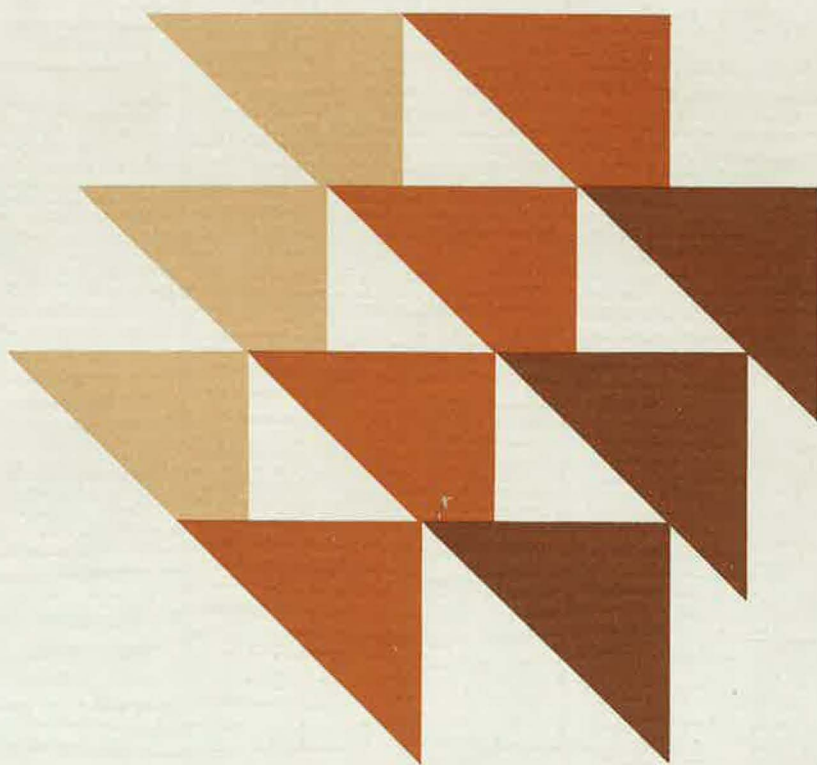


ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS

REIJA LILJA JA YRJÖ VARTIA
KOULUTUSAIKA KOTITALOUKSIEN
TULOEROJEN SELITYSTEKIJÄNÄ



B 25

ELINKEINOELÄMÄN TUTKIMUSLAITOS

THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
ETLA



Lönnrotinkatu 4 B, SF-00120 Helsinki 12 ☎ (90) 601 322

KANNATUSYHDISTYKSEN HALLITUS COUNCIL OF SUPERVISORS

Asko Tarkka pj., president
Veikko Makkonen varapj., vice president
Gay Ehrnrooth
Pekka Herlin
Ahti Hirvonen
Stig H. Hästö
Seppo Konttinen
Jaakko Lassila
Sakari T. Lehto
Matti Pekkanen
Filip Pettersson
Pentti Somerto
Eljas Sukselainen
Heikki Tavela
Heikki Tuominen

VARAMIEHET DEPUTY MEMBERS

Tor Stolpe
Matti Hurme
Fredrik Castrén
Jaakko Ihamuotila
Paavo Laitinen
Juhani Ristimäki
Helge Laakso
Matti L. Aho
Yrjö Pessi
Harri Malmberg
Olli Ikkala
Matti Korhonen
Kalevi Kauniskangas
Heikki Bachmann
Viljo Koponen

JOHTOKUNTA BOARD OF MANAGEMENT

Matti Korhonen pj., chairman
Kimmo Kara varapj., vice chairman
Paavo Grönlund
Kari Jalas
Viljo Koponen
Rauno Niinimäki
Tauno Ranta

Laitoksen henkilökunta on esitetty takakannen sisäsivulla.
Staff of the institute is presented inside the back cover.

Reija Lilja ja Yrjö Vartia

Koulutusaika
kotitalouksien tuloerojen
selitystekijänä

Schooling
as an explanatory factor
of the income differences in households

Helsinki 1980

ESIPUHE

Kotitalouksien tuloeroihin vaikuttavat mm. palkkaerot, verotus ja tulonsiirrot sekä työssäoloajan ja työssäkäyvien henkilöiden lukumäärän vaihtelu. Eri ammattiryhmien palkkaerot johtuvat valtaosalta työn laadun, rasituksen, arvostuksen ja mm. tarvittavan ammattipätevyyden ja koulutuksen eroista. Nyt julkaistavassa tutkimuksessa on pohdittu koulutuksen tuloaikautuksen tilastollisen mittaamisen ongelmia ja pyritty vuoden 1971 kotitaloustiedustelun aineistosta poimitun 1000 kotitalouden otoksen perusteella arvioimaan päämiehen koulutuksen merkitystä kotitalouksien tuloerojen selitystekijänä.

Kotitaloustiedustelun aineiston käyttöön ongelman tutkimisessa liittyy tilastollisia ja tulkinnallisia vaikeuksia, joiden vaikutukset on pyritty eliminoimaan kovarianssianalyysin avulla. Tutkimuksessa onkin erityisesti pohdittu johtopäätösten tekemistä yhteiskuntatieteiden osalta tavanomaisessa tutkimustilanteessa, jossa käytettävissä oleva tilastoaineisto on ei-kokeellista ja tutkimustavoitteen kannalta usein puutteellista.

Tutkimuksen on tehneet KTM Reija Lilja ja apul.prof. Yrjö Vartia. Puhtaaksi kirjoituksen ovat suorittaneet Raili Kosonen ja Tuula Ratapalo sekä kuviot on piirtänyt Arja Virtanen.

Helsingissä 3.6.1980

Tauno Ranta

SISÄLLYSLUETTELO

	sivu
1. JOHDANTO	7
1.1. Tutkimuksen tausta	7
1.2. Tutkimuksen tavoite ja eteneminen	8
2. TULOJEN JAKAUTUMINEN JA KOULUTUKSEN TULOVAIKUTUS	11
2.1. Tulojen jakaumamalleja	11
2.1.1. Pareto jakauma	11
2.1.2. Normaalinen jakauma	13
2.1.3. Lognormaalinen jakauma	14
2.1.4. Yleistetty F-jakauma	16
2.2. Koulutuksen tulovaikutus	19
2.2.1. Muu tietämys ja koulutuksen tulovaikutus	21
2.2.2. Koulutus, lahjakkuus ja sosioekonominen tausta tulon selittäjinä	23
2.2.3. Inhimillisen pääoman teoria - kritiikkiä	28
3. HAVAINTOAINEISTON HANKINTA	31
3.1. Yleistä	31
3.2. Kotitaloustiedustelu 1971	34
3.3. Rakennetiedosto	36
3.3.1. Määritelmiä ja luokituksia	37
3.4. Tilastoaineisto ja tutkimusongelman rajaus	39
4. TUTKIMUKSEN KYSYMYKSENASETELU	40
4.1. Tieteellisen päättelyn problematiikka	40
4.1.1. Selittävä tutkimus ja ei-kokeellinen havainto- materiaali	47
4.2. Tutkimusongelman tarkastelumahdollisuuksia	56

	sivu
5. AINEISTON TILASTOLLINEN ANALYYSI JA MALLIT	70
5.1. Regressio- ja kovarianssianalyysi	70
5.2. Aineiston kuvaus	81
5.3. Estimoitavat mallit	90
6. MALLIEN ESTIMOINTITULOKSET	98
6.1. Perusmallit	98
6.2. Parannetut mallit	113
6.3. Kotitalouden päämiehen koulutuksen tulovaikutus	121
7. LOPPUTOTEAMUKSIA	126

LIITTEET

LIITE 1: Tilastokeskuksen rakennetiedoston tietuekuvaus	129
LIITE 2: Rakennetiedoston luokitteluja numerokoodeineen	130
LIITE 3: Indikaattorimuuttujien muodostaminen	131
LIITE 4: Muuttujien keskiarvot ja hajonnat	133
LIITE 5: Tietokonetunnuksen ja tutkimuksessa käytetyn muuttujatunnuksen vastaavuus	134
LIITE 6: Muuttujien kovarianssimatriisi	135
LIITE 7: Regressiomallien F-arvo, virhetermin hajonnan esti- maatti (s), yhteiskorrelaatiokertoimen neliö (R^2), selittäjien regressiokertoimet ja t-arvot	138
LIITE 8: Regressiomallit graafisesti kuvattuna	141
LIITE 9: Mallin (6) muuttujien korrelaatiomatriisi	142
LIITE 10: Mallin (7) muuttujien korrelaatiomatriisi	143

1. JOHDANTO

1.1. Tutkimuksen tausta

Tuloerojen suuruus ja niihin vaikuttavat tekijät ovat jatkuvasti kiinnostaneet taloustieteilijöitä. T. W. Schultzin toimesta ryhdyttiin 1960-luvulla tutkimaan erityisesti koulutuksen tulovaikutusta, ks. Becker (1964, s. 1). Alkuvaiheessa keskityttiin koulutuksen ja tulojen välisten teoreettisten syy-seuraus-mallien rakentamiseen. Nämä pohdiskelut voidaan tulkita ajatuskokeiksi, joita ei asian luonteen vuoksi kuitenkaan voida käytännössä suorittaa. Tästä huolimatta johtopäätöksiä tehtiin ei-kokeellisen havaintomateriaalin perusteella, eikä useinkaan huomattu, miten vaikeita tieteellisen päättelyn ongelmia johtopäätösten tekemiseen sisältyi.

Tämä tutkimus pohjautuu Reija Liljan (1978) HKKK:n pro gradu-tutkielmaan "Koulutus kotitalouksien tulojen selittäjänä", johon liittyviä kysymyksiä käsillä olevan tutkimuksen tekijät jo silloin yhdessä pohtivat. Asian jatkokehittely tuntui hyödylliseltä, sillä valitettavan vähän yhä kiinnitetään huomiota siihen, miten ei-kokeellinen havaintoaineisto vaikuttaa tieteellisten johtopäätösten tekemiseen. Lähtökohtana ollut Liljan tutkimusta laajemmin on pohdittu selittävän tutkimuksen suorittamismahdollisuuksia ei-kokeellisen havaintomateriaalin avulla yleisesti ja erityisesti koulutuksen tulovaikutuksia mitattaessa. Tutkimuksen suorittamista on auttanut mahdollisuus käyttää Tilastokeskuksen suostumuksella KTT Liisa Uusitalon väitöskirjatyötään varten hankkimaansa vuoden 1971 kotitaloustiedustelun rakennetiedostoa, josta tutkimuksen 1000 kotitalouden otos poimittiin. VTT Matti Vireniä kiitämme kriittisistä huomautuksista tutkimuksen viimeistelyvaiheessa.

1.2. Tutkimuksen tavoite ja eteneminen

Tutkimuksen tarkoituksena on mahdollisimman luotettavasti vastata kysymykseen, missä määrin koulutus selittää tulojen vaihteluja. Käytettävissä oleva havaintoaineisto rajaa tutkimusongelman seuraavan kysymyksen muotoon: miten kotitalouden päämiehen koulutus selittää kotitalouden käytettävissä olevan tulon vaihteluja? Tarkempi tutkimusongelman rajaus tapahtuu havaintoaineiston esittelyn jälkeen luvuissa 3 ja 4.

Kappaleessa 2.1. esitellään eräitä käyttökelpoisia tulojen jakaumamalleja. Erilaisiin olosuhteisiin ja eri ajanjaksoihin liittyvistä havainnoista muodostetut tulojen jakaumat ovat hämmästyttävän säännönmukaisia, mitä ei voi jättää huomiotta etsittäessä tulojakaumassa esiintyville tuloeroille järkeviä selityksiä. Suhteelliset tuloerot näyttävät syntyvän yhteiskuntien ja ajanjaksojen eroista suhteellisen riippumattoman prosessin tuotteena, ja tarkoitus on seuraavassa tutkia, minkälainen asema koulutuksella on tässä prosessissa.

Kappaleessa 2.2. siirrytään tarkastelemaan talousteorian puitteissa, miten koulutus vaikuttaa tuloihin. Inhimillisen pääomateorian perustulokset esitellään lyhyesti ja eräs sen periaatteiden mukainen koulutuksen tulomalli kuvataan. Lisäksi selvitetään, miten mm. lahjakkuuden ja muiden merkittävien selittäjien puuttuminen tulomallista saattaa vaikeuttaa koulutuksen tulovaikutuksen arviointia. Käytettävissä oleva havaintoaineisto rajoituksineen estää tässä tutkimuksessa teoreettisten mallien oikeellisuuden testaamisen. Inhimillisen pääomateorian mukaan on mm. perusteltua selittää logaritmita tuloa koulutusvuosien, työkokemuksen, lahjakkuuden ja sosioekonomisten tekijöiden avulla.

Tutkimuksen empiirinen osa alkaa havaintoaineiston hankinnan kuvauksesta luvussa 3. Havaintoaineiston hankinta on oleellinen osa empiirisestä tutkimustyöstä. Havaintoaineiston hankinta on syytä kuvata huolellisesti, sillä tutkimusongelma on loppujen lopuksi rajattava sellaiseen muotoon, jolle käytettävissä olevan aineiston avulla voidaan vastausta etsiä.

Ei-kokeellisissa tutkimuksissa harvoin selvästi esitetään niihin välttämättä liittyviä perusvaikeuksia, jotka syntyvät tutkittavaan muuttujaan vaikuttavista useista kontrolloimattomista tekijöistä. Näitä ongelmia ei voida ratkaista esim. havaintoaineistoa suurentamalla, menetelmäkirjojen referointia laajentamalla eikä talousteoriaan vetoamalla. Luvussa 4 tarkastellaan yleisesti tieteellisen päättelyn problematiikkaa ja erityisesti ei-kokeellisen havaintomateriaalin vaikutusta tutkimusongelmaan, joka olisi periaatteessa yksinkertaista selvittää kokeellisesti. Luvussa ehdotetaan ratkaisuksi tutkimusstrategiaa, jossa ei-kokeellisten havaintojen perusteella yritetään päätellä, mitä vastaavassa kontrolloidussa kokeessa olisi tapahtunut.

Luvussa 5 esitellään käytettyjä tilastollisia metodeja, tunnuslukuja ja testejä sekä kuvataan erilaisin reunajakaumin käytettyä tilastoaineistoa. Myös estimoitavat mallit esitellään tässä luvussa. Kovarianssianalyysin kuvaus on myös paikallaan, koska sitä näkee tämän laatuissa tutkimuksissa vain harvoin käytettävän huolimatta sen oivasta soveltuvuudesta ei-kokeellisen havaintomateriaalin analysointiin.

Luvussa 6 esitetään varsinaiset estimointitulokset. Edellisessä luvussa esitettyjen mallien antamaa tietoa on pyritty lisäksi selkiinnyttämään kahden parannetun malliversion avulla. Vaikka päähuomio on kiinnitetty

kotitalouden päämiehen koulutuksen tulovaikutukseen, myös muiden malleissa mukana olevien muuttujien saamia kertoimia on pyritty tulkitsemaan, ovathan nekin kotitalouden käytettävissä olevan tulon selittäjiä ja osaltaan siten paljastavat kotitalouden päämiehen koulutusmuuttujan merkityksen kotitalouden tulojen selittäjänä.

Lopputoteamuksissa esitetään lyhyesti tutkimuksen päätulos ja annetaan viitteitä jatkotutkimuksen suunnalle.

2. TULOJEN JAKAUTUMINEN JA KOULUTUKSEN TULOVAIKUTUS

2.1. Tulojen jakaumamalleja

Tulojen jakautuminen on taloudellinen ilmiö, jota on tarkasteltu tilastollisesti ja kuvattu erilaisilla matemaattisilla malleilla jo 1800-luvulla. Mitään yhtä, yleisesti hyväksyttyä tulojen jakaumamallia ei ole esitetty, vaan useita keskenään kilpailevia kuvauksia tulojen jakautumisesta on käytössä. Seuraavassa tarkastellaan lyhyesti eräitä ehdotettuja tulojen jakaumamalleja.

2.1.1. Pareto jakauma

Ensimmäiset järjestelmälliset havainnot tulojen jakauman luonteesta teki Pareto (1897). Hän esitti, että tulon y ja tätä tai suurempaa tuloa ansaitsevien tulonsaajien lukumäärien välillä on tietty suhde. Merkitään $N(y)$ = niiden tulonsaajien lukumäärä, joiden tulot ovat tuloa y korkeammat, siis $N(y)$ on laskeva kumulatiivinen frekvenssi. Pareton laiksi kutsuttu riippuvuus on seuraavaa muotoa:

$$(1) \quad N(y) = \beta y^{-\alpha},$$

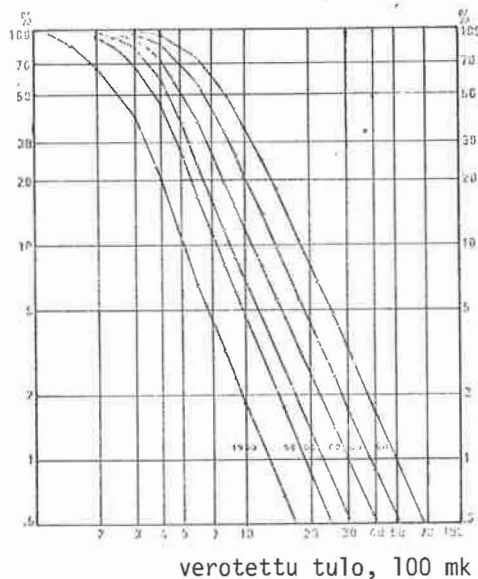
missä β ja α ovat positiivisia vakioita. Esitetty yhtälö merkitsee seuraavaa: kun tulot lisääntyvät yhdellä prosenttiyksiköllä, niin vähintään annetun tulon y ansaitsevien tulonsaajien lukumäärä $N(y)$ vähenee α :n osoitaman prosenttiosuuden, ks. Pen (1971, s. 235). Mitä suurempi α on, sitä nopeammin tulonsaajien lukumäärä laskee valittaessa korkeampi tulo-taso ja sitä pienempi on tulonjaon epätasaisuus. Parametri α on siten tulonjaon tasaisuuden ja esim. $1/\alpha$ epätasaisuuden mittari.

Yksinkertaisin tapa tutkia Pareton lain soveltuvuutta tulonjakaumamallina on muuntaa yhtälö (1) seuraavaan logaritmiseen muotoon:

$$(2) \quad \log N(y) = \log \beta - \alpha \log y$$

Jos laskevat kumulatiiviset frekvenssit $N(y)$ ja tulot y merkitään kaksoislogaritmpaperille, pitäisi tuloksena olla alaspäin laskeutuva suora. Näin ei käytännössä täsmällisesti käy, mutta lineaarisuus näyttää pitävän paikkansa suurella tarkkuudella suurilla tuloilla. Cramer (1971, s. 51-58) on tutkinut useiden empiiristen jakaumien käyttäytymistä ja havainnut, että Pareton laki näyttää kuvaavan tulojen jakautumista tulojakaumien yläpäässä. Suomessa ainakin verotettavan tulon tapauksessa Pareto-jakauma kuvaa erinomaisesti vuonna 1953-68 noin 30 %:n yläpäätä tulojakaumassa, ks. Vartia-Vartia (1972, s. 76). Seuraava kuvio havainnollistaa esitettyä säännönmukaisuutta.

Kuvio 1. Verotetun tulo Pareto-esitys v. 1953-1968: laskevat kertymäfunktiot log-log-paperilla



2.1.2. Normaalinen jakauma

Eräät taloustieteilijät ovat esittäneet, että tutkittaessa riittävän homogeenisia ihmisryhmiä, niihin kuuluvien henkilöiden tulojen jakauma on tai sen pitäisi olla normaalinen, ks. Lydall (1968, s. 25-36). Perusoletuksena on tällöin, että ihmisten lahjakkuus (kyvykkyys, älykkyys) on normaalisesti jakautunut ja että tulot riippuvat lahjakkuudesta. Johtopäätös tulojen normaalista jakautumisesta on tehty viittaamalla esimerkiksi Edgeworth'in teoreemaan: jos jokin muuttuva tekijä noudattaa normalisuuden lakia, niin tuon tekijän funktio noudattaa myös (likimain, jos hajonta on pieni) normalisuuden lakia.

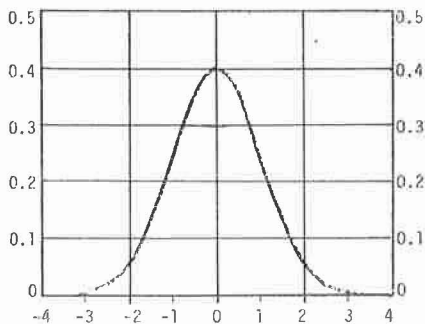
Standardisoidun normaalisen jakauman, $N(0,1)$, tiheys- ja kertymäfunktio ovat seuraavaa muotoa:

$$(3) \quad \rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

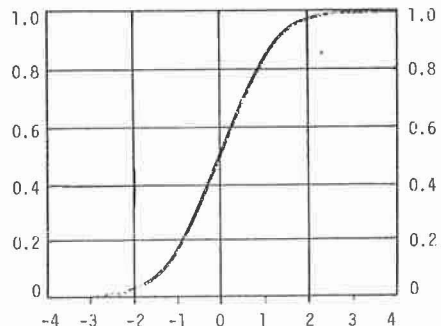
$$(4) \quad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt = N(x; 0,1).$$

Funktioiden kuvaajat ovat seuraavat:

Kuvio 2: $N(0,1)$ -jakauman tiheysfunktio



Kuvio 3: $N(0,1)$ -jakauman kertymäfunktio



Normaalisen tulojakaumateorian yhteensopivuus havaintoihin on todettu yleensä epätydyttäväksi. Teorian perustana oleva oletus lahjakkuuden normaalista jakautumisesta ei välttämättä pidä paikkaansa. Lahjakkuus riippuu useista tekijöistä; geneettisestä perimästä, kotiympäristöstä, koulutuksesta, iästä, kokemuksesta ja monesta muusta seikasta. Näiden tekijöiden normaalille jakautumiselle ei ole vahvoja perusteita. Keskeistä raja-arvolauseetta ei myöskään voida näihin sellaisenaan soveltaa, koska em. tekijät eivät ole toisistaan riippumattomia vaan pikemminkin positiivisesti korreloituneita. Myöskään lahjakkuuden vaikutus tuloihin ei liene lineaarinen. Empiiriset tulojakaumat ovat yleensä selvästi oikealle vinoja, kun normaalijakaumamallin mukaan niiden tulisi olla symmetrisiä.

2.1.3. Lognormaalinen jakauma

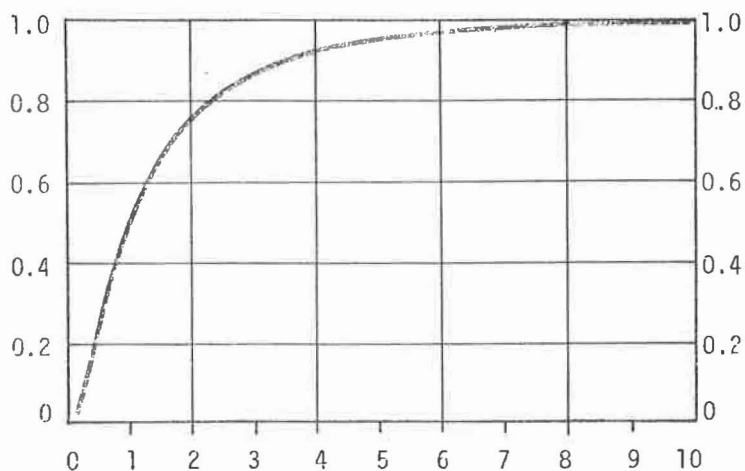
Selvästi parempi yhteensopivuus teorian ja havaintojen välillä saavutetaan sovittamalla normaalijakaumaa tulojen y sijaan logaritmisille tuloille $\log y$. Tämä merkitsee, että aritmeettisille tuloille y käytetään log-normaalista jakaumamallia. Ensimmäisen huomattavan lognormaalisen jakauman sovellutuksen tulonjaon teoriaan on suorittanut Gibrat (1931). Hän oletti, että tulot ovat muodostuneet monien toisistaan riippumattomien positiivisten satunnaistekijöiden tulona, Lydall (1968, s. 37). Tätä kutsutaan Gibratin suhteellisen vaikutuksen laiksi. Kun lakiin sovelletaan keskeistä raja-arvolauseetta, johtaa se tulojen log-normaaliseen jakaumaan rajajakaumana. Keskeisen raja-arvolauseen mukaan muuttujan $Y = \sum_{i=1}^n X_i$ (jossa X_i :t ovat tietyt yleiset ehdot toteuttavia riippumattomia satunnaismuuttujia) jakauma lähenee normaalista jakaumaa,

kun $n \rightarrow \infty$. Tästä seuraa: jos $Y = \prod_{i=1}^n X_i$, niin muuttujan $\log Y = \sum \log X_i$ jakauma lähenee normaalijakaumaa eli Y :n jakauma lähenee tietyillä yleisluontoisilla ehdoilla lognormaalista jakaumaa, kun $n \rightarrow \infty$. Näihin yleisluontoisiin ehtoihin kuuluu mm., että X_i :t ovat positiiviarvoisia ja riippumattomia muuttujia, joiden jakaumat saavat poiketa toisistaan huomattavastikin. Minkä tahansa lognormaalisesti parametrein μ, σ^2 jakauneen muuttujan z kertymäfunktioita merkitään symbolilla $\Lambda(z; \mu, \sigma^2)$ (Cramer (1973, s. 27-28)) ja se johdetaan seuraavasti:

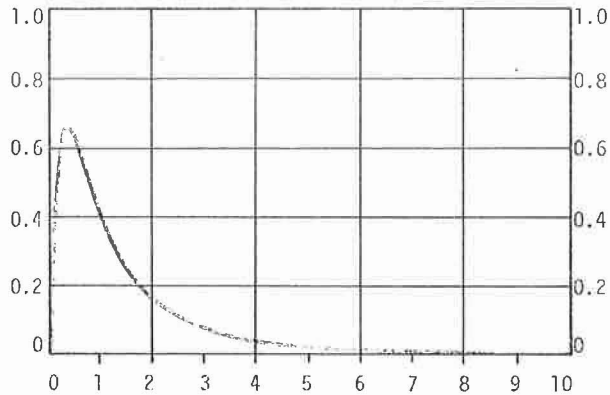
$$\Lambda(z; \mu, \sigma^2) = N(\log z; \mu, \sigma^2) = N\left\{\frac{\log z - \mu}{\sigma}; 0, 1\right\}$$

Standardoidun lognormaalisen jakauman kertymäfunktio $\Lambda(z; 0, 1) = N(\log z; 0, 1)$ ja tiheysfunktio on esitetty seuraavissa kuvioissa.

Kuvio 4: Lognormaalinen kertymäfunktio $\Lambda(z; 0, 1)$



Kuvio 5: Lognormaalinen tiheysfunktio $\frac{d\Lambda(z; 0,1)}{dz}$



Lognormaalisen jakauman tiheysfunktio kuviossa 5 on selvästi oikealle vino jakauma, esim. jakauman moodi on selvästi sen keskiarvoa pienempi.

Cramer (1971, s. 68-75) esitti tutkiessaan empiiristen tulojen jakautumista tulohavainnot ns. logaritmisella todennäköisyyspaperilla, joka oli laadittu siten, että lognormaalisesta käyrästä tällä paperilla tulee suora. Hän havaitsi, että empiiristen tulojakaumien kertymäfunktioit tulojakauman log-normaalisessa esityksessä olivat lievästi S-muotoisia ollen ainoastaan tuloluokkien keskivaiheilla suoraa. Vaikka log-normaalista jakaumaa voidaan käyttää (toisin kuin Pareto-jakaumaa) koko tulojakauman kuvaamiseen näyttää se toimivan parhaiten kuvattaessa keski suurien tulojen jakautumista.

2.1.4. Yleistetty F-jakauma

Pareto-jakauma näyttää kuvaavan suurella tarkkuudella tulojakauman yläpään käyttäytymistä, kun taas lognormaalinen jakauma toimii tyydyttävänä approksimaationa tulon koko variaatioalueella. Kirjallisuudessa on

esitetty useita uusia jakaumamalleja, joihin on yritetty sisällyttää Pareto- ja lognormaalisen jakauman hyviä piirteitä. Mikään näistä mutkikkaammista jakaumamalleista ei ole saanut osakseen yleistä hyväksyntää esim. soveltajien piirissä. Uusimpia tulojen jakaumamalleja on Fisherin tunnettuun F-jakaumaan perustuva ns. yleistetty F-jakauma, ks. Vartia-Vartia (1978). Tämä jakauma on alhaisten tulojen ja keskitulojen osalta lognormaalisen jakauman läheinen approksimaatio ja korkeiden tulojen osalta Pareto-jakauman approksimaatio. Yleistetty F-jakauma, $F(c,A,m,n)$, on muuttujan $\underline{x} = \underline{AF}^c$ jakauma, missä A ja c ovat positiivisia parametreja ja F noudattaa ns. Fisherin F-jakaumaa, $F(m,n)$. $F(c,A,m,n)$:n tiheysfunktio on seuraavaa tyyppiä:

$$(5) \quad f_{\underline{x}}(x) = \text{Vakio} \cdot \left(\frac{x}{A}\right)^{\frac{m}{2c} - 1} \left[1 + \frac{m}{n} \left(\frac{x}{A}\right)^{\frac{1}{c} \frac{m+n}{2}}\right], \quad x > 0$$

Yleistetty F-jakauma $F(c,A,m,n)$ sisältää erikoistapauksenaan useita ehdotettuja malleja. Esimerkiksi $F(1,A,m,\infty) = \text{Gamma}\left(\frac{m}{2A}, \frac{m}{2}\right)$, joten Salemin ja Mountin (1974) ehdottama gammajakauma on yleistetyn F-jakauman erikoistapaus, koska $\text{Gamma}(\lambda, \mu) = F\left(1, \frac{\mu}{\lambda}, 2\mu, \infty\right)$. Myös artikkelissa Ord et al (1979) ehdotettu ns. käänteinen gammajakauma $\text{Gamma}^{-1}(\lambda, \mu)$ on $F(c,A,m,n)$:n erikoistapaus ja $F(1/a_2, (1/a_1 a_3)^{1/a_2}, 2, 2a_3)$ on Singh'in ja Maddalan (1976) ehdottama tulojakauma vain toisella merkintätavalla esitettynä. $F(c,A,m,n)$ lähenee myös lognormaalista jakaumaa (μ, σ^2) , esim. kun $m=n$ ja $m \rightarrow \infty$, sekä $A = e^\mu$ ja $c = \frac{1}{2}\sqrt{m\sigma^2}$. Jos $n = 2\alpha c$ ja $c \rightarrow 0$, approksimoi $F(c,A,m,2\alpha c)$:n yläpää mielivaltaisen tarkasti Pareto (y_0, α) -jakaumaa.

Vartia-Vartia (1978) tarkastelevat erityisesti ns. skaalattua F-jakaumaa, $F(1,A,m,n)$, joka on muuttujan $\underline{x} = A\underline{F}$ jakauma. Myös $F(1,A,m,n)$ on hyvin joustava jakauma sisältäen esimerkiksi gammajakauman erikoistapauksena. Todennäköisyyspaperille piirrettynä skaalatun F-jakauman $F(1,A,m,n)$ kertymäfunktio on S-muotoinen ja sikäli vastaa hyvin Crämerin (1971) havaitsemien empiiristen tulojakaumien muotoa. Voidaankin sanoa, että F-jakaumamalli on vähintään yhtä joustava, kuin ne mallit, jotka se sisältää.

2.2. Koulutuksen tulovaikutus

Analysoitaessa koulutusinvestointien vaikutusta ansiotuloihin tehdään teoreettisessa tarkastelussa tavallisesti yksinkertaistavia alkuoletuksia. Esimerkkinä tarkastellaan seuraavassa inhimillistä pääomateoriaa edustavan Mincerin (1974) esittämää erästä yksinkertaista tulomallia. Yksityiskohtaisia tulomallitarkasteluja ovat esittäneet myös mm. Becker (1964), Chiswick (1974) ja Tinbergen (1975).

Alkuoletuksena on, että koulutuksen päätyttyä ei tapahdu enää sellaista tietomäärän ja taitojen lisäystä, joka voisi vaikuttaa yksilön tuottavuuteen ja sitä kautta tuloihin. Lisäksi oletetaan, että yksilön tulovirta on vakio koko tulonansaitsemisajan. Tämä ei vielä riitä, vaan myös taloudelliset muutokset, jotka vaikuttavat yksilön tuottavuuteen ja ansioihin, jätetään analyysin ulkopuolelle. Seuraavassa tutkitaan koulutusinvestointien vaikutuksia vuosiansioihin. "Investointien arvonalennus" oletetaan koulutusajalta nollaksi, unohtamista ei siis tällöin tapahdu. Tulonansaitsemisajaksi asetetaan koulutuksesta riippumaton vakio. Tällöin korkeasti koulutetut henkilöt siirtyvät pois työelämästä keskimääräistä myöhemmin. Mincer (1974) tarkastelee myös tapausta, jossa jokainen lisäkoulutusvuosi vähentää ansaitsemisaikaa vastaavasti vuodella. Kun tulonansaitsemisaika on pitkä ei ko. tarkastelu oleellisesti poikkea nyt esiteltävästä mallista. Merkittäväntä on koulutuksesta johtuva tulojen lykkääntyminen: tulojen ansaitsemisen siirtyminen vuodella eteenpäin on huomattavasti tärkeämpää kuin neljän tai viiden vuosikymmenen kuluttua tapahtuvan tulojen vähenemisen nykyarvo.

Käytetään seuraavia merkintöjä:

- n = työssäkäyntiajan pituus vuosissa
 Y_s = s vuotta koulutusta saaneen henkilön vuositulot
 V_s = henkilön elinikäisten tulojen nykyarvo koulutuksen alussa
 r = eo. tuloa laskettaessa käytetty diskonttauskerroin
 t = historiallinen aika vuosissa ($t \geq 0$)
 d = koulutusvuosien ero kahden henkilön välillä

Tutkitaan koulutuksen vaikutusta tuloihin seuraavan mallin avulla.

Kun henkilö saa koulutusta s vuotta ja tämän jälkeen (työssäoloajasta riippumattoman) vuositulon Y_s , niin elinikäistulojen nykyarvo koulutuksen aloitushetkellä $t = 0$ on

$$(1) \quad V_s = Y_s \int_0^{n+s} e^{-rt} dt = \frac{Y_s}{r} e^{-rs} (1 - e^{-rn}),$$

kun kysymyksessä on jatkuva diskonttausprosessi. Vastaavasti $s-d$ vuotta koulutusta saaneen henkilön elinikäisten tulojen nykyarvo hetkellä $t = 0$ on

$$(2) \quad V_{s-d} = \frac{Y_{s-d}}{r} (1 - e^{-rn}) e^{-r(s-d)}.$$

Merkitään s vuotta koulutusta saaneen henkilön vuositulon Y_s ja $s-d$ vuotta koulutusta saaneen henkilön vuositulon Y_{s-d} suhdetta symbolilla $k_{s,s-d} = Y_{s-d}$. Elinikäistulon teoriassa ehto koulutuksen kannattavuudelle on, että koulutuksen alkaessa elinikäistulojen nykyarvoille pätee $V_s \geq V_{s-d}$.

Yhtäsuuruuden vallitessa (tämä on rajatapaus, jossa koulutus juuri tulee kannattavaksi) pätee

$$(3) \quad k_{s,s-d} = \frac{Y_s}{Y_{s-d}} = \frac{e^{-r(s-d)}}{e^{-rs}} = e^{rd}.$$

Tässä mallissa tulojen suhde ei lainkaan riipu koulutustasosta s , eikä myöskään työssäkäyntiajasta n . Erityisesti jos $s-d = 0$ niin tulosuhde s vuotta koulutetun ja kouluttamattoman henkilön välillä on $k_{s,0} = Y_s/Y_0 = e^{rs}$.

Tämä voidaan esittää logaritmuodossa seuraavasti:

$$(4) \quad \ln Y_s = \ln Y_0 + rs$$

Tämä yhtälö kertoo, että tulojen suhteellinen lisääntyminen on tarkastellussa tapauksessa suoraan verrannollinen koulutusaikaan s ja koulutuksen tuottoasteeseen r nähden. Positiivisen tuottoasteen r tapauksessa koulutuksella on tuloja lisäävä vaikutus, $Y_s > Y_0$. Yhtälö (4) on kaikkien yksinkertaisin inhimillisen pääomafunktion esitys. Muuttuja Y_s kuvaa niiden henkilöiden tulotasoa, jotka eivät lisää henkistä pääomaansa s vuoden koulutuksensa jälkeen.

2.2.1. Muu tietämys ja koulutuksen tulovaikutus

Useimmat henkilöt jatkavat kykyjensä ja tulonansaitsemiskapasiteettinsa kehittämistä varsinaisen ammattikoulutuksen jälkeen. He ts. lisäävät

"inhimillistä pääomaansa" jatkuvasti iän myötä. Tästä johtuvan tulon kasvun voidaan tulkita olevan seurausta nettoinvestointitoiminnasta, joka jatkuu koulutuksen jälkeen. Mincerin (1974) edustaman inhimillisen pääoman teorian mukaan (1) Kyseiset investoinnit ovat hajaantuneet ajassa, koska niiden rajakustannukset nousevat kunakin periodina. (2) Investoinnit vähenevät jatkuvasti ajan myötä, koska niiden tuottavuus laskee ja rajakustannukset nousevat.

Investointien tuottavuus on sitä korkeampi, mitä pienempi unohtamisaste ja mitä pidempi odotettu työssäkäyntiaika on. Rajakustannukset ovat sitä matalampia mitä oppimiskykyisempi henkilö on. Koska ansiot tulkitaan kumulatiivisten nettoinvestointien tuotokseksi, ne myös kasvavat vähenevällä nopeudella työssäoloajan kuluessa ja vähenevät, kun nettoinvestoinnit tulevat negatiivisiksi. Koulutuksen tulomallin laajentaminen yksityiskohtaisemmaksi tulofunktioksi edellyttää lineaarisen koulutustermin lisäksi koulutuksen jälkeisiä investointeja kuvaavaa termiä. Pulmana on, miten nuo investoinnit voitaisiin ilmaista ajan funktiona. Inhimillisen pääomateorian mukaan koulutuksen jälkeisten investointien täytyy laskea työssäoloajan kuluessa. Mm. seuraavia yksinkertaisia malleja on kirjallisuudessa esitetty: (1) lineaarinen koulutusinvestointien lasku merkitsee parabolista työkokoemusfunktiota (kuvaa koulutuksen jälkeisiä investointeja) ja (2) eksponentiaalinen investointiasteen lasku johtaa Gompertz-tyyppiseen tulofunktioon. Esitetyt tulofunktiot ovat seuraavaa muotoa:

$$(5) \quad \text{Parabolinen: } \ln E_{s,h} = \ln E_0 + r_s s + r_p k_0 h - \frac{r_p k_0}{2T} h^2$$

$$(6) \quad \text{Gompertz: } \ln E_{s,h} = \ln E_0 + r_s s + \frac{r_p k_0}{\beta} (1 - e^{-\beta h}) ,$$

missä $E_{s,h}$ on s vuotta koulutetun ja h vuotta työskennelleen henkilön bruttovuosiansiot vuonna $t = s+h$. Bruttoansiot sisältävät investointikulut. Parametrit $r_s \geq 0$ ja $r_p \geq 0$ ovat koulutuksen ja koulutuksen jälkeisten investointien tuottoasteita, $k_0 > 0$ on investointien suhde bruttotuloihin työelämään siirryttäessä (kun $t = s$) ja β on tämän suhteen vuosittainen lasku. T on positiivinen nettoinvestointiaika. Asettamalla $r_p = 0$ palataan yhtälön (4) tapaukseen. Mallien avulla pyritään poistamaan muiden kuin koulutusinvestointien vaikutukset ansioihin, jotta koulutuksen tulovaikutusta voitaisiin täsmällisemmin analysoida.

2.2.2. Koulutus, lahjakkuus ja sosioekonominen tausta tulon selittäjinä

Yksinkertaistetuissa koulutuksen tulomalleissa eräänä implisiittisenä (joskus myös eksplisiittisenä) oletuksena on, että tulonsaajat ovat geneettiseltä perimältään ja perhetaustaltaan samanlaisia. Tämä luonnollisesti helpottaa teoreettisten mallien konstruointia, mutta samalla ko. mallien empiirinen testattavuus vaikeutuu, koska testattavat populaatiot eivät täytä mallien alkuoletuksia.

Esimerkiksi koulutuksen ja lahjakkuuden (älykkyyden, kyvykkyyden) keskinäinen vuorovaikutus on vaikea pulma tutkittaessa koulutuksen tulovaikutusta. Esitetään perusongelma seuraavassa yksinkertaistettuna. Oletetaan, että tulot riippuvat lineaarisesti koulutuksesta ja lahjakkuudesta:

$$(7) \quad Y = \alpha + \beta_1 S + \beta_2 A + U,$$

missä Y on tarkastellun tilastoyksikön (henkilön, kotitalouden tms.) tulo, S koulutus, A lahjakkuus ja U esittää muita tuloihin vaikuttavia tekijöitä, joiden yleensä oletetaan olevan satunnaisia ja riippumattomia koulutuksesta S ja lahjakkuudesta A . Olkoon $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$ mallin (7) parametrien pns-estimaatit. Jos koulutus ja lahjakkuus korreloivat positiivisesti, niin mallista $Y = \bar{\alpha} + \bar{\beta}_1 S + \bar{U}$ (josta lahjakkuusmuuttuja A puuttuu) saadaan koulutuksen kertoimelle $\bar{\beta}_1$ pns- menetelmällä estimoitaessa tavallisesti $\hat{\beta}_1$:een nähden ylöspäin harhainen arvo $\hat{\beta}_1 + \hat{\beta}_2 b_{AS}$. Termi b_{AS} on lahjakkuuden ja koulutuksen välinen regressiokerroin, nimittäin mallista $A = a + b_{AS} S + u$ estimoitu kerroin. Tämä on vain erikoistapaus yleisestä ongelmasta, poisjätettyjen muuttujien vaikutuksesta regressioanalyysissä, ks. Goldberger (1964, s. 194-201).

Todettakoon, että inhimillisen pääoman teoriassa Becker (1964) kiinnitti huomiota lahjakkuuden ja koulutuksen mahdolliseen positiiviseen korrelaatioon. Inhimillisen pääomateorian tulomalleihin on myöhemmin lisätty lahjakkuusmuuttuja sekä myös muita taustamuuttujia.

Erilaisten taustatekijöiden vaikutusta tuloon, ammatinvalintaan ja koulutukseen on viime vuosina entistä enemmän ryhdytty tutkimaan. Useilla tämänlaatuisilla tutkimuksilla on yhteisenä piirteenä rekursiivinen malli, joka on koostunut yhdestä tai useammasta yhtälöstä

$$(8) \quad A = a_1 X + eG + u_1 \quad (A = \text{Ability, lahjakkuus})$$

$$(9) \quad S = a_2 X + b_2 A + u_2 \quad (S = \text{Schooling, koulutus})$$

$$(10) \quad O = a_3 X + b_3 A + c_3 S + u_3 \quad (O = \text{Occupation, ammatti})$$

$$(11) \quad Y = a_4 X + b_4 A + c_4 S + dO + u_4, \quad (Y = \text{Income, tulot})$$

Sosioekonomisten taustamuuttujien X joukko ja geneettisen perimän muuttuja G ovat mallissa eksogeenisia. Lahjakkuusmuuttuja A riippuu sekä X :stä että G :stä. Koulutusmäärä S riippuu erilaisista taustamuuttujista X , lahjakkuudesta ja satunnaistermistä u_2 . Ammatillinen status O on taustamuuttujien, lahjakkuuden ja koulutusvuosien sekä satunnaismuuttujan u_3 funktio. Tulo Y puolestaan riippuu X :stä, A :sta, S :stä, ammatista ja satunnaistermistä u_4 . Jos esimerkiksi tutkitaan taustamuuttujien kokonaisvaikutusta koulutukseen, kertoimen a_2 tarkastelu ei riitä, vaan myös lahjakkuusmuuttujan kautta tapahtuva epäsuora vaikutus on huomioitava.

Duncan (1968) on esittänyt erään muunnelman esitetystä rekursiivisesta mallityypistä, ns. polkumallin (path model). Hänen mukaansa tulon selittäjistä ammatti oli merkittävin. Tulot määräytyvät pääasiassa yksilöllisten ominaisuuksien perusteella, perhetausta vaikutti lähinnä niihin ominaisuuksiin, jotka puolestaan selittivät tuloja. Tulos on intuitiivisesti selkeä. Vastaavanlaisia mallityyppejä ovat käyttäneet Sewell ja Hauser (1975) ja Conlisk (1971) vaihtelevin tuloksin.

Erääksi tulomallin selittäväksi muuttujaksi ovat ottaneet mm. Wachtel (1975), Solmon (1975) ja Morgenstern (1973) "koulutuksen laatu"-tekijän. Wachtelin tutkimusnäytteessä koulutuksen laatua kuvaa koulun "taso", joka on laskettu joko kuluina oppilasta kohden tai Gourmanin luokituksen mukaan. Leibowitz (1974) ja Lindert (1976) tutkivat tuotantofunktio-näkökulmasta perheen taustamuuttujien merkitystä. Leibowitzin mukaan äidin koulutuksen merkitys selitettäessä lapsuudenaikaista lahjakkuutta saattaa johtua siitä, että lasta kohti uhrattu aika voi kasvaa äidin koulutuksen myötä.

Taustamuuttujien tarkastelun jälkeen käsittelemme seuraavaksi kysymystä kuinka koulutuksen tulovaikutusta tutkittaessa poisjätettyjen ja ei-havaittujen muuttujien vaikutuksia voidaan selvittää ja mitata. Esimerkiksi Griliches ja Mason (1972) ovat tutkineet, mitä lahjakkuuden poisjättäminen on tulomalleissa merkinnyt. Heidän aineistonaan oli vuoden 1964 USA:n väestötutkimuksen näyte asepalveluksensa suorittaneista miehenkilöistä. Lopulliseen aineistoon oli saatu 1454 kokopäiväisesti työskentelevää henkilöä, joiden iät vaihtelivat 21-34 vuoteen. Lahjakkuusmittarina oli armeijan pätevyystesti (Armed Forces Qualification Test, AFQT), jota käytettiin älykkyystestin sijasta.

Griliches ja Mason jakoivat koulutuksen kahteen osaan; armeijaa ennen saavutettuun koulutukseen (SB) ja armeijan jälkeen hankittuun koulutuksen lisäykseen (SI). Koulutuksen ja lahjakkuuden tulovaikutuksen selvittämiseksi he mm. suorittivat regressioanalyysin, jossa tulojen logaritmia selitettiin (a) koulutuksen ja (b) koulutuksen ja lahjakkuuden avulla, kun malleissa selittäjinä olivat lisäksi ikä, asepalveluksen kesto ja rotu. Lahjakkuuden A lisääminen vähensi SI:n kerrointa 10 % ja SB:n kerrointa 17 %. Koulutusmuuttuja oli siten ottanut osan lahjakkuuden selityksestä itselleen, kun lahjakkuustekijä oli puuttunut mallista. Lahjakkuusmuuttujan A lisääminen malliin nosti sen kokonais selitysstetta $100 R^2$ vain alle prosenttiyksikön.

Tutkittuaan myös muiden taustatekijöiden vaikutuksia tuloihin Griliches ja Mason korostivat koulutuksen merkitsevyyttä havaittujen tuloerojen selittäjänä: yksi lisävuosi koulutusta lisää heidän analyysinsä mukaan 4.6 % tuloja, kun synnyinseutu, armeijan palveluksen kesto, lahjakkuus (AFQT-testillä mitattuna) ja isän status ovat vakioituja. Vastaavasti 10 %:n lisäys lahjakkuudessa nostaisi tuloja vain noin prosentilla.

Griffin (1976, s. 121-139) on tutkinut sen spesifiointivirheen vaikutusta, joka syntyy, kun henkilöiden sosioekonomisen taustan (SES-muuttuja) ja lahjakkuuden vaikutukset on jätetty analysoimatta.

Griffin määritteli neljä mallia "sosioekonomisten saavutusten" (socio-economic achievement) tutkimiseksi. Ne ovat

$$(1) \text{ Lahjakkuus} = b_1(\text{SES}) + e_1$$

$$(2) \text{ Koulutus} = b_2(\text{SES}) + b_3(\text{lahjakkuus}) + e_2$$

$$(3) \text{ Ammatillinen status} = b_4(\text{SES}) + b_5(\text{lahjakkuus}) + b_6(\text{koulutus}) + e_3$$

$$(4) \text{ Tulot} = b_7(\text{SES}) + b_8(\text{lahjakkuus}) + b_9(\text{koulutus}) \\ + b_{10}(\text{ammattillinen status}) + e_4$$

Tulojen selitetystä vaihtelusta koulutuksen selitysosuus oli 14-44 %, mutta kun lahjakkuuden ja SES-muuttujan vaikutukset oli otettu mukaan, koulutus selitti vain 1-2 % vuositulojen vaihtelusta. Kun ainoastaan SES-muuttuja "kontrolloitiin" koulutuksen lisäksi, koulutuksen tulovaikutus laski 20-40 % verrattuna tilanteeseen, jossa SES-muuttuja ei ollut selittäjänä. Lahjakkuuden, isän ammatin ja isän koulutuksen vaikutusten eliminointi pienensi koulutuksen kerrointa 13-37 %.

Eräs tapa eliminoida ei-havaittavien geneettisten tekijöiden ja kotiympäristön vaikutuksia tulomallien koulutusmuuttujien kertoimiin on tutkia ko. suhteessa samanlaisia henkilöitä. Identtiset kaksoset ovat geneettiseltä perimältään täysin samanlaisia ja kotiympäristöltään melko samankaltaisia. Myös tavallisten sisarusten ja kaksosten taustatekijät ovat lähempänä toisiaan kuin satunnaisten henkilöparien taustatekijät.

Eräät tutkijat ovat käyttäneet hyväkseen sukulaisuussuhteista johtuvia samankaltaisuuksia. Tutkimalla esim. sisarus- tai kaksosparikkien välisiä eroja tutkijat ovat voineet mitata ja selvittää ei-havaittavien geneettisten tekijöiden ja kotiympäristön merkitystä tulojen määräytymisessä, katso esim. Taubman (1977).

Olneck (1977) tutkittuaan sisarusparikkien välisiä eroja ja Behrman, Taubman ja Wales (1977) vastaavasti kaksosparikkien eroja havaitsivat, että koulutusmuuttujan kerroin tulomallissa laski näissä näytteissä huomattavasti verrattuna 'yksilöiden' väliseen tarkasteluun. Sahota (1978) on laatinut varsin mielenkiintoisen arvioivan yhteenvedon erilaisista tulojakaumateorioista ja tarkastelee myös koulutuksen tulovaikutukseen liittyviä tuloksia.

2.2.3. Inhimillisen pääoman teoria - kritiikkiä

Inhimillisen pääoman teorian lähtökohtana on rationaalisen valinnan teoria. Jotta koulutusinvestointi olisi "järkevä valinta" koulutettavan mielestä, täytyisi tämän uskoa saavansa siitä vähintään sellaisen tuoton, joka korvaa tehdyt uhraukset. Koulutuksen "tuottopuolelle" voidaan laskea taloudellisten hyötyjen lisäksi ei-koulutus-vaihtoehtoon nähden suurempi yhteiskunnallinen arvostus, parempi työn laatu ja pysyvyys, entistä suuremmat mahdollisuudet omaehtoiseen toimintaan, jne. Koulutuksen vaikutukset eivät siis rajoitu vain perustarpeiden tyydytysvälineiden (käytännössä tulojen) hankkimiseen, vaan koulutus usein välittömästi tyydyttää 'henkisiä' tai 'älyllisiä' tarpeita: uteliaisuutta, tiedon halua, itsensä kehittämistä, jne., ks. esim. Ahmavaara (1969, s. 180-196), Maslow

(1954), Vepsä (1973, s. 29-45). Tämän kaltaisten tekijöiden mittaamisen vaikeudesta ilmeisesti johtuu, ettei niitä normaalisti käsitellä koulutusinvestointien teoriassa. Tästä teorian laiminlyönnistä ei voida päätellä, että ne eivät olisi todellisuudessa merkittäviä tekijöitä, ks. esim. Talousneuvosto (1972, s. 74-94 ja Liite 3), Tinbergen (1975).

Mikäli muut kuin koulutuksesta johtuvat tuloerot jätetään huomiotta, niin reaalitytulovirran nykyarvo investoitaessa koulutukseen pitäisi olla vähintään reaalitytulovirran nykyarvo ilman tätä investointia. Jos nämä reaalitytulovirtojen nykyarvot olisivat yhtäsuuria, niin "rationaalinen homo economicus" (vrt. Pareto (1971, s. 12-18)) olisi tällöin indifferentti eri koulutusvaihtoehtojen suhteen. Investoinnin tuottoaste (r) olisi tällöin se diskonttokorko, jolla (odotettujen) reaalitytulovirtojen nykyarvot tulisivat yhtäsuuriksi. Inhimillisen pääoman lähestymistavassa oletetaan usein, että investoinnin tuottoaste pysyisi vakiona ajan kuluessa ja olisi sama kaikille. Tämä on ilmeisen epärealistinen oletus: vuoden lisäkoulutus ei lisää tuloja aina ja kaikkialla samassa suhteessa. Inhimillisen pääoman teoria nojautuu myös muihin oletuksiin. Seuraavat lisäoletukset on tehtävä, jotta teoreettiset mallit kertoisivat sen, mitä niiden on tarkoituskin kertoa, ks. Merret (1971, s. 213):

- (1) Ihmiset ovat yhtä lahjakkaita (kyvykkäitä, älykkäitä)
- (2) Työmarkkinoilla vallitsee täydellinen kilpailu
- (3) Ihmiset tietävät potentiaalisen tulo- ja kustannusvirtansa
- (4) Ihmiset ovat indifferenttejä eri investointivaihtoehtojen suhteen, mikäli ne tuottavat saman reaalitytulovirran.

Näistä oletuksista ilmenee selkeästi, kuinka vaikeata on empiirisesti testata inhimillisen pääoman teorian tulomalleja. Teoreettiset mallit kuvaavat kuviteltua 'Simpletaniaa', josta puuttuvat useimmat esim. suomalaisen arkitodellisuuden piirteet.

Yksilön tulotaso määräytyy vain osaksi koulutuksen kautta. Muista tärkeistä tuloon vaikuttavista tekijöistä mainittakoon sukupuoli, rotu, lahjakkuus, ikä, fyysinen kunto ja ammatin laatu (rasittavuus, vaarallisuus, arvostus). Jos käytetään ainoastaan koulutusta tulojen selittäjänä yliarvioidaan tavallisesti koulutuksen todellinen vaikutus tulonmuodostukseen. Tämä johtuu siitä, että korkeammin koulutettuihin ja hyvätuloisempiin tehtäviin valikoituu normaalisti esim. perityn lahjakkuuden ja aktiivisuuden suhteen keskimääräistä kyvykkäämpiä henkilöitä. Pelkääntään meriitteihin perustuvan järjestelmän terävä analyysi ja satiiri on Young (1967): "Meritokratian nousu 1870-2033, tutkielma koulutuksesta ja tasa-arvosta". Sukupuolesta ja rodusta johtuvien erojen analysoiminen on erityisen problemaattista, koska näiden tekijöiden perusteella tapahtuvaa valikoitumista ohjaavat yhteiskunnassa vallitsevat monimutkaiset normit, säännöt ja etuoikeudet. Analysoinnin eräs vaikeus on, että esim. sukupuolen tapauksessa on olemassa sekä perusteltua että epäoikeudenmukaista erottelua.

Inhimillisen pääoman teorian mallit ovat liian yksinkertaisia testattavaksi tavallisen poikkileikkausaineiston avulla. Ei-kokeellinen havaintomateriaali ei vastaa teorian oletuksia, vaan todellisuus on yksinkertaistettua teoriaa huomattavasti monisärmäisempi. Luvussa 4 tarkastellaan, mitä mahdollisuuksia on ei-kokeellisen aineiston perusteella estimoida inhimillisen pääomateorian mukaisia vaikutuksia.

3. HAVAINTOAINEISTON HANKINTA

3.1. Yleistä

Tarvittavaa tietoaineistoa etsittäessä tutkittiin Tilastokeskuksen suorittamien kotitaloustiedustelujen soveltuvuutta tutkimusaineistoksi ja aineiston yleisrakenteen selvittämisen jälkeen viimeisin julkaistu tutkimus, vuoden 1971 kotitaloustiedustelu, valittiin tarkemman selvityksen kohteeksi.

Kuten tilastojulkaisujen lähemmässä tarkastelussa saattaa käydä, vuoden 1971 kotitaloustiedustelun raportista, ks. Tilastokeskus (1977 a,b,c), ei löytynyt tämän tutkimustyön kannalta riittävän yksityiskohtaisia tietoja. Julkaisussa on tarkasteltu vain osaa tulojen suuruuteen vaikuttavista tekijöistä. Tiedot on esitetty tiivistetyssä muodossa erilaisissa taulukoissa, joiden edelleen tiivistäminen tilastollisin menetelmin ei tuntunut kovin hyödylliseltä. Valmis julkaisu vuoden 1971 kotitaloustiedustelusta ei riittänyt aineistopohjaksi, vaan tarvittiin tietojen hienorakennetta, yksikkökohtaisia tietoja. Yksikkötason tietoja saadaan ainoastaan kotitaloustiedustelun alkuperäisestä perusaineistosta, jota säilytetään Tilastokeskuksessa.

Yksinkertaisesti tarvittavat tiedot saa Tilastokeskuksen tietopalvelun tai julkaisun tehneen osaston kautta. Tilastokeskuksen tietopalvelu on kaikkien tutkijoiden käytettävissä, ja se jakaantuu kolmeen osaan; kirjasto, arkisto ja maksupalvelu. Kirjasto suorittaa tietopalvelua omista aineistoistaan, lähinnä tilastoista ja vastaa puhelintiedusteluihin.

Arkistossa säilytetään tilastotiedustelujen perusaineistoja (esim. alkuperäiset lomakkeet), jotka ovat salaisia n. 10-20 vuoteen tiedustelujen suorittamisen jälkeen. Tietoja voi saada arkistosta käyttöönsä myös niiden ns. "salaisuusaikana" Tilastokeskuksen ylijohdajan suostumuksella. Varsinainen tietopalvelu keskittyy maksupalvelun ympärille. Asiakkaalle voidaan toimittaa tietoja valmiista perusaineistosta joko magneettinauhoina tai taulukkoina. Mikäli tietoja halutaan yksikötason tiedostoista tarvitaan Tilastokeskuksen ylijohdajan erityislupa. Tällöin periaatteena on, että tietoja annetaan vain sellaisiin tilastollisiin tarkoituksiin, joista yksilötason tietoja ei voida saada esiin. Maksupalvelu perii palveluksistaan (suunnittelusta, ohjelmoinnista, tietokoneajoista, jne.) asetuksen mukaiset maksut, jotka tarkistetaan vuosittain.

Tässä tutkimuksessa tarvittavan kotitaloustiedustelun 1971 perusaineisto on talletettu erilaisina ATK-tiedostoina magneettinauhuille konekielisesä muodossa. Koska kysymyksessä on kotitalouskohtainen aineisto, sen käyttöön saamisen edellytyksenä on Tilastokeskuksen ylijohdajan lupa. Periaatteena on, että yksityistä henkilöä tai kotitaloutta ei annetuista tiedoista saa tunnistaa. Erityisluvan hyväksymisen jälkeen voi sopia maksupalvelun kanssa konkreettisista toimenpiteistä.

Eräs ongelma kotitaloustiedustelun 1971 perustietojen hankinnassa on soveliaiden tietojen saamisen kalleus. Tämä edellytti KT-tiedustelun eri ATK-tiedostojen sisällön ja rakenteen tarkkaa selvittämistä. Tavoitteena oli saada valitun näytteen otantakustannukset, etenkin työkustannukset, mahdollisimman alhaisiksi.

Nimensä mukaisesti vuoden 1971 kotitaloustiedustelussa on pääasiassa kotitalouskohtaisia (eikä esim. henkilökohtaisia) tietoja. Yksilötason tietoja on jonkin verran kerätty erikseen kotitalouksien eri jäsenten taustatekijöistä ja tulojen suuruudesta. Tiedustelun perustiedot on talletettu erikseen yksilö- ja ruokakuntatason erittäin laajoihin magneettinauhatieostoihin, joita yhteensä on 12 kappaletta. Yksilötason tietojen käyttö tutkimusnäytteen otannan pohjana olisi edellyttänyt vähintään nk. yksilötietotiedoston ja tulotiedoston läpikäymistä. Tämä olisi kuitenkin tullut liian kalliiksi. Tavoitteena oli löytää tiedosto, johon sopivien taustamuuttujien ohella kuuluisi jokin tulomuuttuja. Suoritetun selvityksen perusteella todettiin, että Tilastokeskuksen ns. ruokakuntakohtaiseen rakennetiedostoon sisältyi oikean laatuista tietoja. Rakennetiedostossa on kotitalouden taustaa kuvaavia muuttujia sekä myös tarpeellinen tulomuuttuja, kotitalouden käytettävissä oleva vuositulo.

Otoksen poimimisessa rakennetiedostosta ei ollut tarpeellista käyttää Tilastokeskuksen maksupalvelun palveluksia, vaan otos poimittiin Kauppa-
korkeakoulussa Tilastokeskuksen suostumuksella KTT Liisa Uusitalon aiemmin hankkimasta kopiosta.

3.2. Kotitaloustiedustelu 1971

Vuoden 1971 kotitaloustiedustelun perusjoukko koostuu maan kaikista yksityisistä kotitalouksista. Laitostaloudet, joihin kuuluvat esimerkiksi vanhain- tai lastenkodeissa, sairaaloissa ja rangaistuslaitoksissa olevat henkilöt, asevelvolliset jne, ovat jääneet tutkimuksen ulkopuolelle, ks. Tilastokeskus (1977 a, s. 4). Näyte on poimittu suorana asuntokuntanäytteenä ilman esiasteen kuntaotosta, joten aineistossa on mukana kotitalouksia lähes kaikista maan kunnista. Näytteitä on kaksi: kuukausitilinpito- ja vuosihaastattelunäytteet. Kuukausitilinpidon näytteen poiminta on suoritettu kahdessa vaiheessa. Ensin on poimittu valtion tietokonekeskuksessa väestörekisteristä ns. perusnäyte asuntokunnista tasavälistä satunnaisotantaa käyttäen. Tilastokeskuksessa näytteen asuntokunnat on jaettu ruokakuntiin nimitietojen perusteella siten, että kaikki asunnossa asuvat ruokakunnat ovat tulleet mukaan. Lopullinen ruokakuntien rajaaminen on jäänyt haastattelijan suoritettavaksi. Lisäksi saatu perusnäyte on ositettu siten, että mukaan on otettu kaikki maanviljelijäruokakunnat. Muista ruokakunnista lopulliseen näytteeseen on otettu joka toinen. Vanhuskotitalouksista on poimittu lisäksi täydennysnäyte. Siihen on otettu mukaan sellaiset alkuperäisessä perusnäytteessä olleet kotitaloudet, joiden kaikki jäsenet olivat syntyneet vuonna 1905 tai aikaisemmin. Vuosihaastattelua varten väestörekisteristä oli poimittu kuukausitilinpidon ruokakuntien täydennykseksi tasavälistä satunnaisotantaa käyttäen noin 8600 ruokakuntaa. Lopullisen näytteen kooksi tuli kuukausitilinpidossa 3512 kotitaloutta (jotka sisältyvät myös vuosiaineistoon) ja vuosiaineistossa kaikkiaan 8817 kotitaloutta.

Kotitaloustiedustelun perustiedot Tilastokeskus on kerännyt yhdistetyllä tilinpito- ja haastattelumenetelmällä. Lisäksi on kerätty tietoja vero- toimistoilta postitiedustelulla sekä Kansaneläkelaitokselta koneellisesti sen aineistoista. Kotitaloustiedusteluissa on perinteisesti keskitytty kulutuksen rakenteen selvittämiseen. Kotitalouksien suorittaman kuukausi-tilinpidon avulla tutkitaan mm. kertakulutustavaroiden, puolikestävien tavaroiden ja palvelusten käyttöä. Myös kotitalouskoneiden yms. omistuksesta ja ostosuunnitelmista sekä nuorison varojenkäytöstä saadaan kuukausi-tilinpidosta tietoja. Vuosihaastattelussa on tutkittu mm. kotitalouksien kestotavaroita, säästämistä, ilmaisipalvelujen käyttöä ja ajankäyttöä sekä tämän tutkimuksen kannalta oleellisia tulo- ja varallisuustietoja. Vuoden 1966 kotitaloustiedustelussa nämä tiedot olivat ensimmäisen kerran mukana. Haastattelulla on kerätty tiedot sellaisista tuloeristä, joita muista lähteistä ei ole voitu saada tai jotka oli määritelty rekisteritiedostoissa tulonjakotilastosta poikkeavalla tavalla. Veroviranomaisilta on saatu osa tulo- ja varallisuustiedoista, ja kansaneläkelaitokselta tiedot kansaneläkejärjestelmään kuuluvista tulonsiirroista. Verrattaessa ruokakunnan koon ja päämiehen iän jakaumia väestölaskennasta 1970 saatuihin jakaumiin, voidaan todeta eri kokoisia ja eri ikäisten päämiesten ruokakuntia olevan näytteessä oikeassa suhteessa. Seuraavasta asetelmasta selviää kotitalouksien jakaantuminen päämiehen iän mukaan.

Päämiehen ikä	Kotitaloustiedustelu	Väestölaskenta
- 24	8.7 %	7.6 %
25 - 44	38.3	38.0
45 - 64	37.3	37.8
65 -	15.7	16.6
Yhteensä	100.0 %	100.0 %

Ruokakunnan keskikooksi saatiin kotitaloustiedustelun aineistosta 2.97 henkilöä. Vastaava luku vuoden 1970 väestölaskennan mukaan oli 2.99.

3.3. Rakennetiedosto

Rakennetiedosto on Tilastokeskuksen omia taulukointeja ja selvityksiä varten kotitaloustiedustelusta saatujen tietojen pohjalta muodostettu ruokakuntakohtainen tiedosto. Sitä voisi luonnehtia kotitaloustiedustelun perustiedostoksi. Useat kotitaloustiedustelun 1971 julkaisun ris-tiintaulukoinnit on suoritettu juuri rakennetiedostoa hyväksikäyttäen. Tiedostossa on mukana kaikki kotitaloustiedustelunäytteen kotitaloudet. Rakennetiedoston tietuekuvaus on esitetty liitteessä 1. Tietuekuvauksessa esiintyvät kuntakoodin tiedot on tässä tutkimuksessa korvattu nolllilla, sillä tämä oli Tilastokeskuksen salassapitovelvollisuuteen liittyvä edellytys tiedoston käytölle.

Rakennetiedostossa on sekä kuukausitilinpidon että vuosihaastattelun tietoja. Kotitalouksiin liittyviä rakennetiedoston taustatekijöitä ovat mm. kotitalouden asuinalue, jäsenluku, päämiehen eräät ominaisuudet ja kotitalouden sosioekonominen asema. Lisäksi tiedostoon kuuluvat tutkimuksen kannalta tärkeät tiedot ruokakuntien käytettävissä olevasta vuositulosta.

Tässä tutkimuksessa ei ole syytä ottaa kaikkia rakennetiedoston kotitalouksia mukaan analyysiin, vaan rakennetiedostosta poimitaan 1000:n aktiivikotitalouden satunnaisotos. Aktiivikotitalouksiksi määritellään ne kotitaloudet, joiden päämies on alle 65 vuotias. Yli 64 vuotiaiden kotitalouksien päämiesten ja kotitalouksien tulon välinen suhde on aivan oma erityisongelmansa, johon tässä tutkimuksessa ei ole tarkoitus puuttua. Ei-aktiivien kotitalouksien tutkiminen jätetään muiden tutkijoiden tehtäväksi. Otettua näytettä voidaan pitää edustavana otoksena Suomen vuoden 1971 aktiivikotitalouksien perusjoukosta.

3.3.1. Määritelmiä ja luokituksia

Tutkitaan seuraavassa lähemmin rakennetiedostossa esiintyvien käsitteiden merkitystä sekä erilaisten luokitusten sisältöä.

Kotitalouden (tai ruokakunnan) määrittelyssä ja rajaamisessa päätuntonmerkkinä on yhdessä asuminen ja osaksi tai kokonaan yhteinen ruokatalous. Kotitalouteen kuuluu myös sellainen perheenpää tai muu jäsen, joka käy työssä muualla ja on kotonaan vain vapaapäivinä eikä ole täysihoidossa missään muussa kotitaloudessa, ks. Tilastokeskus (1977 b, s. 26-).

Tulokäsitteenä on kotitalouden käytettävissä oleva vuositulo. Sen muodostaminen tapahtuu melko työläästi seuraavan asetelman mukaisesti:

Saadut tuotannontekijätulot

 palkkatulot

 yrittäjätulot

 omaisuustulot

Saadut tulonsiirrot (+)

Maksetut tulonsiirrot (-)

Saadut pääomansiirrot (+)

Annetut pääomansiirrot ja saaduista pääomansiirroista
maksetut verot (-)

= Käytettävissä oleva tulo

Kotitaloustiedustelun 1971 tulojen laskentamenetelmästä ilmenee mm., että palkkatuloihin sisältyy arvio erilaisten luontoisetujen arvosta ja että

maatalouden tuloihin sisältyy arvio kotitalouden kulutukseen otettujen maatalouden tuotteiden ja palvelusten arvosta. Käytettävissä olevan tulo on verraten laaja tulokäsite, joka sisältää mm. palkat, voitot, sosiaali-turvaetuudet, veikkausvoitot ja perinnöt. Se kuvaa kotitalouden kulutus-mahdollisuuksia eli maksimituloa, jonka kotitalous voi käyttää varalli-suuttaan vähentämättä tarpeidensa tyydyttämiseen.

Kotitalouden päämiehen koulutus on luokiteltu systemaattisen koulutus-alaluokituksen 1970 mukaisesti. Aloittaiset koulutuskoodit osoittivat vain ao. koulutusalan ja koulutuksen alakohtaisen laajuuden. Ne eivät so-veltuneet sellaisenaan tämän tutkimuksen tarkoituksiin, vaan ne täytyi muuttaa vastaamaan koulutusvuosia. Aloittaiset koulutuskoodit muutettiin vastaaviksi keskimääräisiksi koulutusvuosiksi Tilastokeskuksen Koulutus-luokittelua 1971 ja harkintaa hyväksikäyttäen. Esimerkiksi ekonomin tut-kinnon suorittanut päämies sai koulutusvuosikseen 15 vuotta.

Muut tärkeät analyysiin mukaan tulevat luokittelut, kuten kotitalouden asuinalue, kuntatyyppi, päämiehen työnantaja, elinkeino ja sukupuoli, kotitalouden sosioekonominen asema sekä asuinympäristö on esitetty liit-teessä 2.

3.4. Tilastoaineisto ja tutkimusongelman raja

Kun yleisellä tasolla tutkimusongelmana on, miten koulutus selittää tulojen vaihtelua, rajaa käytettävissä oleva havaintoaineisto melko voimakkaasti kyseisen ongelman käsittelyä. Tilastoyksikköinä ovat kotitaloudet, joiden edustajina ovat niiden päämiehet. Kotitalouden käytettävissä olevan tulojen vaihtelua pyritään selittämään kotitalouden päämiehen koulutuksen sekä kotitalouden taustatietojen avulla. Tutkimusongelma tarkentuu muotoon, missä määrin kotitalouden päämiehen koulutus ja eräät taustatiedot selittävät kotitalouden käytettävissä olevien tulojen vaihtelua. Seuraavassa luvussa tutkimuksen pääongelma puetaan seuraavan kysymyksen muotoon: Montako prosenttia suurempi kotitalouden käytettävissä oleva tulo olisi keskimäärin ollut, jos sen päämiehen koulutusaika olisi ollut vuoden pidempi, ceteris paribus. Havaintoaineisto on saatu vuoden 1971 suomalaisista aktiivikotitalouksista ja tutkimuksen tulokset koskevat tätä perusjoukkoa.

Tutkimuksen havaintomateriaali on osa kotitaloustiedustelun kotitalouksien ominaisuuksia kuvaavista tiedoista ja tutkimus kuuluu siten ei-kokeellisten tutkimusten piiriin. Ei-kokeellinen tutkimus edellyttää, kun selvitetään kotitalouksien käytettävissä olevan tulojen ja niiden päämiesten koulutuksen välistä suhdetta, että muista tuloihin vaikuttavista tekijöistä on saatavissa riittävän yksityiskohtaisia tietoja. Kotitaloustiedustelun rakennetiedostoon sisältyy vain joitakin tämän tyyppisiä luokitteluja, joten vain osa taustatekijöiden aiheuttamasta tulojen vaihtelusta voidaan selvittää. Seuraavassa luvussa tarkastellaan lähemmin, minä tyyppisiä lisäongelmia kokeelliseen tutkimustilanteeseen nähden syntyy, kun johtopäätöksiin pyritään ei-kokeellisen havaintoaineiston perusteella.

4. TUTKIMUKSEN KYSYMYKSEN ASETTELU

4.1. Tieteellisen päättelyn problematiikkaa

Empiiristen tieteiden tehtävänä voidaan nähdä maailmaa koskevien tosiasiaväitteiden esittäminen. Tämän luonnehdinnan perusteella puhdasta matematiikkaa, logiikkaa, filosofiaa, etiikkaa ja esim. teologiaa ei voida pitää ainakaan pääasiallisesti empiirisinä tieteinä. Taloustieteiden kohteena ovat taloudellisiksi luokitellut ilmiöt, joita koskevia invariansseja ja riippuvuuksia niissä pyritään selvittämään. On varsin vaikeaa ja ehkä turhaakin yrittää vetää selvää rajaa esim. taloudellisten ja muiden yhteiskunnallisten tai psykologisten ilmiöiden välille. Valtio-opin, sosiologian, sosiaalipsykologian ja psykologian tavoin taloustieteiden tutkimuskohteena on inhimillinen toiminta.

Tutkimuskohteena erityispiirteiden vuoksi taloustieteissä esiintyy ongelmia, joita luonnontieteissä kuten fysiikassa, kemiassa, tähtitieteessä tai meteorologiassa ei ole vastaavassa määrin havaittavissa. Tällainen on esimerkiksi arvosidonnaisuuden ongelma: tosiasiaväitteiden ja arvojen (arvostusten) erottaminen toisistaan. Taloustieteissä tämä on johtanut pyrkimykseen erottaa normatiivinen ja positiivinen taloustiede toisistaan. Jälkimmäisessä pyritään - vaikka se varmastikin on yhteiskunta- ja käyttäytymistieteissä vaikeaa - välttämään arvostusten esittämistä, esim. toteamasta mitä *tulisi* tehdä tai mikä olisi *hyvä* ratkaisu. Esim. Lipsey (1966, s. XV) toteaa: "Economic theory cannot, of course, ever show us what we ought to do, but only what will happen if we do do certain things".

Tosiasiaväitteiden ja arvostusten erottamiseen liittyy vaikeita filosofisia ongelmia, joihin tässä ei kuitenkaan ole syytä puuttua, ks. esim. Ketonen (1976, s. 141-199), von Wright (1968, s. 245-257).

Taloustieteiden ja luonnontieteiden havaintomateriaalissa on niinikään selviä eroja. Valtaosa luonnontieteiden "datoista" on nykyisin (toisin kuin ennen Galileita!) koetuloksia, jotka lisäksi ovat usein kvantitatiivisia, siis numeerisia mittaustuloksia. Poikkeuksellisessa asemassa luonnontieteilijöiden joukossa tässä suhteessa ovat esim. tähtitieteilijät, geologit ja meteorologit, joiden havainnot eivät pääasiallisesti perustu kontrolloituihin kokeisiin (Näiden tieteiden menettelytavoista voisivat yhteiskuntatieteilijät ottaa mallia puhtaasti kokeellisten luonnontieteiden sijaan). Mutta heilläkin valtaosa havaintoaineistosta on kvantitatiivista ja tarkkoihin mittauksiin perustuvaa.

Yhteiskuntatieteiden havaintoaineisto sensijaan on pääasiassa ei-kokeellista ja suurelta osalta kvalitatiivista, ks. esim. Ahmavaara (1969, s. 74-86). Kvantitatiiviset mittaukset (jos niitä ollenkaan on käytettävissä) eivät normaalisti kuvaa juuri tutkimuksen kohteena olevaa asiaa, vaan mittausten validiteetti on usein heikkoa.

Tämä tilanne asettaa suuria vaikeuksia havaintoaineiston analysoinnille ja johtopäätöksiä tekemiselle, kuten tulemme seuraavassa havaitsemaan. Ilmeisesti ongelmien vaikeudesta juuri johtuu, että tämántapaiset ongelmat jätetään usein tutkimuksissa käsittelemättä. Tavanomainen menettely esimerkiksi opinnäytetöissä on "soveltaa" aihepiiriä käsittelevää teoriaa (jossa havaintoaineiston hankintaa koskevat vaikeudet on tavallisesti jätetty huomiotta) sellaisenaan käsilläolevaan havaintoaineistoon.

Tällöin virheelliset väitteet usein kätkeytyvät viattoman tuntuisiin "tilastollisiin olettamuksiin", jotka on usein sellaisenaan kopioitu jostain metodikirjasta. Tämän tapaisten perustelemattomien olettamuksien ja väitteiden esittäminen on esimerkiksi ekonometrisissä tutkimuksissa pikemminkin sääntö kuin poikkeus, mikä on vähitellen aikaansaanut merkillisen toleranssin näitä hämäryyksiä ja perustelemattomuuksia kohtaan. Asiantuntijoiden vaitioloa ylläpitää ilmeisesti myös keskinäisen puuttumattomuuden periaate, ks. esim. Galbraith (1971, s. 29-40, 55-74), Wiener (1969, s. 132-136), Ketonen (1976, s. 170-189).

Seuraavassa tarkastellaan tilastollisen tutkimuksen ongelmia Woldin (1954, 1956, 1963, 1967) esittämän ns. lipputaulun avulla. Wold jakaa tilastolliset tutkimusongelmat luokkiin niiden tavoitteiden (ts. onko kysymys kuvailevasta vai selittävästä tutkimuksesta) ja havaintoaineiston laadun (ts. kokeellisten ja ei-kokeellisten havaintojen) perusteella. Wold (1967) tarkastelee erityisesti ennustemallien rakentamista ja esittää seuraavan pohjoismaisen lipun muotoon laaditun luokittelun.

Lipputaulussa $S \Rightarrow P \Rightarrow U$ on lyhennysmerkintä otosinformaation S perusteella tapahtuvalle päättelylle joko (yleensä konkreettiseen) populaatioon P tai käsitteelliseen "superpopulaatioon" (Universe) U :

Sample \Rightarrow Population \Rightarrow Universe.

Merkintä $x \Rightarrow y$ taulun alaosassa sensijaan kuvaa muuttujien (tai muuttujaryhmien) x ja y välistä *kausaalisuhdetta*, ks. tarkemmin Wold (1967).

Kuvio 1. Ennustemallien tavoitteiden ja välineiden lipputaulu

	Kokeelliset havainnot	välimuoto	Ei-kokeelliset havainnot
Kuvaileva tavoite	A. Toistot vakio-olosuhteissa B. Toistot ovat riippumattomia Otanta: $S (\Rightarrow P) \Rightarrow U$ I	A tai B ei ole voimassa N	Vakio-olosuhteet (A) eivät vallitse Riippumattomuus (B) häiriintynyt Osittaminen Otanta: $S \Rightarrow P \Rightarrow U$ III
välimuoto	Selittävät mallit: riippumattomia toistoja vakio-olosuhteissa W		E Moniolotteiset ei-kausalliset mallit
Selittävä tavoite	II Kausalliset mallit: $x \Rightarrow y$ x kontrolloitu Testi-kontrolliryhmä-kokeet; Stimulus-response-kokeet. Jokaisessa ryhmässä: riippumattomia toistoja vakio-olosuhteissa	S $x \Rightarrow y$ jotkut x:t kontrolloitu, jotkut eivät Operaatio- tutkimus	IV Kausalliset mallit: $x \Rightarrow y$ X ei kontrolloitu Muita vaikutuksia: spesifiointivirheet.

Tyypillinen kuvaileva kysymys on tyyppiä "Mikä?" Se sisältää joukon muunnelmia, joista "Kuinka monta?", "Kuinka suuri?", "Kuinka usein?" ovat yleisimpiä. Selittävään tutkimusongelmaan liittyvä kysymys on tyyppiä "Miksi?", joka saa usein vastaukseksi kausallisen päättelyn.

Lipputaulun eri alueet on referoinnin helpottamiseksi nimetty luvuin tai ilmansuunnin, nimittäin (a) neljään ennustemallien pääkategoriaan:

- I tai NV: Kuvailevat kokeelliset ennustemallit
- II' tai SW: Selittävät kokeelliset ennustemallit
- III tai NE: Kuvailevat ei-kokeelliset selitysmallit
- IV tai SE: Selittävät ei-kokeelliset ennustemallit

ja (b) neljään välimuotoon, joita merkitään ilmansuuntasymbooleilla: W, N, E, S. Tämä jaottelu perustuu käsitteisiin kuvaileva, selittävä, kokeellinen. Nämä kaikki ovat jokapäiväisessä käytössä olevia termejä, jotka kuitenkin johtavat meidät syviin vesiin yrittäessämme antaa niille täsmällisiä määrittelyjä. Oleellinen kohta on, että lipputaulu pyrkii kausaalisen riippuvuuden operationaaliseen määritelmään, mikä on eräs tietoteorian ikimuistoisia kiistakysymyksiä. Samoin korostetaan, että ei-kokeellisen mallinrakennuksen laaja alue on suhteellisen alikehittynyt, erityisesti selittävien mallien tapauksessa, Wold (1967, s. 10).

Tilastollisen oppikirjansa Wold (1964) esipuheessa luonnehtii lipputauluaan seuraavasti:

"Den dubbla kategoriklyvningen av mål och medel ger en indelning av den tillämpade statistiken i fyra fält, eller rättare sagt fem, om man beaktar att gränsdragningen har flytande övergångar. Indelningen illustreras av ovanstående tablå, den s.k. *flaggtablån*. De fem fälten ger en strukturell indelning av den tillämpade statistikens problematik. Den grundläggande betydelsen av indelningen enligt flaggtablån framgår av att de fyra hörnfälten markerar en vetenskaplig evolution: beskrivande frågeställningar är lättare att behandla än förklarande, och statistiken har lättare att komma tillrätta med problemen i experimentella situationer än när kontrollerade experiment ej är möjliga att genomföra. Denna evolution återspeglas i statistikens historia, i det att genomarbetningen av de fyra fälten markerar successiva faser i utvecklingen".

Tilastollisten menetelmien virheellistä käyttöä esiintyy erityisesti IV neljänneksessä, jossa kausaaliseen selitykseen pyritään ei-kokeellisen aineiston perusteella. Tällaisessa tilanteessa usein esiintyvä virhemahdollisuus on käyttää päättelymenetelmiä, jotka sellaisenaan soveltuisivat vain kokeelliseen aineistoon. Tämän kaltaista virhepäättelyä voi olla vaikeata havaita, koska yksinkertaisemmassa kokeellisessa tilanteessa

päätely on aivan pätevää. Tulemme seuraavassa analysoimaan näitä kysymyksiä konkreettisten esimerkkien avulla. Tarkastellut ongelmat eivät suinkaan rajoitu käsilläolevaan tutkimukseen, vaan kysymys on yleisestä ja samalla vaikeasta tieteellisen päättelyn problematiikasta.

Tilastotieteen metodikirjat jakaantuvat usein seuraaviin osiin:

- havaintoaineiston hankinta (otanta, koesuunnittelu),
- havaintoaineiston muokkaaminen ja kuvaaminen (taulukointitekniikka, tunnuslukujen laskeminen),
- populaation parametrien estimointi (estimaattorit, luottamusvälit),
- matemaattisten mallien konstruointi (todennäköisyyslaskenta),
- matemaattinen tilastotiede,
- hypoteesien testaaminen (testiteoria),
- tulosten raportointi.

Eräät mainituista ovat pääasiallisesti kuvailevia luonteeltaan, mutta esim. tilastollinen päätely (siis estimointi ja testaus) sisältää sekä kuvailevia että selittäviä kysymyksenasetteluja.

Tästä johtuu, ettei esim. tilastollisen testin avulla todettu "erittäin merkitsevä" korrelaatio kahden muuttujan välillä suinkaan osoita, että muuttujien välillä vallitsisi syy-seuraus- suhde suuntaan tai toiseen. Kummankaan muuttajan ei tarvitse toimia toisen vaihteluiden "asiallisena" selittäjänä, vaan havaitulla korrelaatiolla on usein vain kuvaileva merkitys. Lyhyesti sanottuna: oikein suoritettun tilastollisen päättelyn mukaisella tuloksella ei välttämättä ole kausaalista tulkintaa.

Näennäiskorrelaation erottaminen syy-seuraus- suhteesta (tai johonkin teoriaan perustuvasta selityksestä) ei ole varsinaisesti tilastotieteen alaan kuuluva tehtävä, vaan edellyttää ko. sovellutusalan asiantuntemusta. Kyseessä on vaikea ongelma, jota käsitellään esim. seuraavien otsakkeiden alla: näennäiskorrelaatio (spurious correlation), mallien virhespesifikaatio, ennuste/selitysmallit, poisjätettyjen muuttujien vaikutukset.

Tietojen hankkiminen kokeellisen tutkimuksen avulla poikkeaa oleellisesti ei-kokeellisesta havaintoaineiston keräämisestä. Kokeella pyritään selvittämään määrättyjen "käsittelyjen" vaikutuksia eri koeyksiköiden jotakin ominaisuutta kuvaavaan muuttujaan y : tutkijalla on mahdollisuus kontrolloida käsittelyjen eri tasoja ja usein hän voi valita koeyksiköt. Erityisesti Fisherin (1935) kehittämän koesuunnittelun yleiset periaatteet ovat seuraavat (ks. Vasama ja Vartia (1970, s. 125-132)):

(1) *Satunnaistaminen*; eri menetelmän tai saman menetelmän eri tasoja sovelletaan koeyksiköihin satunnaisesti. Satunnaistamisella varmistetaan siitä, etteivät mitkään systemaattiset häiriötekijät dominoi koetilannetta.

(2) *Toistaminen*; koe suoritetaan useita kertoja tai tutkittavia menetelmiä kohdistetaan useaan eri koeyksikköön samanaikaisesti satunnaisten tekijöiden vaikutusten pienentämiseksi ja toisaalta satunnaisvaihteluiden suuruusluokan määrittämiseksi.

(3) *Lohkojen muodostaminen*; lohkojen muodostamisella tarkoitetaan koeyksiköiden ryhmittelyä homogeenisiin osajoukkoihin joidenkin (tutkimuksen kannalta relevanttien) muuttujien mukaisesti. Lohkojen muodostamisella pyritään vähentämään osajoukkojen sisäistä vaihtelua, joka saattaa peit-

tää tutkittavista menetelmistä aiheutuvan vaihtelun (osajoukkojen välisen vaihtelun).

Ei-kokeellisessa tutkimuksessa tutkija voi vain passiivisesti tarkkailla tutkimuskohdetta ja mahdollisesti erilaisten otantamenetelmien avulla kerätä tarvitsemaansa havaintoaineistoa.

Työkentän IV neljännes, jossa ei-kokeellisen havaintoaineiston perusteella pyritään selittävään tutkimustavoitteeseen, vastaa tämän tutkimuksen tilannetta, joten tutkitaan sen problematiikkaa seuraavassa kappaleessa tarkemmin.

4.1.1. Selittävä tutkimus ja ei-kokeellinen havaintomateriaali

Selittävässä tutkimuksessa analysoidaan muuttujien välistä mahdollista kausaalista yhteyttä. Tutkimuksen kohteena on syy-seuraus-suhde, ks.

Wold (1964, s. 39-43) :

x	→	y
syy		seuraus
esim: koulutus		tulo

Woldin (1964, s. 40-41) mukaan ei-kokeelliseen havaintomateriaaliin perustuvalla kausaalianalyysillä on kaksi huomattavaa eroa verrattuna kokeelliseen tilanteeseen.

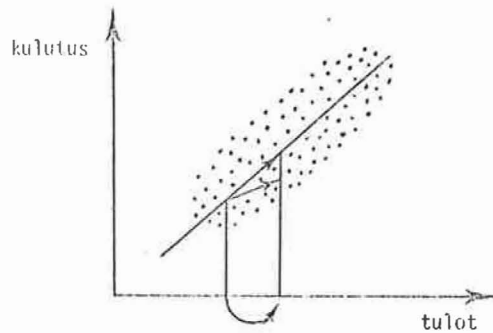
(1) Kokeellisessa tutkimuksessa kokeen suorittaja voi tarkoituksellisesti vaihdella syy-muuttujaa. Ei-kokeellisessa tutkimuksessa on syy-muuttuja vain passiivisen rekisteröinnin kohde. Vaikka tämä ero on useassa suhteessa merkityksellinen, ei sen saisi antaa hämärtää sitä seikkaa, että kausaalihypoteesilla on sama tietoteoreettinen sisältö sekä kokeellisessa että ei-kokeellisessa tilanteessa. Ero on merkittävä mm. itse kausaalihypoteesin asettamiselle. Kokeellisessa tutkimuksessa itse kokeen rakenteesta selviää, mikä on syy- ja mikä on seurausmuuttuja. Ei-kokeellisessa tutkimuksessa on kausaaliyhteys motivoitava muutoin, esimerkiksi olemassa olevaan teoriaan vedoten.

(2) Kokeellisessa tutkimuksessa seurausmuuttujaan vaikuttavat muutkin seikat kuin kokeessa suoritettu "käsittely", mutta näitä tekijöitä voidaan yrittää pitää niin vakioina kuin mahdollista. Ei-kokeellisessa tutkimuksessa taasen seurausmuuttujaan vaikuttavat useat seikat, jotka eivät ole kontrolloituja kokeellisessa merkityksessä. Esimerkiksi tutkimuksen kohteena olevat henkilöt voivat asua kaupungissa tai maaseudulla. Muiden seurausmuuttujaan vaikuttavien tekijöiden olemassa olo vaikuttaa analyysimetodin valintaan. Kokeellisessa tilanteessa voidaan satunnais-taminen suorittaa ennen itse koetta. Ei-kokeellisessa tilanteessa ei voida satunnaistaa ja siksi siinä täytyy ottaa eksplisiittisesti huomioon ne tekijät, jotka voivat vaikuttaa häiritsevästi analyysiin. Havaintoaineistoa analysoitaessa yritetään siis jälkikäteen eliminoida tai arvioida häiriötekijöiden vaikutukset. Näin esimerkiksi tapahtuu monimuuttujaregressioanalyysiä käytettäessä, jolloin näitä häiriötekijöitä on mukana selittäjinä.

Tarkastellaan lyhyesti eräitä tapauksia, joissa ei-kokeellisen aineiston perusteella voidaan tehdä virheellisiä johtopäätöksiä.

A. Väite A1: "Keskimääräistä suurituloisemman kuluttajan (esim. kotitalouden) kulutusmenot tavallisesti ovat myös keskimääräistä suuremmat".

Kuvio 2:

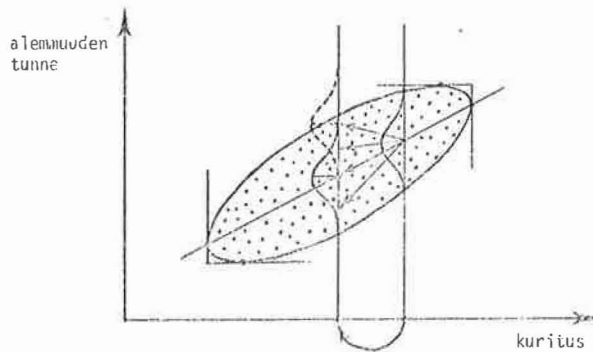


Väite A2: "Tulon lisääminen nostaa kulutusmenoja".

Yleensä katsotaan, että sekä A1 että A2 pitävät paikkansa. Usein ilmeisesti päätellään seuraavasti: "Koska A1 niin A2". Tämä päättely ei ole loogisesti pätevä, vaan on aivan mahdollista, että tulojen noustessa (joku) kuluttaja vähentää kulutustaan. Väitteet A1 ja A2 liittyvät eri tilanteisiin: A1 esim. tietyn vuoden poikkileikkaustietoihin ja A2 esim. kahden peräkkäisen vuoden tai hypoteettisen tilanteen vertailuun. Vain kielellisellä tasolla nämä tilanteet ovat samantapaisia.

B. Väite B1: "Lapsuudessaan runsaasti ruumillista kuritusta saaneet henkilöt kärsivät tavallista enemmän alemmuudentunteesta".

Kuvio 3:



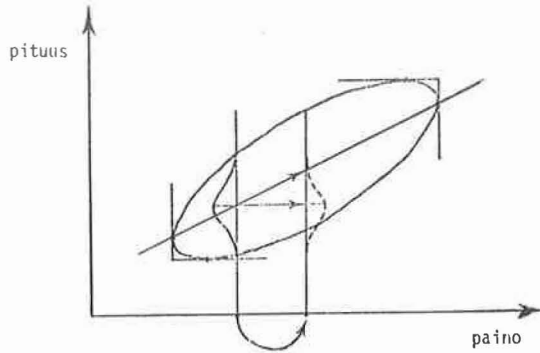
Väite B2: "Jos kuritusta olisi käytetty lapsuudessa vähemmän, ei poikkeuksellista alemmuudentunnetta olisi myöhemmin esiintynyt".

Väite B3: "Lapsuudessa saadun kurituksen vähentämisellä (tai lisäämisellä) ei ole oleellista vaikutusta henkilön itsetuntoon".

Yleinen käsitys lienee, että B1 tms. pitää paikkansa. Kiistaa käydään onko B2, B3 tai jokin muu vastaavanlainen väite tosi. Näin ollen kumpikaan implikaatioista " $B1 \Rightarrow B2$ " tai " $B1 \Rightarrow B3$ " ei ole ainakaan loogisesti tosi väite.

C. Väite C1: "Keskimääräistä painavampi täysi-ikäinen henkilö on yleensä myös keskimääräistä pitempi".

Kuvio 4:



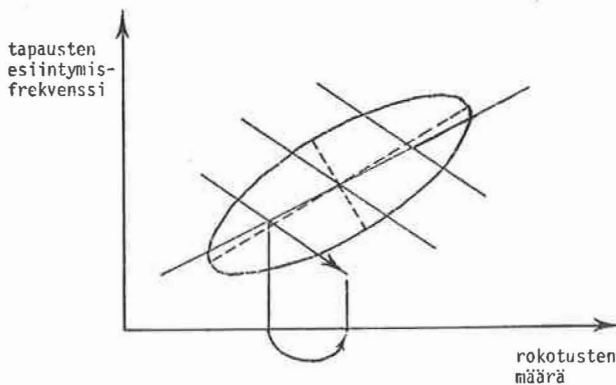
Väite C2: "Täysi-ikäisen painoa lisäämällä lisätään myös pituutta".

Väite C3: "Täysi-ikäisen painon lisääminen ei muuta pituutta".

Väite C1 on tosi, C2 on biologisten tosiasioiden perusteella väärä ja C3 tosi.

D. Väite D1: "Tietyillä alueilla todettujen isorokkotäpausten ja kuu-
kautta aiemmin suoritettujen rokotusten välillä on todettu tilastolli-
sesti erittäin merkitsevä positiivinen korrelaatio".

Kuvio 5:



Väite D2: "Isorokkorokotusten saamisen ja isorokkoon sairastumisen välillä on positiivinen korrelaatio".

Väite D3: "Rokottaminen lisää isorokkoon sairastumisen riskiä".

Väite D4: "Rokottaminen aiheuttaa isorokkoa".

Väite D5: "Jos isorokkorokotuksia olisi suoritettu aiempaa enemmän, olisi sairastumisien määrä lisääntynyt".

Väite D6: "Jos rokotuksia olisi suoritettu aiempaa enemmän, olisi sairastumisien määrä vähentynyt".

Väite D7: "Rokottamisella on isorokkotartuntaa ehkäisevä vaikutus".

Koska rokotuksiin ryhdytään tavallisesti vain kun tautitapauksia on havaittu (eli epidemia on lähtenyt liikkeelle), pitää D1 ilmeisesti paikkansa. D2 on niin epämääräisesti formuloitu, ettei sen totuudesta voi sanoa sitä eikä tätä. Väitteet D3, D4 ja D5 ovat lääketieteen tutkimustulosten perusteella selvästi virheellisiä, D6 sekä D7 ovat tosia.

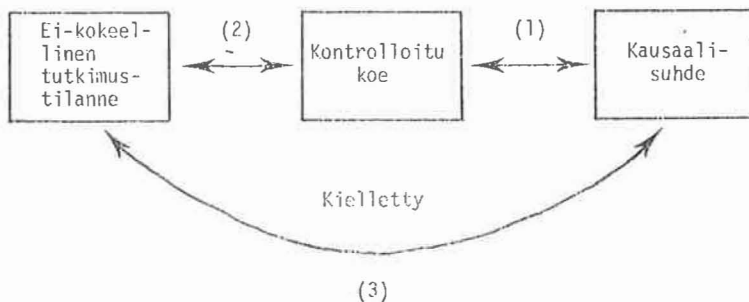
Normaali käytäntö esim. lääketieteessä on perustella kontrolloitujen kliinisten kokeiden avulla, että tietyllä toimenpiteellä on toivottu vaikutus. Vastaavaan menettelyyn tulisi pyrkiä myös taloustieteessä. Usein kuitenkin kontrolloitujen kokeiden suorittaminen on hyvin kallista tai käytännössä lähes mahdotonta tarkastellun ilmiön (esim. kulutuksen ja tulojen riippuvuuden) laadun vuoksi. *Tällöin tulisi*

ei-kokeellisen aineiston avulla yrittää mahdollisimman luotettavasti arvioida, millaisia tuloksia kuvitteellisessa kontrolloidussa kokeessa saataisiin, jos sellainen vaikeuksista huolimatta suoritettaisiin.

Tämä on ideana esim. ei-kokeellisen aineiston regressioanalyysissä, jossa häiriömuuttujien mukaanottamista malliin joskus kutsutaan tosin hieman harhaanjohtavasti "niiden kontrolloimiseksi" - suora viittaus vastaavaan koetilanteeseen.

Siis lyhyesti sanoen: Ei-kokeellisen aineiston perusteella tulisi yrittää ennustaa, mitä vastaavassa koetilanteessa olisi tapahtunut. Tällä tavoin syy-seuraus-analyysi välillisesti (siis käsitteellisesti) liitetään johonkin koetilanteeseen. Näin toimimalla havaitaan, kuinka paljon epävarmimmalla pohjalla ei-kokeellinen kausaalianalyysi on kokeellista kausaalianalyysiä. Ei-kokeellisessa kausaalianalyysissä ei siis saisi kategorisesti väittää, että vaikutus on tietynlainen vaan, että eräässä suorittamattomassa kokeessa saataisiin tietynlaisia tuloksia. Ei-kokeellisessa tutkimustilanteessa suoritettu regressioanalyysi on koetilanteen analysoinnin *korvikeratkaisu*. Seuraava kuvio pyrkii havainnollistamaan esitettyä menettelytapaa.

Kuvio 6:



Symbolilla (1) merkitty nuoli kuvaa kokeellisen tilanteen päättelyä, jossa kontrolloidun kokeen perusteella päätellään kausaalisen riippuvuuden olemassaolo. Symbolilla (2) merkitty nuoli kuvaa edellä ehdotettua tutkimusstrategiaa, jossa ei-kokeellisten havaintojen perusteella yritetään päätellä, mitä vastaavassa kontrolloidussa kokeessa olisi tapahtunut. Nuoli (3) vastaa tavanomaista yritystä, jossa ei-kokeellisen aineiston perusteella yritetään suoraan päätellä kausaalisia vaikutussuhteita. Tämän lähestymistavan suurien virhemahdollisuuksien vuoksi on ko. tutkimusstrategia "kielletty".

Ehdotetulla tutkimusstrategialla on useita etuja traditionaaliseen menettelytapaan verrattuna.

1. Menetelmä paljastaa empiirisesti tyhjät ("metafyysiset") väitteet, koska niiden paikkansapitävyyden selvittämiseksi ei voida keksiä kontrolloitua koetta (eikä muutakaan menettelyä)
2. Empiirisesti mielekkäät, mutta vaikeasti kokeellisesti tutkittavat ongelmat erottuvat omaksi erityisryhmäkseen. Näiden tutkiminen ei-kokeellisen aineiston perusteella vaatii erityistä asiantuntemusta ja huolellisuutta
3. Tutkimusongelma, joka on periaatteessa yksinkertaisesti selvitettävissä kokeellisesti, voi ei-kokeellisen datan tapauksessa olla
 - (a) helppo
 - (b) hankala
 - (c) erittäin vaikea
 - (d) mahdoton

ratkaista, ongelman ja ei-kokeellisen datan laadusta riippuen. Näin tutkimusongelmat jakaantuvat hierarkkisesti vaikeusasteen mukaisiin ryhmiin.

Ehkä on paikallaan esittää esimerkit kustakin tapauksesta:

1. "Ihmisen kohtalosta päättää havaitsematon henkiolento"
(Empiirisesti mieletön väite)
2. "Maan kiertoradan keskietäisyys auringosta olisi nykyistä pienempi, jos auringon massa olisi suurempi"
(Kokeellisesti mahdoton suoraan selvittää)
- 3a. Metallista valmistetun 3 metrin kiskon pituuden lämpölaajenemis-kerroin on määrättävä 120 päivittäin suoritetun pituus-, ilmanpaine- ja lämpötilamittauksen perusteella. Kisko on ulkoilmassa, mittaukset ovat tarkkoja. (Yksinkertainen regressioanalyysin tehtävä: ilmanpaineen regressiokerron osoittautuu nolllaksi.)
- 3b. Tehtävä 3a, kun kiskon lämpötila on pidetty 0.1 asteen tarkkuudella vakiona.
(Selittäjien kovarianssimatriisi lähes singulaarinen, koska lämpötila lähes vakio. Erikoistapaus multikollineaarisuusongelmasta.)
- 3c. Tehtävä 3b, kun lämpötilamittaus sisältää satunnaisvirheitä.
(Multikollineaarisuus + mittausvirheitä selittäjissä.)
- 3d. Tehtävä 3a, mutta lämpömittarin lukemat eivät koske kiskoa vaan ulkoilmaa; kiskon lämpötila vaihtelee ulkoilmasta riippumatta.
(Epärelevantit selittäjät.)

4.2. Tutkimusongelman tarkastelumahdollisuuksia

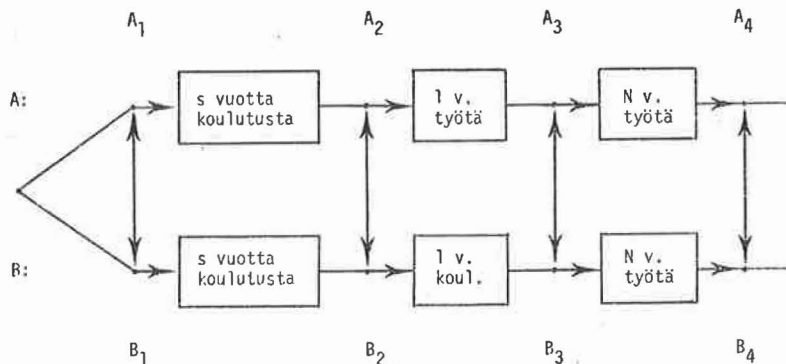
Ei ole suinkaan yksinkertaista vastata kysymykseen, miten kotitalouden päämiehen koulutuksen vaikutusta kotitalouden käytettävissä olevaan tuloon voitaisiin tieteellisesti tutkia. Kontrolloidun kokeen suorittaminen asian tutkimiseksi on käytännössä mahdotonta, eikä todellisuudessa tuloihin vaikuttavien muiden tekijöiden vaikutuksia voida täysin eliminoida. Asian havainnollistamiseksi tarkastellaan tieteellistä kiistaa, jota käytiin 1950-luvulla tupakanpolton vaarallisuudesta. Erityisesti tutkittiin, miten tupakanpoltto vaikuttaa keuhkosyöpäsairastuvuuteen. Esimerkiksi Richard Doll ja A. Bradford Hill, ks. Brown (1972, s. 40-51), suorittivat laajan haastattelututkimuksen englantilaiselle sairaanhoitohenkilöstölle saaden noin 40 000 vastausta. Haastattelussa tiedusteltiin henkilön tupakanpolttotapoja ja muutamia muita ominaisuuksia, kuten ikää ja sukupuolta. Haastateltujen terveydentilaa seurattiin useiden vuosien ajan ja tuloksena havaittiin, että tupakoivilla oli huomattavasti enemmän keuhkosyöpätapauksia kuin tupakoimattomilla. Paljon tupakoivien keuhkosyöpäkuolleisuus oli 1.66/1000 henkilöä vuodessa, kun tupakoimattomien vastaava luku oli 0.07/1000 henkilöä vuodessa. Tupakoivien keuhkosyöpäkuolleisuus oli näin ollen 24 kertaa suurempi kuin tupakoimattomien. Tämä tutkimus tulkittiin erittäin vakuuttavaksi lisätodisteeksi tupakoinnin vaarallisuudesta.

Varsinaiseksi kiistaksi tupakanpoltto-keuhkosyöpätutkimuksissa nousi (tupakkateollisuuden toimesta) kuitenkin kysymys, miten valiideja johtopäätöksiä voidaan tehdä tutkimuksessa, jossa on useita kontrolloimattomia tai vain osittain kontrolloituja epätoivottuja vaihteluja aiheuttavia tekijöitä datassa. Tupakkateollisuus sai puolustajikseen kaksi huomatta-

vaa tilastotieteilijää: R.A. Fisherin, lukuisten modernien tilastomenetelmien kehittäjän ja Joseph Berksonin, tilastotieteen ja lääketieteen erikoistuntijan. Fisher (1959) ja Berkson (1958) toteavat painokkaasti, että tupakan vaarallisuutta koskeva todistusaineisto on vain näennäistä. He osoittavat heikkouksia tutkimusten todistusaineistossa ja puutteita niiden tilastollisessa päättelyssä. Havaittujen säännöllisyyksien voitiin selittää johtuvan erilaisista kontrolloimattomista tekijöistä yhtä hyvin kuin tupakoinnista. Esimerkiksi jos perinnöllisesti tupakointiin taipuvalla henkilöllä on perinnöllinen taipumus sairastua keuhkosityöpään, niin saataisiin vastaavia "todisteita" tupakoinnin vaarallisuudesta, vaikka tupakanpoltolla ei olisikaan vaikutusta keuhkosityöpäsairastuvuuteen. Tupakoinnin lopettaminen ei vähentäisi sairastumisriskiä henkilöllä, jolla keuhkosityövän sairastumistodennäköisyys on perinnöllisesti korkea. Tällöin ei tupakointi vaikuttaisi haitallisesti terveyteen. (Fisher ja Berkson siis muistuttivat, että keuhkosityöpä tutkimusten aineistossa tilanne *voi vastata* edellisen luvun C-tapausta. Vasta myöhemmät tutkimukset ovat osoittaneet, että riippuvuus on kuin onkin A-tapauksen mukainen). Fisher (1959) ehdotti kiistan selvittämiseksi identtisten kaksosten tutkimista, jolloin verrattavien henkilöiden geneettinen perimä olisi samanlainen.

Fisherin esittämä koejärjestely olisi tieteellisesti parhain tapa tutkia myös kotitalouden päämiehen koulutuksen ja kotitalouden käytettävissä olevan tulon yhteyttä. Identtiset päämiehet saisivat määrätyn käsittelyn (tässä: tietyn koulutuksen), jonka vaikutusta kotitalouden käytettävissä olevaan tuloon selvitettäisiin. Suoritetaan ajatuskoe, jossa seurataan kahden identtisen henkilön A ja B elämäntapaa. Tarkastellaan kuvion 7 avulla A:n ja B:n koulutuksen aiheuttamia eroavaisuuksia erityisesti kotitalouksien tuloissa.

Kuvio 7: Ajatuskoe: identtiset kotitalouksien päämiehet



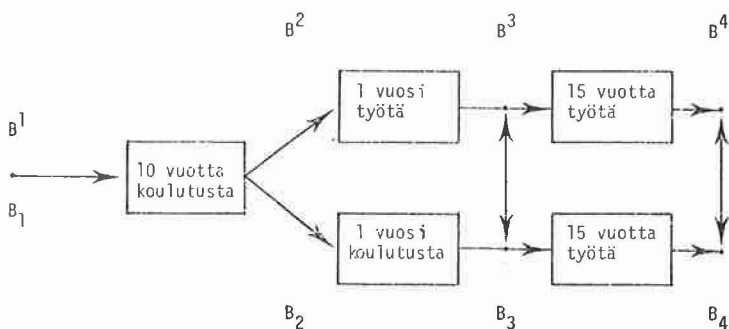
Henkilöt A ja B ovat kuvion 1-vaiheessa identtisiä: heillä on sama menneisyys, perintötekijät jne. Toisessa vaiheessa tapahtuu muutos: A siirtyy työelämään ja B jatkaa opintoja. Kolmannessa vaiheessa A:lla ja B:llä on jo selviä eroja: A₃ on saanut työkokemusta ja B₃ lisäkoulutusta. Tässä vaiheessa A:n tulot voivat hyvinkin olla B:tä suuremmat. Sensijaan muutaman vuoden kuluttua tilanne on toisin päin. Tämän ajatuskokeen mukaisessa tilanteessa voidaan periaatteessa joka vaiheessa selvittää, kannattaako vaihtaa vuosi työtä vuoteen koulutusta. Vertailemalla A:n ja B:n tuloja ongelmaa koulutuksen tulovaikutuksista voitaisiin lähestyä.

Mikäli esitetty ajatuskoe olisi mahdollista suorittaa käytännössä olisi kotitalouden päämiehen koulutuksen tulovaikutus helposti tutkittavissa. Mutta tällaista koejärjestelyä on lähes mahdoton toteuttaa, joten on tyydyttävä ei-kokeelliseen havaintomateriaaliin, jossa useat kontrolloimattomat tekijät vaikuttavat tuloksiin. Näin törmätään samoihin tilastollisen päättelyn ongelmiin kuin esitettyssä tupakanpoltto-keuhkosityöpätutkimuksessa. Nähdäksemme selvemmin seuraavissa luvuissa esiteltävän havain-

tomateriaalin ja tilastollisen analyysin mahdolliset virhelähteet, pohditaan kuinka tilastollinen päättely voitaisiin suorittaa esitellyn ajatuskokeen tapauksessa.

Tarkastellaan henkilöä B, jolla on 10 vuoden pohjakoulutus, jonka jälkeen hän joutuu valitsemaan vuoden lisäkoulutuksen ja työhön siirtymisen kesken. Vain toinen vaihtoehto on luonnollisesti mahdollinen.

Kuvio 8: Kuvitellun koehenkilön B valintatilanne



Valitessaan vuoden lisäkoulutuksen B seuraa elämänskaarta $B_1-B_2-B_3-B_4$, kun taas työhön siirtymistä vastaa polku $B^1-B^2-B^3-B^4$. Vuoden lisäkoulutuksen vaikutukset ilmenevät toisiaan vastaavien tilanteiden (B_3, B^3) ja (B_4, B^4) eroina. Vuoden lisäkoulutusta voidaan pitää edullisena vain jos "sen seurauksena" syntyneet tilanteet B_3, B_4, \dots ovat edullisempia kuin mitä olisi tapahtunut ilman lisäkoulutusta, ts. jos B olisi seurannut ylempää polkua. Käytännössä vain yhtä elämänskaarta kuljetaan, joten edellä sanottu on varsin pitkälle menevää "jossittelua". Mutta valintatilanteissa eri vaihtoehtojen edullisuuden vertailu perustuu juuri päätösvaihtoehtoja vastaavien tulevaisuudenkuvien ("elämänskaarien") vertailuun, kuten esim. Törnqvist ja Nordberg (1968) esittävät. He kutsuvat päättämisprosessia *aidoksi*, jos se sisältää seuraavat viisi vaihetta:

1. päätöstilanteen tutkiminen, toisin sanoen menneisyyden kuvan terävöittäminen ja sen sisältämän informaation ("sanoman") selventäminen
2. eri päätösmahdollisuuksien selvittäminen
3. eri päätösmahdollisuuksia vastaavien tulevaisuudennäkymien muodostaminen
4. eri päätösmahdollisuuksia vastaavien tulevaisuudennäkymien arvostaminen ja vertailu
5. päätöksen valinta eri päätösmahdollisuuksista.

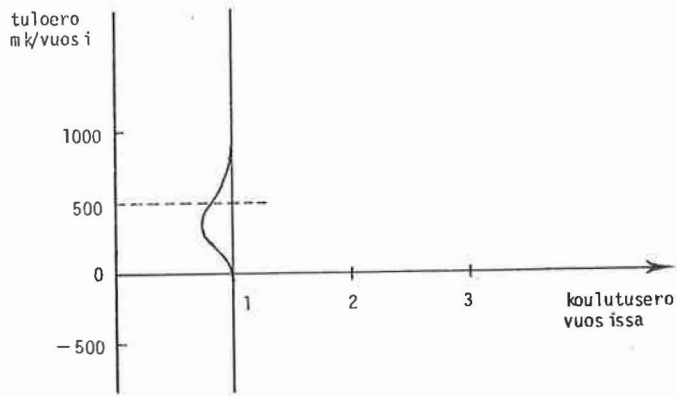
Tieteellisen tutkimuksen tehtävä liittyy erityisesti kohtaan 3, eri päätösmahdollisuuksia vastaavien tulevaisuudennäkymien muodostamiseen.

Edellä on pyritty korostamaan, että luotettavimmat "päättövaihtoehtoja vastaavat tulevaisuudennäkymät" (ts. "eri käsittelyjä vastaavat vaikutukset") muodostetaan kokeellisen tutkimuksen avulla. "Tekemällä kokeita, joissa päätös toteutetaan ainoastaan pienessä mittakaavassa, yrite-tään kerätä tietoa eri päätösten odotettavissa olevista seurauksista", Törnqvist ja Nordberg (1968, s. 14).

Mahdollisimman yksinkertainen koejärjestely perustuisi identtisten henkilöiden käyttöön. Tarkastellaan tätä seuraavassa lähemmin.

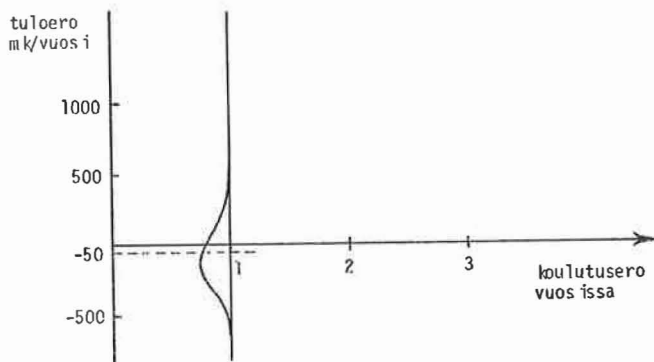
Oletetaan, että kuvion 8 tilannetta vastaava koe olisi suoritettu identtisillä kaksosilla ja toista arvottua kaksosparikkia olisi koulutettu vuoden enemmän kuin toista. Useiden kaksosparien tulokset kuvion 8 vertailukohdassa (B_4, B^4) ilmeisesti jakautuisivat suurin piirtein seuraavan kuvion mukaisesti.

Kuvio 9: Kaksosparikkien tuloerot vertailukohtassa (B_4, B^4)



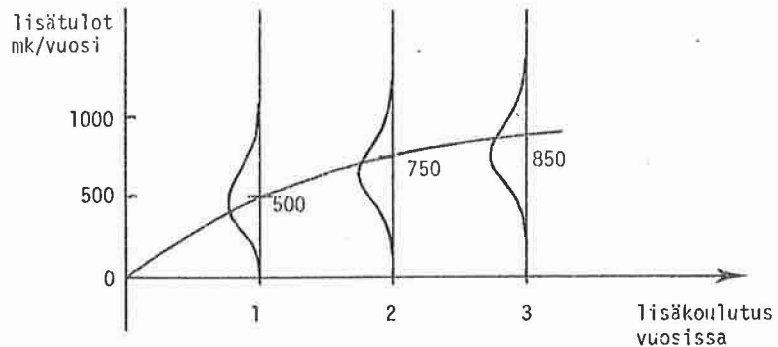
Kaksosten kuukausitulojen ero on kuviossa keskittynyt (hatusta temmatun) 500 mk:n lähetyville: vuoden enemmän koulutusta saaneen kaksosen vuosi-ansiot ovat 15 vuoden työkokemusvuoden jälkeen keskimäärin 500 mk suuremmat kuin vuotta vähemmän koulutusta saaneen (mutta 16 vuotta työssä olleen) kaksosen. Heti koulutuksen jälkeen kaksosia verrattaessa tilanne voisi olla toinen:

Kuvio 10: Kaksosparikkien tuloerot vertailukohtassa (B_3, B^3)



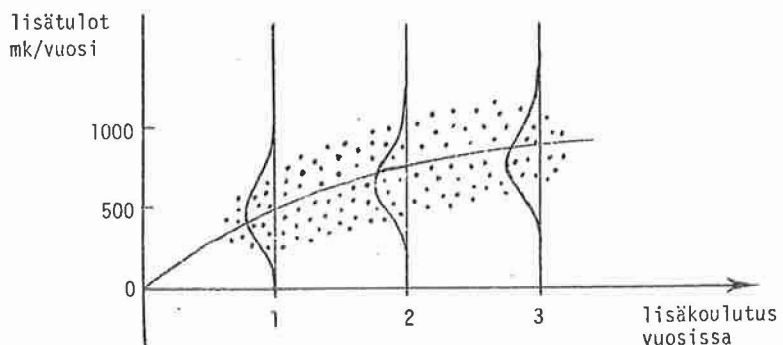
Vuoden enemmän koulutusta saaneiden kaksosten vuosiansiot ovat keskimäärin 50 mk pienemmät kuin vuoden työkokemusta hankkineella kaksosparikilla. Vertailutulokset riippuvat todennäköisesti juuri tähän tapaan valitusta ajankohdasta. Laajennetaan kuviteltua kaksostutkimusta ja oletetaan, että identtisillä kaksosilla olisi suoritettu koe, jossa toista arvottua kaksosparikkia olisi koulutettu vaihtoehtoisesti 1, 2 tai 3 vuotta enemmän kuin toista. Tuloterot voisivat tällöin jakaantua seuraavasti.

Kuvio 11. Kaksosparikkien tuloerot vertailukohdassa (B_4, B^4)



Edellä on pidetty keinotekoisesti kaksosparikkien väliset koulutuserot vuoden monikertoina. Käytettäessä kuukauden välein vaihtelevia koulutusajan eroja saataisiin realistisempi, kuvion 12 kaltainen korrelaatiodiaagramma.

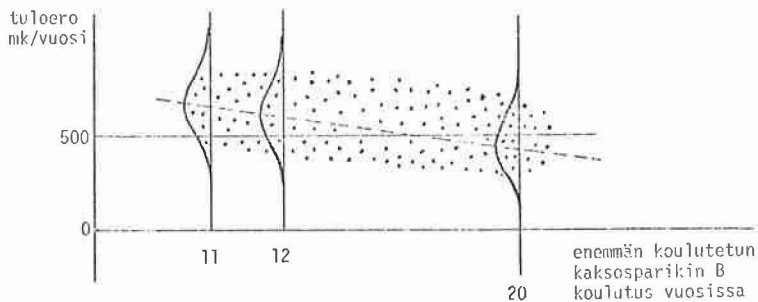
Kuvio 12: Kaksosparikkien tuloerot vertailukohdassa (B_4, B^4)



Näissä kuvioissa on lisäkoulutuksesta johtuvia tuloeroja tarkasteltu sellaisena kuin ne 15 vuoden sopeutumisen jälkeen ilmenevät.

Analysoidaan tarkemmin kuvion 9 tilannetta, jossa kaksosilla oli vuoden koulutusero. Taulukoimalla tuloerot enemmän koulutusta saaneen kaksosparikin koulutuksen mukaisesti voitaisiin saada seuraavanlainen kuvio:

Kuvio 13: Kaksosparikkien tuloerot vertailukohtassa (B_4, B^4)



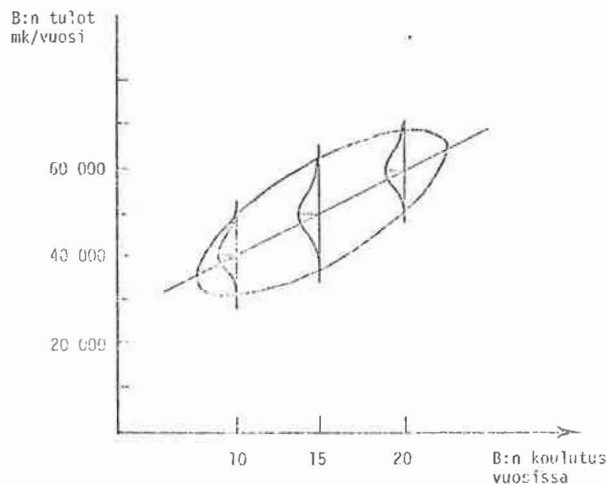
Vuoden lisäkoulutus tuottaisi kuvion 9 mukaan keskimäärin 500 markan tuloeron vuodessa. Kuvion 13 mukaan tuloero pienenisi kaksosparikkien peruskoulutuksen kasvun myötä. Vaikka vuoden lisäkoulutus vaikuttaa tuloeroihin sitä vähemmän, mitä suurempi lähtökohtana oleva peruskoulutus on, niin koulutuksen lisäys aiheuttaa eo. tilanteessa positiivisen tulonlisäyksen kaikilla koulutustasoilla.

Edellä kuvatun, identtisillä kaksosilla suoritetun (hypoteettisen) koulutuskokeen tuloksien voitaisiin katsoa lopullisesti ratkaisevan, kuinka paljon ko. olosuhteissa lisäkoulutus nostaa (tai laskee) tuloja. Esitetyissä kuvioissa on koko ajan tarkasteltu identtisten kaksosten A ja B välistä tuloeroa, joka on havaittu syntyneen satunnaisesti valitun B:n saadessa koulutusta tietyn määrän A:ta enemmän. Tämä lisätulo on esitet-

ty edellä joko lisäkoulutuksen (kuviot 9, 10, 11 ja 12) tai kuviossa 13 B:n koulutuksen mukaisesti. Vastaavalla tavalla voidaan periaatteessa selvittää muitakin koulutuksen vaikutuksia (esim. perhekokoon, elinikään, asuinpaikkaan, sanavarastoon jne.).

Mutta identtisten kaksosten koetuloksista voitaisiin laatia useita muitakin enemmän tai vähemmän hyödyllisiä korrelaatiodiagrammeja. Jos esimerkiksi luokiteltaisiin enemmän koulutusta saaneen kaksosparin B tulot tämän koulutuksen mukaan, saataisiin seuraavanlainen kuvio.

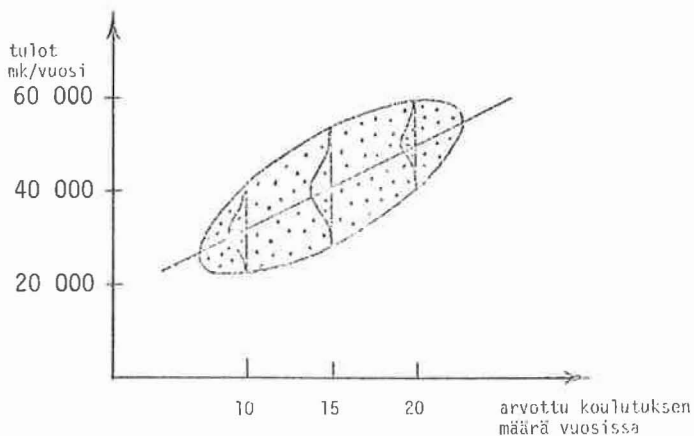
Kuvio 14: Enemmän koulutusta saaneiden kaksosparikkien koulutus ja tulot



Kuviosta 14 ei voida päätellä paljoakaan, kuinka lisäkoulutus vaikuttaa tuloihin, sillä vain toinen kaksosista, B, on mukana analyysissä. Kuviossa mukana olevat B:t ovat aivan erilaisia ihmisiä ja tämän näköisiä "tuloksia" saataisiin ilman kaksosiakin, ristiintaulukoimalla satunnaisesti valittujen henkilöiden tulot ja koulutus. Edellä on esitetty esimerkki sellaisesta kaksostutkimukseen liittyvästä aineiston erittelystä, joka ei mitenkään vie tutkimusta eteenpäin.

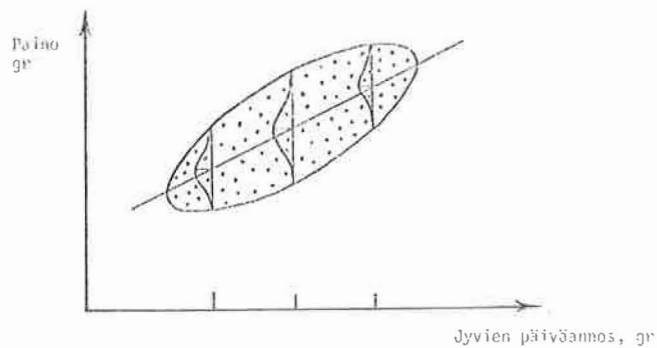
Minkäläaatusessa tilanteessa koulutuksesta aiheutuva tulonlisäys voitaisiin päätellä suoraan kuvion 14 näköisestä kuviosta? Milloin koulutuksen mukaan luokiteltujen tulojen erot harhattomasti kuvastavat vain koulutuksesta (eikä esimerkiksi perintötekijöistä, lahjakkuudesta, ympäristötekijöistä, tms.) johtuvia eroja? Yksinkertaisin ja selkein tilanne, jossa näin tapahtuisi, olisi ilmeisesti seuraavanlainen (äärimmäisen hypoteettinen) koetilanne. Ajatellaan tutkittavan 1000:n samaan aikaan syntyneen identtisen henkilön koulutuksen ja tulojen välistä riippuvuutta. Jokaisen henkilön koulutuksen määrä on arvottu tietyn suuruiseksi ja tulojen määrää tarkkaillaan. Kullakin henkilöllä olisi tällöin 999 dublikaattia, jotka eroaisivat toisistaan vain koulutuksen osalta. (Mainitunlaista monistusta tai kloonausta on suoritettu sammakoille ja tätä kirjoitettaessa huhutaan monistuksen onnistuneen myös ihmiselle). Tämänlaisen aineiston analysointi voitaisiin perustaa kuvioon 15.

Kuvio 15: 1000:n, kokeen alussa identtisen henkilön tulot arvotun koulutusmäärän mukaan



Luonnontieteissä mainitunlaisia dublikaatteja (toistoja) voidaan helposti aikaansaada. Esimerkiksi tutkittaessa kananpoikien painon ja ruokinnan välistä riippuvuutta voitaisiin gramman tarkkuudella valita samanpainoisia ja -ikäisiä kananpoikia, joita ruokittaisiin eri ravintomäärillä. Kokeen alussa voitaisiin kananpoikien katsoa olleen vielä "identtisiä". Käsitys ruokinnan vaikutuksesta kananpojan painoon saataisiin, kun näiden painot mitattaisiin esim. 20 päivän ruokinnan jälkeen.

Kuvio 16: 1000:n, kokeen alussa identtisen kananpojan paino arvotun jyväännoksen määrän mukaan

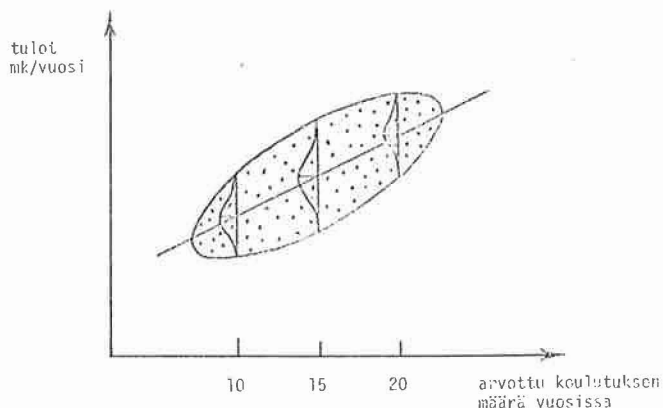


Vastaavanlaista koetta ihmisillä, kun tutkitaan henkisiä tai yhteiskunnallisia asioita, ei ole (ainakaan tällä hetkellä) mahdollista suorittaa. Ominaisuuksiltaan "identtisten" ihmisten 1000 kappaleen otosta ei saada generoitua koetutkimuskäyttöön. Tässä mielessä ihanteellista 1000 henkilön koeaineistoa ei siis voida käytännössä hankkia.

Merkitseekö tämä, ettei koulutuksen aikaansaamaa tulojen lisäystä voida kuvion 15 mukaisella tavalla ollenkaan selvittää? Ei suinkaan. Vaikka identtisiä henkilöitä ei voidakaan hankkia 1000 kappaletta, voitaisiin (luonnollisestikin koetutkimukselle myötämielisten viranomaisten suostumuksella) valita 1000 riittävän samankaltaista henkilöä esimerkiksi esikouluvaiheessa, arpoa näille koulutusmäärät ja seurata tulojen yms. teki-

jöiden kehitystä. Oletetaan, että näin on jonkin erityisjärjestelyn mukaisesti tehty ja tutkitaan, miten valittujen 1000:n koehenkilön on käynyt esimerkiksi 20 vuoden seurannan jälkeen. Koehenkilöiden tulojen ja koulutuksen korrelaatiodiagramma on esitetty seuraavassa kuviossa.

Kuvio 17: 1000:n, kokeen alussa riittävän samankaltaisen koehenkilön tulot arvotun koulutusmäärän mukaan



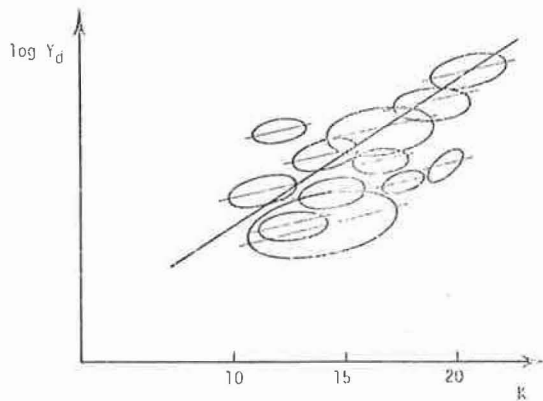
Olisiko kuvio 17 lukijan mielestä riittävä osoitus siitä, että koulutus lisää tuloja? Koehenkilöt ovat lähtökohdiltaan mahdollisimman samankaltaisia ja heille on arvottu eri koulutusmäärät.

Tilanne muistuttaa kaikissa oleellisissa suhteissa kananpoikien koetta. Edellä kuvatun koejärjestelyn tuottama riippuvuus (mikä se sitten olisikin) tulisi hyväksyä koulutuksen tulovaikutukseksi ko. yhteiskunnan olosuhteissa. Henkilöllä, joka katsoisi, ettei ao. aineisto oleellisilta osiltaan selvitä kysymystä, tulisi olla jotakin konkreettista esitettävänä mielipiteensä perusteluksi.

Siirrytään ajatuskokeista suomalaisen arkitodellisuuteen. Ihmiset ovat voineet arpomiseen verrattuna suhteellisen vapaasti valita koulutuksensa. Tämän vuoksi lukemattomat enemmän tai vähemmän systemaattiset tekijät (kuten lahjakkuus, rotu, äidinkieli, asuma-alue, vanhempien sosioekono-

minen asema jne.) aiheuttavat kontrolloimatonta variaatiota sekä koulutuksessa että tuloissa. Jos koulutuksen ja tulojen korrelaatiodiagramma piirrettäisiin suomalaisesta aineistosta, niin se olisi ilmeisesti seuraavan tapainen.

Kuvio 18: Vapaavalintainen koulutus ja tulot, korrelaatiodiagramma



Seuraavassa luvussa esitellään ns. kovarianssianalyysi, joka on erityisesti tämän tapaisiin heterogeenisiin aineistoihin soveltuva tutkimusmenetelmä. Ei-kokeelliseen havaintomateriaaliin sovelletun kovarianssianalyysin perusajatus on, että homogeenisissa ryhmissä (kuviossa 18: \circ) henkilöt ovat riittävän samankaltaisia ("melkein dublikaatteja"). Jos asianlaita on näin, niin kovarianssianalyysin antama kulmakerroin vastaa kuvion 17 keskimääräistä kulmakerrointa, joka kuvaa koulutuksen tulovaiikutusta. Ongelma on, missä määrin kuvion 18 ryhmiä voidaan verrata kuvion 17 tilanteeseen, toisin sanoen ovatko kuvion 18 ryhmät riittävän homogeenisia.

Tutkimuksen tarkoituksena on selvittää kotitalouden päämiehen koulutuksen vaikutusta kotitalouden käytettävissä olevaan tuloon. Kovarianssianalyysissä on tunnettava taustamuuttujat, joiden avulla kotitaloudet saadaan

luokiteltua riittävän homogeenisiin luokkiin. Saatavissa olevat tiedot asettavat omat rajoituksensa kovarianssianalyysin soveltamiselle, mutta mm. kotitalouksien erilaiset rakennetekijät, sosioekonomisen aseman erot ja sijaintitekijät voidaan ottaa analyysiin mukaan.

Kovarianssianalyysin avulla pyritään ei-kokeellisesta havaintomateriaalista tiivistämään melko työläällä tavalla vastaavat tiedot, jotka helpommin saataisiin koetilanteen havaintomateriaalista. Toisin kuin esimerkiksi oppikirjoissa esitetyt koulutusta koskevat yksinkertaistetut teoriat tai mallit ovat seuraavissa luvuissa esitettävät kotitaloustiedustelun tiedoista lasketut tulokset suomalaista arkitodellisuutta kuvaavia *tosiasioita*. Kunkin aineistosta estimoidun regressio- tai kovarianssianalyysimallin kertoimet määrätään kotitaloustiedustelun havaintojen perusteella, joten saadut kerroinestimaatit ovat tosiasioita kuvaavien havaintojen funktioina myös tosiasioita (toisin kuin jokin väärä teoria). Pulmana on saatujen tulosten tarkempi tulkinta. Vaikea kysymys esimerkiksi on, miten kotitaloustiedustelun tiedoista lasketut tulokset auttavat pääättelemään, miten identtisten kaksosten tapauksessa kävisi, jos toista koulutettaisiin enemmän. Tätä kysymystä tuskin voidaan täysin tyydyttävällä tavalla ratkaista, mutta mallien avulla kyllä voidaan simuloida tai arvioida, minkä suuntaisia tuloksia kaksoskokeessa todennäköisesti saataisiin. Tähän pyrittäessä on tilastollinen analyysitekniikka valittava käytettävissä olevan havaintoaineiston mukaisesti, jolloin aineiston tilastollinen analyysi tulee sitä mutkikkaammaksi mitä enemmän tarkasteltavat olosuhteet poikkeavat hypoteettisen koetilanteen ideaaliolosuhteista.

5. AINEISTON TILASTOLLINEN ANALYYSI JA MALLIT

5.1. Regressio- ja kovarianssianalyysi

Regressioanalyysissä tutkitaan, kuinka selittäjien (x-muuttujien) lineaarikombinaatio $\sum \beta_k x_k$ kuvaa y-muuttujan (selitettävän muuttujan) vaihtelua. Seuraavassa esitellään lyhyesti regressioanalyysin peruskäsitteet ja tärkeimmät tulokset. Hyviä johdatuksia aihepiiriin ovat esim. Draper ja Smith (1966), Johnston (1963) sekä Wonnacott ja Wonnacott (1970). Kuvataan muuttujan y riippuvuutta muuttujista x_1, \dots, x_K lineaarisella regressiomallilla. Jos muuttujista y ja x_1, \dots, x_K on N havaintoa, malli voidaan esittää muodossa:

$$(1) \quad \begin{array}{l} y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_K x_{1K} + \varepsilon_1 \\ \cdot \\ \cdot \\ y_N = \beta_0 + \beta_1 x_{N1} + \dots + \beta_K x_{NK} + \varepsilon_N, \end{array}$$

jossa x_{ik} on k:n selittävän muuttujan i:s arvo, y_i on arvoja x_{i1}, \dots, x_{iK} vastaava selitettävä muuttuja ja ε_i on jäännöstermi. Parametrit $\beta_0, \beta_1, \dots, \beta_K$ ovat estimoitavia regressiokertoimia, joista β_0 on vakiotermi. Käytettäessä matriisimerkintöjä (1) saa seuraavan esityksen:

$$(2) \quad \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1K} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & x_{N1} & & x_{NK} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \cdot \\ \cdot \\ \cdot \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix}$$

$$(3) \quad y = X\beta + \varepsilon.$$

Ilman jäännöstermiä koskevia lisäoletuksia voidaan β valita mielivaltaisesti ja malli (3) on siis epäidentifioituva. Yhteinen piirre erilaisissa spesifikaatioissa on, että jäännöstermit $\varepsilon_1, \dots, \varepsilon_N$ ovat keskimäärin "lähellä nollaa" ja spesifikaatio määrittelee tarkemmin jäännöstermin luonteen. Seuraavassa tarkastellaan yksinkertaisinta tapausta, ns. *klassista lineaarista regressiomallia*, ks. Goldberger (1964, s. 161-71). Tässä oletetaan mm., että X on kiinteä (esim. koetilanteessa ennalta valittu) selittäjien havaintomatriisi, parametrit $(\beta_0, \beta_1, \dots, \beta_K)$ ovat niinkään kiinteitä lukuja, kun taas y_i :t ja ε_i :t ovat satunnaismuuttujia. Ennalta valitun X -matriisin lisäksi on havaittu eräät y_i -arvot, mutta β_i ja ε_i -arvot ovat sen sijaan tuntemattomia. Klassisessa lineaarisessa regressioanalyysissä tehdään seuraavat jäännöstermivektorin jakaumaa koskevat perusolettamukset:

$$(4) \quad E(\varepsilon) = [0, \dots, 0]' = \vec{0}.$$

$$(5) \quad \text{On olemassa vakio } \sigma > 0 \text{ s.e. } E(\varepsilon\varepsilon') = \sigma^2 I.$$

Tällöin jokaisen jäännöstermin ε_i odotusarvo on nolla ja hajonta σ sekä eri havaintojen jäännöstermit ovat korreloimattomia. Lisäksi oletetaan selittäjien havaintomatriisista:

$$(6) \quad \text{Matriisin } X \text{ aste} = 1 + K \leq N.$$

Olettamus (6) merkitsee, että havaintomatriisin X vektorit ovat lineaarisesti riippumattomia ja että havaintoja on riittävästi. Olettamus (6) kieltää selittäjien välisen täydellisen lineaarisen riippuvuuden eli täy-

dellisen multikollinearisuuden. Usein oletetaan lisäksi, että vektorin ϵ jakauma on multinormaalinen, siis $N_N(0, \sigma^2 I)$, jolloin kyseessä on *klassinen normaalin lineaarinen regressiomalli*, ks. Goldberger (1964, s. 171).

Regressioanalyysissä on tehtävänä estimoida havainnoista y ja X parametrit β ja virhevarianssi σ^2 . Tarkastellaan pienimmän neliösumman (pns) menetelmää ja olkoon β :n estimaatti b . Minimoitava neliösumma on: $Q = (y - Xb)'(y - Xb)$, jonka minimi saadaan asettamalla osittaisderivaatat muuttujien b_i suhteen nolliksi. Ratkaisuna saadaan $b = (X'X)^{-1}X'y$, joka on β :n harhaton estimaattori. Gauss-Markov-teoreeman mukaan b on lisäksi ns. *paras lineaarinen harhaton estimaattori*, ts. β :n lineaarisesti harhattomien estimaattoreiden joukossa pns-estimaattorin b yleistetty varianssi eli kovarianssimatriisi $E(bb') = \sigma^2(X'X)^{-1}$ on pienin mahdollinen, ks. Goldberger (1964, s. 164). Merkitään mallista lasketuja y :n arvoja symbolilla $\hat{y} = Xb$ ja jäännösten ϵ arvioita $e = y - y^* = y - Xb$. Voidaan osoittaa, että $E(e'e) = \sigma^2(N-K-1)$, joten jäännösvarianssin σ^2 harhaton estimaattori on $s^2 = \sum_{i=1}^N e_i^2 / (N-K-1)$. Normaalijakaumaoletta-
muksen $\epsilon \sim N_N(0, \sigma^2 I)$ vallitessa lisäksi $b \sim N_K(\beta, \sigma^2(X'X)^{-1})$, josta seuraavat tavanomaiset regressiokertoimia koskevat testausmenettelyt.

Eräänlainen varianssi- ja regressioanalyysin välimuoto on ns. kovarianssi-analyysi, jossa osa selittävästä muuttujista on kvantitatiivisia ja osa kvalitatiivisia muuttujia. Kovarianssianalyysi yhdistää kahden hyvin käytökelpoisen menettelyn, regressio- ja varianssianalyysin edut. Varianssi-analyysiä käytetään tutkittaessa, ovatko perusjoukkojen odotusarvot yhtäsuuret ja testaus suoritetaan kahden sopivasti lasketun otosvarianssin suhteen avulla. Havaintoaineistossa esiintyvä kokonaisvariaatiota kuvaa-

va neliösumma hajoitetaan kahteen komponenttiin: ryhmien välistä variaatiota kuvaavaan neliösummaan ja ryhmien sisäistä variaatiota kuvaavaan neliösummaan. Ryhmien välistä variaatiota kuvaava neliösumma paljastaa systemaattiset, kun taas ryhmien sisäistä variaatiota kuvaava neliösumma satunnaiset tekijät, ks. esim. Vasama ja Vartia (1972, s. 661).

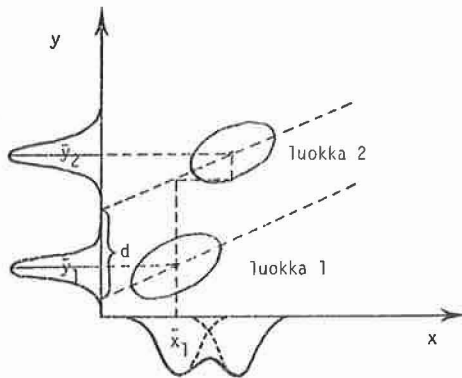
Kovarianssianalyysi kehitettiin aluksi korjaamaan kokeellisessa tutkimuksessa sen laatuisten muuttujien vaikutuksia, joita ei voitu kunnolla kontrolloida eri luokissa. Tutkitaan esimerkin avulla tätä ongelmaa, ks. Johnston (1972, s. 192). Tarkastellaan seuraavaa havaintoaineistoa:

		Luokat	
		1	. . . p
Havainnot	1	y_{11}	y_{1p}
	.		
	.		
	m	y_{1m}	y_{pm}

jossa p on luokkien lukumäärä ja jokaisessa luokassa on sama määrä m havaintoja. Havaintoarvot y_{11}, \dots, y_{1m} voivat olla esim. alueella 1 asuvien $m:n$ kotitalouden tuloja. Luokakeskiarvot $\bar{y}_i = \frac{1}{m} \sum_{j=1}^m y_{ij}$ kussakin luokassa (esim. alueella) $i = 1, \dots, p$ voidaan laskea ja peruskysymyksenä on, osoittavatko nämä otoskeskiarvot, että y -muuttujan odotusarvot vaihtelevat luokasta toiseen. Muuttujan y kokonaisvariaatio jaetaan luokkien väliseen ja luokkien sisäiseen vaihteluun. Varsinaisena kiinnostuksen kohteena on luokkien välinen vaihtelu (joka syntyy juuri luokakeskiarvojen

eroista) ja jonka merkitsevyyttä voidaan tutkia F-testillä. F-testin validiteetti riippuu mm. siitä voidaanko luokkien sisäistä neliösummaa käyttää satunnaistekijöiden vaikutuksen indikaattorina. Oletetaan, että y-muuttujaan vaikuttaa luokkia $1, \dots, p$ erottavien tekijöiden lisäksi muita tekijöitä, joita ei ole pystytty kontrolloimaan ja standardisoimaan eri luokissa. Esimerkiksi kotitalouksien koolla saattaa olla tuloeroja aiheuttava vaikutus varsinaisten aluetekijöiden lisäksi. Tällaisessa tapauksessa tulojen vaihtelu alueiden (luokkien) sisällä muodostuu suuremmaksi kuin jos kotitalouksien koko olisi vakioitu. Jos on kysymys vain yhdestä ei-kontrolloidusta muuttujasta x ja kahdesta luokasta, ongelma on seuraavan kuvion mukainen:

Kuvio 19. Muuttujien y ja x välinen riippuvuus



Kuviossa ellipsit kuvaavat havaintojen vaihtelua luokissa ja yhdensuuntaiset viivat x :n erillisvaikutuksia y -muuttujaan kummassakin luokassa. Luokan 2 havainnoilla on suuremmat x :n arvot kuin luokassa 1. Karkea luokkakeskiarvojen $(\bar{y}_2 - \bar{y}_1)$ ero yliarvioi luokkavaikutusta, joka itse asiassa on yhdensuuntaisten suorien välinen ero d . Tämän suuruinen kes-

keskimääräinen ero luokkien y -arvojen välillä todettaisiin, jos x -arvot olisi voitu vakioida tietyn suuruiseksi (esim. \bar{x}_1 :ksi). Tätä virhettä voidaan kovarianssianalyysin avulla korjata ottamalla mukaan analyysiin x -muuttuja. Tässä esitetty analyysi lähtee siitä hypoteesista (joka lienee yleensä väärä), että kuvioon piirretyt suorat kuvaavat paitsi ko. ellipsien regressiosuoria myös niitä käyriä, joita pitkin ellipsit liikkuisivat, jos keskimääräisiä x -arvoja olisi muutettu. Nämä ovat tyypillisiä ei-kokeellisen ja selittävän tutkimuksen ongelmia.

Luultavasti yleisimmin kovarianssianalyysiä käytetään satunnaistettujen kokeiden tavanomaista tarkempaan analysointiin, mutta kovarianssianalyysiä voidaan käyttää myös ei-kokeellisessa havaintomateriaalissa häiriötekijöiden poistamisen apuvälineenä. Kvantitatiivisten selittäjien ja selitettävän välistä riippuvuutta voidaan tarkentaa ottamalla mukaan analyysiin kvalitatiivisia luokkamuuttujia. Tällöin normaalina oletuksena on, että kvantitatiivisilla selittäjillä on sama vaikutus selitettävän eri luokissa. Tarkastellaan Johnstonin (1972, s. 194-) esimerkin avulla, miten kovarianssianalyysissä tutkitaan luokkamuuttujien merkitsevyyttä selittäjinä.

Havaintoaineisto muodostuu luvuista $y_{ij}, x_{2ij}, \dots, x_{kij}$, jossa $i = 1, \dots, p$ ja $j = 1, \dots, m$. Muuttujien y, x_2, \dots, x_k arvot on siis mitattu pm :n tilastoyksikön a_{ij} (esim. henkilön) kohdalla. Tilastoyksiköt a_{i1}, \dots, a_{im} kuu-

luvut i :nteen luokkaan, ts. juuri näillä tilastoyksiköillä a indikaattorimuuttuja¹⁾ $d_i = \phi(a \text{ kuuluu luokkaan } i)$ saa arvon yksi. Aineistoa voidaan havainnollistaa seuraavanlaisella havaintomatriisilla:

		Muuttujat							
		y	x_2	...	x_k	d_1	d_2	...	d_p
Tilastoyksiköt	a_{11}	y_{11}	x_{211}			1	0		0

	a_{1m}	y_{1m}	x_{21m}			1	0		0
	a_{21}

	a_{2m}
	a_{p1}	y_{p1}	x_{2p1}			0	0		1

	a_{pm}	y_{pm}	x_{2pm}			0	0		1

1) Väitteen p_i indikaattorimuuttuja d_i (eli "totuusarvo", logikkojen käyttämä termi tai "dummy-muuttuja", ekonometrikkojen termi) voidaan määrittellä logiikan totuusfunktioita (eli "validiteetti-funktioita" tai matemaatikkojen "karakteristista funktiota") $\phi(p)$ käyttäen seuraavasti:

$$d_i = \phi(p_i) = \begin{cases} 1, & \text{jos väite } p_i \text{ on tosi} \\ 0, & \text{jos väite } p_i \text{ on epätosi.} \end{cases}$$

Esim. $d_2 = \phi(\text{ko. tilastoyksikkö kuuluu luokkaan } 2)$
 $= \begin{cases} 1, & \text{jos ko. tilastoyksikkö kuuluu luokkaan } 2 \\ 0 & \text{muulloin.} \end{cases}$

Tilastoyksikön käsitettä käyttäen esityksestä saatiin varsin yksinkertainen, vrt. kuitenkin Stenius (1976), Vasama ja Vartia (1972, s. 34-65).

Kvantitatiivista selitettävää muuttujaa y selitetään $(K-1)$ kvantitatiivisen selittäjän x_2, \dots, x_K ja aineiston p luokkaan jakavan luokittelukriteerin avulla. Jokaisessa p :ssä luokassa on m havaintoa, joten havaintoja on yhteensä $N=mp$ kappaletta. Tavanomainen aineistoa koskeva regressiomalli olisi

$$(4) \quad y = X\beta + u,$$

missä y on eo. havaintomatriisissa esitetty p :stä "apuvektorista" muodostettu pystyvektori, jossa kunkin luokan y -havainnot on lueteltu erikseen. X on vakion ($x_1 \equiv 1$) sisältävä $(N \times K)$ matriisi ja β on $(K \times 1)$ vektori. Tämä malli soveltuu erityisesti tilanteeseen, jossa luokittelulla p luokkaan ei ole itsenäistä merkitystä ja y :n vaihtelu tulee selvitetyksi jo x -muuttujien avulla. Luokkavaikutuksen tutkimiseksi muodostetaan yleisempi kovarianssianalyysimalli:

$$(5) \quad y = D\alpha + X\beta + u,$$

missä y , X ja β ovat kuten edellä ja α on $(p-1)$ pystyvektori, $\alpha = (\alpha_2, \alpha_3, \dots, \alpha_p)'$. D on seuraavanlainen indikaattorimuuttujamatriisi:

$$D = [D_2 \ D_3 \ \dots \ D_p] = \begin{array}{cccc} & d_2 & d_3 & \dots & d_p \\ \left. \begin{array}{c} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{array} \right\} m & & 0 & & 0 \\ & & \vdots & & \vdots \\ & & \vdots & & \vdots \\ & & 0 & & 0 \\ & & \vdots & & \vdots \\ & & \vdots & & \vdots \\ & & 0 & & 1 \\ \left. \begin{array}{c} \vdots \\ \vdots \\ 0 \end{array} \right\} m & & \vdots & & \vdots \\ & & 0 & & 1 \end{array} \left. \vphantom{\begin{array}{c} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \\ \vdots \\ \vdots \\ 1 \\ \vdots \\ \vdots \\ 0 \\ \vdots \\ \vdots \\ 0 \end{array}} \right\} mp = N \text{ kp1}$$

p-1 kp1

jossa esim. pystyrivi D_2 muodostuu indikaattorimuuttujan d_2 havaintoarvoista. Ensimmäisen luokan indikaattori d_1 on jätetty pois, koska muutoin indikaattorimuuttujat olisivat lineaarisesti riippuvia. Malli (5) merkitsee seuraavia yhtälöitä eri luokissa:

Luokka 1: $y_{1j} = \beta_1 + \beta_2 x_{21j} + \dots + \beta_k x_{k1j} + u_{1j}$

Luokka 2: $y_{2j} = (\alpha_2 + \beta_1) + \beta_2 x_{22j} + \dots + \beta_k x_{k2j} + u_{2j}$

⋮

Luokka p: $y_{pj} = (\alpha_p + \beta_1) + \beta_2 x_{2pj} + \dots + \beta_k x_{kpj} + u_{pj}$,

jossa jokaisessa $j = 1, \dots, m$.

Vakiotermin muuttuu luokasta toiseen, mutta kaikissa luokissa kertoimet β_2, \dots, β_K pysyvät samoina. Jos kaikki luokkavaikutukset ovat nolliä, niin $\alpha_2 = \dots = \alpha_p = 0$. Testattaessa hypoteesia $\alpha = 0$ kysytään lisäksi malli (5) merkittävästi selitettyä neliösummaa verrattuna malliin (4). Kun sovelletaan malliin (4) pns-menetelmää, saadaan $y = X\hat{\beta} + s$, missä $\hat{\beta} = (X'X)^{-1}X'y$ ja s on pns-residuaalivektori. Tällöin $y'y = \hat{\beta}'X'X\hat{\beta} + s's + 2\hat{\beta}'X's = \hat{\beta}'X'y + s's$, koska $X's = 0$. Vastaavasti voidaan soveltaa pns-menetelmää malliin (5):

$$(6) \quad y = D\hat{\alpha} + X\hat{\beta} + e,$$

missä

$$(7) \quad \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} D'D & D'X \\ X'D & X'X \end{bmatrix}^{-1} \begin{bmatrix} D'y \\ X'y \end{bmatrix}.$$

Kuten edellä saadaan

$$(8) \quad y'y = \hat{\alpha}'D'y + \hat{\beta}'X'y + e'e.$$

Mallien (4) ja (5) neliösummien erotus on siten

$$(9) \quad s's - e'e = \hat{\alpha}'D'y + \hat{\beta}'X'y - \hat{\beta}'X'y.$$

Nämä suureet esitetään usein seuraavanlaisessa kovarianssianalyysi-asetelmassa, johon hypoteesia $\alpha = 0$ koskeva testi voidaan perustaa:

Taulukko 1. Kovarianssianalyysi erisuuruisista vakiotermeistä

Variaatiolähde	Jäännösneliösumma	Vapausasteet	Kovarianssiestimaatti
X ja D	$e'e = y'y - \hat{\alpha}'D'y - \hat{\beta}'X'y$ $= S_2$	mp-p-K+1	$\frac{S_2}{mp-p-K+1}$
D	lisäksi: $s's - e'e =$ $\hat{\alpha}'D'y + \hat{\beta}'X'y - \hat{\beta}'X'y$ $= S_1$	p-1	$\frac{S_1}{p-1}$
X	$s's = y'y - \hat{\beta}'X'y$	mp-K = N-K	

Virhetermejä koskevan normaalisuusolettamuksen vallitessa hypoteesia $\alpha_2 = \dots = \alpha_p = 0$ voidaan testata F-testillä, jonka testisuure saadaan tavanomaiseen tapaan vapausasteella jaettujen neliösummien suhtena:

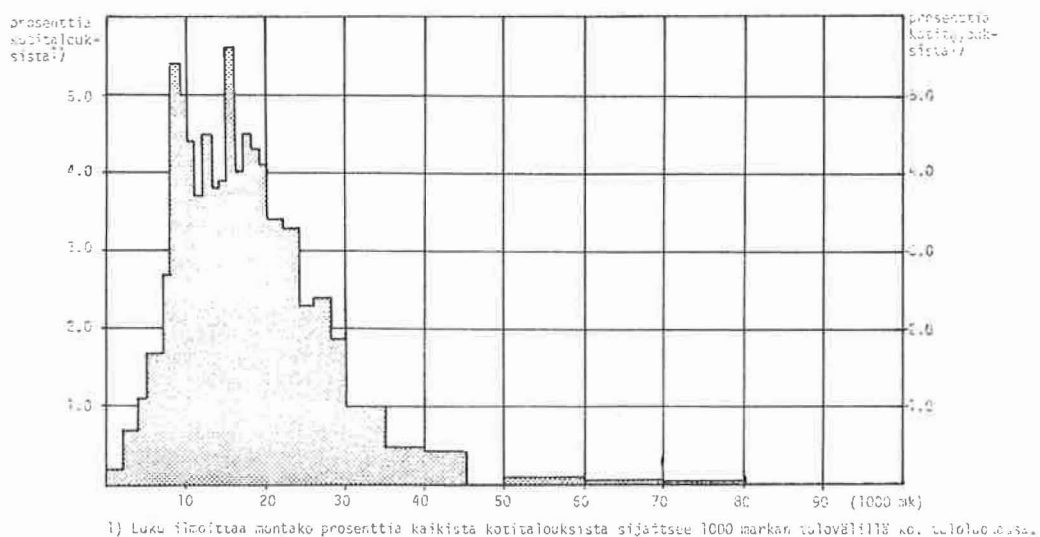
$$(10) \quad F = \frac{S_1/(p-1)}{S_2/(mp-p-K+1)}$$

Nollahypoteesin ollessa tosi testisuure (10) noudattaa F-jakaumaa vapausastein (p-1, mp-p-K+1). Voidaan osoittaa, että esitetty kovarianssianalyysin luokittelun merkitsevyyden testaaminen vastaa regressioanalyysissä indikaattorimuuttujien merkitsevyyden testaamista, ks. Kooyman (1976, s. 118-126). Näin ollen tavallinen regressioanalyysi, jossa selittäjinä on sekä kvantitatiivisia, että kvalitatiivisia muuttujia, merkitsee itse asiassa kovarianssianalyysiä, jossa indikaattorimuuttujien avulla poistetaan epätoivottuja häiriötekijöitä.

5.2. Aineiston kuvaus

Kotitalouksien käytettävissä oleva markkamääräinen vuositulo on jakautunut tutkimusotoksessa kuvion 2 mukaisesti. Jakauma on vahvasti oikealle vino.

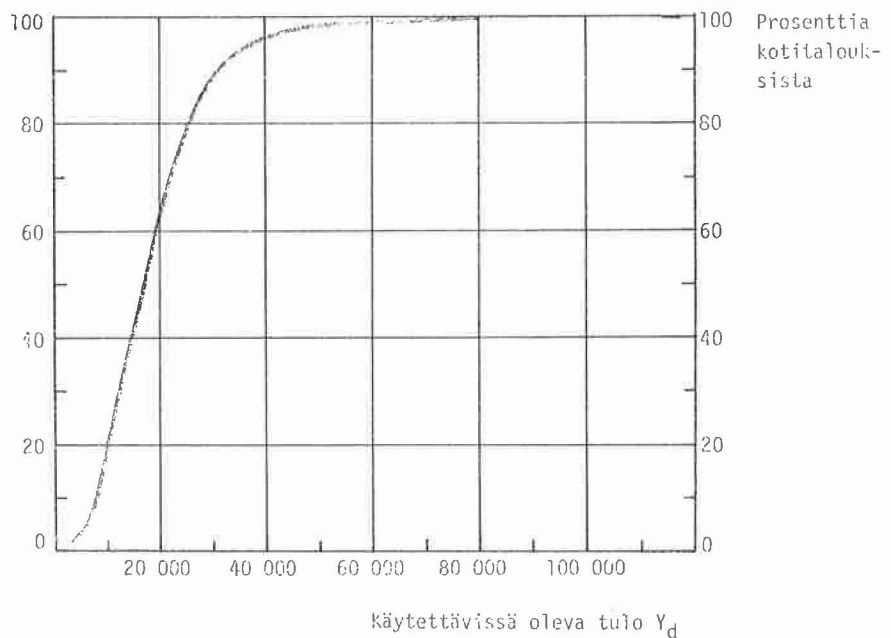
Kuvio 2. Kotitalouksien käytettävissä olevien vuositulojen Y_d jakautuminen



Kuviossa 2 on esitetty kotitalouksien käytettävissä olevan vuositulon prosenttinen frekvenssihistogramma. Pylvään pinta-ala kussakin tulo-
luokassa on verrannollinen ko. tuloaluksaan sattuvien kotitalouksien
suhteellisen osuuden kanssa, ts. pylvään korkeus on ko. suhteellisen
osuuden ja luokkavälin yksikköpituuden (tässä 1000 mk) suhde.

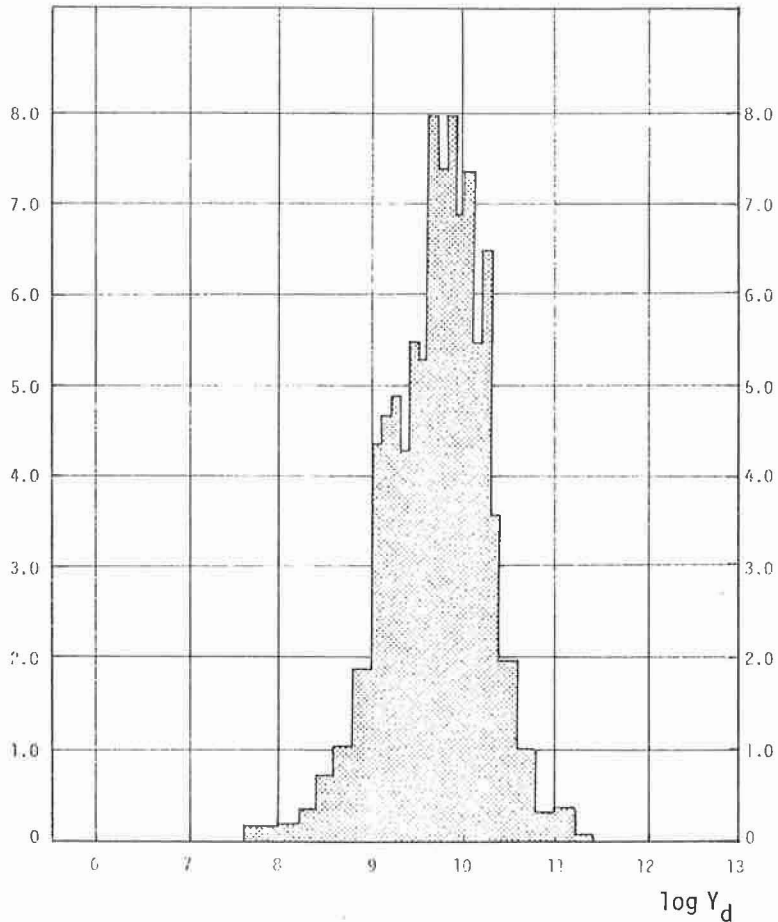
Kotitalouksien käytettävissä olevan tulon otoskertymäfunktioistakin (kuvio 3) voidaan todeta, että 50 prosentilla kotitalouksista käytettävissä oleva tulo jäi alle 17 070 markan vuonna 1971. Otoksessa käytettävissä olevan tulon keskiarvo on 19 473 mk, mikä on 14,1 % em. mediaanituloa suurempi. Myös tämä on osoitus tulojakauman vinoudesta.

Kuvio 3. Kotitalouksien käytettävissä olevan tulon otoskertymäfunktio



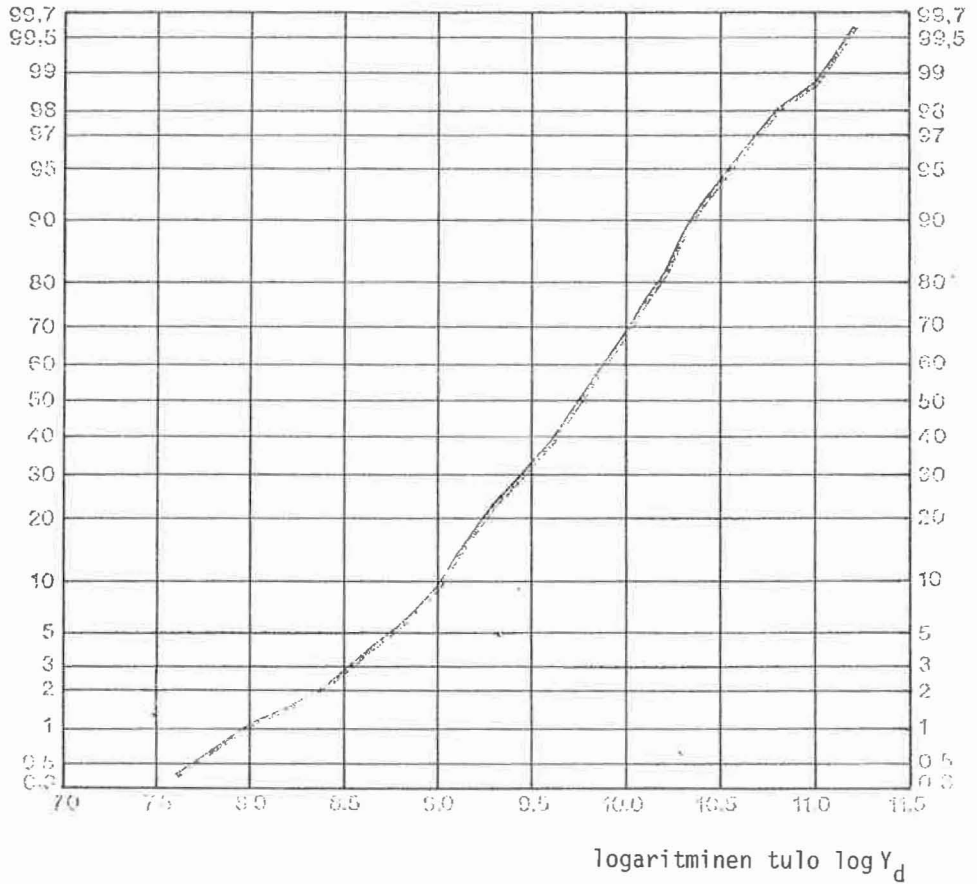
Kuviossa 4 on esitetty kotitalouksien käytettävissä olevan tulon suhteellinen frekvenssihistogramma logaritmisella tuloasteikolla. Sen perusteella voidaan todeta, että tulot jakautuvat logaritmiasteikolla likimain normaalisesti, joten alkuperäisellä tuloasteikolla jakauma on likimain log-normaalinen.

Kuvio 4. Kotitalouksien logaritmistien käytettävissä olevien vuositulojen $\log Y_d$ jakautuminen



Havaintojen poikkeamat log-normaalista jakaumasta ovat odotettavan suuntaiset. Jos muuttujan $\log Y_d$ kertymäfunktio piirretään todennäköisyyspaperille saadaan tulojakauman ns. log-normaalinen esitys. Log-normaalisen jakauman tapauksessa logaritmissen tulon kertymäfunktio olisi todennäköisyyspaperilla nouseva suora; logaritmissen käytettävissä olevan tulon kertymäfunktio on todennäköisyyspaperilla lievästi S:n muotoinen, kuitenkin niin, että 1. desilistä lähtien se on likimain suora.

Kuvio 5. Kotitalouksien käytettävissä olevien tulojen jakauman log-normaalinen esitys

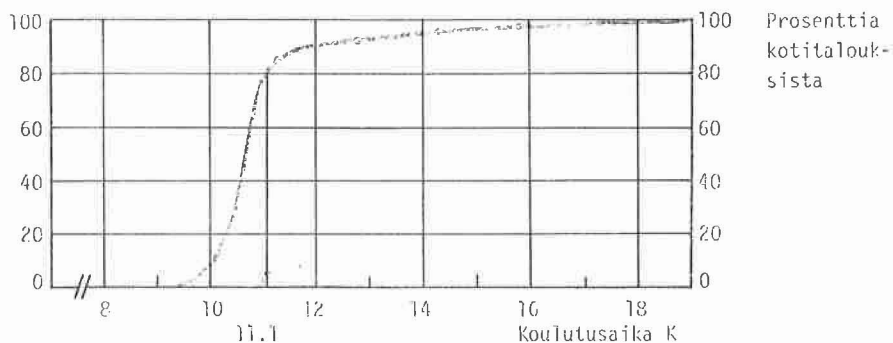


Logaritmisen tulon keskiarvo on 9.697, joten tulojen geometrinen keskiarvo $\text{Geom}(Y_d) \approx e^{9.697} \approx 16\,270$ markkaa. Log-tulojen mediaani on hieman suurempi n. 9.745 ja siis tulojen mediaani $\text{Md}(Y_d)$ on likimain $e^{9.745} \approx 17\,070$ markkaa. Log-tulojen hajonta $s(\log Y_d) = 0.6004$ ja varianssi $s^2(\log Y_d) = 0.360$. Tämän osoittamaa log-tulojen variaatiota tullaan malleilla selittämään.

Kotitalouksien käytettävissä olevien tulojen erilaisia jakaumia tarkasteltaessa on muistettava, että tutkimusotoksessa on mukana vain aktiivikotitaloudet. Jos verrataan kotitaloustiedustelun 1971 kaikkien kotitalouksien käytettävissä olevan tulon keskiarvoa (16 548 mk) otoksen keskiarvoon (19 473 mk), niin havaitaan, että aktiivikotitalouksiin rajaaminen vaikutti nostavasti kotitalouksien käytettävissä olevan tulon keskimääräiseen suuruuteen. Tulojakauman muotoa ja ominaisuuksia ei tässä tutkita tämän tarkemmin, sillä tarkoituksena on yleisten tuloerojen sijaan selvittää, paljonko koulutuksen lisäyksen voidaan katsoa aikaansaavan tulojen lisäystä.

Tarkasteltaessa kotitalouksien päämiesten koulutusvuosien otoskertymäfunktiota (kuvio 6) havaitaan, että 80 % kotitalouksien päämiehistä on saanut korkeintaan n. 11 vuotta kestäneen koulutuksen ja siis n. 20 prosentilla koulutusaika ylittää 11 vuotta.

Kuvio 6. Kotitalouksien päämiesten koulutusvuosien K otoskertymäfunktio



Päämiehet, joiden koulutusaika oli Systemaattisen Koulutusluokituksen mukaan tuntematon, on sijoitettu alle 11.25 vuoden koulutusluokkaan. Vain 2.3 prosentilla koulutusaika jää 10 vuotta lyhyemmäksi. Verrattaessa taulukossa 2 Väestölaskennan (1970, osa VII B, s. 312) ammatissa toimivan väestön koulutuksen mukaista jakaumaa tämän tutkimuksen aktiivikotitalouksien päämiesten koulutuksen otosjakaumaan voidaan todeta niiden vastaavan hyvin toisiaan.

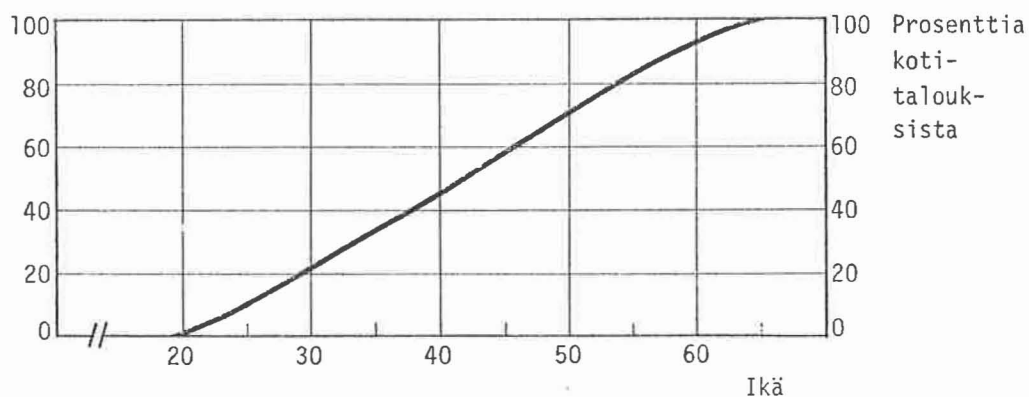
Taulukko 2. Henkilöt koulutusvuosien mukaan väestölaskennassa v. 1970 ja tutkimuksen aineistossa, %

Arvioitu koulutusaika vuosissa	Väestölaskenta 1970 (ammatissa toimiva väestö)	Tutkimuksen aineisto (kotitalouksien päämiehet)
alle 11.25	24.3	85.3
11.25-12.75	7.5	7.5
12.75-14.25	2.9	2.5
14.25-15.75	1.2	1.8
15.75-18.25	2.1	2.4
yli 18.25	0.2	0.5
Luokittelematon	61.8	-
Yhteensä	100.0	100.0

Kuviossa 7 esitetty kotitalouksien päämiehen iän otoskertymäfunktio kertoo mm., että 50 % otoksen päämiehistä oli v. 1971 alle 42 vuotiaita. Kertymäfunktio nousee hyvin tasaisesti, joten päämiehiä on otoksen kaikissa (esim. 5 vuoden pituisissa) ikäluokissa likimain yhtä paljon. Ikäjakauma on häm-

mästyttävän lähellä välin 18-65 tasajakaumaa. Kaikki päämiehet ovat tehdyn rajauksen mukaisesti alle 65 vuotiaita. Väestön ikäjakauma välillä 18-65 ei suinkaan ole yhtä tasainen, vaan nuorempien ikäluokkien suhteellinen osuus on selvästi vanhempien ikäluokkien osuutta suurempi.

Kuvio 7. Kotitalouksien päämiehen iän otoskertymäfunktio haastatteluvuonna 1971



Kotitalouksien keskipöytävuoden 1971 kotitaloustiedustelussa oli 2.97 henkeä. Aktiivikotitalouksiin rajoittuneessa otoksessa kotitalouksien keskipöytävuosi nousi 3.39 henkeä. Suhteellisen pienten vanhustalouksien poisjääminen luonnollisesti näkyy kotitalouksien keskipöytävuoressa.

Kotitalouksien "päämiehistä" oli otoksessa yllättävän monta, nimittäin 218 naispuolista ja vastaavasti 782 miespuolista. Kotitalouksien ammatissa toimivien lukumäärä jakautui taulukon 3 mukaisesti.

Taulukko 3. Kotitaloudet ammatissa toimivien lukumäärän mukaan

Ammatissa toimivien lkm	Kotitalouksien lkm, %
0	5.9
1	36.9
2	43.3
3	9.8
4	3.2
5	.9
Yhteensä	100.0

Lähes 43 prosentissa kotitalouksista ammatissa toimii yksi tai ei yhtään henkilöä. Suurimman ryhmän muodostavat kotitaloudet, joissa ammatissa toimii kaksi henkeä.

Luokittelemalla kotitaloudet sosioekonomisen aseman mukaan saadaan seuraava taulukko.

Taulukko 4. Kotitaloudet sosioekonomisen aseman mukaan

Sosioekonominen asema	Kotitalouksien lkm, %
Maa- ja metsätalouden harjoittajat	18.3
Muut yrittäjät	4.4
Johtajat tai toimihenkilöt	27.9
Työntekijät	41.4
Ammatissa toimimattomat	8.0
Yhteensä	100.0

Työntekijäkotitalouksia on lähes 42 % kaikista otoksen kotitalouksista. Ammatissa toimimattomiksi kotitalouksiksi on määritelty ne kotitaloudet, jotka saavat pääasiallisen toimeentulonsa muusta kuin yritys- tai palkanansaintatoiminnasta. Ammatissa toimimattomassa kotitaloudessa voi sitten olla ammatissa toimivia jäseniä, kunhan tästä toiminnasta saadut tulot ovat vain pieni osa sen käytettävissä olevasta tulosta. Ammatissa toimimattomista kotitalouksista suurin osa on eläkeläistalouksia, joita on 6.5 % kaikista kotitalouksista.

Otoskotitaloudet ovat jakautuneet sijaintinsa suhteen taulukon 5 osoittamalla tavalla.

Taulukko 5. Kotitaloudet suuralueittain

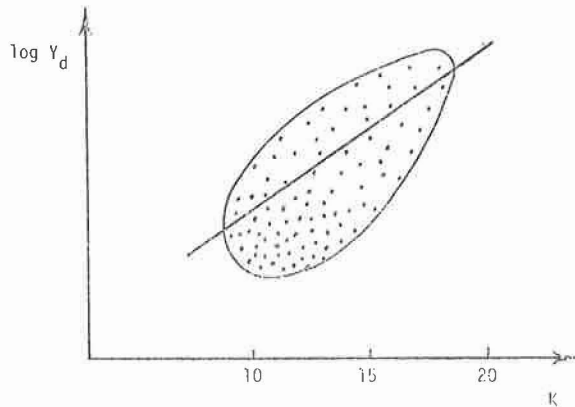
Alue	Kotitalouksien lkm, %
Etelä-Suomi	58.7
Väli-Suomi	28.5
Pohjois-Suomi	12.8
Yhteensä	100.0

Lähes 60 % otoskotitalouksista sijaitsee Etelä-Suomessa. Väli- ja Pohjois-Suomi jakavat 40 % otoskotitalouksista siten, että Väli-Suomi saa niistä suurimman osan. Kaupungeissa ja kauppaloissa asuu 53.3 % kaikista otoskotitalouksista ja maalaiskunnissa loput 46.7 %.

5.3. Estimoitavat mallit

Kotitalouden päämiehen koulutuksen ja kotitalouden käytettävissä olevan tulon keskinäisen riippuvuuden tarkastelu aloitetaan aivan yksinkertaisesta regressiomallista: selitettävänä muuttujana on kotitalouden käytettävissä olevan vuositulon Y_d tuloteorioidenkin suosittama luonnollinen logaritmi $\log Y_d$ ja selittäjänä on kotitalouden päämiehen koulutus K . Tulot on mitattu markoissa ja koulutus vuosissa. Vuositulojen vaihdellessa noin tuhannesta aina 600 000 markkaan tulojen logaritmin vaihteluväliksi tulee 5.5 - 13.5. Mallissa kiinnitetään huomio päämiehen koulutuksen suhteelliseen tulovaikutukseen, ts. montako prosenttia koulutusvuoden lisäys nostaa tuloja. Käytettävissä olevaa havaintomateriaalia käsitellään tässä vaiheessa yhtenä kokonaisuutena, josta yksinkertaisen regressiomallin avulla määrätään kuvion 8 tapaan koulutuksen kulmakerroin.

Kuvio 8. Kotitaloudet päämiehen koulutuksen ja logaritmisien tulon mukaan (Skemaattinen esitys)



Käytetty regressiomalli (1) on seuraavaa muotoa:

$$(1) \quad \log Y_d = \gamma_0^{(1)} + \gamma_1^{(1)} K + U, \text{ missä}$$

Y_d = kotitalouden käytettävissä oleva (kokonais) vuositulo

K = kotitalouden päämiehen koulutus vuosissa

U = jäännöstermi

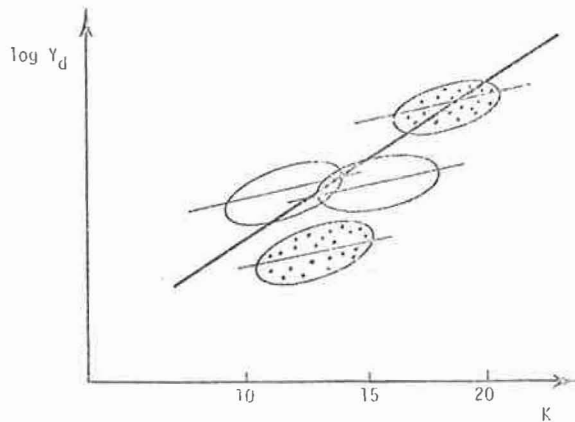
$\gamma_0^{(1)}$ ja $\gamma_1^{(1)}$ ovat 1-mallin parametreja.

Mikäli havaintomateriaalin muodostaisivat luvussa 4.2 esitetyn 1000 dublikaattiin liittyvän ajatuskokeen tulokset, niin tämä malli riittäisi selvittämään päämiehen koulutuksen vaikutuksen kotitalouden käytettävissä olevaan tuloon. Käytettävä havaintoaineisto on kuitenkin kerätty passiivisen rekisteröinnin avulla, joten näin yksinkertainen malli ei sellaisenaan sovi tutkimusongelman selvittämiseen. Tieteellinen tutkimus olisi varsin mekaanista puuhaa, jos ongelmat voitaisiin näin yksinkertaisesti selvittää. Monet muut tekijät koulutuksen lisäksi ovat aiheuttaneet systemaattisia eroja kotitalouksien tuloissa. Vain, jos nämä muut tuloihin vaikuttavat tekijät olisivat korreloimattomia koulutusmuuttujan kanssa - kuten olisi asian laita, jos koulutuksen määrä olisi ratkaistu arvalla hypoteettisen kokeemme tavoin - saisi eo. mallin koulutusmuuttujan kerroin oikean suuruisen arvon, ks. Goldberger (1964, s. 194-201). Todellisuudessa "häiriötekijät" (kuten perityt ominaisuudet, lahjakkuus) ovat enemmän tai vähemmän korreloituneita koulutuksen kanssa. Esimerkiksi sosioekonomisen aseman ja lahjakkuustekijöiden positiivinen korrelaatio koulutuksen kanssa näkyy tuloksissa siten, että kotitalouden päämiehen koulutusmuuttujan kerroin $\gamma_1^{(1)}$ saa mallissa (1) liian suuren arvon todelliseen vaikutuskertoimeen

verrattuna. Epätoivottuja vaikutuksia pitäisi jollakin tavoin selvittää ja eliminoida. Ei-kokeellisessa tutkimuksessa tähän yleensä käytetään monimuuttujaregressioanalyysia. Mallia (1) täytyy siis laajentaa.

Seuraavaksi kuvataan käytetyt merkinnät ja mallien rakenne. Estimointitulokset esitetään seuraavassa luvussa. Laajennettuihin malleihin otetaan mukaan muita kotitalouden päämiehen ominaisuuksia kuvaavia muuttujia, kuten päämiehen potentiaalinen työkokemus T , päämiehen sukupuoli d_1 ja lähinnä päämiehen ammattialaa kuvaavia elinkeinoindikaattoreita. Kotitaloudet tulevat mallia laajennettaessa alustavasti luokiteltua kotitalouksien päämiesten ominaisuuksien mukaan: kuvion 8 tilanteesta siirrytään tarkempaan kuvaukseen, kuvioon 9.

Kuvio 9. Kotitaloudet päämiehen koulutuksen ja logaritmisien tulon mukaan joidenkin taustamuuttujien (indikaattoreiden) määräämissä luokissa (Skemaattinen esitys)



Laajennettu malli (2) on seuraava:

$$(2) \quad \log Y_d = \gamma_0^{(2)} + \gamma_1^{(2)} K + \gamma_2^{(2)} T + \beta_1^{(2)} d_1 + \sum_{i=2}^9 \beta_i^{(2)} d_i + U,$$

missä Y_d , K ja U on jo kuvattu mallissa (1), T = päämiehen potentiaali-
nen työkokemus ja $d_i = \Phi(p_i)$ väitteen p_i indikaattori(muuttuja) eli
totuusarvo.

Väitteet p_1, \dots, p_9 ovat seuraavat:

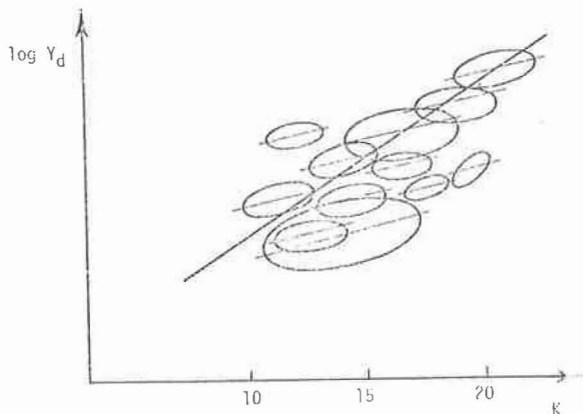
		<u>päämiehen sukupuoli</u>
p_1 = kotitalouden	päämies on nainen	
		<u>päämiehen ammattiala</u>
p_2 = kotitalouden	päämies on luontaistalouden harjoittaja	
p_3 = "	" " kaivannaistoiminnan harjoittaja	
p_4 = "	" " energia- tai vesihuollon tehtävissä	
p_5 = "	" " rakennustoiminnan tehtävissä	
p_6 = "	" " kaupan tehtävissä	
p_7 = "	" " kuljetus- tai tietoliikenteen tehtävissä	
p_8 = "	" " yksityisten palvelusten tehtävissä	
p_9 = "	" " julkisten palvelusten tehtävissä	

Kotitalouden päämiehen potentiaalista työkokemusta¹⁾ mittaava muuttuja T
on määritelty seuraavasti $T = A - K - 7$, siis lapsuus- ja koulutusajan
yli menevinä elinvuosina. Päämiehen työtehtävään liittyvät väitteet
 p_2, \dots, p_9 vastaavat toisensa poissulkevia ammattialoja. Mukaan ei ole
otettu väitettä p_{10}^* = "päämies on teollisuuden palveluksessa", sillä
tällöin indikaattorit d_1, \dots, d_9 , $d_{10}^* = \Phi(p_{10}^*)$ olisivat lineaarisesti riip-
puvia niiden summan ollessa aina 1.

1) T kuvaa niiden vuosien lukumäärää, jotka päämies on voinut opintojensa päättämi-
sen jälkeen olla työssä. Todellinen työkokemus on voinut jäädä näin määriteltyä
potentiaalista työkokemusta pienemmäksi.

Edellä olevat indikaattorimuuttujat jakavat kotitaloudet päämiesten suhteen samankaltaisiin ryhmiin. Samankaltaisten päämiesten kotitaloudet eivät kuitenkaan yleensä ole muissa suhteissa samanlaisia. Kotitaloudet voivat erota lisäksi rakenteensa, sosioekonomisen asemansa ja esim. sijaintinsa suhteen toisistaan. Nämä kotitalouksien ominaisuudet voivat osaltaan vaikuttaa niiden käytettävissä olevaan tuloon. Näiden tekijöiden vaikutuksia ei malleissa (1) ja (2) ole vielä huomioitu. Mallissa (3) tarkennetaan analyysiä ottamalla mukaan muuttujia, jotka kuvaavat esitettyjä kotitalouden keskeisiä piirteitä. Tällöin pyritään muodostamaan entistä homogeenisempia kotitalouksien ryhmityksiä ja kovarianssianalyysin keinoin etsimään vastausta kysymykseen, paljonko koulutuksen lisääminen nostaisi kotitalouden käytettävissä olevaa tuloa (siitä mihin se ilman ko. lisäystä olisi jäänyt). Yritetään muodostaa luvussa 4.2 esitettyyn tapaan kotitalouksista riittävän homogeenisia ryhmiä, joiden "sisäisestä" koulutuksen ja tulojen riippuvuudesta arvioidaan kovarianssianalyysin avulla koulutuksen todellista vaikutuskerrointa.

Kuvio 10. Kotitaloudet päämiehen koulutuksen ja kotitalouden logaritmisestä tulon mukaan useiden taustamuuttujien määrittämässä luokissa (Skemaattinen esitys)



Jotta kotitalouksien eri ominaisuuksien mukaanottamisen vaikutus analyysissä voitaisiin havaita erikseen, täydennetään regressiomallia (2) vaiheittain seuraavasti:

Kotitalouden ammatillisilla rakennetekijöillä d_{10} , d_{11} ja d_{12} laajennettu malli:

$$(3) \quad \log Y_d = \gamma_0^{(3)} + \gamma_1^{(3)}K + \gamma_2^{(3)}T + \gamma_3^{(3)}N + \sum_{i=1}^{12} \beta_i^{(3)}d_i + U$$

Kotitalouden ammatillisella rakennetekijöillä ja sosioekonomisen aseman tekijöillä d_{13}, \dots, d_{19} laajennettu malli:

$$(4) \quad \log Y_d = \gamma_0^{(4)} + \gamma_1^{(4)}K + \gamma_2^{(4)}T + \gamma_3^{(4)}N + \sum_{i=1}^{19} \beta_i^{(4)}d_i + U$$

Kotitalouden rakennetekijöillä, sosioekonomisen aseman tekijöillä ja sijaintitekijöillä d_{20}, \dots, d_{25} laajennettu malli:

$$(5) \quad \log Y_d = \gamma_0^{(5)} + \gamma_1^{(5)}K + \gamma_2^{(5)}T + \gamma_3^{(5)}N + \sum_{i=1}^{25} \beta_i^{(5)}d_i + U$$

Mallien (3)-(5) muuttujat ovat seuraavat: Y_d , K , T ja U kuten malleissa (1) ja (2), N = kotitalouden jäsenmäärä ja $d_i = \Phi(p_i)$ = väitteen p_i totuusarvo.

Väitteet p_1, \dots, p_9 on esitetty mallissa (2).

Kotitalouden ammatilliset rakennetekijät

p_{10} = kotitalous on ammattiaktiivi

p_{11} = " " ammattipassiivi

p_{12} = " " eläkeläistalous

Kotitalouden sosioekonominen asema

p_{13} = kotitalouden päämies on työtön

p_{14} = " " " julkisen sektorin palveluksessa

p_{15} = kotitalous harjoittaa maa- ja metsätaloutta

p_{16} = " " muuta yrittäjätoimintaa

p_{17} = " on johtaja- tai toimihenkilökotitalous

p_{18} = " " pääasiassa ammattissa toimimaton

p_{19} = " " suurtilalliskotitalous

Kotitalouden sijaintitekijät

p_{20} = kotitalous sijaitsee Väli-Suomessa

p_{21} = " " Pohjois-Suomessa

p_{22} = " " maalaiskunnassa

p_{23} = " asuu pientalossa

p_{24} = " " kylämaisessä asutuksessa

p_{25} = " " maja-asutuksessa

Kotitalous on määritelty ammattiaktiiviksi (jolloin p_{10} on tosi eli $d_{10} = \Phi(p_{10}) = 1$) silloin, kun ko. kotitaloudessa on kaksi tai sitä useampia ammatissa toimivia henkilöitä. Ammattipassiivissa kotitaloudessa (jolle $d_{11} = 1$) ammatissa toimivia jäseniä ei ole lainkaan. Kotitalous on suurtilalliskotitalous (jolle $d_{19} = 1$), jos kotitalouden tilan

kokonaispinta-ala on sata hehtaaria tai enemmän. Indikaattorimuuttujilla d_{15} - d_{18} erotellaan ko. toimialan tai ammattiryhmän kotitaloudet jäljelle jäävissä työntekijäkotalouksista. Muuttujien d_{20} ja d_{21} avulla voidaan erottaa Väli- ja Pohjois-Suomen kotitaloudet Etelä-Suomessa sijaitsevista: jos $d_{20} = d_{21} = 0$ niin ko. kotitalous sijaitsee Etelä-Suomessa. Etelä-Suomessa sijaitsemisen indikaattoria ei siis tarvita eikä sitä voitaisi regressioanalyysiin ottaakaan, koska tällöin momenttimatriisi olisi singulaarinen. Indikaattorimuuttujien muodostaminen havaintoaineistosta selviää liitteestä 3.

Teoreettisen pohdiskelun perusteella vain positiiviset kotitalouden päämiehen koulutuksen ja potentiaalisen työkokemuksen regressiokerroinestimaatit voidaan hyväksyä. Muiden mukana olevien muuttujien "oikeita" etumerkkejä ei edellä esitellyn teorian avulla voida päätellä, vaan niiden mielekkyyttä tarkastellaan vasta estimoinnin tapahduttua.

6. MALLIEN ESTIMOINTITULOKSET

6.1. Perusmallit

Luvussa 5.4. esitetyt regressiomallit on estimoitu Helsingin Kauppakorkeakoulun laskentakeskuksen STATS-ohjelmistoa hyväksikäyttäen. Malleille lasketut F-arvot, virhetermien standardipoikkeamien estimaatit, yhteiskorrelaatiokertoimien neliöt, parametrien estimaatit ja regressiokertoimien saamat t-arvot on esitetty liitteessä 7 ja ne on kuvattu graafisesti liitteessä 8. Mallit esitellään merkityssä järjestyksessä.

Tarkastellaan aluksi edellisen luvun yksinkertaisinta mallia (1). Kotitalouden päämiehen koulutus-muuttujan regressiokerroin poikkeaa nollasta tilastollisesti erittäin merkitsevästi. Mallin yhteiskorrelaatiokerroin kuvaa yhden selittäjän tapauksessa selittäjän ja selitettävän välistä korrelaatiota. Mallissa (1) kotitalouden päämiehen koulutuksen ja kotitalouden käytettävissä olevan tulon logaritmin korrelaatio osoittautui 0.305:n suuruiseksi. Tämä merkitsee sitä, että päämiehen koulutus selittää 9 % kotitalouden käytettävissä olevan tulon kokonaisvaihtelusta. Todettakoon, että poikkileikkaustutkimuksissa selitysasteet jäävät säännönmukaisesti pienemmiksi kuin mihin on totuttu aikasarjatutkimuksissa. Vaikka vain 9 % kotitalouksien välisistä tuloeroista voitiin selittää koulutuksen avulla, niin tämä selitysaste poikkeaa nollasta erittäin merkitsevästi.

Kotitalouden päämiehen koulutus-muuttujan regressiokertoimen estimaatiksi saatiin arvo 0.124. Luku 0.124 voidaan tulkita seuraavasti: jos kahden suomalaisen kotitalouden päämiehen vapaasti määräytyneiden koulutusten ero on vuosi, niin kotitalouksien käytettävissä olevat tulot poikkeavat

keskimäärin 12.4 %:lla¹⁾). Koulutusmuuttujan parametrilla on a priori oletuksen mukainen etumerkki, mutta se on lukuarvoltaan ilmeisesti liian suuri. Muiden kotitalouden käytettävissä olevaan tuloon vaikuttavien, koulutuksen kanssa positiivisesti korreloivien tekijöiden vaikutus luultavasti nostaa päämiehen koulutus-muuttujan regressiokerrointa. Todetakaan kuitenkin, että tätä yksinkertaista mallia voidaan käyttää ennustettaessa kotitalouden tuloja, kun ainoa kotitaloudesta tiedetty seikka on sen päämiehen koulutus vuosissa. Tällöin malli tavallaan ottaa huomioon paitsi koulutuksen suoranaiset vaikutukset, myös koulutuksen kanssa korreloivien tekijöiden kautta kulkevat välilliset vaikutukset. Esimerkiksi kotitalouksilla, joilla päämiehen koulutus on 15 vuotta, logaritmisten tulojen keskimääräinen arvo on mallin mukaan $8.340 + 0.124(15) = 10.2$. Jos virhetermi oletetaan normaaliseksi, niin luku 10.2 on normaalijakauman symmetrisyyden vuoksi samalla $\log(Y_d)$:n ehdollinen odotusarvo sekä mediaani. Siis 50 %:lle kotitalouksista, joilla $K = 15$, $\log(Y_d)$ jää pienemmäksi kuin 10.2 eli $Y_d \leq e^{10.2} = 26\ 900$ mk. Tämä 26 900 mk on mallin perusteella arvioitu mediaanitulo kotitalouksille, joiden päämiesten koulutusaika on 15 vuotta. Muuttujan $\log(Y_d)$ virhetermin hajonnan estimaatti on 0.572, joten ehdollisen muuttujan ($\log Y_d | K = 15$) arvioitu jakauma on $N(\mu = 10.2, \sigma^2 = (0.572)^2)$. Tästä jakaumasta voidaan

1) Muuttujan x logaritmisien muutoksen $\Delta \log x = \log(x^1/x^0)$ ($\log = \log_e = \ln$) satakertaa $100\Delta \log x$ kutsutaan x :n loogiseksi muutokseksi ja sen arvot ovat loogisia prosentteja. Erotukseksi tavallisista prosenteista loogisten prosenttien merkinä käytetään prosenttimerkkiä (%) alleviivattuna (%). Esim. 1 %:n suuruinen tulon muutos on loogisissa prosenteissa $100 \log(1.01) = 0.995\%$ (siis 0.995 loogista prosenttia) Pienten suhteellisten muutosten tapauksessa loogiset prosentit poikkeavat vain vähän tavallisista prosenteista, esim. $10\% \approx 9.95\% = 100 \log(1.1)$. Näitä termejä ja %-merkintää suosittelee asiaa pohtinut työryhmä, jonka jäseninä toimivat prof. Leo Törnqvist, dos. Pentti Vartia ja fil.tri Yrjö Vartia. Ehdotus korvaa aiemmat ehdotukset, ks. Herva, Vartia ja Vasama (1973, s. 26), P. Vartia (1974, s. 35), Y. Vartia (1976, s. 6).

laskea esimerkiksi arvio sille, että ko. kotitalouksissa $\log Y_d$ jää pienemmäksi kuin sen keskiarvo 9.70 koko aineistossa. Tämä tapahtuu seuraavasti: Normeerattu muuttuja $z = (x - \mu)/\sigma = (9.7 - 10.2)/0.572 = -0.874$. Normeeratussa normaalijakaumassa pisteen $z = -0.874$ alapuolelle jää noin 18 % havainnoista. Siis vain noin 18 %:ssa niistä kotitalouksista, joissa päämiehen koulutus on 15 vuotta, logaritmitiset tulot jäävät keskiarvon alapuolelle.

Tarkastellaan lyhyesti ehdolliselle muuttujalle $(\log Y_d|K)$ estimoidun mallin $(\log Y_d|K) = 8.340 + 0.124K + U$, jossa $U \sim N(0, (0.572)^2)$ ominaisuuksia. Normaalijakauman perusteella:

$$1. E(\log Y_d|K) = 8.340 + 0.124K$$

$$2. D^2(\log Y_d|K) = (0.572)^2$$

Jos $\log y \sim N(\mu, \sigma^2)$ eli $y \sim \text{log-norm}(\mu, \sigma^2)$, niin odotusarvo $Ey = \exp(\mu + \frac{1}{2}\sigma^2)$, hajonta $D(y) = Ey \sqrt{\exp(\sigma^2) - 1}$ ja mediaani $Md(y) = \text{Geom}(y) = e^\mu$, ks. esim. Aitchison-Brown (1957, s. 8). Näin ollen ehdollisille markkatuloille pätee $(Y_d|K) = e^{8.340 + 0.124K} e^U = 4188e^{0.124K} e^U$, jossa e^U noudattaa lognormaalista jakaumaa, $e^U \sim \text{log-norm}(0, (0.572)^2)$.

Erityisesti:

$$3. Md(Y_d|K) = 4188e^{0.124K} = \text{Geom}(Y_d|K)$$

$$4. E(Y_d|K) = 4188e^{0.124K} e^{\frac{1}{2}\sigma^2} = 4932e^{0.124K}$$

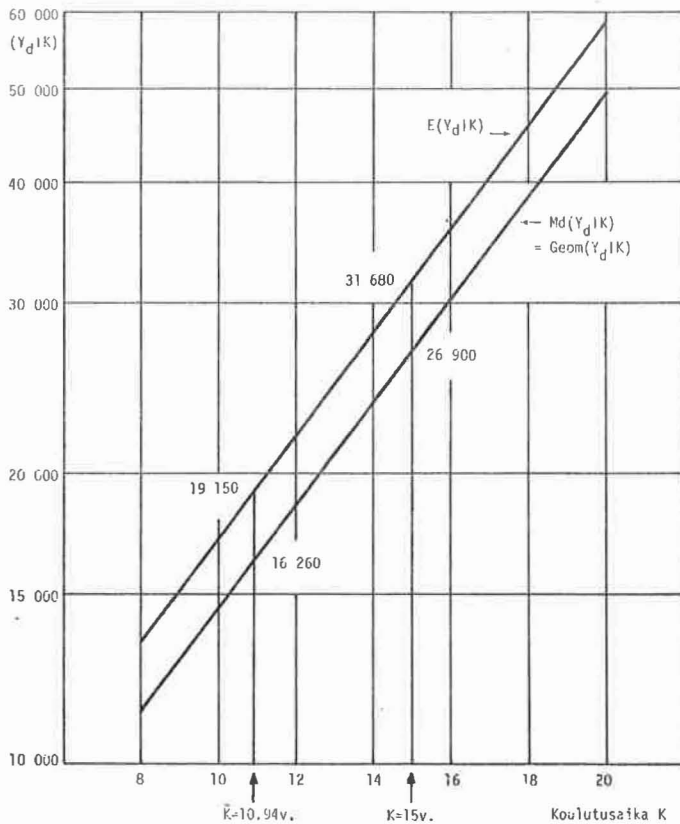
Ehdollisen tulon hajonta ja variaatiokerroin ovat seuraavat:

$$5. D(Y_d|K) = E(Y_d|K) \sqrt{\exp(\sigma^2) - 1} = 0.622E(Y_d|K) = 3068e^{0.124K}$$

$$6. VC(Y_d|K) = \sqrt{\exp(\sigma^2) - 1} = 0.622$$

Ehdollinen mediaani ja odotusarvo ovat molemmat logaritmiasteikolla tulon mukaan nousevia yhdensuuntaisia suoria, kuten kuviossa 1 on esitetty.

Kuvio 1. Mallin (1) mukaisen ehdollisen tulon ($Y_d|K$) mediaani ja odotusarvo koulutuksen K funktiona



Siirrytään tarkastelemaan mallin (2) estimointituloksia. Lisäselittäjien mukaan tulo on nostanut mallin selitysasteen 23 %:n tasolle. Kotitalouden päämiehen koulutuksen tulovaikutus on malliin (1) verrattuna jopa hieman kasvanut. Tämä saattaa johtua kotitalouden päämiehen potentiaalisen työkokemusmuuttujan T ottamisesta mukaan analyysiin. T ei kuitenkaan saanut tässä vaiheessa tilastollisesti merkitsevää regressiokerroimen arvoa. Sen etumerkki on kuitenkin a priori oletuksen mukainen. Mallissa (2) on lisäksi yhdeksän indikaattorimuuttujaa. Naispäämiehi-

syysindikaattorin regressiokertoimen estimaatti on tilastollisesti erittäin merkitsevä, sen arvo -0.466 kuvaa seuraavaa: jos kotitalouden päämiehenä on nainen, niin kotitalouden käytettävissä oleva tulo on 46.6% pienempi verrattuna niiden kotitalouksien tuloon, joiden päämiehenä on mies, mutta joissa mallin muut selittävät muuttujat saavat samat arvot.

Analyysin tässä vaiheessa emme ryhdy tarkemmin selvittämään mistä naispäämiehisyyssindikaattorin kertoimen suuri arvo johtuu. Todetaan vain, että kotitaloudet, joissa päämies on nainen, eroavat rakenteeltaan melko varmasti miespuolisten päämiesten kotitalouksista, joten kotitalouden rakennetekijöiden vaikutus kotitalouden käytettävissä olevaan tuloon heijastuu indikaattorin d_1 kautta. Kotitalouden elinkeinoa kuvaavien indikaattoreiden (d_2, \dots, d_9) merkitys kotitalouden käytettävissä olevan tulon selittäjinä on mallissa suhteellisen vähäinen. Vain indikaattoreiden d_2 ja d_9 parametrien estimaatit saavat tilastollisesti merkitsevät arvot. Muuttujan d_2 kerroin -0.148 kuvaa seuraavaa: jos kotitalouden päämies harjoittaa luontaistaloutta, kotitalouden käytettävissä oleva tulo on 14.8% :a pienempi kuin tulo kotitalouksissa, joissa päämies on teollisuuden palveluksessa, ceteris paribus¹⁾. Vastaava tulkinta voidaan antaa muuttujan d_9 parametrin estimaatille.

Jatketaan analysointia ja tutkitaan mallia (3), jossa lisäselittäjinä ovat erilaiset kotitalouden rakennetekijät. Mallin selitysaste on nousut edellisiin malleihin verrattuna huomattavasti, lähes 47% :iin. Lähes

1) Tämä sanonta viittaa mallin muihin selittäviin muuttujiin, joiden oletetaan pysyvän muuttumattomina, ts. samoina kuin vertailutilanteessa. Mitään oletuksia mallissa mainitsemattomien tekijöiden samana pysymisestä ei luonnollisestikaan voida tehdä, vaan näiden on annettu ja annettava sopeutua tarkasteltavan muuttujakombinaation kulloinkin kuvaamaan tilanteeseen.

puolet kotitalouksien käytettävissä olevan tulon logaritmin vaihtelusta voidaan siis selittää päämiehen koulutuksesta ja kotitalouden sosioekonomisten tekijöiden eroista johtuviksi. Kotitalouden päämiehen koulutuksen lisääntyminen vuodella nostaa mallin mukaan kotitalouden käytettävissä olevaa tuloa 11.6 %:a, ceteris paribus. Koulutuksen tulovaikutus on hieman laskenut edellisiin malleihin verrattuna, mutta muutos on vielä suhteellisen vähäinen. Kotitalouden potentiaalinen työkokemus T on saanut yksisuuntaisessa t -testissä tilastollisesti merkitsevän regressiokertoimen, $\gamma_2^{(3)} = 0.003$. Vuoden lisäys päämiehen potentiaalisessa työkokemuksessa merkitsee (ceteris paribus) kotitalouden käytettävissä olevan tulon kasvua siis vain 0.3 %:lla. Tässä yhteydessä on huomattava kotitalouden päämiehen koulutuksen K ja potentiaalisen työkokemuksen T määritelmällinen riippuvuus. Kotitalouden päämiehen ikä A saadaan potentiaalisen työkokemuksen määritelmästä seuraavasti:

$$(1) \quad A = 7 + K + T$$

Yhtälö (1) merkitsee, että tietyn ikäisellä päämiehellä koulutusaika voi lisääntyä vain potentiaalisen työkokemuksen kustannuksella ja lisäkoulutusvuoden aikana päämies siis menettää vuoden työkokemuksen antaman tulovaikutuksen. Näin mallissa (3) kotitalouden päämiehen lisäkoulutusvuoden antama lisäys kotitalouden käytettävissä olevaan tuloon hieman supistuu, kun päämiehen potentiaalisen työkokemuksen vastaava supistuminen otetaan huomioon. Vuoden lisäkoulutuksen ja siihen liittyvän vuoden työkokemuksen menetyksen aiheuttama tulonlisäys on $11.6 \% - 0.3 \% = 11.3 \%$, ceteris paribus. Tämä laskelma vastaa kappaleessa 4.2. kuvion 7 vertailua (A_3, B_3).

Kotitalouden päämiehen potentiaalisen työkokemuksen vaikutus kotitalouden käytettävissä olevaan tuloon mallin (3) mukaan tuntuu kovin pieneltä. Intuitiivisesti uskoisi työkokemuksella olevan suuremman vaikutuksen tuloihin. On huomattava kuitenkin, että selitettävänä muuttujana on kotitalouden käytettävissä olevan tulon logaritmi, eikä kotitalouden päämiehen tulon logaritmi. Kotitalouden käytettävissä oleva tulo karttuu myös muista tulolähteistä kuin päämiehen tuomista tuloista.

Mincer (1960) on tutkinut kotitalouksien tulojen ja vastaavasti niiden päämiesten tulojen eroavaisuuksia. Hän totesi, että kotitalouden tuloihin sisältyvien muiden kuin päämiehen tulojen vaikutus näkyi verrattaessa päämiesten ikä-tuloprofiileja kotitalouksien vastaaviin profiileihin. Kotitalouksien tulojen ikäprofiileissa säilyi päämiehen koulutuksen aiheuttamat eroavaisuudet, mutta profiilien vaihtelu päämiehen iän suhteen katosi, ts. kotitalouksien tulot eivät päämiehen iän suhteen paljoakaan poikenneet toisistaan. Tällöin kotitalouden päämiehen ikä ei yksinään selitä paljoakaan kotitalouksien tulojen vaihtelusta. Mincerin mielestä tämä saattoi johtua nuorten kotitalouksien päämiesten keskimääräistä korkeammasta koulutus pohjasta ja vastaavasti vanhempien päämiesten keskimääräistä matalammasta koulutus pohjasta. Vaikka näin olisikin asianlaita, voi työkokemus olla merkittävä selittäjä mallissa, jossa työkokemuksen lisäksi koulutus on selittäjänä. Tällöin työkokemus ja koulutus ovat keskenään negatiivisesti korreloituneita selittäjiä, joten on kysymys eräänlaisesta multikollineaarisuusilmiöstä. Koska mallissa (3) on selittäjänä kotitalouden päämiehen työkokemus sekä koulutus, ei Mincerin havainto kotitalouksien tulojen ikäprofiilien suhteellisesta tasaisuudesta tässä selitä, miksi kotitalouden päämiehen potentiaalisen työkokemuksen vaikutus kotitalouden käytettävissä olevaan tuloon jäi niinkin pieneksi.

On ilmeistä, että työkokemuksella olisi koulutuksen lisäksi positiivinen vaikutus päämiehen ja vähäisemmässä määrin myös kotitalouden kokonaistuloihin. Ilmeisesti tulonsiirrot, lapsiperheille kohdistuvat vähennykset verotuksessa ja progressiivinen verotus nostavat lapsiperheiden käytettävissä olevaa tuloa suhteessa näiden kokonaistuloihin. Tämän laatuiset tekijät selittänevät miksi työkokemuksen tuloja lisäävä vaikutus voitiin todeta vain heikkona.

Kotitalouksien rakennetekijöistä ensimmäisenä esiintyy mallissa (3) kotitalouden jäsenluku N . Tämä on erittäin merkittävä selittäjä ja jäsenluvun kasvu yhdellä henkilöllä merkitsee kotitalouden käytettävissä olevan tulon kasvua keskimäärin 10.2 %:lla (*ceteris paribus*, ts. kun muut selittäjät säilyttävät valitut arvonsa ja mallissa mainitsemattomat tekijät ovat sopeutuneet jäsenluvun eroon). Jäsenluvun vaikutus kotitalouden käytettävissä olevaan tuloon on suhteellisen suuri, ja sen tulkinnassa täytyy muistaa, että ruokakuntaan kuuluvat aikuis- ja lapsijäsenet vaikuttavat samalla tavalla jäsenlukua N määrättäessä.

Kotitalouksien rakennetekijöitä kuvaavat myös mukaan tulleet kolme uutta indikaattorimuuttujaa d_{10} , d_{11} , d_{12} . Ammattiaktiivien kotitalouksien ($d_{10} = 1$) käytettävissä oleva tulo on mallin mukaan lähes 39 % korkeampi verrattuna niiden kotitalouksien tuloon, jotka eivät ole ammattiaktiiveja, *ceteris paribus*. Kotitalouksien määriteltiin ammattiaktiiviksi, jos sen ammatissa toimivien jäsenten lukumäärä oli suurempi tai yhtäsuuri kuin 2. Ammattiaktiivisuudella on erittäin merkittävä ja erittäin suuri vaikutus kotitalouden käytettävissä olevaan tuloon. Samoin ammattiaktiivisuusindikaattori d_{11} on erittäin merkittävä selittäjä: jos kotitaloudessa ei ole yhtään ammatissa toimivaa jäsentä, mallin (3) mukaan koti-

talouden käytettävissä oleva tulo on 49 % pienempi kuin niiden kotitalouksien tulo, joissa on ammatissa toimivia jäseniä. Esitettyjen indikaattoreiden regressiokertoimet tuntuvat intuitiivisesti järkeviltä, joskin kovin suurilta.

Tarkastellaan vielä, mitä kotitalouden rakennetekijöiden mukaantulo on vaikuttanut muiden indikaattorimuuttujien regressiokertoimiin. Odotetusti naispäämiehisyyssindikaattorin d_1 regressiokerroin on edelliseen malliin nähden pienentynyt. Kotitalouden päämiehen ollessa nainen jää kotitalouden käytettävissä oleva tulo enää 15.4 % pienemmäksi miespuolisten päämiesten kotitalouksien tuloon verrattuna, ceteris paribus. Mallin (2) mukaan vastaava luku oli 46.6 %. Kotitalouden päämiehen elinkeinoindikaattoreiden kertoimet ovat pysyneet suhteellisen merkityksettöminä. Ainoastaan indikaattori d_2 saa erittäin merkitsevän regressiokertoimen, joka lisäksi on huomattavasti kasvanut edelliseen malliin verrattuna. Muuttujan d_2 kerroin -0.301 merkitsee seuraavaa: mikäli kotitalouden päämies harjoittaa luontaistaloutta, kotitalouden käytettävissä oleva tulo on noin 30 % pienempi kuin niiden kotitalouksien tulo, joissa päämies on teollisuuden palveluksessa. Malli (3) osoittaa siis esitetyn kahden elinkeinohaaran palveluksessa toimivien päämiesten kotitalouksien tulojen välillä olevan huomattavaa eroa. Tämä ero on merkityksellinen, kun ottaa huomioon sen, että useiden muiden kotitalouksien käytettävissä olevaan tuloon vaikuttavien tekijöiden vaikutus on mallissa jo huomioitu.

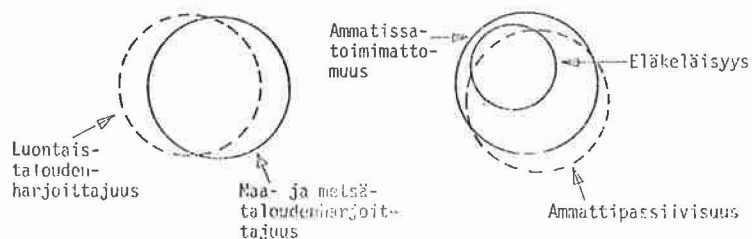
Siirrytään vaihe eteenpäin ja tarkastellaan mallia (4). Lisäselittäjinä aiempiin malleihin verrattuna on tässä erilaisia kotitalouden sosioekonomista asemaa kuvaavia muuttujia. Selitysaste on hieman noussut edelliseen malliin verrattuna, 49 %:iin. Kotitalouden päämiehen koulutus säi-

lyttää asemansa erittäin merkitsevänä kotitalouden käytettävissä olevan tulon selittäjänä. Sen vaikutus on kuitenkin jonkin verran heikentynyt ja mallin mukaan kotitalouden päämiehen lisäkoulutusvuosi nostaa kotitalouden käytettävissä olevaa tuloa 9 %:n verran, ceteris paribus. Kotitalouden päämiehen potentiaalinen työkokemus on saanut merkitsevän regressiokertoimen, jonka suuruus on 0.004. Päämiehen työkokemuksen vaikutus kotitalouden käytettävissä olevaan tuloon on yhä suhteellisen heikko, eikä kovinkaan paljon heikennä päämiehen lisäkoulutusvuoden tuomaa todellista tulohyötyä, joka on $9\% - 0.4\% = 8.6\%$.

Kotitalouden jäsenluvun ja useimpien indikaattoreiden vaikutus kotitalouden käytettävissä olevaan tuloon on pysynyt lähes ennallaan edelliseen malliin verrattuna. Elinkeinoindikaattoreista luontaistaloudenharjoittajuusindikaattori d_2 on menettänyt merkitsevyyttään. Samoin on käynyt ammattipassiivisuusindikaattorille. Tämä johtuu uusien indikaattorimuutusten osittaisesta päällekkäisyydestä ennestään mukana olevien kanssa, mikä aiheuttaa positiivista korrelaatiota ko. indikaattoreiden välille. Jos kotitalous harjoittaa maa- ja metsätaloutta (indikaattori d_{15}), niin sen päämies on melkoisella varmuudella luontaistalouden harjoittaja (indikaattori d_2). Jos taaskotitalous on ammatissa toimimaton kotitalous (indikaattori d_{18}), niin se on hyvin todennäköisesti myös ammattipassiivinen kotitalous (indikaattori d_{11}). Kotitalouden ammatissa toimimattomuudella on hyvin suuri negatiivinen vaikutus kotitalouden käytettävissä olevaan tuloon. Regressiokerroin 0.613 merkitsee noin 61 %:n tuloeroa ammatissa toimimattomien ja työntekijöiden välillä. Vaikutus on vieläkin suurempi, jos ammatissa toimimattomaksi kotitaloudeksi määritellyssä kotitaloudessa ei ole yhtään ammatissa toimivaa jäsentä. Tällöin mukaan tulee vielä ammattipassiivisuusindikaattorin lähes 24 %:n negatiivinen tulovaikutus. Yhdessä tämä negatiivinen vaikutus on $61\% + 24\% = 85$

toogista prosenttia, toisin sanoen tulot putoaisivat aiemmasta 100:sta 43:een (eli $-0.85 = \log(0.43)$). Muista muuttujista määritelmällisesti riippuvainen on myös eläkeläisyysindikaattori d_{12} , jonka regressiokerroimen arvo edelliseen malliin nähden on huomattavasti noussut ja on tässä mallissa tilastollisesti erittäin merkitsevä. Kuviossa 2 havainnollistetaan, miten esitetyt indikaattorimuuttujat voivat kuvata samoja kotitalouden piirteitä.

Kuvio 2: Indikaattoreiden keskinäinen riippuvuus



Indikaattorimuuttujien keskinäinen riippuvuus aiheuttaa tässä erikoislaatuisen multikollinearisuusongelman, jonka vuoksi esim. parametriestimaattien keskivirheet saattavat nousta suuriksi.

Indikaattoreiden regressiokertoimia tulkittaessa on siis muistettava indikaattoreiden mahdollinen määritelmällinen keskinäinen riippuvuus. Esimerkiksi eläkeläisyysindikaattorin d_{12} saamaa positiivista regressiokerrointa 0.582 on tulkittava muistaen, että eläkeläiskotitaloudet ovat (tässä) välttämättä myös ammatissa toimimattomia kotitalouksia ja mahdollisesti ammattipassiiveja kotitalouksia. Tällöin eläkeläiskotitaloudet ovat regressiokertoimensa osoittaman määrän paremmassa asemassa kotitalouden käytettävissä olevan tulon suhteen kuin muut ammatissa toimimattomat kotitaloudet, mutta silti 3.1 % heikommassa asemassa kuin ammatissa toimivat kotitaloudet (koska $0.582 < 0.613$).

Mallin (4) uusista muuttujista sai johtaja- tai toimihenkilökotitalouksia kuvaava indikaattori d_{17} erittäin merkitsevän regressiokertoimen.

Jos kotitalous on johtaja- tai toimihenkilökotitalous, niin kotitalouden käytettävissä oleva tulo on mallin mukaan lähes 18 % suurempi kuin työntekijäkotitalouksien vastaava tulo, ceteris paribus. Tämän indikaattorin mukaan ottaminen on ilmeisesti vaikuttanut kotitalouden päämiehen koulutuksen tulovaikutuksen pieneen laskuun mallissa (4). Tämän ilmiön tulkinta jääköön lukijan tehtäväksi.

Lopuksi tutkitaan mallia (5), jossa kotitalouden käytettävissä olevan tulon selittäjinä on kolme kvantitatiivista muuttujaa sekä 25 kvalitatiivista indikaattorimuuttujaa¹⁾. Kotitalouksia on edellisten mallien indikaattoreiden muodostamien luokkien lisäksi luokiteltu sijaintiteki-
jööidensä suhteen uusien indikaattorimuuttujien avulla. Mallin selity-
aste on noussut lähes 50 %:iin. Kotitalouden päämiehen koulutus-muuttuja on osoittanut hämmästyttävää stabiilisuutta: mallissa (5) kotitalouden päämiehen lisäkoulutusvuosi nostaa kotitalouden käytettävissä olevaa tuloa 9 %:lla, ceteris paribus. Koulutuksen tulovaikutus näyttää asettuvan tälle 9 %:n tasolle, josta sitä eivät entistä yksityiskohtaisemmat-
kaan kotitalouksien luokitukset muuttaneet. Samoin kotitalouden päämiehen potentiaalisen työkokemuksen vaikutus kotitalouden käytettävissä olevaan tuloon on pysynyt hyvin stabiilina, vaikka sen merkitys onkin suhteellisen vähäinen pienen regressiokertoimen vuoksi. Kotitalouden

1) Näin ollen selittäjiä regressiomallissa on 28 kpl ja siis käännettävä momenttimatriisi on dimensiota 29x29. Näin monta muuttujaa sisältävän regressiomallin estimointi vaatii niin paljon laskutyötä, ettei siitä ilman tietokonetta käytännössä suoriuduttaisi. Jopa tietokoneelle näin suurien matriisien kääntäminen on vaativa toimenpide.

käytettävissä oleva tulo kasvaa vain 0.3 %:lla, kun päämiehen potentiaalinen työkokemus lisääntyy vuodella, ceteris paribus. Päämiehen kahdenkymmenen vuoden työkokemus nostaisi siis kotitalouden käytettävissä olevia tuloja ainoastaan 6 % "muuten samanlaiseen", mutta ilman työkokemusta olevan päämiehen kotitalouteen verrattuna. Todettakoon, että näiden "muuten samanlaisten" kotitalouksien päämiesten ikäero on välttämättä 20 vuotta ja että mallissa mainitsemattomien tekijöiden (esim. tulonsiirtojen määrän) on oletettu sopeutuneen näihin eroihin.

Kotitalouden jäsenluvun tulovaikutus on myös pysynyt suhteellisen stabiilina siirryttäessä mallista toiseen. Kotitalouden jäsenluku on erittäin merkitsevä selittäjä, mutta sen heikkona puolena on se, ettei sen avulla voida ottaa huomioon, onko kotitalouden jäsen lapsi vai aikuinen. Malleissa on niinkään oletettu, että esim. jäsenluvun lisääntyminen yhdestä kahteen vaikuttaa yhtä paljon tuloon kuin henkilöluvun kasvu viidestä kuuteen. Malleista voidaan vain lukea, että yksi lisähenkilö nostaa kotitalouden käytettävissä olevaa tuloa keskimäärin 11 %, ceteris paribus. Jatkossa tutkitaan kotitalouden jäsenluku-muuttujan korjaamista tarkoituksenmukaisemmaksi.

Naispäämiehisyyssindikaattorin regressiokerroin on stabilisoitunut malliin (5) mennessä arvoon -0.15 pysyen jatkuvasti tilastollisesti erittäin merkitseväenä. Mallien mukaan kotitalouden päämiehen sukupuolella on kiistaton merkitys kotitalouden käytettävissä olevan tulon vaihtelun selittäjänä "matriarkaalisten" kotitalouksien käytettävissä olevien tulojen jäädessä n. 15 % vastaavanlaisten "patriarkaalisten" kotitalouksien tulojen alapuolelle. "Vastaavanlaisuus" ei kuitenkaan ole tässä itseltään selvä käsite, koska nykyisen käytännön mukaan kotitalous luokitellaan "matriarkaaliseksi" vain kun siihen ei kuulu minkäänlaista miestä.

Tämä johtaa siihen, että yksinäisten naisten lisäksi naispuoliset päämiehet ovat pääasiassa yksinhuoltajia tai leskiä. Kotitalouden päämiehen elinkeino ei mallien mukaan ole kovin merkittävä kotitalouden käytettävissä olevan tulon selittäjä. Elinkeinoindikaattoreiden kertoimet eivät ole tilastollisesti merkitseviä, ainoa poikkeus on luontaistaloudenharjoittajuusindikaattori d_2 , joka vielä mallissa (5) saa melkein merkitsevän kertoimen.

Kotitalouden ammattiaktiivisuudella on erittäin voimakas ja stabiili positiivinen vaikutus kotitalouden käytettävissä olevaan tuloon. Ammatipassiivisuuden negatiivinen tulovaikutus on sitä vastoin heikentynyt uusien muuttajien mukaantulon myötä. Muiden mallissa (4) mukana olleiden indikaattoreiden vaikutus on pysynyt lähes muuttumattomana siirryttäessä malliin (5).

Uusista vain mallissa (5) mukana olevista kotitalouden sijaintia kuvaavista indikaattoreista kaksi on saanut tilastollisesti merkitsevät kertoimet. Indikaattori d_{20} , Väli-Suomalaisuus, on saanut kertoimekseen -0.112 ja indikaattori d_{21} , Pohjois-Suomalaisuus, vastaavasti arvon -0.116 . Kertoimien tulkinta on seuraava: mikäli kotitalous sijaitsee Väli- tai Pohjois-Suomessa, niin kotitalouden käytettävissä oleva tulo on noin 11% pienempi kuin Etelä-Suomessa sijaitsevien kotitalouksien vastaava tulo, ceteris paribus. Kotitalouden sijainnilla näyttää siten olevan merkitystä, kun tarkastellaan kotitalouden tuloja. Sen sijaan kotitalouden asuinyllillä (indikaattorit d_{23} - d_{25}) ei näytä olevan itsenäistä merkitystä tulojen vaihtelua selitettäessä.

Kotitalouksien luokittelijoina käytetyt indikaattorimuuttujat ovat jos-
sakin määrin määritelmällisesti keskenään riippuvia. Tämä häiritsee nii-
den tulkintaa. Seuraavassa luvussa malleja (1) - (5) muokataan edelleen
poistamalla analyysistä määritelmällisesti muiden indikaattoreiden kanssa
riippuvia indikaattoreita sekä osa ei-merkitseviä selittäjiä.

6.2. Parannetut mallit

Estimoidaan kaksi uutta mallia, joissa selittäjinä on edelleenkin kolme kvantitatiivista muuttujaa, mutta indikaattorimuuttujien lukumäärää on laskettu malliin (5) verrattuna yhteentoista. Analyysistä on jätetty pois kotitalouden päämiehen elinkeinoa kuvaavat indikaattorit (d_2, \dots, d_9) osaksi niiden heikon selittävyuden ja osaksi niiden määritelmällisten riippuvuuksien vuoksi lähinnä kotitalouden sosioekonomista asemaa kuvaavien indikaattoreiden kanssa. Lisäksi päämiehen työttömyyttä kuvaava indikaattori d_{13} ja julkisen sektorin palvelusta kuvaava indikaattori d_{14} on poistettu uusista malleista niiden heikon selityskyvyn ja merkittävyyden vuoksi. Samasta syystä yrittäjäyysindikaattori d_{16} ja asuintyyli-indikaattorit (d_{23}, d_{24}, d_{25}) jätettiin pois analyysistä.

Uudet mallit poikkeavat toisistaan ainoastaan kotitalouden jäsenluku-muuttujan suhteen ja ne ovat seuraavaa muotoa:

$$(6) \quad \log Y_d = \gamma_0^{(6)} + \gamma_1^{(6)}K + \gamma_2^{(6)}T + \gamma_3^{(6)}N + \beta_1^{(6)}d_1 + \\ \sum_{i=10}^{12} \beta_i^{(6)}d_i + \beta_{15}^{(6)}d_{15} + \sum_{i=17}^{22} \beta_i^{(6)}d_i + U,$$

$$(7) \quad \log Y_d = \gamma_0^{(7)} + \gamma_1^{(7)}K + \gamma_2^{(7)}T + \gamma_3^{(7)}\log(n) + \beta_1^{(7)}d_1 + \\ \sum_{i=10}^{12} \beta_i^{(7)}d_i + \beta_{15}^{(7)}d_{15} + \sum_{i=17}^{22} \beta_i^{(7)}d_i + U,$$

missä

Y_d = kotitalouden käytettävissä oleva tulo (mk/v)

K = kotitalouden päämiehen koulutus vuosissa

T = kotitalouden päämiehen potentiaalinen työkokemus

N = kotitalouden jäsenluku

n = kotitalouden jäsenten lukumäärä yhteensä, kun aikuisjäsen saa arvon 1 ja lapsijäsen arvon 0.5

U = jäännöstermi

$$d_i = \Phi(p_i) = \begin{cases} 1, & \text{jos väite } p_i \text{ on tosi} \\ 0, & \text{jos väite } p_i \text{ on epätosi} \end{cases}$$

p_1 = kotitalouden päämies on nainen

p_{10} = kotitalous on ammattiaktiivi

p_{11} = kotitalous on ammattipassiivi

p_{12} = kotitalous on eläkeläiskotitalous

p_{15} = kotitalous harjoittaa maa- ja metsätaloutta

p_{17} = kotitalous on johtaja- tai toimihenkilökotitalous

p_{18} = kotitalous on pääasiassa ammatissa toimimaton

p_{19} = kotitalous on suurtilalliskotitalous

p_{20} = kotitalous sijaitsee Väli-Suomessa

p_{21} = kotitalous sijaitsee Pohjois-Suomessa

p_{22} = kotitalous sijaitsee maalaiskunnassa

Mallissa (7) on vanhan jäsenlukumuuttujan N tilalle määritelty uusi muuttuja $\log(n)$, jonka avulla pyritään täsmällisemmin ottamaan huomioon kotitalouden jäsenistön rakenne. Sen tulkintaan palataan tarkasteltaessa estimointituloksia.

Mallien (6) ja (7) estimointitulokset on esitetty liitteessä 7 mallien (1) - (5) tapaan. Lähdetään tutkimaan mallin (6) antamia tuloksia. Selitysaste on hieman laskenut malliin (5) verrattuna pysyen kuitenkin lähellä 50 prosenttia. Tämä olikin odotettua, koska mallissa (6) on 14 indikaattorimuuttujaa vähemmän kuin mallissa (5). Kotitalouden päämiehen koulutuksen vaikutus kotitalouden käytettävissä olevaan tuloon on pysynyt muuttumattomana malliin (5) verrattuna: kotitalouden päämiehen lisäkoulutusvuosi lisää kotitalouden käytettävissä olevaa tuloa 9 %, ceteris paribus. Samoin kotitalouden päämiehen potentiaalinen työkokemus ja kotitalouden jäsenluku ovat pysyneet merkitsevinä selittäjinä. Tarkasteltaessa eri muuttujien kertoimia voidaan havaita, että poisjätettyjen indikaattorimuuttujien poistaminen näkyy ainoastaan kahden indikaattorin d_{15} ja d_{19} muuttuneissa kertoimissa. Aikaisemmissa malleissa tilastollisesti ei-merkitseviä arvoja saanut muuttuja d_{15} (maa- ja metsätaloudenharjoittajuusindikaattori) saa nyt mallissa (6) tilastollisesti erittäin merkitsevän arvon. Tämä johtuu muuttujan d_2 (luontaistaloudenharjoittajuusindikaattori) poistamisesta, jolloin muuttuja d_{15} ottaa sen selittäjän roolin. Muuttujan d_{15} saama kerroin -0.240 merkitsee seuraavaa: jos kotitalous harjoittaa maa- ja metsätaloutta, niin sen käytettävissä oleva tulo on 24 % pienempi kuin vastaava tulo työntekijä- tai yrittäjäkotitalouksissa, ceteris paribus. Mallin (6) mukaan erilaisten kotitalouksien sosioekonomisten ryhmien välillä näyttää olevan tuloeroja, vaikka kotitalouksien useat muut tekijät, kuten esimerkiksi päämiehen koulutus, potentiaalinen työkokemus ja perhekoko, olisivatkin näissä ryhmissä samanlaisia. Maa- ja metsätalouden harjoittajakotitaloudet (indikaattori $d_{15} = 1$) näyttävät olevan keskimääräistä heikommassa ja johtaja- tai toimihenkilökotitaloudet (indikaattori $d_{17} = 1$) keskimääräistä paremmassa asemassa kotitalouden käytettävissä olevan tulon suhteen. Huomattavasti keskimää-

räistä heikommassa asemassa ovat ammatissa toimimattomat kotitaloudet (indikaattori $d_{18} = 1$), kuten jo aikaisemmin esitetyissä malleissa on tullut ilmi. Tässä yhteydessä on syytä pohtia, mitä itse käytettävissä olevan tulon käsite merkitsee sosioekonomisten ryhmien tuloeroavaisuuksia tulkittaessa. Kotitaloustiedustelussa on pyritty erittäin tarkasti määrittelemään ja mittaamaan kaikki mahdolliset tulo- ja kuluerät, joista kotitalouden käytettävissä oleva tulo on muodostettu. Tietoja on kerätty kolmesta eri lähteestä; haastatteleamalla, veroviranomaisilta ja kansaneläkelaitoksen rekisteristä. Saattaa kuitenkin olla, että esimerkiksi maa- ja metsätalouden piirissä on sellaisia arvioon perustuvia tulo- tai kulueriä, joiden vaikutus kotitalouden "varsinaiseen" käytettävissä olevaan tuloon jää odotettua vähäisemmäksi. Tämän tapaiset laksennalliset kotitalouden käytettävissä olevan tulon eroavaisuudet eri sosioekonomisissa ryhmissä antavat aihetta varovaisuuteen tehtäessä eri ryhmien välisiä vertailuja.

Indikaattorin d_{15} lisäksi merkitsevyyttään lisäsi edellisiin malleihin verrattuna indikaattori d_{19} (suurtilallisuusindikaattori), joka nyt saa mallissa (6) tilastollisesti melkein merkitsevän regressiokerroimen. Jos kotitalouden tilan kokonaispinta-ala on vähintään 100 ha, niin kotitalouden käytettävissä olevat tulot jäävät 24 % pienemmäksi kuin niillä kotitalouksilla, joilla tällaista tilaa ei ole muiden mallissa olevien selittäjien saadessa samat arvot. Intuitiivisesti oikeammalta tuntuisi suurtilallisuuden positiivinen tulovaikutus, mutta negatiivisellekin vaikutukselle voidaan kehittää selityksiä. Ensinnäkään indikaattorin d_{19} negatiivinen regressiokerroin ei väitä, että suurtilalliskotitaloudet saivat muita pienemmän käytettävissä olevan tulon. Näin tapahtuisi ainoastaan silloin, jos muut mallissa olevat selittäjät saivat samat arvot.

Tämä merkitsisi esimerkiksi, että työntekijäkotalouksien ja suurtilalliskotalouksien päämiehillä olisi yhtä pitkä koulutus ja että kotitalouksien jäsenluvut olisivat samat, kotitaloudet sijaitisivat samalla alueella jne. Voidaan tuskin olettaa, että asianlaita olisi näin. Muuttujan d_{19} negatiivinen kerroin ehkä kertoo seuraavaa: suurtilalliskotalouksien päämiehet ovat keskimääräistä korkeammin koulutettuja, mutta koulutuksen hyöty kotitalouksien käytettävissä olevan tulon kasvuna jää suurtilallisilla keskimäärin 24 % pienemmäksi kuin ei-suurtilallisilla kotitalouksien päämiehillä. Tällöin -24 % osoittaisi koulutuksen tavanomaista huonompaa hyödynnettävyyttä suurtilallisten kohdalla. Muut selitykset ovat luonnollisesti myös mahdollisia ja kilpailevien selitystyörytysten arviointi edellyttäisi entistä yksityiskohtaisempaa aineistoa ja analyysiä.

Siirrytään tarkastelemaan viimeistä estimoitua mallia (7), joka poikkeaa mallista (6) vain kotitalouden jäsenlukua kuvaavan muuttujan suhteen. Tämä muutos on nostanut mallin selityssasteen yli 52 prosenttiin, mikä on korkein estimoituista malleista laskettu selittävyys. Kotitalouden päämiehen koulutuksen tulovaikutus on stabilisoitunut 9 %:n tienoille, mallissa (7) se on 8.8 %. Uusi jäsenten lukumäärää kuvaava muuttuja ei siten muuttanut koulutuksen kerrointa. Tarkastellaan lähemmin muuttujan $\log(n)$ regressiokertoimen sekä mallin (6) jäsenlukumuuttujan N kertoimen tulkin-toja. Jos jäsenluku kasvaa yhdestä aikuisesta kahteen aikuiseen, niin kotitalouden käytettävissä oleva tulo nousee mallissa (7) $55.4(\log 2) \% = 38 \%$, ceteris paribus, kun se mallin (6) mukaan nousee vain 11 %. Jos taas jäsenluku kasvaa viidestä aikuisesta kuuteen, niin kotitalouden käytettävissä oleva tulo kasvaa mallin (7) mukaan enää vain 10 %, kun se mallin (6) mukaan edelleen nousisi 11 %. Mikäli kotitalous kasvaisi lapsi-

määrän lisääntymisen vuoksi merkitsisi se mallissa (7) jatkuvasti pienenevää kotitalouden käytettävissä olevan tulon lisäystä, kun taas lapsimäärän lisäyksen indusoima tulojen kasvu mallissa (6) pysyy jatkuvasti 11 %:ssa. Kotitalouden jäsenlukumuuttuja $\log(n)$ on suuremman selityskyvyn omaava tulon vaihtelun selittäjä kuin muuttuja N . Tämä merkitsee sitä, että muuttuja $\log(n)$ on vienyt joiltain muilta selittäjiltä selittävyttä, sillä mallin selityssaste nousi suhteellisen vähän.

Ensimmäinen muuttuja, jonka merkitsevyys $\log(n)$:n mukaantulon myötä pieneni oli kotitalouden päämiehen potentiaalinen työkokemus. Sen regressio-kerroin pieneni 0.003:sta 0.002:een, joka kuitenkin yhä poikkeaa yksisuuntaisen t-testin nojalla 'melkein merkitsevästi' nolasta. Vuoden lisäys työkokemukseen nostaisi perheen käytettävissä olevia tuloja vain 0.2 %.

Sen sijaan mallissa (6) erittäin merkitsevän kertoimen saanut naispäämiehisyyssindikaattori ei enää mallissa (7) saa tilastollisesti merkitsevää kerrointa. Tuntuu ymmärrettävältä, että kotitalouden kokoa suhteellisen täsmällisesti kuvaava muuttuja vie naispäämiehisyyssindikaattorilta selittävyttä, mutta että se vie selittävyyden kokonaan, on hämmästyttävää. Mallin (7) mukaan päämiehen sukupuoli ei ole merkitsevä selittäjä selitettäessä kotitalouden käytettävissä olevaa tuloa, kun mallin muut selittäjät saavat samat arvot. Tiettyjen kotitalouksien keskimääräistä pienemmät tulot selittyivät aiemmin naispäämiehisyyssindikaattorin d_1 avulla, mutta nyt selitys tapahtuu jäsenlukumuuttujan $\log(n)$ kautta. Ainoa muutos malliin (6) nähden on kokomuuttujan N korvaaminen muuttujalla $\log(n)$. Uusi muuttuja $\log(n)$ erottelee aiempaa selvemmin pienet perheet (joilla n on 1 tai 1.5) omaksi ryhmäkseen. Ilmeisestikin valta-

osa naispäämiesten kotitalouksista on juuri näitä pieniä kotitalouksia eli vähänkin suuremmasta kotitaloudesta löytyy jonkinlainen mies päämieheksi. Tällaisessa tapauksessa kotitalouden pienuus ja naispäämiehisyydet korreloi voimakkaasti keskenään ja tarjoavat vaihtoehtoiset selitysmahdollisuudet ko. kotitalouksien keskimääräistä pienemmille tuloille. Kyseessä on eräänlainen multikollineaarisuusilmiö.

Kotitalouden ammattiaktiivisuutta kuvaava indikaattori d_{10} on säilyttänyt vahvan tulovaikutuksensa mallissa (7): jos kotitalous on ammattiaktiivinen, niin sen käytettävissä oleva tulo on noin 31 % suurempi verrattuna ei-ammattiaktiivien kotitalouksien vastaavaan tuloon, ceteris paribus. Kotitalouden ammattipassiivisuutta kuvaava indikaattori d_{11} on sen sijaan menettänyt merkitsevyytensä nähtävästi ammatissa toimimattomia kotitalouksia kuvaavalle indikaattorille d_{18} , jonka regressiokerroin on hivenen noussut. Eläkeläiskotitalouksia kuvaava indikaattori d_{12} on erittäin merkitsevä selittäjä ja sen vaikutus voidaan tulkita seuraavasti, kun sen määritelmällinen riippuvuus indikaattorin d_{18} kanssa huomioidaan: mikäli kotitalous on eläkeläiskotitalous ja näin ollen ammatissa toimimaton ($d_{18} = 1$), niin kotitalouden käytettävissä oleva tulo on $0.551 - 0.665 = -0.114$ eli 11.4 % pienempi kuin vastaava tulo ammatissa toimivissa kotitalouksissa, ceteris paribus. Ammatissa toimimattomilla mutta ei-eläkeläiskotitalouksilla tulot jäävät keskimäärin 66.5 % vastaavalaisten ammatissa toimivien kotitalouksien tuloja pienemmiksi. Maataloudenharjoittajuusindikaattorin negatiivinen tulovaikutus on säilynyt lähes ennallaan ja on mallissa 25.3 %. Samoin johtaja- tai toimihenkilökotitalouksia kuvaavan indikaattorin d_{17} 18 %:n positiivinen tulovaikutus näyttää pysyvän mallista toiseen samana. Jälleen kerran tuntuu aiheelliselta muistuttaa, että esitettyihin tuloksiin voi vaikuttaa käytetyn tulokäsitteen mahdollinen heikko validiteetti eri sosioekonomisissa ryhmissä.

Suurtilallisuus- ja maalaiskuntaisuusindikaattorit d_{19} ja d_{22} säilyttävät asemansa mallissa (7) jälkimmäisen jäädessä edelleen ilman tilastollista merkitsevyyttä. Kotitalouden sijaintia kuvaavat indikaattorit d_{20} ja d_{21} kertovat seuraavaa: jos kotitalous sijaitsee Väli- tai Pohjois-Suomessa, niin kotitalouden käytettävissä oleva tulo on noin 12 % pienempi kuin Etelä-Suomen kotitalouksien vastaava tulo, ceteris paribus. Kotitalouksien sijaintitekijät ovat säilyttäneet merkitsevyytensä kotitalouksien tulojen selittäjinä.

Etelä-Suomessa asuvilla kotitalouksilla näyttää olevan keskimääräistä korkeammat käytettävissä olevat tulot. Tämä on merkittävä havainto, kun ottaa huomioon, että mm. kotitalouden päämiehen koulutus, johon usein viitataan alueellisten tuloerojen aiheuttajana, on mallissa vakioitu useiden muiden kotitalouksien ominaisuuksien myötä. Malli (7) kertoo myös sen, että pohjois-suomalaisten ja väli-suomalaisten kotitalouksien käytettävissä oleva tulo on lähes sama, jos mallin muut selittäjät saavat samat arvot. Näiden alueiden välillä ei eroa tässä suhteessa näytä olevan.

Edellä on esitetty suhteellisen laajojen regressiomallien avulla kuvaus käytettävissä olevien tulojen eroista erilaisilla luokitelluissa suomalaisissa kotitalouksissa vuonna 1971. Analyysi on ollut varsin yksityiskohdainen, jonkun mielestä ehkä liiankin pikkutarkka, mutta tämä on ollut välttämätöntä, jotta kotitalouksien päämiesten koulutuksen tulovaikutus olisi tullut mahdollisimman luotettavasti selvitetyksi. Seuraavassa kappaleessa palautetaan mieleen, mitä estimoidut mallit kertovat koulutuksen tulovaikutuksesta ja "unohdetaan" mallien antama muu tieto siinä määrin, kuin se on mahdollista.

6.3. Kotitalouden päämiehen koulutuksen tulovaikutus

Edellä on esitetty seitsemän, kotitalouden käytettävissä olevan tulon logaritmia selittävän regressiomallin estimointitulokset. Kussakin mallissa on yhtenä selittäjänä kotitalouden päämiehen koulutus vuosissa. Ensimmäisessä mallissa, jossa ainoana selittäjänä oli kotitalouden päämiehen koulutus, koulutusmuuttuja jo osoittautui erittäin merkitseväksi selittäjäksi. Teoreettisten pohdiskelujen pohjalta pystyttiin kuitenkin ennakoimaan, että mallin (1) antama koulutuksen positiivinen tulovaikutus 12.4 % koulutusvuotta kohden yliarvioi lisäkoulutusvuoden todellisen vaikutuksen. Konkreettisesti ilmaistuna malli (1) kertoo, kuinka paljon päämiehen koulutusvuosien mukaisesti luokiteltujen kotitalouksien käytettävissä olevat tulot vaihtelevat syystä tai toisesta ryhmästä toiseen. Koulutusvuosien erojen lisäksi useissa muissakin muuttujissa esiintyy systemaattisia eroja ryhmien kesken, joten tuloerojen ei voi väittää johtuvan kokonaan vain koulutusvuosien vaihtelusta. Tästä huolimatta malli (1) kelpaa tulojen ennustamiseen, kun ainoa tieto kotitaloudesta on sen päämiehen koulutus vuosissa.

Mallissa (2) muiden kotitalouden päämiestä kuvaavien selittäjien mukaantulo hieman nosti kotitalouden päämiehen tulovaikutusta 12.7 %:iin. Emme ole nähneet tarpeelliseksi koetella lukijan kärsivällisyyttä havainnon mahdollisilla selityksillä.

Seuraavassa mallissa (3) kotitalouden rakennetekijöiden mukainen luokittelu laski kotitalouden päämiehen tulomuuttujan kertoimen (satakerran) 11.6 %:iin. Odotettu väheneminen jäi kuitenkin pieneksi. Vasta mallissa (4), johon oli lisätty kotitalouden sosioekonomisen aseman mukaisia luokituksia, koti-

talouden päämiehen koulutuksen arvioitu tulovaikutus laski malliin (1) verrattuna merkittävästi 9 %:n tasolle. Tämä osoittautui mallin luokitusten laajentamisen ja mallin muokkauksen jälkeenkin tasoksi, jolle koulutuksen tulovaikutus näyttää stabiloituvan. Termi $0.09\Delta K$ vastaa sitä keskimääräistä tuloeroa, joka aineistossa on havaittu sellaisten kotitalouksien välillä, joiden päämiesten koulutusero on ΔK vuotta ja muut kotitalouksien eroja kuvaavat selittävät muuttujat saavat samat arvot (mallissa mainitsemattomien tekijöiden vaihdellessa tuntemattomalla tavalla). Tässä mielessä voidaan päätellä, että kotitalouden päämiehen yksi lisäkoulutusvuosi nostaa keskimäärin kotitalouden käytettävissä olevaa tuloa noin 9 %:lla, ceteris paribus. Tarkastellaan asiaa lähemmin mallin (7) avulla. Kysytään, paljonko kotitalouden käytettävissä oleva tulo keskimäärin muuttuisi, jos päämies olisi opiskellut vuoden enemmän ja hänen työkokemuksensa olisi siten vuoden pienempi, mutta muut mallissa mukana olevat tekijät pysyisivät ennallaan? Lisävuosi koulutuksessa lisäisi tuloja $(8.8 \pm 1.0) \%$ ja vuosi vähemmän työkokemusta vähentäisi tuloja $(0.2 \pm 0.1) \%$. Muiden selittäjien pysyessä ennallaan eivät nämä muuta tuloja, joten tulot nousisivat keskimäärin $(8.6 \pm 1.0) \%$. (Keskivirhe $s = 1.005$ on laskettu käyttäen kaavaa $s^2 = s_1^2 + s_2^2$ riippumattomien virheiden summan varianssille). Jos samanaikaisesti muissa selittäjissä tapahtuu muutoksia huomioidaan näiden tulovaikutukset erillisten laskelmien avulla. Esimerkiksi kahden aikuisen ja yhden lapsen kotitaloudessa perheenlisäys nostaa jäsenlukumuuttujaa n arvosta 2.5 arvoon 3, joten siitä muodostuva tulovaikutus on $0.554 \log(3/2.5) = 0.554(0.182) = 0.101$ eli 10.1 %. Tämä ei toteudu automaattisesti, vaan edellyttää esimerkiksi kotitalouden työpanoksen (tai muiden mallissa esiintymättömien tekijöiden) sopeutumista uuteen tilanteeseen.

Käytettäessä mallia (7) ennustamiseen tilanteessa, jossa kaikkien selittävien muuttujien arvot tunnetaan poikkeavat tulojen logaritmillemme laaditut ennusteet todellisista arvoista siten, että tavallisissa ("keskimääräisissä") tilanteissa ennustevirheiden hajonta on noin 0.42 eli 42 %. Ennustevirheillä on taipumus kasvaa, kun selittävät muuttujat saavat poikkeuksellisia arvoja. Vertailun vuoksi todettakoon, että ennustettaessa aina kotitalouden logaritmisen tulon saavan keskimääräisen arvon 9.7 on ennustevirheiden hajonta keskimäärin 60 %. Ennusteiden luotettavuus on siis selvästi parantunut, mutta ymmärrettävästi varsin suuret ennustevirheet ovat yhä mahdollisia.

Voidaanko edellä esitettyjä tuloksia pitää riittävänä osoituksena siitä, että kotitalouden päämiehen koulutuksen "todellinen" tulovaikutus suomalaisissa aktiivikotitalouksissa on löytynyt? Luonnollisestikaan lopullista vastausta kysymykseen ei voida kotitaloustiedusteluun sisältyvien tietojen perusteella antaa, mutta uskomusten tai toiveiden ohella voidaan kyllä esittää tosiasia-aineistoon perustuvia mielipiteitä. Kovarianssianalyysin keinoin on annetusta tietoa-aineuksesta muodostettu mahdollisimman homogeenisia kotitalouksien luokkia, jotta päämiehen koulutuksen nettutulovaikutus saataisiin aineistosta esiin. Luokituksen laajentaminen mallista (4) malliin (5), jossa koulutuksen lisäksi oli 2 kvantitatiivista ja 25 kvalitatiivista selittäjää, ei enää muuttanut kotitalouden päämiehen koulutuksen tulovaikutuksen suuruutta. Ei näytä olevan perusteita esittää edes mihin suuntaan mahdollinen yksityiskohtaisempi kotitaloustiedustelun aineiston luokitus muuttaisi koulutuksen tulovaikutusta saadusta (8.8 ± 1.0) %:sta.

Tutkimuksen teoreettisessa osassa viitattiin mahdollisuuteen, että lahjakkuuden (kyvykkyyden, älykkyyden) ja koulutuksen todennäköinen positiivinen korrelaatio aiheuttaa positiivista harhaa koulutuksen kertoimeen tuloyhtälöissä, joissa lahjakkuustekijää ei ole selittäjänä. Lahjakkuustekijää ei esiinny tässä tutkimuksessa estimoiduissa malleissa selittäjänä siitä yksinkertaisesta syystä, ettei kotitaloustiedustelun ns. rakennetiedosto sisällä tämäläisiä tietoja. Suoraan ei siten ole voitu mitata ja eliminoida lahjakkuustekijöiden mahdollisia tulovaikutuksia. Vaikka kyvykkyyttä (esim. älykkyydosamäärällä mitattuna) ei olekaan mukana estimoiduissa yhtälöissä selittäjänä ei tästä välttämättä seuraa, että koulutuksen tulovaikutus on estimoitu harhaisesti. Näin olisi vain mikäli kyvykkyystekijän lisääminen malliin muuttaisi merkittävästi juuri koulutuksen regressiokerrointa. On luultavaa, että kyvykkyyden lisääminen malliin muuttaisi joidenkin sosioekonomisten tekijöiden regressiokertoimia, mutta jättäisi koulutusmuuttujan kertoimen suurin piirtein ennalleen mikäli selittäjinä olevat sosioekonomiset muuttujat ovat toimineet kyvykkyyden korvikkeina siten, että ne ovat selittäneet myös kyvykkyyden vaihtelusta johtuvia tuloeroja.

Kovarianssianalyysin avulla voitaisiin tavanomaiseen tapaan selvittää tietyn vaikutussuhteen (esim. koulutuksen tulovaikutuksen) suuruus siinä tilanteessa, jossa kaikki asiaan vaikuttavat "häiriötekijät" ovat mukana analyysissä. Tutkimuksessa on sovellettu kovarianssianalyysiä hieman tavanomaisesta poikkeavalla tavalla käyttäen ideaalisten "häiriötekijöiden" (esim. kyvykkyyden) sijaan niiden kanssa korreloivia ja niitä edustavia korvikemuuttujia. Vaikka siis eräs oleellinen selittäjä puuttuukin analyysistä ei tästä seuraa, että erään toisen selittäjän

vaikutus on tullut väärin arvioitua, koska tavanomaiset regressiomallin oletukset ovat riittävät, mutta ei välttämätön ehto tietyn regressio-kertoimen harhattomuudelle, vrt. Goldberger (1964, s. 197). Riittävän laajoissa tuloyhtälöissä ei siten lahjakkuustekijän puuttuminen ole merkittävä haitta. Kysymykseksi jää näin ollen, milloin tuloyhtälöä voidaan pitää riittävän laajana.

Käytettävissä ollut tietoaineisto ei anna mahdollisuutta mallien oleelliseen laajentamiselle.

Malleja voitaisiin tosin tarkentaa erottamalla esimerkiksi lasten lukumäärä omaksi selittäjäksi, tarkastelemalla erikseen kotitalouksia, joissa on molemmat vanhemmat, tai tutkimalla tulosten eroja maanviljelijä-kotitalouksien ja muiden kotitalouksien välillä. Edellä suoritettussa analyysissä tämántapaisten tekijöiden vaikutusten on annettu ilmetä vain mallien vakiotermin siirtoina. Oleellista tulosten tulkintaan ja luotettavuuteen vaikuttavaa lisäinformaatiota, jota tutkimusresurssien puitteissa ei ole ollut mahdollista hankkia tai jota ei ole yksinkertaisesti saatavissa, olisimme mielellään käyttäneet esim. kotitalouden työpanoksesta ja erityisesti sen päämiehen lahjakkuudesta. Lahjakkuustekijän mukaanottaminen ilmeisesti pienentäisi koulutuksen tulovaikutusta parhaiden malliemme 9 %:sta edelleen jonkin verran, mahdollisesti jopa pari prosenttiyksikköä. Tässä on kuitenkin tyydyttävä subjektiivisiin arvioihin, koska tarvittavia tietoja analyysin suorittamiseksi ei ole olemassa.

7. LOPPUTOTEAMUKSIA

Tarkastellaan vielä lyhyesti, miten asetettu tutkimustehtävä, kotitalouden päämiehen koulutuksen tulovaikutuksen suuruuden arvioiminen, on suoritettu ja mitä mahdollisesti voitaisiin jatkotutkimuksilla edelleen selvittää.

Parhaissa malleissa kotitalouksien käytettävissä olevan tulon logaritmin vaihtelusta voitiin selittää lähes 50 % päämiesten koulutuseroilla sekä kotitalouksien sosioekonomisilla tekijöillä. Käytetyssä 1000 kotitalouden poikkileikkausaineistossa selitysastetta voidaan pitää hyvänä.

Kotitalouden päämiehen koulutus on erittäin merkitsevä kotitalouden käytettävissä olevan tulon selittäjä. Päämiehen yksi lisäkoulutusvuosi nostaa kotitalouden käytettävissä olevaa tuloa keskimäärin 8.6 prosenttia, kun yhden vuoden työkokemuksen samanaikainen väheneminen otetaan huomioon (kotitalouden muut ominaisuudet pidetään ennallaan). Eksplisiittisen lahjakkuusmuuttujan puuttumisen vuoksi saatu arvio luultavasti antaa jonkin verran liian positiivisen kuvan lisäkoulutuksen vaikutuksesta kotitalouden käytettävissä olevaan tuloon. Koulutuksen tulovaikutus 8.6 % on vuoden lisäkoulutusta vastaava keskimääräinen tulonlisäys niiden kotitalouksien kohdalla, jotka ovat analyysiin sisältyvien sosioekonomisten tekijöiden suhteen samanlaisia, mutta joissa ilmeisesti päämiehen lahjakkuus ja perhetausta jonkin verran mutta systemaattisesti vaihtelevat koulutusvuosien mukana. Koska koulutuksen määrä, lahjakkuus ja 'hyvä perhetausta' korreloivat positiivisesti keskenään, karakterisoi niitä henkilöitä, joilla koulutusmäärä on keskimääräistä korkeampi ilmeisesti myös keskimääräistä suurempi lahjakkuus ja parempi kotitausta verrattuna samanlai-

seen sosioekonomiseen ryhmään kuuluvaan henkilöön, jolla koulutus on vähäisempi. Näin ollen on vuoden lisäkoulutuksen vaikutukseksi analyysissämme laskettu myös osa perityn lahjakkuuden ja perhetaustan aiheuttamasta tulonlisäyksestä eikä vain sitä lisätuloa, jonka 'keskiarvohenkilö' voisi odottaa saavansa vuoden lisäkoulutuksesta. Analyysimme tulkintaa vaikeuttaa siis se, että saadun koulutuksen määrää säätelevät yhteiskunnassamme sellaiset henkilökohtaiset ominaisuudet, jotka josinänsä ovat tulonmuodostuksen kannalta edullisia. Näiden koulutuksen kanssa korreloivien henkilökohtaisten ominaisuuksien aiheuttaman harhan numeerista suuruutta ei kuitenkaan olemassa olevan tietoaaineiston perusteella ole ollut mahdollista selvittää.

Käytettävissä oleva havaintoaineisto on muutoinkin rajannut mahdollisten kovarianssianalysimallien valintaa: eräitä kotitalouksien käytettävissä olevaan tuloon mahdollisesti vaikuttavia tekijöitä ei ole voitu ottaa mukaan analyysiin siitä syystä, että tietoja ei ole ollut saatavilla. Lukija lienee kiinnittänyt huomiota muiden perheenjäsenten koulutusmuuttujien ja ammattien sekä kotitalouden kokonaistyöpanoksen puuttumiseen analyysistä. Voitaneen kuitenkin sanoa, että tutkimusaineistossa havaittu keskimääräinen koulutuksen tulovaikutus antaa hyvän lähtökohdan asian tarkemmalle käsittelylle.

Jatkossa voitaisiin esimerkiksi tutkia, mille tasolle tämän tutkimuksen eri osaryhmillä (esimerkiksi maanviljelijöillä ja muilla kotitalouksilla) koulutuksen tulovaikutus asettuu. Kotitalouksien henkilörakenne voitaisiin niinikään tarkemmin huomioida lisäämällä malleihin lasten lukumäärää kuvaava muuttuja ja erottaa yksinhuoltajakotitaloudet omaksi erityisryhmäkseen. Jatkotutkimuksissa olisi myös syytä kiinnittää enemmän huomiota

työkokemuksen vaikutusten tarkempaan selvittämiseen: työkokemuksen (tai iän) kasvaessa tulojen lisäys ei ilmeisestikään ole suoraviivaista vaan tulot kääntyvät luultavasti tietyssä vaiheessa laskuun. Näin ollen mallien selittäjänä olisi syytä elinkaarihypoteesin puitteissa kokeilla työkokemuksen T lisäksi myös muuttujaa $(T - \bar{T})^2$. Jatkotutkimuksissa olisi toivottavaa saada riittävän yksityiskohtaisia tietoja joko kotitalous- tai yksilötasolla, jotta tästä tutkimuksesta poisjääneiden tekijöiden mahdollisten vaikutusten suuruudet voitaisiin kvantifioida.

Olemme joka tapauksessa voineet osoittaa, että keskimääräistä enemmän koulutettujen päämiesten kotitalouksissa käytettävissä oleva tulo on selvästi keskimääräistä korkeampi. Käytettävissä oleva tulo lisääntyy keskimäärin noin 9 % koulutusajan lisääntyessä vuodella. Tämä tulos koskee tietyn vuoden poikkileikkausaineistoa, siis kotitalouksien välisiä eroja, eikä siitä voida päätellä millainen vaikutus koulutuksen yleisellä lisääntymisellä on ollut tai olisi tulonmuodostukseen. Ei myöskään ole ilmeistä, missä määrin koulutuksen edulliset tulovaikutukset johtuvat lisääntyneestä ammattipätevyydestä tai tehokkuudesta ja missä määrin siitä, että koulutuksen laatua ja määrää käytetään valintaperusteena työnhakijoiden pyrkiessä paremmin palkattuihin tehtäviin. Näin ollen koulutuksen yhteiskunnallisia vaikutuksia selvitettäessä oleellisesti laajalaisempi tutkimusote on tarpeen.

LIITE 1

TILASTOKESKUKSEN RAKENNETIEDOSTON TIETUEKUVAUS

RUOKAKUNNAN NUMERO

kotitaloustiedustelun kuntakoodi
osite

ruokakunnan juokseva numero

HAASTATTELUN ONNISTUMINEN

hyväksytty kuukausihaastattelu

hyväksytty vuosihaastattelu

kk-alkuhaastattelu

säästämistietojen luotettavuus

kh-poisputoama

vh-poisputoama

ALUEET

suuralue

kuntatyyppi

läänikoodi

TILINPITOKUUKAUSI

TIETOJA RUOKAKUNNAN HENKILÖISTÄ

jäsenluku

lasten luku

ammattissa toimivien luku

yli 64-vuotiaiden luku

PÄÄMIEHISTÄ KOSKEVIA TIETOJA

päämiehen sukupuoli

päämiehen syntymävuosi

päämiehen avioliiton solmimisvuosi

päämiehen elinkeino

päämiehen työnantaja

päämiehen koulutus

RUOKAKUNNAN SOSIOEKONOMINEN ASEMA

TILAN KOKONAISPINTA-ALA

TILAN PELTOPINTA-ALA

KULUTUSYKSIKÖT

kulutussyksiköt yhteensä

kulutussyksiköt: vieraiden

kulutussyksiköt: ruokakunnan jäsenten

kotitalouden tv:n kulutussyksiköt

yrittäjätoiminnan kulutussyksiköt

ASUINYMPÄRISTÖ

TYHJÄÄ

KUUKAUSIOTOSKERROIN

VUOSIOTOSKERROIN

Q-KERROIN

KH-KULUTUS

VH-KULUTUS 27

VH-KULUTUS 28

RUOKAKUNNAN KÄYTETTÄVISSÄ OLEVA TULO

RUOKAKUNNAN KÄYTETTÄVISSÄ OLEVA TULO HENKILÖÄ KOHDEN

DESIILILUOKAT

RUOKAKUNTATYYPPI

LIITE 2

RAKENNETIEDOSTON LUOKITTELUJA NUMEROKOODEINEEN

KOTITALOUDEN ASUINALUE

- | | |
|-------------------|---------------------------------|
| - Suuralue | - Kuntatyyppi |
| 1 = Etelä-Suomi | 0 = Helsinki, Turku ja Tampere |
| 2 = Väli-Suomi | 1 = muut kaupungit ja kauppalat |
| 3 = Pohjois-Suomi | 2 = maalaiskunnat |

PÄÄMIEHEN TYÖNANTAJA

- 0 = ei palkansaaja
- 1 = valtio
- 2 = kunta
- 3 = seurakunta
- 4 = säätiö
- 5 = muu työnantaja
- 9 = työtön

PÄÄMIEHEN ELINKEINO

- 0 = maatalous, metsätalous, kalatalous ja metsästys
- 1 = kaivokset, muu kaivannaistoiminta
- 2 = teollisuus
- 3 = "
- 4 = sähkö-, kaasu- ja vesihuolto
- 5 = rakennustoiminta
- 6 = tukku- ja vähittäiskauppa
- 7 = kuljetus ja tietoliikenne
- 8 = rahoitus, vakuutus- ja kiinteistöpalvelukset sekä liike-elämää palveleva toiminta
- 9 = yhteiskunnalliset ja henkilökohtaiset palvelukset

PÄÄMIEHEN SUKUPUOLI

- 1 = mies
- 2 = nainen

KOTITALOUDEN SOSIOEKONOMINEN ASEMA

- 1 = maa- ja metsätalouden harjoittajat
- 2 = muut yrittäjät
- 3 = johtajat ja toimihenkilöt
- 4 = työntekijät
- 5 = ammatissa toimimattomat, jossa kotitalouden pääasiallinen toimeentulolähde tarkennettu
 - 54 = työ- tai virkaeläke
 - 55 = kansaneläke, työkyvyttömyyseläke, perhe-eläke

KOTITALOUDEN ASUINYMPÄRISTÖ

- 1 = kerrostaloalue
- 2 = omakoti- tai muu pientaloalue
- 3 = muu, esim. teollisuusalue
- 4 = kirkonkylä
- 5 = "kylämäinen asutus"
- 6 = maja-asutus

LIITE 3

INDIKAATTORIMUUTTUJUIEN MUODOSTAMINEN

- merkitään x-kirjaimella rakennetiedostossa esiintyvää koodia
(ks. liite 2)

LUOKITTELU: PÄÄMIEHEN SUKUPUOLI

$d_1 = \begin{cases} 1, & \text{jos } x = 2 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "nainen"

LUOKITTELU: PÄÄMIEHEN ELINKEINO

$d_2 = \begin{cases} 1, & \text{jos } x = 0 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "luontaistalous"

$d_3 = \begin{cases} 1, & \text{jos } x = 1 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "kaivannaistoiminta"

$d_4 = \begin{cases} 1, & \text{jos } x = 4 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "energia- ja vesihuolto"

$d_5 = \begin{cases} 1, & \text{jos } x = 5 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "rakennustoiminta"

$d_6 = \begin{cases} 1, & \text{jos } x = 6 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "kauppa"

$d_7 = \begin{cases} 1, & \text{jos } x = 7 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "kuljetus ja tietoliikenne"

$d_8 = \begin{cases} 1, & \text{jos } x = 8 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "yksityiset palvelukset"

$d_9 = \begin{cases} 1, & \text{jos } x = 9 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "julkiset palvelukset"

AMMATISSA TOIMIVIEN LUKUMÄÄRÄ

$d_{10} = \begin{cases} 1, & \text{jos } x \geq 2 \\ 0, & \text{jos } x < 2 \end{cases}$ "ammattiaktiivi"

$d_{11} = \begin{cases} 1, & \text{jos } x = 0 \\ 0, & \text{jos } x \neq 0 \end{cases}$ "ammatissa toimimaton"

LUOKITTELU: KOTITALouden SOSIOEKONOMINEN ASEMA

$d_{12} = \begin{cases} 1, & \text{jos } x = 54 \text{ tai } x = 55 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "eläkeläinen"

LUOKITTELU: PÄÄMIEHEN TYÖNANTAJA

$d_{13} = \begin{cases} 1, & \text{jos } x = 9 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases}$ "työtön"

LIITE 3

LUOKITTELU: PÄÄMIEHEN TYÖNANTAJA

$$d_{14} = \begin{cases} 1, & \text{jos } x = 1 \text{ tai } x = 2 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"julkisen sektorin palveluksessa"}$$

LUOKITTELU: KOTITALOUDEN SOSIOEKONOMINEN ASEMA

$$d_{15} = \begin{cases} 1, & \text{jos } x = 1 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"maa- ja metsätalous"}$$

$$d_{16} = \begin{cases} 1, & \text{jos } x = 2 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"muu yrittäjä"}$$

$$d_{17} = \begin{cases} 1, & \text{jos } x = 3 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"johtaja tai toimihenkilö"}$$

$$d_{18} = \begin{cases} 1, & \text{jos } x = 5 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"ammattissa toimimaton"}$$

TILAN KOKONAISPINTA-ALA

$$d_{19} = \begin{cases} 1, & \text{jos } x \geq 100.0 \\ 0, & \text{jos } x < 100.0 \end{cases} \quad \text{"suurtilallinen"}$$

LUOKITTELU: KOTITALOUDEN ASUINALUE

$$d_{20} = \begin{cases} 1, & \text{jos } x = 2 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"Väli-Suomi"}$$

$$d_{21} = \begin{cases} 1, & \text{jos } x = 3 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"Pohjois-Suomi"}$$

LUOKITTELU: KOTITALOUDEN KUNTATYYPPI

$$d_{22} = \begin{cases} 1, & \text{jos } x = 2 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"maalaiskunta"}$$

LUOKITTELU: KOTITALOUDEN ASUINYMPÄRISTÖ

$$d_{23} = \begin{cases} 1, & \text{jos } x = 2 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"pientaloasutus"}$$

$$d_{24} = \begin{cases} 1, & \text{jos } x = 5 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"kylämäinen asutus"}$$

$$d_{25} = \begin{cases} 1, & \text{jos } x = 6 \\ 0, & \text{muilla } x:n \text{ arvoilla} \end{cases} \quad \text{"maja-asutus"}$$

LIITE 4

MUUTTUJIEN KESKIARVOT JA HAJONNAT

Muuttujan tunnus		Keskiarvo	Vapausaste- korjattu hajonta
logY _d	"tulot"	9.69659	0.60042
K	"koulutus"	10.94250	1.47703
T	"potentiaalinen työkokemus"	24.36250	12.32660
N	"perheen jäsenluku"	3.39000	1.69433
log(n)	"perheen painotettu jäsenluku"	0.92429	0.46695
d ₁	"nainen"	0.21800	0.41309
d ₂	"luontaistalous"	0.20300	0.40243
d ₃	"kaivannaistoiminta"	0.00400	0.06315
d ₄	"energia- ja vesihuolto"	0.01400	0.11755
d ₅	"rakennustoiminta"	0.10900	0.31179
d ₆	"kauppa"	0.09400	0.29197
d ₇	"kuljetus ja tietoliikenne"	0.08300	0.27602
d ₈	"yksityiset palvelukset"	0.03000	0.17067
d ₉	"julkiset palvelukset"	0.22300	0.41647
d ₁₀	"ammattiaktiivi"	0.57200	0.49504
d ₁₁	"ammattissa toimimaton"	0.05900	0.23574
d ₁₂	"eläkeläinen"	0.06500	0.24665
d ₁₃	"työtön"	0.00100	0.03162
d ₁₄	"julkisen sektorin palveluksessa"	0.19200	0.39407
d ₁₅	"maa- ja metsätalous"	0.18300	0.38686
d ₁₆	"muu yrittäjä"	0.04400	0.20520
d ₁₇	"johtaja tai toimihenkilö"	0.27900	0.44873
d ₁₈	"ammattissa toimimaton"	0.08000	0.27143
d ₁₉	"suurtilallinen"	0.01300	0.11333
d ₂₀	"Väli-Suomi"	0.28500	0.45164
d ₂₁	"Pohjois-Suomi"	0.12800	0.33425
d ₂₂	"maalaiskunta"	0.46700	0.49916
d ₂₃	"pientaaloasutus"	0.19700	0.39793
d ₂₄	"kylämäinen asutus"	0.13000	0.33647
d ₂₅	"maja-asutus"	0.17200	0.37757

LIITE 5

TIETOKONETUNNUKSEN JA TUTKIMUKSESSA KÄYTETYN MUUTTUJATUNNUKSEN VASTAAVUUS

Kvantitatiiviset muuttajat:

Tietokonetunnus	Muuttuja
3	N
14	$\ln Y_d$
15	K
127	T
128	$\ln(n)$

Kvalitatiiviset muuttajat:

Tietokonetunnus	Muuttuja
101	d ₂₀
102	d ₂₁
103	d ₁₀
104	d ₁₁
105	d ₁₃
106	d ₁₄
107	d ₂
108	d ₃
109	d ₄
110	d ₅
111	d ₆
112	d ₇
113	d ₈
114	d ₉
115	d ₁₅
116	d ₁₆
117	d ₁₇
118	d ₁₈
119	d ₁₂
120	d ₁
121	d ₁₉
122	e ₁ käytössä
123	d ₂₃
124	d ₂₄
125	d ₂₅
126	d ₂₂

LIITE 6: MUUTTUJIEN KOVARIANSSIMATRIISI¹⁾

Liitteessä 5 on esitetty tietokonetunnuksen ja tutkimuksessa käytetyn muuttujatunnuksen vastaavuus.

VAR	14	15	101	102	103	104	105	106	107	108
	109	110	111	112	113	114	115	116	117	118
	119	120	121	122	123	124	125	126	127	128
14	.35997 .00368 -.03458	.27025 .00875 -.08814	-.02278 -.00365 -.00234	-.00034 .00526 .06594	.13574 .00275 .00317	-.04480 -.03584 -.00032	-.00001 -.00700 -.00599	.03059 .00554 -.00045	-.00738 .05668 -.37216	-.00092 -.05055 .09750
15	.27025 -.00319 -.01026	2.17947 -.02373 -.04196	-.07211 -.01909 -.00575	-.00414 -.01423 -.00814	-.03111 .03223 -.00617	-.01061 .13482 -.05102	-.00044 -.09448 -.06511	.12804 -.00997 -.12615	-.10183 .31554 -2.29034	-.00177 -.01690 -.00079
101	-.02278 -.00299 .00747	-.07211 -.00093 -.01113	.20378 -.00579 .00429	-.03648 -.00166 .24111	.00298 -.00355 -.00715	-.00518 -.00256 .01595	-.00029 .02385 .01998	-.00772 .00146 .03590	.02514 -.02652 .45619	-.00014 -.00020 .02772
102	-.00034 .00421 -.00232	-.00414 .00305 -.00390	-.03648 -.00003 -.00066	.11162 .00338 .03189	-.00022 -.00284 .00178	-.00055 -.00054 .00536	-.00013 .00458 .00298	.01142 -.00163 .01622	.00702 .00129 -.16790	-.00051 -.00024 .03166
103	.13574 -.00001 -.03318	-.03111 .00665 -.07970	.00298 -.01277 .00156	-.00022 .00252 .07431	.24482 -.00616 -.00268	-.03375 .05756 .00864	-.00057 .05232 .01762	-.00082 .00183 .02588	.05588 -.02859 .14565	-.00071 -.04176 .05622
104	-.04480 -.00083 .04517	-.01061 -.00643 .01514	.00518 -.00555 .00123	-.00055 -.00390 .31171	-.03375 -.00177 .00338	.05552 .04484 .00133	-.00006 -.01080 .00385	-.01133 -.00260 .00245	-.01196 -.01646 .68061	-.00024 .05378 -.01337
105	.00001 -.00001 -.00007	-.00044 -.00011 -.00022	-.00029 -.00009 -.00001	-.00013 -.00008 -.00295	-.00057 -.00003 -.00020	-.00006 .00078 -.00013	.00100 -.00018 .00083	-.00019 -.00004 .00053	-.00020 -.00028 .00814	-.00000 .00092 .00239
106	.03059 .00331 -.01148	.12804 .00807 .00414	-.00772 -.01505 -.00050	.01142 .02406 -.07717	-.00082 -.00176 -.00182	-.01133 .05618 -.00896	-.00019 -.03514 -.01402	.15514 -.00745 -.01766	-.03598 .04343 -.40160	-.00077 -.01436 .00276
107	-.00738 -.00284 -.01320	-.10183 -.02213 -.02525	.02514 -.01908 -.00064	.00702 -.01685 .44134	.05588 -.00609 -.03399	-.01196 -.04527 .02061	-.00020 .14585 .06708	-.03598 -.00793 .08920	.16179 -.05364 1.03791	-.00081 -.01624 .01508
108	-.00092 -.00006 -.00026	-.00177 -.00044 .00013	-.00014 -.00038 -.00005	-.00051 -.00033 -.01182	.00071 -.00012 .00021	-.00024 -.00089 -.00052	-.00000 -.00073 .00031	-.00077 -.00018 -.00187	-.00081 .00088 .00055	.00398 -.00032 -.00143

1) Muuttujaa 122 ei ole käytetty tutkimuksessa, muuttuja 3 = N puuttuu korrelaatiomatriisista.

LIITE 6: jatkuu

VAR	14 109 119	15 110 120	101 111 121	102 112 122	103 113 123	104 114 124	105 115 125	106 116 126	107 117 127	108 118 128
109	.00368 .01380 -.00091	-.00319 -.00153 -.00305	-.00299 -.00132 -.00018	.00421 -.00116 .06764	-.00001 -.00042 .00124	-.00083 -.00312 -.00182	-.00001 -.00256 .00059	.00331 -.00062 .00046	-.00284 -.00191 -.03808	-.00006 -.00112 .00299
110	.00875 -.00153 -.00708	-.02373 .09712 -.02076	.00093 -.01025 .00258	.00305 -.00905 -.00299	.00665 -.00327 .01353	-.00643 -.02431 .00083	-.00011 -.01995 -.00475	.00807 -.00480 -.00690	-.02213 -.01141 -.27551	-.00044 -.00872 .02023
111	-.00365 -.00132 -.00611	-.01909 -.01025 .02651	-.00579 .08516 -.00022	-.00003 -.00780 -.10268	-.01277 -.00282 .00248	-.00555 -.02096 -.00722	-.00009 -.01720 -.01217	-.01505 .00586 -.00290	-.01988 .03077 -.35057	-.00038 -.00752 -.00715
112	.00526 -.00116 -.00539	-.01423 -.00905 -.01109	-.00166 -.00780 -.00008	.00338 .07611 -.11118	.00252 -.00249 -.00035	-.00390 -.01851 .00221	-.00008 -.01519 -.00928	.02406 .01135 -.00076	-.01685 -.00616 -.23909	-.00033 -.00664 .00104
113	.00275 -.00042 -.00195	.03223 -.00327 .00446	-.00355 -.00282 -.00039	-.00284 -.00249 -.00362	-.00616 .02910 -.00491	-.00177 -.00669 -.00190	-.00003 -.00549 -.00516	-.00176 -.00032 -.00701	-.00609 .01063 -.17338	-.00012 -.00240 -.00524
114	-.03584 -.00312 .05050	.13482 -.02431 .03639	-.00256 -.02096 .00110	-.00054 -.01851 .11626	-.05756 -.00669 -.00293	.04484 .17327 -.00799	.00078 -.04081 -.00836	.05618 -.00181 -.01814	-.04527 .02578 .67216	-.00089 .06216 -.01508
115	-.00700 -.00256 -.01189	-.09448 -.01995 -.02189	.02385 -.01720 -.00138	.00458 -.01519 .43642	.05232 -.00549 -.03205	-.01080 -.04081 .02121	-.00018 .14951 .06152	-.03514 -.00805 .08254	.14585 -.05106 1.02166	-.00073 -.01464 .01224
116	.00554 -.00062 -.00286	-.00997 -.00480 -.00259	.00146 .00586 .00043	-.00163 .01135 -.10898	.00183 -.00037 -.00067	-.00260 -.00181 .00628	-.00004 -.00805 -.00357	-.00745 .04206 .00545	-.00793 -.01228 .05555	-.00018 -.00352 .00925
117	.05668 -.00191 -.01814	.31354 -.01141 .02918	-.02652 .03077 -.00163	.00129 -.00616 -.39217	-.02859 .01063 -.00296	-.01646 .02578 -.01927	-.00028 -.05106 -.03799	.04343 -.01228 -.05429	-.05364 .20116 -1.27064	.00086 -.02232 -.01213
118	-.05055 -.00112 .05980	-.01690 -.00872 .01456	.00620 -.00752 .00196	-.00174 -.00664 .28868	-.04176 -.00240 .00324	.05328 .06216 .00060	.00092 -.01464 .00824	-.01436 -.00352 .00464	-.01624 -.02232 .97550	-.00032 .07360 -.01714

LIITE 6: jatkuu

VAR	14 109 119	15 110 120	101 111 121	102 112 122	103 113 123	104 114 124	105 115 125	106 116 126	107 117 127	108 118 128
119	-.03458 -.00091 .06077	-.01026 -.00708 .01083	.00747 -.00611 .00215	-.00232 -.00539 .33299	-.03318 -.00195 .00319	.04517 .05050 .00055	-.00007 -.01189 .00582	-.01148 -.00286 .00164	-.01320 -.01814 .97744	-.00026 .05949 -.02102
120	-.08814 -.00305 .01083	-.04196 -.02076 .17048	-.01113 .02651 .00017	-.00390 -.01169 -.15297	-.07970 .00446 -.00195	.01514 .03639 -.00854	-.00022 -.02189 -.02050	.00414 -.00259 -.03681	-.02525 .02018 .20848	.00013 .01456 -.06054
121	-.00234 -.00018 .00215	-.00575 .00258 .00017	.00429 -.00022 .01283	-.00066 -.00008 .00960	.00156 -.00039 -.00256	.00123 .00110 .00331	-.00001 -.00158 .00376	-.00050 .00043 .00393	-.00064 -.00163 .02879	-.00005 .00196 .00310
122	.06594 .06764 .33299	-.00814 -.00299 -.15297	.24111 -.10268 .00960	.03189 -.11118 196.29590	.07451 -.00362 -.31994	.31171 .11626 .07198	-.00295 .43642 .28091	-.07717 -.10898 .65648	.44134 -.39217 11.20517	-.01182 .28868 -.14853
123	.00317 .00124 .00319	-.00617 .01353 -.00195	-.00715 .00248 -.00256	.00178 -.00035 -.31994	-.00268 -.00491 .15819	.00338 -.00293 -.02561	-.00020 -.03205 -.03388	-.00182 -.00067 .08200	-.03399 -.00296 -.08891	.00021 .00324 -.00478
124	-.00032 -.00182 .00055	-.05102 .00083 -.00034	.01595 -.00722 .00331	-.00536 .00221 .07198	.00864 -.00190 -.02561	.00133 -.00799 .11310	-.00013 .02121 -.02236	-.00896 .00628 .06629	.02061 -.01927 .43338	-.00032 .00060 .01496
125	.00599 .00059 .00582	-.06511 -.03475 -.02050	.01998 -.01217 .00376	.00298 -.00928 .28091	.01762 -.00516 -.03388	.00385 -.00836 -.02236	.00083 .06152 .14242	-.01402 -.00357 .08668	.06708 -.03799 .84565	.00031 .00824 .02692
126	-.00045 .00046 .00164	-.12615 -.00690 -.03681	.03490 -.02290 .00393	.01622 -.00076 .65648	.02588 -.00701 -.08200	.00245 -.01814 .06629	.00053 .08254 .08668	-.01766 .00545 .24891	.03920 -.05429 .96471	-.00187 .00464 .04106
127	-.37216 -.03808 .97744	-2.29034 -.27551 .20848	.45619 -.35057 .02879	-.16790 -.23909 11.20517	-.14565 -.17338 -.08891	.68061 .67216 .43338	.00814 1.02166 .84565	-.40160 .05555 .96471	1.03791 -1.27064 151.79333	.00055 .97550 -1.14765
128	.09750 .00299 -.02102	-.00879 .02023 -.06054	.02772 -.00715 .00310	.03168 .00104 -.14853	.05022 -.00524 -.00478	-.01837 -.01808 .01496	.00239 .01224 .02692	.00276 .00925 .04106	.01508 -.01213 -1.14765	-.00143 -.01714 .44284

LIITE 7: REGRESSIOMALLIEN F-ARVO, VIRHETERMIN HAJONNAN ESTIMAATTI (s), YHTEISKORRELAATIOKERTOIMEN NELIÖ (R^2), SELITTÄJIEN REGRESSIOKERTOIMET JA t-ARVOT

Malli	F-arvo	s	R^2	vakio- termi	K	T	N	ln(n)	d_1	d_2	d_3
(1)	102.3***	0.572	0.093	8.340	0.124*** (10.115)						
(2)	27.6***	0.528	0.235	8.467	0.127*** (10.537)	0.0020 (1.430)			-.466*** (10.569)	-.148** (-2.846)	-.266 (-.999)
(3)	49.9***	0.444	0.464	7.934	0.116*** (11.230)	0.0031** (2.415)	0.102*** (10.812)		-.154*** (-3.751)	-.301*** (-6.762)	-.264 (-1.180)
(4)	42.1***	0.435	0.487	8.122	0.091*** (8.199)	0.0036** (2.902)	0.107*** (11.454)		-.157*** (-3.891)	-.229* (-2.246)	-.309 (-1.406)
(5)	34.4***	0.433	0.495	8.133	0.090*** (8.082)	0.0035** (2.775)	0.112*** (11.739)		-.152*** (-3.769)	-.217* (-2.104)	-.289 (-1.317)
(6)	67.4***	0.432	0.489	8.110	0.090*** (8.439)	0.0034** (2.768)	0.113*** (11.960)		-.149*** (-3.863)		
(7)	77.1***	0.418	0.523	8.052	0.088*** (8.521)	0.0020* (1.684)		0.554*** (14.907)	-.036 (-.920)		

LIITE 7: jatkuu

d ₄	d ₅	d ₆	d ₇	d ₈	d ₉	d ₁₀	d ₁₁	d ₁₂	d ₁₃	d ₁₄	d ₁₅
.089 (.610)	-.083 (-1.343)	.019 (.293)	-.078 (-1.159)	-.070 (-.680)	-.274*** (-5.346)						
.055 (.444)	-.128* (-2.416)	.022 (.391)	-.107 (-1.768)	.037 (.423)	-.137* (-2.514)	.389*** (11.482)	-.490*** (-5.017)	.156 (1.586)			
.097 (.802)	-.092 (-1.751)	-.035 (-.631)	-.077 (-1.261)	.016 (.180)	-.084 (-1.483)	.390*** (11.655)	-.239* (-2.198)	.582*** (4.449)	.350 (.765)	.040 (.869)	-.045 (-.436)
.105 (.863)	-.086 (-1.652)	-.024 (-.420)	-.072 (-1.181)	.015 (.174)	-.079 (-1.401)	.382*** (11.393)	-.232* (-2.136)	.601*** (4.611)	.277 (.607)	.050 (1.073)	-.052 (-.502)
						.382*** (11.501)	-.245* (-2.290)	.588*** (4.623)			-.240*** (-5.615)
						.313*** (9.492)	-.159 (-1.527)	.551*** (4.488)			-.253*** (-6.140)

LIITE 7: jatkuu

d ₁₆	d ₁₇	d ₁₈	d ₁₉	d ₂₀	d ₂₁	d ₂₂	d ₂₃	d ₂₄	d ₂₅
.046 (.623)	.176*** (4.316)	-.613*** (-4.206)	-.227 (-1.851)						
.034 (.465)	.175*** (4.302)	-.628*** (-4.312)	-.212 (-1.709)	-.112*** (-3.445)	-.116*** (-2.635)	.057 (1.415)	.026 (.668)	.005 (.103)	-.011 (-.214)
	.176*** (4.680)	-.652*** (-4.953)	-.241* (-1.971)	-.116*** (-3.565)	-.117** (-2.699)	.043 (1.363)			
	.179*** (4.932)	-.665*** (-5.233)	-.241* (.2.043)	-.117*** (-3.745)	-.123** (-2.958)	.046 (1.508)			

LIITE 9: MALLIN (6) MUUTTUJIEN KORRELAATIOMATRIISI

VAR	3 120	15 121	101 126	102 127	103 14	104	115	117	118	119
3	1.00000 -.39761	-.01223 .05177	.10968 .21526	.14684 -.00577	.35913 .43319	-.18297	.15727	-.11561	-.13756	-.14935
15	-.01223 -.06885	1.00000 -.03440	-.10821 -.17127	-.00339 -.12592	-.04259 .30511	-.03049	-.16551	.47353	-.04220	-.02820
101	.10968 -.05972	-.10821 .08400	1.00000 -.15943	-.24189 .08202	.01334 -.08413	.04875	.13661	-.13096	.05063	.06717
102	.14684 -.02830	-.00839 -.01755	-.24189 .09734	1.00000 -.04079	-.00131 -.00167	-.00701	.03542	.00860	-.01368	-.02817
103	.35913 -.39011	-.04259 .02791	.01334 .10482	-.00131 .02389	1.00000 .45724	-.28947	.27349	-.12882	-.31110	-.27202
104	-.18297 .15560	-.03049 .04620	.04875 .02082	-.00701 .23445	-.28947 -.31693	1.00000	-.11851	-.15576	.83350	.77753
115	.18727 -.13714	-.16551 -.03148	.13661 .42786	.03542 .21446	.27349 -.03019	-.11851	1.00000	-.29441	-.13956	-.12479
117	-.11561 .10896	.47353 -.03202	-.13096 -.24263	.00860 -.22995	-.12882 .21063	-.15576	-.29441	1.00000	-.18344	-.16402
118	-.13756 .12998	-.04220 .06378	.05063 .03428	-.01368 .29185	-.31110 -.31054	.83350	-.13956	-.18344	1.00000	.89413
119	-.14935 .10640	-.02820 .07717	.06717 .01337	-.02817 .32181	-.27202 -.25380	.77753	-.12479	-.16402	.89413	1.00000
120	-.39761 1.00000	-.06885 .00355	-.05972 -.17868	-.02830 .04098	-.39011 -.35582	.15560	-.13714	.10896	.12998	.10640
121	.05177 .00355	-.03440 1.00000	.08400 .06952	-.01755 .02063	.02791 -.03441	.04620	-.03148	-.03202	.06378	.07717
126	.21526 -.17868	-.17127 .06952	.15943 1.00000	.09734 .15695	.10482 -.00150	.02082	.42786	-.24263	.03428	.01337
127	-.00577 .04098	-.12592 .02063	.08202 .15695	1.00000	-.05035	.23445	.21446	-.22995	.29185	.32181
14	.43319 -.35582	.30511 -.03441	-.08413 -.00150	-.00167 -.05035	.45724 1.00000	-.31693	-.03019	.21063	-.31054	-.23380

LIITE 10: MALLIN (7) MUUTTUJIEN KORRELAATIOMATRIISI

VAR	15 121	101 126	102 127	103 128	104 14	115	117	118	119	120
15	1.00000 -.03440	-.10821 -.17127	-.00839 -.12592	-.04259 -.00312	-.03049 .30511	-.16551	.47353	-.04220	-.02820	-.06885
101	-.10821 .05400	1.00000 .15943	-.24189 .08202	.01334 .10021	.04875 -.08413	.13661	-.13096	.05063	.06717	-.01972
102	-.00839 -.01755	-.24189 .09734	1.00000 -.04079	-.00131 .11787	-.00701 -.00167	.03542	.00860	-.01368	-.02817	-.02450
103	-.04259 .02791	.01334 .10482	-.00131 .07389	1.00000 .48225	-.28947 .45724	.27349	-.12882	-.31110	-.27202	-.39017
104	-.03049 .04620	.04875 .07082	-.00701 .23445	-.28947 -.22685	1.00000 -.31693	-.11851	-.15576	.83350	.77753	.15560
115	-.16551 -.03148	.13661 .42786	.03542 .21446	.27349 .23464	-.11851 -.03019	1.00000	-.29441	-.13456	-.12479	-.11896
117	.47353 -.03202	-.13096 -.24263	-.00860 -.22995	-.12882 -.13823	-.15576 .21063	-.29441	1.00000	-.18344	-.16402	.11896
118	-.04220 .06378	.05063 .03428	-.01368 .29185	-.31110 -.16868	.83350 -.31054	-.13956	-.18344	1.00000	.89413	.12479
119	-.02820 .07717	.06717 .01337	-.02817 .32181	-.27202 -.15895	.77753 -.23380	-.12479	-.16402	.89413	1.00000	.11896
120	-.06885 .00355	-.05972 -.17468	-.02830 .04098	-.39011 -.53232	.15560 -.35582	-.13714	.10896	.12098	.10640	1.00000
121	-.03440 1.00000	.08400 .06952	-.01755 .02063	.02791 .04154	.04620 -.03441	-.03148	-.03202	.06378	.07717	.00355
126	-.17127 .06952	.15943 1.00000	.09734 .15695	.10482 .21840	.02862 -.00150	.42786	-.24263	.03428	.01337	-.11896
127	-.12592 .02063	.08202 .15695	-.04079 1.00000	.02389 .05850	.23445 -.05035	.21446	-.22995	.29185	.32181	.01755
128	-.00312 .04154	.10021 .21840	.11787 .05850	.48225 1.00000	-.22685 .53218	.23464	-.13823	-.16868	-.15895	-.11896
14	.30511 -.03441	-.08413 -.00150	-.00167 -.05035	.45724 .53218	-.31693 1.00000	-.03019	.21063	-.31054	-.23380	-.35582

LÄHDELUETTELO

AHMAVAARA, Yrjö (1969): "Yhteiskuntatieteiden kyberneettinen metodologia. Positivismin kritiikki", Helsinki.

AITCHISON, J. and BROWN, J. (1957): "The Lognormal distribution", Cambridge University Press, Cambridge.

ATKINSON, A.B. (1975): "The economics of inequality", Oxford University Press, Oxford.

BECKER, G.S. (1964): "Human capital: A theoretical and empirical analysis with special reference to education", Nat. Bur. Econ. Res., New York.

BEHRMAN, J., TAUBMAN, P. and WALES, T. (1977): "Controlling for and measuring the effects of genetics and family environment in equations of schooling and labor market success", in *Kinometrics: determinants of socioeconomic success within and between families*, edited by Taubman, North-Holland Publishing Company, Amsterdam-New York-Oxford.

BERKSON, J. (1958): "Smoking and lung cancer: Some observations on two recent reports", *Journal of the American Statistical Association*, 53, 28-38.

BROWN, B.W. Jr. (1972): "Statistics, scientific method and smoking", in *Statistics: A Guide to the Unknown*, edited by Tanur, Mosteller, Kruskal, Link, Pieters, and Rising, Holden-Day, Inc., San Francisco.

CHISWICK, B.R. (1974): "Income inequality: Regional analyses within a human capital framework", National Bureau of Economic Research, New York.

COCHRAN, W.G. (1957): "Analysis of covariance: its nature and uses", *Biometrics*, 13, 261-268.

CONLISK, J. (1971): "A bit of evidence on the income-education-ability interrelation", *The Journal of Human Resources*, VI, 3, 358-362.

CRAMER, J.S. (1973): "Empirical econometrics", Elsevier-North Holland Publishing Company, Amsterdam.

DRAPER, N.R. and SMITH, H. (1966): "Applied regression analysis", Wiley-Interscience, New York.

DUNCAN, O.D. (1968): "Ability and achievement", *Eugenics Quarterly*, 15, 1-11.

FISHER, R.A. (1959): "Smoking: The cancer controversy", Oliver & Boyd, Edinburgh.

GALBRAITH, J.K. (1974): "Valta ja raha", KY, Helsinki 1974.

GOLDBERGER, A.S. (1964): "Econometric theory", Wiley, New York.

GRIFFIN, L. (1976): "Specification biases in estimates of socioeconomic returns to schooling", *Sociology of Education*, 49, April 121-139.

- GRILICHES, Z. and MASON, W. (1972): "Education, income and ability", *Journal of Political Economy*, 80, Nr 3, Part II, 74-79.
- JOHNSTON, J. (1972): "Econometric methods", 2nd ed. McGraw-Hill Book Company, New York.
- KETONEN, Oiva (1976): "Se pyörii sittenkin. Tieteenfilosofisia peruskysymyksiä", WSOY, Porvoo ja Helsinki.
- KOYMAN, M.A. (1976): "Dummy variables in econometrics", Tilburg University Press, Rotterdam.
- LEIBOWITZ, A. (1974): "Home investments in children", *Journal of Polit. Econ.*, 82, 2, PH, II, S111-S131.
- LILJA, R. (1978): "Koulutus kotitalouksien tulojen selittäjänä", Helsingin kaupakorkeakoulun pro gradu-tutkielma, Helsinki.
- LINDERT, P. (1976): "Sibling position and achievement", University of Wisconsin.
- LIPSEY, R.G. (1966): "An introduction to positive economics", Weidenfeld & Nicolson, London.
- LYDALL, H. (1968): "The structure of earnings", The Clarendon Press, Oxford.
- MASLOW, Abraham M. (1954): "Motivation and personality", New York.
- MERRET, S. (1971). "The rate of return to education: A critique", in *Education and the Economics of Human Capital*, edited by Wykstra, New York.
- MINCER, J. (1960): "Labour supply, family income, and consumption", *American Economic Review*, Papers and Proceedings, 50, 574-583.
- MINCER, J. (1974): "Schooling, experience and earnings", National Bureau of Economic Research, New York.
- MORGENSTERN, R.D. (1973): "Direct and indirect effects on earnings of schooling and socio-economic background", *Rev. Econ. Stat.*, LV, 2, 225-233.
- OLNECK, M.R. (1977): "On the use of sibling data to estimate the effects of family background, cognitive skills, and schooling: results from the Lalamazoo brothers study", in *Kinometrics: determinants of socioeconomic success within and between families*, edited by Taubman, North-Holland Publishing Company, Amsterdam·New York·Oxford.
- ORD, J.K., PATIL, G.P. and TAILLIE, G. (1979): "The choice of a distribution to describe personal incomes", University of Warwick (mimeo).
- PARETO, V. (1971): "Manual of political economy", Macmillan, London.
- PEN, J. (1971): "Income distribution", Praeger, New York.
- SAHOTA, G.S. (1978): "Theories of personal income distribution: A survey", *Journal of Economic Literature*, Vol. 16, 1-55.

- SALEM, A.Z.B. and MOUNT, T.D. (1974): "A Convenient descriptive model of income distribution: The Gamma density", *Econometrica*, 42, 1115-1127.
- SEWELL, W.H. and HAUSER, R.M. (1975): "Education, occupation and earnings: achievement in the early career", Academic Press, New York.
- SINGH, S.K. and MADDALA, G.S. (1976): "A function for size distribution of incomes", *Econometrica*, 46, 963-970.
- SOLMON, L.C. (1975): "The definition of college quality and its impact on earnings", *Explorations in Economic Research*, 2, No. 4, 537-587.
- TALOUSNEUVOSTO (1972): "Elämisen laatu: tavoitteet ja mittaaminen", Helsinki (Liite 3 sisältää koulutuspolitiikan tavoitteita tutkineen työryhmän raportin).
- TAUBMAN, P. (Editor) (1977): "Kinometrics: determinants of socioeconomic success within and between families", North-Holland Publishing Company, Amsterdam-New York-Oxford.
- TILASTOKESKUS (1971): "Käsikirjoja nro I, Koulutusluokittelu", Valtion painatuskeskus, Helsinki.
- TILASTOKESKUS (1974): "Väestölaskenta 1970, osa VII B", Valtion painatuskeskus, Helsinki.
- TILASTOKESKUS (1977): "Tilastollisia tiedonantoja nro 55, Kotitaloustiedustelu 1971, osat I ja II", Valtion painatuskeskus, Helsinki.
- TILASTOLLINEN PÄÄTOIMISTO (1970): "Systemaattinen koulutusalaaluokitus ja aakkosellinen koulutusalahakemisto", Helsinki.
- TINBERGEN, J. (1975): "Income distribution", Elsevier-North Holland Publishing Company, Amsterdam.
- TÖRNQVIST, L. ja NORDBERG, L. (1968): "Päättätieteeseen keskeisiä onglemia", WSOY, Porvoo ja Helsinki.
- WACHTEL, P. (1975): "The effect of school quality on achievement attainment levels, and lifetime earnings", *Explorations In Economic Research*, 2, No. 4, 502-536.
- VASAMA, P.-M. ja VARTIA, Y. (1970): "Johdatus tilastotieteeseen, osa I", Gaudeamus, Helsinki.
- VASAMA, P.-M. ja VARTIA, Y. (1972): "Johdatus tilastotieteeseen, osa II", Gaudeamus, Helsinki.
- VARTIA, P. and VARTIA, Y. (1973): "Comments on the discussion held at the conference on income distribution", *Taloustieteellisen Seuran vuosikirja 1972*, Helsinki 1973.
- VARTIA, Y. (1976): "Relative changes and index numbers", The Research Institute of the Finnish Economy, A4, Helsinki.

VARTIA, P. and VARTIA, Y. (1978): "Description of the income distribution by the scaled F distribution model", ETLAn keskusteluaiheita, 18.

VEPSÄ, Kirsti (1973): "Elintason muutos Suomessa v. 1910-1965", WSOY, Porvoo ja Helsinki.

WIENER, N. (1969): "Ihmisistä, koneista ja kielestä", WSOY, Porvoo ja Helsinki.

WOLD, H. (1954): "Causality and econometrics", *Econometrica* 22, 162-177.

WOLD, H. (1956): "Causal inference from observational data. A review of ends and means", *Journal of the Royal Statistical Society (A)*, 119, 28-61.

WOLD, H. (1963): "Huvudlinjer och grundbegrepp i statistiken", Biblioteks-förlaget, Stockholm.

WOLD, H. (1967): "Forecasting and scientific method", in *Forecasting on a scientific basis*, Proceedings of an international summer institute held in Curia, Portugal, September 1966, Lisboa.

WONNACOTT, R. and WONNACOTT, I. (1970): "Econometrics", Wiley, New York.

von WRIGHT, G.H. (1968): "Logikka, filosofia ja kieli", Otava, Helsinki.

YOUNG, M. (1967): "Meritokratian nousu 1870-2033, tutkielma koulutuksesta ja tasa-arvosta", Weilin & Göös, Helsinki.

ELINKEINOELÄMÄN TUTKIMUSLAITOS (ETLA)

JULKAISUJA — PUBLICATIONS

THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

LÖNNROTINKATU 4 B, 00120 HELSINKI 12

ETLAn edeltäjän, Taloudellisen Tutkimuskeskuksen julkaisuluettelo on saatavissa Elinkeinoelämän Tutkimuslaitoksen toimistosta.

A list of the publications of ETLA's predecessor, the Economic Research Institute of Finnish Industry is available from the Office of ETLA.

SUHDANNE — ECONOMIC PROSPECTS

Suhdanne-ennustetta on julkaistu syksystä 1971 lähtien. Suhdanne ilmestyy keväällä ja syksyllä. Economic Prospects has been published since 1971 twice a year in spring and in autumn.

No 17 SUHDANNE syksy 1979. Economic prospects with English summary. Helsinki 1979. 127 s. nid.

No 18 SUHDANNE kevät 1980. Economic prospects with English summary. Helsinki 1980. 128 s. nid.

KANSANTALouden KEHITYSNÄKYMÄT — MEDIUM-TERM ECONOMIC PROSPECTS

No 1 Kansantalouden kehitysnäkymät 1978—1982, Medium-term economic prospects with English summary. Helsinki 1978. 70 s. nid.

Myynti — For sale at

Akateeminen Kirjakauppa, The Academic Bookstore SF-00100 Helsinki 10.

Suomalainen Kirjakauppa, The Finnish Bookstore SF-00100 Helsinki 10.

SARJA A SERIES

A 1 YRJÖ O VARTIA, Betajakauman parametrien estimointi maximum likelihood-menettämällä. (Estimation of the parameters of the beta distribution by maximum likelihood method). 120 s. nid.

PENTTI L. I. VARTIA, Huomioita teoreettisten apukäsitteiden identifioituvuudesta. (Comments of the identification of auxiliary concepts.) 47 s. nid. Helsinki 1973.

A 2 PENTTI L. I. VARTIA, An Econometric Model for Analyzing and Forecasting Short-term Fluctuations in the Finnish Economy. Helsinki 1974. 249 s. nid.

A 3 ANTTI TANSKANEN, Ulkomaankaupan tasapaino, taloudellinen kasvu ja Suomen velkaantuminen. (Balance in Foreign Trade, Economic Growth and Finland's Indebtedness). Helsinki 1976. 115 s. nid.

A 4 YRJÖ O. VARTIA, Relative Changes and Index Numbers. Helsinki 1976. 230 s. nid.

A 5 HEIKKI HÄMÄLÄINEN, Taloudellisten vaihteluiden seuranta- ja ennakointijärjestelmän laadinta ja valvonta. English summary: Construction and supervision of a system for analysing and forecasting of short-term economic fluctuations. Helsinki 1977. 146 s. nid.

SARJA B SERIES

B 1 KARI SIHTOLA, Teollisuuden välilliset työvoimakustannukset v. 1970, English summary: Indirect labor costs in Finnish manufacturing in 1970. Helsinki 1972. 133 s. nid.

B 2 ROLF MAURY — JUKKA RINNE — REIJO VUORIKALLIO, Kuljetuskustannusten alueelliset erot ja niiden tasoittaminen. English summary: The regional variations of transport costs. Helsinki 1972. 127 s. nid.

B 3 ILPO SEPPÄ — SEPPO WALLENIUS, Tekstiili- ja vaatetustavaroiden kulutus Suomessa vuosina 1954—1977. Helsinki 1972. 95 s. nid.

Separate English summary: The consumption of clothing and household textiles in Finland during 1954—1977. Helsinki 1973. 15 p.

- B 4** KARI HAGFORS, Teollisuuden vapaaehtoisten työvoimakustannusten alueittainen jakaantuminen vuosina 1965 ja 1970. English summary: A study of voluntary personnel expenditure by manufacturing firms in different regions of Finland in 1965 and 1970. Helsinki 1973. 176 s. nid.
- B 5** HEIKKI HÄMÄLÄINEN, Yksityisten kulutusmenojen rakenne ja kehitys Suomessa vuosina 1965—1975. English summary: Private consumption expenditure patterns and trends in Finland 1965—1975. Helsinki 1973. 159 s. nid.
- B 6** PIRKKO HAAVISTO — KARI HAGFORS, Teollisuuden kannattavuus 1960-luvulla. English summary: The development of the profitability of the Finnish industry in the 1960. Helsinki 1973. 101 s. nid.
- B 7** TIMO SUMMA — ANTTI SÄÄSKILAHTI, Suomen koneteollisuus vuosina 1954—1980. English summary: The manufacture of machinery in Finland 1954—1980. Helsinki 1974. 210 s. nid.
- B 8** CHRISTIAN EDGREN, Suomalaisen siirtolaistyövoiman rakenne ja taloudellinen asema Ruotsissa. (The structure and the economic condition of Finnish workers in Sweden). Helsinki 1974. 178 s. nid.
- B 9** RONALD WIMAN, Työvoiman kansainvälisen muuttoliikkeen mekanismi. Tutkimus Suomen ja Ruotsin välisen muuttoliikkeen syistä. English summary: The mechanism on international labour migration. A study of the causes of Finnish emigration to Sweden. Helsinki 1975. 216 s. nid.
- B 10** ILPO SEPPÄ — SEPPO WALLENIOUS — ARJA VÄISÄNEN, Tekstiili- ja vaate-tavaroiden kulutus Suomessa vuosina 1970—1980. (The consumption of textiles and clothing in Finland in 1970—1980). Helsinki 1975. 54 s. + liitt. nid.
- B 11** KARI SIHTOLA, Teollisuuden välilliset työvoimakustannukset vuonna 1973 sekä arvioita niiden kehityksestä vuosina 1974—1976. English summary: Indirect labour costs in Finnish manufacturing in 1973 and prospects for development of labour costs in 1974—1976. Helsinki 1975. 100 s. nid.
- B 12** TAUNO KALLINEN, Teollisuuden rahoitusrakenne toimialoittain vuosina 1960—1973. English summary: Structure of financing in Finnish industry, by industry group, 1960—1973. Helsinki 1975. 104 s. nid.
- B 13** TAUNO KALLINEN — KARI SIHTOLA — TIMO SUMMA — PEKKA YLÄ-ANTTILA, Teollisuuden kustannusrakenne toimialoittain ja kokoluokittain vuosina 1972 ja 1973. The cost structure of Finnish industry, by industrial and size groups, in 1972 and 1973). Helsinki 1975. 101 s. nid.
- B 14** HEIKKI LEHTIMÄKI — PENTTI VARTIA, Tärkeimmät raaka-aineiden hintaindeksit. English summary: Main raw material price indices. Helsinki 1976. 270 s. nid.
- B 15** SINIKKA SALO, Suomen teollisuuden varastot vuosina 1960—1975. (Finnish industrial inventories in 1960—1975). Helsinki 1977. 125 s. nid.
- B 16** JOUKO KOSONEN, Riskien siirtäminen ja vakuutustoiminta. (The transfer of risks and insurance). Helsinki 1977. 113 s. nid.
- B 17** SYNNÖVE VUORI, Toimialarationalisointi Suomen teollisuudessa vuosina 1961—1974. English summary: Branch rationalization in Finnish industry 1961—1974. Helsinki 1977. 186 s.
- B 18** PEKKA YLÄ-ANTTILA, Suomen ja Ruotsin metsäteollisuuden kannattavuus ja rahoitus- asema vuosina 1971—1976. English summary: A comparative study of forest industry profitability and financing in Finland and Sweden in 1971—1976. Helsinki 1978. 124 s.
- B 19** KARI SIHTOLA, Yksikkötyökustannusten kansainvälinen vertailu. English summary: An international comparison of unit labour costs. Helsinki 1978. 130 s. nid.
- B 20** KARI SIHTOLA, Teollisuuden työvoimakustannusten taso ja rakenne vuonna 1977 sekä arvioita niiden kehityksestä vuosina 1978—1980. English summary: The level and structure of labour costs in Finnish manufacturing industry in 1977 and estimates concernig their course in 1978—1980. Helsinki 1979. 125 s.
- B 21** SINIKKA SALO, Sektoreittaiset rahoitusvirrat Suomen kansantaloudessa vuosina 1970—78. English summary: Financial flows by sectors in the Finnish economy, 1970—78. Helsinki 1979. 109 s. nid.

- B 22** KARI ALHO, Rahamäärän kasvun hajottaminen eri tarjontalähteisiin sekä hajotteen soveltaminen eri rahalaitosten ottolainauskasvun selittämiseen. English summary: The decomposition of monetary growth in Finland with an application of the decomposition to the explanation of the growth of deposits in various banks. Helsinki 1979. 84 s. nid.
- B 23** ARJA VÄISÄNEN, Yksityisten kulutusmenojen rakenne ja kehitys Suomessa vuosina 1970—1982. (The Composition and Course of Private Consumption Expenditure in Finland, 1970—82). Helsinki 1980. 93 s. nid.
- B 24** PEKKA YLÄ-ANTTILA — ARMO HEIKKILÄ, Teollisuuden kannattavuuskehitys toimialoittain. English summary: Profitability in the Finnish industry by branches. Helsinki 1980. 110 s. nid.

SARJA C SERIES

- C 1** Suhdannevaihtelut Suomen kansantaloudessa v. 1958—1971. Osat I, II ja III. Helsinki 1972. 370 s.
- C 2** Suomen metalli- ja TEVA-tuotteiden ulkomaiseen kysyntään vaikuttaneet tekijät 1960-luvulla. Helsinki 1973. 86 s.
- C 3** CHRISTIAN EDGREN — RONALD WIMAN, Suomalainen siirtolaistyövoima Ruotsissa vuosina 1972. Helsinki 1973. 119 s.
- C 4** MATTI PURASJOKI, Investoinnit yrityksen finanssipäätöksenteon osana. Ekonometrinen simultaanimalli. Helsinki 1974. 135 s.
- C 5** AARNO SAVIAHO, Sahatavaramarkkinat ja vienti — Englannin markkinoiden lyhyen tähtäyksen ennustemalli. Helsinki 1975. 92 s.
- C 6** HEIKKI HÄMÄLÄINEN, Inflaatio toisen maailmansodan jälkeen Suomessa. Helsinki 1975. 33 s.
- C 7** JOUKO KOSONEN, Raha-aggregaattien soveltuvuudesta bruttokansantuotteen ja inflaation selittämiseen. Ekonometrinen analyysi Suomesta vuosina 1952—73. Helsinki 1975. 98 s.
- C 8** TAUNO KALLINEN, Työn tuottavuuden taso ja ansiotaso Suomen, Norjan, Ruotsin ja Tanskan tehdasteollisuudessa vuonna 1973. Helsinki 1976. 21 s.
- C 9** KARI ALHO, Rahoitusmarkkinoiden ja rahapolitiikan makrotaloudellinen tarkastelu korko- ja luottosäännöstelyn vallitessa. Helsinki 1977. 106 s.
- C 10** KARI ALHO, Ekonometrisen mallin ennustus- ja päätöksentekokäytön sekä mallin spesifioinnin ja estimoinnin välisten tilastollisten yhteyksien analyysi. Helsinki 1977. 191 s.
- C 11** ANTTI TANSKANEN, Dollarin termiinimarkkinat Suomessa 1973—76. Helsinki 1977. 43 s.
- C 12** PEKKA LASTIKKA, Tutkimus Suomen teollisuustyöntekijöiden työvoimapiireittäisistä palkkaeroista vuosina 1960—1971. Helsinki 1977. 177 s.
- C 13** LAURA VAJANNE, Kansantalouden ulkoisen tasapainokäsitteen vaihtoehtoisista tulokintatavoista. Helsinki 1978. 89 s.
- C 14** JUKKA LASSILA, Talonrakennustuotannon lyhyen aikavälin vaihtelut. Helsinki 1979. 125 s.
- C 15** TOM WALLDEN, Suomen Paperi- ja massateollisuuden tuotannon ja viennin rakenne vuosina 1955—1975 sekä kehitysarvio vuoteen 1985. Helsinki 1979. 206 s.
- C 16** TIMO SUMMA, Teollisen toimialan tehokkuusrakenne ja sen muutokset — Salterin teoria ja sen soveltuvuus toimialan kehitysprosessin tarkasteluun. Helsinki 1979. 116 s.
- C 17** TAUNO KALLINEN, Taseanalyysi Suomen teollisuudesta vuosina 1974—1976 ja vertaileva tarkastelu Ruotsin ja Norjan teollisuudesta vuosilta 1974—1976. Helsinki 1979. 69 s.
- C 18** JOUKO KOSONEN, Kansantalouden eläkemenot vuosina 1977—2010. Helsinki 1979. 28 s.
- C 19** HEIKKI VAJANNE — EERO PYLKKÄNEN — KARI SALMI, ETLAn kokonaistaloudellisen mallin ohjelmistokuvaus. Helsinki 1980. 169 s.

KESKUSTELUAIHEITA sarja — DISCUSSION PAPERS series

- No 1** JUKKA LASSILA, Asuntotuotannon ennustaminen rakennuslupien ja aloitettujen asuntojen avulla. 11.05.1976. 27 s.
- No 2** YRJÖ O. VARTIA, Ideal Log-change Index Numbers. 03.06.1976. 17 s.
- No 3** YRJÖ O. VARTIA, Fisher's Five-Tined Fork and Other Quantum Theories of Index Numbers. 19.06.1976. 23 s.
- No 4** Ks. No 3, korjattu versio, 10.08.1976. 23 s. (revised from No 3).
- No 5** PENTTI VARTIA, Huomioita työmarkkinoiden indeksiehdon inflaatiovaikutuksista. 20.09.1976. 15 s.
- No 6** YRJÖ O. VARTIA, Gammajakauman parametrien maximum-likelihood estimaattoreiden approksimoinnista. 10.10.1976. 19 s.
- No 7** HANNELE WALLENIUS — JYRKI WALLENIUS — PENTTI VARTIA, An Approach to Solving Multiple Criteria Macroeconomic Policy Problems and an Application. 15.10.1976. 27 s.
- No 8** JUKKA LASSILA, Lineaaristen simultaaniyhtälömallien parametrien estimointi rajoitetun informaation menetelmillä. 22.11.1976. 59 s.
- No 9** TIMO SUMMA, On Describing and Projecting Industrial Development. A Framework for Studying the Structural Change and Growth of the Finnish Pulp and Paper Industries. 27.11.1976. 42 s.
- No 10** PENTTI VARTIA, Indexed Deposits and Price Expectations. 21.02.1977. 18 s.
- No 11** KARI ALHO, Reaalitulon käsitteestä ja sen mittaamisesta. 16.06.1977. 16 s.
- No 12** ANTTI TANSKANEN, Suomen teollisuuden kansainvälinen erikoistuminen. 01.07.1977. 28 s.
- No 13** ANTTI TANSKANEN, The Effect of an Exogenous Factor Price Ratio on Trade Balance. 05.07.1977. 14 s.
- No 14** JUHANI TURKKILA, Julkisen talouden kasvu ja inflaatio — tutkimusluonnos. 23.11.1977. 13 s.
- No 15** JUHANI TURKKILA, Näkökohtia reaalisien nettoansiotasoindeksin laatimisesta. 12.12.1977. 26 s.
- No 16** CHRISTIAN EDGREN, Veronkannon aiheuttamista viiveistä verotusjärjestelmässä. 12.12.1977. 18 s.
- No 17** CHRISTIAN EDGREN — JUHANI TURKKILA — YRJÖ VARTIA, Tuloverotuksen analysoinnin matemaattisista ongelmista. 15.02.1978. 32 s.
- No 18** PENTTI VARTIA — YRJÖ VARTIA, Description of the Income Distribution by the Scaled F Distribution Model. 18.02.1978. 15 s.
- No 19** HEIKKI VAJANNE — YRJÖ VARTIA, ETLAn suhdannemallin iteratiivinen ratkaisumenetelmä. 26.05.1978. 25 s.
- No 20** YRJÖ O. VARTIA, Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Market Demand Functions. 31.08.1978. 18 s.
- No 21** KARI ALHO, The Econometric Foundation of Optimal Economic Policy: An Analysis of the Static Case. 12.09.1978. 15 s.
- No 22** KARI ALHO, Rahamäärän kasvun hajoittaminen eri tarjontalähteiden osalle. 14.01.1979. 27 s.
- No 23** CHRISTIAN EDGREN, Ett försök att mäta förändringar i offentliga handlingsparametrar: folkpensionerna, en case-study. 15.01.1979. 12 s.
- No 24** CHRISTIAN EDGREN, Sosiaaliturvamaksujen ennustamisesta. 15.01.1979. 11 s.
- No 25** YRJÖ O. VARTIA, Kvadraattisten mikroyhtälöiden aggregoinnista. 24.01.1979. 14 s.
- No 26** CHRISTIAN EDGREN, Selvitys lapsilisistä sekä niiden kehityksestä. 26.01.1979. 20 s.
- No 27** CHRISTIAN EDGREN, Julkisen sektorin tulonsiirtojärjestelmästä. 08.02.1979. 25 s.
- No 28** YRJÖ O. VARTIA, Ansiotasoindeksin "uudistamisesta": Alustus ja muistio työn hinnan sekä työpanoksen muutosten mittaamisesta. 13.02.1979. 22 s.
- No 29** YRJÖ O. VARTIA, Tuonnin yksikköarvoindeksin (1954 = 100) vuoden 1954 puuttuvien neljännesvuosiarvojen arvioiminen. 12.03.1979. 16 s.
- No 30** LAURA VAJANNE, Suomen palvelusten vienti ja tuonti vuosina 1964—1977. 14.03.1979. 46 s.

- No 31** CHRISTIAN EDGREN, En analys av sambandet mellan förändring i statlig inkomst och förändring i beskattad inkomst. 26.03.1979. 35 s.
- No 32** PENTTI VARTIA, Wage Indexation and Price and Wage Changes. 20.03.1979. 31 s.
- No 33** YRJÖ O. VARTIA ja PENTTI L. I. VARTIA, Liikevaihtoveron korottaminen ja tulojen ostovoima. 30.03.1979. 138 s.
- No 34** CHRISTIAN EDGREN, Luettelo tärkeimmistä välittömän verotuksen veroperusteista. 30.03.1979. 40 s.
- No 35** YRJÖ O. VARTIA, About "Exactness" of Index Number Formulas in Demand Worlds. 06.04.1979. 21 s.
- No 36** ARMO HEIKKILÄ — TIMO SUMMA — PEKKA YLÄ-ANTTILA, Metall- ja metsäteollisuusyritysten kannattavuudesta ja rahoitusasemasta 1970-luvulla. 25.04.1979. 27 s.
- No 37** JUKKA LASSILA, Intermediate-run Effects of Fiscal Policy with Balanced Budget. 04.05.1979. 27 s.
- No 38** KARI ALHO, Rahamäärän kasvun jakautuminen eri rahalaitosten talletusten osalle. 16.05.1979. 24 s.
- No 39** SINIKKA SALO, Sektoreittaiset rahoitusvirrat v. 1970—77. 04.06.1979. 89 s.
- No 40** HEIKKI HÄMÄLÄINEN — SAMI TUURNA, Talouskehityksen taustaa. 11.06.1979. 14 s.
- No 41** CHRISTIAN EDGREN, En analys av den undre gränsen för beskattad inkomst som skattepolitiskt instrument. 17.07.1979. 37 s.
- No 42** CHRISTIAN EDGREN, Vähennysjärjestelmän merkitys verotuksessa. 17.07.1979. 51 s.
- No 43** ROLF MAURY, Tuottavuuteen ja työllisyysvaikutuksiin liittyviä näkökohtia. Metalliteollisuutta koskeva tarkastelu. 31.07.1979. 38 s.
- No 44** YRJÖ O. VARTIA — JOHN A. WEYMARK, Four Revealed Preference Tables. 01.08.1979. 19 s.
- No 45** ANTTI SUVANTO, Error Learning and Return-to-normality in Public Forecasts: An Empirical Note. 01.10.1979. 26 s.
- No 46** HEIKKI HÄMÄLÄINEN — SAMI TUURNA, Suomen kansantalouden pitkän aikavälin kehitysmahdollisuudet. 16.10.1979. 56 s.
- No 47** PEKKA YLÄ-ANTTILA, Suomen ja Ruotsin metsäteollisuuden kannattavuusvertailu v. 1971—78. 31.10.1979. 11 s.
- No 48** JUHANI TURKKILA, Reaalitulon verottaminen. 06.11.1979. 20 s.
- No 49** HEIKKI VAJANNE — EERO PYLKKÄNEN — KARI SALMI, ETLAn suhdannemallin ATK-dokumentti. 16.11.1979. 169 s.
- No 50** HEIKKI HÄMÄLÄINEN, Kansantalouden kehitysnäkymät. 27.11.1979. 18 s.
- No 51** KARI ALHO — SINIKKA SALO, Rahoitusvirtamatriisin liittäminen ETLAn suhdannemalliin. 20.12.1979. 13 s.
- No 52** ANTTI SUVANTO, Econometric Studies on the Demand for and the Supply of Money in Finland: A Survey. 26.02.1980. 17 s.
- No 53** YRJÖ O. VARTIA, Comments on W.E. Diewert's Paper "The Economic Theory of Index Numbers: A Survey". 14.03.1980. 8 s.

ERIPAINOSSARJA — REPRINT SERIES

- No 1** ANTTI TANSKANEN, Maan velanhoitokyvystä. *Unitas* 4/1976: 203—210.
- No 2** YRJÖ O. VARTIA, Ideal Log-change Index Numbers. *Scandinavian Journal of Statistics* 1976/3:121—126.
- No 3** ANTTI TANSKANEN, Suomen teollisuuden kansainvälinen erikoistuminen. English Summary: The International Specialization of the Finnish Industry. *Kansantaloudellinen aikakauskirja* 1977/4:416—434.
- No 4** ANTTI TANSKANEN, Stability Conditions in Dynamic Foreign Exchange Markets. *Weltwirtschaftliches Archiv* 1977 Band 113 Heft 4:607—622.
- No 5** YRJÖ O. VARTIA, Fisher's Five-tined Fork and Other Quantum Theories of Index Numbers. Reprint from *Theory and Applications of Economic Indices*, edited by W. Eichorn, R. Henn, O. Opitz and R. W. Shephard, Physica-Verlag, Würzburg 1978; p. 271—295.

- No 6** PENTTI VARTIA, A Note on the Calculation of Arc-elasticities. *Scandinavian Journal of Economics* Vol. 78 1977:485—487.
- No 7** PENTTI VARTIA, Huomioita Elinkeinoelämän Tutkimuslaitoksen mallista. *Taloustieteellisen Seuran vuosikirja* 1977:44—51.
- No 8** HEIKKI HÄMÄLÄINEN, Mono- ja monitieteellisyys taloudellisessa tutkimustyössä. *Kansantaloudellinen aikakauskirja* 1978/2:199—202.
- No 9** TAUNO RANTA, Suomen asema kansainvälisen talouden kentässä. *Areena* 2:336—347, Tekninen Kustannus Oy, Turku 1978.
- No 10** PEKKA YLÄ-ANTTILA, Suomen ja Ruotsin metsäteollisuuden yritysکوhtainen kannattavuus vuosina 1971—76. *Kansantaloudellinen aikakauskirja* 1978/3:247—255.
- No 11** HANNELE WALLENIUS, JYRKI WALLENIUS and PENTTI VARTIA, An Approach to Solving Multiple Criteria Macroeconomic Policy Problems and an Application. *Management Science*, Vol. 24, No 10, June 1978: 1021—1030.
- No 12** TIMO SUMMA ja PEKKA YLÄ-ANTTILA, Metall- ja metsäteollisuusyritysten kannattavuudesta 1970-luvulla. *Unitas* 1/1979:22—30.
- No 13** HANNELE WALLENIUS, JYRKI WALLENIUS and PENTTI VARTIA, An Experimental Investigation of an Interactive Approach to Solving Macroeconomic Policy Problems. *Operational Research '78. Proceedings of the Eighth IFORS International Conference on Operational Research Toronto, Canada, June 19—23, 1978* Edited by K. B. Haley, an *Operational Research*, Edited by K. B. Haley, 878—894.
- No 14** PENTTI VARTIA, Indexed Deposits and Price Expectations. *Weltwirtschaftliches Archiv*, Band 115, 1979 Heft 2, 2242—254.
- No 15** KARI ALHO, Tuotantokelijöiden hintojen työllisyysvaikutuksia, *Taloustieteellisen Seuran vuosikirja* 1979. 27—37.
- No 16** KARI ALHO, Rahamäärästä ja sen säätelystä kansantaloudessa. *Synpunkter på penningmängden och dess reglering inom folkhushållningen*. *Unitas* 4/1979: 225—232.
- No 17** HEIKKI HÄMÄLÄINEN, Kokonaistaloudellisen mallin ja taloustilastojen käyttö talouspolitiikan suunnittelussa. *Kansantaloudellinen aikakauskirja* 1980/1:43—48.