MS-C2132 Systeemianalyysilaboratorio I

Laboratoriotyö 2: Sähkönkulutuksen ennustaminen ja hankinnan optimointi

Aikasarja on joukko peräkkäisiä, toisistaan riippuvia havaintoja. Aikasarja-analyysin tavoitteena on kuvata, selittää, ennustaa ja mahdollisesti ohjata aikasarjan tuottanutta prosessia. Ehkä suosituin lähestymistapa aikasarja-analyysiin on Box-Jenkins-menetelmä, jossa tarkastellaan erästä lineaaristen stokastisten mallien luokkaa. Työssä rakennetaan kyseisellä menetelmällä sähkön kulutukselle malli. Mallilla ennustetaan seuraavan vuorokauden sähkön kulutusta. Sähkön hankinta optimoidaan ennusteeseen perustuen.

1 Box-Jenkins-malleista

Koska Box-Jenkins-malleja on käsitelty kurssilla MS-C2128 Ennustaminen ja aikasarja-analyysi, ei työohjeessa esitellä ARIMA-mallien perusasioita.

1.1 SARIMA-mallit

Mikäli datassa on kausivaihteluja, siirrytään Seasonal ARIMA- eli SARIMA-malleihin. Kausivaihteluita sisältävässä mallissa käytetään kausivaihtelun pituisen viiveen päästä otettuja suureen ja kohinan arvoja selittämään tämänhetkistä suureen arvoa. ARIMA-malli voidaan ymmärtää kausivaihtelumallin erikoistapaukseksi, jossa kausivaihtelun pituus on yksi. Käytetään ARIMA(p,d,q)-mallille seuraavaa notaatiota:

$$\Phi_p(B)\nabla^d z_t = \Theta_q(B)a_t, \tag{1}$$

missä z_t on mallitettava suure, a_t on valkoista kohinaa, viiveoperaattori $Bz_t = z_{t-1}$, $\nabla = (1-B)$ ja polynomit $\Phi_p(B) = (1-\phi_{1,1}B-\phi_{1,2}B^2-...-\phi_{1,p}B^p)$ ja $\Theta_q(B) = (1-\theta_{1,1}B-\theta_{1,2}B^2-...-\theta_{1,q}B^q)$ sisältävät mallin parametrit. Data määrää kertaluvut p, d ja q. Yleensä $d = 1$. Käytännössä mallin AR- ja MA-osiin otetaan tilastollisesti merkitsevät parametrit, ellei ole erityistä syytä muuhun. On tietysti mahdollista, että jokin parametreista p, d tai q on nolla, jolloin aito ARIMA-malli redusoituu ARMA-, ARI- tai IMA-malliksi. Mikäli kausivaihtelun pituus on s, käytetään SARIMA(P,D,Q)-mallia

$$\Phi_P(B^s)\nabla_s^D z_t = \Theta_Q(B^s)a_t, \tag{2}$$

missä viiveoperaattori $B^s z_t = z_{t-s}$, operaattori $\nabla_s = (1-B^s)$ ja polynomit $\Phi_P(B^s) = (1-\phi_{s,1}B^s-\phi_{s,2}B^{2s}-...-\phi_{s,p}B^{Ps})$ ja $\Theta_Q(B) = (1-\theta_{s,1}B^s-\theta_{s,2}B^{2s}-...-\theta_{s,q}B^{Qs})$ sisältävät mallin parametrit. On syytä huomata, että varsinkin jos viive s on pitkä datan kokonaispituuteen nähden, ei edes ole mahdollista käyttää kertaluvuille P, D ja Q suuria arvoja, vaan käytännössä usein P = D = Q = 1. Voi myös käydä niin, että P tai Q saa arvon nolla, jos viiveen s päässä olevat suureen tai kohinan arvot eivät ole merkittäviä selittäjiä.

Jos datassa on useita kausivaihteluita, päädytään multiplikatiiviseen SARIMA(p,d,q)x(P,D,Q)-malliin eli kausivaihtelut jaksoilla 1 ja s johtavat malliin

$$\Phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_t = \Theta_q(B)\Theta_Q(B^s)a_t, \tag{3}$$

missä merkinnät kuten edellä. Tässäkin on mahdollista, että jokin tai jotkin parametreistä p, d, q, P, D ja Q ovat nollia ja usein d=1 ja P ja Q ovat pieniä, ainakin jos viive on pitkä.

Yleisesti kausivaihtelu viiveillä $s_1, s_2, ..., s_i$ johtaa multiplikatiiviseen malliin, jossa on differoinnit viiveillä $s_1, s_2, ..., s_i$, AR-osassa polynomit $\Phi_{p_1}(B^{s_1})$, $\Phi_{p_2}(B^{s_2})$, ..., $\Phi_{p_i}(B^{s_i})$ ja MA-osassa polynomit $\Theta_{q_1}(B^{s_1})$, $\Theta_{q_2}(B^{s_2})$, ..., $\Theta_{q_i}(B^{s_i})$.

Esimerkiksi seuraavassa mallissa on kaksi kausivaihtelua, jotka ovat viikkovaihtelu ja vuosivaihtelu:

$$(1 - \phi_{7,1}B^7 - \phi_{7,2}B^{14} - \phi_{7,3}B^{21})\nabla_7\nabla_{365}z_t = (1 - \theta_{7,1}B^7 - \theta_{7,2}B^{14})a_t. \tag{4}$$

Siis $s_1 = 7$, $p_1 = 1$, $d_1 = 1$, $q_1 = 2$, $s_2 = 365$, $p_2 = 0$, $d_2 = 1$, $q_2 = 0$.

1.2 Mallien identifiointi

Tarkasteltavan ilmiön kuvaamiseen käytettävän SARIMA-mallin luokan valinnan jälkeen tehtävänä on määrittää mahdollisten kausivaihteluiden pituudet ja kertaluvut sekä parametrit p, d, ja q. Lineaarisesti kasvava tai vähenevä trendi poistetaan differoimalla aikasarjaa. Käytännössä differointien kanssa on kuitenkin syytä olla varovainen, sillä signaali–kohina-suhde saattaa differoitaessa olennaisesti heikentyä, varsinkin kohinaisessa datassa. Mikäli kuitenkin AR(1)-parametrin ϕ_1 estimaatti mallissa $(1 - \phi_1 B)z_t = a_t$ on hyvin lähellä ykköstä ($>$0,9), differointia voidaan yleensä pitää perusteltuna. Kausivaihtelun tunnistamisessa käytetään hyväksi *a priori* -tietämystä mallitettavasta prosessista. Usein esiintyviä jaksoja ovat vuorokausivaihtelu, viikkovaihtelu ja vuosivaihtelu. Kausivaihtelumallien käyttöä on harkittava erityisen huolellisesti silloin, kun kausivaihtelun pituus on suuri aikasarjan pituuteen nähden, sillä differointi pitkällä viiveellä syö dataa. Myös ∇_s-operaation järkevyyttä voi testata sovittamalla dataan vastaava $(1 - \phi_s B^s)z_t = a_t$ -malli ja katsomalla kuinka lähellä parametrin ϕ_s estimaatti on ykköstä.

Matemaattisista identifioinnin apuvälineistä tärkeimmät ovat autokorrelaatio ja osittaisautokorrelaatio. Ne kaikki kuvaavat suureen korrelaatio-ominaisuuksia itsensä suhteen eripituisilla viiveillä.

Autokorrelaatio r_k on suureen kovarianssi itsensä kanssa viiveellä k jaettuna r_0:lla eli varianssilla:

$$r_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2} \tag{5}$$

Vaimeneva autokorrelaatio viittaa AR-osaan ja yksittäiset piikit MA-osaan. Jos yksittäisiä, vaimenevia piikkejä esiintyy viiveen s välein, kysymys on luultavasti kausivaihtelusta.

Osittaisautokorrelaatiot ρ_k saadaan autokorrelaatioista ratkaisemalla ns. Yule-Walkerin yhtälöt

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \tag{6}$$

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \tag{7}$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \tag{8}$$

Osittaisautokorrelaatio viiveellä ρ kuvaa sitä, kuinka ρ ajanhetkeä taaksepäin oleva suureen arvo selittää tämänhetkistä suureen arvoa, kun muiden viiveiden selittävä osuus on poistettu. Selvästi nollasta poikkeavat osittaisautokorrelaatiot viiveeseen ρ_1 asti viittaavat siihen, että AR-osan kertaluku on ainakin ρ_1. Yksittäiset nollasta poikkeavat osittaisautokorrelaatiot puolestaan voivat liittyä kausivaihteluun ja johtaa multiplikatiiviseen malliin.

1.3 Mallin diagnostiset tarkistukset

Mallin testaus on olennainen osa mallitusprosessia. Mallin käyttötarkoitus vaikuttaa siihen, millainen testaus on järkevää. Jos mallia on tarkoitus käyttää ennustamiseen, tutkitaan *ex post-* ja mahdollisuuksien mukaan myös *ex ante*-ennustuksien hyvyyttä. Mallin hyvyyden arvioimiseen voidaan käyttää myös erilaisia tilastollisia testejä. On syytä aina tarkistaa, onko mallin generoima kohina riittävän satunnaista ja korreloimatonta. Residuaalien autokorrelaatiofunktion visuaalinen tarkastelu kertoo kohinan valkoisuudesta. Lisäksi käytetään nk. Portmanteau-testiä sen testaamiseksi, ovatko kohinan autokorrelaatiot riittävällä tarkkuudella nollia, eri K:n arvoilla. Testisuure on Q, jossa n on aikasarjan havaintojen lukumäärä ja $\hat{r}_i$ on residuaalien autokorrelaatio.

$$Q_k = \frac{n \cdot (n+2)}{n-k} \cdot \sum_{i=1}^{k} \hat{r}_i^2 \tag{9}$$

1.4 Aikasarjamallin rakentaminen Box-Jenkinsin menetelmällä

1. Identifioidaan kertaluvut.

2. Estimoidaan parametrit.

3. Suoritetaan diagnostiset tarkistukset. Jos malli ei ole hyvä, palataan kohtaan 1.

4. Käytetään mallia ennustamiseen, säätöön jne. Jos malli ei ole hyvä, palataan kohtaan 1 tai harkitaan jonkin muun malliluokan tai menetelmän käyttöönottoa.

2 Yrityksen sähkön hankinta

Monet suuret sähkön käyttäjät hankkivat sähkönsä useista eri lähteistä. Vapailla markkinoilla yrityksille on tärkeää kyetä ennustamaan sähkön kysyntä ja tuleva hintakehitys. Riskejä minimoidaan erilaisilla sopimuksilla. Laboratoriotyön kohdassa 3.4 esitellään tyypillisiä sähkökaupan sopimuksia.

Suomalaisilla sähkömarkkinoilla energiaa voidaan ostaa pohjoismaisesta sähköpörssi Nord Poolista, riippumattomilta sähkömeklareilta ja pienemmistä markkinapaikoista (esim. Voimatori Oy). Nord Poolissa varsinkin pienemmät yritykset käyvät kauppaa meklarien välityksellä.

Jokaisen sähkön käyttäjän (asiakas, sähkön myyjä, verkkoyhtiö jne.) pitää ilmoittaa viranomaisille taho, joka viime kädessä vastaa omilla resursseillaan kunkin sähkön käyttötunnin kysynnän ja tarjonnan tasapainosta. Kunkin asiakkaan ennakkohankintojen ja käyttötuntien toteutuneen kulutuksen välisen erotuksen toimittaa asiakkaalle hänen valitsemansa ns. avoin toimittaja, joka yleensä on sähkön myyntiyhtiö. Sähkön myyntiyhtiöiden taseesta vastaavaa sähköntuottajaa kutsutaan nimellä säätövastaava. Suomen kokonaissähkötaseen tasapainottamiseen tarvittavan säätösähkön markkinapaikkana on kantaverkkoyhtiö Fingrid. Myös kantaverkossa hintataso määräytyy kysynnän ja tarjonnan mukaan siten, että kukin tuottaja tai markkinaosapuoli, jolla on ylimääräistä kapasiteettia, tarjoaa kantaverkolle tuotantoresurssejaan erilaisina kokonaisuuksina sopivaksi katsomallaan hinnalla, ja kantaverkko hankkii sähköä aina halvimmista lähteistä.

3 Laboratoriotyö

Tässä työssä käytetään aikasarjamallinnusta erään yrityksen sähkönkulutuksen mallintamiseen. Muodostetulla mallilla ennustetaan seuraavan vuorokauden kulutus ja ennusteeseen perustuen optimoidaan sähkönhankinta.

3.1 Esiselostustehtävä

Tarkastellaan prosesseja

(i) $$w_t = (1 - \theta B)(1 - \Theta B^s)a_t, \;\; s \geq 3$$

(ii) $$(1 - \Phi B^s)w_t = (1 - \theta B)(1 - \Theta B^s)a_t, \;\; s \geq 3$$

missä $w_t = \nabla_s z_t$.

Millä viiveen k arvoilla prosessin autokorrelaatiofunktio $r_k = \frac{\gamma_k}{\gamma_0}$, missä $\gamma_k = E[w_t w_{t+k}]$, saa nollasta poikkeavia arvoja?

Perustele vastauksesi tarkastelemalla autokovarianssin γ_k lauseketta.

Piirrä autokorrelaatiofunktiot, kun $s = 24$, $\theta = -0.8$, $\Theta = -0.3$, $\Phi = 0.5$.

Johdettujen teoreettisten autokorrelaatiofunktioiden tarkistamisessa ja piirtämisessä auttaa R:n ARMAacf- ja plot-funktiot. Esimerkit funktioiden käytöstä löytyy työselostuksen lopusta kohdasta 3.6.3.

3.2 Mallin rakentaminen

Rakenna yrityksen sähkönkulutukselle hyvä SARIMAX-malli (Seasonal ARIMA with eXternal variable), missä mahdollisena ulkoisena selittäjänä on lämpötila. Käytä mallin rakentamiseen marras-joulukuulta 2014 mitattua yrityksen sähkönkulutusdataa. Tutki aikasarjan stationaarisuutta ja kausivaihteluita. Esitä perustelut valitsemillesi differoinnille ja kausivaihteluiden pituuksille. Tutki differoidun aikasarjan korrelaatiofunktioita. Valitse korrelaatiofunktioiden perusteella AR- ja MA-osien kertaluvut. Kokeile erilaisia vaihtoehtoja ja mieti syitä mallien ennustuskyvyn hyvyyteen. Perustele kaikissa vaiheissa tekemäsi ratkaisut.

Koska lämpötilasignaali on jatkuvasti herättävä noin kertalukua yksi, lämpötilariippuvuus mallitetaan lineaarisena αT_{ulko}. Pohdi kuitenkin, vaikuttaako ulkolämpötila viiveellä ja mistä mahdollinen viive voi johtua. Tutki onko aikasarja oleellisesti nollakeskiarvoista vai tarvitaanko vakiotermi.

Vertaa mallin antamia ex post -ennusteita toteutuneeseen kulutukseen. Analysoi poikkeamia ja tarkastele keskivirhettä. Käytä diagnostisia testejä (esim. Ljung-Boxin χ^2-testi) sekä korrelaatiofunktioita sen selvittämiseen, ovatko residuaalit valkoista kohinaa eli onko kaikki datan informaatio saatu mallitettua.

3.3 Ennuste sähkönkulutukselle

Tee valitsemallasi mallilla ennuste havaintojen jälkeisen vuorokauden (to 4.12.2014) sähkön kulutukselle. Kukin pari palauttaa ennustamansa sähkönkulutuksen ja sen 95 %:n luottamusvälit Excel-tiedostona (enne99.xls, missä 99 on työparin numero) kurssin MyCourses-sivuille. Assistentti vertaa kunkin ryhmän ennustamaa sähkönkulutusdataa toteutuneeseen kulutukseen. Vertaus tehdään sekä graafisesti että laskemalla neliövirhe.

Tässä malli perustuu usean viikon ehjään (ei puuttuvia havaintoja) dataan, mikä käytännössä on harvinaista, ja mallilla ennustetaan yksi vuorokausi eteenpäin. Mitä olisi otettava huomioon, jos mallilla haluttaisiin ennustaa pidemmälle tulevaisuuteen? Mihin olisi erityisesti kiinnitettävä huomiota, jos haluttaisiin ennustaa sähkönkulutusta seuraavalle viikonlopulle la 6.12 - su 7.12? Esitä ehdotuksesi siitä millaisia vaihtoehtoisia lähestymistapoja voitaisiin soveltaa sähkönkulutuksen ennustamiseen.

3.4 Yrityksen sähkönhankintastrategia

Laadi strategia kyseisen yrityksen sähköenergianhankinnalle 4.12.2014 klo 00.00-24.00. Olet siis yrityksesi energiajohtaja. Eletään keskiviikkoa 3.12.2014. Kello on tasan 12 ja olet juuri palannut ruokatunniltasi. Seuraavan päivän sähkönhankintastrategia on ilmoitettava kaikille sähkömarkkinaosapuolille seuraavan tunnin kuluessa. Sinun on hankittava kullekin sähkön käyttötunnille tarvittava määrä sähköä mahdollisimman halpaan hintaan. Yrityksesi toimitusjohtaja on sitonut henkilökohtaiset optiosi vuoden energiahankinnan onnistumiseen.

Useissa sopimuksissa erotetaan päiväaika ja yöaika. Päiväaika käsittää (arkisin ma-pe) tunnit 7.00-22.00 ja loppu on yöaikaa. Kaikissa hinnoissa on mukana kustannukset eli tarvitsemasi sähkö toimitetaan mainittuun hintaan Suomen kantaverkkoon. Sähköä voit ostaa kolmesta lähteestä:

1. Yrityksesi käy kauppaa pohjoismaisessa Nord Poolissa, jonka Elspot-markkinoilta voit hankkia tarvitsemasi määrän sähköä seuraavalle vuorokaudelle. Huomaat, että seuraavana päivänä käydään paljon pieniä kauppoja ja hintatasosta ollaan yksimielisiä. Niinpä pörssin tunneittaiset hinnat 4.12.2014 voidaan olettaa jo kiinteiksi, kunhan ostosi/myyntisi ovat alle 100 000 kWh kutakin vuorokauden tuntia kohden. Ostettaessa hintoihin on lisättävä välityspreemio, joka on 0,009 snt/kWh. Nord Pooliin myytäessä hinnoista on vähennettävä sama preemio. Tiedostossa pörssihinnat.csv on pörssin tunneittaiset sähkön hinnat. Pohdi lyhyesti selostuksessa, pitäisikö sähkönhankintastrategiaa jollain tavalla muuttaa, jos hinnoissa olisi epävarmuutta.

2. Vaihtoehtoisesti voit ostaa sähköä Nord Poolissa toimivalta meklarilta. Meklari tarjoaa sinulle seuraavaksi päiväksi blokkituotetta, jossa ostat kunakin päivän tuntina vakiomäärän x_p ja kunakin yön tuntina vakiomäärän x_y. Päivähinta on 10,00 snt/kWh ja yöhinta 6,00 snt/kWh. Maksimissaan saat ostaa 50 000 kWh per tunti. Sinulla on siis kaksi päätettävää muuttujaa x_p ja x_y ja kumpikin on rajoitettu välille $[0, 50\,000]$. Meklari ei osta sinulta sähköä. Mikäli päädyt ostamaan meklarilta sähköä ($x_p + x_y \neq 0$), on yrityksesi maksettava sähkön hinnan lisäksi meklarisopimuksen sopimuspreemio, joka on kiinteät 2 000 EUR riippumatta ostamasi sähkön määrästä.

3. Yritykselläsi on myös kiinteä sopimus sähkömyyntiyhtiö SVK:n kanssa. SVK:n sopimukseen on neuvoteltu seuraavat hinnat ja määrät. Perushintaa sovelletaan, kun ostat vähemmän kuin 40 000 kWh tunnissa. 40 000 kWh:n ylimenevästä osasta maksat huippuhintaa. Maksimaalinen ostoteho on 70 000 kWh tunnissa. Yöllä perushinta on 5,85 snt/kWh ja huippuhinta 6,06 snt/kWh. Päivällä perushinta on 10,08 snt/kWh ja huippuhinta 10,23 snt/kWh.

Optimoi sähkönhankintasi Excelin Solverilla tai muulla vastaavalla optimointityökalulla. Kun olet optimoinut hankintasi vastaamaan ennustamaasi kysyntää, ilmoita tunneittaiset ostosi ja myyntisi sekä tunneittaiset ostosi meklarilta ja SVK:lta. Ilmoitus tehdään sähkömarkkinoiden ennakkoilmoituslomakkeella (Excel-pohja edi99.xls). Lataa tiedosto kurssin MyCourses-sivuille.

Ennakkohankintasi ja kunkin käyttötunnin toteutuneen kulutuksesi välisen erotuksen tasaajaksi olet valinnut Assivoima Oy:n. Jos olet hankkinut sähköä enemmän kuin yrityksesi tarvitsee

4.12.2014, Assivoima Oy ostaa ylimääräisen sähkön takaisin päiväaikaan hintaan Nord Poolin tuntihinta - 4% ja yöaikaan hintaan Nord Poolin tuntihinta - 3%. Vastaavasti jos olet hankkinut sähköä liian vähän, Assivoima Oy myy sinulle sähköä päiväaikaan hintaan Nord Poolin tuntihinta + 5% ja yöaikaan hintaan Nord Poolin tuntihinta + 4%. Jos et määräaikaan mennessä tilaa sähköä muilta toimittajilta, joudut maksamaan sähköstäsi yllämainitut hinnat. Avoimelta toimittajalta ostettava sähkö on siis kalliimpaa kuin etukäteen tilattu sähkö, mutta toisaalta saat keskimäärin ostohintaa huonomman hinnan myydessäsi ylijäävää sähköä. Joudut punnitsemaan vastakkain toisaalta sitä, että etukäteen ostamasi sähkö ei riitä ja toisaalta sitä, että olet etukäteen hankkinut liikaa sähköä, joista kummastakin koituu ylimääräisiä kustannuksia. Pohdi, millaista riskiasennetta valitsemasi sähkönhankintastrategia kuvastaa.

Assistentti laskee kunkin työparin sähkönhankintastrategian kustannukset, kun otetaan huomioon, että sähköä on hankittava toteutuneen kulutuksen verran. Puuttuva sähkö siis ostetaan kantaverkosta ja ylijäävä sähkö myydään kantaverkkoon.

3.5 Työselostus

Jokainen ryhmä palauttaa **esitehtävän ratkaisun** ja **työselostuksen** paperilla kurssin lokerikkoon. Sähkönkulutuksen ennuste (**enne99.xls**) ja sähkönhankintastrategia (**edi99.xls**) ladataan kurssin MyCourses-sivuille.

Työselostuksen tulee sisältää vähintään:

1. kansilehti
2. sisällysluettelo
3. johdanto
4. aikasarjan analyysi
   - aikasarjan ja sen korrelaatiofunktioiden analyysi
   - perustelut differoinneille
   - perustelut mallin parametrien valinnoille
   - selostus siitä, millaisia malleja kokeilitte, mitä jätitte pois ja mihin lopulta päädyitte
   - diagnostinen testaus valitulle mallille ja perustelut sille, miksi valittu malli on parempi kuin muut
   - kuvia, jotka tukevat tekemiänne ratkaisuja
   - seuraavan vuorokauden ennustettu kulutus
5. sähkönhankintasuunnitelma ja selvitys siitä, miten kyseiseen suunnitelmaan on päädytty, sekä perustelut, miksi ratkaisu on optimaalinen

R-koodia ei tule palauttaa, vaan kaikki mitä on tehty tulee käydä ilmi työselostuksesta.

4 Tiedostot

Tiedostossa sahko.csv on 816 sähkönkulutusmittausta kilowattitunteina (kWh) aikajaksolta 31.10.-3.12.2014 ja 840 ulkolämpötilaa aikajaksolta 31.10.-4.12.2014. Lämpötilahavaintoja on enemmän, sillä aikasarja sisältää torstain havainnot, joita käytetään ennustamiseen. Tiedot ovat tiedostossa riveittäin järjestyksessä sähkönkulutus ja lämpötila, dataerottimena puolipilkku (;).

Tiedostossa pörssihinnat.csv on NordPoolin tunnettaiset hinnat. Data on eroteltu puolipisteellä (;).

esimerkki.r sisältää esimerkin useamman kausidifferoinnin mallin rakentamisesta.

Tiedostossa pohja.r on joukko R-komentoja, joilla käsitellään kulutustietoja ja piirretään kuvia. Voit muunnella pohjaa tarpeen mukaan. Osa komennoista on pantu kommentteihin, joita merkitään R:ssä #-merkeillä. Poista käskyjä kommenteista jos ja kun tarvitset niitä.

Excel-tiedostoissa enne99.xls ja edi99.xls on kummassakin keltainen ruutu, johon kukin työpari täydentää työparin numeron kahdella numerolla, siis pari 3 kirjoittaa 03. Enne-tiedostoon täydennetään myös parin nimet ja opiskelijanumerot. Tiedostot tallennetaan nimillä enne99.xls ja edi99.xls, jossa 99 on parin numero. Enne-tiedostoon täydennetään parin ennuste luottamusväleineen ja edi-tiedostoon sähkönhankintastrategia. Tiedostot palautetaan kurssin MyCourses-sivuille.

5 R-ohjelma

R on avoimeen lähdekoodiin perustuva tilastolliseen analyysiin soveltuva ohjelmointikieli. Tässä työssä käytetään R:n valmiita funktiota mallien rakentamiseen, estimointiin ja ennustamiseen.

R-ohjelmia voi luoda ja muokata valitsemalla R:n file-valikosta New Document tai Open Document. Ohjelman voi ajaa valitsemalla file-valikosta Source File. R:n käytössä kannattaa turvautua R:n omaan dokumentaatioon ja esimerkiksi verkosta löytyvään R:n aikasarja-analyysioppaaseen (http://www.stat.pitt.edu/stoffer/tsa3/).
Hyödyllisiä R:n funktiota:

- `help`
  - Komennolla help(*funktion nimi*) saat kuvauksen funktiosta ja sen vaatimista syötteistä. Esimerkiksi komento `help(arima)` avaa R:n dokumentaatiosta arima-funktiota käsittelevän sivun.

    R:n dokumentaatiossa on määritelty, mitä parametreja funktioille pitää syöttää, kuvattu miten ne toimivat, sekä yleensä annettu muutamia esimerkkejä funktioiden käytöstä.
- `arima`
  - R:n arima-funktiolla estimoidaan arima-malleja. Funktiolle voi antaa myös syötteeksi ulkoisia regressiomuuttujia, jolloin funktio estimoi SARIMAX-mallin. Katso R:n dokumentaatiosta tai komennolla `help(arima)` R:n käyttämä notaatio ARIMA-malleille ja lisäohjeet arima-funktion käytölle. Esim.

    ```
    malli <- arima(aikasarja, order = c(3,0,1), seasonal = list(order = c(1,0,0),
    period = 12)
    ```

    estimoi SARIMA(3,0,1)x(1,0,0) mallin kausivaihtelun pituudella 12 ja tallentaa sen nimellä malli. Esimerkiksi mallin residuaaleihin päästään käsiksi komennolla `malli$residuals`. Parametrien estimaatteja voi tarkastella kirjoittamalla muuttujan nimi `malli` R:n komentoriville. `$` viittaa taulukkomuotoisen muuttujan kenttään.

- `predict`
  - R:n predict-funktiolle annetaan syötteeksi arima-funktion estimoima malli. Funktio laskee ennusteen ja ennusteen keskipoikkeaman. Lisäohjeita predictin käyttöön saa taas komennolla `help(predict.Arima)` (huomaa iso kirjain). Esim. jos malli on aiemmin estimoitu ARIMA-malli, komento

    ```
    enne <- predict(malli, n.ahead = 24)
    ```

    ennustaa aikasarjaa mallin perusteella 24 aika-askelta eteenpäin. Ennusteisiin päästään käsiksi kirjoittamalla `enne$pred`. Ennusteen keskivirhe puolestaan saadaan kirjoittamalla `enne$se`.
- `acf`
  - acf funktiolla voidaan laskea ja piirtää autokorrelaatio- ja osittaisautokorrelaatiofunktioiden kuvat. Esim.

    ```
    acf(sahko, lag.max = 168, type = "partial")
    ```

    laskee aikasarjan "sahko" osittaisautokorrelaatiofunktion viiveeseen 168 asti ja piirtää kuvan.
- `diff`
  - diff-komennolla voidaan differensoida aikasarjoja. Esim.

    ```
    diffsahko = diff(sahko, lag = 1, differences = 2)
    ```

    differensoi sahko-aikasarjan viiveellä 1 kaksi kertaa ja tallentaa differensoidun aikasarjan muuttujaan diffsahko.
- `ts`
  - ts-komennolla luodaan vektoreista aikasarjaobjekteja, joita esimerkiksi arima-funktio käyttää syötteenään. Esim.

    ```
    ts(vektori, start = c(2014, 3), frequency = 12)
    ```

    luo vektorista aikasarjan, jonka jakso on 12 kuukautta ja alkaa vuoden 2014 maaliskuussa. Parametriksi start on annettu vektori komennolla `c( . . . )`.
- `ts.plot`
  - ts.plotilla voidaan piirtää aikasarjoja. Esimerkiksi jos "malli" on arima-funktiolla sovitettu malli ja "enne" on predict-funktiolla mallista saatu ennuste (kts. esimerkit yllä), voidaan sovite, ennuste ja 95 %:n luottamusvälit (1,96 keskipoikkeamaa) piirtää komennolla

    ```
    ts.plot(fitted(malli), enne$pred, enne$pred + 1,96*enne$se, enne$pred - 1,
    96*enne$se, col = c("black", "blue", "blue", "blue"))
    ```

    Parametriksi `col` on annettu vektori, joka määrää aikasarjojen värit piirrettävässä kuvassa. Dollarimerkillä päästään käsiksi ennusteisiin (pred) sekä estimoituhin keskivirheisiin (se).

- `Box.test`
  - Esimerkiksi komennolla

    ```
    Box.test(res, lag = 12, type = "Ljung-Box", fitdf = p+q)
    ```

    voi laskea Ljung-Boxin testisuureen (Portmanteau-testi) 12 viipeeseen asti mallille, jonka AR- ja MA-osien parametrit ovat p ja q, ja res on mallin antaman sovitteen residuaalien aikasarja.

- `ARMAacf`
  - ARMAacf laskee annetun mallin teoreettisen autokorrelaatiofunktion.

    Esimerkiksi SMA-mallin $w(t) = a_t - \theta_{t-1}a_{t-1} - \theta_{t-7}a_{t-7} - \theta_{t-8}a_{t-8}$ autokorrelaatiofunktio saadaan laskettua komennoilla

    ```
    ma1 = c(-theta1, rep(0,5), -theta7, -theta8)
    akf = ARMAacf(ar = 0, ma = ma1)
    ```

    Huomaa, että ensimmäisellä komennolla luodaan vektori $[-\theta_1\ 0\ 0\ 0\ 0\ 0 - \theta_7 - \theta_8]$. Komento `rep(0,5)` siis vain luo viisi peräkkäistä nollaa. SMA-mallia voidaan tässä käsitellä MA-mallina, jossa puuttuvat parametrit asetetaan nolliksi. Kiinnitä huomiota tapaan jolla malli on määritelty, jotta parametrien etumerkit menevät oikein päin.