

Contents

Uvod	4
Časovna vrsta <i>Gostinske nastanitvene dejavnosti</i>	5
Pregled časovne vrste	5
Transformacija	6
Avtokorelogram in parcialni avtokorelogram	7
Diferenciranje	7
Izbira ustreznega modela	10
Napoved	15
Časovna vrsta <i>Dejavnost strežbe jedi in pijač</i>	18
Pregled časovne vrste	18
Transformacija	19
Avtokorelogram in parcialni avtokorelogram	20
Diferenciranje	20
Izbira ustreznega modela	23
Napoved	28
Časovna vrsta <i>Gostinstvo prosta delovna mesta 1+</i>	31
Pregled časovne vrste	31
Transformacija časovne vrste	32
Avtokorelogram in parcialni avtokorelogram	33
Diferenciranje	34
Izbira ustreznega modela	35
Napoved	36
Časovna vrsta <i>Gostinstvo zasedena delovna mesta 1+</i>	40
Pregled časovne vrste	40
Transformacija časovne vrste	41
Analiza avtokoreliranosti	42
Diferenciranje	42
Izbira ustreznega modela	43
Napoved	44

List of Figures

1	Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na leta.	5
2	Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na mesec.	5
3	Časovna vrsta indeksa prihodka za 'Gostinske nastanitvene dejavnosti'.	6
4	Osnovna in transformirana časovna vrsta 'Gostinske nastanitvene dejavnosti'.	6
5	Avtokorelogram in parcialni avtokorelogram časovne vrste.	7
6	Osnovna in diferencirana časovna vrsta.	7
7	Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.	8
8	Osnovna in diferencirana(sezonskost) časovna vrsta.	8
9	Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.	9
10	Diferenciranje trenda in sezonskosti v časovni vrsti.	9
11	Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.	10
12	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.	11
13	Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.	12

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

List of Tables

1	SARIMA modeli in pripadajoče AIC vrednosti.	14
2	SARIMA modeli in pripadajoče AIC vrednosti.	27

Uvod

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 vzročenje.

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 kvartalih. Frekvenca vzorčenja je približno enakomerna in enaka 4, torej gre za četrtno vzročenje.

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

Časovna vrsta *Gostinske nastanitvene dejavnosti*

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 minimalni trend, ki je sprva negativen, nato pa nekoliko pozitiven 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 januarju in decembru, kar je dokaj logično glede na življenje ljudi.

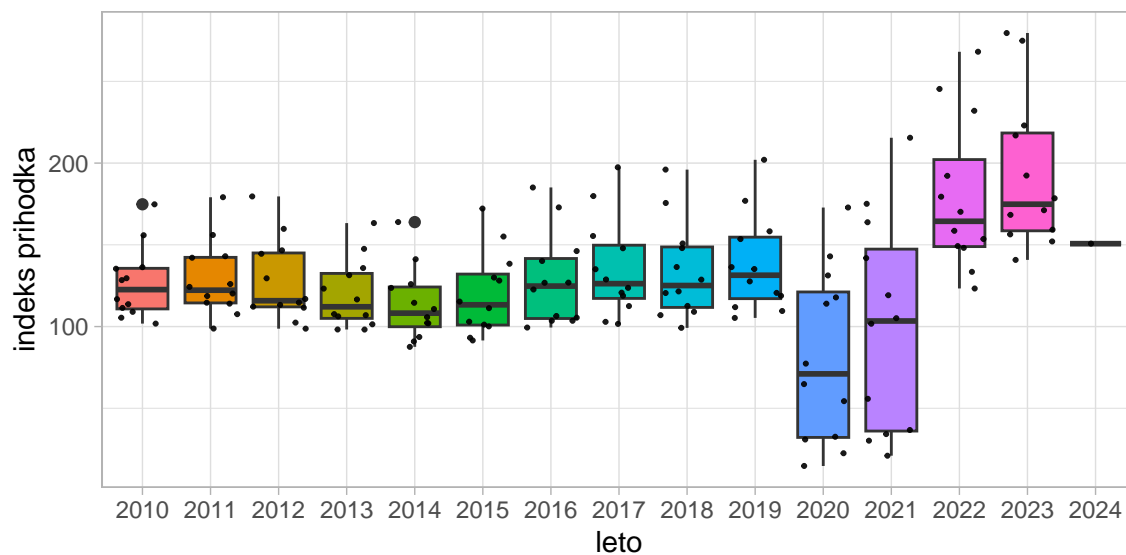


Figure 1: Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na leta.

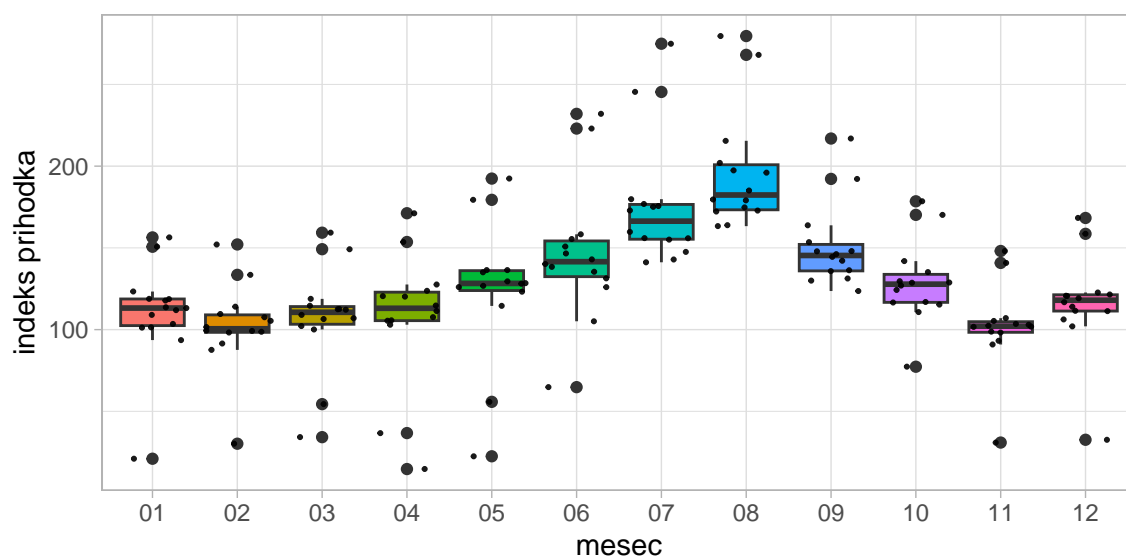


Figure 2: Porazdelitev indeksa prihodka v gostinski nastanitveni dejavnosti glede na mesec.

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

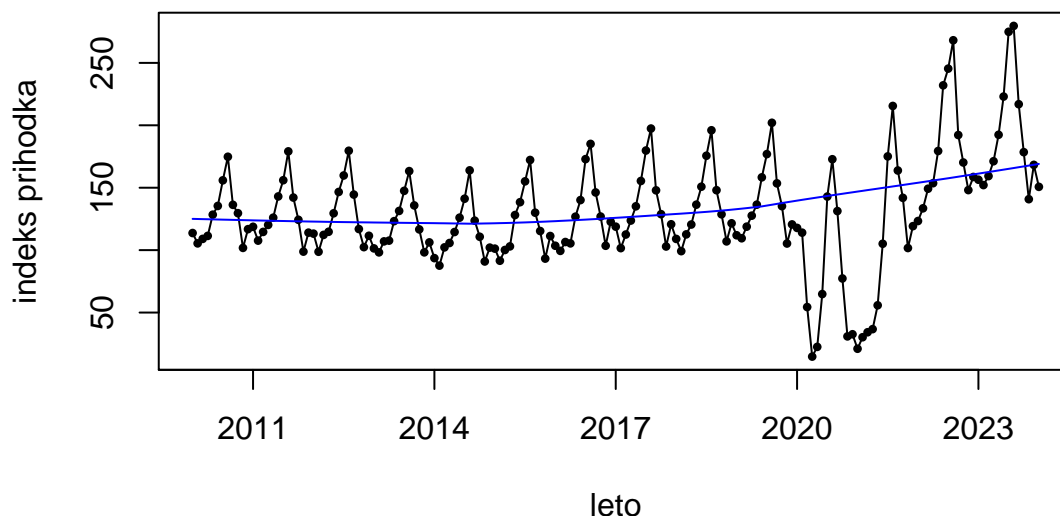


Figure 3: Časovna vrsta indeksa prihodka za 'Gostinske nastanitvene dejavnosti'.

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.

Transformacija

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

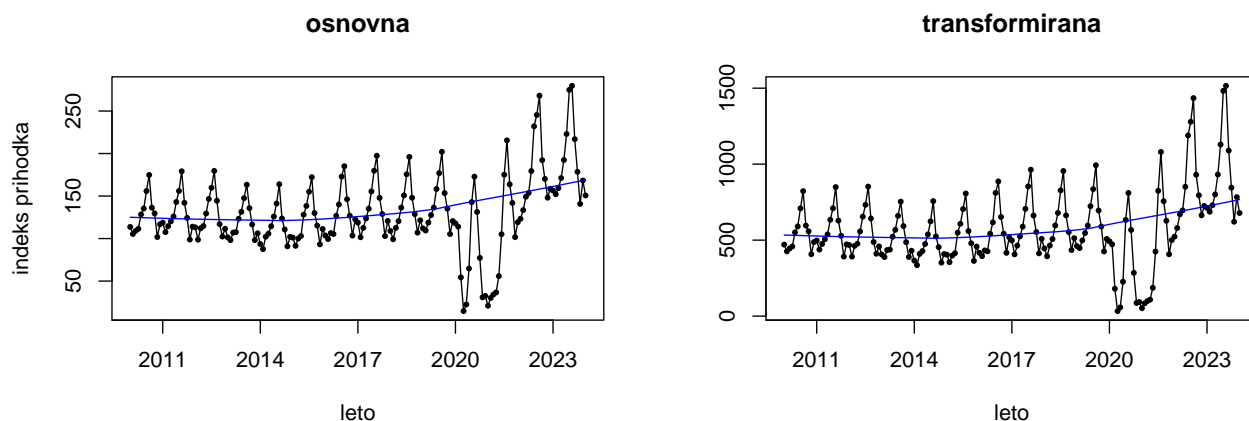


Figure 4: Osnovna in transformirana časovna vrsta 'Gostinske nastanitvene dejavnosti'.

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

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 postvetiti nekaj pozornosti tudi temu.

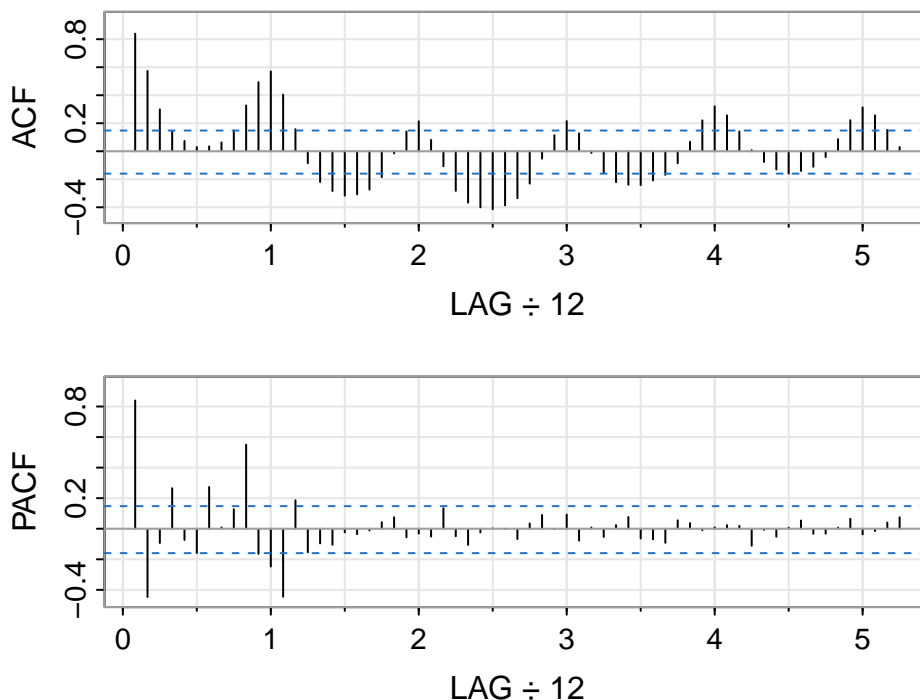


Figure 5: Avtokorelogram in parcialni avtokorelogram časovne vrste.

Diferenciranje

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

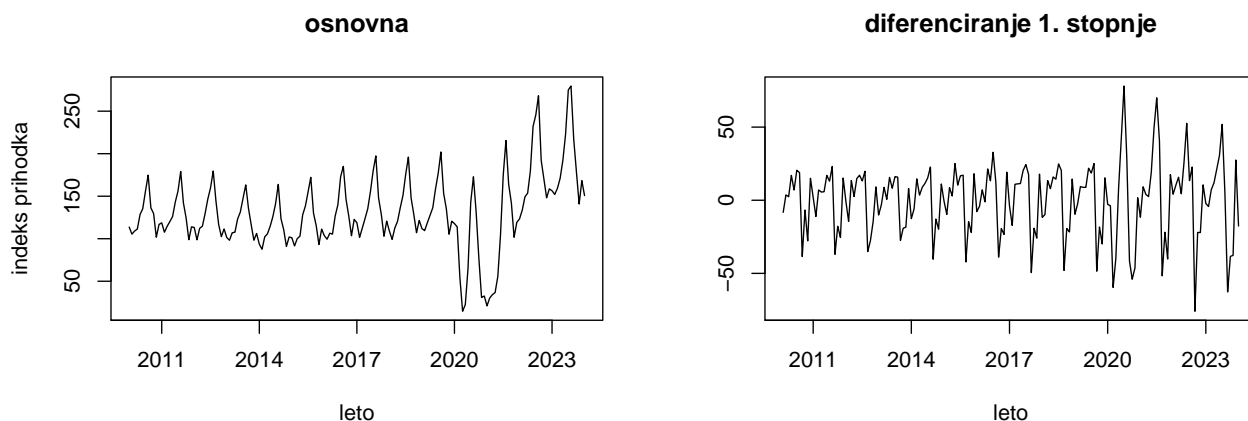


Figure 6: Osnovna in diferencirana časovna vrsta.

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.

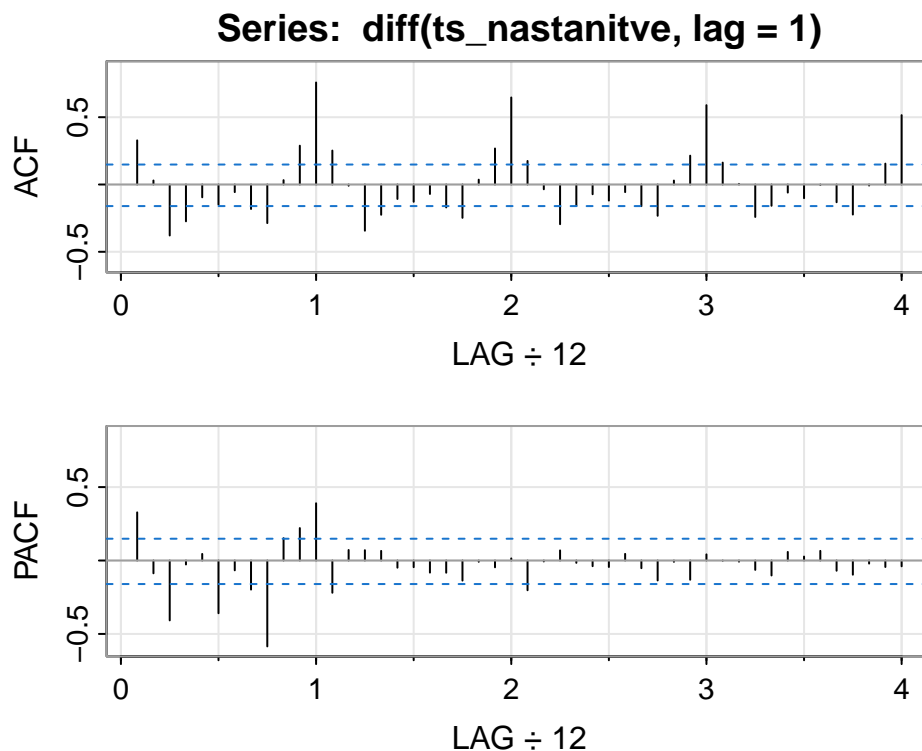


Figure 7: Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.

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.

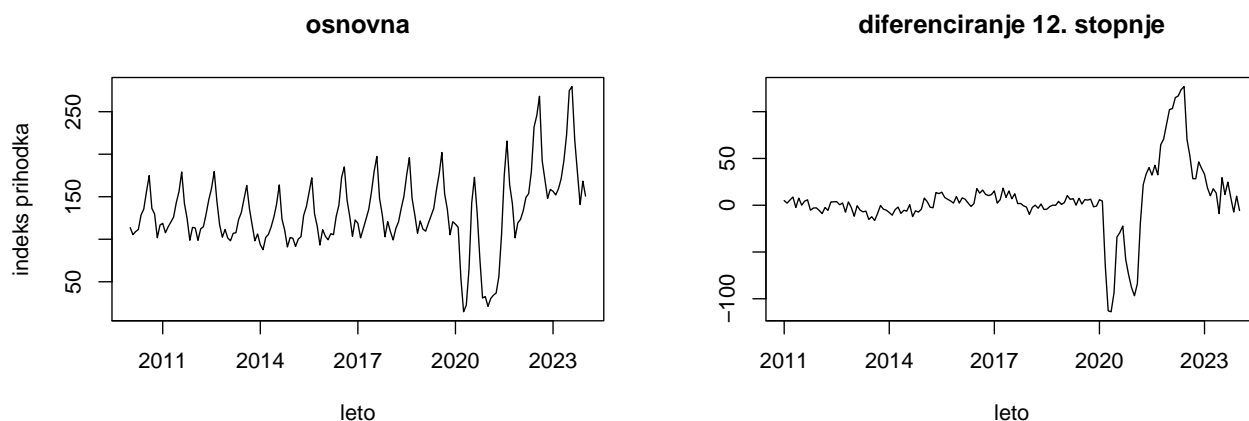


Figure 8: Osnovna in diferencirana(sezonskost) časovna vrsta.

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

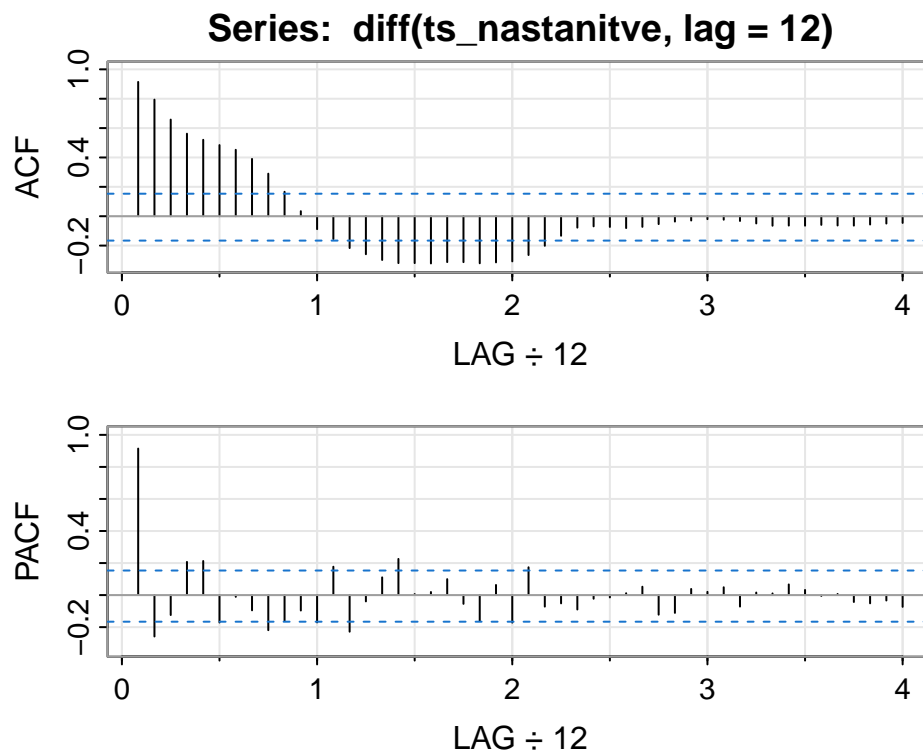


Figure 9: Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.

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.

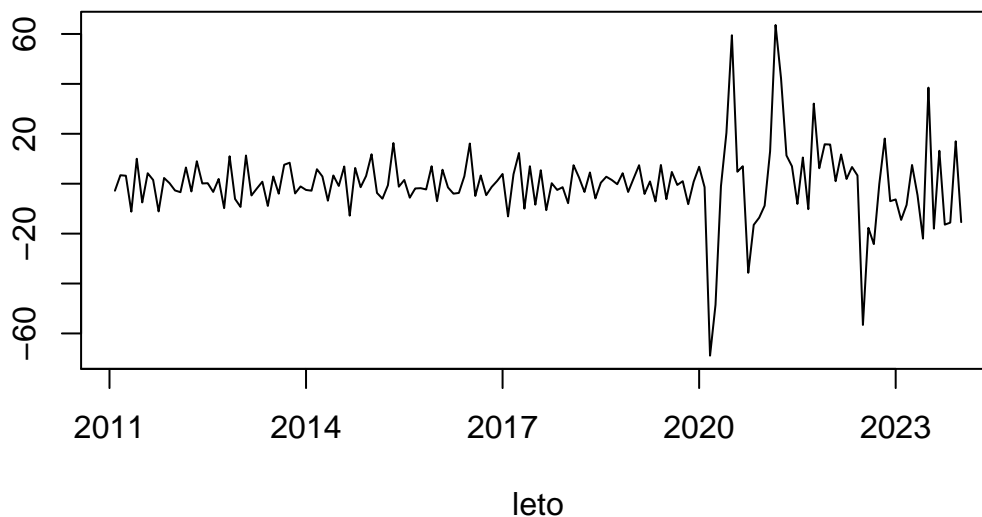


Figure 10: Diferenciranje trenda in sezonskosti v časovni vrsti.

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.

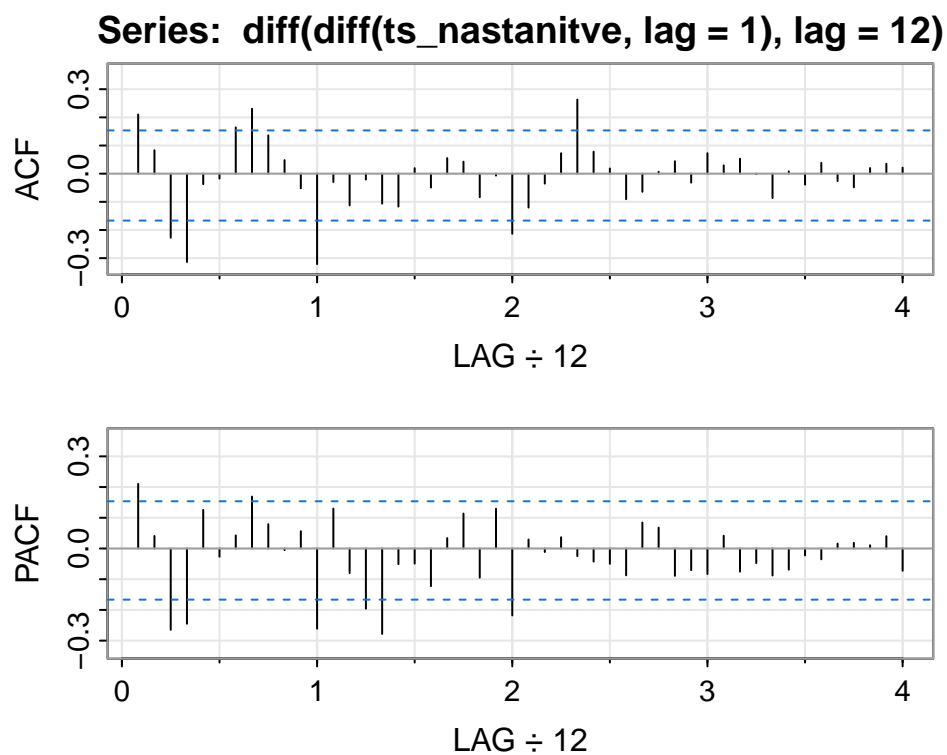


Figure 11: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.

Izbira ustreznega modela

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

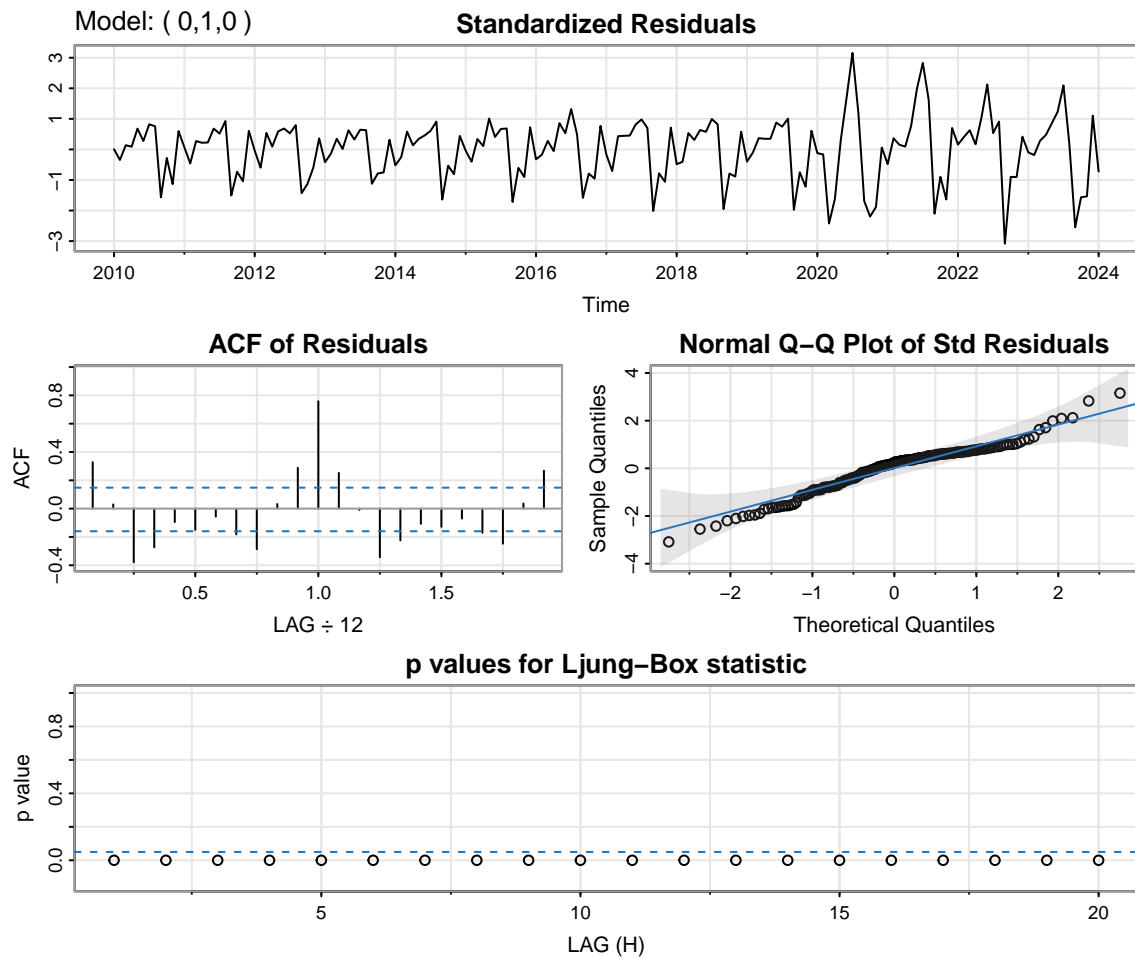


Figure 12: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.

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.

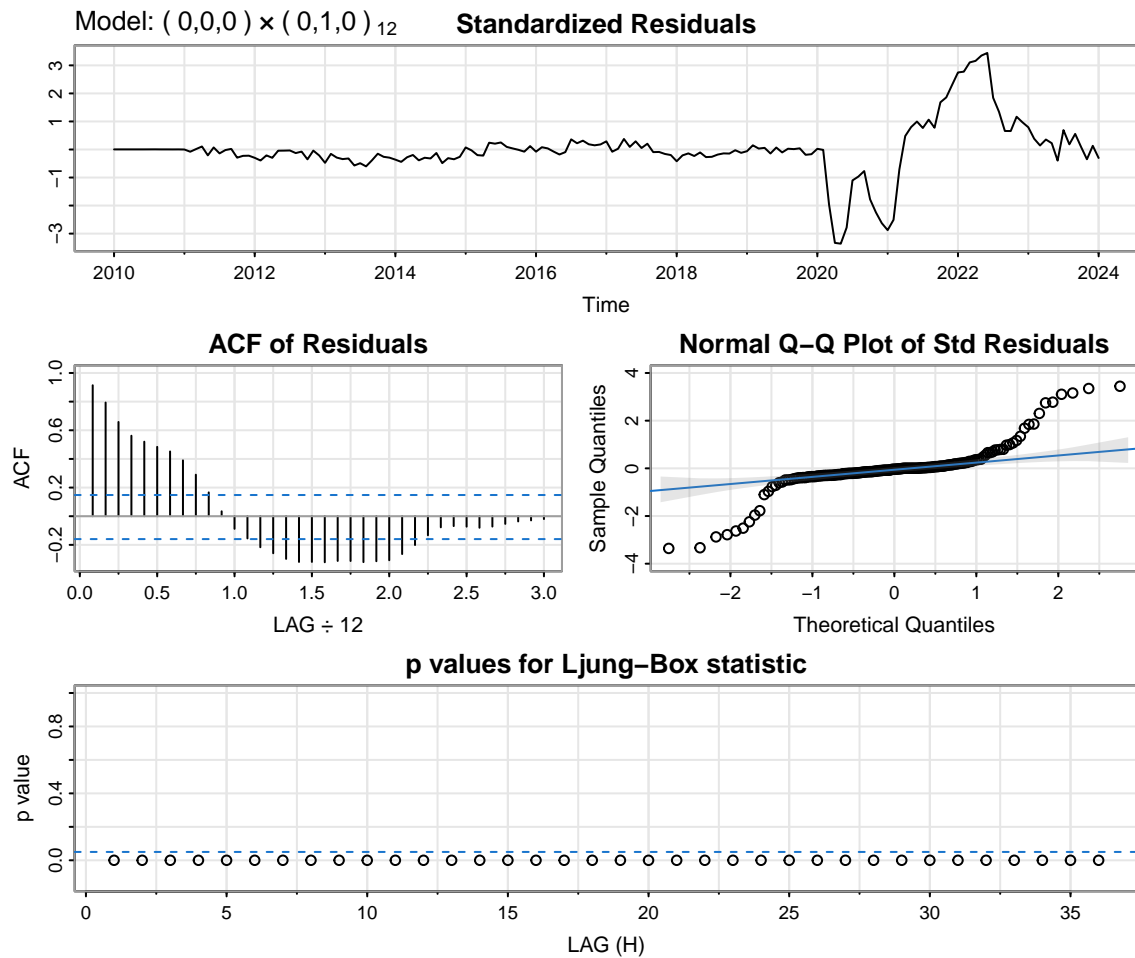


Figure 13: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

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.

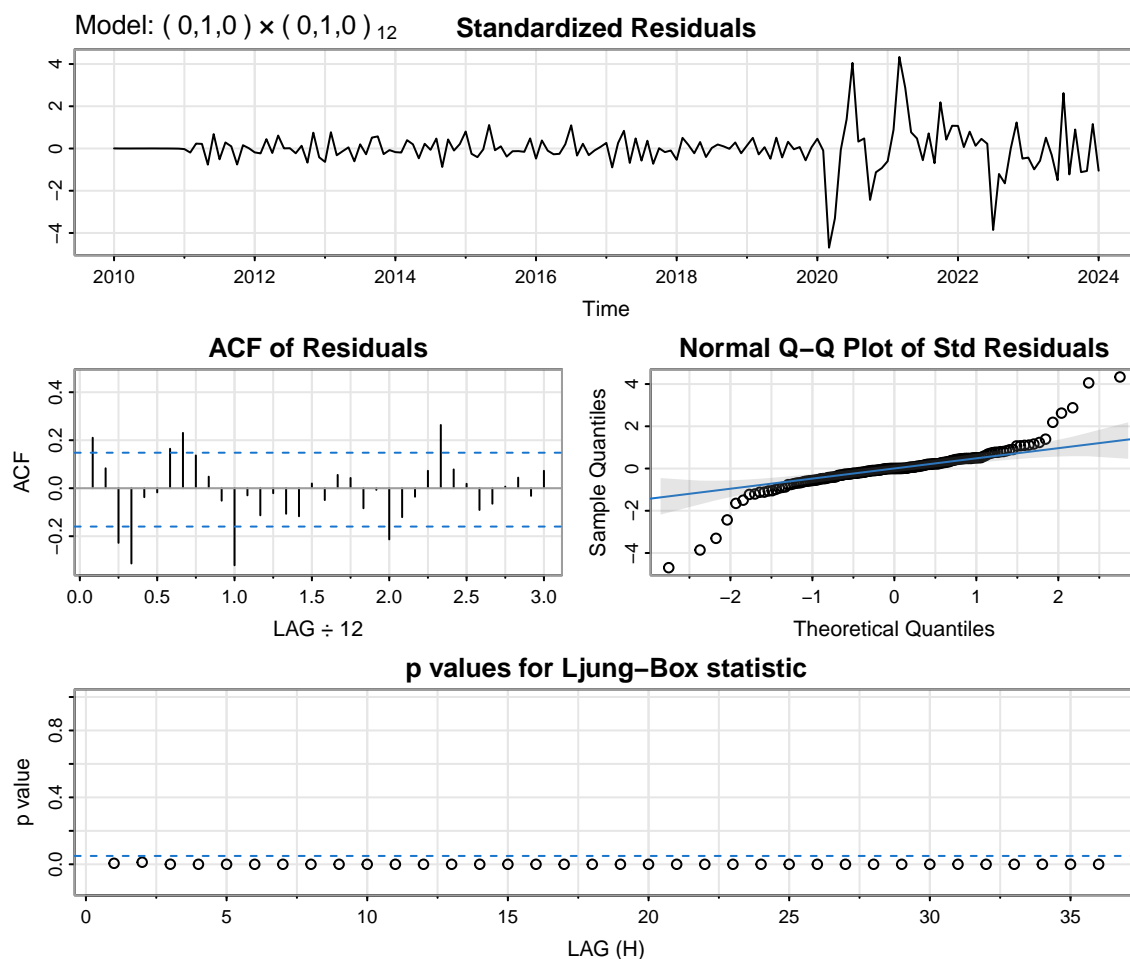


Figure 14: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

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 .

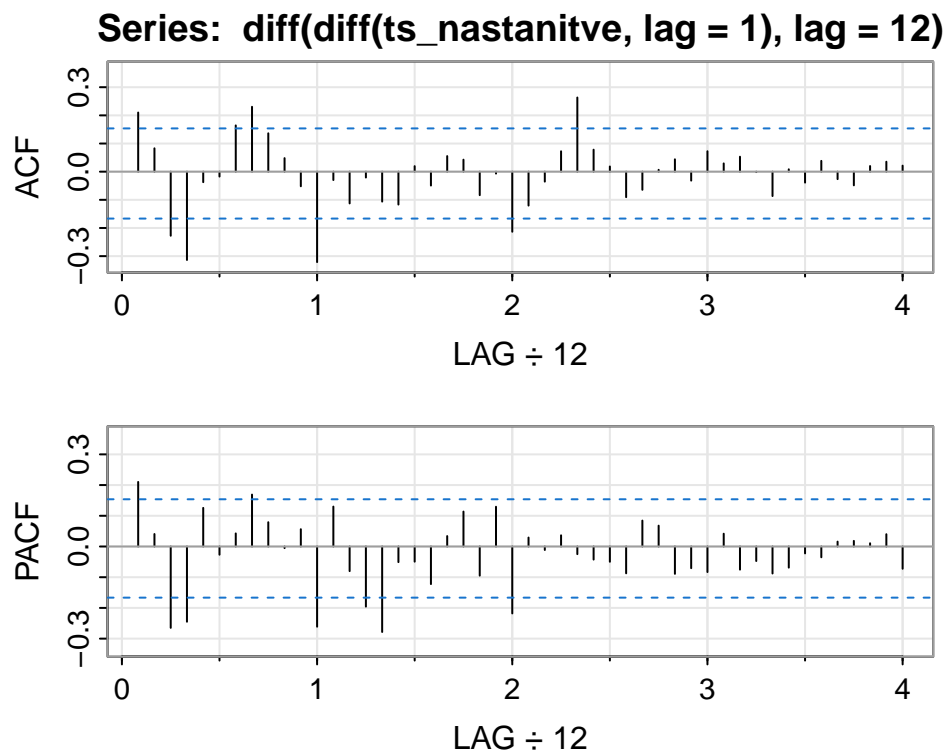


Figure 15: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.

Table 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 bi lahko zavrnili ničelno domnevo.

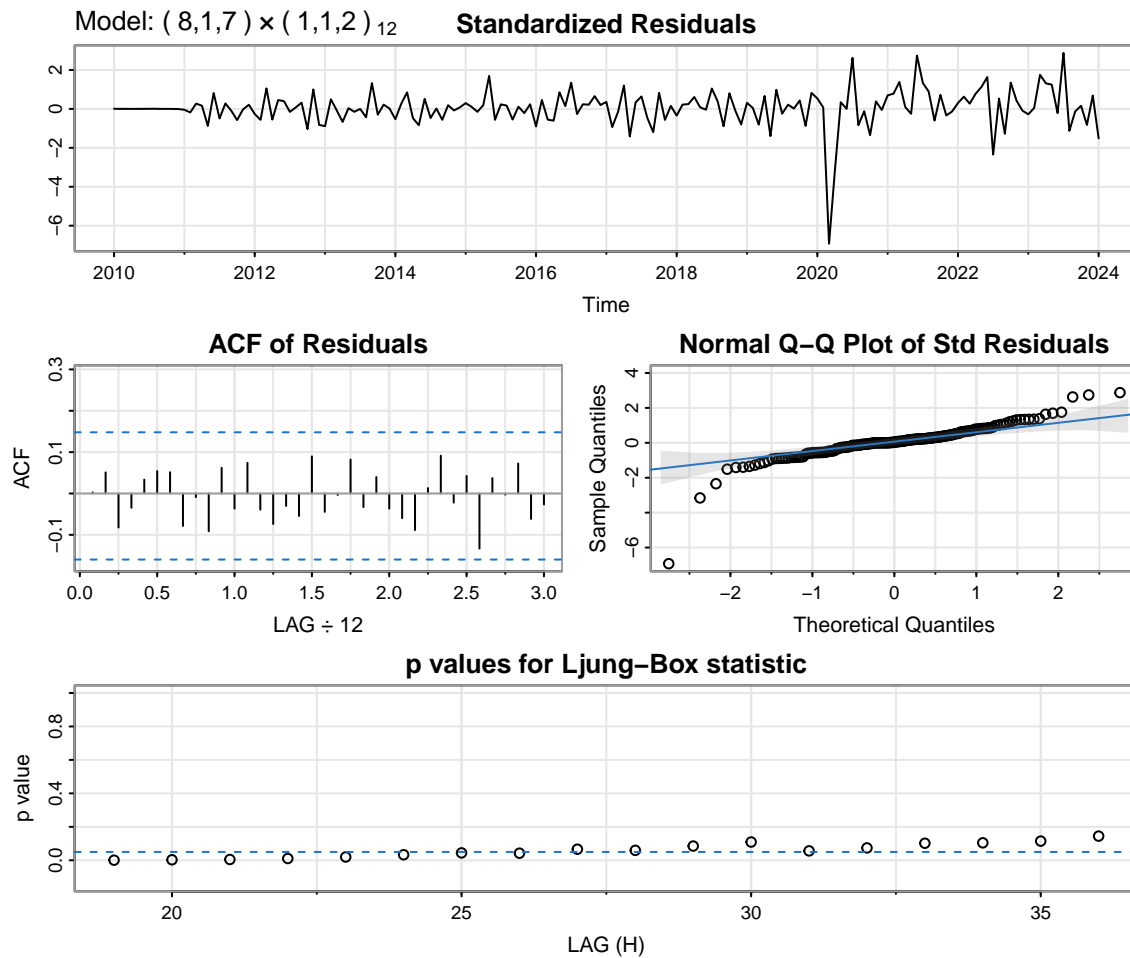


Figure 16: Ostanke modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

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.

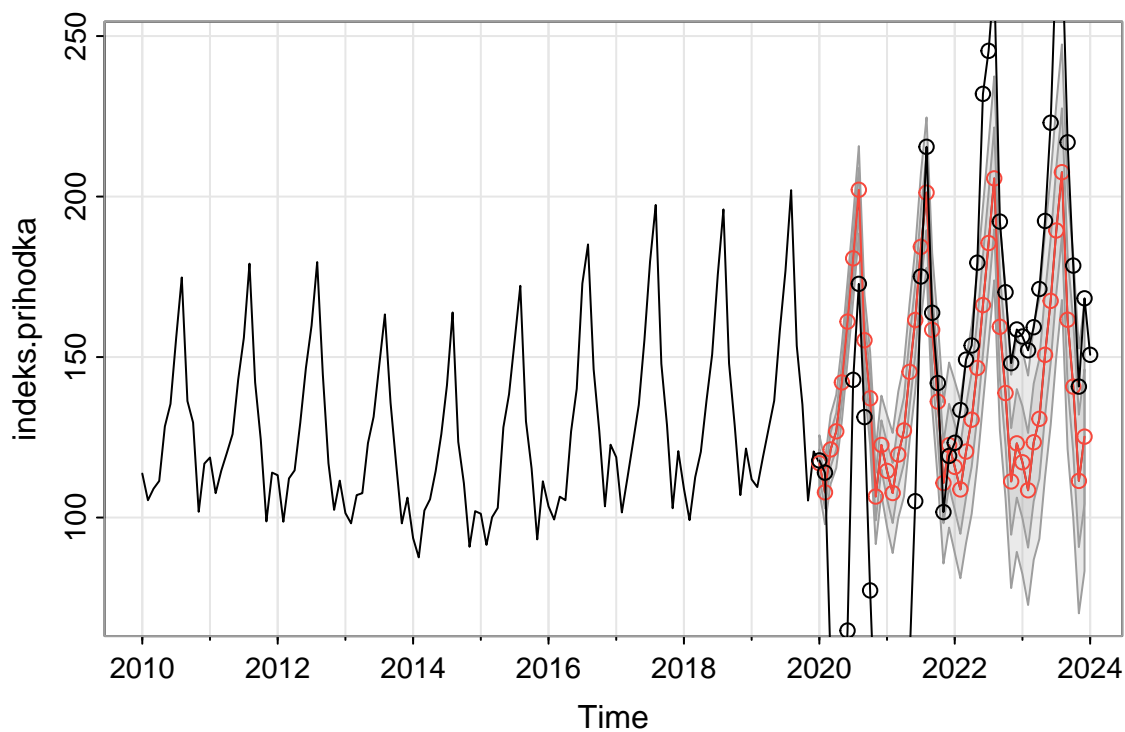


Figure 17: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

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 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.

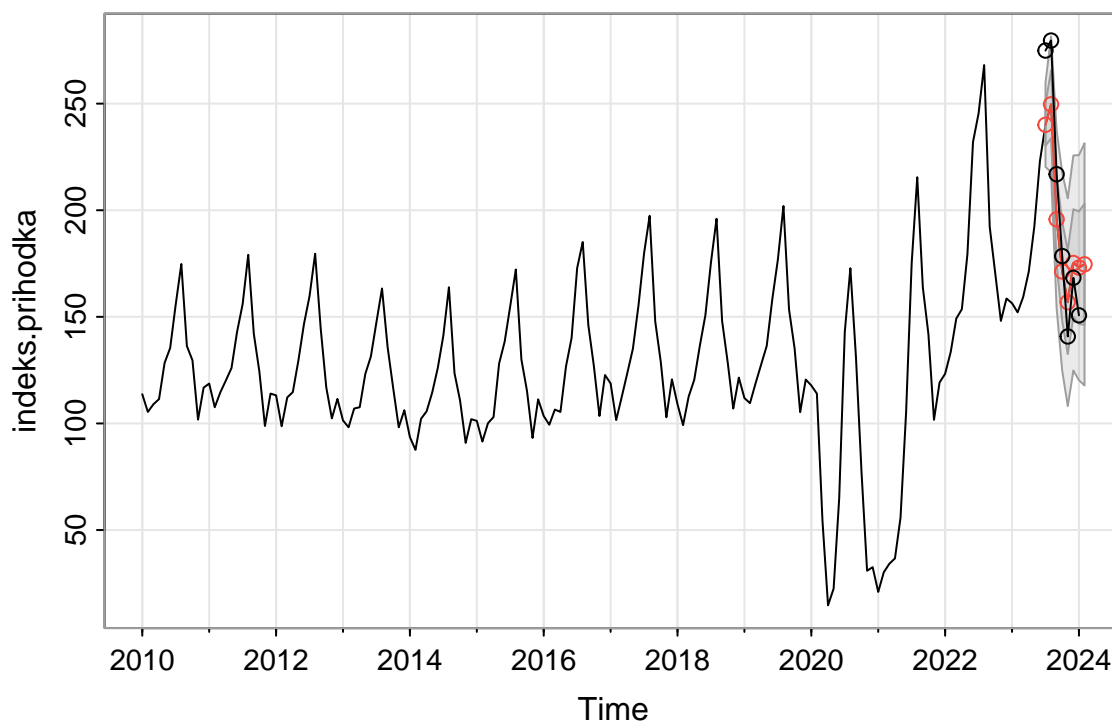


Figure 18: Časovna vrsta in napovedi na podlagi modela od sredine leta 2023 naprej.

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.

Časovna vrsta *Dejavnost strežbe jedi in pijač*

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.

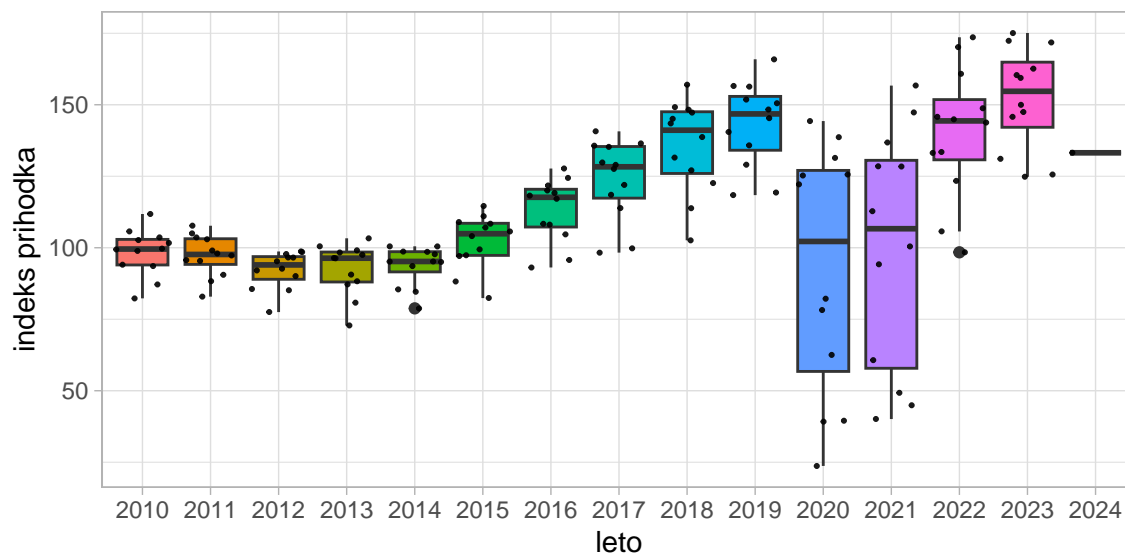


Figure 19: Porazdelitev indeksa prihodka v dejavnosti strežbe jedi in pijač glede na leta.

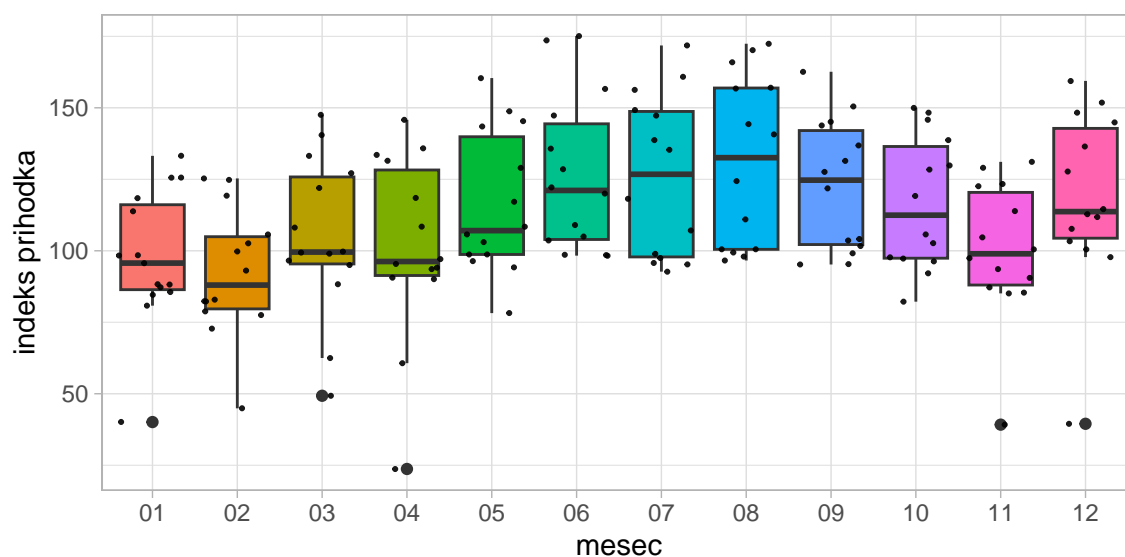


Figure 20: Porazdelitev indeksa prihodka v dejavnosti strežbe jedi in pijač glede na mesec.

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

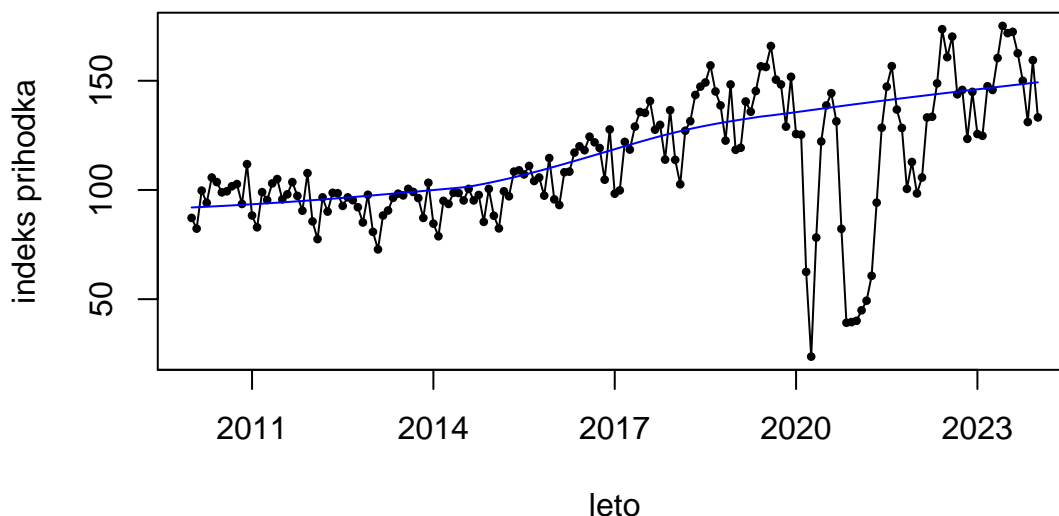


Figure 21: Časovna vrsta indeksa prihodka za 'Dejavnost strežbe jedi in pijač'.

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 predhodnim opazovanim obdobjem, za kar je verjetno ponovno krivo obnašanje in potrebe ljudi po pandemiji in izdaja državnih bonov.

Transformacija

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

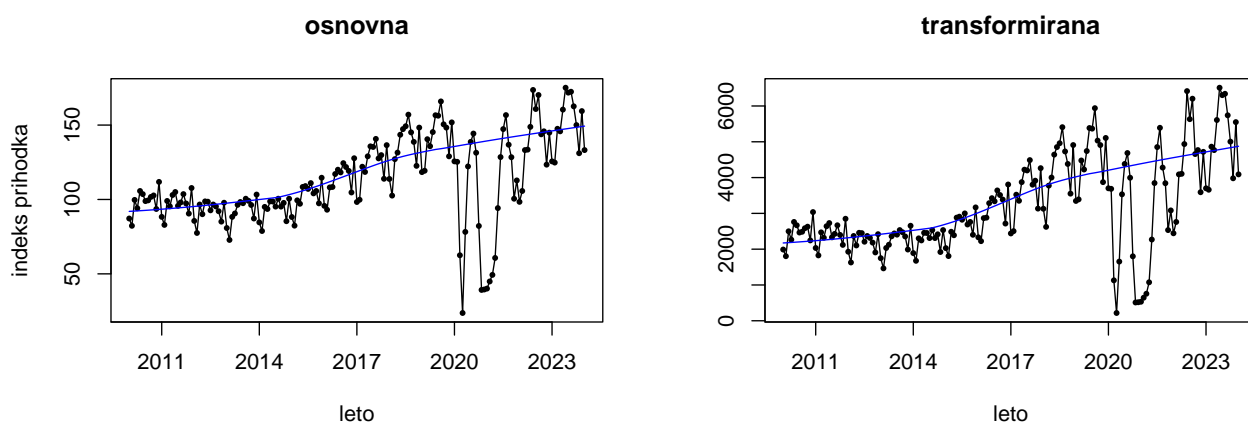


Figure 22: Osnovna in transformirana časovna vrsta 'Dejavnost strežbe jedi in pijač'.

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

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 vključitvi modeliranja sezonskosti, 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.

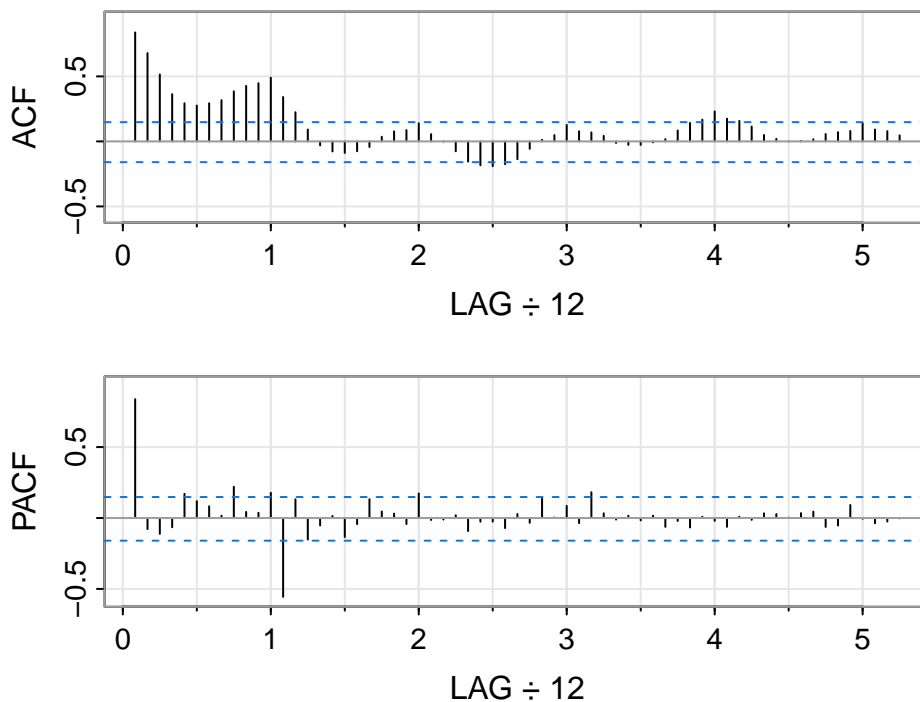


Figure 23: Avtokorelogram in parcialni avtokorelogram časovne vrste.

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.

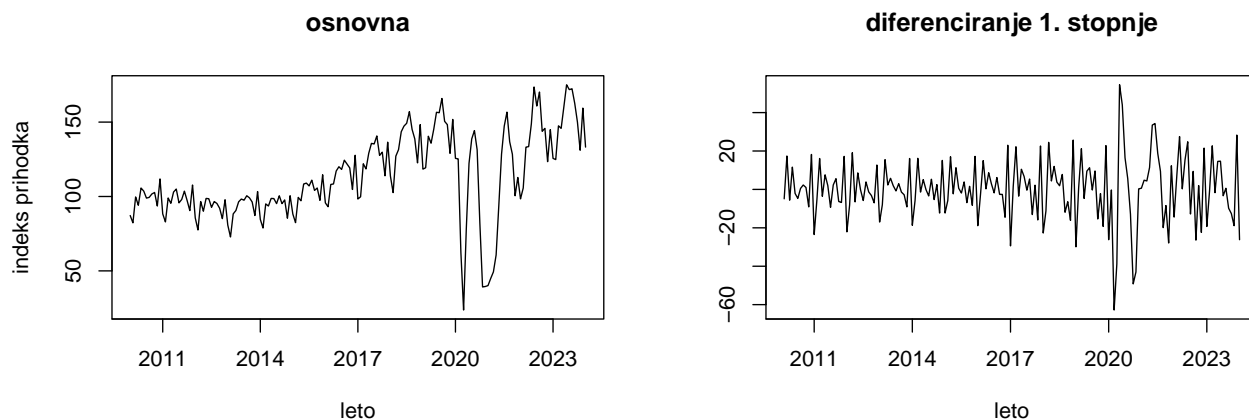


Figure 24: Osnovna in diferencirana časovna vrsta.

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.

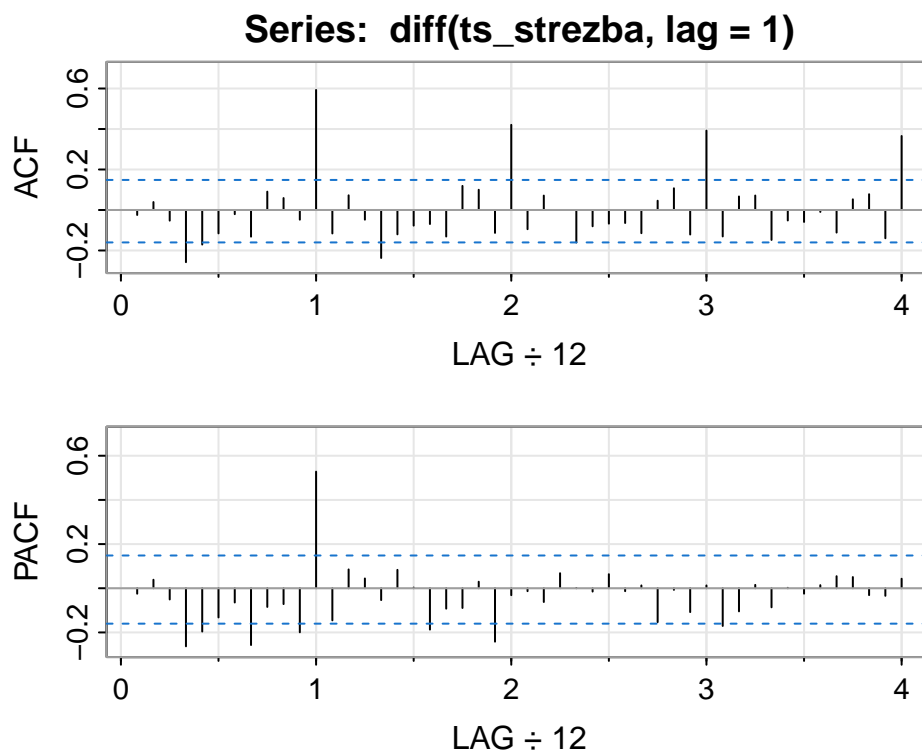


Figure 25: Avtokorelogram in parcialni avtokorelogram diferencirane časovne vrste.

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.

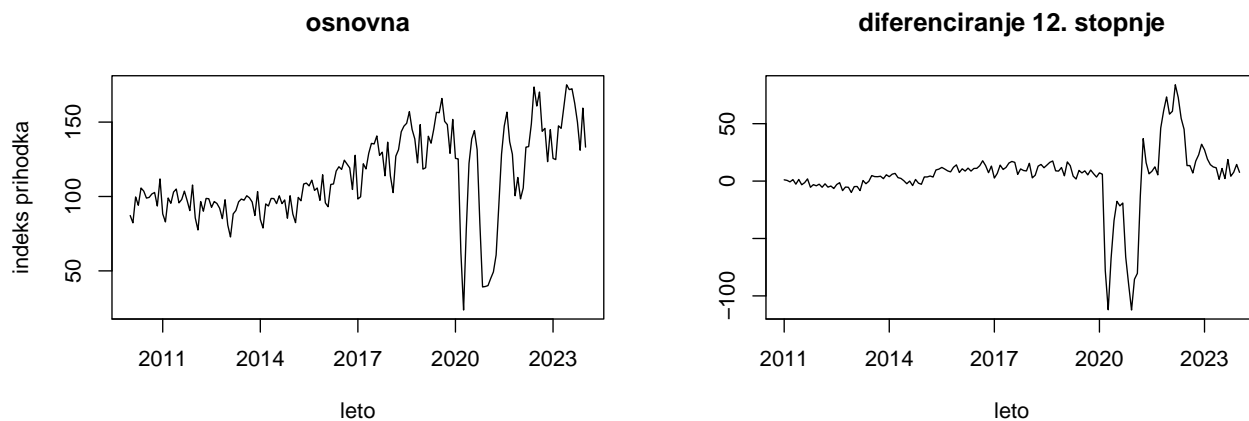


Figure 26: Osnovna in diferencirana(sezonskost) časovna vrsta.

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

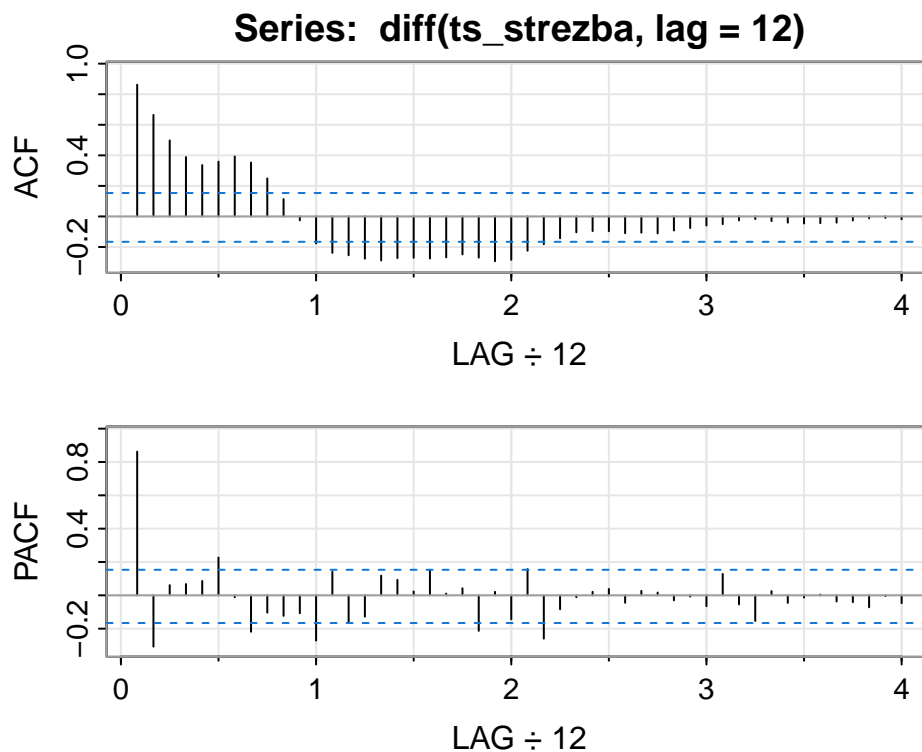


Figure 27: Avtokorelogram in parcialni avtokorelogram diferencirane(sezonskost) časovne vrste.

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.

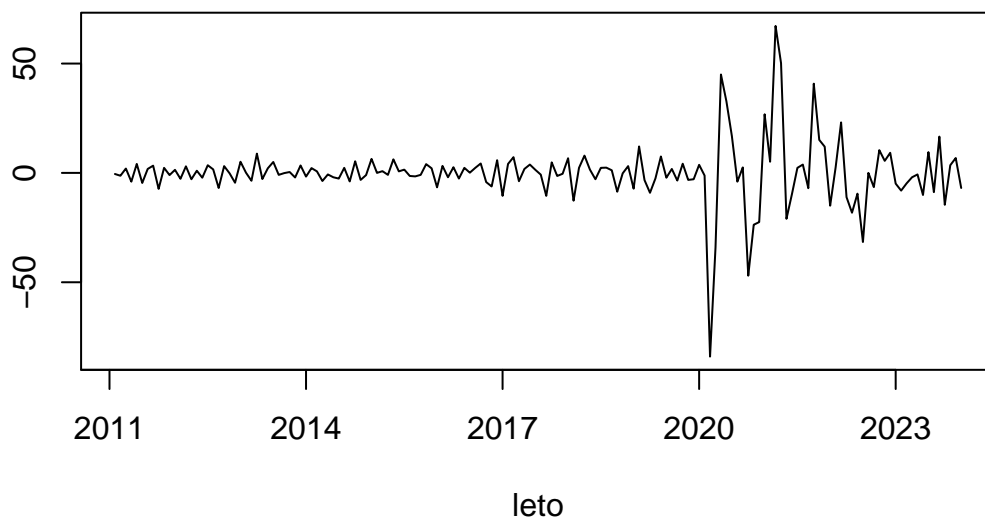


Figure 28: Diferenciranje trenda in sezone v časovni vrsti.

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.

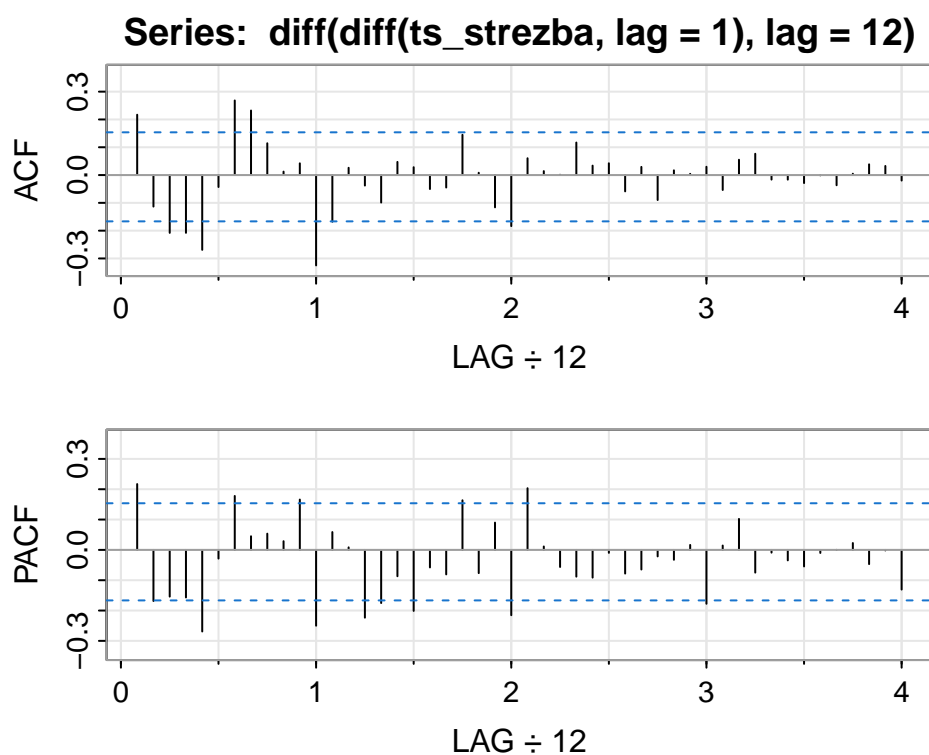


Figure 29: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezone) časovne vrste.

Izbira ustreznega modela

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

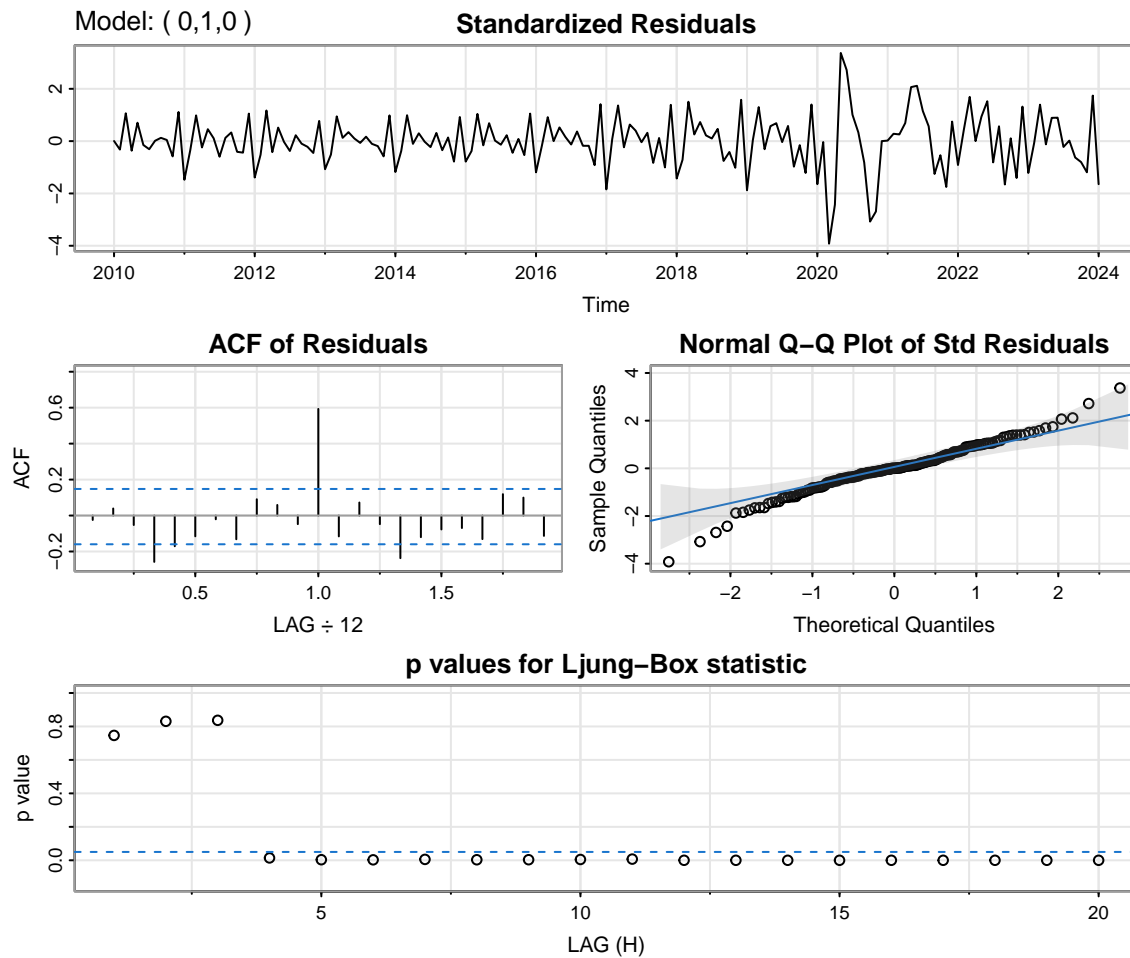


Figure 30: Ostanke modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo trend.

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 imamo večina p-vrednosti manjše od 0.05, torej zavračamo ničelno domnevo, da ni prisotne avtokorelacije med členi časovne vrste. Zato v model vključimo diferenciranje sezonskosti.

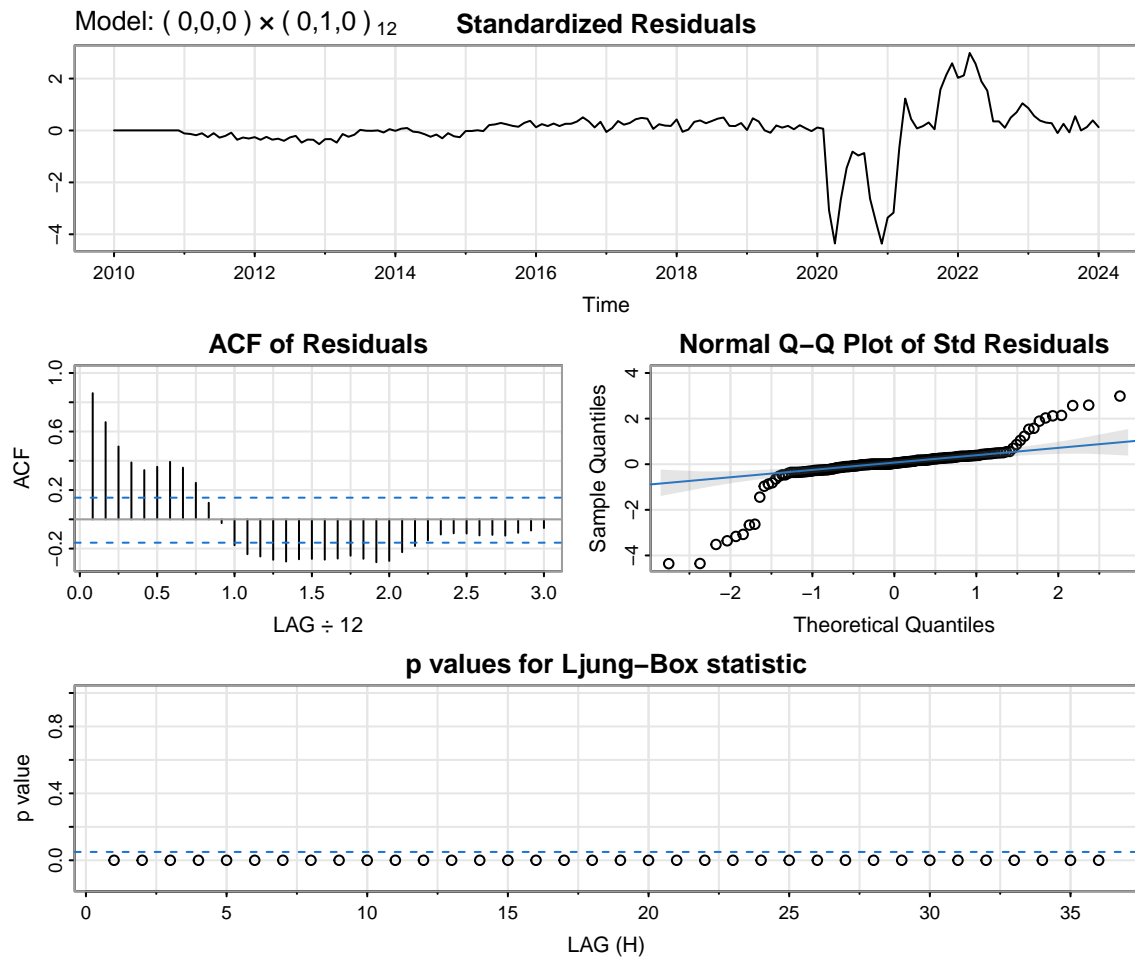


Figure 31: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

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.

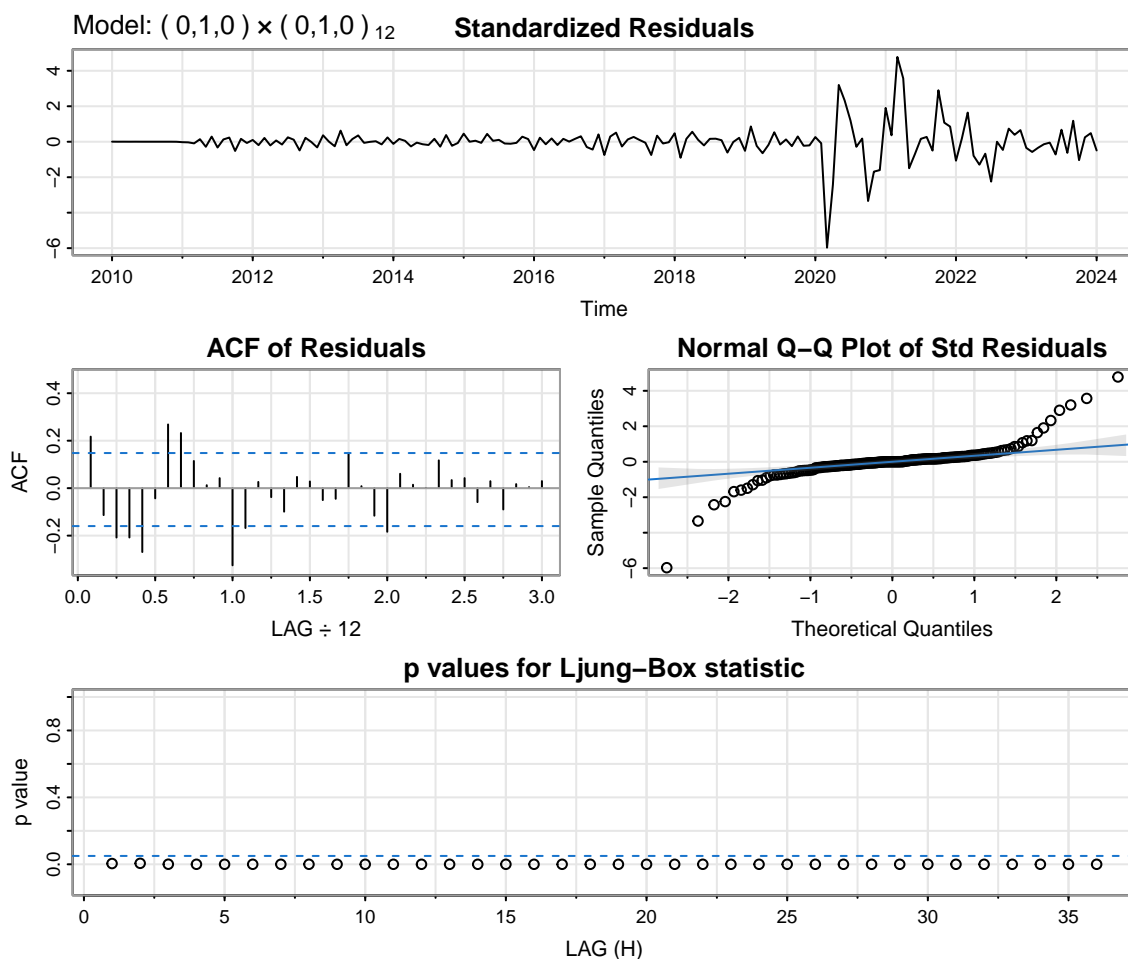


Figure 32: Ostanke modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

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 .

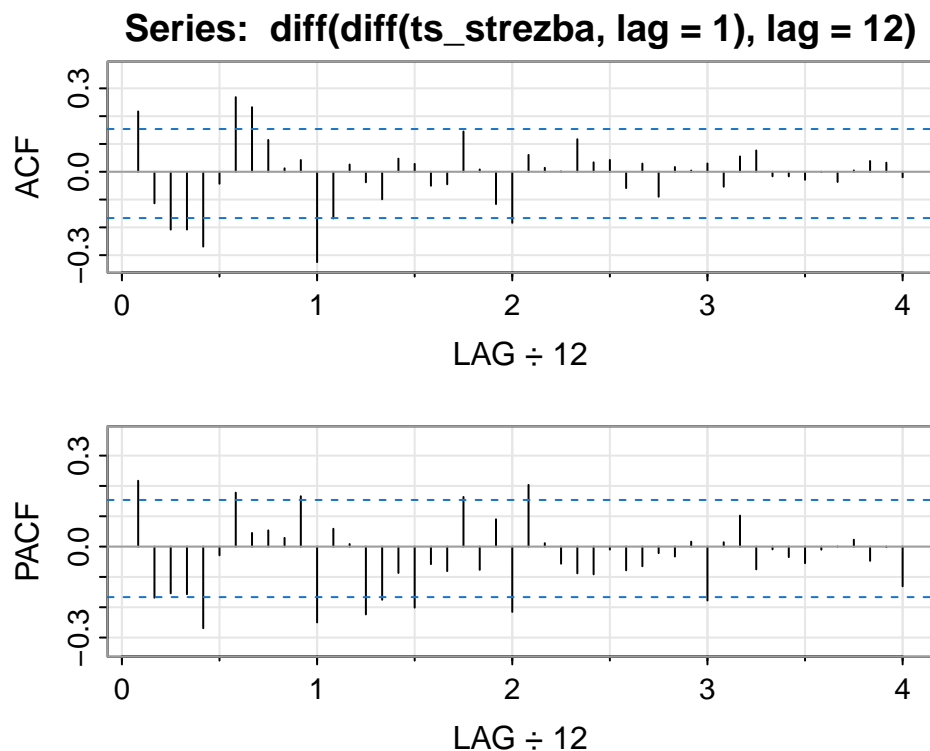


Figure 33: Avtokorelogram in parcialni avtokorelogram diferencirane(trend in sezonskost) časovne vrste.

Table 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 Akaikijevega 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.

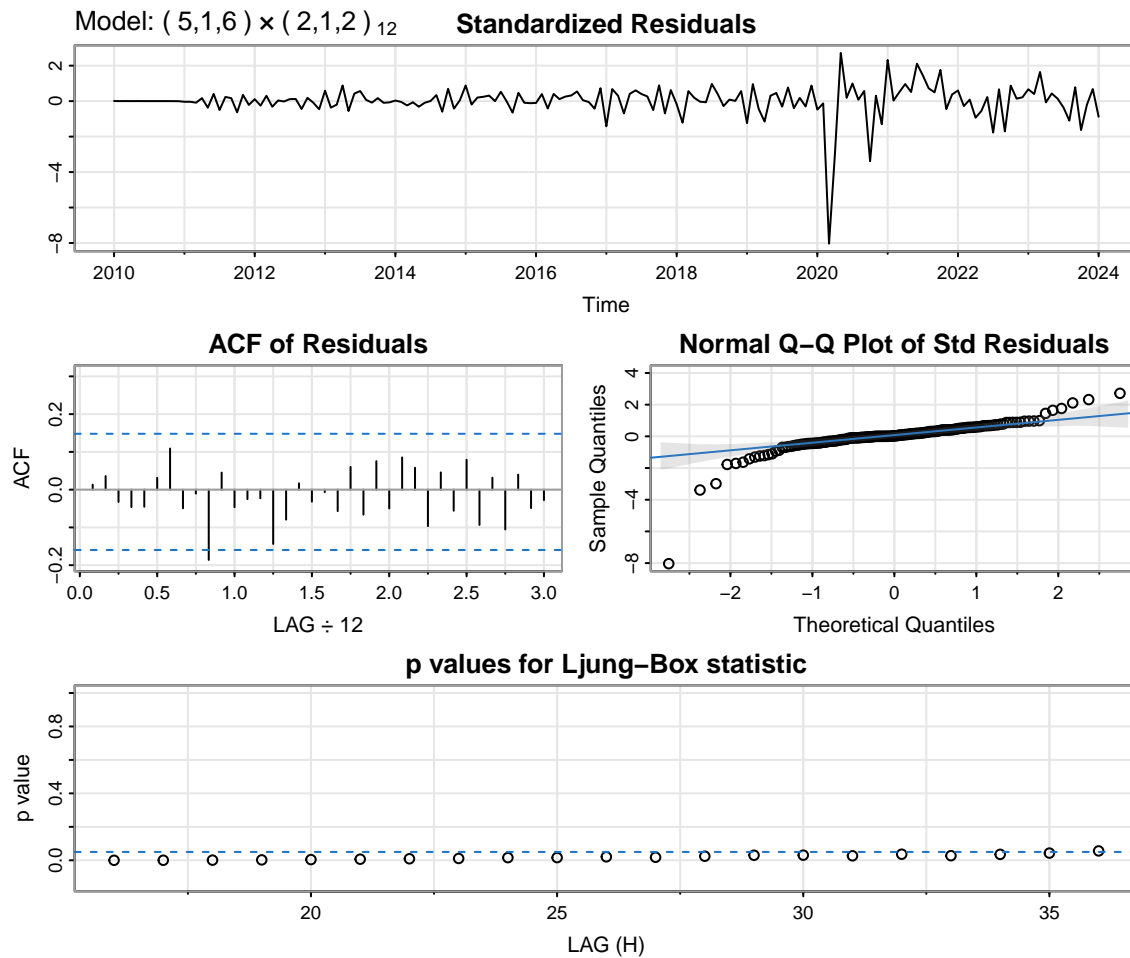


Figure 34: Ostanke modela, avtokorelogram in p-vrednosti Ljung-Box testa za časovno vrsto, kjer diferenciramo sezonskost.

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.

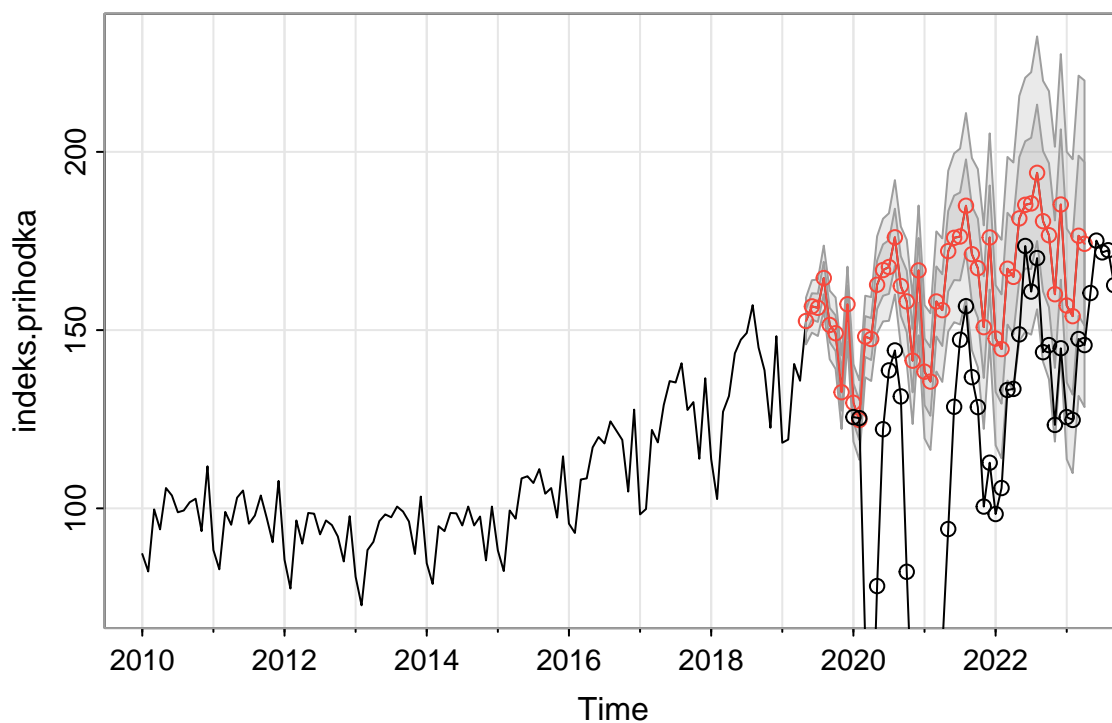


Figure 35: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

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.

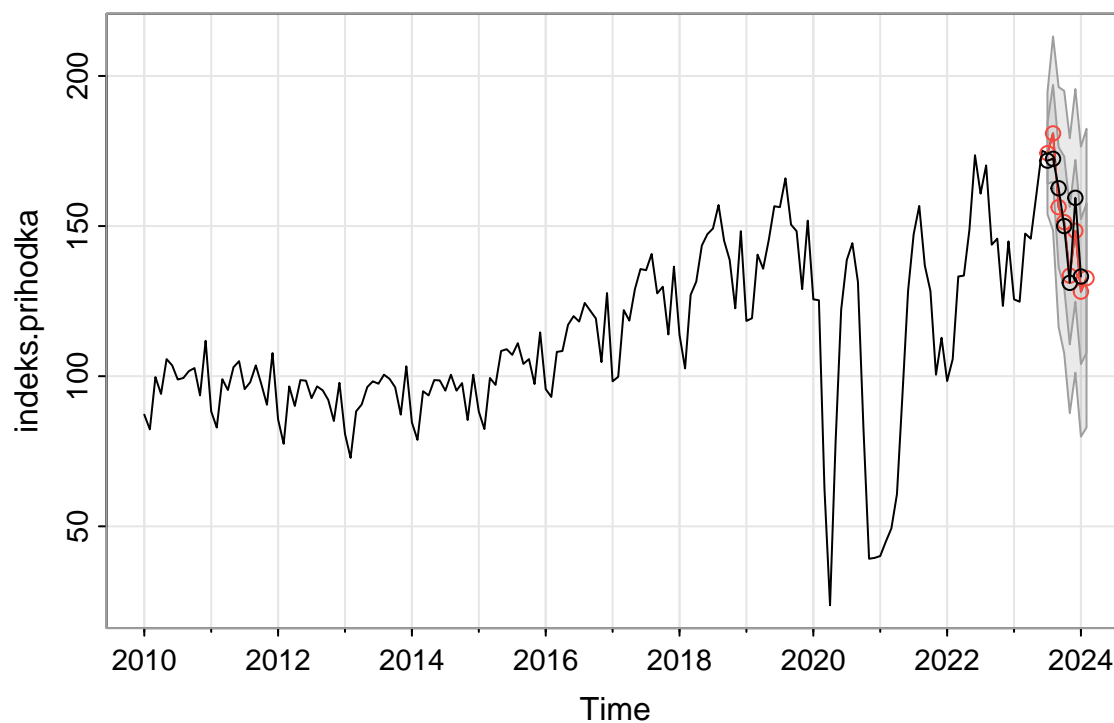


Figure 36: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

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.

Časovna vrsta *Gostinstvo prosta delovna mesta 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.

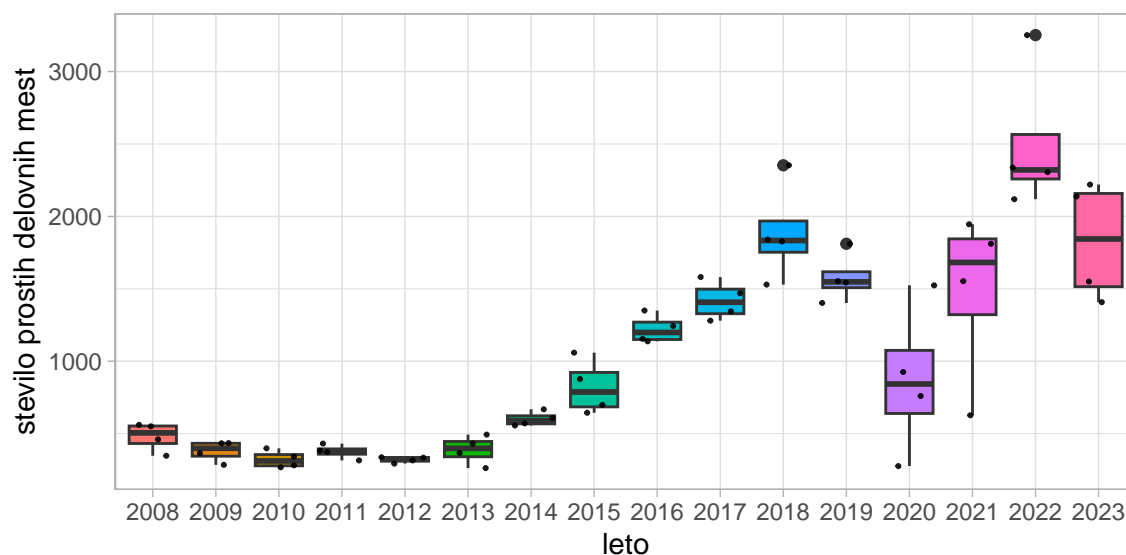


Figure 37: Porazdelitev prostih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na leto.

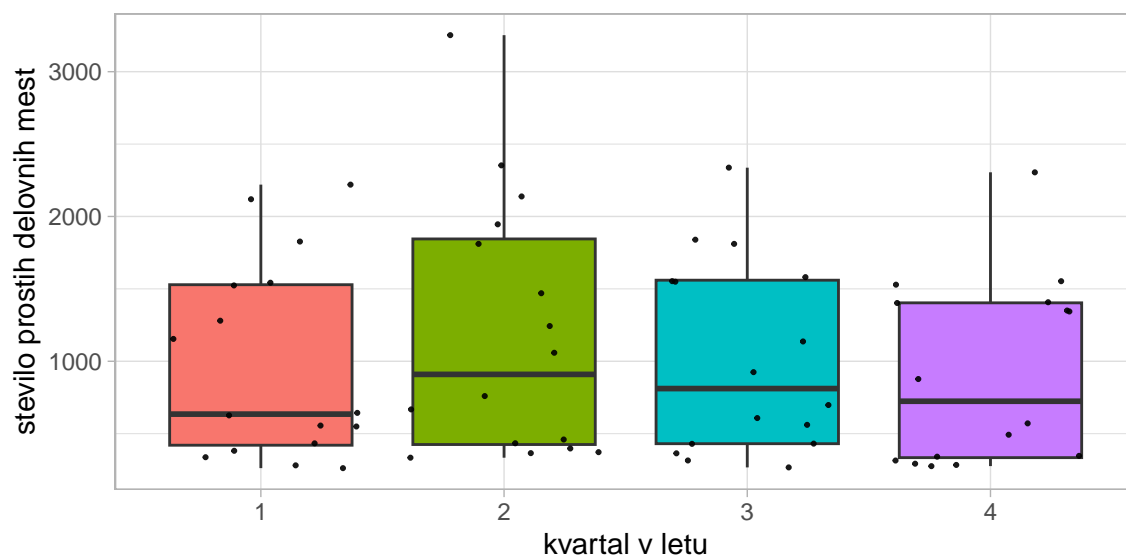


Figure 38: Porazdelitev prostih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na kvartal.

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

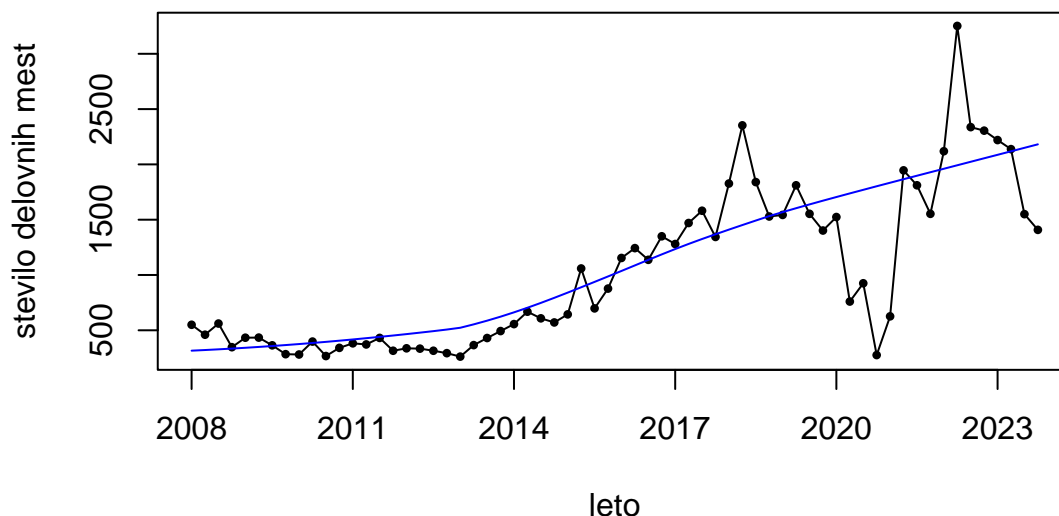


Figure 39: Časovna vrsta 'Gostinstvo prosta delovna mesta 1+'.

Kot sva omenila že zgoraj je iz grafa časovne vrste možno opaziti prirobnost 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 ekstrmno 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.

Transformacija časovne vrste

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

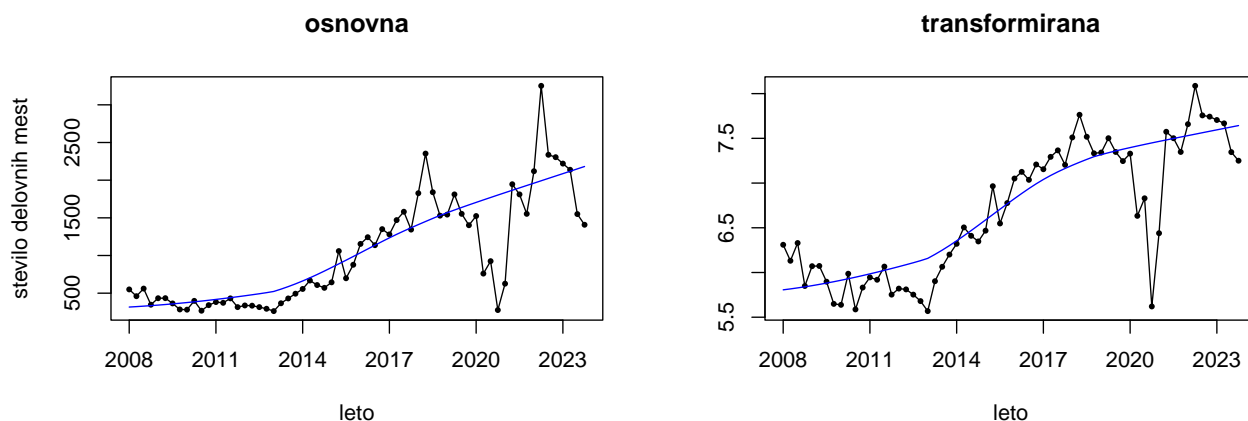


Figure 40: Osnovna in logaritmirana časovna vrsta 'Gostinstvo prosta delovna mesta 1+'.

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

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.

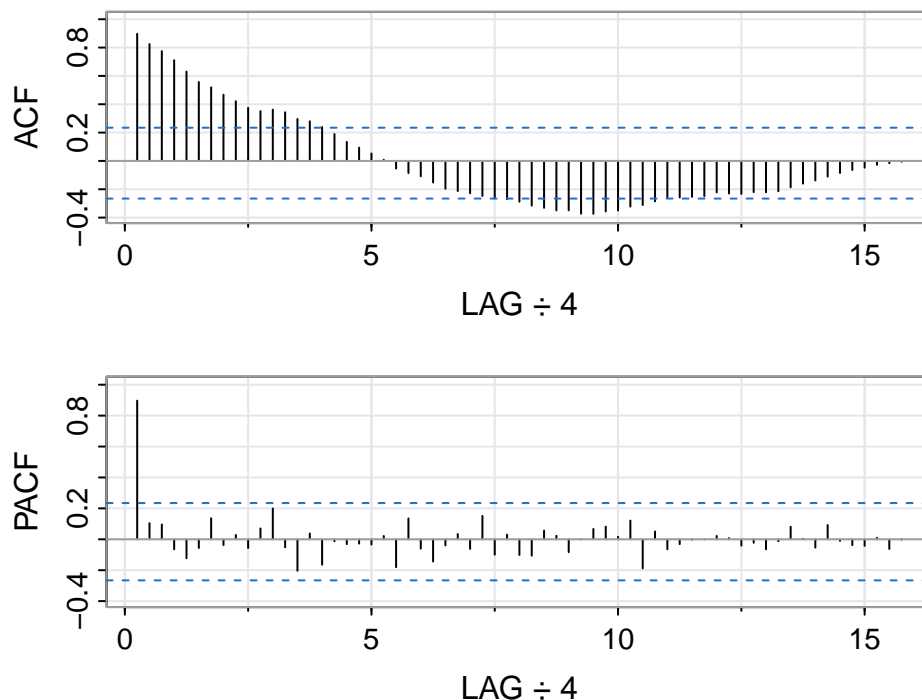


Figure 41: Avtokorelogram in parcialni avtokorelogram logaritmirane časovne vrste.

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

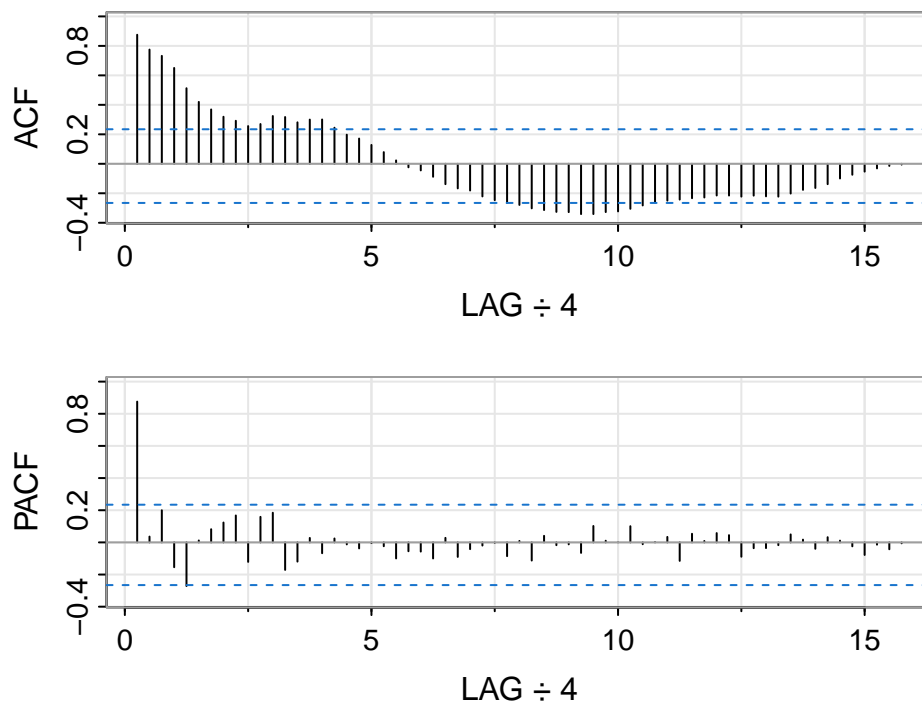


Figure 42: Avtokorelogram in parcialni avtokorelogram osnovne časovne vrste.

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

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.

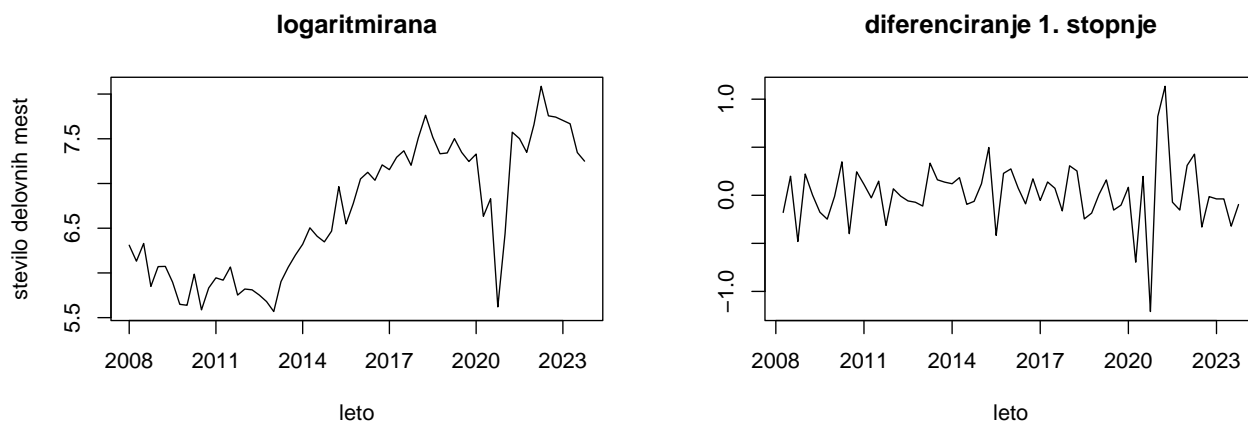


Figure 43: Logaritmirana osnovna in logaritmirana diferencirana časovna vrsta.

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 vedno vpliv Covid-19 epidemije.

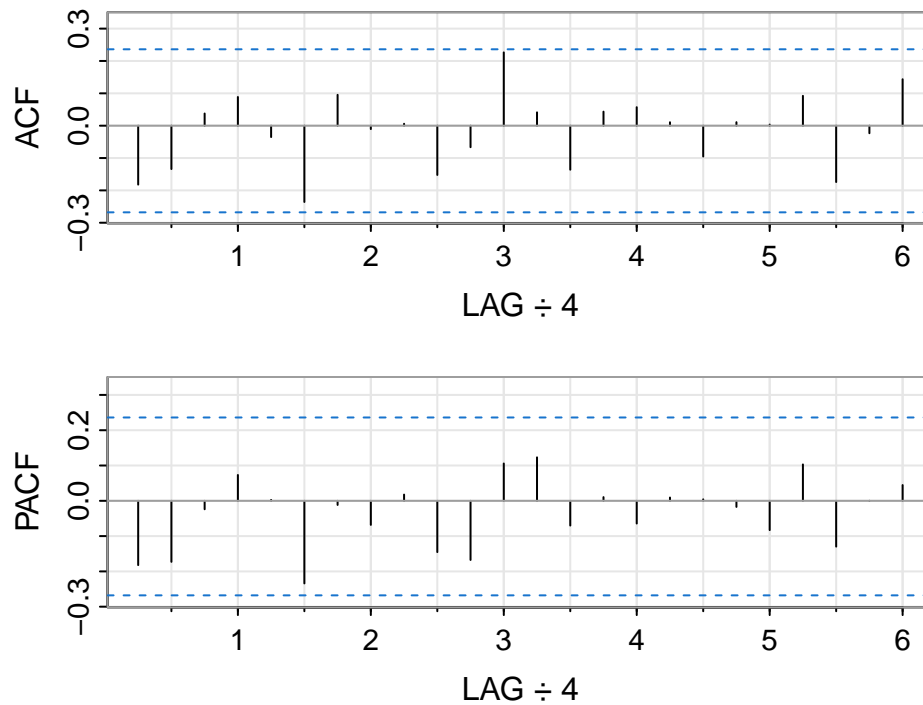


Figure 44: Avtokorelogram in parcialni avtokorelogram logaritmirane diferencirane časovne vrste.

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.

Izbira ustreznega modela

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

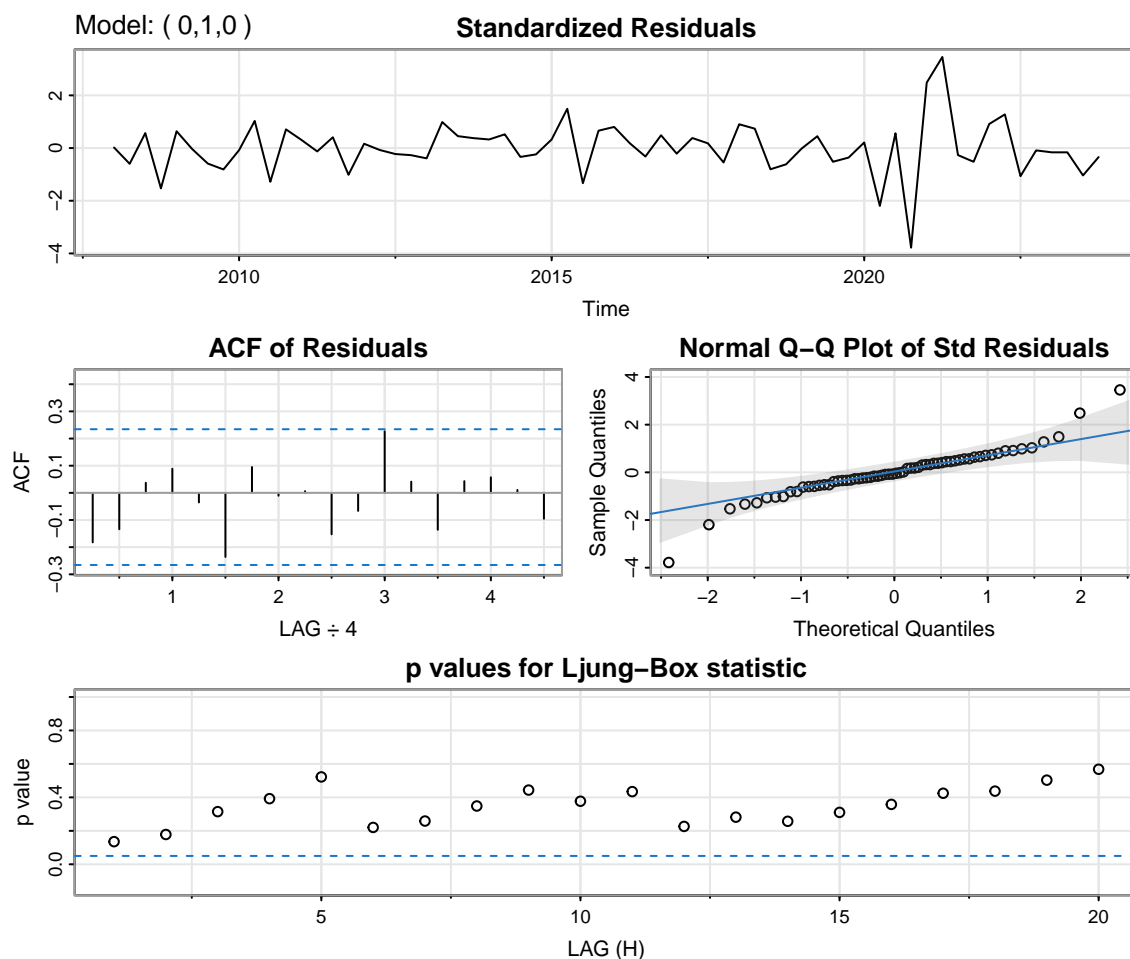


Figure 45: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa.

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.

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.

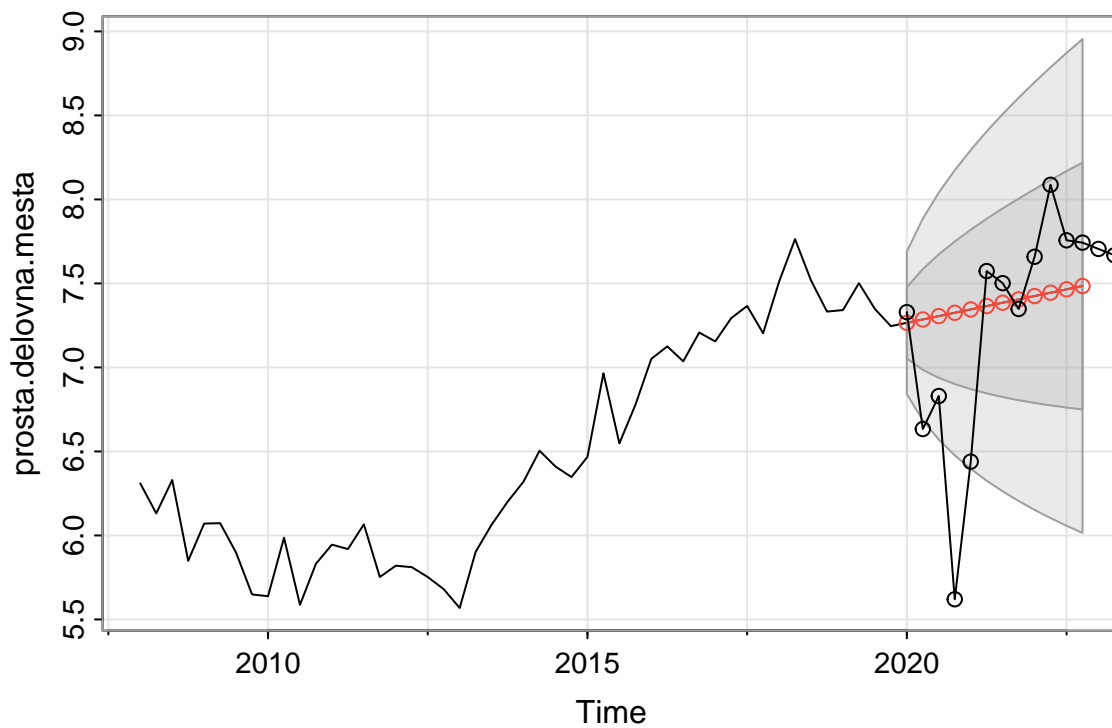


Figure 46: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

Jasno je, da napoved ni najboljša, saj ostanki vrste predstavljajo beli šum in pričakovano će 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.

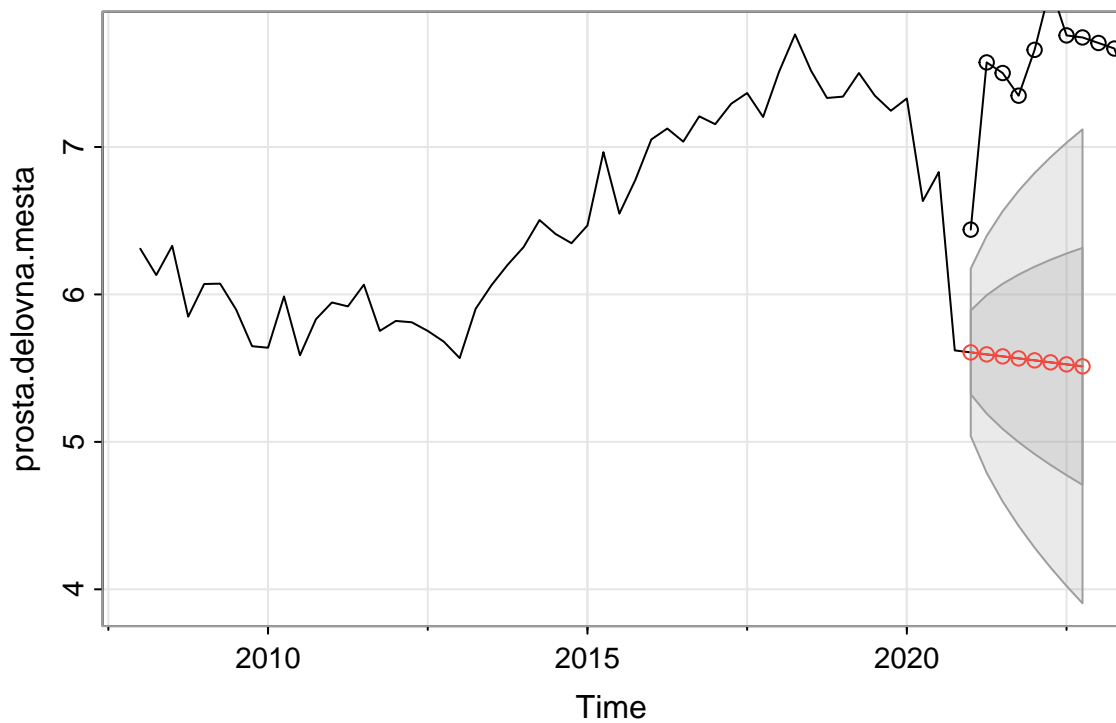


Figure 47: Časovna vrsta in napovedi na podlagi modela od leta 2021 naprej.

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.

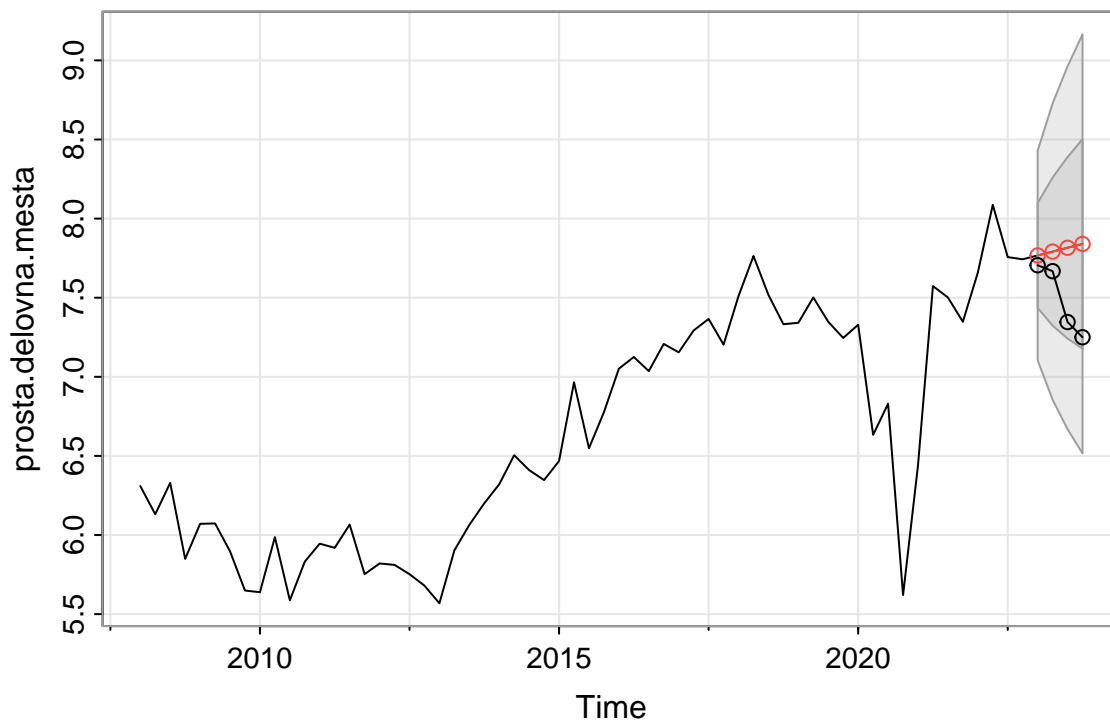


Figure 48: Časovna vrsta in napovedi na podlagi modela od leta 2023 naprej.

Napovedi tudi tukaj niso pravilne in ne ravno najbolj informativne, najboljše zadane samp za en trenutek naprej, 95% interavli 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.

Časovna vrsta *Gostinstvo zasedena delovna mesta 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.

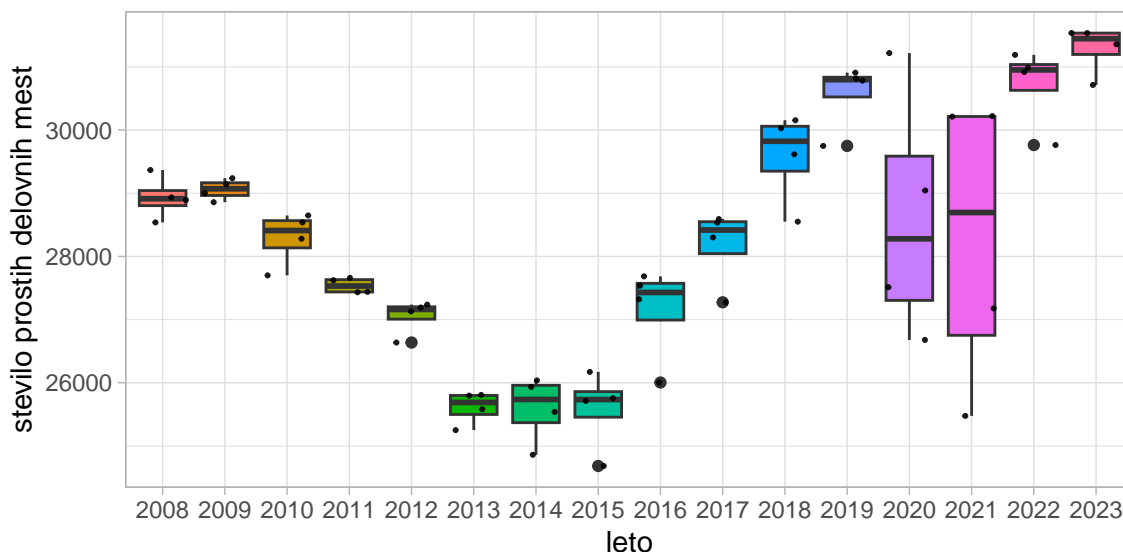


Figure 49: Porazdelitev zasedenih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na leto.

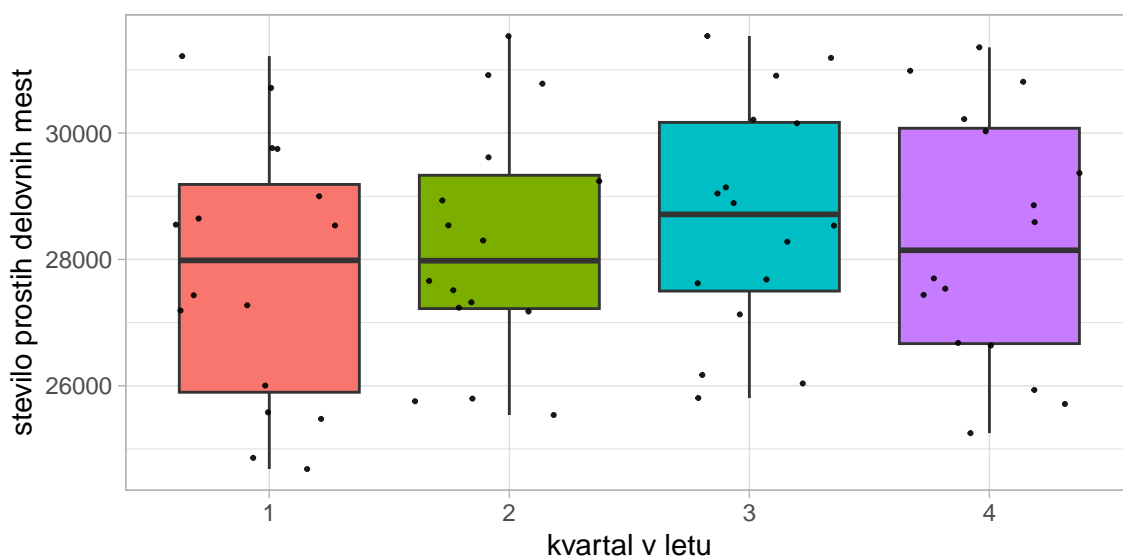


Figure 50: Porazdelitev zasedenih delovnih v gostinstvu, kjer je zaposlena vsaj 1 oseba, glede na kvartal.

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

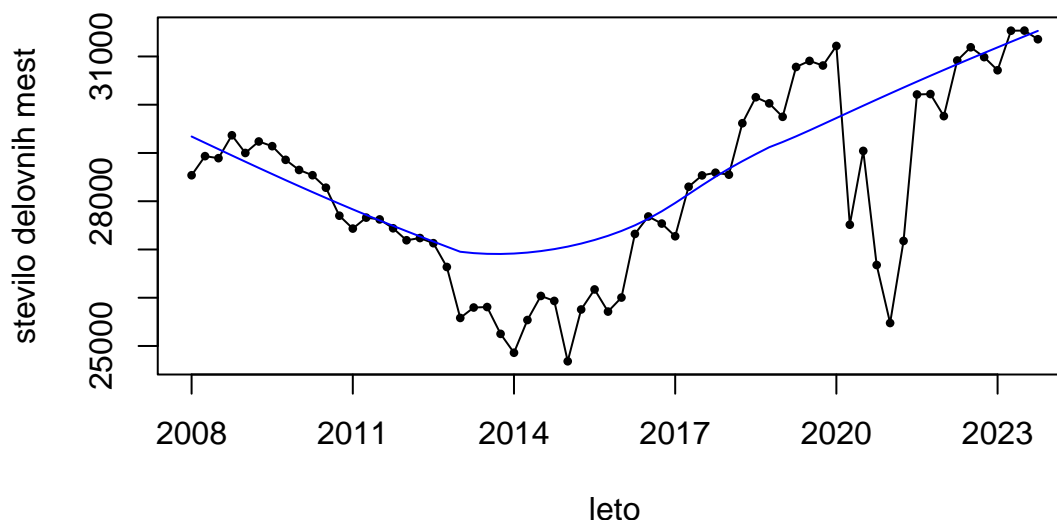


Figure 51: Časovna vrsta 'Gostinstvo zasedena delovna mesta 1+'.

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 ekstremono 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 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.

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.

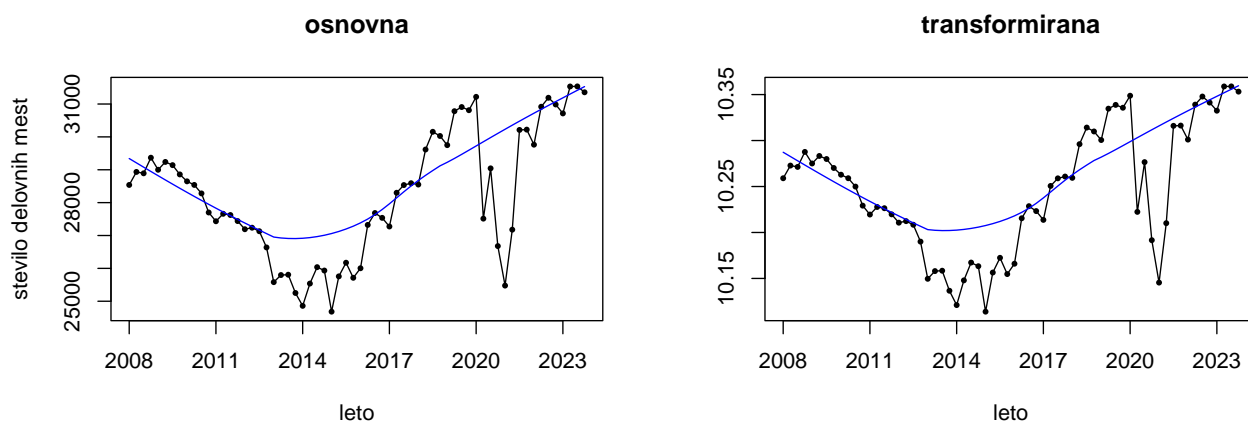


Figure 52: Osnovna in logaritmirana časovna vrsta 'Gostinstvo zasedena delovna mesta 1+'.

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

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 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 (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.

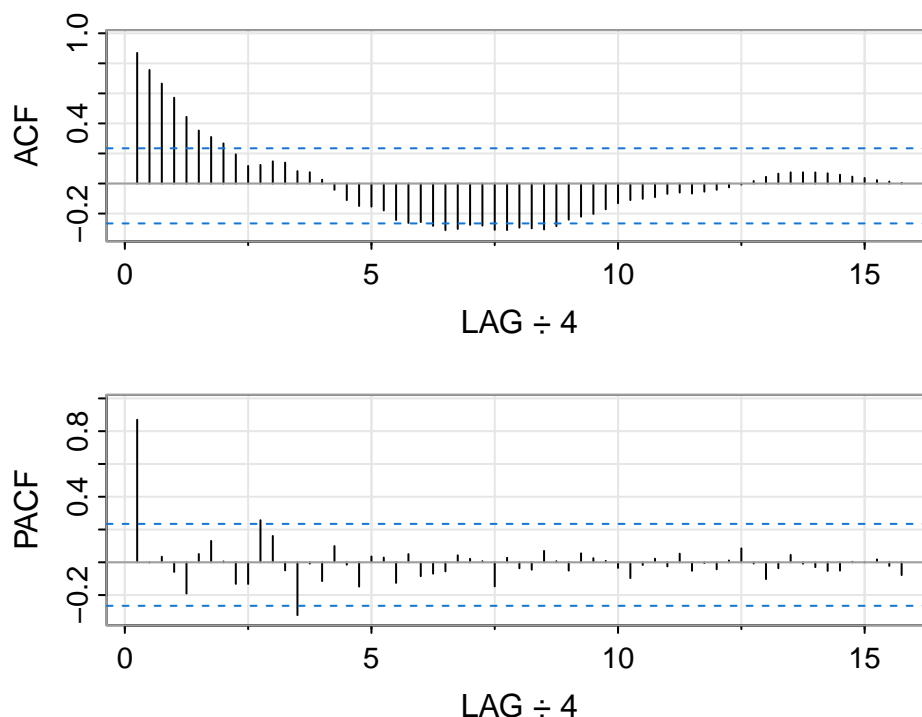


Figure 53: Avtokorelogram in parcialni avtokorelogram logaritmirane časovne vrste.

Diferenciranje

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

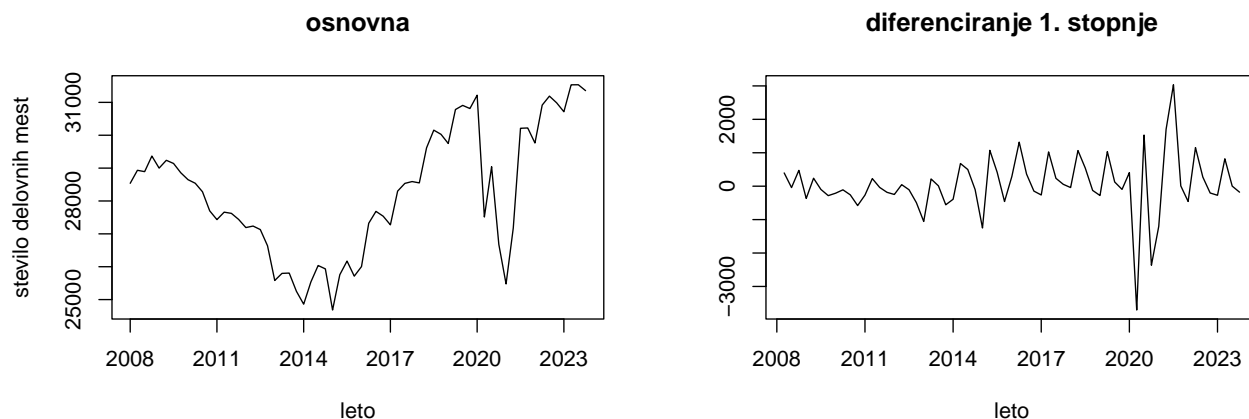


Figure 54: Osnovna in diferencirana časovna vrsta.

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.

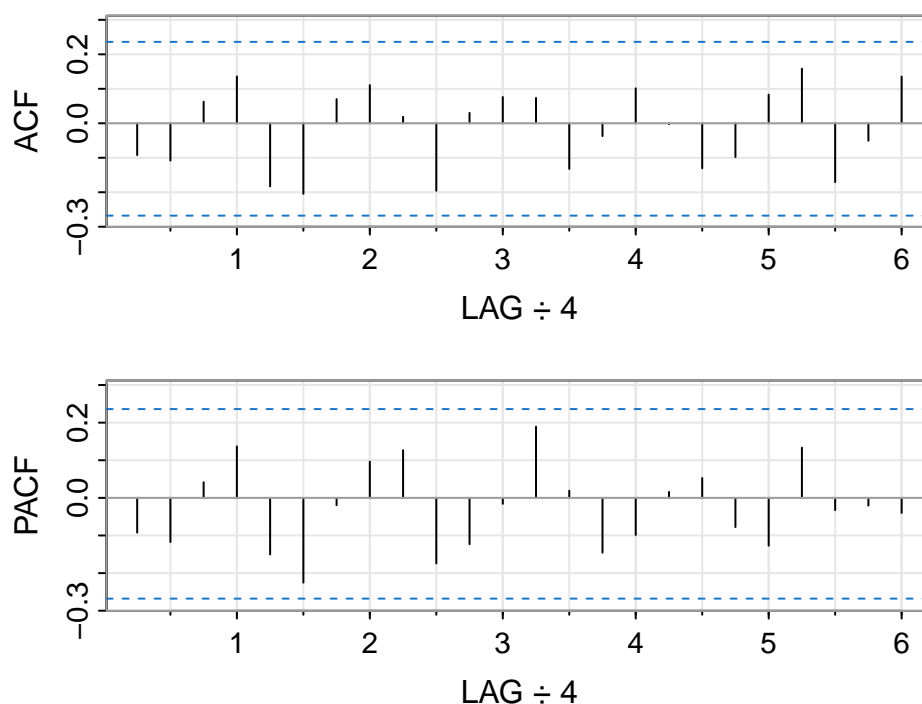


Figure 55: Avtokorelogram in parcialni avtokorelogram logaritmirane diferencirane časovne vrste.

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.

Izbira ustreznega modela

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

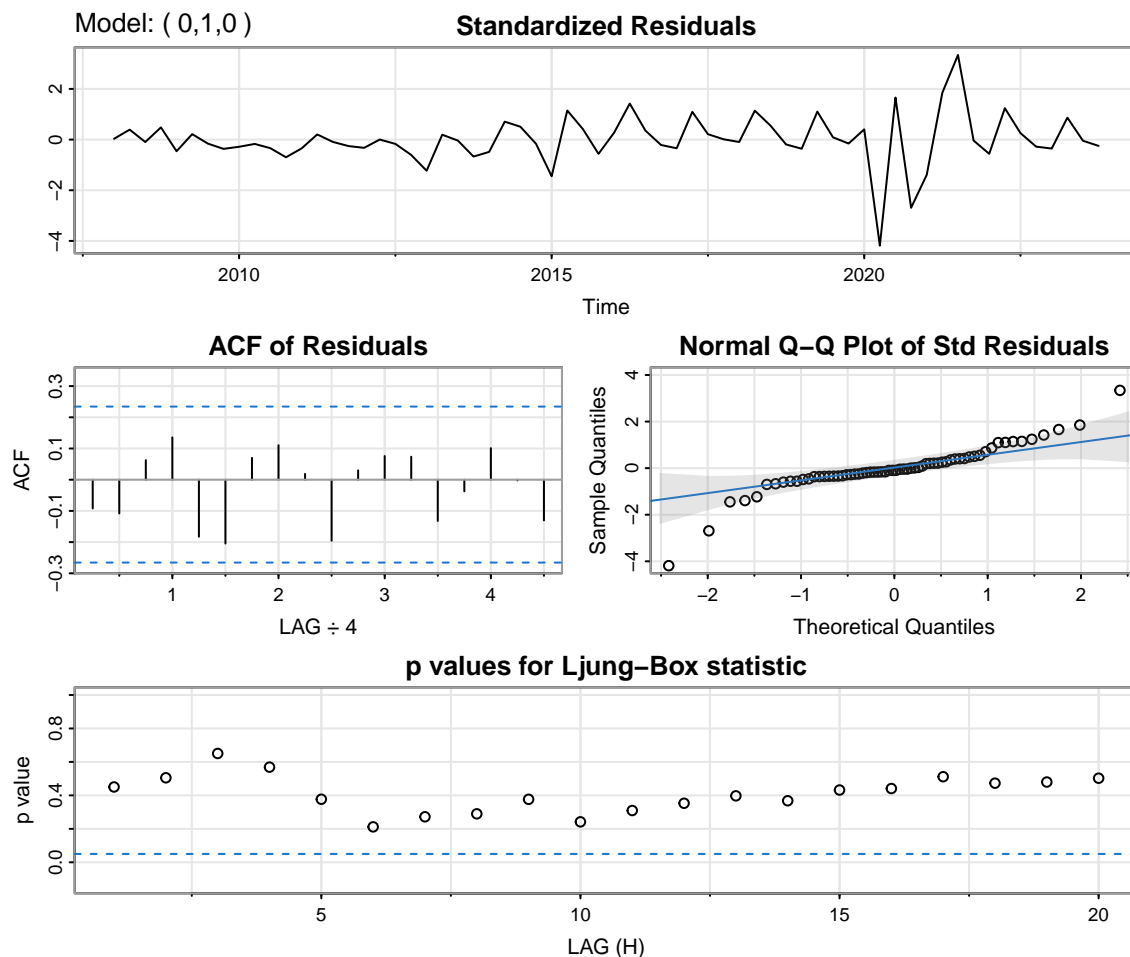


Figure 56: Ostanki modela, avtokorelogram in p-vrednosti Ljung-Box testa.

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.

Napoved

Sedaj bova poskušala z modelom še napovedati število zasedenih 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.

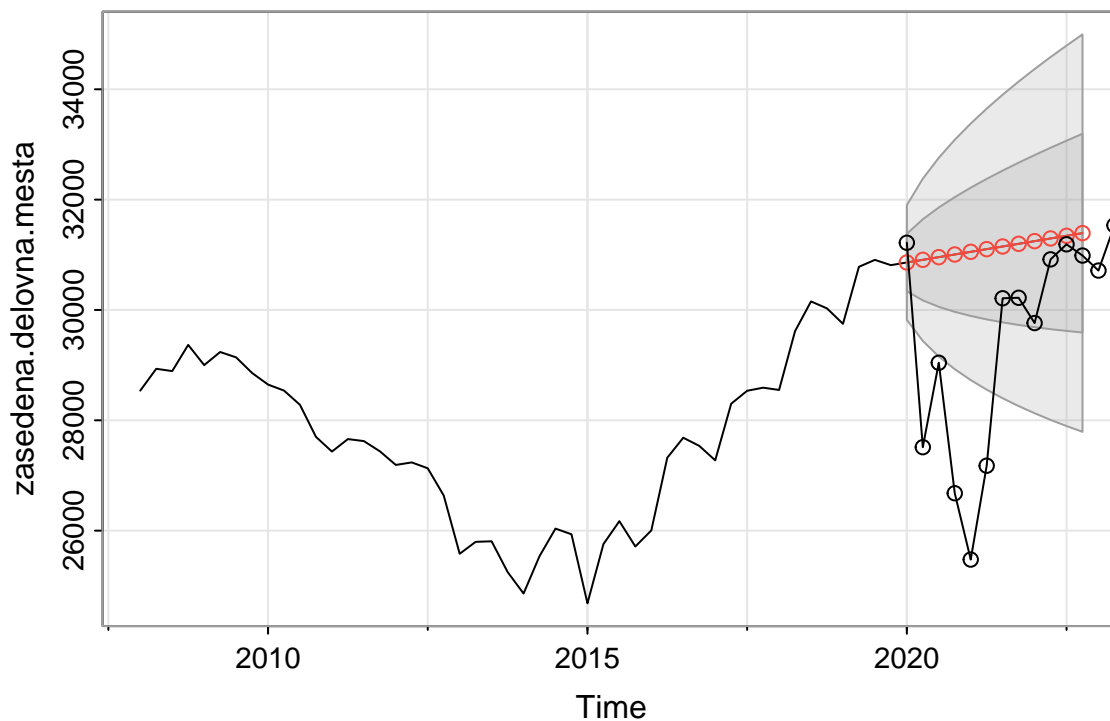


Figure 57: Časovna vrsta in napovedi na podlagi modela od leta 2020 naprej.

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.

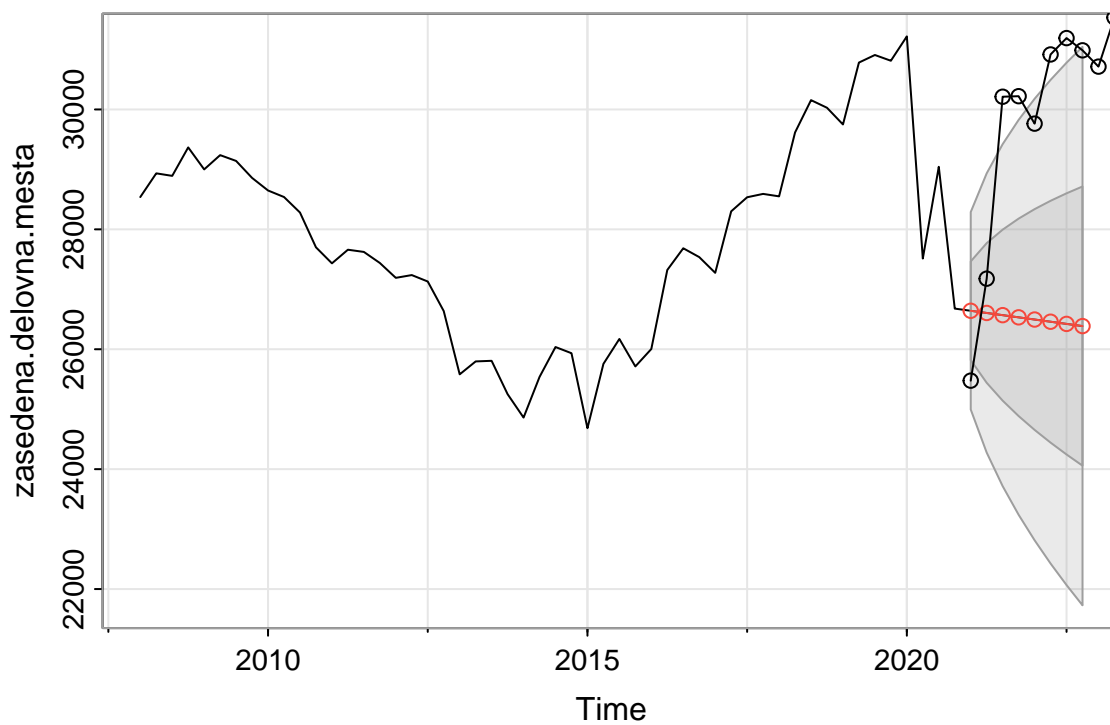


Figure 58: Časovna vrsta in napovedi na podlagi modela od leta 2021 naprej.

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.

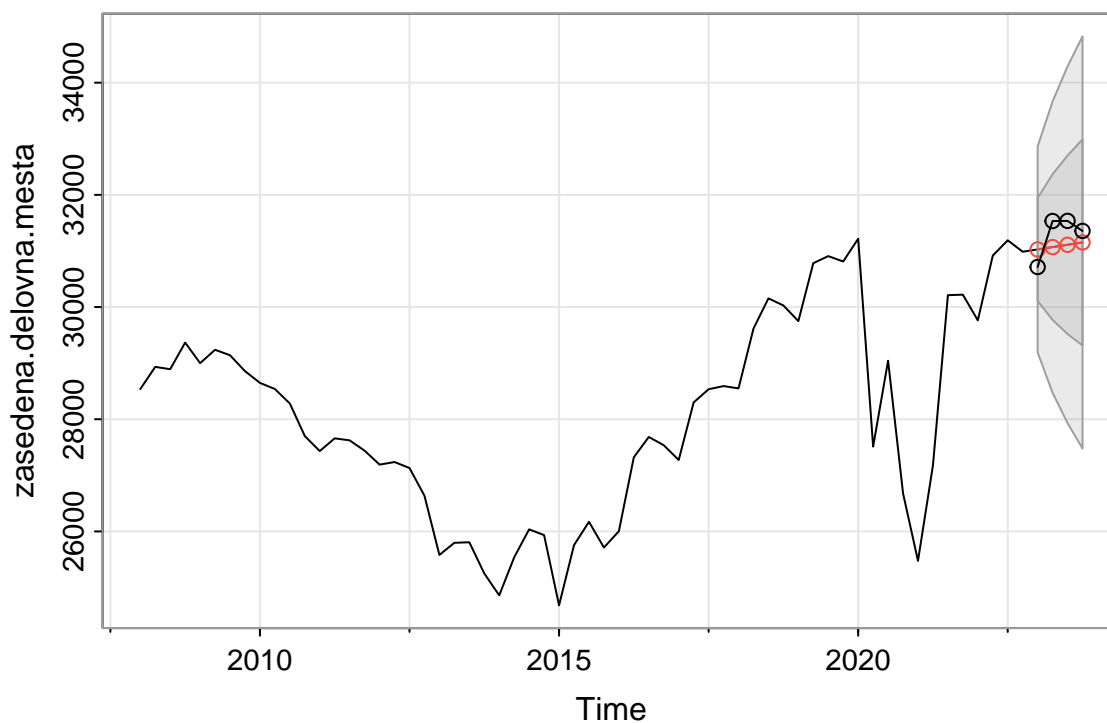


Figure 59: Časovna vrsta in napovedi na podlagi modela od leta 2023 naprej.

Napovedi tudi tukaj niso pravilne in ne ravno najbolj informativne, najboljše zadane samo za en trenutek naprej, 95% interavli 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.