

Aegridade analüüs

Praktikum nr. 3, 2016

Käsitletavad teemad: Eksponentsiaalse silumise, Holti meetodi ja Holt-Wintersi meetodi rakendamine aegrea prognoosimiseks; prognoosivigade mõõdikute arvutamine, autokorrelatsioonifunktsioon, prognoosimudeli sobivuse uurimine.

Silumisvõtteid kasutatakse lisaks andmetest müra eemaldamisele sageli ka prognoosimisel. Tähistame kujul $\hat{z}_{t+p|t}$ aegrea väärtuse prognoosi ajamomendi $t+p$ jaoks, mis on arvutatud ainult ajamomendil t teadaolevate andmete põhjal. Kui ennustame aegrea järgmist väärtust (st $p = 1$), siis kasutame ka lihtustatud tähist $\hat{z}_{t+1}$. Lihtsamad eksponentsiaalsel keskmistamisel põhinevad prognoosimeetodid on järgmised.

- **Eksponentsiaalse keskmistamise meetod**, mida võib kasutada, kui aegreal ei ole trendi ja perioodilist komponenti. Tavaliselt defineeritakse $\hat{z}_1 = z_1$ ning arvutatakse seejärel

$$\hat{z}_{t+1} = \alpha z_t + (1 - \alpha)\hat{z}_t, \quad t = 1, \dots, n.$$

Parameetri α võib valida ise või määrata sobivaim väärtus olemasolevate andmete põhjal nii, et mineviku jaoks leitud ühesammulised ennustused oleks võimalikult lähedased olemasolevate andmetega.

- **Holti meetod**. Kasutatakse ajas muutuva trendiga, kuid ilma perioodilise komponendita aegrea väärtuste prognoosimiseks. Valemid on järgmised:

$$\hat{z}_{t+p|t} = a_t + b_t p,$$

kus trendikõvera hetkeväärtuse a_t arvutamisel kasutatakse z_t väärtust ning eelnevate andmete põhjal tehtud prognoosi:

$$a_t = \alpha z_t + (1 - \alpha)\hat{z}_t = \alpha z_t + (1 - \alpha)(a_{t-1} + b_{t-1})$$

ning trendikõvera tõusu b_t arvutamisel kasutatakse eelmise väärtuse ning a muutuse keskmist:

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}.$$

Meetodi kasutamiseks tuleb valida või andmete põhjal hinnata väärtused a_1, b_1, α, β . Sageli valitakse $a_1 = z_1$, $b_1 = 0$ või siis kasutada nendeks teatud arvu esimeste z väärtuste jaoks leitud lineaarse regressioonikõvera vastavaid väärtusi.

- **Holt-Wintersi meetodid** perioodilise komponendiga ridade prognoosimiseks. Kasutamisel tuleb õigepealt otsustada, kas trend ja perioodiline komponent on omavahel korrutatud või liidetud. Korrutamise eeldus (ehk multiplikatiivne meetod) sobib positiivsete väärtustega ridadele juhul, kui perioodiliste võngete ulatus kasvab koos väärtuste kasvuga. Vastavad valemid on toodud loengukonspektis, kuid intuitiivselt on tähtis aru saada, et leitud parameetritest α kirjeldab seda, kui tugevalt me arvestame viimast väärtust jooksvale ajahetkele vastava trendikõvera väärtuse leidmisel ($\alpha = 1$ puhul loeme selleks viimast vaatlust, $\alpha = 0$ korral on aga trendikõver ajas muutumatu sirge), β kirjeldab lineaarse trendi tõusu muutumise kiirust juhul, kui leitud trendikõvera muutus üle viimase perioodi erineb prognoositud muutuses (ehk eelmisel ajamomendil prognoositud trendikõvera tõusust) ning γ kirjeldab seda, kui kiiresti muutub sesoonne komponent ($\gamma = 1$ puhul arvestame järgmise perioodi vaatluse sesoonse osa puhul ainult käesolevat aastat, $\gamma = 0$ puhul aga sesoonne komponent on ajas muutumatu).

Käesoleva praktikumi ennustusmeetodite rakendamiseks on R-is käsk `HoltWinters`. Vaikimisi kasutab R aditiivset meetodit ja leiab parameetritest α, β ja γ need, mis ei ole käsus otseselt ette antud, ennustusvigade ruutude summat minimiseerides. Näiteid kasutamisevõimalustest olemasoleva aegrea `x` korral:

`m1=HoltWinters(x)`-sobitab aditiivse meetodi reale `x`,

`m1=HoltWinters(x,seasonal="multiplicative")`-sobitab multiplikatiivse meetodi reale `x`,

`m1=HoltWinters(x,gamma=FALSE)`-sobitab Holti meetodi reale `x`,

`m1=HoltWinters(x,beta=FALSE,gamma=FALSE)`-eksponentsiaalse keskmistamise meetod.

Soovi korral võib optimeerimise alustamiseks ise sobivad parameetrite väärtused ette anda. Selleks tuleb käsu `HoltWinters` argumendina edastada

`optim.start=c(alpha=a0,beta=b0,gamma=g0)`, kus a_0 , b_0 ja g_0 on mingid konkreetsed numbrid nulli ja ühe vahel. Detailsemat infot käsu võimaluste kohta saab `help` käsuga. Käsu tulemuse põhjal saab tekitada graafiku `plot` käsuga, leida ennustusvigu, ennustada tulevikuväärtusi `predict` käsuga jpm.

1. Tekitage joonised, kus on majutatud turistide andmestiku esialgse aegrea ning eksponentsiaalse silumise, Holti meetodi ja Holt-Wintersi aditiivse ja multiplikatiivse meetodi poolt leitud ühesammuliste ennustuste graafikud (iga meetodi kohta eraldi joonis). Samuti tehke graafikud iga vaadeldud meetodi ennustusvigadega, mis on kättesaadavad käsuga kujul `residuals(mudel)`.
2. Meetodi headusest aimusaamiseks ja erinevate meetodite omavaheliseks võrdlemiseks arvutatakse mitmesuguseid ennustusvigade mõõdikuid. Arvutage eelmises punktis mainitud meetodite korral keskmine absoluutne viga MAD, ruutjuur vigade ruutude keskmisest RMSE ja keskmine absoluutne suhteline viga MAPE. Selleks on hea teada, et keskmiste arvutamisel saab kasutada R käsku `mean`, mis leiab etteantud vektori aritmeetilise keskmise.
3. Leidke parima RMSE-ga meetodi korral ennustus järgnevakts neljaks kuuks ning ka ennustusvigade hinnangud. Vajadusel uurige selleks käsu `predict.HoltWinters` abimaterjale.

Aegridade ennustamismeetodit võib lugeda sobivaks, kui tekkivad ennustusvead ei ole süstemaatilised, st ei sisalda enam mingit informatsiooni, mida saaks (lihtsalt) kasutada tulevikuproгноoside täpsustamiseks. Sellist vigade vahelist sõltuvust saab uurida nn. autokorrelatsioonifunktsiooniga, mis väljendab hetkevea ja k ajasammu minevikus olnud prognoosivea lineaarset sõltuvust. Matemaatiliselt tähendab see prognoosivigade aegreale vastava juhusliku protsessi $(A_t)_{t \in \mathbb{Z}}$ korral suuruste A_t ja A_{t-k} vahelist korrelatsiooni. Eeldusel, et selline korrelatsioon sõltub ainult nihkest k ja mitte ajamomendist t , on vaadeldav korrelatsioon hinnatav prognoosivigade aegrea $a_1, \dots, a_n$ abil, täpsemad räägitakse hindamisest loengus. Mida suurem on selle kordaja absoluutväärtus, seda täpsemalt saab k sammu tagasi arvutatud vea abil hetkeviga ennustada.

Isegi täiesti juhuslike ennustusvigade puhul võivad andmete põhjal arvutatud autokorrelatsioonide hinnangud olla nullist erinevad, seetõttu võetakse sobivuse üle otsustamisel arvesse ka seda, kui suured võivad vastavad autokorrelatsioonid juhuslike ennustusvigade puhul olla. Üheks sobivuse kriteeriumiks on praktikas näiteks iga autokorrelatsiooni hinnangu jäämine piiridesse, kuhu juhuslike vigade puhul vastava autokorrelatsiooni hinnangud jääksid 95% juhtudest. Samas arvutatakse korraga paljudele erinevatele nihetele vastavaid autokorrelatsioone ning mida suurem on vaadeldavate autokorrelatsioonide hulk, seda suurem on võimalus, et mõni neist juhuslikult veapiiridest välja jääb. Seetõttu on kasulik kontrollida tekkinud autokorrelatsioonide rühma kui terviku vastavust juhuslike vigade eeldusega. Üheks sobivaks vahendiks on **Ljung-Box test**, mille puhul tuleb ette anda korraga vaadeldavate autokorrelatsioonide arv, mille põhjal leitakse sobivalt kaalutud autokorrelatsioonide summa ning leitakse tõenäosus, et vähemalt nii suur väärtus saadakse täiesti juhuslike andmete korral. Kahjuks ei ole universaalseid reegleid selle kohta, kui suurt rühma peaks korraga vaatama. Teada on, et rühma suurus peaks olema oluliselt väiksem vaatluste arvust, kuid samas peaks olema küllalt suur, et arvutatava statistiku jaotus oleks mõistlikult lähendatud p -väärtuse leidmisel kasutatud χ^2 jaotusega. Majanduslike aegridade puhul loetakse sageli mõistlikuks vaadelda rühmi, mis vastavad vähemalt kahele-kolmele võimalikule perioodile (tavaliselt aastale), sest pikemaajalisi sõltuvusi ei peeta enamasti usaldusväärseks.

Praktikumis vaatleme prognoosivigade sõltumatuse uurimist Holt-Wintersi meetodi baasil arvutatud ennustuste korral.

5. Tekitage käsuga `rnorm` 100 sõltumatut standardse normaaljaotusega juhusliku suuruse väärtust ning leidke nendest moodustatud aegrea autokorrelatsioonifunktsiooni väärtused (käsuga `acf`). Testige selle aegrea esimese kümne korrelatsioonikordaja vastavust sõltumatute juhuslike suuruste reale Ljung-Box testiga (käsk `Box.test`, kasutades suvandit `type="Ljung-Box"`).
6. Kontrollige ülesandes 2 sobitatud eksponentsiaalse keskmistamise, Holti meetodi ja Holt-Wintersi meetodite sobivust prognoosimiseks ennustusvigade autokorrelatsioonide uurimise teel. Ljung-Box testi korral soovitatakse aegrea jääkide uurimise korral arvestada ka sobitamisel kasutatud parameetrite arvu, mida saab käsule `Box.test` edastada parameetriga `fitdf=param_arv`.
7. Kontrollige aditiivse ja multiplikatiivse Holt-Wintersi meetodi sobivust kinnisvara tehingumahtude prognoosimiseks.