

# Moniulotteiset aikasarjat

Pentti Saikkonen

Syksy 2011

Päivitetty versio 3.3.2015

## Sisältö

1.	Johdanto	1
1.1.	Taustaa	1
1.2.	Stokastinen prosessi	2
2.	Stationaariset prosessit	4
2.1.	Määritelmiä	4
2.2.	Lineaarinen prosessi	5
2.3.	VAR(p)–prosessi	7
2.4.	VARMA(p,q)–prosessi	16
3.	Stationaaristen prosessien ensimmäisten ja toisten momenttien estimointi	17
4.	Stationaarisen VAR(p)–mallin teoriaa	21
4.1.	Parametrien rajoittamaton estimointi	21
4.2.	Estimointi lineaarisin rajoittein	26
4.3.	Hypoteesien testaus	29
4.4.	Mallinvalinta ja valitun mallin sopivuuden tutkiminen	31
4.5.	Grangerin kausaalisuus	36
4.6.	Impulssivasteanalyysi	40
	Liite A: Matriisilaskentaa	45
	Liite B: Asymptoottisia tuloksia	54

# 1 Johdanto

## 1.1 Taustaa

Kurssilla tarkastellaan moniulotteisten aikasarjojen kuvaamiseen ja ennustamiseen käytettäviä tilastollisia malleja ja niiden parametrien estimointia ja testausta. *Moniulotteinen tai vektoriarvoinen aikasarja* on havaintoaineisto, jossa on vektoriarvoisia havaintoja peräkkäisinä ajankohtina eli

$$y_t = (y_{1t}, \dots, y_{nt}), \quad t = 1, \dots, T,$$

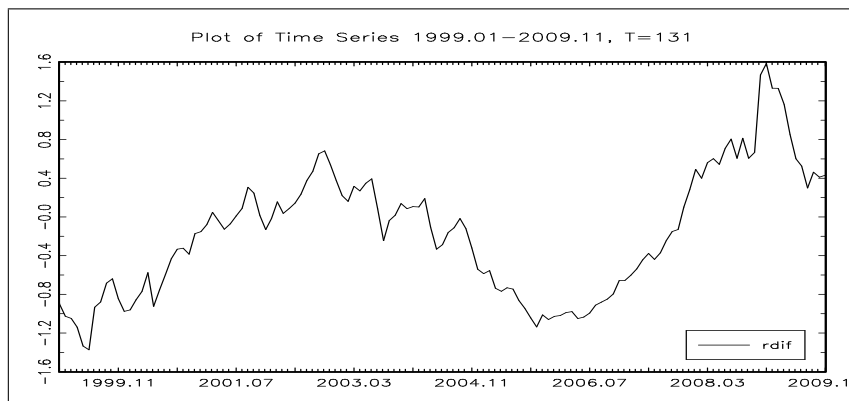
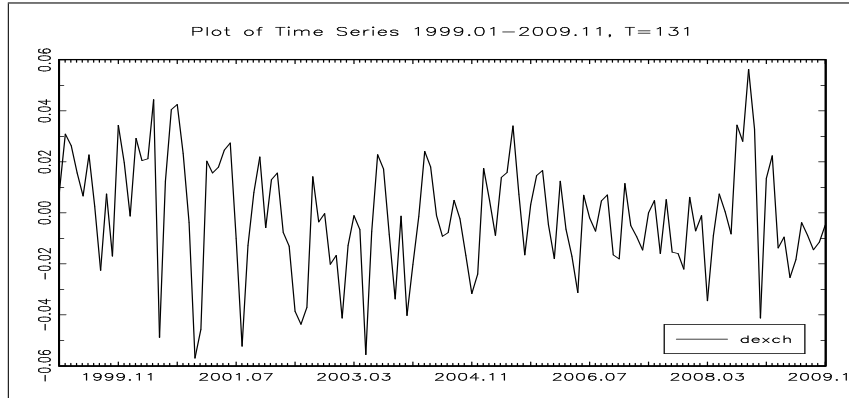
jossa  $t$  on ajankohta,  $n$  tarkasteltavien muuttujien lukumäärä ja  $T$  havaintojen lukumäärä.<sup>1</sup> Tällä kurssilla oletetaan, että havainnot ovat ajallisesti yhtä kaukana toisistaan tai niitä voidaan käsitellä sellaisina. Jos  $n = 1$ , on kysymyksessä *yksiulotteinen tai reaaliarvoinen aikasarja*. Kurssilla oletetaan perustiedot yksiulotteisten aikasarjojen mallintamisesta.

Esimerkki moniulotteisesta aikasarjasta saadaan tarkastelemalla vaikkapa usean pörssikurssin päivittäistä kehitystä Helsingin pörsissä vuonna 2009. Myöhemmin tarkastellaan kaksiulotteista esimerkkiä, jossa toisena komponenttina on euron ja Yhdysvaltojen dollarin välisen kuukausittaisen vaihtokurssin (euro/U.S. dollari) muutos ( $dexch_t$ ) ajanjaksona 1999I - 2009XI. Taloustieteessä esitetyn ns. kattamattoman korkopariteettiteorian mukaan vaihtokurssin odotetun muutoksen pitäisi vastata vertailuvaluuttojen korkoeroa (euroalue/US), jota mitataan euroalueen ja Yhdysvaltojen 10 vuoden valtioiden obligaatioiden keskikorkojen erolla ( $rdif_t$ ). Vaihtokurssin odotettu muutos ajankohtana  $t$  määritellään muuttujan  $dexch_{t+1}$  ehdollisena odotusarvona ehdolla ”kaikki ajankohtana  $t$  käytävissä oleva relevantti informaatio”. Jos tämän odotuksen korvikkeena käytetään ajankohtana  $t$  tiedossa olevaa muutosta  $dexch_t$ , voidaan vaihtokurssin muutosta ja korkoeroa kuvaavien aikasarjojen  $dexch_t$  ja  $rdif_t$  odottaa olevan yhteydessä toisiinsa. Kuviossa 1.1 esitetään näistä muuttujista koostuva aineisto graafisesti. Keskeinen mielenkiinto tässä kuten moniulotteisissa aikasarja-aineistoissa yleensäkin kohdistuu tarkasteltavien muuttujien välisiin riippuvuuksiin, jotka eivät välttämättä ilmene samanaikaisina riippuvuuksina.

Kuten yksiulotteisessa tapauksessakin, vaikuttaa moniulotteisten aikasarja-aineistojen analysointiin ja siinä käytettäviin tekniikoihin se, ovatko aikasarjat stationaarisia vai ”trendimäisesti” epästationaarisia. Seuraavassa tarkastellaan tähän liittyviä yleisiä kysymyksiä moniulotteisten aikasarjojen tapauksessa.

---

<sup>1</sup>Tällä kurssilla käytetään yleensä samaa merkintää satunnaismuuttujista ja niiden havaituista arvoista. Ero on ymmärrettävä asiayhteydestä. Vektorit tulkitaan matriisilaskuissa aina pystyvektoreiksi ja niistä voidaan käyttää myös matriisilaskennan merkintää eli esimerkiksi  $y_t = [y_{1t} \cdots y_{nt}]'$ , jossa pilkku osoittaa vektorin (samoin kuin myöhemmin matriisin) transponointia.



**Kuvio 1.1.** Euron ja Yhdysvaltojen dollarin välisen kuukausittaisen vaihtokurssin (euro/U.S. dollari) muutos (ylhäällä) ja euroalueen ja Yhdysvaltojen 10 vuoden valtioiden obligaatioiden keskkorkojen ero (alhaalla) ajanjaksona 1999I - 2009XI.

## 1.2 Stokastinen prosessi

Aikasarja voidaan tulkita stokastisen prosessin realisaatioksi eli satunnaisilmiön havaituksi kehityskuluksi ajassa. *Stokastinen prosessi* tai usein vain prosessi voidaan tällä kurssilla määritellä joukoksi satunnaisvektoreita (sv) tai yksiulotteisissa tapauksessa satunnaismuuttujia (sm):

$$\{y_t, t \in \mathbb{Z}\} \quad (\text{diskreetti aika}) \quad \text{tai} \quad \{y_t, t \in \mathbb{R}\} \quad (\text{jatkuva aika}).$$

Erityisesti jatkuva-aikaisessa tapauksessa käytetään myös merkintää  $y(t)$ . Aikamuuttuja  $t$  voi kuulua myös johonkin  $\mathbb{Z}$ :n tai  $\mathbb{R}$ :n osajoukkoon kuten  $\mathbb{N}$  tai  $\{0 \leq t \leq 1\}$ .

Jatkossa merkitään yleensä lyhyesti  $\{y_t\}$  tai vain  $y_t$ . Merkintä  $\{y_t\}$  korostaa sitä, että stokastinen prosessi on (yleisesti ottaen) ääretönulotteinen satunnaismuuttuja. Esimerkiksi reaaliarvoinen diskreettiaikainen prosessi  $\{y_t\} = \{\dots, y_{-1}, y_0, y_1, \dots\}$  saa

arvoja joukossa  $\mathbb{R}^\infty = \mathbb{R} \times \mathbb{R} \times \dots$ . Jatkuva-aikainen prosessi voi puolestaan saada arvoja esimerkiksi joukossa  $C[0, 1]$  eli kaikkien välillä  $[0, 1]$  määriteltyjen jatkuvien reaaliarvoisten funktioiden joukossa. Käytännössä havaittu aikasarja on vain yksi ääretönulotteisen satunnaismuuttujan saama arvo tai yleensä vain osa sen saamasta arvosta (esim.  $y_1, \dots, y_T, T < \infty$ ).

Kun stokastinen prosessi tulkitaan ääretönulotteisena satunnaismuuttujana, voidaan kysyä mitä tarkoitetaan sen todennäköisyysjakaumalla, joka on määriteltävä esimerkiksi joukossa  $\mathbb{R}^\infty$  tai  $C[0, 1]$ . Tätä jakaumaa tarvitaan todennäköisyysteoreettisissa tarkasteluissa, kun tutkitaan sellaisten tapahtumien todennäköisyyksiä, joita ei voida ilmaista käyttäen todennäköisyyksiä

$$P \{y_{t_1} \in A_1, \dots, y_{t_m} \in A_m\}, \quad 1 \leq m < \infty, \quad (1.1)$$

eli ns. *äärellisulotteisia jakaumia*. Esimerkki tällaisesta tapahtumasta on vaikkapa  $\{0 \leq y_t \leq 1 : \forall t \geq 100\}$ .

Kysymys stokastisen prosessin todennäköisyysjakaumasta kuuluu kehittyneen todennäköisyyslaskennan piiriin. Keskeinen tulos stokastisten prosessien teoriasta sanoo hieman yksinkertaistaen, että prosessin (ääretönulotteinen) jakauma määräytyy yksikäsitteisesti äärellisulotteisista jakaumista (1.1). Käänteinen tulos eli prosessin äärellisulotteisten jakaumien (1.1) määräytyminen yksikäsitteisesti ääretönulotteisista jakaumista on ilmeinen. Prosessin jakauma voidaan siis karakterisoida käyttäen äärellisulotteisia jakaumia. Tästä seuraa esimerkiksi, että prosessi  $\{y_t\}$  on normaalin täsmälleen silloin, kun sen kaikki äärellisulotteiset jakaumat ovat normaalisia eli sv  $(y_{t_1}, \dots, y_{t_m})$  noudattaa multinormaalijakaumaa kaikilla  $t_1, \dots, t_m$  ja kaikilla  $m \geq 1$  ( $m \in \mathbb{Z}$ , mutta jatkuva-aikaisessa tapauksessa  $t_1, \dots, t_m$  eivät välttämättä ole kokonaislukuja).

Stokastisen prosessin ensimmäiset ja toiset momentit määritellään seuraavasti (olettaen, että ne ovat äärellisinä olemassa). Prosessin  $y_t$  *odotusarvo(funktio)* on

$$\mu_t = \mathbf{E}(y_t) \quad (n \times 1)$$

ja *kovarianssifunktio*

$$\Gamma_{s,t} = \mathbf{Cov}(y_s, y_t) = \mathbf{E}[(y_s - \mu_s)(y_t - \mu_t)'] \quad (n \times n).$$

Valitsemalla viimeksi mainitussa  $s = t$  saadaan prosessin *kovarianssimatriisi*  $\Sigma_t = \Gamma_{t,t}$  eli

$$\Sigma_t = \mathbf{Cov}(y_t) = \mathbf{E}[(y_t - \mu_t)(y_t - \mu_t)'] \quad (n \times n).$$

Jatkuva-aikaisessa tapauksessa alaindeksit  $s$  ja  $t$  voivat saada mitä reaalityyppisiä arvoja tahansa. Käytännössä rajoitutaan usein näiden momenttien selvittämiseen (normaalissa tapauksessa muuta selvitettävää ei olekaan). Teoreettisissa tarkasteluissa oletukset korkeampiasteisista momenteista ovat kuitenkin tavallisia. Ellei toisin mainita, tulkitaan merkintöjen kuten  $\mu_t = \mathbf{E}(y_t)$  ja  $\Sigma_t = \mathbf{Cov}(y_t)$  sisältävän automaattisesti oletuksen, että kyseiset momentit ovat äärellisinä olemassa.

## 2 Stationaariset prosessit

### 2.1 Määritelmiä

Stationaarisuus määritellään moniulotteisessa tapauksessa samalla tavalla kuin yksiulotteisessa tapauksessa. Stokastinen prosessi  $y_t$  ( $n \times 1$ ) on siten (vahvasti) *stationaarinen*, jos

$$\mathbb{P}\{y_{t_1} \in A_1, \dots, y_{t_m} \in A_m\} = \mathbb{P}\{y_{t_1+k} \in A_1, \dots, y_{t_m+k} \in A_m\}$$

kaikilla  $A_1, \dots, A_m$ ,  $t_1, \dots, t_m$ ,  $k$  ja  $m \geq 1$  (jatkuva-aikaisessa tapauksessa vain  $m$ :n tarvitsee olla kokonaisluku). Toisin sanoen, stationaarisen prosessin todennäköisyysstruktuuri on aikainvariantti.

Stationaarisuuden määritelmästä seuraa erityisesti, että sv:t  $y_t$  ovat samoin jakautuneita kaikilla  $t$  ja myös, että sv:t  $(y_t, y_{t+h})$  ja  $(y_s, y_{s+h})$  ovat samoin jakautuneita kaikilla  $t$  ja  $s$  ja jokaisella (kiinteällä)  $h$ . Näiden ominaisuuksien lisäksi (vahvasti) stationaarisilla prosesseilla on se hyödyllinen ominaisuus, että ne säilyttävät luonteensa funktiomuunnoksissa. Erityisesti siis, jos  $z_t$  on (vahvasti) stationaarinen diskreetti-aikainen prosessi ( $t \in \mathbb{Z}$ ), niin samoin on  $y_t = g(z_{t+h}, \dots, z_{t-k})$  ( $h, k \geq 0$ ) millä tahansa ”siistillä” (eli esim. jatkuvalla tai yleisemmin ns. mitallisella) funktiolla  $g$ . Lisäksi  $h$ :n tai  $k$ :n tai molempien paikalla voi olla  $\infty$ . Kun  $h$  ja  $k$  ovat äärellisiä seuraa tämä ominaisuus jokseenkin suoraan (vahvan) stationaarisuuden määritelmästä.

Prosessi  $y_t$  ( $n \times 1$ ) on *heikosti stationaarinen* tai *kovarianssistationaarinen*, jos sen toiset (ja siten myös ensimmäiset) momentit ovat äärellisinä olemassa ja ne ovat aikainvariantteja eli

$$\mathbb{E}(y_t) = \mu \quad \text{ja} \quad \text{Cov}(y_t, y_{t+k}) = \Gamma_k \quad \text{kaikilla } t \text{ ja } k.$$

(Jatkuva-aikaisessa tapauksessa  $t \in \mathbb{R}$  ja myös  $k \in \mathbb{R}$ ).

Vahvasta stationaarisuudesta ja toisten momenttien äärellisyydestä seuraa ilmeisesti heikko stationaarisuus. Käänteinen tulos pitää paikkansa normaalilla prosesseilla, mutta *ei yleisesti*. Huomaa myös, että heikosti stationaariset prosessit eivät yleisesti säilytä luonnettaan funktiomuunnoksissa. Jatkossa pelkkä ’stationaarisuus’ viittaa aina vahvaan stationaarisuuteen (tämä puhetapa ei ole yleisesti käytössä).

Yksinkertaisin esimerkki (diskreettiaikaisesta) stationaarisesta prosessista on jono riippumattomia samoin jakautuneita satunnaisvektoreita (stationaarisuus voidaan todeta suoraan määritelmästä). Jos  $y_t$  ( $n \times 1$ ) on tällainen jono ja lisäksi  $\mathbb{E}(y_t) = \mu$  ja  $\text{Cov}(y_t) = \Sigma$  ovat äärellisinä olemassa, merkitään  $y_t \sim \text{iid}(\mu, \Sigma)$  ( $t \in \mathbb{Z}$ ). Tällaista prosessia sanotaan (vahvaksi) *valkoiseksi kohinaksi*. Jos  $y_t$  on lisäksi normaalisti jakautunut, merkitään  $y_t \sim \text{nid}(\mu, \Sigma)$ . Jos edellä oletetaan riippumattomuuden asemesta (lievemmin) korreloimattomuus eli  $\text{Cov}(y_s, y_t) = 0$ ,  $t \neq s$ , on  $y_t$  heikosti stationaarinen eli ns. *heikko valkoinen kohina*.

Yksiulotteisessa tapauksessa kovarianssifunktio on symmetrinen viipymän  $k$  suhteen. Moniulotteisessa tapauksessa näin ei ole. Sen sijaan pätee

$$\begin{aligned}\Gamma_{-k} &= \mathbf{E} [(y_t - \mu)(y_{t-k} - \mu)'] \\ &= (\mathbf{E} [(y_{t-k} - \mu)(y_t - \mu)'])' \\ &= \Gamma_k' .\end{aligned}$$

Kuten yksiulotteisessa tapauksessakin, riittää siis tuntea  $\Gamma_k$  vain, kun  $k \geq 0$  (tai  $k \leq 0$ ).

Kovarianssifunktiosta saadaan prosessin komponenttien autokovarianssifunktiot ja ristikovarianssifunktiot. Olkoot  $y_t = (y_{1t}, \dots, y_{nt})$  ja  $\Gamma_k = [\gamma_{ab,k}]$  ( $a, b = 1, \dots, n$ ). Tällöin prosessin  $y_{at}$  *autokorrelaatiofunktio* on

$$\rho_{a,k} = \gamma_{a,k} / \gamma_{a,0}, \quad a = 1, \dots, n$$

ja prosessien  $y_{at}$  ja  $y_{bt}$  välinen *ristikorrelaatiofunktio* on

$$\rho_{ab,k} = \gamma_{ab,k} / \sqrt{\gamma_{a,0}\gamma_{b,0}} \quad a, b = 1, \dots, n.$$

Edellä  $\gamma_{a,k} = \gamma_{aa,k}$  ja  $\gamma_{ab,k}$  ( $a \neq b$ ) ovat vastaavasti *autokovarianssi-* ja *ristikovarianssifunktioita*.

Käytännön kannalta stationaariseen prosessiin voidaan yleensä liittää oletus

$$\Gamma_k \rightarrow 0, \quad \text{kun } |k| \rightarrow \infty, \quad (2.1)$$

jossa suppeneminen voidaan tulkita tapahtuvaksi alkioittain (matriisiarvoisten jonojen suppenemista on tarkasteltu Liitteessä A.1). Tällöin  $y_t$  ja  $y_{t+k}$  ovat lähes korreloimattomia, kun  $|k|$  on ”suuri” eli, kun  $y_t$  ja  $y_{t+k}$  ovat ajallisesti ”kaukana” toisistaan. (Huomaa kuitenkin, että ehto (2.1) ei seuraa stationaarisuudesta.) Jos ehto (2.1) pätee, riittää  $y_t$ :n ensimmäisten ja toisten momenttien estimoimiseksi (olenaisesti) estimoida vain äärellinen määrä parametreja (eli  $\mu$  ja  $\Gamma_k$ ,  $k = -K, \dots, K$ , ”suurella”  $K$ :n arvolla).

Ellei toisin mainita, oletetaan tarkasteltavat prosessit seuraavassa diskreettiaikaisiksi.

## 2.2 Lineaarinen prosessi

Tarkastellaan aluksi yksiulotteista (kausaalista) lineaarista prosessia<sup>2</sup>

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \varepsilon_t \sim \text{iid}(0, \omega^2), \quad t \in \mathbb{Z}, \quad (2.2)$$

---

<sup>2</sup>Lineaarinen prosessi voidaan määritellä myös yleisemmin sallimalla yhtälössä (2.2) indeksille  $j$  arvot  $j < 0$ . Tällaista lineaarista prosessia kutsutaan toisinaan ei-kausaaliseksi.

jossa  $\psi_0 = 1$  ja

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty. \quad (2.3)$$

Yhtälön (2.2) oikealla puolella oleva ääretön summalauseke voidaan tulkita osasummien  $\sum_{j=0}^N \psi_j \varepsilon_{t-j}$  jonon kvadraattiseksi raja-arvoksi ( $N \rightarrow \infty$ ). Stokastisten prosessien (kehittyneestä) teoriasta tiedetään, että ehdon (2.3) voimassa ollessa tämä kvadraattinen raja-arvo on olemassa ja lisäksi, että raja-arvolla on äärellinen toinen momentti. Yhtälössä (2.2) olevasta iid-prosessista käytetään hieman tilanteesta riippuen nimitystä *virhe*, *sokki* tai *innovaatio*. Vahvan valkoisen kohinan asemasta voitaisiin  $\varepsilon_t$  olettaa monissa tarkasteluissa lievemmin heikoksi valkoiseksi kohinaksi (samoin sen alempana esitettävä moniulotteinen vastine).

Koska prosessi  $\{\varepsilon_t\}$  on (vahvasti) stationaarinen pätee sama myös prosessille  $\{y_t\}$ . Heikko stationaarisuus nähdään seuraavassa esitettävän moniulotteisen yleistyksen yhteydessä.

Moniulotteisessa tapauksessa  $y_t = (y_{1t}, \dots, y_{nt})$  määritellään (kausaalinen) *lineaarinen prosessi* yhtälöllä

$$y_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad \varepsilon_t \sim \text{iid}(0, \Omega), \quad t \in \mathbb{Z}, \quad (2.4)$$

jossa  $\Psi_0 = I_n$  ( $n \times n$  yksikkömatriisi) ja kerroinmatriisien  $\Psi_j$  täytyy toteuttaa sopiva yhtälön (2.3) vastine. Käyttäen (Euklidisen) matriisnormin käsitettä tämä ehto voidaan esittää kätevästi muodossa

$$\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty. \quad (2.5)$$

Normi  $\|\Psi_j\|$  määritellään yksinkertaisesti muodostamalla matriisin  $\Psi_j$  alkioista vektori ja soveltamalla tavallista  $\mathbb{R}^{n^2}$ :n (Euklidista) vektorinormia. Määritelmä on esitetty yksityiskohtaisemmin Liitteessä A.1, missä todetaan myös, että ehto (2.5) on yhtäpitävää sen kanssa, että ehdon (2.3) vastine pätee matriisijonon  $\Psi_j$  kaikille alkioille.

Yhtälön (2.4) oikea puoli voidaan tulkita komponenteittain kvadraattisina raja-arvoina aivan kuten yksiulotteisessa tapauksessa (2.2). Prosessin  $\{y_t\}$  (vahva) stationaarisuus seuraa prosessin  $\{\varepsilon_t\}$  vastaavasta ominaisuudesta kuten yksiulotteisessa tapauksessakin. Heikko stationaarisuus todetaan laskelmalla

$$\mathbf{E}(y_t) = \sum_{j=0}^{\infty} \mathbf{E}(\Psi_j \varepsilon_{t-j}) = \sum_{j=0}^{\infty} \Psi_j \mathbf{E}(\varepsilon_{t-j}) = 0$$



ja

$$\begin{aligned}
\text{Cov}(y_t, y_{t+k}) &= \mathbf{E}(y_t y'_{t+k}) \\
&= \mathbf{E}\left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Psi_i \varepsilon_{t-i} \varepsilon'_{t+k-j} \Psi'_j\right) \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Psi_i \mathbf{E}(\varepsilon_{t-i} \varepsilon'_{t+k-j}) \Psi'_j \\
&= \sum_{j=0}^{\infty} \Psi_j \Omega \Psi'_{j+k}.
\end{aligned}$$

Laskelmissa on käytetty odotusarvon lineaarisuusominaisuutta ja innovaatioiden  $\varepsilon_t$  riippumattomuutta, josta seuraa  $\mathbf{E}(\varepsilon_{t-i} \varepsilon'_{t+k-j}) = 0$ , kun  $j \neq k + i$ , ja  $\mathbf{E}(\varepsilon_t \varepsilon'_t) = \Omega$ .<sup>3</sup> Koska matriisinormin ominaisuuksien perusteella  $\|\Psi_j \Omega \Psi'_j\| \leq \|\Omega\| \|\Psi_j\|^2$ , nähdään kolmioepäyhtälön avulla, että  $\text{Cov}(y_t)$  on ehdon (2.5) voimassa ollessa äärellinen (ks. Liite A.1), joten  $y_t$  on heikosti stationaarinen.

Edellä esitetyt odotusarvoa  $\mathbf{E}(y_t)$  ja kovarianssia  $\text{Cov}(y_t, y_{t+k})$  koskevat laskelmat pätevät myös, kun innovaatiot ovat vain korreloimattomia (eli  $\text{Cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ ,  $k \neq 0$ ), joten heikko stationaarisuus pätee tällöinkin. Huomaa myös, että tarkastelemalla  $y_t$ :n asemesta prosessia  $y_t - \mu$  voidaan ottaa huomioon nolasta poikkeava odotusarvo.

### 2.3 VAR(p)–prosessi

Astetta  $p$  olevan yksiulotteisen autoregressiivisen prosessin eli  $AR(p)$ –prosessin tai  $AR(p)$ –mallin lähtökohtana on yhtälö

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \omega^2) \quad (t \in \mathbb{Z}). \quad (2.6)$$

Toisin sanoen, havaittavan prosessin  $y_t$  arvo ajankohtana  $t$  riippuu lineaarisesti sen omista viipymistä  $y_{t-1}, \dots, y_{t-p}$  ja ei-havaittavasta virhetermistä  $\varepsilon_t$  aivan kuten lineaarisessa regressiomallissa.

Moniulotteisessa tapauksessa  $y_t = (y_{1t}, \dots, y_{nt})$  otetaan lähtökohdaksi yhtälön (2.6) yleistys

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \Omega) \quad (t \in \mathbb{Z}), \quad (2.7)$$

jossa  $A_1, \dots, A_p$  ovat  $n \times n$  kerroinmatriiseja ja  $\varepsilon_t$  ( $n \times 1$ ) on ei-havaittava virhetermi (tai sokki tai innovaatio). Ellei toisin mainita, oletetaan kovarianssimatriisi  $\Omega$  positiivisesti definiitiksi (ja siten epäsingulaariseksi eli kääntyväksi).<sup>4</sup> Tässä yhteydessä

<sup>3</sup>Lineaarisuusominaisuuden käyttäminen tässä eli odotusarvon ja äärettömän summan järjestyksen vaihtamisen luvallisuus ei ole itsestään selvää, vaan perustuu äärettömän summan (2.4) kvadraattisen konvergenssin tapaan stokastisten prosessien (kehittyneeseen) teoriaan.

<sup>4</sup>Symmetrinen matriisi  $A$  ( $k \times k$ ) on positiivisesti definiitti, jos  $x'Ax > 0$  kaikilla nollasta poikkeavilla vektoreilla  $x$  ( $k \times 1$ ). Positiivisesti definiitti matriisi on tunnetusti epäsingulaarinen. Symmetristä matriisia  $A$  sanotaan positiivisesti semidefiniitiksi, jos  $x'Ax \geq 0$  kaikilla  $x$  ( $k \times 1$ ).

puhutaan  $VAR(p)$ -prosessista tai  $VAR(p)$ -mallista ('V' viittaa sanaan vektori). Ideana on siis selittää jokaista vektorin  $y_t$  komponenttia sen omilla viipymillä ja myös  $y_t$ :n muiden komponenttien viipymillä lineaarisen regressiomallin tapaan. Moniulotteisessa tapauksessa voidaan siten tarkastella myös muuttujien välisiä viivästettyjä riippuvuuksia kuten esimerkiksi sitä, miten  $y_{2t}$  vaikuttaa  $y_{1t}$ :n tulevaan kehitykseen ja miten  $y_{1t}$  vaikuttaa  $y_{2t}$ :n tulevaan kehitykseen.

Jotta yhtälö (2.7) määritteli stationaarisen prosessin (tai edes stokastisen prosessin), täytyy kerroinmatriisien  $A_1, \dots, A_p$  toteuttaa sopiva ehto. Tämän ehdon esittämiseksi on kätevää tarkastella yhtälöön (2.7) läheisesti liittyvää yhtälöä

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad (t \in \mathbb{Z}), \quad (2.8)$$

jossa

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} \quad (np \times np),$$

$$\mathbf{y}_t = [y'_t \cdots y'_{t-p+1}]' \quad (np \times 1) \quad \text{ja} \quad \boldsymbol{\varepsilon}_t = [\varepsilon'_t \ 0' \cdots 0']' \quad (np \times 1).$$

Nähdään helposti, että yhtälö (2.8) pätee jos ja vain jos yhtälö (2.7) pätee. Oletuksesta  $\varepsilon_t \sim \text{iid}(0, \boldsymbol{\Omega})$  seuraa, että

$$\boldsymbol{\varepsilon}_t \sim \text{iid}(0, \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \text{diag}[\Omega \ 0 \ \cdots \ 0],$$

jossa  $\text{diag}[\Omega \ 0 \ \cdots \ 0]$  on lohkodeagonaalinen matriisi diagonaaleilla osoitetut  $(n \times n)$  matriisit. Kovarianssimatriisi  $\boldsymbol{\Omega}$  on siis singulaarinen (ei-kääntyvä) silloinkin, kun  $\Omega$  on positiivisesti definiitti. Tästä huolimatta voidaan prosessia (2.8) käyttää apuvälineenä tutkittaessa  $VAR(p)$ -prosessin (2.7) ominaisuuksia.

**$VAR(p)$ -prosessin stationaarisuusehto.** Seuraavassa todetaan, että yhtälö (2.8) määrittelee stationaarisen prosessin, jos matriisin  $\mathbf{A}$  kaikki ominaisarvot ovat itseisarvoltaan ykköstä pienempiä. On helppo todeta, että tämä ehto voidaan esittää yhtäpitävästi käyttäen determinanttiehtoa

$$\det(I_{np} - \mathbf{A}z) \neq 0, \quad |z| \leq 1 \quad (z \in \mathbb{C}), \quad (2.9)$$

jossa  $|z|$  on kompleksiluvun  $z$  itseisarvo (ks. Liite A.2).<sup>5</sup> Jos yhtälön (2.8) määrittelemä prosessi  $\mathbf{y}_t$  on stationaarinen, niin samoin on yhtälön (2.7) määrittelemä prosessi  $y_t$  ( $y_t$  on lineaarinen muunnos  $\mathbf{y}_t$ :stä). Kun edellä mainittua ehtoa sovelletaan

<sup>5</sup>Jos  $z \in \mathbb{C}$  eli  $z = x + iy$  ( $i^2 = -1$ ), niin  $|z| = \sqrt{x^2 + y^2}$  on  $z$ :n itseisarvo.

yksiulotteiseen AR-prosessiin (2.6) tapauksessa  $p = 1$ , saadaan tuttu AR(1)-prosessin riittävä stationaarisuusehto  $|a_1| < 1$ .

Tarkastellaan ensin yhtälöä (2.8) tapauksessa  $t \geq 1$ , jolloin prosessi  $\mathbf{y}_t$  on hyvin määritelty, kun alkuarvo  $\mathbf{y}_0$  spesifioidaan. Peräkkäisillä sijoituksilla saadaan

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{A}\mathbf{y}_0 + \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2 &= \mathbf{A}\mathbf{y}_1 + \boldsymbol{\varepsilon}_2 \\ &= \mathbf{A}^2\mathbf{y}_0 + \mathbf{A}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 \\ &\vdots \\ \mathbf{y}_t &= \mathbf{A}^t\mathbf{y}_0 + \sum_{j=0}^{t-1} \mathbf{A}^j\boldsymbol{\varepsilon}_{t-j}. \end{aligned}$$

”Sopivan” alkuarvon määrittelemiseksi todetaan, että ehdon (2.9) voimassa ollessa  $\mathbf{A}^j \rightarrow 0$  geometrisesti, kun  $j \rightarrow \infty$ , mistä seuraa  $\sum_{j=0}^{\infty} \|\mathbf{A}^j\|^2 < \infty$  (ks. Liite A.3). Edellisessä jaksossa todetun nojalla voidaan siten määritellä lineaarinen prosessi

$$\mathbf{y}_t^* = \sum_{j=0}^{\infty} \mathbf{A}^j\boldsymbol{\varepsilon}_{t-j}, \quad t \in \mathbb{Z},$$

joka on sekä vahvasti että heikosti stationaarinen (eli vahvasti stationaarinen ja toiset momentit äärellisinä olemassa). Valitsemalla alkuarvoksi  $\mathbf{y}_0 = \mathbf{y}_0^*$  saadaan lineaarinen prosessi

$$\mathbf{y}_t = \mathbf{A}^t \sum_{j=0}^{\infty} \mathbf{A}^j\boldsymbol{\varepsilon}_{-j} + \sum_{j=0}^{t-1} \mathbf{A}^j\boldsymbol{\varepsilon}_{t-j} = \sum_{j=0}^{\infty} \mathbf{A}^j\boldsymbol{\varepsilon}_{t-j}, \quad t \geq 1.$$

Toisin sanoen,  $\{\mathbf{y}_t, t \geq 1\}$  on stationaarinen. Lisäksi on selvää, että näin määritelty prosessi ratkaisee yhtälön (2.8) ja voidaan laajentaa kaikille  $t \in \mathbb{Z}$  eli ehdon (2.9) voimassa ollessa yhtälö (2.8) määrittelee stationaarisen prosessin

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{A}^j\boldsymbol{\varepsilon}_{t-j}, \quad (t \in \mathbb{Z}). \quad (2.10)$$

Määrittelemällä matriisi  $J = [I_n : 0 : \dots : 0]$  ( $n \times np$ ) nähdään, että

$$\mathbf{y}_t = J\mathbf{y}_t \quad \text{ja} \quad \boldsymbol{\varepsilon}_t = J'\boldsymbol{\varepsilon}_t,$$

joten yhtälöstä (2.10) seuraa

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \Psi_j\boldsymbol{\varepsilon}_{t-j}, \quad (t \in \mathbb{Z}), \quad (2.11)$$

jossa  $\Psi_j = J\mathbf{A}^jJ'$  ja siten  $\Psi_0 = I_n$ . Koska  $\mathbf{A}^j \rightarrow 0$  geometrisesti, kun  $j \rightarrow \infty$ , pätee vastaava ominaisuus myös matriiseille  $\Psi_j$ , joten erityisesti  $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$ .

Jos prosessia  $y_t$  tarkastellaan rajoittuen ajankohtiin  $t \geq 1$  ja sallien alkuarvojen  $y_{-p+1}, \dots, y_0$  epästationaarisuus, on  $\{y_t, t \geq 1\}$  epästationaarinen. Edellä esitetystä nähdään, että alkuarvojen vaikutus häviää kuitenkin  $t$ :n kasvaessa ja prosessi ”stationarisoituu”. Tässä yhteydessä puhutaan usein *stabiilisuudesta* tai *asymptoottisesta stationaarisuudesta* ja ehtoa (2.9) kutsutaan *stabiilisuusehdoksi*. Myöhemmin esitettävät estimointi- ja testaustulokset soveltuvat myös tällaisten prosessien analysointiin.

**Stationaarisuusehdon vaihtoehtoinen muotoilu.** Määritellään mihin tahansa (mahdollisesti vektoriarvoiseen) jonoon  $x_t$  liittyvä viivästysoperaattori  $\mathbf{B}$  yhtälöllä  $\mathbf{B}x_t = x_{t-1}$  ja induktiivisesti  $\mathbf{B}^k x_t = x_{t-k}$  ( $\mathbf{B}^0 x_t = x_t$  ja  $\mathbf{B}^{-k} x_t = x_{t+k}$ ). Tällöin VAR(p)–prosessi (2.7) voidaan kirjoittaa lyhyesti

$$A(\mathbf{B})y_t = \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \Omega) \quad (t \in \mathbb{Z}), \quad (2.12)$$

jossa  $A(\mathbf{B}) = I_n - A_1\mathbf{B} - \dots - A_p\mathbf{B}^p$ . Voidaan osoittaa, että stationaarisuusehto (2.9) on yhtäpitävä seuraavan ehdon kanssa (perustelu jätetään tehtäväksi):

$$\det A(z) \neq 0, \quad |z| \leq 1. \quad (2.13)$$

Tässä  $A(z)$  on esimerkki polynomimatriisista eli matriisista, jonka alkiot ovat polynomeja (vastaavasti yhtälössä (2.12), jossa argumenttina on viivästysoperaattori  $\mathbf{B}$ ). Vaihtoehtoisesti  $A(z)$  voidaan tulkita matriisipolynomiksi eli polynomiksi, jonka kertoimet ovat matriiseja. Polynomimatriiseilla laskutoimitukset määritellään kuten tavanomaisilla reaalilukumatriiseilla (ks. Liite A.3). Yhtälö (2.12) voidaan siten ratkaista formaalisti kertomalla vasemmalta käänteismatriisilla  $A(\mathbf{B})^{-1}$ . Ratkaisuksi täytyy tietenkin tulla (2.11) eli lyhyesti ilmaistuna

$$y_t = \Psi(\mathbf{B})\varepsilon_t \quad (t \in \mathbb{Z}),$$

jossa  $\Psi(\mathbf{B}) = \sum_{j=0}^{\infty} \Psi_j \mathbf{B}^j = A(\mathbf{B})^{-1}$ . Stationaarisuusehtoa (2.13) voidaan havainnollistaa käyttämällä tulosta  $A(\mathbf{B})^{-1} = [1/\det(A(\mathbf{B}))] A(\mathbf{B})^*$ , jossa  $A(\mathbf{B})^*$  on (äärellisasteinen) matriisipolynomi (ks. Liite A.4). Ehdon (2.13) riittävyys stationaarisuudelle nähdään siten käyttäen (tunnetuksi oletettua) yksiulotteisen AR(p)–prosessin stationaarisuusehtoa.

Yhtälön (2.11) yhteydessä todettiin, että  $\Psi_j = J\mathbf{A}^j J'$ . Vaihtoehtoinen esitys matriiseille  $\Psi_j$  saadaan käyttäen identiteettiä  $\Psi(\mathbf{B})A(\mathbf{B}) = I_n$  eli

$$\begin{aligned} I_n &= (\Psi_0 + \Psi_1\mathbf{B} + \Psi_2\mathbf{B}^2 + \dots)(I_n - A_1\mathbf{B} - \dots - A_p\mathbf{B}^p) \\ &= \Psi_0 + (\Psi_1 - \Psi_0 A_1)\mathbf{B} + (\Psi_2 - \Psi_1 A_1 - \Psi_0 A_2)\mathbf{B}^2 + \dots \end{aligned}$$

Tästä seuraa

$$\begin{aligned} I_n &= \Psi_0 \\ 0 &= \Psi_1 - \Psi_0 A_1 \\ 0 &= \Psi_2 - \Psi_1 A_1 - \Psi_0 A_2 \\ &\vdots \end{aligned}$$

mistä nähdään edelleen, että matriisit  $\Psi_j$  voidaan laskea rekursiivisesti yhtälöistä

$$\begin{aligned}\Psi_0 &= I_n \\ \Psi_j &= \sum_{i=1}^j \Psi_{j-i} A_i, \quad j = 1, 2, \dots,\end{aligned}$$

jossa  $A_i = 0$ , kun  $i > p$ .

**Impulssivasteet.** Kuten aiemmin jo mainittiin, voidaan VAR(p)-malli ajatella usean lineaarisen regressiomallin systeeminä, jossa selittävät muuttujat ovat selitettävien muuttujien viivästettyjä arvoja. Tavanomaiseen regressiomalliin verrattuna selittävät muuttujat on siten valittu erityisellä tavalla. Ei-havaittavien muuttujien  $\varepsilon_t$  tulkinta jäännös- tai virhetermeinä on kuitenkin pitkälti samanlainen kuin regressiomalleissa. Regressiomallien tapaan voidaan kerroinmatriisien  $A_1, \dots, A_p$  alkioden tulkita kuvaavan selittävien muuttujien  $y_{t-1}, \dots, y_{t-p}$  komponenttien vaikutusta selitettäviin muuttujiin eli  $y_t$ :n komponentteihin. Näiden kerroinmatriisien alkioden tulkinnallinen hyöty on kuitenkin varsin rajallinen, sillä yleensä ollaan kiinnostuneita muuttujien välisistä dynaamisista vaikutuksista eli esimerkiksi siitä, miten  $y_{1t}$ :ssä tapahtuva muutos vaikuttaa kaikkien  $y_t$ :n komponenttien tulevaan kehitykseen. Tällaisten dynaamisten vaikutusten selvittäminen suoraan kerroinmatriisien  $A_1, \dots, A_p$  alkioden avulla on hankalaa. Tämän näkee havainnollisesti jo tapauksessa  $n = 2$  kaaviokuvasta, johon on merkitty nuolilla se, miten muuttujat  $y_{1t}$  ja  $y_{2t}$  vaikuttavat tuleviin arvoihinsa  $y_{1,t+j}$  ja  $y_{2,t+j}$ ,  $j = 1, 2, \dots$ . Vaikka kysymyksessä on vain kaksiulotteinen tapaus, tulee kuviosta varsin sekavan näköinen (piirrä kuva!).

Vaihtoehtoinen tulkinta perustuu ajatukseen, jossa virhetermin  $\varepsilon_t$  komponentit tulkitaan prosessiin  $y_t$  vaikuttaviksi impulsseiksi tai sokeiksi. Tällöin kiinnostava kysymys on, miten tällaisen yhden yksikön kokoisen impulssin vaikutus ilmenee prosessin  $y_t$  komponenteissa ja niiden tulevissa arvoissa. Tähän kysymykseen palataan yksityiskohtaisemmin myöhemmin. Yhtälön (2.11) perusteella on kuitenkin helppo uskoa, että matriisin  $\Psi_j$  alkiot kuvaavat tällaisen ajankohtana  $t_0$  tapahtuneen impulssin vaikutusta prosessin  $y_t$  komponenttien saamiin arvoihin ajankohtana  $t_0 + j$ . Tästä syystä matriisin  $\Psi_j$  alkioita sanotaan *impulssivasteiksi*.

**VAR(p)-prosessin odotusarvo ja kovarianssifunktio.** Koska stationaaraisella VAR(p)-prosessilla on lineaarinen esitys (2.11), voidaan sen odotusarvo ja kovarianssifunktio johtaa jaksossa 2.2 esitetyllä tavalla. Tulokseksi saadaan

$$E(y_t) = 0 \quad \text{ja} \quad \Gamma_k = \sum_{j=0}^{\infty} \Psi_j \Omega \Psi'_{j+k} = \Gamma'_{-k}, \quad k \geq 0,$$

jossa  $\Psi_j$  on kuten yhtälössä (2.11). Koska  $\Psi_j \rightarrow 0$  geometrisesti, kun  $j \rightarrow \infty$ , vaimenee myös kovarianssifunktio  $\Gamma_k$  geometrisesti nolnaan viipymän  $k$  kasvaessa rajatta.

Vaimeneminen on sitä hitaampaa mitä lähempänä ykköstä yhtälössä (2.8) määritellyn matriisin  $\mathbf{A}$  suurin ominaisarvo on itseisarvoltaan (vrt. yhtälöt (2.10) ja (2.11) ja huomaa, että tapauksessa  $n = p = 1$  matriisi  $\mathbf{A}$  on AR-kerroin ja  $\mathbf{A}^k$  on  $k$ . autokorrelaatiokerroin). Kovarianssifunktio voidaan selvittää myös seuraavalla vaihtoehdoisella tavalla.

Kertomalla yhtälö (2.7) oikealta (transponoidulla) vektorilla  $y'_{t-k}$  ja ottamalla odotusarvo puolittain saadusta yhtälöstä saadaan

$$\mathbf{E}(y_t y'_{t-k}) = A_1 \mathbf{E}(y_{t-1} y'_{t-k}) + \cdots + A_p \mathbf{E}(y_{t-p} y'_{t-k}) + \mathbf{E}(\varepsilon_t y'_{t-k}).$$

Kun  $k = 0$ , nähdään yhtälöstä (2.11), että  $\mathbf{E}(\varepsilon_t y'_t) = \mathbf{E}(\varepsilon_t \varepsilon'_t) = \Omega$ , joten ottamalla huomioon tulos  $\mathbf{E}(y_t) = 0$ , tulee edellä johdetusta yhtälöstä

$$\Gamma_0 = A_1 \Gamma_1 + \cdots + A_p \Gamma_p + \Omega.$$

Kun  $k > 0$ , nähdään vastaavasti, että  $\mathbf{E}(\varepsilon_t y'_{t-k}) = 0$  ja edelleen, että

$$\Gamma_{-k} = A_1 \Gamma_{1-k} + \cdots + A_p \Gamma_{p-k}, \quad k > 0,$$

tai yhtäpitävästi

$$\Gamma'_k = A_1 \Gamma'_{k-1} + \cdots + A_p \Gamma'_{k-p}, \quad k > 0.$$

Kuten yksiulotteisessa tapauksessakin, sanotaan näitä yhtälöitä *Yule-Walker -yhtälöiksi*. Niiden avulla voidaan kerroinmatriisit  $A_1, \dots, A_p$  lausua kovarianssimatriisien  $\Gamma'_0, \dots, \Gamma'_{p-1}$  funktiona (perustelu jätetään tehtäväksi). Lisäksi niiden avulla voidaan laskea kovarianssimatriisit  $\Gamma'_k$ ,  $k \geq p$ , kun kerroinmatriisit  $A_1, \dots, A_p$  on annettu ja kovarianssimatriisit  $\Gamma'_0, \dots, \Gamma'_{p-1}$  tunnetaan. Koko kovarianssifunktion  $\Gamma_k$  selvittämiseksi (VAR-mallin parametrien funktiona) on siten vielä selvitettävä kovarianssimatriisien  $\Gamma'_0, \dots, \Gamma'_{p-1}$  riippuvuus kerroinmatriiseista  $A_1, \dots, A_p$  ja (kuten ilmeistä) kovarianssimatriisista  $\Omega$ .

Koska  $\mathbf{E}(y_t) = 0$  ja  $\Gamma_{-k} = \Gamma'_k$ , pätee yhtälössä (2.8) määritellylle (stationaariselle) prosessille  $\mathbf{y}_t$

$$\begin{aligned} \text{Cov}(\mathbf{y}_t) &= \mathbf{E} \left( \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} \begin{bmatrix} y'_t & \cdots & y'_{t-p+1} \end{bmatrix} \right) \\ &= \begin{bmatrix} \Gamma_0 & \Gamma_{-1} & \cdots & \Gamma_{-p+1} \\ \Gamma_1 & \Gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Gamma_{-1} \\ \Gamma_{p-1} & \cdots & \Gamma_1 & \Gamma_0 \end{bmatrix} := \mathbf{\Gamma}_0. \end{aligned}$$

Koska  $E(\varepsilon_t \mathbf{y}'_{t-k}) = 0$ ,  $k > 0$ , on  $E(\varepsilon_t \mathbf{y}'_{t-1}) = 0$ , joten yhtälöstä (2.8) saadaan matriisille  $\Gamma_0$  yhtälö

$$\begin{aligned}\Gamma_0 &= E \left[ (\mathbf{A} \mathbf{y}_{t-1} + \varepsilon_t) (\mathbf{A} \mathbf{y}_{t-1} + \varepsilon_t)' \right] \\ &= \mathbf{A} E(\mathbf{y}_{t-1} \mathbf{y}'_{t-1}) \mathbf{A}' + E(\varepsilon_t \varepsilon_t') \\ &= \mathbf{A} \Gamma_0 \mathbf{A}' + \Omega,\end{aligned}$$

jossa  $\Omega = \text{Cov}(\varepsilon_t)$  kuten aikaisemminkin.

Matriisi  $\Gamma_0$  voidaan ratkaista matriisien  $\mathbf{A}$  ja  $\Omega$  funktiona edellä johdetusta yhtälöstä. Ratkaisukaavan esittämisessä tarvitaan ns.  $\text{vec}(\cdot)$ -operaattoria ja Kroneckerin tuloa. Edellinen muodostaa matriisista vektorin asettamalla sarakkeet alekkain alkaen ensimmäisestä (ks. Liite A.1). Dimensioista riippumatta matriisien  $A$  ja  $B$  Kroneckerin tulo  $A \otimes B$  määritellään yhtälöllä  $A \otimes B = [a_{ij} B]$ , jossa  $a_{ij}$  on matriisin  $A$  tyypillinen alkio (ks. Liite A.5). Jättäen perustelut tehtäväksi todetaan, että mille tahansa matriiseille  $A$ ,  $B$  ja  $C$  pätee  $\text{vec}(A + B) = \text{vec}(A) + \text{vec}(B)$  ja  $\text{vec}(ABC) = (C' \otimes A) \text{vec}(B)$  (olettaen summa ja tulo määritellyiksi). Edellä johdetusta yhtälöstä saadaan siten

$$\text{vec}(\Gamma_0) = (\mathbf{A} \otimes \mathbf{A}) \text{vec}(\Gamma_0) + \text{vec}(\Omega)$$

eli

$$(I_{n^2 p^2} - \mathbf{A} \otimes \mathbf{A}) \text{vec}(\Gamma_0) = \text{vec}(\Omega),$$

josta voidaan ratkaista<sup>6</sup>

$$\text{vec}(\Gamma_0) = (I_{n^2 p^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}(\Omega).$$

Kun matriisit  $A_1, \dots, A_p$  ja  $\Omega$  tunnetaan, tunnetaan tämän yhtälön oikea puoli ja matriisin  $\Gamma_0$  määritelmän nojalla siten myös matriisit  $\Gamma_0, \dots, \Gamma_{p-1}$ .

**VAR(p)-prosessin ennustaminen.** Tarkastellaan VAR(p)-prosessin ennustamista teoreettisessa tilanteessa, jossa prosessin parametrien arvot tunnetaan. Käytännössä joudutaan tietenkin käyttämään estimaatteja.

Todennäköisyyslaskennasta tiedetään, että  $sm:n$   $Y$   $sv:iin$   $X$  perustuva ”optimaalinen” ennuste on  $E(Y|X)$  eli  $Y:n$  ehdollinen odotusarvo ehdolla  $X$  (ks. Liite A.6, jossa motivoidaan myös merkintä  $E(Y|X)$  ja esitetään tarvittavat ehdollisen odotusarvon ominaisuudet). Tässä ”optimaalinen” viittaa ennusteeseen, joka minimoi ennusteen keskineliövirheen. Jos  $Y$  on vektori, pätee sama tulos, kun kriteerinä on keskineliövirhematriisin minimointi. Tällä tarkoitetaan seuraavaa. Olkoon  $\tilde{Y} = g(X)$

---

<sup>6</sup>Käänteismatriisin olemassaolo seuraa, sillä stationaarisuusoletuksen mukaan matriisin  $\mathbf{A}$  ominaisarvot ovat itseisarvoltaan ykköstä pienempiä ja matriisin  $\mathbf{A} \otimes \mathbf{A}$  ominaisarvojen voidaan puolestaan osoittaa olevan  $\mathbf{A}$ :n ominaisarvojen tuloja, joten nekin ovat itseisarvoltaan ykköstä pienempiä. Tästä seuraa erityisesti, että matriisi  $I_{n^2 p^2} - \mathbf{A} \otimes \mathbf{A}$  on epäsingulaarinen ja mainittu (yksikäsitteinen) ratkaisu on olemassa.

mikä tahansa sv:n  $Y$  ennuste, joka on sv:n  $X$  funktio. Tällöin

$$\text{MSE}(\tilde{Y}) = \mathbf{E}[(Y - \tilde{Y})(Y - \tilde{Y})']$$

on ennusteen  $\tilde{Y}$  keskineliövirhematriisi ( $= \mathbf{E}[(Y - \tilde{Y})^2]$  reaaliarvoisessa tapauksessa) ja (perustelu jätetään tehtäväksi)

$$a' \text{MSE}(\tilde{Y}) a \geq a' \text{MSE}(\mathbf{E}(Y|X)) a$$

kaikilla dimensioltaan sopivilla vektoreilla  $a$ . Toisin sanoen, matriisi  $\text{MSE}(\tilde{Y}) - \text{MSE}(\mathbf{E}(Y|X))$  on positiivisesti semidefiniitti eli symbolein  $\text{MSE}(\tilde{Y}) - \text{MSE}(\mathbf{E}(Y|X)) \geq 0$ . Valitsemalla  $a = (0, \dots, 0, 1, 0, \dots, 0)$  nähdään, että ennusteen  $\tilde{Y}$  jokaisen komponentin keskineliövirhe on vähintään yhtäsuuri kuin ennusteella  $\mathbf{E}(Y|X)$ .

Sovelletaan nyt edellä esitettyä VAR(p)-prosessiin (2.7). Muodostetaan ensin ennuste muuttujalle  $y_{t+1}$ , kun prosessin koko edeltävä historia  $y_t, y_{t-1}, \dots$  tunnetaan.<sup>7</sup> Prosessin määrittely-yhtälöstä (2.7) nähdään, että

$$\mathbf{E}_t(y_{t+1}) := \mathbf{E}(y_{t+1}|y_{t-j}, j \geq 0) = \mathbf{E}(y_{t+1}|y_t, \dots, y_{t-p+1}) =: y_t(1).$$

Yhtälöstä (2.11) ja virheiden riippumattomuudesta nähdään lisäksi, että  $\varepsilon_t \perp\!\!\!\perp y_{t-j}$ ,  $j > 0$ . Käyttäen yleisesti merkintää  $\mathbf{E}_t(\cdot) = \mathbf{E}(\cdot|y_{t-j}, j \geq 0)$  saadaan siten  $\mathbf{E}_t(\varepsilon_{t+j}) = \mathbf{E}(\varepsilon_{t+j}) = 0$  ( $j > 0$ ) ja

$$\begin{aligned} y_t(1) &= A_1 \mathbf{E}_t(y_t) + \dots + A_p \mathbf{E}_t(y_{t+1-p}) + \mathbf{E}_t(\varepsilon_{t+1}) \\ &= A_1 y_t + \dots + A_p y_{t+1-p}. \end{aligned}$$

Yleisemmin, kun  $h \geq 1$ , saadaan

$$\mathbf{E}_t(y_{t+h}) = A_1 \mathbf{E}_t(y_{t+h-1}) + \dots + A_p \mathbf{E}_t(y_{t+h-p})$$

eli toisin merkinnöin

$$y_t(h) = A_1 y_t(h-1) + \dots + A_p y_t(h-p),$$

jossa  $y_t(h-j) = \mathbf{E}_t(y_{t+h-j})$  ja  $y_t(h-j) = y_{t+h-j}$ , kun  $h \leq j$ . Toisin sanoen,  $y_t(h-j)$  on prosessin havaittu arvo  $y_{t+h-j}$ , kun sellainen on ennustejankohtana  $t$  tiedossa, ja muulloin ajankohtana  $t$  tehty ennuste  $\mathbf{E}_t(y_{t+h-j})$ .

Käyttäen mallin lineaarista esitystä (2.11) saadaan ennusteelle vaihtoehtoinen esitys

$$y_t(h) = \mathbf{E}_t(y_{t+h}) = \sum_{j=0}^{\infty} \Psi_j \mathbf{E}_t(\varepsilon_{t+h-j}) = \sum_{j=h}^{\infty} \Psi_j \varepsilon_{t+h-j},$$

<sup>7</sup>Vaihtoehtoisesti voitaisiin tarkastella tulevien arvojen  $y_{T+h}$ ,  $h \geq 1$ , ennustamista, kun käytävissä on havaittu aikasarja  $y_1, \dots, y_T$ . VAR(p)-prosessin tapauksessa näillä kahdella tavalla ei ole eroa, kun  $T > p$ .



josta voidaan johtaa ennustevirheelle yhtälö

$$y_{t+h} - y_t(h) = \sum_{j=0}^{h-1} \Psi_j \varepsilon_{t+h-j} \quad (\Psi_0 = I_n).$$

Tästä nähdään, että ennustevirheen odotusarvo on nolla eli ennuste on harhaton ja kovarianssimatriisi eli keskineliövirhematriisi on  $\sum_{j=0}^{h-1} \Psi_j \Omega \Psi_j'$ . Normaaliossa  $\varepsilon_t \sim \text{nid}(0, \Omega)$  saadaan ennustevirheelle siten tulos

$$y_{t+h} - y_t(h) \sim \mathbf{N} \left( 0, \sum_{j=0}^{h-1} \Psi_j \Omega \Psi_j' \right),$$

jonka avulla voidaan muodostaa luottamusvälejä ennustettavan  $y_{t+h}$  komponenteille. Jos  $\sigma_i^2(h)$  on matriisin  $\sum_{j=0}^{h-1} \Psi_j \Omega \Psi_j'$   $i$ . diagonaali-alkio, niin ilmeisin merkinnöin

$$(y_{i,t+h} - y_{i,t}(h)) / \sigma_i(h) \sim \mathbf{N}(0, 1)$$

ja

$$\begin{aligned} 0.95 &= \mathbf{P} \left\{ -1.96 \leq \frac{y_{i,t+h} - y_{i,t}(h)}{\sigma_i(h)} \leq 1.96 \right\} \\ &= \mathbf{P} \{ y_{i,t}(h) - 1.96\sigma_i(h) \leq y_{i,t+h} \leq y_{i,t}(h) + 1.96\sigma_i(h) \}. \end{aligned}$$

Toisin sanoen, 95%:n luottamusväli ennustettavan muuttujan  $y_{t+h}$   $i$ :nnelle komponentille on  $y_{i,t}(h) \pm 1.96\sigma_i(h)$ .

Käytännössä tuntemattomat parametrit korvataan estimaateilla, jolloin edellä esitetty pätee vain likimäärisesti. Huomaa myös, että esitettyä luottamusväliä johdettaessa indeksit  $i$  ja  $h$  oletettiin kiinnitettyiksi. Tarkasteltaessa useita tällaisia luottamusvälejä joudutaan ”tavanomaisiin” yhdistettyjen luottamusvälien ongelmiin.

**VAR(p)–prosessi, kun  $\mathbf{E}(y_t) \neq 0$ .** Käytännössä on yleensä epärealistista olettaa, että havaitut aikasarjat tuottaneen prosessin odotusarvo olisi nolla. Kaikki VAR(p)–prosessista edellä esitetyt tulokset voidaan yleistää helposti tapaukseen  $\mathbf{E}(y_t) \neq 0$ . Prosessin  $y_t$  paikalle asetetaan vain sen keskistetty version  $y_t - \mu$ , jolloin stationaarisuusehdon voimassa ollessa  $\mu = \mathbf{E}(y_t)$ . Koska  $A(\mathbf{B})(y_t - \mu) = A(\mathbf{B})y_t - A(1)\mu$ , voidaan vaihtoehtoisesti tarkastella prosessia

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \Omega) \quad (t \in \mathbb{Z}),$$

jossa  $\nu = A(1)\mu$  ja muut merkinnät ovat kuten aikaisemmin. Kuten stationaarisuusehdosta (2.13) nähdään, on matriisi  $A(1)$  epäsingulaarinen, joten odotusarvon  $\mu$  ja vakiotermin  $\nu$  välinen suhde voidaan kirjoittaa myös  $\mu = A(1)^{-1}\nu$ .

## 2.4 VARMA(p,q)–prosessi

Yksiulotteisessa tapauksessa AR(p)–prosessi voidaan yleistää ARMA(p,q)–prosessiksi lisäämällä yhtälön (2.6) oikealle puolelle innovaatioiden  $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$  lineaarikombinaatio. Vastaava moniulotteinen yleistys on

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t - M_1 \varepsilon_{t-1} - \dots - M_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{iid}(0, \Omega) \quad (t \in \mathbb{Z}), \quad (2.14)$$

jossa  $M_1, \dots, M_q$  ovat  $n \times n$  kerroinmatriiseja ja muut merkinnät ovat kuten yhtälössä (2.7). Tässä yhteydessä käytetään nimitystä *VARMA(p,q)–prosessi* tai *VARMA(p,q)–malli*.

Samaan tapaan kuin yksiulotteisessa tapauksessa voidaan todeta, että VARMA(p,q)–prosessin riittävä stationaarisuusehto on (2.9) tai vaihtoehtoisesti (2.13) eli sama kuin VAR(p)–prosessilla. Käyttäen viivästysoperaattoria voidaan yhtälö (2.14) kirjoittaa

$$A(\mathbf{B}) y_t = M(\mathbf{B}) \varepsilon_t,$$

jossa  $A(\mathbf{B}) = I_n - A_1 \mathbf{B} - \dots - A_p \mathbf{B}^p$  ja  $M(\mathbf{B}) = I_n - M_1 \mathbf{B} - \dots - M_q \mathbf{B}^q$ . Stationaarisuusehdon voimassa ollessa voidaan tässä yhtälössä kertoa vasemmalta käänteismatriisilla  $A(\mathbf{B})^{-1}$  ja saada ratkaisuksi hyvin määritelty lineaarinen prosessi

$$y_t = \Psi(\mathbf{B}) \varepsilon_t,$$

jossa nyt  $\Psi(\mathbf{B}) = \sum_{j=0}^{\infty} \Psi_j \mathbf{B}^j = A(\mathbf{B})^{-1} M(\mathbf{B})$  ja  $\Psi_0 = I_n$ . Kuten VAR(p)–prosessin tapauksessakin,  $\Psi_j \rightarrow 0$  geometrisesti, kun  $j \rightarrow \infty$ , ja siten myös VARMA(p,q)–prosessin kovarianssifunktio vaimenee geometrisesti nolnaan viipymän kasvaessa rajatta.

VARMA(p,q)–prosessien teoria ja soveltaminen on huomattavasti hankalampaa kuin VAR(p)–prosessien. Tämä johtuu pitkälti polynomimatriisien ominaisuuksista, joiden vuoksi VARMA(p,q)–prosessien rakenneteoria ja siihen liittyvä mallinvalinta tulee monimutkaisemmaksi kuin VAR(p)–prosesseilla. Tämä näkyy erityisesti yksikäsitteisyys- tai identifioituvuusehdossa, joka yksiulotteisessa tapauksessa  $n = 1$  vaatii, että (nyt) *polynomeilla*  $A(z)$  ja  $M(z)$  ei ole yhteisiä tekijöitä ja ainakin toinen korkeimman asteen kertoimista on nollasta poikkeava.

Seuraava esimerkki havainnollistaa miksi yksikäsitteisyyskysymys hankaloituu moniulotteisessa tapauksessa  $n > 1$ . Tarkastellaan kaksiulotteista VARMA(1,1)–prosessia

$$\begin{bmatrix} 1 & a\mathbf{B} \\ 0 & 1 \end{bmatrix} y_t = \begin{bmatrix} 1 & m\mathbf{B} \\ 0 & 1 \end{bmatrix} \varepsilon_t, \quad a \neq m. \quad (2.15)$$

Ensimmäisen asteen polynomimatriisit  $A(\mathbf{B})$  ja  $M(\mathbf{B})$  näyttäisivät toteuttavan yksiulotteisen ARMA(1,1)–prosessin yksikäsitteisyys ehdon kaltaisen ehdon. Kertomalla yhtälö (2.15) vasemmalta käänteismatriisilla

$$A(\mathbf{B})^{-1} = \begin{bmatrix} 1 & -a\mathbf{B} \\ 0 & 1 \end{bmatrix}$$

saadaan prosessin lineaarinen esitys

$$y_t = \begin{bmatrix} 1 & (m-a)\mathbf{B} \\ 0 & 1 \end{bmatrix} \varepsilon_t,$$

joka on vektoriarvoinen MA(1)–prosessi eli lyhyesti VMA(1)–prosessi (vastaavasti puhutaan MA(q)–prosessista, jossa  $q = \infty$  on myös mahdollinen). Kertomalla yhtälö (2.15) vasemmalta käänteismatriisilla  $M(\mathbf{B})^{-1}$  saadaan vastaavasti

$$\begin{bmatrix} 1 & (a-m)\mathbf{B} \\ 0 & 1 \end{bmatrix} y_t = \varepsilon_t$$

eli VAR(1)–prosessi. VARMA(1,1)–prosessi (2.15) voidaan siis lausua myös VAR(1)– ja VMA(1)–prosessina. Jos tällaisessa tilanteessa sovitetaan aineistoon VARMA(1,1)–prosessi, joutuu estimointimenetelmä hankaluuksiin, koska löytyy kolme eri mallia, jotka sopivat aineistoon yhtä hyvin. Ongelma ratkeaa tietysti, jos prosessin rakenteesta on tietoa niin, että estimoitavaksi voidaan valita joko VAR(1)– tai VMA(1)–malli. Yleensä tällaista tietoa ei kuitenkaan ole ja monimutkaisemmissa tapauksissa ongelman tunnistaminenkaan ei onnistu yhtä helposti kuin tässä tarkastellussa yksinkertaisessa esimerkissä.

VARMA(p,q)–prosesseille on kehitetty varsin täydellinen yksikäsitteisyysteoria, johon perustuen mallinvalintakin voidaan hoitaa käytännössä ns. mallinvalintakriteerejä käyttäen. Myös (normaalijakaumaan perustuva) suurimman uskottavuuden estimointiteoria on kehitetty. Näiden kysymysten tarkastelu sivuutetaan kuitenkin tällä kurssilla. Esitettyjä VAR(p)–prosesseihin perustuvia tilastollisia menetelmiä voidaan perustella approksimaatioina, sillä käännettävyysehdon

$$\det(M(z)) \neq 0, \quad |z| \leq 1,$$

voimassa ollessa, VARMA(p,q)–prosessi voidaan kirjoittaa

$$M(\mathbf{B})^{-1} A(\mathbf{B}) y_t = \varepsilon_t,$$

jossa  $M(\mathbf{B})^{-1} A(\mathbf{B}) = \sum_{j=0}^{\infty} \Pi_j \mathbf{B}^j$  ja kerroinmatriisit  $\Pi_j$  konvergoivat nolnaan geometrisesti, kun  $j \rightarrow \infty$  ( $\Pi_0 = I_n$ ). VARMA(p,q)–prosessia voidaan siten approksimoida VAR(h)–prosessilla, kun aste  $h$  on ”tarpeeksi suuri”.

### 3 Stationaarisen prosessin ensimmäisten ja toisten momenttien estimointi

**Otoskeskiarvon tarkentuvuus.** Olkoon  $y_t = (y_{1t}, \dots, y_{nt})$  stationaarinen prosessi odotusarvona  $E(y_t) = \mu$  ja kovarianssifunktiona  $\text{Cov}(y_t, y_{t+k}) = \Gamma_k = [\gamma_{ij,k}]$ . Oletetaan, että (vrt. ehto (2.1))

$$\sum_{k=-\infty}^{\infty} \|\Gamma_k\| < \infty. \quad (3.1)$$

Tarkastellaan otoskeskiarvoa  $\bar{y} = T^{-1} \sum_{t=1}^T y_t$  odotusarvon  $\mu$  estimaattorina. Harhattomuus on ilmeisen, sillä

$$\mathbf{E}(\bar{y}) = T^{-1} \sum_{t=1}^T \mathbf{E}(y_t) = \mu.$$

Kun merkitään  $\bar{y} = (\bar{y}_1, \dots, \bar{y}_n)$  ja  $\mu = (\mu_1, \dots, \mu_n)$  saadaan lisäksi

$$\begin{aligned} \text{Var}(\bar{y}_i) &= \mathbf{E} \left[ \left( T^{-1} \sum_{t=1}^T (y_{it} - \mu_i) \right)^2 \right] \\ &= T^{-2} \sum_{t=1}^T \sum_{s=1}^T \mathbf{E} [(y_{it} - \mu_i)(y_{is} - \mu_i)] \\ &= T^{-2} \sum_{t-s=-T}^T (T - |t-s|) \gamma_{i,t-s} \\ &= T^{-1} \sum_{k=-T}^T \left( 1 - \frac{|k|}{T} \right) \gamma_{i,k}. \end{aligned}$$

Tässä kolmannen yhtälön voi todeta huomaamalla, että edeltävä kaksoissumma on matriisin  $[\gamma_{i,t-s}]_{t,s=1,\dots,T}$  alkioiden summa. Kolmioepäyhtälön ja oletuksen (3.1) avulla saadaan tulos

$$\text{Var}(\bar{y}_i) = \mathbf{E} [(\bar{y}_i - \mu_i)^2] \leq T^{-1} \sum_{k=-T}^T \left( 1 - \frac{|k|}{T} \right) |\gamma_{i,k}| \rightarrow 0, \quad \text{kun } T \rightarrow \infty.$$

Yhdistämällä edellä todettu voidaan päätellä (Markovin epäyhtälöä käyttäen), että otoskeskiarvo  $\bar{y}_i$  on odotusarvon  $\mu_i$  tarkentuva estimaattori ja edelleen, että sama pätee otoskeskiarvolle  $\bar{y}$ . Toisin sanoen, on todettu stokastinen konvergenssi

$$\bar{y} \xrightarrow{p} \mu. \quad (3.2)$$

Oletus (3.1) on voimassa erityisesti VARMA(p,q)-prosesseille, koska tällöin  $\Gamma_k \rightarrow 0$  geometrisesti, kun  $|k| \rightarrow \infty$ .

**Otoskovarianssifunktion tarkentuvuus.** Tarkastellaan nyt otoskovarianssifunktiota

$$C_k = [c_{ij,k}] = (T-k)^{-1} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})', \quad k = 0, \dots, T-1,$$

$C_k = C'_{-k}$ ,  $k = -T+1, \dots, -1$ . Kiinnitetään indeksit  $i, j$  ja  $k \geq 0$  ja määritellään yksinkertaisuuden vuoksi muuttuja  $x_t = y_{it}y_{j,t+k}$ , joka on keskeinen osa estimaattoria

$c_{ij,k}$ . Oletetaan  $\mathbf{E}(y_{it}^4) < \infty$  ( $i = 1, \dots, n$ ), jolloin  $x_t$  on stationaarinen ja sillä on äärellinen toinen momentti. Jos oletetaan, että  $x_t$ :n autokovarianssifunktio toteuttaa ehdon (3.1) vastineen, saadaan edellä otoskeskiarvolle todetun perusteella tulos

$$\bar{x} = (T - k)^{-1} \sum_{t=1}^T y_{it} y_{j,t+k} \xrightarrow{p} \mathbf{E}(y_{it} y_{j,t+k}),$$

jossa kovarianssin määritelmän perusteella

$$\mathbf{E}(y_{it} y_{j,t+k}) = \gamma_{ij,k} + \mathbf{E}(y_{it}) \mathbf{E}(y_{j,t+k}) = \gamma_{ij,k} + \mu_i \mu_j.$$

Olettaen, että edellä tehdyt oletukset pätevät kaikilla  $i, j$  ja  $k \geq 0$  voidaan estimaattorin  $\bar{x}$  tarkentuvuudesta ja tuloksesta (3.2) johtaa edelleen tulokset  $c_{ij,k} \xrightarrow{p} \gamma_{ij,k}$  ja

$$C_k \xrightarrow{p} \Gamma_k. \quad (3.3)$$

Yksityiskohdat, joissa tarvitaan lisäksi stokastisen konvergenssin jatkuvan kuvauksen lausetta (ks. Liite B, Lause B.1), jätetään tehtäväksi.

Voidaan siis todeta, että kohtuullisin oletuksin otoskovarianssifunktio on tarkentuva. (Huomaa kuitenkin, että edellä  $k$  on kiinteä.) Tarvittavat oletukset on helppo tarkistaa erikoistapauksessa  $y_t \sim \text{iid}(\mu, \Sigma)$  ja sen jälkeen edelleen VMA(q)-prosessin tapauksessa ( $q < \infty$ ). Yleisillä VARMA(p,q)-prosesseilla tarkentuvuustulos (3.3) voidaan perustella käyttäen approksimaatiota VARMA(p,q)  $\approx$  VMA(h), kun  $h$  on ”suuri”. Yksityiskohdat ovat kuitenkin hieman teknisiä ja sivuutetaan.

**Asymptoottinen normaalisuus.** Edellä esitetyt tulokset havainnollistavat sitä, että stationaarisilla prosesseilla suurten lukujen lait pätevät ”varsin yleisin oletuksin”. Seuraavassa yritetään havainnollistaa sitä, että sama pätee keskeiselle raja-arvolauseelle.

Tarkastellaan yksinkertaisuuden vuoksi ( $n$ -ulotteista) lineaarista prosessia

$$y_t = \mu + \Psi(\mathbf{B}) \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \Omega), \quad (3.4)$$

jossa  $\Psi(\mathbf{B}) = \sum_{j=0}^{\infty} \Psi_j \mathbf{B}^j$  ja  $\Psi_0 = I_n$ . Oletetaan lisäksi vahvemmin kuin jaksossa 2.2 eli

$$\sum_{j=1}^{\infty} j \|\Psi_j\| < \infty.$$

Tällöin voidaan kirjoittaa  $\Psi(\mathbf{B}) = \Psi(1) + \Delta G(\mathbf{B})$ , jossa  $\Delta = 1 - \mathbf{B}$  on differensointioperaattori ja  $G(\mathbf{B}) = \sum_{j=0}^{\infty} G_j \mathbf{B}^j$  toteuttaa ehdon  $\sum_{j=0}^{\infty} \|G_j\| < \infty$  ja siten  $\sum_{j=0}^{\infty} \|G_j\|^2 < \infty$  (ks. Liite A. 4). Soveltamalla tätä yhtälössä (3.4) saadaan

$$y_t - \mu = \Psi(1) \varepsilon_t + \Delta u_t, \quad u_t = G(\mathbf{B}) \varepsilon_t.$$

Edellä todetun nojalla  $u_t$  on lineaarinen prosessi ja sen komponenteille pätee  $\mathbf{E}(u_{it}^2) < \infty$  (ks. jakso 2.2). Otoskeskiarvolle  $\bar{y}$  saadaan siten

$$\begin{aligned}\sqrt{T}(\bar{y} - \mu) &= \Psi(1) \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t + \frac{1}{\sqrt{T}} \sum_{t=1}^T (u_t - u_{t-1}) \\ &= \Psi(1) \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t + \frac{1}{\sqrt{T}} (u_T - u_0).\end{aligned}$$

Koska  $\varepsilon_t \sim \text{iid}(0, \Omega)$ , niin tavanomaisen (moniuulotteisen) keskeisen raja-arvolauseen nojalla pätee jakaumakonvergenssi  $T^{-1/2} \sum_{t=1}^T \varepsilon_t \xrightarrow{d} \mathbf{N}(0, \Omega)$ . Toisaalta, koska  $\mathbf{E}(u_{it}) = 0$  ja  $\mathbf{E}(u_{it}^2) \leq c < \infty$  kaikilla  $i = 1, \dots, n$  ja kaikilla  $t \in \mathbb{Z}$ , saadaan Markovin epäyhtälöä käyttäen  $T^{-1/2} u_T \xrightarrow{p} 0$  ja  $T^{-1/2} u_0 \xrightarrow{p} 0$  eli  $T^{-1/2}(u_T - u_0) \xrightarrow{p} 0$ . Käyttäen tunnettua jakaumakonvergenssia ja stokastista konvergenssia koskevaa jatkuvan kuvauksen lausetta (ks. Liite B, Lause B.3) voidaan kaiken kaikkiaan päätellä, että sv:lla  $\sqrt{T}(\bar{y} - \mu)$  ja  $\Psi(1) T^{-1/2} \sum_{t=1}^T \varepsilon_t$  on sama asymptoottinen jakauma ja

$$\sqrt{T}(\bar{y} - \mu) \xrightarrow{d} \mathbf{N}(0, \Psi(1) \Omega \Psi(1)') . \quad (3.5)$$

Otoskeskiarvon keskeinen raja-arvolause voidaan osoittaa myös ilman lineaarisuusoletusta (3.4), mutta yhtä kätevää perustelua kuin edellä ei ole helppo esittää. Oleellinen vaatimus useissa vastaavissa tuloksissa on, että sv:t  $y_t$  ja  $y_{t-k}$  ”tulevat riippumattomiksi”, kun  $k \rightarrow \infty$ . Lineaarisen prosessin (3.4) tapauksessa  $y_{t-k}$  riippuu virhetermeistä  $\varepsilon_{t-k}, \varepsilon_{t-k-1}, \dots$  ja  $y_t$  riippuu näistä (ja siten  $y_{t-k}$ :sta) vain muunnosten  $\Psi_k \varepsilon_{t-k}, \Psi_{k+1} \varepsilon_{t-k-1}, \dots$  kautta. Koska  $\Psi_k \rightarrow 0$ , kun  $k \rightarrow \infty$ , ”häviää” sv:ien  $y_t$  ja  $y_{t-k}$  riippuvuus  $k$ :n kasvaessa.

Tuloksen (3.5) perustelussa käytettyä tekniikkaa voidaan soveltaa myös otoskovarianssifunktion  $C_k$  tapauksessa, mutta algebralliset laskelmat tulevat ikäviksi. Koska saatava tulos on lisäksi monimutkainen erikoistapauksia (lähinnä VMA(q)- ja iid-tapausta) lukuun ottamatta, ei sitä käsitellä. Käytännössä kovarianssit korvataan yleensä vastaavilla korrelaatioilla eli käytetään otossuureita

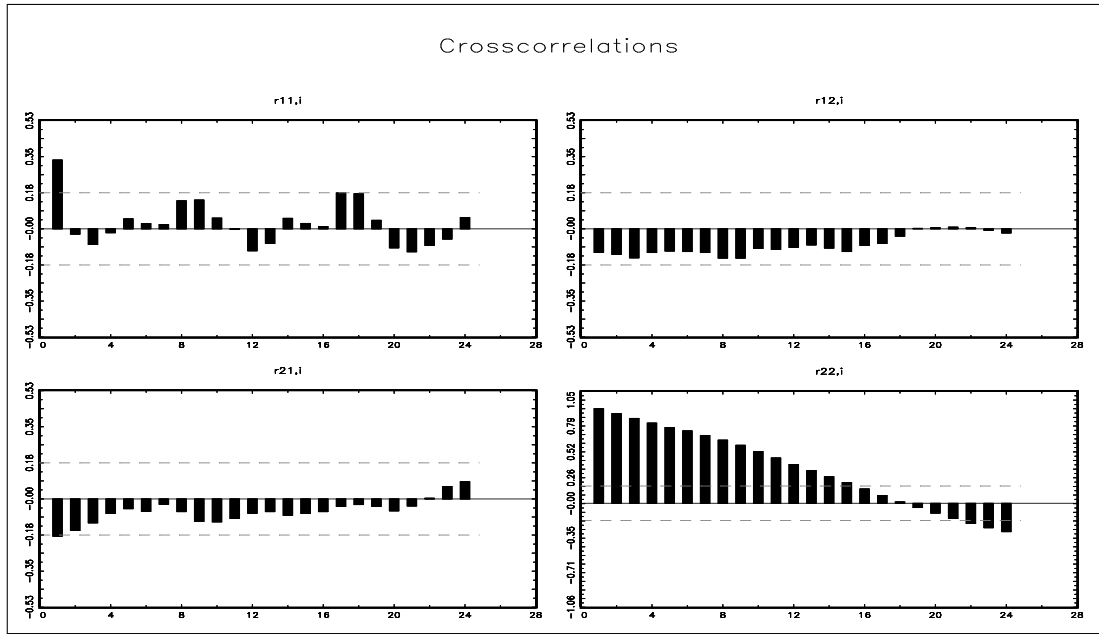
$$r_{ab,k} = \frac{c_{ab,k}}{\sqrt{c_{a,0} c_{b,0}}}, \quad c_{a,0} = c_{aa,0}.$$

Voidaan osoittaa, että tapauksessa  $y_t \sim \text{iid}(\mu, \Sigma)$  pätee asymptoottinen tulos

$$r_{ab,k} \underset{as}{\sim} \mathbf{N}(0, 1/T), \quad k \neq 0. \quad (3.6)$$

Lisäksi, jos  $R_k = [r_{ab,k}]$ , ovat  $R_k$  ja  $R_l$  asymptoottisesti riippumattomia, kun  $k \neq l$ .

**Esimerkki 3.1.** Kuviossa 3.1 nähdään Kuvion 1.1 vaihtokurssi/korkoaineiston kahdesta aikasarjasta lasketut auto- ja ristikorrelaatiofunktiot. Kuvioon on piirretty



**Kuvio 3.1.** Euron ja Yhdysvaltojen dollarin välisen kuukausittaisen vaihtokurssin muutoksen ( $y_{1t}$ ) ja euroalueen ja Yhdysvaltojen 10 vuoden valtioiden obligaatioiden keskiporkkojen eron ( $y_{2t}$ ) estimoidut autokorrelaatio- ja ristikorrelaatiofunktiot ( $t = 1999\text{I}, \dots, 2009\text{XI}$ ,  $T = 131$ ).

myös kriittiset rajat  $\pm 2/\sqrt{T}$ , joiden sisään yksittäisen korrelaation pitäisi jäädä likimain 95%:n todennäköisyydellä, mikäli kysymyksessä olisi iid-prosessista saatu reaalisiaatio. Estimoidut autokorrelaatiot vahvistavat jo Kuvioista 1.1 nähtävän selvän autokorrelotuneisuuden. Estimoidut ristikorrelaatiot näyttävät sen sijaan pieniltä. Huomaa kuitenkin, että komponenttisarjojen autokorrelotuneisuuden vuoksi kuvion kriittisiä rajoja ei voida soveltaa ristikorrelaatioestimaatteihin edellä kuvatulla tavalla.

## 4 Stationaarisen VAR-mallin teoriaa

### 4.1 Parametrien rajoittamaton estimointi

**Mallin määrittely.** Oletetaan, että havaittuun  $n$ :n aikasarjan aineistoon voidaan soveltaa VAR(p)-mallia

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \quad t = 1, \dots, T, \quad (4.1)$$

jossa alkuarvot  $y_{-p+1}, \dots, y_0$  on myös havaittu ja kerroinmatriisit  $A_1, \dots, A_p$  ( $n \times n$ ) toteuttavat stationaarisuus- tai stabiilisuusehdon (2.9) (tai yhtäpitävästi ehdon (2.13)). Jotta parametrien estimoinnissa ja testauksessa päästäisiin soveltamaan suurimman uskottavuuden (SU) teoriaa, liitetään virhetermiin normaalisuusoletus

$$\varepsilon_t \sim \text{nid}(0, \Omega), \quad (4.2)$$

jossa kovarianssimatriisi  $\Omega$  on positiivisesti definiitti. Todettakoon kuitenkin, että esitettävät asymptoottiset tulokset pätevät yleisemmin oletuksella  $\varepsilon_t \sim \text{iid}(0, \Omega)$ .

**Uskottavuusfunktion johto.** Otetaan käyttöön merkinnät

$$x_t = [1 \ y'_{t-1} \ \cdots \ y'_{t-p}]' = [1 \ \mathbf{y}'_{t-1}]' \quad ((np+1) \times 1)$$

ja

$$\Pi = \begin{bmatrix} \pi'_1 \\ \vdots \\ \pi'_n \end{bmatrix} = [\nu : A_1 : \cdots : A_p] \quad (n \times (np+1)).$$

Tällöin yhtälö (4.1) voidaan kirjoittaa

$$y_t = \Pi x_t + \varepsilon_t, \quad t = 1, \dots, T. \quad (4.3)$$

Tämä on yhtälöittäin esitettynä ( $\pi'_i x_t = x'_t \pi_i$ )

$$y_{it} = x'_t \pi_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (4.4)$$

mistä nähdään, että yhtälö (4.3) voidaan kirjoittaa

$$y_t = X'_t \pi + \varepsilon_t, \quad t = 1, \dots, T, \quad (4.5)$$

jossa  $\pi = [\pi'_1 \ \cdots \ \pi'_n]'$  ( $n(np+1) \times 1$ ) ja

$$X'_t = \text{diag}[x'_t \ \cdots \ x'_t] = I_n \otimes x'_t \quad (n \times n(np+1)).$$

Uskottavuusfunktion johtoa varten merkitään  $\mathbf{Y}_t = [y'_{-p+1} \ \cdots \ y'_0 \ y'_1 \ \cdots \ y'_t]'$ ,  $t = 0, 1, \dots, T$ . Jättäen argumentit yksinkertaisuuden vuoksi pois merkitään satunnaisvektorin  $\mathbf{Y}_t$  yhteistiheysfunktioita symbolilla  $f_{\mathbf{Y}_t}$ . Tällöin koko aineistoa vastaavan sv:n  $\mathbf{Y}_T$  yhteistiheysfunktio on  $f_{\mathbf{Y}_T}$  ja käyttäen ehdollisen tiheysfunktion kaavaa saadaan

$$f_{\mathbf{Y}_T} = f_{y_T | \mathbf{Y}_{T-1}} \cdot f_{\mathbf{Y}_{T-1}} = f_{y_T | \mathbf{Y}_{T-1}} \cdot f_{y_{T-1} | \mathbf{Y}_{T-2}} \cdot f_{\mathbf{Y}_{T-2}} = \cdots = \prod_{t=1}^T f_{y_t | \mathbf{Y}_{t-1}} \cdot f_{\mathbf{Y}_0}.$$

Jaksossa 2.3 esitetyistä VAR(p)-prosessin lineaarisesta esityksestä (2.11) seuraa riippumattomuus  $y_{t-j} \perp\!\!\!\perp \varepsilon_t$ ,  $j \geq 1$ , joten erityisesti  $\mathbf{Y}_{t-1} \perp\!\!\!\perp \varepsilon_t$ . Tämän ja malliyhtälön (4.1) perusteella on selvää, että sv:n  $y_t$  ehdollinen jakauma ehdolla  $\mathbf{Y}_{t-1}$  riippuu vain  $\mathbf{Y}_{t-1}$ :n  $p$ :stä ensimmäisestä osavektorista  $y_{t-1}, \dots, y_{t-p}$  eli sv:sta  $\mathbf{y}_{t-1}$ . Käyttäen malliyhtälön (4.1) esitystä (4.5) saadaan siten<sup>8</sup>

$$y_t | \mathbf{Y}_{t-1} \sim y_t | \mathbf{y}_{t-1} \sim \mathbf{N}(X'_t \pi, \Omega), \quad t = 1, \dots, T.$$

<sup>8</sup>Tässä voi käyttää seuraavaa multinormaalijakaumaan liittyvää tulosta. Jos  $Y = X + Z$ , jossa  $X \perp\!\!\!\perp Z$  ja  $Z \sim \mathbf{N}(\mu, \Sigma)$ , niin  $Y | (X = x) \sim \mathbf{N}(\mu + x, \Sigma)$  (perustelu jätetään tehtäväksi).



Multinormaalijakauman tiheysfunktion lausekkeesta saadaan näin ollen

$$f_{y_t|\mathbf{Y}_{t-1}}(y_t|\mathbf{y}_{t-1}) = (2\pi)^{-1/2} \det(\Omega)^{-1/2} \exp\left\{-\frac{1}{2}(y_t - X_t'\pi)' \Omega^{-1} (y_t - X_t'\pi)\right\}. \quad (4.6)$$

Koko aineistoa vastaavan sv:n  $\mathbf{Y}_T$  tiheysfunktion  $f_{\mathbf{Y}_T}$  selvittämiseksi on siten vielä selvitettävä alkuarvovektorin  $\mathbf{Y}_0 = \mathbf{y}_0$  jakauma. Stationaarisessa tapauksessa oletuksesta (4.2) seuraa, että  $\mathbf{y}_0$  on (normaalisti jakautuneiden sv:ien lineaarisena muunnoksena) normaalisti jakautunut. Odotusarvo saadaan tuloksesta  $\mathbf{E}(y_t) = A(1)^{-1}\nu$  (ks. jakson 2.3 loppu) ja kovarianssimatriisi on  $\mathbf{\Gamma}_0 = [\Gamma_{i-j}]$  ( $i, j = 1, \dots, p$ ), joka voidaan lausua parametrien  $\nu, A_1, \dots, A_p$  ja  $\Omega$  funktiona (ks. jakso 2.3).

Seuraavassa alkuarvo  $\mathbf{y}_0$  tulkitaan kiinteäksi vakioksi tai yhtäpitävästi ehdollistaa sen saaman arvon suhteen. Tällöin päädytään ns. *ehdolliseen uskottavuusfunktioon*, joka saadaan edellä todetun mukaan tulkitsemalla tulo  $f_{y_1|\mathbf{Y}_0} \cdots f_{y_T|\mathbf{Y}_{T-1}}$  parametrien funktiona. Käyttäen yhtälöä (4.6) ja ottamalla logaritmi päädytään siten ehdolliseen log-uskottavuusfunktioon (parametrin riippumaton vakio jätetään pois)

$$l(\pi, \Omega) = -\frac{T}{2} \log \det(\Omega) - \frac{1}{2} \sum_{t=1}^T (y_t - X_t'\pi)' \Omega^{-1} (y_t - X_t'\pi). \quad (4.7)$$

On selvää, että suurilla havaintomäärillä alkuarvojen  $y_{-p+1}, \dots, y_0$  vaikutus on mitätön ja uskottavuusfunktioon perustuvien tilastollisten menetelmien asympotoottisiin ominaisuuksiin alkuarvoilla ei ole mitään vaikutusta. Ellei toisin mainita, tarkoitetaan uskottavuusfunktiolla seuraavassa ehdollista uskottavuusfunktiota.

**Uskottavuusfunktion maksimointi.** Maksimoidaan nyt uskottavuusfunktio tavantomaista derivointimenettelyä käyttäen. Derivoimalla saadaan yhtälöstä (4.7) (ks. Liite A.7)

$$\frac{\partial}{\partial \pi} l(\pi, \Omega) = \sum_{t=1}^T X_t \Omega^{-1} (y_t - X_t'\pi). \quad (4.8)$$

Matriisin  $X_t$  määritelmän ja Kroneckerin tulon ominaisuuksia käyttäen saadaan (ks. Liite A.5)

$$X_t \Omega^{-1} = (I_n \otimes x_t) (\Omega^{-1} \otimes 1) = \Omega^{-1} \otimes x_t = (\Omega^{-1} \otimes I_{np+1}) (I_n \otimes x_t),$$

jossa siis  $(I_n \otimes x_t) = X_t$ . Yhtälö (4.8) voidaan näin ollen kirjoittaa

$$\frac{\partial}{\partial \pi} l(\pi, \Omega) = (\Omega^{-1} \otimes I_{np+1}) \sum_{t=1}^T X_t (y_t - X_t'\pi). \quad (4.9)$$

Asettamalla  $\partial l(\pi, \Omega) / \partial \pi = 0$  voidaan (epäsingulaarinen) tekijä  $(\Omega^{-1} \otimes I_{np+1})$  supistaa pois, joten parametrin  $\pi$  uskottavuusyhtälöksi saadaan

$$\sum_{t=1}^T X_t X_t' \pi = \sum_{t=1}^T X_t y_t,$$

jonka yksikäsitteinen ratkaisu on

$$\hat{\pi} = \left( \sum_{t=1}^T X_t X_t' \right)^{-1} \sum_{t=1}^T X_t y_t$$

eli malliyhtälöstä (4.5) muodostettu pienimmän neliösumman (PNS) estimaattori. Viimeksi mainittu seikka tulee havainnollisemmaksi, kun yhtälöryhmä (4.5) kirjoitetaan matriisimuodossa

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} X_1' \\ \vdots \\ X_T' \end{bmatrix} \pi + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

tai lyhyesti

$$y = X\pi + \varepsilon.$$

Suoraviivaisella matriisilaskulla nähdään, että  $\hat{\pi}$  saadaan tavanomaisella PNS-kaavalla

$$\hat{\pi} = (X'X)^{-1} X'y.$$

Huomaa kuitenkin, että  $\text{Cov}(\varepsilon) = I_T \otimes \Omega$ , mikä poikkeaa tavanomaisesta lineaarisen mallin yhteydessä tehtävästä oletuksesta. Syy sille, että SU-menetelmä johtaa tästä huolimatta tavanomaiseen PNS-estimointiin piilee selittäjämatrisin  $X$  erityisessä rakenteessa. Myöhemmin nähdään, että vastaava tulos ei päde, kun matriisin  $X$  rakenne muuttuu. Suoralla (tehtäväksi jätettävällä) laskulla nähdään, että estimaattorivektorin  $\hat{\pi}$  osavektorit  $\hat{\pi}_1, \dots, \hat{\pi}_n$  saadaan soveltamalla PNS-menetelmää yhtälöryhmän (4.4) yksittäisiin yhtälöihin.

Tarkastellaan nyt kovarianssimatriisin  $\Omega$  SU-estimointia. Tässä yhteydessä on kätevää johtaa käänteismatriisin  $\Omega^{-1} =: \Phi$  SU-estimaattori, minkä jälkeen  $\Omega$ :n SU-estimaattori saadaan SU-menetelmän invarianssiominaisuuden nojalla kääntämällä  $\Phi$ :n SU-estimaattori. Käyttäen parametrimatriisia  $\Phi$  ja lyhennysmerkintää  $\varepsilon_t(\pi) = y_t - X_t'\pi$  voidaan log-uskottavuusfunktio kirjoittaa

$$l(\pi, \Phi) = \frac{T}{2} \log \det(\Phi) - \frac{1}{2} \sum_{t=1}^T \varepsilon_t(\pi)' \Phi \varepsilon_t(\pi).$$

Liitteen A.7 lopussa esitettyjä derivointikaavoja käyttäen saadaan <sup>9</sup>

$$\frac{\partial}{\partial \Phi} \varepsilon_t(\pi)' \Phi \varepsilon_t(\pi) = \varepsilon_t(\pi) \varepsilon_t(\pi)' \quad \text{ja} \quad \frac{\partial}{\partial \Phi} \log \det(\Phi) = \Phi^{-1} = \Omega,$$

---

<sup>9</sup>Huomaa, että tässä derivointia on yksinkertaistettu jättämällä matriisin  $\Phi$  symmetrisyys (virheellisesti) huomioon ottamatta. Kuten Liitteessä A.7 todetaan, johtaa symmetrisyyden (oikea) huomioon ottaminen monimutkaisempiin derivaattoihin, mutta kuitenkin yhtäpitävään  $\Phi$ :n uskottavuusyhtälöön ja siten samaan SU-estimaattiin. Intuitiivisesti tämä on uskottavaa, koska maksimoinnissa päädytään symmetriseen (ja positiivisesti definiittiin) ratkaisuun, vaikka symmetrisyysehtoa ei otetakaan huomioon.

joten

$$\frac{\partial}{\partial \Phi} l(\pi, \Phi) = \frac{T}{2} \Omega - \frac{1}{2} \sum_{t=1}^T \varepsilon_t(\pi) \varepsilon_t(\pi)'$$

Asettamalla oikea puoli nolllaksi ja korvaamalla vektorissa  $\varepsilon_t(\pi)$  parametri edellä johdetulla  $\pi$ :n uskottavuusyhtälön ratkaisulla  $\hat{\pi}$  saadaan ratkaisu

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T (y_t - X_t' \hat{\pi})(y_t - X_t' \hat{\pi})'$$

Koska  $\hat{\pi}$  minimoi PNS-estimaattorina log-uskottavuusfunktion  $l(\pi, \Omega)$  summalausekkeen (ks. (4.7)), on SU-menetelmän invarianssiominaisuuden nojalla selvää, että  $\hat{\pi}$  ja  $\hat{\Omega}$  ovat parametrien  $\pi$  ja  $\Omega$  (ehdollisia) SU-estimaattoreita.<sup>10</sup>

**SU-estimaattorien asymptoottiset ominaisuudet.** SU-estimaattorien asymptoottisten ominaisuuksien selvittämisessä nojaututaan tavanomaiseen SU-teoriaan, jonka toimiminen perustuu siihen, että tavanomaiset raja-arvolauseet (eli suurten lukujen laki ja keskeinen raja-arvolause) pätevät tarkasteltavan VAR(p)-mallin tapauksessa. Oletetaan alkuarvot  $y_{-p+1}, \dots, y_0$  merkintöjen yksinkertaistamiseksi stationaariksi.

Estimaattorin  $\hat{\pi}$  asymptoottisen jakauman johtamiseksi lasketaan ensin (ks. (4.9) ja Liite A.7)

$$\frac{\partial^2}{\partial \pi \partial \pi'} l(\pi, \Omega) = -(\Omega^{-1} \otimes I_{np+1}) \sum_{t=1}^T X_t X_t' = -\Omega^{-1} \otimes \sum_{t=1}^T x_t x_t',$$

jossa jälkimmäinen yhtälö perustuu matriisiin  $X_t$  esitykseen  $X_t = I_n \otimes x_t$  ja Kroneckerin tulon ominaisuuksiin (ks. Liite A.5). Koska  $x_t = [1 \ \mathbf{y}'_{t-1}]'$ , on parametrin  $\pi$  Fisherin informaatiomatriisi siten

$$\mathcal{I}_{\pi\pi}(\pi, \Omega) := -\mathbf{E} \left[ \frac{\partial^2}{\partial \pi \partial \pi'} l(\pi, \Omega) \right] = T(\Omega^{-1} \otimes \mathbf{\Gamma}_x),$$

jossa

$$\mathbf{\Gamma}_x = \mathbf{E}(x_t x_t') = \begin{bmatrix} 1 & \mathbf{E}(\mathbf{y}'_{t-1}) \\ \mathbf{E}(\mathbf{y}_{t-1}) & \mathbf{E}(\mathbf{y}_{t-1} \mathbf{y}'_{t-1}) \end{bmatrix}.$$

<sup>10</sup>Mainittakoon, että edellä esitettyä menettelyä soveltaen voidaan johtaa SU-estimaattorit myös usean yhtälön regressiomallissa eli mallissa, jonka määrittelee yhtälö (4.3), kun  $X_t$  on kiinteä (tai voidaan tulkita kiinteäksi) ja virhetermi toteuttaa normaalisuusoletuksen (4.2). Erikoistapaus tästä on malli  $Y_1, \dots, Y_T \sim \mathbf{N}(\mu, \Sigma)$  ||. Sama pätee myös, kun VAR(p)-malliin lisätään (vakiotermin lisäksi) kiinteitä selittäviä muuttujia.

Todetaan seuraavaksi, että (ks. (4.8) ja Liite A.7)

$$\begin{aligned} \mathbf{E} \left[ \frac{\partial^2}{\partial \pi \partial \omega_{ij}} l(\pi, \Omega) \right] &= \sum_{t=1}^T \mathbf{E} \left[ X_t \left( \frac{\partial}{\partial \omega_{ij}} \Omega^{-1} \right) (y_t - X_t' \pi) \right] \\ &= \sum_{t=1}^T \mathbf{E} \left[ X_t \left( \frac{\partial}{\partial \omega_{ij}} \Omega^{-1} \right) \right] \mathbf{E}(\varepsilon_t) \\ &= 0. \end{aligned}$$

Tässä toinen yhtälö perustuu yhtälöön  $\varepsilon_t = y_t - X_t' \pi$  ja siihen, että  $\varepsilon_t$  on riippumaton muuttujista  $y_{t-1}, y_{t-2}, \dots$  ja siten muuttujasta  $X_t$  (ks.  $X_t$ :n määritelmä yhtälössä (4.5)).

Edellä todetusta seuraa, että parametrien  $\pi$  ja  $\Omega$  Fisherin informaatiomatriisi on lohkodeagonaalinen eli parametrit  $\pi$  ja  $\Omega$  ovat ortogonaaliset. Yleisestä SU-estimointiteoriasta voidaan siten päätellä, että estimaattorit  $\hat{\pi}$  ja  $\hat{\Omega}$  ovat asymptoottisesti riippumattomat ja estimaattorin  $\hat{\pi}$  asymptoottinen jakauma on sama kuin tapauksessa, jossa parametrin  $\Omega$  arvo tunnetaisiin. Toisin sanoen,  $\hat{\pi}$  on asymptoottisesti normaalin odotusarvona  $\pi$  ja kovarianssimatriisina  $\mathcal{I}_{\pi\pi}(\pi, \Omega)^{-1}$ . Koska  $(\Omega^{-1} \otimes \mathbf{\Gamma}_x)^{-1} = \Omega \otimes \mathbf{\Gamma}_x^{-1}$  saadaan matriisin  $\mathcal{I}_{\pi\pi}(\pi, \Omega)$  määritelmän nojalla siten tulos

$$\hat{\pi}_{as} \sim \mathbf{N}(\pi, T^{-1}(\Omega \otimes \mathbf{\Gamma}_x^{-1})).$$

Estimaattorin  $\hat{\Omega}$  asymptoottista jakaumaa ei tarkastella. Myöhemmin vedotaan kuitenkin tarkentuvuuteen  $\hat{\Omega} \xrightarrow{p} \Omega$ .

Edellä johdettu tulos pätee myös ilman oletusta stationaarisista alkuarvoista, kunhan matriisi  $T(\Omega^{-1} \otimes \mathbf{\Gamma}_x)$  korvataan Fisherin informaatiomatriisin yleisellä lausekkeella  $\mathcal{I}_{\pi\pi}(\pi, \Omega) = \Omega^{-1} \otimes \sum_{t=1}^T \mathbf{E}(x_t x_t')$ . Sama pätee seuraavassa esitettävään rajoitettuun estimointiin ja testimenetelmiin. On kuitenkin syytä huomata, että asymptoottisten jakauma-approksimaatioiden toimivuuteen vaikuttaa se, kuinka voimakasta prosessin korreloituneisuus on. Kun prosessin stabiilisuusehto (ks. jakso 2.3) on ”lähellä rikkoontua”, voivat asymptoottiset jakaumat olla pienillä havaintomäärillä harhaanjohtavia.

## 4.2 Estimointi lineaarisin rajoittein

Tarkastellaan edellisen jakson mallia, jossa parametreihin  $A_1, \dots, A_p$  (ja mahdollisesti myös  $\nu$ ) liitetään lineaarisia rajoitteita. Nämä rajoitteet on kätevä esittää käyttäen yhtälön (4.5) parametrivektoria  $\pi$ , jonka oletetaan toteuttavan

$$\pi = H\delta + a, \tag{4.10}$$

jossa  $H$  on tunnettu  $n(np+1) \times m$  matriisi,  $a$  on tunnettu  $n(np+1)$  vektori ja  $\delta$  on tuntematon  $m \times 1$  parametrivektori. Lisäksi oletetaan, että matriisi  $H$  on täyttä sarakeastetta eli  $r(H) = m$

Usein lineaariset parametrirajoitteet esitetään vaihtoehtoisella tavalla käyttäen yhtälöä

$$R\pi = b, \quad (4.11)$$

jossa  $R$  ( $q \times n(np+1)$ ) ja  $b$  ( $q \times 1$ ) ovat tunnettuja ja  $r(R) = q$ . Lineaarialgebraa käyttäen rajoitteet (4.10) ja (4.11) voidaan todeta yhtäpitäviksi, kun  $q + m = n(np+1)$  (perustelu jätetään tehtäväksi). Edellinen on kätevä tämän jakson estimointitarkasteluissa ja jälkimmäinen seuraavan jakson Waldin testissä.

Tyypillisin esimerkki edellä esitetyistä rajoitteista rajoittaa osan parametrivektorin  $\pi$  komponenteista nollassi. Jos esimerkiksi valitaan  $H = [I_{n^2p} : 0]'$  ja  $a = 0$ , asetetaan  $\pi$ :n  $n$  viimeistä komponenttia nollassi eli oletetaan, että  $y_{t-p}$  ei selitä  $y_t$ :n viimeistä komponenttia  $y_{nt}$ . Tällöin selvästikin  $R = [0 : I_n]$  ja  $b = 0$ .

Rajoitetun mallin parametrien  $\delta$  ja  $\Omega$  log-uskottavuusfunktio saadaan yksinkertaisesti suorittamalla rajoittamattoman mallin log-uskottavuusfunktiossa (4.7) sijoitus  $\pi \rightarrow H\delta + a$ . Merkitsemällä

$$z_t = y_t - X_t' a \quad \text{ja} \quad W_t = H' X_t$$

saadaan log-uskottavuusfunktioksi

$$l(\delta, \Omega) = -\frac{T}{2} \log \det(\Omega) - \frac{1}{2} \sum_{t=1}^T (z_t - W_t' \delta)' \Omega^{-1} (z_t - W_t' \delta).$$

Uskottavuusyhtälöiden tutkimiseksi derivoidaan parametrien suhteen. Kuten derivaatan (4.8) tapauksessa saadaan

$$\frac{\partial}{\partial \delta} l(\delta, \Omega) = \sum_{t=1}^T W_t \Omega^{-1} (z_t - W_t' \delta). \quad (4.12)$$

Toisin kuin parametrin  $\pi$  tapauksessa (ks. yhtälöt (4.8) ja (4.9)) ei matriisia  $\Omega^{-1}$  voida nyt kuitenkaan ”siirtää” summan eteen. Tämä merkitsee, ettei yhtälöä  $\partial l(\delta, \Omega) / \partial \delta = 0$  voida ratkaista tuntematta matriisia  $\Omega$ , vaan ratkaisu jää (yleensä) muotoon

$$\delta = \left( \sum_{t=1}^T W_t \Omega^{-1} W_t' \right)^{-1} \sum_{t=1}^T W_t \Omega^{-1} z_t. \quad (4.13)$$

Merkitsemällä jälleen  $\Omega^{-1} = \Phi$  voidaan edellisen jakson laskelmat toistaa ja todeta, että  $\Phi$ :n uskottavuusyhtälö  $\partial l(\delta, \Phi) / \partial \Phi = 0$  johtaa yhtälöön

$$\Omega = \frac{1}{T} \sum_{t=1}^T (z_t - W_t' \delta) (z_t - W_t' \delta)'. \quad (4.14)$$

Uskottavuusyhtälöiden ratkaiseminen on yhtäpitävää yhtälöiden (4.13) ja (4.14) ratkaisemisen kanssa. Koska ratkaisu  $\delta$ :n suhteen vaatii  $\Omega$ :n arvon tuntemisen ja

ratkaisu  $\Omega$ :n suhteen vaatii  $\delta$ :n arvon tuntemisen (erikoistapauksia lukuun ottamatta), vaatii ratkaisu (yleensä) numeeristen menetelmien käyttöä. Havainnollinen tapa ratkaista uskottavuusyhtälöt etenee seuraavasti. Korvataan tuntematon  $\Omega$  yhtälön (4.13) oikealla puolella edellisen jakson rajoittamattomalla estimaatilla  $\hat{\Omega}$ , jolloin saadaan

$$\tilde{\delta}^{(1)} = \left( \sum_{t=1}^T W_t \hat{\Omega}^{-1} W_t' \right)^{-1} \sum_{t=1}^T W_t \hat{\Omega}^{-1} z_t.$$

Tämä voidaan seuraavaksi sijoittaa yhtälön (4.14) oikealle puolelle, jolloin saadaan uusi  $\Omega$ :n estimaatti  $\tilde{\Omega}^{(1)}$ , joka voidaan sijoittaa estimaatin  $\tilde{\delta}^{(1)}$ :n yhtälöön  $\hat{\Omega}$ :n paikalle. Näin saadaan uusi  $\delta$ :n estimaatti  $\tilde{\delta}^{(2)}$ . Tätä menettelyä voidaan jatkaa, kunnes estimaatit eivät enää muutu. Tämä menettely esitetään lähinnä sen havainnollisuuden vuoksi. Käytännössä tietokoneohjelmat suorittavat ratkaisun yleensä jotain muuta algoritmia käyttäen. Mainittakoon kuitenkin, että jo yksinkertaisella yhden askeleen estimaattorilla  $\tilde{\delta}^{(1)}$  on hyvät asymptoottiset ominaisuudet, sillä sen voidaan osoittaa olevan asymptoottisesti yhtäpitävä SU-estimaattorin kanssa (selitys piilee parametrien  $\pi$  ja  $\Omega$  ortogonaalisuudessa).

Merkitään parametrien  $\delta$  ja  $\Omega$  SU-estimaattoreita symboleilla  $\tilde{\delta}$  ja  $\tilde{\Omega}$  ja johdetaan edellisen asymptoottinen jakauma vedoten SU-estimoinnin yleiseen teoriaan. Derivoimalla log-uskottavuusfunktiota toisen kerran  $\delta$ :n suhteen saadaan (ks. (4.12) ja Liite A.7)

$$\frac{\partial^2 \delta}{\partial \delta \partial \delta'} l(\delta, \Omega) = - \sum_{t=1}^T W_t \Omega^{-1} W_t',$$

joten parametrin  $\delta$  Fisherin informaatiomatriisi on

$$\mathcal{I}_{\delta\delta}(\delta, \Omega) = \sum_{t=1}^T \mathbf{E}(W_t \Omega^{-1} W_t'),$$

joka stationaarisessa tapauksessa on yhtä kuin  $T \mathbf{E}(W_t \Omega^{-1} W_t')$ . Samanlainen laskelma kuin edellisessä jaksossa osoittaa, että  $\mathbf{E}[\partial^2 l(\delta, \Omega) / \partial \delta \partial \omega_{ij}] = 0$  eli parametrit  $\delta$  ja  $\Omega$  ovat ortogonaaliset. Estimaattorin  $\tilde{\delta}$  asymptoottiseksi jakaumaksi saadaan siten

$$\tilde{\delta} \underset{as}{\sim} \mathbf{N}(\delta, \mathcal{I}_{\delta\delta}(\delta, \Omega)^{-1}).$$

Fisherin informaatiomatriisissa voidaan ottaa huomioon määritelmä  $W_t = H' X_t$ , jossa  $H$  on tunnettu ja kiinteä, mutta kovin havainnollista lauseketta ei silti saada.

Kovarianssimatriisin  $\Omega$  rajoitetun SU-estimaattorin  $\tilde{\Omega}$  asymptoottisista ominaisuuksista mainitaan vain tarkentuvuus, jota käytetään seuraavassa jaksossa. Edellä sanotusta seuraa, että estimaattori  $\tilde{\Omega}$  voidaan kirjoittaa

$$\tilde{\Omega} = \frac{1}{T} \sum_{t=1}^T (z_t - W_t' \tilde{\delta}) (z_t - W_t' \tilde{\delta})'$$

### 4.3 Hypoteesien testaus

Johdetaan Waldin testi lineaariselle hypoteesille (ks. (4.11))

$$H_0 : R\pi = b,$$

jonka voimassaolo seuraavassa oletetaan. Lisäksi oletetaan, että SU-estimaattoreilla  $\hat{\pi}$  ja  $\hat{\Omega}$  on jaksossa 4.1 esitetyt asymptoottiset ominaisuudet.

SU-estimaattorin  $\hat{\pi}$  asymptoottinen normaalisuus on formaalisti esitettynä (oletetaan merkintöjen yksinkertaistamiseksi stationaariset alkuarvot)

$$\sqrt{T}(\hat{\pi} - \pi) \xrightarrow{d} \mathbf{N}(0, \Omega \otimes \Gamma_x^{-1}).$$

Koska  $R\pi = b$ , seuraa tästä (ks. Liite B, Lause B.2)

$$\sqrt{T}(R\hat{\pi} - b) = R\sqrt{T}(\hat{\pi} - \pi) \xrightarrow{d} Z \sim \mathbf{N}_q(0, R(\Omega \otimes \Gamma_x^{-1})R')$$

ja edelleen (ks. Liite B, Lause B.3)

$$T(R\hat{\pi} - b)' [R(\Omega \otimes \Gamma_x^{-1})R']^{-1} (R\hat{\pi} - b) \xrightarrow{d} Z' [R(\Omega \otimes \Gamma_x^{-1})R']^{-1} Z \sim \chi_q^2,$$

jossa viimeksi mainittu tulos seuraa tunnetusta multinormaalijakaumaa ja  $\chi^2$ -jakaumaa koskevasta tuloksesta.<sup>11</sup> Vasemmalla oleva neliömuoto mittaa selväsikin hypoteesin  $R\pi = b$  realistisuutta, mutta sisältää tuntemattoman matriisin  $\Omega \otimes \Gamma_x^{-1}$ , joka on testisuureessa korvattava (tarkentuvalla) estimaattorilla. Koska  $\Gamma_x = \mathbf{E}(x_t x_t')$ , jossa  $x_t = [1 \ \mathbf{y}'_t]'$  ja  $\mathbf{y}'_t = [y'_{t-1} \ \cdots \ y'_{t-p+1}]'$ , ja koska jaksossa 3 todetun perusteella (stationaarisen) VAR(p)-prosessin tapauksessa ensimmäiset ja toiset otosmomentit ovat vastaavien teoreettisten momenttien tarkentuvia estimaattoreita, saadaan tämän ja estimaattorin  $\hat{\Omega}$  tarkentuvuuden nojalla

$$\left[ R(\hat{\Omega} \otimes \hat{\Gamma}_x^{-1})R' \right]^{-1} \xrightarrow{p} \left[ R(\Omega \otimes \Gamma_x^{-1})R' \right]^{-1},$$

jossa

$$\hat{\Gamma}_x = T^{-1} \sum_{t=1}^T x_t x_t'.$$

Yhteenvedonä edellä esitetystä saadaan nyt *Waldin testisuure* (ks. Liite B, Lause B.3)

$$W = T(R\hat{\pi} - b)' \left[ R(\hat{\Omega} \otimes \hat{\Gamma}_x^{-1})R' \right]^{-1} (R\hat{\pi} - b) \xrightarrow{d} \chi_q^2.$$

<sup>11</sup>Edellä (kuten tarvittaessa myöhemminkin) käytetään merkintää  $\mathbf{N}_q(\cdot, \cdot)$  havainnollistamaan multinormaalijakauman dimensiota. Mainitun multinormaalijakaumaa ja  $\chi^2$ -jakaumaa koskevan tuloksen mukaan  $Z \sim \mathbf{N}_k(\mu, \Sigma) \Rightarrow (Z - \mu)' \Sigma^{-1} (Z - \mu) \sim \chi_k^2$  (olettaen, että  $\Sigma$  on positiivisesti definiitti).

Suuret testisuureen arvot ovat tietenkin kriittisiä. Käytännössä testiä sovelletaan laskemalla approksimatiivinen P-arvo

$$P = P_{H_0} \{W \geq W(\mathbf{y})\} \approx P \{\chi_q^2 \geq W(\mathbf{y})\},$$

jossa  $\chi_q^2$  on  $\chi_q^2$ -jakaumaa noudattava satunnaismuuttuja ja  $W(\mathbf{y})$  on (satunnaisen) testisuureen  $W$  aineistosta laskettu arvo.

Waldin testin asemasta voidaan käyttää *uskottavuusosamäärätestiä*. Testisuureen yleinen lauseke on

$$LR = 2 \left[ l(\hat{\pi}, \hat{\Omega}) - l(\tilde{\pi}, \tilde{\Omega}) \right],$$

jossa  $\tilde{\pi}$  ja  $\tilde{\Omega}$  ovat parametrien  $\pi$  ja  $\Omega$  jaksossa 4.2 johdetut rajoitetut SU-estimaattorit. Käyttäen hyväksi estimaattoreiden  $\hat{\Omega}$  ja  $\tilde{\Omega}$  sekä log-uskottavuusfunktion lausekkeita voidaan suoralla laskulla todeta, että (perustelu jätetään tehtäväksi)

$$l(\hat{\pi}, \hat{\Omega}) = -\frac{T}{2} \log \det(\hat{\Omega}) - \frac{Tn}{2} \quad \text{ja} \quad l(\tilde{\pi}, \tilde{\Omega}) = -\frac{T}{2} \log \det(\tilde{\Omega}) - \frac{Tn}{2}.$$

Uskottavuusosamäärätesti voidaan siten perustaa testisuureeseen

$$LR = T \left[ \log \det(\tilde{\Omega}) - \log \det(\hat{\Omega}) \right] \xrightarrow{d} \chi_q^2,$$

jossa asymptoottinen jakauma voidaan perustella samoilla ehdoilla kuin Waldin testin tapauksessa. Suuret testisuureen arvot ovat jälleen kriittisiä, joten approksimatiiviset P-arvot lasketaan kuten Waldin testissä.

Edellä esitetyt testit voidaan yleistää koskemaan myös jaksossa 4.2 tarkastellun rajoitetun mallin parametrivektoria  $\delta$ . Rajoitetun SU-estimaattorin asymptoottinen normaalisuus on formaalisti esitettyä (ks. jakson 4.2 loppu)

$$\sqrt{T}(\tilde{\delta} - \delta) \underset{as}{\sim} \mathbf{N} \left( 0, \left( \sum_{t=1}^T \mathbf{E} (W_t \Omega^{-1} W_t') \right)^{-1} \right),$$

joten Waldin testiä varten tarvitaan asymptoottisesti pätevä empiirinen vastine matriisille  $\sum_{t=1}^T \mathbf{E} (W_t \Omega^{-1} W_t')$ . Ilmeinen vaihtoehto on  $T^{-1} \sum_{t=1}^T W_t \tilde{\Omega}^{-1} W_t'$ , jota käyttäen Waldin testi voidaan johtaa samaan tapaan kuin edellä rajoittamattoman SU-estimaattorin tapauksessa. Matriisin  $\left( \sum_{t=1}^T W_t \tilde{\Omega}^{-1} W_t' \right)^{-1}$  diagonaalialkioiden neliöjuurista saadaan approksimatiiviset keskivirheet estimaattorin  $\tilde{\delta}$  komponenteille. Jos  $i$ :n:n komponentin keskivirheestä käytetään merkintää  $\text{s.e.}(\tilde{\delta}_i)$ , pätee  $\tilde{\delta}_i / \text{s.e.}(\tilde{\delta}_i) \underset{as}{\sim} \mathbf{N}(\delta_i, 1)$ , joten parametrin  $\delta_i$  approksimatiivinen 95%:n luottamusväli on  $\tilde{\delta}_i \pm 1.96 \text{s.e.}(\tilde{\delta}_i)$ . Vastaavalla tavalla voidaan muodostaa approksimatiivisia luottamusvälejä rajoittamattoman mallin parametrivektorin  $\pi$  komponenteille.



#### 4.4 Mallinvalinta ja valitun mallin sopivuuden tutkiminen

**Testien käyttö mallinvalinnassa.** Usein VAR-mallin rakentaminen aloitetaan rajoittamattomasta mallista, koska tarpeeksi vahvaa etukäteistietoa parametrirajoitteista ei ole tai mielessä olevia rajoitteita halutaan testata käyttäen Waldin testiä tai uskottavuusosamäärätestiä. Tällöin ensimmäinen tehtävä on valita mallin (yleensä) tuntematon aste. Jos asteeksi on valittu tavalla tai toisella  $p$ , voidaan valinnan sopivuutta tutkia testaamalla valittua VAR( $p$ )-mallia jotain korkeampaa astetta olevaa vaihtoehtoa vastaan. Jos vaihtoehtona on VAR( $p+s$ )-malli

$$y_t = \nu + \sum_{j=1}^{p+s} A_j y_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim \text{nid}(0, \Omega),$$

testataan (lineaarista) hypoteesia

$$H_0 : A_{p+1} = \dots = A_{p+s} = 0.$$

Jos  $\hat{\Omega}(k) = T^{-1} \sum_{t=1}^T (y_t - X_t' \hat{\pi})(y_t - X_t' \hat{\pi})'$  on rajoittamattoman VAR( $k$ )-mallin virhekovarianssimatriisin  $\Omega$  SU-estimaattori, voidaan käyttää uskottavuusosamäärätestiä ja testisuureta

$$\text{LR} = T \left[ \log \det(\hat{\Omega}(p)) - \log \det(\hat{\Omega}(p+s)) \right] \underset{as}{\sim} \chi_{n^2 s}^2.$$

Huomaa, että tässä (samoin kuin vastaavan tyyppisissä testisuureissa alempana) estimaattorit  $\hat{\Omega}(p)$  ja  $\hat{\Omega}(p+s)$  perustuvat samaan määrään havaintoja, joten (ehdolisessä) SU-estimoinnissa  $y_{-p-s+1}, \dots, y_0$  toimivat alkuarvoina.

Usein testaus suoritetaan peräkkäin aloittaen jostain ”tarpeeksi suuresta” asteesta  $P$ . Jos hypoteesi  $A_P = 0$  jää voimaan, testataan hypoteesia  $A_{P-1} = 0$ . Jos tämäkin hypoteesi jää voimaan, jatketaan testausta, kunnes saadaan ensimmäinen hylkäävä testituloks tai päädytään asteeseen 0. Tämä menetelmä vastaa yksiulotteisesta tapauksesta tutun osittaisautokorrelaatiofunktion käyttämistä AR-mallin asteen valinnassa. Jos ensimmäinen hylkäys tapahtuu suurelta tuntuvan asteen kohdalla, voidaan testausta silti jatkaa samaan tapaan kuin osittaisautokorrelaatiofunktioita käytettäessä välttämättä tiukkaa formaalista testausta. Jonon  $i$ . testisuure on

$$\text{LR}(P-i+1) = T \left[ \log \det(\hat{\Omega}(P-i)) - \log \det(\hat{\Omega}(P-i+1)) \right] \quad (1 \leq i \leq P),$$

jonka saamia arvoja verrataan  $\chi_{n^2}^2$ -jakauman prosenttipisteisiin. On syytä huomata, että yksittäisistä testeistä muodostetun yhteistestin merkitsevyytaso (vastaavasti  $P$ -arvo) riippuu kaikkien suoritettujen testien merkitsevyytasoista (vastaavasti  $P$ -arvoista). Voidaan osoittaa, että jos  $k$  ensimmäistä hypoteesia on voimassa, ovat testisuureet  $\text{LR}(1), \dots, \text{LR}(k)$  asymptoottisesti riippumattomia ja (approksimatiivinen) todennäköisyys hylätä jokin niistä saadaan kaavasta

$$\alpha_k = 1 - (1 - \gamma_1) \cdots (1 - \gamma_k) \quad (1 \leq k \leq P),$$

jossa  $\gamma_i$  on  $i$ :nnessä testissä käytetty merkitsevyystaso. Jos  $\gamma_i = 0.05$  kaikilla  $i$ , on  $\alpha_1 = 0.05$ ,  $\alpha_2 = 0.0975$ ,  $\alpha_3 = 0.143$  ja  $\alpha_4 = 0.185$ .

Kun rajoittamattoman mallin aste on valittu, voidaan Waldin testiä tai uskottavuusosamäärätestiä soveltaa rajoitetun mallin yksityiskohtaisemmassa täsmennyksessä. Tämä on luontevaa, kun esimerkiksi taustateoriaa käyttäen voidaan muotoilla kiinnostavia hypoteeseja. Jos näin ei ole ja verrattavia mallivaihtoehtoja on useita, voivat alempana tarkasteltavat mallinvalintakriteerit olla luontevampia vaihtoehtoja.

**Esimerkki 3.1** (jatkoa). Sovelletaan edellä kuvattua peräkkäistä asteen valintamenetelyä Kuvion 1.1 valuuttakurssi/korkoaineistoon. Valittaessa maksimiasteeksi  $P = 6$  saadaan seuraavat tulokset.

Testisuure :	LR (2)	LR (3)	LR (4)	LR (5)	LR (6)
p-arvo :	0.003	0.678	0.359	0.434	0.866

Tulokset viittaavat (olettaen käytetyn maksimiasteen riittävyys) selvästi asteeseen  $p = 2$ , jota käytettäessä saadaan estimoiduksi malliksi

$$\begin{aligned} \begin{bmatrix} dexch_t \\ rdiff_t \end{bmatrix} &= \begin{bmatrix} -0.002 \\ (0.002) \\ 0.004 \\ (0.014) \end{bmatrix} + \begin{bmatrix} 0.418 & 0.023 \\ (0.089) & (0.012) \\ -0.187 & 1.144 \\ (0.672) & (0.090) \end{bmatrix} \begin{bmatrix} dexch_{t-1} \\ rdiff_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} -0.178 & -0.026 \\ (0.087) & (0.012) \\ 0.603 & -0.179 \\ (0.659) & (0.089) \end{bmatrix} \begin{bmatrix} dexch_{t-2} \\ rdiff_{t-2} \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{bmatrix}. \end{aligned} \quad (4.15)$$

Estimaattien alla suluisissa on (approksimatiiviset) keskivirheet. Residuaalien välinen korrelaatiokerroin on  $-0.259$ , joten mallin virhetermit ovat lievästi korreloituneita. Estimoitu malli toteuttaa stationaarisuusehdon (2.9), sillä matriisin  $\mathbf{A}$  SU-estimaatin suurin ominaisarvo on itseisarvoltaan  $0.953$ . Jotkut estimaatit ovat pieniä keskivirheisiinsä verrattuna, joten mallin rajoittaminen edelleen näyttää mahdolliselta. Tähän kysymykseen palataan myöhemmin.

**Mallivalintakriteerit.** Rajoittamattoman mallin asteen valinnassa voidaan käyttää myös *mallinvalintakriteerejä*, joita on johdettu useita periaatteita käyttäen. Olkoon jälleen  $P$  ”tarpeeksi suuri” ennalta valittu maksimiaste ja  $\hat{\Omega}(k)$  on kuten edellä. Tällöin eräs yleinen kriteerityyppi on

$$C(k) = \log \det(\hat{\Omega}(k)) + \frac{f(T)}{T}d, \quad k = 1, \dots, P, \quad (4.16)$$

jossa  $k$  on mallin aste,  $d$  vastaava vapaasti estimoitujen autoregressiivisten parametrien ja vakiotermin lukumäärä (rajoittamattoman mallin tapauksessa  $d = n(nk + 1)$ )

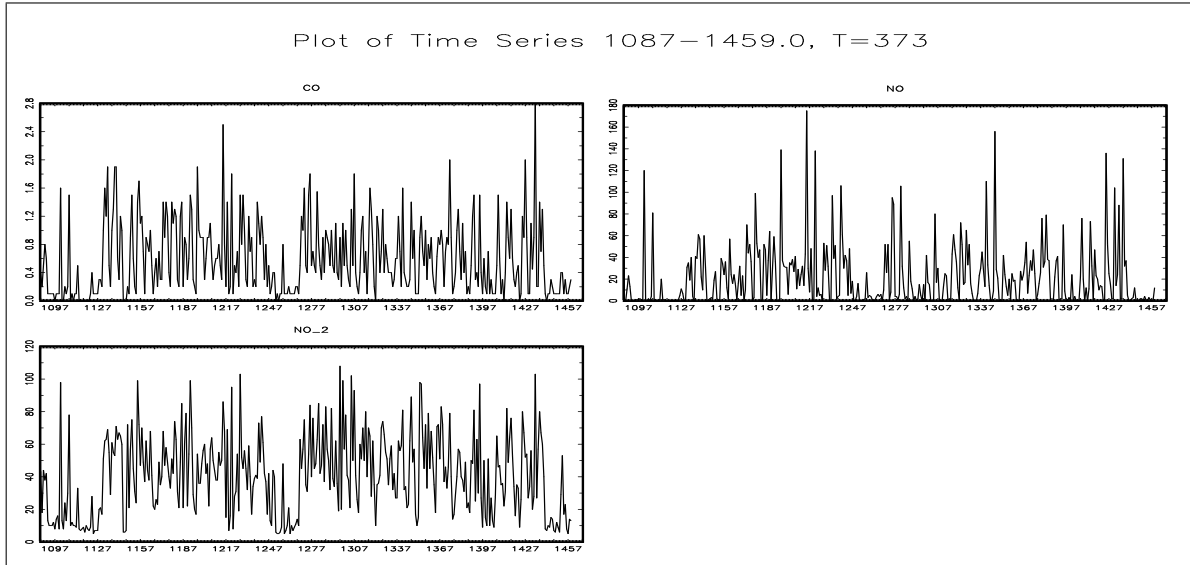
ja ns. *sakkofunktio*  $f$  on positiivinen ja toteuttaa  $f(T)/T \rightarrow 0$ , kun  $T \rightarrow \infty$ . Sakkofunktion idea on rankaista tarpeettoman laajan mallin käyttämisestä. Kriteerifunktion ensimmäinen termi mittaa sovituksen hyvyttä ja, jos asteen  $k$  kasvattaminen ei pienennä sitä tarpeeksi, ei laajempaa mallia suositeta. Tämä tulee ilmeiseksi, kun huomataan, että kriteerifunktion arvoja verrattaessa ensimmäisen termin paikalla voisi olla  $-n - (2/T) \times \log$ -uskottavuusfunktion maksimi-arvo (ks. jakso 4.3). ”Hyvillä” malleilla  $C(k)$  saa siten pienen arvon. Tunnettuja sakkofunktioita ovat

$$\begin{aligned} \text{AIC: } f(T) &= 2 && (\text{Akaike, 1974}) \\ \text{HQ: } f(T) &= 2 \log(\log T) && (\text{Hannan ja Quinn, 1979}) \\ \text{BIC: } f(T) &= \log T && (\text{Schwarz, 1978, Rissanen, 1978}). \end{aligned}$$

Näistä ensimmäinen sakottaa vähiten (suosii laajempia malleja) ja viimeinen eniten (suosii suppeampia malleja). Kriteerifunktiosta HQ on myös versioita, joissa vakion 2 paikalla on joku muu vakio. Näistä kolmesta kriteerifunktiosta kaksi viimeksi mainittua tuottaa asteen tarkentuvan estimaattorin, kun valinta perustetaan kriteerifunktion minimointiin. Käytännössä mallinvalintakriteerejä on kuitenkin suositeltavaa käyttää vain yhtenä mallinvalinnan apuvälineenä eikä valita mallia minimoimalla kriteerifunktiota mekaanisesti. Kun Esimerkin 3.1 tapauksessa maksimiasteeksi valitaan  $P = 6$ , saadaan BIC:tä käyttäen ”parhaaksi” asteeksi  $p = 1$ , kun taas AIC ja HQ suosittelivat valintaa  $p = 2$ .

Kuten edellä mainittiin, voidaan mallinvalintakriteerejä käyttää myös erilaisten rajoitettujen mallien vertailuun. Tällöin yhtälön (4.16) oikealla puolella SU-estimaattori  $\hat{\Omega}(k)$  muodostetaan rajoitetusta VAR(k)-mallista ja  $d$  on rajoitetun mallin vapaasti estimoitujen AR-parametrien ja vakio-termien lukumäärä (jakson 4.2 VAR(p)-mallissa  $d = \dim(\delta)$ ). Tätä menettelyä käytetään erityisesti, kun hypoteesien muotoilu ei ole selkeää ja vertailtavia mallivaihtoehtoja on useita. Yksi vaihtoehto on estimoida kaikki mahdolliset mallikombinaatiot ja perustaa valinta saatuihin kriteerifunktion arvoihin kuten rajoittamattoman mallin tapauksessa. Koska tämä voi johtaa sangen raskaisiin laskelmiin, on esitetty erilaisia askeltavia vaihtoehtoja. Nollarajoitteiden tapauksessa yksi mahdollisuus on poistaa lähtökohtana olevasta mallista vuorotellen yksi selittävä muuttuja (eli viipymän  $y_{t-i}$ ,  $i = 1, \dots, p$  komponentti) ja laskea näin saaduille malleille käytettävän kriteerifunktion arvot. Seuraavassa vaiheessa valitaan lähtökohdaksi suppeammista malleista se, jota vastaava kriteerifunktion arvo on pienentynyt eniten lähtökohtana olevaan malliin verrattuna. Näin jatketaan poistaen aina yksi selittävästä muuttujista kerrallaan, kunnes kriteerifunktion arvo ei pienene edellisen vaiheen malliin verrattuna. Seuraavassa esimerkki tästä menettelystä.

**Esimerkki 4.1.** Kuviossa 4.1 esitetyt kolme aikasarjaa kuvaavat Heidelbergissä heinä- ja elokuussa 1991 tehtyjä päivittäisiä mittauksia hiilimonoksidin (CO), typpioksidin (NO) ja typpidioksidin (NO<sub>2</sub>) määristä (mg/m<sup>3</sup>). Kultakin päivältä on kuusi tasavälistä mittausta ja havaintoja on kaikkiaan 373. Hiilimonoksidin ja typ-



**Kuvio 4.1.** Hiilimonoksidin (CO), typpioksidin (NO) ja typpidioksidin (NO<sub>2</sub>) määrät (mg/m<sup>3</sup>) Heidelbergissä heinä- ja elokuussa 1991. Kuusi tasavälistä mittausta päivittäin.

pioksidin määrään vaikuttavat liikenteen, teollisuuden ja energiatuotannon päästöt, jotka osaltaan kasvattavat typpidioksidin määrää.

On odotettavaa, että peräkkäisten päivien välillä on riippuvuutta, joten aivan pieniasteinen VAR-malli ei välttämättä toimi. Rajoittamattoman mallin astetta valittaessa käytettiin mallinvalintakriteerejä maksimiasteena 15 (kaksi ja puoli päivää). AIC suositteli astetta 8, HQ astetta 6 ja BIC astetta 2. Valitaan asteeksi AIC:n suosituksen mukaan  $p = 8$  ja käytetään edellä kuvattua askeltavaa menettelyä sovellettaessa ankarampaa HQ-kriteeriä. Tulokseksi saadaan parametrimatriiseille  $A_1, \dots, A_8$  seuraava rakenne. Estimaattien asemesta ilmoitetaan ”t-suhteet” eli estimaatit jaettuna approksimatiivisilla keskivirheillään. Lihavoidut nollat osoittavat parametreit, jotka menettely ehdottaa rajoitettavaksi nolaksi.

$$\tilde{A}_1 = \begin{bmatrix} 5.39 & \mathbf{0} & \mathbf{0} \\ -2.58 & 6.10 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 5.36 \end{bmatrix} \quad \tilde{A}_2 = \begin{bmatrix} -2.67 & 2.76 & 3.64 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 5.07 \end{bmatrix}$$

$$\tilde{A}_3 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 2.26 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \tilde{A}_4 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 3.17 & \mathbf{0} & -2.29 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\tilde{A}_5 = \begin{bmatrix} 4.02 & -3.95 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -1.99 & \mathbf{0} \end{bmatrix} \quad \tilde{A}_6 = \begin{bmatrix} \mathbf{0} & 5.14 & 2.49 \\ \mathbf{0} & 9.45 & \mathbf{0} \\ -1.90 & 4.00 & 3.86 \end{bmatrix}$$

$$\tilde{A}_7 = \begin{bmatrix} 2.34 & -2.30 & -2.00 \\ \mathbf{0} & -3.10 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \tilde{A}_8 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2.18 \\ 4.14 & -4.03 & \mathbf{0} \end{bmatrix}$$

Estimoinnissa käytettiin myös vakiotermiä, jonka kaikkien komponenttien estimaatit olivat ”selvästi merkitseviä”. Malli toteuttaa stationaarisuusehdon (2.9), sillä matriisin  $\mathbf{A}$  SU-estimaatin suurin ominaisarvo on itseisarvoltaan 0.893. Virheiden välillä ilmenee selvää korreloituneisuutta, sillä kovarianssimatriisiestimaatista  $\tilde{\Omega}$  muodostetut korrelaatiokertoimet vaihtelevat 0.6:n ja 0.8:n välillä. Estimaateista  $\tilde{A}_1, \dots, \tilde{A}_8$  on hankala tehdä yksityiskohtaisia päätelmiä. Nollat niiden viimeisen rivin kahdella ensimmäisellä sarakkeella osoittavat kuitenkin, että hiilimonoksidin ja typpioksidin vaikutus typpidioksidiin ilmenee vasta viipymiltä 5 ja 6 alkaen. Muut vaikutussuhteet ilmenevät sen sijaan jo viipymilla 1 ja 2.

**Valitun mallin sopivuuden tutkiminen.** Edellä esitettyjen menetelmien lisäksi mallin sopivuutta voidaan tutkia myös käyttäen residuaaleja. Rajoittamattoman mallin tapauksessa residuaalien määritelmä on

$$\hat{\varepsilon}_t = y_t - \hat{\nu} - \hat{A}_1 y_{t-1} - \dots - \hat{A}_p y_{t-p}, \quad t = 1, \dots, T.$$

Residuaalien tulisi muistuttaa ominaisuuksiltaan teoreettisia virheitä  $\varepsilon_t$ . Tämän tarkistamiseksi kannattaa piirtää residuaalisarjojen kuvat ja tutkia löytyykö poikkeavia residuaaleja tai residuaaliryhmiä, varianssin vaihtelua tai muita selviä systemaattisia piirteitä. Myös residuaalien normaalisuutta on hyvä tutkia, vaikka mallin asymptoottinen estimointi- ja testiteoria ei normaalisuutta vaadikaan (iid-oletus riittää).

Residuaalien auto- ja ristikorreloituneisuutta voidaan tutkia laskemalla residuaalisarjoista jaksossa 3 esitetyt auto- ja ristikorrelaatiofunktiot. Jos teoreettisille virheille pätee  $\varepsilon_t \sim \text{iid}(0, \Omega)$ , ovat auto- ja ristikorrelaatioestimaattorit asymptoottisesti normaalisti jakautuneita odotusarvona nolla. Toisin kuin teoreettisten virheiden tapauksessa ne eivät kuitenkaan ole asymptoottisesti riippumattomia eikä jaksossa 3 esitetty  $N(0, 1/T)$ -jakauma-approksimaatio ole pätevä. Oikea asymptoottinen jakauma tunnetaan, mutta sitä ei tarkastella tällä kurssilla. Jotkut tietokoneohjelmat (esim. JMulti) tulostavat residuaalien auto- ja ristikorrelaatioestimaattorien oikeat asymptoottiset keskivirheet tai niihin perustuvat kriittiset rajat, joiden avulla residuaalien mahdollista korreloituneisuutta voidaan arvioida (ks. seuraavaan esimerkkiin liittyvää Kuviota 4.2).

Residuaalien mahdollisen auto- ja ristikorreloituneisuuden tutkimisessa voidaan käyttää apuna myös tilastollisia testejä. Ns. *portmanteau-testi* perustuu suoraan

residuaalien auto- ja ristikorrelaatiofunktioihin. Eräs versio testisuureesta on

$$Q_K^* = T^2 \sum_{k=1}^K \frac{1}{T-k} \text{tr} (\mathcal{S}'_k \mathcal{S}_0^{-1} \mathcal{S}_k \mathcal{S}_0^{-1}), \quad \mathcal{S}_k = \sum_{t=1}^{T-k} \hat{\varepsilon}_t \hat{\varepsilon}'_{t+k}.$$

Jos valitun rajoittamattoman VAR(p)-mallin virheille pätee  $\varepsilon_t \sim \text{iid}(0, \Omega)$ , noudattaa testisuure  $Q_K^*$  asympotoottisesti  $\chi^2$ -jakaumaa vapausastein  $n^2(K-p)$ . Jos kysymyksessä on rajoitettu malli, tulee  $\chi^2$ -jakauman vapausasteluvuksi  $n^2K$  miinus vapaasti estimoitujen autoregressiivisten parametrien lukumäärä. Asympotoottinen tulos edellyttää lisäksi, että  $K$  on ”suuri”.

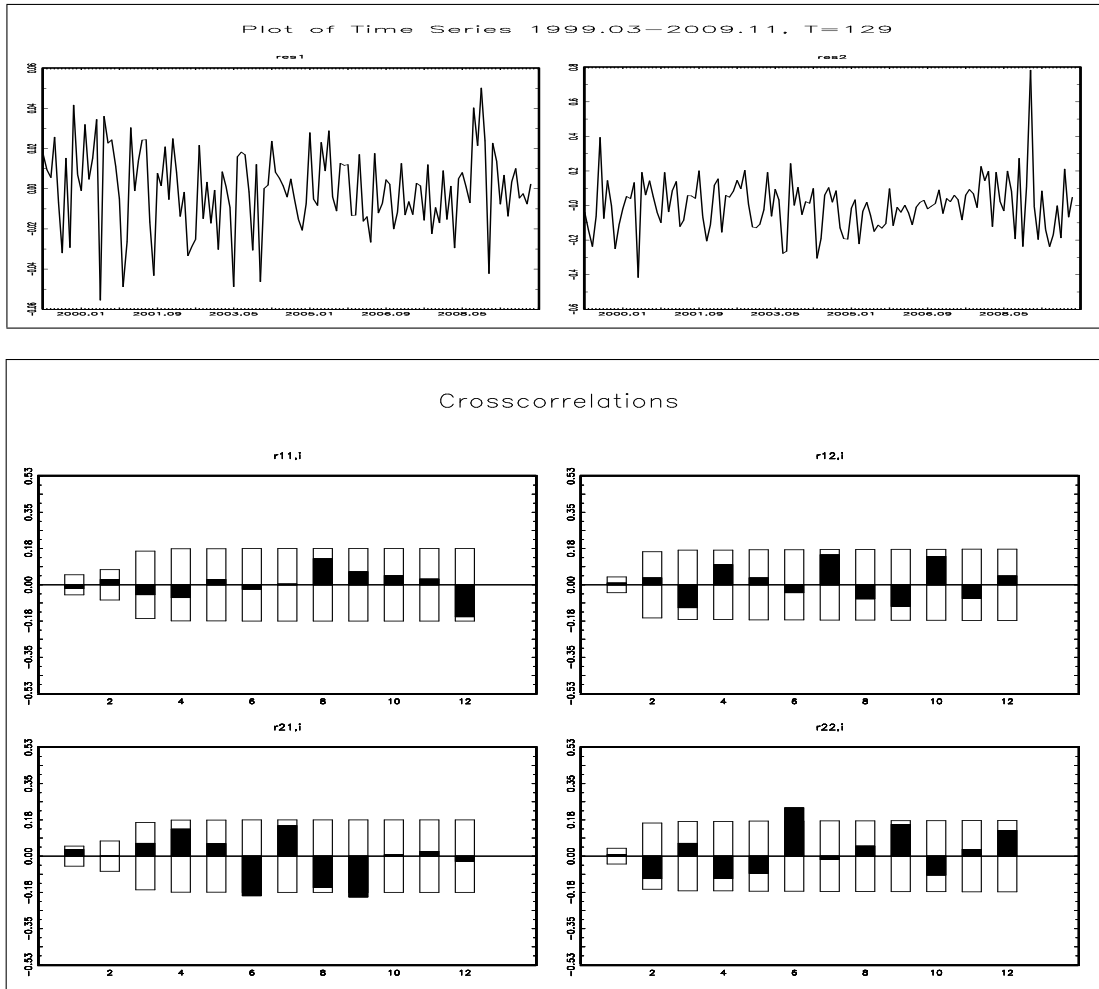
Vaikka residuaalien korreloitunaisuutta ei olisikaan syytä epäillä, eivät residuaalit ole silti välttämättä riippumattomia, sillä korreloimattomuus merkitsee vain lineaarisen riippuvuuden puuttumista. Mahdollisen epälineaarisen riippuvuuden tutkimiseksi voidaan (rajoitetussa mielessä) käyttää neliöidyistä residuaaleista lasketuja auto- ja ristikorrelaatiofunktioita. Jos virheet ovat vain autokorreloimattomia, mutta eivät riippumattomia (esimerkiksi neliöt korreloituneita), eivät estimaattorien  $\hat{\pi}$  ja  $\hat{\delta}$  asympotoottisten jakaumien kovarianssimatriisit ole kuten jaksoissa 4.1 ja 4.2 esitetään (tarkentuvuus ja asympotoottinen normaalisuus voidaan kuitenkin osoittaa yleisin oletuksin). Tällöin esitetyt keskivirheet ja testit eivät myöskään ole päteviä.

**Esimerkki 3.1** (jatkoa). Kuviossa 4.2 nähdään estimoidun mallin (4.15) residuaalit ja niistä lasketut auto- ja ristikorrelaatiofunktiot. Ensimmäisen residuaalin vaihtelu näyttää olevan sarjan alkupuolella hieman voimakkaampaa kuin loppupuolella ja toisessa residuaalisarjassa silmään pistävät muutamat suurehkot arvot. Näistä piirteistä huolimatta residuaalit ovat kuitenkin ”kohtuullisen epämääräisen” näköisiä. Residuaalien auto- ja ristikorrelaatiot ovat pieniä muutamaa suurehkoa ja melko kaukaisilla viipymillä esiintyvää ristikorrelaatiota lukuun ottamatta. Portmanteau-testisuureen arvo ( $Q_{12}^* = 51.38$ ) ei kuitenkaan hälytä, sillä likimääräinen P-arvo on 0.11. Piirtämällä (standardoitujen) residuaalien histogrammit saadaan viitteitä normaalijakaumaa paksuhäntäisemmästä jakaumasta, mihin residuaalien graafinen tarkastelu myös viittaa. Ei-normaalisuuteen (ja ns. ehdolliseen heteroskedastisuuteen) viittaa myös neliöityjen residuaalien lievä korreloituneisuus muutamalla viipymällä. Kaiken kaikkiaan mallia voidaan kuitenkin pitää kohtuullisen onnistuneena.

## 4.5 Grangerin kausaalisuus

**Yleinen määritelmä.** Tarkastellaan aluksi kahta reaaliarvoista muuttujaa  $x$  ja  $z$  ja pohditaan miten voitaisiin määritellä kausaalisuus edellisestä jälkimmäiseen. Menemättä filosofisiin kysymyksiin voidaan ajatella, että kausaalisuudelle on välttämätöntä, ettei syytapahtuma tapahdu seurauksen jälkeen. Jos muuttuja  $x$  on muuttujan  $z$  syy, täytyisi muuttujasta  $x$  olla siten hyötyä ennustettaessa muuttujaa  $z$ .

Olkoon nyt  $x_t$  ja  $z_t$  *reaaliarvoisia* stokastisia prosesseja ja tarkastellaan prosessin  $z_t$  tulevien arvojen ennustamista ajankohtana  $t$ . Merkitään symbolilla  $\mathcal{F}_t$  hypoteettista



**Kuvio 4.2.** Mallin (4.15) residuaalit (ylhäällä vas.  $\hat{\varepsilon}_{1t}$  ja ylhäällä oik.  $\hat{\varepsilon}_{2t}$ ), ja niistä lasketut auto- ja ristikorrelaatiofunktiot. Valkoiset pylväät osoittavat kriittiset rajat ( $\pm 2 \times$  likimääräinen keskivirhe, kun  $\varepsilon_t \sim \text{iid}(0, \Omega)$ ).

muuttujajoukkoa, joka  $x_t$ :n lisäksi sisältää kaikki muuttujat, jotka ovat ajankohtana  $t$  havaittavia ja voivat tulla kysymykseen  $z_t$ :n tulevia arvoja  $z_{t+h}$ ,  $h \geq 1$ , ennustettaessa. Tätä muuttujajoukkoa sanotaan usein informaatiojoukoksi. Siihen perustuvaa  $z_{t+h}$ :n (keskineliövirheen mielessä) optimaalista ennustetta merkitään symbolilla  $z_t(h|\mathcal{F}_t)$ , kun taas  $z_t(h|\mathcal{F}_t \setminus \mathcal{F}_t^x)$  on  $z_{t+h}$ :n optimaalinen ennuste, kun  $\mathcal{F}_t$ :stä on poistettu muuttujajoukko  $\mathcal{F}_t^x = \{x_s, s \leq t\}$ . Olkoot  $\Sigma_z(h|\mathcal{F}_t)$  ja  $\Sigma_z(h|\mathcal{F}_t \setminus \mathcal{F}_t^x)$  vastaavat keskineliövirheet, jotka ovat tässä siis skalaareja (ks. jakso 2.3 tässä mainituista käsitteistä). Prosessien  $x_t$  ja  $z_t$  välillä sanotaan olevan *Grangerin kausaalisuus* edellisestä jälkimmäiseen eli  $x \rightarrow z$ , jos

$$\Sigma_z(h|\mathcal{F}_t) < \Sigma_z(h|\mathcal{F}_t \setminus \mathcal{F}_t^x)$$

ainakin yhdellä  $t$  ja  $h = 1, 2, \dots$ . Toisin sanoen, prosessia  $z_t$  voidaan ennustaa tarkemmin käyttämällä prosessia  $x_t$  kuin ilman sitä.

Okoot nyt  $x_t$  ja  $z_t$  vektoriarvoisia prosesseja ja  $\Sigma_z(h|\mathcal{F}_t)$  ja  $\Sigma_z(h|\mathcal{F}_t \setminus \mathcal{F}_t^x)$  edellä määriteltyjen reaalisten keskineliövirheiden ilmeisiä matriisivastineita. Tällöin pätee välttämättä (vrt. jakso 2.3 ennusteosion alku)

$$\Sigma_z(h|\mathcal{F}_t \setminus \mathcal{F}_t^x) - \Sigma_z(h|\mathcal{F}_t) \geq 0.$$

Koska ei ole kohtuullista vaatia, että vasemmalla oleva matriisi olisi positiivisesti definiitti (miksi?), sanotaan *Grangerin kausaalisuuden* pätevän suuntaan  $x \rightarrow z$ , jos

$$\Sigma_z(h|\mathcal{F}_t) \neq \Sigma_z(h|\mathcal{F}_t \setminus \mathcal{F}_t^x)$$

ainakin yhdellä  $t$  ja  $h = 1, 2, \dots$ . Tulkinta on kuitenkin olennaisesti sama kuin reaalisessa tapauksessa. Huomaa, että esitetty määritelmä ei ota kantaa saman aikaiseen kausaalisuuteen eli siihen voidaanko  $z_t$ :n ennustetta parantaa käyttämällä hyväksi myös muuttujaa  $x_t$ . Tällaista saman aikaista kausaalisuutta on tarkasteltu kirjallisuudessa, mutta siihen ei syvennyttä tällä kurssilla.<sup>12</sup>

**Grangerin kausaalisuus VAR–mallissa.** Grangerin kausaalisuuden määritelmässä käytetty informaatiojoukko  $\mathcal{F}_t$  on käytännön kannalta tietenkin ongelmallinen. Käytännössä se rajataan käytettävissä olevien tai valittujen muuttujien joukoksi, joka VAR(p)–mallin tapauksessa muodostuu vektorin  $y_t$  komponenteista ja niiden viipymistä eli  $\mathcal{F}_t = \{y_s, s \leq t\}$ .

Tarkastellaan nyt stationaarista VAR(p)–prosessia

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \Omega).$$

Ositetaan  $y_t = (z_t, x_t)$  ja kirjoitetaan tämä yhtälö vastaavasti ositettuna

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \sum_{j=1}^p \begin{bmatrix} A_{11,j} & A_{12,j} \\ A_{21,j} & A_{22,j} \end{bmatrix} \begin{bmatrix} z_{t-j} \\ x_{t-j} \end{bmatrix} + \begin{bmatrix} \zeta_t \\ \xi_t \end{bmatrix}. \quad (4.17)$$

Tuntuu varsin selvältä, että Grangerin kausaalisuus suuntaan  $x \rightarrow z$  liittyy siihen onko  $A_{12,j} \neq 0$  jollain  $1 \leq j \leq p$ . Seuraavassa tämä perustellaan formaalisti.

**Lause 4.1.** Jos  $y_t = (z_t, x_t)$  on kuten edellä, niin Grangerin kausaalisuus suuntaan  $x \rightarrow z$  ei päde jos ja vain jos  $A_{12,j} = 0$  kaikilla  $j = 1, \dots, p$ .

<sup>12</sup>Prosessien  $x_t$  ja  $z_t$  välillä sanotaan vallitsevan samanaikainen Grangerin kausaalisuus, jos  $\Sigma_z(1|\mathcal{F}_t) \neq \Sigma_z(1|\mathcal{F}_t \cup \{x_{t+1}\})$ , jossa muuttujajoukko  $\mathcal{F}_t \cup \{x_{t+1}\}$  sisältää joukon  $\mathcal{F}_t$  muuttujien lisäksi muuttujan  $x_{t+1}$ . Toisin sanoen,  $x_{t+1}$ :n lisääminen muuttujajoukkoon  $\mathcal{F}_t$  auttaa parantamaan muuttujan  $z_{t+1}$  ennustetta ajankohtana  $t$ . Voidaan osoittaa, että samanaikainen kausaalisuus on symmetrinen eli jos prosessien  $x_t$  ja  $z_t$  välillä vallitsee samanaikainen kausaalisuus, niin samanaikainen kausaalisuus vallitsee myös prosessien  $z_t$  ja  $x_t$  välillä. Määritelmä ei siten ota kantaa kausaalisuuden suuntaan.



**Todistus:** Oletetaan yksinkertaisuuden vuoksi, että  $p = 1$  ja merkitään  $A_{ij,1} = A_{ij}$  (yleisen tapauksen todistus sujuu samalla tavalla).

Oletetaan  $A_{12} = 0$ . Tällöin prosessille  $z_t$  saadaan yhtälöstä (4.17) stationaarinen VAR(1)–esitys (miksi stationaarinen?)

$$z_t = A_{11}z_{t-1} + \zeta_t.$$

Tästä seuraa (ks. jakso 2.3)

$$\mathbf{E}(z_{t+h}|z_s, s \leq t) = A_{11}^h \mathbf{E}(z_{t+h-1}|z_s, s \leq t), \quad h \geq 1.$$

Toisaalta yhtälöstä (4.17) seuraa tapauksessa  $A_{12} = 0$  myös

$$\mathbf{E}(z_{t+h}|y_s, s \leq t) = A_{11}^h \mathbf{E}(z_{t+h-1}|y_s, s \leq t), \quad h \geq 1.$$

Näistä yhdessä seuraa

$$\begin{aligned} \mathbf{E}(z_{t+1}|z_s, s \leq t) &= A_{11}z_t = \mathbf{E}(z_{t+1}|y_s, s \leq t) \\ \mathbf{E}(z_{t+2}|z_s, s \leq t) &= A_{11}\mathbf{E}(z_{t+1}|z_s, s \leq t) \\ &= A_{11}(A_{11}z_t) \\ &= A_{11}^2\mathbf{E}(z_{t+1}|y_s, s \leq t) \\ &= \mathbf{E}(z_{t+2}|y_s, s \leq t) \\ &\vdots \end{aligned}$$

ja jatkaen induktiivisesti nähdään, että ennusteet  $\mathbf{E}(z_{t+h}|z_s, s \leq t)$  ja  $\mathbf{E}(z_{t+h}|y_s, s \leq t)$  ovat samat kaikilla  $h$  ja  $t$ . Tästä seuraa, ettei Grangerin kausaalisuutta ole.

Oletetaan nyt, että  $A_{12} \neq 0$ . Koska  $\varepsilon_{t+1} \perp\!\!\!\perp \{y_s, s \leq t\}$ , on  $\mathbf{E}(\varepsilon_{t+1}|y_s, s \leq t) = \mathbf{E}(\varepsilon_{t+1}) = 0$  ja yhtälöstä (4.17) saadaan

$$z_{t+1} - \mathbf{E}(z_{t+1}|y_s, s \leq t) = z_{t+1} - A_{11}z_t - A_{12}x_t = \zeta_{t+1}$$

ja

$$\begin{aligned} z_{t+1} - \mathbf{E}(z_{t+1}|z_s, s \leq t) &= z_{t+1} - A_{11}z_t - A_{12}\mathbf{E}(x_t|z_s, s \leq t) \\ &= \zeta_{t+1} + A_{12}[x_t - \mathbf{E}(x_t|z_s, s \leq t)]. \end{aligned}$$

Viimeisellä rivillä olevat kaksi sv:ia ovat riippumattomia (miksi?) ja jälkimmäinen on nolasta poikkeava, koska muutoin  $A_{12}x_t$  olisi sv:ien  $z_s, s \leq t$  funktio. Ennusteen  $\mathbf{E}(z_{t+1}|z_s, s \leq t)$  keskineliövirhe poikkeaa näin ollen optimaalisen ennusteen  $\mathbf{E}(z_{t+1}|y_s, s \leq t)$  keskineliövirheestä eli  $\zeta_{t+1}$ :n kovarianssimatriisista, joten Grangerin kausaalisuus pätee suuntaan  $x \rightarrow z$ .  $\square$

On selvää, että Lauseen 4.1 tulos pätee myös silloin, kun malliin lisätään vakiotermin (tehdään muunnos  $y_t \rightarrow y_t - \mathbf{E}(y_t)$ ). Grangerin kausaalisuutta VAR(p)-mallissa voidaan edellä kuvatussa tilanteessa siten testata testaamalla lineaarista hypoteesia

$$H_0 : A_{12,1} = \dots = A_{12,p} = 0.$$

Näiden rajoitteiden ilmaiseminen yhtälöä (4.10) tai (4.11) käyttäen (ks. jakso 4.2) on kuitenkin hankalaa yleisillä  $n_1$ :n ja  $p$ :n arvoilla. Periaatteessa ongelmia ei kuitenkaan ole. Jos  $z_t$  on  $n_1 \times 1$  vektori ja  $x_t$  on  $n_2 \times 1$  vektori ( $n_1 + n_2 = n$ ) on  $A_{12,j}$   $n_1 \times n_2$  matriisi ja rajoitteiden lukumäärä on  $n_1 n_2 p$ . Waldin testiä tai uskottavuusosamäärätestiä voidaan siten käyttää yhdessä (asymptoottisen)  $\chi^2_{n_1 n_2 p}$ -jakauman kanssa.

**Esimerkki 3.1** (jatkoa). Mallin (4.15) estimointitulosten perusteella on selvää, ettei Grangerin kausaalisuus suuntaan  $dexch_t \rightarrow rdif_t$  saa tukea aineistolta (vertaa relevantteja estimaatteja niiden keskivirheisiin). Käyttäen formaalia uskottavuusosamäärätestiä saadaan (likimääräiseksi)  $P$ -arvoksi 0.65. Grangerin kausaalisuus suuntaan  $rdif_t \rightarrow dexch_t$  näyttää estimointitulosten perusteella sen sijaan realistiselta. Formaalisissa testauksissa  $P$ -arvoksi tulee 0.06. Näiden tarkastelujen perusteella saadun rajoitetun mallin estimointi tuottaa tulokseksi

$$\begin{aligned} \begin{bmatrix} dexch_t \\ rdif_t \end{bmatrix} &= \begin{bmatrix} -0.002 \\ (0.002) \\ 0.003 \\ (0.014) \end{bmatrix} + \begin{bmatrix} 0.411 & 0.023 \\ (0.084) & (0.012) \\ 0 & 1.147 \\ & (0.086) \end{bmatrix} \begin{bmatrix} dexch_{t-1} \\ rdif_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} -0.158 & -0.026 \\ (0.082) & (0.012) \\ 0 & -0.185 \\ & (0.086) \end{bmatrix} \begin{bmatrix} dexch_{t-2} \\ rdif_{t-2} \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_{1t} \\ \tilde{\varepsilon}_{2t} \end{bmatrix}. \end{aligned} \quad (4.18)$$

Verrattuna mallin (4.15) estimointiin on estimointitarkkuus (odotetusti) parantunut, kuten keskivirheitä vertaamalla voidaan todeta.

## 4.6 Impulssivasteanalyysi

**Impulssivasteiden määrittely.** Tarkastellaan samaa VAR(p)-mallia kuin edellisessä jaksossa, jossa esitetyn kausaalisuustarkastelun lisäksi mallia voidaan yrittää tulkita ns. impulssivasteanalyysin avulla. Kuten jaksossa 2.3 mainittiin, impulssivasteanalyysissä mallin virheet tai innovaatiot  $\varepsilon_t$  ajatellaan ”impulsseiksi” tai ”satunnaissokeiksi”, joiden vaikutusta havaittaviin muuttujiin (eli  $y_t$ :n komponentteihin) voidaan tutkia ratkaisemalla  $y_t$ :n VMA( $\infty$ )-esitys (ks. jakso 2.3):

$$y_t = A(\mathbf{B})^{-1} \varepsilon_t = \Psi(\mathbf{B}) \varepsilon_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad (4.19)$$

jossa  $\Psi(\mathbf{B}) = \sum_{j=0}^{\infty} \Psi_j \mathbf{B}^j$ ,  $\Psi_0 = I_n$ .

Oletetaan nyt, että innovaation  $\varepsilon_t$  yhdessä komponentissa  $\varepsilon_{kt}$  ( $1 \leq k \leq n$ ) tapahtuu yhden yksikön muutos (esimerkiksi) ajankohtana  $t = 0$ . Jotta tämän muutoksen tai impulssin vaikutukset voitaisiin eristää ajatellaan (keinotekoisesti), että ennen ajankohtaa  $t = 0$  prosessi on ollut odotusarvossaan eli tässä tapauksessa  $y_t = 0$ ,  $t < 0$ . Olkoon edelleen  $\varepsilon_{k0} = 1$  ja, jotta voitaisiin nähdä miten impulssin vaikutus leviää prosessin  $y_t$  tulevilla arvoilla, oletetaan (edelleen keinotekoisesti), että  $\varepsilon_{i0} = 0$ ,  $i \neq k$ , ja  $\varepsilon_1 = \varepsilon_2 = \dots = 0$ . Jos merkitään  $e_k = (0, \dots, 0, 1, 0, \dots, 0)$  (ykköinen komponentissa  $k$ ), saadaan tehtyjen oletusten voimassa ollessa yhtälöstä (4.19)

$$\begin{aligned} y_0 &= \varepsilon_0 = e_k \\ y_1 &= \Psi_1 \varepsilon_0 = \Psi_1 e_k \\ &\vdots \\ y_h &= \Psi_h \varepsilon_0 = \Psi_h e_k. \end{aligned}$$

Tästä nähdään, että matriisin  $\Psi_h$   $k$ . sarake kuvaa impulssin  $\varepsilon_{k0} = 1$  vaikutusta prosessiin  $y_t$  ajankohtana  $t = h$  eli  $h$  aikayksikköä impulssin tapahtuman jälkeen. Vastaavasti matriisin  $\Psi_h$  alkio eli *impulssivaste*  $\psi_{jk,h}$  kuvaa tämän impulssin vaikutusta prosessin  $y_t$  komponenttiin  $y_{jt}$  ajankohtana  $t = h$ . Koska  $\Psi_h \rightarrow 0$  geometrisesti, kun  $h \rightarrow \infty$ , häipyvät nämä vaikutukset ajan kuluessa. Havainnollisen kuvan asiasta saa piirtämällä impulssivasteista  $\psi_{jk,h}$ ,  $h = 1, \dots, H$ , (tai niiden estimaateista) muodostuva (empiirinen) *impulssivastefunktio* ”tarpeeksi suurella”  $H$ :n arvolla (eli  $\psi_{jk,h} \approx 0$ , kun  $h > H$ ). Näin määriteltäviä impulssivasteita nimitetään myös pidemmin *ennustevirheimpulssivasteiksi* erotuksena alempana määriteltävistä vaihtoehtoisista impulssivasteista.

Matriisin  $\Psi_h$  avulla voidaan siis tutkia miten impulssin  $\varepsilon_t$  vaikutus ilmenee prosessissa  $y_t$  ajankohtana  $t + h$ . Tällaisten yksittäisten vaikutusten lisäksi voidaan tutkia myös kumulatiivisia vaikutuksia ja kokonaisvaikutusta käyttäen matriiseja  $\sum_{j=0}^h \Psi_j$  ja  $\sum_{j=0}^{\infty} \Psi_j = A(1)^{-1}$ . Edellistä käytetään, kun innovaatioissa tapahtunut yhden yksikön muutoksesta on kulunut  $h$  aikayksikköä ja jälkimmäistä, kun tutkitaan kokonaisvaikutusta.

Edellä kuvatulla impulssivasteanalyysillä on yhteys myös edellisen jakson kausaalisuustarkasteluihin. Lauseen 4.1 perusteella kausaalisuutta suuntaan  $x \rightarrow z$  ei ole jos ja vain jos  $A_{12}(\mathbf{B}) = 0$  ( $A(\mathbf{B}) = [A_{ij}(\mathbf{B})]_{i,j=1,2}$ ). Jos  $\Psi(\mathbf{B}) = [\Psi_{ij}(\mathbf{B})]_{i,j=1,2}$  ositetaan  $A(\mathbf{B})$ :n ositusta vastaavalla tavalla, nähdään kääntematriisin määritelmän avulla helposti (ks. Liite A.5), että  $A_{12}(\mathbf{B}) = 0$  jos ja vain jos  $\Psi_{12}(\mathbf{B}) = 0$ . Toisin sanoen, kausaalisuutta voidaan tarkastella myös prosessin VMA( $\infty$ )-esityksen avulla.

Edellä oletettiin, että innovaatioissa (eli  $\varepsilon_t$ :n komponenteissa) tapahtuvat muutokset ovat yhden yksikön suuruisia. Jos innovaatioiden varianssit ovat kovin erisuuruisia, voi olla luontevampaa tarkastella yhden hajontayksikön suuruisia muutoksia. Edellä tarkasteltiin vain yksittäisessä innovaatioissa tapahtuvan muutoksen vaikutuksia, mitä voidaan pitää epäloogisena, jos innovaatiot ovat korreloituneita. Innovaatioiden vä-

linen korrelaatio on kuitenkin varsin hankalasti huomioon otettava kysymys. Eräs ehdotettu ratkaisu on ns. ortogonaalisiin eli korreloimattomiin innovaatioihin siirtyminen (normaalisessa tapauksessa innovaatiot ovat tällöin jopa riippumattomia). Ideana on kirjoittaa VMA( $\infty$ )-esitys muotoon

$$y_t = \sum_{j=0}^{\infty} (\Psi_j C) C^{-1} \varepsilon_{t-j},$$

jossa alakolmiomatriisi tai vaihtoehtoisesti yläkolmiomatriisi  $C$  ( $n \times n$ ) on valittu siten, että  $\text{Cov}(C^{-1}\varepsilon_t) = I_n$ .<sup>13</sup> Muunnetut innovaatiot  $C^{-1}\varepsilon_t$  ovat siten korreloimattomia ja varianssiltaan saman suuruisia, joten niihin liittyvät ns. *ortogonaaliset impulssivasteet* (eli matriisin  $\Psi_j C$  alkiot) välttävät edellä kuvatut ongelmat. Uudeksi hankaluudeksi muodostuu kuitenkin se, että matriisin  $C$  oletettu kolmiorakenne on varsin rajoittava ja impulssivasteanalyysin tulokset riippuvat (mahdollisesti paljonkin) siitä missä järjestyksessä tarkasteltavat aikasarjat esiintyvät vektorissa  $y_t$ . Sopivan järjestyksen valinta vaatii (ei-tilastollista) tietoa tarkasteltavasta ilmiöstä.

**Impulssivastefunktion estimointi.** Tarkastellaan nyt impulssivasteiden estimointia. Olkoon  $\hat{A}_1, \dots, \hat{A}_p$  valitun VAR(p)-mallin kerroinmatriisien (rajoittamattomat tai rajoitetut) SU-estimaattorit. Jaksossa 2.3 esitetyn perusteella on matriisit  $\Psi_j$  tällöin luonteva estimoida käyttäen yhtälöitä

$$\begin{aligned} \hat{\Psi}_1 &= \hat{A}_1 \\ \hat{\Psi}_2 &= \hat{\Psi}_1 \hat{A}_1 + \hat{A}_2 \\ &\vdots \end{aligned}$$

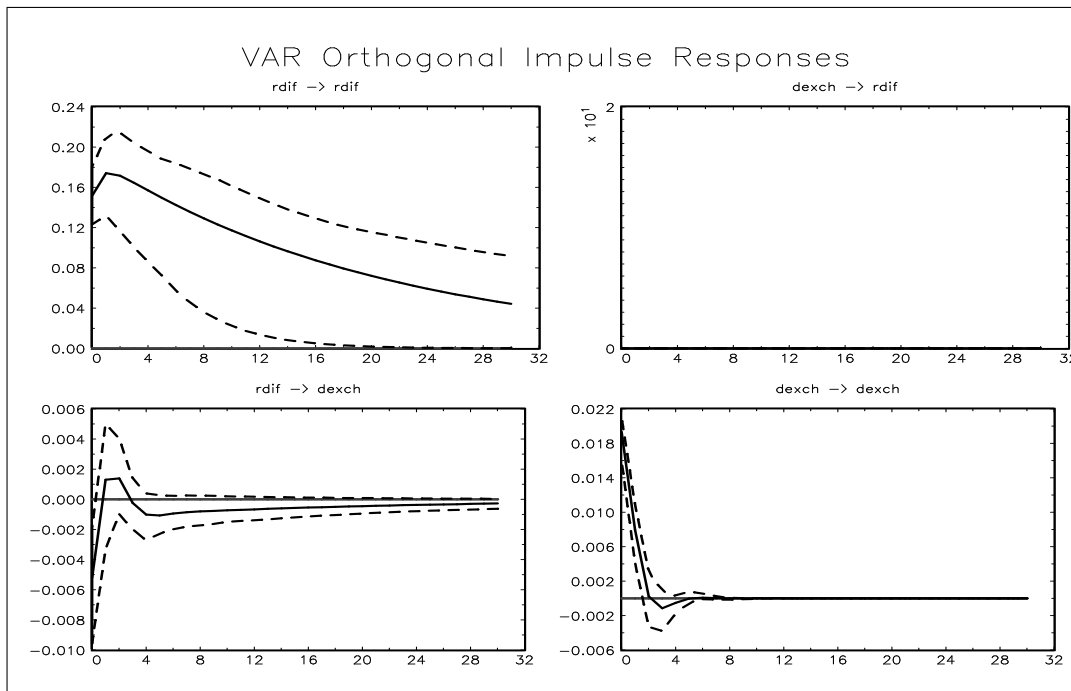
tai yleisesti

$$\hat{\Psi}_j = \sum_{i=1}^j \hat{\Psi}_{j-i} \hat{A}_i, \quad j = 1, 2, \dots, \quad \hat{\Psi}_0 = I_n, \quad \hat{A}_i = 0, \quad i > p.$$

Vaihtoehtoisesti voidaan kirjoittaa  $\hat{\Psi}_j = J \hat{\mathbf{A}}^j J'$ , jossa  $\hat{\mathbf{A}}$  on yhtälössä (2.8) määritellyn matriisin  $\mathbf{A}$  ilmeinen estimaattori ja  $J = [I_n : 0 : \dots : 0]$  ( $n \times np$ ) (ks. jakso 2.3).

Koska estimaattorit  $\hat{\Psi}_j$  saadaan asymptoottisesti normaaleista estimaattoreista jatkuvasti derivoituvan funktion avulla, voidaan ns. deltamenetelmää käyttäen osoittaa, että estimaattorit  $\hat{\Psi}_1, \dots, \hat{\Psi}_H$  ( $H$  kiinteä) ovat asymptoottisesti normaalisti jakautuneita (ks. Liite B). Rajaajakauman kovarianssimatriisi on monimutkainen, joten sen

<sup>13</sup>Koska  $\text{Cov}(\varepsilon_t) = \Omega$  on oletettu positiivisesti definitiksi, tiedetään matriisilaskennasta, että se voidaan esittää tulona  $\Omega = CC'$  (ns. Cholesky-hajotelma), jossa  $C$  on epäsingulaarinen alakolmio- tai yläkolmiomatriisi (ja sama pätee siten myös sen käänteismatriisille  $C^{-1}$ ).

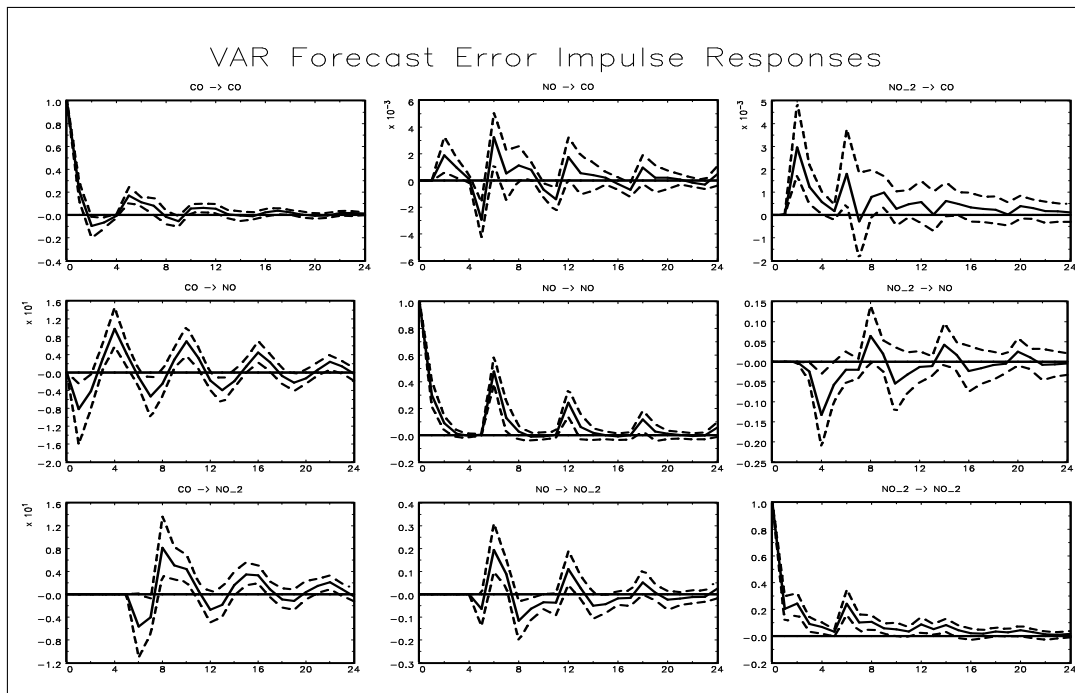


**Kuvio 4.3.** Mallista (4.18) lasketut ortogonaaliset impulssivasteet (oletettu samanaikainen vaikutus  $rdif_t \rightarrow dexch_t$ ). Katkoviivat osoittavat likimääräiset 95%:n luottamusvälit, jotka on muodostettu eräällä bootstrap-menetelmän versiolla.

tarkastelu sivuutetaan. Todetaan kuitenkin, että tapauksessa  $H > p$  rajajakauma on välttämättä singulaarinen ja että näin voi käydä myös, kun  $H < p$ .<sup>14</sup> Tästä syystä impulssivasteille  $\psi_{ij,k}$  muodostetut luottamusvälit saattavat epätarkkoja ja hankalia tulkita. Vaihtoehtoinen tapa muodostaa luottamusvälejä perustuu ns. bootstrap-menetelmien käyttöön.

**Esimerkki 3.1** (jatkoa). Havainnollistetaan impulssivasteiden käyttöä käyttäen Kuvion 1.1 vaihtokurssi/korkoaineistoon sovitettua rajoitettua VAR(2)-mallia (4.18), jossa Granger-kausalisuus ilmenee korkoerosta vaihtokurssiin, mutta ei päinvastoin. Oletetaan, että samanaikainen vaikutus ilmenee myös korkoerosta vaihtokurssiin ja käytetään ortogonaalisia impulssivasteita valiten muunnosmatriisi  $C$  yläkolmiomatriisiksi. Kuvioista 4.3 nähdään, että (mallin oletetusta rakenteesta johtuen) vaihtokurssiin tuleva impulssi ei vaikuta tällöin korkoeron tulevaan kehitykseen (ks. oikealla ylhäällä olevaa kuvion osaa). Se vaikuttaa kuitenkin vaihtokurssin tulevaan kehitykseen  $n$  puolen vuoden ajan (ks. oikealla alhaalla olevaa kuvion osaa). Korkoeroon tuleva impulssi vaikuttaa sen sijaan sekä korkoeron että vaihtokurssin tulevaan kehitykseen.

<sup>14</sup>Tapauksessa  $H > p$  tämä seuraa siitä, että estimaattoreita  $\hat{\Psi}_1, \dots, \hat{\Psi}_H$  on enemmän kuin niiden muodostamisessa käytettäviä estimaattoreita  $\hat{A}_1, \dots, \hat{A}_p$ .



**Kuvio 4.4.** Kuvion 4.1 saasteaineistolle jaksossa 4.4 valitun rajoitetun VAR(8)-mallin tavanomaiset impulssivastefunktiot. Katkoviivat osoittavat likimääräiset 95%:n luottamusvälit, jotka on muodostettu eräällä bootstrap-menetelmän versiolla.

Edellinen vaikutus häviää varsin hitaasti (ks. vasemmalla ylhäällä olevaa kuvion osaa), kun taas jälkimmäinen ilmenee pääasiallisesti muutamien ensimmäisten kuukausien aikana (ks. vasemmalla alhaalla olevaa kuvion osaa). Tuloksia voidaan tulkita niin, että systeemin sopeutuminen muutoksiin tapahtuu vaihtokurssin kautta.

**Esimerkki 4.1** (jatkoa). Tarkastellaan toisena esimerkkinä Kuvion 4.1 saasteaineistolle valitun rajoitetun VAR(8)-mallin tavanomaisia ennustevirheimpulssivastefunktioita, jotka on esitetty Kuviossa 4.4. Impulssivasteista nähdään jaksossa 4.4 todetut viivästetyt vaikutukset. Hiilimonoksidin- ja typpioksidin yhtälöissä ilmenevän impulssin vaikutus typpidioksidiin ilmenee vasta viipymiltä 5 ja 6 alkaen, kun taas muissa tapauksissa impulssien vaikutukset ilmenevät jo viipymilla 1 ja 2.

# Liite

## A Matriisilaskentaa

### A.1 Matriisinormi

Jos  $A = [a_{ij}]$  on dimensiota  $n \times m$  oleva reaalinen matriisi, niin sen (Euklidinen) *normi* on määritelmän mukaan

$$\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2}.$$

On selvää, että määritelmä voidaan esittää vaihtoehtoisesti muodostamalla matriisin  $A$  alkioista  $nm \times 1$  vektori ja käyttämällä tavallista (Euklidisen) vektorinormin määritelmää. Tähän (ja muuhunkin) liittyen määritellään matriisin vektorointiopeeraattori  $\text{vec}(\cdot)$  seuraavasti. Ositetaan  $A = [a_1 \cdots a_m]$  sarakkeittain, jolloin siis  $a_i$  ( $n \times 1$ ) on  $A$ :n  $i$ . sarake ja määritellään vektori (eli vektoroitu  $A$ )

$$\text{vec}(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \quad (nm \times 1).$$

Tällöin on selvää, että  $\|A\|$  on vektorin  $\text{vec}(A)$  normi eli

$$\|A\| = \sqrt{\text{vec}(A)' \text{vec}(A)}.$$

Jälkimmäisestä määritelmästä ja vektorinormin tunnetuista ominaisuuksista nähdään, että matriisinormilla on vektorinormin tavanomaiset ominaisuudet

- (i)  $\|A\| \geq 0$  ja  $\|A\| = 0$  jos ja vain jos  $A = 0$
- (ii)  $\|cA\| = |c| \|A\|$  kaikilla  $c \in \mathbb{R}$
- (iii)  $\|A + B\| \leq \|A\| + \|B\|$  kaikilla  $m \times n$  matriiseilla  $B$  (kolmioepäyhtälö).

Näiden lisäksi pätee

- (iv)  $\|AB\| \leq \|A\| \|B\|$  kaikilla  $m \times l$  matriiseilla  $B$  (ns. submultiplikatiivisuus).

Edellä määritellyn matriisinormin lisäksi käytetään useita muitakin matriisinormeja, joilta vaaditaan edellä mainitut neljä ehtoa (joissakin lähteissä neljättä ehtoa ei vaadita). Näitä matriisinormeja ei kuitenkaan tarvita tällä kurssilla.

Joissakin tapauksissa on kätevää esittää (Euklidinen) matriisinormi käyttäen neliömatriisin jälkeä eli diagonaalialkioiden summaa. Jos  $B = [b_{ij}]$  on  $n \times n$  matriisi, niin sen jälki  $\text{tr}(B)$  on määritelmän mukaan

$$\text{tr}(B) = \sum_{i=1}^n b_{ii}.$$

Määritelmästä seuraa helposti, että matriisin jäljellä on seuraavat ominaisuudet

$$\text{tr}(B + C) = \text{tr}(B) + \text{tr}(C) \quad \text{ja} \quad \text{tr}(BD) = \text{tr}(DB).$$

Edellisessä yhtälössä  $C$  on  $n \times n$  matriisi, mutta jälkimmäisessä  $B$  voi olla yleisemmin  $n \times m$  matriisi ja  $D$  puolestaan  $m \times n$  matriisi.

Jäljen määritelmää käyttäen voidaan todeta, että matriisin  $A$  ( $n \times m$ ) normi voidaan kirjoittaa

$$\|A\| = \sqrt{\text{tr}(A'A)}.$$

Liitteessä A.3 nähdään miten matriisinormia voidaan käyttää tarkasteltaessa matriisijonojen ja sarjojen konvergenssia.

## A.2 Matriisin ominaisarvot ja Jordanin hajotelma

Neliömatriisin  $A$  ( $n \times n$ ) ominaisarvot  $\lambda_1, \dots, \lambda_n$  toteuttavat determinanttiyhtälön  $\det(A - \lambda I_n) = 0$ . Vastaavat ominaisvektorit  $u_1, \dots, u_n$  ( $u_i \neq 0$ ) saadaan yhtälöistä

$$(A - \lambda_i I_n) u_i = 0 \iff Au_i = \lambda_i u_i, \quad i = 1, \dots, n.$$

Jos  $A$  ( $n \times n$ ) ja  $\lambda_1, \dots, \lambda_n$  ovat kuten edellä, on olemassa epäsingulaarinen (eli kääntyvä) matriisi  $P$  ( $n \times n$ ) siten, että

$$A = P\Lambda P^{-1}.$$

Tässä  $\Lambda$  ( $n \times n$ ) on lohkodeagonaalinen ja täsmällisemmin

$$\Lambda = \begin{bmatrix} \Lambda_{n_1} & 0 & \cdots & 0 \\ 0 & \Lambda_{n_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Lambda_{n_k} \end{bmatrix} := \text{diag}[\Lambda_{n_1} \cdots \Lambda_{n_k}],$$

jossa  $n_1 + \cdots + n_k = n$  ja

$$\Lambda_{n_i} = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_i & 1 & & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 0 & \lambda_i & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_i \end{bmatrix}.$$



Matriisin  $\Lambda_{n_i}$  päädiagonaalilla on siis ominaisarvo  $\lambda_i$ , päädiagonaalin yläpuolella olevalla diagonaalilla 1 ja muualla 0 (sama ominaisarvo saattaa esiintyä yhtä useammassa  $\Lambda$ :n lohossa). Matriisin  $A$  edellä esitettyä esitystä sanotaan *Jordanin (kanoniseksi) muodoksi* tai *Jordanin hajotelmaksi*. Jos  $A$ :n ominaisarvot ovat erisuuria, on  $\Lambda$  diagonaalinen diagonaalialkioina  $A$ :n ominaisarvot ja matriisin  $P$  (lineaarisesti riippumattomat) sarakkeet ovat  $A$ :n ominaisvektoreita

Laskemalla voidaan todeta seuraava kurssin kannalta hyödyllinen tulos, jossa  $\binom{k}{j} = 0$ , jos  $j > k$ :

$$\Lambda_{n_i}^k = \begin{bmatrix} \lambda_i^k & \binom{k}{1}\lambda_i^{k-1} & \cdots & \binom{k}{j}\lambda_i^{k-j} & \cdots & \binom{k}{n_i-1}\lambda_i^{k-n_i+1} \\ 0 & \ddots & \ddots & & \ddots & \\ \vdots & 0 & \ddots & \ddots & & \ddots \\ \vdots & & 0 & \ddots & \ddots & \\ \vdots & & & 0 & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & 0 & \ddots \end{bmatrix},$$

jossa diagonaaleilla on aina sama alkio (ja päädiagonaalin alapuolella nolla).

### A.3 Matriisijonot ja matriisisarjat

Matriisijonon  $A_N = [a_{ij,N}]$ ,  $N = 1, 2, \dots$ , konvergenssi kohti (samaa dimensiota olevaa) matriisiä  $A = [a_{ij}]$  voidaan määritellä vaatimalla, että  $a_{ij,N} \rightarrow a_{ij}$  kaikilla  $i, j$ , kun  $N \rightarrow \infty$ . Tällöin merkitään  $\lim_{N \rightarrow \infty} A_N = A$  tai  $A_N \rightarrow A$  (kun  $N \rightarrow \infty$ ).

Matriisiarvoinen sarja  $S = \sum_{i=1}^{\infty} A_i$  määritellään osajonon  $S_N = \sum_{i=1}^N A_i$  raja-arvona eli vaaditaan sellaisen matriisin  $S$  olemassaolo, että  $S_N \rightarrow S$ , jolloin siis  $S = \sum_{i=1}^{\infty} A_i$ .

Jos  $A = [a_{ij}]$  on  $n \times m$  matriisi, niin matriisinnormin määritelmästä seuraa  $\max\{|a_{ij}|\} \leq \|A\| \leq \sqrt{nm} \max\{|a_{ij}|\}$ , jossa maksimi on yli indeksien  $i = 1, \dots, n$  ja  $j = 1, \dots, m$ . Käyttäen tätä ja reaaliulukujonon konvergenssin määritelmää nähdään, että matriisijonolle  $A_N$  pätee  $A_N \rightarrow A$  jos ja vain jos  $\|A_N - A\| \rightarrow 0$ . Vastaava vaihtoehtoinen määritelmä voidaan tietysti esittää myös matriisisarjoille. Edellä todetun ja reaalisten sarjojen suppenemista koskevien tulosten perusteella on myös selvää, että matriisisarjan suppenemiselle on riittävää, että reaalinen sarja  $\sum_{i=1}^{\infty} \|A_i\|$  suppenee (eli itseinen suppeneminen pätee).

Olkoon nyt  $A$  ( $n \times n$ ) neliömatriisi ja  $\lambda_1, \dots, \lambda_n$  sen ominaisarvot. Oletetaan, että  $\rho := \max_{1 \leq i \leq n} |\lambda_i| < 1$ .<sup>15</sup> Jaksossa A.2 esitetyn Jordanin hajotelman avulla nähdään, että

$$A^N = (P\Lambda P^{-1})^N = P\Lambda^N P^{-1}.$$

<sup>15</sup>Jos  $\lambda \in \mathbb{C}$  eli  $\lambda = x + iy$  ( $i^2 = -1$ ), niin  $|\lambda| = \sqrt{x^2 + y^2}$  on  $\lambda$ :n itseisarvo.

Koska tunnetusti  $N^k \rho^N \rightarrow 0$ , kun  $N \rightarrow \infty$  kaikilla  $k \geq 0$ , saadaan jaksossa A.2 todetusta  $\Lambda^N \rightarrow 0$ , joten edellä todetusta yhtälöstä seuraa

$$A^N \rightarrow 0, \quad \text{kun } N \rightarrow \infty.$$

Valitaan nyt  $0 < \bar{\lambda} < 1$  siten, että  $\rho < \bar{\lambda} < 1$  (tällainen  $\bar{\lambda}$  voidaan aina löytää). Tällöin  $N^k \rho^N = N^k (\rho/\bar{\lambda})^N \bar{\lambda}^N$ , jossa  $N^k (\rho/\bar{\lambda})^N \rightarrow 0$ . Näin ollen,  $N^k \rho^N < \bar{\lambda}^N$  kaikilla tarpeeksi suurilla  $N$ :n arvoilla. Konvergenssi  $\Lambda^N \rightarrow 0$  on siten geometrista eli on olemassa vakiot  $C < \infty$  ja  $r < 1$  siten, että matriisin  $\Lambda^N$  alkioille  $[\Lambda^N]_{ij}$  pätee  $\left| [\Lambda^N]_{ij} \right| \leq Cr^N$ . Sama pätee siten matriisin  $A^N = P\Lambda^N P^{-1}$  alkioille. Matriisinormin submultiplikatiivisuudesta seuraa lisäksi, että

$$\|A^N\| = \|P\Lambda^N P^{-1}\| \leq \|P\| \|P^{-1}\| \|\Lambda^N\|,$$

joten myös  $\|A^N\| \rightarrow 0$  geometrisesti (jos  $A$ :n ominaisarvot ovat erisuuria, on  $\Lambda$  diagonaalinen diagonaali-alkioina  $A$ :n itseisarvoltaan ykköstä pienemmät ominaisarvot ja tämän tuloksen perustelu yksinkertaistuu). Edellä todetusta seuraa, että sarja  $\sum_{i=0}^{\infty} A^i$  suppenee ( $A^0 = I_n$ ). Nähdään helposti, että (perustelu jätetään tehtäväksi)

$$\sum_{i=0}^{\infty} A^i = (I_n - A)^{-1}.$$

#### A.4 Polynomit, polynomimatriisit ja potenssisarjat

Astetta  $m$  oleva *polynomi* on funktio

$$p(z) = \sum_{k=0}^m a_k z^k,$$

jossa  $a_k$  ja  $z$  ovat reaali- tai kompleksilukuja ja  $a_m \neq 0$ . Kompleksitapauksessa  $z \in \mathbb{C}$  voidaan polynomille  $p(z)$  johtaa *algebran peruslauseen* avulla esitys

$$p(z) = a_m (z - \zeta_1) \cdots (z - \zeta_m),$$

jossa  $\zeta_1, \dots, \zeta_m$  ovat  $p(z)$ :n nollakohdat tai juuret eli pätee  $p(\zeta_k) = 0$ . Tapaus  $\zeta_k = \zeta_l$  ( $k \neq l$ ) on mahdollinen ja reaalikertoimisessa tapauksessa  $a_k \in \mathbb{R}$  kaikilla  $k$  esiintyvät kompleksiset juuret liittolukuina eli jos  $\zeta_k = x_k + iy_k$  ( $i^2 = -1$ ), niin jollain  $l \neq k$  pätee  $\zeta_l = x_l - iy_l =: \bar{\zeta}_l$ .

*Polynomimatriisi* on matriisi, jonka alkiot ovat polynomeja:

$$P(z) = [p_{ij}(z)], \quad p_{ij}(z) = \sum_{k=0}^{m_{ij}} a_{ij,k} z^k.$$

Asettamalla  $m = \max \{m_{ij}\}$  voidaan vaihtoehtoisesti kirjoittaa

$$P(z) = \sum_{k=0}^m A_k z^k,$$

jossa  $A_k = [a_{ij,k}]$  ja  $a_{ij,k} = 0$ , kun  $k > m_{ij}$ . Tämän esitystavan perusteella voidaan puhua myös *matriisipolynomeista* eli polynomeista, joiden kertoimet ovat matriiseja.

Laskutoimitukset  $A(z)+B(z)$  ja  $A(z)B(z)$  sekä determinantti  $\det(A(z))$  määritellään kuten tavanomaisilla reaalityyppisillä matriiseilla. Samoin käänteismatriisi eli polynomimatriisin  $A(z)$  ( $m \times m$ ) käänteismatriisi  $A(z)^{-1}$  toteuttaa  $A(z)^{-1}A(z) = A(z)A(z)^{-1} = I_m$ . Kuten reaalityyppisillä matriiseilla voidaan käänteismatriisi laskea myös kaavalla

$$A(z)^{-1} = \frac{1}{\det(A(z))} A(z)^* \quad (\det(A(z)) \neq 0),$$

jossa  $A(z)^* = [a_{ij}^*(z)]$  on  $A(z)$ :n ns. adjungaattimatriisi eli  $a_{ij}^*(z) = (-1)^{i+j}$  kertaa determinantti siitä  $A(z)$ :n alimatriisista, joka saadaan poistamalla  $A(z)$ :sta  $j$ . rivi ja  $i$ . sarake (määritelmä on samanlainen myös tavallisilla matriiseilla). Koska determinantti voidaan muodostaa summaamalla matriisin alkiosta laskettuja tuloja, ovat sekä  $\det(A(z))$  että  $a_{ij}^*(z)$  polynomeja. Käänteismatriisin  $A(z)^{-1}$  alkiot ovat siten rationaalifunktioita eli tyyppiä  $\sum_{j=0}^q a_j z^j / \sum_{j=0}^r b_j z^j$  ( $q, r < \infty$ ).

Palataan nyt tavalliseen polynomiin  $p(z) = \sum_{k=0}^m a_k z^k$ . On helppo todeta, että  $p(z)$  voidaan kirjoittaa

$$p(z) = p(1) + (1-z)q(z), \tag{A.1}$$

jossa

$$q(z) = \sum_{k=0}^{m-1} b_k z^k, \quad b_k = - \sum_{j=k+1}^m a_j \quad k = 1, \dots, m-1.$$

Vastaavanlainen vaihtoehtoinen esitys on

$$p(z) = p(1)z + (1-z)r(z), \tag{A.2}$$

jossa

$$r(z) = \sum_{k=0}^{m-1} c_k z^k, \quad c_0 = a_0 \quad \text{ja} \quad c_k = - \sum_{j=k+1}^m a_j, \quad k = 1, \dots, m-1.$$

Nämä esitykset voidaan yleistää edelleen potenssisarjoille eli tapaukseen

$$p(z) = \sum_{k=0}^{\infty} a_k z^k \quad (|z| \leq 1).$$

Jos oletetaan, että

$$\sum_{k=1}^{\infty} k |a_k| < \infty,$$

voidaan helposti todeta, että yhtälöt (A.1) ja (A.2) pätevät, kun  $|z| \leq 1$  ja  $m$  korvataan symbolilla  $\infty$ . Lisäksi pätee  $\sum_{k=1}^{\infty} |b_k| < \infty$  ja  $\sum_{k=1}^{\infty} |c_k| < \infty$ .

Edellä esitetty yleistyy matriisitapaukseen. Olkoon

$$P(z) = [p_{ij}(z)] = \sum_{k=0}^m A_k z^k \quad (|z| \leq 1),$$

jossa  $A_k = [a_{ij,k}]$  toteuttaa ehdon

$$\sum_{k=1}^{\infty} k \|A_k\| < \infty.$$

Soveltamalla aikaisempaa polynomeihin  $p_{ij}(z)$  saadaan tuloksen (A.1) yleistys

$$P(z) = P(1) + (1-z)Q(z), \quad (\text{A.3})$$

jossa

$$Q(z) = \sum_{k=0}^{\infty} B_k z^k, \quad B_k = - \sum_{j=k+1}^{\infty} A_j \quad \text{ja} \quad \sum_{k=1}^{\infty} \|B_k\| < \infty.$$

Vastaavasti saadaan tuloksen (A.2) yleistys

$$P(z) = P(1)z + (1-z)R(z), \quad (\text{A.4})$$

jossa

$$R(z) = \sum_{k=0}^{\infty} C_k z^k, \quad C_0 = A_0, \quad C_k = - \sum_{j=k+1}^{\infty} A_j, \quad k \geq 1 \quad \text{ja} \quad \sum_{k=1}^{\infty} \|C_k\| < \infty.$$

Jos  $P(z)$  on astetta  $m$  oleva polynomimatriisi eli  $A_k = 0$  kaikilla  $k > m$ , niin  $Q(z)$  ja  $R(z)$  ovat astetta  $m-1$  olevia polynomimatriiseja.

## A.5 Ositetut matriisit ja Kroneckerin tulo

Ositettu matriisi on matriisi, jonka alkiot ovat matriiseja eli

$$A = [A_{ij}] = \begin{bmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nm} \end{bmatrix}.$$

Jos  $A = [A_{ij}]$  ja  $B = [B_{ij}]$ , niin (olettaen, että dimensiot täsmäävät laskutoimitusten määritelmien kanssa)

$$A + B = [A_{ij} + B_{ij}] \quad \text{ja} \quad AB = \left[ \sum_k A_{ik} B_{kj} \right].$$

Nämä tulokset seuraavat tavanomaisista matriisien yhteen- ja kertolaskukaavoista. Lisäksi, kun  $\det(A) \neq 0$  ja  $\det(A_{ii}) \neq 0$  ( $i = 1, 2$ ),

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix},$$

jossa  $A_{i,j} = A_{ii} - A_{ij}A_{jj}^{-1}A_{ji}$ . Mainitaan myös tulokset  $\det(A) = \det(A_{11})\det(A_{22})$ , kun  $A_{12} = 0$  tai  $A_{21} = 0$  sekä  $\det(A) = \det(A_{11})\det(A_{2,1})$ , kun  $\det(A_{11}) \neq 0$  ja  $\det(A) = \det(A_{22})\det(A_{1,2})$ , kun  $\det(A_{22}) \neq 0$ .

Matriisien  $A$  ( $n \times m$ ) ja  $B$  ( $p \times q$ ) *Kroneckerin tulo* on  $np \times mq$  matriisi

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nm}B \end{bmatrix} = [a_{ij}B].$$

Kroneckerin tulolla on seuraavat ominaisuudet, jotka voidaan todeta jokseenkin suoraan määritelmästä (olettaen, että laskutoimitukset ovat määriteltyjä).

- $(A \otimes B)(C \otimes D) = (AC \otimes BD)$
- $(A \otimes B)' = A' \otimes B'$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- $A \otimes (B + C) = A \otimes B + A \otimes C$
- Jos matriisien  $A$  ( $n \times n$ ) ja  $B$  ( $m \times m$ ) ominaisarvot ovat  $\lambda_1, \dots, \lambda_n$  ja  $\mu_1, \dots, \mu_m$ , niin matriisin  $(A \otimes B)$  ominaisarvot ovat  $\lambda_i\mu_j$  ( $i = 1, \dots, n, j = 1, \dots, m$ )
- $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$

## A.6 Ehdollisen odotusarvon ominaisuuksia

Tarkastellaan lyhyesti yleistä ennustustilannetta, jossa (reaalisen) satunnaismuuttujan  $Y$  arvoa halutaan ennustaa satunnaisvektorin (sv)  $X$  saaman arvon  $X = x$  avulla. Todennäköisyyslaskennasta tiedetään, että keskineliövirheen mielessä optimaalinen ennuste on  $Y$ :n ehdollinen odotusarvo ehdolla  $X = x$  eli pätee

$$\mathbb{E}[(Y - \mathbb{E}(Y|X=x))^2] \leq \mathbb{E}[(Y - g(x))^2]$$

olipa  $g(x)$  mikä tahansa  $x$ :n funktio (tässä odotusarvojen äärellisyys oletetaan). Jatkuvien jakaumien tapauksessa määritellään ehdollinen odotusarvo kaavalla

$$\mathbb{E}(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y;x) dy = \int_{-\infty}^{\infty} y \frac{f_{Y,X}(y,x)}{f_X(x)} dy,$$

jossa  $f_{Y,X}(y, x)$  on sv:n  $(Y, X)$  yhteistiheysfunktio,  $f_X(x) = \int_{-\infty}^{\infty} f_{Y,X}(y, x) dy$  on  $X$ :n reunajakauman tiheysfunktio ja  $f_{Y|X}(y; x) = f_{Y,X}(y, x) / f_X(x)$  on  $Y$ :n ehdollinen tiheysfunktio ehdolla  $X = x$ . Näissä määritelmissä, samoin kuin seuraavissa tarkasteluissa,  $Y$  voi olla vektoriarvoinen.

Kun  $x$  vaihtelee yli satunnaisvektorin  $X$  mahdollisten arvojen määrittelee  $x$ :n funktio  $E(Y | X = x)$  satunnaismuuttujan, josta on luonteva käyttää merkintää  $E(Y | X)$ . Kehittyneessä todennäköisyyslaskennassa ehdollinen odotusarvo määritelläänkin suoraan satunnaismuuttujana ja siellä se yleistetään myös tapaukseen, jossa satunnaisvektori  $X$  voi olla ääretönulotteinen. Tämä tilanne tulee vastaan, kun tarkastellaan stokastisen prosessin ennustamista olettaen, että prosessin  $\{y_t, t = 0, \pm 1, \dots\}$  kaikki ennusteaikakohtaa edeltävät arvot tunnetaan. Ehdollisesta odotusarvosta riittää kuitenkin tietää vain sen tavanomaiset ominaisuudet, jotka pätevät tällöinkin. Eri-tyisesti, että

**EO1**  $E(aY_1 + bY_2 | X) = aE(Y_1 | X) + bE(Y_2 | X)$ , kun  $a$  ja  $b$  ovat vakioita (lineaarisuus)

**EO2**  $E(Y | X) = E(Y)$ , kun  $Y \perp\!\!\!\perp X$ .

**EO3**  $E(Y) = E[E(Y | X)]$  (ns. iteroidun odotusarvon laki)

**EO4**  $E[g(X)Y | X] = g(X)E(Y | X)$  mille tahansa funktiolle  $g$  (jolla tulo  $g(X)Y$  on määritelty ja sen odotusarvo on äärellisenä olemassa).

## A.7 Derivointikaavoja

Kun  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  on derivoituva, merkitään

$$\frac{\partial}{\partial x} f(x) = \left[ \frac{\partial}{\partial x_1} f(x) \quad \cdots \quad \frac{\partial}{\partial x_n} f(x) \right]' \quad (n \times 1)$$

ja

$$\frac{\partial}{\partial x'} f(x) = \left[ \frac{\partial}{\partial x_1} f(x) \quad \cdots \quad \frac{\partial}{\partial x_n} f(x) \right] \quad (1 \times n).$$

Kun  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  on derivoituva, merkitään

$$\frac{\partial}{\partial x} f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x) & \cdots & \frac{\partial}{\partial x_1} f_m(x) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_n} f_1(x) & \cdots & \frac{\partial}{\partial x_n} f_m(x) \end{bmatrix}$$

On helppo todeta, että  $\partial a'x / \partial x = a$  ( $a$   $n \times 1$ ) ja  $\partial Ax / \partial x = A'$ , kun  $A$  on  $m \times n$ . Edelleen, kun  $A$  on  $n \times n$ ,

$$\begin{aligned} \frac{\partial}{\partial x} x'Ax &= (A + A')x \\ &= 2Ax, \quad \text{kun } A = A'. \end{aligned}$$

Kun  $B = B'$  ( $m \times m$ ), saadaan näiden avulla edelleen

$$\frac{\partial}{\partial x} (c - Ax)' B (c - Ax) = -2A'Bc + 2A'BAx \quad (c \ m \times 1 \text{ ja } A \ m \times n).$$

Mainitaan lisäksi tulokset

$$\frac{\partial}{\partial A} x'Ax := \left[ \frac{\partial}{\partial a_{ij}} x'Ax \right] = [x_i x_j] = xx' \quad (x \ n \times 1 \text{ ja } A \ n \times n)$$

ja

$$\frac{\partial}{\partial A} \log \det(A) = (A^{-1})' = (A')^{-1}.$$

Viimeksi mainittu perustuu determinantin tunnettuun kehityskaavan  $\det(A) = \sum_k a_{ik} a_{ik}^*$ , jossa  $A$ :n adjungaattimatriisiin  $A^*$  (ks. Liite A.4) alkio  $a_{ik}^*$  ei riipu  $A$ :n alkioista  $a_{ij}$ . Ketjusääntöä soveltaen saadaan näin ollen,

$$\frac{\partial}{\partial a_{ij}} \log \det(A) = \frac{1}{\det(A)} \frac{\partial \det(A)}{\partial a_{ij}} = \frac{1}{\det(A)} a_{ij}^* = [A^{-1}]_{ji} = [(A^{-1})']_{ij},$$

jossa viimeistä edellinen yhtälö perustuu käänteismatriisin laskukaavaan  $A^{-1} = (\det(A))^{-1} A^*$  ja adjungaattimatriisin määritelmään (ks. Liite A.4).

Kahdessa viimeksi mainitussa derivointikaavassa oletetaan, että matriisi  $A$  on rajoittamaton. Ne eivät siten päde esimerkiksi  $A$ :n ollessa symmetrinen. *Symmetrisessä tapauksessa*  $A = A'$  ( $n \times n$ ) on  $a_{ij} = a_{ji}$  ja saadaan

$$\frac{\partial}{\partial a_{ij}} x'Ax = \begin{cases} x_i^2, & \text{kun } i = j \\ 2x_i x_j, & \text{kun } i \neq j \end{cases}$$

ja

$$\frac{\partial}{\partial a_{ij}} \log \det(A) = \begin{cases} [A^{-1}]_{ii}, & \text{kun } i = j \\ 2[A^{-1}]_{ji}, & \text{kun } i \neq j \end{cases}.$$

Matriiseja käyttäen nämä kaavat voidaan ilmaista muodossa

$$\frac{\partial}{\partial A} x'Ax = 2xx' - \mathbf{dg}[xx']$$

ja

$$\frac{\partial}{\partial A} \log \det(A) = 2A^{-1} - \mathbf{dg}[A^{-1}],$$

jossa merkinnällä  $\mathbf{dg}[\cdot]$  tarkoitetaan diagonaalimatriisia diagonaalialkioina asianomaisen matriisin diagonaalialkiot (esim.  $\mathbf{dg}[A] = \mathbf{dg}[a_{11} \cdots a_{nn}]$ ).

Kun viimeksi mainittuja tuloksia sovelletaan monisteen jaksossa 4.2 esitettyyn log-uskottavuusfunktioon

$$l(\pi, \Phi) = \frac{T}{2} \log \det(\Phi) - \frac{1}{2} \sum_{t=1}^T \varepsilon_t(\pi)' \Phi \varepsilon_t(\pi)$$

saadaan

$$\frac{\partial}{\partial \Phi} \partial \varepsilon_t(\pi)' \Phi \varepsilon_t(\pi) = 2\varepsilon_t(\pi) \varepsilon_t(\pi)' - \mathbf{d}\mathbf{g} [\varepsilon_t(\pi) \varepsilon_t(\pi)']$$

ja (käytetään jakson 4.2 merkintää  $\Phi^{-1} = \Omega$ )

$$\frac{\partial}{\partial \Phi} \log \det(\Phi) = 2\Phi^{-1} - \mathbf{d}\mathbf{g} [\Phi^{-1}] = 2\Omega - \mathbf{d}\mathbf{g} [\Omega].$$

Parametrin  $\Phi$  uskottavuusyhtälöksi saadaan siten

$$\frac{\partial}{\partial \Phi} l(\pi, \Phi) = \frac{T}{2} \{2\Omega - \mathbf{d}\mathbf{g} [\Omega]\} - \frac{1}{2} \sum_{t=1}^T \{2\varepsilon_t(\pi) \varepsilon_t(\pi)' - \mathbf{d}\mathbf{g} [\varepsilon_t(\pi) \varepsilon_t(\pi)']\} = 0.$$

Tarkastelemalla tätä matriisiyhtälöä alkioittain nähdään, että se voidaan esittää yhtäpitävässä muodossa

$$\frac{\partial}{\partial \Phi} l(\pi, \Phi) = \frac{T}{2} \Omega - \frac{1}{2} \sum_{t=1}^T \varepsilon_t(\pi) \varepsilon_t(\pi)' = 0.$$

Huomataan siis, että sama tulos saadaan, kun log-uskottavuusfunktiota derivoitaessa matriisia  $\Phi$  käsitellään ilman symmetrisyysrajoitetta.

## B Asymptoottisia tuloksia

Seuraavassa esitetään lyhyesti stokastiseen konvergenssiin ja jakaumakonvergenssiin liittyviä tuloksia (mainitut konvergenssikäsitteet oletetaan tunnetuiksi). Ensimmäinen tulos kertoo miten stokastinen konvergenssi käyttäytyy jatkuvissa muunnoksissa. Tässä kuten myöhemminkin tarkoitetaan vakiolla ei-satunnaista (äärellistä) suuretta.

**Lause B.1.** Jos jono sv:ta  $X_1, X_2, \dots$  ( $k \times 1$ ) konvergoi stokastisesti kohti vakiovektoria  $c$  eli  $X_n \xrightarrow{p} c$  ja  $g: \mathbb{R}^k \rightarrow \mathbb{R}^l$  on jatkuva pisteessä  $c$ , niin  $g(X_n) \xrightarrow{p} g(c)$ .

Tulos pätee myös matriisitapauksessa, joka voidaan palauttaa lauseen tilanteeseen suorittamalla vektorointi (ks. Liite A.1). Sama pätee seuraaviin jatkuvan kuvauksen lauseisiin, joista ensimmäinen koskee jakaumakonvergenssia.

**Lause B.2.** Jos jono sv:ta  $X_1, X_2, \dots$  ( $k \times 1$ ) konvergoi jakaumaltaan kohti sv:ia  $Z$  eli  $X_n \xrightarrow{d} Z$  ja  $g: \mathbb{R}^k \rightarrow \mathbb{R}^l$  on jatkuva, niin  $g(X_n) \xrightarrow{d} g(Z)$ .

Mainittakoon, että tämän lauseen jatkuvuusoletusta voidaan lieventää. Funktion  $g$  jatkuvuus riittää vaatia vain joukossa  $C_g$ , jolle pätee  $\mathbf{P}\{Z \in C_g\} = 1$ . Koska jakaumakonvergenssista kohti vakiota seuraa stokastinen konvergenssi, seuraa Lauseesta



B.2 Lauseen B.1 tulos, kun  $Z = c$  on vakio.<sup>16</sup> Seuraava lause sisältää edellä mainitut erikoistapauksenaan.

**Lause B.3.** Oletetaan, että sv-jonoille  $X_1, X_2, \dots$  ( $k \times 1$ ) ja  $Y_1, Y_2, \dots$  ( $l \times 1$ ) pätee  $X_n \xrightarrow{d} Z$  ja  $Y_n \xrightarrow{p} c$ , jossa  $c$  on vakiovektori. Tällöin  $(X_n, Y_n) \xrightarrow{d} (Z, c)$ .

Lauseiden 1.3 ja 1.4 suorina seurauksia saadaan mm. seuraavat tulokset.

**Seuraus B.1.** Oletetaan Lauseen B.3 tilanne eli  $X_n \xrightarrow{d} Z$  ja  $Y_n \xrightarrow{p} c$  (vakio) ja lisäksi että  $k = l = 1$ . Tällöin,

- (i)  $X_n + Y_n \xrightarrow{d} Z + c$
- (ii)  $X_n Y_n \xrightarrow{d} Zc$
- (iii)  $X_n / Y_n \xrightarrow{d} Z/c$ , kun  $c \neq 0$ .

Tässä ensimmäinen tulos pätee tietysti myös vektoritapauksessa, jota tarkastellaan seuraavassa, jossa  $Y_n$  on  $l \times k$  matriisi.

**Seuraus B.2.** Olkoon  $X_1, X_2, \dots$  ja  $Z$  ( $k \times 1$ ) kuten Lauseessa 1.4 eli  $X_n \xrightarrow{d} Z$  ja  $Y_1, Y_2, \dots$  jono  $l \times k$  matriiseja, jolle pätee  $Y_n \xrightarrow{p} A$ , jossa  $A$  ( $l \times k$ ) on vakiomatriisi. Tällöin,

(i)  $Y_n X_n \xrightarrow{d} AZ$

Jos  $k = l$ , pätee myös

(ii)  $X_n' Y_n X_n \xrightarrow{d} Z' A Z$

(iii)  $X_n' Y_n^{-1} X_n \xrightarrow{d} Z' A^{-1} Z$ , kun  $A$  on epäsingulaarinen (eli kääntyvä).

Nämä tulokset seuraavat funktioiden  $(A, x) \mapsto Ax$ ,  $(A, x) \mapsto x' Ax$  ja  $A \mapsto A^{-1}$  ( $A$  epäsingulaarinen) jatkuvuudesta ja Lauseista B.2 ja B.3. Kun  $Z$  on vakio, voidaan jakaumakonvergenssi korvata (vahvemmalla) stokastisella konvergenssilla. Käänteismatriisin jatkuvuus voidaan perustella seuraavasti. Todetaan ensin, että determinanttifunktio  $A \mapsto \det(A)$  on jatkuva, koska determinantti voidaan muodostaa summamalla matriisin alkioista laskettuja tuloja (katso  $2 \times 2$  tapausta). Koska käänteismatriisin  $A^{-1}$  alkioita voidaan lausua  $A$ :n ja sen alimatriisien determinanttien osamäärien

---

<sup>16</sup>Stokastisessa konvergenssista (mahdollisesti kohti satunnaista suuretta) seuraa aina jakaumakonvergenssi, mutta käänteinen tulos pätee vain, kun jakaumakonvergenssissa raja-arvo on vakio.

avulla (ks. Liite A.4 ja A.7), on funktio  $A \mapsto A^{-1}$  jatkuvien funktioiden yhdisteenä jatkuva.

Seurauksessa B.1(iii) on implisiittisesti oletettu, että  $Y_n \neq 0$  kaikilla  $n$  (vastaavasti Seurauksessa B.2(iii) on oletettu  $Y_n$ :n epäsingulaarisuus kaikilla  $n$ ). Vaikka oletuksista  $Y_n \xrightarrow{p} c$  ja  $c \neq 0$  seuraa, että tapahtuman  $\{Y_n \neq 0\}$  todennäköisyys lähestyy ykköstä, ei ehdon  $Y_n \neq 0$  toteutuminen äärellisillä  $n$ :n arvoilla ei ole kuitenkaan taattu ilman lisäoletuksia (kuten  $Y_n$ :n jakauman jatkuvuutta). Tämä teoreettinen hankaluus voitaisiin ottaa huomioon määrittelemällä  $1/Y_n$  jollain (mielivaltaisella) tavalla ilman, että esitetty jakaumakonvergenssi muuttuisi mitenkään. Kuten kirjallisuudessa yleensäkin, ei tätä seikkaa oteta yksinkertaisuuden vuoksi eksplisiittisesti huomioon, vaan Seurauksen B.1(iii) kaltaisissa tilanteissa  $Y_n \neq 0$  oletetaan, jos  $Y_n$ :n stokastinen raja-arvo on nolasta poikkeava. Vastaava tulkinta tehdään ilman eri mainintaa Seurauksen 1.2(iii) kaltaisissa tilanteissa.

Edellä esitettyjen tulosten ja väliarvolauseeseen avulla voidaan perustella ns. *delta-menetelmä*, joka reaaliarvoisessa tapauksessa on seuraava. Olkoon  $\theta \in \Theta \subset \mathbb{R}$  estimoitava parametri ja  $\theta_0$  sen ”todellinen” arvo. Oletetaan (yksinkertaisuuden vuoksi), että parametrivaruus  $\Theta$  on avoin väli ja tarkastellaan  $\theta_0$ :n asympotoottisesti normaalia estimaattoria  $\hat{\theta}_n$  eli

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathbf{N}(0, \sigma^2(\theta_0)) \quad (\sigma^2(\theta_0) > 0).$$

Jos  $h : \Theta \rightarrow \mathbb{R}$  on jatkuvasti derivoituva funktio ja  $h'(\theta_0) \neq 0$ , niin muunnoksen  $h(\theta_0)$  estimaattorille  $h(\hat{\theta}_n)$  pätee

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta_0)) \xrightarrow{d} \mathbf{N}(0, [h'(\theta_0)]^2 \sigma^2(\theta_0)).$$

Yleistys vektoriparametrin tapaukseen on seuraava. Olkoon  $\Theta \subseteq \mathbb{R}^d$  avoin ja konvekksi ja  $h : \Theta \rightarrow \mathbb{R}$  jatkuvasti derivoituva funktio. Oletetaan, että gradienttivektori  $\partial h(\theta)/\partial \theta' = [\partial h(\theta)/\partial \theta_1 \ \cdots \ \partial h(\theta)/\partial \theta_d]$  toteuttaa  $\partial h(\theta_0)/\partial \theta \neq 0$  ja että estimaattorille  $\hat{\theta}_n$  ( $d \times 1$ ) pätee

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathbf{N}_d(0, \Sigma(\theta_0)), \quad \Sigma(\theta_0) \text{ positiivisesti definiitti.}$$

Tällöin,

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta_0)) \xrightarrow{d} \mathbf{N}(0, V(\theta_0)), \quad V(\theta_0) = [\partial h(\theta)/\partial \theta'] \Sigma(\theta_0) [\partial h(\theta)/\partial \theta].$$

Deltamenetelmä voidaan yleistää edelleen tapaukseen, jossa  $h$  on vektoriarvoinen. Väliarvolauseetta täytyy tällöin soveltaa erikseen funktion  $h$  kuhunkin komponenttiin.