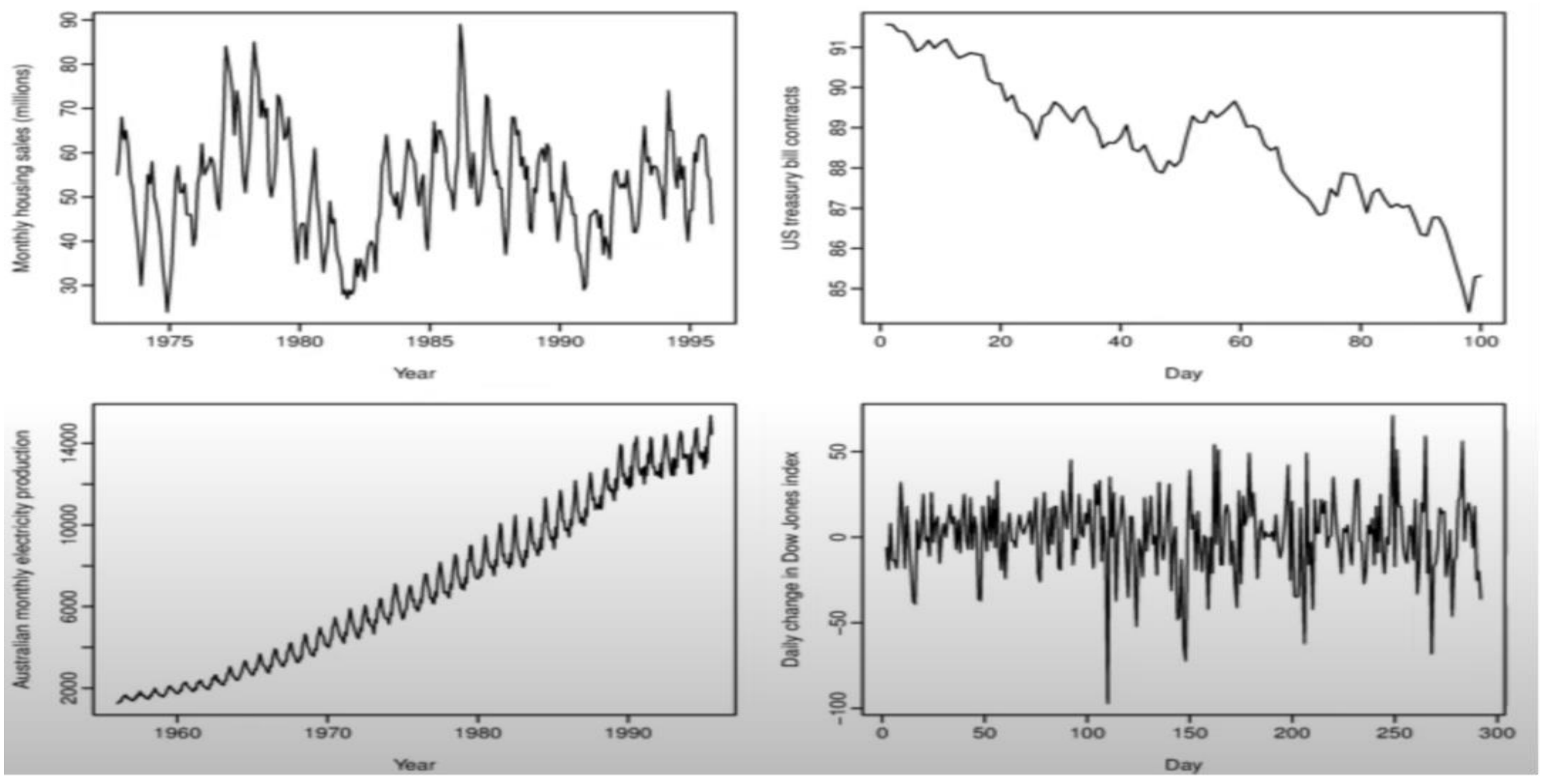
ANALIZA VREMENSKIH SERIJA

predavač:
Aleksandar Kovačević

Definicija vremenske serije

- Vremenska serija (*time series*) je kolekcija opservacija x_t koje su merene sekvencijalno po vremenu.
- Kolekcija podataka uređenih (indekisranih) po vremenu.



Diskretne i kontinualne vremenske serije

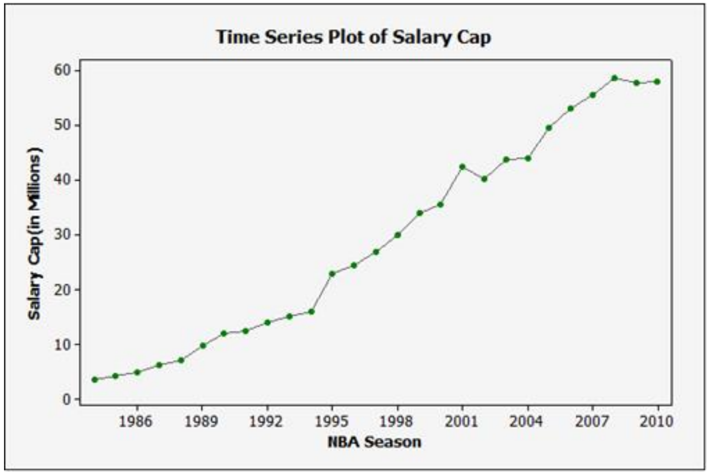
- Diskretna vremenska serija je kolekcija opservacija x_t koje su merene u diskretnom skupu T_0 trenutaka u vremenu.

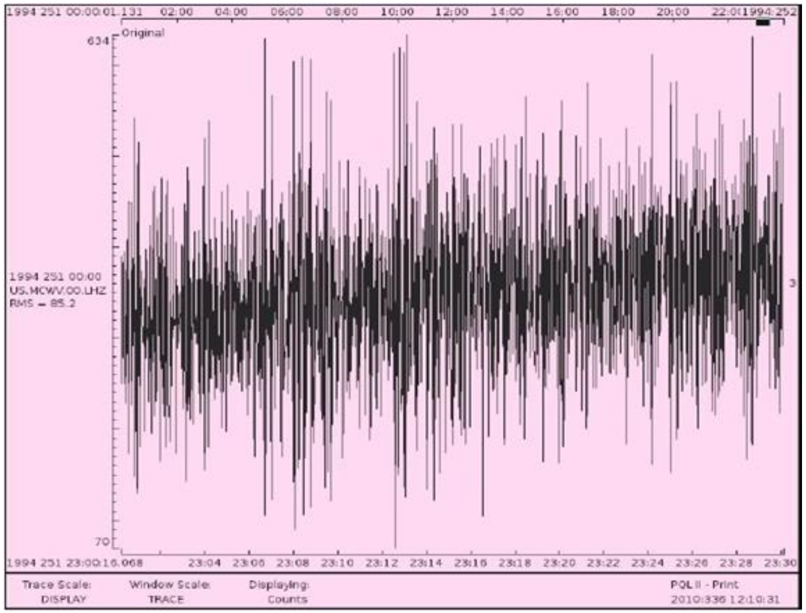
- Kontinualna vremenska serija je kolekcija opservacija x_t koje su merene kontinualno u nekom vremenskom intervalu.

- Predmet ovog kursa su diskretne vremenske serije.

Primeri diskretne i kontinualne vremenske serije

- Diskretna vremenska serija – NBA *Salary Cap* za svaku godinu do 2010.
- Kontinualna vremenska serija – audio snimak





Analiza vremenskih serija

- Oblast koja se bavi vremskim serijama sa <u>dva važna cilja</u>:

1. <u>Analiza prošlosti</u>
   - Proučavanje oblika vremenske serije
   - Rasklapanje vremenske serije na komponente
   - Tumačenje same serije i komponenti

2. <u>Predikcija budućnosti</u>
   - Modelovanje procesa koji generisao datu vremesku seriju
   - Upotreba modela za predikciju narednih vrednosti po vremenu

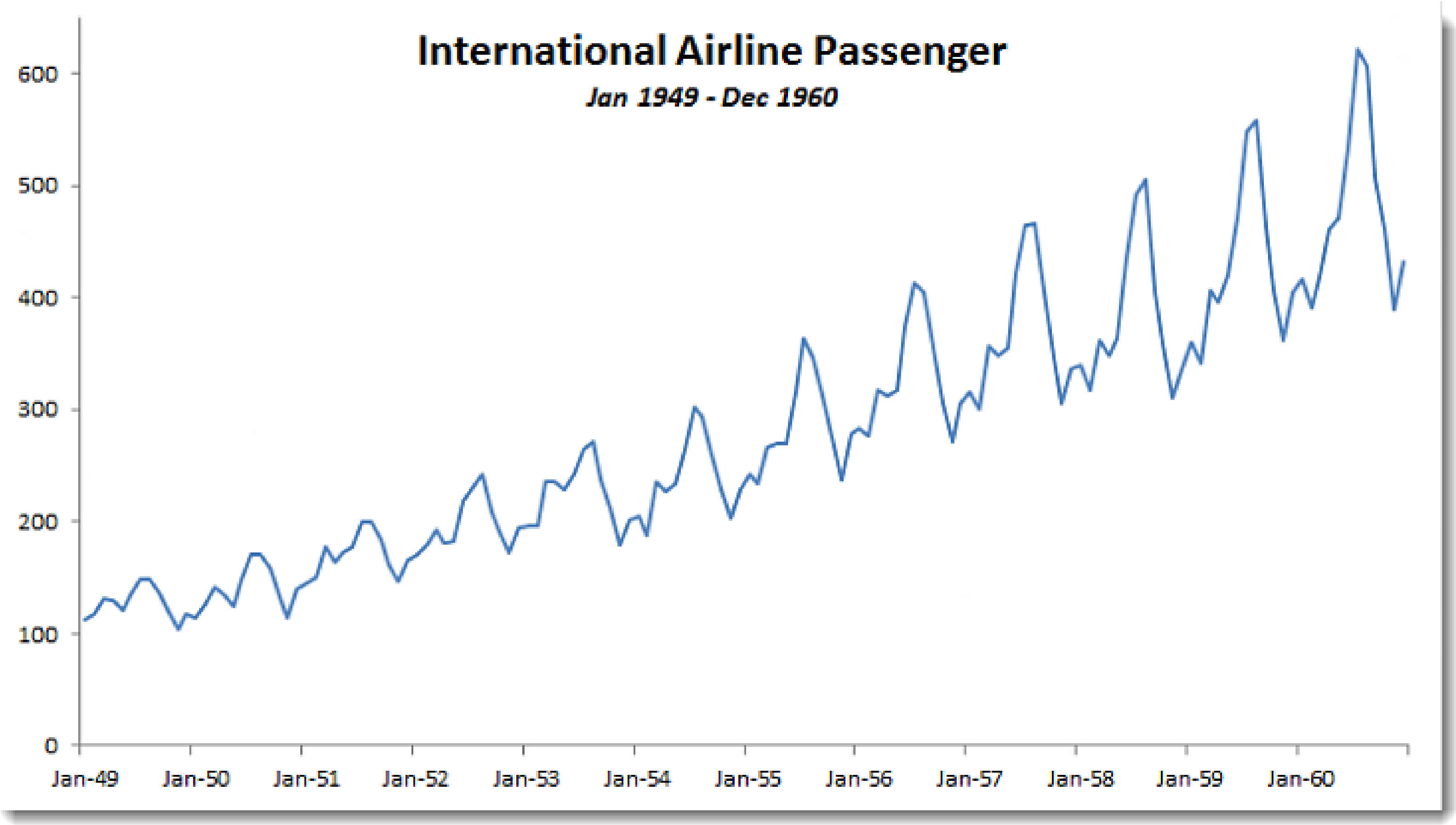
Primeri upotrebe analize vremenskih serija

- Ekonomija:
  - ❑Praćenje prometa na berzi tokom vremena radi donošenja investicionih odluka.
  - ❑Analiza cena potrošačkih proizvoda kako bi se identifikovali sezonski obrasci i predviđanje budućih cena.
  - ❑Praćenje stope inflacije tokom godina kako bi se razumela ekonomska stabilnost.
  - ❑Analiza vremenskih serija bruto domaćeg proizvoda (BDP) radi procene ekonomske aktivnosti.
- Sport:
  - ❑Analiza statističkih podataka o performansama igrača tokom sezona.
  - ❑Praćenje rezultata sportskih ekipa tokom vremena radi identifikacije trendova.
  - ❑Analiza brzine i performansi sportista tokom vremena radi poboljšanja treninga.

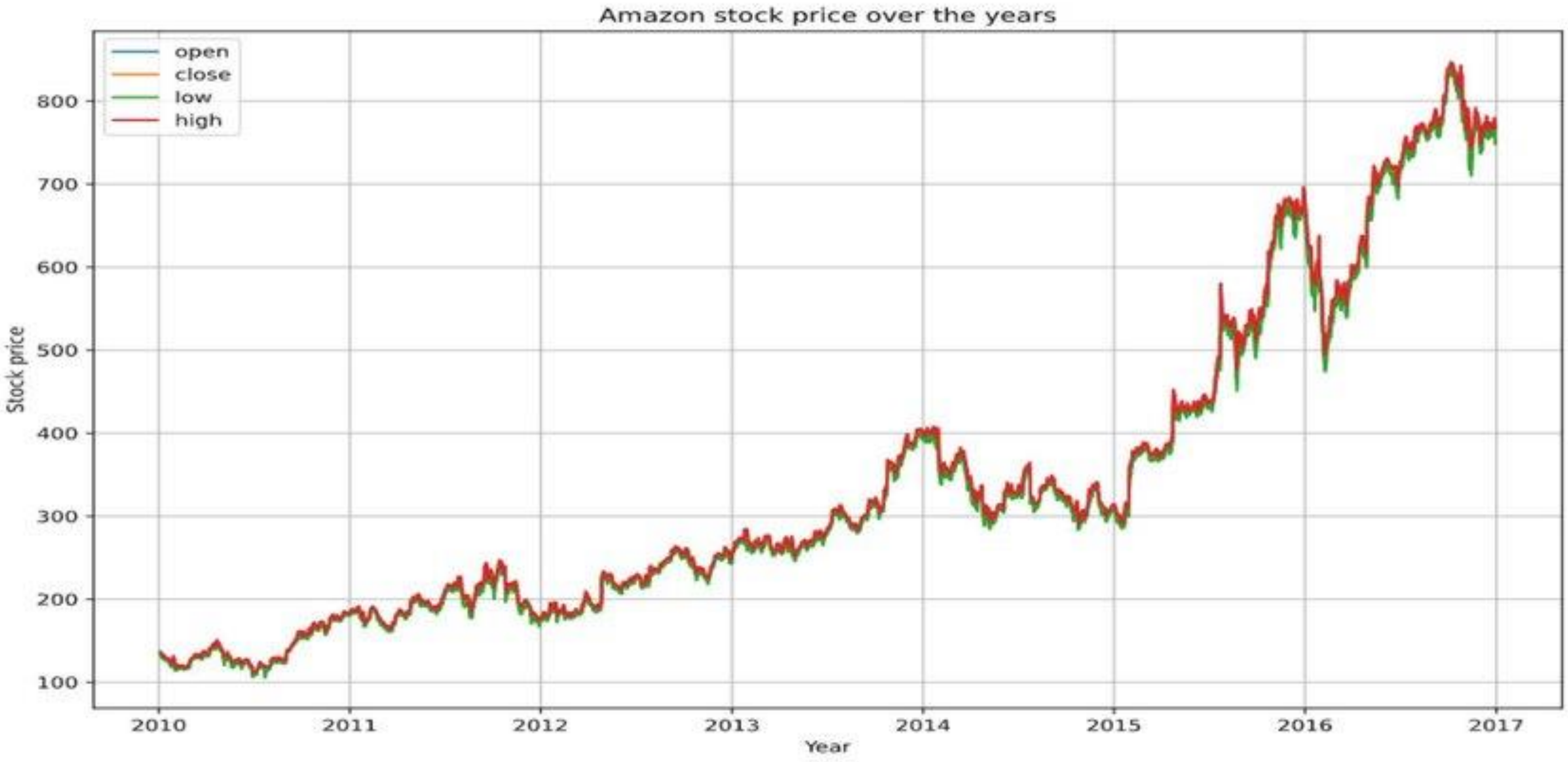
Primeri upotrebe analize vremenskih serija

- Nauka:
  - Praćenje promena u nivou zagađenja vazduha tokom vremena radi procene kvaliteta vazduha.
  - Analiza vremenskih serija meteoroloških podataka radi predviđanja vremenskih uslova.
  - Praćenje evolucije populacija u ekologiji radi boljeg razumevanja dinamike ekosistema.
  - Analiza vremenskih serija u astrofizici za istraživanje svemira.
- Računarske nauke:
  - Praćenje opterećenja servera tokom vremena radi optimizacije resursa.
  - Analiza logova sistema za identifikaciju sigurnosnih incidenata.
  - Praćenje prometa na web sajtovima radi analize ponašanja korisnika i prilagođavanja sadržaja.

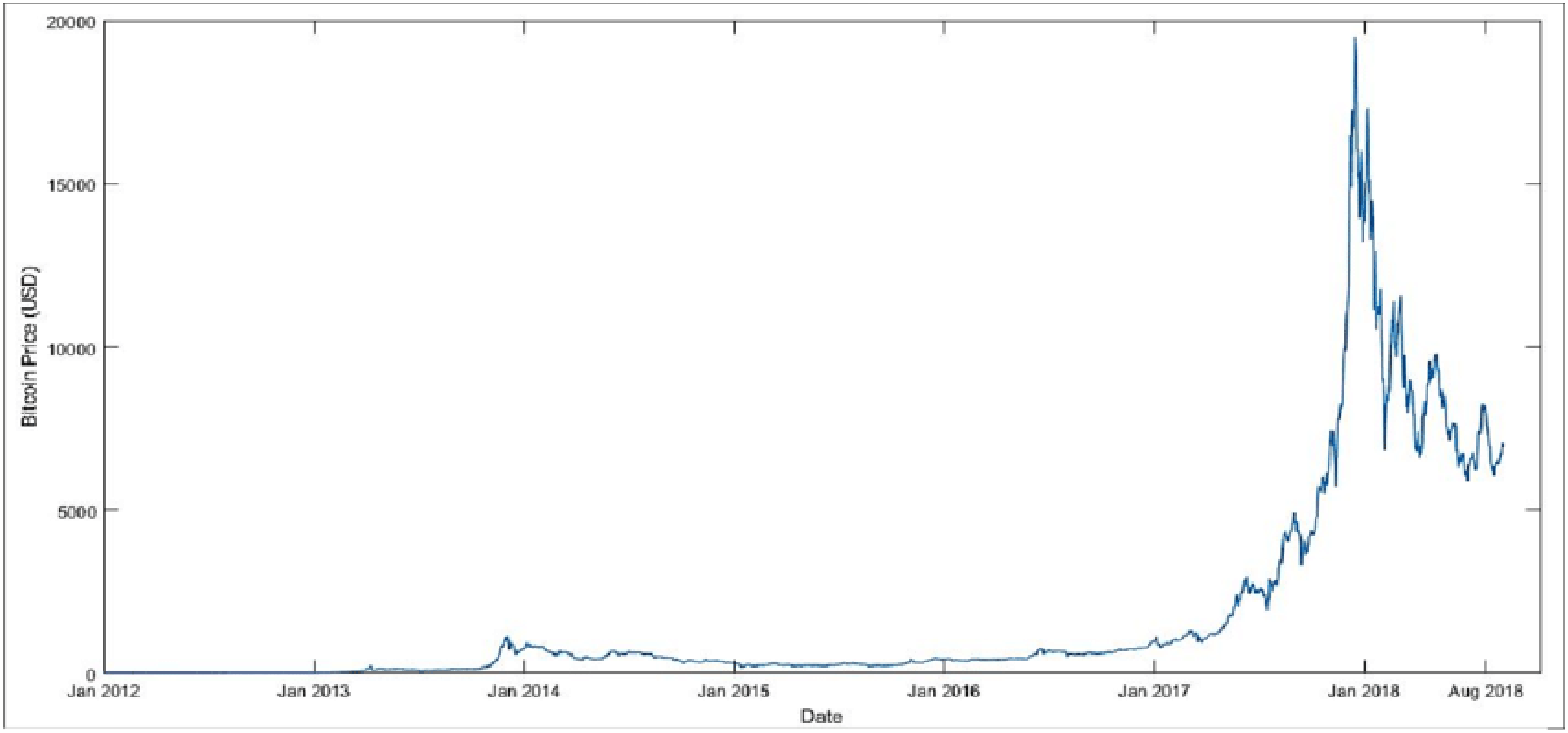
Primeri vremenskih serija – Ekonomija



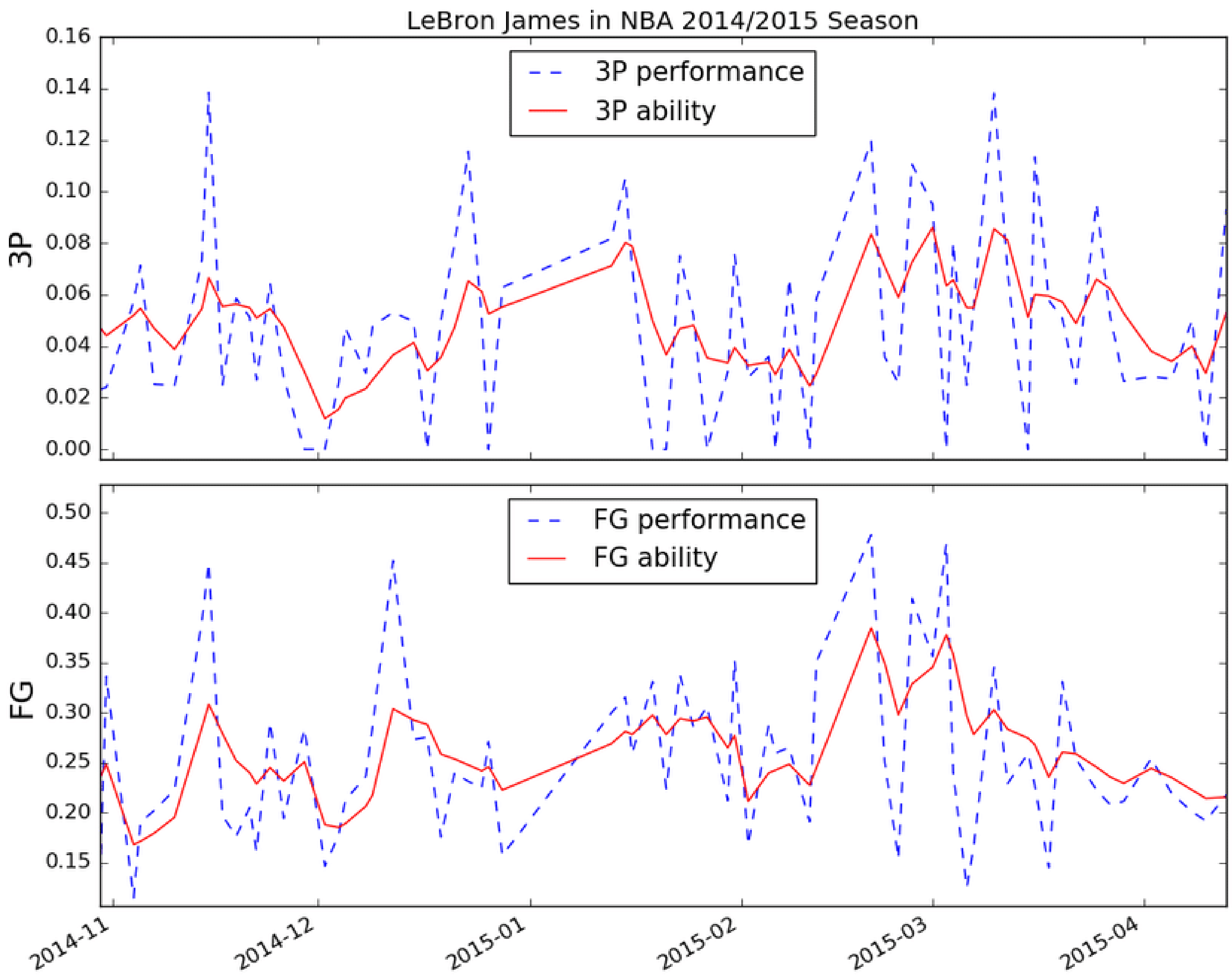
Primeri vremenskih serija – Ekonomija



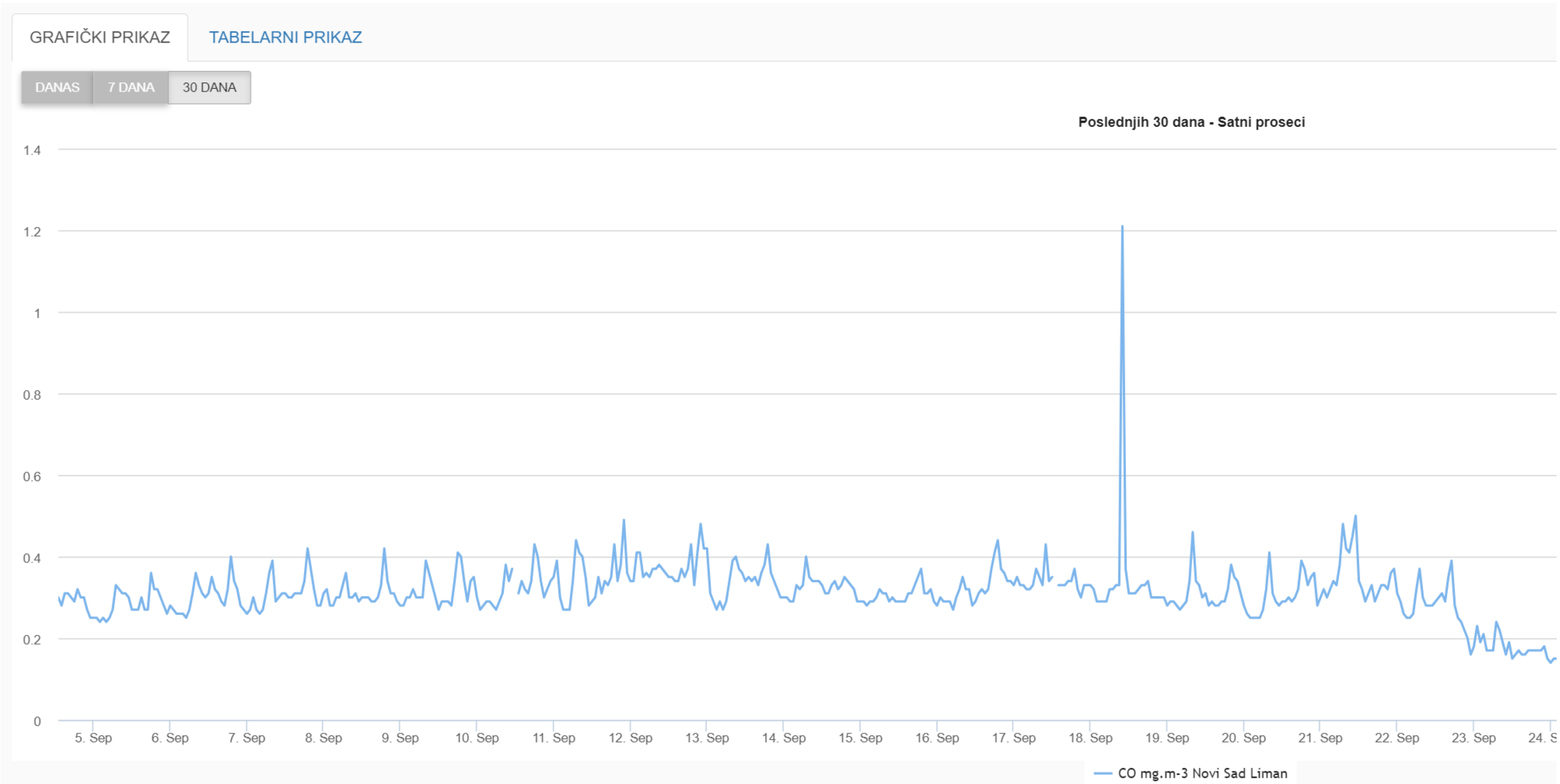
Primeri vremenskih serija – Ekonomija



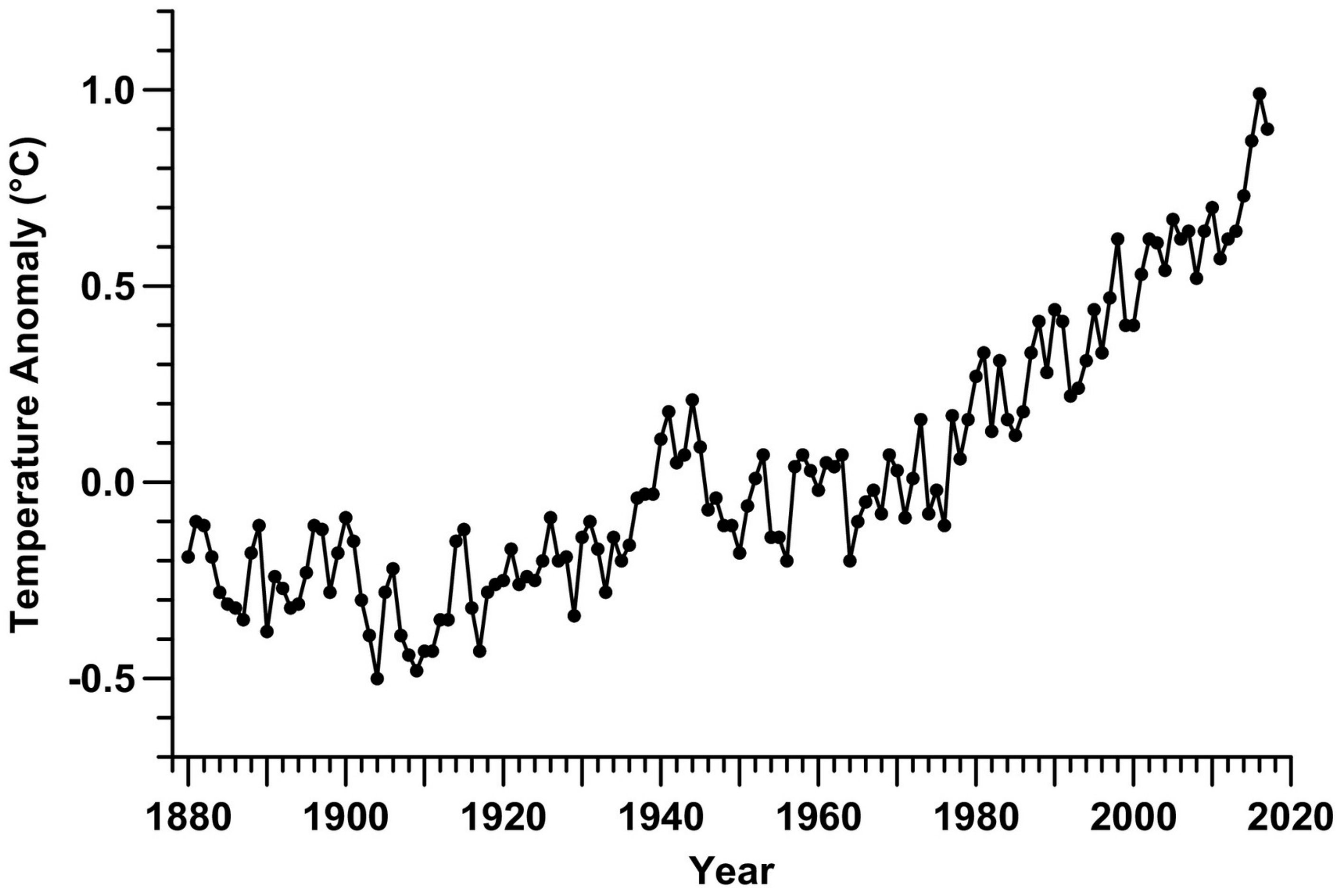
Primeri vremenskih serija – Sport



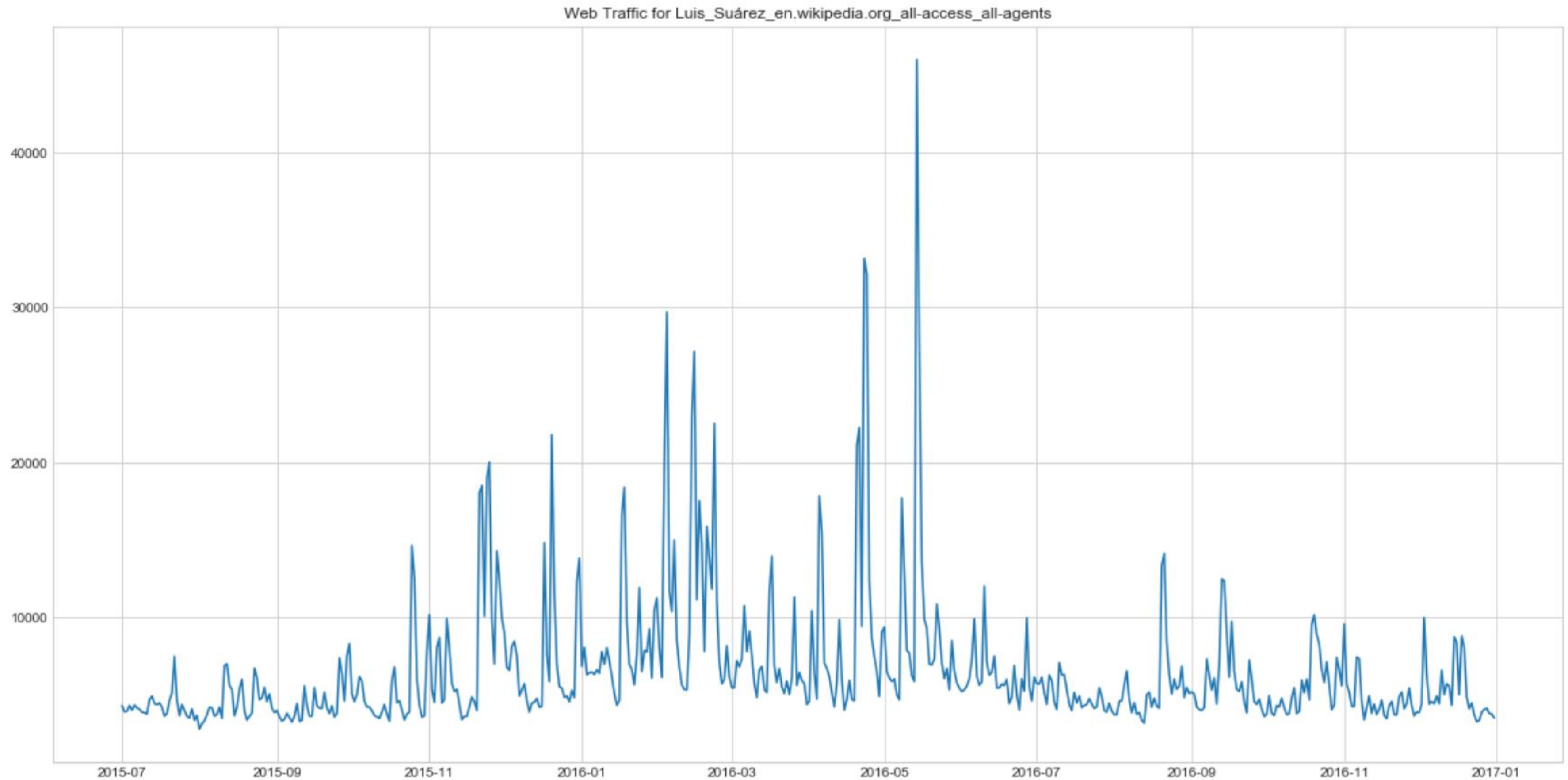
Primeri vremenskih serija – Nauka



Primeri vremenskih serija – Nauka

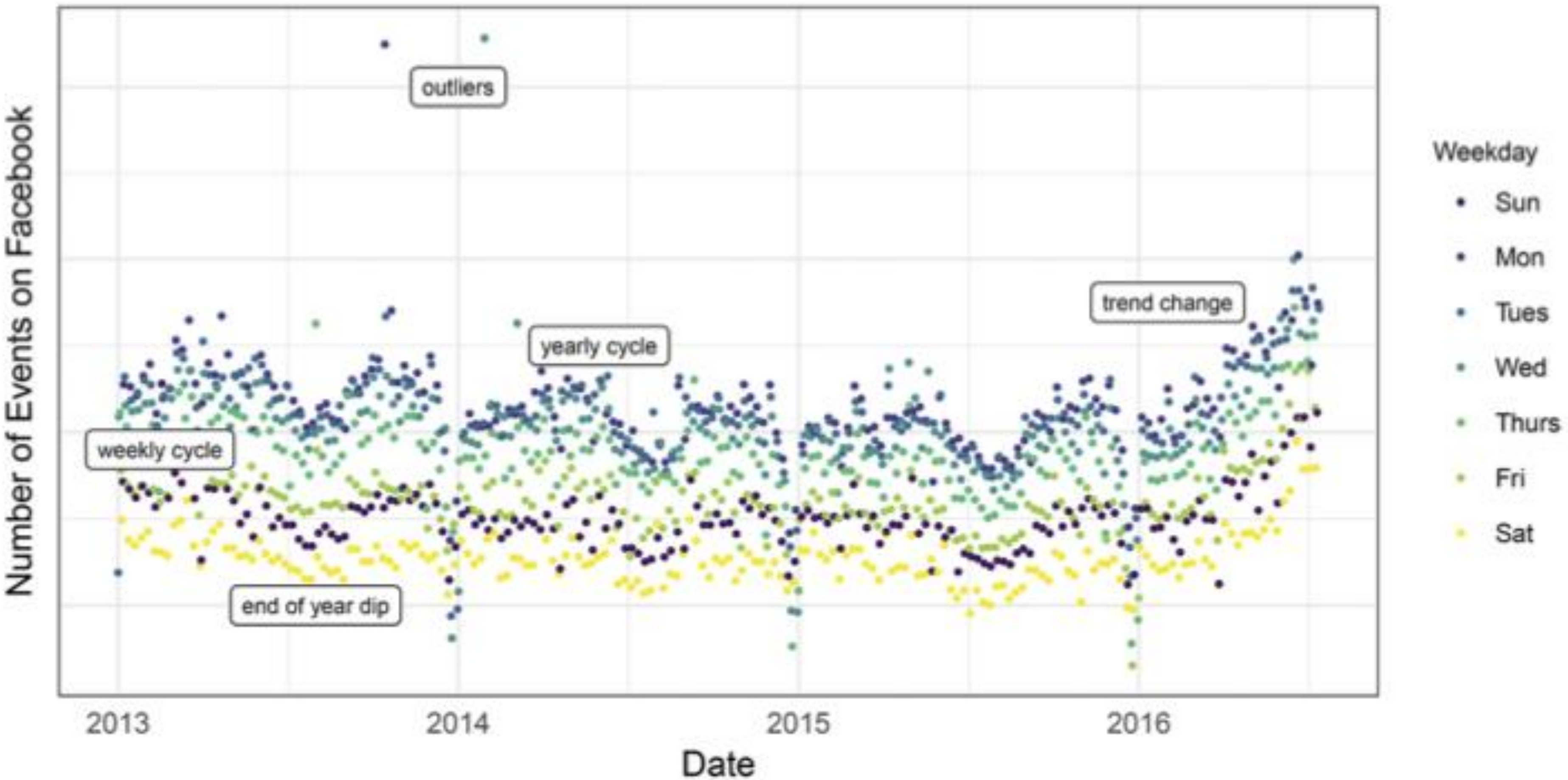


Primeri vremenskih serija – Računarstvo



This timeseries representes the traffic on the Luis Suarez wikipage (he's footabl player for Barcelona). It has quite a few spikes, which might be explained by the footbal season activity and his performance in the matches. For example, detecting these spikes might be useful in forecasting demand for his merchendise.

Primeri vremenskih serija – Računarstvo



Zašto posebno predavanje o vremenskim serijama?

- Problem kojim se bavimo je utvrđivanje odnosa dve promenjljive i predikcija jedne pomoću druge.
- Sada je nezavisna promenjljiva vreme.
- Šta je drugačije, odnosno zašto učimo nove tehnike u odnosu na one koje smo do sada naučili?
- Odgovor na to pitanje daćemo kroz primer prodaje sladoleda.
- Naš zadatak je da procenimo koliko će se sladoleda prodati zadatog dana.

Primer prodaje sladoleda

- Jedan način da predvidimo količinu prodatog sladoleda je da koristimo temperaturu vazduha.

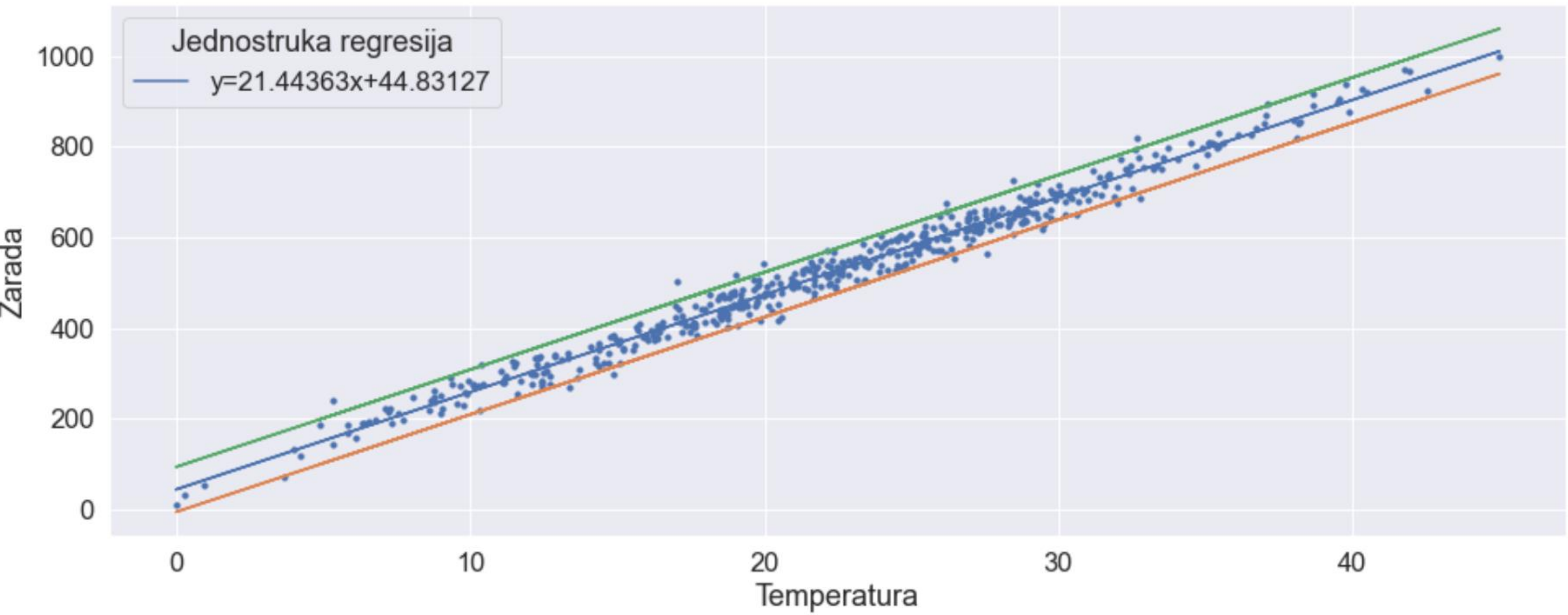
- Znamo da je prodaja sladoleda veća kada je toplije napolju.

- Recimo da smo neko vreme beležili temperaturu i koliko sladoleda je prodato tog dana.

- Nad tim uzorkom podataka onda radimo linearnu regresiju i dobijamo prediktivni model kao na sledećem slajdu.

Primer prodaje sladoleda – nezavisna promenljiva je Temperatura

- Na grafiku je dat rezultat linearne regresije kao i intervali predikcije.
- Vidimo da smo jednako pouzdani u predikcije na celom rasponu vrednosti *Temperature*.

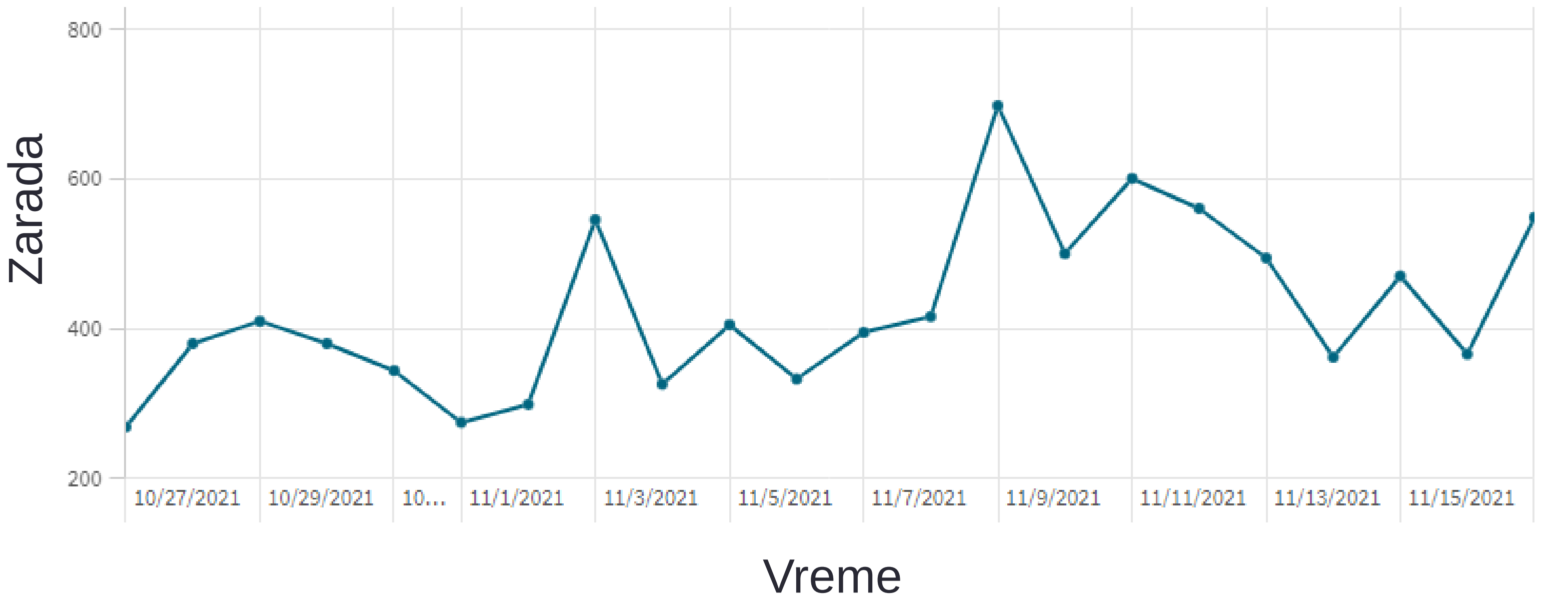


Primer prodaje sladoleda – nezavisna promenljiva je Vreme

- Drugi način da predvidimo količinu prodatog sladoleda je da koristimo vreme kao nezavisnu promenjljivu.
- Znamo da prodaja sladoleda sutra u nekoj meri zavisi od prodaje danas.
- Recimo da smo neko vreme beležili koliko sladoleda je prodato kog dana.
- Dobijamo grafik kao na sledećem slajdu.

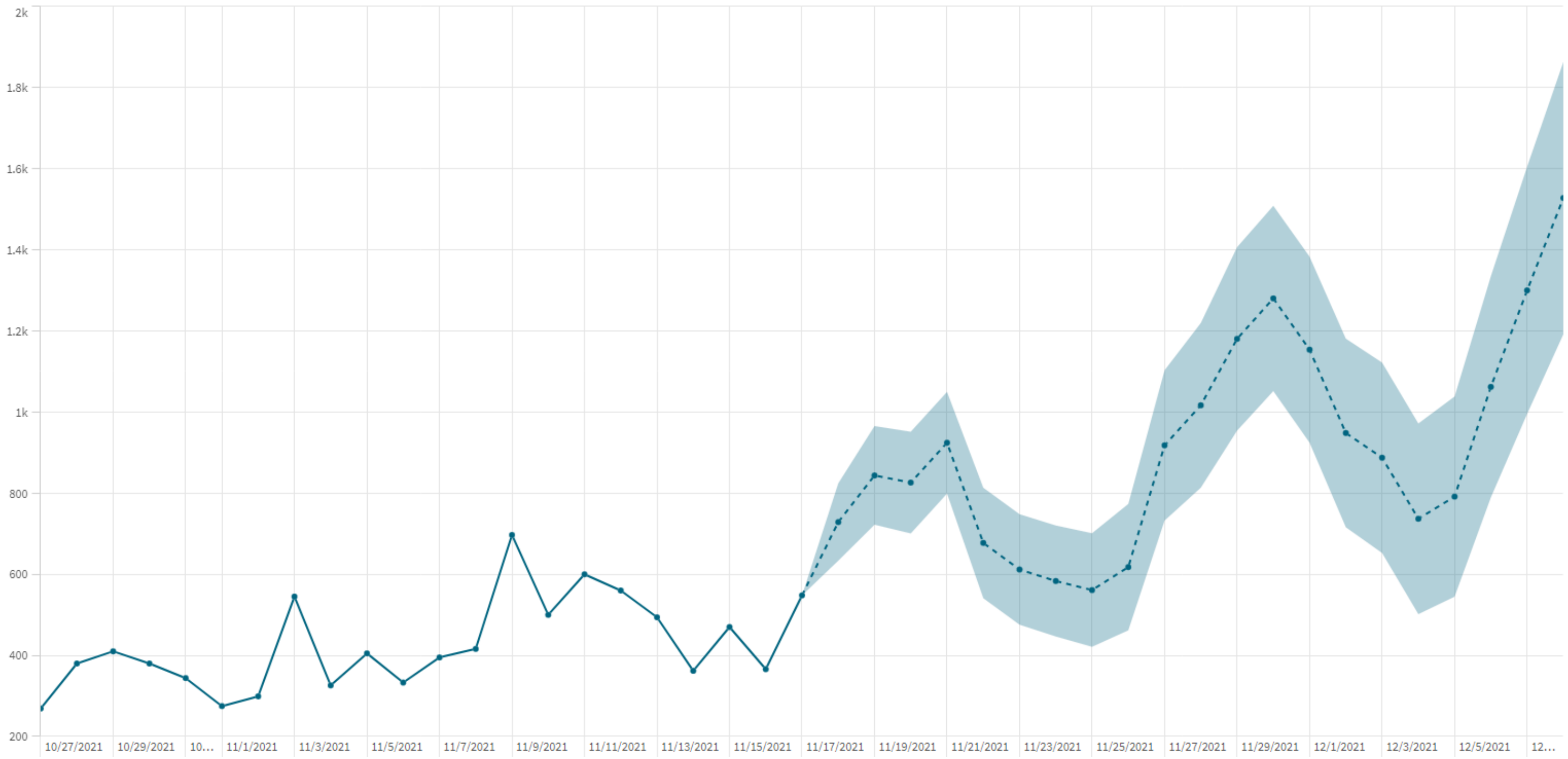
Primer prodaje sladoleda – nezavisna promenljiva je Vreme

- Recimo da želimo da predvidimo zaradu za budući period u odnosu na onaj koji imamo u podacima.
- Za svaki sledeći dan kao nezavisnu promenjljivu koristimo zaradu od prethodnog dana.
- Na sledećem slajdu data je ilustracija intrevala predikcije u tom slučaju.



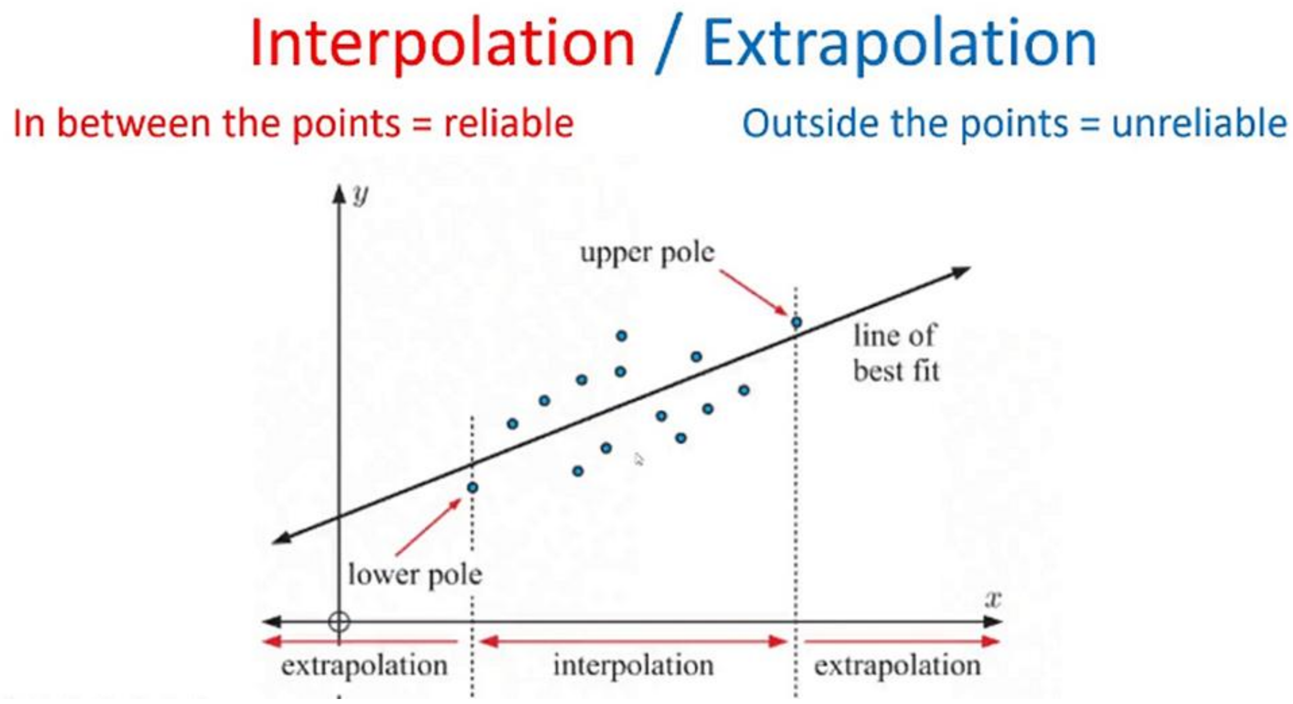
Primer prodaje sladoleda – propagacija greške

- Intervali predikcije postaju sve širi sa vremenom.
- To je zato što se greška predikcije propagira kroz vreme.
  - Predikcija za 18.11.2021. je relativno dobra jer je podatak od 17.11. tačan (dat). Međutim, zaradu za 19.11. predviđamo pomoću predikcije za 18.11. koja nije potpuno tačna, onda zaradu za 20.11. predviđamo pomoću zarade 19.11. itd.
- Zašto ovaj problem nismo imali kada smo koristili temperaturu?



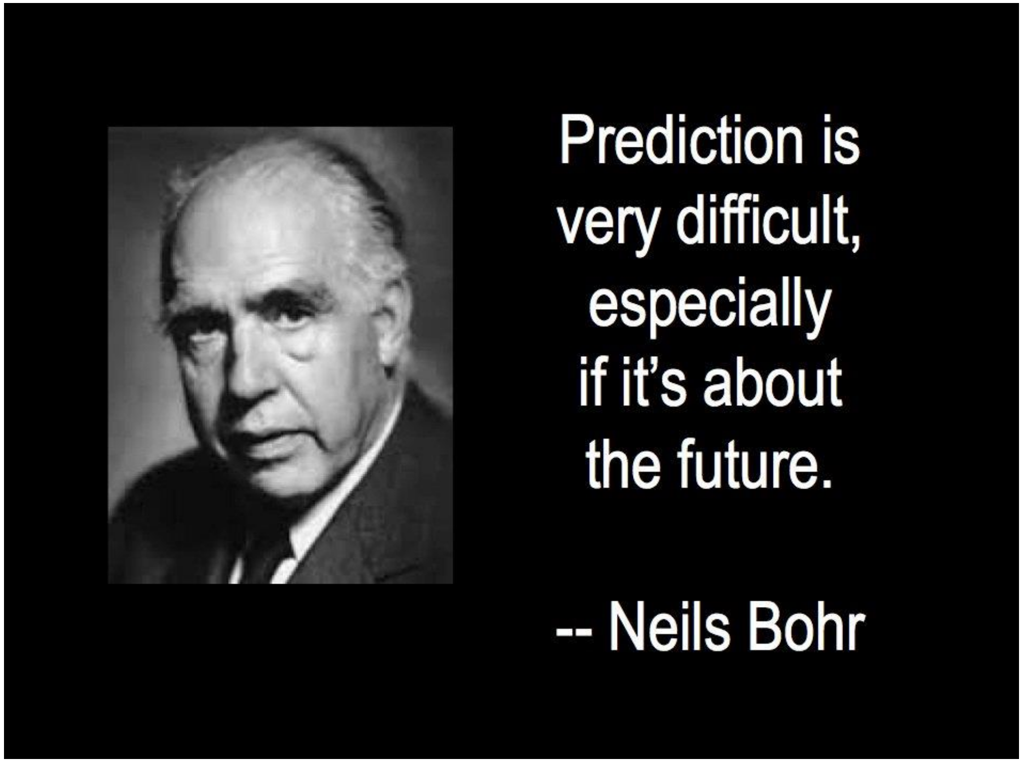
Primer prodaje sladoleda – ekstrapolacija

- Zašto ovaj problem nismo imali kada smo koristili temperaturu?
- Zato što smo u tom slučaju radili interpolaciju odnosno predikciju smo vršili unutar intervala vrednosti koje imamo u podacima.
- Kod predikcije pomoću vremena radimo ekstrapolaciju, tj. predviđamo van intrevala za koji imamo podatke.



Ekstrapolacija

- Ektrapolacija je značajno teža, upravo zbog propagacije greške.
- Zato koristimo posebne modele za vremenske serije kako bi razumeli proces koji je generisao tu vremsku seriju.
- Jer ako dobro modelujemo proces onda možemo bolje da predvidimo šta će biti u budućnosti.



Vremenske serije – sekvencijalnost podataka

- Još jedan važan razlog za razvoj posebnih modela za rad sa vremenskim serijama je to što su one uređene po vremenu, odnosno redosled podataka kod vremenske serije je važan.

- Podatke kod kojih je redosled važan nazivamo sekvencijalno uređeni podaci ili sekvencijalni podaci.
  - Prirodni jezik je primer sekvencijalno uređenih podataka.
    - Redosled reči je važan za semantiku rečenice.

- Podsetimo se modelovanja cena kuća na osnovu površine placa.
  - Redosled kuća u uzorku podataka nije važan.
  - Svaka permutacija uzorka podataka rezultovala bi istim modelom.

- Dakle, modeli koji ne uzimaju sekvencijalnost u obzir nisu pogodni za rad za vremsnkim serija.

- Zato su na potrebni posebni modeli koje razmatramo u nastavku krusa.

Komponente vremenskih serija

- Strukture ili šabloni (paterni) u vremenskim serijama koji utiču na ponašanje vremenske serije kroz vreme.

- Razumevanje komponenti važno je i za fazu analize, kao i za fazu prediktivnog modelovanja.

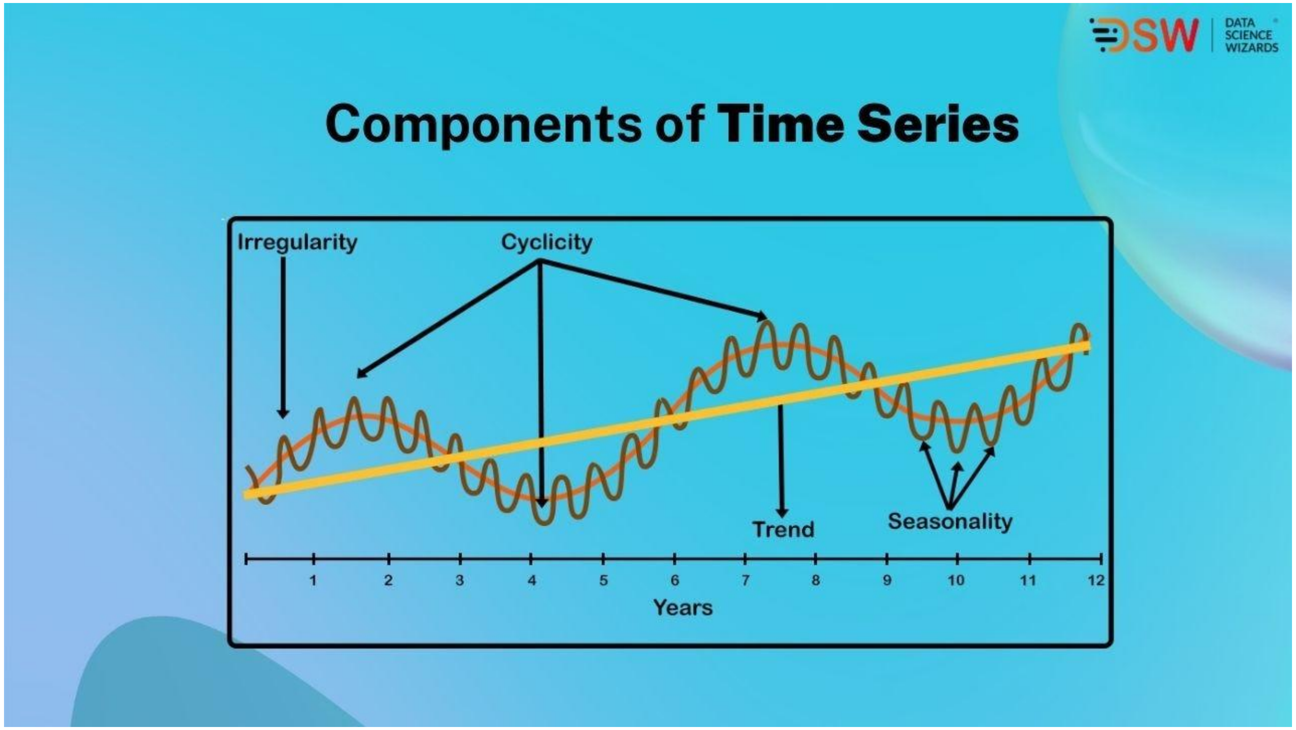
- Tipično se komponente mogu uočiti vizualizacijom vremenske serije.

- Izdvajanje komponenti (rasklapanje vremenske serije) vrši se pomoću različitih numeričkih alata.

Komponente vremenskih serija

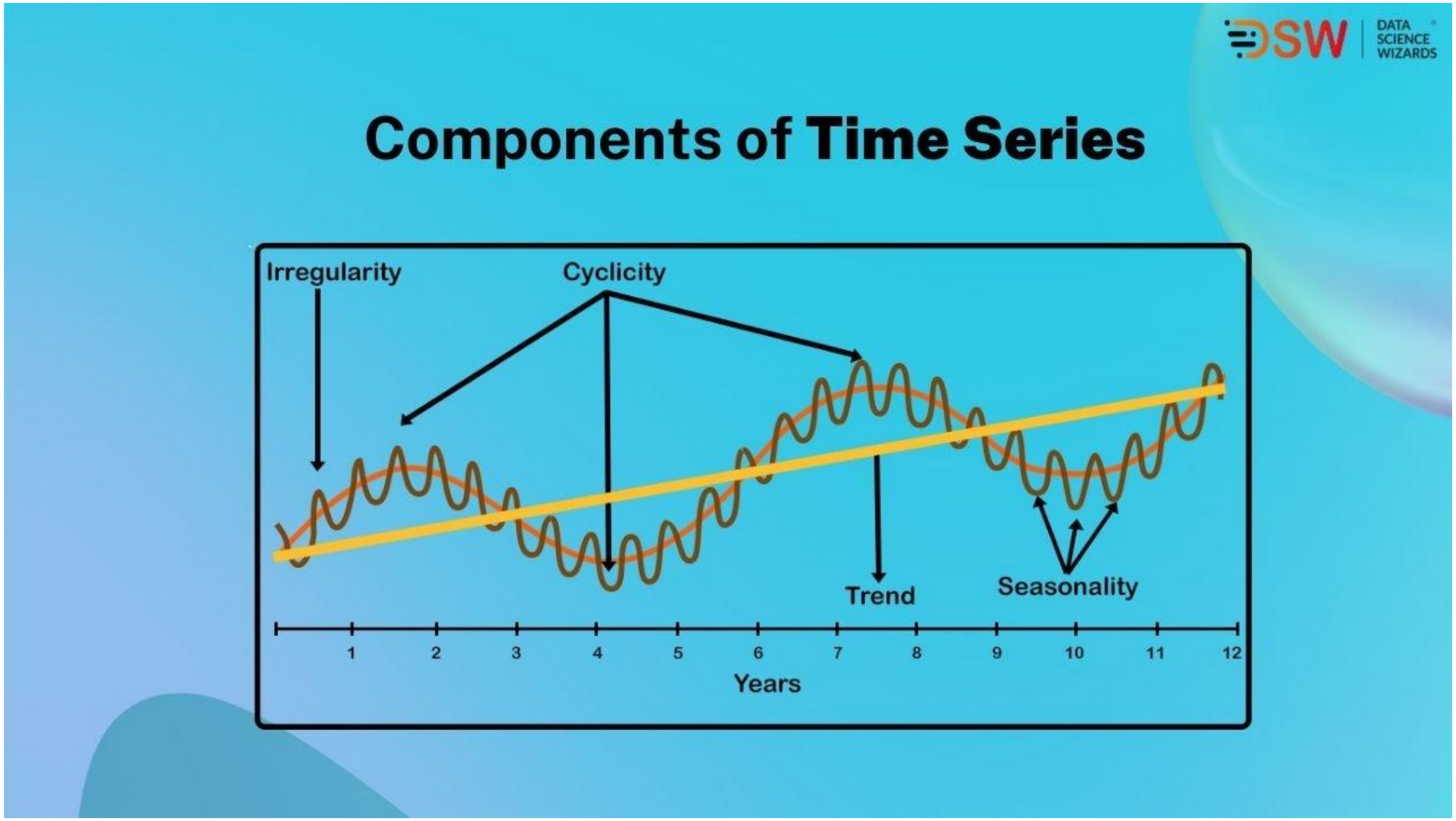
- Četiri glavne komponente vremenskih serija su:

- ❑ **Trend**
- ❑ **Sezonalnost**
- ❑ **Cikličnost**
- ❑ **Šum**



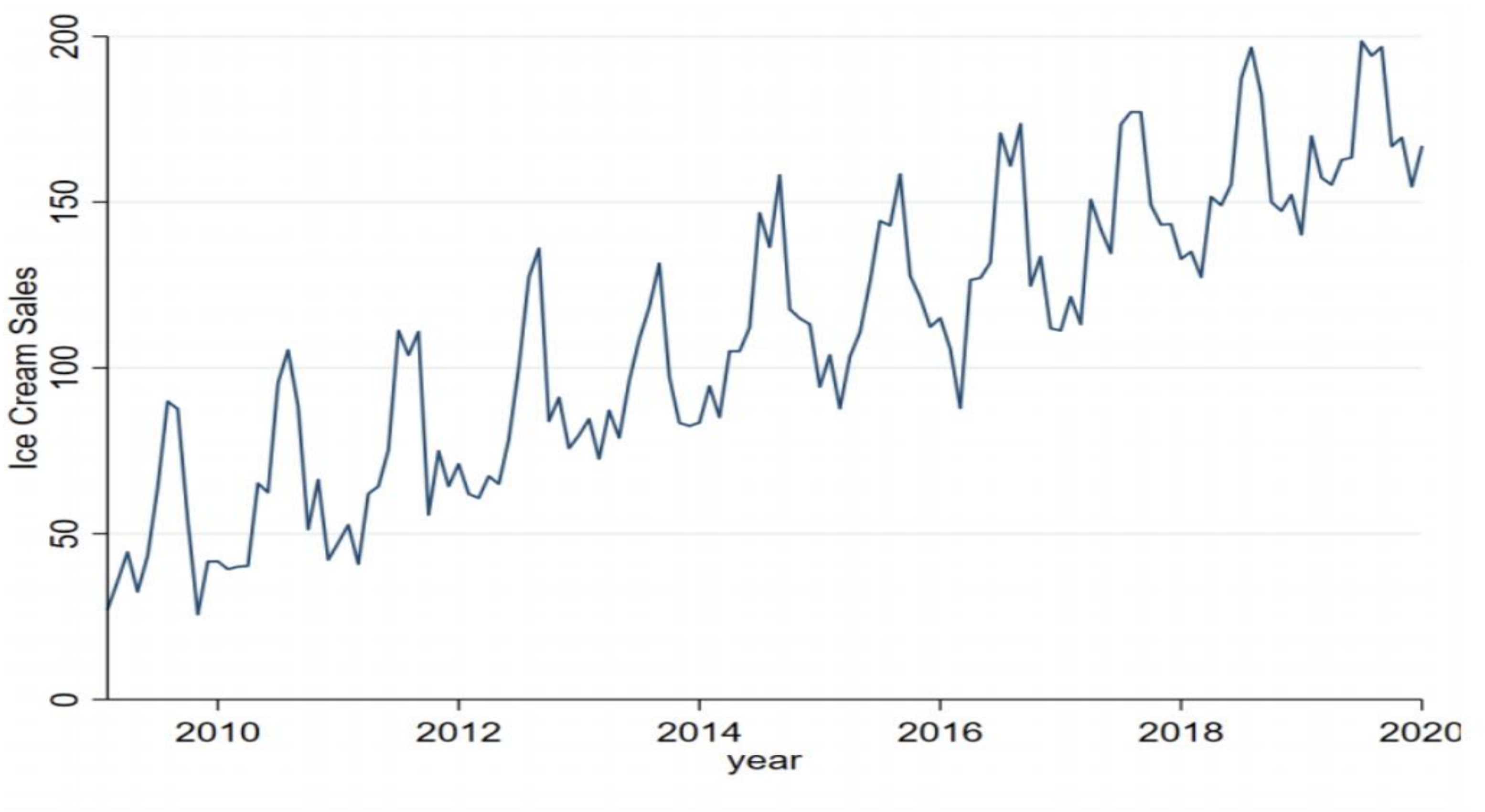
Komponente vremenskih serija – Trend

- Trend je tendencija vremenske serije da postepeno ali konstantno raste ili opada kroz dug vremenski period.
  - Na primer, konstantan rast populacije nekog naroda kroz vreme.



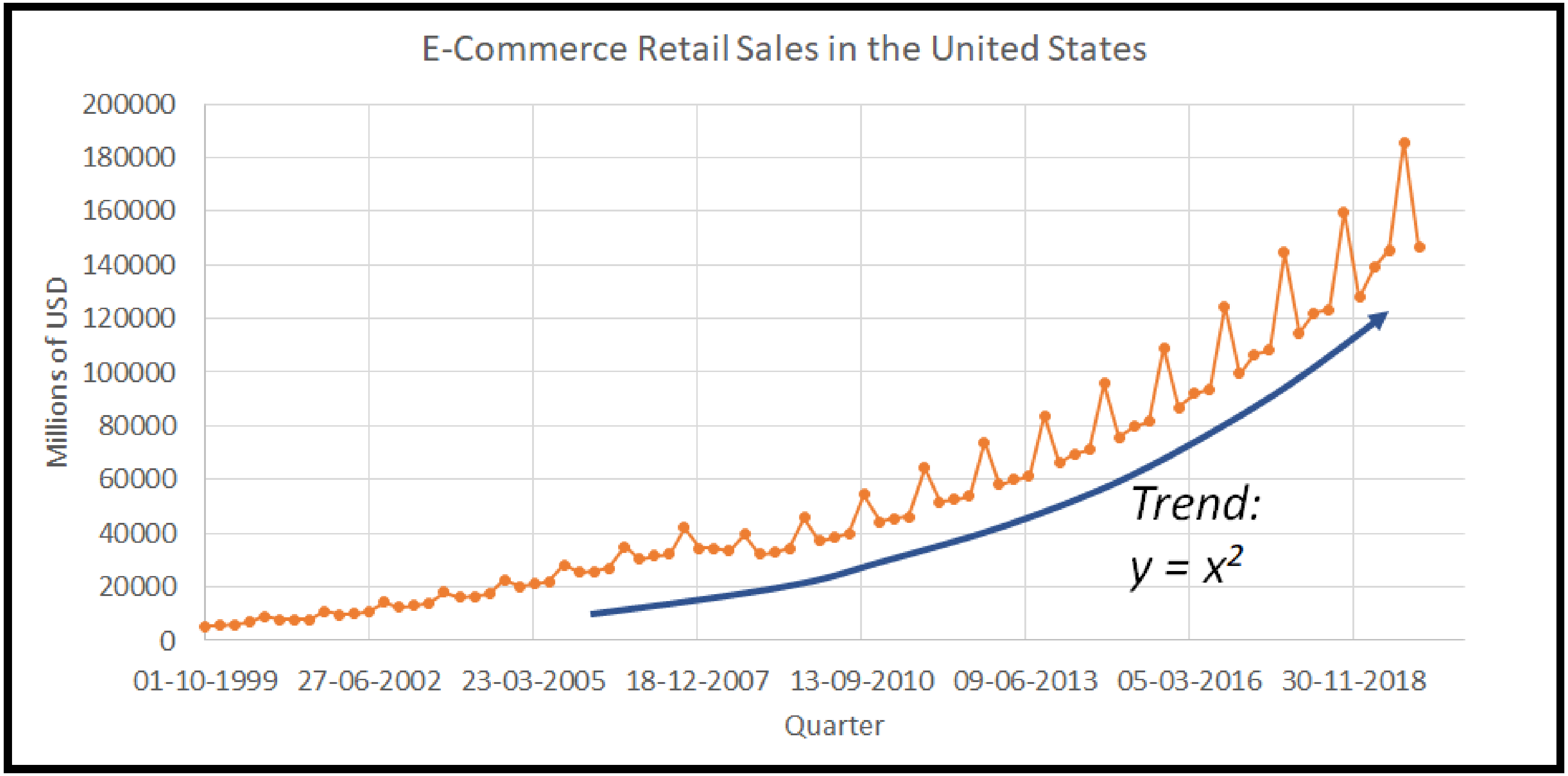
Trend - primeri

Linearni trend



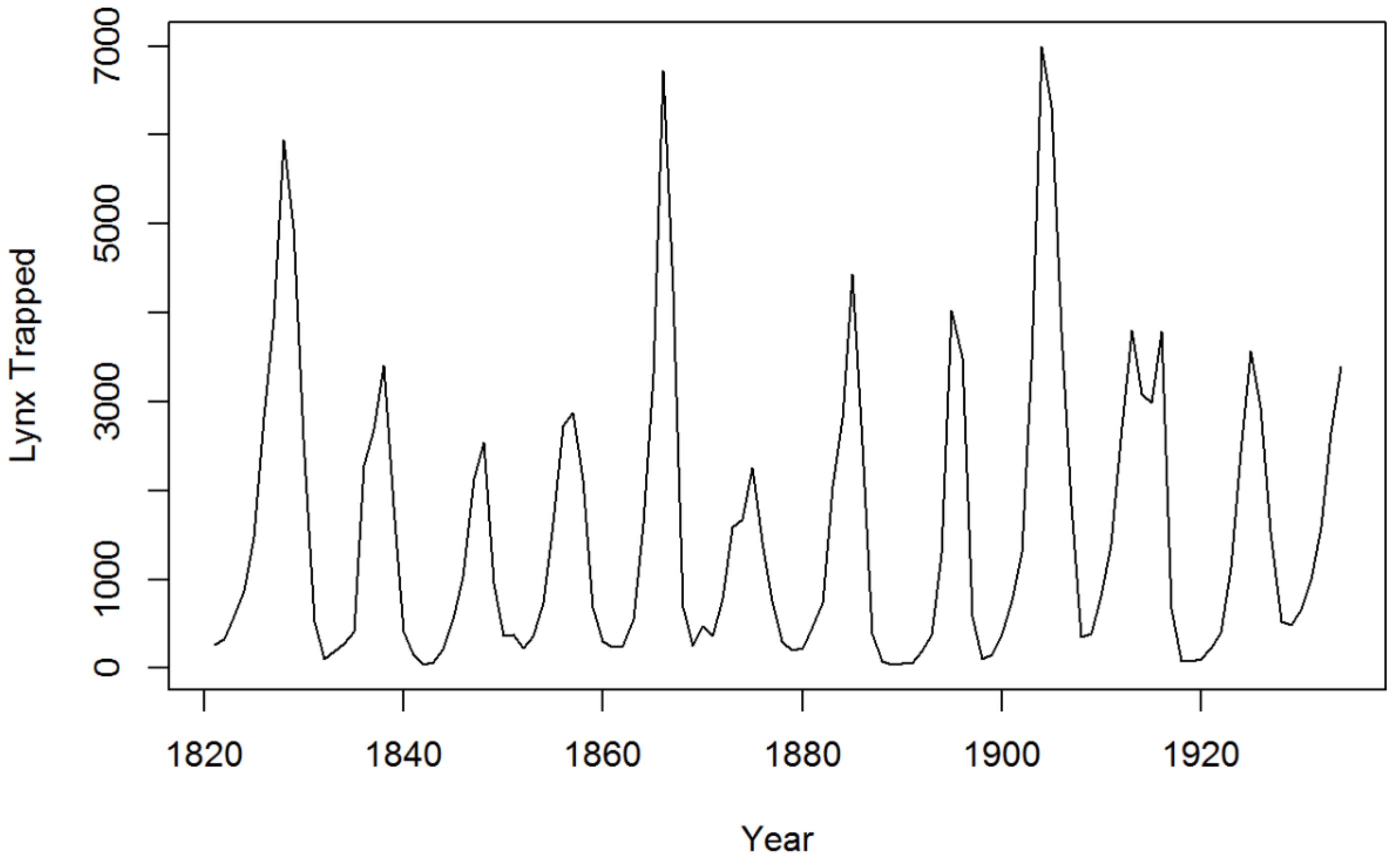
Trend - primeri

Kvadratni trend



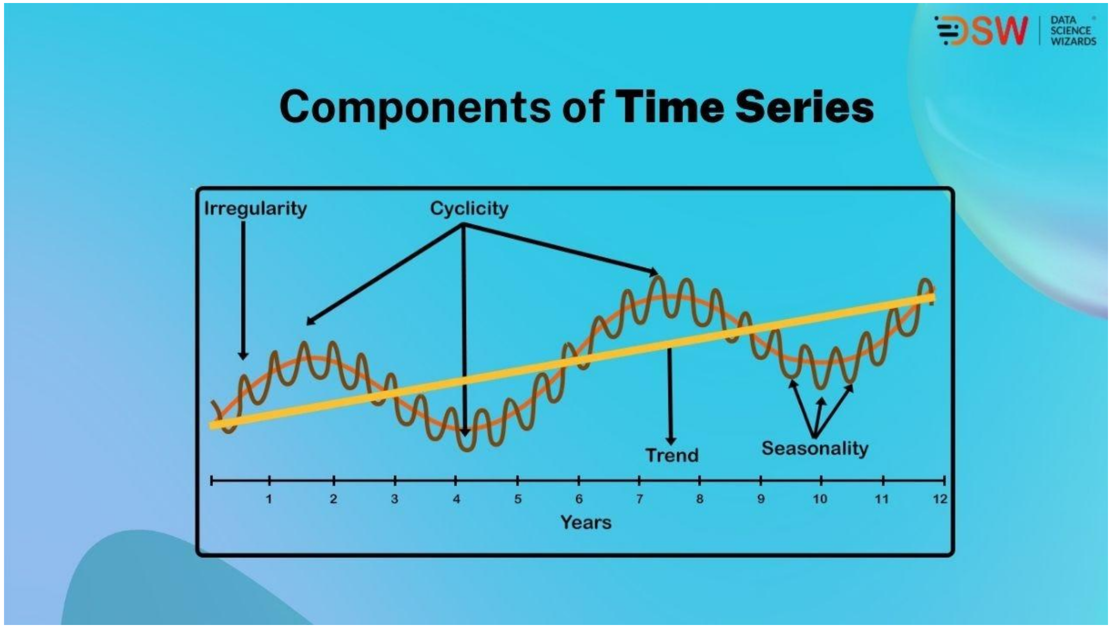
Trend - primeri

Nema trenda

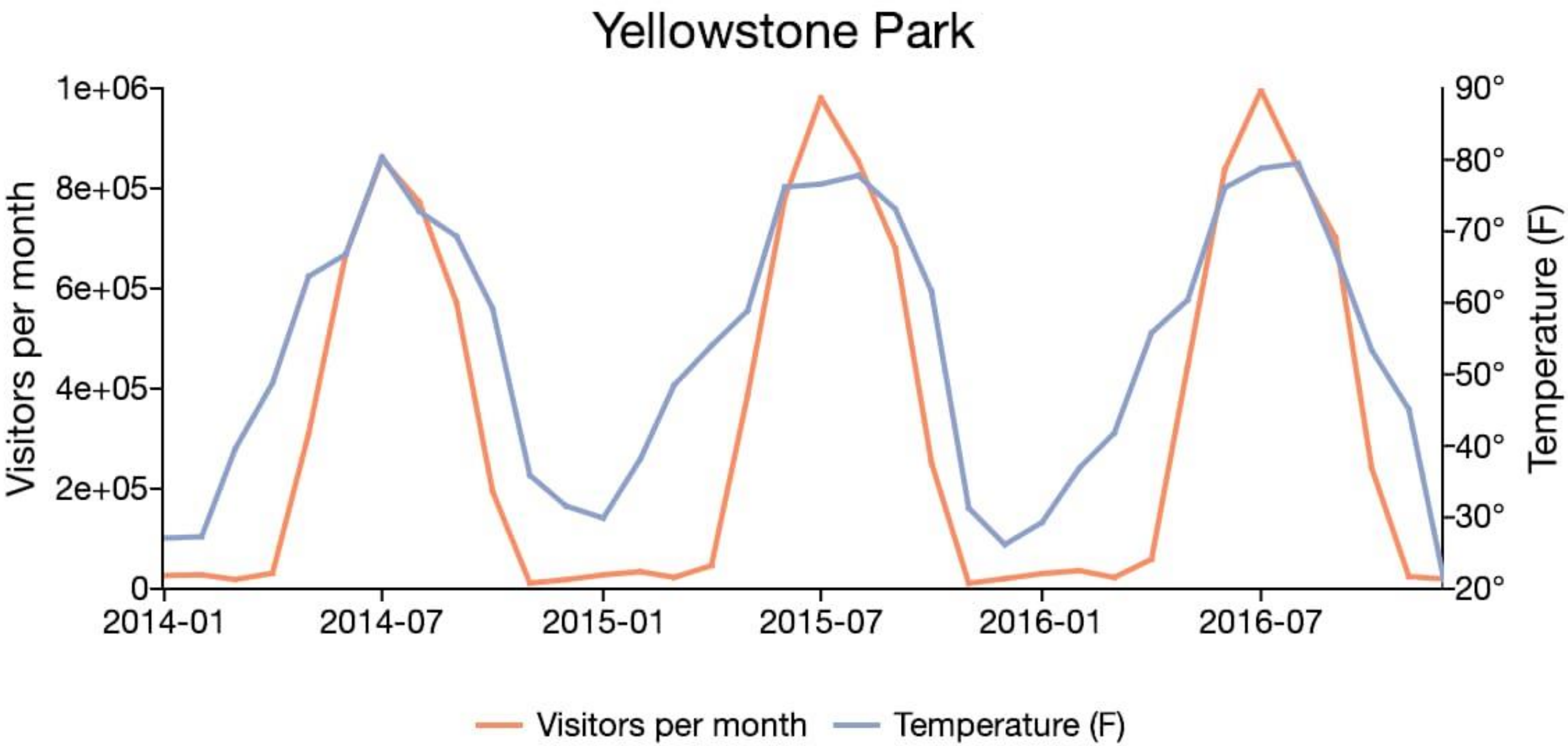


Komponente vremenskih serija – Sezonalnost

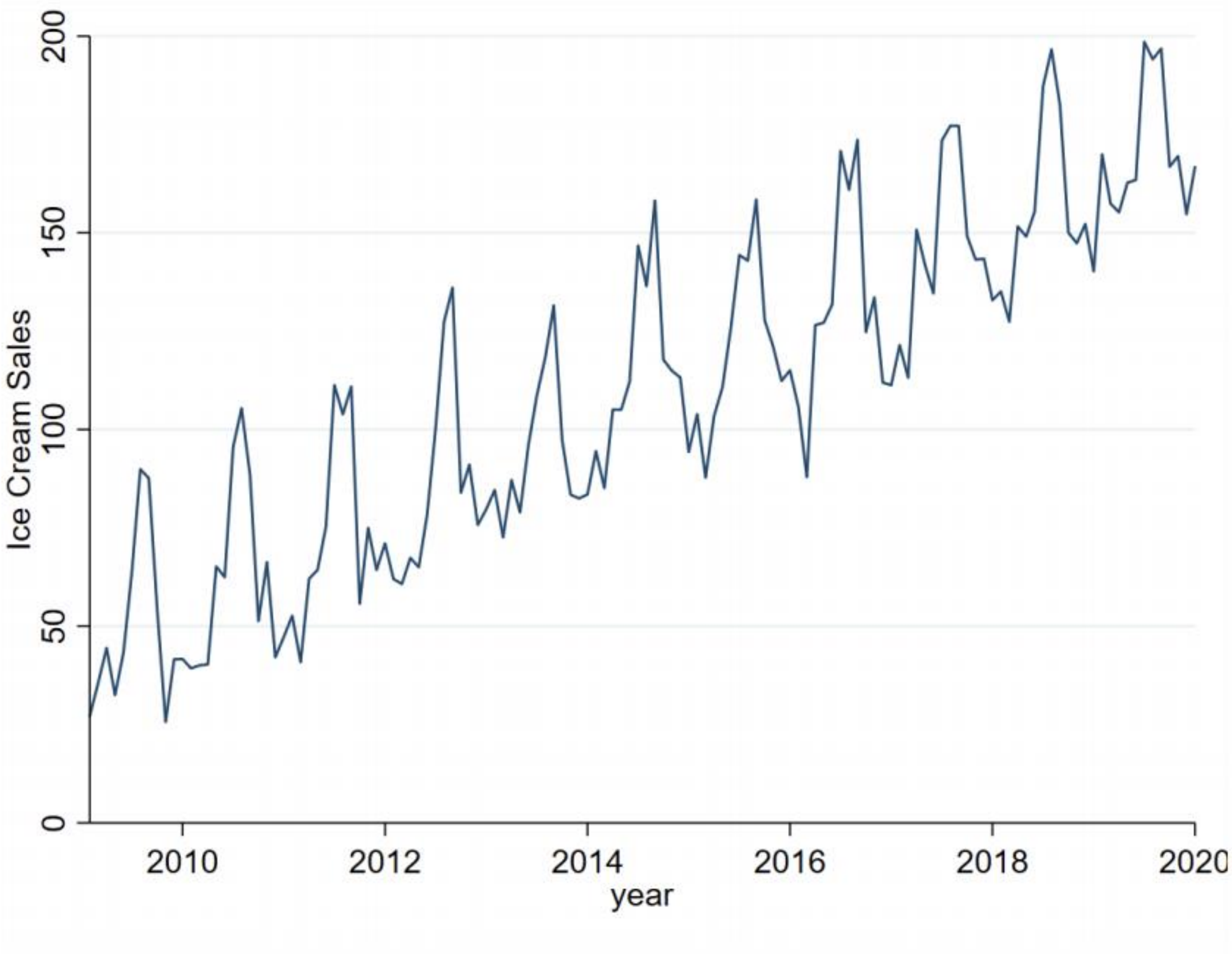
- Sezonalnost su fluktuacije vremenske serije koje se pojavljuju u kraćim vremenskim intervalima (dani ili meseci) u periodu do godinu dana.
- Sezonalnost odlikuju redovno pojavljivanje i uvek isto trajanje.
- Na primer:
  - promena temperature vazduha tokom godišnjih doba u kontinentalnoj klimi.
  - skok prodaje proizvoda svake godine za vreme novogodišnjih praznika.
  - porast prodaje sladoleda tokom proleća pa vrhunac leti, pa onda opadanje do kraja godine.



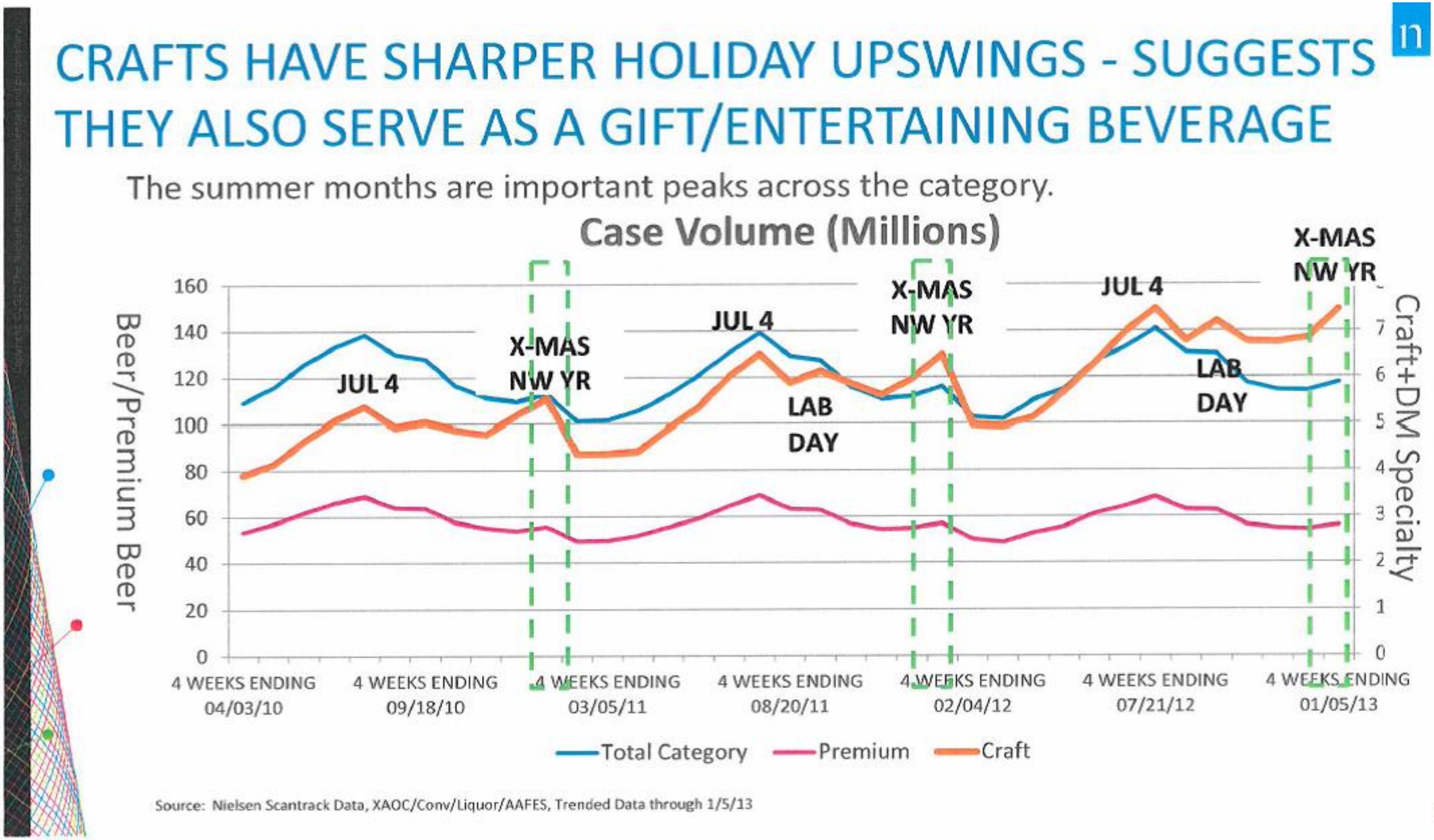
Sezonalnost - Primeri



Sezonalnost – Primeri

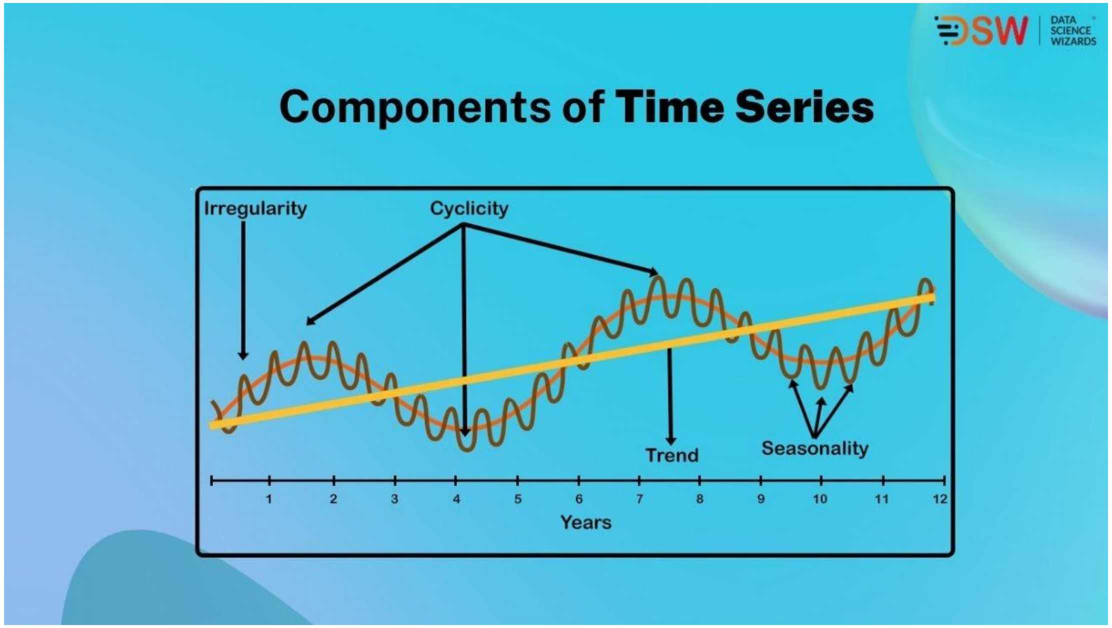


Sezonalnost – Primeri



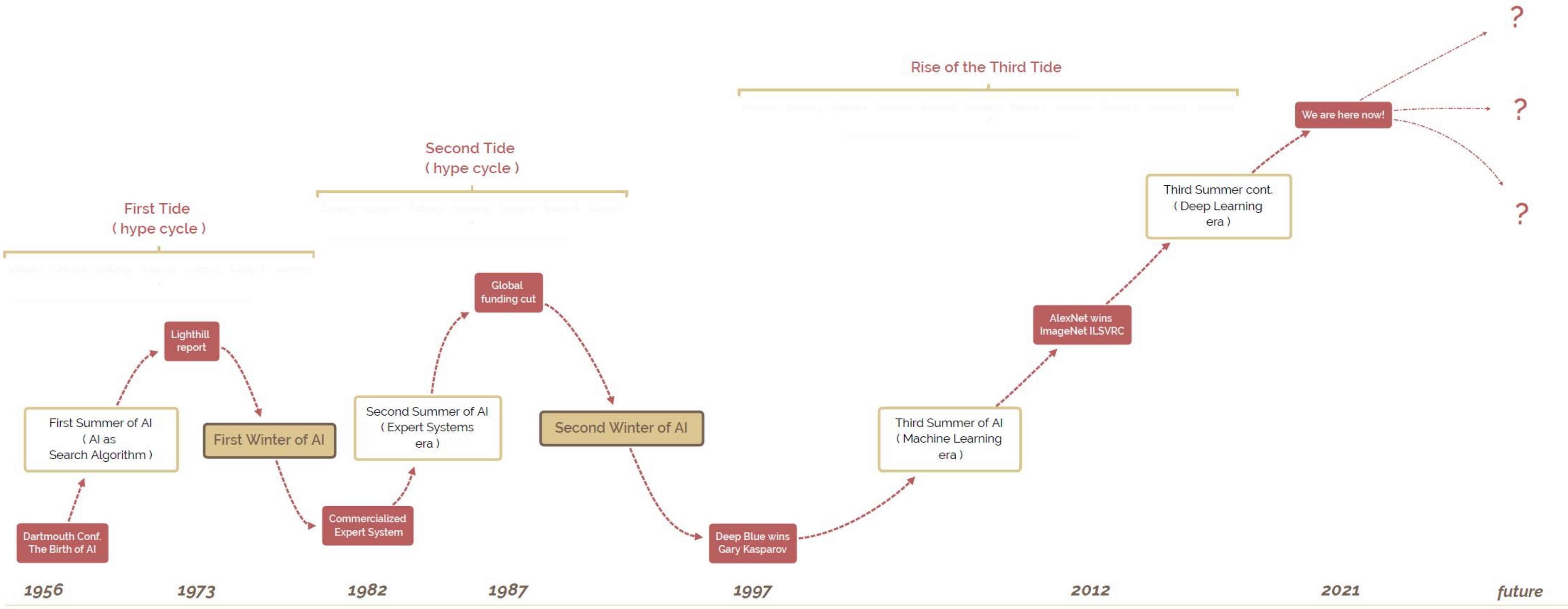
Komponente vremenskih serija – Cikličnost

- Cikličnost su fluktuacije vremenske serije koje se pojavljuju u dužim vremenskim intervalima (godine) u periodu od najmanje dve godine.
- Fluktuacija koja je ciklična ima uvek isti oblik ali se ne pojavljuje redovno i ne traje uvek isto.
- Na primer:
  - razvoj AI oblasti ima ciklus naglog razvoja (uzbuđenja), stagnacije i onda pada interesovanja (*AI winter*).
  - faze ekonomskog razvoja: ekspanzija, stagnacija, recesija i oporavak.

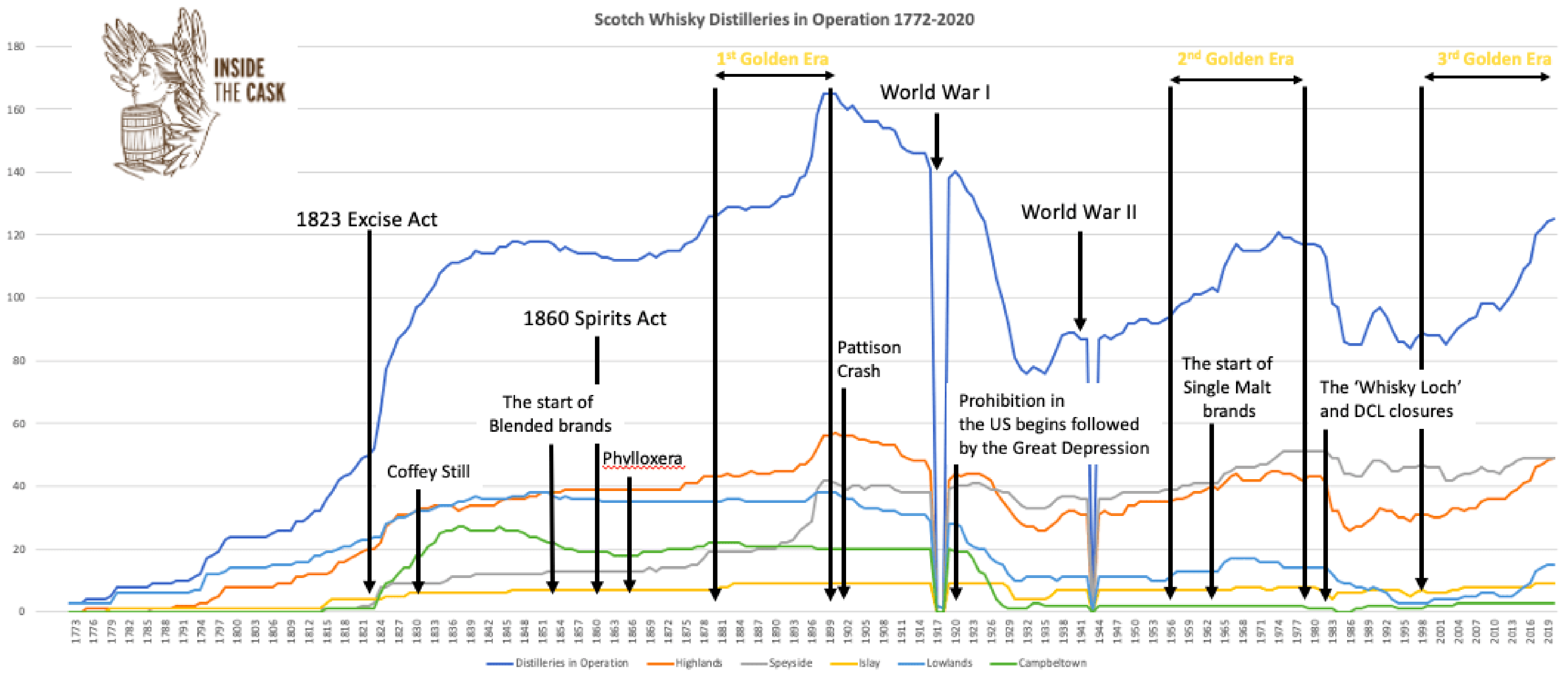


Cikličnost – Primer

Istorija razvoja veštačke inteligencije

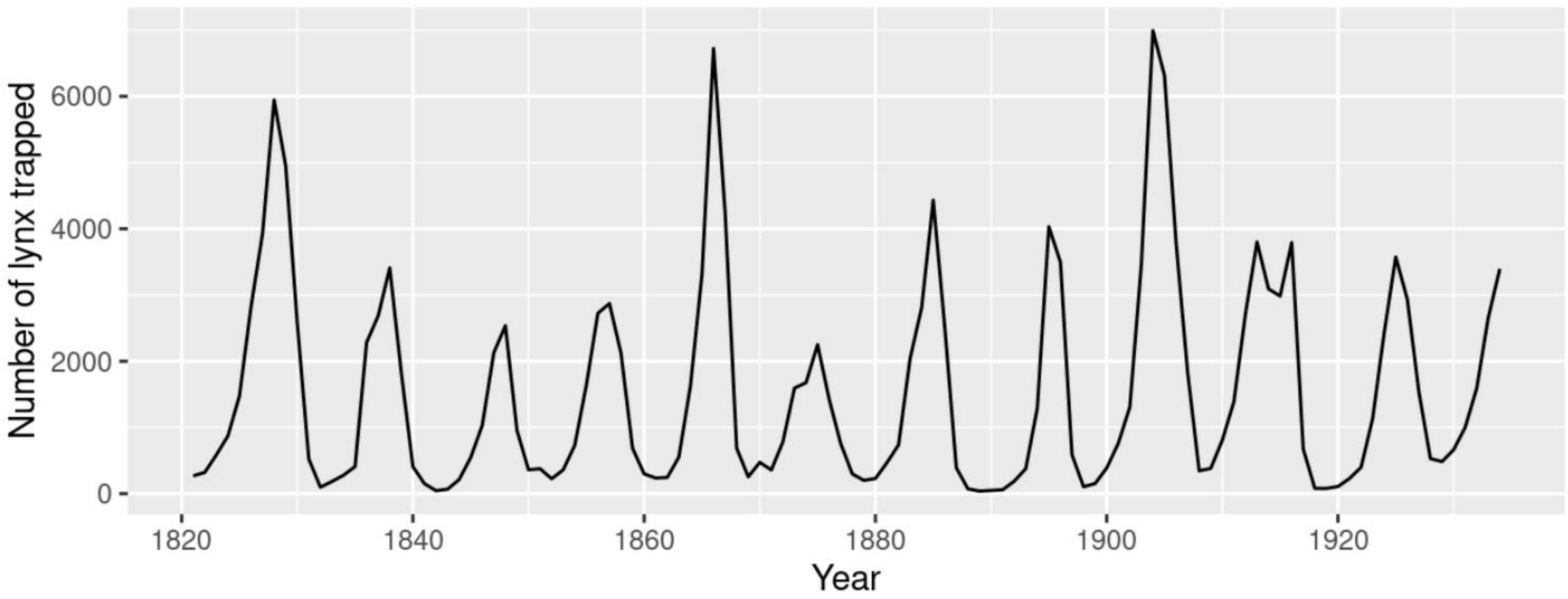


Cikličnost – Primer



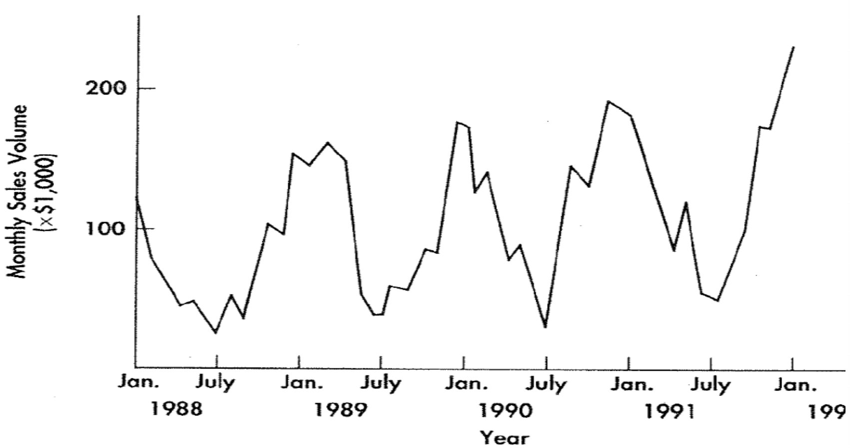
Lynx – primer cikličnosti

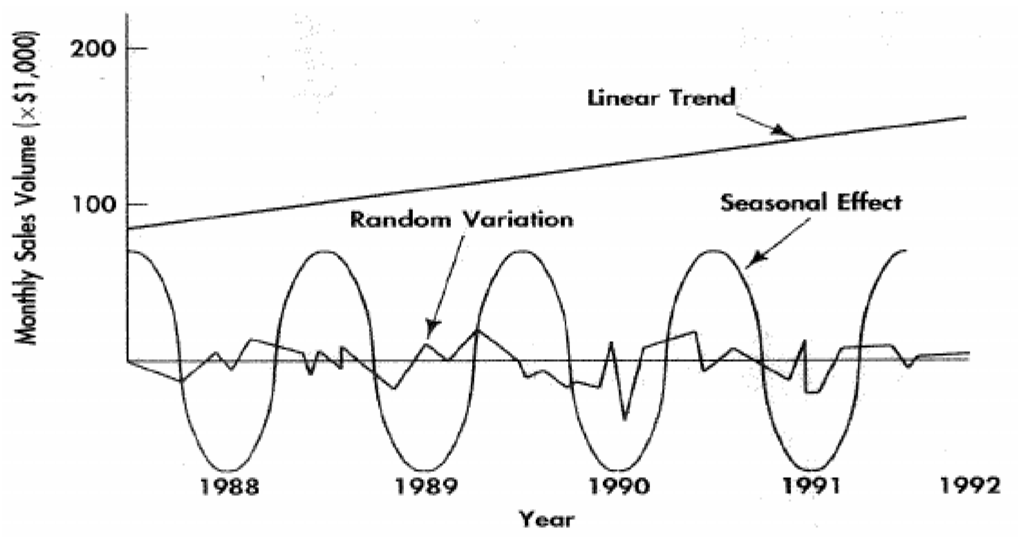
Populacija risova (*lynx*) na severu Kanade od 1821 do 1934.



Komponente vremenskih serija – Šum

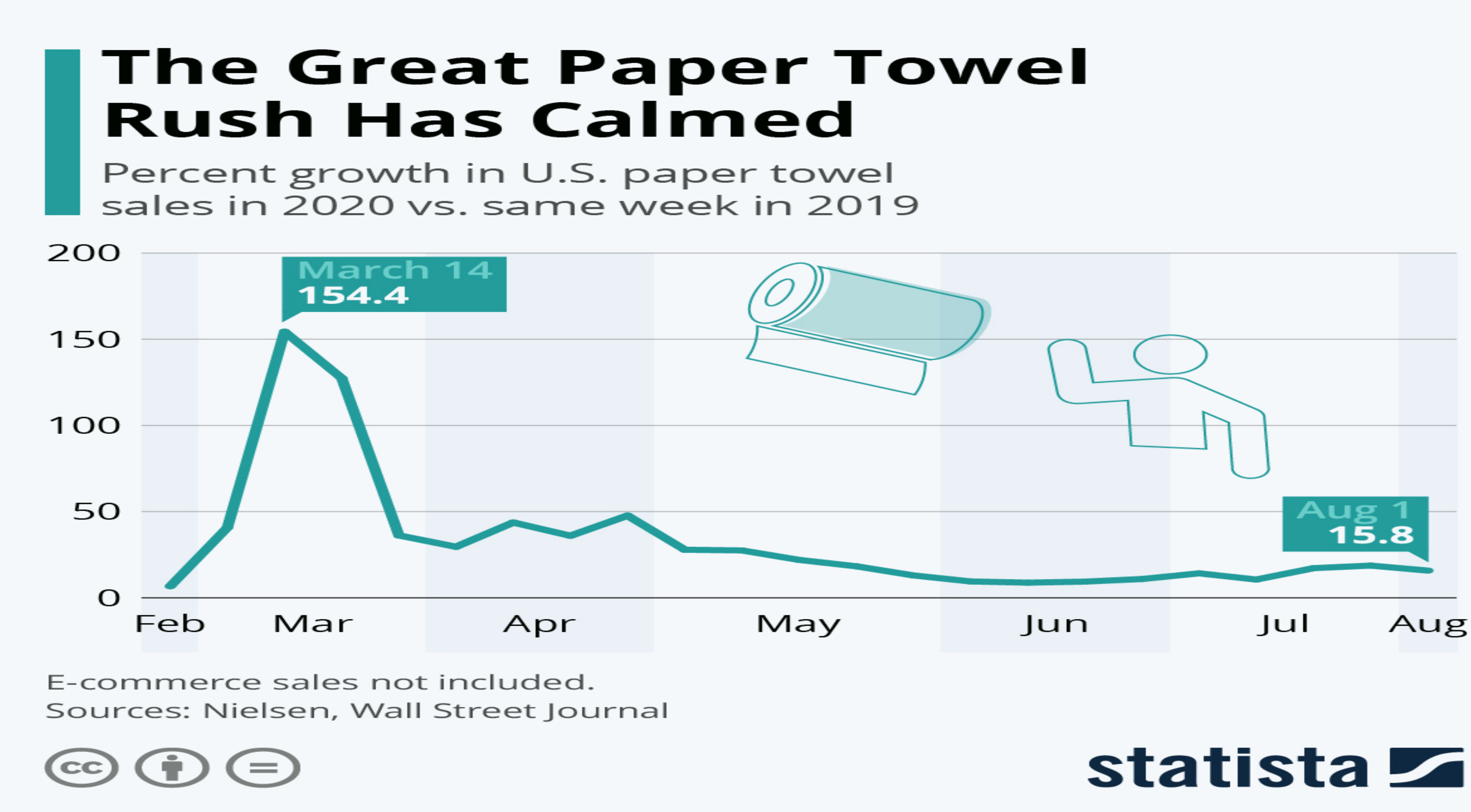
- Pod terminom šum spadaju sve fluktuacije koje nisu trend, sezonalnost ni cikličnost.
- Glavna odlika šuma je to da nema fiksni period pojavljivanja niti fiksni oblik.
- Šum može biti posledica nepredvidih pojava u okruženju ili grešaka u merenju.
- Na primer,
  - skok cena namirnica i usluga kao posledica rata u Ukrajini.
  - skok potraženje toalet papira u zapadnim zemljama za vreme pandemije.
  - pad ili rast cena akcija kao posledica tvitova poznatih osoba.





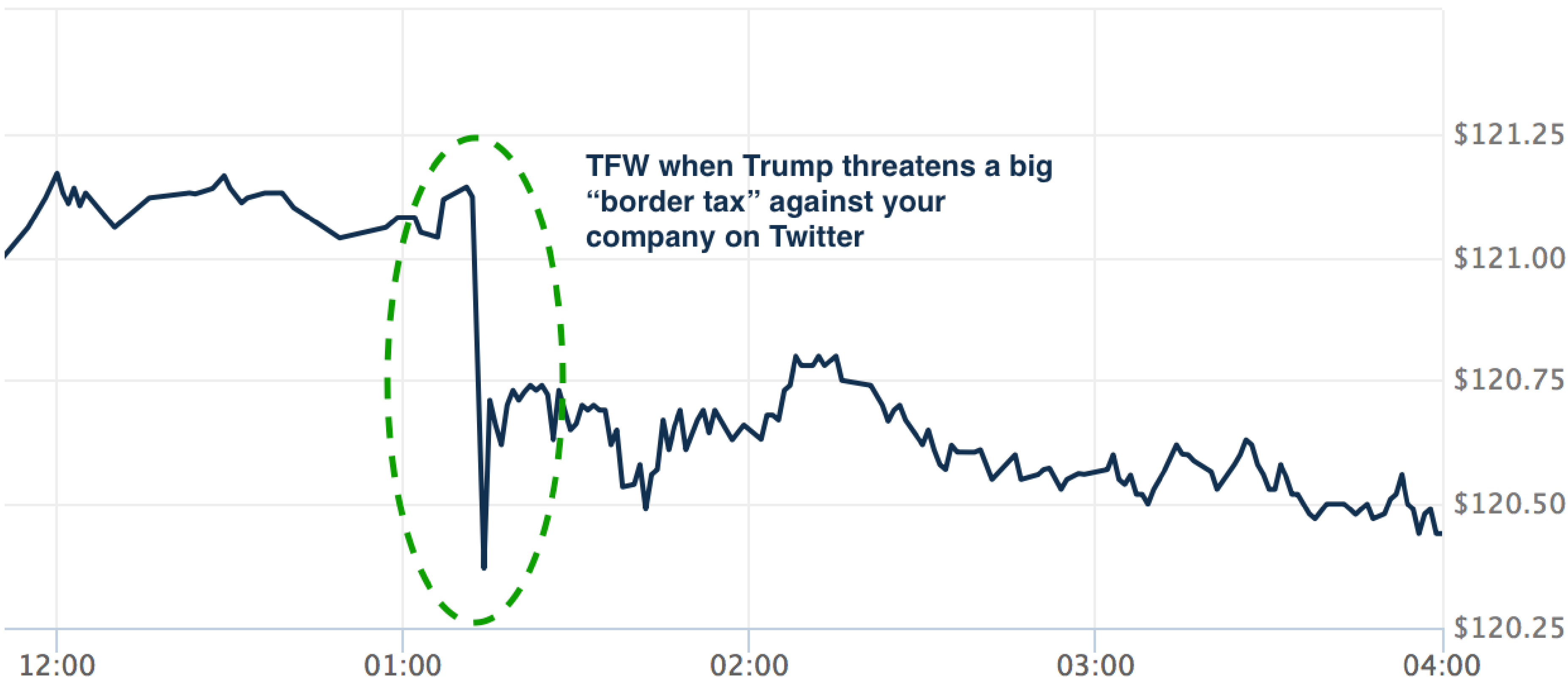
Šum - primer

Skok prodaje toalet papira u toku epidemije COVID-19.



Šum – Primer

Pad vrednosti akcija Tojote nakon jednog tvita.

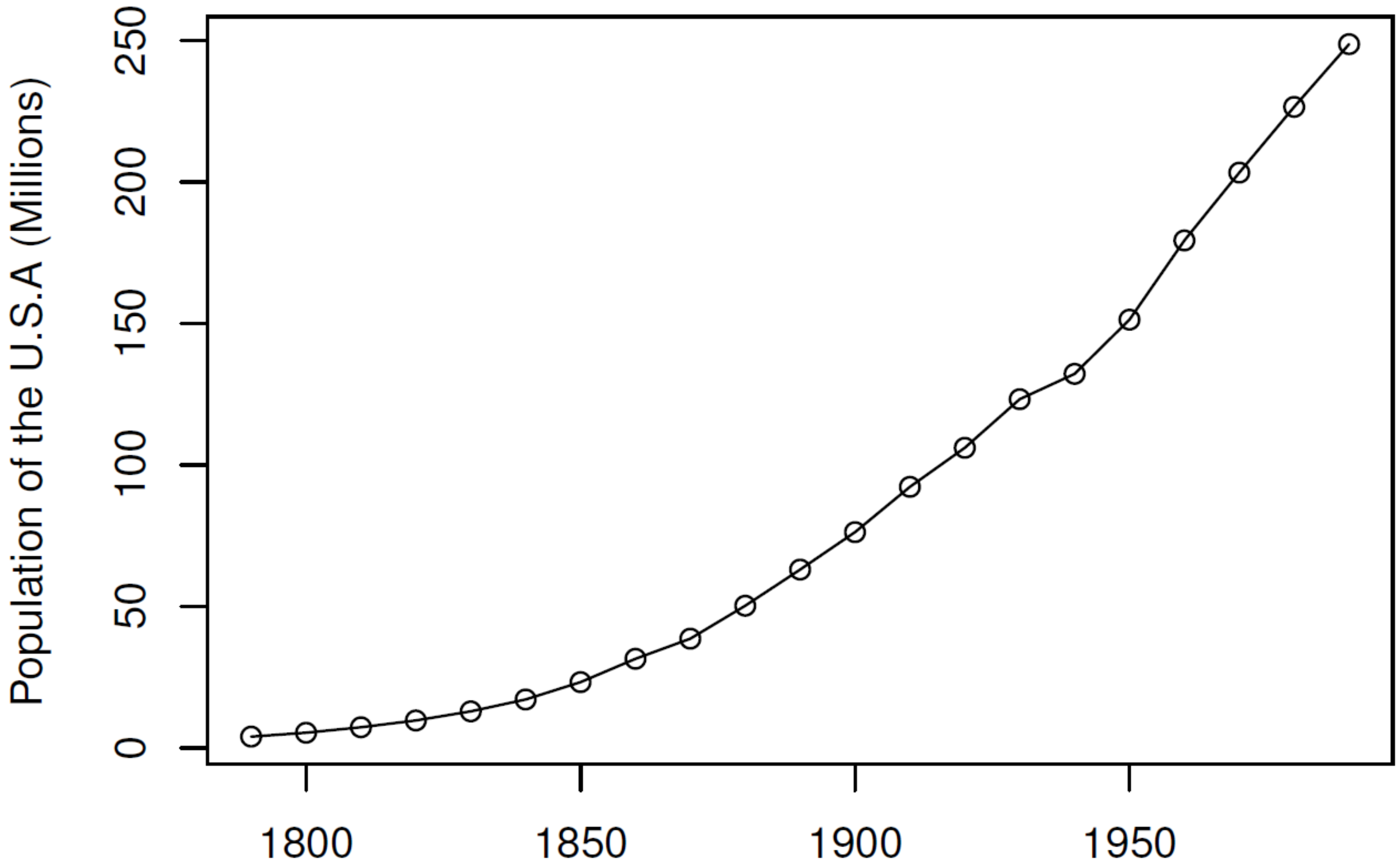


Šum – Primer

Vrednosti kriptovaluta su obično primeri vremskih serija sa puno šuma

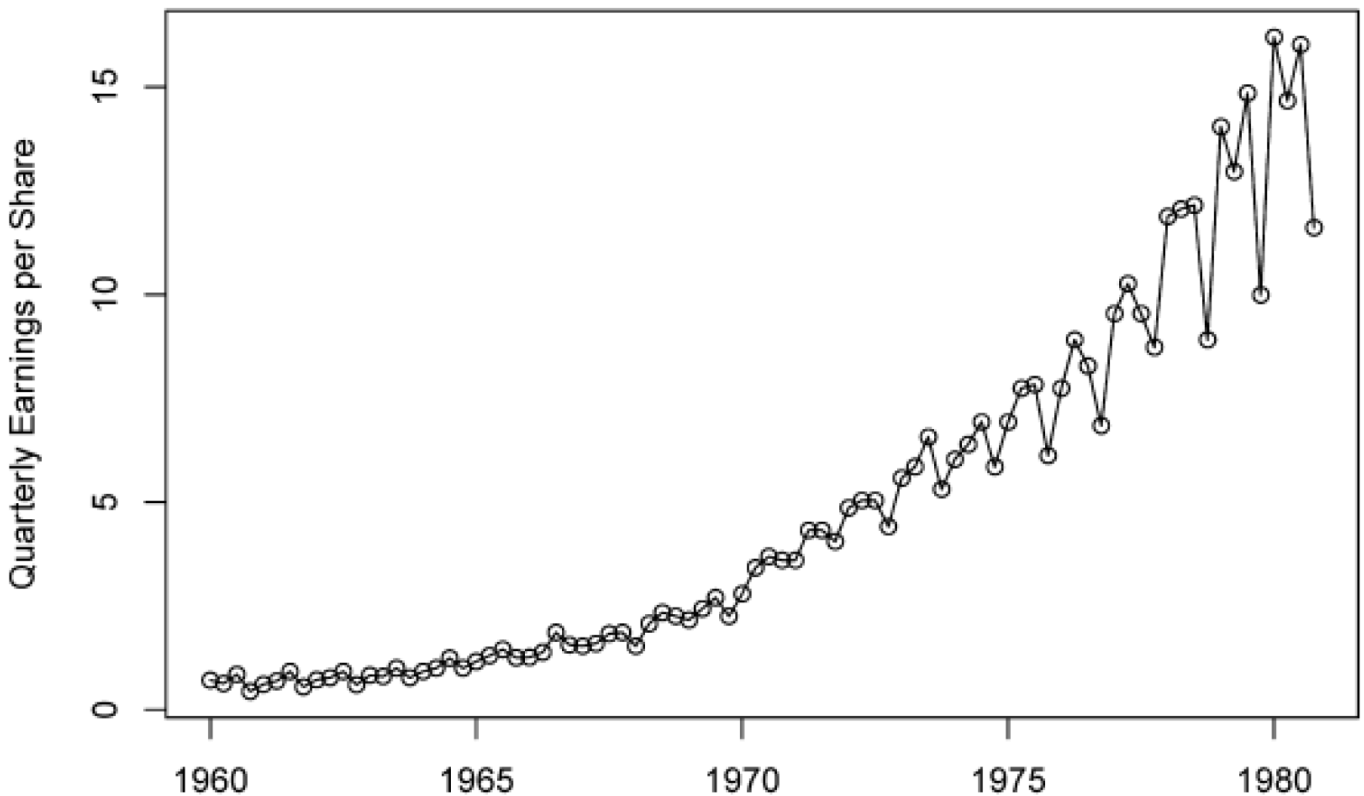
Komponente vremenskih serija – Primeri

- Populacija USA od 1790 – 1990.
- Vidi se nelinearan trend porasta i blaga promena oblika vremenske serije.



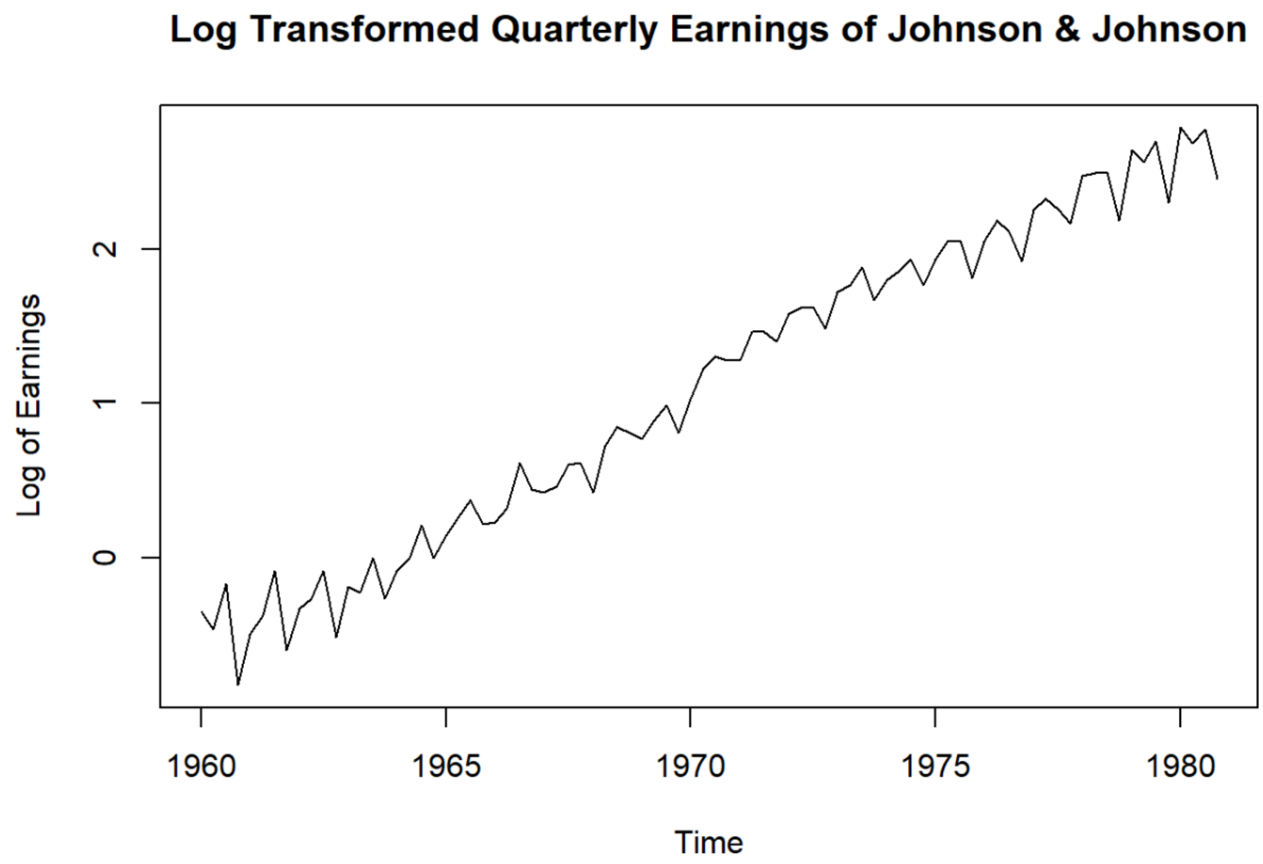
Komponente vremenskih serija – Primeri

- 84 kvartala (21 godina) zarade kompanije „Johnson & Johnson“ od 1960.
- Vidi se trend porasta, varijabilnost takođe raste.
- Sezonalnost se nazire, ali se fluktuacije u prvom delu grafika ne vide dobro.
  - Da bi to korigovali uradićemo logaritmasku transformaciju podataka – sledeći slajd.



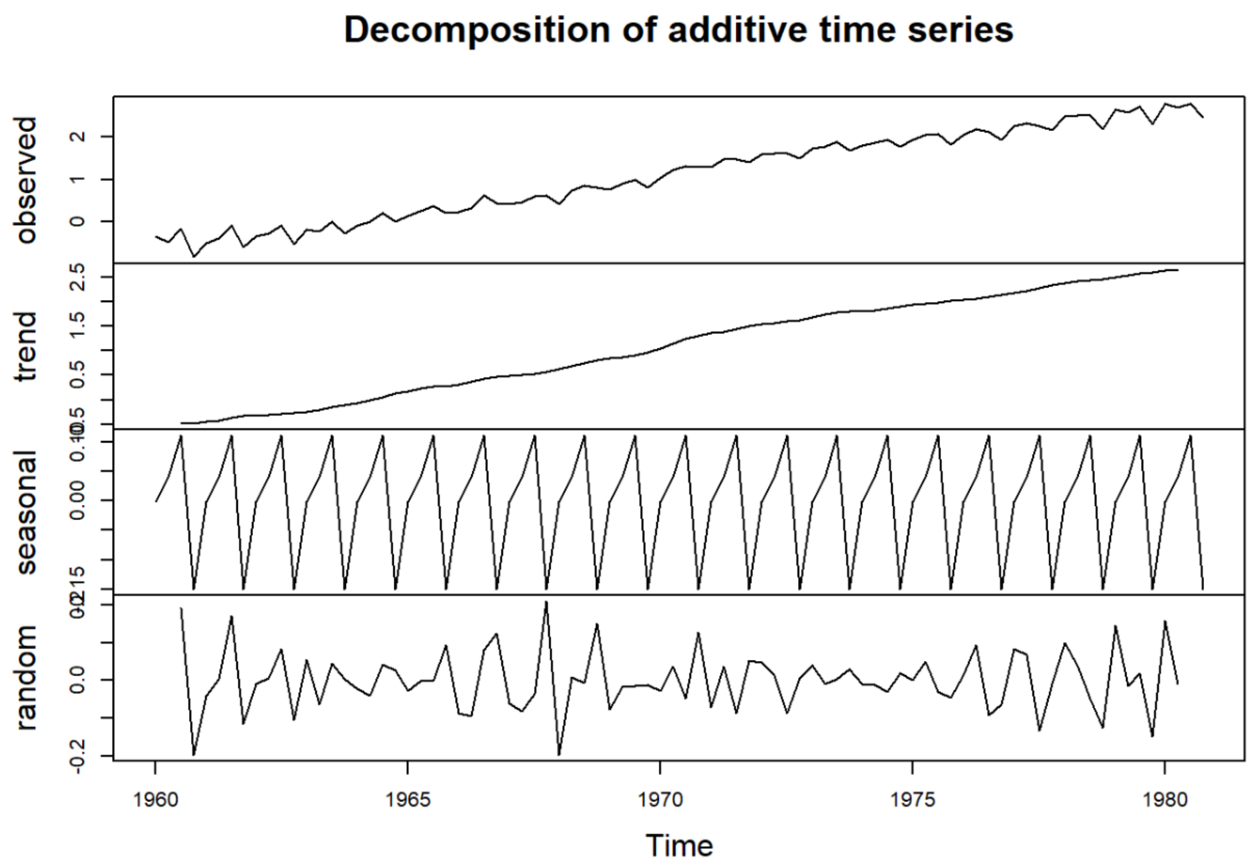
Komponente vremenskih serija – Primeri

- Sada je sezonalnost uočljiva.
- U nekim četvrtinama su prisutni skokovi, a u nekim padovi zarade. Taj šablon ponavlja se u svakoj godini.
- Na sledećem slajdu prikazane su sve komponente ove vremenske serije.



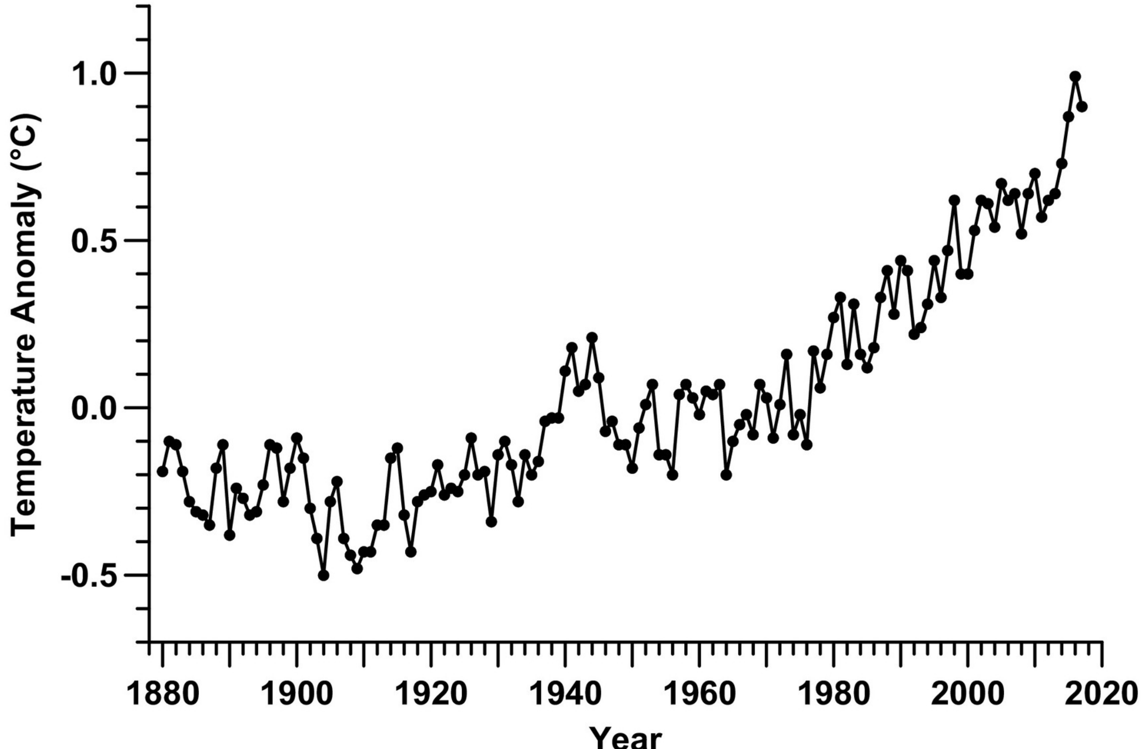
Komponente vremenskih serija – Primeri

- Na grafiku je data originalna vremenska serija kao i trend, sezonalnost i šum koji su dobijeni dekompozicijom (rasklapanje) na komponente.
- Tehnikama rasklapanja bavimo se kasnije tokom kursa.



Komponente vremenskih serija – Primeri

- Globalno zagrevanje – Na x-osi su meseci i godine. Na y-osi je *anomalija temperature*:
  - za svaki mesec izračunata je razlika temperature okeana u tom mesecu (i odgovrajućoj godini) u odnosu na srednju vrednost za taj isti mesec u periodu od 1951 do 1980.
- Očigledan rastući trend, naročito u drugoj polovini XX veka koristi se kao jedan od indikatora globalnog zagrevanja.
- Sa obzirom na to da se temperatura meri mesečno sezonalnost je prisutna i posledica je godišnjih doba.



Modelovanje vremenske serije pomoću komonentni

- Vremenska serija može se modelovati kao interakcija svojih komponenti.

- Postoje dva osnovna modela:

- Aditivni

- Multiplikativni

Aditivni model

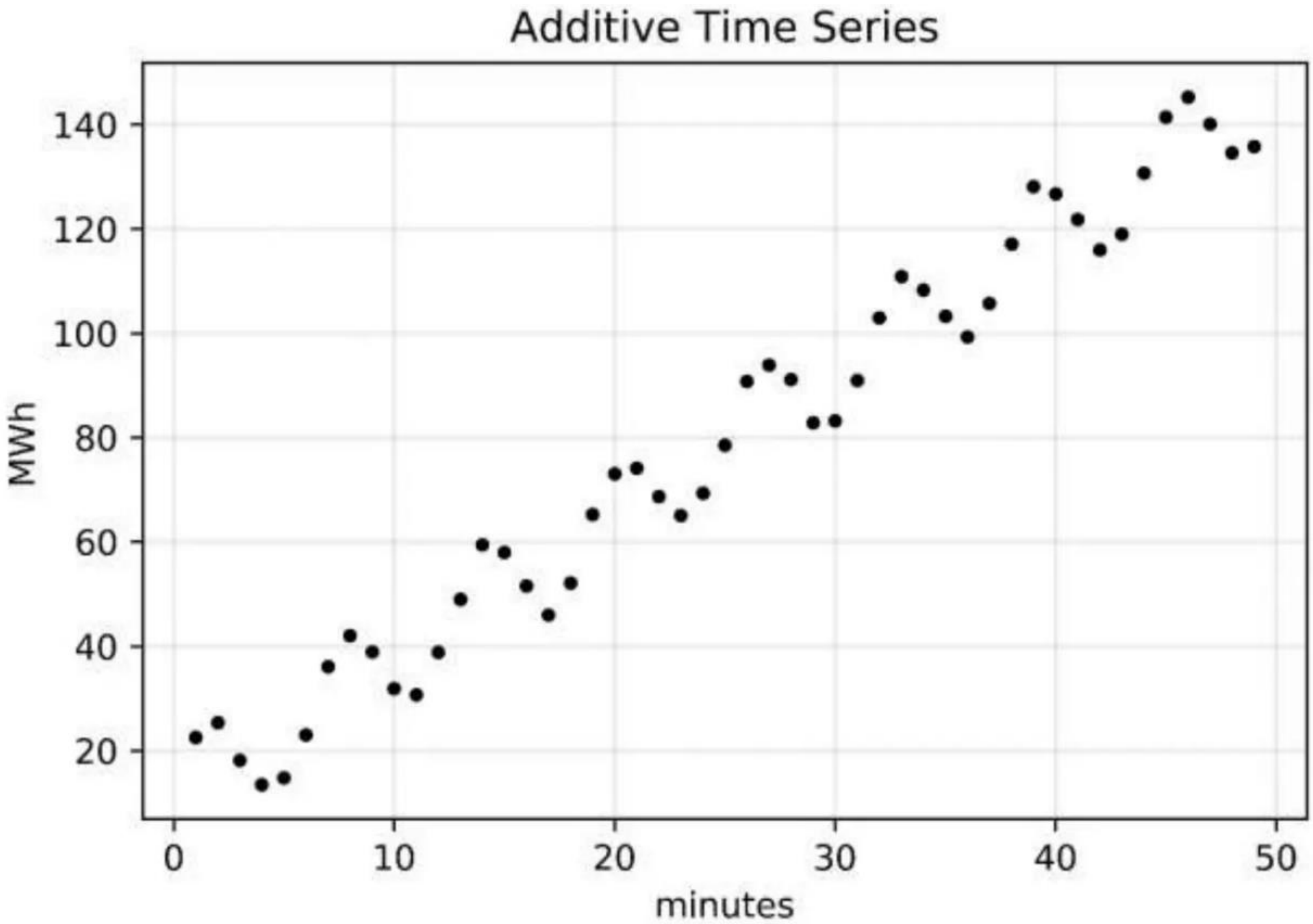
- Pretpostavka je da su podaci u vremenskoj seriji dobijeni kao zbir komponenti.

opservacija = Trend + Sezonalnost + Cikličnost + Šum

$$Y_t = T_t + S_t + C_t + R_t$$

- Koristi se kad jačina (magnituda) sezonalnosti, cikličnosti i šuma ne zavisi od trenda.

Aditivni model – Primer



Multiplikativni model

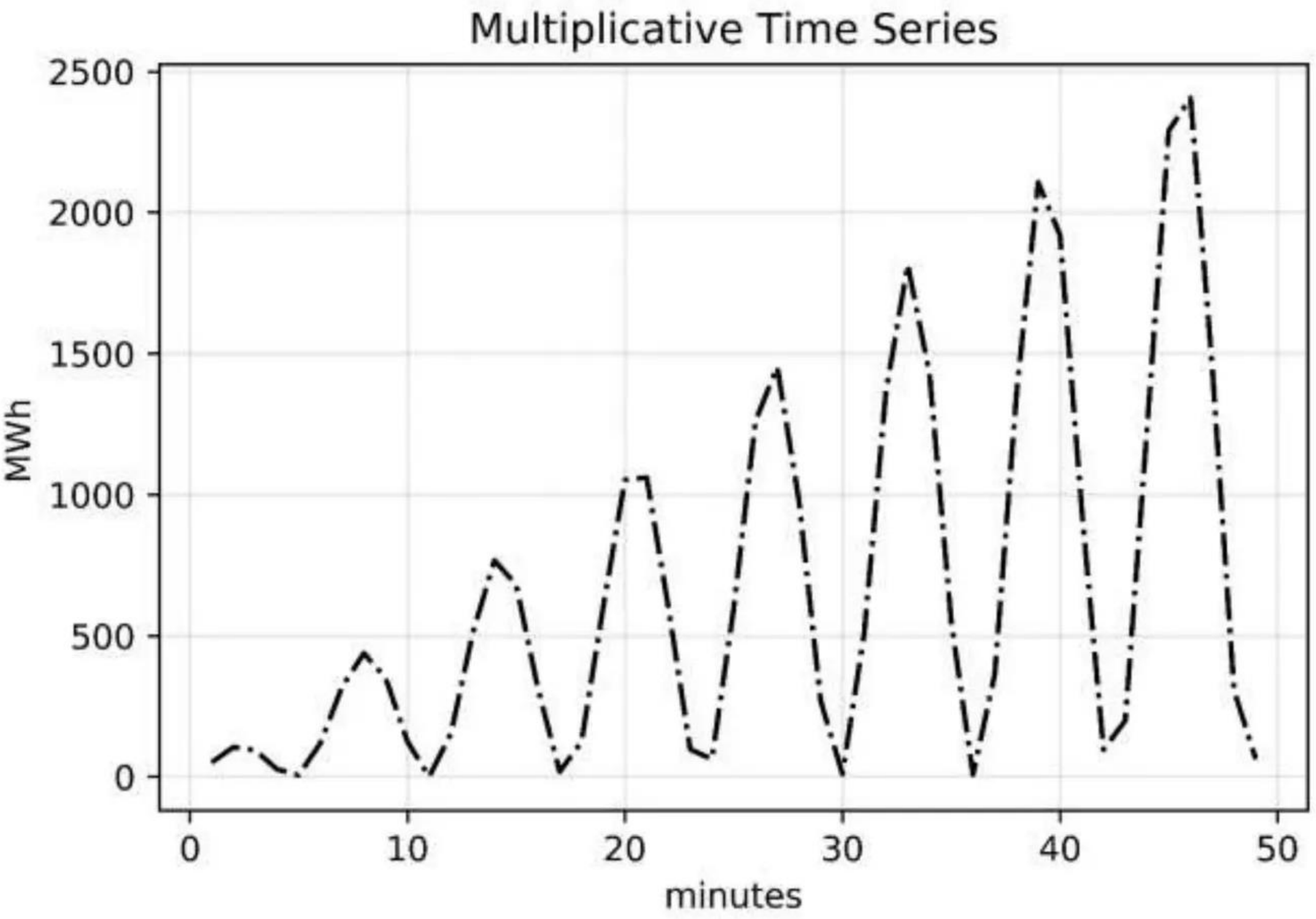
- Pretpostavka je da su podaci u vremenskoj seriji dobijeni kao proizvod komponenti.

opservacija = Trend * Sezonalnost * Cikličnost * Šum

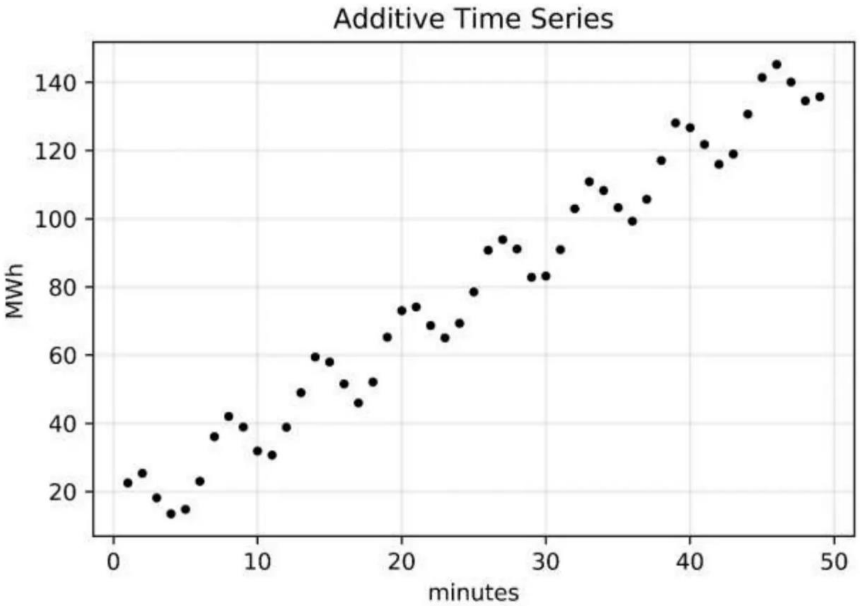
$$Y_t = T_t * S_t * C_t * R_t$$

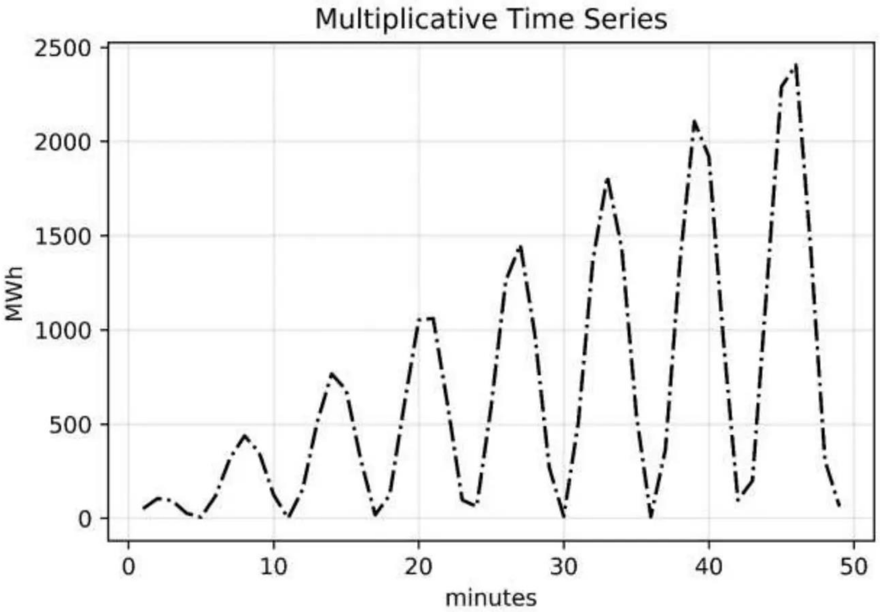
- Koristi se kad jačina (magnituda) sezonalnosti, cikličnosti i šuma zavisi od trenda, odnosno magnituda ovih komponenti se menaj kako se menja trend.

Multiplikativni model – Primer



Multiplikativni vs. Aditivni





Magnituda sezonalnosti, cikličnosti i šuma ne zavisi od trenda.

Magnituda sezonalnosti, cikličnosti i šuma zavisi od trenda.

Konverzija multiplikativnog u aditivni model

- Multiplikativni model se može lako konvertovati u aditivni primenom logaritmovanja:

$$Y_t = T_t * S_t * C_t * R_t$$

$$\log(Y_t) = \log(T_t) + \log(S_t) + \log(C_t) + \log(R_t)$$

- Ova osobina je korisna kada želimo da na vremensku seriju kojoj odgovara multiplikativni model primenimo tehniku koja je prikladna za aditivni.
  - Više o svemu u nastavku predavanja.

Značaj modela interakcije komponenti

- Određivanje načina interakcije komponenti značajno je za dekompozciju vremenske serije.
- Ako ne znamo na koji način je vremenska serija sačinjena od svojih komponenti onda ne možemo pravilno da izdvojimo te komponente.
- Izdvajanje komponenti značajno je i za deskriptivni i za prediktivni aspekt analize vremenskih serija.

Određivanje modela interakcije komponenti

- Ne postoji univerzalna metodologija.
- Analiza grafikona vremenske serije je ključna.
- Na osnovu grafikona se obično može utvrditi koji model je prikladniji.
- Nakon odabira modela vrši se dekompozicija po odabranom modelu.
- Analizom rezultata dekompozicije može se utvrditi da li je izabran odgovarajući model.
- Napomena: postoje i različite kombinacije i modfikacije aditivnog i multiplikativnog modela (npr. pseudo-aditivni model).
  - Obično su prikladne za vremenske serije koje imaju vrlo specifičan oblik.
  - Na ovom kursu se nećemo baviti drugim modelima.

Dekompozicija vremenske serije

- Postoji veliki broj metodologija za dekompoziciju i svi uključuju sledeće osnovne korake:

1. Određivanje trenda
2. Uklanjanje trenda
3. Određivanje sezonalnosti
4. Određivanje šuma uklanjanjem sezonalnosti

- Često se koraci 1. – 3. iterativno ponavljaju nekoliko puta.
- Sezonalnost dobijena u koraku 3. može se oduzeti od orginalne serije pa se tek onda radi korak 1.
  - Motivacija postupka je u tome da će trend biti preciznije određen iz vremenske serije koja nema sezonalnost.

Određivanje trenda

- Postoje dve grupe metoda sa puno varijacija:

- Metode pokretnog proseka (*Moving Average*, MA)
  - Aproksimacija trenda pomoću pomerajućeg prozora (vremenskog intervala) u kome određujemo prosek vrednosti sa *y*-ose. Detaljno obrađujemo u nastavku.

- Metode uklapanja krivih (*curve fitting*)
  - Globalne metode
    - Uklapanje iste krive nad celim rasponom x-ose.
      - Na primer, linearna regresija pravom ili polinomom.
  - Lokalne metode
    - Uklapanje posebnih krivih za svaki podeok x-ose.
      - Na primer, *Lowess*, interpolacija splajnom itd.

Određivanje trenda – pokretni prosek

- Određivanje proseka *k* uzastopnih vrednosti vremenske serije.
  - *k* je veličina prozora koju određuje korisnik

- Postoje tri osnovna tipa pokretnog proseka:

❑ Jednostavni pokretni prosek (*Simple Moving Average*, SMA)

❑ Pokretni prosek sa težinama (*Weighted Moving Average*, WMA)

❑ Eksponencijalni pokretni prosek (*Exponential Moving Average*, EMA)

- Spada u tzv. izravnjavajuće (*smoothing*) metode.

Određivanje trenda – Jednostavni pokretni prosek

- Predstavlja prosek *k* uzastopnih vrednosti vremenske serije:

$$SMA_k = \frac{p_1 + p_2 + \cdots + p_k}{k}$$

- gde je p_i *i*-ta vrednost u prozoru od *k* vrednosti.

- Na primer, ako imamo vremensku seriju veličine *6* za *k*=*3* imamo:

$$SMA_{3,1} = \frac{p_1 + p_2 + p_3}{3}$$

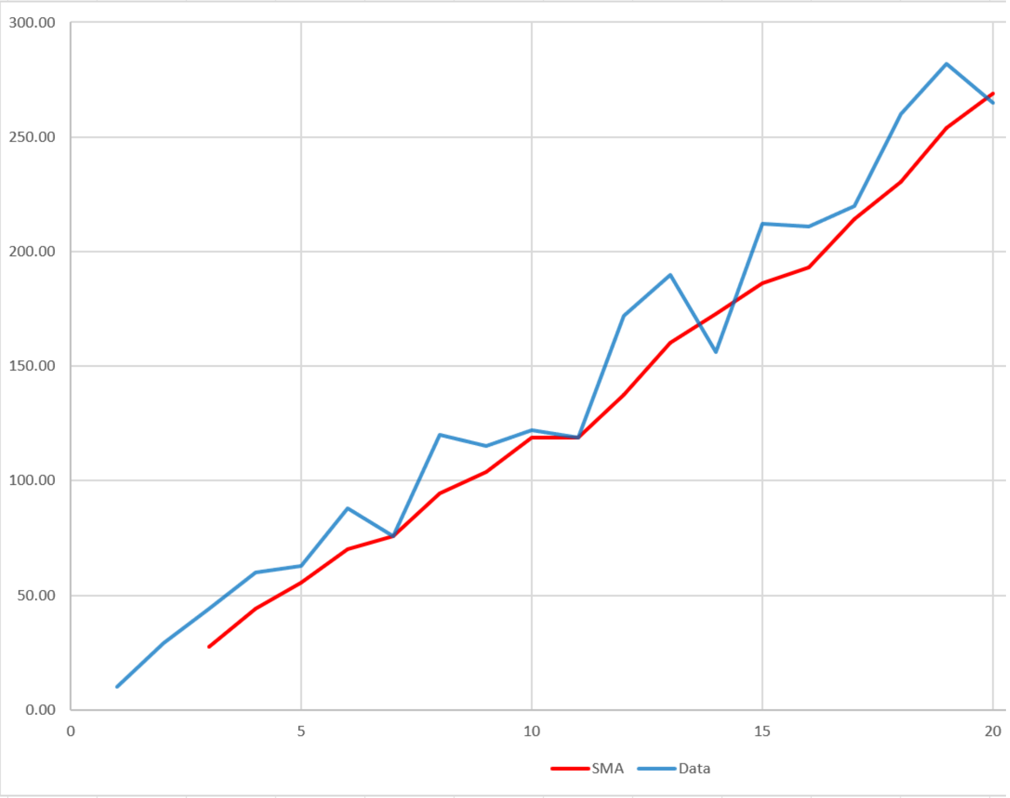
$$SMA_{3,2} = \frac{p_2 + p_3 + p_4}{3}$$

$$SMA_{3,3} = \frac{p_4 + p_5 + p_6}{3}$$

Određivanje trenda – Jednostavni pokretni prosek

- Primer SMA_3 za vremensku seriju veličine 20:

Data	SMA
10	
29	
44	27.66667
60	44.33333
63	55.66667
88	70.33333
76	75.66667
120	94.66667
115	103.6667
122	119
119	118.6667
172	137.6667
190	160.3333
156	172.6667
212	186
211	193
220	214.3333
260	230.3333
282	254
265	269



$$SMA_{3,1} = \frac{10 + 29 + 44}{3} = 27.67 \qquad SMA_{3,2} = \frac{29 + 44 + 60}{3} = 44.33$$

Određivanje trenda – Jednostavni pokretni prosek

- Jednako vrednuje skorije i davne vrednosti u vremenskoj seriji.

- Iz tog razloga je:
  - Dobar izbor za određivanje trenda na duži period.
  - Dobar izbor za uklanjanje šuma (šum će biti „izravnat“ prosekom).
  - Loš izbor za trend u kraćim vremenskim periodima.

Određivanje trenda – Jednostavni pokretni prosek

- Prilikom implementacije koristi se sledeća formula:

$$SMA_{k,trenutni} = SMA_{k,prethodni} + \frac{1}{k}\left(p_{trenutno} - p_{prethodno-k+,1}\right)$$

$$SMA_{3,2} = SMA_{3,1} + \frac{1}{3}(p_4 - p_{3-3+1}) = SMA_{3,1} + \frac{1}{3}(p_4 - p_1)$$

Određivanje trenda – Pokretni prosek sa težinama

- Modfikacija SMA kod koje su skorijim vrednostima dodeljene veće težine.
- Težine su celi brojevi koji odogvaraju poziciji elementa u prozoru.

$$WMA_k = \frac{1 \cdot p_1 + 2 \cdot p_2 + \cdots + k \cdot p_k}{1 + 2 + \cdots + k}$$

- gde je p_i *i*-ta vrednost u prozoru od *k* vrednosti.
- Na primer, ako imamo vremensku seriju veličine *6* za *k=3* imamo*:*

$$WMA_{3,1} = \frac{p_1 + 2p_2 + 3p_3}{3}$$

$$WMA_{3,2} = \frac{p_2 + 2p_3 + 3p_4}{3}$$

$$WMA_{3,3} = \frac{p_4 + 2p_5 + 3p_6}{3}$$

Određivanje trenda – Pokretni prosek sa težinama

- Ako težine posmatramo individualno one imaju sledeće vrednosti:

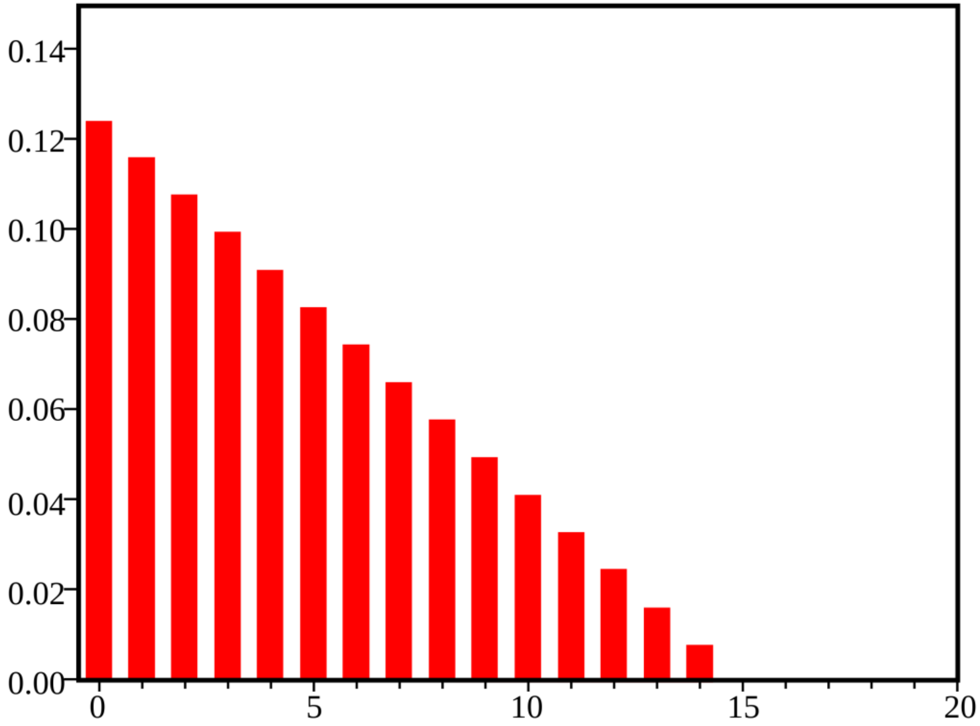
$$W_{k,1} = \frac{1}{1+2+\cdots+k} \quad W_{k,2} = \frac{2}{1+2+\cdots+k} \quad W_{k,k} = \frac{k}{1+2+\cdots+k}$$

- Na primer za *k=3* imamo*:*

$$W_{3,1} = \frac{1}{1+2+3} = 0.17 \quad W_{3,2} = \frac{2}{6} = 0.33 \quad W_{3,3} = \frac{3}{6} = 0.5$$

Određivanje trenda – Pokretni prosek sa težinama

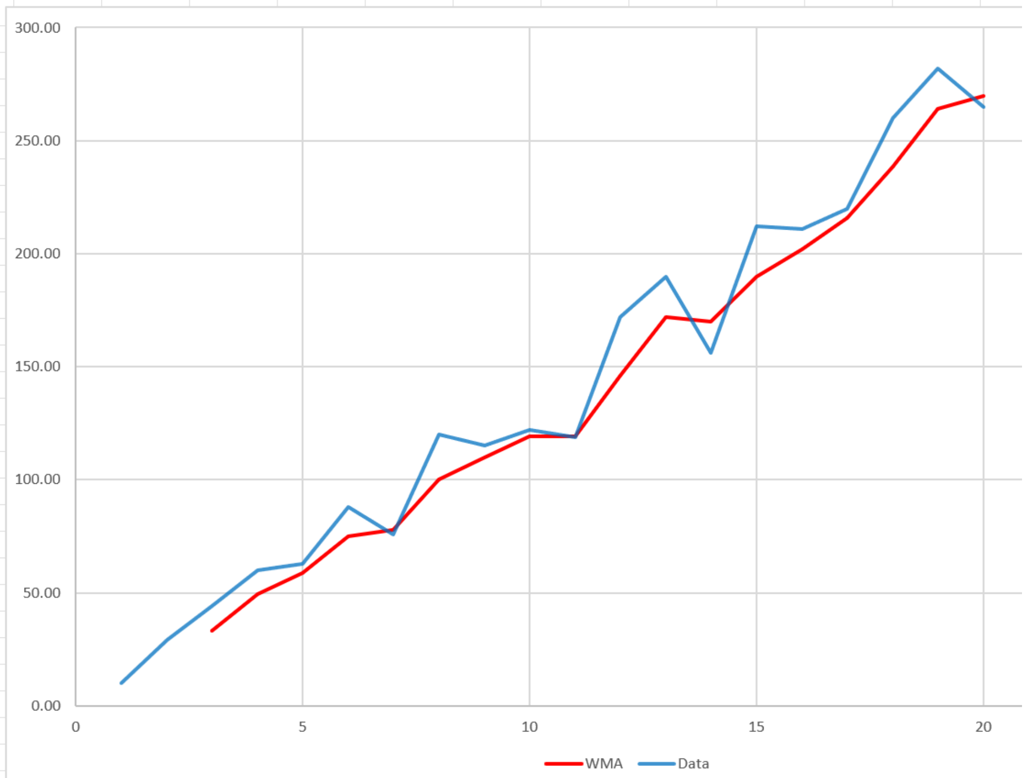
- Na grafiku su date težine za k=15.



Određivanje trenda – Pokretni prosek sa težinama

- Primer WMA_3 za vremensku seriju veličine 20:

Data	Weights	WMA
10	0.166667	
29	0.333333	
44	0.5	33.27
60		49.45
63		58.78
88		74.99
76		77.75
120		100.04
115		110.02
122		119.35
119		119.31
172		146.01
190		171.99
156		169.94
212		189.78
211		201.98
220		215.67
260		238.47
282		264.2
265		269.76



$$WMA_{3,1} = \frac{10 + 2 \cdot 29 + 3 \cdot 44}{3 + 2 + 1} = 33.33 \quad WMA_{3,2} = \frac{29 + 2 \cdot 44 + 3 \cdot 60}{6} = 41.66$$

Određivanje trenda – Eksponencijalni pokretni prosek

- Modifikacija SMA kod koje korisnik kontroliše da li će veća težina biti dodeljena skorijim ili davnijim vrednostima.

- Kontrola se vriši pomoću parametra $\alpha \in (0,1)$:

$$EMA_1 = p_1$$
$$EMA_{trenutni} = \alpha \cdot p_{trenutno} + (1 - \alpha) \cdot EMA_{prethodni}$$

- Uobičajena vrednost za prametar α je:

$$\alpha = \frac{2}{1 + k}$$

- da ponovimo da je *k* veličina prozora

Određivanje trenda – Eksponencijalni pokretni prosek

- Na prethodom slajdu data je semo uobičajena vrednost za prametar α.

- Napomena je da je α parametar koji se može štelovati proizvoljno dok god je zadovoljeno $\alpha \in (0,1)$.

- Formula za α koja ukjljučuje *k* omogućuje korišćenje EMA zadavanjem *k*, a ne α, tj. pruža analogiju sa SMA i WMA.

Određivanje trenda – Eksponencijalni pokretni prosek

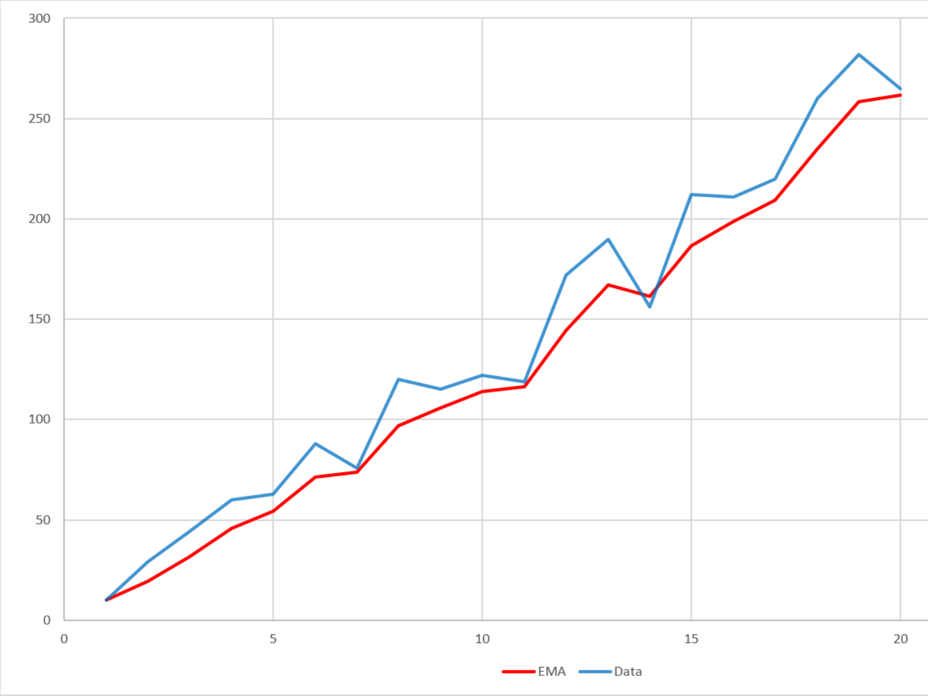
- Na primer, ako imamo vremensku seriju veličine *6:*

$$
\begin{aligned}
EMA_1 &= p_1 \\
EMA_2 &= \alpha \cdot p_2 + (1-\alpha) \cdot EMA_1 \\
EMA_3 &= \alpha \cdot p_3 + (1-\alpha) \cdot EMA_2 \\
EMA_4 &= \alpha \cdot p_4 + (1-\alpha) \cdot EMA_3 \\
EMA_5 &= \alpha \cdot p_5 + (1-\alpha) \cdot EMA_4 \\
EMA_6 &= \alpha \cdot p_6 + (1-\alpha) \cdot EMA_5
\end{aligned}
$$

Određivanje trenda – Eksponencijalni pokretni prosek

- Primer EMA_3 za vremensku seriju veličine 20:

Data	Alpha	EMA
10	0.5	10
29	0.5	19.5
44	0.5	31.75
60	0.5	45.875
63	0.5	54.4375
88	0.5	71.21875
76	0.5	73.60938
120	0.5	96.80469
115	0.5	105.9023
122	0.5	113.9512
119	0.5	116.4756
172	0.5	144.2378
190	0.5	167.1189
156	0.5	161.5594
212	0.5	186.7797
211	0.5	198.8899
220	0.5	209.4449
260	0.5	234.7225
282	0.5	258.3612
265	0.5	261.6806



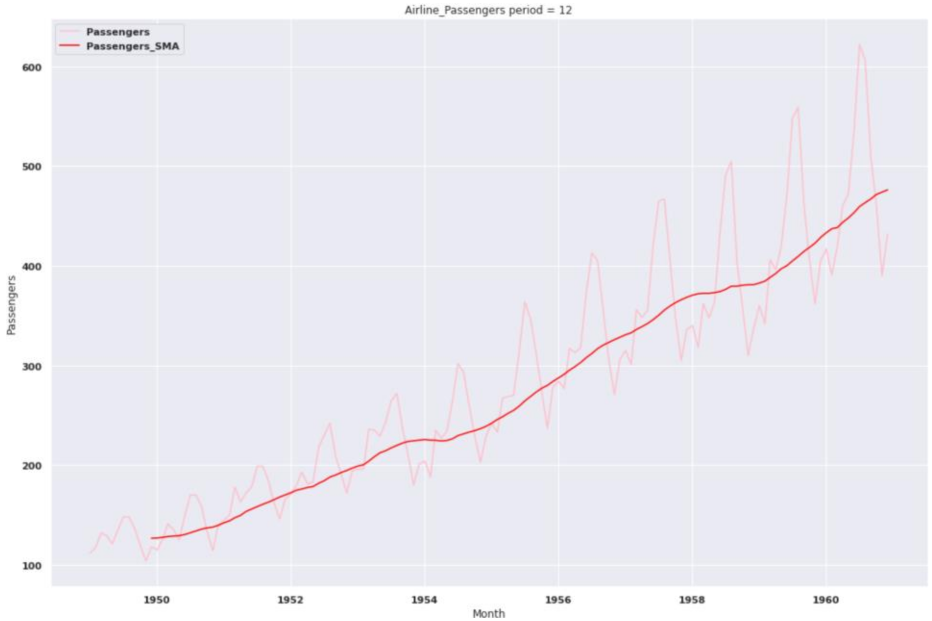
$$\alpha = \frac{2}{1+3} = 0.5$$

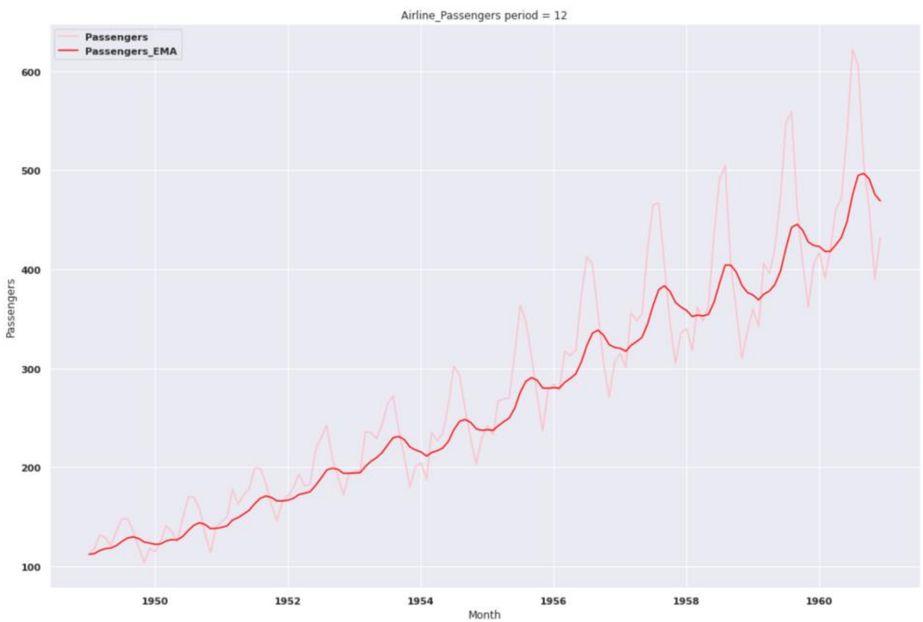
$$EMA_1 = 10$$
$$EMA_2 = 0.5 \cdot 29 + (1 - 0.5) \cdot 10 = 19.5$$
$$EMA_3 = 0.5 \cdot 44 + (1 - 0.5) \cdot 19.5 = 31.75$$

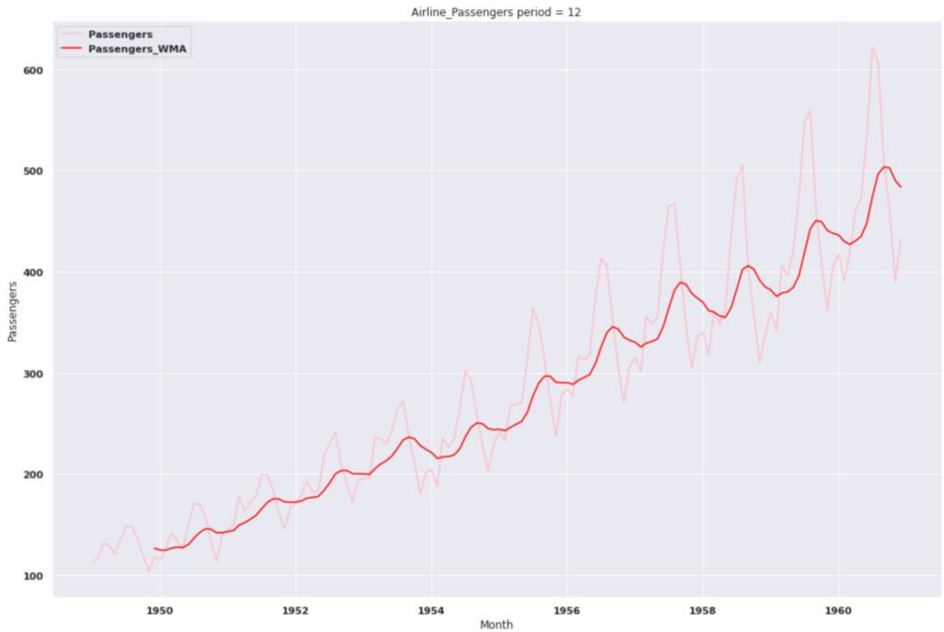
Određivanje trenda – Eksponencijalni pokretni prosek

- Princip modifikacije proseka u EMA koji kombinuje novu vrednost i do tada izračunati prosek je značajan i u drugim oblastima kao što je veštačka inteligencija.
- Iz tog razloga u nastavku je data jedna situcija iz realnog života koja ilustruje ideju na kojoj je princip zasnovan:
- *Otišli ste u svoju omiljenu piceriju i poručili picu koju uvek poručujete.*
- *Međutim, ovog puta ukus pice je bilo loš, odnosno drugačiji nego do sada.*
- *Dogodilo Vam se novo iskustvo, shodno tome svoje dotadašnje iskustvo možete korigovati na više načina:*
  - *Ignorišete novo iskustvo kao da se nije ni dogodilo* - $\alpha = 0$
  - *Potpuno promenite mišljenje i više nikada ne naručujete tu picu* - $\alpha = 1$
  - *Balans između prethodne dve mogućnosti gde uzimate novo iskusto u obzir ali ne odbacujete staro. Na primer, odlučite da ćete još jednom za svaki slučaj probati istu picu pa onda odlučiti da li ćete je još uvek poručivati* - $\alpha \in (0,1)$

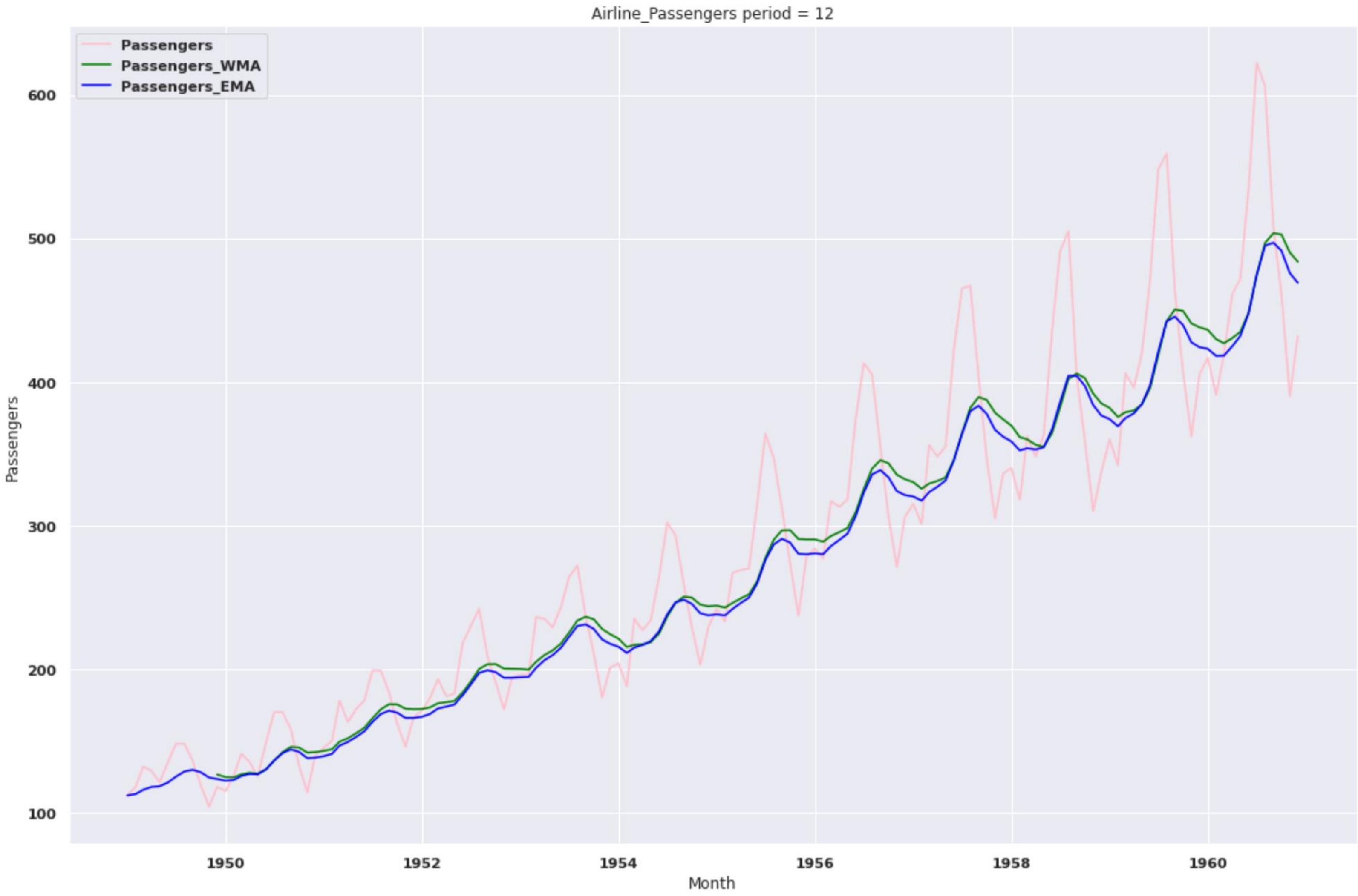
Određivanje trenda – SMA vs. WMA vs. EMA







Određivanje trenda – SMA vs. WMA vs.EMA



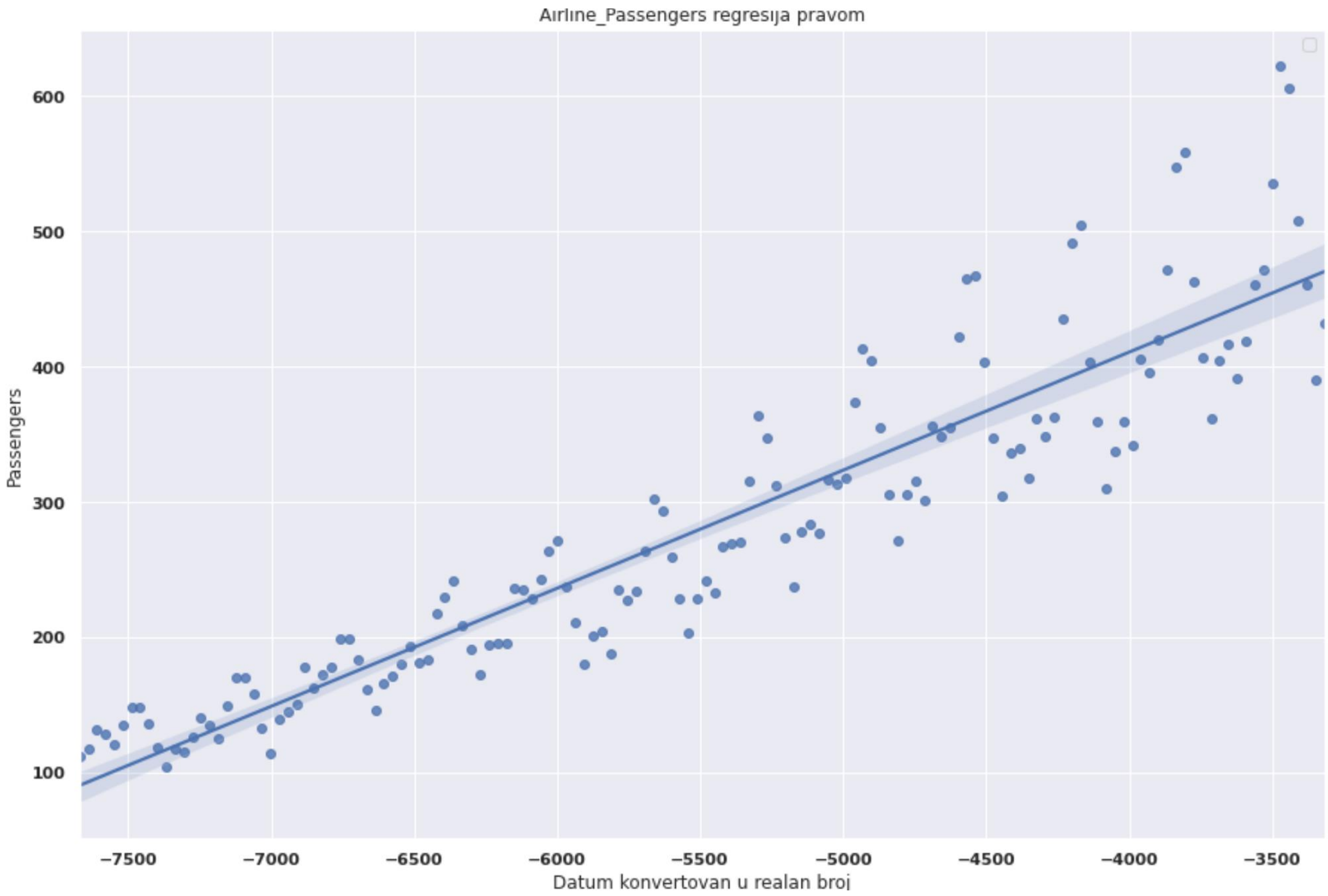
Određivanje trenda – Metode uklapanja krivih – Globalne metode

- Koriste se za dobijanje grubog trenda u vremeskoj seriji.
  - Takav trend nije osetljiv na lokalne varijabilnosti.

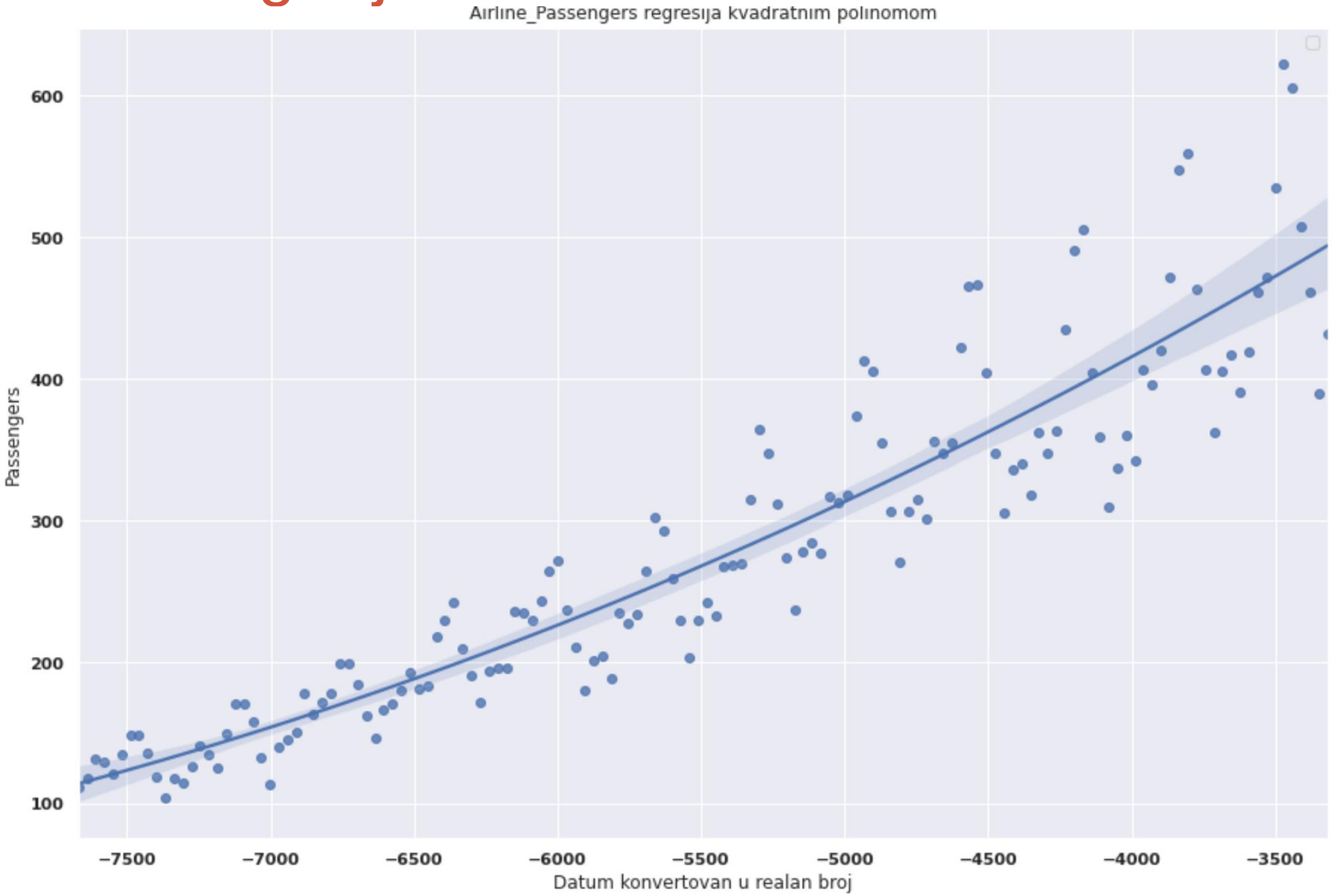
- Ima ih smisla primeniti kada vremenska serija prati jedan trend.
  - Promene trenda je veoma teško detektovatni metodama uklapanja krivih.
    - Bolja opcija u tom slučaju su metode pokretnog proseka.

- Tipično se koristi lineana regresija pomoću prave, kvadratnog ili kubnog polinoma.

Određivanje trenda – Globalne metode – Linearna regesija – Primer



Određivanje trenda – Globalne metode – Linearna regesija – Primer

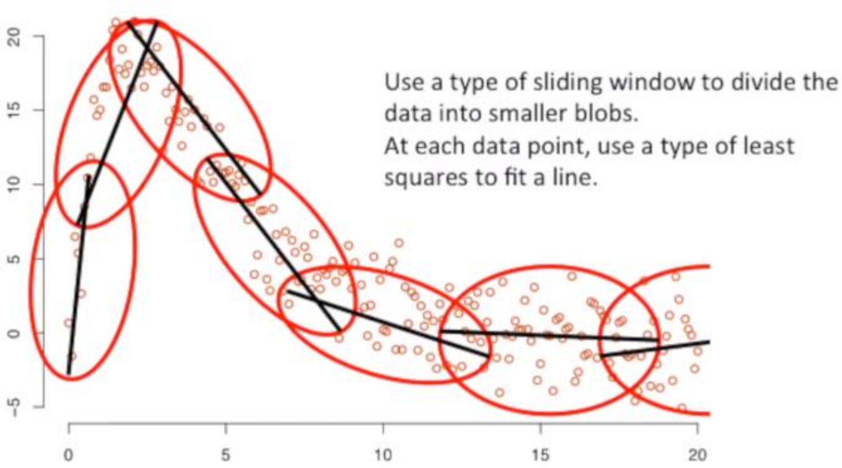


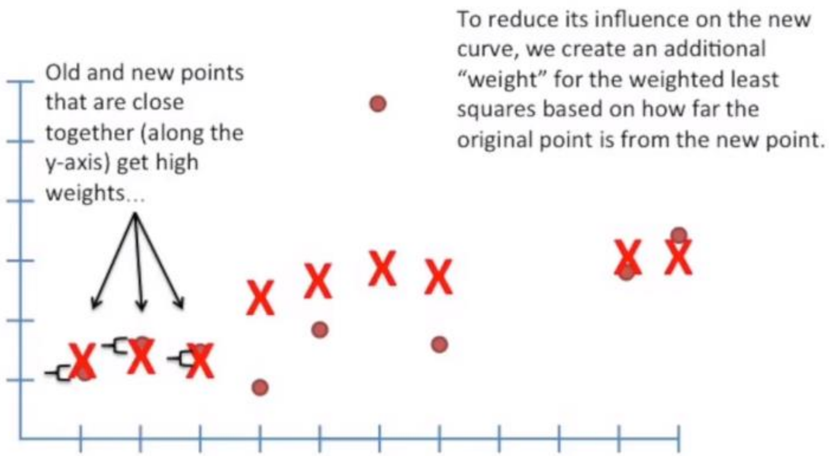
Određivanje trenda – Metode uklapanja krivih – Lokalne metode

- Uklapanje posebnih krivih za svaki prozor (podeok) x-ose.
- Predstavljaju balans između pokretnih proskeka i globalnih metoda.
- Mogu da modeluju promene trenda.
- Spadaju u najbolje metode za detekciju trenda.
- Najćešće korišćene tehnike su *Lowess* i interpolacija kubnim splajnom.

Određivanje trenda – Metode uklapanja krivih – Lokalne metode – LOWESS

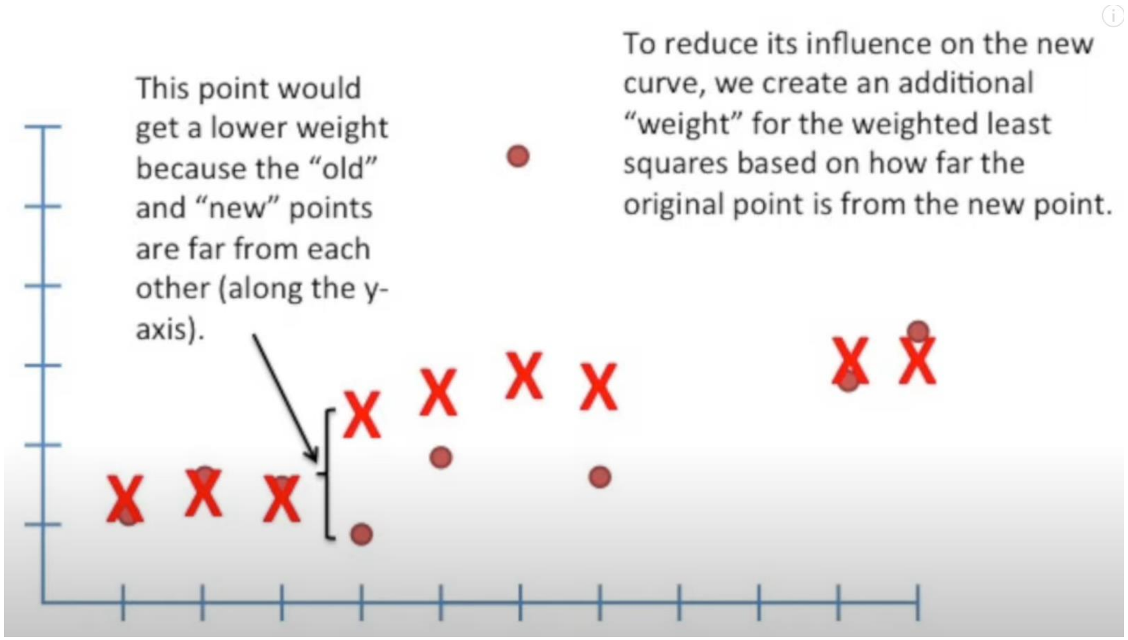
- LOWESS (*Locally Weighted Scatterplot Smoothing*) kriva dobija se interpolacijom reprezentativnih tačaka.





Određivanje trenda – Metode uklapanja krivih – Lokalne metode – LOWESS

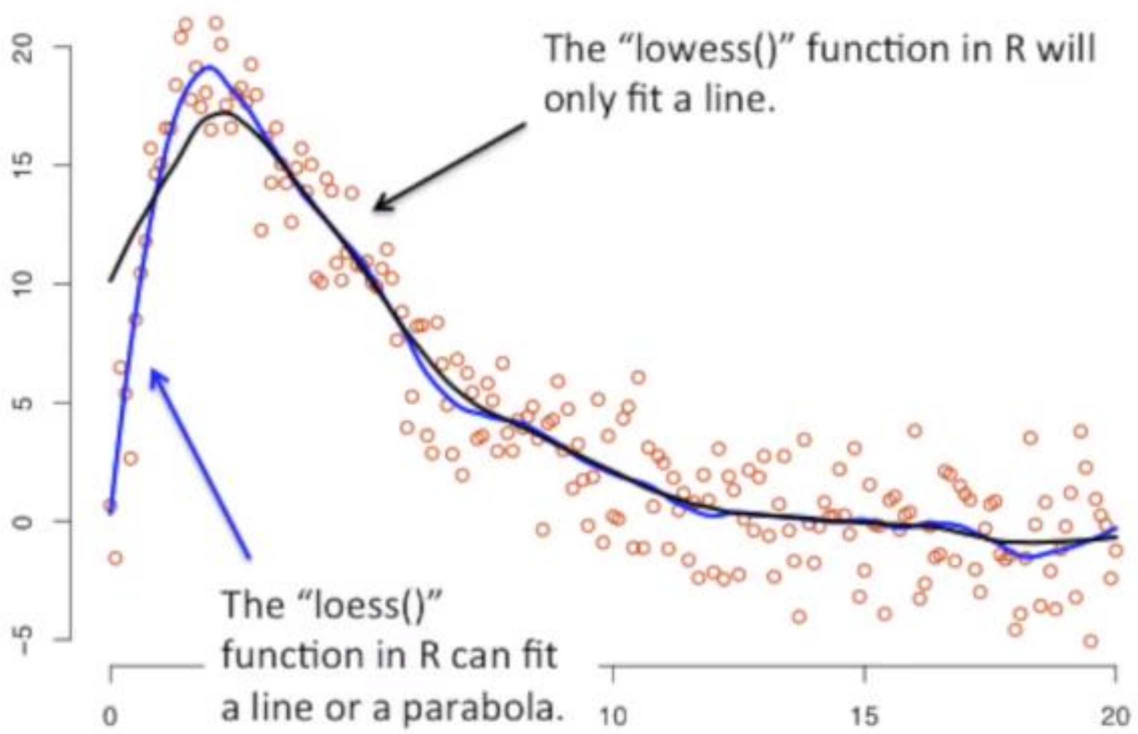
- Reprezentativne tačke nisu vrednosti iz skupa podataka već se dobijaju primenom linearne regresija sa težinama (*Weighted Least Squares, WLS*) redom na vremenske prozore čija se veličina zadaje unapred.
  - WLS nam omogućava da greškama u SSE dodelimo težine i tako utičemo na rezultujuću pravu.
  - Upotreba WLS kod LOWESS smanje uticaj autlajera tako što im se dodeljuju male težine.



Određivanje trenda – Metode uklapanja krivih – Lokalne metode – LOWESS

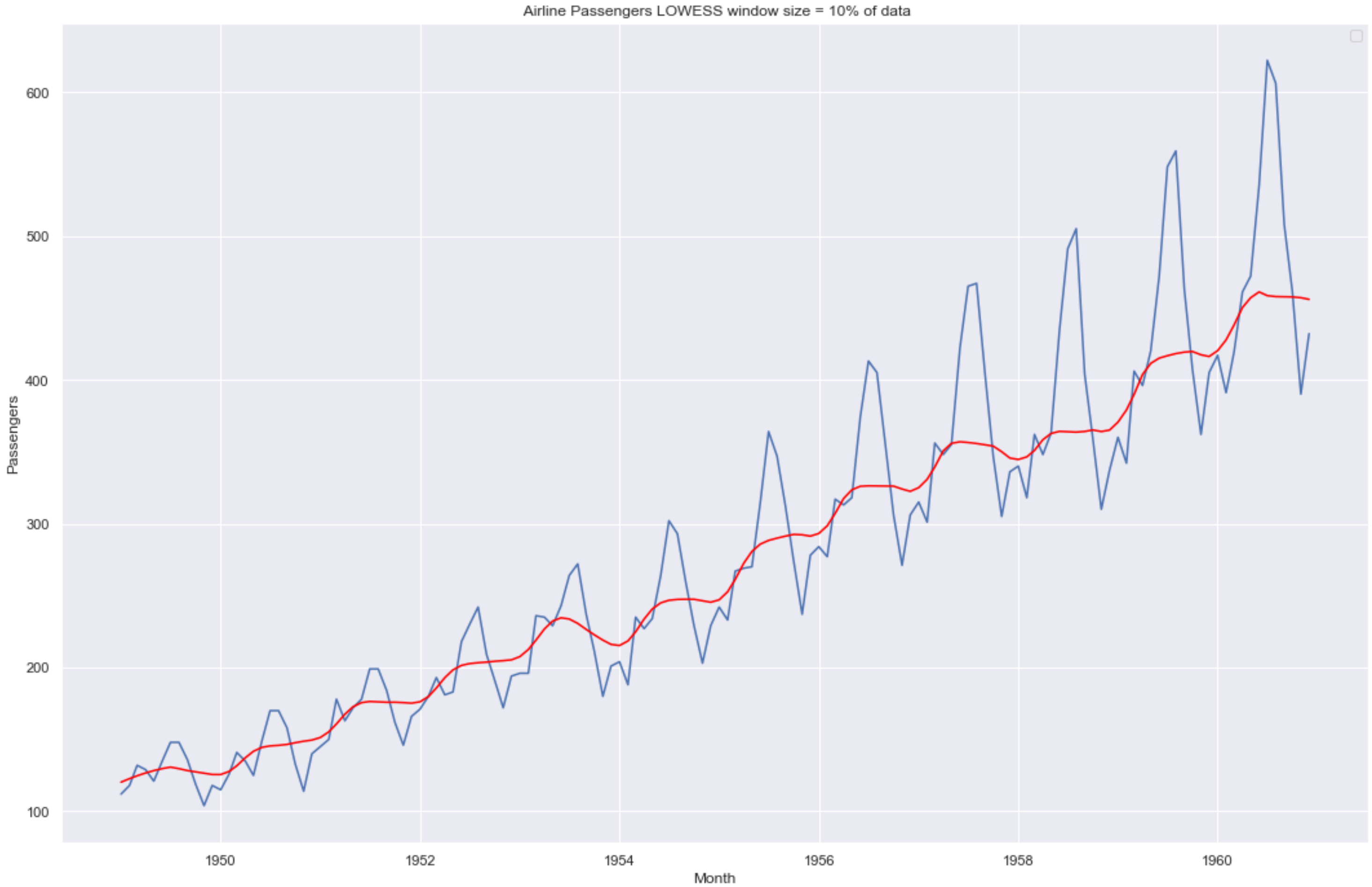
- LOWESS (*Locally Weighted Scatterplot Smoothing*) kriva dobija se interpolacijom reprezentativnih tačaka.

- Za WLS mogu se koristiti prave ili parabole.
  - Izbor zavisi od analize grafika vremenske serije.



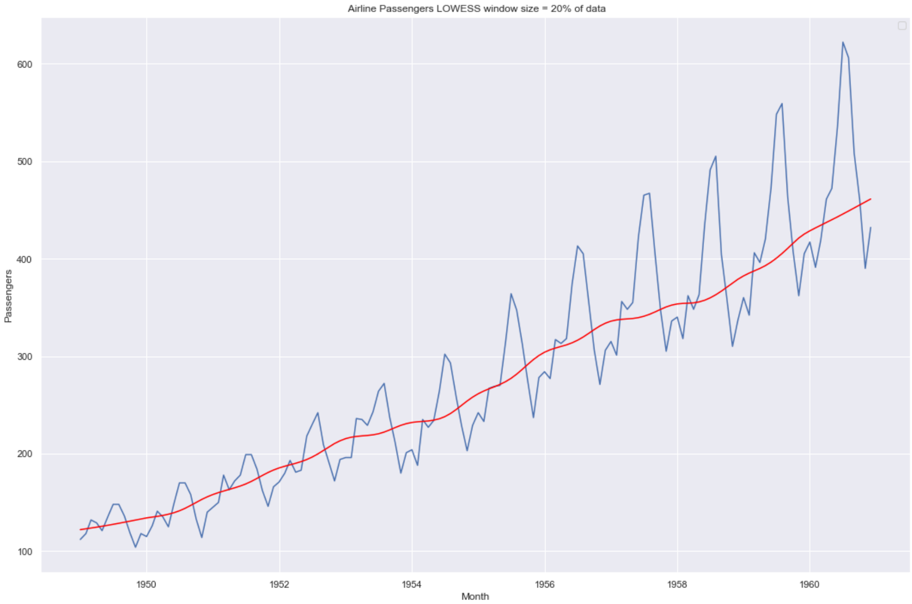
Određivanje trenda – Metode uklapanja krivih – Lokalne metode – LOWESS

- Kod većine implementacija LOWEES veličina prozora zadaje se kao procenat skupa podataka. Podrazumevana vrednost je 10%

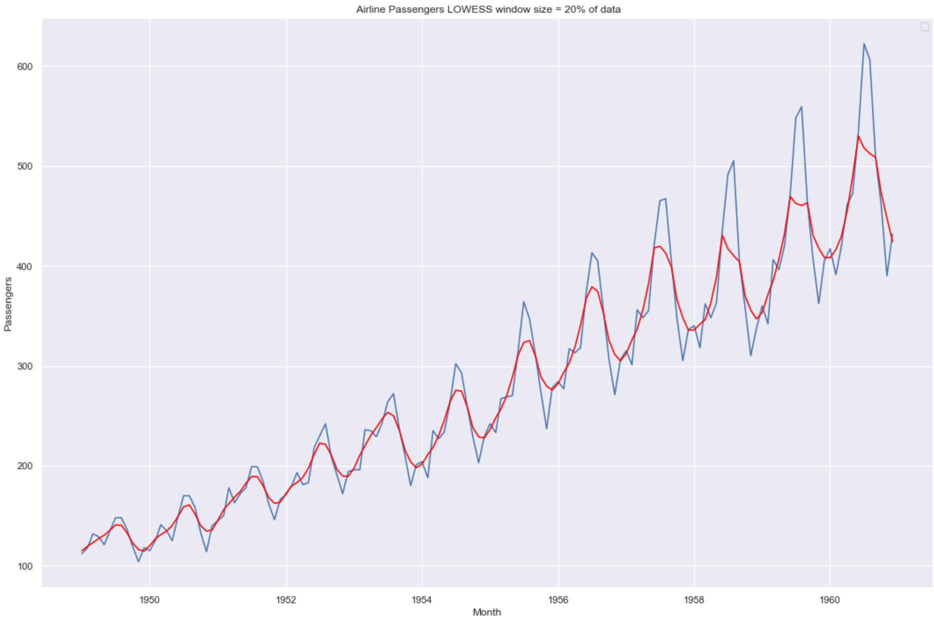


Određivanje trenda – Metode uklapanja krivih – Lokalne metode – LOWESS

veličina prozora 20%

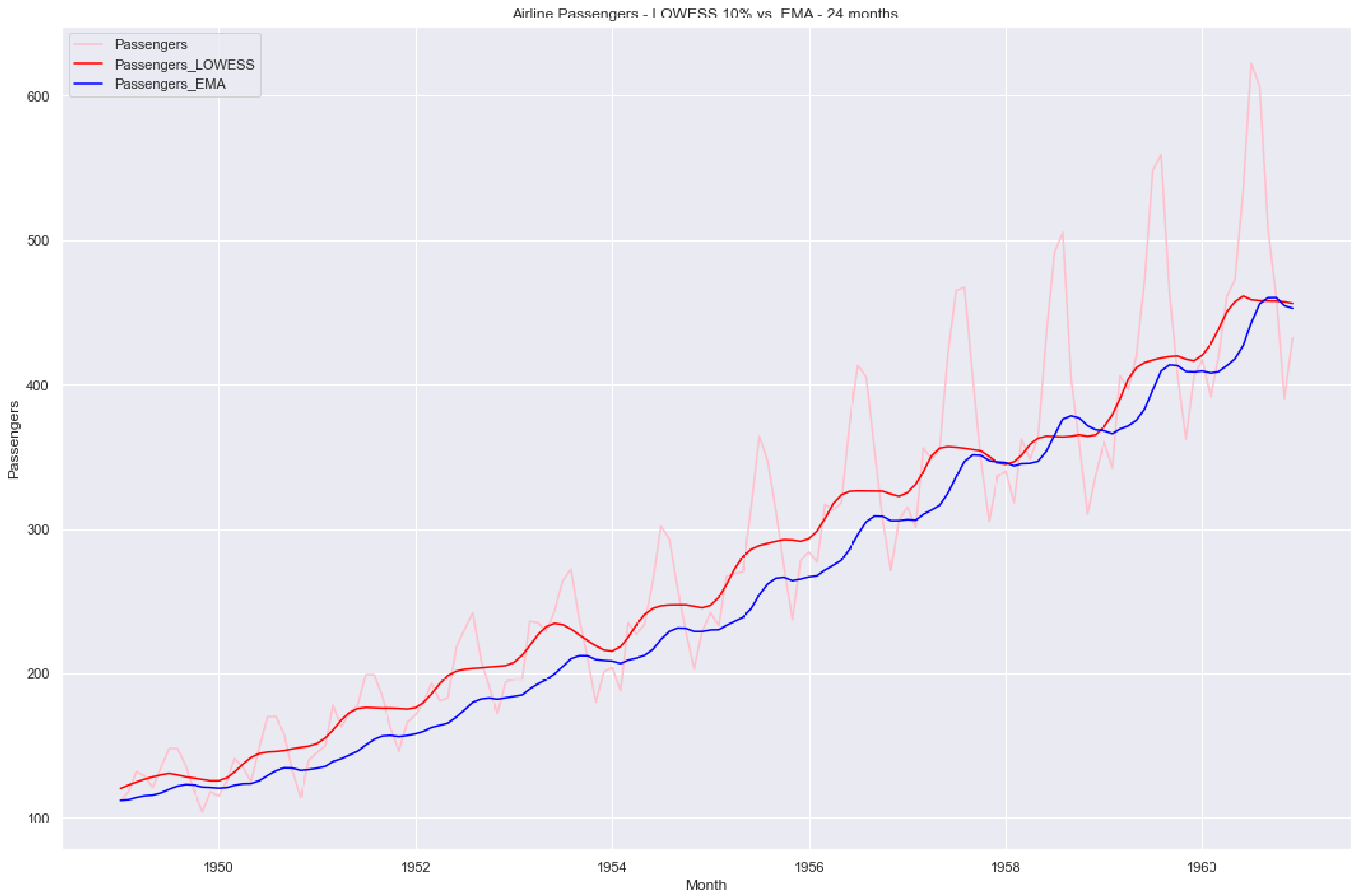


veličina prozora 5%



Određivanje trenda – Metode uklapanja krivih – Lokalne metode – LOWESS vs. EMA

- LOWEES brže reaguje na promene od EMA.



Određivanje trenda – Cikličnost

- Cikličnost se ne pojavljuje u regularnim intrevalima i ima šablon koji nije toliko uniforman kao kod seznonalnosti.
- Iz tog razloga cikličnost je teško detektovati kao zasebnu komponentu.
- Zato se cikličnost detektuje u sklopu detekcije trenda.
- U slučaju da vremenska serija ima cikličnost potrebno je koristiti metode pokretnog proseka ili lokalnog uklapanja krivih da bi se zajedno sa trendom detektovala i cikličnost.
  - Takva komponenta zove se Trend-Ciklus (*Trend-Cycle*).
- Sa obzirom na to, u nastavku prezentacije trend i cikličnost biće označeni zajedno kao T_t.
  - Oznaku C_t koristićemo samo ako je važno da se cikličnost posebno istakne.
  - Takođe ako nije posebno naznačeno pod terminom trend podrazumeva se i cikličnost.

Uklanjanje trenda

- Da bi odredili sezonalost u vremenskoj seriji prvo moramo da uklonimo trend.

- Postupak ukalanja trenda zavisi od odabranog modela:

- Aditivni model – trend se uklanja oduzimanjem od originalne vremenske serije:

$$S_t + R_t = Y_t - T_t$$

- Multiplikativni model – trend se uklanja deljenjem orignialne serije:

$$S_t * R_t = \frac{Y_t}{T_t}$$

Uklanjanje trenda – Odabir modela

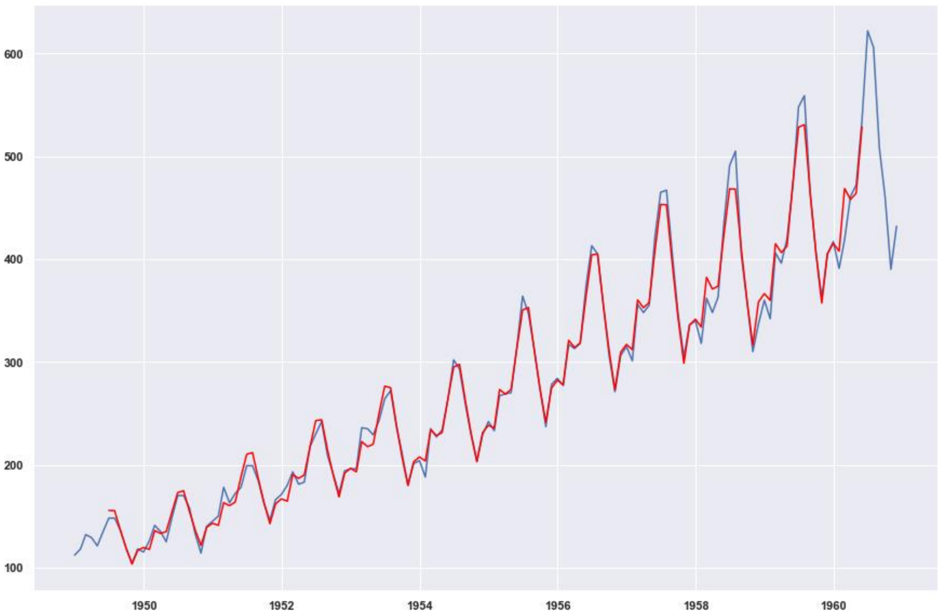
- Vrši se poređenje sledeće dve vremenske serije u odnosu na originalnu seriju Y_t:
  1. $S_t+R_t+T_t$ – za aditivni model
  2. $S_t*R_t*T_t$ – za multiplikativni model

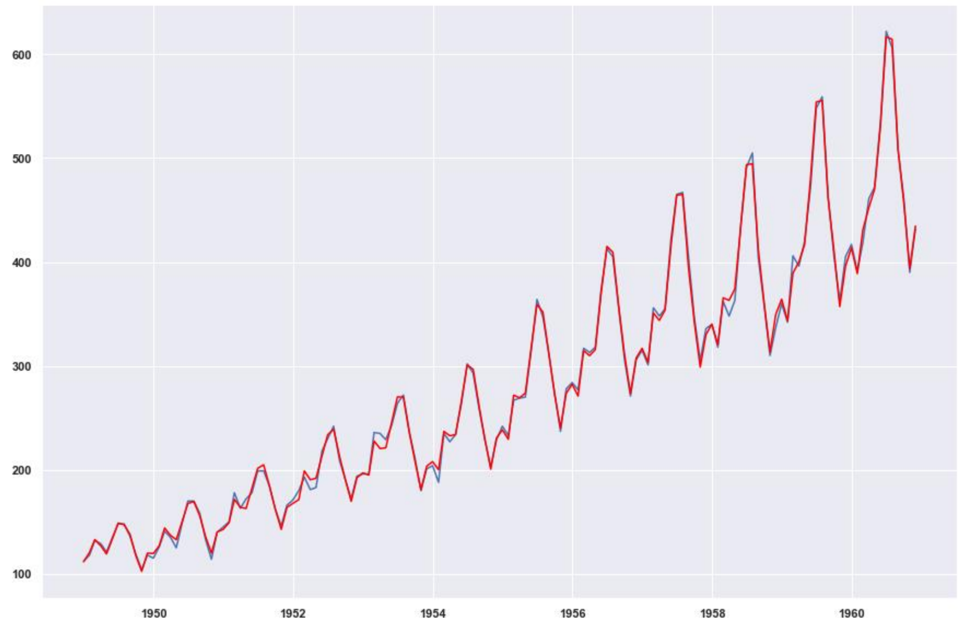
- U zavisnosti od toga koja od serija 1. i 2. se bolje poklapa sa Y_t biramo model koji ćemo koristiti da uklonimo trend.

- Poređenje je najlakše uraditi vizualno (sledeći slajd).

Uklanjanje trenda – Odabir modela

- Kod primera sa avio-komanijom prikladan je multiplikativni model jer seznonalnost povećava magnitudu po vremenu.





Plavo je Y_t

Crveno je $S_t+R_t+T_t$

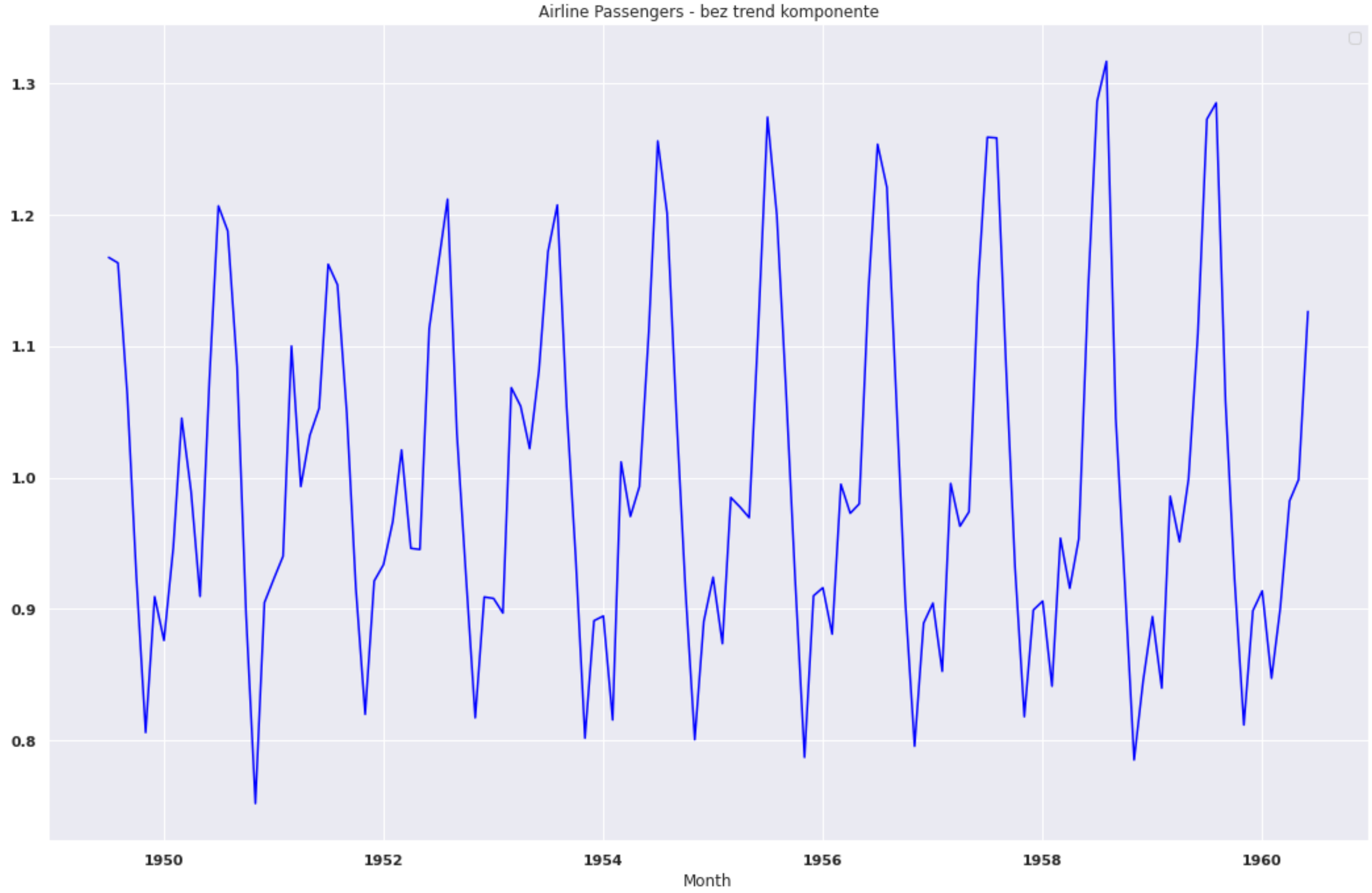
Crveno je $S_t*R_t*T_t$

Određivanje sezonalnosti

- Seznonalnost se određuje iz vremenske serije koja je preostala nakon uklanjanja trenda.
- Prvi korak pri određivanju sezonalnosti je grupisanje vrednosti po jedinicama vremena u kojima je vremenska serija data.
- Na primer, ako su podaci na mesčenom nivou onda grupišemo vrednosti redom za svaki januar u svakoj godini koju imamo, pa onda za februar itd.
- Drugi korak je agregacija grupisanih vrednosti.
- Tipične metode agregacije su prosek ili medijan.

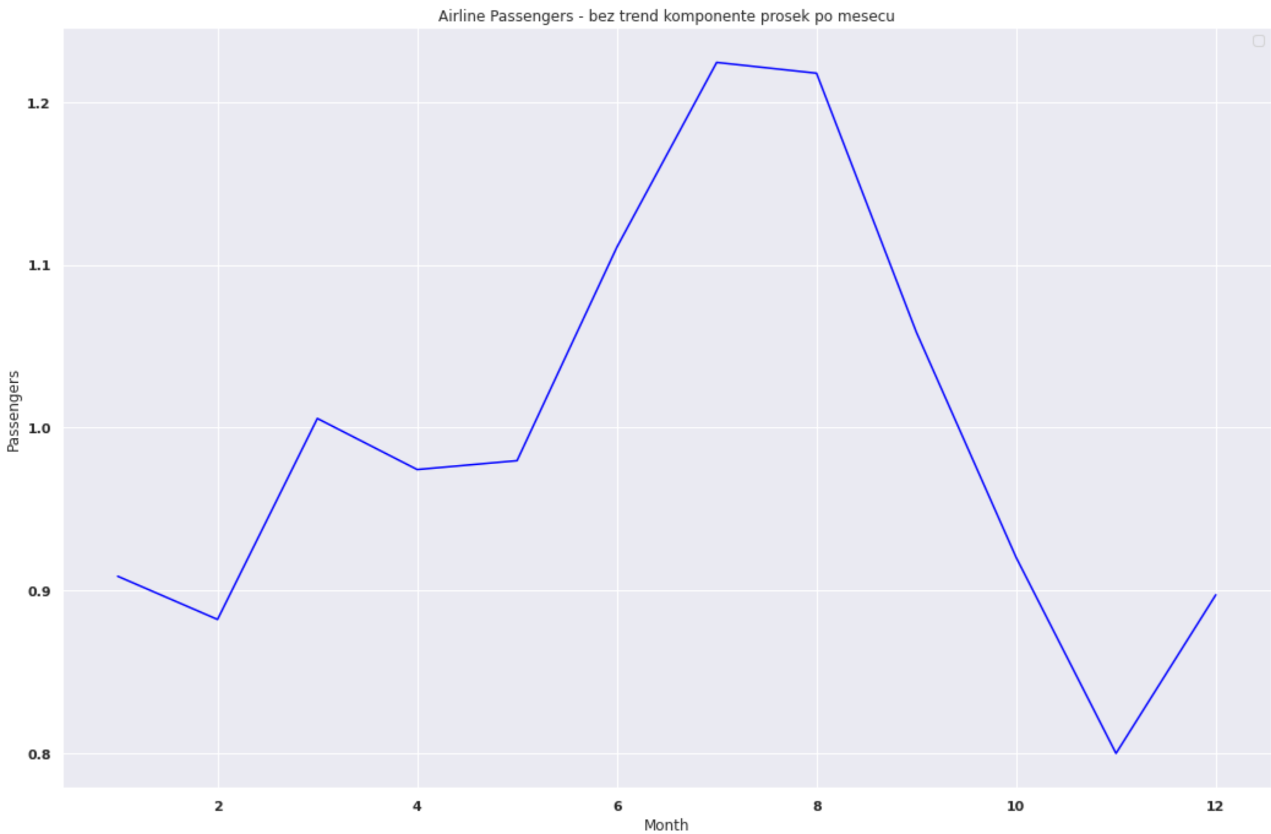
Određivanje sezonalnosti – Primer

- Dat je primer vremenske serije avio-kompanije nakon uklanjanja trenda.
- Grupišemo sada vrednosti za svaki mesec i računamo prosek (sledeći slajd).



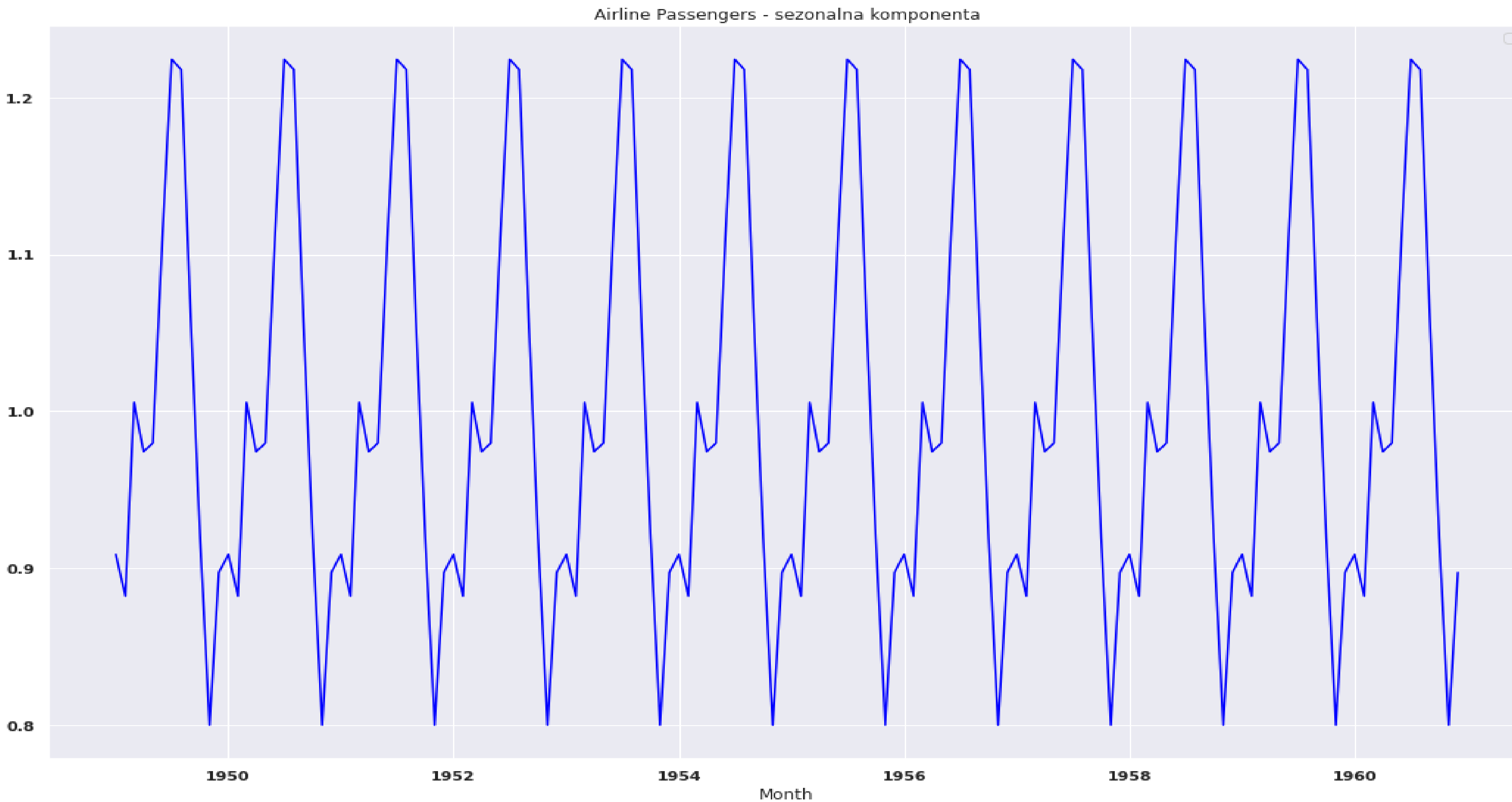
Određivanje sezonalnosti – Primer

- Dobili smo oblik prosečne (tipične) sezonalnosti koja se (u ovom primeru) javlja svake godine.
  - Vidimo da od proleća broj putnika raste i kulminira leti, nakon čega sledi pad i onda skok za Božićne praznike.



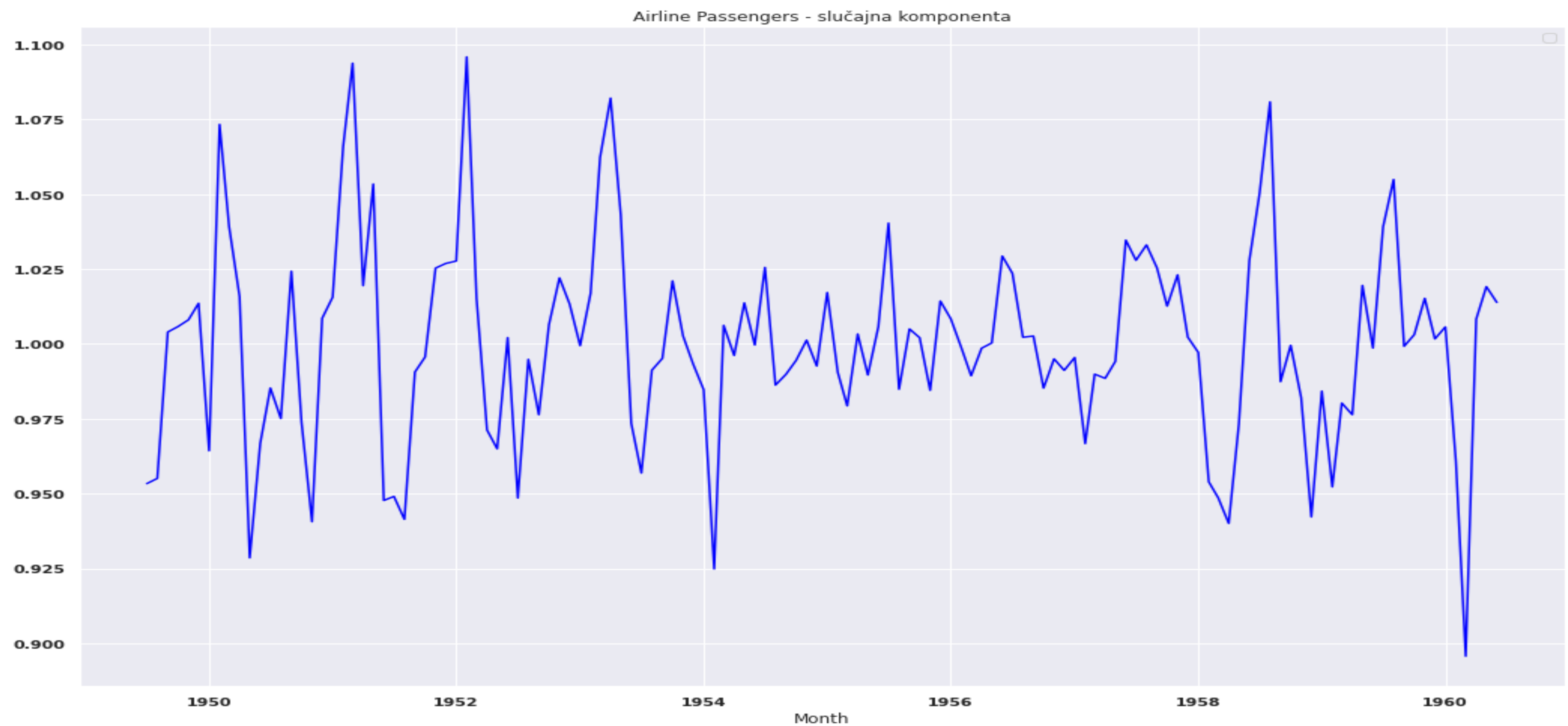
Određivanje sezonalnosti – Primer

- Prosečne sezonalnosti kopiramo za svaki mesec u vremenskoj seriji po godinama redom.
- Tako dobijenu seriju onda uklanjamo redom posle trenda i dobijamo poslednju komponentu odnosno šum.



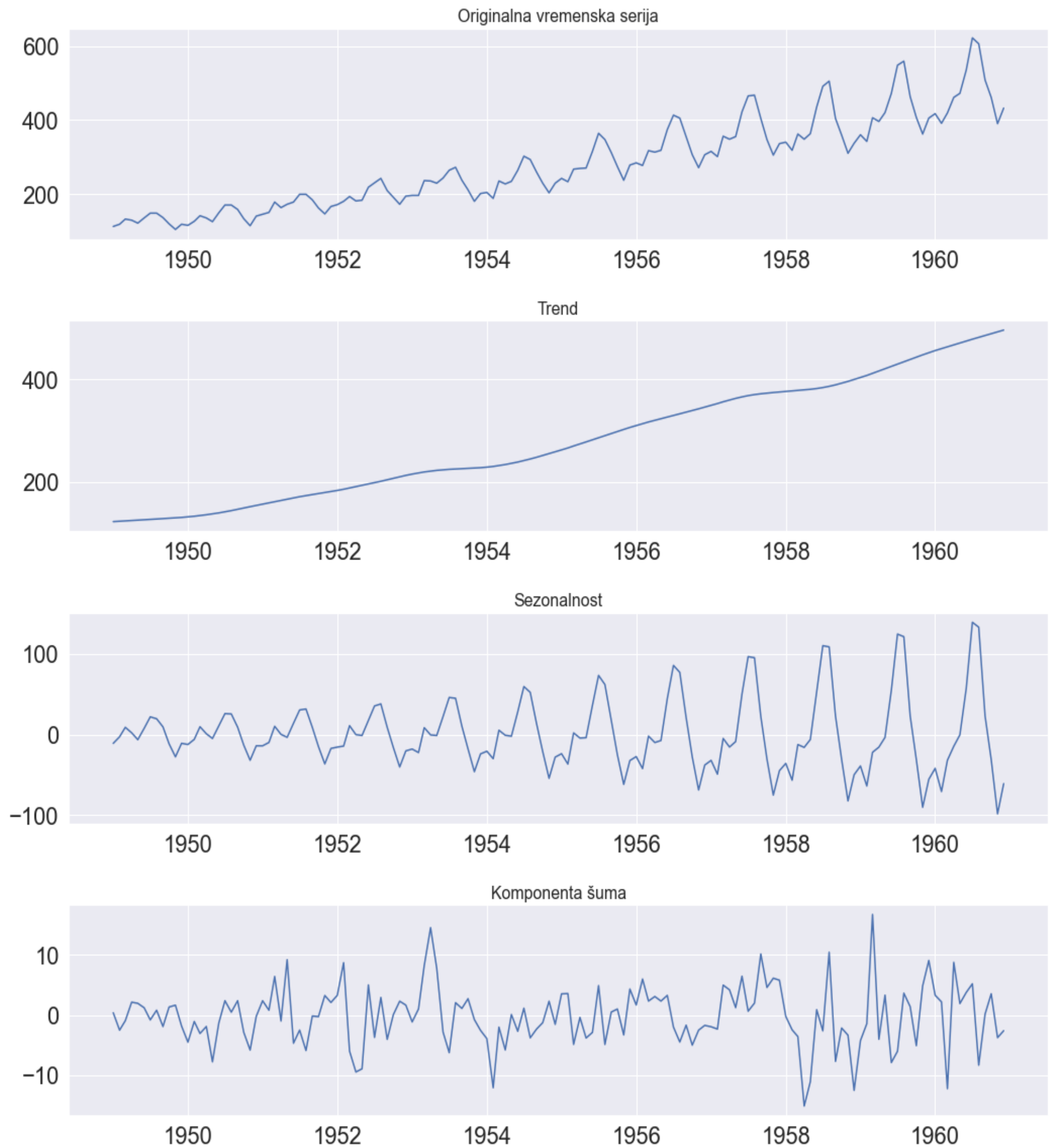
Određivanje komponente Šuma

- Komponenta šuma dobija se nakon uklanjanja trenda i sezonalnosti.
  - Šum predstavlja sve ostale fluktuacije vremenske serije koje se ne mogu objasniti ni pomoću trenda ni pomoću sezonalnosti.
- Na grafiku je primer šuma za podatke o putnicima avio-kompanije.



Dekompozicija vremenske serije – STL metod

- Seasonal-Trend Decomposition Using LOWESS (STL).
- Jedan od najboljih metoda za kompletnu dekompozciju vremenske serije.
- Zasnovan je na LOWEES metodi.
- Postoje gotove implementacije.
  - Na primer u Python statsmodels bibiloteci.



Autokorelacija

- Autokorelacija je mera korelacije između vremenske serije i te iste vremenske serije pomerene za zadati broj *k* vremeskih jedinica unazad.

- Korelacija se računa pomoću koeficijenta korelacije koji smo učili na predavanjima o višestrukoj regresiji:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Autokorelacija – upotreba

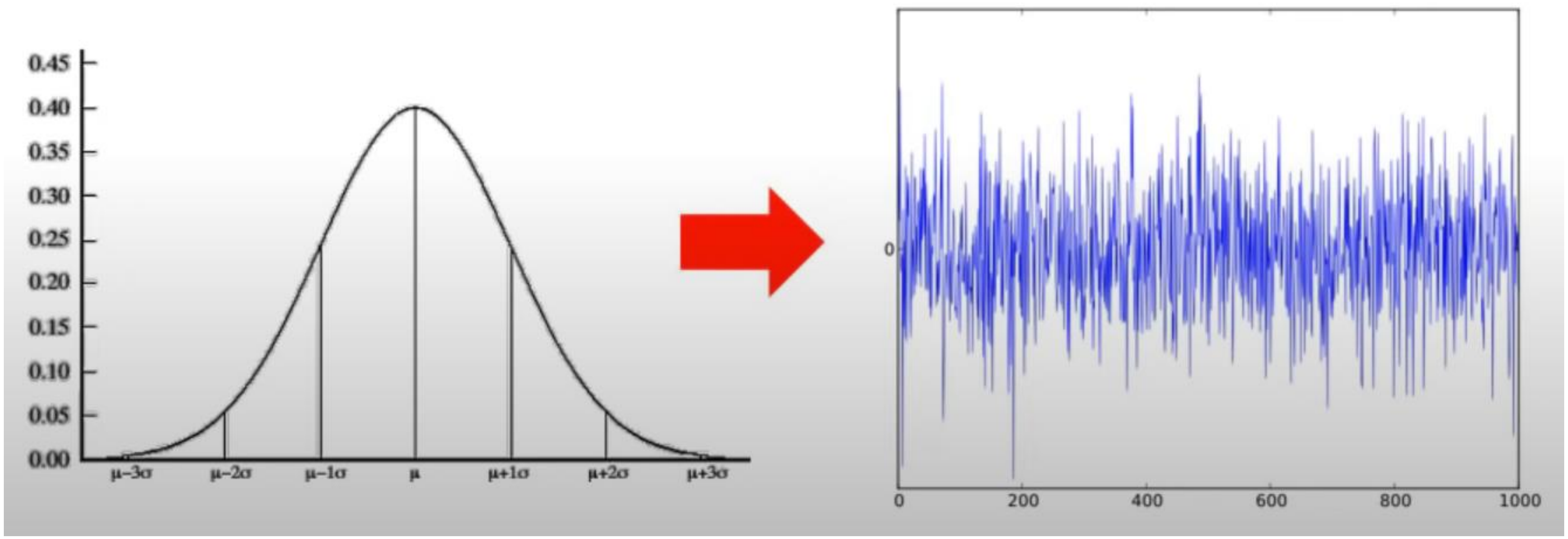
- Autokorelacija se koristi kao prvi dijganostički korak pri modelovanju vremenskih serija.

- Ako vremenska serija nema autokorelacije to znači da se vrednosti kroz vreme menjaju na slučajan način.

- Ako se vrednosti menjaju na slučajan način onda tu vremesku seriju ne možemo da modelujemo jer ne sadrži nijedan šablon koji možemo da otkrijemo.

Autokorelacija – upotreba

- Primer vremenske serije koja nema autokorelacije je pozadinski šum (*white noise*).
- Pozadinski šum dobija se ako izvlačimo (semplujemo) vrednosti iz normalne distribucije i ređamo ih po vremenu.



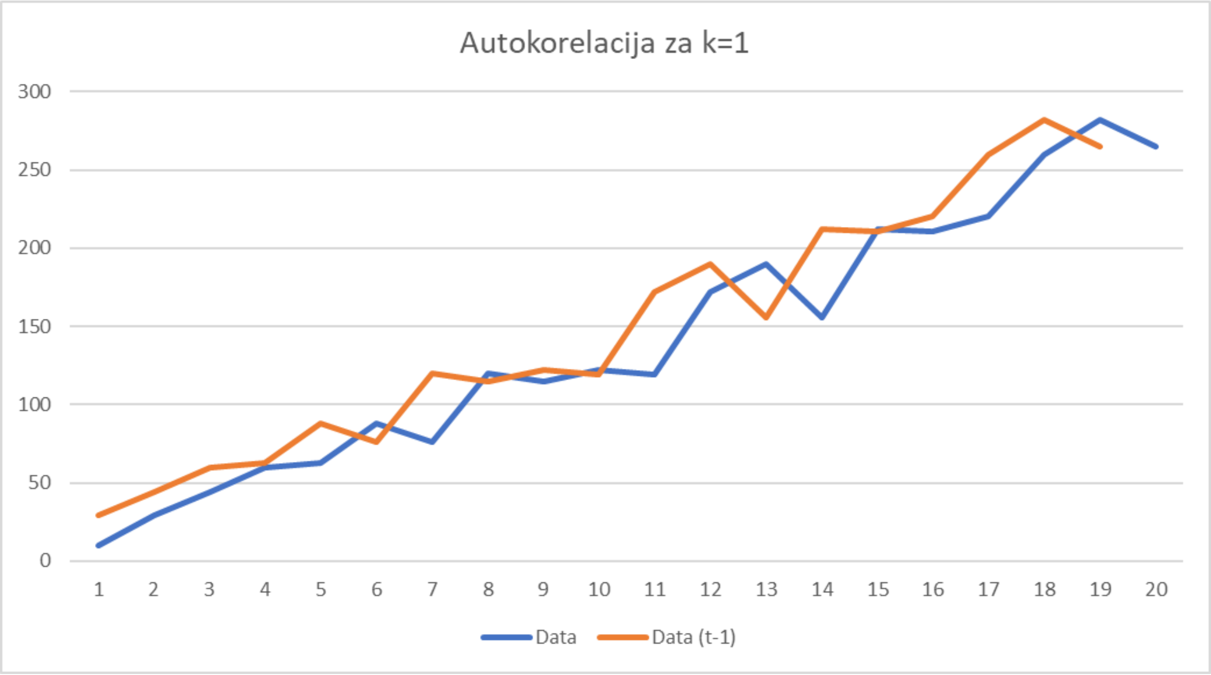
Autokorelacija – Primer

- Za vremensku seriju y_t i jednu vremensku jedinicu unazad imamo sledeću formulu:

srednja vrednost vremske serije pomerene unazad.

$$autocorr\_1_y = \frac{\sum_{t=1}^{n-1}(y_t - \bar{y})(y_{t+1} - \bar{y}_{.1})}{\sqrt{\sum_{t=1}^{n-1}(y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n-1}(y_{t+1} - \bar{y}_{.1})^2}}$$

Data	Data (t-1)	Autocorr_1
10	29	0.9552801
29	44	
44	60	
60	63	
63	88	
88	76	
76	120	
120	115	
115	122	
122	119	
119	172	
172	190	
190	156	
156	212	
212	211	
211	220	
220	260	
260	282	
282	265	
265		

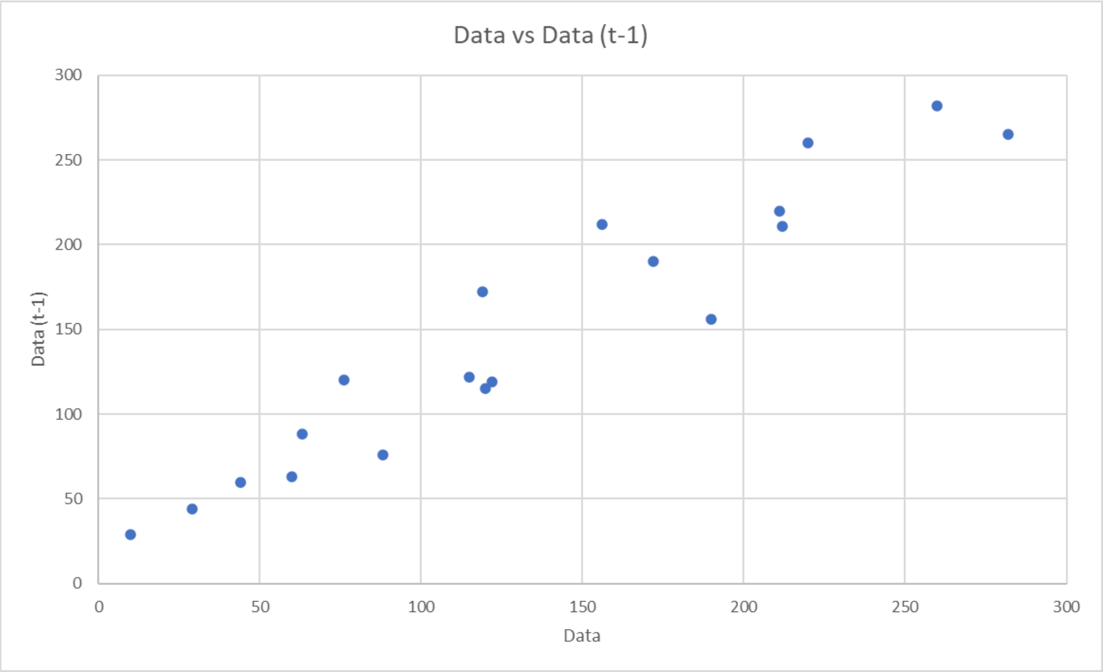


Autokorelacija – Primer

- Sa grafika rasipanja vidi se da postoji linearni odnos sa pozitivnim nagibom između *Data* i *Data(t-1)* pa je zato vrednost autokorelacije 0.95.

$$autocorr\_1_{Data} = 0.95$$

Data	Data (t-1)	Autocorr_1
10	29	0.9552801
29	44	
44	60	
60	63	
63	88	
88	76	
76	120	
120	115	
115	122	
122	119	
119	172	
172	190	
190	156	
156	212	
212	211	
211	220	
220	260	
260	282	
282	265	
265		



Autokorelacija – opšta formula

- Za vremensku seriju y_t i k jedinica unazad imamo sledeću formulu:

$$autocorr\_k_y = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y}_{.k})}{\sqrt{\sum_{t=1}^{n-k}(y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n-k}(y_{t+k} - \bar{y}_{.k})^2}}$$

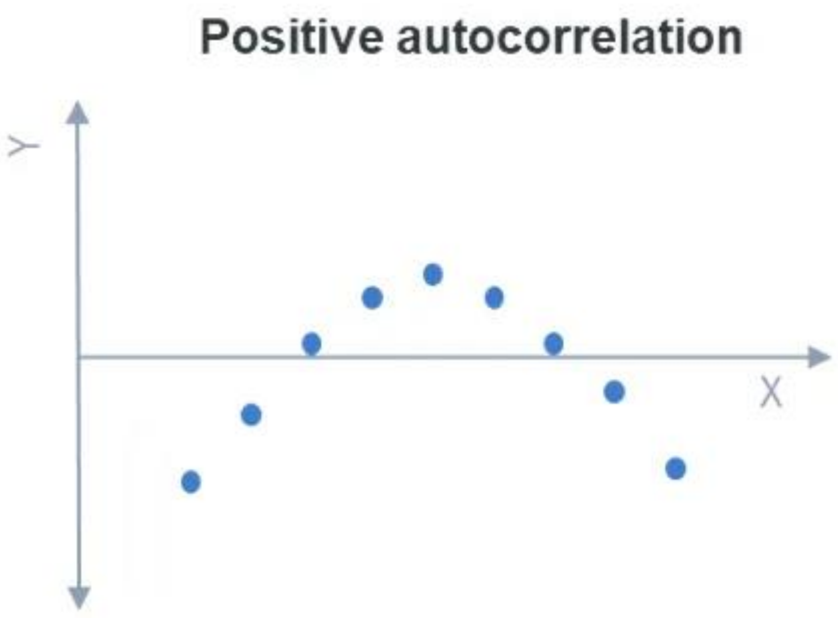
srednja vrednost vremske serije pomerene unazad.

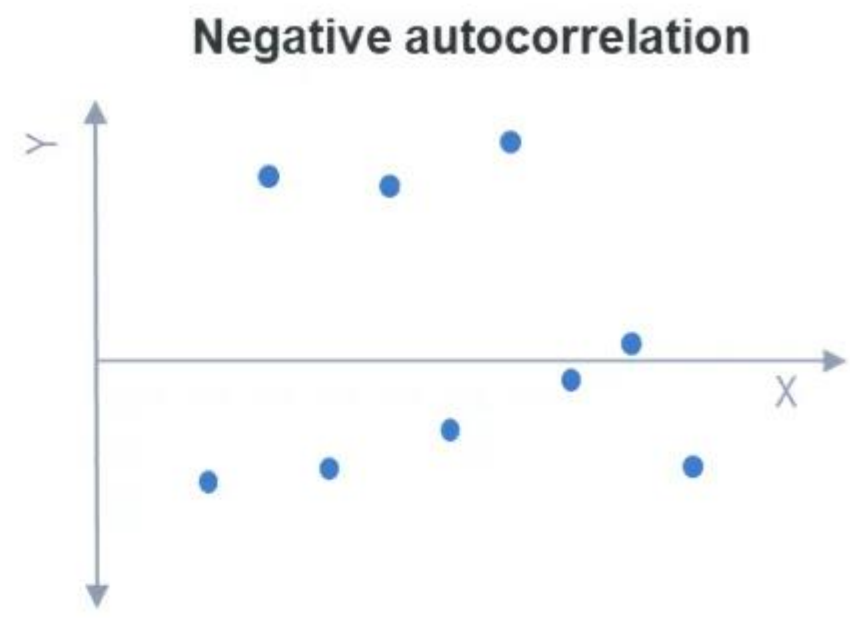
- Za vremenske serije sa velikim brojem elemenata smatra se da je $\bar{y}_{.k} = \bar{y}$ pa se onda koristi uprošćena formula:

$$autocorr\_k_y = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n-k}(y_t - \bar{y})^2}$$

Autokorelacija – pozitivna i negativna

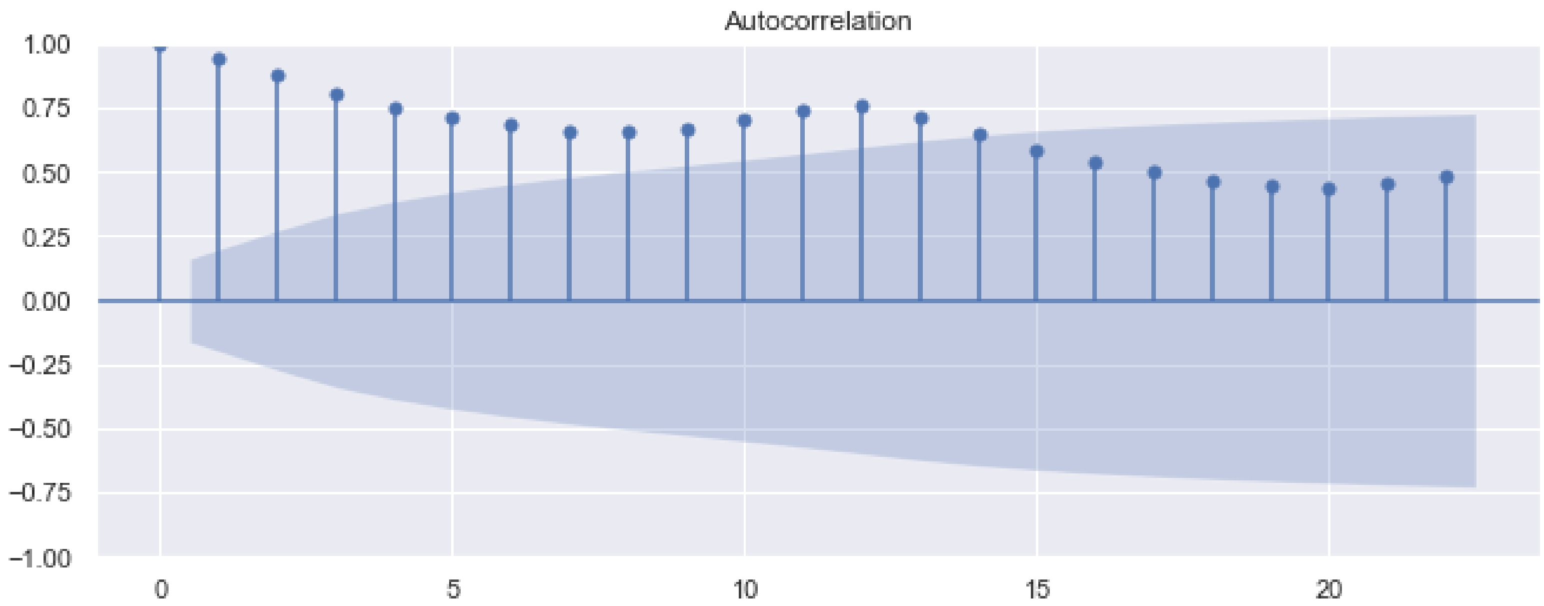
- Pozitivna autokorelacija znači da vrednosti y_{t-k} i y_t opadaju ili rastu zajedno (slika levo).
- Negativna autokorelacija znači da vrednosti y_{t-k} i y_t opadaju i rastu suprotno (grafikon desno).





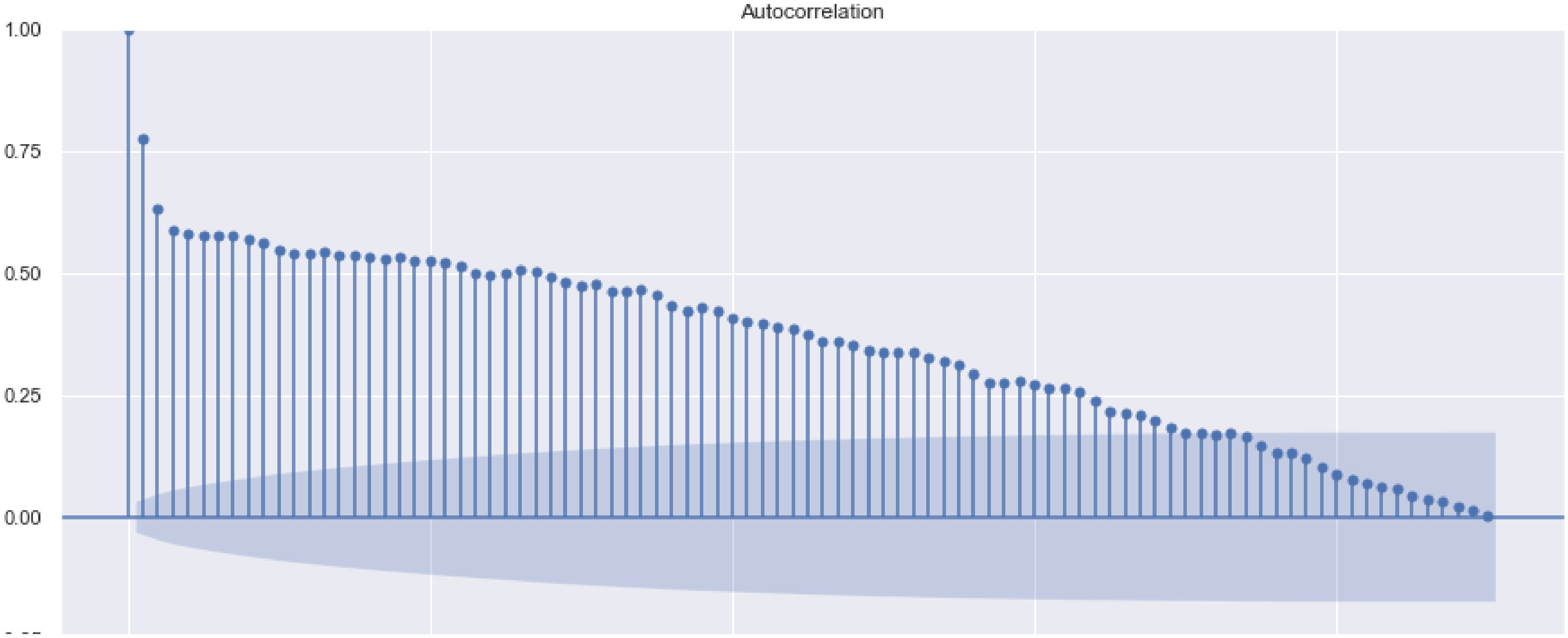
Autokorelacija – Autokorelogram

- Grafik vrednosti autokorelacije za zadati raspon vrednosti *k*.
- Vrednosti u osečenom delu mogu se smatrati beznačajnim.
- Osečen deo su vrednosti autokorelacije koje bi sa 95% verovatnoće dobili za pozadinski šum.



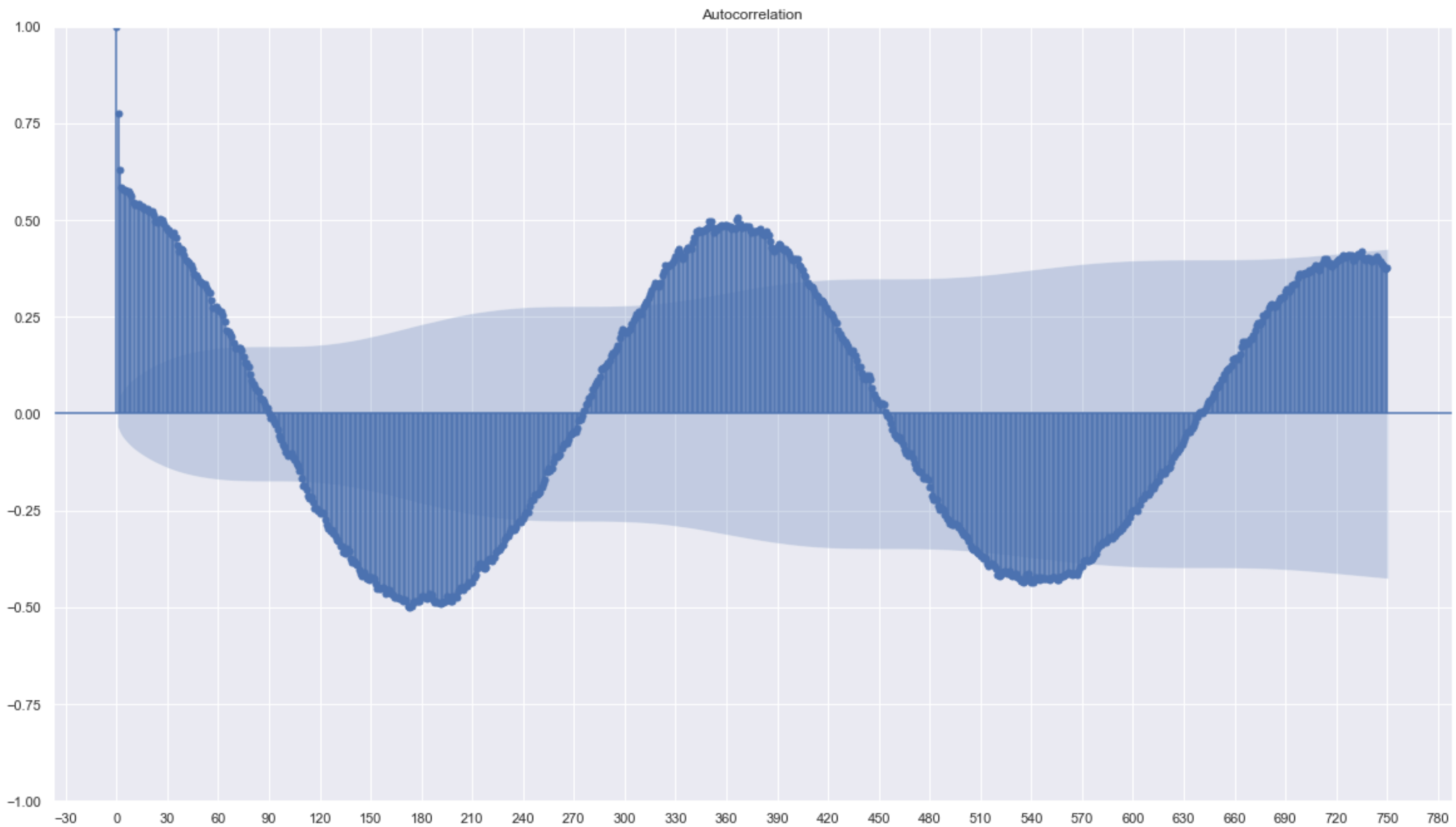
Autokorelacija – Autokorelogram – Tipični šabloni

- Vremenske serije koje nisu slučajni šum obično imaju značajne vrednosti autokorelacije bar za nekoliko vremenskih jedinica unazad.
- Tipičan šablon je postepen pad autokorelacije sa povećanjem broja vremenskih jedinica kao na grafiku ispod.



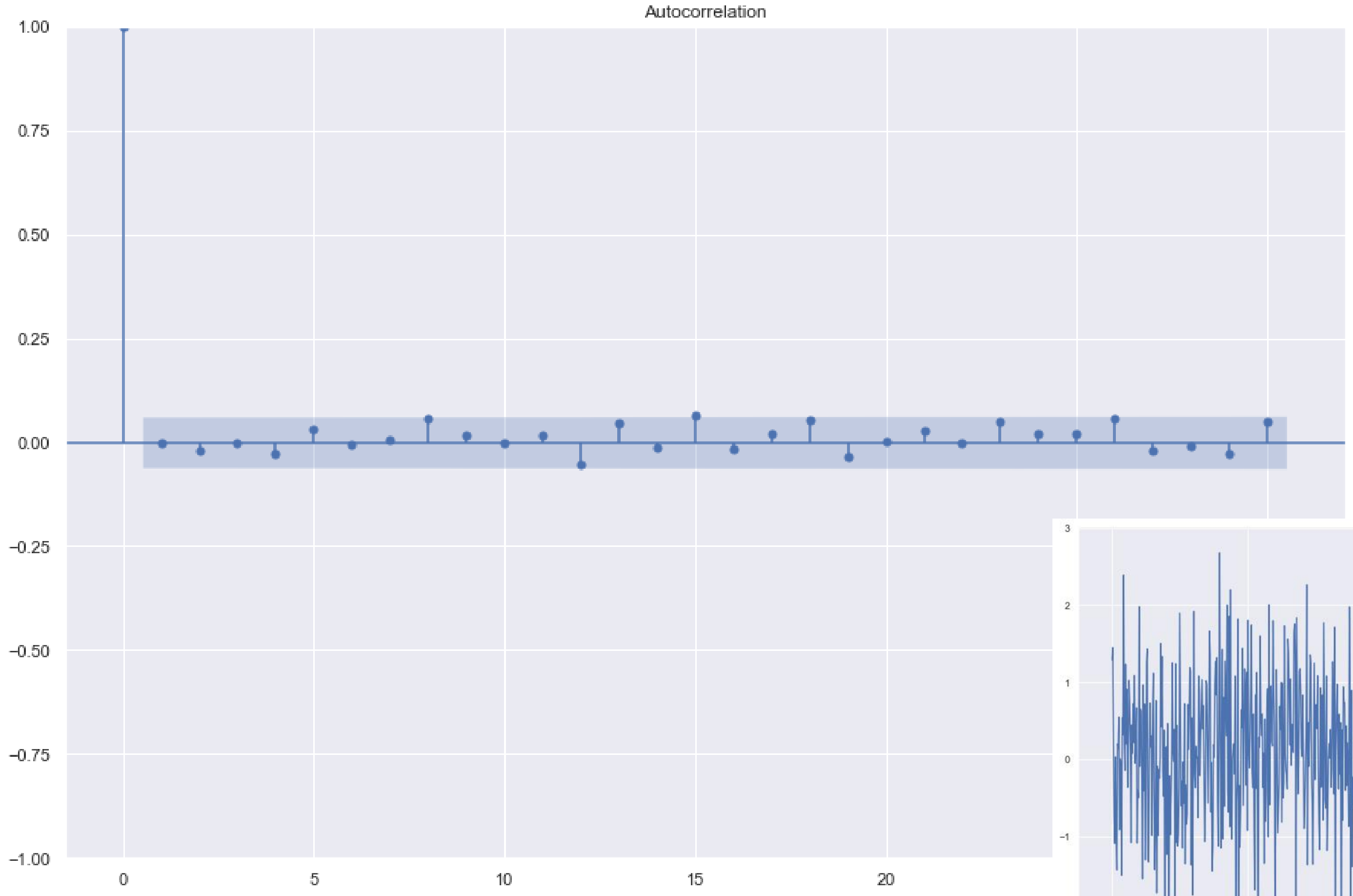
Autokorelacija – Autokorelogram – Tipični šabloni

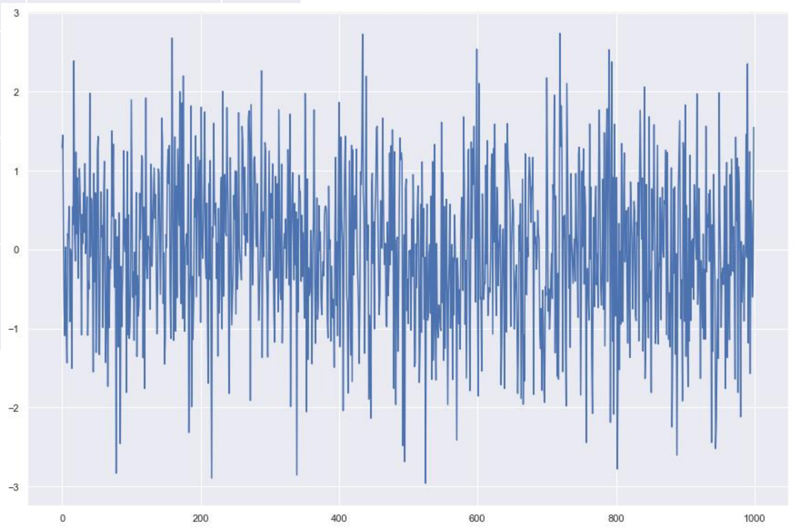
- Ako vremenska serija ima sezonalnost onda se na grafiku obično vidi prvo postepen pad pa onda porast pa onda oped pad itd.
- Dat je autokorelogram dnevne temperature od 1981 - 1991.
- Sa grafika se vidi sezonalnost, gde se na svakih 90 dana menjaju godišnja doba.



Autokorelacija – Autokorelogram – Tipični šabloni

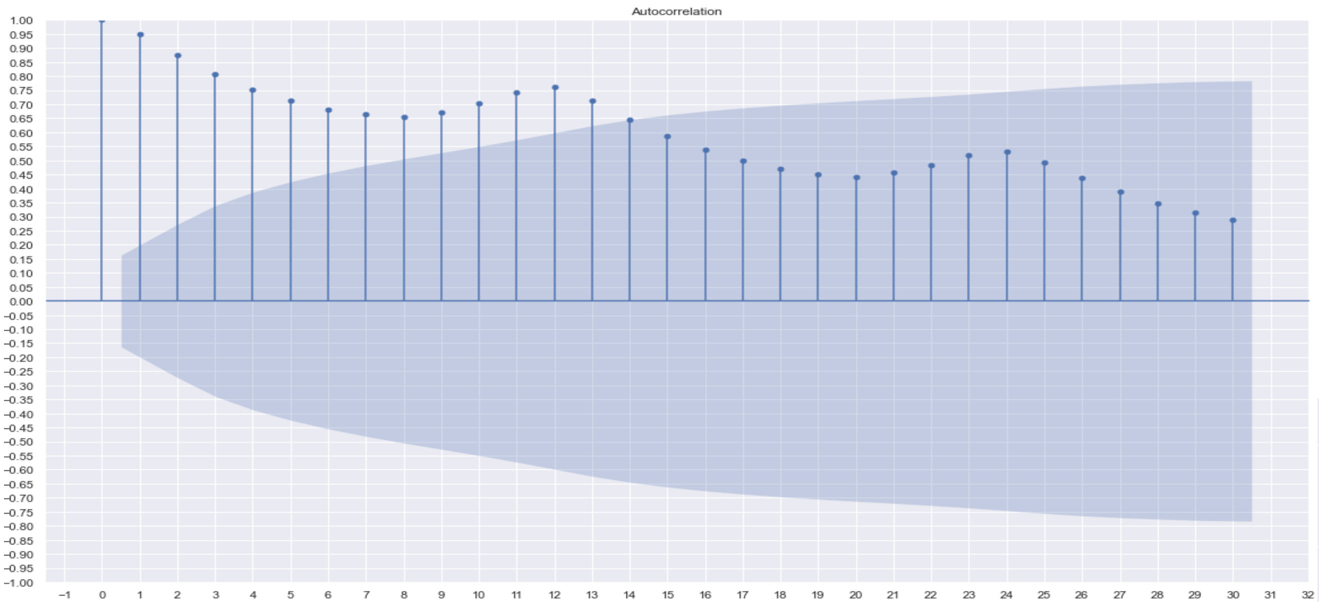
- Autokorelogram za pozadinski šum takođe ima tipičan šablon kod koga su (uglavnom) sve vrednosti unutar osenčenog dela.

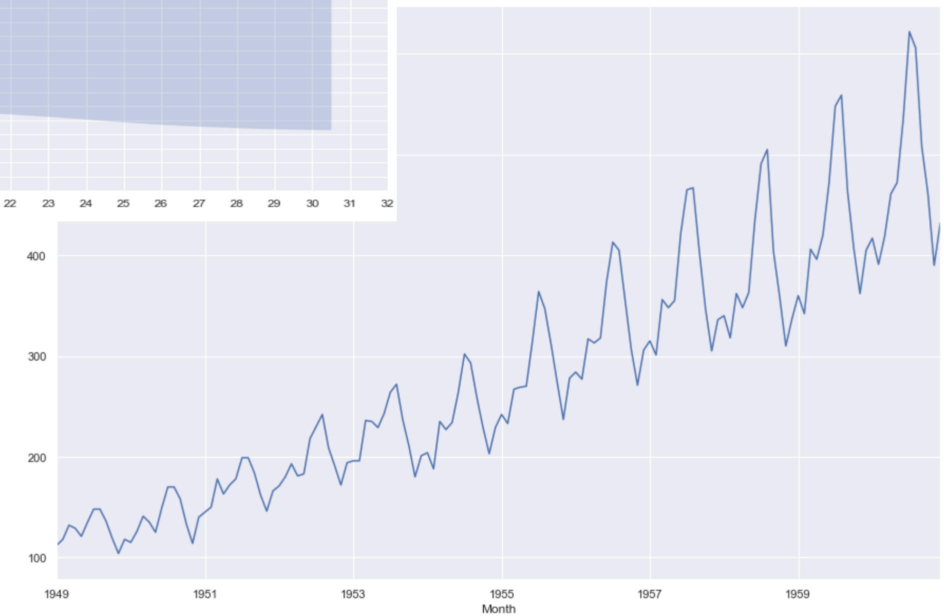




Autokorelacija – Autokorelogram – Primer

- Autokorelogram za podatke o avio-kompaniji.





- Sa grafika se vidi da su autokorelacija posle k=14 beznačajne.

- Takođe se vidi da su sve vrednosti pozitivne.

Parcijalna autokorelacija

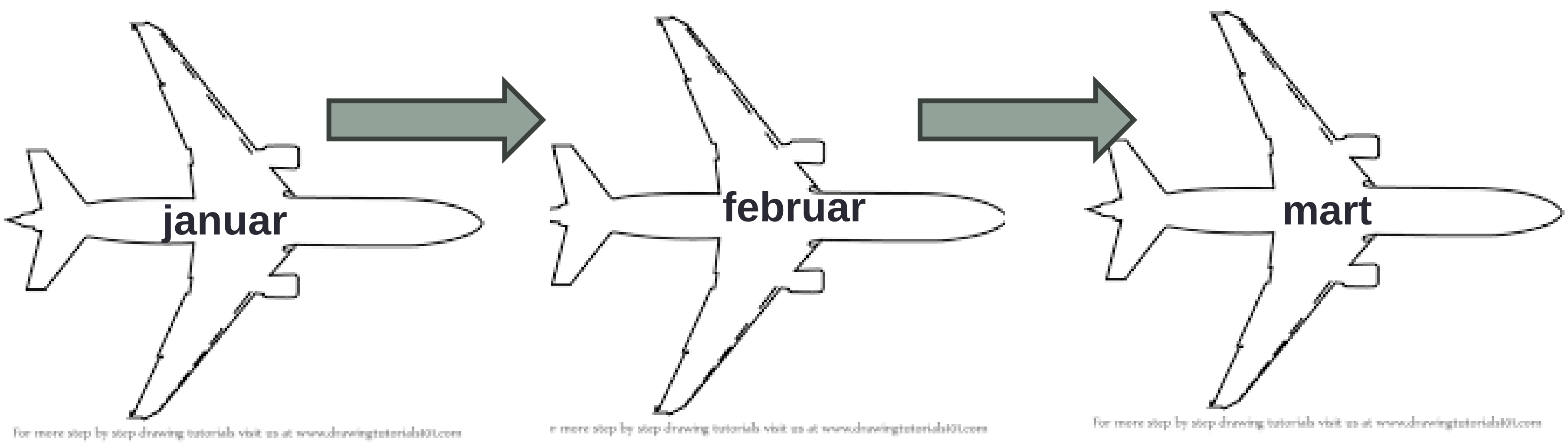
- Problem sa autokorelacijom je što vrednosti koeficijenta za svako k>1 nisu realne.

- Razlog za to je indikretna (posredna) autokorelacija.

- Indikretnu autokorelaciju objasnićemo na primeru avio-kompanije (broj putnika po mesecu).

Parcijalna autokorelacija

- Recimo da imamo veliku autokorelciju za k=1.
- To znači da broj putnika u januaru utiče na broj putnika u februaru, a broj putnika u februaru utiče na broj putnika u martu.
- Iz toga sledi da broj putnika u januaru utiče na broj putnika u martu, ali to nije direketan uticaj već posredan preko februara.



Parcijalna autokorelacija

- Koeficijent autokorelacije ne može da razdvoji direktan uticaj od posrednog i zato njegove vrednosti nisu realne.

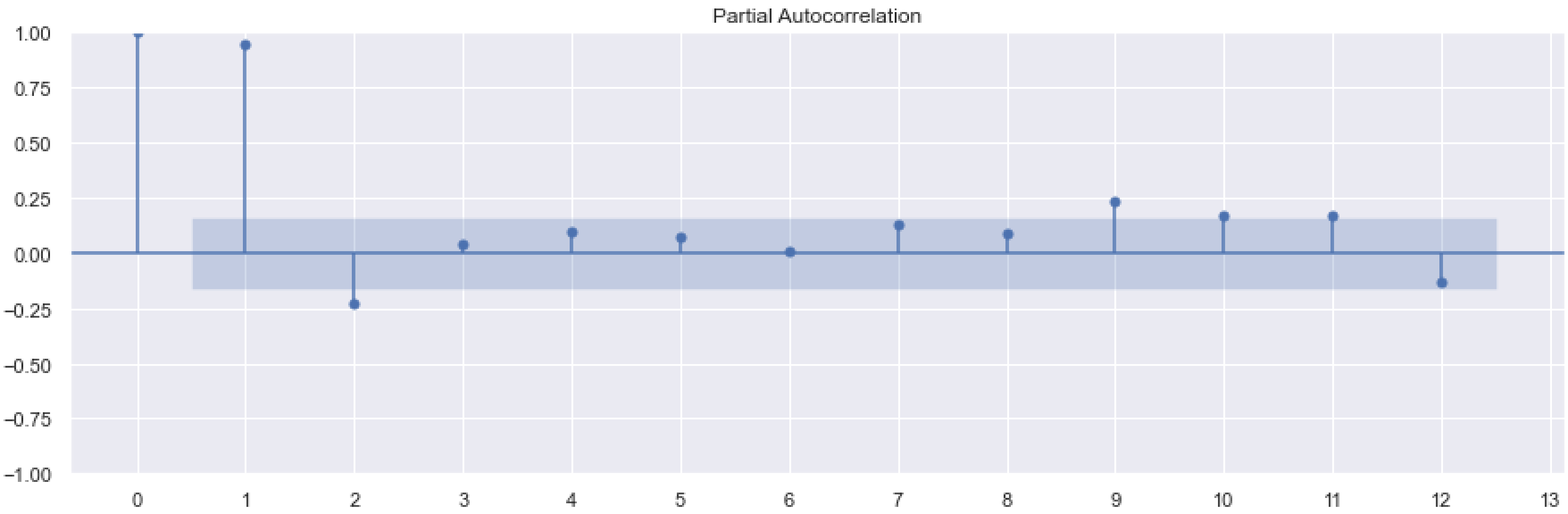
- Direktan uticaj dobićemo pomoću koeficijenta parcijalne autokorelacije.

Parcijalna autokorelacija

- Parcijalna autokorelacija meri se pomoću koeficijenta parcijalne autokorelacije.
- Koeficijent parcijalne autokorelacije izračunava se u dva koraka pomoću linearne regresije.
- Vrednosti koeficijenta su u rasponu [-1, 1].
- Vrednosti koeficijenta prikazuju se pomoću parcijalnog autokorelograma.

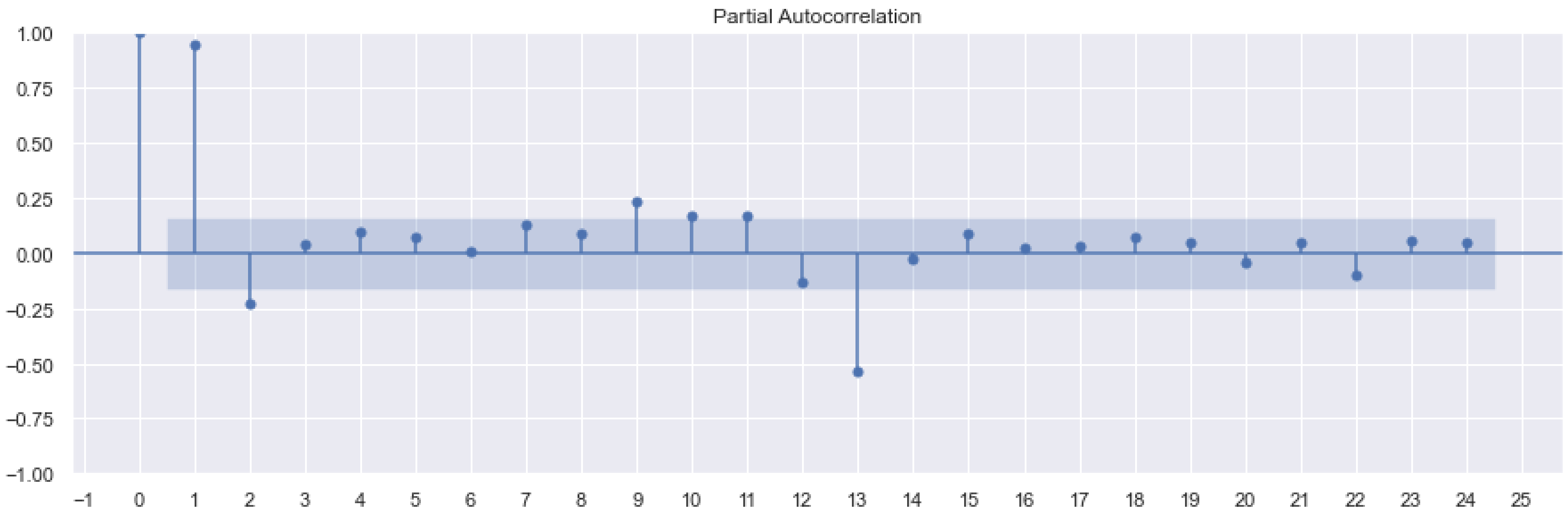
Autokorelacija – Parcijalni Autokorelogram

- Grafik vrednosti parcijalne autokorelacije za zadati raspon vrednosti *k*.
- Osenčen deo su vrednosti parcijalne autokorelacije koje bi sa 95% verovatnoće dobili za pozadinski šum.



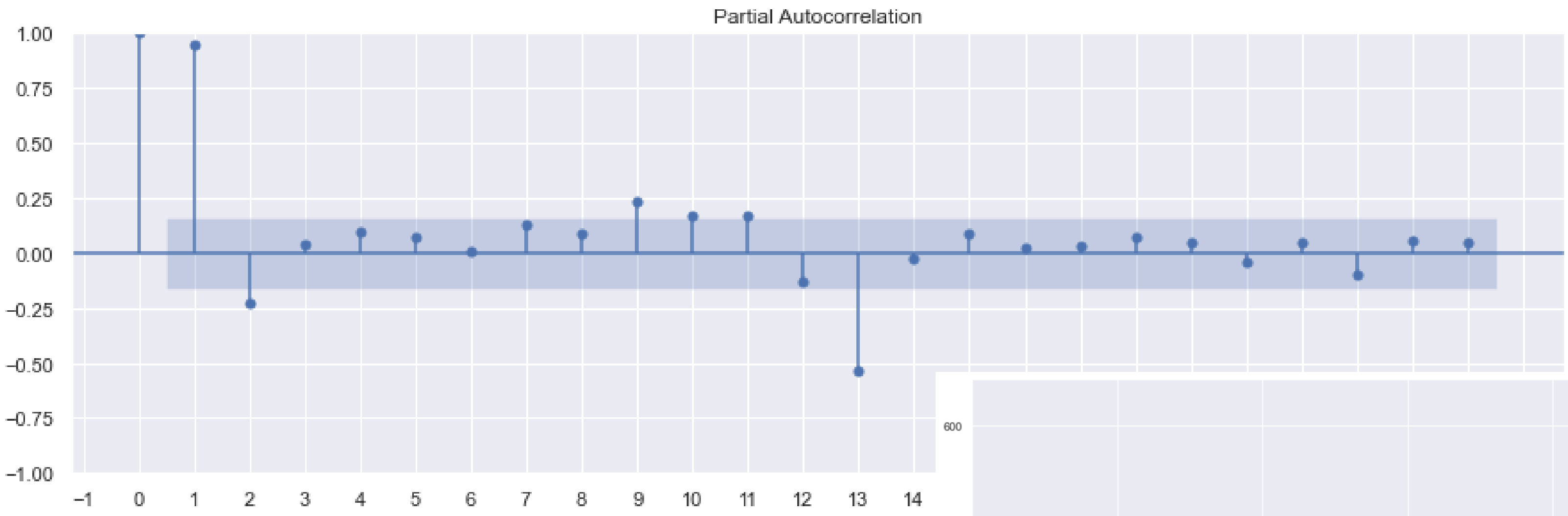
Autokorelacija – Parcijalni Autokorelogram – Tipični šabloni

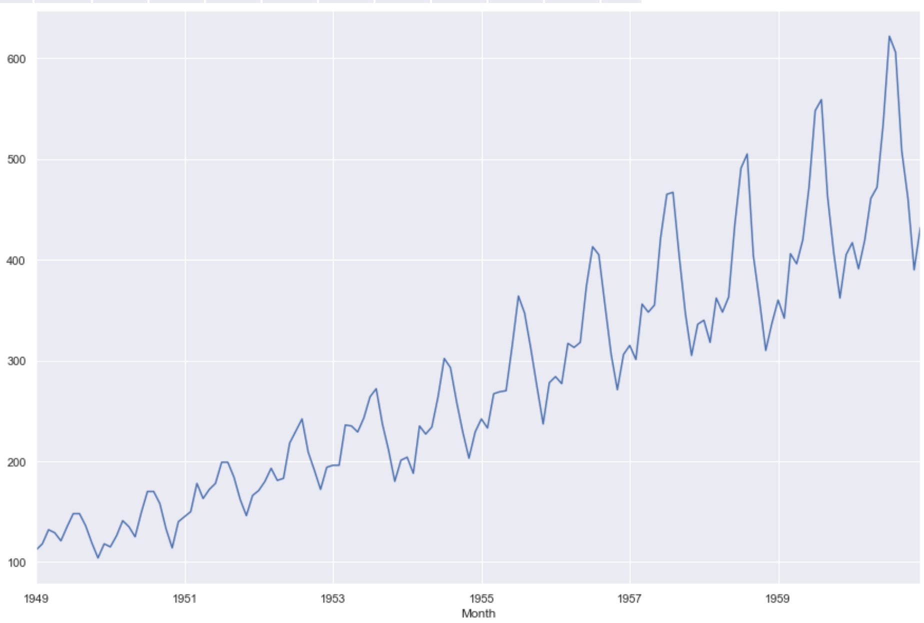
- Vremenske serije koje nisu slučajni šum obično imaju značajne vrednosti autokorelacije bar za nekoliko vremenskih jedinica unazad.
- Tipčan šablon je značajna vrednost za k=1 i potencijalno još nekoliko drugih, kao npr. za k=13 na grafiku ispod.



Autokorelacija – Parcijalni Autokorelogram – Primer

- Autokorelogram za podatke o avio-kompaniji.



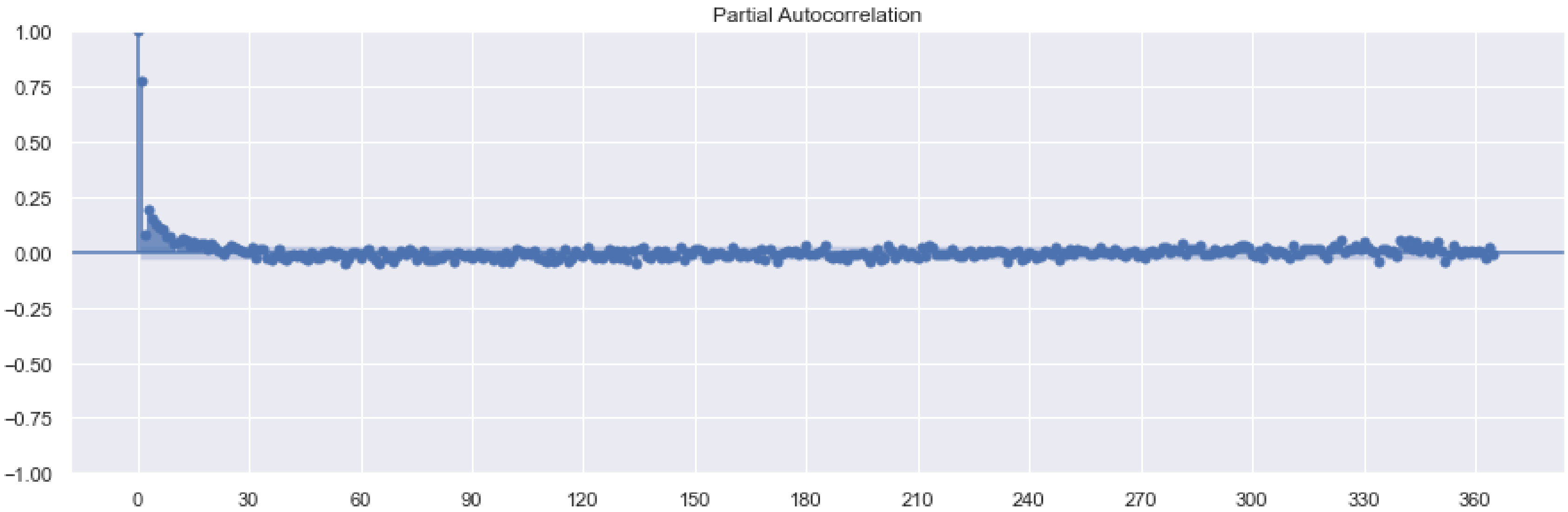


- Interesatno zapažanje je negativna vrednost za k=13.

- Ako je broj putnika pre 13 meseci porastao onda će sada opasti (i obrnuto).

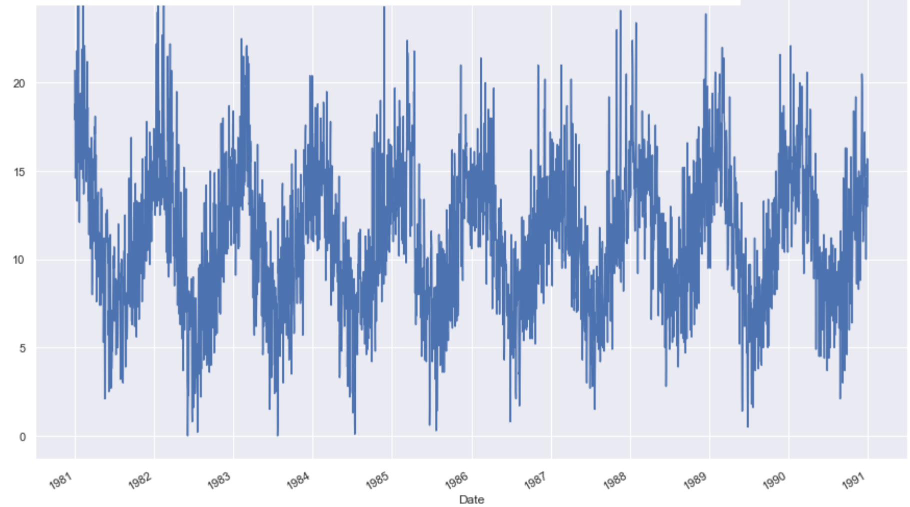
Autokorelacija – Parcijalni Autokorelogram – Primer

- Parcijalni autokorelogram ne otkriva sezonalnost.



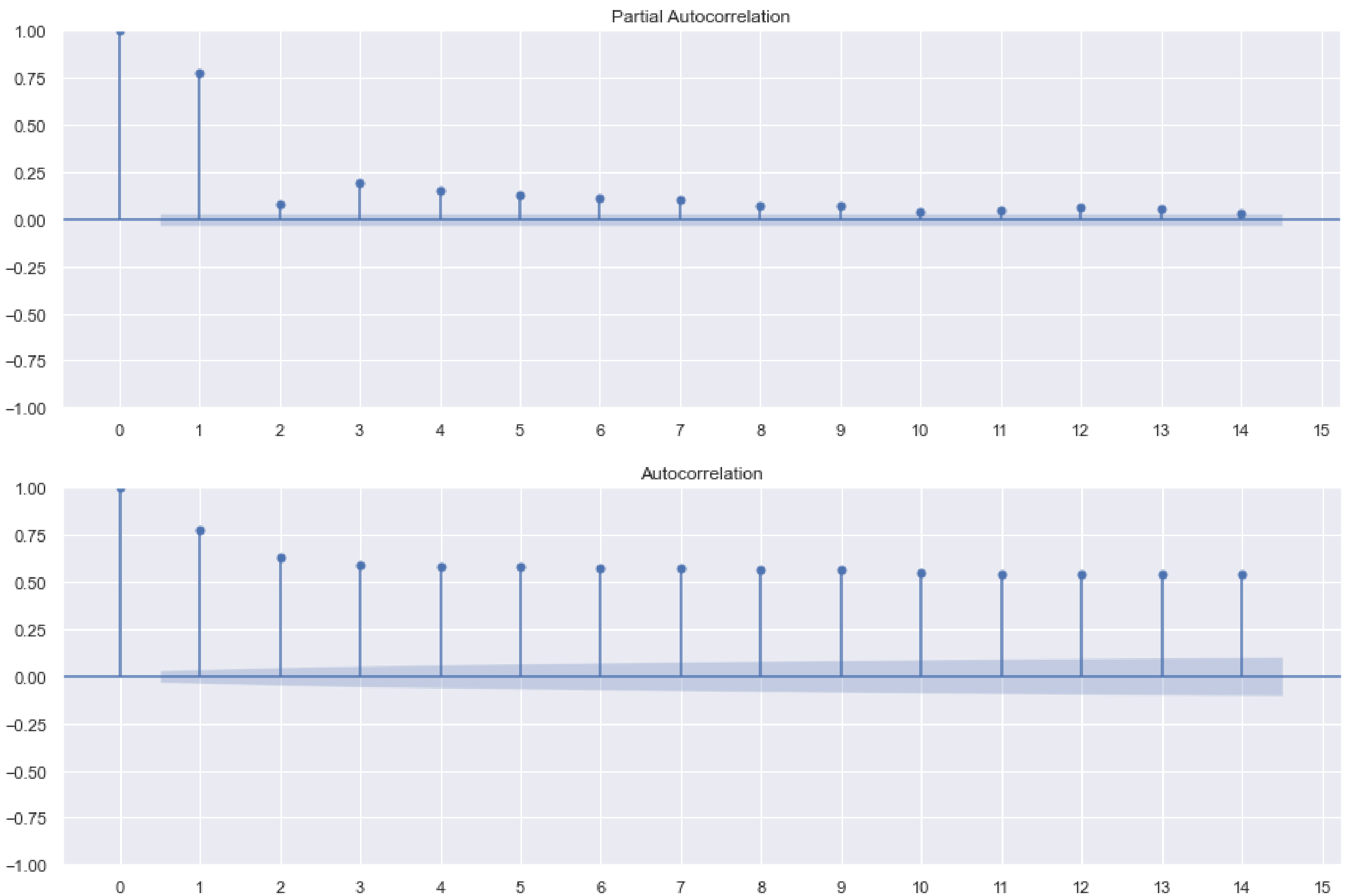
- Iznad su vrednosti parcijalne autokorelacije za jednu godinu.

- Sa desne strane ja grafike temperature za 10 godina.



Autokorelacija – Parcijalni Autokorelogram – Primer

- Sa gornjeg grafika vidi se direktan uticaj temperatura iz prethodnih 14 dana na temperaturu danas.
- Uporedite to sa autokorelogramom (donji grafik).



Autokorelacija – Parcijalni Autokorelogram – Upotreba

- Ima značajnu upotrebu u modelovanju.

- Prilikom modelovanja važno je odabrati koje tačno vremenske jedinice unazad će biti korišćene za predikciju trenutne vrednosti.

- Na primer, da li ćemo današnju temperaturu modelovati samo pomoću jučerašnje ili ćemo koristiti i temepraturu od pre dva, tri ili više dana.

- Na primer, da li ćemo broj putnika avio-kompanije modelovati samo pomoću vrednosti iz prethodnog meseca ili ćemo odabrati i vrednost od pre 13 meseci kao dobar indikator.

- Detalje upotrebe objasnićemo prilikom objašnjavanja samih prediktivnih modela.

Stacionarnost

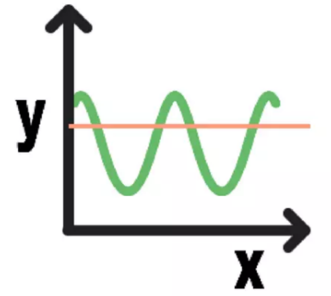
- Prediktivne metode u mašinskom učenju pretpostavljaju da su podaci na osnovu kojih se model formira iz iste distribucije.
- Distribuciju između ostalog karakterišu srednja vrednost, varijansa i u slučaju vremenskih serija kovarijansa.
- Stacionarnost intuitivno znači da se distribucija iz koje su dobijene vrednosti vremenske serije ne menja kroz vreme.
- Stacionarnost je značajna za modelovanje jer većina standarnih prediktivnih modela zahtevaju stacionarnost.

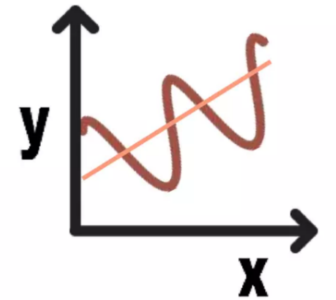
Stacionarnost - definicija

- Vremenska serija je stacionarna ako kroz vreme ima konstantnu:

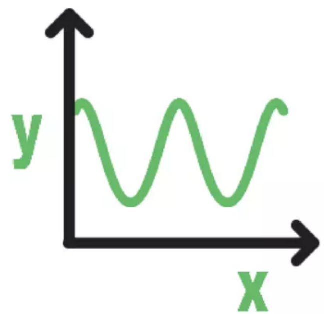
Stacionarnost | **Ne-Stacionarnost**

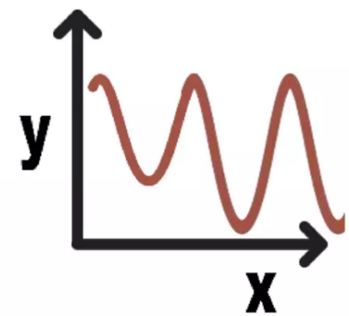
- ❑ **Srednju vrednost**



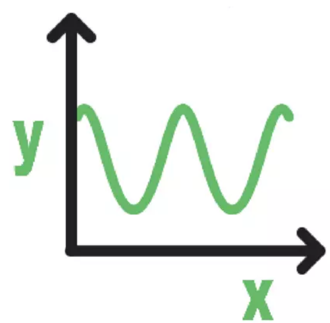


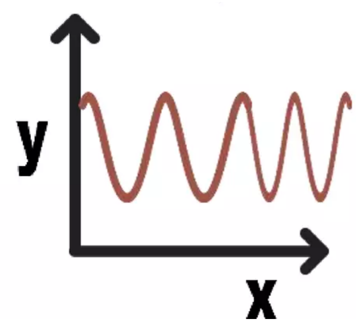
- ❑ **Varijansu**





- ❑ **Ko-varijansu**





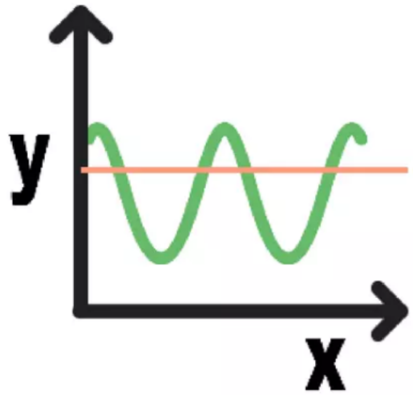
Stacionarnost - definicija

- Konstantna **srednja vrednost**.

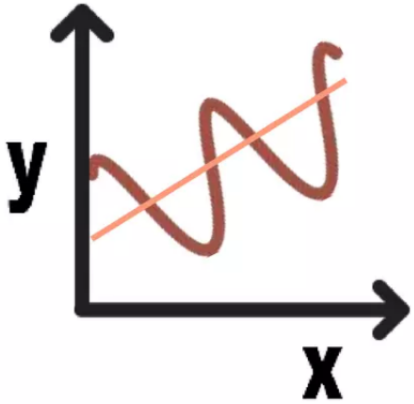
- Ako vremenska serija sadrži trend onda se srednja vrednost te serije menja po tom trendu.

- Vremenska serija koja sadrži trend nije stacionarna.

Stacionarnost



Ne-Stacionarnost



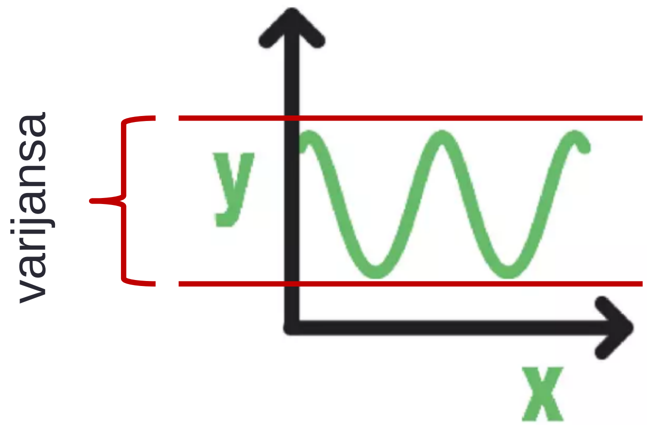
Stacionarnost - definicija

- Konstantna **varijansa**.

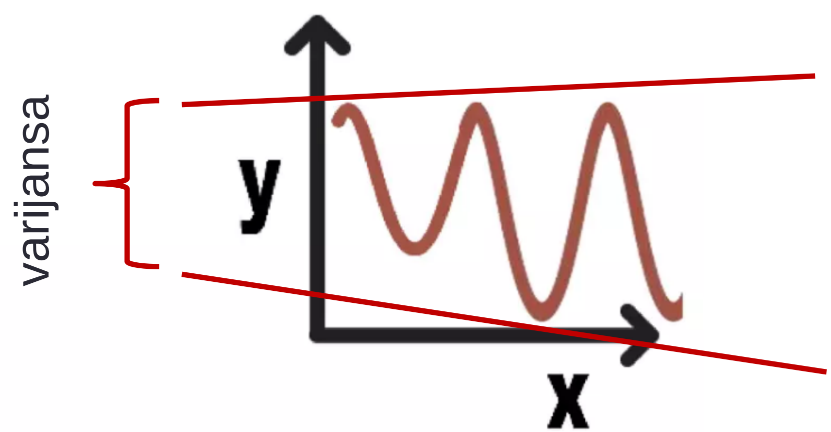
- Varijansa oko srednje vrednosti se ne menja kroz vreme.

- Intuitivno „širina“ vremenske serije po y-osi se ne menja kroz vreme.

Stacionarnost



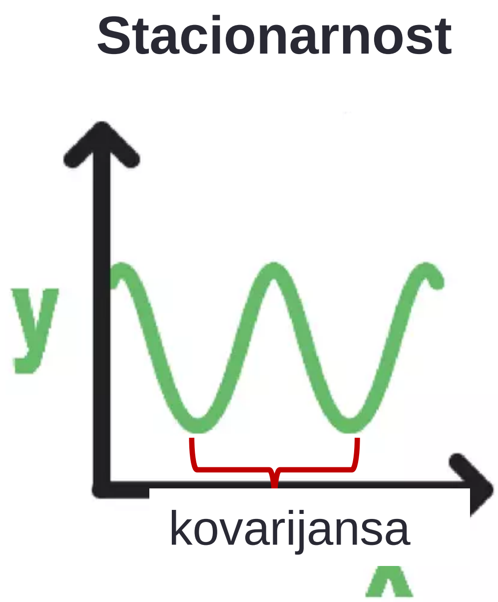
Ne-Stacionarnost

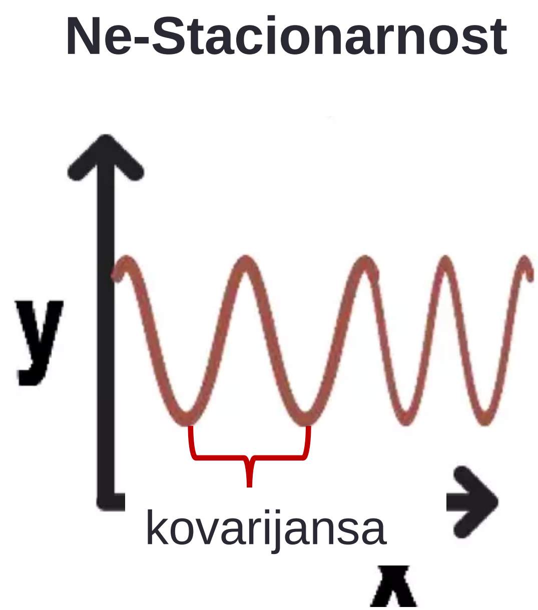


Stacionarnost - definicija

- Konstantna **kovarijansa**.

- Intuitivno „širina“ vremenske serije po x-osi se ne menja kroz vreme.

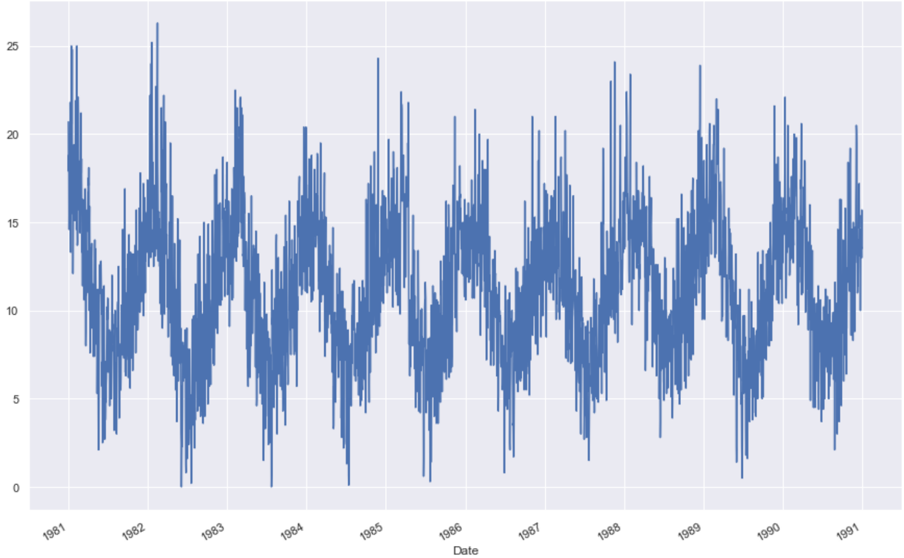




Stacionarnost i ne-stacionarnost

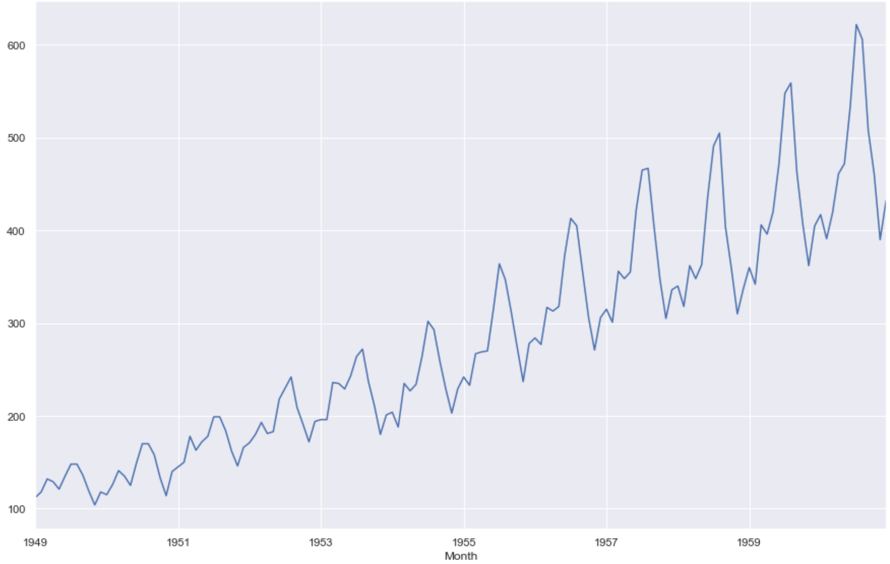
Stacionarnost

1. Konstantna srednja vrednost.
2. Konstantna varijansa.
3. Konstantna ko-varijansa.



Ne-stacionarnost

1. Srednja vrednost se menja po trednu.
2. Varijansa se povećava kroz vreme.
3. Ko-varijansa se smanjuje kroz vreme.



Stacionarnost i Modelovanje

- Kod ne-stacionarnih vremenskih serija odnos y_t i y_{t-k} se menja kroz vreme.

- Recimo da hoćemo da modelujemo odnos y_t i y_{t-k} sa auto-regresivnim modelom:

$$y_t = b_1 y_{t-k} + b_0$$

- Pošto se odnos y_t i y_{t-k} menja kroz vreme onda bi i parametri b_1 i b_0 morali da se menaju kroz vreme odnosno da b_1 i b_0 zavise od t:

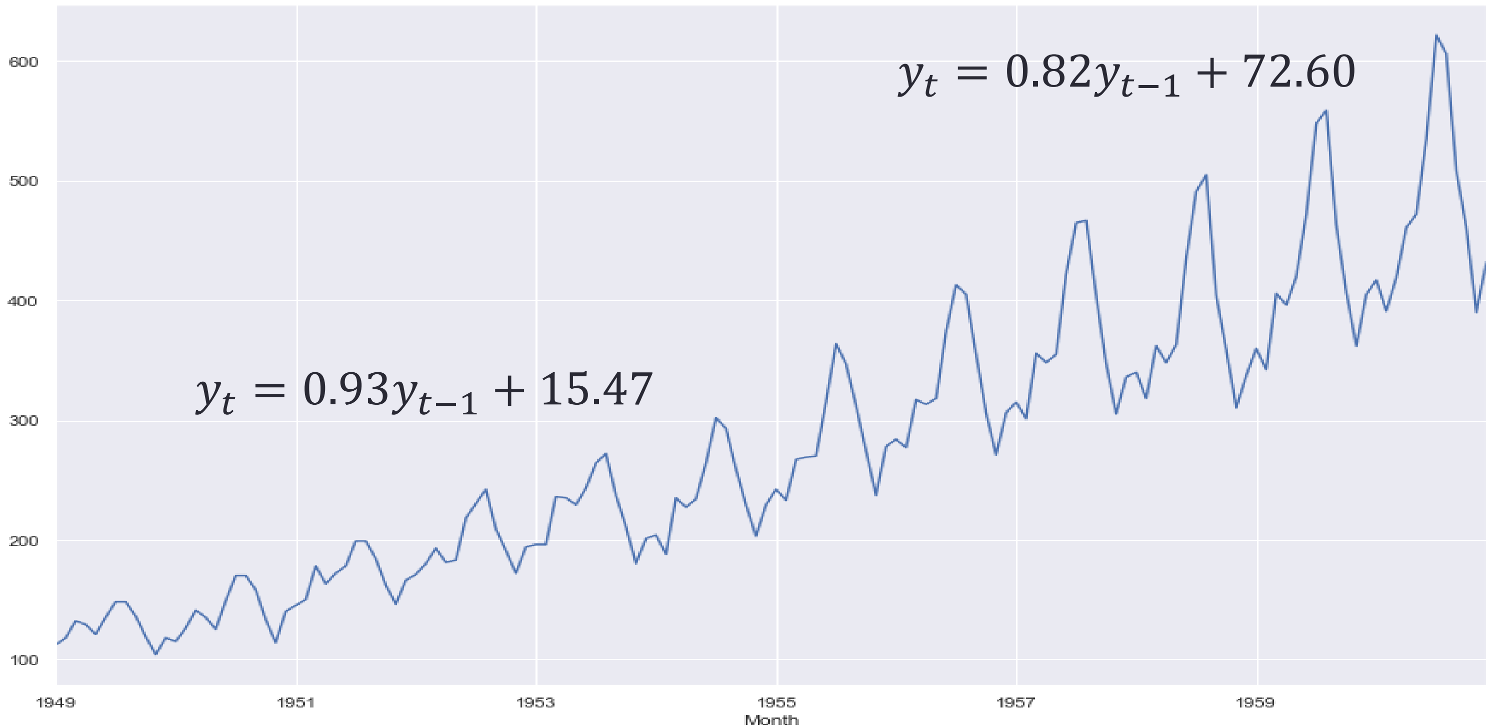
$$y_t = b_1(t) y_{t-k} + b_0(t)$$

- Recimo sada da želimo da uradimo ekstrapolaciju za neko t+k koje nemamo u obučavajućem skupu.
  - To znači da bi morali da znamo vrednosti b1(t+k) i b0(t+k) što nije moguće jer se vrednosti parametara određuju pomoću obučajućeg skupa, a vrednost y_{t+k} nemamo u obučavajućem skupu već želimo da je predvidimo.

- Iz navedenog sledi da ne možemo da obučimo model sa parametrima koji se menjaju kroz vreme.

Stacionarnost i Modelovanje

- Kod ne-stacionarnih vremenskih serija odnos y_t i y_{t-k} se menja kroz vreme.
- Recimo da hoćemo da modelujemo odnos y_t i y_{t-1} za primer sa avio-kompanijom koristeći auto-regresivni model: $y_t = b_1 y_{t-1} + b_0$.
- Formiramo poseban model za period od 1949 do 1955 i od 1955 do 1961.
- Sa grafika se vidi da se modeli značajano razlikuju.



Stacionarnost i Modelovanje

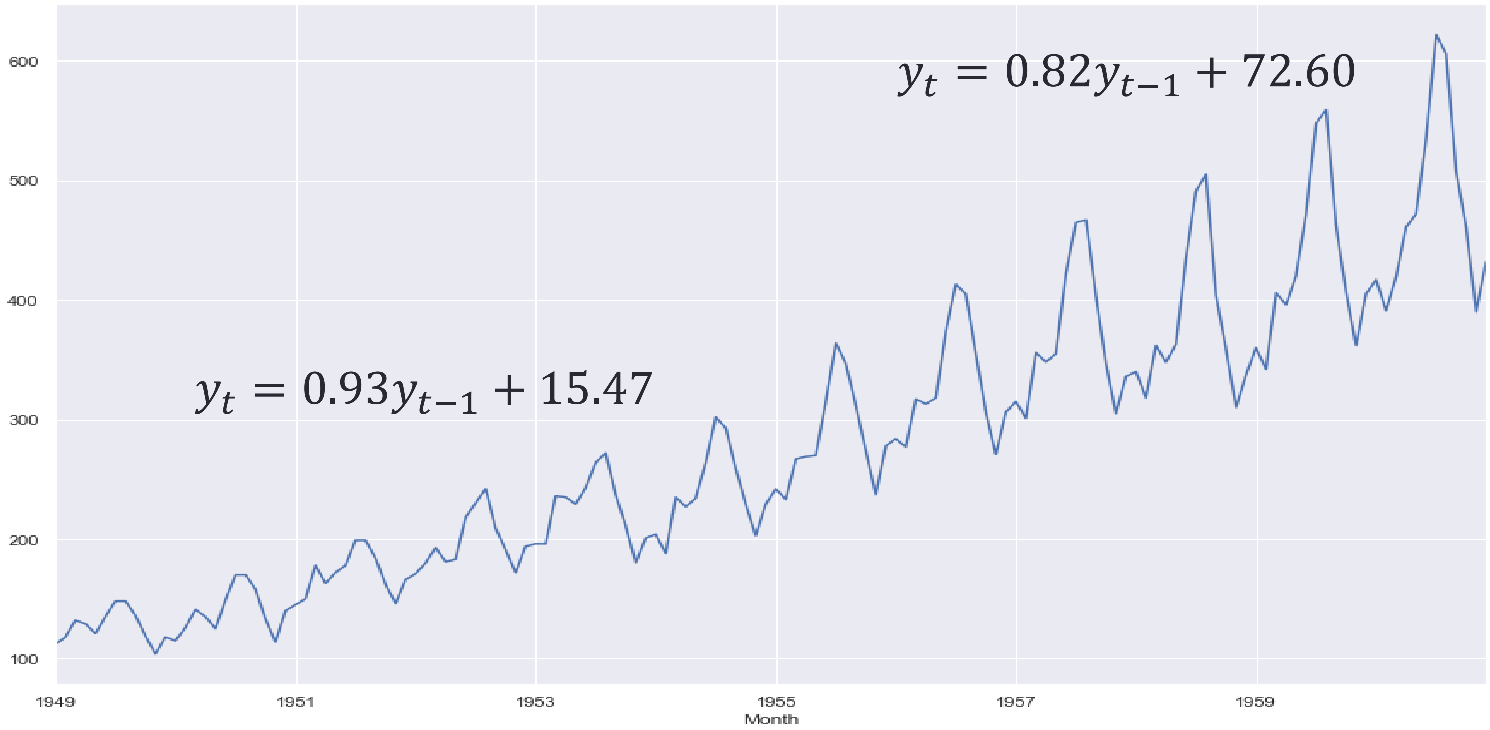
- Modeli za period od 1949. do 1955. i od 1955. do 1961. se razlikuju jer je odnos y_t i y_{t-1} drugačiji u ta dva perioda.

- Razlog za to je što vremenska serija nije stacionarna.

- To znači da bi naš model morao da menja svoje parametre po vremenu, odnosno da bude oblika:

$$y_t = b_1(p_i)y_{t-k} + b_0(p_i)$$

- gde su p_i periodi u kojima se menja odnos y_t i y_{t-1}.



Stacionarnost i Modelovanje

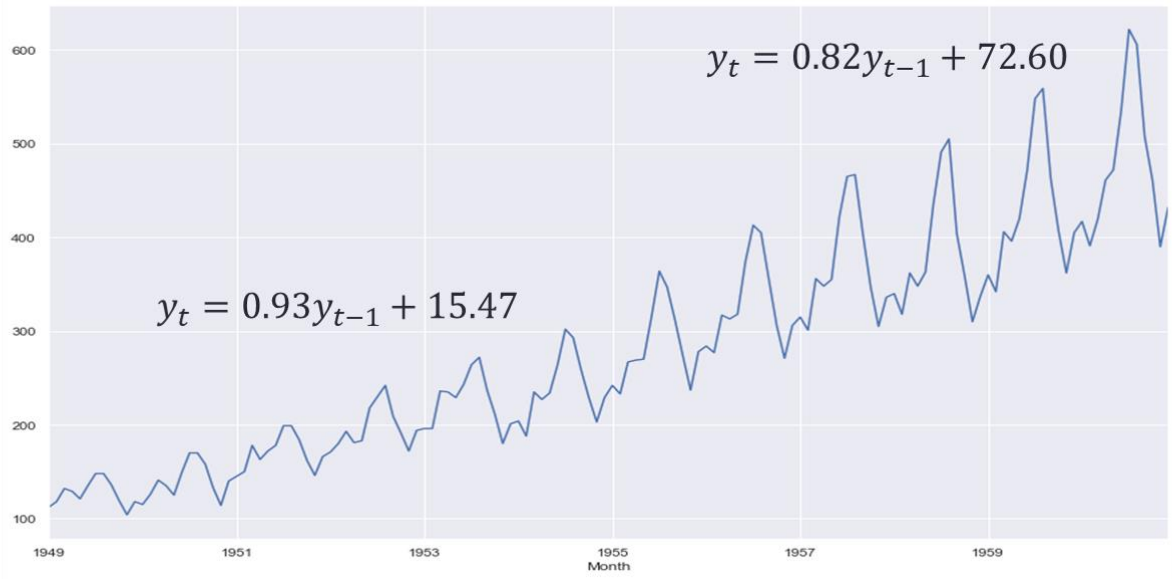
- Recimo sada da želimo da uradimo ekstrapolaciju za neko y_{t+k} koje se nalazi u nekom periodu p_{i+k} koji nemamo u obučavajućem skupu (na primer 1962. godina). Tada bi model imao sledeći oblik:

$$y_{t+k} = b_1(p_{i+k})y_{t+k-1} + b_0(p_{i+k})$$

- To znači da bi morali da znamo vrednosti $b_1(p_{i+k})$ i $b_0(p_{i+k})$ što nije moguće jer se vrednosti parametara određuju pomoću obučajućeg skupa, a vrednosti za period p_{i+k} nemamo u obučavajućem skupu.

- Iz toga sledi da ne možemo da obučimo model sa parametrima koji se menjaju kroz vreme.

- Zato to je stacionarnost važna za modelovanje.

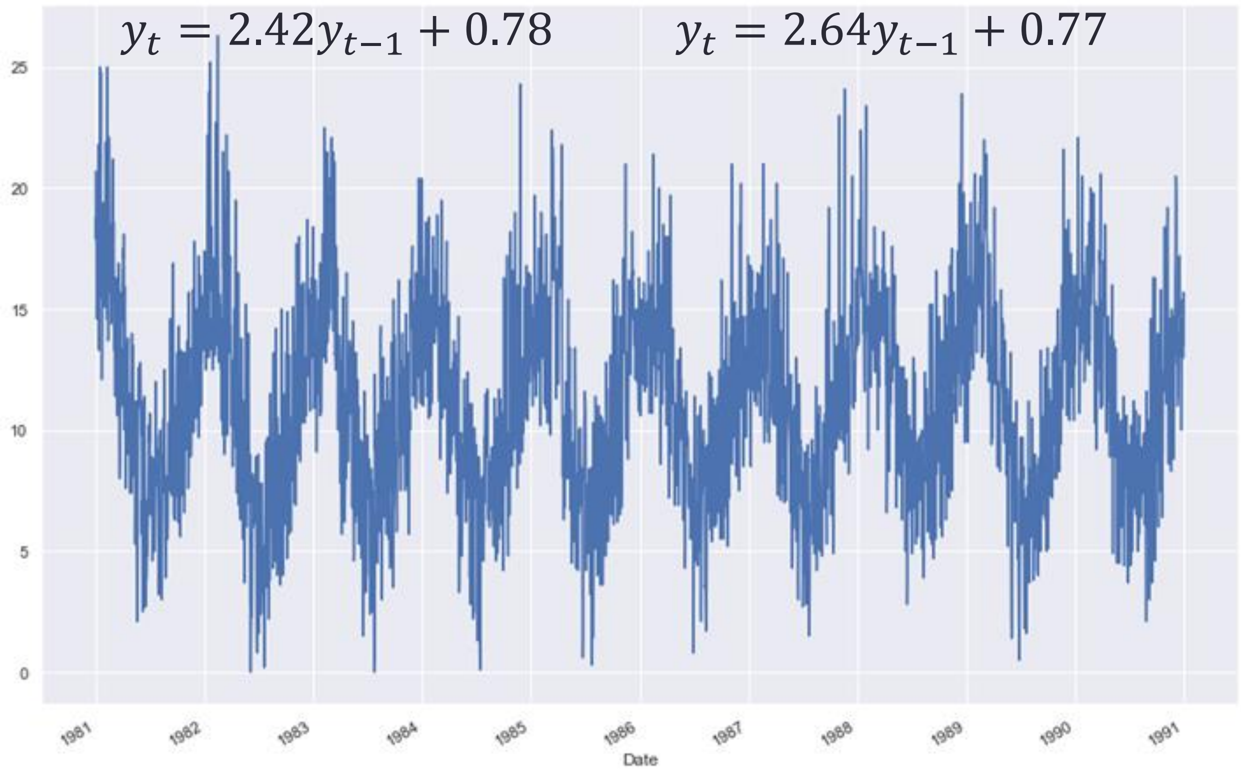


Stacionarnost i Modelovanje – Primer stacionarne serije

- Na grafiku je data temperatura vazduha od 1981. do 1991. koju modelujemo pomoću modela: $y_t = b_1 y_{t-1} + b_0$

- Vidimo da se modeli od 1981. do 1986. i od 1986. do 1991. veoma malo razlikuju.

- To je zato što je ovo primer stacionarne vremenske serije.



Stacionarnost – Testiranje

- Za testiranje vremenske serije na stacionarnost koristi se *Augmented Dickey–Fuller* (ADF) statistički test.

- ADF za hipotezu ima da vremenska serija nije stacionarna.

- Sam test za rezultat ima ADF statistiku koja se tumači na sledeći način:
  - ADF ≤ -3.432 – sa 99% pouzdanosti odbacujemo hipotezu
  - ADF ≤ -2.862 – sa 95% pouzdanosti odbacujemo hipotezu
  - ADF ≤ -2.567 – sa 90% pouzdanosti odbacujemo hipotezu

- Pouzdanost od 95% odnosno p-vrednost ≤ 0.05 dobićemo ako je ADF ≤ -2.862

- Dakle, vrednosti ispod -2.862 su poželjne i što su bliže -∞ to bolje.

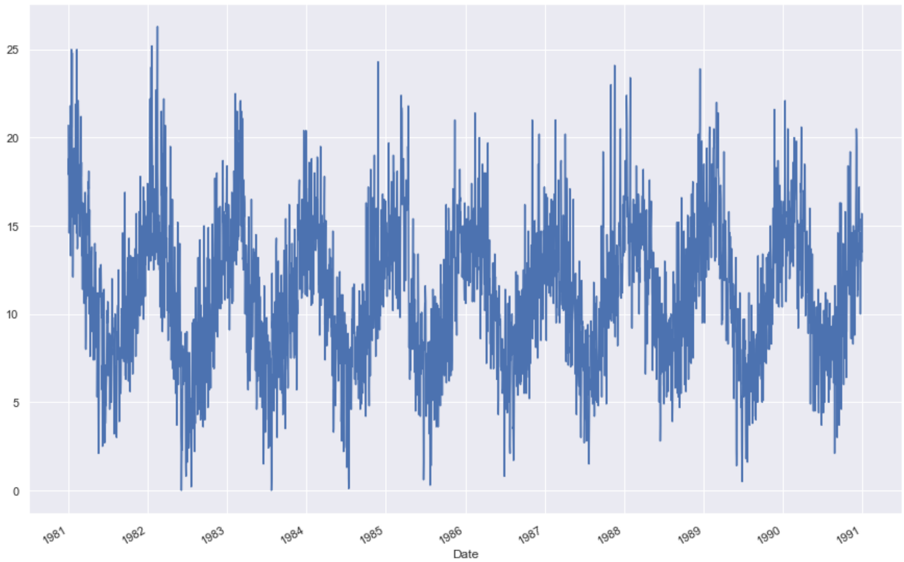
Stacionarnost – Testiranje

Stacionarnost

1. Konstantna srednja vrednost.
2. Konstantna varijansa.
3. Konstantna ko-varijansa.

ADF: -4.44

p-vrednost: 0.000247

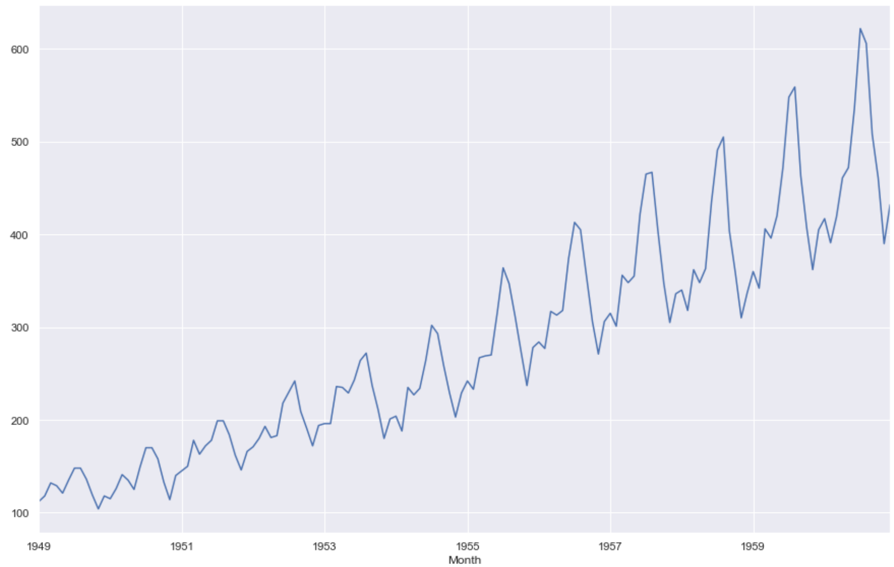


Ne-stacionarnost

1. Srednja vrednost se menja po trednu.
2. Varijansa se povećava kroz vreme.
3. Ko-varijansa se smanjuje kroz vreme.

ADF: 0.815

p-vrednost: 0.991880



Stacionarnost – Postizanje

- Vremensku seriju koja nije stacionarna ima bar jedan od tri problema:

1. Srednja vrednost se menja po trednu.
2. Varijansa se menja kroz vreme.
3. Ko-varijansa se menja kroz vreme.

- Problem 1. može se rešiti uklanjanjem tredna:
  - Otkrivanje trenda dekompozicijom i onda uklanjanje oduzimanjem ili deljenjem (pokazano ranije na prezentaciji).
  - Diferenciranjem vremenske serije (prikazano u nastavku prezentacije).

- Problemi 2. i 3. mogu se rešiti na sledeći način:
  - Uklanjanjem sezonalnosti:
    - Dekompozicijom i uklanjanje (pokazano ranije na prezentaciji).
    - Diferenciranjem vremenske serije (prikazano u nastavku prezentacije).
  - Transformacijama kao što su logaritmovanje ili korenovanje sa ciljem stabilizacije varijanse.

Stacionarnost – Postizanje, Diferenciranje

- Diferenciranje je određivanje numeričkog izvoda (konačnih razlika) za vremensku seriju.
- Za uklanjanje trenda tipično se koristi diferenciranje prvog ili drugog reda (prvi ili drugi izvod):

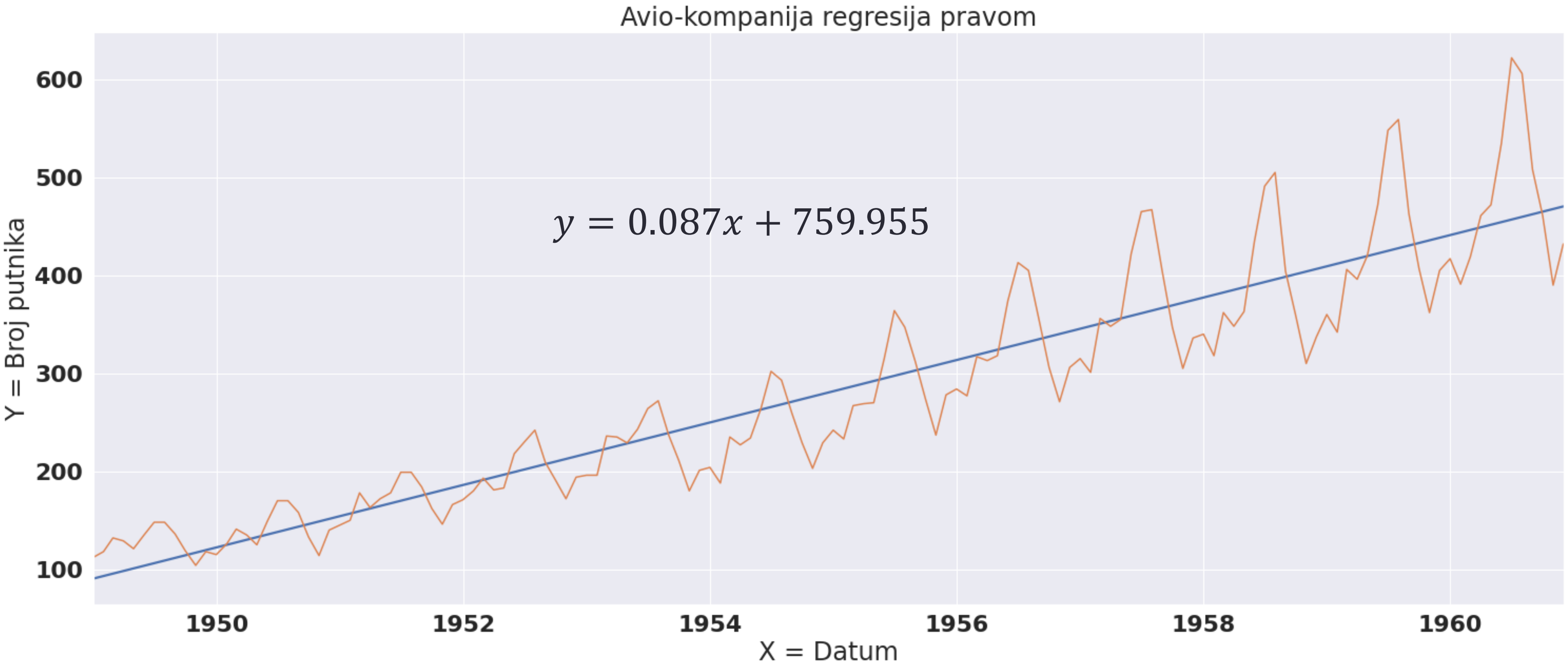
$$y'_t = \frac{y_t - y_{t-1}}{t - (t-1)} = y_t - y_{t-1}$$

$$y''_t = y_t - 2y_{t-1} + y_{t-2}$$

- Ako trend ima zakrivljenost onda se koristi drugi izvod ili se više puta uzastopno primeni diferenciranje prvog reda dok se ne postigne konstantna srednja vrednost.

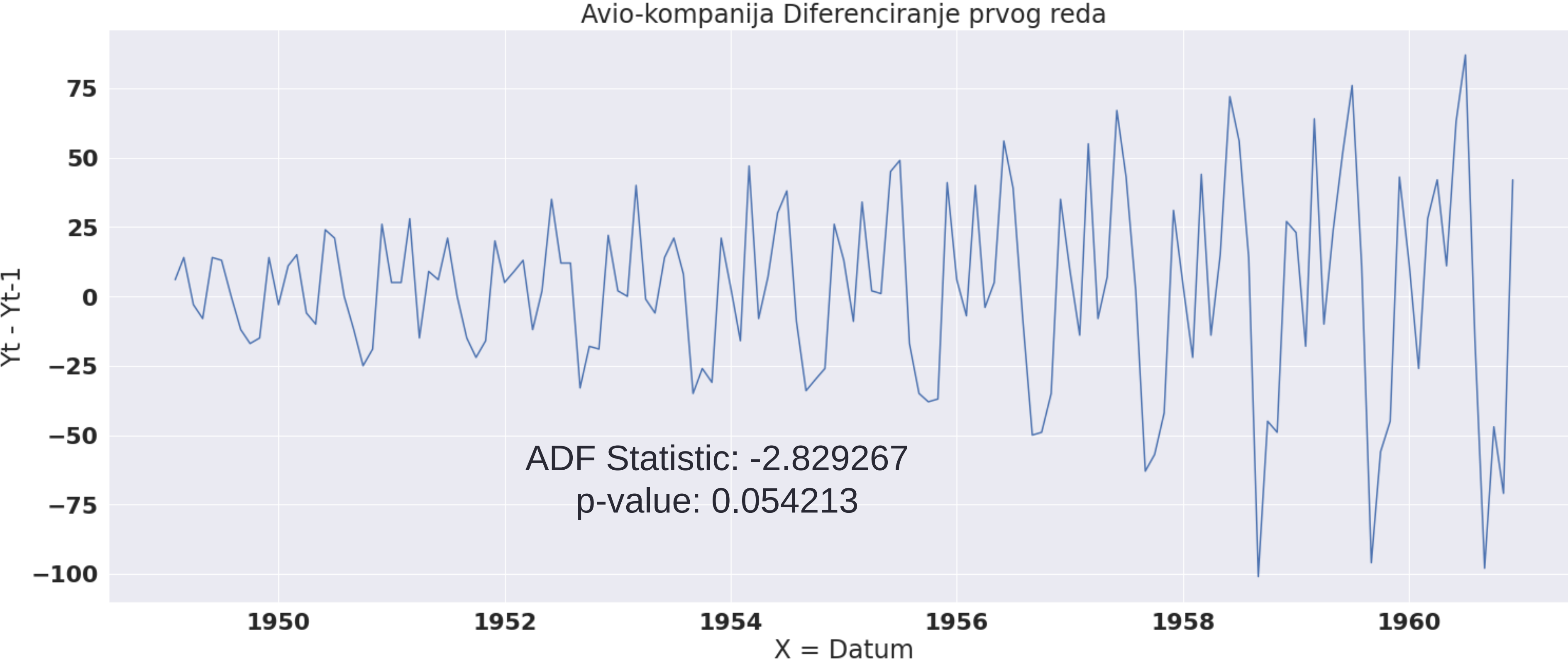
Stacionarnost – Postizanje, Diferenciranje

- Recimo da se srednja vrednost vremenske serije povećava linearno kao na grafiku.
- To znači da je razlika između dve uzastopne vrednosti serije konstantna (otprilike odgovara prvom izvodu tj. 0.087.
- Dakle, diferenciranjem prvog reda dobićemo vremensku seriju čija je srednja vrednost konstantna po vremenu.



Stacionarnost – Postizanje, Diferenciranje

- Nakon diferenciranja prvog reda sa grafika se vidi da vremenska serija ima konstantnu srednju vrednost (oko 0.087).
- Na taj način smo rešili problem 1. za postizanje stacionarnosti.
- Međutim vidi se da je probelm 2. prisutan (p-vrednost je na granici).
- U nastavku pokazujemo na koji način rešavamo probleme 2. i 3.



Stacionarnost – Postizanje, Diferenciranje po sezonalnosti

- Kao i trend, sezonalnost se takođe može ukloniti diferenciranjem po formuli:

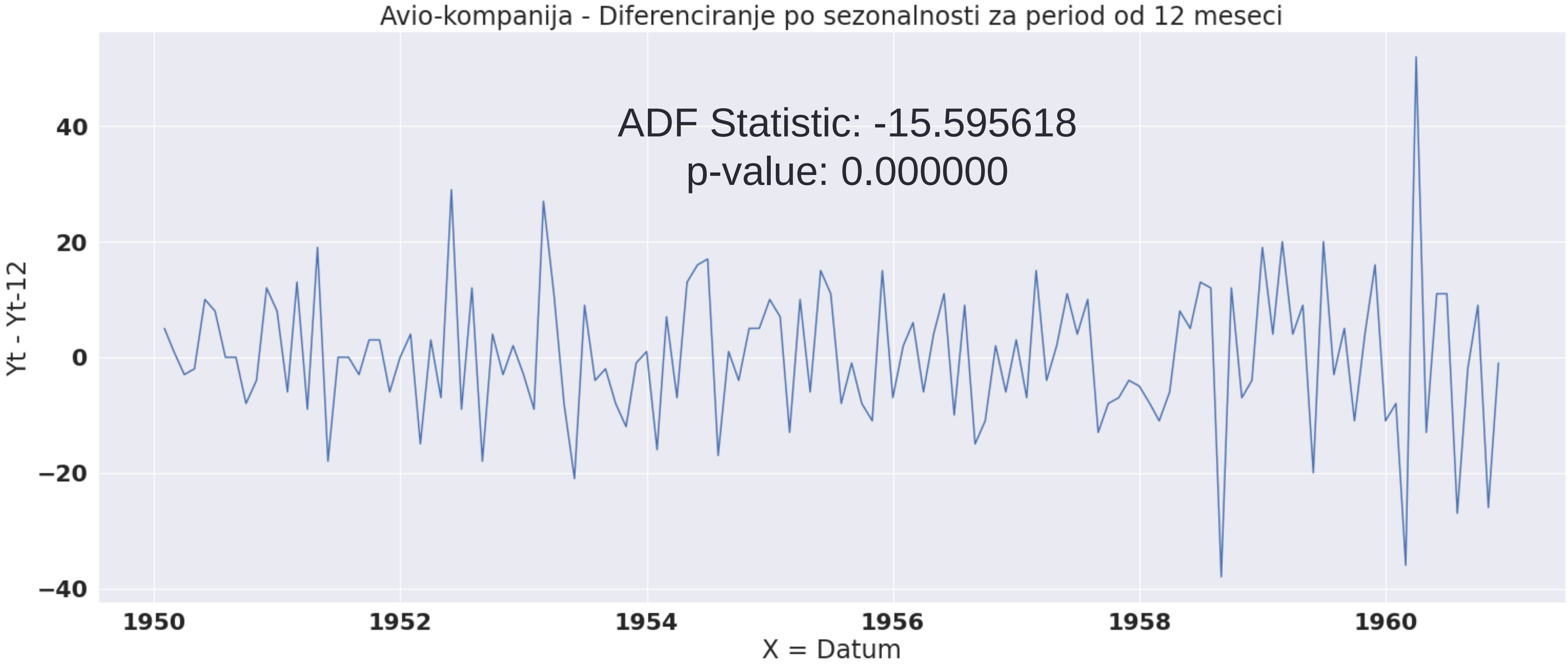
$$y'_t = y_t - y_{t-k}$$

gde je *k* period za koji traje sezonalnost.

- U nastavku vršimo diferenciranje za k=12 za grafik za prethodnog slajda jer su podaci o avio kompaniji dati na mesečnom nivou i imaju sezonalnost na godinu dana.

Stacionarnost – Postizanje, Diferenciranje

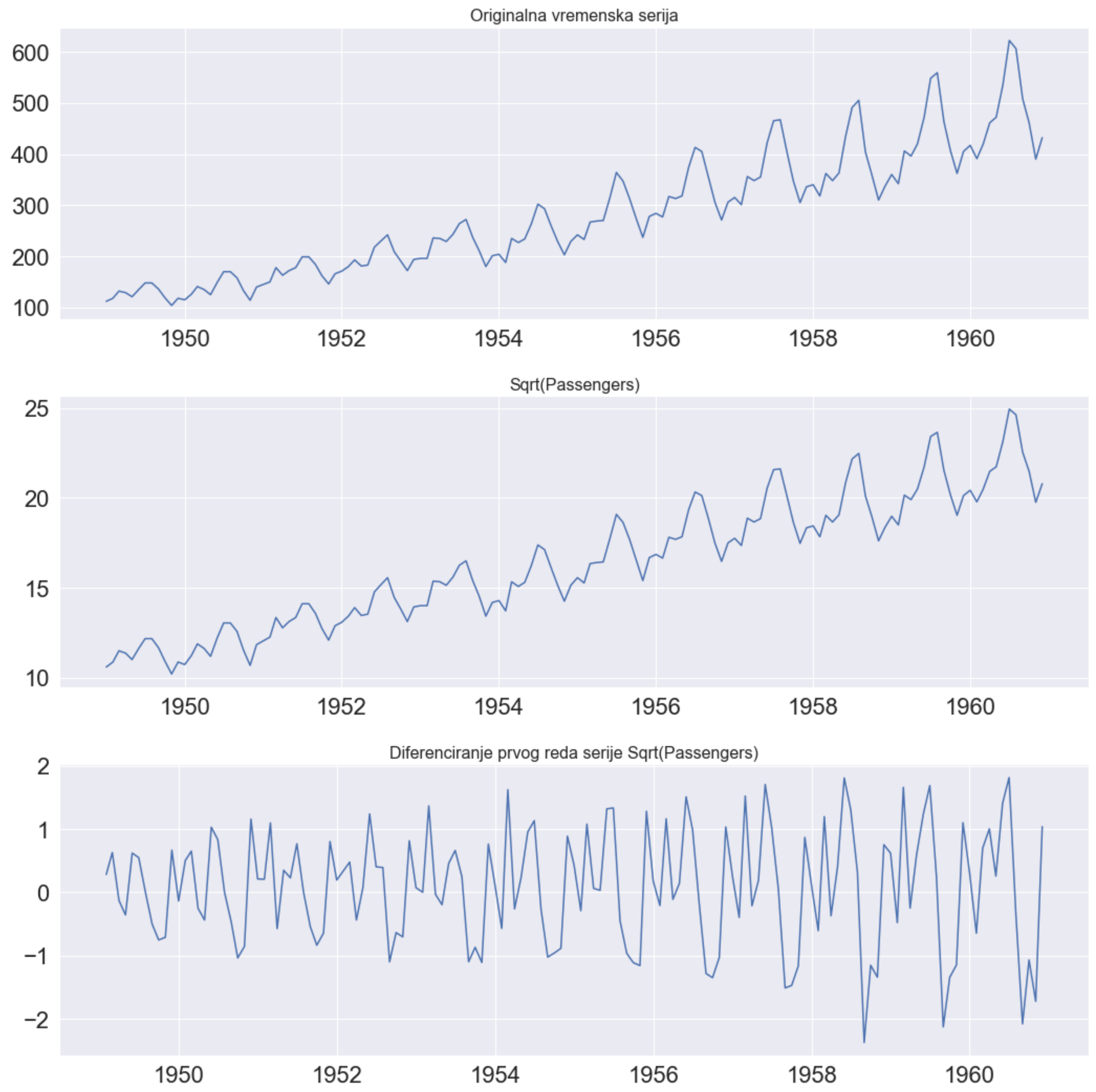
- Nakon diferenciranja prvog reda sa grafika se vidi da vremenska serija ima konstantnu varijansu oko svoje srednje vrednosti.
- Napomena, diferenciranje po sezonalnosti izvršeno je na rezultatima diferenciranja prvog reda.
  - Prvo je uklonjen trend pa onda sezonalnost.



Stacionarnost – Postizanje, Korenovanje

- Upotreba korena ili logaritamske transformacije u nekim slučajevima može da ublaži varijansu i reši probleme 2. i 3.
- Koristimo korenovanje jer su sve vrednosti vremenske serije avio-kompanije pozitivne.
- Tipično se transformacija vrši pre uklanjanja trenda sa ciljem da se nakon uklanjanja trenda dobije stacionarna vremenska serija.

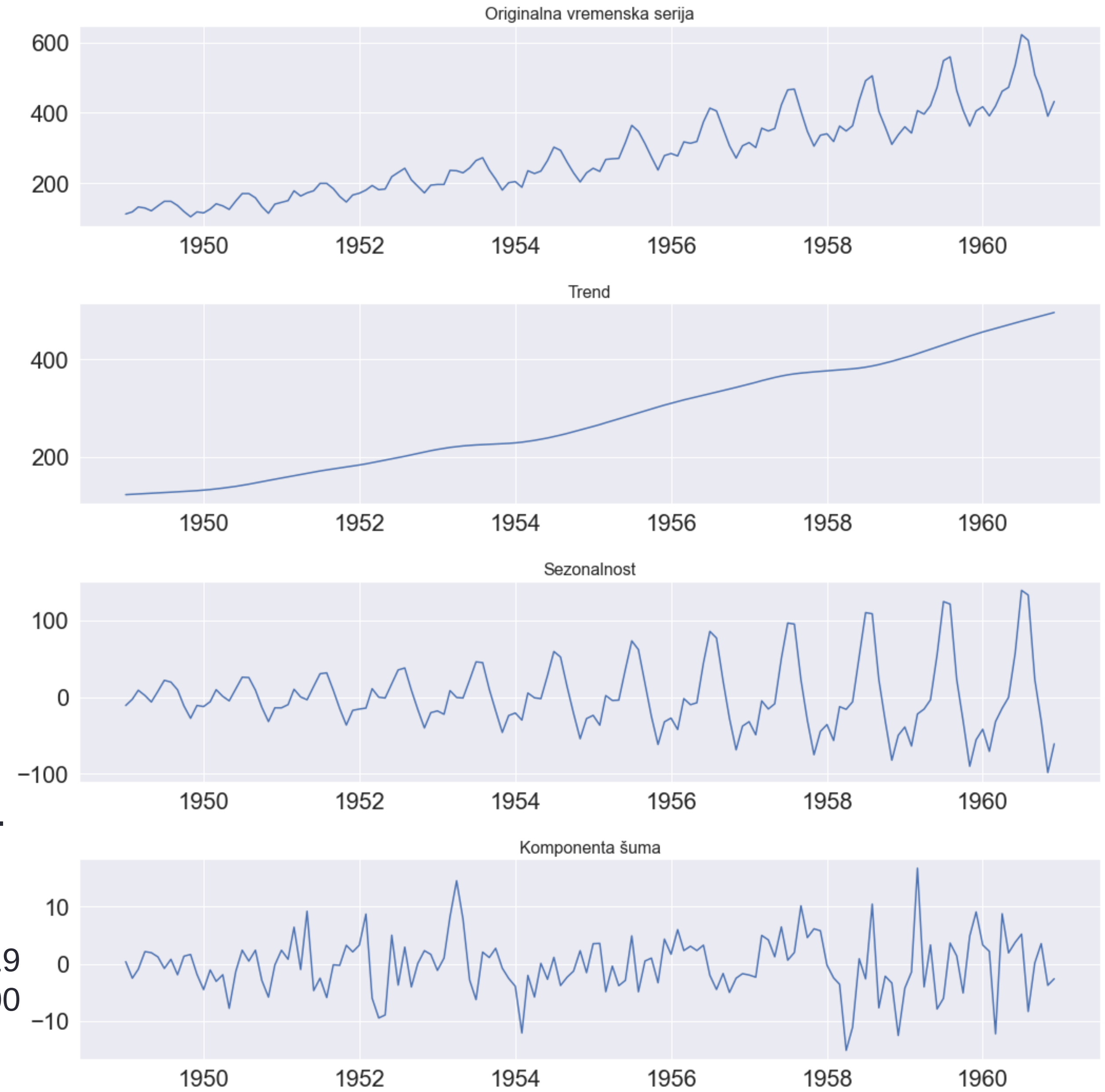
ADF Statistic: -3.186422
p-value: 0.020784



Stacionarnost – Postizanje, Dekompozicija

- Dekompozicijom vremenske serije na komponente može se postići stacionarnost.

- Rezultujuća komponenta šuma će najverovatnije biti stacionarna.
  - Ako to nije slučaj onda je potrebno prvo izvršiti log ili sqrt transformaciju orginalne vremenske serije.



ADF Statistic: -6.007619
p-value: 0.000000

PREDIKTIVNI MODELI ZA VREMENSKE SERIJE

Predikitvni modeli uopšteno

- Prediktivni modeli koji su predstavljeni u nastavku, pretpostavljaju da trenutne vrednosti vremenske serije zavise od prethodnih vrednosti te iste vremenske serije.
- To je vrlo intuitivan način modelovanja za vrednosti koje se menjaju po vremenu.
  - Očekivano je da vrednost danas zavisi od vrednosti juče, prekjuče itd.
- Razlike između modela ogledaće se u tome na koji način se prethodne vrednosti koriste za formiranje prediktivnog modela.

Predikitvni modeli uopšteno

- Prvo ćemo prikazati dva osnovna modela:

- Autoregresivni (AR) model
  - Trenutna vrednost je linearna kombinacija prethodnih.

- Model pokretnog proseka (MA model)
  - Trenutna vrednost je linearna kombinacija prethodnih grešaka i neke vrednosti.

- Nakon toga prikazaćemo dve kombinacije ovih modela:

- ARMA model
  - Trenutna vrednost je zbir AR i MA modela.

- ARIMA model
  - Trenutra vrednost je zbir AR i MA modela uz mogućnost diferenciranja serije sa ciljem uklanjanja trenda.

Autoregresivni model (AR)

- Vremensku seriju modelujemo kao linearnu kombinaciju prethodnih vrednosti:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + \varepsilon_t$$

gde je *k* broj vremenskih jedinica unazad koje koristimo, a ε_t je slučajna greška (kao što smo učili kod linearne regresije).

- Oznaka za autoregresivni model sa *k* vremenskih jedinica unazad je AR(*k*). Na primer AR(1):

$$y_t = \beta_1 y_{t-1} + \varepsilon_t$$

Autoregresivni model (AR)

- AR(*k*) je model linearne regresije kod koga kao nezavisne promenljive koristimo prethodne vrednosti vremenske serije:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + \varepsilon_t$$

- Parametre modela možemo odrediti pomoću Metoda Najmanjih Kvadrata (MNK) i tako dobijemo jednačinu AR(*k*) modela:

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_k y_{t-k}$$

AR model i stacionarnost

- AR model je model čiji se parametri određuju jednom u procesu obučavanja i ne menjaju se kroz vreme.

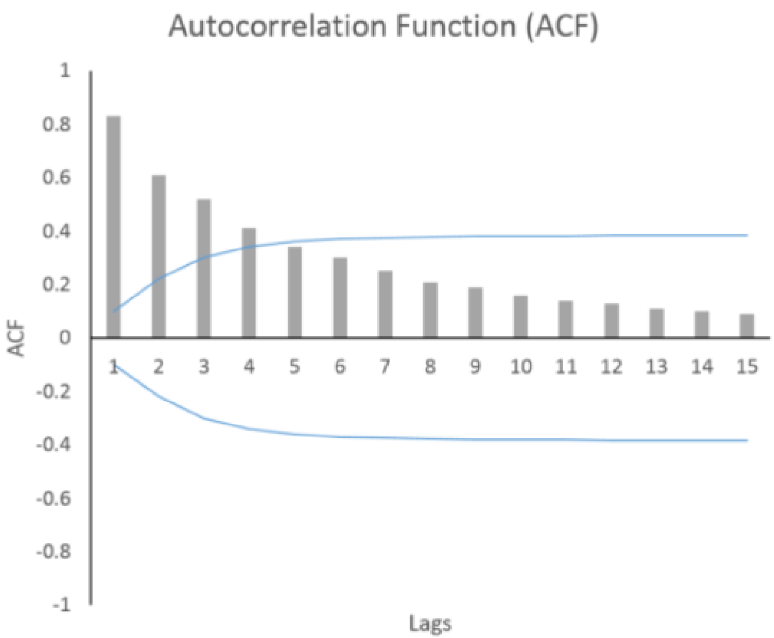
- Odnosno AR podrazumeva da linearna kombinacija pomoću koje računamo y_t uvek ima iste koeficijente, bez obzira na to da li računamo y_1, y_{10}, y_{100} itd.

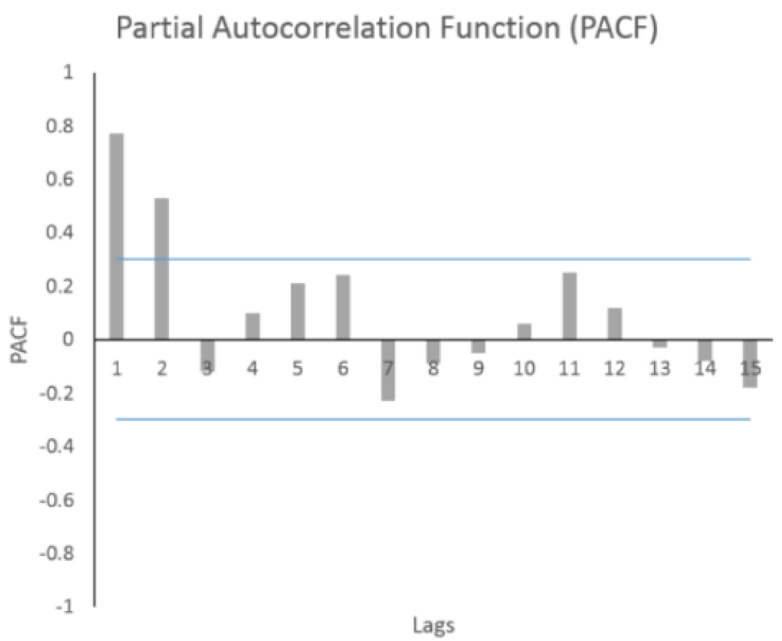
- To znači da se pomoću AR modela validno mogu modelovati samo vremenske serije čije karakteristike se ne menjaju kroz vreme.

- To su stacionarne vremenske serije.

AR model – ACF i PACF

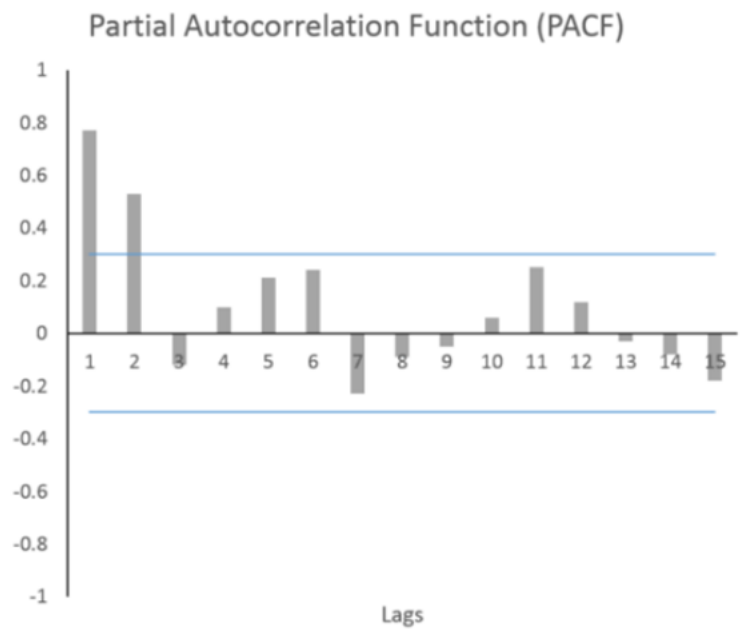
- Autokorelogram i parcijalni autokorelogram koriste se kao indikatori prikladnosti upotrebe AR modela na određenu vremensku seriju.
- Ako postoje značajne vrednosti ACF koje postepeno padaju to je indikator da postoji veza prethodnih i trenutnih vrednosti vremenske serije i da je AR model prikladan.
- Ako postoji samo par značajno većih vrednosti PACF za male vrednosti *k* i onda sledi nagli pad, onda je to indikator da je AR prikladan model.





AR model određivanje *k*

- Vrednost *k* kod AR(*k*) određuje se analizom parcijalnog autokorelograma.
- Sa autokoreloglama uzima se poslednje *k* posle kojeg su vrednosti PACF beznačajne (nalaze se unutra označenog dela).
- Sa grafika ispod zaključjemo da je *k*=2 odgovarajuća vrednost.



AR model – Primer avio-kompanija

- Prikaz Jypter notebook primera primene AR modela na vremensku seriju sa putnicima avio-kompanije.

Model pokretnog proseka (MA model)

- Vremensku seriju modelujemo kao linearnu kombinaciju prethodnih grešaka i srednje vrednosti vremenske serije:

$$y_t = \mu + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_k \varepsilon_{t-k} + \varepsilon_t$$

gde je *k* broj vremenskih jedinica unazad koje koristimo, μ je srednja vrednost vremeske serije, a ε_t je slučajna greška (kao što smo učili kod linearne regresije).

- MA model podrazumeva da slučajna greška ima normalnu distribuciju.

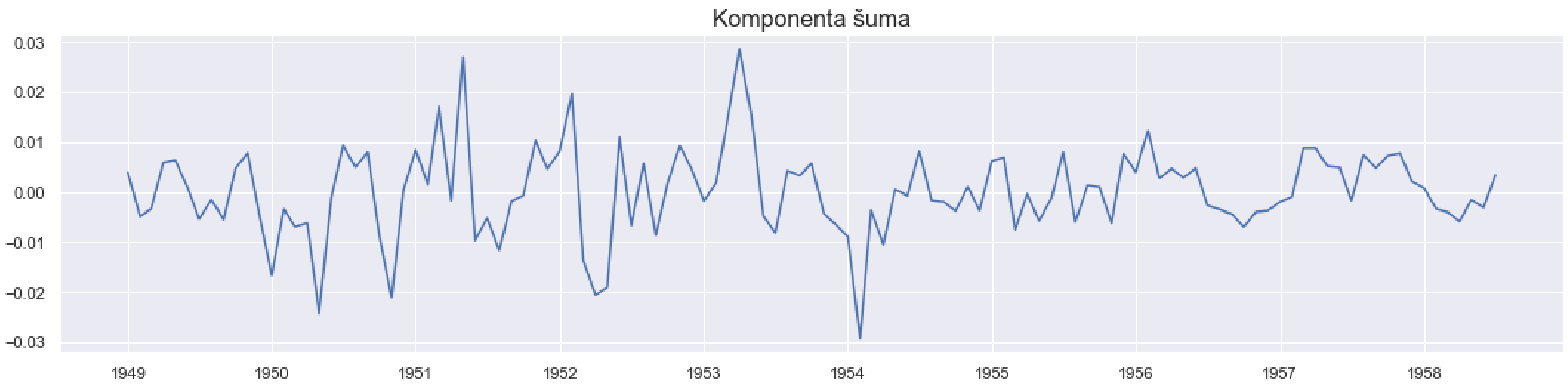
- Oznaka za model pokretnog proseka sa *k* vremenskih jedinica unazad je MA(*k*). Na primer MA(1):

$$y_t = \mu + \beta_1 \varepsilon_{t-1} + \varepsilon_t$$

- Po samoj definciji modela (srednja vrednost se ne menja kroz vreme, varijansa grešaka je konstantna) može zaključiti da MA model podrazumeva stacionarnost vremenske serije.

Model pokretnog proseka (MA model)

- MA model nije toliko intuitivan kao AR model pa ćemo zato u nastavu objasniti ideju MA modela.
- Kao prvi korak pogledaćemo grafik komponente šuma logaritmovane vremenske serije avio-kompanije.
- Sa grafika se vidi da vremenska serija varira oko srednje vrednosti 0.
- Svaka varijacija (skok ili pad) ima neko svoje trajanje posle koga prestaje.
- Ideja MA modela je da je trenutna vrednost vremenske serije zbir srednje vrednosti i linearne kombinacije varijacija u poslednjih *k* vrememenskih jedinica.

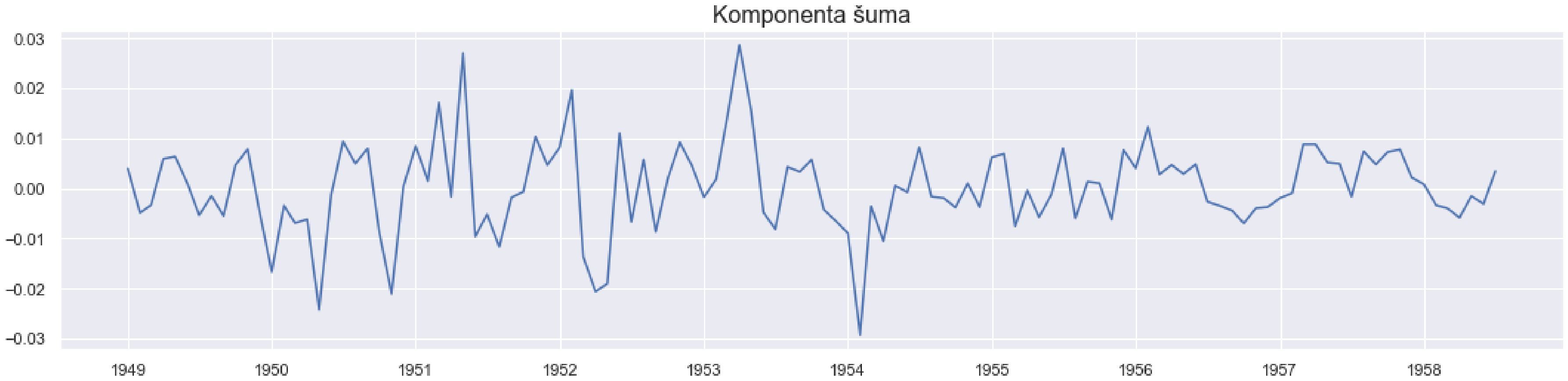


Model pokretnog proseka (MA model)

- Ako zamislimo srednju vrednost kao ravnu liniju, a svaku varijaciju kao neki mali potres ravne linije koji ima neko svoj trajanje (kao npr. zemljotres).

- MA model modeluje na koji način će se poremetiti ravna linija pod uticajem svih potresa u poslednjih *k* vremenskih jedinica.

- Potresi ranije od *k* nemaju uticaj na trenutnu vremensku vrednost jer je se pretpostavlja da je njihovo trajanje isteklo.



Model pokretnog proseka (MA model)

- Ilustrovaćemo sada MA(1) modela na jednom primeru. Recimo da mesec dana svake nedelje organizujete žurku i da za sledeći mesec hoćete procenite koliko piva da kupite za svaku žurku.

- Recimo da ste prošlog meseca u proseku kupovali 20 piva po žurki.

- Rešili ste onda da za prvu žurku ovog meseca kupite 20 piva.

- Onda da za svaku sledeću žurku kupite: 20 piva + 2 * (onoliko piva koliko je ostalo ili je nedostajalo prethodne nedelje):

$$y_t = \mu + \beta_1 \varepsilon_{t-1} + \varepsilon_t \qquad pivo_t = 20 + 2\varepsilon_{t-1} + \varepsilon_t$$

- Količina piva koja je preostala ili je nedostajala prati normalnu distribuciju sa srednjom vrednosti 0 i standardnom devijacijom 1.

- Na taj način formirali ste MA(1) za predikciju piva koje bi trebalo da kupite.

- Na sledećem slajdu prikazujemo kako bi funkcionisao taj model na primeru jednog meseca.

Model pokretnog proseka (MA model)

- Ilustrovaćemo sada MA(1) na primeru sa žurkom:

$$\widehat{pivo}_t = 20 + 2\varepsilon_{t-1}$$

t	$\widehat{pivo}_t$	$\varepsilon_t \sim N(0,1)$	$pivo_t = \widehat{pivo}_t + \varepsilon_t$
1	20	2	22
2	$20 + 2 \cdot 2 = 24$	-1	23
3	$20 + 2 \cdot (-1) = 18$	-2	16
4	$20 + 2 \cdot (-2) = 16$	1	17

- $pivo_t$ - koliko piva je bilo stvarno potrebno
- $\widehat{pivo}_t$ - predikcija piva

Model pokretnog proseka (MA model)

- Proces određivanja parametara za MA model je značajno komplikovaniji u odnosu na AR model. Detalji samog procesa su van opsega ovog kursa.

- Nakon što odredimo parametre, model koji koristimo za predikcije ima sledeći oblik:

$$\hat{y}_t = \mu + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots + b_k\varepsilon_{t-k}$$

- Primećujemo da u modelu nema vrednosti ε_t.

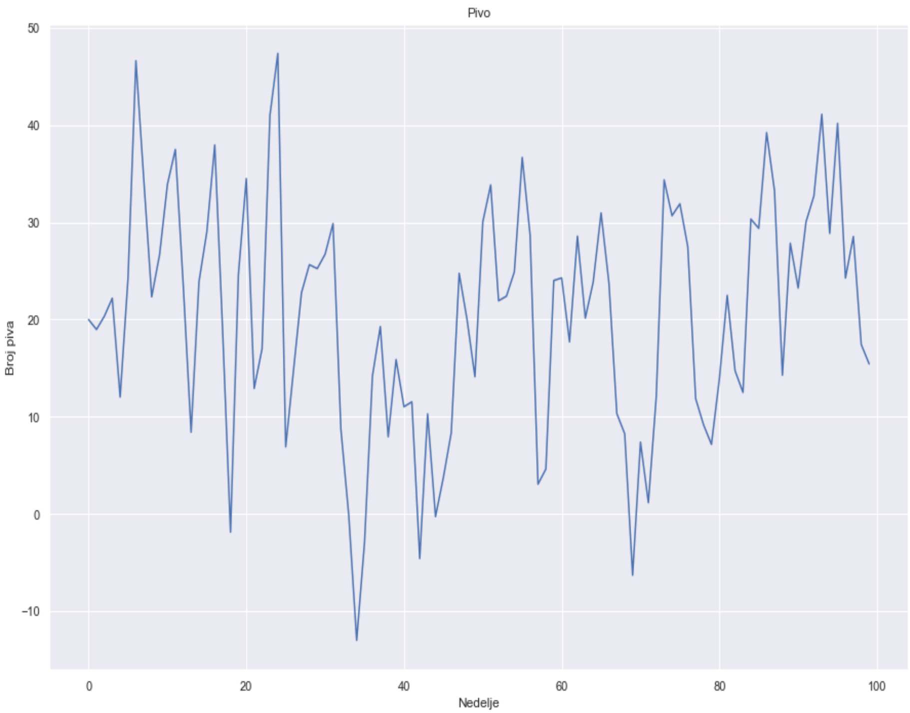
- To je zato što je ε_t deo modela koji reprezentuje grešku za y_t koje predviđamo.

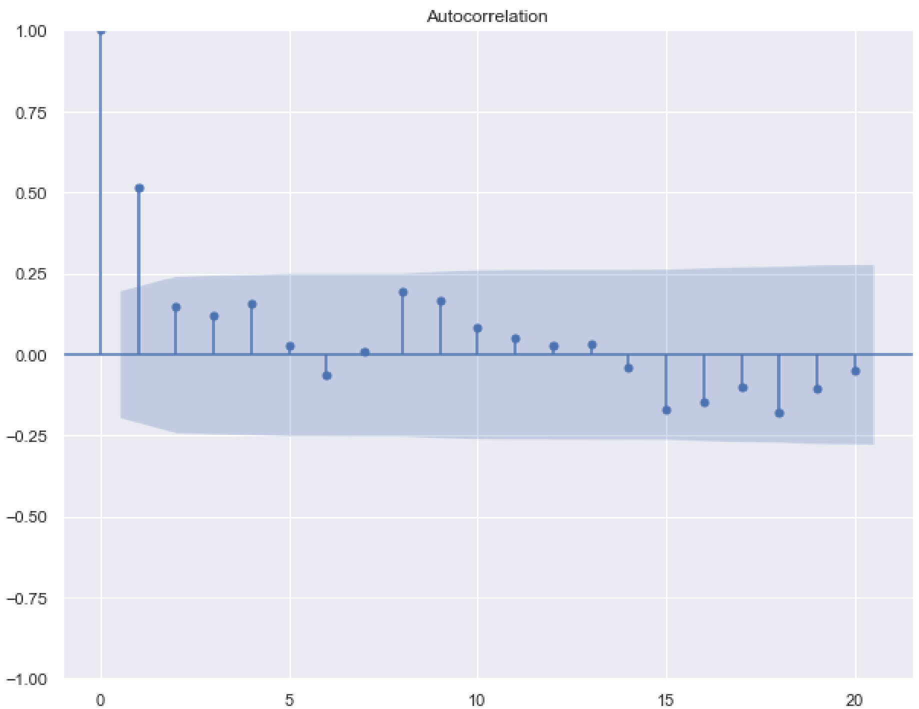
- Ne možemo unapred znati grešku za vrednost koju predviđamo.

- Grešku ε_t izračunavamo kada uradimo predikciju i onda je koristimo u predikciji y_{t+1}.

MA model – ACF i PACF

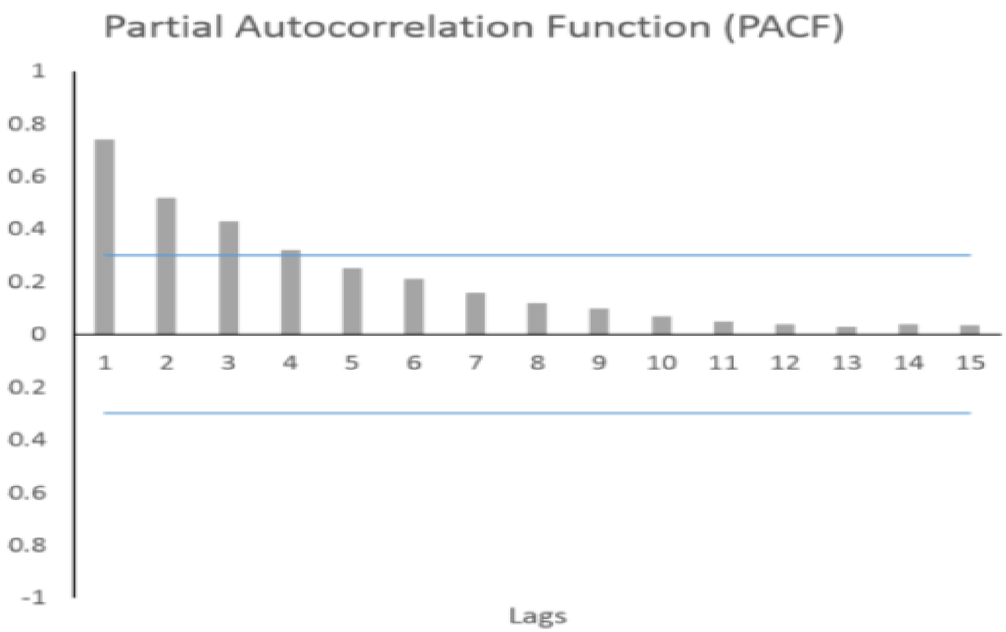
- Recimo da nastavimo proces predikcije količine piva za žurku do 100 predikcija (grafik levo) i nakon tog odredimo autokorelogram tog procesa (grafik desno).
- Autokorelogram ima nagli pad za *k*=1 i posle su vrednosti beznačajne.
- Tipičan indikator za upotrebu MA(*k*) modela je nagli pad posle *k*-te vremenske jedinice unazad.





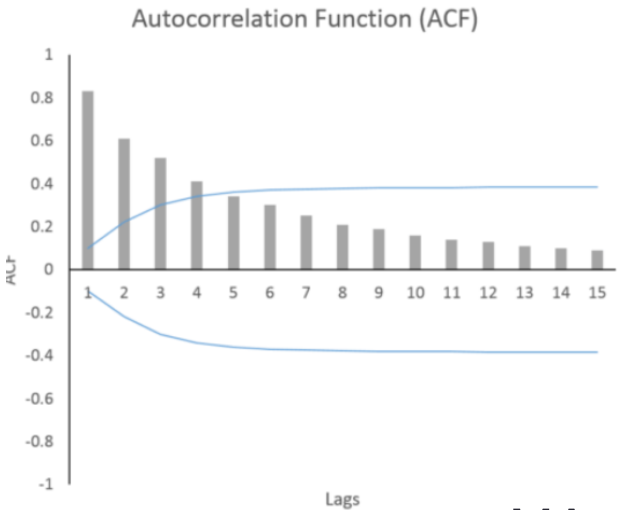
MA model – ACF i PACF

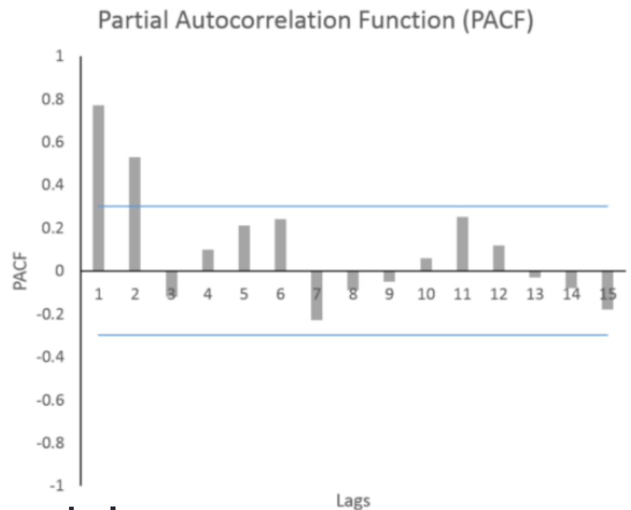
- Parcijalni autokorelogram ima manje važnu ulogu kod MA modela i može da se iskoristi samo indikator da vremenska serija nije pozadinski šum.
- Parcijalni autokorelogram ima postepen pad sa porastom *k*.
- Imamo obrnutu situaciju u odnosu na AR model kod koga je postepen pad bio za ACF.
- Na sledećem sladju dato je poređenje grafikona za AR i MA.



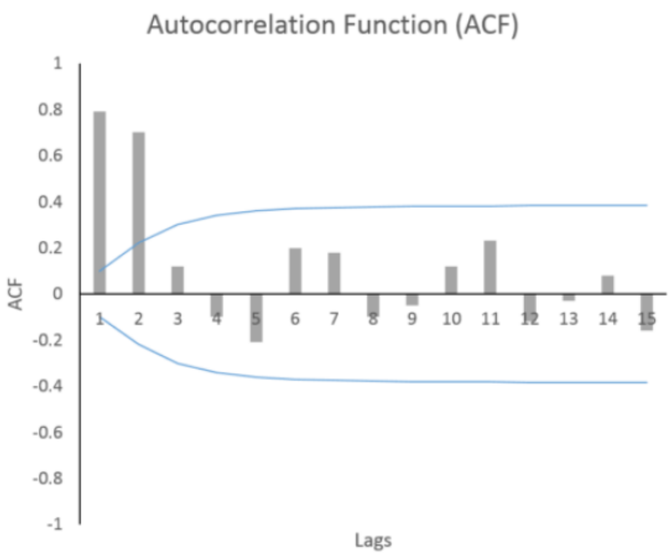
MA i AR model – poređenje ACF i PACF

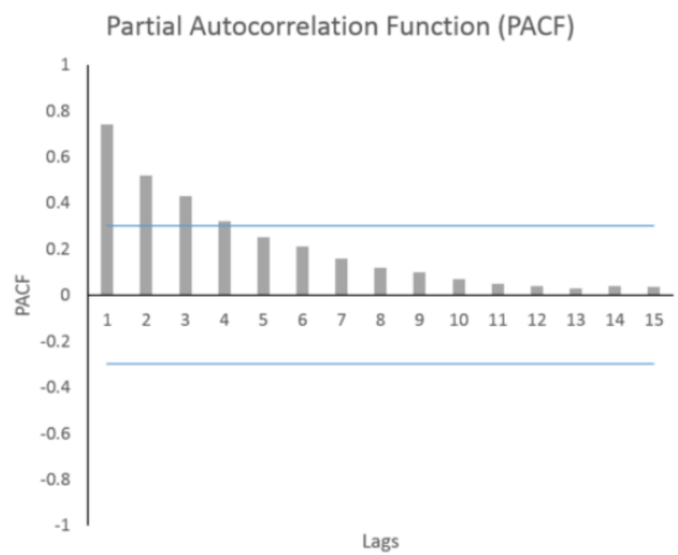
AR model





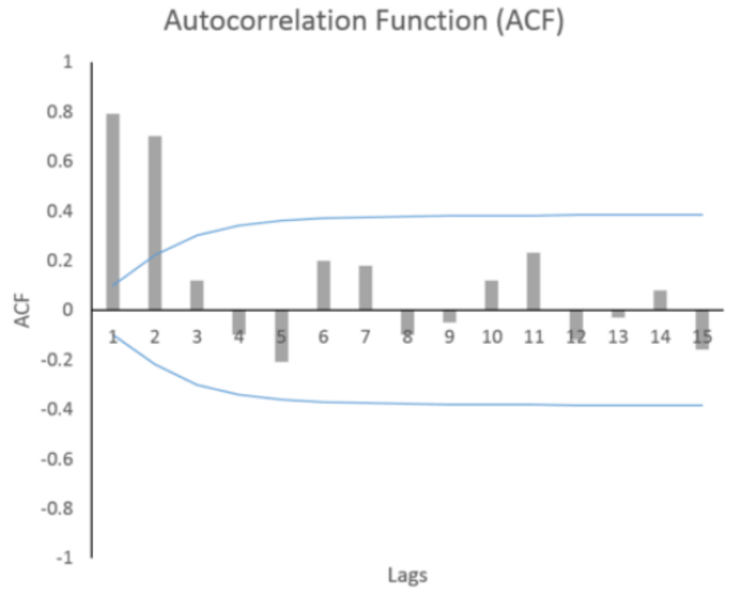
MA model





MA model određivanje *k*

- Vrednost *k* kod MA(*k*) određuje se analizom autokorelograma.
- Sa autokoreloglama uzima se poslednje *k* posle kojeg su vrednosti ACF beznačajne (nalaze se unutra označenog dela).
- Sa grafika ispod zaključjemo da je *k*=2 odgovarajuća vrednost.



MA model – Primer avio-kompanija

- Prikaz Jypter notebook primera primene MA modela na vremensku seriju sa putnicima avio-kompanije.

ARMA model

- Vremensku seriju modelujemo kao zbir AR i MA modela:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

gde je *p* broj vremenskih jedinica unazad koje koristimo za AR model, a *q* za MA model.

- Na primer, ARMA(3,2) je oblika:

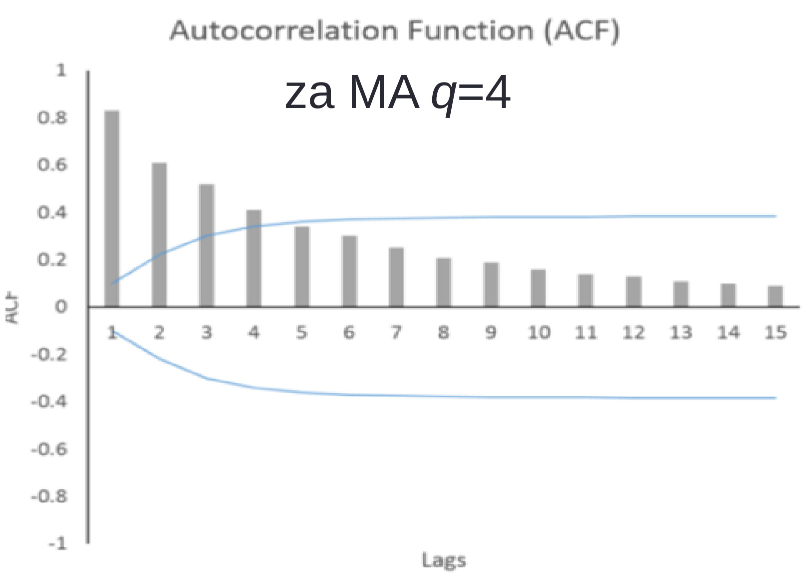
$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_2 y_{t-3} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

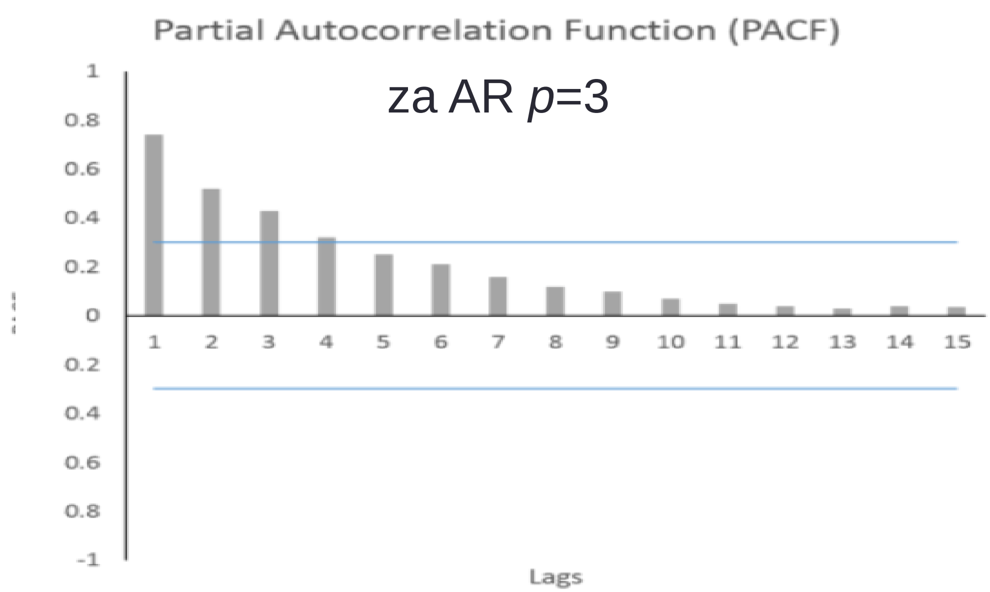
- Koristi se kada se na vremeskoj seriji identifkuju šabloni ACF i PACF koji u isto vreme odgovaraju i AR i MA modelima.

- ARMA model podrazumeva stacionarnost.

ARMA model – ACF i PACF

- Kao kombinacija AR i MA modela, za ARMA model posmatramo oba autokorelolograma.
- Šabloni koje tražimo su uskladu sa do sada pomenutim šablonima za AR i MA modele zasebno.
- Za autokorelologram očekujemo postepen pad, gde je poslednja značajna vrednost indikator *q* za MA deo.
- Za parcijalni autokorelogram takođe očekujemo postepen pad, gde je poslednja značajna vrednost intikator *p* za AR deo.





ARMA model – ACF i PACF – napomena

- ARMA modeli uključuju dva različita načina modelovanja vremenske serije (prethodne vrednosti i ostatke) pa mogu rezlultovati različitim interakcijama AR i MA terminima.

- Na primer, dodavanje MA termina može da poništi AR termin.

- Takve interakcije mogu prouzrokovati različite šablone ACF i PACF koji nisu uvek očigledni indikatori ARMA modela.

- Sa takvih šablona nije uvek lako uvideti vrednosti *p* i *q*.

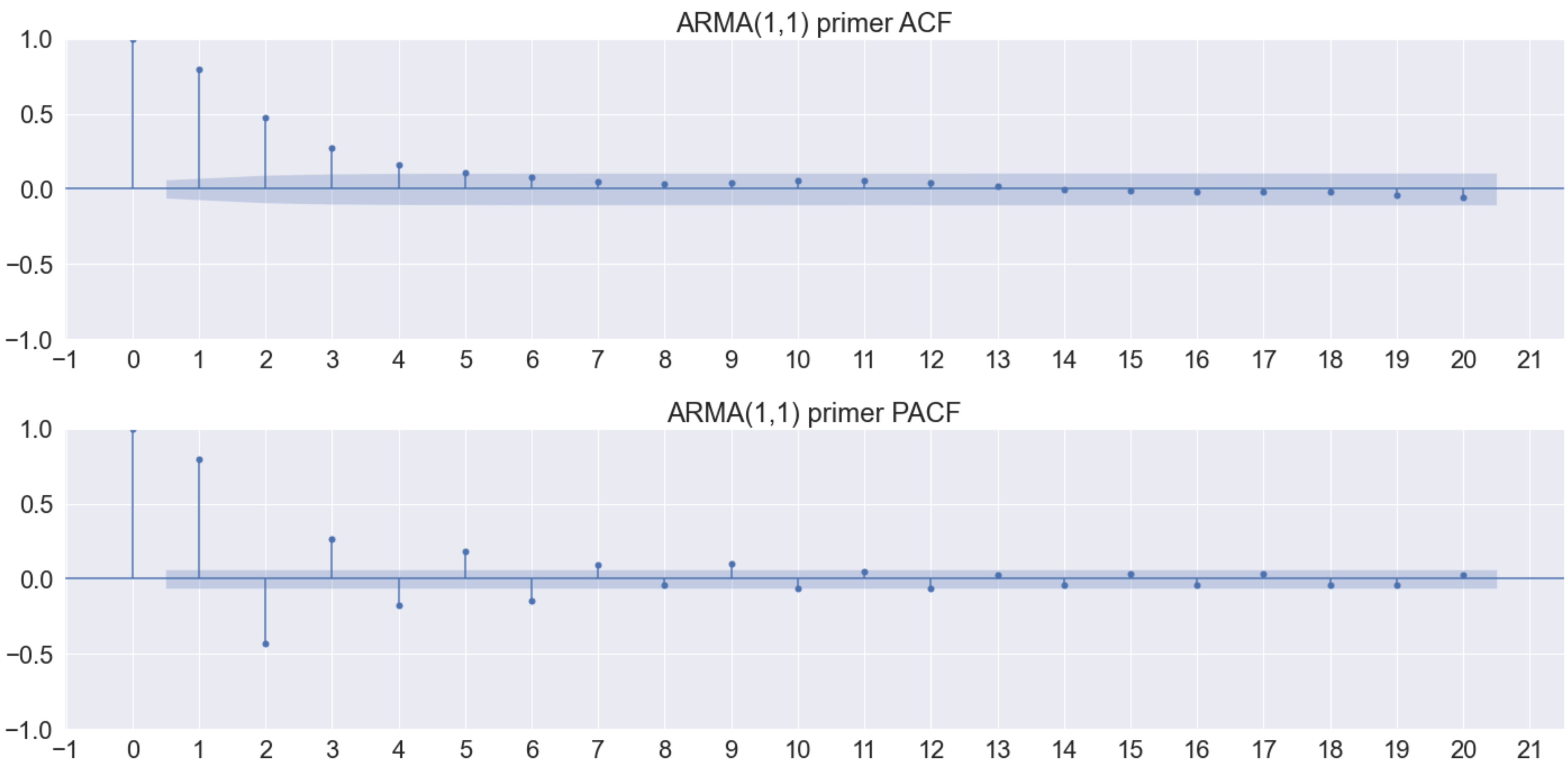
- Iz tog razloga postoje automatizovane metode za određivanje *p* i *q* kao na primer *auto_arima* iz *Python* biblioteke *pmdarima*.

- Na slajdovima u nastavku prikazaćemo nekoliko primera.

ARMA model – ACF i PACF – Primeri

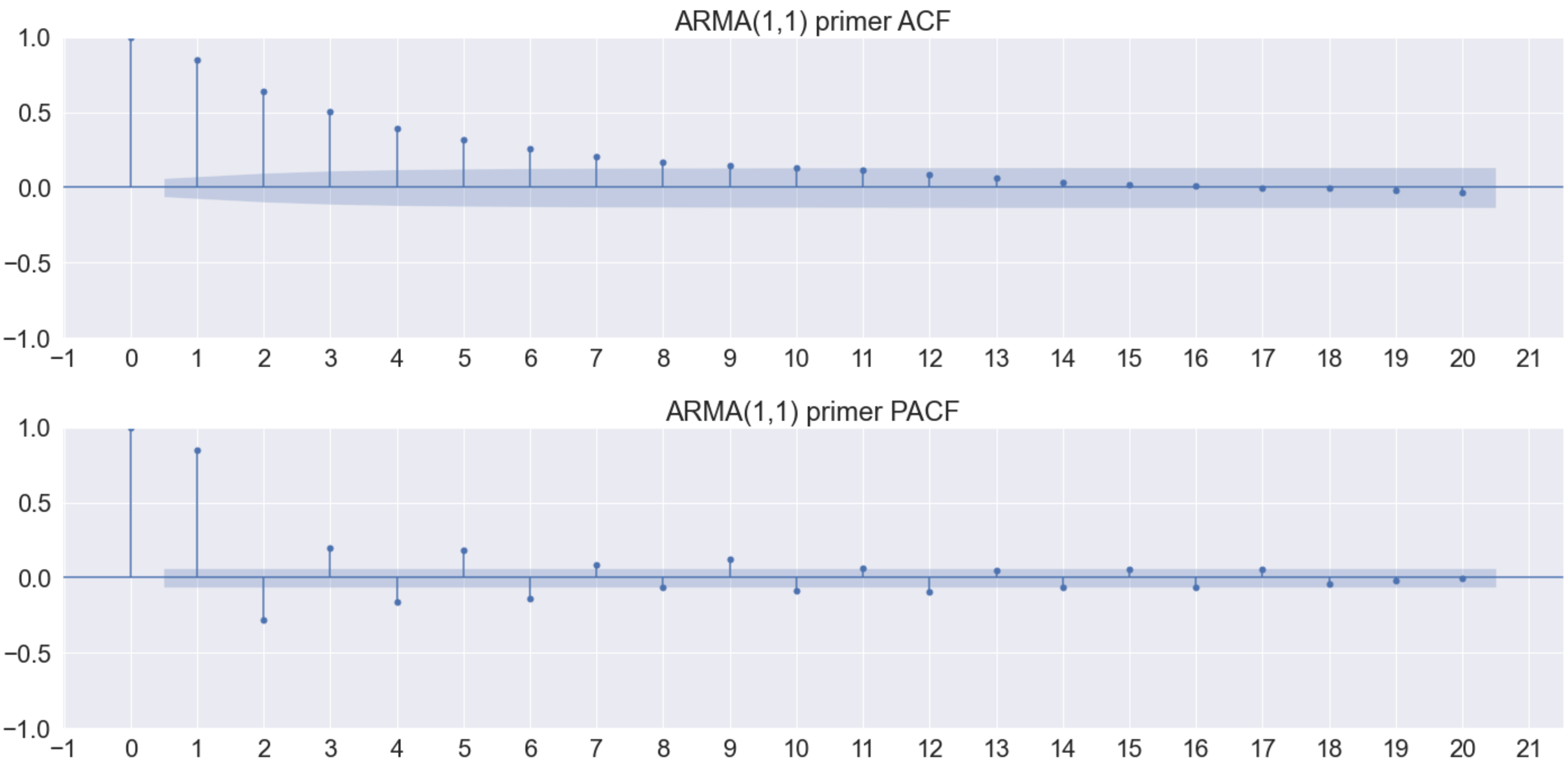
$$y_t = 0.6y_{t-1} + 0.8\varepsilon_{t-1}$$

Sa grafika nije očigledno da je $p = 1$ i $q = 1$.



ARMA model – ACF i PACF – Primeri

$$y_t = 0.4y_{t-1} + 0.3y_{t-2} + 0.9\varepsilon_{t-1}$$



ARMA model – ACF i PACF – auto_arima

- Sa desne strane prikazan je rezultat primene *auto_arima* na vremenskoj seriji sa prethodnog slajda.
- Iz rezultata se može videti da je pravilno identifikovan ARMA(2,1) model*.
- Takođe se može videti i da je obučavanje takvog modela uspešno jer su tačno identifikovane vrednosti parametara.
- *u nastavku ćemo objasniti zašto se u rezultatima model zove ARIMA i šta je vrednost 0 u između 1 i 2.

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=2880.433, Time=0.12 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=4319.232, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=3057.504, Time=0.03 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=3261.524, Time=0.09 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=2803.050, Time=0.13 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=2978.501, Time=0.05 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=2805.049, Time=0.18 sec
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=2805.049, Time=0.25 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=2819.879, Time=0.12 sec
 ARIMA(3,0,0)(0,0,0)[0]             : AIC=2940.229, Time=0.12 sec
 ARIMA(3,0,2)(0,0,0)[0]             : AIC=inf, Time=1.02 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=2804.650, Time=0.23 sec

Best model:  ARIMA(2,0,1)(0,0,0)[0]
Total fit time: 2.361 seconds
```

```
print(model.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1000
Model:               SARIMAX(2, 0, 1)   Log Likelihood               -1397.525
Date:                Mon, 06 Nov 2023   AIC                           2803.050
Time:                        13:53:55   BIC                           2822.681
Sample:                             0   HQIC                          2810.511
                               - 1000
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3686      0.035     10.461      0.000       0.300       0.438
ar.L2          0.3259      0.036      9.006      0.000       0.255       0.397
ma.L1          0.9252      0.016     58.724      0.000       0.894       0.956
sigma2         0.9555      0.042     22.652      0.000       0.873       1.038
```

ARMA model – Primer simulacije

- Prikaz Jypter notebook primera primene simulacija vremenskih serija koje prate redom: AR, MA i ARMA modele.

ARIMA

- ARIMA model je varijanta ARMA modela koja može da se primeni na seriju koja ima trend.

- Funkcioniše tako što se prvo vrši diferenciranje pa se onda na diferenciranu seriju primenjuje ARMA model.

- Na primer, za model ARIMA(3,1,2) prvo se vrši diferenciranje reda 1 za vremensku seriju pa se onda primenjuje sledeći ARMA model:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_2 y_{t-3} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

Evaluacija modela – Metrike

- Postoji veliki broj metrika koje mere razliku predikcija u odnosu na tačne vrednosti.

- Metrike koristimo tokom procesa modelovanja da poboljšamo model i/ili da ga uporedimo sa drugim modelima.

- Nakon procesa modelovanja da:
  - Predstavimo kvalitet rezultata
  - Uporedimo rezultate sa srodnim radovima koji koriste isti model ili isti skup podataka.

- U nastavku navodimo metrike koje se najčešće koriste.

Prosečna Kvadratna Greška (*Mean Squared Error*, MSE)

- Računa se po sledećoj formuli:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Prednosti:
- Pogodna za procese optimizacije koji se oslanjaju na gradijente.
- Osetljiva na ekstremne vrednosti (nagle skokove ili padove) – može biti prednost ili mana u zavisnosti od našeg cilja.
- Mane:
- Nije jednostavna za tumačenje jer su vrednosti u kvadratima jedinicama vremenske serije.

Koren Prosečne Kvadratne Greške (*Root Mean Squared Error*, RMSE)

- Računa se po sledećoj formuli:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- Varijanta MSE kod koje se vrednosti mogu tumačiti jer su u istim jedinicama kao i vremenska serija.

- Tipično se MSE koristi za optimizaciju tokom obučavanja modela, a RMSE za tumačenje i sapoštavanje rezultata.

Prosečna Apsolutna Greška (*Mean Absolute Error*, MAE)

- Računa se po sledećoj formuli:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- <u>Prednosti:</u>
- Jednostavna za računanje.
- Jednostavna za tumačenje jer su vrednosti u istim jedinicama kao vremenska serija.
- <u>Mane:</u>
- Ne pruža informacije o pozitivnim i negativnim greškama.
- Nije osteljiva na ekstremne vrednosti (nagle skokove ili padove) – može biti prednost ili mana u zavisnosti od našeg cilja.

Prosečna Apsolutna Procentualna Greška (*Mean Absolute Percentage Error*, MAPE)

- Računa se po sledećoj formuli:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100$$

- <u>Prednosti:</u>
- Pogodna za poređenje modela na različitim vremenskim serijama jer su vrednosti u procentima.
- <u>Mane:</u>
- Problematična za vremenske serije koje imaju vrednosti jako blizu nule, zbog deljenja sa y_i.
- Asimetričnost – MAPE(y, $\hat{y}$)≠ MAPE($\hat{y}$, y). Na primer MAPE(100, $\widehat{150}$)=50%, MAPE(150, $\widehat{100}$)=33.33%. Dakle, greške u kojima predikcija precenjuje tačnu vrednost su veće nego ako potcenjuje.

Prosečna Apsolutna Skalirana Greška (*Mean Absolute Scaled Error*, MAE)

- Računa se po sledećoj formuli:

$$MASE = \frac{MAE}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|} = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}$$

- Mera koja poredi naš model sa modelom koji koristi samo prethodnu vrednost vremenske serije kao predikciju za trenutnu vrednost.

- <u>Prednosti:</u>

- Invarjantnost na skalu (scale invariance) – vrednosti ne zavise od raspona vrednosti vremenske serije, što omogućava poređenje modela na različitim skupovima podataka.

- Jednostavna za tumačenje jer vrednosti manje od 1 znače da je naš model bolji od onog koji se oslanja samo na prehodnu vrednost.

- Simetričnost - MASE(y, $\hat{y}$)=MASE($\hat{y}$, y)

- <u>Mane:</u>

- Problematična za vrednosti y_i koje su jako blizu nuli, zbog potencijalnog detljenja nulom ili jako malim brojem.

Završna napomena

- Prikazani su osnovni koncepti Analize vremenskih serija.

- Analiza vremenskih serija je široka oblast, pa samim tim određeni broj koncepata i modela nisu spomenuti na ovom kursu.

- Cilj ovog dela kursa je samo uvod u ovu oblast koja je veoma značajna, a često nije deo kurseva koji se tiču prediktivnih modelovanja.