

Model napovedi odjema električne energije

Matematika z računalnikom 2023/24

Karolina Šavli

Maj 2024

1 Uvod

Napovedovanje odjema in proizvodnje električne energije je ključno za zagotavljanje stabilnosti in učinkovitosti elektroenergetskega sistema. Pravilno načrtovanje proizvodnje in distribucije električne energije je pomembno, saj zmanjšuje tveganje za preobremenitve in izpade elektrike. V projektni nalogi bom analizirala in napovedovala odjem električne energije gospodinskih odjemalcev. Skozi celoten projekt sem uporabljala programski jezik Python ter glavne koncepte teorije časovnih vrst.

Glavni cilj projekta je sestaviti model, ki bo kar se da točno napovedal odjem električne energije za celotni nasleni dan, torej za naslednjih 24 ur.

1.1 Teoretična osnova

V projektni nalogi se bomo ukvarjali s časovno vrsto $\{X_t\}_t$, torej nekim zaporedjem podatkov. Zaradi lažjega razumevanja vsebine si najprej pogledajmo nekaj pomembnih konceptov v okviru časovnih vrst.

Pomemben postopek pri sestavljanju modelov je transformacija časovne vrste v stacionarno. Časovna vrsta je stacionarna, če ima podobne statistične značilnosti kot vsaka časovno zamaknjena časovna vrsta. Ne smemo torej imeti trenda ali sezonskosti. Stacionarnost lahko preverimo z različnimi formalnimi testi, kot so na primer Augmented Dickey-Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test in Phillips-Perron test. Pogosti pristopi, da pridemo do stacionarne časovne vrste so logaritmiranje časovne vrste, diferenciranje, sezonsko diferenciranje, ... Pri tem si pogosto pomagamo z obravnavo avtokorelacijske funkcije (ACF) in parcialno avtokorelacijske funkcije (PACF). Ti funkciji sta pomembni, saj opisujeta korelacijo med podatkovnimi točkami v časovni vrsti in prejšnjimi vrednostmi časovne vrste (za različne dolžine zamika). Iz teh funkciji lahko razberemo tudi, če imamo sezonsko komponento. V grobem, ACF meri linearno razmerje med opazovanjem in njegovimi prejšnjimi opazovanji pri različnih zamikih. PACF pa neposredno meri linearno razmerje med opazovanjem in njegovimi prejšnjimi opazovanji ob določenem zamiku, pri čemer so izključeni prispevki iz vmesnih zamikov. [4, 2]

Namen naše naloge je najti najbolj optimalen model. Za določitev le-tega bomo uporabili kriterij AIC (ang. *Akaike information criterion*), ki je ocenjevalec napake napovedi in s tem relativne kakovosti statističnih modelov za dani nabor podatkov.

1.1.1 Družina modelov ARMA

Ker bomo imeli opravka s časovno vrsto, katere lastnosti se s časom ne spreminjajo, bomo obravnavali model ARMA (ang. *autoregressive moving average process*), ki velja za enega najbolj znanih modelov in je osnova vseh modelov s podobnim imenom, ki jih bom v nadaljevanju navedla. ARMA(p,q) je združitev modelov AR(p) (ang. *Autoregressive model*) in MA(q) (ang. *Moving-average model*). Del AR vključuje regresijo spremenljivke na njene pretekle vrednosti. Del MA pa vključuje modeliranje napake kot linearne kombinacije napake, ki se pojavljajo v preteklosti. ARMA model je oblike

$$X_t = \mu + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (1)$$

kjer so $\varphi_1, \dots, \varphi_p$ ter $\theta_1, \dots, \theta_q$ parametri modela, $\varepsilon_t, \varepsilon_{t-1}, \dots$ pa je beli šum, ki je običajno normalno porazdeljen.

ARMA predpostavlja stacionarno vrsto. Če vrsta ni popolnoma stacionarna, kar je običajno razvidno že kar iz ACF in PACF, lahko uporabimo posplošitev modela ARMA, to je model ARIMA(p, d, q) (ang. *autoregressive integrated moving average*). Na kratko, ARIMA vrednosti podatkov nadomesti z diferenciranimi vrednostmi reda d, saj na tak način dobi stacionarne podatke. V primeru, da imamo sezonsko komponento, uporabimo model SARIMA (sezonska ARIMA). Parametre v modelu SARIMA(p, d, q)(P, D, Q)[S] lahko razdelimo na nesezonske (p, d, q) in sezonske (P, D, Q, S). p pove koliko korakov nazaj bomo gledali (preberemo iz grafa PACF), d določi kolikokrat smo diferencirali, q pa pove število zgodovinskih napak, ki vplivajo na trenutno vrednost (preberemo iz grafa ACF). Sezonski P predstavlja red sezonskega AR(P), D pove kolikokrat smo sezonsko diferencirali, Q določa red sezonskega MA(Q), S pa je perioda. [5]

Model SARIMA(p, d, q)(P, D, Q)[S] je odlike

$$X_t = \mu + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^P \phi_i X_{t-S} + \sum_{i=1}^Q \Theta_i \varepsilon_{t-S} + \varepsilon_t, \quad (2)$$

kjer so ϕ_1, \dots, ϕ_P in $\Theta_1, \dots, \Theta_Q$ dodatni parametri modela. Opazimo, da smo formuli (1) dodali moder del iz (2), da smo dobili SARIMA model.

Model lahko še dodatno nadgradimo, in sicer z vključitvijo eksogenih spremenljivk, torej neki zunanjih podatkov. Tako dobimo model SARIMAX(p, d, q)(P, D, Q)[S], ki je oblike

$$X_t = \mu + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^P \phi_i X_{t-S} + \sum_{i=1}^Q \Theta_i \varepsilon_{t-S} + \sum_{r=1}^R \beta_r Y_{t-r} + \varepsilon_t, \quad (3)$$

kjer je ob času t , Y_{t_i} i -ta eksogena spremenljivka s koeficientom β_i . [1]

1.1.2 Model GARCH

V modelih družine ARMA je privzeta konstantna varianca. Ni pa nujno, da je le-ta vedno konstantna, zato vpeljimo model GARCH, ki se običajno uporablja ravno na podatkih, ki nimajo konstantne variance. Model GARCH(p_G, q_G) je definiran kot

$$\sigma_t^2 = \omega + \sum_{i=1}^{p_G} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q_G} \beta_i \sigma_{t-i}^2, \quad (4)$$

kjer so $\alpha_1, \dots, \alpha_{p_G}$ in $\beta_1, \dots, \beta_{q_G}$ koeficienti, ki jih poskušamo določiti. σ_t^2 je pogojna varianca ob času t , ε_t pa predstavlja beli šum ob času t . [3]

1.1.3 Model SARIMAX-GARCH

Da v modelu iz družine ARMA, npr. modelu SARIMAX, opišemo tudi varianco, ki ni konstantna, in s tem dobimo bolj točen model, lahko konstruiramo model SARIMAX-GARCH. Da dobimo model združimo formul (3) in (4) ter dobimo

$$X_t = \mu + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^P \phi_i X_{t-S} + \sum_{i=1}^Q \Theta_i \varepsilon_{t-S} + \sum_{r=1}^R \beta_r Y_{t-r} + \sqrt{\sigma_t} z_t, \quad (5)$$

kjer je $z_t \sim \text{NEP}(0, 1)$. [3]

2 Osnovna analiza časovne vrste

Podjetje GEN-I je v obliki excel razpredelnice pripravilo tabelo podatkov, sestavljeno iz 7 stolpcev:

- `DateTimeStartUTC`: univerzalni koordinirani čas,
- `DateTimeStartCET`: srednjeevropski čas,
- `Odjem ACT`: neto odjem električne energije gospodinjskih odjemalcev (tistih brez sončne elektrarne) v kWh,
- `Temperatura ACT`: dejanska temperatura,
- `Temperatura FC`: s strani GEN-I napovedana temperatura,
- `Sevanje ACT`: dejansko sevanje in
- `Sevanje FC`: s strani GEN-I napovedano sevanje.

Slika 1: Tabela podatkov, 2021-2024

	<code>DateTimeStartUTC</code>	<code>DateTimeStartCET</code>	<code>Odjem ACT</code>	<code>Temperatura ACT</code>	<code>Temperatura FC</code>	<code>Sevanje ACT</code>	<code>Sevanje FC</code>
0	31.10.2021 23:00	2021-01-11 00:00:00	0.000010	3.60	5.300	0.0	0.0
1	31.10.2021 23:15	2021-01-11 00:15:00	0.000009	3.60	5.300	0.0	0.0
2	31.10.2021 23:30	2021-01-11 00:30:00	0.000009	3.60	5.300	0.0	0.0
3	31.10.2021 23:45	2021-01-11 00:45:00	0.000009	3.60	5.300	0.0	0.0
4	1.11.2021 00:00	2021-01-11 01:00:00	0.000008	3.45	5.300	0.0	0.0
...
80059	12.02.2024 21:45	2024-12-02 22:45:00	0.000012	8.00	3.375	0.0	0.0
80060	12.02.2024 22:00	2024-12-02 23:00:00	0.000012	7.90	3.200	0.0	0.0
80061	12.02.2024 22:15	2024-12-02 23:15:00	0.000011	7.80	3.300	0.0	0.0
80062	12.02.2024 22:30	2024-12-02 23:30:00	0.000011	7.15	3.400	0.0	0.0
80063	12.02.2024 22:45	2024-12-02 23:45:00	0.000009	6.50	3.500	0.0	0.0

V analizi sem uporabljala vse stolpce, razen stolpca `DateTimeStartUTC`, saj je v okviru časa bolj relevanten stolpec `DateTimeStartCET`. Podatki so podani za obdobje od 1. novembra 2021 do 12. februarja 2024, na vsakih 15 minut. Tabela ima vsega skupaj 80064 vrstic.

2.1 Odjem električne energije

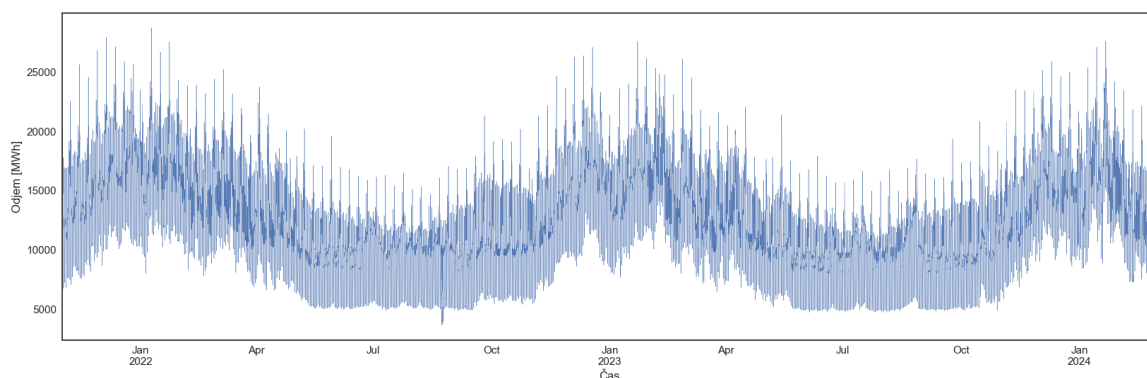
Slika 2 prikazuje odjem električne energije za obdobje od 1. novembra 2021 do 12. februarja 2024. Opazna je sezonskost; odjem je znatno večji jeseni in pozimi, zaradi povečane uporabe energije za ogrevanje in razsvetljevo (saj se število ur dnevne svetlobe podaljša).

S Tabele 1 preberemo, da je povprečna poraba električne energije gospodinjskih odjemalcev okrog 12240,53 MWh, minimalna dosežena vrednost je 3629,32 MWh, maksimalna pa 28736,80 MWh. Vrednosti varirajo za okrog 4167,98 MWh.

Tabela 1: Opisne statistike porabe električne energije, 2021-2024

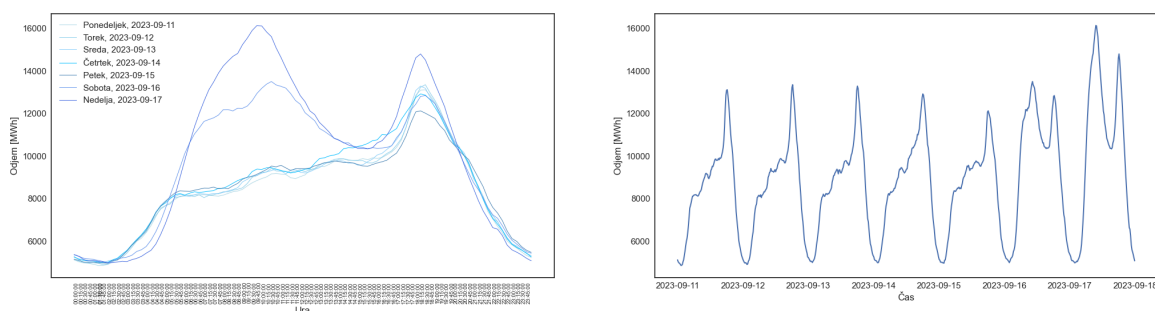
	Min	Max	Povprečje	Mediana	Standardni odklon
Odjem [MWh]	3629,32	28736,80	12240,53	11708,50	4167,98

Slika 2: Odjem električne energije, 2021-2024



Odjem električne energije si pogledjmo še na ravni tedna. Opazimo, da je od ponedeljka do petka odjem najmanjši ponoči in se večja vse do viška okrog 18 ure, nato pa hitro pade. V soboto in nedeljo pa je prvi višek porabe dopoldne, drugi pa okrog 18 ure, kar je opazno s Slike 3, ki prikazuje odjem električne energije v drugem tednu septembra 2023. Imamo torej sezonskost na dnevni ravni.

Slika 3: Odjem električne energije po urah, drugi teden septembra 2023

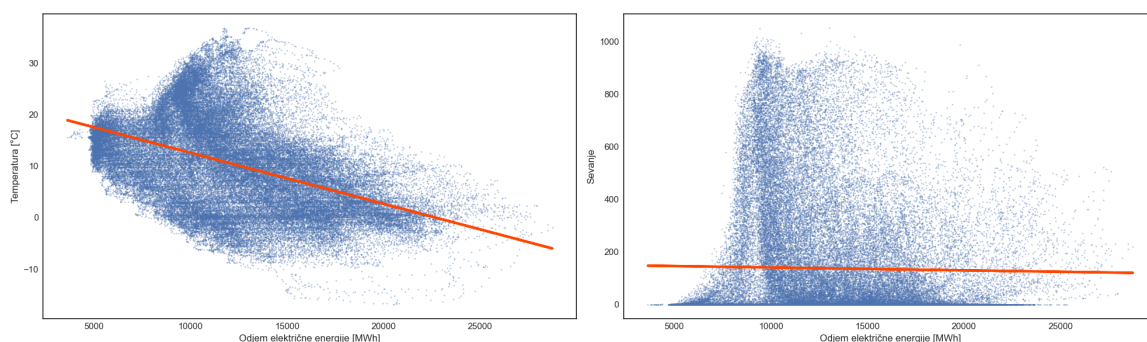


Zaključimo lahko, da je naša časovna vrsta visokofrekvenčna, ima sezonsko komponentno ter njeno povprečje ni konstantno.

2.2 Povezava med odjemom in temperaturo ter sevanjem

Smislno se zdi, da obstaja povezava med odjemom električne energije ter temperaturo in sevanjem. S pomočjo Slike 4 ugotovimo, da nižja temperatura pomeni višji odjem in obratno. Povezave med sevanjem in odjemom pa ni opaziti; sklepam da zato, ker ne analiziramo samooskrbnih odjemalcev (tisti ki imajo svojo sončno elektrarno).

Slika 4: Povezava med odjemom in temperaturo ter sevanjem, 2021-2024



V nadaljevanju bomo v upanju boljše napovedi v naše modele kot eksogeni spremenljivki vključili tudi temperaturo in sevanje.

3 Napredna analiza in izbira modela

3.1 Izbira družine modelov

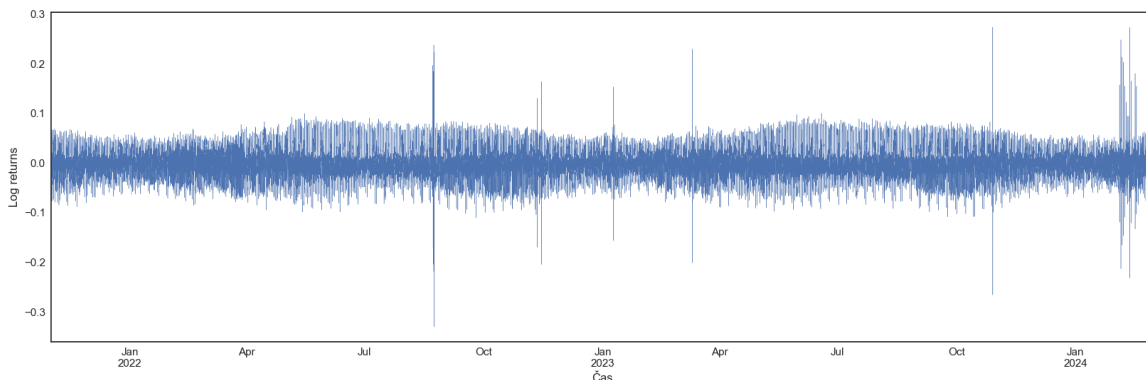
Ker pa imamo v podatkih očitno sezonskost, lahko ARIMO nadgradimo v SARIMO, ki dodatno upošteva še sezone vzorce. Ker bi radi vključili tudi eksogeni spremenljivki (temperatura in sevanje), model dodatno nadgradimo v model SARIMAX. Vemo pa, da ima ta družina modelov težavo predvsem pri napovedih časovnih vrst, ki se jim skozi čas spreminja varianca. Da bo naše napovedovanje bolj učinkovito, bomo modelu SARIMAX povezali z modelom GARCH, ki se uporablja ravno za modeliranje volativnih podatkov. [3] Konstruirati torej želimo model SARIMAX-GARCH, ki smo ga že opisali formulo (5).

3.2 Odstranitev sezonskosti in pridobitev stacionarnosti

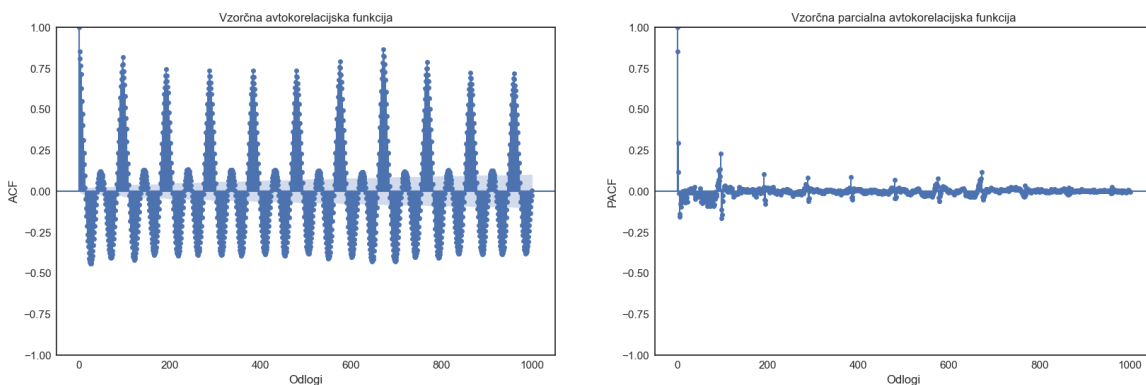
Da bomo lahko indentificiralni potencialne modele iz družine ARMA, moramo najprej originalno časovno vrsto (Slika 2) narediti stacionarno.

Ker so naši podatki volatilni najprej naredimo **logaritmične donose** (ang. *log returns*). Če z W_t označimo originalno časovno vrsto odjema električne energije, so logaritmični donosi definirani kot časovna vrsta $Y_t = \ln\left(\frac{W_t}{W_{t-1}}\right)$. Slednja časovna vrsta je prikazana na Slika 5.

Slika 5: Logaritmični donosi odjema električne energije, 2021-2024



Slika 6: Vzorčna avtokorelacijska in parcialna avtokorelacijska funkcija časovne vrste logaritmičnih donosov

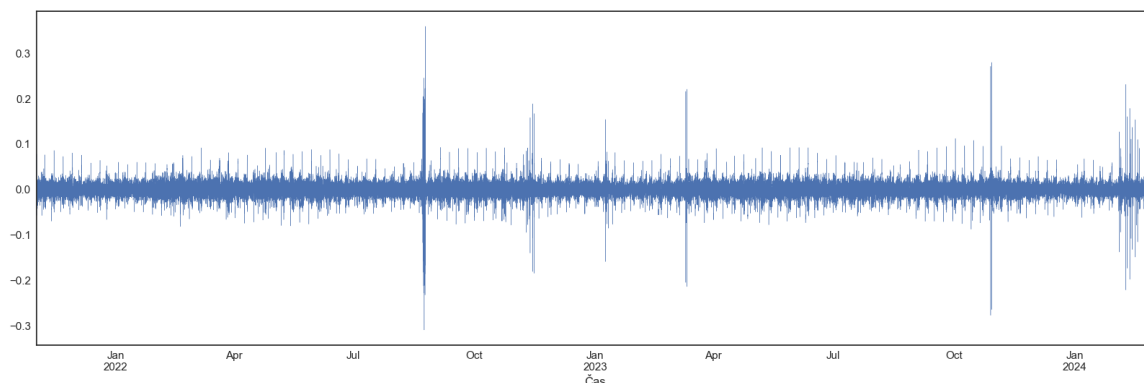


Slika 6 prikazuje ACF in PACF logaritmičnih donosov do odloga 1000 (to je malo manj kot dva tedna). Z grafov je očitno, da časovna vrsta ni stacionarna. Opazna je sezonska komponentna, in sicer 96, kar je ravno en dan ¹.

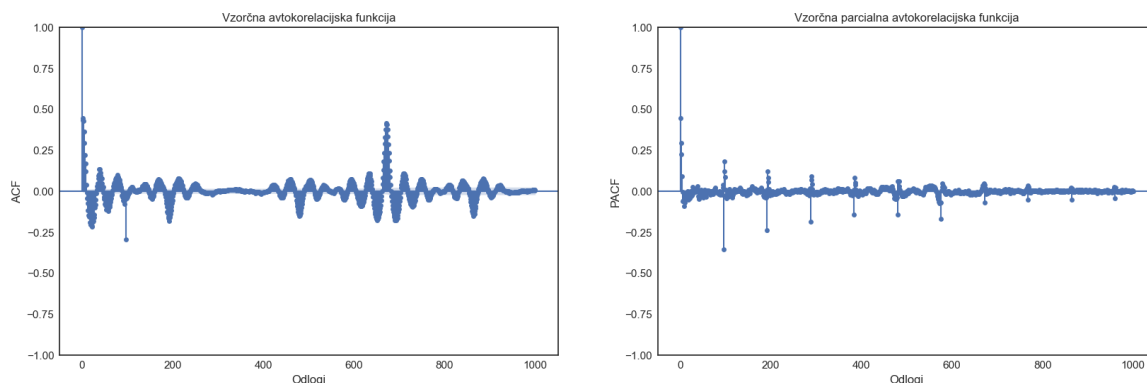
¹Podatki so podani na 15 minut in $15 \text{ min} \cdot 96 = 1440 \text{ min}$, kar je ravno en dan.

Da dosežemo stacionarnost, se je potrebno sezonskosti znebiti, zato bomo podatke **sezonsko diferencirali**. Nova časovna vrsta bo $Z_t = Y_t - Y_{t-96}$. Prikazana je na Sliki 7, njeni ACF in PACF pa na Sliki 8.

Slika 7: Časovna vrsta po sezonskem diferenciranju, 2021-2024

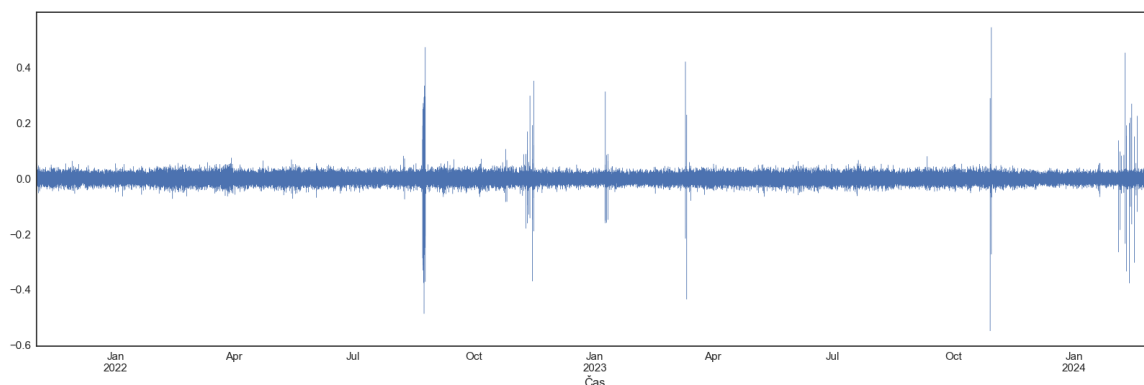


Slika 8: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste po sezonskem diferenciranju



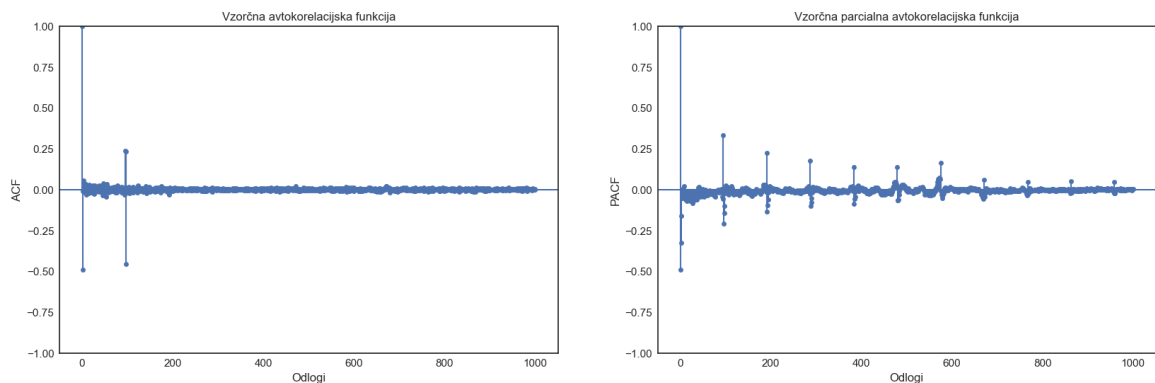
Na grafu ACF opazimo ponavljanje vzorca, kar pomeni, da časovna vrsta Z_t še kar ni stacionarna. Še enkrat jo diferenciramo, tokrat nesezonsko, in dobimo vrsto $X_t = Z_t - Z_{t-1}$. Prikazana je na Sliki 9, njeni ACF in PACF pa na Sliki 10.

Slika 9: Časovna vrsta po prvem nesezonskem diferenciranju, 2021-2024

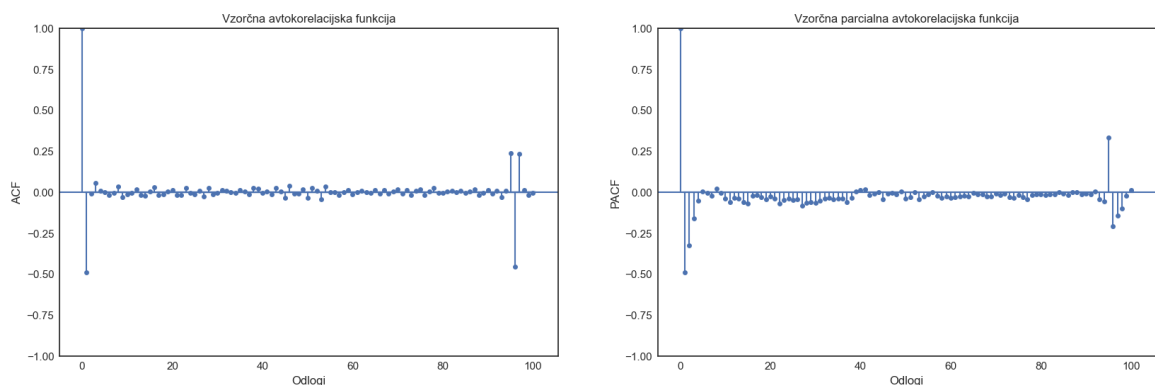


Dobljena časovna vrsta X_t na podlagi ACF in PACF (Slika 10) zgleda stacionarno in tudi formalni testi ADF, KPSS in Phillips-Perron potrdijo njeno stacionarnost. Časovna vrsta je torej primerna za izbero parametrov modela SARIMAX.

Slika 10: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste po prvem nesezonskem diferenciranju



Slika 11: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste po prvem nesezonskem diferenciranju, odlogi do 100



3.3 Identifikacija modela SARIMAX

Na podlagi ACF in PACF (Slika 10) stacionarne časovne vrste X_t , izberimo parametre modela SARIMAX(p, d, q)(P, D, Q)[S], ki je podan z enačbo (3).

Ker smo časovno vrsto enkrat sezonsko diferencirali bo $D = 1$ in ker smo jo enkrat navadno diferencirali bo $d = 1$. Perioda je enaka 96, torej je $S = 96$.

Za določitev sezonskih parametrov P in Q gledamo korelacije pri odlogih, ki so večkratniki periode S . Pri PACF visoko korelacijo opazimo predvsem pri 96 in 192, nato pa se z vsako dodatno periodo manjša. Parameter P je torej 1 ali več; predlagala bi 1, 2 ali 3. Pri ACF pa je občutna korelacija zgolj pri 96, zato izberemo $Q = 1$.

Za lažjo določitev nesezonskih parametrov p in q bomo gledali ACF in PACF do prve periode (torej do odloga 96). Slednje je prikazano na Sliki 11. Tako iz PACF, kot tudi iz ACF, je opazna večja korelacija pri prvih nekaj urah in v uri tik pred periodo. V modelu bomo zato zagotovo vključili prvih nekaj 15-minutnih intervalov, saj le-ti kažejo močan vpliv.

Izbrala sem nekaj potencialnih modelov SARIMAX, kjer sem kot eksogeni spremenljivki podala temperaturo in sevanje. Modele sem trenirala na 75 % podatkov, to je od 1. novembra 2021 do 4. avgusta 2023. Na preostanku podatkov pa sem izvajala teste napovedi. Vrednosti kriterija AIC za posamezne modele so prikazane v Tabeli 2. Najmanjšo vrednostjo kriterija AIC ima model SARIMAX(4, 1, 5)(0, 1, 0)[96].

S Tabele 2 je razvidno, da sem za sezonska parametra P in Q v vseh primerih izbrala vrednost 0, čeprav predvidevam, da bi bile bolj optimalne višje vrednosti. Te modele sem poskusila zagnati, vendar moj računalnik ni vrnil nobenega rezultata, ker očitno ni dovolj zmogljiv.

Tabela 2: Vrednost kriterija AIC za izbrane modele iz družine SARIMAX

Model	AIC
SARIMAX(1,1,0)(0,1,0)[96]	-338588,046
SARIMAX(0,1,1)(0,1,0)[96]	-346718,519
SARIMAX(1,1,1)(0,1,0)[96]	-346768,583
SARIMAX(2,1,1)(0,1,0)[96]	-347555,507
SARIMAX(3,1,2)(0,1,0)[96]	-347573,394
SARIMAX(4,1,3)(0,1,0)[96]	-278981,698
SARIMAX(5,1,4)(0,1,0)[96]	-345287,413
SARIMAX(5,1,5)(0,1,0)[96]	-345310,105
SARIMAX(4,1,5)(0,1,0)[96]	-347619,304
SARIMAX(6,1,5)(0,1,0)[96]	-345325,842
SARIMAX(6,1,6)(0,1,0)[96]	-345342,972
SARIMAX(5,1,6)(0,1,0)[96]	-345351,794

Izbran imamo model SARIMAX, v naslednjem koraku pa na podoben način izberimo še model GARCH.

3.4 Izbira modela SARIMAX-GARCH

Model GARCH(p, q) bomo konstruirali na rezidualih izbranega modela SARIMAX. Preizkusili bomo 6 kombinacij parametrov p in q, in sicer (1, 1), (1, 2), (2, 1), (2, 2), (1, 3) ter (3, 1). Na podlagi vrednosti kriterija AIC se odločimo za model **SARIMAX(4, 1, 5)(0, 1, 0)[96]-GARCH(1, 3)**, saj je iz Tabele 3 razvidno, da ima najmanjšo vrednost kriterija AIC.

Tabela 3: Vrednost kriterija AIC za izbrane modele SARIMAX(4, 1, 5)(0, 1, 0)[96]-GARCH(p, q)

SARIMAX-GARCH(p,q)	AIC
(0, 0)	-347619,304
(1, 1)	-365754,448
(1, 2)	-366056,218
(2, 1)	-365434,577
(2, 2)	-365843,673
(1, 3)	-366134,335
(3, 1)	-365252,576

3.5 Izbran model

Na podlagi vrednosti kriterija AIC smo se odločili za SARIMAX(4, 1, 5)(0, 1, 0)[96]-GARCH(1, 3). Ključni koeficienti izbranega modela v formuli (5) so:

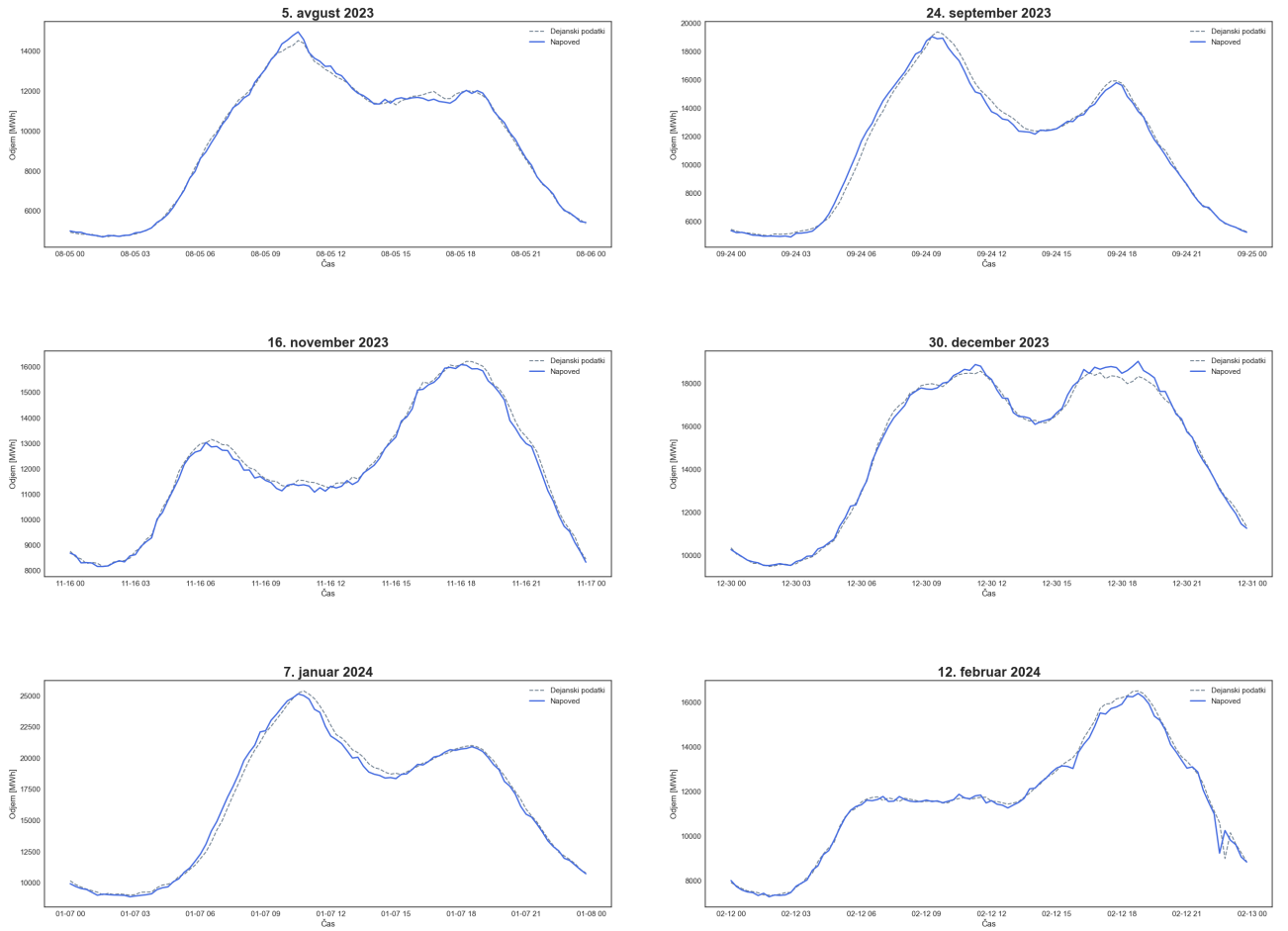
- AR del: $\varphi_1 = -0,3686$, $\varphi_2 = -0,4185$, $\varphi_3 = -0,2894$, $\varphi_4 = 0,0486$
- MA del: $\theta_1 = -0,3311$, $\theta_2 = 0,1863$, $\theta_3 = 0,0931$, $\theta_4 = -0,2059$, $\theta_5 = 0,0583$
- sezonskih del: parametrov nimamo, ker sta P in Q enaka 0
- GARCH del: $\alpha_1 = 0,2$, $\beta_1 = 0,2333$, $\beta_2 = 0,2333$, $\beta_3 = 0,2333$
- Eksogene spremenljivke: koeficient pri temperaturi je enak $-0,0003$, pri sevanju pa $3,747 \cdot 10^{-6}$

Sevanje ima torej pri modelu napovedi zelo majhen vpliv, vendar sem preverila, da je model z vključitvijo sevanja še vedno malce boljši. Model je nasploh z vključitvijo eksogenih spremenljivk preverjeno bolj točen.

4 Testiranje izbranega modela SARIMAX-GARCH

Izbran model SARIMAX-GARCH uporabimo za napovedovanje odjema električne energije za različne dni v testnem obdobju (od 5. avgusta 2023 do 12. februarja 2024). Izbrala sem si 6 različnih dni, narisala njihove napovedi ter jih prikazala na Sliki 12.

Slika 12: Napovedi odjema električne energije za izbrane datume



Prileganje napovedi dejanski vrednosti se zdi zelo dobra, vseeno pa si pogledjmo napaki RMSE (ang. *Root-mean-square deviation*) in MAPE (ang. *Mean absolute percentage error*), ki ju izračunamo po formulah:

$$RMSE = \sqrt{\frac{1}{96} \sum_{t=1}^{96} (W_t - \hat{W}_t)^2} \quad \text{in} \quad MAPE = \frac{1}{96} \sum_{t=1}^{96} \frac{W_t - \hat{W}_t}{W_t},$$

kjer je W_t dejanska vrednost odjema, \hat{W}_t pa napovedana.

Vrednosti napak RMSE in MAPE za izbrane napovedi so zapisane v Tabeli 4 in v povprečju znašajo RMSE 261,926, MAPE pa 1,478 %.

Napaki RMSE in MAPE sem izračunala še posebej za zadnji teden septembra 2023 in prvi teden februarja 2024. Rezultati so prikazani v Tabeli 5 in v Tabeli 6. Iz Tabele 5 zglada kot da se večja napaka pojavi za vikend, vendar glede na rezultat iz Tabele 6 ne moremo tega z gotovostjo trditi.

Tabela 4: Vrednost napake RMSE in MAPE za izbrane datume

Datum	RMSE	MAPE
5.avg.23	156,600	1,104
24.sep.23	365,372	2,392
16.nov.23	176,901	1,199
30.dec.23	224,123	1,046
7.jan.23	399,578	1,839
12.feb.23	248,981	1,287
Povprečje	261,926	1,478

Tabela 5: Vrednost napake RMSE in MAPE zadnji teden septembra 2023

	Ponedeljek	Torek	Sreda	Četrtek	Petek	Sobota	Nedelja	Povprečje
RMSE	98,768	73,987	84,263	90,712	182,816	220,022	336,630	155,314
MAPE	0,813	0,707	0,800	0,773	1,442	1,679	2,406	1,231

Tabela 6: Vrednost napake RMSE in MAPE prvi teden februarja 2024

	Ponedeljek	Torek	Sreda	Četrtek	Petek	Sobota	Nedelja	Povprečje
RMSE	136,451	129,370	418,558	324,463	168,860	353,170	420,640	278,788
MAPE	0,867	0,817	2,041	1,522	1,108	1,670	2,377	1,486

5 Zaključek

S pomočjo transformacije podatkov, preverjanja ACF, PACF ter vrednosti kriterija AIC pridemo do modela napovedi odjema električne energije gospodinjstkih odjemalcev. Kot optimalen model izberemo SARIMAX(4, 1, 5)(0, 1, 0)[96]-GARCH(1, 3).

Napoved izbranega modela dobro opisuje dejanske vrednosti, kar je lepo razvidno iz grafov napovedi. Formalno, povprečna napaka RMSE znaša okrog 260, MAPE pa 1,5 %. Tekom dela ugotovimo tudi, da vključitev eksogenih podatkov prispeva k bolj točni napovedi.

Če bi modele testirala s pomočjo bolj zmogljivega računalnika, sem prepričana, da bi prišla do vsaj malo boljših napovedi. Kljub vsemu pa so rezultati napovedi dobri in predvsem smiselni.

Literatura

- [1] B. Artley. Time Series Forecasting with ARIMA , SARIMA and SARIMAX. *Towards Data Science*, Apr. 2022.
- [2] K. Brilej. *Napovedovanje cen in povpraševanja po električni energiji: magistrsko delo*. PhD thesis, 2021.
- [3] T. Dierckx. ARIMA-GARCH forecasting with Python. *Medium*, Sept. 2020.
- [4] V. K. G. Statistical Tests to Check Stationarity in Time Series. *Medium*, May 2024.
- [5] B. Köseoğlu. Guide to Time Series Analysis with Python — 4: ARIMA and SARIMA. *Medium*, Jan. 2024.