

Dekompozicija vremenskih nizova

Biškup, Ida

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, School of Applied Mathematics and Informatics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet primijenjene matematike i informatike**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:126:411796>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-26**



mathos

Repository / Repozitorij:

[Repository of School of Applied Mathematics and Informatics](#)



DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJ



SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FAKULTET PRIMIJENJENE MATEMATIKE I INFORMATIKE

Sveučilišni diplomski studij matematike
modul: financijska matematika i statistika

Dekompozicija vremenskih nizova

DIPLOMSKI RAD

Mentor:

**izv. prof. dr. sc. Danijel Gra-
hovac**

Student:

Ida Biškup

Osijek, 2024.

Sadržaj

1	Uvod	5
2	Osnovni pojmovi	7
3	Transformacija i prilagodbe vremenskih nizova	15
3.1	Kalendarske prilagodbe	15
3.2	Populacijska prilagodba	16
3.3	Inflacijska prilagodba	17
3.4	Transformacije funkcijama	19
4	Dekompozicija i metode dekompozicije vremenskih nizova	23
4.1	Komponente vremenskog niza	23
4.2	Metode dekompozicije	25
4.2.1	Klasična dekompozicija	25
4.2.2	Metode korištene od strane službenih statističkih agencija	31
4.2.3	STL dekompozicija	37
5	Primjene dekompozicije	41
5.1	Mjerenje snage trenda i sezonalnosti	41
5.2	Predviđanje pomoću dekompozicije	43
	Literatura	45
	Sažetak	47
	Summary	49
	Životopis	51

1 | Uvod

U današnje vrijeme vremenski nizovi su svuda oko nas. Bilo da promatramo oscilacije u vodostaju neke rijeke ili kretanje cijena dionica na burzi sve su to podaci koji se bilježe kroz vrijeme. Iako prvi zapisi vremenskih nizova datiraju još iz doba Kineskog Carstva, njihova analiza počela se razvijati tek u 20. stoljeću. Usprkos tome, ideja dekompozicije vremenskih nizova javila se puno ranije. Naime, još u 17. stoljeću astronomi su koristili dekompoziciju u svojim izračunima. S razvojem analize vremenskih nizova i povećanjem količine i kvalitete podataka javila se potreba primjene dekompozicije kako bi se olakšalo uočavanje različitih sezonalnih uzoraka i trenda. Danas se dekompozicija vremenskih nizova uglavnom koristi u službenim statističkim agencijama upravo kako bi se olakšalo baratanje sve većom količinom podataka.

Prvi dio ovog rada posvećen je uvođenju osnovnih pojmova i rezultata vezanih uz vremenske nizove. Osim toga vidjeti ćemo i neke primjere osnovnih modela te neka njihova svojstva. Nakon toga, upoznat ćemo se s različitim metodama transformacije i prilagodbe podataka. Na primjerima ćemo ilustrirati kako nam takve metode pomažu u pojednostavljivanju samih vremenskih nizova kako bi daljnja analiza bila što preciznija.

Središnje poglavlje daje nam pregled najpoznatijih metoda dekompozicije vremenskih nizova. Počevši od najjednostavnijih metoda poput klasične dekompozicije, prateći povijesni razvoj, upoznat ćemo se s najpopularnijim metodama koje se danas koriste. Osim samog opisa, za svaku metodu dekompozicije navodimo algoritam kao i primjer te ćemo istaknuti prednosti, ali i mane.

Posljednje poglavlje daje kratak uvid u moguće primjene dekompozicije vremenskih nizova. Promotrit ćemo kako se različite metode koriste za mjerenje snage trenda i sezonalnosti te za predviđanje.

2 | Osnovni pojmovi

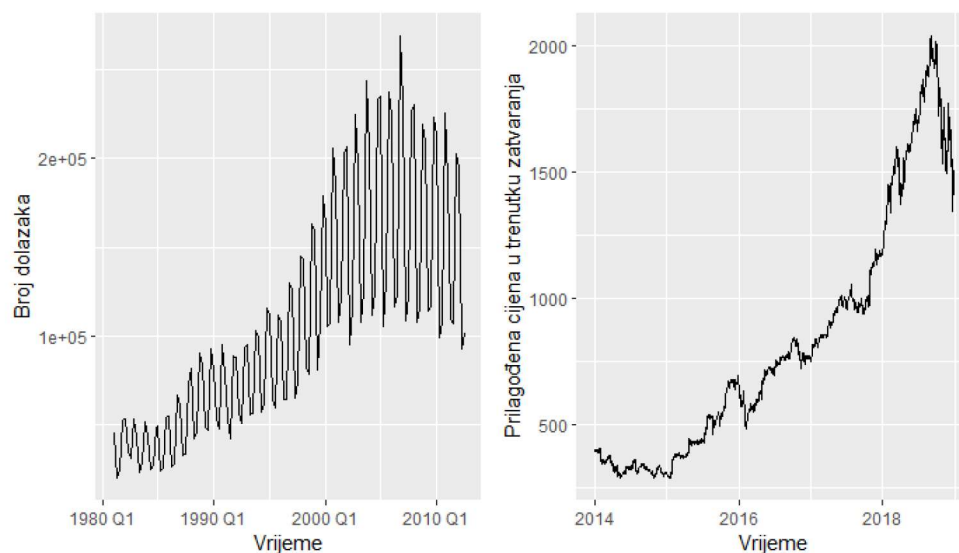
U ovom poglavlju ćemo definirati pojmove koji će nas pratiti kroz cijeli rad. Osim samih definicija, navesti ćemo neka svojstva kao i različite primjere. Započnimo s najjednostavnijom, ali i najbitnijom definicijom, onom vremenskog niza.

Definicija 1. *Vremenski niz je niz podataka*

$$x_{t_1}, \dots, x_{t_n}, \quad (2.1)$$

prikupljenih u uzastopnim vremenskim trenucima $T_0 = \{t_1, \dots, t_n\}$ pri čemu je $t_1 < t_2 < \dots < t_n$ i $x_{t_i} \in \mathbb{R}$.

Ukoliko je skup vremenskih trenutaka T_0 konačan, tada govorimo o vremenskom nizu u **diskretnom vremenu**. Naravno, moguće je da su vrijednosti skupa T_0 , za neki vremenski niz, poznati u svakom trenutku pa se tada radi o vremenskom nizu u **neprekidnom vremenu**. S obzirom da T_0 sadrži vremenske trenutke, oni mogu označavati mjesece, dane, godine itd., a mogu biti raspoređeni **ekvidistantno**, tj. $t_2 - t_1 = \dots = t_n - t_{n-1}$ ili mogu biti raspoređeni **neekvidistantno**. U ovom radu ćemo promatrati vremenske nizove u diskretnom vremenu s ekvidistantnim podacima.



Slika 2.1: Grafički prikaz vremenskih nizova.

Na gornjoj slici možemo vidjeti dva vremenska niza. Lijevi prikazuje broj dolazaka stranih turista u Australiju po kvartalima u razdoblju od 1980. do 2013.

godine. Vremenski niz zdesna prikazuje prilagođenu cijenu dionica Amazona u trenutku zatvaranja za period od 2014. do 2018. godine. Pogledamo li ih detaljnije, vrlo lako uočavamo postojanje nekih uzoraka. Kada imamo dugoročan rast ili pad vrijednosti vremenskog niza, koji ne mora nužno biti linearan, radi se o **trendu**. Tako recimo na oba vremenska niza prikazana na Slici 2.1 možemo zamijetiti rastući trend, ali kod vremenskog niza koji prikazuje kretanje cijena dionica Amazona vidimo i izražen padajući trend u drugoj polovici 2018. godine. Ukoliko imamo varijacije koje se javljaju u određenim pravilnim intervalima manjim od godine, poput tjedna, dana ili kvartala, onda govorimo o **sezonalnosti**. Ona može biti uzrokovana različitim čimbenicima kao što su vremenski uvjeti, godišnji odmori ili praznici. Sastoji se od periodičnih, ponavljajućih i općenito redovitih i predvidivih uzoraka koji se najčešće vrlo lako uočavaju. Sezonalnost možemo vidjeti na lijevom vremenskom nizu budući da se broj dolazaka u Australiju povećava u drugom i trećem kvartalu svake godine, što ima smisla budući da su to ljetni mjeseci na južnoj polutci Zemlje. Pojam sezonalnosti često se miješa s pojmom **cikličnosti** koja se također odnosi na rast i pad među podacima, ali se ne radi o fiksnim frekvencijama. Cikličnost pak možemo vidjeti na grafičkom prikazu cijena dionica Amazona jer imamo fluktuacije, ali one nisu povezane sa sezonalnim faktorom, već ovise o ekonomskim čimbenicima.

Općenito, ako imamo vremenski niz u diskretnom vremenu $\{x_t, t \in T\}$ želimo pronaći matematički objekt, odnosno model čiji bi rezultat mogao dati opažene podatke. Ukoliko bismo to mogli postići korištenjem različitih fizikalnih zakona, takav model bi onda bio u potpunosti **deterministički** jer bismo znali vrijednosti promatrane pojave u svakome trenutku. Ipak, nijedna pojava nije u potpunosti deterministička. Zbog toga promatramo **stohastičke modele** koji nam daju mogućnost procjenjivanja vjerojatnosti da buduća vrijednost bude unutar nekih granica. Stohastički model koji ćemo koristiti za opisivanje vremenskog niza biti će **slučajni proces** koji definiramo u nastavku.

Definicija 2. *Slučajni proces u oznaci $\{X_t, t \in T\}, T \subseteq \mathbb{R}$ je familija slučajnih varijabli definiranih na istom vjerojatnosnom prostoru (Ω, \mathcal{F}, P) .*

Svaki podatak $x_{t_i}, i = 1, \dots, n$ vremenskog niza možemo smatrati realizacijom jedne slučajne varijable X_{t_i} dok sami vremenski niz $\{x_t, t \in T_0\}$ smatramo realizacijom ili dijelom realizacije slučajnog procesa $\{X_t, t \in T\}, T \subseteq T_0$. Već smo spomenuli kako skup trenutaka T može biti diskretan, ali i neprekidan, no mi ćemo uglavnom promatrati modele kod kojih je $T \subseteq \mathbb{Z}$. Sam slučajni proces $\{X_t, t \in T\}$ nazivamo **model vremenskog niza**. Za analizu vremenskih nizova važnu klasu modela čine **stacionarni procesi**. To su procesi koji imaju određenu konstantnost pa na neki način ne mijenjaju svoja distribucijska svojstva kako vrijeme prolazi.

Definicija 3. *Za slučajni proces $\{X_t, t \in T\}$ kažemo da je **strogo stacionaran** ili **stacionaran u užem smislu** ako za svaki $h > 0$ procesi $\{X_t, t \in T\}$ i $\{X_{t+h}, t \in T\}$ imaju jednake konačnodimenzionalne distribucije, tj. vrijedi*

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{h+t_1}, \dots, X_{h+t_n}),$$

za sve $t_1, \dots, t_n \in T$ i $n \in \mathbb{N}$.

Provjera stroge stacionarnosti često može biti vrlo zahtjevna, pogotovo na osnovu samih podataka. Iz tog razloga drugi tip stacionarnosti se definira pomoću momentnih karakteristika pa za početak pogledajmo njihove definicije.

Definicija 4. Neka je $\{X_t, t \in \mathbb{Z}\}$ slučajan proces takav da je $EX_t^2 < \infty$ za sve $t \in \mathbb{Z}$. Definirajmo redom:

(i) **Funkciju očekivanja** $\mu : \mathbb{Z} \rightarrow \mathbb{R}$ slučajnog procesa $\{X_t\}$ koja $t \mapsto EX_t$ sa

$$\mu(t) = EX_t. \quad (2.2)$$

(ii) **Funkciju autokovarijanci** $\gamma : \mathbb{Z}^2 \rightarrow \mathbb{R}$ koja $(t, s) \mapsto Cov(X_t, X_s)$ sa

$$\gamma(t, s) = Cov(X_t, X_s) = E(X_t X_s) - EX_t EX_s. \quad (2.3)$$

(iii) **Funkciju autokorelacija** $\rho : \mathbb{Z}^2 \rightarrow [-1, 1]$ koja $(t, s) \mapsto Corr(X_t, X_s)$ sa

$$\rho(t, s) = Corr(X_t, X_s) = \frac{Cov(X_t, X_s)}{\sqrt{VarX_t VarX_s}} = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t)\gamma(s, s)}}. \quad (2.4)$$

Gore navedene funkcije autokovarijanci i autokorelacija imaju neka korisna svojstva koja prelaze granice ovog rada, ali mogu se pronaći u [3]. Sada možemo definirati **slabo stacionarne procese**.

Definicija 5. Slučajni proces $\{X_t, t \in \mathbb{Z}\}$ je slabo stacionaran ili stacionaran u širem smislu ako vrijedi sljedeće

(i) $EX_t^2 < \infty$ za sve $t \in \mathbb{Z}$,

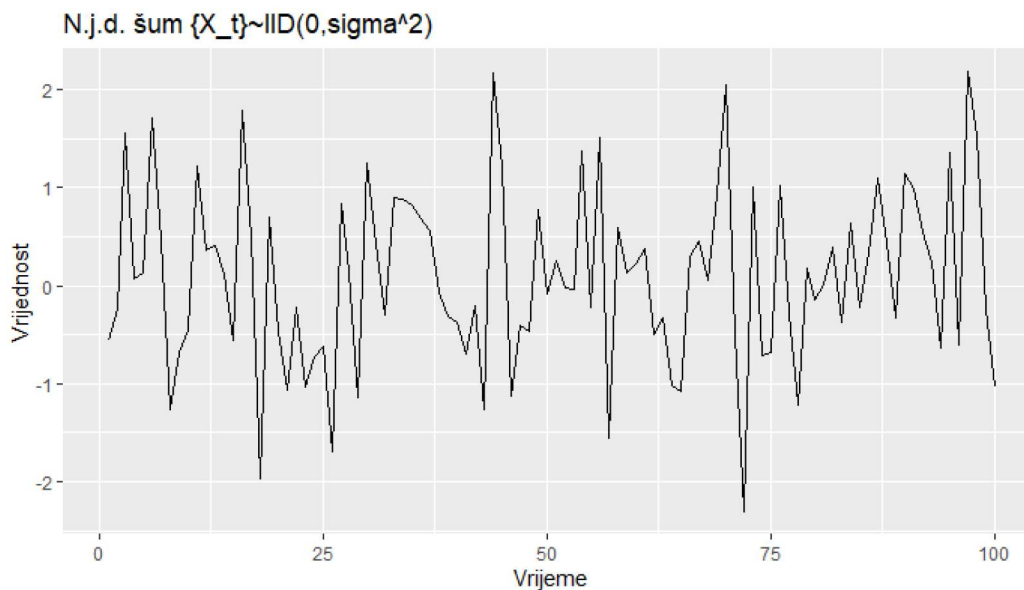
(ii) $\mu(t) = EX_t = c \in \mathbb{R}$ za sve $t \in \mathbb{Z}$,

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

Od sada pa nadalje u radu ćemo pod stacionarnim procesom uvijek misliti na slabo stacionaran proces. Pogledajmo u nastavku neke osnovne primjere modela vremenskih nizova.

N.j.d. šum je niz nezavisnih jednako distribuiranih slučajnih varijabli za koje vrijedi da je $EX_1^2 < \infty$, $EX_1 = 0$, $VarX_1 = \sigma^2$. N.j.d. šum označavamo kao $\{X_t\} \sim \text{IID}(0, \sigma^2)$.

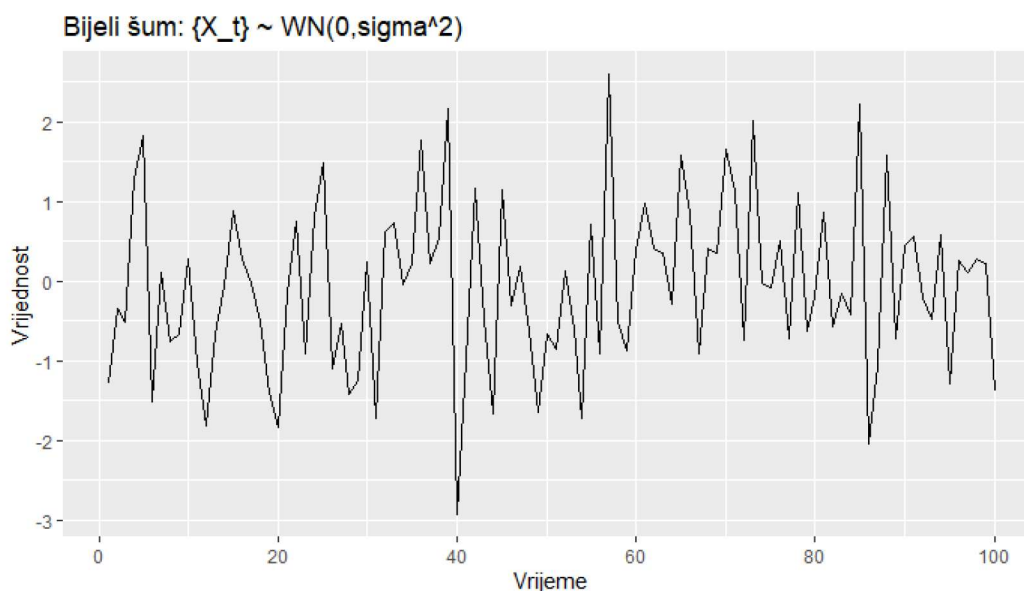
Primjer 1. Kao primjer simulirajmo jedan n.j.d. šum. Pomoću funkcije `rnorm()` generiramo n.j.d. šum $X_t \sim \text{IID}(0, \sigma^2)$ u 100 vremenskih trenutaka. Dobiveni rezultat možemo vidjeti na Slici 2.2.



Slika 2.2: Grafički prikaz n.j.d. šuma.

Bijeli šum je slučajni proces takav da je $EX_t^2 < \infty$, $EX_t = 0$, $\text{Var}X_t = \sigma^2$ za sve $t \in \mathbb{Z}$, te je $\text{Cov}(X_t, X_s) = 0$ za sve $t \neq s$. Pišemo $\{X_t\} \sim \text{WN}(0, \sigma^2)$.

Primjer 2. Simulirajmo sada bijeli šum čije je očekivanje jednako 0, a varijanca σ^2 . Sličnim postupkom, ponovno za 100 vremenskih trenutaka, dobivamo bijeli šum $X_t \sim \text{WN}(0, \sigma^2)$ čiji grafički prikaz vidimo na Slici 2.3.



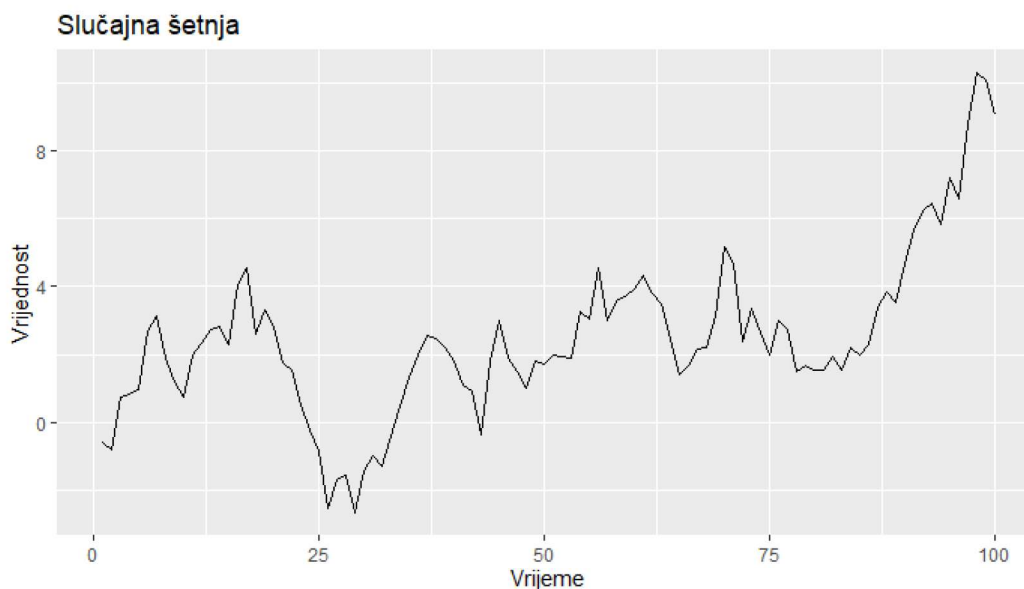
Slika 2.3: Grafički prikaz bijelog šuma.

Slučajna šetnja je proces $\{X_t, t \in \mathbb{N}_0\}$ definiran s $X_0 = 0$ i

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

gdje je $\{Y_t\} \sim \text{IID}(0, \sigma^2)$.

Primjer 3. Malo prije smo vidjeli kako nam za konstrukciju slučajne šetnje treba $\{Y_t\}$ koji je n.j.d. šum. Iskoristimo li n.j.d šum iz Primjera 1 dobivamo slučajnu šetnju koju možemo vidjeti na Slici 2.4.



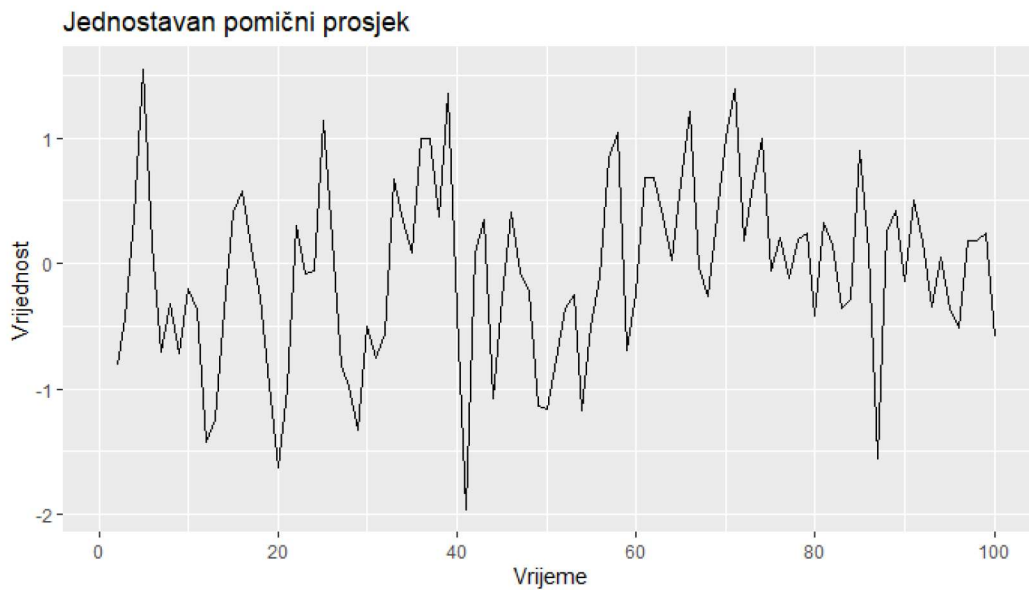
Slika 2.4: Grafički prikaz slučajne šetnje.

Jednostavni pomični prosjek je proces $\{X_t, t \in \mathbb{N}_0\}$ definiran s

$$X_t = \frac{Y_t + Y_{t-1}}{2}$$

gdje je $\{Y_t\} \sim \text{WN}(0, \sigma^2)$.

Primjer 4. Kako bismo simulirali jednostavan pomičan prosjek, iskoristit ćemo bijeli šum iz Primjera 2. Uz pomoć malo prije navedene formule u definiciji, dobivamo jednostavan pomičan prosjek prikazan na Slici 2.5.



Slika 2.5: Grafički prikaz jednostavnog pomičnog prosjeka.

Neke od metoda dekompozicije koje ćemo obraditi u nastavku rada, i njihove izvedenice, često koriste ARIMA modele, tj. procese kako bi poboljšale kvalitetu rezultata. Kako bismo uspješno došli do definicije ARIMA modela i nekih njihovih svojstva, pogledajmo za početak definiciju ARMA procesa.

Definicija 6. *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*

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

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

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z + \dots + \phi_p z^p, \\ \theta(z) &= 1 - \theta_1 z + \dots + \theta_q z^q, \end{aligned}$$

nemaju zajedničkih nultočki. Pišemo $\{X_t\} \sim \text{ARMA}(p, q)$.

ARMA proces pripada klasi **linearnih procesa**, tj. procesa $\{X_t, t \in \mathbb{Z}\}$ koji imaju reprezentaciju

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

gdje je $\{Z_t, t \in \mathbb{Z}\} \sim \text{WN}(0, \sigma^2)$ i $\{\Psi_j, j \in \mathbb{Z}\}$ niz konstanti takav da je $\sum_{j=-\infty}^{\infty} |\Psi_j| < \infty$. Linearne procese možemo jednostavnije zapisati pomoću sljedećeg operatora.

Definicija 7. Za proces $\{X_t, t \in \mathbb{Z}\}$ **operator pomaka unazad** definiran je s

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

tj. općenito za $j \in \mathbb{Z}$ je

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

Sada linearni proces možemo zapisati kao

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

pri čemu operator $\Psi(B) = \sum_{j=-\infty}^{\infty} \Psi_j B^j$ nazivamo **linearni filter**.

Jedno od svojstava ARMA procesa je **kauzalnost** čiju definiciju navodimo u nastavku.

Definicija 8. Za ARMA(p, q) proces $\{X_t\}$ kažemo da je *kauzalan* ako postoji niz konstanti $\{\Psi_j, j \in \mathbb{Z}\}$, takav da je

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

Osim definicije same kauzalnosti, postoji i uvjet koji je iskazan u sljedećem teoremu.

Teorem 1. ARMA(p, q) proces je *kauzalan* ako i samo ako vrijedi

$$\phi(z) \neq 0 \text{ za } |z| \leq 1, z \in \mathbb{C}.$$

Gornji teorem nam govori kako će ARMA proces biti kauzalan ako i samo ako sve nultočke karakterističnog AR polinoma $\phi(z)$ leže izvan jediničnog kruga. Dokaz nećemo navoditi, ali se može pronaći u [3], Theorem 3.1.1., str. 85.

Kao što smo već spomenuli ranije, prilikom promatranja vremenskih nizova možemo uočiti različite uzorke i varijacije. Oni posljedično dovode do nestacionarnosti vremenskog niza koja nije poželjna ukoliko bismo htjeli nastaviti modeliranje i prediktiranje. Iz tog razloga, različitim transformacijama želimo doći do niza koji izgleda kao realizacija nekog stacionarnog procesa. O samim metodama transformacije i prilagodbe vremenskog niza više ćemo reći u nadolazećem poglavlju, no spomenimo sada **diferenciranje** koje će nam trebati kako bismo definirali ARIMA proces, a kojim možemo eliminirati razne oblike nestacionarnosti.

Definicija 9. Za slučajni proces $\{X_t, t \in \mathbb{Z}\}$ **operator diferenciranja** dan je s

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

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

Konačno, navedimo definiciju ARIMA procesa, tj. procesa čijim diferenciranjem dolazimo do nekog ARMA procesa.

Definicija 10. Neka je $d \in \mathbb{N}_0$. Slučajni proces $\{X_t\}$ je **integrirani ARMA proces reda** (p, d, q) ako je

$$Y_t = \Delta^d X_t \tag{2.5}$$

kauzalan ARMA(p, q) proces. Pišemo $X_t \sim \text{ARIMA}(p, d, q)$.

Kao i ranije, iskoristimo li operator pomaka unazad, ARIMA proces možemo jednostavnije zapisati kao $\phi(B)(1 - B)^d X_t = \theta(B)Z_t$, $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. Sam proces $\{X_t\}$ je stacionaran ako i samo ako je $d = 0$, odnosno kada se svede na ARMA(p, q) proces.

3 | Transformacija i prilagodbe vremenskih nizova

Prije no što prijedemo na dekompoziciju vremenskih nizova i različite metode dekompozicije, u ovom poglavlju pozabavit ćemo se njihovom transformacijom i prilagodbom. Naime, vrlo često prilagodba povijesnih podataka može pojednostaviti uzorke uklanjanjem poznatih izvora varijacija. To obično dovodi do preciznijih predviđanja i poboljšanja kvalitete daljnje analize.

3.1 Kalendarske prilagodbe

Kalendarske prilagodbe odnose se na metode kojima uklanjamo **efekt kalendara**. On podrazumijeva varijacije u vremenskom nizu koje su prouzrokovane promjenama u kalendaru. Do njih dolazi u nekoliko slučajeva od kojih svakako valja istaknuti **komponentu pomičnog blagdana**. Pomični blagdani su oni blagdani koji se svake godine održavaju na neki drugi datum. Primjer takvog blagdana je Uskrs. Prema julijanskom kalendaru taj blagdan se obilježava svake godine između 22. ožujka i 25. travnja. Iz tog razloga ukoliko bismo promatrali količine prodanog mladog luka u nekoj trgovini, uočili bismo skok u prodaji ponekad u ožujku, a ponekad u travnju ovisno o datumu Uskrsa.

Još jedan primjer pomičnog blagdana je Lunarna nova godina čiji se datum također mijenja svake godine, ali je uvijek između 21. siječnja i 20. veljače. Osim komponente pomičnog blagdana, možda i važnije je spomenuti **trading day komponentu**. Ona proizlazi iz činjenice da se svake godine mijenja kompozicija kalendara te na taj način dolazi do različitog broja radnih dana po mjesecima u godini. To pak može imati utjecaja na ekonomske aktivnosti što ćemo ilustrirati na sljedećem primjeru.

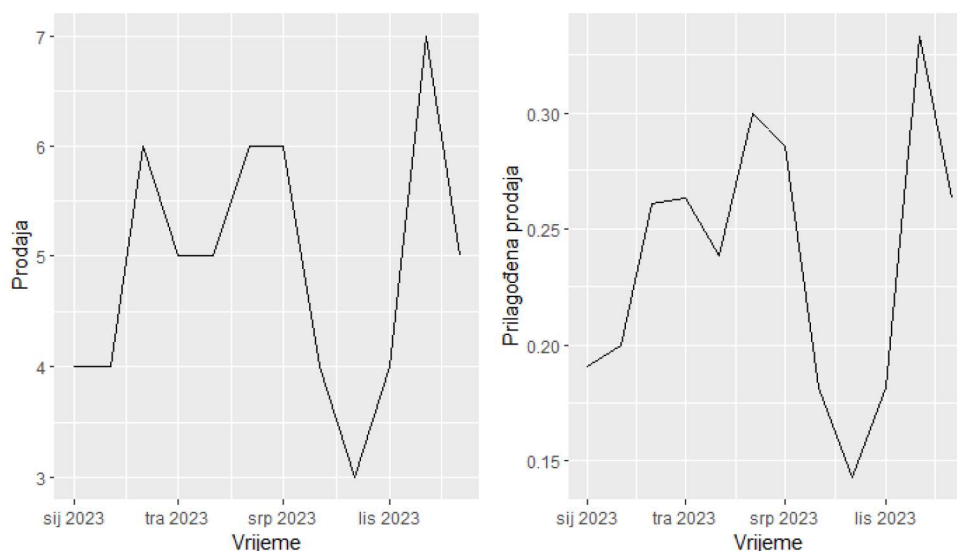
Primjer 5. *Pretpostavimo da promatramo mjesečni broj prodanih automobila u prodavaonici rabljenih vozila za 2023. godinu prikazanu na Slici 3.1 lijevo. Vrlo lako možemo uočiti fluktuaciju prodaje od mjeseca do mjeseca, ali zanima nas je li to utjecalo na poslovanje same prodavaonice. Kako bismo to provjerili moramo ukloniti trading day komponentu, tj. izračunat ćemo omjer broja prodanih vozila i radnih dana u mjesecu.*

U Tablici 3.1 možemo vidjeti broj radnih dana te broj prodanih automobila po mjesecima kao i njihov omjer. Uočimo kako uklanjanjem trading day komponente možemo doći do nešto drugačijeg zaključka o poslovanju promatrane prodavaonice rabljenih vozila. Fokusiramo li se prvo na ožujak i travanj, možemo vidjeti kako je u travnju prodano manje

Mjesec	R. dani	Prodano	Omjer	Mjesec	R. dani	Prodano	Omjer
Siječanj	21	4	0.190	Srpanj	21	6	0.286
Veljača	20	4	0.2	Kolovoz	22	4	0.182
Ožujak	23	6	0.261	Rujan	21	3	0.143
Travanj	19	5	0.263	Listopad	22	4	0.182
Svibanj	21	5	0.238	Studeni	21	7	0.333
Lipanj	20	6	0.3	Prosinac	19	5	0.263

Tablica 3.1: Tablični prikaz broja radnih dana, prodanih automobila i omjera prodanih automobila i radnih dana u 2023. godini.

automobila, što je jasno prikazano padom na Slici 3.1 lijevo. Obratimo li pažnju na broj radnih dana u tim mjesecima pa onda i na izračunati omjer, možemo zaključiti da je prodavaonica rabljenih vozila u travnju zapravo poslovala bolje nego u ožujku što vidimo i na Slici 3.1 desno. U svibnju je pak prodan jednak broj automobila kao i u travnju, no zbog većeg broja radnih dana prodavaonica je poslovala lošije.



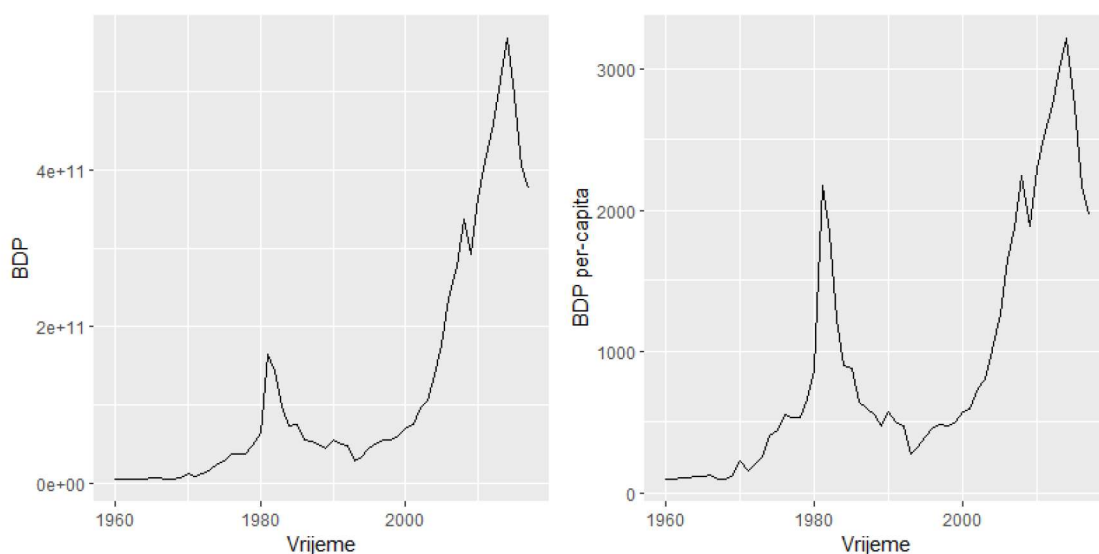
Slika 3.1: Grafički prikaz mjesečne količine prodanih automobila (lijevo) i omjera prodanih automobila i radnih dana (desno).

Isto tako pogledamo li razdoblje lipnja i srpnja možemo uočiti da je prodan jednak broj rabljenih automobila, no zbog većeg broja radnih dana u srpnju omjer nam ukazuje na lošije poslovanje. Dakle, vidimo kako broj radnih dana itekako može utjecati na ekonomsku aktivnost.

3.2 Populacijska prilagodba

Primjer 6. Promotrimo grafički prikaz kretanja bruto domaćeg proizvoda za Nigeriju u razdoblju od 1960. do 2017. na Slici 3.2 lijevo. Početkom 70-tih godina prošlog stoljeća

započinje njegov rast koji svoj vrhunac ima desetak godina kasnije. Nakon toga, do sredine 90-tih BDP pada, a zatim bilježi izrazit rast sve do 2012. kada ponovno počinje opadati. Iako je BDP jedan od vodećih pokazatelja gospodarskog napretka neke zemlje, njegov fokus je ipak na ukupnoj vrijednosti dobara i usluga. Zato se često kao bitniji pokazatelj gospodarske slike ističe BDP per-capita. On služi kao mjerilo prosječnog ekonomskog dohotka po stanovniku te omogućava usporedbu životnog standarda između različitih zemalja što je po nekima puno bitniji faktor. Na Slici 3.2 desno vidimo distribuciju BDP-a per-capita za Nigeriju u istom razdoblju koje smo već naveli. Iako su grafički prikazi zapravo dosta slični, BDP per-capita bilježi snažan rast između 1970. i 1982. kada je gotovo jednak onome iz 2008. Nakon toga trend kretanja BDP-a per-capita gotovo je jednak kretanju BDP-a. Dakle, možemo uočiti da se životni standard stanovništva Nigerije u razdoblju 70-tih drastično poboljšao iako rast BDP-a nije bio toliko izražen.



Slika 3.2: Grafički prikaz kretanja BDP-a (lijevo) i BDP-a per-capita (desno) za Nigeriju u razdoblju od 1960. do 2017.

Gore opisan primjer prikazuje **populacijsku prilagodbu podataka**. Ona se najčešće koristi kada imamo podatke na koje utječu promjene u populaciji. Bit same prilagodbe je prikazati podatke tako da uklonimo utjecaj navedene promjene. Najčešći oblik populacijske prilagodbe je svođenje podataka na per-capita podatke, tj podatke po glavi stanovnika. Osim po glavi stanovnika, podatke također možemo promatrati i na 1000 ili 10000 stanovnika. Ovakvom prilagodbom olakšavamo interpretaciju i daljnju analizu.

3.3 Inflacijska prilagodba

U današnje vrijeme vrlo je teško zaobići tematiku povećanja cijena. Stručnjaci i mediji često rade usporedbe količine proizvoda koju građani mogu kupiti s istim iznosom novca danas i prije nekoliko godina. Zaključak je gotovo uvijek isti, a to je da smo prije, za isti iznos novca, mogli kupiti više, no što možemo danas. Jedan

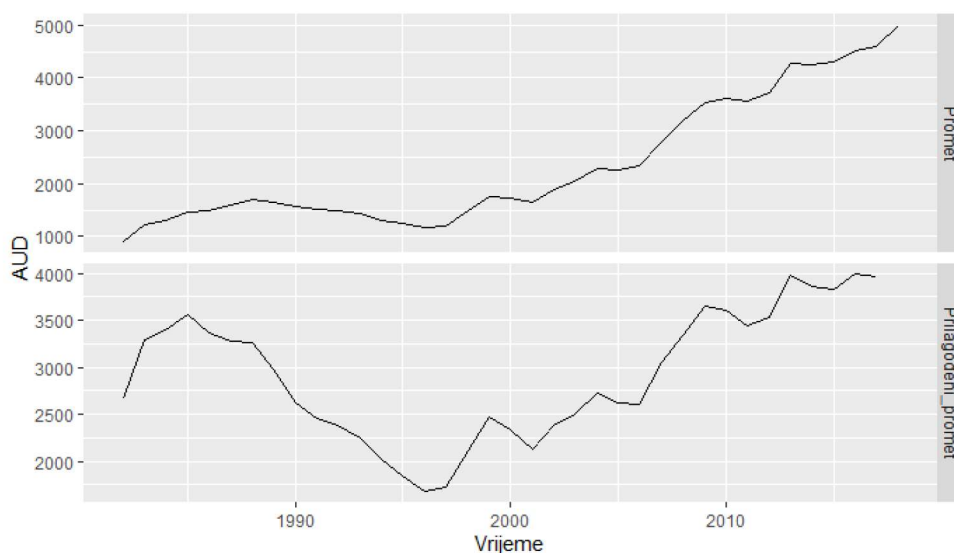
od glavnih uzroka toga je inflacija, tj. kontinuirani rast opće razina cijena. U tom slučaju dolazi do pada kupovne vrijednosti novca pa zato za isti iznos možemo kupiti manje nego što smo mogli prije nekoliko godina.

Prilikom promatranja vremenskih nizova koji sadrže podatke na koje utječe vrijednost novca, prije modeliranja je najbolje iste prilagoditi. To se najčešće radi tako da se sve vrijednosti cijena izraze u novčanoj jedinici, npr. eurima, iz neke godine. Kako bismo to napravili koristimo neki indeks cijena. Najčešće se radi o **indeksu potrošačkih cijena** (eng. *consumer price index*, CPI), no mogu se koristiti i indeks proizvođačkih cijena te deflator BDP-a. Prema [1], CPI indeks odražava fluktuacije cijena proizvoda i usluga korištenih za osobnu potrošnju od strane referentnog urbanog stanovništva unutar određene ekonomije. Pretpostavimo da želimo cijene izraziti u eurima iz 2000. godine. Ako sa i_n označimo cjenovni indeks, a sa p_n originalnu cijenu u n -toj godini, tada

$$x_n = \frac{p_n}{i_n} \cdot i_{2000} \quad (3.1)$$

daje željenu prilagođenu cijenu. Ovakav pristup omogućava nam usporedbu rasta, odnosno pada u industriji. Ilustrirajmo prethodno opisanu metodu na sljedećem primjeru.

Primjer 7. Na Slici 3.3 gore možemo vidjeti promet u maloprodaji (eng. *retail turnover*) za industriju odjeće u australskoj saveznoj državi Victoriji u razdoblju od 1982. do 2017. Vrlo lako uočavamo kontinuirani rast prometa u maloprodaji za odabranu industriju uz povremeni pad. Zanima nas je li ovakvo kretanje prometa uzrokovano isključivo industrijom ili vrijednost novca ima neku ulogu.



Slika 3.3: Grafički prikaz prometa maloprodaje (gore) i prilagođenog prometa maloprodaje (dolje) za industriju odjeće u Victoriji između 1982. i 2017.

Prilagodimo li podatke korištenjem malo prije spomenutog indeksa na Slici 3.3 dolje možemo vidjeti izražen rast sada prilagođenog prometa od 1982. do 1985. Njega prati

snažan pad do sredine 90-tih godina. Nakon toga grafički prikazi su relativno slični, no valja uočiti istaknutije točke rasta i pada u prilagođenom prometu. U konačnici, uočimo da nakon 2010. godine promet bilježi rast što nije slučaj kod prilagođenog prometa. Dakle, vrijednost novca kroz vrijeme zaista može uvelike utjecati na promet unutar neke industrije.

3.4 Transformacije funkcijama

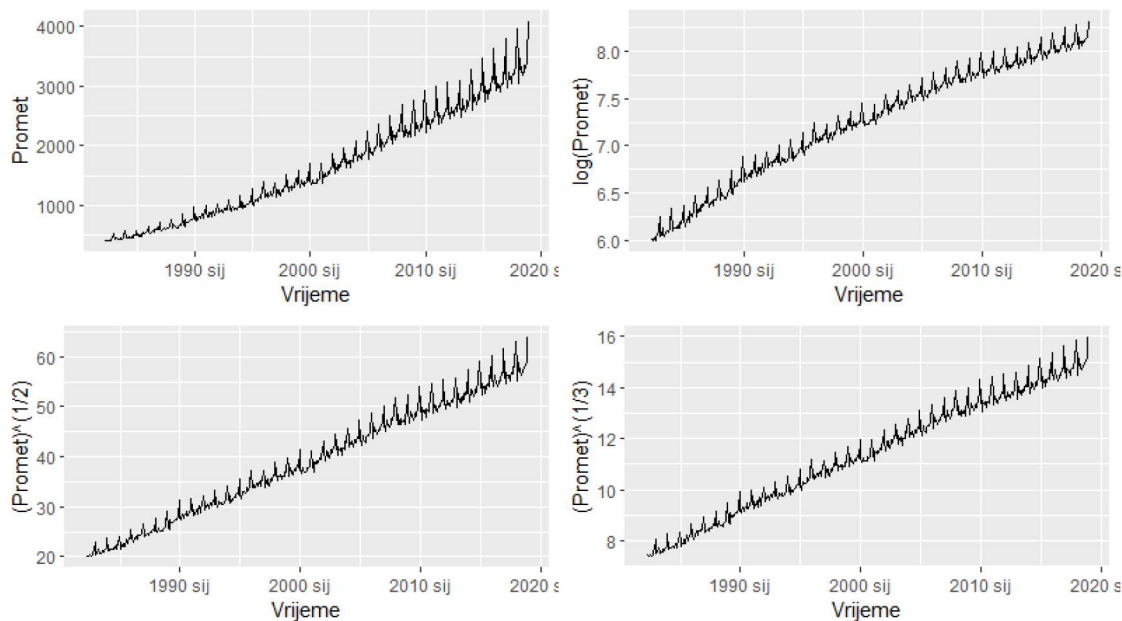
U slučajevima kada imamo nestacionarne vremenske nizove korisno je transformirati podatke kako bismo olakšali daljnje modeliranje. Jedna od najčešćih transformacija je logaritmiranje. Ako sa $x_{t_1}, x_{t_2}, \dots, x_{t_n}$ označimo proizvoljan vremenski niz, tada funkcijom

$$f_{t_i} = \log(x_{t_i}), \quad i = 1, \dots, n \quad (3.2)$$

podatke transformiramo logaritmiranjem. Logaritamska funkcija korisna je za transformacije zato što je lako interpretabilna pa promjene u vrijednosti logaritma odgovaraju relativnoj, tj. postotonoj promjeni na originalnoj skali. Tako primjerice ukoliko transformiramo podatke pomoću logaritma s bazom 10, jedinično povećanje logaritma odgovarati će množenju originalne vrijednosti s 10. Važno je napomenuti da, ukoliko vremenski niz ima jednu ili više komponenti koje su negativne ili jednake nula, ne možemo koristiti logaritamsku transformaciju. Iako nije toliko interpretabilna, još jedna češće korištena transformacija je potenciranje koje je oblika

$$f_{t_i} = x_{t_i}^p, \quad i = 1, \dots, n. \quad (3.3)$$

Primjer 8. Fokusirajmo se na podatke o prometu u maloprodaji za industriju hrane u australskoj saveznoj državi Novi Južni Wales prikazane na Slici 3.4 gore lijevo. Osim izraženog ratućeg trenda, vrlo lako uočavamo i sezonalnost koja, kako vrijeme prolazi, postaje sve izraženija. Provođenjem nekih od malo prije navedenih transformacija podataka možemo vidjeti kako je logaritamska transformacija najbolje stabilizirala varijacije. Kod drugog i trećeg uočavamo stabilizaciju, no podaci s kraja niza i dalje jače variraju od onih s početka niza.



Slika 3.4: Grafički prikaz prometa maloprodaje (gore lijevo) i transformacija prometa maloprodaje (ostalo) za prehrambenu industriju u Novom Južnom Walesu između 1982. i 2017.

Vremenske nizove također možemo transformirati i funkcijom oblika

$$g_{\lambda}(x) = \begin{cases} \frac{x^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \log(x), & \lambda = 0 \end{cases} \quad (3.4)$$

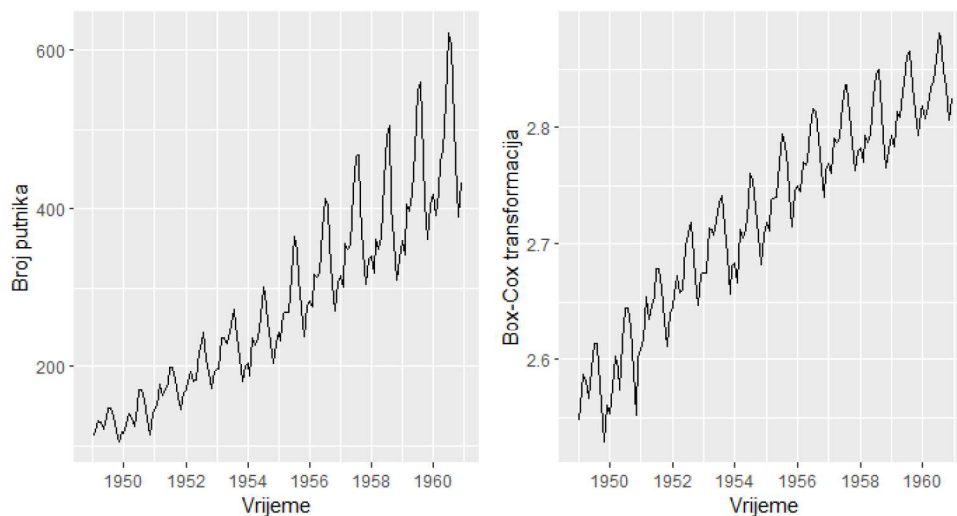
za danu vrijednost $\lambda \in \mathbb{R}$. Ova familija transformacija poznata je i kao Box-Cox transformacije. U prvom izrazu, glavni dio je x^{λ} dok oduzimanje s 1 i dijeljenje s λ osiguravaju neprekidnost familije po λ budući da $g_{\lambda}(x)$ konvergira prema $\log x = g_0(x)$ kada λ konvergira prema nuli. Napomenimo kako za $\lambda = 0$ ova transformacija odgovara već ranije spomenutoj logaritmaskoj transformaciji. Valja istaknuti kako se uz nju još koriste i korijenovanje, tj, $\lambda = \frac{1}{2}$ koja može biti prikladna za brojeće podatke i recipročne vrijednosti kada je $\lambda = -1$. Vrlo lako uočavamo kako se ova transformacija može primijeniti samo za pozitivne podatke, no postoji i modificirana Box-Cox transformacija oblika

$$g_{\lambda}(x) = \begin{cases} \log(x), & \lambda = 0 \\ \frac{\operatorname{sgn}(x)|x|^{\lambda}-1}{\lambda}, & \text{inače} \end{cases} \quad (3.5)$$

koja dozvoljava negativne vrijednosti za x ukoliko je λ pozitivan. Postavlja se pitanje na koji način odabrati najprikladniju λ . Najčešće, prikladna vrijednost je ona koja održava konzistentnu veličinu sezonalnih varijacija kroz čitav vremenski niz čime se pojednostavljuje daljnje prediktiranje.

Primjer 9. Na Slici 3.5 lijevo možemo vidjeti kretanje ukupnog mjesečnog broja zrakoplovnih putnika u Sjedinjenim Američkim Državama od 1949. do 1960. Osim vidljivog

rastućeg trenda, možemo uočiti i sezonalnost koja postaje sve izraženija, pogotovo u ljetnim mjesecima, kako vrijeme prolazi. Kako bismo stabilizirali varijacije, u ovom primjeru, iskoristiti ćemo Box-Cox transformaciju. Za pronalaženje optimalne lambde iskoristili smo funkciju `BoxCox.lambda()`. Sama funkcija temelji se na Box-Cox transformaciji danoj formulom 3.5, tj. varijantom koja dopušta negativne vrijednosti. Kao metodu smo odabrali Guerrero metodu koju je V. M. Guerrero predstavio 1993. godine u radu [11]. Ta metoda odabire λ na način da ona minimizira koeficijent varijacije podnizova polaznog vremenskog niza. Kao optimalan λ dobivamo -0.295 . Prikažemo li ponovno podatke, ali ovaj puta transformirane, na Slici 3.5 desno možemo vidjeti kako ova transformacija vrlo dobro stabilizira varijacije promatranog vremenskog niza te su one sada podjednake.



Slika 3.5: Grafički prikaz kretanja broja zrakoplovnih putnika u SAD-u između 1982. i 2017. (lijevo) i njihova Box-Cox transformacija (desno).

U prethodnom primjeru vidjeli smo kako pomoću Box-Cox transformacije možemo stabilizirati i ujednačiti varijacije u vremenskom nizu. Valja naglasiti da, ukoliko bismo dalje radili predikciju na podacima, radimo sa transformiranim podacima. To je važno za zapamtiti budući da nam treba obratna transformacija kako bismo mogli interpretirati predikciju na početnim podacima. Obrat Box-Cox transformacije dan je formulom

$$g_{\lambda}(x) = \begin{cases} \exp(x), & \lambda = 0 \\ \operatorname{sgn}(\lambda x + 1)|\lambda x + 1|^{\frac{1}{\lambda}}, & \text{inače} \end{cases} \quad (3.6)$$

4 | Dekompozicija i metode dekompozicije vremenskih nizova

Nakon što smo se upoznali s nekim metodama transformacije i prilagodbe vremenskih nizova u ovom poglavlju pozabavit ćemo se samom dekompozicijom vremenskog niza kao i nekim od najpoznatijih metoda za njihovo dekomponiranje.

4.1 Komponente vremenskog niza

Vremenski nizovi mogu pokazivati različite uzorke pa je često korisno podijeliti ih na nekoliko komponenti. U Poglavlju 2 naveli smo tri uzorka koja se pojavljuju kod vremenskih nizova, a to su trend, sezonalnost i cikličnost. Prilikom dekompozicije, radi jednostavnosti, trend i cikličnost često se kombiniraju u jednu komponentu koju zovemo **trend komponenta**.

Uzevši u obzir malo prije napisano, vremenski niz X_t možemo promatrati kao funkciju oblika

$$x_t = f(S_t, T_t, R_t), \quad (4.1)$$

pri čemu x_t predstavlja opažanje vremenskog niza u trenutku t , S_t sezonalnu komponentu, T_t komponentu trenda, a R_t komponentu ostatka u trenutku t pri čemu se t nalazi u skupu vremenskih trenutaka $\{t_1, \dots, t_n\}$. Ukoliko su sezonalne varijacije vremenskog niza relativno konstantne tijekom vremena, prilikom dekompozicije možemo koristiti **aditivnu dekompoziciju**

$$x_t = S_t + T_t + R_t. \quad (4.2)$$

Ako to nije slučaj, tj. ako se sezonalne varijacije s vremenom povećavaju onda koristimo **multiplikativnu dekompoziciju**

$$x_t = S_t \times T_t \times R_t. \quad (4.3)$$

Istaknimo kako su moguće i kombinacije aditivne i multiplikativne dekompozicije poput

$$x_t = S_t \times (1 + T_t) \times (1 + R_t),$$

no one se u praksi rijetko pojavljuju. Također, napomenimo kako je multiplikativna dekompozicija prikladnija ukoliko se radi o ekonomskim podacima koji su prikazani kao vremenski niz. Isto tako korištenje multiplikativne dekompozicije

možemo izbjeći ukoliko iskoristimo aditivnu dekompoziciju. Primjerice, logaritmiramo li vremenski niz vrlo lako prelazimo iz multiplikativne u aditivnu dekompoziciju

$$\log(x_t) = \log(S_t \times T_t \times R_t) = \log S_t + \log T_t + \log R_t. \quad (4.4)$$

Jedna posljedica dekompozicije vremenskog niza je i sezonalna prilagodba podataka do koje dolazi uklanjanjem sezonalne komponente. Za aditivnu dekompoziciju ona je oblika

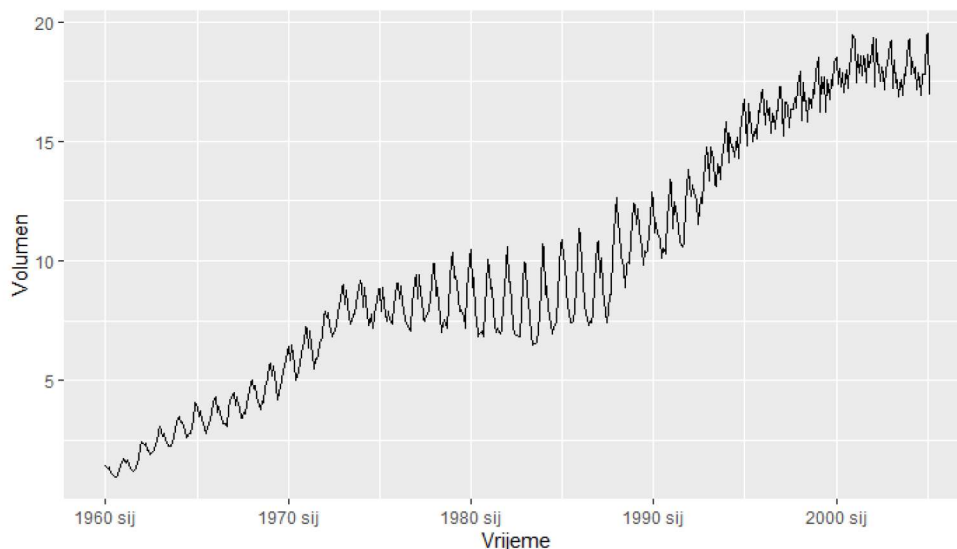
$$x_t - S_t = T_t + R_t,$$

a ukoliko se radi o multiplikativnoj dekompoziciji tada sezonalna prilagodba ima oblik

$$\frac{x_t}{S_t} = T_t \times R_t.$$

Ona može biti posebno korisna ukoliko promatramo podatke o nezaposlenosti kako bi se istaknule varijacije koje ovise o samoj ekonomiji, a ne o sezonalnosti. Na jednostavnom primjeru ilustrirajmo dekompoziciju vremenskog niza na malo prije spomenute komponente.

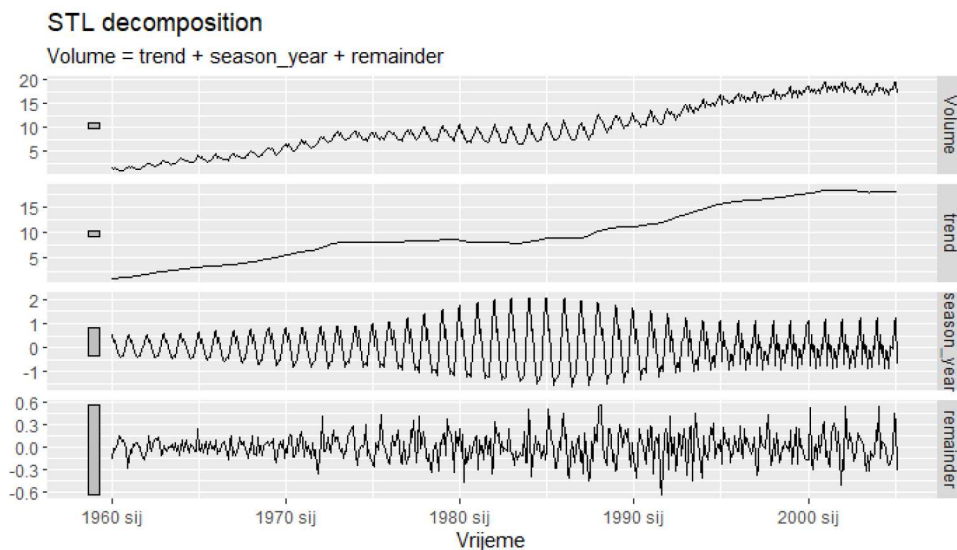
Primjer 10. *Promatramo volumen proizvedenog plina u Kanadi na mjesečnoj bazi za razdoblje od 1960. do 2017. godine prikazan na Slici 4.1. Među podacima vrlo lako možemo uočiti rastući trend i sezonalnost.*



Slika 4.1: Grafički prikaz količine proizvedenog plina u Kanadi u razdoblju od 1960. do 2017. godine

Dalje nas zanima kako bi prethodno opisani vremenski niz izgledao ukoliko ga dekomponiramo na komponente S_t , T_t i R_t . Kako bismo to provjerili iskoristiti ćemo SLT dekompoziciju o kojoj ćemo nešto više reći u jednom od poglavlja koja slijede. Slika 4.2 daje nam grafički pregled dekompozicije promatranog vremenskog niza na malo prije spomenute komponente. Možemo vidjeti kako trend komponenta prati kretanje vremenskog niza, dok se sezonalna komponenta mijenja s vremenom na način da dvije uzastopne godine imaju

sličan uzorak, a udaljenije se podosta razlikuju. Ono što ostane nakon uklanjanja trend komponente i sezonalne komponente, prikazano je u komponenti ostatka.



Slika 4.2: Grafički prikaz SLT dekompozicije podataka o mjesečnoj proizvodnji plina u Kanadi.

4.2 Metode dekompozicije

U ovom potpoglavlju upoznat ćemo se s nekim od najpoznatijih metoda dekompozicije vremenskih nizova. Kao što smo spomenuli u uvodu, prema [13] sama ideja dekompozicije seže još u sedamnaesto stoljeće kada su ju tadašnji astronomi koristili za računanje planetarnih orbita. W. M. Persons je u svojem članku [14] iz 1919. godine prvi postavio pretpostavke o rastavu vremenskih nizova na komponente.

4.2.1 Klasična dekompozicija

Metoda klasične dekompozicije razvila se 20-tih godina prošlog stoljeća te je bila korištena narednih tridesetak godina. Zbog svoje jednostavnosti i danas čini temelj mnogih drugih dekompozicija. Prvi korak u klasičnoj dekompoziciji je procjena trend komponente T_t . To se može napraviti na nekoliko načina, no najčešće se koriste **pomični prosjeci** (eng. *Moving Averages*) ili filteri pomičnog prosjeka.

Pomični prosjeci

Do procjene trend komponente u trenutku t pomoću pomičnog prosjeka dolazimo na način da odredimo prosjek vrijednosti vremenskog niza u k perioda oko t . Za vremenski niz $x_{t_1}, x_{t_2}, \dots, x_{t_n}$ **pomični prosjek reda m** u oznaci m -MA dan je

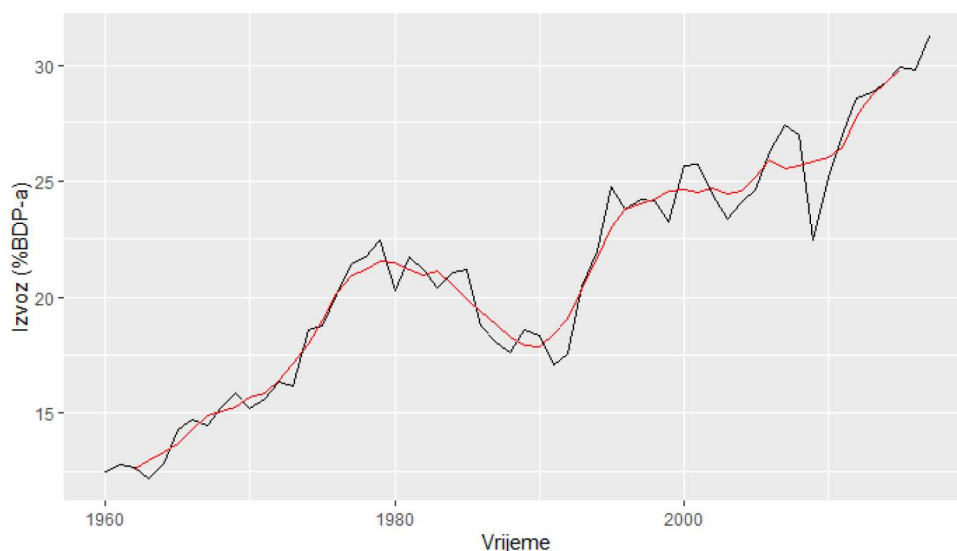
formulom

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k x_{t+j}, \quad t \in \{t_1, \dots, t_n\}, \quad k = \frac{m-1}{2}, \quad (4.5)$$

što ujedno odgovara i procjeni trend komponente. Budući da opažanja u bliskim vremenima uglavnom imaju bliske vrijednosti, njihovo uprosječivanje uklanja određenu razinu slučajnosti u podacima, ostavljajući glatku trend komponentu. Pogledajmo na jednostavnom primjeru kako izgleda korištenje pomičnog prosjeka.

Primjer 11. Promatramo podatke o izvozu, izraženom kao postotak BDP-a, za Italiju. Grafički prikaz kretanja možemo vidjeti na Slici 4.3. Procijenimo trend komponentu pomoću pomičnog prosjeka reda 5. Budući da je $m = 5$ znamo da je $k = 2$, a prema formuli 4.5 procjena trend komponente dana formulom

$$\hat{T}_t = \frac{1}{4}(x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}).$$



Slika 4.3: Grafički prikaz kretanja izvoza za Italiju od 1960. do 2017. godine.

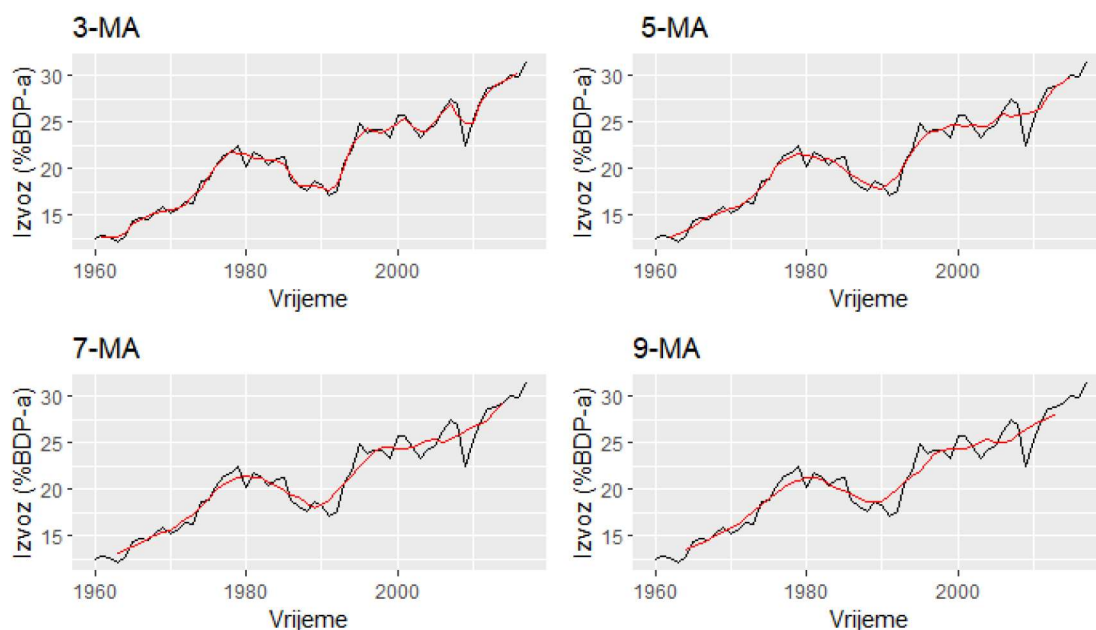
Primijenimo li gornju formulu na podatke redom, dobivamo rezultate koje možemo vidjeti u Tablici 4.1. Na Slici 4.3 možemo vidjeti grafički prikaz procjene trend komponente koja je glađa od originalnog vremenskog niza, ali dobro prati njegovo kretanje

Godina	Izvoz	5-MA	Godina	Izvoz	5-MA
1960.	12.5	NA	⋮	⋮	⋮
1961.	12.8	NA	2014.	29.3	29.3
1962.	12.7	12.6	2015.	29.9	29.8
1963.	12.2	12.9	2016.	29.8	NA
⋮	⋮	⋮	2017.	31.3	NA

Tablica 4.1: Tablični prikaz 5-MA metode na podacima o izvozu za Italiju.

U Tablici 4.1 u prethodnom primjeru lako možemo uočiti kako nemamo vrijednosti pomičnog prosjeka reda 5 za prve dvije i zadnje dvije godine u promatranom intervalu. To se događa zato što za vrijednosti bliske kraju, odnosno početku niza nemamo dovoljno okolnih opservacija kako bismo odredili pomičan prosjek. Isto tako komentirali smo da je krivulja, koja prikazuje procijenjenu trend komponentu, znatno glađa od grafičkog prikaza polaznih podataka, no ipak prati oblik vremenskog niza bez svih fluktuacija. Općenito, veličina reda pomičnog prosjeka utječe na glatkoću krivulje pa tako veći red znači glađu krivulju. Ilustrirajmo to na sljedećem primjeru.

Primjer 12. Za podatke iz Primjera 11 odredimo 3-MA, 7-MA i 9-MA pomične prosjeke.



Slika 4.4: Grafički prikaz raznih pomičnih prosjeka za kretanje izvoza Italije od 1960. do 2017. godine.

Uočavamo spomenuto, tj. s povećanjem reda pomičnog prosjeka, krivulja procijenjene trend komponente postaje glađa i sve manje prati fluktuacije vremenskog niza.

Uočimo kako su svi dosadašnji primjeri sadržavali pomične prosjeke neparnog reda. Razlog tomu je što tada imamo srednju opservaciju u trenutku t kao

i k opservacija s obje strane, odnosno imamo simetričnost. Ukoliko bi m bio paran, što naravno može biti slučaj, simetričnost možemo postići na drugi način. Taj postupak podrazumijeva određivanje pomičnog prosjeka reda n nekog pomičnog prosjeka reda m , tj. $n \times m$ -MA pomičan prosjek.

Primjer 13. *Pretpostavimo da želimo procijeniti trend komponentu proizvoljnog vremenskog niza x_{t_1}, \dots, x_{t_n} 4-MA pomičnim prosjekom. To možemo napraviti na dva načina korištenjem formule 4.5:*

$$\hat{T}_t^{(1)} = \frac{1}{4}(x_{t-2} + x_{t-1} + x_t + x_{t+1}) \text{ ili } \hat{T}_t^{(2)} = \frac{1}{4}(x_{t-1} + x_t + x_{t+1} + x_{t+2}).$$

Kao što možemo vidjeti nemamo simetričnost koju smo imali kod pomičnih prosjeka neparnog reda. Kako bismo ju ipak postigli, napraviti ćemo pomičan prosjek reda 2 dviju varijanti pomičnog prosjeka reda 4. Dobivamo

$$\begin{aligned} \hat{T}_t &= \frac{1}{2} \left[\frac{1}{4}(x_{t-2} + x_{t-1} + x_t + x_{t+1}) + \frac{1}{4}(x_{t-1} + x_t + x_{t+1} + x_{t+2}) \right] \\ &= \frac{1}{8}x_{t-2} + \frac{1}{4}x_{t-1} + \frac{1}{4}x_t + \frac{1}{4}x_{t+1} + \frac{1}{8}x_{t+2}. \end{aligned}$$

U konačnici smo dobili simetričnost, a krajnji rezultat sličan je 5-MA pomičnom prosjeku kojeg smo vidjeli u Primjeru 11

Općenito, određivanje $n \times m$ -MA pomičnog prosjeka ekvivalentno je **težinskom pomičnom prosjeku reda m** koji je dan formulom

$$\hat{T}_t = \sum_{j=-k}^k a_j x_{t+j}, \quad t \in t_1, \dots, t_n, \quad (4.6)$$

gdje je $k = \frac{m-1}{2}$, a $[a_{-k}, \dots, a_k]$ predstavlja težine čija suma mora biti jednaka 1 i imaju svojstvo da je $a_{-j} = a_j$. Kada 2-MA pomičan prosjek prati pomičan prosjek parnog reda nazivamo ga **centrirani pomični prosjek reda m** . Važno je napomenuti da su druge kombinacije pomičnih prosjeka također moguće. Tako recimo pomičan prosjek reda 3 može pratiti još jedan pomičan prosjek reda 3. Nakon što smo se upoznali sa pomičnim prosjecima u nastavku navodimo algoritme koji se nalaze u pozadini dvaju oblika klasične dekompozicije.

Aditivna dekompozicija

1. *Procjena trend komponente.* Odredimo procjenu trend komponente \hat{T}_t u trenutku t pomoću pomičnih prosjeka. Ukoliko je m paran broj koristimo $2 \times m$ -MA metodu, a ukoliko je m neparan koristimo m -MA metodu.
2. *Detrendiranje.* Odredimo vremenski niz bez trenda korištenjem formule

$$x_t - \hat{T}_t, \quad t \in \{t_1, \dots, t_n\}.$$

3. *Procjena sezonalne komponente.* Kako bismo došli do procjene sezonalne komponente \hat{S}_t moramo doći do sezonalnih indeksa. Cilj je odrediti prosjek detrendiranih vrijednosti za sezonalni period. Pretpostavimo primjerice da promatramo vremenski niz koji sadrži podatke o prodaji televizora kroz n godina. Podaci su prikupljeni mjesečno te neka je sezonalni period jednak m . Tada bismo do sezonalnog indeksa za kolovoz došli pomoću formule

$$\bar{S}_8 = \frac{1}{n} \sum_{i=1}^n x_{8_i},$$

tj. napravimo prosjek svih poznatih vrijednosti za kolovoz u promatranom vremenskom nizu. Istaknimo kako nam u gornjoj formuli x_{8_i} predstavlja podatak za kolovoz u i -toj godini. Ponovimo li isti postupak za preostale mjeseci u godini dolazimo do sezonalnih indeksa oblika \bar{S}_t , $t \in 1, \dots, m$. Još je potrebno osigurati da sezonalni indeksi u sumi daju 0. Ukoliko taj uvjet ne vrijedi nakon što odredimo sve indekse, oni se normaliziraju na sljedeći način: izračunamo prosjek svih sezonalnih indeksa kao

$$w = \frac{1}{m} \sum_{t=1}^m \bar{S}_t$$

te ga oduzmemo od samog sezonalnog indeksa, tj. imamo

$$\bar{S}_t - w, t \in \{1, \dots, m\}.$$

Povezivanjem svih vrijednosti sezonalnih indeksa koji u sumi daju 0 dolazimo do tražene procjene \hat{S}_t , $t \in \{1, \dots, n\}$.

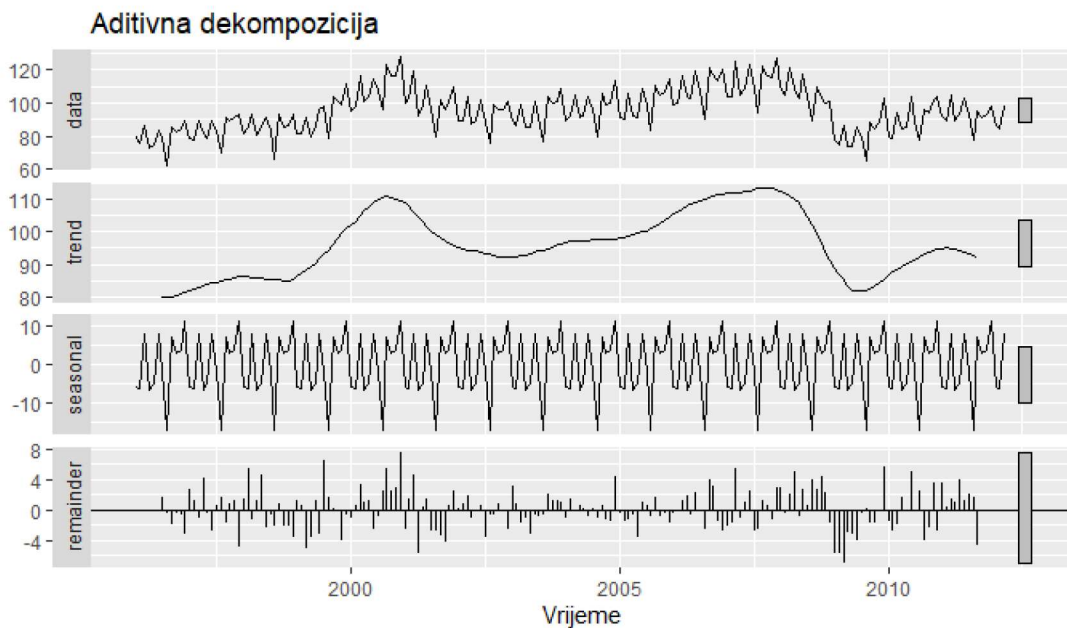
4. *Procjena komponente ostatka.* U konačnici, odredimo procjenu komponente ostatka \hat{R}_t oduzimanjem procijenjenih komponenti trenda i sezonalnosti, odnosno

$$\hat{R}_t = x_t - \hat{T}_t - \hat{S}_t, t \in \{t_1, \dots, t_n\}.$$

Napomenimo još kako prethodni proces primjenjujemo za svaku godinu u vremenskom nizu.

Pogledajmo sada kako aditivna dekompozicija izgleda na praktičnom primjeru.

Primjer 14. *Promatramo podatke o mjesečnoj proizvodnji električnih uređaja na području Europe (17 zemalja) u razdoblju od 1996. do 2012. Želimo rastaviti niz na komponente pomoću aditivne dekompozicije. Koristimo funkciju `decompose()` kojoj prosljedimo originalni vremenski niz te specificiramo željeni oblik dekompozicije. Kao rezultat dobivamo rastav niza prikazan na Slici 4.5.*



Slika 4.5: Grafički prikaz aditivne dekompozicije danog vremenskog niza.

Multiplikativna dekompozicija

Slično kao i malo prije, u nastavku navodimo algoritam multiplikativne dekompozicije.

1. *Procjena trend komponente.* Odredimo procjenu trend komponente \hat{T}_t pomoću pomičnih prosjeka. Ukoliko je m paran broj koristimo $2 \times m$ -MA metodu, a ukoliko je m neparan koristimo m -MA metodu.
2. *Detrendiranje.* Odredimo vremenski niz bez trenda korištenjem formule

$$\frac{x_t}{\hat{T}_t}$$

3. *Procjena sezonalne komponente.* Postupak procjene sezonalne komponente jednak je procesu koji smo vidjeli malo prije u algoritmu aditivne dekompozicije. Razlika je u tome što sada moramo osigurati da sezonalni indeksi \bar{S}_t , $t \in 1, \dots, m$, u sumi daju m . Ukoliko taj uvjet ne vrijedi nakon što odredimo sve indekse, oni se normaliziraju na sljedeći način: izračunamo prosjek svih sezonalnih indeksa kao

$$w = \frac{1}{m} \sum_{t=1}^m \bar{S}_t$$

te sezonalni indeks podijelimo njime, tj. imamo

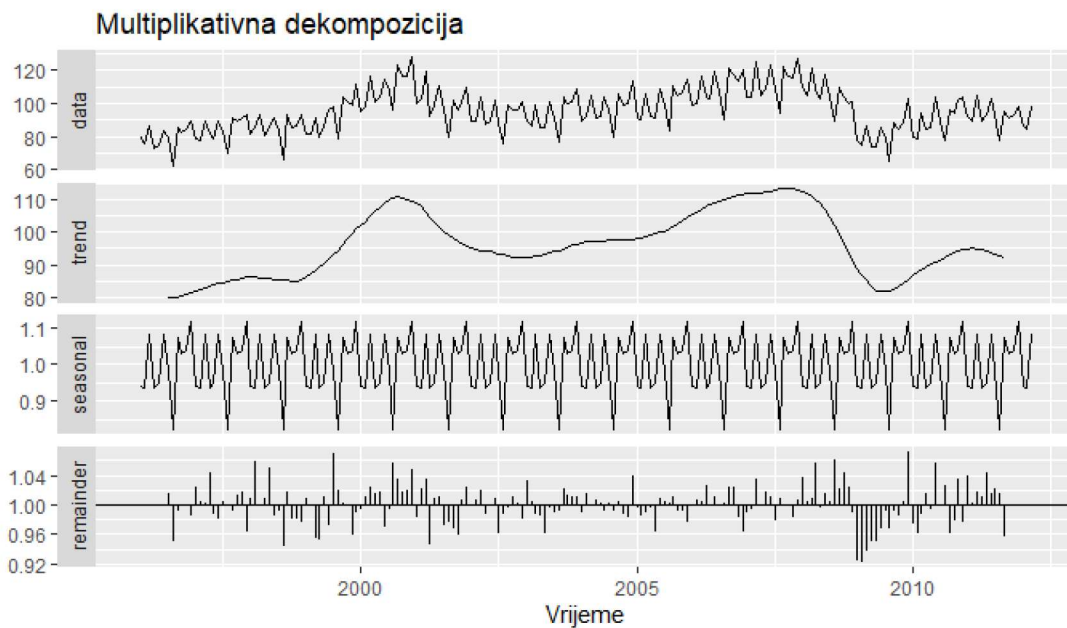
$$\frac{\bar{S}_t}{w}, t \in \{1, \dots, m\}.$$

Kao i ranije, povezivanjem svih vrijednosti sezonalnih indeksa koji u sumi daju 0 dolazimo do tražene procjene \hat{S}_t , $t \in \{1, \dots, n\}$.

4. *Procjena komponente ostatka.* U konačnici, odredimo procjenu komponente ostatka \hat{R}_t oduzimanjem procijenjenih komponenti trenda i sezonalnosti, odnosno

$$\hat{R}_t = \frac{x_t}{\hat{T}_t \cdot \hat{S}_t}.$$

Primjer 15. Na istim podacima kao u Primjeru 14 napravimo multiplikativnu dekompoziciju. Korištenjem iste funkcije dobivamo rezultat grafički prikazan na Slici 4.6. Ponovno, kao i u prethodnom primjeru, trend komponenta dobro prati vremenski niz, a sezonalna komponenta je ujednačena. Manje razlike možemo vidjeti u komponenti ostatka što ćemo prokomentirati u nastavku.



Slika 4.6: Grafički prikaz multiplikativne dekompozicije podataka o proizvodnji elektroničke opreme u Europi.

Usporedimo li obje dekompozicije u prethodnim primjerima možemo uočiti kako su rezultati relativno slični. Razliku možemo uočiti u komponenti ostatka fokusiramo li se na period 2009. godine. Kod multiplikativne dekompozicije vidimo nešto veći pad vrijednosti, ali u oba slučaja taj dio dobro prati trend komponentu. To nam pak može ukazati na lošiju procjenu trend komponente koja se onda "prelila" i na komponentu ostatka.

4.2.2 Metode korištene od strane službenih statističkih agencija

Kao što smo već spomenuli u uvodu prethodnog potpoglavlja, klasična dekompozicija je polako počela izlaziti iz uporabe 50-tih godina prošlog stoljeća. Naime, 1957. godine, američki Ured za popis stanovništva (*eng.* US Census Bureau) predstavio je novu metodu nazvanu **Census II**. Ona je poslužila kao osnova za nadolazeće metode poput **X-11 dekompozicije**, koju su 1965. godine predstavili Sishkin, Young i Musgrave u članku [15], i njenih varijanti koje uključuju X-11-ARIMA,

X-12-ARIMA i X-13-ARIMA-SEATS metode. Osamdesetih godina uvodi se **STL dekompozicija** o kojoj ćemo nešto više reći u idućem potpoglavlju, a razdoblje devedesetih, posebice u Europi, obilježila je pojava **TRAMO-SEATS dekompozicije**.

Budući da postoji mnoštvo različitih metoda dekompozicije vremenskih nizova, među statističkim agencijama nema ustaljene metode koja se koristi. Tako primjerice australski Zavod za statistiku koristi X-12-ARIMA metodu dok, već spomenuti, američki Ured za popis stanovništva koristi X-12ARIMA-SEATS metodu. EuroStat koristi istu metodu dok se britanski Ured za nacionalnu statistiku također opredijelio za X-12-ARIMA metodu. U nastavku ćemo pobliže opisati neke od malo prije spomenutih metoda kao i praktične primjere njihove uporabe.

X-11 metoda dekompozicije i izvedenice

X-11 metoda dekompozicije, poznata i kao Census X-11 metoda, pretpostavlja da glavne komponente vremenskog niza prate multiplikativan ili aditivan model dekompozicije, a njihove kombinacije poput $X_t = T_t \times S_t + R_t$ nisu moguće. Istaknuli smo kako kod klasične dekompozicije nemamo procjenu trend komponente u rubnim točkama što ovdje nije slučaj. Naime, X-11 dekompozicija taj problem rješava korištenjem jednostranih filtera koji se primjenjuju uzastopno, a cijeli proces se ponavlja dvaput. Isto tako ova metoda uzima u obzir trading day komponentu kao i komponentu pomičnog blagdana koje procjenjuje pomoću modela linearne regresije. Algoritam X-11 dekompozicije je sljedeći:

1. *Procjena trend komponente.* Trend komponentu procijenimo pomoću $2 \times m$ -MA metode pri čemu je $m = 12$, tj.

$$\hat{T}_t^{(1)} = M_{2 \times 12}(x_t).$$

2. *Procjena zajedničke sezonalne komponente i komponente ostatka.* Sezonalnu komponentu i komponentu ostatka procijenimo detrendiranjem, tj.

$$(\widehat{S_t + R_t})^{(1)} = x_t - \hat{T}_t^{(1)}.$$

3. *Procjena sezonalne komponente.* Do procjene sezonalne komponente dolazimo na sličan način kao kod aditivne i multiplikativne dekompozicije. Za početak iskoristimo 3×3 -MA metodu kako bismo procijenili sezonalni indeks za svaki mjesec iz zajedničke procjene sezonalne komponente i komponente trenda. Dobivamo sezonalni indeks oblika

$$\tilde{S}_t^{(1)} = M_{3 \times 3} \left[(\widehat{S_t + R_t})^{(1)} \right].$$

Gornji izraz se potom normalizira tako da suma sezonalnih indeksa kroz period od 12 mjeseci bude blizak nuli. Kao rezultat dobivamo procjenu sezonalne komponente danu s

$$\hat{S}_t^{(1)} = \tilde{S}_t^{(1)} - M_{2 \times 12}(\tilde{S}_t^{(1)}).$$

4. *Procjena sezonalno prilagođenog niza.* Napravimo procjenu sezonalno prilagođenog vremenskog niza na način

$$\hat{A}_t^{(1)} = x_t - \hat{S}_t^{(1)}.$$

5. *Procjena trend komponente.* Ponovno procjenjujemo trend komponentu, no ovaj puta koristeći Hendersonov pomični prosjek s 13 članova. Imamo

$$\hat{T}_t^{(2)} = H_{13}(A_t^{(1)}).$$

6. *Procjena zajedničke sezonalne komponente i komponente ostatka.* Sezonalna komponenta i komponenta ostatka sada se procjenjuju kao

$$(\widehat{S_t + R_t})^{(2)} = x_t - \hat{T}_t^{(2)}.$$

7. *Procjena sezonalne komponente.* Slično kao u trećem koraku, ovaj puta korištenjem 3×5 -MA metode procjenjujemo sezonalni indeks za svaki mjesec na temelju zajedničke komponente ostatka i sezonalne komponente kao

$$\tilde{S}_t^{(2)} = M_{3 \times 3} \left[(\widehat{S_t + R_t})^{(2)} \right].$$

Normalizacijom kao i ranije dolazimo do procjene sezonalne komponente koja je dana s

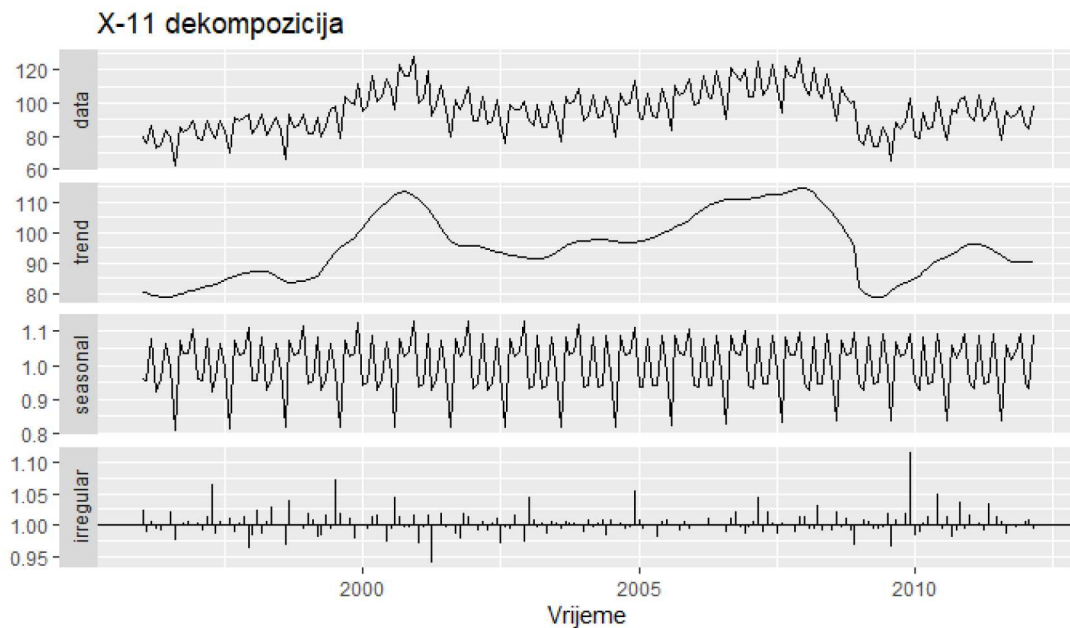
$$\hat{S}_t^{(2)} = \tilde{S}_t^{(2)} - M_{2 \times 12}(\tilde{S}_t^{(2)}).$$

8. *Procjena sezonalno prilagođenog niza.* U konačnici, ponovno napravimo procjenu sezonalno prilagođenog vremenskog niza

$$A_t^{(2)} = (T_t + R_t)^{(2)} = x_t - \hat{S}_t^{(2)}.$$

U 5. koraku gornjeg algoritma iskoristili smo Hendersonov pomičan prosjek. Takva metoda pomičnih prosjeka koristi se u X-11 metodi kako bismo uspješno izdvojili trend komponentu iz procjene sezonalno prilagođenog vremenskog niza. O samoj metodi više se može saznati u [13], str. 36 – 37. Pomični prosjeci vrlo lako reagiraju na ekstremne vrijednosti pa stoga X-11 dekompozicija uključuje alat za detekciju i promjenu takvih vrijednosti. On se koristi kako bi se vremenski niz "očistio" prije sezonalne prilagodbe. Isto tako ranije spomenute trading day komponenta i komponenta pomičnog blagdana mogu uzrokovati različite varijacije u vremenskom nizu. Takve komponente se u X-11 dekompoziciji procjenjuju uz pomoć linearne regresije koja ima temelje u komponenti ostatka o čemu se više može saznati u [13], str. 19. Pogledajmo sada primjenu opisane dekompoziciju u stvarnosti.

Primjer 16. *Nastavljamo s promatranjem podataka o proizvodnji električnih uređaja na području Europe. Nakon što smo vidjeli aditivnu i multiplikativnu dekompoziciju na spomenutom nizu, napravimo X-11 dekompoziciju.*



Slika 4.7: Grafički prikaz X-11 dekompozicije podataka o proizvodnji elektroničke opreme u Europi.

Možemo uočiti kako X-11 metoda puno bolje reagira na vrijednosti iz 2009. godine u odnosu na multiplikativnu i aditivnu dekompoziciju. Nemamo veliki pad u vrijednostima kao što smo imali ranije, a velika promjena lako se uočava i u komponenti ostatka.

Iako je X-11 metoda donjela mnoga poboljšanja u odnosu na prethodnu klasičnu dekompoziciju i ona je imala neke nedostatke. Ova metoda dizajnirana je isključivo za mjesečne i kvartalne podatke. Također sama metoda nije temeljena na nekom modelu što onemogućava konstrukciju pouzdanih intervala.

Rješenje nekih problema dala je **X-11-ARIMA** metoda koju je predstavila B. E. Dagum u svojim člancima [7] i [8]. Kao što joj ime govori, ova dekompozicija temelji se na ARIMA procesu čiju uporabu su svojom knjigom [2] popularizirali Box i Jenkins. To je omogućilo opisivanje stohastičkog ponašanja samog vremenskog niza te računanje njegovog očekivanja i varijance. Isto tako bilo je moguće odrediti pouzdane intervale, a to je pak dovelo do bolje detekcije stršućih vrijednosti na krajevima vremenskog niza što je isticano kao jedna od glavnih mana originalne X-11 dekompozicije.

Nekoliko godina kasnije, Findley i ostali su u članku [9] predstavili **X-12-ARIMA** metodu dekompozicije kao obogaćenu verziju X-11-ARIMA metode. Ona nudi iste mogućnosti kao i prethodna metoda uz dodatak novih opcija kod odabira sezonalnih filtera, nove tablice za trading day komponentu itd. Isto tako, sama metoda se temelji na regARIMA modelu oblika

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D \left(x_t - \sum_{i=0}^r \beta_i y_{it} \right) = \theta(B)\Theta(B^s)Z_t, \quad t \in \mathbb{Z}, \quad (4.7)$$

gdje je za $\{X_t\}$, u općenitom obliku,

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t, \quad t \in \{t_1, \dots, t_n\}$$

sezonalni ARIMA model reda $(p, d, q) \times (P, D, Q)_s$ pri čemu $s \in \mathbb{N}$ označava period. Komad $x_t - \sum_{i=0}^r \beta_i y_{it}$ označava greške z_t linearnog regresijskog modela za vremenski niz koji je dan formulom

$$x_t = \sum_{i=0}^r \beta_i y_{it} + z_t, \quad i = 1, \dots, r, \quad t \in \{t_1, \dots, t_n\}.$$

Dakle, regARIMA model umjesto komponenti vremenskog niza, modelira greške regresijskog modela. Više detalja o konstrukciji regARIMA modela može se pronaći u [6], str. 97.

TRAMO-SEATS dekompozicija

Nakon uključivanja ARIMA modela u dekompoziciju vremenskih nizova razvoj nije stao. P. Burman je u svojem članku [4] prvi razvio metodu sezonalne prilagodbe temeljenu na ARIMA modelu nazvanu SIGEX. Hillmer i Tiao nastavili su daljnji razvoj predstavljanjem sezonalne prilagodbe temeljene na ARIMA modelu, a Gomez i Maravall su u svojim radnim materijalima [10] predstavili **TRAMO-SEATS dekompoziciju** i pripadni software koji je među najkorištenijima u europskim statističkim agencijama i koji ćemo detaljnije opisati u nastavku.

Sama metoda može se podijeliti u dva dijela. Prvo se determinističke komponente poput trading day komponente, komponente pomičnih blagdana, stršećih vrijednosti itd. procjenjuju pomoću TRAMO metode koja za to koristi regresiju, a same komponente se zatim uklanjaju iz početnih podataka. U drugom koraku SEATS metodom procjenjujemo sezonalnu komponentu i komponentu trenda iz ARIMA modela koji je prilagođen na podatke iz kojih su uklonjene determinističke komponente.

TRAMO (*eng. Time Series Regression with ARIMA Noise, Missing Observations, and Outliers*) je regresijska metoda koja obavlja predviđanje, procjene i interpolaciju nedostajućih vrijednosti i grešaka ARIMA modela uz prisutstvo nekoliko tipova stršećih vrijednosti. Ukoliko sa $\mathbf{y} = (y_1, \dots, y_n)$ označimo opservacije, tada program prilagođava funkciju tj. "fita" model, u ovom slučaju regresijski kao

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + v_t,$$

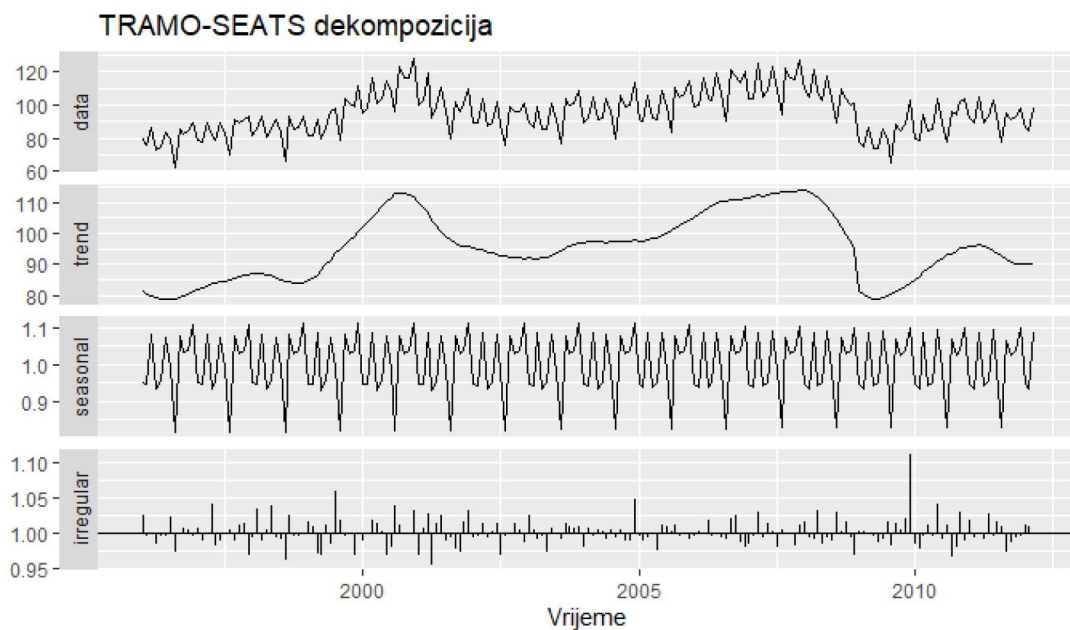
gdje je $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ vektor regresijskih koeficijenata, a $\mathbf{x}_t = (x_{1t}, \dots, x_{pt})$ p regresora kojima definiramo deterministički dio modela. S druge strane, v_t predstavlja stohastički dio modela za koji pretpostavljamo da prati ARIMA model dan s 2.5. Regresore \mathbf{x}_t može generirati program, ali i sam korisnik pri čemu ukoliko program definira regresore oni odgovaraju trading day komponentama, komponentama pomičnog blagdana (prvenstveno Uskrsu) ili stršećim vrijednostima. Varijacije uzrokovane trading day komponentom su povezane s kalendarom određene godine te se koristi šest varijabli kako bi se one modelirale. Efekt pomičnog blagdana, tj. u ovom slučaju **efekt Uskrsa** modelira konstantnu promjenu u razini dnevne aktivnosti d dana prije Uskrsa, a tipična vrijednost varira između tri i šest. Stršeće vrijednosti odražavaju utjecaj nekih posebnih ponavljajućih događaja poput uvođenja nove regulacije, značajnih ekonomskih ili političkih promjena. U nastavku navodimo skraćeni oblik TRAMO metode dok se detaljan opis svih koraka može pronaći u [6], str. 118 – 121.

1. *Predtest za log/level specifikaciju* u kojem različitim testovima pokušavamo utvrditi hoće li se originalni podaci transformirati logaritmiranjem ili će ostati na isti.
2. *Predtest za trading day efekt i efekt Uskrsa*. Koristi se regresija na početnom modelu te se promatra jesu li spomenuti efekti značajni ili ne. Valja napomenuti da predtest ponekad može dati rezultat da nam neki efekt nije značajan, ali on na kraju uđe konačnu procjenu modela i obratno.
3. *Automatska detekcija i korekcija stršućih vrijednosti*. Sam program ima mogućnost pronalaska stršućih vrijednosti i uklanjanja njihovih vrijednosti iako ih i korisnik može unijeti manualno.
4. *Automatski odabir ARIMA modela od strane programa*.
5. *Procjena regARIMA modela*.
6. *Dijagnostika reziduala*. Analiziramo jesu li procijenjeni reziduali \hat{a}_t modela međusobno nezavisni te jesu li reziduali a_t normalno distribuirani sa očekivanjem 0 i varijancom σ_a^2 .
7. *Predviđanje*. Ukoliko su zadovoljene sve stavke dijagnostike, model se koristi za određivanje optimalnog predviđanja vremenskog niza.

SEATS metoda (*eng. Signal Extraction in ARIMA Time Series*) pripada klasi procedura temeljenih na ARIMA modelima za dekompoziciju vremenskih nizova na sezonalnu komponentu, trend komponentu i komponentu ostatka. Ponovno, navodimo skraćenu verziju algoritma, a detalji se mogu pronaći u [6], str.121 – 126.

1. *Procjena ARIMA modela*. SEATS metoda započinje prilagodbom modela na podatke koji nemaju utjecaj determinističkih komponenti.
2. *Izvod ARIMA modela za svaku od komponenti*. Program nastavlja s dekompozicijom niza u nekoliko komponenti, a ona može biti multiplikativna ili aditivna.
3. *Procjena komponenti*. Program koristi minimalnu srednju kvadratnu grešku za procjenu komponenti.
4. *Dijagnostika*.

Primjer 17. *Kao i ranije, na podacima o proizvodnji elektroničke opreme u Europi provest ćemo TRAMO-SEATS dekompoziciju. Rastav niza po komponentama možemo vidjeti na Slici 4.8. Sam rezultat sličan je onome koji smo imali prilikom korištenja X-11 dekompozicije.*



Slika 4.8: Grafički prikaz TRAMO-SEATS dekompozicije podataka o proizvodnji elektroničke opreme u Europi.

4.2.3 STL dekompozicija

STL dekompozicija (*Seasonal and Trend decomposition using Loess*) je raznovrsna i robusna procedura koja dekomponira vremenski niz na trend komponentu, sezonalnu komponentu i komponentu ostatka. Razvijena je od strane R. B. Clevelanda, W. S. Clevelanda, J.E. McRea i I. Terpenning devedesetih godina prošlog stoljeća. Neki od ciljeva izgradnje same metode bili su jednostavniji dizajn i korištenje, mogućnost dekompozicije vremenskih nizova s nedostajućim vrijednostima te jednostavna implementacija i brz proces računanja za duge nizove.

Sama metoda ima nekoliko parametara koje korisnik mora sam birati pri čemu neki mogu biti a priori definirani dok drugi ovise o samoj proceduri. Dekompozicija se sastoji od niza operacija zaglađivanja koje sve, osim jedne, kao zaglađivač koriste lokalnu linearnu regresiju (dalje LOESS regresija).

LOESS regersija

Označimo li sa $x_i, y_i, i = 1, \dots, n$ vrijednosti nezavisne, odnosno zavisne varijable redom, tada krivulja LOESS regresije $\hat{g}(x)$ predstavlja zaglađenje od y za dani x . Valja naglasiti kako je LOESS definirana za svaki x , a ne samo za x_i što omogućava STL dekompoziciji jednostavnije rješavanje problema nedostajućih vrijednosti.

Krivulju $\hat{g}(x)$ određujemo na sljedeći način. Za početak, neka je q pozitivan cijeli broj te pretpostavimo da je $q \leq n$. Odabiremo q vrijednosti od x_i koje su najbliže x -u te svakoj dajemo **susjedsku težinu** (eng. *neighbourhood weight*) koja ovisi o udaljenosti od x . Sa $\lambda_q(x)$ označimo udaljenost q -tog najudaljenijeg x_i od

x te neka je W težinska funkcija oblika

$$W(u) = \begin{cases} (1 - u^3)^3, & 0 \leq u < 1 \\ 0, & u \geq 1 \end{cases}.$$

Tada je susjedska težina v_i dana formulom

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right), \quad i = 1, \dots, n.$$

Istaknimo kako se s povećanjem udaljenosti težine v_i smanjuju te postaju 0 u q -toj najudaljenijoj točki. Sljedeći korak je prilagodba polinoma stupnja d na podatke (x_i, y_i) sa težinom $v_i(x)$. Vrijednost tako lokalno prilagođenog polinoma u x -u je traženi $\hat{g}(x)$.

Sama STL metoda sastoji se od dvije rekurzivne procedure, unutarnje petlje koja je ugniježdjena u vanjsku petlju. Svaki prolazak kroz unutarnju petlju sastoji se od sezonalnog zaglađivanja kojim ćemo ažurirati sezonalnu komponentu te trend zaglađivanja kojim ažuriramo trend komponentu. Označimo sa $S_t^k, T_t^k, t = t_1, \dots, t_n$ sezonalnu i trend komponentu u k -tom prolasku kroz unutarnju petlju. Ažuriranje spomenutih komponenti u $k + 1$ prolasku kroz petlju određuje se na sljedeći način:

1. *Detrendiranje.* Unutarnju petlju započinjemo specificiranjem početnu vrijednosti za trend komponentu, tj. definiramo $T_t^{(0)}$. Često se uzima da je $T_t^{(0)} = 0$. Odredimo vremenski niz bez trenda kao

$$x_t - T_t^k.$$

2. *Zaglađivanje cikličkih podnizova.* Za početak podijelimo detrendirane podatke u podnizove čija je duljina jednaka duljini ciklusa. Korištenjem LOESS regresije, uz parametre $q = n_{(s)}$, pri čemu $n_{(s)}$ označava parametar zaglađivanja sezonalne komponente, i $d = 1$ koje smo spomenuli prilikom opisivanja same regresije, zagladimo podnizove ciklusa. Vrijednosti zaglađenih komponenti računamo u svim vremenima cikličkog podniza, uključujući i trenutke u kojima imamo nedostajuće vrijednosti. Isto tako u izračun uključimo i trenutak prije prvog vremenskog trenutka kao i trenutak nakon posljednjeg vremenskog trenutka. Na primjer, pretpostavimo da promatramo mjesečne podatke. Tada je broj opservacija u svakom ciklusu sezonalne komponente $n_{(p)}$ jednak je 12 te pretpostavimo da ciklički podniz za siječanj traje od siječnja 1968. godine do siječnja 2000. godine s tim da imamo nedostajuću vrijednost u siječnju 1985. godine. Zaglađene vrijednosti računamo u svim vremenskim trenucima počevši od 1967. godine zaključno sa 2001. godinom. Sve zaglađene vrijednosti nalazit će se u privremenom sezonalnom nizu koji označimo s C_t^{k+1} te se sastoji od $t + 2n_{(p)}$ podataka za $t \in \{t_1, \dots, t_n\}$.

3. *Low-pass filtriranje.* Kombinacijom pomičnih prosjeka i LOESS regresije napravimo low-pass filtriranje niza C_t^{k+1} te kao output dobivamo L_t^{k+1} . Budući

da pomični prosjeci ne mogu odrediti vrijednosti u krajevima vremenskih intervala, L_t^{k+1} ima t elemenata za $t \in \{t_1, \dots, t_n\}$, tj. gubimo $n_{(p)}$ vrijednosti sa svake strane što smo i očekivali pa smo zato u prethodnom koraku privremeni niz C_t^{k+1} proširili za isti iznos s obje strane.

4. *Određivanje procjene sezonalne komponente.* Uklonimo trend iz zaglađenih cikličkih podnizova te kao sezonalnu komponentu u $k + 1$ koraku petlje imamo

$$\hat{S}_t^{k+1} = C_t^{k+1} - L_t^{k+1}, \quad t \in \{t_1, \dots, t_n\}.$$

5. *Desezoniranje.* Uklonimo sezonalnost iz vremenskog niza kao

$$x_t - \hat{S}_t^{k+1}.$$

Napomenimo kako u ovom koraku, ako postoje nedostajuće vrijednosti za x_t , $t \in \{t_1, \dots, t_n\}$, one neće postojati niti u desezoniranom nizu.

6. *Zaglađivanje trend komponente.* Primjenom LOESS regresije uz parametre $q = n_{(t)}$, gdje je $n_{(t)}$ parametar zaglađivanja za trend komponentu i $d = 1$ odradimo zaglađivanje trend komponente u svim pozicijama za $t \in \{t_1, \dots, t_n\}$ te dobivamo procjenu \hat{T}_t^{k+1} .

Pretpostavimo sada da smo odradili jedan prolazak kroz unutarnju petlju te došli do procjena \hat{S}_t i \hat{T}_t . U vanjskoj petlji komponentu ostatka odredimo kao

$$\hat{R}_t = x_t - \hat{S}_t - \hat{T}_t.$$

U nastavku, za svaku točku u kojoj promatramo X_t odredimo **robusnu težinu** (eng. *robustness weight*) koja nam sugerira koliko je \hat{R}_t ekstreman. Korištenjem bikvadratne funkcije

$$B(u) = \begin{cases} (1 - u^2)^2, & 0 \leq u < 1 \\ 0, & u \geq 1 \end{cases}$$

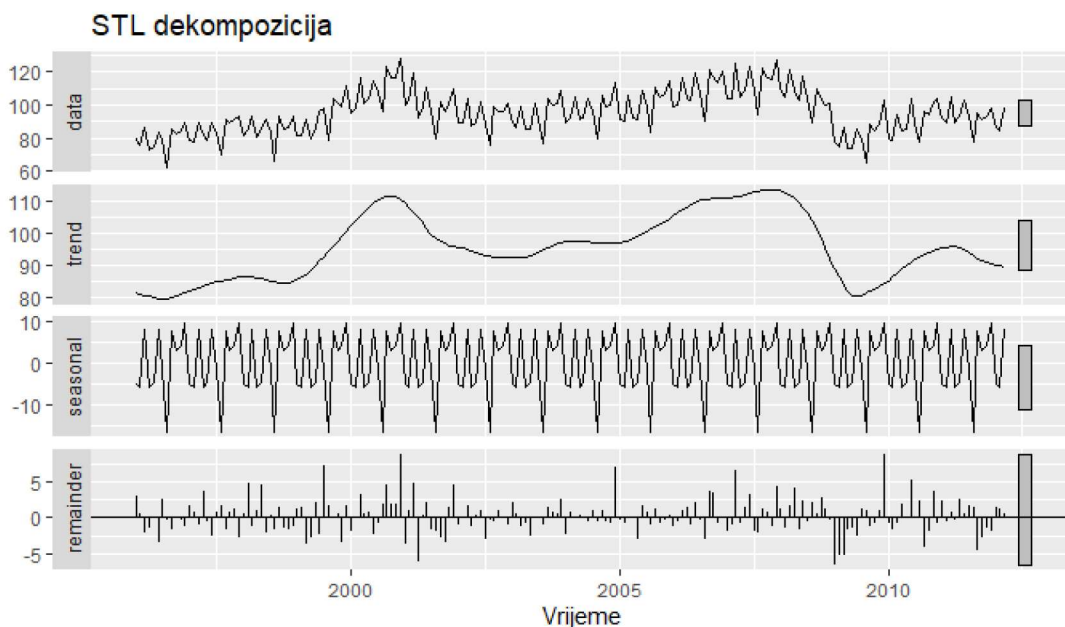
robusnu težinu u trenutku t računamo kao

$$\rho_t = B\left(\frac{|\hat{R}_t|}{h}\right)$$

pri čemu je $h = 6 \text{ median}(|\hat{R}_t|)$. Sada bismo ponovno ušli u unutarnju petlju, no u zaglađivanjima u koracima 2. i 6. bismo susjedsku težinu u trenutku t množili sa robusnom težinom u istom trenutku. Prilikom svakog novog ulaska u unutarnju petlju više ne definiramo $T_t^{(0)}$, već u prvi korak uzimamo \hat{T}_t^{k+1} koji smo odredili u 6. koraku prethodne unutarnje petlje.

Na početku potpoglavlja spomenuli smo kako ova metoda dekompozicije ima parametre koje korisnik smije sam birati. Osim već navedenih $n_{(p)}$, $n_{(s)}$ te $n_{(t)}$ imamo i broj prolazaka kroz unutarnju petlju $n_{(i)}$, broj robusnih iteracija u vanjskoj petlji $n_{(o)}$ i parametar zaglađivanja za low-pass filter $n_{(l)}$. Prilikom odabira parametra zaglađivanja sezonalne komponente valja pripaziti, dok se ostali parametri odabiru vrlo jednostavno. Pogledajmo kako STL dekompozicija izgleda na primjeru.

Primjer 18. Nastavljamo s dekomponiranjem podataka o proizvodnji električnih uređaja. Primjenom `stl()` funkcije na podacima dolazimo do istoimene dekompozicije danog niza. Rezultati kod trend komponente i sezonalne komponente slični su gotovo isti kao i u svim dosadašnjim primjerima, no razlika je vidljiva u komponenti ostatka.



Slika 4.9: Grafički prikaz STL dekompozicije podataka o proizvodnju elektroničke opreme u Europi.

STL dekompozicija također ima neke nedostatke. Konkretno, ona ne rješava efekte kalendara automatski i moguće ju je koristiti samo za aditivnu dekompoziciju. Ipak, multiplikativnu dekompoziciju možemo dobiti logaritmiranjem ili korištenjem Box-Cox transformacije.

5 | Primjene dekompozicije

5.1 Mjerenje snage trenda i sezonalnosti

STL dekompozicija se, osim za rastav vremenskog niza na komponente, može koristiti i kao određena mjera snage trenda i sezonalnosti. Ideju su 2006. godine u članku [16] predstavili Wang, Smith i Hyndman. Prisjetimo se aditivne dekompozicije koja nam omogućava da vremenski niz X_t zapišemo u obliku

$$X_t = T_t + S_t + R_t.$$

Za podatke koji imaju snažno izražen trend, sezonalno prilagođeni podaci bi trebali imati znatno izraženije varijacije od komponente ostatka. Iz tog razloga bi

$$\frac{\text{Var}(R_t)}{\text{Var}(R_t + T_t)}$$

trebala biti relativno mala. S druge strane, za podatke koje imaju malo trenda ili su bez trenda, varijance u brojniku i nazivniku bi mogle imati slične vrijednosti pa je sam razlomak blizak jedinici. Za vremenski niz X_t , $t \in \{t_1, \dots, t_n\}$ **snagu trenda** definiramo kao

$$F_T = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(R_t + T_t)} \right). \quad (5.1)$$

Na sličan način definira se i **snaga sezonalnosti**, no umjesto da promatramo sezonalno prilagođene podatke promatrati ćemo detrendirane podatke. Snaga sezonalnosti dana je formulom

$$F_S = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(R_t + S_t)} \right). \quad (5.2)$$

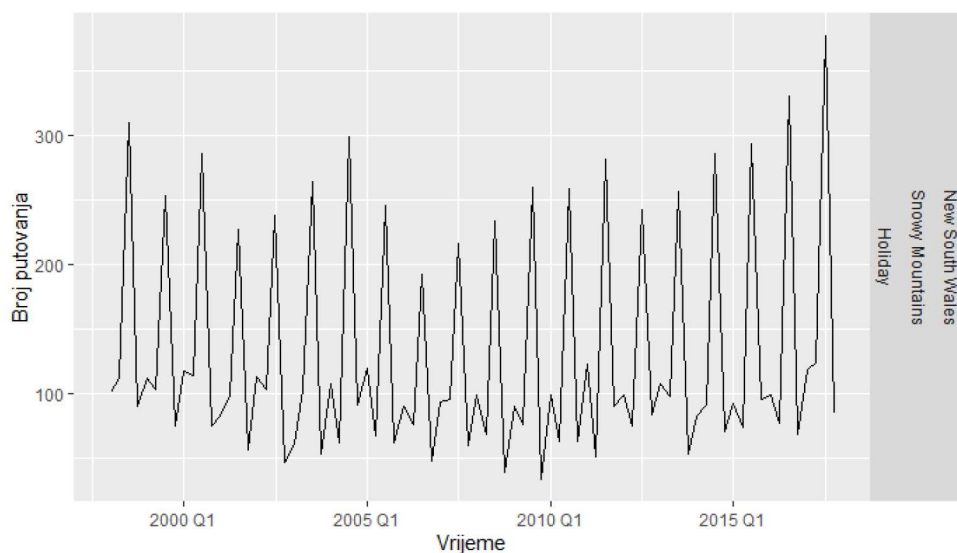
Vremenski nizovi čija je sezonalna snaga bliska nuli prikazuju gotovo nikakvu sezonalnost dok nizovi s izraženom sezonalnosti imaju snagu blisku jedinici. Ove mjere mogu biti od koristi ukoliko imamo veliki skup različitih vremenskih nizova te želimo pronaći niz s najvećim trendom ili sezonalnosti. Na primjeru pogledajmo kako koristimo snagu trenda i sezonalnosti.

Primjer 19. *Promatramo kvartalne podatke o broju putovanja s noćenjem po australskim saveznim državama u periodu od 1998. do 2017. Na Slici 5.1 možemo vidjeti grafički prikaz snaga trenda i sezonalnosti po svrsi putovanja za svaku od australskih saveznih država.*



Slika 5.1: Grafički prikaz snage sezonalnosti i snage trenda za sve vrste putovanja po saveznoj državi.

Uočavamo kako najjaču sezonalnost imaju putovanja kojima je svrha godišnji odmor što ima smisla budući da se ona uglavnom događaju u ljetnim mjesecima. Najjači trend se pak pojavljuje kod vremenskih nizova povezanih sa saveznom državama Victoria i Zapadna Australija. Pokušamo li pronaći niz s najjačom sezonalnosti dobivamo vremenski niz koji prati broj putovanja u planine koja su povezana uz godišnji odmor u saveznoj državi Novi Južni Wales, a možemo ga vidjeti na Slici 5.2.



Slika 5.2: Grafički prikaz vremenskog niza s najjačom sezonalnosti.

5.2 Predviđanje pomoću dekompozicije

Još jedna zanimljiva primjena dekompozicije vremenskih nizova je i korištenje iste za predviđanje. Pretpostavimo da vremenski niz X_t prati aditivnu dekompoziciju danu formulom 4.2. Tada ga možemo zapisati kao

$$X_t = \hat{S}_t + \hat{A}_t,$$

gdje je $\hat{A}_t = \hat{T}_t + \hat{R}_t$ sezonalno prilagođena komponenta. Kako bismo prediktirali dekomponirani vremenski niz moramo zasebno prediktirati sezonalnu komponentu \hat{S}_t i sezonalno prilagođenu komponentu \hat{A}_t . Uobičajeno je pretpostaviti da sezonalna komponenta ostaje nepromijenjena ili da se mijenja izrazito sporo pa se stoga prediktira na način da uzmemo zadnju godinu prediktirane komponente. Takav pristup naziva se još i **naivna metoda predikcije**.

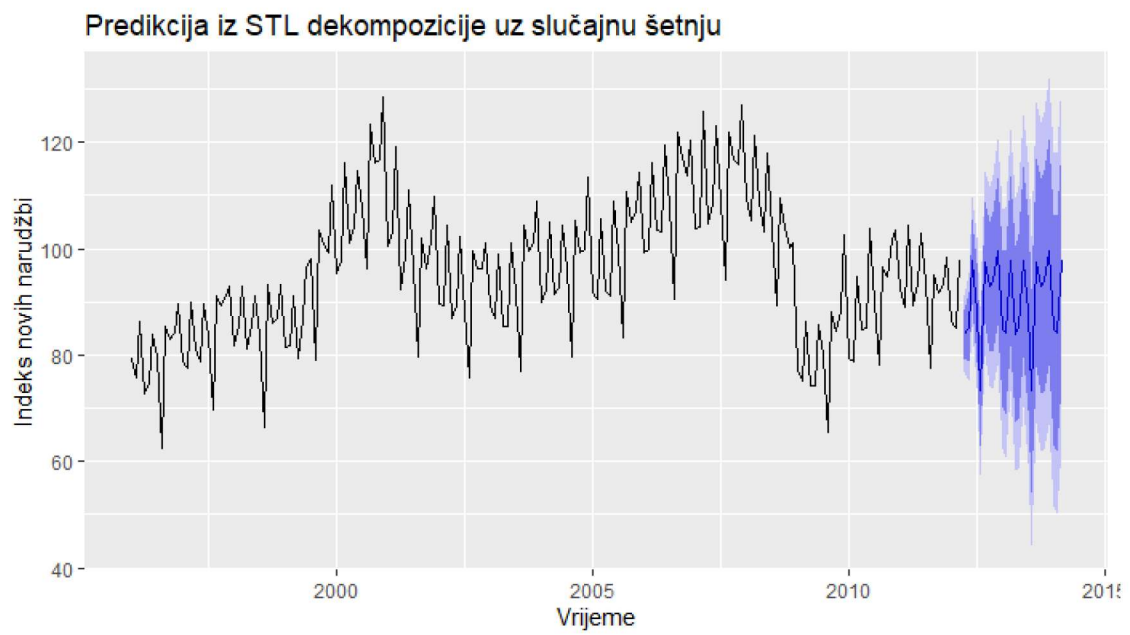
Za prediktiranje sezonalno prilagođene komponente možemo koristiti bilo koju od ne sezonalnih metoda predikcije poput drift metode, Holtove metode itd. o kojima se više može pročitati u [12].

Primjer 20. Vratimo se na podatke o proizvodnji električnih uređaja koje smo koristili u prethodnom poglavlju. Želimo predvidjeti daljnje kretanje tog vremenskog niza pomoću dekompozicije. Za početak ćemo prediktirati sezonalno prilagođenu komponentu korištenjem naivne metode. U pozadini ćemo koristiti STL metodu dekompozicije kako bismo rastavili niz na komponente. Predikciju sezonalno prilagođene komponente možemo vidjeti na Slici 5.3.



Slika 5.3: Grafički prikaz naivne dekompozicije prilagođenih podataka o proizvodnji elektroničke opreme u Europi.

Ovi podaci se zatim "resezonaliziraju" na način da se dodaju sezonalne predikcije dobivene naivnom metodom. Kao rezultat dobivamo predikciju početnog vremenskog niza koju možemo vidjeti na Slici 5.4.



Slika 5.4: Grafički prikaz predikcije podataka o proizvodnji elektroničke opreme u Europi.

Literatura

- [1] Đ. BOROZAN, *Makroekonomija*, Osijek: Ekonomski fakultet u Osijeku Sveučilišta Josipa Jurja Strossmayera u Osijeku, 2019.
- [2] G. E. P. BOX & G. M. JENKINS, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day, 1970.
- [3] P. J. BROCKWELL & R. A. DAVIS, *Time Series: Theory and Methods, 2nd edition*, Springer Science & Business Media, New York, 1991.
- [4] J. P. BURMAN, Seasonal Adjustment by Signal Extraction, *Journal of the Royal Statistical Society, Series A*, **143**(1980), 321-337.
- [5] R. B. CLEVELAND, W. S. CLEVELAND, J. E. McRAE & I. J. TERPENNING, *STL: A seasonal-trend decomposition procedure based on loess*, *Journal of Official Statistics* **6**(1)(1990), 3–33.
- [6] E. B. DAGUM & S. BIANCONCINI, *Seasonal adjustment methods and real time trend-cycle estimation*, Berlin/Heidelberg, Germany: Springer International Publishing, 2016.
- [7] E. B. DAGUM, *Seasonal Factor Forecasts from ARIMA Models*, *Proceedings of the International Institute of Statistics, 40th Session, Contributed Papers*, Warsaw, **3**(1975), 206-219.
- [8] E. B. DAGUM, *The X-11-ARIMA Seasonal Adjustment Method*, *Statistique Canada, Catalogue*(1980) 12-564E.
- [9] D. F. FINDLEY ET AL., *New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program*, *Journal of Business and Economic Statistics*, **16**(1998), 127-177.
- [10] V. GOMEZ & A. MARAVALL, *Programs TRAMO and SEATS*, Banco de Espana, Documento de Trabajo 9628., 1996.
- [11] V. M. GUERRERO, *Time-series analysis supported by power transformations*, *Journal of Forecasting*, **12**(1993), 37–48.
- [12] R. J. HYNDMAN & G. ATHANASOPOULOS, *Forecasting: principles and practice, 3rd edition*, Melbourne, Australia: OTexts., 2021.

-
- [13] D. LADIRAY & B. QUENNEVILLE, *Seasonal Adjustment with the X-11 Method*, New York, Springer-Verlag, 2001.
- [14] W. M. PERSONS, *Indices of Business Conditions: An Index of General Business Conditions*, *Review of Economic Statistics*, **1**(1919) 5-107.
- [15] J. SHISKIN, A. YOUNG & J. MUSUGRAVE, *The X-11 Variant of the Census Method II Seasonal Adjustment Program*, Bureau of the Census. Technical paper no. 15, 1965.
- [16] X. WANG, K. A. SMITH & R. J. HYNDMAN, Characteristic-based clustering for time series data., *Data Mining and Knowledge Discovery*, **13**(3)(2006), 335–364.

Sažetak

U ovom diplomskom radu bavimo se dekompozicijom vremenskih nizova kao i metodama koje se koriste u praksi. Započinjemo s pregledom važnih definicija i rezultata potrebnih za uvođenje novih pojmova. Nakon toga se okrećemo raznim metodama transformacije i prilagodbe vremenskih nizova koje ilustriramo na primjerima te naglašavamo njihovu korisnost za daljnju analizu. Zatim prelazimo na same metode dekompozicije gdje, osim povijesnog razvoja, detaljno opisujemo i analiziramo svaku od najpoznatijih metoda na primjerima. Na samome kraju pažnju dajemo primjenama metoda dekompozicije.

Ključne riječi

vremenski niz, ARIMA model, Box-Cox transformacija, metode dekompozicije, X-11 dekompozicija, TRAMO-SEATS dekompozicija, STL dekompozicija

Time series decomposition

Summary

In this paper, we consider decomposition of time series as well as methods used in practice. We begin by reviewing important definitions and results which are necessary for the introduction of new terms. After that we turn to various methods of transformation and time series adjustments that we will illustrate in examples. We emphasize their usefulness for further analysis. Then we move on to decomposition methods themselves where, in addition to the historical development, we give detailed description and analysis for some of most used decomposition methods. At the very end, we take a look at some applications of methods described earlier.

Keywords

time series, ARIMA model, Box-Cox transformation, decomposition methods, X-11 decomposition, TRAMO-SEATS decomposition, STL decomposition

Životopis

Rođena sam 21. travnja 1998. godine u Karlovcu. Osnovnu školu Braće Seljan, u istom gradu, završila sam 2013. godine te upisala opći smjer u Gimnaziji Karlovac. Prijediplomski studij matematike, na tadašnjem Odjelu za matematiku, upisala sam 2017. godine, a završila 2021. godine s temom završnog rada "Krivuljni integrali i neke njihove primjene" pod mentorstvom prof. dr. sc. Krešimira Burazina i komentorstvom dr. sc. Jelene Jankov Pavlović. U akademskoj godini 2021./2022. upisujem diplomski studij matematike, modul Financijska matematika i statistika te sam u istoj akademskoj godini izabrana za pravobranitelja Odjela za matematiku. U godini 2023./2024. dobila sam Pohvalu za izvannastavne aktivnosti. Za vrijeme studiranja radila sam kao podatkovni znanstvenik u Alpha score d.o.o. u Osijeku.