

Ekonometrian
kurssin
(721339S/805683S/805339A)
luentomuistiinpanot

Syksy 2008

M. Rahiala

1 MUUTTUJIEN VÄLISTEN VAIKUTUSTEN KUVAAMINEN TILASTOLLISTEN MALLIEN AVULLA

1.1 Tilastollisen mallittamisen peruseriaatteet

Ajatellaan esimerkiksi, että halutaan tutkia jotakin taloudellisten toimijoiden (yritysten, kotitalouksien, rahoituslaitosten tms.) käyttäytymiseen liittyvää ilmiötä ja että puhtaasti talousteoreettisten näkökohtien *lisäksi* on olemassa aikaisempaa kokemusta toimijoiden reagoititavoista ympäröivissä olosuhteissa tapahtuneisiin muutoksiin. On selvää, että kaikille mahdollisille reagoititavoille voidaan aina löytää äärettömän monia erilaisia selitystapoja, joten *pelkkien* aikaisempien *havaintojen* perusteella emme voi päätellä paljoakaan. Tästä syystä on välttämätöntä aluksi rajata jonkinlainen *ajatuskehikko* (malli), jonka puitteissa havaintoihin sisältyvää informaatiota yritetään tulkita. Tämän ajatuskehikon valinta ja rajaaminen kannattaa aina perustaa johonkin järkevältä tuntuvaan talousteoriaan, mikäli sellainen suinkin on olemassa tai kehitettävissä.

Joissakin tapauksissa on mahdollista kerätä tietoja *yksittäisten* talouden toimijoiden käyttäytymisestä, kun taas toisissa tapauksissa joudutaan tyytymään koko maata tai kansantaloutta koskeviin *aggregoituihin* tietoihin. Edellisessä tilanteessa puhutaan *paneeliaineistosta* ja jälkimmäisessä tilanteessa *aikasarja-aineistosta*. On tietenkin selvää, että paneeliaineistot ovat aggregoituja aineistoja informatiivisempia, mikäli seurannan kohteena ollut paneeli on riittävän suuri ja mikäli sitä voidaan pitää kaikkia vastaavia talouden toimijoita hyvin edustavana. Tällä kurssilla keskitytään pääasiassa aikasarja-aineistojen tutkimiseen, mutta myös paneeliaineistojen analysointia käsitellään jonkin verran.

Olipa käytettävissä olevan havaintoaineiston tyyppi mikä tahansa, liittyy havaintoihin aina niin monenlaista vaihtelua, että ”deterministiset” käyttäytymismallit eivät koskaan ole käyttökelpoisia, vaan osa havaintoihin sisältyvästä vaihtelusta joudutaan aina tulkitsemaan ”satunnaiseksi”. Tästä syystä joudutaan turvautumaan ns. *tilastollisiin* malleihin, joissa havaintoihin sisältyviä, satunnaisiksi tulkittuja vaihtelukomponentteja kuvataan *matemaattisilla satunnaismuuttujilla*, joiden *jakaumien* muodot mallissa spesifoidaan. Ajatellaan nyt, että kaikki käytettävissä olevat havainnot on koottu *havaintomatriisiksi* \tilde{Z} . Havaintomateriaalin tuottaneelle ilmiölle ”kuvaksi” (matemaattiseksi malliksi) pyritään täsmentämään jokin matemaattinen satunnaismuuttuja $Z = \text{vec}(\tilde{Z}')$ ja siihen liittyvä todennäköisyysjakauma F_Z . (Tässä F_Z toimii satunnaismuuttujan Z *kerätymsymbolina* ja operaattorisymboli ”vec” tarkoittaa matriisin pystyrivit allekkain latovaa operaattoria. Käytetyt merkinnät on määritelty tarkemmin liitteessä A1.) Etukäteen on yleensä pakko varautua moniin vaihtoehtoihin malleihin, joista sitten myöhemmin pyritään valitsemaan ”havaintojen valossa parhaalta näyttävä” yksittäinen malli. Rajataan aluksi ns. *malliperhe*, johon pyritään ottamaan mukaan kaikki ”potentiaalisesti kelvolliset” (talousteoreettisesti järkevät) mallit. Mikäli malliperhe onnistutaan rajaamaan niin tiukasti, että se on samaa mahtavuutta kuin \mathcal{R}^p , voidaan puhua *parametrisesta* mallista ja malliperhe voidaan kirjoittaa formaalisti muotoon

$$\mathcal{F} = \{F_Z^{(\theta)} \mid \theta \in \Theta\} \quad , \quad (1.1)$$

jossa θ edustaa ns. *parametria* (jakauman ”nimilappua” perheen sisällä) ja jossa $\Theta \subseteq \mathcal{R}^p$. Huomattakoon, että sama malliperhe voidaan aina parametroida äärettömän monella eri tavalla. Jos malliperheen kaikki jakaumat ovat keskenään erilaisia, voidaan ajatella havaintoaineiston Z sisältävän *informaatiota* parametrissa θ sikäli, että eri θ :n arvoihin liittyy erilainen käsitys Z :n käyttäytymisestä. Tämän informaation mittaamista ja hyödyntämistä pohtivaa tieteenalaa kutsutaan *tilastotieteeksi*. (Kaikkia tällä kurssilla esiteltäviä päätelyperiaatteita voidaan soveltaa myös muihin sovellusaloihin liittyvien ongelmien yhteydessä, mutta esiteltävä mallityyppivalikoima on valittu erityisesti taloudellisten ilmiöiden kuvaamista ajatellen.)

Havaintoaineiston Z johdannaisia $T = G(Z)$ sanotaan *tunnusluvuiksi*.

Huomautus 1.1: Jos tunnusluku $T = G(Z)$ on *kääntäen yksikäsitteinen* transformatio havaintomateriaalista Z , on syytä ajatella T :n ja Z :n sisältävän *saman* informaation, koska Z voidaan aina rekonstruoida T :stä. Toisaalta, jos T :n jakauma ei riipu θ :sta lainkaan, ei T myöskään sisällä *lainkaan* θ :aa koskevaa *informaatiota*. ■

Määritelmä 1.1: Havaintoaineiston $Z = z$ määräämällä *likelihood-funktiolla* $L_z : \Theta \rightarrow \mathcal{R}^1$ tarkoitetaan funktiota

$$L_z(\theta) = \begin{cases} f_Z^{(\theta)}(z) & \text{jos } Z \text{ on jatkuva} \\ P^{(\theta)}(Z = z) & \text{jos } Z \text{ on diskreetti} \end{cases} \quad (1.2)$$

Mikäli havaintoaineisto edustaa em. perustyyppien sekamuotoa, voidaan määritelmä laajentaa suoraviivaisesti tähänkin tilanteeseen.

Jos *likelihood-funktio* $L_z(\theta)$ maksimoidaan θ :n suhteen, nähdään, mikä perheen jakaumista (malleista) olisi ”parhaiten sopuinnussa” aineiston $Z = z$ kanssa. Jos tätä maksimipistettä merkitään symbolilla $\hat{\theta}(z)$, saadaan samalla päätelysääntö, joka liittyy jokaiseen potentiaaliseen havaintoaineistoon Z ”uskottavimman” jakaumatunnuksen $\hat{\theta}(Z)$. Tätä havaintoaineiston funktiota kutsutaan parametrin θ *maximum likelihood-estimaattoriksi* (ML-estimaattoriksi). ■

Huomautuksessa 1.1 esitetyn periaatteen mukaan *likelihood-funktiossa* oleellista on vain sen *muoto*, eivätkä parametrin θ suhteen vakiona pysyvät kerrointermit saa millään tavalla päästä vaikuttamaan *likelihood-funktion* perusteella vedettäviin johtopäätöksiin.

Huomautus 1.2: Kuten edellä jo todettiin, jakaumaperhe $\mathcal{F} = \{F_Z^{(\theta)} \mid \theta \in \Theta\}$ voidaan aina parametroida äärettömän monella tavalla, sillä \mathcal{F} voidaan yhtä hyvin kirjoittaa muotoon

$$\mathcal{F} = \{F_Z^{(\psi)} \mid \psi = H(\theta) \in \Psi = H(\Theta)\} \quad (1.3)$$

olipa $H : \Theta \rightarrow \Psi$ mikä kääntäen yksikäsitteinen kuvaus tahansa.

Likelihood-funktio muuntuu parametriformaation yhteydessä luonnollisesti muotoon

$$L_Z^*(\psi) = L_Z(H^{-1}(\psi)) \quad \psi \in \Psi \quad , \quad (1.4)$$

mutta itse *jakaumaperhe* \mathcal{F} pysyy siis *samana*.

Mikäli malliperheeseen sisältyy sama jakauma kahdella (tai useammalla) ”nimi-
lapulla” varustettuna (ts. mikäli $F_Z^{(\theta)} \equiv F_Z^{(\theta^*)}$ jollekin $\theta \neq \theta^*$), sanotaan, että
parametri θ ei ole *identifioituva* havaintojen perusteella. Tällöinhän havainnot
eivät nimittäin sisältäisi lainkaan sellaista informaatiota, joka auttaisi meitä te-
kemään eron parametriarvojen θ ja θ^* välillä. ■

1.2 Yksisuuntaisten vaikutusrakenteiden kuvaaminen: Regressiomallit ja niiden käyttöä koskevat rajoitukset

Ajatellaan nyt, että halutaan tutkia muuttujien $X_t = (x_{t1} \dots x_{tm})'$ vaikutuksia
muuttujaan y_t , ja että ollaan vakuuttuneita kyseisten vaikutusten *yksisuuntais-*
suudesta, ts. siitä, ettei y_t :n arvon muuttuminen lainkaan vaikuta muuttujien
 X_t käyttäytymiseen. (Nämä merkinnät viittaavat implisiittisesti siihen, että X_t
ja y_t olisivat aggregaattitason aikasarjoja, mutta havaintoyksikön tunnuksena
toimiva indeksi t voi viitata myös johonkin muuhun asiaan kuin ajankohtaan.
Alaindeksejä voi olla myös useampia kuin yksi; esimerkiksi paneeliaineistojen yh-
teydessä on usein luonnollista käyttää kaksinkertaista indeksointia, jossa ensim-
mäinen indeksi viittaa toimijaan ja jälkimmäinen ajankohtaan.) Jos tiedot näi-
den muuttujien käyttäytymisestä on kerätty ajankohdista (havaintoyksiköistä)
 $t = 1, \dots, n$, olisi luvussa 1.1 mainittu havaintomatriisi siis muotoa

$$\tilde{Z} = \begin{pmatrix} y_1 & x_{11} & \dots & x_{1m} \\ y_2 & x_{21} & \dots & x_{2m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ y_n & x_{n1} & \dots & x_{nm} \end{pmatrix} .$$

Monissa tapauksissa *yksisuuntaisia* vaikutusrakenteita voidaan onnistuneesti ku-
vata muotoa

$$\begin{aligned} y_t &= g(X_t, \beta) + \varepsilon_t \quad , \quad \{\varepsilon_t\} \perp\!\!\!\perp \{X_t\} \\ \varepsilon_t &\sim NID(0, \sigma^2) \quad , \quad t = 1, \dots, n \end{aligned} \tag{1.5}$$

olevilla *regressiomalleilla*. Tässä $\{g(\cdot, \beta)\}$ on jokin riittävän säännöllinen funk-
tioparvi, jonka tehtävänä on siis kuvata y_t :n *regressiofunktioita* X_t :n suhteen,
ts.

$$E(y_t | X_t = X) = g(X, \beta) .$$

(Huom.: Muuttujan X_t reunajakaumamallia ei välttämättä tarvitse spesifioida
lainkaan.)

Tuutin tapaus liittyy epäilemättä *lineaariseen* regressiofunktioon $g(X, \beta) = \beta'X$, jolloin voidaan johtaa eräitä hyvin tunnettuja, β :n estimointiin liittyviä optimaalisuustuloksia (esim. ns. Gauss-Markovin lause). Numeroidaan nyt lineaariseen regressiomalliin liittyvät tavanomaisimmat perusoletukset erikseen, jotta niihin olisi myöhemmin mahdollisimman helppo viitata:

$$y_t = \beta'X_t + \varepsilon_t \quad , \quad (1.6)$$

$$E(\varepsilon_t | X_t) \equiv 0 \quad , \quad t = 1, \dots, n$$

$$\text{cov}((\varepsilon_1 \quad \dots \quad \varepsilon_n)') = \sigma^2 I \quad (1.7)$$

$$\{\varepsilon_t\} \perp\!\!\!\perp \{X_t\} \quad (1.8)$$

$$\varepsilon_t \sim NID(0, \sigma^2) \quad (1.9)$$

Oletus (1.8) tekee tietenkin regressiomallien käytön täysin *mahdottomaksi* tilanteissa, joissa selittävien tekijöiden X_t ja vastemuuttujan y_t välillä esiintyy *vuorovaikutuksia*.

Huomautus 1.3: Kuten tunnettua, voidaan malli (1.6) – (1.9) esittää kaikkien havaintoyksiköiden osalta kompaktisti vektorimuodossa

$$Y = X\beta + \varepsilon \quad , \quad \varepsilon \sim N_n(0, \sigma^2 I) \quad (1.10)$$

jossa

$$Y = (y_1 \quad \dots \quad y_n)'$$

$$X = (X_1 \quad \dots \quad X_n)' = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \cdot & \dots & \cdot \\ x_{n1} & \dots & x_{nm} \end{pmatrix} \quad \text{ja}$$

$$\varepsilon = (\varepsilon_1 \quad \dots \quad \varepsilon_n)'$$

Ns. OLS-estimaattori (pienimmän neliösumman estimaattori)

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (1.11)$$

on oletusten (1.6) – (1.8) vallitessa Gauss-Markovin lauseen mukaan β :n ”paras” (MVU) estimaattori. Jos malliin liittyy myös oletus (1.9), on $\hat{\beta}$ samalla myös β :n ML-estimaattori. Perusmallin (1.6) – (1.9) puitteissa OLS-estimaattorilla (1.11) on tunnetusti seuraavat ominaisuudet:

$$\begin{aligned} \hat{\beta} &\sim N_m(\beta, \sigma^2(X'X)^{-1}) \quad , \\ E = Y - X\hat{\beta} &= (I - P_X)Y = (I - P_X)\varepsilon \sim N_n(0, \sigma^2(I - P_X)) \quad , \quad (1.12) \\ \hat{\beta} &\perp\!\!\!\perp E \end{aligned}$$

■

Huomautus 1.4: Palataan vielä hetkeksi regressiomallin (1.5) yleisempään muotoon, jossa regressiofunktio ei välttämättä olekaan lineaarinen. Esimerkkinä epälineaarisesta regressiomallista mainittakoon ns. kahden panoksen CES-tuotantofunktio (Constant Elasticity of Substitution)

$$y_t = \beta_1 \left[(1 - \beta_2)L_t^{-\beta_3} + \beta_2 K_t^{-\beta_3} \right]^{-\frac{\beta_4}{\beta_3}} + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad (1.13)$$

jossa

- $y_t \leftrightarrow$ tuotannon jalostusarvo kiinteisiin hintoihin periodilla t
- $L_t \leftrightarrow$ työpanos (tehdyt työtunnit) po. toimialalla
- $K_t \leftrightarrow$ pääomapanos
- $\varepsilon_t \leftrightarrow$ virhetermi

Mallia (1.13) ei selvästikään voida muuttaa lineaariseksi minkäänlaisten muuttujatransformaatioiden avulla.

Jos $\beta_3 \rightarrow 0$, saadaan (1.13):n rajatapauksena malli

$$y_t = \beta_1 L_t^{\gamma_1} K_t^{\gamma_2} + \varepsilon_t \quad , \quad (1.14)$$

jossa

$$\gamma_1 = (1 - \beta_2)\beta_4 \quad \text{ja} \quad \gamma_2 = \beta_2\beta_4 \quad .$$

Mallia (1.14) sanotaan Cobb-Douglas-malliksi. Mikäli se muotoiltaisiin virhetermin osalta multiplikatiiviseen muotoon

$$y_t = \beta_1^* L_t^{\gamma_1^*} K_t^{\gamma_2^*} \kappa_t \quad , \quad \log \kappa_t \sim NID(0, \sigma^2) \quad , \quad (1.15)$$

voitaisiin malli ”linearisoida” logaritmiseen skaalaan siirtymällä

$$\log y_t = \gamma_o^* + \gamma_1^* \log L_t + \gamma_2^* \log K_t + \varepsilon_t \quad (1.15')$$

jossa

$$\gamma_o^* = \log \beta_1^* \quad \text{ja} \quad \varepsilon_t = \log \kappa_t \sim NID(0, \sigma^2) \quad .$$

Sen sijaan malleja (1.13) ja (1.14) ei voida linearisoida vastaavilla tempuilla. Vaikka OLS-estimaattoreille ei epälineaaristen mallien (1.5) osalta voidakaan johtaa mitään MVU-optimaalisuusominaisuuksia, on niiden käyttö lähes yhtä luonnollista kuin lineaarisessakin tapauksessa. On helppo todeta, että virhetermien normaalisuutta koskevan oletuksen puitteissa regressioparametrien OLS-estimaattorit ovat samalla po. parametrien ML-estimaattoreita. Mallin (1.5) mukainen, havaintojen y_1, \dots, y_n määräämä likelihood-funktio on nimittäin muotoa

$$\begin{aligned} L_{y_1, \dots, y_n | X_1, \dots, X_n}(\beta, \sigma^2) &= \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(y_t - g(X_t, \beta))^2} \\ &= (2\pi)^{-\frac{n}{2}} \cdot (\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2}Q(\beta)} \end{aligned} \quad (1.16)$$

jossa

$$Q(\beta) = \sum_{t=1}^n (y_t - g(X_t, \beta))^2 \quad .$$

Neliösumman $Q(\beta)$ minimointi β :n suhteen johtaa siis likelihood-funktion (1.16) maksimointiin, joten β :n OLS-estimaattori on samalla myös ML-estimaattori täysin riippumatta siitä, onko malli lineaarinen (regressiofunktio muotoa $g(X_t, \beta) = \beta' X_t$) vai epälineaarinen. Linearisessa tapauksessa neliösumman $Q(\beta)$ minimoiva estimaattori on kirjoitettavissa eksplisiittiseen analyttiseen muotoon (1.11), kun taas epälineaaristen mallien osalta joudutaan turvautumaan numeerisiin optimointialgoritmeihin, joista enemmän luvussa 2.3. ■

Esimerkki 1.1: Tarkastellaan analyttikkojen julkistamia ennusteita Nokia-konsernin tuloksesta vuodelta 2001. Verrataan näitä ennusteita lopulliseen, seuraavassa huhtikuussa valmistuneesta tilinpäätöksestä ilmenevään (satunnaiseristä puhdistettuun) tulokseen, ja kutsutaan ennusteiden ja lopullisten tietojen välisiä eroja *ennustevirheiksi*. Ennustevirheet konvergoivat tietenkin periaatteessa kohti nollaa huhtikuun 2002 lähestyessä. Huomattakoon, että analyttikot seuraavat konsernin toimintaa *jatkuvasti*, mutta heidän odotustensa muutokset julkistetaan ainoastaan silloin tällöin, vaihtelevin väliajoin. Jos halutaan kuvata ennustevirheiden y_t konvergenssiprosessia, on tästä syystä välttämätöntä perustaa malli

jatkuva-akaiseen ajatteluun. Jos oletetaan, että $\{y_t\}$ -prosessin eteneminen riippuisi ainoastaan tilanteesta hetkellä t eikä lainkaan prosessin varhaisemmasta historiasta, olisi yksinkertaisin ajateltavissa oleva malli muotoa

$$\dot{y}_t = \frac{d}{dt}y_t = y_t \cdot \log \phi + \dot{w}_t \quad ,$$

jossa piste muuttujasymbolin yläpuolella tarkoittaa aikaderivaattaa ja ”virhetermin” roolissa olevan termin \dot{w}_t oletetaan käytökseltään muistuttavan ns. Brownin liikkeen (ks. määritelmä 8.6) aikaderivaattaa. Jos analyytikkojen odotusten julkistamisajankohtia merkitään symbolein t_1, \dots, t_n , olisi tällöin

$$E(y_{t_i} | y_{t_{i-1}}) = \phi^{(t_i - t_{i-1})} \cdot y_{t_{i-1}} \quad ,$$

joten päätyisimme (*havaintojen osalta*) malliin

$$y_{t_i} = \phi^{(t_i - t_{i-1})} \cdot y_{t_{i-1}} + \varepsilon_i \quad , \quad i = 2, \dots, n - 1 \quad ,$$

jossa $\varepsilon_i = w_{t_i} - w_{t_{i-1}}$. Vaikka tämä malli on *epälineaarinen* parametrin ϕ suhteen, on ϕ :n estimointi OLS-periaatteella yleisesti saatavilla olevien ohjelmistojen avulla erittäin helppoa:

PcGive:

```
Algebra code for Nokia01.in7:
ennustevirh = tulosennust-tulos;
aika = Vuosi+((kk-1)/12)+(pv-1)/365;
aikaero = aika-lag(aika,1);
```

```
EQ( 1) Modelling actual by NLS (using Nokia01.in7)
The present sample is: 2 to 16
```

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
&1	0.76897	0.39623	1.941	0.0727	0.2120

```
R2 = 0.885405 \sigma = 335.536 DW = 1.61
RSS = 1576178.152 for 1 variables and 15 observations
```

```
Code used for estimation:
actual=ennustevirh;
fitted=lag(ennustevirh,1)*(&1^aikaero);
&1=0.1;
```

Parametrin ϕ tulkintaa voi havainnollistaa esimerkiksi laskemalla ennustevirheen ”puoliintumisajan” $-\log(2)/\log(\hat{\phi}) = -\log(2)/\log(0.76897) \approx 2.6$ vuotta.

SAS:

```

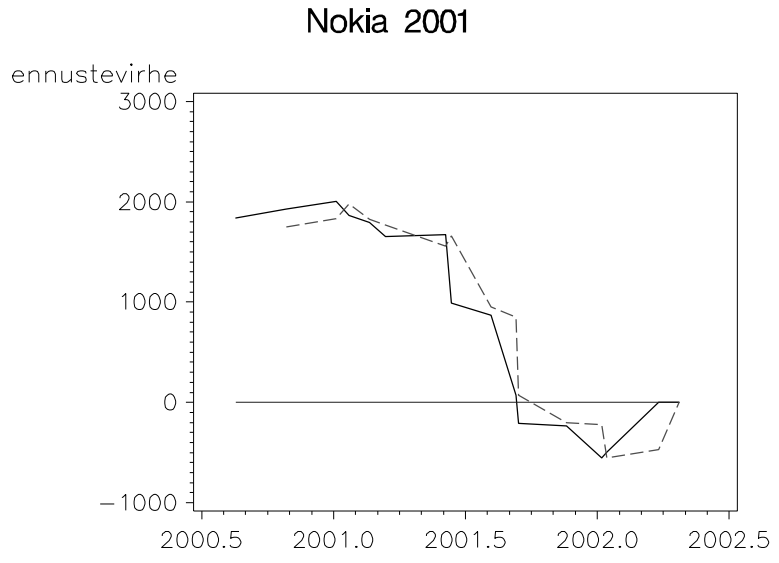
data jvaara;
  infile 'd:\vanhat\vaarala\puhdistetut\nokiapuh01.dat';
  input vuosi kk pv tulosenuste tot tulos ;
  aika=vuosi+((kk-1)/12)+(pv-1)/365; aikaero=aika-lag(aika);
  ennustevirhe=tulosenuste-tulos;
  edennustevirhe=lag(ennustevirhe); nolla=0; run;
proc nlin data=jvaara;
  parameters phi=0.1 ;
  model ennustevirhe=(edennustevirhe)*(phi**aikaero);
  output out=eipain predicted=ensov; run;
data eipain; merge eipain jvaara; run;
proc gplot data=eipain;
  title f=swiss 'Nokia 2001';
  plot (ennustevirhe nolla ensov)*aika / overlay ;
  symbol1 v=none i=join c=green l=1 w=2;
  symbol2 v=none i=join c=blue l=1 w=1;
  symbol3 v=none i=join c=red l=3 w=2; run;

```

The NLIN Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	1	20775350	20775350	184.53	<.0001
Error	14	1576178	112584		
Uncorrected Total	15	22351528			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
phi	0.7690	0.3894	-0.0662	1.6041



1.3 Vuorovaikutussuhteiden kuvaaminen ja siihen liittyvät ongelmat

Eräänlaisena johdantona moniyhtälömallien problematiikkaan tarkastellaan tietyn (hypoteettisen) hyödykkeen kysynnän ja tarjonnan samanaikaista mallittamista. Merkitään q_t^S :llä tarjonnan logaritmia periodilla t , q_t^D :llä kysynnän ja p_t :llä hinnan logaritmia periodilla t . Ajatellaan, että kysyntä määräytyy hintojen ja tarjonta hintojen ja mahdollisten muiden tekijöiden (mm. tuotantokustannusten) X_t perusteella seuraavasti:

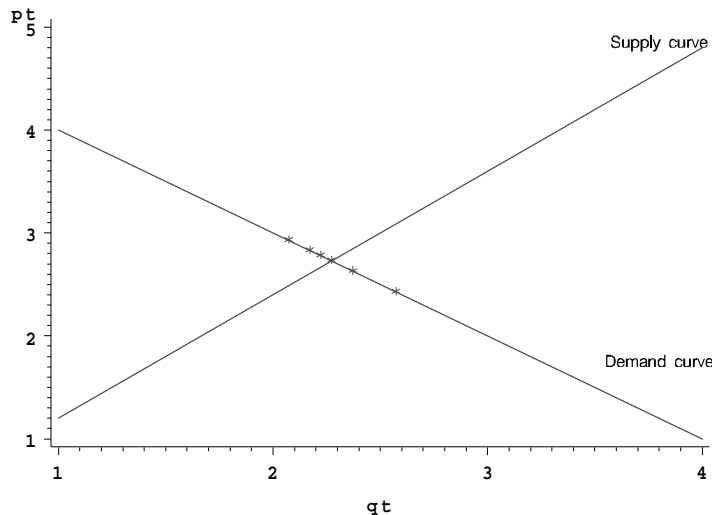
$$\begin{cases} q_t^S = \gamma_{11}p_t + \beta_{11} + \beta'_{12}X_t \\ q_t^D = \gamma_{21}p_t + \beta_{21} \end{cases} \quad (1.17)$$

(Asian yksinkertaistamiseksi tarkastellaan deterministisiä malleja, joista kysyntään ja tarjontaan sisältyvien stokastisten komponenttien kuvaus on jätetty kokonaan pois.)

Jos ajatellaan, että kysyntä ja tarjonta olisivat tasapainossa, olisi

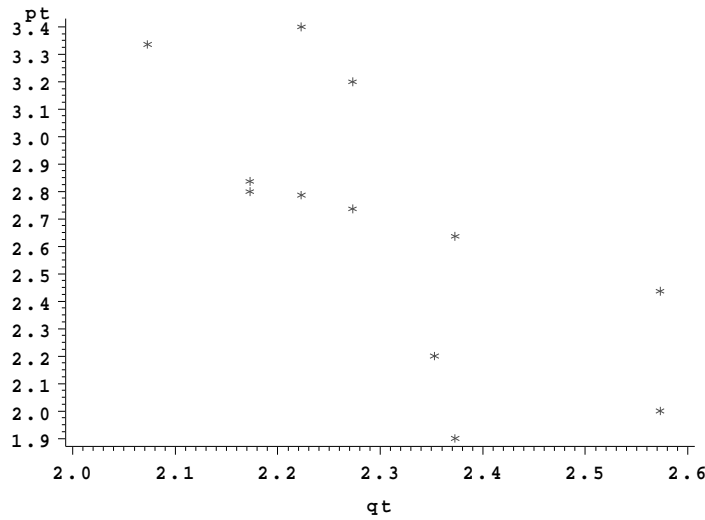
$$q_t^S = q_t^D \quad (1.18)$$

Yhtälöt (1.17) ja (1.18) muodostavat eräänlaisen systeemikuvauksen muuttujien $q_t = q_t^S = q_t^D$ ja p_t vuorovaikutuksista, kun taas X_t -muuttujien ajatellaan määräytyvän systeemin ulkopuolella, ts. niitä pidetään *eksogeenisina*. (Mikäli olisi voimassa hintasäännöstely, voitaisiin p_t :täkin ehkä pitää eksogeenisena.)



Koska kysyntäkäyrä pysyy paikallaan, se voidaan löytää seuraamalla tasapainopisteen liikkeitä eri X_t :n arvoilla. Se sijaan tarjontakäyrän kulmasta kysyntäkäyrään nähden ei saada lainkaan informaatiota, koska havaitut tasapainopisteet aina sijaitsevat kysyntäkäyrällä. Jopa deterministisen systeemin yhteydessä voi siis syntyä identifioituvuusongelmia! ■

Välihuomautus 1.5: Jos mallin (1.7) yhtälöihin liitettäisiin myös stokastiset virhetermit, saattaisi $(p_t \ q_t)'$ -havaintojen plottauskuvio olla seuraavan näköinen:



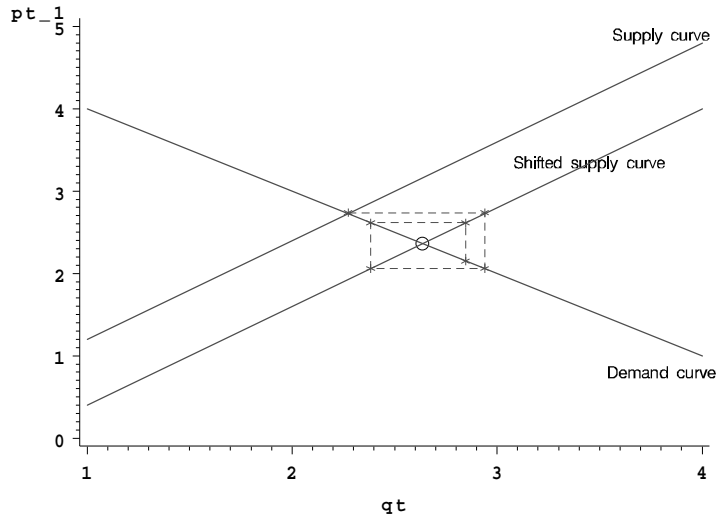
Tällöin voisi tulla mieleen ajatus ”selittää” toisaalta p_t :tä q_t :llä ja toisaalta q_t :tä p_t :llä, jolloin OLS antaisi *kaksi* eri sovitesuoraa. Näitä ei kuitenkaan missään nimessä pidä ruveta tulkitsemaan kysyntä- ja tarjontasuoriksi, sillä edellä esitetyn mukaisesti havainnot eivät sisällä *lainkaan* informaatiota tarjontasuoran kulmakertoimesta. Kyseessä on vain OLS-menetelmän *tekninen* ominaisuus, eikä *kummallakaan* estimoidulla suoralla itse asiassa ole tulkinnallista merkitystä. ■

Mikäli mallia (1.17) muutettaisiin siten, että tarjonnan ajateltaisiin reagoivan hintojen muutoksiin yhden aikayksikön viiveellä, päädyttäisiin malliin

$$\begin{cases} q_t^S = \gamma_{11}p_{t-1} + \beta_{11} + \beta_{12}'X_t \\ q_t^D = \gamma_{21}p_t + \beta_{21} \\ q_t = q_t^S = q_t^D \end{cases} \quad (1.19)$$

Tämän mallin osalta voidaan todeta, että vaikka $(p_t \ q_t)'$ -pisteet edelleenkin sijaitsevat kysyntäsuoralla täyttämättä aidosti kaksiulotteista tasoa, eivät $(p_{t-1} \ q_t)'$ -pisteet enää pysykään tällä suoralla.

Ajatellaan nyt, että hetkellä t_0 eksogeenisissa tekijöissä X_t tapahtuisi muutos, mutta että ne sen jälkeen pysyisivät vakioina. Tällöin systeemin (1.19) mukainen ura $(p_{t-1} \ q_t)'$ -tasossa näyttäisi seuraavalta:



Hetken t_o jälkeen havaitut pisteet sijaitsevat siis uudella tarjontasuoralla S' ja po. suoran paikka identifioidaan (p_{t-1}, q_t) -pisteiden ”seittimäisen” liikkeen ansiosta. Iteraation edetessä tasapainopiste \circ estimoituu yhä tarkemmin. Pisteitä \circ ja $*$ vertailemalla voidaan toisaalta identifioida myös kysyntäsuoran D sijainti. Dynamiikan huomioon ottaminen ratkaisi siis tässä tapauksessa identifiointuongelman, mutta on huomattava, että mallin (1.19) realistisuus riippuu oleellisesti havaintovälin pituudesta. Malliin (1.17) sen sijaan sisältyi identifiointuongelmia nimenomaan *havaintovälin puitteissa* ilmenevän vuorovaikutuksen osalta.

2 LIKELIHOOD-FUNKTION PERUSTUVA ESTIMOINTI JA HYPOTEESEIN TESTAAMINEN

2.1 ML-estimaattoreiden asymptoottiset ominaisuudet

Ajatellaan aluksi, että havainnot Y_1, \dots, Y_n olisivat toisistaan riippumattomia ja noudattaisivat yhteistä, identifiointuvaan jakaumaperheeseen $\mathcal{F} = \{F^{(\theta)} \mid \theta \in \Theta\}$, $\Theta \subseteq \mathcal{R}^p$ kuuluvaa jakaumaa. Merkitään θ_o :lla sitä parametriarvoa, johon liittyvästä jakaumasta havainnot ajatellaan saatuiksi, ja tarkastellaan funktiota

$$h(\theta) = E^{(\theta_o)} \log L_{Y_i}(\theta) \quad , \quad \theta \in \Theta \quad . \quad (2.1)$$

Ns. Jensenin epäyhtälön mukaan

$$-E^{(\theta_o)} \log \frac{L_{Y_i}(\theta)}{L_{Y_i}(\theta_o)} \geq -\log E^{(\theta_o)} \frac{L_{Y_i}(\theta)}{L_{Y_i}(\theta_o)} = 0 \quad ,$$

ts.

$$h(\theta) = E^{(\theta_o)} \log L_{Y_i}(\theta) \leq E^{(\theta_o)} \log L_{Y_i}(\theta_o) = h(\theta_o) \quad (2.2)$$

olipa $\theta \in \Theta$ mikä tahansa. Log-likelihood-funktion ”odotusarvopinta” $h(\theta)$ saa siis maksiminsa ”oikean” parametriarvon θ_o kohdalla. Koska malliperheen parametrit oletettiin identifioituviksi, on $h(\theta) \neq h(\theta_o)$ kaikilla $\theta \neq \theta_o$, joten kyseessä on samalla aito maksimi.

Koska toisaalta suurten lukujen lain nojalla

$$\frac{1}{n} \log L_{Y_1, \dots, Y_n}(\theta) = \frac{1}{n} \sum_{i=1}^n \log L_{Y_i}(\theta) \xrightarrow{p} h(\theta) \quad \text{kun} \quad n \rightarrow \infty,$$

nähdään, että $\frac{1}{n} \sum_{i=1}^n \log L_{Y_i}(\theta)$:n maksimipiste $\hat{\theta}_n$ lähestyy n :n kasvaessa $h(\theta)$:n maksimipistettä θ_o , kunhan funktio h vain on jatkuva. On siis todettu, että n :n kasvaessa

$$\hat{\theta}_n \xrightarrow{p} \theta_o \quad (2.3)$$

ts. että ML-estimaattori $\hat{\theta}_n$ on *tarkentuva*.

Määritelmä 2.1: Oletetaan, että havaintojen Y_1, \dots, Y_n määräämä log-likelihood-funktio

$$l_{Y_1, \dots, Y_n}(\theta) = \log L_{Y_1, \dots, Y_n}(\theta)$$

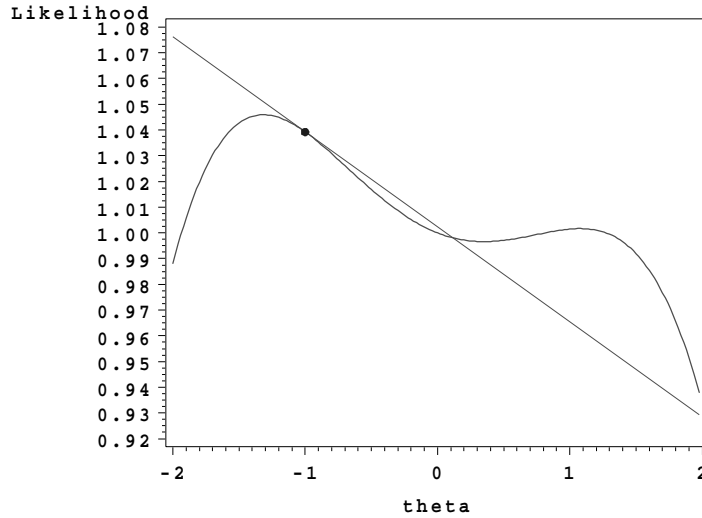
on parametrien $\theta \in \mathcal{R}^p$ suhteen jatkuvasti derivoituva ja että derivaattojen kovarianssimatriisit

$$\mathcal{I}(\theta) = \text{cov}(Dl_{Y_1, \dots, Y_n}(\theta)') \quad (2.4)$$

ovat aina hyvin määriteltyjä. Tällöin matriisia $\mathcal{I}(\theta)$ kutsutaan *Fisherin informaatiomatriisiksi*. Po. matriisin nimi johtuu siitä, että sen elementtien voidaan ajatella tietyllä tavalla mittaavan havaintoihin sisältyvän, eri θ :n komponentteja koskevan informaation määrää. ■

Huomautus 2.1: Merkitään koko havaintoaineistoa lyhyesti merkinnällä $Y = (Y_1 \dots Y_n)'$. Ajatellaan aluksi, että $p = 1$, ts. että $\Theta \subseteq \mathcal{R}^1$. Tyypillisen likelihood-funktion käyttäytymistä on tällöin helppo havainnollistaa graafisen kuvaajan avulla:

$$L_Y(\theta)$$



”Naapurimallit” θ -parametroinnissa erottuvat toisistaan sitä paremmin, mitä suurempi on $|L'_Y(\theta)|$, ts. mitä jyrkempi $L_Y(\theta)$ -funktion tangentti po. pisteessä on. Periaatteessa voidaan ajatella erottuvuutta mitattavan $|\frac{d}{d\theta}g(L_Y(\theta))|^\alpha$ -tyyppisillä lausekkeilla. Erottuvuuden odotusarvo voidaan puolestaan tulkita informaatiomitaksi, joka kertoo, kuinka paljon θ :aa koskevaa informaatiota havaintoihin *keskimäärin* sisältyy. Tuntuu luonnolliselta vaatia informaatiomitan toimivan niin, että toisistaan riippumattomat havaintomateriaalit kantaisivat yhdessä informaatiota osainformaatioiden summan verran. Tämä toivomus johtaa valintaan

$$g(x) = \log x \quad \text{ja} \quad \alpha = 2 \quad .$$

Koska toisaalta $ED \log L_Y(\theta) \equiv 0$, saadaan informaatiomitaksi siis

$$I(\theta) = E \left| \frac{d}{d\theta} \log L_Y(\theta) \right|^2 = \text{var}(D \log L_Y(\theta)) \quad .$$

Vastaavasti useamman parametrin tapauksessa ($p > 1$) on luontevaa määritellä keskimääräinen informaatio kaavalla (2.4). ■

Huomautus 2.2: Jos $\log L_Y(\theta)$ on kolmasti jatkuvasti differentioituva, voidaan osoittaa, että informaatiomatriisi voidaan laskea myös kaavasta

$$I(\theta) = -ED^2 \log L_Y(\theta) \tag{2.5}$$

■

Lause 2.1: Oletetaan, että θ :n estimaattori $T = T(Y_1, \dots, Y_n)$ on harhaton, ts $E^{(\theta)}T \equiv \theta$. Jos $\log L_Y(\theta)$ on kolmasti jatkuvasti differentioituva θ :n suhteen, on

$$\text{cov}(T) - \mathcal{I}(\theta)^{-1} \succeq 0 \quad (2.6)$$

■

Huomautus 2.3: On melko luonnollista, että käytettävissä oleviin havaintoihin sisältyvän informaation määrä ratkaisee, kuinka tarkasti θ :aa voidaan po. havaintojen avulla estimoida. Huomattakoon, että jos havainnot Y_1, \dots, Y_n ovat toisistaan riippumattomia ja identtisesti jakautuneita ($Y_i \sim i.i.d.$), on

$$\mathcal{I}(\theta) = n\mathcal{I}_1(\theta)$$

jossa $\mathcal{I}_1(\theta)$ tarkoittaa yksittäiseen havaintoon liittyvän informaation määrää. ■

Lause 2.2: Jos $Y_i \sim i.i.d.$, jos havaintojen määräämä likelihood-funktio on kolmasti jatkuvasti differentioituva, jos θ_o on Θ :n sisäpiste ja jos $\mathcal{I}(\theta_o)$ on epäsingularinen, on θ :n ML-estimaattorin $\hat{\theta}_n$ asympotoottinen otantajakauma normaallinen,

$$\sqrt{n}(\hat{\theta}_n - \theta_o) \stackrel{\text{asympt.}}{\sim} N_p(0, \mathcal{I}_1(\theta_o)^{-1}) \quad (2.7)$$

Sama asia voidaan kirjoittaa myös matemaattisesti epätäsmällisempään muotoon

$$\hat{\theta}_n \stackrel{\text{asympt.}}{\sim} N_p(\theta_o, \mathcal{I}(\theta_o)^{-1}) \quad (2.8)$$

■

Huomautus 2.4: Sama asympotoottinen jakaumatulos pätee melko yleisesti myös toisistaan riippuvien ja jakaumiltaan vaihtelevien havaintojen tapauksessa.

Mikäli informaatiomatriisin $\mathcal{I}(\theta)$ muodon analyttinen määrittäminen on ylivoimaisista, voidaan $\mathcal{I}(\theta)$ tarvittaessa korvata havaitulla $-D^2 \log L_Y(\hat{\theta}_n)$:n arvolla. Mikäli $Y_i \sim i.i.d.$, voidaan $\mathcal{I}(\theta)$:n estimaattorina käyttää myös ns. *ulkotuloestimaattoria* $\sum_{i=1}^n D \log L_{Y_i}(\hat{\theta}_n)' D \log L_{Y_i}(\hat{\theta}_n)$. ■

Seurauslause 2.1: Mikäli lauseen 2.2 oletukset pitävät paikkansa ja $G : \Theta \rightarrow \mathcal{R}^r$ on jokin θ :n differentioituva funktio, on myös $G(\hat{\theta}_n)$:n asymptoottinen otantajakautuma normaalin, sillä

$$G(\hat{\theta}_n) \stackrel{\text{asympt.}}{\sim} N_r(G(\theta_o), DG(\theta_o) \mathcal{I}(\theta_o)^{-1} DG(\theta_o)') \quad (2.9)$$

2.2 Parametristen hypoteesien testaaminen

Ajatellaan, että yleishypoteesin

H : ”Havainnot $Y = (Y_1 \dots Y_n)'$ on saatu jostakin malliperheen $\mathcal{F} = \{F_Y^{(\theta)} \mid \theta \in \Theta\}$ ($\Theta \subseteq \mathcal{R}^p$) jakaumasta”

puiteissa tutkittava nollahypoteesi

H_o : ”Havainnot $Y = (Y_1 \dots Y_n)'$ on saatu jostakin malliperheen $\mathcal{F}_o = \{F_Y^{(\theta)} \mid \theta \in \Theta_o\}$ ($\Theta_o \subset \Theta$) jakaumasta”

on saatettu joko muotoon

$$\Theta_o = \{\theta(\psi) \mid \psi \in \Psi\} \subset \Theta, \quad \dim(\Psi) = \dim(\Theta_o) = p - r < p = \dim(\Theta)$$

tai muotoon

$$\Theta_o = \{\theta \mid G(\theta) = 0 \in \mathcal{R}^r\} .$$

Likelihood- funktion maksimia

$$\max_{\theta \in \Theta_o} L_Y(\theta) = \max_{G(\theta)=0} L_Y(\theta) = \max_{\psi \in \Psi} L_Y(\theta(\psi))$$

voidaan pitää mittana, joka kertoo, *kuinka paljon tukea aineisto Y antaa hypoteesille H_o .*

Vastaavasti maksimi

$$\max_{\theta \in \Theta} L_Y(\theta) = L_Y(\hat{\theta}_n)$$

kertoo, kuinka paljon tukea aineisto Y antaa yleishypoteesille H .
Tuntuu kohtuulliselta katsoa aineiston olevan ristiriidassa H_o :n kanssa, jos suhde

$$\Lambda_Y = \frac{\max_{\theta \in \Theta_o} L_Y(\theta)}{\max_{\theta \in \Theta} L_Y(\theta)} \quad (2.10)$$

on kovin pieni. Suuretta (2.10) sanotaan *LR-testisuureeksi* ja siihen perustuvat ”kriittiset alueet” ovat tietenkin muotoa

$$\{\Lambda_Y \leq c\} \quad .$$

Mikäli Λ_Y :n tarkkaa otantajakaumaa ei pystytä selvittämään, voidaan testin p -arvojen laskemisessa käyttää hyväksi seuraavaa asymptoottista tulosta:

Lause 2.3: Mikäli havaintoaineiston määräämä likelihood-funktio toteuttaa lauseen 2.2 oletukset kaikilla $\theta \in \Theta_o$, pätee LR-testisuureelle n :n kasvaessa tulos

$$-2 \log \Lambda_Y \stackrel{\text{asympt.}}{\sim} \chi_r^2 \quad (2.11)$$

Toinen, yhtä luonnollinen H_o :n testausmahdollisuus perustuu $G(\hat{\theta}_n)$:n ja origon väliseen etäisyyteen. Tähän ajatukseen perustuvia testejä kutsutaan *Wald-testeiksi*.

Lause 2.4: Mikäli havaintoaineiston määräämä likelihood-funktio toteuttaa lauseen 2.2 oletukset kaikilla $\theta \in \Theta_o$ ja mikäli G on kaikkialla differentioituva, on apulauseen 2.1 mukaan

$$W_n = (G(\hat{\theta}_n) - G(\theta_o))' [DG(\hat{\theta}_n) \mathcal{I}(\hat{\theta}_n)^{-1} DG(\hat{\theta}_n)']^{-1} (G(\hat{\theta}_n) - G(\theta_o)) \quad (2.12)$$

$$\stackrel{\text{asympt.}}{\sim} \chi_r^2$$

Esimerkki 2.1: Tarkastellaan epälineaarisia regressiomalleja (1.5)

$$y_t = g(X_t, \beta) + \varepsilon_t \quad ,$$

$\varepsilon_t \perp\!\!\!\perp X_t$, $\varepsilon_t \sim NID(0, \sigma^2)$, $t = 1, \dots, n$, $\beta = (\beta_1 \dots \beta_m)'$, $\theta = (\beta' \sigma^2)'$.
joista esimerkkeinä johdantoluvussa mainittiin ns. CES- ja Cobb-Douglas-tuotantofunktiomallit. Mallin (1.5) puitteissa

$$\log L_{y_t|X_t}(\beta, \sigma^2) \simeq -\frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y_t - g(X_t, \beta))^2 \quad , \quad (2.13)$$

joten

$$D_\beta \log L_{y_t|X_t}(\beta, \sigma^2)' = \frac{1}{\sigma^2} \varepsilon_t Dg(X_t, \beta)'$$

ja

$$\frac{d}{d\sigma^2} \log L_{y_t|X_t}(\beta, \sigma^2) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \varepsilon_t^2 \quad .$$

Hvaintoon t liittyväksi informaatiomatriisiksi saadaan

$$\mathcal{I}_t(\beta, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} Dg(X_t, \beta)' Dg(X_t, \beta) & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} \quad . \quad (2.14)$$

Vaikka informaation määrä havainnoittain siis vaihtelee, ovat lauseiden 2.2 ja 2.3 tulokset silti edelleen voimassa, kunhan informaatiiovirta ei jossain vaiheessa kokonaan tyrehdy.

Kuten huomautuksessa 1.4 todettiin, on *koko* havaintoaineiston määräämä likelihood-funktio muotoa (1.16) ja β :n ML-estimaattori on samalla OLS-estimaattori, joka minimoi neliösumman $Q(\beta)$. Lisäksi nähdään, että

$$\hat{\sigma}^2 = \frac{1}{n} \sum (Y_t - g(X_t, \hat{\beta}))^2 \quad .$$

Lauseen 2.2 tuloksesta voidaan päätellä, että

$$\hat{\beta} \stackrel{\text{asympt.}}{\sim} N_m(\beta, \sigma^2 (D'_\beta D_\beta)^{-1}) \quad (2.15)$$

jossa on merkitty

$$D_\beta = \begin{pmatrix} Dg(X_1, \beta) \\ \vdots \\ Dg(X_n, \beta) \end{pmatrix} \quad .$$

Ajatellaan nyt, että halutaan testata β :aa koskevia rajoitteita

$$H_o : G(\beta) = (g_1(\beta) \dots g_r(\beta))' = 0 \quad .$$

Merkitään näiden rajoitteiden vallitessa estimoitua β :aa symbolilla $\hat{\beta}_R$ ja vastaavaa virhevarianssiestimaattoria symbolilla $\hat{\sigma}_R^2$.

Koska

$$\max_{\beta, \sigma^2} L_{Y|X}(\beta, \sigma^2) = \left(\frac{1}{2\pi\hat{\sigma}^2} \right)^{\frac{n}{2}} e^{-\frac{n}{2}} ,$$

on

$$\begin{aligned} -2 \log \Lambda_{Y|X} &= n \log \left(\frac{\hat{\sigma}_R^2}{\hat{\sigma}^2} \right) \\ &= n \log \left(\frac{Q(\hat{\beta}_R)}{Q(\hat{\beta})} \right) \\ &\stackrel{\text{asympt.}}{\sim} \chi_r^2 \end{aligned} \tag{2.16}$$

jossa r tarkoittaa rajoitteiden lukumäärää.

Ajatellaan nyt hetkeksi, että rajoitteet $G(\beta) = 0$ olisivat lineaarisia. Koska $\log(1+x) \approx x$ kun $|x|$ on pieni, saadaan kaavasta (2.16) myös approksimaatio

$$\begin{aligned} -2 \log \Lambda_{Y|X} &\approx n \frac{Q(\hat{\beta}_R) - Q(\hat{\beta})}{Q(\hat{\beta})} \\ &\stackrel{\text{asympt.}}{=} r \frac{\frac{1}{r}(Q(\hat{\beta}_R) - Q(\hat{\beta}))}{\frac{1}{n-m}Q(\hat{\beta})} \\ &= r F_{r, n-m} , \end{aligned} \tag{2.17}$$

jossa $F_{r, n-m}$ on konstruoitu täsmälleen samalla periaatteella kuin lineaaristen mallien yhteydessä yleisesti käytetty, vastaavaa rajoitetta testaava F -testisuure. Huomataan siis, että lineaarisiin ja epälineaarisiin regressiomalleihin liittyvillä tarkasteluilla on suhteellisen vähän eroa keskenään! Vain OLS-estimaattien numeerinen laskeminen on epälineaaristen mallien osalta hivenen hankalampaa.

(Esimerkkeinä regressioparametreja koskevien rajoitteiden realistisuuden testaamisesta ajateltakoon ns. vakioskaalatuottohypoteesia CES-mallin (1.13) puitteissa

$$H_o : \beta_4 = 1$$

tai Cobb-Douglas-mallin (1.14) puitteissa

$$H_o : \gamma_1 + \gamma_2 = 1 \quad .)$$

■

2.3 Likelihood- funktion numeerinen optimointi

Kaikkein yksinkertaisimpia tilanteita lukuun ottamatta on ns. likelihood-yhtälöiden

$$D \log L_Y(\theta)' = 0$$

analyttinen ratkaiseminen yleensä mahdotonta. Tällöin on pakko tyytyä log-likelihood-funktion $l_Y(\theta) = \log L_Y(\theta)$ iteratiivisiin maksimointialgoritmeihin, joista useimmat perustuvat toisen asteen Taylor-kehitelmiin

$$l_Y(\theta) = l_Y(\theta^*) + D l_Y(\theta^*)(\theta - \theta^*) + \frac{1}{2!}(\theta - \theta^*)' D^2 l_Y(\theta^*)(\theta - \theta^*) + \|\theta - \theta^*\| o(\theta - \theta^*) \quad , \quad (2.18)$$

jossa θ^* tarkoittaa kehityspistettä. Yhtälöstä (2.18) nähdään heti, että θ^* voi olla $l_Y(\theta)$:n maksimipiste vain, jos ehdot $D l_Y(\theta^*) = 0$ ja $-D^2 l_Y(\theta^*) \succ 0$ ovat täytetyt.

Kehitelmään (2.18) perustuvista maksimointialgoritmeista vanhin on ns. *Newton-Raphson*-algoritmi, jossa iterointi aloitetaan valitusta alkupisteestä $\theta^{(o)}$ ja approksimoidaan sitten $l_Y(\theta)$ -funktiota kehitelmän (2.18) mukaisella toisen asteen approksimaatiolla. Jos $-D^2 l_Y(\theta^{(o)}) \succ 0$, vastaa po. toisen asteen approksimaatio $l_Y(\theta)$ -funktion kuvaaja”pintaa” pisteessä $\theta^{(o)}$ sivuavaa elliptistä parabolodia. Siirrytään seuraavaksi tämän paraboloidin huipulle ja merkitään tätä pistettä symbolilla $\theta^{(1)}$. On helppo todeta, että tällöin

$$\theta^{(1)} = \theta^{(o)} - [D^2 l_Y(\theta^{(o)})]^{-1} D l_Y(\theta^{(o)})' \quad .$$

Valitaan sitten $\theta^{(1)}$ uudeksi kehityspisteeksi ja jatketaan iterointia saman periaatteen mukaan, ts. askelluskaavoja

$$\theta^{(i)} = \theta^{(i-1)} - [D^2 l_Y(\theta^{(i-1)})]^{-1} D l_Y(\theta^{(i-1)})' \quad i = 1, 2, 3, \dots \quad (2.19)$$

noudattaen. Kun jono $\{\theta^{(i)}\}$ on konvergoinut, ts. kun $\|\theta^{(i)} - \theta^{(i-1)}\|$ on (halutulla tarkuudella) $= 0$, on ilmeisesti $D l_Y(\theta^{(i)})' = 0$, joten maksimi on löytynyt, mikäli lisäksi $D^2 l_Y(\theta^{(i)}) \prec 0$.

Newton-Raphson- algoritmilla on kuitenkin seuraavat varjopuolet:

- Hessen matriisi $D^2 l_Y(\theta)$ mielivaltaisella θ :n arvolla ei välttämättä ole definiitti.
- Pitkät askeleet $\theta^{(i-1)} \rightarrow \theta^{(i)}$ voivat johtaa algoritmin täysin harhapoluille, sillä toisen asteen paraboloidiapproksimaatio (2.18) voi tällöin olla hyvin epätarkka.
- Hessen matriisin $D^2 l_Y(\theta)$ laskeminen numeerisesti on usein erittäin työlästä, joten Newton-Raphson -algoritmia kannattaa käyttää lähinnä vain sellaisissa tilanteissa, joissa $D^2 l_Y(\theta)$ hallitaan analyttisesti.

Ensimmäinen ja kolmas haitta voidaan eliminoida korvaamalla $-D^2l_Y(\theta)$ vastaavalla *informaatiomatriisilla* $\mathcal{I}(\theta)$ askelluskaavassa (2.19), sillä informaatiomatriisi on score-vektorin kovarianssimatriisina aina ei-negatiivisesti definiitti. Tällöin askelluskaavaksi saadaan

$$\theta^{(i)} = \theta^{(i-1)} + \mathcal{I}(\theta^{(i-1)})^{-1} D l_Y(\theta^{(i-1)})' \quad i = 1, 2, 3, \dots \quad (2.20)$$

Näin syntynyttä optimointialgoritmia sanotaan *likelihood scoring*-algoritmiksi. Sen käyttö edellyttää tietenkin, että informaatiomatriisi $\mathcal{I}(\theta)$ pystytään analyttisesti laskemaan.

Jos toisaalta informaatiomatriisi korvataan ulkotuloestimaattorillaan huomautuksen 2.4 mukaisesti, päädytään askelluskaavoihin

$$\theta^{(i)} = \theta^{(i-1)} + \left[\sum_{i=1}^n D \log L_{Y_i}(\theta^{(i-1)})' D \log L_{Y_i}(\theta^{(i-1)}) \right]^{-1} D l_Y(\theta^{(i-1)})' \quad (2.21)$$

$$i = 1, 2, 3, \dots$$

Näin syntynyttä algoritmia kutsutaan *BHHH-algoritmiksi* (Berndt-Hall-Hall-Hausman-algoritmiksi). Koska askelluskaavoissa tarvitaan ainoastaan havainnoittaisia score-vektoreita, on BHHH-algoritmi yleensä erityisen helppo implementoida.

Esimerkki 2.1: (jatkoa)

Tarkastellaan epälineaarista regressiomallia

$$y_t = g(X_t, \beta) + \varepsilon_t \quad , \quad (2.22)$$

$\varepsilon_t \perp\!\!\!\perp X_t$, $\varepsilon_t \sim NID(0, \sigma^2)$, $t = 1, \dots, n$, $\beta = (\beta_1 \dots \beta_m)'$, $\theta = (\beta' \sigma^2)'$. Aikaisempien tarkastelujen mukaisesti ML-estimaattori $\hat{\beta}$ minimoi samalla neljösumman $Q(\beta) = \sum_{t=1}^n (y_t - g(X_t, \beta))^2$ ja $\hat{\beta}$:n asymptoottinen otantajakauma on muotoa

$$\hat{\beta} \stackrel{\text{asympt.}}{\sim} N_m(\beta, \mathcal{I}(\beta)^{-1})$$

jossa $\mathcal{I}(\beta) = \frac{1}{\sigma^2} D'_\beta D_\beta$ ja

$$D_\beta = \begin{pmatrix} D g(X_1, \beta) \\ \vdots \\ D g(X_n, \beta) \end{pmatrix} \quad . \quad (2.23)$$

Toisaalta

$$D_\beta \log L_{Y_t|X_t}(\theta)' = \frac{1}{\sigma^2} e_t(\beta) D g(X_t, \beta) \quad \text{ja}$$

$$D_\beta^2 \log L_{Y_t|X_t}(\theta) = -\frac{1}{\sigma^2} D g(X_t, \beta)' D g(X_t, \beta) + \frac{1}{\sigma^2} e_t(\beta) D^2 g(X_t, \beta) \quad ,$$

jossa $e_t(\beta) = y_t - g(X_t, \beta)$. Tästä nähdään, että Newton-Raphson -algoritmissa tarvittavien Hessen matriisien laskeminen olisi melko hankalaa, kun taas likelihood scoring -algoritmissa tarvittavat informaatiomatriisit saadaan yksinkertaisemmin muodosta

$$\begin{aligned} \mathcal{I}(\beta) &= \sum_{t=1}^n -E D_\beta^2 \log L_{Y_t|X_t}(\theta) \\ &= \frac{1}{\sigma^2} \sum_{t=1}^n D g(X_t, \beta)' D g(X_t, \beta) = \frac{1}{\sigma^2} D_\beta' D_\beta \quad . \end{aligned}$$

Koska informaatiomatriisi $\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}(\beta) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$ on blokkidiagonaalinen, saavat scoring- algoritmin askelluskaavat (2.20) β :n osalta muodon

$$\begin{aligned} \beta^{(j)} &= \beta^{(j-1)} + \mathcal{I}(\beta^{(j-1)})^{-1} \sum_{t=1}^n D_\beta \log L_{Y_t|X_t}(\beta^{(j-1)})' \\ &= \beta^{(j-1)} + [D_{\beta^{(j-1)}}' D_{\beta^{(j-1)}}]^{-1} D_{\beta^{(j-1)}}' E(\beta^{(j-1)}) \quad , \quad j = 1, 2, \dots \end{aligned} \quad (2.24)$$

jossa $E(\beta) = (e_1(\beta) \dots e_n(\beta))'$. Muodosta (2.24) nähdään heti, että scoring- proseduuri voidaan tulkita myös siten, että $\beta^{(j)}$ saadaan mallin (2.22) ”lineari-soituun” versioon

$$y_t = g(X_t, \beta^{(j-1)}) + D g(X_t, \beta^{(j-1)}) (\beta - \beta^{(j-1)}) + \varepsilon_t \quad (2.25)$$

liittyvänä pienimmän neliösumman estimaattorina. ■

2.4 Kaksivaiheiset, ML -tekniikkaa jäljittelevät estimointimenetelmät

Lauseen 2.2 tulos perustuu oleellisesti luvulla $\frac{1}{\sqrt{n}}$ kerrotun log-likelihood -funktion gradienttivektorin (*score* -vektorin)

$$\frac{1}{\sqrt{n}} D l_Y(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^n D l_{Y_t}(\theta)$$

asymptoottiseen normaalisuuteen, joka seuraa suoraan keskeisestä raja-arvolauseesta. Scoring-algoritmin askelluskaavaa (2.20) tarkasteltaessa tulee etsimättä mieleen, että yhden askeleen jälkeisen estimaatin $\theta^{(1)}$ stokastiset ominaisuudet määräytyvät score-vektorin $D l_Y(\theta^{(0)})$ perusteella, mikäli aloitusarvo $\theta^{(0)}$ pystytään valitsemaan *tarkentuvasti*. Tähän oivallukseen perustuen on helppo todistaa seuraava lause:

Lause 2.5: Mikäli scoring-algoritmin alkuarvo $\theta^{(0)}$ valitaan sellaisella alkuestimointimenetelmällä, että $\sqrt{n}(\theta^{(0)} - \theta)$ lähestyy jakaumaltaan hyvin määriteltyä rajajakaumaa, pätee tulos

$$\theta^{(1)} \stackrel{\text{asympt.}}{\sim} N_p(\theta, \mathcal{I}(\theta)^{-1}) \quad (2.26)$$

Yhden askeleen jälkeen saadulla estimaattorilla on siis *samat* asymptoottiset ominaisuudet kuin varsinaisilla ML-estimaattoreilla (jotka saataisiin, jos iteraatiota jatkettaisiin loppuun saakka), kunhan alkuestimaatit $\theta^{(0)}$ täyttävät edellä mainitut ehdot!

Huomattakoon samalla, että iteraatioissa voidaan $\mathcal{I}(\theta^{(0)})$:n korvikkeena käyttää mitä tahansa matriisia $\mathcal{I}^*(\theta^{(0)})$, jolla on ominaisuus

$$p \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{I}^*(\theta^{(0)}) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{I}(\theta) \quad .$$

Esimerkki 2.2: Tarkastellaan lineaarista regressiomallia

$$y_t = \beta' X_t + \varepsilon_t \quad , \quad t = 1, \dots, n \quad , \quad \beta \in \mathcal{R}^m \quad , \quad (3.27)$$

jossa ε_t - virhetermisarja oletetaan *autokorreloituneeksi* ns. ARMA -rakenteen

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q} \quad , \quad \alpha_t \sim NID(0, \sigma^2) \quad (2.28)$$

mukaisesti. Otetaan käyttöön lyhyet merkinnät

$$\psi = (\phi_1 \quad \dots \quad \phi_p \quad \theta_1 \quad \dots \quad \theta_q \quad \sigma^2)'$$

ja

$$\Sigma(\psi) = \text{cov}((\varepsilon_1 \quad \dots \quad \varepsilon_n)') \quad .$$

Tällöin

$$l_Y(\beta, \psi) \simeq -\frac{1}{2} \log \det \Sigma(\psi) - \frac{1}{2} (Y - X\beta)' \Sigma(\psi)^{-1} (Y - X\beta) \quad , \quad (2.29)$$

jossa

$$Y = (y_1 \quad \dots \quad y_n)' \quad \text{ja} \quad X = (X_1 \quad \dots \quad X_n)' \quad .$$

Funktion l_Y gradienttivektoriksi ja Hessen matriisiksi saadaan

$$D l_Y(\beta, \psi)' = \begin{pmatrix} X' \Sigma(\psi)^{-1} (Y - X\beta) \\ D_\psi l_Y(\beta, \psi)' \end{pmatrix}$$

ja

$$D^2 l_Y(\beta, \psi) = \begin{pmatrix} -X' \Sigma(\psi)^{-1} X & X' [D_\psi \Sigma(\psi)^{-1}] (Y - X\beta) \\ (Y - X\beta)' [D_\psi \Sigma(\psi)^{-1}] X & D_\psi^2 l_Y(\beta, \psi) \end{pmatrix} \quad .$$

Tästä on helppo nähdä, että koska

$$E(Y - X\beta) = 0 \quad ,$$

on informaatiomatriisi $\mathcal{I}(\beta, \psi) = E - D^2 l_Y(\beta, \psi)$ *blokkidiagonaalinen*.

Tästä puolestaan seuraa, että

$$\begin{aligned} \beta^{(1)} &= \beta^{(0)} + [X' \Sigma(\psi^{(0)})^{-1} X]^{-1} X' \Sigma(\psi^{(0)})^{-1} (Y - X\beta^{(0)}) \\ &= [X' \Sigma(\psi^{(0)})^{-1} X]^{-1} X' \Sigma(\psi^{(0)})^{-1} Y \end{aligned} \quad (2.30)$$

$$\stackrel{\text{asympt.}}{\sim} N_m(\beta, [X' \Sigma(\psi)^{-1} X]^{-1}) \quad .$$

Estimaattori $\beta^{(1)}$ ei siis *riipu alkuarvoista $\beta^{(0)}$ lainkaan!* Se on samalla helppo tunnistaa ns. GLS-estimaattoriksi (yleistetyksi pienimmän neliösumman estimaattoriksi). Se on myös samalla *asymptoottisesti tehokas*, kunhan vain ψ :n alkuestimointi $\psi^{(0)}$ on tarkentuva (\sqrt{n} -tarkentuva). ■

2.5 LM-testit (score-testit)

Palataan tarkastelemaan luvussa 2.2 käsiteltyä parametristen hypoteesien testausongelmaa ja käytetään täsmälleen samoja merkintöjä kuin tuolloin. Oletetaan, että lauseen 2.2 oletukset ovat voimassa ja merkitään rajoitteiden $H_0 : G(\theta) = 0$ puitteissa estimoitua ML-estimaattoria symbolilla $\hat{\theta}_n^{(R)}$. Kaikkien edellä esitettyjen lauseiden 2.2, 2.3 ja 2.4 tulokset perustuvat siihen keskeiseen tietoon, että lauseessa 2.2 esitetyin ehdoin

$$\frac{1}{\sqrt{n}} D l_Y(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^n D l_{Y_t}(\theta) \stackrel{\text{asympt.}}{\sim} N_p \left(0, \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{I}(\theta) \right) .$$

Tätä taustaa vasten on helppo ymmärtää, että samoin edellytyksin pätee myös seuraava lause:

Lause 2.5: Mikäli havaintoaineiston määräämä likelihood-funktio toteuttaa lauseen 2.2 oletukset kaikilla $\theta \in \Theta_0$, pätee n :n kasvaessa tulos

$$LM_n = D \log L_Y(\hat{\theta}_n^{(R)})' \mathcal{I}(\hat{\theta}_n^{(R)})^{-1} D \log L_Y(\hat{\theta}_n^{(R)})' \stackrel{\text{asympt.}}{\sim} \chi_r^2 \quad (2.31)$$

Tällaista score-vektoriin $D \log L_Y(\hat{\theta}_n^{(R)})'$ perustuvaa testiä sanotaan *Lagrangen kertojatestiksi* (LM-testiksi) tai *score-testiksi*. Tämän periaatteen hyvänä puoleena on se, ettei testisuureen (2.31) laskeminen edellytä ”vapaan mallin” mukaisen ML-estimaattorin $\hat{\theta}_n$ muodostamista lainkaan. Tästä on usein valtavaa hyötyä erityisesti ns. mallidiagnostiikan yhteydessä, ts. silloin, kun tutkitaan jonkin aikaisemmassa mallissa tehdyn oletuksen realistisuutta havaintojen valossa. Mikäli tarkastelun kohteena oleva oletus jätetään tekemättä, saattaa mallin parametrien estimointi vaikeutua ratkaisevasti, kuten seuraava esimerkkikin osoittaa:

Esimerkki 2.3: Tarkastellaan esimerkin 2.2 tavoin lineaarista regressiomallia

$$y_t = \beta' X_t + \varepsilon_t \quad , \quad t = 1, \dots, n \quad ,$$

jossa virhetermien ε_t autokorrelaatioon pyritään varautumaan yksinkertaisimman mahdollisen mallin, ts. AR(1)-mallin

$$\varepsilon_t = \phi \varepsilon_{t-1} + \alpha_t \quad , \quad \alpha_t \sim NID(0, \sigma^2)$$

puitteissa. Voidaan helposti todeta, että tällöin $\text{corr}(\varepsilon_t, \varepsilon_{t-\nu}) = \rho(\nu) = \phi^\nu$, $\nu = 1, 2, \dots$, joten malli todella sallii eri ajankohtiin liittyvien virhetermien väliset korrelaatiot. Lisäksi parametri ϕ saa tulkinnan $\phi = \text{corr}(\varepsilon_t, \varepsilon_{t-1}) = \rho(1)$.

Testataan nyt hypoteesin $H_o : \phi = 0$, ts. autokorreloimattomuushypoteesin realistisuutta. Merkitään kaikkien parametrien muodostamaa vektoria symbolilla $\theta = (\beta' \quad \sigma^2 \quad \phi)'$. Tällöin

$$\begin{aligned} \log L_{Y|X}(\theta) &= \sum_{t=2}^n \log L_{y_t|X_t, y_{t-1}, \dots, y_1, X_1}(\theta) \\ &= -\frac{n-1}{2} \log \left(\frac{1}{2\pi\sigma^2} \right) - \frac{1}{2\sigma^2} \sum_{t=2}^n [(y_t - \beta' X_t) - \phi(y_{t-1} - \beta' X_{t-1})]^2 \end{aligned}$$

ja $\hat{\beta}^{(R)} = \hat{\beta} = (X'X)^{-1}X'Y$ ja $\hat{\sigma}_{(R)}^2 = \hat{\sigma}^2 = \frac{1}{n}E'E = \frac{1}{n}\sum_{t=1}^n e_t^2$, jossa $E = Y - X\hat{\beta} = (e_1 \quad \dots \quad e_n)'$ tarkoittaa tavallisista pienimmän neliösumman jäännöstermeistä koostuvaa vektoria. Tästä seuraa, että $\hat{\theta}^{(R)} = (\hat{\beta}' \quad \hat{\sigma}^2 \quad 0)'$ ja että

$$\frac{\partial}{\partial \beta} \log L_{Y|X}(\hat{\theta}^{(R)}) = 0 \quad \text{ja} \quad \frac{\partial}{\partial \sigma^2} \log L_{Y|X}(\hat{\theta}^{(R)}) = 0 \quad .$$

Toisaalta

$$\frac{\partial}{\partial \phi} \log L_{Y|X}(\theta) = \frac{1}{\sigma^2} \sum_{t=2}^n (\varepsilon_t - \phi \varepsilon_{t-1}) \varepsilon_{t-1} \quad ,$$

joten

$$\frac{\partial}{\partial \phi} \log L_{Y|X}(\hat{\theta}^{(R)}) = \frac{1}{\hat{\sigma}^2} \sum_{t=2}^n e_t e_{t-1} = (n-1)\hat{\rho}(1)$$

ja

$$E \left[\frac{\partial^2}{\partial \phi^2} \log L_{Y|X}(\theta) \right] = -\frac{1}{\sigma^2} \sum_{t=2}^n E \varepsilon_{t-1}^2 = -(n-1) \quad .$$

Näin ollen

$$LM_n = (0 \quad 0 \quad (n-1)\hat{\rho}(1)) \begin{pmatrix} \cdot & \cdot & 0 \\ \cdot & \cdot & 0 \\ 0 & 0 & (n-1) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ (n-1)\hat{\rho}(1) \end{pmatrix} = (n-1)\hat{\rho}(1)^2 \quad .$$

LM-periaate (ja lause 2.5) kehottaa siis laskemaan vierekkäisten OLS-jäännöstermien välisen korrelaatiokertoimen $\hat{\rho}(1)$ ja perustamaan autokorreloimattomuushypoteesia H_o koskevan testin tulokseen

$$(n-1)\hat{\rho}(1)^2 \stackrel{\text{asympt.}}{\sim} \chi_1^2 \quad \text{kun} \quad n \rightarrow \infty \quad .$$

■

2.6 Hausman-testit

Toinen mallidiagnostiikan kannalta erittäin käyttökelpoinen testausperiaate on ns. *Hausman*-periaate:

Käytetään edelleen samoja merkintöjä kuin luvussa 2.5. Ajatellaan, että $\hat{\theta}_n^{(R)}$ olisi jokin θ :n H_o :n vallitessa *täystehokas* estimaattori (ei kuitenkaan välttämättä ML-estimaattori), joka toisaalta olisi *epäkonsistentti* aina, kun H_o ei päde. Ajatellaan lisäksi, että olisi käytettävissä myös θ :n estimaattori $\hat{\theta}_n$ (ei välttämättä ML-estimaattori), joka olisi konsistentti koko yleismallin H puitteissa. Lisäksi oletetaan, että sekä $\hat{\theta}_n^{(R)}$ että $\hat{\theta}_n$ olisivat otantajakaumiltaan asympotoottisesti normaalisia,

$$\begin{aligned} \sqrt{n} \left(\hat{\theta}_n^{(R)} - \theta \right) &\stackrel{\text{asympt.}}{\sim} N_p(0, V_R) \\ \sqrt{n} \left(\hat{\theta}_n - \theta \right) &\stackrel{\text{asympt.}}{\sim} N_p(0, V) \quad , \end{aligned} \tag{2.32}$$

jossa V_R vastaa *Cramér-Raon alarajaa* (2.6). Tällöin tuntuisi varsin luonnolliselta perustaa H_o :n realistisuutta arvioiva testisuure erotukseen $\hat{\theta}_n^{(R)} - \hat{\theta}_n$, sillä

$$p \lim_{n \rightarrow \infty} \left(\hat{\theta}_n^{(R)} - \hat{\theta}_n \right) \neq 0$$

aina, kun H_o ei päde. Jotta tätä ajatusta päästäisiin soveltamaan, on ensin selvitettävä $(\hat{\theta}_n^{(R)} - \hat{\theta}_n)$:n asympotoottinen kovarianssimatriisi ja asympotoottinen otantajakauma. Tässä tarkoituksessa todistetaan ensin seuraava apulause:

Apulause 2.1: Merkitään

$$C_n = \text{cov} \left(\sqrt{n} \left(\hat{\theta}_n^{(R)} - \theta \right) , \sqrt{n} \left(\hat{\theta}_n - \theta \right) \right) \quad .$$

Oletusten (2.32) vallitessa pätee

$$C = \lim_{n \rightarrow \infty} C_n = 0 \quad .$$

Todistus: Tehdään vastaoletus, että olisi $C \neq 0$. Tarkastellaan estimaattoreita

$$\tilde{\theta}_n^{(r)} = \hat{\theta}_n^{(R)} - r \cdot C_n (\hat{\theta}_n^{(R)} - \hat{\theta}_n)$$

erilaisilla r :n arvoilla $r \in \mathcal{R}^1$. Tällöin

$$\sqrt{n}(\tilde{\theta}_n^{(r)} - \theta) \stackrel{\text{asympt.}}{\sim} N_p(0, \tilde{V}_r) \quad ,$$

jossa

$$\tilde{V}_r = V_R - 2 \cdot r \cdot C_n C_n' + r^2 \cdot C_n \text{cov} \left(\sqrt{n} \left(\hat{\theta}_n^{(R)} - \theta \right) \right) C_n' \quad .$$

Tarkastellaan funktiota $H(r) = \tilde{V}_r - V_R$, jolle siis pätee

$$H(0) = 0 \quad , \tag{2.33}$$

$$\frac{d}{dr}H(r) = -2 \cdot C_n C_n' + 2 \cdot r \cdot C_n \operatorname{cov} \left(\sqrt{n} \left(\hat{\theta}_n^{(R)} - \theta \right) \right) C_n' \quad , \quad (2.34)$$

$$\left[\frac{d}{dr}H(r) \right]_{r=0} = -2 \cdot C_n C_n' \leq 0 \quad . \quad (2.35)$$

Tuloksista (2.33) , (2.34) ja (2.35) seuraa, että tehdyn vastahypoteesin puitteissa olisi aina olemassa jokin sellainen $r \neq 0$, että $H(r) < 0$, ts. $\tilde{V}_r < V_R$. Tämä ei ole mahdollista, koska $\hat{\theta}_n^{(R)}$:n piti olla täystehokas, ts. kovarianssimatriisin V_R piti olla ”pienin” mahdollinen. Tehty vasta oletus $C \neq 0$ on siis väärä. ■

Lause 2.6: Oletusten (2.32) ja H_o :n vallitessa

$$\lim_{n \rightarrow \infty} \operatorname{cov} \left(\sqrt{n} \left(\hat{\theta}_n^{(R)} - \hat{\theta}_n \right) \right) = V - V_R \quad ,$$

ja siis

$$n \cdot (\hat{\theta}_n^{(R)} - \hat{\theta}_n)' [V - V_R]^{-1} (\hat{\theta}_n^{(R)} - \hat{\theta}_n) \stackrel{\text{asympt.}}{\sim} \chi_p^2 \quad . \quad (2.36)$$

Todistus: Apulauseesta 2.1 seuraa välittömästi, että

$$V = \lim_{n \rightarrow \infty} \operatorname{cov} \left(\sqrt{n} \left(\hat{\theta}_n - \theta \right) \right) = \lim_{n \rightarrow \infty} \operatorname{cov} \left(\sqrt{n} \left(\hat{\theta}_n^{(R)} - \theta \right) \right) + \lim_{n \rightarrow \infty} \operatorname{cov} \left(\sqrt{n} \left(\hat{\theta}_n^{(R)} - \hat{\theta}_n \right) \right) \quad .$$

Estimaattoreiden $\hat{\theta}_n^{(R)}$ ja $\hat{\theta}_n$ asympotoottisen normalisuuden perusteella on siis

$$\sqrt{n} \cdot (\hat{\theta}_n^{(R)} - \hat{\theta}_n) \stackrel{\text{asympt.}}{\sim} N_p(0, V - V_R) \quad , \quad (2.37)$$

joten lauseen väite on todistettu. ■

3 LINEAARIIN REGRESSIOMALLEIHIN LIITTYVÄÄ MALLIDIAGNOSTIIKKA

3.1 Virhetermien autokorrelaation havaitseminen ja huomioon ottaminen

Tarkastellaan yksittäistä, muotoa (1.6) – (1.9) olevaa regressiomallia

$$y_t = \beta' X_t + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad .$$

Jos tavalla tai toisella käy ilmi, että ajatus ε_t -virhetermien välisestä korreloimattomuudesta on epärealistinen, halutaan yleensä silti pitää kiinni oletuksista

- 1^o $E\varepsilon_t \equiv 0$ ja
- 2^o $\{\varepsilon_t\}$ stationäärinen .

Esimerkissä 2.2 yo. mallia täydennettiin virhetermimallilla

$$\begin{aligned} \varepsilon_t &= \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q} \quad , \\ \alpha_t &\sim NID(0, \sigma^2) \quad , \quad \{\varepsilon_t\} \text{ stationäärinen} \end{aligned} \quad (3.1)$$

Tällaisia malleja kutsutaan ARMA-malleiksi ja niiden suosio perustuu siihen, että niiden avulla voidaan approksimoida melkeinpä millaista stationääristä prosessia tahansa mielivaltaisen tarkasti melko vähin parametrein. Tästä syystä tuntuu luonnolliselta laajentaa mallia (1.6) – (1.9) muotoon (1.6) – (1.9) + (3.1), mikäli tarvetta virhetermien autokorrelaation sallimiseen ilmenee. Tarkastellaan nyt mallin (1.6) – (1.9) realistisuuden testaamista laajennuksen (1.6) – (1.9) + (3.1) puitteissa. Merkitään mallin (1.6) – (1.9) OLS-jäännöstermivektoria aikaisempaan tapaan symbolilla

$$E = (e_1 \quad \dots \quad e_n)' = Y - X\hat{\beta} = (I - P_X)Y \quad .$$

Määritelmä 3.1: Jäännöstermeistä lasketuksi *von Neumannin suhteeksi* sanotaan lauseketta

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad . \quad (3.2)$$

(Suuretta d kutsutaan usein myös Durbin-Watson-testisuureksi.) ■

Määritelmä 3.2: Jäännöstermien *autokorrelaatiofunktio*ksi sanotaan lukujonoa

$$r_e(\nu) = \frac{\frac{1}{n-\nu} \sum_{t=1}^{n-\nu} e_t e_{t+\nu}}{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad \nu = 1, 2, \dots, n-1 \quad . \quad (3.3)$$

■

Määritelmä 3.3: Virhetermien $\{\varepsilon_t\}$ *autokorrelaatiofunktio*ksi sanotaan lukujonoa

$$\rho_\varepsilon(\nu) = \text{corr}(\varepsilon_t, \varepsilon_{t+\nu}) \quad \nu = 1, 2, \dots \quad . \quad (3.4)$$

■

Alkuperäinen perusmalli (1.6)–(1.9) voidaan nyt laajennuksen (1.6)–(1.9) + (3.1) puitteissa muotoilla esim. hypoteesin

$$H_o: \quad \rho_\varepsilon(\nu) = 0 \quad \text{kaikilla} \quad \nu = 1, 2, \dots \quad . \quad (3.5)$$

muotoon. Ennen vanhaan hypoteesin testaamiseen käytettiin varsin yleisesti ns. *Durbin-Watson*-testiä, jossa testisuurena käytettiin von Neumannin suhdetta (3.2). Kuten huomataan, on testisuure tulkittavissa kahden jäännöstermivektoriin E liittyvän neliömuodon osamääränä. Tuloksen (1.12)

$$E \sim N_n(0, \sigma^2(I - P_X))$$

mukaan E :n jakauma siis riippuu H_o :n vallitessa selittäjämatrisista X . Näin ollen on selvää, että myös d :n otantajakauma riippuu X :stä. On olemassa laskennallisia menetelmiä (mm. ns. Imhoff-menetelmä) normaalisten vektorien neliömuotojen osamäärien jakaumien kartoittamiseksi, joten d :n jakauman *tarkkojen* prosenttipisteiden laskeminen on periaatteessa mahdollista. Prosenttipisteet pitäisi kuitenkin laskea jokaista tapausta varten erikseen. Toisaalta d :n jakauman prosenttipisteet (eli *fraktiilit*) voivat vaihdella vain tietyissä rajoissa, joten on mahdollista laskea arvo $d_U^{\frac{\alpha}{2}}$, jota suurempi d :n jakauman $\frac{\alpha}{2}$ -fraktiili ei koskaan voi olla. Vastaavasti $d_L^{\frac{\alpha}{2}}$ olkoon raja, jota pienempi d :n jakauman $\frac{\alpha}{2}$ -fraktiili ei koskaan voi olla. Näin päästään Durbin-Watson -testin tavallisimpaan käyttötapaan:

1° Päätetään ”olla hylkäämättä” H_o :aa, jos

$$d_U^{\frac{\alpha}{2}} < d < 4 - d_U^{\frac{\alpha}{2}} \quad .$$

2° ”Hylätään” H_o , jos

$$d < d_L^{\frac{\alpha}{2}} \quad \text{tai} \quad d > 4 - d_L^{\frac{\alpha}{2}} \quad .$$

3° Muissa tapauksissa pidättäydytään jyrkistä johtopäätöksistä.

Huomautus 3.1: Edellä mainitut d_U, d_L -taulukot on konstruoitu oletusta (1.6) hyödyntäen, eikä niitä pidä käyttää, jos selittäjinä on viivästettyjä y_t -muuttujia. Jos malli olisi esimerkiksi muotoa

$$y_t = \alpha y_{t-1} + \beta' X_t + \varepsilon_t \quad , \quad (3.6)$$

jossa ε_t olisi muotoa (3.1), olisi selvää, että $\{\varepsilon_t\}$ - ja $\{(y_{t-1} \ X_t)'\}$ -prosessit eivät voisi olla toisistaan riippumattomia. Voidaan kuitenkin osoittaa, että d :lle pätee seuraava asymptoottinen jakaumatulos:

$$h = \left(1 - \frac{1}{2}d\right) \sqrt{\frac{n}{1 - n \operatorname{var}(\hat{\alpha})}} \stackrel{\text{asympt.}}{\sim} N(0, 1) \quad (3.7)$$

jossa $\hat{\alpha}$ tarkoittaa α :n OLS-estimaattoria. Tulos (3.7) on tietysti käyttökelpoinen vain, jos $\operatorname{var}(\hat{\alpha}) < \frac{1}{n}$. Tulokseen (3.7) perustuva testi tunnetaan *Durbinin h-testin* nimellä. ■

Huomautus 3.2: On helppo huomata, että testisuureen d ja jäännöstermien ensimmäisen autokorrelaation $r_e(1)$ välillä vallitsee läheinen yhteys:

$$\begin{aligned} d &= \frac{\sum_{t=2}^n e_t^2 + \sum_{t=1}^{n-1} e_t^2 - 2 \sum_{t=1}^{n-1} e_t e_{t+1}}{\sum_{t=1}^n e_t^2} \\ &= 2\left(1 - \frac{n-1}{n} r_e(1)\right) - \frac{e_1^2 + e_n^2}{\sum_{t=1}^n e_t^2} \\ &\approx 2(1 - r_e(1)) \quad . \end{aligned}$$

Durbin-Watson-testi kiinnittää siis huomiota vain *vierekkäisten* jäännöstermien väliseen korrelaatioon. Tästä syystä testi on varsin epäherkkä, jopa täysin tunneton monen tyyppisille autokorrelaatioille. ■

Lower and Upper bounds of the 1% points of the Durbin-Watson test statistic

	$m = 2$		$m = 3$		$m = 4$		$m = 5$		$m = 6$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

3.1.2 Portmanteau-testit

Olkoon K jokin kiinteä kokonaisluku. Hypoteesin (3.5) vallitessa voidaan osoittaa, että

$$\sqrt{n} R = \sqrt{n} (r_e(1) \dots r_e(K))' \stackrel{\text{asympt.}}{\sim} N_K(0, I) \quad (3.8)$$

Tuloksen (3.8) voi tehdä ymmärrettäväksi seuraavalla tarkastelulla: Vektorin R *asymptoottisen* jakauman kannalta on yhdentekevää, ajatellaanko autokorrelaatiot lasketuiksi jäännöksistä e_t vai vastaavista virhetermeistä ε_t . Merkitään

$$T_t = \begin{pmatrix} \varepsilon_t \varepsilon_{t-1} \\ \vdots \\ \varepsilon_t \varepsilon_{t-K} \end{pmatrix} \quad t = 1, \dots, n - K \quad .$$

Tällöin $ET_t \equiv 0$ ja T_t :t ovat identtisesti jakautuneita, keskenään korreloimattomia ja

$$\text{cov}(T_t) = ET_t T_t' = \sigma^4 I \quad .$$

Keskeisen raja-arvolauseen mukaan siis

$$Y_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n-K} \frac{1}{s^2} T_t \stackrel{\text{asympt.}}{\sim} N_K(0, I) \quad , \quad (3.9)$$

jossa $s^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2$.

Toisaalta

$$\sqrt{n} R = \text{diag} \left(\left(\frac{n}{n-1} \right), \dots, \left(\frac{n}{n-K} \right) \right) Y_n \quad ,$$

joten tulos (3.8) seuraa suoraan (3.9):stä. Samalla nähdään, että

$$n R' R = n \sum_{\nu=1}^K r_e(\nu)^2 \stackrel{\text{asympt.}}{\sim} \chi_K^2 \quad (3.10)$$

Tähän asymptoottiseen jakaumatulokseen perustuvaa testiä hypoteesille (3.5) kutsutaan yleensä *Box-Pierce-testiksi*.

Pienien havaintoaineistojen osalta päästään kuitenkin parempaan jakauma-aproksimaatioon käyttämällä seuraavaa (3.10):n muunnelmaa:

$$\begin{aligned}
Q &= (n+2) \sum_{\nu=1}^K \left(1 - \frac{\nu}{n}\right) r_e(\nu)^2 \\
&= (n+2) R' \operatorname{diag}\left(\frac{n-1}{n}, \dots, \frac{n-K}{n}\right) R \\
&\underset{\text{asympt.}}{\sim} \chi_K^2
\end{aligned} \tag{3.11}$$

Tulokseen (3.11) perustuvaa testiä sanotaan *Box-Ljung-testiksi*. Nämä testit soveltuvat hypoteesin (3.5) testaamiseen huomattavasti paremmin kuin Durbin-Watson-testit, mutta asymptoottisiin jakaumatuloksiin turvautuminen tietysti edellyttää, että käytettävissä olevien havaintojen lukumäärä on riittävän suuri.

Huomautus 3.3: Mikäli laajennettu malli (1.6)–(1.9)+(3.1) todella ollaan valmiita estimoimaan, voidaan suppeamman mallin (1.6)–(1.9) realistisuutta tietysti tutkia myös vertaamalla parametriestimaatteja $\hat{\phi}_1, \dots, \hat{\phi}_p$ ja $\hat{\theta}_1, \dots, \hat{\theta}_q$ tuloksen (2.8) mukaisesti arvioituihin hajontoihinsa.

Mikäli lisäksi halutaan tutkia laajennetun mallin realistisuutta α_t -termien riippumattomuusoletuksen osalta, voidaan estimoiduista $\hat{\alpha}_t$ -jäännöstermeistä laskea autokorrelaatioita $r_{\hat{\alpha}}(\nu)$ ja niistä voidaan puolestaan panna kokoon Box-Ljung-testisuure (3.11). Tällöin on referenssijakaumana kuitenkin käytettävä χ_{K-p-q}^2 -jakaumaa *eikä* χ_K^2 -jakaumaa. ■

3.1.3 Parametrien estimointi autokorreloituneiden virhetermien tapauksessa

Mikäli luvissa 3.1.1 ja 3.1.2 esitetyt testit antavat aihetta epäillä, että mallin (1.6)–(1.9) virhetermit ε_t oisivat autokorreloituneita, on mallityypin (3.1) piiristä ilmeisesti yritettävä löytää sopiva kuvaustapa havaitulle autokorrelaatiolle. Merkitään virhetermimallin (3.1) parametriblokkeja symbolein $\psi = (\zeta' \ \sigma^2)'$ ja $\zeta = (\phi_1 \ \dots \ \phi_p \ \theta_1 \ \dots \ \theta_q)'$. Kuten esimerkin 2.2 yhteydessä todettiin, on täydennettyyn mallikokonaisuuteen (1.6)–(1.9)+(3.1) liittyvä informaatiomatriisi $\mathcal{I}(\beta, \psi)$ *blokkidiagonaalinen*, joten tuloksen (2.8) mukaan β :n ML-estimaattoria

$$\hat{\beta} = [X' \Omega(\hat{\zeta})^{-1} X]^{-1} X' \Omega(\hat{\zeta})^{-1} Y$$

koskee seuraava asymptoottinen jakaumatulos:

$$\hat{\beta} \stackrel{\text{asympt.}}{\sim} N_m(\beta, \sigma^2 [X' \Omega(\zeta)^{-1} X]^{-1}) \quad . \quad (3.12)$$

jossa

$$\Omega(\zeta) = \frac{1}{\sigma^2} \Sigma(\psi) = \frac{1}{\sigma^2} \text{cov}((\varepsilon_1 \ \dots \ \varepsilon_n)') \quad .$$

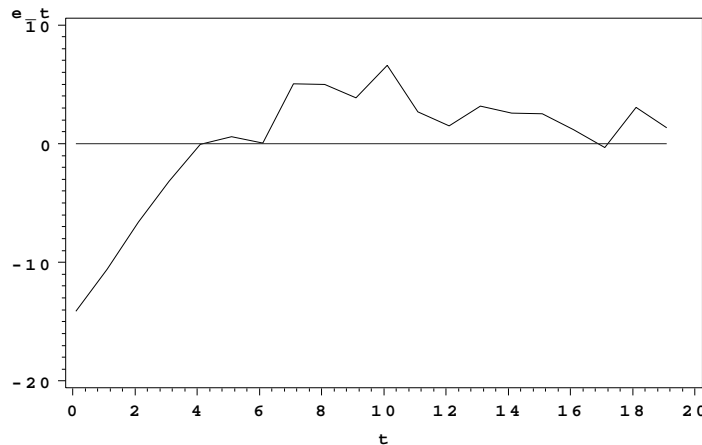
Mallin (3.1) yksinkertaisimman erikoistapauksen

$$\begin{aligned} \varepsilon_t &= \phi \varepsilon_{t-1} + \alpha_t \quad , \\ \alpha_t &\sim NID(0, \sigma^2) \quad , \quad \{\varepsilon_t\} \text{ stationäärinen} \end{aligned} \quad (3.13)$$

osalta voidaan todeta, että parametrien β, ϕ ja σ^2 ML-estimaattorien asymptoottiset jakaumat ovat muotoa (3.12) ja

$$\begin{aligned} \hat{\phi} &\stackrel{\text{asympt.}}{\sim} N\left(\phi, \frac{1 - \phi^2}{n}\right) \\ \hat{\sigma}^2 &\stackrel{\text{asympt.}}{\sim} N\left(\sigma^2, \frac{2\sigma^4}{n}\right) \end{aligned} \quad (3.14)$$

Huomautus 3.4: Kannattaa lopuksi korostaa, että mallin (1.6) – (1.9) realisti- suutta voidaan *kaikkein* parhaiten tutkia visuaalisesti. Virhetermien ajallista säännöttömyyttä arvioitaessa ei koskaan pitäisi laiminlyödä *OLS-residuaalien aikaurakuvion* piirtämistä. Esimerkiksi oheinen aikaurakuvio olisi huomattavasti informatiivisempi kuin vastaavat autokorrelaatiolaskelmat:



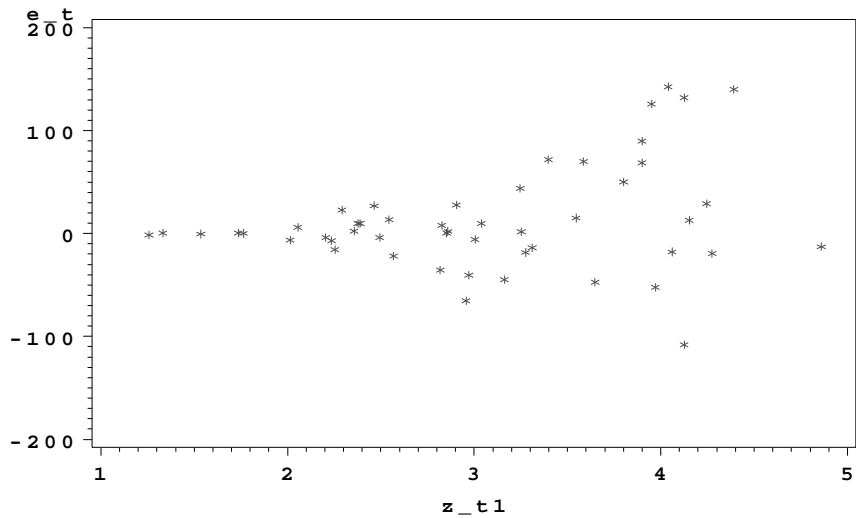
█

3.2 Virhetermien heteroskedastisuuden havaitseminen

Huomautus 3.5: Virhetermien ε_t mahdollista heteroskedastisuutta tutkittaessa on suureksi avuksi, jos on olemassa jonkinlainen käsitys siitä, *minkä tekijöiden epäillään vaikuttavan virhevarianssin $\sigma_t^2 = \text{var}(\varepsilon_t)$ muuttumiseen* havaintoyksiköstä (ajankohdasta) toiseen. Merkitään näistä muuttujista saatuja havaintoja vektorisymboleilla

$$Z_t = (z_{t1} \quad \dots \quad z_{tp})' \quad , \quad t = 1, \dots, n \quad .$$

Tavallisimmin Z_t -vektoriin sisältyy lähinnä mallissa jo olevia selittäviä muuttujia (x -muuttujia) tai niiden transformaatioita, mutta mukana voi olla myös muita tekijöitä (esim. aika t). Tällöin ensimmäisenä tehtävänä on plotata OLS-jäännöstermit e_t kutakin z_{tj} -muuttujaa vastaan vuorollaan, sillä plottauskuviot ovat heteroskedastisuuden havaitsemisessa aina paljon informatiivisempia kuin mitkään yksittäisten testisuureiden arvot. Esimerkiksi seuraavan kaltainen kuvio indikoisi selvää heteroskedastisuutta:



Jos perusmallin (1.6) – (1.9) laajennukseksi heteroskedastisuuden suuntaan ajatellaan muotoa

$$\sigma_t^2 = \text{var}(\varepsilon_t) = E\varepsilon_t^2 = \sigma^2 + \gamma'Z_t \quad (3.15)$$

olevaa väljennystä oletukseen (1.7), päädytään myöhemmin esiteltävän LM-testausperiaatteen avulla seuraavaan testiproseduriin:

Lasketaan OLS-jäännökset e_t ja käytetään perusmallin (1.6) – (1.9) puitteissa virhevarianssiestimaattorina tavanomaista lauseketta

$$s^2 = \frac{1}{n-m} \sum_{t=1}^n e_t^2 \quad (3.16)$$

Muodostetaan sitten muuttujat

$$u_t = \frac{e_t^2}{s^2} \quad t = 1, \dots, n \quad (3.17)$$

ja sovitetaan niihin mallia

$$u_t = \sigma^2 + \gamma'Z_t + \nu_t \quad , \quad E\nu_t = 0 \quad , \quad t = 1, \dots, n \quad (3.18)$$

OLS-estimointimenetelmää käyttäen. Merkitään näin saatuja sovitteita symbolein \hat{u}_t ($t = 1, \dots, n$).

Homoskedastisuusoletusta (1.7) vastaisi laajennuksen (3.15) puitteissa hypoteesi

$$H_o : \quad \gamma = 0 \quad (3.19)$$

jota puolestaan voidaan testata suurella

$$BP = \frac{1}{2} \sum_{t=1}^n (\hat{u}_t - \bar{u})^2 \stackrel{\text{asympt.}}{\sim} \chi_p^2 \quad . \quad (3.20)$$

Tätä testisuuretta ovat ehdottaneet *Breusch ja Pagan*, ja sen otantajakauman pitäisi hypoteesin (3.19) vallitessa lähestyä χ_p^2 -jakaumaa havaintomäärän kasvaessa.

Koska OLS-jäännökset $u_t - \hat{u}_t$ ovat automaattisesti korreloimattomia sovitteiden \hat{u}_t kanssa, on

$$\begin{aligned} \sum_{t=1}^n (\hat{u}_t - \bar{u})^2 &= \sum_{t=1}^n (u_t - \bar{u})^2 - \sum_{t=1}^n (u_t - \hat{u}_t)^2 \\ &= \frac{R^2}{1 - R^2} \sum_{t=1}^n (u_t - \hat{u}_t)^2 \quad . \end{aligned}$$

Tästä nähdään, että *BP*-testisuure (3.20) voidaan poimia kätevästi minkä tahansa regressio-ohjelmiston mallisovituksen (3.18) jälkeisestä tulostuksesta muodossa

$$BP = \frac{1}{2} \frac{R^2}{1 - R^2} RSS \quad .$$

■

Esimerkki 3.1: Oheisessa tulostuksessa on estimoitu PCGIVE:a käyttäen muotoa (1.15') oleva tuotantofunktio-malli Suomen paperi-, paperituote- ja puumassateollisuudelle vuosien 1960 - 1990 aineiston perusteella. Jäännöstermit on plotattu sovitteita vastaan ja apuregressioon (3.18) liittyvät laskelmat on suoritettu. Breuschin ja Paganin testisuureen arvoksi saadaan

$$BP = \frac{1}{2} \frac{0.3247}{1 - 0.3247} 52.5736 = 12.64$$

joka on χ_2^2 -jakaumaan verrattuna kovin suuri!

■

Algebra code for Cdklem.in7:

LYt = log(Yt);

LKt = log(Kt);

LLt = log(Lt);

EQ(1) Modelling LYt by OLS (using Cdklem.in7)

The present sample is: 1960 to 1990

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	-0.51968	1.6373	-0.317	0.7533	0.0036
LKt	1.0915	0.066181	16.493	0.0000	0.9067
LLt	-0.34111	0.17196	-1.984	0.0572	0.1232

R² = 0.937127 F(2,28) = 208.67 [0.0000] \sigma = 0.0903506 DW = 0.934
RSS = 0.2285703706 for 3 variables and 31 observations

Residual saved to Cdklem.in7

Algebra code for Cdklem.in7:

u = Residual*Residual;

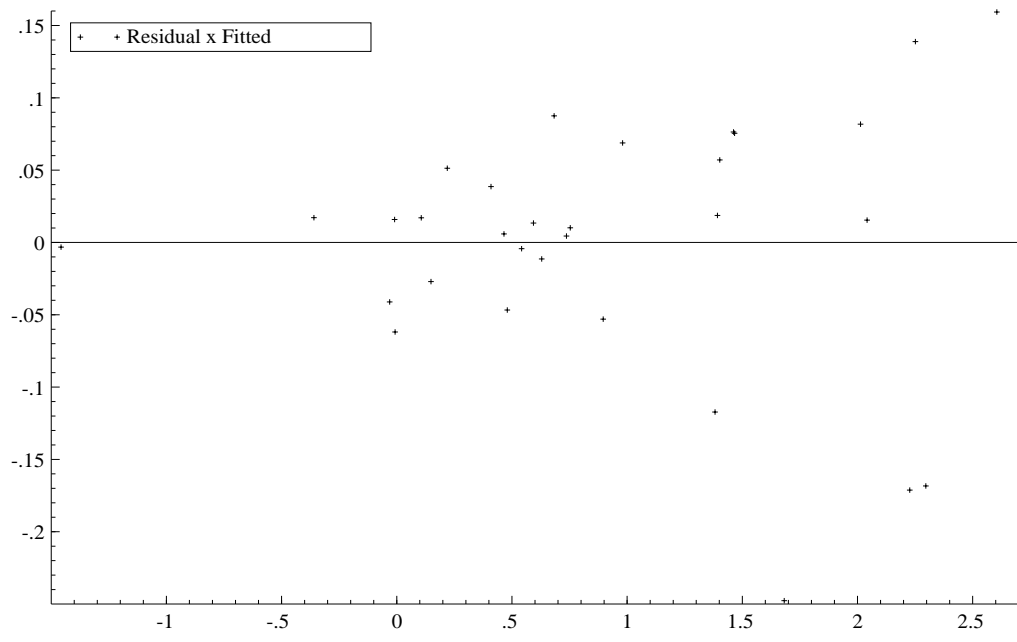
u = u/(0.09035²);

EQ(2) Modelling u by OLS (using Cdklem.in7)

The present sample is: 1960 to 1990

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	-90.126	24.832	-3.629	0.0011	0.3199
LKt	2.8178	1.0037	2.807	0.0090	0.2197
LLt	9.0310	2.6080	3.463	0.0017	0.2998

R² = 0.324732 F(2,28) = 6.7325 [0.0041] \sigma = 1.37027 DW = 1.19
RSS = 52.57364672 for 3 variables and 31 observations



Huomautus 3.6: Jos aineistossa on havaittu esiintyvän häiritsevää heteroskedastisuutta ja tutkijalla on täysin selkeä käsitys heteroskedastisuuden luonteesta (varianssit $\sigma_t^2 = \text{var}(\varepsilon_t) = \sigma^2 \alpha_t$ vaihtelevat havaintoyksiköittäin; luvut $\alpha_1, \dots, \alpha_n$ oletetaan tunnetuiksi), voidaan heteroskedastisuus ottaa huomioon seuraavasti: Määritellään ns. *painokertoimet*

$$w_t = \frac{\sigma^2}{\sigma_t^2} = \frac{1}{\alpha_t} \quad t = 1, \dots, n \quad . \quad (3.21)$$

Jos alkuperäiset, havaintoyksiköittäiset regressiomallit

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_m x_{tm} + \varepsilon_t$$

kerrotaan $\sqrt{w_t}$:llä, päästään muotoon

$$\sqrt{w_t} y_t = \beta_1 \sqrt{w_t} + \beta_2 \sqrt{w_t} x_{t2} + \dots + \beta_m \sqrt{w_t} x_{tm} + \sqrt{w_t} \varepsilon_t \quad t = 1, \dots, n \quad . \quad (3.22)$$

Mallien (3.22) virhetermien varianssit ovat kaikki saman suuruisia, sillä

$$\sqrt{w_t} \varepsilon_t \sim NID(0, \sigma^2) \quad t = 1, \dots, n \quad .$$

Malli (3.22) toteuttaa siis kaikki perusoletukset (1.6) – (1.9), joiden vallitessa OLS todettiin optimaaliseksi estimointimenetelmäksi. Kannattaa siis estimoida β siten, että malliin (3.22) liittyvä jäännösneliösumma

$$Q^*(\beta) = \sum_{t=1}^n (\sqrt{w_t} y_t - \sqrt{w_t} \beta' X_t)^2 = \sum_{t=1}^n w_t (y_t - \beta' X_t)^2 \quad (3.23)$$

minimoituu. Merkitään näin saatavaa β :n estimaattoria symbolilla $\hat{\beta}^*$ ja kutsutaan sitä *painotetuksi* pienimmän neliösumman estimaattoriksi (tai GLS-estimaattoriksi). Merkitään vastaavaa jäännösneliösummaa symbolilla

$$\begin{aligned} Q^* &= Q^*(\hat{\beta}^*) = \sum_{t=1}^n w_t (y_t - \hat{\beta}^{*'} X_t)^2 \\ &= \sum_{t=1}^n w_t e_t^{*2} \end{aligned} \quad (3.24)$$

Kertoimia β koskevien hypoteesien testaaminen pitäisi tietysti perustaa *painotettuihin* jäännösneliösummiin (3.24), koska vain mallimuoto (3.22) täyttää perusoletukset (1.6) – (1.9). Painotetun ja painottamattoman testisuureen mukaisilla johtopäätöksillä saattaa joissakin tilanteissa olla huikea ero!

Joissakin tilanteissa painotuksen tarve voi olla etukäteen täysin selvä, eikä mitään

testejä heteroskedastisuuden havaitsemiseen tarvita. Ajatellaan esimerkiksi, että haluttaisiin tutkia kuntien elinkeinorakenteiden vaikutuksia naisten kunnittaisiin osallistumisasteisiin. Olisi selvästikin syytä olettaa, että osallistumisasteisiin liittyvän satunnaisvaihtelun varianssi olisi kääntäen verrannollinen kunnan asukaslukuun. ■

Huomautus 3.7: Jos mahdollisen heteroskedastisuuden luonteesta ei tiedetä kerrassaan mitään (jos mitään kovariaatteja Z_t ei ole käytettävissä), ei kaikkia σ_t^2 -parametreja tietenkään pystytä estimoimaan. Jos $\Sigma = \text{cov}(\varepsilon) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ on diagonaalimatriisi, pystytään matriisiin

$$\lim_{n \rightarrow \infty} \frac{1}{n} X' \Sigma X$$

elementit estimoimaan konsistentisti estimaattorilla

$$\hat{V} = \frac{1}{n} \sum_{t=1}^n e_t^2 X_t X_t' \quad (3.25)$$

Merkitään $M = \frac{1}{n} X' X$.

Tällöin tavallisen OLS-estimaattorin $\hat{\beta}$ symptoottiseksi otantajakaumaksi saadaan

$$\sqrt{n} (\hat{\beta} - \beta) \stackrel{\text{asympt.}}{\sim} N_m(0, M^{-1} \hat{V} M^{-1}), \quad (3.26)$$

joten esimerkiksi parametrirajoitteiden $R\beta = \gamma \in \mathcal{R}^r$ realistisuutta voidaan testata testisuurella

$$n (R\hat{\beta} - \gamma)' [RM^{-1} \hat{V} M^{-1} R']^{-1} (R\hat{\beta} - \gamma) \stackrel{\text{asympt.}}{\sim} \chi_r^2. \quad (3.27)$$

Virhetermien homoskedastisuutta voidaan testata vertaamalla \hat{V} :n elementtejä homoskedastisuusoletuksen mukaisen vastaavan estimaattorin $s^2 M$ elementteihin:

$$\begin{cases} \psi_t = \text{vec}(X_t X_t') \in \mathcal{R}^{m(m+1)/2} \\ \psi = \text{vec}(M) = \frac{1}{n} \sum_{t=1}^n \psi_t \\ D = \text{vec}(\hat{V} - s^2 M) \\ B = \sum_{t=1}^n (e_t^2 - s^2) (\psi_t - \psi) (\psi_t - \psi)' \end{cases} .$$

Voidaan osoittaa, että homoskedastisuushypoteesin $\sigma_1^2 = \dots = \sigma_n^2 = \sigma^2$ vallitessa

$$n^2 D' B^{-1} D \stackrel{\text{asympt.}}{\sim} \chi_{m(m+1)/2}^2. \quad (3.28)$$

Tätä testiä on ehdottanut Halbert White (1980) ja siihen voidaan päätyä myös ns. IM-periaatteella. Edellä käytettyä OLS-estimaattorin $\hat{\beta}$ kovarianssimatriisin robustia estimaattoria $\widehat{\text{cov}}(\hat{\beta}) = \frac{1}{n} M^{-1} \hat{V} M^{-1}$ kutsutaan usein *sandwich*-estimaattoriksi. ■

3.3 Virhetermien normaalisuusoletuksen realistisuuden tutkiminen

Virhetermien normaalisuudesta tehtyä oletusta (1.9) voidaan tutkia OLS-jäännöstermien e_t avulla ($t = 1, \dots, n$) seuraavasti:

Merkitään Φ :llä $N(0, 1)$ - jakauman kertymäfunktioita. Järjestetään jäännöstermit e_t kasvavaan suuruusjärjestykseen $e_{(1)} \leq \dots \leq e_{(n)}$. Piirretään tasoon pisteet

$$(e_{(j)} \quad \Phi^{-1}(\frac{j-0.5}{n}))' \quad j = 1, \dots, n \quad (3.29)$$

Pisteiden pitäisi sijaita suurin piirtein samalla suoralla, jos e_t -lukujen jakauma olisi normaalinen. Tällaisia kuvioita sanotaan jäännöstermien *probit*-kuvioiksi ja niiden peilikuvia sanotaan *QQ*-kuvioiksi. Niiden tukena voidaan käyttää monenlaisia formaaleja testejä, joista ekonometristen sovellusten yhteydessä erikoisen suosituksi on tullut ns. *Jarque ja Beran* testi. Se perustuu jäännöstermisarjan vinouteen ja huipukkuuteen.

Satunnaismuuttujan z keskusmomenteillahan tarkoitetaan lukuja

$$\mu_\nu = E(z - \mu)^\nu \quad \nu = 2, 3, \dots,$$

jossa $\mu = Ez$. Keskusmomentteihin perustuvina vinous- ja huipukkuusmittareina käytetään tavallisesti suureita

$$\gamma_S = \frac{\mu_3}{(\mu_2)^{3/2}} \quad \text{ja} \quad \gamma_K = \frac{\mu_4}{(\mu_2)^2} - 3 \quad (3.30)$$

Kuten tunnettua, normaalijakaumaa noudattavalle muuttujalle $z \sim N(\mu, \sigma^2)$ pätee $\gamma_S = 0$ ja $\gamma_K = 0$.

Merkitään nyt symboleilla

$$\hat{\mu}_\nu = \frac{1}{n} \sum_{t=1}^n (e_t - \bar{e})^\nu \quad \nu = 2, 3, \dots$$

OLS-jäännöksistä laskettuja keskusmomentteja, ja

$$\hat{\gamma}_S = \frac{\hat{\mu}_3}{(\hat{\mu}_2)^{3/2}} \quad \text{ja} \quad \hat{\gamma}_K = \frac{\hat{\mu}_4}{(\hat{\mu}_2)^2} - 3 \quad (3.31)$$

Jarque ja Bera osoittivat, että virhetermien normaalisuuden testaamiseen voidaan käyttää suurta

$$JB = n \left[\frac{1}{6} \hat{\gamma}_S^2 + \frac{1}{24} \hat{\gamma}_K^2 \right] \stackrel{\text{asympt.}}{\sim} \chi^2 \quad (3.32)$$

Jos normaalisuusoletuksen vaihtoehtona pidetään sitä, että virhetermin jakauma kuuluisi ns. Pearsonin jakaumaperheeseen (joka erikoistapauksinaan sisältää mm. normaali-, beta-, gamma-, t -, F - ja Pareto-jakaumat), voidaan Jarque-Bera-testiä pitää oletusta (1.9) testaavana LM-testinä.

Huomautus 3.8: Jos virhetermin autokorrelloimattomuutta, homoskedastisuutta ja normaalisuutta halutaan testata samanaikaisesti, voidaan testisuurena käyttää Box-Pierce-testisuureen (3.10), Breusch-Pagan-testisuureen (3.20) ja Jarque-Bera-testisuureen (3.32) summaa

$$\begin{aligned}
 & n \sum_{\nu=1}^K r_{\epsilon}(\nu)^2 + \frac{1}{2} \sum_{t=1}^n (\hat{u}_t - \bar{u})^2 \\
 & + n \left[\frac{1}{6} \hat{\gamma}_S^2 + \frac{1}{24} \hat{\gamma}_K^2 \right] \qquad (3.33) \\
 & \underset{\text{asympt.}}{\sim} \chi_{K+p+2}^2
 \end{aligned}$$

Tämä on täysin ymmärrettävää, koska kaikki kolme testiä voidaan tulkita LM-periaatteen mukaisiksi. ■

Huomautus 3.9: On helppo osoittaa, että

$$\begin{aligned}
 x_i & \sim i.i.d.(F_x) \qquad i = 1, \dots, n \\
 \implies z_i = F_x(x_i) & \sim i.i.d.(\text{tas}(0, 1)) \qquad i = 1, \dots, n
 \end{aligned}$$

Koska tasaisen jakauman kertymäfunktio on muotoa $F_z(y) = y$ kun $0 \leq y \leq 1$, on helppo todeta, että

$$KS = \max_{0 \leq y \leq 1} |F_{z_1, \dots, z_n}^*(y) - y| \qquad (3.34)$$

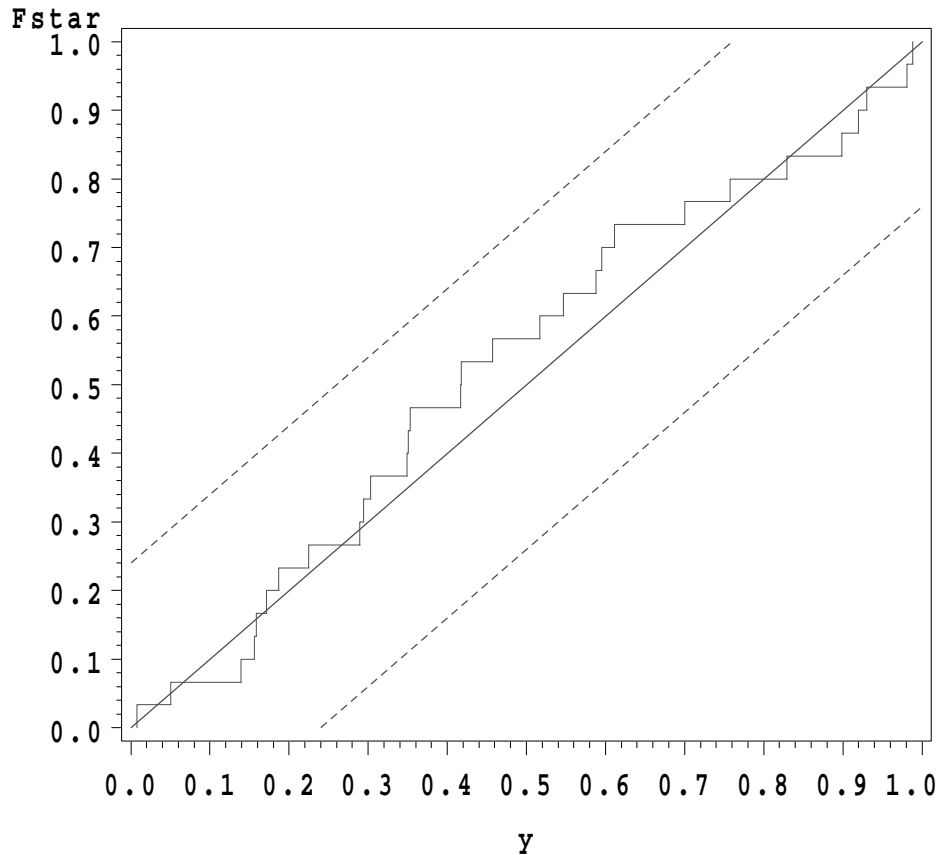
on jakaumaltaan täysin riippumaton F_x :n muodosta. Kaavassa (3.34) $F_{z_1, \dots, z_n}^*(y)$ tarkoittaa aineiston z_1, \dots, z_n kumulatiivista frekvenssifunktiota. Suure KS tunnetaan *Kolmogorov-Smirnov*-testisuurena ja siihen liittyviä jakaumataulukkoita on yleisesti saatavilla.

Jäännöstermien normaalisuutta tutkittaessa transformoidaan residuaalit e_t siis muunnoksella

$$z_t = \Phi\left(\frac{e_t - \bar{e}}{s}\right) \qquad t = 1, \dots, n \qquad (3.35)$$

ja muodostetaan z_t -havaintojen kumulatiivinen frekvenssifunktio $F_{z_1, \dots, z_n}^*(y)$. Hakemalla taulukosta KS -testisuureen haluttua merkitsevyytastasoja vastaava kriittinen arvo voidaan testi esittää visuaalisesti hyvin havainnollisessa muodossa:

Kuva 3.1: Erään transformoidun havaintoaineiston kumulatiivinen frekvenssifunktio $F_{z_1, \dots, z_n}^*(y)$ ($n = 30$)



Koska transformaatiassa (3.35) on jouduttu estimoimaan parametrit σ ja μ , eivät Kolmogorov-Smirnov-taulukoista saatavat kriittiset arvot ole täsmälleen oikeita tähän tarkoitukseen. Liliefors on muodostanut taulukot muunnoksen (3.35) mukaiselle normaalisuustestille, ja näitäkin taulukoita on melko hyvin saatavissa. ■

Kolmogorov-Smirnov-testisuureen (3.34) kriittisiä arvoja

n	Merkitsevyystaso				
	0.20	0.15	0.10	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.510	0.565	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.368	0.410	0.490
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.450
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
> 35	$1.07/\sqrt{n}$	$1.14/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

3.4 Mallin rakenteessa tapahtuvien muutosten paljastaminen

3.4.1 Rekursiivinen estimointi

Tarkastellaan perusmuotoisen lineaarisen regressiomallin (1.6) – (1.9)

$$Y = X\beta + \varepsilon \quad ,$$

$$\varepsilon \sim N_n(0, \sigma^2) \quad , \quad Y = (y_1 \quad \dots \quad y_n)' \quad , \quad X = (X_1 \quad \dots \quad X_n)'$$

parametrien OLS-estimaattien kehitystä, kun havaintoja lisätään estimointiaineistoon yksi kerrallaan aikajärjestyksessä. Otetaan käyttöön seuraavat merkin-

nät:

$$Y_{(t)} = (y_1 \quad \dots \quad y_t)' \quad ,$$

$$X_{(t)} = (X_1 \quad \dots \quad X_t)' \quad ,$$

$$\hat{\beta}_{(t)} = \left(X_{(t)}' X_{(t)} \right)^{-1} X_{(t)}' Y_{(t)} \quad ,$$

$$M_{(t)} = X_{(t)}' X_{(t)} \quad ,$$

$$V_t = M_{(t-1)}^{-1} X_t \quad ,$$

$$g_t = 1 + X_t' M_{(t-1)}^{-1} X_t \quad \text{ja}$$

$$f_t = y_t - X_t' \hat{\beta}_{(t-1)} \quad t = m + 1, \dots, n \quad .$$

Apulause 3.1: Mikäli kaikki seuraavassa kaavassa esiintyvät käänteismatriisit ovat hyvin määriteltyjä, niin

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1} \quad (3.36)$$

Tulos tunnetaan yleensä ”matriisien inversiolemman” tai ”Sherman-Morrison-Woodburyn kaavan” nimellä. ■

Apulauseesta 3.1 seuraa heti, että

$$M_{(t)}^{-1} = (M_{(t-1)} + X_t X_t')^{-1}$$

$$= M_{(t-1)}^{-1} - M_{(t-1)}^{-1} X_t (1 + X_t' M_{(t-1)}^{-1} X_t)^{-1} X_t' M_{(t-1)}^{-1} \quad (3.37)$$

$$= M_{(t-1)}^{-1} - g_t^{-1} V_t V_t' \quad t = m + 1, \dots, n \quad .$$

Kun tämä tulos sijoitetaan OLS-estimaattorin $\widehat{\beta}_{(t)}$ kaavaan, saadaan

$$\begin{aligned}\widehat{\beta}_{(t)} &= \left[M_{(t-1)}^{-1} - g_t^{-1} V_t V_t' \right] \left(X_{(t-1)}' Y_{(t-1)} + y_t X_t \right) \\ &= \widehat{\beta}_{(t-1)} + (y_t - X_t' \widehat{\beta}_{(t-1)}) g_t^{-1} V_t \quad .\end{aligned}\tag{3.38}$$

Kokoamalla ”ylläpitokaavat” (3.37) ja (3.38) yhteen saadaan

$$\begin{aligned}M_{(t)}^{-1} &= M_{(t-1)}^{-1} - g_t^{-1} V_t V_t' \\ \widehat{\beta}_{(t)} &= \widehat{\beta}_{(t-1)} + f_t g_t^{-1} V_t \\ f_t &= y_t - X_t' \widehat{\beta}_{(t-1)} \\ V_t &= M_{(t-1)}^{-1} X_t \\ g_t &= 1 + X_t' V_t\end{aligned}\tag{3.39}$$

joita kutsutaan usein *Plackettin kaavoiksi*.

Kuten huomataan, rekursiivisten estimaattien $\widehat{\beta}_{(t)}$ muodostaminen on laskennallisesti tavattoman vaivatonta. (Samat kaavat voidaan muuten johtaa myös ns. Kalman-suodatuksen sivutuotteina.)

Huomautus 3.10: Estimaattien aikaurien tarkastelu saattaa olla tavattoman valaisevaa, sillä estimaattien äkilliset muutokset heijastelevat tietysti viimeksi tulleiden havaintojen suurta vaikutusvaltaisuutta. Seuraavilla kalvoilla on tarkasteltu viinien kulutuksen vuosimuutoksia koskevan mallin

$$\nabla_3 \log q_t = \beta_1 + \beta_2 \nabla_3 \log p_t + \beta_3 \nabla_3 \log Q_t + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2)$$

β_2 -parametrin OLS-estimaattien aikauraa. (Havaintoaineistona on käytetty neljän kuukauden jaksoissa mitattuja tietoja viinien kulutuksen määristä q_t , viinien reaalihinnoista p_t ja yksityisen kokonaiskulutuksen määrästä Q_t vuosilta 1970-1986. Symboli ∇_3 tarkoittaa *vuosimuutosta*, esim. $\nabla_3 \log q_t = \log q_t - \log q_{t-3}$.) ■

```

Algebra code for Wines.in7:
Lviinivol = log(viinivol);
Lkokkulut = log(kokkulut);
p = viinip/kulhinta;
Lp = log(p);
D3Lviinivol = diff(Lviinivol,3);
D3Lkokkulut = diff(Lkokkulut,3);
D3Lp = diff(Lp,3);

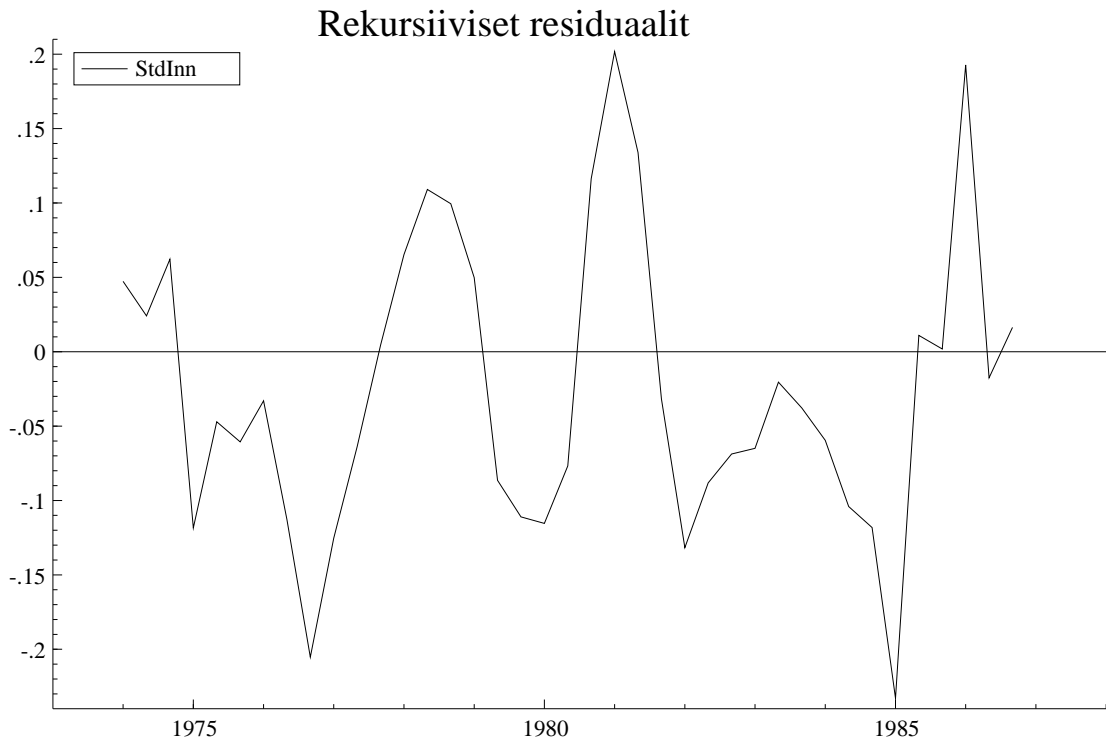
```

EQ(1) Modelling D3Lviinivol by RLS (using Wines.in7)
The present sample is: 1970 (1) to 1986 (3)

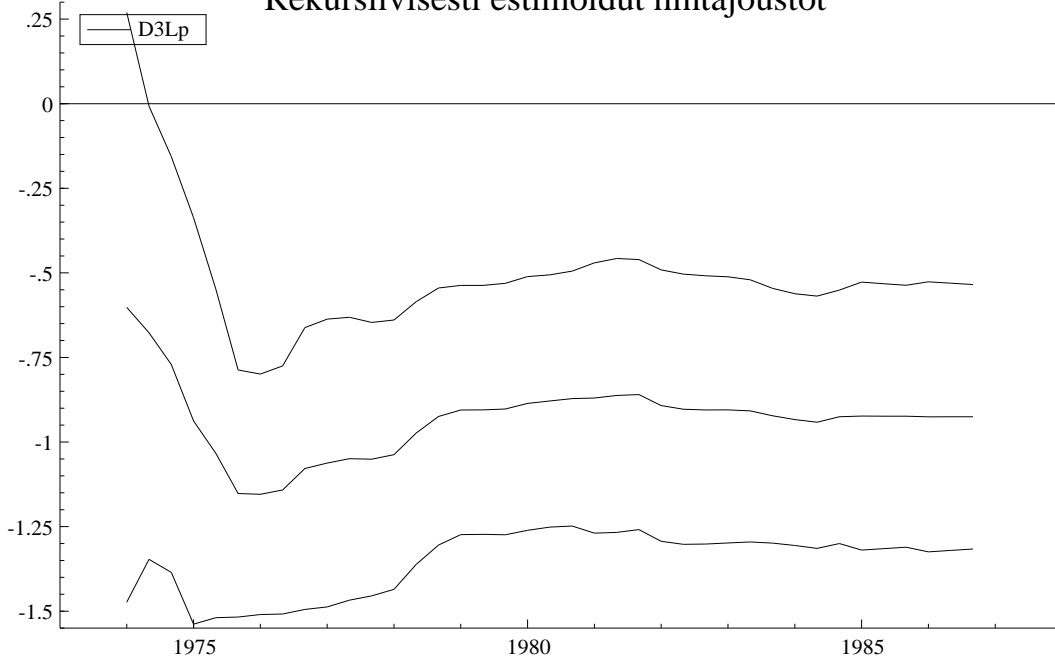
Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	-0.013880	0.022403	-0.620	0.5385	0.0079
D3Lkokkulut	1.2763	0.54612	2.337	0.0237	0.1022
D3Lp	-0.92532	0.19542	-4.735	0.0000	0.3184

R² = 0.333542 F(2,48) = 12.011 [0.0001] \sigma = 0.0986196 DW = 0.973
RSS = 0.466839459 for 3 variables and 51 observations

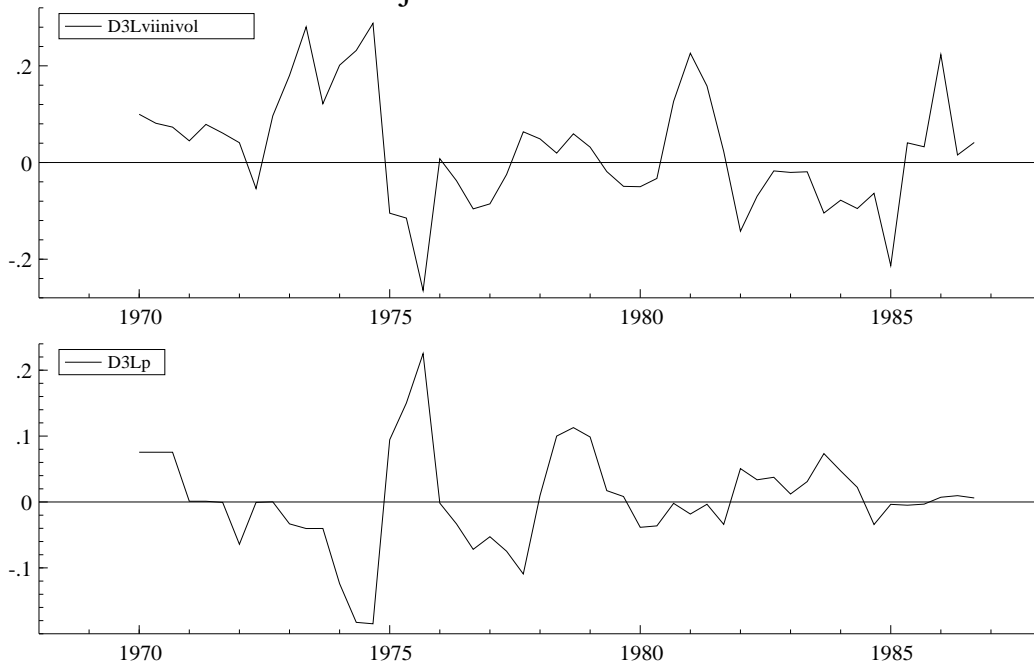
StdInn saved to Wines.in7



Rekursiivisesti estimoidut hintajoustot



Kulutus- ja hintamuutosten aikaurat



Huomautus 3.11: Rekursiokaavat (3.39) tuottavat paljon muitakin mielenkiintoisia sivutuotteita kuin parametrien askeltavat estimaatit $\hat{\beta}_{(t)}$. Mainittakoon erityisesti ”yhden askeleen ennustevirheet” f_t ($t = m + 1, \dots, n$), joilla mallin (1.6) – (1.9) puitteissa on mm. seuraavat ominaisuudet:

$$\begin{aligned} f_t &= X_t' \beta + \varepsilon_t - X_t' M_{(t-1)}^{-1} \sum_{i=1}^{t-1} X_i (X_i' \beta + \varepsilon_i) \\ &= \varepsilon_t - X_t' M_{(t-1)}^{-1} \sum_{i=1}^{t-1} X_i \varepsilon_i \quad , \end{aligned}$$

joten

$$\begin{aligned} E f_t f_{t+\nu} &= E \left(\varepsilon_t - \sum_{i=1}^{t-1} \varepsilon_i X_t' M_{(t-1)}^{-1} X_i \right) \left(\varepsilon_{t+\nu} - \sum_{j=1}^{t+\nu-1} \varepsilon_j X_{t+\nu}' M_{(t+\nu-1)}^{-1} X_j \right) \\ &= -E \varepsilon_t \varepsilon_t X_t' M_{(t+\nu-1)}^{-1} X_{t+\nu} + E \sum_{j=1}^{t-1} \varepsilon_j^2 X_t' M_{(t-1)}^{-1} X_j X_j' M_{(t+\nu-1)}^{-1} X_{t+\nu} \quad . \end{aligned} \tag{3.40}$$

Tästä seuraa välittömästi, että mallin (1.6) – (1.9) perusoletusten vallitessa

$$E f_t f_{t+\nu} = \begin{cases} 0 & \text{jos } \nu \geq 1 \\ \sigma^2 g_t & \text{jos } \nu = 0 \end{cases} \quad .$$

Koska ennustevirheiden f_t pitäisi lisäksi olla (normaalisten muuttujien lineaarilausekkeina) normaalisesti jakautuneita, saadaan

$$f_t \sim NID(0, \sigma^2 g_t) \quad t = m + 1, \dots, n \tag{3.41}$$

Tulos (3.41) antaa aiheen määrittellä ns. (normitetut) *rekursiiviset residuaalit*

$$e_t^* = \frac{f_t}{\sqrt{g_t}} \quad t = m + 1, \dots, n \tag{3.42}$$

jotka ovat siis normitettuja yhden askeleen ennustevirheitä (”standardoituja innovaatioimpulsseja”), ja joille pätee

$$e_t^* \sim NID(0, \sigma^2) \quad t = m + 1, \dots, n \tag{3.43}$$

Koska e_t^* -termit ovat määritelmänsä (3.42) mukaan selvästi ”residuaalin” luonteisia, ja koska niillä on ominaisuus (3.43), ne sopivat kaikkeen tässä luvussa esitettyyn mallidiagnostiikkaan oikeastaan paremmin kuin OLS-residuaalit e_t konsanaan.

Rekursiivisten residuaalien muista ominaisuuksista kannattaa erityisesti mainita, että jäännösneliösummalaskelmalle

$$Q_{(t)} = \| Y_{(t)} - X_{(t)} \widehat{\beta}_{(t)} \|^2 \quad (3.44)$$

pätee rekursiokaava

$$Q_{(t)} = Q_{(t-1)} + e_t^{*2} \quad t = m + 1, \dots, n \quad (3.45)$$

Jos rekursiivisten residuaalien muodostamaa vektoria merkitään symbolilla

$$E^* = (e_{m+1}^{*2} \quad \dots \quad e_n^{*2})' \quad ,$$

on siis erityisesti

$$Q(\widehat{\beta}) = E' E = E^{*'} E^* = \sum_{t=m+1}^n e_t^{*2} \quad . \quad (3.46)$$

Tulos (3.45) seuraa kaavoista (3.39), joiden mukaan

$$\widehat{\beta}_{(t)} = \widehat{\beta}_{(t-1)} + f_t g_t^{-1} V_t \quad ,$$

joten

$$\begin{aligned} Y_{(t-1)} - X_{(t-1)} \widehat{\beta}_{(t)} &= Y_{(t-1)} - X_{(t-1)} \widehat{\beta}_{(t-1)} - f_t g_t^{-1} X_{(t-1)} V_t \\ &= E_{(t-1)} - f_t g_t^{-1} X_{(t-1)} V_t \quad , \end{aligned}$$

jossa $E_{(t-1)} = Y_{(t-1)} - X_{(t-1)} \widehat{\beta}_{(t-1)}$ ja $E_{(t-1)}' X_{(t-1)} = 0$.

Näin ollen

$$\begin{aligned} Q_{(t)} &= \| Y_{(t-1)} - X_{(t-1)} \widehat{\beta}_{(t)} \|^2 + (y_t - X_t' \widehat{\beta}_{(t)})^2 \\ &= E_{(t-1)}' E_{(t-1)} + f_t^2 g_t^{-2} V_t' X_{(t-1)}' X_{(t-1)} V_t + (f_t - f_t g_t^{-1} X_t' V_t)^2 \\ &= Q_{(t-1)} + f_t^2 g_t^{-1} \quad , \end{aligned}$$

joten tulos (3.45) on todistettu. ■

3.4.2 Rakennemuutosten havaitseminen

Ajatellaan nyt, että malli

$$y_t = X_t' \beta + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad t = 1, \dots, n$$

kuvaa adekvaatisti X_t :n ja y_t :n välistä yhteyttä ajankohtina $t = 1, \dots, t_o$, mutta että periodilla $(t_o, t_o + 1]$ mallissa (tai pikemminkin havainnot tuottaneessa mekanismissa) tapahtuisi äkillinen muutos. Ajatellaan aluksi, että regressiokertoimet β äkillisesti muuttuisivat muotoon β^* . Mikäli hetki t_o tunnetaan (lainsäädännöllinen muutos, tilastointitavan muutos, äkillinen markkinahäiriö tms.), ja mikäli regressiokertoimien β^* oletetaan pysyvän vakioina koko jäljellä olevan havaintoperiodin ajan, voidaan hypoteesia $\beta = \beta^*$ testata seuraavasti:

- 1° Jos $n - t_o > m$ (ts. jos uudetkin regressiokertoimet pystytään estimoimaan), voidaan hypoteesi $\beta = \beta^*$ tulkita lineaarisiksi rajoitteiksi yleisemmän

$$\begin{cases} y_t = X_t' \beta + \varepsilon_t & t = 1, \dots, t_o \\ y_t = X_t' \beta^* + \varepsilon_t & t = t_o + 1, \dots, n \\ \varepsilon_t \sim NID(0, \sigma^2) \end{cases} \quad (3.47)$$

puitteissa. Tavanomainen F -testisuure hypoteesin $\beta = \beta^*$ testaamiseksi olisi tietenkin muotoa

$$C_{t_o} = \frac{\frac{1}{m}(Q_o - Q)}{\frac{1}{n-2m}Q} \sim F_{m, n-2m} \quad (3.48)$$

jossa $Q_o = E'_{(n)} E_{(n)}$ ja Q tarkoittaa malliin (3.47) liittyvää jäännösneliösummaa ilman rajoitusta $\beta = \beta^*$. Testisuuretta C_{t_o} kutsutaan yleensä *Chow-testisuureeksi*.

- 2° Jos $n - t_o \leq m$, ei β^* -parametreja pystytä lainkaan estimoimaan. Tällöinkin voidaan kuitenkin tarkkailla estimaattoreihin $\hat{\beta}_{(t_o)}$ perustuvien ennusteiden toimintakykyä periodilla $t_o + 1, \dots, n$. Jos $\beta^* \neq \beta$, olisivat ennusteet ilmeisesti *harhaisia*, ts.

$$E(y_{t+\nu} - X'_{t+\nu} \hat{\beta}_{(t_o)}) \neq 0 \quad \text{kun} \quad \nu = 1, \dots, n - t_o \quad .$$

Ennusteharhan olemassaoloa olisi luontevaa mitata suurella

$$C_{t_o}^{(n)} = \frac{\frac{1}{n-t_o}(E'_{(n)}E_{(n)} - E'_{(t_o)}E_{(t_o)})}{\frac{1}{t_o-m}E'_{(t_o)}E_{(t_o)}} \quad (3.49)$$

jota niinikään kutsutaan *Chow-testisuureeksi*. Koska rekursiokaavan (3.45) mukaan

$$Q_{(t)} = E'_{(t)}E_{(t)} = \sum_{\nu=m+1}^t e_{\nu}^{*2} \quad ,$$

voidaan (3.49) yhtä hyvin kirjoittaa muotoon

$$C_{t_o}^{(n)} = \frac{\frac{1}{n-t_o} \sum_{\nu=t_o+1}^n e_{\nu}^{*2}}{\frac{1}{t_o-m} \sum_{\nu=m+1}^{t_o} e_{\nu}^{*2}} \sim F_{n-t_o, t_o-m} \quad (3.49')$$

Testisuureen (3.49') toiminta perustuu siis mallin toimintakyvyn arviointiin estimointiperiodin ulkopuolella. Tällaista toimintaa sanotaan joskus ”prediktiiviseksi testaamiseksi”.

3° Jos rakennemuutoksen potentiaalisesta ajankohdasta t_o ei ole selvää käsitystä ja jos muuttuneen rakenteen ei oleteta pysyvän (ainakaan heti) vakiona, voidaan käyttää *jatkuvaa* kontrollia varten kehitettyjä päätössääntöjä, joista ensimmäisenä esiteltäköön ns. *CUSUM*-testi:

Kaavoissa (3.42) ja (3.39) määritellyt rekursiiviset residuaalit ovat erittäin soveliaita tarkastimia rakennemuutosten etsimisessä, sillä nehan voitiin tulkita normitetuiksi yhden askeleen ennustevirheiksi. Mikäli hetkellä t_o tapahtuisi β :aa koskeva rakennemuutos, olisi selvästikin

$$Ee_t^* \neq 0 \quad \text{kun} \quad t > t_o \quad . \quad (3.50)$$

Odotusarvoa koskevan hypoteesin tutkimiseen tarkoitettu testisuure kannattaa luonnollisesti perustaa odotusarvon tehokkaimpaan mahdolliseen estimaattoriin eli *keskiarvoon*. Jos nollahypoteesista

$$H_o : \quad Ee_t^* \equiv 0 \quad t = m + 1, \dots, n \quad (3.51)$$

tapahtuvat poikkeamat halutaan havaita mahdollisimman nopeasti, kannattaa keskiarvoa $\bar{e}_{(t)}^* = \frac{1}{t-m} \sum_{\nu=m+1}^t e_{\nu}^*$ tietysti seurata jatkuvasti t :n kasvaessa $t = m + 1, \dots, n$. Koska keskiarvoon liittyvän satunnaisvaihtelun mitakaava pienenee havaintomäärän kasvaessa, on kuitenkin miellyttävämpää seurata e_{ν}^* -termien *kumulatiivista summaa*

$$w_t = \sum_{\nu=m+1}^t e_{\nu}^* \quad t = m+1, \dots, n \quad . \quad (3.52)$$

Formaali, w_t -lukujonoon perustuva testi voidaan konstruoida seuraavan teknisen tuloksen avulla:

Jos $z_1, \dots, z_n \sim NID(0, 1)$, niin yhtälö

$$P\left(\left|\sum_{i=1}^t z_i\right| \leq b\left(\sqrt{n} + \frac{2}{\sqrt{n}}t\right) \quad \text{kaikilla } t = 1, \dots, n\right) = 1 - \alpha \quad (3.53)$$

toteutuu mm. lukupareilla

α	b
0.1	0.850
0.05	0.948
0.01	1.143

Epäyhtälön rajasuorat muodostavat eräänlaisen "V-maskin", jonka sisällä kumulatiivisen summan aikaura pysyttelee todennäköisyydellä $1 - \alpha$.

Tulosta (3.53) voidaan mainiosti soveltaa myös rekursiivisten residuaalien kumulatiivisen summan (3.52) seuraamiseen. V-maskia ei kuitenkaan kannata pitää kiinteänä, ajankohdasta $m+1$ alkavana, vaan "liu'uttaa" w_t -summan aikauraa pitkin. Suositeltavin menettelytapa olisi ehkä seuraava:

- Piirretään $(t \ w_t)'$ -pisteiden muodostama ura $t = m+1, \dots, n$.
- Haetaan silmämääräisesti uran epäilyttävien kohta ja merkitään sitä t_o :lla.
- Piirretään suorien

$$w_{t_o} \pm b s \left(\sqrt{n - t_o} + \frac{2}{\sqrt{n - t_o}}(t - t_o) \right)$$

kuvaajat aikaväliltä $t = t_o + 1, \dots, n$ ja "hylätään" vakiorakennehypoteesi (3.51), jos $(t \ w_t)'$ -ura ajautuu piirrettyjen suorien ulkopuolelle.

Tätä testiä kutsutaan (liukuvaa V-maskia käyttäväksi) CUSUM-testiksi.

4° Rekursiivisten residuaalien *neliöiden* kumulatiivisia summia voidaan niinkään käyttää vakiorakennehypoteesin realistisuuden tutkimiseen. Tällaisia testejä kutsutaan *CUSUMQ*-testeiksi.

Todetaan aluksi eräs tekninen, todennäköisyyslaskentaan liittyvä tulos, jota voidaan hyödyntää useissa eri yhteyksissä:

```
DLQt = "Yksityisen kokonaiskulutuksen vuosimuutos (log)"
DLYt = "Kotitalouksien käytettävissä olevien tulojen vuosimuutos (log)"
LSt_1 = "log(1-St_1/100), ts. edellisen vuoden säästämisaste (kotitaloudet)
        logaritmoituna"
```

Vuosien 1960-1974 vanha aineisto on yhdistetty Suomen kansantalouden uudistetun tilinpidon mukaisesti rekonstruoituihin tietoihin 1975-2005 "ketjuttamalla".

EQ(1) Modelling DLQt by RLS (using CONS.IN7)

The present sample is: 1961 to 2005

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	0.0024893	0.0059656	0.417	0.6786	0.0041
DLYt	0.74162	0.097808	7.582	0.0000	0.5779
LSt_1	-0.25963	0.12930	-2.008	0.0511	0.0876

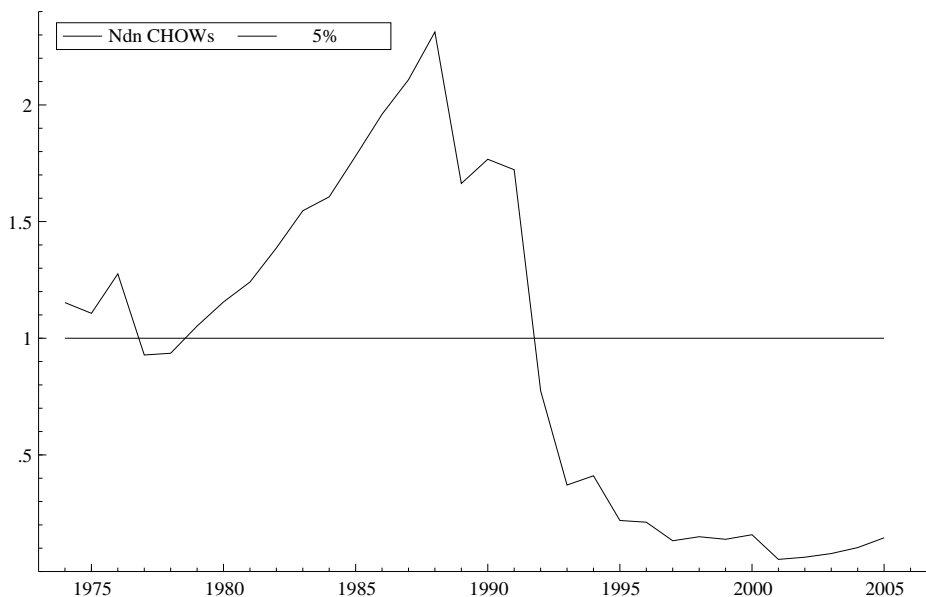
R² = 0.586449 F(2,42) = 29.78 [0.0000] \sigma = 0.0188946 DW = 1.62
 RSS = 0.01499422663 for 3 variables and 45 observations

StdInn saved to CONS.IN7

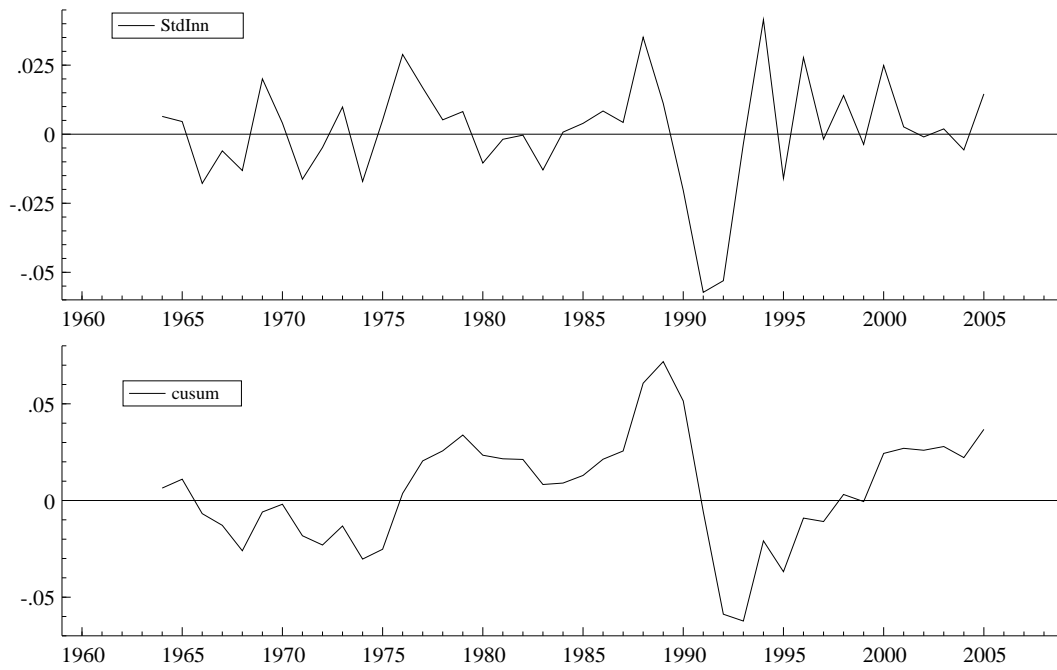
Algebra code for CONS.IN7:

```
cusum = cum(StdInn);
sqr=sqrt(16);
aika=vuosi-1989;
wt0=0.0714932;
raja1=wt0-0.948*0.0188946*(sqr+2*aika/sqr);
raja2=wt0+0.948*0.0188946*(sqr+2*aika/sqr);
```

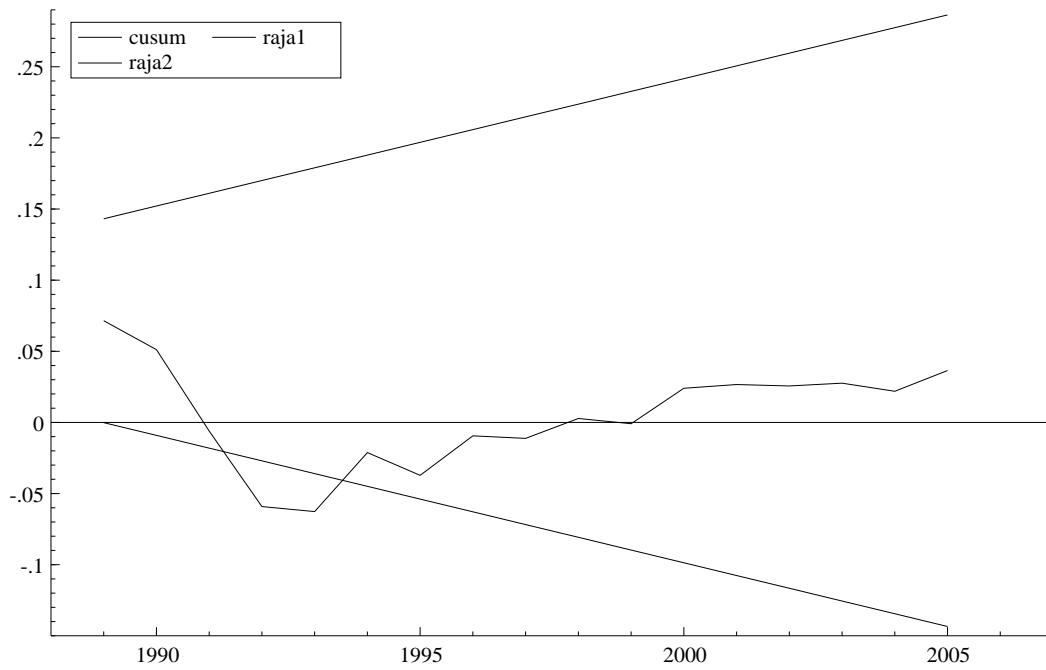
Askeltava Chow-testi (Tässä muodossa testin käyttöä ei suositella!)



Rekursiiviset residuaalit ja niiden kumulatiivinen summa



CUSUM-testi



Jos muuttujat x_1, \dots, x_n ovat eksponentiaalisesti jakautuneita *i.i.d.*-muuttujia, niin muuttujat

$$z_{(t)} = \frac{x_1 + \dots + x_t}{x_1 + \dots + x_n} \quad t = 1, \dots, n-1 \quad (3.54)$$

ovat yhteisjakaumaltaan *täsmälleen* $\text{tas}(0,1)$ - jakaumasta saadun, vastaavan kokoisen yksinkertaisen otoksen kaltaisia.

Kolmogorov-Smirnov-testin (ks. huomautus 3.9) yhteydessä tehtyjä prosenttipistelaskelmia voidaan siis hyödyntää myös muotoa (3.54) olevia $z_{(t)}$ -osamääriä tutkittaessa.

Todettakoon, että

$$\frac{1}{\sigma^2}(e_t^{*2} + e_{t+1}^{*2}) \sim \chi_2^2 = \text{eksp}\left(\frac{1}{2}\right) .$$

Tätä tulosta (ja Kolmogorov-Smirnov- taulukoita) voidaan hyödyntää tarkkailtaessa kumulatiivisia neliösummia

$$q_{(t)} = \frac{Q_{(t)}}{Q_{(n)}} = \frac{\sum_{\nu=m+1}^t e_{\nu}^{*2}}{(n-m)s^2} \quad t = m+1, \dots, n \quad . \quad (3.55)$$

Mikäli

$$\left| q_{(t)} - \frac{t-m}{n-m} \right| > c_{\alpha} \quad (3.56)$$

jollakin $t = m+1, \dots, n-1$, ”hylätään” vakiorakennehypoteesi. (Tässä c_{α} tarkoittaa merkitsevyytystasoon $1-\alpha$ liittyvää, havaintomäärän $\lfloor \frac{n-m}{2} \rfloor + 1$ mukaista Kolmogorov-Smirnov-testisuureen kriittistä arvoa.)

Saatua testiä (3.55) + (3.56) sanotaan CUSUMQ-testiksi, ja se soveltuu mainiosti myös virhetermien variansseissa tapahtuvien muutosten paljastamiseen.

Mikäli CUSUMQ-testillä halutaan testata nimenomaan β -regressioker-toimien muuttumista, on huomattava, että tällöin

$$Ee_t^{*2} > \sigma^2 \quad \text{kun} \quad t > t_o \quad ,$$

jossa t_o tarkoittaa rakennemuutoksen ajankohtaa. Tästä syystä testiä (3.56) on syytä käyttää *yksipuolisena* ja valita kriittiseksi alueeksi

$$q(t) < \frac{t-m}{n-m} - c_\alpha \quad . \quad (3.57)$$

Huomautus 3.12: PCGIVE kannustaa käyttäjiään seuraamaan Chow-testisuureen $C_t^{(n)}$ käyttäytymistä *askeltavasti* käymällä läpi kaikki mahdolliset rakennemuutosajankohdat $t = m+1, \dots, n-1$. Koska Chow-testisuure $C_t^{(n)}$ voidaan kirjoittaa muotoon (3.49')

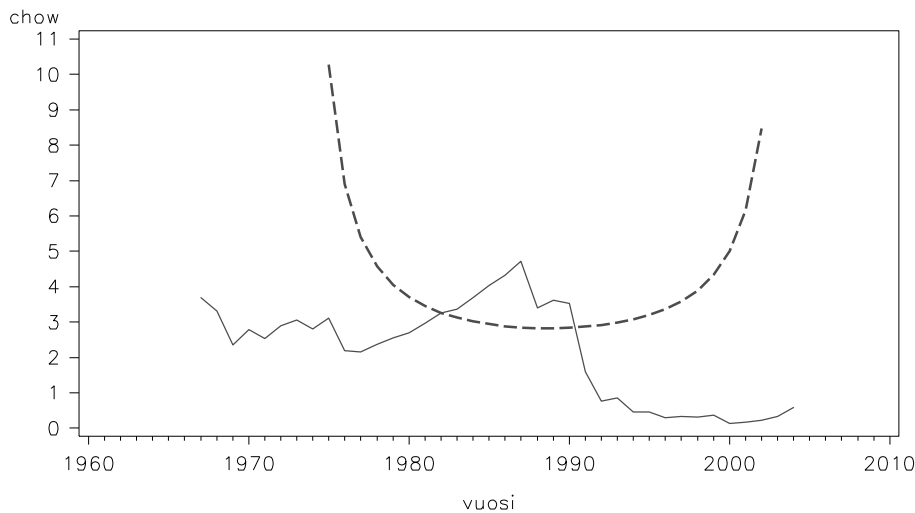
$$C_{t_o}^{(n)} = \frac{\frac{1}{n-t_o} \sum_{\nu=t_o+1}^n e_\nu^{*2}}{\frac{1}{t_o-m} \sum_{\nu=m+1}^{t_o} e_\nu^{*2}} \quad t = m+1, \dots, n-1 \quad ,$$

on helppo havaita, että $C_t^{(n)}$:n ja CUSUMQ-testisuureen $q(t)$ välillä vallitsee seuraava, kääntäen yksikäsitteinen yhteys:

$$C_t^{(n)} = \frac{t-m}{n-t} \cdot \left(\frac{1}{q(t)} - 1 \right) \quad t = m+1, \dots, n-1 \quad . \quad (3.58)$$

CUSUMQ-testin kriittinen alue (3.57) voidaan siis kirjoittaa muotoon

$$C_t^{(n)} > \frac{1 + c_\alpha - \frac{t-m}{n-m}}{\frac{t-m}{n-m} - c_\alpha} \cdot \frac{t-m}{n-t} \quad \text{jollakin} \quad t > m + c_\alpha(n-m) \quad . \quad (3.57')$$



Testi (3.57') on siis periaatteessa ekvivalentti vastaavan CUSUMQ-testin (3.57) kanssa. PCGIVE kuitenkin piirtää $C_t^{(n)}$:n aikaurakuviota *kiinteään* ”kriittisen rajan” (ks. kuvio sivulla 55). Tämä *ei* ole testin sekventiaalisuuden vuoksi korrektia, vaan korrektimpi kriittinen alue saadaan kaavasta (3.57') (ks. yllä oleva kuvio). ■

4 DYNAAMISET REGRESSIOMALLIT

4.1 Siirtofunktiomallit

4.1.1 Adaptiivisten odotusten mallit

Ajatellaan, että vastemuuttuja y_t riippuu tekijän x_t tulevaa kehitystä koskevista *odotuksista* x_{t+1}^* mallin

$$y_t = \beta_1 + \beta_2 x_{t+1}^* + \varepsilon_t \quad , \quad \varepsilon_t \perp\!\!\!\perp x_{t+1}^* \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad (4.1)$$

mukaisesti. Koska odotuksia x_{t+1}^* ei yleensä voida havaita, oletetaan niiden kehittyvän *adaptiivisesti* kaavan

$$x_{t+1}^* - x_t^* = \alpha(x_t - x_t^*) \quad (4.2)$$

mukaisesti. ($0 < \alpha < 1$). Tällöin siis

$$\begin{aligned} x_{t+1}^* &= [1 - (1 - \alpha)L]^{-1} \alpha x_t \\ &= \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j L^j x_t \\ &= \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j x_{t-j} \quad , \end{aligned} \quad (4.3)$$

jossa L tarkoittaa viiveoperaattoria $Lx_t = x_{t-1}$. Sijoittamalla tulos (4.3) malliin (4.1) saadaan

$$y_t = \beta_1 + \beta_2 \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j x_{t-j} + \varepsilon_t \quad ,$$

joten ajatus (4.2) implikoi x_t :n ja y_t :n välille ns. siirtofunktiomallin

$$\begin{aligned} y_t &= \mu + v(L)x_t + \varepsilon_t \quad , \quad \text{jossa} \\ \{\varepsilon_t\} &\perp\!\!\!\perp \{x_t\} \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad , \\ v(L) &= \sum_{j=0}^{\infty} v_j L^j = \alpha \beta_2 [1 - (1 - \alpha)L]^{-1} \quad . \end{aligned} \quad (4.4)$$

Mallissa (4.4) joudutaan siis varautumaan x_t :n *viivästettyihin*, hyvinkin hitaasiin vaikutuksiin. Lisäksi useimmiten joudutaan realismuksen nimissä sallimaan jonkinasteinen ajallinen säännöllisyys (autokorrelaatio) virhetermisarjalle ε_t .

4.1.2 Eniten käytetyt siirtofunktioallityypit

Ajatellaan nyt, että selittäviä muuttujia (aikasarjoja) x_{it} olisi useampia ja että niillä kaikilla saattaisi olla myös viivästeisiä vaikutuksia y_t :hen. Yleensä on syytä alusta asti varautua virhetermisarjan ε_t autokorrelaatioon esim. stationäärisen ARMA-mallin

$$\varepsilon_t = \frac{\theta(L)}{\phi(L)} \alpha_t \quad , \quad \{\varepsilon_t\} \text{ stationäärinen} \quad , \quad (4.5)$$

$$\alpha_t \sim NID(0, \sigma^2)$$

puitteissa. Lausekkeet $\theta(L)$ ja $\phi(L)$ tarkoittavat viiveoperaattorin polynomeja,

$$\begin{cases} \theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q & , \\ \phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p & . \end{cases} \quad (4.6)$$

Tällöin x_{it} -muuttujien *lineaarisia* vaikutuksia y_t -muuttujaan kuvaava, yleisin kuviteltavissa oleva malli olisi muotoa

$$y_t = \mu + v_1(L) x_{1t} + \dots + v_m(L) x_{mt} + \varepsilon_t \quad \text{jossa}$$

$$\{\varepsilon_t\} \perp\!\!\!\perp \{x_{1t}, \dots, x_{mt}\} \quad , \quad \{\varepsilon_t\} \text{ stationäärinen} \quad ,$$

$$\varepsilon_t = \frac{\theta(L)}{\phi(L)} \alpha_t \quad , \quad \alpha_t \sim NID(0, \sigma^2) \quad \text{ja} \quad (4.7)$$

$$v_i(L) = \sum_{j=0}^{\infty} v_{ij} L^j \quad i = 1, \dots, m \quad .$$

Mallin mielekkyyden kannalta kannattaa erityisesti korostaa virhetermisarjan $\{\varepsilon_t\}$ riippumattomuutta $\{x_{1t}, \dots, x_{mt}\}$ -sarjoista, ts. selittäjäsarjojen *eksogeenisuutta*.

Jotta mallin parametrimäärä pysyisi äärellisenä, on *siirtofunktioille* $v_i(L)$ asetettava joitakin rajoituksia. Tavallisimpia ovat rajoitukset

$$v_i(L) = \omega_{i0} - \omega_{i1} L - \dots - \omega_{ir_i} L^{r_i} \quad (4.8)$$

tai

$$v_i(L) = \frac{\omega_i(L)}{\delta_i(L)} \quad , \quad \text{jossa}$$

$$\omega_i(L) = \omega_{i0} - \omega_{i1}L - \dots - \omega_{ir_i}L^{r_i} \quad , \quad \text{ja} \quad (4.9)$$

$$\delta_i(L) = 1 - \delta_{i1}L - \dots - \delta_{is_i}L^{s_i}$$

tai

$$v_i(L) = \sum_{j=0}^{r_i} v_{ij}L^j \quad , \quad \text{jossa} \quad (4.10)$$

$$v_{ij} = \beta_{i0} + \beta_{i1}j + \dots + \beta_{id_i}j^{d_i} \quad j = 0, \dots, r_i$$

Muodot (4.8) ja (4.10) edellyttävät siis, että selittäjän x_{it} vaikutusaika y_t :hen on rajoitettu siten, että kaikki vaikutukset ehtivät ilmetä r_i aikayksikön kuluessa. Muoto (4.9) taas sopii pitkäaikaisten vaikutusten kuvaamiseen, koska v_{ij} -kertoimet eroavat nolasta myös suurilla j :n arvoilla.

Muotoa (4.9) sanotaan *rationaaliseksi* ja muotoa (4.8) *polynomimuotoiseksi* siirtofunktioksi. Muotoa (4.10) sanotaan *Almon-malliksi* ja sen puitteissa voidaan usein selvittää suhteellisen harvoilla parametreilla, vaikka maksimaalinen vaikutusaika r_i olisikin melko pitkä.

Ihanteellista tietysti olisi, mikäli kulloinkin käytettävän mallin (4.7) muoto voitaisiin valita talousteorian perusteella, mutta useimmiten muodon valinta joudutaan tekemään yksinomaan käytettävissä olevien havaintojen perusteella.

Määritelmä 4.1: Kertoimia v_{ij} ($j = 0, 1, 2, \dots$) sanotaan selittäjään x_{it} liittyviksi *impulssivasteiksi* ja v_{i0} :aa *välittömäksi vasteeksi*.

Kumulatiivisia summia

$$c_{ij} = \sum_{k=0}^j v_{ik} \quad j = 0, 1, 2, \dots$$

sanotaan *vastekertymiksi* (interim multipliers) ja summaa $c_{i\infty}$ selittäjään x_{it} liittyväksi *kokonaisvasteeksi* (total multiplier).

Mikäli $v_{ij} \geq 0$ kaikilla $j = 0, 1, 2, \dots$, voidaan lukuja $v_{ij}/c_{i\infty}$ ajatella diskreettiin,

kokonaislukuarvoiseen todennäköisyysjakaumaan liittyvinä pistetodennäköisyysyksinä. Tällöin puhutaan usein viivejakaumamalleista (distributed lag models) ja voidaan määritellä po. ”todennäköisyysjakaumaan” liittyviä tunnuslukuja, kuten esim. ”keskimääräinen vaikutusaika” (mean lag)

$$\frac{1}{c_{i\infty}} \sum_{j=1}^{\infty} j v_{ij} \quad .$$

■

Määriteltyjen termien ja sanontojen tausta on seuraava. Tarkastellaan y_t :n odotusarvouran

$$\bar{y}_t = \mu + v_1(L)x_{1t} + \dots + v_m(L)x_{mt}$$

käyttäytymistä, kun selittävät tekijät vakioituvat tasoille $\bar{x}_1, \dots, \bar{x}_m$. Tällöin \bar{y}_t asettuisi ”tasapainotilaan”

$$\bar{y}_t = \mu + v_1(1)\bar{x}_1 + \dots + v_m(1)\bar{x}_m \quad .$$

Jos \bar{x}_i :lle annettaisiin yksikön suuruinen lisäys, kertoisivat impulssivasteet v_{ij} ja vastekertymät c_{ij} , miten \bar{y}_t asettuisi uuteen tasapainotilaan. Kokonaisvaste $c_{i\infty}$ kertoisi tasapainotilan kokonaissiirtymän.

Esimerkiksi adaptiivisten odotusten mallissa (4.1) + (4.2) vastaisivat impulssivasteet

$$v_j = \alpha\beta_2(1 - \alpha)^j \quad j = 0, 1, 2, \dots$$

ns. *geometriska* jakaumaa, joten keskimääräinen vaikutusaika olisi

$$\alpha \sum_{j=1}^{\infty} j (1 - \alpha)^j = \frac{1 - \alpha}{\alpha} \quad .$$

Vastekertymät olisivat muotoa

$$c_j = \sum_{\nu=0}^j \alpha\beta_2(1 - \alpha)^\nu = \beta_2 [1 - (1 - \alpha)^{j+1}] \quad j = 0, 1, 2, \dots$$

ja kokonaisvaste $c_{i\infty} = \beta_2$.

4.2 Differenssiyhtälön muotoiset mallit

4.2.1 Osittaisen sopeutumisen mallit

Ajatellaan, että tekijät X_t määräisivät muuttujan y_t tavoitetason y_t^* yhtälön

$$y_t^* = \beta' X_t \quad t = 1, \dots, n$$

mukaisesti, mutta että y_t :n kehityksessä olisi tiettyä ”hitausmomenttia” sopeutumisyhtälön

$$y_t - y_{t-1} = \alpha(y_t^* - y_{t-1}) + \varepsilon_t \quad \varepsilon_t \perp\!\!\!\perp X_t$$

mukaisesti ($0 < \alpha < 1$). Tällöin päädytään havaittavissa olevien suureiden osalta malliin

$$\begin{aligned} y_t &= (1 - \alpha)y_{t-1} + \alpha \beta' X_t + \varepsilon_t \quad , \\ \varepsilon_t &\sim NID(0, \sigma^2) \quad , \quad \varepsilon_t \perp\!\!\!\perp X_t \quad , \quad t = 2, \dots, n \end{aligned} \quad (4.15)$$

Mallia (4.15) sanotaan *osittaisen sopeutumisen* malliksi ja se edustaa differenssiyhtälön muotoista mallia yksinkertaisimmillaan. Jos y_t ”ratkaistaan” mallista (4.15), saadaan

$$y_t = \frac{\alpha}{1 - (1 - \alpha)L} \beta' X_t + \frac{1}{1 - (1 - \alpha)L} \varepsilon_t \quad , \quad (4.16)$$

joten *jokaisen* selittäjän impulssivasteet olisivat muotoa

$$v_{ij} = \alpha \beta_i (1 - \alpha)^j \quad j = 0, 1, 2, \dots$$

”keskimääräinen vaikutusaika”

$$\frac{\sum_{j=0}^{\infty} v_{ij} j}{\sum_{j=0}^{\infty} v_{ij}} = \frac{1 - \alpha}{\alpha}$$

olisi sama kaikille muuttujille x_{it} ($i = 1, \dots, m$). Tästä syystä sitä kutsutaankin yleensä keskimääräiseksi *sopeutumisajaksi*.

4.2.2 ARX- ja ARMAX-mallit

Mallin (4.15) sisältö voidaan tulkita myös siten, että y_t pyrkii kehittymään differenssiyhtälömallin

$$y_t = \mu + (1 - \alpha)y_{t-1} + \varepsilon_t \quad (4.17)$$

mukaisesti, mutta "eksogeeniset" tekijät X_t vaikuttavat lineaarisesti siihen tasoon $\frac{\mu}{\alpha}$, jonka lähistölle y_t olisi oman kehitysmekanisminsa (4.17) mukaisesti asettumassa.

Jos sama ajatus yleistetään suoraviivaisesti yhtälöä (4.17) monimutkaisempiin kehitysrakenteisiin (esim. ARMA -rakenteisiin), päädytään luontevasti mallityyppiin

$$\begin{aligned} \alpha(L) y_t &= \mu + \sum_{i=1}^m \omega_i(L) x_{it} + \theta(L) \varepsilon_t \quad , \\ \{\varepsilon_t\} &\perp\!\!\!\perp \{x_{1t}, \dots, x_{mt}\} \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad , \\ \alpha(L) &= 1 - \alpha_1 L - \dots - \alpha_p L^p \quad , \\ \theta(L) &= 1 - \theta_1 L - \dots - \theta_q L^q \quad , \\ \omega_i(L) &= \omega_{i0} - \omega_{i1} L - \dots - \omega_{ir_i} L^{r_i} \quad , \quad i = 1, \dots, m \quad . \end{aligned} \quad (4.18)$$

Tällaisia malleja kutsutaan ARMAX-malleiksi ja niitä käytetään usein pelkästään siksi, että niiden parametrien estimointi on teknisesti helpompaa kuin siirtofunktio mallien (4.7). (Malli (4.7) sisältää toisaalta erikoistapauksinaan kaikki ARMAX-mallit (4.18).)

Parametrien estimointi on vieläkin helpompaa, mikäli $\theta(L) \equiv 1$ (ts. mikäli $q = 0$), sillä mallin

$$\begin{aligned} \alpha(L) y_t &= \mu + \sum_{i=1}^m \omega_i(L) x_{it} + \varepsilon_t \quad , \\ \{\varepsilon_t\} &\perp\!\!\!\perp \{x_{1t}, \dots, x_{mt}\} \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \end{aligned} \quad (4.19)$$

parametreja voidaan hyvällä omallatunnolla estimoida *OLS-tekniikalla*. Malleja (4.19) sanotaan ARX-malleiksi. Koska ε_t on oletettu kaikista selittäjistä (viivästetyt y :n arvot mukaan luettuina) riippumattomaksi, on siis OLS-estimaattoreiden käyttö täysin paikallaan ja estimaatit voidaan saman tien laskea

jopa *askeltavasti* rekursiokaavoilla (3.39). (Tämä on eräs syy ARX-mallien suureen suosioon.) Toisaalta, myös ARMAX-mallien (4.18) parametreja voidaan estimoida rekursiivisesti, mutta tarvittava algoritmi on paljon monimutkaisempi kuin (3.39).

Huomautus 4.1: ARX-malli (4.19) sisältää pohjimmiltaan seuraavan, y_t :n kehitysmekanismia koskevan oletuksen:

Muuttujan y_t kehityksen määrää primääristi sen oma, AR-mallilla kuvattavissa oleva etenemismekanismi, joka voi olla osa suuremmasta, useamman muuttujan muodostamasta vuorovaikutussysteemistä. ”Ulkoiset” (eksogeeniset) tekijät $x_t = (x_{1t} \dots x_{mt})'$ voivat vaikuttaa *lineaarisesti* systeemin ”tasapainotilan” sijaintiin, mutta *eivät* systeemin rakenteeseen.

Tämä ajattelutapa voi joissakin sovelluksissa (esim. osittaisen sopeutumisen mallit) tuntua täysin realistiselta, mutta *ei* varmastikaan kaikissa. Mallista (4.19) nimittäin seuraa, että

$$y_t = \frac{\mu}{\alpha(1)} + \sum_{i=1}^m \frac{\omega_i(L)}{\alpha(L)} x_{it} + \frac{1}{\alpha(L)} \varepsilon_t \quad .$$

Jokaisen selittäjän vaikutustavan y_t :hen pitäisi siis suurilla viiveillä olla *samanlainen*. Tämä ajatus ei kaikissa tapauksissa tunnu realistiselta. ■

4.3 Virheenkorjausmallit

Tarkastellaan aluksi dynaamista mallia

$$\begin{aligned} \log y_t &= \alpha \log y_{t-1} + \mu + \omega_o \log x_t - \omega_1 \log x_{t-1} + \varepsilon_t \quad , \\ \varepsilon_t &\perp\!\!\!\perp x_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \end{aligned} \quad (4.20)$$

Jos x_t stabiloituu tasolle \bar{x} , stabiloituu $\log y_t$:n odotusarvo vastaavasti tasolle

$$\log \bar{y} = \frac{\mu}{1 - \alpha} + \frac{\omega_o - \omega_1}{1 - \alpha} \log \bar{x} \quad (4.21)$$

Tällaista $(\bar{x} \ \bar{y})'$ -tilaa sanotaan mallin (4.20) *tasapainotilaksi*. Eri \bar{x} :n arvoihin liittyvistä tasapainotiloista muodostuu ns. *tasapainoura*, jota voitaisiin kutsua myös $(x_t \ y_t)'$ -muuttujan *attraktoriksi*.

Merkitään differenssioperaattoria symbolilla $\nabla = 1 - L$. Tällöin malli (4.20) voidaan kirjoittaa muotoon

$$\begin{aligned} \nabla \log y_t &= (\alpha - 1) \log y_{t-1} + \mu + \omega_o \nabla \log x_t + (\omega_o - \omega_1) \log x_{t-1} + \varepsilon_t \\ &= \omega_o \nabla \log x_t + \varepsilon_t \\ &\quad + (\alpha - 1) \left[\log y_{t-1} - \frac{\mu}{1 - \alpha} - \frac{\omega_o - \omega_1}{1 - \alpha} \log x_{t-1} \right] \end{aligned} \tag{4.22}$$

Yhtälön (4.22) oikean puolen hakasulkulauseke siis kertoo, kuinka kaukana attraktorista (4.21) muuttuja $(x_t \ y_t)'$ on hetkellä $t - 1$.

Jos muuttuja $(x_t \ y_t)'$ on lähellä attraktoria (tasapainouraa) koko havaintoperiodin ajan, vaihtelee hakasulkulauseke hyvin vähän, ja α :n estimointi saattaa vaikeutua. Jos taas tasapainouralta (4.21) poiketaan paljon, vaikuttaa (4.22):n hakasulkulauseke *tasapainotilaa kohti vetävänä tekijänä*, joka pyrkii nopeuttamaan muuttujan paluuta tasapainouran läheisyyteen.

Muutosten $\nabla \log x_t$ ja $\nabla \log y_t$ välinen relaatio mallissa (4.22) ohjaa siis muuttujan $(x_t \ y_t)'$ *lyhyen tähtäimen* dynamiikkaa, kun taas hakasulkulausekkeen avulla saadaan käsitys *pitkän tähtäimen* tasapainouran sijainnista.

Muotoa (4.22) olevia malleja sanotaan *virheenkorjausmalleiksi* (error correction models). Myöhemmin esiteltävän yhteisintegroituvuuden käsitteen ja virheenkorjausmallien välillä vallitsee läheinen yhteys: Jos $\log x_t$ ja $\log y_t$ ovat yhteisintegroituneita, voidaan niiden välille aina muodostaa mallin (4.22) kaltainen virheenkorjausmalli. Virheenkorjausmalli edustaa samalla yhteisintegroituvuusrelaation tulkinnallisesti hedelmällisintä esitystapaa.

Huomautettakoon samalla, että tasapainorelaatioon (attraktoriin) saattavat vaikuttaa sellaisetkin muuttujat, joita ei lyhyen tähtäimen dynamiikan kuvauksessa lainkaan tarvita.

Esimerkki 4.1: Davidson, Hendry, Srba ja Yeo käyttivät 1978 ilmestyneessä artikkelissaan seuraavan tyyppistä mallia (kyseessä oli todennäköisesti ensimmäinen julkaistu virheenkorjausmallin sovellus) kuvaamaan kotitalouksien reaalityulojen y_t ja yksityisen kulutuksen volyymin c_t välistä yhteyttä:

$$\nabla_4 \log c_t = \beta_1 \nabla_4 \log y_t + \beta_2 \nabla \nabla_4 \log y_t + \varepsilon_t + \beta_3 \log \left(\frac{c_{t-4}}{y_{t-4}} \right), \tag{4.23}$$

jossa $\varepsilon_t \sim NID(0, \sigma^2)$, $\nabla_4 = 1 - L^4$ ja $\nabla = 1 - L$.

Sovellus koski Englannin neljännesvuosiaineistoa. ”Tasapainoura” vastaisi siis vakiona pysyvää säästämisastetta. Toisaalta ”lyhyen tähtäimen rajakulutusalittua” vastaisi mallissa (4.23) parametri $(\beta_1 + \beta_2) - \beta_2 = \beta_1$. ■

5 IVE-, GIVE- ja GMM-ESTIMOINTI

5.1 Epäsuorasti havaitut selittävät muuttujat (proxy-muuttujat)

Niitä muuttujia, joilla talousteoria operoi, ei useinkaan pystytä suoraan (virheettömästi) havaitsemaan. Esim. odotusten havaitseminen on yleensä mahdotonta, samoin ”pysyväistulojen”, jne. Joskus taas tietojen saatavuus voi tuottaa ongelmia.

Ajatellaan nyt, että regressiomallissa

$$y_t = \beta_1 + \beta_2 x_t^* + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad (5.1)$$

olisi vain yksi selittäjä, mutta että sen arvoihin liittyisi havaintovirhettä mallin

$$x_t = x_t^* + \eta_t \quad , \quad \eta_t \sim NID(0, \tau^2) \quad , \quad \eta_t \perp\!\!\!\perp \varepsilon_t \quad (5.2)$$

mukaisesti, jossa x_t edustaa selittävän muuttujan *havaittua* arvoa havaintoyksikössä t ($t = 1, \dots, n$).

Yhdistämällä mallit (5.1) ja (5.2) saadaan

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t + (\varepsilon_t - \beta_2 \eta_t) \\ &= \beta_1 + \beta_2 x_t + \kappa_t \quad , \end{aligned} \quad (5.3)$$

jossa siis $x_t \not\perp \kappa_t$.

Välittömänä seurauksena johdannossa esitetyn perusoletuksen (1.8) rikkoutumisesta on, että OLS-estimaattorit ovat *epäkonsistentteja*. (Ajatellaan yksinkertaisuuden vuoksi, että $\beta_1 = 0$.) Tällöin nimittäin

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} = \frac{\frac{1}{n} \sum_{t=1}^n (\beta_2 x_t^* + \varepsilon_t)(x_t^* + \eta_t)}{\frac{1}{n} \sum_{t=1}^n (x_t^* + \eta_t)^2} \\ &\xrightarrow{p} \frac{\beta_2 \frac{1}{n} \sum_{t=1}^n x_t^{*2}}{\frac{1}{n} \sum_{t=1}^n x_t^{*2} + \tau^2} \neq \beta_2 \quad . \end{aligned} \quad (5.4)$$

Tämä epäkonsistenssi on samalla oire vakavammastakin estimoituvuusongelmasta, sillä mallin (5.1) + (5.2) mukaisen likelihood-funktion logaritmi on muotoa

$$\begin{aligned} &\log L_{Y,X}(X^*, \beta_1, \beta_2, \tau^2, \sigma^2) \\ &\simeq \sum_{t=1}^n \left[-\frac{1}{2} \log \tau^2 - \frac{1}{2} \log \sigma^2 - \frac{1}{2\tau^2} (x_t - x_t^*)^2 - \frac{1}{2\sigma^2} (y_t - \beta_1 - \beta_2 x_t^*)^2 \right] \quad . \end{aligned} \quad (5.5)$$

Jaetaan likelihood-funktio (5.5) kahteen osaan,

$$\log L_{Y,X}(X^*, \beta_1, \beta_2, \tau^2, \sigma^2) = l_{Y,X}^{(1)}(X^*, \beta_1, \beta_2, \tau^2, \sigma^2) + l_{Y,X}^{(2)}(\beta_1, \beta_2, \tau^2, \sigma^2) \quad , \quad (5.6)$$

jossa

$$\begin{aligned}
 l_{Y,X}^{(1)}(X^*, \beta_1, \beta_2, \tau^2, \sigma^2) \\
 = -\frac{1}{2} \left(\frac{1}{\tau^2} + \frac{\beta_2^2}{\sigma^2} \right) \sum_{t=1}^n \left(x_t^* - \frac{\sigma^2 x_t + \beta_2 \tau^2 (y_t - \beta_1)}{\sigma^2 + \beta_2^2 \tau^2} \right)^2
 \end{aligned} \tag{5.7}$$

ja

$$\begin{aligned}
 l_{Y,X}^{(2)}(\beta_1, \beta_2, \tau^2, \sigma^2) \\
 = -\frac{n}{2} (\log \tau^2 + \log \sigma^2) - \frac{1}{2} \frac{1}{\sigma^2 + \beta_2^2 \tau^2} \sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)^2
 \end{aligned} \tag{5.8}$$

On helppo nähdä, että $l_{Y,X}^{(1)}$ kaavassa (5.7) ei sisällä mitään sellaista informaatiota parametreista β_1, β_2, τ^2 tai σ^2 , joka olisi tulkittavissa ilman tietoa $X^* = (x_1^* \dots x_n^*)'$:stä.

Jos taas β_1 ja β_2 ajateltaisiin tunnetuiksi, riippuisi $l_{Y,X}^{(2)}$ kaavassa (5.8) havainnoista vain suureen $\sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)^2$ kautta. Toisaalta, auki olevia parametreja olisi vielä kaksi (τ^2 ja σ^2), jotka siis eivät voi olla identifioitavissa havaintojen perusteella. Ongelma pysyy tietysti samanlaisena, vaikka β -parametreja ei oletteisikaan tunnetuiksi. Malliin (5.1) + (5.2) on siis *lisättävä jotakin informaatiota*, jotta parametrien estimointi onnistuisi. Periaatteessa on olemassa kolme mahdollisuutta:

- 1° Oletetaan $\frac{\tau^2}{\sigma^2}$ tunnetuksi.
- 2° Postuloidaan parametreille priorijakauma ja turvaudutaan bayesiläiseen päättelyyn.
- 3° Oletetaan, että voidaan havaita kolmattakin, samoja x_t^* -arvoja heijastelevaa muuttujaa

$$z_t = \alpha_1 + \alpha_2 x_t^* + \omega_t \quad , \tag{5.9}$$

jossa $\omega_t \perp\!\!\!\perp (\eta_t \ \varepsilon_t)'$ ja $\omega_t \sim NID(0, \delta^2) \quad t = 1, \dots, n$.

Tarkastellaan nyt erityisesti lisäinformaatiotyyppiä 3°. Tällöin havaintoaineiston Y, X, Z määräämän likelihood- funktion logaritmi voidaan aikaisempaan tapaan hajottaa kahden termin summaksi

$$\log L_{Y,X,Z}(X^*, \alpha, \beta, \tau^2, \sigma^2, \delta^2) = l_{Y,X,Z}^{(1)}(X^*, \alpha, \beta, \tau^2, \sigma^2, \delta^2) + l_{Y,X,Z}^{(2)}(\alpha, \beta, \tau^2, \sigma^2, \delta^2) \quad , \tag{5.10}$$

jossa

$$\begin{aligned}
& l_{Y,X,Z}^{(1)}(X^*, \alpha, \beta, \tau^2, \sigma^2, \delta^2) \\
&= -\frac{1}{2} \left(\frac{1}{\tau^2} + \frac{\beta_2^2}{\sigma^2} + \frac{\alpha_2^2}{\delta^2} \right) \sum_{t=1}^n \left[x_t^* - \frac{\frac{x_t}{\tau^2} + \beta_2 \frac{(y_t - \beta_1)}{\sigma^2} + \alpha_2 \frac{(z_t - \alpha_1)}{\delta^2}}{\frac{1}{\tau^2} + \frac{\beta_2^2}{\sigma^2} + \frac{\alpha_2^2}{\delta^2}} \right]^2
\end{aligned} \tag{5.11}$$

ja

$$\begin{aligned}
& l_{Y,X,Z}^{(2)}(\alpha, \beta, \tau^2, \sigma^2, \delta^2) = -\frac{n}{2} (\log \tau^2 + \log \sigma^2 + \log \delta^2) \\
& - \frac{1}{2} \frac{1}{\frac{1}{\tau^2} + \frac{\beta_2^2}{\sigma^2} + \frac{\alpha_2^2}{\delta^2}} \left\{ \frac{1}{\sigma^2} \left(\frac{1}{\tau^2} + \frac{\alpha_2^2}{\delta^2} \right) \sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)^2 \right. \\
& \quad + \frac{1}{\delta^2} \left(\frac{1}{\tau^2} + \frac{\beta_2^2}{\sigma^2} \right) \sum_{t=1}^n (z_t - \alpha_1 - \alpha_2 x_t)^2 \\
& \quad \left. - 2 \frac{\beta_2 \alpha_2}{\sigma^2 \delta^2} \sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)(z_t - \alpha_1 - \alpha_2 x_t) \right\}
\end{aligned} \tag{5.12}$$

Jos jälleen ajatellaan α - ja β -parametrit tunnetuiksi, riippuu $l_{Y,X,Z}^{(2)}$ havainnoista *kolmen* tunnusluvun kautta. Koska varianssiparametreja on niinkään kolme, voidaan ne kaikki estimoida, ja aikaisempi identifioituvuusmurhe on siis poistunut.

5.2 Instrumenttimuuttujiin perustuva estimointi

Edellä todettiin, että malliin (5.1) + (5.2) liittyvät estimointiongelmat johtuivat pohjimmiltaan siitä, että muodossa (5.3) selittävät tekijät ja virhetermit korreloivat keskenään. Tarkastellaan nyt yleisemminkin johdannossa esiteltyä perusmuotoista regressiomallia (1.6) + (1.7) + (1.9), jossa kuitenkin ehto (1.8) on rikottu, ts.

$$E\varepsilon_t X_t \neq 0 \quad .$$

Tällöinhän

$$p \lim_{n \rightarrow \infty} \frac{1}{n} X' \varepsilon \neq 0 \quad ,$$

joten

$$\begin{aligned}
 \text{p} \lim_{n \rightarrow \infty} \hat{\beta} &= \text{p} \lim_{n \rightarrow \infty} \left(\frac{1}{n} X'X \right)^{-1} \frac{1}{n} X'(X\beta + \varepsilon) \\
 &= \beta + \text{p} \lim_{n \rightarrow \infty} \left(\frac{1}{n} X'X \right)^{-1} \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} X'\varepsilon \\
 &\neq \beta \quad .
 \end{aligned} \tag{5.13}$$

Tilanteisiin, joissa oletus (1.8) on rikottu, päädytään *proxy*-muuttujien käytön lisäksi myös eräiden *aikasarjamallien* sekä ns. *moniyhtälömallien* (ks. luku 6) yhteydessä.

Kuten luvussa 5.1 todettiin, on β :n estimoinnin kannalta hyödyllistä, jos käytettävissä on havainnot muuttujista z_1, \dots, z_m , jotka *eivät* korreloi virhetermin ε kanssa, mutta ”heijastelevat samoja asioita kuin selittäjät x_1, \dots, x_m ”. Jos z -muuttujista saatua havaintomatriisia ($n \times m$ -matriisi) merkitään symbolilla Z , voidaan em. toivomukset kirjoittaa muotoon

$$\begin{aligned}
 \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} Z'\varepsilon &= 0 && \text{ja} \\
 \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} Z'X &= Q_{zx} && \text{epäsingulaarinen}
 \end{aligned} \tag{5.14}$$

(Huom.: Osa z_j -muuttujista voi olla *samoja* kuin vastaavat x_j :t.)

Jos ehdot (5.14) ovat voimassa, sanotaan Z :aa *instrumenttimuuttujamatriisiksi*. Lisäksi toivotaan, että

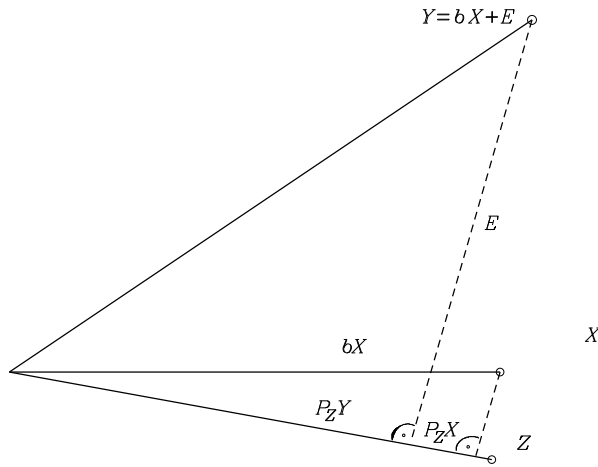
$$\frac{1}{\sqrt{n}} Z'\varepsilon \stackrel{\text{asympt.}}{\sim} N_m(0, \Gamma) \quad . \tag{5.15}$$

(Sivuhuomautus: Jos $\text{cov}(\varepsilon) = \sigma^2 \Omega$, on $\Gamma = \text{p} \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} Z'\Omega Z$.)

Instrumenttimuuttujaestimaattoriksi (IVE) kutsutaan tällöin estimaattoria

$$\hat{\beta}_{IVE} = (Z'X)^{-1} Z'Y \quad . \tag{5.16}$$

Kuva 6.1: Instrumenttiestimoinnin toimintaperiaate yksinkertaistettuna



Nähdään heti, että $\hat{\beta}_{IVE}$ on konsistentti:

$$\begin{aligned} \text{p} \lim_{n \rightarrow \infty} \hat{\beta}_{IVE} &= \text{p} \lim_{n \rightarrow \infty} \left(\frac{1}{n} Z' X \right)^{-1} \text{p} \lim_{n \rightarrow \infty} \left[\frac{1}{n} Z' (X\beta + \varepsilon) \right] \\ &= Q_{zx}^{-1} [Q_{zx} \beta + 0] = \beta \end{aligned} \quad (5.17)$$

Lisäksi oletuksesta (5.15) seuraa, että

$$\begin{aligned} \sqrt{n} (\hat{\beta}_{IVE} - \beta) &= \left(\frac{1}{n} Z' X \right)^{-1} \frac{1}{\sqrt{n}} Z' \varepsilon \\ &\stackrel{\text{asympt.}}{\sim} N_m(0, Q_{zx}^{-1} \Gamma Q_{zx}^{-1} ') \end{aligned} \quad (5.18)$$

Tämän tuloksen mukaisesti voidaan lineaarista hypoteesia $H_o: R\beta = \gamma$ ($\gamma \in \mathcal{R}^r$) testata testisuurella

$$(R\hat{\beta}_{IVE} - \gamma)' \hat{V}^{-1} (R\hat{\beta}_{IVE} - \gamma) \stackrel{\text{asympt.}}{\sim} \chi_r^2 \quad (5.19)$$

jossa

$$\widehat{V} = \widehat{\sigma}^2 R(Z'X)^{-1} Z' \Omega Z (X'Z)^{-1} R' .$$

Jos nyt ajateltaisiin, että β -parametreja olisi vain yksi, olisi $\widehat{\beta}_{IVE}$:n asymptoottinen virhevarianssi muotoa

$$\frac{1}{n} Q_{zx}^{-1} \Gamma Q_{zx}^{-1} ' = \frac{\text{p lim}_{n \rightarrow \infty} \frac{1}{n} Z' Z}{(\text{p lim}_{n \rightarrow \infty} \frac{1}{n} Z' X)^2} ,$$

mikäli $\text{cov}(\varepsilon) = \sigma^2 I$. Tästä nähdään, että Z kannattaa valita (vaatimuksen $Z \perp \varepsilon$ puitteissa) siten, että sen korrelaatio X :n kanssa olisi mahdollisimman suuri. Jos toisaalta Z :n ja X :n välinen korrelaatio on hyvin pieni, muodostuu $\widehat{\beta}_{IVE}$:n virhevarianssi tavattoman suureksi.

Samalla on täysin selvää, että jos on löydettävissä useampia instrumenteiksi kelpaavia muuttujia, kannattaa kaikkien niiden sisältämää lisäinformaatiota tietysti käyttää hyväksi. Sisältäköön Z^* ($n \times k$ -matriisi, $k > m$) nyt kaikki potentiaaliset instrumentit. Oletetaan hetkeksi, että $\text{cov}(\varepsilon) = \sigma^2 I$, ts. $\Omega = I$. Tällöin ilmeisesti ”optimaalinen” (ts. X -muuttujien kanssa maksimaalisesti korreloiva) $n \times m$ -ulotteinen instrumenttimatriisi olisi $Z = P_{Z^*} X = Z^* (Z^{*'} Z^*)^{-1} Z^{*'} X$. Vastaavaa IVE-estimaattoria

$$\begin{aligned} \widehat{\beta}_{GIVE} &= (Z' X)^{-1} Z' Y \\ &= (X' P_{Z^*} X)^{-1} X' P_{Z^*} Y \end{aligned} \tag{5.20}$$

kutsutaa *yleistetyksi* instrumenttiestimaattoriksi (GIVE). myös sille voidaan johtaa (5.18):n kaltainen asymptoottinen jakaumatulos. Regressiokertoimia β koskevia lineaarisia hypoteeseja voidaan tutkia (5.19):n kaltaisilla testisuureilla.

5.3 Yleistetty momenttimenetelmä (GMM)

Tarkastellaan aluksi yksiulotteisesta muuttujasta saatua havaintoaineistoa $Y = (y_1 \dots y_n)'$. Oletetaan, että havaintovektorin Y jakauma sisältyy jakaumaperheeseen $\{F_Y^{(\theta)} \mid \theta \in \Theta\}$, $\Theta \subseteq \mathcal{R}^p$. Lisäksi oletetaan, että prosessi $\{y_t\}$ on *stationäärinen* ja että havaintojen väliset riippuvuudet katoavat ”riittävän nopeasti” havaintojen ajallisen etäisyyden kasvaessa. Varsin usein on löydettävissä jokin kääntäen yksikäsitteinen yhteys y -muuttujan p :n ensimmäisen momentin

$$\alpha_{(p)} = (\alpha_1 \dots \alpha_p)' , \quad \alpha_\nu = E y^\nu \quad \nu = 1, 2, \dots$$

ja parametrin θ välille,

$$\alpha_{(p)} = G(\theta) \quad , \quad \theta = G^{-1}(\alpha_{(p)}) \quad . \quad (5.21)$$

Edellä sanotuin ehdoin $\alpha_{(p)}$ voidaan estimoida tarkentuvasti vastaavilla otosmomenteilla

$$\hat{\alpha}_\nu = \frac{1}{n} \sum_{t=1}^n y_t^\nu \quad \nu = 1, \dots, p \quad ,$$

jolloin (5.21):n mukaisesti

$$\hat{\theta} = G^{-1}(\hat{\alpha}_{(p)}) \quad (5.22)$$

olisi θ :n tarkentuva estimaattori. Estimaattoria (5.22) sanotaan *momenttiperiaatteen* mukaisesti konstruoiduksi.

Toisaalta myös korkeamman kertaluvun momentit $\alpha_\nu \quad \nu = p+1, p+2, \dots$ riippuvat yleensä θ -parametreista, joten tulee etsimättä mieleen käyttää useampia estimoituja momenteja $\hat{\alpha}_1, \dots, \hat{\alpha}_r \quad (r > p)$ hyväksi θ :n estimoinnissa. Jos $r > p$, ei koko jakaumaperheestä tietenkään löydy ainuttakaan jakaumaa, jolle $\hat{\alpha}_{(r)} = \alpha_{(r)}$ pitäisi täsmälleen paikkansa. Sen sijaan voitaisiin yrittää hakea malliperheestä sellaista jakaumaa, jolle $\alpha_{(r)}$ olisi (jossakin metriikassa mitattuna) mahdollisimman lähellä $\hat{\alpha}_{(r)}$:ää. Tätä periaatetta kutsutaan *yleistetyksi momenttimenetelmäksi* (GMM).

Formalisoidaan tämä periaate seuraavasti: (Itse havainnot voivat olla yhtä hyvin vektoriarvoisia arvoalueena \mathcal{R}^k , joten siirrytään tältä osin suuriin kirjaimiin $Y_t \quad t = 1, \dots, n$.) Oletetaan, että on löydetty r -ulotteinen funktio ($r \geq p$) $H(\theta, Y_t)$, jolle

$$E^{(\theta)} H(\theta, Y_t) \equiv 0 \quad . \quad (5.23)$$

Odotusarvoa (5.23) voidaan estimoida (momenttiperiaatteen mukaisesti) suurella

$$G(\theta, Y) = \frac{1}{n} \sum_{t=1}^n H(\theta, Y_t) \quad ,$$

mikäli $\{H(\theta, Y_t)\}$ -prosessi on *ergodinen*. (Näin käy, mikäli Y_t -havaintojen väliset mahdolliset riippuvuudet katoavat ”riittävän nopeasti” havaintojen välisen ajallisen etäisyyden kasvaessa.) Muodostetaan nyt θ :lle ns. GMM-estimaattori minimoimalla lauseketta

$$G(\theta, Y)' M G(\theta, Y) \quad , \quad (5.24)$$

ts. valitsemalla θ siten, että $G(\theta, Y)$ olisi mahdollisimman lähellä origoa M -metriikassa mitattuna.

Voidaan osoittaa, että mikäli peräkkäiset $H(\theta, Y_t)$:t eivät korreloi keskenään, saadaan optimaalinen M -matriisi kovarianssimatriisiin

$$S = \lim_{n \rightarrow \infty} n \operatorname{cov}(G(\theta, Y)) = \lim_{n \rightarrow \infty} n E G(\theta, Y) G(\theta, Y)'$$

inverssinä. Matriisia S voidaan puolestaan estimoida matriisilla

$$\frac{1}{n} \sum_{t=1}^n H(\theta, Y_t) H(\theta, Y_t)' \quad ,$$

mutta θ on tietenkin toistaiseksi tuntematon. Olkoon nyt $\hat{\theta} = \hat{\theta}(Y)$ mikä tahansa \sqrt{n} -tarkentuva θ :n estimaattori. Tällöin

$$\hat{S}_n = \hat{S}_n(\hat{\theta}) = \frac{1}{n} \sum_{t=1}^n H(\hat{\theta}, Y_t) H(\hat{\theta}, Y_t)'$$

on luonnollisesti S :n tarkentuva estimaattori ($\text{p lim}_{n \rightarrow \infty} \hat{S}_n = S$) ja kriteerifunktiona (5.24) voidaan käyttää lauseketta

$$c(\theta) = G(\theta, Y)' \hat{S}_n(\hat{\theta})^{-1} G(\theta, Y) \quad (5.25)$$

jota pyritään siis minimoimaan θ :n suhteen. *Kaksivaiheisella* GMM-estimaattorilla tarkoitetaan tällöin estimaattoria

$$\hat{\theta}_{GMM} = \arg \min_{\theta} G(\theta, Y)' \hat{S}_n(\hat{\theta})^{-1} G(\theta, Y) \quad (5.26)$$

ja *iteratiivisella* GMM-estimaattorilla tarkoitetaan raja-arvoa

$$\hat{\theta}_{ITGMM} = \lim_{j \rightarrow \infty} \hat{\theta}_{GMM}^{(j)}, \quad \text{jossa} \quad \hat{\theta}_{GMM}^{(j)} = \arg \min_{\theta} G(\theta, Y)' \hat{S}_n(\hat{\theta}_{GMM}^{(j-1)})^{-1} G(\theta, Y) \quad . \quad (5.27)$$

Mikäli peräkkäiset $H(\hat{\theta}, Y_t)$:t korreloivat keskenään aikaeroilla $\nu = 1, \dots, q$, voidaan \hat{S}_n :n paikalla käyttää ns. Newey-West-estimaattoria

$$\hat{S}_n = \hat{\Gamma}_{o,n} + \sum_{\nu=1}^q \left[1 - \frac{\nu}{q+1} \right] (\hat{\Gamma}_{\nu,n} + \hat{\Gamma}'_{\nu,n}) \quad , \quad (5.28)$$

jossa

$$\hat{\Gamma}_{\nu,n} = \frac{1}{n} \sum_{t=1}^n H(\hat{\theta}, Y_t) H(\hat{\theta}, Y_{t-\nu})' \quad \nu = 0, \dots, q \quad .$$

Huomautus 5.1: Mikäli H -funktiot toteuttavat seuraavat ehdot:

- $\text{p lim}_{n \rightarrow \infty} \hat{\theta}_{GMM} = \theta$
- $\sqrt{n} G(\theta, Y) \stackrel{\text{asympt.}}{\sim} N_p(0, S)$
- jokaiselle tarkentuvalla estimaattorijonolla $\tilde{\theta}_n \xrightarrow{p} \theta$ pätee

$$\text{p lim}_{n \rightarrow \infty} D_{\theta} G(\tilde{\theta}_n, Y) = \text{p lim}_{n \rightarrow \infty} D_{\theta} G(\theta, Y) = W(\theta) \quad ,$$

pätee sekä iteratiiviselle että kaksivaiheiselle GMM-estimaattorille tulos

$$\sqrt{n} (\hat{\theta}_{GMM} - \theta) \stackrel{\text{asympt.}}{\sim} N_p(0, V) \quad (5.29)$$

jossa

$$V = [W(\theta)'S^{-1}W(\theta)]^{-1} \quad .$$

■

Huomautus 5.2: Kannattaa mainita, että jopa ML-estimaattorit voidaan monessa tapauksessa tulkita GMM-estimaattoreiksi.

Merkitään $\mathcal{F}_t = \{Y_\tau \mid \tau \leq t\}$ ja tarkastellaan ehdolliseen likelihood-funktioon

$$L_{Y_t|\mathcal{F}_{t-1}}(\theta) = f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y_t)$$

liittyvää score-vektoria

$$\begin{aligned} H(\theta, Y_t) &= D_\theta \log L_{Y_t|\mathcal{F}_{t-1}}(\theta)' \\ &= \left(\frac{\partial}{\partial \theta_1} \log L_{Y_t|\mathcal{F}_{t-1}}(\theta) \quad \dots \quad \frac{\partial}{\partial \theta_p} \log L_{Y_t|\mathcal{F}_{t-1}}(\theta) \right)' \quad . \end{aligned} \quad (5.30)$$

Näin valitulle $H(\theta, Y_t)$ -funktiolle nimittäin pätee

$$\begin{aligned} EH(\theta, Y_t) &= \int_{\mathcal{R}^k} \dots \int [D_\theta \log f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y)] \cdot f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y) dy_1 \dots dy_k \\ &= \int_{\mathcal{R}^k} \dots \int \left[\frac{1}{f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y)} D_\theta f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y) \right] \cdot f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y) dy_1 \dots dy_k \\ &= \int_{\mathcal{R}^k} \dots \int D_\theta f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y) dy_1 \dots dy_k \\ &= D_\theta \int_{\mathcal{R}^k} \dots \int f_{Y_t|\mathcal{F}_{t-1}}^{(\theta)}(Y) dy_1 \dots dy_k \\ &= D_\theta 1 = 0 \quad , \end{aligned}$$

jossa $Y = (y_1 \dots y_k)' \in \mathcal{R}^k$.

Huomataan siis, että ehdolliset score-vektorit (5.30) kelpaavat $H(\theta, Y_t)$ -funktioiksi.

Tällä tulkinnalla GMM-menetelmä johtaisi lausekkeen

$$G(\theta, Y)' \hat{S}_n^{-1} G(\theta, Y)$$

minimointiin.

Toisaalta tässä tapauksessa $r = p$, joten vektori

$$G(\theta, Y) = \frac{1}{n} \sum_{t=1}^n D_{\theta} \log L_{Y_t | \mathcal{F}_{t-1}}(\theta)' = \frac{1}{n} D_{\theta} \log L_Y(\theta)'$$

saadaan ML-estimaattorin $\hat{\theta}$ kohdalla täsmälleen nolaksi, joten ML-estimaattori on samalla tulkintaan (5.30) liittyvä GMM-estimaattori.

Toisaalta, jos score-vektori on asymptoottisesti normaalinen, on tuloksen (5.29) perusteella

$$\sqrt{n} (\hat{\theta} - \theta) \overset{\text{asympt.}}{\sim} N_p(0, V) \quad ,$$

jossa $V = [W(\theta)' S^{-1} W(\theta)]^{-1}$. Rajamatriisi S puolestaan voidaan tulkita myös yhteen havaintoon keskimäärin sisältyväksi informaatioksi, sillä

$$S = \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n D_{\theta} \log L_{Y_t | \mathcal{F}_{t-1}}(\theta)' D_{\theta} \log L_{Y_t | \mathcal{F}_{t-1}}(\theta) = \mathcal{I}_1(\theta) \quad .$$

Toisaalta

$$\begin{aligned} \text{p} \lim_{n \rightarrow \infty} W(\hat{\theta}) &= \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n D_{\theta} H(\hat{\theta}, Y_t) \\ &= \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n D_{\theta}^2 \log L_{Y_t | \mathcal{F}_{t-1}}(\theta) \\ &= \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} D_{\theta}^2 \log L_Y(\theta) \\ &= -\mathcal{I}_1(\theta) \quad . \end{aligned}$$

Kaavan (5.29) erikoistapauksena saadaan siis tuttu, ML-estimaattoreita koskeva tulos

$$\sqrt{n} (\hat{\theta} - \theta) \overset{\text{asympt.}}{\sim} N_p(0, \mathcal{I}_1(\theta)^{-1}) \quad .$$

■

Esimerkki 5.1: Tarkastellaan lineaarista regressiomallia

$$y_t = \beta' X_t + \varepsilon_t \quad , \quad \beta \in \mathcal{R}^p \quad , \quad \varepsilon_t \sim i.i.d(0, \sigma^2)$$

jossa (syystä tai toisesta) $EX_t \varepsilon_t \neq 0$. (Lisää esimerkkejä tällaisista tilanteista esitetään luvuissa 6 ja 7.)

Ajatellaan, että voitaisiin löytää p kappaletta ”instrumenteiksi” kelpaavia muuttujia $Z_t = (z_{t1} \dots z_{tp})'$, joille

$$EZ_t \varepsilon_t \equiv 0 \quad ,$$

mutta jotka korreloisivat voimakkaasti X_t -muuttujien kanssa. Tällöin ”ortogonaalisuusehtoina” (5.23) voidaan pitää yhtälöitä

$$EZ_t(y_t - \beta' X_t) \equiv 0 \quad ,$$

joten merkitsemällä $W_t = (y_t \quad X_t' \quad Z_t)'$ voidaan H -funktioksi valita

$$H(\beta, W_t) = Z_t(y_t - \beta' X_t) \quad .$$

Merkitsemällä

$$Y = (y_1 \quad \dots \quad y_n)' \quad , \quad X = (X_1 \quad \dots \quad X_n)'$$

$$Z = (Z_1 \quad \dots \quad Z_n)' \quad \text{ja} \quad W = (Y \quad X \quad Z)$$

saadaan

$$G(\beta, W) = \frac{1}{n} \sum_{t=1}^n Z_t(y_t - X_t' \beta) = \frac{1}{n} [Z'Y - Z'X\beta] \quad .$$

Yhtälön $G(\beta, W) = 0$ ratkaisuna saadaan ns. *instrumenttiestimaattori*

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y \quad . \quad (5.31)$$

Ajatellaan seuraavaksi, että instrumenteiksi kelpaavia muuttujia $Z_t = (z_{t1} \quad \dots \quad z_{tr})'$ olisi löydettävissä vähimmäismäärää enemmän ($r > p$). Tällöin pitäisi edelleen tulos $G(\beta, W) = \frac{1}{n} [Z'Y - Z'X\beta]$, minkä lisäksi olisi

$$W(\beta) = \text{p} \lim_{n \rightarrow \infty} D_\beta G(\beta, W) = -\text{p} \lim_{n \rightarrow \infty} \frac{1}{n} X'Z$$

ja

$$\begin{aligned} S &= \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{t'=1}^n \varepsilon_t \varepsilon_{t'} Z_t Z_{t'}' \\ &= \sigma^2 \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} Z'Z \quad , \end{aligned}$$

mikäli $\varepsilon_t Z_t$ -vektorit eivät korreloi keskenään ($\varepsilon_t \sim i.i.d.(0, \sigma^2)$).

Tätä asymptoottista kovarianssimatriisia S vastaavan GMM-estimaattorin pitäisi siis minimoida kohdefunktio

$$\frac{1}{n} [Z'Y - Z'X\beta]' (Z'Z)^{-1} [Z'Y - Z'X\beta] \quad .$$

Tämä tavoite toteutuu, kun

$$\begin{aligned}\hat{\beta} &= [X'Z(Z'Z)^{-1}Z'X]^{-1} X'Z(Z'Z)^{-1}Z'Y \\ &= [Z*'X]^{-1} Z*'Y \quad ,\end{aligned}\tag{5.32}$$

jossa on merkitty $Z^* = P_Z X = Z(Z'Z)^{-1}Z'X$.

Tuloksesta (5.32) on helppo tunnistaa ns. *yleistetty* instrumenttimestimaattori $\hat{\beta}_{GIVE}$. Voidaan siis todeta, että $\hat{\beta}_{GIVE}$ ja $\hat{\beta}_{GMM}$ ovat läheistä sukua keskenään (merk. $\hat{\beta}_{GIVE} \approx \hat{\beta}_{GMM}$). Erona on vain se, että GMM-estimoinnissa S -matriisiin paikalla käytetään (robustimpaa) ”sandwich”-estimaattoria $\hat{S}_n = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2 Z_t Z_t'$. Tuloksen (5.29) mukaisesti kummankin estimaattorin otantajakaumia voidaan approksimoida jakaumalla

$$N_p(\beta, \hat{\sigma}^2 [X'Z(Z'Z)^{-1}Z'X]^{-1}) \quad .\tag{5.33}$$

Jos toisaalta $\hat{\beta}_{GMM} \approx \hat{\beta}_{GIVE}$:n otantajakauma halutaan arvioida mahdollisen autokorrelaation suhteen robustilla tavalla, voidaan kaavassa (5.33) tarvittava matriisi $\hat{\sigma}^2 \frac{1}{n} Z'Z = \hat{S}_n$ korvata aikaisemmin mainitulla Newey-West-estimaattorilla

$$\hat{S}_n = \hat{\Gamma}_{o,n} + \sum_{\nu=1}^q \left[1 - \frac{\nu}{q+1} \right] (\hat{\Gamma}_{\nu,n} + \hat{\Gamma}'_{\nu,n}) \quad ,$$

jossa

$$\hat{\Gamma}_{\nu,n} = \frac{1}{n} \sum_{t=\nu+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-\nu} Z_t Z_{t-\nu}' \quad \nu = 0, \dots, q$$

ja

$$\hat{\varepsilon}_t = y_t - X_t' \hat{\beta}_{GMM} \quad .$$

■

Huomautus 5.3: Jos $r > p$, voidaan ”ortogonaalisuusehtojen” (5.23) realisuutta periaatteessa tutkia ns. Hansenin testisuurella

$$n G(\hat{\theta}, Y)' \hat{S}_n^{-1} G(\hat{\theta}, Y) \quad ,\tag{5.34}$$

jonka asymptoottisen otantajakauman pitäisi olla tyyppiä χ_{r-p}^2 .

■

Esimerkki 5.1: (jatkoa)

Tarkastellaan tupakkatuotteiden reaalihinnan P_t vaikutuksia tupakkatuotteiden kysyntään Q_t vuosina $t = 1951, \dots, 2001$ ARX-mallin

$$\log(Q_t) = \beta_0 + \beta_1 \log(Q_{t-1}) + \beta_2 \log(FC_t) + \beta_3 \log(P_t) + \varepsilon_t, \varepsilon_t \sim NID(0, \sigma^2)$$

puitteissa. Tässä FC_t tarkoittaa elintarvikkeiden, alkoholijuomien ja tupakan yhteenlaskettua kulutusmäärää (kansantalouden tilinpidon hyödykepääryhmän 1 kulutusvolyymiä). (Tupakkatuotteiden reaalihintaa on laskettu vertaamalla niiden hintoja pääryhmän 1 keskimääräisiin hintoihin.)

Koska tupakointi on todettu terveydelle vaaralliseksi, on tupakkalakia kuitenkin muutettu radikaalisti vuosina 1977 (jolloin mm. kiellettiin tupakan mainonta) ja 1994 (jolloin mm. ryhdyttiin voimakkaasti rajoittamaan työpaikoilla tapahtuvaa tupakointia). (Mainitut vuodet ovat lakimuutosten *voimaantulovuosia*.) Laajennetaan mallia muotoon, joka varautuu tupakkatuotteiden kysyntärakenteen mahdollisiin muutoksiin näiden lakimuutosten seurauksina. (Dummymuuttujille *dum77* ja *dum94* on annettu arvo 1 lakimuutosvuodesta aina vuoteen 2001 saakka.) Lakimuutokset huomioon ottava versio mallista johtaa kuitenkin OLS-jäännöstermeihin, joiden vierekkäiset arvot korreloivat keskenään. Estimoitujen autokorrelaatioiden perusteella näyttäisi siltä, että yksinkertainen liukuvan keskiarvon malli

$$\varepsilon_t = \kappa_t - \theta \kappa_{t-1}, \quad \kappa_t \sim NID(0, \sigma^2)$$

pystyisi kuvaamaan virhetermisarjaan sisältyvää ajallista säännöllisyyttä hyvinkin realistisesti. Tästä tietysti seuraa, että ε_t ja $\log(Q_{t-1})$ korreloivat keskenään, joten mallin parametrien estimoinnin yhteydessä on syytä turvautua GIVE- tai GMM-estimointiin, kunhan vain sopivia instrumentteja voidaan löytää.

Yllä hahmotellun mallikehikon puitteissa sopivia instrumentteja olisivat selvästi ainakin $\log(Q_{t-2})$ ja $\log(Q_{t-3})$.

PcGive: (GIVE-estimointi)

```
P1 = V1/Q1;
P1D = V1D/Q1D;
Pt = P1D/P1;
LPt = log(Pt);
LQt = log(Q1D);
LFCt = log(Q1);
LQt_1 = lag(LQt,1);
dum77 = dummy(1977,1, 2001,1);
dum94 = dummy(1994,1, 2001,1);
d77LPt = dum77*LPt;
d94LPt = dum94*LPt;
```

```
EQ(1) Modelling LQt by OLS (using ALKTUP.IN7)
The present sample is: 1953 to 2001
```

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
LQt_1	0.20660	0.13112	1.576	0.1228	0.0571
Constant	-0.92516	0.38699	-2.391	0.0215	0.1223
LPt	0.11538	0.10184	1.133	0.2638	0.0304
LFct	0.74241	0.12995	5.713	0.0000	0.4432
dum77	-0.31065	0.086088	-3.609	0.0008	0.2410
dum94	-0.18247	0.048527	-3.760	0.0005	0.2564
d77LPt	-0.32682	0.13422	-2.435	0.0193	0.1263
d94LPt	-0.51743	0.23777	-2.176	0.0354	0.1035

R² = 0.967525 F(7,41) = 174.5 [0.0000] \sigma = 0.0397596 DW = 1.20
 RSS = 0.06481396905 for 8 variables and 49 observations

Residual correlogram

Portmanteau statistic for 6 lags and 50 observations: 10.53

Autocorrelation coefficients

0.38939 0.18708 -0.083791 -0.015549 -0.067259 -0.076485

EQ(2) Modelling LQt by IVE (using ALKTUP.IN7)

The present sample is: 1953 to 2001

Variable	Coefficient	Std.Error	t-value	t-prob
LQt_1	-0.21773	0.38438	-0.566	0.5742
Constant	-1.4804	0.63560	-2.329	0.0249
LPt	0.19141	0.13065	1.465	0.1505
LFct	1.1458	0.36769	3.116	0.0033
dum77	-0.46174	0.15906	-2.903	0.0059
dum94	-0.26292	0.086551	-3.038	0.0041
d77LPt	-0.49445	0.20568	-2.404	0.0208
d94LPt	-1.0380	0.51070	-2.032	0.0486

Additional Instruments used:

LQt_2 LQt_3

\sigma = 0.0445495 DW = 0.76

RSS = 0.08137102281 for 8 variables and 49 observations
 2 endogenous and 7 exogenous variables with 9 instruments

SAS: (GMM-estimointi)

```

proc import out=alktup
  datafile="c:\mr\data\alktup.xls"
  dbms=excel2000 replace; getnames=yes ; run;
data alktup; set alktup; drop F1;
  logFC=log(Q1); logQ=log(Q1D); logQ_1=lag(logQ);
  vuosi=1949+_n_;
  P1D=V1D/Q1D; P1=V1/Q1; logP=log(P1D/P1);
  dum77=0; if vuosi>1976 then dum77=1;
  dum94=0; if vuosi>1993 then dum94=1;
  d77logP=dum77*logP; d94logP=dum94*logP;
  l2logQ=lag2(logQ); l3logQ=lag3(logQ); run;
proc reg data=alktup;
  model logQ = logQ_1 logFC logP dum77 dum94 d77logP d94logP ;
  where vuosi>1952 ;
  output out=tupres residual=e; run;
proc arima data=tupres;
  identify var=e ; run;
proc model data=alktup;
  parms beta0-beta7;
  logQ = beta0+beta1*logQ_1+beta2*logFC+beta3*logP+beta4*dum77+beta5*dum94
        +beta6*d77logP+beta7*d94logP;
  exogenous logFC logP dum77 dum94 d77logP d94logP ;
  fit logQ / gmm ;
  instruments l2logQ l3logQ _exog_;
  where vuosi>1952 ; run;

```

The REG Procedure
 Dependent Variable: logQ

Root MSE 0.03976 R-Square 0.9675

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.92516	0.38699	-2.39	0.0215
logQ_1	1	0.20660	0.13112	1.58	0.1228
logFC	1	0.74241	0.12995	5.71	<.0001
logP	1	0.11538	0.10184	1.13	0.2638
dum77	1	-0.31065	0.08609	-3.61	0.0008
dum94	1	-0.18247	0.04853	-3.76	0.0005
d77logP	1	-0.32682	0.13422	-2.43	0.0193
d94logP	1	-0.51743	0.23777	-2.18	0.0354

The ARIMA Procedure
Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	10.18	6	0.1171	0.385	0.182	-0.078	-0.013	-0.054	-0.060

The MODEL Procedure
GMM Estimation Summary

Parameters Estimated 8
Kernel Used PARZEN
l(n) 2.177906
Method Gauss
Iterations 1

Nonlinear GMM Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t
beta0	-1.6772	0.5523	-3.04	0.0041
beta1	-0.31742	0.3593	-0.88	0.3822
beta2	1.252194	0.3421	3.66	0.0007
beta3	0.283838	0.1049	2.71	0.0099
beta4	-0.53987	0.1264	-4.27	0.0001
beta5	-0.28181	0.0708	-3.98	0.0003
beta6	-0.61169	0.1518	-4.03	0.0002
beta7	-1.15725	0.4571	-2.53	0.0153

Number of Observations Statistics for System

Used 49 Objective 0.0342
Missing 0 Objective*N 1.6756

6 PANEELIAINEISTOJEN TUTKIMINEN

6.1 Regressiotyypiset (staattiset) mallit

Tarkastellaan paneeliaineistoa, joka koostuu selittävistä muuttujista x_1, \dots, x_m ja vastemuuttujasta y saaduista havainnoista $X_{it} = (x_{i,t,1} \dots x_{i,t,m})'$ ja y_{it} , jossa indeksi i ($i = 1, \dots, n$) viittaa yksilöön (talouden toimijaan) ja indeksi t viittaa havainnon järjestysnumeroon tai ajankohtaan ($t = 1, \dots, T_i$). Oletetaan, että selittävien muuttujien X_{it} vaikutukset vastemuuttujaan y_{it} ovat *yksisuuntaisia* ja että niitä voidaan kuvata *lineaarisella* regressiomallilla. Useimmiten joudutaan kuitenkin varautumaan vastemuuttujassa ilmenevään *yksilölliseen* tasovaihteluun mallirakenteen

$$y_{it} = u_i + \beta' X_{it} + \varepsilon_{it} \quad , \quad \varepsilon_{it} \perp\!\!\!\perp X_{it} \quad , \quad \varepsilon_{it} \sim N(0, \sigma^2) \quad , \\ t = 1, \dots, T_i \quad , \quad i = 1, \dots, n$$

mukaisesti. Mikäli yksilölliset tasotermit u_i tulkittaisiin *parametreiksi*, päädyttäisiin mallirakennelmaan, jossa havaintomäärän kasvattaminen samalla kasvat-
taisi myös parametrien lukumäärää. Tästä puolestaan seuraisi, että luvussa 2
esitellyt likelihood-päätelyyn liittyvät asymptoottiset tulokset eivät enää olisi
voimassa. Esimerkiksi ML-estimaattorit eivät olisi kaikkien parametrien suh-
teen edes tarkentuvia, vaan *epäkonsistentteja*. Tästä syystä tuntuu usein hou-
kuttelevalta tulkita u_i -tasotermit *satunnaismuuttujiksi*, joilla olisi oma jakau-
mansa tutkittavassa ”yksilöpopulaatiossa”. Teknisesti mukavin tasotermien ja-
kauman muotoa koskeva oletus olisi olettaa ne normaalisesti jakautuneiksi,
 $u_i \sim NID(0, \tau^2)$, ja olettaa havaintokohtaisten virhetermien ε_{it} olevan
 u_i -termeistä riippumattomia:

$$y_{it} = u_i + \beta' X_{it} + \varepsilon_{it} \quad , \quad \varepsilon_{it} \perp\!\!\!\perp X_{it} \quad , \quad \varepsilon_{it} \perp\!\!\!\perp u_i \quad , \\ \varepsilon_{it} \sim N(0, \sigma^2) \quad , \quad u_i \sim NID(0, \tau^2) \quad , \quad (6.1) \\ t = 1, \dots, T_i \quad , \quad i = 1, \dots, n \quad .$$

Virhetermisarjat $\{\varepsilon_{it}\}$ täytyy voida olettaa *stationäärisiksi*, mutta peräkkäisten virhetermien välisiin korrelaatioihin (ts. autokorrelaatioon) on monissa tapauk-
sissa syytä varautua. Jos yksilöstä i saadut havainnot kootaan T_i -ulotteiseksi
vektoriksi

$$Y_i = (y_{i1} \dots y_{iT_i})' \in \mathcal{R}^{T_i}$$

ja $T_i \times m$ -matriisiksi

$$X_i = (X_{i1} \ \dots \ X_{iT_i})' = \begin{pmatrix} x_{i,1,1} & \dots & x_{i,1,m} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ x_{i,T_i,1} & \dots & x_{i,T_i,m} \end{pmatrix},$$

voidaan malli (6.1) kirjoittaa muotoon

$$\begin{aligned} Y_i &= u_i J_i + X_i \beta + \varepsilon_i \quad , \quad \varepsilon_i \perp\!\!\!\perp X_i \quad , \quad \varepsilon_i \perp\!\!\!\perp u_i \quad , \\ \varepsilon_i &\sim N_{T_i}(0, \sigma^2 R_i) \quad , \quad u_i \sim NID(0, \tau^2) \quad , \quad i = 1, \dots, n \quad , \end{aligned} \quad (6.2)$$

jossa $J_i = (1 \ \dots \ 1)' \in \mathcal{R}^{T_i}$ ja R_i tarkoittaa virhetermivektorin $\varepsilon_i = (\varepsilon_{i1} \ \dots \ \varepsilon_{iT_i})'$ korrelaatiomatriisia

$$R_i = \text{corr}(\varepsilon_i) = \begin{pmatrix} 1 & \rho_\varepsilon(1) & \rho_\varepsilon(2) & \cdot & \cdot & \cdot & \rho_\varepsilon(T_i - 1) \\ \rho_\varepsilon(1) & 1 & \rho_\varepsilon(1) & \cdot & \cdot & \cdot & \rho_\varepsilon(T_i - 2) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \rho_\varepsilon(1) & 1 & \rho_\varepsilon(1) \\ \rho_\varepsilon(T_i - 1) & \cdot & \cdot & \cdot & \rho_\varepsilon(2) & \rho_\varepsilon(1) & 1 \end{pmatrix}.$$

(Tällaisia ”nauhamatriiseja”, joissa päädiagonaalin suuntaisilla nauhoilla on samat alkio, sanotaan *Töplitz*-matriiseiksi.) Vaikka kaikki virhetermien autokorrelaatiot $\rho_\nu = \rho_\varepsilon(\nu)$ ($\nu = 1, \dots, T - 1$; $T = \max_i T_i$) jätettäisiin vapaiksi parametreiksi, olisi mallissa (6.2) siis vain $m + T + 1$ tuntematonta parametria. Ja, mikä tärkeintä, parametrien määrä pysyisi muuttumattomana, vaikka seurattujen yksilöiden lukumäärää lisättäisiin.

Yhdistämällä mallin (6.2) satunnaistermit yhdeksi vektoriksi $\omega_i = u_i J_i + \varepsilon_i$ on helppo huomata, että kyseinen mallikokonaisuus voidaan tulkita tavalliseksi (vektorimuotoiseksi) regressiomalliksi latomalla eri yksilöistä saadut tiedot allekkain:

$$Y = (Y_1' \ \dots \ Y_n')' = X\beta + \omega \quad , \quad \omega \sim N_N(0, V) \quad , \quad (6.2')$$

jossa $N = T_1 + \dots + T_n$, $\omega = (\omega_1' \ \dots \ \omega_n')'$ ja V on blokkidiagonaalimatriisi

$$V = \text{diag}(\text{cov}(\omega_1), \dots, \text{cov}(\omega_n)) = \text{diag}(\tau^2 J_1 J_1' + \sigma^2 R_1, \dots, \tau^2 J_n J_n' + \sigma^2 R_n) \quad .$$

Kuten tunnettua, on mallin (6.2') mukaisen multinormaalijakauman tiheysfunk-

tio muotoa

$$f_Y(Y) = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \cdot \frac{1}{\sqrt{\det(V)}} e^{-\frac{1}{2}(Y-X\beta)'V^{-1}(Y-X\beta)} ,$$

joten havaintoaineiston Y määräämän likelihood-funktion logaritmi on muotoa

$$\log L_Y(\beta, \tau^2, \sigma^2, \rho_1, \dots, \rho_{T-1}) = -\frac{1}{2} \log \det(V) - \frac{1}{2}(Y - X\beta)'V^{-1}(Y - X\beta) . \quad (6.3)$$

Jos V (ts. τ^2 , σ^2 ja $\rho_1, \dots, \rho_{T-1}$) tunnettaisiin, olisi β :n ML-estimaattori muotoa

$$\hat{\beta}(V) = (X'V^{-1}X)^{-1}X'V^{-1}Y , \quad (6.4)$$

ts. ns. GLS-estimaattori. Käytännössä V täytyy kuitenkin ensin estimoida maksimoimalla lauseketta (6.3) iteratiivisten optimointialgoritmien avulla.

Esimerkki 6.1: Tarkastellaan 11 pörssi-yhtiön osakkeiden vaihtomäärien (merkitään logaritmoituja määriä symboleilla $lvol_{it}$) ja kurssikehitykseen perustuvien päivätuottojen (merkitään tuottoja symboleilla ret_{it}) vaihteluita Helsingin Arvopaperi- ja johdannaispörssissä kauppapäivittäin vuonna 1995. On hyvä syy uskoa, että päivätuottojen kehitys olisi melko ennakoimaton, joten mahdollisen riippuvuuden $ret_{it} \rightarrow lvol_{it}$ voidaan olettaa olevan yksisuuntainen. Tutkitaan nyt tarkemmin, miten kunkin yhtiön osakkeen tuottokehitys vaikuttaa osakkeiden vaihtomääriin. Kokeillaan yksinkertaista siirtofunktiomallia

$$\begin{aligned} lvol_{it} &= \mu + u_i + \beta_1 ret_{it} + \beta_2 ret_{i,t-1} + \beta_3 ret_{i,t-2} + \varepsilon_{it} , \\ u_i &\sim NID(0, \tau^2) , \quad \varepsilon_{it} \sim N(0, \sigma^2) , \quad u_i \perp\!\!\!\perp \{\varepsilon_{it}\} , \\ t &= 2, \dots, 249 ; \quad i = 1, \dots, 11 \end{aligned}$$

täydennettynä joko ARMA(1,1)- tai MA(4)-virhetermimallilla.

Näyttää siltä, että ARMA(1,1)-virhetermimalli sopii aineistoon kilpailijaansa paremmin. Päivätuottomuutosten vaikutukset osakkeiden vaihtomääriin näyttävät menevän ohi jo vajaassa kahdessa vuorokaudessa.

SAS:

```
data stock; infile 'c:\mr\long05\stock.dat';
input firm date volume Lvol Price Lprice Return ;
retain time 1; time=time+1;
lret=lag(return); time=time+1;
if firm^=lag(firm) then do; lret=missing; time=1; end;
l2ret=lag(lret); if firm^=lag(firm) then l2ret=missing; run;
proc mixed data=stock method=ml; class firm ;
model Lvol = Return lret l2ret / s outpred=sovite ;
random intercept / type=un subject=firm s ;
*repeated / type=TOEP(5) subject=firm ;
repeated / type=ARMA(1,1) subject=firm ;
id firm time Lvol; run;
proc gplot data=sovite;
plot (Lvol pred)*time / overlay nolegend;
by firm; symbol1 v=none c=green i=join w=1 r=1 ;
symbol2 v=none c=red i=join w=2 l=3 r=1 ; run;
```

The Mixed Procedure

Dependent Variable	Lvol
Covariance Structures	Unstructured, Autoregressive Moving Average
Subject Effects	firm, firm
Estimation Method	ML

Dimensions	
Covariance Parameters	4
Columns in X	4
Columns in Z Per Subject	1
Subjects	11
Max Obs Per Subject	247

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	firm	3.4863
Rho	firm	0.8110
Gamma	firm	0.1966
Residual		1.6450

Fit Statistics

-2 Log Likelihood	8752.3
AIC (smaller is better)	8768.3
BIC (smaller is better)	8771.5

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	9.7212	0.5646	10	17.22	<.0001
Return	4.2766	1.1146	2646	3.84	0.0001
lret	3.2546	1.1105	2646	2.93	0.0034
l2ret	0.5678	1.1127	2646	0.51	0.6099

The Mixed Procedure

Dependent Variable Lvol
Covariance Structures Unstructured, Toeplitz
Subject Effects firm, firm
Estimation Method ML

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	firm	3.5050
TOEP(2)	firm	0.3090
TOEP(3)	firm	0.2384
TOEP(4)	firm	0.09654
TOEP(5)	firm	0.09783
Residual		1.6274

Fit Statistics

-2 Log Likelihood	8768.7
AIC (smaller is better)	8788.7
BIC (smaller is better)	8792.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	9.7196	0.5655	10	17.19	<.0001
Return	4.2259	1.1260	2646	3.75	0.0002
lret	3.1982	1.1228	2646	2.85	0.0044
l2ret	0.4706	1.1243	2646	0.42	0.6756

6.3 GMM- menetelmän käyttö ARX- tyyppisten mallien parametrien estimoinnissa

Tarkastellaan nyt ARX- tyyppisiä paneelimalleja, jotka eroavat tyyppistä (6.1) sikäli, että malliyhtälön oikealla puolella määriteltyyn regressio-osaan sisältyy myös *vastemuuttujan* viivästeisiä arvoja. Korvataan lisäksi malliin (6.1) sisältyneiden satunnaistermien jakaumaoletukset $u_i \sim NID(0, \tau^2)$ ja $\varepsilon_{it} \sim NID(0, \sigma^2)$ epämääräisemmillä oletuksilla $u_i \sim i.i.d.(0, \tau^2)$ ja $\varepsilon_{it} \sim i.i.d.(0, \sigma^2)$:

$$\begin{aligned}
 y_{it} &= \beta' X_{it} + \phi_1 y_{i,t-1} + \dots + \phi_p y_{i,t-p} + u_i + \varepsilon_{it} \quad , \\
 u_i &\sim i.i.d.(0, \tau^2) \quad , \quad \varepsilon_{it} \sim i.i.d.(0, \sigma^2) \quad , \\
 \{\varepsilon_{it}\} &\perp\!\!\!\perp \{X_{it}\} \quad , \quad \{\varepsilon_{it}\} \perp\!\!\!\perp \{u_i\} \\
 t &= p + 1, \dots, T_i \quad ; \quad i = 1, \dots, m \quad .
 \end{aligned}
 \tag{6.5}$$

Näin saatu malli on estimointiteknisesti sikäli hankala, että yhdistetty satunnais-termi $\omega_{it} = u_i + \varepsilon_{it}$ korreloi regressio-osaan liittyvien muuttujien $y_{i,t-1}, \dots, y_{i,t-p}$ kanssa, koska yksilöön i liittyvä tasotermi u_i sisältyy kaikkiin näihin muuttujiin. Vaikka ML-estimointiperiaatteenkin käyttö olisi (normaalijakaumaoletusten puitteissa) täysin mahdollista, neuvotaan seuraavassa vain luvussa 5 esiteltyjen GIVE- ja GMM-menetelmien käyttöä.

On helppo havaita, että mallimuodon (6.5) puitteissa ϕ - ja β -parametrit voidaan estimoida GMM-menetelmällä, mikäli $\{y_{it}\}$ -prosessit ovat *stationäärisiä*, sillä tällöin

$$\nabla y_{i,t-\nu} \perp\!\!\!\perp \omega_{it} = u_i + \varepsilon_{it} \quad ,$$

mutta

$$\nabla y_{i,t-\nu} \not\perp\!\!\!\perp y_{i,t-1}, \dots, y_{i,t-p}$$

kaikilla $\nu = 1, \dots, p$. Differenssit $\nabla y_{i,t-\nu} = y_{i,t-\nu} - y_{i,t-\nu-1}$ ($\nu = 1, \dots, p$) kelpaavat siis instrumenteiksi ϕ -parametreja estimoitaessa, kun taas β -parametrien osalta X_{it} -muuttujia voidaan käyttää omina instrumentteinaan.

Itse asiassa kaikki muutkin viivästeiset vastemuuttujan muutokset $\nabla y_{i,t-\nu}$, $\nu > p$ kelpaavat instrumenteiksi, mutta näiden instrumenttien laatu huononee nopeasti ν :n kasvaessa, kun korrelaatio $y_{i,t-p}$:n ja $\nabla y_{i,t-\nu}$:n välillä heikkenee.

Perinteinen (esimerkin 5.1 mukainen) instrumenttimatriisi näyttäisi siis (yksilöä i koskevalta osioltaan) seuraavalta:

$$Z_i = \begin{pmatrix} \nabla y_{i,2} & \dots & \nabla y_{i,p+1} & X'_{i,p+2} \\ \nabla y_{i,3} & \dots & \nabla y_{i,p+2} & X'_{i,p+3} \\ \vdots & \dots & \vdots & \vdots \\ \nabla y_{i,T_i-p} & \dots & \nabla y_{i,T_i-1} & X'_{i,T_i} \end{pmatrix} ,$$

mutta haluttaessa matriisin vasempaan laitaan voidaan toki ottaa enemmänkin $\nabla y_{i,T_i-\nu}$ -loppuisia sarakkeita ($\nu > p$). (Jokainen tällainen ylimääräinen sarake tosin pudottaa yhden havaintorivin pois Z_i -matriisin yläosasta.)

GMM-estimoinnin perusidea voidaan käyttää myös vapaammin (ks. Arellano, s. 111) muodostamalla jokaiselle havainnolle omat instrumenttinsa:

$$Z_i^{(AB)} = \begin{pmatrix} 1 & \nabla y_{i,2} & \dots & \nabla y_{i,p+1} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & X'_{i,p+2} \\ 0 & 0 & \dots & 0 & 1 & \nabla y_{i,3} & \dots & \nabla y_{i,p+2} & \dots & 0 & 0 & \dots & 0 & X'_{i,p+3} \\ \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & \nabla y_{i,T_i-p} & \dots & \nabla y_{i,T_i-1} & X'_{i,T_i} \end{pmatrix}.$$

(Itse asiassa jokaiselle $Z_i^{(AB)}$ -matriisin vaakariville voitaisiin sisällyttää *kaikki* aikaisemmat muutokset $\nabla y_{i,t-\nu}$, $\nu = 1, \dots, t-2$, mutta tällä tavalla saavutettavissa oleva hyöty on yleensä varsin kyseenalainen.)

Huomattakoon lopuksi, että u_i -tasotermien ARX-mallien yhteydessä aiheuttama ongelma voidaan kiertää myös toisella tavalla: Malliyhtälöstä (6.5) seuraa välittömästi, että

$$\begin{aligned} \nabla y_{it} &= \beta' \nabla X_{it} + \phi_1 \nabla y_{i,t-1} + \dots + \phi_p \nabla y_{i,t-p} + \varepsilon_{it} - \varepsilon_{i,t-1} \quad , \\ \varepsilon_{it} &\sim i.i.d.(0, \sigma^2) \quad , \quad \{\varepsilon_{it}\} \perp\!\!\!\perp \{\nabla X_{it}\} \quad (6.6) \\ t &= p+2, \dots, T_i \quad ; \quad i = 1, \dots, m \quad . \end{aligned}$$

Kannattaa erityisesti korostaa, että malleissa (6.5) ja (6.6) esiintyvät ϕ - ja β -parametrit ovat *samat*. Näiden parametrien estimointi voidaan siis perustaa myös muotoon (6.6), jota tosin vaivaa ongelma

$$\nabla y_{i,t-1} \not\perp \kappa_{it} = \varepsilon_{it} - \varepsilon_{i,t-1} \quad .$$

On kuitenkin helppo havaita, että kaikki muuttujat $y_{i,t-2}, \dots, y_{i1}$ kelpaavat instrumenteiksi, sillä

$$y_{i,t-\nu} \perp\!\!\!\perp \kappa_{it} = \varepsilon_{it} - \varepsilon_{i,t-1} \quad ,$$

mutta

$$y_{i,t-\nu} \not\perp \nabla y_{i,t-1}$$

kaikilla $\nu = 2, \dots, t-1$. Parametrit ϕ_1, \dots, ϕ_p ja β voidaan siis estimoida mallin (6.6) perusteella luvussa 5.3 esitellyllä GMM-menetelmällä käyttämällä alkuperäisiä *tasomuuttujia* $y_{i,t-\nu}$ ($\nu \geq 2$) instrumentteina. Huomattakoon, että kelvollisten instrumenttien lukumäärä vaihtelee *yksilöittäin*. Tässäkin tapauksessa on täysin mahdollista yrittää käyttää kaikkien kelvollisten instrumenttien sisältämä

informaatio estimoinnissa apuna, mutta toisaalta on ilmeistä, että muuttujien $\nabla y_{i,t-1}$ ja $y_{i,t-\nu}$ ($\nu \geq 2$) välinen riippuvuus heikkenee nopeasti ν :n kasvaessa. Tästä syystä tavallisesti tyydytäänkin tiettyyn *kiinteään* instrumenttimäärään $y_{i,t-2}, \dots, y_{i,t-\nu_0}$ ($\nu_0 \geq p+1$) ja niiden ”perinteiseen” käyttötapaan.

Huomautus 6.1: Huomattakoon, että X_{it} -selittäjävektorit voivat sisältää myös kaikille havaintoyksiköille yhteiset, ajankohdittaiset dummy-muuttujat (1/0 - indikaattorit). Tällä tavalla sovellettuna mallityyppi (6.5) sisältää myös tapauksen

$$y_{it} = \beta' X_{it} + \phi_1 y_{i,t-1} + \dots + \phi_{p_0} y_{i,t-p_0} + \lambda_t + u_i + \varepsilon_{it} \quad ,$$

jossa λ_t ($t = 1, \dots, T$) ovat kiinteitä parametreja. Malli (6.5) voi siis sisältää myös kaikille havaintoyksiköille yhteisen ajallisen trendin. ■

Huomautus 6.2: Vaikka edellä oletettiin, että $\{X_{it}\}$ -, $\{u_i\}$ - ja $\{\varepsilon_{it}\}$ -prosessit olisivat kaikki toisistaan riippumattomia, on itse asiassa mahdollista sallia joidenkin X_{it} -selittäjien ja u_i -tasotermien riippuvan toisistaan. Mikäli näin halutaan menetellä, voidaan kyseiset selittäjät tulkita ”endogeenisiksi” ja lisätä tämän tulokinnan mukaisesti muuttujien viivästetyt havainnot tai differenssit instrumenttilistaan aivan samoin kuin vastemuuttujan osalta menetellään. ■

Huomautus 6.3: Mikäli käytettyjen instrumenttien lukumäärä r on suurempi kuin estimoitavien kerroinparametrien lukumäärä $m+p$, voidaan ylimääräisten (”yli-identifioivien”) ortogonaalisuusehtojen (5.23) realistisuutta tutkia huomautuksessa 5.3 mainitun Hansenin testisuureen avulla. Mikäli käytetty painomatriisi vastaisi optimaalista painomatriisia, pitäisi testisuureen otantajakauden asympotoottisesti muistuttaa χ^2_{r-p-m} -jakaumaa. Tämä testi on siis suositeltavinta perustaa *iteratiiviseen* GMM-estimaattoriin, koska tällöin \hat{S}_n^{-1} toimii optimaalisen painomatriisin S^{-1} tarkentuvana estimaattorina kaksivaiheista vastinettaan tehokkaammin. ■

Huomautus 6.4: Mikäli yksilöiden lukumäärä on kustakin yksilöstä saatujen havaintojen lukumäärää pienempi ($n < T$), on \hat{S}_n -estimaattori luonnollisesti aina singulaarinen. Tästä syystä sekä iteratiivisen että kaksivaiheisen GMM-estimaattorin käyttö edellyttää, että $n > T$. ■

Huomautus 6.5: Kaikki edellä sanottu on perustunut keskeisesti oletukseen $\varepsilon_{it} \sim i.i.d.(0, \sigma^2)$. Mikäli $\{\varepsilon_{it}\}$ -prosessi olisi autokorreloitunut, ei kelvollisten instrumenttien muodostamisessa käytetty järjely tietenkään päde, eikä mikään edellä esitetystä pidä paikkaansa. Tästä syystä po. oletuksen realistisuutta kannattaa ehdottomasti tutkia, erityisesti differenssimuotoista mallia (6.6) estimoidessa. Apuvälineinä voidaan käyttää esimerkiksi sarjan

$$\nabla \hat{\omega}_{i,t} = \hat{\omega}_{i,t} - \hat{\omega}_{i,t-1} = \hat{\varepsilon}_{i,t} - \hat{\varepsilon}_{i,t-1} \quad t = 2, \dots, T_i$$

autokorrelaatioita $r_{\nabla\omega}(\nu)$ ($\nu = 1, 2, \dots$). Mikäli po. oletus olisi realistinen, pitäisi $r_{\nabla\omega}(1)$:n olla kohtuullisen lähellä -0.5 :ttä ja $r_{\nabla\omega}(\nu)$:n kohtuullisen lähellä 0 :aa kaikilla $\nu \geq 2$. ■

Esimerkki 6.2: Tarkastellaan Englannissa koottua, 140 yrityksestä koostuvaa paneeliaineistoa, joka syntyi, kun tiedot seuraavista muuttujista

N_{it} = yrityksen i työntekijämäärä vuoden t lopussa
 W_{it} = yrityksen i työntekijöiden keskimääräinen reaalipalkka vuonna t
 K_{it} = yrityksen i käyttöpääoman reaaliarvo vuonna t
 YS_{it} = yrityksen i toimialan tuotannon kokonaismäärä (vol.) vuonna t

kirjattiin vuosilta 1976 - 1984.

Pyritään kuvaamaan työn kysynnän kehitystä yrityksittäin ARX-mallilla

$$n_{it} = \mu + u_i + \phi n_{i,t-1} + \beta_1 w_{it} + \beta_2 w_{i,t-1} + \beta_3 k_{it} + \beta_4 k_{i,t-1} + \beta_5 y_{sit} + \beta_6 y_{s_{i,t-1}} + \lambda_t + \varepsilon_{it}$$

$$u_i \sim i.i.d.(0, \sigma^2) \quad , \quad \varepsilon_{it} \sim i.i.d.(0, \tau^2) \quad , \quad u_i \perp \{ \varepsilon_{it} \} \quad , \\ t = 1, \dots, T \quad ; \quad i = 1, \dots, m \quad ,$$

jossa pienet kirjaimet tarkoittavat logaritmoituja muuttujia ($n_{it} = \log N_{it}$ jne.). Instrumentteja valittaessa on tulkittu w_{it} ja k_{it} ”endogeenisiksi” muuttujiksi, jotka saattavat korreloida sekä u_i :n että ε_{it} :n kanssa. Sen sijaan y_{sit} on tulkittu puhtaasti eksogeeniseksi ja niinpä sitä onkin käytetty ”omana instrumenttinaan”.

SAS:

```
axis1 label=(h=1.4 f=simplex) value=(h=1.4 f=simplex) minor=none;
data labeng; infile 'c:\mr\long05\labeng.dat';
input branch year EMP WAGE CAPIT OUTP ;
retain firm 1;
if lag(year)>year then firm=firm+1;
n=log(EMP); w=log(WAGE); k=log(CAPIT); ys=log(OUTP);
n1=lag(n); if lag(year)>year then n1=missing;
w1=lag(w); if lag(year)>year then w1=missing;
k1=lag(k); if lag(year)>year then k1=missing;
ys1=lag(ys); if lag(year)>year then ys1=missing;
n2=lag(n1); if lag(year)>year then n2=missing;
w2=lag(w1); if lag(year)>year then w2=missing;
k2=lag(k1); if lag(year)>year then k2=missing;
n3=lag(n2); if lag(year)>year then n3=missing;
w3=lag(w2); if lag(year)>year then w3=missing;
k3=lag(k2); if lag(year)>year then k3=missing;
n4=lag(n3); if lag(year)>year then n4=missing;
w4=lag(w3); if lag(year)>year then w4=missing;
k4=lag(k3); if lag(year)>year then k4=missing;
drop missing; run;
```

```

data labeng; set labeng;
  dn=n-lag(n); if lag(year)>year then dn=missing;
  dn1=lag(dn); if lag(year)>year then dn1=missing;
  dn2=lag(dn1); if lag(year)>year then dn2=missing;
  dn3=lag(dn2); if lag(year)>year then dn3=missing;
  dn4=lag(dn3); if lag(year)>year then dn4=missing;
  dw=w-lag(w); if lag(year)>year then dw=missing;
  dw1=lag(dw); if lag(year)>year then dw1=missing;
  dw2=lag(dw1); if lag(year)>year then dw2=missing;
  dw3=lag(dw2); if lag(year)>year then dw3=missing;
  dw4=lag(dw3); if lag(year)>year then dw4=missing;
  dk=k-lag(k); if lag(year)>year then dk=missing;
  dk1=lag(dk); if lag(year)>year then dk1=missing;
  dk2=lag(dk1); if lag(year)>year then dk2=missing;
  dk3=lag(dk2); if lag(year)>year then dk3=missing;
  dk4=lag(dk3); if lag(year)>year then dk4=missing;
  dys=ys-lag(ys); if lag(year)>year then dys=missing;
  dys1=lag(dys); if lag(year)>year then dys1=missing;
  v1981=0; if year=1981 then v1981=1;
  v1982=0; if year=1982 then v1982=1;
  v1983=0; if year=1983 then v1983=1;
  v1984=0; if year=1984 then v1984=1;
  dn1v81=dn1*v1981; dw1v81=dw1*v1981; dk1v81=dk1*v1981;
  dn1v82=dn1*v1982; dw1v82=dw1*v1982; dk1v82=dk1*v1982;
  dn1v83=dn1*v1983; dw1v83=dw1*v1983; dk1v83=dk1*v1983;
  dn1v84=dn1*v1984; dw1v84=dw1*v1984; dk1v84=dk1*v1984;
  dn2v81=dn2*v1981; dw2v81=dw2*v1981; dk2v81=dk2*v1981;
  dn2v82=dn2*v1982; dw2v82=dw2*v1982; dk2v82=dk2*v1982;
  dn2v83=dn2*v1983; dw2v83=dw2*v1983; dk2v83=dk2*v1983;
  dn2v84=dn2*v1984; dw2v84=dw2*v1984; dk2v84=dk2*v1984;
  dn3v81=dn3*v1981; dw3v81=dw3*v1981; dk3v81=dk3*v1981;
  dn3v82=dn3*v1982; dw3v82=dw3*v1982; dk3v82=dk3*v1982;
  dn3v83=dn3*v1983; dw3v83=dw3*v1983; dk3v83=dk3*v1983;
  dn3v84=dn3*v1984; dw3v84=dw3*v1984; dk3v84=dk3*v1984;
  dn4v81=dn4*v1981; dw4v81=dw4*v1981; dk4v81=dk4*v1981;
  dn4v82=dn4*v1982; dw4v82=dw4*v1982; dk4v82=dk4*v1982;
  dn4v83=dn4*v1983; dw4v83=dw4*v1983; dk4v83=dk4*v1983;
  dn4v84=dn4*v1984; dw4v84=dw4*v1984; dk4v84=dk4*v1984; run;
proc model data=labeng;
  parms mu phi b1-b6 lambda2-lambda4 ;
  n = mu + phi*n1 + b1*w + b2*w1 + b3*k + b4*k1 + b5*ys + b6*ys1
      + lambda2*v1982 + lambda3*v1983 + lambda4*v1984 ;
  exogenous ys ys1 v1982-v1984 ;
  instruments dn1 dn2 dn3 dn4 dw1 dw2 dw3 dw4 dk1 dk2 dk3 dk4 _exog_;
  fit n / gmm ;
  *fit n / itgmm kernel=(bart,0,0) ; run;

```

The MODEL Procedure

Model Summary

Model Variables	6
Exogenous	5
Parameters	11
Equations	1
Number of Statements	1

Model Variables n ys ys1 v1982 v1983 v1984
 Parameters mu phi b1 b2 b3 b4 b5 b6 lambda2 lambda3 lambda4

The Equation to Estimate is

$$n = F(\mu(1), \phi(n1), b1(w), b2(w1), b3(k), b4(k1), b5(ys), b6(ys1), \lambda2(v1982), \lambda3(v1983), \lambda4(v1984))$$

Instruments 1 dn1 dn2 dn3 dn4 dw1 dw2 dw3 dw4 dk1 dk2 dk3 dk4 ys ys1 v1982 v1983 v1984

GMM Estimation Summary

Minimization Summary

Parameters Estimated	11
Iterations	1

Nonlinear GMM Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t
mu	3.327689	2.6520	1.25	0.2105
phi	0.875218	0.0865	10.12	<.0001
b1	-0.72595	0.3156	-2.30	0.0221
b2	0.782423	0.2234	3.50	0.0005
b3	1.132141	0.2797	4.05	<.0001
b4	-1.05765	0.2837	-3.73	0.0002
b5	-0.70492	0.4018	-1.75	0.0803
b6	-0.02495	0.4085	-0.06	0.9513
lambda2	0.040947	0.0364	1.13	0.2608
lambda3	0.031065	0.0494	0.63	0.5298
lambda4	0.017799	0.0606	0.29	0.7691

Number of Observations Statistics for System

Used	331	Objective	0.009523
Missing	700	Objective*N	3.1521

Koska käytettyjä instrumentteja oli yhteensä 18 kappaletta ja estimoitavia kerroinparametreja oli 11 kappaletta, voidaan Hansenin testisuureta 3.1521 verrata

(asymptoottisessa hengessä) χ^2_7 -jakaumaan. Koska kyseisen jakauman *odotusarvo* on 7, on havaittua Hansenin testisuureen arvoa pidettävä varsin pienenä, ja estimoinnin apuna käytettyjen ylimääräisten instrumenttien osalta ortogonaalisuusoletukset näyttävät siis hyvin realistisilta. Yhteiseen aikafaktoriin varautuvat λ_t -termit näyttävät melko tarpeettomilta. AR-parametrin ϕ estimaatti näyttää melko suurelta, joten työntekijämäärien sopeutuminen selittävässä tekijöissä tapahtuviin muutoksiin näyttää tapahtuvan melko hitaasti.

Mikäli halutaan käyttää havaintokohtaisiin instrumentteihin perustuvia $\hat{\phi}^{(AB)}$ - ja $\hat{\beta}^{(AB)}$ -estimaattoreita, voidaan käyttää seuraavaa ohjelmakoodia:

```
proc model data=labeng;
  parms mu phi b1-b6 lambda2-lambda4 ;
  n = mu + phi*n1 + b1*w + b2*w1 + b3*k + b4*k1 + b5*ys + b6*ys1
      + lambda2*v1982 + lambda3*v1983 + lambda4*v1984 ;
  exogenous ys ys1 v1982-v1984 ;
  instruments dn1v81 dn1v82 dn1v83 dn1v84
              dn2v81 dn2v82 dn2v83 dn2v84
              dn3v81 dn3v82 dn3v83 dn3v84
              dn4v81 dn4v82 dn4v83 dn4v84
              dw1v81 dw1v82 dw1v83 dw1v84
              dw2v81 dw2v82 dw2v83 dw2v84
              dw3v81 dw3v82 dw3v83 dw3v84
              dw4v81 dw4v82 dw4v83 dw4v84
              dk1v81 dk1v82 dk1v83 dk1v84
              dk2v81 dk2v82 dk2v83 dk2v84
              dk3v81 dk3v82 dk3v83 dk3v84
              dk4v81 dk4v82 dk4v83 dk4v84
              v1981 _exog_ ;
  fit n / gmm ;
  *fit n / itgmm kernel=(bart,0,0) ; run;
```

Differenssimalliin (6.6) perustuen samat kerroinparametrit voitaisiin estimoida seuraavasti:

```
proc model data=labeng;
  parms phi b1-b6 lambda1-lambda4 ;
  dn = phi*dn1 + b1*dw + b2*dw1 + b3*dk + b4*dk1 + b5*dys + b6*dys1;
  exogenous dys dys1;
  instruments n2 n3 n4 w2 w3 w4 k2 k3 k4 v1982 v1983 v1984 _exog_ ;
  *fit dn / itgmm kernel=(bart,3,0); ** Newey-West-tyyppinen sandwich-estim. **;
  fit dn / itgmm kernel=(parzen,0,0) ;
  solve dn/data=labeng out=sov outpredict; id firm year; run;
  data sov; set sov; dnsov=dn; keep firm year dnsov; run;
  data yhd; merge labeng sov; by firm year; resi=dn-dnsov; run;
  proc arima data=yhd; identify var=resi outcov=acf; run;
  data acf; set acf; yla=2*stderr; ala=-2*stderr; if lag<4; run;
  proc gplot data=acf; plot (corr yla ala)*lag/overlay haxis=axis1 vaxis=axis1;
  symbol1 i=needle r=1 c=black w=2;
  symbol2 i=join r=2 c=red l=3 w=3; run;
```

The MODEL Procedure

Model Variables dn dys dys1

Parameters phi b1 b2 b3 b4 b5 b6 lambda1 lambda2 lambda3 lambda4

The Equation to Estimate is

$$dn = F(\text{phi}(dn1), b1(dw), b2(dw1), b3(dk), b4(dk1), b5(dys), b6(dys1))$$

Instruments 1 n2 n3 n4 w2 w3 w4 k2 k3 k4 v1982 v1983 v1984 dys dys1

NOTE: At ITGMM Iteration 5 CONVERGE=0.001 Criteria Met.

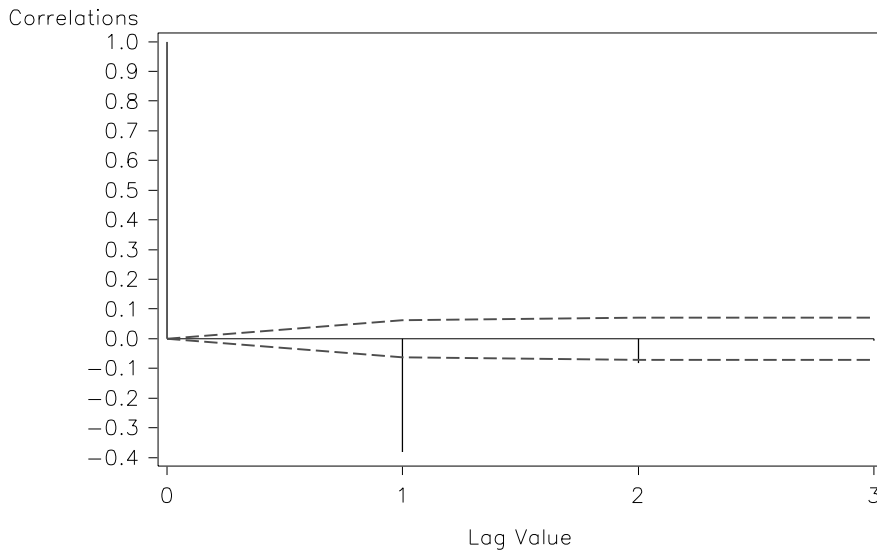
Parameters Estimated	7
Kernel Used	PARZEN
l(n)	0
Method	Gauss
Iterations	5

Nonlinear ITGMM Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t
phi	0.772324	0.1939	3.98	<.0001
b1	-0.88609	0.3570	-2.48	0.0134
b2	0.293637	0.2287	1.28	0.1999
b3	0.429839	0.2013	2.14	0.0333
b4	-0.36088	0.2276	-1.59	0.1136
b5	0.742216	0.1875	3.96	<.0001
b6	-0.61658	0.3242	-1.90	0.0578

Number of Observations	Statistics for System
Used 471	Objective 0.0165
Missing 560	Objective*N 7.7866

Jännöstermien autokorrelaatiot



7 MONIYHTÄLÖMALLIT

7.1 SURE-estimointi

Tarkastellaan nyt samanaikaisesti K :ta näennäisesti toisistaan erillistä regressiomallia

$$y_{kt} = X'_{kt}\beta_{(k)} + \varepsilon_{kt} \quad , \quad t = 1, \dots, n \quad , \quad k = 1, \dots, K \quad , \quad (7.1)$$

joissa X_{kt} :t ($X_{kt} \in \mathcal{R}^{m_k}$) ovat eksogeenisiä selittäjävektoreita. Ajatellaan, että kaikista muuttujista on saatu havainnot *samoilta ajankohdilta* $t = 1, \dots, n$. (Eri malleissa voi olla osittain samojaakin selittäjiä.) Ajatellaan, että eri malleihin liittyvät *samanaikaiset* virhetermit ε_{kt} voisivat korreloida keskenään, mutta että eri ajankohtiin liittyvät virhetermit olisivat riippumattomia toisistaan:

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Kt} \end{pmatrix} \sim i.i.d. \quad , \quad \text{cov}(\varepsilon_t) = \Sigma = (\sigma_{ij}) \quad . \quad (7.2)$$

Merkinnällisesti koko malli (7.1) + (7.2) voidaan ajatella yhdistettäväksi joko muotoon

$$Y = X\beta + \varepsilon \quad , \quad E\varepsilon = 0 \quad , \quad \text{cov}(\varepsilon) = V \quad (7.3)$$

jossa

$$Y_{(k)} = (y_{k1} \quad \dots \quad y_{kn})' \quad , \quad k = 1, \dots, K \quad ,$$

$$Y = (Y'_{(1)} \quad \dots \quad Y'_{(K)})' \in \mathcal{R}^{Kn} \quad ,$$

$$\beta = (\beta'_{(1)} \quad \dots \quad \beta'_{(K)})' \in \mathcal{R}^M \quad , \quad M = m_1 + \dots + m_K \quad ,$$

$$X = \begin{pmatrix} X_{(1)} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & X_{(K)} \end{pmatrix} \quad \text{ja}$$

$$V = \Sigma \otimes I = \begin{pmatrix} \sigma_{11}I & \dots & \sigma_{1K}I \\ \vdots & \dots & \vdots \\ \vdots & \dots & \vdots \\ \sigma_{K1}I & \dots & \sigma_{KK}I \end{pmatrix} \quad ,$$

tai muotoon

$$Y^* = X^*\beta + \varepsilon^* \quad , \quad E\varepsilon^* = 0 \quad , \quad \text{cov}(\varepsilon^*) = V^* \quad (7.4)$$

jossa

$$Y^* = (y_{11} \quad \dots \quad y_{K1} \quad \dots \quad y_{1n} \quad \dots \quad y_{Kn})' \quad ,$$

$$X_t = \begin{pmatrix} X'_{1t} & 0 & \dots & 0 \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & X'_{Kt} \end{pmatrix} \quad , \quad X^* = \begin{pmatrix} X_1 \\ \cdot \\ \cdot \\ X_n \end{pmatrix} \quad ,$$

ja

$$V^* = I \otimes \Sigma = \begin{pmatrix} \Sigma & 0 & \dots & 0 \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & \Sigma \end{pmatrix} \quad .$$

Halu käsitellä yhtälöitä yhdessä voi juontaa juurensa

A) siitä, että selitettävien tekijöiden y_{kt} ($k = 1, \dots, K$) on yhdessä toteutettava joitakin ehtoja, kuten esim. budjettirajoitukset kulutusmalleissa tai vastaavat rajoitukset rahoitusvirtamalleissa.

tai

B) siitä, että eri yhtälöiden virhetermeihin ajatellaan sisältyvän yhteisiä piirteitä, jolloin kaikkiin yhtälöihin liittyvän havaintomateriaalin sisältämä lisäinformaatio kannattaa käyttää estimoinnissa hyväksi ("seemingly unrelated regressions").

Tapauksessa A) voidaan lineaariset rajoitukset $R\beta = \gamma$ ottaa huomioon aivan tavanomaiseen tapaan käyttämällä rajoitettua GLS-estimaattoria

$$\hat{\beta}_R = \hat{\beta}_{GLS} - (X^{*'}V^{*-1}X^*)^{-1} R'[R(X^{*'}V^{*-1}X^*)^{-1}R']^{-1} (R\hat{\beta}_{GLS} - \gamma) \quad (7.5)$$

Tapauksessa B) taas on helppo nähdä (esim. muotoilua (7.3) käyttäen), että

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (7.6)$$

on β :n lineaarisista estimaattoreista tehokkain ja asympotoottisesti normaalin, kunhan

$$p \lim_{n \rightarrow \infty} \frac{1}{n} X'V^{-1}X = Q \quad (7.7)$$

on hyvin määritelty ja positiivisesti definiitti.

Vaikka $\hat{\beta}_{OLS}$ ei olekaan yhtä tehokas kuin $\hat{\beta}_{GLS}$, on se ainakin harhaton ja tarkentuva. Tästä johtuu, että luvun 2.4 tulosten mukaan seuraavalla kaksivaiheisella estimaattorilla on sama asympotoottinen jakauma kuin GLS-estimaattorilla (7.6):

$$\begin{aligned}
\hat{\beta}_{SUR} &= (X'\hat{V}^{-1}X)^{-1} X'\hat{V}^{-1}Y \quad \text{jossa} \\
\hat{V} &= (\hat{\sigma}_{ij}) \otimes I \quad , \\
\hat{\sigma}_{ij} &= \frac{1}{n} E'_{(i)} E_{(j)} \quad , \quad i, j = 1, \dots, K \quad \text{ja} \\
E &= (E'_{(1)} \quad \dots \quad E'_{(K)})' = (I - P_X)Y
\end{aligned} \tag{7.8}$$

(Tarpeen vaatiessa iteraatiota voidaan tietenkin jatkaa ja antaa \hat{V} :n ja $\hat{\beta}_{SUR}$:n vaiheittain sopeutua toisiinsa. Normaalisten virhetermien tapauksessa näin päädyttäisiin lopulta β :n ML-estimaattoriin.)

Nähdään siis, että

$$\sqrt{n} (\hat{\beta}_{SUR} - \beta) \stackrel{\text{asympt.}}{\sim} N_M(0, Q^{-1}) \tag{7.9}$$

jossa Q on määritelty kaavalla (7.7).

Karkeasti voidaan sanoa, että $\hat{\beta}_{SUR}$:n suhteellinen asymptoottinen tehokkuus (ARE) $\hat{\beta}_{OLS}$:iin verrattuna kasvaa K :n kasvaessa ja toisaalta Σ :aan liittyvien korrelaatioiden kasvaessa. Toisaalta, $\hat{\beta}_{OLS}$ saattaa olla jopa tehokkaampi kuin $\hat{\beta}_{SUR}$, jos Σ on lähellä diagonaalimatriisia. (Jos Σ diagonaalinen, on $\hat{\beta}_{OLS}$ tietysti tehokkain lineaarinen estimaattori.)

Huomautus 7.1: Jos kaikissa malleissa on samat selittäjät, ei SUR-estimoinnista ole mitään hyötyä. Tämä nähdään seuraavasti:

Jos $X_{(1)} = \dots = X_{(K)} = \bar{X}$, on $X = I \otimes \bar{X}$ ja

$$X'\hat{V}^{-1}X = (I \otimes \bar{X})'(\hat{\Sigma}^{-1} \otimes I)(I \otimes \bar{X}) = \hat{\Sigma}^{-1} \otimes (\bar{X}'\bar{X}) \quad ,$$

joten

$$\begin{aligned}
\hat{\beta}_{SUR} &= (\hat{\Sigma} \otimes (\bar{X}'\bar{X})^{-1})(I \otimes \bar{X}')(\hat{\Sigma}^{-1} \otimes I)Y \\
&= (I \otimes (\bar{X}'\bar{X})^{-1}\bar{X}')Y = \hat{\beta}_{OLS} \quad .
\end{aligned}$$

■

Huomautus 7.2: Palataan vielä hetkeksi tapaukseen A) . Jos esimerkiksi ”budjettirajoitukset” implikoivat ehdon $\sum_{k=1}^K \varepsilon_{kt} = 0$ kaikilla $t = 1, \dots, n$, on virhetermivektorin ε_t jakauma siis *singulaarinen*. Tästä kiusallisesta ominaisuudesta päästään yleensä kätevimmin eroon eliminoimalla yksi mallin yhtälöistä. Parametrirajoitusten huomioon ottamisen tarve katoaa samalla, mutta on selvää, että jäljelle jäävien yhtälöiden virhetermien välisiin korrelaatioihin on ehdottomasti varauduttava. Tämä tapahtuu parhaiten SUR-estimointia käyttämällä. ■

7.2 Simultaanimallit

7.2.1 Parametrien identifioituvuus

Tarkastellaan nyt vastemuuttujien (endogeenisten muuttujien) y_{kt} ($k = 1, \dots, K$) välisten *vuorovaikutusten* kuvaamista. Aihetta pohjustettiin jo johdantoluvussa 1.3.

Merkitään

$$Y_t = (y_{1t} \dots y_{Kt})' \quad \text{ja} \quad Y = (Y_1 \dots Y_n)'$$

jossa Y on siis $n \times K$ - matriisi. ”Selittäjämatrisin” X ($n \times m$) ajatellaan puolestaan sisältävän eksogeenisten tekijöiden lisäksi myös (mahdollisesti) endogeenisten muuttujien viivästettyjä arvoja. Tarkastellaan nyt moniyhtälömallia

$$Y \quad \Gamma \quad + \quad X \quad \Delta \quad = \quad \varepsilon$$

$$(n \times K) \quad (K \times K) \quad (n \times m) \quad (m \times K) \quad (n \times K)$$

jossa $\varepsilon = (\varepsilon_1 \dots \varepsilon_n)'$ ja $\varepsilon_i \perp \varepsilon_j$ kun $i \neq j$, (7.10)

sekä $\varepsilon \perp X$, $E\varepsilon_t = 0$, $\text{cov}(\varepsilon_t) = \Sigma$ $t = 1, \dots, n$.

Matriisin Γ diagonaalelementit oletetaan ykkösiksi ja Γ oletetaan epäsingulaarisiksi. Tällöin (7.10) muodostaa täydellisen ”systeemikuvauksen” endogeenisten muuttujien Y_t kehitysmekanismista. Mikäli $\Gamma \neq I$, esiintyy endogeenisten muuttujien kesken samanaikaisia vaikutuskytkentöjä, ja tällöin mallia (7.10) sanotaan *simultaanimalliksi*. Tällaisiin malleihin liittyvistä identifioituvuusongelmista oli puhetta jo luvussa 1.3, jolloin kävi ilmi, että ongelmat liittyvät nimenomaan Γ -parametrien estimointiin. On selvää, että Γ -parametrien identifioituvuuden takaamiseksi tarvitaan riittävä määrä ”ennalta määrättyjä” X -muuttujia, jotka heijastuvat kyllin eri tavoin eri endogeenisiin tekijöihin Y .

Esimerkki 7.1: Lawrence Klein rakensi aikoinaan seuraavan kaltaisen (todennäköisesti maailman ensimmäisen, eksplisiittisesti muotoa (7.10) olevan) simulaationmallin kuvaamaan Yhdysvaltojen kansantalouden perustoimintoja:

Malliin sisältyvät seuraavat endogeeniset muuttujat

c_t	\longleftrightarrow	yksityisen sektorin kulutus
p_t	\longleftrightarrow	yksityisen sektorin tuotot
w_t	\longleftrightarrow	yksityisen sektorin palkkasumma
i_t	\longleftrightarrow	investoinnit
x_t	\longleftrightarrow	tuotanto
$wtot_t$	\longleftrightarrow	kokonaispalkkasumma
k_t	\longleftrightarrow	pääomavaranto
y_t	\longleftrightarrow	kansantulo

ja seuraavat eksogeeniset muuttujat

wp_t	\longleftrightarrow	julkisen sektorin palkkasumma
gd_t	\longleftrightarrow	julkisen sektorin kysyntä
tax_t	\longleftrightarrow	verot .

Itse mallin Klein muotoili seuraavasti:

$$\begin{cases} c_t = \mu_1 + \gamma_{21}p_t + \gamma_{61}wtot_t + \delta_{41}p_{t-1} + \varepsilon_{1t} \\ i_t = \mu_2 + \gamma_{22}p_t + \delta_{42}p_{t-1} + \delta_{52}k_{t-1} + \varepsilon_{2t} \\ w_t = \mu_3 + \gamma_{53}x_t + \delta_{63}x_{t-1} + \delta_{73} \cdot t + \varepsilon_{3t} \end{cases}$$

Toiminnallisia rakennytälöitä tarvitaan vain kolme, sillä endogeenisia muuttujia sitovat toisiinsa seuraavat määritelmälliset identiteetit:

$$\begin{aligned} x_t &\equiv c_t + i_t + gd_t \\ y_t &\equiv c_t + i_t + gd_t - tax_t \\ p_t &\equiv y_t - wtot_t \\ k_t &\equiv k_{t-1} + i_t \\ wtot_t &\equiv w_t + wp_t \end{aligned}$$

Rakennytälöihin sisältyvien satunnaistermien luonteesta tehdään seuraavat oletukset:

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} \sim NID_3(0, \Sigma) \quad .$$

■

Simultaanimallin *rakennemuoto* (7.10) voidaan ratkaista Y :n suhteen, jolloin päädytään ns. *redusoituun muotoon*

$$\begin{aligned}
 Y &= X\Pi + V \quad , \quad \text{jossa} \\
 \Pi &= -\Delta\Gamma^{-1} \quad \text{ja} \\
 V &= (V_1 \quad \dots \quad V_n)' = \varepsilon\Gamma^{-1} \quad , \\
 \Omega &= \text{cov}(V_t) = \Gamma^{-1}'\Sigma\Gamma^{-1} \quad .
 \end{aligned}
 \tag{7.11}$$

Lisäksi $X \perp\!\!\!\perp V$, joten Π -parametrien OLS-estimaatit $\hat{\Pi}$ ovat tarkentuvia ja

$$\begin{aligned}
 \sqrt{n} \text{vec}(\hat{\Pi} - \Pi) &\stackrel{\text{asympt.}}{\sim} N_{mK}(0, \Omega \otimes Q^{-1}) \\
 \text{jossa} \quad Q &= \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} X'X \quad .
 \end{aligned}
 \tag{7.12}$$

Määritellään nyt mallin (7.10) parametrien identifioituvuus hieman yksinkertaistaen:

Määritelmä 7.1: Muodossa (7.10) esiintyviä ns. *rakennemuodon* parametreja Γ, Δ sanotaan *identifioituviksi*, jos ne voidaan johtaa redusoidun muodon parametreista Π yhteyden $\Pi = -\Delta\Gamma^{-1}$ perusteella. Rakennemuodon yksittäistä yhtälöä sanotaan identifioituvaksi, jos sen kaikki parametrit ovat identifioituvia. ■

Merkitään nyt

$$B = \begin{pmatrix} \Gamma \\ \Delta \end{pmatrix} = (\beta_{.1} \quad \dots \quad \beta_{.K})$$

ja ajatellaan, että $\beta_{.1}$ täyttää r ehtoa $\Phi\beta_{.1} = 0$, jotka määräävät, mitkä eksogeeniset ja mitkä endogeeniset muuttujat todella ovat mukana mallin (7.10) ensimmäisessä yhtälössä.

Lause 7.1: Rakennemuodon (7.10) 1. yhtälö on identifioituva jos ja vain jos

$$\text{rank}(\Phi B) = K - 1 \quad ,$$

ts. joss

$$\text{rank} \begin{pmatrix} \Pi & I \\ \Phi & \end{pmatrix} = K + m - 1 \quad .$$

Todistus: Matriisissa

$$A = \begin{pmatrix} \Pi & I \\ \Phi & \end{pmatrix}$$

on $K+m$ pystyriviä, joten $\text{rank}(A) \leq K+m$. Yhtälöllä $A\beta = 0$ saa olla vain yksi ratkaisu $\beta = \beta_{.1}$, jotta 1. yhtälö olisi identifioituva. Tämä merkitsee juuri ehtoa $\text{rank}(A) = K+m-1$. Koska toisaalta $\text{rank}(B) = K$ (muutenhan systeemissä olisi redundanssia), voidaan todeta, että

$$\text{rank}(A) = K+m-1 \quad \implies \quad \text{rank}(AB) \geq K-1 \quad .$$

(A :n ytimeen voi kuulua vain yksi vektori.)

Toisaalta

$$AB = \begin{pmatrix} 0 \\ \Phi B \end{pmatrix} \quad \text{ja} \quad \Phi \beta_{.1} = 0 \quad ,$$

joten

$$\text{rank}(\Phi B) = K-1 \quad \iff \quad \text{rank}(A) = K+m-1 \quad .$$

■

Huomautus 7.3: Yhtälö $\text{rank}(\Phi B) = K-1$ voi toteutua vain, jos $r \geq K-1$. (Symboli r tarkoittaa 1. yhtälöstä puuttuvien selittäjien lukumäärää.) ■

Määritelmä 7.2: Ensimmäistä yhtälöä sanotaan

- *ali-identifioituksi*, jos $\text{rank}(\Phi B) < K-1$
- *täsmälleen identifioituksi*, jos $\text{rank}(\Phi B) = K-1 = \text{rank}(\Phi) = r$
- *yli-identifioituksi*, jos $\text{rank}(\Phi B) = K-1 < \text{rank}(\Phi) = r$.

Vastaavia määritelmiä voidaan luonnollisesti soveltaa jokaiseen yhtälöön vuorolaan. ■

7.2.2 Rakennemuodon parametrien estimointi

Ajatellaan, että muuttuja y_{2t} olisi ”selittäjänä” rakennemuodon (7.10) 1. yhtälössä ja y_{1t} vuorostaan 2. yhtälössä. Tällöin on päivänselvää, että $E y_{1t} \varepsilon_{2t} \neq 0$ ja $E y_{2t} \varepsilon_{1t} \neq 0$. Oletuksen (1.8) rikkoutuminen tuo (ks. 5. luku) automaattisesti OLS-estimaattoreihin itsepintaista harhaa, joka ei poistu edes asymptoottisesti. Seuraavassa esitellään eräitä suositeltavampia estimointimenetelmiä:

7.2.2.1 Kaksivaiheinen pns-menetelmä

Tarkastellaan esimerkkinä mallin (7.10) 1. rakenneyhtälöä

$$(Y_{(1)} \quad Y_{\cdot 1}) \begin{pmatrix} 1 \\ \gamma_{\cdot 1} \end{pmatrix} + X_{\cdot 1} \delta_{\cdot 1} = \varepsilon_{\cdot 1} \quad (7.13)$$

jossa $Y_{(1)} = (y_{11} \quad \dots \quad y_{1n})'$ jne.

Matriisi $X_{\cdot 1}$ sisältää yhtälössä mukana olevat eksogeeniset (ja muut ”ennalta määräytyt”) selittäjät, $Y_{\cdot 1}$ endogeeniset selittäjät ja $V_{\cdot 1}$ vastaavat redusoidun muodon jäännöstermit muodosta (7.11). Merkitään

$$Z_{\cdot 1} = (Y_{\cdot 1} \quad X_{\cdot 1}) \quad .$$

Koska $\varepsilon \perp\!\!\!\perp X$, on

$$\varepsilon_{\cdot 1} \perp\!\!\!\perp E(Y_{\cdot 1} | X) = Y_{\cdot 1} - V_{\cdot 1} \quad . \quad (7.14)$$

Jos malli (7.13) kirjoitettaisiin muotoon

$$Y_{(1)} = -E(Y_{\cdot 1} | X) \gamma_{\cdot 1} - X_{\cdot 1} \delta_{\cdot 1} + (\varepsilon_{\cdot 1} - V_{\cdot 1} \gamma_{\cdot 1}) \quad , \quad (7.13')$$

olisi suluissa oleva jäännöstermi riippumaton yhtälössä esiintyvistä selittäjistä, joten $\gamma_{\cdot 1}$ - ja $\delta_{\cdot 1}$ -parametrien estimointi onnistuisi tarkentuvasti OLS:illa. Vaikeutena on vain se, ettei ehdollisia odotusarvoja $E(Y_{\cdot 1} | X)$ tunneta. Ne voidaan kuitenkin estimoida redusoidun muodon (7.11) avulla OLS:ia käyttäen:

$$\widehat{Y}_{\cdot 1} = E(\widehat{Y}_{\cdot 1} | X) = (X\Pi)_{\cdot 1} = Y_{\cdot 1} - \widehat{V}_{\cdot 1} \quad . \quad (7.15)$$

Sijoittamalla tämä $E(Y_{\cdot 1} | X)$:n paikalle yhtälössä (7.13') ja estimoimalla rakenneparametrit sitten OLS:illa päästään ns. *kaksivaiheiseen* pns-estimaattoriin (2SLS)

$$\tilde{\beta}_{\cdot 1}^o = - \begin{pmatrix} \tilde{\gamma}_{\cdot 1} \\ \tilde{\delta}_{\cdot 1} \end{pmatrix} = \begin{pmatrix} \widehat{Y}'_{\cdot 1} \widehat{Y}_{\cdot 1} & \widehat{Y}'_{\cdot 1} X_{\cdot 1} \\ X'_{\cdot 1} \widehat{Y}_{\cdot 1} & X'_{\cdot 1} X_{\cdot 1} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{Y}'_{\cdot 1} Y_{(1)} \\ X'_{\cdot 1} Y_{(1)} \end{pmatrix} \quad . \quad (7.16)$$

Koska toisaalta

$$\begin{aligned}\widehat{Y}'_1 \widehat{Y}_1 &= Y'_1 P_X P_X Y_1 \\ &= Y_1 P_X Y_1 = \widehat{Y}'_1 Y_1 \quad ,\end{aligned}$$

$$\begin{aligned}X'_1 \widehat{Y}_1 &= X'_1 (Y_1 - \widehat{V}_1) \\ &= X'_1 Y_1 \quad , \\ (\widehat{V}_1 &= (I - P_X)Y_1 \perp X)\end{aligned}$$

voidaan 2SLS-estimaattori tulkita myöskin instrumentteihin $\widehat{Z}_1 = (\widehat{Y}_1 \quad X_1)$ liittyvänä instrumenttiestaattorina, sillä

$$\begin{aligned}\tilde{\beta}_{\cdot 1}^o &= \begin{pmatrix} \widehat{Y}'_1 \widehat{Y}_1 & \widehat{Y}'_1 X_1 \\ X'_1 \widehat{Y}_1 & X'_1 X_1 \end{pmatrix}^{-1} \begin{pmatrix} \widehat{Y}'_1 Y_{(1)} \\ X'_1 Y_{(1)} \end{pmatrix} \\ &= \begin{pmatrix} \widehat{Y}'_1 Y_1 & \widehat{Y}'_1 X_1 \\ X'_1 Y_1 & X'_1 X_1 \end{pmatrix}^{-1} \begin{pmatrix} \widehat{Y}'_1 Y_{(1)} \\ X'_1 Y_{(1)} \end{pmatrix} \quad (7.17) \\ &= (\widehat{Z}'_1 Z_1)^{-1} \widehat{Z}'_1 Y_{(1)} \quad .\end{aligned}$$

Kaksivaiheisen estimaattorin $\tilde{\beta}_{\cdot 1}^o$ harha on muotoa

$$E(\tilde{\beta}_{\cdot 1}^o) - \beta_{\cdot 1}^o = E \left[(\widehat{Z}'_1 \widehat{Z}_1)^{-1} \widehat{Z}'_1 \varepsilon_1 \right] \quad ,$$

joka ei häviä, koska \widehat{Y}_1 ja ε_1 korreloivat keskenään. Harha kuitenkin poistuu asymptoottisesti havaintomäärän kasvaessa, joten $\tilde{\beta}_{\cdot 1}^o$ on tarkentuva.

Lisäksi

$$\begin{aligned}\sqrt{n} (\tilde{\beta}_{\cdot 1}^o - \beta_{\cdot 1}^o) &\stackrel{\text{asympt.}}{\sim} N_{K+m-r}(0, \sigma_{11} Q_1^{-1}) \\ \text{jossa} \quad Q_1 &= \text{p} \lim_{n \rightarrow \infty} \frac{1}{n} \widehat{Z}'_1 \widehat{Z}_1 \quad ,\end{aligned} \quad (7.18)$$

joten parametristimaattien hajonta-arviot saadaan normaaliin tapaan \widehat{Z}_1 :n momenttimatriisin inverssinä.

7.2.2.2 Kolmivaiheinen pns- menetelmä

Oleellinen osa mallia (7.10) oli se, että oletettiin ε_t :n komponenttien korreloivan keskenään täysin vapaasti, $\text{cov}(\varepsilon_t) = \Sigma = (\sigma_{ij})$. Tästä syystä on ilmeistä, että estimoinnin tehokkuutta voidaan parantaa turvautumalla SURE-estimointiin. Kaksivaiheisen 2SLS-menetelmän ideaa mukailen ja muotoa (7.3) jäljitellen käytetään selittäjämatrisina sovitteista $\hat{Z}_{\cdot j}$ koostuvaa matriisia

$$\hat{Z} = \begin{pmatrix} \hat{Z}_{\cdot 1} & 0 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & \dots & 0 & \hat{Z}_{\cdot K} \end{pmatrix} \quad (7.19)$$

ja selitettävänä muuttujana vektoria

$$Y_* = \text{vec}(Y) \quad . \quad (7.20)$$

Tällöin päädytään malliin

$$Y_* = \hat{Z}\beta + \varepsilon_* \quad , \quad (7.21)$$

jossa $\beta = (\beta_{\cdot 1}^o \dots \beta_{\cdot K}^o)'$ ja $\varepsilon_* = \text{vec}(\varepsilon)$. Estimoidaan β kaksivaiheisella SURE-estimaattorilla

$$\tilde{\beta} = (\hat{Z}'\hat{V}^{-1}\hat{Z})^{-1}\hat{Z}'\hat{V}^{-1}Y_* \quad , \quad (7.22)$$

jossa

$$\hat{V} = (\hat{v}_{ij}) \otimes I \quad \text{ja}$$

$$\hat{v}_{ij} = \frac{1}{n} (Y_{(i)} - \hat{Z}_{\cdot i}\tilde{\beta}_{\cdot i}^o)'(Y_{(j)} - \hat{Z}_{\cdot j}\tilde{\beta}_{\cdot j}^o) \quad i, j = 1, \dots, K \quad .$$

Virhetermien kovarianssimatriisin estimaattori $\hat{\Sigma}$ muodostetaan siis 2SLS-estimaatteihin liittyvien rakenneyhtälöiden jäännöstermien perusteella. Estimaattoria $\tilde{\beta}$ sanotaan *kolmivaiheiseksi* pns-estimaattoriksi (3SLS).

Huomautus 7.4: Kannattaa ehkä mainita toinenkin 2SLS- ja 3SLS-estimaattorien esitystapa, jossa 1. vaihe voidaan ohittaa muuttujien transformoinnin avulla: Muodostetaan $(X'X)^{-1}$:n Cholesky-dekompositio

$$(X'X)^{-1} = LL' \quad .$$

Merkitään

$$U_* = \text{vec}(L'X'Y) \quad , \quad \Psi = \text{vec}(L'X'\varepsilon)$$

ja

$$W = \begin{pmatrix} \widehat{L}'X'Z_{.1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & L'X'\widehat{Z}_{.K} \end{pmatrix} .$$

Tällöin

$$\begin{aligned} \text{cov}(\Psi) &= V_o = \Sigma \otimes I_{m \times m} && \text{ja} \\ U_* &= W\beta + \Psi && . \end{aligned} \tag{7.23}$$

Kaksi- ja kolmivaiheiset estimaattorit voidaan nyt esittää lyhyesti muodossa

$$\begin{aligned} \text{2SLS} : \quad \tilde{\beta} &= (W'W)^{-1}W'U_* \\ \text{3SLS} : \quad \tilde{\tilde{\beta}} &= (W'\widehat{V}_o^{-1}W)^{-1}W'\widehat{V}_o^{-1}U_* \quad , \\ \widehat{V}_o &= \widehat{\Sigma} \otimes I_{m \times m} \end{aligned} \tag{7.24}$$

Esitystavat (7.24) voidaan johtaa toteamalla, että

$$\begin{aligned} W'_{.1}W_{.1} &= Z'_{.1}XLL'X'Z_{.1} \\ &= Z'_{.1}P_XZ_{.1} = \widehat{Z}'_{.1}\widehat{Z}_{.1} \quad \text{jne.} \end{aligned}$$

Koska Ψ ja W ovat asymptoottisesti toisistaan riippumattomia, on

$$\begin{aligned} \sqrt{n}(\tilde{\beta} - \beta) &\overset{\text{asympt.}}{\sim} N_p(0, p \lim_{n \rightarrow \infty} \left(\frac{1}{n} W'W\right)^{-1}) \quad , \\ \sqrt{n}(\tilde{\tilde{\beta}} - \beta) &\overset{\text{asympt.}}{\sim} N_p(0, p \lim_{n \rightarrow \infty} \left(\frac{1}{n} W'\widehat{V}_o^{-1}W\right)^{-1}) \quad , \end{aligned} \tag{7.25}$$

jossa p tarkoittaa β -parametrien lukumäärää.

Voidaan siis sanoa, että 3SLS on *asymptoottisesti* tehokkaampi kuin 2SLS.

Lisäksi voidaan osoittaa, että 3SLS on *asymptoottisesti tehokas*, mikäli kovarianssimatriisia Σ ei ole mitenkään rajoitettu mallissa (7.10). ■

Huomautus 7.5: Redusoidun muodon (7.11) parametreja estimointiin edellä OLS:illa, joka tietysti toimiikin tarkentuvasti. Toisaalta kuitenkin $\Pi = -\Delta\Gamma$, joten Π :n elementtien välillä on (yli-identifioitujen yhtälöiden seurauksena) epälineaarisia kytkentöjä, joita OLS-estimaatit $\widehat{\Pi}$ eivät toteuta. Tästä syystä joskus

käytetään ns. *johdettuja* redusoidun muodon estimaattoreita $\tilde{\Pi} = -\tilde{\Delta}\tilde{\Gamma}$ tai $\tilde{\tilde{\Pi}} = -\tilde{\tilde{\Delta}}\tilde{\tilde{\Gamma}}$, jotka luonnollisesti ovat konsistentteja. Voidaan osoittaa, että kolmivaiheiseen menetelmään liittyvä $\tilde{\tilde{\Pi}}$ on asympotoottisesti tehokkaampi kuin OLS-estimaattori $\hat{\Pi}$. ■

7.2.2.3 ”Täyden informaation” ML- menetelmä (FIML)

Koska ε_t :n tiheysfunktio on muotoa

$$f_{\varepsilon_t}(\varepsilon) = \left(\frac{1}{\sqrt{2\pi}}\right)^K \frac{1}{\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}\varepsilon'\Sigma^{-1}\varepsilon}$$

ja koska transformaation $\varepsilon_t \rightarrow Y_t$ Jacobin determinantti on $\det\Gamma^{-1}$, on

$$\begin{aligned} \log L_{Y|X}(\Gamma, \Delta, \Sigma) &\simeq -\frac{n}{2} \log \det\Sigma + n \log \det\Gamma \\ &\quad - \frac{1}{2} \sum_{t=1}^n (Y_t'\Gamma + X_t'\Delta)\Sigma^{-1}(\Gamma'Y_t + \Delta'X_t) \end{aligned} \tag{7.26}$$

ML- menetelmän käytölle ei näin ollen ole mitään periaatteellisia esteitä, sillä likelihood (7.26) on täysin säännöllinen ja helposti derivoitavissa (jopa useampia kertoja) kaikkien parametrien suhteen.

Jos rakennemuodon virhetermien kovarianssimatriisi Σ on täysin rajoitteeton, on 3SLS-estimaattoreilla sama asympotoottinen otantajakauma kuin ML-estimaattoreilla.

Välihuomautus 7.1: Palautettakoon nyt mieliin luvussa 2.6. esitelty Hausman-testausperiaate. Mikäli halutaan *testata*, ovatko esimerkiksi muuttujat y_{1t} ja y_{2t} simultaanisessa vuorovaikutussuhteessa toisiinsa, vai olisiko po. muuttujien välinen yhteys *yksisuuntainen* ($y_{2t} \rightarrow y_{1t}$), ts. mallin

$$H_o : \quad y_{1t} = -\delta_{11} - \delta_{21} x_{2t} - \dots - \delta_{m1} x_{mt} - \gamma_{21} y_{2t} + \varepsilon_{1t} \quad \varepsilon_{1t} \sim NID(0, \sigma_1^2)$$

mukainen. Tämän ajatuksen realistisuutta voidaan todellakin empiirisesti tutkia, mikäli pystytään täsmentämään, mikä olisi tämän yksisuuntaisuushypoteesin kanssa kilpaileva *vaihtoehtoinen* hypoteesi. Jos ajateltaisiin, että vaihtoehtoisen hypoteesin mukainen käänteinen riippuvuus olisi muotoa

$$H_1 : \quad y_{2t} = -\delta_{12} - \delta_{22} x_{2t} - \dots - \delta_{m2} x_{mt} - \gamma_{12} y_{1t} + \varepsilon_{2t} \quad \varepsilon_{2t} \sim NID(0, \sigma_2^2)$$

ja että parametrien $\delta_{21}, \dots, \delta_{m1}, \delta_{22}, \dots, \delta_{m2}$ joukossa olisi riittävä määrä nollia, jotta jäljelle jääneet parametrit olisivat estimoitavissa, voitaisiin yksisuuntaisuushypoteesia $y_{2t} \rightarrow y_{1t}$ testata joko Waldin periaatteella tarkastelemalla, onko $\hat{\gamma}_{12}^{(ML)}$ kaukana nolasta, *tai* Hausmanin periaatteella tutkimalla, onko erotus $\hat{\gamma}_{21}^{(ML)} - \hat{\gamma}_{21}^{(OLS)}$ kaukana nolasta. Lauseen 2.6 todistuksen yhteydessä esitetyn tuloksen (2.37) mukaisesti H_o :n vallitessa pitäisi olla asympotoottisesti

$$\text{var} \left(\hat{\gamma}_{21}^{(ML)} - \hat{\gamma}_{21}^{(OLS)} \right) \approx \text{var} \left(\hat{\gamma}_{21}^{(ML)} \right) - \text{var} \left(\hat{\gamma}_{21}^{(OLS)} \right) \quad ,$$

joten erotuksen $\hat{\gamma}_{21}^{(ML)} - \hat{\gamma}_{21}^{(OLS)}$ otantajakauman normaaliapproksimaatioon perustuvan testin suorittaminen olisi todella helppoa. ■

7.3 Dynaamiset vuorovaikutusmallit

Tarve vuorovaikutussysteemien *dynamiikan* huomioon ottamiseen kasvaa yleensä havaintovälin lyhentyessä, samalla kun samanaikaisen simultaanisuuden aiheuttamat identifiointi- ja estimointiongelmat pienenevät. Mallimuodossa (7.10) saattoi ”ennalta määrättyihin” muuttujiin sisältyä myös viivästettyjä endogeenisiä tekijöitä, joten malli (7.10) oli jo eräessä mielessä ”dynaaminen”. Tässä luvussa tarkastellaan hyvin lyhyesti tapausta, jossa dynamiikan rakenne ei ole etukäteen tiedossa, vaan mallin muoto joudutaan valitsemaan havaintojen perusteella. Muodon (7.10) luonteva yleistys olisi ns. *VARMAX*-mallityyppi

$$\Phi(L)Y_t + \Delta(L)X_t = \Theta(L)\varepsilon_t \quad , \quad (7.27)$$

jossa Y_t ja ε_t ovat luonteiltaan saman kaltaisia kuin mallissa (7.10), mutta X_t sisältää vain eksogeenisiä tekijöitä. Operaattorit $\Phi(L)$, $\Delta(L)$ ja $\Theta(L)$ ovat viiveoperaattorin L matriisipolynomeja

$$\begin{aligned} \Phi(L) &= \Phi_o - \Phi_1 L - \dots - \Phi_p L^p \quad , \\ \Delta(L) &= \Delta_o - \Delta_1 L - \dots - \Delta_r L^r \quad \text{ja} \\ \Theta(L) &= I - \Theta_1 L - \dots - \Theta_q L^q \quad . \end{aligned}$$

Redusoitua muotoa (7.11) vastaavaa esitystä

$$\begin{aligned} Y_t &= -\Phi(L)^{-1}\Delta(L)X_t + \Phi(L)^{-1}\Theta(L)\varepsilon_t \\ &= \Pi_o X_t + \Pi_1 X_{t-1} + \dots + \Phi(L)^{-1}\Theta(L)\varepsilon_t \quad , \end{aligned} \quad (7.28)$$

sanotaan nyt mallin (7.27) *finaalimuodoksi*.

Matriiseja $\Pi_0, \Pi_1, \Pi_2, \dots$ sanotaan malliin liittyviksi *impulssivastematriiseiksi* ja matriiseja

$$-\Phi(1)\Delta(1) = \Pi_0 + \Pi_1 + \Pi_2 + \dots$$

sanotaan *kokonaisvastematriiseiksi*. VARMAX-mallien (7.27) stabiilisuusominaisuudet määräytyvät pelkästään mallin autoregressiivisen osan $\Phi(L)$ perusteella. Tästä syystä seuraavassa tarkastellaan lähemmin erityisesti ns. VAR- malleja

$$\begin{aligned} \Phi_0 Y_t &= \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \varepsilon_t \quad , \\ \varepsilon_t &\sim NID_K(0, \Sigma) \quad , \quad \Phi_0 = I \end{aligned} \tag{7.29}$$

joiden määrittely-yhtälö voidaan luonnollisesti kirjoittaa aikaisempia merkintöjä käyttäen myös muotoon $\Phi(L)Y_t = \varepsilon_t$.

Mikäli ”innvoatioimpulssien” ε_t kovarianssirakenne Σ halutaan jättää täysin vapaasti parametroiduksi, on selvää, *etteivät* Y_t -vektorin komponenttien *samaaikaisia* vuorovaikutuksia kuvaavat Φ_0 -parametrit voi koskaan olla identifioituvia. Tästä syystä mallimuodossa (7.29) onkin asetettu rajoite $\Phi_0 = I$.

Lause 7.2: VAR(p)-mallin (7.29) määrittelemä $\{Y_t\}$ -prosessi voi olla stationäärinen (kehityksetön) vain, jos yhtälön

$$\det\Phi(s^{-1}) = 0 \tag{7.30}$$

kaikki ratkaisut s ovat itseisarvoiltaan ykköstä pienempiä.

Todistus: Todistus sisältyy kolmeen seuraavaan huomautukseen 7.6 - 7.8 . ■

Huomautus 7.6: Ajatellaan aluksi yksinkertaisinta mahdollista tilannetta $K = 1$, $p = 1$, jolloin malli (7.29) surkastuu muotoon

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2) \quad . \tag{7.31}$$

Mallin ns. ”karakteristinen yhtälö” (7.30) on tällöin muotoa

$$1 - \phi_1 \frac{1}{s} = 0 \quad ,$$

ja sen ainoa ratkaisu on tietysti $s = \phi_1$.

Jos prosessi $\{y_t\}$ ”käynnistetään” hetkellä $t = 0$ tilasta $y_0 \neq 0$, nähdään määrittely-yhtälöstä (7.31) heti, että

$$\begin{aligned} y_t &= \phi_1^t y_0 + (\varepsilon_t + \phi_1 \varepsilon_{t-1} + \dots + \phi_1^{t-1} \varepsilon_1) \\ &= \phi_1^t y_0 + \kappa_t \quad . \end{aligned}$$

Tästä muodosta nähdään heti, että jos $|\phi_1| > 1$, on y_t -sarjassa eksponentiaalinen trendi. Jos taas $|\phi_1| = 1$, on

$$\lim_{t \rightarrow \infty} \text{var}(y_t) = \lim_{t \rightarrow \infty} \text{var}(\kappa_t) = \infty \quad .$$

Kummassakaan tapauksessa $\{y_t\}$ ei voi olla stationäärinen. ■

Huomautus 7.7: Tarkastellaan seuraavaksi tilannetta $K = 1$, $p > 1$. Tällöin mallin määrittely-yhtälöstä

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad , \quad \varepsilon_t \sim NID(0, \sigma^2)$$

seuraa välittömästi, että odotusarvojen $q_t = Ey_t$ muodostama lukujono toteuttaa differenssiyhtälön

$$Ey_t = \phi_1 Ey_{t-1} + \dots + \phi_p Ey_{t-p} \iff \phi(L)Ey_t \equiv 0 \quad ,$$

jossa $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$. Huomataan heti, että muotoa $q_t = s^t$ olevat lukujonot ovat tämän differenssiyhtälön ratkaisuja, mikäli $\phi(\frac{1}{s}) = 0$, sillä tällöin

$$\phi(L)s^t = s^t - \phi_1 s^{t-1} - \dots - \phi_p s^{t-p} = s^t \phi\left(\frac{1}{s}\right) \equiv 0 \quad .$$

Yhtälöllä $\phi(\frac{1}{s}) = 0$ on periaatteessa p ratkaisua. Toisaalta p alkuarvoa määrää täydellisesti differenssiyhtälön $\phi(L)q_t = 0$ ratkaisuna saatavan lukujonon, joten *kaikkien* ratkaisujonon täytyy olla muotoa

$$q_t = Ey_t = c_1 s_1^t + \dots + c_p s_p^t \quad , \quad \text{jossa} \quad \phi\left(\frac{1}{s_j}\right) = 0 \quad j = 1, \dots, p \quad .$$

Jos yksikin juurista s_j on itseisarvoltaan > 1 , ”räjähtää” odotusarvojen Ey_t ajan t edetessä, ja mallia sanotaan *epästabiiliksi*.

Toisaalta, AR(p)-malli voidaan kirjoittaa myös vektorimuotoon

$$\begin{aligned} \tilde{Y}_t &= \begin{pmatrix} y_t \\ \cdot \\ \cdot \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \dots & \dots & \dots & \phi_p \\ 1 & 0 & \dots & \dots & 0 \\ \cdot & \dots & \dots & \dots & \cdot \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix} \tilde{Y}_{t-1} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \cdot \\ 0 \end{pmatrix} \\ &= \tilde{\Phi} \tilde{Y}_{t-1} + \tilde{\varepsilon}_t \quad , \end{aligned}$$

jolloin huomautuksen 7.6 laskelmien mukaisesti saadaan

$$\tilde{Y}_t = \tilde{\Phi}^t \tilde{Y}_0 + \sum_{\nu=0}^{t-1} \tilde{\Phi}^\nu \tilde{\varepsilon}_{t-\nu} \quad .$$

Tästä nähdään, että matriisin $\tilde{\Phi}$ ominaisarvot määräävät AR(p)-mallin stabiilisuusominaisuudet.

Toisaalta voidaan helposti nähdä, että

$$\det(\tilde{\Phi} - sI) = \det \begin{pmatrix} \phi_1 - s & \phi_2 & \dots & \dots & \phi_p \\ 1 & -s & 0 & \dots & 0 \\ \cdot & \dots & \dots & \dots & \cdot \\ 0 & \dots & 0 & 1 & -s \end{pmatrix} = (-1)^{p+1} s^p \phi\left(\frac{1}{s}\right) \quad ,$$

joten matriisin $\tilde{\Phi}$ ominaisarvot ja karakteristisen yhtälön $\phi\left(\frac{1}{s}\right) = 0$ juuret ovat samat.

(Edellä määritellyn $\tilde{\Phi}$ -matriisin muotoisia matriiseja sanotaan *companion*-matriiseiksi.) ■

Huomautus 7.8: Tarkastellaan nyt tapausta $K > 1$, $p > 1$. Myös VAR(p)-malli (7.29) voidaan kirjoittaa huomautuksen 7.7 mukaisesti muotoon

$$\begin{aligned} \tilde{Y}_t &= \begin{pmatrix} Y_t \\ \cdot \\ \cdot \\ Y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \dots & \dots & \dots & \Phi_p \\ I & 0 & \dots & \dots & 0 \\ \cdot & \dots & \dots & \dots & \cdot \\ 0 & \dots & 0 & I & 0 \end{pmatrix} \tilde{Y}_{t-1} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \cdot \\ 0 \end{pmatrix} \\ &= \tilde{\Phi} \tilde{Y}_{t-1} + \tilde{\varepsilon}_t \quad , \end{aligned}$$

jolloin nähdään, että ”blokki-companion”-matriisin $\tilde{\Phi}$ ominaisarvot saadaan yhtälön

$$\det(\tilde{\Phi} - sI) = \det \begin{pmatrix} \Phi_1 - sI & \Phi_2 & \dots & \dots & \Phi_p \\ I & -sI & 0 & \dots & 0 \\ \cdot & \dots & \dots & \dots & \cdot \\ 0 & \dots & 0 & I & -sI \end{pmatrix} = (-1)^{pK+1} s^{pK} \det \Phi\left(\frac{1}{s}\right)$$

ratkaisuihin, jossa

$$\Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p \quad .$$

Tällä tavalla lauseen 7.2 väite on tullut todistetuksi. ■

7.4 VAR- mallin muodon spesifointi ja parametrien estimointi

Ajatellaan nyt, että havaintosarjaan Y_1, \dots, Y_n olisi tarkoitus sovittaa VAR(p)-mallia (7.29). Voidaan helposti todeta, että havaintoaineiston määräämän likelihood-funktion logaritmi on muotoa

$$\log L_{Y_1, \dots, Y_n}(\Phi, \Sigma) = \log L_{Y_{p+1}, \dots, Y_n | Y_1, \dots, Y_p}(\Phi, \Sigma) + \log L_{Y_1, \dots, Y_p}(\Phi, \Sigma) \quad , \quad (7.32)$$

$$\text{jossa} \quad \Phi = (\Phi_1 \quad \dots \quad \Phi_p)' \quad .$$

Lausekkeen (7.32) ensimmäinen termi on puolestaan muotoa

$$\log L_{Y_{p+1}, \dots, Y_n | Y_1, \dots, Y_p}(\Phi, \Sigma) \simeq -\frac{n-p}{2} \log \det \Sigma - \frac{1}{2} \sum_{t=p+1}^n \varepsilon_t' \Sigma^{-1} \varepsilon_t \quad , \quad (7.33)$$

$$\text{jossa} \quad \varepsilon_t = \Phi(L)Y_t \quad .$$

Mikäli kuvattavat aikasarjat Y_t ovat epästationäärisiä, joudutaan päätelmät tietysti perustamaan yksinomaan *ehdolliseen* likelihood-funktioon (7.33). Jos taas kuvattavan ilmiön voidaan olettaa olevan ”stationäärisessä tilassa” jo havaintoperiodin alkaessa, voidaan päätelmät perustaa ”tarkkaan” likelihood-funktioon (7.32).

Tarkastellaan seuraavaksi lausekkeen (7.32) optimointia parametrien Φ ja Σ suhteen, kun Σ :lle ei ole asetettu mitään etukäteisrajoituksia. Otetaan aluksi käyttöön merkinnät

$$Z = (Y_{p+1} \quad \dots \quad Y_n)' \quad \text{ja} \quad X = \begin{pmatrix} Y_p' & \dots & Y_1' \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ Y_{n-1}' & \dots & Y_{n-p}' \end{pmatrix} \quad .$$

Muotoillaan sitten (7.33) uudelleen käyttämällä Σ :n asemesta parametreina matriisin $A = \Sigma^{-1}$ alkioita:

$$\log L(\Phi, A) \simeq \frac{n-p}{2} \log \det A - \frac{1}{2} \sum_{t=p+1}^n \varepsilon_t' A \varepsilon_t \quad . \quad (7.34)$$

Muodon (7.34) perusteella on helppo huomata, että

$$\frac{\partial}{\partial a_{ij}} \log L(\Phi, A) = \frac{n-p}{2} \cdot \frac{1}{\det A} \cdot (-1)^{i+j} \cdot \det A_{ij} - \frac{1}{2} \sum_{t=p+1}^n \varepsilon_{ti} \varepsilon_{tj} \quad , \quad (7.35)$$

jossa A_{ij} tarkoittaa alkion a_{ij} alimatriisia. Toisaalta

$$\frac{1}{\det A} \cdot (-1)^{i+j} \cdot \det A_{ij} = (A^{-1})_{ij} = \sigma_{ij} \quad ,$$

joten derivaattojen (7.35) nollakohdat vastaavat valintoja

$$\sigma_{ij} = \frac{1}{n-p} \sum_{t=p+1}^n \varepsilon_{ti} \varepsilon_{tj} \quad i, j = 1, \dots, K \quad .$$

Tästä nähdään, että Σ :n ML-estimaattoriksi saadaan jäännöstermien kovarianssimatriisi

$$\begin{aligned} \widehat{\Sigma} &= \frac{1}{n-p} \sum_{t=p+1}^n \varepsilon_t \varepsilon_t' \\ &= \frac{1}{n-p} (Z - X\Phi)' (Z - X\Phi) \quad . \end{aligned} \quad (7.36)$$

Sijoittamalla $\widehat{\Sigma}$:n lauseke kaavaan (7.33) saadaan ”konsentroiduksi” likelihood-funktioksi

$$\begin{aligned} \log L_{\max}(\Phi) &\simeq -\frac{n-p}{2} \log \det \widehat{\Sigma} - \frac{1}{2} \text{tr} \left(\widehat{\Sigma}^{-1} \sum_{t=p+1}^n \varepsilon_t \varepsilon_t' \right) \\ &= -\frac{n-p}{2} \log \det \widehat{\Sigma} - \frac{1}{2} \cdot K \cdot (n-p) \quad , \end{aligned}$$

joten Φ :n ML-estimaattori saadaan minimoimalla jäännöstermien ”yleistetty varianssi” $\det \widehat{\Sigma}$.

Toisaalta (7.36):n mukaan

$$\begin{aligned} (n-p) \widehat{\Sigma} &= (Z - X\Phi)' (Z - X\Phi) \\ &= Z' (I - X(X'X)^{-1}X') Z \\ &\quad + [\Phi - (X'X)^{-1}X'Z]' (X'X) [\Phi - (X'X)^{-1}X'Z] \quad . \end{aligned} \quad (7.37)$$

Matriisi $(n-p)\widehat{\Sigma}$ voidaan siis hajottaa kahden ei-negatiivisesti definiitin matriisin summaksi, joten

$$\det \widehat{\Sigma} \geq \det \frac{1}{n-p} Z' (I - X(X'X)^{-1}X') Z \quad (7.38)$$

kaikilla Φ :n arvoilla.

Yhtäsuuruus epäyhtälössä (7.38) voidaan saavuttaa valitsemalla

$$\hat{\Phi} = (X'X)^{-1}X'Z \quad . \quad (7.39)$$

Tästä muodosta on helppo tunnistaa tavallinen OLS-estimaattori, joten *rajoittamattoman* VAR(p)-mallin parametrien (ehdolliset) ML-estimaattorit saadaan *tavallisina OLS-estimaattoreina*.

Samalla nähdään, että saavutettu likelihood-maksimi on muotoa

$$\begin{aligned} \log L(\hat{\Phi}, \hat{\Sigma}) &\simeq -\frac{n-p}{2} \log \det \hat{\Sigma} \quad , \\ \text{jossa} \quad \hat{\Sigma} &= \frac{1}{n-p} (Z - X\hat{\Phi})'(Z - X\hat{\Phi}) \\ &= \frac{1}{n-p} Z'(I - P_X)Z \\ &= \frac{1}{n-p} \sum_{t=p+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t' \quad . \end{aligned} \quad (7.40)$$

Mikäli (rajoittamattoman) VAR(p)-mallin puitteissa haluttaisiin testata hypoteesia

$$H_o : \quad \Phi_p = 0 \quad ,$$

(ts. ajatusta, että VAR($p-1$)-malli riittäisi aineiston käyttäytymisen kuvaamiseen), saataisiin LR- testisuureksi lauseke

$$-2 \log \Lambda = (n-p) \log \frac{\det \hat{\Sigma}^{(o)}}{\det \hat{\Sigma}} \quad (7.41)$$

jossa $\hat{\Sigma}^{(o)}$ tarkoittaa Σ :n ML-estimaattoria H_o :n (ts. VAR($p-1$)-mallin) puitteissa.

Näiden alkuvalmistelujen jälkeen voidaankin hahmotella yksinkertainen proseduuri VAR-mallin (7.29) kertaluvun p valitsemiseksi:

1° Sovitetaan aineistoon kasvaviin kertalukuihin ν liittyviä VAR(ν)-malleja

$$Y_t = \Phi_1^{(\nu)} Y_{t-1} + \dots + \Phi_\nu^{(\nu)} Y_{t-\nu} + \varepsilon_t^{(\nu)} \quad \nu = 1, 2, 3, \dots \quad (7.42)$$

OLS-estimointitekniikkaa käyttäen. Korkeimman kertaluvun termien kerroinmatriisien estimaattoreista syntyvää jonoa

$$\widehat{\Phi}_\nu^{(\nu)} \quad \nu = 1, 2, \dots \quad (7.43)$$

sanotaan estimoiduksi *osittaisristikorrelaatiofunktioiksi*.

Mikäli aineistoon sopisi VAR(p)-malli, pitäisi estimoitujen $\widehat{\Phi}_\nu^{(\nu)}$ -matriisien olla lähellä nollaa kaikilla $\nu > p$.

Toinen (parempi) tapa sopivan p :n valitsemiseksi olisi testata sekventiaalisesti eri kertalukujen riittävyttä LR-testisuureiden (7.41) avulla. Merkitään mallien (7.42) OLS-residuaaleja symboleilla $\widehat{\varepsilon}_t^{(\nu)}$ ($t = \nu + 1, \dots, n$) ja niistä laskettuja kovarianssimatriiseja symboleilla

$$S(\nu) = \frac{1}{n - \nu} \sum_{t=\nu+1}^n \widehat{\varepsilon}_t^{(\nu)} \widehat{\varepsilon}_t^{(\nu)'} \quad \nu = 1, 2, \dots$$

Testisuureiden

$$m(\nu) = (n - \nu) \log \frac{\det S(\nu - 1)}{\det S(\nu)}, \quad (7.44)$$

pitäisi tällöin noudattaa asympotoottisesti $\chi_{K \cdot K}^2$ -jakaumaa, mikäli $\nu = p + 1$. Kannattaa siis etsiä sellaista kertalukua ν , jolle $m(\nu)$ ei enää saisi ($\chi_{K \cdot K}^2$ -jakaumaan verrattuna) kovin suurta arvoa. Samalla $S(\nu)$ -matriisien diagonaalelementit antavat käsityksen mallin ja havaintojen välisen yhteensovivuuden paranemisesta ”yhtälöittäin” mallin kertaluvun kasvaessa.

- 2° Tärkein VAR(p)-malliin (7.29) sisältyvä oletus koski virhetermien ”kohinaominaisuutta”

$$\varepsilon_t \sim i.i.d._K(0, \Sigma) \quad , \quad (7.45)$$

jonka vallitessa ”autokovarianssimatriisien” $E\varepsilon_t \varepsilon_{t+\tau}'$ pitäisi olla nollia kaikilla $\tau \neq 0$. Tarkastelemalla jäännöstermien $\widehat{\varepsilon}_t^{(\nu)}$ vastaavia ”ristikorrelaatioestimaatteja”

$$(\text{diag } S(\nu))^{-\frac{1}{2}} \cdot \frac{1}{n - \nu - \tau} \sum_{t=\nu+1}^{n-\tau} \widehat{\varepsilon}_t^{(\nu)} \widehat{\varepsilon}_{t+\tau}^{(\nu)'} \cdot (\text{diag } S(\nu))^{-\frac{1}{2}} \quad (7.46)$$

erilaisilla τ :n arvoilla $\tau = 1, 2, \dots$ voidaan yleensä muodostaa jonkinlainen käsitys virhetermien kohinaoletuksen (7.45) realistisuudesta (ja tätä kautta koko VAR-mallin realistisuudesta).

Huomautus 7.9: Edellä kuvattua VAR-kertaluvun valintastrategiaa voidaan käyttää (sopivasti modifioituna) hyväksi myös VARMAX-mallien (7.27) spesifioinnin yhteydessä. ■

Huomautus 7.10: Kun stabiilin VARMAX-mallin (7.27) parametrit on estimoitu, voidaan johtaa mallin redusoitu muoto (7.28)

$$\begin{aligned} Y_t &= \Pi_o X_t + \Pi_1 X_{t-1} + \dots + \Phi(L)^{-1} \Theta(L) \varepsilon_t \\ &= \Pi_o X_t + \Pi_1 X_{t-1} + \dots + \Psi_o \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots \quad , \end{aligned}$$

jossa on käytetty merkintöjä $\Psi(L) = \sum_{\nu=0}^{\infty} \Psi_{\nu} L^{\nu} = \Phi(L)^{-1} \Theta(L)$. Kuten aiemmin todettiin, kutsutaan kertoimia $\Pi_o, \Pi_1, \Pi_2, \dots$ *impulssivasteiksi*, jotka kertovat, miten eksogeenisiin muuttujiin X_t liittyvät muutosimpulssit myöhemmin heijastuvat endogeenisiin muuttujiin Y_t . Aivan samaan tapaan voidaan ajatella, että kertoimet $\Psi_o, \Psi_1, \Psi_2, \dots$ kertovat, miten *endogeenisiin* muuttujiin Y_t liittyvät uutuusimpulssit (innovaatioimpulssit) ε_t myöhemmin heijastuvat muihin endogeenisiin Y_t -vektorin komponentteihin. Tästä syystä myös kerroinmatriiseja Ψ_{ν} ($\nu = 0, 1, 2, \dots$) kutsutaan *impulssivasteiksi*. Ne siis kertovat, miten aikaisemmat innovaatioimpulssit $\varepsilon_{t-\nu}$ heijastuvat Y_t -havaintoihin. Niiden avulla voidaan laskea myös *kumulatiivisia* impulssivasteita $C_{\nu} = \Psi_o + \Psi_1 + \dots + \Psi_{\nu}$ ($\nu = 0, 1, 2, \dots$).

Impulssivasteiden tulkinta on kuitenkin hieman ongelmallista, mikäli innovaatioimpulssien ε_t komponenttien ajatellaan korreloivan keskenään. Tästä syystä onkin yleensä mielenkiintoisempaa käyttää ajattelun pohjana mallin (7.27) sellaista versiota, jossa $\Sigma = \text{cov}(\varepsilon_t)$ oletetaan *diagonaalimatriiksi* ja tarkasteltavien muuttujien välisiin *samanaikaisiin* riippuvuuksiin otetaan kantaa matriisiin Φ_o muotoa koskevien oletusten kautta. (Matriisin Φ_o diagonaalielementit oletetaan aina ykkösiksi.) Matriisi Φ_o voidaan rajoittaa esimerkiksi alakolmiomatriisiksi, jolloin siis itse asiassa oletetaan, että vektorin Y_t komponentit riippuvat toisistaan *rekursiivisesti* niiden luettelujärjestyksessä ylhäältä alaspäin. Tämän ajatuskehikon mukaisia impulssivastekertoimia kutsutaan *ortogonalsoiduiksi* impulssivasteiksi. Vaikka monet ohjelmistot auuliisti laskevatkin rekursiiviseen vaihtustapaan liittyviä impulssivaste-estimaatteja, täytyy soveltajan aina muistaa, että tutkittavan ilmiön kannalta näillä laskelmilla on mielenkiintoa vain, mikäli tutkijalla on vahvat perustelut rekursiivista vaikutustapaa (ja muuttujien keskinäistä *vaikutusjärjestystä*) koskevien oletustensa tueksi. ■

Esimerkki 7.2: Tarkastellaan euroalueen valuuttojen ja Yhdysvaltain dollarin keskimääräisten vaihtokurssien (suhteellisia) muutoksia $Dexch_t$ kuukausittain ajanjaksona 1990:1-2002:5. Ns. ”kattamattoman korkopariteettiteorian” mukaan vaihtokurssien odotetun muutoksen $E_t(dexch_{t+1})$ pitäisi vastata vertailuvaluuttojen *korkoeroa*. Seuraavassa käytetään korkomuuttujina 10 vuoden obligaatiolainojen keskikorkoja $Bondeuro_t$ ja $BondUSA_t$ transformoituina muotoihin $iteu_t = \log(1 + Bondeuro_t/100)$ ja $itus_t = \log(1 + BondUSA_t/100)$. Euroja dollarialueilla vallitsevien pitkien korkojen eroa mitataan sitten muuttujalla $Korkoero_t = iteu_t - itus_t$. Hetkellä t vallitsevien, tulevaa vaihtokurssimuutosta koskevien odotusten $E_t(dexch_{t+1})$ korvikkeena käytetään tuoreinta käytettävissä olevaa muutostietoa $Dexch_t$.

Haetaan aluksi vektorimuuttujan $Y_t = (Korkoero_t \quad Dexch_t)'$ kehityksen kuvaamiseen sopivan VAR(p)-mallin kertalukua ja poistetaan mallista sitten "turhilta" näyttävät parametrit. Estimoidaan mallin parametrit ML-periaatteella ja tutkitaan estimoidun mallin stabiilisuutta laskemalla siihen liittyvän karakterisen yhtälön (7.30) juuret.

Lopuksi on oletettu, että muuttujien $Korkoero_t$ ja $Dexch_t$ väliset, saman kuukauden aikana ilmenevät vuorovaikutukset ilmenevät enimmäkseen suuntaan $Korkoero_t \rightarrow Dexch_t$. Tämän oletuksen mukaisten ortogonaloitujen impulssivasteestimaattien käyttäytymistä on havainnollistettu graafisten kuvaajien avulla.

```
Algebra code for uip.in7:
iteu = log(1+Bondeuro/100);
itus = log(1+BondUSA/100);
Algebra code for uip.in7:
exch = log(Eurodollar);
Dexch = diff(exch,1);
Korkoero = iteu-itus;
```

EQ(1) Estimating the unrestricted reduced form by OLS (using uip.in7)
The present sample is: 1990 (6) to 2002 (5)

URF Equation 1 for Dexch

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	0.34806	0.087226	3.990	0.0001
Dexch_2	-0.28219	0.090244	-3.127	0.0022
Dexch_3	0.13429	0.089848	1.495	0.1373
Dexch_4	-0.15963	0.084454	-1.890	0.0609
Korkoero_1	-2.6297	0.87153	-3.017	0.0030
Korkoero_2	1.4157	1.2040	1.176	0.2417
Korkoero_3	0.91405	1.2067	0.757	0.4501
Korkoero_4	0.015893	0.87592	0.018	0.9856
Constant	0.0017421	0.0018573	0.938	0.3499

\sigma = 0.0220027 RSS = 0.06535607503

URF Equation 2 for Korkoero

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	-0.013540	0.0086966	-1.557	0.1218
Dexch_2	0.0090887	0.0089974	1.010	0.3142
Dexch_3	0.0098030	0.0089580	1.094	0.2758
Dexch_4	0.0034914	0.0084202	0.415	0.6791
Korkoero_1	0.99793	0.086892	11.485	0.0000
Korkoero_2	-0.085798	0.12004	-0.715	0.4760
Korkoero_3	0.045647	0.12031	0.379	0.7050
Korkoero_4	0.016508	0.087331	0.189	0.8504
Constant	-6.1979e-005	0.00018517	-0.335	0.7384

\sigma = 0.0021937 RSS = 0.000649664454

correlation of URF residuals

	Dexch	Korkoero
Dexch	1.0000	
Korkoero	-0.18720	1.0000

standard deviations of URF residuals

Dexch	Korkoero
0.022003	0.0021937

loglik = 1443.0426 log|\Omega| = -20.0423 |\Omega| = 1.97587e-009 T = 144

log|Y'Y/T| = -16.9964

R²(LR) = 0.952445 R²(LM) = 0.580391

correlation of actual and fitted

Dexch	Korkoero
0.49062	0.96836

EQ(2) Estimating the unrestricted reduced form by OLS (using uip.in7)

The present sample is: 1990 (6) to 2002 (5)

URF Equation 1 for Dexch

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	0.32970	0.086238	3.823	0.0002
Dexch_2	-0.24999	0.087741	-2.849	0.0051
Dexch_3	0.075281	0.084191	0.894	0.3728
Korkoero_1	-2.8377	0.86696	-3.273	0.0013
Korkoero_2	1.3898	1.2105	1.148	0.2529
Korkoero_3	1.2284	0.86988	1.412	0.1602
Constant	0.0015351	0.0018629	0.824	0.4113

\sigma = 0.0221289 RSS = 0.06708744864

URF Equation 2 for Korkoero

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	-0.012887	0.0084931	-1.517	0.1315
Dexch_2	0.0086738	0.0086412	1.004	0.3173
Dexch_3	0.011297	0.0082915	1.363	0.1753
Korkoero_1	1.0038	0.085382	11.757	0.0000
Korkoero_2	-0.085700	0.11922	-0.719	0.4735
Korkoero_3	0.054545	0.085670	0.637	0.5254
Constant	-5.8952e-005	0.00018346	-0.321	0.7485

\sigma = 0.00217936 RSS = 0.0006506976959

correlation of URF residuals

	Dexch	Korkoero
Dexch	1.0000	
Korkoero	-0.19044	1.0000

standard deviations of URF residuals

	Dexch	Korkoero
	0.022129	0.0021794

loglik = 1441.137 log|\Omega| = -20.0158 |\Omega| = 2.02886e-009 T = 144
log|Y'Y/T| = -16.9964
R²(LR) = 0.95117 R²(LM) = 0.570126

correlation of actual and fitted

	Dexch	Korkoero
	0.46967	0.96831

LR-testisuure: 2*(1443.0426-1441.137)=3.8 [p-arvo 0.44]
=====

EQ(3) Estimating the unrestricted reduced form by OLS (using uip.in7)
The present sample is: 1990 (6) to 2002 (5)

URF Equation 1 for Dexch

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	0.34056	0.082993	4.104	0.0001
Dexch_2	-0.20976	0.082094	-2.555	0.0117
Korkoero_1	-2.7410	0.86506	-3.169	0.0019
Korkoero_2	2.4588	0.85480	2.876	0.0047
Constant	0.0016093	0.0018636	0.864	0.3893

\sigma = 0.0222052 RSS = 0.06853693083

URF Equation 2 for Korkoero

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	-0.013896	0.0081565	-1.704	0.0907
Dexch_2	0.013388	0.0080682	1.659	0.0993
Korkoero_1	1.0168	0.085018	11.960	0.0000
Korkoero_2	-0.050843	0.084009	-0.605	0.5460
Constant	-4.2524e-005	0.00018315	-0.232	0.8167

\sigma = 0.00218232 RSS = 0.0006619891791

correlation of URF residuals

	Dexch	Korkoero
Dexch	1.0000	
Korkoero	-0.17048	1.0000

standard deviations of URF residuals

Dexch	Korkoero
0.022205	0.0021823

loglik = 1437.8231 log|\Omega| = -19.9698 |\Omega| = 2.12442e-009 T = 144
log|Y'Y/T| = -16.9964
R²(LR) = 0.94887 R²(LM) = 0.560016

correlation of actual and fitted

Dexch	Korkoero
0.45139	0.96775

LR-testisuure: 2*(1441.137-1437.8231)=6.63 [p-arvo 0.16]
=====

EQ(4) Estimating the unrestricted reduced form by OLS (using uip.in7)
The present sample is: 1990 (6) to 2002 (5)

URF Equation 1 for Dexch

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	0.32104	0.081245	3.952	0.0001
Korkoero_1	-0.23875	0.23210	-1.029	0.3054
Constant	0.0012115	0.0019419	0.624	0.5337

\sigma = 0.0231813 RSS = 0.07576979951

URF Equation 2 for Korkoero

Variable	Coefficient	Std.Error	t-value	t-prob
Dexch_1	-0.010772	0.0076784	-1.403	0.1628
Korkoero_1	0.96069	0.021935	43.797	0.0000
Constant	-2.3107e-005	0.00018353	-0.126	0.9000

\sigma = 0.00219083 RSS = 0.0006767616602

correlation of URF residuals

	Dexch	Korkoero
Dexch	1.0000	
Korkoero	-0.20045	1.0000

standard deviations of URF residuals

Dexch	Korkoero
0.023181	0.0021908

loglik = 1429.8395 log|\Omega| = -19.8589 |\Omega| = 2.37354e-009 T = 144
log|Y'Y/T| = -16.9964
R²(LR) = 0.942874 R²(LM) = 0.517497

correlation of actual and fitted

	Dexch	Korkoero
	0.34601	0.96702

LR-testisuure: $2*(1437.8231-1429.8395)=15.97$ [p-arvo 0.0031]

=====

EQ(5) Estimating the model by FIML (using uip2.in7)

The present sample is: 1990 (6) to 2002 (5)

Equation 1 for Korkoero

Variable	Coefficient	Std.Error	t-value	t-prob	HCSE
Korkoero_1	0.96713	0.022144	43.675	0.0000	0.020757
Dexch_1	-0.014783	0.0080057	-1.847	0.0669	0.0073237
Dexch_2	0.013344	0.0080496	1.658	0.0996	0.0080470
Constant	-3.9605e-005	0.00018267	-0.217	0.8287	---

\sigma = 0.00217737

Equation 2 for Dexch

Variable	Coefficient	Std.Error	t-value	t-prob	HCSE
Korkoero_1	-2.6544	0.85022	-3.122	0.0022	0.95181
Korkoero_2	2.3702	0.83927	2.824	0.0054	0.98319
Dexch_1	0.34210	0.082661	4.139	0.0001	0.088735
Dexch_2	-0.20968	0.081804	-2.563	0.0114	0.084545
Constant	0.0016043	0.0018570	0.864	0.3891	---

\sigma = 0.0221266

loglik = 1437.6336 $\log|\Omega| = -19.9671$ $|\Omega| = 2.13002e-009$ T = 144

LR test of over-identifying restrictions: $\chi^2(1) = 0.378955$ [0.5382]

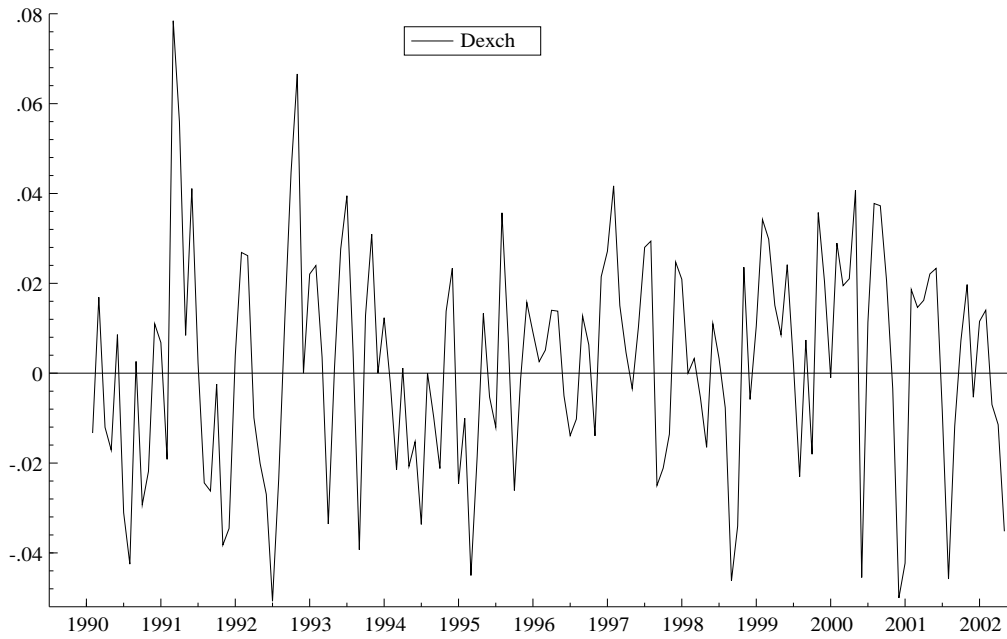
correlation of residuals

	Korkoero	Dexch
Korkoero	1.0000	
Dexch	-0.17070	1.0000

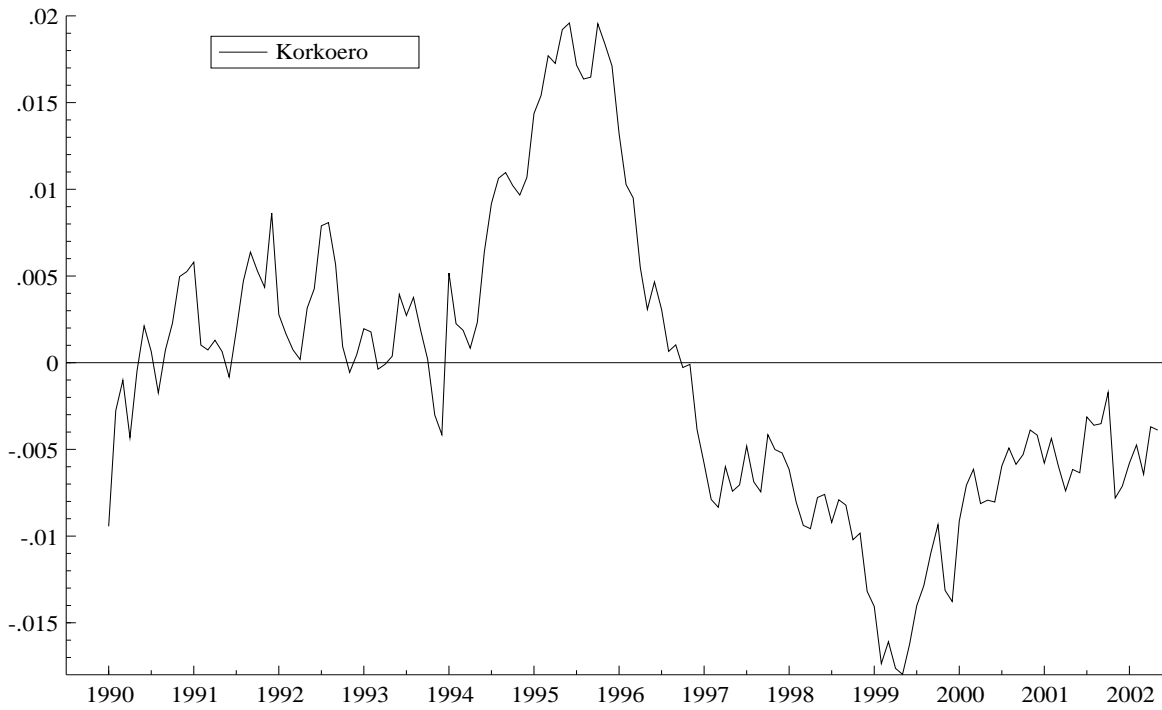
Eigenvalues of companion matrix

real	complex	modulus
0.9674	0.0000	0.9674
-0.1380	0.0000	0.1380
0.2400	0.4234	0.4867
0.2400	-0.4234	0.4867

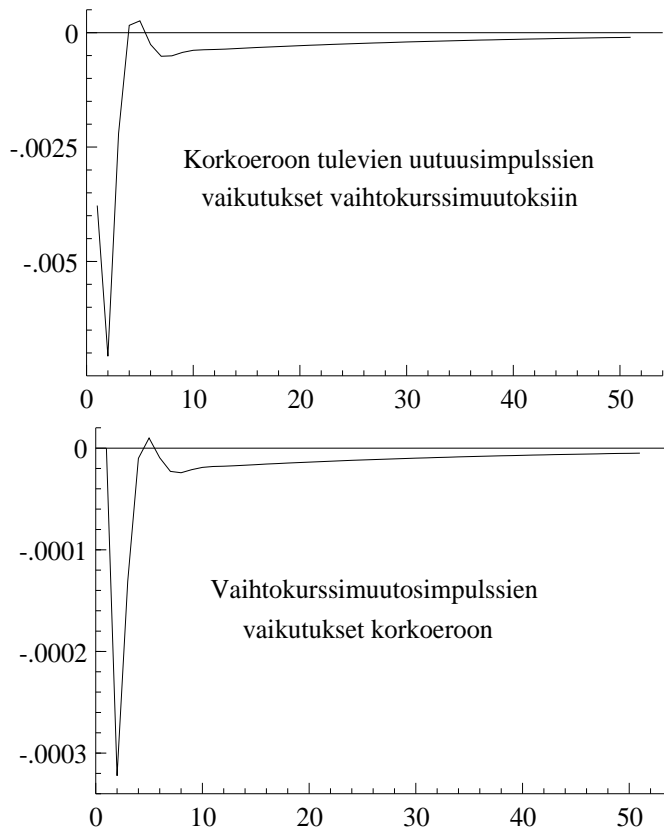
Kuva 7.1: Vaihtokurssimuutosten $Dexch_t$ aikaura



Kuva 7.2: Muuttujan $Korkoero_t$ aikaura



Kuva 7.3: Impulssivaste-estimaatit $\hat{\psi}_{21}^{(\nu)}$ ja $\hat{\psi}_{12}^{(\nu)}$ $\nu = 0, 1, \dots, 50$



■

Huomautus 7.11: Palataan vielä lopuksi VAR-mallien stabiilisuusominaisuuksiin. Merkitään karakteristisen yhtälön (7.30) itseisarvoltaan suurinta juurta symbolilla s_{\max} . Jos $|s_{\max}| < 1$, sanotaan mallia lauseen 7.2 tuloksen mukaisesti *stabiiliksi*. Jos taas $|s_{\max}| > 1$, sanotaan mallia huomautuksen 7.6 mukaisesti *epästabiiliksi*.

Jos taas $s_{\max} = 1$, on siis $\det \Phi(1) = 0$ ja matriisi

$$\Phi(1) = I - \Phi_1 - \dots - \Phi_p$$

on singulaarinen. Mallin (7.29) mukaan odotusarvoista EY_t muodostuva vektorijono toteuttaa ”differenssiyhtälön”

$$\Phi(L) EY_t = 0 \quad t = 1, 2, \dots$$

Näin ollen ”pitkän tähtäimen tasapainotiloina” $Y_\infty = \lim_{t \rightarrow \infty} EY_t$ tulevat kysy-

mykseen kaikki ne vektorit $Y_\infty \in \mathcal{R}^K$, joille

$$(I - \Phi_1 - \dots - \Phi_p)Y_\infty = \Phi(L)Y_\infty = 0 \quad . \quad (7.47)$$

Näihin kysymyksiin palataan seuraavassa luvussa, jossa kaikkien mahdollisten tasapainotilojen $Y_\infty = \lim_{t \rightarrow \infty} EY_t$ muodostamaa joukkoa tullaan kutsumaan $\{Y_t\}$ -prosessin *attraktoriksi*.

Jos siis VAR(p)-prosessi on *stabiili* ($|s_{\max}| < 1$), sisältää attraktori vain *yhden pisteen* (origon). *Epästabiiliin* malliin liittyvällä prosessilla taas *ei ole attraktoria lainkaan*.

Jos $s_{\max} = 1$ on c -kertainen karakteristisen yhtälön (7.30) juuri ($c \geq 1$), muodostuu attraktori *kaikista* yhtälön (7.47) ratkaisuista, ts. lineaarikuvauksen $\Phi(1)$ *ytimeistä* $\mathcal{Y}(\Phi(1))$. Attraktori muodostaa siis \mathcal{R}^K :n c -ulotteisen alivaruuden. ■

8 YHTEISINTEGROITUVUUSTEORIAN PERUSTEET

Tässä luvussa tutkitaan *trendipitoisten* aikasarjojen välisiä riippuvuuksia. VAR-mallit (7.29), joissa $s_{\max} = 1$, tarjoavat trendipitoisten sarjojen kuvaamisessa ai-
van uusia mahdollisuuksia. Eräissä tapauksissa voidaan yksikköjuuria sisältävien VAR-
mallien avulla kuvata (yhdessä ja samassa mallissa) niin vertailtaviin aika-
sarjoihin sisältyvät trendit, sarjojen väliset lyhyen tähtäimen vuorovaikutukset
kuin sarjojen väliset pitkän tähtäimen tasapainorelaatiotkin.

Tarkastellaan muuttujakokonaisuutta $Z_t = (Y_t' \ X_t')'$ ($Z_t \in \mathcal{R}^p$, $Y_t \in \mathcal{R}^K$ ja $X_t \in \mathcal{R}^m$, $p = K + m$), jossa Y_t -muuttujat ajatellaan endogeenisiksi ja X_t -
muuttujat eksogeenisiksi.

Määritelmä 8.1: Stokastista prosessia $\{Z_t\}$ sanotaan *trendipitoiseksi*, jos prosessi $\{Z_t + W\}$ on epästationäärinen, olipa W mikä yksittäinen satunnaismuuttuja tahansa. ■

Määritelmä 8.2: Trendipitoisen sarjan $\{Z_t\}$ sanotaan olevan *integraatioastetta* d ($Z_t \sim I(d)$), jos

$$\nabla^{d-1}(Z_t - EZ_t)$$

on trendipitoinen, mutta

$$\nabla^d(Z_t - EZ_t)$$

on stationäärinen. (Tässä $d \in \mathcal{N}$ on jokin luonnollinen luku ja $\nabla = 1 - L$ tarkoittaa differenssioperaattoria.) ■

Määritelmä 8.3: Integraatioastetta d olevaa prosessia $\{Z_t\}$ ($d \geq 1$) sanotaan *yhteisintegroituneeksi* astein d, b ($Z_t \sim C(d, b)$) yhteisintegraatiovektorina $\beta \in \mathcal{R}^p$, jos

$$\beta' Z_t \sim I(d - b) \quad , \quad b \geq 1 \quad .$$

■

Määritelmä 8.4: Oletetaan, että yksiulotteiset satunnaismuuttujat $\varepsilon_1, \dots, \varepsilon_n$ ovat toisistaan riippumattomia ja samalla tavalla jakautuneita ($\varepsilon_t \sim i.i.d.$, $t = 1, \dots, n$) ja että $E\varepsilon_t = 0$. Tarkastellaan kumulatiivisia summia

$$w_t = \varepsilon_1 + \dots + \varepsilon_t \quad t = 1, 2, \dots \quad . \quad (8.1)$$

Tyyppiä (8.1) olevia prosesseja $\{w_t\}$ sanotaan *random walk*-prosesseiksi. ■

Huomautus 8.1: On helppo huomata, että prosessi $\{w_t\}$ ei ole stationäärinen, vaan *trendipitoinen* määritelmän 8.1 mielessä. Toisaalta $\nabla w_t = \varepsilon_t$ on varmasti stationäärinen, joten $w_t \sim I(1)$. Random walk -prosessit voidaan tietysti ajatella määritellyiksi AR(1)-mallin

$$w_t = w_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim i.i.d. \quad (8.1')$$

avulla. Mallin (8.1') karakteristisen yhtälön ainoa ratkaisu olisi luonnollisesti $s_{\max} = 1$, joten random walk - prosessia $\{w_t\}$ tarkastelemalla voi samalla ymmärtää, millaisia kehityspiirteitä AR- ja VAR -prosesseihin liittyy, mikäli $s_{\max} = 1$. Eräs huomion arvoinen w_t - sarjan ominaisuus on sen *pitkä muisti*:

$$E(w_{t+\tau} | w_t) = w_t \quad \text{kaikilla} \quad \tau = 1, 2, \dots \quad ,$$

joten w_t :n ”regressiokerroin” $w_{t+\tau}$:ta ennustettaessa ei lähestykään nollaa τ :n kasvaessa. Stationääristen sarjojen yhteydessä näin aina tapahtuu (autokorrelaatiot $\rightarrow 0$ vertailtavien havaintojen aikaeron kasvaessa), joten stationäärisiä sarjoja voidaan pitää ”lyhytmuistisina”. ■

Määritelmä 8.5: Jos malliin (8.1') lisätään vakiotermi ($\delta \neq 0$),

$$w_t^* = w_{t-1}^* + \delta + \varepsilon_t \quad \varepsilon_t \sim i.i.d.(0, \sigma^2) \quad , \quad (8.2)$$

kutsutaan syntyvää $\{w_t^*\}$ -prosessia nimellä *random walk with drift*. ■

Huomautus 8.2: Mallin (8.2) mukaan

$$w_t^* = w_o^* + \delta \cdot t + (\varepsilon_1 + \dots + \varepsilon_t) \quad , \quad \varepsilon_t \sim i.i.d.(0, \sigma^2) \quad t = 1, 2, \dots \quad (8.2')$$

joten $w_t = \varepsilon_1 + \dots + \varepsilon_t$:n lisäksi w_t^* :hen voi sisältyä *lineaarinen, deterministinen trendi*. Sarjojen w_t ja w_t^* tyypillisissä aikauria on yritetty havainnollistaa oheisissa kuvioissa.

Jos verrataan mallin (8.2') antamaa käsitystä kuvattavan aikasarjan tulevasta kehityksestä mallin

$$y_t = y_o + \delta \cdot t + \kappa_t \quad , \quad \kappa_t \sim I(0) \quad , \quad E\kappa_t \equiv 0 \quad (8.3)$$

antamaan käsitykseen, huomataan, että malli (8.2') on eräessä mielessä väljempi ja sisältää vähemmän sarjan tulevaa kehitystä koskevia oletuksia. ■

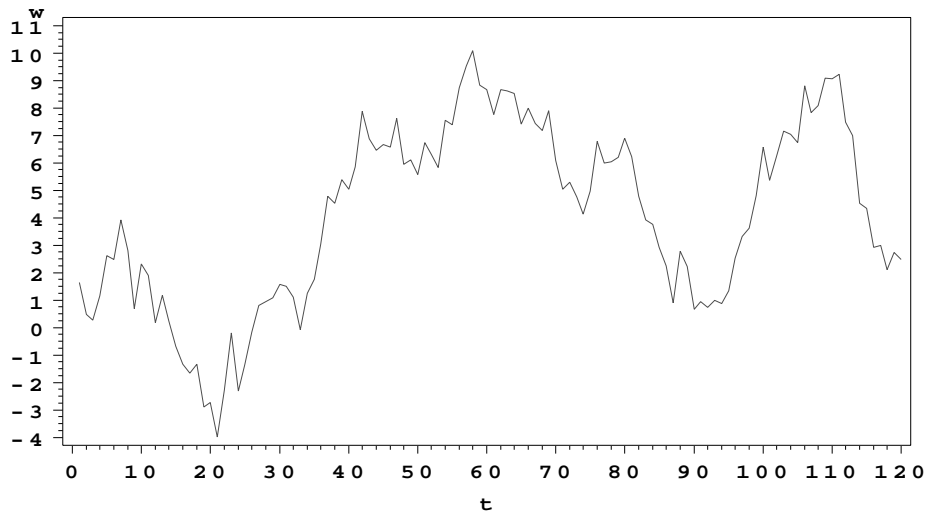
Huomautus 8.3: Tarkastellaan kaksiulotteista aikasarjaa, jonka komponentit ovat muotoa

$$\begin{cases} y_{1t} = \beta w_t^* + \kappa_{1t} \\ y_{2t} = w_t^* + \kappa_{2t} \end{cases} \quad (8.4)$$

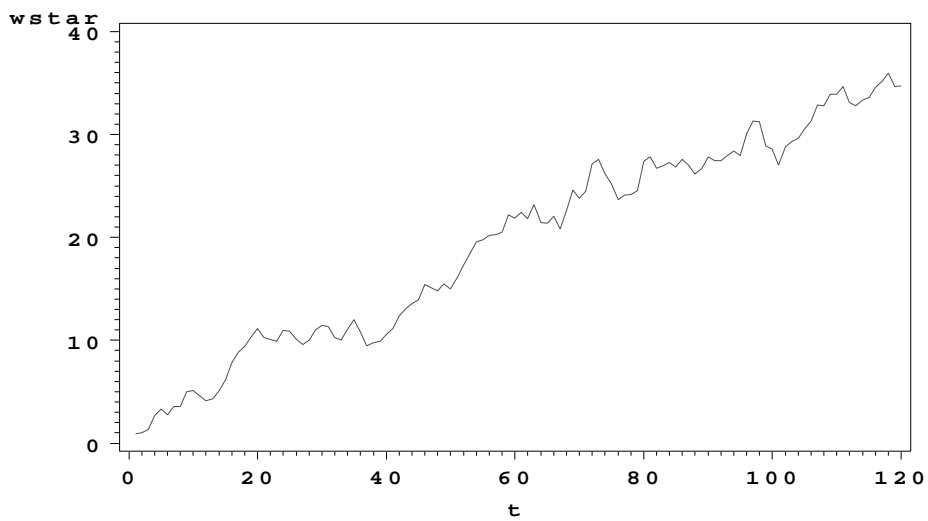
jossa $w_t^* \sim I(1)$ ja $\kappa_{1t}, \kappa_{2t} \sim I(0)$. Tällöin ilmeisesti

$$(1 \quad -\beta) \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \beta(w_t^* - w_t^*) + (\kappa_{1t} - \beta\kappa_{2t}) \sim I(0) \quad ,$$

Kuva 8.1: Tyypillinen random walk -prosessin aikaura (ei driftiä)



Kuva 8.2: Tyypillinen random walk -prosessin aikaura, kun driftiparametri $\delta > 0$



joten yhtälöllä (8.4) määritelty sarja $Y_t = (y_{1t} \ y_{2t})'$ olisi yhteisintegroitunut ($Y_t \sim I(1,1)$) määritelmän 8.3 mielessä. Nähdään siis, että sarjat ovat yhteisintegroituneita ainakin silloin, kun niiden trendiosat (stokastiset ja deterministiset) ovat lineaarisesti toisistaan riippuvia. ■

Määritelmä 8.6: Jatkuva-aikaista p -ulotteista stokastista prosessia $\{B(\tilde{t})\}$ ($0 \leq \tilde{t} \leq 1$) sanotaan p -ulotteiseksi (standardoiduksi) *Brownin liikkeeksi*, jos se toteuttaa seuraavat ehdot:

- 1^o $B(0) = 0$
 2^o Jos $0 \leq \tilde{t}_1 \leq \dots \leq \tilde{t}_k \leq 1$, ovat lisäykset $(B(\tilde{t}_2) - B(\tilde{t}_1)), \dots, (B(\tilde{t}_k) - B(\tilde{t}_{k-1}))$ toisistaan riippumattomia ja

$$B(\tilde{t}_j) - B(\tilde{t}_{j-1}) \sim N_p(0, (\tilde{t}_j - \tilde{t}_{j-1})I) \quad j = 1, \dots, k \quad .$$

- 3^o $B(\tilde{t})$ -realisaatiot ovat jatkuvia todennäköisyydellä 1. ■

Huomautus 8.4: Brownin liike $B(\tilde{t})$ toimii eräänlaisena random walk- prosessin asympotoottisena miniatyyrikuvana, sillä jos diskreettiaikaiset prosessit w_{1t}, \dots, w_{pt} ovat toisistaan riippumattomia ja tyyppiä (8.1), on keskeisen raja-arvolauseen mukaan

$$\frac{1}{\sqrt{n}}W_n = \frac{1}{\sqrt{n}} \left(\frac{w_{1n}}{\sigma_{\varepsilon_1}} \quad \dots \quad \frac{w_{pn}}{\sigma_{\varepsilon_p}} \right)' \underset{\text{asympt.}}{\sim} N_p(0, I) \quad (8.5)$$

kun $n \rightarrow \infty$.

Jos \tilde{t} on mielivaltainen luku väliltä $[0, 1]$, voidaan määritellä prosessi

$$\begin{aligned} B(\tilde{t}) &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}}W_n \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{[\tilde{t}n]} \left(\frac{w_{1t}}{\sigma_{\varepsilon_1}} \quad \dots \quad \frac{w_{pt}}{\sigma_{\varepsilon_p}} \right)' \end{aligned} \quad (8.6)$$

jossa $[\tilde{t}n] = \max \{ \nu \in \mathcal{N} \mid \nu \leq \tilde{t}n \}$, ts. suurin kokonaisluku, joka on $\leq \tilde{t}n$.

Tuloksen (8.5) perusteella kaavalla (8.6) määritelty jatkuva-aikainen prosessi $\{B(\tilde{t})\}$ on p -ulotteinen Brownin liike.

Koska VAR-prosessin käytös muistuttaa random walk-prosessin käytöstä, mikäli $s_{\max} = 1$, voidaan siis Brownin liikettä pitää myös VAR-prosessin ”asympotoottisen käytöksen miniatyyrikuvana” tilanteessa $s_{\max} = 1$. ■

Huomautus 8.5: Ajatellaan, että vektorin Z_t komponentit olisivat yhteisintegroituneita määritelmän 8.3 mielessä, ts. $Z_t \sim C(d, b)$. Tällöin on muistettava, että yhteisintegraatiovektoreita β , joille

$$\beta' Z_t \sim I(d - b) \quad ,$$

saattaa löytyä *useampia*. Toisistaan lineaarisesti riippumattomien yhteisintegraatiovektoreiden lukumäärää sanotaan $\{Z_t\}$:n *yhteisintegraatioasteeksi*. ■

Lause 8.1 (Grangerin esityslause) : Jos $Z_t \sim C(1, 1)$ ja $\{Z_t\}$:n yhteisintegraatioaste on r , niin

1^o Sarjalle Z_t on löydettävissä VARMA-esitys

$$\Phi(L)Z_t = \mu + \theta(L)\varepsilon_t \quad , \quad (8.7)$$

jossa

$$\begin{aligned} \Phi(L) &= I - \Phi_1 L - \dots - \Phi_k L^k && \text{on matriisipolynomi} \quad , \\ \text{cov}(\varepsilon_t, \varepsilon_{t+\nu}) &= 0 && \text{kaikilla} \quad \nu \neq 0 \quad , \\ \theta(L) &= 1 - \theta_1 L - \dots - \theta_q L^q && \text{on skalaaripolynomi} \quad \text{ja} \\ \text{rank}(\Phi(1)) &= r && . \end{aligned}$$

(Apulauseen 1.1 mukaisesti tällöin on olemassa $p \times r$ -matriisit α ja β siten, että $\Phi(1) = -\alpha\beta'$.)

2^o Malli (8.7) voidaan esittää myös virheenkorjausmallin (ks. luku 4.3) muodossa

$$\nabla Z_t = -\Phi(1)Z_{t-1} + \Gamma_1 \nabla Z_{t-1} + \dots + \Gamma_{k-1} \nabla Z_{t-k+1} + \mu + \theta(L)\varepsilon_t \quad , \quad (8.8)$$

jossa

$$\begin{aligned} \Phi(1) &= I - \sum_{j=1}^k \Phi_j = -\alpha\beta' && \text{ja} \\ \Gamma_j &= -(\Phi_{j+1} + \dots + \Phi_k) && j = 1, \dots, k-1 \quad . \end{aligned}$$

Todistus: Puuttumatta esitystavan (8.7) johtamiseen todetaan vain (8.7):n ja (8.8):n välinen yhtäpitävyys:

$$(8.8) \quad \iff$$

$$\begin{aligned} Z_t &= Z_{t-1} - \left(I - \sum_{j=1}^k \Phi_j \right) Z_{t-1} - \sum_{j=1}^{k-1} \left(\sum_{\nu=j+1}^k \Phi_\nu (Z_{t-j} - Z_{t-j-1}) \right) + \mu + \theta(L)\varepsilon_t \\ &= \Phi_1 Z_{t-1} + \dots + \Phi_k Z_{t-k} + \mu + \theta(L)\varepsilon_t \end{aligned}$$

$$\iff (8.7) \quad .$$

■

Huomautus 8.6: Varsin usein voidaan olettaa (ainakin, jos k on valittu riittävän suureksi), että $q = 0$. Tällöin siis lauseen 8.1 mukaan

$$\begin{aligned} Z_t &\sim C(1,1) \\ \iff Z_t &:\text{lle on olemassa VAR-esitys} \\ \iff Z_t &:\text{lle on olemassa virheenkorjausesitys (8.8)} \end{aligned}$$

Myös muunlaisia esitystapoja $C(1,1)$ -prosesseille voidaan johtaa, mutta VAR- ja ECM-esitykset ovat osoittautuneet tulkinnallisesti hyödyllisimmiksi. ■

Huomautus 8.7: Oletetaan jatkossa koko ajan, että $q = 0$. Jos parametrit β (ts. yhteisintegroituvuusrelaatiot) tunnettaisiin muodossa (8.8), voitaisiin loput parametrit $\alpha, \Gamma_1, \dots, \Gamma_{k-1}$ ja μ estimoida OLS:illa. Toisaalta β :n alkiot voitaisiin luonnollisesti estimoida minimoimalla lauseketta

$$\text{tr} \left(\beta' \sum_{t=1}^n (Z_t - \bar{Z})(Z_t - \bar{Z})' \beta \right), \quad (8.9)$$

kunhan β :n sarakkeet on sopivasti normitettu. Huomattakoon, että (8.9):n minimointi vastaa OLS:in käyttöä, kun Z_t :n komponentteja ”selitetään” toisillaan. Kaiken kukkuraksi voidaan todistaa, että jos Z_t :n komponenteissa on lineaarinen trendi, pätee eräin lisäehdoin tulos

$$p \lim_{n \rightarrow \infty} n^{(1-\delta)} \left(\text{vec}(\hat{\beta}_{OLS}) - \text{vec}(\beta) \right) = 0 \quad \text{kaikilla} \quad \delta > 0 \quad . \quad (8.10)$$

Tämä tarkoittaa, että β :n OLS-estimaattorit tarkentuvat huomattavasti tavano- maista nopeammin. (Yleensä $\sqrt{n}(\hat{\theta} - \theta) \overset{\text{asympt.}}{\sim} N(0, \Sigma)$.) Tätä ominaisuutta voidaan kutsua vaikkapa *superkonsistenssiksi*.

Tulokseen (8.10) vedoten Engle ja Granger ovat ehdottaneet seuraavaa menet- telyä:

- 1° Estimoidaan 1. yhteisintegroituvuusrelaatio OLS:illa ”selittämällä” jotakin Z_t :n komponenttia kaikilla muilla komponenteilla, ts. rajoitetaan yksi β :n komponentti ykköseksi ja estimoidaan mallin

$$\beta' Z_t = \mu + \kappa_t$$

parametrit OLS:illa. Talletetaan jäännöstermit $\hat{\kappa}_t$.

- 2° Testataan jäännöstermien stationäärisyyttä Durbin-Watson-testisuureella (ts. von Neumannin suhteella (3.2)). Nollahypoteesina on, että κ_t ei ole stationäärinen, ts. että

$$H_o : \quad \rho(1) = 1 \quad .$$

(Nollahypoteesin vallitessa d -testisuureen odotetaan tietysti olevan hyvin pienen.)

3° Jos nollahypoteesi hylätään, katsotaan, että sarjat *ovat* yhteisintegroitu-
neita, jolloin Γ_j -parametrit sekä α -parametrit voidaan estimoida OLS:illa
mallista (8.8).

Kohdassa 2° ehdotetun Durbin-Watson -testin asemesta voidaan käyttää myös
ns. täydennettyä Dickey-Fuller -testiä, jossa mallista

$$\nabla \widehat{\kappa}_t = \phi \widehat{\kappa}_{t-1} + \delta_1 \nabla \widehat{\kappa}_{t-1} + \dots + \delta_{k-1} \nabla \widehat{\kappa}_{t-k+1} + \varepsilon_t$$

estimoidaan OLS:illa ϕ -parametri ja perustetaan yhteisintegroituvuuden testaa-
minen osamäärään

$$\frac{\widehat{\phi}_{OLS}}{\sqrt{\widehat{\text{var}}(\widehat{\phi}_{OLS})}}$$

■

Huomautuksessa 8.7 esitetyn menettelyn asemesta voidaan mallin (8.8) pa-
rametrit estimoida myös ML-periaatteella, kunhan virhetermien ε_t jakauman
muoto on tarkemmin spesifioitu (Johansen). Malliin (8.8) liittyviä parametrisia
hypoteeseja voidaan samalla testata LR-periaatteella, joten ML-lähestymistapa
on eräissä suhteissa edellä esitettyä menettelyä houkuttelevampi. Täydennetään
aluksi mallia (8.8) ($q = 0$) mahdollisilla kausidummyilla muotoon

$$\begin{aligned} \nabla Z_t &= \mu + \Gamma_1 \nabla Z_{t-1} + \dots + \Gamma_{k-1} \nabla Z_{t-k+1} + \Pi Z_{t-1} + \Psi D_t + \varepsilon_t \\ \Pi &= \alpha \beta' \quad (\alpha \text{ ja } \beta \text{ ovat } p \times r \text{ - matriiseita}) \end{aligned} \tag{8.11}$$

jossa D_t tarkoittaa tarvittavista kausidummyista koostuvaa vektoria. Huomat-
takoon, että malli (8.11) voidaan yhtä hyvin kirjoittaa muotoon

$$\begin{aligned} \nabla Z_t &= \mu + \Gamma_1^* \nabla Z_{t-1} + \dots + \Gamma_{k-1}^* \nabla Z_{t-k+1} + \Pi Z_{t-k} + \Psi D_t + \varepsilon_t \\ \text{jossa } \Gamma_j^* &= \Gamma_j + \Pi \quad \text{ja} \quad \Pi = \alpha \beta' \end{aligned} \tag{8.11'}$$

Jos malli (8.11) (tai (8.11')) täydennetään virhetermien normaalisuusoletuksella

$$\varepsilon_t \sim NID_p(0, \Sigma) \quad , \tag{8.12}$$

voidaan osoittaa, että parametrien $\mu, \Gamma_1^*, \dots, \Gamma_{k-1}^*, \Psi, \alpha, \beta$ ML-estimaatit voidaan löytää seuraavalla proseduurilla:

- 1° ”Selitetään” ∇Z_t :tä ja Z_{t-k} :ta muuttujilla $\nabla Z_{t-1}, \dots, \nabla Z_{t-k}$ ja D_t sekä vakiolla. Talletetaan saadut OLS-residuaalit vektoreihin $E_t^{(o)}$ ja $E_t^{(k)}$. Lasketaan näiden residuaalien kovarianssimatriisit

$$S_{ij} = \frac{1}{n} \sum_{t=1} E_t^{(i)} E_t^{(j)'} \quad i, j = o, k \quad . \quad (8.13)$$

- 2° Kullakin kiinteällä β :n arvolla saadaan α :n ML-estimaatit kaavasta

$$\hat{\alpha}(\beta) = S_{ok}\beta \cdot [\beta' S_{kk}\beta]^{-1} \quad . \quad (8.14)$$

Toisaalta β :n ML-estimaatit saadaan ratkaisemalla p . asteen ominaisarvoyhtälö

$$\det(\rho^2 S_{kk} - S'_{ok} S_{oo}^{-1} S_{ok}) = 0 \quad (8.15)$$

ρ^2 :n suhteen. Merkitään ratkaisuja symbolein $\hat{\rho}_1^2 \geq \dots \geq \hat{\rho}_p^2 > 0$ ja vastaavia ominaisvektoreita symbolein $\hat{V}_1, \dots, \hat{V}_p$, ts.

$$\hat{\rho}_i^2 S_{kk} \hat{V}_i = S'_{ok} S_{oo}^{-1} S_{ok} \hat{V}_i \quad i = 1, \dots, p \quad .$$

Tällöin

$$\hat{\beta} = (\hat{V}_1 \quad \dots \quad \hat{V}_p) \quad . \quad (8.16)$$

Lisäksi voidaan osoittaa, että

$$-2 \log L_{\max}(\hat{\alpha}, \hat{\beta}) = n \log \det S_{oo} + n \sum_{i=1}^r \log (1 - \hat{\rho}_i^2) \quad . \quad (8.17)$$

- 3° Parametrien $\mu, \Gamma_1^*, \dots, \Gamma_{k-1}^*, \Psi$ lopulliset ML-estimaatit saadaan ”selittämällä” muuttujaa $\nabla Z_t - \hat{\alpha}\hat{\beta}' Z_{t-k}$ selittäjillä $\nabla Z_{t-1}, \dots, \nabla Z_{t-k+1}$ ja D_t . Estimointimenetelmänä käytetään OLS:ia.

Huomautus 8.8: Mallin (8.11') parametrien estimointi ML-periaatteella merkitsee itse asiassa ns. *kanonisen analyysin* käyttöä: Muodostetaan differenssien ∇Z_t residuaaleista $E_t^{(o)}$ ja tasojen Z_{t-k} residuaaleista $E_t^{(k)}$ sellaiset lineaarikombinaatiot $\alpha_1' E_t^{(o)}$ ja $\beta_1' E_t^{(k)}$, joiden välinen korrelaatio

$$\frac{\alpha_1' S_{ok} \beta_1}{\sqrt{\alpha_1' S_{oo} \alpha_1 \cdot \beta_1' S_{kk} \beta_1}}$$

olisi *mahdollisimman suuri*. Lagrangen kertojatekniikkaa käyttäen on helppo todeta, että tämä tavoite johtaisi yhtälöihin

$$\begin{cases} S_{ok} \beta_1 - 2\lambda_1 S_{oo} \alpha_1 = 0 \\ S_{ok}' \alpha_1 - 2\lambda_2 S_{kk} \beta_1 = 0 \\ \alpha_1' S_{oo} \alpha_1 = 1 \\ \beta_1' S_{kk} \beta_1 = 1 \end{cases},$$

joista yhdistämällä saadaan

$$\frac{1}{2\lambda_1} S_{ok}' S_{oo}^{-1} S_{ok} \beta_1 - 2\lambda_2 S_{kk} \beta_1 = 0$$

eli

$$(\rho^2 S_{kk} - S_{ok}' S_{oo}^{-1} S_{ok}) \beta_1 = 0, \quad ,$$

jossa on merkitty $\rho^2 = 4\lambda_1 \lambda_2$.

Maksimikorrelaatiota haettaessa päädytään siis *ominaisarvoyhtälöön* (8.15) sekä vastaavaan ominaisvektoriin β_1 , joka toteuttaa ehdon

$$\rho^2 S_{kk} \beta_1 = S_{ok}' S_{oo}^{-1} S_{ok} \beta_1.$$

Tällöin tietenkin

$$\begin{aligned} \alpha_1 &= \frac{1}{2\lambda_1} S_{oo}^{-1} S_{ok} \beta_1, \\ 1 &= \alpha_1' S_{oo} \alpha_1 = \frac{1}{4\lambda_1^2} \beta_1' S_{ok}' S_{oo}^{-1} S_{ok} \beta_1 \end{aligned}$$

ja

$$\begin{aligned} \frac{\alpha_1' S_{ok} \beta_1}{\sqrt{\alpha_1' S_{oo} \alpha_1 \cdot \beta_1' S_{kk} \beta_1}} &= \frac{\frac{1}{2\lambda_1} \beta_1' S_{ok}' S_{oo}^{-1} S_{ok} \beta_1}{\sqrt{\frac{1}{4\lambda_1^2} \beta_1' S_{ok}' S_{oo}^{-1} S_{ok} \beta_1 \cdot \beta_1' S_{kk} \beta_1}} \\ &= \sqrt{\frac{\beta_1' S_{ok}' S_{oo}^{-1} S_{ok} \beta_1}{\beta_1' S_{kk} \beta_1}} = \rho_1. \end{aligned}$$

Tästä syystä ρ_1 :tä sanotaan muuttujablokkien $E_t^{(o)}$ ja $E_t^{(k)}$ väliseksi (ensimmäiseksi) *kanoniseksi korrelaatioksi*.

Tämän jälkeen voidaan hakea α_2 ja β_2 siten, että muuttujien $\alpha_2' E_t^{(o)}$ ja $\beta_2' E_t^{(k)}$ välinen korrelaatio maksimoituisi ehdoilla $\alpha_2' S_{oo} \alpha_1 = 0$ ja $\beta_2' S_{kk} \beta_1 = 0$, jne. ■

Huomautus 8.9: Erilaisia Π -matriisin astetta r koskevia hypoteeseja voidaan testata tuloksen (8.17) avulla seuraavasti:
Jos yleismallina pidetään rajoittamatonta VAR(k)-mallia

$$H : \quad \Phi(L)Z_t = \mu + \Psi D_t + \varepsilon_t \quad , \quad \varepsilon_t \sim NID_p(0, \Sigma) \quad (\text{rank}(\Pi) \leq p)$$

ja halutaan sen puitteissa tutkia ECM- mallin (8.11)

$$H_o : \quad \text{rank}(\Pi) \leq r \quad (r < p)$$

realistisuutta, saadaan LR-testisuureeksi (8.17):n mukaan

$$-2 \log \Lambda = -n \sum_{i=r+1}^p \log (1 - \hat{\rho}_i^2) \quad . \quad (8.18)$$

Tätä suuretta kutsutaan yleensä *trace*-testisuureeksi syistä, jotka ilmenevät seuraavista lauseista. ■

Lause 8.2: Hypoteesin H_o vallitessa suureet $n\hat{\rho}_{r+1}^2, \dots, n\hat{\rho}_p^2$ lähestyvät yhteisjakaumaltaan ominaisarvoyhtälön

$$\det \left(\lambda \int_0^1 B(\tilde{t})B(\tilde{t})' d\tilde{t} - \int_0^1 B(\tilde{t}) [dB(\tilde{t})]' \cdot \int_0^1 [dB(\tilde{t})] B(\tilde{t})' \right) = 0 \quad (8.19)$$

ratkaisujen $\lambda_1, \dots, \lambda_{p-r}$ yhteisjakaumaa, jossa $B(\tilde{t})$ tarkoittaa standardoitua $(p-r)$ -ulotteista Brownin liikettä. ■

Kuten tunnettua, on neliömatriisin jälki (trace) laskettavissa po. matriisin ominaisarvojen summana. Toisaalta

$$\log (1 - \hat{\rho}_i^2) \approx -\hat{\rho}_i^2 \quad , \quad (8.20)$$

kun $\hat{\rho}_i^2$ on pieni, joten lauseen 8.2 seuraukset testisuureen (8.18) jakauman kannalta voidaan muotoilla vaikkapa seuraavasti:

Lause 8.3: Hypoteesin H_o vallitessa testisuureen (8.18) asymptoottinen käyttäytyminen on samanlaista kuin suureen

$$\text{tr} \left(\int_0^1 [dB(\tilde{t})] B(\tilde{t})' \cdot \left[\int_0^1 B(\tilde{t})B(\tilde{t})' d\tilde{t} \right]^{-1} \cdot \int_0^1 B(\tilde{t}) [dB(\tilde{t})]' \right) \quad , \quad (8.21)$$

jossa $B(\tilde{t})$ tarkoittaa standardoitua $(p-r)$ -ulotteista Brownin liikettä. ■

Huomautus 8.10: Suureiden (8.21) käyttäytyminen eri $(p-r)$:n arvoilla on kartoitettu simuloimalla. Tärkeimmät prosenttipisteet ilmenevät oheisista taulukoista.

Huomautus 8.11: Mikäli halutaan tutkia sisäkkäisten hypoteesien

$$H_r : \quad \text{rank}(\Pi) = r \quad r = 1, \dots, p-1$$

uskottavuuksia sekventiaalisesti, olisi hypoteesia H_{r-1} hypoteesin H_r puitteissa testaava LR-testisuure muotoa

$$-2 \log \Lambda = -n \log(1 - \hat{\rho}_r^2) \approx n \hat{\rho}_r^2 \quad . \quad (8.22)$$

Tätä testisuuretta sanotaan $\hat{\lambda}_{\max}$ -testisuureeksi ja sen jakaumataulukot on niinkään muodostettu simuloimalla lauseen 8.2 tuloksen mukaisesti. ■

Huomautus 8.12: Mallin (8.11) vakiovektori $\mu \in \mathcal{R}^p$ voidaan tietysti aina parametroida uudelleen muotoon

$$\mu = \alpha\beta_o + \alpha_{\perp}\gamma \quad , \quad (8.23)$$

jossa vektorit $\beta_o \in \mathcal{R}^r$ ja $\gamma \in \mathcal{R}^{p-r}$ toimivat uusina parametreina. Tulkinallisesti β_o tarkoittaisi attraktoriin

$$\beta_o + \beta'Z = 0 \quad (8.24)$$

liittyvää vakiotermiä, kun taas $\alpha_{\perp}\gamma$ edustaisi ∇Z_t :n kehitystä kuvaavan VAR-mallin ”muuta vakioita”. Jos $\alpha_{\perp}\gamma \neq 0$, olisi siis alkuperäisissä Z_t -muuttujissa *lineaarisia trendejä*. Mikäli tällaisia trendejä *ei* haluta sisällyttää malliin, on ilmeisesti vaadittava, että $\gamma = 0$, ts. että $\alpha_{\perp}\mu = 0$. Tämän rajoituksen huomioon ottaminen kuitenkin muuttaa testisuureiden (8.18) ja (8.22) otantajakaumia jonkin verran. (Ks. oheiset taulukot.) ■

Table 1:
 Percentile points of the *trace* - and $\hat{\lambda}_{\max}$ -test statistics
 Unrestricted μ ($\alpha'_{\perp} \mu \neq 0$)

$p - r$	0.50	0.80	0.90	0.95	0.975	0.99
	$\hat{\lambda}_{\max}$					
1	0.447	1.699	2.816	3.962	5.332	6.936
2	6.852	10.125	12.099	14.036	15.810	17.936
3	12.381	16.324	18.697	20.778	23.002	25.521
4	17.719	22.113	24.712	27.169	29.335	31.943
5	23.211	27.899	30.774	33.178	35.546	38.341
	<i>trace</i>					
1	0.447	1.699	2.816	3.962	5.332	6.936
2	7.638	11.164	13.338	15.197	17.299	19.310
3	18.759	23.868	26.791	29.509	32.313	35.397
4	33.672	40.250	43.964	47.181	50.424	53.792
5	52.588	60.215	65.063	68.905	72.140	76.955

Table 2:
 Percentile points of the *trace* - and $\hat{\lambda}_{\max}$ -test statistics
 Restriction $\alpha'_{\perp} \mu = 0$ (i.e. $\mu = \alpha\beta_o$)

$p - r$	0.50	0.80	0.90	0.95	0.975	0.99
	$\hat{\lambda}_{\max}$					
1	3.474	5.877	7.563	9.094	10.709	12.740
2	8.337	11.628	13.781	15.752	17.622	19.834
3	13.494	17.474	19.796	21.894	23.836	26.409
4	18.592	22.938	25.611	28.167	30.262	33.121
5	23.817	28.643	31.592	34.397	36.625	39.672
	<i>trace</i>					
1	3.474	5.877	7.563	9.094	10.709	12.740
2	11.381	15.359	17.957	20.168	22.202	24.988
3	23.243	28.768	32.093	35.068	37.603	40.198
4	38.844	45.635	49.925	53.347	56.449	60.054
5	58.361	66.624	71.472	75.328	78.857	82.969

Huomautus 8.13: Jos mallissa (8.11) (tai (8.11')) halutaan testata hypoteesia, jonka mukaan X_t olisi heikosti eksogeeninen β :n suhteen ($X_t \in \mathcal{R}^{p-K}$, $Y_t \in \mathcal{R}^K$), pitäisi α :n ilmeisesti olla muotoa

$$H_o: \quad \alpha = \begin{pmatrix} \alpha_Y \\ 0 \end{pmatrix} \quad (\alpha_Y \text{ on } K \times r \text{-matriisi}), \quad (8.25)$$

koska muuten β :t esiintyisivät myös likelihood-hajotelmien

$$L_{Y_t, X_t | \mathcal{F}_{t-1}}(\cdot) = L_{Y_t | X_t, \mathcal{F}_{t-1}}(\cdot) L_{X_t | \mathcal{F}_{t-1}}(\cdot)$$

jälkimmäisissä termeissä. (Merkinnällä \mathcal{F}_{t-1} tarkoitetaan informaatiojoukkoa, joka koostuu hetkeen $t-1$ mennessä saaduista havainnoista.) Jos nimittäin ∇X_t -muuttujia kuvaavissa (8.11):n yhtälöissä esiintyisi α -parametreja, voitaisiin ∇X_t :n lyhyen tähtäimen vaihteluista eristää tasapainotilaan (ts. β :aan) liittyvää informaatiota, jolloin X_t -sarja *ei* olisi heikosti eksogeeninen.

Hypoteesia H_o voidaan testata tavanomaisella LR-testisuureella, joksi (8.17):n mukaan saadaan

$$-2 \log \Lambda = n \sum_{i=1}^r \{ \log(1 - \hat{\rho}_{R,i}^2) - \log(1 - \hat{\rho}_i^2) \} \quad (8.26)$$

asympt. $\underset{\sim}{\sim} \chi_{(p-K) \cdot r}^2$.

■

Huomautus 8.14: Jos oletetaan, etteivät jotkin Z_t -vektoriin sisältyvät heikosti eksogeeniset tekijät osallistu "tasapainorelaatioihin"

$$\beta_o + \beta' Z_t = 0 \quad (8.27)$$

lainkaan, on β -matriisin vastaavat vaakarivit vain rajoitettava nolliksi. Tämän kaltaisia hypoteeseja voidaan testata tavanomaisilla LR-testeillä huomautuksen 8.13 tapaan.

■

Esimerkki 8.1: Seuraavassa eräitä tuloksia artikkelista Johansen & Juselius (1992): "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK", *Journal of Econometrics*, vol. 53, ss. 211-244.

Tarkastellaan seuraavia muuttujia:

$Lpwmf_t \longleftrightarrow$ tukkuhintaindeksi Englannissa vuosineljänneksellä t logaritmoituna

$Lpf6_t \longleftrightarrow$ kuuden Englannin kauppakumppanin tukkuhintaindeksin painotettu keskiarvo logaritmoituna

$Le_t \longleftrightarrow$ Englannin punnan vaihtokurssi (valuuttakoriin verrattuna) logaritmoituna

$rtb_t \longleftrightarrow$ korkotaso Englannissa logaritmoituna ($\log(1 + 0.01 r_t)$)

$reu_t \longleftrightarrow$ eurodollarin korko samalla tavalla transformoituna

$difpo_t \longleftrightarrow$ öljyn reaalihinnan logaritminen muutos

Niputetaan viisi ensiksi mainittua muuttujaa vektoriksi

$$Z_t = \begin{pmatrix} Lpwmf_t \\ Lpf6_t \\ Le_t \\ rtb_t \\ reu_t \end{pmatrix}$$

ja sovitetaan aineistoon mallia

$$\nabla Z_t = \mu + \Gamma_1 \nabla Z_{t-1} + C_o difpo_t + C_1 difpo_{t-1} + \Pi Z_{t-2} + \Psi D_t + \varepsilon_t \quad ,$$

$$\varepsilon_t \sim NID_5(0, \Sigma) \quad ,$$

jossa D_t merkitsee kolmesta kausivaihteludummystä koostuvaa vektoria.

Yhteisintegraatioasteen valintaan liittyvät testit

$\hat{\lambda}_{\max}$ -testi

i	$\hat{\rho}_i^2$	$-n \log(1 - \hat{\rho}_i^2)$	$\hat{\lambda}_{\max}$ -testisuureen 0.95-fraktiili
1	0.407	31.33	33.18
2	0.285	20.16	27.17
3	0.254	17.59	20.78
4	0.102	6.48	14.04
5	0.083	5.19	3.84

trace -testi

i	$\hat{\rho}_i^2$	$-n \sum_{\nu=i+1}^p \log(1 - \hat{\rho}_\nu^2)$	trace-testisuureen 0.95-fraktiili
1	0.407	80.75	68.68
2	0.285	49.42	47.21
3	0.254	29.26	29.34
4	0.102	11.67	15.34
5	0.083	5.19	3.84

Trace-testin mukaan sopiva yhteisintegraatioaste voisi olla $r = 2$ (tai $r = 1$).
Yhteisintegraatiovektorit olisivat

$$\hat{\beta}_1 = \begin{pmatrix} -1.073 \\ 0.975 \\ 1.000 \\ 3.620 \\ 2.028 \end{pmatrix} \quad \hat{\beta}_2 = \begin{pmatrix} -0.297 \\ 0.340 \\ 1.000 \\ -10.480 \\ 9.788 \end{pmatrix} .$$

Esimerkkinä hypoteesien testaamisesta tutkitaan, voitaisiinko Englannin kaup-
pakkumppaneiden keskimääräisen hintason logaritmia $Lpwf6_t$ pitää heikosti ek-
sogeenisena α :n ja β :n suhteen. Kun yhteisintegraatioasteeksi on valittu $r = 2$,
saa testisuure (8.26) muodon

$$-2 \log \Lambda = 60 \cdot \log \left(\frac{(1 - 0.400)}{(1 - 0.407)} \cdot \frac{(1 - 0.285)}{(1 - 0.285)} \right) \approx 0.68 ,$$

joka on χ_2^2 -jakaumaan verrattuna varsin pieni. Näin ollen muuttujaa $Lpwf6_t$
voidaan hyvällä omallatunnolla pitää heikosti eksogeenisena. ■

Huomautus 8.15: On ehdottomasti syytä huomata, että jos yhteisintegroitu-
vuusvektoreita β_1, \dots, β_r on useampia ($r > 1$), liittyy mallin (8.11') parametreihin
seuraava identifioituvuusongelma:

Olkoon A mikä tahansa epäsingulaarinen $r \times r$ -matriisi. Tällöin malliin
(8.11') sisältyvä $\Pi = \alpha\beta'$ -matriisi voidaan yhtä hyvin kirjoittaa muotoon
 $\Pi = \alpha A^{-1} \cdot A\beta'$, joten yhteisintegroituvuusvektorisysteemeitä
 $\beta = (\beta_1 \dots \beta_r)$ ja $\beta A'$ vastaa täsmälleen sama *tilastollinen* malli (8.11').
Tästä johtuu, ettei ”pitkän tähtäimen tasapainorelaatioita”

$$\beta' Z_t = \text{vakio}$$

ja

$$A\beta'Z_t = \text{vakio}$$

voida *havaintojen valossa* erottaa toisistaan. Tästä huolimatta niiden *tulkinnat* saattavat olla täysin erilaiset.

Tästä johtuu, että β -vektoreiden mielekäs tulkinta edellyttää riittävää määrää *rakenteellisia rajoituksia*, jotta avoimiksi jätetyt parametrit olisivat identifioitavissa ja jotta ne pystyttäisiin havaintojen avulla estimoimaan. ■

Seuraava määritelmä tarjoaa erään keinon tutkia, milloin β -vektoreille asetetut rajoitukset riittävät takaamaan jäljelle jäävien parametrien identifioitavuuden:

Määritelmä 8.7: Ajatellaan, että yhteisintegroituvuusvektoreille β_1, \dots, β_r asetetaan rajoitukset

$$R'_i \beta_i = 0 \quad i = 1, \dots, r \quad , \quad (8.28)$$

jossa R_i on täysiasteinen $p \times k_i$ -matriisi. (Rajoitusten yhteismäärä on siis $k_1 + \dots + k_r$.)

Olkoon $H_i = R_{i,\perp}$, ts. $R'_i H_i = 0$, jossa H_i on täysiasteinen $p \times (p - k_i)$ -matriisi ($i = 1, \dots, r$).

Mikäli jokaiselle i :n arvolle, jokaiselle $k = 1, \dots, r - 1$ ja jokaiselle indeksijoukolle

$$J_k = \{i_1, \dots, i_k \mid 1 \leq i_1 < \dots < i_k \leq r ; i_\nu \neq i \quad \forall \nu = 1, \dots, k\}$$

pätee

$$\text{rank}(R'_i H_{i_1} \quad \dots \quad R'_i H_{i_k}) \geq k \quad , \quad (8.29)$$

sanotaan rajoituksia (8.28) *yleisesti identifioiviksi* (generically identifying restrictions). ■

Huomautus 8.16: Määritelmän 8.7 ehtojen täyttymistä voidaan kätevimmin tutkia seuraavaan tapaan:

1° Jos $r = 2$, ovat ehdot (8.29) voimassa, jos

$$r_{i,j} = \text{rank}(R'_i H_j) \geq 1 \quad \text{kun} \quad i, j = 1, 2 \quad ; \quad i \neq j \quad .$$

2° Jos $r = 3$, joudutaan tutkimaan seuraavien ehtojen voimassaoloa:

$$r_{i,j} = \text{rank}(R'_i H_j) \geq 1 \quad \text{kun} \quad i, j = 1, 2, 3 \quad ; \quad i \neq j \quad ,$$

ja

$$r_{i,jm} = \text{rank}(R'_i(H_j \ H_m)) \geq 2 \quad \text{jossa } (i, j, m) \text{ on mikä tahansa} \\ \text{lukujen } (1, 2, 3) \text{ permutaatio .} \quad (8.30)$$

Ehdossa (8.30) esiintyvän matriisin aste voidaan kätevimmin selvittää hakemalla *symmetrisen* matriisin

$$G_{i,jm} = (H_j \ H_m)'(I - H_i(H_i'H_i)^{-1}H_i')(H_j \ H_m)$$

ominaisarvot. On nimittäin helppo nähdä, että

$$\text{rank}(G_{i,jm}) = \text{rank}(R'_i(H_j \ H_m)) \quad .$$

■

Huomautus 8.17: Yhteisintegroituvuusvektoreihin β_1, \dots, β_r liittyvä identifioituvuusongelma voidaan periaatteessa ratkaista kirjoittamalla $p \times r$ -matriisi β muotoon

$$\beta = \begin{pmatrix} \beta^o \\ \beta^{oo} \end{pmatrix} \quad ,$$

jossa β^o on $r \times r$ -matriisi ja β^{oo} on $(p-r) \times r$ -matriisi. Huomautuksen 8.15 merkintöjä käyttäen A -matriisiksi voidaan valita vaikkapa $A = (\beta^{o'})^{-1}$, jolloin

$$\beta A' = \begin{pmatrix} I \\ \beta^{oo}\beta^{o-1} \end{pmatrix} \quad . \quad (8.31)$$

Tätä β -matriisin muotoa kutsutaan *reduoiduksi muodoksi*, koska systeemin ”tasapainorajoitteet” tulevat ilmaistuksi siten, että r ensimmäistä Z_t -vektorin komponenttia on ”ratkaistu” $(p-r)$:n viimeisen elementin suhteen. (Tätä esitystapaa käyttää mm. PCFIML tulostuksessaan.)

Muotoa (8.31) olevat rajoitteet $\beta^o = I$ ovat kylläkin yleisesti identifioivia, mutta kovin harvoin tulkinnallisesti mielenkiintoisia. ■

Huomautus 8.18: Vaikka β -vektoreihin liittyviä identifioituvuusongelmia ei olikaan juurta jaksain ratkaistu, voidaan erilaisia β -vektoreita koskevia *hypoteeseja* testata LR-testisuureen (8.17) avulla varsin helposti. Testisuureen $-2 \log \Lambda$ asympotoottinen otantajakauma on tällöin nimittäin χ^2 -tyyppinen. Referenssijakauman vapausasteiden lukumäärä määräytyy hypoteesissa eliminoidujen (oleellisten) parametrien lukumäärän mukaisesti. Koska oleellisten β -parametrien lukumäärän laskeminen on hieman konstikasta, voidaan eri hypoteeseihin liittyviä parametrimääriä laskeskella esim. seuraavan ajattelutavan mukaisesti:

Vapaaseen (rajoittamattomaan) r -ulotteiseen β -avaruuteen liittyvien *oleellisten* parametrien lukumäärä on $r \cdot (p - r)$.

Tämä voidaan nähdä seuraavasti:

Euklidisen avaruuden \mathcal{R}^p r -ulotteisen aliavaruuden 1. ortonormeerattu kantavektori määräytyy $(p - 1)$ parametrin avulla, koska kyseisen kantavektorin komponenteista yksi ratkeaa normitusehdon perusteella. Toisen kantavektorin valintaan liittyy vastaavasti $(p - 2)$ parametria, koska jäljellä olevia suuntia on yhtä vähemmän kuin 1. kantavektoria valittaessa. Tällä tavalla nähdään, että r :n kantavektorin valintaan liittyy yhteensä $(p - 1) + \dots + (p - r)$ parametria. Valittua koordinaatistoa voidaan toisaalta *rotatoida* mielivaltaisesti itse avaruutta siirtämättä. Koska vastaavasti (edellä esitetyn järjestyksen mukaisesti) r -ulotteisen avaruuden r :n kantavektorin valintaan liittyy $(r - 1) + \dots + 1 + 0$ parametria, saadaan \mathcal{R}^p :n r -ulotteisen aliavaruuden paikan määrävien *oleellisten* parametrien kokonaismääräksi

$$\sum_{i=1}^r (p - i) - \sum_{i=1}^r (i - 1) = r \cdot (p - r) \quad . \quad (8.32)$$

Huomattakoon, että tulos on symmetrinen r :n ja $(p - r)$:n suhteen. Näin täytyy ollakin, sillä β -avaruuden paikka voidaan yhtä hyvin ilmaista sen ortogonaalisen komplementin (ts. Z_t -prosessin attraktorin) avulla. ■

LIITE A1

Eräitä välttämättömiä taustatietoja

A1.1 Todennäköisyytlaskentaan liittyviä peruskäsitteitä ja -tuloksia

Reaaliarvoisia *satunnaismuuttujia* merkitään tällä kurssilla pienillä kirjaimilla. Suurilla kirjaimilla merkitään matriiseja ja vektoreita, joista puhutaan tarkemmin luvussa A1.2.

Määritelmä A1.1: Reaaliarvoisen satunnaismuuttujan x *kertymäfunktio*ksi F_x sanotaan funktiota

$$F_x(t) = P(x \leq t) \quad t \in \mathcal{R}^1 \quad .$$

■

On helppo nähdä, että jokaisella kertymäfunktiolla on ominaisuudet

- (i) F_x on kasvava
- (ii) F_x on oikealle jatkuva
- (iii) $F_x(-\infty) = 0$ ja $F_x(\infty) = 1$.

Toisaalta jokainen funktio, jolla on ominaisuudet (i) - (iii), voidaan tulkita jonkin jakauman kertymäfunktioksi. Kun kertymäfunktio tunnetaan, voidaan kaikkien mahdollisten välien todennäköisyydet $P(a < x < b)$ laskea, ja niiden avulla puolestaan voidaan muodostaa kaikkien *tapahtumien* ($x \in A$) todennäköisyydet, kunhan A on muodostettavissa väleistä numeroituvalla määrällä joukko-opillisia alkeisoperaatioita (yhdiste, leikkaus, komplementti). Tapahtumaa ($x \in A$) sanotaan *melkein varmaksi*, mikäli $P(x \in A) = 1$.

Määritelmä A1.2: Satunnaismuuttujan x jakaumaa sanotaan *jatkuva*ksi, jos F_x on *derivoituva* ”melkein kaikkialla”, ts. jos on olemassa integroituva funktio $f_x : \mathcal{R}^1 \rightarrow \mathcal{R}^1$, jolle pätee

$$F_x(t) = P(x \leq t) = \int_{-\infty}^t f_x(u) du \quad \text{jokaisella } t \in \mathcal{R}^1 \quad ,$$

ts. jolle $f_x(t) \equiv F'_x(t) = \frac{d}{dt} F_x(t)$ (korkeintaan numeroituvaa pistejoukkoa lukuun ottamatta).

■

Määritelmä A1.3: Jatkuvan satunnaismuuttujan x *odotusarvo*ksi sanotaan lukua

$$E(x) = \int_{-\infty}^{\infty} t \cdot f_x(t) dt \quad ,$$

mikäli po. integraali on hyvin määritelty. Transformoidun muuttujan $g(x)$ odotusarvoksi sanotaan (vastaavin ehdoin) lukua

$$E(g(x)) = \int_{-\infty}^{\infty} g(t) \cdot f_x(t) dt \quad .$$

Lukuja

$$\alpha_k = \int_{-\infty}^{\infty} t^k \cdot f_x(t) dt \quad , \quad k = 1, 2, \dots$$

sanotaan x :n *origomomenteiksi*, mikäli ne ovat hyvin määriteltyjä.

Merkitään nyt x :n ensimmäistä momenttia (ts. sen odotusarvoa) symbolilla $\mu = \alpha_1$. Tällöin voidaan määritellä x :n *keskusmomentit*

$$\mu_k = E(x - \mu)^k = \int_{-\infty}^{\infty} (t - \mu)^k \cdot f_x(t) dt \quad , \quad k = 2, 3, \dots \quad .$$

Huomattakoon, että μ_k on olemassa jos ja vain jos α_k on olemassa. Lisäksi voidaan helposti nähdä, että α_k :n olemassaolo takaa myös alempien kertalukujen momenttien $\alpha_1, \dots, \alpha_{k-1}$ ja μ_2, \dots, μ_{k-1} olemassaolon.

Erytysmaininnan ansaitsee toinen keskusmomentti

$$\mu_2 = E(x - \mu)^2 = \text{var}(x) = \sigma_x^2 \quad ,$$

jota kutsutaan x :n *varianssiksi*. Sen neliöjuurta σ_x kutsutaan x :n *hajonnaksi*. ■

Määritelmä A1.4: Tarkastellaan nyt *kahta* samassa mallikokonaisuudessa (samassa ”todennäköisyyskentässä”) määriteltyä satunnaismuuttujaa x ja y . Yhdessä niitä voidaan ajatella eräänlaisena kaksiulotteisena ”vektorimuuttujana” $Z = (x \ y)'$. Tämän vektorimuuttujan jakauma (tai x :n ja y :n ”yhteisjakauma”) voidaan karakterisoida ns. *kertymäfunktion*

$$F_Z(s, t) = F_{x,y}(s, t) = P(x \leq s, y \leq t) \quad s, t \in \mathcal{R}^1$$

avulla, ts. kaikki jakaumaan liittyvät todennäköisyydet voidaan johtaa tästä kahden argumentin reaaliarvoisesta funktiosta.

Mikäli kertymäfunktiolla $F_{x,y}(s, t)$ on hyvin määriteltyt, jatkuvat sekaderivaatat

$$f_Z(s, t) = f_{x,y}(s, t) = \frac{\partial^2}{\partial s \partial t} F_{x,y}(s, t)$$

”melkein kaikkialla”, sanotaan yhteisjakaumaa *jatkuvaksi* ja funktiota $f_{x,y}(s, t)$ sen *tiheysfunktiksi*. Tämä nimitys johtuu tietenkin siitä, että $f_{x,y}(s, t) \geq 0$ kaikilla s, t ja

$$F_{x,y}(s, t) = \int_{-\infty}^s \int_{-\infty}^t f_{x,y}(s', t') ds' dt' \quad .$$

Määritelmä A1.5: Jos muuttujien x ja y yhteisjakauma on jatkuva, ovat myös ns. *ehdolliset* jakaumat ($y | x = s$ ja $x | y = t$) jatkuvia ja niiden tiheysfunktiot ovat verrannollisia $f_{x,y}(s, t)$ -pinnan pystysuoriin poikkileikkauskuvioihin

$$f_{y|x=s}(t) \propto f_{x,y}(s, t) \quad , \quad t \in \mathcal{R}^1 \quad \text{ja} \quad f_{x|y=t}(s) \propto f_{x,y}(s, t) \quad , \quad s \in \mathcal{R}^1 \quad .$$

Jotta näiden ehdollisten tiheysfunktioiden normitukset saataisiin oikein suoritetuksi, on määriteltävä

$$f_{y|x=s}(t) = \frac{f_{x,y}(s,t)}{f_x(s)} \quad , \quad t \in \mathcal{R}^1$$

ja

$$f_{x|y=t}(s) = \frac{f_{x,y}(s,t)}{f_y(t)} \quad , \quad s \in \mathcal{R}^1 \quad .$$

Tässä $f_x(s)$ ja $f_y(t)$ edustavat x :n ja y :n ns. *reunajakaumien* tiheysfunktioita, jotka saadaan yhteisjakauman tiheysfunktioista $f_{x,y}(s,t)$ kaavoilla

$$f_x(s) = \int_{-\infty}^{\infty} f_{x,y}(s,t) dt$$

ja

$$f_y(t) = \int_{-\infty}^{\infty} f_{x,y}(s,t) ds \quad .$$

■

Määritelmä A1.6: Muuttujaa x sanotaan muuttujasta y *riippumattomaksi* ($x \perp\!\!\!\perp y$), mikäli kaikki ehdolliset jakaumat $x | y = t$ (kaikilla t) ovat samanlaisia, ts. mikäli käsitys y :n arvosta ei millään tavalla muuta meidän käsitystämme x :n käyttäytymisestä. Kuten helposti nähdään, tämä ehto toteutuu jatkuvien muuttujien yhteydessä silloin ja vain silloin, kun

$$f_{x,y}(s,t) \equiv f_x(s) \cdot f_y(t) \quad \forall \quad s, t \in \mathcal{R}^1 \quad .$$

Tästä nähdään, että riippumattomuusominaisuus on itse asiassa symmetrinen: Jos $x \perp\!\!\!\perp y$, on myös $y \perp\!\!\!\perp x$. Tästä syystä tällöin yleensä sanotaankin x :n ja y :n olevan *toisistaan riippumattomia*. Mikäli muuttujat eivät ole toisistaan riippumattomia, niitä sanotaan toisistaan riippuviksi. ■

Määritelmä A1.7: Muuttujien x ja y johdannaisten $g(x,y)$ odotusarvo määritellään kaavalla

$$Eg(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(s,t) \cdot f_{x,y}(s,t) ds dt$$

Helposti nähdään, että tulos

$$x \perp\!\!\!\perp y \implies E(h_1(x) \cdot h_2(y)) = Eh_1(x) \cdot Eh_2(y)$$

pätee aina, kun kaavassa esiintyvät odotusarvot ovat hyvin määriteltyjä.

Mikäli x :n ja y :n toiset momentit (varianssit) ovat hyvin määriteltyjä, sanotaan x :n ja y :n väliseksi *kovarianssiksi* lukua

$$\sigma_{x,y} = \text{cov}(x,y) = E(x - Ex)(y - Ey) \quad .$$

Havaitaan heti, että

$$x \perp\!\!\!\perp y \implies \text{cov}(x, y) = 0 \quad ,$$

mutta käänteinen tulos ei missään tapauksessa pidä paikkaansa. ■

Määritelmä A1.8: Koska kovarianssin käsite on mittakaavasidonnainen, käytetään lineaarisen riippuvuuden mittarina useimmiten ns. *korrelaatiokerrointa*

$$\rho_{x,y} = \text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\text{var}(y)}} \quad .$$

Edellisen kommentin mukaisesti

$$x \perp\!\!\!\perp y \implies \text{corr}(x, y) = 0 \quad ,$$

mutta

$$\text{corr}(x, y) = 0 \not\Rightarrow x \perp\!\!\!\perp y \quad .$$

Lisäksi todettakoon, että ns. Schwarzin epäyhtälön mukaisesti

$$(\text{cov}(x, y))^2 \leq \text{var}(x) \cdot \text{var}(y) \quad ,$$

joten $-1 \leq \text{corr}(x, y) \leq 1$. Ääriarvot ± 1 merkitsevät itse asiassa sitä, että y :n ja x :n arvot ovat aivan tarkasti samalla suoralla, joten tällaisessa tilanteessa y :n ja x :n yhteisjakauma olisi singulaarinen eikä oikeasti jatkuva. ■

Määritelmä A1.9: Tarkastellaan nyt useampiulotteista ("vektoriarvoista") satunnaismuuttujaa $X = (x_1 \dots x_m)'$. Muuttujan X yhteisjakauman kertymäfunktioilla tarkoitetaan funktiota

$$F_X(t_1, \dots, t_m) = F_X(T) = F_{x_1, \dots, x_m}(t_1, \dots, t_m) = P(x_1 \leq t_1, \dots, x_m \leq t_m) \quad ,$$

$$T = (t_1 \dots t_m)' \in \mathcal{R}^m \quad .$$

Mikäli kertymäfunktioilla F_X on hyvin määritellyt m . kertaluvun sekaderivaatat

$$f_{x_1, \dots, x_m}(t_1, \dots, t_m) = \frac{\partial^m}{\partial t_1 \dots \partial t_m} F_{x_1, \dots, x_m}(t_1, \dots, t_m)$$

"melkein kaikkialla" \mathcal{R}^m :ssä, sanotaan jakaumaa *jatkuvaaksi*.

Jos X :n komponentit ajatellaan jaetuiksi kahteen ositteeseen

$$X = (X^{(1)'} \quad X^{(2)'})' = (x_1 \dots x_k \quad x_{k+1} \dots x_m)' \quad ,$$

voidaan puhua ositteiden $X^{(1)}$ ja $X^{(2)}$ (moniulotteisista) *reunajakaumista*, joiden tiheysfunktiot ovat muotoa

$$f_{x_1, \dots, x_k}(t_1, \dots, t_k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{x_1, \dots, x_k, x_{k+1}, \dots, x_m}(t_1, \dots, t_k, t_{k+1}, \dots, t_m) dt_{k+1} \dots dt_m$$

ja

$$f_{x_{k+1}, \dots, x_m}(t_{k+1}, \dots, t_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{x_1, \dots, x_k, x_{k+1}, \dots, x_m}(t_1, \dots, t_k, t_{k+1}, \dots, t_m) dt_1 \cdots dt_k \quad .$$

Ositteen $X^{(1)}$ ehdollisen jakauman tiheysfunktio olisi muotoa

$$f_{x_1, \dots, x_k | x_{k+1}=t_{k+1}, \dots, x_m=t_m}(t_1, \dots, t_k) = \frac{f_{x_1, \dots, x_k, x_{k+1}, \dots, x_m}(t_1, \dots, t_k, t_{k+1}, \dots, t_m)}{f_{x_{k+1}, \dots, x_m}(t_{k+1}, \dots, t_m)} \quad .$$

Esimerkiksi kolmen yksiulotteisen satunnaismuuttujan $(x \ y \ z)'$ yhteiskäyt-
täytymistä tarkasteltaessa voidaan määritellä ns. *ehdollinen riippumattomuus*
($\langle x \ \underline{\underline{\parallel}} \ y \rangle | z$) ominaisuutena

$$f_{x,y|z=u}(s, t) \equiv f_{x|z=u}(s) \cdot f_{y|z=u}(t) \quad \forall \ s, t \in \mathcal{R}^1 \quad ,$$

olipa u mikä tahansa. Tämä ominaisuus on käytännön ongelmia tutkittaessa erityisen tärkeä. Esimerkiksi tilanteessa, jossa x :llä ja y :llä kuvattavilla empiirillä suureilla ei ole mitään suoraa yhteyttä keskenään, mutta niihin molempiin vaikuttaa yhteinen, z :lla kuvattava taustatekijä, ilmenee x :n ja y :n käyttäytymisessä usein näennäistä yhdenmukaisuutta. Käytetyn matemaattisen mallin kannalta tämä merkitsisi sitä, että x ja y näyttäisivät kaksiulotteisen yhteisjakaumansa valossa riippuvan toisistaan, mutta siitä huolimatta x ja y olisivat (kolmiulotteisen yhteisjakauman valossa) *ehdollisesti* toisistaan riippumattomia, kun z on kiinnitetty ($x \ \underline{\underline{\not\parallel}} \ y$, mutta $\langle x \ \underline{\underline{\parallel}} \ y \rangle | z$). ■

Määritelmä A1.10: Tarkastellaan (jatkuvaa) vektoriarvoista satunnaismuuttujaa $Y = (y_1 \ \dots \ y_m)'$, jonka komponenttien varianssit $\text{var}(y_j) \quad j = 1, \dots, m$ ovat äärellisiä. Muuttujan Y *kovarianssimatriisilla* tarkoitetaan $m \times m$ -neliomatriisia

$$\text{cov}(Y) = E[(Y - EY)(Y - EY)'] = \begin{pmatrix} \text{var}(y_1) & \dots & \text{cov}(y_1, y_m) \\ \vdots & \dots & \vdots \\ \text{cov}(y_1, y_m) & \dots & \text{var}(y_m) \end{pmatrix}$$

Määritelmänsä mukaisesti kovarianssimatriisit ovat aina symmetrisiä ja ei-negatiivisesti definiittejä.

Jos A on kiinteä $n \times m$ -matriisi, voidaan helposti todeta, että

$$\text{cov}(AY) = E[(AY - A E(Y))(AY - A E(Y))'] = A \text{cov}(Y) A' \quad .$$

(Kovarianssimatriisin ei-negatiivinen definiittisyys nähdään tämän tuloksen perusteella välittömästi ajattelemalla, että $n = 1$, sillä tällöin $\text{cov}(AY) = \text{var}(AY)$, joka on aina ei-negatiivinen.) ■

Määritelmä A1.11: Päätymättömän satunnaismuuttujajonon $x_1, x_2, \dots, x_n, \dots$ sanotaan suppenevan *todennäköisesti* (in probability) kohti vakiota μ , mikäli

$$\lim_{n \rightarrow \infty} P(|x_n - \mu| \geq \varepsilon) = 0 \quad ,$$

olipa $\varepsilon > 0$ kuinka pieni tahansa. Todennäköistä konvergenssia kutsutaan usein myös satunnaismuuttujajonon *heikoksi* konvergenssiksi. Siitä käytetään joko merkintää

$$p \lim_{n \rightarrow \infty} x_n = \mu \quad \text{tai merkintää} \quad x_n \xrightarrow{p} \mu .$$

Satunnaismuuttujajonon $x_1, x_2, \dots, x_n, \dots$ sanotaan suppenevan jakaumaltaan kohti rajamuuttujaa x , mikäli

$$\lim_{n \rightarrow \infty} F_{x_n}(t) = F_x(t) \quad \text{jokaisessa } F_x : \text{n jatkuvuusasteessa } t .$$

(Funktio F_x oletetaan olevan hyvin määritelty kertymäfunktio, mutta rajajakauma ei välttämättä ole jatkuva, vaikka jonon kaikki satunnaismuuttujat x_n olisivatkin jatkuvia.)

Huomattakoon erityisesti, että jos $p \lim_{n \rightarrow \infty} x_n = \mu$, lähestyy kertymäfunktiojono F_{x_n} rajafunktiota

$$\lim_{n \rightarrow \infty} F_{x_n}(t) = \begin{cases} 0 & \text{kun } t < \mu \\ 1 & \text{kun } t \geq \mu \end{cases} ,$$

ts. muuttujajonon jakaumat lähestyvät rajajakaumaa, joka on konsentroitunut yhteen pisteeseen μ , ts. $P(x = \mu) = 1$. ■

Lause A1.1: (Suurten lukujen laki)

Oletetaan, että $x_1, x_2, \dots, x_n, \dots$ on jono toisistaan riippumattomia, samalla tavalla jakautuneita satunnaismuuttujia (merkitään $x_n \sim i.i.d.$), joilla on äärelliset toisen kertaluvun momentit. Merkitään $\mu = EX_n$ ja $\sigma^2 = \text{var}(x_n)$. Tällöin

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \xrightarrow{p} \mu$$

kun $n \rightarrow \infty$. ■

Lause A1.2: (Keskeinen raja-arvolause)

Jos $x_n \sim i.i.d. (F_x)$, $EX_n \equiv \mu$ ja $\text{var}(x_n) = E(x_n - \mu)^2 \equiv \sigma^2 < \infty$, niin

$$y_n = \sqrt{n} \frac{\bar{x}_n - \mu}{\sigma} \underset{\text{asympt.}}{\sim} N(0, 1) ,$$

ts.

$$\lim_{n \rightarrow \infty} F_{y_n}(t) \equiv \Phi(t) \quad , \quad \text{jossa} \quad \Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds .$$

Lause A1.3: (Moniulotteinen keskeinen raja-arvolause)

Jos $X_n = (x_{1,n} \dots x_{m,n})' \sim i.i.d.$, $EX_n \equiv \mu = (\mu_1 \dots \mu_m)'$ ja $\text{cov}(X_n) = E(X_n - \mu)(X_n - \mu)' \equiv \Sigma$ on hyvin määritelty, pätee

$$Y_n = \sqrt{n}(\bar{X}_n - \mu) \underset{\text{asympt.}}{\sim} N_m(0, \Sigma)$$

jossa $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ ja $N_m(0, \Sigma)$ tarkoittaa m -ulotteista multinormaali-jakaumaa. ■

A1.2 Matemaattisia merkintöjä ja määritelmiä

Merkitään matriiseja ja vektoreita suurilla kirjaimilla ja niiden elementtejä pienillä. Esimerkiksi

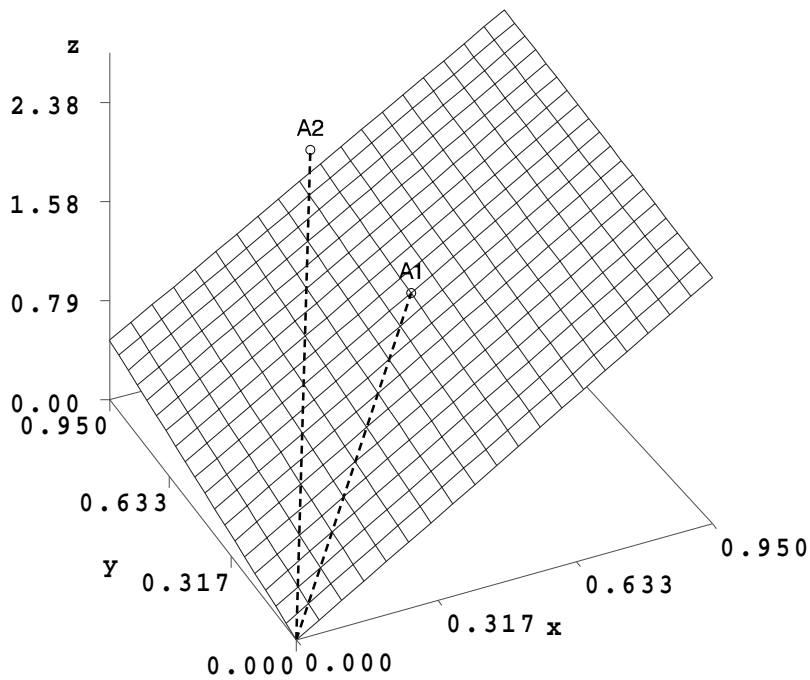
$$X = (x_1 \quad \dots \quad x_n)' \in \mathcal{R}^n \quad , \quad A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \cdot & \dots & \cdot \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \quad n \times m \quad .$$

Kuten tunnettua, voidaan matriisitulon

$$L_A(Z) = AZ \quad Z \in \mathcal{R}^m$$

ajatella vastaavan *linearikuvausta* $L_A : \mathcal{R}^m \rightarrow \mathcal{R}^n$, jonka kuva-avaruus $L_A(\mathcal{R}^m)$ on samalla A :n sarakkeiden virittämä \mathcal{R}^n :n lineaarinen aliavaruus, jota usein merkitään symbolilla $\text{col}(A)$.

Kuva A1.1: Kuva-avaruus $\text{col}(A_1 \ A_2)$ tapauksessa $n = 3$, $m = 2$,
jossa $A_1 = (0.5 \ 0.5 \ 1.25)'$ ja $A_2 = (0.5 \ 1 \ 1.5)'$.



Apulause A1.1: Jos $Z_1 \in \mathcal{R}^m$ ja $Z_2 \in \mathcal{R}^m$ ovat kaksi vektoria, joille pätee $Z_1'Z_2 = 0$, ovat Z_1 ja Z_2 (euklidisen geometrian mielessä) kohtisuorassa toisiaan vastaan ($Z_1 \perp Z_2$).

Todistus: Euklidisessa geometriassa vektorin $Z \in \mathcal{R}^m$ pituus (etäisyys origosta) on (Pythagoraan lauseen mukaisesti) muotoa $\|Z\| = \sqrt{z_1^2 + \dots + z_m^2} = \sqrt{Z'Z}$. Matriisitulon peruslaskusääntöjen mukaisesti taas

$$\|Z_1 + Z_2\|^2 = (Z_1 + Z_2)'(Z_1 + Z_2) = Z_1'Z_1 + Z_2'Z_2 + 2 \cdot Z_1'Z_2 \quad ,$$

joten

$$Z_1 \perp Z_2 \iff Z_1'Z_2 = 0 \quad .$$

■

Apulause A1.2: Jos A on *vajaa-asteinen* (ts. $r = \dim(\text{col}(A)) = \text{rank}(A) < m$), se voidaan aina hajottaa muotoon

$$A = \alpha\beta' \tag{A1.1}$$

jossa α on $n \times r$ -matriisi ja β on $m \times r$ -matriisi. Tällöin siis $\text{col}(A) = \text{col}(\alpha)$. ■

Määritelmä A1.12: Olkoon A täysiasteinen. *Ortogonaaliseksi projektioksi* $\text{col}(A)$:lle sanotaan matriisiin

$$P_A = A(A'A)^{-1}A' \tag{A1.2}$$

liittyvää lineaarikuvausta. Aliavaruuden $\text{col}(A) \subset \mathcal{R}^n$ *ortogonaaliseksi komplementiksi* $\text{col}(A)^\perp$ sanotaan sitä \mathcal{R}^n :n lineaarista aliavaruutta, jonka jokainen vektori on kohtisuorassa $\text{col}(A)$:n jokaista vektoria vastaan.

(Ortogonaalinen projektiio $\text{col}(A)^\perp$:lle vastaa luonnollisesti matriisia $I - P_A = I - A(A'A)^{-1}A'$.)

Mitä tahansa matriisia, jonka sarakkeet virittävät $\text{col}(A)^\perp$:n, merkitään geneerisellä symbolilla A_\perp . ■

Määritelmä A1.13: Lineaarikuvauksen L_A *ytimeksi* $\mathcal{Y}(A)$ sanotaan niiden vektorien $Z \in \mathcal{R}^m$ muodostamaa joukkoa, joille $AZ = 0$.

Jos A on kirjoitettu muotoon $A = \alpha\beta'$, on siis

$$\mathcal{Y}(A) = \text{col}(\beta)^\perp = \text{col}(\beta_\perp) \quad .$$

■

Määritelmä A1.14: Neliömatriisia A sanotaan *positiivisesti definiitiksi*, jos

$$X'AX > 0 \quad \text{kaikilla} \quad X \neq 0 \quad .$$

(Tällöin käytetään usein merkintää $A \succ 0$.)

Vastaavasti neliömatriisia sanotaan *ei-negatiivisesti definitiksi* (tai positiivisesti semidefinitiksi), jos $X'AX \geq 0$ kaikilla $X \in \mathcal{R}^n$.

Neliömatriisin *jäljellä* (trace) tarkoitetaan sen diagonaalelementtien summaa

$$\text{tr}(A) = \sum_{i=1}^n a_{ii} \quad .$$

■

Apulause A1.3: Symmetrinen matriisi A voidaan aina esittää muodossa

$$A = R\Lambda R' = \sum_{i=1}^n \lambda_i R_i R_i' \quad (A1.3)$$

jossa $R = (R_1 \dots R_n)$ on ortogonaalinen ($R'R = I$) ja diagonaalimatriisi $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ koostuu A :n ominaisarvoista.

■

Huomautus A1.1: Symmetrisen matriisin ominaisarvot ovat aina reaalisia, joten *spektraaliesityksen* (A1.3) kaikki osatekijät ovat reaalisia. Hajotelmasta (A1.3) seuraa tietenkin, että

$$A^k = R\Lambda^k R' = \sum_{i=1}^n \lambda_i^k R_i R_i'$$

kaikille kokonaisluvuille k .

Lisäksi

$$\det(A) = \prod_{i=1}^n \lambda_i \quad \text{ja} \quad \text{tr}(A) = \sum_{i=1}^n \lambda_i \quad .$$

■

Apulause A1.4: Jos symmetrinen matriisi on positiivisesti definitti, on aina löydettävissä sellainen alakolmiomatriisi L ja sellainen yläkolmiomatriisi U , että

$$A = UU' = LL' \quad . \quad (A1.4)$$

Tätä hajotelmaa kutsutaan *Cholesky-dekompositioksi*.

■

Määritelmä A1.15: Olkoon $F = (F_1 \dots F_n)' : \mathcal{R}^m \rightarrow \mathcal{R}^n$ kuvaus, jonka komponenttifunktiolla F_j ($j = 1, \dots, n$) on jatkuvat osittaisderivaatat. Tällöin F :n *derivaattamatriisiksi* pisteessä $X \in \mathcal{R}^m$ sanotaan matriisia

$$DF(X) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_m} \\ \vdots & \dots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \dots & \frac{\partial F_n}{\partial x_m} \end{pmatrix} \quad .$$

Jos $n = 1$, sanotaan vektoria $\nabla F(X) = DF(X)'$ funktion F *gradienttivektoriksi* ja matriisiä

$$D^2F(X) = D(DF(X)) = \left(\frac{\partial^2 F}{\partial x_i \partial x_j} \right)$$

F :n *Hessen matriisiksi*.

Jos F on kahdesti jatkuvasti derivoituva, on $D^2F(X)$ aina *symmetrinen*. ■

Huomautus A1.2: Määritelmän A1.15 mukaiset derivoimissäännöt ovat kutakuinkin koulussa opittujen kaltaisia, sillä esimerkiksi

$$D(AX) \equiv A \quad , \quad D(X'AX) = 2X'A$$

ja

$$DG \circ F(X) = DG(F(X)) DF(X) \quad .$$

Mainittakoon myös 2. asteen Taylor-kehittelmä pisteessä X_o , kun $n = 1$:

$$F(X) = F(X_o) + DF(X_o)(X - X_o) + \frac{1}{2!}(X - X_o)'D^2F(X_o)(X - X_o) + \|X - X_o\| o(X - X_o) \quad ,$$

jossa $o(Z)$ toimii geneerisenä symbolina termille, joka on pienempää suuruusluokkaa kuin Z , ts. jolle $\lim_{Z \rightarrow 0} \frac{1}{\|Z\|} o(Z) = 0$. ■

Määritelmä A1.16: Matriisin A ($m \times n$) ja B ($p \times r$) ns. *Kroneckerin tulolla* tarkoitetaan $mp \times nr$ -matriisiä

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \dots & \vdots \\ a_{n1}B & \dots & a_{nn}B \end{pmatrix} \quad .$$

Ns. *vec-operaattori* taas määritellään seuraavasti: Jos A on $m \times n$ -matriisi

$$A = (A_1 \quad \dots \quad A_n) \quad , \quad A_i \in \mathcal{R}^m \quad ,$$

tarkoitetaan $\text{vec}(A)$:lla mn -ulotteista vektoria

$$\text{vec}(A) = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \in \mathcal{R}^{mn} \quad .$$

Jos A on symmetrinen $n \times n$ -matriisi, tarkoittaa $\text{vec}(A)$ sitä $\frac{1}{2}n(n+1)$ - ulotteista vektoria, joka koostuu A :n alakolmion elementeistä. Vastaavasti $\overline{\text{vec}}(A)$ koostuu A :n yläkolmion elementeistä. ■