

Seminarska naloga pri predmetu Modeliranje časovnih in prostorskih procesov

Neža Kržan, Tom Rupnik Medjedovič

Kazalo

Uvod	5
1 Analiza časovnih vrst	5
1.1 Predstavitev podatkov	5
1.2 Časovna vrsta <i>Gostinske nastanitvene dejavnosti</i>	6
1.2.1 Pregled časovne vrste	6
1.2.2 Transformacija	7
1.2.3 Avtokorelogram in parcialni avtokorelogram	8
1.2.4 Diferenciranje	8
1.2.5 Izbira ustreznega modela	11
1.2.6 Napoved	16
1.3 Časovna vrsta <i>Dejavnost strežbe jedi in pijač</i>	19
1.3.1 Pregled časovne vrste	19
1.3.2 Transformacija	20
1.3.3 Avtokorelogram in parcialni avtokorelogram	21
1.3.4 Diferenciranje	21
1.3.5 Izbira ustreznega modela	24
1.3.6 Napoved	29
1.4 Časovna vrsta <i>Gostinstvo prosta delovna mesta 1+</i>	32
1.4.1 Pregled časovne vrste	32
1.4.2 Transformacija časovne vrste	33
1.4.3 Avtokorelogram in parcialni avtokorelogram	34
1.4.4 Diferenciranje	35
1.4.5 Izbira ustreznega modela	36
1.4.6 Napoved	37
1.5 Časovna vrsta <i>Gostinstvo zasedena delovna mesta 1+</i>	41
1.5.1 Pregled časovne vrste	41
1.5.2 Transformacija časovne vrste	42
1.5.3 Analiza avtokoreliranosti	43
1.5.4 Diferenciranje	43
1.5.5 Izbira ustreznega modela	44
1.5.6 Napoved	45
2 Analiza prostorskega procesa	49
2.1 Predstavitev podatkov	49
Vegetacijski indeks NDRE	49
2.1.1 Grafični prikaz parametrov	50
2.2 Grafični prikaz	51
2.3 Analiza prostorske korelacije	52
2.3.1 Razsevni grafikon	52
2.3.2 Oblak semivariograma in vzorčni semivariogram	53
2.3.3 Ovojnice za vzorčni semivariogram	55
2.4 Model semivariograma za dane podatke	56
2.4.1 Sferični model variograma brez zlatega zrna	56
2.4.2 Gaussov model variograma brez zlatega zrna	59

Slike

1	Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na leta.	6
2	Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na mesec.	6
3	Časovna vrsta indeksa prihodka za 'Gostinske nastanitvene dejavnosti'.	7
4	Osnovna in transformirana časovna vrsta 'Gostinske nastanitvene dejavnosti'.	7

5	Avtokorelogram in parcialni avtokorelogram časovne vrste.	8
6	Osnovna in diferencirana časovna vrsta.	8
7	Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.	9
8	Osnovna in diferencirana(sezonskost) časovna vrsta.	9
9	Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.	10
10	Diferenciranje trenda in sezonskosti v časovni vrsti.	10
11	Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.	11
12	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.	12
13	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	13
14	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	14
15	Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.	15
16	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	16
17	Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.	17
18	Časovna vrsta in napovedi na podlagi modela od sredine leta 2023 naprej.	18
19	Porazdelitev indeksa prihodka v dejavnosti strežbe jedi in pijač glede na leta.	19
20	Porazdelitev indeksa prihodka v dejavnosti strežbe jedi in pijač glede na mesec.	19
21	Časovna vrsta indeksa prihodka za 'Dejavnost strežbe jedi in pijač'.	20
22	Osnovna in transformirana časovna vrsta 'Dejavnost strežbe jedi in pijač'.	20
23	Avtokorelogram in parcialni avtokorelogram časovne vrste.	21
24	Osnovna in diferencirana časovna vrsta.	22
25	Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.	22
26	Osnovna in diferencirana(sezonskost) časovna vrsta.	23
27	Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.	23
28	Diferenciranje trenda in sezonskosti v časovni vrsti.	24
29	Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.	24
30	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.	25
31	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	26
32	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	27
33	Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.	28
34	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	29
35	Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.	30
36	Časovna vrsta in napovedi na podlagi modela od sredine leta 2023 naprej.	31
37	Porazdelitev prostih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na leto.	32
38	Porazdelitev prostih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na kvartal.	32
39	Časovna vrsta 'Gostinstvo prosta delovna mesta 1+'.	33
40	Osnovna in logaritmira časovna vrsta 'Gostinstvo prosta delovna mesta 1+'.	33
41	Avtokorelogram in parcialni avtokorelogram logaritmira časovne vrste.	34
42	Avtokorelogram in parcialni avtokorelogram osnovne časovne vrste.	35
43	Logaritmira osnovna in logaritmira diferencirana časovna vrsta.	35
44	Avtokorelogram in parcialni avtokorelogram logaritmira diferencirane časovne vrste.	36
45	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa.	37
46	Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.	38
47	Časovna vrsta in napovedi na podlagi modela od leta 2021 naprej.	39
48	Časovna vrsta in napovedi na podlagi modela od leta 2023 naprej.	40
49	Porazdelitev zasedenih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na leto.	41
50	Porazdelitev zasedenih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na kvartal.	41

51	Časovna vrsta 'Gostinstvo zasedena delovna mesta 1+'.	42
52	Osnovna in logaritmirana časovna vrsta 'Gostinstvo zasedena delovna mesta 1+'.	42
53	Avtokorelogram in parcialni avtokorelogram logaritmirane časovne vrste.	43
54	Oosnovna in diferencirana časovna vrsta.	44
55	Avtokorelogram in parcialni avtokorelogram logaritmirane diferencirane časovne vrste.	44
56	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa.	45
57	Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.	46
58	Časovna vrsta in napovedi na podlagi modela od leta 2021 naprej.	47
59	Časovna vrsta in napovedi na podlagi modela od leta 2023 naprej.	48
60	Porazdelitev parametrov glede na vrsto.	50
61	Porazdelitev parametrov glede na lokacijo.	50
62	Porazdelitev spremenljivke NDRE za mesec julij glede na lokacijo in vrsto.	51
63	Grafični prikaz lokacij in relativne vrednosti spremenljivke 'NDRE'.	51
64	Predstavitev realizacije prostorskega procesa 'NDRE'.	52
65	Prostorska korelacija za 'NDRE(jul)'.	53
66	Oblak semivariograma za 'NDRE(jul)'(levo), okvirji z ročaji za vrednosti klasičnega variaograma po razdaljah med lokacijami(sredina) in za vrednosti robustnega vzorčnega variograma(desno).	53
67	Vzorčni semivariogram za 'NDRE(jul)'.	54
68	Vzorčni variogrami za 'NDRE(jul)' za štiri glavne smeri.	55
69	Monte Carlo ovojnica semivariograma.	55
70	Profile-log-likelihood za parametra sferičnega modela semivariograma brez zlatega zrna kaže na napačno izbiro modela prostorske korelacije.	56
71	Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.	56
72	Profile-log-likelihood za parametre sferičnega modela semivariograma z zlatim zrnem.	57
73	Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.	57
74	Napovedi NDRE za mesec julij za dano območje - model semivariograma fit1.ml(levo).	58
75	Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma fit1.ml.	58
76	Napovedi NDRE za mesec julij za dano območje - model semivariograma fit2.ml.	59
77	Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma fit2.ml.	59
78	Profile-log-likelihood za parametra Gaussovega modela semivariograma brez zlatega zrna.	60
79	Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.	60
80	Profile-log-likelihood za parametre Gaussovega modela semivariograma z zlatim zrnem.	61
81	Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.	61
82	Napovedi NDRE za mesec julij za dano območje - model semivariograma fit1g.ml.	62
83	Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma fit1g.ml.	62
84	Napovedi NDRE za mesec julij za dano območje - model semivariograma fit2g.ml.	63
85	Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma fit2g.ml.	63

Tabele

1	SARIMA modeli in pripadajoče AIC vrednosti.	15
2	SARIMA modeli in pripadajoče AIC vrednosti.	28

Uvod

Seminarska naloga je sestavljena iz dveh delov.

V prvem delu sva analizirala štiri izbrane časovne vrste, ki spadajo pod temo *Gostinstvo*. Obravnavala sva indeks prihodka za *Gostinske nastanitvene dejavnosti* in *Dejavnost strežbe jedi in pijač* ter prosta in zasedena delovna mesta v gostinstvu. Za vse štiri časovne vrste sva naredila podrobnejšo analizo in skušala najti model, ki najbolje napoveduje indeks prihodka v obeh dejavnostih ter prosta in zasedena delovna mesta v gostinstvu.

V drugem delu sledi analiza prostorskega procesa. Podrobneje sva obravnavala vegetacijski indeks *NDRE* v mesecu juliju in izmerjene parametre, ki lahko, poleg lokacije, tudi vplivajo na vegetacijski indeks.

1 Analiza časovnih vrst

1.1 Predstavitev podatkov

Za analizo sva si izbrala podatke, ki spadajo pod temo *Gostinstvo*.

Za indeks prihodka po dejavnosti sva si izbrala *Gostinske nastanitvene dejavnosti* in *Dejavnost strežbe jedi in pijač*. Podatki so podani za časovno obdobje od januarja 2010 do januarja 2024(2010M01 do 2024M01). Frekvenca vzorčenja je približno enakomerna in enaka 12, torej gre za mesečno vzorčenje.

Nato sva analiziral še časovni vrsti za prosta in zasedena delovna mesta v gostinstvu, kjer je zaposlena vsaj 1 oseba. Podatki so podani za časovno obdobje od leta 2008 do leta 2023(2008Q1 do 2023Q4) po kvartalnih. Frekvenca vzorčenja je približno enakomerna in enaka 4, torej gre za četrtletno vzorčenje.

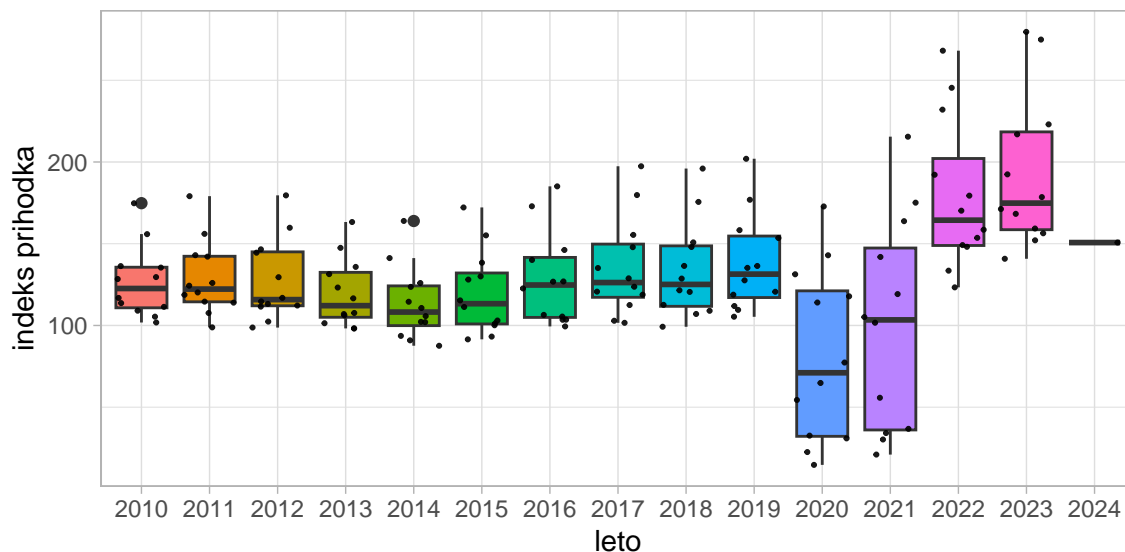
Ker sta frekvenci vzorčenja enakomerni imamo opravka z ekvidistantnimi časovnimi vrstami.

1.2 Časovna vrsta *Gostinske nastanitvene dejavnosti*

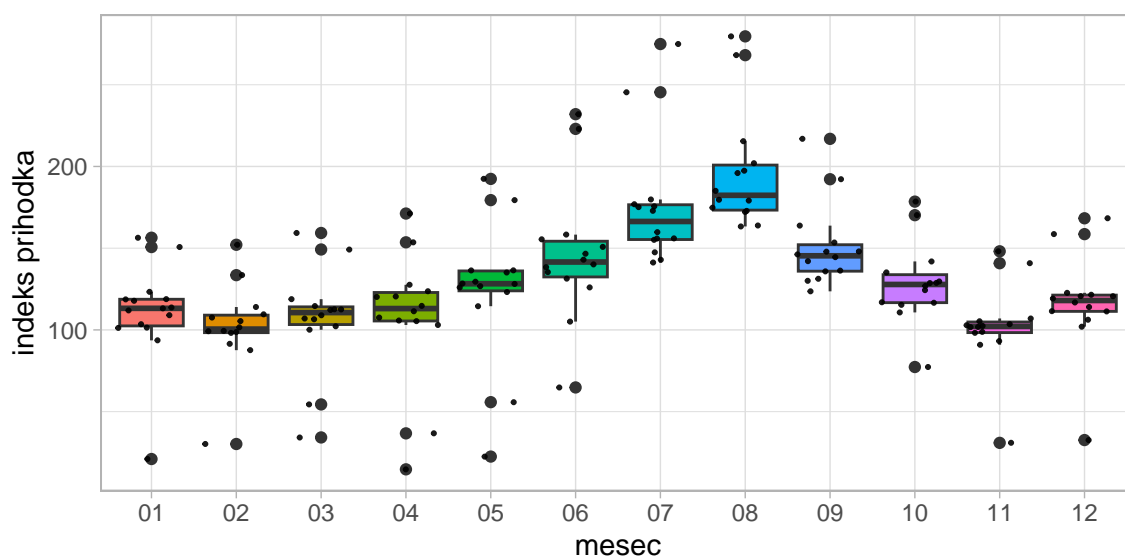
1.2.1 Pregled časovne vrste

Podatke imamo za leta od 2010 do leta 2023 za vsak mesec v letu in za leto 2024 za mesec januar.

Na spodnjem grafu porazdelitev indeksa prihodka glede na leta lahko vidimo, da bo proti koncu variabilnost časovne vrste zelo nekonstantna v primerjavi z ostalimi leti. Prav tako so vrednost indeksa prihodka v zadnjih letih nekoliko višji. Pri časovni vrsti na začetku pričakujemo minimalni trend, ki je sprva negativen, nato nekoliko pozitiven, potem pa sledijo velike spremembe v in po letu 2020. Glede na mesece vidimo, da je prihodkov v dejavnosti največ ravno v poletnih mesecih in pa večje je tudi v mesecu januarju in decembru, kar je dokaj logično glede na življenje ljudi.

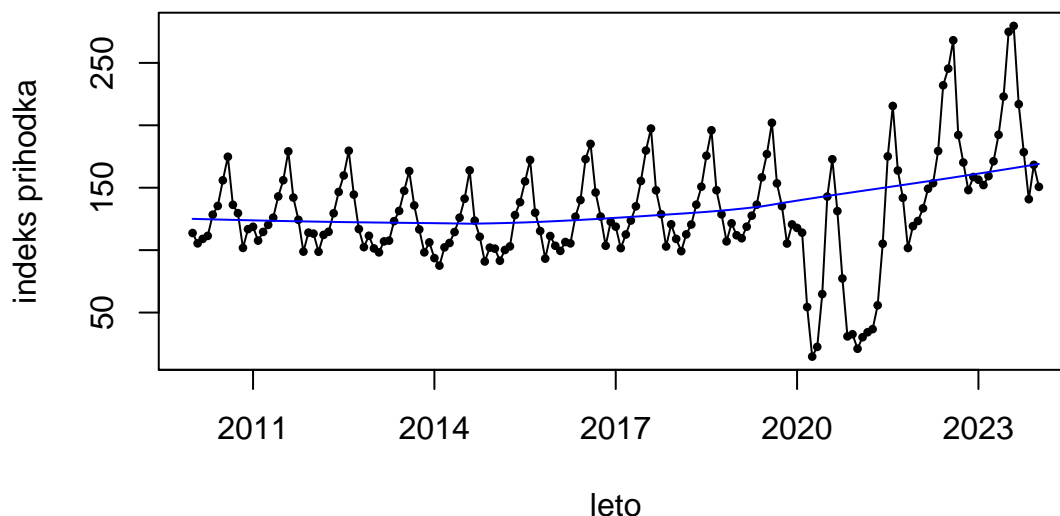


Slika 1: Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na leta.



Slika 2: Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na mesec.

Narišimo si sedaj časovno vrsto, ki jo bomo v nadaljevanju podrobneje analizirali.



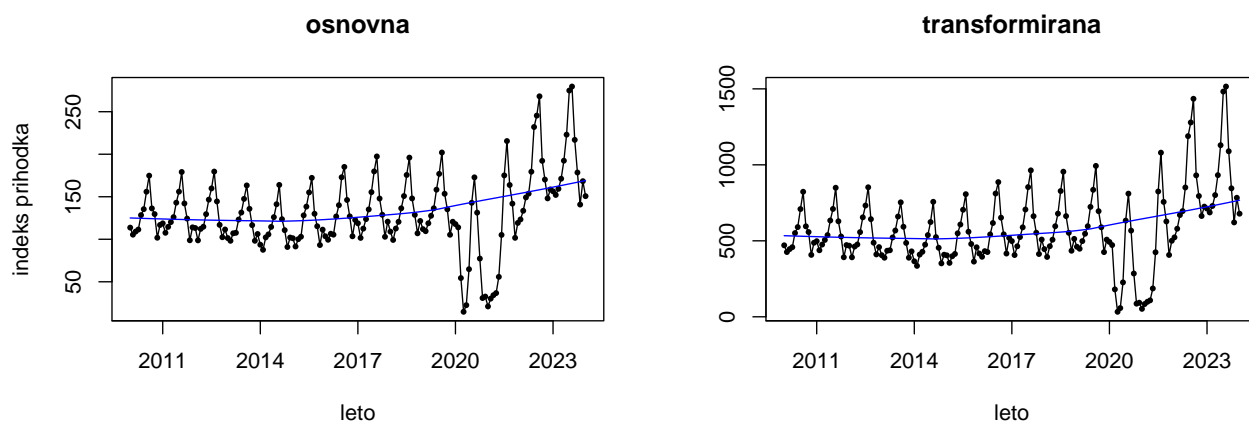
Slika 3: Časovna vrsta indeksa prihodka za 'Gostinske nastanitvene dejavnosti'.

Ponovno opazimo prisotnost trenda, ki je sprva negativen, nato pa se leta 2015 obrne in postane pozitiven, naklon pa je bil v zadnjih letih največji. Dobro je vidna tudi prisotnost sezonskosti. Indeks se v vsakem letu poveča v začetku leta (januarja), nato nekoliko pade in ponovno raste do poletja (avgusta). Potem sledi padanje do konca let, ko je ponovno nekoliko povišanje decembra. Tako gibanje je bilo tudi pričakovano, saj je to obdobje poletne in zimske sezone.

Prav tako je dobro vidno ekstremno gibanje v času pandemije Covid-19, saj se je takrat zgodil velik padec vrednosti indeksa. Tudi amplituda nihanja se je drastično spremenila v primerjavi s predhodnjim opazovanim obdobjem, za kar je verjetno ponovno krivo obnašanje in potrebe ljudi po pandemiji in izdaja državnih bonov.

1.2.2 Transformacija

Ker imamo prisotno nekonstantno varianco, s pomočjo BoxCox testa, ugotovimo, da bi bila primerna transformacija z $\lambda = 1.3$.



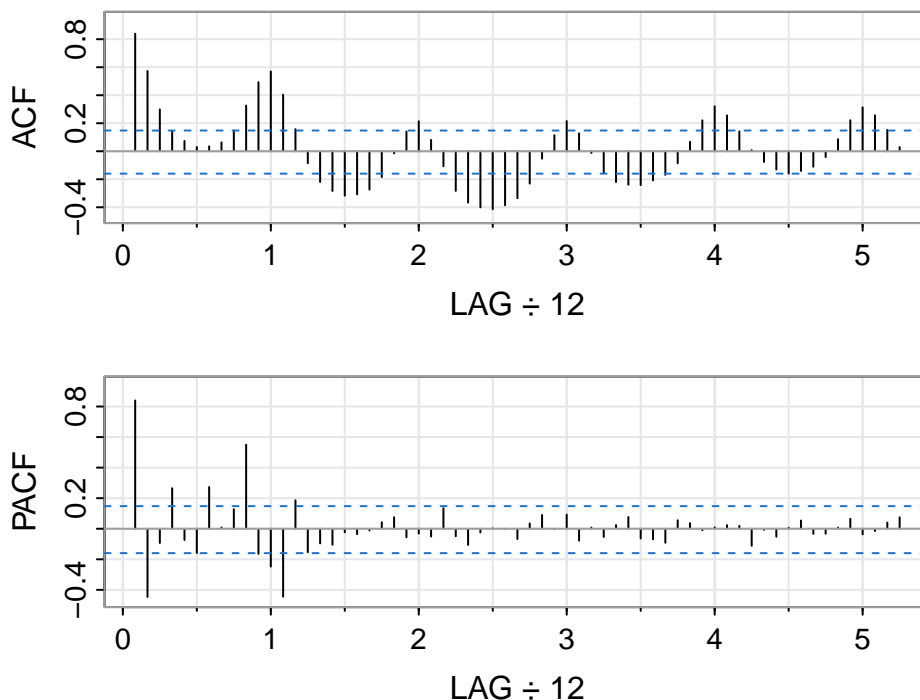
Slika 4: Osnovna in transformirana časovna vrsta 'Gostinske nastanitvene dejavnosti'.

Ker variabilnost osnovne in transformirane časovne vrste izgleda dokaj podobna (razlikujejo se razponi na y-osi), se zaenkrat ne odločiva za transformacijo in analizo nadaljujemo z osnovno časovno vrsto.

1.2.3 Avtokorelogram in parcialni avtokorelogram

Na prvem grafu je vidno, da koeficienti avtokorelacije padajo dokaj počasi, kar nakazuje, da časovna vrsta vsebuje trend, ki ga bi bilo treba modelirati z diferenciranjem. Na avtokorelogramu koeficienti avtokorelacije nakazujejo na nihanje. Sezona je vidna z nihanjem vrednosti koeficientov oz. periodičnostjo in izrazitimi vrhovi, ki se ponavljajo. Trenutno v modelu je statistično značilna korelacija med členi časovne vrste.

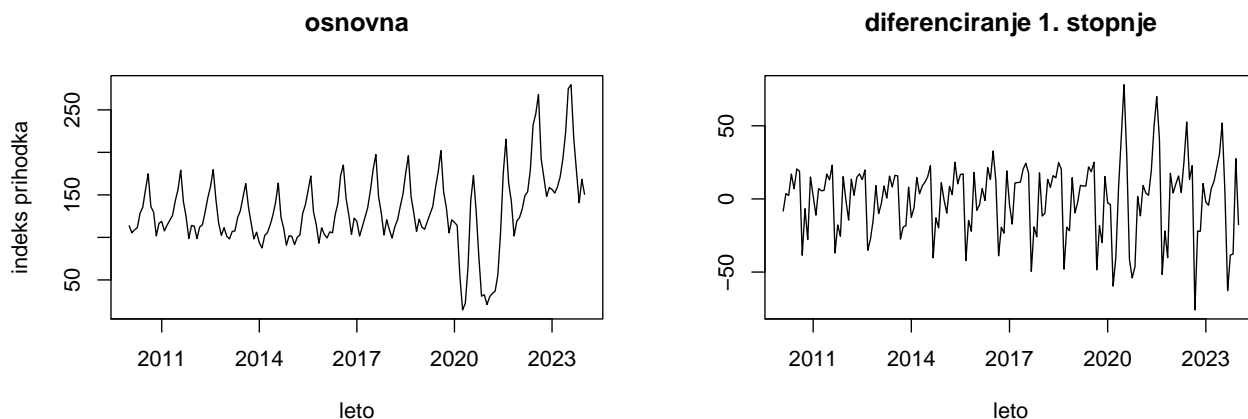
Tudi na parcialnem avtokorelogramu imamo nekaj koeficientov parcialne avtokorelacije, ki so statistično značilni, zato bo potrebno posvetiti nekaj pozornosti tudi temu.



Slika 5: Avtokorelogram in parcialni avtokorelogram časovne vrste.

1.2.4 Diferenciranje

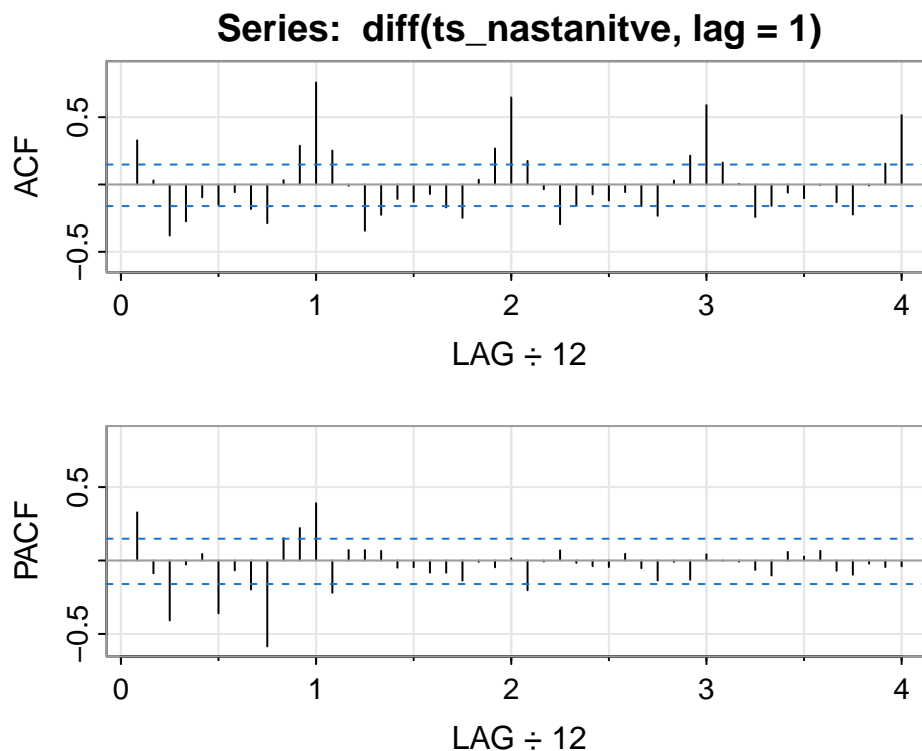
Torej na avtokorelogramu se vidi, da koeficienti avtokorelacije nakazujejo na trend v časovni vrsti, zato jo enkrat diferenciramo in preverimo ali imamo stacionarno časovno vrsto.



Slika 6: Osnovna in diferencirana časovna vrsta.

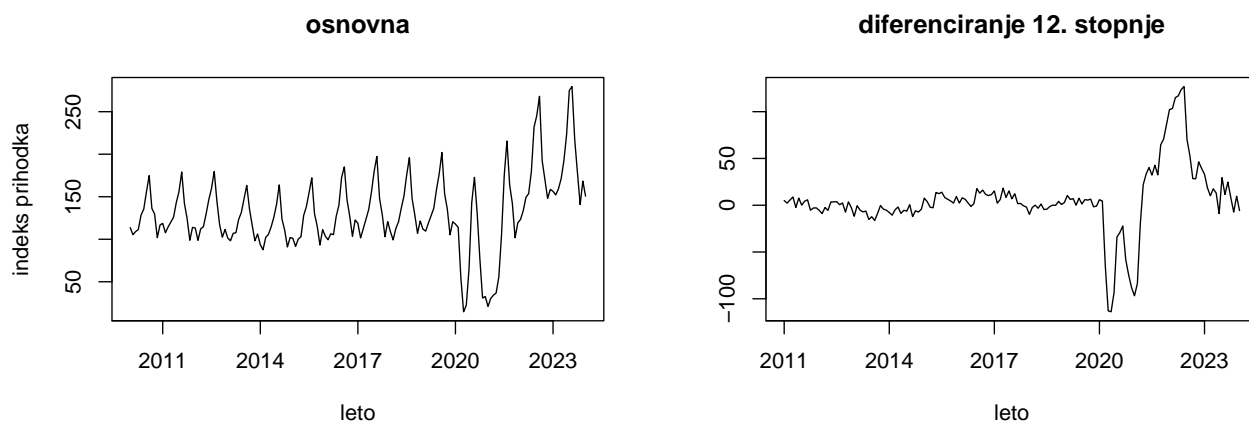
Z diferenciranjem linearnega trenda smo na levi strani drugega grafa ustvarili dokaj končno variabilnost in pričakovano vrednost enako 0, v letu 2020 temu ni tako, potem pa se ponovno “uredi” časovna vrsta.

Na spodnjem avtokorelogramu opazimo, da se pojavljajo statistično značilni koeficienti avtokorelacije na odlogih 1, 2, 3, 4; torej je prisotna izrazita sezonskost.



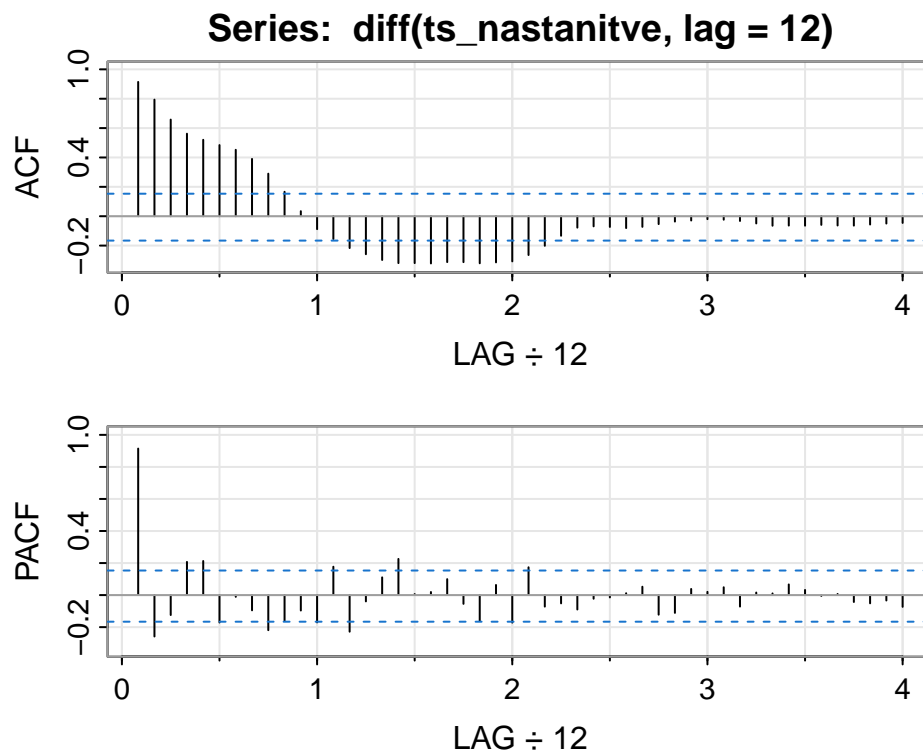
Slika 7: Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.

Zaradi zgornje ugotovitve o sezonskosti poskušajmo diferencirati še to. Na grafu vidimo, da je varianca tokrat bolj nekonstantna na desni polovici desnega grafa, kot pri diferenciranju 1. stopnje. Še vedno se vidi velik vpliv epidemije Covid-19.



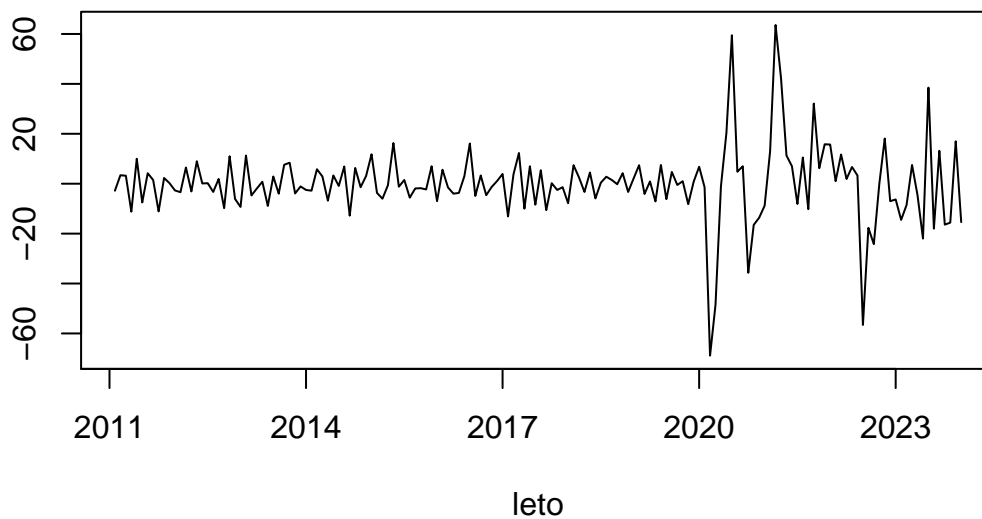
Slika 8: Osnovna in diferencirana(sezonskost) časovna vrsta.

Če diferenciramo samo sezonskost nam avtokorelogram še vedno nakazuje prisotnost trenda, zato preverimo še, kaj se zgodi če diferenciramo trend in sezonskost.



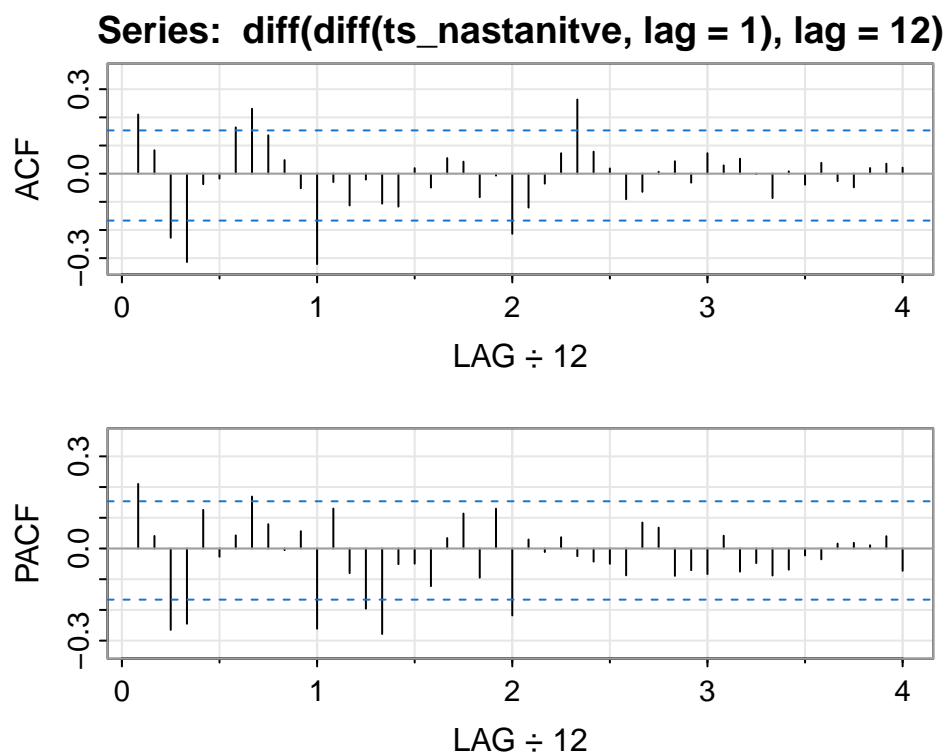
Slika 9: Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.

Na spodnjem grafu vidimo dokaj konstantno varianco in pričakovano vrednost 0 do leta 2020, nato se to, zaradi epidemije Covid-19, nekoliko spremeni, kar bomo težko še bolj modelirali.



Slika 10: Diferenciranje trenda in sezonskosti v časovni vrsti.

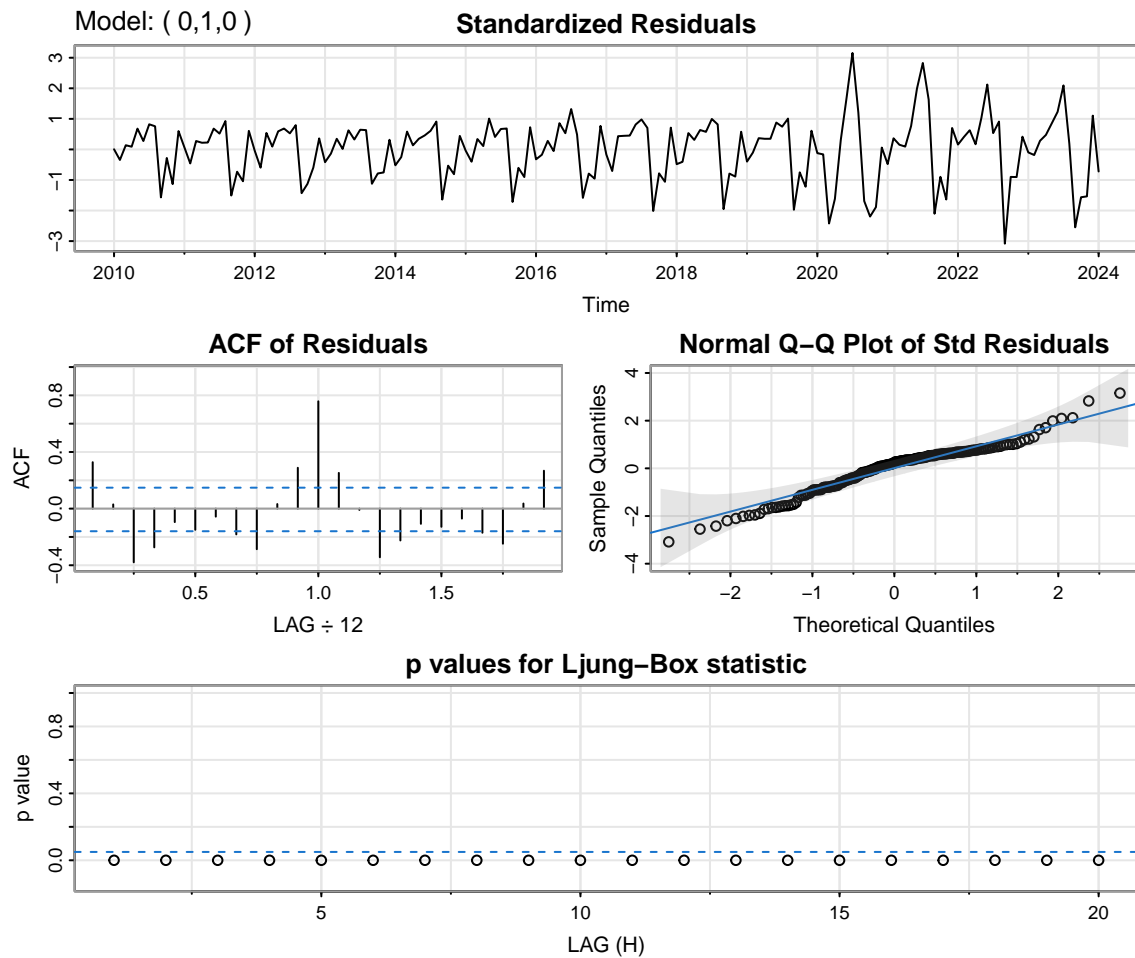
Vidimo, da imamo nekaj statistično značilnih koeficientov avtokorelacije in parcialne avtokorelacije, torej verjetno še vedno obstaja neka majhna avtokorelacija med podatki časovne vrste, ampak poskusimo sedaj narediti model.



Slika 11: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.

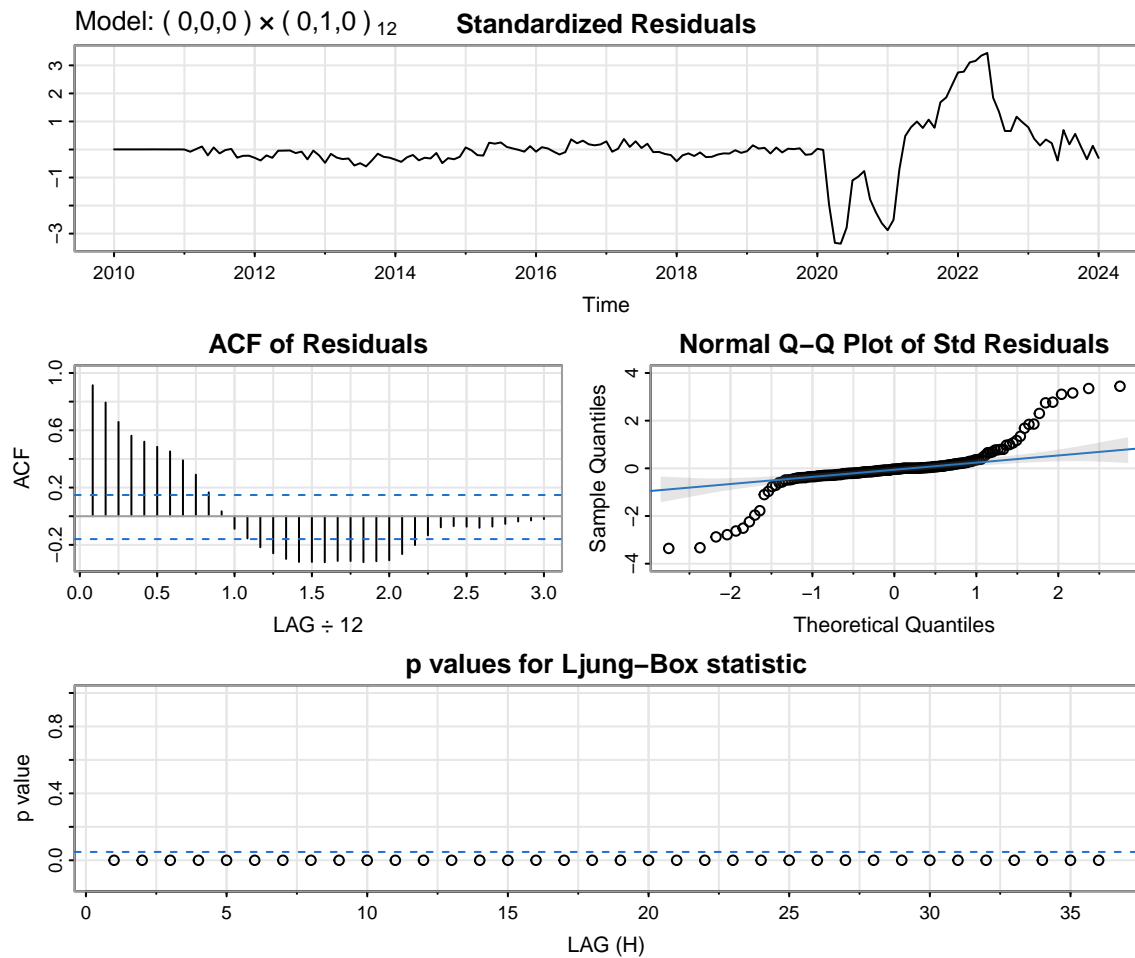
1.2.5 Izbira ustreznega modela

Izbire modela se lotiva postopoma, zato najprej pogledava kakšne rezultate dobiva, če časovni vrsti diferencirava le trend($d = 1$).



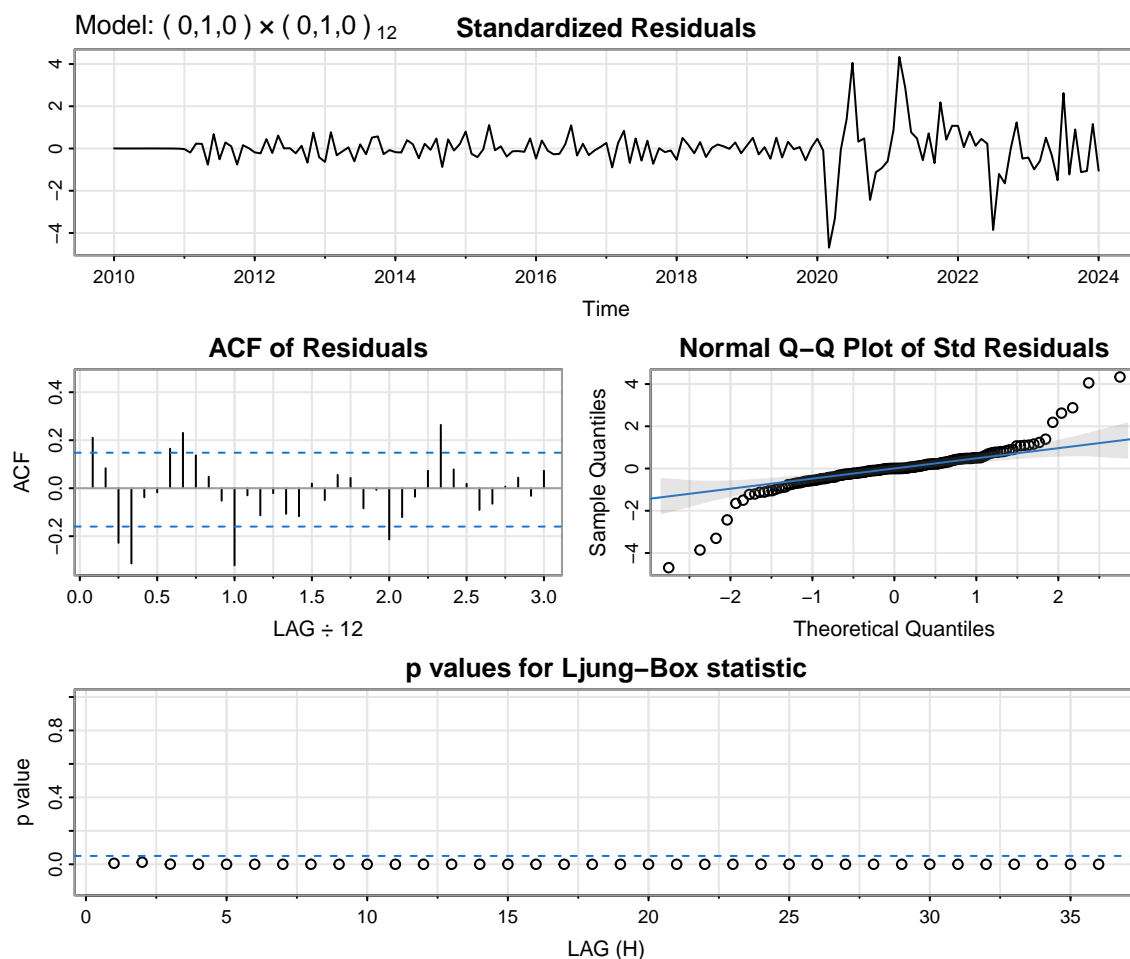
Slika 12: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.

Takoj nam je jasno, da je v časovni vrsti še vedno prisotna avtokorelacija na podlagi zgornje analize, sedaj nam Ljung-Box test to le še potrdi, saj vidimo na spodnjem grafu, da so vse p-vrednosti manjše od 0.05, torej zavračamo ničelno domnevo, da ni prisotne avtokorelacije med členi časovne vrste. Zato v model vključimo diferenciranje sezone.



Slika 13: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

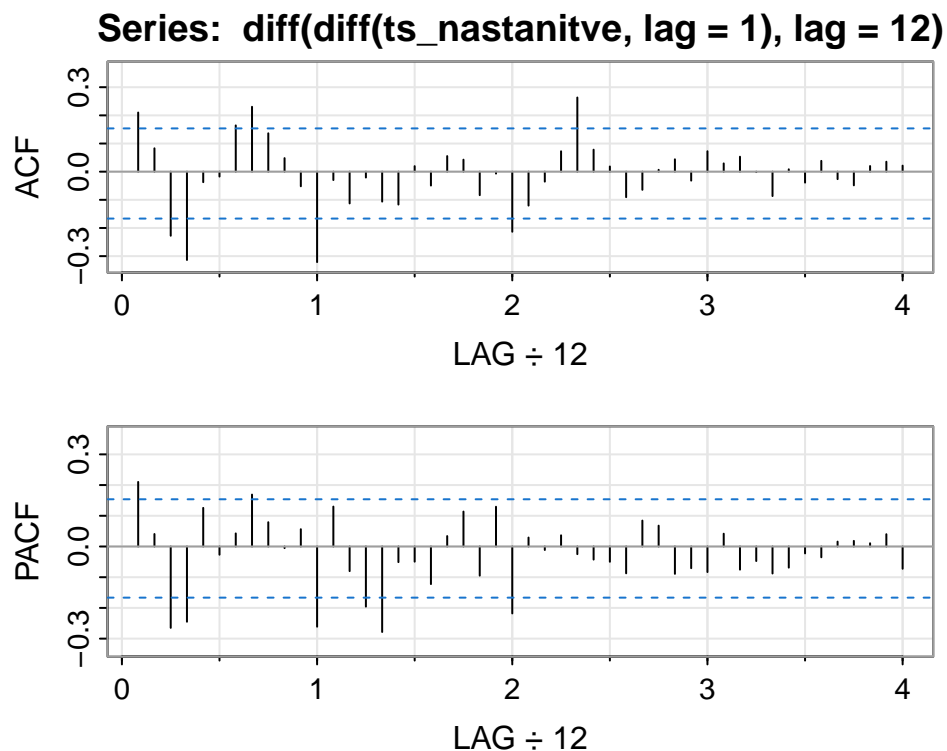
Tudi tukaj nam Ljung-Box test potrdi, da je v časovni vrsti prisoten trend, saj vidimo na spodnjem grafu, da imamo p-vrednosti manjše od 0.05, torej zavračamo ničelno domnevo, da ni prisotne avtokorelacije med členi časovne vrste. Zato v model vključimo diferenciranje trenda in sezonskosti.



Slika 14: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

Imamo prisotnih nekaj statistično značilnih koeficientov avtokorelacije, tudi pr Ljung-Box testu imamo še vedno p-vrednosti manjše od 0.05, torej so členi časovne vrste še vedno avtokorelirani med seboj.

Poglejmo si ponovno avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste z odlogom 1 in 12. Iz ACF razberemo, da bi lahko poskusili parameter q nastaviti za začetek na 4 in iz PACF, da lahko poskusimo parameter p nastaviti na 4 - izberemo na podlagi statistično značilnih koeficientov. Kasneje raziskujem različne vrednosti parametrov p , q , P in Q .

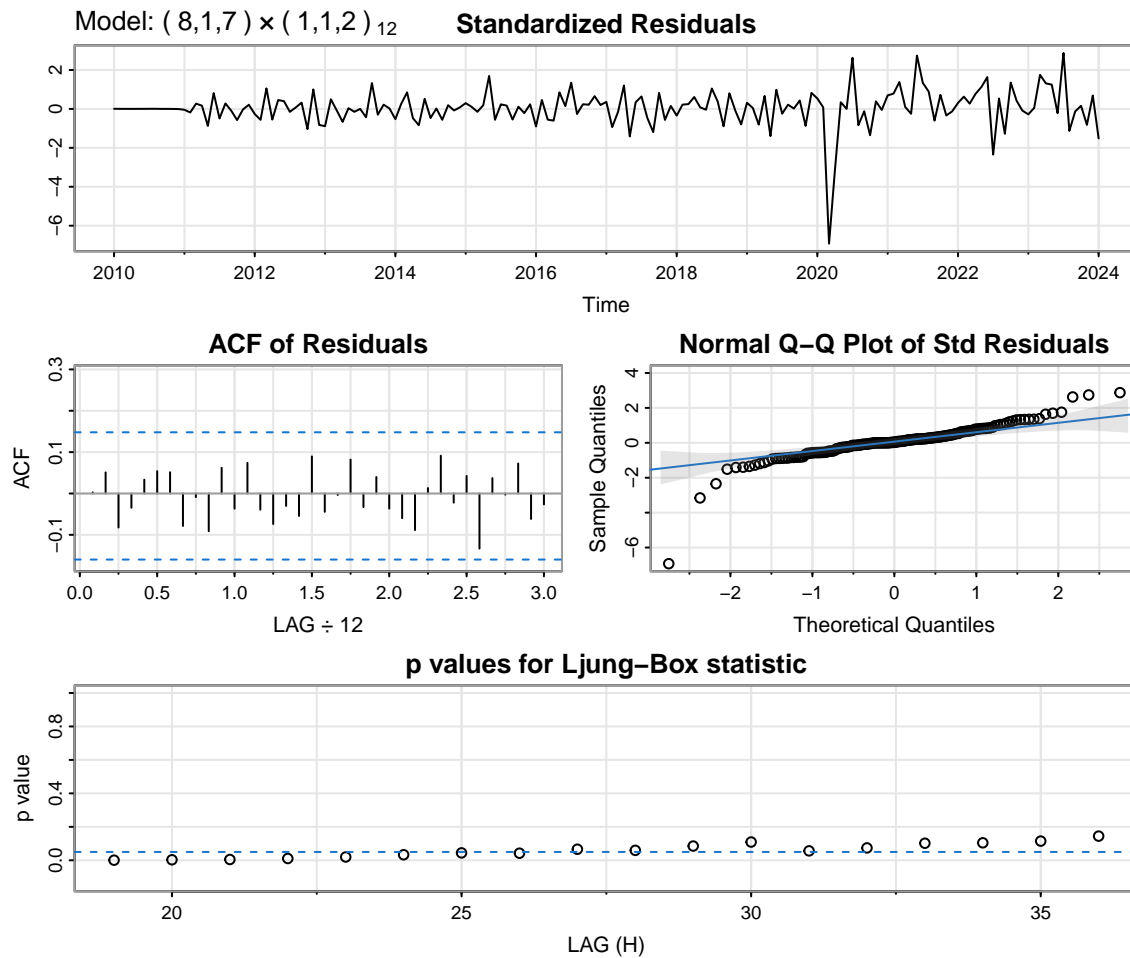


Slika 15: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.

Tabela 1: SARIMA modeli in pripadajoče AIC vrednosti.

indeks	model	AIC
1	SARIMA(4, 1, 4, 0, 1, 0)	1252.806
2	SARIMA(4, 1, 4, 2, 1, 2)	1216.153
3	SARIMA(3, 1, 3, 2, 1, 2)	1218.477
4	SARIMA(8, 1, 7, 2, 1, 2)	1215.847
5	SARIMA(8, 1, 8, 2, 1, 2)	1217.424
6	SARIMA(8, 1, 7, 2, 1, 1)	1221.499
7	SARIMA(8, 1, 7, 1, 1, 2)	1212.329

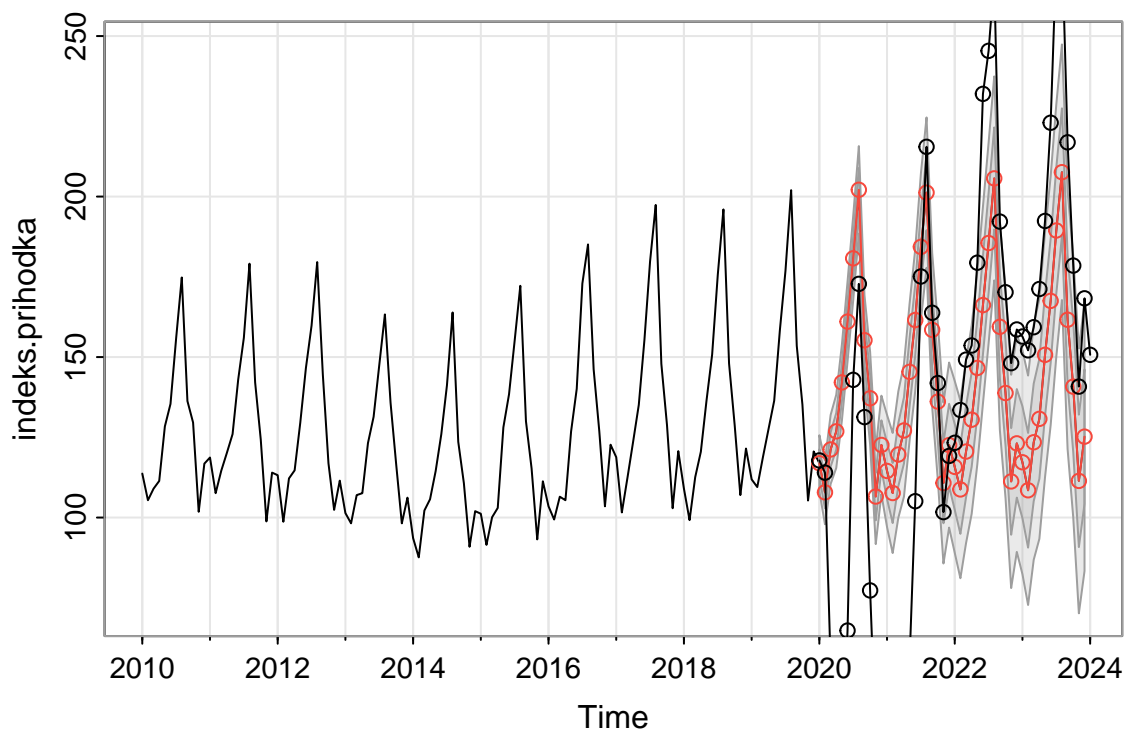
Na podlagi Akaikijevega informacijskega kriterija(AIC) se odločiva za model **SARIMA(8, 1, 7, 1, 1, 2)**. Tudi na spodnjih grafih diagnostike ostankov modela vidimo, da sva se z izbranimi parametri nekoliko bolj približala idealu. p-vrednost Ljung-Box testa kaže na statistično neznačilen rezultat (p-vrednost = 0.477), zato ničelne domneve ne moremo zavrni.



Slika 16: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

1.2.6 Napoved

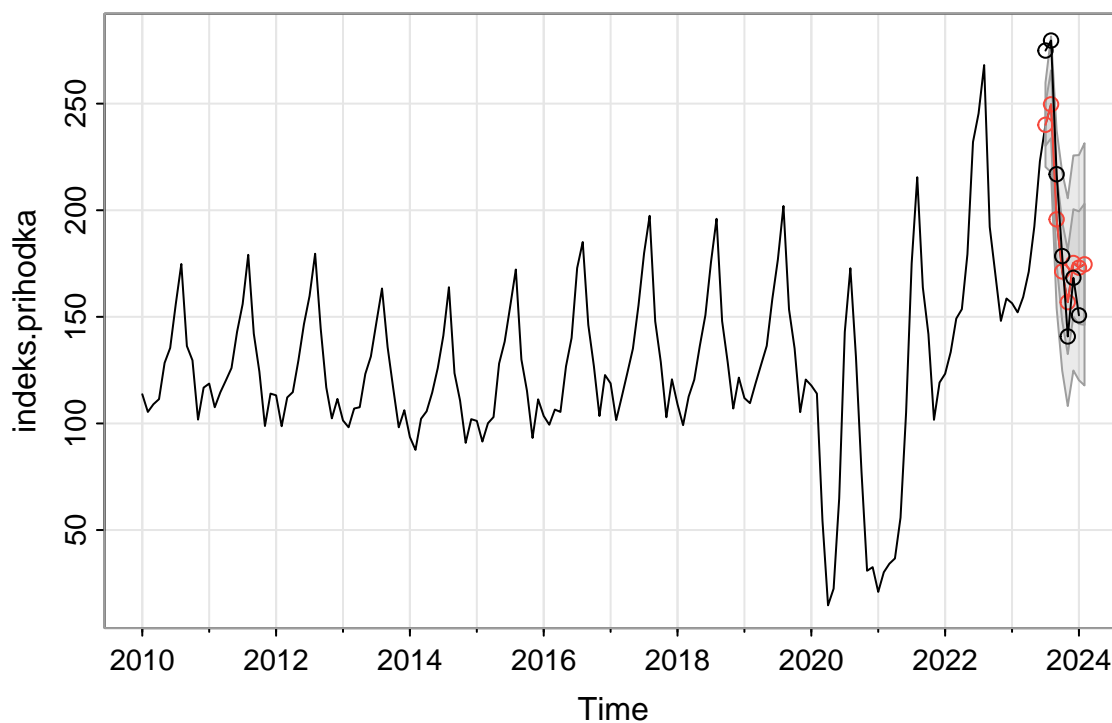
S pomočjo zgornjega modela bova sedaj poskušala napovedati indeks prihodka za nekaj let vnaprej. Ker naju zanima kako dober je najin model, na začetku ponovno podatke razdeliva na testne in učne. Najbolj naju zanima kako bo napovedoval po letu 2020, če tega leta nima v učnih podatkih in kako napoveduje, če ga vključiva v učne podatke in napovedujeva samo za mesec januar v letu 2024.



Slika 17: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

Model prevzame naraščajoč trend in ga predpostavi tudi napovedim. Pričakovano je bilo, da za leto 2020 model ne bo napovedal ustrezno, potem pa za kasnejša leta napovedi tudi niso najboljše, so nižje kot realne, torej model nekoliko podcenjuje indeks prihodka. Tudi intervali zaupanja niso ekstremno široki. Zaenkrat bi sicer rekli, da je napoved nekoliko neinformativna.

Poglejmo si sedaj še, kakšne so napovedi, če v učno množico vključimo podatke iz leta 2020 do sredine leta 2023.



Slika 18: Časovna vrsta in napovedi na podlagi modela od sredine leta 2023 naprej.

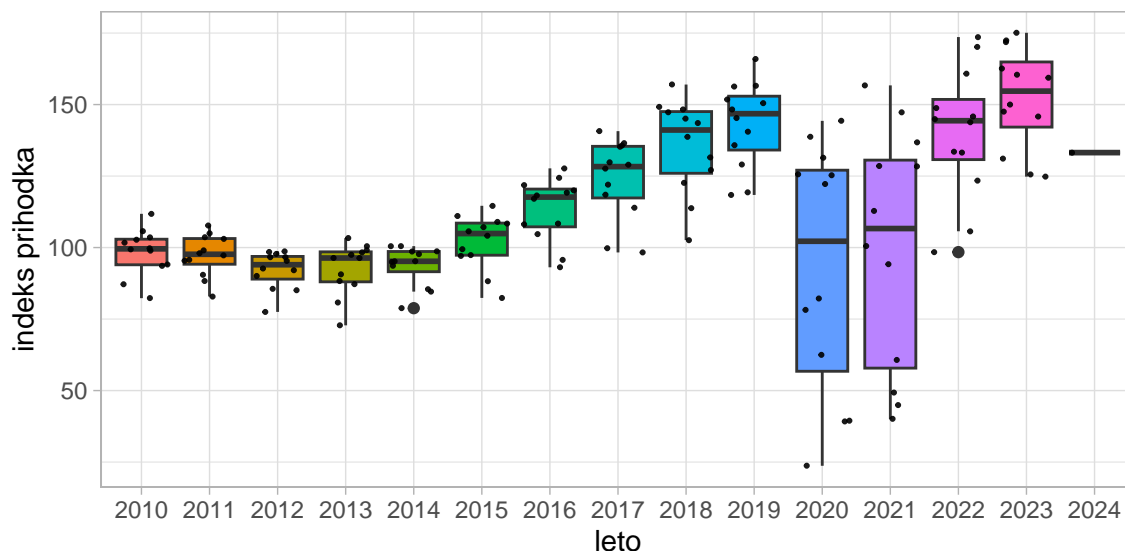
Tokrat pa vidimo, da napovedane vrednosti niso tako zelo slabe, model določene napovedi skoraj zadane z realnimi indeksi prihodka, interval zaupanja sicer še vedno ni pretirano ozek. Torej s tega bi lahko sklepala, da je najin model dober za napovedovanje vrednosti indeksa prihodka za kratke časovne intervale naprej.

1.3 Časovna vrsta *Dejavnost strežbe jedi in pijač*

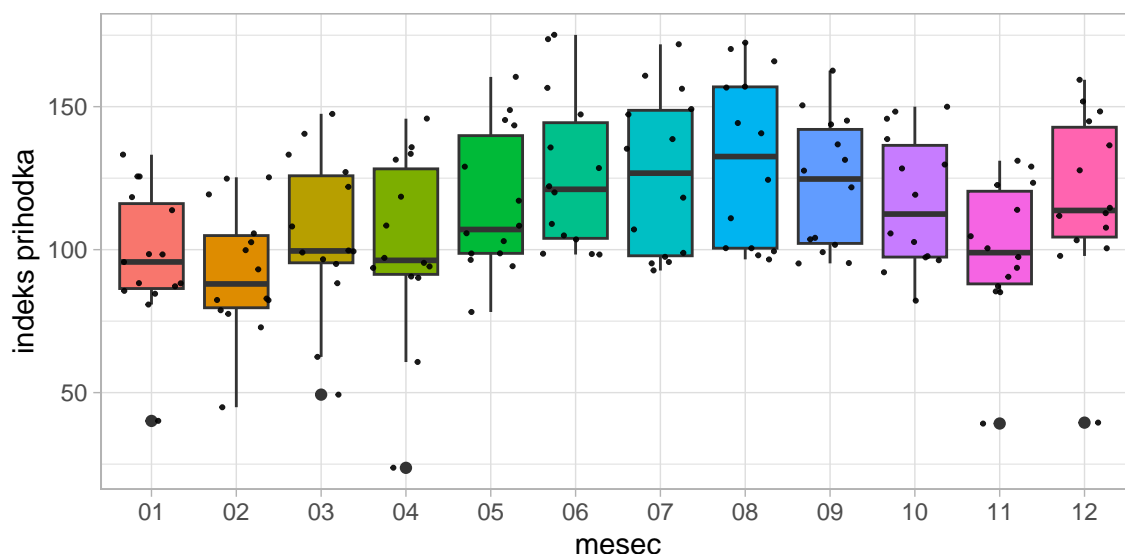
1.3.1 Pregled časovne vrste

Podatke imamo za leta od 2010 do leta 2023 za vsak mesec v letu in za leto 2024 za mesec januar.

Na spodnjem grafu porazdelitev indeksa prihodka glede na leta lahko vidimo, da bo proti koncu variabilnost časovne vrste spet zelo nekonstantna v primerjavi z ostalimi leti. Pri časovni vrsti na začetku pričakujemo naraščajoč trend in potem velike spremembe v in po letu 2020. Glede na mesece vidimo, da je prihodkov v dejavnosti največ ravno v poletnih mesecih in pa večje je tudi v mesecu decembru, kar je dokaj logično glede na življenje ljudi.

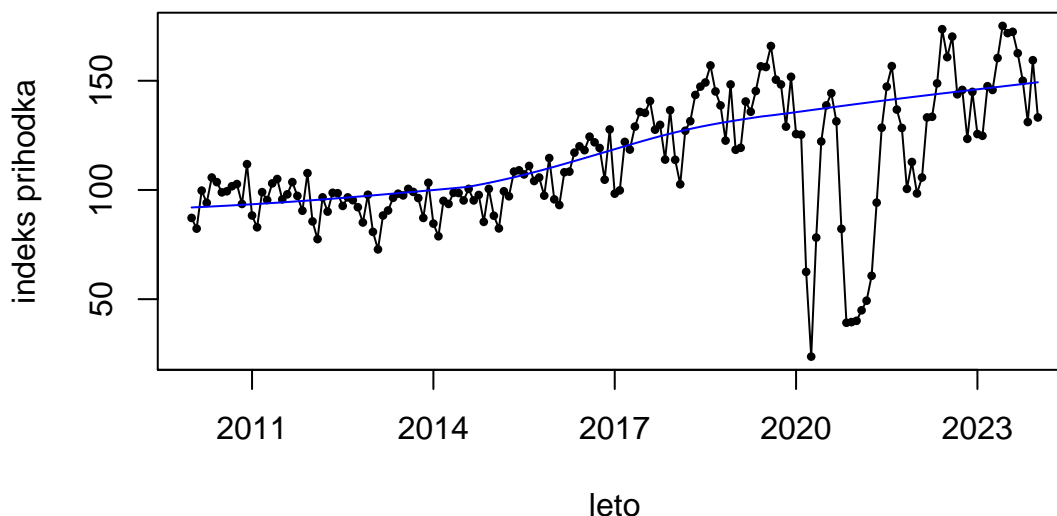


Slika 19: Porazdelitev indeksa prihodka v dejavnosti strežbe jedi in pijač glede na leta.



Slika 20: Porazdelitev indeksa prihodka v dejavnosti strežbe jedi in pijač glede na mesec.

Narišimo si sedaj časovno vrsto, ki jo bomo v nadaljevanju podrobneje analizirali.



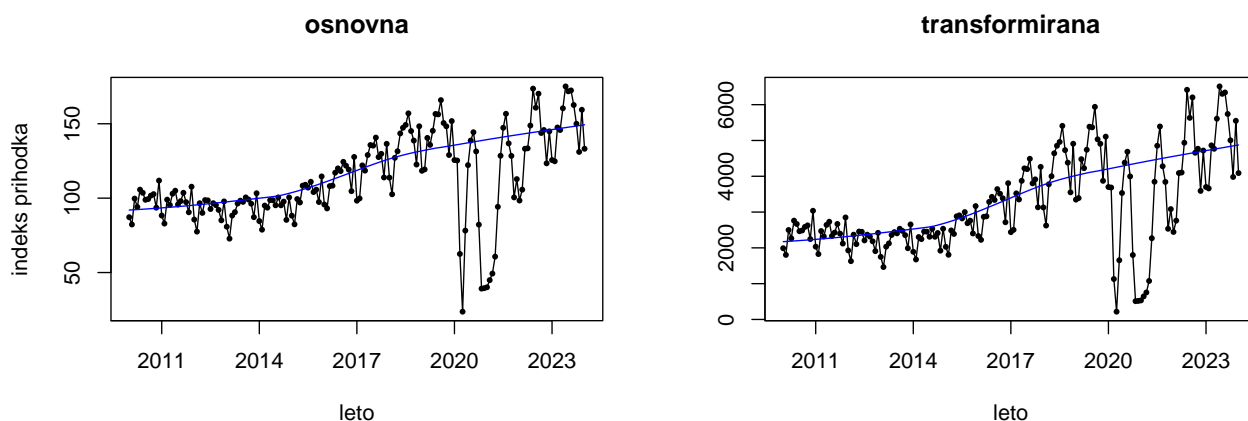
Slika 21: Časovna vrsta indeksa prihodka za 'Dejavnost strežbe jedi in pijač'.

Ponovno opazimo prisotnost trenda, ki je čez celotno obdobje naraščajoče, naklon pa je bil največji med letoma 2015 in 2018. Sprva bi glede na vzorec spreminjanja (nihanja) indeksa lahko rekli, da je prisotna sezonskost. Če natančneje pogledamo sta v posameznem letu ekstremni vrednosti poleti in meseca decembra. Do leta 2018 je bila najvišja vrednost indeksa v posameznem letu dosežena decembra, nato pa je bila višja vrednost dosežena poleti. Iz tega bi lahko rekli, da je nekoliko prisotno ciklično nihanje, vendar ni enako čez celotno opazovano obdobje.

Prav tako je dobro vidno ekstremno gibanje v času pandemije Covid-19, saj se je takrat zgodil velik padec vrednosti indeksa. Tudi amplituda nihanja se je drastično spremenila v primerjavi s predhodnjim opazovanim obdobjem, za kar je verjetno ponovno krivo obnašanje in potrebe ljudi po pandemiji in izdaja državnih bonov.

1.3.2 Transformacija

Ker imamo priročno nekonstantno varianco, s pomočjo BoxCox testa, ugotovimo, da bi bila primerna transformacija z $\lambda = 1.7$.



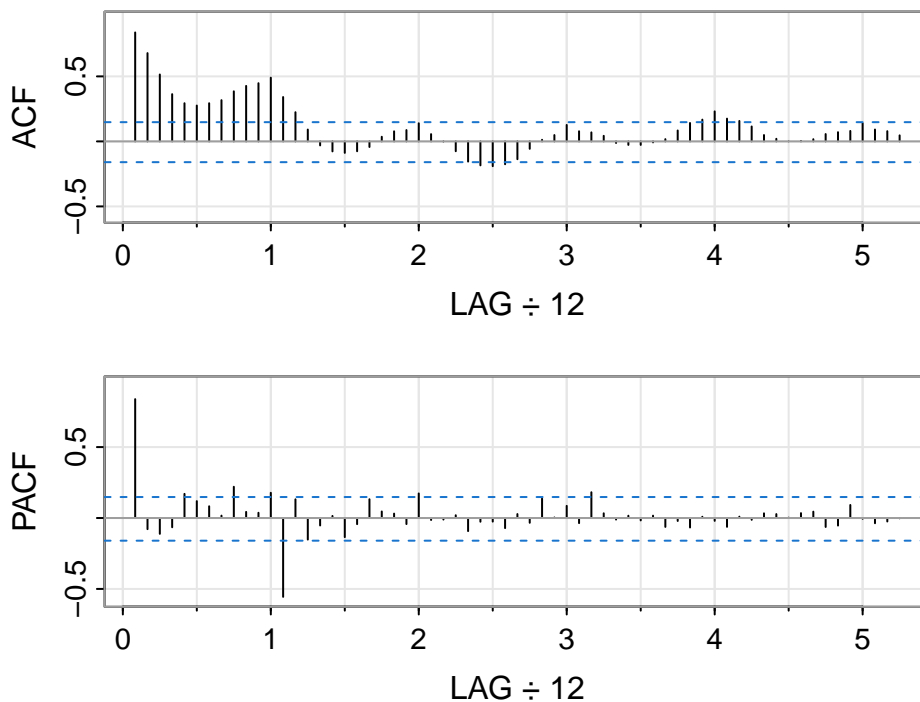
Slika 22: Osnovna in transformirana časovna vrsta 'Dejavnost strežbe jedi in pijač'.

Ker variabilnost osnovne in transformirane časovne vrste izgleda dokaj podobna, se zaenkrat ne odločiva za transformacijo in analizo nadaljujemo z osnovno časovno vrsto.

1.3.3 Avtokorelogram in parcialni avtokorelogram

Na prvem grafu je vidno, da koeficienti avtokorelacije padajo dokaj počasi, kar nakazuje, da časovna vrsta vsebuje trend, ki ga bi bilo treba modelirati z diferenciranjem. Na avtokorelogramu koeficienti avtokorelacije nakazujejo na nihanje, sicer šele pri kasnejših odlogih in tudi težko ocenimo ali gre za počasno padanje ali ne, ampak vseeno to vzamemo v razmislek za vključitev modeliranja sezone, ki se na osnovnem grafu časovne vrste ni videla. Trenutno v modelu je statistično značilna korelacija med členi časovne vrste.

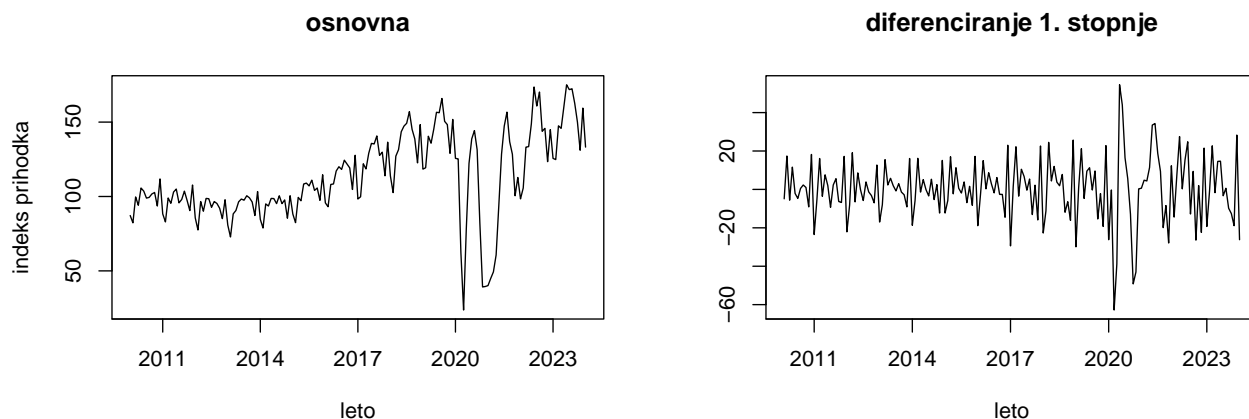
Parcialni avtokorelogram ne prikazuje posebnosti, imamo nekaj koeficientov parcialne avtokorelacije, ki so statistično značilni, ampak se pojavljajo šele pri večjih odlogih, zato temu pred modeliranjem trenda, ne bova posvečala večje pozornosti.



Slika 23: Avtokorelogram in parcialni avtokorelogram časovne vrste.

1.3.4 Diferenciranje

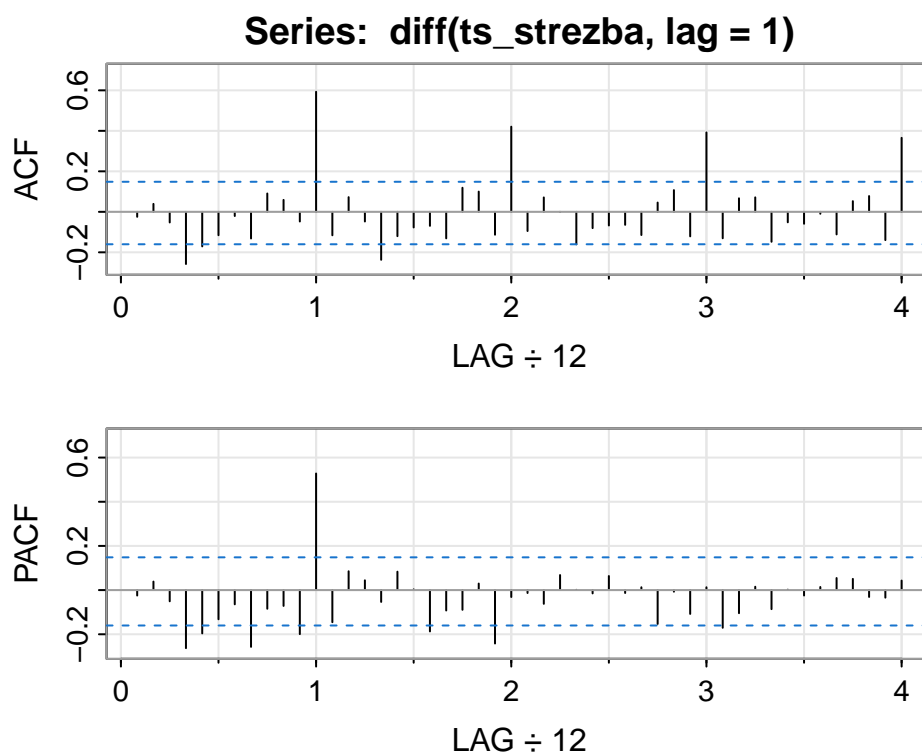
Torej vidi se na avtokorelogramu, da koeficienti avtokorelacije nakazujejo na trend v časovni vrsti, zato jo enkrat diferenciramo in poskušamo dobiti stacionarno časovno vrsto.



Slika 24: Osnovna in diferencirana časovna vrsta.

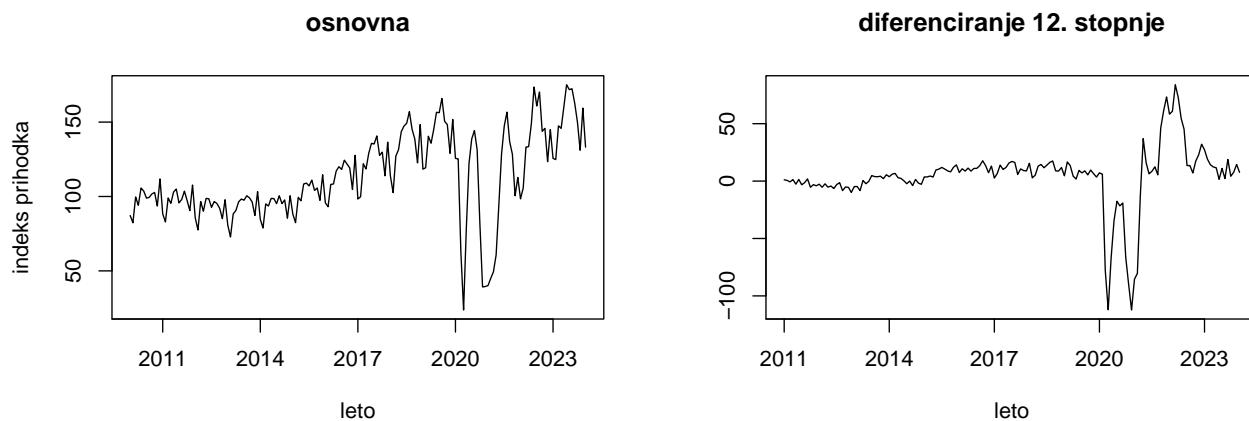
Z diferenciranjem linearnega trenda smo na levi strani drugega grafa ustvarili dokaj končno variabilnost in pričakovano vrednost enako 0, v letu 2020 temu ni tako, potem pa se ponovno “uredi” časovna vrsta.

Na spodnjem avtokorelogramu opazimo, da se pojavljajo statistično značilni koeficienti avtokorelacije na odlogih 1, 2, 3, 4; torej se nama zdi, da je prisotna neka sezonskost.



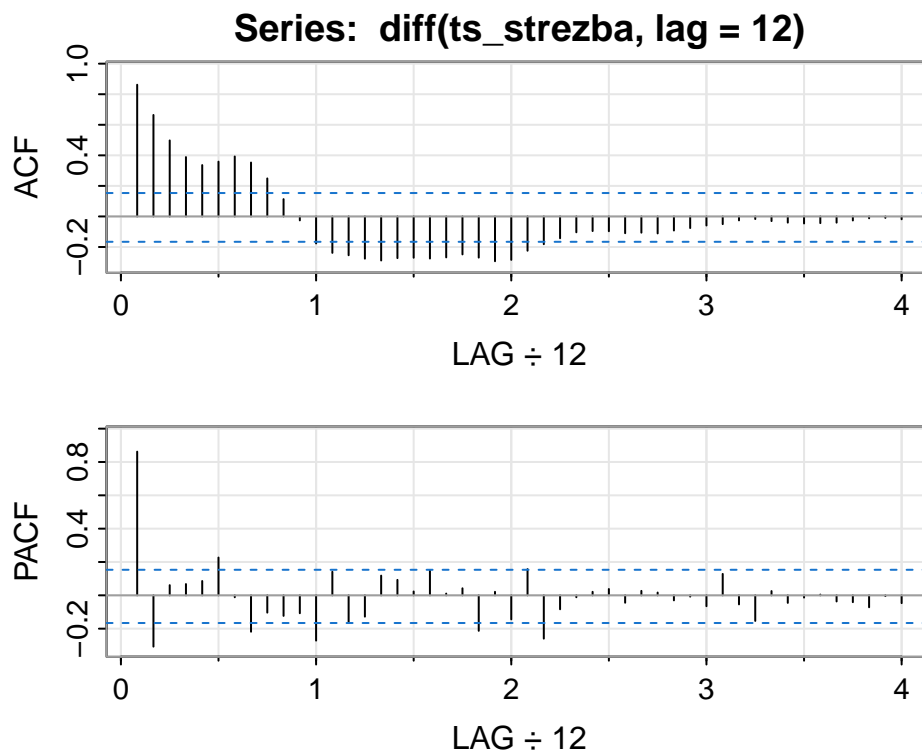
Slika 25: Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.

Zaradi zgornje ugotovitve o sezonskosti poskušajmo diferencirati še to. Na grafu vidimo, da je varianca tokrat bolj nekonstantna na levi strani levega grafa in tudi skrajno desno na levem grafu, kot pri diferenciranju 1. stopnje. Še vedno se vidi velik vpliv epidemije Covid-19.



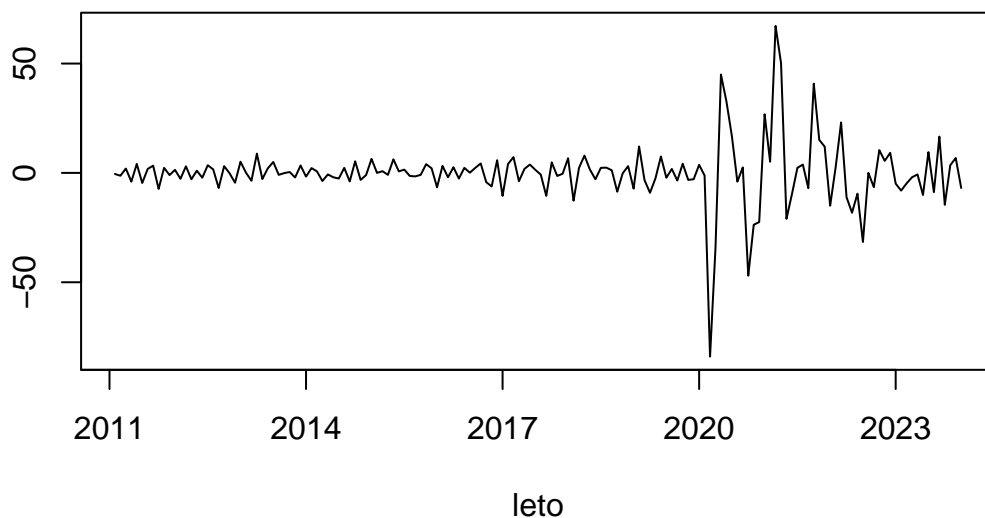
Slika 26: Osnovna in diferencirana(sezonskost) časovna vrsta.

Če diferenciramo samo sezonskost nam avtokorelogram še vedno nakazuje prisotnost trenda, zato preverimo še, kaj se zgodi če diferenciramo trend in sezonskost.



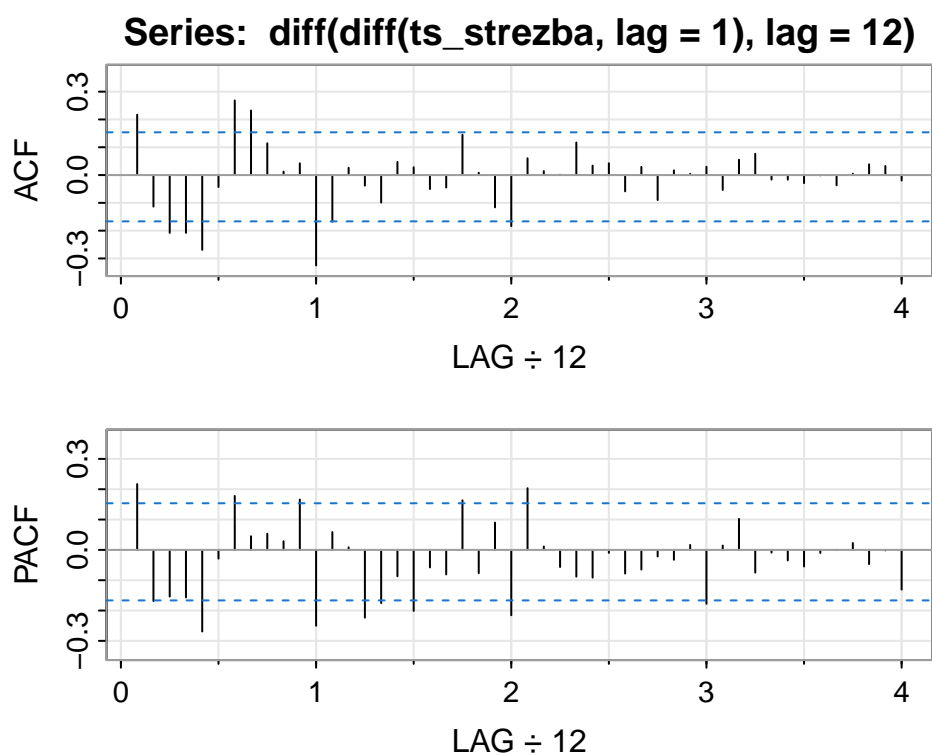
Slika 27: Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.

Na spodnjem grafu vidimo dokaj konstantno varianco in pričakovano vrednost 0 do leta 2020, nato se to, zaradi epidemije Covid-19, nekoliko spremeni, kar bomo težko še bolj modelirali.



Slika 28: Diferenciranje trenda in sezone v časovni vrsti.

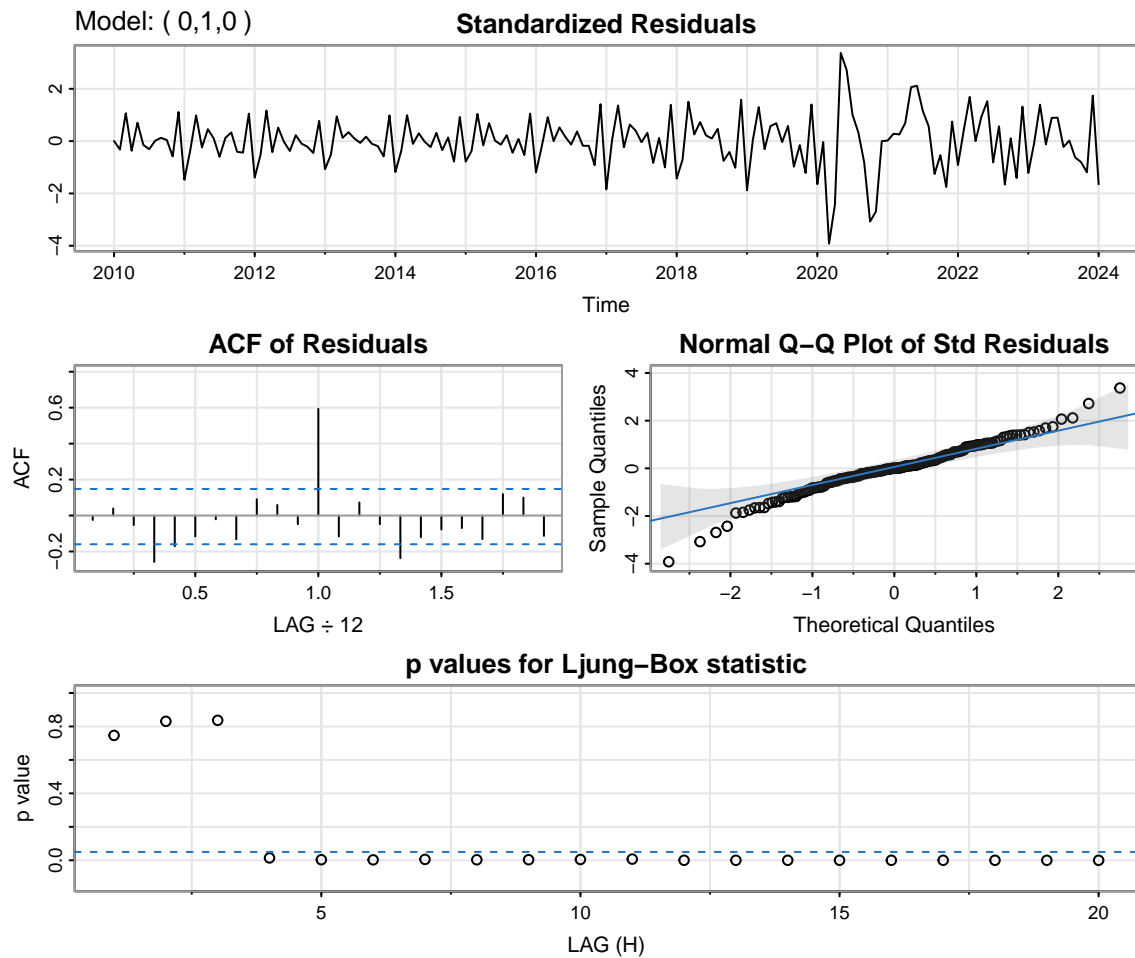
Vidimo, da imamo nekaj statistično značilnih koeficientov avtokorelacije in parcialne avtokorelacije, torej verjetno še vedno obstaja neka majhna avtokorelacija med podatki časovne vrste, ampak poskusimo sedaj narediti model.



Slika 29: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezone) časovne vrste.

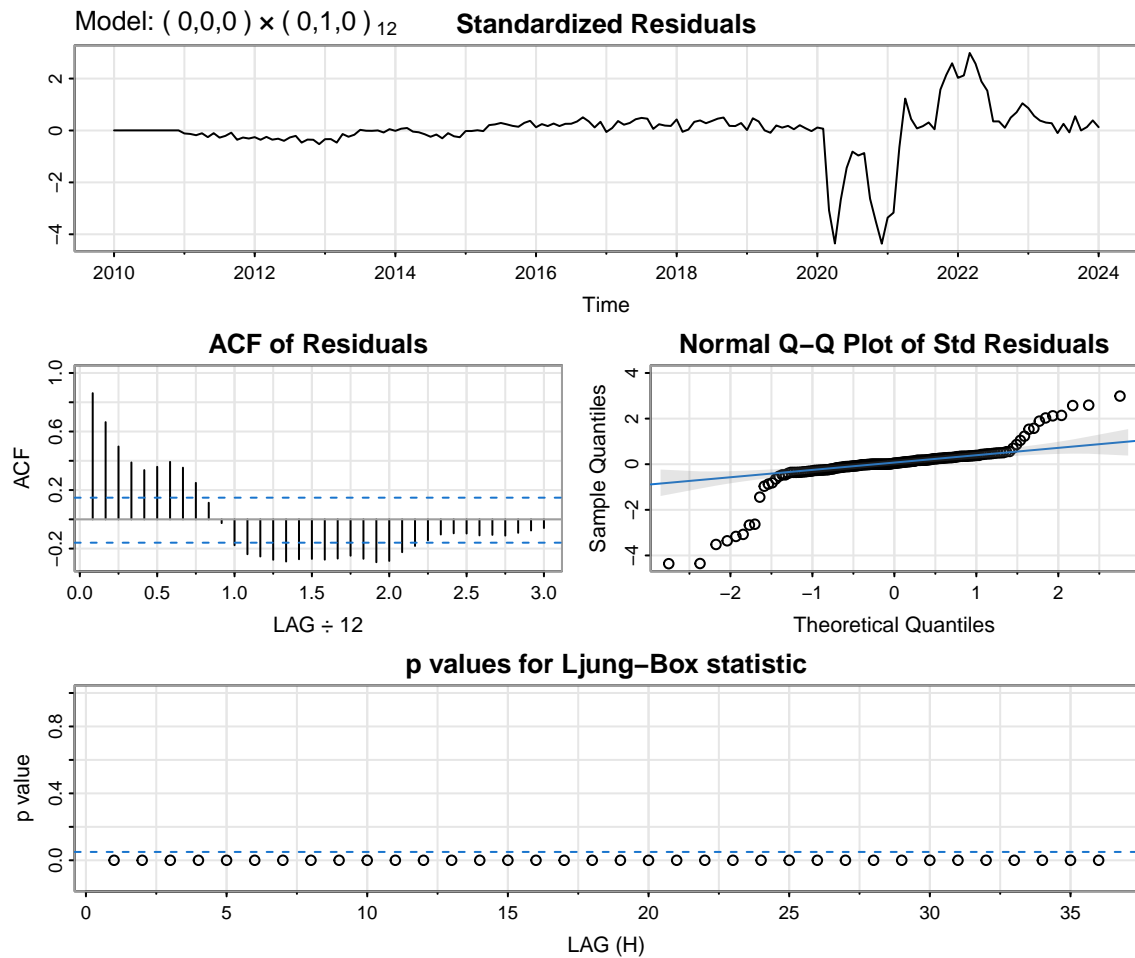
1.3.5 Izbira ustreznega modela

Izbire modela se lotiva postopoma, zato najprej pogledava kakšne rezultate dobiva, če časovni vrsti diferencirava le trend($d = 1$).



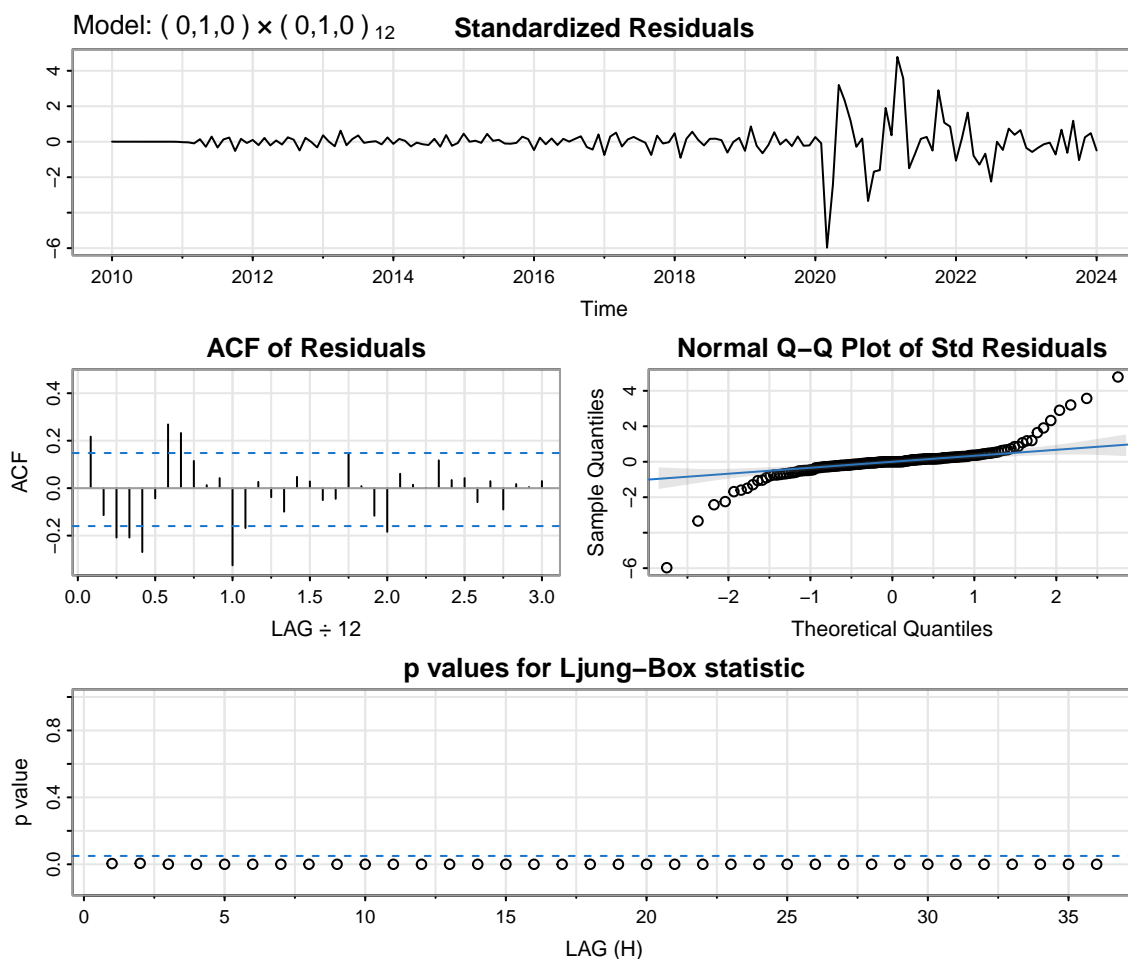
Slika 30: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.

Takoj nam je jasno, da je v časovni vrsti še vedno prisotna avtokorelacija na podlagi zgornje analize, sedaj nam Ljung-Box test to le še potrdi, saj vidimo na spodnjem grafu, da je večina p-vrednosti manjših od 0.05, torej zavračamo ničelno domnevo, da ni prisotne avtokorelacije med členi časovne vrste. Zato v model vključimo diferenciranje sezonskosti.



Slika 31: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

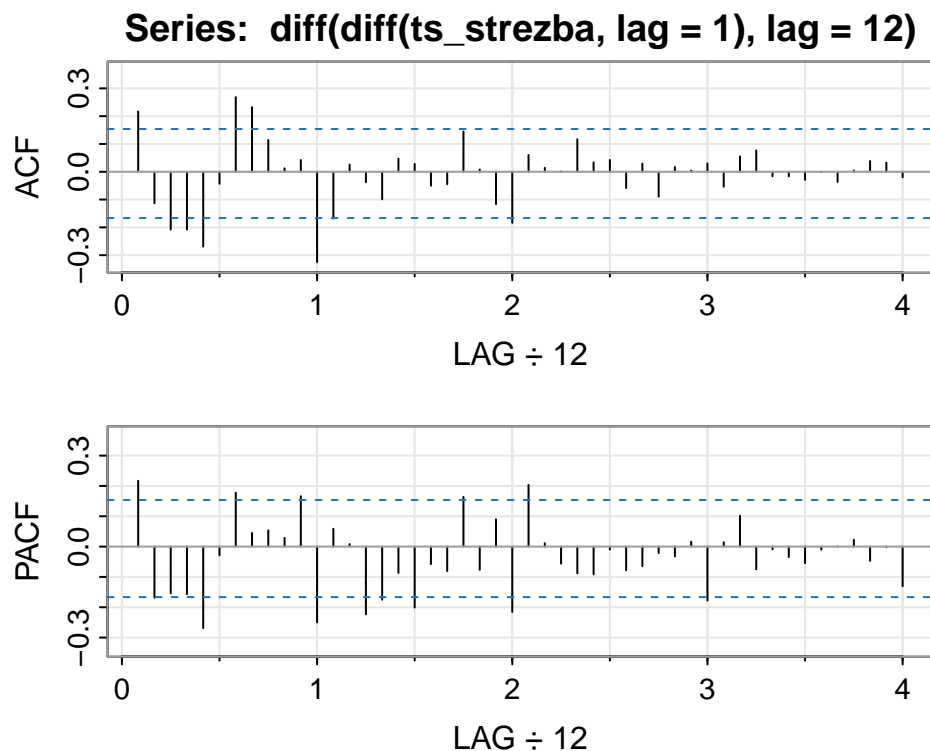
Tudi tukaj nam nam Ljung-Box test potrdi, da je v časovni vrsti prisoten trend, saj vidimo na spodnjem grafu, da imamo p-vrednosti manjše od 0.05, torej zavračamo ničelno domnevo, da ni prisotne avtokorelacije med členi časovne vrste. Zato v model vključimo diferenciranje trenda in sezonskosti.



Slika 32: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

Imamo prisotnih nekaj statistično značilnih koeficientov avtokorelacije, tudi pr Ljung-Box testu imamo še vedno p-vrednosti manjše od 0.05, torej so členi časovne vrste še vedno avtokorelirani med seboj.

Poglejmo si ponovno avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste z odlogom 1 in 12. Iz ACF razberemo, da bi lahko poskusili parameter q nastaviti za začetek na 8 in iz PACF, da lahko poskusimo parameter p nastaviti na 6 - izberemo na podlagi statistično značilnih koeficientov. Kasneje raziskujem različne vrednosti parametrov p , q , P in Q .

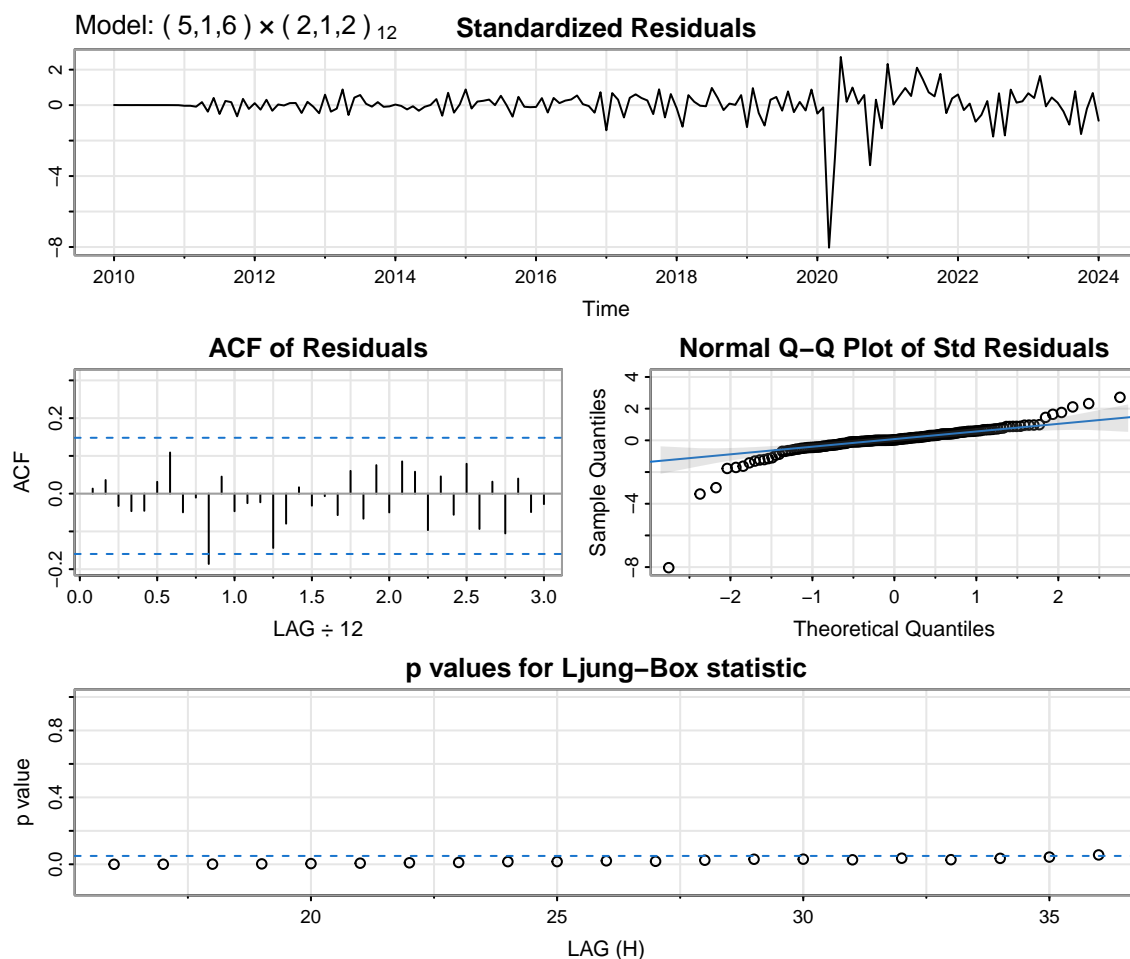


Slika 33: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.

Tabela 2: SARIMA modeli in pripadajoče AIC vrednosti.

indeks	model	AIC
1	SARIMA(8, 1, 6, 0, 1, 0)	1234.608
2	SARIMA(8, 1, 6, 2, 1, 1)	1209.009
3	SARIMA(7, 1, 6, 2, 1, 2)	1204.863
4	SARIMA(7, 1, 5, 2, 1, 2)	1209.447
5	SARIMA(5, 1, 6, 2, 1, 2)	1200.246
6	SARIMA(5, 1, 6, 2, 1, 1)	1209.481
7	SARIMA(5, 1, 8, 2, 1, 2)	1203.572
8	SARIMA(3, 1, 6, 2, 1, 2)	1208.608
9	SARIMA(5, 1, 6, 3, 1, 1)	1211.117

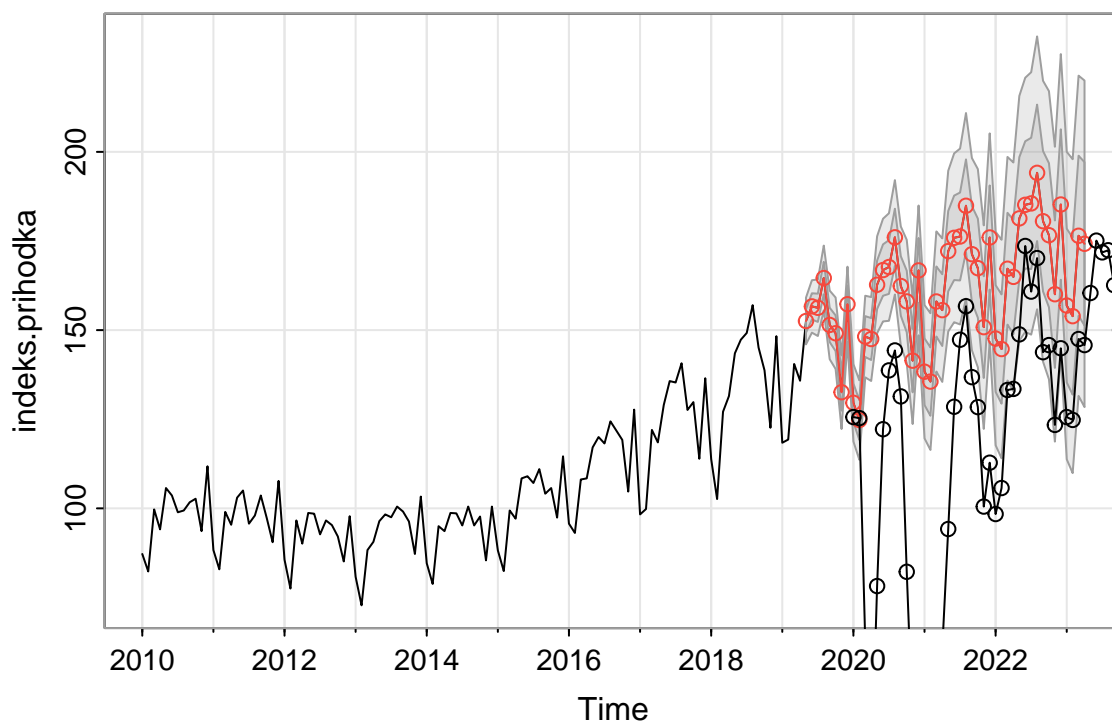
Na podlagi Akaikejevega informacijskega kriterija(AIC) se odločiva za model **SARIMA(5, 1, 6, 2, 1, 2)**. Tudi na spodnjih grafih diagnostike ostankov modela vidimo, da sva se z izbranimi parametri nekoliko bolj približala idealu. p-vrednost Ljung-Box testa pa je tudi ravno na meji(p-vrednost = 0.512), zato bi lahko zavrnili ničelno domnevo.



Slika 34: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

1.3.6 Napoved

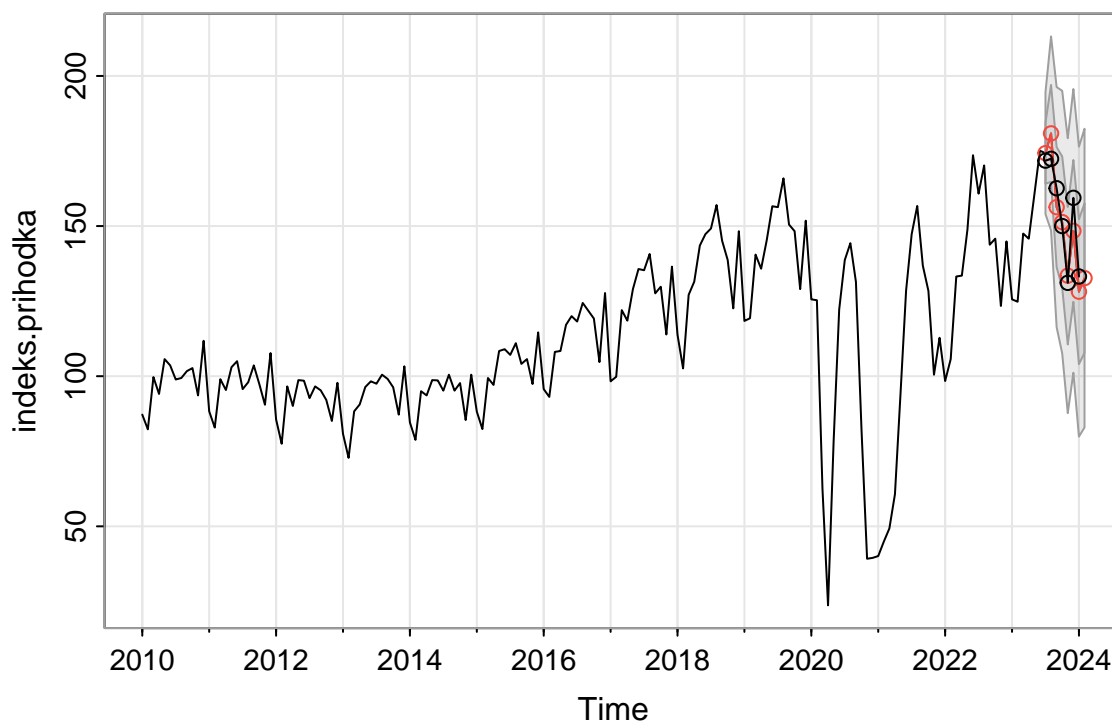
S pomočjo zgornjega modela bova sedaj poskušata napovedati indeks prihodka za nekaj let vnaprej. Ker naju zanima kako dober je najin model, na začetku ponovno podatke razdeliva na testne in učne. Najbolj naju zanima kako bo napovedoval po letu 2020, če tega leta nima v učnih podatkih in kako napoveduje, če ga vključiva v učne podatke in napovedujeva samo za mesec januar v letu 2024.



Slika 35: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

Model prevzame naraščajoč trend in ga predpostavi tudi napovedim. Pričakovano je bilo, da za leto 2020 model ne bo napovedal ustrezno, potem pa za kasnejša leta napovedi niso tako zelo slabe, so nekoliko višje kot realne, torej model nekoliko precenjuje indkes prihodka. Tudi intervali zaupanja niso ekstremno široki. Zaenkrat bi sicer rekli, da je napoved nekoliko neinformativna.

Poglejmo si sedaj še, kakšne so napovedi, če v učno množico vključimo podatke iz leta 2020 do sredine leta 2023.



Slika 36: Časovna vrsta in napovedi na podlagi modela od sredine leta 2023 naprej.

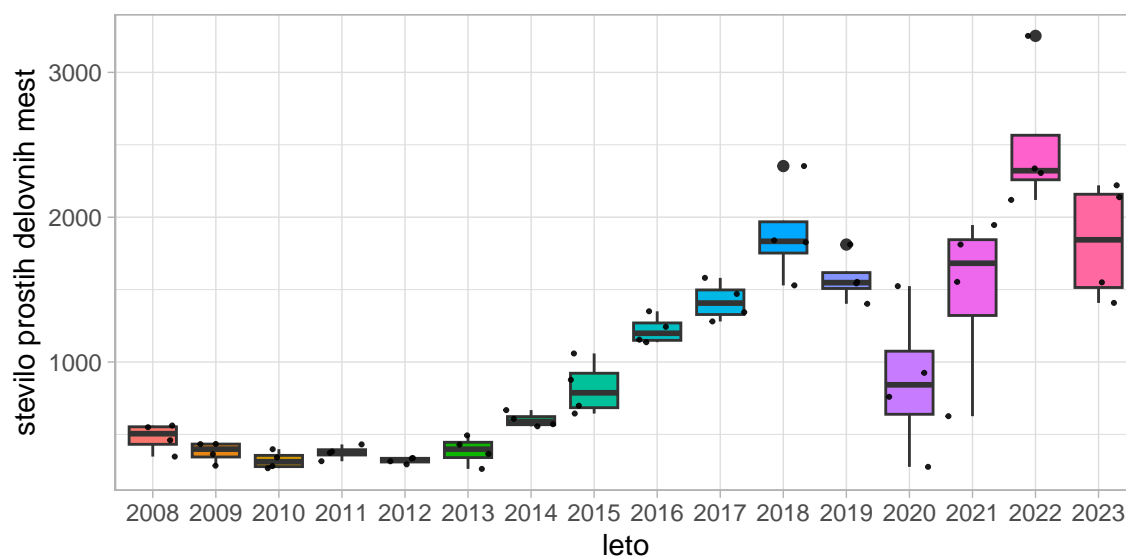
Tokrat pa vidimo, da napovedane vrednosti niso tako zelo slabe, model določene napovedi skoraj zadane z realnimi indeksi prihodka, interval zaupanja sicer še vedno ni pretirano ozek. Torej s tega bi lahko sklepala, da je najin model dober za napovedovanje vrednosti indeksa prihodka za kratke časovne intervale naprej.

1.4 Časovna vrsta *Gostinstvo prosta delovna mesta 1+*

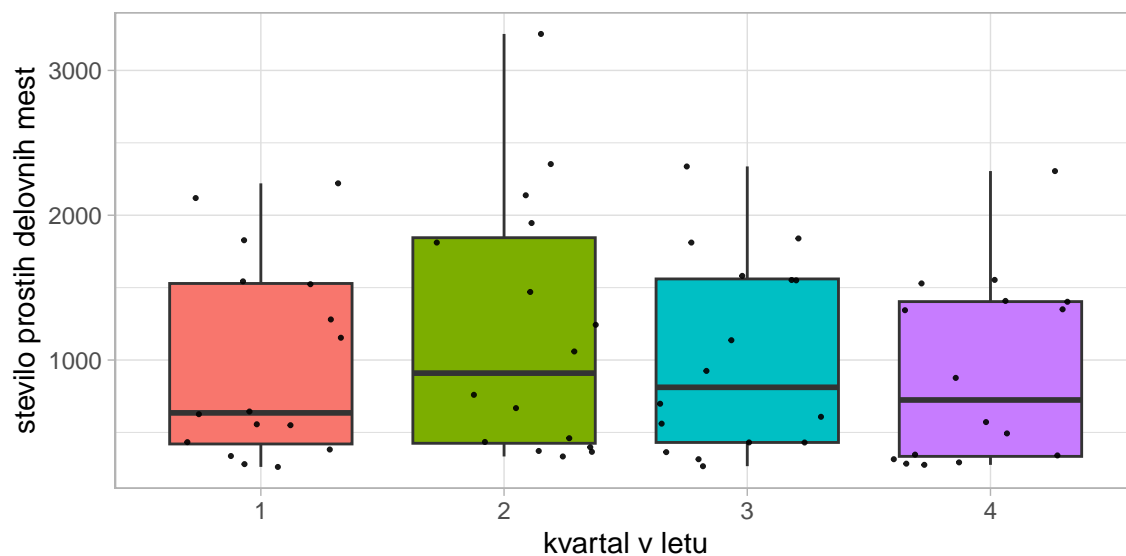
1.4.1 Pregled časovne vrste

Podatke imamo za leta od 2008 do leta 2023 za vsako kvartalno obdobje.

Na spodnjem grafu porazdelitve prostih delovnih mest v gostinstvu glede na leta lahko vidimo, da bo proti koncu variabilnost časovne vrste zelo nekonstantna v primerjavi z ostalimi leti. Pri časovni vrsti na začetku pričakujemo naraščajoč trend in potem velike spremembe v letu 2020. Glede na kvartal pa je porazdelitev dokaj enaka, malenkost odstopa morda le drugo kvartalno obdobje.

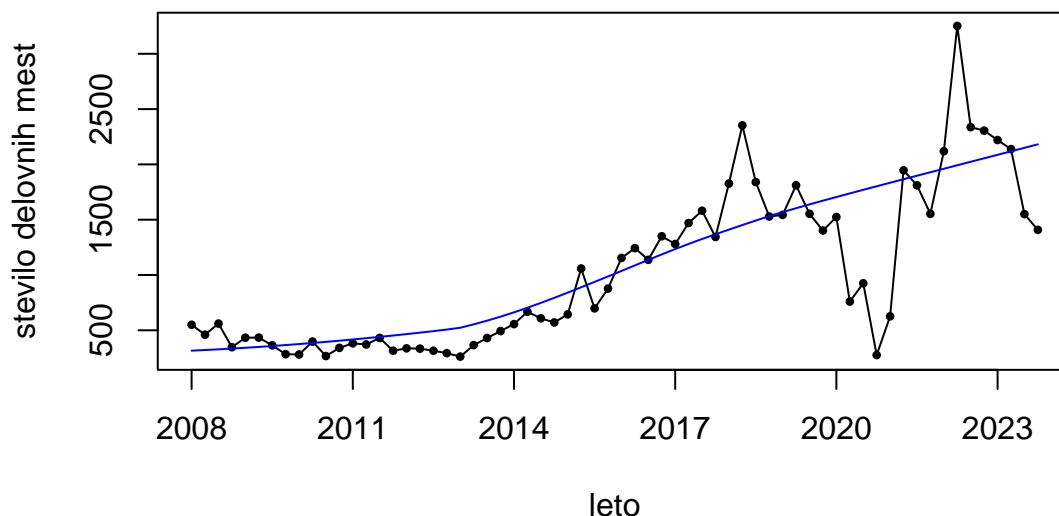


Slika 37: Porazdelitev prostih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na leto.



Slika 38: Porazdelitev prostih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na kvartal.

Narišimo si sedaj časovno vrsto, ki jo bomo v nadaljevanju podrobneje analizirali.



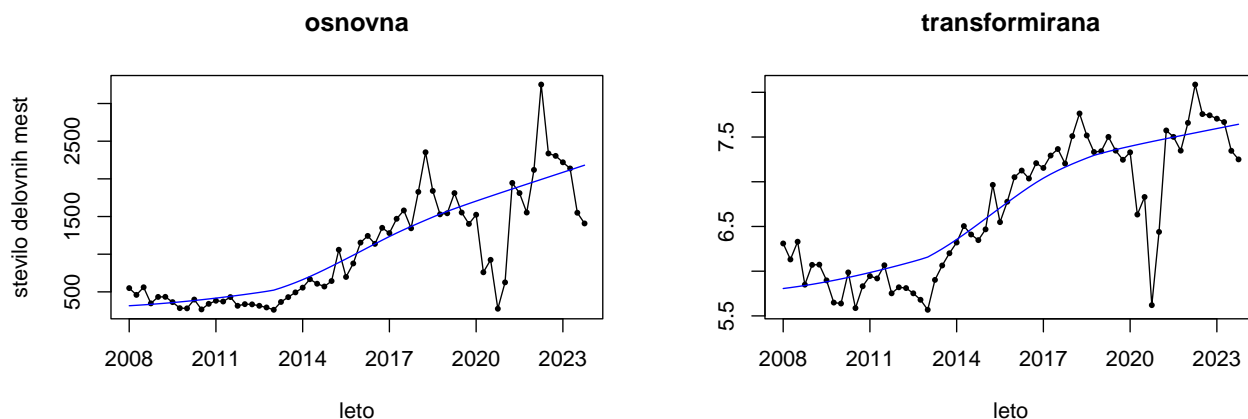
Slika 39: Časovna vrsta 'Gostinstvo prosta delovna mesta 1+'.

Kot sva omenila že zgoraj je iz grafa časovne vrste možno opaziti prisotnost trenda. Ta čez celotno opazovano obdobje narašča. Sprva nekoliko bolj položno, po letu 2013 pa se naklon poveča. Sezonskosti ni mogoče zaznati, saj so lokalni ekstremi bolj naključni, kot da bi se ponavljali v ciklu.

Vidno pa je ekstremno obdobje (čas pandemije Covid-19) in to ravno proti koncu časovne vrste. To obdobje bo verjetno povzročalo težave pri iskanju ustreznega modela za napovedovanje prihodnjih prostih delovnih mest v gostinstvu, kjer je zaposlena 1 oseba. Če bi podate razdelili na učno in testno množico in v testno množico vzamemo zadnja 3 leta v naših podatkih, napovedi modela zagotovo ne bodo pravilne, saj se bo model učil na razmeroma normalnih in predvidljivih podatkih.

1.4.2 Transformacija časovne vrste

S pomočjo BoxCox testa, ugotovimo, da je primerno časovno vrsto transformirati in se s tem morda poskušamo malenkost znebiti nekonstantne variance. Primerna je logaritemska transformacija.



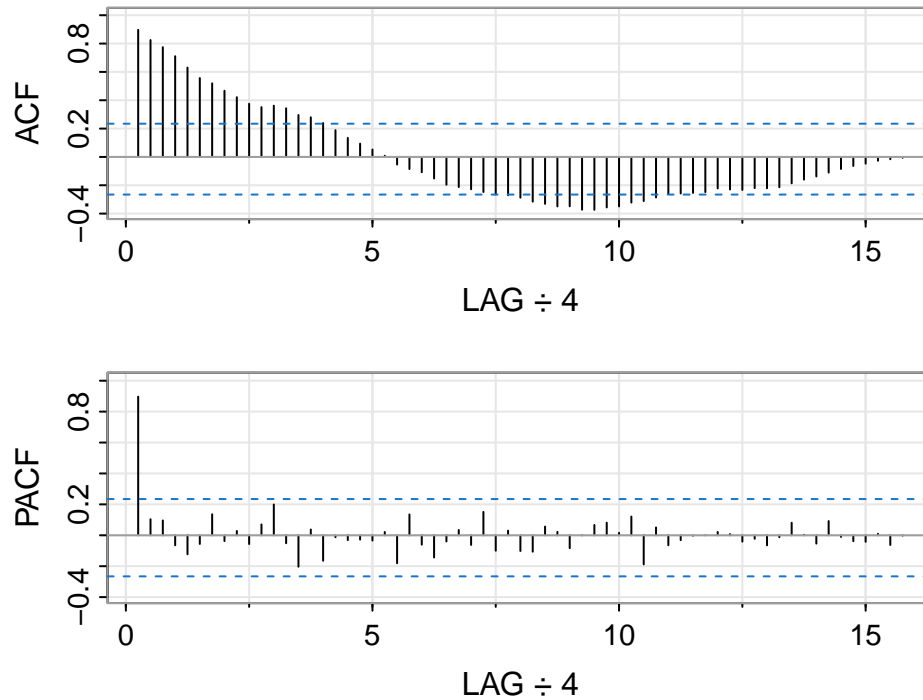
Slika 40: Osnovna in logaritmirana časovna vrsta 'Gostinstvo prosta delovna mesta 1+'.

Krivulji osnovne in transponirane časovne vrste sta si dokaj podobni, ampak nadaljujmo analizo z transformirano časovno vrsto.

1.4.3 Avtokorelogram in parcialni avtokorelogram

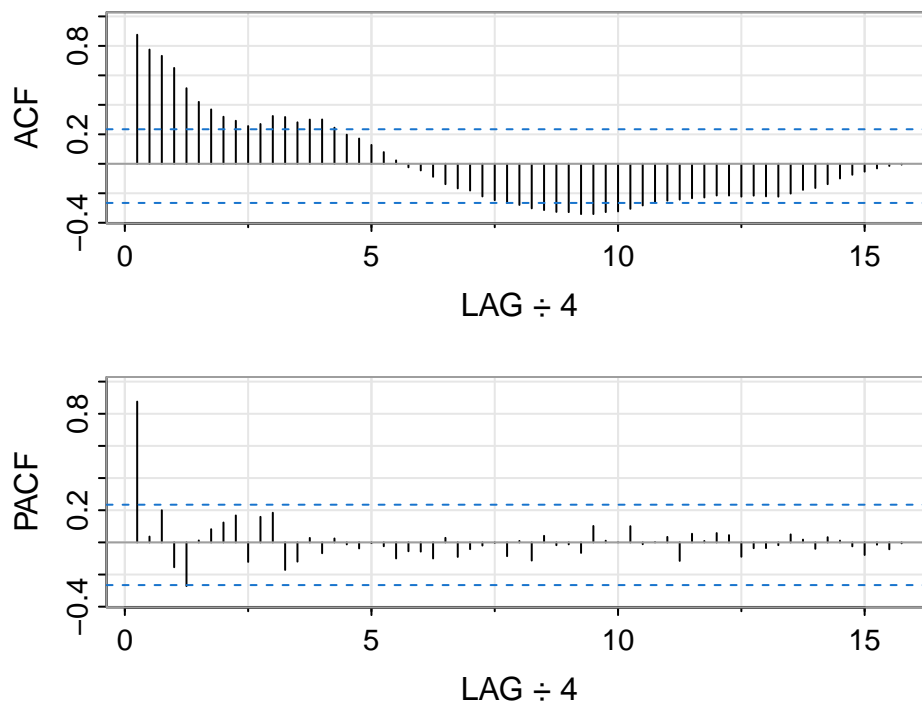
Na prvem grafu je vidno, da koeficienti avtokorelacije počasi padajo, torej časovna vrsta vsebuje trend, ki ga bi bilo treba modelirati. Sprva je bil narisan osnovni avtokorelogram, s katerega se je videlo, da če bi povečali odloge (*Lag*), bi morda bilo videti nihanje. Na spodnjem avtokorelogramu imamo povečane odloge in koeficienti avtokorelacije morda nakazujejo na zelo počasno nihanje.

Parcialni avtokorelogram ne prikazuje posebnosti, imamo samo en koeficient parcialne avtokorelacije statistično značilen, zato sklepava, da bomo v modelu imeli red avtoregresijskega modela enak 1 ali pa celo 0.



Slika 41: Avtokorelogram in parcialni avtokorelogram logaritmirane časovne vrste.

Poglejmo si še avtokorelogram in parcialni avtokorelogram osnovne časovne vrste, torej tiste brez logaritemske transformacije, samo kot zanimanje, kaj točno se dogaja.

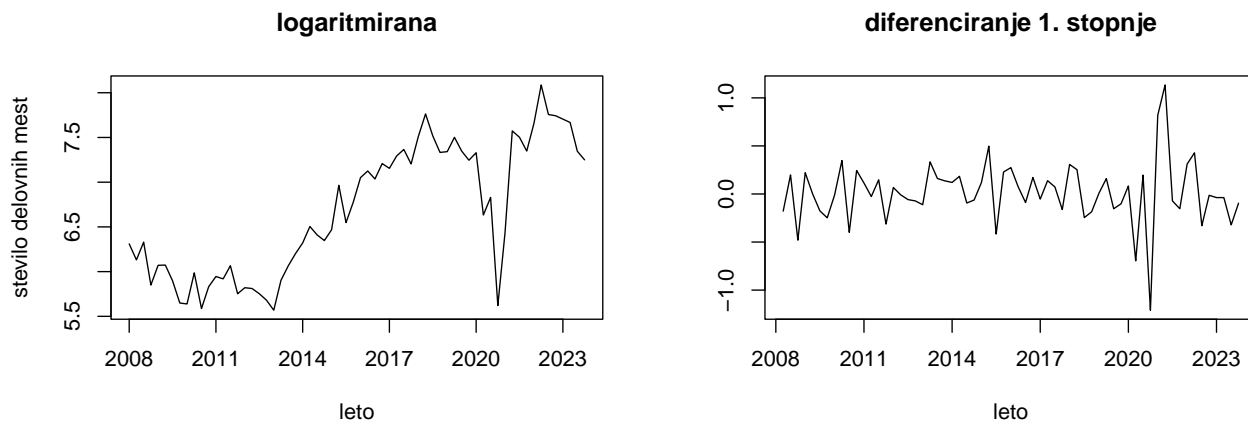


Slika 42: Avtokorelogram in parcialni avtokorelogram osnovne časovne vrste.

Kot pri avtokorelogramu in parcialnem avtokorelogramu je zelo podobna situacija kot pri logaritmirani časovni vrsti.

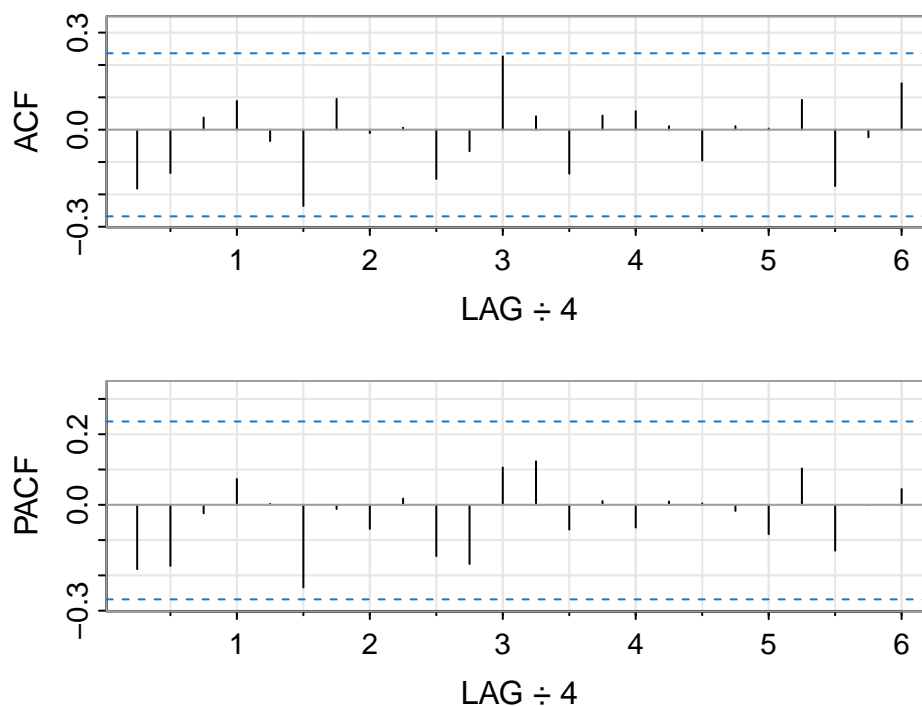
1.4.4 Diferenciranje

Kot sva že napisala zgoraj, koeficienti avtokorelacije nakazujejo na trend v logaritmirani časovni vrsti, zato jo enkrat diferenciramo in poskušamo dobiti stacionarno časovno vrsto.



Slika 43: Logaritmirana osnovna in logaritmirana diferencirana časovna vrsta.

Lahko bi rekli, da smo z diferenciranjem linearnega trenda dobili stacionarno časovno vrsto. Bolje temu zadoščamo sicer na levi strani drugega grafa, saj izgleda da je pričakovana vrednost enaka 0 in variabilnost dokaj končna, na desni strani, pri letu 2020 pa je še veden vpliv Covid-19 epidemije.

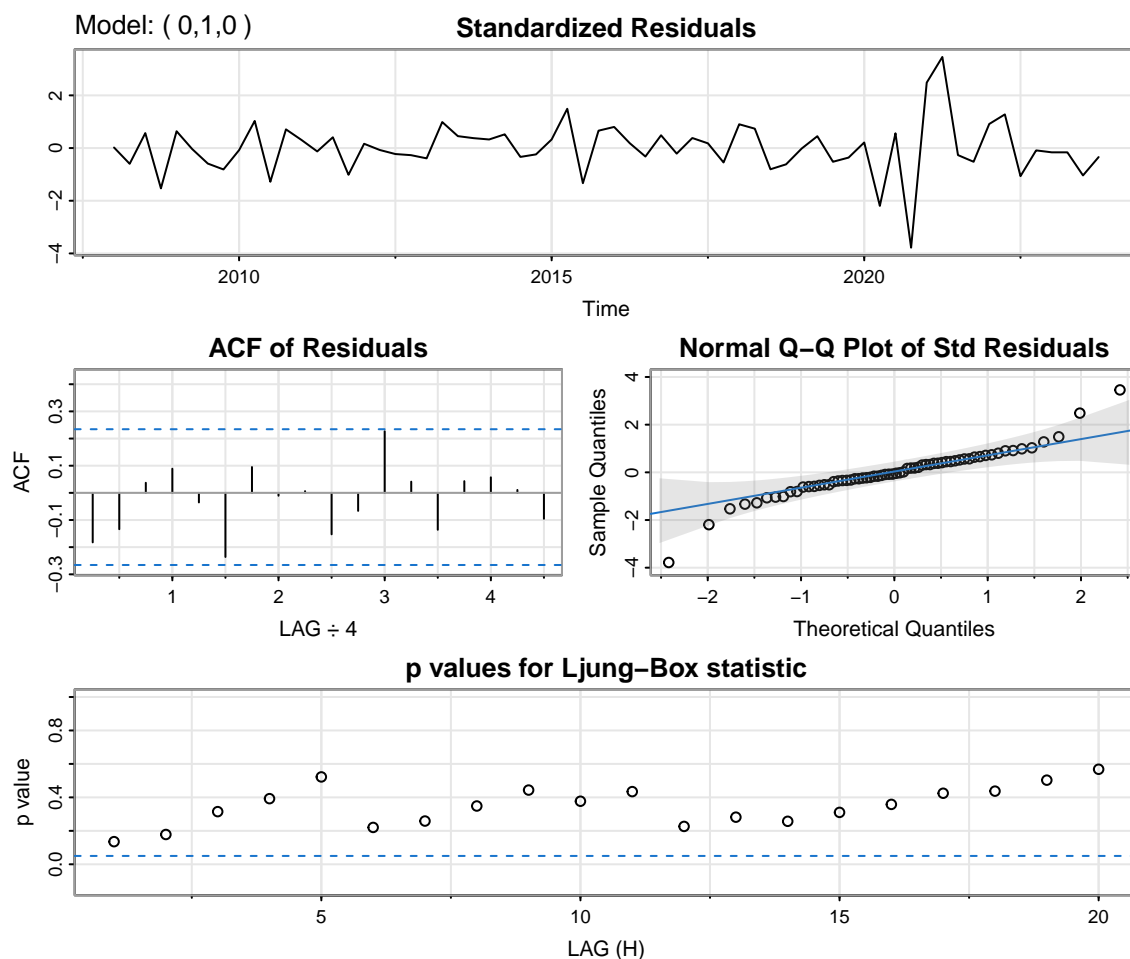


Slika 44: Avtokorelogram in parcialni avtokorelogram logaritmirane diferencirane časovne vrste.

Na zgornjem avtokorelogramu in parcialnem avtokorelogramu vidimo, da z diferenciranjem poskrbimo, da niso več prisotni statistično značilni koeficienti avtokorelacije, kar pomeni, da v modelu ni statistično značilne korelacije med členi časovne vrste. Zanimivo je le videti, da imamo nekoliko večji koeficient od ostalih pri odlogu “1.5” in pri odlogu “3” - morda je to nekaj kar še vedno malo bolj vpliva na avtokorelacijo, ampak je statistično neznačilno.

1.4.5 Izbira ustreznega modela

V model sva najprej vključila diferenciranje prvega reda. Gre za ARIMA model s koeficienti $p = 0$, $d = 1$ in $q = 0$

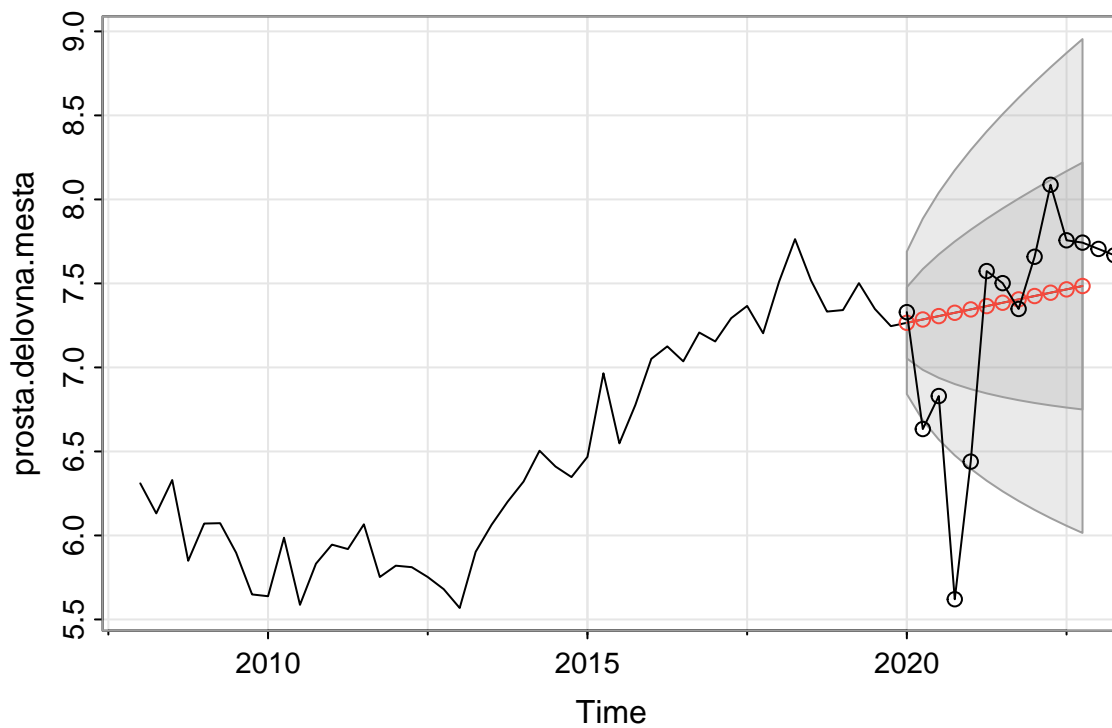


Slika 45: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa.

Koeficienti avtokorelacije so znotraj 95% intervala zaupanja, torej so statistično neznačilni, kar pomeni, da v modelu ni statistično značilne korelacije med ostanki (preverili smo tudi z Ljung-Box testom, kjer nismo mogli zavrniti ničelne domneve, da v tej časovni vrsti ostankov ni več avtokorelacije). Na podlagi tega lahko predpostavljamo, da ostanki so beli šum.

1.4.6 Napoved

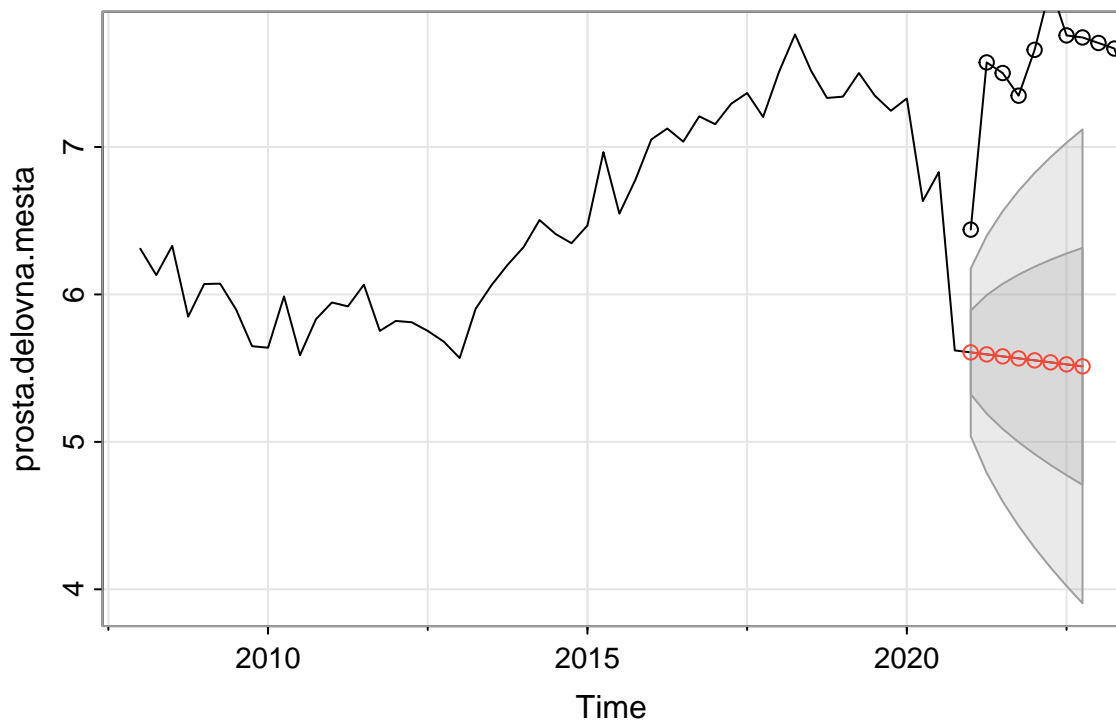
Sedaj bova poskušala z modelom še napovedati število prostih delovnih mest za nekaj let vnaprej. Ker naju zanima kako dober je najin model, na začetku podatke razdeliva na testne in učne. Najbolj naju zanima kako bo napovedoval po letu 2020, če tega leta nima v učnih podatkih in kako napoveduje, če ga vključiva v učne podatke in napovedujeva od leta 2021 naprej.



Slika 46: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

Jasno je, da napoved ni najboljša, saj ostanki vrste predstavljajo beli šum in pričakovano že na začetku poglavja, da model ne zna napovedati padca v letu 2020, saj je naučen po naraščajočem trendu, kar napovedi tudi prikazujejo. Ker so slabe napovedi, so temu primerni tudi izjemno široki 95% intervali zaupanja.

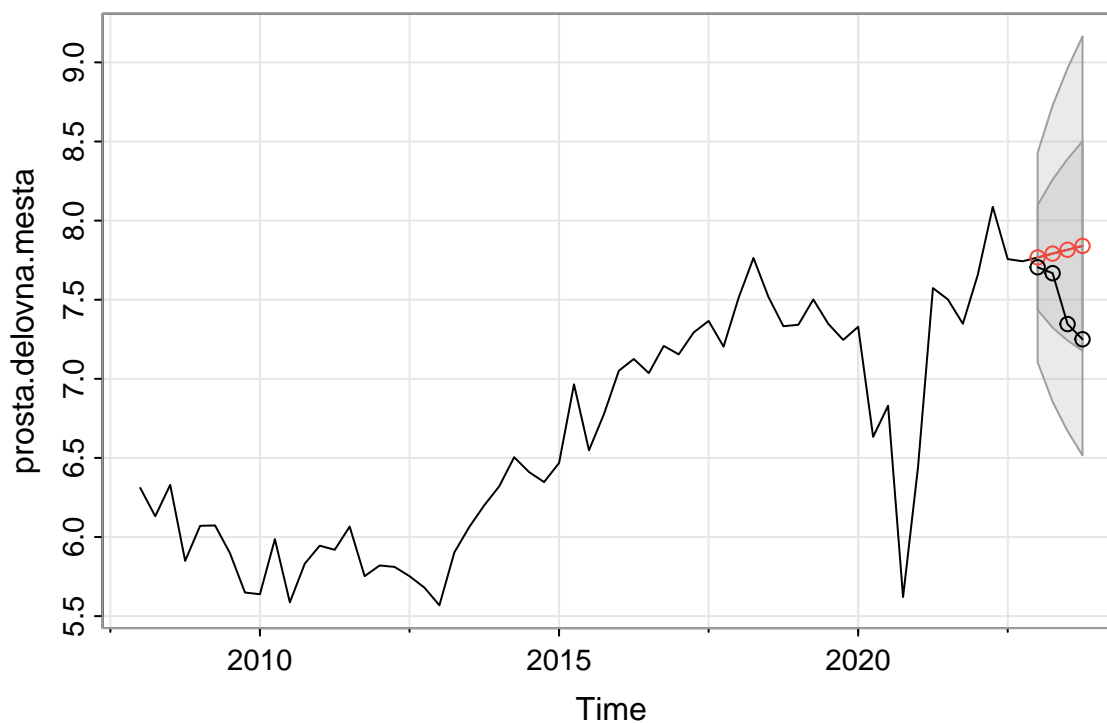
Poglejmo si sedaj še, kakšne so napovedi, če v učno množico vključimo podatke iz leta 2020.



Slika 47: Časovna vrsta in napovedi na podlagi modela od leta 2021 naprej.

Pričakovano tudi sedaj napovedi niso ravno informativne, model je prevzel trend zadnjega leta in ga je nadaljeval tudi naprej.

Sedaj pa napovejmo samo za zadnje leto, leto 2023, da bo časovna vrsta imela v učnih podatkih in padec leta 2020 ter potem ponovno rast.



Slika 48: Časovna vrsta in napovedi na podlagi modela od leta 2023 naprej.

Napovedi tudi tukaj niso pravilne in ne ravno najbolj informativne, najboljše zadane samo za en trenutek naprej, 95% intervali zaupanja pa so še vedno izjemno široki.

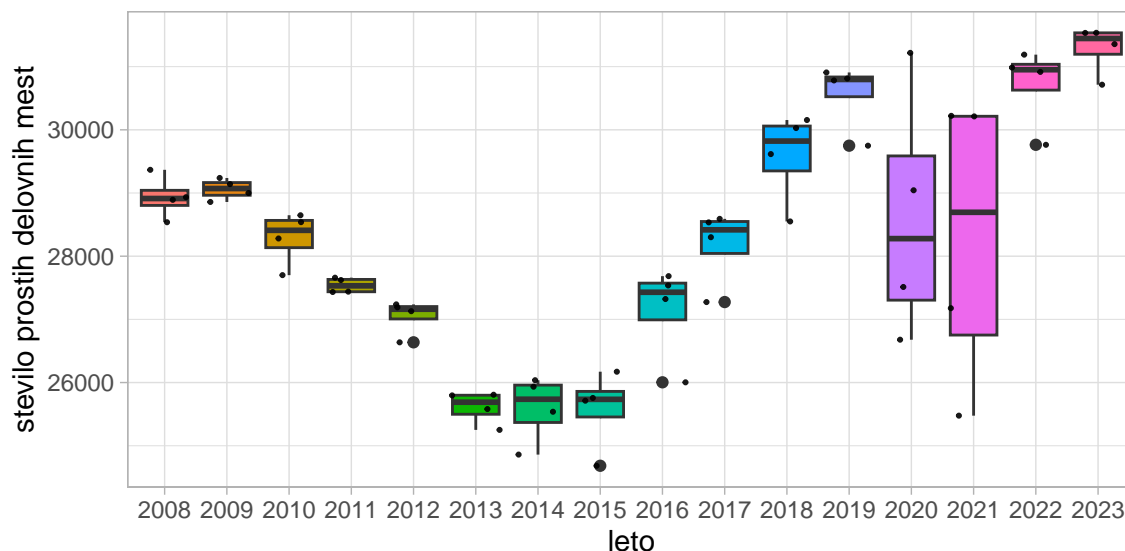
Ker je časovna vrsta tako zelo razgibana in ima veliko variabilnost ravno v zadnjih letih, težko uporabljamo ARIMA oz. SARIMA modele za napovedovanje prihodnjih vrednosti.

1.5 Časovna vrsta *Gostinstvo zasedena delovna mesta 1+*

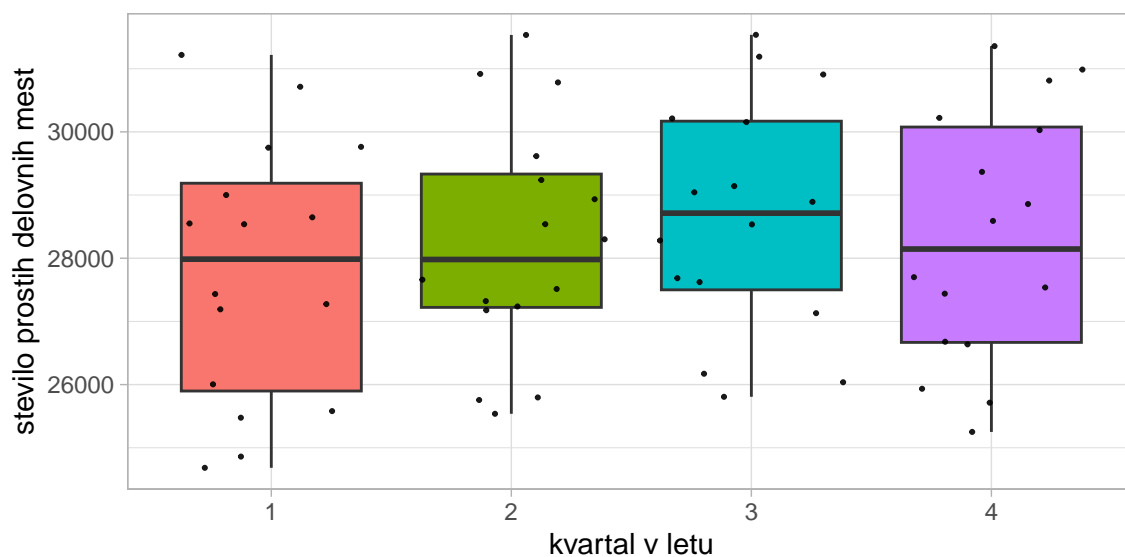
1.5.1 Pregled časovne vrste

Podatke imamo za leta od 2008 do leta 2023 za vsako kvartalno obdobje.

Na spodnjem grafu porazdelitve zasedenih delovnih mest v gostinstvu glede na leta lahko vidimo, da bo proti koncu variabilnost časovne vrste zelo nekonstantna v primerjavi z ostalimi leti, predvsem v letu 2020 in 2021. Pri časovni vrsti na začetku pričakujemo padajoč trend do leta 2014, ki nato postane pozitiven ter potem velike spremembe v letu 2020. Glede na kvartal je srednja vrednost dokaj enaka, malenkost odstopa morda le tretje kvartalno obdobje. Nekaj razlik je tudi v variabilnosti vrednosti znotraj kvartalnega obdobja.

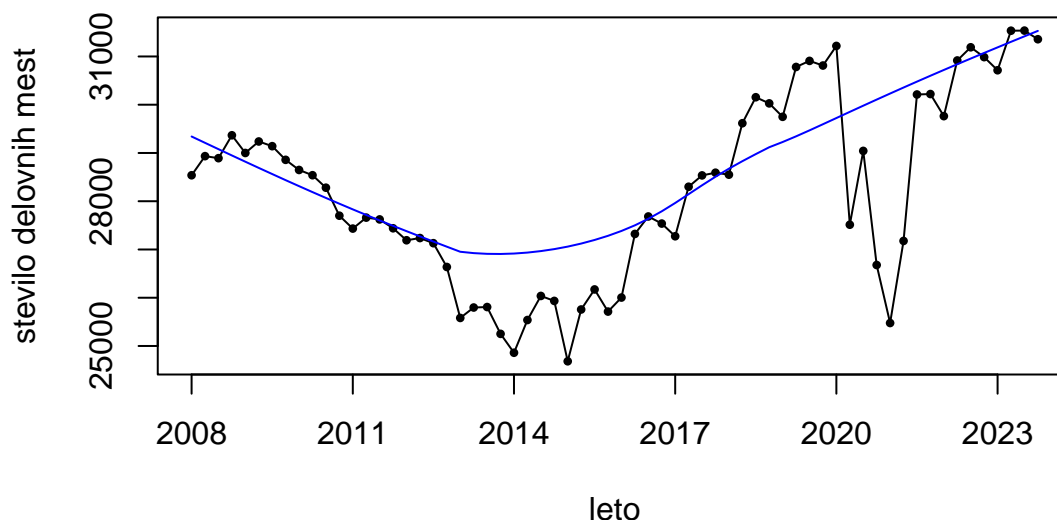


Slika 49: Porazdelitev zasedenih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na leto.



Slika 50: Porazdelitev zasedenih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na kvartal.

Narišimo si sedaj časovno vrsto, ki jo bomo v nadaljevanju podrobneje analizirali.



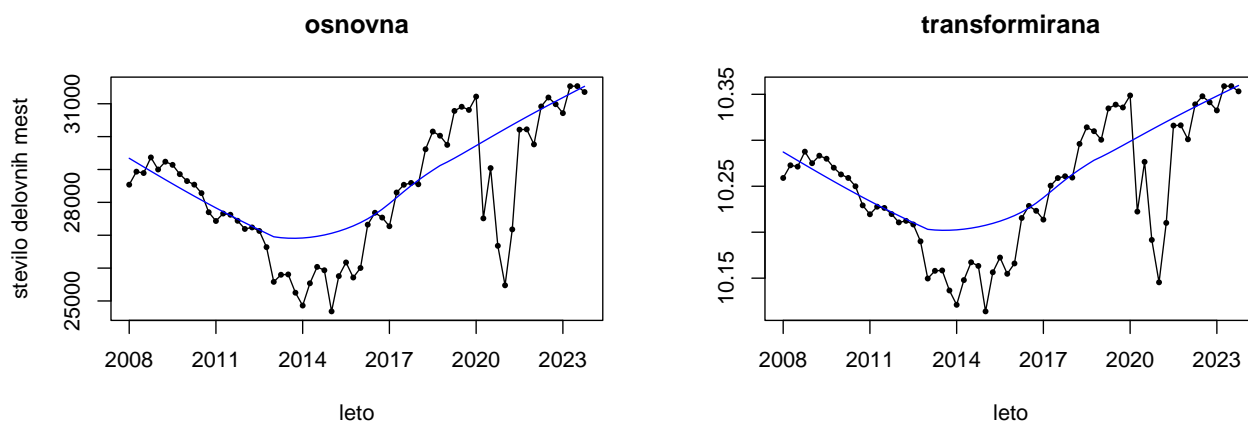
Slika 51: Časovna vrsta 'Gostinstvo zasedena delovna mesta 1+'.

Kot sva omenila že zgoraj je iz grafa časovne vrste možno opaziti prirobnost trenda. Ta je od leta 2008 do 2014 padajoč, nato pa se obrne in začne strmo naraščati. Sezonskosti ni mogoče zaznati, saj so lokalni ekstremi bolj naključni, kot da bi se ponavljali v ciklu.

Vidno pa je ekstremno obdobje (čas pandemije Covid-19) in to ravno proti koncu časovne vrste. To obdobje bo verjetno povzročalo težave pri iskanju ustreznega modela za napovedovanje prihodnjih zasedenih delovnih mest v gostinstvu, kjer je zaposlena 1 oseba. Če bi podatke razdelili na učno in testno množico in v testno množico vzamemo zadnja 3 leta v naših podatkih, napovedi modela zagotovo ne bodo pravilne, saj se bo model učil na razmeroma normalnih in predvidljivih podatkih.

1.5.2 Transformacija časovne vrste

S pomočjo BoxCox testa, ugotovimo, da je 95% interval dokaj širok in vsebuje tako $\lambda = 0$ in $\lambda = 1$. Uporabili bomo logaritemsko transformacijo in časovno vrsto primerjali s časovno vrsto, kjer transformacije ne uporabimo.



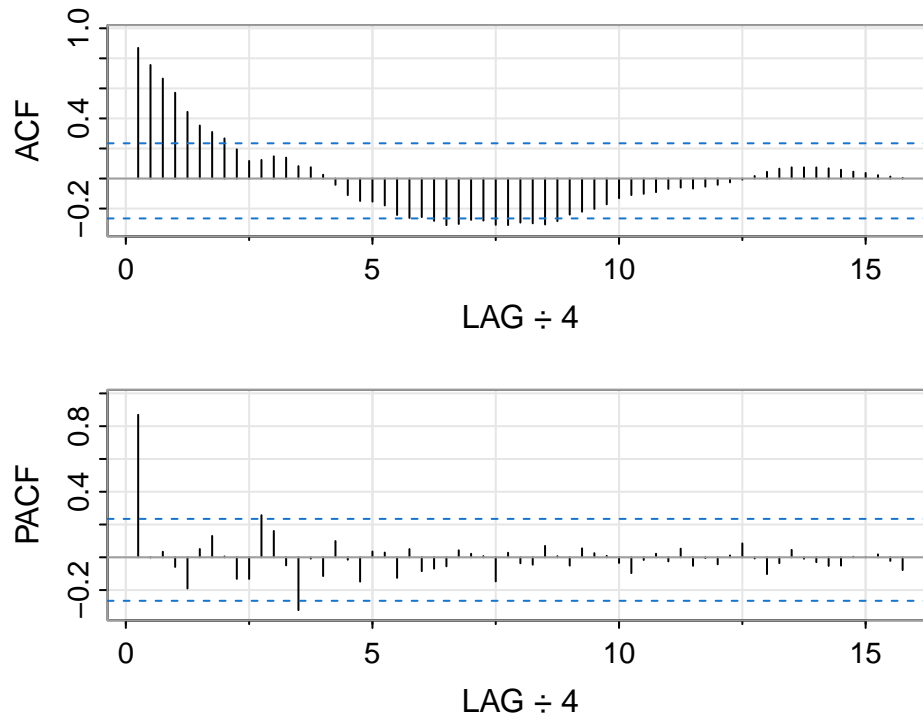
Slika 52: Osnovna in logaritmirana časovna vrsta 'Gostinstvo zasedena delovna mesta 1+'.

Krivulji osnovne in transponirane časovne vrste sta si identični, z izjemo razpona vrednosti na y -osi. Zato nadaljujmo analizo z netransformirano časovno vrsto.

1.5.3 Analiza avtokoreliranosti

Na prvem grafu je vidno, da koeficienti avtokorelacije počasi padajo, torej časovna vrsta vsebuje trend, ki ga bi bilo treba modelirati. Sprva je bil narisani osnovni avtokorelogram, s katerega se je videlo, da če bi povečali odloge (*Lag*), bi morda bilo videti nihanje. Na spodnjem avtokorelogramu imamo povečane odloge in koeficienti avtokorelacije morda nakazujejo na zelo počasno nihanje.

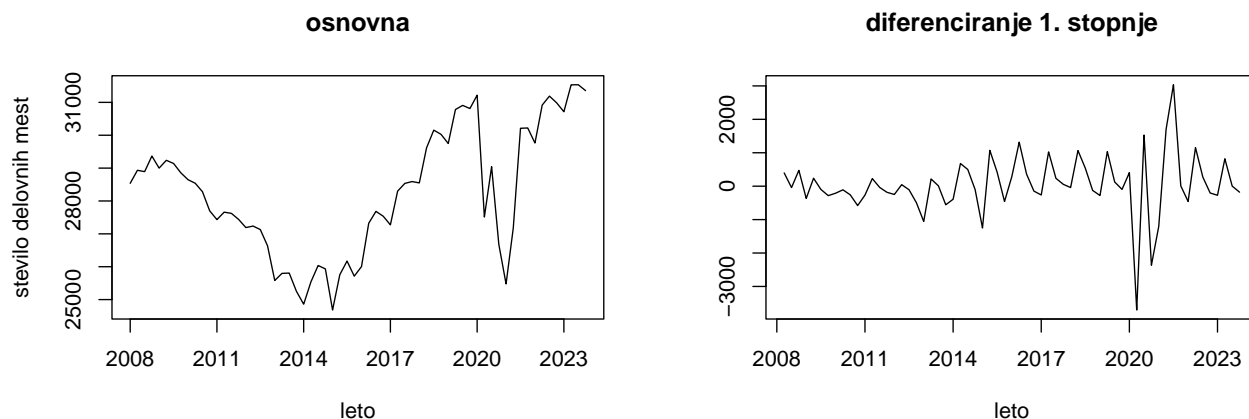
Parcialni avtokorelogram ne prikazuje posebnosti, imamo samo en koeficient parcialne avtokorelacije statistično značilen (kasneje sta se dva vendar jih zaradi oddaljenosti ne upoštevamo), zato sklepava, da bomo v modelu imeli red avtoregresijskega modela enak 1 ali pa celo 0.



Slika 53: Avtokorelogram in parcialni avtokorelogram logaritmirane časovne vrste.

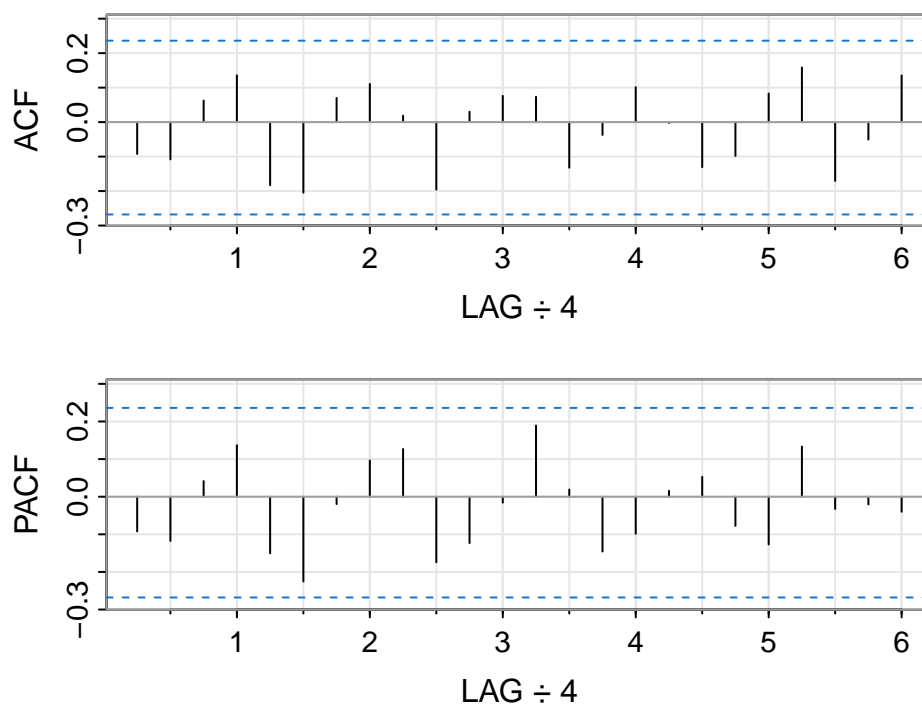
1.5.4 Diferenciranje

Kot sva že napisala zgoraj, koeficienti avtokorelacije nakazujejo na trend v časovni vrsti, zato jo enkrat diferenciramo in poskušamo dobiti stacionarno časovno vrsto.



Slika 54: Oosnovna in diferencirana časovna vrsta.

Lahko bi rekli, da smo z diferenciranjem linearnega trenda dobili stacionarno časovno vrsto. Bolje temu zadoščamo sicer na levi strani drugega grafa, saj izgleda da je pričakovana vrednost enaka 0 in variabilnost dokaj končna, na desni strani, pri letu 2020 pa je še vedno vpliv Covid-19 epidemije.

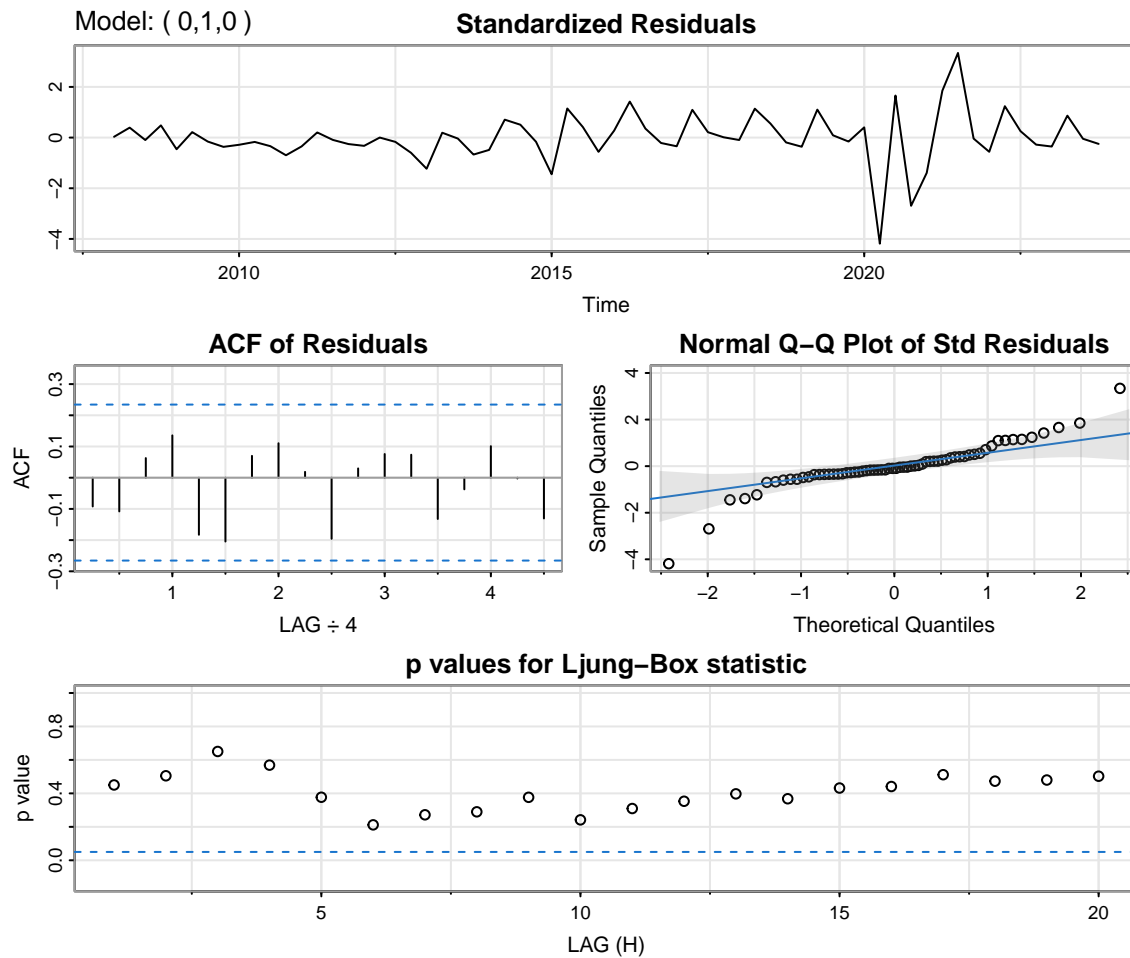


Slika 55: Avtokorelogram in parcialni avtokorelogram logaritmirane diferencirane časovne vrste.

Na zgornjem avtokorelogramu in parcialnem avtokorelogramu vidimo, da z diferenciranjem poskrbimo, da niso več prisotni statistično značilni koeficienti avtokorelacije, kar pomeni, da v modelu ni statistično značilne korelacije med členi časovne vrste.

1.5.5 Izbira ustreznega modela

V model sva najprej vključila diferenciranje prvega reda. Gre za ARIMA model s koeficienti $p = 0$, $d = 1$ in $q = 0$

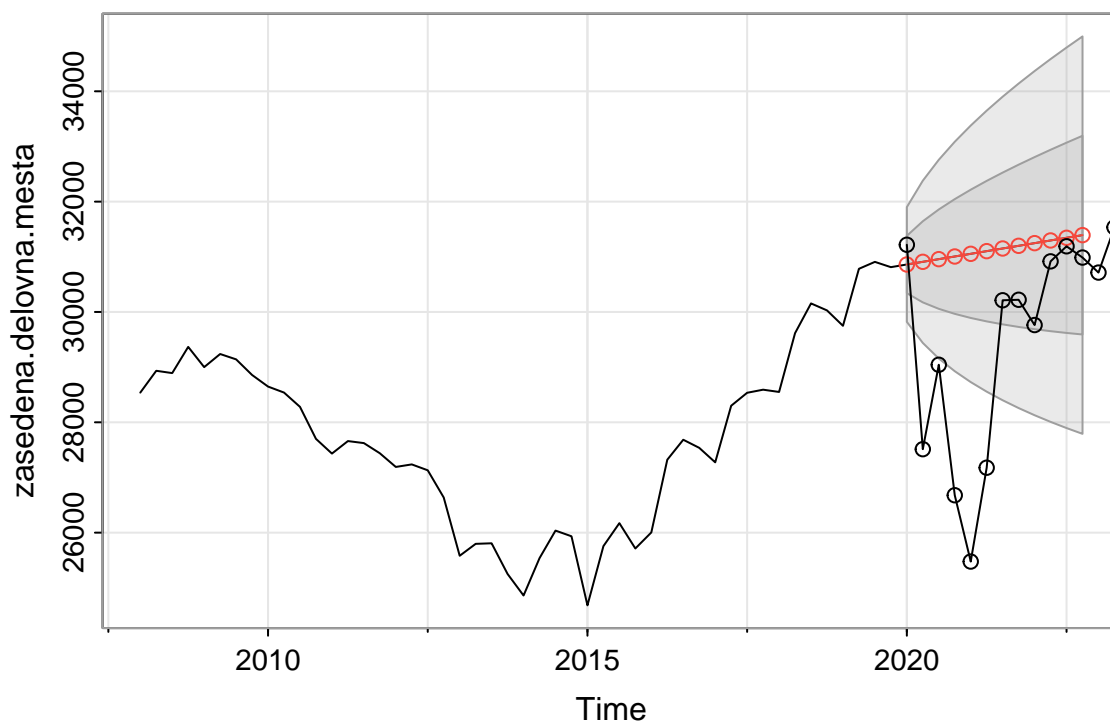


Slika 56: Ostanke modela, avtokorelogram in p-vrednosti Ljung-Box testa.

Koeficienti avtokorelacije so znotraj 95% intervala zaupanja, torej so statistično neznačilni, kar pomeni, da v modelu ni statistično značilne korelacije med ostanki (preverili smo tudi z Ljung-Box testom, kjer nismo mogli zavrniti ničelne domneve, da v tej časovni vrsti ostankov ni več avtokorelacije). Na podlagi tega lahko predpostavljamo, da ostanki so beli šum.

1.5.6 Napoved

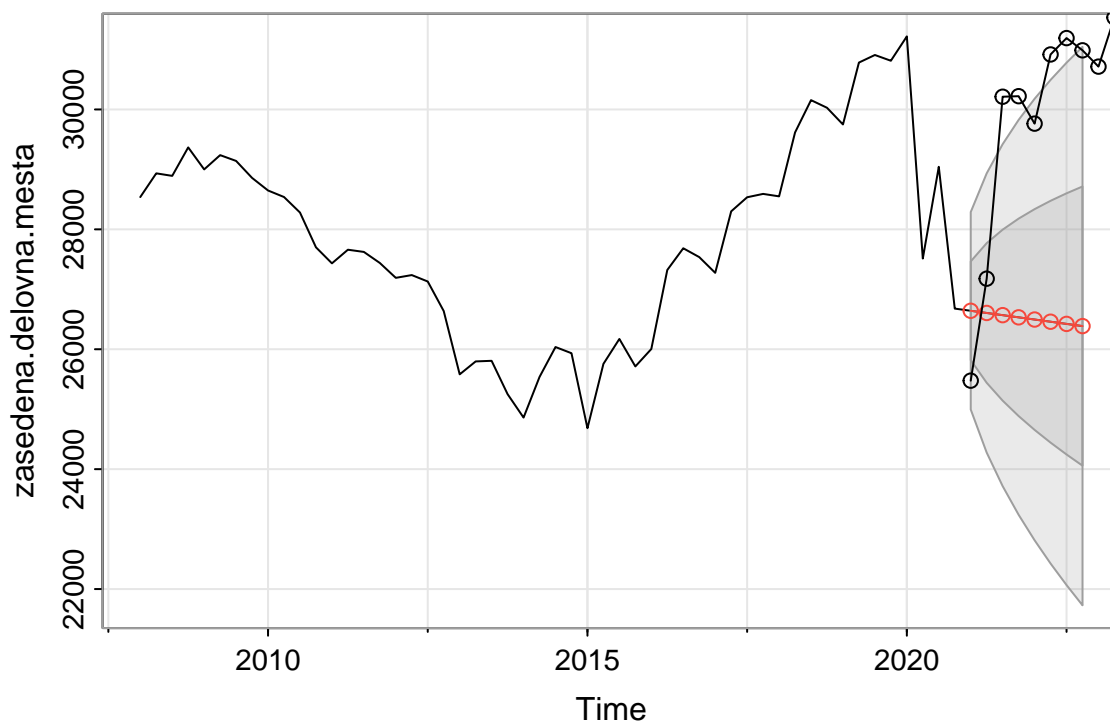
Sedaj bova poskušala z modelom še napovedati število zaposlenih delovnih mest za nekaj let vnaprej. Ker naju zanima kako dober je najin model, na začetku podatke razdeliva na testne in učne. Najbolj naju zanima kako bo napovedoval po letu 2020, če tega leta nima v učnih podatkih in kako napoveduje, če ga vključiva v učne podatke in napovedujeva od leta 2021 naprej.



Slika 57: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

Jasno je, da napoved ni najboljša, saj ostanki vrste predstavljajo beli šum in pričakovano že na začetku poglavja, da model ne zna napovedati padca v letu 2020, saj je naučen po naraščajočem trendu, kar napovedi tudi prikazujejo. Ker so slabe napovedi, so temu primerni tudi izjemno široki 95% intervali zaupanja.

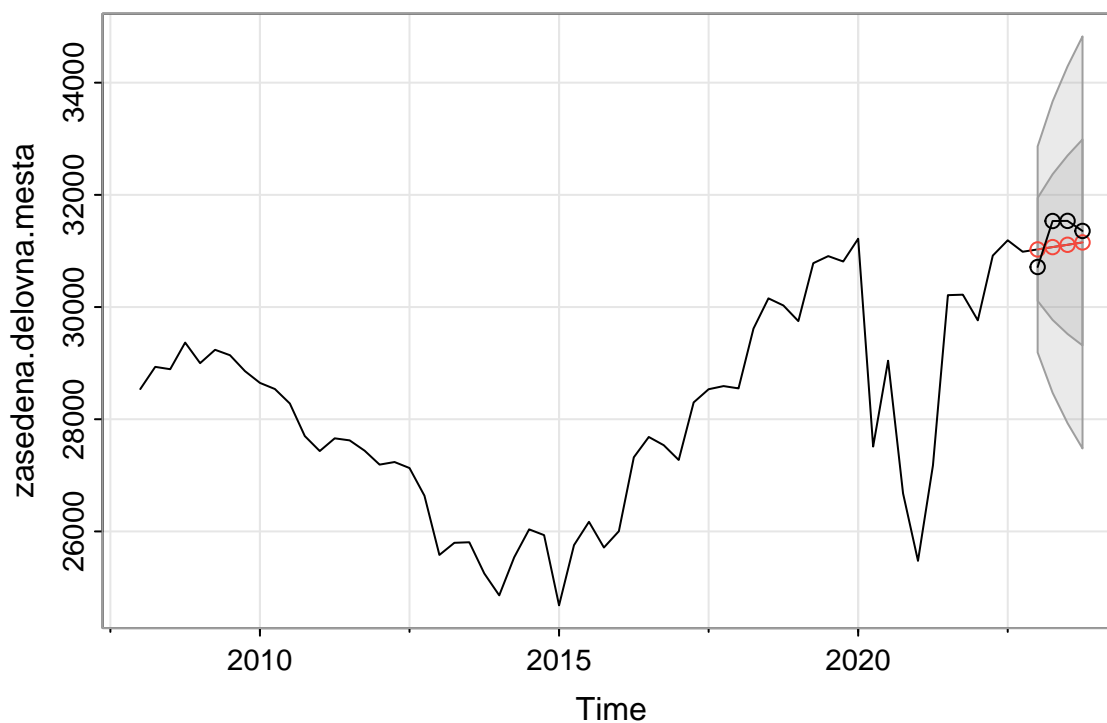
Poglejmo si sedaj še, kakšne so napovedi, če v učno množico vključimo podatke iz leta 2020.



Slika 58: Časovna vrsta in napovedi na podlagi modela od leta 2021 naprej.

Pričakovano tudi sedaj napovedi niso ravno informativne, model je prevzel trend zadnjega leta in ga je nadaljeval tudi naprej.

Sedaj pa napovejmo samo za zadnje leto, leto 2023, da bo časovna vrsta imela v učnih podatkih in padec leta 2020 ter potem ponovno rast.



Slika 59: Časovna vrsta in napovedi na podlagi modela od leta 2023 naprej.

Napovedi tudi tukaj niso pravilne in ne ravno najbolj informativne, najboljše zadane samo za en trenutek naprej, 95% intervali zaupanja pa so še vedno izjemno široki.

Ker je časovna vrsta tako zelo razgibana in ima veliko variabilnost ravno v zadnjih letih, težko uporabljamo ARIMA oz. SARIMA modele za napovedovanje prihodnjih vrednosti.

2 Analiza prostorskega procesa

2.1 Predstavitev podatkov

Podatki so bili pridobljeni kot meritev več različnih spremenljivk na izbrani njivi v Šempasu. Podatki tal so pridobljeni 18. 7. 2023 iz globine tal 20 cm, izmerjeni parametri pa so bili:

- *swc* je gravimetrična vsebnost vode v tleh (vol. %)
- *gostota* je gostota tal (g/cm³)
- *pF2* je vsebnost vode pri matričnem potencialu vode pF 2.0
- *pF4.2* je vsebnost vode pri matričnem potencialu vode pF 4.2
- *aw* je rastlinam razpoložljiva voda v tleh (available water) (pF 2.0 - pF 4.2)

Podrobneje bova obravnala vegetacijski indeks *NDRE* v mesecu juliju in zgoraj napisani izmerjeni parametri lahko, poleg lokacije, tudi vplivajo na vegetacijski indeks.

Na spodnji sliki si lahko ogledamo še načrt meritev.



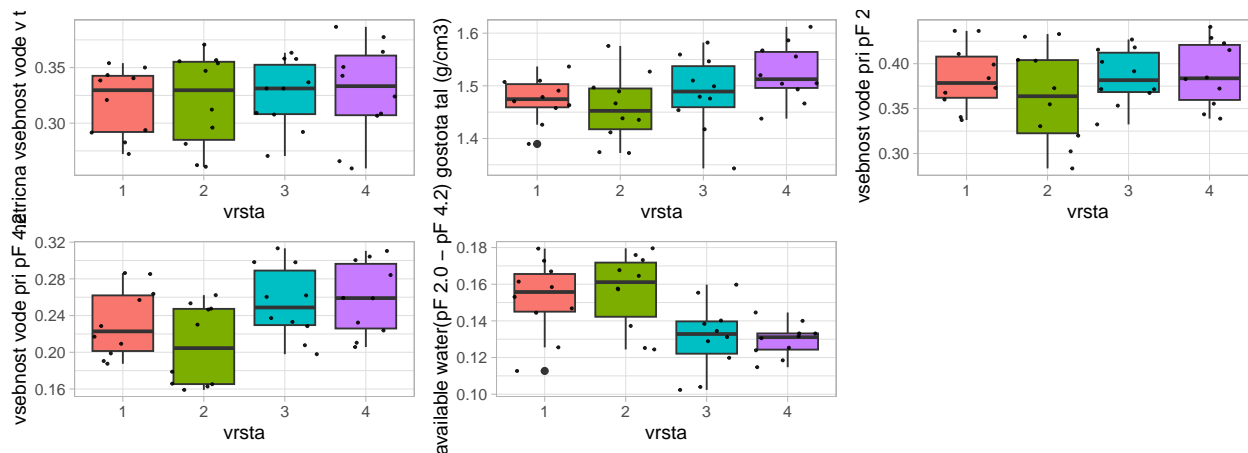
Vegetacijski indeks NDRE

Vegetacijski indeks NDRE (Normalized Difference Red Edge) je spektralni indeks, ki se uporablja v daljinskem zaznavanju za merjenje vsebnosti klorofila v rastlinah. Pogosto se uporablja v preciznem kmetijstvu za

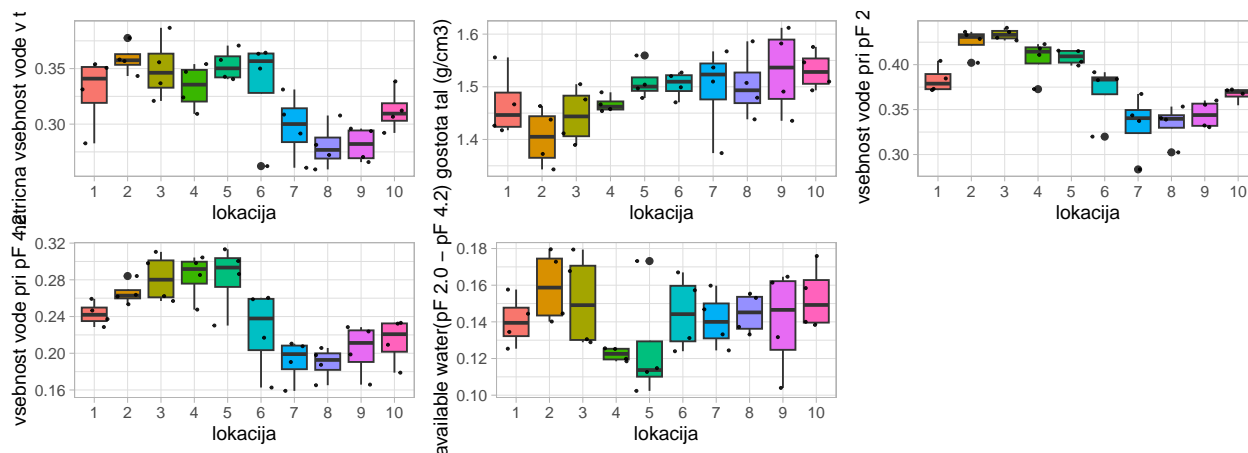
optimizacijo pridelkov, spremljanje zdravja rastlin, določanje potreb po gnojilih in drugih kmetijskih praks. Uporaba tega indeksa omogoča bolj ciljno usmerjene in učinkovite kmetijske posege, kar lahko vodi k večji produktivnosti in trajnosti.

2.1.1 Grafični prikaz parametrov

Najprej si pogledajmo s kakšnimi podatki imamo opravka v našem podatkovnem okvirju.

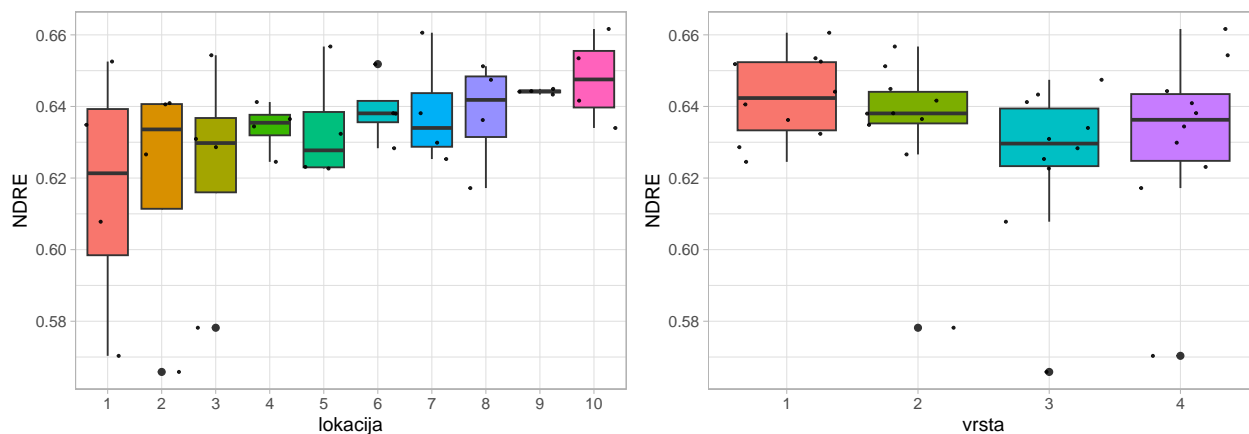


Slika 60: Porazdelitev parametrov glede na vrsto.



Slika 61: Porazdelitev parametrov glede na lokacijo.

Glede na vrsto in lokacijo vidimo, da imajo parametri različne variabilnosti, prav tako je razpon vrednosti, še posebej glede na lokacijo, zelo različen.

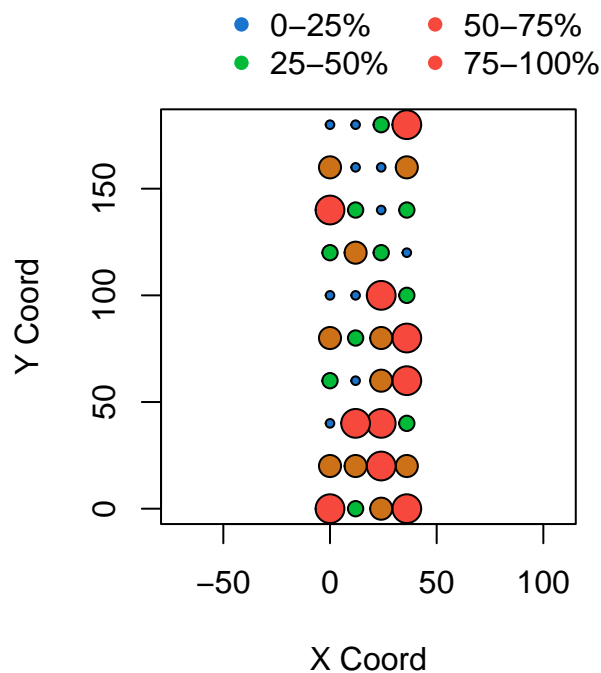


Slika 62: Porazdelitev spremenljivke NDRE za mesec julij glede na lokacijo in vrsto.

Tudi porazdelitve indeksa NDRE se glede na vrsto in lokacijo razlikujejo, njihova variabilnost je različna ter pri določenih vrstah in lokacijah opazimo osamelce ter asimetričnost.

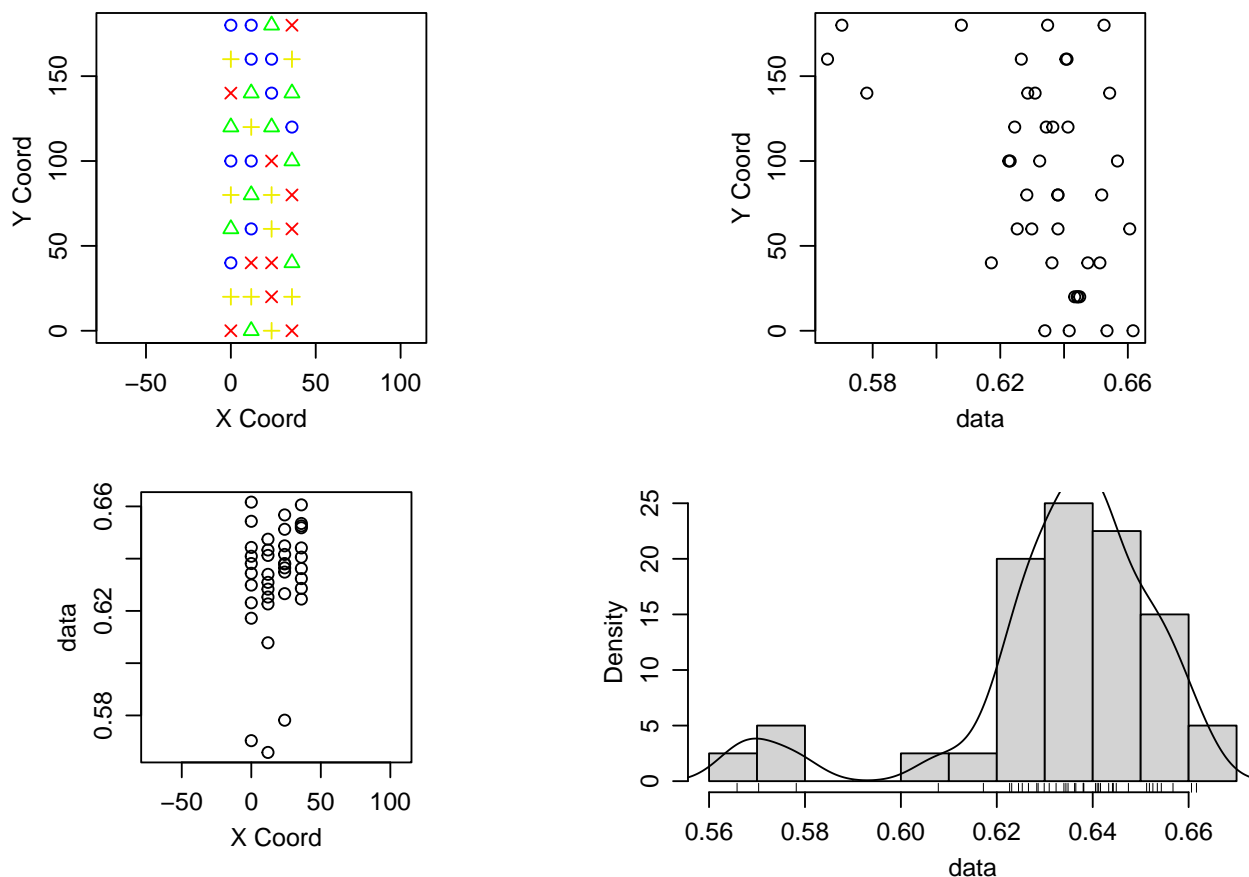
2.2 Grafični prikaz

Na spodnjem grafu lahko vidimo lokacije meritev, na katerem sta ploščina in barva krogcev sorazmerna z vrednostjo spremenljivke *NDRE_jul*, podatki so razdeljeni v štiri razrede, ki so določeni s kvantili.



Slika 63: Grafični prikaz lokacij in relativne vrednosti spremenljivke 'NDRE'.

Poglejmo si še nekoliko podrobnejši prikaz, kjer nas bosta zanimala predvsem razsevana grafikona *vrsta* in *NDRE_jul*, ter *lokacija* in *NDRE_jul*.



Slika 64: Predstavitev realizacije prostorskega procesa 'NDRE'.

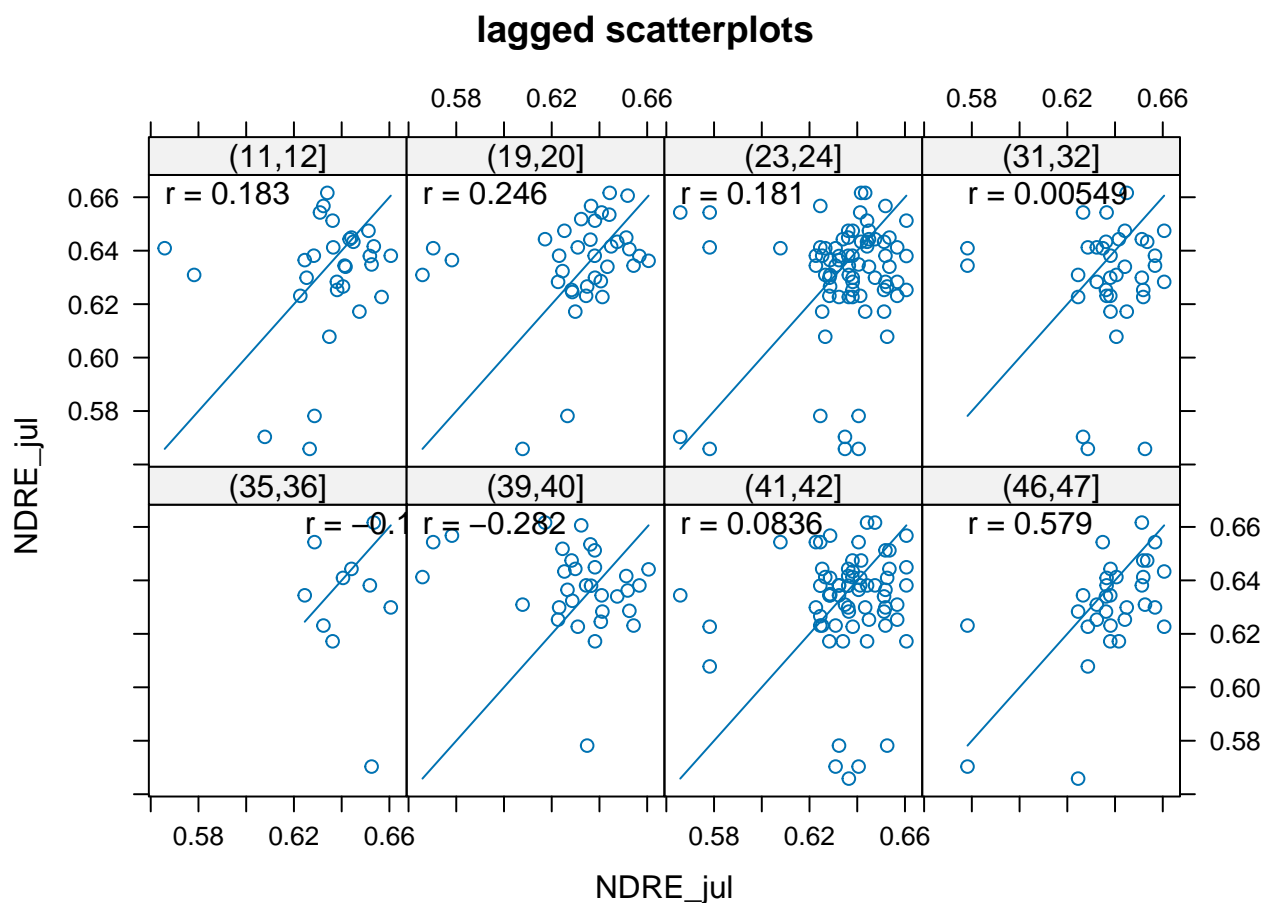
Iz razsevnih grafikonov ni videti odvisnosti spremenljivke *NDRE_jul* od koordinat lokacije. Iz prvega grafa in pa zgornjega levega grafa lahko opazimo nek vzorec, kar nam nakazuje na prisotnost prostorske korelacije.

Torej iz zgornjega desnega grafa (vrednosti v odvisnosti od *lokacija*) in spodnjega levega grafa (vrednosti v odvisnosti od *vrsta*) ni možno razbrati prirobnosti trenda in nekonstantne varinace, zato ne bova naredila podrobnejše analize morebitne nestacionarnosti.

2.3 Analiza prostorske korelacije

2.3.1 Razsevni grafikon

Ker naju natančneje zanima izražanje prostorske korelacije, si nariševa razsevne grafikone, na katerih so prikazani pari podatkov na različnih prostorskih odlogih. Črta nam prikazuje idealno povezanost z vrednostjo korelacijskega koeficienta 1.

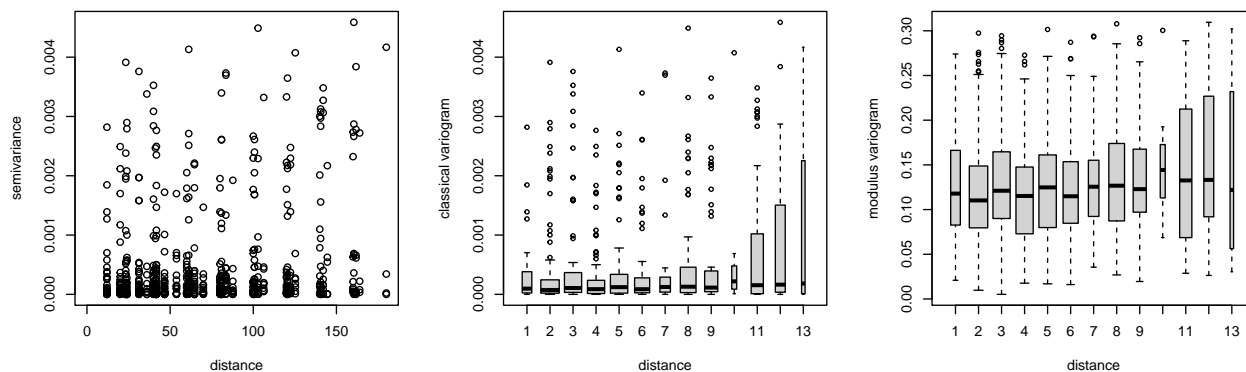


Slika 65: Prostorska korelacija za 'NDRE(jul)'.

Za pare podatkov o **NDRE_jul** glede na razrede oddaljenosti lahko vidimo, da se z oddaljenostjo prostorska korelacija zmanjšuje (Pearsonov koeficient korelacije). Lahko bi torej reki, da na zelo kratkih razdaljah imamo majhno povezanost, ki na daljših razdaljah še pada.

2.3.2 Oblak semivariograma in vzorčni semivariogram

Poglejmo analizo prostorske korelacije še s pomočjo oblaka semivariograma in vzorčnega semivariograma.

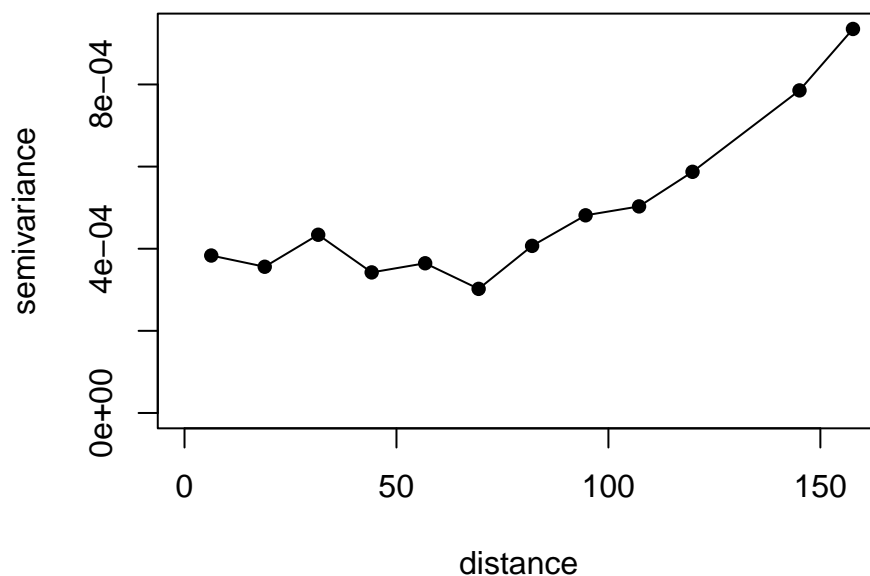


Slika 66: Oblak semivariograma za 'NDRE(jul)'(levo), okvirji z ročaji za vrednosti klasičnega variaograma po razdaljah med lokacijami(sredina) in za vrednosti robustnega vzorčnega variaograma(desno).

Na skrajno levem zgornjem grafu vidimo, da vrednosti *NDRE_jul* predstavljajo meritve v pravilni mreži točk, saj oblak variograma prikazuje diskretne vrednosti razdalj med lokcijami in so točke narisane v črtah. Porazdelitve vrednosti vzorčnega variograma pa so asimetrične v vsakem razredu (razredi predstavljajo razdalje), prav tako pa imamo nekaj osamelcev. Na zadnjem, robustnem vzorčnem variogramu, kjer na nek način “posekamo” vpliv vrednosti, ki zelo odstopajo, vidimo, da so porazdelitve glede na razrede sedaj bistveno drugačne in nekateri osamelci izginejo.

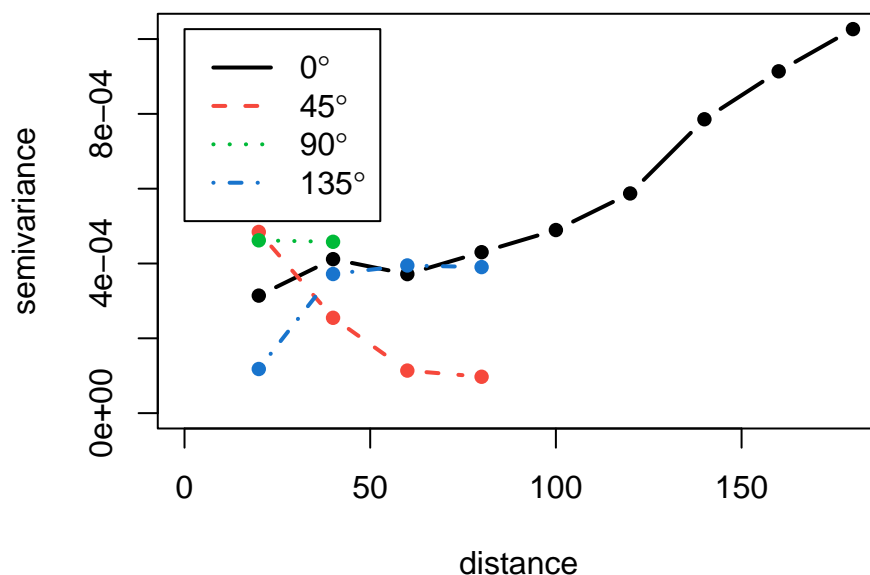
Na spodnjem levem vzorčnem semivariogram s povprečji za posamezne razrede razdalje lahko iz razporeditve točk razberemo, da so pri manjših razdaljah vzorčne variogramske vrednosti manjše, z naraščajočo razdaljo se povečujejo in se z nadaljnjim večanjem razdalje ne ustalijo ravno oz. ne izgleda, kot da bi se približevale določenemu pragu, torej kaže na prisotnost prostorskega trenda v podatkih.

Vzorčni variogram za ostanke prostorskega trenda (desno) pokaže, da je prostorska korelacija med ostanki precej šibkejša.



Slika 67: Vzorčni semivariogram za 'NDRE(jul)'.

Do sedaj sva proučevala prostorsko korelacijo ne glede na smer v prostoru, torej sva predpostavila izotropni prostorski proces. Zato narišimo še usmerjene semivariograme za določene smeri v prostoru in s tem preverimo ali je predpostavka izotropnosti upravičena.

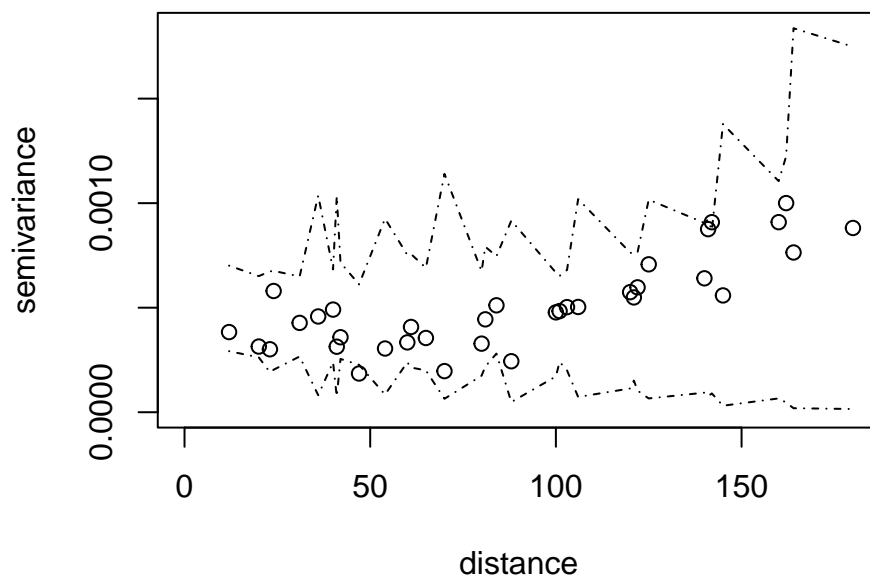


Slika 68: Vzorčni variogrami za 'NDRE(jul)' za štiri glavne smeri.

Vzorčni variogrami za *NDRE_jul* za štiri glavne smeri izraža izotropijo, saj že na krajših razdaljah vidimo razlike v vzorčnih variogramih za različne smeri, ki se potem sicer zelo hitro zmanjša (*pri razdalji 50*). Tudi začetne razlike so dokaj majhne. Torej bi glede na graf morda bilo potrebno modeliranje prostorske korelacije na kratke razdalje.

2.3.3 Ovojnice za vzorčni semivariogram

Da bi dobila širšo sliko o prostorski povezanosti procesa, pri raziskovanju prostorske povezanosti uporabiva Monte Carlo simulacije na podlagi te realizacij prostorskega procesa. Če padejo vse vrednosti osnovnega vzorčnega variograma znotraj ovojnice, ki je določena z robnima variogramoma permutiranih vrednosti, potem podatki ne nakazujejo prostorske povezanosti. Predvsem naju zanima nekaj prvih točk, ki odražajo korelacijo na manjših razdaljah.



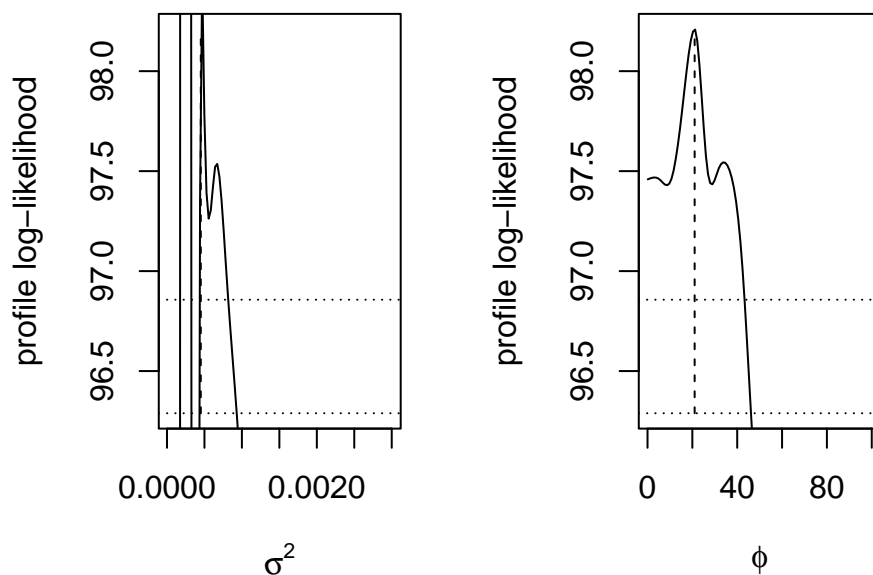
Slika 69: Monte Carlo ovojnice semivariograma.

Vidimo, da ni izrazite koreliranosti oz. podatki niso prostorsko korelirani.

2.4 Model semivariograma za dane podatke

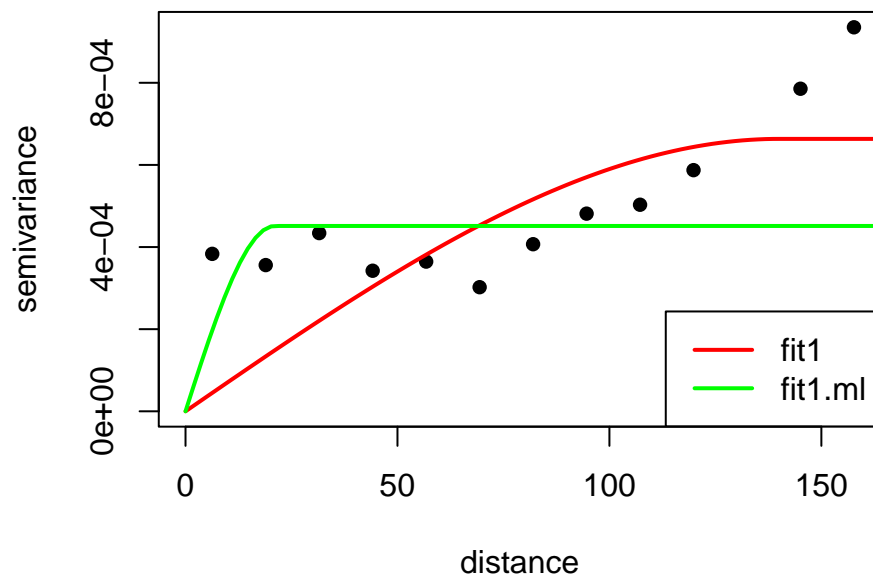
2.4.1 Sferični model variograma brez zlatega zrna

Najprej sva parametre semivariograma ocenila z metodo Maximum Likelihood (ML), ki za začetne vrednosti vzame ocene dobljene s funkcijo `variofit`.



Slika 70: Profile-log-likelihood za parametra sferičnega modela semivariograma brez zlatega zrna kaže na napačno izbiro modela prostorske korelacije.

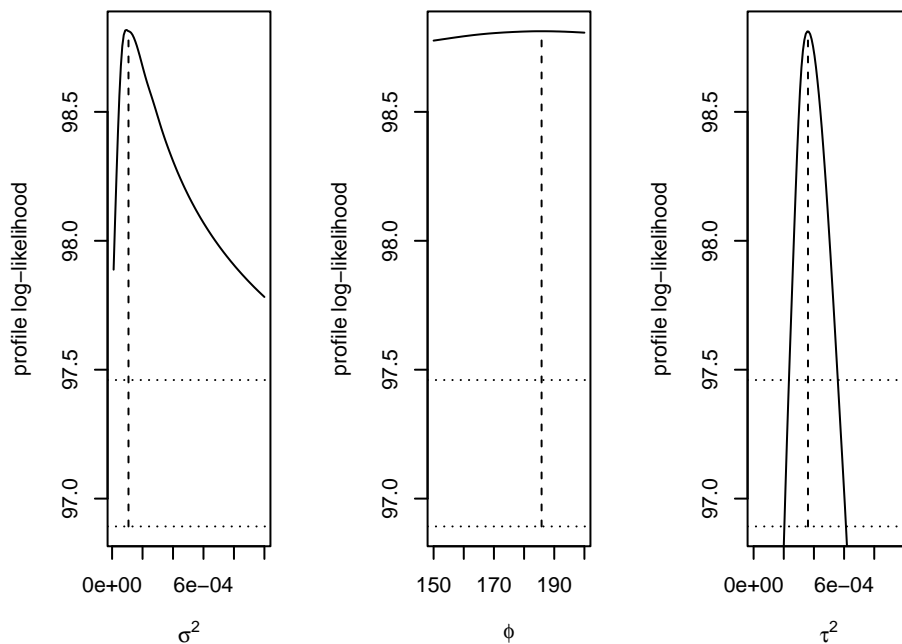
Izgleda kot da po metodi največjega verjetja z začetnimi vrednostmi praga in razmika variograma **fit1** dobimo optimalne rešitve za parametre semivariograma.



Slika 71: Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.

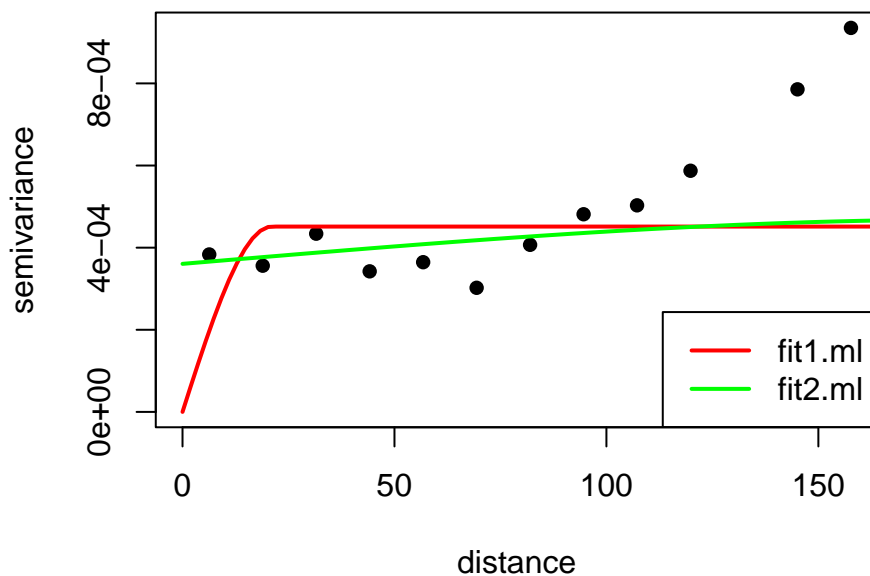
Vidimo, da z modelom *fit1.ml*, kjer smo kot metodo uporabili Maximum likelihood, nekoliko bolje opisali vrednosti semivarianc za manjše razdalje. V celoti pa bi lahko rekli da z modelom *fit1* bolje opišemo vrednosti semivarianc za razdalje od 1 do 10.

V naslednjem koraku pri uporabi funkcije *likfit* dopustiva, da dobimo tudi oceno za zlato zrno.



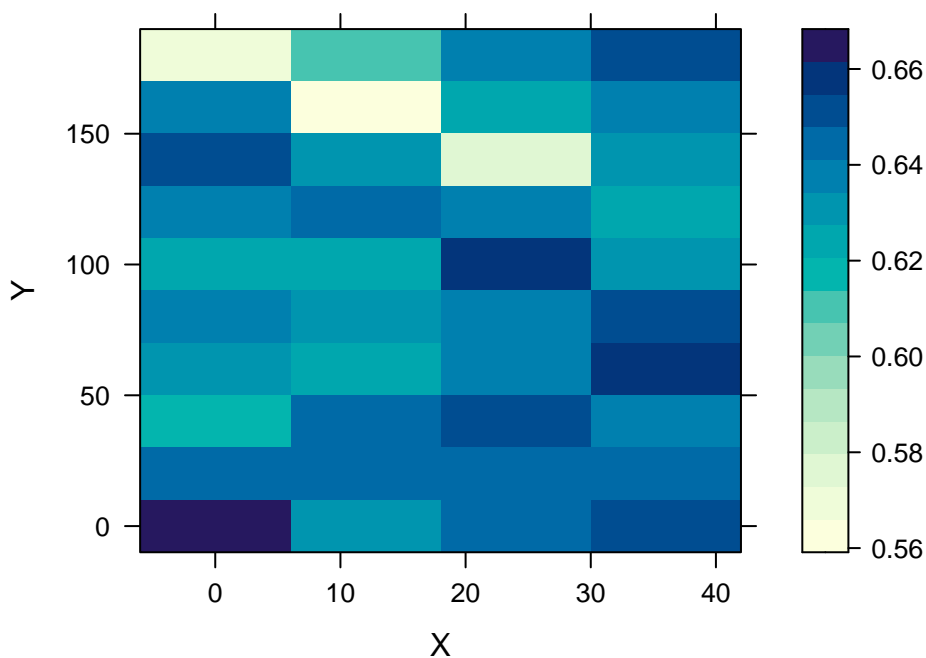
Slika 72: Profile-log-likelihood za parametre sferičnega modela semivariograma z zlatim zrnem.

Iz grafov lahko vidimo da po metodi največjega verjetja z začetnimi vrednostmi praga in razmika variograma **fit1** ne dobimo optimalne rešitve za ϕ (razmik-range). Optimalno rešitev dobimo le za σ^2 (prag semivariograma-sill) in τ (varianca zlatega zrna).

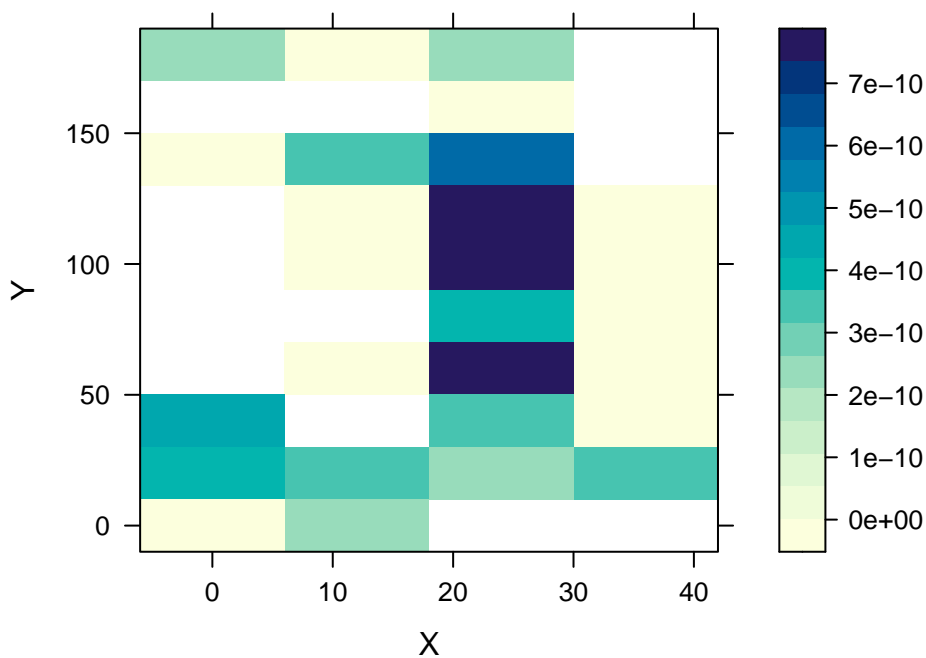


Slika 73: Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.

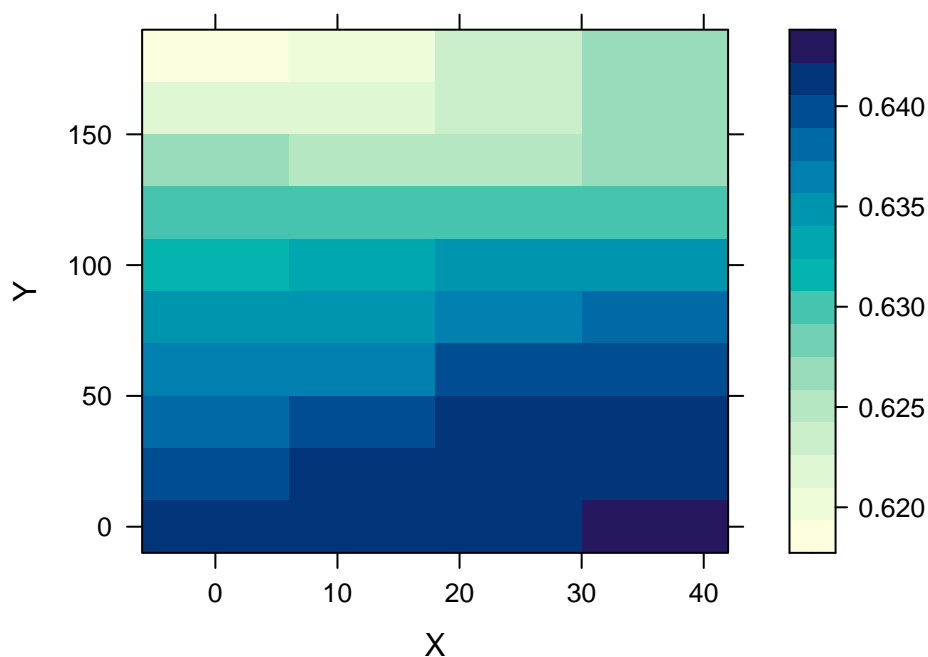
Če med seboj primerjamo modela, pri katerem je vrednost zlatega zrna fiksna (rdeča) in model pri katerem ocenjujemo vrednost zlatega zrna (zelena). Izkaže se, da v primeru, ko ocenjujemo vrednost zlatega zrna bolje opišemo prvo in drugo vrednost semivariance, tretjo pa boljše model s fiksnim zlatim zrnem, torej na zelo kratke razdalje je boljši model pri katerem ocenjujemo vrednost zlatega zrna (zelena, *fit2.ml*).



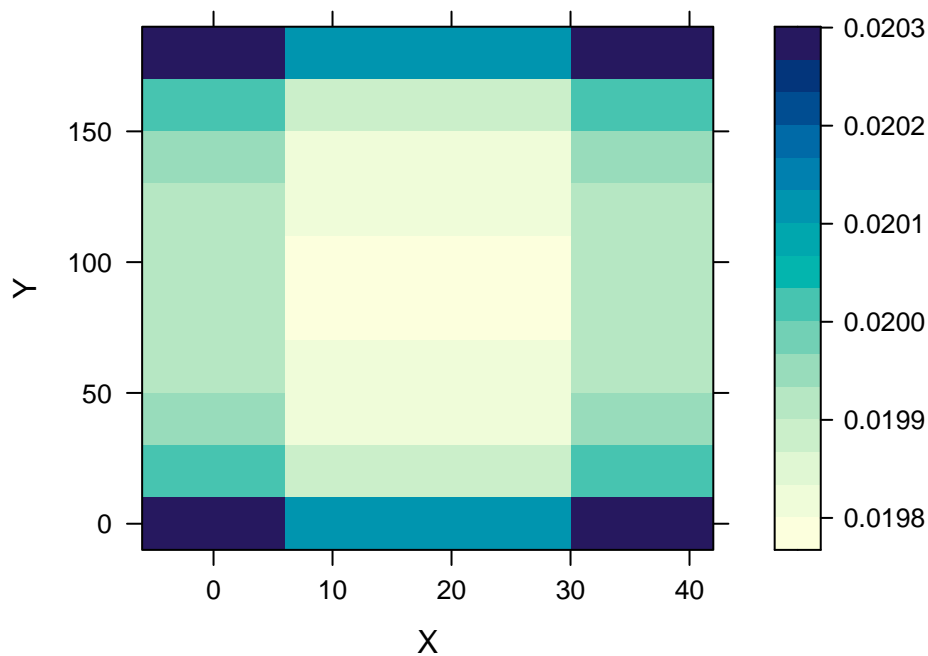
Slika 74: Napovedi NDRE za mesec julij za dano območje - model semivariograma *fit1.ml*(levo).



Slika 75: Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma *fit1.ml*.



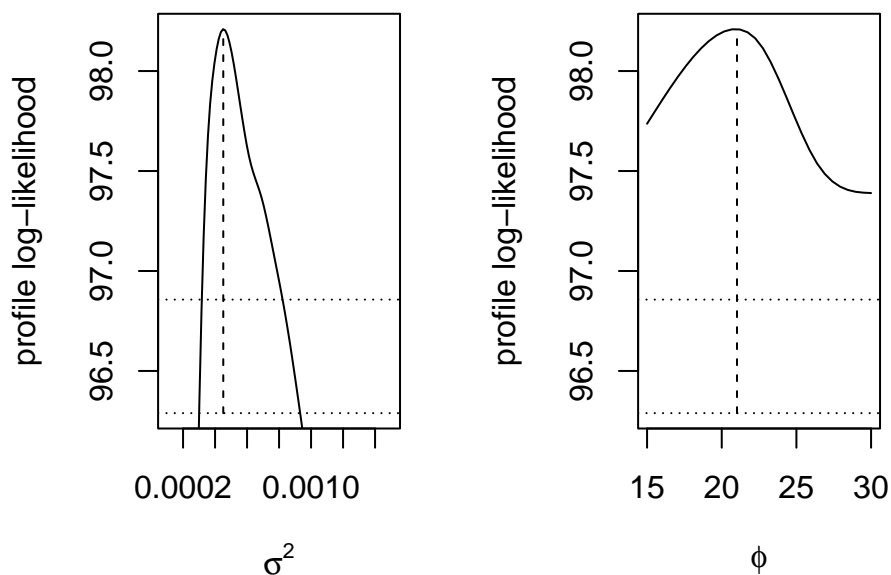
Slika 76: Napovedi NDRE za mesec julij za dano območje - model semivariograma fit2.ml.



Slika 77: Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma fit2.ml.

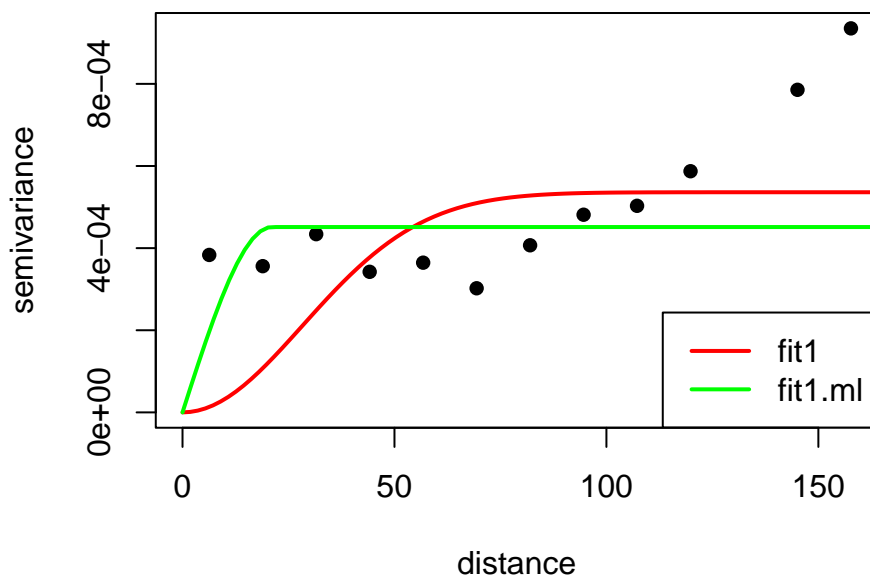
2.4.2 Gaussov model variograma brez zlatega zrna

Poskusila bova modelirati še z uporabo Gaussove kovariančne funkcije. Preverila bova ali z uporabo metode ML dobimo boljše prilaganje vrednostim variograma.



Slika 78: Profile-log-likelihood za parametra Gaussovega modela semivariograma brez zlatega zrna.

Izgleda kot da po metodi največjega verjetja z začetnimi vrednostmi praga in razmika variograma **fit1_g** dobimo optimalne rešitve za parametre semivariograma.



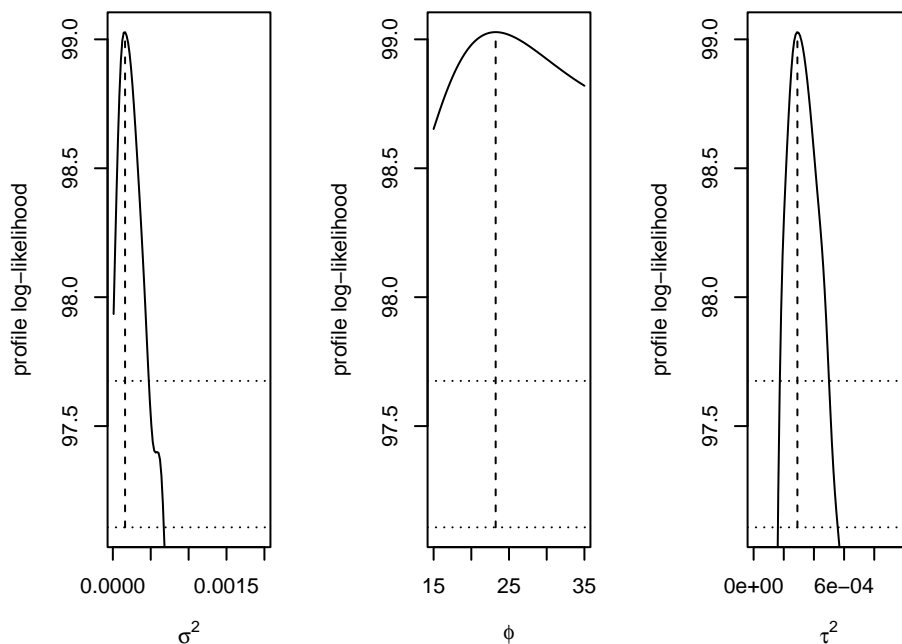
Slika 79: Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.

Ponovno vidimo, da z uporabo metode Maximum likelihood, torej modelom *fit1.ml*, bolje opišemo le prvi dve vrednosti vzorčnega semivariograma. Z modelom *fit1* pa bolje opišemo vrednosti tudi pri večjih razdaljah, ampak pri velikih razdaljah (npr. 130 naprej) pa noben od modelov prav dober.

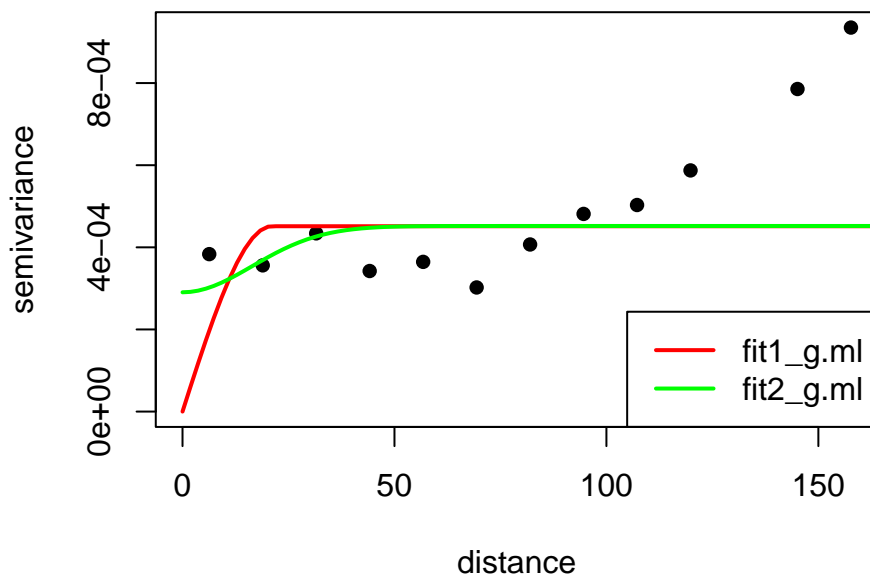
V naslednjem koraku pri uporabi funkcije *likfit* dopustiva, da dobimo tudi oceno za zlato zrno.

Iz spodnjih grafov lahko vidimo da po metodi največjega verjetja z začetnimi vrednostmi praga in razmika variograma **fit1_g** ne dobimo optimalne rešitve za ϕ (razmik-range). Optimalno rešitev dobimo le za σ^2

(prag semivariograma-sill) in τ (varianca zlatega zrna).

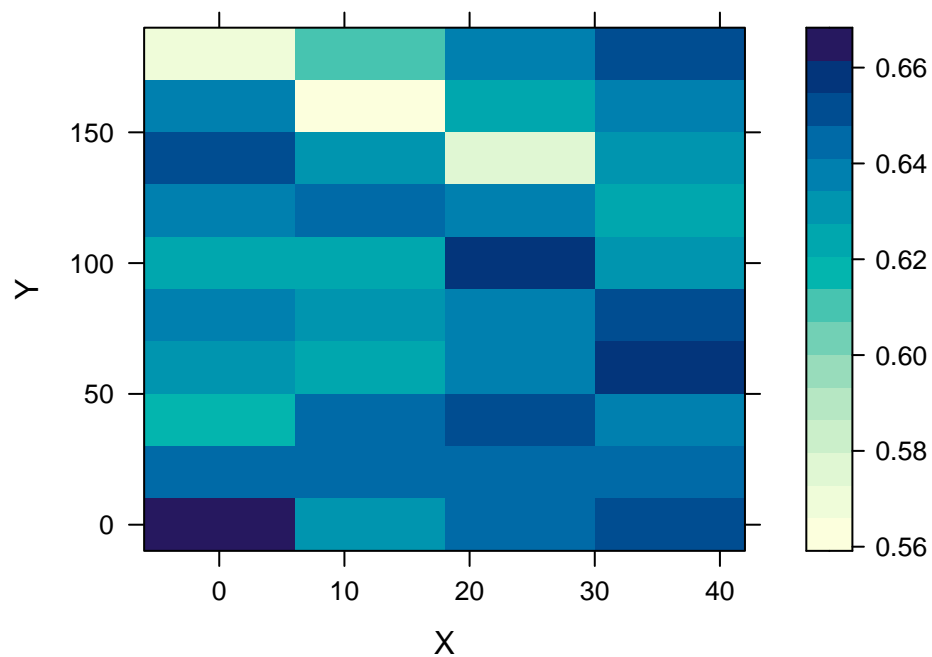


Slika 80: Profile-log-likelihood za parametre Gaussovega modela semivariograma z zlatim zrnom.

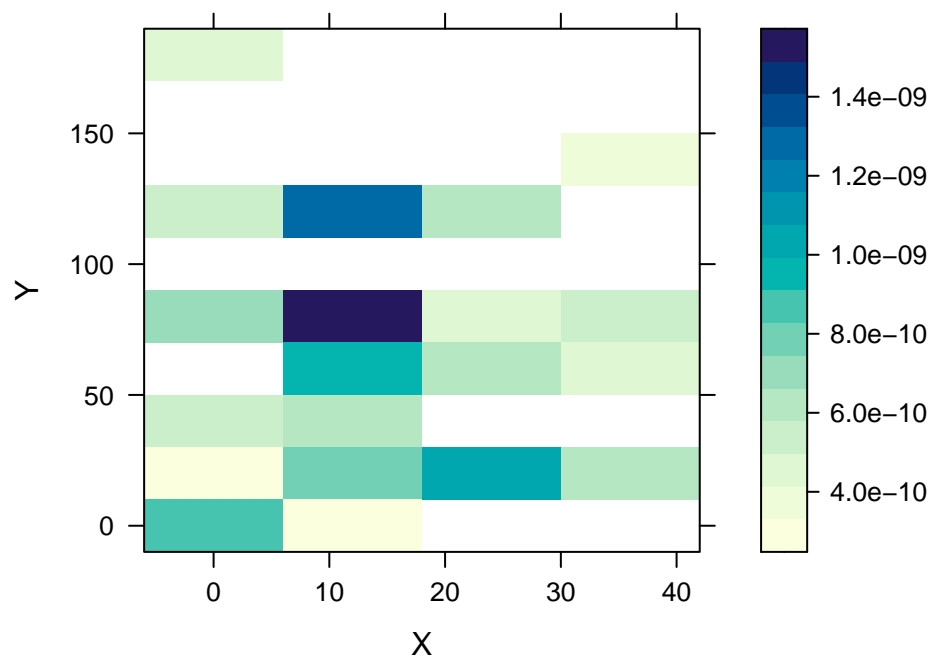


Slika 81: Modela vzorčnega semivariograma za prostorski trend, na podlagi vzorčnega variograma in na podlagi oblaka variograma z ML.

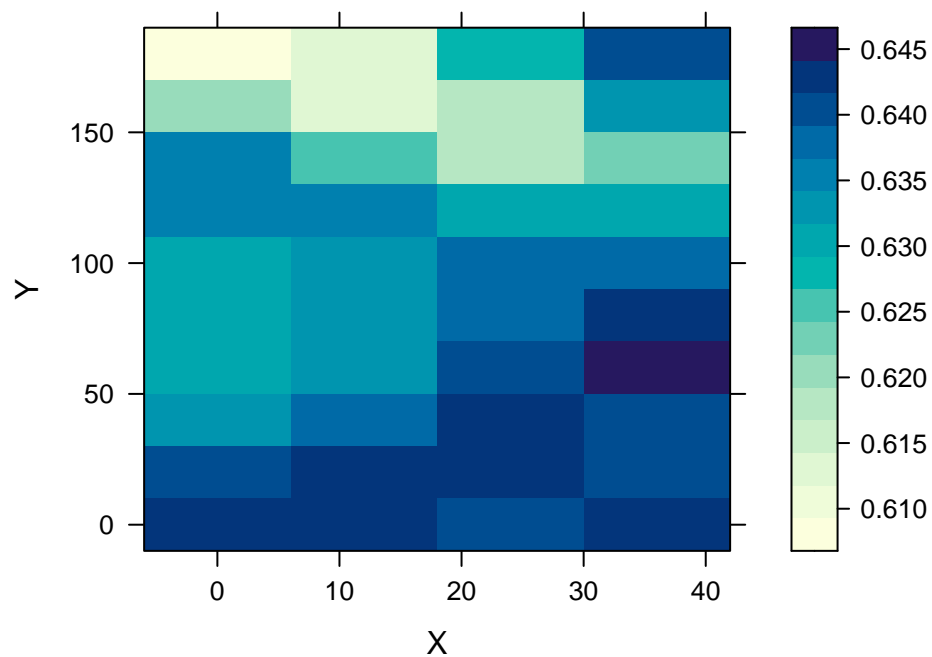
Če med seboj primerjamo modela, pri katerem je vrednost zlatega zrna fiksna (rdeča, *fit1_g.ml*) in model pri katerem ocenjujemo vrednost zlatega zrna (zelena, *fit2_g.ml*). Izkaže se, da v primeru modela, kjer ocenjujemo vrednost zlatega zrna bolje opišemo vrednosti semivarianc za manjše razdalje, kasneje pa sta si modela skoraj enaka za ocenjevanje.



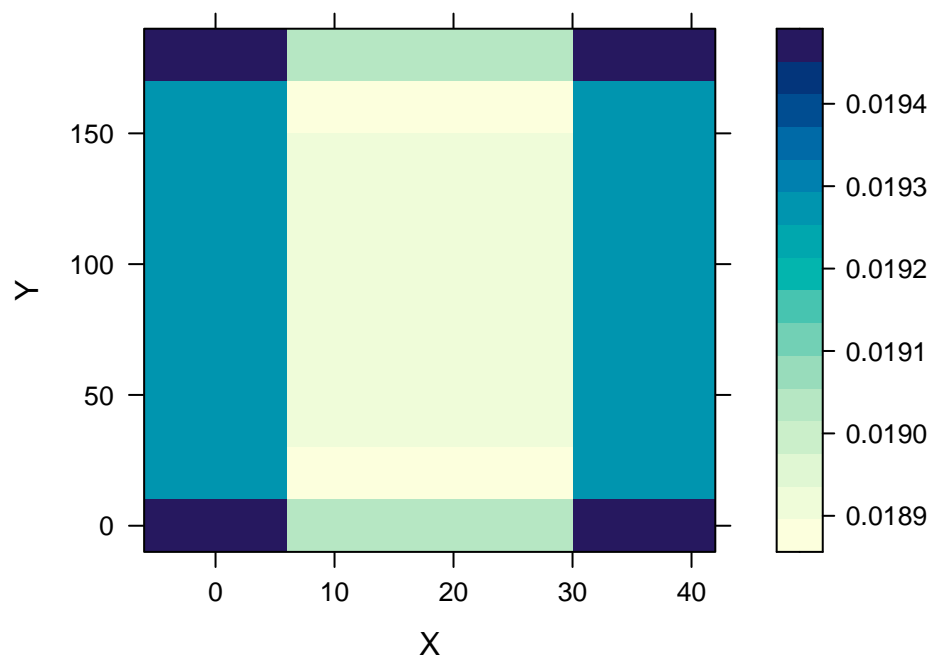
Slika 82: Napovedi NDRE za mesec julij za dano območje - model semivarigrama fit1g.ml.



Slika 83: Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivarigrama fit1g.ml.



Slika 84: Napovedi NDRE za mesec julij za dano območje - model semivariograma fit2g.ml.



Slika 85: Standardna napaka napovedi NDRE za mesec julij za dano območje - model semivariograma fit2g.ml.