

Aegridade analüüs

Praktikum nr. 12, 2016

ARIMA tüüpi mudelite sobitamine empiirilistele aegridadele. ARIMAX tüüpi mudelite sobitamine.

Sel korral kasutame praktikumis R tarkvara. Kõigepealt leiame sobivad mudelid praktikumides 1 ja 2 vaadeldud Eesti kinnisvaratehingute ning majutatud turistide andmetele. ARIMAX mudelite sobitamiseks vajalikud andmed paiknevad failis **prax12.txt**. Tuletame meelde, kuidas kasutada deterministlikku regressorit, st sobitame mudeleid kujul

$$Z_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$

kus $X_t = t$ ning ε_t on mingi ARIMA tüüpi protsess. Mudeli sobitamise protsess on järgmine:

- i) Sobitame reale Z_t lineaarse regressioonimudeli regressoriga X_t ning leiame selle mudeli prognoosivead.
- ii) Teeme kindlaks, mis tüüpi ARIMA mudel prognoosivigadele sobib
- iii) Sobitame vastava ARIMAX mudeli andmetele ning testime saadud mudeli jääkide sõltumatust. Vajadusel muudame vigade jaoks valitud ARIMA mudeli kuju.

Kindlasti tasuks võrrelda leitud ARIMAX mudelit parima ARIMA mudeliga, mis regressoreid ei kasuta.

- 1) Sobitage lineaarse regressoriga ARIMAX mudel ridadele $z1$ ja $z2$. Tehke kindlaks, kas regressori lisamine on õigustatud. Meeldetuletus: lineaarse mudeli sobitamiseks on R-is käsk **arima(rida,xreg=regressor)**; regressioonimudelile vastavad väärtused saab kätte käsuga **residuals(mudel)**, kus mudel on eelneva käsu **lm** tulemus. Kui prognoosivigade analüüsimise tulemusel on püstitatud hüpotees vigadele sobiva ARIMA mudeli tüübi kohta, siis vastava ARIMAX mudeli sobitamiseks saab kasutada **arima** käsku, kus lisaks tavalistele parameetritele on ette antud ka regressor parameetriga **xreg**, näiteks

arima(z,order=c(1,0,0),xreg=aeg),

kus **aeg** sisaldab näiteks vaatluste järjekorranumbreid. Leidke samuti ennustused järgneva 5 perioodi jaoks; arimax mudelis tuleb selleks **predict** käsus anda regressorite vastavad väärtused parameetriga **newxreg**.

Järgnevalt vaatleme juhtu, kus regressoriteks on mingid täiendava aegrea andmed. Sobiva ARIMAX mudeli leidmise protseduur on sel juhul sama nagu varem.

- 2) Vaatleme juhtu, kus meil on teada, et rea $z3$ regressoriks peaks sobima rea $x3$ nihkega 6 väärtused. Seda, kuidas tegelikult sobivat nihet leida, vaatleme hiljem, kuid etteruttavalt võib öelda, et vähemalt statsionaarsete ridade korral annab sobiva nihke kohta informatsiooni ristkorrelatsioon (mida arvutatakse R-is käsuga **ccf**). Kuna esimese 6 $z3$ väärtuse jaoks meil regressoreid ei ole, jätame need vaatluse alt välja:

z31=window(Z3,start=7).

Sejärel moodustame regressorite vektori $x3$ väärtustest (viimased 6 jäävad välja), leiame vastava lineaarse regressioonimudeli vigade põhjal sobiva ARIMA mudeli kuju ning sejärel sobitame ARIMAX mudeli valitud regressoritega. Pärast leitud mudeli sobivuse kindlakstegemist leidke ka prognoos järgmise kuue ajaperioodi jaoks.

- 3) Mitme regressori olemasolul tegevuste iseloom jääb samaks. Vaatleme juhtu, kus rea $z4$ jaoks kasutatakse regressoritena rea $x41$ nihkega 4 väärtusi ja rea $x42$ nihkega 2 väärtusi. Kõige suuremaks erinevuseks eelneva juhuga võrreldes on see, et **arima** käsus antakse regressorid ette kujul `xreg=data.frame(x1,x2)`, kus $x1, x2$ on regressoritele vastavad aegread; samasugune süntaks on ka ennustamisel parameetri **newxreg** etteandmisel. Leidke sobiv mudel ja ennustage selle põhjal $z3$ järgnevad 2 väärtust.

Lõpuks vaatleme lühidalt küsimust, kuidas kindlaks teha, milliseid antud regressoraegrea X_t nihkeid huvipakkuva aegrea Z_t väärtuste ennustamisel kasutada. Lihtsuse mõttes eeldame, et mõlemad aegread on statsionaarsed ning nullkeskmisega, vastasel korral tuleb kas ühest või mõlemast leida diferentsid nii, et need eeldused oleks rahuldatud.

Eelnevalt nägime (nt ülesanne 2), et ristkorrelatsioonid võivad anda sobiva nihke kindlakstegemise jaoks olulist infot. Samas aga ei anna lihtsalt ridade Z_t ja X_t ristkorrelatsioonid tavaliselt kuigi selget infot selle kohta, mitu erinevat X_t nihet tuleks kasutada, et parimat ennustust leida. Oluliseks võtteks täpsema info saamiseks on nn eelvalgendamine (inglise keeles *prewhitening*). See seisneb selles, et kõigepealt leitakse sobiv ARIMA tüüpi mudel regressorrea X_t jaoks ning leitakse mudeli jäägid $A_{x,t}$. Need jäägid on mudeli kehtivuse korral sõltumatud ja sama jaotusega (ehk nn valge müra; sellest ka tehnika nimetus). Seejärel leitakse jäägid $A_{z,t}$, mis vastavad **sama** mudeli kasutamisele suuruste Z_t ennustamiseks ning arvutatakse leitud jääkide $A_{x,t}$ ja $A_{z,t}$ vahelised ristkorrelatsioonid. Osutub (täpsemalt räägitakse sellest loengus), et nende korrelatsioonide abil saab kindlaks teha, milliseid X nihkeid tuleks Z väärtuste ennustamisel kasutada. Erijuhul, kui nullist erinevaid ristkorrelatsioone on lõplik arv (ja need vastavad X minevikuväärtuste kasutamisele), siis tuleks kasutada täpselt selliste nihetega X_t väärtusi regressoritena.

- Kasutame eelpool kirjeldatud tehnikat, et leida parim prognoosimudel rea $z5$ tulevikuväärtuste ennustamiseks selle rea ja rea $x5$ teadaolevate väärtuste abil. Selleks leidke reale $x5$ sobiv ARIMA tüüpi mudel (olgu selle nimeks **m5x**). Leitud mudeli jäägid on kättesaadavad kujul **residuals(m5x)**. Sama mudeli kasutamisel rea $z5$ ennustamiseks tekkivate jääkide leidmiseks sobiama samade kordajatega mudeli sellele reale käsuga

```
m5z=arima(z5,order=sama_mis_m5x,fixed=m5x$coef)
```

Edasi leiame tekkinud jäägid ja nende ristkorrelatsiooni **m5x** jääkidega. Seejärel moodustame nullist erinevatele kordajatele vastavate regressoritega ARIMAX mudeli nii nagu eelnevalt, testime selle sobivust ja ennustame tulevikuväärtusi.