

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 2. stopnja

Katarina Brilej

**NAPOVEDOVANJE CEN IN POVPRAŠEVANJA
PO ELEKTRIČNI ENERGIJI**

Magistrsko delo

Mentor: izr. prof. dr. Mihael Perman

Ljubljana, 2021

Zahvala

Zahvaljujem se mentorju izr. prof. dr. Mihaelu Permanu za strokovno pomoč in nasvete pri pisanju magistrskega dela.

Kazalo

Program dela	ix
1 Uvod	1
2 Trg električne energije	2
3 Analiza časovnih vrst	4
3.1 Osnovni pojmi in modeli	4
3.2 Stacionarnost	5
3.3 Avtokovariančna funkcija stacionarnega procesa	8
3.4 Transformacije časovnih vrst	10
3.5 Modeli ARMA	12
3.5.1 Procesi $\text{ARMA}(p, q)$	12
3.5.2 ACF procesa $\text{ARMA}(p, q)$	16
3.5.3 Napovedovanje procesov ARMA	17
3.5.4 Modeli ARIMA za nestacionarne časovne vrste	21
3.5.5 Sezonski modeli ARIMA	22
3.6 Modeli GARCH	22
3.6.1 Model ARCH	23
3.6.2 Model GARCH	23
3.6.3 Model EGARCH	25
3.7 Identifikacija modela in napovedovanje	26
3.7.1 Box-Jenkinsova metodologija	26
3.7.2 Identifikacija modela	27
3.7.3 Kriterij AIC	29
3.7.4 Diagnostika in Ljung-Boxov test	30
3.7.5 Napovedovanje	31
4 Metoda podpornih vektorjev in jedrne funkcije	35
4.1 Linearni model za binarno klasifikacijo.	35
4.2 Model za podatke, ki niso linearno ločljivi	41
4.3 Regresija	43
4.4 Jedrne funkcije	46
4.5 Merjenje napovedne napake in izbira modela	49
5 Empirično modeliranje in napovedovanje	51
5.1 Predstavitev podatkov	51
5.2 Metodologija	53
5.3 Napovedovanje povpraševanja po električni energiji	55
5.3.1 ARMA-GARCH	55
5.3.2 SVR	63
5.3.3 Rezultati	65
5.4 Napovedovanje cene električne energije	70
5.4.1 ARMA-GARCH	70
5.4.2 SVR	77

5.4.3	Rezultati	79
6	Zaključek	84
	Literatura	85

Kazalo slik

1	Vrednosti NEP $N(0,1)$ šuma	9
2	Vzorčna ACF za NEP $N(0,1)$ šum.	10
3	ACF procesa $MA(1)$	13
4	ACF procesa $AR(1)$	14
5	Box-Jenkinsova metodologija.	27
6	Načelo širokega roba.	36
7	Ločevalna hiperravnina z najširšim robom.	37
8	Izračun širine roba.	38
9	Podporni vektorji in odmiki ξ_i	42
10	ε pas in funkcija izgube V_ε	46
11	Primer binarne klasifikacije v \mathbb{R}^2	47
12	Poraba električne energije v letih 2019 in 2020.	52
13	Cena električne energije v letih 2019 in 2020.	52
14	Povprečna poraba in cena po urah.	53
15	Logaritemska vrednost porabe električne energije za leto 2019.	55
16	Vzorčna ACF časovne vrste $\ln D_t$	56
17	Časovna vrsta $\nabla_{24} \ln D_t$	56
18	Vzorčna ACF časovne vrste $\nabla_{24} \ln D_t$	57
19	Vzorčna ACF časovne vrste $\nabla \nabla_{24} \ln D_t$	57
20	Vzorčna PACF časovne vrste $\nabla \nabla_{24} \ln D_t$	58
21	Residuali modela SARIMA-GARCH- $n(1,3)$ in vzorčna ACF.	62
22	Residuali modela SARIMA-GARCH- $t(3,1)$ in vzorčna ACF.	62
23	Residuali modela SARIMA-EGARCH(1,4) in vzorčna ACF.	63
24	Napoved porabe za tretji teden februarja.	66
25	Napoved porabe za tretji teden julija.	67
26	Napoved porabe za tretji teden septembra.	68
27	Logaritemska vrednost cene električne energije za leto 2019.	71
28	Vzorčna ACF časovne vrste $\ln P_t$	71
29	Časovna vrsta $\nabla_{24} \ln P_t$	72
30	Vzorčna ACF časovne vrste $\nabla_{24} \ln P_t$	72
31	Vzorčna ACF časovne vrste $\nabla \nabla_{24} \ln P_t$	73
32	Vzorčna PACF časovne vrste $\nabla \nabla_{24} \ln P_t$	73
33	Residuali modela SARIMA-GARCH- $n(2,1)$ in vzorčna ACF.	77
34	Napoved cene za tretji teden februarja.	80
35	Napoved cene za tretji teden julija.	81
36	Napoved cene za tretji teden septembra.	82

Kazalo tabel

1	Opisne statistike.	52
2	Vrednost kriterija AIC za 16 izbranih modelov.	60
3	Vrednost kriterija AIC za 10 izbranih modelov.	60
4	Optimalne vrednosti parametrov v primeru linearnega jedra.	64
5	Optimalni parametri v primeru radialnega jedra.	65

6	Napaka $MAPE$ (%) za tretji teden februarja.	66
7	Napaka $MAPE$ (%) za tretji teden julija.	67
8	Napaka $MAPE$ (%) za tretji teden septembra.	68
9	Napaka $MAPE$ (%) za tretji teden vseh mesecev v letu 2020.	69
10	Povprečna napaka $MAPE$ (%).	69
11	Napaka $RMSE$ za tretji teden vseh mesecev v letu 2020.	70
12	Povprečna napaka $RMSE$	70
13	Vrednost kriterija AIC za 13 izbranih modelov.	75
14	Vrednost kriterija AIC za 5 izbranih modelov.	76
15	Optimalne vrednosti parametrov v primeru linearnega jedra.	78
16	Optimalne vrednosti parametrov v primeru radialnega jedra.	79
17	Napaka $MAPE$ (%) za tretji teden februarja.	80
18	Napaka $MAPE$ (%) za tretji teden julija.	81
19	Napaka $MAPE$ (%) za tretji teden septembra.	82
20	Napaka $MAPE$ (%) za tretji teden vseh mesecev v letu 2020.	83
21	Povprečna napaka $MAPE$ (%).	83
22	Napaka $RMSE$ za tretji teden vseh mesecev v letu 2020	83
23	Povprečna napaka $RMSE$	84

Program dela

V magistrskem delu predstavite modeliranje in napovedovanje cen ter povpraševanja po električni energiji s časovnimi vrstami in strojnim učenjem. Metode uporabite na dejanskih podatkih in predstavite rezultate ter presodite uporabnost enega in drugega pristopa.

Osnovna literatura

- [7] P. J. Brockwell in R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, ZDA, 2006
- [17] B. Scholkopf in A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, The MIT Press, Cambridge, 2002
- [10] M. Fałdziński, P. Fiszeder in W. Orzeszko, *Forecasting Volatility of Energy Commodities: Comparison of GARCH Models with Support Vector Regression*, *Energies* **14**(1) (2021), doi: 10.3390/en14010006
- [13] R. C. Garcia in dr., *A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices*, *IEEE Transactions on Power Systems* **20**(2) (2005) 867–874, doi: 10.1109/TPWRS.2005.846044
- [23] R. Weron, *Electricity price forecasting: A review of the state-of-the-art with a look into the future*, *International Journal of Forecasting* **30**(4) (2014) 1030–1081, doi: 10.1016/j.ijforecast.2014.08.008

Podpis mentorja:

Napovedovanje cen in povpraševanja po električni energiji

POVZETEK

Napovedovanje porabe električne energije je priljubljena tema že kar nekaj časa, elektrike se namreč v večjih količinah ne da shraniti, stabilnost elektroenergetskega sistema pa zahteva stalno ravnovesje med proizvedeno in porabljeno električno energijo. Z deregularizacijo trga električne energije je dragoceno orodje postalo tudi napovedovanje cen. Magistrsko delo ponuja pristop k napovedovanju cen in povpraševanja po električni energiji na podlagi modelov ARMA-GARCH (področje časovnih vrst) in metode podpornih vektorjev (področje strojnega učenja). Pri tem gre za ceno pri trgovanju za dan vnaprej (ang. day ahead trading). Tako v primeru cene kot porabe električne energije so podatki dani na urni ravni, napovedujemo pa jih za 24 ur naprej.

Ključni del predstavlja izbira in identifikacija modela, s katerim kasneje izdelamo napovedi. Tako ceno kot porabo napovemo le z uporabo preteklih vrednosti cene in porabe, brez dodatnih zunanjih spremenljivk. Identifikacija poteka po Box-Jenkinsovi metodologiji, ki predstavlja standardno orodje v analizi časovnih vrst. Pri vsakem pristopu se osredotočimo na nekaj različic modelov, optimalne parametre pa izberemo s pomočjo različnih kriterijev. Pri modelih časovnih vrst sta to kriterij AIC in Ljung-Boxov test, pri metodi podpornih vektorjev pa prečno preverjanje.

Na koncu z izbranimi modeli napovemo prihodnje vrednosti cene in porabe električne energije. Modele med seboj primerjamo na podlagi različnih mer kakovosti modelov (*MAPE*, *RMSE*). V obeh primerih rezultati kažejo, da so za dane podatke napovedi modelov ARMA-GARCH boljše. V primeru porabe električne energije izrazitega favorita ni, v primeru cene pa izstopa model ARMA-GARCH- t .

Electricity price and demand forecasting

ABSTRACT

Electricity demand forecasting has been a popular topic for quite some time, as electricity cannot be stored in large quantities, whereas the stability of the electric power system requires a constant balance between electricity produced and consumed. With the deregulation of the electricity market, price forecasting has become valuable. This master's thesis offers an approach to electricity price and demand forecasting based on ARMA-GARCH models (time series) and support vector machines (machine learning). We focus on the day-ahead electricity prices. In both electricity price and demand, the analyzed data are available hourly, and our models provide 24-hour forecasts for the next day.

The crucial part in the elaboration of the forecasts is the model selection and identification. Both price and demand are predicted using only past values of price and demand, without additional external variables. The selection of models follows the Box-Jenkins methodology, a standard tool in time series analysis. We focus on a few (similar) models in each approach, and the optimal parameters are selected using different criteria. For time series models, these are the AIC criterion and the Ljung-Box test, and for the support vector machines, we use cross-validation.

Finally, we predict the future values of electricity price and demand with the previously selected models. We compare the models based on different measures for forecast accuracy ($MAPE$, $RMSE$). In both cases, the results show that the predictions of the ARMA-GARCH models are better for the given data. In the case of electricity demand, no model is considerably better than the rest, and in the case of price, the ARMA-GARCH- t model stands out.

Math. Subj. Class. (2010): 62M10, 68T01, 62P05

Ključne besede: napovedovanje popraševanja po električni energiji, napovedovanje cene električne energije, modeli ARMA, modeli GARCH, metoda podpornih vektorjev

Keywords: electricity demand forecasting, electricity price forecasting, ARMA models, GARCH models, support vector machines

1 Uvod

Proces deregularizacije in uvedba konkurenčnih trgov je od začetka devedsetih let prejšnjega stoletja preoblikovala tradicionalno monopolno podobo energetskega sektorja, nadzorovanega s strani države. V mnogih državah po vsem svetu se zdaj z električno energijo trguje po tržnih pravilih. Vendar pa je elektrika zelo posebna dobrina, v večjih količinah je namreč ne moremo shraniti, stabilnost elektroenergetskega sistema pa zahteva stalno ravnoesje med proizvodnjo in porabo. Hkrati pa je povpraševanje po električni energiji odvisno od vremena (temperatura, hitrost vetra) in intenzitete poslovanja ter vsakodnevnih aktivnosti (razlika med delavniki in vikendom, nočjo in dnevom). Za vzdrževanje ravnovesja je torej pomembno napovedati prihodnjo porabo. Te edinstvene in specifične lastnosti cene električne energije vodijo do dinamik cene, ki jih ni mogoče opaziti na nobenem drugem trgu. Kažejo namreč sezonskost na dnevni, tedenski in letni ravni prav tako pa nenadne, kratkotrajne in navadno nepričakove skoke cen. Ravno ta dinamika cen je raziskovalce spodbudila k intenzivnemu razvoju novih napovedovalnih tehnik [23].

Cilj magistrskega dela je tako modeliranje in nato napovedovanje cen in povpraševanja po električni energiji. Pri tem gre za ceno pri trgovanju za dan vnaprej (ang. day ahead trading). Dve glavni skupini modelov, ki so se v preteklosti uporabljali za napovedovanje cen in povpraševanja po električni energiji, sestavljajo statistični modeli (npr. modeli časovnih vrst, kot so AR, ARX, ARMA in GARCH) in modeli na osnovi strojnega učenja (npr. nevronske mreže, metoda podpornih vektorjev), možno jih je tudi kombinirati med seboj. Magistrsko delo ponuja pristop k napovedovanju na podlagi modelov ARMA-GARCH in metode podpornih vektorjev.

Uvodoma bomo predstavili trg električne energije, kar nam bo služilo kot motivacija za nadaljnje delo. Nadaljevali bomo z analizo časovnih vrst, spoznali bomo osnovne pojme, kot je stacionarnost časovne vrste, to lastnost namreč potrebujemo za modeliranje. Seznanili se bomo z metodami, kako stacionarnost prepoznamo, tu bomo omenili avtokorelacijsko funkcijo, pa tudi, kako časovno vrsto transformiramo v stacionarno. Glavni del tega poglavja bomo namenili modelom ARMA, seznanili se bomo z avtokorelacijsko in parcialno avtokorelacijsko funkcijo procesa $ARMA(p, q)$, obravnavali bomo tudi napovedovanje prihodnjih vrednosti stacionarnega procesa. Ker so resnični podatki velikokrat nestacionarni, potrebujemo modele ARIMA in njihovo sezonsko različico SARIMA (tako poraba kot cena električne energije imata močno sezonsko komponento). Modeli ARMA (in njihove različice) predpostavljajo konstantno varianco, vendar pa se v primerih finančnih podatkov pogosto izkaže, da tej homoskedastični predpostavki ni zadoščeno. Za modeliranje nekonstantne variance bomo predstavili model GARCH in njegove posplošitve. Ob koncu poglavja bomo predstavili Box-Jenkinsovo metodologijo, s katero bomo konstruirali model za obravnavano časovno vrsto. Prvi korak v metodologiji je identifikacija modela, kjer se naslonimo na vzorčno avtokorelacijsko in parcialno avtokorelacijsko funkcijo podatkov, uporaben pa je tudi informacijski kriterij AIC. Ustreznost izbranega modela je seveda potrebno preveriti, mi bomo za ta namen predstavili Ljung-Boxov test.

Sledi poglavje o metodi podpornih vektorjev, eni najbolj uporabljenih metod strojnega učenja. Najprej si bomo ogledali model za binarno klasifikacijo v primeru, ko so podatki linearno ločljivi, kar je bila tudi prvotna ideja metode. Seveda po-

datki v praksi navadno niso linearno ločljivi, zato si bomo ogledali prilagojen model za ta primer. Tako bomo spoznali intuicijo v ozadju metode podpornih vektorjev in nadaljevali na primer regresije, kar bomo uporabljali za napovedovanje cene in povpraševanja po električni energiji. Omenili bomo še jedrne funkcije, ki omogočajo nelinearno razvrščanje. Na koncu bomo opisali, kako merimo napovedno napako modela na danih podatkih ter kako izberemo optimalen model. Spoznali bomo različne mere napak (MSE , $RMSE$ in $MAPE$) ter k -kratno prečno preverjanje.

Nazadnje se bomo posvetili empiričnemu modeliranju in napovedovanju cene in povpraševanja po električni energiji na slovenskem trgu [10], [13]. Tako v primeru cene kot porabe (povpraševanja) električne energije so podatki dani na urni ravni (visoka frekvenčnost), napovedujemo pa jih za 24 ur naprej. Analizirali bomo podatke za leto 2019, napovedovali pa za leto 2020. V vseh primerih bomo ceno in porabo napovedovali le z uporabo preteklih vrednosti cene in porabe, brez dodatnih zunanjih spremenljivk (med te bi sicer lahko šteli temperaturo, hitrost vetra, pa tudi porabo za napoved cene). Pri vsakem pristopu bomo naredili nekaj različic modelov, optimalne parametre pa bomo izbrali s pomočjo različnih kriterijev. Pri modelih časovnih vrst bodo to kriterij AIC in Ljung-Boxov test, pri podpornih vektorjih pa prečno preverjanje. Na začetku je seveda potrebno podatke še obdelati, da so primerni za posamezne modele. Na koncu bomo z izbranimi modeli naredili napovedi za tretji teden vsakega meseca v letu 2020 in si ogledali napake na dnevni in tedenski ravni. Na podlagi različnih mer kvalitete modelov ($MAPE$, $RMSE$) bomo modele primerjali med seboj in sklepali, kateri je boljši.

2 Trg električne energije

Napovedovanje porabe električne energije je priljubljena tema že kar nekaj časa, elektrike se namreč v večjih količinah ne da shraniti, stabilnost elektroenergetskega sistema pa zahteva stalno ravnovesje med proizvedeno in porabljeno električno energijo (tega je potrebno uravnavati v realnem času). V primeru odstopanja od ravnovesja, je potrebno plačati kazen, saj se spremeni frekvenca elektroenergetskega sistema (večja poraba od proizvodnje pomeni nižjo frekvenco in obratno). Za energetska podjetja so tako ključne napovedi porabe električne energije, na podlagi katerih se odločajo o nakupu ali prodaji električne energije (če je proizvodnja večja od napovedane porabe). Za potrošnike električne energije je značilno, da svoje porabe navadno niso pripravljeni prilagajati cenam, prav tako pa so prisotne precejšnje razlike v porabi na dnevni in tedenski ravni (čez dan je povpraševanja več kot ponoči, med delavniki več kot ob vikendu) [23], [13], [24]. Vse te lastnosti je seveda potrebno upoštevati pri napovedovanju.

Z deregularizacijo trga električne energije je dragoceno orodje postalo tudi napovedovanje cen. Podjetja, ki trgujejo na trgih z električno energijo, pogosto uporabljajo tehnike napovedovanja cen za razvijanje ponudbenih strategij ali pogajalskih veščin, da bi povečali dobiček, lahko pa se tudi zavarujejo pred volatilitnostjo cen. Na nihanje cen vplivajo številni dejavniki, med najpomembnejšimi so spremembe v hitrosti vetra in temperaturi ter količina padavin, na volatilitnost cen pa ima vpliv tudi nezmožnost shranjevanja električne energije [13].

S sprejetjem prvega energetskega zakona leta 1999 se je začel odpirati tudi trg

električne energije v Sloveniji. Z vstopom v EU pa je Slovenija postala del enotnega trga tudi na področju energetike, zato je morala vzpostaviti evropsko primerljiv energetskega sistem. Trg električne energije se je tako dokončno odprl leta 2007 z odprtjem trga za gospodinjstva. Pred tem se je z električno energijo trgovalo predvsem bilateralno, prek borznih posrednikov ali neposrednih stikov. Liberalizacija energetskega trga pa je omogočila razvoj konkurence med udeleženci na trgu. Del bilateralnega trgovanja se je tako prenesel na elektronske platforme, kjer se trguje anonimno in avtomatično [22], [1].

V Sloveniji trgovanje z električno energijo poteka na energetski borzi BSP SouthPool. BSP tržnim udeležencem omogoča trgovanje za dan vnaprej (ang. day-ahead) in trgovanje znotraj dneva (ang. intraday) na slovenskem borznem trgu. Možno je tudi sodelovanje na dolgoročnih avkcijah (ang. long-term auctions). Sodelujejo lahko tako pravne kot fizične osebe, potrebujejo pa odobreno vlogo za sodelovanje na trgih [21].

Trgovanje na borznem trgu poteka tako, da se soočata ponudba in povpraševanje za standardizirane produkte električne energije pri trgovanju za dan vnaprej in za standardizirane in nestandardizirane produkte električne energije pri trgovanju znotraj dneva. Mi se bomo osredotočili na trgovanje za dan vnaprej, to na slovenskem borznem trgu poteka na način avkcijskega trgovanja, ki se deli na štiri faze [21]:

1. Med 8:00 in 12:00 poteka faza trgovanja. Tržni udeleženci v trgovalno aplikacijo vnašajo standardizirane urne produkte, ki so omejeni s cenovnim razponom od -500 EUR/MWh do 3000 EUR/MWh in količinskim intervalom 1 MW. Ponudbe je možno vnašati že osem dni pred samim dnevom trgovanja. Možen je vnos, sprememba in preklic ponudb, udeleženci lahko vidijo samo svoje ponudbe.
2. Med 12:00 in 12:05 poteka faza mirovanja. Nadzornik trgovanja preveri vnesene ponudbe in ukrepa v primeru nepravilnosti.
3. Med 12:05 in 12:52 poteka faza po trgovanju. Izvede se izračun marginalne cene, ki temelji na algoritmu trgovalne aplikacije. Članom trgovanja se prikazujejo marginalne cene, izračunane na avkciji.
4. Od 12:52 naprej poteka neaktivna faza. Možen je pregled marginalnih cen in lastnih poslov.

Udeleženci na trgu morajo torej svoje ponudbe izraziti v smislu cen in količin. Podjetje, ki lahko uspešno napove ceno, lahko prilagodi lastno ceno oz. načrt proizvodnje glede na urne cene in lastne proizvodne stroške.

Drugi instrument za olajšanje trgovanja je dvostranski pogodbeni sistem. V tem primeru se kupec in prodajalec dogovorita o določenem znesku, ki ga želita prenesti prek omrežja po določeni fiksni ceni. Za to ceno se predhodno dogovorita obe strani in temelji tudi na napovedih cen. Večina dereguliranih trgov z električno energijo uporablja tako avkcijski sistem kot tudi bilateralne pogodbe. Podjetja morajo optimizirati svoje proizvodne načrte, da se lahko z dvostranskimi pogodbami varujejo pred nihanjem cen na trgu. Zato je dobro poznavanje prihodnjih cen zelo koristno za natančnejše vrednotenje dvostranskih pogodb [13].

Predstavili smo, zakaj je napovedovanje cen in povpraševanja po električni energiji pomembno, v naslednjih poglavjih pa se bomo seznanili z dvema možnima metodama, kako k temu problemu pristopimo.

3 Analiza časovnih vrst

V tem poglavju bomo predstavili nekaj osnovnih idej analize časovnih vrst. Cilj je sklepati na podlagi opažene časovne vrste, pred tem pa moramo vzpostaviti hipotetični verjetnostni model, ki bo predstavljal podatke. Ko izberemo primerno družino modelov, lahko potem ocenimo parametre in preverimo prileganje modela podatkom. Dobljen model lahko uporabimo tudi za izboljšanje razumevanja mehanizma v ozadju časovne vrste. Ko dobimo model, ki zadošča našim kriterijem, ga lahko uporabimo v različne namene, npr. napovedovanje ali pa simulacije, odvisno od področja uporabe.

V analizi časovnih vrst sta posebej pomembna koncepta stacionarnosti ter avtokovariančne in vzorčne avtokovariančne funkcije. Stacionarnost je namreč ena glavnih lastnosti časovnih vrst, ki nas bo zanimala. Osredotočili se bomo predvsem na to, kako časovno vrsto transformiramo v stacionarno in kako njeno stacionarnost nato preverimo. V ta namen bomo opisali nekaj standardnih tehnik za odstranitev trenda in sezone (z znano periodo) iz opazovane časovne vrste. Na koncu si bomo ogledali še različne modele časovnih vrst, kot so ARMA, ARCH in GARCH ter njihove kombinacije. Z njihovo uporabo bomo namreč modelirali časovni vrsti cene in porabe električne energije, opisali bomo tudi postopek izbire modela. Osredotočili se bomo tudi na napovedovanje na podlagi ocenjenega modela, kar je tudi naš glavni cilj. Poglavje je v večini povzeto po [7] in [8].

3.1 Osnovni pojmi in modeli

Uvodoma si oglejmo nekaj osnovnih pojmov in definicij, ki jih bomo potrebovali v nadaljnji analizi.

Časovna vrsta je časovno urejeno zaporedje številčnih podatkov x_t , ki izraža vrednost neke slučajne spremenljivke. V primeru diskretnih časovnih vrst (tem se bomo posvetili) je množica časov T_0 , za katere imamo meritve, diskretna množica. Podatki so navadno izmerjeni v enakih časovnih intervalih, npr. vsako uro, vsak dan, vsak mesec, lahko tudi manj pogosto, npr. vsak mesec, vsako četrtoletje, leto. Primer diskretne časovne vrste je poraba električne energije v Sloveniji, ki je izmerjena vsako uro. V primeru zveznih časovnih vrst so podatki zvezno izmerjeni čez nek časovni interval.

Prvi korak v analizi časovnih vrst je izbira primerne matematičnega modela (ali razreda modelov) za podatke. Da bi vključili morebitno nepredvidljivo naravo prihodnjih opažanj, je naravno privzeti, da je vsako opažanje x_t realizacija neke slučajne spremenljivke X_t .

Definicija 3.1. Model časovnih vrst za opažene podatke $\{x_t\}$ je specifikacija skupne porazdelitve (ali morda samo povprečja in variance) zaporedja slučajnih spremenljivk $\{X_t\}$, kjer domnevamo, da je $\{x_t\}$ realizacija teh slučajnih spremenljivk.

Opomba 3.2. Izraz časovna vrsta bomo uporabljali tako za same podatke kot tudi za proces, katerega realizacija so. V večini praktičnih primerov opazimo samo eno realizacijo, mislimo pa si, da je to ena izmed mnogih, ki bi se lahko zgodile.

Oglejmo si najpreprostejši model časovnih vrst, ki mu pravimo šum.

Primer 3.3 (n. e. p. šum). Najpreprostejši model za časovno vrsto je tak, pri katerem ni opaziti trenda ali sezonske komponente, opažanja pa so preprosto neodvisne in enako porazdeljene (n. e. p.) slučajne spremenljivke z ničelnim povprečjem. Takšno zaporedje slučajnih spremenljivk X_1, X_2, \dots imenujemo n. e. p. šum. V tem modelu med opažanji ni odvisnosti. Poznavanje vrednosti X_1, \dots, X_n (za neko celo število n) ne doda vrednosti pri napovedovanju obnašanja X_{n+h} (za $h \geq 1$). Četudi se zdi n. e. p. šum nezanimiv proces za napovedovanje, pa igra pomembno vlogo kot gradnik zapletenih modelov časovnih vrst, ki jih bomo spoznali v nadaljevanju.

3.2 Stacionarnost

V tem poglavju bomo spoznali pojem stacionarnosti ter avtokorelacijsko funkcijo, ki imata ključno vlogo pri analizi in modeliranju časovnih vrst.

Večkrat smo že omenili stacionarnost kot želeno lastnost časovne vrste. Ohlapno rečeno je časovna vrsta $\{X_t, t = 0, \pm 1, \dots\}$ stacionarna, če ima podobne statistične značilnosti kot časovno zamaknjena časovna vrsta $\{X_{t+h}, t = 0, \pm 1, \dots\}$, za vsako celo število h . Če pozornost omejimo samo na lastnosti, ki so odvisne od prvega in drugega momenta časovne vrste $\{X_t\}$, lahko to idejo natančno opišemo z naslednjimi definicijami.

Definicija 3.4 (Kovariančna funkcija). Naj bo $\{X_t, t \in T\}$ časovna vrsta, za katero je $\text{Var}(X_t) < \infty$ za vsak $t \in T$. Kovariančna funkcija $\gamma_X(\cdot, \cdot)$ časovne vrste $\{X_t\}$ je dana z

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))], \quad r, s \in T. \quad (3.1)$$

Definicija 3.5 (Šibka stacionarnost). Časovna vrsta $\{X_t, t \in \mathbb{Z}\}$ z indeksno množico $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ je šibko stacionarna, če velja naslednje:

- $E(X_t^2) < \infty$ za vse $t \in \mathbb{Z}$,
- $E(X_t) = m$ za vse $t \in \mathbb{Z}$ in
- $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ za vse $r, s, t \in \mathbb{Z}$.

Opomba 3.6. Ko bomo uporabljali izraz stacionarnost, bomo mislili šibko stacionarnost kot v definiciji 3.5.

Opomba 3.7. Če je $\{X_t, t \in \mathbb{Z}\}$ stacionarna časovna vrsta, potem velja $\gamma_X(r, s) = \gamma_X(r - s, 0)$ za vse $r, s \in \mathbb{Z}$. Če bomo izraz kovariančne funkcije uporabili v povezavi s stacionarno časovno vrsto $\{X_t\}$ (kot v tretjem pogoju definicije 3.5), bomo mislili funkcijo $\gamma_X(\cdot)$ ene spremenljivke

$$\gamma_X(h) \equiv \gamma_X(h, 0) = \gamma_X(t + h, t), \quad \text{za vse } t, h \in \mathbb{Z}. \quad (3.2)$$

Imenovali jo bomo avtokovariančna funkcija, $\gamma_X(h)$ je njena vrednost pri odlogu (ang. lag) h .

Oglejmo si še formalni definiciji avtokovariančne in avtokorelacijske funkcije šibko stacionarne časovne vrste.

Definicija 3.8. Naj bo $\{X_t\}$ šibko stacionarna časovna vrsta. Avtokovariančna funkcija (ACVF) časovne vrste $\{X_t\}$ pri odlogu h je dana z

$$\gamma_X(h) \equiv \text{Cov}(X_{t+h}, X_t) \quad \text{za vse } t, h \in \mathbb{Z}. \quad (3.3)$$

Avtokorelacijska funkcija (ACF) časovne vrste $\{X_t\}$ pri odlogu h je dana z

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Cor}(X_{t+h}, X_t) \quad \text{za vse } t, h \in \mathbb{Z}. \quad (3.4)$$

Spoznali smo pojem šibke stacionarnosti, pogosto pa se uporablja tudi stroga stacionarnost, ki je dana z naslednjo definicijo.

Definicija 3.9 (Stroga stacionarnost). Časovna vrsta $\{X_t, t \in \mathbb{Z}\}$ je strogo stacionarna, če imata (X_1, \dots, X_k) in $(X_{1+h}, \dots, X_{k+h})$ enako skupno porazdelitev za vsa cela števila h in pozitivna cela števila k .

Stroga stacionarnost intuitivno pomeni, da imajo grafi realizacije časovne vrste na časovnem intervalu enake dolžine podobne statistične lastnosti. Na primer, delež ordinat, ki ne presegajo ravni x , mora biti približno enak za oba intervala. Med šibko in strogo stacionarnostjo obstaja povezava, preprosto lahko preverimo naslednjo trditev [7]. Če je $\{X_t\}$ strogo stacionarna časovna vrsta in $E(X_t^2) < \infty$ za vse $t \in \mathbb{Z}$, potem je $\{X_t\}$ tudi šibko stacionarna. Obratno ne velja nujno.

Oglejmo si preprosta primera stacionarnih časovnih vrst.

Primer 3.10 (n. e. p. šum). Če je $\{X_t\}$ n. e. p. šum in $E(X_t^2) = \sigma^2 < \infty$, potem je prvi zahtevi v definiciji 3.5 zadoščeno. Zadoščeno je tudi drugi zahtevi, saj je $E(X_t) = 0$ za vse t . Ker smo privzeli neodvisnost, je kovariančna funkcija enaka

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases}, \quad (3.5)$$

kar ni odvisno od t . Torej je n. e. p. šum s končnim drugim momentom stacionarna časovna vrsta. Uporabljali bomo zapis

$$\{X_t\} \sim \text{NEP}(0, \sigma^2),$$

da nakažemo, da so slučajne spremenljivke X_t neodvisne in enako porazdeljene slučajne spremenljivke, vsaka s povprečjem 0 in varianco σ^2 .

Primer 3.11 (Beli šum). Naj bo $\{X_t\}$ zaporedje nekoreliranih slučajnih spremenljivk, vsaka z ničelnim povprečjem in varianco σ^2 . $\{X_t\}$ je očitno stacionarna časovna vrsta z enako kovariančno funkcijo kot n. e. p. šum v primeru 3.10. Takšnemu zaporedju pravimo beli šum (ang. white noise) in označimo

$$\{X_t\} \sim \text{WN}(0, \sigma^2).$$

Očitno je vsako $\text{NEP}(0, \sigma^2)$ zaporedje tudi $\text{WN}(0, \sigma^2)$, obratno pa ne velja nujno.

Oglejmo si še dva primera stacionarnih časovnih vrst, ki ju bomo natančneje obravnavali tudi v nadaljevanju.

Primer 3.12 (Proces MA(1)). Oglejmo si časovno vrsto definirano z enačbo

$$X_t = Z_t + \theta Z_{t-1}, \quad t \in \mathbb{Z}, \quad (3.6)$$

kjer je $\{Z_t\} \sim WN(0, \sigma^2)$, θ je realnoštevilska konstanta. Iz enačbe (3.6) vidimo, da je $E(X_t) = 0$, $E(X_t^2) = \sigma^2(1 + \theta^2) < \infty$ in

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2(1 + \theta^2), & h = 0 \\ \sigma^2\theta, & h = \pm 1 \\ 0, & |h| > 1 \end{cases}. \quad (3.7)$$

Sledi, da je $\{X_t\}$ stacionarna. Avtokorelacijska funkcija časovne vrste $\{X_t\}$ je enaka

$$\rho_X(h) = \begin{cases} 1, & h = 0 \\ \theta/(1 + \theta^2), & h = \pm 1 \\ 0, & |h| > 1 \end{cases}. \quad (3.8)$$

Časovni vrsti $\{X_t\}$ pravimo proces drsečih sredin (ang. moving average process) prvega reda in krajše označimo kot MA(1).

Primer 3.13 (Proces AR(1)). Privzemimo, da je $\{X_t\}$ stacionarna časovna vrsta, ki zadošča enačbi

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \dots, \quad (3.9)$$

kjer je $\{Z_t\} \sim WN(0, \sigma^2)$, $|\phi| < 1$, Z_t in X_s pa sta nekorelirana za vsak $s < t$ (kasneje bomo pokazali, da obstaja natanko ena rešitev enačbe (3.9)). Če izračunamo pričakovano vrednost obeh strani enačbe (3.9) in uporabimo dejstvo, da je $E(Z_t) = 0$, vidimo, da velja

$$E(X_t) = 0.$$

Da dobimo avtokorelacijsko funkcijo časovne vrste $\{X_t\}$, pomnožimo obe strani enačbe (3.9) z X_{t-h} ($h > 0$), izračunamo pričakovano vrednost in dobimo

$$\begin{aligned} \gamma_X(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \text{Cov}(\phi X_{t-1}, X_{t-h}) + \text{Cov}(Z_t, X_{t-h}) \\ &= \phi \gamma_X(h-1) + 0 = \dots = \phi^h \gamma_X(0). \end{aligned}$$

Opazimo, da je $\gamma(h) = \gamma(-h)$ ter uporabimo definicijo avtokorelacijske funkcije, da dobimo

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \quad h = 0, \pm 1, \dots \quad (3.10)$$

Iz linearnosti kovariančne funkcije in dejstva, da sta Z_t in X_{t-1} nekorelirana, sledi

$$\gamma_X(0) = \text{Cov}(X_t, X_t) = \text{Cov}(\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t) = \phi^2 \gamma_X(0) + \sigma^2. \quad (3.11)$$

Od tod pa sledi, da je $\gamma_X(0) = \sigma^2/(1 - \phi^2)$. Časovni vrsti $\{X_t\}$ pravimo avtoregresijski proces (ang. autoregressive process) prvega reda in krajše označimo kot AR(1).

3.3 Avtokovariančna funkcija stacionarnega procesa

V prejšnjem poglavju smo predstavili avtokovariančno funkcijo, zdaj pa si bomo ogledali nekaj njenih lastnosti.

Lema 3.14 (Osnovne lastnosti). *Naj bo γ avtokovariančna funkcija stacionarnega procesa $\{X_t, t \in \mathbb{Z}\}$, potem velja naslednje:*

- $\gamma(0) \geq 0$,
- $|\gamma(h)| \leq \gamma(0)$ za vse $h \in \mathbb{Z}$,
- $\gamma(\cdot)$ je soda funkcija, tj. $\gamma(h) = \gamma(-h)$ za vse $h \in \mathbb{Z}$,
- $\gamma(\cdot)$ je pozitivno semidefinitna funkcija, tj. za vsa pozitivna cela števila n in vsak vektor $(a_1, \dots, a_n) \in \mathbb{R}^n$ velja $\sum_{i=1}^n \sum_{j=1}^n a_i \gamma(i-j) a_j \geq 0$.

Dokaz. Prva lastnost je očitna, saj velja $\text{Var}(X_t) \geq 0$. Druga lastnost je takojšnja posledica Cauchy-Schwarzove neenakosti:

$$|\text{Cov}(X_{t+h}, X_t)| \leq (\text{Var}(X_{t+h}))^{1/2} (\text{Var}(X_t))^{1/2}.$$

Tretja lastnost sledi iz

$$\gamma(-h) = \text{Cov}(X_{t-h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma(h).$$

Pokazati moramo še četrto lastnost. Naj bo a poljuben $n \times 1$ vektor z realnoštevilskimi komponentami a_1, \dots, a_n in naj bo $X_n = (X_1 \dots X_n)^T$. Iz nenegativnosti varianc sledi

$$\text{Var}(a^T X_n) = a^T \Gamma_n a = \sum_{i=1}^n \sum_{j=1}^n a_i \gamma(i-j) a_j \geq 0,$$

kjer je Γ_n variančno kovariančna matrika slučajnega vektorja X_n . Zadnja neenakost ravno pomeni, da je funkcija $\gamma(\cdot)$ pozitivno semidefinitna. \square

Opomba 3.15. Avtokorelacijska funkcija $\rho(\cdot)$ ima vse lastnosti avtokovariančne funkcije, dodatno pa velja še $\rho(0) = 1$.

Ogledali smo si že, kako za preprost model časovnih vrst izračunamo avtokorelacijsko funkcijo. Vendar pa v realnosti ne začnemo z znanim modelom, temveč z opaženimi podatki $\{x_1, x_2, \dots, x_n\}$. Naša naloga je oceniti stopnjo odvisnosti v podatkih in izbrati primeren model, ki to odraža. Pri tem si pomagamo z vzorčno avtokorelacijsko funkcijo. Če verjamemo, da so podatki realizacija stacionarne časovne vrste $\{X_t\}$, potem nam vzorčna avtokorelacijska funkcija da oceno za avtokorelacijsko funkcijo časovne vrste $\{X_t\}$. Ta ocena nakazuje, kateri izmed mnogih stacionarnih modelov časovnih vrst je primeren kandidat za prikaz odvisnosti v podatkih. Na primer, vzorčna avtokorelacijska funkcija, ki je blizu ničli za vse neničelne odloge, nakazuje, da je primeren model za podatke n. e. p. šum. Naslednje definicije so analogne tistim za stacionarne modele časovnih vrst, le da te temeljijo na vzorcu.

Definicija 3.16. Naj bodo x_1, \dots, x_n opažanja časovne vrste. Vzorčna avtokovariacijska funkcija je definirana kot

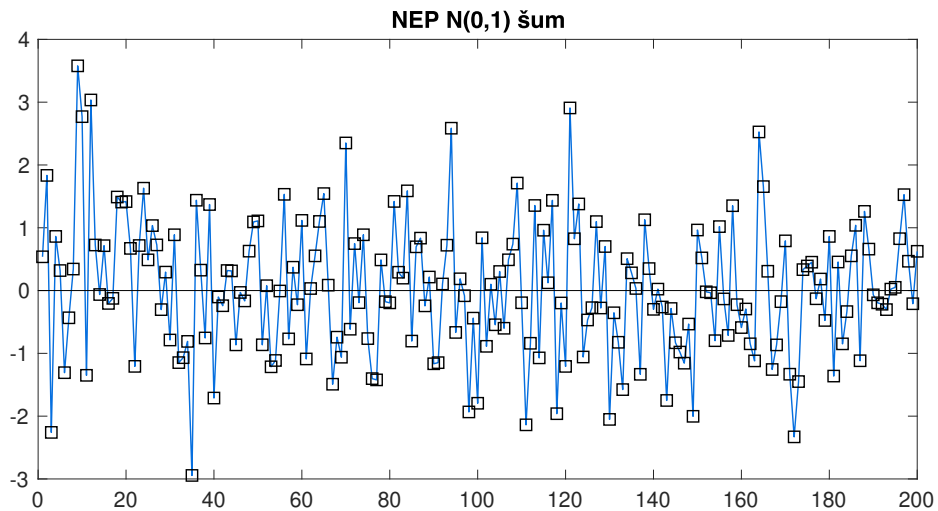
$$\hat{\gamma}(h) := \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x}), \quad 0 \leq h < n, \quad (3.12)$$

in $\hat{\gamma}(h) = \hat{\gamma}(-h)$ za $-n < h \leq 0$. Pri tem je \bar{x} vzorčno povprečje; $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$. Vzorcna avtokorelacijska funkcija je definirana kot

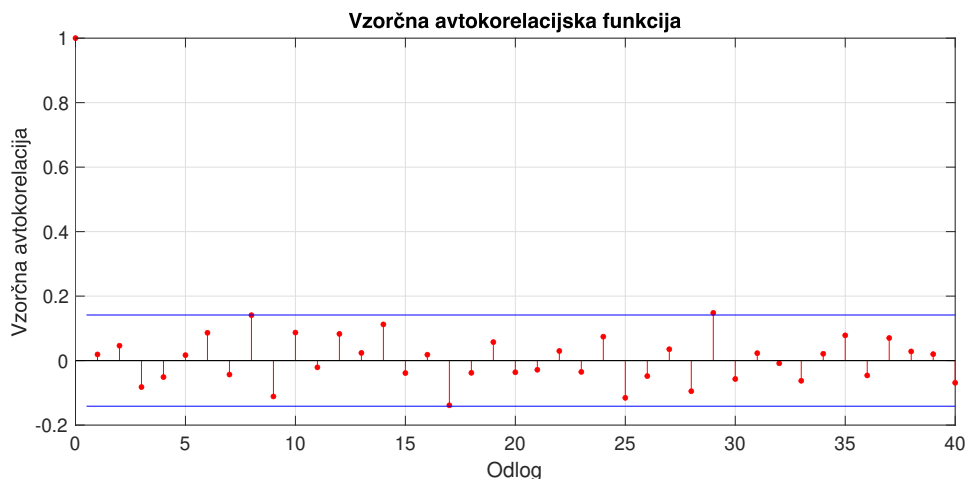
$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad |h| < n. \quad (3.13)$$

Oglejmo si preprost primer vzorčne avtokorelacijske funkcije.

Primer 3.17. Slika 1 prikazuje zaporedje 200 simuliranih vrednosti n. e. p. normalne slučajne spremenljivke s povprečjem 0 in varianco 1 (imenujemo ga NEP $N(0,1)$ zaporedje). Slika 2 prikazuje pripadajočo vzorčno avtokorelacijsko funkcijo pri odlogih $0, 1, \dots, 40$. Ker je $\rho(h) = 0$ za $h > 0$, bi pričakovali, da bo tudi vzorčna avtokorelacijska funkcija blizu 0. Pokazati se da [7], da za n. e. p. šum s končno varianco velja, da so vzorčne avtokorelacije $\hat{\rho}(h)$, $h > 0$, porazdeljene približno NEP $N(0, 1/n)$, za velik n . Zato mora približno 95% vzorčnih avtokorelacij ležati znotraj mej $\pm 1.96/\sqrt{n}$ (1.96 je 0.975-kvantil standardne normalne porazdelitve). Na sliki 2 bi torej pričakovali, da 2 vrednosti izmed 40 padeta izven mej in res vidimo nekaj takšnega.



Slika 1: 200 simuliranih vrednosti NEP $N(0,1)$ šuma.



Slika 2: Vzorčna ACF za podatke iz slike 1. Modra črta označuje meji pri $\pm 1.96/\sqrt{n}$.

Pripomniti velja, da lahko vzorčno avtokovariančno in avtokorelacijsko funkcijo izračunamo za katerokoli množico podatkov $\{x_1, \dots, x_n\}$ in ne samo za opažanja stacionarne časovne vrste. V podatkih, kjer je prisoten trend, $|\hat{\rho}(h)|$ počasi pada z večanjem h . V podatkih z deterministično periodično komponento lahko v $|\hat{\rho}(h)|$ zasledimo podobno obnašanje z enako periodo. $\hat{\rho}(\cdot)$ je tako pomemben indikator nestacionarnosti.

3.4 Transformacije časovnih vrst

Prvi korak v analizi katerekoli časovne vrste je, da podatke narišemo. Ko graf natančno preučimo, lahko morda podatke predstavimo kot realizacijo procesa

$$X_t = m_t + s_t + Y_t, \quad (3.14)$$

kjer je m_t počasi spreminjajoča se funkcija, znana kot komponenta trenda, s_t je funkcija z znano periodo d , imenujemo jo sezonska komponenta, Y_t pa je komponenta naključnega šuma, ki je stacionarna. Zgornjemu procesu pravimo model klasične dekompozicije (ang. the classical decomposition model).

Pri obravnavanju trenda in sezone prevladujeta dva pristopa. Cilj prvega pristopa je oceniti deterministični komponenti m_t in s_t v upanju, da bo komponenta šuma Y_t stacionarna časovna vrsta. Potem lahko uporabimo teorijo takšnih procesov, da poiščemo zadovoljiv verjetnostni model za proces $\{Y_t\}$ ter ga nato skupaj z m_t in s_t uporabimo za napovedovanje vrednosti $\{X_t\}$.

Nekoliko drugačen pristop pa sta leta 1976 razvila Box in Jenkins [6]. Ideja njunega pristopa je, da na časovni vrsti $\{X_t\}$ večkrat uporabimo operatorje diferenciranja, dokler diferencirana opažanja ne spominjajo na realizacijo neke stacionarne časovne vrste $\{W_t\}$. Nato lahko uporabimo teorijo stacionarnih procesov za modeliranje, analizo in napovedovanje $\{W_t\}$ in s tem tudi osnovnega procesa. Mi se bomo osredotočili na način, kot sta ga predlagala Box in Jenkins, trend in sezone bomo torej odstranili z diferenciranjem.

Najprej obravnavajmo situacijo, ko imamo samo trend, ne pa tudi sezonskosti. Model v enačbi (3.14) tako postane

$$X_t = m_t + Y_t, \quad t = 1, \dots, n. \quad (3.15)$$

Brez škode za splošnost lahko privzamemo, da velja $E(Y_t) = 0$ (Če $E(Y_t) \neq 0$, potem m_t in Y_t v enačbi (3.15) zamenjamo z $m_t + E(Y_t)$ in $Y_t - E(Y_t)$). Trend želimo odstraniti z diferenciranjem. Definirajmo operator prva diferenca ∇ kot

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t, \quad (3.16)$$

kjer je B operator pomik nazaj (ang. backward shift operator),

$$BX_t = X_{t-1}. \quad (3.17)$$

Potenci operatorjev B in ∇ sta definirani na očitni način, tj. $B^j(X_t) = X_{t-j}$ in $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$ za $j \geq 1$ in $\nabla^0(X_t) = X_t$. Polinome v B in ∇ obravnavamo na enak način kot polinomske funkcije realnih spremenljivk. Na primer,

$$\begin{aligned} \nabla^2(X_t) &= \nabla(\nabla(X_t)) = (1 - B)(1 - B)X_t = (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned}$$

Če operator ∇ uporabimo na funkciji linearnega trenda $m_t = c_0 + c_1 t$, potem dobimo konstantno funkcijo $\nabla m_t = m_t - m_{t-1} = c_0 + c_1 t - (c_0 + c_1(t-1)) = c_1$. Na enak način lahko polinomski trend stopnje k znižamo do konstante z uporabo operatorja ∇^k . Na primer, če je $X_t = m_t + Y_t$, kjer je $m_t = \sum_{j=0}^k c_j t^j$ in je Y_t stacionarna s povprečjem 0, nam aplikacija ∇^k da

$$\nabla^k(X_t) = k!c_k + \nabla^k Y_t, \quad (3.18)$$

torej stacionarni proces s povprečjem $k!c_k$. Ti premisleki nakazujejo naslednjo možnost: na danem zaporedju podatkov $\{x_t\}$ večkrat uporabimo operator ∇ , dokler ne najdemo zaporedja $\{\nabla^k x_t\}$, ki ga je mogoče modelirati kot realizacijo stacionarnega procesa. V praksi je pogosto potreben red diferenciranja k majhnen, pogosto ena ali dva. To temelji na dejstvu, da lahko na intervalu končne dolžine številne funkcije dobro aproksimiramo s polinomi dokaj nizkih stopenj.

Oglejmo si primer, kjer imamo prisotna tako trend kot sezonskost. Metodo, opisano za odstranitev trenda, je mogoče na naraven način prilagoditi za odstranitev trenda in sezonskosti v splošnem modelu, ki je dan z

$$X_t = m_t + s_t + Y_t, \quad t = 1, \dots, n, \quad (3.19)$$

kjer je $E(Y_t) = 0$, $s_{t+d} = s_t$ in $\sum_{j=1}^d s_j = 0$.

Tehniko diferenciranja, ki smo jo prej uporabili za nesezonske podatke, lahko prilagodimo za obravnavanje sezonskosti periode d z uvedbo operatorja d -te difference ∇_d , ki je definiran kot

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t. \quad (3.20)$$

Pripomnimo, da tega operatorja ne smemo zamenjati s prej definiranim operatorjem $\nabla^d = (1 - B)^d$. Če operator ∇_d uporabimo na modelu

$$X_t = m_t + s_t + Y_t, \quad (3.21)$$

kjer ima $\{s_t\}$ periodo d , dobimo

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}, \quad (3.22)$$

kar nam da dekompozicijo difference $\nabla_d X_t$ na komponento trenda ($m_t - m_{t-d}$) in komponento šuma ($Y_t - Y_{t-d}$). Trend $m_t - m_{t-d}$, lahko nato izločimo s pomočjo že opisanih metod, na primer z uporabo potence operatorja ∇ .

Cilj opisanih transformacij je dobiti časovno vrsto, ki ne odstopa preveč od stacionarnosti, brez očitnega trenda in sezonečnosti.

3.5 Modeli ARMA

Ključno vlogo v analizi časovnih vrst igrajo procesi, katerih lastnosti, ali nekatere izmed njih, se ne spreminjajo s časom. Če želimo izdelati napovedi, moramo očitno privzeti, da se nekaj ne spreminja s časom. Med najbolj znane stacionarne procese sodijo procesi ARMA, katerih definicijo si bomo najprej ogledali.

3.5.1 Proces ARMA(p, q)

Definicija 3.18 (Proces ARMA). Naj bo $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ stacionarni proces in naj za vsak t velja

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (3.23)$$

kjer je $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. $\{X_t\}$ imenujemo avtoregresijski proces drsečih sredin (ang. autoregressive moving average process) reda (p, q) ter krajše označimo kot ARMA(p, q).

Linearno diferenčno enačbo (3.23) lahko zapišemo v kompaktnejši obliki, in sicer

$$\phi(B)X_t = \theta(B)Z_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (3.24)$$

kjer sta ϕ in θ polinoma stopnje p in q :

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \quad (3.25)$$

in

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q. \quad (3.26)$$

Spomnimo se, B je operator pomik nazaj, definiran z

$$B^j X_t = X_{t-j}, \quad j = 0, \pm 1, \pm 2, \dots \quad (3.27)$$

Primer 3.19 (Proces MA(\mathbf{q})). Če je $\phi(z) \equiv 1$, potem velja

$$X_t = \theta(B)Z_t \quad (3.28)$$

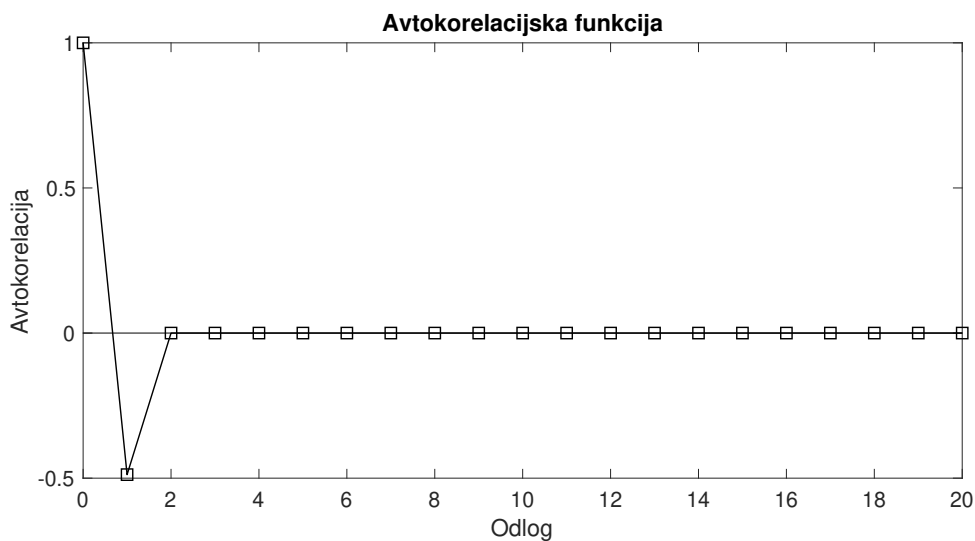
in pravimo, da je časovna vrsta $\{X_t\}$ proces drsečih sredin reda q (ali MA(q)). V tem primeru imajo diferenčne enačbe enolično rešitev (3.28). Velja še, da je rešitev $\{X_t\}$ stacionarni proces. Če definiramo $\theta_0 = 1$ in $\theta_j = 0$ za $j > q$, velja

$$E(X_t) = \sum_{j=0}^q \theta_j E(Z_{t-j}) = 0 \quad (3.29)$$

in

$$\text{Cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & |h| \leq q \\ 0, & |h| > q \end{cases} \quad (3.30)$$

Slika 3 prikazuje ACF procesa $X_t = Z_t - 0.8Z_{t-1}$.

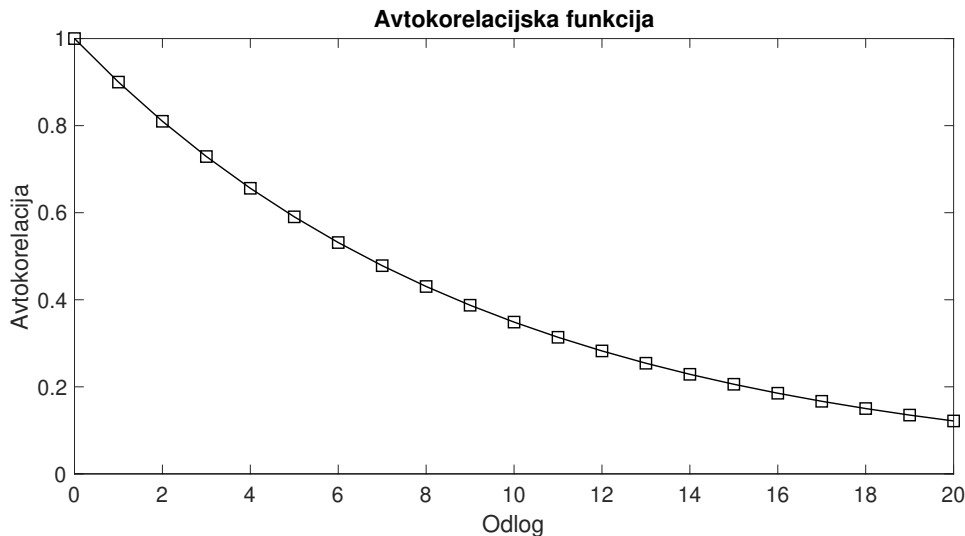


Slika 3: Avtokorelacijska funkcija procesa $X_t = Z_t - 0.8Z_{t-1}$.

Primer 3.20 (Proces AR(\mathbf{p})). Če je $\theta(z) \equiv 1$, potem velja

$$\phi(B)X_t = Z_t \quad (3.31)$$

in pravimo, da je časovna vrsta $\{X_t\}$ avtoregresijski proces reda p (ali AR(p)). V tem primeru (kot tudi v splošnem primeru procesa ARMA, ki ga bomo kasneje obravnavali) obstoj in enoličnost stacionarne rešitve enačbe (3.31) ni očitna. V primeru 3.13 smo že določili ACF procesa AR(1), slika 4 prikazuje ACF procesa $X_t - 0.9X_{t-1} = Z_t$.



Slika 4: Avtokorelacijska funkcija procesa $X_t - 0.9X_{t-1} = Z_t$.

Proces ARMA tako vsebuje komponente avtoregresijskega dela (AR) in komponente drsečih sredin (MA).

Definicija 3.21 (Kavzalnost). Pravimo, da je proces $ARMA(p, q)$, definiran z enačbo $\phi(B)X_t = \theta(B)Z_t$, kavzalen (ali natančneje kavzalna funkcija procesa $\{Z_t\}$), če obstaja tako zaporedje konstant $\{\psi_j\}$, da velja $\sum_{j=0}^{\infty} |\psi_j| < \infty$ in

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \dots \quad (3.32)$$

Opomba 3.22. Iz pogoja $\sum_{j=0}^{\infty} |\psi_j| < \infty$ in neenakosti $E[|Z_t|] \leq \sigma$ sledi

$$E[|X_t|] \leq \sum_{j=0}^{\infty} (|\psi_j| E[|Z_{t-j}|]) \leq \left(\sum_{j=0}^{\infty} |\psi_j| \right) \sigma < \infty. \quad (3.33)$$

Od tod sklepamo, da vrsta v enačbi (3.32) konvergira absolutno (z verjetnostjo 1). Pogoj $\sum_{j=0}^{\infty} |\psi_j| < \infty$ zagotavlja tudi $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Z uporabo Cauchyjevega kriterija se da pokazati, da vrsta v enačbi (3.32) konvergira tudi v L^2 ([7]).

Brez dokaza (na voljo je v [7]) bomo navedli naslednji izrek, ki nam pove, kdaj je proces $ARMA(p, q)$ kavzalen.

Izrek 3.23. Naj bo $\{X_t\}$ proces $ARMA(p, q)$, katerega polinoma $\phi(\cdot)$ in $\theta(\cdot)$ nimata skupnih ničel. Potem je $\{X_t\}$ kavzalen natanko tedaj, ko velja $\phi(z) \neq 0$ za vse $z \in \mathbb{C}$, $|z| \leq 1$. Koeficienti ψ_j v enačbi (3.32) so določeni z zvezo

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z), \quad |z| \leq 1. \quad (3.34)$$

Zgornja zveza za koeficiente ψ_j je ekvivalentna enakosti

$$(1 - \phi_1 z - \dots - \phi_p z^p)(\psi_0 + \psi_1 z + \dots) = 1 + \theta_1 z + \dots + \theta_q z^q. \quad (3.35)$$

Če enačimo koeficiente pred potencami $z^j, j = 0, 1, \dots$, dobimo

$$\begin{aligned} 1 &= \psi_0 \\ \theta_1 &= \psi_1 - \psi_0\phi_1 \\ \theta_2 &= \psi_2 - \psi_0\phi_1 - \psi_0\phi_2 \\ &\vdots \end{aligned}$$

ali ekvivalentno

$$\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = \theta_j, \quad j = 0, 1, \dots, \quad (3.36)$$

kjer je $\theta_0 := 1, \theta_j := 0$ za $j > q$ in $\psi_j := 0$ za $j < 0$.

Opomba 3.24. Pokazati se da [7], da v primeru ko polinoma $\phi(\cdot)$ in $\theta(\cdot)$ nimata skupnih ničel in če je $\phi(z) = 0$ za nek $z \in \mathbb{C}, |z| = 1$, potem enačba $\phi(B)X_t = \theta(B)Z_t$ nima stacionarne rešitve.

Zgornja opomba nam pove, kdaj enačba $\phi(B)X_t = \theta(B)Z_t$ nima stacionarne rešitve. Če pa velja $\phi(z) \neq 0$ za vse $z \in \mathbb{C}, |z| = 1$, dobro poznan rezultat iz kompleksne analize zagotavlja obstoj $r > 1$, tako da velja

$$\theta(z)/\phi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j = \psi(z), \quad r^{-1} < |z| < r, \quad (3.37)$$

Laurentova vrsta je namreč absolutno konvergentna znotraj navedenega kolobarja. Naslednji izrek nam pove, kdaj imajo enačbe, ki določajo proces ARMA, rešitev (dokaz je zopet na voljo v [7]).

Izrek 3.25. *Naj velja $\phi(z) \neq 0$ za vse $z \in \mathbb{C}, |z| = 1$, potem imajo t. i. ARMA enačbe $\phi(B)X_t = \theta(B)Z_t$ enolično stacionarno rešitev*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (3.38)$$

kjer so koeficienti ψ_j določeni z zvezo 3.37. Pri tem zgornja vrsta konvergira absolutno in v L^2 .

Oglejmo si primer izračuna koeficientov ψ_j za proces ARMA(1,1).

Primer 3.26. Naj bo $\{X_t\}$ proces ARMA(1,1), ki zadošča enačbi

$$X_t - 0.5X_{t-1} = Z_t + 0.4Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2). \quad (3.39)$$

Avtoregresijski polinom $\phi(z) = 1 - 0.5z$ ima ničlo pri $z = 2$, ki je zunaj enotskega kroga, zato sklepamo, da obstaja enoličen proces ARMA, ki zadošča zgornji enačbi, ta je tudi kavzalen. Poiskati želimo koeficiente $\{\psi_j\}$ iz zapisa 3.32, ki jih lahko določimo s pomočje zveze (3.36):

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= 0.4 + 0.5 \\ \psi_2 &= 0.5(0.4 + 0.5) \\ \psi_j &= 0.5^{j-1}(0.4 + 0.5), \quad j = 1, 2, \dots \end{aligned}$$

3.5.2 ACF procesa ARMA(p, q)

Spoznali smo že definicijo avtokorelacijske funkcije, sedaj pa jo bomo za proces ARMA(p, q) tudi določili. Najprej bomo določili avtokovariančno funkcijo, saj ACF preprosto dobimo tako, da ACVF delimo z $\gamma(0)$. Tudi parcialna avtokorelacijska funkcija, ki jo bomo spoznali v nadaljevanju, je funkcija $\gamma(\cdot)$.

Naj bo $\{X_t\}$ kavzalni proces ARMA(p, q), definiran z

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2), \quad (3.40)$$

kjer sta $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ in $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. Privzetek kavzalnosti implicira zvezo

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad (3.41)$$

kjer je $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$, $|z| \leq 1$. Izračun zaporedja konstant $\{\psi_j\}$ smo že obravnavali v prejšnjem poglavju (enačba (3.36)).

Določimo ACVF procesa $\{X_t\}$. Najprej razpišimo zvezo (3.40) v

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}. \quad (3.42)$$

Nato obe strani zgornje enakosti pomnožimo z X_{t-k} , $k = 0, 1, 2, \dots$ in izračunamo pričakovano vrednost obeh strani. Dobimo naslednje:

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j} \psi_j, \quad 0 \leq k < m \quad (3.43)$$

in

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = 0, \quad k \geq m, \quad (3.44)$$

kjer je $m = \max(p, q+1)$, $\psi_j := 0$ za $j < 0$, $\theta_0 := 1$ in $\theta_j := 0$ za $j \notin \{0, \dots, q\}$. Pri računanju desne strani v (3.43) smo uporabili enačbo (3.41). Avtokovariance izračunamo tako, da rešimo prvih $p+1$ enačb v (3.43) in (3.44) za $\gamma(0), \dots, \gamma(p)$, te nato uporabimo za zaporedno računanje $\gamma(p+1), \gamma(p+2), \dots$. Načinov za določanje ACVF je sicer več, zgornji je zlasti smiseln, če za računanje uporabljamo računalnik. Oglejmo si primer določitve ACVF za proces ARMA(1,1).

Primer 3.27. Naj bo $\{X_t\}$ kavzalni ARMA(1,1) proces, definiran z enačbo

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2), \quad (3.45)$$

kjer je $|\phi| < 1$. Enačbi (3.43) se glasita:

$$\gamma(0) - \theta \gamma(1) = \sigma^2(1 + \theta(\theta + \phi)) \quad (3.46)$$

in

$$\gamma(1) - \theta \gamma(0) = \sigma^2 \theta. \quad (3.47)$$

Krajši račun nam da

$$\gamma(0) = \sigma^2 \left(1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right) \quad \text{in} \quad \gamma(1) = \sigma^2 \left(\theta + \phi + \frac{(\theta + \phi)^2 \phi}{1 - \phi^2} \right). \quad (3.48)$$

Enačba (3.44) ima obliko

$$\gamma(k) - \theta\gamma(k-1) = 0, \quad k \geq 2, \quad (3.49)$$

zaporedno lahko torej izračunamo $\gamma(2), \gamma(3), \dots$. Določiti znamo ACVF, spomnimo pa se, da ACF procesa $\{X_t\}$ preprosto dobimo iz ACVF $\gamma(\cdot)$ kot

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}. \quad (3.50)$$

Pokazati se da naslednje [7]: za proces MA(q) je ACF enaka 0 za odloge večje od q , za proces AR(p) pa ACF sicer ni ničelna, temveč je omejena z geometrijsko padajočo funkcijo. Omenjeno lastnost bomo uporabili pri izbiri ustreznega modela za dane podatke, kar bomo v nadaljevanju tudi natančno opisali.

3.5.3 Napovedovanje procesov ARMA

Omenili smo že parcialno avtokorelacijsko funkcijo (PACF), ki poleg ACF igra pomembno vlogo pri modeliranju procesov ARMA. Tako kot ACF tudi PACF nosi pomembne informacije o strukturi odvisnosti stacionarnega procesa. Preden jo definiramo, pa se bomo posvetili še napovedovanju stacionarnih časovnih vrst.

Uvodoma bomo obravnavali problem napovedovanja vrednosti X_{n+h} , $h > 0$ za stacionarno časovno vrsto z znanim povprečjem μ in avtokovariančno funkcijo γ na osnovi vrednosti X_n, \dots, X_1 . Cilj je najti linearno kombinacijo $1, X_n, X_{n-1}, \dots, X_1$, s katero napovemo X_{n+h} z najmanjšo srednjo kvadratno napako. Najboljši linearni prediktor za X_{n+h} glede na $1, X_n, \dots, X_1$ bomo označili kot $P_n X_{n+h}$ in ima očitno obliko

$$P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1. \quad (3.51)$$

Določiti moramo koeficiente a_0, a_1, \dots, a_n , tako da minimiziramo

$$S(a_0, \dots, a_n) = E(X_{n+h} - a_0 - a_1 X_n - \dots - a_n X_1)^2. \quad (3.52)$$

S je kvadratna funkcija spremenljivk a_0, \dots, a_n , ki je navzdol omejena z 0, zato očitno obstaja vsaj ena vrednost (a_0, \dots, a_n) , ki minimizira S in minimum (a_0, \dots, a_n) zadošča enačbam

$$\frac{\partial S(a_0, \dots, a_n)}{\partial a_j} = 0, \quad j = 0, \dots, n. \quad (3.53)$$

Iz zgornje enačbe sledi

$$E \left[X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i} \right] = 0, \quad (3.54)$$

$$E \left[(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}) X_{n+1-j} \right] = 0, \quad j = 1, \dots, n. \quad (3.55)$$

Enačbe lahko lepše vektorsko zapišemo kot

$$a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right) \quad (3.56)$$

in

$$\mathbf{\Gamma}_n \mathbf{a}_n = \gamma_n(h), \quad (3.57)$$

kjer so

$$\mathbf{a}_n = (a_1, \dots, a_n)^T, \quad \mathbf{\Gamma}_n = [\gamma(i-j)]_{i,j=1}^n$$

in

$$\gamma_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))^T.$$

Od tod sledi

$$P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu), \quad (3.58)$$

kjer \mathbf{a}_n zadošča enačbi (3.57). Sledi, da je pričakovana vrednost napovedne napake $X_{n+h} - P_n X_{n+h}$ enaka 0, srednja kvadratna napaka napovedne napake pa je enaka

$$\begin{aligned} E[(X_{n+h} - P_n X_{n+h})^2] &= \gamma(0) - 2 \sum_{i=1}^n a_i \gamma(h+i-1) + \sum_{i=1}^n \sum_{j=1}^n a_i \gamma(i-j) a_j \\ &= \gamma(0) - \mathbf{a}_n^T \gamma_n(h), \end{aligned} \quad (3.59)$$

kjer zadnja vrstica sledi iz enačbe (3.57). Če povzamemo lastnosti linearnega prediktorja $P_n X_{n+h}$:

1. $P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$, kjer $\mathbf{a}_n = (a_1, \dots, a_n)^T$ zadošča (3.57),
2. $E[(X_{n+h} - P_n X_{n+h})^2] = \gamma(0) - \mathbf{a}_n^T \gamma_n(h)$,
3. $E(X_{n+h} - P_n X_{n+h}) = 0$,
4. $E[(X_{n+h} - P_n X_{n+h})X_j] = 0, \quad j = 1, \dots, n$.

Oglejmo si primer napovedi procesa AR(1) za en koraj naprej.

Primer 3.28. Naj bo $\{X_t\}$ stacionarna časovna vrsta, definirana z enačbami

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \dots, \quad (3.60)$$

kjer je $|\phi| < 1$ in $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. Zgornji proces AR(1) smo že obravnavali v primeru 3.13, kjer smo ugotovili, da velja $\gamma_X(h) = \phi^h \gamma_X(0)$ in $\gamma_X(0) = \sigma^2/(1 - \phi^2)$. Iz enačb (3.57) in (3.58) sledi, da je najboljši linearni prediktor za X_{n+1} na osnovi $\{1, X_n, \dots, X_1\}$ enak (za $n \geq 1$)

$$P_n X_{n+1} = \mathbf{a}_n^T \mathbf{X}_n, \quad (3.61)$$

kjer je $\mathbf{X}_n = (X_n, \dots, X_1)^T$ in

$$\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & & & & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix} \quad (3.62)$$

Rešitev zgornje enačbe je očitno

$$\mathbf{a}_n = (\phi, 0, \dots, 0)^T, \quad (3.63)$$

in zato je najboljši linearni prediktor za X_{n+1} na osnovi $\{X_n, \dots, X_1\}$ enak

$$P_n X_{n+1} = \mathbf{a}_n^T \mathbf{X}_n = \phi X_n. \quad (3.64)$$

Srednja kvadratna napaka je enaka

$$E[(X_{n+1} - P_n X_{n+1})^2] = \gamma(0) - \mathbf{a}_n^T \gamma_n(1) = \frac{\sigma^2}{1 - \phi^2} - \phi \gamma(1) = \sigma^2. \quad (3.65)$$

Privzemimo sedaj, da so Y in W_n, \dots, W_1 poljubne slučajne spremenljivke s končnimi drugimi momenti ter da so povprečja $\mu_Y = E(Y)$, $\mu_i = E(W_i)$ in kovariance $\text{Cov}(Y, Y)$, $\text{Cov}(Y, W_i)$ in $\text{Cov}(W_i, W_j)$ znane. Označimo slučajni vektor $\mathbf{W} = (W_n, \dots, W_1)^T$, pripadajoč vektor povprečij $\mu_W = (\mu_n, \dots, \mu_1)^T$, vektor kovarianc

$$\gamma = \text{Cov}(Y, \mathbf{W}) = (\text{Cov}(Y, W_n), \text{Cov}(Y, W_{n-1}), \dots, \text{Cov}(Y, W_1))^T \quad (3.66)$$

in kovariančno matriko

$$\mathbf{\Gamma} = \text{Cov}(\mathbf{W}, \mathbf{W}) = [\text{Cov}(W_{n+1-i}, W_{n+1-j})]_{i,j=1}^n. \quad (3.67)$$

Postopamo enako kot pri računanju $P_n X_{n+h}$ in dobimo, da je najboljši linearni prediktor za Y na osnovi $\{1, W_n, \dots, W_1\}$ enak

$$P(Y|\mathbf{W}) = \mu_Y + \mathbf{a}^T (\mathbf{W} - \mu_W), \quad (3.68)$$

kjer je $\mathbf{a} = (a_1, \dots, a_n)^T$ rešitev enačbe

$$\mathbf{\Gamma} \mathbf{a} = \gamma. \quad (3.69)$$

Srednja kvadratna napaka prediktorja je enaka

$$E[(Y - P(Y|\mathbf{W}))^2] = \text{Var}(Y) - \mathbf{a}^T \gamma. \quad (3.70)$$

Za poljubne $\mathbf{W} = (W_n, \dots, W_1)^T$ in Y s končnim drugim momentom znamo izračunati najboljši linearni prediktor $P(Y|\mathbf{W})$ za Y na osnovi $1, W_n, \dots, W_1$. Funkcijo $P(\cdot|\mathbf{W})$, ki pretvori Y v $P(Y|\mathbf{W})$ imenujemo operater napovedovanja (ang. prediction operator). Operator P_n , definiran z enačbama (3.57) in (3.58), je primer, ko je $\mathbf{W} = (X_n, X_{n-1}, \dots, X_1)^T$. Operator napovedovanja ima številne koristne lastnosti, ki jih lahko uporabimo za enostavnejše računanje najboljših linearnih prediktorjev. Našteli bomo nekatere izmed njih.

Predpostavimo, da velja $E(U^2) < \infty$, $E(V^2) < \infty$, $\mathbf{\Gamma} = \text{Cov}(\mathbf{W}, \mathbf{W})$, β, α_i , $i = 1, \dots, n$, pa so konstante. Potem veljajo naslednje lastnosti:

1. $P(U|\mathbf{W}) = E(U) + \mathbf{a}^T(\mathbf{W} - E(\mathbf{W}))$, kjer je $\mathbf{\Gamma}\mathbf{a} = \text{Cov}(U, \mathbf{W})$,
2. $E[(U - P(U|\mathbf{W}))\mathbf{W}] = 0$ in $E[U - P(U|\mathbf{W})] = 0$,
3. $E[(U - P(U|\mathbf{W}))^2] = \text{Var}(U) - \mathbf{a}^T \text{Cov}(U, \mathbf{W})$,
4. $P(\alpha_1 U + \alpha_2 V + \beta | \mathbf{W}) = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta$,
5. $P(\sum_{i=1}^n \alpha_i W_i + \beta | \mathbf{W}) = \sum_{i=1}^n \alpha_i W_i + \beta$,
6. $P(U|\mathbf{W}) = E(U)$, če je $\text{Cov}(U, \mathbf{W}) = 0$.

Oglejmo si še primer napovedovanja procesa $\text{AR}(p)$ za en korak naprej.

Primer 3.29. Naj bo $\{X_t\}$ stacionarna časovna vrsta, dana z enačbami

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t, \quad t = 0, \pm 1, \dots, \quad (3.71)$$

kjer je $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, Z_t in X_s sta nekorelirana za vsak $s < t$. Če je $n > p$, lahko operator napovedovanja P_n apliciramo na obeh straneh zgornje enačbe. Uporabimo lastnosti (4), (5) in (6), da dobimo

$$P_n X_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p}. \quad (3.72)$$

Opomba 3.30. V splošnem velja naslednje. Naj bo $\{Y_t\}$ stacionarna časovna vrsta s povprečjem μ in naj bo $\{X_t\}$ časovna vrsta z ničelnim povprečjem, definirana kot $X_t = Y_t - \mu$. Potem je linearni prediktor poljubne slučajne spremenljivke W na osnovi $1, Y_n, \dots, Y_1$ enak kot linearni prediktor na osnovi $1, X_n, \dots, X_1$. Če ta prediktor označimo s $P_n W$ in apliciramo P_n na enačbo $Y_{n+h} = X_{n+h} + \mu$, dobimo

$$P_n Y_{n+h} = \mu + P_n X_{n+h}. \quad (3.73)$$

Najboljši linearni prediktor za Y_{n+h} lahko torej najdemo tako, da poiščemo najboljši linearni prediktor za X_{n+h} in nato prištejemo μ . Ker je $E(X_t) = 0$, iz enačbe (3.58) sledi, da je $P_n X_{n+h}$ enak kot najboljši linearni prediktor za X_{n+h} na osnovi X_n, \dots, X_1 .

Zgornja opomba nam pove, da se lahko osredotočimo na stacionarne časovne vrste z ničelnim povprečjem. Če je $\{X_t\}$ stacionarna časovna vrsta z ničelnim povprečjem in avtokovariančno funkcijo $\gamma(\cdot)$, potem enačbi (3.68) in (3.69) rešita problem določanja najboljšega linearnega prediktorja $P_n X_{n+h}$ za X_{n+h} na osnovi $\{X_n, \dots, X_1\}$. Direktn pristop zahteva določitev rešitve sistema n linearnih enačb, kar je lahko za velike n zahtevno in zamudno. Koristno bi bilo, če bi prediktor $P_n X_{n+1}$ na osnovi prejšnjih n opazanj uporabili za izračun prediktorja $P_{n+1} X_{n+2}$ na osnovi prejšnjih $n+1$ opazanj. Algoritmom, ki uporabijo to idejo, pravimo rekurzivni. Pomemba primera sta Durbin-Levinsonov algoritem in inovacijski algoritem (ang. the innovations algorithm). Slednjega bomo predstavili v nadaljevanju, oba pa sta podrobno opisana v [7].

Seznanili smo se z linearnimi prediktorji, zato si lahko ogledamo definicijo parcialne avtokorelacijske funkcije.

Definicija 3.31. Parcialna avtokorelacijska funkcija (PACF) stacionarne časovne vrste je definirana z enačbama

$$\alpha(0) = 1 \quad (3.74)$$

in

$$\alpha(h) = \phi_{hh}, \quad h \geq 1, \quad (3.75)$$

kjer je ϕ_{hh} zadnja komponenta

$$\phi_h = \Gamma_h^{-1} \gamma_h, \quad (3.76)$$

$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$ in $\gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]^T$. Vrednosti $\alpha(h)$ pravimo parcialna avtokorelacija pri odlogu h .

Pokazati se da [7], da ϕ_{hh} predstavlja korelacijo med napovednima napakama $X_h - P(X_h|X_1, \dots, X_{h-1})$ in $X_0 - P(X_0|X_1, \dots, X_{h-1})$. Vzorčna parcialna avtokorelacijska funkcija je definirana na podoben način.

Definicija 3.32. Vzorčna parcialna avtokorelacija $\hat{\alpha}(h)$ pri odlogu h za vzorec $\{x_1, \dots, x_n\}$, za katerega velja $x_i \neq x_j$ za neka i in j , je definirana kot

$$\hat{\alpha}(0) = 1 \quad (3.77)$$

in

$$\hat{\alpha}(h) = \hat{\phi}_{hh}, \quad h \geq 1, \quad (3.78)$$

kjer je $\hat{\phi}_{hh}$ zadnja komponenta

$$\hat{\phi}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h. \quad (3.79)$$

Kot smo že omenili, je PACF ključnega pomena pri modeliranju procesov ARMA, pokazati se namreč da naslednje [7]: za proces $AR(p)$ je PACF enaka 0 za odloge večje od p . Za proces $MA(q)$ pa PACF sicer ne izgine za velike odloge, temveč je omejena z geometrijsko padajočo funkcijo.

3.5.4 Modeli ARIMA za nestacionarne časovne vrste

Obravnavali smo že razred modelov ARMA, s katerimi lahko predstavimo stacionarne časovne vrste. Posplošitev tega razreda, ki vključuje velik nabor nestacionarnih časovnih vrst, so procesi ARIMA. To so procesi, ki se po končnem številu diferenciranj reducirajo v proces ARMA. Oglejmo si definicijo procesa ARIMA.

Definicija 3.33 (Proces $ARIMA(p, d, q)$). Naj bo d nenegativno celo število in

$$Y_t := (1 - B)^d X_t \quad (3.80)$$

kavzalni proces $ARMA(p, q)$. Procesu $\{X_t\}$ pravimo avtoregresijski integrirani proces drsečih sredin reda (p, d, q) in krajše označimo kot $ARIMA(p, d, q)$.

Zgornja definicija pomeni, da $\{X_t\}$ zadošča diferenčni enačbi oblike

$$\phi^*(B)X_t \equiv \phi(B)(1 - B)^d X_t = \theta(B)Z_t, \quad (3.81)$$

kjer je $Z_t \sim \text{WN}(0, \sigma^2)$, $\phi(z)$ in $\theta(z)$ sta polinoma stopenj p in q ter velja $\phi(z) \neq 0$ za $|z| \leq 1$. Polinom $\phi^*(z)$ ima ničlo stopnje d pri $z = 1$. Proces $\{X_t\}$ je stacionaren natanko tedaj, ko je $d = 0$, v tem primeru je to kar proces ARMA(p, q).

Če je $d \geq 1$, lahko časovni vrsti $\{X_t\}$ dodamo poljuben polinomski trend stopnje $d - 1$, brez da bi kršili diferenčno enačbo (3.81). Modeli ARIMA so tako primerni za predstavitev podatkov s trendom.

Značilnost podatkov, ki kaže na ustreznost modela ARIMA, je počasi padajoča pozitivna vzorčna avtokorelacijska funkcija. Če bi imeli na voljo samo podatke in bi želeli najti ustrezen model, bi bilo naravno večkrat uporabiti operator $\nabla = 1 - B$, v upanju, da bo imel za nek j proces $\{\nabla^j X_t\}$ hitro padajočo vzorčno avtokorelacijsko funkcijo, ki bi lahko pripadala procesu ARMA, katerega avtoregresijski polinom nima ničel v bližini enotskega kroga.

3.5.5 Sezonski modeli ARIMA

Za sezone časovne vrste je značilna močna avtokorelacija pri sezonskem odlogu (in morda tudi njegovih večkratnikih). Obravnavali smo že dekompozicijo časovne vrste $X_t = m_t + s_t + Y_t$, kjer je m_t komponenta trenda, s_t sezonska komponenta, Y_t pa komponenta naključnega šuma. Vendar v praksi morda ni smiselno predpostaviti, da se sezonska komponenta ponavlja na enak način cikel za ciklom. Sezonski modeli ARIMA omogočajo naključnost v sezonskem vzorcu od enega do drugega cikla.

Definicija 3.34 (Proces SARIMA(p, d, q) \times (P, D, Q) $_S$). Naj bosta d in D ne-negativni celi števili, procesu $\{X_t\}$ pravimo sezonski SARIMA(p, d, q) \times (P, D, Q) $_S$ proces s periodo S , če je diferenciran proces

$$Y_t := (1 - B)^d (1 - B^S)^D X_t \quad (3.82)$$

kavzalni proces ARMA, ki zadošča enačbi

$$\phi(B)\Phi(B^S)Y_t = \theta(B)\Theta(B^S)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2). \quad (3.83)$$

Pri tem so $\phi(z) = 1 - \phi_1(z) - \dots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1(z) - \dots - \Phi_P z^P$, $\theta(z) = 1 + \theta_1(z) + \dots + \theta_q z^q$ in $\Theta(z) = 1 + \Theta_1(z) + \dots + \Theta_Q z^Q$.

Proces $\{Y_t\}$ je kavzalen natanko tedaj, ko velja $\phi(z) \neq 0$ in $\Phi(z) \neq 0$ za $|z| \leq 1$. V praksi je parameter D redko več kot 1, parametra P in Q pa sta tipično manjša od 3.

3.6 Modeli GARCH

V modelih ARMA (in posplošitvah) smo privzeli konstantno varianco, vendar pa ta homoskedastični privzetek ne velja nujno vedno, zlasti pri analizi finančnih podatkov. Zato potrebujemo razred modelov, pri katerih predpostavka konstantne variance ne velja, torej heteroskedastične. Osredotočili se bomo na model GARCH

(in nekaj posplošitev), ki se je razvil iz modela ARCH, zato bomo najprej predstavili še tega. Oba sta bila razvita za potrebe analize finančnih časovnih vrst, da bi odražala značilnosti finančnih podatkov, med katere sodijo težek rep, asimetrija in skupinjenje volatilitosti (ang. volatility clustering). Teh lastnosti ni mogoče zajeti s tradicionalnimi linearnimi modeli časovnih vrst, kot so modeli ARMA, ki smo jih obravnavali v prejšnjem poglavju.

Če s P_t označimo ceno delnice ali drugega finančnega sredstva ob času t , $t \in \mathbb{Z}$, potem je časovna vrsta logaritmskih donosov (ang. log returns) $\{X_t := \log P_t - \log P_{t-1}\}$ tipično modelirana kot stacionarna časovna vrsta. Model ARMA za časovno vrsto $\{X_t\}$ bi imel lastnost, da je pogojna varianca h_t časovne vrste X_t glede na $\{X_s, s < t\}$ neodvisna od t in $\{X_s, s < t\}$. Vendar celo površen pregled večine empiričnih časovnih vrst logaritmskih donosov močno kaže, da temu ni tako.

3.6.1 Model ARCH

Da bi lahko modelirali spreminjajočo se volatilitost, je Robert F. Engle leta 1982 predstavil proces ARCH (avtoregresijska pogojna heteroskedastičnost, ang. autoregressive conditional heteroscedasticity), kjer je pogojna varianca odvisna od časa. Proces ARCH(p) $\{Z_t\}$ je definiran kot stacionarna rešitev enačb

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{NEP } N(0,1), \quad (3.84)$$

kjer je h_t pozitivna funkcija procesa $\{Z_s, s < t\}$ (znana kot volatilitost), definirana z

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2. \quad (3.85)$$

Pri tem je $\alpha_0 > 0$ in $\alpha_i \geq 0$, $i = 1, \dots, p$, h_t pa je pogojna varianca časovne vrste Z_t glede na $\{Z_s, s < t\}$.

Več o obstoju stacionarne rešitve bomo povedali v naslednjem podpoglavju, ko bomo obravnavali proces GARCH.

3.6.2 Model GARCH

Leta 1986 je Tim Bollerslev predstavil proces GARCH(p, q) (posplošen ARCH, ang. generalized ARCH), ki danes predstavlja standardno orodje pri modeliranju volatilitosti. Proces predpostavlja splošnejšo zvezo za varianco, enačba (3.85) je tako zamenjana z

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \quad (3.86)$$

kjer so $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p$ in $\beta_j \geq 0$, $j = 1, \dots, q$. Za $q = 0$ je to proces ARCH(p), za $p = q = 0$ pa je Z_t beli šum. V procesu ARCH(p) je pogojna varianca določena kot linearna funkcija preteklih vzorčnih varianc, medtem ko proces GARCH(p, q) dovoli tudi pretekle pogojne variance.

Enačbo (3.86) lahko z uporabo operatorja pomik nazaj prepišemo v

$$h_t = \alpha_0 + \alpha(B)Z_t^2 + \beta(B)h_t, \quad (3.87)$$

kjer sta

$$\alpha(B) = \sum_{i=1}^p \alpha_i B^i \quad \text{in} \quad \beta(B) = \sum_{i=1}^q \beta_i B^i. \quad (3.88)$$

Če vse ničle $1 - \beta(z) = 0$ ležijo zunaj enotskega kroga, lahko enačbo (3.87) prepišemo v

$$\begin{aligned} h_t &= \alpha_0(1 - \beta(1))^{-1} + \alpha(B)(1 - \beta(B))^{-1} Z_t^2 \\ &= \alpha_0 \left(1 - \sum_{i=1}^q \beta_i\right)^{-1} + \sum_{i=1}^{\infty} \delta_i Z_{t-i}^2, \end{aligned} \quad (3.89)$$

kar si lahko skupaj z enačbo (3.84) predstavljamo kot neskončno-dimenzionalni proces ARCH(∞). Koeficiente δ_i dobimo iz zapisa $\delta(B) = \alpha(B)(1 - \beta(B))^{-1}$ kot potenčne vrste. Naslednji izrek nam pove, kdaj obstaja stacionarna rešitev procesa GARCH (dokaz je na voljo v [5]).

Izrek 3.35. *Stacionarna rešitev Z_t t. i. GARCH enačb (enačbi 3.84 in 3.86) obstaja natanko tedaj, ko velja $\alpha(1) + \beta(1) < 1$. V tem primeru velja še $E(Z_t) = 0$, $\text{Var}(Z_t) = \alpha_0(1 - \alpha(1) - \beta(1))^{-1}$ ter $\text{Cov}(Z_t, Z_s) = 0$ za $t \neq s$.*

Proces GARCH(p, q) lahko alternativno zapišemo kot

$$Z_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2 + \sum_{j=1}^q \beta_j Z_{t-i}^2 - \sum_{j=1}^q \beta_j \nu_{t-j} + \nu_t, \quad (3.90)$$

kjer sta

$$\nu_t = Z_t^2 - h_t = (e_t^2 - 1)h_t \quad \text{in} \quad e_t \sim \text{NEP } N(0, 1). \quad (3.91)$$

Enačbo (3.90) lahko krajše zapišemo kot

$$(1 - \alpha(B) - \beta(B))Z_t^2 = \alpha_0 + (1 - \beta(B))\nu_t. \quad (3.92)$$

Iz definicije sledi, da je $E(\nu_t) = 0$. Proces GARCH(p, q) lahko torej obravnavamo kot proces ARMA za Z_t^2 reda $m = \max\{p, q\}$ in q . To dejstvo bomo uporabili v nadaljevanju v sklopu validacije modela.

Pri analizi empiričnih finančnih podatkov velikokrat boljše prilaganje podatkom dosežemo s sprostitvijo Gaussove predpostavke v (3.84). Namesto tega predpostavimo, da ima porazdelitev Z_t glede na $\{Z_s, s < t\}$ težji rep, še vedno pa povprečje 0, kot npr. Studentova t -porazdelitev. Če namesto pogojne normalne porazdelitve uporabimo Studentovo t -porazdelitev, model označimo kot GARCH- t . Porazdelitev t je uporabna za modeliranje časovnih vrst z večjim številom ekstremnih vrednosti, kot je pričakovano pri Gaussovi porazdelitvi (ekscresna sploščenost oz. kurtosis). Da vključimo tudi takšno porazdelitev, lahko splošni GARCH(p, q) proces definiramo kot stacionarni proces $\{Z_t\}$, ki zadošča enačbi (3.86) in posplošeni različici enačbe (3.84), dani z zvezo

$$Z_t = \sqrt{h_t} e_t, \quad \text{kjer je} \quad \{e_t\} \sim \text{NEP}(0, 1). \quad (3.93)$$

Za namene modeliranja navadno privzamemo, da velja

$$e_t \sim N(0, 1) \quad (3.94)$$

(kot v (3.84)) ali pa

$$\sqrt{\frac{\nu}{\nu - 2}} e_t \sim t_\nu, \quad \nu > 2. \quad (3.95)$$

Pri tem je t_ν Studentova t -porazdelitev z ν prostostnimi stopnjami. Skalirni faktor je dodan z namenom, da je varianca e_t enaka 1. Uporabimo lahko tudi druge porazdelitve za e_t , vendar se bomo osredotočili samo na omenjeni dve.

Za finančne podatke je značilna tudi vztrajnost volatilitosti (ang. persistence of volatility), to je pojav, ko velikim (majhnim) nihanjem v podatkih navadno sledijo nihanja s primerljivo velikostjo. Modeli GARCH to odražajo z vključitvijo korelacije v zaporedju pogojnih varianc $\{h_t\}$, prav tako pa lahko upoštevajo težji rep porazdelitve.

Funkcija pogojne variance standardnega modela GARCH je simetrična v preteklih vrednostih Z_t . To pomeni, da ne more upoštevati asimetrije glede pozitivnih in negativnih vrednosti e_t (imenujemo jih šok). Potrebujemo torej posplošitev modela GARCH, da lahko opišemo še to značilnost.

3.6.3 Model EGARCH

Daniel B. Nelson je leta 1991 predstavil model EGARCH (eksponentni GARCH). Z njim imajo lahko negativne in pozitivne vrednosti e_t v definiciji procesa GARCH različne vplive na nadaljnje volatilitosti h_s , ($s > t$). Opisuje torej asimetrični odziv pogojne variance na pozitivne in negativne vrednosti e_t .

Proces EGARCH(p, q) je definiran na naslednji način. Naj bo proces $\{Z_t\}$ definiran z enačbo

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{NEP}(0, 1), \quad (3.96)$$

kjer je $\{l_t := \ln h_t\}$ šibko in strogo stacionarna rešitev enačbe

$$l_t = c + \alpha(B)g(e_t) + \gamma(B)l_t. \quad (3.97)$$

Pri tem je $c \in \mathbb{R}$,

$$g(e_t) = e_t + \lambda(|e_t| - E(|e_t|)) \quad (3.98)$$

in

$$\alpha(B) = \sum_{i=1}^p \alpha_i B^i, \quad \gamma(B) = \sum_{i=1}^q \gamma_i B^i. \quad (3.99)$$

Očitno so $\{l_t\}$, $\{h_t\}$ in $\{Z_t\}$ strogo stacionarni in kavzalni, če $1 - \gamma(z)$ ni enako 0 za vse kompleksne z , $|z| \leq 1$. e_t ima simetrično porazdelitev okrog ničle, tj. $e_t \stackrel{d}{=} -e_t$.

Proces je definiran na način, da zagotavlja, da je $h_t (= e^{l_t}) > 0$. Enačbo (3.98) lahko prepišemo v

$$g(e_t) = \begin{cases} (1 + \lambda)e_t - \lambda E(|e_t|); & e_t \geq 0 \\ (1 - \lambda)e_t - \lambda E(|e_t|); & e_t < 0 \end{cases} \quad (3.100)$$

Iz zgornjega je razvidno, da je funkcija g odsekoma linearna z naklonom $(1 + \lambda)$ na $(0, \infty)$ in naklonom $(1 - \lambda)$ na $(-\infty, 0)$. Ta asimetrija v funkciji g dovoli različen odziv l_t na pozitivne in negativne šoke e_{t-i} , ki imajo enako velikost. Če je $\lambda = 0$, potem ni asimetrije.

Ko obravnavamo model EGARCH za primer cen delnic, se navadno izkaže, da je ocenjena vrednost λ negativna, kar ustreza temu, da imajo negativni šoki večji vpliv na volatilitnost kot pozitivni šoki enake velikosti.

Omenili smo že časovno vrsto logaritemskih donosov, za katero se v praksi pogosto izkaže, da jo je ugodno modelirati kot proces GARCH (oz. njegovo posplošitev). V primeru časovne vrste cene in povpraševanja po električni energiji (ki nas zanima) pa Z_t (kot v enačbi 3.93 ali 3.96) ne predstavlja celotnega modela za ceno in povpraševanje, temveč je samo gradnik obsežnejšega modela. V našem primeru je to sezonski model ARIMA, katerega komponento šuma Z_t modeliramo kot proces GARCH ali njegovo različico (Z_t v tem primeru ni več beli šum). Še vedno pa velja predpostavka šibke stacionarnosti, tako da celotna teorija procesov ARMA še vedno velja.

3.7 Identifikacija modela in napovedovanje

V tem poglavju se bomo osredotočili na problem izbire primerne modela za dano množico podatkov $\{X_t, t = 1, \dots, n\}$. Uvodoma bomo predstavili Box-Jenkinsovo metodologijo, s pomočjo katere bomo izbrali ustrezen model ARMA. Predstavili bomo tudi informacijski kriterij AIC, ki nam bo pomagal vrednotiti modele in jih primerjati med seboj. Osredotočili se bomo tudi na ocenjevanje parametrov modela in diagnostične teste, s katerimi preverimo, če je izbran model res primeren za naše podatke. Predstavili bomo, kako izdelamo napovedi za prihodnost na podlagi izbranega modela. Na koncu bomo obravnavali še modeliranje variance z modeli GARCH, v primeru ko ta ni konstantna.

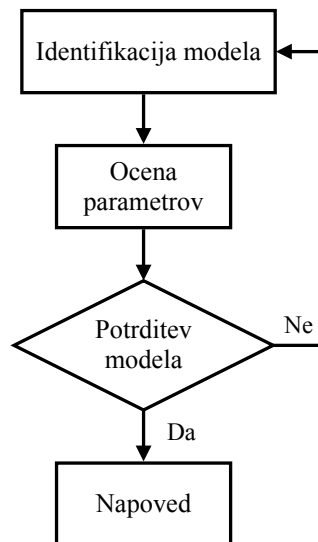
3.7.1 Box-Jenkinsova metodologija

Izbira modelov poteka po Box-Jenkinsovi metodologiji. To je standardna metoda v analizi časovnih vrst, v sklopu katere izberemo model ARMA, ki se najbolj prilega dani časovni vrsti. Iterativni pristop je sestavljen iz treh delov (v nadaljevanju jih bomo natančneje obrazložili):

- Najprej poteka identifikacija in izbira modela. Uvodoma preverimo, če je časovna vrsta stacionarna, identificiramo možne sezonskosti (in sezonsko diferenciramo po potrebi). Pri tem se opremo na grafa ACF in PACF, tudi pri določanju komponent avtoregresijskega dela in dela drsečih sredin.

- Ko smo model identificirali, sledi ocenjevanje parametrov z izbranim algoritmom. Tako dobimo koeficiente, ki se najboljše prilagajajo izbranemu modelu ARMA.
- Na koncu moramo model še potrditi. Preveriti moramo, ali ocenjeni model ustreza specifikacijam stacionarnega procesa. Natančneje, residuali morajo biti nekorelirani, njihovo povprečje in varianca pa se skozi čas ne smeta spreminjati. Tvoriti morajo torej beli šum. Zopet si pomagamo z grafoma ACF in PACF, tokrat za residue. Opravimo tudi Ljung-Boxov test za različne odloge. Če modela ne uspemo potrditi, se moramo vrniti na prvi korak. V nasprotnem je model primeren za napovedovanje.

Če smo v zadnjem koraku model potrdili, ga lahko uporabimo za napovedovanje. Diagram na sliki 5 prikazuje zgoraj opisano metodologijo.



Slika 5: Box-Jenkinsova metodologija.

3.7.2 Identifikacija modela

Prvi korak pri konstrukciji Box-Jenkinsovega modela je ugotoviti, ali je časovna vrsta stacionarna in ali je potrebno modelirati kakšno pomembno sezonskost. V ta namen si navadno ogledamo graf časovne vrste in vzorčno avtokorelacijsko funkcijo. Iz grafa časovne vrste je možno zaznati odvisnost variabilnosti od nivoja, na katerem je časovna vrsta (npr. višja variabilnost za visoke vrednosti časovne vrste) ali pa veliko variabilnost podatkov. V tem primeru je smiselno na podatkih opraviti logaritemsko transformacijo.

Zdaj nas zanima, ali ima časovna vrsta značilnosti, ki kažejo na nestacionarnost. Če si ogledamo graf (morda predhodno transformirane) časovne vrste, morata biti povprečje in varianca konstantna v času. Ne smemo torej zaslediti trenda in sezonskosti. Vendar pa samo pregled grafa ni zagotovilo, da je časovna vrsta res stacionarna. V ta namen si ogledamo še graf vzorčne ACF, prisotnost trenda in sezonskosti nakazuje ACF, ki počasi pada (ali pa če sploh ne pada) in je periodična. V

poglavju o transformacijah časovnih vrst smo že obravnavali, kako z diferenciranjem izločimo trend in sezonskost ter tako dobimo stacionarno časovno vrsto, katere ACF hitro pada.

Ko torej odkrijemo, da je časovna vrsta nestacionarna, si pomagamo z diferenciranjem (ta pristop priporočata Box in Jenkins). Morda je potrebno tudi sezonsko diferenciranje, za številne časovne vrste je perioda znana iz same narave podatkov. Z diferenciranjem lahko določimo parametra d in D v zapisu sezonskega procesa ARIMA, danega z definicijo 3.34, tj. diferencirana opažanja

$$Y_t = (1 - B)^d(1 - B^S)^D X_t \quad (3.101)$$

izgledajo stacionarno. Pri tem je S perioda, X_t pa opažena časovna vrsta. Pri izgradnji modela sezonskosti navadno ne odstranimo eksplicitno preden ocenimo parametre, namesto tega v specifikacijo modela vključimo red sezonskega diferenciranja (večina programov vsebuje to možnost).

Diferencirana časovna vrsta sedaj izgleda stacionarno, naslednji korak vključuje natančno preučitev vzorčne ACF in PACF diferenciranih podatkov. Tako bomo identificirali sezonske (parametra P in Q) in nesezonske (parametra p in q) komponente modela v skladu z definicijo sezonskega modela SARIMA 3.34.

Najprej se bomo osredotočili na sezonskost, tj. na parametra P in Q . V enem izmed predhodnih korakov smo že ugotovili periodo S . V naslednjem koraku preučimo vzorčno ACF in PACF diferencirane časovne vrste $\{Y_t\}$ pri odlogih, ki so večkratniki periode S . Če je $\hat{\rho}(\cdot)$ vzorčna ACF časovne vrste $\{Y_t\}$, potem parametra P in Q izberemo tako, da se $\hat{\rho}(ks)$, $k = 1, 2, \dots$ ujemajo z ACF procesa ARMA(P, Q). Red p in q izberemo tako, da se $\hat{\rho}(1), \dots, \hat{\rho}(s-1)$ ujemajo z ACF procesa ARMA(p, q).

V primerih iz prakse grafi vzorčnih ACF in PACF niso preprosti, zato je več možnosti za izbiro parametrov. Na prvi pogled se zdi, da bomo z višjim redom parametrov dosegli boljše prilaganje modela danim podatkom. Paziti pa moramo na nevarnost pretiranega prilaganja (ang. overfitting), torej da bi model preveč prilagodili točno določenim podatkom, ki smo jih opazili. Pri končni odločitvi za izbiro modela si zato lahko pomagamo z informacijskimi kriteriji. Mi bomo uporabili kriterij AIC (ang. Akaike information criterion), ki ga bomo predstavili v nadaljevanju. Značilnost kriterija AIC je, da kaznuje preveliko število parametrov in tako preprečuje pretirano prilaganje.

Omenili smo, da parametre izberemo tako, da primerjamo graf vzorčne ACF in PACF s teoretičnih obnašanjem teh grafov v primeru, ko je red znan. V poglavju o procesih ARMA smo že omenili, kaj je značilno za ACF in PACF procesa AR(p) in procesa MA(q), izbira čistega AR ali MA modela je torej preprosta. Velikokrat v praksi pa je bolj primeren mešan model ARMA. Za procese ARMA(p, q), kjer sta tako p kot q neničelna, je vzorčno ACF in PACF mnogo težje interpretirati. Tako ACF kot PACF procesa ARMA(p, q) sta omejeni z geometrijsko padajočo funkcijo, vendar pa je red težko določiti. Pomagamo si s kriterijem AIC, vendar pa je iskanje optimalnega modela (takega, ki ima najmanjšo vrednost kriterija AIC) dolgotrajno, če nimamo približne ideje o tem, katere vrednosti bi parametra p in q lahko zavzela.

Če ugotovimo, da tako vzorčna ACF kot vzorčna PACF ne izvenita po nekaj odlogih, je za naše podatke primeren model ARMA. Na grafu vzorčne ACF lahko vidimo, do katerega odloga so vzorčne avtokorelacije izstopajoče (zunaj mej), ta

odlog vzamemo kot maksimalno vrednost parametra q , ki ga bomo preizkusili. Na enak način iz grafa vzorčne PACF preberemo maksimalno vrednost parametra p , ki ga bomo preizkusili (torej po katerem odlogu so vzorčne parcialne avtokorelacije znotraj mej). Tako dobimo vrednosti p in q , ki bi morali zajeti pomembne značilnosti podatkov. Teoretično bi nato preizkusili vse možne kombinacije parametrov $\leq p$ in $\leq q$ in izračunali pripadajoči kriterij AIC. Če je red p in q velik, je težko preiskati prav vse kombinacije, ki bi prinesle najmanjšo vrednost kriterija AIC, zato si še dodatno pomagamo z grafom vzorčne ACF in PACF. Nekatere vmesne kombinacije lahko zato sistematično izpustimo, če se nam ne zdijo optimalne. Če na grafu vzorčne ACF npr. opazimo pomembno avtokorelacijo pri odlogih do 5 in odlogu 20, med 5 in 20 pa ne, lahko preizkusimo model, kjer parametre $\theta_6, \dots, \theta_{19}$ predhodno nastavimo na 0 (Podobno lahko iz grafa vzorčne PACF nekatere parametre ϕ_i nastavimo na 0).

Ko identificiramo model oz. eno izmed verjetnih možnosti, je potrebno parametre oceniti. Metod je več, ena možnost je metoda najmanjših kvadratov, najpogostejša pa je metoda največjega verjetja, ki jo bomo uporabili. V večini statističnih paketov je ta metoda že vgrajena in vrne oceno po metodi največjega verjetja za parametre $\phi = (\phi_1, \dots, \phi_p)$, $\theta = (\theta_1, \dots, \theta_q)$ in σ^2 , poleg tega pa še standarne napake ocen in vrednost logaritma verjetja za izbrane parametre. Več o ocenjevanju parametrov je na voljo v [7].

3.7.3 Kriterij AIC

Ko smo podatke transformirali (diferenciranje, odstranitev trenda in sezonskih komponent) do točke, kjer lahko transformirano časovno vrsto $\{X_t\}$ potencialno modeliramo z modelom ARMA, se moramo odločiti še o ustreznih vrednostih parametrov p in q . Način, kako izbiramo možne vrednosti parametrov, smo opisali v prejšnjem razdelku. Vendar pa v primeru mešanega modela, torej ko sta tako p kot q večja od 0, vzorčno ACF in PACF mnogo težje prepoznamo in ima torej manjšo težo kot v posebnih primerih, ko je $p = 0$ ali pa $q = 0$. Predstavili bomo način, kako se iskanja parametrov p in q lotimo sistematično.

S stališča napovedovanja ni ugodno, če parametra p in q izberemo poljubno velika. Če izberemo model zelo visokega reda, bomo v splošnem dobili majhno ocenjeno vrednost belega šuma. Ko pa je tak model uporabljen za napovedovanje, je srednja kvadratna napaka napovedi odvisna ne le od variance belega šuma temveč tudi od napak, ki nastanejo pri ocenjevanju parametrov modela. Te pa bodo večje pri modelih višjega reda. Iz tega razloga moramo uvesti kazenski faktor, ki nas odvrača od izbire modelov s preveč parametri. V obstoječi literaturi lahko najdemo mnogo takšnih meril, ki temeljijo na takšnih kazenskih faktorjih. Mi se bomo osredotočili na kriterij AIC.

Akaikeov informacijski kriterij (ang. Akaike information criterion, AIC) je predstavil japonski statistik Hirotugu Akaike leta 1973 in je najbolj splošno uporabljeno merilo za izbiro modela. Parametra p in q izberemo na način, da minimiziramo vrednost statistike AIC, ki je dana z enačbo

$$\text{AIC} = -2 \cdot \log L + 2 \cdot \text{numParam} \quad (3.102)$$

kjer je $\log L$ logaritem verjetja (dobimo ga pri ocenjevanju parametrov modela), numParam pa je število ocenjenih parametrov. Za fiksna p in q je iz definicije očitno, da je statistika AIC minimizirana pri ocenah za parametra ϕ in θ po metodi največjega verjetja. Kriterij AIC kaznuje veliko število ocenjenih parametrov in na ta način prepreči pretirano prileganje.

Ko modeliramo resnične podatke, redko obstaja pravi red. Za proces $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ je morda mnogo polinomov $\theta(z)$ in $\phi(z)$ takih, da koeficienti pred z_j v $\theta(z)/\phi(z)$ dobro aproksimirajo ψ_j za razmeroma majhne vrednosti j . Zato lahko obstaja več procesov ARMA, ki imajo lastnosti, podobne časovni vrsti $\{X_t\}$. Vseeno pa nam kriterij AIC predstavlja smiselni kriterij za izbiro primerljivih modelov. Smiselno si je natančneje ogledati nekaj modelov z najmanjšo vrednostjo kriterija AIC in končno izbiro narediti na podlagi pregleda residualov.

3.7.4 Diagnostika in Ljung-Boxov test

Omenili smo, da je zadnji korak pri dokončni izbiri modela še njegova validacija. V ta namen se bomo osredotočili na residue modela, ki jih bomo označili z \hat{W}_t . Če je ocenjen model res pravilen, morajo lastnosti residualov $\{\hat{W}_t\}$ odražati lastnosti belega šuma $\{Z_t\}$, ki generira proces ARMA. V primeru, ko je $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, morajo biti ti nekorelirani. Skalirane residue $\{\hat{R}_t\}$, $t = 1, \dots, n$ dobimo tako, da residue $W_t, t = 1, \dots, n$ delimo z oceno standardnega odklona belega šuma $\hat{\sigma} = \sqrt{(\sum_{i=1}^n \hat{W}_i^2)/n}$, torej

$$\hat{R}_t = \hat{W}_t / \hat{\sigma}. \quad (3.103)$$

Če je ocenjen model primeren, morajo imeti skalirani residuali lastnosti podobne zaporedju $\text{WN}(0,1)$. Če te predpostavke niso izpolnjene, je treba najti ustreznejši model. To pomeni, da se vrnemo na korak identifikacije modela in poskusimo razviti boljši model. Pri tem nam analiza residualov lahko pomaga pri identifikaciji ustrežnejšega modela.

Preverili bi radi, če skalirani residuali ustrezajo belemu šumu. V prvem koraku narišemo graf skaliranih residualov, spominjati mora na beli šum z varianco 1. Iz grafa je sicer težko zaznati strukturo korelacij procesa $\{\hat{R}_t\}$, opazimo pa lahko odklone povprečja od ničle in nekonstantno varianco, katere magnituda je odvisna od t .

V naslednjem koraku si ogledamo vzorčno ACF residualov $\{\hat{W}_t\}$ (ali ekvivalentno $\{\hat{R}_t\}$). Za velike n je vzorčna ACF n. e. p. zaporedja Y_1, \dots, Y_n s končno varianco porazdeljena približno n. e. p. $N(0, 1/n)$ [7]. Če je torej y_1, \dots, y_n realizacija takšnega n. e. p. zaporedja, mora približno 95% vzorčnih avtokorelacij pasti med meji $\pm 1.96/\sqrt{n}$. Če izračunamo npr. 40 vzorčnih avtokorelacij in vidimo, da več kot dve ali tri padejo izven meja ali pa ena zelo daleč stran od meje, zavrnemo hipotezo n. e. p. šuma.

Druga možnost je, da uporabimo t. i. portmanteau test, ki preverja ničelno hipotezo, da časovna vrsta residualov ne kaže avtokorelacije za določeno število odlogov h , in sicer proti alternativni, da je nek koeficient $\hat{\rho}(j), j = 1, \dots, h$ neničln. Testna

statistika je enaka

$$Q = n \sum_{j=1}^h \hat{\rho}^2(j), \quad (3.104)$$

kjer je n velikost vzorca, h je število avtokorelacijskih odlogov, $\hat{\rho}(j)$ je vzorčna korelacija pri odlogu j . Če je Y_1, \dots, Y_n n. e. p. zaporedje s končno varianco, potem je Q porazdeljena kot vsota kvadratov neodvisnih $N(0,1)$ slučajnih spremenljivk, $\sqrt{n}\hat{\rho}(j), j = 1, \dots, h$. Pod ničelno hipotezo je asimptotična porazdelitev Q hi-kvadrat s h prostostnimi stopnjami. Velika vrednost Q kaže, da so vzorčne avtokorelacije na podatkih prevelike, da bi bili podatki vzorec iz nekoreliranega zaporedja. Zavrnilo ničelno hipotezo pri stopnji značilnosti α če je $Q > \chi^2_{1-\alpha}(h)$, kjer je $\chi^2_{1-\alpha}(h)$ $(1 - \alpha)$ -kvantil hi-kvadrat porazdelitve s h prostostnimi stopnjami. Mi bomo uporabili modifikacijo zgornjega testa, ki sta ga formulirala Ljung in Box, kjer je Q zamenjana s statistiko

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j}, \quad (3.105)$$

katere porazdelitev je bolj aproksimirana s porazdelitvijo hi-kvadrat s h prostostnimi stopnjami. Testu pravimo Ljung-Boxov test in testira korelacijo residualov. Specificiramo lahko število odlogov h , pri katerih računamo testno statistiko, to pa vpliva na moč testa:

- Če je h premajhen, test ne zazna avtokorelacij visokega reda.
- Če je h prevelik, potem test izgubi moč, ko je pomembna korelacija na enem odlogu zanemarljiva zaradi nepomembnih korelacij na drugih odlogih.

Z Ljung-Boxovim testom tako lahko preverimo, če so residuali korelirani in bi bilo potrebno najti ustreznejši model. Omenili smo že, da v primerih finančnih podatkov, varianca pogosto ni konstantna in bi jo bilo potrebno modelirati z enim izmed modelov GARCH. Ko identificiramo model v smislu parametrov AR in MA dela ter nato ocenimo parametre, moramo preveriti, če je v residualih zaslediti heteroskedastičnost. Ljung-Boxov test lahko uporabimo za ugotavljanje pogojne heteroskedastičnosti (učinki ARCH) s testiranjem kvadratov residualov. Če se izkaže, da je heteroskedastičnost prisotna, moramo varianco modelirati z modelom GARCH. Ogledamo si lahko tudi graf in vzorčno ACF kvadratov residualov. Pri izbiri parametrov modela GARCH si bomo zopet pomagali s kriterijem AIC. Tudi v tem primeru parametre modelov tipa GARCH navadno ocenimo z metodo največjega verjetja.

3.7.5 Napovedovanje

Ko potrdimo, da je model primeren za naše podatke, ga lahko uporabimo za napovedovanje. Nekaj o napovedovanju smo že povedali v poglavju 3.5.3, zdaj pa bomo predstavili inovacijski algoritem (ang. the innovations algorithm) za rekurzivno računanje prihodnjih vrednosti procesa ARMA (podrobna izpeljava in dokazi pa so na voljo v [7]).

Trditev 3.36 (Inovacijski algoritem.). Naj bo $\{X_t\}$ časovna vrsta z ničelnim povprečjem in $E(X_i X_j) = \kappa(i, j)$, kjer je matrika $[\kappa(i, j)]_{i,j=1}^n$ nesingularna za vsak $n = 1, 2, \dots$. Prediktorji za en korak naprej \hat{X}_{n+1} , $n \geq 0$ in njihove srednje kvadratne napake $\nu_n = E[(X_{n+1} - \hat{X}_{n+1})^2]$, $n \geq 1$ so dani z

$$\hat{X}_{n+1} = \begin{cases} 0, & n = 0, \\ \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq 1, \end{cases} \quad (3.106)$$

in

$$\begin{cases} \nu_0 & = \kappa(1, 1), \\ \theta_{n,n-k} & = \nu_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j \right), \quad k = 0, 1, \dots, n-1, \\ \nu_n & = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j. \end{cases} \quad (3.107)$$

Trivialno je rekurzivno rešiti enačbo (3.107) v zaporedju $\nu_0; \theta_{11}, \nu_1; \theta_{22}, \theta_{21}, \nu_2; \theta_{33}, \theta_{32}, \theta_{31}, \nu_3; \dots$. Dokaz je na voljo v [7].

Zgornji algoritem nam torej da koeficiente t. i. inovacij $(X_j - \hat{X}_j)$, $j = 1, \dots, n$, v primeru procesa ARMA pa ga lahko dodatno poenostavimo. Trditev 3.36 lahko apliciramo na primer napovedi procesa ARMA(p, q),

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2), \quad (3.108)$$

kjer sta kot običajno $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ in $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$. Videli bomo, da se izračuni močno poenostavijo, če trditve 3.36 ne uporabimo neposredno na časovni vrsti $\{X_t\}$, ampak na transformiranem procesu

$$\begin{cases} W_t = \sigma^{-1} X_t, & t = 1, \dots, m, \\ W_t = \sigma^{-1} \phi(B) X_t, & t > m, \end{cases} \quad (3.109)$$

kjer je

$$m = \max(p, q). \quad (3.110)$$

Zaradi lažjega zapisa definirajo $\theta_0 = 1$ in privzemimo, da velja $p \geq 1$ in $q \geq 1$. (v teh predpostavkah ni izgube splošnosti, saj so v nadaljni analizi lahko katerikoli koeficienti ϕ_i in θ_i enaki 0). Izračun avtokovariančne funkcije $\gamma_X(\cdot)$ časovne vrste $\{X_t\}$ smo že obravnavali v poglavju 3.5.2. Avtokovariance $\kappa(i, j) = E(W_i W_j)$ potem dobimo kot

$$\kappa(i, j) = \begin{cases} \sigma^{-2} \gamma_X(i - j), & 1 \leq i, j \leq m, \\ \sigma^{-2} \left[\gamma_X(i - j) - \sum_{r=1}^p \phi_r \gamma_X(r - |i - j|) \right], & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{r=0}^p \theta_r \theta_{r+|i-j|}, & \min(i, j) > m, \\ 0, & \text{sicer,} \end{cases} \quad (3.111)$$

kjer je standardno $\theta_j = 0$ za $j > q$. Apliciramo trditev 3.36 na proces $\{W_t\}$ in dobimo

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^n \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \leq n < m, \\ \hat{W}_{n+1} = \sum_{j=1}^q \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & n \geq m, \end{cases} \quad (3.112)$$

kjer lahko koeficiente θ_{nj} in srednje kvadratne napake $r_n = E[(W_{n+1} - \hat{W}_{n+1})^2]$ rekurzivno dobimo iz enačbe (3.107), κ pa je definirana z enačbo (3.111). Iz enačb (3.109) vidimo, da lahko vsak $X_n, n \geq 1$ zapišemo kot linearno kombinacijo $W_j, 1 \leq j \leq n$. Vsak $W_n, n \geq 1$ pa lahko zapišemo kot linearno kombinacijo $X_j, 1 \leq j \leq n$. To pomeni, da je najboljši linearni prediktor za katerokoli slučajno spremenljivko Y na osnovi $\{1, X_1, \dots, X_n\}$ enak kot najboljši linearni prediktor za Y na osnovi $\{1, W_1, \dots, W_n\}$. Označimo ga s $P_n Y$. Prediktorja za en korak naprej za W_{n+1} in X_{n+1} sta dana z enačbama $\hat{W}_{n+1} = P_n W_{n+1}$ in $\hat{X}_{n+1} = P_n x_{n+1}$. Če uporabimo linearnost prediktorja P_n in enačbo (3.109), vidimo, da velja

$$\begin{cases} \hat{W}_t = \sigma^{-1} \hat{X}_t, & t = 1, \dots, m, \\ \hat{W}_t = \sigma^{-1} [\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}], & t > m. \end{cases} \quad (3.113)$$

Supaj z enačbo (3.109) torej velja

$$X_t - \hat{X}_t = \sigma [W_t - \hat{W}_t], \quad \text{za vse } t \geq 1. \quad (3.114)$$

Če v enačbi (3.111) $(W_j - \hat{W}_j)$ zamenjamo s $\sigma^{-1}(X_j - \hat{X}_j)$ in potem to vstavimo v enačbo (3.112), dobimo

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < m, \\ \phi_1 X_n + \dots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m, \end{cases} \quad (3.115)$$

in

$$E[(X_{n+1} - \hat{X}_{n+1})^2] = \sigma^2 E[(W_{n+1} - \hat{W}_{n+1})^2] = \sigma^2 r_n, \quad (3.116)$$

kjer θ_{nj} in r_n dobimo iz inovacijskega algoritma, kjer je κ kot v enačbi (3.111). Enačbe (3.115) rekurzivno določajo prediktorje za en korak naprej $\hat{X}_2, \hat{X}_3, \dots$

Oglejmo si primer napovedi procesa ARMA(1,1) za en korak naprej.

Primer 3.37. Želimo napovedati proces ARMA(1,1). Če je

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad (3.117)$$

in $|\phi| < 1$, potem se enačbe (3.115) zreducirajo na eno samo enačbo

$$\hat{X}_{n+1} = \phi X_n + \theta_{n1}(X_n - \hat{X}_n), \quad n \geq 1. \quad (3.118)$$

Da izračunamo θ_{n1} , uporabimo primer 3.27, kjer smo določili $\gamma_X(0) = \sigma^2(1 + 2\theta\phi + \theta^2)/(1 - \phi^2)$. Vstavimo v enačbo (3.111) in za $i, j \geq 1$ dobimo

$$\kappa(i, j) = \begin{cases} (1 + 2\theta\phi + \theta^2)/(1 - \phi^2), & i = j = 1, \\ 1 + \theta^2, & i = j \geq 2, \\ \theta, & |i - j| = 1, i \geq 1, \\ 0, & \text{sicer.} \end{cases} \quad (3.119)$$

S temi vrednostmi za $\kappa(i, j)$ se rekurzije inovacijskega algoritma reducirajo v

$$\begin{cases} r_0 &= (1 + 2\theta\phi + \theta^2)/(1 - \phi^2), \\ \theta_{n1} &= \theta/r_{n-1}, \\ r_n &= 1 + \theta^2 - \theta^2/r_{n-1}, \end{cases} \quad (3.120)$$

ki jih lahko preprosto rešimo.

V poglavju 3.5.3 smo obravnavali problem iskanja najboljšega linearnega prediktorja za proces ARMA, v tem poglavju pa smo predstavili inovacijski algoritem za rekurzivno računanje prihodnjih vrednosti procesa ARMA. Tudi v primeru procesov GARCH nas zanima napovedovanje prihodnjih vrednosti. Oglejmo si primer napovedi v primeru procesa GARCH(1,1), ki je dan z enačbo

$$Z_t = \sqrt{h_t}e_t, \quad \{e_t\} \sim \text{NEP } N(0,1), \quad (3.121)$$

kjer je

$$h_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 h_{t-1}, \quad \alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0. \quad (3.122)$$

Napovedali bi radi prihodnje vrednosti pogojne variance h_t . Zaradi preprostejšega zapisa privzemimo, da so parametri modela znani (ocene sicer navadno dobimo z metodo največjega verjetja). Napoved z najmanjšo srednjo kvadratno napako za h_{t+k} na osnovi informacij do časa t je pogojno matematično upanje $E[h_{t+k}|F_t]$, kjer je F_t množica vseh informacij, ki so na voljo ob času t . Za $k = 1$ tako velja

$$\begin{aligned} E[h_{t+1}|F_t] &= \alpha_0 + \alpha_1 E[Z_t^2|F_t] + \beta_1 E[h_t|F_t] \\ &= \alpha_0 + \alpha_1 Z_t^2 + \beta_1 h_t, \end{aligned} \quad (3.123)$$

kjer privzamemo, da sta Z_t^2 in h_t znana ob času t . Podobno za $k = 2$ velja

$$E[h_{t+2}|F_t] = \alpha_0 + \alpha_1 E[Z_{t+1}^2|F_t] + \beta_1 E[h_{t+1}|F_t]. \quad (3.124)$$

Velja $E[Z_{t+1}^2|F_t] = E[h_{t+1}e_{t+1}^2|F_t] = E[h_{t+1}|F_t]$, zgornja enačba tako postane

$$\begin{aligned} E[h_{t+2}|F_t] &= \alpha_0 + (\alpha_1 + \beta_1)E[h_{t+1}|F_t] \\ &= \alpha_0 + (\alpha_1 + \beta_1)(\alpha_0 + \alpha_1 Z_t^2 + \beta_1 h_t). \end{aligned} \quad (3.125)$$

V splošnem za $k \geq 2$ velja

$$E[h_{t+k}|F_t] = \alpha_0 + (\alpha_1 + \beta_1)E[h_{t+k-1}|F_t], \quad (3.126)$$

z rekurzijo pa dobimo zvezo

$$E[h_{t+k}|F_t] = \alpha_0 \sum_{i=0}^{k-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{k-1}(\alpha_1 Z_t^2 + \beta_1 h_t). \quad (3.127)$$

Spomnimo se: časovna vrsta Z_t je stacionarna, če velja $\alpha_1 + \beta_1 < 1$. V tem primeru v limiti velja

$$\lim_{k \rightarrow \infty} \sum_{i=0}^{k-1} (\alpha_1 + \beta_1)^i = \frac{1}{1 - (\alpha_1 + \beta_1)} \quad \text{in} \quad \lim_{k \rightarrow \infty} (\alpha_1 + \beta_1)^{k-1} = 0. \quad (3.128)$$

Sledi

$$\lim_{k \rightarrow \infty} E[h_{t+k}|F_t] = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} = \text{Var}(Z_t), \quad (3.129)$$

$E[h_{t+k}|F_t]$ torej konvergira k brezpogojni varianci procesa Z_t .

4 Metoda podpornih vektorjev in jedrne funkcije

Metoda podpornih vektorjev sodi med metode nadzorovanega strojnega učenja in je med njimi ena najbolj uporabljenih. Prvotno je bila razvita za naloge razvrščanja, še posebej binarne klasifikacije, apliciramo pa jo lahko tudi na primer regresije. Za lažje razumevanje si bomo tako najprej ogledali ozadje in delovanje metode na primeru binarne klasifikacije, nato pa še za primer regresije. Slednjo bomo uporabili pri napovedovanju cene in porabe električne energije. Na koncu poglavja bomo spoznali še jedrne funkcije, ki omogočajo nelinearno razvrščanje. Poglavje je v večini povzeto po [17] in [18].

4.1 Linearni model za binarno klasifikacijo.

Kot smo že omenili, je bila metoda podpornih vektorjev najprej razvita za primer binarne klasifikacije oz. razvrščanja v dve skupini. Oglejmo si motivacijo in izpeljavo metode.

Temeljni problem je naslednji: predpostavimo, da imamo N primerov, ki so predstavljeni z vektorji iz \mathbb{R}^d . Imamo dva razreda, ki ju označimo kot pozitivnega in negativnega. Vsak primer spada v enega od teh dveh razredov. Učno množico predstavljajo pari (x_i, y_i) za $i = 1, \dots, N$, $x_i \in \mathbb{R}^d$ je vektor, $y_i \in \{-1, 1\}$ je njegova oznaka razreda. Pozitivni in negativni razred smo zaradi matematične preprostosti označili z 1 in -1 . Nato se soočimo z novim primerom, ki ga moramo dodeliti enemu izmed dveh razredov. V ta namen bi radi dobili klasifikator (model), ki bi razločeval ta dva razreda. Spremenljivke x_i so v tem primeru napovedne (vhodne), z njimi želimo napovedati vrednost ciljne (izhodne) spremenljivke y_i (v našem primeru je to razred). V osnovnem modelu predpostavimo še, da sta razreda linearno ločljiva.

Pozitivni in negativni razred bi želeli razmejiti s hiperravnino oblike

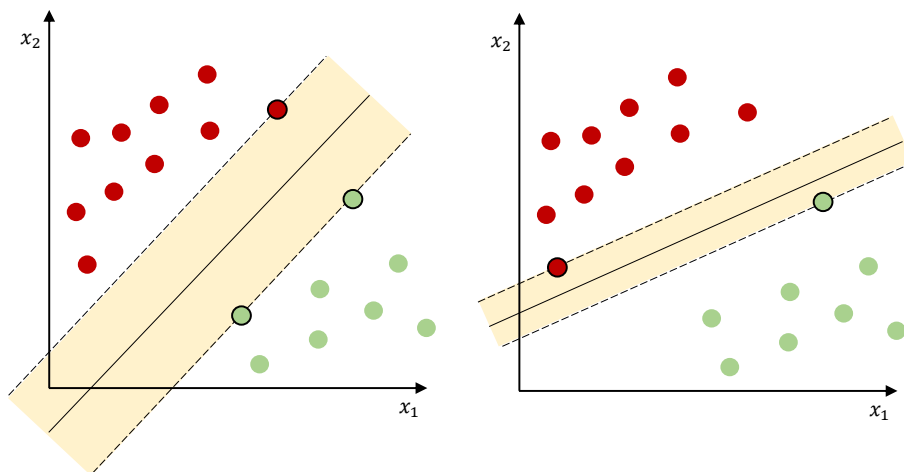
$$\beta^T x + \beta_0 = 0, \quad (4.1)$$

kjer je $\beta \in \mathbb{R}^d$ in $\beta_0 \in \mathbb{R}$. Ravnina razdeli prostor na dva polprostora ($\beta^T x + \beta_0 \geq 0$ in $\beta^T x + \beta_0 < 0$) in za poljubno točko je preprosto ugotoviti, v katerem polprostoru leži. Ravnino bi radi postavili tako, da bi pozitivni primeri ležali na eni strani ravnine, negativni pa na drugi. Odločitveno pravilo je tako enako

$$\hat{y} = f(x) = g(\beta^T x + \beta_0), \quad g(z) = \begin{cases} 1; & z \geq 0 \\ -1; & z < 0 \end{cases}. \quad (4.2)$$

Funkcija g pozitivnim primerom priredi vrednost 1, negativnim pa -1 . Odločitveno mejo, ki loči negativne primere od pozitivnih lahko postavimo na različne načine. Oglejmo si primer v \mathbb{R}^2 , kjer lahko posamezne primere tudi narišemo. Slika 6 prikazuje dva različna načina, kako lahko postavimo ločevalno hiperravnino. V dvodimenzionalnem prostoru je iskana hiperravnina kar premica.

Vprašanje je, kateri model je najboljši, torej kako postaviti hiperravnino, ki bo najboljše ločevala med primeri. V ta namen definiramo rob (ang. margin) kot prostor med ravnini najbližjima primeroma na eni in na drugi strani. Odgovor metode podpornih vektorjev je, da je najboljši model tisti z največjim robom (ang. large margin principle). Na sliki 6 je rob označen s prekinjeno črto, želimo pa si, da bi bil čim širši.



Slika 6: Ilustracija načela širokega roba. Levo: ločevalna hiperravnina z velikim robom. Desno: ločevalna hiperravnina z majhnim robom.

Kot smo že omenili, lahko na sliki 6 vidimo dva zgleda hiperravnin, ki ločujeta pozitivne od negativnih primerov. Obe popolnoma ločita oba razreda, vendar pa je rob v levem primeru bistveno širši kot rob v desnem primeru.

Odločitveno mejo med pozitivnimi in negativnimi primeri iščemo na način, da bo širina roba maksimalna. V ta namen potrebujemo enačbo za širino roba, tako bomo z njo predstavili optimizacijski problem.

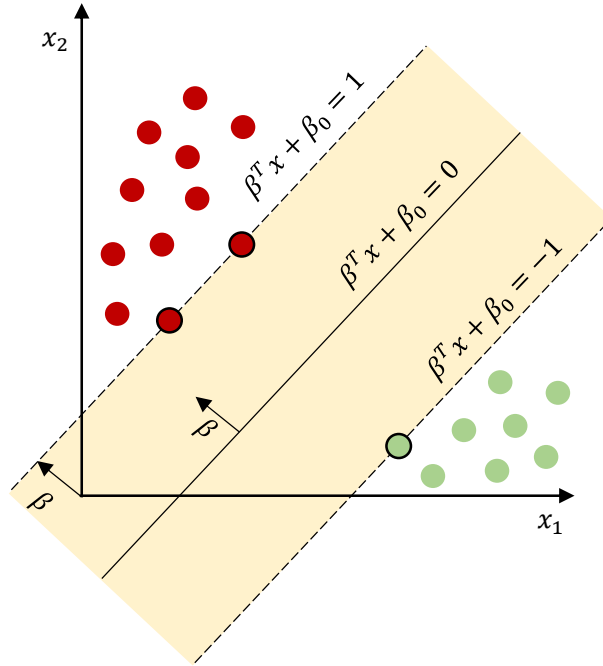
Naj bo rob omejen z ravninama $\beta^T x + \beta_0 = \gamma$ in $\beta^T x + \beta_0 = -\gamma$. Odločitveno pravilo je enako:

- Za pozitivni primer x_{\oplus} : $\beta^T x_{\oplus} + \beta_0 \geq \gamma$
- Za negativni primer x_{\ominus} : $\beta^T x_{\ominus} + \beta_0 \leq -\gamma$

Pozitivni primeri ležijo na pozitivni strani, negativni pa na negativni. Vmes med $-\gamma$ in γ pa je rob. Za pozitivne primere je $y = 1$, za negativne pa $y = -1$, zato lahko zgornji enačbi združimo v

$$(\beta^T x + \beta_0)y - \gamma \geq 0. \quad (4.3)$$

Enačaj velja točno na robu. Razmislimo lahko, da je parameter γ odveč. Če ravnina $\beta^T x + \beta_0$ ustreza pogoju $(\beta^T x + \beta_0)y - \gamma \geq 0$, lahko isto ravnino opišemo kot $\hat{\beta}^T x + \hat{\beta}_0$ za $\hat{\beta} = \beta/\gamma$ in $\hat{\beta}_0 = \beta_0/\gamma$ in bo ustrezala pogoju $(\hat{\beta}^T x + \hat{\beta}_0)y - 1 \geq 0$. Brez škode za splošnost lahko zato vzamemo $\gamma = 1$. Slika 7 prikazuje premico, ki ločuje pozitivne in negativne primere, s prekinjeno črto pa sta označeni premici, ki omejujeta rob.



Slika 7: Ločevalna hiperravnina, ki loči negativne od pozitivnih primerov in ima najširši rob.

Zdaj moramo izračunati širino roba, v pomoč nam bo slika 8. Ta je enaka sliki 7, z izjemo tega, da so prikazani samo primeri točno na robu, ki so odebljeni s črno. Izberemo si poljuben pozitiven in poljuben negativen primer, ki ležita točno na robu. Pripadajoča smerna vektorja naj bosta x_{\oplus} in x_{\ominus} . Oglejmo si razliko $x_{\oplus} - x_{\ominus}$ teh dveh vektorjev. Širino roba dobimo kot dolžino projekcije razlike na vektor β in je dana z enačbo

$$\frac{\beta^T (x_{\oplus} - x_{\ominus})}{\|\beta\|}. \quad (4.4)$$

Ker pozitivni in negativni primer ležita točno na robu, velja

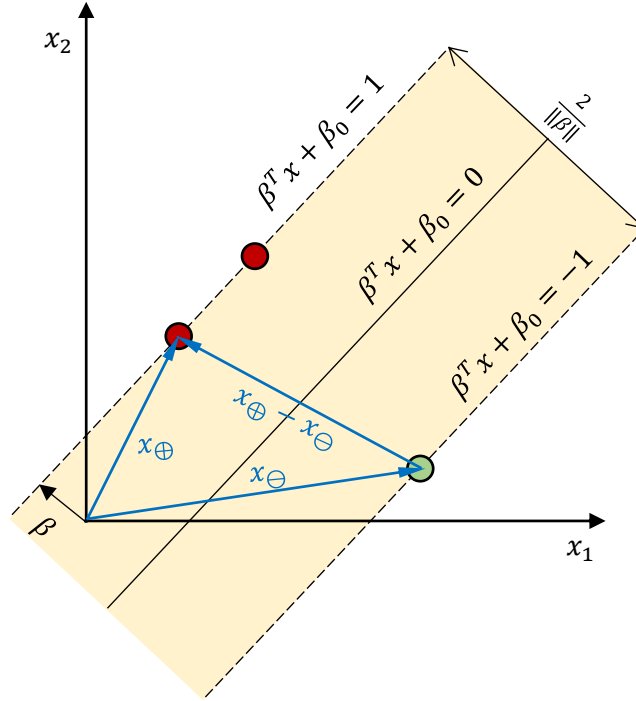
$$\beta^T x_{\oplus} + \beta_0 = 1 \quad \text{in} \quad \beta^T x_{\ominus} + \beta_0 = -1. \quad (4.5)$$

Če enačbi odštejemo, dobimo

$$\beta^T (x_{\oplus} - x_{\ominus}) = 2. \quad (4.6)$$

Dobljeno enakost vstavimo v enačbo (4.4) in dobimo širino roba.

$$\frac{\beta^T (x_{\oplus} - x_{\ominus})}{\|\beta\|} = \frac{2}{\|\beta\|}. \quad (4.7)$$



Slika 8: Izračun širine roba.

Ko poznamo širino roba, bi jo radi maksimizirali. Iskanje ločevalne ravnine lahko torej zastavimo kot optimizacijski problem. Osnovni optimizacijski problem je enak

$$\max_{\beta, \beta_0} \frac{2}{\|\beta\|} \quad (4.8)$$

pri pogoju $y_i(\beta^T x_i + \beta_0) \geq 1, \quad i = 1, \dots, N.$

Imamo N linearnih omejitev, ki nam povedo, da morajo učni primeri ležati izven roba. Vendar pa zgornja funkcija ni konveksna, kot bi si želeli. V ta namen upoštevamo $\max 1/\|\beta\| = 1/\min \|\beta\|$. Upoštevamo še, da je ekvivalentno $\min \|\beta\|$ in $\min \frac{1}{2}\|\beta\|^2$. Faktor $\frac{1}{2}$ tu dodamo zaradi lepšega nadaljevanja izpeljave. Zdaj imamo kvadratno funkcijo, ki je konveksna in lepa za optimizacijo.

Kot smo že omenili, so na sliki 8 prikazani samo vektorji, ki ležijo točno na robu, ostale pa smo izpustili. Ugotovili smo, da jih pri določanju širine roba sploh ne potrebujemo. Na to, kolikšna je širina roba in kako bomo postavili premico, vplivajo samo vektorji točno na robu, imenujemo pa jih podporni vektorji. Končna formulacija za binarno klasifikacijo je sledeča:

$$\min_{\beta, \beta_0} \frac{1}{2} \|\beta\|^2 \quad (4.9)$$

pri pogoju $1 - y_i(\beta^T x_i + \beta_0) \leq 0, \quad i = 1, \dots, N.$

Funkcijo v (4.9), ki jo minimiziramo, imenujemo ciljna funkcija (kriterijska, namen-ska). V našem primeru imamo konveksno kvadratično ciljno funkcijo ter linearne omejitve, kar predstavlja nalogo kvadratnega programiranja. Možno je najti globalni minimum, rešitev pa je par (β, β_0) , ki določa ravnino z najširšim robom.

Rešitev zgornjega optimizacijskega problema bomo dobili preko Lagrangeove dualnosti. Vpeljimo Lagrangeove multiplikatorje (koeficiente) $\alpha_i \geq 0, i = 1, \dots, N$. Vektor koeficientov označimo z $\alpha = (\alpha_1, \dots, \alpha_N)$. Lagrangeova funkcija za zgornji optimizacijski problem je enaka

$$\mathcal{L}(\beta, \beta_0, \alpha) = \frac{1}{2} \|\beta\|^2 + \sum_{i=1}^N \alpha_i (1 - y_i(\beta^T x_i + \beta_0)). \quad (4.10)$$

Vemo, da optimalna rešitev obstaja, zato zanjo veljajo Karush-Kuhn-Tuckerjevi (KKT) pogoji. Več o reševanju optimizacijskih problemov, dualnih programih in KKT pogojih je npr. dostopno v [17]. Optimalna rešitev (β, β_0) zadošča pogojem KKT, če obstajajo $\alpha_1, \alpha_2, \dots, \alpha_N \geq 0$, tako da velja:

$$\frac{\partial \mathcal{L}(\beta, \beta_0, \alpha)}{\partial \beta} = 0 \quad (4.11)$$

$$\frac{\partial \mathcal{L}(\beta, \beta_0, \alpha)}{\partial \beta_0} = 0 \quad (4.12)$$

$$\alpha_i (1 - y_i(\beta^T x_i + \beta_0)) = 0, \quad i = 1, \dots, N \quad (4.13)$$

$$1 - y_i(\beta^T x_i + \beta_0) \leq 0, \quad i = 1, \dots, N. \quad (4.14)$$

Želimo čim manjšo vrednost ciljne funkcije $\frac{1}{2} \|\beta\|^2$, hkrati pa morajo biti izpolnjeni vsi pogoji. Pogoj (4.13) je izpolnjen, če primer x_i leži točno na robu ali pa je ustrezeni Lagrangeov multiplikator enak 0. Najprej obravnavajmo prvi KKT pogoj, ki je dan z enačbama (4.11) in (4.12). Upoštevamo $\|\beta\|^2 = \beta^T \beta$ oziroma $\frac{\partial \|\beta\|^2}{\partial \beta} = 2\beta$. Iz pogoja v (4.11) dobimo enačbo

$$\frac{\partial \mathcal{L}(\beta, \beta_0, \alpha)}{\partial \beta} = \beta - \sum_{i=1}^N \alpha_i y_i x_i = 0,$$

od koder sledi

$$\beta = \sum_{i=1}^N \alpha_i y_i x_i. \quad (4.15)$$

Pogoj v (4.12) nam da enačbo

$$\frac{\partial \mathcal{L}(\beta, \beta_0, \alpha)}{\partial \beta_0} = - \sum_{i=1}^N \alpha_i y_i = 0,$$

od koder sledi

$$\sum_{i=1}^N \alpha_i y_i = 0. \quad (4.16)$$

Najti moramo še optimalne vrednosti parametrov α_i . Z $\langle u, v \rangle = u^T v = v^T u$ označimo skalarni produkt vektorjev u in v . Upoštevamo enačbo (4.15) in osnovno ciljno funkcijo prepisemo v

$$\frac{1}{2} \|\beta\|^2 = \frac{1}{2} \langle \beta, \beta \rangle = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle. \quad (4.17)$$

Zapišemo lahko še

$$\beta^T x = \langle \beta, x \rangle = \sum_{i=1}^N \alpha_i y_i \langle x, x_i \rangle. \quad (4.18)$$

Upoštevamo zgornje ugotovitve in Lagrangeovo funkcijo v (4.10) zapišemo kot

$$\begin{aligned} \mathcal{L}(\beta, \beta_0, \alpha) &= \frac{1}{2} \|\beta\|^2 + \sum_{i=1}^N \alpha_i (1 - y_i (\beta^T x_i + \beta_0)) \\ &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{i=1}^N \alpha_i \\ &\quad - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \beta_0 \sum_{i=1}^N \alpha_i y_i \\ &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle. \end{aligned} \quad (4.19)$$

V prehodu na zadnjo vrstico smo upoštevali $\sum_{i=1}^N \alpha_i y_i = 0$. V funkciji \mathcal{L} so ostale še dualne spremenljivke α_i , β in β_0 pa ne nastopata več. Tako smo dobili dualni optimizacijski problem, ki ga navadno rešujemo v praksi. Končna formulacija dualnega problema je naslednja:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{pri pogojih} \quad & \alpha_i \geq 0, i = 1, \dots, N \\ & \sum_{i=1}^N \alpha_i y_i = 0. \end{aligned} \quad (4.20)$$

Imamo $(N+1)$ linearnih omejitev. Omenimo še, da so vse operacije v prostoru vhodnih spremenljivk skalarni produkti. Rešitev zgornjega problema nam da optimalne α_i . Te vstavimo v enačbo (4.15) in dobimo optimalno β . Ko imamo določeno ločevalno hiperravnino, lahko za primer x_0 napovemo pripadajoč razred:

$$\begin{aligned} f(x_0) &= g(\langle \beta, x_0 \rangle + \beta_0) = g\left(\left\langle \sum_{i=1}^N \alpha_i y_i x_i, x_0 \right\rangle + \beta_0\right) \\ &= g\left(\sum_{i=1}^N \alpha_i y_i \langle x_i, x_0 \rangle + \beta_0\right) \end{aligned} \quad (4.21)$$

Vsoto v zgornji formuli računamo le za primere x_i , kjer je $\alpha_i > 0$. Zaradi drugega KKT pogoja (4.13) mora za te primere veljati $1 - y_i(\langle \beta, x_i \rangle + \beta_0) = 0$. Torej x_i ležijo točno na robu. Podporni vektorji so primeri x_i , ki ležijo točno na robu, zanje velja $\alpha_i > 0$ in samo ti vplivajo na dokončno razvrstitev primera v pozitivni ali negativni razred. Kompleksnost modela je tako določena s številom podpornih vektorjev.

Izračunati moramo še β_0 . Upoštevamo, da velja $y_i(\langle \beta, x_i \rangle + \beta_0) = 1$, kjer so x_i podporni vektorji, nato pa vzamemo povprečje po vseh β_0 .

4.2 Model za podatke, ki niso linearno ločljivi

V prejšnjem primeru smo predpostavili linearno ločljive podatke. Te smo lahko ločili s hiperravnino s trdim robom (ang. hard margin), kar pomeni, da ne dovolimo napak na učni množici. V praksi se seveda zgodi, da razreda nista linearno ločljiva. Če na učni množici še vedno ne dovolimo napak, se lahko zgodi, da prvotna naloga nima dopustne rešitve. Zato uporabimo hiperravnino z mehkim robom (ang. soft margin), kjer na učni množici dovolimo napake. V ta namen vpeljemo kazenske spremenljivke $\xi_i \geq 0$ za $i = 1, \dots, N$ in pogoj

$$y_i(\beta^T x_i + \beta_0) \geq 1$$

omilimo v

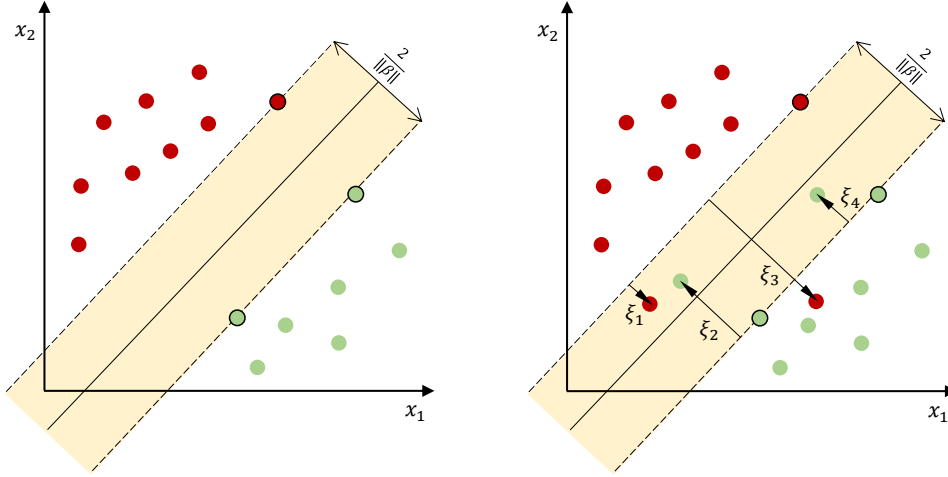
$$y_i(\beta^T x_i + \beta_0) \geq 1 - \xi_i. \quad (4.22)$$

Vrednost ξ_i upoštevamo v kriterijski funkciji. Alternativna formulacija optimizacijskega problema za podatke, ki niso linearno ločljivi je naslednja:

$$\begin{aligned} \min_{\beta, \beta_0} \quad & \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i \\ \text{pri pogojih} \quad & y_i(\beta^T x_i + \beta_0) \geq 1 - \xi_i, \quad i = 1, \dots, N \\ & \xi_i \geq 0, \quad i = 1, \dots, N. \end{aligned} \quad (4.23)$$

Imamo $2N$ linearnih omejitev. C je pozitivna konstanta, ki določa kompromis med maksimizacijo roba in minimizacijo odmikov v napačno smer. Večanje parametra C prenese fokus optimizacije na zmanjševanje odmikov ξ_i tj. na čim boljše ločevanje razredov. Manjšanje vrednosti parametra C dopušča velike vrednosti ξ_i , zato ločevanje razredov ni tako pomembno. Poveča se število podpornih vektorjev znotraj roba.

Kazenske spremenljivke ξ_i merijo odmik od roba v napačno smer. Za vsak primer, ki je na napačni strani roba ali pa znotraj roba, vpeljemo kazen. Če je primer pravilno razvrščen, je $\xi_i = 0$. Slika 9 na levi strani prikazuje linearno ločljive podatke, na desni pa take, ki niso linearno ločljivi. Za vsak primer, ki ni na svoji strani roba (torej znotraj roba ali pa na drugi strani) je označena napaka ξ_i . V tem primeru je le nekaj primerov takšnih, ki odstopajo, večina pa jih je pravilno raziščanih.



Slika 9: Podporni vektorji in odmiki ξ_i . Levo: linearno ločljivi podatki. Desno: podatki, ki niso linearno ločljivi.

Dualni problem je podoben kot v prejšnjem primeru:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{pri pogojih} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, N \\ & \sum_{i=1}^N \alpha_i y_i = 0. \end{aligned} \quad (4.24)$$

Imamo $2N + 1$ linearnih omejitev. Edina razlika v primerjavi z linearno ločljivimi podatki je, da so zdaj Lagrangeovi multiplikatorji α_i navzgor omejeni s C . Na ta način je vpliv posameznih primerov (ki bi lahko bili izstopajoči) omejen. Podporni vektorji x_s so v tem primeru lahko tudi znotraj roba. Na robu, ki je dan z enačbo $y_s(\beta^T x_s + \beta_0) = 1$, velja $0 < \alpha_s < C$. V notranjosti roba, ki je določena z neenakostjo $y_s(\beta^T x_s + \beta_0) < 1$, pa velja $\alpha_s = C$. Tudi v tem primeru ima rešitev enako obliko kot v enačbi (4.21). β_0 lahko izračunamo z uporabo dejstva, da za podporne vektorje x_i na robu, torej tiste z $\alpha_i < C$, velja $\xi_i = 0$ (to zopet sledi iz pogojev KKT). Velja

$$\sum_{j=1}^N \alpha_j y_j \langle x_i, x_j \rangle + \beta_0 = y_i. \quad (4.25)$$

Zdaj imamo več podpornih vektorjev, saj zraven spadajo tudi tisti, ki ležijo znotraj roba. Oglejmo si še enkrat sliko 9. Poleg vseh pozitivnih in negativnih primerov na

robu, podporni vektorji postanejo tudi vsi znotraj rumenega pravokotnika (desna slika).

4.3 Regresija

Metodo podpornih vektorjev lahko posplošimo tudi na primer regresije. Tokrat y nima le dveh vrednosti (1 in -1 oz. pozitivni in negativni razred), ampak je lahko katerokoli realno število. Konstruirali bomo algoritem, ki je analogen primeru z mehkim robom, le da je tokrat $y \in \mathbb{R}$. Napovedovali bi radi na naslednji način:

$$f(x) = \langle \beta, x \rangle + \beta_0, \quad (4.26)$$

kjer je $x \in \mathbb{R}^d$ in $\beta_0 \in \mathbb{R}$. To si lahko predstavljamo kot ravnino v $(d+1)$ -razsežnem prostoru. Ko namesto prave vrednosti y napovemo $f(x)$, naredimo napako $|y - f(x)|$ in želimo, da je ta čim manjša. Za merjenje napake izberemo funkcijo izgube V_ε za $\varepsilon > 0$ (ang. ε -insensitive loss function), ki je definirana kot

$$V_\varepsilon(r) = \begin{cases} 0 & ; |r| \leq \varepsilon \\ |r| - \varepsilon & ; |r| > \varepsilon \end{cases}. \quad (4.27)$$

Funkcija V_ε torej ne kaznuje napak manjših od ε . Utemeljitev te izbire je naslednja. V primeru klasifikacije smo imeli veliko območje, kjer je bila izguba enaka 0: kadar je bil primer na pravi strani odločitvene meje in se ni dotikal roba, ta ni prispeval k ciljni funkciji. Tak primer ne vsebuje nobenih informacij o položaju odločitvene meje, saj se ta izračuna tako, da minimiziramo ciljno funkcijo. To je tudi razlog, zakaj se primer ne pojavi v končni rešitvi. Sklepamo, da mora tudi funkcija izgube v primeru regresije imeti neobčutljivo območje, tj. območje, v katerem primeri ne prispevajo k funkciji izgube.

Zopet želimo, da so parametri hiperravnine čim manjši. Ciljna funkcija, ki jo želimo minimizirati, je enaka

$$\frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N V_\varepsilon(y_i - f(x_i)), \quad (4.28)$$

kjer je $f(x_i) = \langle \beta, x_i \rangle + \beta_0$, C pa je regularizacijska konstanta. Izbrati je torej potrebno, kako velik bo ε in pa konstanto C , ki uravnava, na katerem delu bo fokus optimizacije. Na to lahko gledamo kot na regularizirano linearno regresijo. Zgornja ciljna funkcija je konveksna, vendar pa zaradi absolutne vrednosti v funkciji V_ε ni odvedljiva. Problem razrešimo tako, da uvedemo dve kazenski spremenljivki za oba možna primera odstopanja izven ε pasu, tj. $f(x_i) - y_i > \varepsilon$ in $y_i - f(x_i) > \varepsilon$. Označimo ju z ξ_i in ξ_i^* , velja pa

$$f(x_i) - y_i \leq \varepsilon + \xi_i \quad \text{in} \quad y_i - f(x_i) \leq \varepsilon + \xi_i^*. \quad (4.29)$$

Optimizacijski problem je dan z

$$\begin{aligned} \min_{\beta, \beta_0, \xi, \xi^*} \quad & \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{pri pogojih} \quad & f(x_i) - y_i \leq \varepsilon + \xi_i, \quad i = 1, \dots, N \\ & y_i - f(x_i) \leq \varepsilon + \xi_i^*, \quad i = 1, \dots, N \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N. \end{aligned}$$

Napake, ki so manjše od ε , ne potrebujejo neničelnih ξ_i ali ξ_i^* in tako ne nastopajo v ciljni funkciji. Kot v primeru binarne klasifikacije bomo rešitev dobili preko Lagrangeove dualnosti. Lagrangeova funkcija za zgornji optimizacijski problem je enaka

$$\begin{aligned}\mathcal{L}(\beta, \beta_0, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) &= \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) - \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ &\quad + \sum_{i=1}^N \alpha_i (\langle \beta, x_i \rangle + \beta_0 - y_i - \varepsilon - \xi_i) \\ &\quad + \sum_{i=1}^N \alpha_i^* (y_i - \langle \beta, x_i \rangle - \beta_0 - \varepsilon - \xi_i^*),\end{aligned}\quad (4.30)$$

kjer so $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ Lagrangeovi multiplikatorji, ki morajo zadoščati

$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0. \quad (4.31)$$

Optimalna rešitev mora zadoščati naslednjim enačbam:

$$\frac{\partial \mathcal{L}}{\partial \beta} = \beta - \sum_{i=1}^N (\alpha_i^* - \alpha_i) x_i = 0 \quad (4.32)$$

$$\frac{\partial \mathcal{L}}{\partial \beta_0} = \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \quad (4.33)$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C - \alpha_i - \eta_i = 0, \quad i = 1, \dots, N \quad (4.34)$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0, \quad i = 1, \dots, N. \quad (4.35)$$

Pogoj v (4.32) nam da

$$\beta = \sum_{i=1}^N (\alpha_i^* - \alpha_i) x_i, \quad (4.36)$$

pogoja v (4.34) in (4.35) pa

$$\eta_i = C - \alpha_i \quad \text{in} \quad \eta_i^* = C - \alpha_i^*, \quad i = 1, \dots, N. \quad (4.37)$$

Enačbe (4.33), (4.36) in (4.37) vstavimo v enačbo (4.30) in dobimo

$$\mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*). \quad (4.38)$$

Tako dobimo naslednji dualni problem:

$$\begin{aligned}
\max_{\alpha, \alpha^*} \quad & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \langle x_i, x_j \rangle \\
& -\varepsilon \sum_{i=1}^N (\alpha_i^* + \alpha_i) + \sum_{i=1}^N y_i (\alpha_i^* - \alpha_i) \\
\text{pri pogojih} \quad & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\
& 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, N.
\end{aligned} \tag{4.39}$$

Rešitev nam da optimalne α_i in α_i^* in nato še β . Velja $\beta = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i$, β lahko torej zapišemo kot linearno kombinacijo učnih primerov. Napoved regresijskega modela v točki x_0 je enaka

$$f(x_0) = \langle \beta, x_0 \rangle + \beta_0 = \sum_{i=1}^N (\alpha_i^* - \alpha_i) \langle x_i, x_0 \rangle + \beta_0. \tag{4.40}$$

Kot v primeru klasifikacije, so vse operacije še vedno skalarni produkti. Pri napovedi $f(x)$ nam v resnici ni potrebno eksplicitno izračunati β . Izračunati moramo še konstanto β_0 , za kar bomo uporabili KKT pogoje - produkt med dualnimi spremenljivkami (Lagrangeovimi multiplikatorji) in omejitvami mora biti enak 0:

$$\alpha_i (\langle \beta, x_i \rangle + \beta_0 - y_i - \varepsilon - \xi_i) = 0, \tag{4.41}$$

$$\alpha_i^* (y_i - \langle \beta, x_i \rangle - \beta_0 - \varepsilon - \xi_i^*) = 0, \tag{4.42}$$

$$(C - \alpha_i) \xi_i = 0, \tag{4.43}$$

$$(C - \alpha_i^*) \xi_i^* = 0. \tag{4.44}$$

Zgornje nam omogoča več koristnih zaključkov:

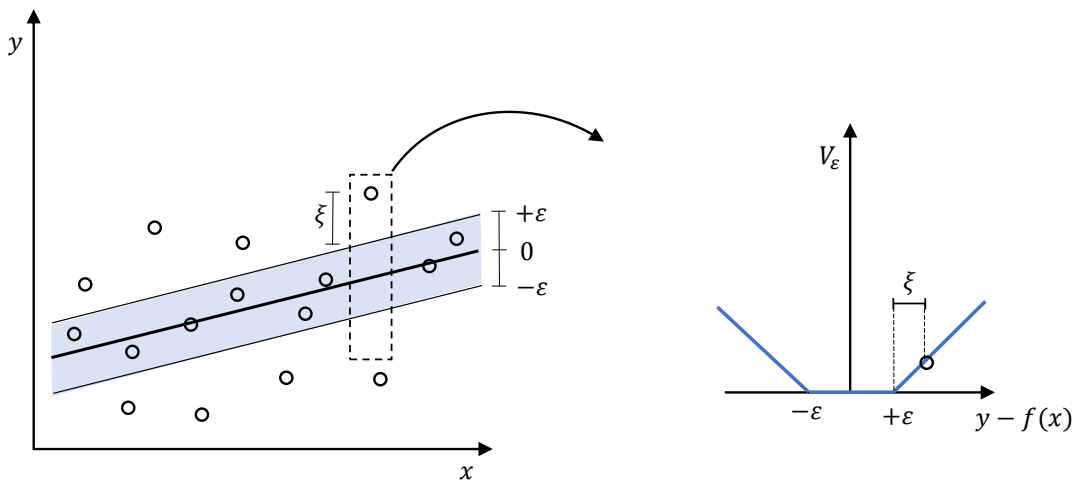
- Samo primeri (x_i, y_i) s pripadajočimi vrednostmi $\alpha_i = C$ (oz. $\alpha_i^* = C$) lahko ležijo zunaj ε pasu (tj. kjer velja $\xi_i > 0$ oz. $\xi_i^* > 0$).
- Velja $\alpha_i \alpha_i^* = 0$. Z drugimi besedami, ne obstaja par dualnih spremenljivk α_i, α_i^* , kjer sta obe hkrati enaki 0.
- Za $\alpha_i \in (0, C)$ (oz. $\alpha_i^* \in (0, C)$) velja $\xi_i = 0$ (oz. $\xi_i^* = 0$). Drugi faktor v (4.41) oz. (4.42) pa mora biti enak 0. β_0 lahko torej izračunamo kot:

$$\begin{aligned}
b &= y_i - \langle \beta, x_i \rangle - \varepsilon \quad \text{za} \quad \alpha_i \in (0, C), \\
b &= y_i - \langle \beta, x_i \rangle + \varepsilon \quad \text{za} \quad \alpha_i^* \in (0, C).
\end{aligned} \tag{4.45}$$

Teoretično zadostuje, če uporabimo katerikoli Lagrangeov multiplikator, ki je strogo med 0 in C . Če imamo možnost izbire med več takšnimi multiplikatorji (običajno obstaja več multiplikatorjev, ki niso točno na meji, torej niso enaki 0 ali pa C), je najvarneje izbrati takšnega, ki ni preblizu 0 ali C . Konstanto β_0 lahko izračunamo tudi na več drugačnih načinov, ki pa jih tu ne bomo obravnavali.

- Iz (4.41) in (4.42) sledi, da so lahko Lagrangeovi multiplikatorji različni od 0 le za $|f(x_i) - y_i| \geq \varepsilon$; z drugimi besedami, za vse primere znotraj ε -pasu velja $\alpha_i, \alpha_i^* = 0$. Kadar je $|f(x_i) - y_i| < \varepsilon$, je drugi faktor v (4.41) in (4.42) neničeln, zato morata biti α_i in α_i^* enaka 0, da bomo izpolnjeni KKT pogoji. Tiste x_i , pri katerih α_i in α_i^* nista oba enaka nič, imenujemo podporni vektorji; to so točke, ki ležijo zunaj ali na ε pasu. Geometrijsko se zdi smiselno, da primeri znotraj ε pasu ne prispevajo k rešitvi: kateregakoli od njih bi lahko odstranili in še vedno dobili isto rešitev, zato o njej ne morejo imeti nobenih informacij.

Slika 10 prikazuje ε pas na levi in graf funkcije izgube V_ε na desni.



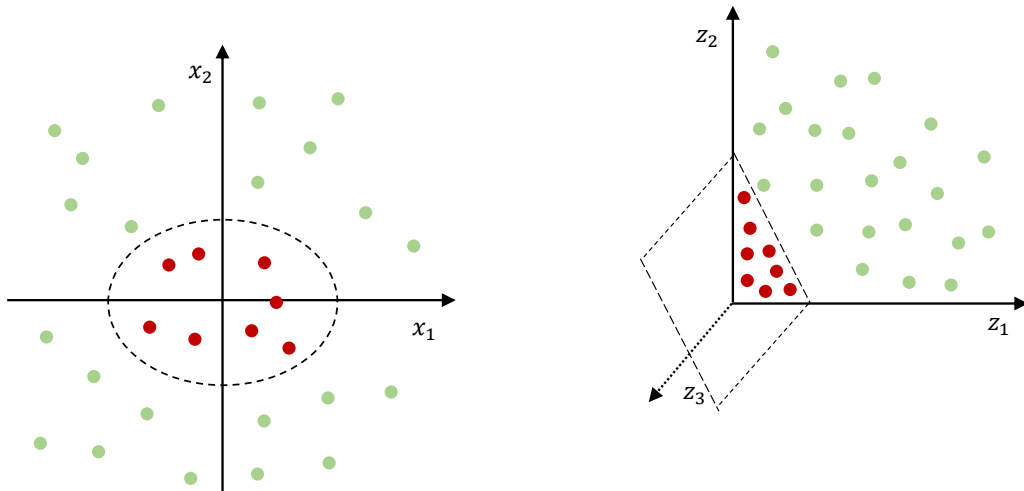
Slika 10: ε pas in funkcija izgube V_ε .

4.4 Jedrne funkcije

Kot smo že omenili, podatki večinoma niso linearno ločljivi. Lahko so skoraj linearno ločljivi in je le nekaj šumnih primerov (desni del slike 9), lahko pa je prisoten nek drug nelinearen vzorec. Ideja je, da podatke preslikamo v nek drug prostor, navadno višje dimenzije, kjer podatke lahko (bolje) ločimo s hiperravnino. Učne primere x_i bi torej preslikali s preslikavo $\phi : \mathbb{R}^d \rightarrow \mathcal{F}$, kjer je \mathcal{F} nek nov vektorski prostor, nato pa apliciramo standardno metodo podpornih vektorjev.

Za zgled si oglejmo preprost primer binarne klasifikacije v \mathbb{R}^2 , ki je prikazan na sliki 11. Imamo dva razreda, rdeče in zele kroge. Levi del slike prikazuje osnovni prostor, kjer točke niso linearno ločljive. Predpostavimo, da je resnična odločitvena meja elipsa. Če prostor napovednih spremenljivk preslikamo z nelinearno transformacijo $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, $\phi(x) = (z_1, z_2, z_3) = ([x]_1^2, [x]_2^2, \sqrt{2}[x]_1[x]_2)$ (desna slika), elipsa postane hiperravnina. $[x]_i$ razumemo kot i -to komponento vektorja $x \in \mathbb{R}^2$. V tem preprostem primeru je hiperravnina vzporedna z osjo z_3 , zato so vse točke narisane v ravnini (z_1, z_2) . Nelinearna odločitvena meja v tem primeru postane linearna, problem smo torej prevedli na iskanje hiperravnine, kar pa smo že obravnavali. V tem primeru smo iz dvodimenzionalnega prostora prešli v trodimenzionalni prostor, vendar pa je dimenzija novega prostora lahko občutno večja. Čeprav se ta pristop v

konkretnem primeru zdi smiselno, lahko izračun hitro postane računsko zahteven ali celo neizvedljiv. Želeli bi torej najti računsko optimalnejši način.



Slika 11: Primer binarne klasifikacije v \mathbb{R}^2 . Levo: osnovni prostor. Desno: prostor, preslikan s transformacijo $\phi(x) = (z_1, z_2, z_3) = ([x]_1^2, [x]_2^2, \sqrt{2}[x]_1[x]_2)$.

Če si ogledamo enačbo za napoved regresijskega modela (4.40), vidimo, da vhodni podatki (tj. učni primeri) nastopajo samo v skalarnem produktu (enako velja v primeru klasifikacije). Zato nam ni potrebno preslikati vseh elementov (kar je lahko računsko zahtevno), znati moramo le izračunati skalarni produkt v transformiranem prostoru. Tu zdaj pridejo v poštev jedrne funkcije, prav z njimi namreč izračunamo skalarne produkte v transformiranem prostoru in nam tako omogočajo nelinearno ločevanje.

Glavna ugotovitev je torej, da lahko z jedrno funkcijo skalarni produkt vektorjev v transformiranem prostoru izračunamo brez potrebe prehoda v transformirani prostor. Z drugimi besedami, metodo podpornih vektorjev lahko izvajamo v več-dimenzionalnih prostorih, ne da bi se nam bilo treba tja premakniti. Dovolj je, da znamo izračunati skalarni produkt dveh vektorjev (kar nam računsko ne vzame veliko časa). Temu pravimo jedrni trik. Oglejmo si definicijo jedrne funkcije.

Definicija 4.1. Jedrna funkcija $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ je definirana z

$$K(u, v) = \langle \phi(u), \phi(v) \rangle. \quad (4.46)$$

Krajše jo imenujemo tudi jedro, vrne pa vrednost skalarnega produkta transformiranih vektorjev.

Jedro funkcijo si lahko predstavljamo tudi kot mero podobnosti med primeri. Za primera u in v vrne realno število, ki ponazarja podobnost med njima. Iz definicije jedrne funkcije sledi, da je ta simetrična, torej $K(u, v) = K(v, u)$, kar se zdi smiselno, če želimo meriti podobnost. Preprost primer mere podobnosti je na primer skalarni produkt, torej $K(u, v) = \langle u, v \rangle$, kasneje bomo to funkcijo imenovali linearno jedro.

Za prejšnji primer velja:

$$\begin{aligned} K(u, v) &= \langle \phi(u), \phi(v) \rangle = \langle (u_1^2, u_2^2, \sqrt{2}u_1u_2), (v_1^2, v_2^2, \sqrt{2}v_1v_2) \rangle \\ &= u_1^2v_1^2 + u_2^2v_2^2 + 2u_1u_2v_1v_2. \end{aligned}$$

Oglejmo si še en primer jedrne funkcije.

Primer 4.2. Naj bo jedrna funkcija enaka $K(u, v) = (\langle u, v \rangle + 1)^2$. Razpišemo:

$$\begin{aligned} (\langle u, v \rangle + 1)^2 &= (u_1v_1 + u_2v_2 + 1)^2 \\ &= u_1^2v_1^2 + u_2^2v_2^2 + 1 + 2u_1u_2v_1v_2 + 2u_1v_1 + 2u_2v_2 \\ &= \langle \phi(u), \phi(v) \rangle. \end{aligned}$$

Od tu vidimo, da je preslikava ϕ enaka

$$\phi(u) = (u_1^2, u_2^2, 1, \sqrt{2}u_1u_2, \sqrt{2}u_1, \sqrt{2}u_2).$$

S pomočjo jedrnega trika lahko tako izračunamo model podpornih vektorjev v 6-dimenzionalnem prostoru, čeprav je osnovni prostor dvodimenzionalen. Kasneje bomo spoznali, da je to polinomsko jedro stopnje 2.

Vprašanje je, kdaj je funkcija K veljavno jedro, katerim lastnostim mora zadoščati. V ta namen definirajmo jedrno matriko.

Definicija 4.3. Jedrna matrika K za učno množico $\{x_1, x_2, \dots, x_N\}$ je matrika dimenzije $N \times N$, kjer je element na mestu (i, j) enak $K_{ij} = K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. Zaradi komutativnosti skalarne produkta je K simetrična.

Ko potrebujemo napoved za nov primer, ki ni iz učne množice, je tako potrebno izračunati le en dodatni stolpec te matrike (zaradi simetrije). Naslednji izrek nam pove, kdaj je funkcija lahko jedrna funkcija.

Izrek 4.4 (Mercer). *Potrebni in zadostni pogoj zato, da matrika K določa veljavno jedro, je to, da je K simetrična in pozitivno semidefinitna, tj. $\forall z \in \mathbb{R}^N : z^T K z \geq 0$. Funkcija $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ je potem jedrna funkcija.*

Oglejmo si nekaj pogosto uporabljenih jedrnih funkcij:

- Linearno jedro:

$$K(u, v) = \langle u, v \rangle. \quad (4.47)$$

- Polinomsko jedro, parameter je stopnja d :

$$K(u, v) = (1 + \langle u, v \rangle)^d. \quad (4.48)$$

- Radialno (RBF) jedro, parameter je $\gamma = 1/(2\sigma^2)$:

$$K(u, v) = e^{-\gamma \|u-v\|^2}. \quad (4.49)$$

Pri napovedovanju cen in porabe električne energije bomo poleg osnovnega linearne jedra uporabili še radialno jedro. V primerjavi s polinomskim je to računsko veliko manj zahtevno, je tudi najbolj splošno, če nismo prepričani, kakšen vzorec je prisoten v podatkih.

4.5 Merjenje napovedne napake in izbira modela

V prejšnjih poglavjih smo opisali idejo metode podpornih vektorjev ter kako dobimo napoved regresijskega modela v izbrani točki. V tem poglavju pa se bomo osredotočili na izbiro optimalnega modela (tj. optimalnih parametrov).

Naloga nadzorovanega strojnega učenja je na osnovi podane učne podatkovne množice najti model m , ki je točen in splošno veljaven. Točen model doseže minimalno napako (maksimalno točnost) na učni množici, splošno veljaven pa doseže majhne napake na poljubni podatkovni množici.

Vprašanje je, kako vrednotimo točnost modela. Napaka na enem primeru je dana s funkcijo izgube $L(y, \hat{y})$, ki vrne razliko med opazovano (y) in napovedano (\hat{y}) vrednostjo ciljne spremenljivke na enem primeru. Napaka na podatkovni množici vrne povprečno vrednost funkcije izgube podanega modela m na primerih iz podatkovne množice.

Običajna funkcija izgube v primeru regresije je kvadratna napaka (ang. squared error, SE), ki je definirana kot

$$L_{SE}(y, \hat{y}) = (y - \hat{y})^2. \quad (4.50)$$

S to funkcijo izgube vrednotimo srednjo kvadratno napako (ang. mean squared error, MSE) modela na podatkovni množici z n primeri, ki je definirana kot

$$MSE = \frac{1}{n} \sum_{i=1}^n L_{SE}(y_i, \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (4.51)$$

Pogosto računamo tudi koren srednje kvadratne napake (ang. root mean squared error, $RMSE$)

$$RMSE = \sqrt{MSE}, \quad (4.52)$$

ta ima namreč enako enoto kot podatki. Omenimo še srednjo absolutno odstotkovno napako (ang. mean absolute percentage error, $MAPE$), ki je dana z

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (4.53)$$

V primeru, ko so resnične vrednosti lahko tudi ničelne, je potrebno zgornjo definicijo nekoliko spremeniti, ena možnost je naslednja:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\bar{y}} \right|, \quad (4.54)$$

kjer je $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Želimo najti model, ki se dobro odreže tako na učni kot poljubni podatkovni množici (tj. ima majhno srednjo kvadratno napako). Zato nam ni v interesu minimizirati napako samo na učni množici, v tem primeru namreč nimamo informacij o kakovosti modela na neki drugi množici. Napovedno napako moramo tako izmeriti na nek drug način, ne torej na celotni učni množici.

V praksi napako navadno ocenimo na posebni testni množici oziroma na primerih, ki niso bili uporabljeni za učenje modela. Učno množico tako razbijemo na dva dela: na podmnožico primerov za učenje modela in podmnožico primerov za ocenjevanje napake (presek obeh podmnožic je prazen). Odločiti se moramo za delež testnih primerov, izbor primerov v množici za učenje pa je naključen.

Na ta način ocenimo napovedno napako, vendar pa ne želimo, da bi imela ocena veliko varianco. Ena možnost, kako znižamo varianco ocene, je prečno preverjanje (ang. cross-validation), ki jo bomo tudi uporabili. Navadno se izvaja 5- ali 10-kratno prečno preverjanje, mi bomo izbrali slednje.

Oglejmo si, kako izvedemo k -kratno prečno preverjanje. Učno množico S naključno razbijemo na k podmnožic $S_i, i = 1, \dots, k$. Pri tem mora veljati $|S_i| \approx |S|/k$, $S = \bigcup_{i=1}^k S_i$ in $\forall i \neq j : S_i \cap S_j = \emptyset$. Želimo še, da ima vsaka podmnožica približno enako porazdelitev. Za vsako podmnožico $S_i, i = 1, \dots, k$, opravimo naslednja koraka:

- naučimo se napovedni model m_i na primerih $S \setminus S_i$ (učna množica),
- izmerimo napako Err_i na primerih iz S_i (testna množica): $Err_i = Err(m_i, S_i)$.

Vsaka izmed k -tih podmnožic je torej enkrat uporabljena kot testna množica (tj. vsak primer je enkrat tudi testni primer). Ocena napovedne napake Err_{CV} je potem enaka

$$Err_{CV} = \frac{1}{k} \sum_{i=1}^k Err_i. \quad (4.55)$$

Izračunamo torej povprečje napak, izračunanih na posamezni testni množici. V našem primeru bomo za Err_i izbrali kar srednjo kvadratno napako, definirano z enačbo (4.51). S takšno metodo preverjamo ustreznost modela, torej nabor nastavitve parametrov. Za konkretno izbiro parametrov znamo torej oceniti kakovost modela tj. srednjo kvadratno napako, dobljeno z 10-kratnim prečnim preverjanjem. Vprašanje je, katere parametre moramo v našem primeru sploh določiti in na kakšen način izberemo vrednosti, ki jih želimo preizkusiti.

V primeru regresije želimo minimizirati ciljno funkcijo

$$\frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N V_\varepsilon(y_i - (\langle \beta, x_i \rangle + \beta_0)). \quad (4.56)$$

Kot smo že omenili, je potrebno izbrati konstanti ε in C , ki nastopata v zgornji ciljni funkciji. V primeru linearnega jedra je potrebno izbrati le ti dve konstanti, v primeru radialnega jedra pa je potrebno izbrati še konstanto γ , ki nastopa v enačbi (4.49).

Možnih vrednosti konstant in njihovih kombinacij je zelo veliko, zato moramo te preizkusiti sistematično. Za vsako izmed konstant bomo izbrali interval, na katerem bomo iskali optimalne vrednosti, znotraj intervala pa bomo vrednosti spreminjali z večkratnikom števila 10. Primer vrednosti, ki jih želimo preizkusiti za en parameter, je $[0.01, 0.1, 1, 10, 100]$. Za vsak parameter na ta način izberemo nabor vrednosti. Izračunamo kartezični produkt vseh naborov, da dobimo vse možne kombinacije

nastavitev parametrov, ki jih želimo preveriti (ang. grid search). Za vsako tako kombinacijo pa znamo izračunati kakovost modela.

Za končni model upoštevamo parametre, ki so privedli do najmanjše napake prečnega preverjanja Err_{CV} . Izbrani model nato naučimo na celotni podatkovni množici. Omenimo še, da vrednosti spremenljivk x_i in y_i običajno najprej standardiziramo.

5 Empirično modeliranje in napovedovanje

V prejšnjih poglavjih smo predstavili glavni metodi, ki ju bomo uporabili za modeliranje in napovedovanje cen ter povpraševanja po električni energiji, tj. modele ARMA-GARCH in metodo podpornih vektorjev. V naslednjih poglavjih se bomo najprej seznanili s podatki o ceni in povpraševanju po električni energiji ter predstavili njihove glavne značilnosti. Nato se bomo najprej posvetili povpraševanju po električni energiji, kjer bomo izbrali napovedni model in na koncu tudi napovedali porabo električne energije. Zatem bomo postopek ponovili tudi za ceno električne energije. V obeh primerih bomo modele primerjali med seboj in sklepali, kateri je boljši.

5.1 Predstavitev podatkov

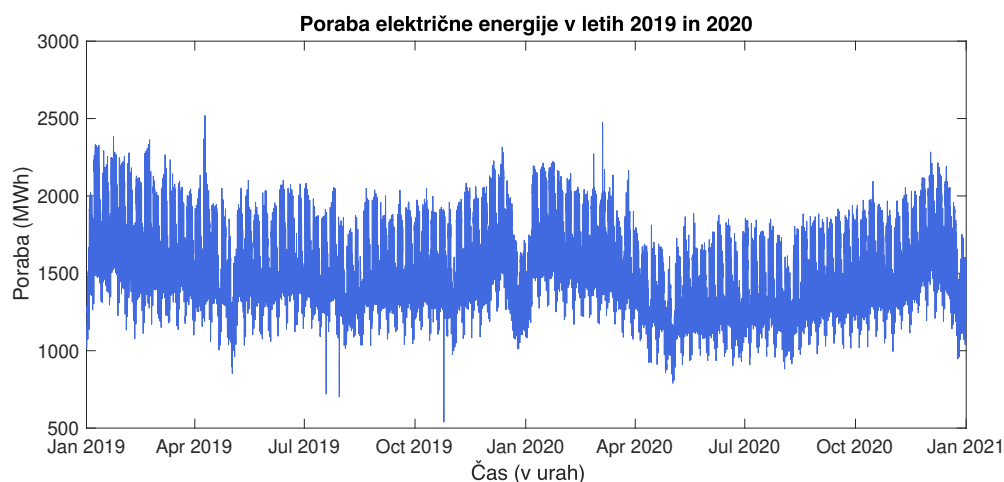
Opisane algoritme bomo uporabili na podatkih o porabi in ceni (za trgovanje dan vnaprej) električne energije v Sloveniji. Analizirali bomo podatke za obdobje med 1.1.2019 in 31.12.2020, v obeh primerih so ti dani za vsako uro v dnevu, kar skupno obsega 17.544 podatkov.

Podatke o porabi električne energije smo pridobili na spletni strani Evropskega združenja sistemskih operaterjev elektroenergetskega omrežja (ang. European Network of Transmission System Operators for Electricity, ENTSO-E, [19]). Podatki so za izbran datum na voljo v obliki razpredelnice v html formatu. V odvisnosti od datuma (za 731 dni) je potrebno uvoziti ustrezno html datoteko, v njej pa poiskati razpredelnico, ki vsebuje podatke o urni porabi. Nato je potrebno vse razpredelnice združiti v eno, ki predstavlja časovno vrsto porabe električne energije, ki jo potrebujemo za nadaljnjo analizo.

Podatke o ceni električne energije smo pridobili na spletni strani BSP SouthPool energetske borze [20]. Podatki so za vsako leto posebej na voljo v obliki razpredelnice v excel formatu, kjer je vsak mesec na svojem listu. Za vsako leto je potrebno uvoziti 12 listov, podatke urediti in jih združiti v eno samo razpredelnico, ki predstavlja časovno vrsto cene električne energije, ki jo bomo v nadaljevanju analizirali.

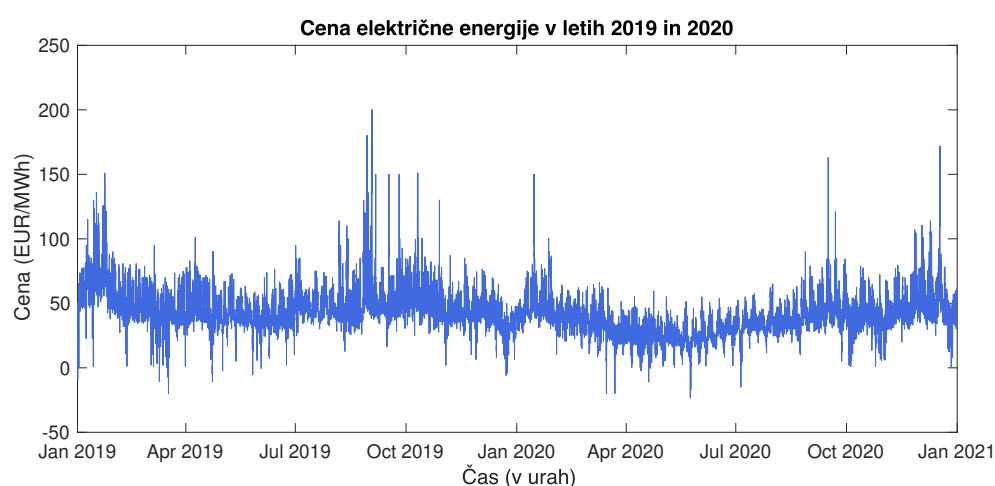
Najprej si podatke grafično ogledjmo. Slika 12 prikazuje urno gibanje porabe električne energije v letih 2019 in 2020.

Opazimo, da je poraba najvišja jeseni in pozimi, v času kurilne sezone. Najnižja je spomladi, nekoliko višja pa zopet poleti, ko so potrebne npr. klimatske naprave, vendar še vedno občutno nižja kot v času kurilne sezone. V času praznikov (npr. konec decembra in aprila) zopet opazimo nekoliko manjšo porabo.



Slika 12: Gibanje porabe električne energije v letih 2019 in 2020.

Oglejmo si sliko 13, ki prikazuje urno gibanje cene električne energije v letih 2019 in 2020.



Slika 13: Gibanje cene električne energije v letih 2019 in 2020.

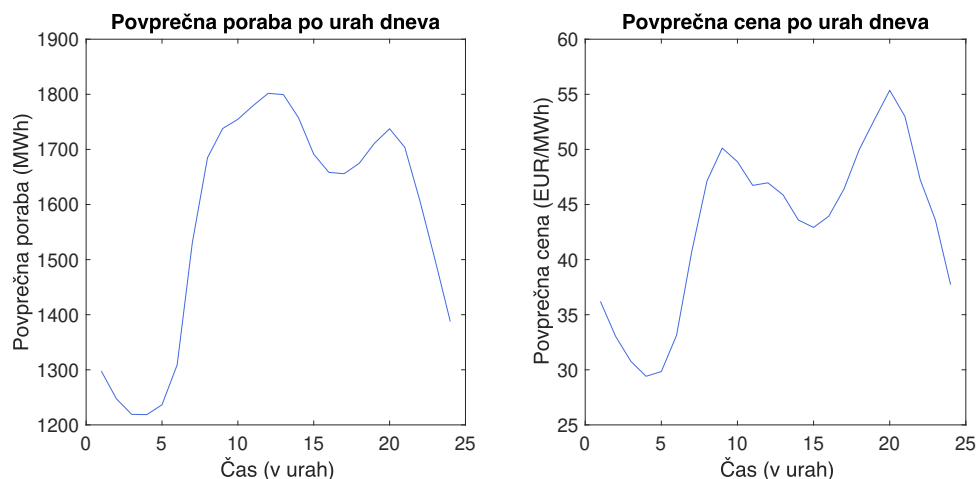
Vidimo, da je cena električne energije veliko bolj volatilna kot pa poraba. Zopet opazimo, da je cena v pomladnih in poletnih mesecih nekoliko nižja, vendar razlika ni tako velika. Nekajkrat je cena tudi negativna. Oglejmo si opisne statistike podatkov, ki so prikazane v tabeli 1.

	Min	Max	Povprečje	Mediana	SD
Poraba (MWh)	539	2520	1570.8	1570	294.68
Cena (EUR/MWh)	-23.45	200.02	43.14	41.22	18.16

Tabela 1: Opisne statistike.

Povprečna poraba električne energije se giblje okoli 1570 MWh, minimalna dosežena vrednost je 539 MWh, maksimalna pa 2520 MWh. Pri ceni res vidimo, da zavzame tudi negativne vrednosti, minimum je namreč -23.45 EUR/MWh. Na sliki 13 smo videli, da se to ni zgodilo samo enkrat. Negativne cene so smiselne v primerih proizvajalcev, kjer so stroški prekinitve proizvodnje električne energije (torej da napravo ugasnejo) večji, kot če električno energije prodajo po negativni ceni (torej zanjo plačajo).

Slika 14 prikazuje povprečno ceno in porabo po urah. Vidimo, da je gibanje dokaj podobno. Ob osmih zvečer začneta tako poraba kot cena padati, to traja do petih zutraj, ko je dosežen minimum. Okrog pete ure se začnejo prebujati gospodinjstva, tovarne začnejo obratovati, zato začneta poraba in cena naraščati. Naraščanje je opaziti približno do enajste/dvanajste ure, nekateri imajo namreč samo dopoldanski delavnik, spet drugi pa odhajajo na malice. Nato je opaziti padec, ki traja do tretje ure popoldne, in spet naraščanje do osme ure zvečer. Na dnevni ravni je med minimalno in maksimalno ceno razlika približno 25 EUR/MWh, med minimalno in maksimalno porabo pa 600 MWh. Poraba svoj maksimum doseže okrog enajste ure dopoldne, cena pa okrog osme ure zvečer.



Slika 14: Levo: povprečna poraba po urah. Desno: povprečna cena po urah.

5.2 Metodologija

V prejšnjem poglavju smo si ogledali gibanje porabe in cene električne energije ter njune osnovne značilnosti. Zdaj bomo obe količini še natančneje raziskali, konstruirali modele in nato še napovedi. Najprej se bomo posvetili porabi električne energije, nato pa še ceni. V obeh primerih bomo postopali v naslednjem vrstnem redu:

1. konstrukcija modelov ARMA-GARCH,
2. konstrukcija modelov SVR in
3. napovedovanje in primerjava modelov.

Za celotno modeliranje, ocenjevanje parametrov in napovedovanje bomo uporabili program MATLAB R2021a. Za modele tipa ARMA-GARCH bomo potrebovali Econometrics Toolbox, model bomo specifikirali s funkcijo *arima* [3], parametre ocenili s funkcijo *estimate* [9] in napovedi naredili s funkcijo *forecast* [12]. Za modele tipa SVR bomo model naučili s funkcijo *fitrsvm* [11], napovedi pa izdelali s funkcijo *predict* [16].

Konstrukcija modelov ARMA-GARCH bo potekala po Box-Jenkinsovi metodologiji, ki je predstavljena v poglavju 3.7.1. Pri konstrukciji se bomo oprli na grafa vzorčnih ACF in PACF, za izbrano število odlogov ju lahko izračunamo s funkcijama *autocorr* [4] in *parcorr* [15]. Dodatno bomo uporabili še informacijski kriterij AIC, ki ga lahko določimo s funkcijo *aicbic* [2]. Za validacijo modela bomo uporabili Ljung-Boxov test, ki ga lahko izvedemo s funkcijo *lbqtest* [14]. Konstruirali bomo tri različne modele v odvisnosti od variance, in sicer ARMA-GARCH- n , ARMA-GARCH- t in ARMA-EGARCH. Funkcija *estimate* parametre modelov oceni z metodo največjega verjetja.

Konstrukcija modelov tipa SVR in izbira hiperparametrov bo potekala na osnovi prečnega preverjanja. Za izbrane kombinacije hiperparametrov bomo uporabili 10-kratno prečno preverjanje, opisano v poglavju 4.5. Konstruirali bomo dva različna modela v odvisnosti od jedra, in sicer linearno (SVR-lin) in radialno jedro (SVR-rbf). Konstruirali bomo avtoregresijske SVR modele, kar pomeni, da bomo prihodnjo porabo/ceno izračunali z uporabo preteklih vrednosti porabe/cene kot napovednih spremenljivk. Kot smo že omenili, je podatke priporočljivo najprej standardizirati, saj se tako bistveno zmanjša čas učenja modelov. To lahko neposredno podamo funkciji *fitrsvm*, tako da dodamo '*Standardize*', *true*.

Zaključni del vsakega sklopa (za ceno in porabo) bo predstavljala primerjava uspešnosti napovedovanja modelov ARMA-GARCH in SVR. Izbiro modelov bomo naredili na podlagi podatkov iz leta 2019, modele pa bomo primerjali na napovedih za leto 2020. Modele bomo natančneje primerjali na podlagi napovedi za tri izbrane tedne, in sicer tretji teden februarja, tretji teden julija in tretji teden septembra. V teh mesecih je poraba navadno visoka, srednja in nizka. Zaradi kompletnosti bomo napovedi naredili tudi za tretji teden preostalih mesecev v letu 2020 in izvedli primerjavo na letni ravni.

Da lahko ocenimo kakovost napovedi, pa je te najprej potrebno dobiti. Konstrukcija obeh modelov nam da osnovne nastavitve modela (koeficiente). Za ocenjevanje parametrov modela bomo uporabili drseč interval (ang. rolling window), to metodo opisuje naslednji postopek. Za začetni vzorec (tj. od 1. januarja 2019 do 31. decembra 2019) ocenimo parametre modela in dobimo napoved za naslednjih 24 ur. Nato ocenjevanemu vzorcu dodamo podatke za naslednjih 24 ur, hkrati pa odstranimo podatke za najstarejših 24 ur. Na podlagi vsakega novega vzorca ponovno ocenimo parametre modela in izdelamo napoved za naslednjih 24 ur. Ta postopek ponavljamo, dokler ne pridobimo napovedi za obdobje od 1. januarja 2020 do 31. decembra 2020. V našem primeru sicer izpustimo vmesne napovedi in modele, saj nas zanima le tretji teden vsakega meseca. Na ta način za vsak tip modela ocenimo parametre 84 modelov za različne časovne odseke podatkov.

Za primerjavo kakovosti modelov bomo uporabili dve meri napake, *MAPE* in *RMSE*. Obe smo že spoznali v poglavju 4.5, napaka *RMSE* je dana z enačbo

(4.52), napaka $MAPE$ pa je dana z enačbo (4.53). Ker lahko cena električne energije zavzame tudi ničelno vrednost (ali pa vrednosti blizu 0), standardna definicija napake $MAPE$ ne ustreza, saj zahteva deljenje z 0. V tem primeru bomo uporabili alternativno definicijo, dano z enačbo (4.54). Za izbrane tri tedne (februar, julij in september) bomo napovedi primerjali glede na napako $MAPE$, ki pa jo bomo izračunali za vsak dan v tednu. Za celostno primerjavo vseh tretjih tednov v letu bomo napako $MAPE$ izračunali na tedenski ravni, prav tako tudi $RMSE$.

5.3 Napovedovanje povpraševanja po električni energiji

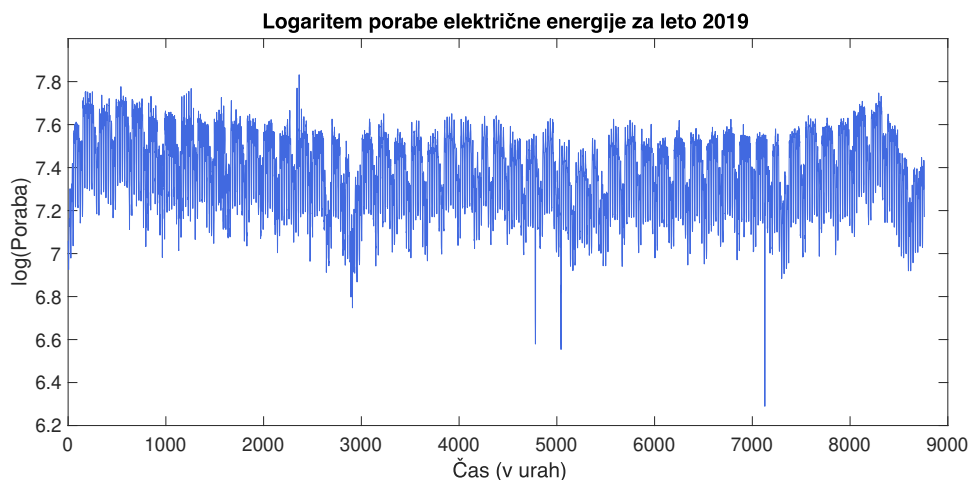
Kot smo že omenili, se bomo najprej osredotočili na napovedovanje porabe električne energije. Najprej bomo konstruirali modele ARMA-GARCH, nato pa še modele SVR. Nazadnje bomo predstavili rezultate in primerjali napovedi modelov med seboj. Časovno vrsto porabe električne energije bomo označili z D_t .

5.3.1 ARMA-GARCH

Za konstrukcijo modela tipa ARMA-GARCH bomo uporabili Box-Jenkinsovo metodologijo. Izbiro modela bomo naredili na podatkih za leto 2019, torej v obdobju med 1. 1. 2019 in 31. 12. 2019.

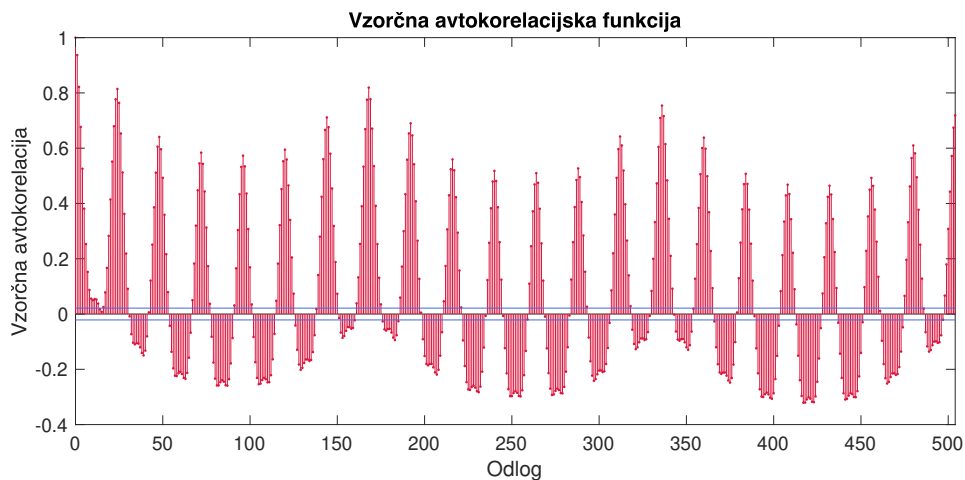
Izbiro naredimo na podlagi podrobne analize urnih podatkov za porabo električne energije. Glavne lastnosti, ki smo jih že opazili v poglavju 5.1 so: visokofrekvenčni podatki, nekonstantno povprečje in varianca ter večkratna sezonskost (dnevna, tedenska). Ti dejavniki so med glavnimi, da se odločimo za model ARMA-GARCH.

V naslednjem koraku moramo na podlagi podatkov identificirati model. Ker so podatki dokaj volatilni, predhodno opravimo še logaritemsko transformacijo, da zgladimo učinek volatilnosti in jo stabiliziramo. Uvodoma nas zanima, ali je logaritmizirana časovna vrsta porabe električne energije stacionarna. Najprej si oglejmo graf časovne vrste $\ln D_t$, ki je prikazan na sliki 15.



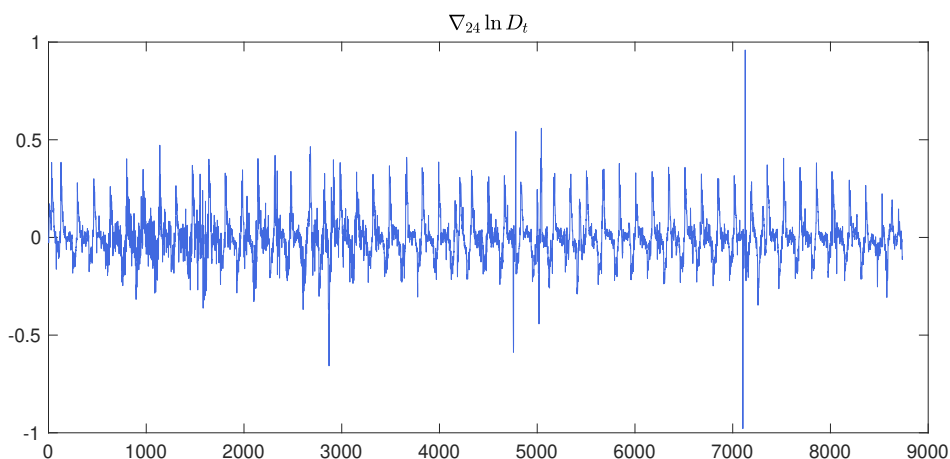
Slika 15: Logaritemska vrednost porabe električne energije za leto 2019.

Tako povprečje kot varianca ne delujeta konstantna, vseeno pa si oglejmo še vzorčno ACF časovne vrste $\ln D_t$, ki jo prikazuje slika 16.



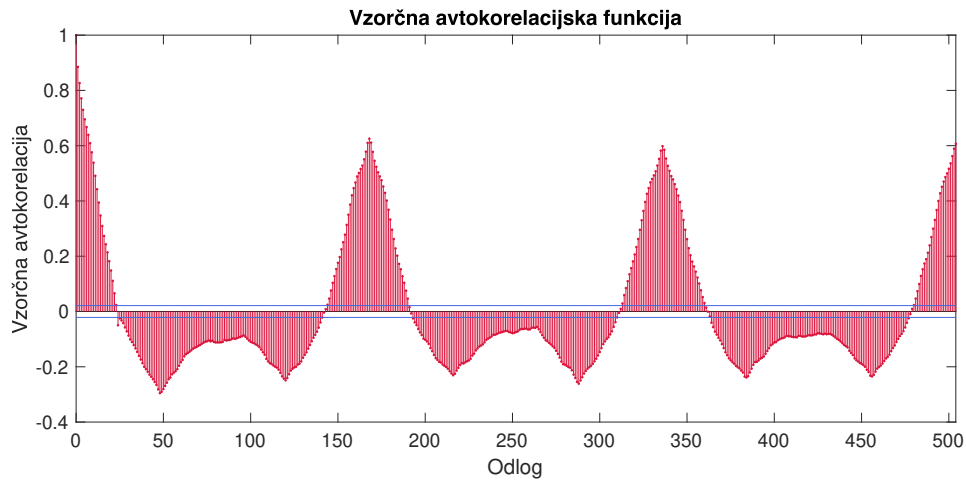
Slika 16: Vzorčna avtokorelacijska funkcija časovne vrste $\ln D_t$.

Vzorčne avtokorelacije so izračunane do odloga 504, kar predstavlja 3 tedne (504 ur). Podatki so očitno nestacionarni, saj ACF sploh ne pada (vidimo, da se vzorec ponavlja vse do treh tednov). Opazimo močno sezonsko komponento, kar smo tudi pričakovali na podlagi predhodnih analiz. Da dosežemo stacionarnost, bo potrebno sezonsko diferenciranje. Perioda je očitno enaka 24 (en dan), to je razvidno iz grafa ACF, pa tudi iz same narave podatkov, saj imamo podatke dane na urni ravni. V ta namen na časovni vrsti $\ln D_t$ uporabimo operator ∇_{24} (spomnimo se: operator d -te difference je definiran kot $\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t$). Logaritmirano časovno vrsto torej sezonsko diferenciramo do prvega reda s periodo 24. Slika 17 prikazuje graf diferencirane časovne vrste.



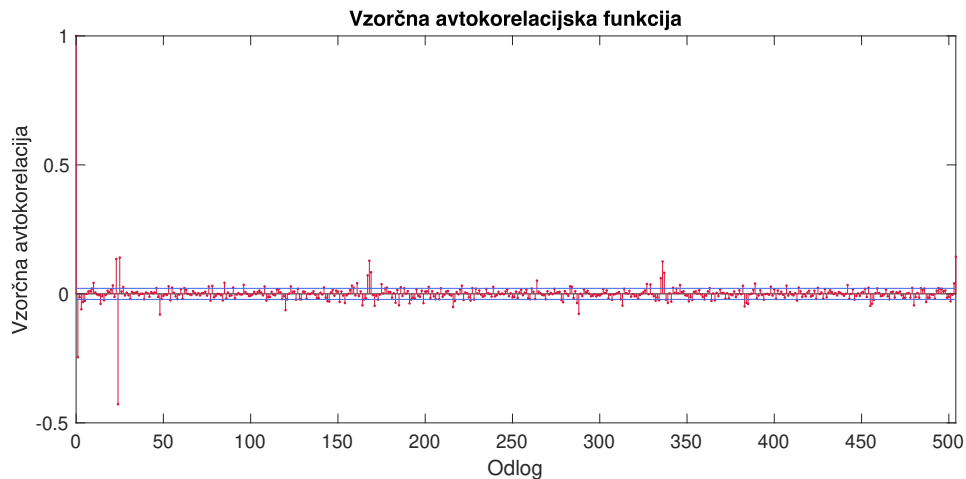
Slika 17: Logaritmirana časovna vrsta porabe električne energije, ki je sezonsko diferencirana do prvega reda s periodo 24.

Iz zgornje slike ne moremo oceniti, ali je časovna vrsta stacionarna ali ne. Deluje pa bolj stacionarno kot prvotna časovna vrsta. Slika 18 prikazuje vzorčno ACF časovne vrste $\nabla_{24} \ln D_t$.



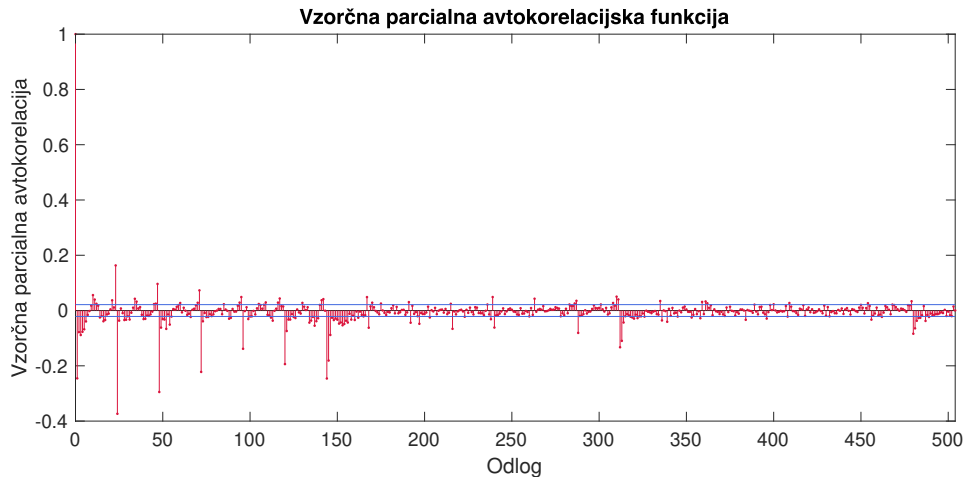
Slika 18: Vzorčna avtokorelacijska funkcija časovne vrste $\nabla_{24} \ln D_t$, ki je prikazana na sliki 17.

Tudi po sezonskem diferenciranju časovna vrsta ne izgleda stacionarno. Uporabimo operator ∇ ($\nabla X_t = X_t - X_{t-1}$), torej časovno vrsto navadno diferenciramo do prvega reda. Slika 19 prikazuje vzorčno ACF časovne vrste, ki je sezonsko diferencirana do prvega reda s periodo 24 in navadno diferencirana do prvega reda, tj. $\nabla \nabla_{24} \ln D_t$.



Slika 19: Vzorčna avtokorelacijska funkcija časovne vrste $\nabla \nabla_{24} \ln D_t$.

Časovno vrsta zdaj izgleda stacionarno in je primerna za model ARMA, torej lahko nadaljujemo z modeliranjem. Še prej pa izračunajmo vzorčno PACF za stacionarno časovno vrsto $\nabla \nabla_{24} \ln D_t$, saj jo bomo potrebovali pri identifikaciji modela. Prikazuje jo slika 20.



Slika 20: Vzorčna parcialna avtokorelacijska funkcija časovne vrste $\nabla\nabla_{24} \ln D_t$.

Pri nadaljni konstrukciji modela se bomo oprli na grafa vzorčne ACF in vzorčne PACF časovne vrste $\nabla\nabla_{24} \ln D_t$, s pomočjo kriterija AIC pa bomo izbrali končne parametre modela. Konstruirali bomo sezonski model ARIMA, ki je dan z definicijo 3.34, ima torej obliko

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D \ln D_t = \theta(B)\Theta(B^S)Z_t. \quad (5.1)$$

Pri tem so $\phi(z) = 1 - \phi_1(z) - \dots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1(z) - \dots - \Phi_P z^P$, $\theta(z) = 1 + \theta_1(z) + \dots + \theta_q z^q$ in $\Theta(z) = 1 + \Theta_1(z) + \dots + \Theta_Q z^Q$, D_t pa je poraba ob času t . Spomnimo se: B je operator premik nazaj, $B^l X_t = X_{t-l}$. Ker smo predhodno ugotovili, da varianca ni konstantna, Z_t ustreza enemu izmed modelov GARCH (modeliranju variance se bomo posvetili v nadaljevanju). S predstavlja periodo sezonskosti. S predhodnim diferenciranjem smo že ugotovili, da je časovna vrsta

$$Y_t := \nabla\nabla_{24} \ln D_t = (1-B)(1-B^{24}) \ln D_t \quad (5.2)$$

stacionarna, vemo torej, da sta parametra d in D enaka 1. Na podlagi prejšnjih ugotovitev prav tako vemo, da je perioda enaka 24 (tj. $S = 24$), torej dnevna perioda. Model lahko zato poenostavimo v

$$\phi(B)\Phi(B^{24})(1-B)(1-B^{24}) \ln D_t = \theta(B)\Theta(B^{24})Z_t. \quad (5.3)$$

Določiti moramo še vrednosti parametrov P in Q za sezonski AR in MA del ter vrednosti parametrov p in q za nesezonski AR in MA del. Številne vmesne vrednosti koeficientov ϕ_i , Φ_i , θ_i in Θ_i so lahko enake 0. Graf vzorčne PACF na sliki 20 nam bo pomagal pri določitvi parametrov AR dela (p in P), graf vzorčne ACF pa pri določitvi parametrov MA dela (q in Q). Pogled na obe sliki ne namiguje na nobeno preprosto obliko modela.

Najprej se bomo osredotočili na sezonska parametra P in Q , pozornost bomo usmerili na korelacije pri odlogih, ki so večkratniki periode S . Če si ogledamo vzorčno PACF, lahko visoko korelacijo opazimo pri odlogu 24, prisotne pa so še pri odlogih 48, 72, 96, 120, 144 in 168, vendar se postopoma manjšajo. Nekoliko izstopa

tudi parcialna avtokorelacija pri odlogu 336 (dva tedna). Tako lahko določimo parameter P (sezonski AR del). Če si ogledamo še graf vzorčne ACF, vidimo mnogo manj visokih avtokorelacij. Občutna je pri odlogu 24 (en dan), opazimo pa jo tudi pri odlogu 168 (en teden) in 336 (trije tedni). Parametra P in Q bosta tako zagotovo enaka vsaj 1, natančneje pa ju bomo določili še s kriterijem AIC, kjer bomo med seboj primerjali različne modele. Pri odlogih, kjer so večje korelacije (konice na grafu) vrednosti koeficientov Φ_i in Θ_i ne bodo ničelne, povsod drugod pa.

Zdaj se lahko osredotočimo še na nesezonska parametra p in q . Zopet si najprej oglejmo vzorčno PACF, visoko korelacijo lahko opazimo pri prvih nekaj urah, to spada v AR del. Opazimo še konice pri odlogih 23, 47, 71, 95, torej v urah pred večkratniki periode, kar zopet spada v AR del. Na grafu vzorčne ACF so večje avtokorelacije zopet vidne v prvih nekaj urah in ob urah pred večkratniki periode, kar spada v MA del. Spet velja, da pri odlogih, kjer so večje korelacije, vrednosti koeficientov ϕ_i in θ_i ne bodo ničelne, povsod drugod pa.

Vzorčni ACF in PACF nam tako pomagata zaznati, katere odloge bi bilo potrebno vključiti. Zagotovo bomo vključili prvih nekaj ur, ki kažejo močan vpliv, koliko dodatnih odlogov bomo še vključili, pa se bomo odločili s pomočjo kriterija AIC. Sprva bomo konstruirali 16 modelov, kjer bomo postopoma vključevali večje število odlogov, ki smo jih zasledili pri grafih vzorčne ACF in PACF. Uvodoma bomo Z_t modelirali kot proces GARCH(1,1), ki velja za najpreprostejšega (vemo, da varianca ni konstantna). Najpreprostejši model bo oblike

$$\phi(B)\Phi(B^{24})(1-B)(1-B^{24})\ln D_t = \theta(B)\Theta(B^{24})Z_t, \quad (5.4)$$

kjer so

$$\begin{aligned} \phi(z) &= (1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \phi_4 z^4), \\ \Phi(z) &= 1 - \Phi_1 z, \\ \theta(z) &= 1 + \theta_1 z + \theta_3 z^3 \quad \text{in} \\ \Theta(z) &= 1 + \Theta_1 z. \end{aligned}$$

Parameter p je torej enak 4, parameter q pa je enak 3, vendar je koeficient θ_2 ničeln. Parametra P in Q sta oba enaka 1. Zgornji model bomo krajše zapisali kot: M1 = {AR = [1-4], MA = [1,3], S = 24, SAR = 1, SMA = 1}. Najbolj kompleksen model ima v zgornjem zapisu obliko M16 = {AR = [1-5,23], MA = [1,3-5,23,25], S = 24, SAR = [1-2], SMA = 1}.

Ocenimo koeficiente modelov in izračunajmo kriterij AIC. Tabela 2 prikazuje specifikacije 16 izbranih modelov in pripadajočo vrednost kriterija AIC. Vrednosti še skaliramo, tako da jih delimo z velikostjo vzorca (funkciji *aicbic* podamo dodaten parameter). Vrstica, ki pripada modelu z najmanjšo vrednostjo kriterija AIC, je odebeljena. Najnižjo vrednost ima model številka 13, ki je med bolj kompleksnimi (vključenih je več odlogov). Četudi kriterij AIC kaznuje veliko število parametrov, se v tem primeru vseeno izkaže, da je bolje vključiti več parametrov.

Model	AR	MA	SAR	SMA	AIC
1	1-4	1,3	1	1	-3.9035
2	1-5	1,3	1	1	-3.9034
3	1-5,23	1,3	1	1	-3.9229
4	1-5,23	1,3-5	1	1	-3.9153
5	1-5,23	1,3-5,23	1	1	-3.9321
6	1-5,23	1,3,23	1	1	-3.9227
7	1-5,23	1,3,23,25	1	1	-3.9204
8	1-5,23	1,3-5,23,25	1	1	-3.9280
9	1-4	1,3	1-2	1	-3.9039
10	1-5	1,3	1-2	1	-3.9038
11	1-5,23	1,3	1-2	1	-3.9148
12	1-5,23	1,3-5	1-2	1	-3.9253
13	1-5,23	1,3-5,23	1-2	1	-3.9352
14	1-5,23	1,3,23	1-2	1	-3.9227
15	1-5,23	1,3,23,25	1-2	1	-3.9306
16	1-5,23	1,3-5,23,25	1-2	1	-3.9281

Tabela 2: Vrednost kriterija AIC za 16 izbranih modelov.

Ker smo na grafih vzorčne ACF in PACF zaznali korelacije tudi pri kasnejših odlogih, se odločimo zgraditi še bolj kompleksne modele. S tem pa se tudi znatno poveča čas ocenjevanja parametrov modela, pri najbolj kompleksnem traja to že več kot pol ure. Skonstruirali bomo 10 modelov. Najpreprostejši ima obliko $M1 = \{\text{AR} = [1-5,23,47], \text{MA} = [1,3-5,23], S = 24, \text{SAR} = [1-2], \text{SMA} = 1\}$, najkompleksnejši pa $M10 = \{\text{AR} = [1-5,23,47,71,95], \text{MA} = [1,3-5,23,167,335], S = 24, \text{SAR} = [1-7], \text{SMA} = [1,7,14,21]\}$. Zopet ocenimo koeficiente modelov in izračunajmo kriterij AIC. Tabela 3 prikazuje specifikacije 10 izbranih modelov in pripadajoče vrednosti kriterija AIC.

Model	AR	MA	SAR	SMA	AIC
1	1-5,23,47	1,3-5,23	1-2	1	-3.9356
2	1-5,23,47,71	1,3-5,23	1-3	1	-3.9357
3	1-5,23,47,71,95	1,3-5,23	1-4	1	-3.9356
4	1-5,23,47,71,95	1,3-5,23	1-5	1	-3.9478
5	1-5,23,47,71,95	1,3-5,23	1-6	1	-3.9401
6	1-5,23,47,71,95	1,3-5,23	1-7	1	-3.9761
7	1-5,23,47,71,95	1,3-5,23,167	1-7	1,7	-4.1183
8	1-5,23,47,71,95	1,3-5,23,167	1-7	1,7,14	-4.1188
9	1-5,23,47,71,95	1,3-5,23,167,335	1-7	1,7,14	-4.1209
10	1-5,23,47,71,95	1,3-5,23,167,335	1-7	1,7,14,21	-4.1207

Tabela 3: Vrednost kriterija AIC za 10 izbranih modelov.

Vrstica, ki pripada modelu z najmanjšo vrednostjo kriterija AIC, je ponovno

odebeljena. Najnižjo vrednost ima model številka 9, to je model oblike

$$\phi(B)\Phi(B^{24})(1-B)(1-B^{24})\ln D_t = \theta(B)\Theta(B^{24})Z_t, \quad (5.5)$$

kjer so

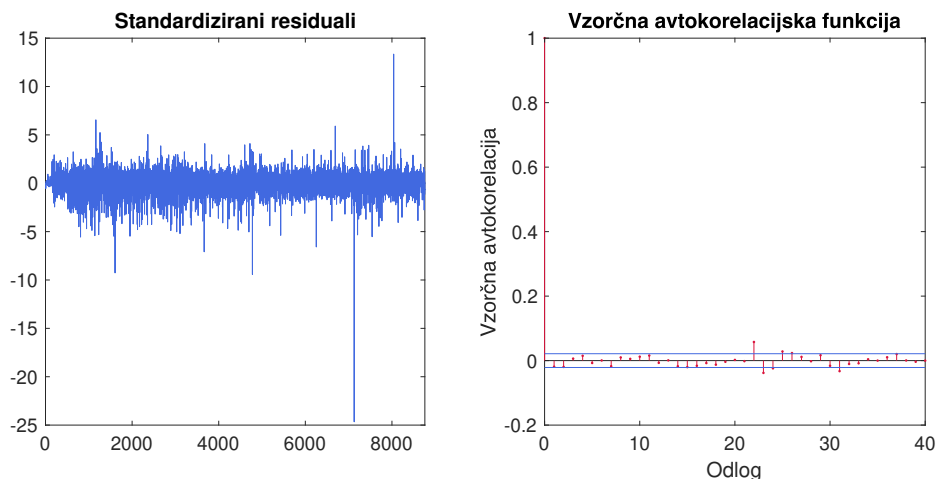
$$\begin{aligned} \phi(z) &= (1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \phi_4 z^4 - \phi_5 z^5 - \phi_{23} z^{23} - \phi_{47} z^{47} - \phi_{71} z^{71} - \phi_{95} z^{95}), \\ \Phi(z) &= 1 - \Phi_1 z - \Phi_2 z^2 - \Phi_3 z^3 - \Phi_4 z^4 - \Phi_5 z^5 - \Phi_6 z^6 - \Phi_7 z^7, \\ \theta(z) &= 1 + \theta_1 z + \theta_3 z^3 + \theta_4 z^4 + \theta_5 z^5 + \theta_{23} z^{23} + \theta_{167} z^{167} + \theta_{335} z^{335} \quad \text{in} \\ \Theta(z) &= 1 + \Theta_1 z + \Theta_7 z^7 + \Theta_{14} z^{14}. \end{aligned}$$

Po izbiri oblike sezonskega modela ARIMA moramo izbrati še model za Z_t . Kot smo že omenili, podatki kažejo, da varianca ni konstantna. Konstruirali bomo tri različne modele GARCH- n , GARCH- t in EGARCH. Za vsakega izmed njih bomo preizkusili 6 kombinacij parametrov p in q ter izračunali vrednost kriterija AIC. Kombinacije so sledeče: (1,1), (1,2), (2,1), (1,3), (3,1) in (2,2).

Med modeli GARCH- n se je za najboljšega izkazal model GARCH- $n(3,1)$. Vrednost kriterija AIC je enaka -4.1410 . Med modeli GARCH- t je imel model GARCH- $t(2,2)$ najnižjo vrednost kriterija AIC, ta je enaka -4.4564 . Model GARCH- t torej nakazuje boljše prilaganje podatkom kot pa model GARCH- n , kar namiguje, da je t -porazdelitev, ki ima težji rep, morda res primernejša za naše podatke. Med modeli EGARCH pa se je za najboljšega izkazal model EGARCH(2,1). Vrednost kriterija AIC je enaka -4.1825 . Najugodnejšo vrednost kriterija AIC ima torej model GARCH- $t(2,2)$.

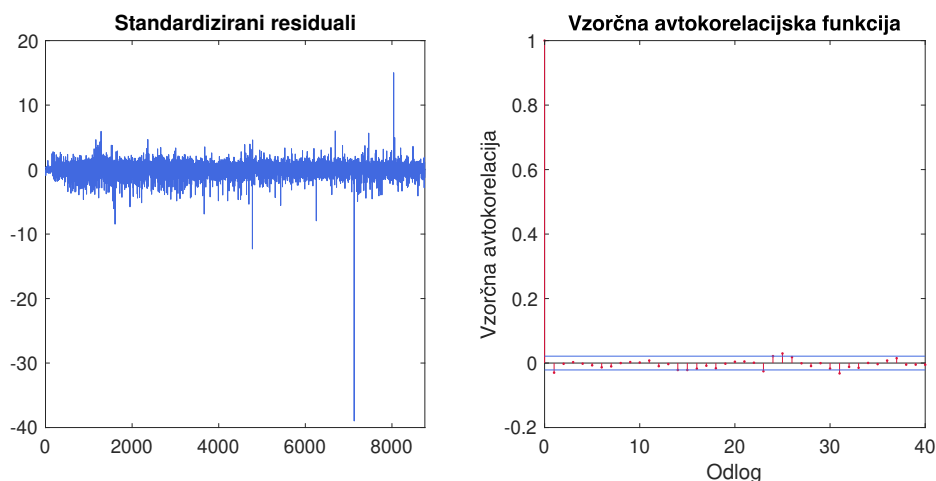
Za vsako vrsto variance smo torej izbrali po en model, katerega napovedi bomo primerjali. Seveda pa moramo prej modele še potrditi. Postopek poteka enako za vse tri modele, ki smo jih predhodno izbrali. Najprej si bomo ogledali standardizirane residue modela. Veljati mora namreč, da residuali tvorijo beli šum. Nekoreliranost bomo preverili z Ljung-Boxovim testom za različne odloge. Heteroskedastičnost residualov bomo preverili z Ljung-Boxovim testom za kvadrate residualov.

Najprej analizirajmo model SARIMA-GARCH- $n(3,1)$, ki smo ga izbrali na podlagi kriterija AIC. Naredimo Ljung-Boxov test na standardiziranih residualih za odloge 5, 10, 15 in 20. Pripadajoče p -vrednosti so enake 0.0087, 0.0275, 0.0126 in 0.0198. V vseh primerih so manjše kot 0.05, torej pri stopnji značilnosti $\alpha = 0.05$ zavrnemo ničelno hipotezo, da so residuali nekorelirani. Ker so vrednosti kriterija AIC blizu tudi za druge kombinacije parametrov modela GARCH- n , naredimo Ljung-Boxov test še za preostale modele. Model GARCH- $n(2,2)$ ima drugo najmanjšo vrednost kriterija AIC, vendar pa tudi v tem primeru zavrnemo ničelno hipotezo, da so residuali nekorelirani. Naslednji je model GARCH- $n(1,3)$, zopet naredimo Ljung-Boxov test. Tokrat so pripadajoče p -vrednosti enake 0.0991, 0.1570, 0.0783 in 0.1182, vse so večje od 0.05. Tokrat ne moremo zavrniti hipoteze, da so residuali nekorelirani. Preverimo še heteroskedastičnost residualov. Naredimo Ljung-Boxov test na standardiziranih kvadratih residualov, p -vrednosti so enake 0.9817, 0.9994, 1.0000 in 1.0000. V residualih torej ni zaznati heteroskedastičnosti. Oglejmo si še graf standardiziranih residualov in vzorčno ACF, prikazuje ju slika 21. Residuali res spominjajo na beli šum (korelacije so večinoma znotraj modrih črt). Model SARIMA-GARCH- $n(1,3)$ lahko torej uporabimo za napovedovanje.



Slika 21: Levo: Standardizirani residuali modela SARIMA-GARCH- $n(1,3)$. Desno: pripadajoča vzorčna avtokorelacijska funkcija.

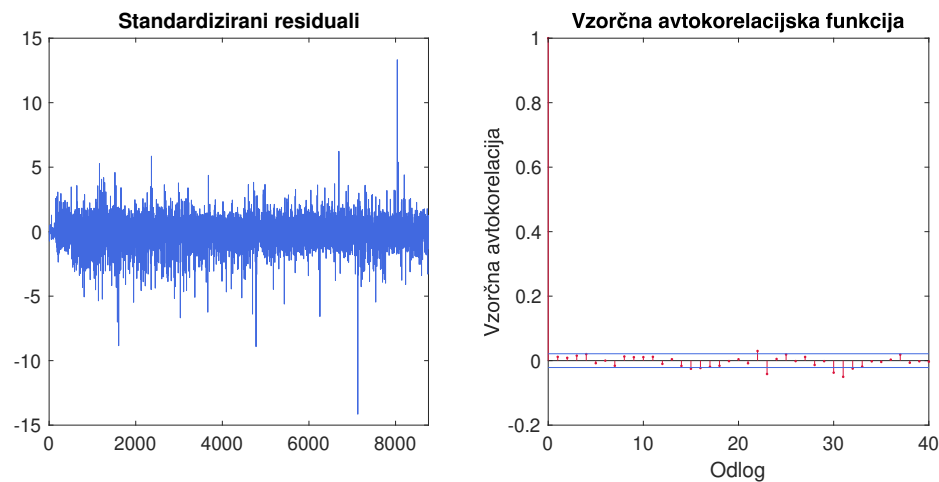
Nadaljujemo z modelom SARIMA-GARCH- $t(2,2)$. Naredimo Ljung-Boxov test za odloge 5, 10, 15 in 20, p -vrednosti so manjše od 0.05, torej zavrnemo ničelno hipotezo, da so residuali nekorelirani. Enak sklep naredimo v primeru modela SARIMA-GARCH- $t(1,1)$. Naslednji model je SARIMA-GARCH- $t(3,1)$, v tem primeru so p -vrednosti enake 0.1352, 0.3558, 0.1503, 0.1520. Ne moremo zavrniti hipoteze, da so residuali nekorelirani. Preverimo še heteroskedastičnost residualov. Ljung-Boxov test na kvadratih residualov vrnete sklep, da v residualih ni prisotne heteroskedastičnosti. Narišimo še graf standardiziranih residualov in vzorčno ACF, prikazana sta na sliki 22. Sklepamo, da je model SARIMA-GARCH- $t(3,1)$ primeren za napovedovanje.



Slika 22: Levo: Standardizirani residuali modela SARIMA-GARCH- $t(3,1)$. Desno: pripadajoča vzorčna avtokorelacijska funkcija.

Nazadnje si oglejmo še model SARIMA-EGARCH(2,1). Zopet naredimo Ljung-

Boxov test za odloge 5, 10, 15 in 20. Zavrnilo ničelno hipotezo, da so residuali nekorelirani. Enak sklep naredimo pri preostalih petih kombinacijah, najbližje ustreznemu modelu pa smo v primeru modela SARIMA-EGARCH(1,3). Zato se odločimo preizkusiti še model EGARCH(1,4). Naredimo Ljung-Boxov test, p -vrednosti so enake 0.1510, 0.1806, 0.0635, 0.0243. Tokrat ne moremo zavrnilo hipoteze, da so residuali nekorelirani, razen za odlog 20, vendar vseeno nadaljujemo z modelom. Kot v obeh modelih do zdaj, na podlagi Ljung-Boxovega testa ne moremo zavrnilo hipoteze, da v residualih ni prisotne heteroskedastičnosti. Narišemo še graf residualov ter vzorčno ACF, kar lahko vidimo na sliki 23. Sklepamo, da je model SARIMA-EGARCH(1,4) primeren za napovedovanje.



Slika 23: Levo: Standardizirani residuali modela SARIMA-EGARCH(1,4)). Desno: pripadajoča vzorčna avtokorelacijska funkcija.

Z uporabo treh izbranih modelov bomo v nadaljevanju izdelali napovedi. Postopek napovedovanja je opisan v poglavju 5.2, rezultati pa so predstavljeni v poglavju 5.3.3. Za celostno primerjavo bomo najprej konstruirali še modele SVR in jih potem na koncu primerjali z modeli ARMA-GARCH.

5.3.2 SVR

Tudi v primeru podpornih vektorjev bo izbira modela potekala na podatkih od 1.1.2019 do 31.12.2019. Kot smo že omenili, bomo konstruirali avtoregresijski SVR model. Prihodnjo porabo bomo izračunali z uporabo preteklih vrednosti porabe kot napovednih spremenljivk. Odločiti pa se moramo, katere pretekle vrednosti upoštevati. Izbrali bomo odloge, ki smo jih kot pomembne zasledili pri analizi vzorčne ACF in PACF v prejšnjem poglavju, ti so: 1, 2, 3, 4, 5, 23, 24, 47, 48, 71, 72, 95, 96, 120, 144, 168, 335 in 336. Vendar pa tokrat za izgradnjo modela ne bo dovolj samo časovna vrsta porabe električne energije, konstruirati moramo namreč ustrezno tabelo. Vrstico v tabeli bo predstavljal par (x_t, y_t) , $t = 1, \dots, 8760$, vsaki urni vrednosti porabe torej pripada ena vrstica tabele. Pri tem x_t predstavlja vektor

napovednih spremenljivk

$$x_t = ([x_t]_1, [x_t]_2, [x_t]_3, [x_t]_4, [x_t]_5, [x_t]_{23}, [x_t]_{24}, [x_t]_{47}, [x_t]_{48}, [x_t]_{71}, [x_t]_{72}, [x_t]_{95}, \\ [x_t]_{96}, [x_t]_{120}, [x_t]_{144}, [x_t]_{168}, [x_t]_{335}, [x_t]_{336}),$$

kjer je $[x_t]_i = D_{t-i}$, imamo torej 18 napovednih spremenljivk (pretekle porabe). Vrednost ciljne spremenljivke y_t je enaka D_t , torej vrednost porabe ob času t . Zapišemo lahko

$$D_t = f(D_{t-1}, D_{t-2}, D_{t-3}, D_{t-4}, D_{t-5}, D_{t-23}, D_{t-24}, D_{t-47}, \\ D_{t-48}, D_{t-71}, D_{t-72}, D_{t-95}, D_{t-96}, D_{t-120}, D_{t-144}, D_{t-168}, D_{t-335}, D_{t-336}),$$

poraba ob času t je torej funkcija preteklih vrednosti porabe.

Napovedi bomo naredili za 24 ur naprej, zgornji model pa za napoved ob času t potrebuje preteklih 5 vrednosti. Te niso dostopne za napovedi od druge ure naprej, zato jih postopoma nadomestimo z napovedmi. Za prvo uro imamo tako vse potrebne podatke, za drugo uro pa manjka podatek za uro nazaj, tega nadomestimo z napovedjo. Podobno še preostale 4 ure.

Pri konstrukciji modelov SVR bomo uporabili dve jedri: linearno in radialno, modela bomo označili kot SVR-lin in SVR-rbf. Najprej moramo določiti vrednosti hiperparametrov ε in C (v primeru rbf jedra pa še parameter γ). Preizkusili bomo različne kombinacije parametrov, optimalno pa bomo izbrali na podlagi 10-kratnega prečnega preverjanja, kot je opisano v poglavju 5.2. V obeh primerih smo začetne kombinacije dobili s kartezičnim produktom vrednosti $C = [0.001, 0.01, 0.1, 1, 10, 100, 1000]$ in $\varepsilon = [0.001, 0.01, 0.1, 1, 10]$ (in v primeru radialnega jedra $\gamma = [0.001, 0.01, 0.1, 1, 10, 100, 1000]$).

Najprej si oglejmo izbiro optimalnih parametrov za primer linearne jedra. Po začetnem iskanju se je kot optimalna izkazala kombinacija $C = 1$ in $\varepsilon = 1$. Zato smo končne parametre iskali v okolici teh vrednosti, tokrat s kartezičnim produktom vrednosti $C = [0.2, 0.3, \dots, 0.9, 1, 2, 3, \dots, 9, 10]$ in $\varepsilon = [0.01, 0.1, 0.5, 1, 2, 3, 4, 5]$. Tabela 4 prikazuje optimalne vrednosti parametrov v primeru linearne jedra. V zadnjem stolpcu je dana srednja kvadratna napaka optimalnega modela, dobljena s prečnim preverjanjem (MSE_{CV}).

Jedro	C	ε	MSE_{CV}
Linearno	1	3	4891.1

Tabela 4: Optimalne vrednosti parametrov v primeru linearne jedra.

Oglejmo si izbiro optimalnih parametrov še za radialno jedro. Uvodoma se je kot optimalna izkazala kombinacija $C = 1000$ in $\gamma = 0.1$, pri ε ni bilo zaslediti opazne razlike v kakovosti modela, z izjemo tega, da najmanjša vrednost ni bila optimalna. Naredimo še eno iskanje, in sicer s kartezičnim produktom vrednosti $C = [500, 600, \dots, 900, 1000, 1100, 1200, 1300, 1400, 1500]$, $\varepsilon = [0.01, 0.1, 1, 5, 10, 15]$, in $\gamma = [0.05, 0.06, \dots, 0.09, 0.10, 0.2, 0.3, 0.4, 0.5]$. Tabela 5 prikazuje optimalne vrednosti parametrov v primeru radialnega jedra. V zadnjem stolpcu je zopet dana srednja kvadratna napaka optimalnega modela, dobljena s prečnim preverjanjem (MSE_{CV}).

Jedro	C	ε	γ	MSE_{CV}
Radialno	600	15	0.09	3314.2

Tabela 5: Optimalni parametri v primeru radialnega jedra.

Konstruirali smo torej dva SVR modela, enega z linearnim in drugega z radialnim jedrom. Vidimo, da je napaka prečnega preverjanja bistveno manjša v primeru radialnega jedra. V naslednjem poglavju bomo izdelali napovedi, pričakujemo pa, da se bo model z radialnim jedrom bolje odrezal.

5.3.3 Rezultati

V prejšnjih dveh poglavjih smo konstruirali tri modele tipa SARIMA-GARCH in dva modela tipa SVR. Sedaj pa jih bomo uporabili za napovedovanje porabe električne energije, na način opisan v poglavju 5.2. Za modele bomo uporabljali naslednje (krajše) oznake:

- GARCH- n : SARIMA-GARCH- n
- GARCH- t : SARIMA-GARCH- t
- EGARCH: SARIMA-EGARCH
- SVM-lin: model SVR z linearnim jedrom
- SVM-rbf: model SVR z radialnim jedrom

Modeli GARCH vrnejo napovedi za logaritem porabe, zato jih je za primerjavo z dejansko vrednostjo porabe potrebno z uporabo eksponentne funkcije transformirati. V nadaljevanju bomo z napovedmi mislili transformirane vrednosti napovedi. Tako napako $MAPE$ kot $RMSE$ računamo na napovedih dejanske porabe in ne na logaritmiranih napovedih.

Uvodoma se bomo posvetili napovedim za tretji teden februarja. Oglejmo si najprej napako $MAPE$ za vseh 5 modelov, ki jo za vsak dan v tednu izračunamo po formuli

$$MAPE = \frac{1}{24} \sum_{t=1}^{24} \frac{D_t - \hat{D}_t}{D_t}, \quad (5.6)$$

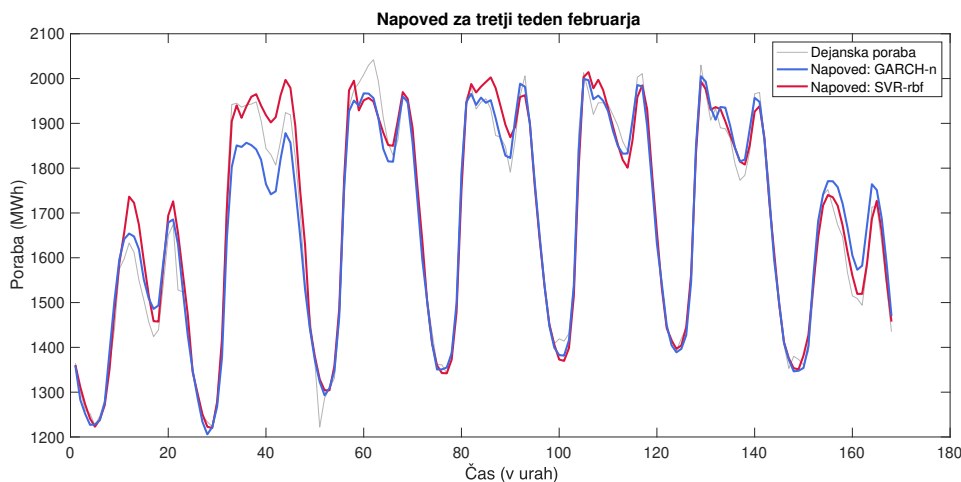
kjer sta D_t in \hat{D}_t dejanska in napovedana vrednost porabe ob času t . Vrednosti napake $MAPE$ so predstavljene v tabeli 6, kjer je v zadnjem stolpcu izračunano tudi povprečje dnevnih napak. Za vsak dan je odebeljena najmanjša napaka (najmanjša vrednost v pripadajočem stolpcu). Omenimo še, da je prvi dan v tem primeru nedelja (sobota je torej sedmi dan). V primeru modelov GARCH se napake napovedi med seboj ne razlikujejo veliko. Največjo napako lahko zasledimo za drugi dan (ponedeljek) in sedmi dan (sobota). Opazimo, da so napovedi modela SVM-rbf boljše kot napovedi modela SVM-lin, kar smo tudi pričakovali. Sicer so napovedi modela SVR-rbf primerljive z modeli GARCH, izstopajo pa ravno na drugi in sedmi

dan, ko so napake modela SVR-rbf precej manjše (ravno ta dva dneva modelom GARCH povzročata težave). Povprečna napaka je tako najmanjša v primeru modela SVM-rbf, sledijo mu modeli GARCH- n , EGARCH, GARCH- t in SVM-lin, ki precej odstopa od ostalih. Če si ogledamo napake še za vsak dan posebej, ima model SVM-rbf najboljši rezultat za 2., 6. in 7. dan. GARCH- n za 1. in 4. dan, EGARCH za 3. dan ter GARCH- t za 5. Model SVM-lin nima prav za noben dan najmanjše vrednosti napake, kar je pričakovano, saj precej odstopa.

	1	2	3	4	5	6	7	Povprečje
GARCH- n	2.21	3.25	1.79	0.88	1.13	1.32	2.86	1.92
GARCH- t	2.53	3.94	1.99	0.92	1.09	1.40	4.09	2.28
EGARCH	2.66	3.52	1.77	0.95	1.16	1.39	3.32	2.11
SVM-lin	5.75	5.12	3.13	2.21	1.88	2.43	4.57	3.58
SVM-rbf	2.87	2.09	2.13	1.68	1.71	0.97	1.46	1.84

Tabela 6: Napaka $MAPE$ (%) za tretji teden februarja.

Med modeli GARCH je tako v ospredju GARCH- n , med modeli SVR pa model SVM-rbf, zato si oglejmo njune napovedi za tretji teden februarja, prikazuje jih slika 24. Kot smo že opazili, je na podlagi napake $MAPE$ napoved modela SVM-rbf boljša za drugi in sedmi dan, sicer pa sta si modela precej blizu. Za drugi dan (ponedeljek) vidimo, da model GARCH- n podceni dejansko vrednost porabe.



Slika 24: Napoved modelov GARCH- n in SVM-rbf za tretji teden februarja.

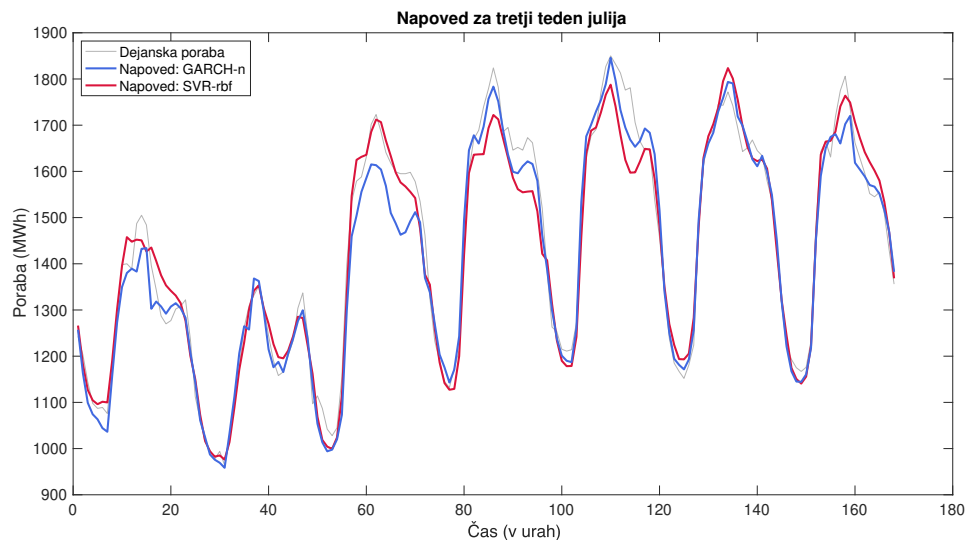
Oglejmo si še napovedi za tretji teden julija. Tabela 7 prikazuje napako $MAPE$, izračunano po formuli (5.6) za vsak dan v tednu. Tokrat je prvi dan sobota. Napovedi modelov GARCH so si zopet blizu, najslabša napoved je ponovno za ponedeljek (tretji dan). V primeru napovedi za soboto in nedeljo (prvi in drugi dan), pa imajo napovedi tokrat dokaj majhno napako. Kot v primeru napovedi za februar so razlike

med modeloma SVR-lin in SVR-rbf večje v primerjavi z modeli GARCH. V tem primeru ima najmanjšo povprečno napako model GARCH- n , vendar je zelo blizu tudi model SVM-rbf. Sledijo jima še modeli EGARCH, GARCH- t in SVM-lin. Podobno kot prej model SVM-lin zelo odstopa od ostalih, zlasti v napovedi za ponedeljek (tretji dan). Preostali štirje modeli pa so primerljivi, noben ne izstopa, razen v primeru napovedi za tretji dan, ko ima model SVM-rbf bistveno manjšo napako od preostalih modelov (več kot dvakrat manjša od druge najboljše).

	1	2	3	4	5	6	7	Povprečje
GARCH- n	2.64	1.43	5.21	2.04	2.35	1.29	1.95	2.42
GARCH- t	2.29	1.57	5.85	1.84	2.31	1.64	1.88	2.48
EGARCH	2.19	1.41	5.29	1.43	2.73	2.43	2.46	2.56
SVM-lin	3.50	3.77	10.14	3.41	6.74	2.49	3.12	4.74
SVM-rbf	2.76	1.39	2.60	3.47	3.22	1.85	1.91	2.46

Tabela 7: Napaka $MAPE$ (%) za tretji teden julija.

V ospredju sta torej zopet modela GARCH- n in SVM-rbf, zato si oglejmo njune napovedi za tretji teden julija, prikazuje ju slika 25. Kot v primeru za februar opazimo, da model GARCH- n podceni napoved za ponedeljek (tretji dan). Modela pa sta sicer primerljiva.



Slika 25: Napoved modelov GARCH- n in SVM-rbf za tretji teden julija.

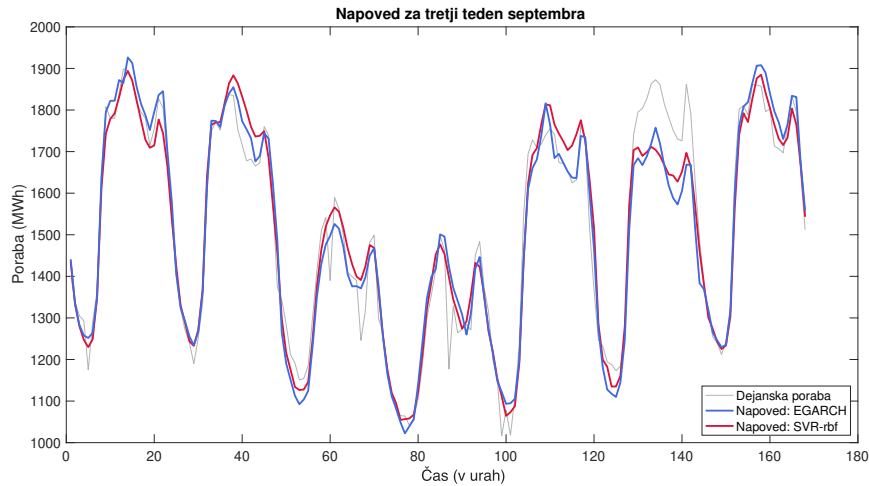
Nazadnje si oglejmo še napovedi za tretji teden septembra. Tabela 8 prikazuje napake $MAPE$, izračunane po formuli (5.6) za vsak dan v tednu. Tokrat je prvi dan četrtek. Razlike med napovedi modelov GARCH so ponovno majhne, najslabša napoved pa je tokrat za šesti dan, kar predstavlja torek. V primerjavi z napovedmi

za februar in julij, modeli tokrat bolje napovejo porabo za ponedeljek. Ponovno opazimo večjo razliko med napovedmi modela SVR-lin in SVR-rbf. Se pa tokrat model SVR-lin bolje izkaže pri napovedi za torek (šesti dan), ki se izkaže kot problematična za model SVR-rbf in vse tri GARCH modele. Za ta dan ima celo najmanjšo napako, razlika do druge najboljše napovedi je kar precejšnja, skoraj 1 %. Model z najmanjšo povprečno napako je tokrat EGARCH, model SVM-rbf je na drugem mestu. Sledijo jima GARCH- t , GARCH- n in SVM-lin. Kot v vseh primerih do zdaj je model SVM-lin na zadnjem mestu.

	1	2	3	4	5	6	7	Povprečje
GARCH- n	1.91	1.82	4.38	3.05	3.86	6.31	1.76	3.16
GARCH- t	1.86	1.94	3.95	2.82	3.39	5.34	1.61	2.99
EGARCH	2.28	1.53	3.76	2.93	3.85	4.02	1.58	2.85
SVM-lin	2.56	3.02	5.34	5.41	7.10	3.19	3.19	4.26
SVM-rbf	1.66	2.65	3.53	2.45	3.79	4.96	1.54	2.94

Tabela 8: Napaka $MAPE$ (%) za tretji teden septembra.

Med modeloma SVR je zopet najboljši SVR-rbf, med modeli GARCH pa tokrat model EGARCH, zato si oglejmo njuni napovedi za tretji teden septembra, ki ju prikazuje slika 26. Modela sta si dokaj enakovredna, oba pa podcenita dejansko vrednost porabe za šesti dan (torek).



Slika 26: Napoved modelov EGARCH in SVM-rbf za tretji teden septembra.

Čeprav so si modeli (z izjemo SVR-lin) dokaj enakovredni, pa bi za najboljšega na podlagi izbranih treh tednov izbrali model SVR-rbf. Vseeno pa si oglejmo še napako $MAPE$, izračunano za tretji teden vseh mesecev v letu 2020, in sicer po formuli

$$MAPE = \frac{1}{168} \sum_{t=1}^{168} \frac{D_t - \hat{D}_t}{D_t}, \quad (5.7)$$

kjer sta D_t in \hat{D}_t dejanska in napovedana vrednost porabe ob času t . Enak rezultat bi dobili tudi, če bi za vsak dan v tednu izračunali napako $MAPE$ po formuli (5.6) in potem vrednosti povprečili. Izračunane napake prikazuje tabela 9, za vsak mesec pa je odebljena najmanjša vrednost napake. Najboljša napoved je za mesec februar in model SVM-rbf in sicer 1.85 %, najslabša pa za mesec december in model SVM-lin in sicer 7.70 %. Napake se v povprečju gibljejo okrog 3 %. V dveh primerih ima najmanjšo vrednost napake model GARCH- n , v štirih model EGARCH, enako model GARCH- t in v treh mesecih model SVM-rbf.

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Jan	3.68	2.47	2.57	2.61	2.25
Feb	3.58	1.85	1.92	2.28	2.11
Mar	5.15	3.83	2.97	3.12	2.76
Apr	5.42	4.01	4.41	4.26	4.35
Maj	4.95	4.73	2.49	2.29	2.24
Jun	5.53	3.49	3.05	3.02	3.30
Jul	4.74	2.46	2.42	2.48	2.56
Avg	5.73	4.03	2.50	2.38	3.10
Sep	4.26	2.94	3.16	2.99	2.85
Okt	3.79	1.92	3.14	2.80	2.87
Nov	3.62	2.87	2.35	2.32	2.54
Dec	7.70	4.53	2.84	3.08	3.25

Tabela 9: Napaka $MAPE$ (%) za tretji teden vseh mesecev v letu 2020.

Tabela 10 prikazuje povprečno napako $MAPE$ za 12 analiziranih tednov. Najnižjo vrednost ima model GARCH- t , zelo blizu sta modela GARCH- n in EGARCH. Čeprav ima model SVM-rbf najboljšo napoved v četrtini primerov, pa je v povprečju šele na četrtem mestu, torej za modelom GARCH- t . Razlog je v precej slabši napovedi za mesece marec, maj, avgust in december, ko se napaka razlikuje za 1 % ali več od modelov tipa GARCH. Model SVM-rbf je torej velikokrat najboljši, ko pa ni najboljši, je precej daleč stran.

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Povprečna $MAPE$	4.85	3.26	2.82	2.80	2.85

Tabela 10: Povprečna napaka $MAPE$ (%).

Tabela 11 prikazuje napako $RMSE$ izračunano za 12 analiziranih tednov po formuli

$$RMSE = \sqrt{\frac{1}{168} \sum_{t=1}^{168} (D_t - \hat{D}_t)^2}, \quad (5.8)$$

kjer sta D_t in \hat{D}_t dejanska in napovedana vrednost porabe ob času t . Najmanjše vrednosti napake dobimo podobno kot prej.

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Jan	99.90	66.05	65.80	70.40	60.33
Feb	81.71	41.80	43.08	51.35	46.75
Mar	98.10	77.06	59.63	65.11	55.75
Apr	110.23	72.54	95.29	92.16	96.32
Maj	99.31	95.43	43.68	41.08	40.87
Jun	106.46	64.28	54.78	56.10	62.60
Jul	94.97	47.58	45.99	47.56	49.02
Avg	115.88	91.51	50.36	47.63	59.70
Sep	87.71	61.61	63.87	60.30	55.76
Okt	83.07	39.73	69.83	59.79	58.80
Nov	90.72	63.00	51.94	50.73	54.49
Dec	147.38	93.17	63.87	67.63	71.31

Tabela 11: Napaka $RMSE$ za tretji teden vseh mesecev v letu 2020.

Tabela 12 prikazuje povprečno napako $RMSE$ na podlagi dvanajstih analiziranih tednov. Najmanjšo vrednost ima model GARCH- n , pri modelu SVM-lin pa vidimo zelo veliko odstopanje.

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Povprečna $RMSE$	101.29	67.81	59.01	59.15	59.31

Tabela 12: Povprečna napaka $RMSE$.

Zaključimo lahko, da med modeli ni izrazitega favorita, z izjemo modela SVR-lin, ki močno odstopa in bi ga trdno uvrstili na zadnje mesto. Izbira preiskovanega tedna pa ima veliko vlogo pri tem, kateri izmed preostalih štirih modelov bo takrat najboljši.

5.4 Napovedovanje cene električne energije

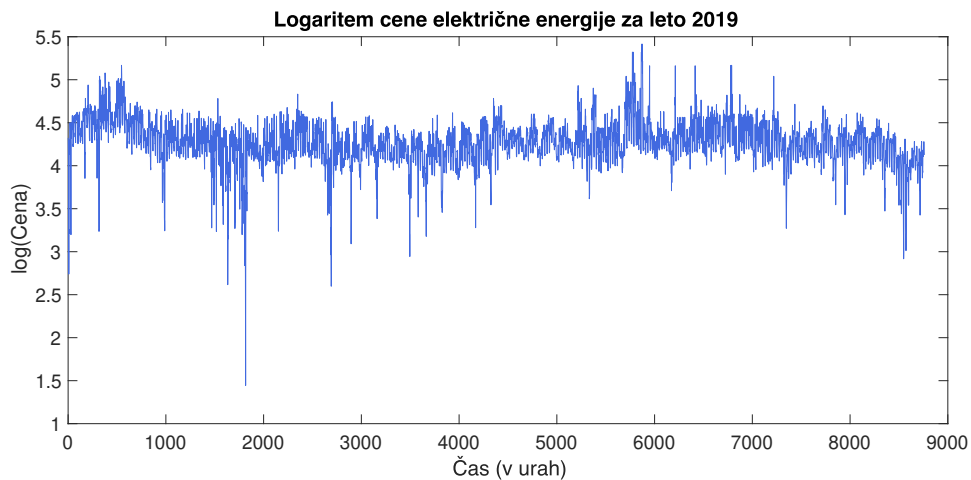
Sedaj si bomo ogledali še konstrukcijo modelov za napovedovanje cene električne energije. Kot v primeru porabe, si bomo najprej ogledali modele ARMA-GARCH, potem pa še modele SVR. Konstrukcija obeh tipov modelov bo potekala na podoben način kot v primeru porabe električne energije, seveda z nekaj spremembami. Na koncu bomo primerjali napovedi ocenjenih modelov. Časovno vrsto cene električne energije bomo označili s P_t .

5.4.1 ARMA-GARCH

Za konstrukcijo modela ARMA-GARCH bomo zopet uporabili Box-Jenkinsovo metodologijo. Tudi v primeru cene električne energije bomo izbiro modela naredili za podatke med 1. 1. 2019 in 31. 12. 2019. Tako bomo lahko opažene vzorce primerjali med seboj.

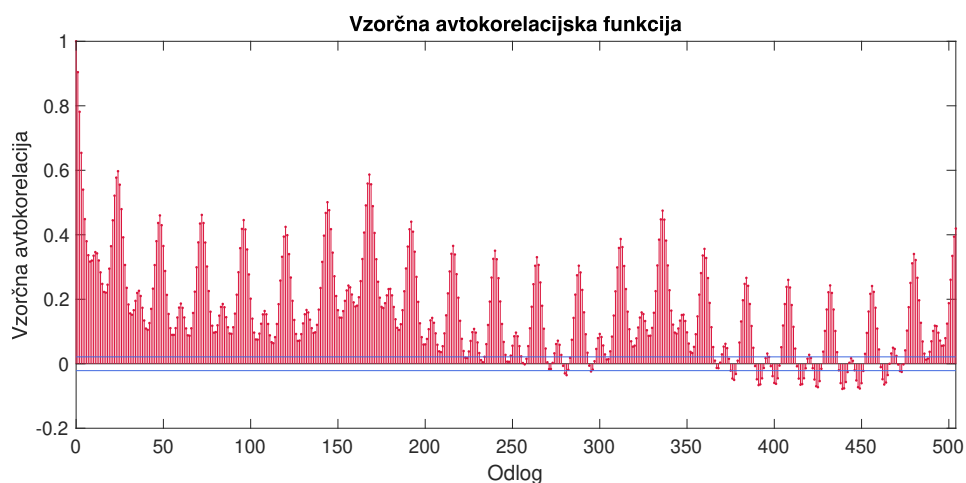
Najprej moramo izbrati razred modelov. Tako kot v primeru porabe, lahko opazimo, da so podatki za ceno električne energije visokofrekvenčni, povprečje in varianca nista konstantna, zasledimo pa tudi večkratno sezonskost (dnevna, tedenska). Zopet bomo torej izbrali model SARIMA-GARCH.

V naslednjem koraku moramo identificirati točno določen model, ki je odvisen od danih podatkov. Te bomo najprej logaritmizirali, saj so urne cene zelo volatilne (bolj kot poraba). Ker pa lahko cena zavzame tudi negativne vrednosti, moramo pred tem še odšteti najmanjšo vrednost cene in prišteti 1. Na ta način bodo vse logaritmizirane vrednosti večje ali enake 0. Preveriti želimo, ali je logaritmizirana časovna vrsta cene električne energije stacionarna. Oglejmo si graf časovne vrste $\ln P_t$, ki je prikazan na sliki 27.



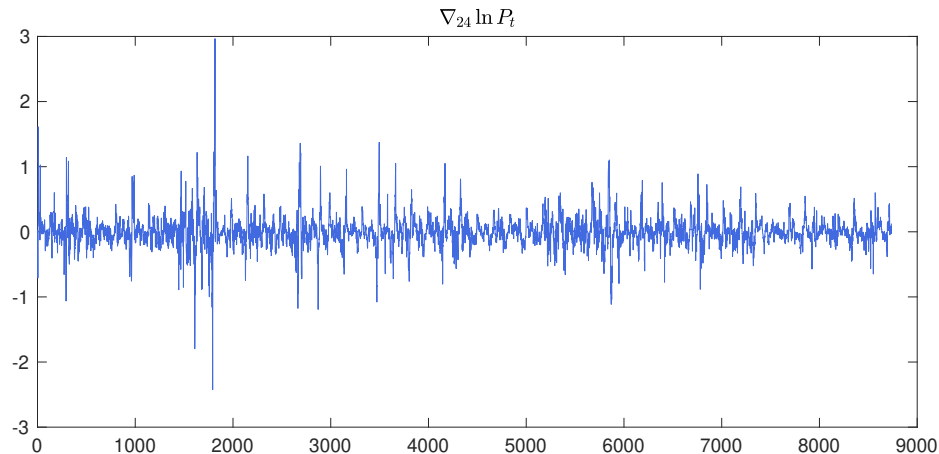
Slika 27: Logaritemska vrednost cene električne energije za leto 2019.

Časovna vrsta očitno ne deluje stacionarno, oglejmo pa si še vzorčno ACF, ki jo prikazuje slika 28. Na ta način bomo lahko zasledili kakšno diferenciranje je potrebno.



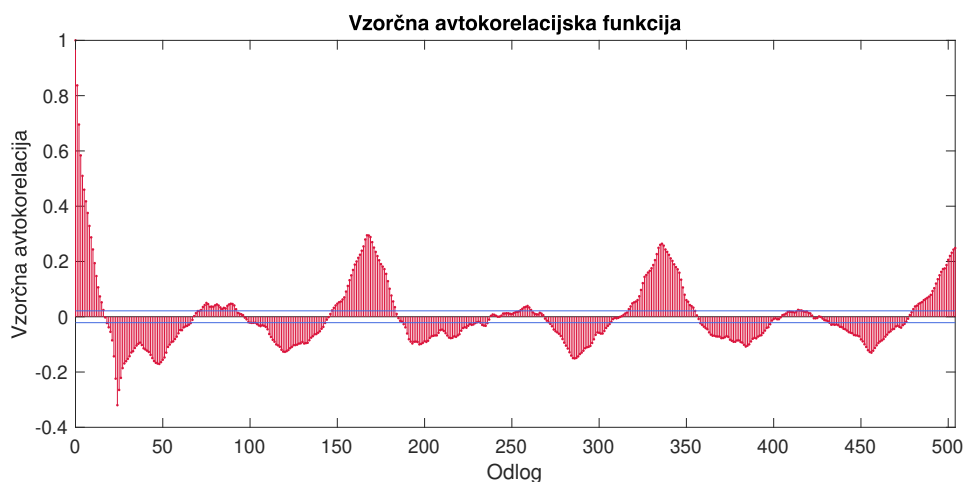
Slika 28: Vzorčna avtokorelacijska funkcija časovne vrste $\ln P_t$.

Iz zgornje vzorčne ACF res lahko potrdimo, da so tako kot podatki za porabo tudi podatki za ceno nestacionarni, vzorčna ACF namreč ne izveni po nekaj odlogih. Pričakovano tudi tu opazimo močno sezonsko komponentno, perioda je enaka 24 (ur). Zopet bo potrebno sezonsko diferenciranje, kot v primeru porabe na časovni vrsti $\ln P_t$ uporabimo operator ∇_{24} . Slika 29 prikazuje graf logaritmirane časovne vrste, ki je sezonsko diferencirana do prvega reda s periodo 24.



Slika 29: Logaritmirana časovna vrsta cene električne energije, ki je sezonsko diferencirana do prvega reda s periodo 24.

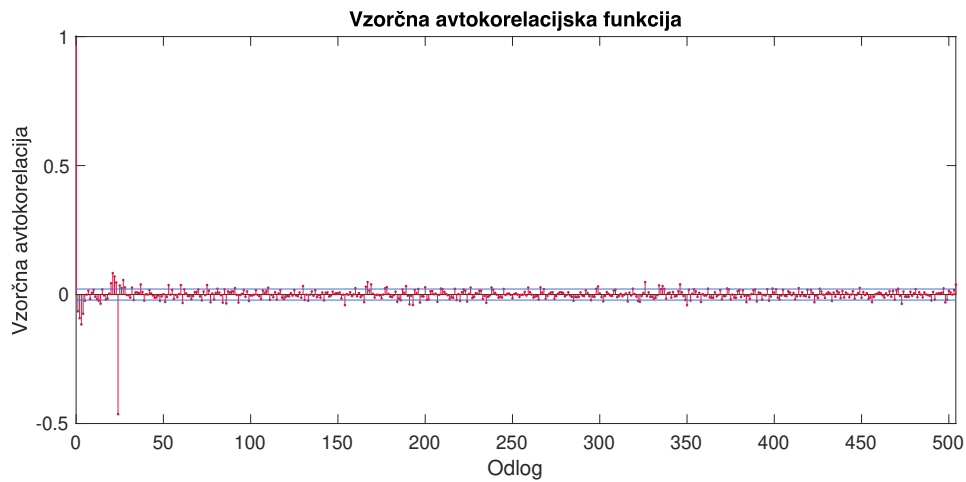
Zgornja časovna vrsta res deluje bolj stacionarno kot prvotna, vendar o stacionarnosti še ne moremo sklepati. Oglejmo si vzorčno ACF časovne vrste $\nabla_{24} \ln P_t$.



Slika 30: Vzorčna avtokorelacijska funkcija časovne vrste $\nabla_{24} \ln P_t$, ki je prikazana na sliki 29.

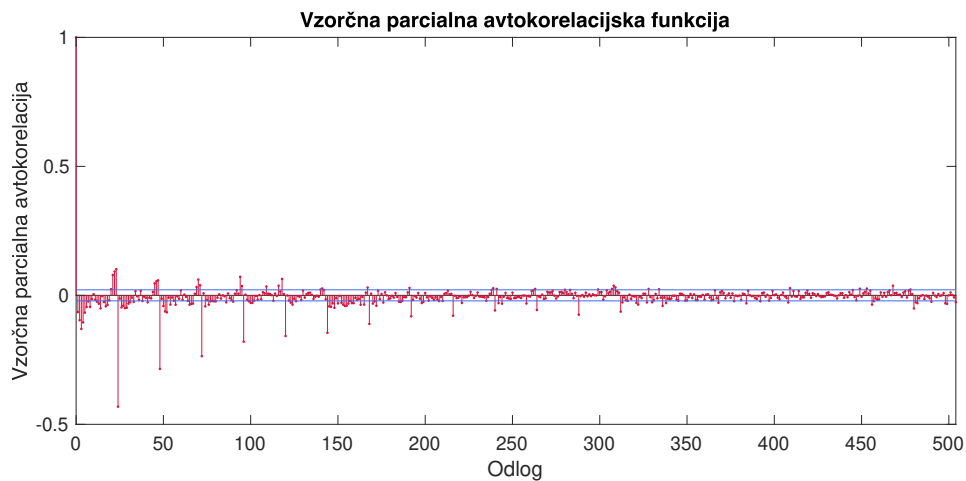
Graf vzorčne ACF nakazuje, da časovna vrsta tudi po diferenciranju še ni stacionarna. Zato na diferencirani časovni vrsti uporabi operator ∇ . Slika 31 prikazuje

vzorčno ACF časovne vrste, ki je sezonsko diferencirana do prvega reda s periodo 24 in navadno diferencirana do prvega reda, tj. $\nabla\nabla_{24} \ln P_t$.



Slika 31: Vzorčna avtokorelacijska funkcija časovne vrste $\nabla\nabla_{24} \ln P_t$.

Časovna vrsta sedaj izgleda stacionarno in je primerna za model ARMA. Pri nadaljnji identifikaciji modela bomo potrebovali še vzorčno PACF časovne vrste $\nabla\nabla_{24} \ln P_t$, prikazana je na sliki 32.



Slika 32: Vzorčna parcialna avtokorelacijska funkcija časovne vrste $\nabla\nabla_{24} \ln P_t$.

Pri konstrukciji modela si bomo ponovno pomagali z grafoma vzorčne ACF in vzorčne PACF za časovno vrsto $\nabla\nabla_{24} \ln P_t$. Nato bomo s pomočjo AIC kriterija izbrali najboljše parametre modela. Zopet bomo konstruirali sezonski model ARIMA oblike

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D \ln^* P_t = \theta(B)\Theta(B^s)Z_t. \quad (5.9)$$

Pri tem so $\phi(z) = 1 - \phi_1(z) - \dots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1(z) - \dots - \Phi_P z^P$, $\theta(z) = 1 + \theta_1(z) + \dots + \theta_q z^q$ in $\Theta(z) = 1 + \Theta_1(z) + \dots + \Theta_Q z^Q$, P_t pa je cena ob času t . Z_t zopet ustreza enemu izmed obravnavanih modelov GARCH (smo prej ugotovili, da varianca ni konstantna). S predstavlja periodo sezonskosti. Velja $\ln^* P_t = \ln(P_t - \min(P_t) + 1)$. Osnovna ideja modela je tako enaka modelu za porabo, z izjemo spremenjene logaritemske transformacije (ker so cene lahko negativne).

Kot v primeru porabe je perioda S enaka 24 (en dan), prav tako sta parametra d in D enaka 1 (smo zasledili iz predhodne analize vzorčne ACF). Ugotovili smo namreč, da je časovna vrsta

$$Y_t := \nabla \nabla_{24} \ln^* P_t = (1 - B)(1 - B^{24}) \ln^* P_t \quad (5.10)$$

stacionarna. Model lahko zopet poenostavimo v

$$\phi(B)\Phi(B^{24})(1 - B)(1 - B^{24}) \ln^* P_t = \theta(B)\Theta(B^{24})Z_t. \quad (5.11)$$

Določiti moramo še 4 parametre, parametra P in Q za sezonski AR in MA del ter parametra p in q za nesezonski AR in MA del. Na podlagi grafa vzorčne PACF (slika 32) bomo lahko sklepali o obliki AR dela, na podlagi grafa vzorčne ACF (slika 31) pa o obliki MA dela. Kot v primeru porabe, tudi zdaj ni zaslediti preproste oblike.

Najprej bomo določili sezonska parametra P in Q , osredotočili se bomo torej na korelacije pri odlogih, ki so večkratniki periode S . Na grafu vzorčne PACF vidimo visoke korelacije (konice na sliki) pri odlogih 24, 48, 72, 96, 120, 144 in 168 (AR del). Najopaznejša je pri periodi 24, potem pa se počasi zmanjšujejo. Na grafu vzorčne ACF ponovno vidimo bistveno manj izstopajočih avtokorelacij, največja pa je ravno ob periodi (MA del). Parametra P in Q bosta tako ponovno enaka vsaj 1, P bo verjetno malo večji, saj smo zasledili več pomembnih avtokorelacij na grafu vzorčne PACF, vendar se bomo na koncu odločili s pomočjo kriterija AIC.

Nazadnje določimo še nesezonska parametra p in q . Na grafu vzorčne PACF vidimo, da ima značilno korelacijo prvih nekaj ur. Prav tako pa so značilne korelacije pri odlogih 21, 22, 23, 46, 47, 70, 71, 94 in 95, torej v urah pred večkratniki periode. Oboje spada v AR del. Vidimo, da so pred vsakim večkratnikom periode značilni po dve ali tri ure, v primeru porabe pa je bila značilna samo ena ura nazaj. Tudi na grafu vzorčne ACF opazimo korelacije v prvih nekaj urah, pa tudi ob urah pred enim dnem, oboje spada v MA del.

Z zgornjimi ugotovitvami nam vzorčni ACF in PACF pomagata zaznati, katere odloge bi bilo smiselno vključiti. Kot v primeru porabe bomo zagotovo vključili prvih nekaj ur, za izbiro naslednjih odlogov pa se bomo odločili s pomočjo kriterija AIC. Začeli bomo s konstrukcijo 13 modelov, nekateri bodo bolj, drugi pa manj preprosti (v smislu, koliko parametrov je potrebno oceniti). Kot v primeru porabe bomo Z_t uvodoma modelirali kot proces GARCH(1,1). Najpreprostejši model bo oblike

$$\phi(B)\Phi(B^{24})(1 - B)(1 - B^{24}) \ln^* P_t = \theta(B)\Theta(B^{24})Z_t, \quad (5.12)$$

kjer so

$$\begin{aligned}\phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \phi_4 z^4, \\ \Phi(z) &= 1 - \Phi_1 z, \\ \theta(z) &= 1 + \theta_1 z + \theta_2 z^2 + \theta_3 z^3 \quad \text{in} \\ \Theta(z) &= 1 + \Theta_1 z.\end{aligned}$$

Parameter p je torej enak 4, parameter q pa je enak 3 (noben vmesni koeficient torej ni ničeln). Parametra P in Q sta oba enaka 1. Zgornji model bomo, enako kot v primeru porabe električne energije, krajše zapisali kot $M1 = \{\text{AR} = [1-4], \text{MA} = [1-3], S = 24, \text{SAR} = 1, \text{SMA} = 1\}$. Nato bomo postopoma dodajali odloge, ki smo jih zasledili pri grafih vzorčne ACF in PACF. Najbolj kompleksen model bo imel tako obliko $M13 = \{\text{AR} = [1-5, 21-23, 46-47], \text{MA} = [1-4, 21-23], \text{SAR} = [1-2], \text{SMA} = 1\}$.

Ocenimo koeficiente modela in izračunajmo kriterij AIC. Tabela 13 prikazuje vrednost kriterija AIC za 13 izbranih modelov. Vrednosti zopet skaliramo, tako da jih delimo z velikostjo vzorca. Vrstica, ki pripada modelu z najmanjšo vrednostjo kriterija AIC, je odebeljena. Najnižjo vrednost ima model številka 13, ki je najbolj kompleksen. Zato se odločimo konstruirati modele, kjer vključimo še več odlogov.

Model	AR	MA	SAR	SMA	AIC
1	1-4	1-3	1	1	-2.2814
2	1-4	1-4	1	1	-2.2812
3	1-5	1-3	1	1	-2.2812
4	1-5	1-4	1	1	-2.2897
5	1-5,21-23	1-4	1	1	-2.2985
6	1-5,21-23	1-4,21-23	1	1	-2.2989
7	1-4	1-3	1-2	1	-2.2817
8	1-4	1-4	1-2	1	-2.2816
9	1-5	1-3	1-2	1	-2.2816
10	1-5	1-4	1-2	1	-2.2928
11	1-5,21-23	1-4	1-2	1	-2.2961
12	1-5,21-23	1-4,21-23	1-2	1	-2.2991
13	1-5,21-23,46-47	1-4,21-23	1-2	1	-2.3019

Tabela 13: Vrednost kriterija AIC za 13 izbranih modelov.

Tokrat konstruiramo 5 modelov. Najpreprostejši ima obliko $M1 = \{\text{AR} = [1-5, 21-23, 46-47, 70-71], \text{MA} = [1-4, 21-23], S = 24, \text{SAR} = [1-3], \text{SMA} = 1\}$, najkompleksnejši pa $M5 = \{\text{AR} = [1-5, 21-23, 46-47, 70-71, 94-95], \text{MA} = [1-4, 21-23], S = 24, \text{SAR} = [1-7, 14, 21], \text{SMA} = 1\}$. Z dodajanjem odlogov se zopet precej poveča čas ocenjevanja parametrov, saj to pri modelu M5 traja že več kot pol ure. Tabela 14 prikazuje vrednost kriterija AIC za 5 izbranih modelov.

Model	AR	MA	SAR	SMA	AIC
1	1-5,21-23,46-47,70-71	1-4,21-23	1-3	1	-2.3055
2	1-5,21-23,46-47,70-71,94-95	1-4,21-23	1-4	1	-2.2998
3	1-5,21-23,46-47,70-71,94-95	1-4,21-23	1-5	1	-2.3089
4	1-5,21-23,46-47,70-71,94-95	1-4,21-23	1-6	1	-2.3009
5	1-5,21-23,46-47,70-71,94-95	1-4,21-23	1-7,14,21	1	-2.3177

Tabela 14: Vrednost kriterija AIC za 5 izbranih modelov.

Najnižjo vrednost ima model številka 5, torej zadnji in najbolj kompleksen. To je model oblike

$$\phi(B)\Phi(B^{24})(1-B)(1-B^{24})\ln^* P_t = \theta(B)\Theta(B^{24})Z_t, \quad (5.13)$$

kjer so

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \phi_4 z^4 - \phi_5 z^5 - \phi_{21} z^{21} - \phi_{22} z^{22} - \phi_{23} z^{23} \\ &\quad - \phi_{46} z^{46} - \phi_{47} z^{47} - \phi_{70} z^{70} - \phi_{71} z^{71} - \phi_{94} z^{94} - \phi_{95} z^{95}, \\ \Phi(z) &= 1 - \Phi_1 z - \Phi_2 z^2 - \Phi_3 z^3 - \Phi_4 z^4 - \Phi_5 z^5 - \Phi_6 z^6 - \Phi_7 z^7 - \Phi_{14} z^{14} - \Phi_{21} z^{21}, \\ \theta(z) &= 1 + \theta_1 z + \theta_2 z^2 + \theta_3 z^3 + \theta_4 z^4 + \theta_{21} z^{21} + \theta_{22} z^{22} + \theta_{23} z^{23} \quad \text{in} \\ \Theta(z) &= 1 + \Theta_1 z. \end{aligned}$$

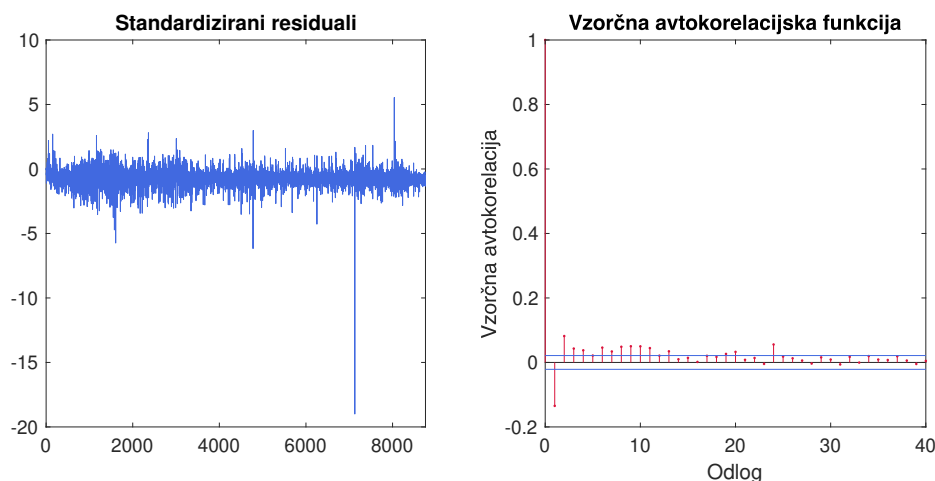
V primerjavi z modelom za porabo vidimo, da je tu S(AR) del bolj kompleksen, (S)MA del pa manj. Izbrali smo obliko sezonskega modela ARIMA, enako kot v primeru porabe pa moramo določiti še model za Z_t , podatki namreč kažejo, da varianca ni konstantna. Preizkusili bomo tri modele, in sicer GARCH- n , GARCH- t in EGARCH, za vsakega pa 6 kombinacij parametrov (p, q) : (1,1), (1,2), (2,1), (1,3), (3,1) in (2,2). Za vsak model bomo izbrali tisto kombinacijo parametrov z najnižjo vrednostjo kriterija AIC.

Med modeli GARCH- n se je za najboljšega izkazal model GARCH- $n(3,1)$, torej enako kot v primeru porabe. Vrednost kriterija AIC je enaka -2.3248 . Med modeli GARCH- t ima najnižjo vrednost kriterija AIC model GARCH- $t(2,2)$, ta je enaka -2.5709 (spet enako kot v primeru porabe). Med modeli EGARCH pa se je za najboljšega izkazal model EGARCH(1,2). Vrednost kriterija AIC je enaka -2.3509 . Najmanjšo vrednost kriterija AIC ima torej model GARCH- $t(2,2)$, enako kot v primeru porabe.

Izbrali smo tri modele, v nadaljevanju pa moramo preveriti, če so primerni za napovedovanje. Kot v primeru porabe, si bomo ogledali standardizirane residue modelov, residuali morajo namreč tvoriti beli šum. Nekoreliranost bomo preverili z Ljung-Boxovim testom, heteroskedastičnost pa z Ljung-Boxovim testom na kvadratih residualov.

Najprej analizirajmo model SARIMA-GARCH- $n(3,1)$, ki smo ga izbrali na podlagi kriterija AIC. Naredimo Ljung-Boxov test na standardiziranih residualih za odloge 5, 10, 15 in 20. Pripadajoče p -vrednosti so v vseh primerih manjše od 0.05, torej pri stopnji značilnosti $\alpha = 0.05$ zavrnilo ničelno hipotezo, da so residuali nekorelirani. Preverimo še preostalih pet kombinacij, pri vseh zavrnilo ničelno hipotezo.

Izberemo pa tisti model, kjer so p -vrednosti največje, to je SARIMA-GARCH- $n(2,1)$. Preverimo še heteroskedastičnost residualov. Naredimo Ljung-Boxov test na standardiziranih kvadratih residualov, p -vrednosti so krepko večje od 0.05. V residualih torej ni zaznati heteroskedastičnosti. Oglejmo si še graf standardiziranih residualov in vzorčno ACF, prikazuje ju slika 33. Residuali res nekoliko odstopajo od belega šuma, vendar pa se vseeno odločimo, da bomo ta model uporabili za napovedovanje. V praktičnih primerih je namreč zelo težko najti model, ki bo popolno opisal naše podatke.



Slika 33: Levo: Standardizirani residuali modela SARIMA-GARCH- $n(2,1)$. Desno: pripadajoča vzorčna avtokorelacijska funkcija.

Enak sklep naredimo tudi v primerih modelov GARCH- t in EGARCH. Pri vseh kombinacijah zavrnemo ničelno hipotezo, da so residuali nekorelirani, odločimo pa se za modela SARIMA-GARCH- $t(1,2)$ in SARIMA-EGARCH(1,1). V nobenem primeru torej ne moremo uradno potrditi modela, vseeno pa se zdi model dovolj dober, da ga bomo uporabili za napovedovanje.

Z uporabo treh izbranih modelov bomo v nadaljevanju izdelali napovedi. Postopek napovedovanja je opisan v poglavju 5.2. Še prej pa bomo konstruirali modele SVR, na koncu pa vse medsebojno primerjali. Rezultati so predstavljeni v poglavju 5.4.3.

5.4.2 SVR

Kot v primeru porabe bomo konstruirali avtoregresijski SVR model. Prihodnjo ceno bomo izračunali z uporabo preteklih vrednosti cene kot napovednih spremenljivk, izbrati pa moramo, katere pretekle vrednosti bomo uporabili. Po zgledu modeliranja porabe izberemo odloge, ki smo jih kot pomembne zasledili pri analizi vzorčne ACF in PACF v prejšnjem poglavju, ti so 1, 2, 3, 4, 5, 21, 22, 23, 24, 46, 47, 48, 70, 71, 72, 94, 95, 96, 120, 144, 168, 336, 504. Na enak način kot pri porabi konstruiramo tabelo. Vrstico v tabeli bo predstavljal par (x_t, y_t) , $t = 1, \dots, 8760$, vsaki urni vrednosti cene

pripada ena vrstica tabele. Pri tem x_t predstavlja vektor napovednih spremenljivk

$$x_t = ([x_t]_1, [x_t]_2, [x_t]_3, [x_t]_4, [x_t]_5, [x_t]_{21}, [x_t]_{22}, [x_t]_{23}, [x_t]_{24}, [x_t]_{46}, [x_t]_{47}, [x_t]_{48}, \\ [x_t]_{70}, [x_t]_{71}, [x_t]_{72}, [x_t]_{94}, [x_t]_{95}, [x_t]_{96}, [x_t]_{120}, [x_t]_{144}, [x_t]_{168}, [x_t]_{336}, [x_t]_{504}),$$

kjer je $[x_t]_i = P_{t-i}$, imamo torej 23 napovednih spremenljivk (pretekle cene). Vrednost ciljne spremenljivke y_t je enaka P_t , torej vrednost cene ob času t . Zapišemo lahko

$$P_t = f(P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}, P_{t-5}, P_{t-21}, P_{t-22}, P_{t-23}, P_{t-24}, P_{t-46}, P_{t-47}, P_{t-48}, \\ P_{t-70}, P_{t-71}, P_{t-72}, P_{t-94}, P_{t-95}, P_{t-96}, P_{t-120}, P_{t-144}, P_{t-168}, P_{t-336}, P_{t-504}),$$

cena ob času t je torej funkcija preteklih vrednosti cene.

Napovedi bomo naredili za 24 ur naprej, zgornji model pa za napoved ob času t potrebuje preteklih 5 vrednosti, enako kot v primeru porabe. Vrednosti, ki ob nekem trenutku niso na voljo, torej nadomestimo z napovedmi. Pri konstrukciji SVR modelov bomo ponovno uporabili dve jedri: linearno in radialno. Najprej moramo določiti vrednosti hiperparametrov ε in C (ter γ v primeru radialnega jedra). Končno izbiro parametrov bomo naredili na podlagi 10-kratnega prečnega preverjanja kot je opisano v poglavju 5.2. Enako kot v primeru porabe, smo v obeh primerih iskanje začeli s kartezičnim produktom vrednosti $C = [0.001, 0.01, 0.1, 1, 10, 100, 1000]$ in $\varepsilon = [0.001, 0.01, 0.1, 1, 10]$ (ter $\gamma = [0.001, 0.01, 0.1, 1, 10, 100, 1000]$ v primeru radialnega jedra).

Najprej izberimo optimalne parametre za primer linearnega jedra. Po začetnem iskanju se je kot optimalna izkazala kombinacija $C = 0.1$ in $\varepsilon = 1$. Nadaljujemo s podrobnejšim iskanjem v okolici teh vrednosti, tokrat s kartezičnim produktom vrednosti $C = [0.01, 0.02, \dots, 0.09, 0.1, 0.2, \dots, 0.9, 1]$ in $\varepsilon = [0.5, 0.75, \dots, 4.75, 5]$. Tabela 15 prikazuje optimalne vrednosti parametrov v primeru linearnega jedra. V zadnjem stolpcu lahko vidimo srednjo kvadratno napako optimalnega modela, dobljeno s prečnim preverjanjem (MSE_{CV}).

Jedro	C	ε	MSE_{CV}
linearno	0.1	3.25	49.8072

Tabela 15: Optimalne vrednosti parametrov v primeru linearnega jedra.

Oglejmo si izbiro optimalnih parametrov še za radialno jedro. Uvodoma se je kot optimalna izkazala kombinacija $C = 100$ in $\gamma = 0.01$, pri ε ni bilo zaslediti opazne razlike v kakovosti modela, z izjemo tega, da največja vrednost ni bila optimalna. Podrobnejše iskanje naredimo z kartezičnim produktom vrednosti $C = [50, 60, \dots, 90, 100, 200, 300]$, $\varepsilon = [0.001, 0.01, 0.1, 1]$ in $\gamma = [0.001, 0.003, \dots, 0.019, 0.021, 0.03, 0.04, \dots, 0.09, 0.1]$.

Tabela 16 prikazuje optimalne vrednosti parametrov v primeru radialnega jedra. V zadnjem stolpcu je srednja kvadratna napaka optimalnega modela, dobljena s prečnim preverjanjem.

Jedro	C	ε	γ	MSE_{CV}
RBF	200	1	0.0130	44.7599

Tabela 16: Optimalne vrednosti parametrov v primeru radialnega jedra.

Konstruirali smo torej dva SVR modela, enega z linearnim in drugega z radialnim jedrom. Napaka prečnega preverjanja je manjša v primeru radialnega jedra, vendar pa je razlika med napakama manjša kot v primeru porabe električne energije. V naslednjem poglavju bomo na podlagi izbranih dveh modelov izdelali napovedi.

5.4.3 Rezultati

Konstruirali smo pet različnih modelov, te bomo zdaj uporabili za napovedovanje cene električne energije, na način opisan v poglavju 5.2. Za modele bomo uporabljali enake oznake kot v primeru porabe. Kot v primeru porabe modeli GARCH vrnejo napovedi za logaritem cene (oz. modifikacijo logaritemske transformacije \ln^*). Za primerjavo z dejansko vrednostjo cene moramo napovedi najprej transformirati z uporabo eksponentne funkcije. V nadaljevanju bomo z napovedmi zopet mislili transformirane vrednosti napovedi. Napaki $MAPE$ in $RMSE$ računamo na napovedih dejanske cene in ne na logaritmiranih napovedih.

Najprej se bomo osredotočili na napovedi za tretji teden februarja. Oglejmo si napako $MAPE$ za vseh 5 modelov, ki jo za vsak dan v tednu izračunamo po formuli

$$MAPE = \frac{1}{24} \sum_{t=1}^{24} \frac{P_t - \hat{P}_t}{\bar{P}}, \quad (5.14)$$

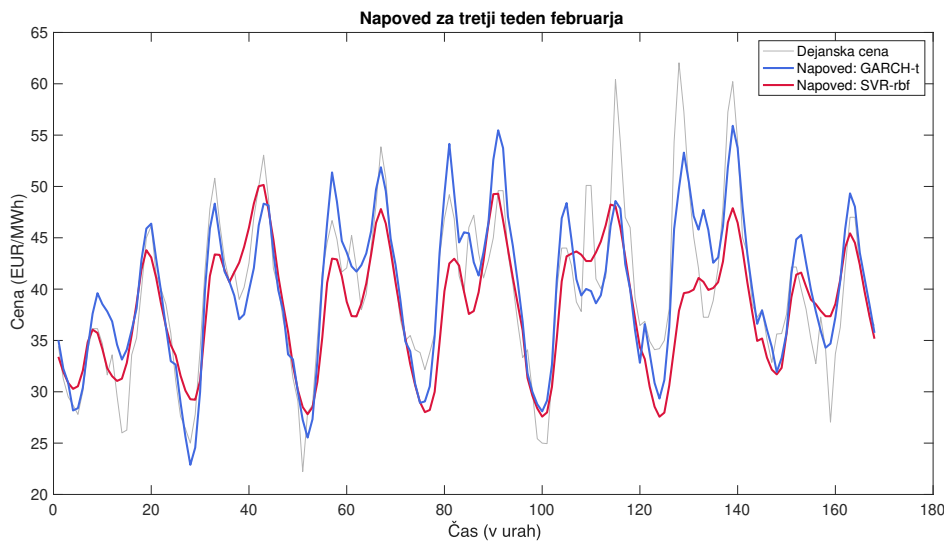
kjer sta P_t in \hat{P}_t dejanska in napovedana vrednost cene ob času t , ter $\bar{P} = \frac{1}{24} \sum_{t=1}^{24} P_t$. Uporabili smo torej alternativno formulacijo napake $MAPE$, saj so cene lahko tudi ničelne. Ker je napaka $MAPE$ izražena v %, bomo napako napovedi za ceno lahko primerjali z napako napovedi za porabo (v primeru $RMSE$ to ni mogoče, saj ima ta enoto samih podatkov). Tabela 17 prikazuje izračunane napake, izračunano je tudi povprečje napak za vse dni v tednu. Prvi dan je v tem primeru nedelja. Napovedi modelov GARCH so si zopet blizu, največja odstopanja od dejanske cene lahko zasledimo peti in šesti dan (četrtek in petek). V primeru modelov SVR lahko vidimo malo večje odstopanje med modeloma, vendar ne tako veliko kot v primeru porabe, ko se je linearno jedro izkazalo za bistveno slabše. Tudi tu je najslabša napoved za peti in šesti dan. Povprečna napaka je najmanjša v primeru modela GARCH- t , sledijo mu EGARCH, GARCH- n , SVM-rbf in SVM-lin. Linearno jedro ne odstopa tako zelo kot v primeru porabe, vseeno pa je zopet na zadnjem mestu. Če si napake ogledamo še za vsak dan posebej, ima model GARCH- t najboljši rezultat za drugi in šesti dan. EGARCH za četrty in peti dan, GARCH- n za tretji dan, SVM-rbf za prvi dan in SVM-lin za sedmi dan. Napovedi modelov GARCH so boljše med drugim in

šestim dnem, torej od ponedeljka do petka. V soboto in nedeljo (sedmi in prvi dan) pa sta rezultata modelov SVR mnogo boljša.

	1	2	3	4	5	6	7	Povprečje
GARCH- n	9.38	7.02	4.78	6.37	9.52	10.65	9.58	8.19
GARCH- t	7.11	5.19	5.36	8.23	9.34	8.20	6.91	7.19
EGARCH	9.20	7.63	4.90	6.31	9.18	10.15	9.87	8.18
SVM-lin	6.47	9.44	8.45	9.15	10.11	14.42	6.07	9.16
SVM-rbf	5.30	6.80	6.48	9.42	9.74	15.54	6.97	8.61

Tabela 17: Napaka $MAPE(\%)$ za tretji teden februarja.

Slika 34 prikazuje napovedi modelov GARCH- t in SVM-rbf za tretji teden februarja. Res opazimo, da so med ponedeljkom in petkom boljše napovedi modela GARCH- t , čez vikend pa se bolje odreže model SVM-rbf. V splošnem pa vidimo, da so napovedi nekoliko manj natančne kot v primeru porabe, to smo sklepali že na podlagi zgornjih napak $MAPE$. To se zdi kar smiselno, saj je gibanje cene veliko bolj nepredvidljivo kot gibanje porabe, tudi iz spodnje slike lahko vidimo, da vzorec ni tako konstanten.



Slika 34: Napoved modelov GARCH- t in SVM-rbf za tretji teden februarja.

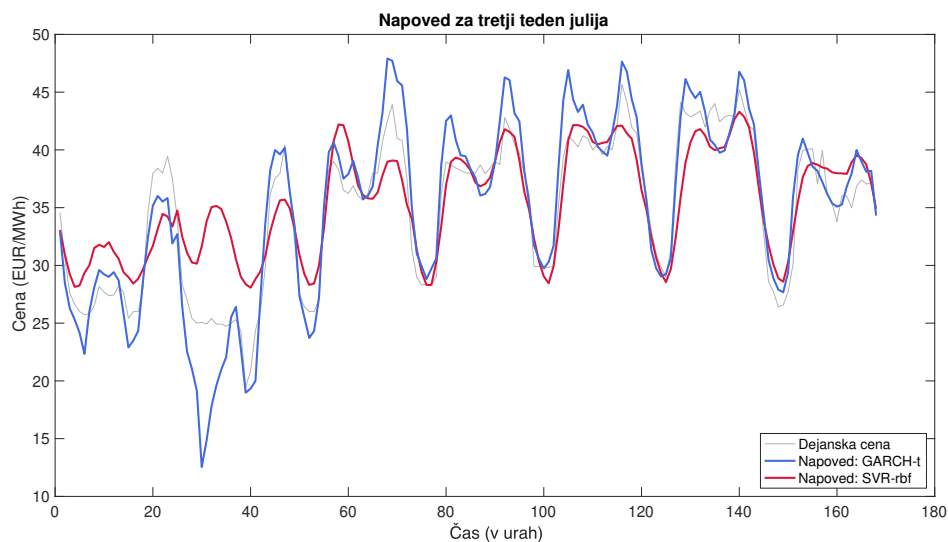
Nadaljujmo z napovedmi za tretji teden julija, tabela 18 prikazuje pripadajočo dnevno napako $MAPE$ in njeno tedensko povprečje. V tem primeru je prvi dan sobota. Tokrat si napovedi modelov GARCH niso tako blizu, od preostalih dveh izstopa model GARCH- t , ki ima bistveno manjšo napako. Tudi razlike med linearnim in radialnim jedrom so večje kot v primeru za februar. Izstopa nenatančna napoved modela SVR-rbf za drugi dan (nedelja), po drugi strani pa je napoved zelo dobra za četrty in peti dan, ko se napaka približa številki, ki smo jih videli v primeru porabe (okoli 3%). Opazimo torej velik razpon napake v odvisnosti od dneva v tednu. Najmanjšo povprečno napako ima ponovno model GARCH- t , sledijo mu

modeli GARCH- n , SVM-rbf, EGARCH in SVM-lin. Vidimo, da so napake za soboto in nedeljo (prvi in drugi dan) bistveno višje, to je bilo opaziti tudi za mesec februar.

	1	2	3	4	5	6	7	Povprečje
GARCH- n	11.77	9.78	8.58	4.06	3.57	6.60	6.00	7.19
GARCH- t	6.77	12.05	5.51	5.76	4.89	4.02	4.30	6.19
EGARCH	12.77	10.57	10.93	3.89	4.11	6.79	6.44	7.93
SVM-lin	13.93	6.15	14.95	8.22	5.03	5.11	8.06	8.79
SVM-rbf	10.93	19.42	7.05	2.95	3.36	5.08	6.12	7.84

Tabela 18: Napaka $MAPE(\%)$ za tretji teden julija.

Slika 35 prikazuje napovedi modelov GARCH- t in SVM-rbf za tretji teden julija. Res vidimo, da je napoved modela SVR-rbf za četrty in peti dan zelo dobra in precej zgrešena za prvi in drugi dan. Model GARCH- t po drugi strani dobro napove ceno za prvi dan, za drugi dan ima nekoliko težav s prvo polovico dneva. Preostale dni sta si modela v povprečju dokaj enakovredna.



Slika 35: Napovedi modelov GARCH- t in SVM-rbf za tretji teden julija.

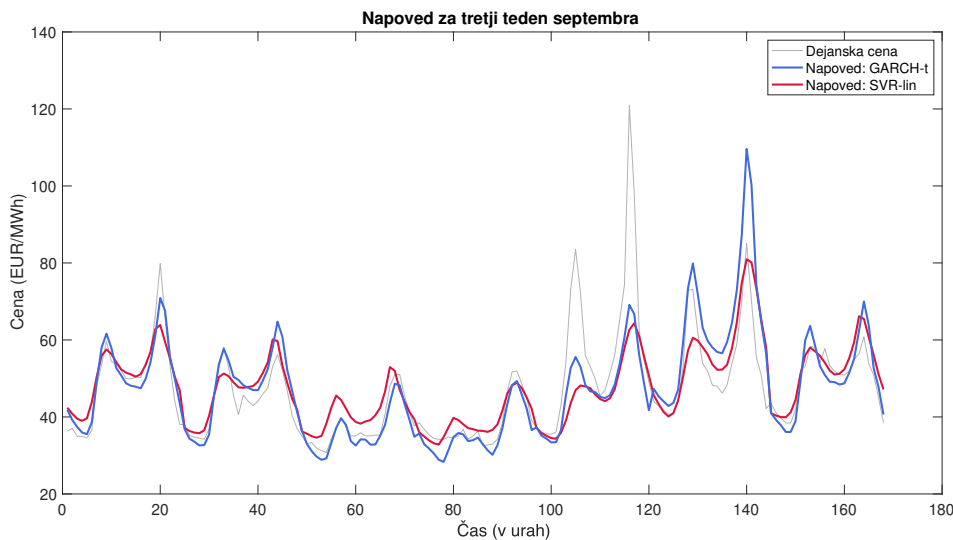
Nazadnje si oglejmo še napovedi za tretji teden septembra. Prvi dan je tokrat četrtek. Tabela 19 prikazuje napake, izračunane za vsak dan v tednu. Napovedi modelov GARCH so spet precej blizu, razen za tretji in četrty dan (sobota in nedelja), ko ima model GARCH- t bistveno boljšo napoved. Slabšo napoved opazimo v ponedeljek, ko je bil nenavaden skok cene (peti dan). V primeru modelov SVR zopet opazimo malo večje razlike med jedroma, zlasti za prvi dan. Tudi v tem primeru napoved za peti dan ni dobra. Model z najmanjšo povprečno napako je ponovno GARCH- t , sledijo mu modeli SVM-lin, EGARCH, GARCH- n in SVM-rbf. Prav za vse tri preiskovane tedne je torej model GARCH- t najboljši. Vidimo, da so za peti

dan napake precejšnje, zato tudi v povprečju. Preostale dni se gibljejo okrog 5-6 %, kar je zelo dobro.

	1	2	3	4	5	6	7	Popvrečje
GARCH- n	6.15	7.33	13.57	12.63	23.44	6.18	9.78	11.30
GARCH- t	6.04	7.98	4.64	6.46	17.97	18.52	6.77	9.77
EGARCH	6.85	7.71	13.79	12.22	24.64	3.79	9.96	11.28
SVM-lin	7.36	7.89	10.54	6.19	20.04	10.96	6.17	9.88
SVM-rbf	20.96	12.56	11.75	9.51	18.26	6.27	6.97	12.33

Tabela 19: Napaka $MAPE(\%)$ za tretji teden septembra.

Slika 36 prikazuje napovedi modelov GARCH- t in SVM-lin za tretji teden septembra. Res opazimo velik skok cene peti dan (ponedeljek), ko so napovedi nenaatančne. Sicer pa so kar blizu dejanski vrednosti cene. Model GARCH- t nekoliko preceni vrednost cene za torek, kot posledica skoka v ponedeljek, vendar pa je že v sredo napoved veliko bolj natančna. Po drugi strani pa model SVR-lin ne reagira tako močno in v torek napove boljše kot model GARCH- t .



Slika 36: Napovedi modelov GARCH- t in SVM-lin za tretji teden septembra.

Oglejmo si še napako $MAPE$, izračunano za tretji teden vseh mesecev v letu 2020, in sicer po formuli

$$MAPE = \frac{1}{168} \sum_{t=1}^{168} \frac{P_t - \hat{P}_t}{\bar{P}}, \quad (5.15)$$

kjer sta P_t in \hat{P}_t dejanska in napovedana vrednost cene ob času t , $\bar{P} = \frac{1}{168} \sum_{t=1}^{168} P_t$. Izračunane napake prikazuje tabela 20, za vsak mesec pa je odebljena najmanjša vrednost napake. Najboljša napoved je za mesec julij in model GARCH- t in sicer 5.95 %, najslabša pa za mesec maj in model SVM-rbf, in sicer 32.04 %. V devetih

primerih ima najboljšo napoved model GARCH- t . Za razliko od primera porabe, bi v primeru cene lahko rekli, da model GARCH- t prevladuje. Tudi za mesec september in november je model GARCH- t zelo blizu najmanjši vrednosti napake.

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Jan	10.95	11.85	10.03	8.84	9.92
Feb	9.32	8.80	8.19	7.22	8.17
Mar	16.59	31.31	18.88	16.24	18.85
Apr	31.65	27.34	24.96	23.59	24.76
Maj	27.09	32.04	26.56	21.83	27.21
Jun	16.37	14.75	14.34	11.24	14.60
Jul	8.64	7.31	6.95	5.95	7.67
Avg	17.11	17.22	16.64	15.61	16.75
Sep	10.28	12.51	11.46	10.42	11.44
Okt	11.87	11.67	11.89	15.20	12.20
Nov	10.98	11.49	11.45	11.29	11.45
Dec	24.54	24.74	18.52	17.58	18.89

Tabela 20: Napaka $MAPE(\%)$ za tretji teden vseh mesecev v letu 2020.

Tabela 21 prikazuje povprečje napak za 12 analiziranih tednov, najmanjšo napako ima model GARCH- t . V primeru porabe pa je bil to model GARCH- n .

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Povprečna $MAPE$	16.28	17.58	14.99	13.75	15.16

Tabela 21: Povprečna napaka $MAPE(\%)$.

Tabela 22 prikazuje izračun $RMSE$ za 12 obravnavanih tednov. Najmanjše vrednosti dobimo podobno kot za napako $MAPE$, izstopa model GARCH- t .

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Jan	7.47	8.19	6.87	6.35	6.93
Feb	5.21	4.95	4.31	3.69	4.26
Mar	5.69	9.98	6.58	5.10	6.65
Apr	10.18	9.41	8.71	7.87	8.65
Maj	7.93	8.79	7.51	6.38	7.65
Jun	6.23	5.56	5.52	4.44	5.61
Jul	3.74	3.30	3.12	2.70	3.41
Avg	9.15	9.00	9.16	8.02	9.25
Sep	8.42	8.87	8.38	8.15	8.58
Okt	7.13	6.69	6.74	8.77	6.88
Nov	6.94	7.08	6.89	7.23	7.00
Dec	17.22	17.94	11.68	11.48	11.84

Tabela 22: Napaka $RMSE$ za tretji teden vseh mesecev v letu 2020

Tabela 23 prikazuje povprečje dvanajstih napak $RMSE$, najmanjšo napako ima pričakovano model GARCH- t .

	SVM-lin	SVM-rbf	GARCH- n	GARCH- t	EGARCH
Povprečna $RMSE$	7.94	8.31	7.12	6.68	7.23

Tabela 23: Povprečna napaka $RMSE$.

Zaključimo lahko, da je najboljši model za napovedovanje cene električne energije model GARCH- t . To nakazuje, da je res primernejša studentova t -porazdelitev s težjim repom kot pa normalna porazdelitev (GARCH- n), kar se v primeru resničnih finančnih podatkov pogosto izkaže. Napovedi so nekoliko manj natančne kot v primeru porabe električne energije, kar pa je tudi pričakovano, saj je obnašanje manj predvidljivo.

6 Zaključek

Tako za porabo kot za ceno električne energije smo primerjali napovedi petih modelov. V primeru porabe ni očitno izstopal noben model, z izjemo modela SVM-lin, ki bi ga trdno uvrstili na zadnje mesto. Vseeno pa lahko rečemo, da so se modeli ARMA-GARCH odrezali bolje kot modeli SVM. Nasprotno pa je v primeru napovedovanja cene precej izstopal model ARMA-GARCH- t . Tudi tokrat so se v splošnem bolje odrezali modeli ARMA-GARCH, med modeli SVR pa se je za boljšo izbiro izkazalo radialno jedro (v primerjavi z linearnim). V splošnem lahko zaključimo, da imajo modeli ARMA-GARCH natančnejše napovedi kot pa modeli podpornih vektorjev. Kot smo že omenili, so napovedi v primeru cene precej slabše, prav tako pa so občutne razlike v kakovosti napovedi v odvisnosti od opazovanega tedna. Ta razlika je smiselna, saj so cene bolj volatilne kot poraba in je zato obnašanje cen manj predvidljivo. Napovedi cen (in porabe) bi lahko izboljšali z modeli, kjer bi kot dodatne spremenljivke vključili npr. hitrost vetra, temperaturo, količino padavin, ceno drugih energentov, in pa tudi porabo oz. napoved porabe pri napovedovanju cene. Preizkusili bi lahko tudi druge modele strojnega učenja kot so npr. nevronske mreže.

Literatura

- [1] *Agencija na slovenskem energetske trgu*, [ogled 2. 8. 2021], dostopno na https://www.agen-rs.si/o_agenciji.
- [2] *aicbic*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/aicbic.html>.
- [3] *arima*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/arima.html>.
- [4] *autocorr*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/autocorr.html>.
- [5] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics **31**(3) (1986) 307–327, doi: 10.1016/0304-4076(86)90063-1.
- [6] G. E. Box in dr., *Time Series Analysis, Forecasting and Control*, Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ, 2016.
- [7] P. J. Brockwell in R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, ZDA, 2006.
- [8] P. J. Brockwell in R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, ZDA, 2012.
- [9] *estimate*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/arima.estimate.html>.
- [10] M. Fałdziński, P. Fiszeder in W. Orzeszko, *Forecasting Volatility of Energy Commodities: Comparison of GARCH Models with Support Vector Regression*, Energies **14**(1) (2021), doi: 10.3390/en14010006.
- [11] *fitrsvm*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/stats/fitrsvm.html>.
- [12] *forecast*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/arima.forecast.html>.
- [13] R. C. Garcia in dr., *A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices*, IEEE Transactions on Power Systems **20**(2) (2005) 867–874, doi: 10.1109/TPWRS.2005.846044.
- [14] *lbqtest*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/lbqtest.html>.
- [15] *parcorr*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/econ/parcorr.html>.
- [16] *predict*, [ogled 4. 7. 2021], dostopno na <https://www.mathworks.com/help/stats/compactregressionsvm.predict.html>.

- [17] B. Scholkopf in A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, The MIT Press, Cambridge, 2002.
- [18] L. Todorovski, *Metoda podpornih vektorjev in jedrne funkcije*, [ogled 9. 5. 2021], dostopno na <http://kt.ijs.si/~ljupco/lectures/itap/itap-06.pdf>.
- [19] *Total load*, [ogled 4. 7. 2021], dostopno na <https://transparency.entsoe.eu/load-domain/r2/totalLoadR2/show>.
- [20] *Rezultati trgovanja za dan vnaprej*, [ogled 4. 7. 2021], dostopno na <https://www.bsp-southpool.com/trgovalni-podatki.html>.
- [21] *Trgovanje za dan vnaprej*, [ogled 4. 7. 2021], dostopno na <https://www.bsp-southpool.com/trgovanje-za-dan-vnaprej.html>.
- [22] *Udeleženci na trgu z električno energijo*, [ogled 2. 8. 2021], dostopno na <https://www.agen-rs.si/gospodinjski/elektrika/udelezenci-na-trgu-z-elektricno-energijo>.
- [23] R. Weron, *Electricity price forecasting: A review of the state-of-the-art with a look into the future*, International Journal of Forecasting **30**(4) (2014) 1030–1081, doi: 10.1016/j.ijforecast.2014.08.008.
- [24] *Zagotavljanje sistemskih storitev*, [ogled 2. 8. 2021], dostopno na <https://www.agen-rs.si/-/zagotavljanje-sistemskih-storitev>.