TURUN YLIOPISTO

TIIVISTELMÄ

	Kandidaatintutkielma
X	Pro gradu -tutkielma
	Lisensiaatintutkielma
	Väitöskirja

Oppiaine	Taloustiede	Päivämäärä	2.6.2020
Tekijä	Tuomas Kaivola	Matrikkelinumero	123456
		Sivumäärä	75+liitteet
Otsikko	Tuloerojen rakenne ja tilastolliset ominaisuudet		
Ohjaaja	Prof. Hannu Salonen		

Tiivistelmä

Tuloerojen selittämiseksi on esitetty useita eri teorioita ja tulojakauman kuvaamiseksi on ehdotettu useita eri matemaattisia malleja. Lähempi matemaattinen tarkastelu alkoi Vilfredo Pareton havainnosta, jossa tulo- ja varallisuuserot noudattavat suurta samankaltaisuutta maasta riippumatta. Tästä kehittyi potenssilain mukainen todennäköisyysjakauma, jota kutsutaan Paretojakaumaksi. Tämän kaltainen jakauma esiintyy monissa eri systeemeissä, joille on yhteistä rikkaat rikastuu -ilmiön kaltaiset mekanismit. Kaupunkien kokoa, rikosten määrää tai suurituloisia voidaan kuvata Paretojakaumalla.

Paretojakauma kuitenkin sopii kuvaamaan vain suurituloisimpia. Perinteisen tuloeroja käsittelevän kirjallisuuden mukaan lognormaalijakauma on paras vaihtoehto kuvaamaan pieni- ja keskituloisia. Lognormaalijakauma syntyy monien riippumattomien satunnaismuuttujien tulosta keskeisen raja-arvolauseen mukaisesti. Andrew Roy havainnoi kuinka yksilön tuottavuus perustuu hänen monien ominaisuuksien tuloon synnyttäen ihmisten tuloerojen lognormaalijakauman. Ekonofyysikot ovat ehdottaneet tulojen noudattavan eksponentti- tai gammajakaumaa lognormaalijakauman sijaan. Heidän mukaansa taloutta voidaan tutkia termodynamiikan oppien avuilla, koska talous koostuu monista toisiinsa vuorovaikuttavista yksiköistä.

Näitä jakaumia sovitetaan Suomen vuosien 2014 – 2018 tuloaineistoon. Pieni- ja keskituloisten palkkatuloja kuvaa paremmin lognormaalijakauma perinteisen näkökulman mukaisesti. Ero gammajakaumaan on kuitenkin pieni. Pääoma- ja osinkotulojen kuvaamiseen sopii gammajakauma, joka muistuttaa ekonofyysikoiden markkinoiden kineettisistä vaihdantamalleja. Pieni- ja keskituloisten tuloerojen vaihtelu vuosien välillä on pientä. Tuhannen suurituloisimman tulot sopivat ennakoidusti Paretojakauman mukaiseen kuvaukseen pääoma- ja ansiotulojen osalta. Ansiotulot jakautuvat tasaisemmin kuin pääomatulot. Suurituloisten tuloerojen vaihtelu vuosien välinen on suurta.

Vain kahden parametrien jakaumilla pystytään kuvaamaan tuloja suhteellisen tarkasti, mikä viittaa säännönmukaisiin mekanismeihin. Näiden tuloerojen ominaisuuksien tilastollinen tarkastelu voi täten tarjota uusia työkaluja taloustieteen käyttöön.

Avainsanat	Tuloerot, Suomi, Tulojakauma

TURUN YLIOPISTO
Kauppakorkeakoulu

TULOEROJEN RAKENNE JA TILASTOLLISET OMINAISUUDET

Taloustieteen
pro gradu -tutkielma

Laatija:
Tuomas Kaivola

Ohjaaja:
Prof. Hannu Salonen

2.6.2020
Turku

SISÄLLYSLUETTELO

KUVIOT

TAULUKOT

1 JOHDANTO

Tutkielmassa tarkastellaan millaisilla matemaattisilla jakaumilla tuloeroja voidaan kuvata ja mitä nämä matemaattiset mallit kertovat tuloerojen syntymekanismeista. Tuloerojen tarkastelu on tärkeää kahdesta syystä. Palkka työstä, voitto yritystoiminnasta tai tuotto sijoituksesta ovat talouden suureita, joita taloustiede käyttää lukuisissa analyyseissään ja teorioissaan. Tulojen jakautumisen tilastollinen tarkastelu siksi parantaa kokonaiskäsitystä talouden rakenteesta. Tulot ovat tärkeä osa ihmisten hyvinvointia talouden rakenteellisen ominaisuuden lisäksi. Absoluuttisten tulojen lisäksi ihmisen suhteellinen tulotaso korreloi hyvinvoinnin kanssa. Suuret tuloerot voivat täten haitata hyvinvointia (Yu & Chen 2016).

Tästä syystä kasvaneet tuloerot ovat jälleen nousseet puheenaiheeksi niin politiikassa kuin taloustieteessä. Tulojakauman rakenne ja tilastollinen tarkastelu on ollut kuitenkin taloustieteen kiinnostuksen kohteena pitkään: ”*Maatalouden tuotteet – kaikki ne, jotka viljellään työvoiman, koneiden ja pääoman yhteiskäytöllä – jaetaan kolmelle yhteiskuntaluokalle; maan omistajalle, viljelyyn tarvittavien varastojen tai pääoman omistajille ja työntekijöillä jotka antavat työpanoksensa viljelylle. Mutta yhteiskuntaluokille ositettu osuus maatalouden tuotannosta, joita kutsutaan vuokraksi, voitoksi ja palkaksi, ovat olennaisesti erilaiset. Se riippuu maaperän hedelmällisyydestä, pääoman ja väkimäärän kertymisestä ja maataloudessa käytetyistä taidoista, kekseliäisyydestä ja välineistä. Näiden jakautumista selittävien tieteellisten lakien määrittely on Kansantalouden ja Verotuksen Periaatteen ensisijainen on kysymys. Tiede on kehittynyt huomattavasti sitten Turgotin, Stuartin, Smithin, Sanonin, Sismodin ja muiden kirjoituksien, mutta ne ovat tarjonneet vain vähän tietoa selittämään vuokran, voittojen ja palkkojen luonteesta”*

Näillä alkusanoilla alkoi David Ricardon klassikkoteos ”*Taloustieteen ja verotuksen periaatteet*” vuodelta 1817. Paljon on muuttunut kahdessa sadassa vuodessa taloustieteen saralla, mutta hieman yllättäen emme vieläkään tarkalleen tiedä tuloerojen luonteesta kaiken kattavasti. Teorioita ja empiirisiä todisteita on laajalti, mutta ne ovat hajanaiset. Taloustieteessä on kiinnitetty enemmän huomioita tuloerojen vaikutuksiin tai keskitytty tarkastelemaan yksittäisten tekijöiden vaikutusta tuloeroihin, joten Ricardon pohtima tuloerojen luonne on jäänyt taka-alalle. Ricardo käytti teoksessaan sanaa luonne, joka tarkoittaa jollekin ilmiölle tai asialle merkittäviä toiminnallisia piirteitä ajasta ja paikasta riippumatta. Tämä on tutkielman tulokulma, jolla tuloerojen rakennetta lähdetään tarkastelemaan. Tuloeroihin vaikuttaa lukuisat pienet ja suuret tekijä yhdessä ja erikseen, eikä tutkielman tarkoituksena ole tarkastella näitä kaiken kattavasti.

Ilmiön ollessa laaja ja kompleksinen on luonnollista, että sitä selittämään löytyy useita toisiaan tukevia ja kilpailevia teorioita. Vilfredo Pareton empiiriset havainnot säännönmukaisista tuloeroista toimi lähtölaukauksena tulojakauman tilastolliselle tarkastelulle. Hyvin nopeasti tutkijat ehdottivat uusia malleja ja kävivät läpi näiden matemaattisten jakaumien tilastollisia ominaisuuksia. Tutkielmassa käydään lyhyesti läpi Pareton, Gibratin, Champernownen, Mandelbrotin, Royn teorioita tuloerojen rakenteesta. Näiden jälkeen esitellään fyysikoiden tuoreempi teoria, joka haastaa perinteisen näkemyksen tuloeroista ottaen inspiraatiota termodynamiikasta. Osa on lähtenyt liikkeelle teoriasta ja edennyt siitä empiriaan. Toiset ovat lähteneet empiirisistä havainnoista ja kehittäneet niistä teoriansa.

Tästä kirjallisuuskatsauksesta nousee esille matemaattisia jakaumia, joita on käytetty kuvaamaan tuloeroja – potenssijakauma, lognormaalijakauma ja eksponentti- tai gammajakauma. Tutkielmassa käydään läpi näiden jakaumien tilastollisia ominaisuuksia ja esiintymisiä eri systeemeissä. Nämä jakaumat toimivat linkkinä tuloerojen ja muiden samankaltaisten systeemien välillä, jolloin se myös auttaa kätevästi intuitiivisesti ymmärtämään näiden tuloerojen luonteesta ja tilastollisia ominaisuuksia. Tällöin saadaan Ricardon peräänkuuluttama selvyys tuloerojen luonteesta. Tämä tuloerojen ”ylhäältä alaspäin” tarkastelu ja matemaattisten jakaumien ominaisuuksien tarkastelu voi tarjota mielenkiintoisia havaintoja tuloerojen muodostumisesta, jotka ehkä muutoin jäisivät piiloon. Matemaattisen jakaumien ominaisuuksien jälkeen tarkastellaan empiirisesti Suomen tuloeroja vuosilta 2014 – 2018. Jakaumia sovitetaan aineistoon, jolloin havaitaan paras jakauma kuvaamaan tuloja. Tämän jälkeen viimeisessä kappaleessa pohditaan analyysin tuloksia ja miten ne heijastuvat tuloeroja käsittelevään kirjallisuuteen.

2 TULOEROJEN TUTKIMUKSEN HISTORIAA

2.1 Perinteinen näkemys

Tuloerojen lähempi tilastollinen tarkastelu voidaan katsoa todella alkaneen vasta 80 vuotta Ricardon klassikkoteoksen jälkeen vuonna 1906 italialaisen taloustieteilijän Vilfredo Pareton tutkimuksista. Hän tilastoi kuinka maanomistus ja tulot jakautuvat ihmisten kesken ja huomasi sen noudattavan jyrkästi vinoutunutta jakaumaa, jossa rikkaimmat omistavat ja tienasivat suhteellisesti enemmän kuin köyhemmät (Pareto 1906). Hänen aineisto osoitti kuinka 20 % Italian suurimmista maanomistajista omisti yhteensä 80 % kaikista yksityisistä maista. Hän jatkoi tutkimuksiaan ja huomasi, kuinka sama jakauma toistuu myös muissa maissa. Pareto kehitti tästä tulo- ja varallisuuserojen havainnosta 80/20-periaatteen havainnoidessaan samantapaisen ilmiön toistuvan puutarhassaan, jossa 20 % hernepalkoista tuotti 80 % kaikista herneistä. Yhdysvaltalainen Joseph. M. Juran nimesi tämän 40 vuotta myöhemmin Pareton periaatteeksi kirjassaan, jossa käsitellään tarkemmin kuinka 80 % tapahtumista johtuu 20 %:n vaikutuksesta (Juran & Godfrey 1941). Hän osoitti miten Pareton periaate pätee myös laaduntarkkailussa, jossa suurin osa tuotteen tai tuotannon vioista aiheutuu vain muutamista virheistä. Pareton periaate muotoutui vähitellen muiden tieteellisten tutkimuksien edetessä matemaattiseen muotoonsa potenssijakauman mukaiseksi todennäköisyysjakaumaksi – Paretojakaumaksi. Vilfredo Pareto itse kutsui jakaumaa luonnolliseksi laiksi, koska hän piti sitä universaalina.

Pareton havainto herätti tiedemaailmassa mielenkiintoa. Suhteellisen nopeasti huomattiin kuitenkin, että potenssijakauman universaalisuus pätee vain rikkaiden osalta. ”*Pareto-lakia ei todellakaan ole. On aika, että se tulisi hylätä kokonaan tuloerojen jakaumaa käsittelevissä tutkimuksissa*” (Shirras 1935). Tarvittiin uusi malli ja tähän haasteeseen tarttui ranskalainen taloustieteilijä Robert Gibrat. Hän ensimmäisenä ehdotti stokastisen prosessin synnyttämää lognormaalijakaumaa kuvaamaan paremmin pieni- ja keskituloisia (1931). Hän kutsui tätä *suhteellisen vaikutuksen laiksi* (eng. *Gibrat's law of proportionate growth*). Tällä hän kuvaili kuinka määrän muutos on riippumaton itse määrästä ja tulojakauman lisäksi hän sovelsi tätä yritysten kokojen jakaumaan. Hän teoretisoi kuinka yrityksen koko ja sen kasvu ovat toisistaan riippumattomia. Tätä prosessia hän sovelsi myös kaupunkien väkimäärän jakaumaan. Sekä kapunkien että yritysten koot on myöhemmin todettu noudattavan kuitenkin potenssijakaumaa. Tämä stokastinen eli sattumanvaraisesti etenevä prosessi lähestyy ajan myötä lognormaalijakaumaa tai tiettyjen ehtojen täyttyessä Paretojakaumaan.

Englantilainen taloustieteilijä Champernowne ehdotti vastaavanlaista stokastista mallia, joka perustuu *Markov-prosessille* (1953). Markov-prosessin kehittymisen määrittää vain määreen nykyarvon ja historialla ei ole merkitystä. Seuraavan jakson tulot määrittää todennäköisyysjakauma, jossa todennäköisyys pienenee tulohyppäyksen kasvaessa – yksilön on todennäköisempää saada pieni kuin suuri palkankorotus. Tämä johtaisi normaalisti lognormaalijakaumaan, mutta Champernownen mallissa tuloille oli minimiraja, joka sai aikaan potenssijakauman hännässä. Tämä malli on myöhemmin kohdannut kritiikkiä, koska siitä puuttuu taloustieteellinen pohja. Markov-prosessissa ikäkohortin tuloerot kasvavat ajan kuluessa ja mistä ei ole todisteita. Toisaalta se sopii ansiotuloja paremmin kuvaamaan pääomatulojen jakautumista (Lydall 1959). Champernowne itsekin totesi ettei hänen mallinsa tuo suurta selvyyttä mekanismeihin, jotka selittävät empiirisiä havaintoja.

Kuuluisa ranskalainen matemaatikko ja fraktaaliteorian popularisoija Benoit Mandelbrot toi oman näkökulmansa tuloerojen tilastolliseen tarkasteluun. Hän keskittyi rikkaimpien tuloja kuvaavan Paretojakauman tilastollisiin ominaisuuksiin ja antoi *Pareto–Lévyjakaumalle* lisänimen mentorinsa Paul Lévyn mukaan (1960). Hän toivoi ekonomistien kiinnittävän enemmän huomioita vakaille ei-gaussilaisille todennäköisyysjakaumille, koska ”*työkalut, joita käytämme, ovat yhtä tärkeitä kuin tulokset, jotka toivomme saavuttavan*”. Vakaus tarkoittaa kuinka riippumattomat Pareto–Lévyjakautuneiden komponenttien summa on myös Pareto–Lévyjakautunut. Mandelbrotin mukaan nämä matemaattiset mallit soveltuvat tulojen lisäksi moniin muihin taloustieteellisiin suureiden mallintamiseen. Hän jatkoi tutkimuksiaan muutamaa vuotta myöhemmin ja ehdotti kuinka kaikkia tuloja voitaisiin kuvata eri jakaumien yhdistelmällä (1963).

Lognormaalijakauma nousi ensimmäisten teoreettisten formalisointien jälkeen ensimmäiseksi matemaattiseksi kandidaatiksi kuvaamaan muiden kuin rikkaiden tuloeroja. Lydall on kritisoinut Champernownen ja Gibratin tapaisia malleja, koska niissä keskitytään liikaa satunnaisuuteen ja liian vähän tekijöihin, joiden tiedetään vaikuttavan tulojen jakautumiseen (1959). Royn tutkimuksessa pyritään vastaamaan tähän kysymykseen ja havainnoimaan miksi ansiotulot käytännössä noudattaa lognormaalijakaumaa (1950). Royn tutkimuksien aikoihin pohdittiin miksi tulojakauma on voimakkaasti vinoutunut eikä normaalijakautunut. Tätä ongelmaa kutsuttiin *Pigoun paradoksiksi* (Pigou 1932), koska tulojen oletettiin perustuvan normaalijakautuneisiin kykyihin. Pigou selitti vinoumaa perityllä varallisuudella ja muulla pääomalla. Pigoun selitys osoitettiin vääräksi Royn osoittaessa myös ansiotulojen vinoutuneen jakauman. Royn tutkimuksen lähtökohta oli Gibratin suhteellisen vaikutuksen laki, jossa stokastisien shokit olivat työntekijän normaalijakautuneita kykyjä tai ominaisuuksia. Näiden normaalijakautuneiden ominaisuuksien tulo määrittää työntekijän tuotoksen. Yksinkertaisen esimerkin oletetaan tuotoksen riippuvan vain kahdesta ominaisuudesta – älykkyydestä ja ahkeruudesta. Tällöin työntekijän ollessa 50 % fiksumpi ja 50 % ahkerampi kuin keskiverto työntekijä, saa hän 2,25-kertaisesti enemmän palkkaa keskiarvoon verrattuna. Tämä tulos on johdettu keskeisestä raja-arvolauseesta; positiivisten satunnaismuuttujien tulon logaritmi on normaalijakautunut.

Roy pohti syitä miksi tämä yksinkertainen mekanismi ansiotulojen jakautumisessa ei pätisi hyvien tieteellisten periaatteiden mukaisesti. Hän pohti komponenttien korrelaation voisi rikkoa keskeisen raja-arvolauseen. Taidot korreloivat iän ja koulutuksen kanssa, jolloin tulot voisivat poiketa lognormaalista. Hän viittaa Haldanen tutkimukseen (1942), jossa kuitenkin osoitetaan kuinka kolmen korreloituneen satunnaismuuttujan tulo ei poikkea merkittävästi riippumattomien satunnaismuuttujien tulosta. Royn toinen pohtima syy lognormaalisuuden rikkojaksi osoittautui hedelmälliseksi taloustieteelliseksi tutkimuskysymykseksi. Vuotta myöhemmin ilmestyneessä tutkimuksessa hän pohti markkinamekanismia, joka tasapainottaa työntekijät työpaikkojen välillä (1951). Tämä ohjaava mekanismi on alunperin johdettu Ricardon suhteellisen edun periaatteesta, missä talouden perusongelma on määritellä kuinka heterogeeniset työntekijät jakautuvat eri työpaikkoihin. Royn tutkimuksessa hän tarkastelin hypoteettista tilannetta, jossa työntekijät jakautuivat metsästäjä- ja kalastaja-ammatteihin. Roy halusi selvittää mitkä ehdot vaikuttivat metsästäjien ja kalastajien jakautumiseen ja oliko lopputulos tehokas. Tätä ajatuskoetta alettiin kutsumaan Royn malliksi, jota taloustieteen Nobelisti James Heckman kutsui yhdeksi taloustieteen tärkeimmiksi malliksi (2010). Tätä työntekijöiden ja yritysten epäsuhdan ilmiötä kutsutaan nyt yleisemmin kohtaanto-ongelmaksi (Jovanovic 1979).

Pitkään ajateltiin tulojakauman muodostuvat ihmisten kokonaisuudessaan tuottavuudesta. Nämä tutkimukset ovat perinteisesti jakautuneet Gibratin ja Champernownen tapaisiin stokastisiin matemaattisiin selityksiin ja toinen koulukunta tutkii taloudellisia, poliittisia ja demografisia tekijöitä. Näiltä tutkimuksilta puuttuu vakuuttava empiirinen pohja. Tämä naiivi lähtökohta on kuitenkin kyseenalaistettu useissa tutkimuksissa (Lydall 1959). Myöhemmin on alettu keskittymään syihin miksi markkinat epäonnistuvat työmarkkinoilla, jolloin selittäviksi tekijöiksi on ehdotettu epäsymmetristä informaatiota tai ulkoisvaikutuksia.

2.2 Fyysikoiden uusi teoria

Tuloerojen tutkimus sai uuden suunnan yllättävästä suunnasta. Taloustiede on ottanut fysiikasta mallia ennenkin, mutta nyt inspiraationa toimi termodynamiikka eli lämpöoppi klassisen mekaniikan sijaan. Termodynamiikassa tutkitaan systeemin lämpöä, energiaa, työtä sekä aineiden mitattavia ominaisuuksia. Näiden väliset riippuvuudet rakentuu termodynamiikan neljän pääsäännön ympärille – nollas, ensimmäinen, toinen ja kolmas. Ensimmäinen ja toinen ovat taloustieteen ja tuloerojen kannalta tärkeimmät. Ensimmäinen pääsäännön mukaan suljetun systeemin energia säilyy vakiona. Tämä on termodynamiikan versio klassisen fysiikan energian säilymislaista. Suljetun laatikon lämpötila pysyy vakiona ellei joku avaa kantta ja lämpö pääse karkaamaan.

Termodynamiikan toinen pääsääntö kertoo kuinka suljetun systeemin *entropia* kasvaa ajan myötä. Entropia tärkeä käsite sillä se kertoo systeemin epäjärjestyksen määrän. Suljetussa systeemissä kaikilla kaasumolekyyleillä ei ole sama nopeus tai energia. Ne törmäilevät satunnaisesti toisiinsa, jolloin molekyylien kineettisen energian säilymislain vuoksi molekyylien energia esitystavasta riippuen noudattaa eksponentti- tai gammajakaumaa – suurimmalla osalla molekyyleistä on vähän energiaa ja pienellä osalla paljon. Tälläistä eksponenttijakaumaa kutsutaan *Boltzmann-Gibbs-jakaumaksi.* Systeemin kaasumolekyylien keskinopeus kertoo systeemin lämpötilan, koska lämpöenergia on kaasumolekyylien liike-energiaa. Toisen pääsäännön takia kahden eri lämpöisen systeemin lämpötilaero tasoittuu niiden ollessa kosketuksessa toisiinsa. Vaikka suljetun systeemin kaasumolekyyleillä olisi aluksi täysin sama nopeus, kaasumolekyylit liikkuu kohti Boltzmann–Gibbsin mukaista tilaa satunnaisen törmäilyn myötä ja entropian kasvaessa. Teoriassa nopeat kaasumolekyylit voisivat pakkautua eri nurkkaan kuin hitaat kaasumolekyylit, jolloin suljetun systeemissä olisi kaksi eri lämpötilaa. Tämä on kuitenkin niin epätodennäköistä ettei niin käytännössä koskaan tapahdu. Todennäköisin kaasumolekyylien tila on tila, jossa entropia eli epäjärjestys maksimoituu. Tämänkaltaisessa suljetussa systeemissä lämpötila on vakio, joka toimii samalla rajoitteena. Suljetun systeemin lämpötila ja samalla molekyylien energian on oltava vakio. Tällöin tämän rajoitteen alaisuudessa Boltzmann-Gibbs-jakauma maksimoi systeemin entropian eli kaasumolekyylien epäjärjestyksen.

Ekonofyysikkoiksi itseään kutsuvien tutkijoiden mukaan taloutta pystytään tutkimaan tilastollisen fysiikan ja termodynamiikan oppien tavoin. Heidän mukaan se tarjoaa uusia välineitä ja näkemyksiä taloustieteen ongelmiin, koska molemmissa tutkitaan suurta kompleksista kokonaisuutta, joka koostuu pienistä toisiinsa reagoivista yksiköistä. Ekonofyysikot alunperin tutkivat rahan jakautumista, koska he ajattelivat rahataloudessa olevan samantapainen säilymislaki kuten termodynamiikassa (Drăgulescu & Yakovenko 2000). Drăgulescun ja Yakovenkon mukaan talousjärjestelmässä rahan kokonaismäärä säilyy vakiona lyhyellä aikavälillä, jolloin taloudellisten agenttien hallussa pitämä rahamäärän jakauman tulisi noudattaa termodynamiikasta tuttua Boltzmann-Gibbs-jakaumaa (2000). Yksinkertaiset simulaatiot osoittivat kuinka rahamäärän eksponenttijakauma syntyy satunnaisen vaihdannan myöden, vaikka kaikilla oli alunperin yhtäsuuruinen alkuvarallisuus. Tutkijoiden mukaan heidän satunnaisen vaihdannan malli sopisi kuvaamaan parhaiten suljettua uhkapelirinkiä. Oikeassa taloudessa agentit eivät vaihda rahaa satunnaisesti vaan deterministisesti pieniä summia omaisuudestaan, mutta koko systeemin tasolla sitä voidaan pitää satunnaisena, jolloin termodynamiikan mukainen tarkastelu on edelleen mielekästä.

Ekonofyysikot aluksi sovelsivat osaamistaan finanssimarkkinoiden puolella, jossa he keskittyivät varallisuuden ja rahan liikkeiden mallintamiseen. Näitä malleja *markkinoiden kineettisiksi vaihdantamalleiksi* (eng. *kinetic exchange mode*l). Näissä taloudelliset agentit vaihtavat rahaa keskenään pitäen systeemin kokonaisvarallisuuden vakiona kuten kaasumolekyylien energioiden tapaan suljetussa systeemissä. Ajan myötä entropian lisääntymisen myötä varallisuuserot asettuvat termodynamiikan mukaiseen tasapainoon kaasumolekyylien tapaan. (Chatterjee ym. 2007). Tätä lähtökohtaa ovat kuitenkin monet taloustieteilijät kritisoineet (Gallegati 2006).

Yleisön hallussa olevaa rahamäärää on kuitenkin vaikea arvioida luotettavasti ja rahan markkinoiden vaihdantamallit vielä vahvasti nojautuvat teoreettisille pohjalle. Tästä huolimatta Yakovenko ja Drăgulescu (2001) tarkastelivat Iso-Britannian ja Yhdysvaltojen osavaltioiden oikeaa dataa kotitalouksien tuloista ja varallisuudesta. He eivät tehneet muita oletuksia aineistosta kuin sen koostuvan joukosta toisiinsa vuorovaikutuksissa olevista taloudellisista agenteista. He havaitsivat rikkaimman 1 – 3 % jakautuvan ennakoidusta potenssijakauman mukaisesti ja pieni- ja keskituloisten osalta he löysivät tutun eksponenttijakauman. Koska he havaitsivat Boltzmann-Gibbs-jakauman, he olettivat myös tulojen noudattavan termodynamiikan kaltaisia mekanismeja. Yakovenko jatkoi tutkimuksia Silvan kanssa (2005) ja he vahvistivat Drăgulescun ja Yakovenkon aiemman tutkimustuloksen. Tällä kertaa he analysoivat Yhdysvaltojen tulojakauman vuosilta 1983–2001. He osoittivat yhteiskunta jakautui kahteen luokkaan – eksponenttijakaumaa noudattavaan keski- ja pienituloisiin sekä potenssijakaumaa noudattavaan suurituloisin 1 – 3%. Tässä tapauksessa pieni- ja keskituloisilla tarkoitetaan 97 – 99% väestön osaa, jota hyvin kuvaava eksponenttijakauma oli vakaa tarkastelu vuosien aikana. Sen sijaan suurituloisimman 1 – 3 %:n potenssijakauma vaihteli suuresti osakekurssien mukaan. Tutkijoiden mukaan suurituloisimpien tuloja hallitsee pääomatulot ja muiden ansiotulot, joka selittää miksi vaihteluiden eroja. Samansuuntaisia tuloksia ja vahvistusta pieni- ja keskituloisten eksponenttijakaumalle on löytynyt muissa tutkimuksissa (Jagielski & Kutner 2013). Myös tulojen gammajakaumalle on löytynyt vahvistusta (Efthimiou & Wearne 2016).

Rahan tai varallisuuden vaihdannassa säilymislaki on helppo hyväksyä, mutta tulojen osalta se on kyseenalaisempaa. Yakovenko ja Rossier tutkimuksessaan kuitenkin huomioivat tämän: ”Vaikka rahaa siirretään säännöllisesti agentilta toiselle transaktion yhteydessä, ei ole tyypillistä agenteille vaihtaa osaa tuloistaan. Tulojen epäsuoraa siirtoa voi kuitenkin tapahtua, kun yksi työntekijä saa palkan korotuksen ja toisen alennetaan kokonaisvuosibudjetin ollessa vakio.” He osoittavat kuinka

Champernownen mallin kaltainen Markov-prosessi voi johtaa myös eksponenttijakauman mukaiseen tuloihin. Tämä syntyy, koska agentit eivät vaihda keskenään palkkaa vaan kaikki agentit satunnaisesti ”vaihtavat” palkkaa vakioisen reservin kanssa (Yakovenko & Rossier 2009).

Ekonofyysikkojen teoriat perustuvat oletukseen talouden säilymislaista, koska he havaitsevat samantapaisen eksponenttijakauman tutkiessaan tuloerojen tilastollisia ominaisuuksia. Tätä lähtökohtaa ovat kuitenkin monet taloustieteilijät kritisoineet (Gallegati 2006). Aikaisemmin tuloeroja tutkinut Mandelbrot totesi miksi fysiikan tarjoama näkökulma on vähintäänkin mielenkiintoinen: ”*On suuri kiusaus pitää rahan vaihdantaa taloudessa samankaltaisena molekyylien fyysisten törmäysten energian vaihdannan kanssa. Löyhimmällä mahdollisella tavalla molempien vuorovaikutusten tulisi johtaa samanlaisiin tasapainotiloihin. Toisin sanoen tulojen jakaumaa pitäisi voida selittää mallilla, joka on samankaltainen kuin tilastollisessa termodynamiikassa.*” Hän myös huomauttaa kuinka monet tutkijat ovat käyttäneet termodynamiikan malleja tietämättään (1960). Kuitenkin koko ekonofyysikkojen oletus talouden säilymislakien ja hyvinkäyttäytyvien eli ennustettavien agenttien käyttäytymisestä on laajalti kritisoitu taloustieteen puolella (Heinrich 2013).

Pareton ajoista sadassa vuodessa tutkijat ovat päässeet yksimielisyyteen vain rikkaiden ihmisten potenssijakaumasta, mutta kysymys on vielä avoin mitä tilastollista jakaumaa enemmistö noudattaa. Lognormaalijakauma viittaisi perinteiseen käsitykseen ihmisten tuottavuudesta ja likimain normaalijakautuneista taidoista. Osa tutkimuksista on kuitenkin osoittanut ettei lognormaalijakauma välttämättä kuvaisikaan tulojakaumaa. Eksponenttijakauman näyttäisi myös sopivan aineistoon hyvin, mutta siltä puuttuu kunnollinen taloustieteellinen pohja. Ihmiset eivät ole toisiinsa törmäileviä kaasumolekyylejä suljetussa systeemissä. Lisäksi ekonofyysikoiden ehdottaman eksponenttijakauman ongelmana on, ettei mikään ominaisuuksien ja inhimillisen pääoman yhdistelmästä syntyvä yksilöiden rajatuottavuus johda tulojen eksponentiaaliseen jakautumiseen. Tämän takia eksponentiaalisesti jakautuneet tulot voivat viitata siihen, ettei työmarkkinoilla pystytä hinnoittelemaan työpanoksen hintaa vastaamaan tuottavuutta. Jos tulot jakautuvat riippumatta tuottavuudesta, taloustieteilijöiden on pakko tutkia tarkkaan joitain heidän keskeisimpiä uskomuksiaan. Uussklassisten mallien kulmakivi on rajatuottavuuksien mukaiset palkat. Uusklassisessa taloustieteessä ymmärretään kuitenkin ettei reaalimaailma toimi täydellisesti. Esimerkiksi ulkoisvaikutukset tai epäsymmetrinen informaatio haittaa markkinoiden toimintaa. Katsomalla muihin systeemeihin, joissa potenssi-, lognormaali- tai eksponentti- tai gammajakaumia esiintyy tutkielma pyrkii tuomaan selkeyttä monimutkaiseen ilmiöön.

3 JAKAUMAT

Kun tarkastellaan tuloeroja makrotasolla, voidaan havainnoida mikä tilastollinen jakauma sopii kuvaamaan niitä. Silloin kun tiedetään mitä jakaumaa tulot noudattavat makrotasolla, voimme ymmärtää paremmin mikrotason mekanismeja, jotka synnyttävät niitä. Vaikka mikrotason selitykset tuloeroille ovat valtavan monimutkaiset ja satunnaiset, noudattavat ne makrotasolla yllättävän säännönmukaisia ominaisuuksia. Tämänkaltaisten emergenttien ilmiöiden tarkastelu on osoittautunut hedelmälliseksi aiheeksi taloustieteen saralla (Harper & Lewis 2012). Talous on luonnon kaltainen kompleksinen systeemi, joka koostuu toisiinsa vuorovaikuttavista yksiköistä kuten ekonofyysikot muistuttavat. Tästä syystä taloudesta ja luonnon ekosysteemeistä voi löytää paljon yhtäläisyyksiä. Ekologia on luonnon ekosysteemien tutkimista kuten taloustiede on talouden. Molemmat tieteenalat on johdettu samasta latinankielisestä kantasanasta *oikos*, joka tarkoittaa taloa tai kotia. Taloustiede saadaan kun kantasanaan lisätään *nemein*, joka tarkoittaa hoitamista tai resurssien hallitsemista – taloustiede on kodin hallintaa. *Logia* tarkoittaa oppia, joten ekologia tarkoittaa kodin oppia.

Ekologia ja taloustiede ovat kuitenkin tieteellisenä alana kasvaneet hyvin erilleen saman kodin tutkimisesta huolimatta. Taloustieteilijät ovat ammentaneet inspiraatiota teorioihinsa ja malleihinsa newtonilaisesta fysiikasta. Ongelmaksi on muodostunut ettei taloustieteen mallit ole samalla tavalla deterministisiä kuin fysiikassa, jolloin teoreettisten mallien ja reaalimaailman välimatka voi kasvaa liian suureksi. Taloustieteen Nobel-palkinnon talouskasvun malleista saanut Paul Romer muistuttaa tästä mallien ja todellisuuden välimatkasta esseessään makrotalouden ongelmista (2016). Ehkä tässä mielessä taloustiede voisi ottaa oppia ekologiasta, jossa pyritään hahmottamaan luonnon kokonaisuutta. Joseph Schumpeter pohti kirjassaan Vilfredo Pareton tuloerojen empiiristen tulosten universaalista luonnetta. Hänen totesi Paretojakauman kaltaisten universaalien havaintojen voivan luoda pohjaa kokonaan uudentyyppiselle teorialle (Schumpeter 1951). Taloustieteen puolella on jo pitkään kvalitatatiivisesti ymmärretty talouden kompleksinen luonne. Adam Smithin näkymättömän käden ohjauksesta ilman erityistä suunnittelua markkinoiden mikrotason kaaoksesta muodostuu toimivia kokonaisuuksia. Tämä on tiedetty jo yli 200 vuotta. Toinen taloustieteen nobelisti Friedrich von Hayek kutsui tätä spontaaniksi järjestykseksi. Ihmisten muodostaman taloudellinen kokonaisuus on täynnä tämäntapaisia emergenttejä ilmiöitä (Harper & Lewis 2012). Näiden makrotason säännönmukaisuuksien tutkimiseen tarvitaan matemaattisia malleja, joita käydään seuraavaksi läpi.

3.1 Potenssijakauma

Potenssilaki tarkoittaa kahden muuttujan välistä yhteyttä, jossa muuttuja muuttuu toisen muuttujan potenssiin ja näiden muuttujien jakaumaa potenssijakaumaksi (eng. *power law distribution*). Potenssijakaumat kuuluvat paksuhäntäisten jakaumien luokkaan, koska ne voivat saada erittäin suuri arvoja. Potenssilain todennäköisyysjakaumaa kutsutaan Vilfredo Pareton mukaan Paretojakaumaksi, vaikka sillä voi olla tilanteesta riippuen myös muita tutkijoiden mukaan nimettyjä nimiä kuten Zipf, Yule, Lotka tai Bradford. Seuraavaksi käydään läpi Paretojakauman tilastollisia ominaisuuksia, koska se on yleisin potenssijakaumista ja sen syntyhistoria pohjautuu tuloerojen mittaamiseen.

Paretojakaumia on 4 eri päätyyppiä, jotka on yksinkertaisesti nimetty järjestysluvun mukaan. Ensimmäisen luokan Paretojakauma on yleisin ja yksinkertaisin. Se määritellään kahden parametrin avulla – skaala parametri x_m ja muoto parametri α. Paretojakauman tiheysfunktio on muotoa (1) ja kertymäfunktio on muotoa (2).

(1) $$\alpha x_m^{\alpha} x^{-\alpha-1}$$

(2) $$1-\left(\frac{x_m}{x}\right)^{\alpha}$$

On tärkeää huomata kuinka parametria x_m merkitään samalla kirjaimella kuin mitattavaa muuttuja x. Se on Paretojakauman pienin arvo eli muuttujan x minimi, joka määrittää jakauman sijainnin. Esimerkiksi kaupunkien koon jakaumassa tämä olisi asukasmäärältään pienin kaupunki. Tämän johdosta Paretojakauma on määritelty välille $[x_m,\infty)$ eikä se voi saada negatiivisia arvoja.

Potenssijakaumien ehkä tärkein ominaisuus on *skaalautumattomuus* tai *mittakaavan invarianssi.* Skaalautumattomuutta voidaan tarkastella muokkaamalla tiheysfunktion (1) parametreja. Merkitsemällä $\alpha x_m^{\alpha}=C,\ b=-\alpha-1$ ja laittamalle se yhtäsuuruiseksi muuttujan y kanssa saadaan funktio (3). Tämä on edelleen teknisesti sama kuin Paretojakauman tiheysfunktion (1), koska vain parametrit on merkitty eri tavalla. Ottamalla funktiosta (3) molemmilta puolilta logaritmi muuttuu se funktioksi (4). Näin Potenssijakauma muuttuu tutuksi suoran yhtälöksi logaritmisella asteikolla. Tämä lineaarisuus kun muuttujista otetaan logaritmi on Paretojakaumien tunnusmerkki. Paretoeksponentti on siis samalla suoran kulmakerroin kun y- ja x-akseli vaihdetaan logaritmiseksi. Tämän matemaattisen ominaisuuden takia Potenssijakaumilla ei ole mitään tiettyä arvoa, jonka ympärille havainnot painottuvat. Tästä tulee nimitys kutsutaan mittakaavan invarianssiksi tai skaalautumattomuus. Ei ole olemassa tiettyä kaupungin kokoa, vaan ne vaihtelevat muutamien sadan asukkaan kylistä miljoonien ihmisten megakaupunkeihin.

(3) $$y=Cx^{b}$$

(4) $$\log y=\log C+b\log x$$

Parametri α:aa kutsutaan häntäparametriksi (eng. *tail index*) tai Paretoeksponentiksi, joka voi saada positiivisa arvoja väliltä (0,∞]. Tästä parametrista ollaan yleensä eniten kiinnostuneita, koska se määrittää jakauman hännän paksuuden eli kuinka suuria arvoja saadaan saa suhteessa koko jakaumaan. Tuloerojen tapauksissa α määrittää siis taloudellisen epätasa-arvon suuruuden. Mitä suurempi Paretoeksponentti α, sitä pienempi häntä eli tulo- tai varallisuuserot. Pareto-jakauman ensimmäinen momentti eli odotusarvo (5) riippuu Paretoeksponentin suuruudesta; kun α on alle 1 jakauman odotusarvo on ääretön. Tämä on mielenkiintoinen ominaisuus, joka voi johtaa harhaan ellei ole varovainen. Jakauman odotusarvon äärettömyys voi olla hankala intuitiivisesti ymmärtää. Tämän takia otoksen keskiarvo ei välttämättä anna oikeaa arvoa koko populaation odotusarvosta eli otoksesta voidaan aina laskea odotusarvo vaikkei se populaatiossa olisi ääretön. Tällöin suurten lukujen laki ei päde tämänkaltaisten jakaumien kanssa, koska otoskoon kasvaessa keskiarvo ei ikinä stabiloidu. Teoriassa on mahdollista saada äärettömän kokoinen havainto, joka nostaa koko otoksen keskiarvon äärettömäksi. Tulojakaumien Paretoeksponentti ei kuitenkaan ole alle 1, jolloin odotusarvo on äärellinen.

(5) $$E(x)=\begin{cases}\infty, & \alpha\leq 1\\ \dfrac{\alpha x_m}{\alpha-1}, & \alpha>1\end{cases}$$

Kaava (6) kertoo Paretojakauman mediaanin ja (7) moodin. Kuten paksuhäntäisen jakauman luonteeseen kuuluu, nämä luvut ovat pieniä. Paretoeksponentin ollessa 2, mikä on mielekäs arvio kaupunkien asukasluvun jakaumalle, mediaanikaupunki on vain √2 eli 1,41-kertainen pienimpään kaupunkiin verrattuna. Esimerkiksi Suomen mediaani kaupungissa on 6 066 asukasta. Moodi eli yleisin arvo Paretojakaumalle on sen minimi. Esimerkiksi voisimme jakaa kaupungit tuhannen asukkaan kokoluokkiin, jolloin eniten kaupunkeja olisi asukasmäärältään pienimmässä kokoluokassa.

(6) $$Mediaani=x_m\sqrt[\alpha]{2}$$

(7) $$Moodi=x_m$$

Kaava (8) kertoo Paretojakauman varianssin. Kuten ensimmäisen momentin tapauksessa, myös toinen momentti voi olla ääretön α:n ollessa alle 2. Tällöin ongelmat ovat samankaltaiset kuin äärettömän odotusarvon kanssa. Suurten lukujen laki toimii erittäin hitaasti, jolloin otoskoon on oltava tarpeeksi suuri varmistuakseen jakauman todellisesta luonteesta. Tämä koskee esimerkiksi varallisuuseroja ja joissain määrin myös tuloeroja. Taloudellisen epätasa-arvon tunnuslukuja mittaavat satunnaisotoksiin perustuvat epäparametriset testit voivat aliarvioida todellisen epätasa-arvon (Fontanari ym. 2018). Tutkimukset voivat perustua kotitalouskyselyyn, joka on satunnaisotos populaatiosta. Äärettömän varianssin tapauksessa otoskoon kasvaessa myös otoksen keskiarvo kasvaa. Laskut voivat osoittaa kuinka taloudellinen epätasa-arvo on kasvanut, vaikka todellisuudessa se on ollut aina siellä. Normaalijakautuneiden muuttujien kuten pituuden tapauksissa yleensä vaaditaan 30 havaintoa, jotta keskiarvo voidaan riittävän luotettavasti määrittää – Kadulla kävellessä otetaan 30 ensimmäisen ihmisen pituus ylös, jolloin voidaan olla suhteellisen luottavaisia ettei otoksen keskiarvo poikkea merkittävästi populaation keskiarvosta. Varallisuus- ja tuloerojen kanssa ei voi toimia näin, koska 31 vastaantulija voi olla mahdollisesti miljardööri. Tällöin otoksen keskiarvo hypähtää huomattavasti ylöspäin. Samaan ongelmaan voidaan törmätä yrittämällä laskea keskimääräinen massa tähdelle. Uusien havaintojen kasvaessa, todennäköisyys saada otokseen keskiarvoa heilauttava massiivinen tähti kasvaa.

$$(8) \qquad Var(x) = \begin{cases} \infty, & \alpha \leq 2 \\ \dfrac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}, & \alpha > 2 \end{cases}$$

Nykyään suosittu taloudellisen epätasa-arvon mitta on rikkaimman prosentin osuus kansantalouden tuloista tai varallisuudesta. Taleb varoittaa kuinka myös tämä tunnusluku on herkkä otoskoon kasvulle (2015). Taloudellinen epätasa-arvo voi näyttää pienemmältä valtiotasolla mitattuna kuin Euroopan tasolla mitattuna. Ratkaisuksi hän ehdottaa epäparametristen testien hylkäämistä ja suosimalla parametrisia testejä. Arvioimalla Paretoeksponentti, josta johdetaan tarvittava taloudellinen epätasa-arvo mittari, voidaan otoskoon aiheuttama vinoumalta välttyä. Yritysten kokojakaumaa arvioitaessa tutkijat havaitsivat saman Paretoeksponentin ja otoskoon välisen negatiivisen riippuvuuden (Segarra & Teruel 2012).

Monia eri ilmiöitä pystyy kuvaamaan Potenssijakaumiin perustuvilla malleilla ja näille ilmiöille on useita nimiä tilanteesta riippuen kuten *Yule prosessi*, *kumulatiivisen etu*, *Matteus-efekti, suosivan kiinnittyminen ja Rosenin supertähti -teoria,* joita käymme lyhyesti läpi. Robert K. Merton tutki ilmiötä, kuinka ansioituneet tutkijat saavat usein enemmän tunnustusta tutkimuksistaan kuin vähemmän tunnetut tutkijat. Hän osoitti, kuinka tutkimuksista palkitaan yleensä sen vanhemmat tutkijat, eikä nuorempia, vaikka suurin työpanos oli nuoremmilla tutkijoilla (Merton 1968). Mertonin mukaan maineella on vaikutus monimutkaisiin sosiaalisiin valintaprosesseihin, mikä johtaa resurssien ja taitojen keskittymiseen. Stephen Stigler vei ajatusta eteenpäin; yhtäkään tieteellistä keksintöä ei nimetä alkuperäisen keksijänsä mukaan. Stigler oli saanut ajatuksen

Mertonilta, ja ironisesti havainto sai myöhemmin nimekseen Stiglerin laki (Gieryn 1980). Myös muissa tutkimuksissa löydettiin useita havaintoja tiedemaailman vinoutuneisuudesta menestyneiden hyväksi. Tutkijat suurissa yliopistoissa saivat enemmän huomiota samalla julkaisumäärällä, kuin tutkijat pienemmissä yliopistoissa (Crane 1965). Amerikkalaisista Nobel-voittajista 49 % työskenteli viidessä (Harvard, Rockefeller, Berkeley, Chicago ja Columbia) yliopistossa (Zuckerman 1977), ja 28 % valtion tuista yliopistoille meni kymmenelle yliopistolle (Merton 1988). Nämä tapahtumat ruokkivat itse itseään. Yliopistot, joilla on hyvä maine, saavat suhteessa enemmän lahjakkaita oppilaita ja resursseja, joilla ne voivat helpommin pysyä tieteellisen kehityksen kärjessä.

Robert K. Merton kuvaili ensimmäisenä tätä tiedemaailmassa tapahtuvaa ilmiötä nimellä *Matteus-efekti* lainaten opetusta Raamatun Matteuksen evankeliumista. ”Jokaiselle, jolla on, annetaan, ja hän on saava yltäkyllin, mutta jolla ei ole, siltä otetaan pois sekin mitä hänellä on.” (Matt. 25:29). Hyvänä esimerkkinä itseään vahvistavasta mekanismista sosiaalisessa kontekstissa toimii Keith A. Stanovichin tutkimus, jossa hän selvitti lasten lukutaitoeroja. Hän huomasi kuinka pienet erot lukutaidossa kasvoivat oppimisen myötä, jota hän kutsui myös Matteus-efektiksi lainaten Mertonin havaitsemaa ilmiötä tiedemaailmasta. Kun lapset aloittavat opettelemaan lukemista, ovat erot yleensä pieniä, mutta nämä erot kasvavat muutamassa vuodessa suuremmiksi. Lapset, jotka oppivat lukemaan nopeammin kehittyvät myös muissa aineissa paremmin. Toisin sanoen hyvän lukutaidon omaavat lapset lukevat jo oppiakseen verrattuna lapsiin, jotka oppivat vielä lukemaan (Stanovich 1986). Tämänkaltaisten oppimiserojen tilastollinen tarkastelu voi olla hankalaa, mutta samankaltainen mekanismi toimii rikollisuudessa, jossa potenssijakauma löydettiin rikosten määrän jakautumisessa (Cook ym. 2004). Heillä oli kaksi eri aineistoa; 14 vuoden seurantatutkimus Iso-Britannian nuorten rikostuomioiden määrästä sekä puolen vuoden ajalta itseraportoidut rikoksien määrä nuorten keskuudessa Yhdysvaltalaisessa kaupungissa. Molemmista he havaitsivat saman suuruisen Paretoeksponentin, vaikka aineistot olivat laadullisesta hyvin erilaisia. Tästä tuloksesta tutkijat päättelivät kuinka vanhat rikokset kasvattivat uuden rikoksen todennäköisyyttä, joten ennaltaehkäisy varhaisessa vaiheessa on erityisen tärkeää. On helppo intuitiivisesti ymmärtää kuinka tavallisella ihmisellä on suuri kynnys tehdä rikos toisin kuin taparikolliselle.

Matteus efektin kaltaista rikkaat rikastuu -ilmiötä tarkasteltiin matemaattisesti ensimmäisenä tilastotieteilijä Udny Yulen artikkelissa *Evoluution matemaattinen teoria perustuen Dr. J. C. Willisin johtopäätöksiin* (1925). Artikkelissa hän kertoo, kuinka stokastiset prosessit, eli satunnaiset tapahtumat, johtavat Potenssijakaumiin. Artikkelin nimen mukaisesti hän tutki asiaa evoluution näkökulmasta. Hän kuvaili matemaattisen mallin avulla, kuinka eläinlajien lukumäärät sukujen sisällä jakautuvat ajan kuluessa epätasaisesti ympäristön muutoksen johdosta. Näiden stokastisten prosessien satunnaisuuksien takia jakauman häntä asymptoottisesti lähenee Potenssijakaumaa. Tästä alettiin myöhemmin käyttämään nimitystä *Yule prosessi* tutkija Herbert Simonin toimesta. Simon vei Yulen ajatusta pidemmälle ja tutki muita Potenssijakaumalla kuvattavia tilanteita Yule prosessia mukaillen; sanojen esiintymistiheyttä teksteissä, tiedemiesten julkaisumäärien jakautumista ja kaupunkien asukasmäärien jakaumaa. Hän myös todisti matemaattisesti, kuinka Yule prosessi johtaa Potenssijakaumiin (Simon 1955). Simonin kontribuution johdosta tätä jakaumaa kutsutaan Yule-Simon jakaumaksi.

Herbert Simonin tutkimuksista inspiroituneena Derek J. de Solla Price tutki tieteellisten tutkimusten välisiä sitaattiverkostoja ja huomasi, kuinka ne noudattavat Yule prosessien mukaista jakaumaa: vain muutamat harvat tutkimukset tietyllä alalla keräsivät suurimman osuuden kaikista tutkimuksien välisistä sitaateista (Price 1965). Hän kuvaili tutkimuksissaan, kuinka tunnettavuus ruokkii menestystä entisestään tiedemaailmassa, mutta erotuksena Mertoniin hän käytti termiä *kumulatiivinen etu*. Myöhemmin verkkoteorian tutkijat Albert-László Barabási ja Réka Albert tutkivat internet-verkon ominaisuuksia. He löysivät kaksi selitystä liittyen siihen, miksi internetsivujen verkostot – Pricen tiedemaailman havainnon tavoin – noudattavat potenssijakaumaa. Heidän mukaansa skaalautumaton verkosto kasvaa jatkuvasti uusilla solmuilla eli verkkosivuilla ja ne kiinnittyvät yleensä niihin, jotka ovat jo hyvin yhteyksissä muihin solmuihin (Barabási & Albert 1999). Kumulatiivisen edun sijasta tutkijat käyttivät termiä *suosiva kiinnittyminen (*eng. *preferential attachment),* joka kuvailee yleisesti, kuinka jokin suure jakautuu yksiköiden kesken riippuen siitä, kuinka paljon yksiköillä jo entuudestaan on. Barabásin ja Albertin mukaan nämä internet-verkon ominaisuudet voidaan yleistää myös muihin skaalautumattomiin verkostoihin.

Diskreettinen erityistapaus Paretojakaumasta on *Zipfin Laki.* Lain teki tunnetuksi yhdysvaltalainen kielitieteilijä George Kingsley Zipf. Hän tutki englannin kielen sanojen esiintymistiheyttä ja huomasi, kuinka englannin kielen yleisin sana ”the” esiintyi teksteissä kaksi kertaa useammin kuin toiseksi yleisin sana ”of” ja kolme kertaa useammin, kuin kolmanneksi suosituin sana ”and” (Zipf 1935). Tätä sanojen esiintymistiheyden ja järjestyksen potenssilain mukaista suhdetta hän nimitti *järjestys vs. esiintymistiheys -säännöksi*, josta myöhemmin alettiin käyttää nimitystä Zipfin laki. Zipf jatkoi tutkimuksiaan ja huomasi, kuinka laki voidaan yleistää koskemaan myös rikkaimpien tulojakaumaan eri maissa (Zipf 1941). Vaikka lähes kaikki kielet noudattavat Zipfin lakia, sen syntymekanismeja ei olla aukottomasti pystytty selittämään. Zipf itse ehdotti pienemmin vaivannäön periaatetta. Tämä poikkitieteellinen konsepti selittää kuinka eri systeemit optimoivat energian kulutustaan ajan kuluessa ja kielten tapauksessa kielen puhuja yrittää kommunikoida käyttäen mahdollisimman vähän sanoja. Tämä konsepti muistuttaa Hayekin spontaania järjestystä, jossa hän korosti nimenomaan kuinka markkinat jakavat resursseja tehokkaasti ilman suunnittelua (1978).

Kaupunkien voidaan ajatella olevan ihmisten välisiä verkostoja, joissa kaupungit ”kilpailevat” asukkaista. Saksalainen fyysikko Felix Auerbach huomasi ensimmäisenä, kuinka Saksan kaupungit asukasluvultaan suuruusjärjestykseen asetettuna noudattavat potenssilain mukaista jakaumaa (Auerbach 1913). Ilmiön myöhemmin teki tunnetuksi Zipfin tutkimukset. Kaupunkien asukasmäärä noudattaa Zipfin lakia hämmästyttävän hyvin useissa eri maissa. Ranskalaisen ekonomistin Xavier Gabaixin tutkimuksessa (Gabaix 1999) havainnoitiin kuinka USA:n 135 suurinta kaupunkia suuruusjärjestyksessä asettuvat lineaarisesti suuruusjärjestykseen, kun käytetään logaritmisia akseleita. Zipfin lain soveltamista kuitenkin vaikeuttaa käsitteen *kaupunki* määrittely; epävarmuutta löytyy mitä tulee asukasluvultaan pienempiin asutuskeskuksiin sekä yhteen kasvaneiden kaupunkien muodostamiin metropolialueisiin.

Kaupunkien kaltainen potenssijakauma esiintyy myös talouden puolella. Robert L. Axtell huomasi, kuinka yritysten koot, kaupunkien tapaan, noudattavat Potenssijakaumaa (Axtell 2001). Yrityksen koko voidaan määritellä työntekijöiden tai liikevaihdon mukaan, mutta molemmat versiot sopivat Potenssijakauman mukaiseen kuvaukseen. Myös markkinat itsessään pystyvät ylläpitämään vain tietyn verran yrityksiä. Axwell sovelsi Zipfin jakaumaa kaikkien yhdysvaltalaisten yritysten kokoon, mutta hän ei huomioinut miten tiettyjen markkinoiden sisällä yritysten koot vaihtelevat. Boston Consulting Groupin perustaja Bruce Henderson ehdotti, että kypsät markkinat pystyvät ylläpitämään kolmea isoa tuottajaa kerrallaan ja loput tuottajat keskittyvät pienempiin niche-markkinoihin (Boston Consulting Group 1976). Hänen mukaansa kolmen suurimman yrityksen koot yleensä ovat 4:2:1 suhteessa.

Tutkimuksia jatkoivat Sheth ja Sisodia, jotka alkoivat käyttää nimitystä *kolmen sääntö* havaitessaan useampien markkinoiden koostuvan kolmesta dominoivasta yrityksestä. Heidän mukaan markkinat kehittyvät kohti ennustettavaa tasapainoa, jossa uudet ja vanhat yritykset seuloontuvat ja lujittavat markkinoilla. He havainnoivat kolmen sääntöä ottamalla esimerkiksi ostoskeskuksen, jossa toimii kolme suurta yleistavarakauppaa ja lisäksi paljon muita kauppoja, jotka keskittyvät pienempiin markkinoihin eivätkä yritä suoranaisesti kilpailla isojen kanssa. (Sheth & Sisodia 2002) Kolmen sääntö sai lisää empiiristä vahvistusta tuoreemmasta tutkimuksesta, jossa tutkittiin 160 yhdysvaltalaisen teollisuusalaa (Uslay ym. 2010). Markkinat voidaan ajatella olevan metsän kaltainen ekosysteemi, joka koostuu lukuisista kilpailevista heterogeenisistä yksiköistä. Markkinoiden voimana toimii voiton etsiminen ja luonnon ekosysteemit toimivat luonnonvalintaan pohjautuva eloonjäämiskamppailu. Amazonin sademetsällä on hämmästyttävän samankaltainen ominaisuus kaupunkien ja markkinoiden kanssa tältä osin. Amazonin 16 000 tunnetusta puulajista, vain 227 eli 1.6% lajeista vastaa puolista sademetsän puista (Ter Steege ym. 2013). Tutkijat kutsuivat näiden puulajien suurta yliedustusta hyperdominanssiksi.

Tästä hyperdominanssin käsitteestä päästään siirtymään viimeisen potenssilain synnyttämään ilmiöön ihmisten puolella – *Supertähtiin*. Vilfredo Pareton empiirisistä tutkimuksista lähtien suhteellisen nopeasti päästiin yksimielisyyteen rikkaimpien tulojen potenssijakaumasta. Kuitenkaan tämä ei vielä kerro syitä miksi näin tapahtuu. Sherwin Rosen tutkimuksissaan etsi syitä näiden supertähtien ylivertaisuuteen. Hänen mukaan on olemassa markkinoita, joilla pieni osa markkinoiden tuottajista saa suurimman osan markkinoiden tuloista. Hän viittasi markkinoilla urheilu- ja musiikkimarkkinoihin, joilla supertähdet dominoivat markkinoita. Hänen mukaansa pienet erot lahjakkuudessa johtavat tilanteeseen, jossa markkinoiden tuottamat tulot vinoutuvat voimakkaasti parhaimpien hyväksi (Rosen 1981). Rosenin mukaan näiden markkinoiden tuottajat olivat keskenään heikosti substituoituja eli yhden supertähden tuotos oli heikosti korvattavissa vähemmän suosittujen tuottajien tarjonnalla; yhden supertähden konsertti ei voi korvata kolmen vähemmän tunnetun konsertilla.

Toinen selittävä tekijä on teknologian kehitys. Esimerkiksi televisio mahdollisti supertähtien tuottamien palveluiden monistamisen yhä useammille kuluttajille yhä alhaisemmilla kustannuksilla. Taloustieteen yrityksen teoriassa voiton maksimoimiseksi työpanosta kasvatetaan kunnes rajatulo vastaa rajakustannusta. Rajakustannus on markkinoilla määräytyvä työn yksikkökustannus, joka on yrityksen kannalta vakio. Tämä selittää miksi parhaimpien teatterinäyttelijöiden palkat eivät ole lähelläkään elokuvanäyttelijöiden palkkaa, koska teknologian kehitys on mahdollistanut suuremmat elokuvamarkkinat kuin teatterimarkkinat. Näin tuottajien väliset pienet erot lahjakkuudessa synnyttävät suuret erot tuloissa. Sama ilmiö on havaittavissa suoratoistopalveluiden yleistyessä. Ennen tv-sarjoja katsottiin vain televisiosta jolloin tv-sarja-markkinat eivät voineet kasvaa tv:n rajallisten kanavapaikkojen takia. Suoratoistoteknologian kehittyessä tv-sarja voidaan jakaa globaalisti miljoonille kuluttajille ja kuluttajat voivat katsoa ne haluamaan aikaan. Tämä teknologian tuoma tv-sarja-markkinoiden kasvu on lisännyt eturivin elokuvatähtien siirtymistä valkokankaalta tähdittämään televisiosarjoja. Krueger kirjassaan pohtii talouden eri alojen kehittymistä kohti tämäntapaisia markkinoita ottaen esimerkkejä musiikkiteollisuudesta kirjassaan Rockonomics (2019).

Supertähtien ilmiö on laaja-alainen ilmiö, jossa supertähteyteen vaikuttavia tekijöitä on useita. Moshe Adler jatkoi Rosenin älkeen tutkimista muutamaa vuotta myöhemmin. Hän analysoi kuinka lahjakkuuden lisäksi maine ja tunnettavuus kasvattavat eroja tuottajien välillä, vaikka he olisivat yhtä lahjakkaita. Adlerin mukaan supertähteys-ilmiö vaatii kuluttajilta tietoa. Tämä kuluttajien tieto supertähdistä synnyttää keskustelua kuluttajien kesken siitä kuka on supertähti. Tämä kuluttajien välinen dialogi johtaa yhteiseen päätökseen, joka kanavoituu tietyn supertähden suosioon (Adler 1985). Tapaus muistuttaa hyvinkin paljon Matteus-efektiä, jossa maineella on voimistava vaikutus kuluttajien päätökseen.

Rosenin mukaan juuri markkinoiden kasvaminen teknologian kehittymisen myötä kasvattaa parhaimpien tuloja. Suurten markkinoiden kuluttajat arvostavat eniten supertähtiä ja kilpailu supertähdistä kasvaa, mikä nostaa näiden palkkoja. Samantapaisia tuloksia löysivät Xavier Gabaix ja Augustine Landier yritysmaailmasta. Työmarkkinoiden supertähtien eli suurten yritysten toimitusjohtajien palkat ovat suoraan verrannollisia yritysten kokoon. Palkkaeroa ei enää selittänyt toimitusjohtajien väliset erot taidossa. Suurimpien yhdysvaltalaisten yrityksien markkina-arvojen kuusinkertaistuminen vuosien 1980 ja 2003 välillä kasvatti toimitusjohtajien palkkaa kuusinkertaiseksi Yhdysvalloissa (Gabaix & Landier 2008). Tutkijat muistuttivat, että pelkkä yrityksen markkina-arvon koko ei välttämättä nosta toimitusjohtajan palkkaa samassa suhteessa. Toiseksi selittäväksi syyksi he ehdottivat *tarttumisefektiä*. Jos pieni osa yrityksistä maksaa toimitusjohtajalle esimerkiksi hinnoitteluvirheen seurauksena enemmän palkkaa kuin muut, kaikkien toimitusjohtajien palkat nousevat kun työmarkkinoiden tasapaino muuttuu. Tämä kilpailu toimitusjohtajista, joiden työpanos on suuri yrityksen tuloksen kannalta, johtaa yritysten väliseen palkkauskilpailuun parhaimmista toimitusjohtajista samaan tapaan kuin elokuvatuottajat kilpailevat parhaimmista näyttelijöistä työmarkkinoilla.

Edellä mainittujen mekanismien kautta voi nähdä jo selkeän yhteyden makrotasolta yksilötasolle. Markkinoiden koon vaikutus yrityksien kokoon ja yrityksien koon vaikutus yksilöiden palkkaan. Näillä kahdella yhteydellä tärkeä implikaatio tuloerojen kehityksen kannalta. Globalisaation kasvaessa kapitalististen prosessien kautta kansalliset markkinat rikkoutuvat ja ne liikkuvat kohti globaalia yhteismarkkinoita. Näin kansalliset yritykset muuttuvat joko monikansalliseksi tai fuusioutuvat keskenään. Joka tapauksessa markkinoiden kasvun takia niissä toimivien yrityksien koko kasvaa ja tämä nostaa toimitusjohtajien ja muiden supertähtien palkkoja.

Edellä on käytä läpi pintapuolisesti potenssijakauman matemaattisia ominaisuuksia ja niiden ilmentymistä eri tilanteissa. Yksinkertaisesti voitaisiin tiivistää, että potenssijakauma syntyy tilanteessa, jossa jokainen yksikkö lisää kasvattaa todennäköisyyttä saada tulevaisuudessa lisää. Tämä lumipallo -efekti alkaa erityisesti toimimaan suurilla muuttujan arvoilla, jolloin suuret arvot ”dominoivat”. Selittävät tekijät ihmisten palkkaeroihin vaihtelevat tilanteesta toiseen, mutta lopputulos on sama – rikkaat rikastuvat.

3.2. Lognormaalijakauma

Lognormaalijakauma pohjautuu nimensä mukaisesti normaalijakaumaan; Jos X on lognormaalijakautunut, niin Y = ln X on normaalijakautunut. Normaalijakauma syntyy monien riippumattomien satunnaismuuttujien summasta ja vastavuoroisesti lognormaalijakauma syntyy monien satunnaismuuttujien tulosta. Tätä tulosta kutsutaan *keskeiseksi raja-arvolauseeksi,* joka on todennäköisyysteorian peruspilareita. Ihmisten pituuteen vaikuttaa lukuisia tekijöitä geeneistä ympäristöön ja kaikkien näiden selvittäminen on haastavaa, ehkä mahdotonta. Se on kuitenkin selvää, että näiden tekijöiden summa synnyttää ihmispopulaation pituuden normaalijakauman. Normaalijakauma syntyy vaikka sen rakentamat satunnaismuuttujat eivät ole normaalijakautuneita. Voimme heittää arpakuutiota 10 kertaa, jolloin odotusarvo tälle on E(arpakuutio) x 10 = 3.5 x 10 = 35. On helppo hahmottaa kuinka tätä koetta toistamalla summat pakkautuvat odotusarvon ympärille tasaisesti. Todennäköisyys saada 10 kertaa peräkkäin 1 tai 6 on pieni mutta yhtä suuri synnyttäen normaalijakauman.

Tämä sama keskeisen raja-arvolauseen mukainen prosessi synnyttää lognormaalijakauman kun osatekijöiden tulon seurauksena. Kuten normaalijakauma myös lognormaalijakauma määritellään sijaintiparametrilla μ ja muotoparametrilla σ^2. Normaali- ja lognormaalijakauman sukulaisuus huomataan myös samankaltaisesta tiheysfunktiosta (9) ja kertymäfunktiosta (10), jossa normaalijakauma on mukana parametreilla $\sigma^2 = 1$ ja $\mu = 0$.

(9) $$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$

(10) $$\frac{1}{2}+\frac{1}{2} erf\left[\frac{\ln x-\mu}{\sqrt{2}\sigma}\right]$$

Normaalijakauman odotusarvon ja mediaanin on yhtä suuri kuin sijaintiparametri μ, mutta lognormaalijakauman tapuksessa hajonta σ^2 vaikuttaa lognormaalin odotusarvoon (11). Monien satunnaismuuttujien tulo saa aikaan hännän ”karkaamisen” oikealle, mikä taas nostaa odotusarvoa. Tämän takia esimerkiksi tuloeroja arvioitaessa on huomataan kuinka suurin osa palkansaajista saa vähemmän tuloja kuin keskipalkka.

(11) $$E(x)=e^{\left(\mu+\frac{\sigma^2}{2}\right)}$$

Tästä syystä lognormaali jakauma on siinä mielessä erikoinen, että se on välimuoto paksuhäntäisen ja lyhythäntäisten jakaumien välissä. Kun lognormaalijakauman muotoparametri σ^2 kasvaa tarpeeksi suureksi, se käyttäytyy hännän osalta kuten Paretojakauma. Joskus näiden jakaumien erottaminen toisistaan aineiston perusteella on hankalaa (Clauset ym. 2009). Oikean matemaattisen mallin valitseminen on ollut tämän haasteen takia kiinnostuksen kohteena viime aikoina (Benguigui & Marinov 2015). Tämä lognormaali- ja Paretojakauman väliltä valitseminen on ollut monien tieteenalojen ongelma kuten esimerkiksi tiedostokokojen mallintamisessa. Tietojärjestelmätieteilijä Mitzenmacher käy läpi tutkimuksessaan kirjallisuutta, joissa pohditaan näiden jakaumien eroja ja käyttökelpoisia tilanteita eri tieteenaloilta (2004).

Lognormaalijakauman mediaaniin hajonta ei vaikuta (12), joten sen paikantamiseen tarvitaan vain μ-parametri. Yleisin arvo eli moodi (13) sijaitsee vastavuoroisesti matalammalla kuin mediaani. Tällöin yleisin vuosipalkka on pienempi kuin tulojakauman keskimmäinen vuosipalkka. Normaalijakauman varianssin määrittää luonnollisesti suoraan hajonta σ^2, mutta lognormaalin tapauksessa mukaan tulee vielä μ-parametri (14).

(12) $$Mediaani=e^{\mu}$$

(13) $$Moodi=e^{\mu-\sigma^2}$$

(14) $$Var(x)=(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$

Lognormaalijakauma nousi ykköskandidaatiksi kuvaamaan keski- ja pienituloisia Gibratin tutkimuksien jälkeen. Roy teoretisoi lognormaalisuuden syitä. Hän ehdotti kuinka työntekijän tulot vastaavat ihmisen tuottavuutta, joka vastavuoroisesti rakentuu työntekijän monien ominaisuuksien tulosta synnyttäen lognormaalijakauman. Lognormaalijakaumia on löydetty taloustieteen lisäksi useissa eri luonnontieteissä kuten geologiassa, epidemiologiassa, ympäristötieteissä, mikrobiologia ja kielitieteissä. Lognormaalijakauman yksinkertaisen syntymekanismin takia sitä esiintyy useissa systeemeissä luonnossa (Limpert ym. 2001). Täten sitä voidaan tästäkin syystä pitää luontaisena kandidaattina kuvaamaan tuottavuutta ja täten palkkaa. Roy oli oikeilla jäljillä, mutta tutkijat ovat jo kauan aikaa ymmärtäneet kuinka lukuisat työntekijästä riippumattomat tekijät vaikuttavat hänen palkkaansa. Toinen kysymys on pohtia voiko edes osoittaa ihmisten tuottavuuden ja siten palkan olevan lognormaalijakautunut.

Yksi tapa on unohtaa markkinoilla määräytyvä palkka ja katsoa muita tilanteita, joissa ihmiset tuottavat jotain. Fysiikan Nobelisti William Shockley huomasi kuinka tutkimuslaboratorioiden tutkijoiden julkaisumäärän logaritmi oli normaalijakautunut – tutkimusmäärä per tutkija noudatti lognormaalijakaumaa (Shockley 1957). Hän pohti tutkijoiden ominaisuuksia, jotka vaikuttivat tämän tuottavuuteen ja tarjosi 8 tutkijoille tärkeää ominaisuutta; kyky ajatella hyvää ongelmaa, kyky työskennellä sen kanssa, kyky tunnistaa arvokas tulos, kyky tehdä päätös siitä milloin lopettaa ja kirjoita tulokset, kyky kirjoittaa riittävästi, kyky hyötyä rakentavasti kritiikistä, päättäväisyys toimittaa paperi tieteelliseen julkaisuun ja päättäväisyys muutosten tekemisessä. Jos yksi tutkija on 50% keskiarvon yläpuolella jokaisissa ominaisuuksissa, on hän $1.5^8 \approx 25,6$x tuottavampi. Tämä tulos antaa viitteitä siitä, että on mielekästä pitää tuottavuutta on lognormaalijakautuneena. Tällöin täydellisessä maailmassa ihmisten palkka tulisi perustua tähän. Shockley vielä vertasi tutkijan julkaisumääriä palkkaan ja huomautti kuinka palkka ei pysy tuottavuuden perässä. Tutkijan on nostettava tuottavuuttaan 30-50% saadakseen 10% palkankorotuksen. (Shockley 1957). Tämä on tärkeä huomio, jonka myös Roy myös ymmärsi. On syytä pohtia tarkkaan vastaako markkinoilla määräytyvä palkka oikeasti ihmisen tuottavuutta.

3.3. Eksponentti- ja gammajakauma

Lognormaalin kilpailijaksi tuloja kuvaamaan on ehdotettu eksponentti- tai gammajakaumaa. Ne molemmat kuuluvat samaan todennäköisyysjakaumaperheeseen ja molemmat pohjautuvat *Poisson-prosessiin*. Tämä tarkoittaa satunnaisten tapahtumien syntymistä ajan kuluessa, jossa tapahtumat ovat keskenään riippumattomia.

Eksponenttijakauma määritellään vain yhdellä asteparametrilla λ (eng. *rate*). Joskus voidaan myös käyttää lambdan tilalla sen käänteislukua, jolloin sitä kutsutaan skaalaparametriksi θ (eng. *scale*). Eksponenttijakauman tiheys- (15) ja kertymäfunktiot (16) ovat yksinkertaisia mikä tekee siitä oivan työvälineen monenlaiseen analyysiin. Huomioitavaa on sen samankaltaisuus potenssijakauman kanssa. Eksponenttijakaumassa muuttuja x on nimensä mukaisesti eksponentissa kun taas Paretojakaumassa parametri on eksponentissa.

Eksponenttijakaumaa käytetään elinaika-analyysissä. Sen ominaisuuksia voidaan tarkastella kuvitteellisen metsän avulla. Metsän puut koostuvat yhdestä puulajista, jonka lajityypillinen vuodenkierron selviämistodennäköisyys on 98 %. Toisin sanoen 2 % puista kuolee vuodessa. Eksponenttijakaumalla on muistittomuusominaisuus, joka tarkoittaa ettei menneisyyden arvoilla ole vaikutusta. Tällöin vanhalla puulla on yhtä suuri kuolemisen todennäköisyys kuin nuorella puulla ja vanhat puut ovat vain onnekkaimpia. Esimerkissä λ-parametri on 2 %, josta voidaan laskea tämän kuvitteellisen metsän tilastollisia tunnuslukuja. Tiheysfunktio kertoo meille puiden ikäjakauman ja kertymäfunktio kertoo tietyn ikäisten puiden osuuden metsästä. Asettamalla λ = 0.02 kaavaan (16) saamme alle 100 vuotaiden puiden osuuden; alle 100 vuotiaita puita on $1-e^{-0.02\cdot 100}\approx 0.86$ metsän puista.

(15) $$\lambda e^{-\lambda x}$$

(16) $$1-e^{-\lambda x}$$

Eksponenttijakauman odotusarvo on asteparametrin käänteisluku (17) ja mediaani on luonnollinen logaritmi kahdesta jaettuna asteparametrillä (18). Odotusarvo puun elinajalle on 0.02^{-1} eli 50 vuotta. Iältään keskimmäinen puu metsässä on $(\ln 2)^{-0.02}$ eli 34.7 vuotias.

(17) $$E(x)=\lambda^{-1}$$

(18) $$Mediaani = \frac{\ln 2}{\lambda}$$

Viimeiset tunnusluvut ovat moodi (19) ja varianssi (20). Moodi eli yleisin puun ikä on luonnollisesti 0, koska kuoleman todennäköisyys ei ole ehtinyt vielä vaikuttamaan puuhun syntymähetkellä.

(19) $$Moodi = 0$$

(20) $$Var(x) = \frac{1}{\lambda^2}$$

Eksponenttijakauman on kuitenkin monotonisesti laskeva funktio, jolloin sillä voidaan kuvata tuloja vain kertymäfunktiolla. Tulojakaumassa on moodin kohdalla kyttyrä, jonka alle jää huomattava osa tuloista. Tästä syystä tulojakauman tiheysfunktiota voidaan kuvata gammajakaumalla, joka määritellään kahdella parametrilla – muotoparametrilla k ja skaalaparametrilla θ. Skaalaparametrin tilalla käytetään sen käänteislukua β (asteparametri) kuten eksponenttijakauman tapauksessa. Silloin β:n kanssa muotoparametri merkitään α:lla pysyen muuten ennallaan.

Gammajakauman tiheys- (21) ja kertymäfunktiot (22) määritellään gammafunktiolla (23), josta jakauma on saanut nimensä. Kun yhtälössä (23) n = 1, gammafunktio saa arvon 1. Tällöin gammajakauman tiheysfunktio (21) muuttuu muotoon $\beta e^{-\beta x}$, jolloin se on sama kuin eksponenttijakauman tiheysfunktion yhtälö (15). Eksponenttijakauma on siis gammajakauman erityistapaus kun muotoparametri saa arvon 1. Eksponenttijakauma kertoo todennäköisyyden tapahtumalle X ja gammajakauma kertoo todennäköisyyden kun X on tapahtunut n-kertaa. Tästä syystä eksponenttijakaumien summa on gammajakautunut.

(21) $$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad tai \quad \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

(22) $$\frac{1}{\Gamma(k)} \gamma(k,\theta) \quad tai \quad \frac{1}{\Gamma(\alpha)} \gamma(\alpha,\beta)$$

(23) $$\Gamma(n) = (n-1)!$$

Gammajakauman odotusarvo (24) on yksinkertainen, mutta mediaanille ei ole yksiselitteistä ratkaisua.Yhtälö (25) kertoo gammajakauman moodin ja yhtälö (26) varianssin.

(24) $$E(x) = k\theta \quad tai \quad E(x) = \frac{\alpha}{\beta}$$

(25) $$Moodi = (k-1)\theta \quad tai \quad Moodi = \left(\frac{\alpha-1}{\beta}\right)$$

(26) $$Var(x)=k\theta^2 \quad tai \quad Var(x)=\frac{\alpha}{\beta^2}$$

Ekonofyysikkojen esikuva tulojen malliksi on nimeltään Boltzmann-Gibbs jakauma (27). Se on eksponenttijakauma, joka kuvaa kaasumolekyylien energian E jakautumista lämpötilassa T. Toisin P(E) kertoo suljetussa systeemissä yksittäisen kaasumolekyylin energian E:n todennäköisyyden vakiolämpötilassa T. Eksponenttijakauman ominaisuuksien mukaisesti kaasumolekyylin energiatilan löytämisen todennäköisyys pienenee mitä suurempi energia sillä on. Suljetussa systeemissä suurimmalla osalla kaasumolekyyleistä on matala energia ja harvalla suuri.

(27) $$P(E)=ce^{-\frac{E}{T}}$$

Tulojen tiheysfunktion kuvaamiseen tarvitaan gammajakauma pitääkseen tulojen yhteyden termodynamiikkaan järkevänä. Tähän soveltuu Maxwell–Boltzmann-jakauma (28). Maxwell–Boltzmann jakauma kertoo kaasumolekyylien kineettisen energian, josta voidaan johtaa kaasumolekyylien nopeuksien jakauma. Tämä jakauma voidaan kirjoittaa myös gammajakauman muodossa asettamalla kT = θ ja 3/2 = k.

(28) $$f(E)=2\sqrt{\frac{E}{\pi}}\left(\frac{1}{kT}\right)^{3/2}\mathrm{e}^{-\frac{E}{kT}}$$

Sosiaalisten ilmiöiden ja termodynamiikan yhdistäminen voi vaikuttaa kaukaiselta teorialta, mutta ekonofyysikot muistuttavat ettei heidän nykyisessä työssään ole mitään uutta. Termodynamiikan pioneeri Gibbs kehitti teoriaansa molekyylien jakautumisesta ottamalla mallia sosiaalisista ilmiöistä (Drăgulescu & Yakovenko 2002). Lisäksi heidän teoriansa perustuu tulojakauman empiirisiin havaintoihin, joiden selittämiseen tarvitaan teoria. Taloustieteilijöillä itsellään ei ole yleisesti hyväksyttyä teoriaa miksi ansiotulot olisivat gamma- tai eksponenttijakautuneet. Ihmisten lognormaalisti jakautunut tuottavuus ei edes pysty teknisesti selittämään tulojen eksponentti- tai gammaluonnetta. Olisi myös outoa, jos markkinoiden häiriöt tuottaisivat gamma- tai eksponenttijakauman kaltaisia säännönmukaisia makrotason ominaisuuksia. Taloustieteilijät ovat tässä mielessä jättäneet hyvin tilaa fyysikoiden omalle teoria.

Tämä termodynaaminen näkökulma tuloeroihin on saanut kuitenkin kritiikkiä sen kunnollisen taloustieteellisien perustelujen puutteen takia (Cho 2014). Vastareaktiona kritiikille tutkijat ovat myöhemmin esittäneet empiiristen havaintojen selittämiseksi perinteiseen taloustieteeseen pohjautuvia teorioita. Yksi teoria tulojen eksponentti- tai gammarakenteen ja yhteyden termodynamiikkaan selittää *Arrow–Debreun yleisen tasapainon -malli* yhdistettynä *Rawlsin oikeudenmukaisuuden periaatteeseen* (Tao ym 2019). Arrow–Debreu-malli on yleisen tasapainoteorian (eng. *general equilibrium theory*) ja uusklassisen taloustieteen kulmakivi. Kenneth

Arrow ja Gérard Debreu todistivat matemaattisesti markkinoiden tehokkuuden tiettyjen tiettyjen ehtojen vallitessa (1954). Tässä Arrow-Debreu -mallissa agentit ovat samalla yrityksiä, jotka tuottavat hyödykkeitä toisille agenteille maksimoiden kulutuksensa ja ostavat hyödykkeitä maksimoiden hyvinvointinsa. Rawlsin oikeudenmukaisuuden periaate takaa kaikilla agenteilla olevan yhtäläinen mahdollisuus osallistua tähän resurssien tuottamiseen ja kuluttamiseen. Näihin oletuksiin pohjautuen ekonofyysikot todistivat kuinka tälläisessä mallissa tulojen kertymäfunktio noudattaa eksponenttijakaumaa, käyttäen tuttua entropian maksimointi tekniikka. Tämän jälkeen tutkijat osoittivat kuinka pieni- ja keskituloisten eksponenttijakauma löytyy 67 eri maasta. He eivät kuitenkaan käytä tilastollisia testejä varmistuakseen tulojen eksponenttiluonteesta vaan tyytyvät graafiseen tarkasteluun.

Toisenlaisen teorian ekonofyysikoiden empiiristen havaintojen selittämiseksi antoi Anwar Shaikh kirjassaan *Kapitalismi, Konflikti, Kilpailu ja Kriisit* (2016). ”*Vaikka tämä [termodynamiikan yhteys rahan vaihdannan mallintamiseen] voi hyvinkin tarjota hedelmällisen lähestymistavan rahavarojen mallintamiseen sosiaalisessa kontekstissa, se ei kuitenkaan koske työn palkkatulojen jakautumista*”. Hänen mukaansa teknologian kehittymisen takia eri yrityksissä on eri teknologian taso, joka saa aikaan kysynnän tietyille töille. Tämän takia yrityksien pitää jatkuvasti päivittää teknologian tasoaan säilyttääkseen voittonsa ja markkina-asemansa. Työntekijöiden liike yrityksien välillä tasoittaa työpaikkojen palkkaerot ammatin sisällä, jolloin yritys ei voi maksaa liian vähän palkkaa verrattuna muihin yrityksiin. Työpanoksen hinnan määrittää ammatin tuottavuus yritykselle, mutta samalla työntekijöiden neuvotteluvoima. Näin teknologian kehittymisen seurauksena eri töiden kysyntä muuttuu satunnaisten kysyntäshokkien seurauksena synnyttäen tulojen eksponentti- tai gammajakauman.

Shaikhin ja Taon teoriat ovat samankaltaisia siinä mielessä, että ne kiinnittävät huomion pois yksilöiden kyvyistä ja keskittyvät työpaikan tai alan tuottavuuteen. He ajattelevat kuinka työpaikka itse määrää palkan eikä niinkään yksilön oma tuottavuus. He voivat osua osin oikeaan, jos ajattelee matalan tuottavuuden ammattia kuten linja-autonkuljettajaa. Kuinka paljon todella taitavan ja keskiverto linja-autonkuljettajan palkka eroaa toisistaan? Lähinnä ainut tapa saada lisää palkkaa on tehdä enemmän töitä, eikä taidolla eli yksilön tuottavuudella ole tässä mielessä juurikaan vaikutusta. Perinteisessä yksilöiden tuottavuuteen keskittyvässä teoriassa ajatellaan kuinka yksilön omat luontaiset ja hankitut taidot määräävät palkan, mutta on eri asia toimiiko maailma näin täydellisesti. Tässä mielessä tulojen gamma- ja eksponenttijakautuneisuus voisi viitata jossain määrin yksilön kannalta satunnaisesti määräytyvään palkkaan. Kaasumolekyylien tavoin palkansaajat ikäänkuin kimpoilevat ammattien välillä talouden sisällä.

4 SUOMEN TULOJAKAUMA

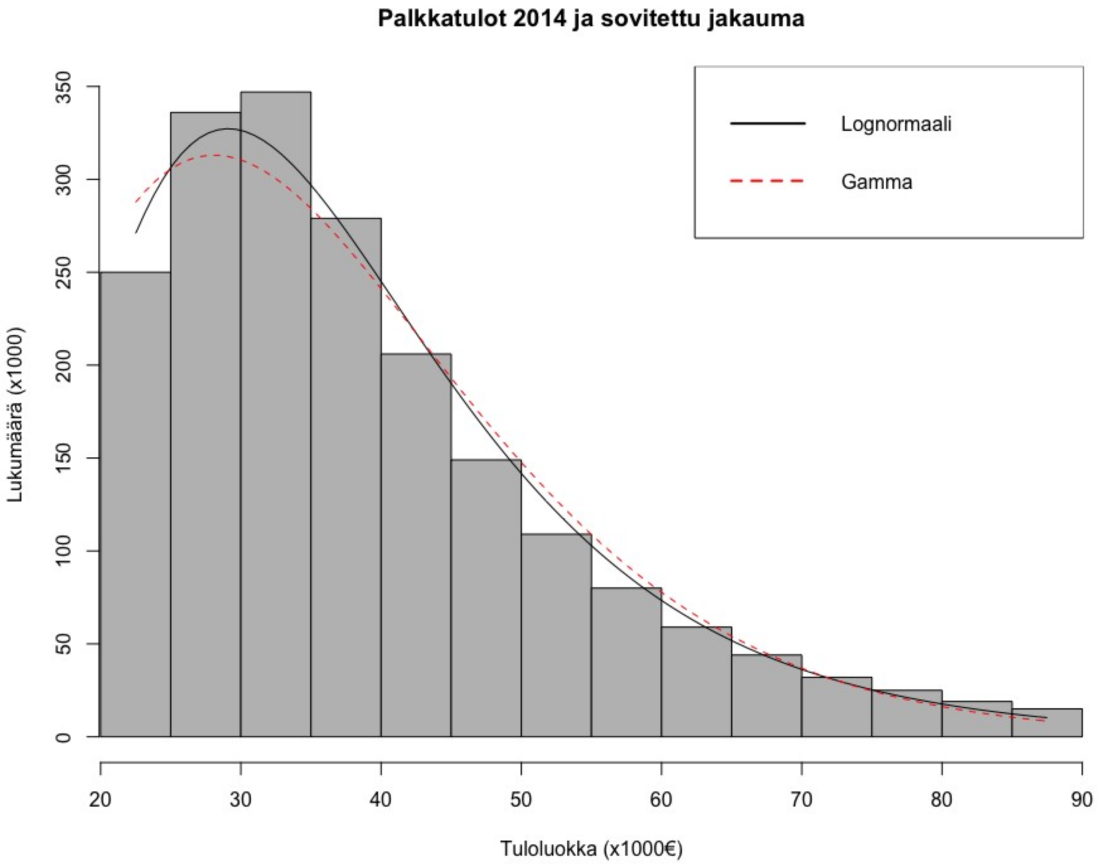
Seuraavaksi tarkastellaan Suomen tuloja aluksi sovittamalla lognormaali- ja gammajakaumaa palkka-, pääoma- ja osinkotuloihin. Lopuksi tarkastellaan erikseen tuhannen suurituloisimman suomalaisen pääoma- ja ansiotulojen jakautumista. Aineisto palkka-, pääoma- ja osinkotuloista on julkisesti saatavilla Verohallinnon tilastotietokannasta (Verohallinto). Tuhannen suurituloisimman ansio- ja pääomatulojen tiedot on julkisesti saatavilla ja ne on kerätty Yleisradion (YLE) ja Iltasanomien (Iltasanomat) internetsivuilta.

Analyysi on suoritettu R-tilasto-ohjelmalla. Jakaumien sovittamiseen on käytetty R:n fitdistrplus- ja MASS-paketteja, joiden avulla jakaumien parametrit on estimoitu suurimman uskottavuuden menetelmällä. Parametrien estimoinnin jälkeen lognormaali- ja gammajakauman yhteensopivuutta aineistoon tarkastellaan eri tunnusluvuin. Nämä ovat tilastolliset testit; Kolmogorov–Smirnov-testi, Anderson–Darling-testi ja Cramer–von Mises-testi. Näiden lisäksi jakaumien vertailuun käytetään kahta mallien hyvyyttä tai ennustusvoimaa kuvaavia ns. informaatiokriteeriä, jotka toimivat samaan tapaan kuten selitysaste regressioanalyysissä.

4.1 Palkkatulot

Palkka-aineisto käsittää veronsa Suomeen maksaneet palkansaajat, jotka on laitettu 5000 euron tuloluokkiin alkaen 20 000 eurosta ja päättyen 90 000 euroon. Vuonna 2014 palkansaajia oli yhteensä 2 784 126 kappaletta, joista yli 90 000 euron vuosipalkan tienasi 84 676 ihmistä eli noin 3% palkansaajista. Alle 20 000 vuosipalkat jätetään pois tarkastelusta yksinkertaisuuden vuoksi. Ansiotuloihin lasketaan palkkatulojen lisäksi mm. eläketulot, etuudet ja tuet kuten opintotuki ja työttömyyskorvaus. On luonnollista odottaa kuinka tulonsiirrot, eläkkeet ja palkat noudattavat eri mekanismeja, joten tästä syystä analyysissa tarkastellaan ansiotulojen sijaan palkkatuloja. Sovitettu jakauma pyrkii leikkaamaan jokaisen pylvään sen keskikohdasta, koska analyysissä on käytetty keskipistemenetelmää. Tässä kaikki 20 000 – 24 999 -tuloluokan palkansaajat saavat kaikki arvon 22 500. Analyysissa tulot on jaettu 1000 yksinkertaisuuden vuoksi. Tämä mittakaavan pienentäminen tuhannella luonnollisesti pienentää myös parametrit 1000:lla, joka on huomioitava. Jakaumien muotoon se ei vaikuta, jolloin se ei myöskään vaikuta mallien sopivuuteen aineiston kanssa.

Kuviossa 1 on esitetty vuoden 2014 palkkajakauma. Ensimmäiseksi visuaalisesti havaitaan kuinka molemmat jakaumat likimain sopivat hyvin kuvaamaan palkkatuloja. Lähempi tarkastelu osoittaa kuitenkin lognormaalijakauman sopivan paremmin kuvaamaan palkkatuloja. Gammajakauman huippu on liian tasainen, joka jää aineiston alapuolelle 20 000 – 30 000 euron välillä. Gammajakauma kulkee lognormaalijakauman yläpuolella välillä 45 000 – 65 000 ja lopulta tippuu lognormaalijakauman alle tästä eteenpäin. Lognormaalijakauma kuvaa paremmin jokaista tuloluokkaa, mutta ero on kuitenkin pieni. Toisaalta alimmat tuloluokat 20 000 – 35 000 poikkeaa lognormaalista.



Kuvio 1

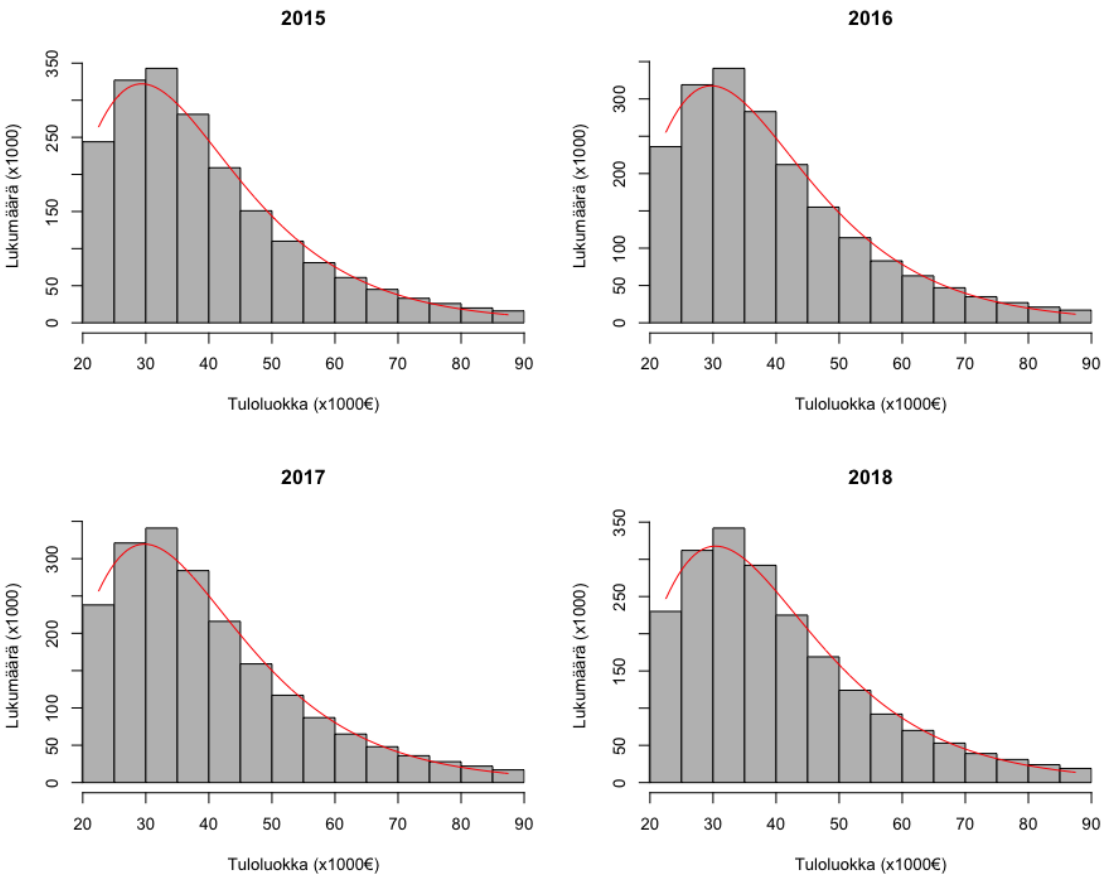
Taulukko 1

Gamma	Lognormaali
Muoto k	Sijainti μ
=4.692188	=3.545627
Skaala θ	Muoto σ
=7.60714859	=0.4186675

Taulukko 1 kertoo lognormaali- ja gammajakauman aineistosta estimoidut parametrit suurimman uskottavuuden menetelmällä. Taulukossa 2 vertaillaan näiden jakaumien yhteensopivuutta aineistoon. Nämä tilastolliset testit vahvistavat lognormaalijakauman paremman yhteensopivuuden verrattuna gammajakaumaan. Kolmogorov-Smirnov, Cramer-von Mises ja Anderson-Darling -testeissä sekä Akaiken ja Bayesin informaatiokriteereissä lognormaalijakauma saa pienempiä arvoja vahvistaen mallin paremmuuden verrattuna gammajakaumaan. Ero on kuitenkin edelleen pieni.

Taulukko 2

Yhteensopivuuden tunnusluvut	Gamma	Lognormaali
Kolmogorov-Smirnov	0.09784816	0.09340067
Cramer-von Mises	3.18215041	2.98466055
Anderson-Darling	20.99132397	19.81020393
Akaiken informaatiokriteeri	15178.58	15156.89
Bayesin informaatiokriteeri	15189.74	15168.04



Kuvio 2

Kuviossa 2 sovitetaan paremmaksi malliksi todettu lognormaalijakauma vuosien 2015 – 2018 palkka-aineistoon. Kuvion 2 graafinen tarkastelu osoittaa kuinka lognormaalijakauma näyttää sopivan aineistoon edelleen hyvin ja vaihtelu vuosien välillä on pientä. Kuitenkin edelleen lognormaali näyttäisi poikkeavan alempien tuloluokkien kohdalla. Taulukko 3 vertailee vuosien 2015 – 2018 lognormaalijakauman yhteensopivuuden tunnuslukuja. Lognormaalijakauma on suhteellisen stabiili viiden tarkasteluvuoden aikana ja suuria poikkeamia ei ole.

Taulukko 3

Yhteensopivuuden tunnusluvut	2015	2016	2017	2018
Kolmogorov-Smirnov	0.09153313	0.09029584	0.0891382	0.08640121
Cramer-von Mises	2.89746401	2.81067000	2.7709939	2.67742315
Anderson-Darling	19.22979801	18.59482957	18.4407909	17.75067450
Akaiken informaatiokriteeri	15177.74	15276.95	15505.20	15927.14
Bayesin informaatiokriteeri	15188.89	15288.10	15516.38	15938.36

Taulukko 4 kertoo eri vuosien lognormaalijakauman parametrit suurimman uskottavuuden menetelmällä. Sijaintiparametri μ ja muotoparametri σ kasvavat tasaisesti 5 vuoden aikana. Taulukossa 5 on laskettu eri vuosien palkkajakauman mediaanin ja jakauman keskiarvon estimoiduilla parametrien arvoilla.

Taulukko 4

Lognormaalijakauman parametrit	Sijainti μ	Muoto σ
2015	3.5552020	0.4195887
2016	3.5679413	0.4196353
2017	3.5716336	0.4216866
2018	3.5920442	0.4221109

Taulukko 5

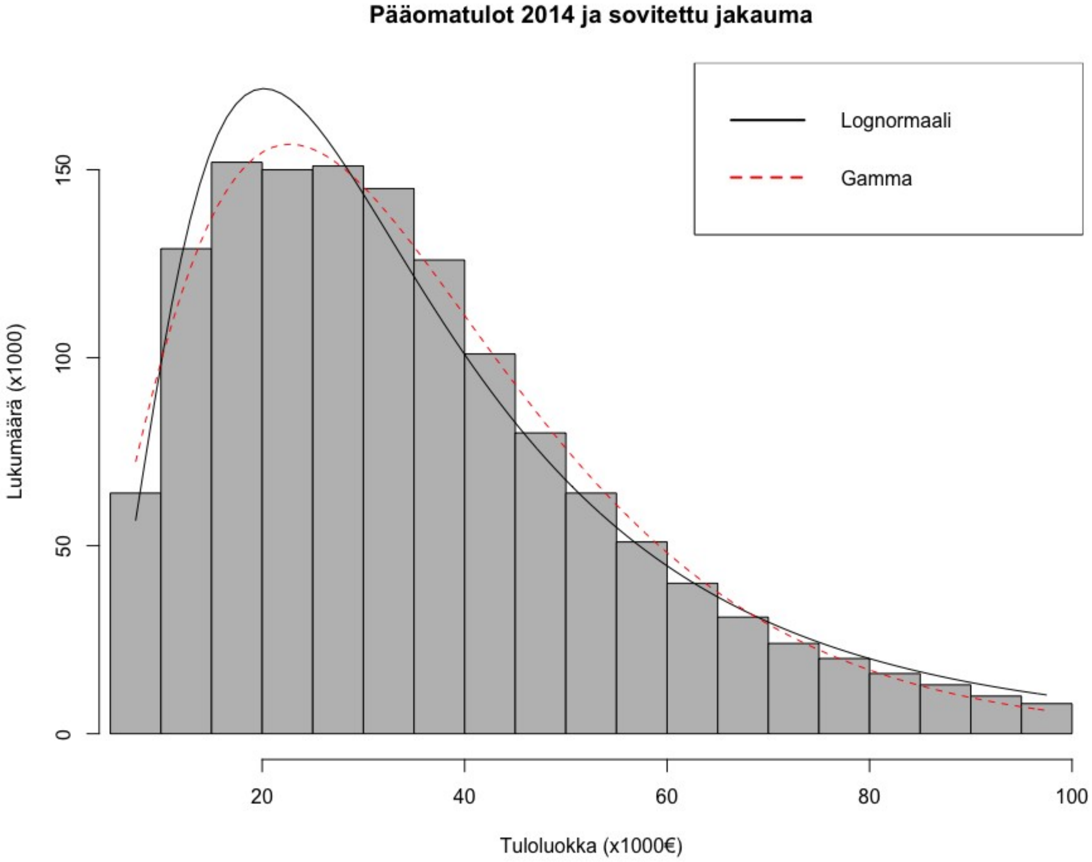
Palkkajakauman tunnusluvut	Odotusarvo	Mediaani
2014	37.836	34.66141
2015	38.215	34.99489
2016	38.706	35.44355
2017	38.882	35.57466
2018	39.691	36.30822

Keskimääräinen palkka eli lognormaalijakauman odotusarvo kasvoi 5 vuoden aikana $\frac{39.691-37.836}{37.836}=0.049$ eli 4.9%. Vastavuoroisesti mediaanipalkka kasvoi $\frac{36.30822-34.66141}{34.66141}=0.048$ eli 4.8%. Nämä luvut eivät luonnollisesti ole todenmukaiset, koska analyysissä on mukana 20 000 – 90 000 euroa vuodessa tienanneet, joka on noin 70% kaikista palkansaajista. Tilastojen mukaan keskimääräinen palkka kasvoi 6.8% ja mediaanipalkka kasvoi 5.8% kyseisenä ajanjaksona (Verohallinto).

4.2 Pääomatulot

Pääomatulojen tarkastelun osalta analyysi on identtinen palkkatulojen kanssa, mutta tuloluokat ovat väliltä 5000 – 100 000 euroa. Ensin tarkastellaan pääomatuloja, minkä jälkeen erikseen osinkotuloja. Vuonna 2014 oli 1 586 322 pääomatulon saajaa, joista yli 100 000 euron vuosittaiset pääomatulot sai 60 230 ihmistä eli 3.8% kaikista pääomatulojen saajista.

Kuviosta 3 graafisesti tarkasteltuna huomataan vastavuoroisesti gammajakauman sopivan paremmin kuvaamaan pääomatuloja. Pääomatuloissa huippu on tasaisempi verrattuna palkkatuloihin. Myös hännässä suurempien tulojen osalta gammajakauma osuu paremmin yhteen aineiston kanssa. Taulukko 6 kertoo vuoden 2014 molempien jakaumien estimoidut parametrit.



Kuvio 3

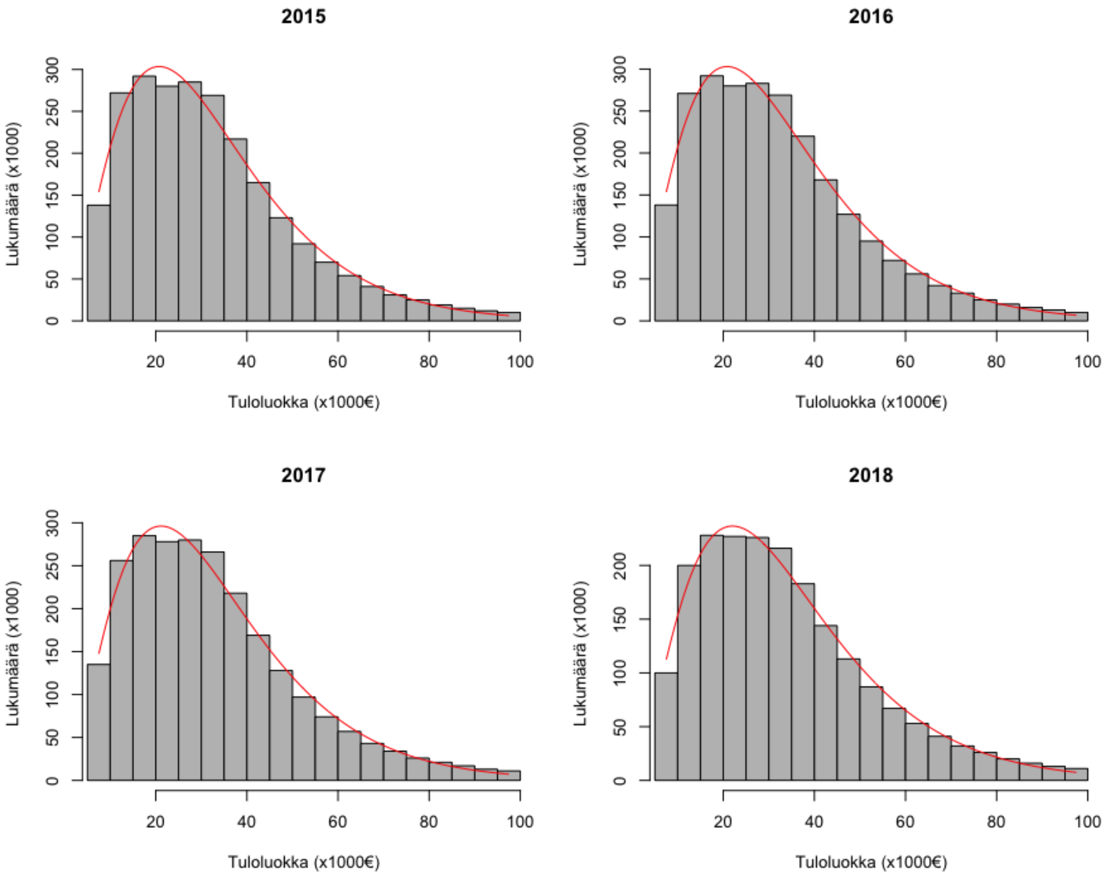
Taulukko 6

Gamma	Lognormaali
Muoto k	Sijainti μ
=2.76460230	=3.445205
Skaala θ	Muoto σ
=12.88147523	=0.664175

Taulukko 7 vahvistaa gammajakauman sopivan pääomatuloihin paremmin kuin lognormaalijakauma. Yhteensopivuuden tunnusluvut osoittavat kuinka gammajakauma sopii kuvaamaan pääomatuloja paremmin kuin lognormaalijakauma kuvasi palkkatuloja. Kuviossa 4 gammajakauma on sovitettu vuosien 2015 – 2018 pääoma-aineistoon, joita kuvaavat gammajakauman parametrit on esitetty taulukossa 8 ja yhteensopivuuden tunnusluvut on esitetty taulukossa 9. Samantapainen stabiilisuus on havaittavissa pääomatulojen osalta kuten palkkatulojen kanssa. Aineiston vuosittaiset vaihtelu on pieniä, joten estimoidun gammajakauman on parametrit ja yhteensopivuus aineiston kanssa ovat vakaita vuosien 2014 – 2018 aikana.

Taulukko 7

Yhteensopivuuden tunnusluvut	Gamma	Lognormaali
Kolmogorov-Smirnov	0.0594325	0.07781119
Cramer-von Mises	0.8916514	1.09264070
Anderson-Darling	5.6943733	7.06369335
Akaiken informaatiokriteeri	11725.98	11732.44
Bayesin informaatiokriteeri	11736.43	11742.89



Kuvio 4

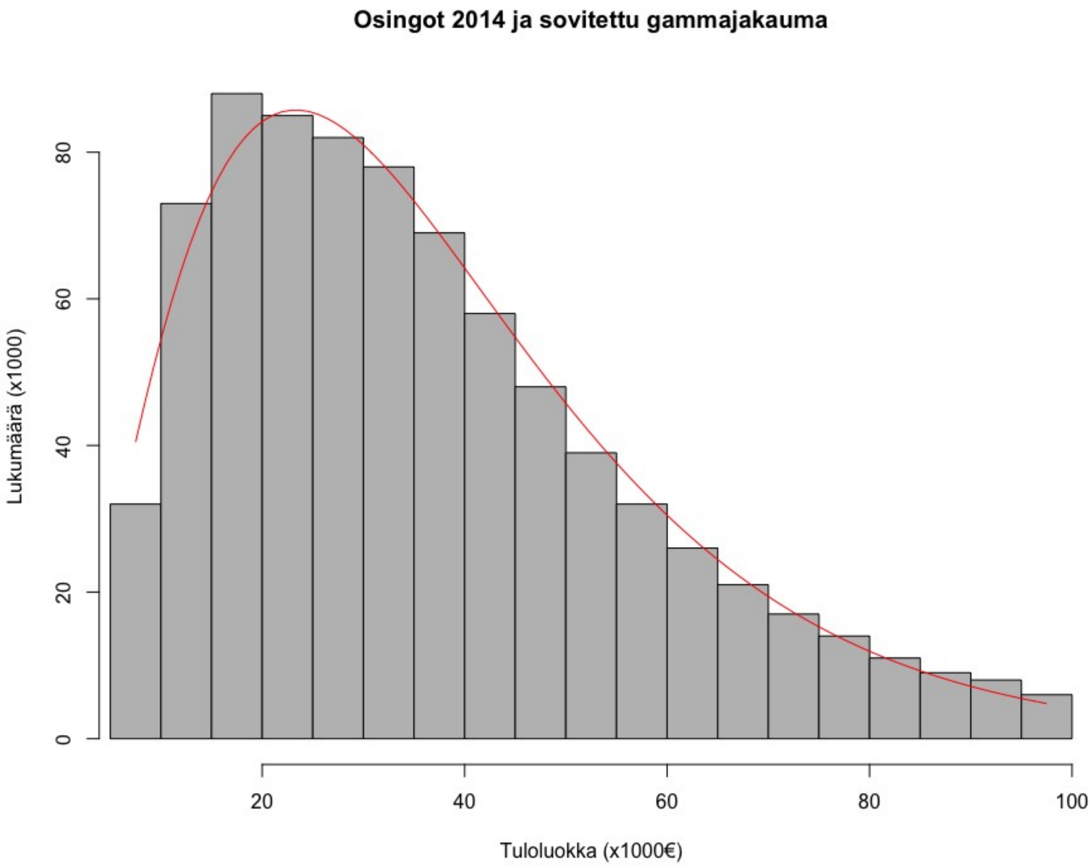
Taulukko 8

Gammajakauman parametrit	Muoto k	Skaala θ
2015	2.799448	11.52123375
2016	2.782538	11.68980586
2017	2.785549	11.84161648
2018	2.782920	12.34919472

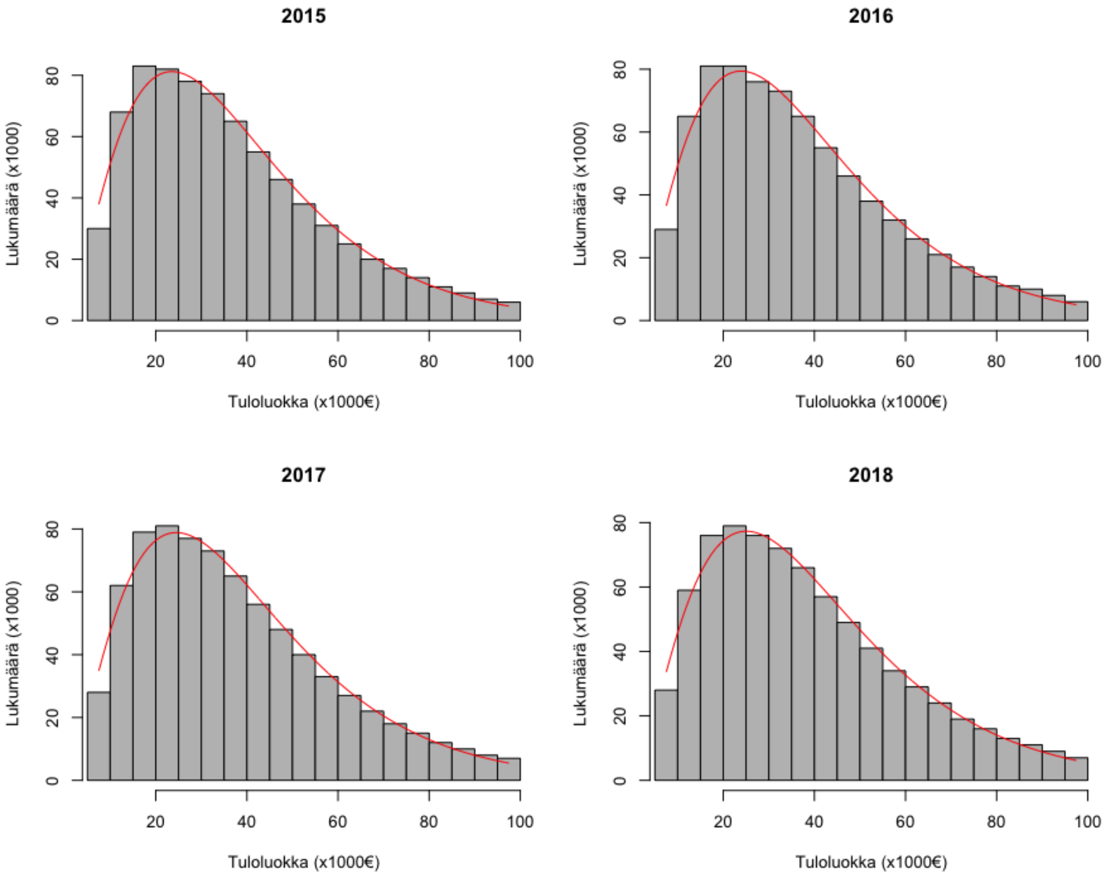
Taulukko 9

Yhteensopivuuden tunnusluvut	2015	2016	2017	2018
Kolmogorov-Smirnov	0.06561889	0.06521973	0.06318765	0.06077424
Cramer-von Mises	1.92371305	1.89256274	1.82352552	1.40093602
Anderson-Darling	12.3261004	12.1602601	11.6775745	8.94618060
Akaiken informaatiokriteeri	20150.51	20359.73	20230.90	16964.74
Bayesin informaatiokriteeri	20162.08	20371.32	20242.48	16975.94

Pääomatulot sisältää eri tulolajeja kuten vuokra-, korko, osinkotulot ja omaisuuden luovutusvoitot. Nämä tulot luonnollisesti noudattavat eri mekanismeja, jolloin niiden kaikkien luokitteleminen pääomatuloiksi voi haitata tietyn jakauman sopimista aineistoon. Juuri tästä syystä ensimmäisessä osiossa tarkasteltiin palkkatuloja ansiotulojen sijaan. Tämän vuoksi kuvioissa 5 ja 6 erikseen sovitetaan gammajakauma osinkotuloihin. Taulukossa 10 on esitetty gammajakauman estimoidut parametrit ja taulukossa 11 näiden yhteensopivuuden tunnusluvut.


Osingot 2014 ja sovitettu gammajakauma
Lukumäärä (x1000)
0
20
40
60
80
20
40
60
80
100
Tuloluokka (x1000€)

Kuvio 5



Kuvio 6

Taulukko 10

Gammajakauman parametrit	Muoto k	Skaala θ
2014	2.64678144	14.15873
2015	2.65070140	14.24675
2016	2.64525324	14.51951
2017	2.66809130	14.68202
2018	2.64860273	15.18328

Taulukko 11

Yhteensopivuuden tunnusluvut	2014	2015	2016	2017	2018
Kolmogorov-Smirnov	0.0623	0.0631	0.0630	0.0614	0.0586
Cramer-von Mises	0.48500847	0.45942419	0.44392139	0.43164743	0.40530574
Anderson-Darling	3.15095606	2.97023219	2.86713596	2.77366087	2.60632207
Akaiken informaatiokriteeri	6859.091	6548.099	6522.110	6599.913	6660.601
Bayesin informaatiokriteeri	6868.450	6557.363	6531.361	6609.183	6669.881

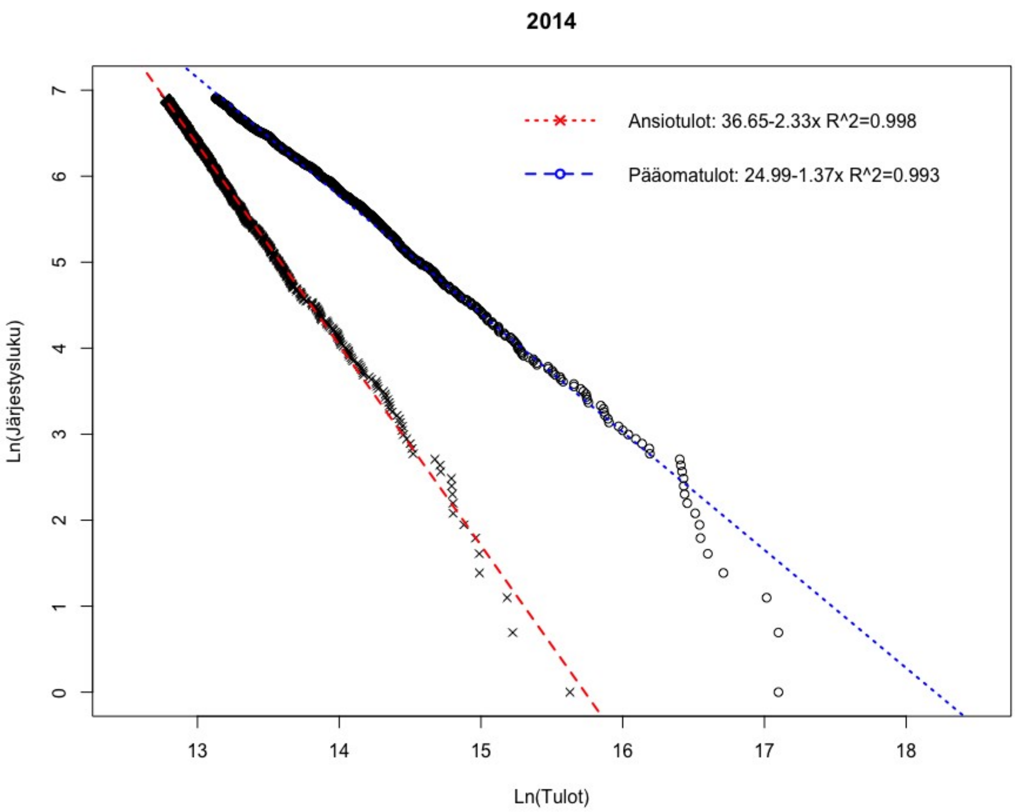
Edellä esitetyt kuviot ja taulukoiden tilastolliset testit vahvistivat kuinka pääoma- ja samalla myös osinkotuloja kuvaamaan sopii paremmin gammajakauma. Osinko- ja pääomatuloja pystyy kuvaamaan likimain saman suuruisella tarkkuudella. Vuosien välillä tuloerot pysyvät vakaana palkkatulojen tapaan.

4.3 Huipputuloiset

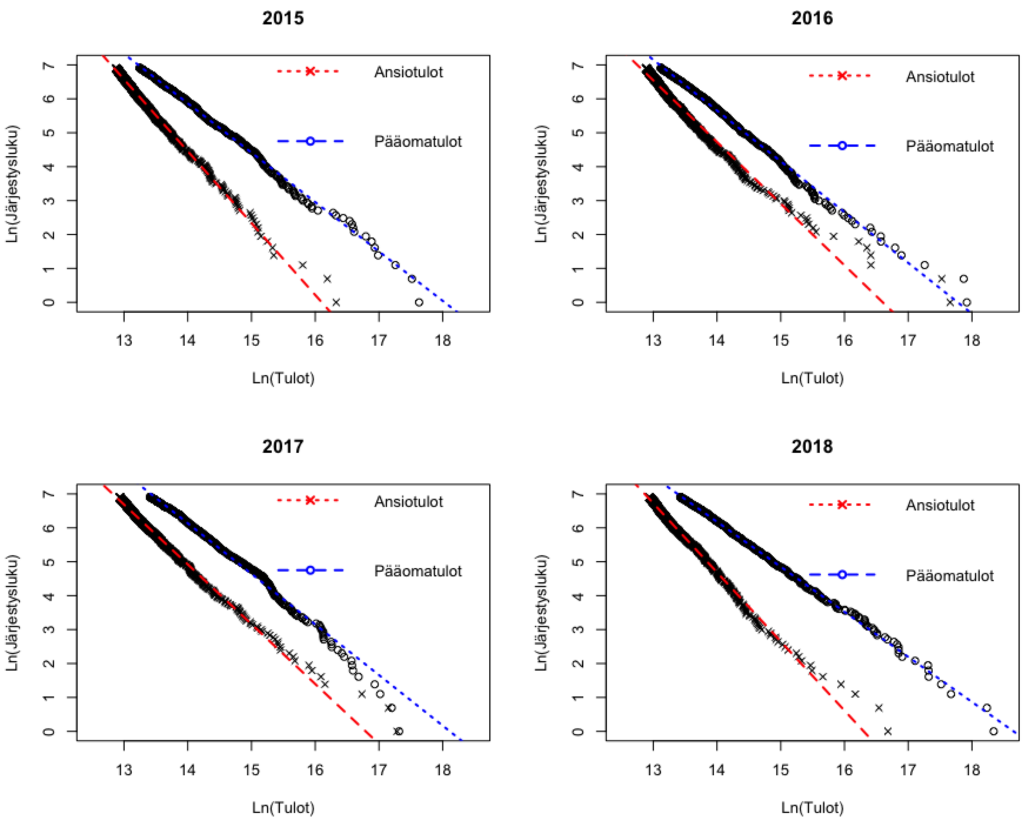
Seuraavaksi tarkastellaan Suomen tuhatta suurituloisinta pääoma- ja ansiotulojen saajaa, joiden osalta analyysi on molemmissa samankaltainen. Tuhat suurinta ansio- ja pääomatulon saajaa asetetaan kuvioon ottamalla luonnollinen logaritmi järjestysluvusta (y-akseli) sekä heidän tuloistaan (x-akseli). Näin tulot asettuvat suoraksi viivaksi log-log-akselistossa potenssijakauman ominaisuuksien mukaisesti.

Vuoden 2014 tuhat suurituloisinta on merkitty kuvioon 7, johon on lisäksi merkitty regressiosuoran yhtälö ja selitysaste. Suurituloisin ansiotulo oli 6 125 472 euroa, jonka luonnollinen logaritmi on Ln(6 125 472) = 15.63. Koska luvun 1 logaritmi on 0, suurituloisimman xy-akselin koordinaatit ovat (0, 15.63). Vastavuoroisesti toiseksi suurimman ansiotulon saajan x-koordinaatti on Ln(4 090 140) = 15.22 ja y-koordinaatti Ln(2) = 0.69. Kuviosta havaitaan kuinka ansiotulojen ja järjestysluvun logaritmimuunnos asettuu lineaarisesti, jonka regressiosuoran yhtälö on 36.65 - 2.33x ja selitysaste on erittäin korkea 0.998. Samaan kuvioon on merkitty myös pääomatulot. Vuoden 2014 suurin pääomatulon x-koordinaatti on Ln(26 662 577) = 17.1. Pääomatulojen regressiosuora on muotoa 24.99 - 1.37x ja selitysaste 0.993. Ansio- ja pääomatulojen regressiosuorien kulmakertoimet kertovat Paretoeksponentin suuruuden kuten aikaisemmin osoitettiin.

Paretojakaumien parametrit ovat $\alpha_{pääoma} = 1.37$ ja $\alpha_{ansio} = 2.33$, jotka ovat samaa luokkaa kuin muissa tutkimuksissa. Pääomatulojen alhaisempi Paretoeksponentti viittaa suurempiin pääomatulojen tuloeroihin. Myös tämä on linjassa muiden tutkimus havaintojen kanssa (Gabaix 2009). Kuviossa 8 on esitetty vuosien 2015 – 2018 tuhat suurituloisinta pääoma- ja ansiotulon saajaa. Taulukoissa 12 ja 13 on esitetty vastaavanlaisesti regressiosuorien yhtälöt ja näiden selitysasteet.



Kuvio 7



Kuvio 8

Taulukko 12

Ansiotulot	Regressiosuora	Selitysaste R^2
2015	34.12 - 2.12x	0.9961
2016	30.17 - 1.82x	0.9731
2017	29.61 - 1.76x	0.9918
2018	33.20 - 2.04x	0.9939

Taulukko 13

Pääomatulot	Regressiosuora	Selitysaste R^2
2015	26.26 - 1.46x	0.9969
2016	26.61 - 1.50x	0.9961
2017	27.02 - 1.49x	0.9925
2018	24.77 - 1.33x	0.9988

Huomioitavaa on kuinka regressiosuoran kulmakerroin ja siten Paretoeksponentin arvo vaihtelee vuodesta toiseen merkittävästi. Suuret tulot eroavat tässä mielessä suhteellisen vakaista matala- ja keskituloisten palkka-, pääoma ja osinkotuloista. Vuosien välisestä heilahtelusta huolimatta regressiosuoran selitysaste pysyy erittäin korkeana. Tässä analyysissa tarkasteltiin vain suurituloisimman tuhannen tuloja, jotka edustavat pientä osaa suurituloisimmasta 1 – 3 %:sta.

Potenssijakaumien mittakaavan skaalattomuuden takia takia voidaan suhteellisen luottavaisesti odottaa regressiosuoran kulmakertoimen pysyvän likimain samansuuruisen, jos analyysiin mukaan yli 100 000 ansio- ja pääomatulot.

5 POHDINTAA

Tämä tutkielma on osa uutta taloustieteellistä tuloeroja käsittelevää nousevaa trendiä. Aikaisemmin tuloeroista ei ole löytynyt riittävän tarkkaa ja luotettavaa aineistoa, jonka päälle olisi voinut rakentaa riittävän syvää analyysiä. Parantunut aineiston laatu ja kiinnostus nousu tuloerojen yhteiskunnallisista vaikutuksista ovat yhdessä lisänneet merkittävästi tutkijoiden kiinnostusta niiden tutkimiseen (Cowell & Flachaire 2008).

Suomen tulojakauma noudatti merkittävää säännönmukaisuutta sekä tuhannen suurituloisimman että pieni- ja keskituloisten osalta. Pieni- ja keskituloisten palkkatuloja kuvasi paremmin lognormaalijakauma gammajakauman sijaan, mikä puoltaisi perinteistä näkemystä palkkojen muodostuksesta. Ero ei kuitenkaan ollut suuri. Lognormaali- ja gammajakauman samankaltaisuus selittää sitä, miksi tutkijat eivät ole yksimielisiä mikä jakauma sopisi kuvaamaan tuloja. Tutkielmassa tarkasteltiin erikseen pääoma- ja palkkatuloja, mikä antaa paremman tarkkuuden tuloerojen rakenteesta. Tämä erottelu voi olla myös syy miksi kirjallisuudessa on poikkeavia näkemyksiä. Tästä huolimatta makrotason palkkatulojen mallintamiseen riittää yksinkertainen kahden parametrin lognormaalijakauma, vaikka mikrotason palkkaan vaikuttavat tekijät ovat monimutkaiset. Tämän palkkojen lognormaalisuus viittaisi markkinoiden pystyvän hinnoittelemaan työpanoksen hinnan likimain vastaamaan ihmisten tuottavuutta. Tämä puoltaisi meritokraatista näkemystä palkan muodostuksesta ja palkkojen yhteydestä yksilöiden tuottavuuseroihin.

On kuitenkin huomioitava muutama asia, jotka voivat muuttaa tuloksia. Aineistona käytettiin vuoden palkkatuloja, minkä takia yhteys tuottavuuteen ei välttämättä ole suoraviivainen. Osa-aikaiset työntekijät, kesätöitä tekevät opiskelijat ja vuoden aikana työttömäksi joutuneet työntekijät vääristävät jakaumaa. Tämä joukko painottuu palkkatulojakauman pienituloisten puolelle, jolloin on odotettavissa matalien palkkatulot voivat poiketa lognormaalijakaumasta. Juuri tästä syystä tarkasteltavan aineiston ulkopuolelle jätettiin nämä erittäin pienituloiset, joten tutkielman pienin tuloluokka oli 20 000 – 25 000 euron vuositulot. Tämän alle jäi huomattava määrä palkkatuloja saaneista. Verohallinnon tietojen mukaan palkkatuloja saaneiden lukumäärä oli 2 784 126 kappaletta vuonna 2014. Keski- ja pienituloisten analyysissä 20 000 – 90 000 euron vuositulon saaneiden lukumäärä oli 1 958 024 eli noin 70% kaikista palkkatuloja saaneista (Verohallinto). Ehkä osin tästä syystä lognormaalijakauma sopi paremmin aineistoon välille 35 000 – 90 000€. Parempi lähtökohta olisi tarkastella tuntipalkkaa tai vastaavaa vuosipalkan sijaan, koska se antaisi paremman kuvan ihmisten tuottavuudesta. Tutkielman aineisto vuosipalkasta on kuitenkin helpommin saatavilla, joten tutkielmassa tarkasteltiin sitä. Toinen huomioitava asia on kuinka

5000€ tuloluokka voi olla liian karkea ja tarkempi aineisto antaisi antaisi paremman analyysin. Tarkempaa aineistoa ei ole kuitenkaan vapaasti saatavilla Verohallinnon internet-sivuilla.

Pääomatulojen osalta gammajakauma muistuttaa ekonofyysikoiden kineettisiä vaihdantamallien tuloksia. Tällöin termodynamiikan kaltainen säilymislaki voisi liittyä pääomatulojen jakaumaan vaikuttaviin mekanismeihin. Esimerkiksi johdannaismarkkinat voisivat noudattaa nolla summa peli -mekanismeja. Tällöin johdannaismarkkinoilla ei tuoteta arvoa vaan arvoa jaetaan markkinoilla ostajien ja myyjien kesken pokeripöydän tavoin. Johdannaisen arvo on johdettu toisesta hyödykkeestä tai arvopaperista. Futuurin, option tai swapin transaktioon tarvitaan myyjä ja ostaja. Tässä transaktiossa ei kuitenkaan tuoteta hyödykettä, arvopaperia, osaketta tai velkakirjaa. Tästä syystä toisen voitto on toisen häviö. Raha ei häviä vaan se vaihtaa vain omistajaa johdannaisen transaktion mukana. Täten rahamäärä systeemissä voisi olla vakio lyhyellä aikavälillä, jolloin entropian kasvaessa ja vaihdannan seuraksena tuloerot liikkuvat kohti gammajakautunutta tasapainoa. Myös osinkotulot, jotka perustuvat pitkälti yritysten liiketoiminnasta saatuihin voittoihin, vaatii kunnollisia perusteluja miksi ne mahdollisesti noudattavat gammajakaumaa.

Tuloerojen ja termodynamiikan yhdistäminen ei välttämättä ole niin kaukainen ajatus. Epätasa-arvoisuutta ja ”tuloeroja” löytyy luonnosta. Jopa erakkoravut kärsivät taloudellista epätasa-arvosta simpukankuorissa mitattuna. Tietyin väliajoin erakkoravun on vaihdettava pieneksi jäänyt simpukankuori, jolloin joukko rapuja kerääntyy kuivalle maalle. Vaihdannan jälkeen erakkoravuilla on sopivan kokoinen simpukankuori suojanaan. Tutkijat mittasivat erakkorapujen simpukankuorien painoja ja havaitsivat niiden jakautuvan epätasa-arvoisesti ihmisten varallisuuserojen tapaan (Chase ym. 2020). Tutkijat ovat löytäneet monia muitakin samankaltaisuuksia luonnossa esiintyvien ilmiöiden ja ihmisten taloudellisen epätasa-arvon välillä. Tutkijoiden mielestä yksinkertaiset matemaattiset mekanismit saavat aikaan samantyyppisiä lopputuloksia niin luonnossa kuin taloudessa. Taloudellinen epätasa-arvo on väistämätön lopputulos ellei aktiivisesti kehitystä estä (Scheffer ym. 2019). Tuloerojen luonnollista selitystä tukee myös se, kuinka niitä voidaan kuvata suhteellisen tarkasti vain kahdella numerolla. Näillä yksinkertaisilla kahden parametrien jakaumilla voidaan myös kuvata eri tyyppisten ja kokoisten kansantalouksien tuloeroja ja säännönmukaisuus säilyy myös ajan kuluessa. Tämä viittaa siihen kuinka on olemassa tiettyjä mekanismeja, jotka synnyttävät nämä säännönmukaisuudet.

On mielenkiintoista pohtia mitä tulojen mahdollinen gammajakauma kertoo tulojen syntymekanismista. Jos lognormaalijakauma viittaa meritokratiaan, niin samantapaista perustelua on vaikeampi löytää tulojen gammajakaumalle. Työntekijät voivat jossain määrin liikkua satunnaisesti työpaikkojen välillä, mikä synnyttäisi tulojen gammajakauman. Mandelbrot kuitenkin muistutti kuinka jakaumia voidaan yhdistää. Ei ole tavatonta kuinka sama mitattava asia noudattaa eri mekanismeja eri koko luokissa. Esimerkiksi sademetsän puiden kasvu noudattaa päinvastaisia mekanismeja kuin tulot. Matalien ja nuorien puiden rungon paksuus noudattaa aluksi potenssijakaumaa, koska sademetsässä puut kilpailevat rajallisesta valosta. Mitä suuremmaksi puu kasvaa, sitä enemmän valoa se saa itselleen, joka auttaa kasvamisessa. Riittävän suurille puille valo ei ole enää merkittävä tekijä, jolloin puiden koko alkaa noudattamaan eksponenttijakaumaa (Farrior ym. 2016). Teknisesti ei ole mitään estettä etteikö tuloja voitaisiin kuvata kolmella jakaumalla. Mahdollisesti pienituloiset voisivat noudattaa gamma-, keskituloiset lognormaali- ja huipputuloiset Paretojakaumaa. Tämänkaltaisen kolmen jakauman malli on saanut tukea kirjallisuudesta (Schneider 2008). Kuitenkin pelkkä tulojen mallintaminen eri jakaumilla ei vielä yksistään riitä antamaan todenmukaista kuvaa niiden syntymekanismista, joten empiiriset todisteet on pystyttävä selittämään hyvällä teorialla.

Suurituloisten tarkastelu rajoittui tässä tutkielmassa vain tuhanteen suurituloisimpaan. Kun siirryttiin tuloissa ylöspäin tarkastelemaan suurituloisia vaihtui tulojen lognormaali- ja gammajakauma potenssijakaumaksi eri lähtökohdista huolimatta. Edelleen samantapainen yksinkertaisuus ja säännönmukaisuus pätee tuhannen suurituloisimman pääoma- ja ansiotulojen osalta. Vaikka pieni- ja keskituloisten pääoma- ja ansiotulot noudattavat eri jakaumia, suurituloisten osalta molemmat tulolajit sopivat potenssijakauman mukaiseen kuvaukseen. Tämä viittaisi siihen kuinka suurituloiset nauttivat mittakaavaetua sekä pääomatulojen että ansiotulojen osalta. Tämä suurituloisten etulyöntiasema ei ole ennenkuulumaton, koska tämänkaltainen Paretojakauma esiintyy systeemissä, jolle on ominaista rikkaat rikastuu -ilmiö. Jokainen tienattu euro lisää tulevaisuuden tienaamisen todennäköisyyttä. Thomas Pikettyn teos *Pääoma 2000-luvulla* (2014) on nostanut suurituloiset puheen aiheeksi ja samalla tuloeroihin kohdistuvan taloustieteellisen kiinnostuksen. Charles Jones muistuttaa Pikettyn työtä käsittelevässä esseessään Paretojakauman tärkeydestä: ”*...suurituloisimman yhden tai 0.1 prosentille päätyvän tulo-osuuden ja Paretojakauman parametrin välillä on vahva yhteys. Ymmärtämällä miksi huipun tuloerot noudattavat Paretojakaumaa ja mitkä taloudelliset voimat vaikuttavat parametriin, on keskeinen asia faktojen ymmärtämisessä.*” (2015). Tämän yhteyden tutkiminen on edelleen avoin taloustieteessä.

6 JOHTOPÄÄTÖKSET

Tutkielmassa tarkasteltiin palkka-, pääoma- ja osinkotulojen jakaumaa vuosilta 2014 – 2018 sekä tuhannen suurituloisimman ansio- ja pääomatuloja vuosilta 2014 – 2016. Tutkielma vahvisti tieteellisen kirjallisuuden konsensuksen, jonka mukaan tuloerojen mallintamiseen tarvitaan kaksi eri matemaattista jakaumaa. Tämä viittaisi siihen kuinka keski- ja pienituloisten tulojen määräytyminen noudattaa eri mekanismeja verrattuna suurituloisimpaan 1 – 3 %.

Palkkatulojen osalta parhaaksi jakaumaksi osoittautui lognormaalijakauma, joka on Gibratin tutkimuksista lähtien monien suosima malli kuvaamaan keski- ja pienituloisia. Lognormaalijakauma lepää teoreettisesti vahvalle pohjalle, koska se pohjautuu ekonomistien yleisesti hyväksyttyyn käsitykseen ihmisen tuottavuuteen pohjautuvasta ansiotulosta ja niiden jakautumiseen populaatiossa. Ero gammajakaumaan on kuitenkin pieni ja toiseen aineistoon voisi sopia paremmin gammajakaumaan perustuva malli. Tässä tutkielmassa tarkoituksella sovitettiin jakaumia palkkatuloaineistoon, koska ansiotuloissa eläkkeet ja tulonsiirrot voivat aiheuttaa mahdollista vinoumaa. Tämä tulisi ottaa huomioon, koska olisi outoa odottaa kuinka eläkkeet, tulonsiirrot ja palkat noudattaisivat samoja mekanismeja. Palkkatulojen lognormaalisuus täten viittaisi palkkojen määräytyvän jossain määrin meritokratisesti, jolloin ihmisen tuottavuus likimain vastaa heidän palkkaansa pieni- ja keskituloisten osalta. On huomioitava kuinka tutkielman vuositulojen tarkastelu voi antaa väärän käsityksen tuottavuuden ja palkan välisestä yhteydestä.

Suomen pieni- ja keskituloisten pääoma- ja osinkotulojen kuvaamiseen sopi paremmin vastavuoroisesti gammajakauma. Tämä tulos on mielenkiintoinen kahdesta syystä. Tilastolliset testit osoittivat gammajakauman sopivan pääomatulojen aineistoon paremmin kuin lognormaalijakauma sopi kuvaamaan palkkatuloja. Tuloeroja käsittelevä kirjallisuus ei ole samoissa määrin kiinnittänyt huomiota pääomatulojen jakautumiseen, kuin tuloihin yleisesti. Toiseksi tulos muistuttaa ekonofyysikoiden termodynamiikkaan pohjautuvaa teoriaa, jossa kaasumolekyylien kineettinen liike-energia säilyy suljetussa systeemissä synnyttäen eksponentti- tai gammajakauman. Tästä syystä tutkielmassa havaittu pääoma- ja osinkotulojen gammajakauma voisi jossain määrin viitata talouden säilymislakiin suljetussa systeemissä. Taloustieteellinen perustelu suljetulle systeemille on tämän osalta vielä ristiriitainen. Ekonofysiikka voi kuitenkin tarjota lupaavia näkökulmia teorian kehittämiseksi, koska he alunperin juuri keskittyivät varallisuuden ja rahan mallintamiseen. Markkinoiden kineettiset vaihdantamallit voivat tästä syystä osoittautua hedelmälliseksi työkaluksi selittämään pääomatulojen jakauman empiirisiä havaintoja.

Suomen tuhannen suurituloisimpien osalta ansio- ja pääomatulojen tarkastelu vahvisti jo pitkään tiedetyn potenssijakauman. Pääomatuloissa havaittiin suuremmat tuloerot kuin ansiotuloissa. Kuitenkin molemmat tulot asettuivat erittäin hyvin potenssijakauman mukaisesti suoraksi viivaksi logaritmisilla asteikoilla. Tulojen potenssijakauma viittaisi suurituloisten kumulatiiviseen etuun – mitä suuremmat tulot sitä suurempi todennäköisyys saada tulevaisuudessa lisää.

Tutkielma johdannossa esiteltiin Ricardon pääteoksen alkusanat, jossa hän painotti tarvetta tutkia tulojen jakautumisen selittäviä tieteellisiä lakeja. Vastaus tähän on vielä jossain määrin avoin, koska teorioita, havaintoja ja malleja on lukuisia. On kuitenkin jossain määrin rohkaisevaa hyvän teoria kehittämisen kannalta havaita kuinka tuloerot vaikuttavat noudattavan jossain määrin universaaleja piirteitä. Näitä tuloerojen makrotason vakaita ja säännönmukaisia ilmiöitä voidaan mahdollisesti käyttää hyödyksi eri analyyseissä ja vertailuissa Schumpeterin ennustuksen mukaisesti. Tässä tutkielmassa vain kahdella parametrin arvolla voitiin suhteellisen tarkasti kuvata tulojen jakautumista koko kansantalouden tasolla niin pääomien ja palkkatulojen osalta. Tuhannen suurituloisimman tulot noudattivat hämmästyttävän suurta säännönmukaisuutta kansantalouden huipulla. Nämä vakaiden havaintojen selittämiseksi toivottavasti saadaan rakennettua toimiva teoria tulevaisuudessa Ricardon toiveiden mukaan.

7 LÄHTEET

Adler, Moshe (1985) "Stardom and talent." American Economic Review 75 (1), 208–212.

Arrow, Kenneth J. & Gerard Debreu (1954): "Existence of an equilibrium for a competitive economy." Econometrica: Journal of the Econometric Society: 265-290.

Auerbach, Felix (1913) "Das Gesetz der Bevölkerungskonzentration." Petermanns Geographische Mitteilungen 59 (1), 74–76.

Axtell, Robert L. (2001) "Zipf distribution of US firm sizes." Science 293 (5536), 1818–1820.

Barabási, Albert-László, & Réka, Albert (1999): "Emergence of scaling in random networks." Science 286 (5439), 509–512.

Benguigui, L., & M. Marinov (2015): "A classification of natural and social distributions Part one: the descriptions." arXiv preprint arXiv:1507.03408.

Champernowne, David G. (1953): "A model of income distribution." The Economic Journal 63.250: 318-351.

Chase, Ivan D., Raphael Douady, & Dianna K. Padilla (2020): "A comparison of wealth inequality in humans and non-humans." Physica A: Statistical Mechanics and its Applications 538 122962.

Chatterjee, Arnab, & Bikas K. Chakrabarti. (2007): "Kinetic exchange models for income and wealth distributions." The European Physical Journal B 60.2: 135-149.

Cho, Adrian (2014): "Physicists say it's simple.": 828-828.

Clauset, Aaron, Cosma Rohilla Shalizi, & Mark EJ Newman (2009): "Power-law distributions in empirical data." SIAM review 51.4: 661-703.

Cook, Will, Paul Ormerod, & Ellie Cooper (2004): "Scaling behaviour in the number of criminal acts committed by individuals." Journal of Statistical Mechanics: Theory and Experiment 2004.07: P07003.

Cowell, Frank A., & Emmanuel Flachaire (2007): "Income distribution and inequality measurement: The problem of extreme values." Journal of Econometrics 141.2: 1044-1072.

Drăgulescu, Adrian A. (2003): "Applications of physics to economics and finance: Money, income, wealth, and the stock market." arXiv preprint cond-mat/0307341.

Drăgulescu, Adrian, & Victor M. Yakovenko (2001): "Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States." Physica A: Statistical Mechanics and its Applications 299.1-2: 213-221.

Dragulescu, Adrian, & Victor M. Yakovenko (2000): "Statistical mechanics of money." The European Physical Journal B-Condensed Matter and Complex Systems 17.4: 723-729.

Efthimiou, Costas J. & Wearne Adam (2016): "Household income distribution in the USA." The European Physical Journal B 89.3: 82.

Farrior, C. E., Bohlman, S. A., Hubbell, S., & Pacala, S. W. (2016): “Dominance of the suppressed: Power-law size structure in tropical forests.” *Science*, *351*(6269), 155-157.

Fontanari, Andrea, Nassim Nicholas Taleb & Pasquale Cirillo (2018): "Gini estimation under infinite variance." Physica A: Statistical Mechanics and its Applications 502: 256-269.

Gabaix, Xavier (1999): "Zipf's law for cities: an explanation." The Quarterly Journal of Economics 114 (3), 739–767.

Gabaix, Xavier (2009): "Power laws in economics and finance." Annual Review of Economics 1 (1), 255–294.

Gabaix, Xavier & Landier, Augustin (2008): "Why has CEO pay increased so much?" The Quarterly Journal of Economics 123 (1), 49–100.

Gallegati, M., Keen, S., Lux, T., & Ormerod, P. (2006): “Worrying trends in econophysics.” Physica A: Statistical Mechanics and its Applications, 370(1), 1-6.

Gibrat, Robert (1931): "On economic inequalities'." INTERNATIONAL LIBRARY OF CRITICAL WRITINGS IN ECONOMICS 158: 497-514.

Gieryn, Thomas (1980): "Science and social structure: A festschrift for Robert K. Merton." New York academy of sciences.

Haldane, J. B. S. (1942): "Moments of the distributions of powers and products of normal variates." Biometrika 32.3/4: 226-242.

Harper, David A., & Paul Lewis (2012): "New perspectives on emergence in economics." 329-337.

Hayek, Friedrich August (1978): “New studies in philosophy, politics, economics, and the history of ideas.” University of Chicago Press.

Heckman, James J. & Christopher Taber (2010): "Roy model." Microeconometrics. Palgrave Macmillan, London, 2010. 221-228.

Heinrich, Torsten. (2013): "The ongoing history of economic conservation laws."

Iltasanomat, Taloussanomat.<https://www.is.fi/verotiedot/>, haettu 14.2020

Jagielski, Maciej & Ryszard Kutner (2013): "Modelling of income distribution in the European Union with the Fokker–Planck equation." Physica A: Statistical Mechanics and its Applications392.9: 2130-2138.

Jones, Charles I. (2015): "Pareto and Piketty: The macroeconomics of top income and wealth inequality." Journal of Economic Perspectives 29.1: 29-46.

Juran, Joseph & Godfrey, Blanton (1941) "Quality handbook." Republished McGraw-Hill.

Krueger, Alan B. (2019): “Rockonomics: A Backstage Tour of What the Music Industry Can Teach Us About Economics and Life.” Broadway Business.

Limpert, Eckhard, Werner A. Stahel, & Markus Abbt (2001): "Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question." BioScience 51.5 : 341-352.

Lydall, Harold F. (1959): "The distribution of employment incomes." Econometrica: Journal of the Econometric Society: 110-115.

Mandelbrot, Benoit (1963); "New methods in statistical economics." Journal of political economy 71.5: 421-440.

Mandelbrot, Benoit (1960): "The Pareto-Levy law and the distribution of income." International Economic Review 1.2: 79-106.

Merton, Robert K. (1968): "The Matthew effect in science." Science 159 (3810), 56–63.

Mitzenmacher, Michael (2004): "A brief history of generative models for power law and lognormal distributions." Internet mathematics 1.2: 226-251.

Pareto, Vilfredo (1964): "Manuale di Economia Politica." Vol. 13. Societa Editrice.

Pigou, Arthur Cecil (1932): “The economics of welfare”. Palgrave Macmillan.

Piketty, Thomas (2014): "Capital in the 21st Century." Harvard University Press 2017

Ricardo, David (1821): “On the principles of political economy”. London: J. Murray

Romer, Paul (2016): "The trouble with macroeconomics." The American Economist 20: 1-20.

Roy, Andrew Donald (1951): "Some thoughts on the distribution of earnings." Oxford economic papers 3.2: 135-146.

Roy, Andrew Donald (1950): "The distribution of earnings and of individual output." The Economic Journal 60.239: 489-505.

Schneider, P. A. (2008): "A Comparative Entropy Analysis of the Distribution of Wages and Salaries." The Annual UMass Amherst/NSSR Workshop.

Schumpeter, Joseph A. (1951): “Ten great economists.” Routledge.

Segarra, Agustí, & Mercedes Teruel (2012): "An appraisal of firm size distribution: Does sample size matter?." Journal of Economic Behavior & Organization 82.1: 314-328.

Shaikh, Anwar (2016): “Capitalism: Competition, conflict, crises.” Oxford University Press.

Shirras, G. Findlay (1935): "The Pareto Law and the distribution of income." The Economic Journal 45.180: 663-681.

Shockley, William (1957): "On the statistics of individual variations of productivity in research laboratories." Proceedings of the IRE45.3 279-290.

Simon, Herbert A. (1955): "On a class of skew distribution functions." Biometrika 42 (3/4), 425–440.

Stanovich, Keith E. (1986): "Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy." Reading research quarterly 21 (4), 360–407.

Taleb, Nassim Nicholas, & Raphael Douady (2015): "On the super-additivity and estimation biases of quantile contributions." Physica A: Statistical Mechanics and its Applications 429: 252-260.

Tao, Y., Wu, X., Zhou, T., Yan, W., Huang, Y., Yu, H., ... & Yakovenko, V. M. (2019): “Exponential structure of income inequality: evidence from 67 countries.” Journal of Economic Interaction and Coordination, 14(2), 345-376.

Ter Steege, H., Pitman, N. C., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., ... & Monteagudo, A. (2013): “Hyperdominance in the Amazonian tree flora.” Science, 342(6156), 1243092.

Uslay, Can, Z. Ayca Altintig, & Robert D. Winsor (2010): "An empirical examination of the “rule of three”: Strategy implications for top management, marketers, and investors." Journal of Marketing74.2: 20-39.

Verohallinto, Verohallinnon tilastotietokanta. <http://vero2.stat.fi/PXWeb/pxweb/fi/Vero/>, haettu 20.3.2020

Yakovenko, Victor M., & A. Christian Silva (2007): "Two-class structure of income distribution in the USA: Exponential bulk and power-law tail." Topological Aspects Of Critical Systems And Networks: (With CD-ROM). 49-58.

Yakovenko, Victor M., & J. Barkley Rosser Jr. (2009): "Colloquium: Statistical mechanics of money, wealth, and income." Reviews of modern physics 81.4: 1703.

Yle, Tässä ovat vuoden 2014 suurituloisimmat – katso koko maan ja maakuntien eniten tienanneet. <https://yle.fi/uutiset/3-8418674>, haettu 25.4.2020

Yle, Tässä ovat vuoden 2015 suurituloisimmat – katso koko maan ja maakuntien eniten tienanneet. <https://yle.fi/uutiset/3-9258065>, haettu 25.4.2020

Yle, Tässä ovat vuoden 2016 suurituloisimmat: Katso, kuka maksaa maakunnassasi eniten veroja. <https://yle.fi/uutiset/3-9908068>, haettu 25.4.2020

Yle, Tässä ovat vuoden 2017 suurituloisimmat – Jopa puolet kymmenestä kärkinimestä Supercellistä. <https://yle.fi/uutiset/3-10473276> , haettu 14.2020

Yu, Zonghuo, & Li Chen (2016): "Income and well-being: Relative income and absolute income weaken negative emotion, but only relative income improves positive emotion." Frontiers in psychology 7.

Yule, G. Udny (1925): "A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS." Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character 213 (1), 21–87.

Zipf, George Kingsley (1935): "The psycho-biology of language. An Introduction to Dynamic Philology." Houghton Mifflin.

Zipf, George Kingsley (1941): "National unity and disunity; the nation as a bio-social organism." Principia Press.

Zuckerman, Harriet (1977): "Scientific elite: Nobel laureates in the United States." Transaction Publishers.

8 R-KOODI

#Katkaistujen lognormaali- ja gammajakaumien muodostaminen välille (20,90)

```
#LOGNORMAALI
dtlnorm <- function(x, meanlog, sdlog, low, upp)
{
  PU <- plnorm(90, meanlog=meanlog, sdlog=sdlog)
  PL <- plnorm(20, meanlog=meanlog, sdlog=sdlog)
  dlnorm(x, meanlog, sdlog) / (PU-PL) * (x >= low) * (x <= upp)
}
ptlnorm <- function(q, meanlog, sdlog, low, upp)
{
  PU <- plnorm(90, meanlog=meanlog, sdlog=sdlog)
  PL <- plnorm(20, meanlog=meanlog, sdlog=sdlog)
  (plnorm(q, meanlog, sdlog)-PL) / (PU-PL) * (q >= low) * (q <= upp) + 1 * (q > upp)
}
#GAMMA
dtgamma <- function(x, shape, rate, low, upp)
{
PU <- pgamma(90, shape=shape, rate=rate)
PL <- pgamma(20, shape=shape, rate=rate)
dgamma(x, shape, rate) / (PU-PL) * (x >= low) * (x <= upp)
}
ptgamma <- function(q, shape, rate, low, upp)
{
  PU <- pgamma(90, shape=shape, rate=rate)
  PL <- pgamma(20, shape=shape, rate=rate)
  (pgamma(q, shape, rate)-PL) / (PU-PL) * (q >= low) * (q <= upp) + 1 * (q > upp)
}
```

#Vuoden 2014 palkkajauman parametrien suurimman uskottavuuden estimointi (maximum likelihood method)

```
#LOGNORMAALI
tlognormaali1 <- fitdist(rep(income/1000,y1/1000), "tlnorm", method = "mle", start = list(meanlog=3.5, sdlog=0.4), fix.arg=list(low=min(rep(income/1000,y1/1000)), upp=max(rep(income/1000,y1/1000))))

#GAMMA
tgamma1 <- fitdist(rep(income/1000,y1/1000),"tgamma", method = "mle", lower=c(0,0), start =list(shape=4.5,rate=0.1), fix.arg=list(low=min(rep(income/1000,y1/1000)), upp=max(rep(income/1000,y1/1000))))
```

#2014 Ansiotulot ja sovitetut jakaumat -kuvio

```
denscomp(list(tlognormaali1, tgamma1), probability = FALSE, xlim = c(22.5, 87.5), fitcol = c(1,10), datacol = "grey", legendtext = NULL, xlab = "Tuloluokka (x1000€)", main = "Palkkatulot 2014 ja sovitettu jakauma", ylab = "Lukumäärä (x1000)")
legend("topright", c("Lognormaali", "Gamma"),col = c("black","red"),lwd = c(2,2),lty = c(1,2))
```

#2015-2018 Palkkatulot -kuviot

```
par(mfrow=c(2,2))
```

denscomp(tlognormaali2, xlim = c(22.5,87.5), datacol = "grey", xlab = "Tuloluokka (x1000€)", legendtext = NULL, main = 2015, ylab = "Lukumäärä (x1000)",probability = FALSE)
denscomp(tlognormaali3 ,xlim = c(22.5,87.5), datacol = "grey", xlab = "Tuloluokka (x1000€)", legendtext = NULL, main = 2016, ylab = "Lukumäärä (x1000)",probability = FALSE)
denscomp(tlognormaali4, xlim = c(22.5,87.5), datacol = "grey", xlab = "Tuloluokka (x1000€)", legendtext = NULL, main = 2017, ylab = "Lukumäärä (x1000)",probability = FALSE)
denscomp(tlognormaali5, xlim = c(22.5,87.5), datacol = "grey", xlab = "Tuloluokka (x1000€)", legendtext = NULL, main = 2018, ylab = "Lukumäärä (x1000)",probability = FALSE)

#Jakaumien ja aineiston yhteensopivuus
gofstat(list(tgamma1,tlognormaali1))
gofstat(list(cgamma1,clognormaali1))

#2014 Pääomatulot ja sovitetut jakaumat -kuvio
denscomp(list(clognormaali1, cgamma1), breaks = 18, xlim = c(7.5, 97.5), datacol = "grey", xlab = "Tuloluokka (x1000€)", legendtext = NULL, main = "Pääomatulot 2014 ja sovitettu jakauma", ylab = "Lukumäärä (x1000)", probability = FALSE,fitcol = c(1,10))
legend("topright", c("Lognormaali", "Gamma"),col = c("black","red"),lwd = c(2,2),lty = c(1,2))