VATT-TUTKIMUKSIA

35

VATT-RESEARCH REPORTS

Saku Aura

LORENZ-KÄYRÄT, HYVINVOINTITEORIAT JA TILASTOLLINEN PÄÄTTELY

VALTION TALOUDELLINEN TUTKIMUSKESKUS

Government Institute for Economic Research

Helsinki 1996

ISBN 951-561-174-1

ISSN 0788-5008

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Hämeentie 3, 00530 Helsinki, Finland

J-Paino Ky

Helsinki 1996

Aura, Saku: Lorenz-käyrät, hyvinvointiteoriat ja tilastollinen päättely. (Lorenz Curves, Welfare and Statistical Inference.) Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, 1996 (B, ISSN 0788-5008; No 35), ISBN 951-561-174-1.

TIIVISTELMÄ: Tutkimuksessa esitellään menetelmä, jolla hyvinvointi- ja eriarvoisuusvertailuja voidaan tehdä tuloaineistojen perusteella ilman vahvoja oletuksia hyvinvointifunktion tai eriarvoisuusindeksin ominaisuuksista. Menetelmä perustuu Lorenz-käyrään ja sen yleistyksiin. Tutkimuksen toinen pääaihe on otoksesta estimoitujen Lorenz-käyrien tilastollisen päättelyn teoria. Lorenz-käyrien tärkeimmät tilastolliset ominaisuudet käsitellään sekä kehitetään simulaatioihin perustuva parannus tällä hetkellä standardiaseman saavuttaneeseen testausmenettelyyn. Tutkimuksen empiirisessä osassa menetelmiä sovelletaan Suomen käytettävissä olevien tulojen jakaumaan vuosilta 1971, 1976, 1981, 1985 ja 1990. Tärkein tässä osassa saatu tulos on, että suhteellinen eriarvoisuus on ollut vuonna 1971 suurimmillaan ja muutokset suhteellisessa eriarvoisuudessa vuosien 1976 ja 1990 välisenä aikana ovat olleet pieniä. Niin sanotun yleistetyn Lorenz-kriteerin mielessä on hyvinvointi tulojakauman valossa kasvanut tarkasteluperiodilla kronologisessa järjestyksessä.

ASIASANAT: Lorenz-käyrät, tulonjako, eriarvoisuus, hyvinvointiteoriat, tilastollinen päättely, ei-parametriset menetelmät

ABSTRACT: The main purpose of this study is to present a method that can be used to make inequality and welfare comparisons from income data with minimal assumption on the properties of welfare function and inequality index. The method is based on the Lorenz curve and its generalisations. The other main theme in the study is statistical inference properties concerning empirical Lorenz curves. Asymptotical sampling properties of the Lorenz curve are discussed. An improvement to the standard testing procedure of dominance relations is suggested. It is based on simulating the distribution of the test statistic. The empirical part of this study is concerned with the Finnish distribution of disposable income in 1971, 1976, 1981, 1985 and 1990. It is found out that relative inequality has been highest on 1971, but the changes in relative inequality during period 1976-1990 have been almost negligible. Using generalised Lorenz criterion it is also found out that welfare in income distribution has been rising in chronological order for the whole observation period.

KEY WORDS: Lorenz Curves, Income Distribution, Inequality, Welfare, Statistical Inference, Non-parametric Methods

ESIPUHE

Tulonjaon ja eriarvoisuuden tutkiminen on korkealle kivunneen työttömyyden takia taloustieteen yhä tärkeämpi osa-alue. Suomessa tämän alueen tutkimiseen on erityisen hyvät mahdollisuudet käytettävissä olevien korkeatasoisten aineistojen ansiosta. Tämä tutkimus jatkaakin uutta luovasti suomalaista vahvaa tulonjakotutkimuksen perinnettä.

Tutkimuksen tärkein anti on sen korkeatasoinen menetelmällinen ote. Tutkimuksessa luodaan katsaus moderneihin taloustieteen menetelmiin, joilla tulonjakoa ja taloudellista hyvinvointia voidaan tutkia. Tutkittaviin suureisiin liittyvän tilastollisen epävarmuuden mallintaminen sillä tarkkuudella kuin tässä tutkimuksessa on myös uutta suomalaisessa tutkimuskentässä.

Tutkimuksen empiirisessä osassa tarkastellaan Suomen käytettävissä olevien tulojen jakaumaa vuosina 1971, 1976, 1981, 1985 ja 1990. Nämä tarkastelut tuovat valoa tapahtuneeseen kehitykseen ja antavat pohjan uusien aineistojen tullessa käyttöön myös tuoreen kehityksen analysointiin.

Tämä tutkimus on osa Valtion taloudellisen tutkimuskeskuksen hyvinvointipalveluja, tulonsiirtoja ja tulonjakoa käsittelevää tutkimusprojektia. Tutkimuksessa esitettävät tulokset pyrkivät täydentämään tämän projektin puitteissa aikaisemmin julkaistuja tutkimuksia.

Valtion taloudellisen tutkimuskeskuksen puolesta haluan kiittää tekijää tästä tutkimuksesta.

Helsinki, 14. kesäkuuta 1996

Seppo Leppänen

SAATTEKSI

Tämä työ perustuu Helsingin yliopistossa toukokuussa 1996 hyväksyttyyn kansantaloutieteen pro gradu-tutkielmaani. Tutkielmani ohjaajana ja tarkastajana toimi professori Yrjö Vartia. Toisena tarkastajana toimi professori Erkki Koskela. Heille haluan esittää lämpimät kiitokset.

Aloitin tämän tutkimuksen tekemisen ollessani harjoittelijana Valtion taloudellisessa tutkimuskeskuksessa. Haluan kiittää VATTia työtiloista sekä tämän tutkimuksen julkaisemisesta. Erityinen kiitos kuuluu johtava ekonomisti Ilpo Suoniemelle tämän tutkimuksen aiheen antamisesta, tutkimusohjauksesta sekä tämän tutkimuksen käsikirjoituksen kommentoinnista.

Kaikista tässä työssä mahdollisesti esiintyvistä virheistä ja virhepäätelmistä vastaan kuitenkin itse.

Saku Aura

Sisällys

1 Johdanto

Tämän tutkimuksena tarkoituksena on esitellä metodi, joka mahdollistaa hyvinvointi- ja eriarvoisuusvertailujen tekemisen tulojakaumien välillä tekemättä epärealistisia oletuksia yhteiskunnan hyvinvointifunktion tai eriarvoisuusindeksin oikeasta funktiomuodosta. Esitettävät tulokset koskevat laajoja hyvinvointifunktio- ja eriarvoisuusindeksiperheitä. Tulojakaumien väliltä pyritään löytämään dominanssirelaatioita, joiden perusteella voidaan tehdä johtopäätös, että kaikki tietyntyyppiset hyvinvointifunktiot tai eriarvoisuusindeksit olisivat päätyneet samaan tulokseen kahden jakauman välisestä järjestyksestä riippumatta siitä, mikä nimenomainen hyvinvointifunktio- tai eriarvoisuusindeksi on valittu. Aina tietenkään tällaista dominanssia ei esiinny. Näissä dominanssivertailuissa keskeisiksi työvälineiksi osoittautuvat Lorenz-käyrä ja sen yleistykset. Esiteltyä metodia sovelletaan tämän tutkimuksen empiirisessä osassa suomalaisen tulojakauman kehitykseen vuosien 1971-1990 välillä.

Tulojakaumien tarkastelun motiivina tässä tutkimuksessa on yksilöiden välisten hyvinvointierojen tarkastelu. Näin siis eksplisiittisesti oletetaan, että jossain mielessä tulotietojen pohjalta yksilöiden välillä voidaan tehdä hyvinvointivertailuja. Esitettävä taloudellinen teoria soveltuisi tältä osin paremmin dynaamiseen analyysin, jossa tarkastelun kohteena olisivat yksilöiden elinkaaritulot, mutta käytettävissä olevan aineiston johdosta empiiriset tarkastelut tehdään kotitalouksien poikkileikkaustulojen perusteella. Poikkileikkaustulojen käyttö hyvinvointivertailuissa vaikeuttaa vertailujen tulkintaa sen takia, että puhtaasti satunnaisen yksilöiden tulojen "vuodesta vuoteen" vaihtelun vaikutusta saatuihin tuloksiin ei pystytä analysoimaan.

Hyvinvointia tulojakauman valossa tarkastellaan tässä tutkimuksessa kahden peruskysymyksen ympärillä: miten arvottaa keskimääräistä tulojen suuruutta ja toisaalta oikeudenmukaista tulonjakoa. Tähän kysymykseen ja siihen, miten tulonjaon oikeudenmukaisuutta pitäisi mitata, esitetään erilaisia mahdollisia vastauksia. Esitettävät tarkastelut eivät sellaisenaan ota kantaa sinänsä tärkeään kysymykseen siitä, että johtaako tasaisempi tulojakauma kenties pienempään kokonaistuloon, vaan tulojakaumien välinen vertailu tehdään erilaisten tulonjako - kokonaistulokombinaatioiden välillä. Tällä tavalla ymmärrettynä tässä esitettävä analyysi on ex post-analyysiä, vertailut tehdään lopputilojen välillä välittämättä siitä, miten lopputiloihin ollaan päädytty.

Tämän tutkimuksen tarkasteluiden toinen pääaihe on vertailuissa hyödyllisiksi osoittautuviin Lorenz-käyriin liittyvä tilastollinen teoria. Käytännön empiirisissä sovelluksissa on usein unohdettu otoksesta estimoitujen Lorenz-käyrien satunnaismuuttujaluonne. Tässä tutkimuksessa esitetään keino huomioida estimoituihin Lorenz-käyriin liittyvä tilastollinen epävarmuus konsistentillä tavalla tarkasteluissa.

Tutkimuksen rakenne

Luvussa kaksi määritellään erilaiset Lorenz-käyrät ja käydään lyhyesti läpi näihin liittyviä ominaisuuksia. Luvun tarkoituksena on esittää Lorenz-käyriin liittyvät matemaattiset ominaisuudet sillä tarkkuudella kun se jatkossa esiintyvän analyysin kannalta on välttämätöntä. Tästä syystä talousteoria on selvästi taustalla tässä luvussa, erityisesti tavanomaisen Lorenz-käyrän yhteydessä kyse voisi olla mielivaltaisesta positiivisesta satunnaismuuttujasta. Tulojakaumien tarkastelu on vain yksi tavanomaisen Lorenz-käyrän sovellusalue.

Luvussa kolme luodaan katsaus siihen, miten yhteiskunta voisi arvottaa erilaisia tulojakaumia. Luvun alkuosan rakenne on aksiomaattinen: siinä lähdetään abstraktista oletuksesta, että yhteiskunta koostuu kokoelmasta identtisiä agentteja ja tarkastellaan erilaisia potentiaalisesti mielenkiintoisia vaihtoehtoja tulojakaumien vertailukriteereiksi tällaisessa tilanteessa. Näiden vertailukriteerien yhteydet erilaisiin Lorenz-käyriin ovat tämän luvun tärkein anti. Tähän lukuun liittyy läheisesti liite yksi, johon osa lukuun liittyvistä todistuksista on koottu tekstin luettavuuden parantamiseksi. Luvun kolme loppuosa sisältää keskustelua siitä, kuinka näitä identtisille agenteille johdettuja hyvinvointiteorioita voitaisiin soveltaa reaalimaailman hyvinvointianalyyseihin, joissa joudutaan huomioimaan ihmisten ja kotitalouksien välisiä eroja.

Neljännessä luvussa luodaan katsaus tuloaineistojen tilastolliseen tutkimukseen ja erityisesti Lorenz-käyriin liittyvään tilastolliseen päättelyyn. Luvun alkuosassa perustellaan tässä valittua ei-parametrista lähestymistapaa, jonka jälkeen esitellään erilaisten Lorenz-käyrien asymptoottiset otantaominaisuudet. Tekstin luettavuuden takia on eräiden tärkeiden tulosten todistaminen jätetty erilliseen liitteeseen kaksi. Lorenz-käyrien otantaominaisuuksia käsittelevän osuuden jälkeen luvussa neljä pohditaan taloudellisesti mielenkiintoisten hypoteesien testaamista kyseisten asymptoottisten tulosten pohjalta. Luvun lopussa esitellään simulointimenettelyyn perustuva parannus kirjallisuudessa tällä hetkellä standardiaseman saavuttaneeseen testausmenettelyyn ja tarkastellaan tämän simulointimenettelyn avulla saavutettuja tehokkuusetuja testauksessa.

Viides luku esittelee lukujen kaksi, kolme ja neljä teorian käyttämistä käytännön tuloaineiston tarkasteluun. Tutkimuksen aineistona on suomalainen kotitaloustiedusteluaineistosta vuosille 1971, 1976, 1981, 1985 ja 1990 laskettu käytettävissä olevien tulojen jakauma. Luvussa viisi saadaan tulos, että erilaisissa hyvinvointidominanssi- ja eriarvoisuustarkasteluissa hyvinvointi tulojakauman valossa on kasvanut näinä vuosina kronologisessa järjestyksessä. Tosin tämä tulos on tutkimuksessa tehtyjen oletusten lisäksi riippuvainen tarkastelijan suhtautumisesta taloudelliseen eriarvoisuuteen. Jos lähtökohdaksi otetaan niin sanottu absoluuttinen eriarvoisuusnäkemys, ei näin vahvoja tuloksia voida esittää. Suhteellinen eriarvoisuus näyttää tämän aineiston valossa pienentyneen vuosien 1971 ja 1976 välillä huomattavasti ja vuoden 1976 jälkeen olleen melkein vakio. Eri tilastollisia riskitasoja merkitsevyyskriteereinä käyttäen saataisiin suhteellisen eriarvoisuuden vuoden 1976 jälkeisestä kehityksestä erilaisia tuloksia, mutta tutkimuksen kvalitatiivinen tulos on, että suhteellinen eriarvoisuus ei ole paljon muuttunut aikavälillä 1976-1990.

Luvussa kuusi esitetään tämän tutkimuksen johtopäätökset sekä annetaan ehdotuksia mahdollisiksi mielenkiintoisiksi jatkotutkimusaiheiksi.

2 Lorenz-käyrät

Tämän luvun tarkoituksena on esitellä erilaiset Lorenz-käyrät ja tarkastella niiden ominaisuuksia. Lorenz-käyrät osoittautuvat luvun kolme hyvinvointiteoreettisessa analyyseissä tärkeiksi työkaluiksi. Tässä luvussa käydään läpi Lorenz-käyrien ominaisuuksia sillä tarkkuudella, kuin se luvun kolme kannalta on välttämätöntä.

2.1 Tavanomainen Lorenz-käyrä

Lorenz-käyrien ja Lorenz-dominanssin merkitys yhteiskunnan hyvinvoinnille on ymmärretty Atkinsonin usein siteeratusta vuoden 1970 artikkelista saakka. Siinä esitetään tulos, jonka mukaan jokainen eriarvoisuutta kaihtava yhteiskunnan hyvinvointifunktio preferoi tulojakaumaa, jonka Lorenz-käyrä on toisen yläpuolella (dominoi toista jakaumaa), jos näiden jakaumien keskiarvot ovat yhtä suuret. Myöskin jokainen suhteellisen eriarvoisuuden indeksi (määritellään myöhemmin) saa pienemmän arvon jakaumalla, jonka Lorenz-käyrä dominoi toista. Jos Lorenz-käyrät leikkaavat, ei näin vahvoja tuloksia voida esittää. Nämä tulokset todistetaan myöhemmin.

Atkinson esittää myös artikkelissaan yhteyden yksilön valintateorian epävarmuuden vallitessa ja yhteiskunnan hyvinvointiteorian välille. Kaikkia näitä tuloksia voisi käyttää epävarmuuden tarkasteluun, kunhan hyötyfunktiokäsite vaihdetaan yksilön hyötyfunktioksi, eriarvoisuuden kaihtaminen riskinkaihtamiseksi ja tulojakauma tuottojakaumaksi.

Määritelmä 1. **(Lorenz-käyrä)** Muuttujan X $(X \geq 0)$[1] Lorenz-käyrä on

(1) $$L(p; X) = \int_0^p X(P)dP/\mu_X,$$

jossa $X(p) = \inf\{x|F(x) \geq p\}$ on kertymäfunktion käänteisfunktio eli fraktiilifunktio, μ_X on muuttujan X keskiarvo ja $p \in [0, 1]$.

Määritelmä 2. **(Lorenz-dominanssi)** Muuttujan X sanotaan vahvasti Lorenz-dominoivan muuttujaa X', jos $\forall\ p \in [0, 1] : L(p; X) \geq L(p; X')$ ja $L(p^*, X) > L(p^*, X')$ ainakin yhdessä

[1] Tässä oletetaan jakaumilla olevan jonkin maksimiarvon $x_{\max} < \infty$, jonka yläpuolella todennäköisyysmassaa ei ole. Tämä on tarpeettomankin vahva matemaattinen oletus tämän tutkimuksen tarpeisiin, mutta koska kyse on reaalimaailman tulojakaumista, ei tämä oletus ole epärealistinen. Vaihtoehtoinen ja riittävä oletus tämän tutkimuksen tarkasteluihin olisi olettaa toisten momenttien olevan kaikissa jakaumissa äärellisiä.

pisteessä p^*. Vahva Lorenz-dominanssi merkitään $X >_L X'$. Heikko Lorenz-dominanssi, jossa tuota aitoa epäyhtälöä ei vaadita, on $\mathcal{R}_+^n$:n aito osittainen (siis transitiivinen, refleksiivinen ja antisymmetrinen) järjestys. Sitä merkitään $X \geq_L X'$.

Lorenz-käyrän tulkinta on seuraava. Olkoon X esimerkiksi käytettävissä olevat tulot. Tällöin $L(p)$ kertoo sen osuuden, jonka köyhimmät $p \cdot 100\,\%$ ihmisistä yhteensä saavat tuloista.[2] Tästä ominaisuudesta ja kaavasta (1) nähdään seuraavat Lorenz-käyrän ominaisuudet:

1) $L(0) = 0$.
2) $L(1) = 1$.
3) $L(p)$ on kasvava p:n funktio. Tämä on helpoin nähdä, jos muuttujan X jakauma on jatkuva:

$$\begin{aligned}\frac{\partial L(p)}{\partial p} &= \frac{\partial}{\partial p}\int_0^{X(p)} xf(x)dx/\mu_X = X(p)f(X(p))\frac{\partial X(p)}{\partial p}/\mu_X = X(p)f(X(p))\frac{\partial F^{-1}(p)}{\partial p}/\mu_X \\ &= X(p)/\mu_X > 0.\end{aligned}$$

4) $L(p)$ on konveksi p:n funktio. Tämä nähdään helpoimmin X:n jakauman ollessa jatkuva ominaisuuden 3) perusteella:

$$\frac{\partial^2 L(p)}{\partial^2 p} = \frac{\partial}{\partial p}X(p)/\mu_X = \frac{\partial}{\partial p}F^{-1}(p)/\mu_X = \frac{1}{f(F^{-1}(p))}/\mu_X > 0.$$

5) Keskiarvo ja Lorenz-käyrä yhdessä määräävät jakauman yksikäsitteisesti. X:n ollessa jatkuva tämä nähdään ominaisuuden 3) perusteella, sillä Lorenz-käyrän derivaatan ja keskiarvon tuntemisen avulla tunnemme fraktiilifunktion, joka kertymäfunktion käänteisfunktiona määrää jakauman yksikäsitteisesti.
6) Lorenz-käyrä on invariantti kaikkien tulojen kertomiselle positiivisella vakiolla. Eli jos $X' = k \cdot X$, jossa $k > 0$, niin $L(p; X') = L(p; X)$.

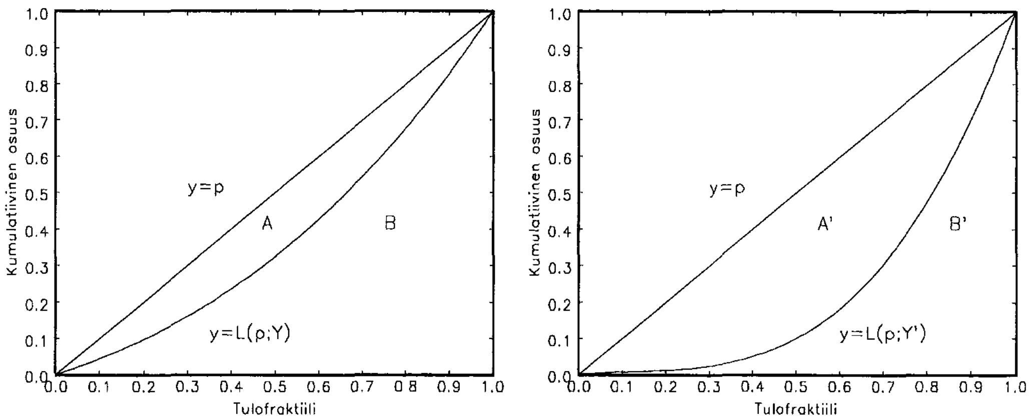
Näistä ominaisuuksista ja Lorenz-käyrän määritelmästä seuraa se, että Lorenz-käyrä on aina 45 asteen suoran ($y = x$) alapuolella, jos tulojakauma ei ole täysin tasainen. Tulojakauman ollessa täysin tasainen[3] on Lorenz-käyrä 45 asteen suora.

Eriarvoisuusmitoista yleisimmin käytetty ginikerroin liittyy Lorenz-käyrään siten, että ginikerroin on Lorenz-käyrän alle jäävän pinta-alan suhde 45 asteen suoran alle jäävään pinta-alaan. Siten vasemmanpuoleisessa kuviossa gini $G(X) = A/(A+B) = A/(1/2) = 2A$ ja oikeanpuoleisessa $G(X') = A'/(A'+B') = 2A'$. Kuviot on piirretty siten, että $G(X) = 0.23$ ja $G(X') = 0.53$.

[2] Täsmällisempi merkintätapa olisi $L(p; X)$, jossa merkitään näkyviin, minkä muuttujan Lorenz-käyrästä on kysymys. Lyhyempi $L(p)$-notaatio on kuitenkin käytössä silloin, kun sekaantumisen vaaraa ei ole.

[3] Tasaisella tulojakaumalla tarkoitetaan jakaumaa, jossa kaikki saavat yhtä suuret tulot. Sillä ei siis ole mitään tekemistä tilastotieteessä esiintyvän tasajakauman kanssa.

Kuvio 1. Lorenz-käyrä ja 45 asteen suora: i) tasaisempi tulojakauma ja ii) eriarvoisempi tulojakauma.



Tulojakaumien tarkastelu on vain yksi mahdollisista tavanomaisen Lorenz-käyrän sovelluskohteista. Muita sovellusaloja, joissa sitä on käytetty, ovat mm. ekologia, bibliografia sekä poliittisen vallan jakauman ja tieteellisten apurahojen jakauman tarkastelut (Goldie 1977, 765). Taloustieteessä esiintyviä muita sovellusalueita ovat tulojakaumatarkasteluja lähellä olevat kulutus- ja varallisuusjakaumatarkastelut sekä esimerkiksi työttömyyden keston (Beach & Kalinski 1986) jakauman tarkastelu.

2.2 Yleistetty Lorenz-käyrä

Yleistetty Lorenz-käyrä mahdollistaa vastaavien hyvinvointivertailujen tekemisen kahden tulojakauman välillä kuin Lorenz-käyrät silloin, kun niiden keskiarvot ovat eri suuret (Shorrocks, 1983). Tämä todistetaan myöhemmin.

Määritelmä 3. (Yleistetty Lorenz-käyrä) Muuttujan X ($X \geq 0$) yleistetty Lorenz-käyrä on

(2) $$GL(p) = \int_0^p X(P)dP = \mu_X L(p),$$

jossa $p \in [0,1]$.

Määritelmä 4. (Yleistetty Lorenz-dominanssi) Muuttujan X sanotaan vahvasti yleistetysti Lorenz-dominoivan muuttujaa X', jos $\forall\ p \in [0,1]$: $GL(p;X) \geq GL(p;X')$ ja $GL(p^*,X) > GL(p^*,X')$ ainakin yhdessä pisteessä p^*. Vahva yleistetty Lorenz-dominanssi merkitään $X >_{GL} X'$. Heikossa yleistetyssä Lorenz-dominanssissa tuota aitoa epäyhtälöä ei vaadita. Sitä merkitään $X \geq_{GL} X'$.

Yleistetty Lorenz-dominanssi on yhtäpitävää rahoitusteorian puolella yleisemmin käytetyn toisen kertaluvun stokastisen dominanssin kanssa (Thistle 1989). Tästä syystä yleistettyyn Lorenz-dominanssiin perustuvaa hyvinvointiteoriaa nimitetään myös toisen kertaluvun stokastisen dominanssin hyvinvointiteoriaksi.

Yleistetyllä Lorenz-käyrällä on seuraavat ominaisuudet:

1) $GL(0) = 0$.
2) $GL(1) = \mu$.
3) $GL(p)$ on kasvava p:n funktio.
4) $GL(p)$ on konveksi p:n funktio.
5) Yleistetty Lorenz-käyrä määrää jakauman yksikäsitteisesti.
6) Jos $X' = k \cdot X + a$, jossa $k > 0$, niin $GL(p, X') = k \cdot GL(p; X) + ap$.

Ominaisuudet 1)-5) seuraavat suoraan Lorenz-käyrän vastaavista ominaisuuksista ja ominaisuus 6) suoralla laskutoimituksella.

Yleistetty Lorenz-käyrä on siis tavanomainen Lorenz-käyrä skaalattuna muuttujan keskiarvolla. Tulojakauman ollessa täysin tasainen, on yleistetty Lorenz-käyrä pisteiden $(0,0)$ ja $(1, \mu_X)$ kautta kulkeva suora. Yleistetty Lorenz-käyrä, toisin kuin tavanomainen Lorenz-käyrä, on siis riippuvainen siitä, missä yksiköissä tulot on mitattu.

2.3 Absoluuttinen Lorenz-käyrä

Absoluuttisella Lorenz-käyrällä on sama merkitys absoluuttisen eriarvoisuuden teoriassa kuin Lorenz-käyrällä suhteellisen eriarvoisuuden teoriassa. Jos jakauman absoluuttinen Lorenz-käyrä on jatkuvasti toisen yläpuolella, on dominoivan jakauman absoluuttinen eriarvoisuus pienempi kuin dominoidun.

Määritelmä 5. (Absoluuttinen Lorenz-käyrä) Muuttujan X ($X \geq 0$) absoluuttinen Lorenz käyrä on

$$AL(p) = \int_0^p X(P)dP - p \cdot \mu_X = GL(p) - p \cdot GL(1) = \mu_X(L(p) - p), \tag{3}$$

jossa $p \in [0, 1]$.

Määritelmä 6. (Absoluuttinen Lorenz-dominanssi) Muuttujan X sanotaan vahvasti absoluuttisesti Lorenz-dominoivan muuttujaa X', jos $\forall\ p \in [0,1]$: $AL(p;X) \geq AL(p;X')$ ja $AL(p^*,X) > AL(p^*,X')$ vähintään yhdessä pisteessä p^*. Vahva absoluuttinen Lorenz-dominanssi merkitään $X >_{AL} X'$. Heikossa absoluuttisessa Lorenz-dominanssissa ei aitoa epäyhtälöä vaadita. Sitä merkitään $X \geq_{AL} X'$.

Absoluuttisella Lorenz-käyrällä on seuraavat ominaisuudet:

1) $AL(0) = 0$
2) $AL(1) = 0$. Tämä seuraa suoraan sijoituksesta määritelmän jälkimmäiseen osaan.
3) $AL(p)$ on vähenevä p:n funktio, kun $p < F(\mu)$, ja kasvava, kun $p > F(\mu)$. Sillä on minimi pisteessä $p = F(\mu)$. Tämä nähdään X:n jakauman ollessa jatkuva seuraavasti:

$$\begin{aligned}\frac{\partial AL}{\partial p} &= X(p) - \mu \\ &\leq 0, \text{ kun } p < F(\mu) \\ &\geq 0, \text{ kun } p > F(\mu).\end{aligned}$$

4) $AL(p)$ on p:n konveksi funktio, koska se on kahden konveksin funktion ($GL(p)$ ja $-p \cdot \mu$) summa.
5) Absoluuttinen Lorenz-käyrä ja keskiarvo määräävät jakauman yksikäsitteisesti.
6) Kaikkien tulojen samansuuruinen lisäys ei vaikuta absoluuttiseen Lorenz-käyrään. Jos $X' = k \cdot X + a$, jossa $k > 0$, niin $AL(p;X') = k \cdot AL(p,X)$.

Absoluuttisen Lorenz-käyrän määritelmän viimeisestä osasta nähdään myös absoluuttisen Lorenz-käyrän yhteys tavanomaiseen Lorenz-käyrään. Absoluuttinen Lorenz-käyrä ilmaisee tavanomaisen Lorenz-käyrän ja 45 asteen suoraan (täysin tasainen tulojakauma) välisen erotuksen skaalattuna muuttujan keskiarvolla. Tulojakauman ollessa täysin tasainen saa absoluuttinen Lorenz-käyrä arvon nolla koko määrittelyvälillään. Absoluuttinen Lorenz-käyrä ei siis ole invariantti sille, missä yksiköissä tulot on mitattu.

Nämä kaikki käyrät voitaisiin yleistää myös negatiivisia arvoja saaville muuttujille, mutta koska tutkimuksen tarkoituksena on käsitellä käytettävissä olevien tulojen jakaumaa, riittävät nämä määritelmät. Käyrien yleistämisestä negatiivisille muuttujille olisi haittana tulkinnallisen selkeyden menettäminen.

3 Hyvinvointiteoriat, kun agentit ovat identtisiä

Tämän luvun tarkoituksena on esittää kriteerit tulojakaumien "hyvyyden" tutkimiselle, joita voitaisiin myös empiirisesti soveltaa. Muodollisesti tämä tapahtuu tarkastelemalla erilaisia yhteiskunnan hyvinvointifunktioita tai yhteiskunnan arvotuskriteerejä, jotka riippuvat yhteiskunnan jäsenten tuloista. Ylevästä nimestään huolimatta yhteiskunnan hyvinvointifunktioiden ei ole tarkoitus olla "kirjaimellisesti" yhteiskunnan hyvinvointifunktioita, vaan niiden tarkoitus on esittää mahdollisia asenteita, joita yhteiskunnalla tai anonyymillä tarkkailijalla voisi olla erilaisia tulojakaumia kohtaan. Filosofisessa mielessä nämä funktiot voisi esimerkiksi ajatella Rawlsin (1973) "Veil of Ignorance"-asetelman avulla, jossa yksilö arvottaa erilaisia tulo- tai hyötyjakaumia tietämättä mikä hänen asemansa tulee kyseisissä jakaumissa olemaan. Tällöin yhteiskunnan hyvinvointiteoria on oikeastaan yksilön valintateoriaa epävarmuuden vallitessa, joten yhteiskunnan eriarvoisuuden kaihtaminen voidaan tulkita yksilön riskinkaihtamiseksi. Pragmaattisempi lähestymistapa on ajatus anonyymistä tarkkailijasta (esimerkiksi tutkija), joka haluaa tehdä vertailuja tulojakaumien välillä siten, että ne painottavat hänen näkemyksiään tulonjaon hyvyydestä tai oikeudenmukaisuudesta. Jatkossa kuitenkin käytetään nimityksiä yhteiskunnan hyvinvointifunktio ja yhteiskunnan päätössääntö.

Yhteiskunnan hyvinvointiteorioiden yllä määritellyssä mielessä ymmärrettynä täytyy ottaa kantaa kahteen tulojakaumaan liittyvään kysymykseen. Arkikielellä sanottuna kyse on siitä, kuinka iso kakku (kokonaistulo) on ja kuinka kakku jaetaan. Englanninkielisessä kirjallisuudessa kakun koosta puhuttaessa puhutaan tehokkuudesta (efficiency) ja kakun jaosta puhuttaessa taas tasa-arvosta (equity). Tätä nimityskäytäntöä noudatetaan myös tässä tutkimuksessa. Yhteiskunnan hyvinvointifunktio ilmaisee sen, miten näitä kahta asiaa painotetaan. Eksplisiittistä hyvinvointifunktioiden määräämistä ei tässä luvussa tehdä, vaan esitetään erilaisia potentiaalisesti mielenkiintoisia hyvinvointifunktioiden luokkia ja tarkastellaan millaisia osittaisia järjestyksiä nämä funktioluokat indusoivat tulojakaumien välille. Näillä tarkasteluilla on suora yhteys edellisessä luvussa määriteltyihin Lorenz-käyriin. Näin pystytään tekemään vertailuja eri tulojakaumien välillä sopimalla ainoastaan perusperiaatteista tehokkuuden ja tasa-arvon painottamisesta siten, että tulokset kattavat kuitenkin laajan perspektiivin näkemyksiä tehokkuuden ja tasa-arvon välisestä vaihtosuhteesta.

Tulojakaumien oletetaan tässä tarkastelussa olevan diskreettejä ja yhteiskunnassa oletetaan olevan äärellisen määrän agentteja. Tästä oletuksesta luovutaan tilastollista päättely koskevassa luvussa neljä, jossa tulojakauman oletetaan olevan jatkuvan satunnaismuuttujan jakauma. Lisäksi kaikkien agenttien oletetaan olevan identtisiä tarpeiltaan, joten yhteiskunnan ei tarvitse päätössäännössään ottaa huomioon eroja erityyppisten agenttien välillä. Tämä on tietenkin erittäin vahva oletus, sillä reaalimaailmassa agentit ovat järjestäytyneet erilaisiksi kokonaisuuksiksi (kotitalouksiksi) ja yhteiskunnallisia päätöksiä arvioitaessa otetaan huomioon erilaisten kotitaloustyyppien erot tarpeissa. Identtiset agentit -oletuksen tarkoitus ei olekaan olla realistinen, vaan mahdollistaa kauniiden matemaattisten tulosten esittäminen ja luoda lähtökohta laajentaa analyysia realistisempaan suuntaan. Tätä problematiikkaa käsitellään tämän luvun lopussa.

Määritelmä 7. Olkoon $X = (x_1, x_2, x_3, \ldots, x_n)$ vektori, jonka alkioina ovat agenttien $i = 1, 2, 3, \ldots, n$ tulot. Vektoria $X_{()} = (x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)})$, jossa $x_{(i+1)} \geq x_{(i)}$, nimitetään (diskreetiksi) järjestetyksi tulojakaumaksi.

Erilaisia vastauksia tehokkuus - tasa-arvoisuus -kysymykseen löytyy kirjallisuudesta monia. Klassinen utilitaristinen vastaus yhteiskunnan hyvinvoinniksi olisi laskea hyvinvointien keskiarvo. Jos lisäksi naiivisti oletetaan agentin hyvinvoinnin olevan yhtä suuri kuin hänen käytettävissään olevat tulot, saadaan yhteiskunnan hyvinvoinniksi $W_{cu}(X) = \sum_{i=1}^{n} x_i/n$. Erisuuruisia populaatioita käsiteltäessä tulojen (hyötyjen) keskiarvo on paremmin eettisesti perusteltavissa oleva suure kuin puhdas tulojen summa, koska se täyttää niin sanotun Senin populaatioperiaatteen (tämä määritellään myöhemmin). Realistisemman hyvinvointikriteerin klassisesta utilitarismista saa, jos olettaa ettei hyvinvointi suoraan ole ilmaistavista ihmisten tuloina, vaan jonain tulojen funktiona. Tällöin päädymme additiivisiin utilitaristisiin hyvinvointifunktioihin $W_u(X) = \sum_{i=1}^{n} u_i(x_i)/n$, missä funktion u_i voidaan esittää olevan joko sosiaalinen arvostus yksilön tuloille tai yksilön oma hyötyfunktio hänen tuloilleen.

Toinen tunnettu tapa ratkaista yhteiskunnan hyvinvointi on samaistaa se huonoimmin voivan yksilön hyvinvointiin. Jos edelleen oletetaan hyvinvoinnin riippuvan vain tuloista, saamme yhteiskunnan hyvinvointifunktioksi $W_R(X) = \min\{x_1, x_2, x_3, \ldots, x_n\} = x_{(1)}$. Tätä hyvinvointifunktiota nimitetään yleensä rawlsilaiseksi filosofi John Rawlsin mukaan.

Näitä kahta kriteeriä voidaan pitää tulojakaumien tarkastelussa ääripäinä. Rawlsin kriteeri on kiinnostunut vain huonoimman asemasta eikä ota millään tavalla huomioon keskimääräistä tulotasoa, kun taas naiivi klassinen utilitarismi, jossa hyödyt samaistetaan tulojen kanssa, on sellaisenaan puhdas tehokkuuskriteeri, joka ei ota mitenkään huomioon tulojen jakautumista.

3.1 Hyvinvointifunktioihin liittyviä määritelmiä

Jotta pääsisimme laajempiin kriteeri- ja hyvinvointifunktioluokkiin, tarvitsemme muutamia teknisiä määritelmiä.

Määritelmä 8. (Bistokastinen neliömatriisi) (Marshal & Olkin 1979, 8) Neliömatriisi $B_{n \times n}$ on bistokastinen, jos sen elementeille ovat seuraavat ehdot voimassa: $b_{ij} \geq 0$ kaikille i ja j, $\sum_{i=1}^{n} b_{ij} = 1$ kaikille j ja $\sum_{j=1}^{n} b_{ij} = 1$ kaikille i.

Bistokastisten matriisien luokka sisältää mm. identiteettimatriisit ja permutaatiomatriisit.

Määritelmä 9. (Schur-konkaavi funktio) (Shorrocks 1983, 4) Funktiota $W : \mathcal{R}^n \rightarrow \mathcal{R}$ sanotaan Schur-konkaaviksi, jos

1) $W(X) = W(\Pi X)$ kaikille $n \times n$ permutaatiomatriiseille Π. Tämä tarkoittaa funktion olevan symmetrinen argumenttiensa suhteen.
2) $W(BX) \geq W(X)$ kaikille $n \times n$ bistokastisille matriiseille B. Funktio on aidosti Schur-konkaavi, jos $W(BX) > W(X)$ kaikille bistokastisille matriiseille B lukuunottamatta identiteetti- ja permutaatiomatriiseja.

Schur-konveksi funktio F määritellään siten, että F on Schur-konveksi jos $-F$ on Schur-konkaavi.

Osan 2) merkitys Schur-konkaavin funktion määrittelyssä ei ole ehkä itsestään selvä, mutta sitä selvittää seuraava määritelmä selityksineen.

Määritelmä 10. (Pigou - Dalton-tulonsiirtoaksiooma) (Marshal & Olkin 1979, 6) Olkoon X tulojakauma. Hyvinvointifunktion W sanotaan täyttävän Pigou-Dalton-tulonsiirtoaksiooman, jos $W(Y) \geq W(X)$, kun Y on saatu jakaumasta X siirtämältä agentin i tuloista osuus $\theta \in]0,1[$ agentille j ja osuus θ agentin j tuloista agentille i. Tällöin $y_i = (1-\theta) \cdot x_i + \theta \cdot x_j$ ja $y_j = (1-\theta) \cdot x_j + \theta \cdot x_i$.

Pigou - Dalton-aksioomassa mainitut tulonsiirrot ovat aina tuloeroja tasoittavia. Pigou - Dalton-aksioomassa tulonsiirrot ovat muodossa, jossa kaksi agenttia vaihtaa keskenään tietyn osuuden tuloistaan. Helposti nähdään, että tällaisen kahden tulonsiirron nettovaikutus on sama, kuin sellaisella tulonsiirrolla, jossa tietty rahasumma siirretään rikkaammalta köyhemmälle. Olkoon nyt $x_j > x_i$. Tällöin Pigou - Dalton-aksiooman tulonsiirrossa rikkaammalta siirretään köyhemmälle summa $\theta \cdot (x_j - x_i)$ eli osuus θ agenttien tulojen erotuksesta.

Merkittävä tulos on se, että sarjan Pigou - Dalton-aksiooman tulonsiirtoja voidaan aina katsoa jonkun bistokastisen matriisin indusoimaksi.[4] Tämän valossa on helppo nähdä, että määritelmän osa 2) tarkoittaa, että hyvinvointifunktio W on eriarvoisuutta kaihtava tai kapulakielellä sanottuna tulonsiirrot hyväksyvä (transfer approving).

Funktiot muotoa $W(X) = \sum_{i=1}^{n} u(x_i)/n$, jossa funktio u on konkaavi, ovat Schur-konkaaveja funktioita (Marshal & Olkin 1979, 11). Nämä von Neumann - Morgenstern-hyvinvointifunktiot ovat tärkeä ryhmä Schur-konkaaveja funktioita. Täten on luonnollista, että kohta esitettävät hyvinvointitulokset ovat laajennettavissa koskemaan jatkuvia jakaumia ja hyvinvointifunktioita, jotka ovat muotoa $W(X) = \int_0^\infty u(x)f(x)dx$ eli ovat jatkuvia von Neumann - Morgenstern-hyvinvointifunktioita.

Toinen tärkeä erikoistapaus Schur-konkaaveista funktioista on Rawlsin kriteerin hyvinvointifunktio $W(X) = min\{x_1, x_2, \ldots, x_n\}$.

Määritelmä 11. **(Senin populaatioperiaate)** (Sen 1979, 35) Yhteiskunnan hyvinvointifunktion sanotaan täyttävän Senin populaatioperiaatteen, jos $W(X) = W(Y)$ jos Y on saatu X:stä monistamalla jokainen X:n alkio n kertaa.

Populaatioperiaate on perusteltavissa silloin, kun halutaan vertailla kahden erikokoisen populaation (esim. maan) tulojakaumia keskenään ja halutaan jättää väestönkasvun hyvinvointivaikutukset vertailun ulkopuolelle. Populaatioperiaate mahdollistaa vertailut kaikkien diskreettien tulojakaumien välille, sillä jos X:ssä on n_1-alkiota ja Y:ssä n_2-alkiota, voidaan vertailut tehdä uusien jakaumien Y' ja X' välillä, jossa Y' on saatu monistamalla jokainen Y:n alkio n_1 kertaa ja X' monistamalla jokainen X:n alkio n_2 kertaa.

---

[4] Tämä nähdään seuraavasti: olkoot X ja Y matriiseja, jotka indusoivat Pigou - Dalton-tulonsiirron. Tällöin on helppo nähdä, että ne ovat bistokastisia. Tällöin tulomatriisi $XY = [xy]_{kl} = [\sum_{i=1}^{n} x_{ki}y_{il}]$ ja tällaisen matriisin rivisumma (sarakesumma) $= \sum_{j=1}^{n} \sum_{i=1}^{n} x_{ki}y_{ij} = \sum_{i=1}^{n} x_{ki} \sum_{j=1}^{n} y_{ij} = \sum_{i=1}^{n} x_{ki} = 1$ eli matriisi XY on bistokastinen. Tällä ja induktiolla näemme, että bistokastisten matriisien tulo (kertoma) on bistokastinen, joten peräkkäisten Pigou - Dalton-tulonsiirtojen voidaan katsoa aina olevan jonkun bistokastisen matriisin indusoima. Myös käänteinen väite "kaikki bistokastiset matriisit indusoivat Pigou - Dalton-tulonsiirtojen sarjan" pätee, katso apulause sivulla 17, mutta sen todistus sivuutetaan.

3.2 Suhteellinen ja absoluuttinen eriarvoisuus sekä eriarvoisuusindeksit

Ennen kuin pääsemme käsittelemään hyvinvointifunktioita on pohdittava mitä tulojakauman eriarvoisuudella tarkoitetaan. Eriarvoisuus määritellään tässä implisiittisesti Pigou - Dalton-aksiooman avulla: eriarvoisuus on asia, jota Pigou - Dalton-tulonsiirto pienentää. Eriarvoisuudesta on kirjallisuudessa esitetty kaksi perusnäkökohtaa, nimittäin suhteellinen eriarvoisuusnäkemys ja absoluuttinen eriarvoisuusnäkemys.

Suhteellisella eriarvoisuusnäkemyksellä tarkoitetaan sitä, että kaikkien tulojen yhtä suuri prosentuaalinen skaalaus ei muuta eriarvoisuutta tulojakaumassa. Tällöin suhteelliset tuloerot pysyvät vakiona ja absoluuttiset kasvavat (pienenevät), jos kaikkien tulot kasvavat (pienenevät).

Absoluuttisen eriarvoisuusnäkemyksen mukaan kaikkien agenttien tulojen yhtä suuri lisäys ei muuta eriarvoisuutta tulojakaumassa. Tulojen yhtä suuri lisäys (pienennys) pitää absoluuttiset tuloerot vakiona, mutta suhteelliset tuloerot pienevät (kasvavat). Absoluuttisen eriarvoisuusnäkemyksen kannalta on tärkeää, että tulokäsitteenä on reaalitulo eikä nimellistulo. Muussa tapauksessa esimerkiksi tasainen inflaatio, joka täysimääräisesti kompensoitaisiin yhteiskunnan jäsenille, lisäisi absoluuttista eriarvoisuutta.

Cowell (1985) pienimuotoisessa empiirisessä kyselytutkimuksessaan tulee siihen tulokseen, etteivät ainakaan kaikki ihmiset sellaisenaan hyväksy suhteellisen eriarvoisuuden määritelmää. Osa koehenkilöistä, jotka olivat joko yhteiskunta- tai taloustieteilijöitä ja taloustieteen opiskelijoita, tuntuivat olevan enemmän absoluuttisen eriarvoisuuden kannalla ja lisäksi osa koehenkilöistä ei ollut koherenttisti kummankaan eriarvoisuuden määritelmän kannalla vastauksissaan.

Määritelmä 12. (Suhteellisen eriarvoisuuden indeksi) Funtiota $I : \mathcal{R}^n_+ \to \mathcal{R}$ nimitetään suhteellisen eriarvoisuuden indeksiksi, jos sillä on seuraavat ominaisuudet:[5]

1) $I = 0$, kun tulojakauma on tasainen.

2) $I(k \cdot X) = I(X)$, kun $k > 0$ eli I on homogeeninen astetta 0.

3) I on Schur-konveksi tulojen funktio.

[5] Huomattava ominaisuus tässä suhteellisen eriarvoisuuden indeksin määritelmässä on sen ordinaalinen luonne. Jos I_1 on suhteellisen eriarvoisuuden indeksi, niin $I_2 = g \circ I_1$, on suhteellisen eriarvoisuuden indeksi, kunhan funktio g on kasvava ja $g(0) = 0$. Tällöin siis erityisesti ei vaadita, että suhteellinen eriarvoisuusindeksi olisi skaalattu välille $[0, 1]$ ja että suhteellinen eriarvoisuus voitaisiin ilmaista prosenttilukuna. Eräät indeksit, kuten ginikerroin, toteuttavat tämän ominaisuuden, kun taas esimerkiksi yleistetyt entropiaindeksit eivät sitä toteuta.

Määritelmä 13. (Absoluuttisen eriarvoisuuden indeksi) Funktiota $I : \mathcal{R}^n_+ \rightarrow \mathcal{R}$ nimitetään absoluuttisen eriarvoisuuden indeksi, jos sillä on seuraavat ominaisuudet:

1) $I = 0$, kun tulojakauma on tasainen.
2) $I(X + k \cdot 1_n) = I(X)$, kun $k > 0$ ja 1_n on yksikkövektori.
3) I on Schur-konveksi tulojen funktio.

Määritelmien kolmas osa on Pigou - Dalton-aksiooma käänteisesti sovellettuna (vrt. hyvinvointifunktiot). Esitetty suhteellisen eriarvoisuuden indeksin määritelmä on aika tiukka ja se sulkee pois eräitä usein eriarvoisuuden kuvaamiseen käytettyjä tunnuslukuja. Esimerkiksi fraktiilisuhteet, joista yleisimmin käytetty lienee niin sanottu $P90/P10$, eivät ole suhteellisen eriarvoisuuden indeksejä tässä käytetyn määritelmän mukaan, koska ne eivät ole Schur-konvekseja funktioita.[6] Luku $P90/P10 = X(0.90)/X(0.10)$ ilmaisee tulojakauman pisteessä $p = 0.90$ saatujen tulojen suhteen tulojakauman pisteessä $p = 0.10$ saatuihin tuloihin. Fraktiilisuhteita käytetään eriarvoisuuden kuvaamiseksi esimerkiksi tuoreessa OECD:n tutkimuksessa (Atkinson et al. 1995, 86).

Hyvällä eriarvoisuusindeksillä on kyseisten määritelmien lisäksi eräitä haluttavia ominaisuuksia. Näitä ovat dekomponoituvuus populaatioryhmittäin ja tulolajeittain, sekä jakauman eri päissä tapahtuvien tulonsiirtojen halutunlainen painottaminen. Jo mainittu ginikerroin on dekomponoituva tulolajeittain, mutta ei populaatioryhmittäin.

Ginikertoimen käyttöön liittyy kuitenkin ongelmia. Otetaan tarkastelun kohteeksi kaksi erillistä tulonsiirtoa. Ensiksi köyhältä siirretään tietty summa rahaa melkein köyhälle. Tämä lisää luonnollisesti tuloeroja. Toinen tulonsiirto tapahtuu tulojakauman yläpäässä, vähän rikas saa rikkaalta yhtäsuuren tulonsiirron kuin äsken vähän köyhä köyhältä. Tämä jälkimmäinen tulonsiirto on luonteltaan tuloeroja tasoittava. Ginikerroin ei käyttäydy tässä tilanteessa ennustettavasti, vaan se voi joko kasvaa, pienetä tai pysyä vakiona. Ginikerroin ei siis toteuta ns. "vähenevän tulonsiirron periaatetta" (principle of diminishing transfer, Lambert 1989, 45), jonka mukaan absoluuttisesti samansuuruisen tulonsiirron vaikutuksen eriarvoisuusindeksiin pitäisi olla sitä pienempi, mitä korkeammalla tulotasolla se tapahtuu.[7]

[6] Tämä voidaan $P90/P10$:n kohdalla todeta tarkastelemalla tulonsiirtoa köyhimmältä toisiksi köyhimmälle taloudessa, jossa on kaksikymmentä jäsentä. Tämä eriarvoisuutta lisäävä tulonsiirto pienentää lukua $P90/P10$. Käytännön tuloaineistoihin sovellettuna fraktiilisuhteet tai muut ei-Schur-konveksit eriarvoisuusindeksit voivat kuitenkin käyttäytyä hyvinkin samantapaisesti kuin Schur-konveksit eriarvoisuusindeksit.

[7] Ginikerroin voidaan esittää diskreetissä muodossa $G = 1 + \frac{1}{n} - \frac{2(x_{(n)} + 2x_{(n-1)} + 3x_{(n-2)} + \cdots + nx_{(1)})}{n^2\mu}$ (Lambert 1989, 44). Jos nyt ajatellaan esimerkiksi tulonsiirtoa, jossa rikas antaa kolmanneksi rikkaimmalle yhtä suuren

Tämä ei ole haluttava ominaisuus eriarvoisuusindeksille ja ginikertoimestakin on esitetty laajennus, joka painottaa halutulla tavalla (valitsemalla erään reaalilukuparametrin arvo) tulojakauman eri osia (Yitzhaki 1983). Tämä indeksiperhe on erityisen käyttökelpoinen, jos halutaan tutkia eri tuloerien vaikutusta kokonaiseriarvoisuuteen, sillä se on tavanomaisen ginikertoimen tapaan dekomponoituva tulolajeittain.

Muitakin eriarvoisuusindeksiperheitä on kirjallisuudessa esitetty ja esimerkiksi Shorrocksin (1980) esittämä suhteellisen eriarvoisuuden indeksiperhe, joka tunnetaan yleistettynä entropiaperheenä, on erityisen käyttökelpoinen, jos halutaan tutkia eri osapopulaatioiden sisäistä ja osapopulaatioden välistä eriarvoisuutta, koska se on dekomponoituva populaatioryhmittäin ja sillä on haluttavat painottamisominaisuudet.

3.3 Tulojakaumiin liittyviä hyvinvointivertailukriteerejä ja hyvinvointifunktioluokkia

Kaikkein yleisimmin taloustieteellisessä tekstissä esiintyvä hyvinvointivertailukriteeri on Pareto-kriteeri.

Määritelmä 14. (Pareto-kriteeri) Tulojakauma X on Pareto-kriteerillä heikosti parempi kuin X', jos $x_i \geq x_i' \; \forall \; i = 1 \ldots n$. Pareto-paremmuus merkitään $X \geq_P X'$.

Jos tulojakauma X on Pareto-parempi kuin X', niin $W(X) \geq W(X')$ kaikilla argumenttiensa suhteen kasvavilla hyvinvointifunktioilla. Pareto-kriteerin huono puoli on, ettei se ole kovin operationaalinen. Käytännössä puhtaita Pareto-parannuksia esiintyy harvoin. Pareto-kriteeristä saa operationaalisemman seuraavalla muutoksella.

Määritelmä 15. (Anonyymi Pareto-kriteeri) Tulojakauma X on anonyymillä Pareto-kriteerillä vähintään yhtä hyvä kuin X', jos $x_{(i)} \geq x'_{(i)} \; \forall \; i = 1, 2, \ldots, n$. Tämä merkitään $X \geq_{AP} X'$.

Anonyymi Pareto-preferenssi vaatii siis järjestetyn tulojakauman jokaisen alkion olevan vähintään yhtä suuri kuin toisen. Jos X on anonyymillä Pareto-kriteerillä vähintään yhtä hyvä kuin X', niin $W(X) \geq W(X')$ kaikilla argumenttiensa suhteen symmetrisillä kasvavilla[8] hyvinvointifunktioilla.

rahasumman kuin köyhin toiseksi köyhimmälle, niin ginikerroin pienenee.

[8] Termillä kasvava tarkoitetaan tässä tutkimuksessa oikeastaan ei-vähenevää. Kasvava funktio ei siis ole välttämättä aidosti kasvava.

Määritelmä 16. (Eriarvoisuutta kaihtava symmetrinen kasvava hyvinvointifunktio) Hyvinvointifunktio W on eriarvoisuutta kaihtava kasvava symmetrinen hyvinvointifunktio, jos se on Schur-konkaavi ja kasvava argumenttiensa suhteen.

Tämä määritelmä ei sisällä kovin voimakasta tulonjaon oikeudenmukaisuuden vaatimusta: jos X' on saatu X:stä lisäämällä vain rikkaimman agentin tuloja, on $W(X') \geq W(X)$. Eriarvoisuutta kaihtavaa symmetristä kasvavaa hyvinvointifunktiota voidaan kuitenkin pitää eräänlaisena standardioletuksena hyvinvointiteoriassa.

Jos kuitenkin halutaan painottaa tasa-arvoisuutta enemmän ja hyväksyä hyvinvoinnin parannukseksi ainoastaan kaikkien agenttien tulojen yhtä suuri prosentuaalinen skaalaus, päädymme seuraavaan funktioluokkaan.

Määritelmä 17. (Skaalaparannusta preferoivat hyvinvointifunktiot) Hyvinvointifunktio W on skaalaparannusta preferoiva hyvinvointifunktio, jos se on Schur-konkaavi ja $W(X') \geq W(X)$, kun $X' = k \cdot X$ ja $k \geq 1$.

Skaalaparannus tulojakaumassa pitää tulojakaman suhteellisen eriarvoisuuden vakiona. Vielä enemmän tasa-arvoisuutta painottaa seuraa hyvinvointifunktioluokka.

Määritelmä 18. (Tasaparannusta preferoivat hyvinvointifunktiot) Hyvinvointifunktio W on tasaparannusta preferoiva hyvinvointifunktio, jos se on Schur-konkaavi ja $W(X') \geq W(X)$, kun $X' = X + k \cdot 1_n$, jossa $k \geq 0$.

Tasaparannus tulojakaumassa pitää absoluuttisen eriarvoisuuden vakiona.

3.4 Lorenz-käyriin liittyviä hyvinvointilauseita

Edellisessä osassa määriteltyjen hyvinvointifunktioluokkien yhteydet Lorenz-käyriin on ensimmäisen kerran tyhjentävästi esittänyt Shorrocks (1983). Näissä lauseissa oletetaan joko populaatioiden olevan yhtä suuria tai että ne on transformoitu Senin populaatioperiaatteen mukaisesti.

Seuraava apulause, joka esitetään ilman todistusta, on tehokas apuväline Lorenz-dominansseihin liittyvissä todistuksissa. Sen ovat alunperin esittäneet Hardy, Littlewood ja Polya.

Apulause 1. (Dasgupta et al. 1970, 182) Olkoot X ja Y vektoreita $\mathcal{R}^n$:ssä ja olkoon $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Tällöin seuraavat ehdot ovat yhtäpitäviä:

1) On olemassa bistokastinen matriisi Π, jolle pätee $Y = \Pi X$.
2) $y_{(1)} + y_{(2)} + \ldots + y_{(k)} \geq x_{(1)} + x_{(2)} + \ldots + x_{(k)}$ kaikille $k \leq n$.
3) Y voidaan saada X:stä äärellisellä sarjalla transformaatioita muotoa:[9]

$$\begin{aligned} x_{(i)}^{\alpha+1} &= x_{(i)}^{\alpha} + e^{\alpha} \leq x_{(j)}^{\alpha}, \; j > i, \\ x_{(j)}^{\alpha+1} &= x_{(j)}^{\alpha} - e^{\alpha} \geq x_{(i)}^{\alpha}, \; e^{\alpha} > 0, \\ x_{(k)}^{\alpha+1} &= x_{(k)}^{\alpha}, \; k \neq i, j \end{aligned}$$

4) Kaikilla konkaaveilla $\mathcal{R}^1$:n funktioilla U pätee $U(y_1) + U(y_2) + \ldots + U(y_n) \geq U(x_1) + U(x_2) + \ldots + U(x_n)$.

Lause 1. **(Lorenz-dominanssin hyvinvointilause)** (Dasgupta et al. 1973, Rotschild et al. 1973) Olkoot X ja X' tulojakaumia ja olkoon $\mu_X = \mu_{X'}$. Tällöin $L(p; X) \geq L(p; X')$ kaikille $p \in [0,1]$ jos ja vain jos $W(X) \geq W(X')$ kaikilla Schur-konkaaveilla hyvinvointifunktioilla W.

Todistus: Diskreetti Lorenz-dominanssi voidaan kirjoittaa muodossa $(x_{(1)} + x_{(2)} + \ldots + x_{(k)})/\mu_X \geq (x'_{(1)} + x'_{(2)} + \ldots + x'_{(k)})/\mu_{X'}$ kaikille $k \leq n$. Tällöin koska $\mu_X = \mu_{X'}$ on Lorenz-dominanssi yhtäpitävää apulauseen ehdon 2) kanssa. Tulos seuraa apulauseen ehdosta 1) ja Schur-konkaavin funktion määritelmästä.

Atkinson alunperin todisti äsken esitetyn tuloksen, mutta hän käytti jatkuvia jakaumia ja von Neumann - Morgenstern-hyötyfunktioita. Yleisin mahdollinen todistus löytyy Thistlen (1989) artikkelista, jossa hän todistaa yleisen Lorenz-dominanssin ja sitä kautta tavanomaisen Lorenz-dominanssin hyvinvointilauseet sekä diskreeteille, jatkuville että sekajakaumille. Diskreeteissä äärellisen populaation jakaumissa pysyminen on perusteltavaa todistuksien eleganttiuden takia, yleisimmissä tai jatkuvissa tilanteissa jouduttaisiin hankaliin osittaisintegraalitarkasteluihin. Lisäksi reaalimaailman tulojakaumien on luontevaa väittää olevan pohjimmiltaan diskreettejä, vaikka niitä on järkevää usein approksimoida jatkuvilla jakaumilla.

Korollaari 1. Olkoot X ja X' tulojakaumia. Tällöin $X \geq_L X'$ jos ja vain jos $I_R(X) \leq I_R(X')$ kaikilla suhteellisen eriarvoisuuden indekseillä I_R.

Korollaarin todistus esitetty liitteessä yksi.

[9] Nämä transformaatiot voidaan helposti tulkita Pigou - Dalton-tulonsiirroiksi.

Korollaarin merkitys tulojakaumien tutkimiselle on ollut tärkeä. Se on mahdollistanut eriarvoisuuden tutkimisen konsistentisti tulojakaumien välillä joiden keskiarvot ovat erisuuret. Kuitenkaan se ei sinällään ole hyvinvointilause, sillä se ei kerro mitään kahden tulojakauman välisistä hyvinvointieroista.

Ennen Shorrocksin vuoden 1983 artikkelia käytettiin empiirisissä tutkimuksissa hyvinvointidominanssin käsitteenä yhdistettyä keskiarvo - Lorenz-dominanssia, eli vaadittiin samanaikaisesti ehtojen $\mu_X \geq \mu_{X'}$ ja $X \geq_L X'$ toteutumista, jotta hyvinvointidominanssista voitaisiin puhua. Shorrocksin esittämä yleistetyn Lorenz-dominanssin käsite mahdollisti hyvinvointivertailujen tekemisen laajempien tulojakaumajoukkojen välillä.

Lause 2. **(Yleistetyn Lorenz-dominanssin hyvinvointilause)** (Shorrocks 1983) Olkoot X ja X' tulojakaumia. Tällöin $X \geq_{GL} X'$ jos ja vain jos $W(X) \geq W(X')$ kaikilla kasvavilla eriarvoisuutta kaihtavilla hyvinvointifunktioilla W.

Todistus: i) Riittävyys: Olkoon X'' sellainen, että $x''_{(i)} = x'_{(i)}$, $i = 1, \ldots n-1$; $x''_{(n)} = x'_{(n)} + n(\mu_X - \mu_{X'})$. Tällöin $W(X'') \geq W(X')$ kaikille kasvaville hyvinvointifunktioilla. Lisäksi $\mu_X = \mu_{X''}$ ja $X \geq_L X''$, joten $W(X) \geq W(X'')$ kaikilla Schur-konvekseilla hyvinvointifunktiolla. Nämä yhdistämällä saamme $W(X) \geq W(X')$ kaikilla kasvavilla Schur-konkaaveilla hyvinvointifunktioilla.

ii) Välttämättömyys: Tämä seuraa toteamalla, että funktio $W_k(X) = \sum_{i=1}^{k} x_{(i)}/n$, $k = 1, 2, \ldots, n$ kuuluu Schur-konkaaveihin kasvaviin funktioihin. Tällöin ehdosta $W_k(X) \geq W_k(X')$ $\forall k = 1, 2, \ldots, n$ seuraa suoraa diskreetin heikon yleistetyn Lorenz-dominanssin määritelmä.

Yleistetyn Lorenz-dominanssin hyvinvointimerkitystä ei ymmärretty ennen Shorrocksin artikkelia. Rotschild ja Stiglitz oikeastaan vuoden 1973 artikkelissaan esittävät yleistetyn Lorenz-dominanssin kriteerin, mutta jättävät sen ilman sen suurempaa huomiota.

Yleistetty Lorenz-dominanssi on käytännön tutkimustyössä osoittautunut usein empiirisesti lähes ekvivalentiksi anonyymin Pareto-kriteerin kanssa, joten $\geq_{GL}$-relaatio painottaa enemmän tehokkuus- kuin tasa-arvoisuuskriteeriä todellisiin tulodatoihin sovellettuna (Bishop et al. 1989b, 72).

Yleistettyä Lorenz-dominanssia vaativampi kriteeri on edellyttää, että jakauma on keskiarvoltaan vähintään yhtä suuri ja Lorenz-dominoi toista jakaumaa. Tätä kriteeriä käytettiin yleisesti empiirisenä kriteerinä hyvinvointivertailuissa tulojakaumien välillä ennen kuin yleistetty Lorenz-dominanssi otettiin käyttöön.

Lause 3. (Keskiarvo - Lorenz-dominanssin hyvinvointilause) (Shorrocks 1983) Olkoot $\mu_X \geq \mu_{X'}$ ja $X \geq_L X'$. Tämän kanssa yhtäpitävä ehto on $W(X) \geq W(X')$ kaikilla skaalaparannusta preferoivilla hyvinvointifunktioilla W.

Todistus on esitetty liitteessä yksi.

Skaalaparannuksen preferoiminen tarkoittaa vaatimusta, että tulojen kasvamisen on pidettävä suhteellinen eriarvoisuus vakioina tai pienennettävä sitä. Toisaalta jos ollaan kiinnostuneita absoluuttisesta eriarvoisuudesta, voidaan käyttää seuraava kriteeriä.

Lause 4. (Keskiarvo - absoluuttinen Lorenz-dominanssin hyvinvointilause) (Shorrocks 1983) Olkoot $\mu_X \geq \mu_{X'}$ ja $X \geq_{AL} X'$. Tämän kanssa yhtäpitävä ehto on $W(X) \geq W(X')$ kaikilla tasaparannusta preferoivilla hyvinvointifunktioilla W.

Todistus on esitetty liitteessä yksi.

Korollaari 2. $X \geq_{AL} X'$ jos ja vain jos $I_A(X) \leq I_A(X')$ kaikilla absoluuttisen eriarvoisuuden indekseillä I_A.

Todistus on esitetty liitteessä yksi.

Eri hyvinvointi- ja dominanssikriteerien väliset yhteydet on koottu seuraavaan lauseeseen.

Lause 5. (Eri hyvinvointi- ja dominanssikriteerien väliset yhteydet) Olkoot X ja Y tulojakaumia. Tällöin seuraavat väittämät pitävät paikkaansa:

1) $X \geq_{AP} Y \Rightarrow X \geq_{GL} Y$ eli ensimmäisen kertaluvun stokastinen dominanssi implikoi toisen kertaluvun stokastisen dominanssin.
2) $\mu_X \geq \mu_Y$ ja $X \geq_L Y \Rightarrow X \geq_{GL} Y$ eli keskiarvo - Lorenz-dominanssi implikoi yleistetyn Lorenz-dominanssin.
3) $\mu_X \geq \mu_Y$ ja $X \geq_{AL} Y \Rightarrow X \geq_L Y$ eli keskiarvo - absoluuttinen Lorenz-dominanssi implikoi keskiarvo - Lorenz-dominanssin.
4) Anonyymi Pareto-kriteeri ei ole riittävä eikä välttämätön ehto keskiarvo - Lorenz- tai keskiarvo - absoluuttinen Lorenz-dominanssille.

Kohdat 1)-3) ovat Shorrocksin (1983) artikkelista ja kohta 4) artikkelista Bishop et al. (1989b). Näiden todistukset on esitetty liitteessä yksi.

3.5 Tulosten laajentaminen, kun erot agenttien välillä otetaan huomioon

Tässä luvussa esitetyn teorian soveltaminen käytännön vertailuihin tulojakaumien välillä on yksinkertaista, jos oletetaan reaalimaailman koostuvan kokoelmasta agentteja, joiden välillä ei ole muita eroja, kuin heidän tulonsa. Tämä on kuitenkin erittäin voimakas oletus ja on olemassa erilaisia ratkaisuja miten realistisesti voitaisiin ottaa huomioon eroja eri ihmis- ja kotitaloustyyppien välillä. Pohjimmiltaan kyse on yhteiskunnan hyvinvointifunktion ja erilaisten eriarvoisuusindeksien symmetrisyysaksioomasta, eli siitä miten eri agentit tuloineen ovat yhteiskunnan hyvinvointifunktion argumentteina. Tässä käsitellään lyhyesti kahta lähestymistapaa, jossa toisessa pidetään kiinni yhteiskunnan hyvinvointifunktion symmetrisyydestä, mutta määritellään käsiteltävä tulokäsite uudella tavalla, kun taas toisessa jaetaan populaatio ryhmiin, joiden sisällä noudatetaan symmetrisyyttä, mutta ryhmien välillä vallitsee epäsymmetria.

3.5.1 Ekvivalenssiskaalat

Yleisimmin käytetty menetelmä tulojen vertailuissa lienee ekvivalenssiskaalojen käyttö. Ekvivalenssiskaaloja käytettäessä eri agenttien tulot muunnetaan keskenään vertailukelpoisiksi jollain funktiolla ja näitä tuloja käytetään vertailuissa kuin ne olisivat alkuperäinen tulokäsite. Tällöin siis pidetään kiinni eriarvoisuusindeksien ja yhteiskunnan hyvinvointifunktion symmetrisyysoletuksesta. Tässä lyhyessä katsauksessa käsitellään ekvivalenssiskaaloja hyvin pragmaattisella tavalla menemättä kovin syvälle niihin liittyviin kysymyksiin. Täten esimerkiksi erilaiset menetelmät ekvivalenssiskaalojen identifioimiseksi joko kulutusaineistojen pohjalta tai muilla menetelmillä sivuutetaan.

Koostukoon yhteiskunta joukosta yksilöitä, jotka ovat järjestäytyneet erilaisiksi kotitalouksiksi. Oletetaan, että kotitaloustyyppien väliset erot voidaan kuvata vektorilla h, joka sisältää tietoja kotitalouden koosta, sen jäsenten terveydestä, heidän ikänsä ja tietoja muista demografisista tekijöistä. Olkoon h^* referenssikotitalouden (esimerkiksi terve nuori yksinäinen aikuinen) tyyppivektori, johon muita kotitalouksia halutaan verrata. Olkoot lisäksi p vallitsevien hintojen vektori, u vertailun kohteena oleva tavoitehyötytaso ja $C(u, p, h)$ niin sanottu menofunktio, joka kertoo kotitalouden kokonaisminimikustannukset, jolla tietyntyyppisessä kotitaloudessa elävät henkilöt[10] pääsevät tavoitehyötytasolle u vallitsevilla hinnoilla. Tällöin

[10] Kriittinen oletus tämän analyysin kannalta on, että kotitalouksien sisällä resurssit jaetaan siten, että

ekvivalenssiskaala tai ekvivalenttien aikuisten lukumäärä kotitaloudessa h on

$$(4) \qquad M(u,p,h,h^*) = \frac{C(u,p,h)}{C(u,p,h^*)},$$

eli se kertoo kuinka moninkertaiset tulot vallitsevilla hinnoilla pitäisi kotitaloudella tyyppiä h olla, jotta sen jäsenet pääsisivät samalle hyötytasolle kuin kotitalouden tyyppiä h^* jäsenet. Kotitalouden jäsenten ekvivalentit tulot x saadaan jakamalla kotitalouden todelliset tulot y ekvivalenssiskaalalla M. Tällä menettelyllä pyritään ottamaan huomioon suuremmissa kotitalouksissa elävien skaalaedut esimerkiksi asumisessa ja kestokulutushyödykkeiden hankinnassa. Käsitykset näiden skaalaetujen koosta muodostavat suurimmat erot eri ekvivalenssiskaalojen välille. (Coulter et al. 1992, 84.)

Hyvinvointivertailuja tehdessä tehdään ekvivalenssiskaalasta yleensä hyvin rajoittavia oletuksia. Ensinnäkin saman ekvivalenssiskaalan oletetaan olevan oikea hintatasosta riippumatta. Tulojen oletetaan siis tulevan hintojen suhteen vertailukelpoisiksi hintaindeksillä deflatoinnilla (tai kansainvälisissä vertailuissa reaalisella valuutakurssilla jakamisella). Niin on myös tehty tämän tutkimuksen empiirisessä osassa, vaikka tietenkään tämä ei ole talousteorian kannalta täysin ongelmaton ratkaisu (erilaisilla suhteellisten hintojen muutoksilla voi olla hyvinkin erilaisia hyvinvointivaikutuksia erityyppisiin kotitalouksiin).

Lisäksi yleensä oletetaan hyvinvointivertailujen yhteydessä ekvivalenssiskaala riippumattomaksi hyötytasosta, jolla suhde M määritellään. Tähänkin perustelu on enemmän pragmaattinen kuin teoreettinen, tulojen vertailukelpoiseksi tekemisestä halutaan tehdä yksinkertainen prosessi. Lisäksi kotitalouksien ominaisuuksista kertovaan vektoriin h yleensä otetaan mukaan vain kotitalouden koko- ja jäsenten ikätietoja. Tällöin päästään ekvivalenssiskaalaan

$$(5) \qquad \bar{M} = \bar{M}(h,h^*),$$

jollaisen kaavan 4 menofunktioajattelun perusteella päästäisiin vain erittäin voimakkaiden menofunktion (ja siten hyötyfunktion) rakennettava koskevien oletusten avulla.

Yleistä konsensusta miten ekvivalenssiskaalat pitäisi muodostaa ei ole (Coulter et al. 1992, 119). Usein ekvivalenssiskaalat, kuten esimerkiksi tässä tutkimuksessa käytetty OECD-skaala, muodostetaan jonkin järkevän kuuloisen "nyrkkisäännön" avulla. Luvun alkuosan perusteella tällöin voidaan ekvivalenssiskaaloja pitää tutkijan käsityksenä järkevästä tai oikeudenmukaisesta vertailuista erilaisten kotitaloustyyppien välillä. Tällöin ollaan jo aika kaukana formaalista menofunktioajattelusta. Käytännössä kuitenkin käytetyt ekvivalenssiskaalat riippuvat

kaikki kotitalouden jäsenet pääsevät samalle hyötytasolle.

lähinnä kotitalouden koosta.[11] Ongelmaksi jää silti, että kuinka perusteltua on käyttää ekvivalentteja tuloja (eli ekvivalenssiskaalalla jaettuja tuloja) vertailun lähtökohtana niin kuin ne olisivat tavanomaisia tuloja ja kuinka paljon tulokset riippuvat käytetystä ekvivalenssiskaalasta.

Tämän luvun alkuosan teoreettiset tarkastelut eivät ole riippuvaisia siitä, mikä niissä käsitelty tulokäsite itse asiassa on yhtä huomattavaa poikkeusta lukuunottamatta. Pigou - Dalton-aksiooma ei sellaisenaan ole välttämättä järkevä ekvivalentteja tuloja käytettäessä. Ongelman ydin on tulokäsitteessä, sillä vaikka tuloja mitattaisiin ekvivalenteissa tuloissa, niin tulonsiirrot tapahtuvat todellisessa rahassa. Tällöin tulonsiirto voi tulonjaon lisäksi muuttaa ekvivalenttia kokonaistuloa.[12] Tätä ilmiötä hyväksikäyttäen voidaan johtaa paradokseja (tehdä tulonsiirto köyhältä rikkaalle ja samalla pienentää eriarvoisuutta), mutta tämän ilmiön käytännön merkitys lienee vähäinen. (Glewwe, 1991.)

Suurempi ongelma lienee kuitenkin tulosten riippuvuus käytettävästä ekvivalenssiskaalasta. Luonnollisesti olisi toivottavaa, etteivät saadut tulokset riippuisi ratkaisevasti juuri tietystä valitusta ekvivalenssiskaalasta. Mitään yleisiä vahvoja tuloksia Lorenz-käyrien suhteen ei ole pystytty esittämään, koska ekvivalenssiskaalan muutos vaikuttaa kahta eri kautta erilaisiin Lorenz-käyriin: ekvivalenttien tulojen arvoihin ja eri yksilöiden suhteelliseen sijaintiin järjestetyssä jakaumassa. Tästä syystä Coulter et al. (1992, 120) suosittavat tulosten riippuvuuden käytetystä ekvivalenssiskaalasta tutkimista tekemällä analyysit useammalla kuin yhdellä ekvivalenssiskaalalla.

[11] Buhman et al. (1988) vertailevat erilaisia ekvivalenssiskaaloja ja niiden vaikutuksia eriarvoisuus- ja köyhyysvertailuissa. Vaikka eri ekvivalenssiskaalat voivat ottaa huomioon useita tekijöitä kotitalouden rakenteesta, niin reaalimaailman aineistoon käytettyinä kaikki ekvivalenssiskaalat olivat "lähellä" pelkkää kokotietoa hyväksi käyttäviä skaaloja: ekvivalenttien aikuisten lukumäärän logaritmin korrelaatio kotitalouden jäsenmäärän logaritmin kanssa vaihteli kansainvälisessä aineistossa ekvivalenssiskaalasta riippuen välillä $[0.95, 1.00]$.

[12] Ajatellaan tulonsiirto kotitaloudelta A kotitaloudelle B. Olkoon kotitalous A yhden hengen muodostama ja skaalatekijä siis yksi. Olkoon kotitaloudessa B viisi jäsentä ja skaalatekijä kolme. Siirretään summa sata markkaa A:lta B:lle. Tällöin ekvivalenttien kokonaistulojen muutos on $-100 + 5 \cdot 100/3 = 66\frac{2}{3}$. Tämä ongelma ei ole riippuvainen siitä, että tässä esimerkissä ekvivalentteja tuloja on painotettu kotitalouden jäsenmäärillä. Myös puhdas per kotitalous tarkastelu johtaisi vastaavanlaisiin tulokseen. Ainoastaan, jos tarkasteluissa ekvivalenttien tulojen painoina käytettäisiin ekvivalentteja aikuisia, ei tätä ilmiötä esiintyisi. Näin ei kuitenkaan yleensä tehdä, koska yleensä ollaan kiinnostuneita yksilöiden eikä siis kotitalouksien tai ekvivalenttien aikuisten hyvinvoinnista.

3.5.2 Atkinson-Bourguignon lähestymistapa

Toinen lähestymistapa, joka on huomattavasti uudempi kuin ekvivalenssiskaalojen käyttö, on ottaa hyvinvointivertailuissa suoraan huomioon kotitalouksien erilaisuus. Tämä menettely perustuu Atkinsonin ja Bourguignonin vuonna 1987 esittämään menetelmään, jossa luovutaan yhteiskunnan hyvinvointifunktion symmetrisyysaksioomasta. Tällöin on oleellista, että kotitaloudet pystytään jakamaan ryhmiin, jotka ovat keskenään vertailukelpoisia. Tällainen jako voisi olla esimerkiksi jako yksinäisiin aikuisiin, lapsettomiin pareihin, erikokoisiin lapsiperheisiin ja muihin kotitalouksiin. Tämän jaon lisäksi tarvitaan käsitys eri tyyppisten kotitalouksien asemasta "tarvehierarkiassa". Tarvehierarkkia voisi olla seuraavanlainen: kaksilapsiset perheet tarvitsevat enemmän rahaa kuin yksilapsiset perheet samalle hyötytasolle päästäkseen, jotka taas tarvitsevat enemmän kuin lapsettomat parit jne. Oletetaan, että tällaisesta tarvehierarkiasta on olemassa käsitys ja oletetaan, että yhteiskunnan hyvinvointifunktio on muotoa

$$W = \sum_{k=1}^{l} p_k \int_0^{y_{max}} u_k(y) f_k(y) dy, \tag{6}$$

missä l on ryhmien lukumäärä, p_k on k:nneksi eniten tarpeita omaavan ryhmän kotitalouksien osuus kaikista kotitalouksista,[13] u_k yhteiskunnan hyötyfunktio heidän tuloilleen, y_{max} on maksimitulo, jota yksikään kotitalous missään luokassa ei ylitä, ja $f_k(y)$ tulomuuttujan tiheysfunktio ryhmässä k. Tässä, toisin kuin luvun alkuosassa, oletetaan yhteiskunnan hyötyfunktion olevan additiivisesti separoituva ja lisäksi oletetaan tulon Y olevan jatkuva satunnaismuuttuja.

Olkoon nyt kotitalouksien osuudet p_k eri ryhmissä samat tarkasteltavissa tulojakaumissa. Tällöin Atkinson ja Bourguignon (1987) osoittavat, että tekemällä sopivia oletuksia hyötyfunktioiden u_k ensimmäisistä ja toisista derivaatoista sekä näiden derivaattojen suuruusjärjestyksistä (siis ryhmien välisestä tarvehierarkiasta) eri ryhmille, päädytään anonyymin Pareto-kriiterin ja yleistettyn Lorenz-kriteerin tapaisiin dominanssikriteereihin, jotka eivät muuten riipu hyötyfunktioiden u_k spesifikaatiosta.

Triviaalia on tietenkin todeta kaavasta 6, että jos joko anonyymi Pareto-kriteeri tai yleistetty Lorenz-kriteeri toteutuu jokaisella ryhmällä erikseen, niin myös vastaavan hyvinvointidominanssin täytyy toteutua. Tarvehierarkkiaa koskevia oletuksia käyttäen Atkinson ja Bourguig-

[13] Jenkins ja Lambert (1993, 339) huomauttavat, etteivät tässä esitetyt tulokset ole riippuvaisia painojen p_k tulkinnasta. Niiden ei siis välttämättä tarvitse olla mitään laskennallisia osuuksia, niin kuin Atkinsonin ja Bourguignonin esityksessä, vaan ne voivat myös esittää "sosiaalisia painoarvoja" eri ryhmille. Jenkins ja Lambert kuitenkin pitävät relevanteimpana tapausta, jossa p_k-painot todella ovat eri ryhmien osuuksia kaikista kotitalouksista.

non kuitenkin osoittavat, että riittävä ehto hyvinvointidominanssille hyvin laajalle funktioluokalle on, että dominanssirelaatiot toteutuvat "järjestetysti." Järjestetyllä dominanssirelaatiolla tarkoitetaan sitä, että dominanssi toteutuu kaikissa vertailuissa kun ensiksi tarkastellaan kaikkein eniten tarpeita omaavaa kotitalousryhmää (esim. monilapsiset perheet), sen jälkeen tarkastellaan toiseksi eniten ja eniten tarpeita omaavan ryhmän yhdistelmää jne. Hyötyfunktion derivaattoja koskevien oletusten avulla saadaan tällä menettelyllä muodostettua vastaavia hyvinvointitulkintoja järjestetyille dominanssirelaatioille kuin anonyymin Pareto- ja yleistetyn Lorenz-kriteerin yhteydessä. Jenkins ja Lambert (1993) laajentavat Atkinsonin ja Bourguignonin tulokset myös tilanteeseen, jossa osuuksien p_k ei tarvitse olla samoja vertailtavissa jakaumissa.

Merkittävä etu Atkinson-Bourguignon lähestymistavan mukaisissa tarkasteluissa on se, että kotitalouksien välisestä hyötytasojen vertailtavuudesta tehdään lievempiä oletuksia kuin ekvivalenssiskaaloja käytettäessä. Tarvehierarkkia määrää ainoastaan järjestyksen eri kotitaloustyyppien välille (ordinaalinen vertailu), kun ekvivalenssiskaaloja käytettäessä ainakin periaatteessa oletetaan, että eri kotitaloustyyppien menofunktioiden erot tunnetaan (kardinaalinen vertailu).

Vaikka Atkinsonin ja Bourguignonin sekä Jenkinsin ja Lambertin tulokset ovat teoreettisesti mielenkiintoisia, niitä ei ole hyödynnetty tässä tutkimuksessa eikä tämä esitys niistä ole missään muodossa kattava. Niiden sijasta on käytetty traditionaalista ekvivalenssiskaalamenettelyä. Tähän on ollut syynä lähinnä halu pitää luvussa neljä esitettävä tilastollinen aspekti voimakkaasti mukana empiirisissä tarkasteluissa. Tilastollisen analyysin validiteetti heikkenisi ja jo sellaisenaan monimutkaisista tilastollisista tarkasteluista tulisi entistä monimutkaisempia, jos järjestettyä dominanssia ruvettaisiin testaamaan aineistosta. Otoksen jakaminen useaan erilliseen ryhmään monikertaistaisi tilastollisessa tarkastelussa vaadittavien tarkastelupisteiden määrän (katso luku 4) ja osa ryhmistä jäisi todennäköisesti niin pieniksi joissain otoksissa, että asymptoottisten tulosten käyttäminen olisi vähintäänkin epäilyttävää. Tällöin todennäköisesti joitain ryhmiä jouduttaisiin yhdistämään, jotta analyysit pystyttäisiin luotettavasti suorittamaan.[14] Tämä yhdistäminen jouduttaisiin toteuttamaan käyttäen ekvivalenssiskaaloja, joten ekvivalenssiskaalojen käyttöä tuskin pystyttäisiin kokonaan välttämään. Lisäksi tarvehierarkkian rakentaminen kaikille kotitaloustyypeille ei välttämättä ole yksinkertaista ja käytännön tutkimustilanteessa jouduttaisiin kokeilemaan tulosten robustisuutta usealle vaih-

[14] Esimerkiksi vuoden 1990 kotitaloustiedusteluaineistossa on vain 101 yhden aikuisen ja yhden lapsen kotitaloutta ja vuoden 1971 aineistossa niitä oli 36. Näin pienten osaotosten perusteelta tehtävä Lorenz-dominansseja koskeva asymptoottinen tilastollinen päättely useassa tarkastelupisteessä olisi erittäin arveluttavaa.

toehtoiselle hierarkkialle.[15] Tästä syystä tämä tutkimus käyttää vähemmän sofistikoitunutta, mutta helpommin toteuttavaa, ekvivalenssiskaalalähestymistapaa vaikka talousteoreettisesti se ei olekaan parempi menettely. Lisäperusteluna ekvivalenssiskaalojen käytölle on myös se, että Atkinsonin ja Bourguignonin tulokset koskevat ainoastaan hyvinvointidominanssitarkasteluja, jotka ovat läheistä sukua anonyymille Pareto- ja yleistetylle Lorenz-kriteerille, mutta eivät esimerkiksi eriarvoisuustarkasteluja.

[15] Jenkins ja Lambert (1993) ratkaisevat nämä ongelmat tarkastelemalla empiirisessä osassaan vain kotitalouksia, joissa on kaksi aikuista ja vaihteleva määrä lapsia. Tällöin tarvehierarkkiasta on helppo päästä yksimielisyyteen: kotitaloustyypin vaatimat resurssit ovat lapsiluvun kasvava funktio eikä ekvivalenssiskaaloja jouduta käyttämään.

4 Lorenz-käyriin liittyvä tilastollinen päättely

4.1 Tuloaineistojen tilastollisista analyysimuodoista

Maasoumi (1995, 15) jakaa tuloaineistojen tilastollisen tarkastelun kolmeen vaihtoehtoiseen luokkaan. Ensimmäisenä vaihtoehtona on yksinkertaisesti estimoida tietyt mielenkiintotunnusluvut, kuten eriarvoisuusindeksit sekä näiden varianssit aineiston perusteella. Toinen vaihtoehto, joka on myös tämän tutkimuksen tavoitteena, on tutkia muutoksia tulojakaumassa ja tulojakaumien välisiä dominanssirelaatioita. Kolmas tarkasteluvaihtoehto on yrittää kuvata tulojakaumaa jollain parametrisella tilastollisella jakaumaperheellä. Kolmas tavoite on tavallaan kaikkein vaativin ja kunnianhimoisin, sillä jos siinä onnistuu riittävällä tarkkuudella, niin ainakin periaatteessa kahden ensiksi mainitun analyysimuodon toteuttaminen on mahdollista parametrisen mallin perusteella.

Kahta ensimmäistä ongelmaa voi lähestyä sekä parametrisesti että ei-parametrisesti. Ei-parametrisen vaihtoehdon käyttö vaatii usein laajan mikrotason aineiston käyttöä, kun taas parametrisen mallin sovittaminen on mahdollista aggregoituunkin dataan. Menetelmän valinnassa onkin ratkaisevaa, mitä aineistosta halutaan tutkia ja mikä on aineiston aggregoinnin taso. Silloin, kun aineisto on käytettävissä vain aggregoituna, niin parametrisen mallin sovittaminen jää usein ainoaksi vaihtoehdoksi. Jos taas tutkijalla on käytettävissään koko mikrotason aineisto, ei hänellä ole ainakaan mitään pakottavaa syytä käyttää parametrisia malleja.

Parametrisessä menetelmässä spesifioidaan jokin tietty todennäköisyysmalli, josta tuloaineiston uskotaan olevan otos. Tämän todennäköisyysmallin parametrit estimoidaan jollain sopivalla estimointimenetelmällä ja näiden estimoitujen parametriarvojen funktiona lasketaan halutut suuret (esim. köyhyys- ja eriarvoisuusindeksit, Lorenz-käyrät jne). Tulojakaumien parametrisissä malleissa on lähdetty alunperin yksi- tai kaksiparametrisistä malleista (esim. lognormaali-, Pareto- ja eksponentiaalijakauma) edeten kolmi- ja neliparametrisiin malleihin (esim. Singh-Maddala- sekä yleistetyt gamma- ja betajakaumat) (McDonald 1984). McDonaldin mukaan erityisesti toisen tyypin yleistetty betajakauma kuvasi Yhdysvaltain tuloaineistoa hyvin. Sen ovat alunperin tulojakaumien yhteydessä esittäneet vuonna 1972 Vartia ja Vartia (1980) nimellä yleistetty skaalattu F-jakauma.

Parametristen mallien käyttöön liittyy kuitenkin ongelmia. Useiden eriarvoisuusindeksiperheiden sekä Lorenz-käyrien estimointiin on kehitetty ei-parametrisia menetelmiä, jotka mahdollistavat asymptoottisen tilastollisen päättelyn. Näitä käyttämällä vältetään se mahdollinen harha, joka seuraa parametrisen mallin sovittamisesta tuloaineistoon. Usein tuloaineistot ovat hyvin suuria (tuhansista havannoista kymmeniin tuhansiin havaintoihin), joten tällaisen aineiston kaiken oleellisen vaihtelun kuvaaminen neljällä tyhjentävällä tunnusluvulla ei välttämättä ole kovin uskottavaa. McDonaldin (1984, 659) tutkimuksessa, joka käsittelee erilaisten parametristen tiheysfunktioiden sovittamista Yhdysvaltojen luokiteltuun tuloaineistoon, onkin mielenkiintoista huomata, että χ^2-yhteensopivustesti hylkää kaikki kokeillut vähäparametriset malliperheet. Tämä ei tarkoita sitä, etteikö parametrinen malliperhe voisi olla riittävän hyvä aproksimaatio todellisuudelle monessa tilanteessa, vaan sitä, että parametrista mallia käyttäessä on aina muistettava, että mallispesifikaatiosta seuraa aina yksi mahdollinen lisävirhelähde tuloksiin.

Lorenz-käyrien yhteydessä parametristen tilastollisten jakaumaperhemallien lisäksi on kehitetty ehdotuksia Lorenz-käyrän mahdolliseksi funktiomuodoksi. Vaikka Lorenz-käyrän tunteminen on keskiarvon tuntemista vaille sama asia kuin jakauman tunteminen, niin Lorenz-käyrän parametroinnissa on kuitenkin mallintamisfilosofia ollut lähempänä epälineaarista regressioanalyysiä kuin jakaumien parametrointia. Erilaisia funktiomuotoehdotuksia ja näiden kykyä estimoida perusjoukon parametreja ryhmitellystä aineistosta ovat käsitelleet Schader & Schmid (1994) ja Suoniemi (1994). Schader & Schmid tutkimuksessaan vertaavat 13:n eri funktiomuodon kykyä estimoida ginikertoimen arvoa 16:sta eri vuosilta olevasta saksalaisesta ryhmitellystä poikkileikkausaineistosta. Näitä he vertaavat Gaswirthin (1972) ei-parametrisiin ylä- ja alarajoihin ginikertoimelle, joka on estimoitu ryhmitellystä aineistosta. Heidän tuloksensa suosittavat ainakin ginikertoimen estimointiin ei-parametrisen menetelmän käyttöä, sillä ainoastaan kaksi parametrista käyräperhettä tuotti aina ginikertoimen piste-estimaatin, joka oli ei-parametristen ylä- ja alarajojen välissä. Täten he toteavat, että ei-parametrisen menetelmän käyttö voi olla suositeltavaa myös ryhmitellyn aineiston ollessa kyseessä. Vastaavanlaisia tuloksia saa myös Suoniemi tutkimuksessaan, jossa hän vertaillessaan ei-parametrisen polynomiaalisen lokaalisen regression suorituskykyä Lorenz-käyrän estimoinnissa lokaalisesti sovitettujen parametristen mallien suorituskykyyn, toteaa ei-parametrisen vaihtoehdon olevan suorituskyvyltään lokaalisesti[16] yhtä hyvä parametristen mallien kanssa.

[16]Tämä tulos koskee siis vain Lorenz-käyrän lokaalia sovitetta. Jotta jokin funktiomuoto olisi globaalisesti oikea, täytyisi sen lokaalisesti sovitettujen parametrien olla kohtuullisen vakioita koko tarkasteluvälillä. Näin ei kuitenkaan ollut Suoniemen tutkimuksessa minkään lokaalisen parametrisen sovitteen yhteydessä.

Jottei lukijalle jäisi liian kriittistä kuvaa parametristen menetelmien käytöstä tuloaineistojen yhteydessä, lienee muutama niiden hyvä ominaisuus syytä tuoda esille. Ensinnäkin parametrisen mallin parametrien muutokset voivat esimerkiksi aikasarja-aineistossa antaa tiivistettyä informaatiota tulojakauman muodossa tapahtuneista muutoksista (Schader & Schmid 1994, 368). Parametrisen mallin estimointi suurimman uskottavuuden menetelmällä mahdollistaa yleensä asymptoottisen tilastollisen päättelyn eriarvoisuusindeksien tai muiden mielenkiintosuureiden yhteydessä, kun taas ei-parametrisille estimaattoreille ei aina tälläistä valmista teoriaa ole löydettävissä (Maasoumi 1995, 16). Lisäksi eräiden parametristen tilastollisten mallien ympärille on rakennettu talousteoreettisia malleja, jotka johtavat juuri näihin tulojakaumiin (Maasoumi 1995, 27).

4.2 Eräitä huomioita Lorenz-käyriin liittyvästä tilastollisesta päättelystä

Lorenz-käyriin liittyvässä tilastollisessa analyysissä on kyse siitä, että halutaan erotella todelliset merkitsevät erot Lorenz-käyrien välillä niistä eroista, jotka voisivat olla otantaan liityvän satunnaisvaihtelun aiheuttamia. Asetelma on siis sama kuin yleensäkin tilastollisessa analyysissä. Satunnaisvaihtelun rajoihin mahtuvat Lorenz-käyrien erot katsotaan ei-merkitseviksi ja jätetään siksi huomioimatta. Tällä menettelyllä on huomattava merkitys luvun viisi analyyseissä, usein erilaisten Lorenz-käyrien leikkaamisen voidaan katsoa olevan otantavaihtelun rajojen sisällä ja siten ei-merkitsevää.

Lorenz-käyriin liittyvän tilastollisen päättelyn mukaan tuominen taloudelliseen tutkimukseen on oikeastaan alkanut vasta 1980-luvulla. Edelleenkin tehdään paljon tutkimuksia, joissa tilastollinen analyysi sivuutetaan kokonaan. Näin tehdään esimerkiksi tuoreessa OECD:n julkaisemassa tutkimuksessa (Atkinson et al. 1995). Formaalin tilastollisen analyysin sijasta he käyttävät nyrkkisääntöä "jos Lorenz-käyrien ero ei ole vähintään yhtä prosenttiyksikköä, niin mitään eroa ei esiinny" (Ibid, 87). Tätä menettelyä voi kritisoida sen alkeellisuuden takia, mutta aivan ylikriittinen sitä kohtaan ei kannata olla. Atkinson et al.:n tutkimuksessa otoskoot ovat niin suuria, että yhden prosenttiyksikön erot ovat varmasti myös tilastollisesti merkitseviä. Toisaalta tämän tutkimuksen viidennessä luvussa ehkä liikaakin pohditaan sitä, että onko jokin itseisarvollisesti mitättömän pieni ero tilastollisesti merkitsevä. Tässä kuitenkin otetaan käyttöön ajattelutapa, jonka mukaan ensiksi tutkitaan esiintyykö tilastollisesti merkitseviä eroja ja tämän informaation valossa pohditaan, onko näillä eroilla mitään käytännön merkitystä.

Tyhjentävää ja lopullista teoriaa erilaisiin Lorenz-käyriin liittyvään tilastolliseen päättelyyn ei tässä esitetä. Ongelmaksi muodostuu Lorenz-käyrien funktioluonne, ne ovat jatkuvia kuvauksia väliltä $[0,1]$. Tässä esitettävä tilastollinen teoria kuitenkin käsittelee Lorenz-käyrien otantaominaisuuksia ennaltamääritellyissä pisteissä. Ajattelutapana on, että tarkastelemalla jatkuvaa funktiota pisteittäin riittävän tiheällä seulalla päädytään hyvin lähelle jatkuvan funktion tarkastelua. Riittävän tiheänä seulana luvun viisi empiirisessä esimerkissä pidetään tarkastelua viiden prosentin välein. Näin tarkastelupisteitä on siis kaksikymmentä.[17]

4.3 Estimaattorien asymptoottinen teoria

4.3.1 Yleistetty Lorenz-käyrä

Luvussa kaksi on esitetty yleistetyn Lorenz-käyrän määritelmä. Siinä se on esitetty perusjoukolle soveltuvassa muodossa, mutta sen muuttaminen otossuureeksi on suoraviivaista. Tässä luvussa noudatetaan hiukan lyhyempää notaatiota kuin luvuissa kaksi ja kolme. Olkoot G_i yleistetty Lorenz-ordinaatta pisteessä p_i ja G yleistettyjen Lorenz-ordinaattojen vektori ennaltamääritellyissä tarkastelupisteissä. Olkoon lisäksi käytettävissä satunnaisotos $x_1, x_2 \ldots x_n$ taustalla olevasta jakaumasta. Tässä esitettävä teoria perustuu oleellisesti järjestystunnuslukujen eli järjestetyn otoksen $x_{(1)}, x_{(2)} \ldots x_{(n)}$ teoriaan. Liitteessä kaksi todistetaan, että yleisten oletusten vallitessa järjestystunnuslukuihin perustuva estimaattori

$$\hat{G}_i = \frac{1}{n} \sum_{i=1}^{[np_i]} x_{(i)}, \tag{7}$$

missä $[np_i]$ tarkoittaa lukua np_i lähinnä olevaa pienempää kokonaislukua, on asymptoottisesti tarkentuva ja harhaton G_i:n estimaattori, jolla on asymptoottisesti normaalinen jakauma. Lisäksi liitteessä kaksi todistetaan, että vektoriestimaattorin $\hat{G}$ asymptoottinen jakauma on multinormaalinen. Vektoriestimaattorin $\hat{G}$ asymptoottinen kovarianssimatriisi [18] Σ voidaan myös estimoida tarkentuvasti. Liitteessä kaksi on esitetty $\hat{G}$ kovarianssimatriisin rakenne. Jatkon kannalta on oleellista, että käytettävissä on estimaattorivektori $\hat{G}$, joka sisältää yleis-

[17] Tiheämpi tarkasteluseula ei olisi muuttanut lopputuloksia. Luvun viisi tilastolliset analyysit tehtiin kokeeksi myös sadan vertailupisteen avulla, mutta yksikään analyysin kvalitatiivinen johtopäätös ei olisi muuttunut yhden prosentin riskitasoa käytettäessä. Seulan rajaton tihentäminen ei ole järkevää, koska se tekee tarkasteluista laskennallisesti raskaita ja toisaalta asymptoottisten tulosten käyttäminen mielivaltaisen tiheällä seulalla ei ole validia, koska asymptoottisen teorian voidaan silloin katsoa soveltuvan huonosti lähekkäisten pisteiden välisten riippuvuuksien kuvaamiseen.

[18] Asymptoottisella kovarianssimatriisilla tarkoitetaan tässä suureen $\sqrt{n}(\hat{G} - G)$ kovarianssimatriisia. Siis asymptoottiseen teoriaan vedottaessa on $\hat{G}$ likimain multinormaalinen parametrein G ja Σ/n.

tettyjen Lorenz-käyrien estimaattorit ennaltavalituissa pisteissä p_i[19] sekä estimaattorivektorin asymptoottinen kovarianssimatriisi Σ tai tämän tarkentuva estimaattori $\hat{\Sigma}$. Huomionarvoista on, että $\hat{G}_k = \hat{\mu}$, eli viimeisen yleistetyn Lorenz-ordinaatan estimaattori on tavanomainen otoskeskiarvo.

4.3.2 Tavanomainen Lorenz-käyrä

Lorenz-ordinaattojen asymptoottinen jakauma saadaan yleistettyjen Lorenz-ordinaattojen asymptoottisesta jakaumasta käyttämällä niin sanottua Raon (1972, 385-386) δ-menetelmää, jolla voidaan johtaa asymptoottisesti normaalisten suureiden säännöllisten funktioiden asymptoottinen jakauma. Olkoon L^* vektori, joka sisältää Lorenz-ordinaatat pisteissä $p_i, i < k$ ja olkoon $L_k^* = \mu$. Otetaan käyttöön kuvaus $h : \mathcal{R}^k \to \mathcal{R}^k$ ja määritellään $h_i(x) = x_i/x_k$, kun $i < k$ ja $h_k(x) = x_k$. Tällöin $h(G) = L^*$. Lisäksi, koska $h(G)$ on differentioituva funktio, sillä tulojakaumasta tehdyn oletuksen mukaan $G_k = \mu > 0$, niin δ-menetelmän soveltaminen antaa tulokseksi että estimaattori $\hat{L}^* = h(\hat{G})$ on asymptoottisesti multinormaalinen odotusarvovektorilla $h(G) = L^*$ ja kovarianssimatriisilla $H\Sigma H'$, missä matriisi

$$H = \begin{bmatrix} 1/\mu & 0 & 0 & \dots & 0 & -G_1/\mu^2 \\ 0 & 1/\mu & 0 & \dots & 0 & -G_2/\mu^2 \\ 0 & 0 & 1/\mu & & 0 & -G_3/\mu^2 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1/\mu & -G_{k-1}/\mu^2 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

on kuvauksen h derivaatta- eli Jacobin matriisi laskettuna pisteessä G. Matriisi $H\Sigma H'$ voidaan tilastollisen päättelyn yhteydessä korvata tarkentuvalla estimaattorillaan. Koska ohessa on johdettu Lorenz-ordinaattojen estimaattorien ja otoskeskiarvon yhteinen asymptoottinen multinormaalijakauma, niin pelkkien Lorenz-ordinaattojen estimaattorien asymptoottisen multinormaalijakauman parametrit saadaan ylläolevasta jättämällä pois vektorista L^* sen k:nnes alkio sekä kovarianssimatriisista k:nnes rivi ja sarake. (Beach & Davidson 1983.)

[19] Tässä $p_i > p_{i-1}$, $p_1 > 0$ ja $p_k = 1$, $k = dim(\hat{G})$. Yleensä vektori p sisältää jonkun tasavälisen jonon, kuten $p = [0.1, 0.2, 0.3, \dots, 0.9, 1.0]'$ tai kuten luvussa viisi $p = [0.05, 0.10, 0.15, \dots, 0.95, 1.0]'$.

4.3.3 Absoluuttinen Lorenz-käyrä

Absoluuttisten Lorenz-ordinaattojen estimaattorien ja otoskeskiarvon yhteisjakauma saadaan helposti yleistettyjen Lorenz-ordinaattojen jakaumasta, koska niiden välillä vallitsee lineaarinen riippuvuus. Olkoot A^* todellisten absoluuttisten Lorenz-ordinaattojen vektori, jonka viimeinen alkio (joka on absoluuttisen Lorenz-käyrän määritelmän mukaan 0) on korvattu populaatiokeskiarvolla ja G todellisten yleistettyjen Lorenz-ordinaattojen vektori. Tällöin $A^* = BG$, missä matriisi

$$B = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & -p_1 \\ 0 & 1 & 0 & \dots & 0 & -p_2 \\ 0 & 0 & 1 & & 0 & -p_3 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & -p_{k-1} \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}.$$

Tällöin lineaarimuunnosten teoriasta tiedetään, että $\hat{A}^* = B\hat{G}$ on asymptoottisesti normaalinen estimaattori A^*:lle. Estimaattorin $\hat{A}^*$ asymptoottinen kovarianssimatriisi on $B\Sigma B'$. Absoluuttisten Lorenz-ordinaattojen estimaattorien asymptoottisen jakauman parametrit saadaan jättämällä pois k:nnes rivi vektorista A^* sekä k:nnes sarake ja rivi matriisista $B\Sigma B'$. (Bishop et al. 1994.)

4.3.4 Ryhmäkeskiarvojen asymptoottinen jakauma

Anonyymin Pareto-kriteerin mukaan olisi mielenkiintoista tarkastella kahden jakauman fraktiilikuvaajien välisiä dominanssisuhteita. Fraktiilifunktion sijaan tässä tutkimuksessa on kuitenkin tarkasteltu ryhmäkeskiarvojen $M_i = E[X|p_{i-1} \leq F(X) \leq p_i]$ jakaumia, jotka kertovat sen, mikä on esimerkiksi tiettyjen desiiliryhmän tulojen keskiarvo. Ryhmäkeskiarvodominanssi on välttämätön, muttei riittävä ehto fraktiilifunktioiden dominanssille, mutta kun ryhmäkeskiarvoja verrataan riittävän tiheällä seulalla (viiden prosentin välein luvussa viisi), niin voidaan ryhmäkeskiarvodominanssitarkasteluita pitää hyvänä korvikkeena fraktiilifunktion dominanssitarkasteluille. Syy miksi tässä tutkimuksessa tarkastellaan ryhmäkeskiarvoja on niiden suora yhteys yleistettyyn Lorenz-käyrään. Tätä menettelyä ovat alunperin esittäneet Bishop et al. (1989b).

Kuten absoluuttisten Lorenz-käyrien tapauksessa, on vektorien G ja M välillä lineaarinen riippuvuus. Olkoon lisäksi nyt vektori p sellainen, että $p_i = i\Delta p$, $i = 1, 2, 3, \dots, k$. Tällöin

M=CG, missä matriisi

$$C = \frac{1}{\Delta p}\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & & -1 & 1 & 0 \\ 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

Estimaattori $\hat{M} = C\hat{G}$ on asymptoottisesti normaalinen odotusarvovektorilla M ja kovarianssimatriisilla $C\Sigma C'$.

4.3.5 Ryhmäosuuksien asymptoottinen jakauma

Vaikka talousteorian mielessa ryhmäosuuksien tarkastelu ei ole minkään tässä esityksessä käsitellyn dominanssirelaation mielessä mielekästä, niin liitteessä kolme on myös esitetty estimaatit vinttiiliryhmien osuuksille kokonaistulosta. Näiden asymptoottinen jakauma saadan helposti Lorenz-ordinaattojen jakaumasta, koska ryhmän k osuus kokonaistulosta on $L_k - L_{k-1}$, kun määritellään $L_0 = 0$, eli kyseessä on Lorenz-ordinaattojen lineaarimuunnos. Tätä tosiasiaa on käytetty hyväksi, kun liiteessä kolme esitetään ryhmäosuuksien asymptoottiset keskihajonnat. (Beach & Davidson 1983.)

4.4 Lorenz-käyriin liittyvistä hypoteeseista ja niiden testaamisesta

Tämän luvun alkuosassa on näytetty, että erilaisten Lorenz-käyrien estimaattoreille voidaan johtaa pisteittäin asymptoottinen normaalijakauma. Tämän osan tarkoituksena on näyttää, miten tuota asymptoottista tulosta voidaan käyttää hyväksi talousteorian kannalta mielenkiintoisten hypoteesien testaamisessa. Tässä osassa käytetään esimerkkinä yleistettyjen Lorenz-käyrien testaamista, mutta sama metodologia soveltuu kaikkien muidenkin edellä käsiteltyihin käyriin liittyvien hypoteesien testaukseen. Hypoteesien muodostamisen ja dominanssirelaatioiden testaamisen päälähteenä on Savin (1984), joka käsittelee yleistä useamman hypoteesin samanaikaista testaamista lähinnä lineaarisen mallin tapauksessa.

4.4.1 Jakaumien yhtäsuuruuden testaus

Olkoon käytettävissä riippumattomat otokset kahdesta eri tulojakaumasta. Olkoot nyt n_1 ja n_2 otoskoot, Σ_1 ja Σ_2 yleistettyjen Lorenz-ordinaattojen asymptoottiset kovarianssimatriisit sekä G_1 ja G_2 yleistetyt Lorenz-ordinaatat ennalta asetetuissa tarkastelupisteissä. Hatulliset suureet kuvaavat jälleen otosestimaattoreita, joten $\hat{G}$ on G:n estimaattori. Olkoon nyt $\hat{d} = \hat{G}_1 - \hat{G}_2$. Tällöin $\hat{d}$ noudattaa asymptoottisesti normaalijakaumaa odotusarvovektorilla $d = G_1 - G_2$ ja kovarianssimatriisilla $\Sigma_1/n_1 + \Sigma_2/n_2$. Jos mielenkiintonollahypoteesina on, että kahden jakauman yleistetyt Lorenz-käyrät yhtäpitäviä eli $d = 0$, niin testisuure voidaan perustaa tavanomaisiin normaalijakauman tuloksiin. Asymptoottisesti päteväksi testisuureeksi muodostuu tällöin $\hat{d}'(\hat{\Sigma}_1/n_1 + \hat{\Sigma}_2/n_2)^{-1}\hat{d}$, jolla on nollahypoteesin vallitessa $\chi^2(k)$-jakauma, missä $k = \dim(d)$.[20]

Tämä menettely ei kuitenkaan ole tyydyttävä. Vaikka χ^2-testillä voidaan näyttää olevan monia hyviä tilastollisia ominaisuuksia, se ei vastaa talousteorian kannalta mielenkiintoiseen kysymykseen. Ongelmana on se, ettei χ^2-testin tulosta pystytä tulkitsemaan dominanssihypoteesien valossa. Itse asiassa χ^2-testi testaa, että onko olemassa sellaista vektorin $\hat{d}$ lineaarikombinaatiota, joka poikkeaisi nollasta tilastollisesti merkitsevästi (Savin 1984, 851-852). Tämä tietenkin on välttämätön ehto, että vahvoja dominanssirelaatioita voi esiintyä, muttei kuitenkaan riittävä. Nollahypoteesin hylkäämisen χ^2-testissä aiheuttava erotusten lineaarikombinaatio ei välttämättä ole dominanssirelaatioiden kannalta mielenkiintoinen. Dominanssirelaatiotestissä halutaan periaatteessa rajoittaa noiden lineaarikombinaatioiden määrää siten, että testataan ainoastaan niitä lineaarikombinaatioita, joiden avulla pystytään identifioimaan kumpi käyrä on toisen yläpuolella jossain pisteessä. Käytännössä tämä tarkoittaa vektorin $\hat{d}$ yksittäisten alkioiden merkitsevyyyden testaamista.

Hintana tästä hypoteesin selkeyttämisestä siten, että dominanssirelaatiot pystytään identifioimaan, on testin voimakkuuden huomattava aleneminen. Tästä syystä luvun viisi empiirisen osan taulukoissa on esitetty sekä dominanssitestisuuret että χ^2-testisuuret. Monissa tapauksessa dominanssitestisuureen antamat erotusten riskitasot ovat luokkaa yhdestä kymmeneen prosenttia, kun taas χ^2-testisuureen antamat riskitasot käyrien eroavuudelle ovat alle yhden kymmenestuhannesosan.

Toinen ongelma χ^2-testissä Lorenz-käyriin liittyvässä päättelyssä on se, että se ei ole tehokkain mahdollinen testiproseduuri siihen kysymykseen vastaamiseen, mihin se vastaa. χ^2-testi

[20] Tämä tietenkin edellyttäen, että matriisi $(\hat{\Sigma}_1/n_1 + \hat{\Sigma}_2/n_2)$ on täysiasteinen.

yleistettyjen Lorenz-ordinaattojen yhtäsuuruudelle testaa kahden jakauman yhtäpitävyyttä eli kyseessä on kahden otoksen yhteensopivuustesti. Tällöin χ^2-testille on olemassa parempia vaihtoehtoja kuten Kolmogorov-Smirnovin testi, jotka käyttävät koko aineiston sisältämän informaation hyväkseen, kun χ^2-testi käyttää ainoastaan hyväkseen ennalta valittujen pisteiden kohdalta laskettujen yleistettyjen Lorenz-ordinaattojen sisältämän informaation. (Bishop et al. 1989a, 726.)

Tulkitsemalla χ^2-testi kahden jakauman yhteensopivuustestinä voidaan ymmärtää luvun viisi taulukoissa silmiinpistävä ilmiö. Testattaessa yleistettyjen Lorenz-ordinaattojen, ryhmäkeskiarvojen tai simultaanista absoluuttisten Lorenz-ordinaattojen ja keskiarvojen yhtäsuuruutta, huomataan että χ^2-testisuureet ovat samoja riippumatta siitä mikä näistä kolmesta on testattava hypoteesi. Tämän on ymmärrettävissä tämän luvun alkuosan perusteella, jossa yleistettyjen Lorenz-ordinaattojen asymptoottisen jakauman perusteella johdettiin sekä ryhmäkeskiarvojen että absoluuttisten Lorenz-ordinaattojen ja keskiarvon asymptoottinen yhteisjakauma. Koska nämä molemmat suureet saadaan yleistetyistä Lorenz-ordinaatoista täysiasteisella lineaarimuunnoksella, niin χ^2-testisuure säilyy muuttumattomana. Tämä on helppo nähdä seuraavasti:

$$\begin{aligned}
\chi^2 &= (\hat{G}_1 - \hat{G}_2)'(\hat{\Sigma}_1/n_1 + \hat{\Sigma}_2/n_2)^{-1}(\hat{G}_1 - \hat{G}_2) \\
&= (D\hat{G}_1 - D\hat{G}_2)'(D\hat{\Sigma}_1/n_1 D' + D\hat{\Sigma}_2/n_2 D')^{-1}(D\hat{G}_1 - D\hat{G}_2) \\
&= (\hat{C}_1 - \hat{C}_2)'(\hat{\Sigma}_1(C)/n_1 + \hat{\Sigma}_2(C)/n_2)^{-1}(\hat{C}_1 - \hat{C}_2),
\end{aligned}$$

missä D:llä merkitään lineaarimuunnosta, jolla siirrytään jompaan kumpaan vaihtoehtoisista suureista, C:llä tätä suuretta ja $\hat{\Sigma}(C)$:llä tämän kovarianssimatriisin estimaattoria.

Edellisestä nähdään siis, että yleistettyjen Lorenz-ordinaattojen, ryhmäkeskiarvojen tai simultaaninen keskiarvojen ja absoluuttisten Lorenz-ordinaattojen testaaminen χ^2-testisuurella johtaa samaan lopputulokseen. Nämä kaikki ovat kahden otoksen yhteensopivuustestejä, jotka voidaan tulkita asymptoottisina Waldin testeinä. Tämä herättää kuitenkin sen kysymyksen, miten simultaaninen Lorenz-ordinaattojen ja keskiarvojen vertaaminen on nähtävä tämän asian valossa. Tarkoittaahan Lorenz-ordinaattojen ja keskiarvojen yhtäsuuruus myös jakaumien yhtäsuuruutta. Tähän vastaus löytyy standardista asymptoottisesta testiteoriasta. Koska Lorenz-ordinaatat saadaan yleistetyistä Lorenz-ordinaatoista epälineaarisella muunnoksella, mutta testattava nollahypoteesi pysyy loogisesti ekvivalenttina, on Lorenz-ordinaattojen ja keskiarvojen yhtäsuuruutta testaava χ^2-testisuure nähtävä saman hypoteesin toisena Waldin testisuureena, jossa nollahypoteesiä muunnettu epälineaarisella muunnoksella. Nollahypoteesin ollessa totta näillä kahdella eri Waldin testillä on sama asymptoottinen χ^2-jakauma, mutta äärellisissä otoksissa ne voivat johtaa eri tuloksiin. Tästä syystä ne luvun viisi tarkasteluissa eroavat toisistaan.

4.4.2 Dominanssirelaatioiden testaamisesta

Dominanssirelaatioiden testaamisessa ei primäärisenä mielenkiinnon kohteena ole se, että onko jokin poikkeamavektorin $\hat{d}$ mielivaltainen lineaarikombinaatio tilastollisesti merkitsevästi nollasti poikkeava. Dominanssirelaation identifioinnin kannalta on tärkeää, mitkä $\hat{d}$:n komponentit poikkeavat merkitsevästi nollasta ja ovatko merkittävät poikkeamat ainoastaan tietyn suuntaisia. Jos ainoastaan positiiviset poikkeamat ovat merkitseviä, niin ensiksi mainittu otos dominoi jälkimmäistä, jos taas vain negatiiviset poikkeamat ovat merkitseviä, niin jälkimmäinen otos dominoi ensimmäistä. Jos sekä positiiviset että negatiiviset poikkeamat ovat merkitseviä, niin vertailtavien käyrien välillä tapahtuu leikkaaminen. Testin pitää siis pystyä erottamaan seuraavat vaihtoehdot toisistaan:[21]

H0:	$d_i = 0\ \forall\, i$	Nollahypoteesi: käyrät ovat samoja
H1:	$d_i \geq 0\ \forall\, i$	Ensimmäinen jakauma dominoi toista
H2:	$d_i \leq 0\ \forall\, i$	Toinen jakauma dominoi ensimmäistä
H3:	$\exists\, i, j : d_i > 0$ ja $d_j < 0$	Käyrät leikkaavat.

Dominanssirelaation testaaminen perustuu siis yksittäisten poikkeamien tarkasteluun. Keskihajonnallaan standardoiduista erotuksista

$$t_i = \frac{\hat{d}_i}{\sqrt{\hat{\sigma}_{1ii}/n_1 + \hat{\sigma}_{2ii}/n_2}} \tag{8}$$

otetaan itseisarvoltaan suurimmat positiiviset ja negatiiviset testisuureiksi, jolloin testisuureiksi muodostuvat

$$t_+ = \max_{\{i\,:\,t_i>0\}} \{0, t_i\} \tag{9}$$

ja

$$t_- = \min_{\{i\,:\,t_i<0\}} \{0, t_i\}. \tag{10}$$

Dominanssirelaatiotestaus perustuu testisuureisiin t_+ ja t_-.[22] Näiden testisuureiden käyttö on kehitetty varianssianalyysin yhteydessä. Varianssianalyysin kehikossa halutaan identifioi-

[21] Symbolien $\geq$ ja $\leq$ katsotaan tässä tapauksessa sisältävän sen ehdon, että aito epäyhtälö toteutuu ainakin yhdellä indeksillä i. Muussa tapauksessa joukko-opin merkinnöin H1$\cap$H2 = H0, joten hypoteesit eivät olisi erillisiä.

[22] Vaikka tässä esitellyssä määritelmässä $t_+ \geq 0$ ja $t_- \leq 0$, niin luvun viisi testausten yhteenvetotaulukoissa on tapaus $t_+ = 0$ tai $t_- = 0$ merkitty puuttuvaksi havainnoksi. Testisuureen matemaattisessa määritelmässä tuo nollavaihtoehto on sallittu, jottei max- ja min-operaatioita jouduttaisi tekemään tyhjälle joukolle.

da kaikkien mahdollisten kontrastien (ryhmäkeskiarvojen erotusten) lineaarikombinaatioiden joukosta ne, jotka ovat tulkinnallisesti mielenkiintoisia. Tällöin niiden standardoitujen kontrastien, joiden itseisarvo ylittää testisuureen

$$t^* = max\{t_+, |t_-|\} \tag{11}$$

kriittiset arvot katsotaan olevan merkitseviä. Varianssianalyysin puolella on testisuureen t^*-kriittisten arvojen laskemiseksi esitetty useita eri vaihtoehtoja. Näitä ovat Bonferroni-epäyhtälön käyttäminen, Šidak-epäyhtälön käyttäminen ja niin sanotun Studentized Maximum Modulus (SMM)-jakauman taulukoitujen kriittisten arvojen käyttäminen. Varianssianalyysin tapauksessa vertailutilanne poikkeaa tässä käsiteltävästä dominanssirelaation testauksessa sillä tavalla, että varianssianalyysissä kontrastien yhteisjakaumana on moniulotteinen t-jakauma, kun dominanssirelaatiotesteissä ainoastaan ordinaattojen erotusten asymptoottinen jakauma tunnetaan. Asymptoottisesti ekvivalenttien Šidak-epäyhtälön ja SMM-jakauman käytön tehokkuuseroja on vertailtu varianssianalyysin yhteydessä: äärellisissä otoksissa SMM-jakauma johtaa tehokkaampaan testiin. Ehkä tästä syystä myös dominanssirelaatiokirjallisuudessa puhutaan SMM-jakaumasta, vaikka saadut tulokset voidaan johtaa käyttämällä Šidak-epäyhtälöä. Tätä nimityskäytäntöä noudatetaan tässäkin esityksessä, vaikka testisuureen kriittisen arvon johtamiseen käytetään Šidak-epäyhtälöä eikä SMM-jakauman ominaisuuksia käytetä missään vaiheessa hyväksi. (Savin 1984, 834-835.)

Olkoon dominanssitestauksessa nollahypoteesi totta eli $d = 0$ ja olkoon otoskoko riittävän suuri, että asymptoottiset tulokset ovat riittävällä tarkkuudella voimassa. Tällöin riippumatta testisuureiden välisestä kovarianssirakenteesta Bonferroni-epäyhtälön mukaan

$$P(t^* \leq z_{\delta/2}) \geq 1 - \delta k, \tag{12}$$

missä $z_{\delta/2}$-tarkoittaa standardoidun normaalijakauman $1 - \delta/2$-fraktiilia. Bonferroni-epäyhtälö lienee tunnetuin tapa sopeuttaa riskitasoja simultaanisessa päättelytilanteessa. Tällöin testisuureen t^* kriittiset arvot tasolla α saadaan standardoidun normaalijakauman $1 - \alpha/(2k)$ -fraktiileina. Tätä tehokkaampi todennäköisyysepäyhtälö on niin sanottu Šidak-epäyhtälö, jonka mukaan multinormaalijakaumalle pätee

$$P(t^* \leq z_{\delta/2}) \geq (1 - \delta)^k \tag{13}$$

riippumatta alkioiden välisestä kovarianssirakenteesta. Yhtäsuuruus Šidak-epäyhtälössä on voimassa, kun kovarianssimatriisi on diagonaalinen. Šidak-epäyhtälö antaa testisuureen t^* riskitason α kriittisiksi arvoiksi standardoidun normaalijakauman $1 - \delta/2$-fraktiilin, jossa $\delta = 1 - (1 - \alpha)^{1/k}$. Saman kriittisen arvon antaisi SMM-jakauma äärettömällä vapausasteella, joten jatkossa puhutaan SMM-jakaumasta. (Savin 1982, 834-835.)

Dominanssirelaatioiden yhteydessä ensimmäisen kerran tätä SMM-jakaumaan perustuvaa päättelyä sovelsivat Beach & Richmond (1985). Standardimenetelmä nykyään dominanssien testauksessa onkin verrata testisuureiden t_+ ja t_- merkitsevyyyttä SMM-jakauman kriittisiin arvoihin. Hypoteesin testauksessa on neljä vaihtoehtoista lopputulosta.

Taulukko 1. Dominanssirelaatioita koskevan päättelyn johtopäätökset testisuureiden t_+ ja t_- tilastollisen merkitsevyyden funktioina.

t_+	t_-	Päättelyn lopputulos
Ei-merkitsevä	Ei-merkitsevä	H0: Käyrät ovat samoja
Merkitsevä	Ei-merkitsevä	H1: Ensimmäinen jakauma dominoi toista
Ei-merkitsevä	Merkitsevä	H2: Toinen jakauma dominoi ensimmäistä
Merkitsevä	Merkitsevä	H3: Käyrät leikkaavat

Testisuureiden t_+ ja t_- käyttö mahdollistaa siis helpon tavan identifioida mikä mahdollisesta neljästä erillisestä hypoteesista on kyseessä. Lisäksi SMM-jakauman α-tason kriittisen arvon käyttäminen johtaa konservatiiviseen asymptoottiseen testiin, jonka riskitaso on korkeintaan α. Tällä tavalla muodostettua testiä Savin (1984, 829) nimittää äärelliseksi indusoiduksi testiksi tai testiksi joka perustuu leikkaus-yhdiste -periaattelle (union-intersect -principle).

Savin (1984, 835-845) esittää tarkastelun, jossa äärellisen indusoidun testin hylkäys- ja hyväksymisalueita verrataan χ^2-testin hyväksymys- ja hylkäysalueisiin kaksiulotteisessa tapauksessa. Näissä tarkasteluissa käy selvästi ilmi, että kaksiulotteisessa tapauksessa äärellisen indusoidun testin voima pienenee, kun korrelaatiokertoimen itseisarvo taustalla olevassa multinormaalijakaumassa kasvaa. Moniulotteisessa tilanteessa äärellisen indusoidun testin käyttö voi siis johtaa liian konservatiiseen testiin, jos yksittäisten testisuureiden välillä on lineaarisia riippuvuuksia. Tätä on pyritty välttämään tässä tutkimuksessa integroimalla numeerisesti Monte Carlo -menetelmällä nollahypoteesin mukaisesta multinormaalijakaumasta testisuureen riskitaso. Silti muutamassa tapauksessa χ^2-testi hylkää nollahypoteesin kirkkaasti, kun taas äärellinen indusoitu testi ei anna tilastollisesti selkeää ja yksikäsitteistä vastausta.

4.4.3 Eräitä vaihtoehtoisia testausperiaatteita

Sekä dominanssitestausproseduuriin että χ^2-testiin liittyy eräitä käytännön ongelmia. Kuten jo aikaisemmin on esitetty, on χ^2-testille parempi vaihtoehto Kolmogorov - Smirnov-testi. Kolmogorov - Smirnov-testin yksisuuntainen versio on mahdollinen ensimmäisen asteen stokastisen dominanssin testaamiseen ja myös toisen asteen stokastiselle dominanssille on Maasoumin (1995, 26) mukaan kehitetty (esim. McFadden 1989) testiä, joka käyttää hyväkseen koko otoksen informaation. MacFaddenin toisen asteen stokastisen dominanssin testi perustuu puhtaasti simulointiperiaatteelle. Tämän lisäksi toisen asteen stokastisen dominanssin testaukseen ovat Kaur et al. (1994) esittäneet asymptoottisen testin, jonka kriittiset arvot saadaan normaalijakaumasta. Tutkimus koko otoksen informaation sisällyttämiseksi erilaisiin dominanssitesteihin on vielä kesken, eikä esim. Kaur et al.:n esittämä testi sellaisenaan olisi soveltunut tähän tutkimukseen, koska sen yleistäminen painotettuun aineistoon tai muille kuin yleistetylle Lorenz-käyrälle ei olisi ollut suoraviivaista.

Toinen, tässä käytettyä menettelyä lähempänä oleva testausperiaate, on Xun (1995) esittämä menetelmä. Xun menetelmä perustuu Kodden ja Palmin (1986) esittämään Waldin testiin samanaikaisille yhtälö- ja epäyhtälörajoituksille. Xu käyttää tätä menetelmää yleistettyjen Lorenz-ordinaattojen testaamiseen ja tätä testausperiaatetta olisi helppo laajentaa kaikkiin tässä tutkimuksessa käsiteltyihin dominanssitesteihin. Ero tässä esitettyyn testausperiaatteeseen Xun menetelmässä on, että sen sijaan että testisuureeksi otettaisiin yksittäisten ordinaattojen standardoidun erotuksen itsearvolliset maksimit, otetaan testisuureeksi vapaan ja sidotun estimaatin[23] erotuksen pituuden neliö kovarianssimatriisin määrämässä metriikassa. Tällä tavalla muodostetun testisuureen jakaumaa ei kuitenkaan pystytä analyyttisesti laskemaan, mutta testisuureen riskitasoille pystytään laskemaan analyyttiset ylä- ja alarajat. Tämän tutkimuksen yhteydessä kokeiltiin Xun menetelmää Lorenz-käyrien yhtäsuuruuden testaamiseen ja todettiin analyyttiset testisuureen ylä- ja alarajat erittäin epäinformatiivisiksi. Täten testisuureiden riskitaso oltaisiin myös tässä menetelmässä jouduttu simuloimaan ja koska Xun menetelmä edellyttää kvadraattisen optimointitehtävän suorittamista testisuureen muodostamisen yhteydessä, päädyttiin SMM-jakaumaan perustuvan testisuureen käyttöön. Xun menetelmän tulkinnallinen heikkous SMM-jakaumaan perustuvaan testaukseen verrattuna on myös siinä, että SMM-jakauman perustuvassa testauksessa identifioituvat selkeästi ne jakauman pisteet, joissa tilastollisesti merkitsevää eroa esiintyy.

4.4.4 Käytetystä simulointiproseduurista

Luvun viisi empiirisessä esimerkissä on testisuureiden riskitasot laskettu simuloimalla testisuureen jakaumaa sen nollahypoteesin mukaisesta asymptooottisesta normaalijakaumasta. Tässä poiketaan dominanssitestauksessa yleisesti käytetystä menettelystä, jossa testisuureita verrataan SMM-jakauman kriittisiin arvoihin. SMM-jakaumaa käytettäessä menetetään

[23] Sidottu estimaatti lasketaan vapaan estimaatin avulla. Se saadaan vapaan estimaatin projektiona kovarianssimatriisin määrämässä metriikassa nollahypoteesin määrittelemään osaan parametriavaruudessa. Käytännössä sidottu estimaatti lasketaan kvadraattisen optimointiongelman ratkaisuna.

se lisäinformaatio, joka sisältyy testisuureiden välisiin asymptoottisiin kovariansseihin, sillä SMM-jakaumaa käytettäessä oletetaan testisuureiden kovarianssimatriisin olevan diagonaalinen. Voidaan kuitenkin näyttää, että SMM-jakauma muodostaa ylärajan kriittiselle testisuureen arvolle (Stoline & Ury 1979, 89).

Kovarianssien huomioonottaminen kuitenkin tehostaa testiproseduuria. Luvun viisi aineiston yhteydessä testisuureet ovat yleensä voimakkaasti (muttei täydellisesti) kollineaarisia. Tämä pystytään toteamaan kahdella alkeellisella mittarilla, jotka testaukseen kirjoitettu tietokoneohjelma tulosti: testisuurematriisin korrelaatiomatriisin suurimman ominaisvektorin selitysprosentilla ja testisuurematriisin korrelaatiomatriisin kuntoisuusluvulla. Vaikka nämä eivät välttämättä ole mitenkään optimaalisia kollineaarisuuden mittareita, niin niiden perusteella voidaan havainnollistaa kuitenkin testisuureiden dimensionaalisuutta. Korrelaatiomatriisin suurimman ominaisvektorin selitysprosentti vaihteli likimain välillä $[50, 85]$ ja kuntoisuusluku likimain välillä $[200, 250000]$ riippuen suoritetusta testistä. Täten ajatus diagonaalisesta korrelaatio- ja kovarianssimatriisista ei ollut luonteva ja tästä syystä kovarianssien sisältämä informaatio haluttiin sisällyttää testiproseduuriin. SMM-jakauman kriittiset arvot riippuvat nimittäin tarkasteluvektorin dimensiosta ja voimakkaan kovarianssirakenteen testisuurevektorin alkioiden välillä voidaan nähdä pienentävän tuota dimensiota.[24]

Simulointiproseduurina oli tavanomainen Monte Carlo -simulaatio. Ensimmäisessä vaiheessa muodostettiin riippumaton standardoitu multinormaalinen otos, jonka dimensio oli sama kuin testisuurevektorilla. Tämän jälkeen tämä riippumaton otosvektori kerrottiin testisuureiden kovarianssimatriisin eräällä hajoitelmalla,[25] jolloin päädytään satunnaisotokseen, jolla on haluttu kovarianssirakenne. Tätä proseduuria toistettiin satatuhatta kertaa ja sen jälkeen verrattiin kuinka monta itseisarvoltaan havaittua testisuuretta suurempaa havaintoa saatiin tästä nollahypoteesin mukaisesta simulaatiosta. Olkoon nyt x itseisarvoltaan testisuuretta suurempien havaintojen lukumäärä. Tällöin testisuureen simuloiduksi riskitasoksi saatiin $\hat{p} = x/100000$.

[24] Todellisuudessa mitään dimension oikeaa pienenemistä ei tapahdu, jollei testisuurevektori ole täysin kollineaarinen. Kuitenkin kovarianssirakenteen huomioonottaminen tarkasteluissa näennäisesti pienentää tuota dimensiota, koska se ottaa huomioon sen, ettei vaihtelu kaikissa dimensioissa ole yhtä merkittävää. Tilanne on tässä suhteessa analoginen pääkomponenttianalyysin kanssa.

[25] Usein multinormaalisen otoksen generoinnin yhteydessä hajoitelmana käytetään kovarianssimatriisin Cholensky-hajoitelmaa. Tässä yhteydessä on kuitenkin käytetty kovarianssimatriisin singulaariarvohajoitelman avulla saatavaa matriisia, koska singulaariarvohajoitelman laskeminen oli numeerisesti luotettavammin toteutettu käytetyssä Gauss-ohjelmistossa. Käytetty matriisihajoitelma ei vaikuta saatuihin tuloksiin, sillä mikä tahansa positiivisesti definiitin neliömatriisin "neliöjuuri" johtaa samaan lopputulokseen.

4.4.5 Simuloitujen riskitasojen luotettavuudesta

Simuloinneissa on tietenkin otettava huomioon, että simuloinnin tuloksina saadut riskitasot ovat nekin satunnaismuuttujia, joihin liittyy satunnaisvaihtelua. Simulointeja on siis suoritettava riittävä määrä, ettei hypoteesien testauksessa riskitasojen satunnaisvaihtelun osuus ole niin suuri, että se voisi vaikuttaa testin lopputulokseen. Tässä tutkimuksessa saatuja riskitasoja voidaan kuitenkin pitää riittävän luotettavina käytännön testaamisen suorittamiselle kymmenen, viiden ja yhden prosentin riskitasoilla. Vaikka testisuureen jakaumaa ei hallitakaan kuin numeerisesti, niin se, että SMM-jakaumasta saadaan ylärajat testisuureiden merkitsevyydelle lisää tulosten luotettavuutta. Simulointeja ei siis jouduttu suorittamaan ilman ennakkokäsityksiä odotettavista tuloksista. Toisaalta satatuhatta simulaatiota on riittävän suuri määrä, jotta tulosten riskitasot saadaan selville riittävällä tarkkuudella.

Havaittujen riskitasojen luotettavuuden selvittämiseksi suoritettiin eräitä suuntaa-antavia laskelmia. Olkoon nyt p testisuureen todellinen riskitaso ja $\hat{p}$ simuloinneissa havaittu suotuisten tapahtumien suhde eli empiirinen riskitaso. Tällöin binomijakaumasta tiedetään, että $\hat{p}$:n odotusarvo on p ja varianssi $\frac{p(1-p)}{n}$, missä n:llä luonnollisesti merkitään otoskokoa. Tällöin voidaan n:n ja p:n funktioina muodostaa symmetrisiä välejä, joihin havaittu $\hat{p}$ osuu halutulla varmuudella. Näitä välejä kutsutaan jatkossa hiukan virheellisesti luottamusväleiksi.

Riskitasojen luottamusväleinä laskelmissa käytettiin 99 prosentin luottamusvälejä. Näiden muodostamisessa käytettiin aluksi kahta frekventistiseen tilastolliseen päättelyyn perustuvaa menetelmää. Keskeisen raja-arvolauseen avulla voidaan muodostaa asymptoottinen väli $[p - z\sqrt{\frac{p(1-p)}{n}}, p + z\sqrt{\frac{p(1-p)}{n}}]$, jossa z:lla merkitään normaalijakauman kriittistä arvoa. Vaihtoehtoiset luottamusvälit saatiin Tsebysevin epäyhtälöstä, josta saadaan luottamusväli $[p - \sqrt{\frac{p(1-p)}{\alpha n}}, p + \sqrt{\frac{p(1-p)}{\alpha n}}]$, missä α:lla merkitään luottamustasoon liittyvää hylkäämisvirheen todennäköisyyttä (tässä tapauksessa yksi prosentti).

Testisuureen tarkasteluiden kannalta mielenkiintoinen alue on, kun todellinen $p \leq 0.1$. Oheisessa taulukossa on laskettu luottamusvälien pituuksien puolikas n ja p:n funktioina, kun $n \in \{1000, 10000, 10000\}$ ja $p \in \{0.01, 0.01\}$. Taulukon mukaan ainakin keskeisen raja-arvolauseen antamat luottamusvälit ovat riittävän lyhyitä käytännön riskitasojen tarkasteluiden kannalta kun $n = 100000$. Ongelmaksi jää kuitenkin se, että kun p pienenee, tarvitaan yhä enemmän havaintoja, jotta keskeisen raja-arvolauseen tuloksia voitaisiin hyödyntää.

Taulukko 2. Tsebysevin epäyhtälön, keskeisen raja-arvolauseen sekä posterioribetajakauman antamat 99 prosentin luottamusvälit n:n ja p:n funktioina.[26]

n	p	Tsebysevin epäyhtälö	Keskeinen raja-arvolause	Posteriori-beta-jakauma
1000	0.01	0.01 ± 0.03146	0.01 ± 0.00812	0.00402 - 0.02058
10000	0.01	0.01 ± 0.00995	0.01 ± 0.00256	0.00766 - 0.01280
100000	0.01	0.01 ± 0.00315	0.01 ± 0.00081	0.00921 - 0.01083
1000	0.10	0.10 ± 0.09487	0.10 ± 0.02448	0.07747 - 0.12633
10000	0.10	0.10 ± 0.03000	0.10 ± 0.00774	0.09246 - 0.10709
100000	0.10	0.10 ± 0.00947	0.10 ± 0.00245	0.09758 - 0.10246

Käyttämällä bayesiläistä lähestymistapaa pystytään tekemään valinta keskeisen raja-arvolauseen ja Tsebysevin epäyhtälön antamien luottamusvälien välillä. Tässä käännetään tarkastelu toisinpäin eli ei enää lasketa sitä, kuinka suurella todennäköisyydellä $\hat{p}$ osuu jollekin tietylle välille, vaan katsotaan millaisen luottamusvälin todelliselle parametriarvolle bayesiläinen havainnoitsija muodostaisi, kun on havaittu tietty määrä $x = \hat{p}n$ suotuisia havaintoja. Olkoon päätöksentekijällä käytettävissään ei-informatiivinen beta$(\frac{1}{2}, \frac{1}{2})$-priori eli ns. Jeffreysin ei-informatiivinen priori parametrin p jakaumasta. Olkoon otoskoko n ja suotuisten havaintojen määrä x. Tällöin päätöksentekijän posteriorijakauma parametrille p on beta$(\frac{1}{2} + x, \frac{1}{2} + n - x)$. Tämän posteriorijakauman suhteen voidaan muodostaa luottamusvälejä. Näissä laskelmissa käytettiin luottamusvälejä joiden ulkopuolelle jäi puoli prosenttia todennäköisyysmasta sekä jakauman ala- että yläpäähän. Nämä eivät ole tavanomaisia bayesiläisiä luottamusvälejä eli niin sanottuja hpd-luottamusvälejä (highest posterior density), mutta kyseessä olevan käyttötarkoitukseen ne ovat riittäviä, koska hpd-luottamusvälit ovat korkeintaan yhtä leveitä kuin näin muodostetut välit. (Box & Tiao 1973, 34-46, 122-124.)

Betajakaumasta saadut tulokset vahvistavat $\hat{p}$:n arvon ollessa sekä 0.1 ja 0.01 keskeisen raja-arvolauseen antamat luottamusvälien suuruusluokat. Kun $\hat{p} = 0.01$ ja $n = 100000$, ei luottamusväli ei ole vielä täysin symmetrinen odotusarvon suhteen, joten keskeinen raja-arvolause ei sinällään anna täydellistä kuvaa parametrin p posteriorijakaumasta. Kuitenkin bayesiläisen analyysin tulos on se, että simulointien tuloksena saadut riskitasot ovat suurella todennäköisyydellä käytännön tarkasteluiden kannalta riittävällä tarkkuudella oikein. Tämä johtopäätös tietenkin sillä varauksella, että saadut testisuureet noudattavat asymptoottista jakaumaansa, eli että otoskoko on riittävä tarkasteltavassa tuloaineistossa.

[26] Tässä on kuitenkin otettava huomioon p:n ja luottamusvälien erilainen tulkinta bayesiläisessä ja kahdessa muussa tarkastelussa. Bayesiläisessä analyysissä p on itse asiassa $\hat{p}$ eli havaittu osuus, kun muissa tarkasteluissa se on parametrin p todellinen arvo.

Tämä posterioribetajakaumaan perustuva analyysi on täysin yleinen, jota voi soveltaa muihinkin vastaavanlaisiin riskitasotarkasteluihin. Oleellista vain on, että simuloitujen luottamustasojen tarkastelu ymmärretään otannaksi binomijakaumasta. Johtopäätöksenä voidaan pitää, että jos halutaan simuloimalla saada jokin tuntematon riskitaso riittävällä varmuudella tunnetuksi, niin satatuhatta simulaatiota on ainakin riittävä määrä ja kymmenenkin tuhatta simulaatiota antaa aika tarkan kuvan todellisesta tilanteesta. Tuhat simulaatiota ei vielä anna riittävän tarkkaa kuvaa todellisesta riskitasosta, jotta sen perusteella voitaisiin suorittaa testausta halutulla riskitasolla.

4.4.6 Simuloinneilla saavutetuista testin tehokkuuseduista

Simuloimalla testisuureiden riskitasot saavutettiin testauksessa käytännön tilanteiden kannalta merkittävää tehokkuusetua. Luvun viisi tarkasteluissa simuloidut testisuureiden kriittisten arvojen vaihteluvälit sekä SMM-jakamauman kriittiset arvot on esitetty oheisessa taulukossa.

Taulukko 3. SMM-jakauman kriittiset arvot äärettömällä vapausasteella 19- ja 20-ulotteisissa tarkasteluissa, kun riskitaso $\alpha \in \{0.10, 0.05, 0.01\}$. Nämä arvot vastaavat standardoidusta normaalijakaumasta saatavia $1-\delta/2$-fraktiileja, missä k on tarkasteltavan testisuureen dimensio ja $\delta = 1-(1-\alpha)^{1/k}$. Lisäksi simuloimalla saatujen testisuureiden kriittisten arvojen vaihteluvälit luvun viisi tarkasteluissa. Simulointi 1 -rivi kattaa kaikki muut kuin ryhmäkeskiarvotarkastelut ja Simulointi 2 -rivi ryhmäkeskiarvotarkastelut.

Vertailu-tilanne	Riski-taso 0.10	Riski-taso 0.05	Riski-taso 0.01	Riski-taso 0.001
SMM k=19	2.774	3.004	3.466	4.044
SMM k=20	2.791	3.016	3.479	4.056
Simulointi 1	[2.143,2.315]	[2.440,2.586]	[3.006,3.123]	[3.651,3.819]
Simulointi 2	[2.541,2.552]	[2.805,2.817]	[3.326,3.336]	[3.933,3.982]

Luvun viisi testaustilanteissa on simuloinnit jätetty suorittamatta, jos testisuuret ylittivät itseisarvoltaan 4. Tähän perustelu saadaan SMM-jakaumasta: tällöin testisuureen riskitaso on korkeintaan promillen kokoluokkaa ja siksi nollahypoteesi hylätään suurella tilastollisella varmuudella. Muussa tapauksessa simuloinnit suoritettiin ja saatuja riskitasoja pystyttiin vertailemaan SMM-jakaumasta saatuihin.

Tulokset simuloinneista osoittavat SMM-jakauman mekaanisen käyttämisen johtavan lievästi heikompaan asymptoottiseen testiin. Simuloinneissa, jotka ottavat testisuureiden välisen korrelaation huomioon olivat kriittisten arvojen erot SMM-jakauman kriittisiin arvoihin joskus jopa yli 0.5. Näin ainakin muutamat luvun viisi tavallisiin Lorenz-käyriin liittyvistä testeistä olisivat päätyneet eri johtopäätöksiin SMM-jakaumaa käytettäessä.

Taulukossa on erotettu ryhmäkeskiarvojen vertailuissa saadut simuloidut kriittiset arvot muista kriittisistä arvoista. Tähän on syynä se, että ne olivat systemaattisesti suurempia ja siten lähempänä SMM-jakauman kriittisiä arvoja. Tälle on selityksenä se, että ryhmäkeskiarvoja testattaessa kollineaarisuutta mittaavat tunnusluvut olivat kertaluokkaa pienempiä kuin muissa testaustilanteissa. Täten ryhmäkeskiarvojen testauksessa oltiin lähempänä SMM-jakauman johtamisessa käytettyä riippumatonta testisuurevektoria.

Näiden tulosten perusteella voi vetää varovaisesti sen johtopäätöksen, että dominanssitestauksessa testisuureiden eksaktin asymptoottisen riskitason etsiminen simuloimalla on kannattavaa. Näin päästään hiukan voimakkaampaan testiin. Tämä on kuitenkin huomioitava sillä varauksella, että otoskokojen on oltava riittävän suuria, jotta asymptoottinen päättely olisi validia. Pienissä otoksissa ei kummankaan testausmenettelyn validiutta pystytä helposti varmentamaan. Yksi mahdollisuus olisi tehdä simulaatiotutkimus, joka selvittäisi näiden kahden testausperiaatteen pienen otoksen ominaisuuksia. Tämä onkin mahdollinen pienimuotoisen jatkotutkimuksen aihe.

5 Empiirinen esimerkki Suomen aineistolla

Luvuissa kaksi ja kolme on esitetty taloudellinen teoria hyvinvointi- ja eriarvoisuuserojen tutkimiselle tuloaineistoista. Tämän luvun tarkoituksena on havainnollistaa käsiteltyjen kriteerien soveltamista suomalaisella kotitaloustiedusteluaineistolla. Samalla pyritään havainnollistamaan millainen tilastollinen epävarmuus liittyy tehtyihin johtopäätöksiin.

5.1 Aineisto

Kotitaloustiedusteluaineisto on Tilastokeskuksen satunnaisotannalla keräämä haastatteluihin ja rekistereihin perustuva aineisto, joka sisältää muun muassa tietoja kotitalouksien erilaisista tulokäsitteistä. Tässä esimerkissä tarkasteltavaksi tulokäsitteeksi on valittu kotitalouden käytettävissä olevat tulot, johon sisältyvät kotitalouden tuotantotekijätulot ja nettotulonsiirrot, mutta ei kotitalouden saamien julkisten palveluiden arvoa. Nettotulonsiirtoihin on laskettu myös toisilta kotitalouksilta saadut avustukset, mutta ei toisille kotitalouksille maksettuja avustuksia (nämä katsotaan kotitalouden kulutukseksi). Tuotantotekijätuloihin on sisällytetty laskennallinen asumistulo omistusasujille. Nämä tiedot on kerätty aineistoon hallinnollisten rekisterien (esim. verotustiedot) ja vuosihaastatteluiden avulla.

Käytettävissä olevien tulojen valintaa mielenkiintomuuttujaksi voidaan perustella kahdelta eri näkökannalta. Kulutusta kuvaava muuttuja voisi olla talousteorian kannalta ihmisten hyvinvoinnin mittaamisessa relevantimpi kuin tuloja kuvaava muuttuja, koska kulutuksen muutosten voidaan katsoa heijastelevan ihmisten odotettavissa olevien elinkaaritulojen muutoksia. Kotitalousaineistossa kuitenkin kulutusta kuvaavien muuttujien voidaan epäillä sisältävän huomattavan määrän mittavirhettä. Tämä johtuu siitä, että aineistossa kulutusta kuvaavat muuttujat on kerätty suureksi osaksi lyhyiden (kaksi viikkoa tai kuukausi) havainnointiperiodien aikana, jolloin kohdekotitaloutta on pyydetty kirjaamaan erilaiset kulutusmenonsa. Tästä syystä kulutusmuuttujissa voidaan pelätä olevan enemmän tarpeetonta satunnaisvaihtelua kuin tulomuuttujista, jotka on kerätty vuositasolla.

Tämän ylimääräisen mittavirheen vaikutus ei jossain toisessa analysointimenetelmässä, kuten regressioon perustuvassa mallintamisessa, olisi niin merkittävää, koska on kehitetty menetelmiä, joilla mittavirheen vaikutusta tuloksiin voidaan kontrolloida (mm. instrumenttimuuttujamenetelmät). Koska tutkimuksen kohteena on muuttujan jakauman, eli muuttujan vaihtelun, tarkastelu, on tärkeää, että muuttujan arvojen vaihtelu on todellista, ei mittavirhettä. Tästä syystä mahdollisimman laaja tulokäsite, jonka toivotaan mahdollisimman hyvin kuvaavan kotitalouden taloudellista asemaa, sopii tarkastelun kohteeksi paremmin kuin jokin kulutusta kuvaava muuttuja.

Päätöstä jättää julkisten palvelujen arvo pois käytettävissä olevista tuloista voidaan perustella suurin piirtein samoilla mittavirheargumenteilla. Julkisten palvelujen käyttö on aineistossa havainnoitu suhteellisen lyhyellä periodilla kuten kulutuskin ja tästä periodin lyhyydestä aiheutuu ylimääräistä vaihtelua havaintoihin. Lisäksi, koska julkisten palvelujen arvo on aineistossa imputoituna niiden keskimääräisten tuotantokustannusten avulla, on mahdollista että niiden arvioinnissa on mukana merkittävä systemaattinen virhekomponentti.

Kotitaloustiedusteluaineisto on tutkimuksessa ollut käytettävissä vuosille 1971, 1976, 1981, 1985 ja 1990. Näiden otoskoot vaihtelivat seuraavan taulukon mukaisesti, jossa on myös esitetty käytettävissä olevien tulojen tunnuslukuja eri vuosille. Nämä on deflatoitu elinkustannusindeksillä vuoden 1990 rahaksi.

Taulukko 4. Aineiston tunnusluvut. Tulokäsitteenä käytettävissä olevat tulot OECD-yksikköä kohden vuoden 1990 rahaksi deflatoituna. Havainnot on painotettu kotitalouksien jäsenmäärillä ja otantapainoilla.

Vuosi	Otoskoko	Tulojen keskiarvo	Tulojen keskihajonta	Tulojen minimi	Tulojen maksimi	Gini-kerroin
1971	2986	42060.14	22306.01	151.10	290112.44	0.2698
1976	3348	48672.23	19681.97	511.34	247686.77	0.2130
1981	7386	51717.42	20002.49	416.31	282010.34	0.2071
1985	8200	57928.45	21941.42	115.36	229472.82	0.1998
1990	8253	70872.11	31042.99	1671.18	1246873.70	0.2050

Vastauskadosta ja muista otantaan liittyvistä ongelmista johtuen kotitaloustiedusteluaineisto sisältää jokaiselle kotitaloudelle arvioidun painon, joka kuvastaa kuinka monta samankaltaista kotitaloutta perusjoukossa kyseinen havaintoyksikkö edustaa. Näitä painoja on johdonmukaisesti käytetty kaikissa laskelmissa. Painotusta kotitaloustiedusteluaineistossa on käsitellyt Laaksonen (1988).

Alkuperäisen aineiston otoskoko vuodelle 1990 oli 8258. Näistä viisi kotitaloutta on jätetty tarkastelujen ulkopuolelle. Näillä kotitalouksilla käytettävissä olevien tulojen arvo oli negatiivinen mm. negatiivisten pääomatulojen takia. Koska ei ole uskottavaa, että kotitalouden hyvinvoinnille relevantti tulokäsite voisi olla negatiivinen kokonaisen vuoden ajan, lienee syytä uskoa, että ko. kotitalouksien taloudellisesta asemasta on jokin hyvinvoinnille oleellinen tekijä jäänyt havaitsematta.[27] Toisaalta esim. Lorenz-käyrän tulkinta muuttujien saadessa negatiivisia arvoja ei ole enää suoraviivaista. Tästä syystä nämä kotitaloudet on jätetty tarkastelujen ulkopuolelle.

5.2 Käytetty ekvivalenssiskaala ja kotitalouksien painotus hyvinvointitarkasteluissa

Tässä empiirisessä esimerkissä käytetään ekvivalenssiskaalana niin sanottua OECD-skaalaa. Tämä tarkoittaa sitä, että kotitalouden käytettävissä olevat tulot jaetaan ekvivalenttien aikuisten lukumäärällä. Ekvivalenttien aikuisten lukumäärä saadaan OECD-skaalassa seuraavasti: ensimmäinen aikuinen on yksi ekvivalentti aikuinen, lisäaikuiset 0.7 ekvivalenttia aikuista kukin ja lapset 0.5 ekvivalenttia aikuista kukin.[28] OECD-skaalan toivotaan muuntavan kotitalouksien väliset tuloerot kotitalouksien välisiksi hyvinvointieroiksi, koska se ottaa huomioon sekä kotitalouksien kokoerot, kotitalouden rakenteen jakautumisen aikuisiin ja lapsiin sekä suurempien kotitalouksien mahdolliset skaalaedut kulutuksessa. Parhaimmillaankin OECD-skaalan voidaan vain toivoa approksimoivan jotain todellista ekvivalenssiskaalaa.

OECD-skaalatut tulot on painotettu tarkasteluissa kotitalouksien jäsenmäärällä. Tämä tarkoittaa sitä, että esimerkiksi nelihenkinen perhe, jonka ekvivalentit tulot ovat kymmenen tuhatta markkaa, otetaan huomioon tarkasteluissa kuin se olisi neljä erillistä kymmenen tuhannen markan tulot saavaa yksikköä. Tämä menettely johtaa siihen, ettei käytettävissä olevien tulojen keskiarvoa voida pitää minkään kirjanpidollisen tulokäsitteen estimaattina, mutta tälle painotukselle on olemassa hyvinvointiteoreettiset perustelut. Ekvivalenssiskaalalla deflatoinnin toivotaan tehneen eri kotitalouksissa elävien yksilöiden käytettävissä olevat tulot, ja vahvoja oletuksia käyttäen myöskin heidän hyötynsä, keskenään vertailukelpoiseksi. Toisaalta koska yhteiskunnan hyvinvointifunktion voidaan perustella olevan individualistinen (eli sen argumentteina ovat yksilöiden hyödyt, ei kotitalouksien hyödyt) täytyy hyvinvointitarkasteluissa painottaa yksilöitä, ei kotitalouksia. Tällöin hyödyn oletetaan kotitalouden

[27] Esimerkiksi Nygård ja Sandström (1980, 7) toteavat, että hyvin määritellyn hyvinvointianalyysille relevantin tulokäsitteen pitää aina saada vain ei-negatiivisia arvoja. He lainaavat Vilfredo Paretoa (1896): "We should observe that when researching into the distribution of income, we are not concerned with the sources of income. Even the poorest man must be regarded as having sufficient income to keep him alive." (Englanninkielinen käännös ranskankielisestä alkutekstistä)

[28] Tämä skaala on yleisesti empiirisessä tutkimuksessa käytetty, mutta sen hyvyys muihin ekvivalenssiskaaloihin verrattuna on epäselvä kysymys. OECD-skaalaa voidaan perustella sillä, että sen käyttö on yksinkertaista. Talousteorian mielessä parempia voisivat olla mikroaineistoiden perusteella estimoidut skaalat, joissa hyödynnettäisiin muutakin kotitalouteen liittyvää informaatiota kuin pelkkää jakoa aikuisiin ja lapsiin. Ekvivalenssiskaalojen estimointiin ja niiden käyttöön kuitenkin liittyy hyvin paljon avoimia kysymyksiä ja tästä syystä tässä esimerkissä käytetään yleisesti käytettyä yksinkertaista skaalaa.

sisällä jakautuvan tasaisesti. (Uusitalo 1980, 25 ja Cowell 1984, 359.)

5.3 Saadut tulokset

Saatuja tuloksia käsitellään tässä luvussa kolme esitellyn teorian valossa. Periaatteessa tässä pyritään esittämään vastauksia kysymyksiin "miten hyvinvointi on muuttunut vuosina 1971-1990 tulojakauman valossa" ja "onko eriarvoisuus tulojakaumassa muuttunut tuona ajanjaksona." Tulosten tulkinnassa käytetään kahta eri kriteeri: tilastollista dominanssi- ja tavallista dominanssikriteeriä. Liitteessä 3 on esitetty yhteenvedot estimoiduista Lorenz-käyristä ja näiden estimaattien asymptoottiset hajontaestimaatit. Lisäksi liitteessä 3 on esitetty tulojen ryhmäkeskiarvot sekä ryhmien osuudet kokonaistuloista vinttiiliryhmittäin, sekä näiden asymptoottiset keskihajonnat. Kaikkia tämän tutkimuksen empiirisiä tuloksia tulkittaessa on huomiotava se, että kaikki markkamääräiset suureet on deflatoitu vuoden 1990 rahaksi.

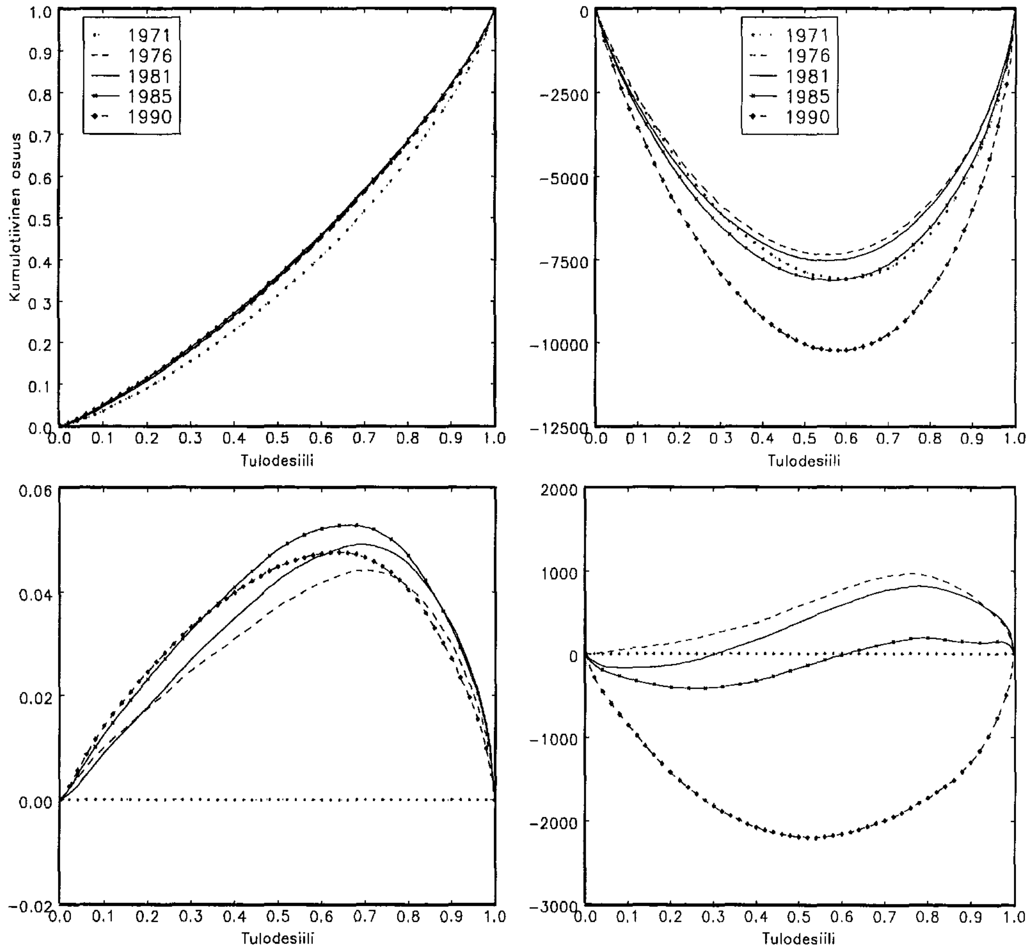
5.3.1 Suhteellisen eriarvoisuuden tarkastelut

Kuviossa 2 on esitetty Lorenz-käyrien kuvaajat eri vuosille sekä tarkasteluiden helpottamiseksi käyrien erotukset vuoden 1971 kuvaajasta. Ensimmäinen havaittava asia on, että vuosien 1976-1990 käyriä ei silmämääräisesti tavanomaisesta Lorenz-kuviosta pysty erottamaan toisistaan, mutta vuoden 1971 käyrä on muiden alapuolella. Kaikki muut vuodet siten dominoivat vuotta 1971. Tästä voi vetää sen johtopäätöksen, ettei vuoden 1971 jälkeen tulojakauman suhteellisessa eriarvoisuudessa ole tapahtunut kovin radikaaleja muutoksia.[29]

Käyrien erotuksia kuvaavan kuvion tarkastelu paljastaa eron vuoden 1971 ja muiden vuosien välillä selvästi. Lisäksi vuoden 1985 käyrä on vuoden 1976 käyrän yläpuolella koko ajan. Kaikki muut käyrät näyttävät leikkaavan tulojakauman jossain pisteessä, joten muiden vuosien välillä ei puhdasta dominanssia esiinny. Kuviosta voidaan kuitenkin huomata, että vuodet 1990 ja 1985 dominoivat vuosia 1981 ja 1976 tulojakauman köyhemmässä puoliskossa. Tästä voi vetää sen johtopäätöksen, että köyhimmän 50 prosentin suhteellinen asema on näinä vuosina 1985 ja 1990 ollut parempi kuin muina vuosina. Vuosi 1990 näyttää lisäksi dominoivan vuotta 1985 aivan jakauman alapäässä, kuten myös vuosi 1976 vuotta 1981.

[29] Tässä ja muissakin tulosten tulkinnassa täytyy kuitenkin huomioida se, että käytetty ekvivalenssiskaala voi vaikuttaa hyvinkin paljon tuloksiin ja jollain eri ekvivalenssiskaalalla olisi voinut saada erilaisia tuloksia. Myöskään 1990-luvun laman vaikutukset tulojakaumaan eivät näy näissä tarkasteluissa koska viimeinen tarkasteluvuosi on 1990 eikä näistä tarkasteluista voi siten vetää suoraan johtopäätöksiä nykytilanteesta.

Kuvio 2. Ylhäällä: i) Lorenz-käyrät eri vuosille, ii) absoluutiset Lorenz-käyrät, alhaalla: iii) Lorenz-käyrien erotukset vuoden 1971 käyrästä ja iv) absoluuttisten Lorenz-käyrien erotukset vuoden 1971 käyrästä.



Jos merkitsevän eron kriteeriksi otettaisiin OECD:n tutkimuksessa (Atkinson et al. 1995, 87) käytetty "vain yli prosenttiyksikön erotukset ovat merkitseviä", olisi johtopäätös se, että vuosi 1971 on kaikkien dominoima, vuosi 1985 dominoi vuotta 1976 ja kaikki muut Lorenz-käyrät ovat ekvivalentteja. Lorenz-käyrien ja ryhmien osuuksien kokonaistulosta erotuksien maksimit ja minimit eri vuosien välisissä vertailuissa on listattu oheisessa taulukossa. Tämä tuo perspektiiviä tässä esitettävälle tilastolliselle analyysille. Hyvällä syyllä voitaisiin kysyä, että ovatko eräät tässä analyysissä tilastolliseksi merkitseviksi havaitut erot kuitenkin niin pieniä, ettei niihin pitäisi kiinnittää huomiota.

Taulukko 5. Lorenz-käyrien erotusten maksimit ja minimit sekä vinttiiliryhmäosuuksien erotusten maksimit ja minimit eri vuosille prosenttiyksiköinä ilmaistuna.

Vuosi	Vuosi	Lorenz-käyrien erotusten maksimi	Lorenz-käyrien erotusten minimi	Ryhmäosuuksien erotusten maksimi	Ryhmäosuuksien erotusten minimi
1990	1985	0.1649	-0.6564	0.5197	-0.1143
1990	1981	0.6720	-0.6599	0.6018	-0.1376
1990	1976	0.9052	-0.2739	0.2535	-0.1628
1990	1971	4.7666	-	0.7103	-1.7620
1985	1981	0.6030	-0.0820	0.2337	-0.1211
1985	1976	1.0985	-	0.1704	-0.2266
1985	1971	5.2959	-	0.6486	-2.2282
1981	1976	0.5445	-0.1566	0.1135	-0.3482
1981	1971	4.9125	-	0.5171	-2.3638
1976	1971	4.4416	-	0.5258	-2.0156

Huomautuksia taulukkoon: Käyrien erotus on laskettu vähentämällä ensiksi mainitun vuoden käyrästä toiseksi mainitun vuoden käyrä tarkastelupisteissä. Jos maksimierotus tai minimierotus on ollut 0, on se ilmaistu puuttuvana tietona taulukossa.

Tilastollisen päättelyn mukaan tuominen muuttaa tarkasteluja huomattavasti. Jos hylkäämisvirheen riski asetetaan yhteen prosenttiin, niin dominanssirelaatio toteutuu kahdeksassa vertailussa kymmenestä suoritetusta vertailusta. Vuosien 1981 ja 1976 Lorenz-käyrät ovat yhden prosentin riskitasolla ekvivalentteja, kuten myös vuosien 1985 ja 1990. Muiden vertailuparien välillä dominanssirelaatio toteutuu kronologisessa järjestyksessä. Näin suhteellinen eriarvoisuus suomalaisessa tulojakaumassa näyttää pienentyneen tai pysyneen samana havaintovuodelta toiselle kronologisessa järjestyksessä.

Yhden prosentin riskitason käyttäminen voi kuitenkin olla järkevää useastakin syystä tässä tutkimuksessa. Viiden ja kymmenen prosentin riskitasolla merkitseviksi tulevat erot ovat niin pieniä, ettei niihin välttämättä kannata kiinnittää huomiota, koska niillä ei käytännössä olisi mitään merkitystä. Vaikka nämä erot olisivatkin muodollisesti tilastollisesti merkitseviä, niin aineistoon ja analyysiin liittyvät muut ongelmat (vastauskato aineistossa, kotitalouksien tulokäsitteiden määritelmät eri vuosille sekä ennen kaikkea tulosten riippuvuus käytetystä ekvivalenssiskaalasta) johtavat tekijän mielestä siihen, ettei muutaman promillen eroihin kahden Lorenz-käyrän välillä kannata kiinnittää suurta huomiota. On kuitenkin mielenkiintoista tarkastella, miten johtopäätökset olisivat muuttuneet, jos yhden prosentin riskitason sijasto oltaisiin käytetty viiden tai kymmenen prosentin riskitasoa. Nämä lisätarkastelut on syytä ymmärtää eräänlaisiksi herkkyysanalyyseiksi, ne kertovat, ovatko yhden prosentin riskitasolla saadut tulokset robusteja riskitason muutokselle. Viiden ja kymmenen prosentin riskitasojen

Taulukko 6. Pareittainen tilastollinen analyysi OECD-skaalattujen tulojen Lorenz-käyrille.

Vuosi	Vuosi	Positiivinen testisuure	Riskitaso	Negatiivinen testisuure	Riskitaso	χ^2-testisuure	Riskitaso
1990	1985	2.369	0.0670	-2.585	0.0384	44.01	< 0.001
1990	1981	7.036	< 0.01	-2.516	0.0459	96.82	< 0.001
1990	1976	5.427	< 0.01	-0.847	0.8155	86.86	< 0.001
1990	1971	16.448	< 0.01	-	-	351.11	< 0.001
1985	1981	4.928	< 0.01	-0.500	0.9718	47.38	< 0.001
1985	1976	4.947	< 0.01	-	-	60.86	< 0.001
1985	1971	16.891	< 0.01	-	-	367.07	< 0.001
1981	1976	2.162	0.1142	-2.527	0.0479	41.35	0.0022
1981	1971	14.576	< 0.01	-	-	271.17	< 0.001
1976	1971	11.410	< 0.01	-	-	201.17	< 0.001

Huomautuksia taulukkoon: Positiivinen testisuure testaa onko käyrissä tilastollisesti merkitsevä poikkeama siten, että ensiksi mainitun vuoden Lorenz-käyrä on toiseksi mainitun vuoden Lorenz-käyrän yläpuolella jossain pisteessä. Positiivisen ja negatiivisen testisuureen riskitasot on saatu simuloimalla 100 000 riippumatonta otosta asymptoottisesta normaalijakaumasta. Merkintä < 0.01 tarkoittaa sitä, että simulointeja ei ole suoritettu, koska testisuure ylittää SMM-jakauman (Studentized Maximum Modulus -jakauman) kriittisen arvon 0.01 riskitasolla. χ^2-testisuure testaa Lorenz-ordinaattojen yhtäsuuruutta samanaikaisesti kaikissa 20 vertailupisteessä. Sen riskitasoja voi verrata dominanssitestiriskitasoihin, jos haluaa todeta kuinka paljon testin voimakkuudesta menetetään kun testi muodostetaan "yhdiste-leikkaus"-periaatteella (union-intersect-principle).

käyttäminen johtaisi toiseen johtopäätökseen kolmessa vertailutilanteessa.

Vaikka yhden prosentin riskitasolla vuosien 1976 ja 1981 Lorenz-käyrät ovat ekvivalentteja, niin viiden ja kymmenen prosentin riskitasoilla vuosi 1976 Lorenz-dominoi vuotta 1981. Tämä dominanssi tapahtuu vielä köyhimmässä vinttiilissä, joten se on eriarvoisuustarkasteluiden kannalta mielenkiintoista. Tämän tuloksen tulkintaa tosin vaikeuttaa vielä se, että vuoden 1981 Lorenz-käyrä on melkein merkitsevästi (riskitaso 0.1142) vuoden 1976 käyrän yläpuolella jakauman keskivaiheilla. Koska havaitut erot kuitenkin ovat absoluuttisesti niin pieniä, voidaan näiden vuosien tulojakaumia pitää likimain Lorenz-ekvivalentteina.

Vuosien 1981 ja 1990 vertailu myös päätyisi eri lopputulokseen, jos käytettäisiin yhden prosentin sijasta viiden tai kymmenen prosentin riskitasoa. Tällöin Lorenz-käyrät leikkaisivat. Vuoden 1990 Lorenz-käyrä on kuitenkin vuoden 1981 käyrän yläpuolella aina jakauman pisteeseen $p = 0.65$, joten eriarvoisuustarkasteluiden kannalta tärkeässä jakauman alapäässä vuosi 1990 kiistatta dominoi vuotta 1981.

Taulukko 7. Tavanomaisen ja tilastollisen päättelyn johtopäätökset kymmenen, viiden ja yhden prosentin riskitasoilla Lorenz-vertailuista eri vuosien välillä.

Vuosi	Vuosi	Ei-tilastollinen analyysi	Riskitaso 0.10	Riskitaso 0.05	Riskitaso 0.01
1990	1985	X_L	X_L	$<_L$	$\sim_L$
1990	1981	X_L	X_L	X_L	$>_L$
1990	1976	X_L	$>_L$	$>_L$	$>_L$
1990	1971	$>_L$	$>_L$	$>_L$	$>_L$
1985	1981	X_L	$>_L$	$>_L$	$>_L$
1985	1976	$>_L$	$>_L$	$>_L$	$>_L$
1985	1971	$>_L$	$>_L$	$>_L$	$>_L$
1981	1976	X_L	$<_L$	$<_L$	$\sim_L$
1981	1971	$>_L$	$>_L$	$>_L$	$>_L$
1976	1971	$>_L$	$>_L$	$>_L$	$>_L$

Huomautuksia taulukkoon: Merkintä $>_L$ tarkoittaa ensinnä mainitun vuoden dominoivan toiseksi mainittua, merkintä $<_L$ toiseksi mainitun dominoivan ensiksi mainittua, merkintä $\sim_L$ tarkoittaa ettei käyrien välillä ole eroa ja X_L tarkoittaa käyrien leikkaavan.

Vuosien 1985 ja 1990 välisessä vertailuissa tulokset jäävät kaikkein epäselvimmiksi. Yhden prosentin riskitasolla Lorenz-käyrät ovat ekvivalentteja, kun viiden prosentin riskitasolla vuosi 1985 dominoi vuotta 1990 ja kymmenen prosentin riskitasolla käyrät leikkaavat. Käyrien leikkaaminen tapahtuu jakauman pisteessä $p = 0.35$, jota ennen vuoden 1990 käyrä näyttää olevan vuoden 1985 käyrän yläpuolella. Näiden kahden vuoden välisestä suhteellisesta asemasta on siis hyvin vaikea sanoa mitään suurella varmuudella, mutta jos huomioidaan mahdollisten erojen absoluuttinen pienuus, lienee turvallista pitää näitä kahta vuotta Lorenz-ekvivalentteina.

Tämän luvun alussa sivulla 45 olevassa taulukossa 4 on esitetty myös ginikertoimen piste-estimaatit tarkasteluvuosille. Tämä on tehty lähinnä ginikertoimen käytön yleisyyden vuoksi. Ginikertoimien hajontoja ei ole estimoitu, koska tässä tutkimuksessa on tilastollisen tarkastelun lähtökohdaksi otettu fundamentaalimpi Lorenz-dominanssin käsite. Lisäksi kuten luvussa kolme on todettu, ei ginikerroin välttämättä ole optimaalinen eriarvoisuusindeksi. Ginikertoimen piste-estimaatit kuitenkin vahvistavat tässä esitettyä näkemystä, että suhteellisessa eriarvoisuudessa on vuoden 1976 jälkeen tapahtunut vain vähän muutoksia. Kronologinen dominanssi ei ginikertoimenkaan piste-estimaattien tapauksessa toteudu, sillä vuoden 1985 ginikertoimen piste-estimaatti on pienempi kuin vuoden 1990.

Vaikka tulosten tilastollinen analysointi siis osoittaa, että dominanssirelaatioita voi esiintyä myös sellaisissa tarkasteluissa, joissa toinen osapuoli ei ole vuosi 1971, on todettava, että analyysin kvalitatiivinen tulos, jonka voi myös kuviosta todeta, on että vuosi 1971 poikkeaa muista vuosista selvästi ja muiden vuosien Lorenz-käyrien erot ovat pieniä. Vuoden 1976 jälkeinen kehitys jää hiukan epäselväksi, tosin selvää on, että suhteelliselta eriarvoisuudeltaan vuosi 1985 on lievästi parempi kuin 1981 ja 1976 ja lisäksi vuosi 1990 on lievästi parempi kuin 1976.

5.3.2 Absoluuttisen eriarvoisuuden tarkastelut

Absoluuttisen eriarvoisuuden kehittymisestä tulojakaumassa silmiinpistävin havainto on se, että vuoden 1990 tulojakaumaa dominoivat kaikkien muiden vuosien tulojakaumat. Tämä on luonnollista, koska vuosien 1976-1990 suhteellinen eriarvoisuus oli likimain yhtä suurta Lorenz-kriteerin mielessä ja vuoden 1990 tulojen keskiarvo on huomattavasti muita suurempi. Siis, kun keskiarvo kasvaa ja suhteellinen eriarvoisuus on likimain vakio, kasvaa absoluuttinen eriarvoisuus. Tämän lisäksi vuosien 1976 ja 1981 absoluuttiset Lorenz-käyrät dominoivat vuoden 1985 absoluuttista Lorenz-käyrää ja vuoden 1976 käyrä dominoi vuoden 1971 käyrää. Muiden vuosien käyrät leikkaavat jossain pisteessä.

Tilastollisen analyysin johtapäätökset ovat riippumattomia siitä, valitaanko tilastollisen merkitsevyyden rajaksi yhden, viiden vai kymmenen prosentin riskitaso. Tilastollinen tarkastelu vahvistaa ei-tilastollisen analyysin dominanssirelaatiot. Vuosi 1990 on kaikkien muiden dominoima. Samoin vuosi 1985 on sitä edeltävien vuosien dominoima, koska vuosien 1985 ja 1971 käyrien leikkaaminen jakauman loppupäässä ei ole tilastollisesti merkitsevää. Vuoden

Taulukko 8. Pareittainen tilastollinen analyysi OECD-skaalattujen tulojen absoluuttisille Lorenz-käyrille.

Vuosi	Vuosi	Positiivinen testisuure	Riskitaso	Negatiivinen testisuure	Riskitaso	χ^2-testisuure	Riskitaso
1990	1985	-	-	-12.902	< 0.01	235.57	< 0.001
1990	1981	-	-	-16.793	< 0.01	387.44	< 0.001
1990	1976	-	-	-16.697	< 0.01	386.09	< 0.001
1990	1971	-	-	-14.208	< 0.01	334.69	< 0.001
1985	1981	-	-	-4.984	< 0.01	37.02	0.0041
1985	1976	-	-	-7.245	< 0.01	84.57	< 0.001
1985	1971	0.977	0.6832	-6.498	< 0.01	114.44	< 0.001
1981	1976	0.124	0.9999	-5.319	< 0.01	54.65	< 0.001
1981	1971	4.190	< 0.01	-4.647	< 0.01	112.38	< 0.001
1976	1971	4.337	< 0.01	-	-	60.81	< 0.001

Taulukko 9. Tavanomaisen ja tilastollisen päättelyn johtopäätökset kymmenen, viiden ja yhden prosentin riskitasoilla absoluuttisista Lorenz-vertailuista eri vuosien välillä.

Vuosi	Vuosi	Ei-tilastollinen analyysi	Riskitaso 0.10	Riskitaso 0.05	Riskitaso 0.01
1990	1985	$<_{AL}$	$<_{AL}$	$<_{AL}$	$<_{AL}$
1990	1981	$<_{AL}$	$<_{AL}$	$<_{AL}$	$<_{AL}$
1990	1976	$<_{AL}$	$<_{AL}$	$<_{AL}$	$<_{AL}$
1990	1971	$<_{AL}$	$<_{AL}$	$<_{AL}$	$<_{AL}$
1985	1981	$<_{AL}$	$<_{AL}$	$<_{AL}$	$<_{AL}$
1985	1976	$<_{AL}$	$<_{AL}$	$<_{AL}$	$<_{AL}$
1985	1971	X_{AL}	$<_{AL}$	$<_{AL}$	$<_{AL}$
1981	1976	X_{AL}	$<_{AL}$	$<_{AL}$	$<_{AL}$
1981	1971	X_{AL}	X_{AL}	X_{AL}	X_{AL}
1976	1971	$>_{AL}$	$>_{AL}$	$>_{AL}$	$>_{AL}$

1981 käyrän ja 1976 käyrän leikkaaminen ei ole tilastollisesti merkitsevää, joten vuosi 1976 dominoi vuotta 1981. Vuosien 1981 ja 1971 käyrien leikkaaminen on tilastollisesti erittäin merkitsevää. Tämä on ainoa vertailtava pari, jossa tilastollisella kriteerillä dominanssirelaatioita ei esiinny. Vuosi 1976 dominoi myös vuotta 1971 erittäin merkitsevästi.

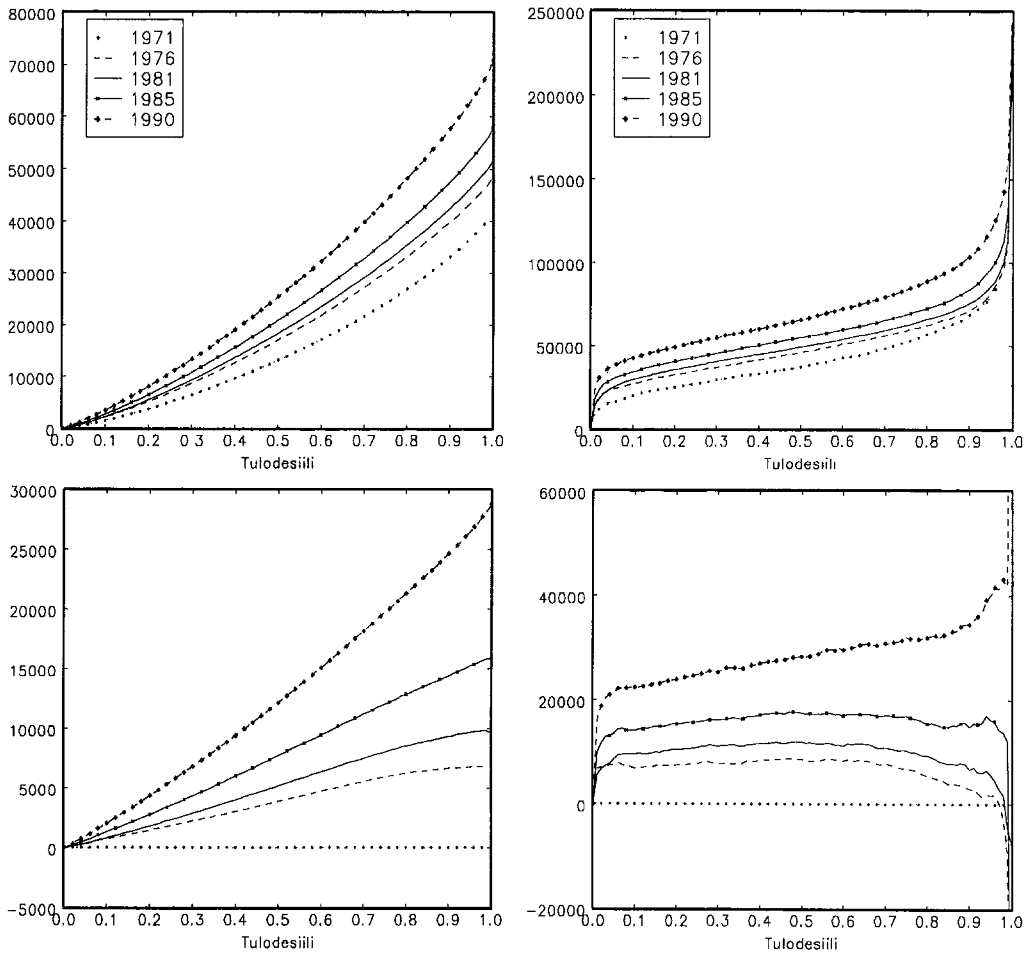
Tilastollinen analyysi absoluuttisesta eriarvoisuudesta johtaa siihen johtopäätökseen, että absoluuttinen eriarvoisuus on ollut suurimmillaan vuonna 1990 ja pienimmillään vuonna 1976. Vuoden 1976 jälkeen tulojen kasvu on tapahtunut siten, että suhteellinen eriarvoisuus tulojakaumassa on ollut kohtuullisen vakio, jolloin absoluuttinen eriarvoisuus on kasvanut.

5.3.3 Anonyymin Pareto-periaateen mukaiset hyvinvointitarkastelut

Anonyymin Pareto-periaatteen mukaan jakauma on parempi kuin toinen, jos sen fraktiilifunktio on kokonaan toisen jakauman fraktiilifunktion yläpuolella. Fraktiilifunktioiden sijasta tässä on tarkasteltu ryhmäkeskiarvoja, joita riittävän tiheästi tarkasteltuina voidaan pitää diskreettinä approksimaationa jatkuvalle fraktiilifunktiolle. Ryhmäkeskiarvojen dominanssi on välttämätön, muttei riittävä ehto fraktiilifunktioiden dominanssille. Tätä menettelyä käyttävät Bishop et al. (1989b).

Graafinen tarkastelu fraktiilifunktion dominanssirelaatioista antaa kuitenkin syytä epäillä että dominanssirelaatioita esiintyy. Vuoden 1990 fraktiilikuvaaja dominoi kaikkia muita kuvaajia. Vuoden 1971 käyrä leikkaa muut käyrät jakauman loppupäässä ja jakauman aivan alkupäässä

Kuvio 3. Ylhäällä: i) yleistetyt Lorenz-käyrät eri vuosille ii) fraktiilifunktiokuvaajat, alhaalla: iii) yleistettyjen Lorenz-käyrien erotus vuoden 1971 käyrästä ja iv) fraktiilifunktioiden erotus vuoden 1971 käyrästä.



tapahtuu useita leikkaamisia. Fraktiilifunktion tarkastelu jakauman loppu- tai alkupäässä on tilastollisesti epävarmaa, sillä normaalit asymptoottiset tulokset eivät päde, kun $p \to 0$ tai $p \to 1$, koska tällöin estimoidaan jakauman pienintä tai suurinta arvoa. Ryhmäkeskiarvokuvaajaksi tasoitettunakin tapahtuu kolme leikkausta: vuoden 1971 käyrä leikkaa vuosien 1976 ja 1981 käyrät ylimmässä vinttiilissä sekä samoin leikkaavat vuosien 1981 ja 1976 käyrät.

Tilastollisessa analyysissä kaikki kolme käytettyä riskitasoa johtavat samoihin päätöksiin. Tilastollinen tarkastelu ryhmäkeskiarvoja käyttäen vahvistaa vuoden 1990 olevan anonyymillä Pareto-periaatteella muita vuosia parempi. Vuosi 1985 dominoi myöskin merkitsevästi ai-

Taulukko 10. Pareittainen tilastollinen analyysi OECD-skaalattujen tulojen ryhmäkeskiarvoille.

Vuosi	Vuosi	Positiivinen testisuure	Riskitaso	Negatiivinen testisuure	Riskitaso	χ^2-testisuure	Riskitaso
1990	1985	29.474	< 0.01	-	-	1441.04	< 0.001
1990	1981	41.841	< 0.01	-	-	3074.17	< 0.001
1990	1976	43.292	< 0.01	-	-	3745.40	< 0.001
1990	1971	66.550	< 0.01	-	-	7841.10	< 0.001
1985	1981	16.669	< 0.01	-	-	536.75	< 0.001
1985	1976	20.598	< 0.01	-	-	1013.52	< 0.001
1985	1971	44.462	< 0.01	-	-	4313.03	< 0.001
1981	1976	8.477	< 0.01	-0.672	0.9976	188.12	< 0.001
1981	1971	30.585	< 0.01	-0.137	1.0000	1877.49	< 0.001
1976	1971	17.210	< 0.01	-0.809	0.9887	711.17	< 0.001

Taulukko 11. Tavanomaisen ja tilastollisen päättelyn johtopäätökset kymmenen, viiden ja yhden prosentin riskitasoilla ryhmäkeskiarvo-vertailuista eri vuosien välillä.

Vuosi	Vuosi	Ei-tilastollinen analyysi	Riskitaso 0.10	Riskitaso 0.05	Riskitaso 0.01
1990	1985	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1990	1981	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1990	1976	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1990	1971	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1985	1981	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1985	1976	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1985	1971	$>_{AP}$	$>_{AP}$	$>_{AP}$	$>_{AP}$
1981	1976	X_{AP}	$>_{AP}$	$>_{AP}$	$>_{AP}$
1981	1971	X_{AP}	$>_{AP}$	$>_{AP}$	$>_{AP}$
1976	1971	X_{AP}	$>_{AP}$	$>_{AP}$	$>_{AP}$

kaisempia vuosia. Vuosi 1981 dominoi vuosia 1976 ja 1971, vaikka näiden käyrät leikkaavat jossain jakauman pisteessä. Ryhmäkeskiarvon estimaatti vuodelle 1981 on pienempi kuin vuodelle 1976 ensimmäisessä vintiilissä, mutta tämä erotus ei ole merkitsevä. Ryhmäkeskiarvon estimaatti viimeiselle vintiiliryhmälle vuonna 1971 on suurempi kuin vuonna 1981 ja 1976, mutta nämä erot eivät kuitenkaan ole tilastollisesti merkitseviä. Vuosi 1976 dominoi siis vuotta 1971.

Näistä tuloksista voi vetää sen johtopäätöksen, että jos hyväksytään OECD-skaala oikeaksi ekvivalenssiskaalaksi ja uskotaan tulojen deflatoinnin elinkustannusindeksillä muuttavan tulot "oikein" vertailukelpoisiksi, niin suomalaisessa tulojakaumassa on tapahtunut paranemista anonyymin Pareto-kriteerin mielessä vuodesta 1971 vuoteen 1990 siten, että edeltäneet vertailuvuodet ovat olleet aina seuraajiaan huonompia.

5.3.4 Yleistettyyn Lorenz-käyrään perustuvat hyvinvointitarkastelut

Koska edellisissä vertailuissa saatiin tulokseksi, että hyvinvointi tulojakauman valossa kasvoi aina siirryttäessä myöhäisempään vertailuvuoteen anonyymillä Pareto-periaatteella, ovat yleistettyyn Lorenz-käyrään perustuvat vertailut redundantteja. Tämä johtuu siitä, että anonyymi Pareto-dominanssi on yhtäpitävää niin sanottuun ensimmäisen kertaluvun stokastisen dominanssin kanssa, joka implikoi toisen kertaluvun stokastisen dominanssin eli yleistetyn Lorenz-dominanssin. Näin ollen 1990 $>_{GL}$ 1985 $>_{GL}$ 1981 $>_{GL}$ 1976 $>_{GL}$ 1971. Tämä relaatio osoittautui myös tilastollisesti merkitseväksi. Ainoa käyrien leikkaaminen tapahtui vuosien 1981 ja 1976 välillä ensimmäisessä vintiilissä. Tämä erotus ei kuitenkaan ollut tilastollisesti merkitsevä.

Näin siis voidaan sanoa, että aineiston valossa suomalaisessa tulojakaumassa on tapahtunut paranemista yleistetyn Lorenz-dominanssin mielessä kronologisessa järjestyksessä tarkasteluperiodilla. Eriarvoisuutta kaihtavan kasvavan hyvinvointifunktion omaava havainnoitsija siis preferoisi tulojakaumia tässä järjestyksessä.

Yleistetty Lorenz-dominanssikriteeri ei siis tässä vertailussa tuonut mitään uutta tarkasteluihin. Empiirisissä vertailuissa on myös havaittu, että vaikka anonyymi Pareto-kriteeri ei mahdollistaisi vertailuja kaikkien tulojakaumien välillä, niin yleistetyn Lorenz-kriteerin käyttöönotto ei välttämättä lisää vertailtavien parien määrää merkittävästi. Bishop et al. (1989b) vertaillessaan Yhdysvaltojen 51:n osavaltion tulojakaumaa koko maan jakaumaan, toteavat anonyymin Pareto-kriteerin mahdollistavan vertailujen tekemisen 46:n osavaltion osalta.[30] Yleistetty Lorenz-kriteeri lisää mahdollisten vertailtavien osavaltioiden määrää vain kahdella. He toteavat (s. 72), että yleistetty Lorenz-kriteeri on enemmän tehokkuus kuin tasa-arvoisuuskriteeri, koska se voi pitää parempana jakaumaa, joka on suhteelliselta eriarvoisuudeltaan huomattavasti huonompi kuin toinen, kunhan sen keskiarvo on riittävän paljon tasaisempaa jakaumaa parempi. Täten vaikka yleistettyä Lorenz-kriteeriä voidaan pitää teoreettisesti mielenkiintoisena hyvinvointikriteerinä, niin se ei kuitenkaan tee erillisiä puhtaita eriarvoisuusvertailuja tarpeettomiksi. Suhteellisen eriarvoisuuden vertailujen (siis tavanomaisten Lorenz-vertailujen) hyvä puoli on myös siinä, että ne soveltuvat aikasarja-aineistoihin ja

[30] He käyttävät kuitenkin termiä "rank criterion" termin anonyymi Pareto-kriteeri sijasta.

Taulukko 12. Pareittainen tilastollinen analyysi OECD-skaalattujen tulojen yleistetyille Lorenz-käyrille.

Vuosi	Vuosi	Positiivinen testisuure	Riskitaso	Negatiivinen testisuure	Riskitaso	χ^2-testisuure	Riskitaso
1990	1985	34.399	< 0.01	-	-	1441.04	< 0.001
1990	1981	51.677	< 0.01	-	-	3074.17	< 0.001
1990	1976	51.422	< 0.01	-	-	3745.40	< 0.001
1990	1971	68.556	< 0.01	-	-	7841.10	< 0.001
1985	1981	19.180	< 0.01	-	-	536.75	< 0.001
1985	1976	23.968	< 0.01	-	-	1013.52	< 0.001
1985	1971	45.429	< 0.01	-	-	4313.03	< 0.001
1981	1976	8.328	< 0.01	-0.6721	0.8898	188.12	< 0.001
1981	1971	30.303	< 0.01	-	-	1877.49	< 0.001
1976	1971	19.622	< 0.01	-	-	711.17	< 0.001

Huomautuksia taulukkoon: χ^2-testisuureet ovat samoja kuin ryhmäkeskiarvojen vertailuissa. Tämä johtuu siitä, että ryhmäkeskiarvot saadaan yleistetyistä Lorenz-kordinaateista täysiasteisella lineaarimuunnoksella. Syy siihen, ettei negatiivisen testisuureen -0.672 riskitaso näissä vertailuissa ole sama kuin ryhmäkeskiarvojen vertailuissa on testisuureiden erilaisessa kovarianssirakenteessa. Yleistetyt Lorenz-ordinaatat ovat enemmän kollineaarisia kuin ryhmäkeskiarvot.

kansainvälisiin tarkasteluihin paremmin kuin hyvinvointivertailut, koska ne ovat skaalainvariantteja. Tällöin mahdolliset virheet, jotka seuraavat tulojen deflatoinnista joksikin reaalisuureeksi, eivät vaikuta tuloksiin. Lisäksi eriarvoisuusvertailuiden hyvänä puolena on se, ettei niiden yhteydessä tarvitse eksplisiittisesti puhua yhteiskunnan hyvinvointifunktioista vaan voidaan paljon lievemmin oletuksin lähinnä mitata empiirisesti mielenkiintoista ilmiötä.

Suomen aineistossa ei yleistetty Lorenz-kriteeri pidä parempana jakaumaa, jonka suhteellinen eriarvoisuus on suurempaa kuin dominoidun, jos Lorenz-vertailuissa tilastollisesti merkitsevän poikkeaman rajana pidetään yhden prosentin riskitasoa.[31] Tällöin vuosien 1981 ja 1976 Lorenz-käyrät ovat ekvivalentteja, samoin kuin vuosien 1990 ja 1985 Lorenz-käyrät. Viiden ja kymmenen prosentin riskitasolla kuitenkin vuosi 1976 Lorenz-dominoi vuotta 1981, mutta vuosi 1981 yleistetysti Lorenz-dominoi vuotta 1976. Viiden prosentin riskitasolla myös vuosi 1985 Lorenz-dominoisi vuotta 1990, mutta vuosi 1990 yleistetysti Lorenz-dominoisi vuotta 1985. Näillä riskitasoilla siis yleistetty Lorenz-dominanssikriteeri preferoi tulojakaumaa, jon-

[31] Tässä hieman virheellisesti ei oteta päättelyssä tapahtuvaa kaksinkertaista simultaanista päättelytilannetta huomioon vaan pidetään sekä Lorenz- että yleistetty Lorenz -vertailuissa saatuja riskitasoja oikeina. Oikeampi ratkaisu olisi perustaa päättelyt Lorenz-ordinaattojen ja yleistettyjen Lorenz-ordinaattojen yhteisjakaumasta simuloituihin testisuureiden riskitasoihin. Tehdyllä yksinkertaistuksella on kuitenkin tuskin vaikutusta tässä esitettyihin tuloksiin.

Taulukko 13. Tavanomaisen ja tilastollisen päättelyn johtopäätökset kymmenen, viiden ja yhden prosentin riskitasoilla yleistetyistä Lorenz-vertailuista eri vuosien välillä.

Vuosi	Vuosi	Ei-tilastollinen analyysi	Riskitaso 0.10	Riskitaso 0.05	Riskitaso 0.01
1990	1985	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1990	1981	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1990	1976	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1990	1971	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1985	1981	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1985	1976	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1985	1971	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1981	1976	X_{GL}	$>_{GL}$	$>_{GL}$	$>_{GL}$
1981	1971	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$
1976	1971	$>_{GL}$	$>_{GL}$	$>_{GL}$	$>_{GL}$

ka suhteellinen eriarvoisuus on marginaalisesti suurempi kuin preferoidun. Tässä on kuitenkin syytä muistaa suhteelisen eriarvoisuuden tarkasteluiden yhteydessä käyty keskutelu, jossa argumentoitiin sen puolesta, että yhden prosentin riskitason käyttäminen olisi järkevintä tässä tutkimuksessa.

5.3.5 Keskiarvo - Lorenz-kriteeriin perustuvat hyvinvointivertailut

Keskiarvo - Lorenz-dominanssiin perustuvat vertailut edellyttävät, että parempana pidetyssä tulojakaumassa on vähintään yhtä suuri keskiarvo ja korkeintaan yhtä suuri suhteellinen eriarvoisuus kuin huonommassa jakaumassa. Aineistossa keskiarvot kasvavat kronologisessa järjestyksessä siten, että keskitulo oli vuonna 1971 kaikkein pienin ja vuonna 1990 kaikkein suurin. Tämä järjestys osoittautuu kaikissa vertailuissa tilastollisesti erittäin merkitseväksi. Tästä seuraa suoraan se, että vuosi 1971 on kaikkien muiden vuosien dominoima tällä kriteerillä, joka siis yleistettyä Lorenz-kriteeriä voimakkaammin ottaa tulojen jakautumisen huomioon. Vuosi 1985 myös dominoi tällä kriteerillä vuotta 1976. Muut tarkastelut jäävät tilastollisen kriteerin varaan.

Yhden prosentin riskitasoa käytettäessä keskiarvo - Lorenz-dominanssi toteutuu kaikissa vertailuissa kronologisessa järjestyksessä. Viiden ja kymmenen prosentin riskitasoja käyttäen ei näin vahvoja tuloksia voitaisi esittää, epäselviksi tapauksiksi jäisivät kuten tavallisten Lorenz-vertailujen yhteydessä vuosiparit 1990 ja 1985, 1990 ja 1981 sekä 1981 ja 1976. Näiden vuosien Lorenz-käyrien välisiä suhteita on käsitelty suhteellisen eriarvoisuuden tarkasteluita koskevas-

Taulukko 14. Pareittainen tilastollinen analyysi OECD-skaalattujen tulojen keskiarvo - Lorenz-dominanssille.

Vuosi	Vuosi	Positiivinen testisuure	Riskitaso	Negatiivinen testisuure	Riskitaso	χ^2-testisuure	Riskitaso
1990	1985	30.899*	< 0.01	-2.585	0.0468	1420.38	< 0.001
1990	1981	46.312*	< 0.01	-2.516	0.0568	3035.77	< 0.001
1990	1976	46.043*	< 0.01	-0.847	0.8802	3733.12	< 0.001
1990	1971	54.122*	< 0.01	-	-	7817.54	< 0.001
1985	1981	18.476*	< 0.01	-0.500	0.9886	538.78	< 0.001
1985	1976	22.164*	< 0.01	-	-	1009.93	< 0.001
1985	1971	33.428*	< 0.01	-	-	4206.76	< 0.001
1981	1976	7.386*	< 0.01	-2.527	0.0584	185.44	< 0.001
1981	1971	20.546*	< 0.01	-	-	1841.01	< 0.001
1976	1971	12.444*	< 0.01	-	-	708.63	< 0.001

Huomautuksia taulukkoon: *-merkintä testisuureen jäljessä tarkoittaa testisuureen perustuvan keskiarvojen väliseen vertailuun. Tässä taulukossa esitetyissä tarkasteluissa testisuureiden riskitasot eroavat pelkkien Lorenz-käyrien tarkasteluissa esitettyjen testisuureiden riskitasoista, koska simultaanisessa tarkastelutilanteessa on Lorenz-ordinaattojen lisäksi yksi ylimääräinen tarkasteltava suure: keskiarvo.

sa kappaleessa. Muissa vertailuissa kaikilla kolmella tavanomaisella riskitasolla päädytään samoihin johtopäätöksiin.

Tilastollinen keskiarvo - Lorenz-dominanssi osoittautui yllättävän usein toteutuvaksi empiiriseksi kriteeriksi Suomen aineistolla. Kaikkiaan kymmenestä vertailtavasta parista dominanssi toteutuu yhden prosentin riskitasolla kaikissa, viiden prosentin tasolla yhdeksässä ja kymmenen prosentin tasolla seitsemässä. Vastaavanlaisia tuloksia saivat myös Bishop et al. (1989b) Yhdysvaltain aineistolla, joissa 51:stä tehdystä vertailusta 42:ssa toteutui tilastollinen keskiarvo - Lorenz-dominanssi. He käyttivät kiinteätä viiden prosentin riskitasoa. Täten keskiarvo - Lorenz-dominanssi voidaan pitää empiirisesti käyttökelpoisena kriteerinä, kunhan Lorenz-käyrien leikkaamisista otetaan huomioon vain ne, jotka ovat tilastollisesti merkitseviä. Tätä voidaan verrata Shorrocksin (1983) maittaisten vertailuiden tuloksiin, jossa hän käyttämällä ei-tilastollista keskiarvo-Lorenz -kriteeriä, toteaa sen toteutuvan vain 53:ssa tapauksessa 190:stä mahdollisesta pareittaisesta vertailuista. Täten tilastollisen merkitsevyyden vaatiminen vertailun lähtökohdaksi vahvistaa tutkijan mahdollisuutta tehdä vertailuja tällä aika tiukalla hyvinvointikriteerillä.

Johtopäätöksenä näistä vertailuista on, että tulojen muutos vertailuvuodelta toisella aikavälillä 1971-1985 on ollut sekä tuloja kasvattavaa, että suhteellisia tuloeroja pienentävää. Tosin

Taulukko 15. Tavanomaisen ja tilastollisen päättelyn johtopäätökset kymmenen, viiden ja yhden prosentin riskitasoilla keskiarvo - Lorenz -vertailuista eri vuosien välillä.

Vuosi	Vuosi	Ei-tilas-tollinen analyysi	Riski-taso 0.10	Riski-taso 0.05	Riski-taso 0.01
1990	1985	X_{ML}	X_{ML}	X_{ML}	$>_{ML}$
1990	1981	X_{ML}	X_{ML}	$>_{ML}$	$>_{ML}$
1990	1976	X_{ML}	$>_{ML}$	$>_{ML}$	$>_{ML}$
1990	1971	$>_{ML}$	$>_{ML}$	$>_{ML}$	$>_{ML}$
1985	1981	X_{ML}	$>_{ML}$	$>_{ML}$	$>_{ML}$
1985	1976	$>_{ML}$	$>_{ML}$	$>_{ML}$	$>_{ML}$
1985	1971	$>_{ML}$	$>_{ML}$	$>_{ML}$	$>_{ML}$
1981	1976	X_{ML}	X_{ML}	$>_{ML}$	$>_{ML}$
1981	1971	$>_{ML}$	$>_{ML}$	$>_{ML}$	$>_{ML}$
1976	1971	$>_{ML}$	$>_{ML}$	$>_{ML}$	$>_{ML}$

Huomautus taulukkoon: Lyhenteellä ML tarkoitetaan keskiarvo-Lorenz -dominanssia (Mean-Lorenz).

suhteellisten tuloerojen muutos vuodesta 1976 vuoteen 1981 jää analyysissä hiukan epäselväksi. Jos muutoksia näiden vuosien välillä on tapahtunut suhteellisissa tuloeroissa, ovat ne olleet kuitenkin absoluuttisesti hyvin pieniä. Vuoden 1990 ja vuosien 1981 ja 1985 keskinäiset vertailut taas osoittavat tulojen kasvaneen vuoteen 1990 saakka ja tuloerojen tulojakauman alapäässä marginaalisesti pienentyneen. Täten voidaan kohtuullisen hyvällä syyllä sanoa hyvinvoinnin tulojakauman valossa kasvaneen suuremmat tulot - pienemmät suhteelliset tuloerot -kriteerillä koko aineistossa kronologisessa järjestyksessä.

5.3.6 Keskiarvo - absoluuttinen Lorenz-kriteerin perustuvat hyvinvointivertailut

Tämä kriteeri ei toteudu aineistossa ilman tilastollisia tarkasteluita kuin vuosien 1976 ja 1971 välillä. Tilastollinen tarkastelukaan ei muuta tätä johtopäätöstä: joko absoluuttiset Lorenz-käyrät leikkaavat merkitsevästi tai absoluuttinen Lorenz-dominanssi ja keskiarvojen suuruusjärjestys menevät eri suuntiin. Vuosina 1976-1990 keskitulo on kasvanut suhteellisen eriarvoisuuden ollessa suurinpiirtein vakio, joten absoluuttiset tuloerot ovat kasvaneet. Täten tämä kaikkein vaativin tarkasteltavista hyvinvointikriteereistä ei toteudu tällä periodilla.

Taulukko 16. Pareittainen tilastollinen analyysi OECD-skaalattujen tulojen absoluuttinen Lorenz - keskiarvo -dominanssille.

Vuosi	Vuosi	Positiivinen testisuure	Riskitaso	Negatiivinen testisuure	Riskitaso	χ^2-testisuure	Riskitaso
1990	1985	30.899*	< 0.01	-12.902	< 0.01	1441.04	< 0.001
1990	1981	46.312*	< 0.01	-16.793	< 0.01	3074.17	< 0.001
1990	1976	46.043*	< 0.01	-16.697	< 0.01	3745.40	< 0.001
1990	1971	54.122*	< 0.01	-14.208	< 0.01	7841.10	< 0.001
1985	1981	18.476*	< 0.01	-4.984	< 0.01	536.75	< 0.001
1985	1976	22.164*	< 0.01	-7.245	< 0.01	1013.52	< 0.001
1985	1971	33.428*	< 0.01	-6.498	< 0.01	4313.03	< 0.001
1981	1976	7.386*	< 0.01	-5.319	< 0.01	188.12	< 0.001
1981	1971	20.546*	< 0.01	-4.648	< 0.01	1877.49	< 0.001
1976	1971	12.444*	< 0.01	-	-	711.17	< 0.001

Huomautuksia taulukkoon: *-merkinnällä tarkoitetaan testisuureen perustuvan keskiarvojen väliseen vertailuun. χ^2-testisuureet ovat samoja kuin ryhmäkeskiarvojen ja yleistettyjen Lorenz-ordinaattojen vertailuissa, koska absoluuttiset Lorenz-ordinaatat ja keskiarvo saadaan yleistetyistä Lorenz-ordinaatoista täysiasteisella lineaarimuunnoksella.

Taulukko 17. Tavanomaisen ja tilastollisen päättelyn johtopäätökset kymmenen, viiden ja yhden prosentin riskitasoilla keskiarvo - absoluuttisen Lorenzdominanssivertailuista eri vuosien välillä.

Vuosi	Vuosi	Ei-tilastollinen analyysi	Riskitaso 0.10	Riskitaso 0.05	Riskitaso 0.01
1990	1985	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1990	1981	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1990	1976	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1990	1971	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1985	1981	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1985	1976	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1985	1971	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1981	1976	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1981	1971	X_{MAL}	X_{MAL}	X_{MAL}	X_{MAL}
1976	1971	$>_{MAL}$	$>_{MAL}$	$>_{MAL}$	$>_{MAL}$

Huomautus taulukkoon: Merkinnällä MAL tarkoitetaan keskiarvo - absoluuttinen Lorenzdominanssia.

Bishop et al. (1989b) Yhdysvaltain aineistossa toteavat tilastollisen keskiarvo - absoluuttinen Lorenz -vertailtavuuden toteutuvan 17 osavaltion kohdalta 51:stä osavaltiosta. Se miksi Suomen aineistolla vertailtavuus toteutuu suhteessa vieläkin harvemmin saattaa johtua suuremmista suhteellisista eroista jakaumien keskiarvojen välillä suomalaisella aikasarja-aineistolla kuin tuossa Yhdysvaltain poikkileikkausaineistossa. Johtopäätöksenä voidaan kuitenkin esittää, että keskiarvo - absoluuttinen Lorenz-dominanssi on empiirisesti harvoin toteutuva ominaisuus.

5.4 Empiirisen analyysin johtopäätökset

Menetelmällisesti tämä empiirinen analyysi osoittaa, että tilastollisen kriteerin käyttö lisää tutkijan mahdollisuuksia tehdä erilaisia dominanssirelaatiovertailuja eri tulojakaumien välillä. Tulosten tulkintojen kannalta ei tilastollinen analyysi ole pelkästään helpottava asia. Vaikka talousteorian kannalta oletettaisiin tutkimusasetelman olevan täysin kohdallaaan, eli että tarkasteltavana on juuri oikea tulokäsite ja käytössä juuri oikea ekvivalenssiskaala, niin tilastollinen hypoteesien testaus tuo yhden epävarmuustekijän tuloksiin, joka tavanomaisessa analyysissä oletetaan pois.

Tästä voidaan esimerkiksi ottaa vuosien 1990 ja 1985 Lorenz-käyrien vertailu. Ei-tilastollinen analyysi olisi päätynyt yksiselitteiseen johtopäätökseen että Lorenz-käyrät leikkaavat. Tällöin oltaisiin epärealistisesti oletettu, että otoksesta estimoidut Lorenz-käyrät olisivat kyseisten vuosien todellisia Lorenz-käyriä ilman mitään tilastollista epävarmuutta. Tilastollinen analyysi taas päätyy johtopäätökseen, että otamme noin neljän prosentin riskin olla väärässä väittäessämme vuoden 1985 Lorenz-käyrän olevan jossain vaiheessa vuoden 1990 käyrän yläpuolella ja toisaalta otamme noin seitsemän prosentin riskin olla väärässä väittäessämme, että vuoden 1990 käyrä on jossain vaiheessa vuoden 1985 käyrän yläpuolella. Siispä käyrät voivat olla ekvivalentteja keskenään tai vuosi 1985 Lorenz-dominoi vuotta 1990 tai käyrät leikkaavat. Selkeämmän tuloksen saaminen vaatisi isompaa aineistoa ja uutta analyysiä, mutta tällä aineistolla joudutaan tyytymään epäselvään johtopäätökseen. Jos kuitenkin eroja vuosien 1985 ja 1990 Lorenz-käyrissä esiintyy, niin ne ovat absoluuttisesti hyvin pieniä. Tästä syystä lienee turvallisinta pitää näiden vuosien Lorenz-käyriä ekvivalentteina.

Selkeään ja kaikilla kolmella käytetyllä riskitasolla kiistattomaan johtopäätökseen päädytään yhteensä 54:ssä pareittaisessa vertailussa kaiken kaikkiaan 60:stä suoritetusta vertailusta. Näistä dominanssirelaatioita oli 44 ja kiistattomia käyrien leikkaamisia tai keskiarvorelaatioiden menemistä ristiin käyrien dominanssirelaatioiden kanssa oli 10. Vain kuudessa tapauksessa

Taulukko 18. Yhteenveto tavanomaisen ja tilastollisen analyysin johtopäätöksien luonteesta.

Kriteeri	Suoritettuja vertailuja	Käyrien leikkaamisia tavallisella kriteerillä	Käyrien leikkaamisia tilastollisella kriteerillä	Dominansseja tilastollisella kriteerillä	Epäselviä tuloksia tilastollisella kriteerillä
Lorenz-dominanssi	10	5	0	7	3
Absoluuttinen Lorenz	10	3	1	9	0
Ryhmäkeskiarvo	10	3	0	10	0
Yleistetty Lorenz	10	1	0	10	0
Keskiarvo - Lorenz	10	5	0	7	3
Keskiarvo - absoluuttinen Lorenz	10	9	9	1	0

Huomautus taulukkoon: Tavanomaisella kriteerillä on vain kaksi vaihtoehtoa, joko dominanssi toteutuu tai sitten käyrät leikkaavat. Tavanomaisella kriteerillä toteutuneiden dominanssien määrän saa siis taulukosta vähentämällä vertailujen kokonaismäärästä käyrien leikkaamisten lukunmäärän. Esimerkiksi Lorenz-dominanssivertailuissa tavanomaisella kriteerillä havaittiin $10 - 5 = 5$ dominanssirelaatiota.

ei saatu selkeää tulosta ja nämä kaikki liittyivät tavallisten Lorenz-käyrien vertailuihin joko Lorenz-dominanssi- tai keskiarvo - Lorenz-dominanssirelaatioiden yhteydessä. Tilastollisen kriteerin käyttö nosti havaittujen kiistattomien dominanssirelaatioiden määrän 35:stä 44:ään. Täten tilastollisen analyysin käyttö paransi kaikissa vertailuissa mahdollisuuksia havaita mielenkiintoisia relaatioita aineistossa noin neljänneksellä.

Suomalaista tulojakaumaa koskevat tulokset osoittavat ensinnäkin periodilla 1971-1990 taloudellisen kasvun tapahtuneen siten, että suhteellinen eriarvoisuus on pienentynyt vuosien 1971 ja 1976 välillä ja vuoden 1976 jälkeen pysynyt suhteellisen vakiona, kuitenkin hiukan pienentyen. Absoluuttinen eriarvoisuus taas on kasvanut vuoden 1976 jälkeen, jolloin se oli alimmillaan. Anonyymi Pareto- ja yleistetty Lorenz-kriteerien mielessä ovat suomalaiset tulojakaumat parantuneet tarkasteluperiodilla kronologisessa järjestyksessä. Myöskin keskiarvo - Lorenz-kriteerillä parantumista on tapahtunut, tosin kronologisen järjestyksen toteutuminen on riippuvaista tilastollisen merkitsevyyden kriteerinä pidetystä riskitasosta. Keskiarvo - absoluuttinen Lorenz-kriteerillä ei suomalaisessa tulojakaumassa toteudu dominanssi kuin vuosien 1976 ja 1971 välillä.

Näiden tulosten tulkinnassa on kuitenkin otettava huomioon tutkimuksen mahdolliset virhelähteet. Aineistoa voidaan pitää niin luotettavana kuin se yleensä rekisteri- ja kyselypohjaiselle aineistolle on mahdollista. Yleisesti tunnettu ongelma kuitenkin on, että kotitalous- ja tulotiedusteluiden vastauskato on kaikkein suurinta kaikkein köyhimmissä ja rikkaimmissa kotitalouksissa. Tämä on tietenkin kiusallista eriarvoisuustarkasteluiden kannalta, mutta voidaan toivoa, että vastauskatotodennäköisyyksiin perustuva painotus estää ainakin suurten systemaattisten virheiden synnyn.

Aineiston tilastollista edustavuutta merkittävämpi kysymys on se, onko tässä käytetty tulokäsite ihmisten hyvinvoinnin kannalta kaikkein relevantein vaihtoehto ja onko se oikein mitattu riittävällä tarkkuudella aineistossa. Myöskin valittuun ekvivalenssiskaalan liittyy ongelmia, koska sen voi parhammaissakin tapauksessa ainoastaan toivoa olevan hyvä approksimaatio oikealle ekvivalenssiskaalalle. Toisella ekvivalenssiskaalalla saatavat tulokset olisivat voineet olla erilaisia.

Jotta ekvivalenssiskaalan vaikutuksen merkittävyyttä tuloksiin pystyttäisiin arvioimaan, tehtiin tämän tutkimuksen yhteydessä myös laskelmat yksinkertaisella per henki -skaalalla, jossa kotitalouden tulot jaettiin tasan jokaiselle kotitalouden jäsenelle sekä per kotitalous -skaalalla, jossa tarkasteltiin kotitalouksia kiinnittämättä mitenkään huomiota niiden kokoeroihin. Näiden tarkasteluiden tuloksia ei kuitenkaan raportoida muuten kuin graafisessa liitteessä. Per henki -skaalalla laskettuna tulokset olivat kvalitatiivisesti samanlaisia kuin OECD-skaalallakin, mutta per kotitalous -skaalalla saatiin hyvinkin poikkeavia tuloksia. Koska kuitenkin per kotitalous -skaalaa voidaan pitää ei-relevanttina ääripään skaalana, koska se ei ota kotitalouksien kokoeroja mitenkään huomioon, voidaan näiden kahden tarkistuslaskelman tuloksia pitää tässä esitettyjä tuloksia vahvistavina: kahdella hyvinvoinnin kannalta relevantilla ekvivalenssiskaalalla saadaan hyvin samantapaisia tuloksia.

Empiirisen aineiston johtopäätöksiä tulkittaessa on otettava myös huomioon, että tässä esimerkissä on pyritty ainoastaan tulojakauman muutosten kuvaamiseen. Laskelmat edustavat siis eräänlaista sofistikoitunutta mittaamista. Tulojakaumassa tapahtuneiden muutosten syitä tai niiden suhdanne- tai rakenneluonteisuutta ei ole pyritty analysoimaan. Tästä syystä julkisen vallan eri toimenpiteiden vaikutuksia saatuihin tuloksiin ei pyritä tässä arvioimaan. Politiikkatekijöiden sekä suhdanne- ja rakennetekijöiden vaikutuksen selvittäminen tulojakaumaan edellyttäisi laajempaa analyysiä, jossa selvitettäisiin eri tuloerien vaikutusta kokonaistulojakaumaan sekä niiden kehitystä yli ajan. Havaintoperiodien pitkä väli (viisi tai neljä vuotta) saattaisi tällä aineistolla vaikeuttaa tällaisen analyysin tekemistä, koska käytössä ei ole ajassa lähekkäisiä poikkileikkausaineistoja, joilla suhdannesyklin eri osien[32] vaikutuksia pystyttäisiin identifioimaan ja erottelemaan ne rakenne- ja poliitikkamuutoksien vaikutuksista. Tämän analyysin perusteella pystytään siis ainoastaan havaitsemaan tulojakaumassa tapahtunutta kehitystä, ei analysoimaan kehityksen syitä.

[32] Riihelän ja Sullströmin (1993, 10) mukaan kaikki kotitaloustiedustelut on tehty matalasuhdannevuosina.

6 Lopuksi

Tutkimuksen lopuksi lienee hiukan syytä kommentoida tässä tutkimuksessa saatuja tuloksia ja tutkimusaiheessa tehtyjä rajauksia. Tutkimuksen tarkoituksena oli esitellä sekä taloudellinen että tilastollinen teoria, joka soveltuisi eriarvoisuus- ja hyvinvointitarkasteluihin reaalimaailman tuloaineistojen yhteydessä, mutta joka vaatisi mahdollisimman vähän tulosten validiutta heikentäviä oletuksia. Tästä syystä tässä tutkimuksessa esitetty teoria on tavallaan kaksinkertaisesti ei-parametrinen. Taloudellisen teorian puolella ei oteta kantaa hyvinvointifunktioiden tai eriarvoisuusindeksien funktiomuotoihin, vaan käsitellään laajoja funktioluokkia ja johdetaan näihin liittyviä dominanssikriteerejä. Tilastollisen teorian puolella taas ei tehdä mitään parametrista oletusta tulojakauman Lorenz-käyrän muodosta vaan esitetään asymptoottinen teoria, joka pätee hyvin yleisten ja tutkimusasetelman kannalta järkevien oletusten vallitessa.

Talousteoreettisessa osuudessa tässä tutkimuksessa oleellisessa asemassa ovat Lorenz-käyrät ja eri jakaumien eri tyyppisten Lorenz-käyrien väliset dominanssirelaatiot. Nämä osoittautuvat myös empiirisessä osuudessa käyttökelpoisiksi työvälineiksi käytännön tutkimuksen kannalta. Erilaisia Lorenz-käyriä käsiteltäessä ollaan kuitenkin rajoitettu puhtaiden dominanssirelaatioiden tarkasteluun, vaikka kirjallisuudessa on esitetty myös erilaisia vertailukriteerejä tilanteisiin (katso esim. Lambert 1989, 73-82), jossa Lorenz-käyrät leikkaavat. Tätä "kolmannen asteen" stokastisen dominanssin teoriaa ei ole tässä tutkimuksessa käsitelty. Lisäksi Lorenz-käyrien lähisukulaiset konsentraatiokäyrät on jätetty kokonaan tämän tutkimuksen ulkopuolelle.

Menetelmällisessä mielessä tässä tutkimuksessa saadaan positiivinen tulos. Tilastollisen epävarmuuden huomioiminen Lorenz-käyriin liittyvissä tarkasteluissa lisää tutkijan mahdollisuutta havaita dominanssirelaatioita. Tämä ilmenee siitä tosiasta, että tilastollista kriteeriä käyttäen luvun viisi vertailuissa todetaan 44 puhdasta dominanssirelaatiota. Ei-tilastollista kriteeriä käyttäen olisi dominanssirelaatioiden määrä ollut vain 35. Näin siis tilastollisen kriteerin käyttö johti vahvempiin ja selkeämpiin tuloksiin kuin ei-tilastollisen kriteerin käyttö.

Empiirisen osan päätulokset voidaan ilmaista lyhyesti. Vuodesta 1971 vuoteen 1990 on keskimääräinen tulotaso noussut selvästi ja suhteellinen eriarvoisuus on ollut vuoden 1976 jälkeen kohtuullisen vakio. Vuosi 1971 on suhteellisen eriarvoisuuden tarkasteluissa huomattavasti huonompi kuin muut tarkasteluvuodet. Absoluuttinen eriarvoisuus on ollut pienimmillään vuonna 1976, jonka jälkeen se on kasvanut.

Empiirisessä osassa saadaan myös tulos, että OECD-ekvivalenssiskaalalla tehtyjen tarkasteluiden kvalitatiiviset johtopäätökset säilyvät samankaltaisina, jos tarkastelut tehdään per henki -ekvivalenssiskaalalla. Tämä vahvistaa saatujen tulosten yleistettävyyttä, ainakin kahdella relevantilla ekvivalenssiskaalalla saadut tulokset ovat hyvin samanlaisia. Per kotitalous -tarkasteluissa tosin johtopäätökset muuttuisivat rajusti.

Tämän tutkimuksen tilastollisessa osuudessa esitetään simulointiin perustuva parannus tällä hetkellä standardiaseman saavuttaneeseen menettelyyn dominanssitestauksessa. Tämän testin pienen otoksen ominaisuuksien selvittäminen simuloimalla ja sen ominaisuuksien vertailu standarditestiin olisi mielenkiintoinen jatkotutkimuksen aihe. Myöskin Lorenz-käyriin läheisesti liittyvien konsentraatiokäyrien tilastollisten ominaisuuksien tarkastelu on potentiaalisesti mielenkiintoinen tutkimusaihe.

Mahdollisia empiirisiä jatkotutkimusaiheita tämän tutkimuksen ympäriltä on useita. Mielenkiintoista olisi esimerkiksi tarkastella tämän tutkimuksen aineistolla suhteellisen eriarvoisuuden dekomponoimissa joko tulolajeittain tai populaatioryhmittäin. Tämä sekä eri tuloerien ja eri tavalla määriteltyjen tulokäsitteiden vertailu Lorenz- ja konsentraatiokäyräkehikossa voisi antaa informaatiota tulojakaumassa tapahtuneiden muutosten syistä. Näitä aihepiirejä on käsitellyt mm. Uusitalo (1988), mutta tilastollisen aspektin mukaan tuominen näihin tarkasteluihin voisi tuoda lisäinformaatiota. Uudemman aineiston tai tiheammällä havaintovälillä kerätyn tuloaineiston analysointi olisi myös kiinnostava potentiaalisen jatkotutkimuksen aihe, koska tiheämmin kerätyllä aineistolla päästäisiin tarkastelemaan suhdannevaihtelun aiheuttamia muutoksia tulojakaumassa ja uudemmalla aineistolla taas saataisiin nykytilanteen kannalta relevantimpaa tietoa.

7 Lähteet

ATKINSON, A. B. (1971): On the Measurement of Inequality. Journal of Economic Theory 2, 244-263.

ATKINSON, A. B. - BOURGUIGNON, F. (1987): Income Distribution and Differences in Needs. Teoksessa FEIWEL G. R.: Arrow and the Foundations of the Theory of Economic Policy. Macmillan Academic and Professional Ltd, London.

ATKINSON, A. B. - RAINWATER, L. - SMEEDING, T. M. (1995): Income Distribution in OECD Countries. OECD Social Policy Studies No. 18, Paris.

BEACH, C. M. - DAVIDSON, R. (1983): Distribution-Free Statistical Inference with Lorenz Curves and Income Shares. Review of Economic Studies 4:163, 723-735.

BEACH, C. M. - KALINSKI, S. F. (1986): Lorenz Curve Inference with Sample Weights: an Application to the Distribution of Unemployment Experience. Applied Statistics 35:1, 38-45.

BEACH, C. M. - RICHMOND, J. (1985): Joint Confidence Intervals for Income Shares and Lorenz Curves. International Economic Review 26:2, 439-450.

BISHOP, J. A. - CHAKRABORTI, S. - THISTLE, P. D. (1994): Relative Inequality, Absolute Inequality, and Welfare: Large Sample Test for Partial Orders. Bulletin of Economic Research 46:1, 41-59.

BISHOP, J. A. - FORMBY, J. P. - THISTLE, P. D. (1989a): Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves. Review of Economics and Statistics 71, 725-727.

BISHOP, J. A. - FORMBY, J. P. - THISTLE, P. D. (1989b): Statistical Inference, Income Distributions, and Social Welfare. S. 49-81. Teoksessa SLOTTJE, D. J. (ed.): Research on Economic Inequality. JAI Press, London.

BOX, G. E. P - TIAO, G. C. (1973): Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, Massachusetts.

BUHMAN, B. - RAINWATER, L. - SCHMAUS, G. - SMEEDING, T. M. (1988): Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates Across Ten Countries Using the Luxembourg Income Study (LIS) Database. Review of Income and Wealth 34,115-142.

COULTER, F. A. E. - COWELL, F. A. - JENKINS, S. P. (1992): Differences in Needs and Assessment of Income Distributions. Bulletin of Economic Research 44:2, 77-124.

COWELL, F. A. (1984): The Structure of American Income Inequality. The Review of Income and Wealth 30, 351-375.

COWELL, F. A. (1985): 'A Fair Suck of the Sauce Bottle' or What Do You Mean by Inequality? Economic Record, 567-579.

DASGUPTA, P. - SEN, A. - STARRET, D. (1973): Notes on the Measurement of Inequality. Journal of Economic Theory 6, 180-187.

GASWIRTH, J. L. (1972): The Estimation of the Lorenz Curve and Gini Index. Review of Economics and Statistics 54, 306-316.

GLEWWE, P. (1991): Household Equivalence Scales and the Measurement of Inequality: Transfers from the Poor to the Rich Could Decrease Inequality. Journal of Public Economics 44, 211-26.

GOLDIE, C. M. (1977): Convergence Theorems for Empirical Lorenz Curves and their Inverses. Adv. Appl. Prob. 9, 765-791.

JENKINS, S. P. - LAMBERT, P. J. (1993): Ranking Income Distributions when Needs Differ. Review of Income and Wealth 39:4, 337-356.

KAUR, A. - RAO, B. L. S. P. - SINGH, H. (1994): Testing for Second-Order Stochastic Dominance of Two Distributions. Econometric Theory 10, 849-866.

KODDE, D. A. - PALM, F. C. (1986): Wald Criteria for Jointly Testing Equality and Inequality Restrictions. Econometrica 54:5, 1243-1248.

LAAKSONEN, S. (1988): Katovirheen korjaus kotitalousaineistossa. Tilastokeskus, Tutkimuksia 147. Helsinki.

LAMBERT, P. J. (1989): The Distribution and Redistribution of Income: a Mathematical Analysis. Basil Blacwell, Oxford.

MAASOUMI, E. (1995): Empirical Analyses of Inequality and Welfare. Paper Presented at Econometric Society's World Conference, Tokyo.

MARSHALL, A. W. - OLKIN, I. (1979): Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

McDONALD, J. B. (1984): Some Generalised Functions for the Size Distribution of Income. Econometrica 52:3, 647-663.

NYGÅRD, F. - SANDSTRÖM, A. (1981): Measuring Income Inequality. Almqvist & Wicksell International. Stockholm.

RAO, C. R. (1973): Linear Statistical Inference and Its Applications. Second Edition. John Wiley & Sons, New York.

RAWLS, J. (1973): A Theory of Justice. Oxford University Press, Oxford.

RIIHELÄ, M. - SULLSTRÖM, R. (1993): Life-cycle Profiles of Consumption and Income: An Analysis of the Finnish Household Budget Surveys in 1966-90. VATT-Keskustelualoitteita no. 50, Helsinki.

ROTSCHILD, M. - STIGLITZ, J. E. (1973): Some Futher Results on the Measurement of Inequality. Journal of Economic Theory 6, 188-204.

SAVIN, N. E (1984): Multiple Hypothesis Testing. S. 827-879. Teoksessa GRILICHES, Z. - INTRILIGATOR, M. D. (eds.): Handbook of Econometrics. Volume II. North-Holland, Amsterdam.

SCHADER, M. - SCHMID, F. (1994): Fitting Parametric Lorenz Curves to Grouped Income Distributions - a Critical Note. Empirical Economics 19, 361-370.

SEN, A. (1979): Survey on Welfare and Real Income. Journal of Economic Literature 17, 1-45.

SHORROCKS, A. F. (1980): The Class of Additively Decomposable Inequality Measures. Econometrica 48:3, 613-625.

SHORROCKS, A. F. (1983): Ranking Income Distributions. Economica 50, 3-17.

STIGLER, S. M. (1974): Linear Functions of Order Statistics with Smooth Weight Functions. Annals of Statistics 2:4, 676-693.

STOLINE, M. R. - URY, H. K. (1979): Tables of the Studentized Maximum Modulus Distribution. Technometrics 21:1, 87-93.

SUONIEMI, I. (1994): Non-parametric Estimation of Lorenz Curves Using Locally Weighted Regression. VATT-Keskustelualoitteita no. 59, Helsinki.

THISTLE, P. D. (1989): Ranking Distributions with Generalised Lorenz Curves. Southern Economic Journal 56:1, 1-12.

UUSITALO, H. (1988): Muuttuva tulonjako: Hyvinvointivaltion ja yhteiskunnan rakennemuutosten vaikutukset tulonjakoon 1966-1985. Tilastokeskus, Tutkimuksia 148. Helsinki.

VARTIA, P. L. I. - VARTIA, Y. O. (1980): Description of Income Distribution by the Scaled F Distribution Model. Teoksessa KLEVMARKEN, N. A. - LYBECK, J. A. (eds.): The Statics and Dynamics of Income. Tieto Ltd, Bristol.

XU, K. (1995): Asymptotically Distribution-Free Statistical Test for Generalised Lorenz Curves: an Alternative Approach. Paper Presented at Econometric Society's World Conference, Tokyo.

YITZHAKI, S. (1983): On an Extension of the Gini Index. International Economic Review 24, 617-628.

Liite 1: Eräitä lukuun kolme liittyviä todistuksia

Tässä esitetyt todistukset ovat hyvinvointilauseiden osalta Shorrocksin (1983) artikkelista, joskaan notaatio ja nimitykset eivät ole alkuperäisiä. Indeksejä koskevat todistukset ovat mukailtu artikkeleista Dasgupta et al. (1973) ja Shorrocks (1983). Näissä lähteissä ei kuitenkaan eksplisiittisesti puhuta indekseistä, mutta todistustekniikka on näille artikkeleille ominaista ja kyseiset ominaisuudet on indekseillä todistettu useassa eri lähteessä. Lauseen 5 todistukset on muokattu lähteistä Shorrocks (1983) ja Bishop et al. (1989b).

Korollaarin 1 todistus sivulta 17:
(Lorenz-dominanssi on yhtäpitävää pienemmän suhteellisen eriarvoisuuden kanssa.)

i) Riittävyys:
Olkoon $X \geq_L X'$ ja $X'' = \mu_X/\mu_{X'} \cdot X'$. Tällöin $L(p; X') = L(p, X'')$ ja suhteellisen eriarvoisuuden indeksin määritelmän mukaan $I_R(X') = I_R(X'')$. Tällöin $X \geq_L X''$ ja $\mu_X = \mu_{X''}$, joten suhteellisen eriarvoisuusindeksin Schur-konveksisuudesta seuraa $I_R(X) \leq I_R(X'')$, joten $I_R(X) \leq I_R(X')$.

ii) Välttämättömyys:
Olkoon $I(X) \leq I(X')$ kaikilla suhteellisen eriarvoisuuden indekseillä. Funktiot $G_k(X) = 1-(\sum_{i=1}^{k} x_{(i)}/(k \cdot \mu))$, $k = 1, 2 \ldots n-1$ kuuluvat suhteellisen eriarvoisuuden indeksien joukkoon. Tästä helposti nähdään, että ehto $G_k(X) \leq G_k(X') \, \forall k$ on yhtäpitävä ehdon $X \geq_L X'$ kanssa.

Lauseen 3 todistus sivulta 19:
(Keskiarvo - Lorenz-dominanssi on yhtäpitävää kaikkien skaalaparannusta preferoivien hyvinvointifunktioiden dominanssin kanssa.)

i) Riittävyys:
Olkoon $\mu_X \geq \mu_{X'}$, $X \geq_L X'$ ja $X'' = \mu_X/\mu_{X'} \cdot X'$. Tällöin $W(X) \geq W(X'')$ koska $X \geq_L X''$ ja $\mu_X = \mu_{X''}$. Toisaalta $W(X'') \geq W(X')$, koska hyvinvointifunktio W preferoi skaalaparannusta, joten $W(X) \geq W(X')$.

ii) Välttämättömyys:
Olkoon $W(X) \geq W(X')$ kaikilla skaalaparannusta preferoivilla hyvinvointifunktioilla. Lisäksi olkoon $G(X) = \mu^\alpha f(X/\mu)$, jossa $f(\cdot)$ on Schur-konkaavi funktio ja $\alpha \geq 0$. G kuuluu skaalaparannusta preferoiviin hyvinvointifunktioihin. Valinnalla $f(X/\mu) \equiv 1$ seuraa ehdosta $G(X) \geq G(X')$ ehto $\mu_X \geq \mu_{X'}$. Toisaalta valinnalla $\alpha = 0$ saamme ehdon $f(X/\mu_X) \geq f(X'/\mu_{X'})$ kai-

killa Schur-konkaaveilla funktiolla $f(\cdot)$. Tämä taas on yhtäpitävää ehdon $X/\mu_X \geq_L X'/\mu_{X'}$ kanssa, joka taas on yhtäpitävä ehdon $X \geq_L X'$ kanssa.

Lauseen 4 todistus sivulta 19 :

(Keskiarvo - absoluuttinen Lorenz-dominanssi on yhtäpitävää kaikkien tasaparannusta preferoivien hyvinvointifunktioiden dominanssin kanssa.)

i) Riittävyys:

Olkoon $\mu_X \geq \mu_{X'}$ ja $X \geq_{AL} X'$. Lisäksi olkoon $X'' = X' + (\mu_X - \mu_{X'}) \cdot 1_n$. Tällöin $W(X'') \geq W(X')$, koska W preferoi tasaparannusta. Toisaalta $AL(p; X') = AL(p; X'')$ absoluuttisen Lorenz-käyrän ominaisuuksien mukaan. Koska $X \geq_{AL} X''$ ja $\mu_X = \mu_{X''}$, niin $X \geq_L X''$. Tästä taas seuraa, että $W(X) \geq W(X'')$ kaikilla Schur-konkaaveilla hyvinvointifunktiolla, joten $W(X) \geq W(X')$.

ii) Välttämättömyys:

Olkoon $W(X) \geq W(X')$ kaikilla tasaparannusta preferoivilla hyvinvointifunktioilla. Lisäksi olkoon $G_{\alpha,\lambda}(X) = \mu^\alpha f(X + (\lambda - \mu) \cdot 1_n)$, jossa $\alpha \geq 0$, $\lambda > 0$ ja $f(\cdot)$ on Schur-konkaavi. Parametri λ, jonka on oltava riittävän suuri, on mukana vain takaamassa, että funktion f argumenttivektori on positiivinen. Tällöin $G_{\alpha,\lambda}$ on tasaparannusta preferoiva hyvinvointifunktio. Valitsemalla $f(X) \equiv 1$ saamme ehdon $\mu_X \geq \mu_{X'}$.
Toisaalta valitsemalla $\alpha = 0$ saamme ehdon $f(X + (\lambda - \mu_X) \cdot 1_n) \geq f(X' + (\lambda - \mu_{X'}) \cdot 1_n)$ kaikilla Schur-konkaaveilla funktioilla f. Tämä taas on yhtäpitävää ehdon $(X + (\lambda - \mu_X) \cdot 1_n) \geq_L (X' + (\lambda - \mu_{X'}) \cdot 1_n)$ kanssa, koska molempien muuttujien keskiarvo on yhtä suuri ($= \lambda$). Täten kaikilla k:n arvoilla $\sum_{i=1}^k x_{(i)}/n + k/n(\lambda - \mu_X) \geq \sum_{i=1}^k x'_{(i)}/n + k/n(\lambda - \mu_{X'})$. Supistamalla tuosta muodosta λ-termit pois, päädymme absoluuttisen Lorenz-dominanssin määritelmään.

Korollaarin 2 todistus sivulta 19 :

(Absoluuttinen Lorenz-dominanssi on yhtäpitävää pienemmän absoluuttisen eriarvoisuuden kanssa.)

i) Riittävyys:

Olkoon $X \geq_{AL} X'$. X ja X' voidaan transformoida keskiarvoltaan yhtä suuriksi lisäämällä niihin sopivat vakiot. Tämä pitää niiden absoluuttiset Lorenz-käyrät muuttumattomina ja toisaalta ei myöskään vaikuta absoluuttisen eriarvoisuuden indekseihin. Tämän jälkeen, toteamalla että $X \geq_{AL} X'$ on yhtäpitävää ehdon $X \geq_L X'$ kanssa, jos $\mu_X = \mu_{X'}$, ja käyttämällä absoluuttisen eriarvoisuuden indeksin määritelmästä Schur-konveksisuutta seuraa $I(X) \leq I(X')$.

Liite 1: Eräitä lukuun kolme liittyviä todistuksia (jatkoa)

ii) Välttämättömyys:

Olkoot $I(X) \leq I(X')$ kaikilla absoluuttisilla eriarvoisuuden indekseillä. Lisäksi olkoon $G_k(X) = k/n \cdot \mu - 1/n \cdot \sum_{i=1}^{k} x_{(i)}$. Tällöin G_k:t ovat absoluuttisen eriarvoisuuden indeksejä. Ehdosta $\forall\, k: G_k(X) \leq G_k(X')$ seuraa suoraan $X \geq_{AL} X'$.

Lauseen 5 todistus sivulta 19.

1) Ensimmäisen kertaluvun stokastinen dominanssi implikoi yleistetyn Lorenz-dominanssin: $X \geq_{AP} Y \Leftrightarrow \forall p \in [0,1] : X(p) \geq Y(p) \Rightarrow \forall p \in [0,1] : \int_0^p X(P)dP \geq \int_0^p Y(P)dP$.

2) Keskiarvo - Lorenz-dominanssi implikoi yleistetyn Lorenz-dominanssin: Yleistetty Lorenz-käyrä voidaan kirjoittaa muodossa $GL(p) = \mu L(p)$, tulos seuraa suoraan tuosta muodosta.

3) Keskiarvo - absoluuttinen Lorenz-dominanssi implikoi keskiarvo - Lorenz-dominanssin: Keskiarvodominanssi on itsestään selvä. Olkoot nyt $\mu_X \geq \mu_Y$ ja $X \geq_{AL} Y$. Tarkastellaan muuttujaa $Y' = Y + (\mu_X - \mu_Y) \cdot 1_n$. On helppo todeta ensinnäkin, että $Y' \geq_L Y$ eli tasaparannus pienentää suhteellista eriarvoisuutta. Toisaalta Y' absoluuttinen Lorenz-käyrä on sama kuin Y:n, joten $X \geq_{AL} Y'$. Toisaalta koska $\mu_X = \mu_{Y'}$, niin $X \geq_L Y'$, joten $X \geq_L Y' \geq_L Y$.

4) Anonyymi Pareto-kriteeri ei ole riittävä eikä välttämätön ehto keskiarvo - Lorenz- tai keskiarvo - absoluuttinen Lorenz-dominanssille: Tarkastellaan tulojakaumia X ja Y, jotka ovat muuten identtisiä, mutta X:ssä rikkain saa enemmän tuloja kuin Y:ssä. Tällöin $X \geq_{AP} Y$, mutta $Y \geq_L X$ ja $Y \geq_{AL} X$. Täten anonyymi Pareto-kriteeri ei implikoi Lorenz- tai absoluuttinen Lorenz-dominanssia. Toisaalta, jos otetaan taas tarkastelun kohteeksi tulojakauma X ja tehdään tulonsiirto toisiksi rikkaimmalta rikkaimmalle, siten että rikkausjärjestys tulojakaumassa ei kuitenkaan muutu, ja kutsutaan tätä jakaumaa Y:ksi. Selvää on, että $X \geq_L Y$ ja $X \geq_{AL} Y$ ja lisäksi $\mu_X \geq \mu_Y$. Kuitenkin nyt ei päde $X \geq_{AP} Y$, sillä rikkain saa vähemmän tuloja jakaumassa X kuin jakaumassa Y. Täten keskiarvo - Lorenz- tai keskiarvo - absoluuttinen Lorenz-dominanssi ei implikoi anonymiä Pareto-dominanssia.

Lauseen 5 kohdista 1-3 on vielä todettava, että implikaatiot eivät päde toiseen suuntaan. Sopivien vastaesimerkkien keksimien on mahdollista.

Liite 2: Eräitä lukuun neljä liittyviä todistuksia

Taloustieteellisessä kirjallisuudessa ensimmäisen kerran Lorenz-ordinaattojen asymptoottisen jakauman ovat esittäneet Beach ja Davidson (1983). He myös olivat ensimmäiset, jotka esittivät Lorenz-ordinaattojen välisen kovarianssirakenteen ja toivat esille sen, että Lorenz-ordinaattojen välinen kovarianssimatriisi voidaan myös estimoida ei-parametrisesti. Tässä liitteessä kuitenkin otetaan lähtökohdaksi Stiglerin (1974) artikkeli, joka käsittelee yleistä järjestystunnuslukujen lineaarikombinaatioiden teoriaa. Tälle on perusteena se, että tekijän mielestä Stiglerin tulosten perusteella yleistettyjen Lorenz-ordinaattojen asymptoottisen jakauman johto on helpompaa kuin käyttämällä Beachin ja Davidsonin lähetysmistapaa. Tässä liitteessä käytetään Stiglerin valmiita tuloksia ja osoitetaan, että yleistetyt Lorenz-ordinaatat kuuluvat niiden otossuureiden luokkaan, joita Stiglerin tulokset koskevat. Etuna tästä menettelystä voidaan todeta se, että yleistettyjen Lorenz-ordinaattojen asymptoottinen multinormaalisuus saadaan helposti todettua reunajakaumien asymptoottisen normaalisuuden lisäksi.

Käytetyt aputulokset

Olkoot $x_1, x_2, x_3 \ldots, x_n$ yksinkertainen satunnaisotos jostain todennäköisyysjakaumasta ja olkoot $x_{(1)}, x_{(2)}, x_{(3)} \ldots, x_{(n)}$ vastaava järjestetty otos. Olkoot mielenkiinnon kohteena tunnusluvut, jotka kuuluvat luokkaan

$$S_n = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) x_{(i)}, \tag{1}$$

missä funkio J on niin sanottu paino- eli kernel-funktio. Olkoot seuraavat ehdot voimassa:

i) J on rajoitettu funktio ja jatkuva melkein kaikkialla. Funktiolla ei J saa olla yhteisiä epäjatkuvuuskohtia satunnaismuuttuja X:n fraktiilifunktion kanssa.
ii) Satunnaismuuttujalla X on äärellinen varianssi.

Tällöin otossuure S_n on asymptoottisesti normaalinen odotusarvolla[1]

$$E(S) = \lim_{n\to\infty} E(S_n) = \int_0^1 J(P)X(P)dP \tag{2}$$

ja varianssilla

$$\sigma^2(S) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F(x))J(F(y))[F(\min(x,y)) - F(x)F(y)]dxdy. \tag{3}$$

(Stigler 1974, 677-684.)

[1] Täsmällisemmin sanottuna $\sqrt{n}(S_n - E(S))$ on asymptoottisesti normaalinen parametrein 0 ja $\sigma^2(S)$. Jatkossa käytetään myös hiukan virheellistä ilmaisua asymptoottinen varianssi, kun tarkoitetaan suuretta $\sigma^2(S)$ ja merkintää Var(S), kun tarkoitetaan suuretta $n\sigma^2(S)$.

Liite 2: Eräitä lukuun neljä liittyviä todistuksia (jatkoa)

Stigler osoittaa myös (s. 688), että otossuureilla muotoa

$$(4) \qquad S'_n = \frac{1}{n}\sum_{i=1}^{n} J_n\left(\frac{i}{n+1}\right)x_{(i)},$$

jossa funktio J_n on rajoitettu ja suppenee funktiota J kohti, on sama asymptotiikka kuin suureella S_n. Riittävä rajankäynnin muoto funktiolle J_n on, että jokaista funktion J jatkuvuuspistettä p kohden on olemassa p:n avoin ympäristö, jossa pätee tasaisesti $J_n(u) \rightarrow J(u)$.

Yleistettyn Lorenz-käyrän estimaattori

Yleistetyn Lorenz-käyrän estimaattori (kaava 7 sivulta 29) kuuluu otossuureiden S' luokkaan. Kirjoitetaan estimaattori muotoon

$$(5) \qquad \hat{G}_j = \frac{1}{n}\sum_{i=1}^{n} I^{p_j}(i/n)x_{(i)},$$

missä funktio I^{p_j} on niin sanottu Heavisiden funktio, joka palauttaa arvon 1, jos sen argumentti $\leq p_j$, ja arvon 0 muulloin. Tuosta muodosta voidaan todeta, että siinä

$$(6) \qquad J_n(u) = I^{p_j}\left(\frac{n+1}{n}u\right),$$

jolle pätee rajankäyntiehto $J_n(u) \rightarrow J(u) = I^{p_j}(u)$ ylläesitetyssä mielessä.[2]

Stiglerin tulokset pätevät siis yleistetylle Lorenz-käyrän estimaattorille, kunhan piste p ei ole fraktiilifunktion epäjatkuvuuskohta (eli pisteessä p ei saa olla nollasta poikkeavaa todennäköisyysmassaa).

Tehdään tulojakaumasta oletus, että kyseessä on jatkuvan positiivisen satunnaismuuttujan jakauma, jonka kaksi ensimmäistä momenttia ovat äärelliset. Tällöin voidaan Stiglerin tuloksia soveltaa yleistetyn Lorenz-käyrän estimaattorin ominaisuuksien johtamiseen.

Lause 1. Olkoon ylläannetut oletukset voimassa. Tällöin yleistetyn Lorenz-käyrän estimaattorille on pätee:
i) Estimaattorin $\hat{G}_j$ otantajakauma on asymptoottisesti normaalinen.
ii) $\lim_{n\to\infty} E(\hat{G}_j) = G_j$ eli estimaattori on asymptoottisesti harhaton.
iii) Estimaattorin $\hat{G}_j$ asymptoottinen varianssi on

$$(7) \qquad \sigma_{jj} = n\mathrm{Var}(\hat{G}_j) = p_j[\lambda_j^2 + (1-p_j)(X(p_j) - \mu_j)^2],$$

[2] Rajankäyntiehto on selvä, jos $u > p_j$ tai piste u on "paljon pienempi" kuin p_j. Olkoon piste u lähellä arvoa p_j eli olkoon $p_j - u = \varepsilon$. Tällöin pisteen u avoimessa $\varepsilon/2$ ympäristössä pätee $J_n = J = 1$, kunhan $n > (2p_j - \varepsilon)/\varepsilon$. Tällöin rajankäyntiehto on voimassa, eli jokaisella J:n jatkuvuuspisteellä on avoin ympäristö, missä J_n suppenee tasaisesti kohti J:tä.

Liite 2: Eräitä lukuun neljä liittyviä todistuksia (jatkoa)

missä $X(p_j)$ on X:n fraktiilifunktion arvo pisteessä p_j, $\mu_j = E[X|F(X) \leq p_j]$ on X:n ehdollinen odotusarvo ja $\lambda_j^2 = E[X - E[X|F(X) \leq p_j]|F(X) \leq p_j]^2$ on ehdollinen X:n varianssi.
iv) Estimaattorivektorin $\hat{G}$ otantajakauma on asymptoottisesti multinormaalinen.
v) Olkoon $\hat{G}_j$ ja $\hat{G}_k$ kaksi estimaattorivektorin alkiota ja olkoon $p_k \geq p_j$. Tällöin $\hat{G}_j$:n ja $\hat{G}_k$:n välinen asymptoottinen kovarianssi on

$$\sigma_{jk} = p_j[\lambda_j^2 + (1 - p_k)(X(p_j) - \mu_j)(X(p_k) - \mu_k) + (X(p_j) - \mu_j)(\mu_k - \mu_j)]. \tag{8}$$

Todistus:

i) Seuraa suoraan aputuloksesta ja alaviitteessä käsitellystä funktion rajankäyntiehdosta.
ii) Seuraa suoraan aputuloksen kaavasta 2.
iii) Aputuloksen kaavan 3 mukaan

$$\begin{aligned}
\sigma_{jj} &= \int_0^\infty \int_0^\infty I^{p_j}(x) I^{p_j}(y)[F(\min(x,y)) - F(x)F(y)]dxdy \\
&= \int_0^{X(p_j)} \int_0^{X(p_j)} [F(\min(x,y)) - F(x)F(y)]dxdy \\
&= \int_0^{X(p_j)} \left(\int_0^y F(x)dx + \int_y^{X(p_j)} F(y)dx - \int_0^{X(p_j)} F(x)F(y)dx \right) dy \\
&= \int_0^{X(p_j)} \left(yF(y) - \int_0^{X(p_j)} xf(x)dx + \int_y^{X(p_j)} xF(y) - F(y)(p_jX(p_j) - p_j\mu_j) \right) dy \\
&= \int_0^{X(p_j)} \left(X(p_j) - p_jX(p_j) + p_j\mu_j)F(y) - \int_0^y xf(x)dx \right) dy \\
&= (X(p_j) - p_jX(p_j) + p_j\mu_j)(p_jX(p_j) - p_j\mu_j) - \int_0^{X(p_j)} \int_0^y xf(x)dxdy \\
&= p_j(1-p_j)(X(p_j) - \mu_j)^2 + p_j\mu_jX(p_j) - p_j\mu_j^2 - \int_0^{X(p_j)} \int_x^{X(p_j)} xf(x)dydx \\
&= p_j(1-p_j)(X(p_j) - \mu_j)^2 + p_j\mu_jX(p_j) - p_j\mu_j^2 - p_jX(p_j)\mu_j + p_j(\lambda_j^2 + \mu_j^2) \\
&= p_j[\lambda^2 + (1-p_j)(X(p_j) - \mu_j)^2]
\end{aligned}$$

iv) Tarkastellaan kahden eri pisteen p_j ja p_k, yleistettyjen Lorenz-ordinaattojen estimaattoreita. Tarkastellaan näiden estimaattorien mielivaltaista lineaarikombinaatiota

$$T = \theta_1\hat{G}_j + \theta_2\hat{G}_k = \frac{1}{n}\sum_{i=1}^n (\theta_1 I^{p_j} + \theta_2 I^{p_k})(i/n)x_{(i)}.$$

Jälkimmäisestä muodosta nähdään, että mielivaltaisen lineaarikombinaation estimaattori kuulu myös Stiglerin esittämien tunnuslukujen luokkaan (painofunktiona $\theta_1 I^{p_j} + \theta_2 I^{p_k}$). Koska mielivaltaisen lineaarikombinaation asymptoottinen jakauma on normaalinen, on asymptoottinen yhteisjakauma multinormaalinen (Rao 1972, 128). Tämä tarkastelu yleistyy useammankin kuin kahden pisteen välisiin tarkasteluihin.

Liite 2: Eräitä lukuun neljä liittyviä todistuksia (jatkoa)

v) Olkoon $\hat{G}_j$ ja $\hat{G}_k$ kuten edellisessä kohdassaja lisäksi olkoot $p_k \geq p_j$. Tarkastellaan näiden summaa eli lineaarikombinaatiota T, jossa $\theta_1 = \theta_2 = 1$ ja käytetään hyväksi identiteettiä $\sigma_{jk} = n\mathrm{Cov}(\hat{G}_j, \hat{G}_k) = \frac{1}{2}n(\mathrm{Var}(\hat{G}_j + \hat{G}_k) - \mathrm{Var}(\hat{G}_j) - \mathrm{Var}(\hat{G}_k))$. Kohdan iv) ja kaavan 3 mukaan

$$
\begin{aligned}
\sigma_{jk} &= \frac{1}{2}\Big(\int_0^\infty \int_0^\infty (I^{p_j}(x) + I^{p_k}(x))(I^{p_j}(y) + I^{p_k}(y))[F(\min(x,y)) - F(x)F(y)]dxdy \\
&\quad -n\mathrm{Var}(\hat{G}_j) - n\mathrm{Var}(\hat{G}_k)\Big) \\
&= \frac{1}{2}\Big(\int_0^{X(p_k)} \int_0^{X(p_k)} (I^{p_j}(x) + 1)(I^{p_j}(y) + 1)[F(\min(x,y)) - F(x)F(y)]dxdy \\
&\quad -n\mathrm{Var}(\hat{G}_j) - n\mathrm{Var}(\hat{G}_k)\Big) \\
&= \frac{1}{2}\Big(\int_0^{X(p_k)} \int_0^{X(p_k)} (I^{p_j}(x)I^{p_j}(y) + I^{p_j}(x) + I^{p_j}(y) + 1)[F(\min(x,y)) \\
&\quad -F(x)F(y)]dxdy - n\mathrm{Var}(\hat{G}_j) - n\mathrm{Var}(\hat{G}_k)\Big) \\
&\overset{(iii)}{=} \frac{1}{2}\int_0^{X(p_k)} \int_0^{X(p_k)} (I^{p_j}(x) + I^{p_j}(y))[F(\min(x,y)) - F(x)F(y)]dxdy \\
&= \int_0^{X(p_k)} \int_0^{X(p_k)} I^{p_j}(x)[F(\min(x,y)) - F(x)F(y)]dxdy \\
&= n\mathrm{Var}(\hat{G}_j) + \int_0^{X(p_j)} \int_{X(p_j)}^{X(p_k)} [F(y) - F(x)F(y)]dxdy \\
&= n\mathrm{Var}(\hat{G}_j) + \int_0^{X(p_j)} F(Y)dy \int_{X(p_j)}^{X(p_k)} [1 - F(x)]dx \\
&= n\mathrm{Var}(\hat{G}_j) + pj(X(p_j) - \mu_j)[(1-p_k)X(p_k) - (1-p_j)X(p_j) + p_k\mu_k - p_j\mu_j] \\
&= p_j[\lambda^2 + (1-p_k)(X(p_j) - \mu_j)(X(p_k) - \mu_k) + (X(p_j) - \mu_j)(\mu_k - \mu_j)]
\end{aligned}
$$

Ylläolevasta lauseesta näkee, että myös kovarianssimatriisin alkiot voidaan estimoida otoksesta ilman jakauman parametrista muotoa koskevia oletuksia. Tätä varten otoksesta täytyy estimoida ehdolliset odotusarvot, ehdolliset varianssit sekä fraktiilifunktion arvot tarkastelupisteissä. Nämä kaikki voidaan estimoida ei-parametrisesti.

Käytännön tutkimustilanteessa on vielä huomioitava se, että usein aineistoa ei ole kerätty yksinkertaisella satunnaisotannalla, vaan se sisältää otantapainot eri havaintoyksiköille, jotka pitäisi ottaa huomioon. Beach & Kalinski (1986) osoittavat kuitenkin, että jos otantapainot oletetaan ei-stokastisiksi (siis kiinteiksi ja ennaltatunnetuiksi), niin nämä tulokset laajenevat myös painotettuun tilanteeseen. Tällöin kaikki painottamattomat suureet korvataan vastaavilla painotetuilla suureilla. Tätä tulosta on käytetty hyväksi tässä tutkimuksessa vaikka oletus kiinteistä ja ennaltatunnetuista otantapainoista ei välttämättä ole kovin realistinen, sillä todellisuudessa otantapainot on estimoitu otannan suorittamisen jälkeen. Tämän tuskin kuitenkaan on merkittävä virhelähde.

Liite 2: Eräitä lukuun neljä liittyviä todistuksia (jatkoa)

Toinen monimutkaistus, jota ei sen tarkemmin tässä käsitellä, on se, että kaikkien pisteittäisten estimaattorien (yleistetyt Lorenz-ordinaatit, ehdolliset odotusarvot ja varianssit sekä fraktiilifunktion arvot) laskemisessa on käytetty lineaarista interpolaatiota kahden havainnon välissä, kun luku np_j ei ole ollut kokonaisluku. Tällä on toivottu saavutettavan parempia pienen otoksen ominaisuuksia estimaattorille. Tässä liitteessä esitetyn asymptoottisen teorian kannalta ovat tavanomainen (katkaistu) sekä lineaarista interpolaatiota käyttävä estimaattori ekvivalentteja.

Liite 3: Tarkasteltavat käyrät taulukoina eri vuosille

Taulukko 1. Vuoden 1971 yleistettyjen, tavanomaisten ja absoluuttisten Lorenz-ordinaattojen sekä ryhmäkeskiarvojen ja ryhmien osuuden kokonaistulosta estimaatit vintiiliryhmittäin vuoden 1990 rahaksi deflatoituina. Estimaatin alapuolella oleva luku on sen asymptoottinen keskihajonta.

Prosentti-piste	GL-ordinaatta	Lorenz-ordinaatta (%)	AL-ordinaatta	Ryhmä-keskiarvo	Ryhmä-osuus (%)
5.0	597.41	1.42	-1505.59	11948.29	1.42
	22.42	0.0515	26.19	448.31	0.0515
10.0	1497.36	3.56	-2708.66	17998.82	2.14
	37.23	0.0830	44.61	377.24	0.0415
15.0	2582.18	6.14	-3726.85	21696.41	2.58
	49.59	0.1085	61.27	311.83	0.0341
20.0	3792.64	9.02	-4619.39	24209.21	2.88
	61.98	0.1329	76.89	319.41	0.0339
25.0	5108.96	12.15	-5406.08	26326.44	3.13
	73.61	0.1550	91.50	305.02	0.0320
30.0	6519.91	15.50	-6098.13	28219.04	3.35
	86.91	0.1780	105.02	352.55	0.0345
35.0	8040.01	19.12	-6681.04	30401.94	3.61
	100.25	0.2001	117.80	351.28	0.0337
40.0	9660.15	22.97	-7163.91	32402.86	3.85
	111.06	0.2190	130.32	286.89	0.0299
45.0	11370.89	27.03	-7556.18	34214.71	4.07
	124.59	0.2389	141.27	363.79	0.0333
50.0	13185.77	31.35	-7844.30	36297.78	4.31
	138.90	0.2579	151.10	386.07	0.0339
55.0	15121.63	35.95	-8011.44	38717.19	4.60
	155.26	0.2768	159.42	440.47	0.0367
60.0	17177.48	40.84	-8058.61	41116.91	4.89
	174.36	0.2941	165.44	518.78	0.0412
65.0	19361.23	46.03	-7977.86	43674.96	5.19
	188.30	0.3093	172.71	386.87	0.0339
70.0	21700.16	51.59	-7741.94	46778.72	5.56
	209.17	0.3222	175.44	580.73	0.0442
75.0	24214.55	57.57	-7330.56	50287.67	5.98
	229.12	0.3333	177.46	564.15	0.0429
80.0	26936.47	64.04	-6711.64	54438.44	6.47
	254.89	0.3386	174.43	743.08	0.0561
85.0	29910.16	71.11	-5840.96	59473.80	7.07
	279.13	0.3414	170.59	723.86	0.0556
90.0	33158.44	78.84	-4695.69	64965.66	7.72
	302.78	0.3373	163.53	759.80	0.0609
95.0	36840.04	87.59	-3117.09	73631.97	8.75
	331.02	0.3154	146.80	994.30	0.0832
100.0	42060.14	100.00	0.00	104402.01	12.41
	408.20	-	-	3178.05	0.3154

Liite 3: Tarkasteltavat käyrät taulukoina eri vuosille (jatkoa)

Taulukko 2. Vuoden 1976 yleistettyjen, tavanomaisten ja absoluuttisten Lorenz-ordinaattojen sekä ryhmäkeskiarvojen ja ryhmien osuuden kokonaistulosta estimaatit vintiiliryhmittäin vuoden 1990 rahaksi deflatoituina. Estimaatin alapuolella oleva luku on sen asymptoottinen keskihajonta.

Prosentti-piste	GL-ordinaatta	Lorenz-ordinaatta (%)	AL-ordinaatta	Ryhmä-keskiarvo	Ryhmä-osuus (%)
5.0	947.25 24.46	1.95 0.0481	-1486.36 25.43	18945.03 489.18	1.95 0.0481
10.0	2221.41 35.33	4.56 0.0675	-2645.82 38.87	25483.12 285.58	2.62 0.0272
15.0	3662.61 48.73	7.53 0.0897	-3638.23 52.35	28823.97 349.25	2.96 0.0311
20.0	5241.69 60.80	10.77 0.1085	-4492.76 64.37	31581.61 313.22	3.24 0.0270
25.0	6940.06 73.17	14.26 0.1263	-5228.00 75.47	33967.52 322.63	3.49 0.0267
30.0	8751.08 84.99	17.98 0.1423	-5850.59 85.69	36220.34 308.91	3.72 0.0250
35.0	10659.80 98.27	21.90 0.1580	-6375.48 94.78	38174.35 351.37	3.92 0.0267
40.0	12682.38 113.05	26.06 0.1737	-6786.52 102.89	40451.59 389.72	4.16 0.0284
45.0	14819.24 126.93	30.45 0.1873	-7083.27 110.29	42737.16 366.82	4.39 0.0261
50.0	17070.66 141.09	35.07 0.1998	-7265.46 116.83	45028.43 376.48	4.63 0.0261
55.0	19428.10 155.65	39.92 0.2107	-7341.63 122.19	47148.79 395.08	4.84 0.0267
60.0	21916.44 171.14	45.03 0.2206	-7286.89 126.57	49766.96 422.26	5.11 0.0280
65.0	24522.88 185.33	50.38 0.2285	-7114.07 130.33	52128.73 396.17	5.36 0.0267
70.0	27261.03 201.57	56.01 0.2346	-6809.54 132.18	54762.90 463.26	5.63 0.0304
75.0	30138.70 215.02	61.92 0.2393	-6365.48 134.24	57553.44 395.44	5.91 0.0280
80.0	33173.09 231.38	68.16 0.2408	-5764.69 133.37	60687.89 498.09	6.23 0.0337
85.0	36387.66 246.62	74.76 0.2403	-4983.73 131.50	64291.42 488.67	6.60 0.0349
90.0	39821.86 263.77	81.82 0.2332	-3983.15 125.45	68683.92 587.82	7.06 0.0423
95.0	43612.53 286.31	89.60 0.2089	-2626.09 109.33	75813.35 846.12	7.79 0.0636
100.0	48672.23 340.15	100.00 -	0.00 -	101194.09 2370.86	10.40 0.2089

Liite 3: Tarkasteltavat käyrät taulukoina eri vuosille (jatkoa)

Taulukko 3. Vuoden 1981 yleistettyjen, tavanomaisten ja absoluuttisten Lorenz-ordinaattojen sekä ryhmäkeskiarvojen ja ryhmien osuuden kokonaistulosta estimaatit vintiiliryhmittäin vuoden 1990 rahaksi deflatoituina. Estimaatin alapuolella oleva luku on sen asymptoottinen keskihajonta.

Prosentti-piste	GL-ordinaatta	Lorenz-ordinaatta (%)	AL-ordinaatta	Ryhmä-keskiarvo	Ryhmä-osuus (%)
5.0	925.50	1.79	-1660.37	18510.06	1.79
	21.18	0.0391	20.58	423.67	0.0391
10.0	2299.55	4.45	-2872.19	27481.03	2.66
	32.42	0.0573	31.07	285.93	0.0246
15.0	3870.21	7.48	-3887.40	31413.11	3.04
	40.94	0.0697	39.17	226.26	0.0186
20.0	5590.53	10.81	-4752.95	34406.46	3.33
	50.07	0.0820	46.83	244.35	0.0192
25.0	7437.80	14.38	-5491.56	36945.25	3.57
	58.46	0.0922	53.60	226.78	0.0172
30.0	9408.90	18.19	-6106.33	39422.05	3.81
	67.85	0.1027	59.93	252.74	0.0183
35.0	11495.75	22.23	-6605.35	41736.94	4.04
	76.38	0.1114	65.59	230.54	0.0163
40.0	13689.77	26.47	-6997.20	43880.42	4.24
	85.26	0.1194	70.64	241.91	0.0165
45.0	15989.80	30.92	-7283.04	46000.72	4.45
	94.07	0.1265	75.08	242.52	0.0161
50.0	18400.80	35.58	-7457.91	48219.97	4.66
	103.81	0.1330	78.74	269.89	0.0173
55.0	20925.22	40.46	-7519.36	50488.46	4.88
	112.13	0.1381	82.19	232.84	0.0153
60.0	23560.92	45.56	-7469.53	52713.90	5.10
	122.17	0.1425	84.41	285.42	0.0176
65.0	26326.56	50.90	-7289.77	55312.75	5.35
	132.17	0.1458	86.02	287.47	0.0177
70.0	29223.32	56.51	-6978.87	57935.29	5.60
	142.36	0.1478	86.76	298.87	0.0184
75.0	32265.04	62.39	-6523.02	60834.47	5.88
	152.44	0.1485	86.75	303.58	0.0190
80.0	35474.26	68.59	-5899.68	64184.28	6.21
	164.62	0.1464	84.47	377.32	0.0235
85.0	38873.76	75.17	-5086.05	67990.01	6.57
	174.59	0.1437	82.42	324.97	0.0219
90.0	42512.95	82.20	-4032.73	72783.73	7.04
	186.67	0.1360	76.96	416.39	0.0281
95.0	46521.28	89.95	-2610.27	80166.63	7.75
	201.87	0.1176	64.99	574.13	0.0410
100.0	51717.42	100.00	0.00	103922.90	10.05
	233.03	-	-	1416.55	0.1176

Liite 3: Tarkasteltavat käyrät taulukoina eri vuosille (jatkoa)

Taulukko 4. Vuoden 1985 yleistettyjen, tavanomaisten ja absoluuttisten Lorenz-ordinaattojen sekä ryhmäkeskiarvojen ja ryhmien osuuden kokonaistulosta estimaatit vintiiliryhmittäin vuoden 1990 rahaksi deflatoituina. Estimaatin alapuolella oleva luku on sen asymptoottinen keskihajonta.

Prosentti-piste	GL-ordinaatta	Lorenz-ordinaatta (%)	AL-ordinaatta	Ryhmä-keskiarvo	Ryhmä-osuus (%)
5.0	1172.03	2.02	-1724.39	23440.60	2.02
	21.36	0.0353	21.16	427.11	0.0353
10.0	2787.23	4.81	-3005.61	32304.01	2.79
	30.68	0.0489	30.92	248.31	0.0194
15.0	4599.73	7.94	-4089.54	36249.92	3.13
	39.73	0.0611	39.86	244.37	0.0183
20.0	6564.73	11.33	-5020.96	39300.10	3.39
	47.99	0.0716	47.89	223.45	0.0163
25.0	8661.24	14.95	-5820.87	41930.17	3.62
	56.34	0.0815	55.36	226.30	0.0159
30.0	10878.48	18.78	-6500.06	44344.74	3.83
	65.40	0.0912	62.26	245.27	0.0164
35.0	13219.72	22.82	-7055.24	46824.84	4.04
	74.19	0.1000	68.63	237.19	0.0155
40.0	15680.91	27.07	-7490.47	49223.80	4.25
	83.32	0.1083	74.43	246.62	0.0155
45.0	18259.39	31.52	-7808.42	51569.60	4.45
	92.65	0.1161	79.64	253.61	0.0155
50.0	20953.39	36.17	-8010.83	53880.06	4.65
	101.65	0.1229	84.33	246.61	0.0149
55.0	23758.76	41.01	-8101.89	56107.38	4.84
	110.39	0.1288	88.45	243.87	0.0147
60.0	26683.31	46.06	-8073.76	58491.01	5.05
	119.98	0.1340	91.65	271.35	0.0156
65.0	29733.62	51.33	-7919.87	61006.20	5.27
	129.14	0.1383	94.34	263.66	0.0153
70.0	32928.94	56.84	-7620.98	63906.35	5.52
	140.56	0.1411	95.22	333.67	0.0183
75.0	36289.25	62.64	-7157.09	67206.24	5.80
	151.29	0.1431	95.69	319.27	0.0178
80.0	39825.36	68.75	-6517.40	70722.18	6.10
	161.89	0.1438	95.27	325.64	0.0185
85.0	43562.95	75.20	-5676.24	74751.80	6.45
	173.89	0.1414	92.35	385.73	0.0219
90.0	47582.93	82.14	-4552.67	80399.68	6.94
	187.82	0.1339	85.87	470.58	0.0275
95.0	52060.72	89.87	-2971.31	89555.74	7.73
	206.55	0.1144	71.19	679.69	0.0429
100.0	57928.45	100.00	0.00	117354.64	10.13
	242.30	-	-	1552.40	0.1144

Liite 3: Tarkasteltavat käyrät taulukoina eri vuosille (jatkoa)

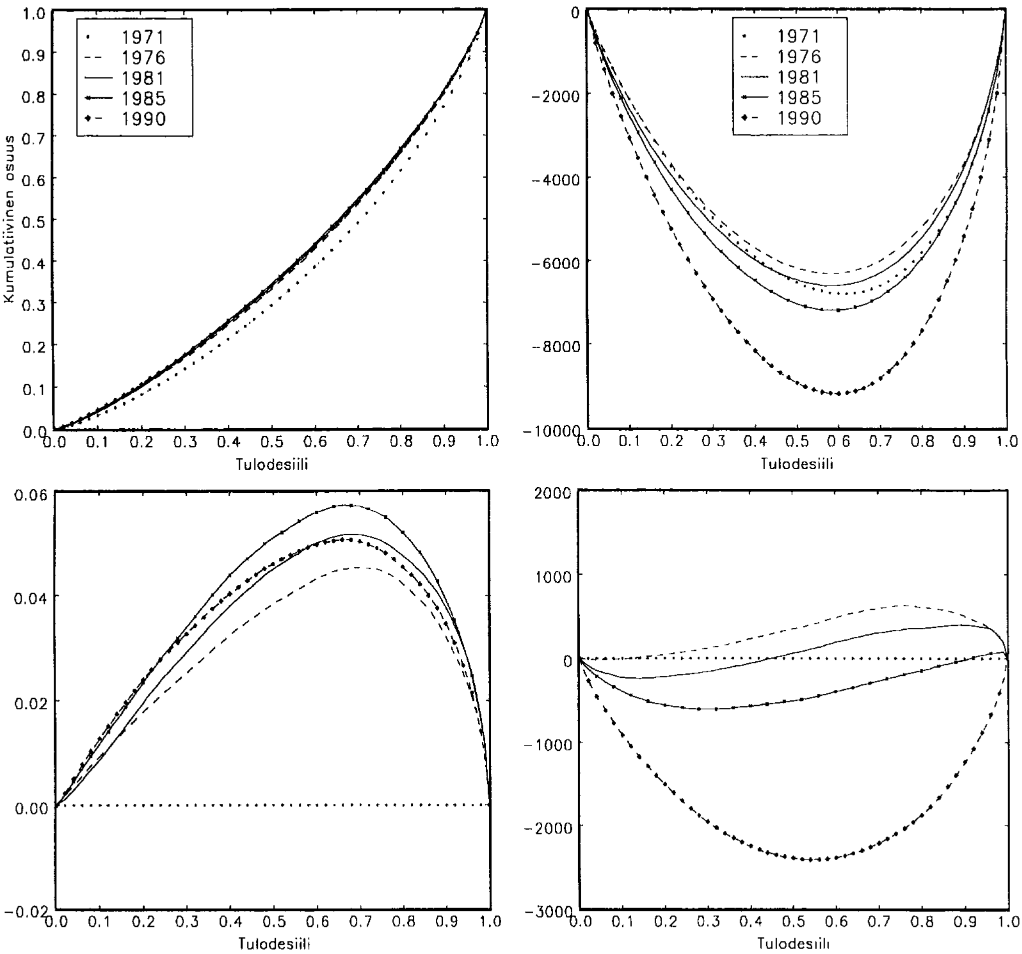
Taulukko 5. Vuoden 1990 yleistettyjen, tavanomaisten ja absoluuttisten Lorenz-ordinaattojen sekä ryhmäkeskiarvojen ja ryhmien osuuden kokonaistulosta estimaatit vintiiliryhmittäin vuoden 1990 rahaksi deflatoituina. Estimaatin alapuolella oleva luku on sen asymptoottinen keskihajonta.

Prosentti-piste	GL-ordinaatta	Lorenz-ordinaatta (%)	AL-ordinaatta	Ryhmä-keskiarvo	Ryhmä-osuus (%)
5.0	1510.12	2.13	-2033.49	30202.34	2.13
	25.95	0.0355	27.34	519.07	0.0355
10.0	3525.57	4.97	-3561.64	40309.01	2.84
	35.72	0.0485	40.87	271.02	0.0190
15.0	5744.38	8.11	-4886.44	44376.24	3.13
	44.65	0.0604	53.92	251.45	0.0177
20.0	8133.95	11.48	-6040.47	47791.41	3.37
	53.46	0.0723	66.70	243.67	0.0171
25.0	10668.88	15.05	-7049.15	50698.60	3.58
	63.34	0.0848	79.10	271.36	0.0179
30.0	13345.63	18.83	-7916.00	53535.03	3.78
	72.67	0.0969	91.04	254.05	0.0171
35.0	16160.90	22.80	-8644.33	56305.45	3.97
	82.94	0.1092	102.47	278.03	0.0178
40.0	19108.45	26.96	-9240.40	58950.86	4.16
	93.16	0.1213	113.38	277.19	0.0176
45.0	22187.54	31.31	-9704.91	61581.88	4.34
	103.53	0.1332	123.74	282.58	0.0176
50.0	25402.50	35.84	-10033.56	64299.08	4.54
	114.76	0.1449	133.35	307.64	0.0183
55.0	28768.61	40.59	-10211.05	67322.24	4.75
	128.15	0.1566	141.81	367.66	0.0202
60.0	32305.69	45.58	-10217.58	70741.65	4.99
	140.68	0.1681	150.09	344.21	0.0194
65.0	36002.23	50.80	-10064.64	73930.85	5.22
	153.18	0.1793	157.71	350.39	0.0197
70.0	39874.38	56.26	-9736.09	77443.01	5.46
	166.07	0.1900	164.51	367.53	0.0204
75.0	43949.16	62.01	-9204.92	81495.62	5.75
	181.15	0.1998	169.59	439.44	0.0230
80.0	48258.77	68.09	-8438.91	86192.21	6.08
	196.03	0.2093	173.95	446.01	0.0237
85.0	52848.25	74.57	-7393.04	91789.55	6.48
	211.74	0.2180	176.88	487.46	0.0258
90.0	57790.84	81.54	-5994.06	98851.69	6.97
	229.74	0.2243	176.90	592.22	0.0310
95.0	63324.96	89.35	-4003.54	110682.53	7.81
	254.92	0.2238	169.99	887.92	0.0473
100.0	70872.11	100.00	0.00	150942.97	10.65
	341.71	-	-	3624.19	0.2238

Liite 4: Käyrien kuvaajat kahdelle muulle ekvivalenssiskaalalle

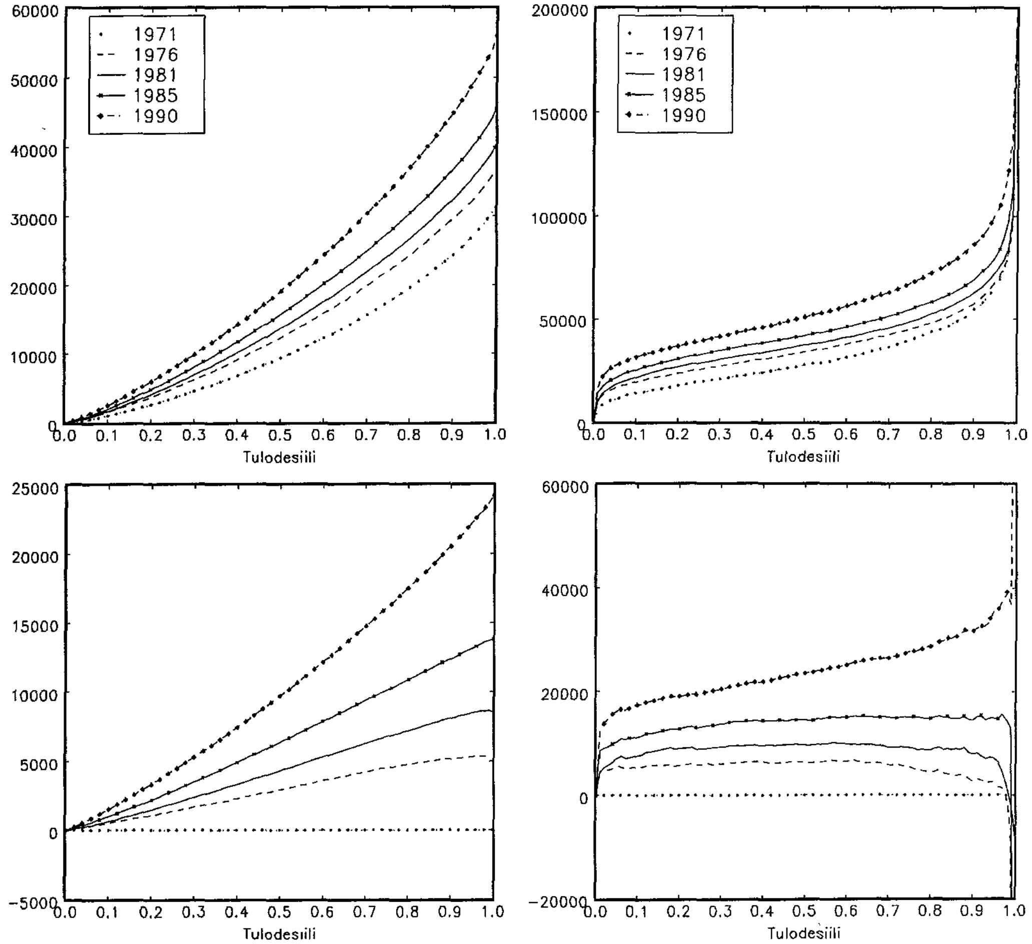
Tässä liitteessä esitetään graafisesti tässä tutkimuksessa tarkastellut käyrät sekä per henki -ekvivalenssiskaalalle laskettuna että per kotitalous -ekvivalenssiskaalalle laskettuna. Kaikki tässä liitteessä esitetyt kuviot ovat vuoden 1990 rahassa.

Liitekuvio 1. Ylhäällä: i) Lorenz-käyrät eri vuosille, ii) absoluutiset Lorenz-käyrät, alhaalla: iii) Lorenz-käyrien erotukset vuoden 1971 käyrästä ja iv) absoluuttisten Lorenz-käyrien erotukset vuoden 1971 käyrästä laskettuna per henki -ekvivalenssiskaalalla.



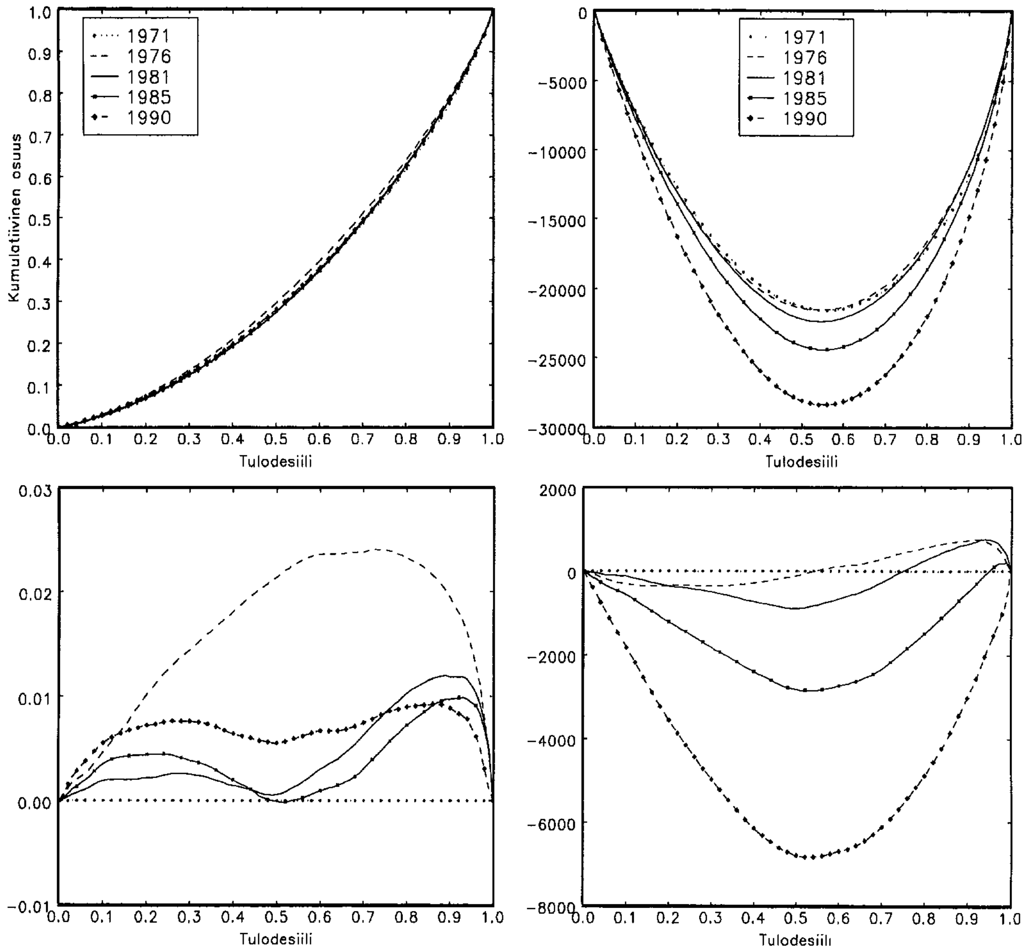
Liite 4: Käyrien kuvaajat kahdelle muulle ekvivalenssiskaalalle (jatkoa)

Liitekuvio 2. Ylhäällä: i) yleistetyt Lorenz-käyrät eri vuosille, ii) fraktiilifunktiokuvaajat, alhaalla: iii) yleistettyjen Lorenz-käyrien erotukset vuoden 1971 käyrästä ja iv) fraktiilifunktioiden erotukset vuoden 1971 käyrästä laskettuna per henki -ekvivalenssiskaalalla.



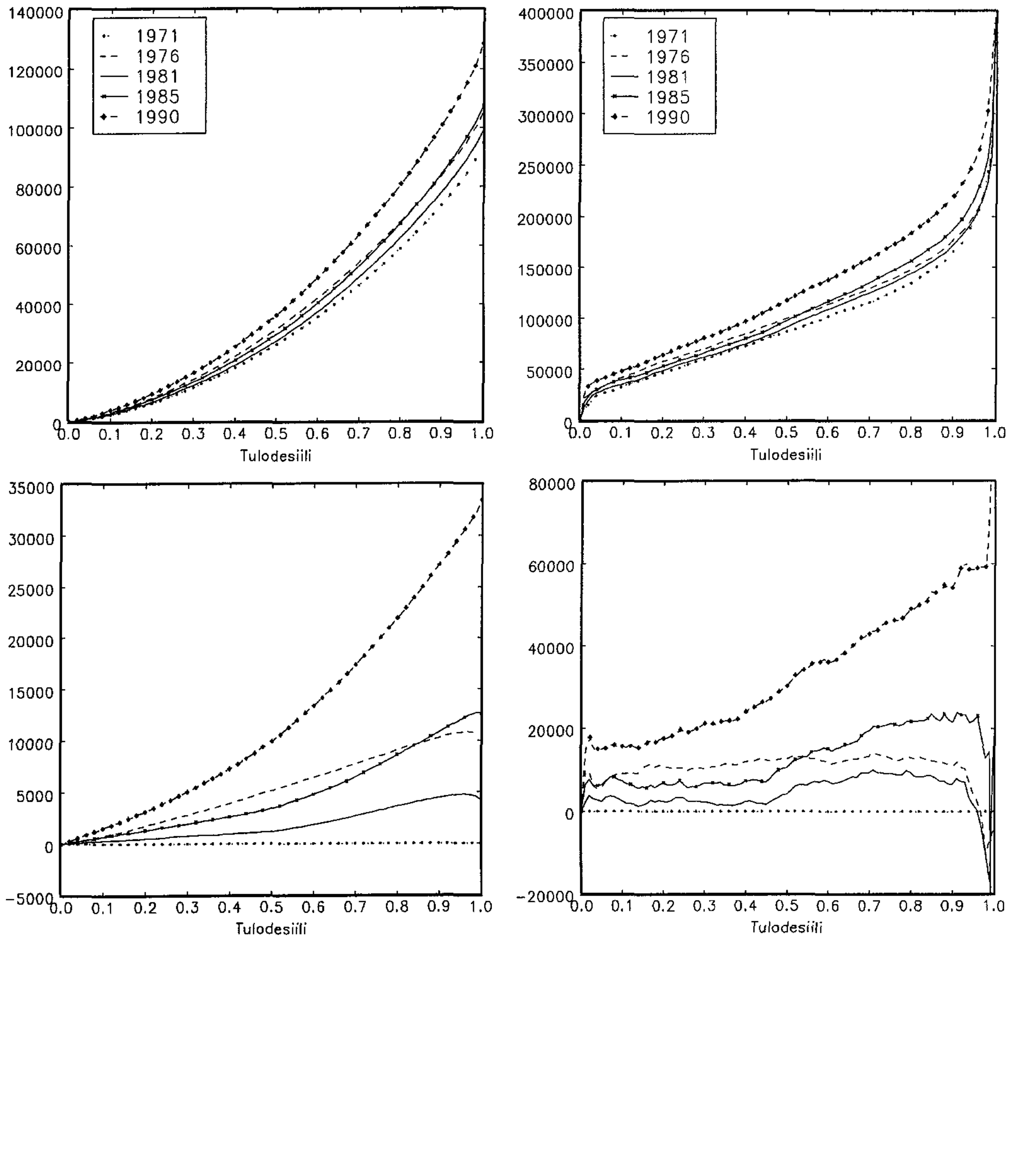
Liite 4: Käyrien kuvaajat kahdelle muulle ekvivalenssiskaalalle (jatkoa)

Liitekuvio 3. Ylhäällä: i) Lorenz-käyrät eri vuosille, ii) absoluutiset Lorenz-käyrät, alhaalla: iii) Lorenz-käyrien erotukset vuoden 1971 käyrästä ja iv) absoluuttisten Lorenz-käyrien erotukset vuoden 1971 käyrästä laskettuna per kotitalous ekvivalenssiskaalalla.



Liite 4: Käyrien kuvaajat kahdelle muulle ekvivalenssiskaalalle (jatkoa)

Liitekuvio 4. Ylhäällä: i) yleistetyt Lorenz-käyrät eri vuosille, ii) fraktiilifunktiokuvaajat, alhaalla: iii) yleistettyjen Lorenz-käyrien erotukset vuoden 1971 käyrästä ja iv) fraktiilifunktioiden erotukset vuoden 1971 käyrästä laskettuna per kotitalous ekvivalenssiskaalalla.



VATT-TUTKIMUKSIA -SARJASSA AIEMMIN ILMESTYNEET JULKAISUT

PUBLISHED VATT-RESEARCH REPORTS

1. Kuusi Osmo: Uusi biotekniikka, mahdollisuuksien ja uhkien teknologia. Helsinki: Tammi 1991.

2. Parviainen Seija: The Effects of European Integration on the Finnish Labour Market. Helsinki 1991.

3. Mustonen Esko: Julkiset palvelut: Tehokkuus ja tulonjako. Helsinki 1991.

4. Rantala Juha: Työpaikan avoinnaolon keston mittaaminen. Helsinki 1991.

5. Mäki Tuomo: Työvoiman riittävyys ja kohdentuminen 1990-luvulla. Helsinki 1991.

6. Hetemäki Martti: On Open Economy Tax Policy. Helsinki 1991.

7. Kirjavainen Tanja: Koulutuksen oppilaskohtaisten käyttömenojen eroista. Helsinki 1991.

8. Puoskari Pentti: Talouspolitiikan funktiot ja instituutiot. Helsinki 1992.

9. Parkkinen Pekka: Koulutusmenojen kehityspiirteitä vuoteen 2030. Helsinki 1992.

10. Laakso Seppo: Kotitalouksien sijoittuminen, asuinkiinteistöjen hinnat ja alueelliset julkiset investoinnit kaupunkialueella. Helsinki 1992.

11. Kirjavainen Tanja - Loikkanen Heikki A.: Ollin oppivuosi 13 000 - 56 000 markkaa. Helsinki 1992.

12. Junka Teuvo: Suurten teollisuusyritysten toimintasopeutus 1980-luvulla. Helsinki 1993.

13. Törmä Hannu - Rutherford Thomas: Integrating Finnish Agriculture into EC´s Common Agricultural Policy. Helsinki 1993.

14. Kuismanen Mika: Progressiivisen tuloverotuksen vaikutus miesten työn tarjontaan. Helsinki 1993.

15. Estonia and Finland - A Retrospective Socioeconomic Comparison. Helsinki 1993.

16. Kirjavainen Tanja - Loikkanen Heikki A.: Lukioiden tehokkuuseroista. DEA-menetelmän sovellus lukioiden tehokkuuserojen arvioimiseksi. Helsinki 1993.

17. Räsänen Mikko: Pankkien talletusvakuuden arvo ja riskikäyttäytyminen vuosina 1982 - 1992: optiohinnoittelumallin sovellus. Helsinki 1994.

18. Holm Pasi: Essays on International Trade and Tax Policy in Vertically Related Markets. Helsinki 1994.

19. Mäkelä Pekka: Markkinat ja ympäristö - Euroopan unionin ympäristöpolitiikan tarkastelua. Helsinki 1994.

20. Vartiainen Hannu: Rahoitusmarkkinat ja talouden tasapaino informaation taloustieteen näkökulmasta. Helsinki 1994.

21. Mäki Tuomo: Julkisen sektorin laajuus ja kasvu OECD-maissa. Helsinki 1995.

22. Pyy Marjo: Nuorten työllistymisen kuvaaminen elinaika-analyysin menetelmin. Helsinki 1994.

23. Lehtinen Teemu: Välittömän verotuksen tulonjakovaikutukset. Helsinki 1994.

24. Oroza Gonzalo: The CIS Mining Industry in a Transition Period - with special reference to Finnish mining prospects. Helsinki 1994.

25. Rantala Juha: Aktiivisten työvoimapoliittisten toimenpiteiden työllistävyys. Helsinki 1995.

26. Lappeteläinen Antti: General Equilibrium Models - Numerical Method and Stability. Helsinki 1995.

27. Suoniemi Ilpo - Sullström Risto: The Structure of Household Consumption in Finland, 1966-1990. Helsinki 1995.

28. Viitamäki Heikki: Vähimmäis- ja ansioturva vuonna 1995. Helsinki 1995.

29. Verouudistukset - yrittäjien sosiaalietuudet ja niiden maksupohja. Työryhmäraportti. Helsinki 1995.

30. Piekkola Hannu: Capital Income Taxation and Tax Criteria in International Capital Markets. Helsinki 1995.

31. Myhrman Rolf - Heikkilä Tuomo: Maatalouden sopeutumistarve EU-jäsenyyteen. Helsinki 1996.

32. Heikkilä Tuomo - Lang Markku - Myhrman Rolf: Maatalouden ensimmäiset vuodet Euroopan unionin jäsenenä. Helsinki 1996.

33. Ollikainen Markku: Essays on Timber Supply and Forest Taxation. Helsinki 1996.

34. Somervuori Elina: Aktiivinen työvoimapolitiikka ja työttömyys OECD-maissa. Helsinki 1996.

VATT

Valtion taloudellinen tutkimuskeskus

Hämeentie 3
PL 269
00531 HELSINKI

Seppo Leppänen,
Ylijohtaja, Valtion taloudellinen tutkimuskeskus