

MODEL NAPOVEDI ODJEMA ELEKTRIČNE ENERGIJE

Matematika z računalnikom 2023/24

Karolina Šavli

Mentorstvo: dr. Blaž Krese in Rok Geršak iz GEN-I

6. junij 2024

Kratek pregled predstavitve

- 1 Podatki in cilj projektne naloge
- 2 Osnovna analiza podatkov
 - Odjem električne energije
 - Povezava med odjemom in temperaturo ter sevanjem
- 3 Napredna analiza podatkov
 - Izbira družine modelov
 - Odstranitev sezonskosti in pridobitev stacionarnosti
 - Identifikacija modela SARIMAX
 - Izbira modela SARIMAX-GARCH
 - Izbiran model
- 4 Testiranje modela
- 5 Zaključek

Predstavitev podatkov

- Podjetje GEN-I je pripravilo tabelo podatkov:

| | DateTimeStartUTC | DateTimeStartCET | Odjem ACT | Temperatura ACT | Temperatura FC | Sevanje ACT | Sevanje FC |
|-------|------------------|------------------|-----------|-----------------|----------------|-------------|------------|
| 0 | 31.10.2021 23:00 | 1.11.2021 00:00 | 0.000010 | 3.60 | 5.300 | 0.0 | 0.0 |
| 1 | 31.10.2021 23:15 | 1.11.2021 00:15 | 0.000009 | 3.60 | 5.300 | 0.0 | 0.0 |
| 2 | 31.10.2021 23:30 | 1.11.2021 00:30 | 0.000009 | 3.60 | 5.300 | 0.0 | 0.0 |
| 3 | 31.10.2021 23:45 | 1.11.2021 00:45 | 0.000009 | 3.60 | 5.300 | 0.0 | 0.0 |
| 4 | 1.11.2021 00:00 | 1.11.2021 01:00 | 0.000008 | 3.45 | 5.300 | 0.0 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 81691 | 29.02.2024 21:45 | 29.02.2024 22:45 | 0.000012 | 6.80 | 7.475 | 0.0 | 0.0 |
| 81692 | 29.02.2024 22:00 | 29.02.2024 23:00 | 0.000011 | 6.65 | 7.300 | 0.0 | 0.0 |
| 81693 | 29.02.2024 22:15 | 29.02.2024 23:15 | 0.000011 | 6.50 | 7.300 | 0.0 | 0.0 |
| 81694 | 29.02.2024 22:30 | 29.02.2024 23:30 | 0.000010 | 6.50 | 7.300 | 0.0 | 0.0 |
| 81695 | 29.02.2024 22:45 | 29.02.2024 23:45 | 0.000010 | 6.50 | 7.300 | 0.0 | 0.0 |

- Obdobje od 1. novembra 2021 do 29. februarja 2024, na 15 minut
- Stolpec Odjem ACT sem pomnožila s 10^6

Predstavitev podatkov

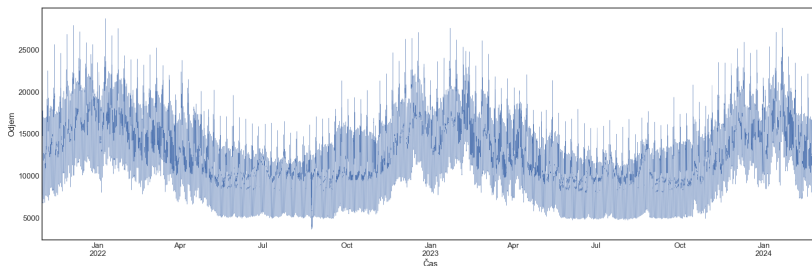
- Podjetje GEN-I je pripravilo tabelo podatkov:

| | DateTimeStartUTC | DateTimeStartCET | Odjem ACT | Temperatura ACT | Temperatura FC | Sevanje ACT | Sevanje FC |
|-------|------------------|------------------|-----------|-----------------|----------------|-------------|------------|
| 0 | 31.10.2021 23:00 | 1.11.2021 00:00 | 0.000010 | 3.60 | 5.300 | 0.0 | 0.0 |
| 1 | 31.10.2021 23:15 | 1.11.2021 00:15 | 0.000009 | 3.60 | 5.300 | 0.0 | 0.0 |
| 2 | 31.10.2021 23:30 | 1.11.2021 00:30 | 0.000009 | 3.60 | 5.300 | 0.0 | 0.0 |
| 3 | 31.10.2021 23:45 | 1.11.2021 00:45 | 0.000009 | 3.60 | 5.300 | 0.0 | 0.0 |
| 4 | 1.11.2021 00:00 | 1.11.2021 01:00 | 0.000008 | 3.45 | 5.300 | 0.0 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 81691 | 29.02.2024 21:45 | 29.02.2024 22:45 | 0.000012 | 6.80 | 7.475 | 0.0 | 0.0 |
| 81692 | 29.02.2024 22:00 | 29.02.2024 23:00 | 0.000011 | 6.65 | 7.300 | 0.0 | 0.0 |
| 81693 | 29.02.2024 22:15 | 29.02.2024 23:15 | 0.000011 | 6.50 | 7.300 | 0.0 | 0.0 |
| 81694 | 29.02.2024 22:30 | 29.02.2024 23:30 | 0.000010 | 6.50 | 7.300 | 0.0 | 0.0 |
| 81695 | 29.02.2024 22:45 | 29.02.2024 23:45 | 0.000010 | 6.50 | 7.300 | 0.0 | 0.0 |

- Obdobje od 1. novembra 2021 do 29. februarja 2024, na 15 minut
- Stolpec Odjem ACT sem pomnožila s 10^6
- Cilj projekta:** sestaviti model, ki bo kar se da točno napovedal odjem električne energije **gospodinjskih odjemalcev** za naslednji dan

Odjem električne energije

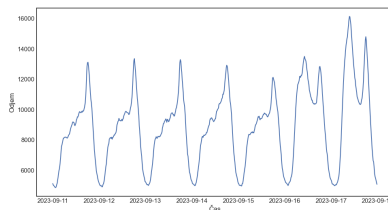
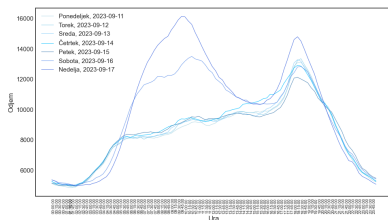
Slika: Odjem električne energije, 2021-2024



- Odjem je znatno večji jeseni in pozimi

Odjem električne energije na ravni tedna

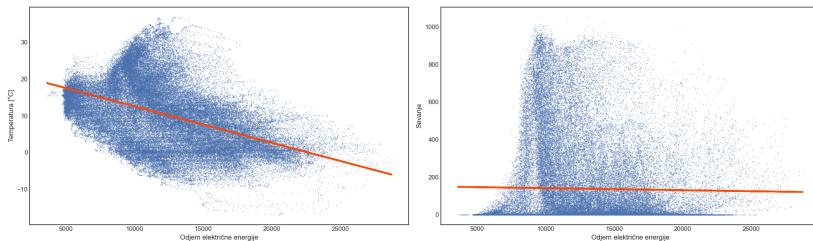
Slika: Odjem električne energije po urah, drugi teden septembra 2023



- Med tednom: en višek
- Vikend: dva viška
- Sezonskost na dnevni ravni

Povezava med odjemom in temperaturo ter sevanjem

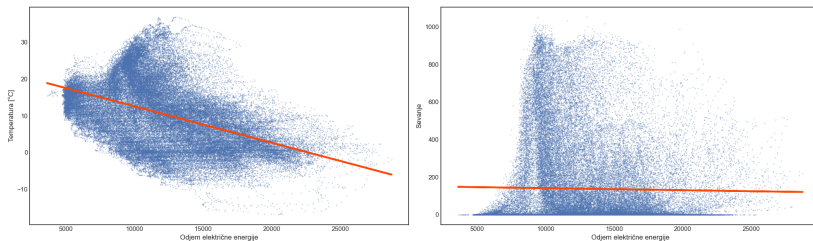
Slika: Povezava med odjemom in temperaturo ter sevanjem, 2021-2024



- \uparrow temperatura \rightarrow odjem \downarrow
- Povezava s sevanjem ni razvidna

Povezava med odjemom in temperaturo ter sevanjem

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



- \uparrow temperatura \rightarrow odjem \downarrow
- Povezava s sevanjem ni razvidna
- Temperaturo in sevanje bomo v model vključili kot eksogeni spremenljivki

Izbira družine modelov

- Gre za časovno vrsto → uporaba teorije časovnih vrst

Izbira družine modelov

- Gre za časovno vrsto → uporaba teorije časovnih vrst
- Časovna vrsta odjema: visofrekvenčna, ima sezonsko komponento in njeno povprečje ni konstantno
- Družina modelov ARMA

Družina modelov ARMA

- Model ARMA

Družina modelov ARMA

- Model ARMA
- Nimamo popolnoma stacionarnih podatkov \rightarrow ARIMA

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

Družina modelov ARMA

- Model ARMA
- Nimamo popolnoma stacionarnih podatkov \rightarrow ARIMA

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

- Imamo sezonsko komponentno \rightarrow SARIMA

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

Družina modelov ARMA

- Model ARMA
- Nimamo popolnoma stacionarnih podatkov \rightarrow ARIMA

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

- Imamo sezonsko komponentno \rightarrow SARIMA

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

- Vključitev eksogenih podatkov \rightarrow SARIMAX

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

Model SARIMAX-GARCH

- Družina modelov ARMA ima težavo predvsem pri napovedih časovnih vrst, ki se jim skozi čas spreminja varianca → povezava z modelom GARCH

Model SARIMAX-GARCH

- Družina modelov ARMA ima težavo predvsem pri napovedih časovnih vrst, ki se jim skozi čas spreminja varianca → povezava z modelom **GARCH**
- Dobimo model SARIMAX(p,d,q)(P,D,Q)[S]-GARCH(p_G, q_G):

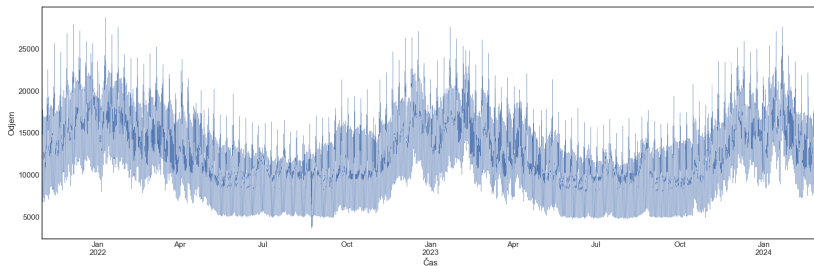
$$X_t = \mu + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^P \phi_i X_{t-S} + \sum_{i=1}^Q \Theta_i \varepsilon_{t-S} + \sum_{r=1}^R \beta_r Y_{it} + \sqrt{\sigma_t} z_t,$$

kjer je $\sigma_t^2 = \omega + \sum_{i=1}^{p_G} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q_G} \beta_i \sigma_{t-i}^2$, σ_t^2 je pogojna varianca ob času t , ε_t pa beli šum ob času t in $z_t \sim \text{NEP}(0, 1)$.

Originalna časovna vrsta

- $W_t \dots$ originalna časovna vrsta odjema električne energije

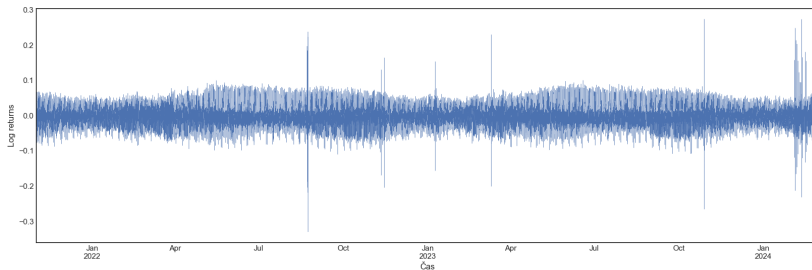
Slika: Odjem električne energije, 2021-2024



- Podatki so volatilni \rightarrow naredimo **logaritmične donose** (ang. *log returns*): $Y_t = \ln \left(\frac{W_t}{W_{t-1}} \right)$

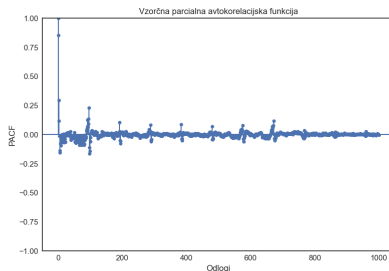
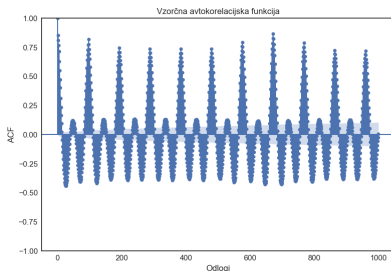
Log returns

Slika: Logaritmični donosi odjema električne energije, 2021-2024



ACF in PACF za Log returns

Slika: Vzorčna avtokorelacijska in parcialna avtokorelacijska funkcija logaritmičnih donosov

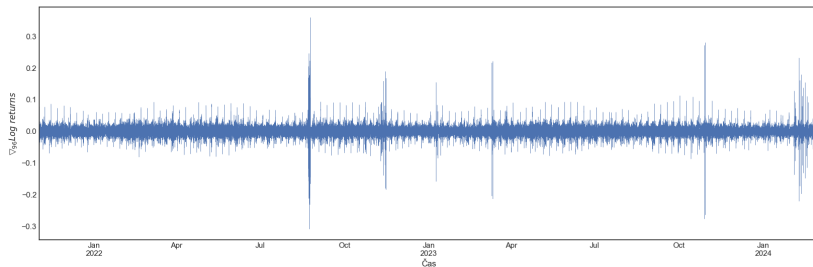


- Časovna vrsta **ni** stacionarna
- Sezonska komponentna: 96 (ravno en dan) → sezonsko diferenciramo

Sezonska diferenciacija

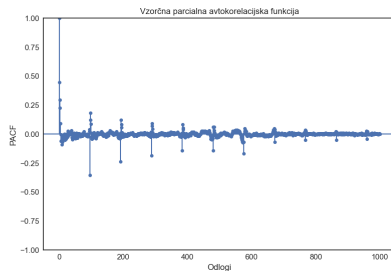
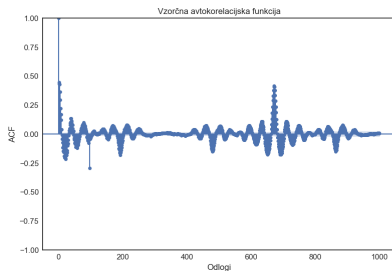
- Podatke **sezonsko diferenciramo**
- Nova časovna vrsta: $Z_t = Y_t - Y_{t-96} = \nabla_{96}Y_t$.

Slika: Časovna vrsta Z_t , 2021-2024



ACF in PACF časovne vrste Z_t

Slika: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste Z_t

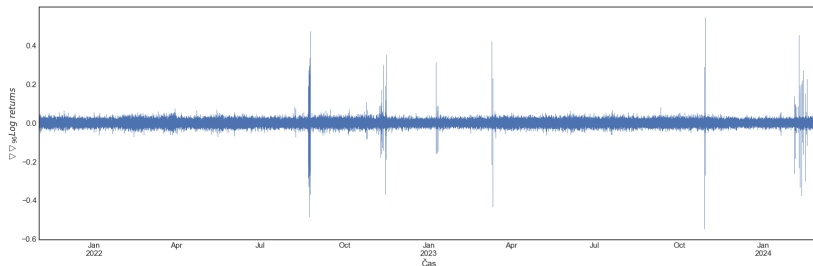


- Ponavljanje vzorca \rightarrow vrsta še kar ni stacionarna \rightarrow *navadno* diferenciramo

Navadna diferenciacija

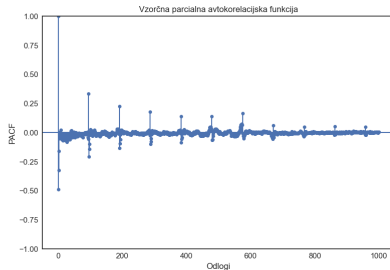
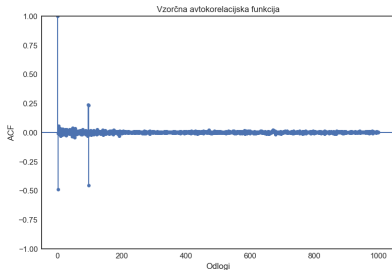
- Dobimo časovno vrsto $X_t = Z_t - Z_{t-1} = \nabla Z_t$

Slika: Časovna vrsta X_t , 2021-2024



ACF in PACF časovne vrste X_t

Slika: Vzorčna avtokorelacijska in parcialna avtokorelacijska funkcija časovne vrste X_t



- Časovna vrsta je stacionarna

Kaj smo torej naredili z originalno časovno vrsto?

Originalna časovna vrsta → Log returns → sezonsko diferencirali →
Navadno diferencirali

Kaj smo torej naredili z originalno časovno vrsto?

Originalna časovna vrsta → Log returns → sezonsko diferencirali →
Navadno diferencirali

Naslednji korak:

S pomočjo ACF in PACF stacionarne časovne vrste X_t določimo parametre modela SARIMAX

Določitev parametrov modela SARIMAX

SARIMAX(p,d,q)(P,D,Q)[S]

- Enkrat smo sezonsko diferencirali $\rightarrow D = 1$
- Enkrat smo navadno diferencirali $\rightarrow d = 1$
- Perioda je 96 $\rightarrow S = 96$

Določitev parametrov modela SARIMAX

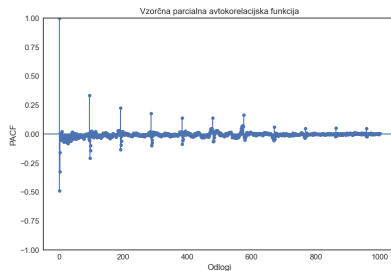
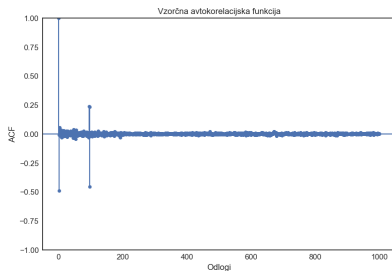
SARIMAX(p, d, q)(P, D, Q)[S]

- Enkrat smo sezonsko diferencirali $\rightarrow D = 1$
- Enkrat smo navadno diferencirali $\rightarrow d = 1$
- Perioda je 96 $\rightarrow S = 96$

Za določitev **sezonskih** parametrov P in Q gledamo korelacije pri odlogih, ki so večkratniki periode S

SARIMAX(**p,1,q**)(**P,1,Q**)[96]

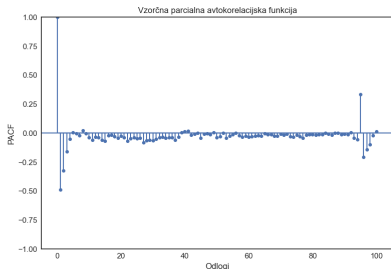
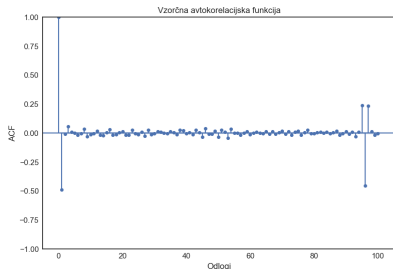
Slika: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste X_t



- PACF \rightarrow **P** je vsaj 1
- ACF \rightarrow **Q** = 1

SARIMAX(p,1,q)(P,1,1)[96]

Slika: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija vrste X_t , odlogi do 100



Za določitev nesezonskih parametrov p in q gledamo ACF in PACF do prve periode (torej do odloga 96)

Večja korelacija pri prvih nekaj urah in v uri tik pred periodo → vključimo prvih nekaj 15-minutnih intervalov

Izbira modela SARIMAX na podlagi vrednosti kriterija AIC

- Eksogeni spremenljivki: temperatura in sevanje
- Model sem trenirala na 75 % podatkih

Izbira modela SARIMAX na podlagi vrednosti kriterija AIC

- Eksogeni spremenljivki: temperatura in sevanje
- Model sem trenirala na 75 % podatkih
- Vrednosti kriterije AIC za izbrane modele:

| 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 |

Izbira modela SARIMAX-GARCH

- Izbran model: **SARIMAX(4,1,5)(0,1,0)[96]**
- Model GARCH(p,q) konstruiramo na rezidualih izbranega modela

Izbira modela SARIMAX-GARCH

- Izbran model: **SARIMAX(4,1,5)(0,1,0)[96]**
- Model GARCH(p,q) konstruiramo na rezidualih izbranega modela
- Poiskujemo 6 kombinacij parametrov:

| 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 |

- Izbran model: **SARIMAX(4,1,5)(0,1,0)[96]-GARCH(1,3)**

SARIMAX(4,1,5)(0,1,0)[96]-GARCH(1,3)

- Spomnimo se:

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

kjer je $\sigma_t^2 = \omega + \sum_{i=1}^{p_G} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q_G} \beta_i \sigma_{t-i}^2$

SARIMAX(4,1,5)(0,1,0)[96]-GARCH(1,3)

- Spomnimo se:

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

kjer je $\sigma_t^2 = \omega + \sum_{i=1}^{p_G} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q_G} \beta_i \sigma_{t-i}^2$

- Vstavimo izbrane parametre:

$$X_t = \sum_{i=1}^4 \varphi_i X_{t-i} + \sum_{i=1}^5 \theta_i \varepsilon_{t-i} + 0 + \beta_1 Y_{1_t} + \beta_2 Y_{2_t} + \sqrt{\sigma_t} z_t,$$

kjer je $\sigma_t^2 = \omega + \sum_{i=1}^1 \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^3 \beta_i \sigma_{t-i}^2$, Y_{2_t} je eksogena spremenljivka temperature Y_{2_t} pa sevanja

Koeficienti modela

$$X_t = \sum_{i=1}^4 \varphi_i X_{t-i} + \sum_{i=1}^5 \theta_i \varepsilon_{t-i} + 0 + \beta_1 Y_{1t} + \beta_2 Y_{2t} + \sqrt{\sigma_t^2} z_t,$$

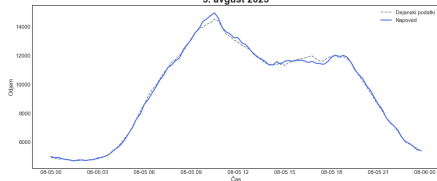
kjer je $\sigma_t^2 = \omega + \sum_{i=1}^1 \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^3 \beta_i \sigma_{t-i}^2$

Koeficienti:

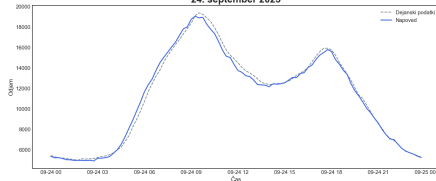
- AR del: $\varphi_1 = -0,3686$, $\varphi_2 = -0,4185$, $\varphi_3 = -0,2894$, $\varphi_4 = 0,0486$
- MA del: $\theta_1 = -0,3311$, $\theta_2 = 0,1863$, $\theta_3 = 0,0931$, $\theta_4 = -0,2059$, $\theta_5 = 0,0583$
- sezonskih del: parametrov nimamo, ker sta P in Q enaka 0
- GARCH del: $\alpha_1 = 0,2$,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

Testiranje modela

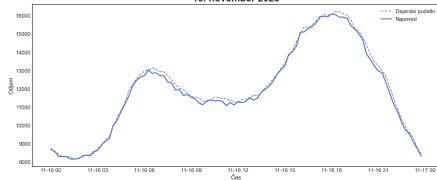
5. avgust 2023



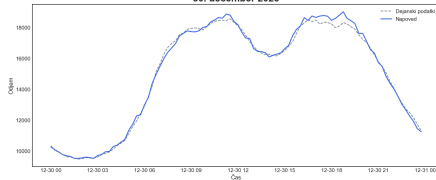
24. september 2023



16. november 2023



30. december 2023



Napaki MAPE in RMSE

Napaki RMSE (ang. *Root-mean-square deviation*) in MAPE (ang. *Mean absolute percentage error*) 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

Napaki MAPE in RMSE za izbrane datume

| Datum | RMSE | MAPE |
|--------------------|---------|-------|
| 5. avgust 2023 | 156,600 | 1,104 |
| 24. september 2023 | 365,372 | 2,392 |
| 16. november 2023 | 176,901 | 1,199 |
| 30. december 2023 | 224,123 | 1,046 |
| 7. januar 2024 | 399,578 | 1,839 |
| 12. februar 2024 | 248,981 | 1,287 |
| Povprečje | 261,926 | 1,478 |

- Z vključitvijo še drugih primerov ugotovimo, da RMSE znaša okrog 260, MAPE pa 1,5 %
- Vključitev eksogenih podatkov prispeva k bolj točni napovedi

Zaključek

- Izbran model: SARIMAX(4,1,5)(0,1,0)[96]-GARCH(1,3)
- Dobre napovedi
- Z bolj zmogljivim računalnikom bi prišli do boljših rezultatov
- Funkcija `napoved_z_SARIMA_GARCH`

Funkcija `napoved_z_SARIMA_GARCH`

Funkcija `napoved_z_SARIMA_GARCH`

Kako funkcijo uporabiti?

V funkcijo `napoved_z_SARIMA_GARCH` damo datum v obliki 'yyyy-mm-dd'; npr. `napoved_z_SARIMA_GARCH("2023-10-10")` za napoved dne 10. oktobra 2023. Najbolj je smiselno, da je datum iz testnega obdobja, torej med '2023-08-05' in '2024-02-28'. Rezultat se sicer dobi za vsak datum od '2021-11-03' naprej.

Kaj funkcija vrne?

Funkcije izriše prilaganje grafa napovedi (po modelu $\text{SARIMA}(4,1,5)(0,1,0)_{[96]}$ - $\text{GARCH}(1,3)$ z upoštevanjem eksogenih podatkov temperature in sevanja) dejanskemu odjemu. Prav tako izpiše napaki RMSE in MAPE. Do napak lahko tudi dostopamo, saj je return funkcije `napoved_z_SARIMA_GARCH` v obliki seznama [napovedane vrednosti po urah, napaka RMSE, napak MAPE].

Koliko časa funkcija potrebuje da vrne rezultat?

Okrog 1 minuto.

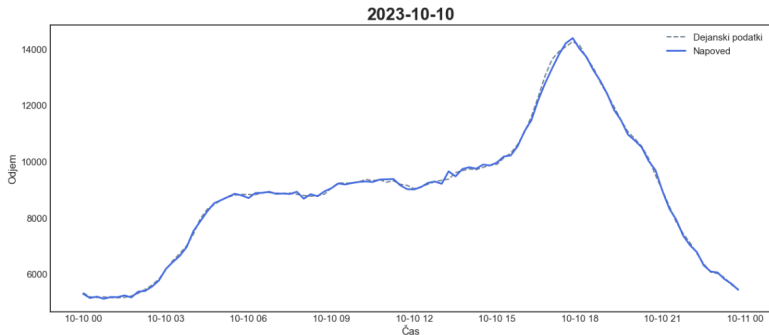
Primer uporabe napoved_z_SARIMA_GARCH

1. Napoved za 10. oktober 2023:

```
napoved_1 = napoved_z_SARIMA_GARCH('2023-10-10')
```

✓ 1m 12.1s

Python



RMSE: 85.8307826939613

MAPE: 0.7302950497184921