

842

JYVÄSKYLÄN YLIOPISTO  
Taloustieteellinen osasto

**USIEN TEOLLISUUSYRITYSTEN PERUSTAMIS-  
STEIDEN SEUTUKUNNITTAINEN VAIHTELU**

Kansantaloustieteen pro gradu -tutkielma  
Huhtikuu 1998

Tekijä: Jussi Kaulio  
Ohjaaja: Hannu Tervo

## 1 Uusien teollisuusyritysten perustamisasteiden seutukunnittainen vaihtelu

Tekijä: Jussi Kaulio  
Ohjaaja: Hannu Tervo

Tieteenala: Kansantaloustiede  
Aika: Huhtikuu 1998  
Paikka: Jyväskylän yliopisto  
Sivumäärä: 51 + liitteet

Tämän pro gradu -tutkielman tarkoituksena oli löytää niitä tekijöitä, jotka ovat uusien teollisuusyritysten perustamisasteiden maantieteellisten eroavaisuuksien taustalla. Maantieteellisenä aluejakona käytettiin Suomen 88 seutukuntaa ja tiedot teollisuusyritysten perustamisasteista oli kerätty tarkasteluajanjaksolla 1.7.1994 - 31.12.1995.

Tutkielman keskeisiä lähteitä olivat Spillingin (1996) *Regional Variation of New Firm Formation: the Norwegian Case*, Audretschin ja Fritschin (1994) *The Geography of Firm Births in Germany*, Krugmanin (1991) *Geography and Trade* ja Gujaratin (1995) *Basic Econometrics*.

Tutkielman empiirissä osuudessa vastemuuttujan ja selittävien muuttujien välisiä riippuvuuksia tarkasteltiin ensin korrelaatiotarkasteluiden avulla. Vastemuuttujan eli uusien teollisuusyritysten perustamisasteiden vaihtelua seutukunnittain tarkasteltiin lineaarisen regressioanalyysin avulla. Vastemuuttujan arvon vaihtelua selittäviä muuttujia löydettiin kolme kappaletta. Nämä olivat palvelualan työpaikkojen osuus kaikista seutukunnan työpaikoista, yritystiheys eli yritysten määrä seutukunnassa sen väkilukuun suhteutettuna ja Maaseudun Puolueen osuus seutukunnan valtuustopaikoista. Näistä ensimmäisen muuttujan suuri arvo vaikutti teollisuusyritysten perustamisasteisiin negatiivisesti ja kahden jälkimmäisen muuttujan suuri arvo taas positiivisesti.

Avainsanat: uusien teollisuusyritysten perustamisaste, seutukunnittainen vaihtelu, lineaarinen regressioanalyysi.

# SISÄLLYS

<b>1 JOHDANTO</b> .....	<b>2</b>
1.1 PERUSTAMISASTEIDEN MITTAAMINEN .....	3
1.2 TUTKIMUKSEN TAVOITTEET JA RAJAUKSET .....	4
<b>2 TEORIOITA UUSIEN YRITYSTEN SYNNYSTÄ</b> .....	<b>5</b>
2.1 TYÖMARKKINAMALLI.....	5
2.2 RAKENNE-KÄYTÖS-SUORITUSKYKY -PARADIGMA.....	7
2.3 YMPÄRISTÖMALLI .....	8
2.4 LOKALISAATIO .....	10
<b>3 AIEMPIÄ TUTKIMUKSIA AIHEESTA</b> .....	<b>16</b>
3.1 UUSIEN YRITYSTEN PERUSTAMISASTEIDEN VAIHTELU NORJASSA .....	16
3.2 UUSIEN YRITYSTEN PERUSTAMISASTEIDEN VAIHTELU LÄNSI-SAKSASSA .....	20
<b>4 TEOLLISUUSYRITYSTEN SYNTYÄ SELITTÄVÄT EROT SEUTUKUNNITTAIN 1994-95</b> .....	<b>24</b>
4.1 TUTKIMUSASETELMA.....	24
4.1 PERUSTAMISASTEET SUOMEN MAAKUNNISSA.....	26
4.2 MALLISSA KÄYTETTÄVÄT MUUTTUJAT .....	27
4.3 MUUTTUJEN KORRELAATIOTARKASTELUT .....	30
4.4 KLASSINEN LINEAARINEN REGRESSIOMALLI .....	35
<b>5. YHTEENVETO</b> .....	<b>49</b>
<b>LÄHDEKIRJALLISUUS</b> .....	<b>53</b>
<b>LIITE 1 . UUSIEN TEOLLISUUSYRITYSTEN PERUSTAMISASTEET SEUTUKUNNITTAIN</b> .....	<b>55</b>

# 1 JOHDANTO

Taloudellisen hyvinvoinnin epätasainen jakautuminen maantieteellisesti on yleisesti tunnustettu tosiasia. Eroja taloudellisessa hyvinvoinnissa löytyy sekä valtakuntien väliltä että niiden sisältä. Taloudellisen hyvinvoinnin tasoa jollakin tietyllä alueella on perinteisesti mitattu erilaisilla talouden tunnusluvuilla, kuten alueen työttömyysasteella ja väkilukuun suhteutetulla bruttokansantuotteella ja varallisuudella. Taloustieteilijät ovat omalta osaltaan pyrkineet löytämään näiden tunnuslukujen taustalla olevia tekijöitä. Yhdeksi mielenkiintoiseksi tekijäksi on havaittu uusien yritysten muodostuminen alueelle. Työpaikat ovat yrityksissä ja niissä tuotetaan myös suuri osa alueen bruttokansantuotteesta. Yritysten olemassaolo on tärkeä tekijä alueen taloudelliselle kehitykselle. Osa yrityksistä lopettaa toimintansa ajan kuluessa, jotkut laajenevat ja työllistävät lisää ihmisiä ja jotkut yritykset taas pysyvät vuodesta toiseen melko staattisessa tilassa. Jotta yritykset voisivat menestyä, täytyy niiden ensin kuitenkin syntyä alueelle. Mitkä sitten ovat ne dynaamiset prosessit, jotka johtavat uusien yritysten syntyyn? Ovatko tällaiset prosessit universaaleja, vai toimivatko ne eri tavoin eri alueilla? Useissa eri maissa tehdyt tutkimukset ovat osoittaneet, että dynaamiset prosessit eivät ole maantieteellisesti riippumattomia, eli ne eivät toimi kaikkialla samalla tavoin. Monissa maissa voidaan tunnistaa alueita, joissa dynaamiset prosessit toimivat erityisen hyvin ja paljon tutkimuksia on tehty, jotta saataisiin selvitettyä, miksi jotkut alueet menestyvät paremmin kuin toiset (Frisk 1992, 1990, Keeble & Walker 1994 ja Westhead & Moyes 1992). Huomiota on kiinnitetty mm. alueen toimipaikkarakenteen ja paikalliskulttuurin vaikutuksiin uusien yritysten synnylle. Jos uusien yritysten syntyyn vaikuttavia tekijöitä pystytään luotettavasti löytämään, niin taloustieteilijöiden tuottama tieto on arvokasta päätöksentekijöille mm. aluepolitiikan linjoja muodostettaessa.

Tämän pro gradu -tutkielman pääpaino on uusien teollisuusyritysten synnyn tutkimisessa Suomessa. Uusien yritysten on havaittu olevan erittäin tärkeitä tekijöitä alueen taloudelliselle kehittymiselle. Davidsson ym. (1994) löysi positiivisen korrelaation uusien yritysten syntymisen ja taloudellisen kasvun, korkean hyvinvoinnin ja alhaisen työttömyyden väliltä. Korkean uusien yritysten perustamisasteen ja taloudellisen kasvun väliltä näyttääkin löytyvän positiivinen korrelaatio niin, että useimmilla alueilla, joissa on taloudellista kehitystä on, myös korkea uusien yritysten perustamisaste. Kaikilla alueilla, joissa on korkea uusien yritysten perustamisaste ei ole korkea taloudellisen hyvinvoinnin taso, mutta on harvinaista,

että alueella, jossa on korkea taloudellinen hyvinvointi, on myös alhainen uusien yritysten perustamisaste. Näin ollen korkea uusien yritysten perustamisaste on välttämätön, mutta ei yksinään riittävä tekijä taloudelliselle kasvulle.

### 1.1 Perustamisasteiden mittaaminen

Vertailtaessa uusien yritysten perustamisasteita eri alueiden välillä tarvitaan jokin standardoitu mittari, jotta alueiden välinen vertailu olisi ylipäättään mielekästä. Pelkkä uusien yritysten absoluuttinen määrä alueella ei täytä tällaiselta mittarilta vaadittavia ominaisuuksia, sillä eri alueet eroavat huomattavasti toisistaan ominaisuuksiltaan, kuten väestömääriltään ja toimialarakenteiltaan. Uusien yritysten syntyä käsittelevässä kirjallisuudessa voidaan löytää kaksi yleistä tapaa mitata uusien yritysten perustamisastetta jollakin alueella (Reynolds ym. 1994). Ensimmäinen niistä on työmarkkinamalli (labour market model) ja toinen yrityskantamalli (business population model).

Yrityskantamallin logiikka perustuu olettamukselle, että uusia yrityksiä syntyy jo olemassa olevista yrityksistä. Uusien yritysten perustamisaste saadaan suhteuttamalla alueelle tietyllä ajanjaksolla perustetut yritykset alueella jo oleviin yrityksiin (yleensä per 100 yritystä). Alueelle jo perustetut yritykset toimivat mallin mukaan hautomoina tuleville yrityksille ja ne toimivat myös tulevien yrittäjien harjoittelukenttinä valmistaen heitä oman yrityksen perustamiseen. Tehdyt tutkimukset ovat osoittaneet, että alueella jo olemassa oleva yrityskanta on tärkeä tekijä uusien yritysten syntymiselle. Mitä enemmän yrityksiä alueella jo on, niin sitä paremmat ovat edellykset myös uusien yritysten muodostumiselle.

Työmarkkinamallissa alueelle perustetut uudet yritykset suhteutetaan alueen väestömäärään. Audretschin ja Fritschin (1994) mukaan työmarkkinamallin teoreettinen vetovoima perustuu ajatukselle, että yrittäjäyys on aina yksilön valinta. Jokaisen uuden yrityksen perustaa aina joku ja mitä suuremmat ovat alueen työvoimaresurssit, niin sitä suurempi on myös potentiaalisten yrittäjien lukumäärä. Malli olettaa epäsuorasti, että uudet yrittäjät tulevat alueen omilta työvoimamarkkinoilta. Työmarkkinamallin mukaisesti lasketut uusien yritysten perustamisasteet saavat arvoja nollan ja ykkösen väliltä, eli alarajana olisi tilanne, että kukaan alueen asukkaista ei perustaisi omaa yritystä ja ylärajana taas tilanne, jossa jokainen

alueen asukas perustaisi oman yrityksen. Tällaiset tilanteet ovat kuitenkin teoreettisia ja käytännössä täysin absurdeja. Yleensä alueelle perustettujen uusien yritysten lukumäärä suhteutetaan per 1000 tai per 10 000 alueen asukkaaseen, jotta saadaan vertailukelpoisia lukuja eri alueiden välille.

Mitattaessa uusien yritysten perustamisasteita yrityskantamallin mukaisesti todetaan väistämättä mallin heikko kohta: se vääristää mittaustuloksia riippuen siitä, ovatko alueella sijaitsevat yritykset keskimääräistä suurempia vai pienempiä. Jos yritykset ovat keskimääräistä suurempia, niin saadaan korkeita uusien yritysten perustamisasteita ja yritysten ollessa keskimääräistä pienempiä saadaan matalia perustamisasteita. Erityisen korkeita uusien yritysten perustamisasteita saadaan ns. yhden tai muutaman tehtaan kaupungeissa.

## **1.2 Tutkimuksen tavoitteet ja rajaukset**

Pro gradu -tutkielmani tavoitteena on löytää syitä, miksi uusia yrityksiä syntyy epätasaisesti eri alueille Suomessa. Tulen esittelemään erilaisia uusien yritysten syntyteorioita kuvaavia malleja ja raportoin myös tärkeimmistä tutkimustuloksista, joita kyseisten mallien testaaminen on tuottanut. Pyrin myös johtamaan teorioista erilaisia hypoteeseja uusien yritysten syntymisten selittämiseksi ja tulen testaamaan hypoteeseja uusien teollisuusyritysten osalta.

Tutkielmani empiirinen osuus tarkastelee uusien teollisuusyritysten syntymistä Suomeen ajanjaksolla 1.7.1994 - 31.12.1995. Tiedot uusien teollisuusyritysten rekisteröinneistä on saatu Tilastokeskuksen yritysrekisteristä. Ajanjakso on valittu tarkasteluun sillä perusteella, että Tilastokeskus muutti rekisteröintiperusteita uusille yrityksille 1.7.1994 alkaen, eikä tätä aikaisempien tietojen käyttö olisi enää mielekäästä. Tietoja 31.12.1995 jälkeen perustetuista yrityksistä ei ollut käytettävissäni, joten edellämainitun ajanjakson valinta uusien teollisuusyritysten syntymisen tutkimiseksi oli varsin johdonmukaista.

## 2 TEORIOITA UUSIEN YRITYSTEN SYNNYSTÄ

Uusien yritysten syntyä käsittelevästä kirjallisuudesta voidaan löytää kolme erilaista teoreettista lähtökohtaa. Ensimmäinen tapa lähestyä ilmiötä on tutkia sitä työvoimamarkkinoiden kautta, jolloin painotetaan yrittäjyyttä yksilön tekemänä tietoisena valintana. Toinen teoreettinen lähtökohta on ns. rakenne-käytös-tulos -paradigma, jonka mukaan toimialan rakenne vaikuttaa yritysten käyttäytymiseen, joka vuorostaan vaikuttaa niiden tulokseen. Kolmas tapa hahmottaa ilmiötä on uusien yritysten synnyn ja alueellisen kehityksen välisen suhteen tutkiminen.

### 2.1 Työmarkkinamalli

Taloustieteen historiassa uusien yritysten markkinoille tulon tutkiminen on ollut perinteisesti toimialataloustieteilijöiden maaperää, kun taas itsensä työllistämisen tutkiminen on kuulunut työntaloustieteilijöille. Heidän käyttämänsä terminologiatkin ovat eronneet toisistaan. Toimialataloustieteilijät puhuvat markkinoille tulijasta ja työmarkkinaekonomistit itsensä työllistämisestä (Storey 1995, 61). Toimialaekonomistit ovat oletaneet, että uudet markkinoille tulijat odottavat jonossa tilaisuuttaan yrityksen perustamiseksi, kunhan vain tuotteen hinta nousisi pitkän aikavälin keskimääräisiä kustannuksia korkeammaksi. Tämä olettaus ei kuitenkaan ole tyydyttänyt kaikkia työntaloustieteilijöitä, vaan on alettu tutkia myös yksilön henkilökohtaisten ominaisuuksien vaikutusta uusien yritysten syntyyn.

Alueellisten taloudellisten olosuhteiden ja paikallisten toimintaympäristöjen lisäksi ihmisten henkilökohtaisten luonteenpiirteiden ja motivaatioiden on tunnistettu vaikuttavan keskeisesti uusien yritysten muodostumiseen. Edellyttäen, että nämä ominaisuudet vaihtelevat populaatioiden välillä eri alueilla, niin myös aktiivisuus yrittäjyyteen vaihtelee alueittain. Käsitys, että yrittäjät jakavat joitakin yhteisiä ominaisuuksia ei ole yllättävä ja se on johtanut tutkijoita etsimään yrittäjille yhteisiä luonteenpiirrejoukkoja. Huolimatta luonteenpiir-

remalliin kohdistetusta kritiikistä, sen käyttö on yleistä tutkimuksissa. Esimerkiksi Brockhaus (1982) on esittänyt lähes 50 yrittäjiä yhdistävää luonteenpiirrettä, joista suoritusmotivaatio ja kontrolliorientaatio ovat kaikkein yleisimmät. (Niittykangas & Tervo 1997, 124)

Suoritusmotivaatio kuvaa yksilön halua saavuttaa jotain elämässään tiettyjen omien standardien mukaan. Tällaisia voivat olla esimerkiksi halu toimia johtajana tai halu perustaa oma menestyvä yritys, luoda jotain tyhjästä. Yksilön kontrolliorientaatio kuvaa, miten hyvin hän pystyy omasta mielestään vaikuttamaan oman elämänsä tapahtumiin. Ne, joilla on vahva sisäinen kontrolliorientaatio, tuntevat pystyvänsä vaikuttamaan oman elämänsä kulkuun teoillaan, taidoillaan ja kyvyillään. Vastaavasti ihmiset joilla on vahva ulkoinen kontrolliorientaatio tuntevat, että kohtalo määrää heidän elämänsä kulun erilaisten tapahtumien muodossa, joihin heillä ei ole vaikutusmahdollisuuksia.

Gibbin ja Ritchien (1982) sosiaalisen kehityksen malli kritisoi näkemystä, että ihmisellä olisi jotkut syntymälahjana saadut luonteenpiirteet, joiden perusteella yrittäjät eroaisivat muusta väestöstä. Heidän mallinsa mukaan yrittäjyys voidaan paremmin ymmärtää yksilön kohtaamien tilanteiden ja hänen sosiaalisen viiteryhmänsä ehdoilla. Tämän näkökulman mukaan ihminen muuttuu koko elämänsä ajan ja hän muovautuu kohtaamiensa sosiaalisten tilanteiden ja sosiaalisten viiteryhmien mukana. Yksilön yhteiskuntaluokka, perheympäristö, koulutus, urakehitys ja sosiaaliset linkit vaikuttavat näin suuresti hänen kehitykseensä ja potentiaaliseen yrittäjyyteensä. Myös ihmisen elämänsä vaiheella on vaikutusta hänen motiiveihinsa perustaa oma yritys. Usein mainittu motiivi omalle yritykselle on halu työskennellä itsenäisesti, olla oman itsensä herra. Motiivit yrittäjyyteen voivat olla sekä positiivisia tai negatiivisia, vaihtoehtoisesti voidaan myös puhua veto- tai työntekijöistä yrittäjyyteen. Perinteisen talousteorian kanssa yhdenmukainen motiivi on varmastikin tiedostettu voiton mahdollisuus (Tervo & Niittykangas 1994, 40). Muita mahdollisia vetotekijöitä yrittäjyyteen voivat olla mm. itsensä toteuttamisen tarve, halu jatkaa perheyrittäjästä tai tunne siitä, että pystyy parempaan suoritukseen kuin oma työnantaja. Negatiivisia motiiveja eli työntekijöitä yrittäjyyteen ovat asiat, jotka saavat yksilön tuntemaan, että hänestä on tullut yrittäjä olosuhteiden pakosta. Tuleva yrittäjä kokee oman yrityksen perustamisen parhaana vaihtoehtonaan tai ainakin parempana vaihtoehtona, kuin pysymisen nykyisessä epätyytyttävässä elämäntilanteessaan. Tyypillinen työntekijä yrittäjyyteen on työttömyys tai sen uhka. Jos työttömyysaste on korkea tai se on kasvanut nopeasti jollakin alueella, niin yhä useampien ihmisten voidaan olettaa harkitsevan itsensä työllistämistä perustamalla oma



yritys avoimien työpaikkojen vähyden vuoksi. Toisaalta voidaan asettaa myös vastakkainen hypoteesi, jonka mukaan korkea työttömyysaste kuvaa talouden kykenemättömyyttä sopeutua muuttuviin olosuhteisiin ja tällöin kenties alueen väestöstä puuttuu yrittäjäainesta. Jos yrittäjäaines puuttuu väestöstä, niin korkeiden työttömyysasteiden alueille voidaan olettaa olevan matalat uusien yritysten perustamisasteet.

## 2.2 Rakenne-käytös-suorituskyky -paradigma

Rakenne-käytös-suorituskyky -paradigma on perinteisin tapa hahmottaa uusien yritysten syntymekanismeja. Teoria painottaa toimialalla jo toimivien yritysten hinnoittelu- ja tuotantopäätösten merkitystä toimialalle pyrkivälle uudelle yritykselle. Eri toimialoilla toimivat yritykset mielletään useimmiten teollisuusyrityksiksi, jotka tuottavat jotain tiettyä tuotetta.

Paradigmassa yrityksen suorituskyky markkinoilla on määritelty sangen laajasti, sisältäen dimensioita kuten allokatiivinen ja teknologinen tehokkuus sekä innovatiivisuus (Porter 1981). Suorituskyky riippuu ostajien ja myyjien käyttäytymisestä tuotteiden hinnoittelupoliitikassa, yritysten välisessä yhteistyössä, tutkimus- ja tuotekehittelyinvestoinneissa, tuotantolinjastrategioissa ym. Yritysten käytös taas riippuu toimialan markkinoiden rakenteesta, jonka määrittelevät mm. seuraavat ominaisuudet: ostajien ja myyjien lukumäärän ja koon jakauma, tuotteiden erikoistumisen aste ja markkinoille tulon esteiden suuruus uusille yrityksille.

Markkinoiden rakenteeseen ja käyttäytymiseen vaikuttavat myös tietyt perusolosuhteet. Kysyntä koostuu tekijöistä, kuten substituuttituotteiden saatavuus, markkinoiden kasvu, kuluttajien käyttäytyminen ja tuotteiden hintajousto. Tarjonta taas koostuu tekijöistä, kuten raaka-aineiden saatavuus, käytettävissä oleva teknologia, valittu tuotantoprosessi, poliittiset päätökset ym. Paradigman kausaalisuus kulkee rakenteesta käytökseen ja edelleen suorituskykyyn lukuunottamatta joitakin markkinoiden häiriötilanteita, jolloin suunta voi olla päinvastainen. (Niittykangas & Tervo 1997, 122)

Toimialalle tulevat uudet yritykset eivät välttämättä ole juuri äskettäin perustettuja, vaan niitä voi olla erilaisia. Mueller (1992) määrittelee uuden yrityksen sellaiseksi, joka valmis-

taa tuotetta jollakin toimialalla tietyssä vuonna, mutta ei valmistanut sitä edellisessä vuonna. Uuden yrityksen ei välttämättä tarvitse olla juuri äskettäin perustettu, vaan se voi olla myös toimialalla jo toimivan yrityksen perustama tai ostama uusi tuotantolaitos.

Storey (1995) on esittänyt seuraavan empiirisen mallin toimialalle tulevalle yritykselle:

$$E = f(\pi, BE, GR, C).$$

Mallissa E (entry) kuvaa jollekin toimialalle tuloa. Se on funktio voitoista ( $\pi$ ), markkinoille tulon esteistä (BE), toimialan kasvusta (GR) ja toimialan keskittyneisyydestä (C). Toimialalle tulon jälkeisten odotettujen voittojen kasvu laskee yrityksen markkinoille tulon kynnystä, kuten tekee myös toimialan odotettu kasvu. Markkinoille tulon esteet, kuten mitta-kaavaetujen olemassaolo ja raaka-aineden rajoitettu saanti laskevat uuden yrityksen halukkuutta tulla toimialalle. Vastaava vaikutus on myös hyvin keskittyneillä toimialoilla, joilla jo toimivilla yrityksillä on mahdollisuus minimoida uuden yrityksen toimialalle tulon mahdollisuudet.

### 2.3 Ympäristömalli

Yrittäjyyden ja uusien yritysten synnyn tutkiminen on osoittanut, että kaikki alueet eivät ole potentiaaliltaan samanlaisia tuottaakseen uusia yrittäjiä (Malecki 1994, 125). Ympäristömalli on läheistä sukua rakenne-käytös-suorituskyky-paradigmalle, ja se korostaa paikallisen toimintaympäristön merkitystä uusien yritysten synnylle.

Ympäristömalli olettaa korkeiden uusien yritysten perustamisasteiden edellyttävän pienyritysvaltaista toimintaympäristöä tuleville yrityksille. Suurten yritysten oletetaan olevan huonoja hautomoita tuleville yrittäjille verrattuna pieniin yrityksiin. Tämän oletetaan johtuvan mm. siitä, että suuryrityksessä traditio työskennellä toisen alaisuudessa on suurempi kuin pienyrityksessä ja näinollen kynnys oman yrityksen perustamiselle voi suuryrityksen työntekijälle olla pienyrityksen työntekijää korkeampi. Pienyrityksessä työntekijä voi nähdä urakehityksensä kulkevan luonnollisesti kohti oman yrityksen perustamista hänen ollessa siihen

valmis. Pienyritysvaltaisuus viittaa myös sellaiseen paikalliseen toimialarakenteeseen, jossa markkinoille tulon kynnyks on uudelle yrittäjälle alhainen. Suuryritysten dominoivilla alueilla markkinolle tulon esteet voivat olla yrittäjälle ylivoimaisia.

Myös paikallisen työvoiman koulutustasolla voidaan katsoa olevan yhteys uusien yritysten perustamisasteisiin. Mitä paremmin koulutettua ja hallinnollisia valmiuksia omaavaa paikallinen työvoima on, sitä paremmat edellytykset alueella on myös uusien yritysten synnylle. Ainoastaan mekaanisia liukuhihnavalmiuksia omaavien työntekijöiden mahdollisuudet perustaa oma yritys arvioidaan mataliksi. Toisaalta voidaan myös esittää hypoteesi, jonka mukaan tällaisilla alueilla ihmisillä on paremmat mahdollisuudet saada työpaikka suuryrityksestä tarvitsematta ottaa yrittäjyydestä aiheutuvia riskejä.

Ympäristömalli liittyy läheisesti myös paikalliseen kulttuuriin, jonka merkitys uusien yritysten synnylle voi olla hyvinkin suuri. Yrityksen menestymiseen vaikuttavat työnjohdon ja työntekijöiden taidot ovat ainakin osaksi paikallisen kulttuurin vaikutuspiirissä ja ne siirtyvät sukupolvelta toiselle. Samoin paikallinen arvostus yrittäjiä kohtaan kannustaa oman yrityksen perustamiseen, kuten tekee myös heidän toimintansa roolimalleina tuleville yrittäjille. Mikäli tällaiset paikalliskulttuuriin sidotut ominaispiirteet puuttuvat alueelta, voi niitä olla vaikea tuoda sinne ulkopuolelta ja uusien yritysten perustamisasteet pysyvät matalina.

Alueen vauraudella voidaan olettaa myös olevan merkitystä uusien yritysten synnylle. Mitä vauraampi alue on, sitä korkeampien voidaan uusien yritysten perustamisasteiden ajatella siellä olevan. Tämä johtuu siitä, että vauraiden alueiden asukkailla on muita alueita keskimääräistä korkeammat käytettävissä olevat tulot ja tulojoustavien hyödykkeiden kysynnän määrä on alueella korkea mahdollistaen niitä tarjoavien pienyritysten synnyn. Varakkaiden ihmisten on myös helpompi saada rahoitusta oman yrityksen perustamiseen kuin köyhempien ja kenties vielä pienemmällä pääomakustannuksilla, joten heidän on helpompaa perustaa oma yritys.

Paikallinen poliittinen ilmapiiri voidaan myös liittää uusien yritysten syntyä koskevaan tarkasteluun. Jos paikallishallinnon menot ovat suuret, niin voidaan esittää hypoteesi, jonka mukaan tämä johtaa alueen kokonaiskysynnän kasvuun ja näin myös korkeampiin uusien yritysten perustamisasteisiin kuin alueilla, joilla paikallishallinnon menot ovat pienemmät.

Paikallishallinnon koostumuksella saattaa myös olla merkitystä uusien yritysten perustamisasteisiin, mutta vaikutuksen suunta on epäselvä. Vasemmistovoittoisen paikallishallinnon voidaan ajatella kasvattavan julkisia menoja oikeistovoittoista enemmän johtaen näin kokonaiskysynnän kasvuun ja korkeisiin uusien yritysten perustamisasteisiin. Toisaalta julkisten menojen kasvattaminen pitää rahoittaa kenties korkeammalla kunnallisveroasteella, joka laskee paikallista kokonaiskysyntää ja myös uusien yritysten perustamista. Voidaan myös esittää hypoteesi, jonka mukaan vasemmistovetoisia paikallishallintoja ei todennäköisesti löydy pienyritysvaltaisilta alueilta.

## 2.4 Lokalisaatio

Lokalisaatio eli saman toimialan tuotantotoiminnan taipumus pyrkiä keskittymään alueellisesti on jo pitkään ollut taloustieteessä tunnettu ilmiö. Marshall (1920) esitteli kolme lokalisoitumiseen vaikuttavaa tekijää, nimittäin yhteiset työmarkkinat, yhteiset välituotemarkkinat ja teknologiset vuodot.

### *Yhteiset työmarkkinat*

Marshallin (1920) alkuperäisen ajatuksen mukaan yhteisistä työmarkkinoista hyötyvät sekä työnantajat että työntekijät. Työnantajat sijoittuvat mieluusti alueille, joissa on saatavilla heidän toimialallaan vaadittavia taitoja omaavia työntekijöitä ja työntekijöiden kannattaa muuttaa alueille, jossa heidän taidoilleen on kysyntää. Eristyksissä toimivan työnantajan on hyvinä aikoina vaikea saada lisää ammattitaitoista työvoimaa, kun taas huonoina aikoina työntekijöiden on vaikea löytää töitä.

Marshall ei tarjonnut hypoteeseilleen minkäänlaista formaalia esitystä. Vasta Krugman (1991) pystyi sellaisen taloustieteelle tarjoamaan. Krugmanin malli yhteisille työmarkkinoille lähtee liikkeelle olettamuksesta, että on olemassa joukko yrityksiä. Jokaista yritystä edustaa tuottofunktio  $R_i$ , missä työvoima on ainoa muuttuja. Tuottofunktion oletetaan yksinkertaisuuden vuoksi olevan toisen asteen yhtälömuotoa.

$$R_i = \alpha + (\beta + \varepsilon_i)L_i - \frac{\gamma}{2}L_i^2. \quad (1)$$

Jokaisen yrityksen tuotot riippuvat nyt vakiotuloista  $\alpha$  ja työvoiman rajatuottavuudesta ( $\beta + \varepsilon_i - \gamma L_i$ ).  $L_i$  on yrityksen työvoiman kysyntä ja  $\varepsilon_i$  kuvaa työvoiman rajatuottavuudessa tapahtuvia yrityskohtaisia shokkeja. Yritysten mahdollisuus toimia oligopolisteina työmarkkinoilla on jätetty mallissa tarkastelun ulkopuolelle. Jokainen yritys asettaa työvoiman rajatuottavuuden vallitsevan palkkatason ( $w$ ) suuruiseksi:

$$w = \beta + \varepsilon_i - \gamma L_i \quad (2)$$

Tästä voidaan johtaa työvoiman kysyntäfunktio jokaiselle yritykselle:

$$L_i = \frac{\beta + \varepsilon_i - w}{\gamma}. \quad (3)$$

Työmarkkinoiden oletetaan selviävän :

$$\sum_{i=1}^n L_i = L. \quad (4)$$

missä  $\sum_{i=1}^n L_i$  on kaikkien työnantajien työvoiman kysyntä ja  $L$  on työvoiman tarjonta.

Vallitseva palkkataso riippuu tarjolla olevien työntekijöiden määrästä yritysten, yritysten määrästä ja yksittäisten yritysten kohtaamista shokeista:

$$w = \beta - \frac{\gamma L}{n} + \frac{1}{n} \sum_i \varepsilon_i. \quad (5)$$

Yksittäisten yritysten kohtaamien shokkien oletetaan olevan korreloimattomia keskenään varianssilla  $\sigma^2$ . Odotettu palkkataso riippuu tällöin vain työvoiman ja yritysten lukumäärän suhteesta:

$$Ew = \beta - \frac{\gamma L}{n}. \quad (6)$$

Odotetun palkkatason varianssi riippuu yritysten lukumäärästä:

$$\text{var}(w) = \frac{\sigma^2}{n}. \quad (7)$$

Yrityskohtaisten shokkien ja palkkatason kovarianssi riippuu myös yritysten lukumäärästä:

$$\text{cov}(w, \varepsilon_1) = \frac{\sigma^2}{n}. \quad (8)$$

Yrityksen voitot ovat tulot vähennettyinä palkkakustannuksilla:

$$\pi = R_i - wL_i. \quad (9)$$

Yhtälöistä 1, 5 ja 9 voidaan ratkaista yritykselle voitot työvoiman, yritysten lukumäärän ja yrityskohtaisten shokkien funktiona. Sieventämisen jälkeen saadaan yhtälö odotetuille voitoille:

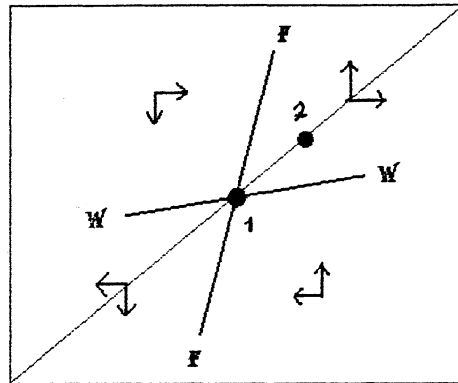
$$E\pi = \alpha + \frac{1}{\gamma} \left( \frac{L}{n} \right)^2 + \frac{1}{2} \frac{n-1}{n} \sigma^2. \quad (10)$$

Näin on johdettu yhtälöt odotetulle palkkatasolle ja odotetuille voitoille yritysten ja työvoiman lukumäärän funktiona tietyllä alueella.

Kuviossa 1 on esitetty kuvitteellinen tilanne, jossa on olemassa kaksi aluetta, itä ja länsi, joiden väliltä sekä yritysten että työnantajien tulee valita. Horisontaalisella akselilla on esitetty lännen osuus työvoimasta ja vertikaalisella lännen osuus yrityksistä. Yritysten kokonaismäärä on  $n$ , josta  $n_1$  yritystä sijoittuu länteen ja  $n_2$  yritystä itään. Vastaavasti  $L$  on työntekijöiden kokonaismäärä ja  $L_1$  työntekijää sijoittuu länteen ja  $L_2$  työntekijää itään.

Lännen osuus

yrittäjistä



Kuva 1: työvoiman ja  
yrittäjien siirtäminen

Lännen osuus

yrittäjistä

Kuviossa 1 piste 1 kuvaa tilannetta, jossa työntekijät ja yritykset ovat jakautuneet puoliksi idän ja lännen välille. Käyrä  $WW$  yhdistää ne pisteet, joissa työntekijät ovat indifferenttejä idän ja lännen välillä. Koska odotettu palkkataso riippuu vain yritysten ja työvoiman lukumäärien suhteesta, niin  $WW$  on suora, jonka kulmakerroin on  $L/n$ .

Lännessä toimivan yrityksen voittojen erotus idässä toimivan yritykseen voittoihin:

$$\pi_1 - \pi_2 = \frac{1}{\gamma} \left[ \left( \frac{L_1}{n_1} \right)^2 - \left( \frac{L_2}{n_2} \right)^2 \right] + \frac{1}{2} \sigma^2 \left[ \frac{n_1 - 1}{n_1} - \frac{n_2 - 1}{n_2} \right]. \quad (11)$$

Kun yritykset ja työntekijät ovat jakautuneet likimain puoliksi idän ja lännen välille, eli  $n_1 = n_2$  ja  $L_1 = L_2$ , niin yritysten ja työntekijöiden uudelleen allokoinnin vaikutukset lännelle ovat:

$$\frac{d(\pi_1 - \pi_2)}{dL_1} = \frac{4L}{\gamma n^2} > 0 \text{ ja}$$

$$\frac{d(\pi_1 - \pi_2)}{dn_1} = -\frac{4L^2}{\gamma n^3} - \frac{1}{2}\sigma^2 \frac{1}{n_1^2} < 0.$$

Käyrä  $FF$  muodostuu niistä pisteistä, joissa yritykset ovat indifferenttejä idän ja lännen välillä eli  $\pi_1 = \pi_2$ .  $FF$  on ylöspäin nouseva ja se on myös jyrkempi kuin  $WW$ . Tämä voidaan todistaa joko yhtälöistä (12) ja (13) tai kuvioista 1 tarkastelemalla. Pisteessä 2 yritysten ja työntekijöiden suhde on sama sekä idässä että lännessä, mutta lännessä on enemmän molempia. Näin ollen länsi tarjoaa paremmat edut yhteisistä työmarkkinoista sekä työntekijöille että yrityksille ja molemmat preferoivat länttä. Yhtälöstä (10) voidaan nyt nähdä, että yritykset ovat voitokkaampia lännessä. Näin ollen  $FF:n$  täytyy olla  $WW:n$  yläpuolella pisteessä 2.

Nuolet kuviossa 1 kuvaavat mallin dynamiikkaa. Työntekijät siirtyvät kohti aluetta, jolla odotettu palkkataso on korkeampi ja yritykset siirtyvät kohti aluetta, jolla odotetut voitot ovat suuremmat. Piste 1 on epävakaa; systeemi ajautuu kohti kulmaa eli kaikki työntekijät ja yritykset keskittyvät samaan paikkaan.

### ***Yhteiset välituotemarkkinat***

Marshallin (1920) toinen syy agglomeraatioon on yhteisten välituotemarkkinoiden olemassaolo. Lokalisoitunut toimiala voi tukea enemmän erikoistuneita paikallisia alihankkijoita, mikä puolestaan parantaa toimialan tehokkuutta ja vahvistaa lokalisaatiota. (Krugman 1991, 49)

Yhteisten välituotemarkkinoiden olemassaolo riippuu ratkaisevasti tuotannon mittakaavaetujen olemassaolosta. Jos välituotteiden valmistamisessa ei olisi tuotannon mittakaavaetuja, niin silloin pienikin tuotannon maantieteellinen keskittymä voisi replikoida suurta keskittymää pienoiskoossa ja silti saavuttaa saman tehokkuuden. Ainoastaan mittakaavaetujen olemassaolo mahdollistaa sen, että suurilla tuotannon maantieteellisillä keskittymillä on tehokkaammat ja monipuolisemmat alihankkijat kuin pienillä tuotantokeskuksilla.

Krugmanin (1991) mukaan välituotteiden kuljetuskustannukset voivat heikentää lokalisoitumisen edellytyksiä ainoastaan, jos ne ovat erittäin alhaiset verrattuna lopputuotteen kul-



jetuskustannuksiin. Yleinen kuljetuskustannusten hintatason lasku rohkaisee lokalisoitumista Krugmanin mukaan pikemminkin kuin heikentää sen mahdollisuuksia.

Todistaakseen edellä olevat väittämät Krugman (1991) kehitti mallin, jossa välituotteet ja lopputuotteet mielletään samaksi asiaksi. Mallissa ajatellaan olevan ryhmä tuotteiden valmistajia, joista jokaisen valmistamaa tuotetta käytetään sekä lopputuotteena että välituotteena (panoksena) kaikille muille ryhmän tuotteiden valmistajille. Jos esimerkiksi ajatellaan, että tyypillisen ryhmän tuotteen myynnit ovat 10, mutta näistä 4 on myyntiä muille ryhmän valmistajille panokseksi. Vastaavasti oletetaan, että näiden 10 yksikön tuottamiseksi tarvitaan 4 yksikköä välituotteita panoksiksi, jotka myös hankitaan samalta toimialalta (muilta ryhmän valmistajilta). Näin jokainen tuote on sekä välituote että lopputuote ja niillä käytävä kauppa on luonteeltaan samanlaista. Seuraavaksi oletetaan, että on olemassa kaksi mahdollista paikkaa tuotannolle, itä ja länsi. Kummassakin on puolet lopputuotteena käytettävän hyödykkeen kysynnästä, eli 3 yksikköä kutakin tuotetta kohden. Jos kaikki yritykset ovat sijoittuneet itään, niin myös markkinoille tuleva yritys haluaa sijoittua itään. Tämä johtuu siitä, että idässä on 7 yksikköä tuotteen kokonaiskysynnästä (3 lopputuotteena ja 4 välituotteen), mikä kannustaa yritystä sijoittamaan tuotantonsa sinne. Yrityksen sijoittumista itään kannustaa myös se, että panoksina käytettävät välituotteet valmistetaan idässä ja ne ovat halvempia siellä hankittuina. (Krugman 1991, 50-51)

### *Teknologiset vuodot*

Teknologisilla vuodoilla tarkoitetaan niitä ulkoishyötyjä, jotka muodostuvat toistensa läheisyydessä sijaitsevien yritysten tietojen ja taitojen vuodoista toisilleen. Paikalliset yritykset hyötyvät teknisten valmistusmenetelmien maantieteellisestä läheisyydestä ja teknologian ylivuotaminen kyvykkäiltä yrityksiltä alueen muille yrityksille parantaa heikompien yritysten kilpailukykyä. Teknologiseen kyvykkyyteen liittyvä tieto ja osaaminen voi vuotaa helposti niillekin, jotka eivät ole osallistuneet sen kehittämisen kustannuksiin. Koska tiedon ja teknologian merkitys on nykyään tuotannossa suuri, niin teknologiset vuodot voidaan nähdä merkittävänä tekijänä lokalisoitumiskehityksessä.

### **3 Aiempia tutkimuksia aiheesta**

Uusien yritysten synty on ollut viime vuosina varsin tutkittu ilmiö taloustieteessä. Aiheesta löytyy runsaasti kansainvälisiä tutkimustuloksia, jotka ovat tuottaneet mielenkiintoisia tuloksia eri maiden aineistoilla. Tässä luvussa esitellään kahden ulkomaisen tutkimuksen metodiikka ja raportoidaan niiden tuottamista tuloksista. Tutkimuksista ensimmäinen (Spilling 1996) käsittelee uusien yritysten syntyä Norjassa ja se on valittu mukaan esiteltäväksi Norjan maantieteellisen läheisyyden vuoksi. Jälkimmäinen esiteltävistä tutkimuksista (Audretsch ja Fritsch 1994) tarkastelee aiemmin esitellyn Krugmanin (1991) lokalisoitumista käsittelevän teorian yhteensopivuutta Länsi-Saksasta peräisin olevaan aineistoon.

#### **3.1 Uusien yritysten perustamisasteiden vaihtelu Norjassa**

Spilling (1996) tutki uusien yritysten perustamisasteiden alueellista vaihtelua Norjassa. Tutkimuksessa Spilling jakoi Norjan 104:ään itsenäiseen työmarkkina-alueeseen. Jokaiselle alueelle laskettiin uusien yritysten perustamisaste viiden vuoden ajanjaksolla 1988-92, sekä työmarkkinamallilla että yrityskantamallilla. Työmarkkinamallissa uusien yritysten perustamisaste laskettiin suhteuttamalla alueelle ajanjaksolla perustettujen yritysten määrä alueen väestömäärään. Yrityskantamallissa taas perustamisaste saatiin suhteuttamalla alueelle ajanjaksona perustettujen yritysten määrä alueella jo toimiviin yrityksiin. Alkutuotantosektori, julkiset palvelut, sekä pankki-, rahoitus- ja vakuutus toiminta talouden yksityiseltä sektorilta jätettiin tarkastelun ulkopuolelle.

Yrityskantamallin mukaan lasketut korkeimmat uusien yritysten perustamisteet löytyivät suurilta urbaaneilta alueilta ja pienistä kaupungeista, mutta myös joillakin kaikkein periferisimmillä alueilla oli korkeita perustamisasteita. Matalimmat uusien yritysten perustamisasteet löytyivät periferia-alueilta, raskaan teollisuuden alueilta sekä teollisen monokulttuurin alueilta, kuten esimerkiksi yhden kalatehtaan dominoimilta paikkakunnilta.

Työmarkkinamallilla lasketut korkeimmat uusien yritysten perustamisasteet löytyivät kaikkein urbanisoiduimmilta sekä kaikkein periferisimmiltä alueilta. Matalimmat perustamisasteet olivat monissa yhden yrityksen kaupungeissa, sekä joillakin periferisillä alueilla ja pienissä kaupungeissa.

Spilling teki lineaariset regressioanalyysit sekä työmarkkinamallille että yrityskantamallille siten, että molemmissa oli vastemuuttujana uusien yritysten perustamisaste. Koko talouspoislukien aiemmin mainitut toimialat kuului tarkastelun piiriin. Työmarkkinamallille saatiin selitysasteeksi noin 70% ja yrityskantamallille noin 40% (taulukko1). Työmarkkinamallin tärkeimmät selittävät muuttujat olivat yritystiheys (yrityksiä per asukas alueella), 24-44 -vuotiaiden osuus väestöstä, konservatiivisten puolueiden 1989 parlamenttivaaleissa saama osuus kaikista äänistä sekä pienissä yrityksissä (alle 20 työntekijää) työskentelevien työntekijöiden osuus alueen työvoimasta. Yrityskantamallin tärkeimmät selittävät muuttujat olivat samat kuin työmarkkinamallin, mutta sillä merkittävällä erolla, että yritystiheys sai negatiivisen arvon. Tälle nähtiin selitykseksi yrityskantamallin jo aiemmin kommentoidut heikkoudet.

**Taulukko 1: Perustamisasteiden alueellisia eroja selittävät tekijät sekä yrityskantamallin että työmarkkinamallin mukaan.**

Selittävä muuttuja	Yrityskantamalli		Työmarkkinamalli	
	Regressiokerroin	Merk.taso	Regressiokerroin	Merk.taso
25-44 -vuotiaiden osuus väestöstä	.358	***	.515	****
Yritystiheys	-.263	*	.750	****
Työvoiman osuus pienissä yrityksissä			.187	**
Konservatiivisten puolueiden äänisaalis	.215	*	.287	****
Mukautettu selitysaste R <sup>2</sup>	.398		.702	
F-arvo	23.7	****	61.5	****

Koska työmarkkinamalli antoi korkeamman selitysasteen kuin yrityskantamalli, niin tästä eteenpäin tutkimusta jatkettiin vain sen näkökulmasta. Aiemmin talous oli käsitelty yhtenä kokonaisuutena regressiomalleja muodostettaessa, nyt se jaettiin kuuteen sektoriin. Vastemuuttujana oli edelleen uusien yritysten perustamisaste.

**Taulukko 2: Regressiomallit toimialoittain työmarkkinamallin mukaan.**

Selittävä muuttuja	Teollisuus	Rakentaminen	Vähittäismyynti/ Hotelli- ja ravintolapalvelut	Kuljetus	Yrityspalvelut	Henkilökohtaiset palvelut
Väestö						
% Korkea-asteen tutkinto Kasvu tutkinnoissa 1985-89	.165*				.595****	
Työttömyys						
Työttömyysaste 1990		-.172*				
Työttömyysasteen nousu 1985-90		.309****				.344****
Talous						
Keskim.hlö.koht. tulot 1990	-.149*					
Sijainti						
Alueindeksi					-.336****	
Työvoiman rakenne						
% Julkisella sektorilla						.219**
% Omalla sektorilla	.589****	.412****	.387****	.337***		
Heterogeenisuus indeksi		-.147*				
Toimipaikkarakenne						
Yritystiheys	.275***	.353****	.331****	.235*	.186**	.267**
% työvoimasta pienillä toimipaikoilla	.601****	.156*				
% työvoimasta suurilla toimipaikoilla	-.281*					
Poliittinen ilmapiiri						
Konservatiiviset puolueet			.374***		.198*	
Keskusta puolueet			-.198*			
Valtion avustusohjelmat						
Tuet, lainat ja vakuudet					.268****	
Adjustoitu selitysaste R2	.527	.514	.622	.134	.646	.133
F-arvo	20.1****	19.1****	43.4****	8.9***	38.6****	6.3****

Jokaiselle muodostetulle talouden sektorille tehtiin oma regressioanalyysi ja neljälle sektorille kuudesta saatiin selitysasteeksi yli 50% (teollisuus, rakentaminen, vähittäismyynti / hotelli- ja ravintolapalvelut ja yrityspalvelut). Vaikkakin eri sektoreiden välisistä regressiomalleista löytyi eroja, niin niissä on myös havaittavissa samankaltaisuuksia.

Yritystiheys eli yritysten kokonaismäärä alueella suhteutettuna 1000 alueen asukkaaseen oli kaikkein yleisin muuttuja eri malleissa ja se oli tilastollisesti merkitsevä kaikille talouden sektoreille. Koko talouden käsittävässä mallissa oli yritystiheys tilastollisesti kaikkein merkitsevin selittävä muuttuja ja se sai näin suurimman myös regressiokertoimen estimaatin arvon. Tämä havainto viittaa siihen, että korkealla yritystiheydellä on positiivinen vaikutus uusien yritysten perustamisasteeseen. Koska korkea yritystiheys vaatii suuren pienten yritysten suhteellisen määrän kaikista alueen yrityksistä, niin se voidaan myös tulkita pienten yritysten muuttujaksi. Tämä tulkinta vahvistui teollisuus- ja rakennussektoreiden kohdalla, kun havaittiin että työvoiman korkea suhteellinen määrä pienillä toimipaikoilla edesauttoi uusien yritysten perustamista. Teollisuuden osalta havaittiin, että työvoiman korkea suh-

teellinen määrä suurilla toimipaikoilla sai negatiivisen regressiokertoimen arvon, eli se heikensi uusien yritysten syntymahdollisuuksia.

Yritystiheyden jälkeen seuraavaksi tärkein muuttuja oli toimialoittaisen erikoistumisen taso eli kuinka suuri osuus alueen koko työvoimasta työskentelee kullakin sektorilla. Työvoiman keskittyminen jollekin sektorille lisäsi uusien yritysten syntymistä kyseiselle sektorille. Tämä muuttuja oli mukana neljän sektorin mallissa.

Väestön ominaispiirteitä kuvaavat muuttujat eivät olleet kovin merkittäviä talouden eri sektoreiden regressiomalleissa. Korkea-asteen tutkinnon suorittaneiden osuus väestöstä oli kaikkein tärkein selittävä muuttuja yrityspalvelusektorille ja kasvu korkea-asteen tutkintojen määrässä oli myös selittävänä muuttujana teollisuudelle. Edellisen havainnon selittää hyvin yrityspalvelujen keskittyminen suuriin asutuskeskuksiin, mistä löytyy myös suurin suhteellinen osuus korkea-asteen tutkinnon suorittaneista henkilöistä. Jälkimmäistä havaintoa on melkoa vaikea selittää, mutta yksi syy korkea-asteen tutkintojen kasvun ja uusien yritysten synnyn positiivisen korrelaation välillä voi olla se, että korkeita uusien yritysten perustamisasteita ilmeni periferisillä alueilla, missä on yleisesti matala koulutustaso, mutta missä on viime aikoina ollut kasvua korkea-asteen tutkintojen määrässä (Spilling 1996, 234).

Työttömyyttä kuvaavat muuttujat ovat mukana kolmessa eri mallissa. Työttömyysaste korreloi negatiivisesti uusien yritysten perustamisasteen kanssa rakennussektorilla ja kasvu työttömyysasteessa positiivisesti rakennussektorilla ja yrityspalveluissa. Voidaankin tulkita, että vain muutos työttömyydessä toimi työntekijänä stimuloiden uusien yritysten syntyä (Spilling 1996, 234). Työttömyyttä kuvaavilla muuttujilla oli merkitystä vain yksittäisillä sektoreilla, ei silloin kuin talous käsitetään yhtenä kokonaisuutena.

Poliittista ilmapiiriä kuvaavat muuttujat eivät olleet kovin merkittäviä yksittäisille talouden sektoreille. Malleista löytyykin vain kolme tilastollisesti merkitsevää regressiokertoimen arvoa. Kaksi näistä liittyi vähittäismyynti-, hotelli- ja ravintolasektoriin, missä oli positiivinen korrelaatio konservatiivisten puolueiden saamien äänien ja uusien yritysten perustamisasteen välillä, sekä negatiivinen korrelaatio keskustapuolueiden saamien äänien ja uusien yritysten perustamisasteen välillä. Tämä voidaan tulkita siten, että näillä sektoreilla oli korkeimmat perustamisasteet urbaaneilla alueilla, missä konservatiivisten puolueiden kannatus on korkea, kun taas keskustapuolueiden kannatus on vahvempi maaseutumaisilla alueilla.

Valtion avustusohjelmat ja uusien yritysten perustamisaste korreloivat positiivisesti yritys-palvelusektorilla. Tämän selittäminen on hieman epäselvää, sillä avustusohjelmat on suunnattu pääasiassa teollisuuden kehittämiseen. Eräs selitys voi kuitenkin olla, että alueet, jotka saavat tukea avustusohjelmista, tarjoavat markkinoita myös yrityspalveluille. (Spilling 1996, 235).

### 3.2 Uusien yritysten perustamisasteiden vaihtelu Länsi-Saksassa

Audretsch ja Fritsch (1994) ovat tutkineet Krugmanin (1991) lokalisoitumista käsittelevän teorian yhteensopivuutta Länsi-Saksasta peräisin olevaan empiiriseen aineistoon. Selitettävänä muuttujana heidän tutkimuksessaan oli uusien yritysten perustamisaste, jota tarkasteltiin sekä yrityskanta- että työmarkkinamallin mukaan. Audretsch ja Fritsch katsovat korkean uusien yritysten perustamisasteen indikoivan osaltaan taloudellisten voimavarojen keskittymistä spatiaalisesti, kun taas matala uusien yritysten perustamisaste ilmaisee päinvastaista prosessia. Korkea uusien yritysten perustamisaste kuvaa siis lokalisoitumista, jonka lähteitä ovat Krugmanin mukaan yhteiset työmarkkinat, yhteiset välituotemarkkinat ja teknologiset vuodot.

**Taulukko 3: Selittävät muuttujat**

Selittävä muuttuja	Hypoteesi
Työttömyysaste, 1985 (%)	?
Muutos työttömyysasteessa, 1984-86 (%)	?
Väestötiheys, 1985	+
Väestönkasvu, 1980-85 (%)	+
Ammattitaidottomien työntekijöiden osuus työvoimasta, 1985 (%)	-
Bruttoarvonlisäys per asukas, 1986	+
Toimipaikan keskikoko, 1986	-

Taulukossa 3 ovat ne selittävät muuttujat, joilla Audretsch ja Fritsch pyrkivät selittämään uusien yritysten perustamisasteen vaihtelua vuosina 1986-89 75 alueella Länsi-Saksassa. Jokaiselle muuttujalle on asetettu hypoteesi siitä, miten kyseisen muuttujan arvon kasvu vaikuttaa vastemuuttujan arvoon. Väestötiheyden ja varsinkin väestön kasvun katsotaan kuvaavan lokalisoitumista ja urbanisoitumista, jotka tuottavat agglomeraatiohyötyjä pienentyneiden kaupankäyntikustannusten muodossa. Agglomeraatiohyötyjä oletetaan löytyvän myös alueilta, joilla BKT per asukas on suhteellisen korkea. Työttömyysasteen vaikutuksesta uusien yritysten perustamisasteeseen ei asetettu täysin yksiselitteistä hypoteesia. Korkea työttömyysaste viestii Audreschin ja Fritschin mukaan alueen hitaasta taloudellisesta kasvusta heikentäen yritysten kannustimia sijoittua alueelle. Toisaalta korkea työttömyysaste tarjoaa paremmat potentiaaliset työvoimaresurssit alueelle tulevalle yritykselle. Ammattitaidottoman työvoiman korkean osuuden koko alueen työvoimasta oletettiin vaikuttavan negatiivisesti uusien yritysten perustamisasteisiin.

**Taulukko 4: Regressiomallin tulokset yrityskantamallin mukaan.**

Selittävä muuttuja	Koko talous	Teollisuus	Palvelut
Työttömyysaste	0.002 (2.41) <sup>1</sup>	0.003 (2.80)	0.001 (0.58)
Muutos työttömyysasteessa	0.002 (3.33)	0.001 (2.04)	0.001 (2.23)
Väestötiheys	0.001 (3.40)	0.001 (2.54)	0.001 (1.79)
Väestönkasvu	0.049 (2.94)	0.031 (1.78)	0.072 (3.91)
Ammattitaidottoman työvoiman osuus	-0.002 (-1.95)	-0.001 (-2.31)	-0.001 (-1.63)
Bruttoarvonlisäys per asukas <sup>2</sup>	0.014 (2.25)	0.007 (1.76)	0.002 (0.43)
Toimipaikan keskikoko	0.003 (1.84)	0.003 (1.46)	0.006 (3.30)
Vakio	0.253 (6.79)	0.031 (1.78)	0.332 (8.05)
Selitysaste R <sup>2</sup>	0.669	0.552	0.473
F-arvo	22.32	14.025	10.475

1. t-arvot suluisissa 2. Kerrottu 1000:lla esityksen selventämiseksi

Taulukossa 4 on esitetty Audretschin ja Fritschin tekemän regressiomallin tulokset yrityskantamallille. Sekä työttömyysasteella että sen muutoksella vuosina 1984-86 oli positiivinen vaikutus uusien yritysten perustamisasteisiin. Myös väestötiheys ja västönkasvu vuosina 1980-85 vaikuttivat positiivisesti vastemuuttujaan. Jälkimmäiset tulokset ovat yhtäpitäviä Krugmanin (1991) teorian kanssa yhdistäen agglomeraatiot taloudellisen aktiviteetin keskittymiseen.

Ammattitaidottoman työvoiman osuus vuoden 1985 kokonaistyövoimasta sai negatiivisen regressiokertoimen arvon. Tämä viittaa siihen, että yrityksillä on suurempi taipumus sijoittua alueille, joilla työvoima on ammattitaitoista, kuin alueille, joilla työvoima koostuu pääasiassa ammattitaidottomista työntekijöistä. Tulos on konsistentti Krugmanin teorian kanssa, että yhteisiin työmarkkinoihin ja teknologiavuotoihin liittyvät edut ovat merkittävämmässä asemassa ammattitaitoisen kuin ammattitaidottoman työvoiman kohdalla.

Uusien yritysten perustamisasteet olivat myös korkeita alueilla, joilla bruttoarvonlisäys per asukas oli suuri vuonna 1986. Tämä tulos viittaa siihen, että uusi taloudellinen aktiviteetti keskittyy alueille, joilla työvoiman tuottavuus on suurinta. Toimipaikan keskikoko vaikutti myös uusien yritysten perustamisasteisiin. Mitä suurempi toimipaikan keskikoko on alueella, niin sitä vähemmän on työnantajia työvoimalle. Uusien yritysten perustamisasteet olivat systemaattisesti suurempia alueilla, joilla toimipaikan keskikoko oli suuri. (Audretsch & Fritsch 1994, 362).

Regressiomallin tulokset vaihtelivat huomattavasti eri talouden sektoreilla. Esimerkiksi työttömyysaste ei vaikuttanut juuri mitenkään uusien yritysten perustamisasteisiin palvelusektorilla, mutta taas teollisuudessa ja koko talous huomioituna sen positiivinen vaikutus on ilmeinen. Vastaavasti myöskään bruttoarvonlisäys per henkilö ei ollut tilastollisesti merkitsevä palvelusektorilla, mutta sillä oli suuri merkitys teollisuudessa ja koko taloudessa.



**Taulukko 5: Regressiomallin tulokset työmarkkinamallin mukaan.**

Selittävä muuttuja	Koko talous	Teollisuus	Palvelut
Työttömyysaste	-0.074 (-0.97) <sup>1</sup>	-0.061 (-2.77)	-0.149 (-2.15)
Muutos työttömyysasteessa	0.113 (2.83)	-0.015 (-1.32)	0.106 (2.93)
Väestötiheys	0.111 (3.81)	0.013 (1.53)	0.096 (3.63)
Väestönkasvu	3.026 (2.23)	0.211 (0.55)	1.602 (1.31)
Ammattitaidottoman työvoiman osuus	-0.139 (-2.17)	-0.005 (-0.27)	-0.119 (-2.05)
Bruttoarvonlisäys per asukas <sup>2</sup>	0.099 (1.95)	-0.027 (-1.86)	0.090 (1.97)
Toimipaikan keskikoko	-1.668 (-11.99)	-0.167 (-4.21)	-0.811 (-6.45)
Vakio	45.546 (14.99)	6.448 (7.42)	26.873 (9.79)
Selitysaste R <sup>2</sup>	0.839	0.471	0.635
F-arvo	56.271	10.405	19.411

1. T-arvot suluissa

2. Kerrottu 1000:lla esityksen selventämiseksi

Taulukossa 5 on esitetty regressiomallin tulokset työmarkkinamallin mukaan laskettuna. Vastemuuttuja on edelleen uusien yritysten perustamisaste vuosina 1985-89. Työmarkkinamallin mukaan lasketut regressiokertoimien arvot poikkeavat kahdella merkittävällä tavalla yrityskantamallin vastaavista. Ehkä merkittävin ero on se, että työttömyysasteen kertoimet ovat negatiivisia. Tämä viittaa siihen, että sellaisille alueille, joilla on korkeat työttömyysasteet syntyy vähän uusia yrityksiä (koko taloudelle laskettua kerrointa ei voida pitää tilastollisesti merkitsevänä). Korkean työttömyysasteen ja matalan uusien yritysten perustamisten välistä yhteyttä voidaan tulkita kahdella eri tavalla. Ensimmäinen tulkinta on, että alttius oman yrityksen perustamiseen on matalampi työttömällä kuin työssä käyvillä. Toinen tulkinta on se, että alttius oman yrityksen perustamiselle on matalampi korkean työttömyyden alueilla riippumatta työllisyysstatuksesta.

Toinen merkittävä ero yrityskantamalliin verrattuna on se, että toimipaikan keskikoon regressiokertoimen arvo on negatiivinen työmarkkinamallissa, mutta positiivinen yrityskantamallissa. Tämä voidaan tulkita niin, että kynnyks on oman yrityksen perustamiselle on matalampi työntekijöille, joilla on työkokemusta pienistä yrityksistä.

## **4 Teollisuusyritysten syntyä selittävät erot seutukunnittain 1994-95**

### **4.1 Tutkimusasetelma**

Tässä tutkielmassa muodostetun empiirisen osuuden tarkoituksena on selvittää, mitkä tekijät vaikuttavat Suomen eri seutukuntien välisiin eroihin uusien teollisuusyritysten perustamisasteissa. Edellä esitellyistä erilaisten uusien yritysten syntyä selittävistä teorioista merkittävämmäksi nousee erityisesti työmarkkinamalli, jonka perusteella on laskettu uusien teollisuusyritysten perustamisasteet. Yrityskantamalli on jätetty tarkastelun ulkopuolelle sen jo aiemmin esiteltyjen teoreettisten heikkouksien perusteella. Sekä Spilling (1996) että Audretsch ja Fritsch (1994) löysivät uusien teollisuusyritysten syntyyn vaikuttavia tekijöitä, joiden merkitystä on testattu tutkielmassa käytettävissä olleella Suomen aineistolla.

Aineisto uusien teollisuusyritysten rekisteröinneistä on saatu Tilastokeskuksen yritysrekisteristä. Ajanjakso (1.7.1994 - 31.12.1995) on valittu tarkasteluun sillä perusteella, että Tilastokeskus muutti rekisteröintiperusteita uusille yrityksille 1.7.1994 alkaen, eikä tätä aikaisempien tietojen käyttö olisi enää mielekäästä. Tietoja 31.12.1995 jälkeen perustetuista yrityksistä ei ollut käytettävissäni, joten edellämainitun ajanjakson valinta uusien teollisuusyritysten syntymisen tutkimiseksi oli varsin johdonmukaista.

Uusien teollisuusyritysten perustamisasteita vertailtaessa on aluejaon perustaksi valittu seutukuntataso. Muita mahdollisia vaihtoehtoja aluejaoksi olisivat olleet kuntataso ja maakuntataso, joista tekijä kuitenkin piti ensimmäistä liian pienenä ja jälkimmäistä liian suurena yksikkönä aluetalouden perustaksi. Seutukunnan ollessa aluejaon perustana oletetaan

myös, että uudet yrittäjät tulevat seutukunnan sisältä. Näin ei todellisuudessa varmastikaan aivan tarkkaan ole, mutta tässä tutkielmassa ei asiaan kiinnitetä enempää huomiota.

Tässä tutkielmassa tarkastellaan siis nimenomaan uusien teollisuusyritysten perustamisasteita seutukunnittain. Teollisuus on valittu tarkastelun kohteeksi talouden sektoreista sillä perusteella, että sitä ovat tutkineet mm. Spilling (1996) että Audretsch ja Fritsch (1994) ja näin voidaan vertailla Suomen aineistolla saatuja tuloksia ulkomaisiin tutkimustuloksiin.

Tutkielman empiirisessä osassa rakennetaan klassinen lineaarinen regressiomalli, jossa vastemuuttujana on seutukunnittainen uusien teollisuusyritysten perustamisaste eli *syntytyvo*. Teollisuusyritysten perustamisasteet seutukunnittain (liite 1) laskettiin seuraavan kaavan mukaan:

$$\text{perustamisaste} = \frac{\left( \frac{\text{aloit942} + \text{aloit95}}{\text{väkiluku90}} \right) * 10000}{1.5}$$

Muuttuja *aloit942* on 1.8.1994 ja 31.12.94 välisenä aikana perustettujen teollisuusyritysten lukumäärä ja *aloit95* on 1.1.1995 ja 31.12.1995 välisenä aikana perustettujen teollisuusyritysten lukumäärä eri seutukunnissa. Näiden muuttujien summa on jaettu seutukunnan vuoden 1990 väestömäärällä. Paras mahdollinen vaihtoehto olisi tietysti ollut käyttää periodin alun väestömäärää, mutta koska sitä ei ollut käytettävissä, niin valitsin aineistosta seuraavaksi parhaan vaihtoehdon. Laskettu osamäärä on kerrottu kymmenellä tuhannella, jotta saataisiin esityksen kannalta helpommin vertailtavia lukuja. Koska uusien teollisuusyritysten rekisteröinnit ovat puolentoista vuoden pituiselta ajanjaksolta, niin perustamisasteet on vielä muutettu vuosikeskiarvoiksi jakamalla edellä laskettu luku 1.5:llä.

Regressiomallin kehittämissä on käytetty mahdollisimman paljon samoja selittäviä muuttujia, kuin Spilling (1996) ja Audretsch ja Fritsch (1994) käyttivät omissa tutkimuksissaan. Käytettävissä ollut aineisto rajasi pois joitakin selittäviä muuttujia. Kaikki tarkastelussa mukana olleet selittävät muuttujat on aggregoitu kuntatasolta seutukuntatasolle ja ne on esitelty taulukossa 7.

Regressiomallin rakentaminen aloitettiin vastemuuttujan ja selittävien muuttujien välisillä korrelaatiotarkasteluilla (taulukko 8). Tästä jatkettiin selittävien muuttujien välisiin korrelaatiotarkasteluihin, jonka jälkeen aloitettiin varsinaisen regressiomallin rakentaminen lisäämällä siihen selittäviä muuttujia yksi kerrallaan.

#### **4.1 Perustamisasteet Suomen maakunnissa**

Koska seutukunta on melko pieni aluetalouden yksikkö, niin tutkielman empiirisen osuuden aluksi tarkastellaan miten perustamisasteet vaihtelevat maakunnittain Suomessa, jotta saataisiin jonkinlainen yleiskäsitys perustamisasteiden maantieteellisistä eroavaisuuksista. Suomen 19 eri maakunnan väliltä löytyi huomattavia eroja uusien teollisuusyritysten perustamisasteiden kesken työmarkkinamallin mukaan laskettuna. Teollisuusyritysten perustamisasteet maakunnissa laskettiin seuraavan saman kaavan mukaan kuin perustamisasteet seutukunnittainkin. Erona oli luonnollisesti se, että kaavan muuttujien arvoina käytettiin maakuntakohtaisia tunnuslukuja.

Taulukossa 6 on esitetty uusien teollisuusyritysten perustamisasteet Suomen maakunnissa 1.8.1994 ja 31.12.1995 välisenä aikana vuosikeskiarvoina. Kaikkien maakuntien yhteenlaskettu perustamisasteiden keskiarvo oli 5.93 uutta teollisuusyritystä 10 000 asukasta kohti. Suurimman ja pienimmän perustamisasteen eli Etelä-Pohjanmaan ja Kainuun välinen ero oli noin 2.8 -kertainen. Etelä- ja Keski-Pohjanmaan perustamisasteet olivat omaa luokkaansa. Seuraavaan ryhmään suuruusjärjestyksessä voidaan katsoa kuuluvan Pirkanmaan ja Satakunnan. Koko maan keskiarvon yläpuolella olivat vielä Vaasan rannikkoseudun, Varsinais-Suomen, Päijät-Hämeen ja Keski-Suomen perustamisasteet. Maakunnittain lasketut perustamisasteet eivät kerro mitään kokonaistaloudellisesta uusien yritysten perustamisaktiiviteetista, vaan ne kertovat kehityksestä vain teollisuusyritysten osalta. Jotkut muut toimialat voivat siis kukoistaa maakunnassa, vaikka teollisuusyritysten osalta tilanne näyttääkin heikolta. Näin ei voida tehdä johtopäätöksiä aluetaloudellisesta kehityksestä vain yhden, vaikkakin suuren talouden sektorin perusteella.

**Taulukko 6: Perustamisasteet maakunnittain.**

Maakunta	Perustamisaste
Etelä-Pohjanmaa	9.72
Keski-Pohjanmaa	9.41
Pirkanmaa	7.44
Satakunta	7.00
Vaasan rannikkoseutu	6.53
Varsinais-Suomi	6.41
Päijät-Häme	6.22
Keski-Suomi	6.22
Uusimaa	5.58
Pohjois-Karjala	5.43
Lappi	5.28
Etelä-Savo	5.19
Kymenlaakso	5.12
Ahvenanmaa	4.88
Häme	4.60
Pohjois-Pohjanmaa	4.41
Etelä-Karjala	4.34
Kainuu	3.44

#### 4.2 Mallissa käytettävät muuttujat

Väkilukuun suhteutettua uusien teollisuusyritysten perustamisastetta Suomen seutukunnissa, eli muuttujaa *syntytyo*, pyritään selittämään regressioanalyysillä taulukon 7 selittävien muuttujien avulla. Selittävät muuttujat on sijoitettu eri kategorioihin niin, että niillä pyritään kuvailemaan kunkin kategorian edustamaa ilmiötä. Jokaiselle selittävälle muuttujalle on asetettu hypoteesi (H) siitä, miten sen oletetaan vaikuttavan vastemuuttujaan *syntytyo*. Suurimmalle osalle selittävistä muuttujista hypoteesi on positiivinen, eli mitä suurempi muuttujan arvo, sitä paremmat ovat edellytykset uusien teollisuusyritysten synnylle. Joidenkin muuttujien kohdalle on hypoteesiksi merkitty kysymysmerkki, eli niiden vaikutuksesta selitettävään muuttujaan ei ole asetettu mitään hypoteesia. Negatiivinen hypoteesi taas olettaa

muuttujan korkean arvon heikentävän uusien teollisuusyritysten syntymahdollisuuksia. Selittävät muuttujat on pyritty valitsemaan siten, että niiden mittaamisajankohta edeltäisi vastermuuttujan *syntytyö* mittaussajankohtaa. Useimpien muuttujien kohdalla tämä ei ole kuitenkaan ollut mahdollista aineiston saatavuudesta johtuen. Selittävät muuttujat ovat kuitenkin luonteeltaan sellaisia, että useimmissa tapauksissa vain kuuden kuukauden viivästyminen mittaussajankohdassa tuskin aiheuttaa merkittävää muutosta niiden arvoissa.

**Taulukko 7: Selittävät muuttujat**

Kategoria	H	Muuttuja	Selitys
Väestö	+	Vakil90	Väkiluku 31.12.1990
	+	Kasvu	Väkiluvun kasvu 1990-1995
	+	Vaika25p	25-44 -vuotiaiden osuus väestöstä, %, 1995
	+	Vakitih	Väestötiheys, as./km <sup>2</sup> , 1.1.1995
	+?	Vatutk1	Keskiasteen tutkinnon suorittaneiden osuus väestöstä, 31.12.1994
	+?	Vatutk2	Korkea-asteen tutkinnon suorittaneiden osuus väestöstä, 31.12.1994
Työttömyys	+	Tyotv93	Työttömyysasteen kuukausikeskiarvo 1993
	+	Tyotnous	Työttömyysasteen kasvu 1993 -1995
Aluetalous	+	Bktas	Bruttokansantuote asukasta kohti
Toimipaikkarakenne	+	Yrtiheys	Yrityksiä per väestö 1.7.1994
	-	Yrkeski	Toimipaikkojen keskikoko, 31.12.1994
Työvoiman rakenne	-	Tytpel1p	Alkutuotannon työpaikkojen osuus, %, 31.12.1994
	+	Tytpel2p	Jalostuksen työpaikkojen osuus, %, 31.12.1994
	?	Tytpel3p	Palvelujen työpaikkojen osuus, %, 31.12.1994
Poliittinen ilmapiiri	?	Valtkesk	Keskustan osuus valtuustopaikoista, 1.1.1996
	+	Valtkok	Kokoomuksen osuus valtuustopaikoista 1.1.1996
	?	Valtrkp	Ruotsal. kansanpuolueen osuus valtuustopaikoista,
	-	Valtsdp	Sosialidemokraattien osuus valtuustopaikoista
	?	Valtskl	Kristillisen liiton osuus valtuustopaikoista
	?	Valtsmp	Maaseudun puolueen osuus valtuustopaikoista
	-	Valtvas	Vasemmistoliiton osuus valtuustopaikoista
?	Valtvih	Vihreiden osuus valtuustopaikoista	

Väestöä kuvaavien muuttujien oletetaan pääosin olevan positiivisessa vuorovaikutussuhteessa uusien teollisuusyritysten synnyn kanssa. Muuttujan *vaika25p* sisältämä ikäryhmä väestöstä on valittu sillä ennako-oletuksella, että kyseisen ikäryhmän oletetaan olevan sen ikäistä, jolloin oma yritys todennäköisimmin perustetaan. Väestön kouluttautuneisuutta kuvaavien muuttujien *vatuk1* ja *vatuk2* vaikutuksesta vastemuuttujaan *syntytyo* ei ole asetettu täysin yksiselitteistä hypoteesia. Joidenkin teollisuusyritysten synnylle, kuten tietoliikenteenkniikan toimialaan kuuluvien, korkea koulutusaste on varmaankin olennainen edellytys. Toisaalta joillakin teollisuuden toimialoilla korkeasta koulutusasteesta saattaa olla suoranaista haittaa uusien yritysten synnylle. Tämän vuoksi muuttujien *vatuk1* ja *vatuk2* hypoteesien perään on asetettu kysymysmerkit.

Työttömyyden vaikutusta vastemuuttujaan *syntytyo* käsitellään muuttujilla *tyotv93* ja *tyotnous*. Molemmilla oletetaan olevan positiivinen vaikutus uusien teollisuusyritysten syntyyn. Työttömyyden tai sen uhkan ajatellaan toimivan työntötekijänä kohti yrittäjyyttä, eli ulkoiset olosuhteet ajavat yksilöä kohti yrittäjyyttä.

Aluetaloutta kuvaavan muuttujan *bktas* oletetaan vaikuttavan positiivisesti uusien teollisuusyritysten syntyyn. Ajatuksen taustalla on oletus, että uusien teollisuusyrityksen perustaminen vaatii taloudellisia resursseja ja mitä paremmat ne ovat alueella, niin sitä todennäköisemmin uusia yrityksiä myös syntyy.

Toimipaikkarakennetta kuvaillaan muuttujalla *yrittystiheys* ja *yrkesk*. *Yrtiheys* kertoo kuinka paljon seutukunnassa on yrityksiä väkilukuun suhteutettuna. *Yrittystiheys* kertoo myös epäsuorasti seutukunnan toimipaikkarakenteen pien- tai suuryritysvaltaisuudesta. Korkea *yrittystiheyden* arvo viittaa korkeaan pienyritysvaltaisuuteen ja matala arvo taas suuryritysvaltaisuuteen. Seutukunnan pienyritysvaltaisuuden oletetaan edistävän uusien teollisuusyritysten syntyä ja suuryritysvaltaisuuden heikentävän sitä. *Yrkesk* kertoo mikä on yritystoimipaikkojen keskikoko henkilökunnan määrällä mitattuna seutukunnissa. Oletuksena on, että muuttujan arvon kasvu vaikuttaa negatiivisesti vastemuuttujaan *syntytyo*.

Työvoiman rakennetta kuvaavat selittävät muuttujat *tytpellp* - *tytpel3p* kertovat, kuinka suuri osuus työvoimasta seutukunnittain työskentelee kullakin talouden sektorilla. Talous on jaettu kolmeen sektoriin eli alkutuotantoon, jalostukseen ja palveluihin.. Alkutuotannossa työskentelevien ei oleteta edesauttavan uusien teollisuusyritysten syntyä, eli hypoteesia on

merkitty negatiivisella etumerkillä. Jalostuksessa ja palveluissa työskentelevien on oletettu vaikuttavan positiivisesti teollisuusyritysten perustamisasteisiin. Käytettävissä olleen tilastoaineiston pohjalta ei valitettavasti ole voitu tehdä jakoa sen perusteella, kuuluvatko työpaikat talouden yksityiselle vai julkiselle sektorille. Yksityisen sektorin työpaikkojen voitaisiin ajatella vaikuttavan enemmän uusien teollisuusyritysten syntyyn kuin julkisen sektorin työpaikkojen.

Poliittista ilmapiiriä kuvaillaan eri puolueiden suhteellisilla osuuksilla valtuustopaikkojen lukumääristä 1.1.1996, jotka on aggregoitu kuntatasolta seutukuntatasolle. Oikeistopuolueista Kokoomuksen kannatuksen kasvun on oletettu olevan positiivisessa vuorovaikutussuhteessa uusien teollisuusyritysten synnyn kanssa, kun taas vasemmistopuolueiden kannatuksen on ajateltu heikentävän niiden syntymahdollisuuksia. Rkp:n, Keskustan, Smp:n, Skl:n ja Vihreiden kannatuksesta ei ole asetettu hypoteesia.

#### 4.3 Muuttujien korrelaatiotarkastelut

Jotta regressioanalyysin suorittaminen olisi mielekästä, tulee selitettävän ja selittävien muuttujien välillä olla joko positiivista tai negatiivista korrelaatiota. Koska sekä vastemuuttuja *syntytyö* että taulukon 7 selittävät muuttujat ovat kaikki vähintään välimatkaasteikkollisia, niin korrelaatiokertoimien arvot lasketaan käyttämällä Pearsonin tulomomenttikorrelaatiokerrointa:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}.$$

Taulukossa 8 on esitetty vastemuuttujan *syntytyö* ja selittävien muuttujien väliset korrelaatiokertoimet ja niiden tilastolliset merkitsevyydet. Yhdellä tähdellä (\*) merkityt korrelaatiokertoimien arvot ovat tilastollisesti merkitseviä 95%:n luottamusvälillä ja kahdella tähdellä merkityt 99%:n luottamusvälillä.



**Taulukko 8: Vastemuuttujan ja selittävien muuttujien väliset korrelaatiot.**

Selittävä muuttuja	Korrelaatiokerroin / Syntytyo	p-arvo
Vakil90	-.070	.515
Kasvu	-.059	.586
Vaika25p	-.268*	.012
Vakitih	-.057	.597
Vatutk1	-.292**	.006
Vatutk2	-.252*	.018
Tyot93	-.095	.381
Tyotnous	-.195	.068
Bktas	-.224*	.036
Yrtiheys	.325**	.002
Yrkeski	-.348**	.001
Tytpel1p	-.350**	.001
Tytpel2p	.012	.909
Tytpel3p	-.398**	.000
Valtkesk	.212*	.050
Valtkok	.166	.126
Valtrkp	-.076	.486
Valtsdp	-.215*	.047
Valtskl	.068	.534
Valtsmp	.231*	.033
Valtvas	-.181	.095
Valtvihr	-.341**	.001

Väestön rakennetta kuvaavat selittävät muuttujat *vakil90*, *kasvu*, *vaika25p* ja *vakitih* korreloivat negatiivisesti vastemuuttujan *syntytyo* kanssa, eli niiden arvon kasvu seutukunnissa näyttäisi heikentävän uusien teollisuusyritysten syntymahdollisuuksia. Nämä tulokset ovat päinvastaisia kuin taulukossa 7 muuttujista asetetut hypoteesit. Myös muuttujat *vatutk1* ja *vatutk2* korreloivat negatiivisesti vastemuuttujan kanssa, eli väestön koulutustason nousu ei nosta uusien teollisuusyritysten perustamisastetta seutukunnissa. Tämäkin tulos on ainakin lievästi päinvastainen kuin muuttujista asetetut hypoteesit olettavat.

Työttömyyttä kuvaavat muuttajat *tyot93* ja *tyotnous*, eli työttömyysasteen kuukausikeskiarvo vuonna 1993 ja työttömyysasteen kasvu vuosina 1990 - 1995 korreloivat molemmat negatiivisesti vastemuuttujan kanssa. Nämä tulokset eivät tue työntöteorioita yrittäjyyteen ainakaan teollisuusyritysten kohdalla. Muista talouden sektoreista ei voida teollisuusyritysten perusteella tehdä vastaavia päätelmiä.

Aluetalouden vahvuutta kuvaava muuttuja *bktas* eli bruttokansantuote asukasta kohti seutukunnassa, korreloi negatiivisesti vastemuuttujan kanssa. Tulos on päinvastainen muuttujalle asetetun hypoteesin kanssa, eli seutukunnan vahva talous ei edesauta teollisuusyritysten perustamista. Yhtenä mahdollisena syynä tähän voi olla mm. valtion avustusohjelmat, jotka houkuttelevat yrityksiä taantuneille alueille. Mitään kovin varmoja johtopäätöksiä ei tuloksen perusteella voida kuitenkaan tehdä.

Seutukunnan toimipaikkarakennetta kuvaavat selittävät muuttajat *yrtiheys* ja *yrkeski* korreloivat molemmat niille asetetun hypoteesin mukaisesti uusien teollisuusyritysten perustamisasteen kanssa. *Yrtiheys* eli yritysten määrä per väestö seutukunnalla korreloi positiivisesti vastemuuttujan *syntytyo* kanssa. Tämän tuloksen mukaan mitä enemmän seutukunnassa on yritystoimipaikkoja per väestö, niin sitä paremmat mahdollisuudet on myös uusilla teollisuusyrityksillä syntyä. Selittävä muuttuja *yrkeski* eli toimipaikan keskikoko seutukunnassa korreloi negatiivisesti vastemuuttujan *syntytyo* kanssa. Tämän tuloksen mukaan yritysten keskikoon kasvaessa henkilökunnan määrällä mitattuna seutukunnassa heikkenevät myös uusien teollisuusyritysten syntymahdollisuudet. Näinollen alueen pienyritysvaltaisuus olisi tärkeä tekijä uusien yritysten muodostumiselle.

Työvoiman rakennetta kuvaavista muuttujista vain *tytpel2:n* kohdalla asetettu hypoteesi piti paikkansa, eli jalostussektorin työpaikkojen suhteellinen osuus ja teollisuusyritysten perustamisaste korreloivat positiivisesti keskenään. Alkutuotannon työpaikkojen osuus kaikista työpaikoista eli *tytpel1* korreloi asetetun hypoteesin vastaisesti positiivisesti vastemuuttujan kanssa. Palvelusektorin työpaikkojen osuus kaikista työpaikoista eli *tytpel3p* korreloi negatiivisesti vastemuuttujan kanssa.

Poliittista ilmapiiriä kuvailtiin eri puolueiden kuntatasolta seutukuntatasolle aggregoiduilla suhteellisilla valtuustopaikkojen lukumäärillä. Asetetut hypoteesit pitivät paikkansa, eli Kokoomuksen kannatus kasvu korreloi positiivisesti ja Sdp:n ja Vasemmistoliiton negatiivi-

sesti uusien teollisuusyritysten perustamisasteiden kanssa. Muista puolueista Keskustan, Skl:n ja Smp:n valtuustopaikkojen määrän kasvu korreloi positiivisesti ja Rkp:n ja Vihreiden negatiivisesti vastemuuttujan syntytyö kanssa.

Kaikkien selittävien muuttujien korrelaatiokertoimet vastemuuttujan *syntytyö* kanssa eivät poikenneet tilastollisesti merkitsevästi nolasta. Regressiomallin jatkokehittelyyn otettiin mukaan vain ne selittävät muuttujat, joiden korrelaatiokertoimet vastemuuttujan syntytyö kanssa olivat tilastollisesti merkitseviä. Poikkeuksena edellämämainitusta on selittävä muuttuja *tyotnous*, joka otettiin mukaan mallin jatkotarkasteluihin, vaikka se ei p-arvonsa osalta täyttänytkaan edellämämainittua kriteeriä. Sen saama p-arvo (0.68) on kuitenkin melko lähellä tilastollisesti merkitsevän rajaa (0.5), joten muuttuja otettiin mukaan huolimatta sen puutteista. Toinen toivottava ominaisuus sekä vastemuuttujalle että selittäville muuttujille on niiden ainakin likimain normaalijakautuneisuus. Normaalijakautuneisuutta tutkittiin Kolmogorov-Smirnovin sovitettestillä, jota voidaan pitää tilastollisesti varsin ”tiukkana” testinä. Allaolevassa taulukossa 9 on esitetty sekä mukaan otettujen selittävien muuttujien korrelaatiokertoimet vastemuuttujan kanssa että niiden saamat p-arvot Kolmogorov-Smirnovin testistä.

**Taulukko 9: Vastemuuttujan ja selittävien muuttujien väliset korrelaatiot sekä Kolmogorov-Smirnov -testin p -arvot.**

Muuttuja	Korrelaatiokerroin / Syntytyö	Kolmogorov-Smirnov -testin p-arvo
Syntytyö	1.000	.302
Tytpel3p	-.398**	.090
Tytpel1p	.350**	.329
Yrkeski	-.348**	.637
Valtvihr	-.341**	.005
Yrtiheys	.325**	.483
Vatutk1	-.292**	.993
Vaika25p	-.268*	.012
Vatutk2	-.252*	.063
Valtsmp	.231*	.049
Bktas	-.224*	.127
Valtsdp	-.215*	.231
Valtkesk	.212*	.885
Tyotnous	-.195	.687

Taulukossa 9 selittävät muuttujat on laitettu järjestykseen niiden vastemuuttujan kanssa saaman korrelaatiokertoimen tilastollisen merkitsevyyden mukaan. Ensimmäisenä taulukossa on esitetty vastemuuttuja, jonka jälkeen seuraavat selittävät muuttujat. Kolmogorov-Smirnov -testin nollahypoteesinä on, että muuttuja noudattaa normaalijakaumaa. Riskitasolla  $\alpha = 0.05$  eli 95% luottamusvälillä kolmen muuttujan nollahypoteesi hylätään, eli muuttujat eivät noudata normaalijakaumaa. Nämä ovat *valtvihr*, *vaika25p* ja *valtsmp*. Kaikki muuttujat ovat selittäviä muuttujia ja koska Kolmogorov-Smirnov -testi on varsin ”tiukka”, muuttujat pidetään edelleen mukana regressiomallin kehittämissä. Lisäksi vastemuuttujan normaalijakautuneisuus on regressioanalyysissä tärkeämpää kuin selittävien muuttujien, joten edellämainittujen selittävien muuttujien pitäminen mukana mallin kehittämissä on varsin perusteltua.

Regressioanalyysin seuraava vaihe oli tutkia, millaisia ovat selittävien muuttujien keskinäisten korrelaatiokertoimien arvot. Regressioanalyysissä selittävät muuttujat eivät saisi korreloida keskenään ainakaan voimakkaammin kuin ne korreloivat selitettävän muuttujan kanssa. Jos selittävien muuttujien väliset korrelaatiot ovat voimakkaita, niin ongelmana on multikollinearisuus. Matemaattisesti tämä tarkoittaa sitä, että jokin mallin selittävästä muuttujasta voidaan esittää muiden selittävien muuttujien lineaarikombinaationa. Multikollinearisuus aiheuttaa sellaisen ongelman, että vaikka ilmiötä  $y$  pystytään selittämään esimerkiksi muuttujien  $x_1$  ja  $x_2$  avulla, niin emme enää tiedä mikä osa selityksestä kuuluu kummallekin selittäväälle muuttujalle.

Taulukossa 10 on esitetty regressiomallin jatkokehittämissä mukana pidettävien selittävien muuttujien välisten Pearsonin tulomomenttikorrelaatiokertoimien arvot. Selittävät muuttujat on valittu seuraavasti. Ensimmäiseksi selittäväksi muuttujaksi regressiomalliin on valittu muuttuja *tytpel3p* sillä perusteella, että sillä on suurin korrelaatiokertoimen arvo vastemuuttujan syntytyö kanssa. Seuraavaksi taulukkoon 10 on valittu ne selittävät muuttujat, jotka korreloivat vähemmän muuttujan *tytpel3p* kanssa, kuin tämä korreloi vastemuuttujan kanssa. Näin on saatu yhteensä viisi regressiomallissa mahdollisesti käytettävää selittävä muuttujaa. Muuttujat on esitetty taulukon 10 ensimmäisessä sarakkeessa siten, että ylimmäisenä on muuttuja *tytpel3p* eli selittävä muuttuja, jolla on suurin korrelaatio vastemuuttujan syntytyö kanssa. Tämän jälkeen sarakkeessa on esitetty muut selittävät muuttujat vastaavassa järjestyksessä.

**Taulukko 10: Selittävien muuttujien väliset korrelaatiot.**

	Tytpel3p	Yrtiheys	Valtsmp	Valtsdp	Valtkesk
Tytpel3p	1.000				
Yrtiheys	-.141	1.000			
Valtsmp	-.168	-.014	1.000		
Valtsdp	.105	-.184	-.296**	1.000	
Valtkesk	-.237*	-.118	.474**	-.646**	1.000

Regressiomallin kehittelyyn lopullisesti valitut selittävät muuttujat ovat *tytpel3p*, *yrtiheys* ja *valtsmp*. Ne eivät korreloi keskenään enempää kuin ne korreloivat selitettävän muuttujan kanssa. Muuttuja *valtsdp* korreloi voimakkaammin muuttujien *valtsmp* ja *valtkesk* kanssa kuin se korreloi selitettävän muuttujan kanssa ja tästä syystä sitä ei otettu enää mukaan mallin jatkokehittelyyn. Myös muuttuja *valtkesk* korreloi voimakkaammin selittävien muuttujien *tytpel3p*, *valtsdp* ja *valtkesk* kanssa kuin selitettävän muuttujan kanssa ja tästä johtuen sekin poistettiin regressiomallin jatkokehittelystä.

#### 4.4 Klassinen lineaarinen regressiomalli

Tässä luvussa on tarkoituksenaan kehitellä klassinen lineaarinen regressiomalli käyttämällä selittävinä muuttujina muuttujia *tytpel3p*, *yrtiheys* ja *valtsmp*. Vastemuuttujana on edelleen *syntytyo*. Klassisen lineaarisen regressiomallin lineaarisuudella tarkoitetaan regressiomallin sitä ominaisuutta, että malli on parametrien suhteen lineaarinen, mutta selittävät muuttujat ovat kvantitatiivisia ja niiden arvojen suuruus otetaan huomioon (Ranta, Rita & Kouki 1997). Regressiomalli on rakennettu SPSS -ohjelmistolla lisäämällä malliin yksi kerrallaan selittäviä muuttujia. Ensimmäinen muodostettu regressiomalli sisältää siis vain yhden selitettävän muuttujan, joka oli *tytpel3p*. Se on valittu malliin ensimmäisenä sillä perusteella, että se korreloi voimakkaimmin selittävistä muuttujista vastemuuttujan kanssa. Mallin pa-

rametrieni estimoinnissa on käytetty pns -menetelmää. Regressiomalli voidaan esittää muodossa:

$$Y = \alpha + \beta X + \varepsilon,$$

missä  $Y$  = vastemuuttuja *syntytyö*,  $\alpha$  = vakiotermi,  $\beta$  = regressiokerroin ja  $\varepsilon$  = virhetermi. Osa  $\alpha + \beta X$  on mallin systemaattinen osa, ja se ilmaisee tavan, jolla vastemuuttujan  $Y$  keskiarvot riippuvat muuttujan  $X$  arvosta. Tekijä  $\varepsilon$  on virhetermi, jonka avulla kuvataan havaintojen vaihtelua vastaavan keskiarvon ympärillä.

Mallin alkuoletukset ovat seuraavat:

1. Selittävillä muuttujilla on kiinteät (ei stokastiset) arvot.
2. Virhetermien odotusarvo (keskiarvo) on nolla:  $E(\varepsilon_i) = 0$ .
3. Virhetermin varianssi on vakio eli homoskedastinen:  $\text{var}(\varepsilon_i) = \sigma^2$ .
4. Virhetermien kesken ei ole autokorrelaatiota:  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ .
5. Virhermit noudattavat jakaumaa:  $\varepsilon_i \sim N(0, \sigma^2)$ .
7. Selittävien muuttujien  $x_i$  ja  $x_j$  välillä ei vallitse täydellinen lineaarinen riippuvuus, eli selittävät muuttujat eivät ole täydellisen multikollineaarisia.

**Taulukko 11: Varianssitaulu yhdelle selittävälle muuttujalle.**

Vaihtelun lähde	Vapausasteet	Neliösumma	Keskineliö	F -testisuure	p -arvo
Regressio	1	66.442	66.442	16.197	.000
Jäännös	86	352.780	4.102		
Kokonaisvaihtelu	87	419.222			

Taulukossa 11 on esitetty varianssitaulu. Regressioneliösumma on systemaattista vaihtelua kuvaava kokonaisvaihteluneliösumman osa eli se osa vastemuuttujan *syntytyo* vaihtelusta, joka pystytään selittämään selittävän muuttujan *tytpel3p* avulla. Jäännöseliösumma on satunnaisvaihtelua eli se osa vastemuuttujan *syntytyo* vaihtelusta, joka jää selittämättä muuttujan *tytpel3p* avulla. Regressioneliösumman ja kokonaisvaihteluneliösumman osamäärä on sama kuin mallin selitysaste  $R^2$ , joksi saatiin .158. Prosenteiksi muunnettuna selittävän muuttujan *tytpel3p* avulla saatiin siis selitettyä 15.8% vastemuuttujan *syntytyo* kokonaisvaihtelusta. Keskineliöt saadaan, kun regressio- ja jäännöseliösummat jaetaan vapausasteillaan.

F -testisuureella testataan hypoteesia:

$H_0$  : ”mallissa ei ole regressiota”, eli vastemuuttujan vaihtelu ei riipu selittävän muuttujan vaihtelusta

$H_1$  : ”mallissa on regressiota”, eli vastemuuttujan vaihtelu riippuu selittävän muuttujan vaihtelusta.

F -testisuureen arvo lasketaan seuraavasti: 
$$F = \frac{SS_R}{SS_E / (n - 2)}$$
,

missä  $SS_R$  on regressioneliösumma ja  $SS_E$  on jäännöseliösumma. Jäännöseliösumma on jaettu vapausasteillaan, joka on havaintojen lukumäärä miinus kaksi. Taulukossa 11 esitetyn F -testisuureen arvon perusteella  $H_0$  -hypoteesi voidaan hylätä ja todeta, että mallissa on havaittavissa regressiota.

**Taulukko 12: Regressioanalyysin tulokset yhdelle selittävälle muuttujalle.**

	$a, b_1$	Keskivirhe	t -testi	p -arvo	95% luottamusvälin alaraja $\beta$ :lle	95% luottamusvälin yläraja $\beta$ :lle
Vakio	11.455	1.312	8.729	.000	8.846	14.064
Tytpel3	-0.098	.024	-4.025	.000	-.146	-.050

Taulukossa 12 on esitetty suoritettujen regressioanalyysien tulokset. Yhden selittävän muuttujan regressiomalliksi saatiin:

$$\text{Syntytyo} = 11.455 + (-0.098) \text{ Tytpel3p}.$$

Vakiotekijän  $\alpha$  estimaatiksi saatiin  $a = 11.455$  ja selittävän muuttujan *tytpel3p* regressiokertoimen  $\beta_1$  estimaatiksi  $b_1 = -0.098$ . Tämä merkitsee sitä, että aineistosta estimoituna selittävän muuttujan *tytpel3* yhden yksikön kasvu pienentää vastemuuttujan *syntytyo* arvoa keskimäärin 0.098 yksikköä.

Taulukossa ilmenevällä kaksisuuntaisella t -testisuureella testattiin sekä regressioyhtälön vakiotekijän  $\alpha$  että regressiokertoimen  $\beta_1$  arvojen nolasta poikkeavuutta. Nollahypoteesina oli, että sekä vakiotekijän että regressiokertoimen arvo on nolla. Kummankin kohdalla nollahypoteesi hylättiin ja todettiin, että niiden arvot poikkeavat nolasta tilastollisesti erittäin merkitsevästi.

Regressiokertoimen estimaatin  $b_1$  keskivirheeksi saatiin 0.024. Tämän perusteella voitiin laskea regressiokertoimen  $\beta$  95%:n luottamusväli. Luottamusväli laskettiin seuraavasti:

$$-0.098 \pm 0.024 \times 2.000 = [-0.146, -0.050],$$

eli regressiokertoimen estimaattiin on lisätty (vähennetty) sen keskivirheen ja taulukoidun Studentin t -jakauman arvon tulo. Studentin t -jakauman arvo, kun vaihtoehtohypoteesi on kaksisuuntainen ja vapausasteiden lukumäärä on 86, on 2.00. Vakiotekijän 95%:n luottamusväli on vastaavasti [8.846, 14.064].

Regressiomallin antama ennuste on mallin antama arvio muuttujan X arvoon liittyvän Y:n populaatiokeskiarvolle. Muuttujan Y yksittäisen arvon ennustaminen tehdään samalla tavalla, mutta ennusteen keskivirhe on suurempi.

Regressiomalleja tehtäessä jäännöstarkastelujen osuus korostuu. Jäännöksiä tutkimalla voidaan havaita, onko jotain systemaattista vaihtelua jäänyt selittämättä. Eräs estimoitun regressioyhtälön jäännöksiin olennaisesti liittyvä ilmiö on autokorrelaatio. Autokorrelaatiolla



tarkoitetaan tilannetta, jossa regressioyhtälön virhetermit ovat keskenään korreloituneita. Tämä on yleistä sekä aikasarja-aineistoille että maantieteellisesti läheisille havainnoille. Autokorreloituneiden jäännösten tapauksessa PNS -estimaatit säilyvät lineaarisina ja harhattomina, mutta ne eivät enää ole tehokkaita, sillä niillä ei ole minimivarianssia. Autokorrelaatio voi olla joko positiivista tai negatiivista sekä ensimmäistä, toista, kolmatta jne. astetta. Ensimmäisen asteen autokorrelaatio voidaan esittää seuraavasti:

$\varepsilon_t = \rho\varepsilon_{t-1} + \mu$ , missä  $\varepsilon$  on autokorreloitunut virhetermi ja  $\mu$  on normaalisti jakautunut satunnaismuuttuja, jonka keskiarvo on nolla ja varianssi  $\sigma^2$ . Jos  $\rho$  saa arvon nolla niin autokorrelaatiota ei esiinny ja malli palautuu takaisin klassiseksi lineaariseksi regressiomalliksi. Negatiivinen 1. asteen autokorrelaatio merkitsee sitä, että peräkkäiset virhetermit ovat yleensä vastakkaismerkkisiä. Positiivisesti autokorreloitunut virhetermi ”muistaa” menneisyytensä ja pyrkii säilymään positiivisena, jos se juuri edellä oli positiivinen.

Autokorrelaation olemassaolon testausta varten on kehitelty erilaisia testejä, joista yksi tunnetuimmista on ns. Durbin-Watson -testisuure. Sillä testataan, esiintyykö estimoidussa mallissa ensimmäiseen asteen autokorrelaatiota. Durbin-Watson -testisuureen arvo lasketaan kaavasta:

$$DW = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{j=1}^n \varepsilon_j^2}, \text{ missä } i = 2, 3, \dots, n \text{ ja } j = 1, 2, \dots, n.$$

DW -testisuureen arvolle on taulukoitu kriittisiä ylä- ja ala-arvoja, joiden välisellä alueella testi ei päde. Yhden selittävän muuttujan (tyypel3) ja 88 havainnon tapauksessa testi ei päde DW -testin arvojen sijoituessa välille [1.635, 1.679]. DW -testisuureen arvoksi saatiin 1.963, minkä perusteella voidaan päätellä, että mallissa ei esiinny ensimmäisen asteen autokorrelaatiota.

Eräs regressioanalyysin perusolettamuksista on se, että virhetermi  $\varepsilon$  on luonteeltaan satunnaismuuttuja, jonka tn -jakauma on sama kaikkialla näytteessä. Erityisesti sen on oltava sama käytettäessä mitä hyvänsä selittävän muuttujan arvoa vastemuuttujan laskemiseen.

$$\text{var}(\varepsilon_i) = E\left[\left(\varepsilon_i - E(\varepsilon_i)\right)^2\right] = E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = \text{vakio}$$

Edellä on esitetty virhetermin varianssia koskeva homoskedastisuusoletus. Jollei se pidä paikkaansa, niin virhetermit ovat heteroskedastisia. Homoskedastisuudella tarkoitetaan, että virhetermi  $\varepsilon_i$  vaihtelee keskiarvonsa nollan ympärillä, eikä se riipu selittävän muuttujan arvoista. Eräs tärkeimmistä heteroskedastisuuden syistä on se, että mallista poisjätetyt muuttujat kasvavat silloin kuin mallissa jo mukana olevat selittävät muuttujat kasvavat. Tästä syystä virhetermin varianssi kasvaa, kun mukana olevat selittäjät kasvavat.

Jos kaikki muut klassisen lineaarisen regressioanalyysin oletukset paitsi homoskedastisuusoletus ovat voimassa, niin siitä seuraa, että PNS -estimaatit ovat edelleen lineaarisia ja harhattomia, mutta niillä ei ole minimivarianssia. Ne eivät ole enää tehokkaita, eli ne eivät ole BLUE (best linear unbiased estimator).

Käytettävissä ei ole mitään varmaa keinoa heteroskedastisuuden havaitsemiseksi. On vain olemassa joukko diagnosoivavälineitä, joista saattaa olla apua. Tarkasteltavan aineiston luonne antaa viitteitä heteroskedastisuus -ongelman olemassaolosta. Heteroskedastisuuden on todettu olevan pikemminkin sääntö kuin poikkeus sellaisessa poikkileikkausaineistossa, johon kuuluu heterogeenisiä yksiköitä, kuten mm. kotitaloustiedustelun aineistot.

Eräs heteroskedastisuusongelman ilmenemistä tutkiva testi on ns. Parkin testi. Testissä estimoidaan yhtälö:

$$\ln \varepsilon_i^2 = \alpha + \beta \ln x_i + v_i,$$

missä  $x$  viittaa siihen selittävään muuttujaan, jonka suhteen mallin jäännösten  $\varepsilon_i$  epäillään olevan heteroskedastisia. Parkin testi suoritetaan seuraavasti:

1. Suoritetaan alkuperäinen regressioanalyysi mahdollisesta heteroskedastisuusongelmasta välittämättä.
2. Otetaan tästä regressiyhtälöstä jäännöstermit  $\varepsilon_i$ , neliöidään ja logaritmoidaan ne.
3. Lasketaan Parkin testin regressioanalyysi käyttämällä vuoron perään jokaista alkuperäisen mallin selittävän muuttujan logaritmia.

4. Testataan  $H_0: \beta_1 = 0$  eli että aineistossa ei ole heteroskedastisuutta. Jos  $H_0$  hylätään ja  $\ln \varepsilon_i^2$ :n ja  $\ln x_i$ :n välillä on merkitsevä riippuvuus, niin todetaan heteroskedastisuuden olemassaolo mallissa.

Aiemmin estimoidulle regressiomallille suoritettiin Parkin testi virhetermin homoskedastisuusoletuksen testaamiseksi. Mallin jäännökset neliöitiin ja logaritmoitiin ja suoritettiin regressioanalyysi käyttäen selittävänä muuttujana muuttujan *tytpel3p* logaritmia. Estimoitava yhtälö oli

$$\ln \varepsilon_i^2 = \alpha + \beta_1 \ln tytpel3p + v_i.$$

Regressioanalyysin tuloksiksi saatiin

$$\ln \varepsilon_i^2 = 0.00037 - 0.026 \ln tytpel3p.$$

$$kv = (6.594) \quad (1.663)$$

$$t = (.000) \quad (-.016) \quad R^2 = .000$$

Parkin testin tulokset osoittavat, että regressikertoimen estimaatti  $\beta$  ei ole tilastollisesti merkitsevä ja näinollen nollahypoteesi hyväksytään ja todetaan, että mallissa ei ole heteroskedastisuusongelmaa.

Viimeinen testaamatta oleva klassista yhden selittävän muuttujan regressioanalyysia koskeva alkuoletus on oletus virhetermien normaalijakautuneisuudesta. Tätä oletusta testattiin jo aiemmin esitellyllä Kolmogorov-Smirnov -testillä. 95%:n luottamusvälillä saatiin testin p - arvoksi .206 eli jäännökset noudattavat normaalijakaumaa.

Seuraavaksi regressiomallin rakentamista jatkettiin lisäämällä siihen toinen selittävä muuttuja eli *yrtiheys*. Selittävällä muuttujalla *yrtiheys* oli toiseksi korkein korrelaatio vaste muuttujan *syntytyo* kanssa. Alla olevassa taulukossa 13 on esitetty varianssitaulu ja sen tulokset kahden selittävän muuttujan (*tytpel3p*, *yrtiheys*) tapauksessa.

**Taulukko 13: Varianssitaulu kahdelle selittävälle muuttujalle.**

Vaihtelun lähde	Vapausasteet	Neliösumma	Keskineliö	F -testisuure	P -arvo
Regressio	2	97.421	48.710	12.866	.000
Jäännös	85	321.801	3.786		
Kokonaisvaihtelu	87	419.222			

Malliin valittujen selittävien muuttujien avulla saatiin regressioneliösumman arvoksi 97.421 ja tämän avulla laskettu mallin selitysaste  $R^2$  prosentteina on 23.2%. Uusi selittävä muuttuja *yrtiheys* kasvatti siis mallin selitystasetta 7.4 prosenttia yhden selittävän muuttujan regressiomalliin verrattuna. F -testisuureen arvon perusteella voidaan nollahypoteesi hylätä ja todetaan, että mallissa on regressiota tilastollisesti erittäin merkitsevästi.

**Taulukko 14: Regressioanalyysin tulokset kahdelle selittävälle muuttujalle.**

	a, b <sub>1</sub> , b <sub>2</sub>	Keskivirhe	t -testi	p -arvo	95%:n luot- tamusvälin alaraja β <sub>i</sub> :lle	95%:n luot- tamusvälin yläraja β <sub>i</sub> :lle
Vakio	7.000	2.004	3.493	.001	3.016	10.984
Tytpel3p	-0.088	.024	-3.743	.000	-.135	-.041
Yrtiheys	95.253	33.299	2.861	.005	29.046	161.459

Taulukossa 14 on esitetty suoritetun kahden selittävän muuttujan regressioanalyysin tulokset. Regressiomalliksi saatiin:

$$\text{Syntytyo} = 7.000 + (-0.088) \text{ Tytpel3p} + 95.253 \text{ Yrtiheys}.$$

Vakiotermin  $\alpha$  estimaatiksi saatiin  $a = 7.000$  ja regressiokertoimien  $\beta_1$  ja  $\beta_2$  estimaateiksi tulivat  $b_1 = -0.088$  ja  $b_2 = 95.253$ . Sekä vakiotermin estimaatin että regressiokertoimien estimaattien arvot poikkesivat tilastollisesti merkitsevästi nolasta niille suoritetun kaksisuuntaisen t -testin perusteella. Molempien selittävien muuttujien pitämien mukana mallissa oli siis mielekästä. Taulukosta 14 löytyvät myös sekä vakiotermin että regressiokertoimien keskivirheiden avulla lasketut 95%:n luottamusvälit.

Jäännöstarkasteluissa Durbin-Watson -testi antoi tulokseksi 1.835, minkä perusteella voidaan todeta, että mallissa ei esiinny ensimmäisen asteen autokorrelaatiota. DW -testi ei kahden selittävän muuttujan ja 88 havainnon tapauksessa ole voimassa välillä [1.600, 1.696] merkitsevyytasolla  $\alpha = 0.05$ .

Mallin virhetermin homoskedastisuutta testattiin jo aiemmin esitellyllä Parkin testillä. Esitimoitava yhtälö oli:

$$\ln \varepsilon_i^2 = \alpha + \beta_2 \ln \text{yrtheys} + v_i.$$

Regressioanalyysin tuloksiksi saatiin:

$$\ln \varepsilon_i^2 = 8.375 + 2.712 \ln \text{yrtheys}.$$

$$t = \quad (1.505) \quad (1.558) \qquad R^2 = .027$$

Regressiokertoimen estimaatille suoritetun kaksisuuntaisen t -testin arvo osoittaa, ettei regressiokertoimen estimaatti ole tilastollisesti merkitsevä. Mallissa ei siis ole havaittavissa heteroskedastisuusongelmaa selittävän muuttujan *yrtheys* suhteen.

Virhetermien normaalijakautuneisuutta tutkittiin jälleen Kolmogorov-Smirnov -testillä. Testin p -arvoksi saatiin .716 eli jäännökset noudattivat normaalijakaumaa.

PNS -menetelmän käytön edellytys on, etteivät selittävät muuttujat ole keskenään lineaarisesti täysin korreloituneita eli multikollineaarisia. Jos näin olisi, regressiokertoimia ei voi-

taisi estimoida lainkaan. Multikollinearisuus on ilmiö, joka kuuluu taloudellisten muuttujien luonteeseen. Ei myöskään ole olemassa selkeää mittaa sille, milloin mahdollisesti vallitseva multikollinearisuus vakavasti heikentää parametriestimaatteja.

Multikollinearisuus voi syntyä siten, että taloudellisten muuttujien takana on yhteisiä tekijöitä, jotka aiheuttavat saman suuntaisia muutoksia tarkasteltavissa muuttujissa. Kasvutekijät ja trendinomaiset muutokset ovat huomattavia multikollinearisuuden aiheuttajia. Multikollinearisuus onkin nimenomaan aikasarja-aineistoissa esiintyvä ongelma, mutta sitä saattaa esiintyä myös poikkileikkausaineistoissa.

#### Multikollinearisuuden teoreettiset seuraukset

PNS -estimaattorit ovat edelleen BLUE ja klassisen lineaarisen regressiomallin oletukset pitävät paikkansa. Tämä ominaisuus säilyy, vaikka joku yksittäinen regressiokerroin ei olisikaan tilastollisesti merkitsevä. Multikollinearisuudesta täytyy kuitenkin olla huolissaan, koska:

1. On muistettava, että harhattomuus on ominaisuus, joka liittyy toistettuun näytteen ottoon. Harvoin on mahdollista ottaa useita näytteitä ja niiden kautta varmistaa hyvä estimaatti.
2. Voimakaskaan multikollinearisuus ei tuhoa PNS -estimaattorin minimivarianssiominaisuutta. Tämä ei kuitenkaan merkitse, että jostakin tietystä näytteestä laskettu PNS -varianssi olisi pieni.

Multikollinearisuus on kaikkea näytteeseen liittyvä ilmiö. Vaikkeivat selittävät muuttujat olisikaan populaatiossa toisistaan lineaarisesti riippuvia, niin ne voivat olla sitä tietystä näytteessä. (Gujarati 1995, 325-326)

#### Multikollinearisuuden käytännön seuraamukset (Gujarati 1995, 237)

1. Estimaateille suuri varianssi ja standardipoikkeama.
2. Luottamusvälit laajempia.
3. Yhden tai useamman estimaatin t -testin arvot ovat tilastollisesti merkityksettömiä.

4. Mallille korkea selitysaste  $R^2$ , mutta t -arvojen merkitsevyys alhainen.
5. PNS -estimaattorit ja niiden keskivirheiden arvot ovat herkkiä pienille aineiston muutoksille.
6. Regressiokertoimet voivat saada vääriä etumerkkejä.
7. Yksittäisten selittävien muuttujien kontribuution arvioiminen on vaikeaa.

Selittävien muuttujien *tytpel3p* ja *yrtiheys* tapauksessa mallissa ei ole havaittavissa merkkejä vahvasta multikollineaarisuudesta. F -testi hyväksyi regression ja t -testikin hyväksyi molemmat selittävät muuttujat. Selittäjien välisen korrelaatiokertoimen arvo oli -0.141, mikä ei viittaa vahvaan multikollineaarisuuden olemassaoloon. Malliin ensimmäisenä sijoitetun selittävän muuttujan *tytpel3p* regressiokertoimen estimaatin arvo ei muuttunut merkittävästi ja näin myös sen etumerkki pysyi negatiivisena, joten nämä havainnot eivät viittaa multikollineaarisuuden olemassaoloon.

Multikollineaarisuuden esiintymistä voidaan tarkastella myös niin sanotulla VIF -luvulla (variance inflator factor). VIF -luku lasketaan seuraavasti:

$$VIF_i = \frac{1}{(1 - R_{di}^2)},$$

missä  $R_{di}^2$  on nk.  $R^2$  -delete, joka on saatu regressoimalla selittäjä  $i$  kaikilla muilla selittävillä muuttujilla. Jos  $R_{di}^2$  on nolla saa  $VIF_i$  arvon 1 ja jos  $R_{di}^2$  on 0.99, niin  $VIF_i$  saa arvon 100. Mitä suuremman arvon VIF -luku saa, sitä ongelmallisempaan voidaan selittävää muuttujaan pitää multikollineaarisuuden kannalta tarkasteltuna. Peukalosääntönä voidaan haitallisen multikollineaarisuuden raja-arvona pitää selittävän muuttujan VIF -luvun arvoa 10, jolloin  $R_{di}^2 = 0.9$  (Gujarati 1995, 339). Kahden selittävän muuttujan tapauksessa on VIF -luku sama molemmille ja muuttujille *tytpel3p* ja *yrtiheys* on  $VIF = 1.020$  eli multikollineaarisuus ei ole ongelmana regressiomallissa.

Viimeinen regressiomalliin lisättävä selittävä muuttuja oli *valtsmp*, eli Maaseudun puoleen valtuustopaikkojen osuus kaikista seutukunnan valtuustopaikoista. Taulukossa 15 on esitetty varianssitaulu kolmen selittävän muuttujan tapauksessa.

**Taulukko 15: Varianssitaulu kolmelle selittäväälle muuttujalle.**

Vaihtelun lähde	Vapausasteet	Neliösumma	Keskineliö	F -testisuure	p -arvo
Regressio	3	127.846	42.615	12.103	.000
Jäännös	82	288.728	3.521		
Kokonaisvaihtelu	85	416.574			

Selittävillä muuttujilla *tytpel3p*, *yrtiheys* ja *valtsmp* saatiin regressioneliösumman arvoksi 127.846 ja mallin selitysaste  $R^2$  oli prosentteina 30.7. Malliin uutena tuotu selittävä muuttuja *valtsmp* kasvatti mallin selitysastetta 7.5 prosenttia kahden selittävän muuttujan malliin verrattuna. F -testisuureen arvon perusteella voidaan todeta, että mallissa on regressiota tilastollisesti erittäin merkitsevästi.

Taulukossa 16 on esitetty kolmen selittävän muuttujan regressioanalyysin tulokset.

**Taulukko 16: Regressioanalyysin tulokset kolmelle selittäväälle muuttujalle.**

	a, b <sub>1</sub> , b <sub>2</sub> , b <sub>3</sub>	Keskivirhe	t -testi	p -arvo	95%:n luottamusväli β <sub>i</sub> :lle	95%:n luottamusväli β <sub>i</sub> :lle
Vakio	4.100	2.257	1.816	.073	-.390	8.591
Tytpel3p	-0.075	.024	-3.072	.003	-.123	-.026
Yrtiheys	139.385	37.225	3.744	.000	65.333	213.437
Valtsmp	14.293	7.161	1.996	.049	.047	28.539



Regressiomalliksi saatiin:

$$\text{Syntytyo} = 4.100 + (-0.088)\text{Tytpel3p} + 139.385\text{Yrtiheys} + 14.293\text{Valtsmp}.$$

Vakiotermin  $\alpha$  estimaatiksi saatiin  $a = 4.100$  ja regressiokertoimien  $\beta_1$ ,  $\beta_2$  ja  $\beta_3$  estimaateiksi tulivat  $b_1 = -0.075$ ,  $b_2 = 139.385$  ja  $b_3 = 14.293$ . Sekä vakiotermin että regressiokertoimien estimaattien arvot poikkesivat tilastollisesti merkitsevästi nolasta kaksisuuntaisen t -testin perusteella merkitsevyystasolla  $\alpha = 0.05$ . Sekä vakiotermin että regressiokertoimien 95% luottamusvälit on esitetty taulukossa 16.

Durbin-Watson -testin tulokseksi saatiin 1.859, minkä perusteella voidaan päätellä, että mallissa ei esiinny ensimmäisen asteen autokorrelaatiota. Kolmen selittävän muuttujan ja 88 havainnon tapauksessa DW -testi ei ole voimassa välillä [1.575, 1.721] merkitsevyystasolla  $\alpha = 0.05$ .

Kolmen selittävän muuttujan mallin virhetermin homoskedastisuutta testattiin jälleen Parkin testillä. Estimoitava yhtälö oli:

$$\ln \varepsilon_i^2 = \alpha + \beta_3 \ln \text{valtsmp} + v_i.$$

Regressioanalyysin tuloksiksi saatiin:

$$\ln \varepsilon_i^2 = 0.649 + 0.315 \ln \text{valtsmp}.$$

$$t = (0.423) (.757)$$

$$R^2 = .008$$

Regressiokertoimen estimaatin kaksisuuntaisen t -testin arvo osoittaa, ettei regressiokertoimen estimaatti ole tilastollisesti merkitsevä. Mallissa ei ole siis havaittavissa heteroskedastisuusongelmaa selittävän muuttujan *valtsmp* suhteen.

Virhetermien normaalijakautuneisuutta tutkittiin jälleen Kolmogorov-Smirnov -testillä. Testin p -arvo oli .368, eli jäännökset noudattavat normaalijakaumaa.

Kolmen selittävän muuttujan muuttujan regressiomallissa ei ollut havaittavissa merkkejä vahvan multikollineaarisuuden olemassaolosta. F -testi hyväksyi regression ja sekä vakio-  
kijän että regressiokertoimien estimaatit olivat t -testin perusteella tilastollisesti merkitseviä. Selittävien muuttujien välisten korrelaatiokertoimien arvot (taulukko 10) eivät viitanneet vahvaan multikollineaarisuuteen. Regressiokertoimien estimaattien arvot eivät muuttuneet merkittävästi, mikä myös tukee päätelmää vahvan multikollineaarisuuden puuttumisesta mallista.

Selittävien muuttujien multikollineaarisuutta tutkittiin jälleen myös jo aiemmin esitellyllä VIF -luvulla. VIF -luvut eri selittäjille on esitetty taulukossa 17.

**Taulukko 17: Selittävien muuttujien VIF -luvut.**

Muuttuja	VIF -luku
Tytpel3p	1.082
Yrtiheys	1.052
Valtsmp	1.032

Selittävien muuttujien VIF -lukujen arvoista voidaan päätellä, että regressiomallissa ei ole multikollineaarisuus ongelmaa.

## 5. YHTEENVETO

Tämän pro gradu -tutkielman lähtökohtana oli monissa sekä koti- että ulkomaisissa tutkimuksissa tehty havainto, että uudet yritykset eivät synny maantieteellisesti tasaisesti valtakunnan rajojen sisällä. Useissa maissa on löydetty alueita, joissa uusia yrityksiä syntyy huomattavasti enemmän tai vähemmän valtakunnan keskiarvoon verrattuna. Yleisimmät tavat mitata uusien yritysten syntyä ovat ns. yrityskantamalli ja työmarkkinamalli. Edellisessä alueelle syntyneet uudet yritykset suhteutetaan alueella jo toimiviin yrityksiin ja jälkimmäisessä taas alueen väestömäärään.

Tutkielman tavoitteena oli ensinnäkin selvittää, löytyykö Suomesta vastaavia maantieteellisiä eroja uusien yritysten perustamisasteissa kuin ulkomaisissa tutkimuksissa on havaittu. Tämän jälkeen tavoitteena oli löytää perustamisasteiden maantieteellisten erojen taustalla olevia tekijöitä. Selvitettäessä eroja uusien yritysten perustamisasteissa rajoituttiin tutkimaan niitä vain teollisuusyritysten osalta. Uusien teollisuusyritysten perustamisasteiden välisiä eroja tarkasteltiin Suomen 88 seutukunnan välillä ajanjaksolla 1.7.1994 - 31.12.1995.

Uusien yritysten syntyä käsitteleviä teorioita valittiin tutkielmaan tarkasteltavaksi neljä kappaletta. Ensimmäinen niistä oli ns. työmarkkinamalli, joka painottaa yksilön henkilökohdainten ominaisuuksien merkitystä oman yrityksen perustamispäätökselle. Jos nämä ominaisuudet vaihtelevat populaatioiden välillä eri alueilla, silloin myös aktiivisuuden yrittäjyyteen voidaan olettaa vaihtelevan alueittain. Toinen uusien yritysten syntyä käsittelevä teoria oli rakenne-käytös-suorituskyky -paradigma, joka on perinteisin tapa hahmottaa uusien yritysten syntymekanismeja. Rakenne-käytös-suorituskyky -paradigma painottaa toimialalla jo toimivien yritysten hinnoittelu- ja tuotantopäätösten merkitystä sinne pyrkiville uusille yrityksille. Kolmantena tarkasteltavana ollut uusien yritysten syntyä käsittelevä teoria oli ns. ympäristömalli. Ympäristömalli korostaa paikallisen toimintaympäristön merkitystä uusien yritysten synnylle. Tärkeimmäksi paikallisen toimintaympäristön ominaisuudeksi uusien yritysten synnylle ympäristömallissa todetaan alueen pienyritysvaltaisuus. Neljäntenä tutkielmassa tarkasteltavana ollut teoria oli lokalisaatio eli saman toimialan tuotantotoiminnan

taipumus keskittyä alueellisesti. Lokalisoitumiseen vaikuttavat tekijät ovat Marshallin (1920) mukaan yhteiset työmarkkinat, yhteiset välituotemarkkinat ja teknologiset vuodot.

Tutkielmassa käsiteltiin tarkemmin kahta ulkomaista tutkimusta, jotka molemmat tarkastelivat uusien yritysten syntyä. Ensimmäinen tutkimuksista oli norjalaisen Spillingin (1996) tutkimus uusien yritysten perustamisasteiden välisestä alueellisista eroista Norjassa. Spilling jakoi Norjan 104:ään itsenäiseen työmarkkina-alueeseen, joista jokaisella laskettiin uusien yritysten perustamisaste ajanjaksolla 1988-92. Eroja eri alueiden välillä pyrittiin selittämään lineaarisen regressiomallin avulla, joka suoritettiin sekä koko talouden ollessa tarkastelun piirissä että talouden eri sektoreille erikseen. Tärkeimmäksi selittäväksi tekijäksi teollisuuden uusille yrityksille saatiin alueen pienyritysvaltaisuus.

Toinen tarkastelluista tutkimuksista oli saksalaisten Audretschin ja Fritschin (1991) tekemä tutkimus uusien yritysten perustamisasteiden välisestä alueellisista eroista Länsi-Saksassa vuosina 1986-89 kerätyllä aineistolla. Tutkimuksessa testattiin Krugmanin (1991) lokalisoitumista käsittelevän teorian yhteensopivuutta edellä mainitun tarkasteluajanjakson aineistoon. Eroja alueiden välillä tarkasteltiin lineaarisen regressiomallin avulla, niin että regressiot suoritettiin koko taloudelle, teollisuudelle ja palveluille. Teollisuuden kohdalla työmarkkinamallin mukaan laskettuna tärkeimmiksi uusien yritysten syntyä selittäviksi tekijöiksi löydettiin väestönkasvu alueella ja toimipaikan keskikoko.

Tutkielman empiirisessä osuudessa pyrittiin selvittämään, miksi Suomen 88 seutukunnassa on erilaisia uusien teollisuusyritysten perustamisasteita työmarkkinamallin mukaan laskettuna. Jokaiselle seutukunnalle laskettiin uusien teollisuusyritysten perustamisaste tarkasteluajanjaksolla 1.7.1994 - 31.12.1995 ja eroja eri seutukuntien välillä selvitettiin lineaarisella regressioanalyysillä. Regressiomallin vastemuuttujana oli *syntytyo*, eli uusien teollisuusyritysten perustamisaste seutukunnittain työmarkkinamallin mukaan laskettuna. Selittävät muuttujat valittiin malliin käytettävissä olleen Kuntafakta -tilastoaineiston pohjalta. Vastemuuttujan ja selittävien muuttujien välisten korrelaatiotarkasteluiden perusteella valittiin mukaan mallin jatkokehittelyyn ne selittävät muuttujat, jotka korreloivat tilastollisesti merkittävästi vastemuuttujan kanssa. Seuraavaksi tarkasteltiin selittävien muuttujien keskinäisiä korrelaatioita. Regressiomallin selittäviksi muuttujiksi valittiin lopulta muuttujat *tytpe13p* (palvelun osuus työpaikoista), *yrtiheys* (yrityksiä per väestö) ja *valtsmp* (maaseudun puolueen osuus valtuustopaikoista). Regressiomalli rakennettiin lisäämällä siihen yksi ker-

rallaan selittäviä muuttujia aloittaen muuttujasta *tytpel3p*, joka korreloi voimakkaimmin vastemuuttujan *syntyö* kanssa. Seuraavaksi malliin lisättiin selittävät muuttujat *yrtiheys* ja *valtsmp* ja lopulliseksi lineaariseksi regressiomalliksi saatiin:

$$\text{Syntytyo} = 4.100 + (-0.088)\text{Tytpel3p} + 139.385\text{Yrtiheys} + 14.293\text{Valtsmp}.$$

Regressiomallin selitysasteeksi saatiin  $R^2 = 0.307$  eli selittävillä muuttujilla saatiin selitettyä 30.7% vastemuuttujan havaitusta vaihtelusta. Selitysaste jäi siis varsin pieneksi, eikä saadun regressiomallin perusteella voida näinollen kovin tarkasti päätellä, mistä johtuvat seutukuntien väliset erot uusien teollisuusyritysten perustamisasteissa.

Regressiomallin selittävästä muuttujista *tytpel3p* eli palvelualan työpaikkojen osuus kaikista työpaikoista pienensi uusien teollisuusyritysten syntymahdollisuuksia seutukunnassa. Selityksenä tälle voi se, että mitä enemmän seutukunnassa on palvelualan työpaikkoja, niin sitä teollistuneempi seutukunta on ja uusia teollisuusyrityksiä syntyy näinollen vähemmän kuin seutukuntiin, joissa on vähän palvelualan työpaikkoja ja jotka ovat näin vähemmän teollistuneita.

Selittävän muuttujan *yrtiheys* suuri arvo eli seutukunnan yritysten määrä sen väkilukuun suhteutettuna vaikutti positiivisesti uusien teollisuusyritysten syntyyn. *Yrtiheys* kertoo myös seutukunnan pienyritysvaltaisuudesta, sillä sen korkea arvo viittaa siihen että seudulla on paljon pienyrityksiä. Myös Spilling (1996) sekä Audretsch ja Fritsch (1994) löysivät yhteyden alueen pienyritysvaltaisuuden ja korkean uusien teollisuusyritysten perustamisasteen väliltä. Tämä tulos on konsistentti luvussa 2.2 esitellyn ympäristömallin olettamusten kanssa. Pienyritysvaltainen toimintaympäristö nähdään siinä perusedellytykseksi uusien yritysten korkeille perustamisasteille. Pienet yritykset ovat ympäristömallin mukaan hyviä hautoja tuleville yrittäjille.

Kolmannen regressiomallin selittävän muuttujan *valtsmp* suuri arvo eli maaseudun puolueen suuri osuus seutukunnan valtuustopaikoista vaikutti positiivisesti uusien teollisuusyritysten perustamisasteisiin. Tälle tulokselle on vaikea esittää mitään täsmällistä syytä. Voidaan tietysti ajatella, että alueilla, joilla puolueen kannatus on suuri asuu myös tavallista yritteliäämpää väestöä. Sitä vastoin olisi varsin rohkeata väittää, että puolueen vahva läsnä-

olo seutukunnassa vaikuttaisi alueen asukkaisiin niin, että heidän yrittäjähenkisyytensä kasvaisi.

Tässä pro gradu -tutkielmassa tarkasteltiin, mitkä tekijät vaikuttavat uusien teollisuusyritysten perustamisasteiden välisiin eroihin seutukuntatasolla tarkasteltuna. Tutkielmassa ei kiinnitetty lainkaan huomiota siihen, mitkä syyt sitten vaikuttavat ilmiön kääntöpuoleen eli teollisuusyritysten tai vain yleisemmin yritysten kuolemiin. Yritysten kuolemien tutkiminen auttaisi kenties ymmärtämään myös yritysten syntymekanismeja hieman paremmin.

## LÄHDEKIRJALLISUUS

Audretsch, D.B., Fritsch, M. 1994. The geography of firm births in Germany. *Regional Studies* 28: 359-365.

Birch, D.L. 1987. *Job creation in America*. New York: Free Press

Brockhaus, R. (1982). *The Psychology of entrepreneurs*. Teoksessa Kent, C.D. Sexton & K. Vesper (toim.) *Encyclopedia of Entrepreneurship*. New Jersey: Englewood Cliffs.

Brusco, S. 1989. *Small firms and industrial districts: The experience of Italy*. Teoksessa *New Firms and Regional Development in Europe*, D. Keeble, E. Wever (toim.). Lontoo: Croom Helm.

Chinitz, B. 1961. Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review* 51: 289-89.

Davidsson, P., Lindmark, L., Olofsson, C. 1994. New firm formation and regional development in Sweden. *Regional Studies* 28: 395-410.

Eisinger, P. 1991. The state of venture capitalism. *Economic Development Quarterly* 5: 64-76

Fritsch, M. 1992. Regional differences in new firm formation: evidence from West Germany. *Regional Studies* 26: 233-241.

Gibb, A., Ritchie, J. 1982. Understanding the process of starting small businesses. *International Small Business Journal* 1(1): 26-45

Gujarati, D.N. 1995. *Basic econometrics*. New York: McGraw-Hill

Keeble, D. 1990. Small firms, new firms and uneven regional development in United Kingdom. *Area* 22: 234-245.

Keeble, D. , Walker, S. 1994. New firms, small firms and dead firms: spatial patterns and determinants in the United Kingdom. *Regional Studies* 28: 411-427.

Krugman, P. 1991. *Geography and trade*. Cambridge, MA: MIT Press

Malecki, E.J. 1994. Entrepreneurship in regional and local development. *International Regional Science Review* 16: 119-253.

Marshall, A. 1920. *Principles of Economics*. Lontoo: Macmillan

- Mueller, D.C. 1992. Entry, Exit and the Competitive Process in P. Geroski and J. Schwalback (eds) *Entry and Market Contestability: An International Comparison*. Oxford:Blackwell
- Niittykangas, H., Tervo, H. Entrepreneurship and local milieu. Julkaisussa Eskelinen, H. (toim.) *Regional specialization and local environment - learning and competitiveness*. Nordrefo 1997:3 Tanska
- Porter, M. 1981. The contributions of industrial organization to strategic management. *Academy of Management Review* 6(4): 609-619.
- Ranta, E., Rita, H., Kouki J. 1997. *Biometria: Tilastotiedettä ekologeille*. Helsinki: Yliopistopaino.
- Reynolds, P., Storey, D.J., Westhead, P. 1994. Cross-national comparisons of the variation in new firm formation rates. *Regional Studies* 21: 13-23.
- Spilling, O.R. 1996. *Regional variation of new firm formation: the Norwegian case*
- Tervo, H., Niittykangas, H. 1994. The impact of unemployment on new firm formation in Finland. *International Small Business Journal* 13:1, 38-53
- Storey, D.J. 1995. *Understanding the small business sector*. New York: Routledge.
- Sweeney, G. P. 1987. *Innovation, entrepreneurs, and regional development*. New York: St. Martin's Press
- Westhead, P., Moyes, T. 1992. Reflections on Thatcher's Britain: evidence from new production firms registrations 1980-88. *Entrepreneurship & Regional development* 4:21-56.



**LIITE 1 . Uusien teollisuusyritysten perustamisasteet seutukunnittain.**

Seutukunta	per.aste	Seutukunta	per.aste	Seutukunta	per.aste
Suupohjan	13.98	Kuusiokuntien	7.08	Pielisen Karjalan	5.10
Eteläisten seinänaapurien	12.91	Sisä-Savon	6.82	Jyväskylän	5.00
Kaustisen	12.04	Kouvolan	6.81	Porvoon	4.99
Kyrönmaan	11.64	Rauman	6.69	Kuopion	4.95
Järvisseudun	11.19	Saarijärven	6.65	Koillismaan	4.92
Siikalatvan	9.85	Kärkikuntien	6.44	Koillis-Lapin	4.85
Kaakkois-Satakunnan	9.38	Porin	6.25	Varkauden	4.85
Torniolaakson	9.32	Pieksämäen	6.17	Maarianhaminan	4.68
Härmänmaan	9.28	Posio-Ranuan	6.15	Savonlinnan	4.60
Viitasaaren	9.07	Lahden	6.14	Joensuun	4.57
Salon	8.22	Riihimäen	5.95	Ilomantsin	4.47
Kokkolan	8.18	Kaakkois-Pirkanmaan	5.94	Loviisan	4.45
Sydösterbotten kustregionen	8.12	Savitaipaleen	5.94	Lappeenrannan	4.16
Keski-Karjalan	8.10	Luoteispirkanmaan	5.87	Rovaniemen	4.09
Loimaan	8.08	Koillis-Savon	5.85	Hämeenlinnan	4.08
Pohjois-Lapin	7.94	Ylä-Savon	5.84	Iin	4.05
Etelä-Pirkanmaan	7.90	Orimattilan	5.83	Raahen	4.03
Kaakkoisen Keski-Suomen	7.81	Heinolan	5.75	Forssan	3.99
Tampereen	7.79	Pohjois-Pirkanmaan	5.64	Imatran	3.98
Sysmän	7.76	Keuruun	5.59	Kehys-Kainuun	3.95
Outokummun	7.75	Vaasan	5.58	Oulun	3.62
Nivala-Haapajärven	7.49	Turun	5.50	Tunturi-Lapin	3.38
Pohjois-Satakunnan	7.47	Pietarsaaren	5.48	Kotkan-Haminan	3.32
Äänekosken	7.34	Föglö	5.48	Kajaanin	3.21
Juvan	7.30	Turunmaan	5.36	Joroisten	3.06
Lohjan	7.30	Koillis-Pirkanmaan	5.27	Itä-Pirkanmaan	3.00
Lounais-Pirkanmaan	7.26	Kemi-Tornion	5.19	Lakeuden	2.80
Jämsän	7.26	Tammisaaren	5.13	Ylivieskan	2.46
Vakka-Suomen	7.21	Helsingin	5.13		
Pohjoisten seinänaapurien	7.11	Mikkelin	5.12		