

Primjena statističkih metoda u predviđanju prodaje

Šutalo, Karla

Master's thesis / Diplomski rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:189000>

Rights / Prava: [In copyright](#)

Download date / Datum preuzimanja: **2021-09-22**



Repository / Repozitorij:

[Repository of Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Karla Šutalo

PRIMJENA STATISTIČKIH METODA U
PREDVIĐANJU PRODAJE

Diplomski rad

Voditelj rada:
prof. dr. sc. Siniša Slijepčević

Zagreb, rujan, 2017.godina

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Zahvaljujem se svom mentoru na pomoći oko izrade ovog diplomskog rada, mojim roditeljima i obitelji koji su mi bili potpora i omogućili ugodno studiranje te Vladimiru koji mi je bio najveća podrška.

Sadržaj

Sadržaj	iv
Uvod	1
1 Stacionarni vremenski nizovi	2
1.1 Slučajni procesi	2
1.2 Stacionarnost vremenskih nizova	3
1.3 Trend i sezonalnost	6
1.4 Autokovarijacijska i autokorelacijska funkcija stacionarnog procesa	11
2 Stacionarni ARMA procesi	13
2.1 Kauzalni i invertibilni ARMA procesi	13
2.2 Proces pomičnih prosjeka reda beskonačno	19
2.3 Računanje autokovarijacijske funkcije ARMA(p,q) procesa	21
2.4 Parcijalna autokorelacijska funkcija	23
2.5 Generirajuća autokovarijacijska funkcija	24
3 Opisna statistika	26
4 Primjena ARMA modela na podacima	33
5 Dodatak	39
5.1 Opisna statistika	39
5.2 Primjena ARMA modela na podacima	41
Bibliografija	44

Uvod

Predviđanje prodaje važno je za svako poduzeće, kako bi sukladno predviđanjima moglo planirati svoje buduće poslovanje. Informacije vezane uz predviđanje mogu se upotrijebiti u različitim sferama poslovanja kao što su: proizvodnja, prodaja, marketing i financije. Mogu se koristiti i za određivanje konkretnih planova vezanih uz planiranje zaliha, distribuciju, troškove i dr.

U 1. poglavlju ovog rada definirat ćemo važne pojmove vezane uz vremenske nizove kao što su stacionarnost te autokovarijacijska i autokorelacijska funkcija. Opisat ćemo klasičnu dekompoziciju vremenskog niza te metode uklanjanja trenda i sezonalnosti. U 2. poglavlju predstavljamo stacionarne ARMA procese koji imaju ključnu ulogu u modeliranju vremenskih nizova, dajući najvažnije definicije i teorijske rezultate vezane uz njih.

3. poglavlje je opisna statistika koja je prvi korak kod analize podataka. U tom ćemo poglavlju podatke o prodaji 3 proizvoda iz iste grupe proizvoda jedne od vodećih hrvatskih kompanija prikazati raznim tablicama i grafovima kako bi ih što bolje razumjeli. U 4. poglavlju, na dijelu dobivenih podataka, tražimo onaj ARMA proces koji najbolje opisuje podatke i računamo predviđanje prodaje. To ćemo predviđanje usporediti sa stvarnim podacima, kako bi odredili koliko dobro predviđanje prodaje daje naš model. U zadnjem dijelu rada dajemo kodove u programskom jeziku R koje ćemo koristiti pri obradi podataka za ovaj diplomski rad.

Poglavlje 1

Stacionarni vremenski nizovi

U prvom dijelu ovog rada opisat ćemo slučajne procese. Definirat ćemo slabu i jaku stacionarnost i dati primjere vremenskih nizova. Zatim ćemo definirati autokovarijacijsku i autokorelacijsku funkciju i navesti njihova svojstva. Opisat ćemo model klasične dekompozicije te metode za uklanjanje trenda i sezonalnosti promatranog vremenskog niza.

Vremenski niz je skup opažanja x_t prikupljenih u vremenskim trenucima t iz skupa vremena T_0 .

Razlikujemo diskretan vremenski niz u kojem je skup vremena u kojima opažamo (T_0) diskretan te neprekidan vremenski niz u kojem je skup vremena u kojem opažamo neprekidan vremenski interval. U ovome radu bavit ćemo se diskretnim vremenskim nizovima.

1.1 Slučajni procesi

Fiksirajmo vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbb{P})$. Ω je neprazan skup čije elemente zovemo elementarnim događajima. \mathcal{F} je σ -algebra događaja na skupu Ω , tj. neprazna familija podskupova od Ω koja sadrži Ω i zatvorena je na komplementiranje i prebrojive unije.

$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ je σ -aditivna funkcija na \mathcal{F} takva da je $\mathbb{P}(\Omega) = 1$, a zovemo je vjerojatnost. Označimo s $\mathcal{B} = \mathcal{B}(\mathbb{R})$ Borelovu σ -algebru na \mathbb{R} . To je najmanja σ -algebra podskupova od \mathbb{R} koja sadrži sve poluotvorene intervale.

Definicija 1.1.1. *Slučajna varijabla na (Ω, \mathcal{F}) je funkcija $X : \Omega \rightarrow \mathbb{R}$ izmjeriva u paru σ -algebri $(\mathcal{F}, \mathcal{B})$.*

Pretpostavljamo da je svako opažanje x_t realizacija slučajne varijable X_t . Tada vrijedi da je vremenski niz $\{x_t, t \in T_0\}$ realizacija familije slučajnih varijabli $\{X_t, t \in T_0\}$, pa podatke modeliramo kao realizaciju slučajnog procesa $\{X_t, t \in T\}$ gdje je $T \supset T_0$. U ovom radu koristit ćemo pojam vremenski niz i za podatke i za proces čija su realizacija.

Definicija 1.1.2. *Slučajni proces je familija slučajnih varijabli $\{X_t, t \in T\}$ definiranih na istom vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$.*

U definiciji slučajne varijable (1.1.1.) fiksirali smo vrijeme i definirali funkcije $X_t(\cdot)$ na skupu Ω . Umjesto da fiksiramo vrijeme i definiramo funkcije $X_t(\cdot)$ na skupu Ω , možemo fiksirati elementarni događaj $\omega \in \Omega$ te promatrati funkciju $X(\omega)$ na T .

Definicija 1.1.3. *Funkcije $\{X(\omega), \omega \in \Omega\}$ na T zovemo realizacija procesa $\{X_t, t \in T\}$.*

Distribuciju slučajnog procesa $\{X_t, t \in T\}$ opisujemo korištenjem njegovih konačno-dimenzionalnih distribucija, tj. distribucija vektora $(X_{t_1}, \dots, X_{t_n}), t = (t_1, \dots, t_n) \in T^n, n \in \{1, 2, \dots\}$.

Kod nekih slučajnih procesa morat ćemo odrediti zajedničke distribucije svih konačno-dimenzionalnih vektora. U tom slučaju, morat ćemo provjeriti da slučajni proces s odabranom distribucijom postoji. Zato iskazujemo Kolmogorovljev teorem. Kolmogorovljev teorem osigurava da postoji slučajni proces uz minimalne uvjete na danu funkciju distribucije.

Pretpostavljamo da je T podskup skupa realnih brojeva \mathbb{R} i da je linearno uređen.

Definicija 1.1.4. *Neka je \mathcal{T} skup svih vektora $\{t = (t_1, \dots, t_n)' \in T^n : t_1 < t_2 < \dots < t_n, n = 1, 2, \dots\}$. Tada su konačnodimenzionalne funkcije distribucije od $\{X_t, t \in T\}$ funkcije $\{F_t(\cdot), t \in \mathcal{T}\}$ definirane za $t = (t_1, \dots, t_n)'$ sa*

$$F_t(\mathbf{x}) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n), \mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n.$$

Teorem 1.1.5. *(Kolmogorov)*

Familija funkcija distribucije $\{F_t, t \in \mathcal{T}\}$ je familija konačnodimenzionalnih distribucija nekog slučajnog procesa ako i samo ako $(\forall n \in \mathbb{N})(\forall t = (t_1, \dots, t_n)' \in \mathcal{T}) i 1 \leq i \leq n,$

$$\lim_{x_i \rightarrow \infty} F_t(\mathbf{x}) = F_{t(i)}(\mathbf{x}(i)) \quad (1.1)$$

gdje je x_i i -ta komponenta vektora \mathbf{x} , a $t(i)$ i $\mathbf{x}(i)$ su vektori t i \mathbf{x} bez i -te komponente.

Dokaz. Dokaz teorema se nalazi u [3, str. 72]. □

1.2 Stacionarnost vremenskih nizova

Stacionarni procesi imaju jako važnu ulogu u analizi vremenskih nizova. Mnogi promatrani vremenski nizovi neće izgledom biti stacionarni, ali ih često raznim tehnikama transformiramo kako bi ih mogli modelirati kao realizacije nekih stacionarnih procesa.

Kada imamo konačan broj slučajnih varijabli, korisno je znati njihovu međusobnu zavisnost. Zato računamo kovarijacijsku matricu. Za računanje kovarijacijske matrice vremenskih nizova trebat će nam autokovarijacijska funkcija.

Definicija 1.2.1. Neka su $X = (X_1, \dots, X_n)$ i $Y = (Y_1, \dots, Y_m)$ slučajni vektori takvi da je $\mathbb{E}|X_i| < \infty, i = 1, \dots, n$ i $\mathbb{E}|Y_i| < \infty, i = 1, \dots, m$. Tada definiramo kovarijacijsku matricu od X i Y sa

$$\Sigma_{xy} = \text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)'] = \mathbb{E}(XY') - (\mathbb{E}X)(\mathbb{E}Y)'$$

gdje je (i, j) -ti element matrice Σ_{xy} jednak $\text{Cov}(X_i, Y_j) = \mathbb{E}(X_i Y_j) - \mathbb{E}(X_i)\mathbb{E}(Y_j)$.

Definicija 1.2.2. Neka je $\{X_t, t \in \mathbb{Z}\}$ slučajni proces koji zadovoljava $\text{Var}X_t < \infty$ za svaki $t \in T$. Autokovarijacijska funkcija $\gamma_x : T \times T \rightarrow \mathbb{R}$ procesa $\{X_t\}$ definira se kao

$$\gamma_x(s, t) = \text{Cov}(X_s, X_t), \quad s, t \in T. \quad (1.2)$$

Definicija 1.2.3. Vremenski niz $\{X_t, t \in \mathbb{Z}\}$ je (slabo) stacionaran ako:

- (i) $\mathbb{E}|X_t|^2 < \infty$, za svaki $t \in \mathbb{Z}$,
- (ii) $\mathbb{E}X_t = m$, za svaki $t \in \mathbb{Z}$,
- (iii) $\gamma_x(s, t) = \gamma_x(s + r, t + r)$, za svaki $s, t, r \in \mathbb{Z}$.

Iz trećeg uvjeta (iii) možemo zaključiti da kovarijanca stacionarnog niza ne ovisi o trenutku kada smo počeli promatrati, već ovisi o odmaku ($s - t$), tj. vrijedi $\gamma_x(s, t) = \gamma_x(s - t, 0)$, za svaki $s, t \in \mathbb{Z}$. Dakle, za stacionarni niz autokovarijacijska funkcija je zapravo funkcija jedne varijable, a vrijednost te funkcije u pomaku h se definira kao

$$\gamma_x(h) \equiv \gamma_x(h, 0) = \text{Cov}(X_{t+h}, X_t), \quad \text{za svaki } t, h \in \mathbb{Z}. \quad (1.3)$$

Napomena : Stacionarnost smo definirali samo za $T = \mathbb{Z}$. Kod modeliranja realizacije stacionarnog procesa $\{X_t, t \in T \subset \mathbb{Z}\}$, možemo ju smatrati dijelom realizacije stacionarnog procesa $\{X_t, t \in \mathbb{Z}\}$.

Definicija 1.2.4. Vremenski niz $\{X_t, t \in \mathbb{Z}\}$ je jako (striktno) stacionaran ako su zajedničke distribucije od $(X_{t_1}, \dots, X_{t_k})$ i $(X_{t_1+h}, \dots, X_{t_k+h})$ jednake za svaki pozitivan cijeli broj k i za sve $t_1, t_2, \dots, t_k, h \in \mathbb{N}$.

Definicija 1.2.5. Neka je $\{X_t, t \in \mathbb{Z}\}$ (slabo) stacionaran niz takav da je $\text{Var}X_t \in \langle 0, \infty \rangle$. Autokorelacijska funkcija od $\{X_t\}$ se definira kao

$$\rho_x(h) \equiv \frac{\gamma_x(h)}{\gamma_x(0)} = \text{Corr}(X_{t+h}, X_t), \quad \text{za svaki } t, h \in \mathbb{Z}. \quad (1.4)$$

Jako stacionaran vremenski niz za koji vrijedi $\mathbb{E}|X_t|^2 < \infty$ za svaki $t \in \mathbb{Z}$ je stacionaran niz. To vidimo iz definicije 1.2.3. stacionarnog niza, gdje je prvi uvjet (i) zadovoljen iz pretpostavke. Drugi uvjet (ii) slijedi iz definicije 1.2.4. jako stacionarnog niza kad uzmemo $k = 1$. Iz toga slijedi da su očekivanje od X_t i varijanca od X_t konstantni. Treći uvjet (iii) dobijemo kada u definiciju jako stacionarnog niza 1.2.4. uvrstimo $k = 2$. Vidimo da tada X_{t+h} i X_t imaju istu zajedničku distribuciju i istu kovarijancu za svaki $h \in \mathbb{Z}$. Stacionaran niz nije nužno i jako stacionaran.

Definicija 1.2.6. *Slučajni vektor $Y = (Y_1, \dots, Y_n)$ ima višedimenzionalnu normalnu distribuciju ako i samo ako postoji vektor a , matrica B i slučajni vektor $X = (X_1, \dots, X_m)$ s nezavisnim jednako distribuiranim komponentama $X_t \sim N(0,1)$ i vrijedi $Y = a + BX$.*

Napomena : Za slučajni vektor Y koji ima višedimenzionalnu normalnu distribuciju vrijedi $\mathbb{E}Y = a + B\mathbb{E}(X) = a + 0 = a$, dok je kovarijacijska matrica $\Sigma_Y = BB'$.

Definicija 1.2.7. *Slučajni proces je Gaussov proces ako su mu sve konačnodimenzionalne distribucije višedimenzionalno normalne.*

Iz definicije 1.2.6. Gaussovog procesa vidimo da za svaki $n \in \{1, 2, \dots\}$ i za svaki $h, t_1, t_2, \dots \in \mathbb{Z}$, slučajni vektori $(X_{t_1}, \dots, X_{t_n})$ i $(X_{t_1+h}, \dots, X_{t_n+h})$ imaju isto očekivanje i istu kovarijacijsku matricu, pa i istu distribuciju. Odnosno, stacionaran Gaussov niz je i jako stacionaran.

Definicija 1.2.8. *Neka su $X_i, i=1, 2, \dots$ nezavisne jednako distribuirane slučajne varijable. Slučajna šetnja je vremenski niz $\{S_n, n = 0, 1, 2, \dots\}$ definiran sa $S_0 = 0$ i*

$$S_t = \sum_{i=1}^t X_i, \quad t = 1, 2, \dots$$

Primjer 1.2.1. Neka su $X_i, i = 1, 2, \dots$ nezavisne jednako distribuirane slučajne varijable s očekivanjem 0 i varijancom σ^2 i neka je $\{S_n, n = 0, 1, 2, \dots\}$ slučajna šetnja. Očito je da vrijede svojstva (i) i (ii) iz definicije 1.2.3. Preostaje provjeriti svojstvo (iii). Za $h > 0$ vrijedi

$$\text{Cov}(S_{t+h}, S_t) = \text{Cov}\left(\sum_{i=1}^{t+h} X_i, \sum_{j=1}^t X_j\right) = \text{Cov}\left(\sum_{i=1}^t X_i, \sum_{j=1}^t X_j\right) = \sigma^2 t.$$

Vidimo da $\text{Cov}(S_{t+h}, S_t)$ ovisi o t , pa zaključujemo da je ovaj proces stacionaran samo kad je $\sigma^2 = 0$. Za ostale vrijednosti σ^2 , slučajna šetnja nije stacionarna.

Primjer 1.2.2. Neka su A_1 i A_2 nekorelirane slučajne varijable takve da je $\mathbb{E}A_1 = \mathbb{E}A_2 = 0$ i $\text{Var}A_1 = \text{Var}A_2 = \sigma^2$. Provjeravamo stacionarnost procesa $X_t = A_1 \cos(\theta t) + A_2 \sin(\theta t)$, $t \in \mathbb{Z}$, $\theta \in [-\pi, \pi]$. Svojstva (i) i (ii) iz definicije 1.2.3. su zadovoljena jer vrijedi

$$\begin{aligned}\text{Var}X_t &= \cos^2(\theta t)\sigma^2 + \sin^2(\theta t)\sigma^2 = \sigma^2 < \infty \\ \mathbb{E}X_t &= \cos(\theta t)\mathbb{E}A_1 + \sin(\theta t)\mathbb{E}A_2 = 0.\end{aligned}$$

Preostalo je provjeriti svojstvo (iii). Za $h > 0$ vrijedi

$$\begin{aligned}\text{Cov}(X_{t+h}, X_t) &= \text{Cov}(A_1 \cos(\theta(t+h)) + A_2 \sin(\theta(t+h)), A_1 \cos(\theta t) + A_2 \sin(\theta t)) \\ &= \text{Cov}(A_1 \cos(\theta(t+h)), A_1 \cos(\theta t)) + \text{Cov}(A_2 \sin(\theta(t+h)), A_2 \sin(\theta t)) \\ &= \cos(\theta(t+h)) \cos(\theta t) \sigma^2 + \sin(\theta(t+h)) \sin(\theta t) \sigma^2 \\ &= \sigma^2 \cos(\theta(t+h) - \theta t) = \sigma^2 \cos(\theta h).\end{aligned}$$

Vidimo da $\text{Cov}(X_{t+h}, X_t)$ ne ovisi o t . Dakle, pokazali smo da je ovaj periodičan proces stacionaran.

1.3 Trend i sezonalnost

Kod analize vremenskih nizova prvo trebamo grafički prikazati podatke. Ako postoje jasne nepravilnosti u nizu, niz možemo podijeliti na više samostalnih segmenata. Podatke možemo prikazati kao realizaciju procesa

$$X_t = m_t + s_t + Y_t, \quad t = 1, \dots, n \quad (\text{model klasične dekompozicije}). \quad (1.5)$$

m_t je polako promjenjiva funkcija koju nazivamo komponenta trenda. s_t je periodična funkcija poznatog perioda d koju nazivamo sezonalna komponenta. Y_t je komponenta slučajnog šuma koja je slabo stacionarna. U nastavku ove cjeline opisat ćemo tehnike za određivanje determinističkih komponenti m_t i s_t . Cilj je da komponenta slučajnog šuma Y_t bude stacionaran slučajan proces jer tada kod modeliranja možemo koristiti teoriju stacionarnih procesa.

Nizove koji ne izgledaju stacionarno možemo transformirati i načinom koji su razvili Box i Jenkins (1970.). Njihov se pristup temelji na višestrukoj primjeni diferencijskih operatora na podacima $\{x_t\}$. Diferencijski operator primjenjuje se sve dok podaci ne počnu sličiti realizaciji nekog slučajnog procesa $\{W_t\}$. Tada možemo iskoristiti teoriju stacionarnih procesa za modeliranje, analizu i predviđanje procesa $\{W_t\}$, pa i za originalni proces. U nastavku ćemo opisati metode uklanjanja trenda uz odsutnost sezonalnosti, a potom metode uklanjanja trenda i sezonalnosti.

Uklanjanje trenda uz odsutnost sezonalnosti

Model klasične dekompozicije (1.5) bez komponente sezonalnosti poprima oblik

$$X_t = m_t + Y_t, t = 1, \dots, n \quad (1.6)$$

gdje bez smanjenja općenitosti pretpostavljamo da je $\mathbb{E}Y_t = 0$. U nastavku opisujemo 3 metode: metoda najmanjih kvadrata, zaglađivanje pomičnim prosjecima i diferenciranje.

Metoda T1: Metoda najmanjih kvadrata

U ovoj metodi komponentu trenda m_t želimo modelirati polinomijalnim izrazom.

Na primjer $m_t = a_0 + a_1t + a_2t^2$ gdje parametre a_0, a_1 i a_2 biramo tako da minimiziramo $\sum_t (x_t - m_t)^2$. Tada procjena komponente trenda u ovom primjeru glasi: $\hat{m}_t = \hat{a}_0 + \hat{a}_1t + \hat{a}_2t^2$.

Metoda T2: Zaglađivanje pomičnim prosjecima

Neka je q nenegativan cijeli broj i X_t , za $t = 1, \dots, n$ kao u (1.6). Vidimo da za $t < 1$ i $t > n$ proces nema opažanja, pa dvostrani pomični prosjek W_t procesa $\{X_t\}$ možemo definirati samo za $q + 1 \leq t \leq n - q$,

$$W_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t+j}.$$

Ako iskoristimo dekompoziciju procesa X_t iz (1.5) tada dobivamo

$$W_t = (2q + 1)^{-1} \sum_{j=-q}^q m_{t+j} + (2q + 1)^{-1} \sum_{j=-q}^q Y_{t+j} \approx m_t$$

gdje pretpostavljamo da je m_t aproksimativno linearan na intervalu $[t - q, t + q]$ i da je prosječna greška na tom intervalu približno nula. Zaključujemo da komponentu trenda m_t možemo procijeniti sa

$$\hat{m}_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t+j}, \quad q + 1 \leq t \leq n - q \quad (1.7)$$

odnosno sa

$$\hat{m}_t = W_t, \quad q + 1 \leq t \leq n - q.$$

O $\{\hat{m}_t\}$ iz (1.7) možemo razmišljati kao o procesu dobivenom iz procesa $\{X_t\}$ primjenom linearnog filtra. Tada je

$$\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t+j}, \quad \text{za } a_j = (2q + 1)^{-1}, -q \leq j \leq q \text{ i } a_j = 0, |j| > q. \quad (1.8)$$

Filter dan s (1.8) je samo jedan primjer koji možemo upotrijebiti za zaglađivanje. Kada m_t nije linearan moramo biti oprezni kod biranja broja q jer s izborom prevelikog broja q , filtrirani proces neće biti dobar za procjenu m_t . Dobro odabrani filter će biti efikasan u uklanjanju šuma iz podataka.

Jednostrani pomični prosjek \hat{m}_t za $t = 1, \dots, n$ možemo rekursivno definirati za fiksni $a \in [0, 1]$ sa

$$\begin{aligned}\hat{m}_t &= aX_t + (1 - a)\hat{m}_{t-1}, \quad t = 2, \dots, n \\ \hat{m}_1 &= X_1.\end{aligned}$$

Navedeni postupak nazivamo eksponencijalno zaglađivanje. Rekursivne jednadžbe glase:

$$\hat{m}_t = \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1, \quad t \geq 2.$$

Metoda T3: Diferenciranje

Za razliku od prošle metode, kada nam je cilj bio uklanjanje šuma zaglađivanjem, sada će nam cilj biti uklanjanje trenda diferenciranjem. Operator diferenciranja ∇ definiramo sa

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

gdje je B operator pomaka unatrag kojeg definiramo sa

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

Za operatore ∇ i B vrijedi

$$B^j(X_t) = X_{t-j} \text{ i } \nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t)), \quad j \geq 1 \text{ gdje je } \nabla^0(X_t) = X_t.$$

Pretpostavimo da je funkcija trenda linearna, tj. $m_t = at + b$ i na nju primijenimo operator diferenciranja. Tada je $\nabla m_t = m_t - m_{t-1} = at + b - a(t-1) - b = a$. Odnosno, dobili smo konstantnu funkciju. Na isti način možemo trend stupnja k reducirati na konstantu primjenom operatora ∇^k .

Neka je Y_t stacionaran s očekivanjem 0, te neka je $X_t = m_t + Y_t$, gdje je $m_t = \sum_{j=0}^k a_j t^j$. Tada je

$$\nabla^k X_t = k! a_k + \nabla^k Y_t,$$

tj. dobili smo stacionaran proces s aritmetičkom sredinom $k! a_k$. Možemo zaključiti da se svaki polinomijalni trend može ukloniti diferenciranjem.

Uklanjanje trenda i sezonalnosti

U prošlom odjeljku smo opisali metode uklanjanja trenda uz odsutnost sezonalnosti, a sada ćemo ih prilagoditi za uklanjanje trenda i sezonalnosti. Pretpostavljamo model klasične dekompozicije (1.5) i dodajemo uvjete na $\{s_t\}$

$$s_{t+d} = s_t \quad \text{i} \quad \sum_{j=1}^d s_j = 0,$$

gdje je d period sezonalne komponente. U nastavku ćemo opisati tri metode: metoda malog trenda, metoda pomičnog prosjeka te diferenciranje. U prvoj metodi ćemo promijeniti i oznake za podatke na način

$$x_{j,k} = x_{k+d(j-1)}.$$

Prirodno se nameće primjer gdje nam j predstavlja godine, a k mjeseci pa $x_{j,k}$ predstavlja k -ti mjesec j -te godine. U tom slučaju period komponente sezonalnosti iznosi 12.

Metoda S1: Metoda malog trenda

U slučaju malih promjena u komponenti trenda možemo pretpostaviti da je trend konstantan kroz period d , pa ga procjenjujemo sa

$$\hat{m}_j = \frac{1}{d} \sum_{k=1}^d x_{j,k}.$$

Sezonalnost procjenjujemo sa

$$\hat{s}_k = \frac{1}{N} \sum_{j=1}^N (x_{j,k} - \hat{m}_j) \quad (1.9)$$

gdje je $N = \frac{n}{d}$, a podaci su prikupljeni kroz N punih ciklusa. Iz (1.9) vidimo da je zadovoljen uvjet $\sum_{k=1}^d \hat{s}_k = 0$.

Procijenjena greška je $\hat{Y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k$, $k = 1, \dots, d$, $j = 1, \dots, N$.

Metoda S2: Metoda pomičnog prosjeka

Metoda malog trenda (Metoda S1) ograničena je pretpostavkom da je trend konstantan kroz period d , pa se metoda pomičnog prosjeka koristi češće jer ne ovisi o tom ograničenju. Pretpostavimo da imamo opažanja $\{x_1, \dots, x_n\}$. Primjenom linearnog filtra procjenjujemo trend

koji ćemo iskoristiti za procjenu sezonalnosti. Procjenjujemo ga:

$$\text{za neparne } d = 2q + 1: \hat{m}_t = \frac{1}{d} \sum_{j=-q}^q X_{t+j}$$

$$\text{za parne } d = 2q: \hat{m}_t = \frac{0.5x_{t-q} + x_{t-q+1} + \dots + x_{t+q-1} + 0.5x_{t+q}}{d}, \quad q < t \leq n - q.$$

Za svaki $k = 1, \dots, d$ računamo prosjek w_k skupa $\{(x_{k+jd} - \hat{m}_{k+jd}), \quad q < k + jd \leq n - q\}$. Kako suma w_k -ova nije nužno 0, sezonalnost procjenjujemo sa

$$\hat{s}_k = w_k - d^{-1} \sum_{i=1}^d w_i, \quad k = 1, \dots, d,$$

$$\hat{s}_k = \hat{s}_{k-d}, \quad k > d.$$

Uklanjanjem sezonalnosti dobivamo niz bez sezonalne komponente

$$d_t = x_t - \hat{s}_t, \quad t = 1, \dots, n.$$

Još nam preostaje ukloniti trend od $\{d_t\}$ što možemo učiniti jednom od ranije opisanih metoda u odjeljku: Uklanjanje trenda uz odsutnost sezonalnosti. Procijenjena greška je:

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \dots, n.$$

Metoda S3: Diferenciranje

U odjeljku: Uklanjanje trenda uz odsutnost sezonalnosti opisali smo uklanjanje trenda diferenciranjem i definirali operator $\nabla^d = (1 - B)^d$. Tu metodu proširujemo definiranjem operatora diferenciranja s pomakom d

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t, \quad t = d + 1, \dots, n.$$

Primjenom tog operatora na model klasične dekompozicije (1.5) dobivamo

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}, \quad t = d + 1, \dots, n.$$

Dobivena dekompozicija $\nabla_d X_t$ sastoji se od komponente trenda $(m_t - m_{t-d})$ i komponente šuma $(Y_t - Y_{t-d})$, a sezonalne komponente nema zbog $s_t = s_{t-d}$. Novodobivenu komponentu trenda $(m_t - m_{t-d})$ možemo ukloniti jednom od ranije opisanih metoda u odjeljku: Uklanjanje trenda uz odsutnost sezonalnosti.

1.4 Autokovarijacijska i autokorelacijska funkcija stacionarnog procesa

U prvom odjeljku smo s (1.2) definirali autokovarijacijsku funkciju i s (1.4) autokorelacijsku funkciju. U nastavku ćemo navesti neka njihova svojstva. Definirat ćemo i uzoračku autokovarijacijsku i uzoračku autokorelacijsku funkciju koje su važan dio konstrukcije modela koji najbolje opisuje podatke.

Propozicija 1.4.1. *Autokovarijacijska funkcija $\gamma(\cdot)$ stacionarnog procesa $\{X_t, t \in \mathbb{Z}\}$ zadovoljava:*

- (i) $\gamma(0) \geq 0$,
- (ii) $|\gamma(h)| \leq \gamma(0)$, za svaki $h \in \mathbb{Z}$,
- (iii) $\gamma(h) = \gamma(-h)$, za svaki $h \in \mathbb{Z}$, tj. $\gamma(\cdot)$ je parna funkcija.

Dokaz.

- (i) $\gamma(0) = \text{Cov}(X_t, X_t) = \text{Var}(X_t) \geq 0$.
- (ii) Koristeći Cauchy-Schwarzovu nejednakost dobivamo $|\text{Cov}(X_{t+h}, X_t)| \leq (\text{Var}(X_{t+h}))^{1/2} (\text{Var}(X_t))^{1/2}$.
- (iii) Po (1.3) vrijedi $\gamma(-h) = \text{Cov}(X_{t-h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma(h)$.

□

Definicija 1.4.2. *Za funkciju $f : \mathbb{Z} \rightarrow \mathbb{R}$ kažemo da je pozitivno semidefinitna ako i samo ako vrijedi*

$$\sum_{i,j=1}^n a_i f(t_i - t_j) a_j \geq 0$$

za sve $n \in \mathbb{N}$ i za sve vektore $\mathbf{a} = (a_1, \dots, a_n)' \in \mathbb{R}^n$ i $\mathbf{t} = (t_1, \dots, t_n)' \in \mathbb{Z}^n$ ili ako i samo ako vrijedi $\sum_{i,j=1}^n a_i f(i - j) a_j \geq 0$ za sve takve n i \mathbf{a} .

Teorem 1.4.3. *Funkcija $f : \mathbb{Z} \rightarrow \mathbb{R}$ je autokovarijacijska funkcija nekog stacionarnog niza ako i samo ako je f parna i pozitivno semidefinitna.*

Dokaz. Dokaz teorema se nalazi u [1, str. 27].

□

Napomena : Ako želimo provjeriti da je neka funkcija pozitivno semidefinitna ponekad je lakše naći stacionaran niz s danom autokovarijacijskom funkcijom i iskoristiti teorem 1.4.3., nego pozitivnu semidefinitnost pokazivati po definiciji 1.4.2..

Napomena : Autokorelacijska funkcija ima ista svojstva kao i autokovarijacijska funkcija uz dodatni uvjet $\rho(0) = 1$.

U praksi autokovarijacijsku i autokorelacijsku funkciju moramo procijeniti. To je važan korak kod konstruiranja modela koji najbolje opisuje podatke. Neka je $\{X_t, t \in \mathbb{Z}\}$ stacionaran vremenski niz i $\{x_1, \dots, x_n\}$ njegova realizacija. U nastavku ćemo definirati uzoračku autokovarijacijsku i uzoračku autokorelacijsku funkciju.

Definicija 1.4.4. *Uzoračka autokovarijacijska funkcija uzorka $\{x_1, \dots, x_n\}$ se definira kao*

$$\hat{\gamma}(h) := n^{-1} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x}), \quad 0 \leq h < n,$$

$$\hat{\gamma}(h) = \hat{\gamma}(-h), \quad -n < h \leq 0,$$

gdje je $\bar{x} = n^{-1} \sum_{j=1}^n x_j$ aritmetička sredina.

Definicija 1.4.5. *Uzoračka autokorelacijska funkcija uzorka $\{x_1, \dots, x_n\}$ se definira kao*

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad |h| < n.$$

Napomena : Uzoračka autokorelacijska funkcija može biti koristan pokazatelj nestacionarnosti podataka. Kada $|\hat{\rho}(h)|$ polako pada s rastom h , možemo zaključiti kako podaci sadrže trend. U slučaju da se $\hat{\rho}(h)$ ponaša slično s istim periodom, tada je kod podataka prisutna periodična komponenta.

Poglavlje 2

Stacionarni ARMA procesi

ARMA procesi ili autoregresivni procesi pomičnih prosjeka imaju ključnu ulogu u modeliranju vremenskih nizova. Jedan od glavnih razloga tome je što za svaku autokovarijacijsku funkciju $\gamma(\cdot)$ takvu da je $\lim_{h \rightarrow \infty} \gamma(h) = 0$ i za svaki cijeli broj $k > 0$ možemo naći ARMA proces s autokovarijacijskom funkcijom $\gamma_x(\cdot)$ takvom da vrijedi $\gamma_x(h) = \gamma(h)$, $h = 0, 1, \dots, k$.

U ovom ćemo poglavlju definirati kada je stacionaran proces ARMA proces, te neka važna svojstva ARMA procesa kao što su kauzalnost i invertibilnost. Opisat ćemo 3 metode za računanje autokovarijacijske funkcije ARMA procesa, te definirati parcijalnu autokorelacijsku funkciju i generirajuću autokovarijacijsku funkciju.

2.1 Kauzalni i invertibilni ARMA procesi

Definicija 2.1.1. *Slučajni proces $\{Z_t, t \in T\}$ se naziva bijeli šum s očekivanjem 0 i varijancom σ^2 ako vrijedi $\mathbb{E}Z_t = 0$, za svaki $t \in T$ i*

$$\gamma(h) = \begin{cases} \sigma^2, & \text{ako je } h = 0, \\ 0, & \text{ako je } h \neq 0. \end{cases} \quad (2.1)$$

Tada pišemo $\{Z_t\} \sim WN(0, \sigma^2)$.

Za slučajne varijable Z_t koje su nezavisne i jednako distribuirane, te za koje vrijedi $\mathbb{E}Z_t = 0$ i $\text{Var}Z_t = \sigma^2$, pišemo $\{Z_t\} \sim \text{NJD}(0, \sigma^2)$. Za takav proces vrijedi: $\mathbb{E}(Z_t^2) = \sigma^2 < \infty$; $\mathbb{E}(Z_t) = 0$ za svaki t ; zbog nezavisnosti slučajnih varijabli je autokovarijacijska funkcija ista kao u (2.1), odnosno ne ovisi o t . Po definiciji smo pokazali da je $\{Z_t\} \sim \text{NJD}(0, \sigma^2)$ stacionaran, pa je onda svaki $\{Z_t\} \sim \text{NJD}(0, \sigma^2)$ i bijeli šum.

Koristeći bijeli šum možemo generirati široku klasu stacionarnih procesa.

Definicija 2.1.2. Stacionaran proces $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ se naziva ARMA(p,q) proces ako za svaki $t \in \mathbb{Z}$ zadovoljava ARMA jednadžbe

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (2.2)$$

gdje je $\{Z_t\} \sim WN(0, \sigma^2)$ i $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ neki realni brojevi.

Kažemo da je $\{X_t\}$ ARMA(p,q) proces s očekivanjem μ ako je $\{X_t - \mu\}$ ARMA(p,q) proces. Jednadžbu (2.2) možemo zapisati u kompaktnijem obliku

$$\phi(B)X_t = \theta(B)Z_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (2.3)$$

gdje je ϕ polinom stupnja p

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p,$$

θ je polinom stupnja q

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

a B je operator pomaka unatrag definiran sa

$$B^j X_t = X_{t-j}, \quad j = 0, \pm 1, \pm 2, \dots$$

Polinom ϕ ćemo zvati autoregresivni polinom, a polinom θ polinom pomičnih prosjeka.

Primjer 2.1.1. Ako je $\phi(z) \equiv 1$ tada jednadžba (2.3) poprima oblik

$$X_t = \theta(B)Z_t.$$

Takav proces nazivamo proces pomičnih prosjeka reda q, odnosno MA(q) proces. Uz pretpostavku $\theta_0 = 1$ i $\theta_j = 0$ za $j > q$, diferencijska jednadžba

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

ima jedinstveno rješenje $\{X_t\}$ koje je stacionaran proces. Stacionarnost slijedi po definiciji jer je očekivanje konstantno,

$$\mathbb{E}X_t = \sum_{j=0}^q \theta_j \mathbb{E}Z_{t-j} = 0$$

i autokovarijacijska funkcija ne ovisi o t,

$$\text{Cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & \text{ako je } |h| \leq q, \\ 0, & \text{ako je } |h| > q. \end{cases}$$

Primjer 2.1.2. Ako je $\theta(z) \equiv 1$ tada jednadžba (2.3) poprima oblik

$$\phi(B)X_t = Z_t.$$

Takav proces nazivamo autoregresivni proces reda p , odnosno AR(p) proces. Da odredimo rješenje ovog procesa i njegovu stacionarnost, prvo promatramo AR(1) proces

$$X_t = Z_t + \phi_1 X_{t-1}. \quad (2.4)$$

Iteriranjem unatrag dobivamo

$$\begin{aligned} X_t &= Z_t + \phi_1 Z_{t-1} + \phi_1^2 X_{t-2} \\ &= \dots \\ &= Z_t + \phi_1 Z_{t-1} + \dots + \phi_1^k Z_{t-k} + \phi_1^{k+1} X_{t-k-1}. \end{aligned}$$

Ovisno o vrijednosti ϕ_1 , promatrat ćemo 3 slučaja: $|\phi_1| < 1$, $|\phi_1| = 1$ i $|\phi_1| > 1$.

i) $|\phi_1| < 1$

U slučaju kada je $|\phi_1| < 1$ i $\{X_t\}$ stacionaran, vrijedi da je $\mathbb{E}(X_t^2)$ konačno i ne ovisi o t . Dakle $\|X_t\|^2 = \mathbb{E}(X_t^2)$ je konstanta pa vrijedi

$$\|X_t - \sum_{j=0}^k \phi_1^j Z_{t-j}\|^2 = \phi_1^{2k+2} \|X_{t-k-1}\|^2 \rightarrow 0 \text{ kada } k \rightarrow \infty.$$

Budući da $\sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$ konvergira u srednjem, slijedi da je

$$X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}. \quad (2.5)$$

Propozicija 2.1.3. Neka je $\{Z_t\}$ proces takav da je $\sup_t \mathbb{E}|Z_t| < \infty$ i neka je $\sum_{j=-\infty}^{\infty} |\phi_j| < \infty$, tada niz

$$\phi(B)Z_t = \sum_{j=-\infty}^{\infty} \phi_j B^j Z_t = \sum_{j=-\infty}^{\infty} \phi_j Z_{t-j}$$

konvergira apsolutno s vjerojatnosti 1. Ako dodatno vrijedi $\sup_t \mathbb{E}|Z_t|^2 < \infty$ tada niz konvergira i u srednjem prema istom limesu.

Dokaz. Dokaz propozicije se nalazi u [1, str. 83]. \square

Vidimo da po ovoj propoziciji, red (2.5) ne konvergira samo u srednjem nego i apsolutno s vjerojatnosti 1. Proces $\{X_t\}$ definiran u (2.5) je stacionaran jer ima konstantno očekivanje,

$$\mathbb{E}X_t = \sum_{j=0}^{\infty} \phi_1^j \mathbb{E}Z_{t-j} = 0$$

i autokovarijacijska funkcija ne ovisi o t ,

$$\begin{aligned} \text{Cov}(X_{t+h}, X_t) &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{j=0}^n \phi_1^j Z_{t+h-j} \right) \left(\sum_{k=0}^n \phi_1^k Z_{t-k} \right) \right] \\ &= \sigma^2 \phi_1^{|h|} \sum_{j=0}^{\infty} \phi_1^{2j} \\ &= \sigma^2 \frac{\phi_1^{|h|}}{1 - \phi_1^2}. \end{aligned}$$

Taj stacionaran proces $\{X_t\}$ zadovoljava diferencijsku jednadžbu (2.4), pa je stoga jedinstveno stacionarno rješenje.

ii) $|\phi_1| > 1$

U slučaju kada je $|\phi_1| > 1$, niz (2.5) ne konvergira u L^2 . Tada AR(1) proces $X_{t+1} = Z_{t+1} + \phi_1 X_t$ možemo zapisati na ovaj način

$$X_t = -\phi_1^{-1} Z_{t+1} + \phi_1^{-1} X_{t+1}.$$

Iteriranjem dobivamo

$$\begin{aligned} X_t &= -\phi_1^{-1} Z_{t+1} - \phi_1^{-2} Z_{t+2} + \phi_1^{-2} X_{t+2} \\ &= \dots \\ &= -\phi_1^{-1} Z_{t+1} - \dots - \phi_1^{-k-1} Z_{t+k+1} + \phi_1^{-k-1} X_{t+k+1}. \end{aligned}$$

Na sličan način kao i u prethodnom slučaju, pokazuje se da je

$$X_t = - \sum_{j=1}^{\infty} \phi_1^{-j} Z_{t+j} \tag{2.6}$$

jedinstveno stacionarno rješenje za (2.4). Ovo se rješenje često smatra neprirodnim jer je ovako definiran X_t koreliran sa $\{Z_s, s > t\}$, a to je svojstvo koje ne vrijedi u slučaju i). Iz tog se razloga često kod modeliranja stacionarnih vremenskih nizova

napravi restrikcija na AR(1) proces s $|\phi_1| < 1$. Na taj način ništa nije izgubljeno jer se svaki AR(1) proces s $|\phi_1| > 1$ može izraziti kao AR(1) proces s $|\phi_1| < 1$ i novim bijelim šumom.

Stacionarno rješenje od (2.4) je (2.6), a nestacionarno rješenje $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ dobijemo kada je X_0 bilo koja slučajna varijabla koja nije korelirana sa $\{Z_t\}$.

iii) $|\phi_1| = 1$

U slučaju kada je $|\phi_1| = 1$, ne postoji stacionarno rješenje od (2.4). Tada po definiciji 2.1.2. ne postoji ni AR(1) proces za $|\phi_1| = 1$.

Definicija 2.1.4. ARMA(p, q) proces definiran jednađbom $\phi(B)X_t = \theta(B)Z_t$ je kauzalan (kauzalna funkcija od $\{Z_t\}$) ako postoji konstantan niz $\{\psi_j\}$ takav da je $\sum_{j=0}^{\infty} |\psi_j| < \infty$ i

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots$$

Kauzalnost nije svojstvo samog procesa $\{X_t\}$, nego je svojstvo veze između procesa $\{X_t\}$ i $\{Z_t\}$ koji se pojavljuju u ARMA jednađbi.

U nastavku iskazujemo teorem koji nam daje nužne i dovoljne uvjete da ARMA proces bude kauzalan.

Teorem 2.1.5. Neka je $\{X_t\}$ ARMA(p, q) proces za koji polinomi $\phi(\cdot)$ i $\theta(\cdot)$ nemaju zajedničkih nultočaka. Tada je $\{X_t\}$ kauzalan ako i samo ako je $\phi(z) \neq 0$ za svaki $z \in \mathbb{C}$ takav da je $|z| \leq 1$.

Dokaz. Dokaz teorema se nalazi u [1, str. 85]. □

Koeficijenti $\{\psi_j\}$ iz definicije 2.1.4 su određeni iz jednakosti

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1. \quad (2.7)$$

Iz teorema slijedi da za $\{X_t\}$ stacionarno rješenje ARMA jednađbe, takvo da je $\phi(z) \neq 0$ za $|z| \leq 1$, vrijedi $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ gdje je $\{\psi_j\}$ definiran s (2.7).

Iz teorema slijedi i obrat, odnosno ako je $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ tada je $\phi(B)X_t = \phi(B)\psi(B)Z_t = \theta(B)Z_t$ gdje je proces $\{\psi(B)Z_t\}$ jedinstveno stacionarno rješenje ARMA jednađbi u slučaju $\phi(z) \neq 0$ za $|z| \leq 1$.

Napomena : Ako $\phi(\cdot)$ i $\theta(\cdot)$ nemaju zajedničke nultočke i ako je $\phi(z) = 0$ za neke $z \in \mathbb{C}$ takve da je $|z| = 1$, tada ne postoji stacionarno rješenje od $\phi(B)X_t = \theta(B)Z_t$.

Rijetko promatramo ARMA proces $\{X_t\}$ s polinomima $\phi(\cdot)$ i $\theta(\cdot)$ koji imaju zajedničke nultočke, jer u tom slučaju postoje dvije mogućnosti:

- i. Niti jedna od zajedničkih nultočaka ne leži na jediničnoj kružnici. U tom slučaju je $\{X_t\}$ jedinstveno stacionarno rješenje ARMA jednadžbi.
- ii. Barem jedna od zajedničkih nultočaka leži na jediničnoj kružnici. U tom slučaju ARMA jednadžbe imaju više od jednog stacionarnog rješenja.

Definicija 2.1.6. Za ARMA(p, q) proces definiran s $\phi(B)X_t = \theta(B)Z_t$ kažemo da je invertibilan ako postoji niz konstanti $\{\pi_j\}$ takav da vrijedi $\sum_{j=0}^{\infty} |\pi_j| < \infty$ i

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.8)$$

Invertibilnost nije svojstvo samog procesa $\{X_t\}$, nego je svojstvo veze između procesa $\{X_t\}$ i $\{Z_t\}$ koji se pojavljuju u ARMA jednadžbi.

U nastavku iskazujemo teorem koji nam daje nužne i dovoljne uvjete za invertibilnost procesa.

Teorem 2.1.7. Neka je $\{X_t\}$ ARMA(p, q) proces za koji polinomi $\phi(\cdot)$ i $\theta(\cdot)$ nemaju zajedničkih nultočaka. Tada je $\{X_t\}$ invertibilan ako i samo ako $\theta(z) \neq 0$ za svaki $z \in \mathbb{C}$ takav da je $|z| \leq 1$.

Dokaz. Dokaz teorema se nalazi u [1, str. 87]. □

Koeficijenti $\{\pi_j\}$ iz definicije 2.1.6. su određeni iz jednakosti

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1. \quad (2.9)$$

Ako je $\{X_t\}$ stacionarno rješenje jednadžbi

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

i ako vrijedi $\phi(z)\theta(z) \neq 0$ za $|z| \leq 1$, tada je

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \text{i} \quad Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

gdje je $\sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$ i $\sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}$, $|z| \leq 1$.

Napomena : Ako je $\{X_t\}$ ARMA proces $\phi(B)X_t = \theta(B)Z_t$, gdje je $\phi(z)$ ne-nul polinom za svaki z takav da je $|z| = 1$, tada je moguće naći polinome $\tilde{\phi}(\cdot)$, $\tilde{\theta}(\cdot)$ i bijeli šum $\{Z_t^*\}$ takav da je $\tilde{\phi}(B)X_t = \tilde{\theta}(B)Z_t^*$ i $\{X_t\}$ kauzalna funkcija od $\{Z_t^*\}$. Ako je i $\theta(z)$ ne-nul polinom za $|z| = 1$, tada možemo izabrati $\tilde{\theta}(\cdot)$ za koji je $\{X_t\}$ i invertibilna funkcija od $\{Z_t^*\}$.

U idućem ćemo primjeru odrediti koeficijente $\{\psi_j\}$ i $\{\pi_j\}$ ARMA(1,1) procesa.

Primjer 2.1.3. Neka je $\{X_t\}$ ARMA(1,1) proces koji zadovoljava sljedeće jednadžbe

$$X_t - 0.5X_{t-1} = Z_t + 0.4Z_{t-1}, \{Z_t\} \sim WN(0, \sigma^2).$$

Vidimo da je tada autoregresivni polinom

$$\phi(z) = 1 - 0.5z.$$

Ovaj polinom ima nultočku $z = 2$, što je izvan jedinične kružnice. Tada po teoremu 2.1.5. zaključujemo da je $\{X_t\}$ kauzalan, a koeficijenti $\{\psi_j\}$ su određeni s (2.7) i iznose

$$\begin{aligned} \psi_0 &= 1, \\ \psi_1 &= 0.4 + 0.5, \\ \psi_2 &= 0.5(0.4 + 0.5), \\ \psi_j &= 0.5^{j-1}(0.4 + 0.5), j = 1, 2, \dots \end{aligned}$$

Polinom pomičnih prosjeka je

$$\theta(z) = 1 + 0.4z.$$

Ovaj polinom ima nultočku $z = -2.5$, što je izvan jedinične kružnice. Tada po teoremu 2.1.7. zaključujemo da je $\{X_t\}$ invertibilan, a koeficijenti $\{\pi_j\}$ su određeni s (2.9) i iznose

$$\begin{aligned} \pi_0 &= 1, \\ \pi_1 &= -(0.4 + 0.5), \\ \pi_2 &= -(0.4 + 0.5)(-0.4), \\ \pi_j &= -(0.4 + 0.5)(-0.4)^{j-1}, j = 1, 2, \dots \end{aligned}$$

U nastavku ovog rada fokusirat ćemo se na kauzalne invertibilne ARMA procese.

2.2 Proces pomičnih prosjeka reda beskonačno

U prethodnom poglavlju smo definirali MA(q) proces, a sada tu definiciju proširujemo za $q = \infty$.

Definicija 2.2.1. Neka je $\{Z_t\} \sim WN(0, \sigma^2)$. Kažemo da je $\{X_t\}$ pomični prosjek (MA(∞)) od $\{Z_t\}$ ako postoji niz $\{\psi_j\}$ takav da vrijedi $\sum_{j=0}^{\infty} |\psi_j| < \infty$ i

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.10)$$

U nastavku navodimo neke primjere pomičnih prosjeka od $\{Z_t\}$.

i) MA(q) proces je pomični prosjek od $\{Z_t\}$ s koeficijentima

$$\psi_j = \theta_j, j = 0, 1, \dots, q \text{ i } \psi_j = 0, j > q.$$

ii) AR(1) proces za koji vrijedi $|\phi| < 1$ je pomični prosjek od $\{Z_t\}$ s koeficijentima

$$\psi_j = \phi^j, j = 0, 1, 2, \dots$$

iii) Kauzalan ARMA(p,q) proces je pomični prosjek od $\{Z_t\}$ s koeficijentima

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, |z| \leq 1.$$

U primjeru iii) ARMA proces mora biti kauzalan zbog definicije pomičnog prosjeka od $\{Z_t\}$. Po definiciji 2.2.1. X_t moramo moći izraziti preko $Z_s, s \leq t$. U slučaju da ARMA proces nije kauzalan, možemo naći bijeli šum $\{Z_t^*\}$ takav da je X_t pomični prosjek od $\{Z_t^*\}$.

Propozicija 2.2.2. *Ako je $\{X_t\}$ stacionaran proces s očekivanjem 0 i autokovarijacijskom funkcijom $\gamma(\cdot)$ takvom da je $\gamma(h) = 0$ za $|h| > q$ i $\gamma(q) \neq 0$, tada je $\{X_t\}$ MA(q) proces. Odnosno, postoji bijeli šum $\{Z_t\}$ takav da vrijedi*

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}.$$

Dokaz. Dokaz propozicije se nalazi u [1, str. 90]. □

Napomena : Ako $\{X_t\}$ ima istu autokovarijacijsku funkciju kao i neki ARMA(p,q) proces, tada je i $\{X_t\}$ ARMA(p,q) proces. Odnosno, tada postoji bijeli šum $\{Z_t\}$ i koeficijenti $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ takvi da vrijedi

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}.$$

Teorem 2.2.3. *MA(∞) proces definiran s (2.10) je stacionaran proces s očekivanjem 0 i autokovarijacijskom funkcijom*

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}.$$

Dokaz. Ovaj teorem je posljedica propozicije čiji se iskaz i dokaz nalaze u [1, str. 84]. □

Ovaj teorem i primjer iii) u potpunosti određuju autokovarijacijsku funkciju γ bilo kojeg kauzalnog ARMA(p,q) procesa.

Definiciju AR(p) procesa iz primjera 2.1.4. možemo proširiti na slučaj kada je $p = \infty$. Tada se pokaže da svaki invertibilni ARMA(p,q) proces ima istu formu kao i AR(p) jednačbe za $p = \infty$,

$$X_t + \sum_{j=1}^{\infty} \pi_j X_{t-j} = Z_t, \quad t = 0, \pm 1, \pm 2, \dots$$

2.3 Računanje autokovarijacijske funkcije ARMA(p,q) procesa

U ovom ćemo poglavlju opisati 3 metode za računanje autokovarijacijske funkcije ARMA procesa.

Metoda $\gamma 1$: U prethodnoj cjelini smo pokazali da autokovarijacijska funkcija kauzalnog ARMA(p,q) procesa $\phi(B)X_t = \theta(B)Z_t$ zadovoljava

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|},$$

gdje je

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)} \quad \text{za } |z| \leq 1, \quad (2.11)$$

$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ i $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$. Dakle, da odredimo autokovarijacijsku funkciju trebamo odrediti koeficijente ψ_j . Uz pretpostavku da su $\theta_0 = 1, \theta_j = 0$ za $j > q$ i $\phi_j = 0$ za $j > p$, (2.11) možemo zapisati u drugoj formi

$$\psi(z)\phi(z) = \theta(z).$$

Nakon izjednačavanja koeficijenata uz z^j dobivamo jednačbe

$$\psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j < \max(p, q + 1)$$

i

$$\psi_j - \sum_{0 < k \leq p} \phi_k \psi_{j-k} = \theta_j, \quad j \geq \max(p, q + 1).$$

Rješavanjem ovih jednačbi dobivamo koeficijente ψ_j :

$$\begin{aligned}\psi_0 &= \theta_0 = 1, \\ \psi_1 &= \theta_1 + \psi_0\phi_1 = \theta_1 + \phi_1, \\ \psi_2 &= \theta_2 + \psi_0\phi_2 + \psi_1\phi_1 = \theta_2 + \phi_2 + \theta_1\phi_1 + \phi_1^2, \\ &\dots\end{aligned}$$

Primjer 2.3.1. Koristeći opisanu metodu γ_1 izračunat ćemo autokovarijacijsku funkciju ARMA(1,1) procesa,

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \{Z_t\} \sim WN(0, \sigma^2),$$

gdje je $|\phi| < 1$. Tada je

$$\begin{aligned}\gamma(0) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\ &= \sigma^2 \left[1 + (\phi + \theta)^2 \sum_{j=0}^{\infty} \phi^{2j} \right] \\ &= \sigma^2 \left[1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right], \\ \gamma(1) &= \sigma^2 \sum_{j=0}^{\infty} \psi_{j+1} \psi_j \\ &= \sigma^2 \left[\theta + \phi + (\phi + \theta)^2 \phi \sum_{j=0}^{\infty} \phi^{2j} \right] \\ &= \sigma^2 \left[\theta + \phi + \frac{(\phi + \theta)^2 \phi}{1 - \phi^2} \right],\end{aligned}$$

i

$$\gamma(h) = \phi^{h-1} \gamma(1), \quad h \geq 2.$$

Metoda γ_2 : Ova se metoda temelji na diferencijskim jednačbama za $\gamma(k)$, $k = 0, 1, 2, \dots$ koje dobijemo tako da pomnožimo svaku stranu jednačbe $\phi(B)X_t = \theta(B)Z_t$ sa X_{t-k} i izračunamo očekivanje. Dobivene diferencijske jednačbe su

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k}, \quad 0 \leq k < \max(p, q+1), \quad (2.12)$$

i

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = 0, \quad k \geq \max(p, q+1), \quad (2.13)$$

gdje koristimo $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$. Ova metoda je najprikladnija ako želimo dobiti rješenje u zatvorenom obliku.

Metoda γ_3 : Ovom metodom numerički određujemo autokovarijacijsku funkciju iz jednadžbi (2.12) i (2.13). Prvo pronađemo $\gamma(0), \dots, \gamma(p)$ iz jednadžbi za $k = 0, \dots, p$, pa onda rekursivno odredimo $\gamma(p+1), \gamma(p+2), \dots$. Ova metoda je najprikladnija ako želimo dobiti numeričke vrijednosti.

2.4 Parcijalna autokorelacijska funkcija

U nastavku ćemo definirati parcijalnu autokorelacijsku funkciju koja nam kao i autokorelacijska funkcija daje važne informacije o zavisnosti stacionarnog procesa.

Definicija 2.4.1. *Parcijalna autokorelacijska funkcija (pacf) stacionarnog vremenskog niza $\{X_t\}$ je funkcija $\alpha(\cdot)$ definirana s jednakostima*

$$\alpha(1) = \text{Corr}(X_2, X_1) = \rho(1)$$

i

$$\alpha(k) = \text{Corr}(X_{k+1} - P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_1), k \geq 2,$$

gdje su $P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_{k+1}$ i $P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_1$ projekcije.

Vrijednost $\alpha(k)$ nazivamo parcijalna autokorelacija u pomaku k .

Parcijalna autokorelacija $\alpha(k)$, $k \geq 2$ je prema definiciji korelacija dva reziduala koja dobijemo nakon regresije X_{k+1} i X_1 na opažanjima X_2, \dots, X_k .

Napomena: Ako stacionaran proces ima očekivanje 0 tada vrijedi $P_{\overline{sp}\{1, X_2, \dots, X_k\}}(\cdot) = P_{\overline{sp}\{X_2, \dots, X_k\}}(\cdot)$.

U nastavku ćemo dati ekvivalentnu definiciju parcijalne autokorelacijske funkcije. Neka je $\{X_t\}$ stacionaran proces s očekivanjem 0 i autokovarijacijskom funkcijom $\gamma(\cdot)$ takvom da $\gamma(h) \rightarrow 0$ kada $h \rightarrow \infty$ i pretpostavimo da su ϕ_{kj} , $j = 1, \dots, k$, $k = 1, 2, \dots$ koeficijenti u izrazu

$$P_{\overline{sp}\{X_1, \dots, X_k\}} X_{k+1} = \sum_{j=1}^k \phi_{kj} X_{k+1-j},$$

tada iz jednadžbi

$$\langle X_{k+1} - P_{\overline{sp}\{X_1, \dots, X_k\}} X_{k+1}, X_j \rangle = 0, j = k, \dots, 1$$

dobivamo

$$\begin{bmatrix} \rho(0) & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & \rho(0) & \rho(1) & \dots & \rho(k-2) \\ \vdots & & & & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{bmatrix}, k \geq 1. \quad (2.14)$$

Definicija 2.4.2. *Parcijalna autokorelacija $\alpha(k)$ u pomaku k od $\{X_t\}$ je*

$$\alpha(k) = \phi_{kk}, k \geq 1,$$

gdje su ϕ_{kk} jedinstveno određeni s (2.14).

Na sličan način definiramo i uzoračku parcijalnu autokorelacijsku funkciju.

Definicija 2.4.3. *Uzoračka parcijalna autokorelacijska funkcija $\hat{\alpha}(k)$ u pomaku k od $\{x_1, \dots, x_n\}$, gdje su $x_i \neq x_j$ za neke i i j , definirana je sa*

$$\hat{\alpha}(k) = \hat{\phi}_{kk}, 1 \leq k < n,$$

gdje su koeficijenti $\hat{\phi}_{kk}$ jedinstveno određeni s (2.14), a svaki $\rho(j)$ je zamijenjen uzoračkom autokorelacijom $\hat{\rho}(j)$.

2.5 Generirajuća autokovarijacijska funkcija

Definicija 2.5.1. *Generirajuća autokovarijacijska funkcija stacionarnog procesa $\{X_t\}$ s autokovarijacijskom funkcijom $\gamma(\cdot)$ je definirana sa*

$$G(z) = \sum_{k=-\infty}^{\infty} \gamma(k)z^k, \quad (2.15)$$

pod uvjetom da niz konvergira za svaki z takav da je $r^{-1} < |z| < r$ za $r > 1$.

Izraz (2.15) pokazuje da određivanjem koeficijenata uz z^k ili z^{-k} možemo lako iz generirajuće autokovarijacijske funkcije odrediti autokovarijaciju u pomaku k , što je praktično koristiti u slučaju kada možemo jednostavno odrediti generirajuću funkciju.

Napomena : $\{X_t\}$ je bijeli šum ako i samo ako je generirajuća autokovarijacijska funkcija konstantna za svaki z .

Ako je

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \{Z_t\} \sim WN(0, \sigma^2)$$

i ako postoji $r > 1$ takav da

$$\sum_{j=-\infty}^{\infty} |\psi_j|z^j < \infty, r^{-1} < |z| < r$$

tada za autokovarijaciju u pomaku k vrijedi

$$\gamma(k) = \text{Cov}(X_{t+k}, X_t) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|k|},$$

pa generirajuća autokovarijacijska funkcija $G(z)$ poprima jednostavni oblik

$$\begin{aligned} G(z) &= \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|k|} z^k \\ &= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \psi_j^2 + \sum_{k=1}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} (z^k + z^{-k}) \right] \\ &= \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j z^j \right) \left(\sum_{k=-\infty}^{\infty} \psi_k z^{-k} \right) \\ &= \sigma^2 \psi(z) \psi(z^{-1}), \quad r^{-1} < |z| < r, \end{aligned}$$

gdje smo koristili

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j, \quad r^{-1} < |z| < r.$$

Poglavlje 3

Opisna statistika

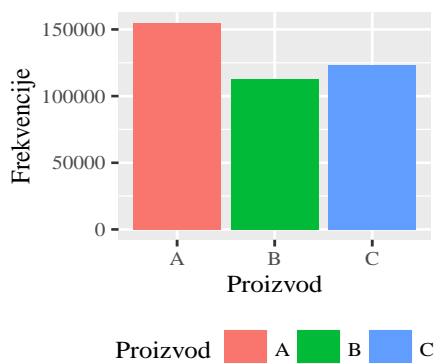
U nastavku ovog rada ćemo obrađenu teoriju primijeniti na stvarnim podacima. Cilj će biti naći model koji najbolje opisuje podatke i napraviti predviđanje buduće prodaje. Predviđanje prodaje je važno za svako poduzeće, kako bi sukladno predviđanjima moglo planirati svoje buduće poslovanje. Informacije vezane uz predviđanje mogu se koristiti za planiranje zaliha, distribuciju, troškove i dr. U ovom poglavlju ćemo se upoznati s podacima, a u idućem na podacima primijeniti ARMA model.

Opisna statistika je prvi korak kod analize podataka. Podatke ćemo opisati pomoću tablica i grafova kako bi ih što boljem razumjeli, pa onda i što bolje analizirali i modelirali. U ovom ćemo radu za analizu i modeliranje podataka koristiti statistički program R. Podaci s kojima ćemo raditi su podaci o prodaji 3 proizvoda iz iste grupe proizvoda koje ćemo zvati: proizvod A, proizvod B i proizvod C. Opisat ćemo podatke iz vremenskog intervala od 1. 3. 2015. do 20. 3. 2016. godine. Za svaku prodaju nekog od ovih proizvoda u navedenom vremenskom intervalu imamo podatke: šifra računa, datum, šifra artikla, količina i sati. U nastavku dajemo prikaz dijela tablice s podacima.

Tablica 3.1: Prikaz dijela podataka

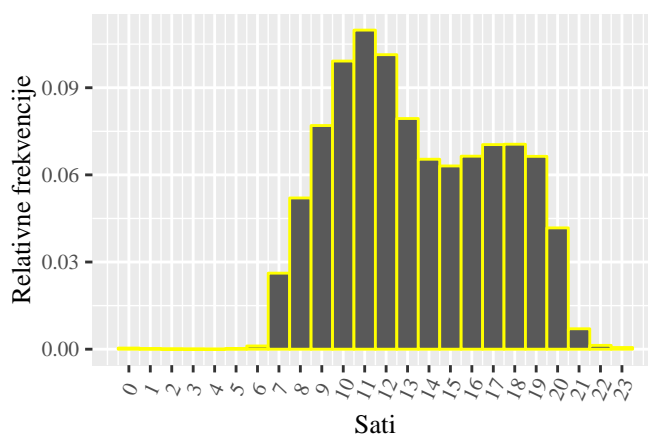
<i>sifra_racuna</i>	<i>datum</i>	<i>artikl_sifra</i>	<i>kolicina</i>	<i>sati</i>
21840/1818/2	2015-03-01	A	1	17
9156/0277/18	2015-03-01	A	1	9
4966/0271/8	2015-03-01	A	1	8
4966/0271/8	2015-03-01	A	1	8
16151/0277/10	2015-03-01	A	1	11

Šifra računa na jedinstven način određuje svaki zapis o prodaji proizvoda iz promatrane grupe proizvoda. Datum je oblika godina-mjesec-dan i sadrži vrijednosti između 2015-03-01 i 2016-03-20. Šifra artikla označava koji je proizvod iz grupe proizvoda kupljen, pa može imati vrijednosti A, B ili C. Količina sadrži informaciju o količini kupljenoga proizvoda, a podatak sati prikazuje sat u kojem se prodaja izvršila.

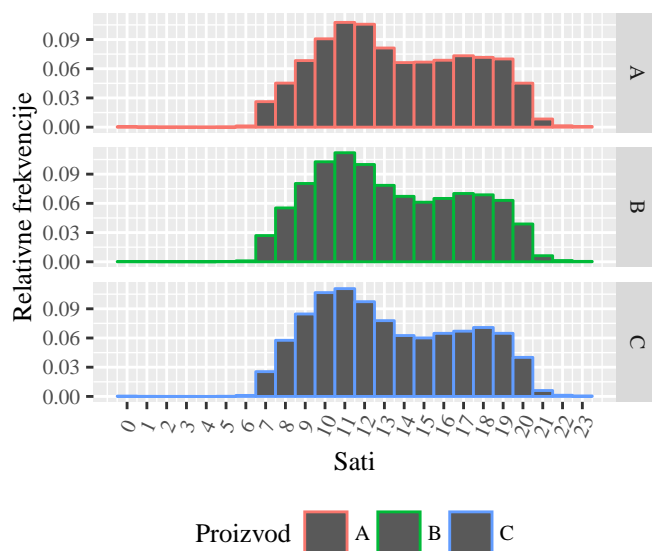


Slika 3.1: Ukupne frekvencije po proizvodima

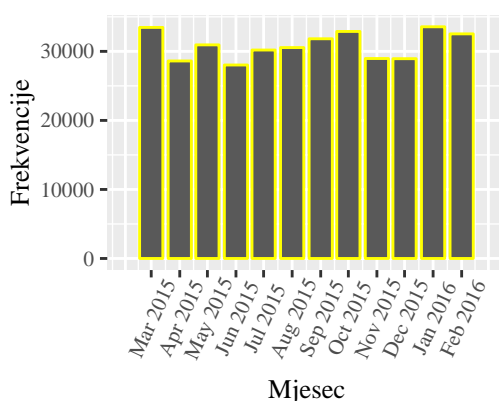
Na slici 3.1 možemo vidjeti da se u razdoblju od 3. mjeseca 2015. do 4. mjeseca 2016. godine najviše prodalo proizvoda A (154 991 komada), zatim proizvoda B (123 264 komada), a najmanje proizvoda C (113 123 komada). Iz promatrane grupe proizvoda ukupno je prodano 391 378 proizvoda.



Slika 3.2: Histogram prodaje svih proizvoda po satima

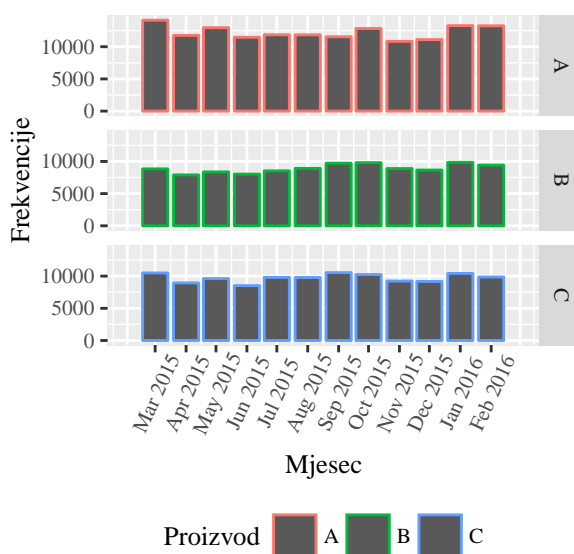


Slika 3.3: Histogrami prodaje različitih proizvoda po satima

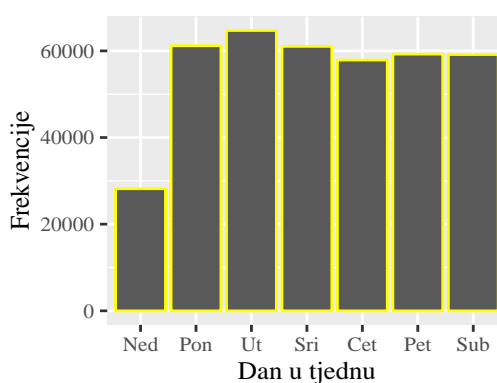


Slika 3.4: Ukupne frekvencije prodaje svih proizvoda po mjesecima

Na temelju slike 3.2 možemo zaključiti da se ova grupa proizvoda najviše kupuje pri-jepodne i to posebno u razdoblju između 9 i 13 sati. Potom slijedi pad u kupnji sve do razdoblja između 16 i 19 sati kada dolazi do blagog porasta kupnje. Na slici 3.3 vidimo da se histogrami prodaje različitih proizvoda po satima ne razlikuju značajno. Svaki se proizvod najviše prodaje u razdoblju između 11:00 i 11:59.

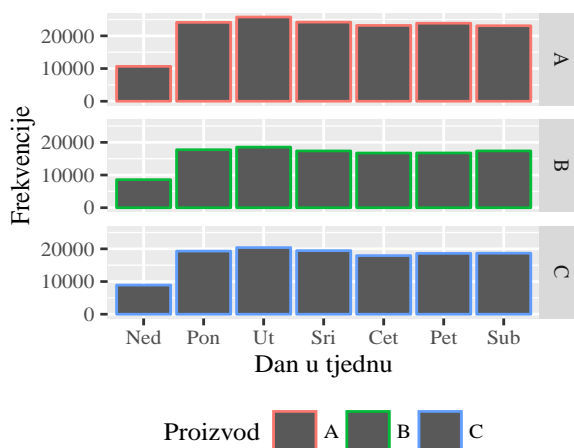


Slika 3.5: Frekvencije prodaje različitih proizvoda po mjesecima



Slika 3.6: Ukupne frekvencije prodaje svih proizvoda po danima u tjednu

Na slici 3.4 su prikazane ukupne frekvencije prodaje svih proizvoda po mjesecima. Vidimo da se najviše proizvoda prodalo u 3. mjesecu 2015. i 1. mjesecu 2016. godine, a najmanje u 6. mjesecu 2015. godine. Kada promatramo prodane količine pojedinih proizvoda na slici 3.5, vidimo da se one razlikuju za različite proizvode. Proizvoda A se najviše prodalo u 3. mjesecu 2015. godine, proizvoda B u 10. mjesecu 2015. godine dok se proizvoda C najviše prodalo u 9. mjesecu 2015. godine. Ukupne frekvencije prodaje svih proizvoda po danima u tjednu smo prikazali na slici 3.6.



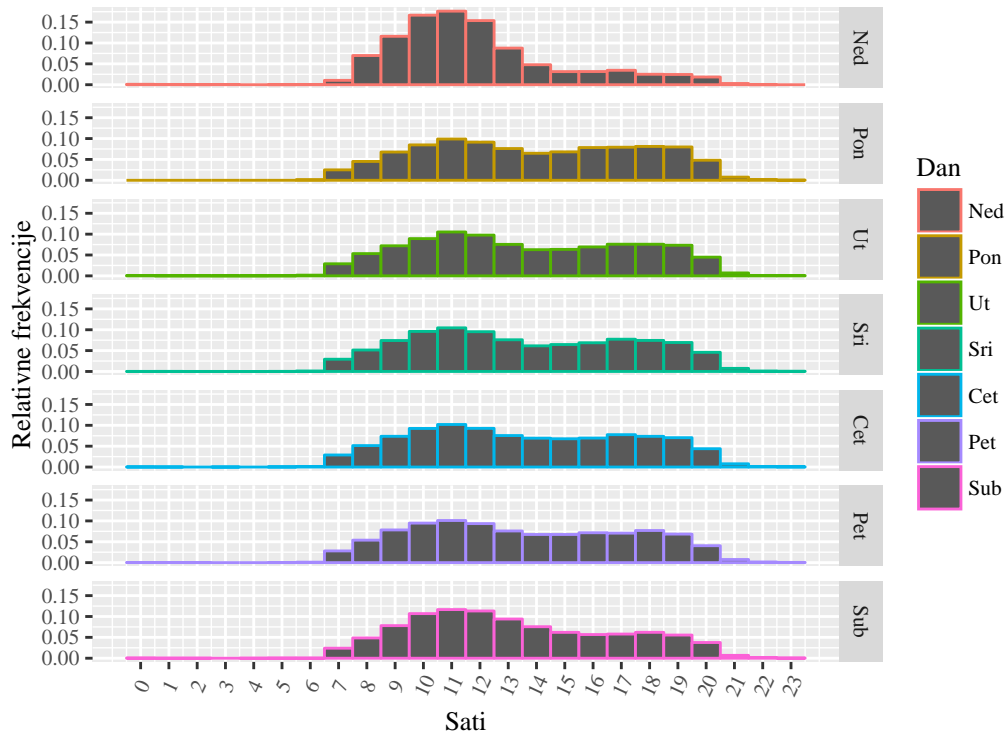
Slika 3.7: Frekvencije prodaje različitih proizvoda po danima u tjednu

Na slici 3.7 smo prikazali frekvencije prodaje različitih proizvoda po danima u tjednu. Možemo uočiti da se nedjeljom prodaje značajno manje proizvoda što je i očekivano s obzirom da dio prodavaonica nedjeljom ne radi ili radi kraće nego ostale dane. To je posebno naglašeno na slici 3.8, koja prikazuje histograme prodaje svih proizvoda po satima za različite dane u tjednu.

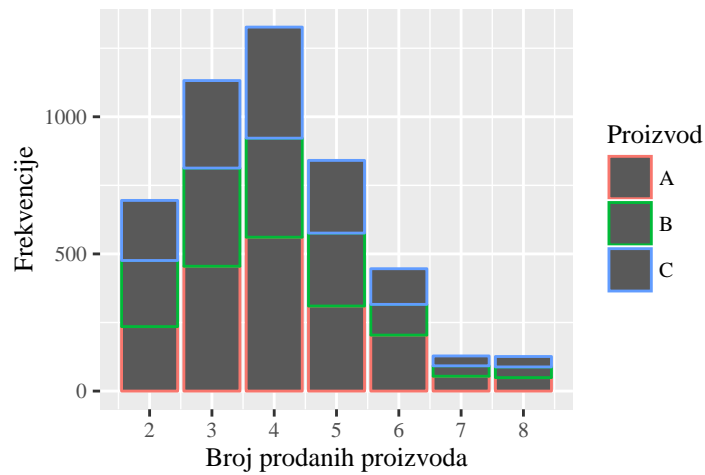
Na slici 3.8 vidimo da se histogram prodaje po satima za nedjelju razlikuje od ostalih histograma. Nedjeljom je većina proizvoda prodana u razdoblju od 9 do 12 sati, a manji dio u ostalim satima. Količina prodanih proizvoda se radnim danima ne razlikuje značajno, iako možemo uočiti da se najviše proizvoda prodalo utorkom, a najmanje četvrtkom. Radnim danima su osim prodane količine podjednaki i histogrami prodaje po satima koje vidimo na slici 3.8. Zaključujemo da se ova grupa proizvoda radnim danima najviše prodaje prijepodne i to posebno u razdoblju između 9 i 13 sati. Potom slijedi pad u prodaji sve do razdoblja između 16 i 19 sati kada dolazi do blagog porasta prodaje.

Na slici 3.6 vidimo da je količina svih prodanih proizvoda subotom podjednaka količini svih prodanih proizvoda radnim danima. Međutim, ako promotrimo histogram prodaje po satima za subotu (slika 3.8) vidimo da se on razlikuje od histograma prodaje po satima za radne dane. Primjećujemo da se subotom većina proizvoda prodava u razdoblju od 9 do 14 sati, a u usporedbi s radnim danima, količina prodaje u poslijepodnevnim satima je manja.

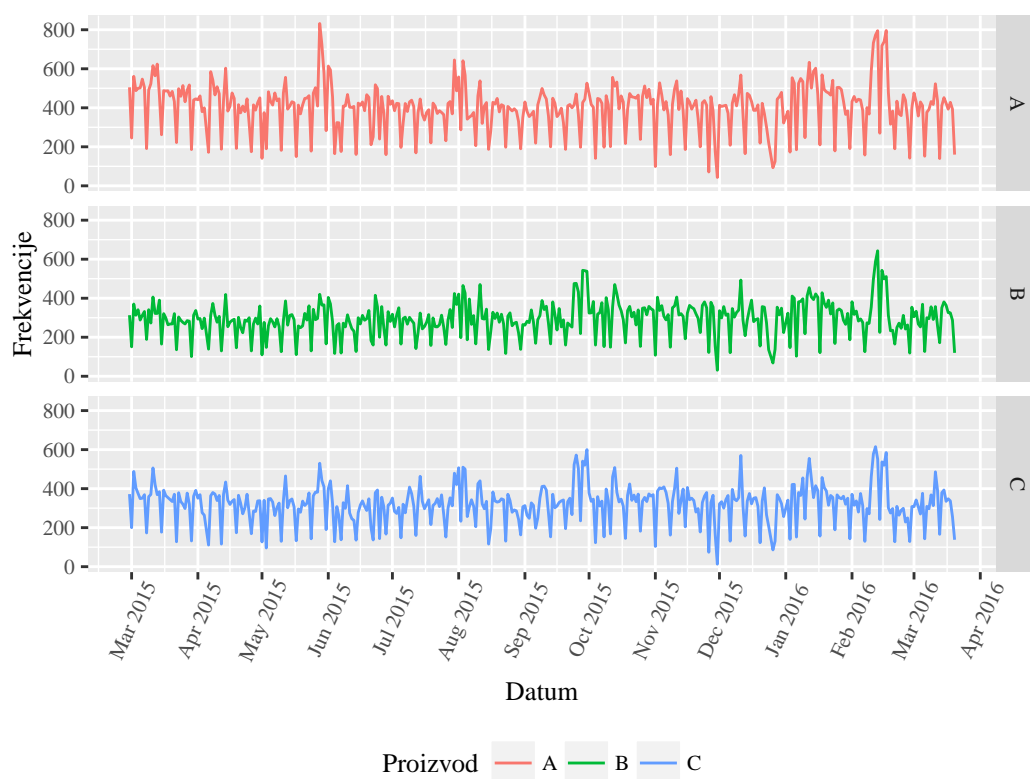
Ako promatramo količinu prodanih proizvoda na jednom računu, uočiti ćemo da je na čak 98.6% računa ta količina 1. Dakle, proizvodi iz ove grupe proizvoda se najviše prodaju po jedan komad. Na slici 3.9 ćemo stoga prikazati samo frekvencije broja prodanih 2, 3, 4, 5, 6, 7 i 8 proizvoda. Zanimljivo je da, ako prodana količina proizvoda nije 1, najčešće će biti 4, zatim 3, pa 5 i tek onda 2.



Slika 3.8: Histogrami prodaje svih proizvoda po satima za različite dane u tjednu



Slika 3.9: Frekvencije količine prodanih proizvoda na jednom računu



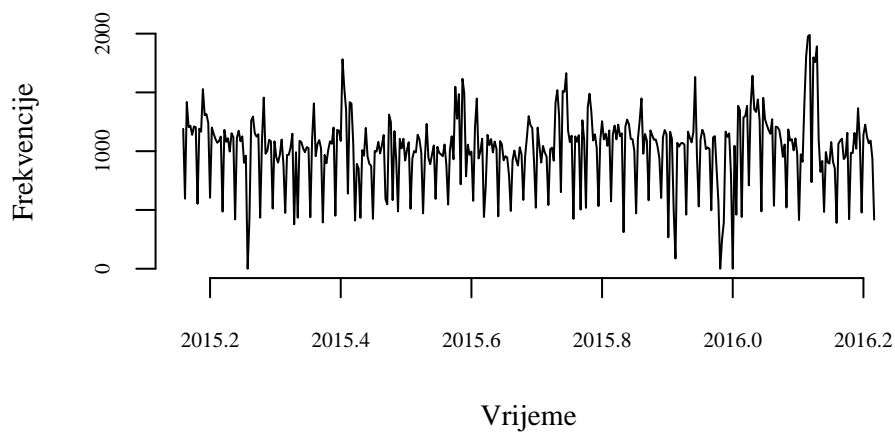
Slika 3.10: Frekvencije prodaje različitih proizvoda po datumima

Slika 3.10 prikazuje frekvencije prodaje različitih proizvoda po datumima. Za svaki proizvod smo zbrojili količinu koja je prodana svakog dana u razdoblju od 1. 3. 2015. godine do 20. 3. 2016. godine. U nastavku ovog rada ćemo koristiti samo ukupne frekvencije prodaje svih proizvoda po datumima.

Poglavlje 4

Primjena ARMA modela na podacima

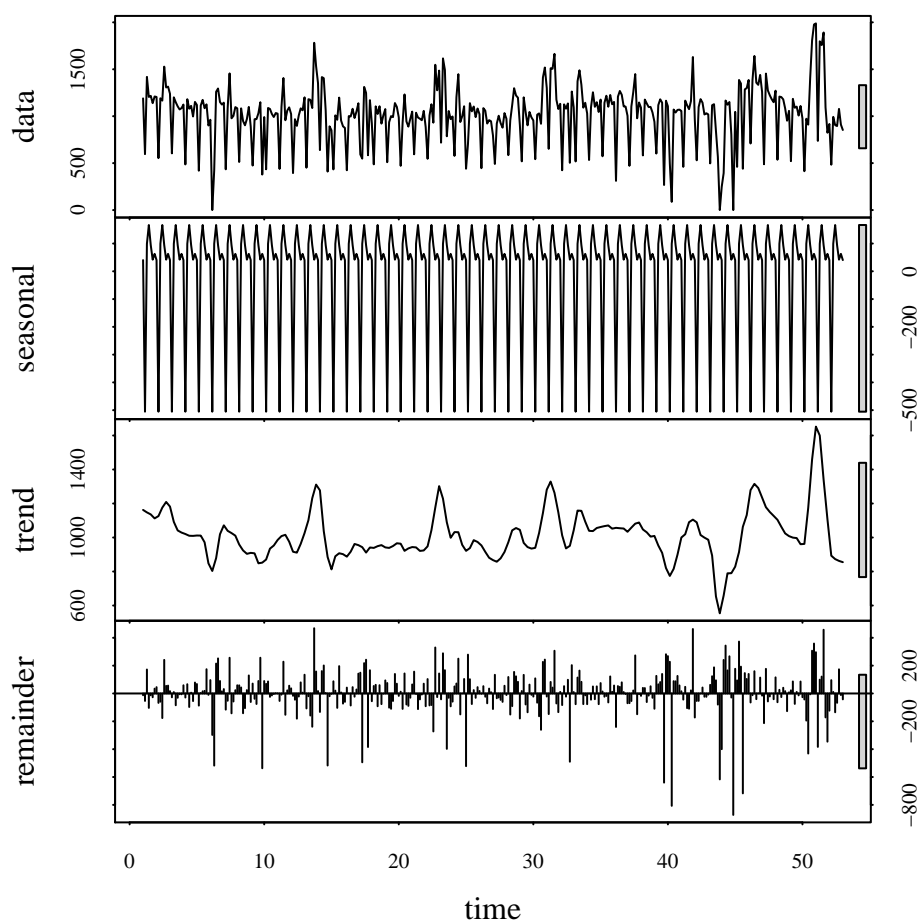
U prethodnom poglavlju smo opisali podatke, a sada ćemo na tim podacima primijeniti ARMA model. Cilj nam je na temelju dijela podataka koje imamo naći onaj model koji najbolje opisuje podatke i pomoću njega napraviti predviđanje prodaje grupe proizvoda. To ćemo predviđanje usporediti sa stvarnim podacima. Kod modeliranja ćemo koristiti podatke o ukupnim frekvencijama prodaje grupe proizvoda po datumima od 28. 2. 2015. do 27. 2. 2016. godine, a predviđanje ćemo provjeriti za datume od 28. 2. 2016. do 20. 3. 2016. godine. Na slici 4.1 vidimo prikaz podataka koje smo zapisali kao vremenski niz.



Slika 4.1: Ukupne frekvencije prodaje kroz vrijeme

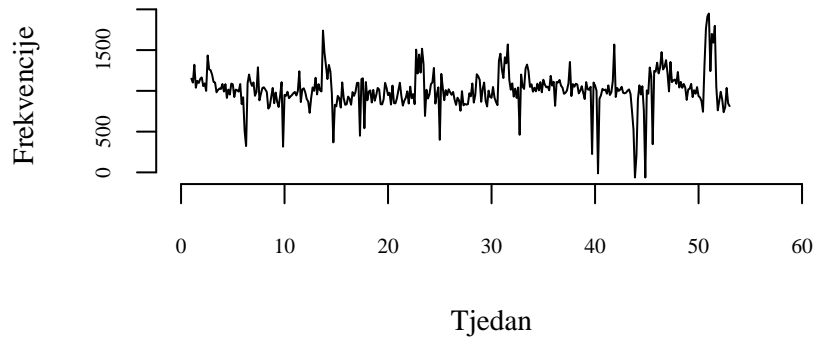
Sa slike 4.1 lako je uočiti da postoji tjedna sezonalnost, a to nam potvrđuje i slika 4.2 koja prikazuje klasičnu dekompoziciju promatranog niza. Klasičnu dekompoziciju smo napravili u R-u s naredbom

```
stl(grupa_proizvoda_ARMA,"periodic").
```



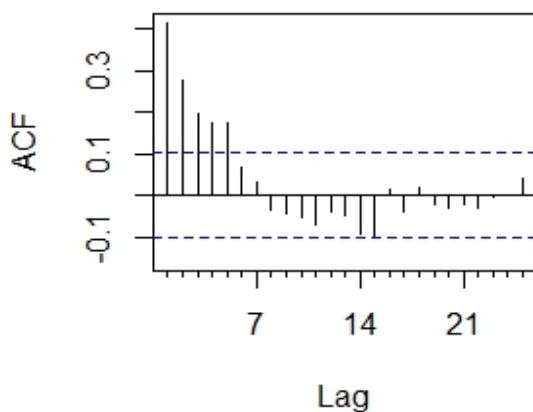
Slika 4.2: Klasična dekompozicija promatranog vremenskog niza

Sa slike 4.2 vidimo prisutnu sezonalnu komponentu s periodom 7 i zaključujemo da nije prisutna komponenta trenda. Niz smo nakon uklonjene sezonalne komponente prikazali na slici 4.3.

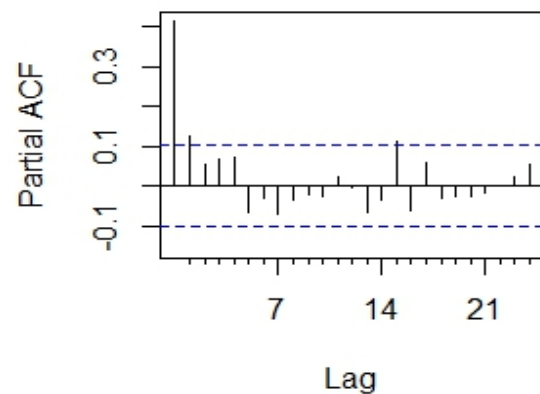


Slika 4.3: Vremenski niz bez sezonalne komponente

Nakon što smo iz vremenskog niza uklonili komponentu sezonalnosti koju je odredila funkcija *stl()*, ostaju nam reziduali. Reziduali su stacionaran niz za koji ćemo tražiti ARMA model koji ih najbolje opisuje. Stacionarnost smo analizirali korištenjem ADF testa (engl. Augmented Dickey-Fuller Unit Root test). Na temelju ovog testa odbacujemo nultu hipotezu da postoji jedinični korijen što bi povlačilo nestacionarnost niza. Važan korak u određivanju modela su autokorelacijska funkcija i parcijalna autokorelacijska funkcija koje prikazujemo na slikama 4.4 i 4.5.



Slika 4.4: ACF reziduala



Slika 4.5: PACF reziduala

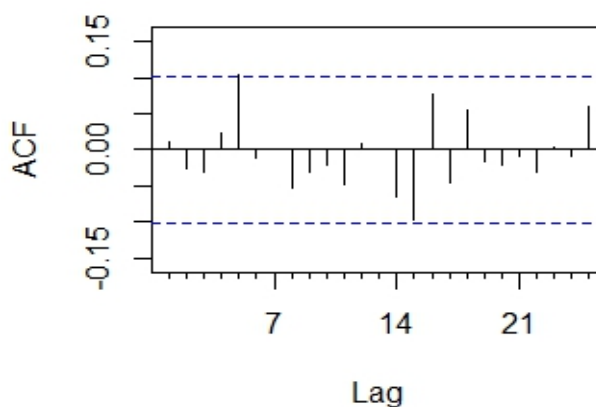
Sa slike 4.4 koja prikazuje autokorelacijsku funkciju primjećujemo kako je ona blizu 0 nakon pomaka 5. Vrijednost funkcije blizu 0 nakon nekoliko koraka ukazuje na stacionarnost reziduala i zaključak da red AR ili MA dijela modela nije veći od 5. Također, možemo zaključiti kako je najznačajniji pad vrijednosti funkcije nakon pomaka 1 kada ostale vrijednosti nastavljaju opadati. Na temelju slike 4.5 koja prikazuje parcijalnu autokorelacijsku funkciju možemo zaključiti da ona pada blizu 0 odmah nakon pomaka 1, nakon kojeg su ostale vrijednosti blizu 0.

Dakle, možemo naslutiti da je jedan od kandidata za model koji najbolje opisuje podatke ARMA(1,1) model. Međutim, ACF i PACF ne daju jasne odgovore kada se radi o određivanju reda ARMA modela, pa ćemo najbolji model odrediti računajući AIC (engl. Akaike information criterion). AIC je mjera koja se temelji na ocjeni kompleksnosti modela. Svakom će modelu pridružiti numeričku vrijednost na temelju koje možemo uspoređivati modele i odrediti najbolji kao onaj s najnižom vrijednosti AIC-a. U R-u radimo funkciju koja prima rezidualne i prirodan broj n te za svaki ARMA(p,q) model gdje je $p + q < n$, računa vrijednost AIC i vraća onaj model kojem je ta vrijednost najmanja. Ovu funkciju smo pozvali za rezidualne i $n = 10$ jer smo sa slike 4.4 zaključili da red AR ili MA dijela modela nije veći od 5. Funkcija je izračunala da ARMA(1,1) model ima najmanji AIC, pa se odlučujemo za taj model.

Koeficijente i rezidualne ARMA(1,1) modela dobivamo naredbom

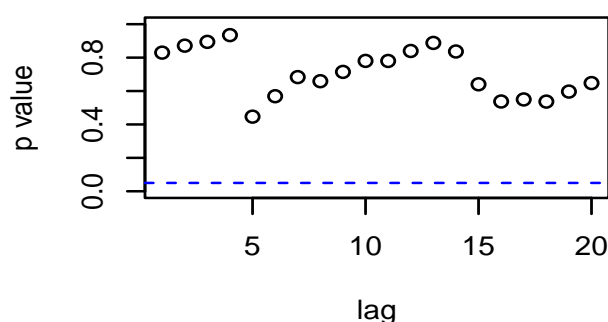
```
arima(reziduali, order = c(1, 0, 1), method = "ML").
```

Za dobivene rezidualne odabranog modela prikazujemo ACF na slici 4.6.



Slika 4.6: ACF reziduala odabranog modela

Na temelju slike 4.6 zaključujemo da su reziduali odabranog modela nekorelirani. Nekoreliranost reziduala provjerit ćemo i Ljung-Box testom, za koji prikazujemo p-vrijednosti do pomaka 20 koje smo dobili naredbom

$$tsdiag(y, \text{gof.lag} = 20).$$


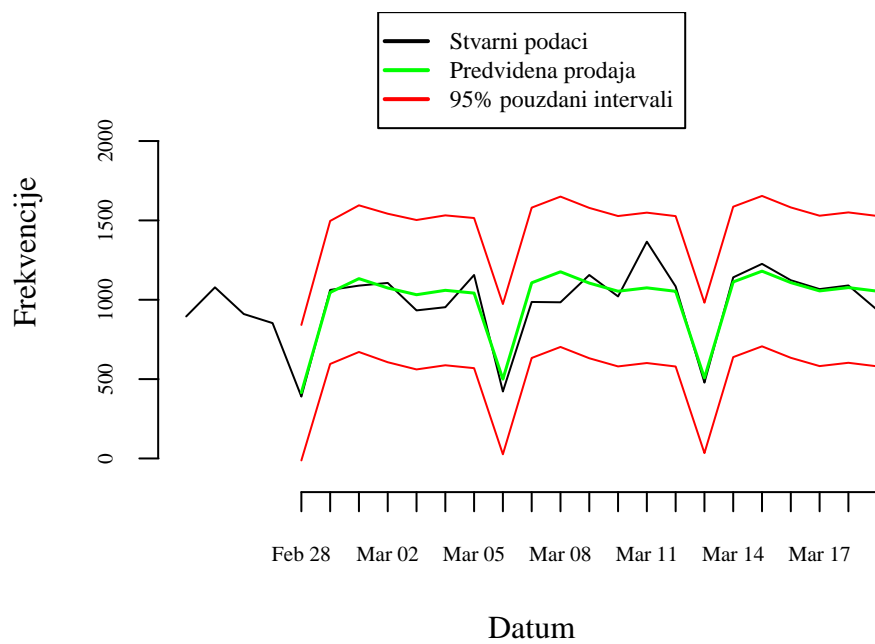
Slika 4.7: p-vrijednosti za Ljung-Box test

Promatrajući p-vrijednosti na slici 4.7 zaključujemo da ne odbacujemo nultu hipotezu o nekoreliranosti reziduala modela. Iz tog razloga prihvaćamo odabrani model i koristimo ga za predviđanje prodaje grupe proizvoda za iduća 3 tjedna.

Naredbom

$$\text{predict}(y, \text{n.ahead} = 21)$$

procjenjujemo budućih 21 reziduala i standardne greške za svaku procjenu. Tim procijenjenim rezidualima dodajemo sezonalnu komponentu i tako dobivamo procjene prodaje. Dobiveno predviđanje prikazujemo na slici 4.8 zelenom linijom zajedno sa stvarnim podacima koji su prikazani crnom linijom. Crvenim linijama smo prikazali 95% pouzdane intervale za dobivena predviđanja. Možemo zaključiti da dobivena predviđanja dobro opisuju stvarne podatke.



Slika 4.8: Predviđena prodaja i 95% pouzdani intervali za iduća 3 tjedna zajedno sa stvarnim podacima

Poglavlje 5

Dodatak

U ovom dijelu rada ćemo navesti kodove u programu R koje smo koristili za dobivanje grafova i rezultata. U [4] se može saznati više o korištenom programu R.

Koristimo razne pakete kako bi imali što više mogućnosti kod prikaza i analize podataka.

```
1 library(ggplot2)
2 library(scales)
3 library(plyr)
4 library(forecast)
```

Podatke učítavamo, spajamo u jednu tablicu i ispisujemo njihov prikaz.

```
5 proizvod_A <- read.csv("proizvod_A.csv", sep=",")
6 proizvod_B <- read.csv("proizvod_B.csv", sep=",")
7 proizvod_C <- read.csv("proizvod_C.csv", sep=",")
8 proizvodi <- rbind(proizvod_A, proizvod_B, proizvod_C)
9 head(proizvodi)
```

5.1 Opisna statistika

U poglavlju 3 Opisna statistika, podatke smo opisali pomoću grafova. Za izradu grafova koristimo paket ggplot2 o čijim se mogućnostima može više pročitati u [8]. U nastavku su kodovi korišteni za dobivanje navedenih slika.

Slika 3.1: Ukupne frekvencije po proizvodima

```
10 ggplot(proizvodi) + geom_bar(aes(ARTIKL_SIFRA, weight=KOLICINA_OSNOVNA_JM, fill=
  ARTIKL_SIFRA)) + xlab("Proizvod") + ylab("Frekvencije") + labs(fill="Proizvod") +
  theme(legend.position="bottom", text=element_text(size=10))
```

Slika 3.2: Histogram prodaje svih proizvoda po satima

```
11 ggplot(proizvodi) + geom_histogram(aes(Sati, y=..density.., weight=KOLICINA_OSNOVNA_JM),
  colour="yellow", binwidth=1) + ylab("Relativne frekvencije") + scale_x_continuous(
  breaks=0:23) + theme(axis.text.x = element_text(angle=65, vjust=0.6), text=
  element_text(size=10))
```

Slika 3.3: Histogrami prodaje različitih proizvoda po satima

```
12 ggplot(proizvodi) + geom_histogram(aes(Sati, y=..density.., weight=KOLICINA_OSNOVNA_JM,
  colour=ARTIKL_SIFRA), binwidth=1) + ylab("Relativne frekvencije") +
  scale_x_continuous(breaks=0:23) + facet_grid(ARTIKL_SIFRA~.) + labs(colour="Proizvod")
  + theme(axis.text.x = element_text(angle=65, vjust=0.6), legend.position="bottom",
  text=element_text(size=10))
```

Slika 3.4: Ukupne frekvencije prodaje svih proizvoda po mjesecima

```
13 proizvodi$datum<-as.Date(proizvodi$datum)
14 proizvodi$Mjesec <- as.Date(cut(proizvodi$datum,breaks = "month"))
15
16 ggplot(proizvodi) + geom_bar(aes(Mjesec, weight=KOLICINA_OSNOVNA_JM), colour="yellow") +
  xlab("Mjesec") + ylab("Frekvencije") + scale_x_date(limits=as.Date(c('2015-02-1',
  '2016-03-01')), date_labels = "%b %Y", date_breaks = "1 month") + theme(axis.text.x =
  element_text(angle=65, vjust=0.6), text=element_text(size=10))
```

Slika 3.5: Frekvencije prodaje različitih proizvoda po mjesecima

```
17 ggplot(proizvodi) + geom_bar(aes(Mjesec, weight=KOLICINA_OSNOVNA_JM, colour=ARTIKL_SIFRA))
  + xlab("Mjesec") + ylab("Frekvencije") + scale_x_date(limits=as.Date(c('2015-02-01',
  '2016-03-01')), date_labels = "%b %Y", date_breaks = "1 month") + theme(axis.text.x
  = element_text(angle=65, vjust=0.6), legend.position="bottom", text=element_text(size
  =10)) + facet_grid(ARTIKL_SIFRA~.) + labs(colour="Proizvod")
```

Slika 3.6: Ukupne frekvencije prodaje svih proizvoda po danima u tjednu

```
18 proizvodi$Dan <- as.POSIXlt(proizvodi$datum)$wday
19 proizvodi$Dan<-as.factor(proizvodi$Dan)
20 levels(proizvodi$Dan) <- c("Ned","Pon","Ut","Sri","Cet","Pet","Sub")
21
22 ggplot(proizvodi) + geom_bar(aes(Dan, weight=KOLICINA_OSNOVNA_JM), colour="yellow") + xlab
  ("Dan u tjednu") + ylab("Frekvencije") + theme(text=element_text(size=10))
```

Slika 3.7: Frekvencije prodaje različitih proizvoda po danima u tjednu

```
23 ggplot(proizvodi) + geom_bar(aes(Dan, weight=KOLICINA_OSNOVNA_JM, colour=ARTIKL_SIFRA)) +
  xlab("Dan u tjednu") + ylab("Frekvencije") + facet_grid(ARTIKL_SIFRA~.) + labs(colour
  ="Proizvod") + theme(legend.position="bottom", text=element_text(size=10))
```

Slika 3.8: Histogrami prodaje svih proizvoda po satima za različite dane u tjednu

```
24 ggplot(proizvodi) + geom_histogram(aes(Sati, y=..density.., weight=KOLICINA_OSNOVNA_JM,
  colour=Dan), binwidth=1) + ylab("Relativne frekvencije") + scale_x_continuous(breaks
  =0:23) + facet_grid(Dan~.) + theme(axis.text.x = element_text(angle=65, vjust=0.6),
  text=element_text(size=10), legend.title=element_text(size=10))
```

Slika 3.9: Frekvencije količine prodanih proizvoda na jednom računu

```
25 ggplot(proizvodi[proizvodi$KOLICINA_OSNOVNA_JM %in% c("2","3","4","5","6","7","8"),]) +
  geom_bar(aes(KOLICINA_OSNOVNA_JM, colour=ARTIKL_SIFRA)) + xlab("Broj prodanih
  proizvoda") + ylab("Frekvencije") + scale_x_continuous(breaks=2:8) + labs(colour="
  Proizvod", text=element_text(size=10)) + theme(text=element_text(size=10))
```

Slika 3.10: Frekvencije prodaje različitih proizvoda po datumima

```
26 ggplot(data = proizvodi, aes(datum, KOLICINA_OSNOVNA_JM, colour=ARTIKL_SIFRA)) +
  stat_summary(fun.y = sum, geom = "line") + xlab("Datum") + ylab("Frekvencije") +
  facet_grid(ARTIKL_SIFRA~.) + labs(colour="Proizvod") + scale_x_date(date_labels = "%b
  %Y", date_breaks = "1 month") + theme(axis.text.x = element_text(angle=65, vjust
  =0.6), legend.position="bottom", text=element_text(size=10))
```

5.2 Primjena ARMA modela na podacima

U nastavku rada koristimo ukupne frekvencije prodaje svih proizvoda po datumima, pa te podatke računamo i pripremamo kako bi ih mogli koristiti za modeliranje.

```
27 datum<-seq(as.Date("2015/02/28"), as.Date("2016/03/20"), by = "day")
28 datum<-data.frame(datum)
29
30 Dnevna_prodaja<-aggregate(KOLICINA_OSNOVNA_JM ~ datum,data=proizvodi, sum)
31 temp<-merge(x = datum, y = Dnevna_prodaja, by = "datum", all.x = TRUE)
32 temp$KOLICINA_OSNOVNA_JM[is.na(temp$KOLICINA_OSNOVNA_JM)] <- 0
33 names(temp)<-c('datum', 'ukupno')
34
35 ts_za_ARMA<-ts(temp$ukupno, start=2015+58/365, freq=365)
```

Slika 4.1: Ukupne frekvencije prodaje kroz vrijeme

```
36 plot(ts_za_ARMA, xlab="Vrijeme", ylab="Frekvencije", axes=F)
37 axis(1, cex.axis = 0.7)
38 axis(2, cex.axis = 0.7)
```

Uzimamo razdoblje od 28.2.2015. do 27.2.2016. (1 godina) za razvoj ARMA modela i zapisujemo ga kao vremenski niz s periodom 7.

```
39 grupa_proizvoda_ARMA<-window(ts_za_ARMA, 2015+58/365, 2016+57/365)
40 grupa_proizvoda_ARMA<-ts(grupa_proizvoda_ARMA, start=1, freq=7)
```

Slika 4.2: Klasična dekompozicija promatranog vremenskog niza

```
41 x=stl(grupa_proizvoda_ARMA,"periodic")
42 plot(x)
```

Iz niza uklanjamo sezonalnost i crtamo novi niz.

Slika 4.3: Vremenski niz bez sezonalne komponente

```
43 reziduali=grupa_proizvoda_ARMA-x$time[,1]
44 plot(reziduali,xlab="Tjedan",ylab="Frekvencije",xlim=c(0,60),axes=F)
45 axis(1,cex.axis = 0.7)
46 axis(2,cex.axis = 0.7)
```

Provodimo Augmented Dickey Fuller Unit Root test.

Slika 4.4: ACF reziduala

Slika 4.5: PACF reziduala

```
47 library("tseries")
48 adf.test(reziduali,alternative="stationary")
49 Acf(reziduali,main=NA)
50 Pacf(reziduali,main=NA)
```

Definiramo funkciju s kojom pronalazimo najbolji ARMA model u smislu AIC-a.

```
51 arma_AIC=function(X,n){
52   i=-1
53   j=-1
54   MAX=1000000
55   for(p in 0:n){
56     for(q in 0:(n-p)){
57       A=arima(x=X,order=c(p,0,q),include.mean=T)$aic
58       D=c(A,p,q)
59       print(D)
60       if(A<MAX){
61         MAX=A
62         i=p
63         j=q
64       }
65     }
66   }
67   print(i)
68   print(j)
69   print(MAX)
70   return(c(MAX,i,j))
71 }
```

Pozivamo definiranu funkciju za rezidualne niza i na temelju rezultata odabiremo model ARMA(1,1).

```
72 arma_AIC(reziduali,10)
73 # 1 1 4959.4
```

Slika 4.7: p-vrijednosti za Ljung-Box test

```

74 y=arima(reziduali,order=c(1,0,1),method="ML")
75 confint(y)
76 tsdiag(y,gof.lag =20)

```

Slika 4.6: ACF reziduala odabranog modela

```

77 reziduali_modela<-y$residuals
78 Acf(reziduali_modela,main=NA)

```

Procjenjujemo prodaju za iduća 3 tjedna i 95% pouzdane intervale za dobivene procjene.

```

79 procjena.sezonalnosti=(x$time[,1])[2:22]
80 procjena.arma=predict(y,n.ahead=21)
81 procjena.ostaci=procjena.arma$pred
82 prodaja.procjena=procjena.sezonalnosti+procjena.ostaci
83
84 greska=procjena.arma$se
85 donja=prodaja.procjena-2*greska
86 gornja=prodaja.procjena+2*greska

```

Slika 4.8: Predviđena prodaja i 95% pouzdani intervali za iduća 3 tjedna zajedno sa stvarnim podacima

```

87 plot(seq(as.Date("2016/02/24"), as.Date("2016/03/19"), by = "day"), axes=F, xlab="Datum",
      ylab="Frekvencije", ylim=c(-500,2500), as.numeric(window(ts_za_ARMA, 2016+54/365,
      2016+78/365)), type="l")
88 axis(2,cex.axis = 0.7)
89 axis.Date(1,at=seq(as.Date("2016/02/28"), as.Date("2016/03/19"), by = "day"),cex.axis =
      0.7)
90 lines(seq(as.Date("2016/02/28"), as.Date("2016/03/19"), by = "day"), prodaja.procjena, lwd
      =1.5, col="green", type="l")
91 lines(seq(as.Date("2016/02/28"), as.Date("2016/03/19"), by = "day"), donja, lwd=1, col="
      red")
92 lines(seq(as.Date("2016/02/28"), as.Date("2016/03/19"), by = "day"), gornja, lwd=1, col="
      red")
93 legend("top", c("Stvarni podaci", "Predvidena prodaja", "95% pouzdani intervali"), col=c("
      black","green","red"), pch=c("-", "-", "-"), lwd=2, cex=0.75)

```


Bibliografija

- [1] Brockwell, Peter J. and Davis, Richard A., *Time Series: Theory and Methods*, Springer, New York, 2006.
- [2] Brockwell, Peter J. and Davis, Richard A., *Introduction to Time Series and Forecasting*, Springer, New York, 2002.
- [3] Lamperti J., *Probability: A Survey of the Mathematical Theory*, Wiley, 1996.
- [4] Venables W. N., Smith D. M., the R Core Team, *An Introduction to R*, dostupno na <https://cran.r-project.org/doc/manuals/R-intro.pdf> (kolovoz 2017.)
- [5] Wagner V., *Vremenski nizovi I*, dostupno na https://web.math.pmf.unizg.hr/~wagner/userfiles/nastava/sp2_vjezbe5.pdf (kolovoz 2017.)
- [6] Wagner V., *Vremenski nizovi I*, dostupno na https://web.math.pmf.unizg.hr/~wagner/userfiles/nastava/sp2_vjezbe6.pdf (kolovoz 2017.)
- [7] Wagner V., *Vremenski nizovi I*, dostupno na https://web.math.pmf.unizg.hr/~wagner/userfiles/nastava/sp2_vjezbe7.pdf (kolovoz 2017.)
- [8] Wickham H., *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York, 2009.

Sažetak

Cilj ovog rada je naći model koji najbolje opisuje podatke o prodaji grupe proizvoda i napraviti predviđanje buduće prodaje. Podaci o prodaji su vremenski niz za koji definiramo najvažnije pojmove: stacionarnost, autokovarijacijsku i autokorelacijsku funkciju. Kod modeliranja podataka koristimo teoriju vezanu uz stacionarne procese pa opisujemo metode uklanjanja trenda i sezonalnosti iz podataka kako bi dobili stacionaran niz koji dalje modeliramo. Definiramo ARMA procese, odnosno autoregresivne procese pomičnih prosjeka i navodimo neka njihova važna svojstva. Obradenu teoriju primjenjujemo na stvarnim podacima o prodaji grupe proizvoda, gdje dio podataka modeliramo ARMA procesom i predviđamo njihovu prodaju. Na kraju dobiveno predviđanje uspoređujemo sa stvarnim podacima o prodaji i zaključujemo kako ARMA proces daje dosta dobra predviđanja prodaje.

Summary

The goal of this thesis was to develop a model which describes the sales data for the group of three products and predicts group's future sales. Sales data is time series for which we define the most important terms: stationarity followed by autocovariance and autocorrelation function. Using theoretical knowledge we defined statistical methods for removing trend and seasonal component from the time series which then leaves us with stationary time series used for modeling. We defined ARMA process, which stands for the Autoregressive-moving-average process, and define some of its properties. We applied theoretical knowledge on actual sales data for the product group by developing an ARMA model on data sample and predicting future sales. Finally, we compared predicted future sales with realized sales and came to the conclusion that ARMA process describes our data well and gives good sales prediction.

Životopis

Rođena sam 2. rujna 1992. godine u Karlovcu. U Karlovcu sam završila osnovnu školu i Gimnaziju Karlovac, opći smjer. Godine 2011. upisujem preddiplomski sveučilišni studij Matematika na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu, koji završavam 2015. godine te stičem naziv sveučilišna prvostupnica matematike. Iste godine upisujem diplomski studij Financijska i poslovna matematika, kojeg završavam ovim radom.