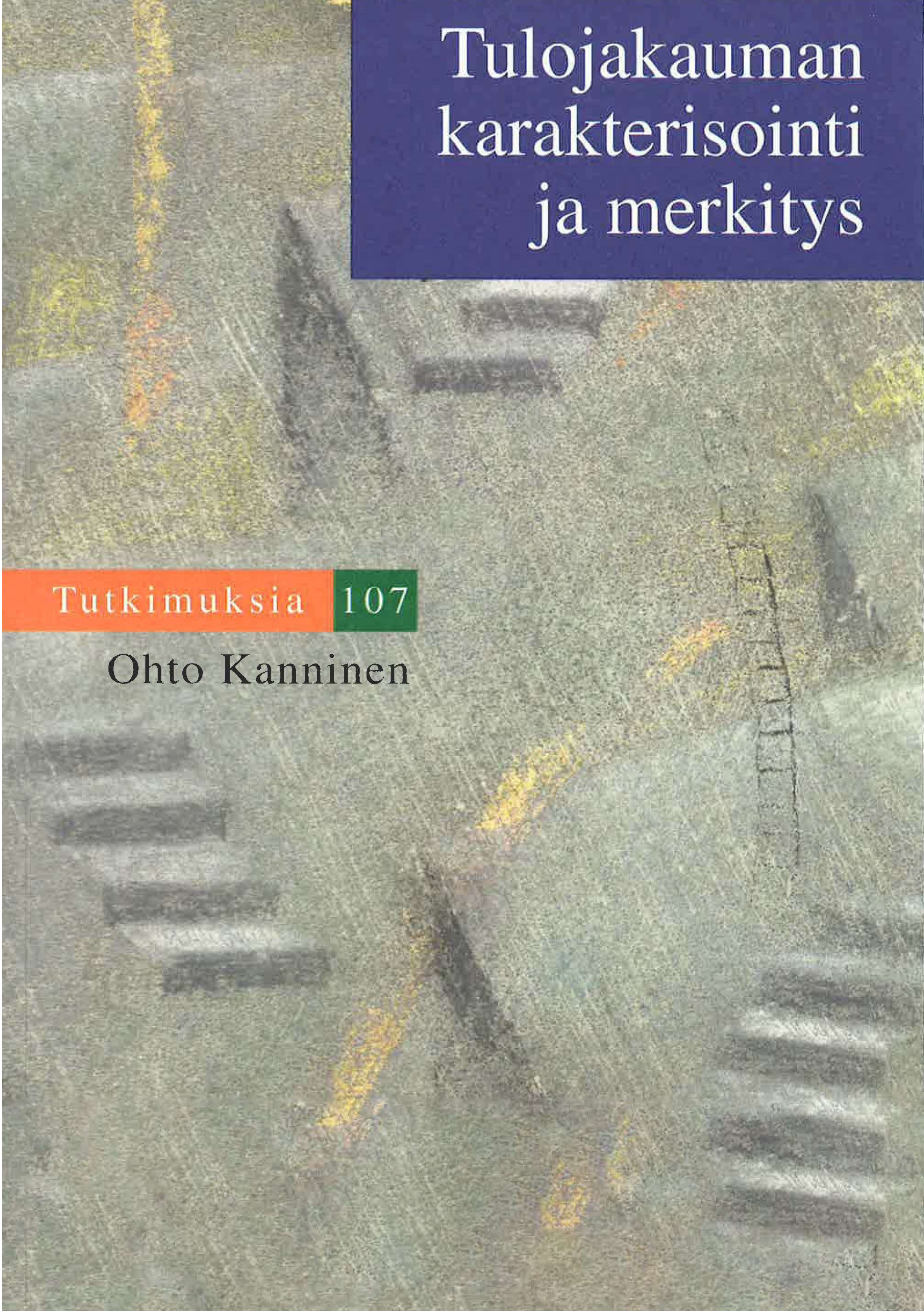
Tulojakauman karakterisointi ja merkitys



Tutkimuksia 107

Ohto Kanninen

Tulojakauman karakterisointi ja merkitys

Ohto Kanninen

Palkansaajien tutkimuslaitos
Tutkimuksia 107
Helsinki 2009

ISBN 978-952-209-068-3
ISSN 1236-7176

Sisältö

Esipuhe

Tämä tutkimus on kirjoitettu pro gradu -tutkielmaksi Helsingin yliopiston kansantaloustieteen laitokselle Palkansaajien tutkimuslaitoksen ja Ilpo Suoniemen avustuksella nimellä "Tuloerojen mittaaminen ja tulojakauman karakterisointi". Työ hyväksyttiin vuoden 2007 syksyllä arvosanalla eximia cum laude approbatur ja sai laitoksen vuoden 2007 parhaan pro gradu -työn palkinnon. Ohjaajana toimi professori Yrjö Vartia. Kiitän Yrjö Vartian lisäksi Ilpo Suoniemeä, Pasi Niskasta, Tatu Westlingiä, Jussi Lindgreniä, Hanna Hakalaa ja monia muita kollegoja ja ystäviä keskusteluista, jotka ovat auttaneet saattamaan tämän työn lopulliseen muotoonsa. Kiitokset myös Reija Liljalle, joka mahdollisti tämän julkaisun syntymisen.

Ohto Kanninen

Helsinki, kesäkuu 2009

Tiivistelmä

Tässä tutkimuksessa pyritään lisäämään ymmärrystämme mahdollisista tavoistamme karakterisoida tulojakaumaa ja hahmottaa valitun kuvailutavan merkitystä tehtyihin johtopäätöksiin. Lisäksi tavoitteena on hahmottaa tulojakauman itsensä merkitystä muuhun talouteen.

Metodologia on yhdistelmä teoriaa ja empiriaa. Keskeisenä metodologisena tavoitteena on pyrkiä jatkuvasti havainnollistamaan, mitä teoreettisia oletuksia tulosten taustalla on ja selkeästi kategorisoimaan ja esittämään mahdollisten vaihtoehtoisten lähestymisten vaikutus tutkimustuloksiin.

Tutkimuksessa saavutettiin kaksi päätulosta. Ensimmäinen tulos on ennen kaikkea deskriptiivinen ja empiirinen. Se nostaa esille selkeän kehityskulun Suomen tulojakaumassa 90-luvulla ja 2000-luvun alussa. Tämä jo aiemmin havaittu tuloeroja millä tahansa mittarilla kasvattanut kehitys on myös muuttanut koko tulojakauman muotoa kvalitatiivisella tavalla. Tähän tulokseen yritetään vastata tuomalla keskusteluun uusi monimutkaisempi teoreettinen jakaumamalli, stabiilit jakaumat, jotka mahdollistavat raskaan 'superparetohännän' mallintamisen ainakin likimain.

Toinen merkittävä tulos nousee esiin tarkasteltaessa kulutuksen, tulonsiirtojen ja tulojen välistä yhteyttä. Melko alustavalla lähestymisellä jo havaitaan monimutkaisia epälineaarisia yhteyksiä. Tämä tarkastelu osoittaa, että tulojakauman tyhjentävä kuvaileminen vaatii kehittyneitä työkaluja, jotta tulojakaumatutkimusta voidaan parhaalla tavalla käyttää muiden taloudellisten ilmiöiden ymmärryksen tukena.

English summary

In this study, we aim to increase our knowledge about the possible ways to characterize the income distribution and to outline the significance of the chosen characterization on the conclusions we make. Another, a more ambitious, aim of the study is to outline the significance of the income distribution itself on other economic outcomes.

The methodology used in this study is partly empirical and partly theoretical. A further methodological goal is to be able to illustrate to as great a length as feasible the significance of the assumptions that we make and the possible other conclusions we could make from the data given a different set of assumptions.

We reach two main results in this study. The first one of these results is mainly descriptive and empirical. We find that the development that in the late nineties and early 2000's increased dispersion in the income distribution, also changed some features of the distribution enough to induce qualitative changes, namely in the form of the heavy upper tail that we denominate 'super Pareto tail'. We propose the use of the stable distribution family to characterize the income distribution and its qualitative properties at least approximatively.

Another result arises in the part where we examine the relationship between consumption, income transfers and income. By using a rather preliminary approach, we conclude that this relationship is non-linear in a complicated way. This indicates that it requires rigorous tools to depict the income distribution in a sufficient manner to support our understanding of other economic phenomena as well.

1 Johdanto

Tuloeroja ja tulojakaumaa on tutkittu taloustieteessä Cournot'sta ja Paretosta lähtien. Atkinsonin (1970) ja Senin (1976) tutkimukset aloittivat 70-luvulla uuden kiinnostuksen tulojakaumaan. 80-luvun puoliväliin tultaessa nykyisin käytössä oleva filosofinen ja ekonometrinen metodologia oli jo suurelta osin kehitetty. 80-luvun puolivälistä lähtien tutkijat ovat keskittyneet teoreettisen kehikon soveltamisen tuottamien ongelmien ratkomiseen. (Kanbur 2003.)

Tulojakaumaan ja tuloeroihin liittyvät kysymykset ovat usein kuitenkin jääneet taloustieteen erityisalojen kysymyksiksi. Niitä ei ole onnistuneesti integroitu yleisempään taloustieteen teoriakehikkoon. Myös 70–80-lukujen teoreettisten edistysten empiiriset ongelmat on suurelta osin ratkottu. Ilmeisesti tuloerokysymyksissä pitäisi seuraavaksi kyetä syventämään tulojakauman teoreettista ymmärrystä ja yhdistämään jakaumakysymykset yleisemmin talousteoriaan.

Jälkimmäinen tavoite on erittäin kunnianhimoinen johtuen talousteorian reduktiivisesta luonteesta: teoriat on usein kehitetty edustavan yksilön abstraktion pohjalle. Paljon taloustieteen piirissä pohdittu kysymys kuuluukin, voidaanko tällaista edustavan agentin yksinkertaistusta tehdä jos ja kun jakaumat vaihtelevat aikakaudesta ja maasta toiseen. Ja jos voidaankin tehdä, minkälaisten reunaehtojen vallitessa se on järkevää.

Edward Leamer (1983) jakaa väittämät ekonometriassa olennaisesti kolmeen kategoriaan: faktoihin, mielipiteisiin ja konventioihin. Hän asettaa pohjalle konventiot. Niitä tulisi välttää. Faktoja, tässä tapauksessa vaikkapa Gini-kerroin tietylle otokselle, tulisi käyttää niin usein kuin mahdollista. Jokaisessa ekonometrisessä tutkimuksessa joudutaan kuitenkin turvautumaan useaan kertaan mielipiteisiin. Mielipiteiden käyttämisen tuottaman vahingon minimoimiseksi on Leamerin mukaan raportoitava muidenkin mahdollisten mielipidejoukkojen implikoimat päätelmät. Tämä älyllisesti rehellinen tavoite on tosin useasti vaikea saavuttaa.

Tuloerojen mittaamiseen liittyen faktaa ovat eri tuloeromittareiden vahvuudet ja heikkoudet ja mielipidettä on valita jokin tuloeromittari painottaen vahvuuksia ja heikkouksia. Valittu muuttuja tai yksikkö, jonka tasa-arvoa mitataan, on myös mielipide. Pyrin tässä tutkielmassa kyseenalaistamaan konventiot, raportoimaan mielipiteet ja tutkimaan faktoja.

1.1 Deskriptiivisyys, preskriptiivisyys ja prediktiivisyys

Nygård – Sandström (1981, 3–6) ja Amartya Sen (1980) kuvailevat kolme tapaa käsittää tuloerot.

Ensimmäinen tapa on deskriptiivisessä mielessä. Tämä viittaa tulojakauman dispersioon ja dispersioiden eroihin. Tähän osastoon kuuluu perinteinen tuloerojen mittaaminen ja vertailu.

Toinen aspekti tulojen tasa-arvoon ja epätasa-arvoon on normatiivinen eli preskriptiivinen. Preskriptiivinen lähestyminen tarkastelee tuloeroja jonkin tai joidenkin oikeudenmukaisuuskäsitteiden kautta ja kysyy, onko tuloeroja niin paljon tai vähän, että se vaikuttaa tämän oikeudenmukaisuusperiaatteen toteutumiseen.

Kolmas mainittu näkökulma antaa epätasa-arvolle prediktiivisen roolin. Siinä tulojakaumaa käsitellään asioita selittävästä, instrumentaalisesta näkökulmasta. Esimerkiksi kokonaiskulutuksen voitaisiin ajatella olevan funktio koko tulojakauman muodosta, vaikka makrotaloudelliset mallit perinteisesti lähtevätkin pelkästään oletuksesta, että tulojakauman keskiarvo tai jokin muu skalaarinen tunnusluku on ainoa tarvittava tieto.

Näiden kolmen kategorian välinen raja ei aina ole ilmeisen selvä. Silti jako näiden kolmen ryhmän välillä on suuressa osassa tapauksista tarpeeksi selvä ollakseen järkevä tapa kategorisoida tuloerotutkimus. Myös tässä tutkielmassa tätä jakoa käytetään sen suhteellisen selkeyden vuoksi.

1.2 Tutkielman rakenne ja tavoitteet

Tässä tutkielmassa jatketaan keskustelua tulojakauman luonteesta ja merkityksestä. Tulojakaumien ja tuloerojen teoria on melko hajanaista. Tulojakaumien tutkimuksen ja talousteorian yhteys on myös jäänyt melko heikoksi. Jotta tutkielman otsikon asettamaan haasteeseen voidaan kunnolla vastata, aihetta on pakko käsitellä laaja-alaisesti. Tässä työssä pyritään mahdollisimman laaja-alaiseen lähestymiseen menemättä kuitenkaan epäolennaisuuksien puolelle. Tutkielma sisältää osia, joissa käydään läpi vakiintunutta käsitteistöä. Osittain se on luonteeltaan teoriaa empiirisesti soveltavaa, osittain teoriaa itseään pyritään kyseenalaistamaan ja kehittämään. Työn toisessa luvussa esitellään aineisto. Luvussa käydään läpi mahdollisia vaihtoehtoisia metodeja muokata aineistoa kyseessä olevaan käyttöön sopivaksi. Samalla myös raportoidaan valitut metodit ja valintojen perustelut. Koska aineiston muokkaamisella voidaan vaikuttaa saatuihin lopputuloksiin, on tärkeää, että Leamerin vaatima muiden vaihtoehtoisten mielipiteiden ja niiden mahdollisten vaikutusten raportointi on mahdollisimman laaja-alaista. Erityisesti siihen pyritään toisessa luvussa tilan sallimissa rajoissa.

Kolmannessa luvussa pureudutaan tulonsaajien muodostaman tulovektorin luonteeseen ja tulkintoihin. Kolmannen luvun jälkimmäisessä puoliskossa todistetaan, kuinka osa perinteisistä teoreettisista jakaumasovitteista saattaa olla sellaisia, joista voisi olla hyvä luopua vanhentuneina konventioina Leamerin suosituksen mukaisesti. Tähän päätelmään antavat aiheen ainakin käytössä olevaan aineistoon tehdyt sovitukset. Ennen kaikkea tulee haastetuksi perinteinen näkemys tavallisesta tulojakauman Pareto-hännästä. Tämä käsite korvataan tähän yhteyteen sopivammalla niin sanotulla superparetohännällä. Sovitetut stabiilit jakaumat tuntuvat mallintavan käytettyjen kriteerien mukaan hyvin tällaista superparetohäntää.

Neljäs ja viides luku sisältävät keskustelua vakiintuneista tuloeromittareista ja niiden deskriptiivisestä ja preskriptiivisestä perustelusta. Näitä mittareita ja niiden käyttökelpoisuutta pohditaan suhteessa aiemmin esitettyyn tietoon tulojakauman luonteesta. Näissä kahdessa luvussa nostetaan esille mahdollisia ongelmia, joita kohdataan tulojakaumaa kuvattaessa yhdellä skalaarisella tunnusluvulla.

Kuudes luku päättää tutkielman rungon. Siinä pureudutaan tämän tutkielman aiheista monimutkaisimpaan eli siihen, mikä on tulojakauman prediktiivinen merkitys muulle taloudelle. Kuudennessa luvussa esitellään myös tässä yhteydessä välttämättömiä taloustieteellisiä käsitteitä ja hieman myös niiden mahdollisia heikkouksia. Luvussa yritetään myös pohtia mahdollisia tapoja yhdistää näitä kahta erillistä taloustieteellisen tutkimuksen haaraa. Luvussa esitellään simuloimalla tulonsiirtoihin ja kulutukseen liittyvä esimerkki, jossa tulojakauman muoto vaikuttaa merkittävällä tavalla muuhun talouteen. Tämä vaikutus johtuu niistä monimutkaisista epälineaarisuuksista, joita taloudessa on ja joista eräitä nostetaan tässäkin työssä esiin.

2 Aineisto

2.1 Aineiston kuvailu

Empiirisissä esimerkeissä ja illustraatioissa käyttämäni aineisto on Tilastokeskuksen Suomen tulonjakotilasto[1] vuosilta 1990–2004. Tulonjakotilasto on otostutkimus. Vuosittainen otoskoko on noin 10 000 kotitaloutta eli noin 30 000 ihmistä. Aineisto on myös kahden vuoden paneelitutkimus, eli samaa kotitaloutta on seurattu aina kahden vuoden ajan. Aineistossa on tarkkaan eritelty kotitalouksia esimerkiksi tuloluokkien, elinvaiheiden, asuinalueiden, koon ja desiiliryhmien mukaan. Tämän lisäksi yksilötasolla on vielä erikseen annettu mm. yksilön ikä, koulutusaste, ammatti ja sukupuoli.

Tiedot on kerätty haastattelemalla sekä hallinnollisista rekistereistä. Suurin osa muuttujista tulee suoraan rekistereistä. Jokaisella kotitaloudella on toteutuneen otoksen suhteen laskettu korotuskerroin, jolla on yritetty korjata kadosta johtuvaa otoksen harhaisuutta. Korotuskertoimet summautuvat Suomen väestön kokoon eivätkä siis otoksen kokoon. Korotuskerroin on kalibroitu muun muassa seuraavien muuttujien suhteen: asuntokuntien jakauma alueittain, tilastollinen kuntaryhmitys, väestön jakauma 5-vuotisikäryhmittäin ikäluokan ja sukupuolen mukaan sekä työttömyyskorvauksia saaneiden määrä.

Tulot on eritelty yksityiskohtaisesti kotitaloustasolla. Tilastossa kuvataan kotitalouksien brutto-, palkka-, yrittäjä- ja omaisuustuloja sekä kotitalouksien saamia ja maksamia tulonsiirtoja. Näistä tulokomponenteista on laskettu käytettävissä oleva tulo. Olkoot palkkatulot PT, omaisuustulot OT, kotitalouden saamat tulonsiirrot STS ja kotitalouden maksamat tulonsiirrot MTS. Käytettävissä oleva tulo KT lasketaan seuraavalla kaavalla:

$$KT = PT + OT + STS - MTS.$$

2.2 Tulotyypit ja tulonsaajat

Tuloeroihin liittyvissä kysymyksissä on olennaista keskustella siitä, kenen ja mitä tuloja käsitellään. Näiden käsitteiden kohdalla kiinnitettävä valinta on normatiivinen luonteeltaan ja voi vaikuttaa suuresti saatuihin tutkimustuloksiin. Tuloiksi voidaan valita esimerkiksi mikä tahansa yllä annetuista tulokomponenteista riippuen kysymyksenasettelusta.

[1]Ks. http://www.stat.fi/meta/til/tjt.html.

Tulonsaajan määritelmä on monimutkaisempi, sillä tulon käsitteen merkitykset eroavat suuresti. Joissakin teoreettisissa lähestymisissä tuloilla tarkoitetaan abstraktiota, joka merkitsee yksilöllistä hyvinvointia. Tulot voivat tarkoittaa myös jotakin huomattavasti käytännöllisempää käsitettä, kuten kotitalouden käytettävissä olevat tulot. Tähän määritelmään taas vaikuttaa käytännön tilastointityössä käytetyt konventiot. Tavanomainen ratkaisu näiden määritelmien lähentämiseen on korjata tilastollista aineistoa jollakin ekvivalenssiskaalalla. (Cowell 2000, 93–95.) Tällaiset skaalat määrittävät jonkun vaihtosuhteen konventionaalisesti määritellyn tulon x ja teoreettisen tai korjatun tulon y välille. Tämä korjattu tulo y toimii parempana mittana hyödylle tai muulle käsitteelle, jota on helpompi lähestyä teoreettiselta kantilta[2]. Voidaan siis ajatella, että nämä uudet tulot ovat tuloja korjattua kotitalouden kulutusyksikköä kohden.

Jos kotitalouden ominaisuuksia voidaan kuvata "täydellisesti" joillakin attribuuteilla $\mathbf{b}$, niin voidaan olettaa olevan myös jokin funktio χ siten, että

$$y = \chi(\mathbf{b}, x). \tag{1}$$

Yleensä tämä yhteys määritetään muotoon

$$y = \frac{x}{v(\mathbf{b})}, \tag{2}$$

missä $v(\mathbf{b})$ on funktio, joka määrää ekvivalenttien aikuisten määrän kyseisessä kotitaloudessa. Määritelmässä (2) on tietenkin ongelmansa, mutta joka tapauksessa (1) tulee spesifioida johonkin muotoon ja yhtälöistä jälkimmäinen on se, jota yleisesti käytetään, kuten Cowell (2000) asian ilmaisee.

Tulojen tekeminen samanarvoisiksi edellä esitetyllä tavalla sisältää luonnollisesti huomattavan määrän ongelmia. Silti se on yleensä välttämätön askel, jotta voidaan tehdä edes jollakin tavalla järkevää tulojen vertailua. (kts. Sen 2000) Talousteoria perustuu suurelta osin yksilön tai yksittäisen kuluttajan käyttäytymisen analysoimiseen, kun taas aineistot ovat yleensä kotitalouskohtaisia.

Kotitalouskohtainen analyysi on yleensä lähes välttämätön tuloeroista tai elintasosta keskusteltaessa. Jos esimerkiksi tutkittaisiin vain alle 15-vuotiaiden suomalaisten tuloja, tuloksena olisi, että lähes koko tämä ryhmä elää köyhyydessä. Huomattavasti parempi mittari tässä tapauksessa on kotitalouksien käytettävissä olevat kokonaistulot kulutusyksikköä kohden jaettuna kunkin kotitalouden ekvivalenttien aikuisten määrällä.

[2]Tässä on myös relevanttia, kuinka pitkän aikavälin tuloista on kyse. Tutkimuksen kohteena voisivat myös olla koko elinajan tulot tai vaikka kuukausitulot. Tässä tutkimuksessa mielenkiinnon kohteena ovat kuitenkin yhden vuoden tulot. Myös käytössä oleva aineisto sopii tähän tarkoitukseen parhaiten.

Kirjallisuudessa on vakiintunut kaksi määritelmää sopivalle ekvivalenssiskaalalle, jolla kulutusyksikkö voidaan määrittä. Ensimmäistä kutsutaan OECD-skaalaksi ja toista modifioiduksi OECD-skaalaksi. OECD:n määritelmillä kulutusyksiköiden määrä on funktio

$$v : \mathbb{N}^2 \to \mathbb{R}$$

kotitalouden koostumuksesta. Funktio saa argumenteiksensa attribuutit $\mathbf{b}$, jotka kuvaavat mahdollisimman täydellisesti kotitalouden tilaa. Käsiteltävillä olevissa OECD-skaaloissa nämä attribuutit ovat yksinkertaisesti aikuisten määrä n ja lasten määrä m. Olkoon S kuvaa yleisesti kotitalouden kulutusyksikköjen määrä eli kaikkien kotitalouden jäsenten suhdelukujen summa:

$$S = v(\mathbf{b}),$$

missä tässä tapauksessa $\mathbf{b} \in \{n, m\}$.

Ensimmäisessä määritelmässä ensimmäinen aikuinen saa suhdeluvun yksi ja seuraavat saavat suhdeluvun 0,7. Alle 18-vuotiaat lapset saavat suhdeluvun 0,5. Olkoon S_{OECD} koko kotitalouden suhdeluku

$$S_{OECD} = T(n \geqslant 1)[1 + 0,7(n-1)] + 0,5m,$$

missä n on aikuisten lukumäärä ja m on alle 18-vuotiaiden lasten lukumäärä. T tarkoittaa totuusfunktiota, eli funktiota, joka saa arvokseen yksi, jos sen argumenttina oleva lause on totta ja nolla, jos se ei ole totta. Tässä tapauksessa argumenttina on lause “kotitaloudessa on vähintään yksi aikuinen”.

Toisessa eli modifioidussa OECD:n määritelmässä ensimmäinen aikuinen saa suhdeluvun yksi ja sitä seuraavat saavat suhdeluvun 0,5. Alle 14-vuotiaat lapset saavat suhdeluvun 0,3. Jos S_{mod} kuvaa kotitalouden jälkimmäistä suhdelukua, n on aikuisten lukumäärä ja m on alle 14-vuotiaiden lasten lukumäärä niin

$$S_{mod} = T(n \geqslant 1)[1 + 0,5(n-1)] + 0,3m.$$

Kulutusyksiköiden määritelmä on riippuvainen siitä, mikä todetaan sopivaksi skaalaeduksi lisääntyvästä perheen koosta. Aasness – Benedictow – Hussein (2003) analysoivat veropolitiikan näkökulmasta neljää erilaista ekvivalenssiskaalamääritelmää ja niiden valinnan herkkyysanalyyttistä merkitystä saatuihin tuloksiin eri politiikkatoimenpiteiden valitsemisessa norjalaisella aineistolla. Heillä on käytössään “ykkösskaala”, molemmat OECD-skaalat ja “nollaskaala”. Ykkösskaala ja nollaskaala edustavat illustroivia ääripäitä. Tässä työssä käytetään vertailun vuoksi Aasness – Benedictow – Husseinin ykkösskaalaa.

Ykkösskaala

$$S_1 = n + m$$

tarkoittaa sitä, että mitään skaalatuottoja ei ole, eli jokainen uusi lisäys kotitalouteen aiheuttaa samat kustannukset kuin jos tämä olisi yksin kotitaloudessa. Hän saa siis suhdeluvukseen yksi.

Heidän tuloksensa ovat kohtuullisen ilmeisiä. Mitä suuremmat skaalaedut, sitä suurempi merkitys keskituloon on lapsilisillä. Toisaalta pienillä skaalaeduilla lapsilisät pienentävät eniten tuloeroja, sillä lapsiperheet ovat niille määritellyn tehottomuuden ansiosta tulojakauman alkupäässä. Suurimmaksi osaksi kuitenkin tulokset ovat aika vakaita yli kaikkien skaalausvaihtoehtojen, mutta ainakin joissakin tutkimusasetelmissa tällä normatiivisella ekvivalenssiskaalan valinnalla on kuitenkin merkitsevä vaikutus saatuihin tutkimustuloksiin.

Tarkastellaan tilannetta Suomessa vuonna 2004. Käytetään valittuja skaalausvaihtoehtoja ja tarkastellaan lyhyesti niiden merkitystä tulojen Gini-kertoimeen, keskiarvoon ja väestön köyhyystasoon. Lähdetään oletuksesta, että keskiarvo on olemassa Suomen vuoden 2004 tulojakaumalle.[3] Köyhyystasot K_{50} ja K_{60} lasketaan seuraavilla kaavoilla

$$K_{50} = \frac{1}{n} \sum_{i=1}^{n} T\left(y(a_i) < 0,5 M_{\mathbf{y}}\right)$$

ja

$$K_{60} = \frac{1}{n} \sum_{i=1}^{n} T\left(y(a_i) < 0,6 M_{\mathbf{y}}\right),$$

missä $M_{\mathbf{y}}$ tarkoittaa tulovektorin $\mathbf{y}$ mediaania ja $y(a_i)$ tarkoittaa henkilön a_i käytettävissä olevia tuloja. n on populaation koko. K_{50} tarkoittaa siis sitä osuutta väestöstä, jonka tulot jäävät alle puoleen mediaanin tuloista ja K_{60} tarkoittaa sitä osuutta väestöstä, jonka tulot jäävät alle 60 %:iin mediaanin tuloista. Gini-kerroin saadaan seuraavalla tavalla:

$$G = \frac{n}{n-1} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \mid x(a_i) - x(a_j) \mid}{2n^2 \mu},$$

missä μ on tulojen keskiarvo, $x(a_i)$ ja $x(a_j)$ on i:nen ja j:nen henkilön tulot ja $i, j \in \{1, 2, \ldots, n\}$.

Eri ekvivalenssiskaaloilla lasketut tunnusluvut on kerätty taulukkoon 2.1. Skaalausvaihtoehdoista ykkösskaala antaa OECD-skaaloista poikkeavia tuloksia. Yk-

[3] Tämä oletuksen järkevyyttä käsitellään seuraavassa luvussa tarkemmin.

Taulukko 2.1: Tunnuslukuja Suomen vuoden 2004 tuloille eri ekvivalenssiskaaloilla. Elintaso tarkoittaa tuloja kyseisellä ekvivalenssiskaalalla korjattua kotitalouden kulutusyksikköä kohden.

Ekvivalenssiskaala	Elintaso	Gini-kerroin	K_{50}	K_{60}
S_1	15264,44	0,2867984	0,066	0,126
S_{OECD}	18992,05	0,2645210	0,050	0,107
S_{Mod}	21944,50	0,2638881	0,051	0,119

kösskaalalla elintaso jää alhaisemmaksi kuin muilla vaihtoehdoilla. Tämä johtuu siitä, että perheiden elintaso jää alhaiseksi, koska mitään skaalaetuja ei ole. OECD-skaalat eivät anna huiman erilaisia tuloksia. Tosin elintasot eroavat siten, että modifioidulla OECD-skaalalla elintaso on noin 16 % korkeampi kuin OECD-skaalalla. Eritelläksemme vielä hieman lisää, minkälaisia vaikutuksia ekvivalenssiskaalan valinnalla on, katsotaan yksinelävien ja yli kahden hengen kotitalouksien K_{50} köyhyystasoja.

Taulukko 2.2: Yhden ja kahden hengen kotitalouksien K_{50} köyhyysosuudet Suomen vuoden 2004 kotitalouksien käytettävissä oleville tuloille kulutusyksikköä kohti eri ekvivalenssiskaaloilla.

Ekvivalenssiskaala	Elintaso	1 hengen kotit.	yli 2 hengen kotit.
S_1	15264.44	0.028	0.303
S_{OECD}	18992.05	0.076	0.155
S_{Mod}	21944.50	0.141	0.093

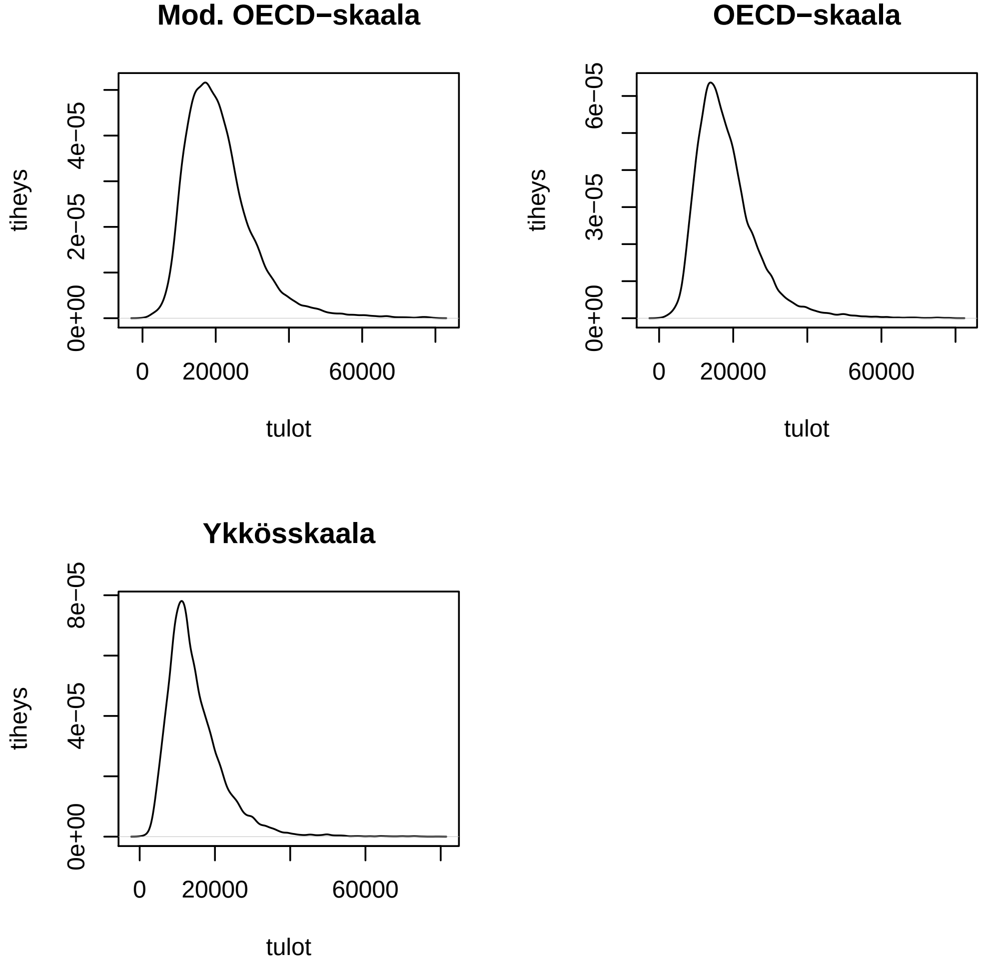
Taulukosta 2.2 nähdään, että köyhyysrajojen tapauksessa ekvivalenssiskaaloilla on huomattavia eroja. Niin sanottu ykkösskaala laittaa suuret kotitaloudet huonoon asemaan, ja näistä 30 % jää köyhiksi. Mielenkiintoisinta näissä laskelmissa on se, että nyt OECD-skaalojen köyhyyksien välille alkaa syntyä merkittäviä eroja. Nämä kaksi yleisintä kirjallisuudessa esiintyvää ekvivalenssiskaala tarjoavat siis huomattavan erilaisia tuloksia, kun köyhyyttä spesifioidaan tarkemmin ja liikutaan alueella, missä köyhyysmittari on skaalan valinnan suhteen herkimmillään. Tällä normatiivisella valinnalla voidaan siis vaikuttaa tutkimustuloksiin radikaalisti.

Kuva 2.1 sisältää kaikkien näiden neljän eri määritelmän mukaan R-ohjelmalla estimoidut tiheysfunktiot gaussisella ydinfunktiolla[4]. Tästä eteenpäin tässä työssä käytetään yleisesti tällaisissa tutkimuksissa käytettyä modifioidulla OECD-

[4]Metodi jakaa tulovektorin empiirisen kertymäfunktion 512:een yhtä suureen osaan, joita käyttäen se arvioi tiheyden eri jakauman pisteissä. Tästä lisää tietoa löytyy esimerkiksi Silvermanin (1986) teoksesta.

skaalalla korjattuja kotitalouden käytettävissä olevia inflaatiokorjattuja tuloja kulutusyksikköä kohti, koska siten tulokset ovat parhaiten verrattavissa muualla tehtyihin vastaaviin tutkimuksiin. Jokaista kotitaloutta painotetaan ekvivalenssikorjauksen jälkeen jäsentensä määrällä. Jatkossa siis empiirisessä yhteydessä tuloista ja tulovektorista puhuttaessa tarkoitetaan tällä tavoin saatua lopullista tulovektoria, jollei muuta mainita.

Kuva 2.1: Kolmella eri ekvivalenssiasteikolla skaalatut Suomen vuoden 2004 tulojen tiheysfunktioestimaatit.



3 Tulovektori ja tulojakauma

Olkoon tulonsaajia n kappaletta. Niitä kutsutaan nimellä $a_1, a_2, \ldots, a_n$. Jokaisella tulonsaajalla a_i on tulot $y(a_i)$. Näille tuloille ja niiden muodostamalle tulovektorille on olemassa kaksi erilaista tulkintaa, deterministinen ja stokastinen. (Nygård – Sandström 1981, 35–37.)

3.1 Tulovektori deterministisessä mielessä

Deterministisen tulkinnan mukaan tulonsaajilla a_i on kiinnitetyt tulot $y(a_i)$. Nämä tulot muodostavat tulovektorin $\mathbf{y} = (y(a_1), y(a_2), \ldots, y(a_n))$, joka nähdään pisteenä $\mathbb{R}^n$ -avaruudessa. Jokainen muutos minkä tahansa tulonsaajan a_i tuloissa $y(a_i)$ antaa uuden tulovektorin. Tämän vektorin perusteella voimme muodostaa empiirisen diskreetin tulojakauman laskemalla, kuinka monta havaintoa milläkin tulojen arvolla on. Näin menetämme luonnollisesti tiedon siitä, mitkä tulot yhdistyvät mihin tulonsaajaan.

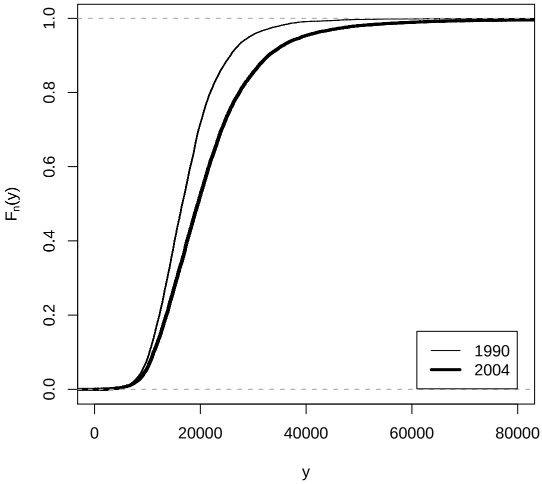
Empiirinen kertymäfunktio $F_n(y)$ määritellään

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} T[y(a_i) \leq y],$$

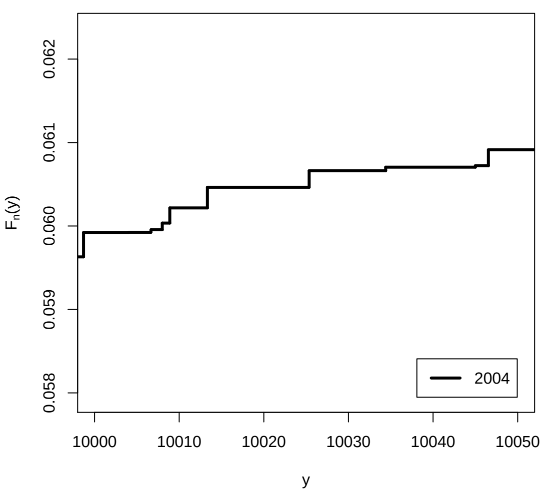
missä n tarkoittaa jälleen tulovektorin astetta, $i \in \{1, 2, \ldots, n\}$ ja T on totuusfunktio. Empiirinen kertymäfunktio on siis porrasfunktio, joka hyppää aina y:n niissä kohdissa, joissa havaintoja on ja pysyy paikallaan niissä väleissä, joissa havaintoja ei ole. Kuvat 3.1 ja 3.2 visualisoivat empiirisen kertymäfunktion. Johtuen tulovektorin suuresta asteesta n ja kuvaajien pienestä tarkkuudesta kuvaajat vaikuttavat olevan jatkuvia käyriä, mutta todellisuudessa kyseessä on porrasfunktio.

Tällä tavoin lasketun tulojakauman antama informaatio on siis verrattavissa tilanteeseen, jossa järjestämme tulonsaajat a_i tulojen $y(a_i)$. mukaan järjestykseen pienimmästä tulosta suurimpaan, siis otamme järjestystunnusluvun. Jos meitä ei kiinnosta kenelle yksittäiselle tulonsaajalle tulot kuuluvat, tällainen järjestetty tulovektori riittää antamaan kaiken relevantin tiedon. Etuna on se, että suurin osa järjestämättömän tulovektorin permutaatioista on eri pisteitä $\mathbb{R}^n$ -avaruudessa, mutta tulovektorin järjestäminen rajoittaa sallitut permutaatiot vain siihen joukkoon, joka on sama piste $\mathbb{R}^n$ -avaruudessa. Siten jos $\mathbf{y_1}$ ja $\mathbf{y_2}$ ovat järjestettyjä tulovektoreita ja $\mathbf{y_1} \neq \mathbf{y_2}$, silloin niistä muodostetut jakaumat ovat erilaisia.

Kuva 3.1: Vuosien 1990 ja 2004 tulojen empiiriset kertymäfunktiot.



Kuva 3.2: Vuoden 2004 empiirinen kertymäfunktio kun $y \in (10000, 10050)$.



Järjestystunnusluku on siis asteeltaan samaa kokoa kuin koko tulovektori. Kuitenkin, kuten edellä todettiin, se on hieman vähemmän kömpelö työkalu kuin koko tulovektori. Yksi tapa pienentää järjestystunnusluvun astetta on käsitellä asiaa tuloluokkien kannalta. Olkoot $x_1, x_2, \ldots, x_n$ eri tuloluokat ja $\mathbf{y}$ on $n \times 2$ tulomatriisi siten että $\mathbf{y} = (\mathbf{x}, \mathbf{w})$, missä $\mathbf{w}$ on painokertoimet eli ne määrät mitä kutakin tuloluokkaa esiintyy. Jos jako eri tuloluokkiin x_i, missä $i \in \{1, 2, \ldots, n\}$, on tehty tarpeeksi karkeasti, tämän tulomatriisin $\mathbf{y}$ aste on kohtuullisen pieni ja se voi toimia ihan järkevänä yksinkertaistuksena uudeksi tunnusluvuksi. Jos jako on tehty fraktiilien[5] mukaan, on ilmeistä, että painokertoimia $\mathbf{w}$ ei tarvita, sillä jokaisessa luokassa on fraktiilin mukaan määrätty osuus havainnoista.

Taulukko 3.1: Modifioidulla OECD-skaalalla korjattujen kulutusyksikköjen käytettävissä olevat vuoden 2004 mukaan inflaatiokorjatut tulot Suomessa vuonna 1990 desiiliryhmittäin.

Tulohaarukka	Frek.	Osuus	Kum. osuus	Keskitulo	Osuus tuloista	Kum. osuus tuloista
0 - 10503	497003	0,1	0,1	8654	0,05	0,05
10503 - 12410	498131	0,1	0,2	11495	0,06	0,11
12410 - 13933	497803	0,1	0,3	13186	0,07	0,19
13733 - 15309	498596	0,1	0,4	14636	0,08	0,27
15309 - 16793	497654	0,1	0,5	16046	0,09	0,36
16793 - 18334	498313	0,1	0,6	17573	0,10	0,46
18334 - 19916	498164	0,1	0,7	19139	0,11	0,57
19916 - 22166	498813	0,1	0,8	20995	0,12	0,68
22166 - 26055	497635	0,1	0,9	23921	0,13	0,82
26055 - 155745	497915	0,1	1,0	33034	0,18	1,00

Taulukoihin 3.1 ja 3.2 on kerätty Nygård – Sandströmiä (1981) mukaillen desiileittäin jaoteltujen tulovektorien joitakin tunnuslukuja. Frekvenssit eivät ole joka desiilissä täysin identtisiä. Tämä johtuu siitä, että jokainen otoksen havainto edustaa noin 200 suomalaista. Kotitalous voi siis edustaa yli tuhatta suomalaista. Näillä kaikilla on samat tulot. Siten se tulotaso, jonka alapuolelle jää esimerkiksi kymmenes kaikista tulonsaajista, on lähes poikkeuksetta sellainen, jonka alapuolelle jää myös hieman yli kymmenes kaikista tulonsaajista. Näistä tunnusluvuista saa kuitenkin jo aika hyvän käsityksen tulojakaumien eroista.

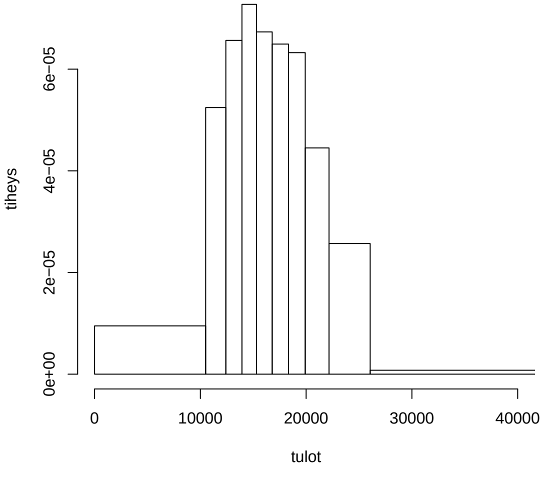
Kuvat 3.3 ja 3.4 illustroivat saman informaation graafisessa muodossa. Histogrammin pinta-ala summautuu yhteen. Jokainen laatikko edustaa yhtä desiiliä eli kymmenystä väestöstä, eli jokainen laatikko edustaa kymmenystä koko histogrammin pinta-alasta. Viimeinen eli kymmenes desiili, joka on katkaistu suurimmaksi osaksi pois molemmista kuvista, on kannaltaan vuonna 2004 niin suuri,

[5]Fraktiilit eli kvantiilit ovat suomeksi sanottuna osuuspisteitä. Prosenttipisteet eli persentiilit, desiilit, kvintiilit ja kvartiilit ovat esimerkkejä yleisesti käytetyistä kvantiileista.

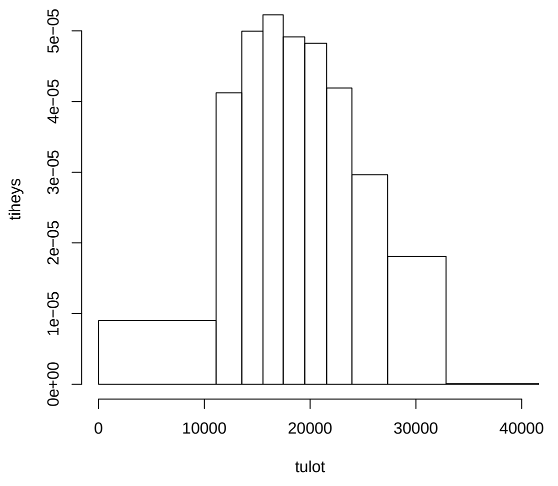
Taulukko 3.2: Modifioidulla OECD-skaalalla korjattujen kulutusyksiköiden käytettävissä olevat tulot Suomessa vuonna 2004 desiiliryhmittäin.

Tulohaarukka	Frek.	Osuus	Kum. osuus	Keskitulo	Osuus tuloista	Kum. osuus tuloista
0 - 11114	516434	0,1	0,1	8842	0,04	0,04
11114 - 13540	516409	0,1	0,2	12401	0,06	0,10
13540 - 15542	516397	0,1	0,3	14542	0,07	0,16
15542 - 17457	516908	0,1	0,4	16521	0,08	0,24
17457 - 19490	515854	0,1	0,5	18476	0,08	0,32
19490 - 21562	516167	0,1	0,6	20504	0,09	0,42
21562 - 23950	516771	0,1	0,7	22731	0,10	0,52
23950 - 27327	516553	0,1	0,8	25546	0,12	0,64
27327 - 32851	516410	0,1	0,9	29865	0,14	0,77
32851 - 1336217	515834	0,1	1,0	48891	0,23	1,00

Kuva 3.3: Histogrammi vuoden 1990 tulojakaumasta desiiliryhmittäin.



Kuva 3.4: Histogrammi vuoden 2004 tulojakaumasta desiiliryhmittäin.



että kuvan antama tarkkuus ei riitä juurikaan erottamaan sen korkeutta tuloakselista. Nämä desiileittäin lasketut histogrammit antavat hieman karkeammin saman kuvan kuin edellisen luvun kuvassa 2.1 esitellyt tiheysfunktioestimaatit.

Kuviin 3.5 ja 3.6 on laskettu absoluuttisesti ja keskiarvoon suhteutettuna tulojakaumien 1., 5., 10., 20., 50., 80., 90., 95. ja 99. prosenttipisteen[6] inflaatiokorjatut tulot aikasarjana vuosille 1990–2004. Aikasarjoista huomataan, että ennen kaikkea ylimmän prosenttipisteen yläpuolella olevat tulot ovat eriytyneet muusta populaatiosta absoluuttisesti ja suhteellisesti.

3.2 Tulovektori stokastisessa mielessä

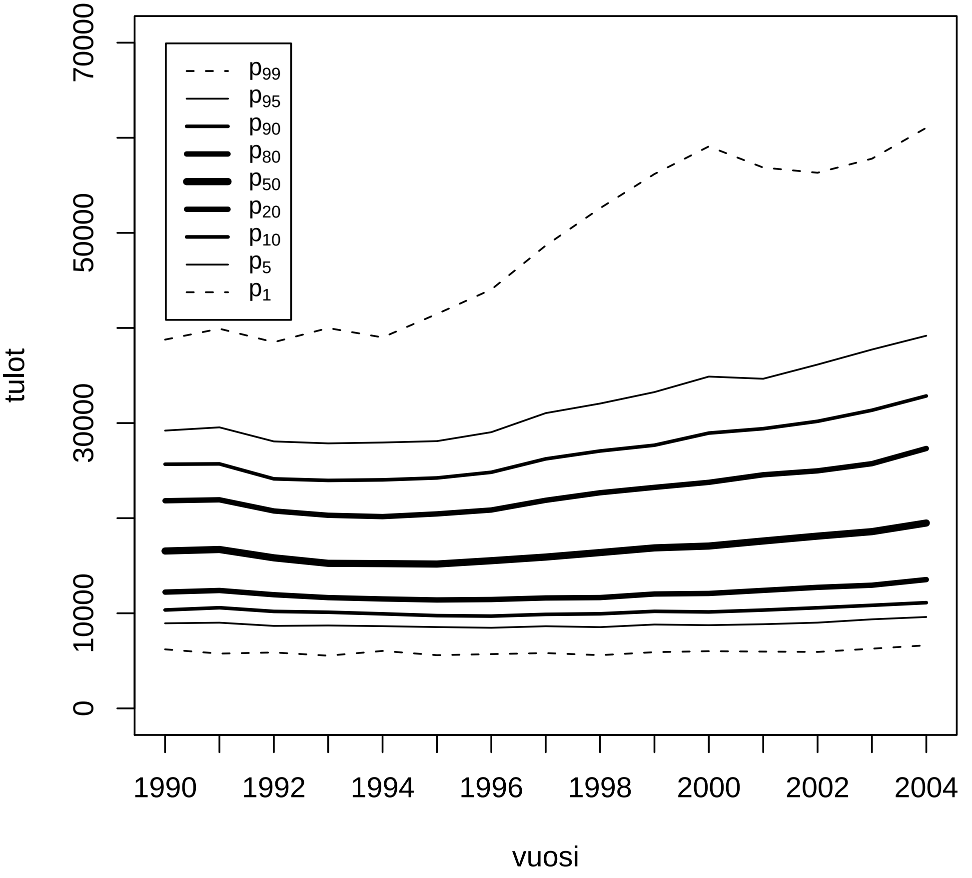
Toinen tapa tulkita tulovektori on nähdä se vektorina

$$\mathbf{Y} = (Y(a_1), Y(a_2), \ldots, Y(a_n)),$$

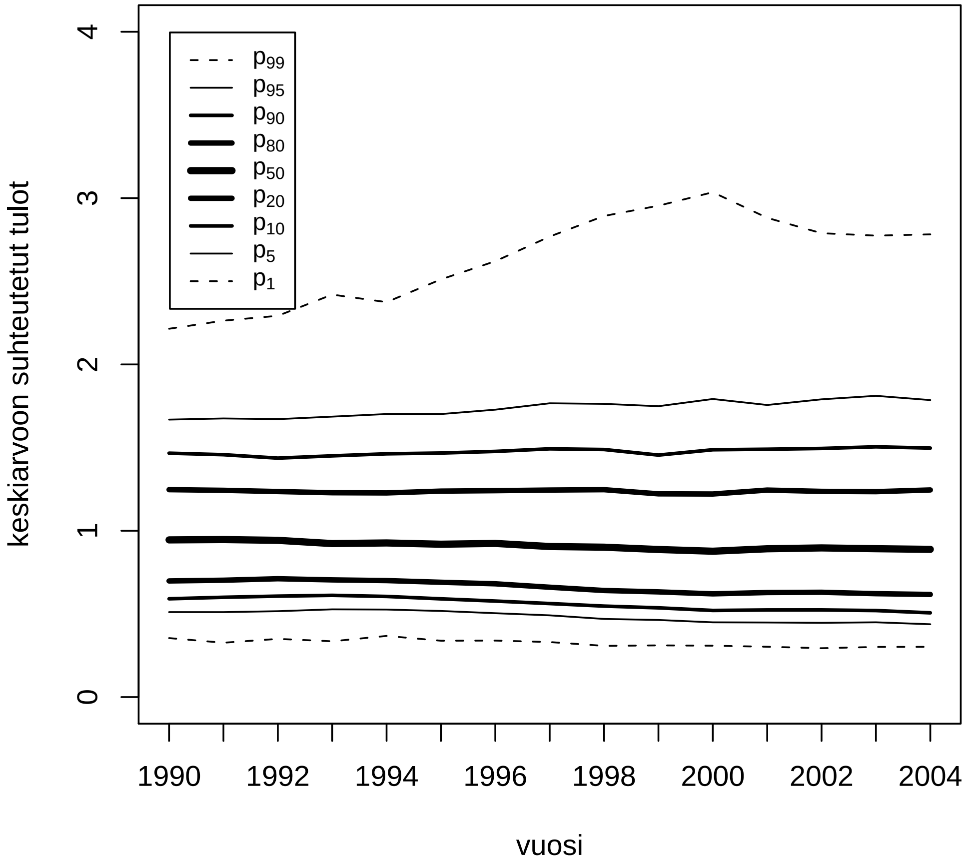
missä $Y(a_i)$ tarkoittaa satunnaismuuttujaa, joka kuvaa henkilön a_i tulonmuodostusta. Jos henkilön a_i tuloja kuvaava satunnaismuuttuja $Y(a_i)$ muuttuu, se muodostaa uuden tulovektorin. Uuden tulovektorin syntymiseen ei siis tämän stokastisen tulkinnan mukaan riitä se, että jonkin tulonsaajan tulojen realisaatio muuttuu. Taustalla siis oletetaan olevan jokin dataa generoiva prosessi, jonka

[6] Merkitsen prosenttipisteitä siten, että esimerkiksi 70. prosenttipiste (tai myös 70-prosenttipiste) on p_{70}.

Kuva 3.5: Eri tuloprosenttipisteet Suomen tulojakaumalle vuosille 1990–2004.



Kuva 3.6: Eri keskiarvoon suhteutetut tuloprosenttipisteet Suomen tulojakaumalle vuosille 1990–2004.



realisaatioista meillä on n-kokoinen otos. Tästä otoksesta voidaan muodostaa empiirinen tulojakauma samoin kuin determinististen lähestymisen tapauksessa. Tällä tavalla tulkittuna tulovektoria siis analysoidaan vain realisaationa taustalla olevista todellisista tulojakaumista.

Taustalla olevat todelliset tulojakaumat voivat siis periaatteessa olla mitä vain, ja vaikka jokaisella tulonsaajalla voi olla oma todennäköisyysjakaumansa. Näin löyhästä lähtökohdasta on kuitenkin mahdotonta tehdä mitään päätelmiä taustalla olevista jakaumista. Jakaumaa ei voi mitenkään tutkia, jos siitä on aina vain yksi realisaatio. Ongelma on analoginen stokastisten aikasarjaprosessien tilanteen kanssa. Jos ei oleteta mitään rajoitteita prosessin heterogeenisyydelle, sitä ei voida myöskään kovin järkevästi tutkia, sillä käsillä on aina vain yksi realisaatio yhdestä aikasarjasta.

Toisaalta toinen ääripää, eli tilanne jossa tulkitaan kaikkien tulonsaajien tulojen generoituvan samasta todennäköisyysjakaumasta, on myös ongelmallinen, sillä voidaan asiallisesti kysyä, onko esimerkiksi yhden vuoden tulojen todennäköisyysjakauma sama kahdella henkilöllä, joista ensimmäinen on edellisenä vuonna ansainnut kymmenkertaisesti toiseen verrattuna. Siis pitääkö paikkansa että

$$F_{Y(a_1)_t|Y(a_1)_{t-1}}(y(a_1)_t|y(a_1)_{t-1}) = F_{Y(a_2)_t|Y(a_2)_{t-1}}(y(a_2)_t|y(a_2)_{t-1}), \qquad (3)$$

kun $y(a_1)_{t-1} = 10y(a_2)_{t-1}$ ja F tarkoittaa kertymäfunktiota. Paras ennuste ilman muuta informaatiota on, että molemmat ansaitsevat edellisen vuoden tulot plus tulojen keskimääräinen kasvu eli yleensä noin kolme prosenttia, eli $E[y(a_i)_t] = 1,03y(a_i)_{t-1}$, $i \in \{1, 2\}$. a_1:n odotusarvo on siis

$$E[y(a_1)_t] = 1,03y(a_1)_{t-1} = 10,3y(a_2)_{t-1} \neq E[y(a_2)_t].$$

Jakauma ei siis ole sama ehdolla edellisen vuoden tulot, koska odotusarvokaan ei ole sama.

Tämä lähestyminen on ehkä paremmin perusteltu, jos puhutaan esimerkiksi koko elinajan tuloista. Tällöin voidaan ehkä ajatella, että ihminen on “tabula rasa”, tyhjä taulu, jonka elinajan tuloilla on sama todennäköisyysjakauma kuin kaikilla muillakin. Tämäkin lähestyminen sulkee kuitenkin silmänsä siltä tosiasialta, että myös koko elinajan tulojen todennäköisyysjakauma on vahvasti ehdollinen muun muassa perhetaustalle ja perimälle.

Yhden vuoden yhteistodennäköisyysjakaumaa analysoitaessa on siis hyväksyttävä lähtökohta, että jokaisen tulonsaajien tulot ovat ikään kuin satunnaismuuttujia, joilla on oma todennäköisyysjakaumansa. Koko otoksen jakauma on näiden yksittäisten jakaumien yhteistodennäköisyysjakauma. Tulosaajien tuloja voitaisiin esimerkiksi ajatella riippumattomana, mutta ei samoin jakautuneena otoksena (Spanos 1986, 216–218), jonka yhteistiheysfunktio on

$$f(y(a_1), y(a_2), \ldots, y(a_n); \boldsymbol{\theta}) = \prod_{i=i}^{n} f_{Y(a_i)}(y(a_i); \boldsymbol{\theta}(a_i)).$$

Jos oletetaan lisäksi funktiomuodon olevan sama jokaisella tulonsaajalla, otosmalli yksinkertaistuu muotoon

$$f(y(a_1), y(a_2) \ldots, y(a_n); \boldsymbol{\theta}) = \prod_{i=i}^{n} f_Y(y(a_i); \boldsymbol{\theta}(a_i)).$$

Ongelma tässä ajattelussa on kuitenkin se, että jokaisen tulonsaajan jokaisen vuoden tuloista on vain yksi realisaatio. Emme siis voi tietää, mitkä ovat taustalla olevat yksittäiset jakaumat. Tästä syystä käsillä on itse asiassa tilanne, jossa empiirisesti ikään kuin mallinnetaan näiden kaikkien satunnaismuuttujien summan todennäköisyysjakaumaa, josta meillä on paljon realisaatioita. Tämä on pätevä tapa saada ymmärrystä koko populaation tuloja muodostavasta stokastisesta prosessista, vaikka väestö olisi erittäin heterogeeninen.

Voidaan myös ajatella, että koko heterogeeninen populaatio koostuu osituksista, joilla tietenkin on omat satunnaisjakaumansa. Tällaisia ositteita voivat olla esimerkiksi kunnat, läänit tai vaikka sosiaaliluokat. Kertymäfunktioiden konveksi kombinaatio on myös kertymäfunktio, eli jos $\theta \in [0, 1]$ ja F_1 ja F_2 ovat kertymäfunktioita, niin

$$\theta F_1 + (1 - \theta) F_2 = F,$$

missä F on myös kertymäfunktio. Tästä syystä myös näiden ositteiden kertymäfunkiot muodostavat väestönsä koolla painotettuna koko populaation kertymäfunktion.

Aggregaattitulojakauman mahdollisesta funktiomuodosta on paljon keskustelua ja monia ehdotuksia on tarjottu (kts. Nygård – Sandström 1980, 113). Esimerkiksi Singh – Maddala (1977), Majunder – Chakravarty (1990) ja McDonald (1984) sovittavat kolmen ja neljän parametrin malleja tulojakaumaan. Näillä jakaumilla saadaan hyviä sovituksia tulojakaumaan, mutta parametrien tulkittavuus kärsii, kun niiden määrä kasvaa. Joka tapauksessa tulojakaumassa tuntuu olevan niin paljon eroavaisuuksia aikakausien ja alueiden välillä, että teoreettisen jakauman sovittaminen tulojakaumaan on vaikea tehtävä, eikä mitään lopullista yhteisymmärrystä vallitsevasta funktiomuodosta ole.

Frank Cowell (2000, 145–147) mainitsee kolme funktiomuotoa jotka ovat “eräitä merkittävimpiä”. Merkittävyys tarkoittanee tässä yhteydessä sitä, että näitä kyseisiä jakaumia on sovitettu eniten empiirisiin jakaumiin ja että niissä on hallittava määrä parametreja. Nämä kolme mallia ovat Pareto-malli, lognormaali malli ja gamma-malli.

3.2.1 Lognormaali jakauma

Lognormaalin mallin Cowell mainitsee soveltuvan parhaiten homogeenisten väestöjen tulojakauman mallintamiseen. Lognormaalin jakauman tulkittavuuden etu on se, että pienten riippumattomien satunnaisten suhteellisten muutosten tuloksena syntynyt satunnaismuuttuja on asymptoottisesti lognormaalisti jakautunut[7]. Tällaisen prosessin voidaan ajatella vastaavan ajassa tapahtuvaa tulonmuodostusta samantapaisesti kuin yhtälössä (3). Tässä tapauksessa vain muutokset ovat satunnaisia. Kokeillaan lognormaalin jakauman sopivuutta Suomen tulojakaumaan.

Lognormaalin jakauman tiheysfunktio on muotoa

$$f_Y(y;\mu,\sigma^2) = \frac{1}{y\sigma\sqrt{2\pi}} exp\left(-\frac{ln\, y-\mu^2}{2\sigma^2}\right).$$

Jos $Y_i \sim logN(\mu, \sigma_i^2)$, $i \in 1, ..., n$, ovat riippumattomia lognormaalisti jakautuneita satunnaismuuttujia, ja jos $X = \prod_{i=1}^{n} Y_i$, niin siinä tapauksessa

$$X \sim logN(n\mu, \sum_{i=1}^{n} \sigma_i^2).$$

Riippumattomien lognormaalisti jakautuneiden satunnaismuuttujien tulon jakauma on siis lognormaalinen, jos odotusarvot ovat samat. Kuitenkin tilanteessa, missä odotusarvot varioivat, tulon jakauma ei ole enää lognormaalinen, vaikka sitä voidaan approksimoida toisella lognormaalilla jakaumalla.

Lognormaalisti jakautuneiden satunnaismuuttujien summa on kuitenkin tässä keskustelussa kiinnostavampi. Summa on myös lognormaalinen, kun kyseessä ovat identtisesti jakautuneet satunnaismuuttujat. Kun parametrien annetaan varioida, summa ei pysy lognormaalina (Wu – Mehta – Zhang 2005). Lognormaalin mallin käyttäminen aiheuttaa siis tulkinnallisia ongelmia. Tämä ei kuitenkaan estä sovittamasta lognormaalia jakaumaa tulovektoriin.

[7] Lognormaaleista jakaumista ja niitä tuottavista prosesseista kirjoittavat lisää Rao (1973, 212–213), Vasama – Vartia (1972, 174–186) ja Nygård – Sandström (1981, 109–112).

Lognormaalin jakauman käyttö empiirisissä sovituksissa aiheuttaa myös ongelman nollatulojen kohdalla. Koska parametrien estimaattoreissa esiintyvää logaritmifunktiota ei ole määritelty argumentin ollessa nolla, nollatuloista on tavalla tai toisella päästävä eroon. Yksi tapa ratkaista ongelma on jättää nollatulot pois. Toinen tapa on tehdä kaikille tuloille jokin affiini muunnos $x(a_i) = c + \alpha y(a_i)$, jonka jälkeen voidaan käsitellä koko tulovektoria. Valitsen näistä jälkimmäisen tavan parametreilla $c = 1$ ja $\alpha = 1$.

Lognormaalin jakauman suurimman uskottavuuden estimaattorit parametreille μ ja σ ovat logaritmimuunnoksen jälkeen samat kuin normaalijakaumalle (Limpert – Stahel – Abbt 2001), joten estimaattorit ovat

$$\overline{y}^* = \frac{1}{n}\sum_{i=1}^{n} ln\, y_{a_i}$$

ja

$$s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(ln\, y_{a_i} - \overline{y}^*)^2.$$

Kuvassa 3.7 illustridaan lognormaalia sovitusta, kun $\overline{y}* = 9,87$ ja $s*^2 = 0,216$. Vertailuksi alle on piirretty histogrammi pienillä intervalleilla. Histogrammiesitys on sopiva, koska se on helposti tulkittavissa, toisin kuin esimerkiksi monimutkaisesti estimoitu tiheysfunktio. Lognormaali jakauma ei tuntuisi seuraavan kovin hyvin jakauman ylähäntää. Tämä havainto on samansuuntainen kuin aiemmatkin havainnot (mm. Cowell 1995).

3.2.2 Gamma-jakauma

Toinen Cowellin mainitsemista tärkeimmistä funktiomuodoista on gamma-jakauma. Gamma-jakauman tiheysfunktio on muotoa (Rao 1973, 164)

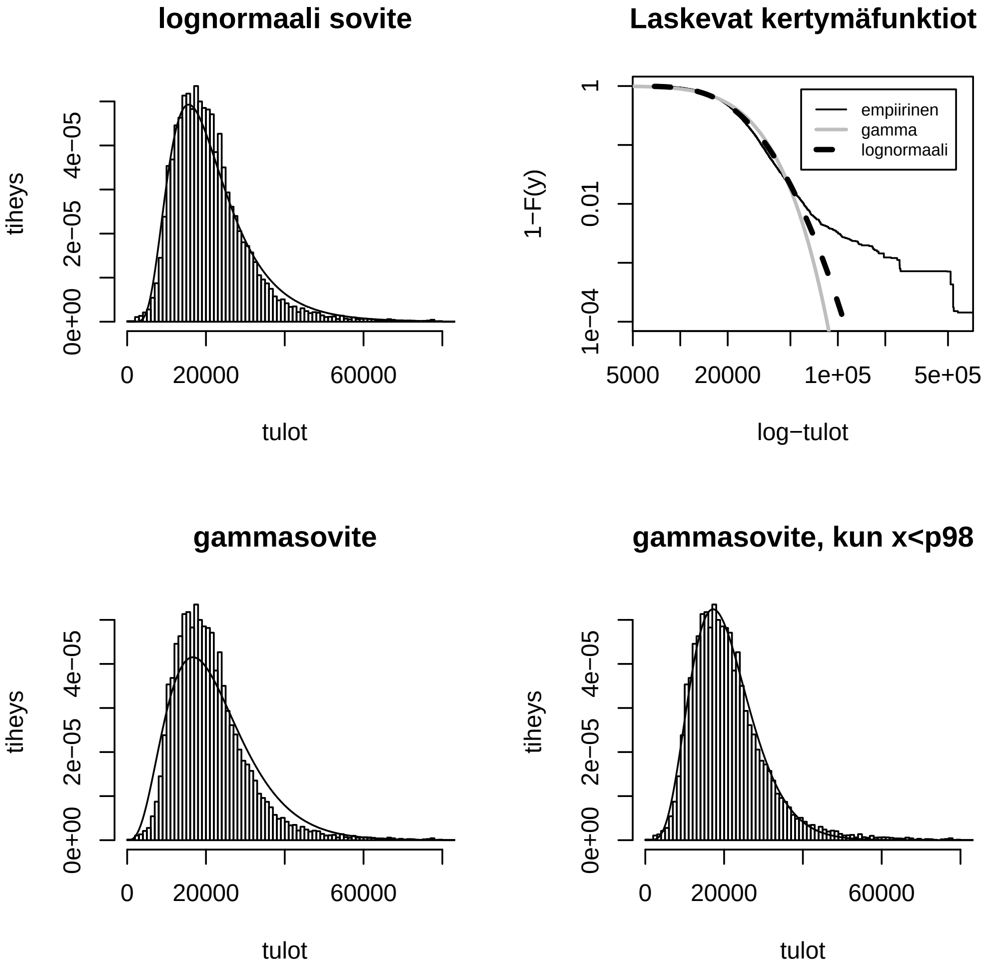
$$f_Y(y;\alpha,p) = \frac{\alpha^p}{\Gamma(p)} e^{-\alpha p} y^{p-1},$$

missä y>0 ja $\Gamma(p)$ on gamma-funktio. p on niin sanottu muotoparametri ja α on käänteinen skaalaparametri. Riippumattomien satunnaismuuttujien $x_i \sim G(\alpha, p_i)$, $i \in \{1, 2, \ldots, n\}$ summa noudattaa myös gamma-jakaumaa

$$\sum_{i=1}^{n} x_i = y \sim G(\alpha, p = \sum_{i=1}^{n} p_i),$$

mutta ei molempien parametrien variaation suhteen. Tilanne on siis sama kuin lognormaalillakin jakaumalla. Jos kaikki yksittäiset tulojakaumat ovat gamma-jakauman muotoa sillä tavalla, että niiden parametrit vaihtelevat, niiden summa ei ole gamma-jakautunut. Toisin sanoen, jos havaitaan tulojen summan olevan gamma-jakautunut, voidaan kohtuullisen suurella varmuudella todeta, että populaation ositteet tai yksittäisten tulonsaajien tulot eivät ole gamma-jakautuneita.

Kuva 3.7: Lognormaalin ja gammajakauman sovitteet vuoden 2004 tulojakaumalle.



Koska gamma-jakauma on määritelty vain positiivisille tuloille, on empiirisissä sovituksissa tehtävä samanlainen affiini muuttujamuunnos kuin lognormaalin jakauman kohdalla.

Parametrien estimointi tapahtuu suurimman uskottavuuden funktion perusteella. n:n havainnon logaritminen SU-funktio on muotoa

$$l(\alpha, p) = np \cdot log(\alpha) - n \cdot log\Gamma(p) + (p-1)\sum_{i=1}^{n} log(x_i) - \alpha \sum_{i=1}^{n}(x_i).$$

Tämän funktion maksimiarvolle ei ole analyyttistä ratkaisua. Tässä työssä tehdyt estimaatit ovat R:n tekemiä numeerisia estimaatteja kyseisestä uskottavuusfunktiosta.

Kuvasta 3.7 näkyy, että gamma-jakauma ei sovi erityisen hyvin Suomen vuoden 2004 tulojen sovittamiseen. Kun jätetään pois ne tulot, jotka ovat suuremmat kuin 98. prosenttipiste, sovitus paranee huomattavasti. Vaikka ylimpien tulojen jättäminen pois on hankalasti perusteltavissa, antaa se kuitenkin ymmärrystä siitä, että gamma-jakaumalla ei ole sellaista korkeaa ylähäntää, mikä on ominaista tulojakaumalle. Jälkimmäinen sovitus on tehty jakauman keskivaiheen sovittamisen korostamiseksi, vaikka metodologisesti tulojakauman katkaiseminen mielivaltaisesta pisteestä ei välttämättä tarjoakaan hyvää mahdollisuutta perehtyä kyseiseen tulojakaumaan.

Kuva 3.7 näyttää myös, kuinka oikea häntä ei tule mallinnetuksi oikealla tavalla gamma- tai lognormaalilla jakaumalla. Nämä jakaumat eivät näyttäisi sopivan noin viimeiseen 20 %:iin todennäköisyysmassasta. Korkeaa ylähäntää kutsutaan usein nimellä Pareto-häntä, koska se noudattaa likimain Pareto-jakaumaa.

3.2.3 Pareto-häntä

Kuten jo ylempänä on todettu, ylähäntä on se osa tulojakaumasta, johon gamma- ja lognormaali jakauma eivät sovi hyvin. Pareto-jakauma on Vilfredo Pareton kehittämä jakauma, jota hän sovelsi varallisuuden jakauman ylähännän mallintamiseen. Pareto-jakaumaa on käytetty myöhemminkin tulonjakotutkimuksessa korkean ylähännän mallintamiseen. Seuraan Pareto-jakauman esityksessäni paljolti Vasama – Vartiaa (Vasama – Vartia 1972, 491–496).

Pareto-jakauman kertymäfunktio on muotoa

$$F_X(x) = P(X < x) = \begin{cases} 1 - \left(\frac{x_0}{x}\right)^{\alpha} & , \quad kun\, x \geq x_0 \\ 0 & , \quad kun\, x < x_0. \end{cases}$$

Pareto-jakauman tiheysfunktio on muotoa (myös Rao 1973, 212)

$$f_X(x) = \begin{cases} \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} = \frac{\alpha x_0^\alpha}{x^{\alpha+1}} & , \quad kun\, x \geq x_0 \\ 0 & , \quad kun\, x < x_0. \end{cases}$$

Pareto-jakaumalla $\alpha > 0$. Jakaumalla on äärellinen odotusarvo vain, kun $\alpha > 1$ ja äärellinen varianssi vain kun $\alpha > 2$. Jakaumalla on k:s momentti vain jos $\alpha > k$. Parametri x_0 on jakauman katkaisupiste.

Kun otoksen koko on n, logaritminen suurimman uskottavuuden funktio saa muodon:

$$l(\alpha, x_0) = n[log\, \alpha + \alpha \cdot log\, x_0 - (\alpha+1)] \sum_{i=1}^{n} log\, x_i,\ i \in \{1, 2, \ldots, n\}.$$

Koska

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - log\, x_0 - \sum_{i=1}^{n} log\, x_i \ ja \ \frac{\partial^2 l}{\partial \alpha^2} = -\frac{n}{\alpha^2} < 0$$

niin α-parametrin suurimman uskottavuuden estimaattori

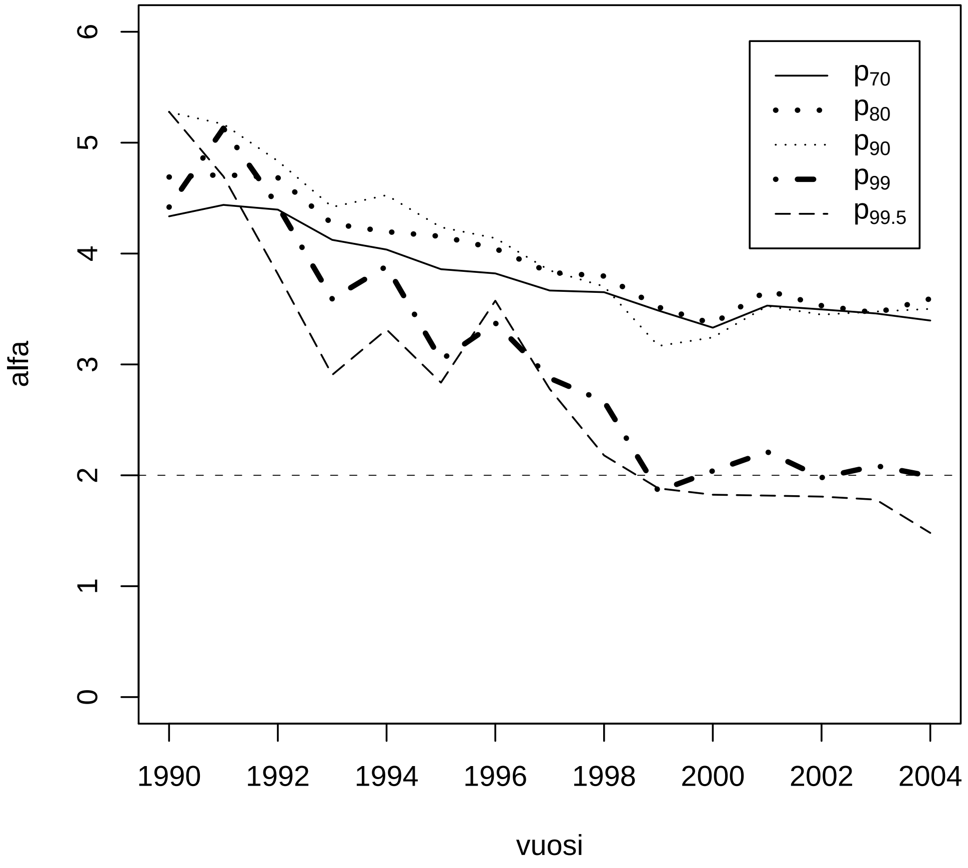
$$\alpha^* = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} log\, x_0 - log\, x_0} = \frac{1}{log\, G(x) - log\, x_0},$$

missä G(x) on geometrinen keskiarvo

$$G(x) = \left(\sum_{i=1}^{n} x_i\right)^{\frac{1}{n}} = exp\left(\frac{1}{n}\sum_{i=1}^{n} log\, x_i\right)$$

Monissa tutkimuksissa (mm. Cowell 1995 ja Vartia–Vartia 1973) on todettu, että Pareto-häntä kuvaa tulojakauman ylähäntää erittäin tarkasti noin 70.–80. prosenttipisteen jälkeen. Kuvasta 3.8 nähdään aikasarjana Pareto-sovitusten α-parametrien kehittyminen. Tämän aikasarjan perusteella näyttäisi siltä, että tulojakauman Pareto-häntäsääntö tuntuu pitävän melko hyvin vielä vuosiin 1996–1997 tultaessa. Tämän jälkeen kuitenkin tulojakauman ylimmän prosenttipisteen ja “99,5.” prosenttipisteen eli 99. ja 100. prosenttipisteen puolivälissä olevan kvantiilin kohdalta katkaistuissa sovituksissa α-parametrin arvot eriytyvät alaspäin muista sovitetuista α:n arvoista. Tämä havainto poikkeaa aiemmissa tutkimuksissa havaitusta tuloksesta, että tulojakauman ylähäntä noudattaisi yhtä jatkuvaa Pareto-häntää.

Kuva 3.8: Suomen vuosien 1990–2004 tulojakaumaan tehtyjen Pareto-hännän sovitusten arvoja, kun katkaisupiste x_0 on tulojakauman 70., 80., 90., 99. ja "99,5." prosenttipiste.



Taulukko 3.3: Neljän vuoden Pareto-sovitusten α-parametrin estimaatteja eri katkaisupisteillä x_0.

prosenttipiste	1990	1996	1999	2004
$x_0 = p_{50}$	3,623	3,356	3,008	2,929
$x_0 = p_{70}$	4,336	3,821	3,486	3,396
$x_0 = p_{80}$	4,690	4,045	3,517	3,592
$x_0 = p_{90}$	5,276	4,139	3,167	3,501
$x_0 = p_{95}$	5,150	3,953	2,753	3,156
$x_0 = p_{99}$	4,418	3,374	1,857	1,992
$x_0 = p_{99,5}$	5,279	3,574	1,882	1,480

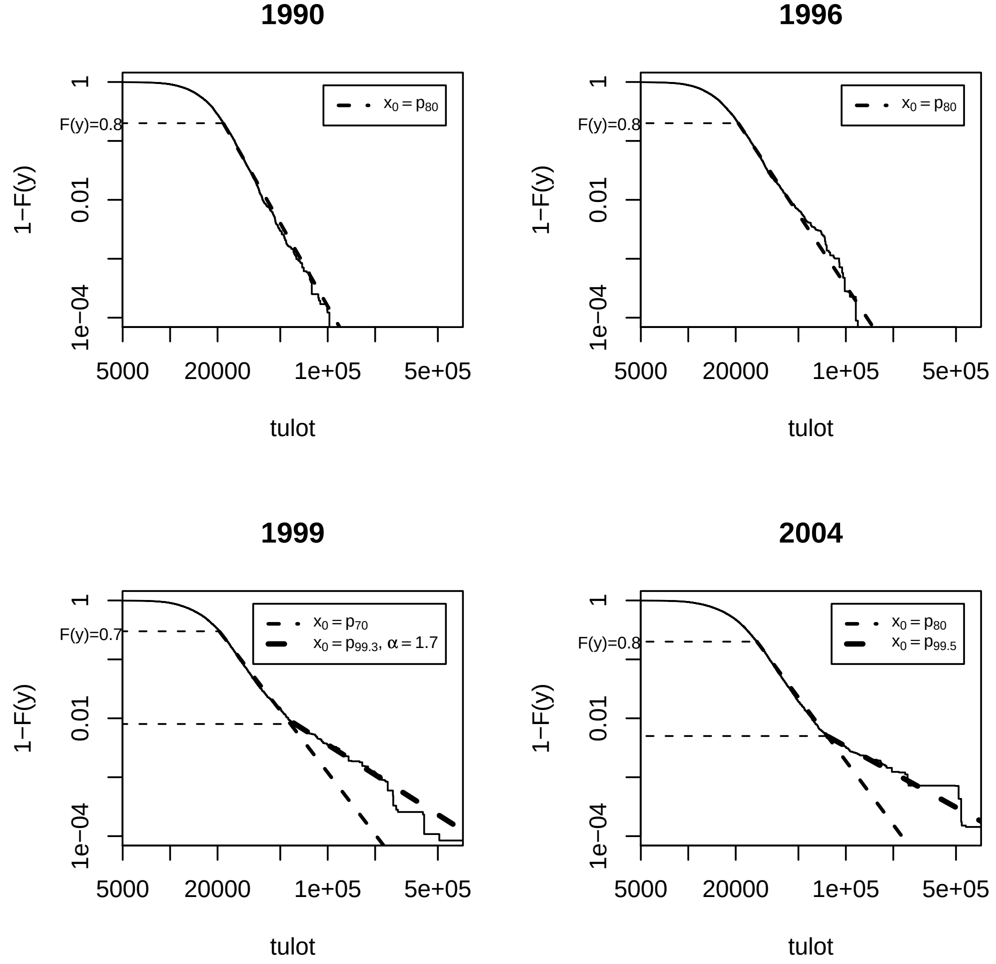
Aikajakson vuosista mielenkiintoisimpia ovat vuosien 1990, 1996, 1999 ja 2004 tulot. Niiden α-estimaatit on esitetty taulukossa 3.3. Kuvan 3.8 katkaisupisteiden lisäksi taulukkoon on lisätty 50. ja 5. prosenttipiste katkaisupisteeksi. Ensimmäinen olennainen huomio on, että 50. prosenttipiste eli mediaani katkaisupisteenä antaa systemaattisesti seuraavia katkaisupisteitä pienempiä arvoja α-parametrille. Paretohäntä ei siis ala niin aikaisin. Kun x_0 on 70. tai 80. prosenttipiste, α-parametrit ovat melko samankaltaisia kaikkina vuosina. Paretohäntä alkaa siis luultavasti jostakin näiden pisteiden välillä.

Mielenkiintoisin havainto on kuitenkin, että vuosina 1999 ja 2004 tuntuu tulojakaumalle syntyvän niin sanottu "superparetohäntä". Tämä superparetohäntä löytyy myös aineistosta kaikilta välissä olevilta vuosilta. Tänä aikana tuntuisi siis tapahtuneen kvalitatiivinen muutos tuloja tuottavassa prosessissa.

Log-log asteikolla Pareto-jakauman kertymäfunktio on suora viiva. Kuvassa 3.9 on sovitettuina log-log asteikolle laskevat kertymäfunktiot ja vastaavat Pareto-suorat. Kuvat valaisevat, kuinka vielä vuonna 1990 yli 80. prosenttipisteeseen sovitettu Pareto-häntä kuvaa erittäin hyvin tulojakaumaa. Vuonna 1996 tilanne on vielä samankaltainen, mutta vuonna 1999 ja erityisesti vuonna 2004 jakaumasta erottuu selkeästi toinen Pareto-häntä. Kutsutaan siis tällaista jakaumahäntää nimellä superparetohäntä. Vuonna 1999 superparetohännän toinen osa alkaa jostakin hieman 99. prosenttipisteen yläpuolelta tulojakaumassa. Vuonna 2004 sen alku on melko tarkalleen niin kutsutulla 99,5. prosenttipisteellä.

Toinen kvalitatiivinen ero tarkastellun ajanjakson alku- ja loppupään välillä on, että vuonna 1999 Pareto-sovitusten α-parametrit laskevat alle kahden, kun katkaisupisteenä on 99. ja 99,5. prosenttipiste. Ensimmäiset α-estimaatit jäävät vuodesta 1999 eteenpäin pyörimään kahden molemmin puolin ollen vuonna 2004 hieman alle kaksi. Estimaatit, jotka on tehty yli 99,5 prosenttipisteen meneville tuloille, eivät enää nouse yli kahden, vaan päätyvät vuonna 2004 1,48:aan. Ai-

Kuva 3.9: Vuosien 1990, 1996, 1999 ja 2004 laskevat tulojakaumat log-log -asteikolla ja sovitetut Pareto-hännät.



emmin jo todettiin, että Pareto-jakaumilla ei ole varianssia, jos $\alpha < 2$. Tästä syystä viimeaikainen muutos tuloaineistossa on ladullinen. Vaikuttaisi siltä, että on syntynyt uusi, erittäin suuria tuloja nauttivien joukko, joka rikkoo aiemmat lainalaisuudet. Muutos sattuu samalle kaudelle, jolloin Suomi liittyi Euroopan Unioniin. Samoihin aikoihin myös uusi informaatioteknologiateollisuus ja suuret optio-ohjelmat alkoivat olla merkittäviä tekijöitä Suomessa. Syitä tämän muutoksen taustalla voi olla monia, mutta muutos vaikuttaisi olevan melko pysyvä.

Testataan tämän hypoteesin kestävyyttä vielä tilastollisesti. Nollahypoteesi H_0 on siis, että $\alpha = 2$. Yksisuuntaisen testin vastahypoteesi H_1 on, että $\alpha < 2$. Tehdään aluksi Pareto-jakaumalle muuttujamuunnos $y = log\left(\frac{x}{x_0}\right)$, eli normeerataan muuttujat siten, että alkuperäisen Pareto-jakauman pienimmät x:n arvot $x = x_0$ saavat arvon 0 ja muiden arvo kasvaa logaritmisesti. Jos $F_Y(y)$ ja $F_X(x)$ ovat eksponentti -ja Pareto-jakauman kertymäfunktiot, niin

$$F_Y(y) = P(Y \le y) = P(log\left(\frac{X}{x_0}\right) \le y) = P(X \le x_0 \cdot e^y) = F_X(x_0 \cdot e^y).$$

Koska

$$F_X(x) = P(X < x) = 1 - \left(\frac{x_0}{x}\right)^\alpha,$$

niin sijoitettaessa $x = x_0 \cdot e^y$ pitää paikkansa, että

$$F_X(x_0 \cdot e^y) = 1 - \left(\frac{1}{e^y}\right)^\alpha = 1 - e^{-\alpha y} = F_Y(y),$$

joista toiseksi viimeinen osa on siis eksponenttijakauman kertymäfunktio standardimuodossaan, kun $x \ge 0$. Muulloin se saa arvon 0.

Tämän muunnoksen jälkeen voimme tarkastella parametrin α jakaumaa. Koska tiedämme eksponenttijakaumasta, että $\alpha^* \sim N(\alpha, \frac{\alpha^2}{n})$, kun H_0 on voimassa, missä α^* on α-parametrin suurimman uskottavuuden estimaatti ja siis vastaa Pareto-jakauman α-parametrin suurimman uskottavuuden estimaattia. Taulukkoon 3.4 on kerätty olennaisia tunnuslukuja, joista alin kertoo t-arvon eli kuinka monta keskihajontaa on matkaa nollahypoteesin mukaiseen tilanteeseen. Näiden lukujen mukaan vuonna 1999 yksisuuntainen nollahypoteesi kahden α:sta tulee hylättyä viiden prosentin riskitasolla. Vuonna 2004 p-arvo muodostuu niin matalaksi, että nollahypoteesi tulee hylättyä millä tahansa järjellisellä riskitasolla.

Tilastollisesti testatessa hypoteesi superparetohännän olemassaolosta ja ainakin vuonna 2004 myös hypoteesi superparetohännän varianssittomuudesta tulevat siis kirkkaasti hyväksytyiksi huolimatta siitä, että otokset ovat kohtuullisen pieniä jakauman ylähännässä.

Taulukko 3.4: Joitakin tunnuslukuja vuosille 1999 ja 2004.

	1999	2004
x_0	p99,3	p99,5
$\alpha*$	1,700	1,480
n	100	93
$\sqrt{n}$	10	9,64365
keskivirhe $\sigma_{\alpha*} = \frac{\alpha}{\sqrt{n}}$	0,170	0,153
t-arvo= $\frac{\alpha-2}{\sigma_{\alpha*}}$	$-1,765$	$-3,399$

Johtuen tästä tulojakauman ominaisuudesta, joka on ilmestynyt viime vuosina, gamma-jakauma ja lognormaali jakauma ovat huonoja sovituksia viime vuosien empiirisiin jakaumiin. Tämä johtuu siitä, että näillä kahdella jakaumatyypillä on vättämättä olemassa varianssi. Seuraavissa luvuissa sovitetaan empiiriseen tulojakaumaan sopivampia jakaumatyyppejä, joilla ei välttämättä ole varianssia.

3.2.4 Skaalattu F-jakauma

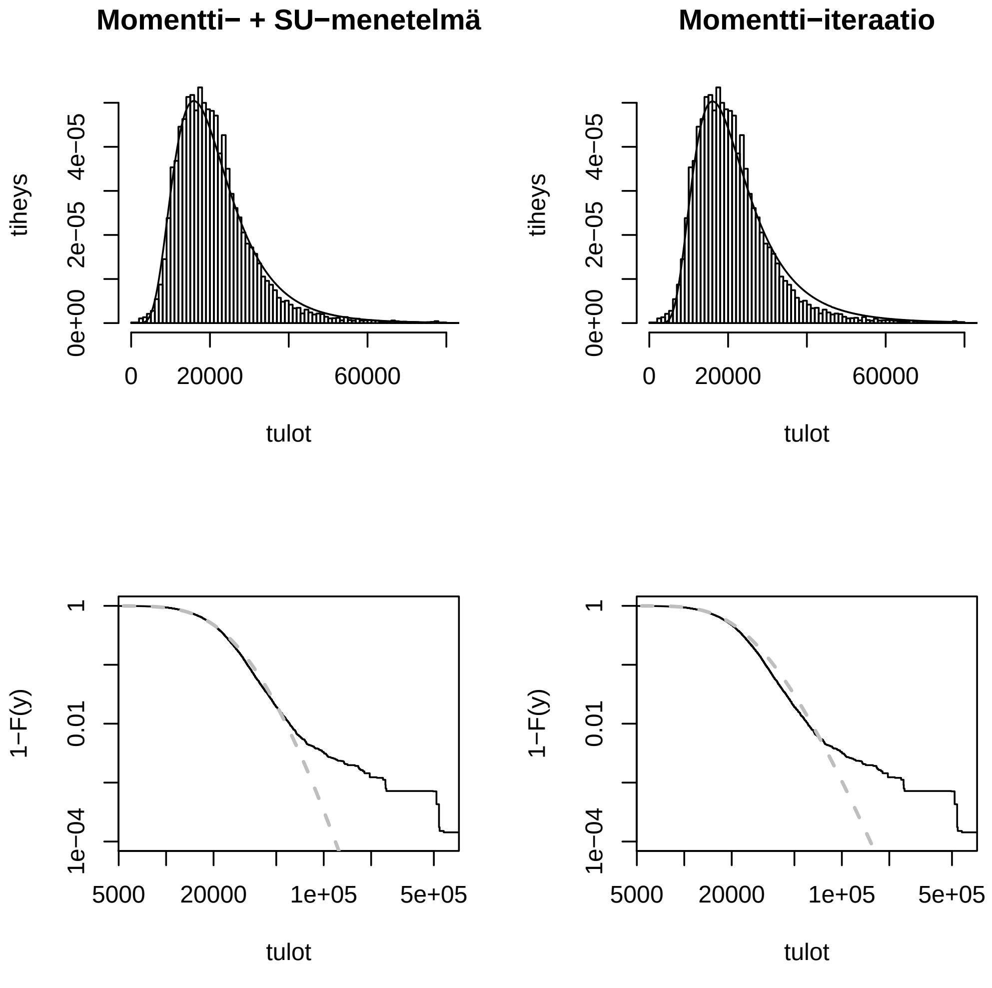
Esittelen lyhyesti skaalatun F-jakauman, jota Vartia – Vartia (1981) ovat sovittaneet tulojakaumaan. Jos F(A, m, n) on skaalattu F-jakauma, niin F(1, m, n) on F-jakauma ja jos $Y \sim F(A, m, n)$, niin $Y/A \sim F(m, n)$[8]. Funktion ylähäntä noudattaa Pareto-jakaumaa, missä Pareto-jakauman α-parametri on skaalatun F-jakauman $\frac{n}{2}$. Skaalatun F-jakauman ylähäntä voi siis saada arvoja, joilla ei ole varianssia. Skaalattu F-jakauma sisältää erikoistapauksenaan gamma-jakauman, joten sovitus on aina vähintään yhtä hyvä kuin gamma-jakaumalla.

Tässä työssä tehdään skaalatun F-jakauman sovite kahdella eri tavalla. Ensimmäisessä tavassa käytetään hyödyksi tietoa, että $E[log\, y] = logA + 2E[z] = log\, A$, jos $m = n$ ja $E[log\, y] \approx logA$ aina. Skaalaparametrin likiarvon selvittämisen jälkeen F-jakaumaa sovitetaan numeerisesti suurimman uskottavuuden metodilla R-ohjelmalla. Tämä on siis momenttimenetelmän ja suurimman uskottavuuden menetelmän sekoitus.

Toinen menetelmä on puhdas momenttimenetelmä. Vartia – Vartia (1981) johtavat iteratiivisen tavan etsiä skaalatun F-jakauman parametrit momenttimenetelmällä. He päätyvät seuraavanlaisiin iteratiivisiin yhtälöihin.

[8]Funktion tarkemmat määritelmät löytyvät Vartia –Vartiasta (1981, 200–203)

Kuva 3.10: Vuoden 2004 tulojakaumaan tehdyt F-sovitteet.



$$
\begin{aligned}
(r_1+r_2)^{(n+1)} &= 2H^{(n)}/(1+\sqrt{1+4H^{(n)}}) \\
(r_1-r_2)^{(n+1)} &= m_3/[4(r_1+r_2)^{(n+1)}+8(r_1^{(n)}r_1^{(n)}+r_1^{(n)}r_2^{(n)}+r_2^{(n)}r_2^{(n)})] \\
r_1^{(n+1)} &= \frac{1}{2}[(r_1+r_2)^{(n+1)}+(r_1-r_2)^{(n+1)}] \\
r_2^{(n+1)} &= \frac{1}{2}[(r_1+r_2)^{(n+1)}-(r_1-r_2)^{(n+1)}],
\end{aligned}
$$

missä

$$
H^{(n)} = \frac{1}{2}s^2 + r_1^{(n)}r_2^{(n)} - \frac{2}{3}[(r_1^{(n)})^3 + (r_1^{(n)})^3] \; ja \; r_1^{(0)} = r_2^{(0)} = 0,
$$

n on iteraatiokierros ja s^2 ja m_3 ovat logaritmien toinen ja kolmas keskistetty momentti. Logaritmimuunnoksilla nämä ovat jakaumalla olemassa huolimatta korkeasta superparetohännästä. A-parametri saadaan vielä jakaumasta seuraavalla tavalla:

$$
log\,A = log\,G + (r_1 - r_2) + \frac{1}{3}(r_1^2 - r_2^2).
$$

Kuvasta 3.10 nähdään, että ensimmäinen sovite on parempi jakauman keskivaiheille, kun taas jälkimmäinen metodi ottaa ylähännän paremmin huomioon. Kumpikaan ei kuitenkaan kykene seuraamaan jakauman superparetohäntää. Taulukkoon 3.5 on kerätty molempien sovitteiden parametriestimaatit. Koska n-parametrit ovat suuria, sovituksilla on varianssit, eivätkä ne siten sovi hyvin ylähännän mallintamiseen.

Taulukko 3.5: Tehtyjen skaalatun F-jakauman sovitteiden parametriestimaatit.

	Momentti + SU	Momentti-iteraatio
A	19379,75	19381,71
m	21,556	38,806
n	18,968	13,712

3.2.5 Stabiilit jakaumat

Eräs useamman parametrin jakaumaperhe, joka soveltuu tulojakauman mallintamiseen, on Levy-vinot alfa-stabiilit jakaumat (eng. Levy skew alpha-stable distribution), joita yleensä kutsutaan vain nimellä stabiilit jakaumat. Mandelbrot (mm. 1960) on tehnyt aikaisia tutkimuksia tämän perheen sovittamiseksi tulojakaumaan. Sittemmin stabiilit jakaumat ovat ilmeisesti menettäneet suosionsa tässä yhteydessä syystä tai toisesta.

Yleisesti stabiileja jakaumia ei voi kuvata suljetussa muodossa tiheysfunktiolla. Stabiilien jakaumien pahin heikkous on neljän parametrin olemassa olo, mikä heikentää jokaisen niistä tulkittavuutta. Stabiilit jakaumat tuodaan tässä työssä mukaan tulojakauman sovittamisen diskurssiin useastakin syystä, vaikka ne eivät kuulukaan Cowellin mainitsemiin merkittävimpiin jakaumiin. Stabiilien jakaumien keskeinen ominaisuus, jota käsitellään tarkemmin alempana, on niiden stabiilisuus. Lisäksi ne ovat hyvin yleinen jakaumaperhe, ja ne myös osoittautuvat kohtalaisen hyväksi sovitukseksi tulojakaumaan. Esitän seuraavaksi lyhyesti stabiilien jakaumien määritelmän Nolania (2007) ja Voitia (2000, 110–112) mukaillen.

Paras tapa kuvata koko stabiilien jakaumien joukkoa on karakteristisen funktion avulla[9]. Satunnaismuuttujalle X, jonka kertymäfunktio on $F_X(x)$, saadaan karakteristinen funktio $\phi(u)$ muunnoksella

$$\phi(u) = E\left[e^{iuX}\right] = \int_{-\infty}^{\infty} e^{iux}\, dF_X(x).$$

Karakteristinen funktio määrittää täysin satunnaismuuttujan, samoin kuin tiheysfunktio silloin, kun se on satunnaismuuttujalle olemassa. Määritellään sign-funktio seuraavalla tavalla:

$$sign\, u = \begin{cases} -1, & u < 0 \\ 0, & u = 0 \\ 1, & u > 0. \end{cases}$$

Merkitään stabiilia jakaumaa $S(\alpha, \beta, \gamma, \delta)$. Satunnaismuuttuja

$$X \sim S(\alpha, \beta, \gamma, \delta),$$

jos X:llä on karakteristinen funktio

$$\phi(u) = E\left[e^{iuX}\right] = e^{-\gamma^{\alpha}|u|^{\alpha}[1+i\beta(tan\,\frac{\pi\alpha}{2})(sign\,u)(|\gamma u|^{1-\alpha}-1)]+i\delta u},$$

missä $\alpha \in (0,1)\bigcup(1,2]$, $\beta \in [-1,1]$, $\gamma \in \mathbb{R}_{++}$, $\delta \in \mathbb{R}$. Erityistapauksessa $\alpha = 1$ karakteristinen funktio saa hieman eri muodon, mutta sitä ei tässä työssä esitellä, koska tulojakauman α on selvästi suurempi.

[9] Stabiilien jakaumien karakteristiselle funktiolle on useita eri parametrisaatioita. Tässä työssä käytetään vain yhtä selkeyden nimissä. Nolan (2007) esittelee niistä useita.

α ja β ovat niin sanottuja muotoparametreja. α:aa kutsutaan stabiilisuusindeksiksi. Sen ollessa 2 jakauma on normaalinen ja vinousparametri β:lla ei ole vaikutusta jakaumaan. Ainoastaan tällöin jakaumalle on myös määritelty varianssi. Kun $\alpha < 2$, varianssia ei ole määritelty ja vinousparametrilla on sitä suurempi vaikutus jakauman muotoon mitä pienempi α on. Kun $\beta > 0$, jakauman oikea häntä on korkeampi kuin vasen. Kun $\beta < 0$, vasen häntä on oikeaa korkeampi.

Stabiilien jakaumien tapauksessa voi kuitenkin olla hyödyllistä käyttää absoluuttisia fraktionaalimomentteja $E[\mid X\mid^p]$. Ne määritellään siten että

$$E[\mid X \mid^p] = \int_{-\infty}^{\infty} \mid x \mid^p f(x)\, dx,$$

missä p on mikä tahansa reaaliluku. Kaikille $\alpha \in (0,2)$ $E[\mid X\mid^p]$ on äärellinen kun $p < \alpha$ ja ääretön muissa tapauksissa. Jos siis $\alpha \leq 1$, niin edes odotusarvoa $E[X]$ ei ole määritetty.

Edellä esitetyssä karakteristisen funktion muodossa γ on skaalaparametri ja δ on siirtoparametri.

Stabiileilla jakaumilla on joitakin erityistapauksia, joille on löydetty tiheysfunktioesitys. Normaalijakauma $N(\mu, \sigma^2)$ on stabiili parametreilla ($\alpha = 2, \beta = 0, \gamma = \frac{\sigma}{\sqrt{2}}, \delta = \mu$). $Cauchy(\gamma_C, \delta_C)$-jakauma on stabiili parametreilla ($\alpha = 1, \beta = 0, \gamma = \gamma_C, \delta = \delta_C$) ja Lévy($\gamma_L, \delta_L$)-jakauma on stabiili parametreilla ($\alpha = \frac{1}{2}, \beta = 1, \gamma = \gamma_L, \delta = \delta_L$).

Stabiilien jakaumien eräs mielenkiintoinen ominaisuus on niiden stabiilisuus. Jos $X \sim S(\alpha, \beta, \gamma, \delta)$, niin kaikille $a \neq 0$, $b \in \mathbb{R}$,

$$aX + b \sim S(\alpha, sign\, a \cdot \beta, \mid a \mid \gamma, a\delta + b) \tag{4}$$

Lisäksi jos $X_i \sim S(\alpha, \beta_i, \gamma_i, \delta_i)$, $i \in \{1, 2, \ldots, n\}$ ovat riippumattomia vakiot $w_i \in \mathbb{R}_{++}$ ovat mielivaltaisia painokertoimia, niin

$$\sum_{i=1}^{n} w_i X_i \sim S(\alpha, \beta, \gamma, \delta),$$

missä

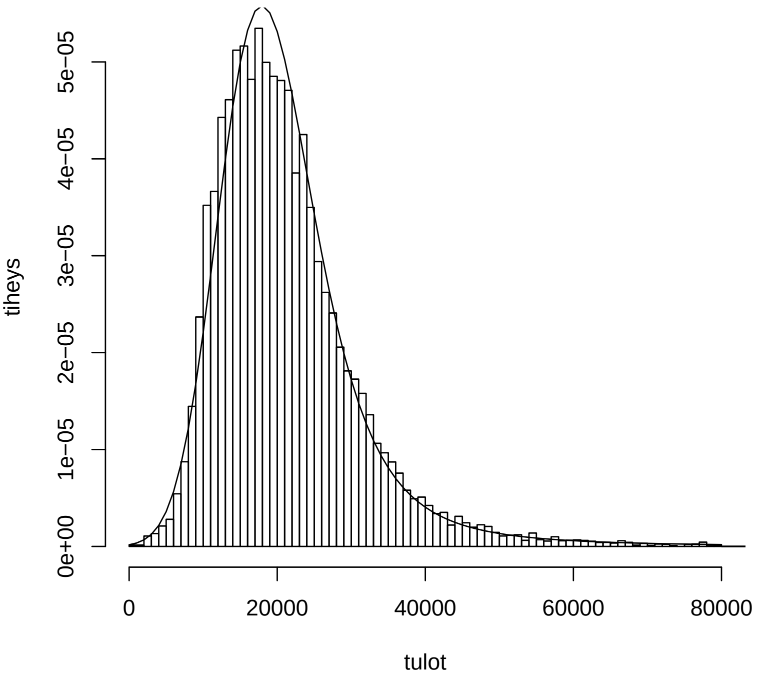
$$\beta = \sum_{i=1}^{n} \frac{(sign\, w_i)\beta_i \mid w_i\gamma_i \mid^\alpha}{\gamma^\alpha}, \; \gamma^\alpha = \sum_{i=1}^{n} \mid w_i \gamma_1 \mid^\alpha$$

ja

$$\delta = \sum_{i=1}^{n} w_i\delta_i + (tan\, \frac{\pi\alpha}{2})[\beta\gamma - \sum_{i=1}^{n} w_i\beta_i\gamma_i],\; kun\, \alpha \neq 1.$$

α-parametrin pysyessä samana stabiilia jakaumaa seuraavien satunnaismuuttujien affiini muunnos ja lineaarikombinaatio pysyvät siis samassa α:n määrittämässä stabiilien jakaumien perheessä. Tunnetusti normaalijakaumalla on samat ominaisuudet. Normaalijakauma on siis stabiili jakauma siten että $\alpha = 2$.

Kuva 3.11: Histogrammi Suomen vuoden 2004 tuloista ja sovitettu stabiili jakauma parametreilla $(\alpha; \beta; \gamma; \delta) = (1,6; 1; 5016,05; 18525,2)$.



Jos α:n sallitaan vaihtelevan tilanne on hieman monimutkaisempi. Keskeinen raja-arvolause (CLT, eng. Central Limit Theorem, kts. esim. Kendall – Stuart 1977, 206–208) toteaa tietyillä ehdoilla riippumattomien satunnaismuuttujien odotusarvon ja hajonnan suhteen normeeratun summan jakauman lähenevän rajatta normaalijakaumaa, erään ehdon ollessa, että satunnaismuuttujilla on varianssi. Normaalijakauma on siis suljettu riippumattomien normaalisten muuttujien lineaaristen lausekkeiden suhteen. Jos vähintäänkin yhdellä summattavista satunnaismuuttujista ei ole varianssia, siirrytään yleistetyn keskeisen raja-arvolauseen (GCLT, eng. Generalized Central Limit Theorem, kts. Gnedenko – Kolmogorov 1954) piiriin. Sen mukaan odotusarvon ja dispersion suhteen normeeratun satunnaismuuttujien summan jakauma lähenee stabiilien jakaumien perhettä, jos se ylipäätänsä lähenee jotakin. GCLT ei siis tee mitään vaatimuksia satunnaismuuttujien variansseista. Jos varianssien vaaditaan olevan olemassa, siirrytään siis takaisin CLT:n piiriin. Kuten yllä todetaan, Suomen viime vuosien tulojakaumalla ei ole varianssia. Tämä johtuu yksinomaan siitä pienestä osasta tuloja, jotka ovat superparetohännällä, jonka α-parametri on alle 2.

Jos mallinnetaan tulojakaumaa stabiileilla jakaumilla, voidaan ajatella että kaikki populaation ositteet tai yksittäisten tulonsaajien tulot seuraavat stabiilia jakaumaa jollakin kiinnitetyllä α:lla. Kaikki muut parametrit voivat vaihdella vapaasti. Tällaista tulovektoria voidaan mallintaa stabiililla jakaumalla, jolla on sama kiinnitetty α muiden parametrien ollessa funktioita summattavien satunnaismuuttujien α, β, γ ja δ -parametreista siten, kuin yllä on esitetty.

Toinen tapa ajatella tulojakaumaa on, että mahdollisesti jokaisella ositteesta saadulla osajoukolla on oma α-parametrinsakin. Tässä tapauksessa siis näiden ositteiden konveksi kombinaatio ei noudata stabiilia jakaumaa. Riippuen kuinka paljon α:t vaihtelevat ja kuinka suuria ositteet ovat, tämä uusi ositteiden konveksina kombinaationa saatu satunnaismuuttuja noudattaa karkeasti stabiileja jakaumia. Koska kyseessä on joka tapauksessa approksimaatio, kummatkin ajattelutavat tarjoavat riittävän uskottavat perusteet sovittaa stabiilia jakaumaa tulojakaumaan. Tässä työssä keskitytään kuitenkin koko populaation tulojen poikkileikkaukseen, kuten tähänkin asti. Ositteiden tarkasteluun olisi suotavaa myös, että tutkittavan aineiston otos olisi suurempi.

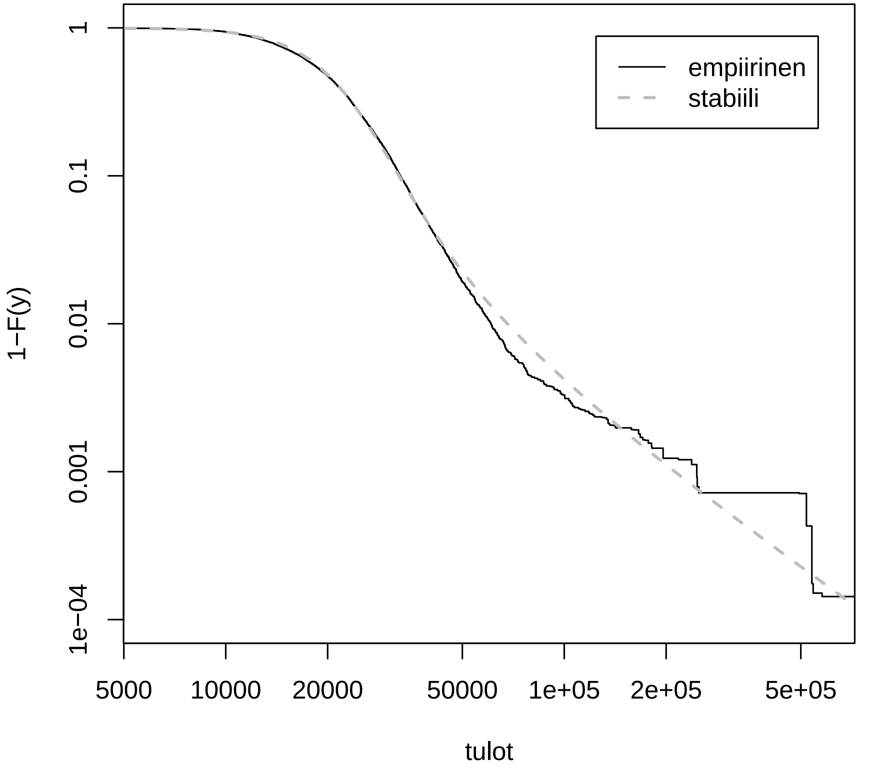
Sovituksia tehtäessä aineistoa on muokattu samanlaisella affiinilla muutoksella kuin lognormaalin ja gamma-jakaumankin kohdalla. Sovitus on ohjelmalla "stable.exe"[10] tehty suurimman uskottavuuden sovitus.

Kuvasta 3.11 nähdään, että stabiilien jakaumien perhe vaikuttaisi kohtuullisen hyvältä tavalta mallintaa tulojakaumaa sen keskivaiheilla.Ylähäntä näyttäisi myös seuraavan kohtuullisen tarkasti empiiristä histogrammia. Toisaalta kuvasta 3.12 huomataan, että sovitettu stabiili jakaumakaan ei aivan pysty tulojakauman hännän täydelliseen mallinnukseen huolimatta neljästä käytössä olevasta parametrista. Näistä neljästä parametrista täytyy tosin huomioida, että β-parametri on tehdyissä sovituksissa lähes jatkuvasti ääripäässään 1. Raskaan ylähännän mallintaminen on se ominaisuus, jossa stabiilit jakaumat päihittävän selvästi lognormaalin ja gamma-jakauman. Tämä pätee erityisesti viime vuosiin, jolloin jakaumaan on syntynyt superparetohäntä. Stabiili sovitus tuntuu jopa kohtuullisen hyvin seuraavan superparetohäntää.

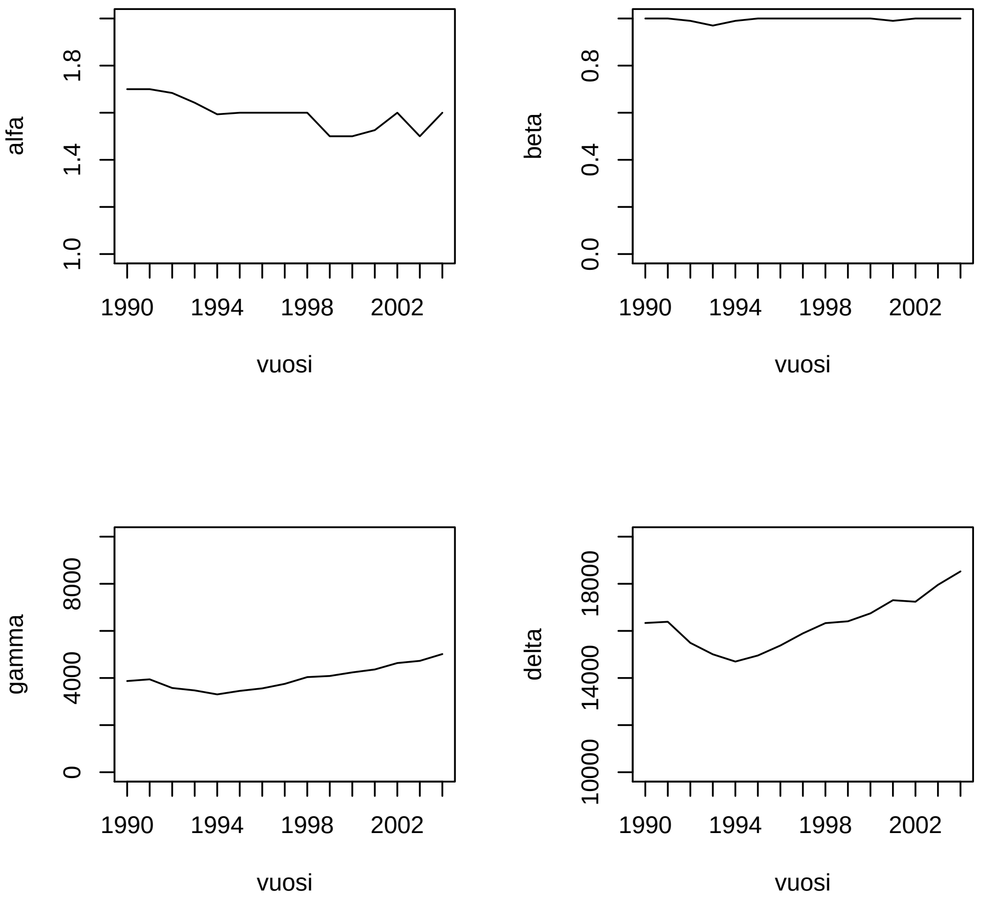
Kuvat 3.13 ja 3.14 illustroivat stabiilien sovitusten parametrien kehitystä absoluuttisesti ja suhteellisesti vuosina 1990–2004.

[10] http://academic2.american.edu/ jpnolan/stable/stable.exe

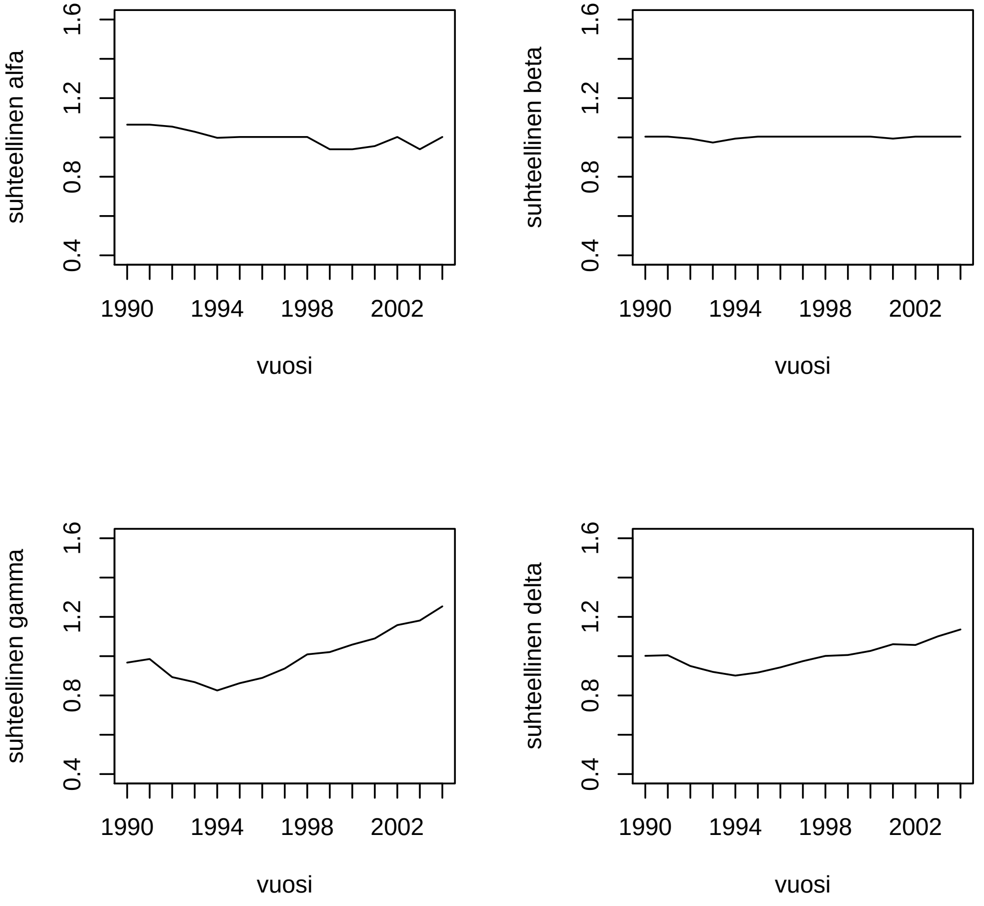
Kuva 3.12: Empiirinen laskeva kertymäfunktio Suomen vuoden 2004 tulojakaumasta ja siihen sovitettu laskeva stabiilin jakauman kertymäfunktio log-log -asteikossa parametreillä $(\alpha;\beta;\gamma;\delta) = (1,6;1;5016,05;18525,2)$.



Kuva 3.13: Vuosien 1990-2004 empiiriseen tulovektoriin tehtyjen stabiilien sovitusten parametrien kehitys.



Kuva 3.14: Vuosien 1990-2004 empiiriseen tulovektoriin tehtyjen stabiilien sovitusten parametrien keskiarvoon suhteutettu kehitys.



3.2.6 Sovitusten hyvyys

Edellä on esitelty kolme eri todennäköisyysjakaumaa, joita on sovitettu havaittuun tulojakaumaan. Silmämääräisesti näyttäisi, että stabiilit jakaumat kuvaavat parhaiten tulojakaumaa. Silmämääräistä tarkastelua analyyttisempi tapa arvioida sovitusten hyvyyttä on osittaa aineisto mielivaltaisesti valittuun n:ään luokkaan, joiden kokoa kuvaa n_i, missä $i \in \{1, 2, \ldots, l\}$. Luokat muodostetaan väleistä. Jokaiselle tällaiselle välille saadaan myös teoreettisesta todennäköisyysjakaumasta laskettu vastaava teoreettinen todennäköisyys p_i. Näitä käsitteitä käyttäen voidaan laskea neliövirheiden summa

$$SSE_Y = \sum_{i=1}^{l} \left(\frac{n_i}{n} - p_i \right)^2,$$

missä l on luokkien määrä, n on otoksen koko ja p_1 on kyseisen luokan teoreettinen todennäköisyys. Tämä mittari ei kuitenkaan sinällään toimi järkevänä sovitusten mittana, vaikka sitä usein kirjallisuudessa siihen tarkoitukseen käytetäänkin.

SSE:n pohjalle rakentuva Pearsonin khiin neliöestimaattori (Pearson's chi-squared) on hyvä sovituksen hyvyyden mittari ja se määritellään seuraavalla tavalla:

$$\chi^2 = \sum_{i=1}^{l} \frac{(o_1 - e_1)^2}{e_1} = n \sum_{i=1}^{l} \frac{\left(\frac{n_i}{n} - p_i\right)^2}{p_i},$$

missä o_1 on havaittu frekvenssi luokassa i ja e_i on odotettu frekvenssi tai sen estimaatti luokassa i. (Vartia – Vartia 1981, 31–34; Vasama – Vartia 1972, 534–543; McDonald – Ransom 1979.)

Aineisto jaettiin 43:een luokkaan, ja otoskoko n on 11 229. Taulukosta 3.6 nähdään sovituksen hyvyys neliövirheiden ja χ^2:n mielessä. Korkea χ^2:n arvo tarkoittaa huonoa sovitusta. 38 vapausasteella[11] 0,05 %:n merkitsevyystasolla kriittinen arvo on noin 53. Nollahypoteesi jakauman muodosta tulee hylätyksi kaikilla sovituksilla. Stabiililla sovituksella testisuure saa alhaisimman arvonsa. Skaalatun F-jakauman sovituksista momenttimenetelmän ja SU-menetelmän yhdistelmä pääsee lähelle stabiilien jakaumien tarkkuutta χ^2:n mielessä. Puhtaasti momenttimenetelmällä tehty sovitus tuntuu korostavan niin paljon ylähäntää, että parempi ylähännän sovite tulee muun sovitteen tarkkuuden kustannuksella.

Parhaan sovituksen valitseminen on siis monella kriteerillä tässä tapauksessa melko helppoa. Stabiili perhe on tarkin kuvaus tulojakaumasta χ^2:n mielessä.

[11] Nollahypoteesin vallitessa χ^2-testisuure noudattaa jakaumaa $\chi^2(l - p - 1)$, missä l on luokkien määrä ja p on parametrien määrä.

Stabiilin perheen erityisyyksiin kuuluu, että se kykenee mallintamaan raskasta ylähäntää. Neljän parametrin olemassaolo tekee tietenkin stabiilien jakaumien käyttämisen hankalaksi. Näiden neljän parametrin tulkitseminen on myös hankalaa siksi, että eri α:n arvoilla muut parametrit saavat uuden tulkinnan. Koska stabiilit jakaumat ovat kuitenkin sovitetuista malleista ainoita, jotka kykenevät mallintamaan tulojakaumalle ominaisen korkean ylähännän käyttäytymistä edes kohtuullisesti, muiden jakaumien käyttäminen muuhun kuin jakauman keskivaiheen ja alahännän mallintamiseen olisi vaikea perustella[12].

Taulukko 3.6: Neliövirheiden summat kolmella eri tulojakauman teoreettisella sovituksella 43:n luokkaan jaetuille Suomen vuoden 2004 tulojakaumalle, kun $n = 11229$. Kriittinen arvo on noin 24.

Jakauma	chi^2
Stabiili	173,1
Skaalattu F momentti/SU	190,6
Skaalattu F momentti-iteraatiolla	283,1
Lognormaali	279,8
Gamma	1113,2

Stabiilien jakaumien on siis monelta ominaisuudeltaan todettu kuvaavan käytetyistä jakaumatyypeistä parhaiten tulojakaumaa. Seuraavaa ilmeinen kysymys onkin, kuvaako se tulojakaumaa tarpeeksi hyvin. Keskeinen laadullinen vaatimus jakauman varianssittomuudesta tulee täytettyä. On kuitenkin ilmeistä, että monissa kohdissa teoreettinen ja empiirinen tulojakauma muodostavat erilaiset kuviot ja tuleehan χ^2-testi hylättyä 11 229:n otoksella.

Tässä vaiheessa on hyvä nostaa esille harmittoman yksinkertaistuksen käsite (Vartia 1988, 307). Tällä käsitteellä tarkoitetaan sitä, että mallissa voidaan hyväksyä pientä yksinkertaistamista tai totuudenvastaisia piirteitä, "jos virheen korjaaminen todenmukaiseksi ei muuta johtopäätöksiä." Tässä yhteydessä keskeinen laadullinen seikka, jota mallilta on vaadittava tarpeeksi tarkan sovitteen lisäksi, on varianssittomuus. Stabiilit sovitukset siis ainakin täyttävät tämän ehdon, mutta lopullinen tuomio sovitteen hyvyydestä jää lopulta hieman epävarmaksi.

[12] Myös muita, lognormaalin ja Pareto-jakauman ominaisuuksia yhdisteleviä jakaumia, kuten kaksoispareto–lognormaali-jakauma, on kirjallisuudessa ehdotettu. Katso esimerkiksi Reed (2003), Reed – Jorgensen (2004) ja Montroll – Shlesinger (1982).

4 Tuloerot deskriptiivisessä mielessä

Kuvailevaa yhteiskuntatieteellistä tutkimusta pidetään Amartya Senin (1980) mukaan yleensä epähaasteellisena tai vähintäänkin tylsänä. Sen argumentoi kuitenkin että näin ei ole, sillä hyvä kuvaus vaatii aina kykyä valikoida oikein. Pitää osata valita kuvailutavoista paras.

Totuudenmukaisuus ei esimerkiksi ole riittävä hyvän kuvailun ehto. Kuvailun on oltava myös merkityksellistä. Itse asiassa totuudenmukaisuus ei yleisesti käytetyillä taloustieteellisillä kriteereillä ole edes välttämätön ehto. Milton Friedman (1953) on kehittänyt metodologiaa, jonka mukaan taloustieteellisen mallin deskriptiivinen hyvyys ei riipu sen totuudenmukaisuudesta vaan ennustuskyvystä. Makrotalousteoriassa on yleisesti myös tapana vaatia, että malli on yhdenmukainen "tyyliteltyjen tosiasioiden" kanssa, joista monet eivät todistetusti ole tosiasioita kuin likimain.

Tuloerojen tarkastelu deskriptiivisessä mielessä vaatii kykyä valita kyseiseen kuvailuun sopiva abstraktiotaso ja sen jälkeen kykyä valita tarjolla olevista metodeista ja mittareista paras. Monessa tilanteessa edellisessä luvussa esitellyt visuaaliset tavat kuvata koko tulovektoria voivat olla huomattavasti havainnollisempia kuin yhden yksittäisen tulovektorista johdetun tunnusluvun antaminen. Tämä johtuu osittain siitä, että ymmärryksemme käsitteestä tuloerot on jo sinällään niin sumea ja moniulotteinen, että siitä on mahdotonta esittää eksaktia yksiulotteista kvantifiointia[13].

4.1 Tunnusluvut ja niiden tyhjentävyys

Kaikki tässä työssä esitettävät tuloeroindeksit ovat dispersiota kuvaavia tunnuslukuja annetusta tulovektorista. Jokaisen niistä käyttämiselle on olemassa perusteet. Eri tilanteet vaativat eri tunnuslukuja. Hyvältä tuloeromittarilta vaaditaan kykyä kuvailla dispersio-omininaisuuksia tulovektorin eri pisteissä.

Tyhjentävän tunnusluvun käsite liittyy parametrisen tilastollisen päättelyn asetelmaan, jossa satunnaismuuttujan tai -muuttujien tiheysfunktiot oletetaan tunnetuiksi. Näissä tapauksissa aineistosta laskettu tyhjentävä tunnusluku antaa kaiken relevantin informaation, joka aineistosta on saatavilla suhteessa johonkin tiettyyn tiheysfunktion parametriin (Nieminen – Saikkonen 2006). Tämä

[13] Taloustieteessä ongelma on yleisempikin. Sen (1973) pohtii tätä problematiikkaa havainnollisesti.

vaatii sen, että taustalla olevia todellisia tiheysfunktioita tutkitaan, tai että niistä tehdään joitakin oletuksia (kts. luku 3.2). Tällaisessa tapauksessa yksi parametri voi tyhjentävästi kuvata koko dispersiota.

Ainoa tyhjentävä tunnusluku kaikissa tilastollisen päättelyn tilanteissa on koko tulovektori itse. Jos tuloerotutkimuksen kohteena on ainoastaan tuloerojen merkitys irrallaan muista muuttujista, järjestetty tulovektori eli järjestystunnusluku tarjoaa kaiken tarvittavan informaation. Koko tulovektori ja järjestetty tulovektori ovat vektorimuotoisia tunnuslukuja.

4.2 Tuloerojen mittaaminen ad hoc -indekseillä

Käytän deskriptiivisten tuloeroindeksien jakoa kahteen kategoriaan: ad hoc -indekseihin ja aksiomaattisiin indekseihin. Käsittelen tässä tarkimmin jälkimmäistä kategoriaa. Monet ad hoc -kategorian indeksit ovat intuitiivisia, mutta tarkemmassa tarkastelussa ne tuottavat vaikeasti tulkittavia tuloksia.

4.2.1 Lorenz-käyrä ja Gini-kerroin

Gini-kerroin on tunnettu ja laajalti käytetty mittari, joka kuuluu ad hoc -kategoriaan. Gini-kerroin kertoo suhteen Lorenz-käyrän ja täysin tasaisen tulonjaon käyrän välillä. (Lorenz 1905, Cowell 1977 ja Gastwirth 1971.) Jos siis alla olevan kuvan 4.1 alue Lorenz-käyrän ja tasaisen tulojaon käyrän eli 45 asteen käyrän välillä on alue A ja Lorenz-käyrän alapuolinen alue on B, Gini-kerroin muodostuu yhtälöstä

$$G = \frac{A}{A+B}.$$

Lorenz-käyrä määritellään seuraavalla tavalla:

$$L(y) = \frac{\int_{\underline{y}}^{y} x \, dF(x)}{\overline{x}}.$$

Funktionaali on Riemann-Stieltjes -integraali, jolla voidaan käsitellä sekä diskreettiä että jatkuvaa tapausta. (Spanos 1986, 69.) Diskreetissä, empiirisessä tapauksessa, yhtälö saa muodon

$$L(y) = \frac{\sum_{i=1}^{n} f[x(a_i)]x(a_i)}{\overline{x}},$$

missä $f[x(a_i)]$ on tulojen tai vaurauden mukaan järjestettyjen yksilöiden a_i tulojen tai vaurauden pistetodennäköisyysfunktio ja $\overline{x}$ on keskimääräiset tulot tai vauraus, riippuen kumpaa mitataan. Kun on tiedossa jatkuva tiheysfunktio tuloille, Lorenz-käyrä lasketaan seuraavalla tavalla: (Weisstein a)

$$L(y) = \frac{\displaystyle\int_{\underline{y}}^{y} xf(x)\,dx}{\overline{x}}.$$

Yleensä pitää paikkansa, että $\underline{y} \geq 0$. Nollatuloja pienempien tulojen ollessa mahdollisia pitää sekä Riemann että Riemann-Stieltjes -integraalit laskea pienimmästä tulosta $\underline{y} < 0$ lähtien. Tämä hankaloittaa koko Lorenz-käyrän käsitteen tulkintaa, sillä siinä tapauksessa Lorenz-käyrä on aluksi laskeva.

Lorez-dominanssilla tarkoitetaan sitä tilannetta, jossa kaksi Lorenz-käyrää eivät leikkaa. Tässä tapauksessa käyristä se, joka on koko ajan toisen yläpuolella, Lorenz-dominoi toista käyrää. Kuva 4.1 illustroi tilannetta, jossa toinen käyrä näyttäisi Lorenz-dominoivan toista. Itse asiassa näin ei tarkasti ottaen ole, jos tarkennetaan tarkastelua tulojakauman alahäntään (Kuva 4.2).

Empiirisessä tapauksessa Gini-kerroin voidaan laskea kaavalla

$$G = \frac{n}{n-1}\frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} \mid x(a_i) - x(a_j) \mid}{2n^2\mu},$$

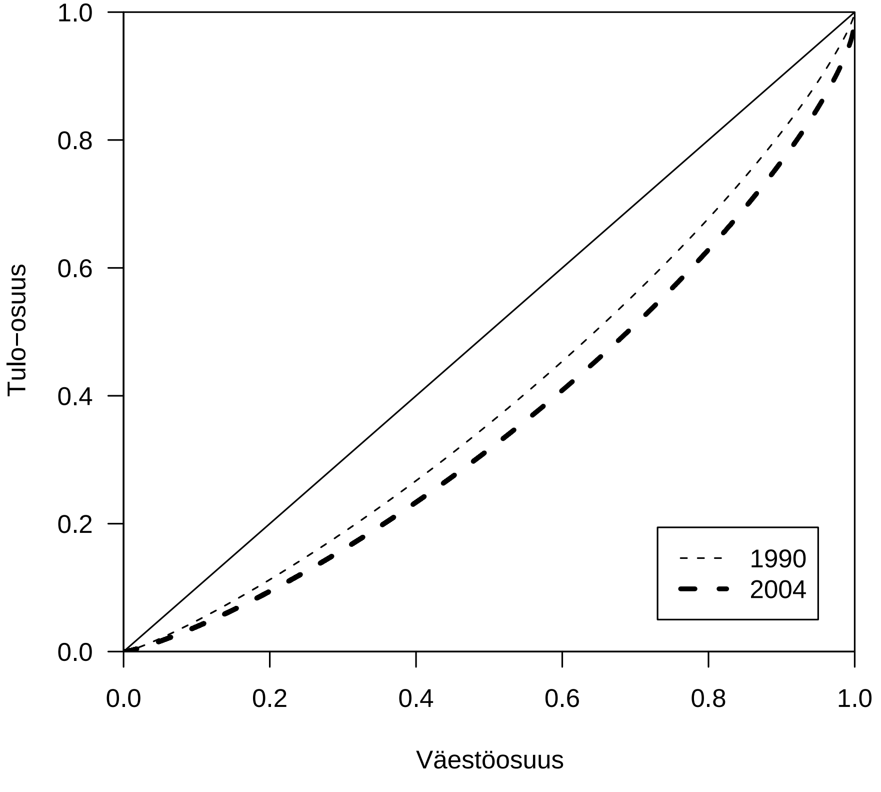
eli intuitiivisemmin esitettynä lasketaan kaikkien mahdollisten yksilöparien välisten etäisyyksien keskiarvo jaettuna keskiarvolla μ. (Weisstein b.)

Kuvat 4.3 ja 4.4 tuntuisivat paljastavan, että jonkinlainen rakenteellinen muutos Suomessa on tapahtunut vuosien 1996 ja 1999 välillä.

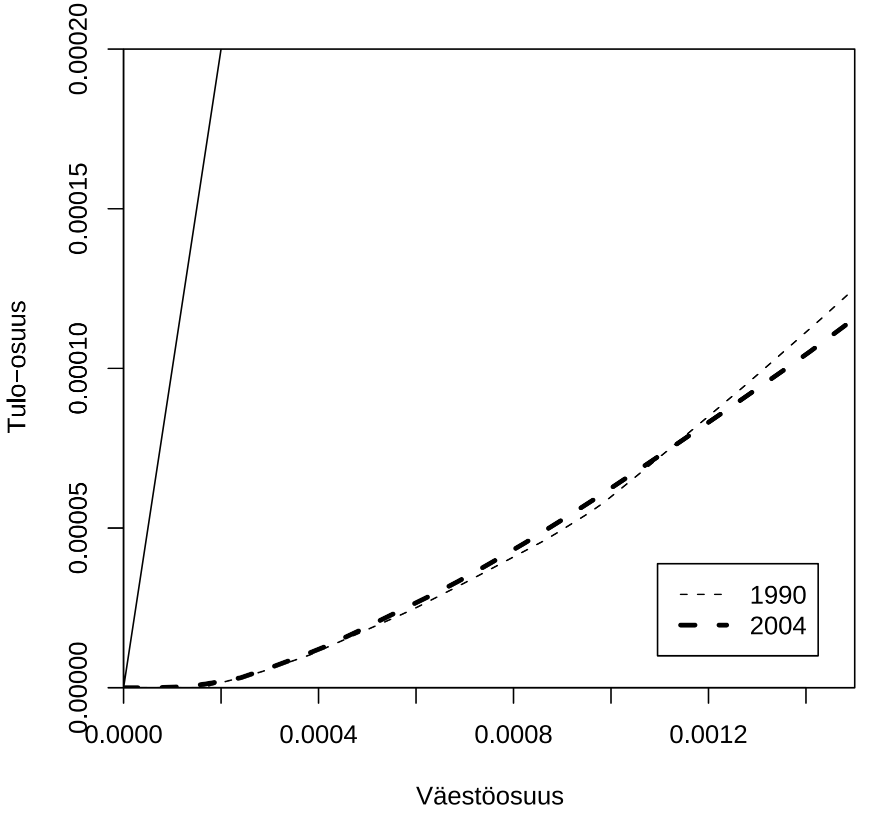
Gini-kerroin on ylivoimaisesti suosituin käytössä oleva tuloeromittari (Cowell 2000). Gini-kertoimen käyttämisessä on kuitenkin huomattava määrä ongelmia. Eräs Gini-kerroinn heikkous on se, että muutokset keskivaiheilla tulojärjestystä vaikuttavat indeksiin enemmän kuin sen ääripäissä. Gini-kertoimen antama arvo ei myöskään kerro mitään käyrän muodosta, eli kaksi samaa Gini-kertoimen lukemaa voi ilmaantua hyvinkin erilaisista Lorenz-käyristä. Otetaan tästä ongelmasta pieni imaginäärinen esimerkki.

Oletetaan että meillä on kaksi maata. Kutsutaan ensimmäistä maata nimellä Antikreik. Antikreikissä väestöstä tasan puolet ovat orjia ja toisen puolikkaan

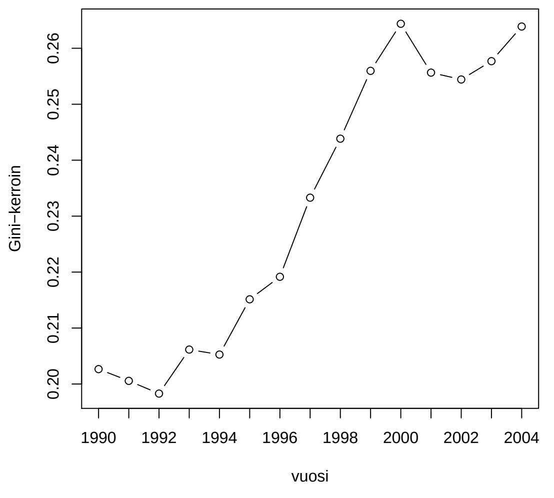
Kuva 4.1: Suomen vuosien 1990 ja 2004 käytettävissä olevien tulojen Lorenz-käyrä laskettu kotitalouden modifioitua OECD-kulutusyksikköä kohden.



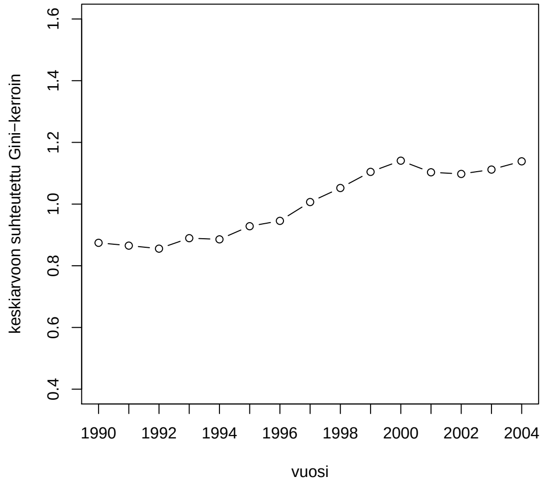
Kuva 4.2: Suomen vuosien 1990 ja 2004 käytettävissä olevien tulojen Lorenz-käyrä laskettu kotitalouden modifioitua OECD-kulutusyksikköä kohden, kun väestöosuus $\in (0; 0,0015)$.



Kuva 4.3: Suomen käytettävissä olevien tulojen gini-kerroinestimaatit vuosille 1990-2004.



Kuva 4.4: Suomen keskiarvoon suhteutetut käytettävissä olevien tulojen Gini-kerroinestimaatit vuosille 1990-2004.

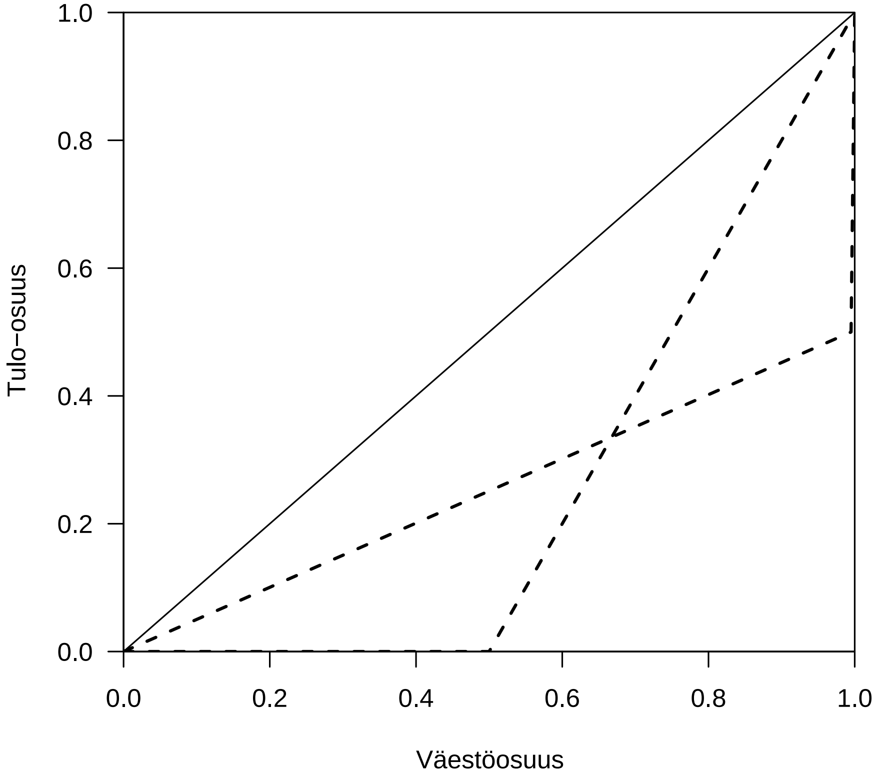


muodostavat vapaat kansalaiset. Orjien tulot ovat nolla, ja kaikilla vapailla kansalaisilla tulot ovat neljä. Gini-kerroin on 0,5. Väestön koon on oltava parillinen luku.

Toinen maa saakoon nimen Komdikt. Komdiktissä kaikilla muilla tulot ovat yksi ja yhdellä tulot ovat populaation koko plus yksi. Tämäkin Gini-kerroin on likimain 0,5, kun kyseessä on suuri populaatio. Voidaan kuitenkin aiheellisesti kysyä, onko nyt todella kyseessä kaksi populaatiota, joilla on samat tai samanlaiset tuloerot. Jos ainoana tunnuslukuna on Gini-kerroin 0,5, emme voi antaa tähän kysymykseen yksiselitteistä vastausta. Gini-kertoimen määritelmä sisältää siis tietynlaisen tavan arvottaa erilaisia tulovektoreita samanarvoisiksi.

Lorenz-käyrä (Kuva 4.5) ja vaikkapa tulojen kertymäfunktio kertoisivat meille suoraan, että kyseessä on kaksi populaatiota, joilla on huiman erilaiset tuloerot.

Kuva 4.5: 2 Lorenz-käyrää Gini-kertoimella 0,5.



Gini-kertoimen puutteiden joukkoon kuuluu myös, että se ei täytä myöhemmin käsiteltävää vaatimusta, jonka mukaan tuloeroindeksin tulee olla purettavissa osiin, eli että jokainen osaryhmä operoi erikseen.

Muita Lorenz-käyrää käyttäviä tuloeroja karakterisoivia mittareita ovat pienimmän enemmistön (minimal majority) ja samojen osuuksien kertoimen (equal shares coefficient) mittarit. Molemmat yrittävät käsitellä epätasa-arvoa vallan jakaantumisen näkökulmasta. Näistä ensin mainittu käyttää tulotasoa y*, jonka molemmilla puolin on yhtä paljon kumulatiivista tuloa. Silloin tämän pisteen yläpuolelle asettuvat ihmiset edustavat pienintä enemmistöä, joka voi saada enemmistöpäätöksen läpi valtaelimessä, jossa valta jakaantuu epätasaisesti Lorenz-käyrän mukaan.

Saman osuuden kerroin lähestyy asiaa päinvastaisesta näkökulmasta eli siitä, kuinka suurella osalla ihmisistä on keskimääräinen määrä valtaa tai vähemmän.

Näitä molempia voidaan soveltaa tuloeroihin siitä näkökulmasta, että tulot edustavat resurssien hallintaa eli valtaa. Kuitenkin varsinkaan ensimmäinen eli pienimmän enemmistön mittari ei sovellu hyvin talouteen, sillä on hyvin epätodennäköistä, että syntyisi tilanne, jossa juuri puolet tuloista tai vauraudesta saavalla osalla väestöstä sattuisi olemaan eri päämäärät kuin koko loppuosalla väestöä. Teknisemmällä tasolla nämä molemmat mittarit ovat siinä mielessä vajavaisia, että mikä tahansa siirto tuloissa, joka jää määritellyn keskipisteen samalle puolelle, ei vaikuta mitenkään näihin mittareihin.

4.2.2 Varianssiin perustuvat tuloeromittarit

Eräs tuloeromittari on luonnollisesti tulojen varianssi, jossa on monta ongelmaa tuloerojen mittaamisessa. Ensimmäinen on se, että kaikilla jakaumilla ei välttämättä ole varianssia, jos niillä on korkeat hännät. Tämä tuntuisi olevan totta myös Suomen tulojakauman kohdalla.

Toinen ongelma varianssissa tuloerojen mittaamisen välineenä on se, että jos kaksinkertaistamme kaikkien tulonsaajien tulot, tämä tuloeromittari nelinkertaistuu. Esimerkiksi Suomessa tulopoliittiset kokonaisratkaisut tehdään sillä tavalla, että kaikki tulonsaajat a_i, $i \in \{1, 2, \ldots, n\}$ saavat periodin t tuloistaan $y_t(a_i)$ osuuden b lisää tuloa periodille $t+1$. Tämä tietenkin nostaa todennäköisyysjakauman odotusarvoa uuteen arvoon

$$E[y_{t+1}] = E[(1+b)y_t] = (1+b)E[y_t],$$

mutta se skaalaa koko jakaumaa samalla b:llä. Varianssi siis kasvaa

$$Var[y_{t+1}] = Var[(1+b)y_t] = (1+b)^2 Var[y_t].$$

Tämän ongelman voi kiertää standardisoimalla tulojen $\mathbf{y}$ varianssin $Var(\mathbf{y})$ siten, että variaatiokerroin

$$c = \sqrt{\frac{Var(\mathbf{y})}{\overline{\mathbf{y}}}},$$

missä $\overline{\mathbf{y}}$ on tulojen keskiarvo. Toinen tapa muokata varianssia on käsitellä tulojen logaritmien varianssia. Näillä muunnoksilla varianssi saadaan suhteellisten tulojen muutosten suhteen muuttumattomaksi ja siten paremmaksi tuloeroindeksiksi. Variaatiokertoimen c yksi ominaisuus on, että yhden yksikön muutos jakauman ylä- ja alapäässä vaikuttaa aina yhtä paljon c:hen, joten se sopii hyvin tulojakauman yläpään tuloerojen mittaamiseen, mutta ei niinkään alapään. Tämä johtuu siitä, että samankokoinen suhteellinen muutos ylä- ja alapäässä tulojakaumaa aiheuttaa absoluuttisesti huomattavasti suuremman muutoksen jakauman yläpäässä. Tulojen logaritmien varianssi taas toimii juuri päinvastoin suorastaan liioitteluun asti, sillä suurissa tuloissa siirto rikkaammalta köyhemmälle saattaa jopa kasvattaa sitä (Cowell 1977).

Tässä työssä jätetään laskematta variaatiokerroin- ja varianssiestimaatit. Tämä johtuu siitä, että tulojakaumat yleensä noudattavat jakaumaa, jolla ei ole varianssia määritelty. Samoin on kolmannen luvun tulosten valossa Suomen tulojakauman laita.

4.3 Aksiomaattiset indeksit

Toinen kategoria deskriptiivisiä tuloeroindeksejä perustuu aksiomeihin, joita pidetään olennaisina hyvälle tuloeromittarille. Tähän kategoriaan kuuluvat kehittyneimmät skalaariset tuloeromittarit, joita nykyään käytetään tieteellisessä tutkimuksessa. Seuraan tässä Cowellin (2000) esitystä.

4.3.1 Anonymiteetin periaate

Aikaisemmin tässä työssä keskusteltiin tulovektorin permutoinnista ja järjestämisestä. Anonymiteetin periaate vaatii, että tuloeromittari tuottaa saman tuloksen riippumatta tulonsaajien vektorin **a** permutaatiosta. Tämä periaate ei pidä siinä tapauksessa, että tahdotaan käyttää hyväksi muitakin muuttujia jokaista tulonsaajaa kohden kuin tulot. Silloin tulee merkitykselliseksi tietää kehen tulonsaajaan a_i tulot $y(a_i)$ ja muuta tietoa liitetään.

4.3.2 Populaation periaate

Populaation periaate (Principle of Population) tarkoittaa, että epätasa-arvon ei tulisi riippua siitä, kuinka moni ihminen on jakamassa kakkua. Jos esimerkiksi yhdistämme kaksi identtistä populaatiota identtisillä tuloeroilla, mittarin pitäisi antaa aina sama arvo. Tämän periaatteen pitäessä voimme tarkastella eri populaatioiden tulovektoreiden tunnuslukuja samanmerkityksisinä riippumatta populaation koosta.

4.3.3 Pigou-Dalton siirtoperiaate

Pigou-Dalton siirtoperiaate (Pigou-Dalton Transfer Principle, Dalton 1920) vaatii hyvältä tuloeromittarilta, että se saa suuremman arvon, kun siirretään keskiarvoa muuttamatta tuloja köyhemmältä rikkaammalle. Useimmat mittarit luonnollisesti täyttävät tämän vaatimuksen. Mutta kuten jo aiemmin todettiin, joskus tuloerojen mittaamisessa käytetty tulojen logaritmien varianssi ei täytä tätä vaatimusta.

4.3.4 Tulojen skaalainvarianssin periaate

Tulojen skaalainvarianssin periaate (Income Scale Invariance Principle) toteaa, että hyvän tuloeromittarin tulisi olla nollannen asteen homogeeninen. Jos siis skaalaamme kaikkien tuloja samalla mitalla, tuloeromittarin tulisi antaa sama luku. Varianssi tuloeromittarina ei täytä tätä vaatimusta. Tähän aksioomaan viitattiin aiemmin analyysissä tulojen varianssista tuloeromittarina.

Tästä periaatteesta heikompi siirtoinvarianssin (Translation Invariance) periaate varmistaa sen, että jos siirretään jakaumaa eli lisätään kaikkien tuloihin äärellinen x määrä tuloja siten, että $x \in \mathbb{R}$, niin sen ei tulisi vaikuttaa tuloerosuureeseen. Varianssi täyttää tämän heikomman periaatteen antaman vaatimuksen.

4.3.5 Osiinpurettavuuden periaate

Osiinpurettavuuden periaate (Principle of Decomposability) vaati indeksiltä sen ominaisuuden, että epätasa-arvo on laskettavissa ryhmien sisäisen ja ryhmien välisen epätasa-arvon kombinaationa (Shorrocks 1980, Bourguignon 1979, Cowell 1980). Ryhmillä voidaan tarkoittaa esimerkiksi eri maita. Tätä ehtoa kutsutaan myös nimellä aggregoitavuus.

Tiukempi versio edellisestä ehdosta on additiivisen osiinpurettavuuden ehto, joka nimensä mukaan vaatii, että yhteisepätasa-arvo tulee olla laskettavissa ryhmien sisäisen ja ryhmien välisen epätasa-arvon summana (Shorrocks 1984). Käyn tästä eteenpäin läpi joitakin tämän tiukemman ehdon alaisia indeksejä.

Usein on olemassa arvioita alueiden sisäisistä ja välisistä tuloeroista. Jos indeksi täyttää additiivisen osiinpurettavuuden ehdon, näitä indeksejä summaamalla voidaan laskea tuloerot aggregoidulla tasolla. Esimerkiksi Gini-kerroin ja jatkossa esiteltävä Atkinsonin indeksi eivät täytä tätä ehtoa. Tämän ehdon ongelma on kuitenkin se, että usein käsityksemme tuloeroista on vahvasti riippuvainen kaikkien muiden tulonsaajien tuloista (Sen 1973, 41). Tulomittari, joka voidaan laskea erikseen operoivien ryhmien yhdistelmänä ei sovi yhteen tämän intuitiivisen käsityksen kanssa.

4.4 Yleistetyt entropiaindeksit

4.4.1 Määritelmä

Theilin (1967) kehittämät yleistetyt entropiaindeksit (YEI, Generalized Entropy Index, GEI) ovat diskreetissä tapauksessa seuraavaa yleistä muotoa:

$$GEI(\theta) = \frac{1}{n}\frac{1}{\theta - \theta^2}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\overline{y}}\right)^{\theta} - 1\right], \tag{5}$$

missä i = 1,...,n indeksoi populaatiota ja $\theta \notin \{0,1\}$. Tapaukset $\theta \in \{0,1\}$ ratkaistaan raja-arvoina.

Yleisimmät erikoistapaukset YE-indekseistä ovat θ:n arvot 0, 1 ja 2. Kun θ lähenee nollaa, l'Hôpitalin säännöllä saamme:

$$GEI(0) = MLD = \frac{1}{n}\sum_{i=1}^{n} ln\left(\frac{\overline{y}}{y_i}\right).$$

MLD tulee sanoista Mean Logarithmic Deviation eli keskimääräinen logaritmien poikkeama. Se voidaan myös kirjoittaa muotoon:

$$GEI(0) = MLD = ln(\overline{y}) - \frac{1}{n}\sum_{i=1}^{n} ln\, y_i,$$

mistä ilmenee hyvin, että se on erotus keskimääräisten tulojen logaritmista ja tulojen logaritmien keskiarvosta.

Kun θ lähenee yhtä, niin taas l'Hôpitalin sääntöä käyttämällä

$$GEI(1) = T = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\overline{y}} ln\left(\frac{y_i}{\overline{y}}\right)\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\overline{y}} ln\, y_i\right) - ln\,\overline{y}$$

GEI(1):tä kutsutaan myös nimellä Theilin indeksi T.

Viimeinen erityistapaus, joka käydään läpi, on kun $\theta = 2$. Silloin

$$GEI(2) = \frac{1}{2}c^2 = \frac{1}{2n\overline{y}^2}\left(\sum_{i=1}^{n} y_i^2 - \overline{y}^2\right).$$

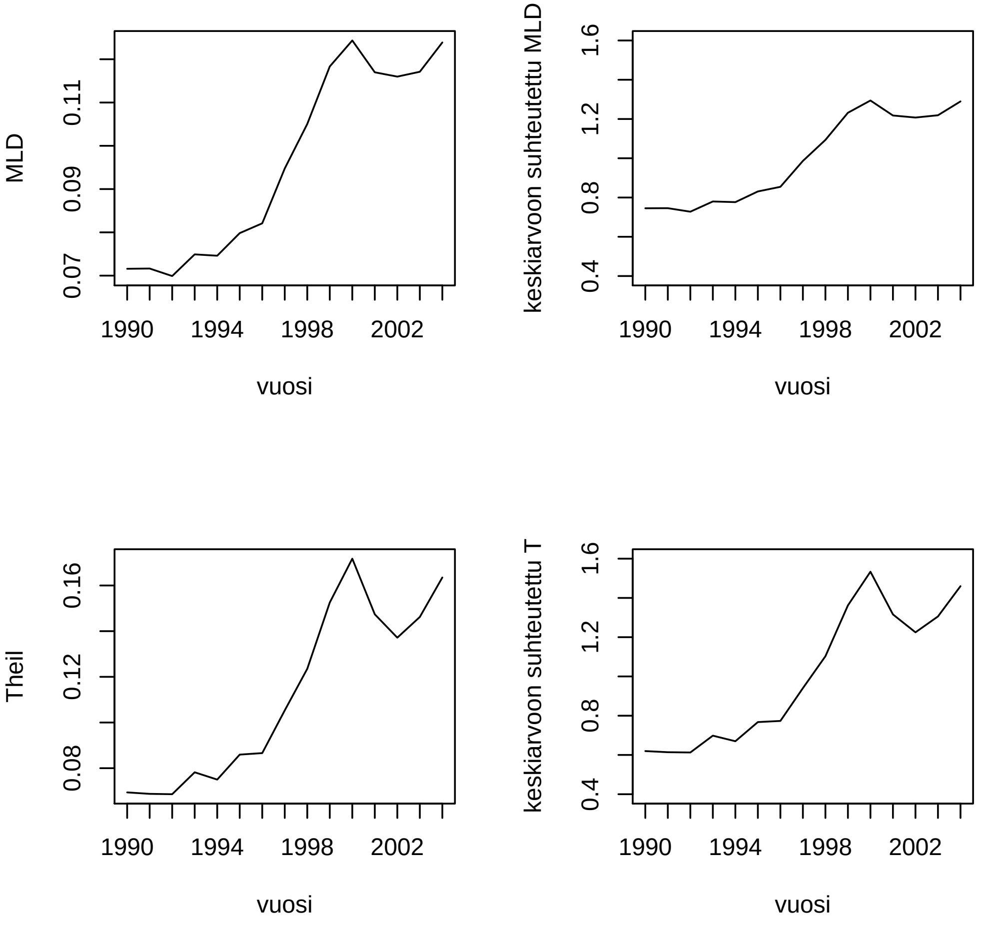
Tämä on puolet variaatiokertoimen neliöstä.

Parametri $\theta \in (-\infty, \infty)$ määrittää indeksin herkkyyden muutoksille eri osissa jakaumaa. Suurella θ:n arvolla indeksi on herkkä jakauman yläpäässä tapahtuville muutoksille ja pienellä θ:n arvolla se on herkkä jakauman alapäässä tapahtuville muutoksille.

YE-indeksit täyttävät yllä annetuista aksioomista Pigou-Dalton siirtoperiaatteen, skaalainvarianssin periatteen ja osiinpurettavuuden periaatteen. Itse asiassa mikä tahansa indeksi täyttää nämä periaatteet, jos ja vain jos se on ordinaalisesti ekvivalentti yhtälön (5) kanssa jollakin parametrin θ arvolla. YE-luokan kanssa ordinaalisesti ekvivalentteja mittareita ovat muiden muassa variaatiokerroin, varianssi ja momentit (Cowell 2000).

Kuvaan 4.6 on kerätty YE-indeksien tulokset kun $\theta \in \{0, 1\}$. Nämä mittarit ovat ovat olemassa stabiileillakin jakaumilla, sillä näissä käsitellään tulojen logaritmejä. Kuten huomaamme, tulojakauman yläpäässä herkän Theilin indeksin suhteelliset muutokset ovat huomattavasti suurempia kuin MLD:llä tai Gini-kertoimella mitattuna. Näiden kardinaalisten erojen olemassaolo tuloeromittareiden välillä vaatii tarkkaavaisuutta mittareiden käyttäjältä.

Kuva 4.6: Suomen inflaatiokorjattujen käytettävissä olevien tulojen $GEI(\theta = 0) = MLD$:t ja $GEI(\theta = 1) = T$:t vuosille 1990-2004 absoluuttisesti ja keskiarvoonsa suhteutettuna.



4.4.2 Entropiaindeksit ja informaatioteoria

Edellä esitettyjen entropiaindeksien pohjana on informaatioteorian lähestyminen. Siinä tapahtumille annetaan todennäköisyydet $P_1, P_2, P_3, \ldots$ siten, että $0 \leq P_i \leq 1$ ja $i \in \{1, 2, \ldots, n\}$. Annamme tapahtuneiden tapahtumien tapahtumiselle informaatioarvon $h(P_i)$. h:n derivaatta $h' < 0$, joka on funktio tapahtuman todennäköisyydestä. Funktio, joka täyttää myös lisäksi vaaditun ehdon $h(P_1P_2) = h(P_1) + h(P_2)$ on $h = -log(P)$. Näiden informaatioarvojen todennäköisyyksillään painotettu summa on järjestelmän entropia eli

$$\sum_{i=1}^{n} P_i h(P_i) = -\sum_{i=1}^{n} P_i (log P_i).$$

Tulkitaan tapahtumat n ihmisinä ja todennäköisyys P_i yksilön i tulo-osuutena vaikkapa nimellä s_i, missä

$$s_i = \frac{y_i}{n\overline{y}}.$$

Nyt vähennetään järjestelmän eli talouden entropia teoreettisesta maksimistaan, joka saadaan kun $s_i = \frac{1}{n}$. Saamme siis

$$\sum_{i=1}^{n} \frac{1}{n} h\left(\frac{1}{n}\right) - \sum_{i=1}^{n} s_i h(s_i) = \sum_{i=1}^{n} s_i \left[h\left(\frac{1}{n}\right) - h(s_i)\right]$$

$$= \sum_{i=1}^{n} s_i \left[log(s_i) - log\left(\frac{1}{n}\right)\right] = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{y_i}{\overline{y}} log\left(\frac{y_i}{\overline{y}}\right)\right] = T.$$

Näistä saman Theilin indeksin esittämismuodoista viimeinen on täysin sama muoto kuin jo ylempänä esitetty. (Shorrocks 1980, Bourguignon 1979, Cowell 1977, Theil 1967.)

4.4.3 YE-indeksien purkaminen ryhmien väliseen ja sisäiseen komponenttiin

YE-indeksien vahvuus on se, että jokainen tulonsaaja toimii riippumatta muista tulonsaajista, jolloin YE-indeksit voidaan laskea minkälaisesta tahansa osituksesta. Theilin indeksi voidaan kirjoittaa ositetussa muodossa

$$T = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{N_i} \left[\frac{y_{ij}}{\overline{y}} ln\left(\frac{y_{ij}}{\overline{y}}\right)\right] = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{N_i} \frac{y_{ij}}{\overline{y}} ln\, y_{ij} - ln\, \overline{y},$$

missä $i \in \{i, 2, \ldots, m\}$, m on osajoukkojen määrä, N_i on jokaisen osaryhmän koko ja n on yhteen lasketun otoksen koko. Indeksi voidaan hajottaa alueiden väliseen ja sisäiseen komponenttiin vähentämällä ja lisäämällä siihen

$$\sum_{i=1}^{m} \frac{N_i y_i}{n\overline{y}} ln\, y_i.$$

Saamme siis

$$T = \sum_{i=1}^{m} \frac{N_i y_i}{n\overline{y}} ln \left(\frac{y_i}{\overline{y}}\right) + \sum_{i=1}^{m} \frac{N_i y_i}{n\overline{y}} T_i,$$

missä Theilin indeksi kussakin ryhmässä i on

$$T_i = \frac{1}{N_i} \sum_{i=1}^{N_i} \frac{y_{ij}}{\overline{y_i}} ln\, y_{ij} - ln\, \overline{y_i}.$$

T_i siis antaa ryhmien sisäiset Theilin indeksit, kun muu osa yhtälöstä viittaa ryhmien välisiin tuloeroihin. Theilin indeksin painot ovat talouksien kokoja väestöjen kokojen sijaan. Rikkaiden ryhmien sisäiset tuloerot ovat siis tällä mittarilla merkittävämpiä kuin köyhien ryhmien.

Keksimääräinen logaritmien hajonta eli MLD painottaa ryhmiä väestön kokojen mukaan, ei talouden kokojen mukaan niin kuin Theilin indeksi. MLD:n osiinsa hajoitettu versio on

$$GEI(0) = MLD = \left[ln\, \overline{y} - \frac{N_i}{n} \sum_{i=1}^{m} ln\, y_i \right] + \sum_{i=1} m \frac{N_i}{n} \left[ln\, y_i - \frac{1}{N_i} \sum_{j=1}^{N_i} ln\, y_{ij} \right].$$

Ensimmäinen osa kertoo ryhmien välisen tuloepätasa-arvon ja toinen osa väestön koolla painotetut ryhmien sisäiset tuloerot.

5 Tuloerot preskriptiivisessä mielessä

Tuloeroista puhuttaessa oikeudenmukaisuuden käsite nousee usein esiin. Tästä tuloerojen merkityksen tulkinnasta käytetään sanaa preskriptiivinen. Seuraavassa luvussa käsitellään lyhyesti preskriptiiviset tavat karakterisoida tuloeroja.

5.1 Tuloerot, tasa-arvo ja oikeudenmukaisuus

Preskriptiivinen tuloerotutkimus tutkii tuloeroja oikeudenmukaisuuden näkökulmasta. Taustalla on jokin normatiivinen käsitys siitä, mikä on oikeudenmukainen tulojakauma. Jokaisessa poliittisessa filosofiassa on omanlaisensa lähestyminen epätasa-arvon käsitteeseen. Kuitenkin ihmisten tasa-arvoinen kohtelu ainakin jonkun muuttujan suhteen on kaikissa poliittisissa filosofioissa yleisesti hyväksytty päämäärä (Sen 1992). Yleensä koetaan tärkeäksi tavoitteeksi myös hyvinvoinnin korkea aggregaattitaso. Näiden kahden usein osittain toistensa kanssa ristiriidassa olevan tavoitteen suhteelliset painot vaihtelevat eri poliittisissa filosofioissa. Samoin vaihtelevat hyvinvoinnin määritelmät.

Poliittiset filosofiat kuitenkin eroavat toisistaan selvästi sen mukaan, minkä muuttujien suhteen tasa-arvoa vaaditaan. Jotkut polittiset filosofiat vaativat vain joidenkin perustavanlaatuisten oikeuksien ja vapauksien tasaista jakautumista. Toisissa filosofioissa, kuten pohjoismaisessa sosiaalidemokratiassa, tasa-arvon vaatimus ulottuu niinkin pitkälle kuin esimerkiksi koulutukseen, terveydenhuoltoon ja osittain myös tuloihin. Sosiaalidemokratiassa tasa-arvon vaatimus siis ylettyy hyvinvoinnin tasaamiseen saakka.

Koska yksilölliset kyvyt ja ominaisuudet vaihtelevat suunnattomasti, ihmisillä on epätasaisesti jakautuneet kyvyt muuttaa esimerkiksi tulot tai mitkä tahansa muut resurssit hyvinvoinniksi. Yrjö Vartia (luento) jakaa tasa-arvon vaatimuksen esimerkin nimissä neljään eri doktriiniin. Nämä neljä tasa-arvon määritelmän puhdasta idealisaatiota rajaavat tasa-arvon käsitteen siten, että ne määrittävät tasa-arvon kysymyksen tulojen kannalta olennaisiin osiinsa. Vartian mainitsemat doktriinit ovat:

1. Sama hyvinvointi kaikille

2. Samat tulot kaikille

3. Sama tuntipalkka kaikista töistä

4. Sama tuntipalkka samasta työstä

Tiukimman tasa-arvon vaatimuksen mukaan kaikille pitäisi saada sama hyvinvointi. Tämän tason vaatimus lienee lähes mahdotonta toteuttaa. Vaikka kaikilla olisi täysin samat tulot, se ei vielä riittäisi läheskään tuottamaan samaa hyvinvoinnin tasoa kaikille johtuen yksilöllisistä eroista kyvyissä muuntaa tulot hyvinvoinniksi.

Toinen tämän lähestymisen ongelma on se, että hyvinvointi käsitteenä on niin hankala. Erityisesti hyvinvoinnin kvantifiointi tuottaa vaikeuksia. Onnellisuuden tutkimus on yrittänyt ratkaista onnellisuuden käsitteellä näitä ongelmia (mm. Layard 2002). Ongelmallisuudestaan johtuen tämä ensimmäinen ehdotus toimii lähinnä mittapuuna seuraaville tasoille.

Toisella ja kolmannella tasolla ovat vaatimukset tulojen täydellisestä tasaamisesta ja samasta tuntipalkasta kaikille. Työn ulkoiset kannustimet työn tekemiseen ovat pieniä. Tulojen täydellinen tasajakautuminen ei kannusta taloudellisesti lainkaan töihin. Sama tuntipalkka kaikille kyllä kannustaa tekemään töitä, mutta oman työn tuottavuuden parantamisen kannustimet puuttuvat. Sosialismi perustuu tämänkaltaisille tasa-arvon vaatimuksille. Näitä vaatimuksia vaativassa järjestelmässä työn tekemisen kannustimien on tultava muualta kuin taloudesta. Tulojen vaikutus hyvinvointiin on suhteellisen tasaisesti jakautunut. Hyvinvoinnin jakauman määrittää samojen tulojen mallissa kyky muuntaa tulot hyvinvoinniksi ja muut hyvinvointiin vaikuttavat muuttujat.

Neljäs ja viimeinen Vartian mainitsema taso on vaatimus saman palkan saamiseksi samasta työstä. Vaikeuksia tuottaa tietenkin saman työn määritteleminen, mutta jos saman työn määritelmäksi otetaan kaikki saman alan työt, joissa on sama tuottavuus, vaatimus on selkeästi määriteltävissä siten, että siinä on konsistentti työn tekijän työn ulkopuolisista tekijöistä riippumaton tasa-arvon vaatimus. Tämä löyhä tasa-arvon määritelmä ei ota juurikaan kantaa siihen, kuinka tulot jakautuvat ihmisten välillä. Hyvinvointi myös tulojen funktiona saattaa siis olla hyvinkin epätasaisesti jakautunut.

5.2 Aksiomatiikkaa

Preskriptiiviseen tuloerojen mittaamiseen on kehitetty omia tuloeromittoja. Aksiomaattisessa lähestymisessä preskriptiivisiltä tuloeromittareilta oletetaan usein vielä edellisessä luvussa esitettyjen aksioomien lisäksi, että se täyttää niin sanotun monotonisuuden periaatteen. Monotonisuuden periaate vaatii, että preskriptiivinen tuloeromittari antaa suuremman hyvinvoinnin, jos jakaumasta siirretään mitä tahansa osaa oikealle, siis kenelle tahansa tulonsaajalle annetaan lisää tuloja.

Monotonisuuden periaatteesta heikompi versio eli yhtäläisen tulojen kasvun (Uniform Income Growth) periaate vaatii vain, että koko jakauman siirtämisen oikealle pitäisi tuottaa hyvinvoinnin kasvua.

5.3 Sosiaalisen hyvinvoinnin funktion indeksit

Eräs puhtaasti preskriptiivinen tapa lähestyä epätasa-arvoa ja sen mittaamista ovat sosiaalisen hyvinvoinnin funktiot (SHF, eng. Social Welfare Function) ja niistä johdetut epätasa-arvomittarit, jotka perustuvat yhteiskunnan antamiin arvoihin epätasa-arvosta ja oikeudenmukaisuudesta. Nämä indeksit siis integroivat eksplisiittisesti normatiivisen pohjan epätasa-arvon mittaamiseen.

SHF tarkoittaa funktiota, joka laittaa kaikki yhteiskunnan olotilat arvojärjestykseen jollakin kriteerillä. Olotilat voivat olla funktioita lähes mistä vaan, mutta yleensä järkevää on tarkastella tuloja tai vaurautta tai muuta relevanttia ja kvantifioitavaa määrettä.

Atkinsonin indeksi on eräs SHF-indeksi. Se suhteuttaa tasan jaetun ekvivalentin tulotason keskimääräiseen tulotasoon. Tasan jaettu ekvivalentti tulotaso tarkoittaa sitä samaa tulotasoa, joka kaikilla olisi oltava, jotta yhteiskunta pääsisi samaan hyvinvoinnin tasoon kuin nykyisellä tulojaolla. Atkinsonin indeksin arvo riippuu siis yhteiskunnan koetusta hyvinvoinnista eri tulotasoilla, eli riskiaversion parametrit yritetään löytää kansalaisten omista mieltymyksistä sen sijaan, että lähdettäisiin arvailemaan tai olettamaan mitään mielivaltaista. Ryhmiin purettu Atkinsonin (tulo)indeksin laskukaava on seuraavanlainen:

$$A_\varepsilon = 1 - \frac{1}{N}\left[\sum_{i=1}^{N}\left(\frac{y(a_i)}{\overline{y}}\right)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}},$$

missä i = 1,...,N ja N on jälleen populaation koko ja ε on suhteellisen epätasa-arvon aversion kerroin eli se suhteellinen haitta, joka tuloepätasa-arvolle asetetaan. $y(a_i)$ on tulonsaajan a_i tulot ja $\overline{y}$ on tulojen keskiarvo. (Cowell 1977, Sala-i-Martin 2002.)

Normatiivisten käsitysten sisällyttäminen tuloeroindeksiin aiheuttaa kuitenkin ongelmia. Atkinsonin indeksin lähestyminen esimerkiksi olettaa, että kaikilla yksilöillä on samanlaiset mieltymykset suhteessa tulojen epätasa-arvoon. Toinen ongelma on se, että kun lähdetään vertailemaan eri yksilöiden hyvinvointia, huomioon pitäisi ottaa muitakin muuttujia kuin tulot. Kysymys siis muuttuu moniulotteiseksi. Koko yhteiskunnan hyvinvoinnin palauttaminen tulovektorin funktiona saaduksi skalaariksi ei vain ole riittävää. Esimerkiksi varallisuus on keskeinen muuttuja hyvinvointia määriteltäessä.

Eräs keskeinen ongelma liittyy myös siihen vahvaan oletukseen, että tulojakauman tasaaminen maksimoisi hyvinvointia joka tilanteessa. Kanbur (2003) muistuttaa, että kehitysmaiden tapauksessa köyhien kansalaisten kuollessa nälkään tuloerot pienenevät. Atkinsonin indeksillä mitattuna tämä olisi hyvää kehitystä kaikilla ϵ-parametrin arvoilla.

Lisäksi ongelmia tuottaa, että jos annetaan jotakin preskriptiivistä merkitystä tulovektorin skalaarina saadulle tuloeroja kuvaavalle funktiolle, on erittäin suurta väliä, mikä otetaan siksi joukoksi, jonka tulovektorille näitä funktioita lasketaan. Onko relevanttia mitata maailman, Euroopan, Suomen, Uudenmaan läänin, Helsingin, Kallion tai vaikka Pengerkadun tuloeroja. Kaikki nämä valinnat antaisivat hyvin erilaisia lukuja. Harvoin esimerkiksi suomalaisessa keskustelussa tulojen tasa-arvosta tarkoitetaan tasa-arvoa suomalaisten ja kehitysmaiden asukkaiden välillä.

Tässä työssä ei tehdä empiirisiä laskelmia Atkinsonin indeksistä, koska sen informaatioarvo on kyseenalainen.

6 Tulojakauman prediktiivinen merkitys

Teoreettisesti keskustelu tulojakauman prediktiivisestä merkityksestä eli tulojakauman merkityksestä muihin reaalitaloudellisiin suureisiin liittyy yleisemmin keskusteluun talousteoriassa usein tehdyistä oletuksista ja abstraktioista. Esimerkiksi Pareto-optimaalisuuden käsite on kehitetty Amartya Senin (1973, 6–7) sanoin "juuri jakaumakysymysten poistamiseksi". Kaikki mahdolliset tavat jakaa kakku – kunhan koko kakku tulee jaettua – ovat Pareto-optimaalisia, sillä kennellekään ei voida antaa lisää ottamatta pois joltakin toiselta. Edustavan agentin käsite liittyy samaan keskusteluun. Suuri osa dynaamisesta yleisen tasapainon makrotalousteoriasta perustuu olettamukselle, että talous käyttäytyy, kuin olisi olemassa vain yksi tai yhdenlaisia kuluttajia. Tämä aiheuttaa ongelmia, sillä samaan aikaan talousteoriassa yritetään ottaa makroteorian mikroperusteet vakavasti. (Krusell – Smith 1998.)

Jakaumakysymykset ja heterogeenisyys tuovat lisää monimutkaisuutta talousteoriaan, eivätkä ne ole vielä löytäneet tietään vakiintuneeseen kansantaloustieteelliseen diskurssiin ja sitä kautta esimerkiksi oppikirjallisuuden piiriin. Kolmesta satunnaisesta perusoppikirjasta (Varian 2003, Burda – Wyplosz 2001 ja Sørensen – Whitta-Jacobsen 2005) etsittäessä sanaa heterogeenisyys ei löydy hakemistosta lainkaan. Sana jakauma on mainittu kerran kahdessa kirjoista. Tuloerotutkimus on kuin oma erillinen tieteenalansa, koska sitä ei ole osattu yhdistää muuhun teoriakehikkoon. Tämän luvun tarkoitus on kyetä hahmottamaan sitä merkitystä, joka tulojakaumalla, sen ominaisuuksilla ja niiden ominaisuuksien oikeanlaisella karakterisoinnilla voisi olla talouteeen ja taloustieteeseen.

6.1 Aiempaa kirjallisuutta

Ensimmäisiä taloustieteilijöitä, jotka ovat käsitelleet tulojakauman merkitystä muiden taloudellisten muuttujien selittäjänä, on Antoine Augustin Cournot, joka totesi jo vuonna 1863, että kysynnän laki “riippuu olennaisesti väestön koosta, vaurauden jakautumisesta, kuluttajien varallisuudesta, mauista ja tottumuksista, markkina-alueiden määrästä ja kuljetuskustannusten alenemisesta johtuvasta markkinoiden leviämisestä” (Cournot 1863, 100).

Vilfredo Pareton mukaan taas “tulojakauma on yksi keskeisiä asioita, joka tulee tietää, jotta voidaan ymmärtää kysynnän ja tarjonnan lakeja”. Sittemmin nämä kysymykset on usein sivuutettu. Jopa Pareto itse totesi myöhemmin, että kysymys on turha, sillä hänen mielestään tulojakauma oli ajasta ja paikasta riippumaton. Staehle ilmoittaa kuitenkin löytäneensä tulojakaumasta lasketun tuloeromittarin ja kysynnän välille yhteyden. Staehle sai seuraavanlaisen tuloksen:

$$X_1 = 197,4 - 0,33X_2 - 2,31X_3,$$

missä X_1 on vähittäiskaupan myynti-indeksi jaettuna työtulojen indeksillä, X_2 on työtulojen indeksi jaettuna palkkojen yleisellä indeksillä ja X_3 on hänen käyttämänsä tuloeromittari. Menemättä syvemmin hänen tuloksiinsa ja metodeihinsa voidaan kuitenkin todeta, että tilastollisesti merkitseviä tuloksia on saatu jo aikaisessa vaiheessa empiria sta. (Staehle 1937.)

Myöhempiä tutkimuksia aiheesta on esimerkiksi Musgroven (1980) tutkimus, jossa hän tutkii tulojakauman vaikutusta keskimääräiseen kulutusalttiuteen. Hän toteaa että 30 maan aineistolla ei löydy yhteyttä tulojen tasa-arvon ja keskikulutusalttiuden välillä. Hän myös toteaa, että jos jotakin, niin näiden muuttujien välillä on aiemmin löydetty ainoastaan negatiivisia yhteyksiä: kun tasa-arvo on lisääntynyt, keskikulutusalttius on vain pienentynyt.

Steckel (1995) löytää kansainvälisellä aineistolla negatiivisen yhteyden pituuden ja gini-kertoimen välillä. Per capita tulojen logaritmilla ei hänen tutkimuksensa valossa ole tilastollisesti merkitsevää merkitystä pituuskasvuun. Tässä tapauksessa siis tulojen keskiarvolla ei tuntuisi olevan merkitystä pituuskasvuun. Selittävä tekijä onkin itse asiassa tulojen dispersio. Vaikka tämä ei olekaan suoranaisesti taloustieteellinen tutkimus, niin se osoittaa hyvin, että tulojakauman muodolla saattaa olla aitoa merkitystä erittäin konkreettisiinkin muuttujiin.

Toinen yhtä konkreettinen muuttuja, jonka kanssa Robert Waldmann (1992) löytää yhteyden tulojen dispersioon, on lapsikuolleisuus. Hänen tulostensa mukaan globaalilla aineistolla lapsikuolleisuus on sitä alhaisempi, mitä pienempi on tulojen dispersio.

Alesina – Rodrik (1994) etsivät maaomistusten ja tulojen epätasa-arvon prediktiivistä roolia talouskasvuun. He siis tutkivat sitä, kuinka tulojen dispersio vaikuttaa niiden keskiarvon muutokseen. Alesina – Rodrik löytävät käyttämällään kansainvälisellä aineistolla negatiivisen yhteyden näiden muuttujien väillä.

Tuloerojen dynaamisista vaikutuksista talouskasvuun kirjoittavat Cornia – Addison – Kiiski (2003), että erittäin suurten suurten ja erittäin pienten tuloerojen maissa syntyy kannustinongelmia. Optimi talouskasvun kannalta on heidän kansainvälisellä aineistolla tekemien empiiristen tutkimuksen mukaan jossakin ääripäiden välissä

6.2 Edustava agentti

Tulojakaumalle on monessa yhteydessä onnistuttu antamaan ainakin tilastollisesti merkitsevä rooli muiden muuttujien selittäjänä. Yleisempää keskustelua heterogeenisyyden merkityksestä käydään kuitenkin yhä. Edustavan agentin oletuksen järkevyydestä kirjoitetaan edelleen (mm. Carroll 2000, Carroll – Kimball 1996, Attanasio – Davis 1996 ja Hartley 1996.), eikä oletuksen järkevyydestä ole päästy yhteisymmärrykseen.

James Hartley (1996) kirjoittaa edustavan agentin historiasta, että se on 1800-luvun lopulla Alfred Marshallin kehittämä metodi tutkia teollisuuden tarjontakäyrää. Edustavan yrityksen abstraktio kohtasi vahvaa vastustusta heti, kun se esiteltiin. Eräs kritiikki liittyi siihen, että teollisuuden kasvaessa edustava yritys ei enää edusta teollisuutta, koska teollisuudenalalla skaalan kasvaessa työnjako lisääntyy ja edustava yritys ei enää olekaan edustava, koska koko teollisuudenalan luonne on muuttunut.

Jos oletetaan, että teollisuuden sisällä ei tapahdu erikoistumista, ja että kaikki yritykset ovat identtisiä, alan kasvu voi johtua joko jo alalla olevien yritysten tuotannon kasvusta tai alalle tulevista uusista yrityksistä. Jos kasvu johtuu ensimmäisestä, edustava yritys kasvaa. Jos kasvu johtuu jälkimmäisestä, edustava yritys ei kasva. Ilman tietoa kasvun tavasta edustavan yrityksen dynamiikkaa ei voi ennustaa. Kovan kritiikin jälkeen ajatus edustavasta agentista hylättiin joksikin aikaa. Sittemmin se on tullut takaisin ja on vakiinnuttanut paikkansa talousteoriassa.

Tutkimuksessaan, joka kritisoi edustavan agentin perusteluja, Alan Kirman (1992) mainitsee, että esimerkiksi työttömyyttä ei voi järkevästi mallintaa edustavan agentin toiminnalla, koska kyseessä on erityisesti koordinaatio-ongelma. Tällaisen ongelman mallintamiseen tarvitaan monta toimijaa. Toinen hänen nostamansa ongelma on, että edustavan agentin malleissa ei usein tapahdu mitään vaihtokauppaa, vaikka kyseessä ovat taloudelliset mallit. “Sellaisessa maailmassa ei olisi merkityksekkäitä osakemarkkinoita, jakaumakysymykset eivät kuuluisi hallituksen poliittisiin huolen kohteisiin, eikä asymmetrisen informaation käsitteessä olisi juurikaan järkeä”, hän kirjoittaa.

Kirman kiteyttää ongelman neljään, osittain päällekkäiseen, kohtaan. Ensinnäkin hän toteaa, että vaikka yksilöiden aggregaatti toimisikin optimaalisesti, sama ei välttämättä päde yksilöihin itseensä. Jos siis edustava yksilö toimii rationaalisesti, sen muodostavat yksittäiset yksilöt eivät valttämättä toimi rationaalisesti. Sama argumentti pätee myös toisin päin.

Toisaalta aggregaattiagentti ei välttämättä reagoi samalla tavalla jonkin esimerkiksi hallituksen politiikan määrittämän parametrin muutoksiin kuin yksilöidensä summa. Siis vaikka aggregoitu agentti edustaisikin hyvin kaikkia yksilöitä ennen parametrin muutosta, näin ei välttämättä ole muutoksen jälkeen. Tämä kohta on samankaltainen kuin aiemmin mainittu Marshallin edustavaa yritystä kohtaan nostettu kritiikki.

Kolmantena kohtana hän nostaa esille konstruoidun esimerkin, jossa kukin yksittäinen agentti preferoi yhtä asiaa, mutta aggregaatti preferoi toista. Toinen ja kolmas kohta kyseenalaistavat aggregoidun yksilön pohjalta tehdyt politiikkasuositukset.

Neljäs Kirmanin esille nostama kohta kohta on, että johtuen aggregaattiagentin taustalla olevasta monimutkaisesta yksilöiden dynamiikasta aggregaattiagentti saattaa saada melko luonnottomia ominaisuuksia. Kaikki näihin ominaisuuksiin liittyviin hypoteeseihin liittyvät päättelyt voivat johtua joko oikeasta päättelystä tai päättelystä, joka johtuu niistä lisähypoteeseista, joita on tehtävä perustellakseen tasapainoanalyysit ja Kirmanin sanoin "pseudomikroperusteet". Edustavan agentin malleja testatessa siis testataan aina yhdistettyä hypoteesia. Yhtäältä testataan mielenkiinnon kohteena olevaa käyttäytymiseen liittyvää hypoteesia. Toisaalta samaan aikaan testataan sitä hypoteesia, että aggregaatin toimintaa voidaan kuvata yhden ainoan edustavan agentin toimilla. Yhdistetyn hypoteesin tullessa hylätyksi emme tiedä, hylätäänkö itse hypoteesi vai edustavan agentin edustavuuden hypoteesi. Tämä päättely pätee ylipäätänsä väärin spesifioidun mallin ongelmiin.

Kirman jatkaa vielä esittäen, että tavallisessa edustavan agentin aggregointitilanteessa oletetaan, että yksilöillä on homoteettiset[14] kysyntäfunktiot ja että suhteellinen tulojakauma on vakio ja riippumaton hintatasosta. Tulojakauman vakioisuuden oletus sopii yhteen Pareton jo sata vuotta sitten tekemän havainnon kanssa. Empiria kuitenkin tuntuisi olevan ristiriidassa tämän oletuksen kanssa. Tässäkin työssä on osoitettu[15], että esimerkiksi Suomessa viimeisten vuosien aikana tulojakauman muutokset ovat olleet suhteellisen radikaaleja. Tulojakauma on jopa menettänyt varianssinsa. Nämä oletukset tulojakaumasta ovat siis melko vahvoja, eivätkä ne tule täytetyksi juuri millään empiirisellä aineistolla.

[14]Homoteettinen tarkoittaa sitä, että kuluttajia valitsee kulutuskorinsa hyödykkeiden suhteellisen määrän mukaan. Positiiviset lineaariset muunnokset hyödykekoreista eivät muuta niiden paremmuusjärjestystä.

[15]Muun muassa kuvat 3.1, 3.13, 4.3 ja 4.6.

Kirman ehdottaa edustavan agentin abstraktion tilalle heterogeenisyyden hyväksymistä ja luettelee monta esimerkkiä ja tutkimusta, joissa on saatu tuloksena, että heterogeenisyys itse asiassa tuo stabiilisuutta moneen prosessiin ja on jopa edellytyksenä tasapainotilojen saavuttamiseen. Lisäksi Kirman suosittelee peliteoreettisen kehikon suosimista, koska siinä monien agenttien interaktio on lähtökohtana.

Kuten James Hartley (1996) tutkimuksessaan toteaa, joudutaan edustavan agentin tapauksessa pohtimaan, onko kyseessä vain keskiarvo kaikista agenteista vai kenties jotakin muuta. Tällaisessa tapauksessa joudutaan tietenkin pohtimaan, millä tavoin tällaista aineistoa tulisi aggregoida. Jos tulojakaumasta yrittää etsiä yhteen lukuun palautettavaa edustavaa kuluttajaa, jonka perusteella yleisen tasapainon analyysit ja niistä johdetut politiikkasuositukset tehdään, joudutaan esimerkiksi miettimään, otetaanko populaatiota edustavaksi tunnusluvuksi moodi, mediaani vai keskiarvo.

Tulojakauman tapauksessa moodi $<$ mediaani $<$ keskiarvo[16], koska jakauma on oikealle vino. Keskiarvo on luultavasti paras tunnusluku ennustamaan koko populaation toimintaa. Ennustamiseen siis kannattaisi luultavasti käyttää keskiarvoa. Toisaalta moodi on se yksittäinen tunnusluku, joka edustaa suurinta osajoukkoa populaatiosta. Politiikkasuositukset, jotka parantavat moodin hyvinvointia, voisivat siis tuottaa tehokkaamman tuloksen.

6.3 Epälineaarisuuksien merkitys

Aggregointikysymyksissä Hartley (1996) kirjoittaa vielä, että “kun edustavia agentteja käytetään makrotaloudellisissa malleissa, tehdään toinen kahdesta yhtä kestämättömästä oletuksesta. Joko kaiken oletetaan olevan lineaarista, tai toimijoiden määrää rajoitetaan eksogeenisesti.” Lineaarisuuden oletuksen merkitystä voidaan tutkia tarkemmin tulonsiirtojen ja kulutuksen kehikossa.

[16] Moodi voi itse asiassa olla lähes mitä vain johtuen suhteellisen pienistä otoksista. Jos kuitenkin lasketaan moodi joistakin tarpeeksi suurista vakiokokoisista väleistä, tämä pitää paikkansa.

6.3.1 Kulutus

Olkoon kaikilla tulonsaajilla kulutusfunktio muotoa[17]

$$C = g(y) = \alpha_c + \beta_c y + \gamma_c log\,(1 + \delta_c y).$$

Rajakulutusalttius on

$$C_y = g'(y) = \beta_c + \frac{\gamma_c \delta_c}{1 + \delta_c y},\ g'(0) = \beta_c + \gamma_c \delta_c,\ \lim_{y \to \infty} g'(y) = \beta_c.$$

Jos asetetaan parametrien arvoksi $(\alpha_c; \beta_c; \gamma_c; \delta_c) = (0; 0,2; 8000; 0,0002)$, saadaan rajakulutusasteiksi

$$g'(0) = 0,2 + 8000 \cdot 0,0002 = 1,\ \lim_{y \to \infty} g'(y) = 0,2.$$

Tämä kulutusfunktion muoto siis takaa sen, että kenenkään rajakulutusalttius eli kulutus saaduista lisätuloista ei voi kohota yli yhden. Lisäksi aivan suurituloisimpien tulonsaajien rajakulutusalttius ei muodostu liian pieneksi. Nämä luvut approksimoivat rajakulutusalttiuksia ensimmäisenä vuonna saadusta lisätulosta.

Tämän muotoinen kulutusfunktio täyttää kvalitatiivisesti kulutusfunktioilta yleensä vaaditut ominaisuudet. Rajakulutusalttiuden ei pitäisi voida nousta yli yhden ja kulutuksen tulisi olla konkaavi funktio tuloista. Logaritmifunktion tuominen yhtälöön mahdollistaa kulutusfunktion määrittämisen sopivalla tavalla kasvavaksi ja konkaaviksi.

Olkoon nyt y_{ij} satunnainen havainto teoreettisesta käytettävissä olevien tulojen jakaumasta

$$Y_{ij} \sim S(\alpha; \beta; \gamma_j; \delta),$$

missä $(\alpha; \beta; \delta)$ arvoina ovat vakiot $(1,6; 1; 18525,2)$ ja γ-parametri vaihtelee seuraavasti: $\gamma_j \in \{3216,05; 3516,05; \ldots; 5916,05\}$, $j \in \{1, 2, \ldots, 10\}$ Jokaisella γ:n arvolla γ_j generoidaan kaksi miljoonaa havaintoa yllä esitetyistä jakaumista eli

[17]Tämä Yrjö Vartian ehdottama funktiomuoto on konkaavi funktio tuloista. Konkaavisuus on kulutusfunktion keskeinen ominaisuus. Lisäksi kyseinen funktiomuoto mahdollistaa erittäin tarkan hienosäädon funktion muodolle δ_c-parametrin avulla. Kuluttajien homogeenisyys kulutusfunktioiden suhteen ei ole realistinen oletus. Tämän oletuksen löyhentämisen tuottamien vaikutusten tutkiminen jätetään kuitenkin myöhemmille tutkimuksille.

$i \in \{1, 2, \ldots, 2000000\}$. Saamme siis kymmenen tulovektoria, jotka kaikki ovat kahden miljoonan havainnon kokoisia otoksia stabiileista jakaumista. Jakauman karakteristisen funktion funktiomuoto ja kolme muuta parametria paitsi γ-parametri pysyvät vakioina.

γ-parametri on, kuten luvussa kolme kerrottiin, stabiilien jakaumien skaalaparametri tai dispersioparametri, eli se ei vaikuta tulojakauman keskiarvoon. Normaalijakaumalla, eli stabiililla jakaumalla kun α-parametri saa arvon 2, on sellainen ominaisuus, että $\gamma = \frac{\sigma}{\sqrt{2}}$, eli varianssi on johdettavissa γ:sta. Vastaavalla tavalla muilla α:n kiinnitetyillä arvoilla γ määrittää jakauman dispersio-ominaisuudet.

Kun annetuilla kulutusfunktiolla lasketaan keskikulutus näillä kaikilla kymmenellä satunnaisotannalla saadulla tulovektorilla[18], kuvasta 6.1 havaitaan, että dispersion kasvattaminen johtaa tässä tapauksessa pienenevään keskikulutukseen.

Tämä tekninen tulos johtuu kulutusfunktion konkaavisuudesta. Jensenin epäyhtälö toteaa, että kun f on konkaavi funktio ja x_i kuuluu funktion f lähtöjoukkoon, missä $i \in \{1, 2, \ldots, n\}$, niin

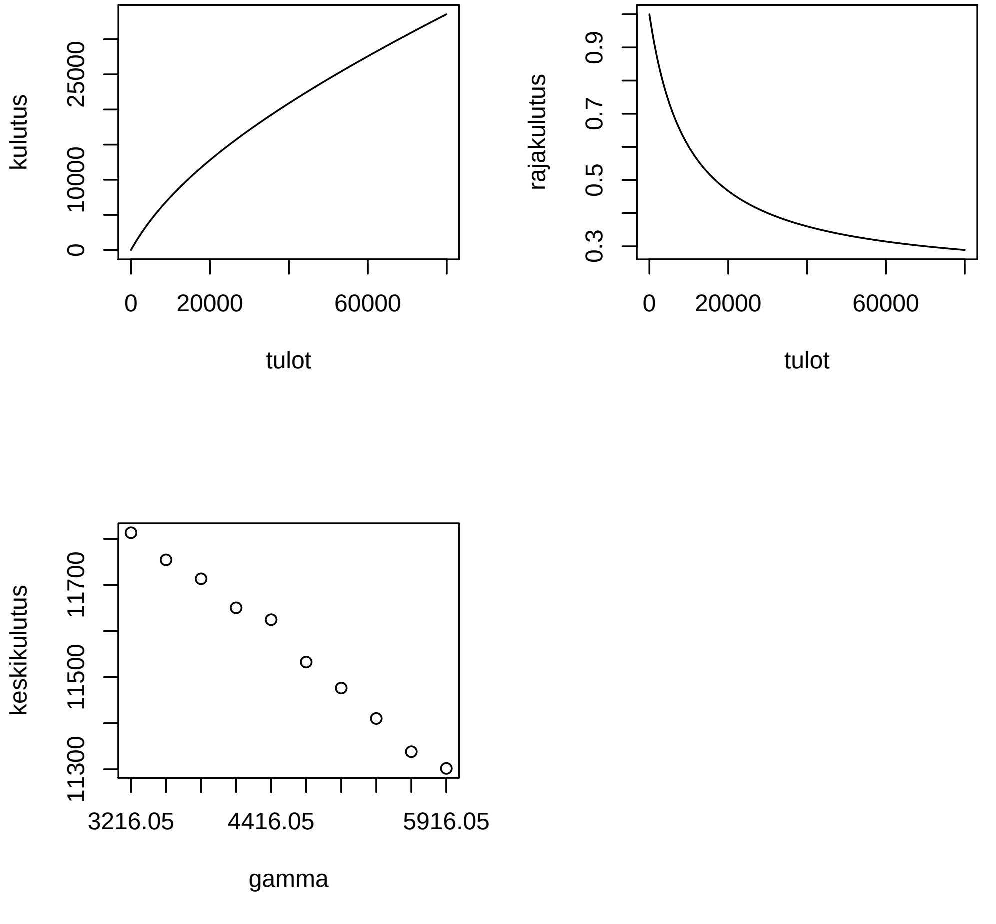
$$f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \geq \frac{1}{n}\sum_{i=1}^{n} f(x_i).$$

Eli konkaavi funktio muuttujien keskiarvosta saa aina vähintään yhtä suuren arvon kuin muuttujien funktioiden arvojen keskiarvo. Yhtäsuuruus pitää paikkansa ainoastaan siinä tapauksessa, että kaikki muuttujat ovat yhtä suuria. Epäyhtälön oikea puoli siis maksimoituu, kun dispersio on minimaalinen eli nolla. Mitä suuremmaksi dispersio kasvaa, sitä pienemmäksi epäyhtälön oikea puoli muuttuu. Konvekseille funktioille pätee päinvastainen tulos.

Tästä syystä keskikulutus kasvaa γ-parametria pienennettäessä. Jos kulutusfunktio olisi lineaarinen, jakauman dispersiolla ei olisi mitään merkitystä keskikulutukseen. Se olisi ainoastaan keskiarvon funktio. Esimerkiksi tässä simulaatiossa käytetyillä tulovektoreilla samalla keskiarvolla, mutta eri tulovektorin dispersioilla keskikulutus on 4,7 % suurempi γ:n arvolla 3216,05, kuin γ:n ar-

[18] Kyseinen jakauma γ-parametrin arvolla 5016,05 on siis paras sovite kotitalouksien käytettävissä oleville tuloille ekvivalenssiskaalalla laskettua kulutusyksikköä kohden vuodelle 2004. Simulaatiossa käytetyt γ:n arvot ovat suunnilleen samalta vaihteluväliltä, kuin Suomen aineistoon vuosille 1990–2004 tehtyjen sovitusten γ-parametrit. Tässä ei siis tehdä arvioita todellisesta Suomen kulutuksesta, vaan enemmänkin kyseessä on havainnollistava esimerkki, jossa käytetään hyväksi aiemmin hankittua ymmärrystä tulojakaumien luonteesta. Tulokset ovat luonteeltaan kvalitatiivisia.

Kuva 6.1: Simulaatiossa käytetyn kulutusfunktion ja sen derivaatan visualisaatio ja keskikulutuksen määriä eri γ-parametreilla simuloiduilla tulovektoreilla.



volla 5916,05. Vaikka kyseessä on vain kvalitatiivinen ja valitusta kulutusfunktiosta riippuvainen tulos, niin vaikutuksen suunta on teorian mukainen. Koska viime vuosina jakauman dispersio on kasvanut talouskasvun kanssa, on kulutusfunktion konkaavisuus johtanut siihen, että keskikulutus on luultavasti kasvanut hitaammin kuin tulojen keskiarvon kasvu on antanut olettaa.

Seuraavissa luvuissa käsitellään vastaavaa yhteyttä tulojen ja tulonsiirtojen välillä.

6.3.2 Maksetut tulonsiirrot

Määritellään perustulo $y_p = y + T_n$, missä y kotitalouden käytettävissä olevat tulot ja T_n on kotitalouden maksamat nettotulonsiirrot ja $T_n = T - T_r$, missä T on maksetut tulonsiirrot ja T_r on saadut tulonsiirrot. Maksetut ja saadut tulonsiirrot liittyvät perustuloon tavalla tai toisella. Oletuksena voidaan pitää ainakin, että keskimäärin perustulojen kasvaessa myös maksetut tulonsiirrot kasvavat ja saadut pienenevät. Aineistossa kahdeksalla kotitaloudella perustuloiksi tuli alle nolla. Ne jätetään huomiotta seuraavissa analyyseissä.

Aineistosta estimoidut yhteydet perustulon ja maksettujen ja saatujen tulonsiirtojen välillä vaikuttavat mutkikkailta ja sisältävät itse asiassa konvekseja, konkaaveja ja affiineja[19] osia ainakin kuvan 6.2 keskiarvopisteiden perusteella. Tällainen funktiohan ei ole silloin koko lähtöjoukossaan konkaavi eikä konveksi. Maksettujen tulonsiirtojen yhteyttä perustuloon estimoitaessa estimointi aloitetaan 5400 eurosta ylöspäin, koska tilannetta tätä ennen voidaan kuvata melko hyvin affiinilla kasvavalla funktiolla. Kuvan 6.2 oikean ylänurkan ruudun keskiarvopisteitä seuraten huomataan, että 5400 euroa ylittävältä osalta maksettujen tulonsiirtojen keskiarvopisteet kasvavat kiihtyvällä eli konveksilla tavalla suhteessa perustuloihin. Toisaalta aivan jakauman loppupäässä kasvu tuntuisi muuttuvan konkaaviksi.

Tätä yhteyttä sovitetaan kulutusfunktion estimoinnissa käytettyjen funktiomuotojen avulla johtuen siitä, että niitä on äärimmäisen helppoa muotoilla reagoimaan eri argumenttiensa alueilla. Konveksi osa vastaa kulutusfunktion logaritmisen osan derivaattaa. Konkaavi osa taas saadaan kulusfunktion logaritmisesta osasta. Sovitettu lineaarinen malli tulee muotoon:

[19] Affiini funktio on samaan aikaan sekä heikosti konveksi että konkaavi.

$$T = \beta_0 + \beta_1 y_p + \beta_2 \frac{\delta_1}{1+\delta_1 y_p} + \beta_3 log\,(1+\delta_2 y_p) + \epsilon.$$

Parametrit $(\delta_1; \delta_2) = (0,0001; 0,00001)$ on määritetty kokeilemalla. Tällöin β-parametrit voidaan estimoida lineaarisella regressiolla[20]. Tehdyn sovituksen parametrien arvot näkyvät taulukosta 6.1. δ-parametrit on valikoitu siten, että konveksi osa vaikuttaa pienillä argumenttinsa arvoilla ja konkaavi osa suurilla argumenttinsa arvoilla. ϵ-parametri on mallin virhetermi tai epäsystemaattinen termi, joka sisältää kaiken vaihtelun, jota mallin selittävät muuttujat eli systemaattinen osa ei kykene selittämään. Mallin huomattavan korkea selitysaste $R^2 = 95,3$ % kertoo sen, että estimoitu kuvaus perustuloilta maksettuihin tulonsiirtoihin on kohtalaisen tarkka. Tätä epälineaarista riippuvuutta T:n ja y_p:n välillä ei ole syytä enää tarkentaa lisätermein.

Taulukko 6.1: Maksettuja tulonsiirtoja perustuloilla selittävän regressiomallin kertoimien estimaatit. Kaikki estimaatit ovat tilastollisesti merkitseviä normaaleilla riskitasoilla. $R^2 = 0,9532$.

kerroin	estimaatti	keskivirhe	t-arvo
vakio	-9775	159,7	-61,22
y_p	0,282	0,00115	245,63
$\frac{\delta_1}{1+\delta_1 y_p}$	$1,602 \cdot 10^8$	$2,919 \cdot 10^6$	54,87
$log\,(1+\delta_2 y_p)$	22180	415,5	53,37

Funktion epälineaarisen osan konkaavisuutta ja konveksisuutta voidaan tarkastella tutkimalla sen toista derivaattaa. Se saa muodon

$$f''(y_p) = -\beta_2 \delta_1^2 \frac{1}{(1+\delta_1 x)^2} + 2\beta_3 \delta_2^3 \frac{1}{(1+\delta_2 x)^3}.$$

Kokeilemalla voidaan havaita, että tämä funktio nollaantuu, kun x on hieman yli 50 000. Sitä ennen se saa positiivisia arvoja ja sen jälkeen negatiivisia arvoja. Siinä pisteessä funktio siis muuttuu konveksista konkaaviksi.

Maksetut tulonsiirrot ovat siis seuraavanlainen jatkuva funktio $f : \mathbb{R}_+ \to \mathbb{R}_+$ perustulosta:

$$\hat{T} = f(y_p)$$

$$T = \hat{T} + e(a_i) = f(y_p) + e(a_i) =$$

$$\begin{cases} 1318,09316 + 0,37 y_p + e(a_i), & y_p \leq 5400 \\ -9775 + 0,282 y_p + 1,602 \cdot \frac{10^4}{1+\frac{y_p}{10^4}} + 22180 log\,(1+\frac{y_p}{10^5}) + e(a_i), & y_p \geq 5400 \end{cases}$$

[20] Lineaarinen regressio on parametriensa suhteen lineaarinen, vaikka muuttujille on tehty epälineaarisia muuttujamuunnoksia.

$$= \begin{cases} 1318,09316 + 0,37 y_p + e(a_i), & y_p \leq 5400 \\ -9775 + 0,282 y_p + \frac{16020}{1+\frac{y_p}{10^4}} + 22180 log\,(1 + \frac{y_p}{10^5}) + e(a_i), & y_p \geq 5400, \end{cases}$$

missä a_i on tuloja saava kotitalous ja $e(a_i)$ kuvaa mallissa jokaisen kotitalouden stokastista elementtiä ja $E[e(a_i)] = 0$. Maksettuja tulonsiirtoja siis mallinnetaan tässä ikään kuin ne olisivat deterministinen funktio perustuloista, jonka päälle tulee stokastinen elementti, joka mahdollista aineistossa olevan hajonnan. Todellisuudessa maksetut tulonsiirrot ovat vielä deterministisempi funktio, mutta muut selittävät tekijät kuin perustulot on tässä yhteydessä jätetty pois ja niitä mallinnetaan stokastisella komponentilla.

6.3.3 Saadut tulonsiirrot

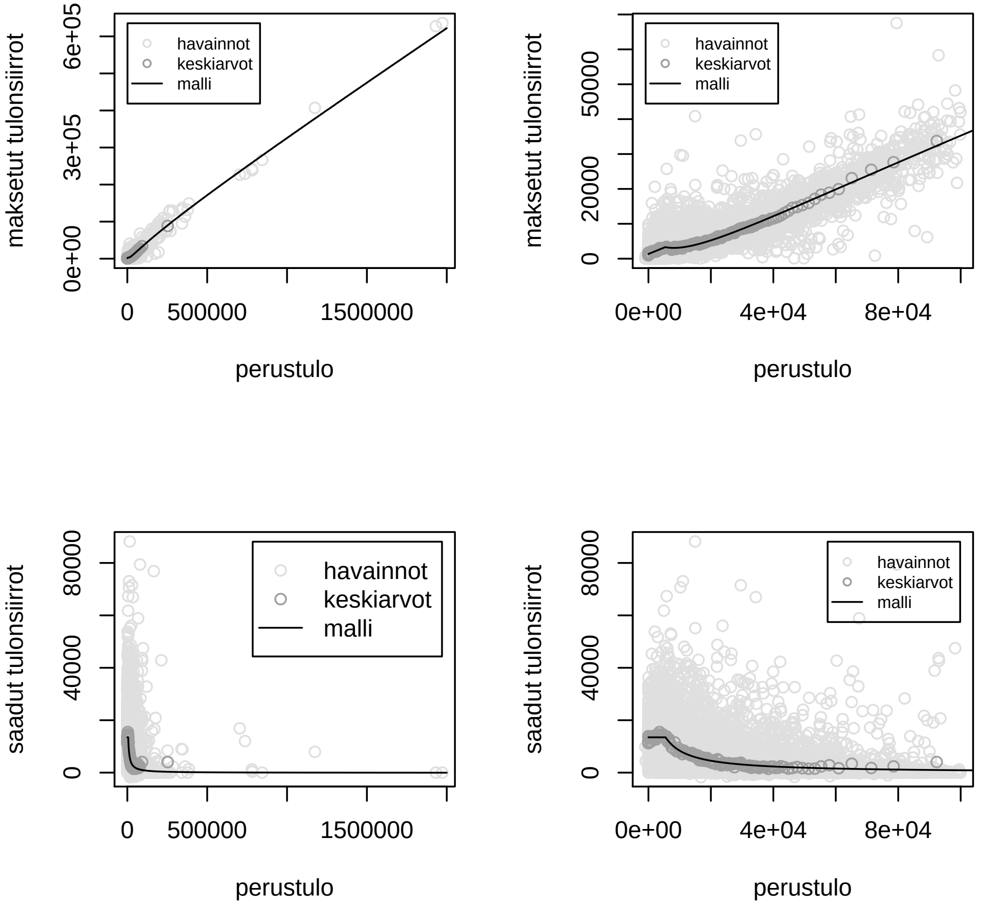
Kotitalouden saamia tulonsiirtoja T_r estimoitaessa kuvasta 6.2 nähdään, että aluksi yhteys tuntuisi olevan affiini. Tämän jälkeen se muuttuu samansuuntaiseksi kuin rajakulutusalttiuksia illustroiva kuvaaja kuvassa 6.1 eli konveksiksi. Estimaatissa oletetaan jälleen affiini alku aina 5500 euron perustuloihin asti. Tämän yli meneviin tuloihin sovitettiin lineaarisella regressiolla funktio

$$T_r = \beta_0 + \beta 1 \frac{\delta_r}{1 + \delta_r y_p} + \epsilon,$$

missä $\delta_r = 0,0005$. Tämän funktion ominaisuuksia on, että se lähenee vakiota β_0 y_p:n kasvaessa rajatta. Tämä on toivottu ominaisuus saatuja tulonsiirtoja kuvaavalle funktiolle, jos β_0 on pieni. Tehdyn sovituksen parametriestimaatit näkyvät taulukosta 6.2.

Mallista saatu kohtuullisen matala alle 0,2:n selitysaste R^2 tarkoittaa sitä, että saatuja tulonsiirtoja selitettäessä esimerkiksi kotitalouden koko tai lasten määrä ovat melko merkittäviä muuttujia, jotka tässä yhteydessä tulevat mallinnettua stokastiikalla. Tehdyissä simulaatioissa näitä muuttujia ei kuitenkaan olisi käytetty hyväksi, joten tämä lähestyminen kelpaa käsillä olevaan tarkoitukseen. Perustulojen ja saatujen tulonsiirtojen välinen keskimääräinen yhteys tulee kuitenkin melko hyvin mallinnettua. Jatkossa kuitenkin tähänkin mallinnukseen ilmeinen laajennus olisi sellainen, joka kykenisi tarkemmin kontrolloimaan muut muuttujan pois mallista. Oletuksena tässä työssä on, että tämä laajennus muuttaisi laadullisesti saatuja tuloksia.

Kuva 6.2: Perustulot ja saadut ja maksetut tulonsiirrot estimoituine funktioineen. Keskiarvot on laskettu aina sadan järjestetyn perättäisen havainnon keskiarvoina.



Koska mallin vakio ei ole tilastollisesti merkitsevästi nollasta poikkeava, suurituloiset eivät tämän mallin mukaan saa odotusarvoisesti tilastollisesti merkitsevästi nollasta poikkeavia tulonsiirtoja.

Taulukko 6.2: Saatuja tulonsiirtoja perustuloilla selittävän regressiomallin kertoimien estimaatit. Vakion estimaatti ei ole tilastollisesti merkitsevä. $R^2 = 0,1893$.

kerroin	estimaatti	keskivirhe	t-arvo
vakio	-81,76	70,79	-1,155
$\frac{\delta_1}{1+\delta_1 y_p}$	$1,017 \cdot 10^8$	$1,339 \cdot 10^6$	75,925

Saadut tulonsiirrot ovat yllä olevan analyysin mukaan seuraavanlainen jatkuva funktio $g : \mathbb{R}_+ \to \mathbb{R}_+$ perustulosta:

$$\hat{T}_r = g(y_p)$$

$$T_r = \hat{T}_r + e(a_i) = g(Y_p) + e(a_i) =$$

$$\begin{cases} 13473,89 + e(a_i), & y_p \leq 5500 \\ -81,76 + 1,017 \cdot 10^8 \frac{0,0005}{1+\frac{5y_p}{10^4}} + e(a_i), & y_p \geq 5500 \end{cases}$$

$$= \begin{cases} 13473,89 + e(a_i), & y_p \leq 5500 \\ -81,76 + \frac{50850}{1+\frac{5y_p}{10^4}} + e(a_i), & y_p \geq 5500, \end{cases}$$

missä jälleen kerran $e(a_i)$ kuvaa mallissa jokaisen kotitalouden stokastista elementtiä ja $E[e(a_i)] = 0$.

6.3.4 Nettotulonsiirrot ja mallien simulointi

Yhdistetään yllä olevat kaksi funktiota sillä tavalla, että laskemme kotitalouden maksamien nettotulonsiirtojen odotusarvon

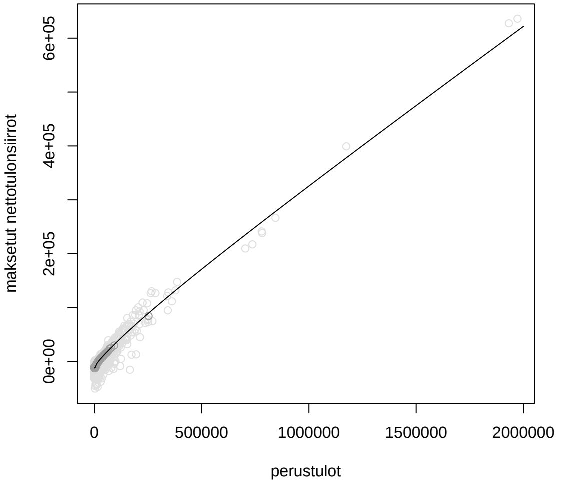
$$E[T_n] = E[T] - E[T_r] = \hat{T} - \hat{T}_r =$$

$$\begin{cases} 1318,09316 + 0,37y_p - 13473,89, & y_p \leq 5400 \\ -9775 + 0,282y_p + \frac{16020}{1+\frac{y_p}{10^4}} \\ +22180 log\,(1 + \frac{y_p}{10^5}) - 13473,89, & 5400 \leq y_p \leq 5500 \\ -9775 + 0,282y_p + \frac{16020}{1+\frac{y_p}{10^4}} \\ +22180 log\,(1 + \frac{y_p}{10^5}) - (-81,76 + \frac{50850}{1+\frac{5y_p}{10^4}}), & y_p \geq 5500, \end{cases}$$

$$
= \begin{cases} -12155,80 + 0,37y_p, & y_p \leq 5400 \\ -23248,89 + 0,282y_p + \frac{16020}{1+\frac{y_p}{10^4}} \\ \quad +22180 log\,(1 + \frac{y_p}{10^5}), & 5400 \leq y_p \leq 5500 \\ -9693.24 + 0,282y_p + \frac{16020}{1+\frac{y_p}{10^4}} \\ +22180 log\,(1 + \frac{y_p}{10^5}) - \frac{50850}{1+\frac{5y_p}{10^4}}, & y_p \geq 5500, \end{cases}
$$

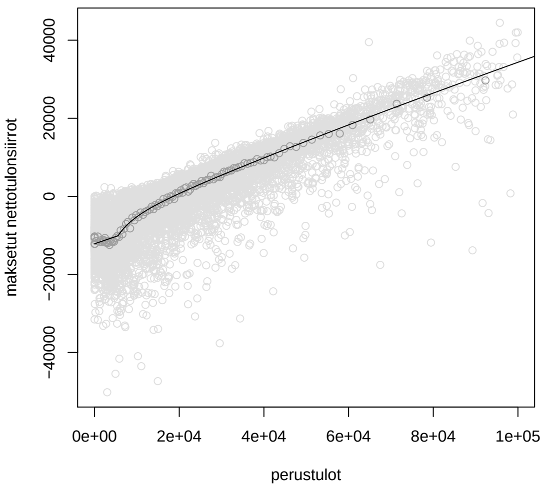
Kuvissa 6.3 ja 6.4 on esitettynä nettotulonsiirrot perustulojen funktiona. Tämä maksettujen tulonsiirtojen ja saatujen tulonsiirtojen erotuksena saatu funktio on muodoltaan jo huomattavan monimutkainen, mutta kuvaa hyvin nettotulonsiirtojen odotusarvoja eri perustulojen pisteissä.

Kuva 6.3: Nettotulonsiirrot $T_n = T - T_r$ y_p:n funktiona.



Tehtäessä perustuloihin perustuvia simulaatioita, on aluksi mallinnettava perustuloja generoivaa prosessia. Jotta vältyttäisiin vastaavanlaiselta analyysiltä, kuin käytettävissä olevien tulojen kohdalla on tehty luvussa 3, simuloidaan käytettävissä olevia tuloja, joiden funktiona lasketaan perustulot. Kuten kuvasta 6.5 nähdään, näiden kahden muuttujan välinen yhteys on lähes täysin affiini. Tästä syystä tehtiin affiini sovite $y_p = \beta_0 + \beta_1 y$, missä y on käytettävissä olevat tulot ja y_p on perustulot. Taulukossa 6.3 on sovitteen parametriestimaatit. Selitysaste $R^2 = 0,936$, joten kuvaus on suhteellisen tarkka. Vaikka kyseessä on vain karkeahko approksimaatio, se riittää käsillä olevaan simulaatioon.

Kuva 6.4: Nettotulonsiirrot $T_n = T - T_r$ y_p:n funktiona siten että $y_p < 100000$.



Taulukko 6.3: Käytettävissä olevien tulojen perustuloja selittävän regressiomallin kertoimien estimaatit. Muuttujien välinen yhteys on estimoitu muuttujien suhteen affiinisti. $R^2 = 0,936$.

kerroin	estimaatti	keskivirhe	t-arvo
vakio	-10660	70,70	-150,8
y	1,510	0,002314	652,5

Kyseessä on siis stabiilin satunnaisjakauman affiini muunnos. Stabiileilla jakaumilla tämä tarkoittaa niiden stabiilisuusominaisuuden johdosta sitä, että myös muunnettu muuttuja noudattaa stabiilia jakaumaa tavalla, joka näkyy luvun 3.2.5 yhtälöstä (4) eli jos $Y \sim S(\alpha, \beta, \gamma, \delta)$, niin kaikille $a \neq 0$, $b \in \mathbb{R}$,

$$aY + b \sim S(\alpha, sign\, a \cdot \beta, \mid a \mid \gamma, a\delta + b).$$

Koska siis $(a; b) = (1,51; -10660)$ ja $(\alpha; \beta; \gamma; \delta) = (1,6; 1, 5016,05; 18525,2)$, niin kotitalouksien käytettävissä olevia tuloja y kuvaavan stabiilia jakaumaanoudattavan satunnaismuuttujan Y affiinilla muunnoksella saatu perustulojen y_p todennäköisyysjakaumaa kuvaavan satunnaismuuttuja

$$Y_p = aY + b = 1,51Y - 10660 \sim S(1,6; 1; 1,51 \cdot 5016,05; 1,51 \cdot 18525,2 - 10660).$$

Siis

$$Y_p \sim S(1,6; 1; 7574,236; 17313,05).$$

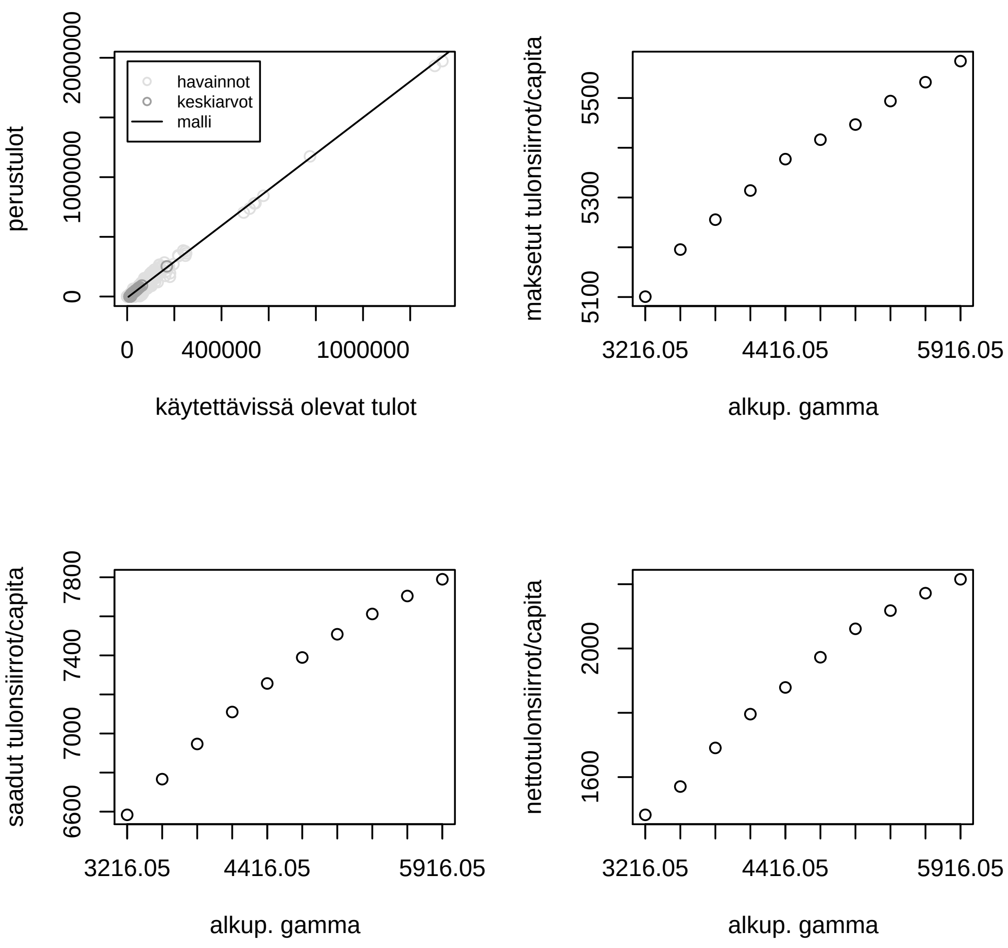
Dispersio on siis perustuloilla suurempaa kuin käytettävissä olevilla tuloilla. Sama muuttujamuunnos on luonnollisesti tehtävä kaikilla kymmenellä simulaatioissa käytetyillä γ-parametrin arvolla. Tässä työssä perustulojen simulaatiot on kuitenkin tehty käytettävissä olevien tulojen simuloitujen tulovektoreiden muunnoksina, koska tällöin on voitu käyttää samoja, jo kertaalleen arvottuja vektoreita. Tulokset eivät olisi muuttuneet, vaikka olisikin arvottu täysin uudet perustulojen satunnaisvektorit.

Simulaatiot siis tehtiin siten, että käytettiin jo edellä käytettyjä käytettävissä olevien tulojen tulovektoreita y. Näille tehtiin muuttujamuunnos, jotta saatiin perustulot $y_p = 1,51y - 10660$. Tulokset on kerätty kuvaan 6.5. Nettotulosiirrot on kuvassa esitetty kotitalouksien saamina nettotulonsiirtoina.

Dispersion kasvaessa sekä kotitalouksien keskimääräiset maksetut että saadut tulonsiirrot lisääntyvät. Tämä johtuu näissä tapauksissa siitä, että annetuilla tulovektoreilla funktioiden konveksit osat dominoivat. Funktiot ovat siis konvekseja merkittävällä alueella, vaikka kumpikaan niistä ei ole konveksi eikä konkaavi koko lähtöjoukossaan. Tulojakauman dispersion kasvaessa siis julkisen sektorin tasoittava rooli kasvaa. Dispersion pienentyessä julkisen sektorin tasoittava rooli pienenee. Yhteisvaikutuksena kotitalouksien saamat nettotulonsiirrot hieman yllättäen vähenevät dispersion pienentyessä. Tässä työssä tehtyjen mallinnusten mukaan kotitalouksien maksamat nettotulonsiirrot ovat itse asiassa suuressa osassa lähtöjoukkoaan konkaavi funktio perustuloista. Kuten aiemmin todettiin, tällaiset funktiot maksimoituvat dispersion pienentyessä. Kotitalouksien maksamat tulonsiirrot siis kasvavat dispersion lisääntymisen funktiona hitaammin, kuin kotitalouksien saamat tulonsiirrot.

Tulokset ovat karkeita siitä syystä, että aluksi on estimoitu teoreettinen käytössä olevien tulojen tulojakauman muoto, jonka jälkeen on estimoitu käytettävissä olevien tulojen ja perustulojen yhteys. Lisää epätarkkuutta analyysiin tuo se, että tulonsiirtojen keskiarvofunktiot ovat vain estimaatteja todellisista. Simulaation suuresta otoskoosta johtuen simuloidut empiiriset jakaumat itsessään ovat kohtuullisen tarkkoja kuvauksia niiden taustalla olevista teoreettisista vastineista, mutta tässäkin kohtaa analyysiin syntyy hieman lisää epävarmuutta. Loppujen lopuksi kuitenkin minkään näistä epävarmuutta aiheuttavista tekijöistä ei pitäisi tuoda niin suurta karkeutta tarkasteluun, että laadullisesti jouduttaisiin hylkäämään jotakin yllä olevasta päättelystä. Kuitenkin jokainen yllä mainituista kohdista vaatii lisätarkasteluja, jotta epätarkkuutta voitaisiin vielä merkittävästi vähentää. Yllä olevassa tarkastelussa merkittävin tulos on, että käsitellyt taloudelliset yhteydet ovat kaukana lineaarisesta ja että itse asiassa yhteydet ovat kovin monimutkaisia. Tarvitaan kehittynyttä teknistä välineistöä, jotta taloustieteen ymmärrystä näistä asioista voidaan syventää.

Kuva 6.5: Perustulojen ja käytettävissä olevien tulojen yhteys ja simulaatioiden tulokset. Nettotulonsiirrot tarkoittaa tässä kuvassa kotitalouksien saamia nettotulonsiirtoja, jotta päästään positiiviselle skaalalle.



7 Päätelmät

Tässä tutkielmassa tarkasteltiin tuloerojen ja tulojakaumien luonnetta. Tutkimusta motivoitiin sekä deskriptiivisen tulojakaumatutkimuksen merkityksellä itsessään, että sen mahdollisilla prediktiivisillä implikaatioilla. Havaittiin, että on merkityksellistä tuntea tulojakauman luonne, jotta voidaan ymmärtää taloudessa olevien epälineaarisuuksien merkitys.

Tutkielman päätuloksena voidaan pitää sitä, että tulojakaumaa generoivaa stokastista mekanismia tutkittaessa havaittiin, että annetulla aineistolla on anomalinen ylähäntä, jota kutsuttiin nimellä superparetohäntä. Niiden tulonsaajien tulot, jotka kuuluvat 99. ja 100. prosenttipisteen väliin, asettuvat vielä uudelle, loivemmalle Pareto-hännälle, joka nostaa ylimmät tulot erittäin suuriksi. Superparetohännän todettiin muodostuneen Suomessa yhdeksänkymmentäluvun lopulla. Tämä ominaisuus lisää tulojakauman mallintamisen vaikeutta. Superparetohännän todettiin myös muodostuneen niin radikaaliksi, että tulojakaumilla ei enää ole varianssia. Jakaumasovituksia tehtäessä havaittiin, että stabiili jakaumaperhe tuottaa hyvän sovituksen samaan aikaan tulojakauman keskivaiheille ja ylähäntään. Erityisesti stabiilien jakaumien kyky mallintaa ylähännän superpareto-ominaisuutta nosti ne esiin suhteessa muihin tehtyihin sovitteisiin erityisesti siksi, että saaduilla stabiileilla sovitteilla ei myöskään ollut varianssia.

Vakiintuneista skalaarisista tuloeromitoista todettiin, että niitä tulee käyttää tarkkaavaisesti kahdesta syystä. Ensinnäkin jokainen näistä mittareista painottaa tulojakauman dispersio-ominaisuuksia eri tavalla, joten tuloeroja tutkittaessa olisi hyvä raportoida myös näiden vaihtoehtoisten mittareiden antamat tulokset, kuten johdannossa todettiin Edward Leamerin suositelleen. Suomessa annetulla aineistolla Theilin indeksiä seuratessa voidaan havaita, että tällä mittarilla tuloerot ovat yli kaksinkertaistuneet kymmenessä vuodessa, kun taas Gini-kertoimella mitattuna tuloerot eivät ole kasvaneet edes puolitoistakertaisiksi. Toinen syy olla tarkkaavainen skalaaristen dispersio-mittojen kanssa on se, että nämä mittarit voivat antaa saman luvun kahden kovin erilaisen tulovektorin funktiona.

Preskriptiivisen lähestymisen tulovektoriin todetaan olevan liian kapea. Pelkän tulojakauman tunteminen ei anna riittävää tietoa asioiden tilasta, jotta voitaisiin arvioida niiden hyvyyttä tai huonoutta. Ainakin varallisuutta tulisi tutkia tulojen lisäksi. Myös ihmisten mieltymykset suhteessa vaurauteen tulisi mielellään tuntea yksilöllisesti, ennen kuin voidaan sanoa, onko jokin oikeudenmukaisuusperiaate toteutunut. Tuloerot voivat johtua siitä, että toiset arvostavat enemmän rahaa kuin toiset ja tekevät siksi enemmän töitä. Toisaalta esimerkiksi voi olla tilanteita, joissa suurta varallisuutta hallinnoivat ihmiset eivät halua ansaita enempää ainakaan töitä tekemällä. Pienen varallisuuden ihmiset taas mahdollisesti haluavat nostaa varallisuuttaan ja siksi ansaita enemmän. Kuten jo aiemmin todettiin, kehitysmaissa pienituloisten ihmisten kuollessa nälkään tuloerot pienenevät.

Tutkielman lopussa käsiteltiin tulojakauman merkitystä muuhun talouteen. Tämä keskustelu liittyy sopivan abstraktiotason löytämiseen taloudellisten toimijoiden heterogeenisyyden mallinnuksessa. Tulojakauman prediktiivistä roolia tutkittiin simulaatiolla, jossa käytettiin hyödyksi jakaumasovituksia tehtäessä saatua tietämystä Suomen tulojakaumasta. Simulaatio perustui tulonsiirtojen funktioiden ja kulutusfunktion epälineaarisuuteen. Toinen tämän työs merkittävä tulos oli, että tulojen yhteydet tulonsiirtoihin ja kulutukseen olivat kovin monimutkaisella tavalla epälineaarisia. Tästä syystä saatiin simuloimalla vahvistettua tulos, että tulojakauman dispersio vaikuttaa merkittävällä tavalla näihin muihin taloudellisiin suureisiin. Näiden yhteyksien merkitys havaittiin, vaikka tulonsaajia mallinnettiin homogeenisinä kulutusfunktioittensa suhteen. Havaittiin siis, että epälineaarisuuksien vallitessa jakaumilla voi olla merkitystä staattisessakin kontekstissa. Tulojakaumien kohdalla ilmeinen keskustelu kannustimista jätettiin tässä yhteydessä kokonaan käsittelemättä, joskin se olisi mahdollinen seuraava laajennus tähän kehikkoon.

Joitakin jatkotutkimuksen aiheita nousee tutkielman aihepiireistä. Tulojakauman luonne jää vieläkin hieman hämärän peittoon. Tutkielmassa onnistuttiin kuitenkin löytämään tulojakaumasta Pareto-hännän sijaan niin sanottu superparetohäntä. Stabiilien jakaumien todettiin tosin kohtuullisen hyvin kykenevän tällaisen häntäkäyttäytymisen mallinnukseen. Ilmeinen seuraava tutkimuksen aihe voisi olla saman ilmiön etsiminen kansainvälisestä aineistosta. Hypoteesina voisi olla ajatus, että tulojakaumaa generoiva mekanismi on muuttunut ja että nykyisiä tulojakaumia luova mekanismi tuottaa näitä superparetohäntiä globaalisti.

Toisena lisätutkimuksen aiheena voisi olla tässä tutkielmassa tehdyn simulaation laajentaminen siten, että kysyntäfunktioiden sallittaisiin vaihtelevan. Tällöin saadaan mukaan toisenlaista tulonsaajien heterogeenisyyttä. Mallin laajentaminen dynaamiseksi voisi myös tuoda lisää valaistusta heterogeenisyyden vaikutusten ymmärrykseen. Tulojen heterogeenisyyden ja taloudellisten syy-seuraus -suhteiden epälineaarisuuksien vaikutusten tutkimus on tämän tutkielman valossa alue, johon taloustieteen kannattaisi keskittyä syvemmin.

Lähteet

Aasness, J. – Benedictow, A. - Hussein, M. F. (2003): Distributional Efficiency of Direct and Indirect Taxes, WIDER Conference on Inequality, Poverty and Human Well-being.

Atkinson, T. (1970): On the Measurement of Inequality, Journal of Economic Theory, Vol. 2, 244–263.

Attanasio, O. – Davis, S. J. (1996): Relative Wage Movements and the Distribution of Consumption, Journal of Political Economy, University of Chicago Press, 104:6, 1227-62.

Alesina, A. – Rodrik, D. (1994): Distributive politics and Economic Growth, The Quarterly Journal of Economics, Vol. 109, No. 2., 465–490.

Burda, M. – Wyplosz, C. (2001): Macroeconomics: A European Text, Oxford : Oxford University Press.

Bourguignon, F. (1979): Decomposable Income Inequality Measures, Econometrica 47, 901–920.

Carroll, C. D. (2000): Requiem for the Representative Consumer? Aggregate Implications of Microeconomic Consumption Behavior, American Economic Review, American Economic Association, 90:2, 110-115.

Carroll, C. D. – Kimball, M. S. (1996): On the Concavity of the Consumption Function, Econometrica, Econometric Society, 64:4, 981-92.

Cornia – Addison – Kiiski (2003): Income Distribution Changes and their Impact in the Post-World War II Period, Discussion Paper No. 2003/28, WIDER. Saatavilla osoitteesta http://www.wider.unu.edu/publications/dps/dps2003/dp2003-28.pdf.

Cournot, A. (1863): Principes de la théorie des richesses, L´Hachette, Pariisi.

Cowell, F. A. (1977): Measuring Inequality, Philip Allan Publishers Limited.

Cowell, F. A. (1980): On the Structure of Additive Inequality Measures, The Review of Economic Studies, Vol. 47, No. 3, 521–531.

Cowell, F. A. (1995): Measuring Inequality, toinen painos, Harvester Wheatsheaf, Hemel Hempstead.

Cowell, F. A. (2000): Measurement of Inequality, teoksessa Atkinson, A.B. - Bourguignon, F.: Handbook of Income Distribution Vol. 1, Amsterdam: Elsevier.

Dalton, H. (1920): Measurment of the Inequality of Incomes, Economic Journal, 30:9, 348–361.

Friedman, M. (1953): Essays in Positive Economics, Chicago University Press, Chicago.

Gastwirth, J. L. (1971): A General Definition of the Lorenz Curve, Econometrica, Vol. 39, No. 6, 1037–1039.

Gnedenko, B. V. – Kolmogorov, A. N. (1952): Limit distrubtions for sums of independent random variables, Addison-Wesley, Reading.

Hartley, J. E. (1996): Retrospectives: The origins of the representative agent, Journal of Economic Perspectives ,10, 169-177.

Kanbur, R. (2003): Conceptual Challenges in Poverty and Inequality: One Development Economist's Perspective, WIDER Conference on Inequality, Poverty and Human Well-being.

Kiander, J. – Kröger, O. – Romppanen, A. (2006): Finnish Economy: Structural Indicators 2006, Government Institute for Economic Research, Helsinki

Kirman, A. P. (1992): Whom or What Does the Representative Individual Represent?, Journal of Economic Perspectives, Vol. 6, 117-136.

Krusell, P. – Smith, A. A. Jr. (1998): Income and Wealth Heterogeneity in the Macroeconomy, Journal of Political Economy, University of Chicago Press, 106:5, 867-896.

Kendall, M. - Stuart, A. (1977): The Advanced Theory of Statistics, Vol 1 - Distribution Theory, Butler & Tanner Ltd., London

Leamer, E. (1983): Let's Take the Con Out of Econometrics, The American Economic Review, Vol. 73, No. 1, 31–43.

Limpert, E. – Stahel, W. – Abbt, M. (2001): Log-normal Distributions across the Sciences: Keys and Clues, BioScience, 51:5, 341–352.

Layard, R. (2002): Happiness: Has Social Science a Clue?, Lionel Robbins Memorial Lectures.

Lorenz, M. O. (1905): Methods of measuring the concentration of wealth, Publications of the American Statistical Association, Vol. 9, 209–219.

Majunder, A. – Chakravarty, S. R. (1990): Distribution of Personal Income: Development of a New Model and Its Application to U.S. Income Data, Journal of Applied Econometrics, Vol. 5, 189–196.

Mandelbrot, B. B. (1960): The Pareto-Lévy law and the distribution of income, International Economic Review, 1, 79–106.

McDonald, J. B. – Ransom, R. M. (1979): Functional Forms, Estimation Techniques and the Distribution of Income, Econometrica, 47:6.

McDonald, J. B. (1984): Some Generalized Functions for the Size Distribtion of Income, Economterica, Vol. 52, No. 3.

Montroll, E. W. – Shlesinger, M. F.(1982): On 1/f noise and other distributions with long tails, Proceedings of the National Academy of Sciences, USA, 79:3380-3383.

Musgrove, P. (1980): Income Distribution and the Aggregate Consumption Function, The Journal of Political Economy, Vol. 88, No. 3, 504–525.

Nieminen, P. – Saikkonen, P. (2006): Tilastollisen päättelyn kurssi, Helsingin yliopisto, Tilastotieteen laitos.

Nolan, J. P. (2007): Stable Distributions - Models for Heavy Tailed Data, Birkhäuser, Boston. Ei julkaistu, ensimmäinen luku osoiteessa
http://academic2.american.edu/~jpnolan. Otettu 26.6.2007.

Nygård, F. – Sandström, A. (1981): Measuring Income Inequality, Almqvist & Wiksell International.

Reed, W. J. (2003): The Pareto law of incomes - an explanation and an extension, Physica A, 319, 469–486. Saatavilla osoitteesta
http://www.math.uvic.ca/faculty/reed/index.html.

Reed, W. J. – Jorgensen, M. (2004): The double Pareto-lognormal distribution - A new parametric model for size distribution, Com. Stats -Theory & Methods, Vol. 33, No. 8., 1733–1753. Saatavilla osoitteesta http://www.math.uvic.ca/faculty/reed/index.html.

Sala-i-Martin, X. (2002): NBER Working Paper No. 8904 julkaistu huhtikuussa 2002. Saatavilla osoitteesta
http://www.nber.org/papers/w8904.

Sen, A. (1973): On Economic Inequality, Clarendon Press, Oxford.

Sen, A. (1976): Poverty: An Ordinal Approach to Measurement, Econometrica, 44: 219–231.

Sen, A. (1980): Description as Choice, Oxford Economic Papers, New Series, Vol.32, No. 3. 353–369.

Sen, A. (1992): Inequality Reexamined, New York: Russell Sage Foundation, Oxford: Clarendon Press.

Sen, A. (2000): Social Justice and the Distribution of Income, teoksessa Atkinson, A.B. - Bourguignon, F.: Handbook of Income Distribution Vol. 1, Amsterdam: Elsevier.

Shorrocks, A. F. (1980): The Class of Additively Decomposable Inequality Measures, Econometrica 48, 613–625.

Shorrocks, A., F. (1984): Inequality Decomposition by Population Subgroups, Econometrica, Vol. 52, No. 6, 1369–1386.

Silverman, B. W. (1986): Density estimation for statistics and data analysis, Chapman & Hall, London.

Singh, S.K. – Maddala, G.S. (1977): A Function for a Size Distribution of Incomes, Econometerica, Vol. 44, No, 5.

Spanos, A. (1986): Statistical Foundations of Econometric Modelling, Cambridge University Press.

Staehle, H. (1937): Short-Period Variations in the Distribution of Incomes, The Review of Economic Statistics, Vol. 19, No. 3., pp. 133–143.

Steckel, R. (1995): Stature and the Standard of Living, Journal of Economic Literature, Vol. 33, No. 4, 1903–1940.

Sørensen, P. B. – Whitta-Jacobsen, H. J. (2005): Introducing Advanced Macroeconomics: Growth and Business Cycles, London : McGraw-Hill Education.

Theil, H. (1967): Economics and Information Theory, North-Holland, Amsterdam.

Varian, H. L. (2003): Introducing Microeconomics: A Modern Approach, New York : W. W. Norton.

Vartia, P. L. I. – Vartia, Y. O. (1973): Comments on the Discussion Held at the Conference on Income Distribution.

Vartia, P. L. I. – Vartia, Y. O. (1981): Description of the Income Distribution by the Scaled F Distribution Model, teoksessa Klevmarken, N. A. – Lybeck, J. A. : The Statics and Dynamics of Income, Tieto limited.

Vartia, Y. O. (luento): Lentomuistiinpanot kurssilta “tuloerojen mittaaminen”.

Vartia, Y. O. (1988): Ovatko tilastolliset tutkimustulokset vain monimutkaisesti perusteltuja mielipiteitä?, Kansantaloudellinen aikakausikirja, 1988:3.

Vasama, P. M. – Vartia, Y. (1972): Johdatus Tilastotieteeseen II, Helsinki: Gaudeamus.

Voit, J. (2000): The Statistical Mechanics of Financial Markets, Springer, Berlin.

Waldmann, R. (1992): Income Distribution and Infant Mortalilty, Quarterly Journal of Economics, 107:4, 1283–1302.

Weisstein, E. W. ym. (a): Lorenz Curve. Sivuilta MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/LorenzCurve.html. Otettu 18.11.2006.

Weisstein, E. W. ym. (b): Gini Coefficient. Sivuilta MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/GiniCoefficient.html. Otettu 18.11.2006.

Wu, J. – Mehta, N. B. – Zhang, J.(2005): Flexible Lognormal Sum Approximation Method, IEEE Global Telecommunications Conference (GLOBECOM), Vol. 6, 3413–3417.