

Aegridade analüüs

Praktikum nr. 7, 2016

Käsitletavad teemad: ARIMA mudelite sobitamine mittestatsionaarsetele ja perioodilist (sesoonset) komponenti mitteomavatele aegridadele, tulevikuväärtuste prognoosimine ARIMA mudelite korral

Sobiva mudeli valimise protseduur on järgmine:

1. Vaatleme aegrida ja autokorrelatsioonifunktsiooni. Mittestatsionaarsuse tunnusteks on keskmise ja/või varieeruvuse selge muutumine ajas (ka sarnane käitumine erinevatel tasemetel) ning autokorrelatsioonide väga aeglane kahanemine järgu kasvades ning väga suur (praktiliselt ühega võrdne) esimene autokorrelatsioon.
2. Mittestatsionaarsuse olemasolul tuleks teha otsus selle kohta, et kas on mõistlik eeldada globaalse (lineaarse, ruutfunktsioonina või mingi teise funktsiooni abil kirjeldatava) trendi olemasolu. Kui on, siis eemaldame selle ennem sobiva ARIMA mudeli kuju kindlakstegemist. Eemaldamiseks võib sobitada andmetele vastava trendikõvera vähimruutude meetodil.
3. Kui globaalset trendi ei ole mõistlik eeldada või pärast selle eemaldamist on ikkagi tegemist mittestatsionaarse aegreaga, siis järgmisena uuritakse aegrea muutusid ehk diferentsitud aegrida. Diferentsimise vajadusele on rea ebaselge käitumise korral võimalik saada kinnitust ka mitmete nn ühikjuurte testidega; meie vaatleme Phillips-Perroni testi. Viimase nullhüpoteesiks on see, et vaadeldava rea muutu ei saa paremini ennustada, kui lisaks varasematele muutudele teame ka rea viimast väärtust (ehk et diferentsitud reale mudeli sobitamisel saame parema tulemuse, kui esialgsele reale mudeli sobitamisel). SAS tarkvaras on sellel testil erinevad versioonid sõltuvalt sellest, kas esialgse rea ennustamisel lubada nullist erinevat keskmist või lineaarset trendi, R tarkvara versioonis vaadeldakse ainult viimast juhtu. Kui ka diferentsitud rida ei vasta statsionaarsuse tingimustele, siis vaadeldakse kaks korda diferentsitud aegrida jne. NB! Lihtsalt igaks juhuks diferentside leidmine on vale lähenemine, see vähendab mudeli sobitamisel kasutatava info hulga ja muudab prognoose ebatäpsemaks.
4. Kui mingi arvu d korda diferentsitud aegrea korral paistavad statsionaarsuse tingimused olevat täidetud, siis leiame selle jaoks sobiva ARMA(p, q) mudeli; kokkuvõttes tähendab see esialgse rea jaoks ARIMA(p, d, q) tüüpi mudeli sobitamist. Kui globaalne trend on eemaldatud või kui selle olemasolu esialgse rea jaoks ei ole mõistlik eeldada, siis diferentsitud reale mudeli sobitamisel on õige eeldada nullkeskmist. Kui d korda diferentsitud rea sobitamisel lubada nullist erinevat keskmist, siis see tähendab tegelikult d järku polünoomiga kirjeldatava globaalse trendi olemasolu lubamist.

Praktikumi tegevused R-is:

1. Tutvume kõigepealt mittestatsionaarsete ARIMA tüüpi aegridade ning nende diferentsitud versioonide võimaliku käitumisega. Diferentsitud aegrida saab tekitada käsuga `diff`. Phillips-Perroni testi saab teostada käsuga `PP.test`.
 - Genereerige käsu `arima.sim(n=500,model=list(order=(0,1,1),ma=0))` abil aegrida (see on lihtne juhuslik ekslemine). Kas on selge, millega põhjendada globaalse trendi puudumist vaadeldud rea korral? Kas Phillips-Perroni test aitab jõuda õigele otsusele diferentsi leidmise vajalikkuse osas? Korra tegevust mitu korda, varieerides veidi ka `ma` parameetri väärtust või lisades mingi väikese `ar` parameetri.
 - Korra eelnevaid tegevusi ARIMA($p, 2, q$) tüüpi mudelite korral (sel juhul `order` parameetri teine komponent on 2).
2. Uurime globaalset trendi sisaldavate ridade omadusi ning trendi eemaldamist. Trendi eemaldamine R-is on oluline, kuna ARIMA(p, d, q) mudeli sobitamisel $d > 0$ korral eeldatakse, et diferentsitud rea keskmine on null.

- Moodustage lineaarsele trendile vastav aegrida käsuga `linear=ts(1:500)` ning tekitage trendiga aegrida kujul
`z=arima.sim(n=500,list(ar=0.6))+0.05*linear`.
 Vaadeldge rea ning selle diferentsitud versioonide graafikuid, autokorrelatsioone ja osa-autokorrelatsioone. Kas Phillips-Perroni test aitab selle rea puhul teha õige otsuse?
 - Trendi eemaldamiseks sobitame andmetele trendikõvera käsuga `arima` kujul
`trend=arima(z,xreg=trendi_rida)`
 ning lahutame reast trendi:
`trendita_z=residuals(trend)`.
 Pärast trendita rea uurimist on kõige otstarbekam sobitada mudel esialgsele reale nii, et `arima` käsk leiaks koos nii trendi kui ka mudeli parameetrid. Selleks tuleb trendile vastav rida anda ette parameetriga `xreg=trendi_rida`.
 - Korra eelnevat juhul, kui trend on liidetud Teie poolt valitud ARIMA(p,1,q) tüüpi mudelile.
3. Rakendage eelnevaid teadmisi andmestikus `praktikum7.txt` olevate aegridadele sobiva mudeli leidmiseks ja nelja järgneva väärtuse ennustamiseks. Kui `arima` mudeli sobitamisel on trend eemaldatud suvandit `xreg=trendi_rida` kasutades, siis prognoosimiseks tuleb `predict` käsuga anda ette trendi uued väärtused suvandiga `newxreg=uued_väärtused`.