog modela.

**Ključne riječi**: vremenski niz, stacionarnost, MA, AR, ARMA, ARIMA, SARMA, SARIMA, VAR, alergija, pelud, ambrozija, koncentracija

# Modelling of ambrosia pollen concentration with the vector autoregressive model

## Summary

This graduate thesis presents basic terms, properties and results of time series analysis. Models of stationarity and non-stationarity univariate time series are described – moving average processes (MA), autoregressive processes (AR), autoregressive (integrated) moving average processes (ARMA and ARIMA) and seasonal autoregressive (integrated) moving average processes (SARMA and SARIMA). Some examples of these processes are also shown. Out of many multivariate time series models, the vector autoregressive process (VAR) is described. Methods of parameter estimation and order determination are shown and Granger causality is analysed. The diagnostic of models based on the residual analysis and the prediction by the chosen model are also described.

Based on the data of pollen concentration in Sombor and the meteorological data from the measuring station Rimski Šančevi, a vector autoregressive model is made with the aim of predicting the maximum daily pollen concentration. For the needs of the model, the choosen variables are also air humidity and average daily temperature and the order of the model is 16, meaning that for the prediction of the maximum pollen concentration of a single day, we use the values od pollen concentration, air humidity and average temperature od the previous 16 days. The obtained values are divided into the following categories - low, moderate, high and very high concentrations. Since the assumptions on model residuals are not satisfied, the model can not be used for the purpose of interpretation, but for prediction purposes it would be accurate for 92.49 % of observed days. Possible correction of the used model is also mentioned.

**Keywords**: time series, stationarity, MA, AR, ARMA, ARIMA, SARMA, SARIMA, VAR, allergy, pollen, ambrosia, concentration

# Životopis

Rođena sam 24. siječnja 1994. godine u Osijeku u Republici Hrvatskoj. Pohađala sam OŠ "Tenja" u Tenji (2000.-2008. godine) te potom i I. gimnaziju u Osijeku (2008.-2012. godine). Tijekom osnovnoškolskog i srednjoškolskog obrazovanja sam sudjelovala na natjecanjima iz matematike i hrvatskog jezika.
2012. godine upisujem Preddiplomski studij matematike na Odjelu za matematiku u Osijeku te 2015. godine stječem naziv *prvostupnice matematike* uz završni rad *Vigenèreova i Playfairova šifra* pod mentorstvom doc.dr.sc. Mirele Jukić Bokun. U jesen iste godine upisujem Diplomski studij matematike, smjer Financijska matematika i statistika na Odjelu za matematiku. Istaknula bih obavljanje stručne prakse u tvrtci Farmeron d.o.o. gdje sam zajedno s mentorima radila na problemima modeliranja podataka iz područja mliječne industrije.

# Dodatak

U sljedećoj tablici su prikazani procijenjeni koeficijenati uz *dummy* varijable u modelima za prosječnu dnevnu temperaturu, prosječnu dnevnu vlažnost zraka i transformiranu maksimalnu dnevnu koncentraciju peludi.

| Dummy varijabla | Prosječna temperatura | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| sd1 | 0.55303 | 0.06544 | 0.02359 |
| sd2 | 2.05683 | -0.05822 | 0.04394 |
| sd3 | 1.42657 | -0.24830 | 0.27550 |
| sd4 | 1.65356 | -2.84600 | 0.19860 |
| sd5 | 0.92578 | -1.85000 | -0.14380 |
| sd6 | 0.62096 | 4.47900 | -0.08916 |
| sd7 | 0.61358 | -1.48400 | 0.02699 |
| sd8 | 1.92830 | 1.77200 | 0.31980 |
| sd9 | 1.47284 | 1.68000 | -0.12820 |
| sd10 | 0.61775 | 5.86200 | -0.01568 |
| sd11 | 2.5433915* | -2.51300 | -0.05248 |
| sd12 | -0.46681 | 1.17000 | -0.00648 |
| sd13 | 1.29746 | -1.04500 | 0.27530 |
| sd14 | 2.5205626* | -2.21700 | 0.17720 |
| sd15 | 1.30794 | 2.26600 | 0.08166 |
| sd16 | 0.57019 | 1.58500 | -0.14330 |
| sd17 | 2.24069 | -2.01400 | 0.20420 |
| sd18 | 2.24318 | 2.27300 | -0.14340 |
| sd19 | 1.50935 | -0.28370 | -0.07359 |
| sd20 | 2.9084479* | 0.57420 | 0.19540 |
| sd21 | 1.18923 | -1.00600 | -0.12800 |
| sd22 | -0.01193 | 1.24800 | -0.11160 |
| sd23 | 0.63914 | -1.45500 | -0.05749 |
| sd24 | -0.18448 | 3.01300 | 0.01860 |
| sd25 | -1.40705 | -0.53650 | -0.00222 |
| sd26 | 0.44074 | -0.67950 | 0.07252 |
| sd27 | 1.39297 | 1.90700 | 0.05808 |
| sd28 | -0.87625 | -0.27090 | 0.07190 |
| sd29 | 1.05951 | -2.24700 | 0.26550 |
| sd30 | 1.92909 | -5.40800 | 0.22500 |
| sd31 | 0.10502 | -0.15190 | 0.13880 |
| sd32 | 0.91910 | -4.90900 | 0.10900 |
| sd33 | 1.61242 | -3.90000 | 0.17390 |
| sd34 | 0.48623 | -0.15840 | -0.14300 |
| sd35 | 1.79273 | 0.92840 | 0.19710 |
| sd36 | 1.69650 | -1.93400 | -0.10780 |
| sd37 | 1.79333 | -0.41430 | 0.00473 |
| sd38 | 0.85528 | 3.34700 | -0.06222 |
| sd39 | -0.00551 | -0.49850 | 0.26440 |
| sd40 | -0.51508 | 1.89900 | -0.03856 |
| sd41 | 1.10646 | 0.92400 | -0.05875 |
| sd42 | 2.27017 | -0.03957 | 0.26480 |
| sd43 | 0.80623 | -0.60170 | 0.20870 |
| sd44 | 0.40232 | -0.30280 | -0.13660 |
| sd45 | 0.83338 | 3.63100 | -0.05030 |
| sd46 | 2.07913 | -3.00800 | -0.02268 |
| sd47 | 1.44562 | -1.90200 | 0.19620 |
| sd48 | 0.60582 | 1.41800 | -0.05144 |
| sd49 | 2.23240 | -0.11240 | 0.29290 |
| sd50 | 2.8449018* | -3.71200 | -0.16940 |

| Dummy varijabla | Prosječna temperatura | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| sd51 | 0.19073 | 4.21000 | -0.06436 |
| sd52 | 1.99327 | -1.63800 | 0.26760 |
| sd53 | 1.32184 | 1.70600 | 0.14200 |
| sd54 | 2.3901837* | 1.41000 | -0.17200 |
| sd55 | 2.9059376* | -0.35780 | 0.73590 |
| sd56 | 1.70475 | -2.17200 | -0.31970 |
| sd57 | 1.88124 | -2.49700 | 0.06316 |
| sd58 | 1.02308 | -3.03400 | 0.28430 |
| sd59 | 1.86481 | -1.82200 | 0.01545 |
| sd60 | 2.9767707* | -1.27900 | 0.04377 |
| sd61 | 2.26385 | -3.139 | -0.22690 |
| sd62 | 1.21765 | -1.44700 | 0.17780 |
| sd63 | 0.61825 | -1.74600 | -0.08432 |
| sd64 | 1.17673 | -2.28900 | 0.38150 |
| sd65 | 2.5344631* | -7.08300 | -0.16050 |
| sd66 | 0.69991 | 2.04500 | 0.19300 |
| sd67 | 1.65550 | 0.87120 | -0.09617 |
| sd68 | 2.8208705* | -8.991* | 0.28850 |
| sd69 | 2.22801 | -5.77200 | 0.08572 |
| sd70 | 2.3033382* | -1.75100 | 0.09788 |
| sd71 | 2.05514 | -2.68000 | -0.14770 |
| sd72 | 2.6065446* | -4.62100 | 0.19490 |
| sd73 | 2.9374012* | -3.57600 | 0.12190 |
| sd74 | 1.30348 | -5.47500 | 0.12190 |
| sd75 | 3.2208386* | -6.51100 | 0.34210 |
| sd76 | 3.1770394* | -9.961* | -0.04846 |
| sd77 | 2.21784 | -6.47400 | 0.14030 |
| sd78 | 1.18869 | 0.60620 | -0.14030 |
| sd79 | 2.4586295* | -6.38800 | -0.07178 |
| sd80 | 3.743296* | -6.68600 | 0.02136 |
| sd81 | 2.8984091* | -9.212* | 0.26670 |
| sd82 | 2.5535951* | -1.71200 | 0.40870 |
| sd83 | 1.98885 | -3.66600 | -0.14140 |
| sd84 | 2.6405517* | -7.49500 | 0.41960 |
| sd85 | 3.765049* | -9.713* | 0.32920 |
| sd86 | 2.12063 | 0.98990 | 0.07419 |
| sd87 | 3.7935549* | -2.91300 | 0.05941 |
| sd88 | 4.0685481* | -7.30800 | 0.13740 |
| sd89 | 3.3742804* | -2.92300 | -0.16830 |
| sd90 | 3.3299849* | -2.51600 | 0.27540 |
| sd91 | 1.39994 | -0.78980 | 0.04407 |
| sd92 | 3.4634142* | -3.74700 | 0.35830 |
| sd93 | 3.5663516* | -6.303 | 0.55910 |
| sd94 | 4.4011919* | -10.57* | 0.14810 |
| sd95 | 2.3707421* | 5.72300 | -0.03475 |
| sd96 | 1.7195582 | -2.03300 | 0.35930 |
| sd97 | 4.7510066* | -5.55000 | 0.48380 |
| sd98 | 2.6016395* | -11.31* | -0.40270 |
| sd99 | 2.944545* | -5.41400 | 0.05013 |
| sd100 | 3.1429276* | -9.168* | 0.10400 |
| sd101 | 4.5924729* | -4.50300 | 0.10660 |
| sd102 | 4.3332413* | -5.11400 | 0.04431 |
| sd103 | 1.45427 | -2.07700 | 0.38700 |
| sd104 | 3.1286727* | -0.42570 | 0.25840 |
| sd105 | 2.8029789* | -1.15500 | -0.15210 |
| sd106 | 4.1125574* | -8.603* | 0.71570 |
| sd107 | 2.8922276* | -1.19000 | -0.24960 |
| sd108 | 4.1141384* | -4.78200 | -0.17050 |
| sd109 | 3.2044936* | -4.28600 | -0.11900 |
| sd110 | 4.6998969* | -8.636* | 0.18960 |
| sd111 | 4.5908535* | -9.347* | -0.17100 |
| sd112 | 2.4003023* | -3.71800 | 0.42830 |
| sd113 | 3.7325638* | -5.01400 | -0.23470 |

| Dummy varijabla | Prosječna temperatura | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| sd114 | 4.7361092* | -9.348* | 0.27890 |
| sd115 | 3.7031065* | -6.59900 | -0.05617 |
| sd116 | 3.7331706* | -10.02* | -0.06823 |
| sd117 | 4.9891416* | -15.45* | 0.39650 |
| sd118 | 3.5242857* | -5.70200 | -0.17610 |
| sd119 | 3.2301889* | -2.90700 | -0.04107 |
| sd120 | 4.9089237* | -5.37000 | 0.03891 |
| sd121 | 4.7508327* | -7.26000 | 0.03768 |
| sd122 | 3.4590799* | -2.47500 | 0.27110 |
| sd123 | 4.2333026* | -3.16200 | 0.27590 |
| sd124 | 2.985531* | -3.66800 | -0.09454 |
| sd125 | 4.5371441* | -9.225* | 0.22010 |
| sd126 | 3.6647452* | -7.63100 | 0.18510 |
| sd127 | 4.2664034* | -5.83800 | -0.10550 |
| sd128 | 3.1264281* | -3.30800 | 0.19450 |
| sd129 | 4.9195571* | -9.244* | -0.11160 |
| sd130 | 5.0912264* | -10.9* | 0.29050 |
| sd131 | 3.7945768* | -4.36000 | -0.09813 |
| sd132 | 3.8924093* | -7.89300 | 0.22290 |
| sd133 | 1.725459 | 1.88800 | -0.01201 |
| sd134 | 4.2932576* | -4.09300 | 0.26550 |
| sd135 | 3.4408805* | -0.16710 | 0.48600 |
| sd136 | 3.2471967* | -5.62200 | 0.13200 |
| sd137 | 4.7087822* | -4.61500 | 0.10310 |
| sd138 | 5.3151641* | -12.02* | -0.13350 |
| sd139 | 5.031173* | -12.03* | -0.07755 |
| sd140 | 4.3947768* | -4.00200 | 0.15860 |
| sd141 | 4.0241577* | -4.63900 | 0.10140 |
| sd142 | 4.1133331* | 0.77200 | -0.17570 |
| sd143 | 3.3723822* | -3.70400 | -0.02734 |
| sd144 | 4.6755008* | -5.80300 | 0.03940 |
| sd145 | 4.4193085* | -6.59500 | 0.10310 |
| sd146 | 3.7271046* | -7.73700 | -0.03296 |
| sd147 | 4.7079303* | 9.349* | 0.00582 |
| sd148 | 2.8105809* | -7.35 | 0.59330 |
| sd149 | 3.0898293* | -5.85600 | -0.28570 |
| sd150 | 4.6714972* | -6.03100 | -0.03899 |
| sd151 | 3.2702503* | -6.83200 | 0.52070 |
| sd152 | 4.0509368* | -0.99810 | 0.06182 |
| sd153 | 3.9179625* | -4.51900 | -0.05143 |
| sd154 | 5.1756883* | -13.15* | -0.03433 |
| sd155 | 3.586544* | -3.33900 | -0.06578 |
| sd156 | 4.5038363* | -2.98200 | -0.00816 |
| sd157 | 5.2365065* | -10.5* | 0.19130 |
| sd158 | 5.1588839* | -12.4* | 0.11830 |
| sd159 | 5.4419153* | -9.829* | 0.33930 |
| sd160 | 5.4570777* | -10.57* | -0.19150 |
| sd161 | 4.6558676* | -10.58* | -0.09063 |
| sd162 | 2.33132 | -2.60100 | 0.38100 |
| sd163 | 5.3574718* | -5.37800 | 0.41430 |
| sd164 | 4.1922541* | -10.74* | 0.20090 |
| sd165 | 4.3593888* | -8.711* | -0.29260 |
| sd166 | 4.9133605* | -10.95* | 0.43930 |
| sd167 | 4.7715645* | -7.63600 | 0.02528 |
| sd168 | 4.2817155* | -5.76900 | -0.23180 |
| sd169 | 4.7081293* | -9.342* | 0.55090 |
| sd170 | 3.9602209* | -6.07200 | -0.03457 |
| sd171 | 4.7112879* | -11.15* | 0.52520 |
| sd172 | 4.5223008* | -9.863* | -0.02869 |
| sd173 | 4.695978* | -8.455* | 0.05408 |
| sd174 | 5.2188657* | -9.835* | 0.63960 |
| sd175 | 2.7719736* | -1.44900 | -0.21430 |
| sd176 | 2.9395157* | -4.39500 | 0.03202 |
| sd177 | 4.3298901* | -9.683* | 0.33410 |
| sd178 | 4.6543566* | -7.36500 | -0.28590 |

| Dummy varijabla | Prosječna temperatura | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| sd179 | 4.494896* | -11.19* | 0.20480 |
| sd180 | 5.605587* | -7.95300 | -0.11080 |
| sd181 | 4.467572* | -6.38300 | -0.12540 |
| sd182 | 4.2669051* | -10.94* | 0.23050 |
| sd183 | 5.4018438* | -10.14* | -0.13960 |
| sd184 | 4.7201316* | -14.14* | 0.21920 |
| sd185 | 4.6980838* | -9.631* | 0.39120 |
| sd186 | 5.0355103* | -9.703* | 0.11560 |
| sd187 | 5.7001817* | -10.21* | -0.14620 |
| sd188 | 5.358987* | -9.355* | 0.57970 |
| sd189 | 4.4160544* | -5.524 | 0.42770 |
| sd190 | 3.5314515* | -9.648* | -0.09263 |
| sd191 | 5.2811483* | -9.853* | -0.23260 |
| sd192 | 4.1190639* | -5.92100 | 0.59750 |
| sd193 | 5.3104792* | -13.61* | 0.24160 |
| sd194 | 5.994053* | -11.48* | 0.56200 |
| sd195 | 4.6272218* | -7.57600 | 0.21690 |
| sd196 | 4.4335663* | -5.61700 | -0.19580 |
| sd197 | 5.0605773* | -10.54* | 0.47840 |
| sd198 | 5.9051955* | -11.7* | 0.56220 |
| sd199 | 4.5275683* | -6.14600 | 0.03187 |
| sd200 | 5.1017971* | -10.77* | 1.177* |
| sd201 | 5.4224487* | -10.4* | 1.113* |
| sd202 | 3.9197028* | -12.1* | 0.78080 |
| sd203 | 4.5899625* | -4.08600 | -0.22020 |
| sd204 | 5.2196278* | -7.38400 | 0.67410 |
| sd205 | 4.5252626* | -5.52400 | 1.232* |
| sd206 | 3.4177369* | -6.04900 | 0.73610 |
| sd207 | 4.5172734* | -7.42900 | -0.00790 |
| sd208 | 5.0410897* | -8.13800 | 0.32250 |
| sd209 | 5.8417632* | -11.12* | 1.797* |
| sd210 | 5.6091904* | -9.836* | 0.59320 |
| sd211 | 4.1427633* | -8.43000 | 0.75590 |
| sd212 | 3.798872* | -5.76000 | 0.42020 |
| sd213 | 6.2226794* | -11.51* | 1.133* |
| sd214 | 6.2979502* | -11.57* | 0.52540 |
| sd215 | 5.5764893* | -10.9* | 1.423* |
| sd216 | 4.3140915* | -7.41500 | 1.778* |
| sd217 | 4.9124985* | -8.558* | 1.946* |
| sd218 | 5.1487976* | -9.436* | 0.9936* |
| sd219 | 5.0672347* | -12.67* | 1.316* |
| sd220 | 6.145216* | -14.63* | 1.305* |
| sd221 | 3.2662396* | -11.04* | 1.659* |
| sd222 | 4.6693939* | -12.91* | 1.342* |
| sd223 | 4.4074205* | -10.59* | 1.233* |
| sd224 | 5.7566792* | -9.932* | 1.559* |
| sd225 | 6.8774856* | -14.01* | 1.587* |
| sd226 | 3.8017368* | -8.64200 | 1.753* |
| sd227 | 5.822527* | -13.86* | 1.722* |
| sd228 | 4.1616446* | -12.39* | 1.912* |
| sd229 | 4.0081626* | -9.014* | 2.22* |
| sd230 | 5.8981977* | -13.23* | 2.091* |
| sd231 | 5.6648002* | -15.62* | 2.201* |
| sd232 | 5.0971983* | -6.12200 | 2.228* |
| sd233 | 4.4332927* | -8.88500 | 2.158* |
| sd234 | 6.039444* | -14.9* | 2.109* |
| sd235 | 5.2284492* | -10.71* | 1.894* |
| sd236 | 4.6765262* | -5.79800 | 1.263* |
| sd237 | 5.0531638* | -12.52* | 2.151* |
| sd238 | 5.6580092* | -8.71100 | 2.197* |
| sd239 | 5.5996565* | -10.5* | 2.053* |
| sd240 | 3.6681217* | -8.56200 | 2.112* |
| sd241 | 4.091476* | -10.86* | 1.945* |
| sd242 | 4.8836295* | -12.29* | 1.077* |
| sd243 | 5.4314947* | -14.41* | 0.69850 |

| Dummy varijabla | Prosječna temperatura | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| sd244 | 4.7946713* | -9.251* | 1.714* |
| sd245 | 4.5506798* | -8.05700 | 1.355* |
| sd246 | 5.4836372* | -12.77* | 1.665* |
| sd247 | 5.770089* | -13.49* | 1.452* |
| sd248 | 3.0474271* | -8.27800 | 1.447* |
| sd249 | 4.1054727* | -10.15* | 1.371* |
| sd250 | 4.0354031* | -14.69* | 2.201* |
| sd251 | 4.4692746* | -8.45800 | 1.555* |
| sd252 | 4.5715225* | -11.02* | 1.62* |
| sd253 | 4.8997202* | -5.86000 | 1.198* |
| sd254 | 4.9741089* | -9.345* | 0.59450 |
| sd255 | 4.722033* | -10.28* | 0.77290 |
| sd256 | 3.2846848* | -3.65800 | 0.939* |
| sd257 | 3.391702* | -4.13800 | 1.833* |
| sd258 | 4.8043167* | -6.52600 | 0.70170 |
| sd259 | 4.8653843* | -12.4* | 1.016* |
| sd260 | 4.0248478* | -7.86400 | 0.55220 |
| sd261 | 4.0236155* | -9.035* | 0.37000 |
| sd262 | 4.1012511* | -7.59200 | 0.03340 |
| sd263 | 1.2624224 | -0.28460 | 0.04039 |
| sd264 | 3.6405701* | -10.63* | 0.69460 |
| sd265 | 3.8534036* | -8.48000 | 0.72510 |
| sd266 | 4.2726686* | -11.08* | 0.55280 |
| sd267 | 4.6113981* | -15.27* | 1.499* |
| sd268 | 2.6235387* | -3.03500 | 0.9469* |
| sd269 | 4.1526701* | -6.94100 | 0.2486 |
| sd270 | 3.8901685* | -11* | 0.48480 |
| sd271 | 3.0313661* | -6.13400 | 0.59180 |
| sd272 | 2.8286947* | -7.33800 | 0.69360 |
| sd273 | 3.2893836* | -7.21200 | 0.17670 |
| sd274 | 4.4013294* | -9.14* | 0.72470 |
| sd275 | 2.4846377* | -1.69400 | 0.45960 |
| sd276 | 3.0274774* | -12.52* | 0.61400 |
| sd277 | 2.7839939* | -5.68400 | 0.27580 |
| sd278 | 3.0992352* | -9.639* | 0.77770 |
| sd279 | 4.5320482* | -9.07* | 1.511* |
| sd280 | 3.6943668* | -6.18800 | 0.52370 |
| sd281 | 1.73442 | -7.69600 | 0.22310 |
| sd282 | 3.4958367* | -2.90700 | 0.38370 |
| sd283 | 3.3413756* | -3.48300 | 0.75670 |
| sd284 | 3.2201853* | -0.50770 | 0.62220 |
| sd285 | 3.6571644* | -7.38900 | 0.73200 |
| sd286 | 2.14493 | -2.18200 | 0.831* |
| sd287 | 2.886248* | -3.60800 | 0.55570 |
| sd288 | 3.2819864* | -6.94000 | 0.77690 |
| sd289 | 2.18945 | -1.93400 | -0.05981 |
| sd290 | 1.49883 | -1.69000 | 0.77140 |
| sd291 | 3.2075354* | -5.54900 | -0.11310 |
| sd292 | 3.256502* | -3.60100 | 0.08227 |
| sd293 | 2.5210561* | -6.08000 | 0.34580 |
| sd294 | 3.1061112* | -4.65600 | 0.62360 |
| sd295 | 2.4654283* | -4.68300 | 0.61430 |
| sd296 | 2.9455285* | -1.63800 | 0.846* |
| sd297 | 2.7727654* | -3.59700 | 0.26590 |
| sd298 | 2.6344468* | -2.32900 | 0.67110 |
| sd299 | 1.8086039 | -4.08500 | 0.39700 |
| sd300 | 2.486751* | -8.296* | 0.9032* |
| sd301 | 2.04685 | -5.66400 | -0.20140 |
| sd302 | 2.4152296* | -7.05200 | 0.28530 |
| sd303 | 1.70206 | -4.02800 | 0.37330 |
| sd304 | 1.95924 | -7.07900 | 0.66860 |
| sd305 | 3.0424146* | -2.70300 | 0.20330 |
| sd306 | 2.3623564* | -7.19100 | -0.61310 |
| sd307 | 1.62320 | -0.66440 | -0.22800 |
| sd308 | 3.4447788* | -4.94200 | -0.15460 |

| Dummy varijabla | Prosječna temperatura | Vlažnost zraka | Transf. konc. peludi |
|---|---|---|---|
| sd309 | 4.3941554* | -7.71600 | 0.02558 |
| sd310 | 1.37592 | -1.59500 | -0.08546 |
| sd311 | 3.3637228* | -6.51800 | -0.45990 |
| sd312 | 4.0157783* | -6.08900 | -0.00974 |
| sd313 | 1.7861004 | -2.27100 | -0.19720 |
| sd314 | 1.78113 | -0.25330 | -0.03545 |
| sd315 | 2.3666757* | -9,023* | 0.01784 |
| sd316 | 1.37097 | -1.75800 | -0.16270 |
| sd317 | 2.4529829* | -3.44300 | -0.08244 |
| sd318 | 1.32571 | -3.05600 | -0.19470 |
| sd319 | 2.5671258* | -6.10400 | 0.02603 |
| sd320 | 1.70609 | 3.04100 | -0.00812 |
| sd321 | 2.4929594* | -5.89600 | -0.18550 |
| sd322 | 1.32509 | 0.48870 | 0.20330 |
| sd323 | 1.87429 | -1.14800 | -0.18800 |
| sd324 | 2.08498 | -1.40900 | -0.01292 |
| sd325 | 1.75595 | 1.27600 | -0.00192 |
| sd326 | -0.20812 | -0.44590 | 0.17350 |
| sd327 | 1.66734 | -1.34900 | -0.05145 |
| sd328 | 0.80264 | -1.88000 | -0.01604 |
| sd329 | 0.53229 | 3.95600 | -0.06679 |
| sd330 | 1.91538 | -2.40600 | 0.34560 |
| sd331 | 1.19233 | -3.86000 | 0.16400 |
| sd332 | 1.47300 | 1.97200 | -0.14510 |
| sd333 | 2.22699 | -2.83200 | 0.23060 |
| sd334 | 1.46533 | -1.28700 | 0.14920 |
| sd335 | 2.19836 | -0.64390 | -0.14250 |
| sd336 | 0.80759 | 2.65900 | -0.05390 |
| sd337 | 2.10225 | -1.99800 | 0.30020 |
| sd338 | 1.50542 | -1.28400 | -0.13910 |
| sd339 | 1.45119 | -2.22200 | -0.07309 |
| sd340 | 1.02096 | 2.97300 | -0.00183 |
| sd341 | 1.08454 | -0.89910 | 0.19970 |
| sd342 | 2.2958846* | -0.59500 | 0.24900 |
| sd343 | -0.45444 | -2.65300 | 0.10110 |
| sd344 | 0.96581 | -1.90000 | 0.08237 |
| sd345 | 0.34854 | 0.90670 | 0.10350 |
| sd346 | 1.14584 | -0.65960 | -0.10210 |
| sd347 | 0.30355 | 0.01450 | -0.04324 |
| sd348 | 0.98148 | 0.08719 | -0.00109 |
| sd349 | 1.37951 | 1.03700 | -0.04640 |
| sd350 | 1.00546 | 0.79550 | 0.03611 |
| sd351 | 0.23816 | -0.58490 | -0.01246 |
| sd352 | 0.64063 | 1.36100 | 0.03016 |
| sd353 | 0.78203 | 0.29050 | 0.01211 |
| sd354 | 1.07861 | 1.10100 | 0.01638 |
| sd355 | 1.40899 | -4.02200 | 0.05756 |
| sd356 | 2.9939957* | -1.87300 | -0.02130 |
| sd357 | 3.2422659* | -4.69300 | -0.04836 |
| sd358 | 1.3319516 | -2.11300 | -0.04558 |
| sd359 | 1.19507 | 0.98690 | -0.04262 |
| sd360 | 0.08464 | 0.23600 | -0.03482 |
| sd361 | -0.64808 | 2.85100 | -0.01132 |
| sd362 | 0.73832 | 2.88000 | 0.06366 |
| sd363 | -1.26559 | 2.09000 | 0.32010 |
| sd364 | 0.98722 | -1.38900 | -0.01344 |

Tablica : Procijenjeni koeficijenti uz *dummy* varijable u modelima za prosječnu dnevnu temperaturu, prosječnu dnevnu vlažnost zraka i transformiranu maksimalnu dnevnu koncentraciju peludi, $\star$ - značajni na nivou značajnosti 0.05