

Keskusteluaiheita Discussion papers

Yrjö O. Vartia

KVADRAATTISTEN MIKROYHTÄLÖIDEN
AGGREGOINNISTA

No 25

24.1.1979

17.4.1979 mennessä havaitut
painovirheet yms. korjattu

This series consists of papers with limited circulation, intended to stimulate discussion. The papers must not be referred or quoted without the authors' permission.



Esipuhe

Tämä työ liittyy Etlan ekonometrisen ryhmän (aiemmin suhdanneryhmä) ja ulkomaantalouden ryhmän tutkimustyön ohjaamiseen. Sektoriyhtälöiden hyväksikäyttöä makroyhtälöiden estimoinnissa on Etlassa pohdiskeltu monen vuoden aikana ja pohdintojen erästä vaihetta kuvaavat ohjaamani johdannossa mainitut tilastotieteen proseminaarit. Aggregointi ja väitöskirjassani käsittelemäni indeksiongelma ovat niinikään lähisukulaisia; tätä sukulaisuutta on käytetty hyväksi esim. julkaisussa Edgren, Turkkila ja Vartia (1978).

Aggregointiongelman yksityiskohtaista ja mahdollisimman yksinkertaista esitystä on kaivattu esim. ulkomaantalouden ryhmässä, jossa sektoriyhtälöiden hyväksikäyttöä suurempien vienti- ja tuontiryhmien käyttäytymisen selittämisessä on sovittu lähitulevaisuudessa kokeiltavan. Tämän vuoksi on nähty tarpeelliseksi kerätä samoihin kansiin aiempien julkaisemattomien käsikirjoitusten ja uudempien pohdintojen tuloksia. Tämä tutkimus perustuu aiempaan vuoden 1978 marraskuun käsikirjoitukseen ja on laajemman aggregointia käsittelevän tutkimuksen 1. osa.

Helsingissä 24.1.1979

Yrjö O. Vartia

KVADRAATTISTEN MIKROYHTÄLÖIDEN AGGREGOINNISTA

1. Johdanto

Tässä paperissa pohditaan edelleen aggregointiongelmia niistä lähtökohdista, jonka muodostavat ohjaamani tilastotieteen proseminarityöt Ranta (1973), Ojamo ja Hyötyläinen (1973), Saarinen ja Laesvuori (1973), keskusteluaiheessa Edgren, Turkkila, Vartia (1978) esitetyt tulokset sekä Ilkan (1978) riippumattomasti johtamat kauniit tulokset. Esitystapa on tarkoituksellisen elementaarinen matemaattisessa suhteessa, jolloin nähdään konkreettisesti mistä aggregoinnin vaikeudet ja makroyhtälön eri termit syntyvät.

2. Mikrodata ja mikroyhtälöt

Tarkastellaan n :ää erillistä tilastoyksikköä $a_1, \dots, a_n$ (esim. henkilöitä, kotitalouksia, yrityksiä, toimialoja tai mahd. atomeja, kärpäsiä, maita). Kunkin tilastoyksikön kohdalla on tiedossa muuttujien ψ , x ja y arvot, jotka muodostavat seuraavan havaintomatriisin

$$\begin{array}{c} \psi \quad x \quad y \\ \begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{matrix} \begin{bmatrix} \psi_1 & x_1 & y_1 \\ \psi_2 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ \psi_n & x_n & y_n \end{bmatrix} \end{array}$$

Muuttuja ψ merkitsee tässä selitettävää muuttujaa (esim. kotitalouden a kulutusta), x ja y ovat selittävät muuttujat (esim. tulot ja kotitalouden kokoa kuvaava muuttuja).

Kaikkien muuttujien oletetaan olevan kvantitatiivisia, siis intervalli-, suhde- tai absoluuttisen asteikon muuttujia. Järjestys- tai luokitteluasteikon muuttujia ei tässä tarkastella, vaikka saman tapaisia tuloksia voidaan johtaa myös niille, ks. Edgren, Turkkila, Vartia (1978, s. 52-61). Niinikään muuttujia tarkastellaan jatkuvina, vaikka esim. lukumäärämuuttujille useat tulokset pätevät sellaisenaan.

Aluksi tarkastellaan yksinkertaista tapausta. Oletetaan, että mikrotasolla muuttujien (ψ_i, x_i, y_i) välillä vallitsee lineaarinen vakiokertominen riippuvuus:

$$(1) \quad \psi_i = \alpha + \beta_1 x_i + \beta_2 y_i, \quad i = 1, \dots, n$$

Siis kunkin kotitalouden a_i kohdalla ψ_i on muuttujien (x_i, y_i) lineaarikombinaatio, jossa kertoimet $(\alpha, \beta_1, \beta_2)$ ovat kaikille kotitalouksille samat. (Tarkastelemme aluksi vain deterministisiä yhtälöitä; myöhemmin on helppo liittää mukaan tarpeelliset virhetermit). Spesifikaatio (1) merkitsee, että on olemassa funktio $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ s.e.

$$(2) \quad g(x, y) = \alpha + \beta_1 x + \beta_2 y$$

ja kaikille $i = 1, \dots, n$ pätee

$$(3) \quad \psi_i = g(x_i, y_i).$$

Tämä on tavattoman restriktiivinen hypoteesi, jota pian kevennetään.

3. Makrodata

Tarkastelemme tässä kahta tapausta. Ensimmäisessä tapauksessa makrotason muuttujat ovat yksinkertaisesti mikrotason muuttujien summia, joita merkitään isoilla kirjaimilla:

$$(4) \quad \Psi = \sum_{i=1}^n \psi_i$$

$$X = \sum x_i$$

$$Y = \sum y_i$$

Jos ψ_i ja x_i ovat kotitalouden a_i kulutus ja tulot markkoissa niin Ψ ja X ovat kotitalouksien (esim. koko kansantalouden) kokonaiskulutus ja kokonaistulot markkoissa. Esimerkiksi kansantalouden tilinpidossa käytetään (kuten yleensäkin tilinpitojärjestelmissä) tätä yksinkertaista aggregointitapaa, suoraa yhteenlaskua. Siksi sen tarkasteleminen on niin tärkeää. Myöhemmin voidaan tutkia mutkikkaampia makrotason muuttujien määrittelyjä, kuten geometrisia keskiarvoja, lineaarikombinaatioita jne., jotka tulevat myös kysymykseen hieman toisenlaisissa tilanteissa.

Suorat summat (4) kuvaavat "kokonaiskäsitteitä" tai "totaaleja" makrotasolla; yleensä esim. X :n suuruusluokka on oleellisesti suurempi kuin minkä tahansa x_i :n. Usein on luontevaa ja yksinkertaista tarkastella summien (4) sijaan (aritmeettisia) keskiarvoja:

$$(5) \quad \bar{\psi} = \frac{1}{n} \sum \psi_i = \Psi/n$$

$$\bar{x} = \frac{1}{n} \sum x_i = X/n$$

$$\bar{y} = \frac{1}{n} \sum y_i = Y/n.$$

Tällöin usein ajatellaan tutkittavan "keskiarvoperhettä", "keskiarvo-kuluttajaa" tms. Per capita-tarkasteluissa käytetään juuri keskiarvoja (5) summien (4) sijaan. Huomattakoon, ettei muunnos summista keskiarvoihin välttämättä ole mikään triviaali operaatio. Vain jos n on (esim. ajassa) vakio ja voidaan luontevasti valita vain yhdellä tavalla on asia lähes triviaali. Usein ei ole itsestään selvää, mitkä tilastoyksiköt $a_1, \dots, a_n$ ovat ja mikä on niiden lukumäärä n : esimerkiksi kotitaloudet tai toimialat voidaan määritellä eri tavoin.

Tavanomainen tutkimusongelma on selvittää makrotason muuttujien (Ψ, X, Y) tai $(\bar{\psi}, \bar{x}, \bar{y})$ välisiä riippuvuuksia. Tavanomainen ennakkoluulo ja virhe-käsitys on, että jos mikrotasolla muuttujien välillä pätee riippuvuus

$$(6) \quad \psi_i = g(x_i, y_i)$$

niin sama (tai samantapainen) riippuvuus pätee myös makrotasolla:

$$(7) \quad \Psi = G(X, Y) \text{ tai}$$

$$(8) \quad \bar{\psi} = \bar{g}(\bar{x}, \bar{y})$$

Näin ei asia suinkaan ole kuin erityisen yksinkertaisissa tapauksissa, joita tarkastelemme heti seuraavassa. Yleensä Ψ riippuu muistakin tekijöistä kuin totaaleista X ja Y ja $\bar{\psi}$ riippuu muistakin suureista kuin keskiarvoista $\bar{x}$ ja $\bar{y}$. Erityisesti x - ja y -tekijöiden varianssit ja kovarianssit vaikuttavat makromuuttujien Ψ ja $\bar{\psi}$ suuruuteen ja vaihteluihin, kuten tulemme seuraavassa osoittamaan.

4. Erikoistapaus, jossa yksinkertainen aggregointi "onnistuu"

Yhtälöt (1)-(3) kuvaavat sellaisen yksinkertaisen tapauksen, jossa makrotasolla päästään hyvin yksinkertaisiin, mikrotason riippuvuuksia vastaaviin yhtälöihin. Tarkastellaan mistä tekijöistä Ψ riippuu:

$$\begin{aligned}(9) \quad \Psi &= \sum \psi_i \\ &= \sum (\alpha + \beta_1 x_i + \beta_2 y_i) \\ &= \sum \alpha + \sum \beta_1 x_i + \sum \beta_2 y_i \\ &= \sum \alpha + \beta_1 \sum x_i + \beta_2 \sum y_i \\ &= n\alpha + \beta_1 X + \beta_2 Y.\end{aligned}$$

Havaitsemme, että on olemassa funktio $G: \mathbb{R}^2 \rightarrow \mathbb{R}$ s.e.

$$(10) \quad G(x,y) = n\alpha + \beta_1 x + \beta_2 y$$

ja kaikilla (Ψ, X, Y) -arvoilla pätee

$$(11) \quad \Psi = G(X, Y).$$

Tässä olemme oletaneet, etteivät $(\alpha, \beta_1, \beta_2)$ -parametrit tai lukumäärä n muutu tarkastelutilanteesta toiseen. Jos esim. lukumäärä n muuttuu vaikkapa vuodesta toiseen on se otettava mukaan G -funktion argumentteihin. Mutta näiden äärimmäisen restriktiivisten olettamusten vallitessa todellakin makrotasolla totaali muuttujien (Ψ, X, Y) välillä vallitsee mikrotason riippuvuutta $\psi_i = g(x_i, y_i)$ vastaava riippuvuus (10)-(11). Tämä ns. analogiaperiaate ei siis ole aina loogisesti ristiriitainen, vaan on olemassa "Simpletanioita" (mahdollisia maailmoita), joissa se realisoituu. Tulos ei ole suinkaan uusi. Sen esittää mm. J.S. Cramer (1971, s. 175-181) ja Chipman (1976, s. 618-626), joka formuloi abstraktilla matemaattisella käsitteistöllä ns. "täydellisen aggregoinnin" ehdot, ks. myös Klein (1946 a, b), Malinvaud (1954), Theil (1965). Pulmana on, miten edetä silloin kun tilanne on mutkikkaampi.

Samalla tavalla saadaan aritmeettisille keskiarvoille:

$$\begin{aligned} (12) \quad \bar{\psi} &= \sum \psi_i / n = \Psi / n \\ &= \sum (\alpha + \beta_1 x_i + \beta_2 y_i) / n \\ &= (n\alpha + \beta_1 \sum x_i + \beta_2 \sum y_i) / n \\ &= n\alpha + \beta_1 X/n + \beta_2 Y/n \\ &= \alpha + \beta_1 \bar{x} + \beta_2 \bar{y}. \end{aligned}$$

Olemme kerranneet välivaiheet, että myöhemmät komplikaatiot näkyisivät oikeassa valossa. Siis on olemassa funktio $\bar{g}: \mathbb{R}^2 \rightarrow \mathbb{R}$ s.e.

$$(13) \quad \bar{g}(x,y) = \alpha + \beta_1 x + \beta_2 y$$

ja kaikilla $(\bar{\psi}, \bar{x}, \bar{y})$ -arvoilla pätee

$$(14) \quad \bar{\psi} = \bar{g}(\bar{x}, \bar{y}).$$

Vieläpä $\bar{g}$ -funktio (13):ssa ja g -funktio (2):ssa ovat sama funktio:

$$(15) \quad g(x,y) = \bar{g}(x,y) = \alpha + \beta_1 x + \beta_2 y.$$

Keskiarvojen $(\bar{\psi}, \bar{x}, \bar{y})$ ja kaikkien mikromuuttujien (ψ_i, x_i, y_i) välillä vallitsee siis tässä Simpletaniassa täsmälleen sama riippuvuus. Tämä tulos on toiminut ajattelun ohjenuorana aggregoinnin ns. analogiaperiaatetta sovellettaessa. On kuviteltu, että yleisemmissäkin tilanteissa makroyhtälön tulisi olla samantapainen kuin mikroyhtälö. Havaitsemme, ettei tavallisesti näin ole, vaan makroyhtälöön tulee aivan uusia aggregaattitason selittäjiä, joita mikrotasolla ei voi esiintyä.

5. Lineaaristen mikroyhtälöiden aggregointi

Tarkastellaan esityksen yksinkertaistamiseksi seuraavaksi tapausta, jossa mikroyhtälöt ovat edelleen lineaarisia, mutta niiden parametrit vaihtelevat mikroyhtälöstä toiseen:

$$(16) \quad \psi_i = \alpha_i + \beta_{1i}x_i + \beta_{2i}y_i.$$

Edellä tarkasteltiin tapausta, jossa kaikkien mikroyksiköiden a_i parametrit olivat samat: $\alpha_i = \alpha$, $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$, $i = 1, \dots, n$. Nyt tarkastellaan lyhyesti niitä komplikaatioita, joita syntyy tällaisesta mutkistuksesta. Jokaisella mikroyksiköllä a_i on siis oma käyttäytymisfunktionsa $g_i: \mathbb{R}^2 \rightarrow \mathbb{R}$ s.e.

$$(17) \quad g_i(x, y) = \alpha_i + \beta_{1i}x + \beta_{2i}y$$

ja kaikilla (ψ_i, x_i, y_i) arvoilla pätee

$$(18) \quad \psi_i = g_i(x_i, y_i).$$

En ryhdy tässä pohtimaan eräitä funktiokäsitteeseen liittyviä melko yleisiä väärinkäsityksiä. Totean vain, että yhtälöt (17) ja (18) vasta antavat täsmällisen merkityksen (16):lle. Voimme todeta, että näiden yhtälöiden kuvaama Simpletania on yhä erittäin "spesiaali", tarkastelemme jatkuvasti merkillistä erityistapausta.

Tarkastellaan jälleen ensin makrototaalia Ψ :

$$\begin{aligned}
 (19) \quad \Psi &= \sum \psi_i \\
 &= \sum (\alpha_i + \beta_{1i}x_i + \beta_{2i}y_i) \\
 &= \sum \alpha_i + \sum \beta_{1i}x_i + \sum \beta_{2i}y_i.
 \end{aligned}$$

Tähän on moni muukin pysähtynyt, ks. esim. J.S. Cramer (1971, s. 177).
 Kuinka edetä? Tässä vaiheessa tarjoutuu käyttöön ns. aggregoinnin perus-
lause, joksi Edgren, Turkkila, Vartia (1978, s. 15) ristivät seuraavan
 identiteetin:

$$(20) \quad \sum_{i=1}^n w_i x_i y_i = \bar{x} \bar{y} + \text{cov}(x, y),$$

jossa $\bar{x} = \sum w_i x_i$, $\bar{y} = \sum w_i y_i$, $\text{cov}(x, y) = \sum w_i (x_i - \bar{x})(y_i - \bar{y})$ ja painojen w_i summa on yksi. Painojen ei välttämättä tarvitse olla ei-negatiivisia, sillä $\sum w_i = 1$ riittää (20):n johtamiseen. Suoritetaan johto tässä:

$$\begin{aligned}
 (21) \quad \text{cov}(x, y) &= \sum w_i (x_i - \bar{x})(y_i - \bar{y}) \\
 &= \sum w_i (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\
 &= \sum w_i x_i y_i - \bar{y} \sum w_i x_i - \bar{x} \sum w_i y_i + \bar{x} \bar{y} \sum w_i \\
 &= \sum w_i x_i y_i - \bar{x} \bar{y} - \bar{x} \bar{y} + \bar{x} \bar{y} \\
 &= \sum w_i x_i y_i - \bar{x} \bar{y}.
 \end{aligned}$$

Todellakin vain määritelmät sekä ehto $\sum w_i = 1$ tulivat käyttöön. Luvut x_i ja y_i voivat olla mielivaltaisia reaalitykkuja (tai lausekkeitä) ja painot w_i samoin mielivaltaisia kunhan vain $\sum_{i=1}^n w_i = 1$. Erityisesti valitsemalla $w_i = 1/n$ saadaan seuraavat identiteetit:

$$\begin{aligned} (22) \quad \frac{1}{n} \sum x_i y_i &= \left(\frac{1}{n} \sum x_i \right) \left(\frac{1}{n} \sum y_i \right) + \frac{1}{n} \sum \left(x_i - \frac{1}{n} \sum x_i \right) \left(y_i - \frac{1}{n} \sum y_i \right) \\ &= \bar{x} \bar{y} + \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= \bar{x} \bar{y} + \text{cov}(x, y). \end{aligned}$$

$$\begin{aligned} (23) \quad \sum x_i y_i &= (\sum x_i)(\sum y_i)/n + \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= XY/n + n \text{cov}(x, y), \end{aligned}$$

jossa $X = \sum x_i$ ja $Y = \sum y_i$. Näiden tulosten ymmärtäminen edellyttää, että tuntee joitakin tärkeimpiä kovarianssin $\text{cov}(x, y)$ ominaisuuksia. Yleisesti voidaan "kirjoittaa" (vaikka symboleilla on tässä hieman tavanomaista yleisempi merkitys):

$$\begin{aligned} (24) \quad \text{cov}(x, y) &= \sum w_i (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{\text{cov}(x, y)}{\sqrt{\text{cov}(x, x) \text{cov}(y, y)}} \sqrt{\text{cov}(x, x) \text{cov}(y, y)} \\ &= \frac{\text{cov}(x, y)}{s(x)s(y)} s(x)s(y) \\ &= r(x, y) s(x)s(y), \end{aligned}$$

jossa $s(x) = \sqrt{\text{cov}(x,x)} = \sqrt{\sum w_i (x_i - \bar{x})^2}$, $s(y) = \sqrt{\text{cov}(y,y)} = \sqrt{\sum w_i (y_i - \bar{y})^2}$
 ja $r(x,y) = \text{cov}(x,y)/s(x)s(y)$. Jos painot w_i ovat ei-negatiivisia niin
 $\bar{x}$ ja $\bar{y}$ ovat x :n ja y :n painotetut keskiarvot, $s(x)$ ja $s(y)$ niiden hajon-
 nat, $\text{cov}(x,y)$ niiden kovarianssi ja $r(x,y)$ niiden korrelaatiokerroin.
 Tällöin $s(x)$ ja $s(y)$ ovat reaalisia ja ei-negatiivisia sekä $-1 \leq r(x,y) \leq 1$.
 Kovarianssin ja korrelaatiokertoimen ominaisuuksista ks. Vasama-Vartia
 (1973, s. 422-427).

Palataan yhtälöön (19) ja sovelletaan identiteettiä (23) termiin $\sum \beta_{1i} x_i$:

$$(25) \quad \sum \beta_{1i} x_i = (\sum \beta_{1i})(\sum x_i)/n + n \text{cov}(\beta_{1i}, x_i) \\ = \bar{\beta}_1 X + n \text{cov}(\beta_{1i}, x_i).$$

Käytämme joskus kovarianssin argumenteissa alaindeksejä selvyiden vuoksi.
 Samalla tavalla kehitellään termiä $\sum \beta_{2i} y_i$ ja saadaan:

$$(26) \quad \Psi = \sum \alpha_i + \sum \beta_{1i} x_i + \sum \beta_{2i} y_i \\ = n\bar{\alpha} + \bar{\beta}_1 X + \bar{\beta}_2 Y \\ + n \text{cov}(\beta_{1i}, x_i) + n \text{cov}(\beta_{2i}, y_i).$$

Tämä on haluttu esitys. Totaalimuuttujien X ja Y kertoimina ovat β_{1i} - ja β_{2i} -parametrien keskiarvot $\bar{\beta}_1 = \frac{1}{n} \sum \beta_{1i}$, $\bar{\beta}_2 = \frac{1}{n} \sum \beta_{2i}$. Lisäksi on makro-
 yhtälöön tullut lisää kaksi kovariansstermiä n :lla kerrottuna; nämä ko-
 varianssit häviävät, jos β -kertoimet eivät vaihtelee mikroyksiköistä toi-
 seen. Niinikään esim. kovarianssi $\text{cov}(\beta_{1i}, x_i)$ on nolla, jos x_i -arvot

(esim. tulot) ovat samat kaikissa mikroyksiköissä (jolloin siis $s(x) = 0$), tai β_{1i} - ja x_i -arvot eivät korreloi keskenään ($r(\beta_{1i}, x_i) = 0$). Tämä kaikki näkyy yhtälöstä

$$(27) \quad \text{cov}(\beta_{1i}, x_i) = s(\beta_{1i})s(x_i)r(\beta_{1i}, x_i).$$

Olemme siis saaneet tarkastellussa Simpletaniassa seuraavan tuloksen.

Totaalimuuttuja Ψ riippuu totaaleista X ja Y sekä kovariansseista $\text{cov}(\beta_1, x)$ ja $\text{cov}(\beta_2, y)$ siten, että on olemassa funktio $H: \mathbb{R}^4 \rightarrow \mathbb{R}$ s.e.

$$(28) \quad H(x, y, c_1, c_2) = n\bar{\alpha} + \bar{\beta}_1 x + \beta_2 y + nc_1 + nc_2$$

jossa $\bar{\alpha} = \frac{1}{n} \sum \alpha_i$, $\bar{\beta}_1 = \frac{1}{n} \sum \beta_{1i}$, $\bar{\beta}_2 = \frac{1}{n} \sum \beta_{2i}$ ja kaikille viisikoille $(\Psi, X, Y, \text{cov}(\beta_1, x), \text{cov}(\beta_2, y))$ pätee

$$(29) \quad \Psi = H(X, Y, \text{cov}(\beta_1, x), \text{cov}(\beta_2, y)).$$

Tässä täsmällisessä mielessä mainitut kaksi kovarianssia ovat täysin totaalimuuttujiin X ja Y rinnastettavia selittäviä muuttujia, kun totaalin Ψ vaihtelua halutaan selittää.

Keskiarvoille tulokset ovat vielä yksinkertaisempia. Ne voitaisiin johtaa eo. tapaan myös suoraan, mutta tulos saadaan helpoimmin suoraan (26):sta jakamalla n :llä. Keskiarvoille siis pätee tarkastellussa Simpletaniassa:

$$(30) \quad \bar{\psi} = \bar{\alpha} + \bar{\beta}_1 \bar{x} + \bar{\beta}_2 \bar{y}$$

$$+ \text{cov}(\beta_1, x) + \text{cov}(\beta_2, y).$$

Tämä palautuu yhtälöksi (12), jos kaikilla i :n arvoilla $\beta_{1i} = \beta_1 = \bar{\beta}_1$ eli $s(\beta_1) = 0$ ja $\beta_{2i} = \beta_2 = \bar{\beta}_2$ eli $s(\beta_2) = 0$. Täsmällisemmin ilmaistuna, on olemassa funktio $h: \mathbb{R}^4 \rightarrow \mathbb{R}$ s.e.

$$(31) \quad h(x, y, c_1, c_2) = \bar{\alpha} + \bar{\beta}_1 x + \bar{\beta}_2 y + c_1 + c_2$$

jossa $\bar{\alpha} = \frac{1}{n} \sum \alpha_i$, $\bar{\beta}_1 = \frac{1}{n} \sum \beta_{1i}$, $\bar{\beta}_2 = \frac{1}{n} \sum \beta_{2i}$ ja kaikille viisikoille $(\bar{\psi}, \bar{x}, \bar{y}, \text{cov}(\beta_1, x), \text{cov}(\beta_2, y))$ pätee

$$(32) \quad \bar{\psi} = h(\bar{x}, \bar{y}, \text{cov}(\beta_1, x), \text{cov}(\beta_2, y)).$$

Tässä täsmällisessä mielessä mainitut kovarianssit ovat selittäjinä $\bar{\psi}$:n käyttäytymisyhtälössä.

Vastaavanlaisia tuloksia johdettiin verofunktioille julkaisussa Edgren, Turkkila, Vartia (1978). Muualla aggregointikirjallisuudessa en tiedä vastaavia tuloksia johdetun. Esimerkiksi J.S. Cramer (1971, s. 177). etenee yhtälöstä (19) vikasuuntaan. Ilkka (1978) käsittelee asiaa samantapaisesti, mutta olettaa, että mikrotasolla kaikki käyttäytymisyhtälöt ovat identtisiä, ts. tähän tapaukseen sovellettuna, että $\beta_{1i} = \beta_1$ ja $\beta_{2i} = \beta_2$ kaikilla i :n arvoilla. Emme pysähdy kommentoimaan näitä tuloksia yksityiskohtaisemmin, vaan jatkamme yleisempiin tilanteisiin.

6. Kvadraattisten mikroyhtälöiden aggregointi

Luovumme mikroyhtälöiden lineaarisuusolettamuksesta ja tarkastelemme oleellisesti yleisempää kvadraattista tapausta:

$$(33) \quad \psi_i = \alpha_i + \beta_{1i}x_i + \beta_{2i}y_i + \beta_{11i}x_i^2 + \beta_{22i}y_i^2 + 2\beta_{12i}x_iy_i.$$

Tarkastelemme $\bar{\psi}$:n määräytymistä:

$$(34) \quad \begin{aligned} \bar{\psi} &= \frac{1}{n} \sum \psi_i \\ &= \frac{1}{n} \sum (\alpha_i + \beta_{1i}x_i + \beta_{2i}y_i + \beta_{11i}x_i^2 + \beta_{22i}y_i^2 + 2\beta_{12i}x_iy_i) \\ &= \frac{1}{n} (\sum \alpha_i + \sum \beta_{1i}x_i + \sum \beta_{2i}y_i + \sum \beta_{11i}x_i^2 + \sum \beta_{22i}y_i^2 + 2\sum \beta_{12i}x_iy_i). \end{aligned}$$

Tämä alkaa jo olla melkoista "harakanvarvasta". Lineaariset termit käsitellään edellä esitetyllä tavalla, joten niiden kohdalla ei tule uusia vaikeuksia. Kvadraattisten termien kohdalla tarkastellaan erikseen termiä $\frac{1}{n}(\sum \beta_{12i}x_iy_i)$. Sovelletaan tähän aggregoinnin peruslausetta (22) seuraavasti:

$$(35) \quad \begin{aligned} \frac{1}{n} \sum \beta_{12i}x_iy_i &= (\frac{1}{n} \sum \beta_{12i})(\frac{1}{n} \sum x_iy_i) + \text{cov}(\beta_{12}, xy) \\ &= \bar{\beta}_{12}(\frac{1}{n} \sum x_iy_i) + \text{cov}(\beta_{12}, xy). \end{aligned}$$

Mutta tässä esiintyvä summa $(\frac{1}{n} \sum x_i y_i)$ on eri asia kuin $(\frac{1}{n} \sum x_i)(\frac{1}{n} \sum y_i) = \bar{x} \bar{y}$, joka halutaan selittäjäksi makroyhtälöön. Siis muokataan (34):ää edelleen soveltaen toisen kerran aggregoinnin peruslausetta (22):

$$(36) \quad \frac{1}{n} \sum \beta_{12i} x_i y_i = \bar{\beta}_{12} (\bar{x} \bar{y} + \text{cov}(x, y)) + \text{cov}(\beta_{12}, xy) \\ = \bar{\beta}_{12} \bar{x} \bar{y} + \bar{\beta}_{12} \text{cov}(x, y) + \text{cov}(\beta_{12}, xy).$$

Tämä on haluttu esitys! Termi $\bar{\beta}_{12} \bar{x} \bar{y}$ vastaa makrotasolla mikroyhtälöiden "ristitermejä" $\beta_{12i} x_i y_i$. Lisäksi on tullut makroyhtälöön kaksi kovarianssitermiä $\bar{\beta}_{12} \text{cov}(x, y)$ ja $\text{cov}(\beta_{12}, xy)$. Edellinen riippuu parametrien β_{12i} keskimääräisestä suuruudesta ja selittävien muuttujien x ja y keskinäisestä variaatiosta $\text{cov}(x, y)$ ja jälkimmäinen parametrien β_{12i} ja ristitulojen $x_i y_i$ keskinäisestä variaatiosta. Kumpikin termi on lähellä nollaa, jos ristitermillä $x_i y_i$ ei ole merkittävää vaikutusta ψ_i :hin, ts. jos β_{12i} :t ovat yleensä pieniä. Vastaavasti saadaan neliötermeille, esim.

$$(37) \quad \frac{1}{n} \sum \beta_{11i} x_i^2 = \bar{\beta}_{11} \bar{x}^2 + \bar{\beta}_{11} \text{cov}(x, x) + \text{cov}(\beta_{11}, x^2) \\ = \bar{\beta}_{11} \bar{x}^2 + \bar{\beta}_{11} s^2(x) + \text{cov}(\beta_{11}, x^2).$$

Kun tulokset kootaan yhteen saadaan $\bar{\psi}$:lle selitysyhtälöksi

$$(38) \quad \bar{\psi} = \bar{\alpha} + \bar{\beta}_1 \bar{x} + \bar{\beta}_2 \bar{y} \\ + \bar{\beta}_{11} \bar{x}^2 + \bar{\beta}_{22} \bar{y}^2 + 2\bar{\beta}_{12} \bar{x} \bar{y} \\ + \text{cov}(\beta_1, x) + \text{cov}(\beta_2, y)$$

$$\begin{aligned}
 &+ \bar{\beta}_{11}s^2(x) + \bar{\beta}_{22}s^2(y) + 2\bar{\beta}_{12}\text{cov}(x, y) \\
 &+ \text{cov}(\beta_{11}, x^2) + \text{cov}(\beta_{22}, y^2) + 2\text{cov}(\beta_{12}, xy).
 \end{aligned}$$

Kahdella ensimmäisellä rivillä on normaalit analogiamenetelmän mukaiset termit, joissa kertoimina on intuition mukaiset keskimääräiset parametrit. Kolmannella rivillä on lineaarisessa tapauksessa jo esiintyneet kovarianssit, jotka ilmaisevat missä määrin selittävien muuttujien keskimääräistä suuremmat tai pienemmät arvot ovat sattuneet kotitalouksille, joiden reaktiot näihin tekijöihin nähden ovat poikkeukselliset. Neljännellä rivillä esiintyvät varsinaiset uutuudet lineaariseen tapaukseen verrattuna, nimittäin selittävien muuttujien variaatiosta ja kovariaatiosta johtuvat $\bar{\psi}$ -tekijässä näkyvät vaikutukset. Jos x - ja y -selittäjät olisivat keskittyneet keskiarvojen kohdalle häviäisivät kaikki nämä vaikutukset, koska tällöin $s^2(x) = s^2(y) = \text{cov}(x, y) = 0$ (viimeinen termi voi hävitä muulloinkin). Yleensä nämä vaikutukset on kokonaan jätetty huomiotta, mikä on tietenkin loogisesti virheellistä. Termin $\bar{\beta}_{11}s^2(x)$ merkitystä on syytä vertailla vastaavaan "tavalliseen" termiin $\bar{\beta}_{11}\bar{x}^2$. Useiden taloudellisten muuttujien (esim. tulojen, kulutuksen) tapauksessa varianssit $s^2(x)$ ovat suuria, jakaumien vinouden vuoksi mahdollisesti jopa suurempia kuin $\bar{x}^2$. ($s^2(x) > \bar{x}^2 \iff s(x) > \bar{x} \iff s(x)/\bar{x} = \text{variaatiokerroin} > 1$. Tämä ei ole mahdoton tapaus!) Koska termeillä $\bar{\beta}_{11}\bar{x}^2$ ja $\bar{\beta}_{11}s^2(x)$ on sama kerroin, ovat $\bar{x}^2$:n ja $s^2(x)$:n vaikutukset $\bar{\psi}$ -muuttujaan suoraan verrannollisia niiden keskinäiseen suuruuteen. Jos $s^2(x)$ on jätetty selittävien muuttujien listasta pois $\bar{\psi}$:tä selitettäessä, tulevat muidenkin muuttujien kertoimet todennäköisesti virheellisesti arvioitua. Erityisesti on odotettavissa, että $\bar{x}^2$ (joka luultavasti korreloi positiivisesti x :n varianssin $s^2(x)$ kanssa) saisi liian suuren kertoimen esim. regressioanalyysissä.

Vastaavasti voidaan tulkita termit $\bar{\beta}_{22}s^2(y)$ ja $2\bar{\beta}_{12}\text{cov}(x, y)$. Jälkimmäinen termi osoittaa, että selittävien muuttujien kovarianssin muuttumisella on välitön vaikutus $\bar{\psi}$ -muuttujaan, vaikka muuttujien keskiarvot ja varianssit pysyisivät ennallaan (siis "ceteris paribus").

Viidennen rivin kovarianssit vastaavat tulkinnaltaan kolmannen rivin kovariansseja. Nämä häviävät jos vastaavat kertoimet eivät vaihtelee mikroyksiköstä toiseen.

Formuloimme tämän matemaattisen tuloksemme teoreemana.

Teoreema 1. Tarkastellaan mikroyksikköjä $a_1, \dots, a_n$ ja näiden ominaisuuksia kuvaavia muuttujia ψ , x ja y . Oletetaan, että kunkin mikroyksikön tapauksessa on olemassa funktio $g_i: \mathbb{R}^2 \rightarrow \mathbb{R}$ siten, että

$$(38) \quad \psi_i = g_i(x_i, y_i).$$

Jos $g_i(x, y)$ on yleinen kvadraattinen funktio, jonka kertoimet saavat muuttua mikroyksiköstä toiseen

$$(39) \quad g_i(x, y) = \alpha_i + \beta_{1i}x + \beta_{2i}y + \beta_{11i}x^2 + \beta_{22i}y^2 + 2\beta_{12i}xy,$$

niin muuttujien (ψ_i, x_i, y_i) aritmeettisten keskiarvojen välillä valitsee yhtälö

$$\begin{aligned}
 (40) \quad \bar{\psi} &= \bar{\alpha} + \bar{\beta}_1 \bar{x} + \bar{\beta}_2 \bar{y} \\
 &+ \bar{\beta}_{11} \bar{x}^2 + \bar{\beta}_{22} \bar{y}^2 + 2\bar{\beta}_{12} \bar{x}\bar{y} \\
 &+ \text{cov}(\beta_1, x) + \text{cov}(\beta_2, y) \\
 &+ \bar{\beta}_{11} s^2(x) + \bar{\beta}_{22} s^2(y) + 2\bar{\beta}_{12} \text{cov}(x, y) \\
 &+ \text{cov}(\beta_{11}, x^2) + \text{cov}(\beta_{22}, y^2) + 2\text{cov}(\beta_{12}, xy).
 \end{aligned}$$

Siis on olemassa funktio $H: \mathbb{R}^{10} \rightarrow \mathbb{R}$ s.e.

$$(41) \quad H \begin{bmatrix} x & y \\ c_1 & c_2 \\ s_1 & s_2 & c_3 \\ c_4 & c_5 & c_6 \end{bmatrix} = \begin{aligned} &\bar{\alpha} + \bar{\beta}_1 x + \bar{\beta}_2 y \\ &+ \bar{\beta}_{11} x^2 + \bar{\beta}_{22} y^2 + 2\bar{\beta}_{12} xy \\ &+ c_1 + c_2 \\ &+ \bar{\beta}_{11} s_1^2 + \bar{\beta}_{22} s_2^2 + 2\bar{\beta}_{12} c_3 \\ &+ c_4 + c_5 + c_6 \end{aligned}$$

ja jolle pätee identtisesti

$$(42) \quad \bar{\psi} = H \begin{bmatrix} \bar{x} & \bar{y} \\ \text{cov}(\beta_1, x) & \text{cov}(\beta_2, y) \\ s(x) & s(y) & \text{cov}(x, y) \\ \text{cov}(\beta_{11}, x^2) & \text{cov}(\beta_{22}, y^2) & \text{cov}(\beta_{12}, xy) \end{bmatrix}$$

Teoreema 1 antaa täydellisen ratkaisun kvadraattisten mikroyhtälöiden aggregoinnille. Teoreeman mukaan analogiaperiaatteen mukaisten selittäjien rinnalla on käytettävä yhtälössä (42) mainittuja variansseja ja kovariansseja.

Jos mikrotason yhtälöissä parametrit ovat i:stä riippumattomia häviävät 2. ja neljännen rivin kovarianssit ja saadaan seuraava teoreema.

Teoreema 2. Tarkastellaan mikroyksikköjä $a_1, \dots, a_n$ ja näiden ominaisuuksia kuvaavia muuttujia. Oletetaan, että on olemassa mikroyksiköistä riippumaton funktio $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ siten, että

$$(43) \quad \psi_i = g(x_i, y_i).$$

Jos $g(x, y)$ on yleinen kvadraattinen funktio

$$(44) \quad g(x, y) = \alpha + \beta_1 x + \beta_2 y, \\ + \beta_{11} x^2 + \beta_{22} y^2 + 2\beta_{12} xy,$$

niin muuttujien (ψ_i, x_i, y_i) aritmeettisten keskiarvojen välillä vallitsee yhtälö

$$(45) \quad \bar{\psi} = \alpha + \beta_1 \bar{x} + \beta_2 \bar{y} \\ + \beta_{11} \bar{x}^2 + \beta_{22} \bar{y}^2 + 2\beta_{12} \bar{x}\bar{y} \\ + \beta_{11} s^2(x) + \beta_{22} s^2(y) + 2\beta_{12} \text{cov}(x, y).$$

Siis on olemassa funktio $H: \mathbb{R}^5 \rightarrow \mathbb{R}$ s.e.

$$(46) \quad H \begin{bmatrix} x & y \\ s_1 & s_2 & c_3 \end{bmatrix} = \begin{aligned} &\alpha + \beta_1 x + \beta_2 y \\ &+ \beta_{11} x^2 + \beta_{22} y^2 + 2\beta_{12} xy \\ &+ \beta_{11} s_1^2 + \beta_{22} s_2^2 + 2\beta_{12} c_3 \end{aligned}$$

ja jolle pätee identtisesti

$$(47) \quad \bar{\psi} = H \begin{bmatrix} \bar{x} & \bar{y} \\ s(x) & s(y) & \text{cov}(x, y) \end{bmatrix}.$$

Teoreema 2 osoittaa, että mikroyksiköiden identtisten ja kvadraattisten käyttäytymisyhtälöiden (44) parametrit $(\alpha, \beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12})$ esiintyvät sellaisenaan vastaavien per-capita-muuttujien $\bar{x}, \bar{y}, \bar{x}^2, \bar{y}^2$ ja $\bar{x} \bar{y}$ kertoimina makroyhtälössä.

Jos mikroyhtälöihin (44) lisätään additiiviset virhetermit ϵ_i , joiden odotusarvot ovat nollia ja jotka ovat toisistaan riippumattomia (kovia oletuksia!) ilmestyy makroyhtälöön (45) vain vastaava keskiarvo $\epsilon = \frac{1}{n} \sum \epsilon_i$. Jos tilannetta on havaittu T :nä eri periodina (joille kullekin erikseen ja kaikille yhdessä) pätevät tehdyt oletukset ja joihin liittyvät virhetermit ovat toisistaan riippumattomia, voidaan mikroparametrit (jotka ovat myös makroparametreja) estimoida seuraaventyypistä havaintomatriisia käyttäen:

	$\bar{x}$	$\bar{y}$	$s(x)$	$s(y)$	$cov(x, y)$	
Periodi	1	$\bar{x}_1$	$\bar{y}_1$	$s_1(x)$	$s_2(y)$	$cov_1(x, y)$
	2	.	.	.	.	.
	.	.	.	.	.	.
	T	$\bar{x}_T$	$\bar{y}_T$	$s_T(x)$	$s_T(y)$	$cov_T(x, y)$

Koska parametrit ovat sidottuja on OLS-estimointi kätevää suorittaa yhtälöstä

$$\begin{aligned}
 (48) \quad \bar{\psi}_t &= \alpha + \beta_1 \bar{x}_t + \beta_2 \bar{y}_t \\
 &+ \beta_{11}(\bar{x}_t^2 + s_t^2(x)) + \beta_{22}(\bar{y}_t^2 + s_t^2(y)) \\
 &+ \beta_{12}(2\bar{x}_t \bar{y}_t + 2cov_t(x, y)) + \bar{\varepsilon}_t.
 \end{aligned}$$

Yhtälön estimoinnissa on varmaankin käytännössä vaikeuksia muuttujien multikollinearisuuden vuoksi, mutta periaatteessa asia on selvä. Jos $s_t^2(x)$, $s_t^2(y)$ ja $cov_t(x, y)$ poistetaan yhtälöstä saadaan ainakin α -parametrille harhainen ja epäkonsistentti estimaattori. Koska ko. varianssit ja kovarianssi eivät normaalisti pysy ajassa vakioina, vaan korreloivat $\bar{x}_t$:n ja $\bar{y}_t$:n kanssa, tulevat ainakin β_{11} , β_{22} ja β_{12} ja mahdollisesti myös β_1 ja β_2 harhaisesti ja epäkonsistentisti estimoitua, ks. Goldberger (1964, s. 196). Tämä olkoon esimerkkinä niistä vaaroista, joita liittyy kun jätetään selittäjien varianssit ja kovarianssit pois makroyhtälöistä, ts. kun makroyhtälöt spesifioidaan virheellisesti.

Yhtälöä (48) voidaan käyttää myös estimoitaessa teoreemassa 1 kuvattua muuttuvaparametrasta mallia, johon on lisätty eo. tavalla virhetermit. Saadut estimaatit ovat tällöin keskimääraisten parametrien $(\bar{\alpha}, \bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_{11}, \bar{\beta}_{22}, \bar{\beta}_{12})$ arvioita, ja niiden luotettavuus riippuu siitä, missä määrin kovarianssit $\text{cov}_t(\beta_1, x)$, $\text{cov}_t(\beta_2, y)$, $\text{cov}_t(\beta_{11}, x^2)$, $\text{cov}_t(\beta_{22}, y^2)$ ja $\text{cov}_t(\beta_{12}, xy)$ poikkeavat nolasta ja ovat selittävien muuttujien kanssa korreloituneita. Nämä tekijät ovat normaalisti lähes vakioita ja niiden suuruus riippuu vastaavien parametrien hajonnoista. Esimerkiksi

$$\begin{aligned} (49) \quad \text{cov}_t(\beta_{11}, x^2) &= \text{cov}_t(\beta_{11i}, x_i^2) \\ &= s_t^2(\beta_{11i}) s_t^2(x_i^2) r(\beta_{11i}, x_i^2) \end{aligned}$$

menee nolnaan yhdessä $s_t(\beta_{11i})$:n kanssa. Nämä termit riippuvat yhteiskunnassa vallitsevista rakenteista, joissa tapahtuu tavallisissa oloissa vain hidasta muuttumista. Näin ollen on odotettavaa, että vain $\bar{\alpha}$ -parametri tulee esitetyllä tavalla harhaisesti estimoitua, muut parametrit voidaan näin arvioida melko luotettavasti.

Jos mainitusta poisjätetyistä parametrien ja selittäjien kovarianssien suuruuksista on saatu tietoa esim. erillisten otantatutkimusten avulla, voidaan tämä tieto luonnollisesti hyödyntää lisäämällä kovarianssit mallin selittäjiin.

7. Odotusarvoja generoiva funktio

Vertailemme edellisessä luvussa esiteltyjä tuloksia Ilkan (1978) samantapaisiin mielenkiintoisiin tuloksiin, jotka ovat eräissä suhteissa johtamiamme tuloksia yleisempiä ja eräissä suhteissa niiden erikoistapauksia. Ilkka (1978) tarkastelee satunnaisvektoria $\underline{x} = (x_1, \dots, x_k)$, jonka kertymäfunktio on $F(x) = F(x_1, \dots, x_k)$, sekä reaaliarvoista funktiota $g: \mathbb{R}^k \rightarrow \mathbb{R}$. Funktio $g(x) = g(x_1, \dots, x_k)$ oletetaan ainakin kolmasti jatkuvasti derivoituvaksi, $Eg(\underline{x}) = \int g(x) dF(x)$ ja vektorit $E\underline{x}$, $E\underline{x}^2$ ja $E\underline{x}^3$ oletetaan olemassaoleviksi. $\mathbb{R}^k$ Tehtävänä on johtaa $g(\underline{x})$:n odotusarvon $Eg(\underline{x})$ ja $g(x)$:n $E\underline{x}$ -pisteessä lasketun arvon $g(E\underline{x})$ keskinäinen riippuvuus, jota kuvaa tekijä d yhtälössä

$$(50) \quad Eg(\underline{x}) = dg(E\underline{x}).$$

Tässä siis d on suhde $Eg(\underline{x})/g(E\underline{x})$, jolle Ilkka johtaa¹⁾ esityksen

$$(51) \quad d = 1 + \sum_{k=1}^K \frac{1}{2} v_k (e_{kk \cdot 0} + 2 \sum_{k=k+1}^K \beta_{ik} e_{ki \cdot 0}) + S, \text{ jossa}$$

S = jäännöstermi

$$v_k = E(\underline{x}_k - E\underline{x}_k)^2 / (E\underline{x}_k)^2$$

= muuttujan $\underline{x}_k$ variaatiosuhde

= muuttujan $\underline{x}_k$ variaatiokertoimen neliö

$$\beta_{ik} = r(\underline{x}_i, \underline{x}_k) (v_i/v_k)^{1/2}$$

= muuttujien $\underline{x}_i$ ja $\underline{x}_k$ välinen betaregressiokerroin

$$e_{ki \cdot 0} = \left(\frac{\partial^2 g(x)}{\partial \log x_k \partial \log x_i} / g(x) \right) \text{ pisteessä } \underline{x} = E\underline{x}$$

= g :n toisen kertaluvun osittaisjousto x_k :n ja x_i :n suhteen pisteessä $\underline{x} = E\underline{x}$.

1) Ilkan oletukset ovat toiset, mutta ne eivät ole riittäviä.

Rekonstruoiimme seuraavassa Ilkan todistuksen:

Kehitetään $g(x) = g(x_1, \dots, x_n)$ Taylor-sarjaksi pisteen $E_x = (E_{x_1}, \dots, E_{x_k})$ ympäristössä

$$(52) \quad g(x) = g(E_x) + \sum_{k=1}^K g_k(E_x)(x_k - E_{x_k}) \\ + \frac{1}{2!} \sum_{i=1}^K \sum_{k=1}^K g_{ik}(E_x)(x_i - E_{x_i})(x_k - E_{x_k}) \\ + R_3,$$

missä $R_3 = \frac{1}{3!} \sum_i \sum_k \sum_l g_{ikl}(\xi)(x_i - E_{x_i})(x_k - E_{x_k})(x_l - E_{x_l})$ on jäännöstermi ja ξ on sopiva pisteiden x ja E_x yhdysjanelalla oleva piste, ks. esim. Apostol (1957, s. 123-4). Tulkitsemalla x satunnaismuuttujaksi ja ottamalla odotusarvot puolittain saadaan

$$(53) \quad E g(x) = E g(E_x) + \sum_{k=1}^K g_k(E_x) E(x_k - E_{x_k}) \\ + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^K g_{ik}(E_x) E(x_i - E_{x_i})(x_k - E_{x_k}) \\ + E R_3 \\ = g(E_x) + \frac{1}{2} \sum_i \sum_k g_{ik}(E_x) \text{cov}(x_i, x_k) + E R_3.$$

Huomattakoon, että kaikki ensimmäisen asteen termit häviävät. Jakamalla suureella $g(E_x)$ saadaan

$$(54) \quad d = E g(\underline{x}) / g(E \underline{x})$$

$$\begin{aligned} &= 1 + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^K \frac{g_{ik}(E \underline{x})}{g(E \underline{x})} \frac{\text{cov}(\underline{x}_i, \underline{x}_k)}{D \underline{x}_i, D \underline{x}_k} D \underline{x}_i D \underline{x}_k \\ &\quad + E \underline{R}_3 / g(E \underline{x}) \\ &= 1 + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^K \frac{g_{ik}}{g} r(\underline{x}_i, \underline{x}_k) D \underline{x}_i D \underline{x}_k + S_2 \\ &= 1 + \frac{1}{2} \sum_{k=1}^K \left(\frac{g_{kk}}{g} D^2 \underline{x}_k + 2 \sum_{i=k+1}^K \frac{g_{ik}}{g} r(\underline{x}_i, \underline{x}_k) D \underline{x}_i D \underline{x}_k \right) + S_3 \\ &= 1 + \sum_{k=1}^K \frac{1}{2} v_k \left(\frac{g_{kk}}{g} (E \underline{x}_k)^2 + 2 \sum_{i=k+1}^K \frac{g_{ik}}{g} \frac{D \underline{x}_i D \underline{x}_k}{v_k} r(\underline{x}_i, \underline{x}_k) \right) + S_3 \\ &= 1 + \sum_{k=1}^K \frac{1}{2} v_k (e_{kk \cdot 0} + 2 \sum_{i=k+1}^K \beta_{ik} e_{ik \cdot 0}) + S_3. \end{aligned}$$

Tässä on käytetty hyväksi tulosta

$$g(\underline{x}) e_{ik \cdot 0}(\underline{x}) = \partial^2 g(\underline{x}) / \partial \log x_i \partial \log x_k = g_{ik} x_i x_k = g_{ki} v_k x_i.$$

Tämä todistaa väitteen.

Esityksessä (52) kaikki tekijät, variaatiosuhteet v_k , betaregressioker-
toimet β_{ik} ja toisen kertaluvun osittaisjoustot $e_{ki \cdot 0}$ ovat dimensiotto-
mia ("paljaita") lukuja. Tämä on sen hyvä puoli. Todistuksen välivaihee-
na saatiin seuraavat vaihtoehtoiset, joissakin tilanteissa helpommin
tulkittavat esitykset:

$$(55) \quad d = E g(\underline{x}) / g(E\underline{x})$$

$$= 1 + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^K \frac{g_{ik}(E\underline{x})}{g(E\underline{x})} D\underline{x}_i D\underline{x}_k r(\underline{x}_i, \underline{x}_k) + S_3$$

$$(56) \quad d = 1 + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^K \frac{g_{ik}(E\underline{x})}{g(E\underline{x})} \text{cov}(\underline{x}_i, \underline{x}_k) + S_3$$

$$(57) \quad E g(\underline{x}) = g(E\underline{x}) + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^K g_{ik}(E\underline{x}) \text{cov}(\underline{x}_i, \underline{x}_k) + R_3$$

Erityisesti (57) on käyttökelpoinen verrattaessa edellisessä luvussa esiteltyjä tuloksia Ilkan tuloksiin. Yhtälö (57) sanoo, etteivät yleensä $E g(\underline{x})$ ja $g(E\underline{x})$ ole yhtä suuria, vaan poikkeavat toisistaan sitä enemmän mitä suurempia muuttujien varianssit $s(\underline{x}_i)$ ja kovarianssit $\text{cov}(\underline{x}_i, \underline{x}_k)$ sekä funktion $g: \mathbb{R}^K \rightarrow \mathbb{R}$ toiset osittaisderivaatat $g_{ik}(E\underline{x}) = \partial^2 g(E\underline{x}) / \partial x_i \partial x_k$ (siis kaarevuudet) ovat. Lineaarisilla funktioilla $g_{ik}(x) = 0$ kaikilla pareilla (i, k) ja näille pätee yleisesti $E g(\underline{x}) = g(E\underline{x})$, sillä tällöin $g(x) = \sum_{k=1}^K \beta_k x_k$ ja

$$(58) \quad E g(\underline{x}) = E(\sum \beta_k x_k)$$

$$= \sum \beta_k E x_k$$

$$= g(E\underline{x}).$$

Kuten esim. Chung (1968, s. 44) osoittaa pätee konveksille g :lle

$$(59) \quad E g(\underline{x}) \leq g(E\underline{x}) \quad (\text{Jensenin epäyhtälö})$$

Vastaavasti konkaaville g :lle saadaan

$$(60) \quad E g(\underline{x}) \geq g(E \underline{x}).$$

Näin ollen asiaa on kyllä kirjallisuudessa tarkasteltu, mutta Ilkan formulaatio on yleisin tuntemani. On mahdollista, että lause on esitetty aiemminkin jossain¹⁾, sen verran suoraviivaiselta tuntuvat esim. formulaatiot (55)-(57). Sovelletaan yhtälöä (57) kvadraattiseen funktioon

$$(61) \quad g(x) = \alpha + \sum_{k=1}^K \beta_k x_k + \sum_{i=1}^K \sum_{k=1}^K \beta_{ik} x_i x_k$$

jolle siis $g_{ik}(x) = \beta_{ik}$:

$$(62) \quad E g(\underline{x}) = g(E \underline{x}) + \sum_{i=1}^K \sum_{k=1}^K \beta_{ik} \operatorname{cov}(x_i, x_k).$$

Tämä on täsmällinen tulos ilman mitään jäännöstermejä.

Erityisesti valitsemalla $K = 2$, $x_1 = x$ ja $x_2 = y$ saadaan tulokset edellisessä luvussa tarkastelluille kahden muuttujan kvadraattiselle funktiolle

$$(63) \quad g(x, y) = \alpha + \beta_1 x + \beta_2 y \\ + \beta_{11} x^2 + \beta_{22} y^2 + 2\beta_{12} xy$$

1) Lisäys 20.2.1979: Goldberger (1964, s. 123) esittää ja johtaa Ilkan soveltamalla tavalla yhtälön (57) ja soveltaa sitä satunnaismuuttujien funktioiden asymptoottisen keskiarvon ja varianssin johtamiseen.

$$\begin{aligned}
 (64) \quad E g(\underline{x}, \underline{y}) &= g(E\underline{x}, E\underline{y}) + \\
 &+ \beta_{11} D^2 \underline{x} + \beta_{22} D^2 \underline{y} + 2\beta_{12} \text{cov}(\underline{x}, \underline{y}) \\
 &= \alpha + \beta_1 E\underline{x} + \beta_2 E\underline{y} \\
 &+ \beta_{11} (E\underline{x})^2 + \beta_{22} (E\underline{y})^2 + 2\beta_{12} E\underline{x} E\underline{y} \\
 &+ \beta_{11} D^2 \underline{x} + \beta_{22} D^2 \underline{y} + 2\beta_{12} \text{cov}(\underline{x}, \underline{y}).
 \end{aligned}$$

Tämä tulos vastaa teoreeman 2 tilannetta, jossa vain odotusarvojen $E\underline{x}$ ja $E\underline{y}$ sijaan on (painottamattomat) keskiarvot $\bar{x}$ ja $\bar{y}$. Ilkan tulokset ovat edellisen luvun tuloksia oleellisesti yleisempiä siinä, että g -funktion muotoa ei ole rajoitettu esimerkiksi kvadraattiseksi. Edellisessä luvussa esitetyt tarkastelut ovat taas siinä mielessä yleisempiä, ettei mikroyksiköiden käyttäytymisyhtälöiden oletettu olevan samoja kaikille mikroyksiköille, vaan niiden parametrit saavat vaihdella mikroyksiköstä toiseen. Sen sijaan tämän luvun tarkastelut koskevat yhtä kiinteää funktiota $g(x)$, joka vastaa esim. edellisen luvun teoreeman 2 mikroyksiköstä toiseen muuttumattomaksi oletettua käyttäytymisyhtälöä $g(x, y)$.

LÄHDELUETTELO:

APOSTOL, T.M., (1957): *"Mathematical analysis"*, Addison-Wesley Publishing Company, Inc., Massachusetts.

CHIPMAN, John S., (1976): "Estimation and aggregation in econometrics: an application of the theory of generalized inverses", in *"Generalized inverses and applications"*, Academic Press, Inc., New York-San Francisco-London.

CHUNG, Kai Lai, (1968): *"A course in probability theory"*, Harcourt, Brace & World, Inc., New York etc.

CRAMER, J.S., (1971): *"Empirical econometrics"*, North-Holland Publishing Company, Amsterdam-London.

GOLDBERGER, Arthur S., (1964): *"Econometric theory"*, John Wiley & Sons, Inc., New York etc.

EDGREN, C., TURKKILA, J. ja VARTIA, Y., (1978): "Tuloerotuksen analysoinnin matemaattisista ongelmista", Keskusteluaiheita No 17, Elinkeinoelämän Tutkimuslaitos, Helsinki.

ILKKA, Tuomo I., (1978): "Odotusarvoja generoiva funktio ja sen tärkeimmät perussovellutukset", tutkimuskäsikirjoitus.

KLEIN, Lawrence R., (1946): "Remarks on the theory of aggregation", *Econometrica*, 14, 303-312.

MALINVAUD, Edmond, (1954): "Aggregation problems in input-output models", in *The Structural Interdependence of the Economy*, Proceedings of an International Conference on Input-Output Analysis, Varenna, edited by Tibor Barna, John Wiley & Sons, Inc., New York.

OJAMO, Markku ja HÖYTYLÄINEN, Raimo, (1973): "Tavaratuonnin pääryhmien regressioanalyttinen tarkastelu", tilastotieteen proseminaariesitelmä syksyllä 1973 ass. Tavailan ryhmässä.

RANTA, Jaakko, (1973): "Tavaratuonnin pääryhmien regressioanalyttinen tarkastelu", tilastotieteen proseminaariesitelmä 28.6.1973 ass. Linnilän ryhmässä.

SAARINEN, Ari ja LAESVUORI, Arto, (1973): "Tavaraviennin pääryhmien regressioanalyttinen tarkastelu vuosina 1949-1970", tilastotieteen proseminaariesitelmä 4.12.1973 ass. Linnilän ryhmässä.

THEIL, H., (1965): *"Linear aggregation of economic relations"*, North-Holland Publishing Company, Amsterdam.