

*Pellervon taloudellisen tutkimuslaitoksen
työpapereita*

*Pellervo Economic Research Institute
Working Papers*

N:o 81 (Tammikuu 2006)

**PALVELUJEN ALUEELLINEN
TUOTANTO SUOMESSA**

Mikko Pakkanen



Pellervon taloudellinen tutkimuslaitos PTT

Pellervo Ekonomiska Forskningsinstitut

Pellervo Economic Research Institute

Eerikinkatu 28 A, 00180 Helsinki

Puh. (09) 348 8844, fax (09) 3488 8500

E-mail: econ.res@ptt.fi

www.ptt.fi

*Pellervon taloudellisen tutkimuslaitoksen
työpapereita*

*Pellervo Economic Research Institute
Working Papers*

N:o 81 (Tammikuu 2006)

**PALVELUJEN ALUEELLINEN
TUOTANTO SUOMESSA***

Mikko Pakkanen

Helsinki, Tammikuu 2006

* Kiitän Janne Huovaria ja Raija Volkia hyödyllisistä kommenteista ja parannusehdotuksista tutkimuksen eri vaiheissa. Työ liittyy Valtioneuvoston kanslian rahoittamaan palvelualan selvitykseen ”Palvelualojen kehitys, tuottavuus ja kilpailu”, joka on julkaistu Valtioneuvoston kanslian julkaisusarjassa 11/2005. Tuloksia on esitelty myös ao. raportin artikkelissa: Janne Huovari, Mikko Pakkanen, Raija Volk ”Palveluiden alueellinen tarjonta ja tuottavuus.” Pellervon taloudellinen tutkimuslaitos kiittää VNK:aa rahoituksesta ja projektin eri vaiheissa saaduista kommenteista hankkeen ohjausryhmää, muita hankkeen tekemiseen osallistuneita sekä erityisesti Vesa Vihriälää ja Pekka Sinkoa VNK:sta.

ISBN 952-5594-08-4 (PAP)
ISBN 952-5594-09-2 (PDF)
ISSN 1455-4623

Pellervon taloudellinen tutkimuslaitos PTT
Pellervo Economic Research Institute PTT
Eerikinkatu 28 A
00180 Helsinki

Helsinki 2006

MIKKO PAKKANEN. 2006. PALVELUJEN ALUEELLINEN TUOTANTO SUOMESSA. Pellervon taloudellisen tutkimuslaitoksen työpapereita n:o 81. 25 s. ISBN 952-5594-08-4 (PAP), ISBN 952-5594-09-2 (PDF), ISSN 1455-4623.

Tiivistelmä: Tutkimuksessa tarkastellaan alueellisesta näkökulmasta yksityisten palvelujen tuotantoa Suomessa. Tulokset tukevat ennako-odotuksia siitä, että paikkasidonaiset alat, kuten vähittäiskauppa, majoitus- ja ravitsemustoiminta, maaliikenne sekä henkilökohtaiset palvelut ovat väestö huomioon ottaen tasaisesti jakautuneita kaikkialle maahan ja vähemmän paikkaan sidotut tietointensiiviset palvelut kuten tietojenkäsittely sekä tutkimus ja kehittäminen puolestaan maantieteellisesti keskittyneitä. Väkirikkaimmissa kunnissa on palveluiden tuotantoa myös väestöön suhteutettuna enemmän kuin pienissä kunnissa.

Muutamilla toimialoilla tarkasteltiin lisäksi ordered logit-mallin avulla, mitkä ovat kynnysväkiluvut sille, että alan toimipaikkoja esiintyy kunnassa. Useimmilla tarkastelluista toimialoista kynnysväkiluvut kasvavat likimäärin suorassa suhteessa toimipaikkojen määrään eli toimipaikkojen kaksinkertaistumiseksi tarvitaan kaksinkertainen väestö. Lisäksi tarkasteltiin palvelujen tuotannon riippuvuutta etäisyyksistä muista asutuskeskuksista. Useimmilla tarkastelluista toimialoista merkitsevää riippuvuutta ei esiinny. Muutamilla toimialoilla, kuten vähittäiskaupassa sekä majoitus- ja ravitsemustoiminnassa suurten väestökeskittymien läheisyys vähentää tarjontaa.

Avainsanat: palvelujen tarjonta, keskittyminen, maantieteellinen läheisyys

MIKKO PAKKANEN. 2006. REGIONAL ASPECTS ON THE SERVICE PRODUCTION IN FINLAND. Pellervo Economic Research Institute Working Papers No. 81. 25 p. ISBN 952-5594-08-4 (PAP), ISBN 952-5594-09-2 (PDF), ISSN 1455-4623

Abstract: This study examines the production of private services in Finland from regional point of view. The results support the expectations that services bounded to local markets sectors, such as retailing, hotels and restaurants, land transport and personal services are evenly produced in relation to the population across the country and knowledge intensive services, such as computer and related services or research and development are geographically concentrated. Usually big municipalities provide more service production than small ones also in relation to the population.

An ordered logit model was used in order to define threshold population for the existence of the first, second etc. outlet. In the most sectors the threshold population increases in the same proportion as the number of outlets, that is, when the number of population doubles also the number of production units in the respective sector doubles.

In addition, the dependence of the production of the proximity to population centres was studied. In most sectors there was no significant dependence. In some sectors, like retailing, hotels and restaurants the vicinity of big cities decreases the production.

Keywords: supply of services, concentration, geographic proximity

SISÄLTÖ

Tiivistelmä	i
Johdanto	1
1 Palvelujen keskittyminen alueittain	2
1.1 Tutkimusmenetelmä	2
1.2 Tuloksia	3
2 Palvelujen tarjonnan riippuvuus etäisyyksistä ja väestöstä	6
2.1 Väkiluvun ja tarjonnan yhteys	6
2.1.1 Motivointi	6
2.1.2 Malli työntekijöiden määrille	6
2.1.3 Malli toimipaikkojen määrille	11
2.1.4 Malli ja väkilukukynnykset toimipaikkojen esiintymiselle.....	13
2.2 Etäisyyksien ja tarjonnan yhteys	17
3 Yhteenveto	20
Lähteet	21
Liite	22

TIIVISTELMÄ

Palvelut työllistävät Suomessa noin kaksi kolmesta työssä käyvästä. Osa palveluista liittyy kiinteästi tavaratuotantoon, osa puolestaan on suunnattu suoraan kuluttajille. Jälkimmäisessä tapauksessa palvelujen tuotanto sijoittuu yleensä lähelle asiakasta. Suomen harva väestörakenne ja monien alueiden väestön väheneminen herättää kuitenkin kysymyksen siitä, ovatko palvelut enää sijoittuneet sinne missä väestö sijaitsee vai ovatko ne keskittyneet.

Monet palvelut ovat tässä saatujen tulosten mukaan jakautuneet tasaisesti eri puolille maata. Ennako-odotusten mukaan paikkasidonnaiset alat, kuten vähittäiskauppa, majoitus- ja ravitsemistoiminta, maaliikenne sekä henkilökohtaiset palvelut ovat varsin tasaisesti jakautuneita ja vähemmän paikkaan sidotut alat, kuten tietojenkäsittely sekä tutkimus ja kehittäminen, puolestaan keskittyneitä. Hieman yllättäen pankkitoiminta kuuluu tasaisemmin jakautuneisiin toimialoihin konttorien merkittävästä vähentymisestä huolimatta. Pankkitoiminnassa varsin pieni väestömäärä riittää pitämään yhden toimipaikan alueella, mikä selittää tasaisen jakautumisen.

Tarkempi tarkastelu osoittaa kuitenkin, että tarjontaa on useimmilla toimialoilla väkilukuun suhteutettuna enemmän väkiluvultaan suurissa kunnissa. Selvimmin tämä tulee esiin tukkukaupassa, rahoitusta palvelevassa toiminnassa ja tietojenkäsittelypalveluissa.

Maaliikenteessä ja pankkitoiminnassa sen sijaan palvelujen tuotantoa on melko tarkalleen samassa suhteessa kuin väestöä. Maaliikenteessä näyttää toimipaikkoja olevan väkiluvultaan pienemmissä kunnissa itse asiassa jopa enemmän kuin suurissa kunnissa. Tämä heijastanee sitä, että pieniä kuljetusyrityksiä on runsaasti eri puolilla maata.

Paikalliset väestökeskittymät ovat usein niin pieniä, ettei niissä ole kovin monta toimipaikkaa. Useimmilla toimialoilla toimipaikkojen määrä kasvaa suunnilleen samassa tahdissa väestömäärän kanssa, mutta ensimmäisen toimipaikan esiintymiselle tarvitaan luonnollisesti erilainen väestömäärä. Pankkialalla esimerkiksi yksi toimipaikka on hyvinkin pienissä kunnissa (jopa 400 asukasta riittää toimipaikan ylläpitoon), mutta useimmat toimipaikat ”edellyttävät” suhteessa huomattavasti suurempaa (seitsenkertaista) asukasmäärää. Myös posti- ja teleliikenteessä useaan toimipaikkaan vaadittava väestömäärä kasvaa varsin jyrkästi. Ensimmäinen toimipaikka alkaa esiintyä kuitenkin keskimäärin vasta noin 2300 asukkaan kohdalla.

Kunnan oman väkiluvun lisäksi myös ympäröivien paikkakuntien väestömäärä voi vaikuttaa palvelujen tarjontaan. Saatujen tulosten mukaan useimmilla toimialoilla sijainti suhteessa muihin väestökeskittyymiin ei vaikuta merkittävästi palvelujen tarjontaan. On

kuitenkin muutamia aloja, kuten vähittäiskauppa sekä ravitsemus- ja ravitsemistoiminta, joilla suurten väestökeskittymien läheisyys vähentää tarjontaa. Tukkukaupassa sen sijaan riippuvuus on päinvastainen, minkä voisi tulkita tarkoittavan sitä, että tukkukauppa sijoittuu väkimäärältään suurten kuntien läheisyyteen, muttei kuitenkaan näiden kuntien alueille.

Johdanto

Ruokakauppa voi sijaita korkeintaan lyhyen automatkan päässä asiakkaasta, kun taas tietotekniikkayritys voi helposti myydä kehittämänsä ohjelmiston asiakkaalle toiselle puolelle maata tapaamatta tätä koskaan. Onkin selvää, että alueellisesta näkökulmasta katsottuna palvelutoimialat voivat olla hyvin erilaisia. Keskeisenä tekijänä on se, miten palvelun tuottaja ja asiakas kohtaavat toisensa. Alat, joilla palvelun ostajan on välttämätöntä tavata palvelun tuottaja henkilökohtaisesti, ovat suurelta osin *paikkasidonnaisia*. Nämä alat ovat usein jakautuneet alueellisesti melko tasaisesti koko maahan. Toisaalta taas aloilla, joilla henkilökohtainen kontakti ei ole välttämätöntä, on asiakassuhteiden kannalta miltei samantekevää, missä toimipaikat sijaitsevat. Tällöin muut syyt, kuten pätevän työvoiman saatavuus tai liikenneyhteydet, voivat johtaa toimintojen keskittymiseen. Tietotekniikan kehityksen ansiosta yrityksen sijaintipaikan merkitys on vähentynyt myös monilla perinteisesti paikkasidonnaisilla aloilla. Esimerkiksi eräät vähittäiskaupan alat, matkatoimistot ja pankit ovat voineet siirtää merkittävän osan toiminnastaan sähköiseen muotoon. Kaikilla aloilla tämänkaltainen kehitys ei tietenkään ole mahdollista, esimerkiksi kampaamopalvelut vaativat luonnollisesti asiakkaan läsnäoloa, ja siivous on suoritettava asiakkaan tiloissa. Lisäksi erilaisia kuljetuspalveluita tarvitaan miltei kaikkialla, missä on edes vähäistä asutusta.

Tutkimuksessa on tarkasteltu *yksityisten* palvelutoimialojen tarjontaa Suomessa. Tavoitteena on ollut selvittää ensinnäkin, millä aloilla tarjonta on jakautunut tasaisesti koko maahan, millä aloilla tarjonta on keskittyneempää, ja miten toimialat eroavat toisistaan alueellisen keskittymisen kannalta. Tarjonnan riippuvuutta alueellisesta väestöstä on tarkasteltu kahdella eri tavalla: estimoimalla *tarjonnan joustoja väkiluvun suhteen* sekä tutkimalla, miten tarjonnan (toimipaikkojen) *esiintymistodennäköisyydet* riippuvat väkiluvusta. Lisäksi on perehdytty tarjonnan riippuvuuteen maantieteellisistä etäisyyksistä selvittäen, miten tarjonta on yhteydessä alueen sijaintiin suhteessa suuriin väestökeskittymiin. Aineistona on käytetty Tilastokeskuksen tuottamia yritysrekisterin tietoja yritysten toimipaikoista kunnissa eri toimialoilla vuosilta 2002 ja 2003, työssäkäyntitilaston tietoja toimialoittaisista työntekijöiden määristä kunnissa vuodelta 2002 sekä vuoden 2004 väkilukutietoja. Koska työssäkäyntitilaston tiedot sisältävät myös julkisen sektorin työntekijöitä, tarkasteltavien toimialojen joukosta poistettiin ne toimialat, joilla julkisen sektorin työntekijöiden määrällä on merkitystä. Näihin kuuluvat mm. koulutus, sosiaali- ja terveystieteiden palvelut sekä julkinen hallinto. Jäljelle jääneillä toimialoilla julkisen palveluntuotannon osuudet lienevät erittäin pieniä.

Ensimmäisessä luvussa käsitellään palvelutoimialojen alueellista keskittymistä. Toisessa luvussa tarkastellaan tarjonnan riippuvuutta väestöstä tarkastellen erikseen työntekijä- ja toimipaikkamäärien kannalta. Lisäksi perehdytään tarjonnan riippuvuuteen etäisyyksistä suhteessa väestökeskittymiin. Yhteenveto tuloksista on esitetty kolmannessa luvussa ja liitteessä käsitellään muutamia tutkimukseen liittyviä teknisiä yksityiskohtia.

1 Palvelujen keskittyminen alueittain

1.1 Tutkimusmenetelmä

Tässä luvussa tarkastellaan, miten palvelutoimialojen tarjonta on jakautunut Suomessa. Tarjontaa arvioidaan sekä toimipaikkojen että työntekijöiden määrien suhteen. Kullekin toimialalle on kunnittain laskettu kunnan väkilukuun suhteutetut toimipaikkojen ja työntekijöiden määrät. Niiden avulla toimialoittaista palvelujen keskittymistä on arvioitu näille suhteellisille toimipaikka- ja työntekijämäärille laskettujen *Gini-kertoimien* avulla. Gini-kertoimien ohella on havainnollista tarkastella toimialoittain, monestako kunnasta alan tarjonta puuttuu kokonaan. Lisäksi toimipaikka- ja työntekijäaineistojen tietoja yhdistelemällä saadaan informaatiota toimipaikkojen keskikoosta sekä toimipaikkojen keskikokojen vaihte- lusta eri kuntien välillä.

Gini-kerroin on nollan ja yhden väliltä arvoja saava tilastollinen tunnusluku, jota on perinteisesti käytetty mm. tuloerotutkimuksessa epätasaisuuden mittarina. Tällöin tulkinta on tunnetusti seuraava: mikäli kaikilla henkilöillä otoksessa on yhtä suuret tulot, kerroin saa arvon nolla, ja vastaavasti kun yksi henkilö saa kaikki tulot ja muut eivät mitään, arvo on yksi. Arvioitaessa palvelujen tarjonnan keskittymistä eri ole tietenkään mielekästä laskea Gini-kerrointa suoraan absoluuttisille tarjonnoille kunnissa, sillä eihän ideaalitapauksessa palvelujen tarjonta olisi suinkaan yhtä suurta kaikissa kunnissa. Siksi kunnittaiset tarjonnan suhteutetaan kuntien väkilukuihin. Näihin lukuihin sovellettuna Gini-kerroin on mielekäs mittari, sillä arvo nolla vastaa tilannetta, jossa toimialan palvelujen *suhteellinen tarjonta* on yhtä suurta kaikissa kunnissa, ja arvo yksi sitä, että koko toimialan tarjonta on keskittynyt yhteen kuntaan. Keskittymisen mittaamiseen usein käytetyillä *Herfindahl-indekseillä* päädyttäisiin käytyn aineiston tapauksessa oleellisesti samoihin tuloksiin arvioitaessa suhteellisen tarjonnan keskittymistä.

Suhteellisten tarjontojen käyttämiseen tässä yhteydessä sisältyy kuitenkin potentiaalinen ongelma: hyvin pienen väestön omaavissa kunnissa erityisesti toimialoilla, joilla työ- ja toimipaikkoja on kaikissa kunnissa melko vähän, voivat suhteelliset tarjonnat muodostua keinotekoisien suuriksi. Tällöin laskettu Gini-kerroin voi olla jossain mielessä vääristynyt. Toisaalta voidaan myös ajatella, että aineiston taustalla on populaation Gini-kerroin G ja aineistosta laskettu tunnusluku \hat{G} on sen estimaattori. Mutta kun käytössä on kaikkien kuntien tiedot, eli periaatteessa kokonaisotanta, väistämättä $\hat{G} = G$, eli estimaatti ja populaatioparametri yhtyvät. Tämän argumentin voi kuitenkin väistää sillä, että kiinnostuksena kohteena ei varsinaisesti olekaan G . Tarkasteltaessa alueellisen tarjonnan kokonaiskuvaa muutamien pikkukuntien havainnot eivät oikeastaan ole kiinnostavia, ja onkin luontevampaa pohtia, miten Gini-kerroin (estimaatti) vaihtelisi, jos kuvitteellinen otanta tuottaisikin erilaisia aineistoja, joissa näitä poikkeavia havaintoja ei ehkä olisi. Tämän motivoinnin kannalta Gini-kertoimen vakauden arvioinnissa voidaan käyttää keskivirheitä, joilla ei siis tässä tapauksessa ole perinteistä otannan epävarmuudesta johdettua tulkintaa.

Taulukko 1: Palvelujen tarjonnan alueellinen keskittyminen, Gini-kertoimet ja toimi- sekä työpaikkojen puuttuminen kunnista.

Toimiala	Työntekijät		Toimipaikat	
	Gini ^a	Ei tarjontaa	Gini ^a	Ei tarjontaa
Autokauppa	0,322 (0,029)	10	0,260 (0,028)	13
Tukkukauppa	0,536 (0,043)	27	0,338 (0,027)	22
Vähittäiskauppa	0,257 (0,035)	0	0,203 (0,028)	0
Majoitus- ja ravitsemistoiminta	0,395 (0,034)	3	0,346 (0,038)	3
Maaliikenne	0,257 (0,030)	0	0,211 (0,028)	0
Vesiliikenne	0,964 (0,547)	292	0,928 (0,194)	327
Ilmaliikenne	0,987 (0,560)	400	0,964 (0,229)	398
Liikennettä palveleva toiminta	0,752 (0,088)	106	0,587 (0,040)	123
Posti- ja teleliikenne	0,412 (0,034)	19	0,444 (0,036)	62
Rahoituksen välitystoiminta	0,270 (0,033)	4	0,370 (0,038)	4
Vakuutustoiminta	0,738 (0,062)	206	0,668 (0,032)	220
Rahoitusta palveleva toiminta	0,745 (0,078)	231	0,708 (0,048)	227
Kiinteistöalan palvelut	0,335 (0,026)	17	0,336 (0,027)	36
Vuokrauspalvelut	0,781 (0,129)	230	0,620 (0,028)	190
Tietojenkäsittelypalvelu	0,797 (0,111)	167	0,601 (0,044)	154
Tutkimus ja kehittäminen	0,901 (0,234)	271	0,921 (0,140)	369
Muu liike-elämää palveleva toiminta	0,324 (0,031)	3	0,253 (0,028)	4
Ympäristönhuolto	0,557 (0,030)	99	0,538 (0,025)	109
Vapaa-ajan palvelut	0,403 (0,033)	5	0,415 (0,028)	64
Muut palvelut	0,237 (0,029)	5	0,218 (0,027)	14

^aKeskivirheiden estimaatit suluissa, raportoitu arvo on $\max\{s.e.(\hat{G}_1), s.e.(\hat{G}_2)\}$, kts. liite. Kovarianssimatriisit on estimoitu käyttäen HC3-estimaattoria (MacKinnon-White, 1985).

1.2 Tuloksia

Yleisesti ottaen taulukossa 1 esitetyt tulokset¹ tukevat vahvasti intuitiivisia ennako-odotuksia siitä, että yhtäältä asiakkaan henkilökohtaista kohtaamista edellyttävät, perinteisesti vahvasti paikkasidonnaiset toimialat olisivat tasaisesti ja kautuneita koko maahan, ja toisaalta sitä, että modernia informaatioteknologiaa hyödyntävät vähemmän paikkaan sidotut alat olisivat vastaavasti enemmän keskittyneitä tietyille alueille. Johtopäätökset sekä väkilukuun suhteutetuista työpaikkojen että työntekijöiden määristä ovat kaiken kaikkiaan hyvin samansuuntaisia. Keskivirheet ovat pääosin varsin pieniä, joten kertoimet eivät siis vaikuttaisi olevan vääristyneitä. Vesi- ja ilmaliikenteen sekä tutkimuksen ja kehittämisen kertoimien virheet ovat tosin suhteellisen suuria, mutta toisaalta näiden alojen tarjonta myös puuttuu useimmista kunnista, joten suurehkot Gini-kertoimet ovat joka tapauksessa odotettuja.

Vähittäiskaupan toimipaikkojen on luonnollisesti sijaittava lähellä asiakkaita.

¹Tutkimuksen tilastolliset analyysit suoritettiin R-ohjelmiston versioilla 2.0.1, 2.1.0 ja 2.1.1 (R Development Core Team, 2005).

Taulukko 2: Toimipaikkojen keskikoot, työntekijät toimipaikkaa kohden.

Toimiala	Koko maassa	Kuntien keskikokojen hajonta
Autokauppa	4,06	1,38
Tukkukauppa	5,23	1,88
Vähittäiskauppa	4,95	1,73
Majoitus- ja ravitsemistoiminta	5,19	2,29
Maaliikenne	3,79	1,38
Vesiliikenne	28,30	18,86
Ilmaliikenne	49,95	41,94
Liikennettä palveleva toiminta	10,69	6,29
Posti- ja teleliikenne	26,54	12,08
Rahoituksen välitystoiminta	10,25	3,70
Vakuutustoiminta	13,99	7,87
Rahoitusta palveleva toiminta	3,03	1,19
Kiinteistöalan palvelut	3,28	2,25
Vuokrauspalvelut	2,88	3,82
Tietojenkäsittelypalvelu	7,91	3,36
Tutkimus ja kehittäminen	45,24	115,95
Muu liike-elämää palveleva toiminta	5,51	1,90
Ympäristöhuolto	8,28	6,14
Vapaa-ajan palvelut	8,26	3,59
Muut palvelut	1,70	0,62

Siksi onkin loogista, että vähittäiskauppa on yksi kaikkein tasaisimmin jakautuneista toimialoista. Vähittäiskaupan toimipaikkoja on jokaisessa Suomen kunnassa. Majoitus- ja ravitsemustoiminta on paikkasidonnaisena myös varsin tasaisesti jakautunutta. Hieman erikoistuneemmat kaupan alat, kuten autokauppa sekä tukkukauppa eivät luonnollisesti ole aivan yhtä sidonnaisia paikkaan kuin vähittäiskauppa, joten ne ovatkin hieman keskittyneempiä. Toimipaikat työllistävät näillä aloilla keskimäärin 4–5 työntekijää, eikä toimipaikkojen keskikoossa ole havaittavissa suurta alueellista vaihtelua (taulukko 2).

Liikennealoilla alueellinen keskittyminen riippuu keskeisesti liikenteen luonteesta. Maaliikenteen palveluja tarvitaan koko maassa, joten se on hyvin tasaisesti jakautunutta, toimipaikkoja on kaikissa Suomen kunnissa. Maaliikenteen toimipaikat ovat pienehköjä koko maassa työllistäen keskimäärin 3–4 henkeä. Toisaalta taas vesiliikenne on luonnollisesti keskittynyt lähinnä rannikolle ja ilmaliikenne lentokenttäpaikkakunnille, erityisesti lentotoiminnan työpaikoista suurin osa Vantaalle. Toimipaikkojen keskikokojen vaihtelu on hyvin suurta kuntien välillä, mikä ei ole yllättävää, sillä näillä aloilla on sekä suuria varustamoja ja lentoyhtiöitä että pienempiä paikallisia vesiliikenteen harjoittajia ja yleisilmailuun keskittyneitä pienyrityksiä. Liikennettä palveleva toiminta sekä matkatoimistot ovat myös kohtalaisen keskittyneitä. Posti- ja teleliikenteen kohdalla erityisesti posti-toimintaa on vielä melko tasaisesti koko maassa, joskin teleliikenteen voisi otaksua olevan keskittyneempää.

Verkkopankkitoiminnan yleistymisestä ja konttoriverkostojen voimakkaasta karsimisesta huolimatta rahoituksen välitystoimintaa on vieläkin varsin tasaisesti koko maassa. Ainoastaan neljässä Ahvenanmaan saaristokunnassa ei ole lainkaan rahoituksen välityksen toimipaikkoja. Toimipaikkojen keskimääräisissä työntekijämäärissä ei esiinny suurta vaihtelua alueiden välillä. Sen sijaan vakuutustoiminta ja rahoitusta palveleva toiminta ovat huomattavasti keskittyneempiä; miltei puolessa Suomen kunnista ei ole lainkaan näiden alojen omia toimipaikkoja.

Kiinteistöalan palveluille ja pääasiassa henkilökohtaisia palveluja käsittävälle muiden palvelujen kategorialle on yhteistä se, että molemmat ovat varsin paikkasidonniaisia. Erityisesti henkilökohtaiset palvelut ovat hyvin tasaisesti jakautuneita, ja toimipaikat ovat koko maassa pieniä, keskimäärin muutaman työntekijän työllistäviä.

Tietojenkäsittelypalvelut ovat luonnollisesti vähemmän paikkasidonniaisia ja teknologiaa hyödyntävänä keskittyneitä. Myös tutkimus- ja kehitystoiminta on erittäin keskittynyttä; toimipaikkoja on vain noin joka seitsemännessä kunnassa. Erityisen huomattavaa on se, miten toimipaikkojen keskikoot vaihtelevat alueittain. Kaupunkien keskimäärin suurempien yksiköiden vastapainona ovat maaseudun kuntien usein vain muutaman henkilön työllistävät pienemmät toimipaikat. Osittain hajontaa selittää tosin myös Jokioisten kunnan ainoa tutkimuksen ja kehittämisen toimipaikka, yli 500 henkeä työllistävä Maa- ja elintarviketalouden tutkimuskeskus, joka on selvästi poikkeava havainto aineistossa.

2 Palvelujen tarjonnan riippuvuus etäisyyksistä ja väestöstä

Tässä osassa tarkastellaan, miten palvelujen tarjonta vaihtelee kunnissa väkiluvun ja kunnan sijainnin mukaan. Erityisesti vastausta haetaan kysymykseen, onko väkiluvultaan suurissa kunnissa tarjontaa väkilukuun suhteutettunakin enemmän kuin pienissä. Lisäksi tutkitaan, kuinka suurta väestöpohjaa eräiden palvelualojen toimipaikkojen esiintyminen kunnassa edellyttää. Kunnan sijainnin osalta selvitetään, minkälainen yhteys on palvelujen tarjonnalla ja kunnan sijainnilla suhteessa suuriin väestökeskittyymiin.

2.1 Väkiluvun ja tarjonnan yhteys

2.1.1 Motivointi

Väkiluvun ja palvelutoimialan tarjonnan yhteys kunnassa riippuu merkittävästi toimialan luonteesta. Paikkasidonnaisille palveluille keskeistä on kunnassa oleva asiakas pohja. Laaja asiakas pohja mahdollistaa myös tarjonnan erikoistumisen, mikä osaltaan lisää tarjonnan määrää. Väkiluvultaan suuriin kuntiin on myös paremmat liikenne yhteydet, jolloin palvelut pystyvät paremmin houkuttelemaan asiakkaita myös kunnan ulkopuolelta.

Vähemmän paikkasidonnaisissa palveluissa yhteys voi olla erilainen. Palvelutoimiala saattaa esimerkiksi tarvita korkeammin koulutettua työvoimaa, jolloin sen on luontevaa sijoittaa paikkakunnalle, jossa on relevantin koulutuksen tarjontaa. Toisaalta koulutus on taas vahvasti keskittynyt väkiluvultaan suuriin kuntiin. Myös mikäli palvelutoimiala on vahvasti sidoksissa johonkin muuhun toimialaan, esimerkiksi teollisuuteen, on sen hyödyllistä sijoittaa lähelle sitä, mikä voi osaltaan lisätä keskittymistä. Vaikkei toimiala olisikaan erityisen paikkasidonnainen, ei liikenne yhteyksien merkitystään pidä vähätellä. Erityisesti tietoliikenne yhteyksien laatu voi olla keskeinen kriteeri varsinkin pienten yritysten valitessa sijaintipaikkaansa, mikä selvästi lisää kaupunkien suosiosta, sillä väkiluvultaan pienissä kunnissa, mahdollisesti harvaan asutulla alueella, tietoliikenneinfrastruktuuri on selvästi vähemmän kehittynyttä.

2.1.2 Malli työntekijöiden määrille

Olkoot $E_{i,j}$ työntekijöiden määrä toimialalla i kunnassa j ja P_j kunnan j väkiluku. Silmäilemällä työntekijöiden määrän ja väkiluvun hajontakuvioita voi päätellä, että toimialoittain riippuvuutta voisi mallintaa epälineaarilla regressiolla

$$E_{i,j} = \varphi_i P_j^{\delta_i} + e_{i,j}. \quad (1)$$

Epälineaarisen regressio-ongelman ratkaiseminen on kuitenkin oleellisesti lineaarista tapausta hankalampaa edellyttäen alkuarvojen antamista ja konvergenssin

tutkimista. Jo toimialojen määrän vuoksi on käytännöllisempää käsitellä mallin (1) log-linearisoitua² muotoa

$$\log E_{i,j} = \alpha_i + \delta_i \log P_j + \varepsilon_{i,j}, \quad (2)$$

missä $\alpha_i \doteq \log \varphi_i$. Parametri δ_i on toimialan i työntekijöiden määrän jousto väkiluvun suhteen. Jos $\delta_i = 1$, toimialan i palvelujen tarjontaa, arvioituna työntekijöiden määrällä, suhteessa väkilukuun on keskimäärin yhtä paljon niin suurissa kuin pienissäkin kunnissa. Parametroimalla mallia uudelleen, siten että $\beta_i \doteq \delta_i - 1$, saadaan estimoinnin kannalta täysin yhtäpitävä muoto

$$\log \frac{E_{i,j}}{P_j} = \alpha_i + \beta_i \log P_j + \varepsilon_{i,j}, \quad (3)$$

jonka avulla voidaan testata luontevaa nollahypoteesia yksikköjoustavuudesta, eli $H_0 : \beta_i = 0$.

On syytä hieman pohdiskella, miten mallin (3) regressiokertoimien estimaatteja tulisi tulkita. Malli on erittäin voimakas yksinkertaistus siinä mielessä, että tarjontaan vaikuttaa väkiluvun ohella varmaan lukemattomia muitakin tekijöitä, jotka voivat olla korreloituneita väkiluvun kanssa. Merkintöjen yksinkertaistamiseksi jätetään tässä vaiheessa tarpeettomat indeksit pois. Tarkastellaan tilannetta, jossa todellinen riippuvuus onkin

$$\log \frac{E}{P} = \tilde{\alpha} + \tilde{\beta} \log P + \boldsymbol{\theta}' \mathbf{x} + u, \quad (4)$$

missä $\mathbf{x} = (x_1, \dots, x_k)$ on muista tarjontaan vaikuttavista tekijöistä koostuva vektori ja $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ sen kerroinvektori. Todellisuudessa \mathbf{x} :n dimensio ei siis varmaankaan ole äärellinen, mutta riittävän suuri k tarjoaa luultavasti hyvän ap-proksimaation. Koska \mathbf{x} :n komponentit voivat olla keskenään tai P :n logaritmin kanssa korreloituneita, tarvitaan jokaista x_i :tä kohti apuregressio

$$x_i = \gamma_{i,0} + \rho_i \log P + \sum_{\substack{j=1 \\ i \neq j}}^k \gamma_{i,j} x_j + v_i.$$

Kootaan apuregressiot, ilmeisin merkinnöin, matriisimuotoon

$$\mathbf{x} = \boldsymbol{\gamma}_0 + \boldsymbol{\rho} \log P + \boldsymbol{\Gamma} \mathbf{x} + \mathbf{v},$$

missä $(\boldsymbol{\Gamma})_{i,j} = \gamma_{i,j}$, kun $i \neq j$ ja $(\boldsymbol{\Gamma})_{i,i} \equiv 0$. Tällöin

$$(\mathbf{I}_k - \boldsymbol{\Gamma}) \mathbf{x} = \boldsymbol{\gamma}_0 + \boldsymbol{\rho} \log P + \mathbf{v}.$$

²Tarkalleen ottaen virhetermin pitäisi mallissa (1) olla tietenkin multiplikatiivinen, eikä additiivinen, jotta puolittaisella logaritmoinnilla todellakin päädyttäisiin malliin (2).

Jos $I_k - \Gamma$ on täysiasteinen, voidaan x ratkaista:

$$x = (I_k - \Gamma)^{-1}(\gamma_0 + \rho \log P + v)$$

Regressio (3) saadaan nyt sijoittamalla ratkaisu yhtälöön (4):

$$\log \frac{E}{P} = \underbrace{\tilde{\alpha} + \theta'(I_k - \Gamma)^{-1}\gamma_0}_{=\alpha} + \underbrace{(\tilde{\beta} + \theta'(I_k - \Gamma)^{-1}\rho)}_{=\beta} \log P + \underbrace{u + \theta'(I_k - \Gamma)^{-1}v}_{=\varepsilon} \quad (5)$$

Selvästi $\beta = \tilde{\beta}$, jos ja vain jos $\theta = \mathbf{0}$ tai $\rho = \mathbf{0}$, eli muut tekijät eivät oikeasti vaikuta tarjontaan tai ne ovat väkiluvun (logaritmin) kanssa korreloimattomia.³ Regression (3) mukaisen mallin ei siis käytännössä voi odottaa tuottavan $\tilde{\beta}$:n harhattomia tai edes tarkentuvia estimaatteja. Mutta kun pyrkimyksenä on tarkastella tarjonnan vaihtelua väkiluvun mukaan, varsinaisen kiinnostuksen kohteena onkin β eikä $\tilde{\beta}$. Silloin tosin ei enää voi puhua väkiluvun vaikutuksesta $\tilde{\beta}$, jolle ei tässä tilanteessa kuitenkaan olisi mitään havainnollista *ceteris paribus*-tulkintaa. Itse asiassa $\tilde{\beta}$:n tarkentuva estimointi saattaa olla ylipäätään mahdotonta, sillä se oikeastaan edellyttäisi, että x :n kaikista komponenteista olisi aineistot käytettävissä.

Aineiston kannalta on huomattavaa, että työntekijöiden määrä on kokonaisluokarvoinen muuttuja, joka voi myös saada arvon nolla. Logaritmoinnin kannalta nollahavainnot ovat tietenkin ongelmallisia, sillä nollan logaritmi ei ole äärellinen. Jotta malli (3) voitaisiin estimoida pienimmän neliösumman menetelmällä (PNS), joudutaan nollahavainnot siis poistamaan, mikä ei *a priori* ole harmitonta. Lisäksi, jos joillakin toimialoilla työntekijämäärät ovat yleisestikin pieniä, tavallisen regressiomallin oletamus vastemuuttujan (työntekijämäärien) jatkuvuudesta on virheellinen. Ongelmaan voidaan hakea ratkaisua sensuroimalla vastemuuttujaa osittain ja soveltamalla siihen *tobit*-estimointia. Liitteessä on tarkemmin perusteltu, miksi näin voidaan tehdä, ja tutkittu *Monte Carlo*-simulaatiokokeen avulla, miten kyseinen estimointimenetelmä käyttäytyy ja verrattu sitä PNS-estimointiin. Simulaation tuloksista voi lyhyesti todeta, että mikäli α ja β ovat pieniä, β :n PNS-estimaattori on voimakkaasti alaspäin harhainen, kun *tobit*-estimaattori on tällöin puolestaan selvästi vähemmän harhainen ja myös keskineliövirheen mielessä optimaalisempi. Jos regression parametrit ovat puolestaan suuria, mikä vastaa tilannetta, jossa havainnot ovat pääasiassa suuria, ja nollahavainnoja on vähän, estimaattorit käyttäytyvät suunnilleen yhtä hyvin.

Palvelutoimialan työntekijöiden määrällä kunnassa on luonnollisesti vaikutus kunnan väkilukuun. Erityisesti, jos toimiala on hyvin merkittävä työllistäjä kunnassa, voi tämä vaikutus olla hyvin suuri, jolloin väkilukua voi pitää *ennaltamääräytyneenä* muuttujana. Tällöin käytetyt estimaattorit ovat harhaisia ja tarkentumattomia. Selvitetessä tämän ongelman vakavuutta malli (3) estimointiin

³On syytä huomioida, että mallin (3) mielekkyyden kannalta on välttämätöntä olettaa, että θ , ρ ja Γ vaihtelevat ainoastaan toimialoittain. Kuntatason vaihtelu aiheuttaisi sen, että mallin (3) regressiokertoimien pitäisi vaihdella myös kunnittain, mikä ei tietenkään ole toivottava ominaisuus jos siksi, että silloin mallissa olisi parametrejä enemmän kuin käytettävissä olevia havaintoja.

käyttämällä väkiluvulle instrumenttina historiallista, vuoden 1975 väkilukua, joka on hyvin voimakkaasti korreloitunut nykyisen väkiluvun kanssa. Instrumenttimuuttujaestimaatit poikkesivat hieman yllättäen PNS-estimaateista niin vähän, että lähes kaikissa tapauksissa PNS-estimaatti mahtui instrumenttimuuttujaestimaatin keskivirheen väliin. Käytetyn instrumentin validius on kuitenkin hieman epäselvä. On nimittäin mahdollista, että esimerkiksi suuresta muuttotappiosta viimeisen parin kymmenen vuoden aikana kärsineessä kunnassa on vielä historiallisena jäänteinä suhteellisen korkea palvelutarjonta. Tällöin historiallinen väkiluku saattaa selittää osaa virhetermin vaihtelusta, jolloin tärkeä virhetermin ja instrumentin ortogonaalisuutta koskeva oletamus rikkoutuu. Ongelmaa voisi mahdollisesti korjata käyttämällä vieläkin vanhempaa väkilukua instrumenttina — valitettavasti vuosi 1975 oli kuitenkin vanhin ajankohta, jolta väkilukutiedot sopivalla kuntajaolla olivat kohtuullisella vaivalla saatavissa.

Mallin estimointitulokset PNS- ja tobit-menetelmillä on esitetty taulukossa 3. Tobit-estimointi on suoritettu mallille (2), joten raportoitu estimaatti on muunneltu, siten että $\hat{\beta}_{\text{tobit}} = \hat{\delta}_{\text{tobit}} - 1$. Hajontakuvioiden perusteella tehtyjen arvioiden perusteella regressioissa ei vaikuttaisi esiytyvän merkittävää virhetermien heteroskedastisuutta. Taulukossa raportoidut keskivirheet on kuitenkin johdettu kovarianssimatriisin HC3-estimaateista (MacKinnon–White, 1985), jotka ovat tarkentuvia myös heteroskedastisten virheiden tapauksessa.⁴ Itse asiassa HC3-estimaatit ovat estimoiduissa regressioissa hyvin lähellä tavallisia PNS-estimaatteja, mikä tukee päätelmiä virhetermien homoskedastiduideesta. Tobit-estimointi perustuu *suurimman uskottavuuden menetelmään* edellyttäen oletusta virhetermien identtisestä ja riippumattomasta normaalisuudesta. Tieto virhetermien homoskedastisuudesta on siis varsin tarpeellinen menetelmän toimivuuden kannalta. Virhetermien normaalisuutta tarkasteltiin tobit-regressioiden residuaaleja tutkimalla, ja hieman harmillisesti normaalisuusolettamus ei varauksetta ole voimassa kaikissa regressioissa. Tobit-estimaattori voi olla asymptoottisestikin harhainen, eli siten tarkentumaton, jos virhetermit eivät olekaan normaalijakautuneita, etenkin kun havainnoista suurin osa on sensuroituja (Arabmazar–Schmidt, 1982).

Puolella toimialoista PNS- ja tobit-estimaatit ovat hyvin lähellä toisiaan. Joustot ovat hieman ykköstä suurempia kaikilla kaupan aloilla, majoitus- ja ravitsemistoiminnassa sekä liike-elämää palvelevassa toiminassa ja muissa palveluissa. Tulkinallisesti tämä siis tarkoittaa, että väkiluvultaan suurissa kunnissa on näiden alojen tarjontaa väkilukuun suhteutettunakin enemmän kuin pienissä kunnissa. Näillä toimialoilla sensuroituiksi päätyneitä havaintoja on muutenkin vähän, minkä perusteella myös PNS-estimaatteja voi pitää luotettavina. Yksikkö-

⁴HC3-estimaattori ei välttämättä ole tarkentuva, jos virheet ovat keskenään korreloituineita. Tässä tapauksessa relevanttia olisi lähinnä virhetermin spatiaalinen autokorreloituminen. Tutkimuksen alkuvaiheissa kokeiltiin mallin (3) muotoa, jossa virheet ovat realisaatioita spatiaalisesta satunnaiskentästä. Korrelaatorakenne tuli kuitenkin hyläytyksi uskottavuusosamäärättestä tukku-kauppaa sekä majoitus- ja ravitsemistoimintaa lukuunottamatta kaikilla toimialoilla, mikä viittaisi siihen että autokorreloitumista ei ilmeisesti esiinny virhetermissä. Minkäänlaisena tyhjentyvänä testinä tätä ei kuitenkaan voi pitää.

Taulukko 3: Työntekijämäärien mallin estimointitulokset.

Toimiala	Tobit ($\lambda = 10$) ^{a,b}			PNS ^a		
	α	β	Sens. ^b	α	β	<i>N</i>
Autokauppa	-7,031*** (0,223)	0,207*** (0,025)	27	-7,309*** (0,216)	0,236*** (0,024)	422
Tukkukauppa	-9,888*** (0,419)	0,490*** (0,046)	67	-9,315*** (0,394)	0,427*** (0,044)	405
Vähittäiskauppa	-5,951*** (0,157)	0,220*** (0,018)	1	-5,702*** (0,181)	0,193*** (0,020)	432
Majoitus- ja ravitsemistoiminta	-6,708*** (0,278)	0,207*** (0,032)	17	-6,596*** (0,300)	0,192*** (0,033)	429
Maaliikenne	-4,480*** (0,184)	0,014 (0,021)	4	-4,669*** (0,177)	0,035 (0,020)	432
Vesiliikenne	-19,504*** (4,067)	0,843* (0,365)	372	-3,116*** (0,967)	-0,512*** (0,113)	140
Ilmaliikenne	-22,101*** (5,939)	1,039* (0,509)	412	-5,690* (2,767)	-0,251 (0,282)	32
Liikennettä palveleva toiminta	-12,123*** (0,885)	0,586*** (0,093)	205	-9,611*** (0,649)	0,333*** (0,071)	326
Posti- ja teleliikenne	-9,392*** (0,331)	0,407*** (0,036)	41	-7,610*** (0,319)	0,218*** (0,036)	413
Rahoituksen välitystoiminta	-6,154*** (0,248)	0,045 (0,028)	23	-5,751*** (0,248)	0,002 (0,029)	428
Vakuutustoiminta	-13,535*** (1,056)	0,628*** (0,104)	270	-8,825*** (0,668)	0,164* (0,073)	226
Rahoitusta palveleva toiminta	-14,899*** (1,448)	0,641*** (0,133)	341	-7,305*** (0,841)	-0,098 (0,092)	201
Kiinteistöalan palvelut	-7,651*** (0,268)	0,242*** (0,030)	45	-7,306*** (0,244)	0,202*** (0,027)	415
Vuokrauspalvelut	-11,512*** (1,002)	0,406*** (0,097)	315	-9,228*** (0,715)	0,169* (0,074)	202
Tietojenkäsittelypalvelu	-17,418*** (1,073)	1,073*** (0,107)	271	-12,212*** (0,699)	0,548*** (0,076)	265
Tutkimus ja kehittäminen	-13,435*** (1,807)	0,506** (0,178)	321	-4,507*** (1,291)	-0,290* (0,143)	161
Muu liike-elämää palv. toim.	-7,569*** (0,175)	0,363*** (0,020)	7	-7,314*** (0,197)	0,335*** (0,022)	429
Ympäristönhuolto	-10,242*** (0,746)	0,337*** (0,076)	208	-8,227*** (0,375)	0,129*** (0,043)	333
Vapaa-ajan palvelut	-9,476*** (0,344)	0,412*** (0,037)	50	-8,440*** (0,276)	0,302*** (0,031)	427
Muut palvelut	-6,787*** (0,191)	0,120*** (0,021)	35	-7,187*** (0,168)	0,161*** (0,019)	427

^aKeskivirheiden estimaatit suluissa, PNS-regressioiden kovarianssimatriisit on estimoitu käyttäen HC3-estimaattoria (MacKinnon–White, 1985). Tähdillä merkitty t-testin merkitsevyydet, riskitasot 5 % (*), 1 % (**) ja 0,5 % (***).

^bTobit-regressiota varten sensuroitujen havaintojen lukumäärä käytetyllä kynnyksparametrilla λ (kts. liite).

joustavuus näyttäisi puolestaan jäävän voimaan maaliikenteessä ja rahoituksen välitystoiminnassa molempien estimointimenetelmien perusteella. Tulos ei välttämättä ole kovin yllättävä, sillä Gini-kertoimien perusteella näiden alojen tarjonta näyttäisi joka tapauksessa olevan alueellisesti varsin tasaisesti jakautunutta.

Toimialat, joilla jouston PNS- ja tobit-estimaatit poikkeavat merkittävästi toisistaan voi jakaa kahteen ryhmään: aloihin, joilla estimaatit ovat saman suuntaisia, mutta hieman eri suuruusluokkaa ja aloihin, joilla estimaatit poikkeavat jopa etumerkin osalta. Aiemmin mainittuun ryhmään kuuluvat mm. liikennettä palveleva toiminta, posti- ja teleliikenne sekä vakuutustoiminta, jonka osalta PNS-estimaatti vaikuttaa hieman alhaiselta verrattuna esimerkiksi tasaisemmin jakautuneisiin kaupan aloihin. Jälkimmäisessä neljä toimialaa käsittävässä ryhmässä PNS-estimaatit ovat negatiivisia ja tobit-estimaatit selvästi positiivisia. Näiden toimialojen PNS-estimaatteja voi joka tapauksessa pitää epäluotettavina estimaattorin harhaisuuden vuoksi ja myös siksi, että intuitiivisesti on hyvin vaikea uskoa, että työntekijöiden määrä suhteessa väkilukuun olisi systemaattisesti suurempi pienissä kunnissa. Tobit-estimaatitkaan eivät välttämättä ole kovin luotettavia, sillä havainnoista on kaikissa neljässä tapauksessa sensuroitu suurin osa, jolloin tobit-estimaattorin varianssi on melko suuri ja oletamus virhetermin normaalisuudesta kriittinen. Varovaisuussyistä näiden toimialojen osalta ainoa mahdollinen johtopäätös onkin, että yksiselitteistä näyttöä nollahypoteesin hylkäämiselle ei ole.

Käytetyt toimialaluokat ovat varsin laajoja, joten poikkeamia yksikköjoustavuudesta voi osittain selittää se, että suuremmissa kunnissa tarjonta on monipuolisempaa toimialaluokan sisällä. Erityisesti luokka muu liike-elämää palveleva toiminta sisältää hyvin erilaisia palveluja ulottuen lakiasiantoimistoista siivousalan yrityksiin. Toisaalta taas rahoituksen välitystoimintaa voi pitää verrattain rajattuna luokkana, sillä se sisältää lähinnä pankkitoimintaa, mikä voi osittain vaikuttaa siihen, että sen jousto on melko alhainen verrattuna muihin laajalle levinneisiin toimialoihin.

2.1.3 Malli toimipaikkojen määrille

Tarjontaa voidaan tarkastella myös toimipaikkojen määrien suhteen samalla tavalla kuin työntekijämäärien tapauksessa. Voidaan tutkia toimipaikkojen määrän joustoja väkiluvun suhteen, ja tulosten tulkinta on samanlainen: lähellä yhtä oleva jousto viittaa siihen, että toimipaikkojen määrä asukasta kohden on keskimäärin samaa luokkaa sekä suurissa että pienissä kunnissa, ja yhtä suurempi puolestaan viittaa siihen että suurissa kunnissa toimipaikkoja on suhteessakin enemmän. Ainoa huomattava ero työntekijöitä koskevaan tarkasteluun nähden on se, useilla toimialoilla kunnittaiset toimipaikkojen määrät ovat maan laajuisestikin hyvin pieniä, minkä vuoksi PNS- ja tobit-estimoinnit ovat mielekkäitä vain osalle toimialoissa. Toimialoilla, joilla toimipaikkoja on vähän, on toisaalta mahdollista juuri pienten lukumäärien vuoksi mallintaa toimipaikkojen *esiintymistodennäköisyyksien* riippuvuutta väkiluvusta, mitä käsitellään seuraavassa aliluvussa.

Taulukko 4: Toimipaikkamäärien mallin estimointitulokset.

Toimiala	Tobit ($\lambda = 10$) ^{a,b}			PNS ^a		
	α	β	Sens. ^b	α	β	N
Autokauppa	-6,105*** (0,243)	-0,006 (0,026)	223	-6,340*** (0,177)	0,017 (0,020)	419
Tukkukauppa	-8,862*** (0,328)	0,283*** (0,035)	258	-7,732*** (0,267)	0,162*** (0,030)	410
Vähittäiskauppa	-5,712*** (0,151)	0,047** (0,017)	98	-5,690*** (0,172)	0,044 (0,019)	432
Majoitus- ja ravitsemistoiminta	-5,713*** (0,275)	-0,043 (0,030)	204	-5,105*** (0,273)	-0,110*** (0,030)	429
Maaliikenne	-3,473*** (0,121)	-0,208*** (0,014)	65	-3,726*** (0,123)	-0,181*** (0,013)	432
Kiinteistöalan palvelut	-7,806*** (0,316)	0,149*** (0,033)	271	-7,387*** (0,209)	0,101*** (0,023)	396
Muu liike-elämää palv. toim.	-7,833*** (0,201)	0,239*** (0,022)	171	-7,103*** (0,200)	0,160*** (0,022)	428
Muut palvelut	-6,687*** (0,191)	0,059*** (0,020)	218	-6,921*** (0,169)	0,082*** (0,019)	418

^aKeskivirheiden estimaatit suluisia, PNS-regressioiden kovarianssimatriisit on estimoitu käyttäen HC3-estimaattoria (MacKinnon–White, 1985). Tähdillä merkitty t-testin merkitsevyydet, riskitasot 5 % (*), 1 % (**) ja 0,5 % (***).

^bTobit-regressiota varten sensuroitujen havaintojen lukumäärä käytetyllä kynnyksparametrilla λ (kts. liite).

Estimoidut mallit olivat oleellisesti muotoa (3), sillä erotuksella, että työntekijämäärien logaritmin sijaan regressoitiiin siis logaritmoituja toimipaikkojen määriä. Tulokset on esitetty taulukossa 4. PNS- ja tobit-estimaatien välillä ei esiinny yhtä suuria poikkeamia kuin työntekijämäärien tapauksessa, mikä johtuu varsinaisesti siitä, että ongelmalliset toimialat on tässä kokonaan sivuutettu. Verrattaessa estimaatteja vastaaviin työntekijöiden määrille saatuihin tuloksiin, voidaan huomata, että ensinnäkin toimialojen keskinäiset erot ovat samansuuntaisia molemmissa tapauksissa, mutta toisaalta toimipaikkojen joustot ovat järjestelmällisesti pienempiä, mikä selittyy sillä, että toimipaikkojen koot (työntekijämäärien mielessä) ovat suurempia väkiluvultaan suuremmissa kunnissa. Autokaupassa sekä ravitsemus- ja majoitustoiminnassa tämä poikkeama konkretisoituu jo siinä, että toimipaikkojen määrissä yksikköjoustavuutta ei voida hylätä, mutta työntekijöiden määrissä vastaavat joustot ovat selvästi yli yhden. Erityisesti näiden toimialojen osalta esitetty selitys toimipaikkojen koon vaihtelusta väkiluvun mukaan on varsin uskottava.

Hieman ehkä yllättävänä voi pitää maaliikenteen negatiivista poikkeamaa yksikköjoustavuudesta. Aiemmin työntekijöiden määrien mallin kohdalla joillakin toimialoilla saatiin kyllä negatiivisia jouston estimaatteja, mutta ainoastaan PNS-menetelmällä, ja näissäkin tapauksissa tobit-estimointi antoi niiden kanssa ristirii-

dassa olevia tuloksia. Nyt kuitenkin maaliikenteessä sekä PNS- että tobit-estimaatit ovat merkitsevästi negatiivisia ja varsin lähellä toisiaan. Negatiivinen poikkeaman tarkoittaisi sitä, että pienissä kunnissa on väkilukuun suhteutettuna enemmän tarjontaa kuin suurissa. Tässä tilanteessa arvoja voi kuitenkin pitää mielekkäinä, sillä maaliikenteen tarve voi hyvinkin olla suhteessa suurempaa väkiluvultaan pienissä, mahdollisesti harvaan asutetuissa kunnissa kuin usein tiheämmin asutetuissa väkiluvultaan suuremmissa kunnissa. Maaliikenne nimittäin käsittää useita pienille paikkakunnille hyvin ominaisia liikennöintimuotoja kuten taksi- ja linja-autoliikenteen, ja on alueiden välistä, eikä sisäistä, joten toimipaikan sijainnilla on vähemmän merkitystä.

2.1.4 Malli ja väkilukukynnykset toimipaikkojen esiintymiselle

Edellä esitettyjen regressiomallien avulla ei oikeastaan voi tehdä päätelmiä hyvin pienistä tarjonnan määristä. Esimerkiksi kysymykseen, paljonko kunnassa ”pitäisi olla” asukkaita, jotta siellä olisi vähintään yksi toimialan i toimipaikka, ei voi saada luotettavaa vastausta jo pelkästään mallien estimointia varten suoritettujen pienten havaintojen sensuroinnin tai nollahavaintojen poistamisen vuoksi. Sen sijaan käyttämällä diskreetin vastemuuttujan regressiota voidaan ottaa täysin formaalilla tavalla huomioon se, että tutkittava muuttuja on kokonaislukuarvoinen. Tutkitaan todennäköisyyttä, että kunnassa esiintyy jonkin tietyn toimialan toimipaikkoja vähintään $k = 1, \dots, K$ kappaletta. Intuitiivisesti on melko selvää, että kyseinen todennäköisyys saattaisi hyvinkin vaihdella kunnan väkiluvun mukaan. Tällöin olisi mahdollista selvittää, kuinka suurissa kunnissa esiintyy ylipäättään toimialan tarjontaa ($k = 1$) tai kilpailua ($k = 2, 3, \dots$).⁵ Merkitään $B_{i,j}$:llä toimialan i toimipaikkojen lukumäärää kunnassa j ja, kuten aikaisemminkin, P_j :llä kyseisen kunnan väkilukua. Toimipaikkojen määrää mallinnetaan *ordered logit*-regressiolla

$$\begin{aligned} \mathbf{P}(B_{i,j} \geq K) &= F(\alpha_{K,i} + \beta_i \log P_j) \\ \mathbf{P}(B_{i,j} \geq K - 1) &= F(\alpha_{K-1,i} + \beta_i \log P_j) \\ &\vdots \\ \mathbf{P}(B_{i,j} \geq 1) &= F(\alpha_{1,i} + \beta_i \log P_j) \\ \mathbf{P}(B_{i,j} = 0) &= 1 - F(\alpha_{1,i} + \beta_i \log P_j), \end{aligned} \tag{6}$$

missä $\mathbf{P}(\cdot)$ on todennäköisyyssmitta ja $F(\cdot)$ *logistisen jakauman* kertymäfunktio

$$F(x) = \frac{1}{1 + \exp(-x)}.$$

Ordered logit-mallin sijaan olisi mahdollista käyttää myös *ordered probit*-mallia, jossa logistisen jakauman sijaan käytetään normaalijakaumaa. Probit-muotoisen

⁵Tietysti, se että kunnassa on kaksi toimialan toimipaikkaa, ei välttämättä takaa, että ne kilpailisivat keskenään. Muutamat käsitellyistä toimialoista ovat määritelmällisestikin niin laajoja ja varsin erilaista toimintaa sisältäviä, ettei kaikkien niiden yritysten tarjoamia palveluita voi pitää keskenään kilpailevina.

Taulukko 5: Toimipaikkamäärien ordered logit-mallin estimointitulokset.

Toimiala	SU-estimaatit ^a						C ^b
	β	α_1	α_2	α_3	α_4	α_5	
Tukkukauppa	2,68 (0,22)	-17,43 (1,61)	-18,84 (1,66)	-19,91 (1,70)	-20,76 (1,73)	-21,25 (1,75)	0,72
Liikennettä palveleva toim.	1,97 (0,14)	-15,20 (1,15)	-16,65 (1,19)	-17,64 (1,23)	-18,21 (1,25)	-18,61 (1,26)	0,51
Posti- ja teleliikenne	2,44 (0,16)	-17,51 (1,23)	-21,29 (1,37)	-22,50 (1,42)	-23,38 (1,46)	-24,25 (1,49)	0,55
Rahoituksen välitystoiminta	2,36 (0,15)	-12,74 (1,18)	-17,48 (1,23)	-19,47 (1,29)	-20,85 (1,34)	-21,71 (1,37)	0,52
Vakuutustoiminta	2,08 (0,15)	-17,67 (1,31)	-19,62 (1,39)	-20,37 (1,42)	-20,80 (1,44)	-21,31 (1,47)	0,60
Rahoitusta palveleva toim.	1,98 (0,15)	-16,95 (1,26)	-18,32 (1,30)	-19,32 (1,35)	-20,19 (1,39)	-20,54 (1,41)	0,62
Kiinteistöalan palvelut	2,84 (0,22)	-19,43 (1,64)	-20,66 (1,67)	-21,53 (1,71)	-22,15 (1,73)	-23,04 (1,77)	0,67
Tietojenkäsittelypalvelu	2,35 (0,16)	-18,84 (1,32)	-20,48 (1,37)	-21,41 (1,41)	-22,08 (1,44)	-22,50 (1,46)	0,56
Ympäristönhuolto	1,79 (0,12)	-13,56 (1,00)	-15,30 (1,05)	-16,66 (1,10)	-17,36 (1,12)	-18,16 (1,15)	0,44
Vapaa-ajan palvelut	2,37 (0,16)	-17,02 (1,29)	-18,48 (1,33)	-19,62 (1,37)	-20,38 (1,40)	-20,90 (1,42)	0,51

^aKeskivirheiden estimaatit suluissa, kaikki kerroinestimaatit ovat tilastollisesti merkitseviä 0,5 %-n riskitasolla.

^bOsuus vastemuuttujan arvoista, jotka malli ennustaa oikein.

mallin estimoinnissa ilmenneiden numeeristen ongelmien vuoksi tässä päädyttiin kuitenkin logit-malliin. Itse asiassa sillä, kumpaa mallia käyttää, ei ole suurtaakaan merkitystä; jakaumat ovat hyvin samankaltaiset, ja valintaa logit- ja probitmuotojen välillä pidetään yleisestikin vaikeana, ellei käytössä ole hyvin suurta määrää havaintoja, jolloin jakaumat erottavia häntätodennäköisyyksiä voidaan luotettavammin arvioida (Amemiya, 1981).

Malli (6) estimoidaan toimialoittain suurimman uskottavuuden menetelmällä. Parametrien estimaatit on esitetty taulukossa 5. Malliin on valittu pääasiassa sellaisia toimialoja, joilla huomattavassa määrässä kuntia toimipaikkojen määrä on hyvin pieni tai nolla. Toimialat, joilla väkiluvun yhteys tarjontaan vaikuttaa ylipäätään epäselvältä, kuten tutkimustoiminta sekä vesi- ja ilmaliikenne, sivuutettiin. Mallia estimoidaessa vähimmäismäärien ylärajaksi valittiin $K = 5$. Koska logit-mallien parametriestimaatit eivät sinänsä ole tulkinnallisesti kovin havainnollisia, tarkastellaan väkiluvun kynnsarvoja toimipaikkojen esiintymiselle. Olkoon $T_{k,p,i}$ väkiluku, jolla kunnassa esiintyy vähintään k toimialan i toimipaikkaa

todennäköisyydellä p . Se voidaan ratkaista kertymäfunktioista helposti:

$$T_{k,p,i} = \left(\frac{1-p}{p} \exp(\alpha_{k,i}) \right)^{-\frac{1}{\beta_i}} \quad (7)$$

Kynnysarvojen ohella on havainnollista lisäksi tarkastella niiden keskinäisiä suhteita (vrt. Bresnahan–Reiss, 1991). Määritellään *normalisoidut suhdeluvut* $R_{k,i}$, $k = 1, \dots, K - 1$ kaavalla

$$R_{k,i} = \left(\frac{k}{k+1} \right) \frac{T_{k+1,i}}{T_{k,i}}.$$

Suhdelukujen tulkinnan kannalta käyttökelpoinen tieto on se, että kynnysarvot kasvavat lineaarisesti, jos ja vain jos suhdeluvut ovat ykkösiä.⁶ Ykkösestä poikkeavat suhdeluvut kertovat siis kasvun epälineaarisuuksista. Kaavasta (7) on hyödyllistä huomata, että käytetystä jakaumasta johtuen kyseiset suhdeluvut ovat riippumattomia valitusta varmuustasosta p .

Kynnysarvot kolmella varmuustasolla sekä niihin liittyvät suhdeluvut on esitetty taulukossa 6. Tarkastellut toimialat poikkeavat voimakkaasti jo ensimmäisen toimipaikat kynnysarvon kannalta. Rahoituksen välitystoiminnassa kynnys on erittäin alhaalla, toimipaikkoja esiintyy 80 prosentin varmuudella jo 400 asukkaan kunnissa, mikä ei ole tietenkään ole yllättävää, sillä alan toimipaikkoja on neljää Ahvenanmaan saaristokuntaa lukuun ottamatta kaikissa Suomen kunnissa. Sen sijaan yllättävämpää on se, miten jyrkästi rahoituksen välitystoiminnan kynnykset nousevat, kun toimipaikkoja on enemmän; jo se, että kunnassa olisi vähintään kaksi toimipaikkaa, edellyttää alempaan kynnykseen verrattuna yli seitsenkertaista väkilukua. Tuloksen voisi katsoa heijastelevan sitä, miten kattava paikallispankkien toiminta edelleenkin on maaseudulla. Suurempien kuntien asukkaat joutuvat tyytymään suhteessa selvästi niukempaan konttorien määrään. Toisaalta taas kaupungeissa sähköisten pankkipalvelujen käyttö on mahdollisesti paljon yleisempää, joten varsinaisia konttoripalveluja käyttäviin suhteutettuna erot eivät välttämättä olisi aivan yhtä suuria. Myös posti- ja teleliikenteessä kynnykset kasvavat varsin jyrkästi, mutta erona rahoituksen välitystoimintaan kannatta huomata, miten ensimmäinen kynnysarvo on huomattavasti suurempi. Postitoimipaikkoja on myös siirtynyt asiamiesposteiksi muun toiminnan, esimerkiksi vähittäiskaupan yhteyteen. Tällöin ne eivät määrity itsenäisiksi toimipaikoiksi.

Muista toimialoista kannattaa huomioida vakuutustoiminta ja rahoitusta palveleva toiminta, joiden toimipaikkoja esiintyy lähinnä suuremmissa, yli 10000 asukkaan kunnissa 80 prosentin varmuudella. Liikennettä palvelevan toiminnan, tietojenkäsittelypalvelujen ja ympäristönhuollon toimipaikkoja puolestaan esiintyy 80 prosentin varmuudella yli 4000–5000 asukkaan kunnissa. Suhdeluvut ovat ympäristönhuoltoa ja aiemmin mainittuja rahoituksen välitystä sekä posti- ja teleliikennettä lukuun ottamatta yllättävän lähellä yhtä. Tästä voisi päätellä, että toimipaikkojen esiintymiselle ”vaadittu” väkiluku kasvaa todellakin lineaarisesti toimipaikkojen määrän suhteen näillä toimialoilla.

⁶Helppo päätellä esimerkiksi käyttämällä induktiota.

Taulukko 6: Toimipaikkojen esiintymisen kynnysarvot ja niiden suhdeluvut.

Toimiala	p	Kynnysarvot (1000 asukasta)					Suhdeluvut			
		T_1	T_2	T_3	T_4	T_5	R_1	R_2	R_3	R_4
Tukkukauppa	0,5	0,7	1,1	1,7	2,3	2,8	0,8	1,0	1,0	1,0
	0,8	1,1	1,9	2,8	3,9	4,7				
	0,95	2,0	3,4	5,1	6,9	8,3				
Liikennettä palveleva toiminta	0,5	2,2	4,7	7,7	10,3	12,6	1,0	1,1	1,0	1,0
	0,8	4,5	9,4	15,6	20,7	25,5				
	0,95	9,9	20,8	34,3	45,7	56,2				
Posti- ja teleliikenne	0,5	1,3	6,2	10,3	14,7	21,0	2,4	1,1	1,1	1,1
	0,8	2,3	11,0	18,1	25,9	37,1				
	0,95	4,4	20,9	34,3	49,2	70,3				
Rahoituksen välitystoiminta	0,5	0,2	1,7	3,9	7,0	10,0	3,7	1,6	1,3	1,2
	0,8	0,4	3,0	7,0	12,5	18,1				
	0,95	0,8	5,8	13,5	24,3	35,0				
Vakuutustoiminta	0,5	4,8	12,3	17,6	21,6	27,6	1,3	1,0	0,9	1,0
	0,8	9,4	23,9	34,3	42,1	53,6				
	0,95	19,8	50,4	72,4	88,8	113,2				
Rahoitusta palveleva toiminta	0,5	5,2	10,4	17,1	26,6	31,8	1,0	1,1	1,2	1,0
	0,8	10,4	20,9	34,5	53,5	63,9				
	0,95	22,9	45,8	75,8	117,3	140,3				
Kiinteistöalan palvelut	0,5	0,9	1,4	1,9	2,4	3,3	0,8	0,9	0,9	1,1
	0,8	1,5	2,3	3,2	3,9	5,4				
	0,95	2,6	4,0	5,5	6,8	9,3				
Tietojenkäsittelypalvelu	0,5	3,0	6,0	8,9	11,9	14,3	1,0	1,0	1,0	1,0
	0,8	5,4	10,9	16,1	21,5	25,7				
	0,95	10,5	21,1	31,2	41,6	49,8				
Ympäristöhuolto	0,5	1,9	5,0	10,8	15,9	24,8	1,3	1,4	1,1	1,2
	0,8	4,1	10,9	23,4	34,4	53,7				
	0,95	9,9	26,0	55,7	82,0	128,0				
Vapaa-ajan palvelut	0,5	1,3	2,4	3,9	5,4	6,8	0,9	1,1	1,0	1,0
	0,8	2,4	4,4	7,1	9,7	12,2				
	0,95	4,6	8,4	13,6	18,8	23,5				

2.2 Etäisyyksien ja tarjonnan yhteys

Edellisessä luvussa tarkasteltiin, miten palvelujen tarjonta riippuu kunnan *omasta väkiluvusta*. Tietenkin on keinotekoisia olettaa, että vain kunnan omalla väkiluvulla olisi merkitystä, sillä eivät asiakasvirrat kunnioita kuntarajoja. Siksi onkin kiinnostavaa selvittää, mikä on tarjonnan riippuvuus kunnan sijainnista suhteessa muihin väestökeskittymiin. Pitkillä etäisyyksillä voi olla sinänsä suurikin merkitys palvelujen tarjonnalle. Pitkien välimatkojen päässä oleville alueille kuljetuskustannukset voivat muodostua merkittäväksi rasitteeksi, ja yhteydenpito voi muutenkin olla keskimääräistä vaikeampaa.

Mallia varten kuntien sijainnit karakterisoidtiin tasoon projisoidusta kuntarajat sisältävästä kartasta laskettujen kuntien *geometrinen massakeskipisteiden* (centroid) avulla. Voisi otaksua, että tällä tavoin laskettu sijainnin tunnusluku approksimoi varsin harhattomasti taloudellisen toiminnan keskimääräistä alueellista sijoittumista ainakin Etelä- ja Keski-Suomen kunnissa. Laajoja, harvaan asuttuja alueita käsittävissä kunnissa, esimerkiksi Lapissa, menetelmä saattaa toimia heikommin. Kuntien i ja j väliseksi etäisyydeksi $d_{i,j}$ asetettiin massakeskipisteiden tavallinen euklidinen etäisyys.

Etäisyyksien ja sijainnin merkityksen arviointia varten kunnalle j on muodostettu tunnusluku

$$L_j = \sum_{\substack{i=1 \\ i \neq j}}^{432} \frac{P_i}{d_{i,j}},$$

jota voidaan kutsua *läheisyysindeksiksi*. Selvästi L on pieni kunnissa, jotka sijaitsevat kaukana suurista väestökeskittymistä ja suuri kunnissa, joiden lähellä on väkirikkaita kuntia. Läheisyysindeksin vaikutusta tarjontaan on tutkittu lisäämällä⁷ sen logaritmi uudeksi regressoriksi malliin (3):

$$\log \frac{E_{i,j}}{P_j} = \alpha_i + \beta_i \log P_j + \gamma_i \log L_j + \varepsilon'_{i,j} \quad (8)$$

Mallin estimointitulokset on esitetty taulukossa 7. Parametrien estimointi suoritettiin sekä PNS- että tobit-menetelmillä, mutta taulukossa raportoidaan ainoastaan PNS-estimaatit. Nimittäin toimialoilla, joilla saatiin selvää näyttöä nollassa hypoteesin $H_0 : \gamma_i = 0$ hylkäämiselle, PNS- ja tobit-estimointien tulokset olivat hyvin lähellä toisiaan.

Tarjonta näyttäisi riippuvan läheisyydestä kahdella eri tavalla: liikennettä palvelevassa toiminnassa, vähittäiskaupassa, vapaa-ajan palveluissa sekä majoitus- ja ravitsemistoiminnassa kerroin on negatiivinen, minkä voi tulkita tarkoittavan sitä, että suurten väestökeskittymien läheisyys vähentää näiden alojen tarjontaa. Mahdollisesti kunnan asukkaat käyvät ostoksilla ja ravintoloissa mieluummin

⁷Koska P :n ja L :n ovat logaritmit ovat lievästi korreloituneita, ei mallin (8) parametri β ole tarkalleen ottaen sama kuin mallissa (3), mikä on kuitenkin lähinnä tekninen sivuseikka.

suuremmassa keskuksessa kuin kotikunnassa. Vastaavasti, jos kunta sijaitsee kaukana suurista väestökeskittymistä, näiden alojen tarjonta on suhteellisesti suurempaa. Majoitus- ja ravitsemisalalla tämä voi osittain heijastaa myös joidenkin vapaa-ajanviettokeskusten sijaintia harvemmin asutuilla alueilla, kuten Lapissa. Tukkukaupassa tilanne näyttäisi toisaalta olevan päinvastainen, läheisyysindeksin kerroin onkin negatiivinen. Tulkinallisesti voisi päätellä, että suurten väestökeskittymien läheisyys kasvattaa tarjontaa, eli toimiala haluaa mahdollisesti sijoittaa lähelle keskusta, muttei aivan keskuksen sisään. Varsin monella toimialalla kerroin ei kuitenkaan poikkea merkitsevästi nolasta, eli riittävää näyttöä läheisyyden yhteydestä tarjontaan ei ole.

Taulukko 7: Estimointitulokset työntekijämäärien mallille, jossa on mukana läheisyysindeksi.

Toimiala	PNS ^a			N
	α	β	γ	
Autokauppa	-7,303*** (0,274)	0,236*** (0,024)	-0,002 (0,061)	422
Tukkukauppa	-11,595*** (0,488)	0,372*** (0,040)	0,814*** (0,100)	405
Vähittäiskauppa	-5,148*** (0,213)	0,206*** (0,020)	-0,200*** (0,042)	432
Majoitus- ja ravitsemistoiminta	-5,598*** (0,423)	0,217*** (0,033)	-0,359*** (0,103)	429
Maaliikenne	-4,585*** (0,251)	0,037 (0,020)	-0,030 (0,066)	432
Vesiliikenne	-2,254 (1,587)	-0,490*** (0,110)	-0,305 (0,377)	140
Ilmaliikenne	-4,974 (3,376)	-0,143 (0,262)	-0,517 (0,558)	32
Liikennettä palveleva toiminta	-8,277*** (0,754)	0,387*** (0,072)	-0,539*** (0,140)	326
Posti- ja teleliikenne	-7,281*** (0,396)	0,227*** (0,037)	-0,120 (0,091)	413
Rahoituksen välitystoiminta	-5,989*** (0,298)	-0,004 (0,029)	0,086 (0,062)	428
Vakuutustoiminta	-8,687*** (0,832)	0,170 (0,072)	-0,055 (0,179)	226
Rahoitusta palveleva toiminta	-6,902*** (0,886)	-0,079 (0,089)	-0,170 (0,129)	201
Kiinteistöalan palvelut	-6,887*** (0,296)	0,212*** (0,027)	-0,149* (0,067)	415
Vuokrauspalvelut	-9,788*** (0,865)	0,153* (0,071)	0,206 (0,162)	202
Tietojenkäsittelypalvelu	-12,368*** (0,778)	0,543*** (0,077)	0,059 (0,142)	265
Tutkimus ja kehittäminen	-3,499** (1,334)	-0,211 (0,145)	-0,520* (0,245)	161
Muu liike-elämää palveleva toiminta	-7,734*** (0,264)	0,325*** (0,022)	0,150* (0,062)	429
Ympäristöhuolto	-7,353*** (0,532)	0,156*** (0,043)	-0,328** (0,120)	333
Vapaa-ajan palvelut	-7,678*** (0,357)	0,322*** (0,031)	-0,277*** (0,085)	427
Muut palvelut	-6,821*** (0,235)	0,170*** (0,018)	-0,131* (0,053)	427

^aKeskivirheiden estimaatit suluissa, PNS-regressioiden kovarianssimatriisit on estimoitu käyttäen HC3-estimaattoria (MacKinnon–White, 1985). Tähdillä merkitty t-testin merkitsevyydet, riskitasot 5 % (*), 1 % (**) ja 0,5 % (***).

3 Yhteenveto

Tässä tutkimuksessa on tarkasteltu yksityisten palvelujen tarjontaa alueellisesta näkökulmasta. Aluksi on tutkittu tarjonnan keskittymistä toimialoittain käyttäen mittareina Gini-kertoimia ja tarjonnan puuttumista kunnista. Tulokset tulevat hyvin vahvasti ennako-odotuksia siitä, että paikkasidonnaiset alat, kuten vähittäiskauppa, majoitus- ja ravitsemistoiminta, maaliikenne sekä henkilökohdaiset palvelut ovat varsin tasaisesti jakautuneita ja vähemmän paikkaan sidotut alat, kuten tietojenkäsittely sekä tutkimus ja kehittäminen, puolestaan keskittyneitä. Hieman yllättäen pankkitoiminta kuuluu tasaisemmin jakautuneisiin toimialoihin konttorien merkittävästä vähentymisestä huolimatta.

Keskittymisen ohella on tarkasteltu myös tarjonnan riippuvuutta alueen väestöstä estimoimalla tarjonnan joustoja väkiluvun suhteen. Keskeisenä tavoitteena on ollut selvittää, onko tarjontaa väkilukuunkin suhteutettuna enemmän väkiluvultaan suurissa kunnissa, eli onko tarjonnan jousto suurempi kuin yksi. Useimmilla toimialoilla joustot ovat hiukan ykköstä suurempia. Joustoiltaan suurimpia toimialoja ovat tukkukauppa, rahoitusta palveleva toiminta ja tietojenkäsittelypalvelu. Merkittäviä yksikköjoustavia toimialoja ovat sen sijaan maaliikenne ja pankkitoiminta. Itse asiassa maaliikenteessä, tarkasteltuna toimipaikkojen määrän kannalta, tarjonnan jousto jää jopa merkitsevästi alle yhden, mikä viittaisi siihen, että alan suhteellista tarjontaa on väkiluvultaan pienemmissä kunnissa enemmän. Muutamilla toimialoilla on lisäksi tarkasteltu ordered logit-mallin avulla, mitkä ovat kynnysväkiluvut sille, että alan toimipaikkoja esiintyy kunnassa. Useimmilla tarkastelluista toimialoista kynnysväkiluvut kasvavat likimäärin lineaarisesti toimipaikkojen määrän suhteen. Merkittävä poikkeus tästä on pankkitoiminta, jossa yksittäisiä toimipaikkoja esiintyy väkiluvultaan hyvinkin pienissä kunnissa, mutta useammat toimipaikat ”edellyttävät” suhteessa huomattavasti suurempaa asukasmäärää.

Maantieteellisten etäisyyksien merkitystä on selvitetty konstruoimalla kunnille läheisyysindeksi, joka kuvaa kunnan sijoittumista suhteessa suuriin väestökeskittymiin. Läheisyysindeksin avulla on sitten tutkittu, miten palvelujen tarjonta riippuu etäisyyksistä. Useimmilla tarkastelluista toimialoista merkitsevää riippuvuutta ei esiinny. Muutamilla toimialoilla, kuten vähittäiskaupassa sekä majoitus- ja ravitsemistoiminnassa suurten väestökeskittymien läheisyys vähentää tarjontaa. Tukkukaupassa sen sijaan riippuvuus on päinvastainen, minkä voisi tulkita tarkoittavan sitä, että tukkukauppa sijoittuu väkimäärältään suurten kuntien läheisyyteen, muttei kuitenkaan näiden kuntien alueille.

Lähteet

- AMEMIYA, T. (1973): Regression Analysis When the Dependent Variable Is Truncated Normal. *Econometrica* 41, 997–1016.
- AMEMIYA, T. (1981): Qualitative Response Models: A Survey. *Journal of Economic Literature* 19, 1483–1536.
- ARABMAZAR, A.–SCHMIDT, P. (1982): An Investigation of the Robustness of the Tobit Estimator to Non-normality. *Econometrica* 50, 1055–1063.
- BRESNAHAN, T. F.–REISS, P. C. (1991): Entry and Competition in Concentrated Markets. *Journal of Political Economy* 99, 977–1009.
- GILES, D. E. A. (2004): Calculating a Standard Error for the Gini Coefficient: Some Further Results. *Oxford Bulletin of Economics and Statistics* 66, 425–433.
- MACKINNON, J. G.–WHITE, H. (1985): Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics* 29, 305–325.
- NYGÅRD, F.–SANDSTRÖM, A. (1981): *Measuring Income Inequality*. Almqvist & Wiksell, Stockholm.
- OGWANG, T. (2000): A Convenient Method of Computing the Gini Index and its Standard Error. *Oxford Bulletin of Economics and Statistics* 62, 123–129.
- OGWANG, T. (2004): Calculating a Standard Error for the Gini Coefficient: Some Further Results: Reply. *Oxford Bulletin of Economics and Statistics* 66, 435–437.
- R DEVELOPMENT CORE TEAM (2005): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
<http://www.r-project.org/>

Liite

Gini-kertoimien keskivirhe-estimointi

Gini-kertoimien keskivirhe-estimointia on syytä esitellä hieman, sillä kirjallisuudessa Gini-kertoimille on esitetty keskivirhearvioita varsin harvoin, ja tarvittavat menetelmät eivät liene kovinkaan yleisesti tunnettuja. Menettely perustuu siihen, että tutkittavalle aineistolle annetaan *stokastinen tulkinta*, toisin sanoen oletetaan, että se on realisaatio jostakin satunnaisprosessista (Nygård–Sandström, 1981, s. 36). Olkoon y_1, \dots, y_n aineisto suuruusjärjestyksessä pienimmästä suurimpaan. Tällöin otoksen Gini-kertoimen määrittäminen voidaan palauttaa (Giles, 2004) pienimmän neliösumman regressio-ongelmaksi

$$i\sqrt{y_i} = \theta\sqrt{y_i} + u_i, \quad (9)$$

ja Gini-kerroin muodostetaan PNS-estimaatin $\hat{\theta}$ avulla:

$$\hat{G}_1 = \frac{2\hat{\theta}}{n} - 1 - \frac{1}{n}. \quad (10)$$

Nyt Gini-kertoimen keskivirhe saadaan Gilesin mukaan kaavan (10) perusteella PNS-estimaatin keskivirheestä:

$$\text{s.e.}(\hat{G}_1) = \frac{2 \text{s.e.}(\hat{\theta})}{n}.$$

Ogwang (2004) on kritisoinut menetelmää sillä, että ensinnäkin regressiossa (9) virheet ovat useimmissa tapauksissa heteroskedastisia ja toiseksi, estimointiongelma voidaan esittää myös vaihtoehtoisella tavalla:

$$\frac{y_i}{\bar{y}} = \alpha + \beta i + v_i,$$

missä \bar{y} on havaintojen keskiarvo, Gini-kertoimen ja keskivirheen ollessa

$$\hat{G}_2 = \frac{n^2 - 1}{6n} \hat{\beta} \quad \text{ja} \quad \text{s.e.}(\hat{G}_2) = \frac{n^2 - 1}{6n} \text{s.e.}(\hat{\beta}).$$

Ogwang (2000) on osoittanut, että $\hat{G}_1 = \hat{G}_2$, mutta yleisesti $\text{s.e.}(\hat{G}_1) \neq \text{s.e.}(\hat{G}_2)$.

Ensimmäiseen ongelmaan voidaan helposti vastata käyttämällä sopivaa kovarianssimatriisin estimaattoria (MacKinnon–White, 1985), mutta estimaattien erisuuruutta koskevaan ongelmaan ei kirjallisuudessa ilmeisesti ole ainakaan toistaiseksi esitetty ratkaisua. Useimmissa tapauksissa molemmat keskivirheen estimaatit vaikuttaisivat olevan samaa suuruusluokkaa, mikä on tarkkuudeltaan oleellisesti riittävää tämän tutkimuksen tarpeisiin. Varovaisuuden vuoksi tulosten yhteydessä kuitenkin raportoidaan estimaateista suurempi.

Tobit-estimaattorin ominaisuudet

Tässä osiossa esitetään perustelut tutkimuksessa käytetyn *tobit*-tyyppisen estimointimenetelmän käytölle. Aluksi esitetään intuitiivinen perustelu menetelmälle ja lopuksi tarkastellaan menetelmän pienotosominaisuuksia verrattuna PNS-estimointiin Monte Carlo-simulaation avulla.

Tarkastellaan yksinkertaista lineaarista regressiomallia

$$\log y_i = \alpha + \beta x_i + u_i, \quad u_i \sim \text{NID}(0, \sigma^2), \quad (11)$$

missä muuttuja y_i on latentti (ei-havaittava). Sen sijaan havaitaan $\tilde{y}_i = r(y_i)$, missä $r : \mathbb{R} \rightarrow \mathbb{Z}$ on arvot kokonaisluvuiksi pyöristävä funktio. Koska y_i on satunnaismuuttuja, myös \tilde{y}_i on satunnaismuuttuja, sillä r on kasvava funktiona Borel-mitallinen. Koska $|y_i - \tilde{y}_i| \leq \frac{1}{2}$, riittävän suurilla y_i pätee approksimaatio $\log y_i \approx \log \tilde{y}_i$. Ongelmana on tietenkin se, että kaikki y_i :t eivät ole riittävän suuria, ja jos $\tilde{y}_i = 0$, sen logaritmi ei ole edes äärellinen. Ongelmalle voi hakea ratkaisua sensuroimalla ongelmalliset havainnot. Muodostetaan siis uusi muuttuja $y'_i = \max\{\tilde{y}_i, \lambda\}$, jonka logaritmia käytetään vastemuuttujana. Kun sensurointikynnyksen määräävä vakio $\lambda > 0$ on riittävän suuri, sensuroimattomassa osassa havaintoja pätee $\log y'_i \approx \log y_i$ ja muulloin $\log y'_i = \log \lambda$. Tällöin mallin (11) parametrit voidaan estimoida (likimäärin) tarkentuvasti suurimman uskottavuuden menetelmään perustuvalla tobit-estimoinnilla (Amemiya, 1973). Keskeinen kysymys on tietenkin λ :n valinta. Sensurointikynnyksen pitäisi olla riittävän suuri, jotta pyöristysvirheiden merkitys saataisiin riittävän vähäiseksi, mutta toisaalta mitä suurempi λ on, sitä enemmän sensuroinnissa menetetään aineiston informaatiota, mikä puolestaan kasvattaa estimoinnin epävarmuutta.

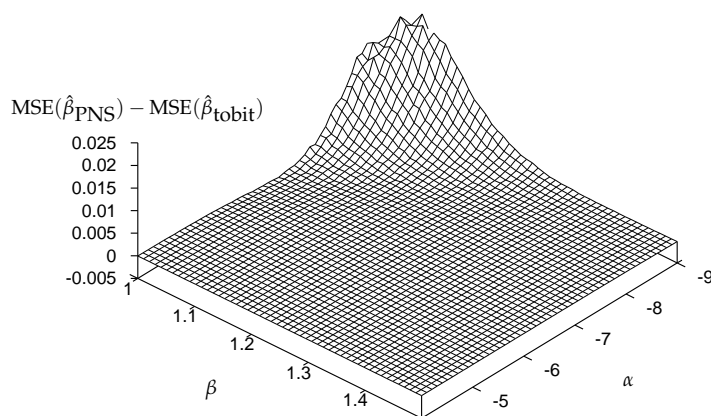
Tobit-estimoinnin käyttöä voidaan näin perustella ainakin heuristisesti, mutta mikään ei vielä oikeastaan takaa menetelmän käyttökelpoisuutta. Jäljelle jääneiden, joskin pienten, pyöristysvirheiden vaikutus on tuntematon, ja estimaattorin tarkentuvuus ei tietenkään kerro mitään käyttäytymisestä pienissä otoksissa. Siksi menetelmän toimivuutta on tarpeellista tutkia simulaatiokokeen avulla, jossa luonnollisena vertailukohtana toimii vaihtoehtoinen PNS-estimaattori. Simulaatiossa on keskeisenä tavoitteena se, että simuloitu aineisto jäljittelisi mahdollisimman hyvin tutkimuksessa käytettyjä oikeita aineistoja. Käytetty *datan generoiva prosessi*, DGP on kaavan (11) mukainen virhetermin hajonnalla $\sigma = \frac{1}{2}$, joka on keskimääräinen arvio estimoiduista regressioista. Regressori on määritelty siten, että sen jakauma on Suomen kuntien logaritmoitujen väkilukujen jakauman kaltainen, käytännössä $x_i = z_i^2$ ja $z_i \sim \text{NID}(\mu_z, \sigma_z^2)$, missä $\mu_z \approx 2,92$ ja $\sigma_z \approx 0,19$ ovat aineistosta estimoidut parametrit. Otokokona on aineiston otoskoko 432, ja toistojen lukumääränä 3000.⁸ Parametreja on muuteltu väleillä $\alpha \in [-9, -4]$ ja $\beta \in [1, 1\frac{1}{2}]$. Koska tutkimuksessa on ollut kiinnostuksen kohteena lähinnä regressorin kerroin, on simulaatiossa tutkittu ainoastaan β :n estimaattoreita. Käytetty

⁸Toistojen määrää ei voi pitää kovin suurena, mikä näkyy kuvassa 1 lievänä karkeutena "vuoren" huipulla. Käytetyn tobit-estimointirutiinin hitauden vuoksi toistojen määrää ei kuitenkaan ollut mielekästä nostaa.

Taulukko 8: Tobit- ja PNS-estimaattorien ominaisuuksia.

α	β	Keskiarha		Hajonta		$\sqrt{\text{Keskineliövirhe}}$	
		PNS	Tobit	PNS	Tobit	PNS	Tobit
-8,5	1,0	-0,22964	0,06217	0,03297	0,18851	0,23199	0,19850
	1,1	-0,11585	0,03327	0,02607	0,08239	0,11875	0,08886
	1,2	-0,05079	0,02025	0,02255	0,04885	0,05557	0,05288
-8,0	1,0	-0,15324	0,04603	0,02632	0,11717	0,15548	0,12589
	1,1	-0,06724	0,02682	0,02359	0,06050	0,07126	0,06618
	1,2	-0,02536	0,01569	0,02209	0,03961	0,03363	0,04261
-7,0	1,0	-0,04637	0,02694	0,02149	0,05365	0,05111	0,06004
	1,1	-0,01405	0,01584	0,02173	0,03652	0,02588	0,03981
	1,2	-0,00354	0,00819	0,02158	0,02865	0,02186	0,02980

Kuva 1: Tobit- ja PNS-estimaattorien keskineliövirheiden vertailu.



kynnysparametrin arvo on tutkimuksessa käytetty, alun perin hajontakuviosta silmämääräisesti arvioitu $\lambda = 10$. PNS-estimointi on suoritettu tavalliseen tapaan poistamalla nollihavainnot.

Simulaatiokokeen tuloksia on raportoitu kahdessa muodossa: kuvassa 1 on esitetty PNS- ja tobit-estimaattorien simuloitujen keskineliövirheiden erotukset tutkitussa parametriavaruuden osassa, ja taulukossa 8 on tarkemmin esitetty estimaattorien keskiarhat ja hajonnat yhdeksässä pisteessä. Kuvassa 1 esiintyväs- tä "vuoresta" voi päätellä, että pienillä parametrien arvoilla tobit-estimaattori on selvästi PNS-estimaattoria optimaalisempi keskineliövirheen perusteella. Taulu- kon 8 mukaan tällöin PNS-estimaattori on hyvin voimakkaasti alaspäin harhai- nen. Tobit-estimaattorin hajonta on kylläkin voimakkaasti PNS-estimaattorin hajon- ta suurempi, mutta keskineliövirheessä merkitykseltään PNS:n harhaa pienem- pi. Itse asiassa tutkimuksen johtopäätösten kannalta pienempi harha saattaisi olla

pientä hajontaa toivottavampi ominaisuus, jolloin keskineliövirheen sijaan pitäisi käyttää jotakin muuta, enemmän harhaa painottavaa tappiofunktiota, jolloin tulokset olisivat luonnollisesti tobit-estimaattorin kannalta vieläkin suotuisampia. Muualla tutkimuksessa alueessa estimaattorin näyttäisivät käyttäytyvän yhtä hyvin, mikä johtuu siitä, että kertoimien ollessa suurempia myös vastemuuttujan arvot ovat suuria, jolloin pyöristysvirheiden vaikutus pienenee. Tämä on itse asiassa PNS-estimoinnin kannalta hyödyllinen tieto, sillä se osoittaa varsin selvästi, että nollahavaintojen poistaminen (joita on silloin varsin vähän) on tällöin harmitonta.

Tehtäessä päätöstä estimaattorin valinnasta tämän simulaatiokokeen perusteella kannattaa huomioida se, että tässä ei ole lainkaan tarkasteltu (hyvinkin todennäköistä) tilannetta, jossa DGP:n virhetermi ei olekaan normaalijakautunut. Tällöin tobit-estimaattori ei ole välttämättä tarkentuva (Arabmazar–Schmidt, 1982), ja pienotosominaisuudet saattavat olla heikompia. Siksi tobit-estimaattoria on tutkimuksessa käytettykin lähinnä PNS-estimaattien robustisuuden tarkistamiseen. Tilanteissa, joissa estimaattorit ovat antaneet hyvin poikkeavia tuloksia, tulkinta onkin lähinnä ollut, että luotettavia johtopäätöksiä ei käytetyn aineiston valossa voi tehdä.



Pellervon taloudellisen tutkimuslaitoksen julkaisuja, publikationer, Publications

18. Vesa Silaskivi. 2004. Tutkimus kilpailuoikeuden ja maatalouden sääntelyn yhteensovittamisesta
17. Aki Kangasharju. 1998. Regional Economic Differences in Finland: Variations in Income Growth and Firm Formation.
16. Pertti Kukkonen. 1997. Rahapolitiikka ja Suomen kriisi

Pellervon taloudellisen tutkimuslaitoksen raportteja, forskningsrapporter, Reports

194. Marko Mäki-Hakola – Mikko Toropainen. 2005. Metsien suojelun vaikutukset tuotantoon ja työllisyyteen – Alueellinen ja valtakunnallinen panos-tuotosanalyysi
193. Ari Peltoniemi. 2005. Työllisten työkyky vuonna 2004. Työpoliittinen tutkimus 273, Työministeriö, Pellervon taloudellisen tutkimuslaitoksen raportteja n:o 193.
192. Terhi Latvala – Antti Suokannas. 2005. Automaattisen lypsyjärjestelmän käyttöönotto: kannattavuus ja hankintaan vaikuttavat tekijät
191. Anna-Kaisa Rämö – Tapio Tilli – Ritva Toivonen – Pekka Ripatti – Karoliina Lindroos – Hanna Ruohola. 2005. Metsäverojärjestelmän muutos ja yksityismetsänomistajien puunmyyntiaikeet vuosina 2004 -2007
190. Janne Huovari – Raija Volk. 2004. Ikääntyminen ja maaseudun työmarkkinat
189. Martti Patjas. 2004. Production costs of milk, beef and pig meat in Finland, Sweden, Denmark and Germany
188. Tapio Tilli – Sten-Gunnar Skutin. 2004. Roundwood markets in the Baltic Sea region

Pellervon taloudellisen tutkimuslaitoksen työpapereita, diskussionsunderlag, Working Papers

80. Jukka Jalava. 2006. Production, primary, secondary, and tertiary: Finnish growth and structural change, 1860-2004
79. Meri Virolainen. 2005. Venäjän maatalous- ja elintarvikesektori muutoksessa
78. Heikki Lehtonen – Perttu Pyykkönen. 2005. Maatalouden rakennekehitys-näkymät vuoteen 2013
77. Janne Huovari – Raija Volk. 2005. Alueellisten asuntomarkkinoiden kehitys vuoteen 2008
76. Karoliina Lindroos. 2005. Metsänomistajien tavoitteet sekä tieto- ja neuvonta-tarpeet
75. Tapio Tilli. 2005. Itämeren alueen ja Venäjän metsäsektorin kehitysskenaariot vuoteen 2010 mennessä
74. Janne Huovari – Mikko Pakkanen – Raija Volk. 2005. Alueellisten asuntomarkkinoiden kehitys vuoteen 2007
73. Marko Mäki-Hakola. 2004. Metsien suojelun vaikutukset puumarkkinoilla - Mallitarkastelu
72. Eriikka Peltonen . 2004. Korkeakoulusta valmistuneiden alueellinen sijoittuminen
71. Jaakko Pulli – Marko Mäki-Hakola. 2004. Metsien suojelun taloudelliset vaikutukset. Kirjallisuuskatsaus.