

# Yritysten tilinpäätöstietojen toimipaikkaistami- sen ongelma

Yleistetyn regressioestimoinnin sovellus rekisteritutki-  
muksessa

Elina Nykyri

Tilastotieteen pro gradu -tutkielma  
31. lokakuuta 2003

Jyväskylän yliopisto

Matematiikan ja tilastotieteen laitos

Tilastotoimen menetelmien maisteriohjelma

## ESIPUHE

Kädessäsi oleva tutkimus on tilastotieteen pro gradu -tutkielma. Tein sen Jyväskylän yliopiston tilastotieteen ja matematiikan laitoksen tilastotoimen menetelmien maisteriohjelmassa. Siihen liittyy pro gradu -työn tekeminen yhteistyössä jonkun yrityksen kanssa. Teinkin graduni osin Tilastokeskuksen rahoituksella ja ohjauksessa. Tutkimusmateriaalina tarvitsin paljon tietoja Tilastokeskuksessa tuotetuista yrityksiin liittyvistä rekistereistä ja tietokannoista (yritys- ja toimipaikkarekisteri sekä rakennetilastojen kyselykanta) sekä yritysten tilinpäätöstiedoista. Suurena apuna ovat olleet keskustelut Tilastokeskuksen henkilökuntaan kuuluvien kanssa.

Haluan kiittää kaikkia pro gradu -työni tekoon osallistuneita. Ohjaajinani toimivat professori Risto Lehtonen Jyväskylän yliopistosta ja kehittämisspäällikkö, yhteiskuntatieteiden lisensiaatti Kari Djerf Tilastokeskuksesta. Tietoa yritysten toiminnasta ja käyttämistäni Tilastokeskuksen tuottamista yritys- ja toimipaikkatiedoista ovat antaneet yliaktuaari Marjatta Laine, tutkimuspäällikkö Mika Maliranta, tilastonlaatija Pirkko Nurmela, kehittämisspäällikkö Matti Parkkinen, tilastopäällikkö Heikki Pihlaja, tilastopäällikkö Leena Storgårds, yliaktuaari Ville Tolkki, yliaktuaari Janne Virtanen ja tilastopäällikkö Ritva Wuoristo. Yliaktuaari Janne Ikäheimonen auttoi SAS-ohjelmistoon liittyvissä erityiskysymyksissä ja antoi tietoa rakennetilastojen kyselykannasta. Kiitän teitä kaikkia yhdessä ja erikseen.

Lisäksi kiitän kaikkia, jotka ovat tutkimuksen eri vaiheissa lukeneet graduani. Heitä ovat Tilastokeskuksesta yliaktuaarit Aku Alanen ja Tero Karttunen sekä Tilastokeskuksen ulkopuolelta Kari Molkänen, yhteiskuntatieteiden tohtori Terttu Nupponen ja filosofian ylioppilas Tiina Taponen. Myös suuri osa jo aiemmin mainituista luki jossain vaiheessa graduani. Olette kaikki antaneet paljon hyödyllisiä kommentteja. Kiitän myös kaikkia niitä, jotka ovat jollakin tavalla auttaneet graduni teossa, mutta eivät ole tulleet mainittua.

Jyväskylässä 31. lokakuuta 2003

Elina Nykyri

## Tiivistelmä

Elina Nykyri: *Yritysten tilinpäätöstietojen toimipaikkaistamisen ongelma. Yleistetyin regressioestimoinnin sovel-  
lus rekisteritutkimuksessa.* Tilastotieteen pro gradu -tutkielma, Jyväskylän yliopisto, 31. lokakuuta 2003. Sivuja  
109, liitteitä 9.

Tutkimuksessa kehitettiin tilastollisten mallien avulla menetelmä yritysten tilinpäätöstietojen toimipaikkaista-  
miseen. Toimipaikkaistamisella tarkoitetaan yritystasojen tietojen disaggregointia toimipaikkatasoisiksi tie-  
doiksi. Kehitettyä menetelmää sovellettiin jalostusarvon toimipaikkaistamiseen. Jalostusarvo valittiin toimipaik-  
kaistettavaksi tilinpäätöstiedoksi, koska se kuuluu tärkeimpiin yritysten ja toimipaikkojen toimintaa kuvaaviin tie-  
toihin. Käytettävissä olivat Tilastokeskuksen tiedot Suomen yrityksistä ja toimipaikoista vuodelta 2000: yritys- ja  
toimipaikkarekisteri sekä rakennetilastojen kyselykanta. Ne yhdistämällä saatiin tutkimusaineisto.

Suomessa ei voida tuottaa tarkkoja koko toimialakenttää koskevia ja/tai alueellisesti ja hallinnollisesti kattavia  
tilastoja tilinpäätöstiedoista, koska niitä voidaan laatia vain toimipaikkakohtaisten tilinpäätöstietojen perusteella  
ja Tilastokeskuksessa ei tuoteta kaikkien toimipaikkojen tietoja. Näitä tilastoja kuitenkin arvioidaan Tilastokes-  
kuksessa yritystietojen avulla. Kaikki toimipaikkatiedot tuotetaan monitoimipaikkaisten yritysten osalta vain teol-  
lisuuden ja rakentamisen toimialoille sekä kaikille toimialoille yksitoimipaikkaisten yritysten osalta. Teollisuuteen  
ja rakentamiseen luetaan myös mineraalien kaivuun sekä sähkön-, kaasun- ja vesihuolto.

Tutkimustehtävänä oli arvioida olemassa olevien yritys- ja toimipaikkatietojen avulla toimipaikkatasoiset tie-  
dot niille toimialoille, joilta tiedot puuttuivat. Nämä toimialat olivat alkutuotannon ja palvelujen toimialoja. Mene-  
telmänä käytettiin imputointia, jossa käytettiin hyväksi saatavilla olevaa lisäinformaatiota. Malliluovuttaja-impu-  
tointimenetelmässä puuttuva tieto korvataan estimaatilla, joka on muodostettu käyttäen tilastollista mallia. Mal-  
lin parametrit estimoidaan käyttämällä niiden havaintoyksiköiden tietoja, joiden tulosuuttujan ja apumuuttujien  
havaintoarvot ovat tiedossa. Puuttuvat tiedot imputoitiin näiden mallien avulla niille toimialoille, joilta toimipaik-  
katasoiset tiedot puuttuivat. Imputoinnissa käytettävä malli riippuu siitä mikä on käytettävien apumuuttujien  
luonne. Tutkimuksessa mallina/luovuttajana käytettiin lineaarista mallia, joka oli määritelty toimialakohtaisesti  
tai kaikille toimialoille yhteisenä.

Tutkimuksessa huomattiin, että malliluovuttaja-imputoinnin onnistuneisuus riippui pitkälti toimialasta ja siitä  
onko toimipaikan omistava yritys yksi- vai monitoimipaikkainen (eli yrityksen lajista). Kaikilla toimialoilla toimia-  
lakohtainen malli paransi estimaattien laatua. Tämä tarkoitti sitä, että suoriin estimaattoreihin perustuva impu-  
tointi tuotti todellisia arvoja lähempänä olevia estimaatteja toimialakohtaisille jalostusarvon kokonaismäärille  
kuin epäsuoriin perustuva. Teollisuuden ja rakentamisen toimialoilla riippui toimipaikan toimialasta olivatko yri-  
tyksen laji- ja toimialakohtaiset estimaatit jalostusarvon totaalille parempia kuin toimialakohtaiset. Mineraalien  
kaivuun sekä rakentamisen toimialoilla olivat toimialakohtaiset totaalin estimaatit parempia ja teollisuuden sekä  
sähkön-, kaasun- ja vesihuollon toimialoilla yrityksen laji- ja toimialakohtaiset totaalin estimaatit. Teollisuuden ja  
rakentamisen toimialoilla eivät myöskään samat mallit toimineet sekä yksi- että monitoimipaikkaisten yritysten  
toimipaikkojen joukossa. Siksi ei ole aihetta olettaa, että kaikilla toimialoilla samat mallit toimisivat sekä yksi- et-  
tä monitoimipaikkaisten yritysten joukossa. Tarvittaisiinkin tietoja monitoimipaikkaisten yritysten toimipaikoista  
toimialoilta A, B ja G-O eli alkutuotannon ja palvelujen toimialoilta.

*Avainkäsitteitä:* lisäinformaation käyttö otannassa ja estimoinnissa, malliluovuttaja-imputointi, jalostusarvo,  
rekisteritutkimus, toimipaikkaistaminen, yleistetty regressioestimointi.

# Sisällysluettelo

|  |            |
|--|------------|
| <b>1 Johdanto</b>  | <b>2</b>   |
| 1.1 Yritysten tilinpäätöstietojen toimipaikkaistamisen määrittely ja tutkimuksen lähtökohdat.....                  | 2          |
| 1.2 Tutkimustehtävät, tutkimusvaiheet ja tutkimusaineiston muodostaminen.....                                      | 6          |
| <b>2 Tulosmuuttujan määrittely ja tulosmuuttujaan vaikuttavat tekijät</b>  | <b>11</b>  |
| 2.1 Jalostusarvon määrittely, yrityksen tilinpäätös ja toimipaikan laskennallinen käyttökate .....                 | 11         |
| 2.2 Toimipaikan tilinpäätöstietoihin vaikuttavat tekijät.....  | 16         |
| <b>3 Tutkimusaineisto</b>  | <b>21</b>  |
| 3.1 Yritys- ja toimipaikkarekisteri sekä rakennetilastojen kyselykanta .....                                       | 21         |
| 3.2 Tutkimuksessa käytettävät muuttujat ja perusjoukko .....   | 27         |
| 3.3 Tutkimusaineisto tiivistettynä.....  | 32         |
| <b>4 Liikeryritysten toimipaikat Suomessa</b>  | <b>34</b>  |
| 4.1 Liiketoimipaikat vuonna 2000.....  | 35         |
| 4.2 Tilinpäätöstietojen yhteisvaihtelu ja multikollineaarisuus aineistossa.....                                    | 41         |
| <b>5 Tutkimusmenetelmät</b>  | <b>45</b>  |
| 5.1 Lisäinformaation käyttö otannassa ja estimoinnissa .....   | 46         |
| 5.2 Otanta-asetelma.....   | 51         |
| 5.3 Estimointiasetelmat- ja strategiat.....  | 60         |
| 5.4 Otantavirhe ja otannasta riippumattomat virheet.....   | 66         |
| 5.5 Malliluovuttaja-menetelmällä imputoiminen/ennustaminen.....  | 71         |
| 5.6 Diagnostiset tarkastelutavat.....  | 75         |
| <b>6 Toimialakohtaiset tilinpäätöstiedot</b>   | <b>77</b>  |
| 6.1 Mallin ja estimaattorin rooli sekä estimaattien vertailu .....   | 78         |
| 6.2 Yritysten jalostusarvon toimipaikkaistaminen ja toimipaikkaistamisen rajoitteet.....                           | 87         |
| 6.3 Toimialojen merkittävyys Suomessa vuonna 2000.....   | 90         |
| <b>7 Yhteenvetoa</b>   | <b>93</b>  |
| 7.1 Keskeisiä tuloksia .....   | 93         |
| 7.2 Suosituksia toimipaikkojen tilinpäätöstietojen tuottamiselle tulevaisuudessa .....                             | 99         |
| <b>8 Tutkimuksen laatuselvitys</b>   | <b>101</b> |
| <b>Lähteet</b>   | <b>108</b> |
| <b>Liitteet</b>  | <b>110</b> |
| Liite 1. Yrityksen toiminta .....  | 110        |
| Liite 2. Toimialaluokitus TOL 95 ja yrityksen jako yksiköihin .....  | 116        |
| Liite 3. Kululajipohjainen tuloslaskelma .....   | 121        |
| Liite 4. Tutkimusaineiston muodostaminen SAS-ohjelmistolla yhdistämällä YTR ja rakennetilastojen kyselykanta ..... | 125        |
| Liite 5. Tutkimusaineiston tarkastelua SAS-ohjelmistolla .....   | 127        |
| Liite 6. Otoksien poiminta ja otoksien ominaisuuksia .....   | 128        |
| Liite 7. Toimialakohtaisten tilinpäätöstietojen estimaattien muodostaminen SAS-ohjelmistolla .....                 | 133        |
| Liite 8. Valitut toimialakohtaiset estimaatit.....   | 138        |
| Liite 9. Yritysten tilinpäätöstietojen toimipaikkaistaminen SAS-ohjelmistolla.....                                 | 143        |

# 1 Johdanto

Tämä pro gradu -tutkimus lähti liikkeelle Tilastokeskuksessa ilmenneestä kehitystarpeesta, joka liittyi yritysten toiminnan tilastointiin. Tarvittiin toimipaikkojen tilinpäätöstietojen arvioita niiden toimialojen toimipaikkojen osalta, joille Tilastokeskuksessa ei tuoteta kaikkia tilinpäätöstietoja. Kattavat toimipaikkatasoiset tiedot tuotetaan vain teollisuuden ja rakentamisen toimialoille monitoimipaikkaisten yritysten osalta sekä yksitoimipaikkaisten yritysten osalta kaikille toimialoille. Teollisuuteen ja rakentamiseen luetaan myös mineraalien kaivuu ja sähkön-, kaasun- ja vesihuolto. Toimipaikkakohtaisia tilinpäätöstietoja tarvitaan, koska vain toimipaikkatietojen perusteella voidaan tuottaa tarkkoja koko toimialakenttää koskevia ja/tai alueellisesti/hallinnollisesti kattavia tilinpäätöstilastoja. Näitä tilastoja ei voida siis laatia suoraan yritystiedoista, mutta niitä voidaan arvioida yritystietojen avulla, kuten Tilastokeskuksessa on tähän asti tehty.

Toimipaikkojen tietoja tuotetaan toimipaikkaistamalla yritysten tilinpäätöstietoja. Se tarkoittaa yritystasoisien tietojen disaggregointia toimipaikkatasoisiksi tiedoiksi. Tässä tutkimuksessa kehitettiin tilastollisten mallien avulla menetelmä yritysten tilinpäätöstietojen toimipaikkaistamiseen, ja menetelmää sovellettiin jalostusarvon toimipaikkaistamiseen. Jalostusarvo valittiin toimipaikkaistettavaksi tilinpäätöstiedoksi, koska se kuuluu tärkeimpiin yritysten ja toimipaikkojen toimintaa kuvaaviin tietoihin. Käytettävissä olivat Tilastokeskuksen tiedot Suomen yrityksistä ja toimipaikoista vuodelta 2000: yritys- ja toimipaikkarekisteri sekä rakennetilastojen kyselykanta. Ne yhdistämällä saatiin tutkimusaineisto.

Toimipaikkaistamismenetelmänä käytettiin saatavilla olevaa lisäinformaatiota hyödyntävää imputointimenetelmää. Malliluovuttaja-imputointimenetelmässä (model-donor imputation) puuttuva tieto korvataan tilastollisen mallin perusteella muodostetulla estimaatilla. Mallin parametrit estimoidaan käyttämällä niiden havaintoyksiköiden tietoja, joiden tulomuuttujan ja apumuuttujien havaintoarvot ovat tiedossa. Puuttuvat tiedot imputoitiin näiden mallien avulla niille toimialoille, joilta toimipaikkatasoiset tiedot puuttuivat. Imputoinnissa käytettävä malli riippuu siitä, mikä on käytettävien apumuuttujien luonne. Tutkimuksessa mallina/luovuttajana käytettiin lineaarista mallia, joka on määritelty toimialakohtaisesti tai kaikille toimialoille yhteisenä.

## 1.1 Yritysten tilinpäätöstietojen toimipaikkaistamisen määrittely ja tutkimuksen lähtökohdat

Yrityksillä on Suomessa tilinpäätöstensä tiedonantovelvollisuus Tilastokeskukselle ja verohallinnolle. Tämän lisäksi Tilastokeskuksen yritysten rakenteet -yksikkö suorittaa erilaisia tiedusteluja (TILKES-lomake). Kyselyyn tulleita yrityksiä pyydetään erittelemään tilinpäätöstietojaan. Lisäksi yritysten rakenteet -yksikkö lähettää kyselyyn niille monitoimipaikkaisille yrityksille ja toimipaikoille, joiden toimiala on teollisuus tai rakentaminen. Teollisuuden ja rakentamiseen luetaan myös mineraalien kaivuu ja sähkön-, kaasun- ja vesihuolto (teollisuus ja rakentaminen käsittävät toimialat C-F, katso liite 2). Kyselyyn tulleita yrityksiä pyydetään ilmoittamaan tilinpäätöstietonsa toimipaikkakohtaisesti - siis toimipaikkaistamaan tilinpäätöstietojaan. Toimipaikkaistaminen, eli yritysta-

soisten tietojen disaggregointi toimipaikkatasoisiksi tiedoiksi, vaatii käytännössä yritysten tilinpäätöstietojen muodostumisen selvittämistä. Esimerkiksi selvitetään mitkä ovat toimipaikkojen osuudet yrityksen kokonaisliikvaihdosta ja missä toimipaikoissa yrityksen henkilökunta työskentelee. Tämä vaatii sitä, että jokainen toimipaikka pitää kirjanpitoa toiminnastaan. Tilastokeskuksessa pyritään vähentämään yritysten suurta vastausrasitetta siirtymällä suoran kyselyn sijasta estimointiin, joka tehdään osalle tiedoista.

Koska Tilastokeskus ei tee kyselyä toimipaikkojen tilinpäätöstiedoista muille kuin teollisuuden ja rakentamisen toimipaikoille, kaikkien toimipaikkojen tilinpäätöstietoja ei ole saatavilla. Kattavia tilinpäätöstietoja on Tilastokeskuksessa vain yksitoimipaikkaisten yritysten toimipaikoista ja monitoimipaikkaisten yritysten teollisuuden ja rakentamisen toimipaikoista. Koska toimipaikkakohtaisia tilinpäätöstietoja ei ole saatu tai kysytty kaikilta yrityksiltä, ei Suomessa voida tuottaa tarkkoja koko toimialakentää koskevia ja/tai alueellisesti/hallinnollisesti kattavia tilinpäätöstilastoja. Näitä tilastoja voidaan kuitenkin arvioida yritystietojen avulla, kuten Tilastokeskuksessa tehdään.

Monitoimipaikkaiset yritykset eivät pysty toimipaikkaistamaan tietojaan, jos ne eivät ole pitäneet toimipaikka-kohtaista kirjanpitoa. Kaikki yritykset eivät ole halukkaitakaan toimipaikkaistamaan tietojaan. Yritysten tilinpäätöstietoja on kuitenkin mahdollista toimipaikkaistaa likimääräisesti esimerkiksi tilastollisten mallien avulla. *Tutkimuksen tavoitteena onkin kehittää menetelmä, jolla voitaisiin käytettävissä olevien yritys- ja toimipaikkatietojen avulla yritysten tilinpäätöstietoja disaggregoida toimipaikkatasolle tilastollisten mallien avulla. Tutkimuksessa siis kehitetään menetelmä, jolla voidaan estimoida toimipaikkojen tilinpäätöstietoja tilastollisten mallien avulla. Kehitettyä menetelmää sovelletaan toimipaikkojen jalostusarvojen toimipaikkaistamiseen, koska jalostusarvo on yksi tärkeimmistä yritysten ja toimipaikkojen toimintaa kuvaavista tiedoista.* Tutkimuksessa kehitettyä menetelmää on tarkoitus käyttää ensisijaisesti toimiala- ja aluetilastojen laatimisessa. Uuden menetelmän avulla voidaan kehittää Tilastokeskuksen tuottamien toimiala- ja aluekohtaisten tilastojen ja toimialakatsausten laatua - siis myös parantaa tietopalvelua.

Ennen tätä tutkimusta toimiala- ja aluekohtaiset tilastot sekä toimialakatsaukset perustuivat yrityksen päätoimialaan (päätoimialan käsite on liitteessä 2), eivätkä toimipaikkakohtaisiin tietoihin. Tämä merkitsi sitä, että toimialakohtaiset tiedot estimoitiin ottamalla huomioon kaikki yksitoimialaiset yritykset ja monitoimialaiset yritykset, joiden päätoimialan toiminnasta muodostui suurin osa arvonnäyksestä. Toimialakohtaisten estimaattien laadinnassa huomioitiin siis vain yritykset, jotka toimivat pääasiasiassa yhdellä toimialalla. Siksi toimialakohtaiset estimaatit eivät sisältäneet tietoja sellaisista yrityksistä, joiden minkään toimialan toiminnasta ei muodostunut yli 90 % arvonnäyksestä. Estimaattien rajoitteena oli myös se, että yritysten toiminta tilastoitiin yritysten päätoimialoille, ja siksi toimialakohtaiset arviot eivät huomioineet yritysten sivutoimialojen osuuksia. (Tilastokeskus, 1993, Tilastokeskus, 2002c).

Toimialakohtaisia tilastoja voidaan käyttää yritysten tilinpäätöstietojen vertailussa niiden toimialoja vastaaviin tietoihin. Toimialakohtaisten tilastojen avulla voidaan esimerkiksi tutkia onko yrityksen jalostusarvo keskimääräistä samalla toimialalla toimivaa yritystä "parempi". Tämä tehdään vertaamalla yrityksen jalostusarvoa toimialakohtaisiin tilastoihin eli yritystietoa verrataan keskimääräiseen jalostusarvoon esim. yrityksen päätoimialalla.

Aiemmissa tutkimuksissa on todettu toimialojen välillä olevan eroja tilinpäätöstiedoissa, ja siksi ei ole järkevää verrata yksittäisen yrityksen/toimipaikan tietoja kaikkien yritysten/toimipaikkojen tietoihin, vaan pelkästään samalla toimialalla toimivien yritysten ja niiden toimipaikkojen tietoihin (Kinnunen ym., 2000, 162-167, Maliranta, 1996).

Toimialakohtaisten tilinpäätöstietojen avulla voidaan myös tutkia toimialojen välisiä eroja sekä erottaa toimialakohtaisista tekijöistä yleistaloudelliset ja yrityskohtaiset tekijät. Toimipaikkojen tilinpäätöstietojen perusteella on mahdollista laatia ns. toimialakatsauksia, joissa tutkitaan mm. toimialan toiminnan laajuutta, kokoa, kannattavuutta ja vakavaraisuutta. Toimialakohtaisten tilinpäätöstietojen estimaatit mahdollistavat myös yleistaloudellisten, toimiala- ja yrityskohtaisten tekijöiden merkityksen arvioimisen yksittäisen yritysten kannalta. ***Toimipaikan tilinpäätöstietoja voidaan siis käyttää moneen tarkoitukseen (mm. tilinpäätösanalyysi, tunnuslukuanalyysi ja yritystutkimus).*** (Kinnunen, ym. 2000).

Estimoitujen toimipaikkojen tilinpäätöstietojen avulla saadaan myös arvioita eri alueilla ja toimialoilla toimivien yritysten tilinpäätöstiedoille (ts. voidaan yhdistää alue- ja toimialadimensiot), voidaan mm. estimoida teollisuuden toimialan toiminnan jalostusarvon kokonaismäärä Keski-Suomessa. Alue- ja toimialakohtaisten estimaattien yhdistäminen mahdollistaa sen, että voidaan tutkia sitä miten alueiden välinen työnjako ja toimipaikan tai yrityksen toimiala vaikuttavat yrityksen tai toimipaikan toimintaan. Aiemmasta tutkimuksesta on selvinnyt, että sekä alueellisuus että toimiala vaikuttavat siihen miten yritys toimii. (Maliranta 1996, Kinnunen ym., 2000, 162-167 ja Virkkala 1987).

Tämä tutkimus liittyy Tilastokeskuksessa meneillään olevaan projektiin nimeltään alue- ja toimialatilastojen kehittäminen rakennetilastoissa. Hankkeessa pyritään kehittämään yritystoiminnan perusaineistojen kattavuutta tuottamalla palvelualoilta alue- ja toimialakohtaista tietoa. Palvelualoja koskevia tietoja halutaan kehittää, koska palvelujen osuus yritystoiminnasta on kasvanut pitkällä aikavälillä ja yritystoiminta ylipäänsä on tärkeä osa talous- ja yhteiskuntakehitystä. Lisäksi Eurostat on pannut alulle hankkeen, ”jonka tavoitteena on kerätä tietoja jäsenmaista palvelualojen hyödykkeiden tuotannosta sekä asiakasryhmistä.” Itse asiassa Tilastokeskuksen hanke on osa EU-hanketta. (Koivuniemi 2002, 4).

Projektin tavoitteena on palvelualojen alue- ja toimialakohtaisten tietojen lisäksi selvittää voiko teollisuuden ja rakentamisen toimialojen toimiala- ja aluetilastojen tuottamisessa käyttää nykyistä paremmin jo olemassa olevia tietoja. Lisäksi projektin tavoitteena on vähentää tietojen kyselyn tarvetta teollisuuden ja rakentamisen toimialojen osalta (toimialat C-F, liite 2) ja kehittää Tilastokeskuksen yritysten rakenteet -yksikön tuottamien tietojen asiakaslähtöisyyttä mm. yhdenmukaistamalla yritys- ja toimipaikkarekisterin (yritysten suhdanteet -yksikön tuottamat yritys- ja toimipaikkatiedot) ja rakennetilastojen kyselykannan (yritysten rakenteet -yksikön tuottamat yritys- ja toimipaikkatiedot) käyttämiä toimintatapoja (luku 3.1). Palvelualojen alue- ja toimialakohtaisia tietoja voitaisiin tuottaa jo olemassa olevien tietojen avulla käyttämällä tilastollisia malleja tai voitaisiin kysellä monitoimipaikkaisten yritysten palvelualojen toimipaikoilta joitakin tärkeimpiä tilinpäätöstietoja. (Koivuniemi, 2002).

Alue- ja toimialakohtaisia tilastoja käyttävät Tilastokeskuksen kansantalouden tilinpito, OECD, Eurostat, eri toimialajärjestöt, tutkimuslaitokset, kansainväliset järjestöt, Tilastokeskuksen yritysten suhdanteet -yksikkö ja yritysten rakenteet -yksikkö sekä niitä pyytävät erilaiset Tilastokeskuksen ulkoiset asiakkaat (esim. yritykset). (Koivuniemi, 2002).

Tutkimus on jäsennetty seuraavasti. Luvussa 1 dokumentoidaan tutkimusongelma ja tutkimusvaiheet. Luku 2 kuvaa yritysten tilinpäätöstä sekä toimipaikan laskennallisen käyttökattteen muodostamista ja niihin vaikuttavia taustatekijöitä. Luvussa 3 määritellään tutkimuksessa käytettävät muuttujat ja perusjoukko. Lisäksi raportoidaan tutkimusaineiston muodostaminen yhdistämällä yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan tietoja. Luvussa 4 kuvataan Suomen liiketoimipaikkoja vuonna 2000, mm. niiden demografisten ominaisuuksien kautta. Lisäksi kerrotaan tutkimuksessa käytettävien muuttujien riippuvuussuhteista aineistossa. Luvussa 5 raportoidaan tutkimuksessa käytetyt tutkimusmenetelmät eli kuvataan sitä millainen toimipaikkojen tilinpäätöstietojen toimipaikkaistamismenetelmä on ja mitä teoreettisia perusteita tutkimuksessa laaditulle menetelmälle on. Luvussa 6 raportoidaan toimialakohtaisia jalostusarvon totaalin estimaatteja. Lisäksi tutkitaan laskettujen estimaattien ominaisuuksia ja sitä miten hyviä valitulla menetelmällä laaditut estimaatit toimipaikkojen jalostusarvolle ovat. Hyvyyden kriteereinä käytetään mm. estimaatin keskivirhettä, variaatiokerrointa, keskineliövirheen neliöjuurta (RMSE, Root Mean Squared Error) sekä estimaatin poikkeamaa todellisesta arvosta. Luvussa 6 arvioidaan lisäksi toimialojen merkittävyyttä Suomessa vuonna 2000 eri ominaisuuksien suhteen ja esitetään arvioita/estimaatteja eri toimialojen jalostusarvoille. Haluttaessa saatuja tietoja voidaan verrata kansantalouden tilinpidon antamiin lukuihin. Luvussa 7 on koottu tutkimuksessa saadut tulokset ja esitetty suosituksia sille miten voisi olla tarkoituksenmukaista toimia yritysten tilinpäätöstietojen toimipaikkaistamisessa ja toimiala- ja aluekohtaisten tilinpäätöstietojen laatimisessa. Luku 8 sisältää erillisen tutkimuksen laaturaportin, joka on tehty arvioimalla Tilastokeskuksessa käytettävien laatukriteereiden avulla tutkimuksen laatua ja laatuun vaikuttavia tekijöitä.

Tutkimusraportti sisältää 9 liitettä. Liitteissä 1-3 kuvataan yrityksen toimintaa, yritysten jakamista toimialoihin niiden toiminnan luonteen mukaan sekä kululajipohjaista tuloslaskelmaa. Liitteissä 1-3 raportoidaan myös tapaa jakaa yritys erilaisiin yksiköihin määrittelemällä toimipaikka sekä paikallis- ja toimialayksiköt. Kaiken kaikkiaan liitteissä 1-3 on määritelty tutkimuksessa käytettävät käsitteet ja kuvattu tutkimuskohdetta. Tutkimustyö on sisältänyt paljon SAS-ohjelmakoodien kirjoittamista, joista pieni osa on valittu tutkimusraportin liitteiksi 4-7 ja 9. Koska nämä SAS-koodit on tehty makrojen avulla, voidaan lähes kaikkia SAS-koodeja käyttää myös muilla aineistolla ja muuttujilla.

Liitteessä 4 raportoidaan tutkimusaineiston muodostaminen SAS-ohjelmaa käyttämällä. Liite 5 kuvaa sitä miten tutkimusaineistoa tarkasteltiin SAS-ohjelmalla. Liitteessä 6 dokumentoidaan tutkimuksessa käytettyjen otosten poimintaan tarvittavat SAS-koodit ja verrataan otoksien ominaisuuksia niitä vastaaviin perusjoukon ominaisuuksiin. Liite 7 kuvaa sitä miten tutkimuksessa estimoitavat toimialakohtaisten jalostusarvojen estimaatit on laskettu. Liitteessä 8 raportoidaan mallit, joiden perusteella toimipaikkojen jalostusarvot toimipaikkaistettiin. Liitteessä 9 raportoidaan yritysten tilinpäätöstietojen toimipaikkaistamiseen laaditun menetelmän soveltaminen toimipaikkojen jalostusarvojen estimointiin SAS-ohjelmalla.



## 1.2 Tutkimustehtävät, tutkimusvaiheet ja tutkimusaineiston muodostaminen

### Tutkimustehtävä

Koska Tilastokeskuksessa kaikkien toimipaikkojen kaikki tilinpäätöstiedot ovat tiedossa vain yksitoimipaikkaisilta yrityksiltä ja monitoimipaikkaisten yritysten teollisuuden ja rakentamisen toimipaikoilta, **tutkimustehtävä on kehittää ja tutkia menetelmiä, joilla voidaan tuottaa mahdollisimman tarkkoja arvioita toimipaikkojen tilinpäätöstiedoille tilastollisten mallien avulla (eli miten disaggregoida yritysten tilinpäätöstietoja toimipaikkatasolle) käyttämällä saatavilla olevaa informaatiota.** Tutkimustehtävä voidaan jakaa tutkimusongelmiin, joita ovat seuraavat.

- **1. Pääongelma: Menetelmien kehittäminen yritysten tilinpäätöstietojen dissagregointiin toimipaikkatasolle.** Alaongelmat:

1. **Lisäinformaation saatavuus.** Selvitetään sitä onko mahdollista käyttää tietoja, jotka parantaisivat toimipaikkojen tilinpäätöstietojen estimaattien laatua.

2. **Toimipaikan tilinpäätöstietoihin vaikuttavat tekijät.** Tutkitaan toimialan vaikutusta vaikuttaa toimipaikan toimintaa, millaisia ovat toimialojen väliset erot ja miten ottaa toimipaikan toimiala huomioon yritysten tilinpäätöstietojen toimipaikkaistamisessa. Tutkitaan myös olisiko tarkoituksenmukaista käyttää toimialakohtaisia malleja vai kaikille toimialoille yhteisiä malleja disaggregoinnin taustalla ja riippuuko tämä valinta toimialasta. Selvitetään sitä miten toimipaikkaa vastaavan yrityksen laji (eli se onko toimipaikkaa vastaava yritys yksi- vai monitoimipaikkainen) vaikuttaa toimipaikan tilinpäätöstietoihin ja miten ne voitaisiin ottaa menetelmien laatimisessa huomioon. Lisäksi tutkitaan toimipaikan sijainnin ja koon vaikutusta toimipaikan tilinpäätöstietoihin ja niiden vaikutuksen huomioonottamista menetelmien kehittämisessä.

-**2. Pääongelma: Menetelmien kehittäminen toimipaikkatason tietojen laadun tarkastelulle teollisuuden ja rakentamisen toimialoilla.**

- **3. Pääongelma: Päätelmien teko toimipaikkaistamisen toimivuudesta muiden kuin teollisuuden ja rakentamisen toimialojen kohdalla (toimialat C – F) ja laaditun laskentamenetelmän käytön rajoitteet.** Alaongelmat:

1. **Toimipaikan toimialan vaikutus tilinpäätöstietoihin.** Tarkkoja toimialakohtaisia tilinpäätöstietoja ei saada laskettua yrityskohtaisten tilinpäätöstietojen avulla, koska niissä kaikki yrityksen toiminta on tilastoitu sen päätoimialalle. Yrityskohtaisissa tilinpäätöstiedoissa ei siis näy yritysten sivutoimialojen osuuksia yrityksen tilinpäätöstiedoista, esimerkiksi teollisuuden alan yrityksen jalostusarvo ei kerro siitä kuinka suuri osa tuosta jalostusarvosta koostuu tuon yrityksen kaupan alan toiminnasta.

2. **Toimipaikkaa vastaavan yrityksen lajin (eli se onko yritys yksi- vai monitoimipaikkainen) vaikutus toimipaikan tilinpäätöstietoihin ja estimaattien laatiminen niille toimipaikoille, joille ei toimipaikan tilinpäätöstietoja ole saatavilla.** Tutkitaan sitä miten monitoimipaikkaisten yritysten toimipaikkojen poisjäntti estimaattien laadinnasta vaikuttaa ja voiko käyttämällä vain yksitoimipaikkaisten yritysten tilinpäätöstietoja saada hyviä estimaatteja monitoimipaikkaisten yritysten toimipaikkojen tilinpäätöstiedoille. Jos yksi- ja monitoimipaikkaisten

yritysten toimipaikat eivät eroa toisistaan henkilöstömäärän, liikevaihdon ja toimipaikalle estimoitujen palkkojen suhteen, voidaan ajatella etteivät yksi- ja monitoimipaikkaiset yritykset eroa toisistaan tilinpäätöstiedoissa ja voidaan pitää toimipaikkojen tilinpäätöstiedoille laadittuja estimaatteja "hyvinä" ja tehdä diagnostisia tarkasteluja yksitoimipaikkaisten yritysten tilinpäätöstietojen perusteella.

3. **Diagnostisten tarkastelutapojen valinta.** Otoksen poimiminen monitoimipaikkaisten yritysten alkutuotannon ja palvelujen toimialojen toimipaikoista ja kyselyn tekeminen otokseen tulleille toimipaikoille voisi olla mahdollinen keino tarkastella toimipaikkojen tilinpäätöstietojen laatua. Otosyksiköiltä täytyy kuitenkin saada hyväksyntä, siis kyselyn täytyy tehdä yhteistyössä yritysten kanssa. Tämä aiheuttaa sen, että ei voida välttämättä käyttää otosta, vaan joudutaan tyytymään näytteeseen. Muutaman monitoimipaikkaisen yrityksenkin toimipaikkojen tiedot (ns. case-esimerkit) olisivat kuitenkin hyödyllisiä (katso luvut 8.2 ja 8.3).

#### **- 4. Pääongelma: Rekistereissä ja tietokannoissa olevien toimipaikkojen tietojen luotettavuus**

Yritykset, jotka ovat pieniä (palkatun henkilöstön määrä on alle 20) on tilastoitu rakennetilastojen toimipaikkatiedoissa (TKANNASSA) yksitoimipaikkaisiksi, vaikka ne saattavat tosiasiallisesti olla monitoimipaikkaisia. Tätä ei tutkimuksessa pystytä tutkimaan, mutta se on mahdollinen jatkotutkimuksen kohde. Lisäksi yritys- ja toimipaikkarekisterissä on kaksi tapaa määrittää toimipaikkaa vastaavan yrityksen laji ja määrittelytavat voivat joissain tapauksissa olla ristiriidassa keskenään. Toimipaikan toimialan määrittelykin riippuu siitä käyttääkö rakennetilastojen kyselykantaa vai toimipaikkarekisteriä. Toimipaikkojen tilinpäätöstiedoista osa on estimaatteja ts. impu-toituja havaintoarvoja ja tämä aiheuttaa epävarmuutta siitä ovatko toimipaikkojen tiedot luotettavia. Lisäksi se vaikeuttaa estimointia. Näistä asioista on kerrottu tarkemmin luvuissa 3-8.

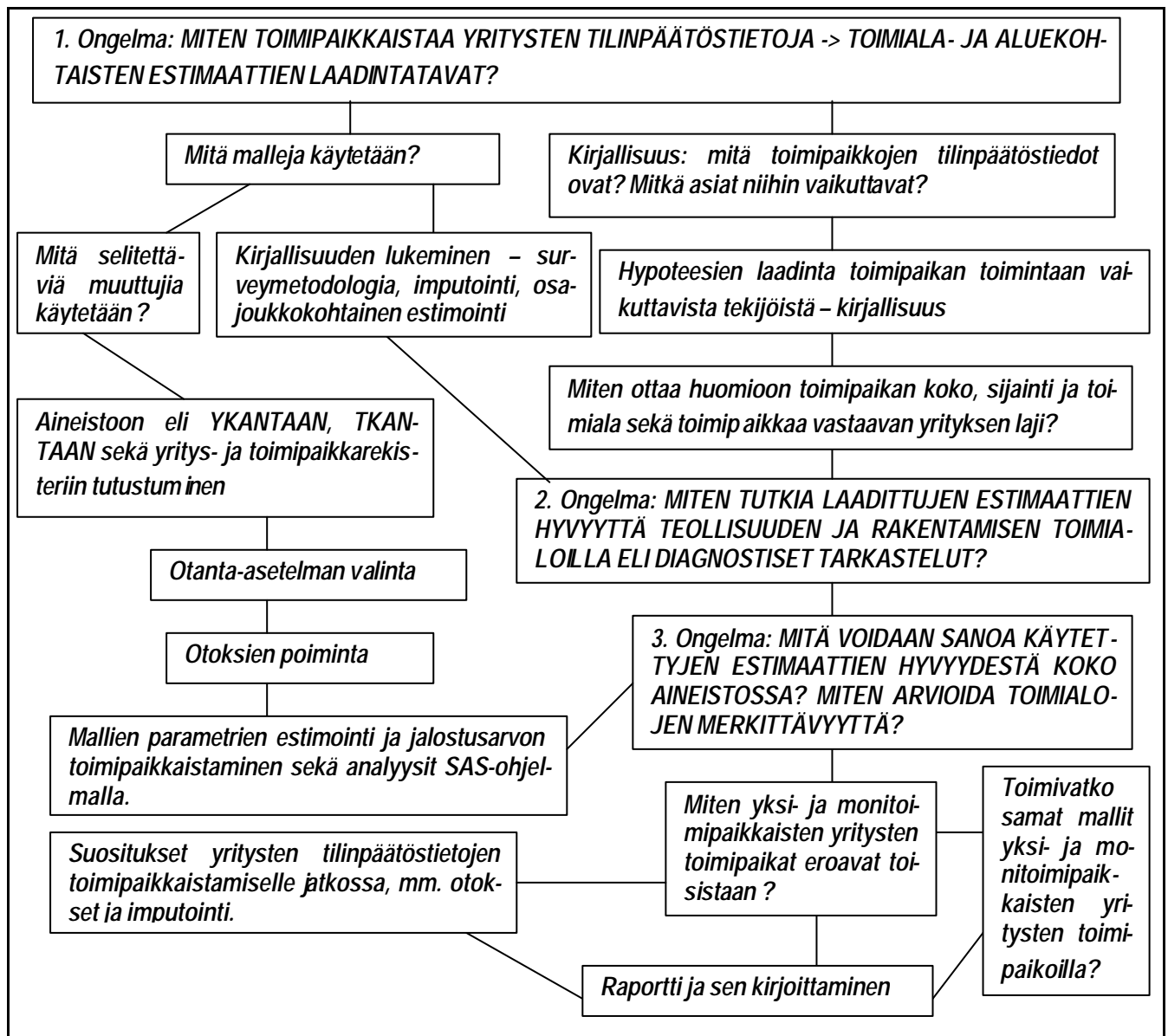
TUTKIMUKSEN HYPOTEESI, JOTKA OVAT KIRJALLISUUDEN PERUSTEELLA OLETETTUJA TOIMIPAIKAN TILINPÄÄTÖSTIETOIHIN VAIKUTTAVIA TEKIJÖITÄ JA EROJA TOIMIALOJEN VÄLILLÄ :

1. **Toimipaikan tilinpäätöstietoihin ja niihin muuttuja-arvoihin, joille estimaatteja laaditaan, vaikuttavat yleistoloudelliset, toimialakohtaiset, alueelliset, toimiala- ja aluekohtaiset, yrityskohtaiset ja toimipaikkakohtaiset tekijät.** Toimipaikan tilinpäätöstietoihin vaikuttavista taustatekijöistä on kerrottu luvussa 2.2.
2. Eri toimialoilla toimivat yritykset/toimipaikat tuottavat erilaisia hyödykkeitä ja siksi tarvitsevat erilaisia tuotantokelijöitä. Sidosryhmien merkityksessä on myös eroja ja yrityksen/toimipaikan kilpailijoita ovatkin ne yritykset ja toimipaikat, jotka toimivat samalla toimialalla. Yrityksen asema tässä kilpailussa vaikuttaa myös sen tilinpäätöksen muodostumiseen.
3. Työn tuottavuus vaihtelee Suomessa toimialoittain (Maliranta 1996, 5). Esimerkiksi teollisuudessa paperiteollisuus ja metalliteollisuus ovat tuottavampia kuin elintarviketeollisuuden alat. Tuottavuuden kasvu on ollut Suomessa nopeaa teollisuuden alalla, mutta ei niin nopeaa metallin perusteollisuudessa sekä toimialaryhmässä, johon kuuluu mm. paperiteollisuus (Maliranta, 1996, 5). Teollisuus vaikuttaa kaiken kaikkiaan kansakunnan tuottavuuteen muita aloja enemmän, mutta on riippuvainen muiden toimialojen yrityksistä.
4. Saman toimialan sisällä yritysten/toimipaikkojen tilinpäätöstiedot ovat verrattavissa keskenään. Mutta eri toimialoilla toimivien yritysten/toimipaikkojen tilinpäätöstiedot eivät ole verrannollisia, koska eri toimialoilla toimivien yritysten ja toimipaikkojen toiminta on erilaista (liite 2). Eri toimialoilla toimivat yritykset/toimipaikat eroavat toisistaan mm. tuottamiensa suoritteiden, työvoima- ja pääomavaltaisuuden sekä kiinteiden ja muuttuvien kustannusten osuuksien suhteen. (Kinnunen ym., 2000, 162 - 167)

5. Se onko yritys tai toimipaikka pääoma- vai työvoimavaltainen, vaikuttaa erityisesti yrityksen tai toimipaikan jalostusarvoon, henkilöstökuluihin ja kiinteisiin/muuttuviin kuluihin. Jalostusarvo on pääomavaltaisilla yrityksillä ja toimipaikoilla työvoimavaltaisia suurempi, kun taas henkilöstökulut ovat suuremmat työvoimavaltaisissa yrityksissä ja toimipaikoissa kuin pääomavaltaisissa. Kiinteät kulut ovat pääomavaltaisissa yrityksissä ja toimipaikoissa suuremmat kuin työvoimavaltaisissa, ja muuttuvat kulut taas suuremmat työvoimavaltaisissa yrityksissä ja toimipaikoissa. Kaiken kaikkiaan työvoima- ja pääomavaltaiset toimipaikat eroavat toisistaan meno- ja tulorakenteeltaan. Se onko yritys tai toimipaikka työvoima- vai pääomavaltainen, määräytyy pitkälti toimialan mukaan. Ne toimipaikat, jotka toimivat palvelualalla (toimialat G-Q) ovat työvoimavaltaisia ja ne, jotka toimivat alkutuotannossa ja jalostuksessa (toimialat D-F) ovat pääomavaltaisia. Koska kaikki toimipaikkojen tilinpäätöstiedot ovat tiedossa teollisuuden ja rakentamisen toimialoilta (pääomavaltaisia toimialoja), voidaan pitää laadittavia toimipaikkojen tilinpäätöstietoja parempina pääomavaltaisille toimipaikoille kuin työvoimavaltaisille.

## Tutkimuksen eteneminen

Kuviossa 1.1 on tiivistetty tutkimuksen kulku osavaiheineen. Kuvion jälkeen on kuvattu jokaista vaihetta yksityiskohtaisesti ja mainittu missä luvussa kustakin vaiheesta on raportti.



Kuvio 2.1. Tutkimuksen kulku.

## Tutkimusaineiston muodostaminen, analysointi, toimipaikkaistaminen sekä diagnostiset tarkastelut

Tutkimus eteni vaiheittain aineiston muodostamisesta aina yritysten jalostusarvon toimipaikkaistamiseen saakka. Tutkimuksen kuluessa jouduttiin palaamaan aiempiin vaiheisiin ja tutkimusongelma tarkentui. Ensimmäiseksi muodostettiin käytettävä aineisto ja sitä analysoitiin. Tämän jälkeen aineistosta poimittiin otoksia ja niiden perusteella laadittiin toimialakohtaisia estimaatteja jalostusarvolle. Lopuksi valittujen toimialakohtaisten estimaattien pohjalta tehtiin yritysten jalostusarvon toimipaikkaistaminen ja diagnostisia tarkasteluja.

Seuraavassa ovat tutkimuksen vaiheet lyhyesti kuvattuna:

1. Muodostetaan aineisto, joka sisältää kaikki toimipaikat, joiden tilinpäätöstietoja on tarkoitus toimipaikkaistaa. Lisäksi aineisto sisältää toimipaikkoja vastaavien yritysten tiedot. Se tehdään siten, että yhdistetään yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan toimipaikka- ja yritystason tiedot eli yhdistetään TKANTA, YKANTA (eli rakennetilastojen kyselykanta) ja YTR (eli yritys- ja toimipaikkarekisteri). Rekistereistä kertoo luku 3. Muodostettava aineisto koostuu kahdenlaisista havaintoyksiköistä. Rakennetilastojen mukaan määriteltyjen teollisuuden ja rakentamisen toimialoihin kuuluvilla toimipaikoilla ovat toimipaikkojen tilinpäätöstiedot tiedossa kaikille toimipaikoille, mutta muiden toimipaikkojen tiedot ovat tiedossa vain, jos toimipaikkaa vastaavalla yrityksellä on yksi toimipaikka ja YTR:n mukaan toimipaikan toimiala = yrityksen toimiala. Siksi tätä yhdistettyä aineistoa muodostettaessa siirretään yrityksen tilinpäätöstiedot toimipaikan tiedoiksi, jos toimipaikkarekisterin mukaan toimipaikkaa vastaavalla yrityksellä on yksi toimipaikka ja toimipaikan toimiala = yrityksen toimiala YTR:n mukaan. Tämä tehdään vain, jos toimipaikan tiedot eivät ole TKANNASSA. Aineiston laadinta on liitteessä 4.
2. Muodostetaan aineisto, joka sisältää ne toimipaikat, joille ovat tiedossa kaikki toimipaikan laskennallisen käyttökateen laskemiseen tarvittavat tiedot. Siis aineisto sisältää vain ne toimipaikat, joiden laskennallisen käyttökateen eri erien tiedot ovat TKANNASSA tai joita vastaavalla yrityksellä on yksi toimipaikka ja yrityksen toimiala = toimipaikan toimiala toimipaikkarekisterin mukaan. Puuttuvia havaintoarvoja on rakennetilastojen kyselykannan toimipaikkatason tiedoissa, ja siksi TKANNASSA olevien toimipaikkojen tiedoissa on puuttuvia arvoja eli tutkimusaineistossa on eräkatoa. Aineiston laadinta on liitteessä 4. Toimipaikan laskennallisesta käyttökatteesta kertoo luku 2.1.
3. Yhdistetään niiden toimipaikkojen yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan tiedot, joiden toimiala on TKANNAN mukaan teollisuus tai rakentaminen. Muodostetaan myös muita aineistoja.
4. Tutkitaan sitä miten toimipaikkojen tilinpäätöstiedot ovat jakaantuneet Suomessa vuonna 2000 (esim. mitkä ovat toimialojen väliset erot toimipaikkojen tietojen mediaaneissa, vinoudessa, huipukkuudessa ja ovatko toimipaikkojen tilinpäätöstiedot normaalisti jakaantuneet). Selvitetään millaisia ovat erot (jos niitä on) yksi- ja monitoimipaikkaisten yritysten toimipaikkojen välillä toimialoittain liikevaihdossa, henkilöstömäärissä ja toimipaikoille estimoiduissa palkoissa. Tutkitaan myös sitä onko käytettävien tulosuuttujien ja apumuuttujien välillä riippuvuutta ja jos on niin kuinka suurta se on Pearsonin korrelaatiokertoimella mitattuna ja onko tässä eroja toimialojen välillä. Lisäksi tutkitaan sitä miten vahvoja ovat käytettävien apumuuttujien väliset riippuvuus-suhteet Pearsonin korrelaatiokertoimella mitattuna ja onko riippuvuus-suhteiden voimakkuudessa eroja toimialojen ja/tai

- monitoimipaikkaisten ja yksitoimipaikkaisten yritysten toimipaikkojen välillä. Tarkastelujen tulokset ovat luvussa 4.
5. Poimitaan ositetulla otannalla käyttäen yksinkertaista satunnaisotantaa palauttamatta (stratified simple random sampling - STRSRS) 3 eri otosta. Ensimmäinen otos poimitaan teollisuuden ja rakentamisen toimialojen monitoimipaikkaisen yritysten toimipaikoista. Toinen otos poimitaan teollisuuden ja rakentamisen yksitoimipaikkaisten yritysten toimipaikoista. Kolmas otos poimitaan yksitoimipaikkaisista yrityksistä, joiden toimiala ei ole rakennetilastojen kyselykannan mukaan teollisuus tai rakentaminen (mutta saattaa olla toimipaikkarekisterin mukaan). Otosten poiminnassa käytettävät otanta-asetelmat ovat luvussa 5.2.
6. Poimitaan edellä mainittujen otosten lisäksi 3 otosta ositetulla otannalla kohdassa 5 mainitulla tavalla. Käytännössä tämä tehdään yhdistämällä edellä mainittuja otoksia. Neljäs otos poimitaan TKANNASSA olevista teollisuuden tai rakentamisen toimipaikoista. Käytännössä tämä tehdään yhdistämällä otokset 1 ja 2. Viides otos poimitaan kaikista toimipaikoista, joiden tilinpäätöstiedot ovat tiedossa. Käytännössä tämä tehdään yhdistämällä otokset 1, 2 ja 3. Kuudes otos poimitaan yksitoimipaikkaisista yrityksistä, joiden kaikki toimipaikkojen tilinpäätöstiedot ovat tiedossa. Käytännössä tämä tehdään yhdistämällä otokset 2 ja 3. Otosten poiminnassa käytettävät otanta-asetelmat ovat luvussa 5.2.
7. Laaditaan vaihtoehtoisia estimaatteja toimialakohtaisille tilinpäätöstiedoille käyttäen otoksia 1-6 (luku 5.2) ja estimaattoreita, jotka on mainittu luvussa 5.3. Malleissa riippuva muuttuja on jalostusarvo. Riippumattomat muuttujat ovat keskimääräinen henkilöstömäärä toimipaikalla, liikevaihto toimipaikalla ja toimipaikalle estimoidut palkat. Tutkimuksessa käytettävistä muuttujista kertovat luvut 3.2 ja 3.3 ja estimaattien laskennassa käytettävistä estimaattoreista luku 5.3. Yleisesti tilinpäätöksestä ja sen taustatekijöistä kertoo luku 2 sekä liite 3.
8. Valitaan toimialakohtaisten tilinpäätöstietojen estimaateista "parhaat". Tästä kerrotaan luvuissa 5 ja 6.1.
9. Imputoidaan malliluovuttaja-menetelmällä valittujen toimialakohtaisten tilinpäätöstietojen estimaattien pohjalta toimipaikkojen jalostusarvot. Toimipaikkojen jalostusarvojen estimaattien laadintatavoista kertoo luku 5. Laadittuja estimaatteja kuvataan luvuissa 6-8.
10. Tutkitaan laadittujen toimipaikkojen jalostusarvojen estimaattien hyvyyttä erilaisin diagnostisin tarkasteluin. Tämä tarkoittaa sitä, että verrataan toimipaikkojen tilinpäätöstietojen estimaatteja oikeisiin arvoihin eli lasketaan ennustusvirheitä silloin kun se on mahdollista. Diagnostisista tarkasteluista kertovat luvut 4, 5.6 ja 6-8.
11. Arvioidaan eri toimialojen merkittävyyttä Suomessa vuonna 2000 liikevaihdon, henkilöstön määrän, palkkojen ja laadittujen toimipaikkojen jalostusarvojen estimaattien avulla. Lisäksi arvioidaan toimialojen merkittävyyttä yritysten tilinpäätöstietojen avulla (ts. tällöin toimialalla tarkoitetaan yrityksen päätoimialaa). Tästä kertoo luku 6.3.
12. Esitetään suosituksia sille miten yritysten tilinpäätöstietoja voisi olla tarkoituksenmukaista toimipaikkaistaa ja mitkä ovat tutkimuksessa käytettyjen laskentamenetelmien rajoitteet. Näistä kertovat luvut 7 ja 8.

## 2 Tulosmuuttujan määrittely ja tulosmuuttujaan vaikuttavat tekijät

*Kaikki ne tiedot, joiden estimoimiseen tutkimuksessa laadittavaa menetelmää Tilastokeskuksessa tarvitaan, ovat toimipaikan laskennallisen käyttökateen eriä tai siitä laskettavia tunnuslukuja.* Tutkimuksen tulosmuuttujakin (jalostusarvo) on laskettavissa toimipaikan laskennallisesta käyttökateesta, jota vastaa yritystasolla tuloslaskelman liikevoittoa/tappiota edeltävä osa. Lisäksi tutkimuksessa estimoinnin vaiheessa käytettävät apumuuttujat ovat toimipaikan laskennallisen käyttökateen eriä (liikevaihto ja palkat) tai toimipaikan tilinpäätöksen liitetietoja (henkilöstön määrä). Siksi luku 2.1 on raportti tuloslaskelmasta ja toimipaikan laskennallisesta käyttökateesta. Ja koska tuloslaskelma on osa tilinpäätöstä, käsitellään luvussa 2.1 myös lyhyesti tilinpäätöstä kokonaisuudessaan. Luvussa 2.2 raportoidaan toimipaikan tilinpäätöstietoihin vaikuttavia taustatekijöitä. Luku 2 on siis taloustieteellistä taustatietoa tutkimuskohteesta. Tulosmuuttujan muodostamisen tekninen kuvaus on luvussa 2.2. Kululajipohjaisesta tuloslaskelmasta on kerrottu tarkemmin liitteessä 3.

### 2.1 Jalostusarvon määrittely, yrityksen tilinpäätös ja toimipaikan laskennallinen käyttökate

*Jalostusarvo (tutkimuksen tulosmuuttuja)* mittaa kansantaloudessa yrityksen tai toimipaikan osuutta (brutto) kansantuotteen kerryttäjänä. *Bruttokansantuotteella* tarkoitetaan kansantaloudessa vuoden aikana tuotettujen hyödykkeiden arvoa rahana. Kansantuote kertoo sen kuinka paljon tavaroita ja palveluksia kansantaloudessa tuotetaan. Laskelmissa tuotetun tuotteen/palvelun arvosta vähennetään tuotannontekijöiden ostot muilta tuottajilta. Se on toimipaikkatasolla laskennallisesta käyttökateesta laskettavissa oleva tunnusluku ja yritystasolla se on tuloslaskelman liikevoittoa/tappiota edeltävistä eristä laskettavissa oleva tunnusluku. (Taloustieto Oy, 1994, Autio, 1995).

*Jokaisella yrityksellä on kirjanpito- ja tilinpäätösvelvollisuus.* Kuitenkin kirjanpito- ja tilinpäätöskäytännöt ovat erilaiset eri puolilla maailmaa, vaikkakin ovat yhdenmukaistuneet EU:n sisällä (ns. tilinpäätösdirektiivien noudattamisen vuoksi). *Suomessa tilinpäätöksen tekoa säätelevät kirjanpitolaki ja EU:n tilinpäätösdirektiivit, neljäs ja seitsemäs yhteisöoikeudellinen direktiivi.* Suomessa yritysten tulee toimittaa tilinpäätöksensä patenttirekisteriin, jossa se on kaikkien nähtävissä. (Leppiniemi ja Leppiniemi, 2000, Kinnunen ym., 2000, 105).

*Tilinpäätös* kertoo siitä, mikä on yrityksen toiminnan tulos ja miten se on muodostunut sekä mitkä ovat yrityksen varat ja velat tilikauden lopussa tarkemmin sanottuna tilinpäätöspäivänä. Se kertoo siis liiketapahtumista. *Se koostuu tuloslaskelmasta, taseesta, tilintarkastus- ja toimintakertomuksesta sekä erilaisista liitetiedoista ja laaditaan kirjanpidosta joltain tilikaudelta.* (Leppiniemi ja Leppiniemi, 2000, Kinnunen ym., 2000, Taloustieto Oy, 1994, ja Autio, 1995).

**Liiketapahtumilla** tarkoitetaan tapahtumia, jotka on kirjattu tuloina, menoina ja rahoitustapahtumina kirjanpitoon. Elinkeinoverolain mukaan sellaiset tulot, jotka on merkitty tilinpäätöksessä, ovat veronalaisia. Niistä voidaan vähentää vain sellaiset menot, jotka on merkitty menoiksi tilinpäätöksessä. Täten se mitkä ovat yrityksen tulot ja menot määrittävät sen mitkä ovat yrityksen verotettavat tulot. **Rahoitustapahtumilla** tarkoitetaan pääoman lisäyksiä ja palautuksia sekä voitonjakoa eli korkoja, veroja ja osinkoja. (Kinnunen ym., 2000, 103, Leppiniemi ja Leppiniemi, 2000).

Kaikkien yritysten tulee kirjanpitolain mukaan käyttää **suoriteperusteista tilinpäätöstä**. Suoriteperusteisuus tarkoittaa sitä, että liiketapahtuma merkitään kirjanpitoon suoritteen luovutushetkellä. Jos tavara tai palvelus on luovutettu tai saatu yrityksen haltuun, katsotaan se kyseisen tilikauden tuloksi tai menoksi. Esimerkiksi, jos yritys on saanut jonkin tavaran tilikauden aikana, muttei sitä tilikauden aikana maksanut, merkitään se suoriteperusteisuuden vuoksi tilinpäätöksessä tilikauden menoksi. Jos yritys on luovuttanut jonkin tavaransa tai palvelunsa (ts. myynyt suoritteitaan tilikauden aikana), eikä tavaran/palvelun saaja ole vielä kyseisen tilikauden aikana sitä maksanut, merkitään tuosta tavarasta/palvelusta pyydetty hinta kyseisen tilikauden tuotoksi suoriteperusteisuuden vuoksi. Suoriteperusteisuus tilinpäätöksessä tarkoittaa sitä, että ennakoon saatuja tuloja ei merkitä kyseisen tilikauden tuloksi, koska tavaroita tai palveluksia, joista ennakko koostuu, ei ole vielä luovutettu ostajalle. Se tarkoittaa myös sitä, että ennakoon maksettuja suorituksia ei merkitä kyseisen tilikauden menoksi, koska tavaroita tai palveluksia, josta yritys on maksanut, ei ole vielä luovutettu yritykselle. (Laitinen ja Luotonen, 1996, Kinnunen ym., 2000, 76-78).

**Tilikausi** on se laskentakausi, jolta yrityksen tilinpäätös laaditaan. Sen pituus on yleensä 12 kuukautta, mutta se ei ole välttämättä kalenterivuosi. Yrityksen tilikauden pituuden tulee olla enintään 18 kuukautta, vähimmäispituutta ei ole säädetty. Tilastokeskuksen yritysten rakenteet -yksikkö mukauttaa saamani tiedon mukaan yrityksen toimittamat tilinpäätöstiedot vastaamaan 12 kuukauden tilikautta. Jos yritys lopettaa tai aloittaa toimintansa kalenterivuoden aikana, niin yritysten rakenteet -yksikkö tallentaa vain ne tiedot, jotka kyseiseltä vuodelta yritys on antanut. Jos yrityksen tilikausi on voitollinen ja kannattava, niin voitonjako tehdään juuri tilikaudelta (merkitystä siis omistajille). Tilikautta lyhyemmältä ajanjaksolta ei voittoa voi jakaa. (Kinnunen ym., 2000).

Tuloslaskelmassa tulot ja menot esitetään siten, että ensin on esitetty tulo ja sen jälkeen sen hankkimisesta johtuneet kulut. **Tuloslaskelman** lähtökohta on liikevaihto. Siitä vähennetään tuloslaskelmaa muodostettaessa sovitussa järjestyksessä kunkin menolajin asianomaiselle tilikaudelle kuuluva osuus. Tätä tilikaudelle jaksotettua menon osaa, joka on jo tuottanut tulonsa, kutsutaan kuluksi. Liikevaihtoon lisätään asianomaiselle tilikaudelle kuuluvat tulot, joita ei ole liikevaihtoa laskettaessa otettu huomioon sovitussa järjestyksessä. Miinusmerkki tarkoittaa kuluja ja plusmerkki tuloa. Sovittu järjestys, jolla tilikaudelle kuuluvia osuuksia kuluista vähennetään ja tuloista lisätään liikevaihtoon ilmenee kululajipohjaisen tuloslaskelman kaavasta. **Kululajipohjaisen tuloslaskelman koko kaava on liitteenä 3 ja se sisältää myös sen eri erien kuvaukset**. Kululajipohjaisen tuloslaskelman kaava liikevoittoon/tappioon saakka on taulukossa 2.1. (mm. Leppiniemi ja Leppiniemi, 2000, Yritystutkimusneuvottelukunta 1999).

*Tase* kertoo yrityksen taloudellisen aseman tilinpäätöspäivänä, ja siksi siihen vaikuttavat kyseessä olevan tilikauden toiminnan muutosten lisäksi edelliset tilikaudet. Taseeseen vaikuttavat siis kaikki yrityksen siihen astisen toiminta-ajan liiketapahtumat eli koko yrityksen toiminta-aika suoraan ja taseen eri tilikausien erät eivät ole laskettavissa yhteen. Lisäksi tase kuvaa yrityksen taloudellista tilaa tilinpäätöksen tekohetkellä, ei koko tilikauden tai edeltävien tilikausien ajalta. Yrityksen taseen laatimista on kuvattu mm. kirjassa Leppiniemi ja Leppiniemi (2000) ja se sisältää myös taseen eri erien kuvaukset.

Varsinaisen tilinpäätöksen lisäksi yrityksen on ilmoitettava verohallinnolle ja Tilastokeskukselle erilaisia muita tietoja toiminnastaan *liitetietoina*. Yrityksen pitää mm. ilmoittaa liikevaihto toimipaikoittain ja ulkomaan toimintojen osalta markkina-alueittain, jos nämä alueet eroavat toiminnan laadun suhteen olennaisesti toisistaan eivätkä käy ilmi muista tiedoista. Yrityksen pitää myös ilmoittaa tilikauteen kohdistetut palkat sekä palkkaan rinnastettavien luontoisetujen (työsuhdeasunto, matkapuhelin, työsuhdeauto yms.) yhteenlaskettu raha-arvo sekä eläkekulut, muut henkilösivukulut sekä monia muita tietoja. Näillä liitetiedoilla on suuri merkitys yrityksen taloudellisen tilanteen arvioinnissa. (mm. Leppiniemi ja Leppiniemi, 2000, Kinnunen ym. , 2000).

Tilinpäätöksestä ja sen liitetiedoista koostuu se informaatio, joka leviää yrityksen ulkopuolelle ja minkä perusteella ulkopuoliset luovat yrityksestä kuvaa (toki muutkin asiat tähän vaikuttavat, esim. hyödykkeet, joita yritys myy). Tilinpäätös antaa tietoja yrityksen tilasta mm. yrityksen omistajille, luotonantajille, työntekijöille ja verottajalle sekä muille sidosryhmille. Tilinpäätös on heille päätöksenteon, edunvalvonnan ja yrityksen valvonnan väline. Tilinpäätös mm. kertoo jakokelpoisen voiton eli sen mistä yritystä verotetaan ja mistä omistajille jaetaan osinkoa. Tilinpäätöksen avulla voidaan tehdä myös " säännönmukaisesti kohtuullisen vertailukelpoisia ja luotettavia yrityksen luottokelpoisuutta koskevia arvioita" (Leppiniemi ja Leppiniemi 2000, 15). Näin yrityksen ns. luottokelpoisuus voidaan määrittellä tilinpäätöksestä, vaikka ei yritystä tunnettaisikaan. Verohallinto käyttää Suomessa tilinpäätöstietoja aputietoina. Leppiniemi ja Leppiniemi (2000, 11) määrittelee tilinpäätöksen siten, että " Se sisältää lakisääteisesti kurinalaisen yhteenvedon tilikauden tuloksesta ja yrityksen saavuttamasta asemasta." (Leppiniemi ja Leppiniemi 2000).

Tilinpäätöksen tulkintaa vaikeuttaa se, että tuloslaskelman ja taseen erät ovat abstrakteja. Tällä tarkoitetaan sitä, että ei tiedetä mistä erät todellisuudessa koostuvat ja niitä ei voida tarkistaa. Koska varastossa olevien valmiiden ja keskeneräisten tavaroiden arvo määritellään sopimukseen (mm. tuotteilla valittuihin hintoihin) nojautuen, sitä ei suoraan varastossa olevista tavaroista voi päätellä. Tuotteille valitut hinnat vaikuttavat mm. tilikauden voittoon, omaan pääomaan sekä tuloslaskelman ja taseen rakenteeseen. (Leppiniemi ja Leppiniemi, 2000).

Jokaisen yrityksen tulee toimittaa Suomessa tilinpäätöksensä verohallinnolle. Verohallinnolta Tilastokeskuksen yritysten rakenteet -yksikkö saa pyynnöstä yritysten tilinpäätökset. Lisäksi yritysten rakenteet -yksikkö saa yritysten tietoja suoraan niiltä yrityksiltä, jotka kuuluvat suoraan kyselyyn. Tilastokeskus käyttää yritysten tilinpäätöstietoja Suomen yritysten taloudellisen tilanteen analysoimiseen sekä yritysten rakenteiden tutkimiseen (esim. yritysten toimialat, kokoluokat ja toimipaikat). Yrityksiltä saatuja tietoja käytetään myös apuna Suomen



taloudellisen tilanteen tutkimisessa (esim. bruttokansantuotetta laskettaessa osa tiedoista koostuu yritysten jalostusarvoista) sekä ihmisten elinolojen analysoimisessa.

Yrityksillä on paljon harkintavalttaa tilinpäätöksiä laatiessaan, koska EU:n ns. tilinpäätösdirektiivin mukaan ”koko tilinpäätöksen tulee antaa oikea ja riittävä kuva sekä kirjanpitovelvollisen tuloksesta että taloudellisesta asemasta” (Leppiniemi ja Leppiniemi 2000, 97). Tämä tarkoittaa sitä, että tilinpäätöksen on annettava oikea ja riittävä kuva yrityksen tai muun kirjanpitovelvollisen varoista ja veloista, taloudellisesta asemasta sekä voitosta tai tappiosta ja se varmistetaan liitetietojen avulla. Tilinpäätöksessä tehdyistä valinnoista on ilmoitettava liitetiedoissa ja noudattava niissä jatkuvuuden periaatetta (valintojen on pysyttävä ennallaan mikäli mahdollista ja niissä tapahtuvista muutoksista on kerrottava). Tilinpäätöksen laatimisessa tehtyjä valintoja sanotaan ns. ns. tilinpäätöksen laadintaperiaatteiksi. (Leppiniemi ja Leppiniemi, 2000, 97, Alhola ym., 2000, Kinnunen ym., 2000).

***Tilinpäätöksen laadintaperiaatteita ovat tilinpäätöksen tekemisessä tehdyt valinnat, jotka koskevat tilinpäätöksen laajuutta, jaksottamista ja arvostusta.*** Tilinpäätöksen laajuutta koskevat ratkaisut määrittävät sen mitkä asiat esitetään varsinaisessa tilinpäätöksessä ja mitkä liitetiedoissa. Jaksottamista koskevat ratkaisut ovat sen päättämistä mitkä tulot ja kulut ovat tilikaudelle kuuluvia (esim. kuinka monelle tilikaudelle investointimenot poistoina jaksotetaan tai millä tavoin ulkomaanvaluuttana tapahtuneet liiketapahtumat ja kurssierot kirjataan). Arvostamista koskevat ratkaisut liittyvät siihen mitkä hinnat yrityksen tuottamille suoritteille ja yrityksen hallussa olevalle omaisuudelle merkitään esim. kirjataanko yrityksen hallussa oleva omaisuus alkuperäisen hankintamenon vai tätä suuremman todennäköisen luovutushinnan mukaisena. Tilinpäätöksen laadintaperiaatteita voidaan pitää myös ns. yrityskohtaisina tekijöinä. (Kinnunen ym., 2000,105).

Tilinpäätöksestä voidaan laskea erilaisia tunnuslukuja, kuten tuotannon bruttoarvo, jalostusarvo ja käyttökate. Tutkimuksen tulosuuttuja, jalostusarvo, on siis yksi tilinpäätöksestä laskettava tunnusluku. Näitä ja tilinpäätöksestä suoraan saatavia tietoja voidaan suhteuttaa esimerkiksi yrityksen liikevaihtoon tai henkilöstömäärään. Täten voidaan vertailla eri yritysten taloudellisia tilanteita keskenään, vaikka kyseiset yritykset olisivatkin eri kokoisia. Yrityksen omasta kehityksestä (esim. siitä onko yrityksen taloudellinen tilanne parantunut edellisestä tilikaudesta) tehdään päätelmiä indeksoimalla. Tätä on esim. tilikauden voiton jakaminen liikevaihdolla. (Leppiniemi ja Leppiniemi, 2000).

Yrityksen toimipaikoille voidaan tehdä oma tilinpäätös, jos jokainen yrityksen toimipaikka on pitänyt omaa kirjanpitoaan. Toimipaikan tilinpäätös koostuu toimipaikan laskennallisesta käyttökatteesta, toimintakertomuksesta ja liitetiedoista. Laskennallinen käyttökate poikkeaa yrityksen tilinpäätöksestä rakenteeltaan, koska se ei sisällä poistoja, mutta sisältää toimipaikan toimitukset muiden yritysten toimipaikkojen käyttöön. Toimipaikan tilinpäätös ei sisällä myöskään tasetta ja on muutenkin suppeampi. Kotimaassa toimivien toimipaikkojen tilinpäätöstietoihin ei sisälly myöskään yrityksen toiminta ulkomailla. ***Tutkimuksen tulosuuttuja ja apumuuttajat ovat toimipaikan laskennallisen käyttökateen eriä ja niistä laskettavia tunnuslukuja.*** Toimipaikan laskennallisen käyttökateen kaava on taulukossa 2.2.

**Taulukko 2.1.** Kululajipohjaisen tuloslaskelman kaava liikevoittoon/tappioon saakka, jota mm. verohallinto ja Tilastokeskus käyttävät. (Leppiniemi ja Leppiniemi, 2000, Yritystutkimusneuvottelukunta, 1999).

|   |     |       |
|---|-----|-------|
| <b>LIKEVAIHTO</b>   | +   | _____ |
| <b>VALMIIDEN JA KESKENERÄISTEN TUOTTEIDEN VARASTOJEN MUUTOS</b> | +/- | _____ |
| <b>VALMISTUS OMAAN KÄYTTÖÖN</b>                                 | +   | _____ |
| <b>LIIKETOIMINNAN MUUT TUOTOT</b>                               | +   | _____ |
| <b>MATERIAALIT JA PALVELUT:</b>                                 | -   |       |
| Ostot tilikauden aikana   | -   | _____ |
| Varastojen muutos   | +/- | _____ |
| Ulkopuoliset palvelut   | -   | _____ |
| <b>HENKILÖSTÖKULUT:</b>   | -   |       |
| Palkat ja palkkiot  | -   | _____ |
| <b>henkilösivukulut:</b>  | -   |       |
| eläkekulut  | -   | _____ |
| muut henkilösivukulut   | -   | _____ |
| <b>POISTOT JA ARVONALENTAMISET:</b>                             | -   |       |
| Suunnitelman mukaiset poistot                                   | -   | _____ |
| Arvon alentamiset pysyvien vastaavien hyödykkeistä              | -   | _____ |
| Vaihtuvien vastaavien poikkeukselliset arvonalentamiset         | -   | _____ |
| <b>LIIKETOIMINNAN MUUT KULUT</b>                                | -   | _____ |
| <b>LIIKEVOITTO/TAPPIO</b>                                       | +/- |       |

**Taulukko 2.2.** Toimipaikan laskennallisen käyttökatteen kaava Tilastokeskuksen esitystavan mukaan.

|   |     |       |
|---|-----|-------|
| <b>1.TOIMIPAIKAN LIKEVAIHTO</b>                                   | +   | _____ |
| <b>2.TOIMITUKSET YRITYKSEN MUIDEN TOIMIPAIKKOJEN KÄYTTÖÖN</b>     | +   | _____ |
| <b>3.VALMIIDEN JA KESKENERÄISTEN TUOTTEIDEN VARASTOJEN MUUTOS</b> | +/- | _____ |
| <b>4.VALMISTUS OMAAN KÄYTTÖÖN</b>                                 | -   | _____ |
| <b>5.LIIKETOIMINNAN MUUT TUOTOT</b>                               | +   | _____ |
| <b>6. MATERIAALIT JA PALVELUT:</b>                                |     |       |
| <b>AINEET TARVIKKEET JA TAVARAT:</b>                              |     |       |
| Ostot tilikauden aikana   | -   | _____ |
| Ostot yrityksen multa toimipaikoilta                              | -   | _____ |
| Varastojen muutos   | +/- | _____ |
| ULKOPUOLISET PALVELUT   | -   | _____ |
| <b>Materiaalit ja palvelut yhteensä</b>                           | +/- | _____ |
| <b>7. HENKILÖSTÖKULUT:</b>  |     |       |
| Palkat ja palkkiot  | -   | _____ |
| <b>HENKILÖSIVUKULUT:</b>  |     |       |
| Eläkekulut  | -   | _____ |
| Muut henkilösivukulut   | -   | _____ |
| <b>Henkilöstökulut yhteensä</b>                                   | -   | _____ |
| <b>8. LIIKETOIMINNAN MUUT KULUT</b>                               | -   | _____ |
| <b>9. LASKENNALLINEN KÄYTTÖKATE</b>                               | +/- | _____ |

Toimipaikkojen tilinpäätöstietoja yhteen laskemalla ei aina saada yrityksen tilinpäätöstietoja, koska yrityksen tilinpäätöstiedot sisältävät erilliskustannusten eli välittömien kustannusten lisäksi yhteiskustannukset eli välilliset kustannukset sekä yrityksen ulkomaisen toiminnan. Erilliskustannuksia ovat menot, jotka voidaan katsoa aiheutuneen tietyistä laskentakohteista ja niiden kohdalla tiedetään minkä toimipaikan tulonhankintaan ne osallistuvat. Yhteiskustannuksia ovat Kinnusen ym. mukaan (2000,253) ”ne kustannukset, joita ei voida aiheuttamisperusteella kohdistaa laskentakohteelle.” Ne jaetaan toimipaikoille harkinnanvaraisesti. Tällaisia kuluja ovat mm. hallintokulut. (Kinnunen ym. 2000, 253-254).

## 2.2 Toimipaikan tilinpäätöstietoihin vaikuttavat tekijät

Toimipaikan tilinpäätöstietoihin vaikuttavat monet **taustatekijät**. Niitä ovat kuuden eri tason tekijät eli **yleistaloudelliset, toimialakohtaiset, alueelliset, toimiala- ja aluekohtaiset, yrityskohtaiset ja toimipaikkakohtaiset tekijät**. **Yleistaloudelliset tekijät** ovat kaikkiin samassa kansantaloudessa toimiviin toimipaikkoihin ja yrityksiin vaikuttavia tekijöitä. Ne vaikuttavat suoraan tai alempitasoisten tekijöiden välityksellä epäsuorasti. Yleistaloudellisia tekijöitä ovat taloudelliset suhdanteet, keskimääräinen yritys- ja toimipaikkakoko kansantaloudessa, kansantalouden toimialarakenteen ja talousyksiköiden kulutuskäyttäytyminen. (Maliranta, 1996, 89-95, Kinnunen ym., 2000, 103-109).

Muita yleistaloudellisia tekijöitä ovat mm. julkisen vallan talouspoliittiset toimenpiteet, yritysveroitus, kaikki yritysten toimintaa säätelevä lainsäädäntö (esim. kirjalaki) ja tulopoliittiset ratkaisut (TUPO). Myös EU:n lainsäädäntö ja talouspoliittiset toimenpiteet vaikuttavat yritysten ja toimipaikkojen toimintaan. Esimerkkinä yleistaloudellisten tekijöiden suorasta vaikutuksesta toimipaikan toimintaan voisi olla TUPO:n vaikutukset henkilöstökuluihin, koska toimipaikan täytyy maksaa vähintään tuloehtosopimuksen määrittämiä palkkoja. (Kinnunen ym., 2000, 103-109).

**Toimialakohtaiset tekijät** ovat tekijöitä, jotka vaikuttavat samansuuntaisesti kaikkien samalla toimialalla toimivien toimipaikkojen ja yritysten toimintaan. Toimialakohtaisten tekijöiden vaikutus on suoraa tai ne voivat vaikuttaa epäsuorasti yritys- ja toimipaikkakohtaisten tekijöiden kautta. Toimialakohtaisia tekijöitä ovat mm. toimialan taloudellinen kehitys ja toimialalla tuotettavien tuotteiden kysynnän ja hintojen kehitys, toimialan työvoima- ja pääomavaltaisuuden aste sekä toimialan yleinen tulo- ja menorakenne (mm. muuttuvien ja kiinteiden menojen osuudet). (Kinnunen ym., 2000, 103-109).

Toimipaikkojen tilinpäätöstietoihin vaikuttaa myös toimipaikan sijainti alueiden välisen työnjaon (alueen asema alueiden välisessä työnajossa vaikuttaa alueen yleiseen taloudelliseen asemaan taloudessa), alueen sisäisen työnjaon (toimipaikan asema alueen sisällä taloudellisessa kilpailussa) ja paikallisen konvergenssin kautta. Ne ovat ns. **aluekohtaisia tai alueellisia tekijöitä**. Alueelliset tekijät ovat sellaisia, jotka vaikuttavat samansuuntaisesti kaikkien samalla alueella toimivien toimipaikkojen ja yritysten toimintaan suoraan tai epäsuorasti yritys- tai toimipaikkatason tekijöiden kautta. Jonkun kansantalouden maantieteellinen sijainti vaikuttaa mm. siihen

millä toimialoilla kansantalous toimii, valtioiden väliseen työnjakoon ja paikalliseen konvergenssiin. (Kinnunen ym., 2000, 103-109, Maliranta, 1996, 89-95).

**Työnjaolla** tarkoitetaan tuotannon jakamista erilaisiin osatehtäviin. Työnjaon periaatteena on, että sen avulla voidaan käyttää hyväksi maapallon ja kunkin valtion eri alueiden erilaisia tuotantomahdollisuuksia. Sitä voidaan tarkastella monelta eri kannalta: sukupuolten, valtioiden, valtion eri alueiden ja yritysten välisenä. Työnjaon periaatteena on, että kukin valtio, valtion sisäinen alue, yritys ja ihminenkin pyrkii valmistamaan tuotteita, joihin sillä on parhaat edellytykset. Tästä on pohjimmiltaan kyse silloinkin kun ihminen erikoistuu johonkin ammattiin. Ammattia valitessaan ihminen valitsee itselleen sellaisen ammatin, jossa tuotetaan hyödykkeitä, joiden tuottamiseen hänellä on parhaat edellytykset (lahjakkuus ja motivaatio yms. psyykkiset, fyysiset ja sosiaaliset ominaisuudet). Työnjakoa tapahtuu myös yrityksen sisällä. (Autio, 1995).

Työnjaosta on seurannut kansainvälinen yhteistyö, erilaisten elinkeinojen muodostuminen, yritysten erikoistuminen erilaisten tuotteiden valmistamiseen ja ihmisten erikoistuminen erilaisiin ammatteihin. Yrityksen kannalta työnjako on erikoistumista. Erikoistuminen tuo mm. seuraavia etuja: hyödykkeitä valmistuu aikayksikössä enemmän ts. tuottavuus kasvaa, hyödykkeet halpenevat, hyödykkeiden laatu paranee (ammattitaitoisemmat valmistajat), ammattikoulutus ei vie kohtuuttoman paljon aikaa ja ihmiset viihtyvät työssään, koska heillä on vankka ammattitaito. Lisäksi ihmisten elintason nousu käy mahdolliseksi, koska tuotanto kasvaa työnjaon seurauksena. (Autio, 1995).

**Alueiden välinen työnjako** tarkoittaa eri alueiden erikoistumista eri tehtäviin. Se on yksi yhteiskunnallisen työnjaon muoto ja ilmenee alueiden erilaisina toimiala- ja elinkeinorakenteina ja se vaikuttaa paljon ihmisten elinolojen alueellisiin eroihin. Alueiden välisen työnjaon takia syntyy erilaisia elinympäristöjä, joissa hyvinvoinnin toteutumisen edellytykset ja esteet ovat erilaiset. Alueelliset erot ovat paljolti seurausta taloudellisten rakenteiden alueellisista eroista (alueilla toimivat organisaatiot) sekä aikaisemmasta kehityksestä. Myös luonnonolot, kulttuuriset erot sekä elinkeino- ja toimialarakenteissa esiintyvät erot alueiden välillä vaikuttavat alueiden väliseen työnjakoon. (Virkkala, 1987, 3).

Alueiden välistä työnjakoa ja eri alueiden erilaisia elinkeino- ja toimialarakenteita pyritään selittämään ns. kansainvälisen kaupan teorian kautta. Teorian mukaan eri alueet pyrkivät erikoistumaan sellaisiin tuotteisiin, joiden suhteelliset tuotantokustannukset ovat pienimmät. Siihen mitkä alueet ovat yritykselle kannattavia sijaintipaikkoja, vaikuttavat erilaiset sijaintiedellytykset, kuten mm. alueiden luonnonvarat (raaka-aineiden riittävyys), kuljetuskustannukset, työvoiman määrä ja ammattitaito sekä työvoimakustannukset, väestö-, elinkeino- ja toimialarakenne, asiakaskunnan suuruus, laajentumismahdollisuudet ja muiden yritysten toiminta alueilla. Laajemmin ajateltuna yritysten sijoittumiseen alueelle vaikuttavat kannattavuus ja yhteiskunta- ja kansantaloudelliset tekijät. (Virkkala, 1987, 3).

Alueiden välinen työnjako on jatkuvassa muutoksentilassa. Yritykset reagoivat ja vaikuttavat alueiden välisiin eroihin sijaintiedellytyksissä lähinnä organisaationsa ja sen muutosten kautta. Alueellisen toimipaikkaverkon muodostaminen on osa yrityksen tapaa reagoida ja vaikuttaa alueellisiin eroihin. Se kertoo alueiden välisen

työnjaon vaikutuksesta yritykseen ja yrityksen koon kasvusta. Yritykset vaikuttavat organisaationsa alueiden väliseen työnjakoon. Toisaalta yritysten alueellisten organisaatioiden piirteet ovat sidoksissa alueiden väliseen työnjakoon. Alueellisen toimipaikkaverkon kautta yritykset voivat jakaa tuotannon eri alueiden kesken eli harjoittaa yrityksen sisäistä alueiden välistä työnjakoa. (Virkkala, 1987, 15-16, Autio, 1995).

Yritysten vaikutuksesta alueellisiin eroihin ollaan aluetieteen parissa hieman eri mieltä. Yritysmaantieteen mukaan yritykset ovat keskeisiä alueellisiin eroihin vaikuttavia yhteiskunnallisia toimijoita ja alueellinen kehitys on selitettävissä yritysten piirteistä ja sijaintipäätöksistä käsin. Toisaalta ns. teollisuuden klassisen sijaintiteorian mukaan yritykset eivät vaikuta alueellisiin eroihin, vaan reagoivat niihin. Yrityksen pyrkimyksenä on löytää tuotelleen ns. optimaalinen sijaintipaikka. Tämä sijaintipaikka on sellainen, jossa tuotantokustannukset ovat mahdollisimman pienet ja tuotot sekä voitot maksimaaliset. Optimaalinen sijaintipaikka valitaan sijaintitekijöissä ilmenevien erojen perusteella. (Virkkala, 1987, 18-21).

Yrityksen johto päättää muutoksista alueellisessa toimipaikkaverkossa ns. eksplisiittisten ja implisiittisten sijaintipäätösten kautta. Eksplisiittiset sijaintipäätökset ovat päätöksiä, joissa päätetään varsinaisesti sijainnista ja jotka ovat siis alueellisista eroista johtuvia. Ne ovat ns. varsinaisia sijaintipäätöksiä. Implisiittiset sijaintipäätökset taas ovat sellaisia päätöksiä, jotka eivät tietoisesti kohdistu johonkin alueeseen tai toimipaikkojen sijaintiin, mutta kuitenkin vaikuttavat alueellisesti. Tällaisia päätöksiä ovat mm. yrityskaupat. Implisiittisissä päätöksissä toimipaikan sijainti ei ole tärkeintä, vaan sen liittyminen yrityksen strategiaan. Varsinainen sijaintiin liittyvä päätöksenteko on siis vain osa siitä yrityksessä tapahtuvassa päätöksenteosta, jolla on alueellisia vaikutuksia. (Virkkala, 1987, 35-37).

Monitoimipaikkainen yritys voi vaikuttaa alueiden väliseen työnjakoon eniten, kun se on kooltaan suuri eli mitä enemmän se työllistää ja mitä suuremmasta osasta tuotantoa se vastaa. Lisäksi mitä laajemmalla alueella yritys toimii, mitä useampia alueita sen toimipaikkaverkko sitoo toisiinsa, mitä syvempää yrityksen sisäinen alueiden välinen työnjako on ja mitä enemmän tämä työnjako noudattaa alueellisia kehittyneisyyseroja sekä kaupunkijärjestelmän hierarkiaa, sitä enemmän yritys voi vaikuttaa alueiden välisiin eroihin. Tämä tarkoittaa sitä, että suuret monitoimipaikkaiset yritykset, jotka toimivat monella toimialalla ja alueella, voivat vaikuttaa paljon alueiden välisiin eroihin. Erityisen suuret vaikuttamismahdollisuudet ovat suurilla kansainvälisillä yrityksillä, kuten Nokialla (ne voivat vaikuttaa valtioiden väliseen työnjakoon). (Virkkala, 1987, 7).

Julkinen valta (valtio, kunnat ja kuntayhtymät sekä seurakunnat) pyrkii vaikuttamaan yritysten sijoittumiseen aluepolitiikan kautta, koska yritykset vaikuttavat alueellisten erojen kehittymiseen vahvasti. **Aluepolitiikka** pyrkii edistämään sitä, että koko valtio kehittyisi mahdollisimman tasapuolisesti eli ehkäisemään tai vähentämään alueellisia eroja. Aluepolitiikkaa suoritetaan erityisesti tukemalla yritysten toimintaa sellaisilla alueilla, joissa taloudellista ja muunlaista kehitystä estävät olennaisesti vaikeammat olosuhteet kuin muualla maassa ja jotka ovat muista maan alueista taloudellisesti jälkeenjääneitä. Näitä alueita sanotaan kehitysalueiksi. Finnvera muodostui 1.1.1999 kun KERA ja Takuukeskus yhdistyivät. Sen yksi tehtävä on tukea kehitysalueilla sijaitsevien, niille perustettavien ja niihin siirtyvien yritysten toimintaa. Finnvera myöntää edellä mainituille yrityksille lainoja, ostaa niiden osakkeita, avustaa niitä markkinoinnin tehostamisessa ja tuotteiden kehittämistyössä. Lisäksi

Finnvera tekee kehitysalueiden yritystoiminnan edellytyksiä selvittävää tutkimusta ja aloitteita elinkeinotoiminnan kehittämiseksi kehitysalueilla. Finnvera on valtioenemmistöinen erityisluottolaitos. (Autio, 1995, 83-86, Finnvera, 2002).

**Paikallisella konvergenssilla** tarkoitetaan maantieteellisesti toisiaan lähellä olevien maiden elintasojen pyrkimystä lähentyä toisiaan. Paikallisesta konvergenssista johtuen Suomen elintasaan ja sitä kautta yrityksiin ja toimipaikkoihin vaikuttaa se millainen elintaso ja sen kehitys on Suomen lähialueilla kauempana olevia maita enemmän. Paikallinen konvergenssi johtuu ainakin osaltaan siitä, että lähellä toisiaan olevilla mailla on samantyyppiset luonnonolot, lainsäädäntö sekä talouspolitiikka. Paikallisen konvergenssin johdosta maantieteellisesti toisiaan lähellä olevat kansantaloudet erikoistuvat usein samantyyppiseen tuotantotoimintaan, ts. harjoittavat samojen toimialojen toimintoja. Siksi lähellä toisiaan olevat valtiot yleensä kilpailevat keskenään. (Maliranta, 1996, 89-95).

**Toimiala- ja aluekohtaiset tekijät** vaikuttavat samansuuntaisesti kaikkien samalla toimialalla ja alueella toimivien yritysten ja toimipaikkojen toimintaan. Niitä ovat alueiden välinen työnjako ja paikallinen konvergenssi (mm. tietyllä alueella ja toimialalla toimivien toimipaikkojen ja yritysten taloudellinen kehitys), tulo- ja menorakenne, tuotettujen tuotteiden kysyntä- ja hintakehitys sekä työvoima- ja pääomavaltaisuuden asteet ja asema alueen sisäisessä työnjaossa. Sijainti suosii joitakin toimialoja ja vaikuttaa siihen mihin maa vie tai mistä tuo tuotannon tekijöitä. Toimiala- ja aluekohtaiset tekijät ovat siis eri asia kuin toimialakohtaiset tai aluekohtaiset tekijät. (Kinnunen ym., 2000, 103-109, Maliranta, 1996, 89-95).

**Yrityskohtaiset tekijät** ovat tekijöitä, jotka vaikuttavat samansuuntaisesti kaikkien yrityksen toimipaikkojen toimintaan suoraan tai toimipaikkakohtaisten tekijöiden kautta. Ne koskevat siis vain yksittäistä yritystä ja sen toimipaikkoja. Yrityskohtaisia tekijöitä ovat yrityksen johdon päätökset ja toimenpiteet juoksevaan liiketoimintaan, investointeihin ja rahoitukseen liittyen. Yrityskohtaisia tekijöitä ovat myös yrityksessä käytettävät **tilinpäätöksen laadintaperiaatteet** (laajuus-, arvostus- ja jaksotusratkaisut). Ne koskevat kaikkia yrityksen toimipaikkoja. Toimipaikan tilinpäätöstä koskevat ratkaisut voivat olla myös toimipaikkakohtaisia ja tällöin tilinpäätöksen laadintaperiaatteet ovat toimipaikkakohtaisia tekijöitä. Tilinpäätöksen laadintaperiaatteista luvussa 2.1. Tilinpäätöksestä on kerrottu luvun 2.1 lisäksi liitteessä 3. Yrityskohtaisia tekijöitä ovat myös yrityksen koko ja laji (se onko yritys yksi- vai monitoimipaikkainen). (Kinnunen ym., 2000, 103-109).

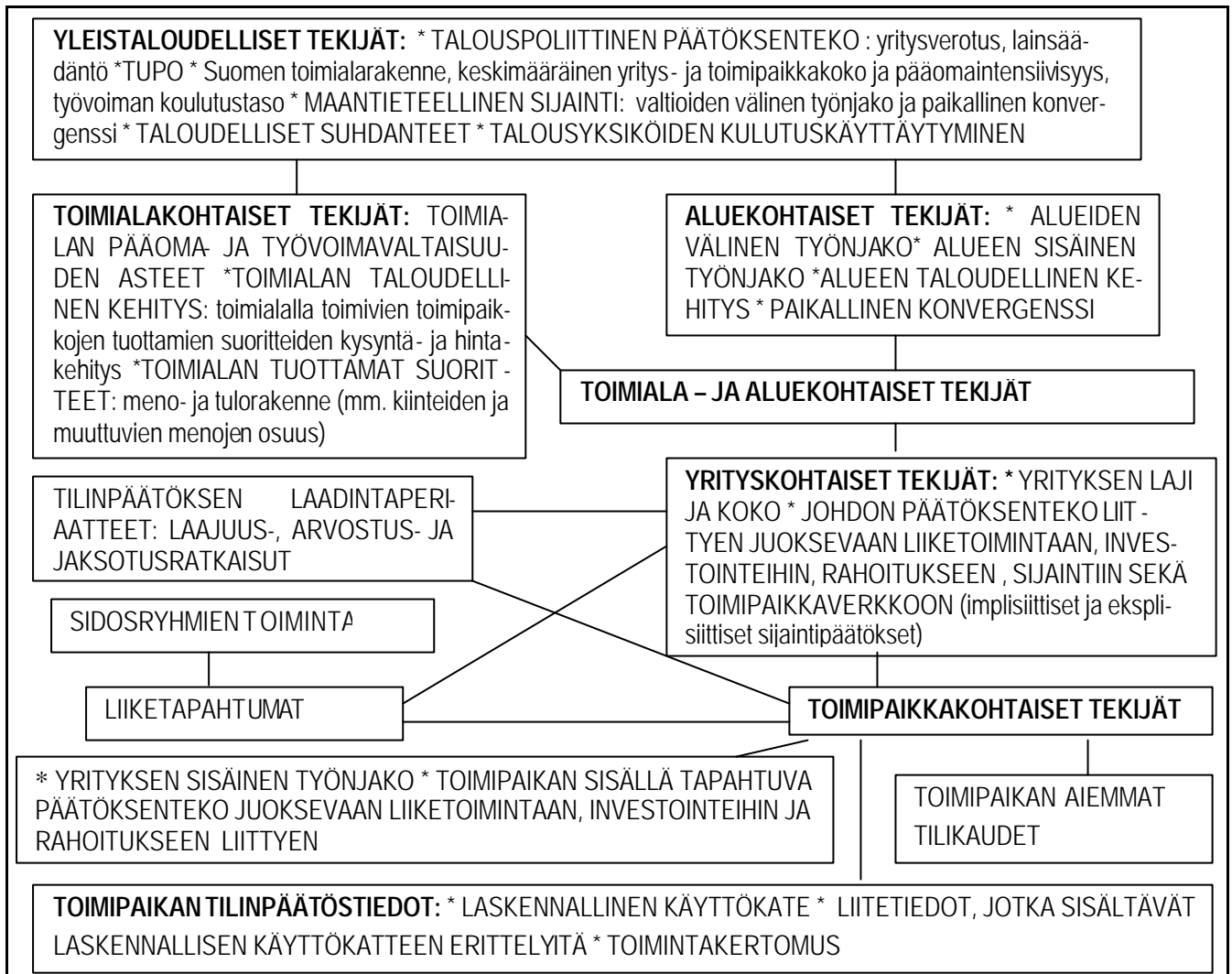
Yrityksen **kokoa** voidaan arvioida keskimääräisen henkilöstön määrän ja liikevaihdon avulla (ne kertovat myös toiminnan laajuudesta). Yrityksiä luokitellaankin niiden koon mukaan mikroyrityksiin, pieniin, keskisuuriin ja suuriin yrityksiin. EU:n komissio on antanut vuonna 1996 suosituksen pienten ja keskisuurten yritysten määritelmästä. Tilastokeskus (Tilastokeskus, 2002e) käyttää tämän suosituksen mukaisia määritelmiä. **Mikroyritys** on yritys, jonka palveluksessa on alle 10 työntekijää ja jonka vuosiliikevaihto tai taseen loppusumma on alle 2 miljoonaa euroa. **Pieni yritys** on riippumaton yritys, jonka palveluksessa on yli 10-49 työntekijää ja jonka vuosiliikevaihto on 2-7 miljoonaa euroa tai taseen loppusumma on 2-5 miljoonaa euroa. Yritys on riippumaton, jos mikään suuryritys tai niiden joukko (konserni) ei omista sen pääomasta tai äänivaltaisista osakkeista yli 25 prosenttia. **Keskisuuri yritys** on riippumaton yritys, jolla on 50-249 työntekijää ja jonka vuosiliikevaihto on 7-40

miljoonaa euroa tai taseen loppusumma on 5-27 miljoonaa euroa. **Suuri yritys** on puolestaan riippumaton yritys, jolla on ainakin 250 työntekijää ja jonka vuosiliikevaihto on yli 40 miljoonaa euroa tai taseen loppusumma on yli 27 miljoonaa euroa. **PK-yritys on pieni tai keskisuuri yritys** eli riippumaton yritys, jolla on alle 250 työntekijää ja jonka vuosiliikevaihto on enintään 40 miljoonaa euroa tai taseen loppusumma on enintään 27 miljoonaa euroa. Suurin osa Suomen liikeyrityksistä oli vuonna 2000 pieniä liikeyrityksiä. Niitä oli 216 183. Siksi myös PK-yritysten osuus liikeyrityksistä oli suurempi kuin suurten liikeyritysten. Kaikkein vähiten Suomessa oli kuitenkin keskisuuria liikeyrityksiä (2514). Suuria liikeyrityksiä oli Suomessa vuonna 2000 4120. (Tilastokeskus 2002 e, 46, Euroopan yhteisöjen komissio 2002, 6).

Se onko yritys yksi- vai monitoimipaikkainen (eli **yrityksen laji**) vaikuttaa tässä tutkimuksessa saatujen tulosten mukaan (raportoitu yksityiskohtaisemmin luvussa 3.2) sen toimipaikkojen tilinpäätöstietoihin siten, että monitoimipaikkaisten yritysten toimipaikkojen tilinpäätöstiedot ovat yksitoimipaikkaisten yritysten tilinpäätöstietoja suuremmat. Monitoimipaikkaiset yritykset ja myös niiden toimipaikat ovat toiminnan laajuudeltaan, merkittävyydeltään ja kooltaan suurempia kuin yksitoimipaikkaiset yritykset.

Toimipaikkakohtaiset tekijät ovat tekijöitä, jotka vaikuttavat vain yksittäiseen toimipaikkaan. Niitä ovat toimipaikan asema yrityksen sisäisessä työnjaossa, toimipaikan koko, toimipaikan sisällä tehty päätöksenteko liittyen juoksevaan liiketoimintaan, investointeihin ja rahoitukseen, toimipaikan tilinpäätöksen laadintaperiaatteet ja liiketapahtumat, joihin vaikuttavat toimipaikan sidosryhmät. Toimipaikan **työvoima- tai pääomavaltaisuus** vaikuttaa sen tilinpäätöstietoihin. Henkilöstökulut ovat työvoimavaltaisilla toimipaikoilla suuremmat kuin pääomavaltaisilla toimipaikoilla ja jalostusarvo on suurempi pääomavaltaisilla toimipaikoilla kuin työvoimavaltaisilla toimipaikoilla.

Kuvio 2.1 kuvaa toimipaikan tilinpäätöstietoihin vaikuttavia taustatekijöitä. Se on laadittu kirjallisuuden pohjalta (Autio, 1995, Kinnunen ym., 2000, Maliranta, 1996 ja Virkkala, 1987), mutta sen muotoilu ja esille otettujen asioiden valinnat ovat kirjoittajan tekemiä.



Kuvio 2.1. Toimipaikan tilinpäätöstietojen muodostuminen.

### 3 Tutkimusaineisto

Luku 3 kuvaa tietokantoja ja rekisterejä, joiden pohjalta tutkimusaineisto laadittiin. Näitä ovat yritys- ja toimipaikkarekisteri sekä rakennetilastojen kyselykanta vuodelta 2000. Luvussa 3.1 raportoidaan niiden muodostamista ja yhdistämisessä esiintyvistä kehittämistarpeista. Luvuissa 3.2 ja 3.3 määritellään tutkimuksessa käytetyt muuttujat sekä niiden tietolähteet. Luvussa 3 kerrotaan myös siitä miten tutkimusaineistoa on käytetty.

#### 3.1 Yritys- ja toimipaikkarekisteri sekä rakennetilastojen kyselykanta

Tilastokeskuksessa on useita eri tietokantoja ja rekisterejä, joihin on koottu laajat tiedot Suomen yrityksistä ja toimipaikoista. Tutkimuksessani käytän Tilastokeskuksen yritysten suhdanteet -yksikön tuottamaa yritys- ja toimipaikkarekisteriä sekä yritysten rakenteet -yksikön tuottamaa rakennetilastojen kyselykantaa vuodelta 2000. Niiden sisältämiä muuttujia raportoidaan luvuissa 3.2 ja 3.3.



**Yritys- ja toimipaikkarekisteri eli YTR** on Tilastokeskuksen yritysten suhdanteet -yksikössä muodostettava jatkuvasti päivitettävä rekisteri. Se käsittelee Suomen **yrityksiä** ja toimipaikkoja. *Yritys- ja toimipaikkarekisterin muodostavat kuitenkin vain ne yritykset, jotka ovat toimineet yli puoli vuotta ja joiden liikevaihto ylittää asetetun kynnyksärajan, joka oli vuonna 2000 8914 EUR (53 000 mk, raja nousee joka vuosi inflaatioprosentin verran) tai joiden toiminta vastaa vähintään puolen henkilön työpanosta. Tilastokeskuksen yritysrekisterin mukaan **liikeyrityksellä** tarkoitetaan talousyksikköä, joka kattaa menonsa kokonaan tai ainakin pääosin myynti-, korko-, osinko- tai vakuutusmaksutuloilla. Vuonna 2000 Suomessa toimi 222 817 liikeyritystä, jotka täyttivät Tilastokeskuksen asettamat ehdot liikeyritysten tilastojen laatimiselle. Niistä 7852 oli maatiloja (maa-, riista-, kala- ja eläintalouden yrityksiä). Maatilat eivät ole arvonlisäverovelvollisia ja siksi sisältyvät yritysrekisteriin vain, jos ovat työnantajia. Maatilojen osuus yrityksistä jää vähäiseksi, koska valtaosa niistä ei toimi työnantajina. Yritysrekisterissä on tietoja liikeyrityksistä ja toimipaikkarekisterissä on tietoja liiketoimipaikoista. Yritysrekisterin tietoja muodostettaessa kaikki yrityksen toiminta tilastoidaan yrityksen päätoimialalle ts. tilastoyksikkö on yritys. Toimipaikkarekisterin tilastoyksikkö on toimipaikka liitteessä 2 annetun määritelmän mukaan. (Tilastokeskus, 2002e).*

Yritys- ja toimipaikkarekisteri sisältää myös yhteisöjä, yksityisiä elinkeinonharjoittajia sekä valtion liikelaitoksia, jotka ovat arvonlisäverovelvollisia tai työnantajia ja joilla on ollut toimintaa kyseisen tilikauden aikana. Yritys- ja toimipaikkarekisteri sisältää lisäksi ns. julkisyhteisöjen rekisterin, joka kattaa kaikki julkisen vallan yhteisöt eli valtion virastot, kunnat ja kuntayhtymät sekä niiden toimipaikat. Julkisyhteisöillä ei ole liikevaihtotietoa, mutta muutoin julkisyhteisöjen rekisterin tietosisältö on lähes sama kuin yrityksiä koskevan yritys- ja toimipaikkarekisterin. Tutkimuksessa käsitellään kuitenkin vain liikeyrityksiä, jotka ovat yritysrekisterissä ja rakennetilastojen kyselykannassa vuonna 2000. Näistä liikeyrityksistä ja niiden toimipaikoista muodostuu tutkimuksen perusjoukko. Siksi tutkimuksessa ei käsitellä julkisyhteisöjen rekisteriä.

Yritysrekisterin tiedot kootaan tiedustelemalla tietoja kaikilta monitoimipaikkaisilta yrityksiltä ja tilikauden aikana perustetuilta yksitoimipaikkaisilta yrityksiltä. Lisäksi tietoja kysellään yksitoimipaikkaisilta yrityksiltä, jotka eivät ole tilikauden aikana perustettuja, mutta työllistävät yli 20 henkeä. Muilta yrityksiltä kysytään tietoja ns. rotaatioperiaatteella (otos). Yritysrekisterin tietoja saadaan myös erilaisista hallinnollisista rekistereistä (EVR, Suomen pankki, tullihallitus ja kaupparekisteri). Verohallinnon rekistereistä (EVR) saadaan yrityksen aloitus- ja lopetus-tiedot, kotikunta, alustava toimialatieto uusille yrityksille sekä liikevaihto ja palkat. Suomen pankissa on tietoja Suomessa toimivista ulkomaalaisomisteisista yrityksistä, tullihallituksella tuonti- ja vientirytyksistä ja kaupparekisterissä fuusioista ja konserneista. Puuttuvat tiedot estimoidaan esim. käyttämällä edellisen vuoden tietoja tai muita muuttujia. Yritysrekisterin tietolähteinä ovat siis Tilastokeskuksen omat tiedustelut, erilaiset hallinnolliset rekisterit ja edellisten vuosien tiedot. (Jouko Rajaniemen esitelmä yritysaineistot aluerakenteen- ja kehityksen tutkimuksessa -seminaarissa 17.10.2002).

**Toimipaikkarekisterissä ovat liikevaihto, henkilöstön määrä ja palkat jaettu toimipaikoille joko käyttäen suoraan kyselyä tai imputointia/estimointia.** Tilastokeskuksen yritysten suhdanteet -yksikkö kerää toimipaikkarekisteriin vuosittain tietoja suoraan yrityksiltä lähettämällä niille lomakkeen, jossa kysytään mm. yrityksen toimipaikkojen toimialat, niiden tuottama liikevaihto ja niissä työskentelevä henkilöstömäärä. Jos tietoja ei toimipaikoilta saada kyselyllä, niin tiedot imputoidaan luvussa 3.2 kuvatulla tavalla. Luvussa 3.2 on lisäksi kerrottu

mitä toimipaikan liikevaihdolla, palkoilla ja henkilöstön määrällä yritysten suhdanteet -yksikössä Tilastokeskuksessa tarkoitetaan. Tilinpäätöstilastoista saadaan selville liikevaihdon jakautuminen päätoimialoittain, kulujen erittely kululajeittain ja keskimääräinen henkilöstömäärä vuodessa. Niiltä yrityksiltä, joille ei kyselyä ole tehty, haetaan tietoja elinkeinoverotuksesta.

**Imputointi** tarkoittaa puuttuvan tiedon korvaamista jollain tavalla estimoidulla arvolla tai arvoilla. Imputointi aiheuttaa lisävarianssia, joka tulisi ottaa huomioon estimoinnissa. Yritys- ja toimipaikkarekisterissä ja rakennetilastojen kyselykannassa ei ole dokumentoitu sitä millä menetelmillä imputointi on tehty ja merkitty sitä mitkä havainnot ovat imputoituja. Tämä voi aiheuttaa epäluotettavuutta tuloksiin, jotka on saatu yritys- ja toimipaikkarekisteriä sekä rakennetilastojen kyselykanta käyttämällä. Imputoinnista on kerrottu luvussa 5.5.

**Rakennetilastojen kyselykanta** jakautuu kahteen osaan. Toinen on yritystasolla oleva kyselykanta, joka on nimeltään YKANTA ja toinen toimipaikkatasolla oleva kyselykanta, joka on nimeltään TKANTA.

YKANTA sisältää kaikilla toimialoilla toimivien yritysten tietoja. Se on kokonaisuineen ja sisältää lähes kaikkien Suomessa toimivien yritysten tiedot. Tilastoyksikkönä on yksittäinen yritys ja kaikki yrityksen toiminta on tilastoitu yrityksen päätoimialalle. Rakennetilastojen tietolähteinä ovat Tilastokeskuksen yritys- ja toimipaikkarekisteri ja verohallinnon yritysten tilinpäätökset ts. elinkeinoverorekisteri eli EVR. Edellä mainituista tietolähteistä saatuja tietoja täydennetään ns. suoralla tiedustelulla. Suora tiedustelu tarkoittaa suoran kyselyn lomakkeen eli TILKES-lomakkeen lähettämistä osalle Suomen yrityksistä (suuret yritykset). Niitä yrityksiä, joille TILKES-lomake lähetettiin, oli vuonna 2000 9158. Näistä yrityksistä hyväksytyyn vastauksen antoi 7432. YKANTAAN otetaan sellaiset yritykset, jotka ovat joko yritysrekisterissä tai mukana rakennetilastojen suorassa tiedustelussa (niille on lähetetty TILKES-lomake) tai jotka ovat uudenmuotoisia liikelaitoksia. (Koivuniemi, 2002, 10).

TILKES-lomake lähetetään niille yrityksille, joiden päätoimiala on C-I tai K, jos kyseinen yritys työllistää vähintään 20 henkilöä. Lisäksi TILKES-lomake lähetetään niille yrityksille, joiden päätoimiala on M, N tai O, jos kyseinen yritys työllistää vähintään 50 henkilöä. Lisäksi TILKES-lomake lähetetään muutamille yrityksille, jotka työllistävät vähemmän kuin 50 henkilöä. TILKES-lomake on jokaiselle toimialalle erilainen ja se sisältää tilinpäätöstietojen erittelyjä. Jollei yritys vastaa TILKES-lomakkeella olevaan kyselyyn tai TILKES-lomaketta ei ole yritykselle lähetetty, niin TILKES-lomakkeen tiedot imputoidaan käyttämällä elinkeinoverorekisterin (EVR), edellisten tilikautien, yritysten toimintakertomuksien tai hyödyketilastojen tietoja. (Koivuniemi, 2002, 10-13).

**Toimipaikkatason tietoja eli TKANTAA** muodostettaessa eritellään yrityksen toiminta yrityksen eri toimipaikoille. Tilastoyksikkönä on toimipaikka, joka on Tilastokeskuksen toimipaikkarekisterissä tai jota vastaava yritys on mukana suorassa kyselyssä. Rakennetilaston toimipaikkatason kyselykanta sisältää vain rakentamisen, teollisuuden, mineraalien kaivuun, sähkön-, kaasun- ja vesihuollon (TOL 95:n pääluokat C, D, E ja F, liite 2) toimialojen toimipaikkojen tiedot. Siksi rakennetilastojen toimipaikkatason kyselykanta sisältää vähemmän toimipaikkoja kuin toimipaikkarekisteri. (Koivuniemi, 2002).

**TKANTAA muodostettaessa yhdistetään suoran kyselyn, yritys- ja toimipaikkarekisterin ja elinkeinoverotuksen (EVR) tiedot. Verotustiedot ja yritys- ja toimipaikkarekisterin tiedot eivät kuitenkaan kata kaikkia sitä mitä kyselyssä on, ja ne suoran kyselyn tiedot, jotka puuttuvat, imputoidaan.** Koska rakennetilastojen toimipaikkatason kyselykanta sisältää vain toimialojen C, D, E ja F toimipaikat, sen sisältämät tiedot ovat Tilastokeskuksessa tiedossa monitoimipaikkaisten yritysten kohdalla vain teollisuuden tai rakentamisen toimipaikoille. YKANNASTA ovat saatavissa yksitoimipaikkaisten yritysten tiedot. Tilinpäätöstiedoista vain liikevaihto, henkilöstön määrä ja palkat ovat tiedossa kaikille toimipaikoille. Ne saadaan toimipaikkarekisteristä ja ovat osittain estimoituja tietoja (luvut 3 ja 8). (Koivuniemi, 2002, Karttunen, 2002).

Rakentamisen toimialalla toimipaikkojen tietojen kysely on suppeampi kuin teollisuuden toimialoilla (siis toimialat C, D ja E), ja siksi rakentamisen toimialalla ne tiedot, joita ei tiedustella, imputoidaan, eli toimipaikkalomake laajennetaan vastaamaan teollisuuden toimialan toimipaikkalomaketta imputoimalla. Lisäksi rakentamisen toimialalla toimipaikkojen (rakennusmaiden) sijainti vaihtuu nopeasti ja olemassaoloaika voi olla lyhyt, eivätkä toimipaikkalomakkeella tiedustellut tiedot ole siksi yhtä luotettavia kuin muilla toimialoilla. Toimipaikkalomake postitetaan myös vähintään 20 henkilöä työllistävälle palvelualalla toimiville yrityksille ja kunnallisille energialaitoksille sekä osalle kunnallisia vesilaitoksia ja kunnallisia rakentamisen ja teollisuuden toimipaikkoja. (Koivuniemi, 2002).

Vuonna 2000 toimipaikkalomakkeita lähetettiin 2132:lle toimipaikalle, joista 10% ei vastannut kyselyyn. Toimipaikkalomakkeen tietoja verrataan toimipaikkaa vastaavan yrityksen tietoihin, hyödyketilastojen tietoihin ja yritys- ja toimipaikkarekisterin tietoihin ennen kuin tiedot talletetaan TKANTAAN. Jos toimipaikkalomakkeella annetut tiedot ovat puutteellisia, niin tiedot imputoidaan. Vastaamatta jättäneiden yritysten tiedot imputoitiin käyttämällä elinkeinoveroaineistoa, yritys- ja toimipaikkarekisteriä ja YKANTAA. (Koivuniemi, 2002).

Syitä sille miksi kaikki toimipaikan tietojen kyselyn piiriin kuuluvat toimipaikat eivät vastaa, on monia. Perussyynä on kuitenkin se, että monitoimipaikkainen yritys ei pysty jakamaan toimintaansa toimipaikoille, jos ei ole käytetty toimipaikkakohtaista kirjanpitoa. Kaikki yritykset eivät käytä toimipaikkakohtaista kirjanpitoa mm. siksi, että yrityksen tulosityksikkö on eri kuin toimipaikka. Voi myös olla, ettei yrityksessä kukaan ehdi, pysty tai halua täyttää lomaketta. Toimipaikkatietojen tulkinnassa on ongelmana kadon lisäksi se, että yritys saattaa antaa jonkun toimipaikan tiedot jonkun muun toimipaikan kohdalla. Tästä seuraa mm. se, että sellaistenkin toimipaikkojen tiedot, jotka eivät toimi teollisuuden ja rakentamisen toimialalla, saattavat sisältyä TKANTAAN. Rakennetilastojen toimipaikkatason kyselykannassa voi siksi esiintyä kehikkovirheitä eli yli- tai alipeittoa sekä kaksinkertaista listausta (katso luku 8 ja 5.4). (Koivuniemi 2002, 12- 15, 17).

Rakennetilastojen toimipaikkatason kysely tehdään vain teollisuuden ja rakentamisen toimialojen toimipaikoille luultavasti siksi, että teollisuuden ja rakentamisen osuus yrityksistä ja yritysten tilinpäätöstiedoista on suuri (luku 6). Itse asiassa suuri osa yrityksistä on teollisuuden tai rakentamisen alan yrityksiä ja näiden yritysten osuudet liikevaihdosta ja henkilöstömäärästä ovat merkittäviä. Teollisuus vaikuttaa kansakunnan tuottavuuteen aika paljon, koska se tuottaa pääosan maan vientituloista ja siellä tehtävä teknologinen kehitystyö hyödyttää muitakin toimialoja. Toisaalta teollisuuden tuottavuus on riippuvainen muista toimialoista, koska tehtaot tarvitsevat

tuotantoonsa mm. koulutettua työvoimaa, toimivaa infrastruktuuria ja monenlaisia tuotannontekijöitä (erityisesti raaka-aineita) muualta. Voidaan siis ajatella teollisuudella olevan ratkaiseva rooli kansantuotteen ja yritysten tilinpäätöstietojen (etenkin Suomen yritysten yhteen lasketun jalostusarvon) muodostumisessa. (Maliranta 1996, 11).

Toimipaikan alueellisena yksikkönä rakennetilastossa on yleensä kunta. Se tarkoittaa sitä, että yrityksen saman kunnan alueella sijaitsevat saman toimialan toimipaikat yhdistetään ts. niistä muodostetaan yksi toimipaikka. Toisaalta tästä ja tilaston tulostustarpeista seuraa, että yrityksen eri kunnissa sijaitsevat toimipaikat katsotaan rakennetilastoissa erillisiksi tilastoyksiköiksi. Rakentamisen toimialan yrityksillä pienin alueellinen yksikkö käsittää yrityksen toiminnan useamman kunnan alueella. Siksi rakentamisen toimialan yrityksillä yksittäinen toimipaikka voi käsittää useamman kunnan alueella olevia toimipaikkoja ts. tilastoissa yhtenä toimipaikkana oleva yksikkö voikin tosiasiallisesti sisältää useamman toimipaikan. Lisäksi kaikkien rakennetilastojen yritysten kohdalla alla 20 henkilön yrityksiä kohdellaan yksitoimipaikkaisina eli niille yritystiedot = toimipaikkatiedot (toimipaikkatiedot saadaan siis YKANNASTA). Tämä tarkoittaa sitä, että alle 20 henkilön yrityksistä osa voi olla tosiasiallisesti monitoimipaikkaisia. Niinpä rakennetilastoissa olevat yritysten toimipaikkojen tiedot voivat poiketa siitä mikä tilanne yrityksissä tosiasiaassa on. (Tilastokeskus, 2002c).

Rakennetilastoissa toimipaikan ei tarvitse käsittää vain rakentamiseen, teollisuuteen, mineraalien kaivuuseen tai sähkön-, kaasun- ja vesihuoltoon liittyviä toimintoja. Rakennetilastot sisältävät myös ns. aputoimintoja ja palveluksia, jotka sijaintinsa puolesta liittyvät teollisuuden ja rakentamisen toimipaikkojen toimintaan. Teollisia aputoimintoja ja palveluksia ovat korjaus ja kunnossapito, energian tuotanto ja jakelu, varastointi, kuljetus, tutkimus- ja kehittämistoiminta, tietojenkäsittely, hallinto sekä myynti. Jos apu- tai palvelutoiminta liittyy kiinteästi varsinaisen toimipaikan toimintaan ja palvelee pääasiassa kyseistä toimipaikkaa, ei näitä toimintoja ole välttämättä erotettu erillisiksi toimipaikoiksi, vaan toiminnot on yhdistetty valmistustoimintoihin. Varsinaisen teollisen toimipaikan yhteydessä toimivat energiaa tuottavat toimipaikat luokitellaan tilastoissa apuyksikköinä sille toimialalle, jota kyseinen toimipaikka pääasiallisesti palvelee. (Tilastokeskus 2002 b).

Rakennetilastojen kyselykannan tietojen ja yritys- ja toimipaikkarekisterin tietojen yhdistämisessä ilmenee kehittämistarpeita, jotka johtuvat siitä ettei rakennetilastoissa noudateta samoja ratkaisuja kuin yritys- ja toimipaikkarekisterissä, ts. rekisterit eivät ole yhteneviä. Näitä ovat Koivuniemen (2002, 16-18) mukaan seuraavat.

1. Aputoimipaikan määrittely. Rakennetilastoissa voivat aputoimipaikkoja olla yritys- ja toimipaikkarekisterin varsinaisiksi toimipaikoiksi määrittelemät yksiköt. Seurauksena on se, että toimipaikkojen toimialat eivät ole aina samat yritys- ja toimipaikkarekisterissä ja rakennetilastojen kyselykannassa. Tutkimuksessa käytetään rakennetilastojen kyselykannan tapaa määrittellä toimipaikan toimiala, jos toimipaikan toimiala on rakennetilastoissa määritelty. Muuten käytetään toimipaikkarekisterin tapaa määrittellä toimipaikan toimiala.

2. Teollisuuteen tai rakentamiseen (toimialat C- F) kuulumattoman toimipaikan tiedot sisältyvät teollisuuden tai rakentamisen toimipaikan tietoihin, jos yritys ei halua tai pysty jakamaan tietojaan toimipaikkatasolle yritys- ja

toimipaikkarekisterin käyttämän tavan mukaan. Tämä ongelma on tavallinen, jos yrityksellä on monta toimipaikka saman kunnan alueella ja erityisesti, jos toimipaikat toimivat samalla toimialalla.

3. Epäitsenäisten yritysten toimialan määrittely ja yrityksen päätoimiala. Rakennetilastojen kyselykannan ja yritys- ja toimipaikkarekisterin muodostamisessa käytetään eri kriteerejä ns. epäitsenäisten yritysten määrittelyssä. Epäitsenäisten yritysten määrittelyerojen vuoksi myös yritysten päätoimialan määrittelytavat poikkeavat yritysrekisterissä ja rakennetilastojen kyselykannassa. Tämä johtuu siitä, että Tilastokeskuksessa epäitsenäisen yrityksen päätoimiala on sen omistavan yrityksen päätoimiala ja epäitsenäisen yrityksen päätoimiala voi olla siis eri kuin mitä se olisi sen toiminnan luonteen perusteella.

4. Toimialaluokituksen 1995 käsikirjan ohjeita (Tilastokeskus, 1993) tulkitaan eri tavoin. Tästä syntyy eroja yritysten ja toimipaikkojen toimialoissa yritys- ja toimipaikkarekisterin ja rakennetilastojen kyselykannan välillä. Tutkimuksessa käytetään rakennetilastojen kyselykannan tapaa tulkita TOL 95:tä, koska TKANTA on muodostettu sen perusteella.

5. Tietojen yhdistämisestä johtuvat ongelmat. Rakennetilastojen kyselykannassa kahden yrityksen tiedot yhdistetään, jos ne ovat fuusioituneet. Yhdistämistä tehdään myös, jos yrityksellä on ollut kalenterivuoden aikana kaksi tilikautta. Sen sijaan yritysrekisterissä ei yhdistelyä tehdä.

6. Erot siinä miten käsitetään se milloin toimipaikan toiminta on alkanut. Rakennetilastojen kyselykannassa ovat sellaisetkin toimipaikat, joiden rakentaminen on aloitettu tilastovuoden aikana ts. toimipaikat lisätään rakennetilastojen kyselykantaan jo niiden investointivaiheessa. Toimipaikkarekisteriin uusi toimipaikka lisätään vasta kun se on valmis ts. ei vielä investointivaiheessa. Tästä johtuen rakennetilastojen kyselykanta sisältää sellaisia toimipaikkoja, joita ei toimipaikkarekisterissä ole. Koska tutkimusaineistossa on vain sellaisia liikeyrityksiä ja niiden toimipaikkoja, jotka ovat sekä yritys- ja toimipaikkarekisterissä että rakennetilastojen kyselykannassa, ei aineisto sisällä toimipaikkoja, jotka ovat olleet investointivaiheessa vuonna 2000. Noudatetaan siis toimipaikkarekisterin tapaa lisätä uusi toimipaikka rekisteriin.

Koivuniemi (2002) ehdottaa, että yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan käyttämiä toimintatapoja yhdenmukaistettaisiin mahdollisimman paljon esim. siten, että yritys- ja toimipaikkarekisteriin lisättäisiin uudet toimipaikat jo niiden investointivaiheissa ja muodostettaisiin yhteiset tavat tulkita toimialaluokitusta 1995. Tätä on jo tehty ja työtä jatketaan Tilastokeskuksessa. Koivuniemi ehdottaa, että rakennetilastojen kyselykannassa alettaisiin noudattaa yritys- ja toimipaikkarekisterin tapaa määrittellä yrityksen tai toimipaikan toimiala. (Koivuniemi 2002, 16–18).

Havaittavissa oli myös muita YTR:n ja rakennetilastojen kyselykannan tietojen yhdistämiseen liittyviä kehittämistarpeita. Ne muuttajat, jotka ovat molemmissa rekistereissä (toimiala, liikevaihto, palkat, henkilöstön määrä) on määritetty YTR:ssä erilailla kuin rakennetilastojen kyselykannassa tai sitten tietolähteet ovat eriävät. Näistä eroista on kerrottu luvussa 8. Määrittely- ja tietolähde-erot johtavat siihen, että samoista asioista kertovien muuttajien havaintoarvot ovat YTR:ssä ja rakennetilastojen kyselykannassa erisuuruisia. YTR:ssä ei puuttuvia

havaintoja ole lainkaan, koska osa tiedoista on imputoitu. Rakennetilastojen kyselykannassa vuodelta 2000 puuttuvia havaintoja on jalostusarvon kohdalla vain toimipaikkatason kyselykannassa eli TKANNASSA (siis yrityksen jalostusarvot ovat tiedossa kaikilta yrityksiltä). TKANNASSA vuodelta 2000 jalostusarvo puuttui viideltä toimipaikalta. Puuttuvia havaintoja on siis vain tulosuuttujan arvoissa, ei lisäinformaatiassa.

### 3.2 Tutkimuksessa käytettävät muuttujat ja perusjoukko

*Tutkimuksessa muodostetaan jalostusarvojen estimaatit niille toimipaikoille, joille se ei ole tiedossa.* Jalostusarvo on valittu tutkimuksen tulosuuttujaksi, koska se kertoo toimipaikan koosta, toiminnan laajuudesta ja kansantaloudellisesta merkittävydestä. Se on yksi tärkeimmistä tilinpäätöksestä laskettavista tunnusluvuista ja se heijastaa kaikkea tuloslaskelmasta saatavaa informaatiota (kuvaa yrityksen osuutta bruttokansantuotteesta). Estimaattien laatimisessa käytetään apumuuttujina seuraavia Tilastokeskuksen toimipaikkarekisterissä olevia muuttujia: toimipaikan toimialan pääluokka, toimipaikan liikevaihto, toimipaikan keskimääräinen henkilöstömäärä tilikauden aikana, toimipaikalle estimoidut palkat, toimipaikkaa vastaavan yrityksen laji ja PK-suuruusluokka. Näiden apumuuttujien havaintoarvoja käytetään lisäinformaationa sekä estimoinnin että otoksen poiminnan vaiheissa ja ne ovat ns. riippumattomia muuttujia. Teollisuuden toimialalla tutkitaan myös toimipaikan sijainnin vaikutusta toimipaikan tilinpäätöstietojen estimointiin ja siksi toimipaikan sijainti mitattuna toimipaikan sijaintiläänillä on estimoinnissa käytettävä apumuuttuja.

Tutkimuksessa käytettävä tulosuuttuja ja apumuuttujat mittaavat toimipaikan kansantaloudellista merkittävyyttä, kokoa ja toiminnan laajuutta. Tutkimuksen tulosuuttuja, jalostusarvo, on toimipaikan laskennallisesta käyttökatteesta laskettava tieto. Siksi se pitää sisällään määritelmänsä perusteella toimipaikan liikevaihdon ja palkat. Tämä antaa aiheen olettaa, että määritelmästä johtuen liikevaihto ja palkat korreloivat vahvasti jalostusarvon kanssa. Liikevaihto ja palkat ovat siksi tilastollisesti perusteltu apumuuttuja estimoinnissa. Koska tilinpäätöstiedot riippuvat paljolti siitä mikä kokoinen yritys/toimipaikka on, voidaan ajatella, että henkilöstömäärä on myös tilastollisesti perusteltu apumuuttuja estimoinnissa ja korreloi mm. vahvasti jalostusarvon (tutkimuksen tulosuuttujan) kanssa.

Tutkittavat yritystason muuttujat tutkimuksessani ovat lähinnä yritysten tilinpäätöksiin sisältyviä tai niistä johdettavia tietoja. ***Henkilöstön määrää lukuun ottamatta tutkimuksen yritystason muuttujat ovat kululajipohjaisen tuloslaskelman liikevoittoa/tappiota edeltävään osan eriä tai niistä laskettavia tunnuslukuja. Toimipaikkatasolla kululajipohjaista tuloslaskelmaa vastaa toimipaikan laskennallinen käyttökate.*** Henkilöstön lukumäärää lukuun ottamatta tutkimuksen toimipaikkatason muuttujat ovatkin toimipaikan laskennallisen käyttökateen eriä tai siitä laskettavissa olevia tunnuslukuja. Kululajipohjaisen tuloslaskelman liikevoittoa/tappiota edeltävien erien kuvaukset ovat liitteessä 3. (Leppiniemi ja Leppiniemi, 2000, Kinnunen ym., 2000).

***Tutkimuksessa käytettävä tulosuuttuja, jalostusarvo.*** Jalostusarvo mittaa kansantaloudessa yrityksen tai toimipaikan osuutta (brutto)kansantuotteen kerryttäjänä. ***Bruttokansantuotteella*** tarkoitetaan kansantaloudessa vuoden aikana tuotettujen hyödykkeiden arvoa rahana. Kansantuote kertoo sen kuinka paljon tavaroita

ja palveluksia kansantaloudessa tuotetaan. Laskelmissa tuotetun tuotteen/palvelun arvosta vähennetään tuotantotehtäjäiden ostot muilta tuottajilta. (Taloustieto Oy 1994, Autio, 1995).

Jalostusarvo kuvaa sitä arvonlisäystä, jonka tuote esim. raaka-aine tai puolivalmiste saa kulkiessaan yrityksen tuotantoprosessin läpi myytäväksi tuotteeksi. Tuotteen jalostusarvo, arvonlisä, saadaan vähentämällä sen myynnin arvosta muilta ostetut välituote-erät esim. raaka-aineet ja puolivalmisteet. Jalostusarvoa käytetään yritysten välisessä koon ja toiminnan merkityksen vertailussa ja sen perusteella on myös saatavissa käsitys yrityksen toiminnan luonteesta. Yritysten tilinpäätöstiedot poikkeavat toisistaan mm. siksi, että yritykset ovat eri kokoisia ja toimivat eri toimialoilla. Lisäksi yritykset ovat usein pääoma- tai työvoimavaltaisia. Pääomavaltaisilla yrityksillä jalostusarvo on suurempi kuin työvoimavaltaisilla. (Leppiniemi ja Leppiniemi 2000, 174, Taloustieto Oy, 1994, Tilastokeskus, 2002 b).

Voidaan ajatella toimipaikkakohtaisten jalostusarvojen sekä arvonlisäyksien kertovan siitä, mikä on eri toimipaikkojen osuus yritysten jalostusarvosta ja kansantuotteesta. Jos yritys toimii monella eri alalla, saadaan toimipaikkakohtaisten jalostusarvojen avulla tietoa myös siitä, mikä on eri toimialojen osuus kansantuotteen kertyjänä. Yrityskohtainen jalostusarvo ei tästä kerro, koska siinä kaikki yrityksen toiminta tilastoidaan yrityksen päätoimialalle (siis ajatellaan monitoimialaistenkin yrityksen toimivan vain yhdellä toimialalla). Siksi ei voida arvioida yrityskohtaisesta jalostusarvosta mm. sitä mitkä ovat eri sivutoimialojen osuudet yrityksen jalostusarvosta.

Tilastokeskuksen ja yritystutkimusneuvottelukunnan laskema jalostusarvo saadaan tuloslaskelman ja sen eritelyjen avulla seuraavasti: **Jalostusarvo = tuotannon bruttoarvo - välituotepanokset**. (Yritystutkimusneuvottelukunta 1999).

**Tuotannon bruttoarvo** mittaa yrityksen tai toimipaikan tuotannon arvoa. Tuotantotoimintaan lasketaan mukaan kaikki tuotannosta saadut tuotot ja niitä oikaistaan käyttöomaisuuden luovutusvoitoilla, koska käyttöomaisuuden luovutusvoittoja ei Tilastokeskuksessa pidetä tuotannollisina erinä, vaan satunnaisina tuloina. Kauppatavaroiden hankinnat vähennetään tuotantotoiminnan muista tuotoista, jotta tuotantotoimintaan saadaan mukaan vain kauppatavaroiden myynnistä yritykselle/toimipaikalle syntyvä marginaali. (Yritystutkimusneuvottelukunta, 1999).

Yrityksen tuotannon bruttoarvo lasketaan yrityksen ilmoittaman tuloslaskelman avulla ja toimipaikan tuotannon bruttoarvo toimipaikan ilmoittaman laskennallisen käyttökateen avulla seuraavasti: **Bruttoarvo = liikevaihto + (toimitukset yrityksen muiden toimipaikkojen käyttöön) + valmisteverastojen muutos + valmius omaan käyttöön + liiketoiminnan muut tuotot - käyttöomaisuuden luovutusvoitot - kauppatavaroiden hankinta - myynnin perusteella saadut avustukset**. (Yritystutkimusneuvottelukunta, 1999).

Yrityksen **välituotepanokset** lasketaan seuraavasti: aine- ja tarvikeostot + varastojen muutos + ulkopuoliset palvelut + muut kiinteät ja muuttuvat kulut pois lukien henkilöstökulut = tuotannon bruttoarvo - jalostusarvo. Toimipaikan välituotepanokset lasketaan seuraavasti: aine- ja tarvikeostot + ostot yrityksen muilta toimipaikoil-

ta + varastojen muutos + ulkopuoliset palvelut + muut kiinteät ja muuttuvat kulut ml. henkilöstökulut = tuotannon bruttoarvo – jalostusarvo. (Yritystutkimusneuvottelukunta, 1999).

***Käytettävät muuttujat estimaattien laskemisessa ts. riippumattomat/selittävät muuttujat X:***

Saamani informaation mukaan yrityksen ja toimipaikan **keskimääräiset henkilöstömäärät tilikauden aikana** on yritys- ja toimipaikkarekisterissä ja rakennetilastojen kyselykannassa määritelty eri tavoin. Yritys- ja toimipaikkarekisterissä keskimääräinen henkilöstön määrä ilmoitetaan ns. **fulltime equivalent** eli kokopäiväisten työntekijöiden käsitteen mukaisesti (Karttunen, 2002). Rakennetilastojen kyselykannassa keskimääräinen henkilöstömäärä ilmoitetaan yrityksen tai toimipaikan antaman tiedon mukaan.

Kokopäiväisten työntekijöiden käsitteen käyttäminen yritysrekisterissä tarkoittaa sitä, että yrityksessä työskentelevän henkilöstön määrä on mitattu siten, että osa-aikaista työtä tekevien työntekijöiden työpanos on muutettu kokoaikaiseksi eli on yhdistetty eri työntekijöiden työpanokset kokonaisiksi työvuosiksi, jolleivät kaikki työntekijät työskentele koko vuotta. Esimerkiksi kaksi henkilöä, jotka ovat työskennelleet yrityksessä puoli vuotta (heidän työpanoksensa on puoli henkilötyövuotta), ilmoitetaan yhdeksi henkilöksi. Kaikki henkilöt, jotka ovat työskennelleet tilikauden aikana yrityksessä tulisi merkitä yritysrekisteriin joko kokonaisina työvuotena tai osatyövuotena. Henkilöstön keskimääräiseen lukumäärään tilikauden aikana lasketaan kaikki yrityksen palveluksessa olevat, myös harjoittelijat ja tilapäiset työntekijät ja toimihenkilöt, vuosilomalla olevat, sairauslomalla olevat, äitiyslomalla olevat, lakossa tai lomautettuna olevat ja kertausharjoituksissa olevat. (Karttunen, 2002).

Se mitä palkansaajan kohdalle merkitään henkilötyövuodeksi riippuu hänen työpanoksestaan. Se määritellään suhteessa ns. viitetietoihin, joita ovat työsuhdekohtainen palkka, ammatti, koulutus, ikä, sukupuoli ja kotikunta sekä työntajayrityksen toimiala ja sektori. Viitetietojen perusteella estimoidaan työntekijälle hänen viiteryhmänsä mukainen keskiarvopalkka ja työntekijän todellinen palkka suhteutetaan edellä olevaan keskiarvopalkkaan. Suhteuttaminen tehdään siten, että normaali maksimi on yksi henkilötyövuosi. Tätä tarkoitetaan työntekijän **työpanoksella**. Nämä henkilötyövuodet sitten summataan työnantajayrityksille. Tästä koostuu yritysrekisterin henkilöstön lukumäärä keskimäärin tilikauden aikana. Viitetiedot saadaan työnantajien vuosi-ilmoitusaineistosta, yritysrekisterin omista tiedoista ja työssäkäyntitilastojen aineistoista. (Karttunen, 2002).

Yritysrekisterin tapa laskea keskimääräinen henkilöstön lukumäärä poikkeaa joskus siitä miten yritys sen ilmoittaa. Yritykset saattavat ilmoittaa myös määräaikaisesti tilivuonna firmassa työskennelleet henkilöt kokonaisina henkilötyövuosina tai jättää lukumäärästä pois kaikki osa-aikaisesti työskennelleet työntekijät (mm. kesätyöntekijät ja puolipäiväisesti työskentelevät sekä harjoittelijat). (Karttunen, 2002).

Yritysten henkilöstön jakaminen toimipaikkoihin ts. toimipaikkojen henkilöstömäärien tiedot tehdään keräämällä tietoja yrityksiltä ja estimoimalla. Kaikilta monitoimipaikkaisilta yrityksiltä kysellään keskimääräinen henkilöstömäärä tilivuoden aikana. Yksitoimipaikkaisilta yrityksiltä tiedot kysellään vain, jos niiden ajatellaan muuttuneen (fuusio, yrityksen muuttuminen monitoimipaikkaiseksi, toimialan tai sijainnin muutos yms.). Muutoin keskimääräinen henkilöstömäärä toimipaikalla on yksitoimipaikkaisille yrityksille sama kuin koko yrityksen keskimääräinen



nen henkilöstömäärä (ajatellaan, että kun yritys on yksitoimipaikkainen, on toimipaikka sama kuin yritys). (Karttunen, 2002).

Jos estimoitu ja yritykseltä saatu tieto poikkeavat toisistaan, niin yritykseltä saatua tulosta tarkistetaan ja saatetaan korjata. Ristiriidat estimoitujen henkilöstömäärien ja yritysten ilmoittamien henkilöstömäärien välillä johtuvat yleensä siitä, että yritykset laskevat henkilöstön määrän eritavoin kuin Tilastokeskus. Jollei monitoimipaikkainen yritys ole ilmoittanut työntekijöidensä jakautumista toimipaikkoihin, niin yritykselle estimoitu keskimääräinen henkilöstömäärä jaetaan toimipaikkoihin käyttämällä edellisen tilivuoden tietoja. Edellä esitetty antaa aiheen olettaa, että osalla yrityksistä on keskimääräinen henkilöstömäärä estimoitu (jos on suuria ristiriitoja yrityksen antaman tiedon kanssa) ja osalla yrityksissä on keskimääräinen henkilöstömäärä yrityksen ilmoituksen mukainen. Samoin on myös toimipaikkatasolla, vaikkakin siellä estimoitujen tietojen osuus on oletettavasti suurempi kuin yritystasolla. (Karttunen, 2002).

**Liikevaihto** koostuu kirjanpitolain mukaan yrityksen varsinaiseen toimintaan (sen toimialan toimintaa, jolla yritys toimii) kuuluvien tuotteiden tai palvelujen tuotoista, joista vähennetään annetut arvonlennukset sekä arvonalisävero ja muut välittömästi myynnin määrään perustuvat verot mm. arvonalisäverot. Se on yleisemmin käytetty yrityksen toiminnan laajuuden sekä yrityksen merkittävyyden ja markkinaosuuden mitta (verrataan yrityksen liikevaihtoa muiden yritysten liikevaihtoon tai muiden saman toimialan yritysten liikevaihtoon tms.) ja se on osatekijänä monissa yritystä kuvaavissa tunnusluvuissa. Esimerkiksi eri kuluerien suuruutta tutkitaan usein suhteuttamalla ne liikevaihtoon. Liikevaihdon muutosten perusteella tehdään päätelmiä yritysten riskistä, mm. erittäin suurta liikevaihdon kasvua on pidetty konkurssiuhan merkinä. Samoin konkurssiuhan merkinä on pidetty myös liikevaihdon kasvun tyrehtymistä. (Leppiniemi ja Leppiniemi, 2000).

Samaan tapaan kuin keskimääräisen henkilöstömäärän jakaminen toimipaikkoihin, suoritetaan yritysrekisterissä myös liikevaihdon jakaminen toimipaikkoihin. Kaikilta monitoimipaikkaisilta yrityksiltä tiedot kysellään ja yksitoimipaikkaisille yrityksille tiedot kysellään vain, jos epäillään tietojen muuttuneen. Jollei monitoimipaikkainen yritys ilmoita liikevaihdon jakaantumista toimipaikoittain, lasketaan edellisen tilivuoden liikevaihtotietojen perusteella estimaatit toimipaikkakohtaisille liikevaihtoille. Vaihtoehtoisesti voidaan myös käyttää henkilöstön jakamaa toimipaikoittain tai liikevaihto per henkilö -tunnuslukua toimialalla. Liikevaihdon jakautuminen toimipaikoittain on siis osassa yrityksiä estimoitu tieto ja osassa yrityksen ilmoituksen mukainen. Kaikilla yksitoimipaikkaisilla yrityksillä liikevaihtotieto on yrityksen ilmoituksen mukainen (tilinpäätöksestä suoraan saatava tieto). (Karttunen, 2002).

Rakennetilastojen kyselykannassa taas liikevaihtotiedot perustuvat suoriin kyselyihin, elinkeinoverorekisteriin tai imputointiin. Täten toimipaikan liikevaihtotiedot eroavat toisistaan yritys- ja toimipaikkarekisterissä sekä rakennetilastojen kyselykannassa.

**Palkalla** tarkoitetaan työntekijän ja työnantajan välisen työsuhteen perusteella työnantajan maksamaa korvausta työntekijälle tekemästään työstä. Se määräytyy yleensä työhön käytettävän ajan, työn tuloksen tai muun sovitun perusteen mukaan. Tavallisesti palkka maksetaan rahana, mutta voidaan sitä myös sopimuksen mukaan

maksaa ns. luontoisetuina, esim. vapaa, työsuhdeasunto (jossa tavallista pienempi vuokra), työsuhdepuhelin tai työsuhdeauto ovat luontoisetuja. Luontoisedut rinnastetaan verotuksessa palkkaan eli ne ovat verotettavaa tuloa työntekijälle. Yrittäjäkin voi maksaa itselleen palkkaa määräämässä summan verran. Yksityisyrietykset eivät maksa lainkaan palkkaa, vaan niiden omistajat (yrittäjät) maksavat itselleen korvauksia suoraan omasta pääomastaan ts. omaisuudestaan. (Taloustieto Oy, 1994).

**Toimipaikan palkat** käsittävät kaikki toimipaikan maksamat ennakkopidätyksen alaiset erät. Palkkasumma on laskettu erikseen toimihenkilöille ja työntekijöille. Palkkasummaan lasketaan varsinainen rahapalkka, luontoisedut, loma-ajan palkka, lomarahat ja tuotantopalkkiot sekä muut lisäpalkkiot. Sairaus- ja äitiysloma-ajan palkkojen maksusta saadut sairausvakuutuskorvaukset vähennetään palkkasummasta. Yritys- ja toimipaikkarekisterissä toimipaikan palkat on estimoitu muodostamalla ne yrityksen palkkatiedoista jakamalla.

**Käytettävät muuttajat otoksen poiminnassa, ts. osituskriteerien laatimisessa käytettävät muuttajat (luku 5.2).**

**Toimipaikan toimialan pääluokka** ts. toimipaikan toimiala yhden kirjaimen tasolla TOL95:n mukaan. Toimialaluokitus 1995 on kuvattu liitteessä 2. Sitä käytetään osajoukkoihin eli domaineihin jaon kriteerinä tutkimuksessa (katso luku 5.2) osittamisen lisäksi.

**Toimipaikan omistavan yrityksen PK-suuruusluokka (tarkempi kuvaus luvussa 2.2):**

0 Määrittelemätön

1 **Mikroyritys**

2 **Pieni yritys**

3 **Keskisuuri yritys**

4 **Suuri yritys**

Edellisestä on muodostettu muuttuja PKSLY2, joka erottaa toisistaan PK-yritykset ja suuret yritykset.

(0 = määrittelemätön, 1 = **PK-yritys** 2 = **Suuri yritys**).

**Toimipaikkaa vastaavan yrityksen laji: yksi/monitoimipaikkainen yritys toimipaikkojen lukumäärän mukaan.** Käytetään kiinnostuksen kohteena oleviin osajoukkoihin jaon kriteerinä ja osittamisen.

**Informaatioryhmiin jaossa käytettävä muuttuja on toimipaikan sijainnista kertova muuttuja eli toimipaikan sijaintilääni**

1 Etelä – Suomen lääni

2 Länsi – Suomen lääni

3 Itä-Suomen lääni

4 Oulun lääni

5 Lapin lääni

6 Ahvenanmaa

Tutkimuksen **kohdeperusjoukko** on Suomessa vuonna 2000 toimineet liiketoimipaikat eli liikeyritysten toimipaikat yritysrekisterin asettamin ehdoin. **Kehikkoperusjoukon** muodostavat yritys- ja toimipaikkarekisterissä ja

rakennetilastojen kyselykannassa olevat toimipaikat, jotka ovat liikeyritysten toimipaikkoja yritysrekisterin aset-  
tamin ehdoin.

### 3.3 Tutkimusaineisto tiivistettynä

Kuviossa 3.1 on kuvattu aineistona käytettävien tietokantojen ja rekisterien välisiä yhteyksiä. Siinä kuvataan etenkin sitä mistä tietolähteistä (niitä on todella monia) muodostetaan Tilastokeskuksen yritysten rakenteet -yksikön rakennetilastojen kyselykanta ja yritysten suhdanteet -yksikön yritys- ja toimipaikkarekisteri ts. mitkä ovat Tilastokeskuksen tietolähteet muodostettaessa rekistereitä ja tietokantoja Suomessa toimiviin yrityksiin ja toimi-  
paikkoihin liittyen. Huomattava on, että Tilastokeskus ei saa kaikkea tietoa kyselämällä yrityksiltä, vaan myös mm. verohallinnon elinkeinoverorekisteristä (EVR). Toimipaikkojen tietoja ei kysellä kaikilta toimipaikoilta, vaan osa tiedoista joudutaan Tilastokeskuksessa estimoimaan/imputoimaan. Toimipaikkojen tilinpäätöstietojen impu-  
toinnissa käytettyjen menetelmien dokumentointi osoittautui kuitenkin puutteelliseksi (luvut 5.5 ja 8).

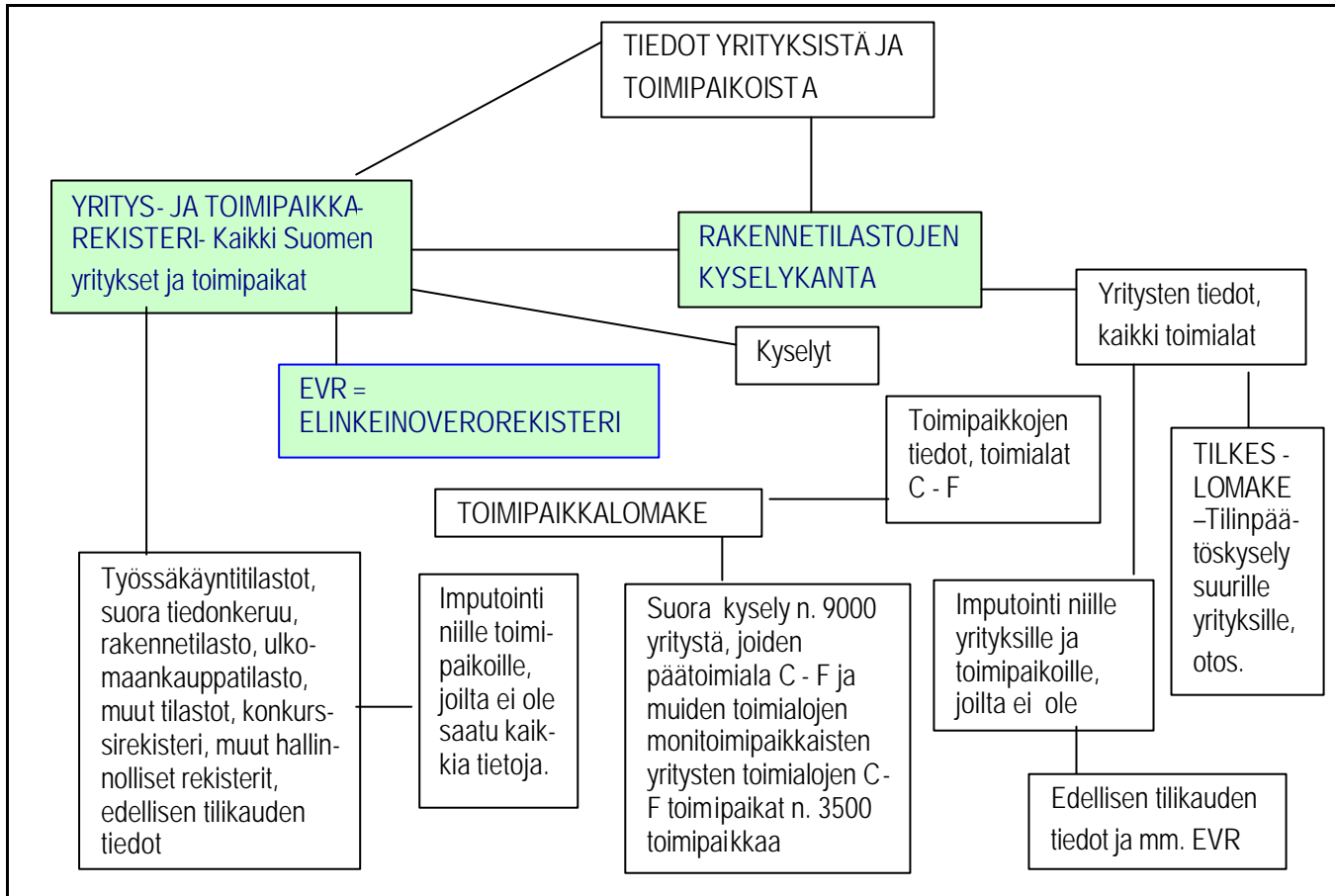
***Analyyysien laatimista varten muodostetaan yhdistämällä YTR, YKANTA ja TKANTA kaksi pääaineistoa*** ja näiden osajoukkoista koostuvia aineistoja. Toinen pääaineisto sisältää kaikki liikeyritykset ja niitä vastaavat toimipaikat, jotka ovat sekä YTR:ssä että YKANNASSA. Toinen aineisto sisältää ne toimipaikat, joille ovat tie-  
dossa kaikki tilinpäätöstiedot. Tämä aineisto sisältää siksi kaikki yksitoimipaikkaiset liikeyritykset, jotka ovat se-  
kä YTR:ssä että YKANNASSA sekä TKANNASSA ja toimipaikkarekisterissä olevat monitoimipaikkaisten yritys-  
ten toimipaikat eli monitoimipaikkaisten yritysten toimipaikat, joiden toimiala on rakennetilastojen mukaan teolli-  
suus tai rakentaminen. Aineistot sisältävät myös tutkimuksessa käytettävien rekistereissä olevien muuttujien  
pohjalta muodostettuja uusia muuttujia (toimipaikan toimialan pääluokka jne.). Rekisterien yhdistämisessä on  
ongelmia, joista on puhuttu luvuissa 3.1 ja 8. YTR:ssä ja YKANNASSA on liiketoimipaikkoja yhteensä 241 987  
kappaletta. Niistä yksitoimipaikkaisten yritysten ja monitoimipaikkaisten yritysten teollisuuden ja rakentamisen  
liiketoimipaikkoja on YTR:ssä ja YKANNASSA yhteensä 213 443 eli n. 88 % kaikista sekä yritys- ja toimipaikka-  
rekisterissä että rakennetilastojen kyselykannassa olevista toimipaikoista. TKANNAN toimipaikkoja eli teollisuu-  
den ja rakentamisen toimipaikkoja niistä on 57 091. Tutkimusaineisto laadittiin yhdistämällä yritys- ja toimipaik-  
karekisterin ja rakennetilastojen kyselykannan tiedot käyttäen liitteessä 4 kuvattua muodostamistapaa.

#### **TUTKIMUKSESSA KÄYTETTÄVIEN MUUTTUJIEN HAVAINTOARVOJEN TIETOLÄHTEET: VAIN YRITYS- REKISTERISSÄ OLEVAT MUUTTUJAT ( YRITYSTASOLLA, Tilastokeskus, 2002 d)**

- | <b><i>Muuttuja</i></b>                       | <b><i>SAS-koodissa annettu muuttujan nimi:</i></b> |
|--|--|
| • Yritystunnus                               | YRTUN  |
| • Yrityksen toimipaikkojen lukumäärä         | TPLKM  |
| • Yrityksen kotilääni                        | ULAANIYRIT   |
| • Yrityksen päätoimiala                      | NTALAYR  |
| • Yrityksen liikevaihto                      | LVMKYR   |
| • Yrityksen maksamat palkat                  | PALKKAMKYR   |
| • Yrityksen keskimääräinen henkilöstön määrä | HK2YRIT  |
| • Yrityksen PK-suuruusluokka                 | PKSLYR   |

VAIN RAKENNETILASTOJEN KYSELYKANNASSA – YKANTA - OLEVAT MUUTTUJAT (YRITYSTASOLLA, Tilastokeskus, 2002 b)

- **Muuttuja** SAS-koodissa annettu muuttujan nimi:
- Yritystunnus YRTUNNUS
- Jalostusarvo JALEU
- Yrityksen päätoimiala TOL 95



Kuvio 3.1. Aineistona käytettävien tietokantojen ja rekisterien välisiä yhteyksiä.

VAIN RAKENNETILASTOJEN KYSELYKANNASSA – TKANNASSA - OLEVAT MUUTTUJAT (TOIMIPAIKKATASOLLA), yksitoimipaikkaisille yrityksille sekä monitoimipaikkaisten yritysten teollisuuden ja rakentamisen toimipaikoille (Tilastokeskus, 2002 b)

- **Muuttuja** SAS-koodissa annettu muuttujan nimi:
- Toimipaikkatunnus TPTUNN
- Jalostusarvo JALEUTP
- Toimipaikan toimiala TOL95TPM

**TOIMIPAIKKAREKISTERISSÄ OLEVAT MUUTTUJAT (vain ne, joita tutkimuksessa käytetään) (TOIMIPAIKKATASOLLA), kaikki toimialat (Tilastokeskus, 2002 d, Tilastokeskus, 2002 b):**

| <b>Muuttuja</b>   | <b>SAS-koodissa annettu muuttujan nimi:</b> |
|---|---|
| • Toimipaikan toimipaikkatunnus   | YKSTUN                                      |
| • Toimipaikan sijaintikuntalääni  | ULAANI                                      |
| • Toimipaikan toimiala  | NTALATP                                     |
| • Toimipaikan liikevaihto (osalla toimipaikoista estimoitu tieto)       | LVMK  |
| • Toimipaikan henkilöstön määrä (osalla toimipaikoista estimoitu tieto) | HK2   |
| • Toimipaikkaa vastaavan yrityksen PK-suuruusluokka                     | PKSLY                                       |
| • Toimipaikalle estimoidut palkat                                       | PALKKAMK                                    |

**TUTKIMUKSESSA KÄYTETTÄVÄT YRITYS- JA TOIMIPAIKKAREKISTERISSÄ JA RAKENNETILASTOJEN KYSEL YKANNASSA OLEVIEN MUUTTUJIEN POHJALTA MUODOSTETUT MUUTTUJAT:**

| <b>Muuttuja:</b>  | <b>Muodostamisessa tarvittavat muuttujat:</b> | <b>Muuttujan lyhenne:</b> |
|---|---|---------------------------|
| Toimipaikan toimialan pääluokka                                   | TOL95TMP ja NTALATP                           | PAALUO                    |
| Toimipaikan toimialan pääluokka TKANNASSA                         | TOL95TMP                                      | PAALUO2                   |
| Toimipaikan toimialan pääluokka YTR:ssä                           | NTALATP                                       | PAALUO3                   |
| Toimipaikkaa vastaavan yrityksen päätoimiala                      | TOL95 ja NTALAYR                              | PAALUOYR                  |
| Toimipaikkaa vastaavan yrityksen päätoimiala YKANNASSA            | TOL95   | PAALUOYR2                 |
| Toimipaikkaa vastaavan yrityksen päätoimiala YTR:ssä              | NTALAYR                                       | PAALUOYR3                 |
| Toimipaikkaa vastaavien yritysten jako PK- ja suuriin yrityksiin  | PKSLY   | PKSLY2                    |
| Toimipaikkaa vastaavan yrityksen laji toimipaikkojen lkm:n mukaan | TPLKM   | YKSITOIMIPAIKKAINEN       |
| Ositetieto  | PAALUO, YKSITOIMIPAIKKAINEN ja PKSLY2         | OSITE                     |

Jalostusarvo (JALEU, JALEUTP), liikevaihto (LVMK, LVMKYR), toimipaikalle estimoidut palkat (PALKKAMK), yrityksen maksamat palkat (PALKKAMKYR) ja henkilöstön määrä (HK2, HK2YRIT) ovat suhdeasteikollisia muuttujia. Yrityksen PK-suuruusluokka (PKSLYR, PKSLY) ja toimipaikkaa vastaavan yrityksen jako PK- ja suuriin yrityksiin (PKSLY2) on järjestysasteikon muuttuja. Kaikki muut tutkimuksessa käytetyt muuttujat ovat luokiteluasteikon muuttujia.

## 4 Liikeryitysten toimipaikat Suomessa

Luvut 4.1 ja 4.2 kertovat tutkimuksen perusjoukon eli Suomessa vuonna 2000 toimineiden liikeryitysten toimipaikkojen (liiketoimipaikkojen) ominaisuuksista. Luvussa 4.2 keskitytään asiaan, joka vaikuttaa toimipaikkojen tilinpäätöstietojen toimipaikkaistamisen pohjalla olevien tilastollisten mallien valintaan – multikollineaarisuuteen. Luku 4 on kokonaisuudessaan pohjatietoa luvulle 5, koska se käsittelee menetelmiä, joilla toimipaikkaistaminen tehdään. Perusjoukon ominaisuudet vaikuttivat paljon siihen mitkä menetelmät valittiin.

## 4.1 Liiketoimipaikat vuonna 2000

Liiketoimipaikat vuonna 2000 olivat suurimmaksi osaksi yksitoimipaikkaisten yritysten toimipaikkoja eli yksitoimipaikkaisia liikeyrityksiä (joille yrityksen tiedot = toimipaikan tiedot) ja pieniä kooltaan. Yksitoimipaikkaisten liikeyritysten toimipaikkoja rakennetilastojen kyselykannan ja yritys- ja toimipaikkarekisterin liiketoimipaikoista oli vuonna 2000 87 % ja monitoimipaikkaisten yritysten liiketoimipaikkoja 13 %. Suurin osa toimipaikoista oli mikroyritysten toimipaikkoja (n. 85% kaikista toimipaikoista). PK-yritysten toimipaikkoja oli vuonna 2000 n. 93% kaikista liiketoimipaikoista.

Toimipaikan toimiala määritellään erilailla yritys- ja toimipaikkarekisterissä ja rakennetilastojen kyselykannassa (rekisterien yhdistämiseen liittyvistä ongelmista on kerrottu enemmän luvuissa 3.1 ja 8). Koska toimipaikkojen tilinpäätöstiedot on saatu rakennetilastojen kyselykannasta, niin yritys- ja toimipaikkarekisteriä yhdistettäessä on toimipaikan toimiala määritelty rakennetilastojen kyselykannan mukaan, jos toimipaikka on TKANNASSA ja muutoin toimipaikkarekisterin mukaan.

Toimipaikkojen toimialojen pääluokat ovat taulukossa 4.1. Suluissa on mainittu niiden toimipaikkojen määrät ja osuudet, joille kaikki toimipaikan tilinpäätöstiedot ovat tiedossa. Suurin osa liiketoimipaikoista vuonna 2000 toimi taulukon 4.1 perusteella toimialalla G eli tukku- ja vähittäiskaupassa. Toiseksi eniten liiketoimipaikkoja oli toimialalla K eli kiinteistö-, vuokraus- ja tutkimuspalveluissa. Kolmanneksi eniten liiketoimipaikkoja oli toimialalla F eli rakentamisessa ja neljänneksi eniten toimialalla D eli teollisuudessa. Suurimmat toimialat vuonna 2000 olivat siis G, K, F ja D. Yksitoimipaikkaisten liikeyritysten osuudet kaikista toimipaikoista vaihtelivat toimialoittain ja olivat välillä 63.46 % ja 95.74 %. Eniten yksitoimipaikkaisia liikeyrityksiä oli toimialalla F (rakentamisessa). Myös toimialoilla O (muut yhteiskunnalliset ja henkilökohtaiset palvelut), N (terveydenhuolto- ja sosiaalipalvelut) ja K (kiinteistö-, vuokraus- ja tutkimuspalvelut) oli paljon yksitoimipaikkaisia yrityksiä. Kaikkein eniten monitoimipaikkaisia yrityksiä oli toimialalla E (sähkön-, kaasun- ja vesihuolto). Paljon monitoimipaikkaisia yrityksiä oli myös toimialalla G, H ja J.

Taulukosta 4.1 havaitaan myös, että toimialoilla L, P ja Q ei toiminut yhtään liiketoimipaikkaa tai liikeyritystä vuonna 2000 yritys- ja toimipaikkarekisterin eikä rakennetilastojen kyselykannan mukaan. Siksi tutkimustuloksia ei voidakaan käyttää jatkossa hyväksi, jos toimipaikka toimii toimialalla L, P tai Q. Yhdenkään liiketoimipaikkaa vastaavan yrityksen päätoimiala ei myöskään ollut L, P tai Q yritysrekisterin eikä rakennetilastojen kyselykannan mukaan. Yhdenkään monitoimipaikkaisen yrityksen toimipaikan toimiala ei ollut tuntematon (X). Osajoukkoja tutkimuksessa on siis 29 (katso luku 5.2).

Taulukko 4.2 kertoo toimipaikkojen jakautumisesta eri lääneihin toimialoittain. Luvut on laskettu käyttämällä kaikkien toimipaikkojen tietoja ja suluissa on alueellinen jakautuminen niiden toimipaikkojen joukossa, joille laskennallisen käyttökatteen tiedot ovat tiedossa. Taulukko 4.3 kuvaa toimipaikkojen toimialojen jakaumaa lääneittäin.

Toimipaikkojen sijainnissa on havaittavissa keskittymistä Etelä- ja Länsi-Suomen lääneihin, sillä taulukon 4.2 perusteella vuonna 2000 Etelä-Suomen tai Länsi-Suomen läänissä sijaitsi n. 78 % kaikista liiketoimipaikoista (n. 42 % kaikista liiketoimipaikoista vuonna 2000 sijaitsi Etelä-Suomen läänissä ja n. 37 % Länsi-Suomen läänissä, n. 10 % Itä-Suomen läänissä ja n. 7 % Oulun läänissä). Uudellamaalla sijaitsi n. 28 % kaikista liiketoimipaikoista, Varsinais-Suomessa n. 9 % ja Pohjois-Pohjanmaalla n. 6 % toimipaikoista. Jos tarkastellaan vain niitä liiketoimipaikkoja, joille kaikki laskennallisen käyttökateen tiedot ovat tiedossa, niin liiketoimipaikkojen alueellinen jakautuminen on melkein samanlainen kuin kaikkien toimipaikkojen joukossa. Kuitenkin vaikuttaa siltä, että Etelä- ja Länsi-Suomen liiketoimipaikoista useammalla on tiedossa laskennallisen käyttökateen tiedot kuin muissa lääneissä sijaitseville liiketoimipaikoilla. Tämä saattaa johtua siitä, että Etelä- tai Länsi-Suomessa sijaitsevista liiketoimipaikoista suurempi osa on yksitoimipaikkaisten liikeyritysten toimipaikkoja kuin muissa lääneissä sijaitsevista. Toimipaikkoja vastaavien yritysten kotipaikka oli myös yleensä Etelä- tai Länsi-Suomen läänissä (n. 45 %:lla toimipaikoista yrityksen kotipaikka oli Etelä-Suomen läänissä ja 35 % Länsi-Suomen läänissä, n. 32%:lla toimipaikoista yrityksen kotipaikka oli Uudellamaalla).

Toimipaikkojen keskittyminen Etelä- ja Länsi-Suomen lääneihin on havaittavissa myös jokaisella toimialalla erikseen. Taulukosta 4.2 huomataan, että kaikilla toimialoilla toimipaikat keskittyivät Etelä- tai Länsi-Suomen lääneihin. Toimialoilla AE (eli alkutuotannon ja jalostuksen toimialoilla lukuun ottamatta rakentamista) suurin osa toimipaikoista sijaitsi Länsi-Suomen läänissä. Toimialoilla F-O (eli rakentamisessa ja palvelualoilla) suurin osa toimipaikoista sijaitsi vuonna 2000 Etelä-Suomen läänissä. Maakunnista eniten toimipaikkoja oli Uudellamaalla. Erityinen piirre toimipaikkojen sijainnissa on se, että lähes kaikilla toimialoilla toimipaikat olivat vuonna 2000 keskittyneitä Uudellemaalle, Pirkanmaalle ja Varsinais-Suomeen (toimialat D ja F-O). Pohjanmaalle ja Keski-Suomeen olivat keskittyneet mineraalien kaivuun toimialan toimipaikat sekä Pohjanmaalle ja Varsinais-Suomeen toimialojen B ja E toimipaikat.

Tämä antaa aiheen olettaa, että suurin osa Suomen liiketoimipaikoista sijaitsee aika suurissa kaupungeissa tai kunnissa, koska suurin osa Suomen suurista kunnista ja kaupungeista on Etelä- ja Länsi-Suomen lääneissä. Huomataan myös se, että toimipaikkaa vastaava yrityskin sijaitsee useimmiten Etelä- tai Länsi-Suomen läänissä, itse asiassa toimipaikkaa vastaavista yrityksistä suurempi osa pitää kotipaikkanaan Etelä-Suomen lääniä ja siellä Uuttamaata kuin toimipaikoista.

Monitoimipaikkaisten liikeyritysten toimipaikoista n. 76 % toimi Etelä- tai Länsi-Suomen läänissä (n. 41% Etelä-Suomen läänissä ja 34 % Länsi-Suomen läänissä. Uudellamaalla toimi peräti 26% monitoimipaikkaisten liikeyritysten toimipaikoista). Monitoimipaikkaisten liikeyritysten kotipaikka oli yleensä Etelä-Suomen läänissä (Etelä- tai Länsi-Suomen lääni oli kotipaikkana n. 86% monitoimipaikkaisista liikeyrityksistä ja peräti n. 55 % oli kotipaikkana Uusimaa). Yksitoimipaikkaiset liikeyritykset toimivat myöskin Etelä- tai Länsi-Suomen läänissä (n. 79 % yksitoimipaikkaisista liikeyrityksistä, n. 42 % toimi Etelä-Suomen läänissä, 37 % Länsi-Suomen läänissä ja n. 10 % Itä-Suomen läänissä. Peräti n. 28 % yksitoimipaikkaisista liikeyrityksistä toimi Uudellamaalla ja n. 9 % Pirkanmaalla).

**Taulukko 4.1.** Toimipaikkojen toimialojen pääluokat sekä yksi- ja monitoimipaikkaisten yritysten toimipaikkojen osuudet toimialoittain

| <i>TOIMIALA</i>   | <i>N</i>         | <i>%</i>     | <i>YKSITOIMI-<br/>PAIKKAISIA</i> | <i>YKSITOIMI-<br/>PAIKKAISET %</i> | <i>MONITOIMI-<br/>PAIKKAISIA</i> | <i>MONITOIMIPAIK-<br/>KAISET %</i> |
|---|------------------|--------------|----------------------------------|------------------------------------|----------------------------------|------------------------------------|
| <i>A Maatalous, riistata-<br/>lous ja metsätalous</i>           | 7106 (6222)      | 2.94(2.92)   | 6223 (6222)                      | 87.57 (100)                        | 883 (0)                          | 12.43 (0)                          |
| <i>B Kalatalous</i>   | 669(626)         | 0.28(0.29)   | 626 (626)                        | 93.57 (100)                        | 43 (0)                           | 6.43 (0)                           |
| <i>C Mineraalien kaivuu</i>                                     | 1306 (1215)      | 0.54 (0.57)  | 1116 (1116)                      | 85.45 (91.85)                      | 190 (99)                         | 14.55 (8.15)                       |
| <i>D Teollisuus</i>   | 28039 (26457)    | 11.59(12.40) | 24419(24417)                     | 87.09 (92.29)                      | 3620 (2040)                      | 12.91 (7.71)                       |
| <i>E Sähkön-, kaasun-<br/>ja vesihuolto</i>                     | 1185 (950)       | 0.49 (0.45)  | 752 (752)                        | 63.46 (79.16)                      | 433 (198)                        | 36.54 (20.84)                      |
| <i>F Rakentaminen</i>   | 29491 (28 684)   | 12.19(13.44) | 28236(28235)                     | 95.74 (98.43)                      | 1255 (449)                       | 4.26 (1.57)                        |
| <i>G Tukku- ja vähittäis-<br/>kauppa</i>                        | 570 11 (45 071)  | 23.77(21.12) | 45079(45071)                     | 78.38 (100)                        | 12 432(0)                        | 21.62(0)                           |
| <i>H Majoitus- ja ravit-<br/>semistoiminta</i>                  | 12 920 (9416)    | 5.34 (4.41)  | 9418 (9416)                      | 72.89 (100)                        | 3502 (0)                         | 27.11 (0)                          |
| <i>I Kuljetus-, varastoin-<br/>ti- ja tietoliikenne</i>         | 26030 (22 794)   | 10.76(10.68) | 22794(22794)                     | 87.57 (100)                        | 3236 (0)                         | 12.43 (0)                          |
| <i>J Rahoitustoiminta</i>                                       | 2794 (2365)      | 1.15 (1.11)  | 2366 (2365)                      | 84.68 (100)                        | 428 (0)                          | 15.32 (0)                          |
| <i>K Kiinteistö-, vuok-<br/>raus- ja tutkimuspal-<br/>velut</i> | 42887 (39 561)   | 17.72(18.53) | 39569(39561)                     | 92.26 (100)                        | 3318 (0)                         | 7.74 (0)                           |
| <i>M Koulutus</i>   | 1687 (1468)      | 0.70 (0.69)  | 1468 (1468)                      | 87.02 (100)                        | 219 (0)                          | 12.98 (0)                          |
| <i>N Terveystieteiden<br/>ja sosiaalipalvelut</i>               | 12 985 (12 008)  | 5.37 (5.63)  | 12008(12008)                     | 92.48 (100)                        | 977 (0)                          | 7.52 (0)                           |
| <i>O Muut palvelut</i>  | 17 366 (16 595)  | 7.18 (7.77)  | 16596 (16595)                    | 95.57 (100)                        | 770 (0)                          | 4.43 (0)                           |
| <i>X Toimiala tuntema-<br/>ton</i>                              | 11 (11)          | 0.00 (0.01)  | 11 (11)                          | 100 (100)                          | 0 (0)                            | 0.00 (0)                           |
| <b>YHTEENSÄ</b>   | 241 987 (213443) | 100(88.204)  | 210681(210657)                   | 87.06 (98.69)                      | 31306 (2786)                     | 12.94 (1.31)                       |



Taulukosta 4.3 huomataan, että Etelä-Suomen läänissä ja Ahvenanmaalla sijaitsevat toimipaikat ovat keskittyneet toimialoille G, K ja F. Länsi-Suomen läänissä ja Oulun läänissä sijaitsevat toimipaikat ovat puolestaan keskittyneet toimialoille G, K ja D ja Itä-Suomen läänin toimipaikat ovat keskittyneet toimialoille G, K, I ja D. Lapsissa toimivat lähinnä toimialojen G, I ja K toimipaikat. Eli kaikissa muissa lääneissä paitsi Itä-Suomen läänissä oli eniten toimialojen G ja K toimipaikkoja ja Itä-Suomen läänissäkin on kolmanneksi eniten toimialan K toimipaikkoja. Toisin sanoen näyttää siltä, että toimipaikkojen lukumäärältään suurimmat toimialat ovat Suomessa ja kaikissa lääneissäkin toimialat G (n. 24 % kaikista toimipaikoista) ja K (n. 18 % kaikista toimipaikoista). Teollisuus ja rakentaminen ovat myös suuria toimialoja, mutta niiden toimipaikat ovat suurelta osaltaan keskittyneet Etelä- ja Länsi-Suomen lääneihin. Kaikissa lääneissä toimii kuitenkin kaikkien toimialojen toimipaikkoja.

Se miksi yritykset toimivat suurissa kaupungeissa, johtuu osaltaan siitä, että niissä on osatyömarkkinoiden koko, eli tietyllä alueella ja tietyllä toimialalla toimivien yritysten ja työntekijöiden määrä, suurempi kuin muualla maassa. Ja koska suurilla alueilla on enemmän työtä etsiviä työntekijöitä kuin pienillä kannattaa yrityksen harjoittaa toimintaansa suurilla alueilla, koska näillä alueilla rekrytointiprosessi on helpompi. Toisaalta se, että yritykset toimivat suurissa kaupungeissa aiheuttaa sen, että niissä on parempi työvoiman saatavuus kuin muualla maassa. Työntekijät hakeutuvat sinne missä työtä on ts. työvoimaa on siellä missä sitä tarvitsevia yrityksiä tai toimipaikkojakin. Alueellisiin eroihin vaikuttavat siis kokonaisuudessaan siellä toimivien yritysten/ toimipaikkojen määrä ja toimialat sekä työvoiman määrä ja laatu ts. alueen työvoimatilanne. (Rantala, 1999, 17-18, 26).

Toimipaikkojen tilinpäätöstietojen aritmeettiset keskiarvot ovat suurempia kuin mediaanit ja mediaanit suurempia kuin moodit. Se johtuu siitä, että toimipaikkojen tilinpäätöstietojen jakaumat ovat oikealle vinoja (lukuun ottamatta toimialaa X) ja pieniä havaintoarvoja on suuria enemmän. Myös toimipaikkojen tilinpäätöstietojen huippukuusluvut ovat positiivisia eli toimipaikkojen tilinpäätöstietojen jakaumilla on normaalia korkeammat huiput. Toimipaikkojen tilinpäätöstiedot vaihtelevat myös toimipaikan toimialan ja sijainnin mukaan. Tilinpäätöstietojen jakaumat ovat kuitenkin oikealle vinoja toimialojen sisälläkin. Täten voidaan pitää mediaania parempana keskimääräisen Suomessa vuonna 2000 toimineen liiketoimipaikan tilinpäätöstiedon mittana kuin aritmeettista keskiarvoa ja voidaan ajatella etteivät toimipaikkojen tilinpäätöstietojen jakaumat ole välttämättä normaalisia.

Toimipaikan toimialasta riippumatta yksitoimipaikkaiset liikeyritykset olivat pienempiä kooltaan, toiminnan laajuudeltaan ja merkittävyydeltään kuin monitoimipaikkaisten yritysten toimipaikat, sillä yksitoimipaikkaisten yritysten tilinpäätöstietojen aritmeettiset keskiarvot sekä mediaanit olivat pienempiä kuin monitoimipaikkaisten yritysten toimipaikkojen toimialasta riippumatta.

Myös jalostusarvon havaintoarvoissa oli samanlainen vaihtelu kuin liikevaihdossa, keskimääräisessä henkilöstömäärässä ja toimipaikalle estimoiduissa palkoissa niissä toimipaikoissa, joiden jalostusarvot olivat tiedossa (eli 88 % toimipaikoista). Monitoimipaikkaisten yritysten toimipaikkojen tilinpäätöstietojen aritmeettiset keskiarvot ja mediaanit olivat siis suurempia, kuin jos yksitoimipaikkaisten yritysten tilinpäätöstietojen. Tämä ero näkyi toimialojen sisälläkin (esim. monitoimipaikkaisten yritysten teollisuuden toimipaikkojen tilinpäätöstietojen aritmeettiset keskiarvot ja mediaanit olivat suuremmat kuin yksitoimipaikkaisten liikeyritysten, joiden toimiala oli teollisuus). Yksi- ja monitoimipaikkaisten yritysten tilinpäätöstiedoissa oleva vaihtelu ei ollut yhtä

suurta, vaan yksitoimipaikkaisilla liikeyrityksillä oli suurempi vaihtelu kuin monitoimipaikkaisten yritysten toimipaikoilla, koska yksitoimipaikkaisilla liikeyrityksillä olivat tilinpäätöstietojen variaatiokertoimet suurempia kuin monitoimipaikkaisten yritysten toimipaikkojen tilinpäätöstietojen variaatiokertoimet.

**Taulukko 4.2.** Toimipaikkojen alueellinen jakauma (suluissa jakauma niiden toimipaikkojen joukossa, joiden jalostusarvo tunnetaan)

| <i>TOIMI-<br/>ALA</i> | <i>ETELÄ-<br/>SUOMEN LÄÄNI<br/>N %</i> | <i>LÄNSI-<br/>SUOMEN<br/>LÄÄNI N %</i> | <i>ITÄ-SUOMEN<br/>LÄÄNI N %</i> | <i>OULUN<br/>LÄÄNIN %</i>  | <i>LAPIN LÄÄNI N<br/>%</i> | <i>AHVENAN-<br/>MAA N%</i> | <i>YHT.</i>                   |
|-----------------------|--|--|---------------------------------|----------------------------|----------------------------|----------------------------|-------------------------------|
| <b>A</b>              | 1239,17.4%<br>(1070,17.2%)             | 3566,50.2%<br>(3257,52.4 %)            | 1227,17.3%<br>(1001,16.1%)      | 705,9.9%<br>(589,9.5%)     | 326,4.6%<br>(262,4.2%)     | 43,0.6%<br>(43,0.7%)       | 7106,2.9%<br>(6222,2.9%)      |
| <b>B</b>              | 117,17.5%<br>(112,17.9 %)              | 299,44.7%<br>(280,44.7%)               | 60,9.0%<br>(55,8.8%)            | 104,15.6%<br>(99,15.8%)    | 44,6.6%<br>(37,6%)         | 45,6.7%<br>(43,6.8%)       | 669,0.3%<br>(626,0.3%)        |
| <b>C</b>              | 185,14.2%<br>(175,14.4 %)              | 564,43.2%<br>(516, 42.5 %)             | 197,15.1%<br>(184, 15.1%)       | 287,22.0%<br>(272,22.4%)   | 70,5.4%<br>(65,5.4%)       | 3,0.2%<br>(3,0.3%)         | 1306,0.5%<br>(1215,0.6%)      |
| <b>D</b>              | 9786,34.9%<br>(9144,34.6%)             | 12485,44.5%<br>(11893,45.0%)           | 2853,10.2%<br>(2712,10.3 %)     | 1901,6.8%<br>(1754,6.6%)   | 853,3.0%<br>(799,3.0%)     | 161,0.6%<br>(155,0.6%)     | 28039,11.6%<br>(26457,12.4%)  |
| <b>E</b>              | 274,23.1%<br>(219, 23. 1%)             | 461,38.9%<br>(392,41.3%)               | 176,14.9%<br>(118, 12.4%)       | 162,13.7%<br>(130,13.7%)   | 100,8.4%<br>(79, 8.3%)     | 12,1.0%<br>(12,1.3%)       | 1185,0.5%<br>(950,0.5%)       |
| <b>F</b>              | 11876,40.3%<br>(11617,40.5 %)          | 11278,38.2%<br>(10976,38.3%)           | 2827,9.6%<br>(2711,9.5 %)       | 2134,7.2%<br>(2046,7.1%)   | 1084,3.7%<br>(1043,3.6%)   | 292,1.0%<br>(291,1.0%)     | 29491,12.2%<br>(28684,13.4%)  |
| <b>G</b>              | 24835,43.2%<br>(19650,43.6%)           | 20684,36.0%<br>(16377,36.3%)           | 5745,10.0%<br>(4328, 9.6 %)     | 3987,6.9%<br>(2973,6.6%)   | 1866,3.2%<br>(1396, 3.1%)  | 394,0.7%<br>(347,0.8%)     | 57511,23.8%<br>(45071,21.1%)  |
| <b>H</b>              | 5247,40.6%<br>(3555,37.8%)             | 4436,34.3%<br>(3412, 36.2%)            | 1417,11.0%<br>(1047,11.1%)      | 998,7.7%<br>(727,7.7%)     | 671,5.2%<br>(536, 5.7%)    | 151,1.2%<br>(139,1.5%)     | 12920,5.3%<br>(9416, 4.4%)    |
| <b>I</b>              | 9654,37.1%<br>(8357,36.7%)             | 9400,36.1%<br>(8293,36.4 %)            | 3209,12.3%<br>(2833,12.4%)      | 2245,8.6%<br>(1979,8.7%)   | 1330,5.1%<br>(1156,5.1%)   | 192,0.7%<br>(176,0.8%)     | 26030,10.8%<br>(22794,10.7%)  |
| <b>J</b>              | 1603,57.4%<br>(1474,62.3 %)            | 830,29.7%<br>(604,25.5%)               | 153,5.5%<br>(115,4.9 %)         | 105,3.8%<br>(84,3.6%)      | 59,2.1%<br>(48,2.0%)       | 44,1.6%<br>(40,1.7%)       | 2794,1.2%<br>(2365, 1.1%)     |
| <b>K</b>              | 22402,52.2%<br>(20906,52.8%)           | 13294,31.0%<br>(12187,30.8%)           | 3258,7.6%<br>(2965,7.5%)        | 2461,5.7%<br>(2156,5.5%)   | 1255,2.9%<br>(1139,2.9%)   | 217,0.5%<br>(208,0.5%)     | 42887,17.7%<br>(39561,18.5%)  |
| <b>M</b>              | 817,48.4%<br>(732, 49.9 %)             | 555,32.9%<br>(487, 33.2 %)             | 137,8.1%<br>(104,7.1 %)         | 117,6.9%<br>(89,6.1 %)     | 51,3.0%<br>(46,3.1 %)      | 10,0.6%<br>(10,0.7%)       | 1687,0.7%<br>(1468 ,0.7%)     |
| <b>N</b>              | 5794,44.6%<br>(5314,44.3 %)            | 4316,33.2%<br>(4038,33.6%)             | 1394,10.7%<br>(1281,10.7 %)     | 1076,8.3%<br>(997,8.3 %)   | 365,2.8%<br>(339, 2.8%)    | 40,0.3%<br>(39 ,0.3 %)     | 12985,5.4%<br>(12008, 5.6%)   |
| <b>O</b>              | 7386,42.5%<br>(6980,42.1%)             | 6117,35.2%<br>(5885,35.5%)             | 1849,10.7%<br>(1780,10.7 %)     | 1250,7.2%<br>(1215,7.3 %)  | 666,3.8%<br>(642, 3.9 %)   | 98,0.6%<br>(93,0.6%)       | 17366,7.2%<br>(16595, 7.8%)   |
| <b>X</b>              | 5,45.5%<br>(5,45.5 %)                  | 3,27.3%<br>(3,27.3%)                   | 0, 0.0%(0,0.0 %)                | 2,18.2%<br>(2,18.2%)       | 1,9.1%(1,9.1%)             | 0,0.0%<br>(0,0.0%)         | 11,0.0%<br>(11,0.01%)         |
| <b>YHT.</b>           | 101220,41.8%<br>(89310,41.8%)          | 88288,36.5%<br>(78600,36.8%)           | 24502,10.1%<br>(21234,10.0 %)   | 17534,7.3%<br>(15112,7.1%) | 8741,3.6%<br>(7588,3.6%)   | 1702,0.7%<br>(1599,0.8)    | 241987,100%<br>(213443, 100%) |

Taulukko 4.3. Toimipaikkojen jakautuminen eri toimialoille alueellisesti (ts. toimialajakaumat lääneittäin)

| TOI-<br>MI-<br>ALA | Etelä-<br>Suomen<br>lääni | % Etelä-<br>Suomen<br>toimipai-<br>koista | Länsi-<br>Suomen<br>lääni | % Länsi-<br>Suomen<br>toimipai-<br>koista | Itä-Suo-<br>men lääni | % Itä-<br>Suomen<br>toimipai-<br>koista | Oulun<br>lääni | % Oulun<br>läänin toi-<br>mipaikois-<br>ta | Lapin<br>lääni | % La-<br>pin toi-<br>mipai-<br>koista | Ahve-<br>nan-<br>maa | % Ahve-<br>nanmaan<br>toimipai-<br>koista |
|--------------------|---------------------------|---|---------------------------|---|-----------------------|---|----------------|--|----------------|---------------------------------------|----------------------|---|
| A                  | 1239                      | 1.2                                       | 3566                      | 4.0                                       | 1227                  | 5.0                                     | 705            | 4.0  | 326            | 3.7                                   | 43                   | 2.5                                       |
| B                  | 117                       | 0.1                                       | 299                       | 0.3                                       | 60                    | 0.3                                     | 104            | 0.6  | 44             | 0.5                                   | 45                   | 2.6                                       |
| C                  | 185                       | 0.2                                       | 564                       | 0.6                                       | 197                   | 0.8                                     | 287            | 1.6  | 70             | 0.8                                   | 3                    | 0.2                                       |
| D                  | 9786                      | 9.7                                       | 12485                     | 14.1                                      | 2853                  | 11.6                                    | 1901           | 10.8                                       | 853            | 9.8                                   | 161                  | 9.5                                       |
| E                  | 274                       | 0.3                                       | 461                       | 0.5                                       | 176                   | 0.7                                     | 162            | 0.9  | 100            | 1.1                                   | 12                   | 0.7                                       |
| F                  | 11876                     | 11.7                                      | 11278                     | 12.8                                      | 2827                  | 11.5                                    | 2134           | 12.2                                       | 1084           | 12.4                                  | 292                  | 17.2                                      |
| G                  | 24835                     | 24.5                                      | 20684                     | 23.4                                      | 5745                  | 23.5                                    | 3987           | 22.7                                       | 1866           | 21.4                                  | 394                  | 23.2                                      |
| H                  | 5247                      | 5.2                                       | 4436                      | 5.0                                       | 1417                  | 5.8                                     | 998            | 5.7  | 671            | 7.7                                   | 151                  | 8.9                                       |
| I                  | 9654                      | 9.5                                       | 9400                      | 10.7                                      | 3209                  | 13.1                                    | 2245           | 12.8                                       | 1330           | 15.2                                  | 192                  | 11.3                                      |
| J                  | 1603                      | 1.6                                       | 830                       | 0.9                                       | 153                   | 0.6                                     | 105            | 0.6  | 59             | 0.7                                   | 44                   | 2.6                                       |
| K                  | 22402                     | 22.1                                      | 13294                     | 15.1                                      | 3258                  | 13.3                                    | 2461           | 14.0                                       | 1255           | 14.4                                  | 217                  | 12.8                                      |
| M                  | 817                       | 0.8                                       | 555                       | 0.6                                       | 137                   | 0.6                                     | 117            | 0.7  | 51             | 0.6                                   | 10                   | 0.6                                       |
| N                  | 5794                      | 5.7                                       | 4316                      | 4.9                                       | 1394                  | 5.7                                     | 1076           | 6.1  | 365            | 4.2                                   | 40                   | 2.4                                       |
| O                  | 7386                      | 7.3                                       | 6117                      | 6.9                                       | 1849                  | 7.6                                     | 1250           | 7.1  | 666            | 7.6                                   | 98                   | 5.8                                       |
| X                  | 5                         | 0.0                                       | 3                         | 0.0                                       | 0                     | 0                                       | 2              | 0.0  | 1              | 0.0                                   | 0                    | 0   |
| YHT.               | 101220                    | 100                                       | 88288                     | 100                                       | 24502                 | 100                                     | 17534          | 100  | 8741           | 100                                   | 1702                 | 100                                       |

Toimipaikkojen tilinpäätöstietojen mediaanit vaihtelivat toimialoittain siten, että

- Keskimääräisen henkilöstömäärän mediaanit olivat välillä 0.1 ja 1.9. Suurin mediaani oli toimialalla H eli majoitus- ja ravitsemistoiminnassa. Toiseksi suurin mediaani oli toimialalla D ja kolmanneksi suurin toimialalla I. Tämä antaa aiheen olettaa, että nämä toimialat (H, D ja I) ovat muita toimialoja työvoimavaltaisempia. Keskimääräiset henkilöstömäärät olivat vuonna 2000 kuitenkin pieniä, koska suurin osa liiketoimipaikoista oli vuonna 2000 alle 4.5 henkilöä työllistäviä (n. 83 % liiketoimipaikoista).

- Toimipaikan liikevaihdon mediaanit olivat välillä 108 000 ja 1500 000 mk. Toimialalla J eli rahoitustoiminnassa ei liikevaihtoa määritellä, vaan se on kaikille toimialalla J toimiville toimipaikoille 0. Suurin liikevaihdon mediaani oli toimialalla E eli sähkön-, kaasun- ja vesihuollossa, toiseksi suurin toimialalla G eli tukku- ja vähittäiskaupassa (975 000 mk) ja kolmanneksi suurin toimialalla D eli teollisuudessa (650 000 mk).

- Estimoidut toimipaikan maksamien palkkojen mediaanit olivat välillä 0 ja 136 000 mk. Suurimmat palkkakulut olivat toimialalla C eli mineraalien kaivuussa, toiseksi suurimmat toimialalla H eli majoitus- ja ravitsemistoiminnassa ja kolmanneksi suurimmat toimialalla E eli sähkön-, kaasun- ja vesihuollossa. Pienimmät palkkakulut olivat toimialoilla B, N, O ja X.

- Jalostusarvon mediaanit olivat välillä 64 000 mk ja 288 000 mk (mukana ne toimipaikat, joille jalostusarvo on tiedossa eli n. 88 % liiketoimipaikoista). Suurimmat jalostusarvojen mediaanit olivat toimialalla I eli kuljetuksen, varastoinnin ja tietoliikenteen toimialalla sekä teollisuuden ja rakentamisen toimialoilla (toimialoilla C-F). Kaikkein suurin jalostusarvon mediaani oli toimialalla I. Pienimmät jalostusarvojen mediaanit olivat toimialoilla X (64000 mk), toimialalla B (80 000 mk) ja toimialalla O (82 000 mk). Teollisuuden ja rakentamisen sekä toimialan I toimipaikkoja voidaan pitää pääomavaltaisina toimialoina ja toimialoja B, O ja X työvoimavaltaisina toimialoina.

## 4.2 Tilinpäätöstietojen yhteisvaihtelu ja multikollineaarisuus aineistossa

Toimipaikan tilinpäätöstiedot ovat vahvasti sidoksissa toisiinsa ja niillä on yhteisvaihtelua, esimerkiksi toimipaikan liikevaihto antaa tietoa siitä mitkä ovat toimipaikan henkilöstökulut. Aineistoa tutkittaessa huomattiin, että kaikkien tilinpäätöstietojen välillä vallitsi vahva lineaarinen riippuvuus, koska Pearsonin korrelaatiokertoimet tilinpäätöstietojen välillä olivat lähellä ykköstä. Tämä tarkoittaa myös sitä, että tutkimuksessa käytettävä lisäinformaatio ja tulosmuuttujat korreloivat vahvasti keskenään ja käytettävää lisäinformaatiota voidaan pitää tilastollisesti perusteltuna.

Kuitenkin voimakkaat korrelaatiot saattavat myös huonontaa toimialakohtaisten tilinpäätöstietojen estimaattien laatua, jos estimoinnissa käytettävät apumuuttujat korreloivat voimakkaasti keskenään. Tällöin on kyseessä ns. **multikollineaarisuuden** tilanne, eivätkä estimointitulokset ole välttämättä luotettavia. Tilannetta voidaan parantaa esimerkiksi poistamalla keskenään merkitsevästi korreloivista apumuuttujista toinen.

Perusjoukon suuresta koosta johtuen on luonnollista, että yritysaineistossa apumuuttujat korreloivat vahvasti keskenään. Tässä tutkimuksessa etenkin toimipaikan keskimääräisen henkilöstömäärän ja toimipaikalla estimoitujen palkkojen välinen Pearsonin korrelaatiokerroin on korkea. Täten voidaan ajatella, että ei ole tilastollisesti perusteltua käyttää sekä toimipaikan keskimääräistä henkilöstömäärää että toimipaikalle estimoituja palkkoja apumuuttujina, vaan vain toista niistä.

Taulukossa 4.4 ovat toimialakohtaiset Pearsonin korrelaatiokertoimet tutkimuksessa apumuuttujien välillä. Perusjoukon koon suuruuden vuoksi on luonnollista, että yritysaineistossa kaikki korrelaatiokertoimet olivat erittäin merkitsevästi nolasta poikkeavia kaikkien muuttujien välillä ja kaikilla toimialoilla. Vahvimmat riippuvuussuhteet ovat havaittavissa keskimääräisen henkilöstömäärän tilikauden aikana ja toimipaikalle estimoitujen palkkojen välillä, koska niiden välinen korrelaatiokerroin oli suurin kaikilla toimialoilla. Korrelaatiokertoimien voimakkuudessa on toimialojen välillä suuriakin eroja. Korrelaatiokertoimet on laskettu käyttäen kaikkia Suomen liiketoimipaikkoja, jotka ovat YTR:ssä ja rakennetilastojen kyselykannassa eli  $N = 241\,987$ .

Keskimääräisen henkilöstömäärän ja toimipaikalle estimoitujen palkkojen väliset Pearsonin korrelaatiokertoimet eri toimialoilla olivat välillä 0.36024 ja 0.99452, toimipaikan liikevaihdon ja keskimääräisen henkilöstömäärän 0.21158 ja 0.89642 sekä toimipaikan liikevaihdon ja toimipaikalle estimoitujen palkkojen 0.24138 ja 0.91328. Toimialoilla B, C, D, F, H, I ja N olivat kaikki apumuuttujien väliset korrelaatiokertoimet suurempia kuin kaikissa liiketoimipaikoissa. Muilla toimialoilla olivat ainakin jotkut apumuuttujien väliset korrelaatiokertoimet pienempiä kuin kaikissa liiketoimipaikoissa.

**Taulukko 4.4.** Henkilöstön määrän, liikevaihdon ja toimipaikoille estimoitujen palkkojen väliset Pearsonin korrelaatiokertoimet toimialoittain ja toimipaikkaa vastaavan yrityksen lajin mukaan

| <i>TOIMIALA</i> | <i>MUUTTUJA</i> | <i>HK2</i> | <i>LVMK</i> | <i>MONITOIMIPAIK-<br/>KAISSET HK2</i> | <i>MONITOIMIPAIKKAISET<br/>LVMK</i> | <i>YKSITOIMIPAIK-<br/>KAISSET HK2</i> | <i>YKSITOIMIPAIK-<br/>KAISSET LVMK</i> |
|-----------------|-----------------|------------|-------------|---------------------------------------|-------------------------------------|---------------------------------------|--|
| <b>A</b>        | LVMK            | 0.21       | 1           | 0.2                                   | 1                                   | 0.75                                  | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.24        | 0.99                                  | 0.24                                | 0.98                                  | 0.71                                   |
| <b>B</b>        | LVMK            | 0.79       | 1           | 0.91                                  | 1                                   | 0.82                                  | 1                                      |
|                 | PALKKAMK        | 0.95       | 0.77        | 0.96                                  | 0.93                                | 0.92                                  | 0.77                                   |
| <b>C</b>        | LVMK            | 0.76       | 1           | 0.72                                  | 1                                   | 0.92                                  | 1                                      |
|                 | PALKKAMK        | 0.97       | 0.73        | 0.97                                  | 0.69                                | 0.99                                  | 0.91                                   |
| <b>D</b>        | LVMK            | 0.62       | 1           | 0.63                                  | 1                                   | 0.8                                   | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.64        | 0.99                                  | 0.64                                | 0.97                                  | 0.81                                   |
| <b>E</b>        | LVMK            | 0.49       | 1           | 0.47                                  | 1                                   | 0.52                                  | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.52        | 0.99                                  | 0.5                                 | 0.995                                 | 0.55                                   |
| <b>F</b>        | LVMK            | 0.82       | 1           | 0.87                                  | 1                                   | 0.72                                  | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.85        | 0.997                                 | 0.88                                | 0.96                                  | 0.80                                   |
| <b>G</b>        | LVMK            | 0.47       | 1           | 0.46                                  | 1                                   | 0.48                                  | 1                                      |
|                 | PALKKAMK        | 0.93       | 0.52        | 0.93                                  | 0.51                                | 0.93                                  | 0.56                                   |
| <b>H</b>        | LVMK            | 0.84       | 1           | 0.81                                  | 1                                   | 0.89                                  | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.86        | 0.99                                  | 0.84                                | 0.99                                  | 0.90                                   |
| <b>I</b>        | LVMK            | 0.87       | 1           | 0.89                                  | 1                                   | 0.64                                  | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.90        | 0.99                                  | 0.90                                | 0.98                                  | 0.7                                    |
| <b>J</b>        | PALKKAMK        | 0.90       | *           | 0.94                                  | *                                   | 0.90                                  | *                                      |
| <b>K</b>        | LVMK            | 0.53       | 1           | 0.50                                  | 1                                   | 0.58                                  | 1                                      |
|                 | PALKKAMK        | 0.91       | 0.60        | 0.90                                  | 0.56                                | 0.92                                  | 0.66                                   |
| <b>M</b>        | LVMK            | 0.46       | 1           | 0.40                                  | 1                                   | 0.51                                  | 1                                      |
|                 | PALKKAMK        | 0.96       | 0.53        | 0.95                                  | 0.50                                | 0.97                                  | 0.55                                   |
| <b>N</b>        | LVMK            | 0.90       | 1           | 0.96                                  | 1                                   | 0.82                                  | 1                                      |
|                 | PALKKAMK        | 0.98       | 0.91        | 0.99                                  | 0.96                                | 0.96                                  | 0.86                                   |
| <b>O</b>        | LVMK            | 0.34       | 1           | 0.33                                  | 1                                   | 0.52                                  | 1                                      |
|                 | PALKKAMK        | 0.99       | 0.35        | 0.995                                 | 0.34                                | 0.94                                  | 0.63                                   |
| <b>X</b>        | LVMK            | 0.68       | 1           | 0.68                                  | 1                                   | .                                     | .                                      |
|                 | PALKKAMK        | 0.36       | 0.25        | 0.36                                  | 0.25                                | .                                     | .                                      |
| <b>KAIKKI</b>   | LVMK            | 0.57       | 1           | 0.58                                  | 1                                   | 0.57                                  | 1                                      |
|                 | PALKKAMK        | 0.98       | 0.59        | 0.99                                  | 0.59                                | 0.96                                  | 0.6                                    |

. = Puuttuva tieto, koska tutkimusaineistossa ei ole kyseiseen osajoukkoon kuuluvia toimipaikkoja.

\* = Liikevaihtoa ei ole kyseisellä toimialalla määritely, vaan se on aina arvoltaan nolla.

Siinä joukossa, josta ovat tiedossa kaikki toimipaikan tilinpäätöstiedot (n. 88 % kaikista liiketoimipaikoista, jotka ovat sekä YTR:ssä että rakennetilastojen kyselykannassa, N = 213 443) olivat apumuuttujien väliset Pearsonin korrelaatiokertoimet hiukan suuremmat kuin kaikkien liiketoimipaikkojen joukossa, mutta erot eivät olleet suuria. Nämä korrelaatiokertoimet ovat taulukossa 4.5. Eri toimialoilla olivat Pearsonin korrelaatiokertoimet yleensä suurempia niillä toimipaikoilla, joille olivat kaikki toimipaikan tilinpäätöstiedot tiedossa kuin kaikille liiketoimipaikoille. Toimialojen väliset erot apumuuttujien välisten riippuvuussuhteiden suuruudessa eli Pearsonin korrelaatiokertoimissa olivat samanlaiset kuin kaikilla liiketoimipaikoilla. Multikollinearisuus on siis ilmeisempää siinä liiketoimipaikkojen joukosta, jolle ovat tiedossa kaikki toimipaikkojen tilinpäätöstiedot.

Taulukossa 4.6 ovat jalostusarvon ja tutkimuksessa käytettävien apumuuttujien väliset korrelaatiokertoimet niiden toimipaikkojen joukossa, jonka jalostusarvo tunnetaan. Taulukosta huomataan, että jalostusarvo korreloi (ainakin niiden toimipaikkojen joukossa, joille jalostusarvo on tiedossa) vahvasti kaikkien apumuuttujien kanssa kaikilla toimialoilla. Mutta jalostusarvo korreloi liikevaihdon kanssa parhaiten ja henkilöstön määrän kanssa huonoiten. Kuitenkin se kuinka suuria apumuuttujien ja jalostusarvon väliset riippuvuudet ovat, on kiinni toimialasta ja toimipaikkaa vastaavan yrityksen lajista. Korrelaatiokertoimista voidaan päätellä, että paras yksittäinen selittäjä jalostusarvolle yksitoimipaikkaisten yritysten toimipaikkojen toimialojen A, B, D, E, G, K ja X toimipaikkojen tapauksessa on liikevaihto, toimialojen H, I, M, N ja O toimipaikkojen tapauksessa palkat ja toimialan C, F ja J tapauksessa henkilöstön määrä. Monitoimipaikkaisten yritysten toimipaikoille jalostusarvon paras yksittäinen selittäjä on toimialoilla C, D ja E palkat ja toimialalla F henkilöstön määrä. Koska korrelaatiokertoimet ovat kaikilla toimialoilla vahvoja, niin voidaan apumuuttujia pitää hyvinä selittäjinä jalostusarvolle.

Se kuinka hyviä apumuuttujat ovat jalostusarvon selittämisessä, riippuu kuitenkin toimialasta ja toimipaikkaa vastaavan yrityksen lajista (ainakin toimialoilla C-F ovat Pearsonin korrelaatiokertoimet kiinni toimipaikkaa vastaavan yrityksen lajista, ja siksi voidaan olettaa, että riippuvuussuhteet muuttujien välillä ovat erilaisia riippuen toimipaikkaa vastaavan yrityksen lajista muillakin toimialoilla). Toimialalla C monitoimipaikkaisten liikeyritysten toimipaikkojen ja toimialoilla X, K ja I yksitoimipaikkaisten liikeyritysten joukossa ovat korrelaatiokertoimet heikompia kuin muilla toimialoilla. Näissä joukoissa voidaan ajatella jalostusarvon ennustamisen onnistuvan huonommin kuin muilla toimialoilla.

**Taulukko 4. 5.** Pearsonin korrelaatiokertoimet niiden toimipaikkojen joukossa, joiden jalostusarvo on tiedossa.

| <i>TOIMIALA</i> | <i>MUUTTUJA</i> | <i>HK2</i> | <i>LVMK</i> | <i>MONITOIMIPAIKKAISET HK2</i> | <i>MONITOIMIPAIKKAISET LVMK</i> | <i>YKSITOIMIPAIKKAISET HK2</i> | <i>YKSITOIMIPAIKKAISET LVMK</i> |
|-----------------|-----------------|------------|-------------|--------------------------------|---------------------------------|--------------------------------|---------------------------------|
| <b>A</b>        | LVMK            | 0.75       | 1           | .                              | .                               | 0.75                           | 1                               |
|                 | PALKKAMK        | 0.98       | 0.71        | .                              | .                               | 0.98                           | 0.71                            |
| <b>B</b>        | LVMK            | 0.82       | 1           | .                              | .                               | 0.82                           | 1                               |
|                 | PALKKAMK        | 0.92       | 0.77        | .                              | .                               | 0.921                          | 0.77                            |
| <b>C</b>        | LVMK            | 0.78       | 1           | 0.73                           | 1                               | 0.92                           | 1                               |
|                 | PALKKAMK        | 0.98       | 0.75        | 0.97                           | 0.70                            | 0.99                           | 0.91                            |
| <b>D</b>        | LVMK            | 0.63       | 1           | 0.64                           | 1                               | 0.78                           | 1                               |
|                 | PALKKAMK        | 0.99       | 0.64        | 0.99                           | 0.64                            | 0.97                           | 0.81                            |
| <b>E</b>        | LVMK            | 0.52       | 1           | 0.51                           | 1                               | 0.52                           | 1                               |
|                 | PALKKAMK        | 0.99       | 0.56        | 0.99                           | 0.54                            | 0.995                          | 0.55                            |
| <b>F</b>        | LVMK            | 0.85       | 1           | 0.91                           | 1                               | 0.72                           | 1                               |
|                 | PALKKAMK        | 0.99       | 0.88        | 0.998                          | 0.92                            | 0.96                           | 0.80                            |
| <b>G</b>        | LVMK            | 0.48       | 1           | .                              | .                               | 0.48                           | 1                               |
|                 | PALKKAMK        | 0.94       | 0.56        | .                              | .                               | 0.94                           | 0.56                            |
| <b>H</b>        | LVMK            | 0.89       | 1           | .                              | .                               | 0.89                           | 1                               |
|                 | PALKKAMK        | 0.99       | 0.90        | .                              | .                               | 0.99                           | 0.90                            |
| <b>I</b>        | LVMK            | 0.64       | 1           | .                              | .                               | 0.64                           | 1                               |
|                 | PALKKAMK        | 0.98       | 0.70        | .                              | .                               | 0.98                           | 0.71                            |
| <b>J</b>        | PALKKAMK        | 0.90       | *           | .                              | .                               | 0.90                           | *                               |
| <b>K</b>        | LVMK            | 0.58       | 1           | .                              | .                               | 0.58                           | 1                               |
|                 | PALKKAMK        | 0.92       | 0.66        | .                              | .                               | 0.92                           | 0.66                            |
| <b>M</b>        | LVMK            | 0.51       | 1           | .                              | .                               | 0.51                           | 1                               |
|                 | PALKKAMK        | 0.97       | 0.55        | .                              | .                               | 0.97                           | 0.55                            |
| <b>N</b>        | LVMK            | 0.82       | 1           | .                              | .                               | 0.82                           | 1                               |
|                 | PALKKAMK        | 0.96       | 0.86        | .                              | .                               | 0.96                           | 0.86                            |
| <b>O</b>        | LVMK            | 0.52       | 1           | .                              | .                               | 0.52                           | 1                               |
|                 | PALKKAMK        | 0.94       | 0.63        | .                              | .                               | 0.94                           | 0.63                            |
| <b>X</b>        | LVMK            | 0.68       | 1           | .                              | .                               | 0.68                           | 1                               |
|                 | PALKKAMK        | 0.36       | 0.25        | .                              | .                               | 0.36                           | 0.25                            |
| <b>KAIKKI</b>   | LVMK            | 0.62       | 1           | 0.64                           | 1                               | 0.57                           | 1                               |
|                 | PALKKAMK        | 0.98       | 0.63        | 0.99                           | 0.64                            | 0.96                           | 0.60                            |

. = Puuttuva tieto, koska tutkimusaineistossa ei ole kyseiseen osajoukkoon kuuluvia toimipaikkoja.

\* = Liikevaihtoa ei ole kyseisellä toimialalla määritelty, vaan se on aina arvoltaan nolla.

Taulukko 4.6. Jalostusarvon ja apumuuttujien väliset korrelaatiokertoimet aineistossa perusjoukko

| TOIMI-<br>ALA | HK2   | LVMK | PALKKAMK | MONITOIMI-<br>PAIKKAISET<br>HK2 | MONITOIMI-<br>PAIKKAISET<br>LVMK | MONITOIMI-<br>PAIKKAISET<br>PALKKAMK | YKSITOIMI-<br>PAIKKAI-<br>SET HK2 | YKSITOIMI-<br>PAIKKAI-<br>SET LVMK | YKSITOIMI-<br>PAIKKAISET<br>PALKKAMK |
|---------------|-------|------|----------|---------------------------------|----------------------------------|--------------------------------------|-----------------------------------|------------------------------------|--------------------------------------|
| A             | 0.82  | 0.84 | 0.81     | .                               | .                                | .                                    | 0.82                              | 0.84                               | 0.81                                 |
| B             | 0.80  | 0.86 | 0.79     | .                               | .                                | .                                    | 0.80                              | 0.86                               | 0.79                                 |
| C             | 0.38  | 0.43 | 0.45     | 0.25                            | 0.33                             | 0.36                                 | 0.7                               | 0.76                               | 0.77                                 |
| D             | 0.74  | 0.89 | 0.76     | 0.76                            | 0.89                             | 0.77                                 | 0.73                              | 0.88                               | 0.77                                 |
| E             | 0.65  | 0.55 | 0.69     | 0.74                            | 0.47                             | 0.78                                 | 0.47                              | 0.74                               | 0.51                                 |
| F             | 0.84  | 0.78 | 0.82     | 0.84                            | 0.79                             | 0.84                                 | 0.88                              | 0.78                               | 0.84                                 |
| G             | 0.67  | 0.64 | 0.73     | .                               | .                                | .                                    | 0.67                              | 0.64                               | 0.73                                 |
| H             | 0.95  | 0.94 | 0.96     | .                               | .                                | .                                    | 0.95                              | 0.94                               | 0.96                                 |
| I             | 0.64  | 0.53 | 0.70     | .                               | .                                | .                                    | 0.64                              | 0.53                               | 0.70                                 |
| J             | 0.80  | *    | 0.68     | .                               | .                                | .                                    | 0.81                              | *                                  | 0.68                                 |
| K             | 0.51  | 0.62 | 0.56     | .                               | .                                | .                                    | 0.51                              | 0.62                               | 0.56                                 |
| M             | 0.92  | 0.59 | 0.94     | .                               | .                                | .                                    | 0.92                              | 0.59                               | 0.94                                 |
| N             | 0.90  | 0.88 | 0.93     | .                               | .                                | .                                    | 0.90                              | 0.88                               | 0.93                                 |
| O             | 0.83  | 0.60 | 0.86     | .                               | .                                | .                                    | 0.83                              | 0.60                               | 0.86                                 |
| X             | -0.05 | 0.35 | 0.12     | .                               | .                                | .                                    | -0.05                             | 0.35                               | 0.12                                 |
| KAIKKI        | 0.72  | 0.85 | 0.74     | 0.75                            | 0.89                             | 0.77                                 | 0.52                              | 0.44                               | 0.56                                 |

. = Puuttuva tieto, koska tutkimusaineistossa ei ole kyseiseen osajoukkoon kuuluvia toimipaikkoja.

\* = Liikevaihtoa ei ole kyseisellä toimialalla määritetty, vaan se on aina arvoltaan nolla.

## 5 Tutkimusmenetelmät

Luku 5 on raportti tutkimuksessa laaditusta yritysten tilinpäätöstietojen toimipaikkaistamismenetelmästä, joka perustuu tilastollisiin malleihin ja hyödyntää saatavilla olevaa lisäinformaatiota. Toimipaikkaistaminen tarkoittaa yritystasojen tietojen disaggregointia toimipaikkatasoisiksi tiedoiksi. Jalostusarvo valittiin toimipaikkaistettavaksi tilinpäätöstiedoksi, koska se kuuluu tärkeimpiin yritysten ja toimipaikkojen toimintaa kuvaaviin tietoihin. Tutkimuksessa laadittua toimipaikkaistamismenetelmää voidaan kuitenkin soveltaa muidenkin tilinpäätöstietojen toimipaikkaistamiseen. Menetelmän laadinnassa ja sen toimivuuden tutkimisessa olivat käytävissä Tilastokeskuksen tiedot Suomen yrityksistä ja toimipaikoista vuodelta 2000: yritys- ja toimipaikkarekisteri sekä rakennetilastojen kyselykanta. Laaditun menetelmän soveltaminen vaatiikin, että soveltajalla on nämä tiedot käytettävissään.

Kehitetty toimipaikkaistamismenetelmä on malliluovuttaja-imputointimenetelmä (model-donor imputation). Siinä puuttuva tieto korvataan tilastollisen mallin perusteella muodostetulla estimaatilla. Mallin parametrit estimoidaan käyttämällä niiden havaintoyksiköiden tietoja, joiden tulosmuuttujan ja apumuuttujien havaintoarvot ovat tiedossa. Imputoinnissa käytettävä malli riippuu siitä, mikä on käytettävien apumuuttujien luonne. Tutki-



muksessa mallina/luovuttajana käytettiin lineaarista mallia, joka oli määritelty otoksen perusteella toimialakohtaisesti tai kaikille toimialoille yhteisenä. Se mikä malli toimipaikkaistamiseen valitaan, riippuu tulosmuuttujan toimialakohtaisien yleistettyyn regressioestimointiin perustuvien kokonaismäärien estimaattien tarkkuudesta. Valintakriteereinä käytetään mm. estimaatin ja todellisen arvon erotusta, estimaatin pohjalla olevan mallin selitysasetta ja keskineliövirhettä, estimaatin variaatiokerrointa sekä keskihajontaa. Teollisuuden ja rakentamisen toimialoilla (Tilastokeskuksessa teollisuuteen ja rakentamiseen luetaan myös mineraalien kaivu ja sähkön-, kaasun- ja vesihuolto) voidaan laskea myös toimiala- ja toimipaikkaa vastaavan yrityksen laji-kohtaisia tulosmuuttujan totaalin estimaatteja ja tutkia sitä parantaako toimipaikkaa vastaavan yrityksen lajin huomioon ottaminen estimointia. Tämä tarkoittaa sitä, että teollisuuden ja rakentamisen toimialoilla lasketaan toimialakohtaisesti yhteen erilliset estimaatit yksi- ja monitoimipaikkaisten yritysten toimipaikkojen tulosmuuttujan kokonaismäärälle ja verrataan sitä vastaavaan toimialakohtaiseen estimaattiin. Samaan tapaan teollisuudessa ja rakentamisessa voidaan tutkia mm. toimipaikan sijaintiläänin vaikutusta estimointiin ja yleensä kiinnostuksen kohteena oleviin osajoukkoihin jaon kriteereiden vaikutusta estimointiin.

Tutkimuksen menetelmien kannalta on tärkeää yleistettyihin regressioestimaattoreihin, estimointiin (erityisesti osajoukkokohtaiseen), otanta-asetelmiin, survey-metodologiaan ja imputointiin liittyvä kirjallisuus. Niistä on kerrottu etenkin kirjassa Särndal ym. (1992). Se on perusteos asetelmaperusteisista estimointimenetelmistä. Lisäksi niistä kertovat kirjat Pahkinen ja Lehtonen (1989) sekä Lehtonen ja Pahkinen (2004). Tutkimuksen apuna ovat olleet myös kirjat Djerf (2001) ja artikkelit Estevao ym. (1995) sekä Särndal (2001).

Luku 5.1 kertoo lisäinformaation käytöstä otannassa ja estimoinnissa, luku 5.2 käytetystä otanta-asetelmasta ja luku 5.3 tavasta, jolla estimoidaan toimialakohtaisia tilinpäätöstietoja. Luku 5.4 keskittyy otannasta riippumattomiin virheisiin, erityisesti siihen mistä tutkimus lähti liikkeelle eli katoon (joka ilmenee siis toimipaikkatie-doissa). Käytetty imputointimenetelmä eli yritysten tilinpäätöstietojen toimipaikkaistaminen valitun lineaarisen mallin perusteella on selostettu luvussa 5.5 ja diagnostiset tarkastelut ovat luvussa 5.6. Laaditun menetelmän soveltaminen jalostusarvoon ja päätelmät menetelmän toimivuudesta ovat luvuissa 6-8.

## 5.1 Lisäinformaation käyttö otannassa ja estimoinnissa

**Otoksesta** on kyse kun perusjoukosta on valittu osajoukko todennäköisyysotantaa käyttämällä. Todennäköisyysotanta on kyseessä kun kaikki mahdolliset otokset (eli otosavaruus) voidaan luetella ja tunnetaan jokaisen otoksen toteutumisen todennäköisyys eli kaikkien otosten poimintatodennäköisyydet. Perusjoukon alkion todennäköisyyttä tulla otokseen kutsutaan sisältymistodennäköisyydeksi. Todennäköisyysotannassa sisältymistodennäköisyydet ovat tunnettuja ja nollan ja ykkösen välillä, mutta voivat poiketa toisistaan. Tämä tarkoittaa sitä, että sisältymistodennäköisyyksien ei tarvitse olla vakio, vaan perusjoukon alkiolla voi olla erisuuret sisältymistodennäköisyydet kuten ositetussa otannassa ja otannassa otosyksikön koon mukaan. (Särndal ym., 1992, 8, Lehtonen ja Pahkinen, 2004).

**Otantamenetelmiä** on useita ja monet tilastotieteen parissa käytetyt menetelmät ovat muiden otantamenetelmien muunnoksia, yhdistelmiä tai täydennyksiä. Otannan perusmenetelmiä ovat yksikertainen satunnaisotanta palauttaen tai palauttamatta (SRSWR, SRSWOR), systemaattinen otanta (SYS) ja otanta otosyksikön koon mukaan (PPS). (Pahkinen ja Lehtonen, 1989, luku 2).

**Otanta-asetelmassa** määritellään kaikkien mahdollisten otosten poimintatodennäköisyydet ja alkioiden sisällymistodennäköisyydet. Otanta-asetelmaa merkitään  $p(\cdot)$ :llä, otoksen  $s$  poimintatodennäköisyyttä  $p(s)$ :llä ja alkion  $k$  sisällymistodennäköisyyttä  $p_k$ :lla. Otanta-asetelma määrittää tärkeimmät tilastolliset ominaisuudet kuten otantajakauman, odotusarvon ja varianssin. Lisäksi se määrittää satunnaisia ominaisuuksia, joita on laskettu otoksesta. Niitä ovat mm. otoksesta lasketut tunnusluvut kuten mediaani, keskiarvo, totaali ja otosvariassi. Se kertoo sen mikä on perusjoukko, mitkä ovat otanta-asteet, mitä otantamenetelmiä käytetään eri otanta-asteilla ja miten otos käytännössä poimitaan. Otanta-asetelman muodostavat lyhyesti ne säännöt ja menetelmät, joita käyttäen otos poimitaan jostain äärellisestä perusjoukosta. (Särndal ym., 1992, 27, Lehtonen ja Pahkinen, 2004).

Otanta-asetelmasta ovat lisäksi johdettavissa tunnuslukujen laskennassa käytettävät estimaattorit ja niiden keskivirheiden estimaattorit. Perusjoukon tuntematonta arvoa kutsutaan parametriksi ja sen arviointia tai enustamista matemaattisin keinoin **estimoinniksi**. Tarkalleen ottaen estimointi tarkoittaa tapaa jolla otanta-asetelmaan liitetään jokin laskukaava tai laskenta-algoritmi. Tätä laskukaavaa tai laskenta-algoritmia kutsutaan perusjoukon tuntemattoman parametrin  $q$  **estimaattoriksi** ja sitä merkitään  $\hat{q}$ :lla. Estimaattori riippuu käytettävästä otanta-asetelmasta. Kun estimaattoriin liitetään otoksesta saadut lukuarvot, saadaan **estimaatti** perusjoukon tuntemattomalle parametrille  $q$ . Yleensä estimointi sisältää myös sen tarkkuuden arvioinnin varianssin ja keskivirheen estimaattorin avulla. Tämä on tärkeä osa koko perusjoukkoa koskevan päättelyn epävarmuuden tutkimisessa. Varianssin estimaattorikin riippuu käytetystä otanta-asetelmasta. (Pahkinen ja Lehtonen, 1989, 3, Lehtonen ja Pahkinen, 2004, Särndal ym., 1992, 27).

**Apumuuttujiksi** kutsutaan muuttujia, joiden havaintoarvoissa esiintyvä vaihtelu liittyy tulomuuttujien havaintoarvoissa esiintyvään vaihteluun ja joiden havaintoarvot otoksessa ovat tiedossa. Lisäksi apumuuttujien havaintoarvot ja/tai kokonaismäärät perusjoukossa tai perusjoukon osajoukoissa tulee olla tiedossa. Apumuuttujien havaintoarvoja ja totaaleja kutsutaan **lisäinformaatioksi**. Voimakas lisäinformaatio on tärkeätä etenkin silloin kun on kyseessä osajoukkokohtainen estimointi. (Särndal ym. 1992, 26, Lehtonen ja Pahkinen, 2004).

**Lisäinformaatiota voidaan käyttää joko otannan tai estimoinnin vaiheissa.** Jos lisäinformaatiota on käytettävissä ennen otoksen poimintaa, tulevat kyseeseen otanta otosyksikön koon mukaan (PPS), ositettu otanta (STR), systemaattinen otanta (SYS) ja ryväotanta (CLU). Tällöin on kyseessä lisäinformaation käyttö otannan vaiheessa. Se mitä otantamenetelmää käytetään, riippuu paljon siitä minkälainen perusjoukko on, mitä tutkitaan (eli mitkä ovat tulomuuttujat ja tuntemattomat parametrit) ja mitä perusjoukosta tiedetään (eli se onko käytettävissä lisäinformaatiota ja miten tehokasta se on tulomuuttujien estimoinnin kannalta). Jollei otosta poimittaessa ole käytettävissä lainkaan lisäinformaatiota, niin otosta ei voida poimia kuin yksinkertai-

sella satunnaisotannalla (SRS). Tämä antaa aiheen olettaa, että yksinkertaisella otannalla saatuja tuloksia vertaamalla muita otantamenetelmiä käyttämällä saatuihin tuloksiin saadaan arvioita sille miten lisäinformaation käyttö otannan vaiheessa on parantanut estimointia. (Lehtonen ja Pahkinen, 2004).

Systemaattisessa otannassa lisäinformaatiota voidaan käyttää perusjoukon alkioiden listauksessa siten, että alkion järjestys listassa riippuu tulosuuttujan arvosta. PPS-otannassa lisäinformaationa on muuttuja, joka mittaa otosyksikön kokoa. Tätä muuttujaa sanotaan kokomuuttujaksi ja alkioiden sisällymistodennäköisyydet määritellään PPS-otannassa suhteessa kokomuuttujaan. PPS-otanta on sitä tehokkaampi kuin yksinkertainen satunnaisotanta, mitä vahvempi riippuvuussuhde kokomuuttujan ja tulosuuttujan välillä on. Tuon riippuvuussuhteen täytyy myös olla vakio koko perusjoukossa. Ryväotannassa perusjoukko voidaan jakaa lisäinformaation suhteen luonnollisesti muodostuviin osajoukkoihin, joita kutsutaan rypäiksi. Jos rypäät ovat sisäisesti homogeenisia, on ryväotanta yksinkertaista satunnaisotantaa tehottomampi. Ositetussa otannassa lisäinformaatiota käytetään jakamalla perusjoukko apumuuttujien suhteen osajoukkoihin. Perusjoukon osajoukkoja kutsutaan ositteiksi ja käytetyt apumuuttujat ovat ns. osituskriteereitä. Ositetussa otannassa jokaisesta ositteesta poimitaan otos erikseen ja ositetun otannan keskihajonta on ositekohtaisten keskihajontojen summa. Jos suuri osa kokonaisvaihtelusta selittyy ositteiden välisellä vaihtelulla, ositettu otanta saattaa olla yksinkertaista satunnaisotantaa tehokkaampi. Ositetun otannan tehokkuus riippuu siis siitä miten sisäisesti homogeenisia ositteet ovat (ts. ovatko alkiot ositteen sisällä tulosuuttujan suhteen samanlaisia, mutta erilaisia tulosuuttujan suhteen verrattuna muiden ositteiden alkioihin). (Lehtonen ja Pahkinen, 2004).

Lisäinformaation käytöllä estimointivaiheessa tarkoitetaan sitä, että lisäinformaatiota käytetään laskettaessa estimaatteja jo poimitulla otoksella. Lisäinformaatiota käytetään regressio- ja suhde-estimoinnissa. Nämä ovat ns. yleistetyt regressioestimoinnin erikoistapauksia (GREG, Generalized Regression Estimator). Ne ovat tehokkaita, jos tulosuuttujien ja apumuuttujien välillä on voimakas korrelaatio. Ilman lisäinformaatiota ei itse asiassa voida laskea muuta kuin Horvitz-Thompson estimaatteja. Lisäinformaatiota voidaan käyttää myös kadon adjustointiin uudelleenpainotusmenetelmien tai imputoinnin avulla. Kadosta ja sen adjustoinnista on puhuttu luvuissa 5.4 ja 5.5. Jos aineistossa on katoa, niin ilman lisäinformaatiota voidaan estimoida vain, jos kato on täysin harmitonta (missing completely at random, MCAR). Kato, joka ei ole täysin harmitonta, voi aiheuttaa helposti harhaisia estimaatteja, jos ei ole käytetty lisäinformaatiota. Kato aiheuttaa lisäksi aina estimaattien keskivirheiden kasvua, koska otoskoko on pienempi kuin jos katoa ei olisi. (Lehtonen ja Pahkinen, 2004).

Lisäinformaation käytöllä pyritään pienentämään laadittavan estimaatin varianssia sekä tulosuuttujan todellisen arvon ja estimaatin erotusta. Jos tämä toteutuu, niin lisäinformaation käytöllä saadaan estimaatteja, jotka ovat tarkempia kuin ilman lisäinformaation käyttöä saadut estimaatit. Tällöin lisäinformaation käyttö on parantanut estimointia ja käytetty lisäinformaatio on ollut tehokasta. Jos tutkimuksessa on tarkoituksena estimoida yhteen tulosuuttujaan liittyviä parametreja (esim. totaalia), voidaan estimointia parantaa käyttämällä lisäinformaatiota otannan yhteydessä. Jos tutkimuksen tavoitteena on estimoida useampiin tulosuuttujiin liittyviä parametreja, tulisi lisäinformaatiota käyttää otannan lisäksi myös estimoinnin vaiheessa, koska yksi otanta-asetelma ei välttämättä ole tehokas kaikkien tulosuuttujien suhteen. Lisäinformaation käyttö esti-

moinnissa mahdollistaa myös sellaisten apumuuttujien käytön, joiden havaintoarvoja perusjoukon tasolla ei ole tiedossa vielä otosta poimittaessa. Lisäksi lisäinformaationa voidaan käyttää sellaisiakin muuttujia, joiden havaintoarvoja ei ole perusjoukon tasolla tiedossa, mutta joiden perusjoukosta lasketut totaalit (osajoukko-kohtaiset ja/tai perusjoukon tasoiset) ja otoksen havaintoarvot ovat tiedossa. (Lehtonen ja Pahkinen, 2004).

**Estimointiasetelma kertoo sen mitä estimaattoria estimoinnissa käytetään. Estimointistrategialla** tarkoitetaan otanta-asetelman ja estimointiasetelman yhdistelmää. Se kertoo sen millä otanta-asetelmalla otos on poimittu ja mitä totaalin estimaattoria käytetään. Yksi strategia tutkimuksessa on esimerkiksi STRSRS+GREG, joka tarkoittaa sitä, että poimitaan otos käyttäen ositettua otantaa ja estimaatit lasketaan yleistettyjä regressioestimaattoreita käyttämällä. Edellä mainitussa strategiassa lisäinformaatiota käytetään sekä otannan että estimoinnin yhteydessä. Tässä tutkimuksessa käytetään useampaa estimointistrategiaa, jotka on raportoitu luvussa 5.3. Estimointistrategian tehokkuutta voidaan tutkia ns. **asetelmakertoimen** (design effect) avulla. (Lehtonen ja Pahkinen, 2004).

Asetelmakerroin (design effect, DEFF) kertoo kuinka paljon lisäinformaation käyttö on tehostanut estimointia keskivirheen pienenemisen mielessä. Siinä valitun estimointistrategian mukaista varianssin estimaattoria verrataan tilanteeseen, jossa otos olisi poimittu yksinkertaisella satunnaisotannalla. Se lasketaan seuraavasti:

$$DEFF_{p(s)}(\hat{t}^*) = \frac{V_{p(s)}(\hat{t}^*)}{V_{srs}(\bar{N}\bar{y})}, \text{ missä} \quad (5.1)$$

$p(s)$  = käytetty otanta-asetelma,  $\hat{t}^*$  = käytetyn strategian mukainen asetelmaperusteinen totaalin estimaattori (kuten GREG-estimaattori),  $V_{srs}(\bar{N}\bar{y})$  = totaalin ( $\bar{N}\bar{y}$ ) asetelmavarianssi kun oletetaan, että käytössä on yksinkertainen satunnaisotanta palauttamatta ja  $V_{p(s)}(\hat{t}^*)$  = valitun estimointistrategian mukainen totaalin estimaattorin varianssi (ts. todellinen asetelmavarianssi käytettäessä valittua totaalin estimaattoria ja otosta).

Asetelmakertoimen estimaattori on *deff*, joka on muotoa:

$$deff_{p(s)}(\hat{t}^*) = \frac{V_{p(s)}(\hat{t}^*)}{V_{srs}(\bar{N}\bar{y})}, \quad (5.2)$$

missä merkinnät ovat samat kuin kaavassa 5.1 (paitsi, että asetelmavarianssit on korvattu estimaattoreillaan). (Pahkinen ja Lehtonen, 1989, Lehtonen ja Pahkinen, 2004).

Valittu estimointistrategia on tehokkaampi kuin strategia, johon sitä verrataan, jos asetelmakerroin on alle yksi. Tutkimuksessa voidaan tutkia eri estimaattoreiden tehokkuutta vertaamalla eri estimointistrategioiden tuottamia variansseja valitulla otanta-asetelmalla poimittuun otanta-asetelmaan, jossa estimaattorina on Horvitz-Thompson estimaattori. Jos tutkimuksessa totaalin synteettisten estimaattoreiden ja yleistettyjen regressioestimaattoreiden *deff* on alle yksi, voidaan sanoa niiden olevan HT-estimaattoreita parempia ja käytetyn lisäinformaation parantaneen estimointia. (Pahkinen ja Lehtonen, 1989, 19-20, Lehtonen ja Pahkinen, 2004).

Parhaat estimaatit saadaan, jos valitaan paras tilanteeseen sopiva otanta-asetelman ja estimaattorin yhdistelmä, jossa käytetään olemassa olevaa lisäinformaatiota sekä otannan että estimoinnin vaiheissa mikäli

mahdollista. Tämä tarkoittaa parhaan mahdollisen estimointistrategian valitsemista. Tiettyä apumuuttujaa voi käyttää sekä otannan että estimoinnin vaiheessa lisäinformaationa. Ennen otoksen poimimista ja estimointia olisikin selvitettävä se, onko löydettävissä sellaisia muuttujia, joita voitaisiin käyttää apumuuttujina otannassa, estimoinnissa tai molemmissa. Näitä muuttujia kutsutaan potentiaalisiksi apumuuttujiksi. Jos potentiaalisia apumuuttujia on paljon, tulee niistä valita sellaiset, jotka korreloivat kaikkien vahvimmin tulosmuuttujien kanssa. Näiden valittujen apumuuttujien havaintoarvot otoksessa tai perusjoukossa sekä tunnetut perusjoukon tai perusjoukon osajoukon totaalit apumuuttujille muodostavat tutkimuksessa käytettävän lisäinformaation. (Lehtonen ja Pahkinen, 2004).

Se kuinka paljon lisäinformaation käyttö parantaa estimointia keskivirheen pienentymisen mielessä, riippuu siitä kuinka suuri riippuvuus lisäinformaation ja tulosmuuttujan välillä on. Mitä suurempi riippuvuus tulosmuuttujan ja lisäinformaation välillä on, sitä parempi estimaatti saadaan. Nykyään käytetään usein lisäinformaatiota erityisesti estimoinnin vaiheessa. Se onko estimoinnissa käytettävissä lisäinformaatiota, vaikuttaa myös siihen mikä otanta-asetelma antaa parhaat tulokset (saattaa esim. käydä niin, että yksinkertainen satunnaisotanta on parempi kuin PPS-otanta, kun on käytössä tehokasta lisäinformaatiota estimointivaiheessa). (Lehtonen ja Pahkinen, 2004).

Lisäinformaatiota on löydettävissä mm. erilaisista rekistereistä (niitä voidaan kutsua myös rekisterimuuttujiksi), joita Suomessa on monia (mm. Väestörekisterikeskuksen ylläpitämä väestörekisteri ja Tilastokeskuksen yritysten suhdanteet -yksikön ylläpitämä yritys- ja toimipaikkarekisteri). Tutkimusaineisto koostuu neljästä Tilastokeskuksen ylläpitämästä rekisteristä, ja siksi kyseessä on rekisteritutkimus. Näitä rekisterejä ovat yritysrekisteri, toimipaikkarekisteri, TKANTA ja YKANTA. Siksi tutkimuksessa on käytettävissä paljon yritys- ja toimipaikkatietoja. Tutkimuksessa on mahdollista käyttää monia apumuuttujia ja niitä voidaan käyttää lisäinformaationa sekä otannan että estimoinnin vaiheessa.

Estimaattien laatimisessa käytetään apumuuttujina  $x$  (eli lisäinformaationa) toimipaikan liikevaihtoa, toimipaikan keskimääräistä henkilöstömäärää tilikauden aikana ja toimipaikalle estimoituja palkkoja. Kiinnostuksen kohteena oleviin osajoukkoihin jaon kriteereinä ovat toimipaikan sijaintilääni (vain teollisuuden toimialalla), toimipaikkaa vastaavan yrityksen laji ja toimipaikan toimialan pääluokka TOL 95:n mukaan. Osituskriteereinä ovat toimipaikkaa vastaavan yrityksen laji, toimipaikan toimialan pääluokka TOL 95:n mukaan ja toimipaikan PK-suuruusluokka. Näiden muuttujien kuvaukset ovat luvussa 3.2.

Luvussa 4.2 todettiin, että apumuuttujien (toimipaikan liikevaihto ja keskimääräinen henkilöstömäärä tilikauden aikana sekä toimipaikalle estimoidut palkat) ja tulosmuuttujien väliset Pearsonin korrelaatiokertoimet ovat voimakkaita käytettävissä aineistossa (Suomen liiketoimipaikat vuonna 2000). Tämä tarkoittaa sitä, että käytettävä lisäinformaatio voi parantaa toimialakohtaisten tilinpäätöstietojen estimointia pienentämällä estimaattien keskivirheitä ja luottamusvälejä. Kuitenkin myös tutkimuksessa käytettävien apumuuttujien väliset korrelaatiokertoimet ovat aineistossa voimakkaita ja tämä aiheuttaa multikollineaarisuuden. Multikollineaarisuus aiheuttaa rajoitteita toimipaikkaistamisen taustalla olevien lineaaristen mallien valinnassa, koska mallit on valittava yhden apumuuttujan mallien joukosta.

## 5.2 Otanta -asetelma

Tutkimuksessa pyritään estimoimaan tulosuuttujan  $y$  (eli toimipaikan jalostusarvon) kokonaismääriä (totaaleja) perusjoukon osajoukoissa. Kyseessä on siis ns. osajoukkokohtaisen estimoinnin tilanne (domain estimation/estimation for domains). Estimoinnista on puhuttu myös luvussa 5.1. Perusjoukon osajoukkoja, joille parametrien piste- ja väliestimaatteja (ts. luottamusvälejä) lasketaan osajoukkokohtaisessa estimoinnissa, kutsutaan **domaineiksi tai kiinnostuksen kohteena oleviksi osajoukoiksi**. Mielivaltaista domainia  $d$  eli perusjoukon  $U = \{1, 2, \dots, k, \dots, N\}$  osajoukkoa merkitään  $U_d$ :llä, missä  $U_d \subseteq U$ . Tämä tarkoittaa myös sitä, että

$$U = \bigcup_{d=1}^D U_d \text{ ja } N = \sum_{d=1}^D N_d. \text{ Osajoukkokohtaisia totaaleja merkitään } t_d = \sum_{U_d} y_k, \text{ missä } d = 1, \dots, D. \text{ (Särndal, 2001, 13-14, Särndal ym., 1992, 386-390, Lehtonen ja Pahkinen, 2004).}$$

2001, 13-14, Särndal ym., 1992, 386-390, Lehtonen ja Pahkinen, 2004).

Osajoukkoihin jaon kriteerinä tutkimuksessa on toimialaluokituksen 1995 pääluokka, joka on myös yksi osituskriteereistä. Tutkimuksessa estimoidaan siis toimialakohtaisia tilinpäätöstietoja. Osajoukkoihin jaon kriteerinä voidaan käyttää toimipaikan toimialan lisäksi toimipaikkaa vastaavan yrityksen lajia (sitä onko yritys yksi- vai monitoimipaikkainen). Tällä jaottelulla osajoukkoja on 29 (koska toimialoilla L, P ja Q ei vuonna 2000 toiminut yhtään toimipaikkaa ja minkään monitoimipaikkaisen yrityksen toimipaikan toimiala ei ollut tuntematon). On kuitenkin huomattava, että kaikkien osajoukkojen jalostusarvon totaaleja ei ole tiedossa, koska monitoimipaikkaisten liikeyritysten toimipaikoista jalostusarvo on tiedossa vain teollisuuden ja rakentamisen toimialojen toimipaikoille. Tiedossa ovat siksi 19:n osajoukon jalostusarvon totaalit ja kymmenen osajoukon totaalit eivät ole tiedossa (eli täytyy imputoida niiden toimipaikkojen jalostusarvot). Toimipaikkarekisterin tiedot ovat kuitenkin olemassa kaikilta osajoukoilta. Koska kiinnostuksen kohteena oleva osajoukko on toimiala, voidaan tutkimuksessa myös puhua toimialakohtaisesta estimoinnista.

Tutkimuksessa laadittavat estimaatit toimipaikkojen tilinpäätöstiedoille pohjautuvat otoksilla saatuihin tuloksiin, koska otoksien avulla voidaan mallien toimintakyky tutkia siinä joukossa, josta on tietoa. Saadaan siis tietoa siitä miten hyvin laaditut estimaatit ja niiden pohjalla olevat mallit toimivat yksitoimipaikkaisissa liikeyrityksissä ja monitoimipaikkaisten liikeyritysten teollisuuden ja rakentamisen toimipaikoilla. Lisäksi saadaan tietoa siitä onko estimaattien "hyvydessä" eroja yksi- ja monitoimipaikkaisten liikeyritysten toimipaikkojen välillä teollisuuden ja rakentamisen toimipaikkojen kohdalla. Otos, otantamenetelmä ja otanta-asetelma on määritelty luvussa 5.1.

Estimoinnissa voidaan käyttää kahta erilaista perusjoukon osajoukkoa kiinnostuksen kohteena olevien osajoukkojen lisäksi. Niitä ovat informaatioryhmät ja ositteet. Ne on erotettava toisistaan, koska ne tarkoittavat eri asiaa, vaikkakin voivat olla jossain tapauksessa samojakin (esimerkiksi ositteet saattavat olla kiinnostuksen kohteena olevia osajoukkoja). (Särndal, 2001, 29, Lehtonen ja Pahkinen, 2004).

**Informaatioryhmiksi** kutsutaan perusjoukon osajoukkoja, joille lisäinformaatio on määritelty. Tutkimuksessa perusjoukko (YTR:ssä ja rakennetilastojen kyselykannassa olevat toimipaikat) voidaan jakaa informaatioryh-

miin toimipaikan sijaintiläänin mukaan ts. informaatioryhmiin jaon kriteerinä on toimipaikan sijaintilääni. Kuitenkin tämä tehdään vain teollisuuden toimialan toimipaikkojen kohdalla. Tutkimuksessa voidaan ajatella myös, että kiinnostuksen kohteena olevat osajoukotkin ovat informaatioryhmiä. Tämä johtuu siitä, että lisäinformaatio on tiedossa myös osajoukkokohtaisesti (esim. toimialakohtaisesti), koska YTR sisältää käytettävien apumuuttujien arvot kaikille toimipaikoille. Jatkossa merkitään Särndalin (2001) tavoin mielivaltaista informaatioryhmää  $i$ :llä, missä  $i = 1, \dots, I$ . (Särndal, 2001, 28-30).

Estimoinnissa käytettävä otantamenetelmä on **ositettu yksinkertainen satunnaisotanta palauttamatta eli STRSRS**. Ositetun otannan menetelmän käyttämisen työvaiheet ovat: 1. osituskriteerin valinta, 2. perusjoukon jakaminen ositteisiin, 3. kiintiöintimenetelmän valinta ja 4. otoksen poiminta jollain otantamenetelmällä. Ositetussa otannassa äärellinen perusjoukko, jota merkitään  $U = \{1, \dots, k, \dots, N\}$ , jaetaan sisäisesti homogeenisiin toisensa poissulkeviin (kukin havaintoyksikkö voi kuulua vain yhteen ositteeseen) osajoukkoihin halutun osituskriteerin mukaan. Näitä sanotaan **ositteiksi** ja niitä merkitään  $\{U_1, \dots, U_H\}$ . Tämän jälkeen poimitaan jollakin perusotantamenetelmällä (systemaattinen otanta eli SYS, yksinkertainen satunnaisotanta eli SRS tai otanta otosyksikön koon mukaan eli PPS) otos alkioista (kun ei ole kyseessä ryväotanta) tai rypäistä (kun kyseessä on ryväotanta) otos jokaisesta ositteesta erikseen. Nämä otokset poimitaan toisistaan riippumatta ja eri ositteista voidaan otokset poimia eri menetelmillä. (Särndal ym., 1992, 101-104, Lehtonen ja Pahkinen, 2004).

Se kuinka monta havaintoyksikköä mistäkin ositteesta poimitaan otokseen määrätään käytettävällä **kiintiöintimenetelmällä**. Tutkimuksessa käytetään suhteellista kiintiöintiä. Suhteellinen kiintiöinti tarkoittaa sitä, että jokaisesta ositteesta otetaan havaintoyksiköitä otokseen samassa suhteessa kuin niitä on perusjoukossa ts.

alkioiden  $k$  sisällysmistodennäköisyydet ovat samat jokaiselle alkioille  $k$  eli  $p_k = \frac{n}{N}$ . Tällöin ositteen  $h$  otos-

koko  $n_h = \frac{N_h}{N} * n$ . Muita kiintiöintimenetelmiä ovat mm. Bankier- ja Neyman kiintiöinti sekä optimaalinen kiintiöinti. (Lehtonen ja Pahkinen, 2004).

Suhteellisessa kiintiöinnissä otetaan siis huomioon ositteen koko ja suuremmasta ositteesta otetaan otokseen enemmän havaintoyksiköitä kuin pienemmästä ositteesta. Se on hyvä menetelmä, jos voidaan ajatella perusjoukon varianssin olevan suurempi isoissa ositteissa kuin pienissä. Kaikissa kiintiöintimenetelmissä

ositteiden otoskoot  $n_h$  valitaan siten, että  $n = \sum_{h=1}^H n_h$ . (Särndal ym., 1992, 106, Lehtonen ja Pahkinen, 2004).

Tutkimuksessa jaetaan osituskriteerien (toimipaikan toimiala, PK-suuruusluokka ja toimipaikkaa vastaavan yrityksen laji) mukaan ne tutkimusaineistossa olevat toimipaikat, jolle toimipaikan laskennallisen käyttökatteen tiedot ovat tiedossa (n. 88 % toimipaikoista) ositteisiin. Jokaisesta ositteesta poimitaan erikseen otos yksinkertaisella satunnaisotannalla. Ositus on tapa kohdistaa otosta halutulla tavalla ja kriteereillä perusjoukkoon. Ositetun otannan tehokkuus riippuu siitä miten perusjoukko jaetaan osajoukkoihin ts. osituskriteeristä.

Jos osituskriteeriä käyttämällä saadaan jaettua perusjoukko sisäisesti homogeenisiin osajoukkoihin, on ositettu otanta tehokas menetelmä. (Särndal ym., 1992, 101-104, Lehtonen ja Pahkinen, 2004).

Osittamisella on monia tavoitteita. Yksi tärkeimmistä on estimaatin tehokkuuden parantaminen keskivirheen pienenemisen mielessä. Keskivirhe pienenee osittamisen seurauksena, jos ositekohtaisten varianssien summa on pienempi kuin koko joukosta laskettu varianssi. Tällöin ositteet ovat sisäisesti homogeenisia ja ositekohtaiset varianssit ovat siksi pieniä. Kiintiöntimenetelmä ja menetelmät, joilla otokset ositteista poimitaan vaikuttavat ositetun otannan tehokkuuteen, koska ositekohtaiset varianssit riippuvat ositteen otoskoosta. Osittamisen käytön muita tärkeitä syitä ovat: erilaisten otantamenetelmien soveltamisen mahdollisuus eri osajoukoissa ts. ositteissa, joustava lisäinformaation käyttö ositekohtaisesti sekä otannassa että estimoinnissa ja se, että osittamisella voidaan perusjoukon osajoukoille (ositteille) varmistaa riittävän suuri otoskoko. (Pahkinen ja Lehtonen, 1989, 60, Lehtonen ja Pahkinen, 2004).

Mikäli mahdollista tulisi tärkeimpien ositteiden olla myös kiinnostuksen kohteena olevia osajoukkoja, sillä siten voidaan asettaa osajoukon koko otoksessa eli  $n_d$  kiinteäksi otanta-asetelmassa ja sellaiseksi, että estimaatille saadaan haluttu täsmällisyys. Tällöin puhutaan ns. **suunnitelluista osajoukoista** (planned domains). Muuten  $n_d$ :t ovat satunnaislukuja eli niitä ei ole asetettu otanta-asetelmassa ja estimaattoreiden varianssit kasvavat. Tällöin on kyseessä ns. **suunnittelemattomien osajoukkojen** tilanne (unplanned domains). Jotta kiinnostuksen kohteena olevat osajoukot voidaan asettaa ositteiksi, tulee tietää jo otanta-asetelmaa suunniteltaessa mille perusjoukon osajoukoille estimointia halutaan tehdä ja mitkä niistä ovat tärkeitä. Käytännön syistä (osajoukkokohtaisen estimoinnin tarvetta ei ole tiedetty otanta-asetelmaa laadittaessa tai kiinnostuksen kohteena olevien osajoukkojen  $D$  määrä on niin suuri, että niiden asettaminen ositteiksi olisi kallista) osittamista kiinnostuksen kohteena olevien osajoukkojen mukaan ei aina voida tehdä. (Särndal, 2001, 13, Särndal ym., 1992, 388, Lehtonen ja Pahkinen, 2004).

Jos käytetään suunnittelemattomia osajoukkoja, voi käydä niin, että joistakin osajoukoista tulee otokseen hyvin vähän alkia, etenkin jos otoskoko kokonaisuudessaan ei ole riittävän suuri. Tämä aiheuttaa epätarkkaa estimointia estimaattoreiden varianssien kasvun lisäksi. Suunnittelemattomien osajoukkojen tapauksessa käytettäessä yksinkertaista satunnaisotantaa palauttamatta osajoukkokohtaisten otoskokojen  $n_d$  odotusarvot  $E(n_d) = n * (N_d / N)$ . Tämä pätee myös ositetun otannan tapauksessa suhteellista kiintiöntiä käytettäessä, siis tässäkin tutkimuksessa. (Lehtonen ja Pahkinen, 2004).

Päädettiin siihen, että otosyksiköt poimitaan ositetulla otannalla, koska Suomen liikeyrityksistä ja niiden toimipaikoista muodostuva tutkimuksen kohdeperusjoukko on hyvin heterogeeninen, mutta se voidaan jakaa sisäisesti homogeenisiin osajoukkoihin. Tämä tarkoittaa sitä, että tutkittavissa ominaisuuksissa on paljon vaihtelua perusjoukossa, mutta kun perusjoukko jaetaan osajoukkoihin, ei näiden osajoukkojen sisällä ole niin suurta vaihtelua tutkittavissa ominaisuuksissa kuin perusjoukossa. Tällöin voidaan myös otoskoot kiinnostuksen kohteena olevissa osajoukoissa (tutkimuksessa toimialan ja toimipaikkaa vastaavan lajin mukaisissa



osajoukoissa) saada kiinteiksi, kun ne muuten olisivat satunnaislukuja. (Pahkinen ja Lehtonen, 1989, luku 2.5, Lehtonen ja Pahkinen, 2004).

Yritykset ja toimipaikat ovat hyvin eri kokoisia, sijaitsevat eri alueilla ja toimivat eri toimialoilla. Toimipaikkatason tilinpäätöstiedot (tutkittavat ominaisuudet) taas riippuvat aiemmin tehtyjen tutkimusten ja luvussa 4 esitetyn tarkastelun mukaan myös käytettävässä aineistossa paljolti toimipaikan toimialasta, koosta ja toimipaikkaa vastaavan yrityksen lajista (ts. siitä onko yritys yksi- vai monitoimipaikkainen). Siksi valittiin osituskriteereiksi toimipaikan toimiala, toimipaikkaa vastaavan yrityksen laji ja toimipaikan suuruusluokka (jako PK-yrityksiin ja suuriin yrityksiin). (Kinnunen ym., 2000, 103-109, Maliranta, 1996 ja Virkkala, 1987).

Tutkimuksessa yritysten jalostusarvojen toimipaikkaistamista rajoittaa se, että muille kuin monitoimipaikkaisten liikeyritysten teollisuuden ja rakentamisen toimialojen toimipaikoille ja kaikkien toimialojen yksitoimipaikkaisille liikeyrityksille laaditut estimaatit ovat sellaisia, joiden "hyvyyttä" ei voi tarkistaa. Se, jos tutkimuksessa rakennettavat mallit toimivat monitoimipaikkaisten liikeyritysten teollisuuden ja rakentamisen toimialojen toimipaikkojen ja kaikkien toimialojen yksitoimipaikkaisten liikeyritysten kohdalla, antaa uskottavuutta niiden toimintaan koko perusjoukossa ts. kaikilla toimipaikoilla. Jollei jokin malli toimi siinä perusjoukon osajoukossa, josta on tietoa, on oletettavaa ettei kyseinen malli toimi koko perusjoukossakaan. Jos taas jokin malli toimii siinä perusjoukon osajoukossa, josta on tietoa, aiheuttaa se epäuskoa siihen, että kyseinen malli toimisi koko perusjoukossa.

Jotta voitaisiin laatia mahdollisimman hyvät estimaatit toimialakohtaisille tilinpäätöstiedoille teollisuuden ja rakentamisen toimialoilla, tulee yksi otos poimia teollisuuden ja rakentamisen toimialojen toimipaikoista ts. perusjoukon osajoukosta. Koska tavoitteena on myös tutkia sitä miten yksitoimipaikkaisten yritysten toimipaikkojen tilinpäätöstietojen ottaminen mukaan estimointiin vaikuttaa estimaatteihin, tulee ottaa kaksi erilaista otosta (ensimmäinen teollisuuden ja rakentamisen monitoimipaikkaisten yritysten toimipaikoista ja toinen teollisuuden ja rakentamisen yksitoimipaikkaisista yrityksistä). Lisäksi tulee tutkia sitä miten toimipaikan toimiala vaikuttaa estimaatteihin, joten tulee ottaa otos sellaisista yksitoimipaikkaisista yrityksistä, joiden toimiala ei ole teollisuus tai rakentaminen.

Tutkimuksessa toimialat poikkeavat toisistaan paljonkin tutkittavien ominaisuuksien suhteen, koska toimipaikan tilinpäätöstiedot riippuvat mm. siitä millä toimialalla toimipaikka toimii. Lisäksi yksitoimipaikkaisten yritysten toimipaikoille ja monitoimipaikkaisten yritysten toimialojen C - F toimipaikoille on tiedossa niiden toimipaikkojen tilinpäätöstietojen todelliset arvot, joita on tarkoitus ennustaa. Siksi voidaan käyttää toimialoilla C-F toimivien monitoimipaikkaisten yritysten toimipaikkojen ja kaikkien toimialojen yksitoimipaikkaisten yritysten tietoja estimoinnissa. Tämä tapahtuu käyttämällä epäsuoraa estimointia. Estimointi tehdään siten, että ensin laaditaan estimaatit tilinpäätöstietojen toimialakohtaisille kokonaismäärille eli totaaleille (ts. totaalien estimaatit osajoukoissa otoksen perusteella) ja näitä käytetään hyväksi laadittaessa toimipaikkakohtaisten tilinpäätöstietojen estimaatteja koko aineistolle.

Yksinkertainen satunnaisotanta ei ole tutkimuksessa paras mahdollinen menetelmä otoksen poimintaan, koska se ei käytä otoksen poiminnassa mitään lisäinformaatiota. Ja mitä enemmän lisäinformaatiota otoksen poiminnassa käytetään, sitä parempia otoksen perusteella lasketut Horvitz-Thompson estimaatit ovat (ts. mitä parempaa lisäinformaatiota otoksen poiminnan vaiheessa on käytettävissä, sitä pienempiä ovat HT-estimaattien keskivirheet ja varianssit). Yksinkertaista satunnaisotantaa ei suositella käytettäväksi muutoin kuin, jos kohdeperusjoukko on homogeeninen (tutkittavissa ominaisuuksissa ei ole perusjoukossa paljon vaihtelua) tai jollei kohdeperusjoukosta ole tietoa ennen otoksen poimintaa (ei ole käytettävissä lainkaan lisäinformaatiota). Myös systemaattista otantaa voidaan käyttää kun perusjoukko on homogeeninen tai sen ominaisuuksia ei tunneta, jollei perusjoukossa esiinny jaksoittaista vaihtelua. Koska tutkimuksessa on käytettävissä lisäinformaatiota, ei myöskään systemaattinen otanta ole välttämättä paras mahdollinen menetelmä tässä tutkimuksessa otoksen poimintaan. (Heikkilä, 1998, 35).

Otanta otosyksikön koon mukaan (PPS) ei myöskään ole välttämättä paras mahdollinen menetelmä, jolla otostoimipaikat tutkimuksessa poimittaisiin. Se johtuu siitä, että toimipaikan koko ei yksin vaikuta tutkittavissa ominaisuuksissa ilmenevään vaihteluun (luku 4). Toimipaikan koon lisäksi tilinpäätöstietoihin vaikuttavat toimipaikan toimiala ja toimipaikkaa vastaavan yrityksen laji (katso luku 4.1). Lisäksi mahdollisten kokomuuttujien (henkilöstön määrä, liikevaihto ja palkat) ja jalostusarvon väliset riippuvuussuhteet eivät välttämättä ole vakioita (PPS-otannan käyttämisen edellytys, luku 4.1), koska kaikkien tutkimuksessa käytettyjen muuttujien väliset riippuvuussuhteet vaihtelevat toimialoittain. Kuitenkin otosyksikön koko (tutkimuksessa toimipaikan koko) on otettava huomioon otosyksiköitä poimittaessa ja osittamiskriteeriksi on niistä valittu toimipaikan PK-suuruusluokka (ts. toimipaikkojen jako PK-yrityksiin ja suuriin yrityksiin.).

Otoskoon valintaan vaikuttavia tekijöitä ovat perusjoukon heterogeenisuus, tulosten yksityiskohtaisuus, valittu luottamustaso, valittu virhemarginaali, tutkittavien tapausten suhteellinen osuus, odotettavissa oleva kato ja perusjoukon koko. Perusjoukon koko ei vaikuta muuta kuin siihen tehdäänkö kokonais- vai otantatutkimus. Ei siis ole niin, että mitä suurempi perusjoukko on, sitä suurempi tulee myös otoskoon olla. Otoskoon määrittävätkin erilaiset perusjoukon muut ominaisuudet (etenkin perusjoukon heterogeenisuus). Mitä enemmän vaihtelua tutkittavissa ominaisuuksissa on, sitä suurempi tulee olla otoskoon. Otoskokoon vaikuttavat myös valittu tulosten tarkkuustaso ja tutkimuksen resurssit (mm. raha ja aika). (Heikkilä, 1998, 40-43).

***Jotta voidaan tehdä tutkimuksessa tarvittavia vertailuja, poimitaan aineistosta kuusi eri otosta ositeulla otannalla käyttäen suhteellista kiintiöintiä. Osituskriteerit, tulosuuttuja, tunnistemuuttujat ja lisäinformaatiomuuttujat ovat jokaisessa otoksessa samat.*** Tunnistemuuttujat: Yritystunnus ja toimipaikatunnus. Osituskriteerit: toimipaikan toimialan pääluokka, toimipaikkaa vastaavan yrityksen laji ja toimipaikan suuruusluokka (jako PK- ja suuriin toimipaikkoihin YTR:n perusteella). Kiinnostuksen kohteena oleviin osajoukkoihin jaon perusteet: Toimipaikan toimialan pääluokka sekä toimipaikkaa vastaavan yrityksen laji. Tulosuuttuja: jalostusarvo. Apumuuttujat eli ns. lisäinformaatiomuuttujat, joita käytetään estimoinnin yhteydessä: Toimipaikan liikevaihto, keskimääräinen henkilöstön määrä tilikauden aikana ja toimipaikalle estimoidut palkat.

### 1. OTANTA-ASETELMA. OTOKSEN POIMINTA TEOLLISUUDEN JA RAKENTAMISEN MONITOIMIPAIKKAISTEN YRITYSTEN TOIMIPAIKOISTA:

Perusjoukko: Toimipaikkarekisterissä ja rakennetilastojen kyselykannassa TKANNASSA olevat monitoimipaikkaisten yritysten toimipaikat, joiden toimiala on rakennetilastojen mukaan teollisuus tai rakentaminen (toimialat C-F, liite 2). Otosyksikkö: toimipaikka, otantamenetelmä: STRSRS eli ositettu otanta käyttäen yksinkertaista satunnaisotantaa palauttamatta. Kiintiöintimenetelmä: suhteellinen kiintiöinti,  $N = 2786$ ,  $n = 1395$ , ositteiden lukumäärä  $H = 8$ . Yksittäisen toimipaikan  $k$  sisällymistodennäköisyys  $p_k = 0.5$ . Ositteet:

| Osite   | $h$ | $N_h$ | $n_h$ | $E(n_h)$ |
|---|-----|-------|-------|----------|
| PK mineraalien kaivuun toimipaikka                | 1   | 56    | 28    | 28.0     |
| suuri mineraalien kaivuun toimipaikka             | 2   | 43    | 22    | 21.5     |
| PK teollisuuden toimipaikka                       | 3   | 972   | 486   | 486,7    |
| suuri teollisuuden toimipaikka                    | 4   | 1068  | 534   | 534.8    |
| PK sähkön-, kaasun- ja vesihuollon toimipaikka    | 5   | 25    | 13    | 12.5     |
| suuri sähkön-, kaasun- ja vesihuollon toimipaikka | 6   | 173   | 87    | 86.6     |
| PK rakentamisen toimipaikka                       | 7   | 275   | 138   | 137.7    |
| suuri rakentamisen toimipaikka                    | 8   | 174   | 87    | 87.1     |

### 2. OTANTA-ASETELMA. OTOKSEN POIMINTA YKSITOIMIPAIKKAISISTA YRITYKSISTÄ, JOIDEN TOIMIALA ON TEOLLISUUS TAI RAKENTAMINEN

Perusjoukko: yritysrekisterissä ja rakennetilastojen kyselykannassa olevat teollisuuden ja rakentamisen toimialojen yksitoimipaikkaiset yritykset (liite 2). Otosyksikkö: yksitoimipaikkainen yritys, otantamenetelmä: STRSRS, kiintiöintimenetelmä: suhteellinen kiintiöinti,  $N = 54\ 305$ ,  $n = 21725$ , ositteiden lukumäärä  $H = 8$ .

Yksittäisen liikeyrityksen  $k$  sisällymistodennäköisyys  $p_k = 0.4$ . Ositteet:

| Osite   | $h$ | $N_h$ | $n_h$ | $E(n_h)$ |
|---|-----|-------|-------|----------|
| PK mineraalien kaivuun toimipaikka                | 1   | 1111  | 445   | 444.5    |
| suuri mineraalien kaivuun toimipaikka             | 2   | 5.0   | 2     | 2.0      |
| PK teollisuuden toimipaikka                       | 3   | 23733 | 9494  | 9494.5   |
| suuri teollisuuden toimipaikka                    | 4   | 508   | 204   | 203.2    |
| PK sähkön-, kaasun- ja vesihuollon toimipaikka    | 5   | 530   | 212   | 212.0    |
| suuri sähkön-, kaasun- ja vesihuollon toimipaikka | 6   | 209   | 84    | 83.6     |
| PK rakentamisen toimipaikka                       | 7   | 28130 | 11252 | 11253.6  |
| suuri rakentamisen toimipaikka                    | 8   | 79    | 32    | 31.6     |

### 3. OTANTA-ASETELMA. OTOKSEN POIMINTA MUIDEN KUIN TEOLLISUUDEN JA RAKENTAMISEN TOIMIALOJEN YKSITOIMIPAIKKAISISTA YRITYKSISTÄ

Perusjoukko: Yritysrekisterissä ja rakennetilastojen kyselykannassa olevat muut kuin teollisuuden ja rakentamisen toimialojen yksitoimipaikkaiset liikeyritykset. Otosyksikkö: yksitoimipaikkainen liikeyritys, jonka toimiala ei rakennetilastojen kyselykannan mukaan ole teollisuus tai rakentaminen. Otantamenetelmä: STRSRS, kiintiöintimenetelmä: suhteellinen kiintiöinti,  $N = 156352$ ,  $n = 62550$ , ositteiden lukumäärä  $H = 26$ . Yksittäisen liikeyrityksen sisällymistodennäköisyys  $p_k = 0.4$

Ositekriteerin pohjalta muodostetut ositteet:

| Osit   | $h$ | $N_h$ | $n_h$ | $E(n_h)$ |
|--|-----|-------|-------|----------|
| PK teollisuuden toimipaikka  | 1   | 139   | 56    | 55.6     |
| suuri teollisuuden toimipaikka                                     | 2   | 37    | 15    | 14.8     |
| PK sähkön-, kaasun- ja vesihuollon toimipaikka                     | 3   | 5     | 2     | 2.0      |
| suuri sähkön-, kaasun- ja vesihuollon toimipaikka                  | 4   | 8     | 4     | 3.2      |
| PK rakentamisen toimipaikka  | 5   | 22    | 9     | 8.8      |
| suuri rakentamisen toimipaikka                                     | 6   | 4     | 2     | 1.6      |
| PK maatila   | 7   | 6213  | 2486  | 2485.6   |
| suuri maatila  | 8   | 9     | 4     | 3.6      |
| PK kalatalouden toimipaikka  | 9   | 626   | 251   | 250.4    |
| PK tukku- ja vähittäiskauppa                                       | 10  | 44820 | 17928 | 17930.6  |
| suuri tukku- ja vähittäiskauppa                                    | 11  | 251   | 101   | 100.4    |
| PK majoitus- ja ravitsemistoiminnan toimipaikka                    | 12  | 9364  | 3746  | 3746.2   |
| suuri majoitus- ja ravitsemistoiminnan toimipaikka                 | 13  | 52    | 21    | 20.8     |
| PK kuljetus-, varastointi- ja tietoliikenteen toimipaikka          | 14  | 22602 | 9041  | 9042.1   |
| suuri kuljetus-, varastointi- ja tietoliikenteen toimipaikka       | 15  | 192   | 77    | 76.8     |
| PK rahoitustoiminnan toimipaikka                                   | 16  | 2255  | 902   | 902.1    |
| suuri rahoitustoiminnan toimipaikka                                | 17  | 110   | 44    | 44.0     |
| PK kiinteistö-, vuokraus- ja tutkimuspalvelujen toimipaikka        | 18  | 38449 | 15380 | 15381.9  |
| suuri kiinteistö-, vuokraus- ja tutkimuspalvelujen toimipaikka     | 19  | 1112  | 445   | 444.9    |
| PK koulutuksen toimipaikka   | 20  | 1456  | 583   | 582.5    |
| suuri koulutuksen toimipaikka                                      | 21  | 12    | 5     | 4.8      |
| PK terveydenhuolto- ja sosiaalipalvelujen toimipaikka              | 22  | 11986 | 4795  | 4795.1   |
| suuri terveydenhuolto- ja sosiaalipalvelujen toimipaikka           | 23  | 22    | 9     | 8.8      |
| PK muiden yhteiskunnallisten ja henkilökohtaisten palvelujen tp    | 24  | 16482 | 6593  | 6593.8   |
| suuri muiden yhteiskunnallisten ja henkilökohtaisten palvelujen tp | 25  | 113   | 46    | 45.2     |
| PK toimipaikka, jonka toimiala on tuntematon                       | 26  | 11    | 5     | 4.4      |

Poimituista kolmesta otoksesta voidaan muodostaa yhdistämällä uusia otoksia. Ensimmäinen niistä on otos, jota voidaan pitää otoksena teollisuuden ja rakentamisen toimipaikoista. Se saadaan käytännössä yhdistämällä otokset 1 ja 2. Otanta-asetelma on tällöin seuraavassa esitetty (4. otanta-asetelma).

#### 4. OTANTA-ASETELMA. OTOKSEN POIMINTA TEOLLISUUDEN JA RAKENTAMISEN TOIMIPAIKOISTA

Perusjoukko: Toimipaikkarekisterissä ja rakennetilastojen kyselykannassa eli TKANNASSA olevat toimipaikat, joiden toimiala on rakennetilastojen mukaan teollisuus tai rakentaminen (toimialat C-F, liite 2). Otosyksikkö: teollisuuden tai rakentamisen toimialan toimipaikka, otantamenetelmä: STRSRS eli ositettu otanta käyttäen yksinkertaista satunnaisotantaa palauttamatta, kiintiöintimenetelmä: suhteellinen kiintiöinti,  $N = 57091$ ,  $n = 23120$  ja ositteiden lukumäärä  $H = 16$ . Yksittäisen toimipaikan  $k$  sisällymistodennäköisyys  $p_k =$  Monitoimipaikkaisten yritysten toimipaikoille eli ositteiden 1-8 toimipaikoille  $p_k = 0.5$ . Yksitoimipaikkaisten yritysten toimipaikoille eli ositteiden 9-16 toimipaikoille  $p_k = 0.4$ .

Toinen yhdistetty otos muodostetaan yhdistämällä kaikki kolme poimittua otosta. Tämä otanta-asetelma on esitetty seuraavassa (5. otanta-asetelma).

## 5. OTANTA-ASETELMA. OTOKSEN POIMINTA YKSITOIMIPAIKKAISISTA LIIKEYRITYKSISTÄ JA MONITOIMIPAIKKAISTEN LIIKEYRITYSTEN TEOLLISUUDEN JA RAKENTAMISEN TOIMIPAIKOISTA

Perusjoukko: Toimipaikkarekisterissä ja rakennetilastojen kyselykannassa olevat yksitoimipaikkaiset liikeyritykset sekä monitoimipaikkaisten liikeyritysten toimipaikat, joiden toimiala on TKANNAN mukaan teollisuus tai rakentaminen (toimialat C-F, liite 1). Otosyksikkö: Liiketoimipaikka (yksitoimipaikkainen liikeyritys = liiketoimipaikka), otantamenetelmä: STRSRS eli ositettu otanta käyttäen yksinkertaista satunnaisotantaa palauttamatta, kiintiöintimenetelmä: suhteellinen kiintiöinti,  $N = 213\,467$ ,  $n = 85670$  ja ositteiden lukumäärä  $H = 42$ . Yksittäisen liiketoimipaikan  $k$  sisältymistodennäköisyys  $p_k =$  Monitoimipaikkaisten liikeyritysten teollisuuden ja rakentamisen toimipaikoille eli ositteiden 1-8 toimipaikoille  $p_k = 0.5$  ja yksitoimipaikkaisille liikeyrityksille eli ositteille 9-42  $p_k = 0.4$ .

Voidaan myös muodostaa otos yhdistämällä otokset 2 ja 3 ja saadaan otos yksitoimipaikkaisista liikeyrityksistä Tämä otanta-asetelma on seuraavassa ja on 6. otanta-asetelma.

## 6. OTANTA-ASETELMA

Saadaan yhdistämällä otokset 2 ja 3. Perusjoukko: Toimipaikkarekisterissä ja rakennetilastojen kyselykannassa olevat yksitoimipaikkaiset liikeyritykset, joiden tilinpäätöstiedot ovat YKANNASSA tai TKANNASSA. Otosyksikkö: Liiketoimipaikka (yksitoimipaikkainen liikeyritys = liiketoimipaikka), otantamenetelmä: STRSRS eli ositettu otanta käyttäen yksinkertaista satunnaisotantaa palauttamatta, kiintiöintimenetelmä: suhteellinen kiintiöinti,  $N = 210\,657$ ,  $n = 84275$ , ositteiden lukumäärä  $H = 34$ . Yksittäisen liiketoimipaikan sisältymistodennäköisyys  $p_k = 0.4$

Koska kaikissa edellä mainituissa otanta-asetelmissä käytetään ositettua otantaa, ovat estimaattorit toimialakohtaisille tilinpäätöstiedoille seuraavaa muotoa (Särndal ym., 1992, 103):

Estimaattori totaalille (kokonaismäärälle),  $\hat{t}_d = \sum_{h=1}^H \hat{t}_{hd}$ , missä  $d = 1, \dots, D$  viittaa toimialaan ja  $h = 1, \dots, H$  ositteeseen. Varianssin estimaattori  $\hat{V}(\hat{t}_d) = \sum_{h=1}^H \hat{V}(\hat{t}_{hd})$ . Keskivirhe  $s.e._d = \sqrt{\hat{V}(\hat{t}_d)}$ .

Kun edellä mainittuja otoksia, eli otoksia 1 - 6, vertaa niiden perusjoukkoihin, niin huomataan, että toimipaikan toimialojen, toimipaikkaa vastaavan yrityksen lajin ja toimipaikan PK-suuruusluokan jakaumat otoksissa vastaavat niiden jakaumaa perusjoukoissaan. Tämä johtuu siitä, että on käytetty suhteellista kiintiöintiä ja osittaminen on tehty toimialan, PK-suuruusluokan ja toimipaikkaa vastaavan yrityksen lajin mukaan. Muissa muuttujissa otosjakaumat poikkeavat jonkin verran perusjoukkojen jakaumista.

Liitteessä 5 on kuvattu toimipaikan sijaintiläänin jakautumista toimialoittain otoksissa ja perusjoukossaan. Siellä ovat myös otoksista lasketut Pearsonin korrelaatiokertoimet toimialoittain ja niiden vertailu vastaaviin Pearsonin korrelaatiokertoimiin perusjoukossa. Niitä tarkasteltaessa huomataan, että Pearsonin korrelaatiot

otoksissa ovat lähellä vastaavia korrelaatiota perusjoukossa. On huomattava, että otoskorrelaatiot eivät ole kaikissa osajoukoissa vastaavia perusjoukon korrelaatiota suurempia, vaan joissakin osajoukoissa otoskorrelaatio on vastaavaa perusjoukon korrelaatiota pienempi. Toimipaikkojen alueelliset jakaumat ovat lähes identtiset verrattaessa otoksien jakaumia niitä vastaaviin perusjoukkoihin. Tämä saattaa kertoa siitä, että toimipaikan toimiala määrittää pitkälle toimipaikan sijainnin. Samalla se kertoo toimipaikkojen keskittymisestä Etelä- ja Länsi-Suomen lääneihin.

Toimialojen koot toimipaikkojen lukumäärällä mitattuna vaihtelevat suuresti, pienimmällä toimialalla (X) toimii 11 toimipaikkaa ja suurimmalla toimialalla (G) toimii 45071 toimipaikkaa. Toimialan X kohdalla ei voidakaan tehdä mitään päätelmiä pienen perusjoukon koon vuoksi. Toimialakohtaisten kokoerojen lisäksi pienten ja keskisuurten yritysten toimipaikkojen ja suurten yritysten toimipaikkojen välillä on kokoeroja kaikilla toimialoilla. Pienten ja keskisuurten yritysten toimipaikkoja on paljon suuria enemmän. Kaiken kaikkiaan ositteiden koot vaihtelevat perusjoukossa välillä 11 (toimiala X) ja 44820 (toimiala G). Suhteellisen kiintiöinnin vuoksi osittaminen ei tasoita ositekohtaisissa otoskoissa esiintyviä eroja, vaan toteutuneet otoskoot ovat lähes samat kuin odotusarvonsa.

Taulukossa 5.1 ovat toimipaikkojen lukumäärät eri toimialoilla siinä joukossa, jonka jalostusarvot ovat tiedossa ( $N = 213\,443$ , n. 88 % kaikista toimipaikoista). Taulukko sisältää myös toimialakohtaiset otoskoot ja näiden odotusarvot. Niitä tarkasteltaessa huomataan, että ne poikkeavat toteutuneista toimialakohtaisista otoskoista siten, että teollisuuden ja rakentamisen toimialojen otoskoot ovat odotusarvoaan suurempia. Tämä on tarpeellista, koska tutkimus kohdistuu teollisuuteen ja rakentamiseen ja näillä toimialoilla on kiinnostuksen kohteena oleva osajoukko toimialan lisäksi toimipaikkaa vastaavan yrityksen laji eli osajoukkoja on toimialoilla C-F useampi kuin muilla toimialoilla. Osittaminen kiinnostuksen kohteena olevien osajoukkojen suhteen yleensä tasoittaa osajoukkojen kokoeroja, mutta riippuu käytetystä kiintiöintimenetelmästä ja menetelmästä, joilla otokset ositteista poimitaan (Lehtonen ja Pahkinen, 2004).

**Taulukko 5.1.** Toimialojen koot perusjoukossa ( $N_d$ ) ja otoksessa ( $n_d$ ) niiden toimipaikkojen joukossa, jonka jalostusarvo on tiedossa. Lisäksi otoskokojen odotusarvot  $E(n_d) = n * (N_d / N)$ .

| TOIMIALA | $N_d$  | $n_d$ | $E(n_d)$ |
|----------|--------|-------|----------|
| A        | 6222   | 2490  | 2497.34  |
| B        | 626    | 251   | 251.26   |
| C        | 1215   | 497   | 487.67   |
| D        | 26457  | 10789 | 10619.09 |
| E        | 950    | 402   | 381.30   |
| F        | 28684  | 11520 | 11512.95 |
| G        | 45071  | 18029 | 18090.23 |
| H        | 9416   | 3767  | 3779.32  |
| I        | 22794  | 9118  | 9148.87  |
| J        | 2365   | 946   | 949.244  |
| K        | 39561  | 15825 | 15878.67 |
| M        | 1468   | 588   | 589.21   |
| N        | 12008  | 4804  | 4819.67  |
| O        | 16595  | 6639  | 6660.77  |
| X        | 11     | 5     | 4.42     |
| YHT.     | 213443 | 85670 | 85670    |

### 5.3 Estimointiasetelmat- ja strategiat

Estimoinnin sekä estimointiasetelman- ja strategian käsitteet on määritelty luvussa 5.1. Tutkimuksessa estimoidaan totaaleja erilaisissa perusjoukon osajoukoissa. Tuntematon parametri on siis totaali jossain osajoukossa eli  $t_{yd} = \sum_{U_d} y_k$ , missä  $d = 1, \dots, D$ . Osajoukkokohtaisesta estimoinnista on puhuttu luvussa 5.2.

Estimoitavat totaalit ovat toimiala- ja toimipaikkaa vastaavan yrityksen lajikohtaiset totaalit jalostusarvolle. Estimointi tehdään käyttämällä luvussa 3.2 mainittuja muuttujia ja luvussa 5.2 mainittujen otanta-asetelmien perusteella poimituista otoksia. Koska kyseessä on osajoukkokohtainen estimointi ja käytettävissä on tehokasta lisäinformaatiota (luku 5.1), käytetään joitakin otantateorian keskuudessa yleisesti hyväksytyistä malliavusteisista osajoukkokohtaisien totaalien estimaattoreista ja verrataan niitä toisiinsa. Niitä ovat suora ja epäsuora Horvitz-Thompson estimaattori (HT-estimaattori), synteettinen estimaattori ja yleistetty regressioestimaattori (GREG-estimaattori). Tutkimuksessa lasketaan siis estimaatteja kuutta eri estimaattoria käyttämällä. Ne kaikki estimoivat osajoukkojen totaaleja. Osajoukkokohtaisesta estimoinnista on kerrottu artikkelissa Särndal (2001) ja kirjoissa Särndal ym. (1992) sekä Lehtonen ja Pahkinen (2004).

Horvitz-Thompson estimaattori poikkeaa muista tutkimuksessa käytetyistä estimaattoreista, koska HT-estimaattien laskemiseen ei välttämättä tarvita lisäinformaatiota estimoinnin vaiheessa. Sitä käytetäänkin yleensä vain silloin kun lisäinformaatiota ei ole estimointivaiheessa saatavilla. Jos lisäinformaatiota on käytettävissä, HT-estimaattien avulla voidaan tutkia onko lisäinformaation käyttö tehostanut estimointia (esim. onko lisäinformaation käyttö pienentänyt keskivirhettä) ja miten hyvä käytetty otanta-asetelma on ollut (mitä parempi otanta-asetelma on, sitä pienempiä ovat HT-estimaattien keskivirheet). Yleisesti estimointistrategian tehokkuutta voidaan tutkia asetelmakertoimen avulla. Siitä on puhuttu luvussa 5.1. (Lehtonen ja Pahkinen, 2004).

Estimoinnissa käytetyn lisäinformaation ja estimaattorin lisäksi otoksen poiminnassa käytetty otanta-asetelma (käytännössä se riippuu siitä millaista lisäinformaatiota on käytössä ennen otoksen poimintaa ja miten sitä käytetään) vaikuttaa estimoinnin tehokkuuteen. Mitä parempi otanta-asetelma on, sitä parempia ovat laaditut HT-estimaatit keskivirheen mielessä. Kuitenkin estimoinnin tehokkuuteen ei vaikuta siis pelkästään käytetty estimaattori tai otanta-asetelma, vaan niiden yhdistelmä (ts. käytettävän estimaattorin, lisäinformaation ja otanta-asetelman vuorovaikutus). Se kuinka hyvä jokin otanta-asetelma on, riippuukin pitkälti siitä millaista lisäinformaatiota käytetään estimointivaiheessa, mikä on tutkimuksen kohde (tuntematon parametri) ja tulosmuuttujat (joskus voi jopa käydä niin, että yksinkertainen satunnaisotanta on PPS-otantaa parempi, jos käytetään vahvaa lisäinformaatiota estimointivaiheessa). Ei voidakaan sanoa sitä mikä otanta-asetelma olisi paras, vaan se riippuu käytettävästä lisäinformaatiosta otannan ja estimoinnin vaiheissa sekä tutkimuksen kohteesta. Lisäinformaation käytöstä otannassa ja estimoinnissa on kerrottu luvussa 5.1. (Lehtonen ja Pahkinen, 2004).

Tutkimuksessa laaditaan sekä suoria että epäsuoria jalostusarvon totaalin toimialakohtaisia GREG- ja synteettisiä estimaatteja. Jos malli määrittää kaikille osajoukoille yhteiset  $\beta$ -parametrit, on kyseessä **epäsuora**

totaalin estimaattori. **Suoraa** totaalin estimaattoria käytettäessä taustalla on malli, jossa määritellään jokaiselle osajoukolle erilliset  $\beta$ -parametrit. Suoran totaalin estimaatin laskemisessa  $\beta$ -parametrien estimoinnissa käytetään tulosmuuttujan havaintoarvoja vain siitä osajoukosta, jonka kokonaismäärää pyritään estimoimaan, kun taas epäsuoraa estimaattoria käytettäessä  $\beta$ -parametrien estimaatit lasketaan käyttämällä kaikkia tulosmuuttujan havaintoarvoja. (Särndal, 2001, 26, Lehtonen ja Pahkinen, 2004).

Ainoa ero suoran ja epäsuoran totaalin estimaattorin välillä on siis mallin  $\beta$ -parametreissa. Kuitenkin sekä epäsuoran että suoran totaalin estimaattorin käyttäminen tarkoittaa sitä, että apumuuttujien totaalit tarkoittavat osajoukkokohtaisia apumuuttujien totaaleja ja harhankorjaustermit lasketaan käyttämällä otoshavaintoja vain siitä osajoukosta, jolle ollaan estimaattia laskemassa. Tästä johtuu se, että jokaiselle osajoukolle laaditaan myös epäsuoraa totaalin estimaattoria käytettäessä erilliset totaaliestimaatit.

Yleistettyjen regressioestimaattoreiden ja synteettisten estimaattoreiden tehokkuus riippuu käytettävän tulosmuuttujan ja apumuuttujan välisen riippuvuuden suuruudesta, eli apumuuttujan ja tulosmuuttujan välisen Pearsonin korrelaatiokerroimen suuruudesta. Mitä suurempi korrelaatiokerroin on, sitä tehokkaampi totaalin estimaattori on. (Pahkinen ja Lehtonen, 1989, 89–96).

Estimoinnissa käytettävässä olevasta aineistosta on poimittu 6 eri otosta samalla otanta-asetelmalla eli STR-SRS-otannalla käyttäen suhteellista kiintiöintiä (luku 5.2). Voidaankin ajatella, että strategiat eroavat toisistaan siinä, että estimaatit lasketaan eri estimaattoreilla (näitä estimaattoreita on kuusi). Täten ajatellen tutkimuksessa on käytetty kuutta estimointistrategiaa. Ne ovat taulukossa 5.2.

Estimointistrategioissa käytettyjen totaalin estimaattoreiden kaavat ovat taulukossa 5.3 ja niiden varianssien estimaattorit taulukossa 5.4. Estimointistrategioiden pohjalla ovat seuraavat oletukset. Oletetaan, että on käytettävissä äärellinen perusjoukko  $U = \{1, \dots, k, \dots, N\}$ . Tästä perusjoukosta poimitaan otos  $s$  jollain otanta-asetelmalla. Alkion  $k$  todennäköisyyttä tulla otokseen  $s$  eli alkion  $k$  sisällymistodennäköisyyttä merkitään

$\mathbf{p}_k$  :lla. Tällöin alkion  $k$  asetelmapainoa ts. sisällymistodennäköisyyden käänteislukua merkitään  $a_k = \frac{1}{\mathbf{p}_k}$ ,

joka on tutkimuksessa  $= \frac{N_H}{n_h}$  ja  $y_k$  = riippuvan muuttujan  $y$  havaintoarvo alkion  $k$ ,  $k \in s$ . Käytettävissä

on jonkun/joidenkin apumuuttujien ( $J$  kappaletta) havaintoarvot perusjoukossa tai jonkun/joidenkin apumuuttujien totaalit perusjoukossa (tutkimuksessa toimipaikan liikevaihto, keskimääräinen henkilöstön määrä tilikauden aikana ja toimipaikalle estimoidut palkat). Tätä sanotaan lisäinformaatioksi estimoinnin vaiheessa.

Merkitään  $\mathbf{x} = (1, x_{1k}, \dots, x_{jk}, \dots, x_{jk})'$  muuttujien  $x_1 \dots x_J$  havaintoarvoja alkion  $k$ ,  $k \in U$ . Tällöin jokaiselle

alkion  $k \in s$  on tiedossa  $(y_k, \mathbf{x}_k)$ , ja aputietoa  $\mathbf{x}$  on tiedossa kaikille  $k \in U$ . Merkitään apumuuttujien to-

taaleja perusjoukossa  $t_{xd} = \sum_{U_d} x_k$ , missä  $d = 1, \dots, D$ . (Esteveo ym. 1995, 183).



Oletetaan vielä, että edellä mainittu äärellinen perusjoukko  $U$  voidaan jakaa osajoukkoihin ts. muodostaa jälkiositteita  $U_1, \dots, U_d, \dots, U_D$ . Lisäksi oletetaan, että tiedetään apumuuttujien havaintoarvot perusjoukossa. Voi myös olla niin, että tiedetään kyseisten apumuuttujien havaintoarvot otoksessa sekä edellä mainituissa perusjoukon osajoukoissa  $U_1, \dots, U_d, \dots, U_D$  jostain rekisteristä. Tällöin voidaan merkitä

$\mathbf{x}_d = (1, x_{1dk}, \dots, x_{jdk}, \dots, x_{jdk})'$  apumuuttujien havaintoarvoja osajoukossa  $d$  ja  $\mathbf{t}_{xd}$ :llä apumuuttujien totaaleja osajoukossa  $d$   $\mathbf{t}_{xd} = \sum_{k \in U} \mathbf{x}_{dk} = (\sum_{k \in U} x_{1dk}, \dots, \sum_{k \in U} x_{jdk})'$ ,  $d = 1, \dots, D$ . Kun nämä oletukset toteutuvat, voi-

daan tulosmuuttujan  $y$  totaali jossain perusjoukon osajoukossa estimoida taulukossa 5.3 mainituilla estimaattoreilla. Seuraavassa selostetaan tarkemmin käytettävät estimointistrategiat. Käytettävien estimaattoreiden kaavat on laadittu tutkimuksessa käytettävän otanta-asetelman mukaisiksi (STRSRS). (Estevao ym., 1995, 186).

Estimaattoreiden yhteisiä ominaisuuksia: A) Estimaattorin keskivirheen estimaattori:  $s.e_d = \sqrt{\hat{V}(\hat{t}_{ydp})}$

B) Estimaattorin variaatiokertoimen estimaattori:  $cv(\hat{t}_{ydp}) = \sqrt{\hat{V}(\hat{t}_{ydp})} / \hat{t}_{ydp}$  C) 95%:n luottamusväli tuntemattomalle tulosmuuttujan  $y$  totaalille osajoukossa  $d$  (esim. toimialalla  $d$ ) on seuraavaa muotoa:

$$\hat{t}_{ydp} \pm 1.96 \sqrt{\hat{V}(\hat{t}_{ydp})}$$

D) Estimaattori käyttäen informaatioryhmiä: Apumuuttujien totaalit, harhankorjaustermit ja HT-estimaattorin keskivirheet lasketaan haluttuun informaatioryhmään ja kiinnostuksen kohteena olevaan osajoukkoon kuuluvien joukosta. Suoraa estimaattoria käytettäessä lasketaan parametrien estimaatit ja selitysasteet haluttuun osajoukkoon ja informaatioryhmään kuuluvien joukosta. Epäsuoraa estimaattoria käytettäessä lasketaan parametrien estimaatit ja selitysasteet käyttäen koko otosta.

**Taulukko 5.2.** Tutkimuksessa käytetyt estimointistrategiat (Lähteet: Särndal ym. 1992, Särndal, 2001, Lehtonen ja Pahkinen, 2004).

| TOTAALIN ESTIMAATTORIN TYYPPI + OTANTA-ASETELMA | TOTAALIN ESTIMAATTORIN LAJI | TOTAALIN ESTIMAATTORI     | STRATEGIAN NUMERO JA NIMI |
|---|-----------------------------|---------------------------|---------------------------|
| SUORA + SRSSTR                                  | ASETELMAPERUSTEINEN         | HT-ESTIMAATTORI           | 1 HT-S                    |
|   | MALLIAVUSTEINEN             | GREG-ESTIMAATTORI         | 2 GREG-S                  |
|   | MALLIPERUSTEINEN            | SYNTEETTINEN ESTIMAATTORI | 3 SYN-S                   |
| EPÄSUORA + SRSSTR                               | ASETELMAPERUSTEINEN         | HT-ESTIMAATTORI           | 4 HT-ES                   |
|   | MALLIAVUSTEINEN             | GREG-ESTIMAATTORI         | 5 GREG-ES                 |
|   | MALLIPERUSTEINEN            | SYNTEETTINEN ESTIMAATTORI | 6 SYN-ES                  |

Suoria estimaattoreita ovat estimointistrategioiden 1–3 käyttämät totaalin estimaattorit (suorat HT-, GREG- ja synteettiset totaalin estimaattorit) ja epäsuoria estimaattoreita ovat estimointistrategioiden 4-6 käyttämät totaalin estimaattorit (epäsuorat HT-, GREG- ja synteettiset estimaattorit). Asetelmaperusteisia estimaattorei-

ta ovat strategioiden 1 ja 4 käyttämät totaalin estimaattorit (suorat ja epäsuora HT-estimaattorit), malliavusteisia estimaattoreita ovat strategioiden 2 ja 5 (suorat ja epäsuorat GREG-estimaattorit) käyttämät totaalin estimaattorit ja mallista riippuvia ovat strategioiden 3 ja 6 käyttämät totaalin estimaattorit (suorat ja epäsuorat synteettiset totaalin estimaattorit).

**Taulukko 5.3.** Tutkimuksessa käytettävät totaalin estimaattorit (Lähteet: Särndal ym., 1992, Särndal, 2001, Lehtonen ja Pahkinen, 2004).

| Estimointistrategian numero ja lyhenne: | Käytetyn totaalin estimaattorin ja kaava:   |
|---|---|
| 1 = HT-S                                | $\hat{t}_{ydp} = \sum_{k \in S_d} a_k y_k$ , missä $a_k = p_k^{-1}$   |
| 4 = HT-ES                               | $\hat{t}_{ydp} = N_d \bar{y}$ , missä $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$  |
| 2 = GREG-S                              | $\hat{t}_{ydgREG} = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in S_d} a_k (y_k - \hat{y}_k) = \mathbf{t}'_{xd} \hat{\mathbf{B}}_d + \sum_{k \in S_d} a_k (y_k - \hat{y}_k)$ , missä<br>$\hat{y}_k = \mathbf{x}'_k \hat{\mathbf{B}}_d$ , $\hat{\mathbf{B}}_d = \left( \sum_{k \in S_d} a_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{k \in S_d} a_k \mathbf{x}_k y_k$ ja $\sum_{k \in S_d} a_k (y_k - \hat{y}_k) =$ harhankorjaustermi |
| 5 = GREG-ES                             | $\hat{t}_{ydgREG} = \mathbf{t}'_{xd} \hat{\mathbf{B}} + \sum_{k \in S_d} a_k (y_k - \hat{y}_k) = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in S_d} a_k (y_k - \hat{y}_k)$ , missä $\sum_{k \in S_d} a_k (y_k - \hat{y}_k) =$ harhankorjaustermi, $\hat{y}_k = \mathbf{x}'_k \hat{\mathbf{B}}$ ja $\hat{\mathbf{B}} = \left( \sum_{k \in S} a_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{k \in S} a_k \mathbf{x}_k y_k$               |
| 3 = SYN-S                               | $\hat{t}_{ydsYN} = \mathbf{t}'_{xd} \hat{\mathbf{B}}_d = \sum_{k \in U_d} \hat{y}_k$ , missä $\hat{\mathbf{B}}_d = \left( \sum_{k \in S_d} a_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{k \in S_d} a_k \mathbf{x}_k y_k$ ja $\hat{y}_k = \mathbf{x}'_k \hat{\mathbf{B}}_d$   |
| 6 = SYN-ES                              | $\hat{t}_{ydsYN} = \mathbf{t}'_{xd} \hat{\mathbf{B}} = \sum_{k \in U_d} \hat{y}_k$ , missä $\hat{\mathbf{B}} = \left( \sum_{k \in S} a_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{k \in S} a_k \mathbf{x}_k y_k$ ja $\hat{y}_k = \mathbf{x}'_k \hat{\mathbf{B}}$   |

Lisäinformaatiota käyttäviä estimaattoreita ovat GREG-estimaattori ja synteettinen estimaattori. GREG-estimaattori on tehokkuudeltaan aina ainakin yhtä hyvä kuin Horvitz-Thompson estimaattori, koska sen keskivirhe (varianssin neliöjuuri) on pienempi tai yhtä suuri kuin Horvitz-Thompson estimaattorin. Jos lisäinformaation ja tulosmuuttujan riippuvuus on täydellistä (yhteiskorrelaatiokerroin on 1 tai -1), niin GREG-estimaattorin varianssin estimaatti on nolla. Jos taas tulosmuuttujan ja lisäinformaation välillä ei ole riippuvuutta (yhteiskorrelaatiokerroin on nolla), niin GREG-estimaattorin varianssi on sama kuin Horvitz-Thompson estimaattorin varianssi. Silloin lisäinformaation käytöstä estimoinnin yhteydessä (GREG-estimaattorin) ei ole hyötyä varianssin ja keskivirheen pienenemisen mielessä. Sitä kuinka paljon GREG-estimaattorin käyttäminen pienentää estimaattorin keskivirhettä voidaan tutkia asetelmakertoimen estimaattorin *deff* avulla. Se kertoo käytetyn estimointistrategian tehokkuudesta (katso luku 5.1).

**Taulukko 5.4.** Tutkimuksessa käytettyjen totaalin estimaattoreiden varianssien estimaattorit. GREG-estimaattorien varianssien estimaattorit ovat approksimatiivisia, ja on olemassa myös muita tapoja approksimoida totaalin GREG-estimaattorin varianssia. (Lähteet: Särndal ym., 1992, Särndal, 2001, Lehtonen ja Pahkinen, 2004).

| Estimointstrategian numero ja lyhenne: | Käytetyn totaalin varianssin estimaattorin kaava:  |
|--|--|
| 1 = HT-S                               | $\hat{V}(\hat{t}_{ydp}) = \sum_{h=1}^H \hat{V}(\hat{t}_{yhdp}) = N_d^2 \left( \frac{1}{n_d} - \frac{1}{N_d} \right) \frac{1}{n_d - 1} \sum_{k=1}^{n_d} (y_{dk} - \bar{y}_d)^2$   |
| 4 = HT-ES                              | $\hat{V}(\hat{t}_{ydp}) = N_d^2 \left( \frac{1}{n_d} - \frac{1}{N_d} \right) \frac{1}{n_d - 1} \sum_{k=1}^{n_d} (y_{dk} - \bar{y})^2$  |
| 2 = GREG-S                             | $\hat{V}(\hat{t}_{ydGREG}) = (1 - R_d^2) \hat{V}_{ydp}$ , missä $\hat{V}_{ydp} = N_d^2 \left( \frac{1}{n_d} - \frac{1}{N_d} \right) \frac{1}{n_d - 1} \sum_{k=1}^{n_d} (y_{dk} - \bar{y}_d)^2$ ja $R_d^2 =$ apumuuttujien ja tulosmuuttujan välisen yhteiskorrelaatiokertoimen neliön estimaattori (selityksasteen estimaattori) osajoukossa $d$ .                       |
| 5 = GREG-ES                            | $\hat{V}(\hat{t}_{ydGREG}) = (1 - R^2) \hat{V}_{ydp}$ , missä $\hat{V}(\hat{t}_{ydp}) = N_d^2 \left( \frac{1}{n_d} - \frac{1}{N_d} \right) \frac{1}{n_d - 1} \sum_{k=1}^{n_d} (y_{dk} - \bar{y})^2$ ja $R^2 =$ apumuuttujien ja tulosmuuttujan välisen yhteiskorrelaatiokertoimen neliön estimaattori (selityksasteen estimaattori), joka on sama kaikissa osajoukoissa. |

Estimointistrategian valinnassa on ongelmana se, että etenkin epäsuorat synteettiset estimaattorit ovat harhaisia, jos  $\beta$ -parametrit eivät ole samoja kaikissa osajoukoissa (tutkimuksessa kaikilla toimialoilla). Toisaalta niillä on pienempi varianssi kuin suorilla estimaattoreilla, jos havaintoja osajoukossa on vähän. Tämä antaa aiheen olettaa, että epäsuorien totaalin estimaattorien käyttäminen tutkimuksessa saattaa johtaa harhaisiin tuloksiin, mutta niiden avulla saaduilla estimaateilla saattaa olla pienempi varianssi, etenkin jos otoskoko osajoukossa on pieni. Suurissa osajoukossa näin ei välttämättä tapahdu. Epäsuorat synteettiset estimaatit ovat ainoa vaihtoehto joissakin kiinnostuksen kohteena olevissa osajoukoissa, koska laskennallisen käyttökäteen tietoja ei ole saatavilla kaikissa kiinnostuksen kohteena olevissa osajoukoissa, (niissä havaintojen lukumäärä on nolla). Särndal (2001, 42-45).

Jos epäsuoralla totaalin estimaattorilla on pienempi varianssi kuin suoralla, on kyseessä **voiman lainaaminen (borrowing strength)** muilta osajoukoilta. Voiman lainaamista GREG-estimaattorin tapauksessa tapahtuu, jos otoksesta estimoitu mallin selityksaste  $R^2$  on suurempi kuin osajoukkokohtaisen selityksasteen estimaatti  $R_d^2$ . Tämä on pääteltävissä suoran ja epäsuoran totaalin GREG-estimaattorin varianssien estimaattien kaavoista (taulukko 5.4). Yleensä pienissä osajoukoissa (ts. niissä osajoukoissa, joilla on pieni otoskoko) tapahtuu voiman lainaamista eli epäsuoran synteettisen estimaatin MSE ja GREG-estimaatin keskivirhe ovat pienempiä kuin suoran estimaatin. Mutta suurissa osajoukoissa näin ei välttämättä tapahdu. Vaikka voiman lainaamista tapahtuisikin, epäsuora synteettinen estimaatti saattaa olla harhaisempi kuin suora. (Särndal, 2001, 42-45).

Estimointistrategian valinnassa on otettava myös huomioon erot GREG-estimaattorien ja synteettisten estimaattorien välillä. Totaalin synteettiset estimaattorit ja yleistetyt regressioestimaattorit ovat sikäli samanlaisia, että niillä on samat estimaattorit  $\beta$ -parametreille, eli ne perustuvat samaan malliin. Totaalin GREG-estimaatti poikkeaa kuitenkin synteettisestä estimaatista, jos GREG-estimaatin harhankorjaustermi ei ole nolla. Tämä tarkoittaa sitä, että synteettinen estimaatti poikkeaa GREG-estimaatista (sekä suorien että epäsuorien estimaattorien tapauksessa), jos  $\sum_{S_d} a_k (y_k - \hat{y}_k) \neq 0$ . Harhankorjaustermi on siis mallin jäännöstermi painotettuna otantapainolla ja se lasketaan otoksessa olevista haluttuun osajoukkoon kuuluvista havainoista.

GREG-estimaattori on ainakin likimain harhaton, mutta synteettinen estimaattori ei ole. Synteettisen estimaattorin toimivuus riippuukin täysin valitun mallin toimivuudesta eli siitä kuinka hyvin käytetyt apumuuttujat selittävät tulosmuuttujien arvoissa esiintyvää vaihtelua, mutta GREG-estimaattorin tapauksessa harhankorjaustermi korjaa mallin vaikutusta. Voidaan ajatella epäsuoran synteettisen estimaattorin olevan harhaisempi kuin suora, koska epäsuoran synteettisen estimaattorin toimivuus riippuu siitä ovatko apumuuttujien ja tulosmuuttujan väliset riippuvuussuhteet samoja kaikissa osajoukoissa. Synteettinen estimaattori on harhaton, jos

$$\sum_{k \in U_d} E_k = \sum_{k \in U_d} (y_k - \mathbf{x}'_k \boldsymbol{\beta}) = 0. \text{ Harha on seuraavaa muotoa: } \text{Bias}(\hat{t}_{yd,SYN}) \approx - \sum_{k \in U_d} E_k = - \sum_{k \in U_d} (y_k - \mathbf{x}'_k \boldsymbol{\beta}).$$

Kuitenkin synteettisellä estimaattorilla saattaa olla pienempi varianssi kuin GREG-estimaattorilla. Synteettisen estimaattorin tapauksessa varianssin estimoinnin tilalla tulisi käyttää keskineliövirhettä (MSE, Mean Squared Error), mutta GREG-estimaattorilla keskineliövirhe eli MSE on sama kuin varianssi, koska GREG-estimaattori on likimain harhaton. Siitä onko GREG-estimaattorin varianssi pienempi kuin synteettisen estimaattorin MSE, riippuu se antaako GREG-estimaattorin vai synteettisen estimaattorin käyttäminen paremmat tulokset keskineliövirheen mielessä. Yhden otoksen perusteella ei kuitenkaan saada tietoa synteettisen estimaattorin harhan suuruudesta, koska ei tiedetä estimaattien jakaumaa (tätä voitaisiin arvioida ns. Monte Carlo -kokeiden avulla). Synteettisen estimaattorin harhasta ja GREG-estimaattorin sekä synteettisen estimaattorin keskivirheistä riippuu se kumpi estimaattori antaa parempia estimaatteja totaaleille. Jos synteettisen estimaattorin harha on suuri, on GREG-estimaattori MSE:n mielessä yleensä parempi, vaikka GREG-estimaattorin varianssi onkin synteettisen estimaattorin varianssia suurempi.

Tutkimuksessa pyritään valitsemaan malleja, joilla synteettisten estimaattien ja GREG-estimaattien väliset erot olisivat mahdollisimman pieniä. Tämä johtui siitä, että tutkimuksessa käytettiin estimoituja mallien parametreja yritysten tilinpäätöstietojen toimipaikkaistamiseen ts. toimipaikkojen tilinpäätöstietojen imputoimiseen malliluovuttaja-menetelmällä. Tällöin imputoinnin tehokkuus riippuu täysin käytetyn mallin paikkaansapitävyydestä. Harhankorjaustermien itseisarvon suuruus kertoo mallin tehokkuudesta otoksessa ja malli, joka toimii hyvin otoksessa johtaa siihen, että jäännöstermit ovat nollia ja harhankorjaustermikin nolla. Tämä antaa aiheen olettaa, että imputointi on sitä tehokkaampaa mitä pienemmät erot synteettisten ja GREG-estimaattien välillä ovat.

## 5.4 Otantavirhe ja otannasta riippumattomat virheet

*Tilastollisissa tutkimuksissa esiintyy erilaisia vaihtelunlähteitä. Ne jaetaan yleensä **A) otantavirheeseen (sampling error) ja B) otannasta riippumattomiin virheisiin (nonsampling error)**.* Otantavirhe tarkoittaa perusjoukon osajoukon tutkimisesta (otostutkimuksesta) aiheutuvaa epävarmuutta tutkimuksen tuloksiin. Otantavirheen suuruutta arvioidaan yleensä laskemalla ns. asetelmavarianssi eli varianssin estimoinnin käyttämällä estimaattoria, joka ottaa huomioon käytetyn otanta-asetelman. Tutkimuksessa käytetyt menetelmät otantavirheen hallintaan ovat luvussa 5.3. (Lehtonen ja Pahkinen, 2004).

Otannasta riippumattomat virheet käsittävät vaihtelunlähteet, jotka johtuvat muusta kuin otannasta ja voivat siten ilmetä myös kokonaistutkimuksissa. ***Otannasta riippumattomat virheet voidaan jakaa A) kehikkovirheisiin, B) mittausvirheisiin, C) prosessointivirheisiin ja D) katoon.*** Otantavirheiden hallintaan on tällä hetkellä käytettävissä hyviä menetelmiä (mm. asetelmavarianssit ja variaatiokertoimet), mutta otannasta riippumattomien virheiden hallintamenetelmät eivät ole yhtä kehittyneitä. Kuitenkin tutkimustulosten epävarmuudesta suurin osa aiheutuu usein otannasta riippumattomista virheistä. (Djerf, 2001, 11-12, Särndal ym. 1992, 13-14, Lehtonen ja Pahkinen, 2004).

***Kehikkovirheillä** tarkoitetaan otantakehikon poikkeamista kohdeperusjoukosta. Se siis tarkoittaa tiedossa olevan perusjoukon eli kehikkoperusjoukon (esim. yritysrekisteri on kehikkoperusjoukko) poikkeamista tuntemattomasta todellisesta perusjoukosta eli kohdeperusjoukosta. **Kehikkovirheet jaetaan yleensä ylipeittoon (overcoverage), alipeittoon (undercoverage) ja kaksinkertaiseen listaukseen (duplicate listings)**.* (Särndal ym., 1992, 13-14, Lehtonen ja Pahkinen, 2004).

***Ylipeitto** tarkoittaa sitä osaa otantakehikosta, joka ei kuulu kohdeperusjoukkoon. Se johtuu siitä, että kohdeperusjoukkoon aiemmin kuulunut havaintoyksikkö on lakannut olemasta (esim. yritys on tehnyt konkurssin, toimipaikka lakkautettu tai yritys on fuusioitunut, eikä tätä ole kehikkoperusjoukko koottaessa tiedetty) tai havaintoyksikkö ei täytä kohdeperusjoukolle asetettuja ehtoja (esim. tutkittaessa vain niitä yrityksiä, joiden päätoimiala on teollisuus, ylipeittoa aiheuttavat yritykset, joiden päätoimiala ei todellisuudessa ole teollisuus).* (Djerf, 2001, 11-12 ja Särndal ym. 1992, 13-14).

***Kaksinkertainen listaus** tarkoittaa sitä, että sama havaintoyksikkö on kehikkoperusjoukossa mukana useamman kerran. Jos yrityksen nimi muuttuu, saattaa yritys tulla rekisteriin merkityksi kaksi kertaa, koska ei olla huomattu poistaa tietoja, jotka koskevat vanhalla nimellä toiminutta yritystä. Myös fuusiot voivat aiheuttaa kaksinkertaista listausta. Kaksinkertainen listaus siis itse asiassa aiheuttaa ylipeittoa ja se voidaankin nähdä yhtenä ylipeiton tyyppinä. Ylipeitto voidaan mitata ja se yleensä havaitaan jo tietoja kerätettäessä, jolloin havaitut ylipeittotapaukset voidaan poistaa otantakehikosta.* (Särndal ym. 1992, 13-14).

***Alipeittoa** aiheuttavat ne kohdeperusjoukkoon kuuluvat havaintoyksiköt, jotka eivät ole mukana kehikkoperusjoukossa. Ne ovat siis sellaisia kehikkoperusjoukkoon kuulumattomia havaintoyksiköitä, jotka täyttävät*

kohdeperusjoukkoon kuulumisen ehdot. Alipeitto johtuu usein päivityksen tai tutkimuksen aloittamisen myöhästyisestä. Päivityksen myöhästyminen tarkoittaa sitä, että otantakehikkoa ei ole päivitetty tai voitu päivittää heti tietojen muuttumisen jälkeen. Tietoja ei voida päivittää jos havaintoyksikkö, jota tieto koskee, ei ole ilmoittanut tietojen muuttumisesta. Tätä tapahtuu mm. jos fuusioituneet yritykset eivät ole ilmoittaneet fuusios- ta tai uusi yritys toimintansa alkamisesta. Tutkimuksen aloittamisen myöhästyminen tarkoittaa tutkimuksen aloittamista otantakehikon keräämistä myöhemmin. Alipeittoa ei ole niin helppo arvioida kuin ylipeittoa, koska alipeittotapaukset eivät ole tiedossa. Sitä voidaan arvioida vain tutkimalla sellaisia aineistoja, joista otantake- hikosta puuttuvat havaintoyksiköt voidaan kenties löytää. (Djerf, 2001, 11-12 ja Särndal ym. 1992, 13-14).

**Mittausvirhe** tapahtuu silloin kun annettu tieto jonkun muuttujan havaintoarvosta poikkeaa todellisesta ha- vaintoarvosta ainakin yhden havaintoyksikön kohdalla. Syitä tälle ovat mm. väärin muistaminen, väärin laske- minen, haastattelijan kysymyksen ymmärtäminen eri tavalla kuin tutkija on sen tarkoittanut ymmärrettävän ja valehtelu. **Prosessointivirhe** tarkoittaa sitä, että tallennettu tieto poikkeaa annetusta tiedosta. Syitä tälle ovat mm. näppäilyvirheet ja haastattelijan tiedon kirjaaminen väärin. Nykyään tietokone täydentää tietoja myös automaattisesti ja virheellinen polutus tai koodivirhe aiheuttaa systemaattisen virheen kaikkiin alkioihin, joihin toimenpide kohdistuu. (Djerf, 2001, 14).

Mittaus- ja prosessointivirheiden tutkiminen tapahtuu yleensä siten, että tutkitaan jokaisen havaintoyksikön antamia tietoja erikseen ja katsotaan ovatko ne ristiriidassa. Tutkitaan myös ns. poikkeavia havaintoarvoja (outliers) eli havaintoarvoja, jotka ovat poikkeuksellisen suuria tai pieniä ja arvioidaan sitä voisivatko ne kuu- lua muuttujan luonnolliselle vaihteluvälille vai onko annettu tieto kenties väärä. Edellä mainittuja tapoja arvioi- da mittaus- ja prosessointivirheitä kutsutaan loogiseksi testaukseksi. Yritysten rakenteet -yksikön rakenneti- lastojen kyselykantaa muodostettaessa loogista testausta tehdään paljon tutkimalla tilinpäätösten eri erien yhdenmukaisuutta mm. tutkimalla tuleeko tuloslaskelman eri erien summaksi yrityksen ilmoittama liikevoitto/ tappio ja onko ilmoitettu tieto, kuten jalostusarvo, saatavissa tuloslaskelman eristä Tilastokeskuksen käyttä- män laskemistavan mukaan. Yleensä yritykset käyttävät eri tapaa määrittellä kysytyjä asioita (esim. henkilös- tön määrää ei ilmoiteta keskimääräisen henkilöstön määrän käsitteen mukaan, vaan mukana ovat vain ne työntekijät, jotka ovat työskennelleet koko tilikauden ajan yrityksessä ja siksi esim. kesätyöntekijät ja harjoit- telijat puuttuvat laskelmasta.). (Karttunen, 2002, Lehtonen ja Pahkinen, 2004, Särndal ym., 1992).

**Kato** tarkoittaa sitä, että kaikilta havaintoyksiköiltä ei ole saatu tietoa kaikista tulosmuuttujista eli aineistossa on puuttuvia tulosmuuttujien havaintoarvoja. **Kato jaetaan eräkattoon (item nonresponse) ja yksikkökatoon (unit nonresponse)**. Yksikkökato on kyseessä silloin kun ainakin yhdeltä havaintoyksiköltä puuttuvat kaikkien tulosmuuttujien havaintoarvot. On hyvä huomata, että yksikkökadonkin tapauksessa voi olla saata- villa joitain tietoja (yleensä rekistereistä saatavia) koskien havaintoyksiköitä, joiden tietoja puuttuu. Näitä tie- toja voidaan käyttää lisäinformaationa estimoinnin vaiheessa. (Djerf, 2001, 13-14, Lehtonen ja Pahkinen, 2004).

Katoa aiheuttavat monet eri syyt. Pääsyyinä ovat tavoittamatta jäämiset ja kieltäytymiset. Katoa voivat aiheut- ta myös muut otannasta riippumattomat virheet (peittovirheet sekä mittaus- ja prosessointivirheet). Otan-

nasta riippumattomat virheet ovatkin vuorovaikutuksessa keskenään. Yksikkökadon syyt jaetaan neljään päätyyppiin, joita ovat:

- A) tavoittamatta jääneet (virheelliset osoitteet ja puhelinnumerot, puuttuvat tiedot puhelimista, puhelujen seurlonta ja tilapäiset poissalot)
- B) kieltäytymiset (johtuen mm. periaatteista, ajanpuutteesta, tilapäisestä sairaudesta, matkasta tai kieliongelmistä)
- C) Peittovirheet (kuten konkurssit, fuusiot sekä yrityksen toiminnan siirtyminen kokonaan ulkomaille) ja
- D) Muut syyt, kuten mittaus- tai prosessointivirheet, tekniset ongelmat (kuten lomakkeiden tuhoutuminen tai tietokoneen hajoaminen), posti ja muut tietoliikenneyhteydet, haastattelijan sairastuminen tai kenttätyöajan päättyminen ennen kuin tiedot on saatu. Eräkatoa aiheuttavat mittaus- ja prosessointivirheet, kieltäytymiset ja ongelmat tietoliikenneyhteyksissä.

***Kadon suuruuteen voidaan vaikuttaa tiedonkeruuvaiheessa ja kadon vaikutus ottaa huomioon estimointivaiheessa (kadon adjustointi kerätyssä aineistossa uudelleenpainotusmenetelmin tai imputoimalla).*** Kadon suuruuteen vaikuttavat kyselytutkimuksessa yhteiskunnan ilmapiiri, tutkimuksen aihe ja kiinnostavuus, tutkimuksen rasittavuus, tutkimusajankohta, tutkimuksen aikataulutus, tiedonkeruutapa ja haastattelijaeefekti. Kadon suuruuteen voidaan siis vaikuttaa ennakkoon tutkimusprosessin ja sen aikataulun huolellisella suunnittelulla. (Tilastokeskus, 2002 a, 48-53, Djerf, 2001, 17, Lehtonen ja Pahkinen, 2004).

***Tutkimuksen aikataulu*** vaikuttaa kadon suuruuteen siten, että mitä kauemmin aikaa tietojen keräämiseen on ja mitä isompi tutkimusorganisaatio on kooltaan, sitä pienempi on kato. Tämä johtuu siitä, että kun aikaa on tarpeeksi, voidaan kadon suuruuteen vaikuttaa vielä kenttätyövaiheessakin. Näitä vaikuttamistapoja ovat tavoittamattomien ja kieltäytyneiden uudelleentavoittamisyrietykset (callbacks) ja vain tärkeimpien kysymysten kysyminen kieltäytyneiltä. Lisäksi voidaan jo ennen tietojen palauttamiselle sovittua ajankohtaa muistuttaa tutkimukseen vastaamisesta ja vastaamisen tärkeydestä tutkimuksen tulosten laadun kannalta (reminders) ja poimia otos ei-vastanneista (subsampling of nonrespondents). Se kuinka kauan käytetään tietojen keräämiseen eli kenttätyövaiheeseen, riippuu käytettävässä olevasta rahamäärästä. Siksi kustannusten ja kenttätyövaiheen pituuden välillä onkin tehtävä kompromisseja. Sanotaankin, että alhaisista kustannuksista, nopeasta aikataulusta ja pienestä kadosta voi vain kaksi toteutua samassa tutkimuksessa. (Tilastokeskus, 2002a, 48-53, Djerf, 2001, 17).

***Yhteiskunnan ilmapiiri*** vaikuttaa katoon siksi, että eri maissa suhtaudutaan eri tavalla kyselyihin vastaamiseen ja vastausprosentit vaihtelevatkin maasta toiseen. Lisäksi vastausprosentti riippuu ***ajankohdasta*** jolloin tutkimus tehdään, esimerkiksi tutkimuksen ajoittuminen kesäaikaan lisää katoa. On huomattavissa, että tutkimusten vastausprosentit ovat laskussa eli kato on kasvava ongelma. Kadon suuruus ja luonne ovat hyvin riippuvaisia ***tutkimuksen aiheesta***. Jos aihe koetaan yrityksessä haitalliseksi yrityksen toiminnalle (mm. jos ajatellaan tiedon julkistamisella olevan hyötyä kilpaileville yrityksille), on kato suurta, mutta jos aihe koetaan yrityksen toiminnan kannalta hyödylliseksi mm. siksi, että sillä saadaan yrityksen taloudellisesta tilasta tietoa, kato on pientä. ***Yleensä kato on sitä suurempi mitä rasittavampi tutkimus on*** siihen osallistuvan kannalta. Tutkimus on yrityksen kannalta rasittava, jos siltä kysytään paljon sellaisia tietoja, joita ei tilinpäätöksestä

ole suoraan saatavissa (mm. toimipaikan tilinpäätöstietojen kysyminen kun yrityksessä ei ole käytetty toimipaikkakohtaista kirjanpitoa). **Tiedonkeruutapa** vaikuttaa katoon siten, että postikyselyissä on yleensä suurempi kato kuin käyntihaastatteluissa. Lisäksi postikyselyn tuloksissa on usein enemmän mittausvirheitä kuin käyntihaastattelun tai puhelinkyselyn tuloksissa. Puhelinkyselyihin vastaamisessa ei ole eroa muuta kuin tutkittavien havaintoyksiköiden tavoittamisen suhteen. Tiedonkeruutapa kannattaa valita tutkimuksen aiheen mukaan. Yleensä yritystutkimuksessa käytetään postikyselyä, jota täydennetään tarvittaessa puhelinkyselyllä. (Tilastokeskus, 2002 a, 48-53, Djerf, 2001, 17).

Puhelin- ja käyntihaastatteluissa haastattelija vaikuttaa tutkimuksen tuloksiin ja katon suuruuteen. Tätä kutsutaan **haastattelijaefektiksi**. Koska yritystutkimuksessakin tietoja joudutaan tiedustelemaan puhelimella, jos niitä ei postikyselyllä saatu, niin haastattelijaefekti esiintyy myös yritystutkimuksissa. Mitä enemmän haastattelijalla on kokemusta haastattelemisesta ja mitä koulutetumpi hän on, sitä pienempi kato on. Myös haastattelijan asenteet (etenkin asenne suostuttelemista kohtaan) ja vuorovaikutus haastateltavan ja haastattelijan välillä vaikuttavat katon suuruuteen ja koostumukseen. (Lehtonen, 1996).

Vaikka katon pienentämiseksi käytettäisiin kaikki käytössä olevat keinot, niin katoa yleensä esiintyy. Kato aiheuttaa tutkimusjoukon pienenemiseen perusjoukkoon nähden. Tämä johtaa siihen, että estimaattien keskiarvot kasvavat. Lisäksi vastanneiden joukon perusteella lasketut estimaatit voivat olla harhaisia. Seuraavassa on tästä esimerkki. (Tilastokeskus, 2002 a, 48-53, Djerf, 2001, 17-21, Lehtonen ja Pahkinen, 2004).

Oletetaan, että tarkoitus on estimoida tuntematonta parametria, joka on jonkun tulosmuuttujan  $y$  totaali perusjoukossa eli  $t = \sum_{k \in U} y_k$  tai perusjoukon osajoukossa eli  $t_d = \sum_{k \in U_d} y_k$ , missä  $d = 1, \dots, D$ . Kuitenkaan ei ole

saatu tulosmuuttujan  $y$  havaintoarvoja kaikilta tutkimukseen kuuluvilta havaintoyksiköiltä eli on katoa. Oso  $s$  (tai kokonaistutkimuksessa perusjoukko  $U$ ) voidaan jakaa kahteen osaan eli jälkiositteeseen, jota merkitään  $U = (U_r, U_{n-r})$ . Tämä tapahtuu siten, että ensimmäisen ositteeseen muodostavat ne havaintoyksiköt, joille on tiedossa tulosmuuttujan havaintoarvo (vastanneiden joukko on kokoa  $n_r < n$ ) ja toisen ositteeseen muodostavat ne havaintoyksiköt, joille ei ole tiedossa kyseisen tulosmuuttujan havaintoarvoa (katojoukko on kokoa  $n_{r-n} = n - n_r$ ). Toisin sanoen vastaustodennäköisyys ositteessa 1 on yksi ja ositteessa 2 nolla. Tällöin HT-estimaattori harhainen, jos vastanneiden joukon keskiarvo tulosmuuttujalle poikkeaa tuntemattomasta katojoukon keskiarvosta kyseiselle tulosmuuttujalle. Harhan suuruus riippuu kyseisten keskiarvojen erotuksesta ja katon suuruudesta ts. siitä kuinka monelta havaintoyksiköltä puuttuu kyseisen tulosmuuttujan havaintoarvo. Toinen näkökulma asiaan on, että HT-estimaattori harhan suuruus riippuu vastaustodennäköisyyden ja tulosmuuttujan välisestä korrelaatiokertoimesta ja vastaustodennäköisyyden sekä tulosmuuttujan variaatiokertoimista. Ja koska variaatiokertoimet ovat positiivisia, riippuu se onko HT-estimaattori harhaton, tulosmuuttujan ja vastaustodennäköisyyden korrelaatiokertoimesta. (Djerf, 2001, 17-21).



Katoa on luonteeltaan kolmea eri tyyppiä: A) Täysin harmiton kato (missing completely at random, MCAR), B) Harmiton kato (missing at random, MAR) ja C) Harmillinen kato (Non-Ignorable nonresponse). Kato on täysin harmitonta, jos vastaustodennäköisyys ei riipu millään tavalla tulosmuuttujasta  $y$ , eikä lisäinformaatiomuuttujista  $x$ . Tämä tarkoittaa sitä, että vastaustodennäköisyyden ja tutkimuksen tulosmuuttujan sekä vastaustodennäköisyyden ja apumuuttujien väliset korrelaatiokertoimet ovat nolliä. Tällöin katoa voidaan pitää täysin satunnaisena toisen vaiheen otantana ja vastaustodennäköisyyttä vakiona kaikille havaintoyksiköille. Harmiton kato kasvattaa estimaattien keskivirheitä, mutta HT-estimaatti totaalille on harhaton. Kaikki analyysit voidaan tehdä tutkimalla vain niitä havaintoyksiköitä, joille ovat tulosmuuttujan havaintoarvot tiedossa eli vastanneiden joukkoa. Tätä sanotaan ns. list wise deletion -menetelmäksi. Tietokoneohjelmat tekevät yleensä sitä. (Lehtonen ja Pahkinen, 2004).

Harmiton kato (MAR) on kyseessä silloin, jos tulosmuuttujan ja vastaustodennäköisyyden välinen korrelaatiokerroin on nolla kun muiden muuttujien vaikutus on vakioitu. Jos kato on harmitonta, niin lisäinformaatiomuuttujien ja vastaustodennäköisyyden välinen korrelaatiokerroin saattaa olla eri kuin nolla. Harmiton kato on kyseessä esim. silloin kun tulotiedossa ilmenevä kato riippuu siviilisäädystä, mutta eri siviilisäädyn luokissa kato on riippumatonta tulotiedosta. Täysin harmiton kato on harmittoman kadon erikoistapaus. (Lehtonen ja Pahkinen, 2004, Djerf, 2001).

Kato on harmillista, jos vastaustodennäköisyys riippuu tulosmuuttujan arvosta, vaikka muiden muuttujien vaikutus olisikin vakioitu. Tällöin vastaustodennäköisyyden ja tulosmuuttujan välinen korrelaatiokerroin ei ole nolla, eivätkä myöskään lisäinformaatiomuuttujien ja vastaustodennäköisyyden väliset korrelaatiokertoimet. (Djerf, 2001, Lehtonen ja Pahkinen, 2004).

On huomattava, että ei voida tietää onko kato täysin harmitonta, harmitonta vai harmillista, koska ei tiedetä tulosmuuttujan havaintoarvoja katojoukossa. Kadon luonnetta voidaan arvioida siis vain epäsuoraan. Yleensä täysin harmiton kato on epärealistinen oletus ja ei voida siksi käyttää suoraan vastanneiden joukosta saatavia tuloksia (ei voida käyttää list wise deletion -menetelmää). Kadon aiheuttamaa virhettä tuleekin oikaista erilaisin menetelmin. Näitä ovat uudelleenpainotus ja imputointi. (Lehtonen ja Pahkinen, 2004, Särndal ym., 1992, Djerf, 2001).

Uudelleenpainotusta käytetään yleensä silloin kun aineistossa on yksikkökatoa ja imputointia kun aineistossa on eräkatoa. Sekä imputointi että uudelleenpainotus tuovat lisäepävarmuutta estimointituloksiin, ja siksi varianssin estimoinnissa on asetelmavarianssin lisäksi laskettava imputoinnista tai uudelleenpainotuksesta johutuva komponentti. Imputoinnista puhutaan enemmän luvussa 5.5. Tutkimuksessa ei käytetä uudelleenpainotusta ja siksi uudelleenpainotusmenetelmien raportointia ei tutkimuksessa ole tehty. Lukija voi halutessaan perehtyä uudelleenpainotusmenetelmiin mm. kirjojen Djerf (2001), Särndal ym. (1992), Tilastokeskus (2002a, luvut 2.8 ja 2.10) sekä Lehtonen ja Pahkinen (2004, Chapter 4) avulla.

## 5.5 Malliluovuttaja -menetelmällä imputoiminen/ennustaminen

**Ennustamisella** tarkoitetaan Gourierouxin ja Monfordin (1995, 361) mukaan sellaisen muuttujan havaintoarvojen arvioimista, jonka kaikki tai osa havaintoarvoista puuttuvat. Arviointi tehdään etsimällä puuttuville havainnoille mahdollisimman hyviä approksimaatiota, jotka ovat havaintojen funktiota. Ennustaminen voi tapahtua menneisyyteen tai tulevaisuuteen. Menneisyyteen ennustamisella tarkoitetaan sitä, että pyritään arvioimaan olemassa olevia muuttuja-arvoja, jotka eivät ole kuitenkaan tiedossa. Imputointikin on menneisyyteen ennustamista. Tulevaisuuteen ennustaminen on taas sellaisten muuttuja-arvojen arviointia, joita ei ole vielä olemassa, vaan jotka esiintyvät tulevaisuudessa. Esimerkiksi yrityksen seuraavan tilikauden tuloksen arviointi viimeisimmän tiedossa olevan liikevoiton/tappion kautta, on tulevaisuuteen ennustamista. Ennustamistehtäviä on taloustieteen alalla paljon. Yleistä on minkä tahansa aikasarjan täydentäminen (mm. kun yhden vuoden havainnot puuttuvat). Tutkimuksessa tehdään menneisyyteen ennustamista, sillä siinä etsitään puuttuville toimipaikkatason tilinpäätöstiedoille, jotka ovat olemassa, mahdollisimman hyviä approksimaatiota.

**Imputointi** tarkoittaa puuttuvan tiedon korvaamista estimoidulla arvolla tai arvoilla. Näitä estimoituja arvoja kutsutaan imputoiduiksi tai estimoiduiksi arvoiksi. Jos puuttuvat tieto korvataan yhdellä estimoidulla arvolla, on kyseessä yksinkertainen imputointi (single imputation), ja jos puuttuvat tiedot korvataan vähintään kahdella estimoidulla arvolla, on kyseessä moni-imputointi (multiple imputation). Imputointi perustuu oletukseen, että kato on täysin harmitonta (MCAR, missing completely at random) tai harmitonta (MAR, missing at random). Kadosta ja sen eri tyypeistä on kerrottu luvussa 5.4. (Lehtonen ja Pahkinen, 2004, Särndal ym., 1992, Tilastokeskus, 2002 a).

Imputoinnin tavoitteena on se, että saadaan käyttöön täydellinen aineisto eli aineisto, josta puuttuvat vain ne havaintoarvot, joiden kuuluisi loogisestikin puuttua. Imputoidut havaintoarvot pyritään muodostamaan siten, että imputoinnin jälkeen aineisto vastaisi mahdollisimman hyvin aineistoa, jossa katoa ei ole. Tämä tarkoittaa sitä, että imputoinnin avulla muodostetusta täydellisestä aineistosta lasketut piste-estimaatit, mutta eivät niiden varianssit, olisivat samat kuin, jos katoa ei olisi (imputointia ei olisi tarvittu). (Lehtonen ja Pahkinen, 2004, Särndal ym., 1992, Tilastokeskus, 2002 a).

Vaikka imputoinnin tuloksena on täydellinen aineisto, on sillä myös haittoja. Haittana on se, että imputoinnin pohjalla olevat mallit perustuvat vain vastaajien tietoihin. Jos vastaajat ja ei-vastanneet poikkeavat toisistaan, voivat imputoidut havaintoarvot olla harhaisia. Lisäksi imputointi perustuu usein mutkikkaisiin malleihin ja otosvariانسien estimointi saattaa olla hankalaa. Imputoinnin tehokkuus riippuu täysin siitä miten hyvä on malli, johon imputoidut havaintoarvot perustuvat. Imputointi vaikeuttaa estimointia, koska imputoituja havaintoja ei saisi käyttää kuin oikeita havaintoja estimoinnissa ja analyyseissa (kuten regressioanalyysi). Imputointi yleensä kasvattaa varianssia. Varianssia ei voikaan laskea imputoinnin läsnäollessa asetelmavarianssilla. Tämä johtuu siitä, että imputoidut havaintoarvot ovat myös estimaatteja ja niillä on oma varianssi. Siksi varianssi koostuu kahdesta osasta: asetelmavarianssista ja imputointivarianssista (imputoitujen havaintoarvojen

oma varianssi). Jos estimoidaan varianssi pelkästään asetelmavarianssilla, niin tuotetut estimaatit varianssille ovat aliestimaatteja. (Särndal ym., 1992, 591, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Imputoinnin tapauksessa estimaattoreiden varianssien estimaattorit ovatkin (imputointimenetelmästä riippumatta) seuraavaa muotoa (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

$$\hat{V}(\hat{t}^*) = \hat{V}_{sam}(\hat{t}^*) + \hat{V}_{IMP}(\hat{t}^*) , \text{ missä} \quad (5.3)$$

$\hat{V}_{sam}(\hat{t}^*) =$  Asetelmavarianssi ja  $\hat{t}^* =$  käytetyn strategian mukainen asetelmaperusteinen totaalin estimaattori (esimerkiksi GREG -estimaattori).

Esimerkiksi yksinkertaisen satunnaisotannan tapauksessa HT-estimaattorin asetelmavarianssi on (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004) seuraava.

$$\hat{V}(\hat{t}_{ydp}) = \sum_{h=1}^H \hat{V}(\hat{t}_{ydp}) = N_d^2 \left( \frac{1}{n_d} - \frac{1}{N_d} \right) \frac{1}{n_d - 1} \sum_{k \in S_d} (y_k - \bar{y}_d)^2 , \text{ missä} \quad (5.4)$$

$N_d =$  kiinnostuksen kohteena olevan osajoukon  $d = 1, \dots, D$  koko perusjoukossa  $n_d$  tämän kiinnostuksen kohteena olevan osajoukon koko otoksessa ja  $\bar{y}_d = \frac{1}{n_d} \sum_{k \in S_d} y_k$ ,

ja approksimaatio suoran GREG -estimaatin asetelmavarianssille on:  $\hat{V}(\hat{t}_{ydpGREG}) = (1 - R^2_d) \hat{V}(\hat{t}_{ydp})$ . (5.5)

Imputointivarianssia laskennassa käytetään vain vastanneiden joukon havaintoyksiköiden tietoja. Se lasketaan yksinkertaisen imputoinnin tilanteessa kun on käytetty yksinkertaista satunnaisotantaa seuraavasti. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004).

$$\hat{V}_{imp}(\hat{t}_{HT}) = N^2 \left( 1 - \frac{n_{(r)}}{n} \right) * \frac{\sum_{k=1}^{n_{(r)}} (\hat{e}_k - \bar{\hat{e}}_k)^2}{n_{(r)} - 1} / n_{(r)} , \quad (5.6)$$

missä  $\hat{e}_k = y_k - \hat{y}_k =$  todellisen arvon ja imputoidun havaintoarvon erotus (ns. ennustevirhe tai mallin jäännöstermi) ja riippuu siis käytettävästä imputointimenetelmästä, ( $\hat{e}_k$  on esim. regressioimputoinnin tapauksessa

$y_k - x'_k \hat{\mathbf{b}}$  ja keskiarvoimputoinnin tapauksessa  $y_k - \bar{y}$ ),  $\bar{\hat{e}}_k = \frac{\sum_{k=1}^{n_{(r)}} (\hat{e}_k)}{n_{(r)}}$  = jäännöstermien (ennustevir-

heiden) keskiarvo ja  $n_{(r)} =$  niiden havaintojen lukumäärä, jolle tulomuuttujan havaintoarvot ovat tiedossa (eli vastanneiden joukon otoskoko). Imputointivarianssin neliöjuurta kutsutaan imputointikeskivirheeksi. Imputointikeskivirheitä on tutkimuksessa laskettu ja ne on raportoitu luvussa 6.2.

Jotta imputointivarianssi voitaisiin laskea, tulisi olla tiedossa se mitkä havainnot ovat imputoituja ja miten imputointi on tehty (käytetty imputointimenetelmä). Tämä johtuu siitä, että imputointivarianssin laskemista varten tulisi tietää mitkä ovat imputoidut havaintoarvot vastanneiden joukossa ja voida laskea imputoitujen havaintoarvojen ja todellisten arvojen erotukset eli mallin jäännöstermit vastanneiden joukossa. Imputointivari-

anssin laskeminen vaatii siis tarkkaa tietoa käytetystä imputointimenetelmästä ja siitä mitkä havainnot ovat imputoituja. (Särndal ym., 1992, 591, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Yksinkertaisen imputoinnin menetelmiä ovat esimerkiksi looginen imputointi, keskiarvoimputointi (RM, mean imputation), hot deck -menetelmä (HD), lähimmän naapurin menetelmä (NN, nearest neighbour) ja erilaiset malliluovuttaja-menetelmään perustuvat imputointitavat, kuten suhde-estimointi (RA, ratio imputation). Koska moni-imputointia ei tutkimuksessa käytetä, ei sen menetelmiä raportoida tutkimuksessa. Lukija voi halutessaan tutustua moni-imputointiin mm. kirjan Lehtonen ja Pahkinen (2004) avulla. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Loogisessa imputoinnissa puuttuva havainto korvataan havaintoarvolla, joka koetaan järkeväksi (ns. cold deck -menetelmä, jota käytetään mm. kun toimipaikan edellisen tilikauden tietoa käytetään imputoinnissa). Hot deck -menetelmällä imputointi tarkoittaa sitä, että puuttuva havainto korvataan havaintoarvolla, joka arvotaan niiden havaintoyksiköiden joukosta, joilla on samanlaisia piirteitä kuin sillä havaintoyksiköllä, jonka havaintoarvoa imputoidaan (otos tulee siis olla jaettavissa jonkun apumuuttujan suhteen sisäisesti homogeenisiin ositteisiin). (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Keskiarvoimputoinnissa puuttuvat havainnot korvataan otoksen keskiarvolla (käyttöä ei suositella, jos lisäinformaatiota on saatavilla) tai otoksen osajoukon (ositteen) keskiarvolla (otos tulee voida jakaa jonkun apumuuttujan suhteen sisäisesti homogeenisiin ositteisiin). Lähimmän naapurin menetelmällä imputoinnissa ja malliluovuttaja-imputointimenetelmää käytettäessä tulee olla tiedossa jonkun/joidenkin jatkuvan lisäinformaatiomuuttujan/muuttujien havaintoarvot kaikilta otosalkiolta. Lähimmän naapurin menetelmässä puuttuva havaintoarvo korvataan sellaisen vastanneen havaintoyksikön arvolla, jolle puuttuvaa havaintoa vastaavan havaintoyksikön lisäinformaatiomuuttujan ja vastanneen havaintoyksikön lisäinformaatiomuuttujan havaintoarvojen erotus on itseisarvoltaan pienin mahdollinen. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Tutkimuksessa muodostettavia toimipaikkojen jalostusarvojen ennusteita/estimaatteja voidaan kutsua **malliluovuttaja-menetelmällä (model donor imputation)** imputoiduiksi toimipaikkojen jalostusarvoiksi. Malliluovuttaja-menetelmä on Tilastokeskuksessa käytettävä suomennos. Puhutaan myös ns. regressioperusteisesta imputoinnista. Se tarkoittaa puuttuvan (tai tiedossa olevan) tiedon korvaamista tilastollisten mallien avulla saadulla estimaatilla. Malliluovuttaja-menetelmässä muodostetaan aluksi siis tilastollinen malli ja estimoidaan tämän mallin parametrit. Käytettävän mallin parametrien estimaatit muodostetaan vain vastanneiden ryhmän havaintoarvojen perusteella (mallin parametrit estimoidaan käyttämällä laskennassa vain niitä havaintoyksiköitä, joiden tulosmuuttujien havaintoarvot ovat tiedossa). Imputointi tehdään kertomalla laadittujen mallien parametrien estimaatit niitä vastaavien apumuuttujien arvoilla. Malliluovuttaja-imputoinnissa tarvitaan siis aina lisäinformaatiota (apumuuttujia, joiden havaintoarvot ovat tiedossa myös niille havaintoyksiköille, joiden tulosmuuttujien havaintoarvot eivät ole tiedossa). Imputoinnissa käytettävä malli riippuu käytettävien apumuuttujien luonteesta. Suhde-estimoinnissa käytetään mallina lineaarista mallia ilman vakiotermiä. Yleistettyihin regressioestimaattoreihin perustuvassa imputoinnissa käytetään esimerkiksi lineaarista mallia vakion

kanssa. Tässä tutkimuksessa käytetään yleistettyihin regressioestimaattoreihin perustuvaa imputointia yritysten tilinpäätöstietoja toimipaikkaistettaessa. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002a, 62).

Tutkimuksessa mallina/luovuttajana käytetään lineaarista mallia, jossa  $\beta$ -parametrien estimaatit ovat suoran tai epäsuoran synteettisen estimaattorin mukaan laskettuja. Malliluovuttaja-menetelmän rajoitteena on se, että sen avulla voidaan saada imputoituja arvoja, jotka ovat tulosuuttujan arvoalueen ulkopuolella, ja että se perustuu vahvaan oletukseen. Oletus on, että niillä havaintoyksiköillä, joille ei tulosuuttujan havaintoarvoa ole tiedossa (ns. katoryhmässä), ovat tulosuuttujan ja apumuuttujan väliset yhteydet samanlaisia kuin vastanneiden ryhmässä, eli voidaan käyttää samoja parametriestimaatteja sekä katoryhmässä että vastanneiden ryhmässä. Jos tämä oletus ei päde, voivat imputoidut havaintoarvot olla systemaattisesti harhaisia. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Imputointi on sitä tehokkaampaa mitä parempaa lisäinformaatiota on käytettävistä otoksesta ja/tai perusjoukosta. Lisäinformaatiota tehokkaammin käyttävät menetelmät ovat parempia kuin menetelmät, joissa ei lisäinformaatiota käytetä. Siksi voidaan malliluovuttaja-menetelmään perustuvia imputointimenetelmiä yleensä pitää muita parempana, jos käytössä on hyviä apumuuttujia ja niiden totaalit ovat tiedossa otoksen tai perusjoukon tasolla. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Koska voidaan olettaa, että yritystiedot ovat toimipaikkojen tietojen summia (tällöin ei yritystiedoissa ole otettu huomioon yrityksen ulkomaista toimintaa eikä poistoja), **tulee estimoidut havaintoarvot tai malliluovuttaja-menetelmällä imputoidut havaintoarvot kalibroida (normeerata)** siten, että yrityksen toimipaikkojen laskennallisen käyttökateen tietojen estimaattien summaksi tulee toimipaikat omistavan yrityksen tieto (ts. toimipaikkatietojen tilinpäätöstietojen estimaattien summa = yritystieto). Tämä tarkoittaa sitä, että estimoidut toimipaikkojen jalostusarvot kalibroidaan. Kalibroiminen tarkoittaa sitä, että estimaatit muunnetaan vastaamaan perusjoukon reunajakaumia. Tämä tapahtuu painottamalla  $\beta$ -parametrien estimaattien laskemisessa havaintoarvoja perusjoukon reunajakauman avulla saaduilla luvuilla, painoina ei käytetä sisällysmistodennäköisyyden käänteislukuja suoraan. (Särndal ym., 1992, Lehtonen ja Pahkinen, 2004, Tilastokeskus, 2002 a).

Tutkimuksessa kalibrointi tarkoittaa sitä, että toimipaikkojen jalostusarvojen estimaatit saatetaan vastaamaan yritystietoja eli yritysten jalostusarvoja. Kalibroinnista johtuen yksitoimipaikkaisten liikeyritysten jalostusarvojen estimaatit ovat yhtä kuin kyseisen yrityksen jalostusarvo (ne ovat todellisia arvoja). Imputoinnin pohjalla olevien mallin  $\beta$ -parametrien estimaatteja ei kalibroida, koska vain siten voidaan tutkia miten hyvin eri apumuuttujat selittävät toimipaikan jalostusarvoa yksitoimipaikkaisilla liikeyrityksillä.

Kalibroiminen tapahtuu seuraavasti. Oletetaan, että yrityksen  $i$  toimipaikkojen laskennallisen käyttökateen tiedon  $j$  estimaattien summa on  $C$  ja yrityksen  $i$  tilinpäätöstieto  $j$  = yrityksen toimipaikkojen laskennallisen käyttökateen tietojen  $j$  summa todellisuudessa =  $D$ . Yksittäisen toimipaikkojen tietoja ei siis tiedetä, mutta kuitenkin niiden summa, jota voidaan käyttää lisäinformaationa imputoinnissa. Tällöin kalibroinnin tavoitteena

on saada yrityksen  $i$  toimipaikkojen laskennallisen käyttökateen tiedon  $j$  estimaattien summa  $= D$ . Tämä tehdään kertomalla jokainen yrityksen  $i$  toimipaikan laskennallisen käyttökateen tiedon  $j$  estimaatti  $D$ :llä ja jakamalla  $C$ :llä. Kun kaikki estimaatit on kalibroitu, saadaan toimipaikkojen laskennallisen käyttökateen tiedon  $j$  summaksi yritysten tilinpäätöstiedon  $j$  summa (esim. Suomen yritysten jalostusarvo). Esim. jalostusarvon estimaattien kohdalla kalibroiminen suoritetaan kertomalla toimipaikan jalostusarvon estimaatti toimipaikkaa vastaavan yrityksen jalostusarvolla ja jakamalla niiden toimipaikkojen jalostusarvojen estimaattien summalla, joiden omistaja sama.

**Luvussa 5.3 on kerrottu siitä miten tutkimuksessa estimoitiiin toimialakohtaisia tilinpäätöstietoja (toimipaikka tulosmuuttujille  $y$  osajoukossa  $d = 1, \dots, D$ ). Näiden perusteella saadaan toimipaikkatasoisille tilinpäätöstiedoille  $n_s$  estimoituja arvoja (Särndal, 2001, 40) eli predicted  $y$ -values. Nämä laaditut estimaatit sitten lopuksi kalibroidaan ja kalibroitujen estimaattien hyvyttä tutkitaan diagnostisten tarkastelujen avulla (luku 5.6).**

**KALIBROITAVAT ENNUSTEET/ESTIMAATIT (mallien avulla laskettavat estimaatit toimipaikkojen tilinpäätöstiedoille):**

1. Suorien yleistettyjen regressioestimaattoreiden ja synteettisten estimaattoreiden pohjalta laadittavat estimaatit:

$$\hat{y}_k = \mathbf{x}_k' \hat{\boldsymbol{\beta}}_d, \text{ missä } \hat{\boldsymbol{\beta}}_d = \left( \sum_{s_d} a_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_{s_d} a_k \mathbf{x}_k y_k.$$

Taustalla on lineaarinen regressiomalli muotoa  $y_k = \mathbf{x}_k' \boldsymbol{\beta}_d + \mathbf{e}_k$ , missä  $k$  viittaa toimipaikkaan ja  $d = 1, \dots, D$ .

2. Epäsuorien synteettisten estimaattorien ja yleistettyjen regressioestimaattoreiden pohjalta laadittavat estimaatit:

$$\hat{y}_k = \mathbf{x}_k' \hat{\boldsymbol{\beta}}, \text{ missä } \hat{\boldsymbol{\beta}} = \left( \sum_s a_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_s a_k \mathbf{x}_k y_k$$

Taustalla on lineaarinen regressiomalli muotoa  $y_k = \mathbf{x}_k' \boldsymbol{\beta} + \mathbf{e}_k$ , missä  $k$  viittaa toimipaikkaan ja  $d = 1, \dots, D$ .

## 5.6 Diagnostiset tarkastelutavat

Se miten hyviä mallien avulla laskettavat estimaatit toimipaikkojen tilinpäätöstiedoille (eli imputoidut havaintoarvot) ovat, riippuu siitä miten vahva riippuvuus tulosmuuttujien ja apumuuttujien välillä on. Tätä voidaan arvioida tutkimalla tulosmuuttujan ja apumuuttujan välisiä Pearsonin korrelaatiokertoimia. Pearsonin korrelaatiokertoimen käyttämisen syy on se, että korrelaatiokertoimien laskeminen mallin riippuvien ja riippumattomien muuttujien välille on hyödyllistä regressiota tulkittaessa. Tämä johtuu siitä, että yhden riippumattoman muuttujan mallin tapauksessa  $\hat{\mathbf{b}} = r \frac{s_y}{s_x}$ , missä  $r$  on  $x$ :n ja  $y$ :n välinen Pearsonin otoskorrelaatiokerroin.

(Wonnacott ym., 1979, 152).

*Imputoinnin pohjalta olevien mallien valintakriteereitä ovat seuraavat (J kappaletta).*

A) Mallin riippumattomien muuttujien (eli muuttujien  $x$ ) tulisi olla toisistaan likimain riippumattomia (käytettävien apumuuttujien välillä ei saa olla suurta korrelaatiota). Tämä tarkoittaa sitä, että ei ole multikollinearisuutta.

den ongelmaa. Jokaisella riippumattomalla muuttujalla tulee myös olla tilastollisesti perusteltu tulkinta (mm. jokaisen riippumattoman muuttujan  $x$  tulee korreloida mahdollisimman voimakkaasti tulosmuuttujan  $y$  kanssa).

B) Mallin tulee olla mahdollisimman yksinkertainen, eli riippumattomia muuttujia tulee olla mahdollisimman vähän. Tämä tarkoittaa sitä, että malliin lisätään uusi muuttuja vain, jos sen lisääminen pienentää keskivirhetä ja kasvattaa vapausastekorjattua selitysstetta (adjusted coefficient of determination). Vapausastekorjattua selitysstetta käytetään, koska tavallinen selitysstete (coefficient of determination) kasvaa riippumattomien muuttujien lukumäärän kasvaessa, mutta vapausastekorjattu selitysstete kasvaa vain, jos uusi riippumaton muuttuja parantaa mallia. Täten vapausastekorjattu selitysstete ottaa huomioon riippumattomien muuttujien lukumäärän, mutta tavallinen selitysstete ei.

C) Mallin parametriestimaattien tulee olla tilastollisesti merkitseviä (vakion estimaatti voi olla tai ei olla merkitsevä).

E) Tulosmuuttujan tuntemattoman parametrin arvon on mahdollista mallin avulla lasketulle estimaatin luottamusvälille.

F) Mallin vapausastekorjatun selitysstteen on oltava hyvä.

G) Harhankorjaustermin eli synteettisen estimaatin ja GREG-estimaatin erotuksen tulee olla mahdollisimman pieni.

H) GREG-estimaatin keskivirheen ja GREG-estimaatin sekä todellisen arvon erotuksen itseisarvon tulee olla mahdollisimman pieni.

I) Toimiala- ja yrityksen lajikohtainen malli on parempi kuin toimialakohtainen malli vain, jos toimiala- ja yrityksen lajikohtaisien mallien avulla laskettu toimialakohtainen estimaatti on lähempänä todellista arvoaan kuin toimialakohtaisella mallilla laskettu estimaatti.

J) Mallia käyttämällä voidaan laskea ns. ennustevirheitä tai mallin jäännöstermejä ja tutkia niiden avulla mallin järkevyyttä ja riittävyyttä. Jäännöstermeillä tarkoitetaan todellisen arvon ja mallilla laaditun estimaatin eroja  $e_k = y_k - \hat{y}_k$ . Ne kuvaavat mallin ns. satunnaista osaa eli sitä osaa havaintoarvojen vaihtelusta, jota ei malli pysty selittämään. Se pitää sisällään mm. malliin kuulumattomien muuttujien vaikutuksen, mittaus- ja prosessointivirheet sekä satunnaisvaihtelun. Termi  $\hat{y}_k$  (eli havaintoyksikön  $k$  muuttujan  $y$  havaintoarvon estimaatti) kuvaa mallin determinististä osaa, eli sitä osaa havaintoarvojen vaihtelusta, jonka malli pystyy selittämään. Jotta valittu malli olisi tilastollisesti perusteltu, tulee jäännöstermien olla normaalisti jakautuneita ja toisistaan riippumattomia (tätä arvioidaan mm. normaalisuustestauksien ja korrelaatiokertoimien tutkimisen avulla). Kun jäännöstermeistä piirretään hajontakuviota, ei siinä tulisi olla säännönmukaisuuksia. Tämä tarkoittaa sitä, että pistejoukko sijaitsee tasaisesti nollan molemmiin puoliin ja on muutenkin satunnainen. Jos hajontakuviossa on säännönmukaisuutta, osoittaa se, että malli ei ole ehkä tilastollisesti perusteltu. (Holopainen, Pulkkinen, 1995, 186-194)

Koska kyse on yksinkertaisesta imputoinnista, riippuu toimipaikkaistamisen onnistuminen valittujen mallien "hyvyyden" lisäksi siitä miten ne toimipaikat, joiden jalostusarvo ei ole tiedossa, poikkeavat niistä, joiden jalostusarvo on tiedossa. Jos voidaan olettaa, että kummassakin joukossa jalostusarvon ja apumuuttujien väliset riippuvuussuhteet ovat samanlaisia (ts. parametrien estimaatit ovat samoja), niin voidaan olettaa impu-

toinnin onnistuneen, jos valittu malli antaa hyvät estimaatit siinä joukossa, joiden jalostusarvo on tiedossa. Tämä tarkoittaa sitä, että aineistossa oleva kato on luonteeltaan harmitonta (MAR) tai täysin harmitonta (MCAR, luku 5.4). Tätä ei voida kuitenkaan arvioida kuin epäsuoraan.

Tutkimuksen kato on luonteeltaan sekä toimialasta että toimipaikkaa vastaavan yrityksen lajista riippuvaa, koska tiedossa ovat kaikkien yksitoimipaikkaisten yritysten toimipaikkojen tiedot ja monitoimipaikkaisten yritysten toimipaikkojen tiedot kun kyseessä on teollisuuden tai rakentamisen toimialalla toimiva toimipaikka. Eli ne toimipaikat, joiden kaikki laskennallisen käyttökäteen tiedot eivät ole tiedossa, ovat suurimmalta osaltaan monitoimipaikkaisten yritysten toimialojen A, B, G-O toimipaikkojen tietoja. Tämä antaa aiheen olettaa ettei tutkimuksessa esiintyvä kato ole täysin harmitonta. Luvussa 4.1 nimittäin todettiin, että kaikilla toimialoilla monitoimipaikkaisten yritysten toimipaikat ovat keskimäärin suurempia tilinpäätöstiedoiltään kuin yksitoimipaikkaisten yritysten toimipaikat mediaanin, keskiarvon ja moodin avulla mitattuna. Voidaan siis ajatella, että myös niillä toimialoilla, joiden jalostusarvot eivät ole tiedossa, ovat jalostusarvot suurempia monitoimipaikkaisten yritysten toimipaikoilla kuin yksitoimipaikkaisten yritysten toimipaikoilla. Tutkimuksessa voidaankin tutkia mallien toimivuutta monitoimipaikkaisten yritysten kohdalla vain teollisuuden ja rakentamisen toimialoilla, mutta yksitoimipaikkaisten yritysten kohdalla kaikilla toimialoilla. Jos toimialakohtaiset mallit toimivat noissa osajoukoissa, antaa se uskoa mallien toimivuudesta myös niissä osajoukoissa, joista ei ole tietoa (monitoimipaikkaisten yritysten toimialojen A, B ja G-O toimipaikat). Mallien ja myös imputoinnin toimivuutta tutkitaan jäännöstermien ja imputointivarianssien avulla, kuten on tehty luvussa 6.2.

## 6 Toimialakohtaiset tilinpäätöstiedot

Tutkimuksessa laadittiin erilaisia toimialakohtaisia estimaatteja jalostusarvolle käyttämällä luvussa 5.3 raportoituja estimaattoreita (epäsuorat ja suorat Horvitz-Thompson-, GREG- ja synteettiset estimaattorit), luvussa 5.2 esitetyjä otanta-asetelmia ja lisäinformaatiota (luvut 3 ja 5.1). Luvussa 6.1 tutkitaan erityisesti sitä miten toimiala ja toimipaikkaa vastaavan yrityksen laji vaikuttavat siihen miten erilaiset estimaattorit ja apumuuttujat toimivat toimialakohtaisen jalostusarvon totaalien estimoinnissa. Vertaillaan siis estimaattorien toimivuuden riippuvuutta käytetystä lisäinformaatiosta ja kiinnostuksen kohteena olevasta joukosta niiden toimipaikkojen joukossa, joiden jalostusarvo on tiedossa. Lisäksi tutkitaan synteettisten- ja GREG-estimaattien välisiä eroja eli harhankorjaustermien suuruutta ja riippuvuutta käytettävästä mallista ja osajoukosta. Tutkitaan myös sitä parantaako lisäinformaation käyttö estimointia vertaamalla suoraa ja epäsuoraa GREG- ja synteettisiä estimaatteja suoriin ja epäsuoriin Horvitz-Thompson estimaatteihin. Tarkastelun perusteella valitaan luvun 5.6 kriteerein toimipaikkojen tilinpäätöstietojen estimoinnissa käytettävät mallit ja mallien parametriestimaatit. Imputoinnissa käytetyt mallit ja niiden pohjalta lasketut estimaatit jalostusarvojen toimialakohtaisille totaaleille raportoidaan luvuissa 6.2. Luvussa 6.3 arvioidaan toimialojen merkittävyyttä henkilöstön määrän, liikevaihdon ja palkkojen avulla. Tutkitaan myös toimialojen merkittävyyttä yritystietojen ja jalostusarvon arvioiden avulla. Tutkimuksessa tuotettujen toimipaikkojen jalostusarvon estimaattien laatua on kuvattu luvuissa 3, 5 ja 8.



## 6.1 Mallin ja estimaattorin rooli sekä estimaattien vertailu

Tutkimuksessa vertaillaan luvussa 5.3 raportoitujen kuuden eri estimointistrategian toimivuutta. Nämä strategiat ovat tiivistettynä taulukossa 6.1.

**Taulukko 6.1.** Tutkimuksessa käytetyt estimointistrategiat

| <i>ESTIMAATTORIN TYYPPI<br/>+ OTANTA-ASETELMA</i> | <i>TOTAALIN ESTIMAATTORIN<br/>LAJI</i> | <i>ESTIMAATTORI</i>       | <i>STRATEGIAN<br/>NUMERO JA NIMI</i> |
|---|--|---------------------------|--------------------------------------|
| <b>SUORA + SRSSTR</b>                             | ASETELMAPERUSTEINEN                    | HT-ESTIMAATTORI           | 1 HT-S                               |
|   | MALLIAVUSTEINEN                        | GREG-ESTIMAATTORI         | 2 GREG-S                             |
|   | MALLIPERUSTEINEN                       | SYNTEETTINEN ESTIMAATTORI | 3 SYN -S                             |
| <b>EPÄSUORA + SRSSTR</b>                          | ASETELMAPERUSTEINEN                    | HT-ESTIMAATTORI           | 4 HT-ES                              |
|   | MALLIAVUSTEINEN                        | GREG-ESTIMAATTORI         | 5 GREG-ES                            |
|   | MALLIPERUSTEINEN                       | SYNTEETTINEN ESTIMAATTORI | 6 SYN-ES                             |

Luvun 5.3 ja taulukon 6.1. mukaan kaikki 6 estimointistrategiaa käyttävät ositettua otantaa suhteellisella kiintiöinnillä ja poikkeavat toisistaan vain käytettävän estimaattorin suhteen. Seuraavassa eri strategioissa käytetyt estimaattorit. 1. Estimointistrategia: HT-S eli suora Horvitz-Thompson estimaattori, 2. Estimointistrategia: GREG-S eli suora GREG -estimaattori, 3. Estimointistrategia: SYN-S eli suora synteettinen estimaattori, 4. Estimointistrategia: HT-ES eli epäsuora Horvitz-Thompson estimaattori, 5. Estimointistrategia: GREG-ES eli epäsuora GREG -estimaattori ja 6. Estimointistrategia: SYN-ES eli epäsuora synteettinen estimaattori. Tutkimuksessa verrataankin näitä estimointistrategioita toisiinsa toimialakohtaisten jalostusarvon totaalien estimoinnin tilanteessa.

Eri estimointistrategioiden lisäksi tutkimuksessa tutkitaan lisäinformaation käytön merkitystä tutkimalla otannan ja estimoinnin vaiheissa käytetyn lisäinformaation merkitystä toimialakohtaisten jalostusarvon kokonaismäärien estimaattien kannalta. Lisäinformaation käytöstä otannan ja estimoinnin vaiheissa on kerrottu luvussa 5.1. Estimoinnin vaiheessa käytettävissä oleva lisäinformaatio koostuu toimipaikan liikevaihdosta (LVMK), toimipaikan keskimääräisestä henkilöstön määrästä (HK2) ja toimipaikalle estimoiduista palkoista (PALKKA-MK). Otannan vaiheessa käytettävissä oleva lisäinformaatio koostuu toimipaikan sijaintiläänistä, toimipaikan toimialan pääluokasta (käyttäen toimialaluokitusta 1995 eli TOL 95) ja toimipaikkaa vastaavan yrityksen lajista (se onko toimipaikkaa vastaava yritys yksi- vai monitoimipaikkainen). Lisäinformaatio on suurimmaksi osaksi Tilastokeskuksen yritysten suhdanteet -yksikön toimipaikkarekisteristä saatavaa tietoa. Käytettävien muuttujien määritelmät ja tietolähteet ovat luvussa 3.

Liitteessä 8 ovat suorat Horvitz-Thompson estimaatit toimialakohtaisille jalostusarvon totaaleille. Lisäksi liitteessä 8 on HT -estimaatteja, jotka ovat jalostusarvon totaalin estimaatteja erilaisissa toimialakohtaisissa osajoukoissa (esim. jalostusarvon totaalin HT -estimaatti monitoimipaikkaisten yritysten teollisuuden toimialan toimipaikoille). Liite 8 sisältää myös imputointiin valitut toimialakohtaiset mallit sekä niiden ominaisuuksia (mm. GREG-estimaatti toimialakohtaiselle tai toimiala- ja yrityksen lajikohtaiselle jalostusarvon totaalille). Liitteessä 8 raportoitujen tulosten mukaan lisäinformaation käyttö sekä otannan että estimoinnin vaiheessa tehostaa

estimointia. Ositus parantaa estimointia, koska ositetun otannan mukaisen Horvitz-Thompson estimaatin asetelmavarianssi on kaikilla toimialoilla pienempi kuin yksinkertaisen satunnaisotannan mukaisen Horvitz-Thompson estimaatin asetelmavarianssi, eli asetelmakerroin  $deff$  on alle 1. Kaikilla toimialoilla lisäinformaation käyttö estimointivaiheessa parantaa estimointia, sillä sekä suorien että epäsuorien Horvitz-Thompson estimaattien keskivirheet ovat suurempia kuin suorien ja epäsuorien GREG-estimaattien keskivirheet, eli asetelmakerroin  $deff$  on alle 1. Tämä antaa aiheen olettaa, että käytetty lisäinformaatio on tehokasta.

Käytetyn lisäinformaation tehokkuus johtuu siitä, että käytetyt jatkuvat apumuuttajat (LVMK, HK2 ja PALKKAMK) korreloivat vahvasti tulosmuuttujan (eli jalostusarvon) kanssa. Lisäksi jalostusarvo vaihtelee toimialan ja toimipaikkaa vastaavan lajin mukaan. Se on yksi syy osittamisen tehostavaan vaikutukseen. Kaikilla toimialoilla kaikki apumuuttajat korreloivat keskenään vahvasti (luku 4.2). Jotta voidaan välttää vakava multikollineaarisuus, niin yritysten jalostusarvon toimipaikkaistamisessa käytetään yhden apumuuttujan mallia, eli valitaan apumuuttujaksi jokin kolmesta valittavissa olevista apumuuttujista eli HK2, LVMK ja PALKKAMK.

Liite 8 sisältää myös raportin epäsuorien (käytetään kaikille toimialoille yhteistä mallia) GREG- ja synteettisten estimaattien ominaisuuksista. Pienimmän GREG-estimaatin yhden apumuuttujan malleista (eli malleista, joissa jalostusarvoa selitetään joko HK2:lla, LVMK:lla tai PALKKAMK:lla) keskivirheen antaa malli, jossa jalostusarvoa selitetään liikevaihdolla ja liitteessä 8 käsitelläänkin sitä. Kahden muuttujan malleista malli, jossa jalostusarvoa selitettiin liikevaihdolla ja palkoilla antoi paremmat tulokset, mutta tämä saattoi johtua multikollineaarisuudesta, koska perusjoukossa liikevaihto ja palkat korreloivat vahvasti. Kun verrataan GREG-estimaatteja suoriin ja epäsuoriin Horvitz-Thompson estimaatteihin, huomataan, että sekä suorat että epäsuorat GREG-estimaatit (estimointistrategiat 2 ja 5) ovat keskivirheen mielessä parempia kuin estimointistrategiat 1 ja 4 eli suorat ja epäsuorat Horvitz-Thompson estimaatit. Lisäksi sekä suorat että epäsuorat GREG-estimaatit ja synteettiset estimaatit poikkeavat vähemmän todellisista arvoista kuin suorat tai epäsuorat Horvitz-Thompson estimaatit.

Suurimmat keskivirheet ja poikkeamat todellisista arvoista antaa estimointistrategia 4 (eli HT-ES), eli voidaan sanoa, että epäsuorat HT-estimaatit toimivat huonoiten (ja siksi sen tuloksia ei ole raportoitu). Strategian 4 suuret keskivirheet johtuvat varmaan siitä, että tulosmuuttujan jakauma on vino ja toimialoittaiset jalostusarvojen keskiarvot poikkeavat paljon toisistaan. Lisäinformaationa käytetään vain toimialojen kokoa perusjoukossa. Estimointistrategiassa 1 eli suorassa HT-estimaattorissa käytetään hyväksi tietoa toimipaikan pääluokasta käyttämällä toimialoittaisia jalostusarvojen keskiarvoja otoksissa sekä toimialojen kokoa perusjoukossa. HT-S käyttää siis enemmän lisäinformaatiota kuin HT-ES.

Taulukossa 6.2 esitellään suorien GREG- ja synteettisten estimaattoreiden ominaisuuksia, eli tutkitaan mallin vaikutusta suoriin synteettisiin ja GREG-estimaattoreihin. Taulukon 6.2 mukaan mallin valintaan vaikuttaa joillakin toimialoilla se mikä valitaan valintakriteeriksi ja eri valintakriteeriä käyttämällä päädyttiin eri malliin kuin toista käyttämällä. Valintakriteereinä olivat harhankorjaustermi, (eli synteettisen estimaatin ja GREG-estimaatin erotus), joka on yhtä kuin  $\sum_{s_d} a_k (y_k - \hat{y}_k)$ , GREG-estimaatin variaatiokerroin ja keskineliövirhe

(MSE, Mean Squared Error), mallin selitysaste sekä todellisen arvon ja GREG-estimaatin erotus. Todellisen arvon ja GREG-estimaatin erotus ei kuitenkaan kerro todellisesta harhasta mitään, eikä harhankorjaustermi GREG-estimaatin todellisesta harhankorjaustermistä mitään, koska käytettävissä on vain yksi otos ja ei voida siksi tietää GREG-estimaattien odotusarvoja ja harhankorjaustermien odotusarvoja. Jotta voisimme sanoa estimaattien harhasta jotain, niin pitäisi poimia useampi otos (esim. 1000 otosta) samasta perusjoukosta samalla otanta-asetelmalla, eli suorittaa Monte Carlo -simulointikoe. Monte Carlo -koetta ei tehty, koska GREG-estimaatin odotusarvo ja harhankorjaustermien odotusarvo voi olla eri vastanneiden joukossa kuin katojoukossa (luku 5.4) ja siksi ei Monte Carlo -kokeella saataisi tietoa estimaattien harhasta perusjoukossa, eli joukossa, jonka jalostusarvot toimipaikkaistetaan.

Taulukossa 6.2 on mainittu mallit, joiden parametrien estimaatit ovat merkitseviä (poislukien vakio). Näistä malleista tutkitaan erikseen yhden, kahden ja kolmen apumuuttujan mallien ominaisuuksia (harhankorjaustermien itseisarvo, tod.arvon ja GREG-estimaatin erotus, GREG-estimaatin variaatiokerroin ja mallin vapausastekorjattu selityaste). Esimerkiksi yhden apumuuttujan mallilla tarkoitetaan mallia, jossa tulosmuuttujaa (esim. jalostusarvoa) selitetään mallilla, jossa on yksi apumuuttuja (tutkimuksessa joko HK2, LVMK tai PALK-KAMK) ja kahden muuttujan mallilla taas mallia, jossa tulosmuuttujaa selitetään mallilla, jossa on kaksi apumuuttujaa (kuten malli, jossa sekä HK2 että LVMK, eli HK2+ LVMK). Tarkastelun pohjalta on valittu multikollineaarisuuden vuoksi ne yhden apumuuttujan mallit, joiden parametriestimaatit ovat merkitseviä, GREG-estimaatin keskivirhe ja RMSE (keskineliövirheen neliöjuuri, Root Mean Squared Error) ovat pienimpiä mahdollisia ja vapausastekorjattu selitysaste suurin mahdollinen (taulukko 6.3). Kriteerinä pidetään siis mahdollisimman kapeaa luottamusväliä GREG-estimaatille, eikä harhankorjaustermiä, todellisen arvon sekä GREG-estimaatin erotusta tai GREG-estimaatin variaatiokerrointa.

On huomattavissa taulukosta 6.2, että vain toimialoilla C, F ja H ovat kolmen apumuuttujan mallit sellaisia, joissa on merkitsevät parametriestimaatit. Näilläkin toimialoilla kolmen apumuuttujan mallit ovat sellaisia, että niillä on suurempi harhankorjaustermien itseisarvo ja todellisen arvon sekä GREG-estimaatin erotus kuin yhden tai kahden apumuuttujan malleilla. Vähintään yhdellä kahden apumuuttujan mallilla on merkitsevät parametriestimaatit kaikilla toimialoilla lukuun ottamatta toimialoja E, I ja X. Mutta monella toimialalla kahden apumuuttujan malleilla on suurempi harhankorjaustermien itseisarvo ja todellisen arvon sekä GREG-estimaatin erotus kuin yhden apumuuttujan malleilla.

Taulukossa 6.3 ovat keskivirheen, RMSE:n ja selitysasteen mielessä parhaiten toimivat suorat GREG-estimaatit ja suorat synteettiset estimaatit, joiden parametriestimaatit (poislukien vakio) ovat merkitseviä, sekä näiden ominaisuuksia. Taulukossa 6.3 mainittuja malleja käytettiin yritysten jalostusarvojen toimipaikkaistamiseen eli toimipaikkojen jalostusarvojen imputoimiseen (luvut 5.4, 5.5 ja 6.2).

Taulukko 6.2. Suorien GREG- ja synteettisten estimaattien ominaisuuksia.

| <b>TOIMIALA (tai kiinnostuksen kohteena oleva osajoukko)</b> | <b>MALLIT, JOILLA MERKITSEVÄT PARAMETRIESTIMAATIT (vakiotermit pois lukien). KAIKKI MAHDOLLISET MALLIT OVAT JALOSTUSARVON SELITTÄMINEN SEURAAVILLA APUMUUTTUJILLA: HK2 (henkilöstö), LVMK (liikevaihto), PALKKAMK (palkat), HK2 + LVMK, HK2 + PALKKAMK, LVMK + PALKKAMK, HK2 + LVMK + PALKKAMK.</b> | <b>PIENIN GREG-ESTIMAATIN VARIATIOKERROIN YHDEN APUMUUTTUJAN MALLEISTA</b> (eli malleista, joissa jalostusarvoa selitetään joko HK2:lla, LVMK:lla tai PALKKAMK:lla). | <b>PARAS VAPAAUSAS-TEKORJATTU SELITYSASTE YHDEN APUMUUTTUJAN MALLEISTA</b> | <b>PIENIN HARHANKORJAUSTERMI YHDEN APUMUUTTUJAN MALLEISTA</b> | <b>PIENIN TOTAALIN JA ESTIMAATIN EROTUS YHDEN APUMUUTTUJAN MALLEISTA</b> |
|--|---|--|--|---|--|
| A  | HK2+LVMK, LVMK+ PALKKAMK, HK2, LVMK, PALKKAMK   | HK2 (0.01541)  | LVMK(0.7428)   | HK2 (-0.00207)  | LVMK (71840.28)  |
| B  | LVMK + PALKKAMK, HK2, LVMK, PALKKAMK  | LVMK (0.04804)   | LVMK(0.8311)   | PALKKAMK(0.00083)   | LVMK (2097.9)  |
| C  | HK2 + LVMK + PALKKAMK, HK2+PALKKAMK, HK2, LVMK, PALKKAMK  | LVMK(0.14429)  | LVMK(0.4971)   | HK2 (-0.00206)  | HK2 (17149.18)   |
| D  | HK2+LVMK,LVMK+ PALKKAMK, HK2, LVMK, PALKKAMK  | LVMK (0.04869)   | LVMK (0.8857)  | HK2(-16.9)  | PALKKAMK(10255945)   |
| E  | HK2, LVMK, PALKKAMK   | PALKKAMK(0.10555)  | PALKKAMK(0.4952)   | HK2(-0.01)  | PALKKAMK(173651.74)  |
| F  | HK2 + LVMK + PALKKAMK, HK2 + PALKKAMK, LVMK + PALKKAMK, HK2, LVMK, PALKKAMK   | HK2 (0.01838)  | HK2 (0.8650)   | PALKKAMK(-0.19087)  | LVMK (-500187.5)   |
| G  | HK2+LVMK, LVMK + PALKKAMK, HK2, LVMK, PALKKAMK  | LVMK(0.02455)  | LVMK (0.6050)  | PALKKAMK(0.05381)   | PALKKAMK(546797.4)   |
| H  | HK2+LVMK+PALKKAMK, HK2+LVMK, HK2+PALKKAMK, LVMK + PALKKAMK, HK2, LVMK, PALKKAMK   | PALKKAMK (0.00848)   | PALKKAMK (0.9277)  | HK2 (0.00858)   | LVMK (-7029.54)  |
| I  | HK2, PALKKAMK   | PALKKAMK (0.05568)   | PALKKAMK (0.2145)  | HK2 (-0.02168)  | PALKKAMK(-393038.5)  |
| J  | HK2 + PALKKAMK, HK2, PALKKAMK   | HK2 (0.19079), MUTTA PIENIN GREG-ESTIMAATIN KESKIVIRHE MALLILLA, JOSSA APUMUUTTUJA ON PALKKAMK   | PALKKAMK (0.2364)  | HK2 (0.02765)   | HK2(-4445964.87)   |
| K  | HK2 + LVMK, LVMK + PALKKAMK, HK2, LVMK, PALKKAMK  | PALKKAMK(0.0285)   | LVMK (0.4723)  | HK2 (0.02094)   | LVMK(-1540468)   |
| M  | HK2+LVMK, LVMK + PALKKAMK, HK2, LVMK, PALKKAMK  | LVMK (0.06908)   | HK2 (0.6676)   | HK2(-0.00042)   | HK2 (-55972.012)   |
| N  | HK2 + LVMK, LVMK+ PALKKAMK, HK2, LVMK, PALKKAMK   | PALKKAMK (0.01049)   | PALKKAMK (0.8837)  | HK2 (-0.01096)  | PALKKAMK (19273.52)  |
| O  | HK2+LVMK, HK2 + PALKKAMK, HK2, LVMK, PALKKAMK   | PALKKAMK (0.01970)   | PALKKAMK(0.76)   | PALKKAMK(0.03317)   | HK2(19577.07)  |
| X  | HK2   | HK2(0.29119)   | HK2(0.0063)  | HK2 (0)   | HK2 (-427.6)   |
| E (vain monitoimipaikkaisten yritysten toimipaikat)          | HK2, LVMK, PALKKAMK   | PALKKAMK(0.10356)  | PALKKAMK(0.4220)   | HK2(0.00561)  | LVMK(-592793.92)   |
| E (vain yksitoimipaikkaiset yritykset)                       | HK2, PALKKAMK   | HK2(0.16515)   | HK2(0.5744)  | HK2(0.01447),   | HK2(-1252012.03)   |
| D (vain monitoimipaikkaisten yritysten toimipaikat)          | HK2 + LVMK, LVMK + PALKKAMK, HK2, LVMK, PALKKAMK  | LVMK (0.05742)   | LVMK (0.8857)  | HK2(-77.92)   | LVMK(3461559)  |
| D(vain yksitoimipaikkaiset yritykset)                        | HK2, LVMK, PALKKAMK   | LVMK(0.04072)  | LVMK(0.8695)   | HK2(-6.97060)   | LVMK(430961)   |
| KAIKKI   | HK2, LVMK, PALKKAMK, HK2 + LVMK, LVMK + PALKKAMK  | LVMK(0.03067)  | LVMK(0.8678)   | HK2(3.68)   | HK2(-1545323)  |

| <b>TOIMIALA</b> (tai kiinnostuksen kohteena oleva osajoukko) | <b>PIENIN GREG-ESTIMAATIN VARIATIOKERROIN KAHDEN APUMUUTTUJAN MALLEISTA</b> -eli malleista, joissa apumuuttujina HK2+LVMK, HK2+PALKKAMK, LVMK + PALKKAMK | <b>PARAS VAPAUSSASTEKORJATTU SELITYSASTE KAHDEN APUMUUTTUJAN MALLEISTA</b> | <b>PIENIN HARHANKORJAUSTERMI KAHDEN APUMUUTTUJAN MALLEISTA</b> | <b>PIENIN TOTAALIN JA ESTIMAATIN EROTUS KAHDEN APUMUUTTUJAN MALLEISTA</b> |
|--|--|--|--|---|
| A  | LVMK+ PALKKAMK (0.01208)   | LVMK + PALKKAMK(0.8427)  | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| B  | LVMK + PALKKAMK (0.04697)  | LVMK + PALKKAMK(0.8487)  | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | LVMK + PALKKAMK(429.409)  |
| C  | YHDEN MUUTTUJAN MALLILLA PIENEMPI VARIATIOKERROIN  | YHDEN MUUTTUJAN MALLILLA SUUREMPI VAPAUSSASTEKORJATTU SELITYSASTE          | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| D  | LVMK + PALKKAMK (0.04337)  | LVMK + PALKKAMK (0.9116)   | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | LVMK + PALKKAMK(4763561)  |
| F  | HK2+PALKKAMK (0.01781)   | HK2 +PALKKAMK (0.8840)   | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| G  | LVMK + PALKKAMK(0.02115)   | LVMK+ PALKKAMK (0.7105)  | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| H  | LVMK + PALKKAMK (0.00755)  | LVMK+ PALKKAMK (0.9493)  | LVMK + PALKKAMK(0.00768)                                       | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| J  | YHDEN MUUTTUJAN MALLEILLA PIENEMPI VARIATIOKERROIN   | HK2+PALKKAMK(0.2433)   | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| K  | LVMK + PALKKAMK(0.02754)   | HK2+PALKKAMK(0.5788)   | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| M  | HK2+LVMK(0.06551)  | HK2+LVMK(0.8022)   | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| N  | LVMK + PALKKAMK (0.00984)  | LVMK + PALKKAMK (0.8921)   | LVMK + PALKKAMK(0.00366)                                       | LVMK + PALKKAMK(4453.34)  |
| O  | YHDEN MUUTTUJAN MALLEILLA PIENEMPI VARIATIOKERROIN   | HK2+PALKKAMK(0.7641)   | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | HK2 + LVMK (-11353.04)  |
| D, monitoimipaikkaisten yritt. TP:t                          | LVMK + PALKKAMK (0.04942)  | LVMK + PALKKAMK(0.9135)  | YHDEN MUUTTUJAN MALLILLA PIENEMPI HARHANKORJAUSTERMI           | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |
| KAIKKI   | LVMK + PALKKAMK(0.02678)   | LVMK+ PALKKAMK(0.8953)   | HK2+LVMK(0)  | YHDEN MUUTTUJAN MALLILLA PIENEMPI TOD. ARVON JA ESTIMAATIN ERO            |

| <b>TOIMIALA</b> | <b>PIENIN GREG-ESTIMAATIN VARIATIOKERROIN KOLMEN APUMUUTTUJAN MALLISTA</b> -eli mallista, jossa apumuuttujina HK2+LVMK+PALKKAMK | <b>PARAS VAPAUSSASTEKORJATTU SELITYSASTE, KOLME APUMUUTTUJAA</b> | <b>PIENIN HARHANKORJAUSTERMI, KOLME APUMUUTTUJAA</b> | <b>PIENIN TOTAALIN JA ESTIMAATIN EROTUS, KOLME APUMUUTTUJAA</b> |
|-----------------|---|--|--|---|
| C               | YHDEN MUUTTUJAN MALLILLA PIENEMPI   | 0.5504   | YHDEN MUUTTUJAN MALLILLA PIENEMPI                    | YHDEN MUUTTUJAN MALLILLA PIENEMPI                               |
| F               | 0.01653   | 0.8994   | YHDEN MUUTTUJAN MALLILLA PIENEMPI                    | YHDEN MUUTTUJAN MALLILLA PIENEMPI                               |
| H               | 0.00744   | 0.9502   | KAHDEN MUUTTUJAN MALLILLA PIENEMPI                   | YHDEN MUUTTUJAN MALLILLA PIENEMPI                               |

Taulukko 6.3. Parhaat suorat GREG -estimaatit ja suorat synteettiset estimaatit sekä niiden ominaisuuksia

| TOIMI-<br>ALA | PARAS YHDEN MUUTTUJAN MALLI,<br>JOSSA MERKITSEVÄT PARAMETRI-<br>ESTIMAATIT (poislukien vakio)  | GREG-ESTIMAATTI<br>SYNTEETTINEN ES-<br>TIMAATTI | TOD.ARVO, TOD.<br>ARVON JA GREG-<br>ESTIMAATIN ERO | SELITYSASTE<br>VARIATIO-<br>KERROIN  | HARHAN-<br>KORJAUS-<br>TERMI |
|---------------|--|---|--|--|------------------------------|
| A             | LIKEVAIHTO   | 3293444.28<br>3293443.54                        | 3221604<br>71840.28                                | 0.7429<br>0.01595  | 0.73847                      |
| B             | LIKEVAIHTO   | 157336.9<br>157336.88                           | 155239<br>2097.9                                   | 0.8318<br>0.04804  | 0.01877                      |
| C             | LIKEVAIHTO. TOIMIALAKOHTAINEN<br>ESTIMAATTI TOIMII PARHAITEN.  | 1108780.78<br>1108781.81                        | 1205056<br>-96275.22                               | 0.4981<br>0.14429  | -1.03251                     |
| D             | OTOS 3 : PALKAT, MUUT OTOKSET :<br>LIKEVAIHTO. TOIMIALA - JA<br>YRITYKSEN LAJI KOHTAINEN<br>ESTIMAATTI TOIMII PARHAITEN.                           | 177965834<br>177982479.34                       | 174073314<br>-3892520                              | MONITOIMIP:<br>0.8858, 0.05742<br>YKSITOIMIP:<br>0.8695, 0.04072<br>Kaikki c.v=0.0526  | 16645.34                     |
| E             | MONITOIMIPAIKKAISET: PALKAT.<br>YKSITOIMIPAIKKAISET: HENKILÖSTÖN<br>MÄÄRÄ TOIMIALA- JA YRITYKSEN LAJI<br>KOHTAINEN ESTIMAATTI TOIMII<br>PARHAITEN. | 8354323.38<br>8354328.18                        | 8362603<br>8279.62                                 | MONITOIMIP:<br>0.4279, 0.010356<br>YKSITOIMIP:<br>0.5758, 0.16515<br>Kaikki c.v=0.1172 | 4.81                         |
| F             | OTOS 3: LIKEVAIHTO, MUUT:<br>HENKILÖSTÖ, TOIMIALAKOHTAINEN<br>ESTIMAATTI   | 27718409.4<br>27718409.78                       | 28667956<br>-949546.6                              | 0.8650<br>0.01838  | -0.38172                     |
| G             | LIKEVAIHTO   | 29793040.9<br>29793084.72                       | 28778095<br>1014945.9                              | 0.6050<br>0.02455  | -43.821                      |
| H             | PALKAT   | 4053612.78<br>4053612.87                        | 4157427<br>-103814.22                              | 0.9278<br>0.00848  | -0.08952                     |
| I             | PALKAT   | 15286748.5<br>15286750.66                       | 15679787<br>-393038.5                              | 0.2145<br>0.05568  | -2.15933                     |
| J             | PALKAT   | 7719553.61<br>7719548.78                        | 12596235<br>-4876681.39                            | 0.2372<br>0.19617  | 4.82778                      |
| K             | LIKEVAIHTO   | 29584886<br>29584865.52                         | 31125354<br>-1540468                               | 0.4724<br>0.02978  | 20.47618                     |
| M             | HENKILÖSTÖN MÄÄRÄ  | 838731.99<br>838731.99                          | 894704<br>-55972.01                                | 0.6681<br>0.07981  | -0.00042                     |
| N             | PALKAT   | 4068121.52<br>4068121.33                        | 4048848<br>19273.52                                | 0.8839<br>0.01049  | 0.18918                      |
| O             | PALKAT   | 3851352.52<br>3851352.49                        | 3930011<br>-78658.48                               | 0.7586<br>0.01970  | 0.03317                      |
| X             | HENKILÖSTÖN MÄÄRÄ  | 338.43<br>338.43                                | 766<br>-427.57                                     | 0.2547<br>0.29119  | 0                            |

Taulukon 6.3 mukaan parhaiten toimiva malli riippuu toimialasta. Paras yksittäinen selittäjä jalostusarvolle on toimialoilla A-D, G ja K liikevaihto, toimialoilla F, M ja X henkilöstön määrä, toimialoilla H-J, N ja O palkat. Toimialalla E yksitoimipaikkaisten liikeyritysten toimipaikoille paras yksittäinen selittäjä jalostusarvolle on henkilöstön määrä ja monitoimipaikkaisten yritysten toimipaikoille palkat. Tuloksista voidaan päätellä, että toimialat A-D, G ja K eli alkutuotanto ja jalostus (lukuun ottamatta toimialoja E ja F) sekä palveluista toimialat G ja K ovat pääomavaltaisia, ja muut toimialat ovat työvoimavaltaisia eli palvelujen toimialat lukuun ottamatta toimialoja G ja K sekä jalostuksen toimialat E ja F ovat työvoimavaltaisia.

Mallien toimivuuskin vaihtelee toimialasta toiseen. Parhaiten jalostusarvon estimointi toimii toimialoilla A, B, F, H, N ja O sekä vapausastekorjatun selityksasteen, RMSE:n ja keskivirheen mielessä. Näillä toimialoilla vapausastekorjattu selityksaste parhaalla mallilla on vähintään 0.74 (toimiala A) ja GREG-estimaatin variaatiokerroin maksimissaan 0.04804 (toimiala B). Suurin vapausastekorjattu selityksaste parhaalla mallilla on toimialalla H (0.9277, apumuuttujana palkat) ja pienin variaatiokerroin myös toimialalla H (0.00848, apumuuttuja-

na palkat). Tästä voidaan päätellä, että jalostusarvon estimointi on luotettavinta ja käytetty lisäinformaatio tehokkainta toimialalla H.

Taulukosta 6.2 huomataan, että huonoiten estimointi toimii kun toimiala on tuntematon. Toimialalla X vain malli, jossa jalostusarvoa selitetään henkilöstön määrällä on sellainen, jolla on merkitsevät parametriestimaatit (pois lukien vakio) ja senkin vapausastekorjattu selitysaste on 0.0063 ja GREG-estimaatin variaatiokerroin 0.29. Estimoinnin huono onnistuminen toimialalla X johtuu siitä, että vuonna 2000 on vain 11 toimipaikkaa, joiden toimiala oli tuntematon ja otoksessa niistä 4. Havaintoja oli siis hyvin vähän, eivätkä estimointitulokset ole luotettavia. Myös toimialoilla I ja J estimointi toimi huonosti. Toimialoilla I ja J vain henkilöstön määrä ja palkat olivat tilastollisesti perusteltuja selittäjiä jalostusarvolle, koska vain malleilla, joissa jalostusarvoa selitettiin palkoilla tai henkilöstön määrällä, oli merkitsevät parametriestimaatit (poislukien vakio). Toimialalla I parhaan mallin eli mallin, jossa jalostusarvoa selitetään palkoilla, vapausastekorjattu selitysaste oli vain 0.2145 ja variaatiokerroin 0.05568. Toimialalla J parhaan mallin eli mallin, jossa jalostusarvoa selitetään palkoilla, vapausastekorjattu selitysaste on 0.2364 ja variaatiokerroin 0.19617.

Epäsuoran GREG-estimaatin keskivirhe on pienempi kuin suoran silloin kun selitysaste, joka on laskettu koko joukosta on suurempi kuin selitysaste, joka on laskettu toimialakohtaisesti. Tällöin tapahtuu voiman lainaamista (borrowing strength). Vaikka voiman lainaamista tapahtuisikin saattaa epäsuora GREG-estimaatti poiketa enemmän todellisesta arvosta ja harhankorjaustermi olla itseisarvoltaan suurempi kuin suoran estimaatin harhankorjaustermi. Epäsuora synteettinen estimaatti voi olla lisäksi hyvin harhainen. Näin tapahtui tutkimuksessakin. Voiman lainaamista oli toimialoilla A-C, E, G, I-K, M, O ja X, eli kaikilla muilla toimialoilla paitsi toimialoilla D, F, H ja N. Ne toimialat, joilla voiman lainaamista ei tapahtunut olivatkin kooltaan suuria toimialoja, ja niissä toimialakohtaiset otoskoot ovat suuret. Niillä toimialoilla, joilla tapahtui voiman lainaamista, oli parhaan suoran mallin selitysaste pienempi kuin epäsuoran mallin, jossa jalostusarvoa selitetään liikevaihdolla. Kuitenkin estimaattien harhankorjaustermit olivat itseisarvoltaan suuria, epäsuora GREG-estimaatti poikkesi suoraa enemmän todellisesta arvosta ja epäsuorat synteettiset estimaatit olivat hyvin harhaisia (antoivat joskus jopa negatiivisen toimialakohtaisen jalostusarvon kokonaismäärän estimaatin). **Näistä syistä voidaan sanoa suorien estimaattien olevan epäsuoria parempia kaikilla toimialoilla.** Särndal (2001, 42-45).

**Harhankorjaustermin vaikutus oli kaikille toimialoilla jalostusarvon totaalin estimointia parantava, koska GREG-estimaatti oli kaikilla toimialoilla lähempänä todellista arvoa kuin synteettinen estimaatti.** Tämä piti paikkansa etenkin epäsuorien GREG- ja synteettisten estimaattien tapauksessa. Harhankorjaustermin suuruus vaihteli toimialasta ja lisäinformaatiosta toiseen. Toimialalla M eli koulutuksessa oli itseisarvoltaan kaikista pienin harhankorjaustermi (-0.00042, apumuuttujana HK2) ja X (harhankorjaustermi 0, apumuuttujana HK2), myös mineraalien kaivuun toimialalla (etenkin kun tutkittiin erikseen yksitoimipaikkaisten ja monitoimipaikkaisten yritysten toimipaikkoja) ja toimialalla B harhankorjaustermit olivat itseisarvoltaan pieniä. Pienimmät harhankorjaustermit (eli pienimmät erot synteettisten ja GREG-estimaattien välillä) suorista estimaateista saatiin malleilla, jotka on mainittu taulukossa 6.2. Kaikilla toimialoilla paitsi toimialoilla B, F, G, H ja N itseisarvoltaan pienimmät harhankorjaustermit saatiin käyttämällä apumuuttujana toimipaikan keskimää-

räistä henkilöstön määrä tilikauden aikana. Kaiken kaikkiaan harhankorjaustermin suuruus vaihteli kaikilla toimialoilla apumuuttujasta riippuen paljon, mutta harhankorjaustermin suhteen parhaiten toimivilla malleilla olivat harhankorjaustermien itseisarvot suuruudeltaan pieniä (eli näillä malleilla olivat synteettiset estimaatit ja GREG-estimaatit lähellä toisiaan). Estimaattien ja todellisten arvojen erotusten itseisarvojen suuruus eli estimaattien "osuvuus" vaihteli myös toimialasta toiseen. Tästä on kerrottu taulukoissa 6.2 ja 6.3. Pienin GREG-estimaatin ja todellisen arvon erotus itseisarvoltaan oli toimialalla B (-0.01366, apumuuttujat LVMK+ PALKKAMK).

Teollisuuden ja rakentamisen toimialoilla tutkittiin toimialan vaikutuksen lisäksi toimipaikkaa vastaavan yrityksen lajin vaikutusta estimoinnin tehokkuuteen. Mineraalien kaivuun toimialalla ja teollisuudessa (ts. toimialoilla C ja D) sekä yksi- että monitoimipaikkaisten yritysten toimipaikkojen joukossa toimi parhaiten malli, jossa jalostusarvoa selitettiin liikevaihdolla. Toimialalla E yksitoimipaikkaisten yritysten joukossa toimi parhaiten malli, jossa jalostusarvoa selitettiin henkilöstön määrällä ja monitoimipaikkaisten yritysten toimipaikkojen joukossa malli, jossa jalostusarvoa selitettiin palkoilla. Rakentamisen toimialalla sekä yksi- että monitoimipaikkaisten yritysten toimipaikkojen joukossa toimi sama malli eli jalostusarvon selittäminen henkilöstön määrällä (taulukko 6.2). Monitoimipaikkaisten yritysten toimipaikoilla estimointi onnistui kuitenkin selvästi paremmin kuin yksitoimipaikkaisten yritysten joukossa toimialoilla C, D ja F (taulukko 6.2). Toimialalla E (eli sähkön-, kaasun-, ja vesihuollossa) tilanne oli päinvastainen eli estimointi onnistui paremmin yksitoimipaikkaisten yritysten joukossa kuin monitoimipaikkaisten yritysten toimipaikkojen joukossa (taulukko 6.3).

Teollisuudessa ja rakentamisessa pystyttiin myös tutkimaan sitä antaako toimipaikkaa vastaavan yrityksen lajikohtainen- ja toimialakohtainen estimointi paremmat tulokset kuin toimialakohtainen (eli tutkia sitä mikä kannattaisi valita kiinnostuksen kohteena olevaksi osajoukoksi estimoinnissa). Koska toimialoilla C ja F toimialakohtaiset GREG-estimaatit ja synteettiset estimaatit olivat lähempänä todellista arvoa ja GREG-estimaatin keskivirhe pienempi kuin laskettaessa yhteen yksitoimipaikkaisten liikeyritysten jalostusarvon estimaatti ja monitoimipaikkaisten liikeyritysten jalostusarvon estimaatti toimialan sisällä, niin voidaan pitää toimialakohtaista estimaattia parempana toimialoilla C ja F kuin toimipaikkaa vastaavan yrityksen laji- ja toimialakohtaisia estimaatteja. Toimialoilla D ja E puolestaan toimialakohtaiset GREG-estimaatit ja synteettiset estimaatit olivat kauempana todellisesta arvosta kuin yhteen laskien toimiala- ja yrityksen lajikohtaiset GREG-estimaatit ja synteettiset estimaatit toimialoittain. Toimialoilla D ja E voidaan siis pitää toimiala- ja yrityksen lajikohtaisia estimaatteja parempina kuin toimialakohtaisia. Tulokset merkitsevät sitä, että se mikä kannattaa valita kiinnostuksen kohteena olevaksi osajoukoksi yrityksen tilinpäätöstietoja toimipaikkaistettaessa riippuu toimialasta. Ei voida siis sanoa sitä mikä kannattaisi valita kiinnostuksen kohteena olevaksi osajoukoksi yritysten tilinpäätöstietojen toimipaikkaistamiseen toimialoilla A, B ja G-X, koska tutkimuksessa ei ollut käytettävissä jalostusarvon tietoja monitoimipaikkaisten yritysten toimipaikoilta näillä toimialoilla.

Tutkimuksessa pyrittiin myös tutkimaan toimipaikan sijainnin merkitystä toimialakohtaisten estimaattien laadinnassa tutkimalla sitä miten toimipaikan sijaintiläänin huomioon ottaminen vaikuttaa estimaattien tarkkuuteen. Tarkastelu tehtiin kuitenkin vain teollisuuden toimialan toimipaikoille. Teollisuuden toimiala valittiin, koska toimipaikat olivat keskittyneet sekä kahdelle toimialalle (teollisuus eli toimiala D sekä tukku- ja vä-



hittäiskauppa eli toimiala G) että kahteen lääniin (Etelä-Suomen lääni ja Länsi-Suomen lääni), ja siksi toimiala- ja aluekohtaiset toimipaikkojen lukumäärät voivat olla joissain lääni- ja toimialakohtaisissa osajoukoissa pieniä. Kaikilla toimialoilla kannata siis muodostaa toimiala- ja aluekohtaisia malleja. Teollisuuden toimialalla olivat toimiala- ja läänikohtaiset otoskoot kuitenkin riittävän suuria.

Taulukossa 6.4 ovat luvun 5.6 kriteerein valitut tilastollisesti perustelluimmat estimaatit teollisuuden toimialan läänikohtaisille toimipaikkojen jalostusarvon totaaleille käyttäen kiinnostuksen kohteena oleviin osajoukkoihin jaon kriteereinä toimialaa, toimipaikkaa vastaavan yrityksen lajia ja toimipaikan sijaintilääniä tai osaa niistä. Taulukossa 6.5. tilastollisesti perustelluimmat estimaatit teollisuuden toimialan toimipaikkojen läänikohtaisille jalostusarvoille.

**Taulukko 6.4.** Toimiala- ja aluekohtaisien jalostusarvon totaalien estimaattien ominaisuuksia. Tarkasteltu vain teollisuuden toimialaa.

| <b>LÄÄNI JA OSA-JOUKKO</b> | <b>MALLIT,JOISSA MERKITSEVÄT PARAMETRIESTIMAATIT(poislukien vakio)</b>                 | <b>PIENIN HARHANKORJAUS-TERMI<br/>PIENIN TOTAALIN JA GREG-ESTIMAATIN EROTUS</b> | <b>PIENIN VARIAATIOKERROIN</b>  | <b>PARAS VAPAAUSASTE-KORJATTU SELITYSASTE</b>                    |
|----------------------------|--|---|---|--|
| <b>ETELÄ-S</b>             |  |   |   |  |
| OTOS 1                     | HK2+PALKKAMK, HK2, LVMK, PALKKAMK  | HK2(17.983)<br>HK2 +PALKKAMK(- 398836.5),<br>PALKKAMK (1127168.1)               | HK2+PALKKAMK(0.06963),<br>PALKKAMK (0.07594)  | HK2 + PALKKAMK (0.7916),<br>PALKKAMK(0.7721)                     |
| OTOS 2                     | HK2+LVMK, LVMK+ PALKKAMK,<br>HK2, LVMK, PALKKAMK                                       | PALKKAMK(-3556610.87474)<br>LVMK + PALKKAMK( 37269),<br>LVMK (190311.5)         | HK2+ LVMK (0.02548),<br>KESKIV. LVMK+PALKKAMK,<br>LVMK (0.03151)                        | LVMK + PALKKAMK(0.9575),<br>LVMK (0.9435)                        |
| OTOS3                      | HK2+LVMK+PALKKAMK,<br>HK2 + PALKKAMK, HK2, LVMK,<br>PALKKAMK                           | PALKKAMK(1.02484)<br>LVMK(1498.43)  | 3 MJAA(0.22094),<br>HK2+ PALKKAMK(0.23311),<br>PALKKAMK (0.32353)                       | 3 MJAA (0.7784),<br>HK2+ PALKKAMK (0.7339),<br>PALKKAMK (0.3775) |
| <b>LÄNSI - S</b>           |  |   |   |  |
| MONITP.                    | HK2, LVMK, PALKKAMK  | HK2 (227.11)<br>LVMK (1099803.4)  | LVMK (0.04598)  | LVMK (0.9843)  |
| YKSITP.                    | LVMK + PALKKAMK, HK2,<br>PALKKAMK  | HK2(-0.01293)<br>HK2(88949.4)   | KESKIV. LVMK+PALKKAMK,<br>HK2 (0.05208)   | LVMK + PALKKAMK(0.7612),<br>HK2(0.6894)                          |
| <b>ITÄ- S</b>              |  |   |   |  |
| OTOS1                      | HK2 + LVMK, LVMK+ PALKKAMK,<br>HK2, LVMK, PALKKAMK                                     | HK2 (-0.0044)<br>LVMK (-117084.05)  | LVMK+PALKKAMK(0.07782),<br>PALKKAMK(0.09319)  | LVMK + PALKKAMK(0.6855),<br>PALKKAMK(0.6729)                     |
| OTOS2                      | HK2 + LVMK, LVMK + PALKKAMK,<br>HK2, LVMK, PALKKAMK                                    | HK2 (0.05078)<br>LVMK (-60713.7)  | PALKKAMK (0.02940)  | LVMK + PALKKAMK<br>(0.9364), PALKKAMK(0.9139)                    |
| OTOS3                      | LVMK+ PALKKAMK, HK2, LVMK,<br>PALKKAMK   | HK2(-0.00006)<br>PALKKAMK(3081.269)   | LVMK(0.03009),<br>KESKIV. PALKKAMK  | LVMK + PALKKAMK<br>(0.9807), LVMK (0.9722)                       |
| <b>OULUN L</b>             |  |   |   |  |
| MONITP.                    | HK2+LVMK, HK2+ PALKKAMK,<br>LVMK+ PALKKAMK, HK2, LVMK,<br>PALKKAMK                     | HK2 (-6.657)<br>HK2 (2362531.4)   | RMSE LVMK + PALKKAMK,<br>HK2+PALKKAMK(0.05024),<br>PALKKAMK(0.05818)                    | LVMK+PALKKAMK(0.9842),<br>PALKKAMK(0.9749)                       |
| YKSITP.                    | HK2+LVMK, LVMK+ PALKKAMK,<br>HK2, LVMK, PALKKAMK                                       | HK2(0.00029)<br>LVMK+PALKKAMK (3860.54),<br>HK2 (49828.27)                      | HK2+LVMK (0.02840),<br>HK2 (0.04021),<br>RMSE PALKKAMK                                  | HK2 + LVMK(0.9347),<br>PALKKAMK(0.8969)                          |
| <b>LAPPI</b>               | HK2, LVMK, PALKKAMK  | HK2(0.44465)<br>PALKKAMK(328460.57)   | PALKKAMK(0.15004)   | PALKKAMK(0.9447)   |
| <b>AHVEN-MAA</b>           | HK2+LVMK+PALKKAMK,<br>HK2+LVMK, HK2+PALKKAMK,<br>LVMK+PALKKAMK, HK2, LVMK,<br>PALKKAMK | HK2(-0.00027)<br>HK2 + LVMK(1404.489),<br>PALKKAMK (7608.947)                   | 3 MJAA (0.08155),<br>HK2+ PALKKAMK (0.1025),<br>RMSE LVMK+ PALKKAMK,<br>PALKKAMK (0.13) | 3 MJAA(0.9531),<br>LVMK+ PALKKAMK(0.9163),<br>PALKKAMK(0.8316)   |

Toimiala- ja aluekohtaiset estimaatit muodostuvat kaikissa muissa lääneissä paitsi Ahvenanmaalla ja Lapissa laskettaessa yhteen yksi- ja monitoimipaikkaisten yritysten toimipaikkojen jalostusarvon totaalien tiedot. Tarkasteltaessa taulukkoja 6.4. ja 6.5 huomataan, että **kun toimipaikan sijaintilääni otetaan huomioon estimoinnissa (ts. se on yksi kiinnostuksen kohteena olevista osajoukoista), saadaan vain GREG-estimaatin ja todellisen arvon erotuksen mielessä tarkempi estimaatti teollisuuden toimialalla toimivien toimipaikkojen jalostusarvojen totaalille. Toimialakohtainen estimaatti ei kuitenkaan parantunut variatiokerroimen, harhankorjaustermien ja keskivirheen mielessä otettaessa huomioon toimipaikan sijaintilääni.** Lisäksi se, että teollisuuden toimialalla toimii paljon toimipaikkoja, vaikuttaa estimoinnin tarkkuuteen. Edellä mainittujen syiden vuoksi ei voidakaan sanoa parantaisiko toimipaikan sijaintiläänin huomioon ottaminen estimointia muilla toimialoilla (etenkin pienillä toimialoilla). Koska toimipaikan sijainti ei ole ollut mukana estimaattien laadinnassa muille kuin teollisuuden toimialan toimipaikoille, ei sitä oteta huomioon toimipaikkaistettaessa yritysten jalostusarvoja. Tällöin voidaan tutkia sitä miten tarkkoja estimaatteja toimipaikkojen jalostusarvolle antavat toimialakohtaiset mallit. Teollisuuden ja rakentamisen toimialoilla käytetään toimipaikkojen jalostusarvon estimoinnissa myös toimipaikka vastaavan yrityksen laji -tietoa, jos sen on huomattu parantavan toimialakohtaisten estimaattien tarkkuutta.

**Taulukko 6.5.** Tilastollisesti perustelluimmat estimaatit lääni- ja toimialakohtaisille jalostusarvon totaaleille sekä toimialakohtaiselle jalostusarvon totaalille. Tarkasteltu vain teollisuuden toimialaa.

| OTOS  | LÄÄNI  | GREG- ESTI-<br>MAATTI | SYNTEETTI-<br>NEN ESTI-<br>MAATTI | TODELLI-<br>NEN<br>ARVO | KESKIVIRHE  | TOD.ARVON<br>JA GREG-ES-<br>TIMAATIN ERO | HARHAN-<br>KORJAUS-<br>TERMI | VARIAA-<br>TIOKER-<br>ROIN | APUMJA                              |
|-------|--------|-----------------------|-----------------------------------|-------------------------|-------------|--|------------------------------|----------------------------|-------------------------------------|
| 1+2+3 | 1      | 66349239.63           | 70511957.69                       | 65213454                | 4897066.91  | -1135785.63                              | 4162718.06                   | 0.07                       | OTOKSET 1 JA 3: PALKKAMK, MUUT LVMK |
| 1+6   | 2      | 74744581.80           | 74747072.04                       | 73555829                | 3567250.17  | -1188752.80                              | 2490.24                      | 0.05                       | MONITOIMIP. LVMK, YKSITOIMIP. HK2   |
| 1+2+3 | 3      | 10091276.68           | 10091275.42                       | 12601520                | 705541.30   | 2510243.32                               | -1.26                        | 0.07                       | PALKKAMK                            |
| 1+6   | 4      | 13369536.04           | 13371282.49                       | 15792306                | 734005.40   | 2422769.96                               | 1746.45                      | 0.05                       | PALKKAMK                            |
| 5     | 5      | 6933517.57            | 6933461.72                        | 6605057                 | 1040336.48  | -328460.57                               | -55.85                       | 0.15                       | PALKKAMK                            |
| 5     | 6      | 312755.95             | 312756.34                         | 305147                  | 40581.83    | -7608.95                                 | 0.39                         | 0.13                       | PALKKAMK                            |
| YHT.  | KAIKKI | 171800907.67          | 175967805.7                       | 174073313               | 10984782.09 | 2272405.33                               | 4166898.04                   | 0.06                       |                                     |

## 6.2 Yritysten jalostusarvon toimipaikkaistaminen ja toimipaikkaistamisen rajoitteet

Yritysten jalostusarvot toimipaikkaistetaan tutkimuksessa regressioperusteisen imputoinnin avulla. Luvussa 6.1 ja liitteessä 8 on kerrottu mallit, joilla puuttuvat tiedot imputoidaan. Apumuuttujat ovat toimialoilla A -D, G ja K liikevaihto, toimialoilla F, M ja X henkilöstön määrä, toimialoilla H-J, N ja O palkat, toimialalla E yksitoimipaikkaisten liikeyritysten toimipaikoille henkilöstön määrä ja monitoimipaikkaisten yritysten toimipaikoille palkat. Toimialoilla D ja E käytetään imputoinnissa toimipaikkaa vastaavan yrityksen laji- ja toimialakohtaisia  $\beta$ -parametrien estimaatteja, muilla toimialoilla toimialakohtaisia  $\beta$ -parametrien estimaatteja. SAS-koodi, jolla imputointi on tehty, on liitteenä 9. Vertailukohtana on imputoitu kaikki puuttuvat tiedot samalla mallilla (apumuuttujana liikevaihto, liite 8 ja luku 5.3) eli käyttäen kaikille toimipaikoille samoja  $\beta$ -parametrien estimaatteja. Pohjalla ovat siis epäsuorat synteettiset estimaatit ja GREG-estimaatit.

Koska puuttuvat tiedot, eli niiden toimipaikkojen tilinpäätöstiedot, joiden laskennallisen käyttökäteen tiedot täytyy imputoida, ovat lähes täysin monitoimipaikkaisten yritysten toimipaikkojen tietoja, niin imputoinnin onnistuminen riippuu toimialoilla A, B ja G-O hyvin pitkälti siitä miten hyvin yksitoimipaikkaisten yritysten tietojen pohjalta luodut mallit selittävät monitoimipaikkaisten yritysten toimipaikkojen tietoja, eli sopivatko samat  $\beta$ -parametrien estimaatit toimialojen sisällä yksitoimipaikkaisten liikeyritysten jalostusarvojen imputoimisen lisäksi monitoimipaikkaisten liikeyritysten toimipaikkojen jalostusarvojen imputoimiseen. Jos näin on, niin tutkimuksessa oleva kato on harmitonta tai täysin harmitonta (luku 5.4). Tätä ei voida tutkimuksella selvittää, koska ei voida laskea todellisten arvojen ja imputoitujen arvojen erotuksia kaikille toimipaikoille. Toisaalta tutkimuksella voidaan selvittää se miten imputointi onnistuu käytetyillä malleilla siinä joukossa, jonka jalostusarvo on tiedossa, eli voidaan laskea imputointivariansseja toimialoittain sekä tutkia todellisten arvojen ja imputoitujen arvojen erotuksia. Imputoinnin onnistumisen rajoituksista on lisää luvuissa 3, 5.4, 5.5 ja 8.

Taulukossa 6.6 ovat toimialakohtaiset estimaatit jalostusarvojen totaalille käyttäen toimiala tai toimiala- ja yrityksen lajikohtaisia  $\beta$ -parametreja, niiden asetelma- ja imputointivarianssit (luku 5.5) sekä todellisen totaalin ja imputoidun totaalin erotukset niiden toimipaikkojen joukossa, joille jalostusarvo on tiedossa, sekä käytetty apumuuttuja(jat). Taulukossa 6.7 ovat toimialakohtaiset estimaatit jalostusarvojen totaalille käyttäen samoja  $\beta$ -parametrien estimaatteja kaikille toimipaikoille (apumuuttuja liikevaihto), niiden asetelma- ja imputointivarianssit (luku 5.5) sekä todellisen totaalin ja imputoidun totaalin erotukset niiden toimipaikkojen joukossa, joille jalostusarvo on tiedossa. Taulukoissa 6.6 ja 6.7 verrataan kalibroituja tietoja kalibroimattomiin.

Taulukkojen 6.6. ja 6.7 perusteella sekä suorien että epäsuorien estimaattien pohjalta laaditun mallin toimivuus vastanneiden joukossa näyttää paranevan kun imputoidut havaintoarvot kalibroidaan. Tällöin yksitoimipaikkaisille liikeyrityksille imputoitu havaintoarvo = todellinen havaintoarvo = yrityksen jalostusarvo. Tällöin imputointivarianssikin on nolla. Koska tutkimusaineistossa oli tiedossa jalostusarvo toimialoilla A, B ja G-X, vain, jos toimipaikkaa vastaava yritys on yksitoimipaikkainen, niin imputointivarianssin tulisi tutkimuksessa näillä toimialoilla olla nolla, kun imputoidut havaintoarvot on kalibroitu. Näin kävikin lukuun ottamatta toimialoja K ja G.

Monitoimipaikkaisten liikeyritysten toimipaikoille kalibrointi ei tuota tulosta, jossa vastanneiden joukossa imputoitu havaintoarvo olisi sama kuin todellinen arvo. Kuitenkin lähes kaikilla toimialoilla imputointivarianssi on pienempi kun imputoidut havaintoarvot on kalibroitu, kuin jos kalibrointia ei olisi tehty, poikkeuksena ovat toimialat C ja D. Myös todellisten arvojen (vastanneiden joukossa) ja imputoitujen arvojen erotukset pienenevät itseisarvoltaan kun imputoidut havaintoarvot kalibroidaan, poikkeuksena ovat epäsuoriin estimaatteihin perustuvassa imputoinnissa toimialat C ja D ja suoriin estimaatteihin perustuvassa imputoinnissa toimiala C.

**Taulukko 6.6.** Kalibroidut ja kalibroimattomat toimialakohtaiset estimaatit jalostusarvojen totaalle (1000 mk) käyttäen toimiala tai toimiala- ja yrityksen lajikohtaisia  $\beta$ -parametrien estimaatteja.

| TOIMI-ALA | KALIBROITU ESTIMAATTI JALOSTUS-ARVOLLE | KALIBROIMATON ESTIMAATTI JALOSTUS-ARVOLLE | KALIBROIDUN ESTIMAATTIN IMPUTOINTIKESKIVIRHE | KALIBROIMATTOMAN ESTIMAATTIN IMPUTOINTIKESKIVIRHE | JALOSTUS-ARVON TOTAALI (VASTANNEET) | TOD. ARVON JA KALIBROIDUN ESTIMAATTIN ERO (VAST.) | TOD.ARVON JA KALIBROIMATTOMAN ESTIMAATTIN ERO (VAST.) | ASETELMAVARIANSSIN NELIÖJUURI |
|-----------|--|---|--|---|-------------------------------------|---|---|-------------------------------|
| A         | 5456110.32                             | 6663626.08                                | 0  | 43675   | 3221604                             | 0   | -71840.276  | 81021                         |
| B         | 190287.351                             | 194910.583                                | 0  | 8030.172227                                       | 155239                              | 0   | -2097.9004  | 14527                         |
| C         | 1316824.26                             | 1357413.38                                | 196773                                       | 190700  | 1205056.39                          | -19653.819  | 96275.6059  | 200798                        |
| D         | 193975670                              | 209847969                                 | 6896613                                      | 7064875   | 174073314                           | 12713693.7  | -3876525.8  | 15367082                      |
| E         | 10744390.2                             | 12862269                                  | 590951                                       | 1035520   | 8362603.12                          | 1761610.93  | 48003.7138  | 1412932                       |
| F         | 31985398.5                             | 33548102.9                                | 574852                                       | 578034  | 28667956.4                          | 3093025.61  | 954767.568  | 1065835                       |
| G         | 60261194.7                             | 54290591                                  | 466.777625                                   | 614791  | 28778094.9                          | 466.777625  | -1014945.9  | 697411                        |
| H         | 9372260.88                             | 8975289.68                                | 0  | 32365   | 4157426.88                          | 0   | 103814.103  | 115721                        |
| I         | 42704655.8                             | 34523074.7                                | 0  | 435076  | 15679787.4                          | 0   | 393038.842  | 600648                        |
| J         | 13946352.2                             | 15500049.7                                | 0  | 3430189   | 12596235                            | 0   | 4876681.39  | 3820903                       |
| K         | 48885767.2                             | 4430659.4                                 | 78026  | 810866  | 31125353.9                          | 78025.5638  | 1540467.98  | 1028571                       |
| M         | 1526632.62                             | 1479597.66                                | 0  | 38823   | 894704.107                          | 0   | 55972.1183  | 91518                         |
| N         | 5273890.8                              | 5368195.31                                | 0  | 32167   | 4048848.17                          | 0   | -19273.354  | 87674                         |
| O         | 11471734.3                             | 7224043.07                                | 0  | 59831   | 3930011.3                           | 0   | 78658.7852  | 116160                        |
| X         | 766                                    | 766                                       | 0  | 224.311377  | 766                                 | 0   | 427.567568  | 212.63819                     |
| YHT.      | 437111935.1                            | 396265558.3                               | 6949391                                      | 8022362   | 316897000.5                         | 17627168.8  | 3163424.38  | 16023842                      |

**Taulukko 6.7.** Kalibroidut ja kalibroimattomat toimialakohtaiset estimaatit jalostusarvojen totaalle (1000 mk) käyttäen samoja  $\beta$ -parametrien estimaatteja kaikille toimipaikoille sekä niiden ominaisuuksia.

| TOIMI-ALA | KALIBROITU ESTIMAATTI JALOSTUS-ARVOLLE | KALIBROIMATON ESTIMAATTI JALOSTUS-ARVOLLE | KALIBROIDUN ESTIMAATTIN IMPUTOINTIKESKIVIRHE | KALIBROIMATTOMAN ESTIMAATTIN IMPUTOINTIKESKIVIRHE | JALOSTUS-ARVON TOTAALI (VASTANNEET) | TOD. ARVON JA KALIBROIDUN ESTIMAATTIN ERO (VAST.) | KALIBROIMATTOMAN ESTIMAATTIN JA TOD. ARVON EROTUS (VAST.) | ASETELMAVARIANSSIN NELIÖJUURI |
|-----------|--|---|--|---|-------------------------------------|---|---|-------------------------------|
| A         | 5118142.2                              | 5410181.12                                | 0  | 52090   | 3221604                             | 0   | -426862.59  | 90459                         |
| B         | 199333.894                             | 193679.51                                 | 0  | 10164   | 155239                              | 0   | -155722.65  | 15782                         |
| C         | 1329547.85                             | 1365468.72                                | 231278                                       | 187162  | 1205056.39                          | -200960.88  | -93409.668  | 203655                        |
| D         | 192587813                              | 185023499                                 | 7096670                                      | 7093603   | 174073314                           | 11497286.2  | 35874769.7  | 15399441                      |
| E         | 9062862.69                             | 9260296.98                                | 761708                                       | 1228553   | 8362603.12                          | 1104258.63  | 921576.053  | 1437958                       |
| F         | 31219510.7                             | 31458419.9                                | 537991                                       | 711472  | 28667956.4                          | 2644874.7   | 85611.9325  | 1077392                       |
| G         | 65751355.8                             | 107868077                                 | 158.822835                                   | 2287236   | 28778094.9                          | 158.822835  | -35815919   | 707721                        |
| H         | 8750279.08                             | 8164784.57                                | 0  | 59651   | 4157426.88                          | 0   | -1941252.4  | 123063                        |
| I         | 42891389                               | 34211684.3                                | 0  | 509891  | 15679787.4                          | 0   | -1251311  | 608604                        |
| J         | 13866683.9                             | 12744514.1                                | 0  | 3827698   | 12596235                            | 0   | 11778798.9  | 3828777                       |
| K         | 48597162.3                             | 41054178.9                                | 30989  | 830651  | 31125353.9                          | 30989.3013  | 2765794.67  | 1038213                       |
| M         | 1483317.87                             | 1085277.55                                | 0  | 78898   | 894704.107                          | 0   | 89158.7103  | 94401                         |
| N         | 5237323.93                             | 4849401.3                                 | 0  | 59094   | 4048848.17                          | 0   | -1665441.6  | 94734                         |
| O         | 11491127.4                             | 8177126.29                                | 0  | 104608  | 3930011.3                           | 0   | -3923028  | 119795                        |
| X         | 766                                    | 766                                       | 0  | 1029.224035                                       | 766                                 | 0   | -3353.8516  | 307.316927                    |
| YHT.      | 437586615.6                            | 450867355.2                               | 7161449                                      | 8558095   | 316897000.5                         | 15076606.77                                       | 6239409.206   | 16023842                      |

*Tulosten mukaan suorien estimaattoreiden pohjalta tehty imputoiminen on luotettavampaa kuin epäsuorien pohjalta tehty, koska suorien estimaattoreiden pohjalta laaditussa imputoinnissa ovat imputointivarianssit pienempiä kuin epäsuorien perusteella, samoin poikkeavat suorien estimaattoreiden pohjalta laaditut imputoidut havaintoarvot vähemmän todellisista arvoista kuin epäsuorien pohjalta laaditut imputoidut havaintoarvot.*

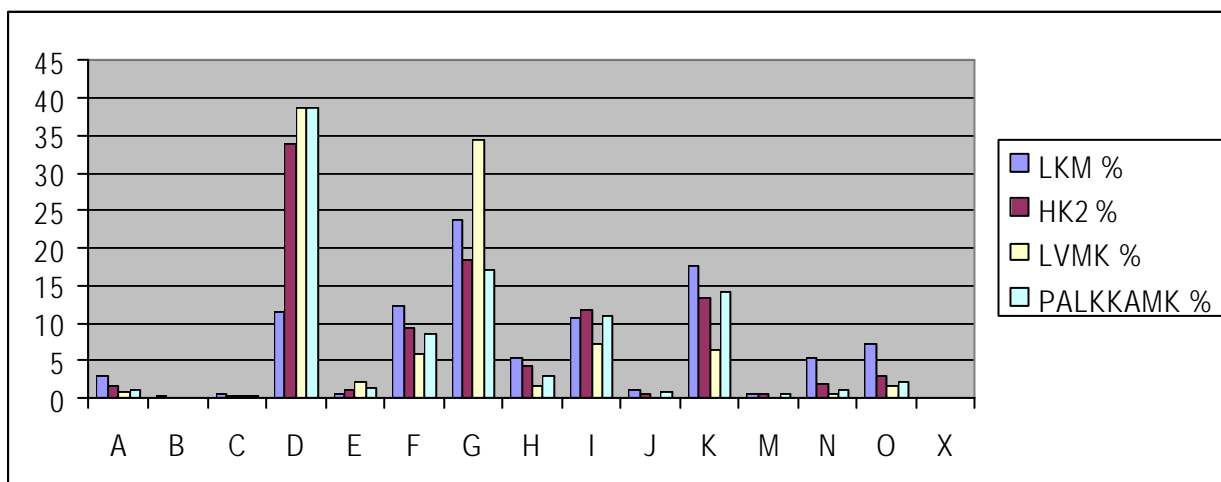
## 6.3 Toimialojen merkittävyys vuonna 2000

Luvussa 6.3. tutkitaan toimialojen pääluokkien merkittävyyttä toimipaikkojen tilinpäätöstietojen avulla. Näitä tilinpäätöstietoja ovat henkilöstön määrä, palkat, liikevaihto ja jalostusarvo. Eri toimialojen suhteelliset osuudet toimipaikoista, henkilöstön määrästä, liikevaihdosta ja palkoista ovat taulukossa 6.8. Sen alla on näitä suhteellisia osuuksia esittävä pylväskuvio.

Taulukon 6.8 mukaan kaikkien vuonna 2000 toimineiden toimipaikkojen joukossa näyttää siltä, että suurimmat toimialat toimipaikkojen lukumäärän, henkilöstön määrän, liikevaihdon ja palkkojen suhteen ovat D ja G (teollisuus sekä tukku- ja vähittäiskauppa). Teollisuus on suurin toimiala kaikissa muissa ominaisuuksissa paitsi toimipaikkojen lukumäärässä, jossa se on neljänneksi suurin toimiala. Kolmanneksi suurin toimiala on K eli kiinteistö-, vuokraus- ja tutkimuspalvelut. Myös kuljetus-, varastointi- ja tietoliikenne sekä rakentaminen eli toimialat I ja F ovat suuria toimialoja toimipaikkojen lukumäärän, henkilöstön määrän, palkkojen ja liikevaihdon suhteen.

**Taulukko 6.8.** Toimialojen prosentuaaliset osuudet toimipaikkojen lukumäärästä, henkilöstöstä, liikevaihdosta ja palkoista vuonna 2000 toimipaikkatietojen mukaan.

| TOIMIALA | N      | %     | HK2       | %      | LVMK       | %       | PALKKAMK  | %       |
|----------|--------|-------|-----------|--------|------------|---------|-----------|---------|
| A        | 7106   | 2.94  | 20013.9   | 1.62   | 14126969   | 0.94    | 2207866   | 1.21    |
| B        | 669    | 0.28  | 735.7     | 0.06   | 493763     | 0.03    | 59252     | 0.03    |
| C        | 1306   | 0.54  | 4291.2    | 0.35   | 4209652    | 0.28    | 652128    | 0.36    |
| D        | 28039  | 11.59 | 418180.3  | 33.93  | 582711805  | 38.63   | 70493250  | 38.56   |
| E        | 1185   | 0.49  | 12700     | 1.03   | 33131294   | 2.20    | 2487066   | 1.36    |
| F        | 29491  | 12.19 | 116861.9  | 9.48   | 88497197   | 5.87    | 15808474  | 8.65    |
| G        | 57511  | 23.77 | 226378.8  | 18.37  | 517313620  | 34.30   | 31243636  | 17.09   |
| H        | 12920  | 5.34  | 51532.8   | 4.18   | 23567900   | 1.56    | 5335372   | 2.92    |
| I        | 26030  | 10.76 | 143928.1  | 11.68  | 110586194  | 7.33    | 20128654  | 11.01   |
| J        | 2794   | 1.16  | 7423.2    | 0.60   | 0          | 0.00    | 1725278   | 0.94    |
| K        | 42887  | 17.72 | 164433.1  | 13.34  | 98046552   | 6.50    | 25665350  | 14.04   |
| M        | 1687   | 0.70  | 6332.6    | 0.51   | 1725797    | 0.11    | 1021466   | 0.56    |
| N        | 12985  | 5.37  | 23259.9   | 1.89   | 8468498    | 0.56    | 2233647   | 1.22    |
| O        | 17366  | 7.18  | 36372.3   | 2.95   | 25479138   | 1.69    | 3763018   | 2.06    |
| X        | 11     | 0.005 | 2.4       | 0.0002 | 1328       | 0.00009 | 60        | 0.00003 |
| YHT.     | 241987 | 100   | 1232446.2 | 100    | 1508359707 | 100     | 182824517 | 100     |



Toimialoilla A, B, F, H, I ja N-O ovat henkilöstön määrän osuudet suurempia kuin palkkojen ja liikevaihdon osuudet. Toimialoilla C ja JM ovat palkkojen osuudet suurempia kuin henkilöstön ja liikevaihdon osuudet. Toimialoilla D, E ja G ovat liikevaihdon osuudet suuremmat kuin henkilöstön ja palkkojen osuudet ja niitä voidaan pitää pääomavaltaisina. Kaikkia muita toimialoja voidaan pitää työvoimavaltaisina. Kaikilla muilla toimialoilla paitsi teollisuudessa, tukku- ja vähittäiskaupassa ja kuljetus-, varastointi- ja tietoliikenteessä (ts. toimialat D, G ja I) ovat toimipaikkojen lukumäärän osuudet suuremmat kuin henkilöstön, palkkojen ja liikevaihdon osuudet.

Kun tutkitaan eri elinkeinojen osuuksia edellä mainituista ominaisuuksista niin huomataan, että jalostuksen toimialojen osuudet liikevaihdosta, henkilöstön määrästä ja palkoista ovat niiden toimipaikkojen lukumäärän osuutta selvästi suuremmat (kaikilla muilla toimialoilla paitsi rakentamisessa) kun taas palvelujen toimialoilla ja alkutuotannossa toimipaikkojen lukumäärän osuudet ovat suurempia kuin liikevaihdon, henkilöstön määrän tai palkkojen osuudet. Toisin sanoen voidaan pitää jalostuksen toimialoilla toimivia toimialojen merkittävyydeltään suurempina kuin palvelujen tai alkutuotannon toimialoja (poikkeuksena toimialat G ja I). Erot elinkeinojen välillä voivat johtua siitä, että jalostuksen elinkeinoa harjoittavat toimipaikoilla olisi korkeampi kokonaistuottavuus kuin alkutuotantoa ja palveluita harjoittavat toimipaikoilla (poikkeuksena toimialat G ja I).

Taulukko 6.9 esittää arvioita toimialojen osuuksille jalostusarvosta. Taulukosta huomataan, että myös toimipaikan jalostusarvon osuuksien arvioiden perusteella on tehtävissä saman suuntaisia arvioita toimialojen merkittävyydestä kuin henkilöstön määrän, liikevaihdon ja maksettujen palkkojenkin suhteen. Toimipaikan jalostusarvon osuudet ovatkin suurimmat toimialoilla D ja G.

**Taulukko 6.9.** Arvioita toimialojen osuuksille jalostusarvosta. Jalostusarvon kokonaismäärät tuhansina markkoina.

| TOIMI-<br>ALA | JALOSTUSAR-<br>VO. TOIMIPAI-<br>KAT, JOIDEN<br>JALOSTUSARVO<br>ON TIEDOSSA | %     | KALIBROITU ES-<br>TIMAATTI, TOIMI-<br>ALAKOHTAISET<br>MALLIT | %     | KALIBROIMA-<br>TON ESTIMAAT-<br>TI, TOIMIALA-<br>KOHTAISET<br>MALLIT | %     | KALIBROI-<br>TU ESTI-<br>MAATTI,<br>YLEINEN<br>MALLI | %     | KALIBROI-<br>MATON ES-<br>TIMAATTI,<br>YLEINEN<br>MALLI. | %     |
|---------------|--|-------|--|-------|--|-------|--|-------|--|-------|
| A             | 3221604  | 1.02  | 5456110.3  | 1.25  | 6663626.1  | 1.68  | 5118142.2  | 1.17  | 5410181.1  | 1.2   |
| B             | 155239   | 0.05  | 190287.35  | 0.04  | 194910.58  | 0.05  | 199333.89  | 0.05  | 193679.51  | 0.04  |
| C             | 1205056.39   | 0.38  | 1316824.3  | 0.3   | 1357413.4  | 0.34  | 1329547.9  | 0.3   | 1365468.7  | 0.3   |
| D             | 174073314  | 54.93 | 193975670  | 44.38 | 209847969  | 52.96 | 192587813  | 44.01 | 185023499  | 41.04 |
| E             | 8362603.12   | 2.64  | 10744390   | 2.46  | 12862269   | 3.25  | 9062862.7  | 2.07  | 9260297  | 2.05  |
| F             | 28667956.4   | 9.05  | 31985399   | 7.32  | 33548103   | 8.47  | 31219511   | 7.14  | 31458420   | 6.98  |
| G             | 28778094.9   | 9.08  | 60261195   | 13.79 | 54290591   | 13.7  | 65751356   | 15.03 | 107868077  | 23.92 |
| H             | 4157426.68   | 1.31  | 9372260.9  | 2.14  | 8975289.7  | 2.26  | 8750279.1  | 2     | 8164784.6  | 1.81  |
| I             | 15679787.4   | 4.95  | 42704656   | 9.77  | 34523075   | 8.71  | 42891389   | 9.8   | 34211684   | 7.59  |
| J             | 12596235   | 3.97  | 13946352   | 3.19  | 15500050   | 3.91  | 13866684   | 3.17  | 12744514   | 2.83  |
| K             | 31125353.39  | 9.82  | 48885767   | 11.18 | 4430659.4  | 1.12  | 48597162   | 11.11 | 41054179   | 9.11  |
| M             | 894704.107   | 0.28  | 1526632.6  | 0.35  | 1479597.7  | 0.37  | 1483317.9  | 0.34  | 1085277.6  | 0.24  |
| N             | 4048848.17   | 1.28  | 5273890.8  | 1.21  | 5368195.3  | 1.36  | 5237323.9  | 1.2   | 4849401.3  | 1.08  |
| O             | 3930011.3  | 1.24  | 11471734   | 2.62  | 7224043.1  | 1.82  | 11491127   | 2.63  | 8177126.3  | 1.81  |
| X             | 766  | 0     | 766  | 0     | 766  | 0     | 766  | 0     | 766  | 0     |
| YHT.          | 316897000.5  | 100   | 437111935.1  | 100   | 396266558.3  | 100   | 437586616  | 100   | 450867356  | 100   |

Koska monitoimipaikkaisten yritysten toimipaikkojen tilinpäätöstiedot ovat yleensä suurempia kuin yksitoimipaikkaisten liikeyritysten, voidaan monitoimipaikkaisten yritysten toimipaikkoja pitää yksitoimipaikkaisia liikeyrityksiä merkittävimpinä. Kuitenkin toimialojen osuudet ovat likipitään samat kuin kaikkien toimipaikkojen joukossa sekä yksitoimipaikkaisten yritysten toimipaikoilla että monitoimipaikkaisten yritysten toimipaikoilla.

Taulukossa 6.10 on esitetty yritysten päätoimialojen osuudet yrityksistä, henkilöstöstä, liikevaihdosta ja jalostusarvosta yrityksen päätoimialan mukaan. Tässä tarkastelussa ovat mukana liikeyritykset, jotka ovat yritysrekisterissä (N = 222 817) ja niiden päätoimiala on määritelty yritysrekisterin mukaan.

Taulukon 6.10 mukaan liikeyrityksiä joiden päätoimiala on G (tukku- ja vähittäiskauppa) oli vuonna 2000 yritysrekisterissä eniten. Kuitenkin henkilöstömäärästä, liikevaihdosta ja jalostusarvosta oli teollisuuden yritysten osuus suurin vuonna 2000. Toiseksi suurimmat osuudet olivat toimialan G yrityksellä ja muiden toimialojen osuudet olivat pieniä. Edellä esitetty antaa aiheen olettaa, että vuonna 2000 oli suurin osa yrityksistä sellaisia, joiden päätoimiala oli teollisuus tai tukku- ja vähittäiskauppa mitattuna yritysten määrällä, henkilöstön määrällä tai liikevaihdolla. Toimialoilla P-X ei toiminut yhtään liikeyritystä. (Tilastokeskus, 2002 e, 21).

Taulukkoja 6.8 ja 6.10 vertaamalla huomataan, että yritysrekisterin mukaan toimialojen osuudet yritysten lukumäärälle, henkilöstön määrälle ja liikevaihdolle ovat suuruusluokaltaan samoja eli teollisuus sekä tukku- ja vähittäiskauppa ovat senkin mukaan merkittävimmät toimialat Suomessa. Toisin sanoen se kuinka merkittävä toimiala on, ei riipu siitä tutkitaanko asiaa toimipaikan toimialan kautta vai yrityksen päätoimialan kautta. Yrityksen päätoimialan suhteen tutkittaessa on siis huomattavissa, että jalostuksen elinkeinon osuudet henkilöstön määrässä, liikevaihdossa, palkoissa ja jalostusarvossa ovat niiden lukumäärää suurempia, mutta alku- tuotannossa ja palveluissa pienempiä (lukuun ottamatta toimialoja G ja I).

Kaiken kaikkiaan toimipaikkojen ja yritysten lukumäärä on jakautunut tasaisemmin eri toimialoille kuin yritysten ja toimipaikkojen tilinpäätöstiedot. Etenkin jalostusarvo ja liikevaihto (ne ovat keskittyneet vielä enemmän kuin henkilöstö ja palkat) ovat keskittyneet teollisuuden ja tukku- ja vähittäiskaupan toimialoille. Jalostusarvossa keskittyminen on kaikkein suurinta yritysten päätoimialan mukaan katsottuna että toimipaikan toimialan mukaan katsottuna (teollisuuden osuus on n. 45 % yrityksen päätoimialan mukaan ja n. 55 % toimipaikan toimialan mukaan). Huomattava on sekin, että tukku- ja vähittäiskaupan osuudet palkoista ja henkilöstön määrästä ovat pienempiä kuin osuus liikevaihdosta, mutta teollisuudessa palkkojen, henkilöstön määrän ja liikevaihdon osuudet ovat likimain yhtä suuret. Ero voi johtua siitä, että tukku- ja vähittäiskaupassa maksetaan pienempiä palkkoja ja työskennellään useammin osa-aikaisena kuin teollisuudessa.

Kansantalouden tilinpitokin tuottaa erilaisia toimiala- ja aluekohtaisia tilinpäätöstietoja ja tutkimuksessa saatuja tuloksia toimialojen merkittävydestä sekä toimialojen tilinpäätöstietojen estimaateista voidaan verrata kansantalouden tilinpidon antamiin tietoihin. Kansantalouden tilinpito käyttää tietojensa pohjalla lähinnä yrityksen päätoimialan mukaisia tietoja. Täsmällisempi vertailu jää kuitenkin tämän tutkimuksen ulkopuolelle.

**Taulukko 6.10.** Toimialojen osuudet yrityksistä, henkilöstöstä, liikevaihdosta ja jalostusarvosta. Mukana liikeyritykset (N = 222 817), joiden päätoimiala on määritelty yritysrekisterin mukaan. (Tilastokeskus, 2002e, 21).

| TOIMIALA | YRITYSTEN LKM | %   | HENKILÖSTÖ | %   | LIKEVAIHTO MILJ. EUR | %   | JALOSTUSARVO MILJ. EUR | %    |
|----------|---------------|-----|------------|-----|----------------------|-----|------------------------|------|
| A ja B   | 7852          | 3.5 | 22096      | 1.7 | 2559                 | 1   | 661                    | 1    |
| C        | 1154          | 0.5 | 3512       | 0.3 | 622                  | 0.2 | 207                    | 0.3  |
| D        | 25687         | 12  | 423 491    | 33  | 102 232              | 39  | 30 816                 | 44.7 |
| E        | 838           | 0.4 | 12 871     | 1   | 7019                 | 2.7 | 1579                   | 2.3  |
| F        | 28758         | 13  | 116 926    | 9   | 14 905               | 5.7 | 4 906                  | 7.1  |
| G        | 48114         | 22  | 230 463    | 18  | 85 446               | 33  | 10 110                 | 14.7 |
| H        | 10054         | 4.5 | 49 084     | 3.8 | 3928                 | 1.5 | 1 418                  | 2.1  |
| I        | 23297         | 11  | 152 246    | 12  | 20 310               | 7.8 | 8 432                  | 12   |
| J        | 3053          | 1.4 | 43 774     | 3.4 | 0                    | 0   | 0                      | 0    |
| K        | 42622         | 19  | 174 110    | 13  | 18 932               | 7.2 | 8 460                  | 12.3 |
| L        | 36            | 0   | 2559       | 0.2 | 0                    | 0   | 0                      | 0    |
| M        | 1546          | 0.7 | 7460       | 0.6 | 289                  | 0.1 | 146                    | 0.2  |
| N        | 12482         | 5.6 | 23 508     | 1.8 | 1 414                | 0.5 | 938                    | 1.4  |
| O        | 17324         | 7.8 | 39 317     | 3   | 4 339                | 1.7 | 1 304                  | 1.9  |
| YHT.     | 222 817       | 100 | 1 301 418  | 100 | 261 995              | 100 | 68 977                 | 100  |

## 7 Yhteenvetoa

Luku 7 on tutkimuksessa saatujen tulosten arviointia ja yhteenvetoa. Luku 7.1 on tutkimuksen yhteenveto ja sisältää tutkimuksen päätulokset sekä johtopäätökset niihin liittyen. Luvussa 7.1 raportoidaan myös lyhyesti vastauksia luvussa 1.2 esitettyihin tutkimustehtäviin sekä arvioidaan tutkimuksen alussa tehtyjen hypoteesien paikkaansa pitävyyttä. Luvussa 7.2 arvioidaan sitä miten yritysten tilinpäätöstietoja tulisi tuottaa jatkossa sekä sitä miten käytössä olevia toimintatapoja tulisi kehittää, jotta saataisiin parannettua toimiala- ja aluekohtaisten tilastojen laatua Tilastokeskuksessa.

### 7.1 Keskeisiä tuloksia

#### **1. Pääongelma: Menetelmien kehittäminen yritysten tilinpäätöstietojen toimipaikkaistamista varten**

Tutkimustehtävänä oli kehittää vartenotettava menetelmä yritysten tilinpäätöstietojen toimipaikkaistamiseen tilastollisten mallien avulla - siis vartenotettava toimipaikkojen tilinpäätöstietojen malliluovuttaja-imputointimenetelmä. Käytettävissä olivat Tilastokeskuksessa tuotettavat yritys- ja toimipaikkatiedot vuodelta 2000, joita ovat yritysten suhdanteet -yksikön tuottama yritys- ja toimipaikkarekisteri (YTR) sekä yritysten rakenteet -yksikön tuottama rakennetilastojen kyselykanta (YKANTA ja TKANTA). Niistä oli löydettävissä lisäinformaatiota, joka parantaa tulosten mukaan toimipaikkojen tilinpäätöstietojen toimialakohtaisten kokonaismäärien estimaattien laatua. Oli löydettävissä sekä otannan että estimoinnin vaiheessa käytettävissä olevaa lisäinformaatiota. Tulosten mukaan kummassakin vaiheessa käytettävä lisäinformaatio paransi toimialakohtaisten jalostusarvon kokonaismäärien estimointia.

Tutkimuksessa käytettävissä olevat apumuuttujat (henkilöstön määrä, palkat ja liikevaihto) korreloivat vahvasti jalostusarvon kanssa. Tämä oli odotettavissa, koska määritelmänsä mukaan jalostusarvo muodostuu tuloslaskelman liikevoittoa/tappiota edeltävästä osasta ja tutkimuksen apumuuttujat ovat (henkilöstön määrä



poislukien) tuloslaskelman liikevoittoa/tappiota edeltäviä eriä. Itse asiassa kaikki Tilastokeskuksessa tuotettavat toimipaikkojen tilinpäätöstiedot ovat laskettavissa toimipaikan laskennallisesta käyttökatteesta tai toimipaikan tilinpäätöksen liitetiedoista. Toimipaikan laskennallista käyttökattetta vastaa yritystasolla tuloslaskelman liikevoittoa/tappiota edeltävä osa. Tämä antaa aiheen olettaa, että tutkimuksessa käytetyt apumuuttujat korreloivat vahvasti myös muiden sellaisten tilinpäätöstietojen kanssa, joiden toimipaikkaistamista tarvitaan. Tämä on yksi syy siihen, että voidaan pitää malliluovuttaja-imputointia tehokkaana myös muiden mahdollisten tilinpäätöstietojen toimipaikkaistamisen tapauksessa.

Lisäinformaation käytön rajoituksena tilastollisissa malleissa oli käytettävissä olevien apumuuttujien voimakas korreloiminen keskenään - siis multikollineaarisuus. Tämä oli odotettavissa, koska tilinpäätöstietojen on todettu olevan riippuvaisia toisistaan. Lisäinformaation saatavuudesta johtuen estimoinnissa voitiin käyttää lineaarista mallia, jonka  $\beta$ -parametrit oli määritetty toimialakohtaisesti tai kaikille toimialoille yhteisenä. Käytössä oli vakavan multikollineaarisuuden välttämisen vuoksi yhden apumuuttujan lineaarinen malli.

Toimipaikan koko, toimiala ja toimipaikkaa vastaavan yrityksen laji voitiin ottaa huomioon toimipaikkaistamisessa käyttämällä niitä osituskriteereinä. Lisäksi voitiin käyttää toimipaikan toimialaa ja toimipaikkaa vastaavan yrityksen lajia kiinnostuksen kohteena oleviin osajoukkoihin jaon kriteereinä (täten tutkimukseen muodostui 29 osajoukkoa).

### ***Mallin ja estimaattorin rooli toimipaikkojen tilinpäätöstietojen estimoinnissa tutkimusaineiston perusteella.***

Käytettävien apumuuttujien selitysvoima ja mallien toimivuus jalostusarvon selittämisessä riippui toimialasta. Paras yksittäinen selittäjä jalostusarvolle on toimialoilla A-D, G ja K liikevaihto, toimialoilla F, M ja X henkilöstön määrä, toimialoilla H-J, N ja O palkat. Toimialalla E oli yksitoimipaikkaisten liikeyritysten toimipaikoille paras yksittäinen selittäjä jalostusarvolle henkilöstön määrä, ja monitoimipaikkaisten yritysten toimipaikoille palkat.

Toimialakohtaiset erot apumuuttujien selitysvoimassa voivat johtua pääoma- ja työvoimavaltaisuuden asteen eroista. Se tarkoittaa sitä, että toimialoja, joilla jalostusarvoa selitti liikevaihto, voidaan pitää pääomavaltaisina. Nämä toimialat olivat alkutuotannon ja jalostuksen toimialoja (lukuun ottamatta toimialoja G ja K). Muita toimialoja voidaan pitää työvoimavaltaisina toimialoina. Nämä olivat (toimialaa E lukuun ottamatta) palvelujen toimialoja. Voidaankin sanoa, että liikevaihto selittää jalostusarvoa, jos toimipaikan elinkeinona on alkutuotanto tai jalostus. Jalostusarvoa selittää henkilöstön määrä tai palkat, jos toimipaikan elinkeinona on palvelut. Kuitenkin toimialat E, G ja K olivat tälle säännölle poikkeuksia. Lisäksi se, että toimialalla E paras selittäjä jalostusarvolle riippui toimipaikkaa vastaavan yrityksen lajista, antaa aiheita olettaa, että myös muilla toimialoilla saattaa paras selittäjä riippua toimipaikkaa vastaavan yrityksen lajista.

Teollisuudessa ja rakentamisessa tutkittiin toimialan vaikutuksen lisäksi toimipaikkaa vastaavan yrityksen lajin huomioon ottamisen vaikutusta estimoinnin tehokkuuteen. Monitoimipaikkaisten yritysten toimipaikoilla estimointi onnistui selvästi paremmin kuin yksitoimipaikkaisten yritysten joukossa toimialoilla C, D ja F. Toimia-

lalla E, eli sähkön-, kaasun- ja vesihuollossa, estimointi onnistui paremmin yksitoimipaikkaisten yritysten joukossa kuin monitoimipaikkaisten yritysten toimipaikkojen joukossa.

Koska mineraalien kaivuun ja rakentamisen toimialoilla toimialakohtaiset GREG -estimaatit ja synteettiset estimaatit olivat lähempänä todellista arvoja ja GREG -estimaatin keskivirhe oli pienempi kuin laskettaessa yhteen yksitoimipaikkaisten liikeyritysten toimipaikkojen ja monitoimipaikkaisten liikeyritysten toimipaikkojen estimaatit, voidaan pitää toimialakohtaista estimaattia parempana toimialoilla C ja F kuin toimipaikkaa vastaavan yrityksen laji- ja toimialakohtaisia estimaatteja. Teollisuudessa ja sähkön-, kaasun- ja vesihuollossa voidaan pitää toimiala- ja yrityksen lajikohtaisia estimaatteja parempina kuin toimialakohtaisia. Tulokset antavat aiheen olettaa, että se mikä kannattaa valita kiinnostuksen kohteena olevaksi osajoukoksi tilinpäätöstietoja toimipaikkaistettaessa, riippuu toimialasta. Ei voida siis sanoa mikä kannattaisi valita kiinnostuksen kohteena olevaksi osajoukoksi toimialoilla A, B ja G-X, koska tutkimuksessa ei ollut käytettävissä jalostusarvon tietoja monitoimipaikkaisten yritysten toimipaikoilta näillä toimialoilla.

Teollisuuden toimialalla tutkittiin lisäksi toimipaikan sijaintiläänin huomioon ottamisen vaikutusta estimointiin. Erilaisia osajoukkokohtaisia jalostusarvon totaalin estimaatteja yhteenlaskettaessa huomattiin, että teollisuuden toimialalla toimialakohtaisen jalostusarvon totaalin estimaatin tarkkuus parani toimipaikan sijaintiläänin huomioon otettaessa vain todellisen arvon ja GREG -estimaatin erotuksen mielessä, ei esimerkiksi GREG -estimaatin keskivirheen tai variaatiokertoimen mielessä. Ei voida siis sanoa, että toimipaikan sijaintiläänin huomioon ottaminen välttämättä parantaisi estimaatteja muillakaan toimialoilla.

Epäsuoran GREG -estimaatin keskivirhe on pienempi kuin suoran, kun kaikille toimialoille yhteisen mallin selityksaste on suurempi kuin toimialakohtaisen mallin selityksaste. Tällöin epäsuorat estimaatit ovat keskivirheen mielessä luotettavimpia kuin suorat. Tätä kutsutaan voiman lainaamiseksi (borrowing strength). Epäsuora synteettinen estimaatti voi olla kuitenkin hyvin harhainen. Näin tapahtui tutkimuksessakin. Voiman lainaamista tapahtui kaikilla muilla toimialoilla paitsi toimialoilla D, F, H ja N. Kuitenkin estimaattien harhankorjaustermit olivat itseisarvoltaan suuria, epäsuora GREG -estimaatti poikkesi suoraa enemmän todellisesta arvosta ja epäsuorat synteettiset estimaatit olivat hyvin harhaisia (antoivat joskus jopa negatiivisen jalostusarvon). Voidaan sanoa suorien estimaattien olevan epäsuoria parempia kaikilla toimialoilla, koska epäsuorat estimaatit tuottavat hyvin harhaisia estimaatteja. (Särndal, 2001, 42-45).

## ***2. Pääongelma. Kehitetyn toimipaikkaistamismenetelmän laadun arviointi teollisuuden ja rakentamisen toimialoilla.***

Teollisuuden ja rakentamisen toimialan toimipaikkojen jalostusarvon imputoimisessa malliluovuttaja-imputointi toimi varsin hyvin. Kuitenkin yhden muuttujan lineaarisen mallin toimivuus riippui toimialasta ja toimipaikkaa vastaavan yrityksen lajista. Monitoimipaikkaisten yritysten toimipaikkojen toimipaikoilla estimointi onnistui selvästi paremmin kuin yksitoimipaikkaisten yritysten joukossa toimialoilla C, D ja F. Toimialalla E tilanne oli päinvastainen, eli estimointi onnistui paremmin yksitoimipaikkaisten yritysten joukossa kuin monitoimipaikkaisten yritysten toimipaikkojen joukossa.

Käytetyn menetelmän laadun arviointia rajoitti kuitenkin se, että sekä Tilastokeskuksen yritysten suhdanteet -yksikön yritys- ja toimipaikkarekisterin että Tilastokeskuksen yritysten rakenteet -yksikön rakennetilastojen kyselykannan toimipaikkatietoja oli imputoitu jo ennen tätä tutkimusta. Tämä tarkoittaa sitä, että tutkimusaineiston sisältämiä muuttuja-arvoja oli arvioitu, eivätkä ne siis aina ole todellisia arvoja. Siksi myös jalostusarvon imputoinnissa käytettyjen apumuuttujien (henkilöstön määrä, liikevaihto ja palkat) havaintoarvoja oli imputoitu jo ennen tutkimusta. Lisäksi myös osa tulosmuuttujan, jalostusarvon, havaintoarvoista oli imputoitu ennen tutkimusta. Tilastokeskuksen yritysten rakenteet -yksikön toimipaikkatietojen (TKANNAN) imputointi on ollut Tilastokeskuksesta saadun tiedon mukaan erityisen laajaa rakentamisen toimialalla, koska rakentamisen alan toimipaikoilta on kysely vain osa niistä tiedoista, joita kysellään mineraalien kaivuun, teollisuuden ja sähkön-, kaasun-, ja vesihuollon toimipaikoilta. Tämä johtaa siihen, että tutkimuksessa laaditun imputointimenetelmän toimivuudesta tehdyt johtopäätökset eivät ole rakentamisen toimialalla niin luotettavia kuin muilla teollisuuden ja rakentamisen toimialoilla. Imputoinnista johtuvat rajoitteet eivät olisi niin suuria, jos imputointi olisi dokumentoitu tarkemmin, koska tällöin imputointi voitaisiin ottaa huomioon analyysejä tehtäessä.

### ***3. Pääongelma. Tutkimuksessa laaditun toimipaikkaistamismenetelmän laadun arviointi muilla kuin teollisuuden ja rakentamisen toimialoilla.***

Malliluovuttaja-imputointi toimi tulosten mukaan varsin hyvin kaikilla toimialoilla yksitoimipaikkaisten yritysten jalostusarvon imputoisessa lukuun ottamatta toimialoja I ja J. Parhaiten jalostusarvon estimointi toimii toimialoilla A, B, F, H, N ja O vapausastekorjatun selityksasteen, RMSE:n ja keskivirheen mielessä. Lähes kaikilla toimialoilla imputoitujen havaintoarvojen kalibroiminen, luvussa 5.5 raportoidulla tavalla, paransi jalostusarvon estimointia.

On kuitenkin huomattava, että tutkimuksella ei saatu tietoa regressioperusteisen imputoinnin toimivuudesta monitoimipaikkaisten liikeyritysten toimipaikkojen joukossa toimialoilla A, B, G-O. Tämä johtui siitä, että teollisuudessa ja rakentamisessa huomattiin yksitoimipaikkaisten yritysten poikkeavan monitoimipaikkaisten yritysten toimipaikoista siinä miten hyvin lineaariset mallit toimivat ja mitkä olivat parhaiten toimivat apumuuttajat. Ei voidakaan sanoa, että kaikilla toimialoilla sama malli toimisi sekä yksitoimipaikkaisten liikeyritysten joukossa että monitoimipaikkaisten liikeyritysten toimipaikkojen joukossa (teollisuudessa ja rakentamisessa niin oli toimialoilla C ja F, mutta ei toimialoilla D ja E). Siksi toimipaikkatietojen kyselyn laajentaminen teollisuuden ja rakentamisen ulkopuolelle, olisi tärkeää. On huomattava myös, että toimialakohtaisiin malleihin perustuva imputointi johti aina parempiin tuloksiin kuin kaikille toimialoille yhteisen mallin käyttäminen. Toimipaikkojen tiedot ovat rekistereissä ja tietokannoissa tällä hetkellä osittain estimoitua tietoa ja siksi niiden käytössä on rajoitteita. Lisäksi rekistereiden ja tietokantojen yhdistämisessä on ongelmia, koska niiden antamat tiedot voivat olla ristiriidassa keskenään.

### ***4. Pääongelma. Käytettävissä olevien toimipaikka- ja yritystietojen eli yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan laadun arvioiminen.***

***Tutkimuksen laatuun vaikutti lähinnä käytetty aineisto.*** Siinä oli monia rajoitteita (luvat 3.1 ja 8). Niistä tärkeimpiä olivat yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan muodostamisen käytän-

töjen erot. Olisikin tarpeen yhdenmukaistaa Tilastokeskuksen sisällä toimintatapoja yritysten ja toimipaikkojen tietojen muodostamisessa. Toinen rajoitus oli se, että tutkimuksessa oli kaksi tapaa määritellä se onko toimipaikkaa vastaava yritys yksi- vai monitoimipaikkainen. Ensimmäinen tapa oli se, että yritys on monitoimipaikkainen, jos yrityksellä on ollut tilikauden aikana useampi kuin yksi toimipaikka. Toinen tapa on se, että yritys on monitoimipaikkainen, jos sillä on tilikauden lopussa vähintään kaksi toimipaikkaa. Nämä määrittelytavat eivät ole aina yhtäpitävät. Jotta toimipaikkojen tiedot tulisivat oikein merkityiksi, tulisi siksi selkeyttää määritelmää sille milloin yritys on monitoimipaikkainen.

Edellisiä rajoituksia vielä suurempi tulosten laatuun vaikuttava seikka oli se, että sekä yritys- ja toimipaikkarekisterissä että rakennetilastojen kyselykannassa oli imputoitu toimipaikkojen tilinpäätöstietoja jo ennen tätä tutkimusta. Imputointi koski tutkimuksessa käytettäviä apumuuttujia (HK2, LVMK ja PALKKAMK) sekä tulosmuuttujaa, eli toimipaikan jalostusarvoa. Myös muita Tilastokeskuksen yritysten rakenteet- yksikön toimipaikkatietoja (TKANNAN tietoja) oli imputoitu, ja siksi imputointia on käytetty myös muiden Tilastokeskuksessa tarvittavien toimipaikkatietojen havaintoarvojen muodostamisessa. Imputointi rajoittaa siis muidenkin mahdollisten yritysten tilinpäätöstietojen (muidenkin mahdollisten tulosmuuttujien) toimipaikkaistamista.

***Imputoinnin dokumentointi osoittautui kuitenkin puutteelliseksi erityisesti siinä mitkä ovat imputoitujen havaintoarvojen luovuttajat (eli imputoidut havaintoarvot) vastanneiden joukossa.*** Toimipaikkarekisterissä ei ollut merkitty sitäkään mitkä havaintoarvot ovat imputoituja. Ja koska estimaateille ei voida edellä mainittujen syiden vuoksi laskea imputointivarianssia, tulevat tutkimuksessa estimaattien varianssit, keski- virheet ja variaatiokertoimet aliestimoiduiksi. Siksi se, että imputointia on käytetty rakennetilastojen kyselykantaan ja yritys- ja toimipaikkarekisteriä muodostettaessa, ja että käytetyistä imputointimenetelmistä on vähän tietoa saatavilla, on ollut suuri rajoitus tutkimuksen teossa. Erityisen laajaa imputoinnin käyttäminen on ollut rakentamisen toimialalla ja sitä koskevat päätelmät ovatkin muita toimialoja koskevia päätelmiä epäluottavampia.

Jatkossa onkin erittäin tärkeää raportoida imputoinnin vaikutus ”liputtamalla” imputoidut havaintoarvot, selostamalla käytetty imputointimenetelmä tarkkaan ja laskemalla varianssin estimaatti asetelmavarianssin ja imputointivarianssin summana. On siis muistettava laskea imputointivarianssi, ja sen laskemiseen tarvitaan imputointimenetelmän tarkkaa raportointia ja imputoitujen havaintoarvojen ”liputtamista” .

***Toimipaikkojen tilinpäätöstietojen estimaattien laatu riippuu käytettävän imputointimenetelmän lisäksi kadon suuruudesta ja tyypistä (eli siitä miten harmillista kato on).*** Koska tällä hetkellä monitoimipaikkaisten yritysten toimipaikkojen tiedot puuttuvat kaikilta muilta paitsi toimialojen C-F toimipaikoilta, voidaan ajatella niiden havaintojen, joista on tietoa poikkeavan niistä, joista ei ole tietoa. Tämä johtuu siitä, että monitoimipaikkaisten ja yksitoimipaikkaisten yritysten toimipaikat poikkeavat toisistaan (luku 4). Tämä tarkoittaa sitä, että kato on harmillista eli vastanneet ja ei-vastanneet poikkeavat tulosmuuttujien suhteen toisistaan. Kato on myös monitoimipaikkaisten yritysten toimipaikkojen kohdalla suurta, koska toimipaikan tiedot kyselyyn vain osalta toimipaikkoja myös toimialoilla C-F.

Kadon harmillisuutta voitaisiin tulevaisuudessa vähentää sillä, että tiedosteltaisiin muuttakin kuin monitoimipaikkaisten yritysten teollisuuden ja rakentamisen toimipaikoilta tärkeimpiä tilinpäätöstietoja (mm. jalostusarvo ja tuotannon bruttoarvo). Tämä tarkoittaisi sitä, että tärkeimmät toimipaikkojen tilinpäätöstiedot pyrittäisiin saamaan kaikilta toimipaikoilta ja osa TKANNAN tiedoista koskisi kaikkia toimialoja. Olisi siis tärkeää laajentaa toimipaikkatietojen kyselyä koskemaan kaikkia monitoimipaikkaisia yrityksiä. Kysyttäviä tietoja voisi olla nykyistä vähemmän. Kyselyn lisäksi voitaisiin TKANTAAN liittää toimipaikkarekisterin tiedot. Toimipaikkojen tietojen saaminen on kuitenkin vaikeaa (luku 1). Tärkeää on, että ainakin yksi monitoimipaikkainen yritys jokaiselta toimialalta vastaisi kyselyyn.

### ***Toimipaikan tilinpäätöstietoihin vaikuttavat tekijät ja tutkimuksen hypoteesit***

Tutkimusaineisto tuki kaikkia tutkimusprosessin alussa kirjallisuuden perusteella asetettuja hypoteeseja, jotka on raportoitu luvussa 1.2. Tutkimus vahvisti hypoteesin 1, jonka mukaan toimipaikan tilinpäätöstietoihin vaikuttavat yleistoloudelliset, toimialakohtaiset, alueelliset, toimiala- ja aluekohtaiset sekä yritys- ja toimipaikkakohtaiset tekijät, koska huomattiin toimipaikkojen tilinpäätöstiedoissa olevan vaihtelun selittyvän hyvin pitkälle toimipaikan toimialalla, koolla ja toimipaikkaa vastaan yrityksen lajilla.

Aluekohtaiset sekä toimiala- ja aluekohtaiset tekijät vaikuttivat toimipaikan toimintaan, koska oli havaittavissa toimipaikkojen keskittymistä sekä kahdelle toimialalle (teollisuus, eli toimiala D, sekä tukku- ja vähittäiskauppa, eli toimiala G) että kahteen lääniin (Etelä- ja Länsi-Suomen läänit). Toimipaikkojen sijainnin keskittymisen vuoksi voivat toimiala- ja aluekohtaiset toimipaikkojen lukumäärät olla joissain lääni- ja toimialakohtaisissa osajoukoissa pieniä. Ja siksi ei kaikilla toimialoilla kannata muodostaa toimiala- ja aluekohtaisia malleja. Läänien välillä ei ollut merkittäviä eroja toimialajakaumissa ja tilinpäätöstiedoissa. Sitä, että Suomessa yritystoiminta on (ainakin tutkimusaineiston toimipaikkojen ja yritysten joukossa) keskittynyt kahteen lääniin, voidaan pitää merkinä alueiden välisestä työnjaosta Suomessa. Toimipaikan tilinpäätöstietoihin vaikuttavista tekijöistä, kuten alueiden välisestä työnjaosta, on puhuttu luvussa 2.2.

Yrityskohtaiset tekijät vaikuttavat tutkimusaineiston perusteella toimipaikkojen toimintaan, sillä toimipaikkaa vastaavan yrityksen lajin (yrityskohtainen tekijä) huomattiin vaikuttavan toimipaikkojen tilinpäätöstietoihin. Toimialasta riippumatta (jokaisella toimialalla) olivat monitoimipaikkaisten yritysten toimipaikat merkitykseltään, toiminnan laajuudeltaan ja kooltaan suurempia kuin yksitoimipaikkaiset yritykset toimipaikan liikevaihdon, henkilöstön määrän ja toimipaikalle estimoitujen palkkojen keskiarvon, mediaanin ja moodin avulla mitattuna. Myös toimipaikan jalostusarvon mukaan (niiden toimipaikkojen joukossa, joiden jalostusarvo oli tiedossa) olivat monitoimipaikkaisten yritysten toimipaikat merkitykseltään, toiminnan laajuudeltaan ja kooltaan suurempia kuin yksitoimipaikkaiset yritykset.

Toimialakohtaiset tekijät vaikuttavat toimipaikan tilinpäätöstietoihin siten, että toimialoilla D, G ja I näytti olevan muita toimialoja selvästi suurempi osuus henkilöstön määrästä, palkoista, liikevaihdosta ja jalostusarvosta sekä toimipaikan toimialan mukaan katsottuna että yrityksen päätoimialan mukaan katsottuna. Kun tutkitaan eri elinkeinojen osuuksia edellä mainituista ominaisuuksista, niin huomataan, että jalostuksen toimialojen osuudet liikevaihdosta, henkilöstön määrästä ja palkoista ovat niiden toimipaikkojen lukumäärän osuutta

selvästi suuremmat (kaikilla muilla toimialoilla paitsi rakentamisessa) kun taas palvelujen toimialoilla ja alkutuotannossa toimipaikkojen lukumäärän osuudet ovat suurempia kuin liikevaihdon, henkilöstön määrän tai palkkojen osuudet. Jalostuksen toimialoja voidaan siis pitää merkittävyydeltään suurempina kuin palvelujen tai alkutuotannon toimialoja (poikkeuksena toimialat G ja I).

Se, että toimialoittaiset osuudet henkilöstön määrässä, palkoissa, liikevaihdossa ja erilaisissa jalostusarvon arvioissa eroavat selvästi toisistaan, voi olla merkki toimialojen välisistä eroista tuottavuudessa, työvoima- ja pääomavaltaisuuden asteissa, tuotettavissa hyödykkeissä sekä tarvittavissa tuotannontekijöissä. Tämä johtuu siitä, että tuottavuutta ja työvoima- ja pääomavaltaisuuden astetta voidaan mitata mm. liikevaihdon ja jalostusarvon avulla. Tutkimusaineisto siis tuki hypoteeseja 2, 3 ja 5. Toimialoittaisien erojen vuoksi voidaan myös sanoa, ettei eri toimialoilla toimivia yrityksiä tai toimipaikkoja voi verrata keskenään, ja tutkimus antoi tukea hypoteesille 4.

## 7.2 Suosituksia toimipaikkojen tilinpäätöstietojen tuottamiselle tulevaisuudessa

*Tutkimuksessa käytettyjen apumuuttujien ja tulosuuttujan (jalostusarvon) imputoinnista johtuvat rajoitteet (luvut 3, 5.5 ja 8) voitaisiin tulevaisuudessa välttää raportoimalla tarkkaan käytetyt imputointimenetelmät, merkitsemällä ("liputtamalla") kaikki imputoidut havaintoarvot ja kirjaamalla jokaisen havaintoyksikön kohdalla luovuttaja tai käytetty imputointimalli.* Imputoinnissa voidaan käyttää mm. toimipaikan tietoa edellisellä tilikaudelta. Se on YTR:ssä yleinen imputointitapa. Luovuttaja on tällöin kyseinen toimipaikka edellisellä tilikaudella ja kyseessä on YTR:n tapa imputoida. Tällöin voitaisiin laskea estimaateille imputointivarianssit, tutkia mm. sitä miten laajaa imputointia on ollut sekä vertailla eri imputointimenetelmiä. *Jos imputointivarianssia ei lasketa, tulee varianssi aliestimoitua.*

*Toinen vaihtoehto on, että kato paikattaisiin muuten kuin imputoimalla. Tällaisia menetelmiä ovat erilaiset uudelleenpainotustekniikat* (mm. painotusluokkaestimaattori, jälkiositus, raking, eksplisiittiset vastaustodennäköisyysmallit ja kalibrointiestimaattori). Nämäkin menetelmät vaativat kuitenkin sen, että varianssien estimaatteja ei lasketa asetelmavarianssien avulla, koska kato lisää aina varianssia. Lisäksi uudelleenpainotustekniikoiden käyttö vaatisi sitä, että toimipaikkatiedoissa ryhdyttäisiin käyttämään vain otostietoja. Jotta uudelleenpainotusmenetelmien avulla lasketut toimialakohtaisten tietojen estimaatit olisivat mahdollisimman hyviä, tulisi tärkeimpiä tilinpäätöstietoja kysellä muiltakin kuin toimialojen C-F toimipaikoilta. (Djerf, 2001, Tilastokeskus, 2002a, 64-67, Särndal ym., 1992).

Imputointi ja uudelleenpainotusmenetelmät voivat joskus tuottaa saman tuloksen (kuten suhdetehosteinen imputointi ja uudelleenpainotus suhde-estimoinnilla.) Imputoinnin etuna on se, että saadaan täydelliset havaintomatriisit. Siksi on suositeltavaa käyttää imputointia toimialakohtaisten tietojen tuotannossa. Toimipaikan tilinpäätöstietojen imputointi voidaan tehdä monella eri tavalla, koska imputointimenetelmiä on olemassa monenlaisia (luku 5.5). Tutkimuksessa käytettiin malliluovuttaja-imputointia. Tulevaisuudessa voitaisiinkin tutkia

sitä miten hyvin muut imputointimenetelmät toimivat toimipaikan tilinpäätöstietojen estimoinnissa. Lisäksi voidaan tutkia sitä kuinka hyvin regressioperusteinen imputointi toimii muiden tilinpäätöstietojen kuin jalostusarvon estimoinnissa. Tuloksia toimialakohtaisista tilinpäätöstiedoista voitaisiin sitten verrata kansantalouden tilinpidon tuloksiin.

Koska toimipaikoista on olemassa monenlaista tietoa, jota voidaan käyttää lisäinformaationa, imputointia ei kannata tehdä käyttämällä menetelmiä, joissa ei lisäinformaatiota käytetä tehokkaasti (mm. lähimmän naapurin menetelmä ja keskiarvoimputointi). Keskiarvoimputointi ei ole suositeltava myöskään siksi, että toimipaikkojen tilinpäätöstiedot ovat vinoja (toimialojen sisälläkin, on siis hankala jakaa toimipaikkoja sisäisesti homogeenisiin ryhmiin). Ja jos havaintojen jakautuminen on vinoa, ei keskiarvoimputointi tuota hyviä tuloksia.

Toimipaikkojen tilinpäätöstietojen estimaattien laatu riippuu käytettävän imputointimenetelmän lisäksi kadon suuruudesta ja tyypistä (eli miten harmillista kato on). Koska tällä hetkellä monitoimipaikkaisten yritysten toimipaikkojen tiedot puuttuvat kaikilta muilta paitsi toimialojen C-F toimipaikoilta, voidaan ajatella niiden havaintojen, joista on tietoa, poikkeavan niistä, joista ei ole tietoa. Tämä johtuu siitä, että monitoimipaikkaisten ja yksitoimipaikkaisten yritysten toimipaikat poikkeavat toisistaan (luku 4). Tämä tarkoittaa sitä, että kato on harmillista eli vastanneet ja ei-vastanneet poikkeavat tulosmuuttujien suhteen toisistaan. Kato on myös monitoimipaikkaisten yritysten teollisuuden ja rakentamisen toimipaikkojen kohdalla suurta, koska toimipaikan tiedot kysellään vain osalta toimipaikkoja myös toimialoilla C- F, eivätkä kaikki kyselyyn tulleet yrityksetkään vastaa (mm. siksi, etteivät ole pitäneet toimipaikkakohtaista kirjanpitoa).

***Kadon harmillisuutta voitaisiin tulevaisuudessa vähentää sillä, että tiedosteltaisiin muiltakin kuin toimialojen C-F monitoimipaikkaisten yritysten toimipaikoilta tärkeimpiä tilinpäätöstietoja*** (mm. jalostusarvo ja tuotannon bruttoarvo). Tämä tarkoittaisi sitä, että TKANNAN tiedoista osa koskisi kaikkia toimialoja ja pyrittäisiin saamaan tärkeimmät tilinpäätöstiedot kaikilta toimipaikoilta. Olisi siis tärkeää laajentaa toimipaikkatietojen kyselyä monitoimipaikkaisten yritysten osalta siten, että toimipaikkatietoja kyseltäisiin myös muiltakin kuin monitoimipaikkaisten yritysten teollisuuden ja rakentamisen toimipaikoilta. Kysyttäviä tietoja voisi olla nykyistä vähemmän. Kyselyn lisäksi voitaisiin TKANTAAN liittää toimipaikkarekisterin tiedot. Toimipaikkojen tietojen saaminen on kuitenkin vaikeaa (luku 1). Siksi päämääränä voisi olla se, että tietoja pyrittäisiin saamaan jokaiselta toimialalta, mutta ei kaikilta toimipaikoilta. Tärkeää on, että ainakin yksi monitoimipaikkainen yritys jokaiselta toimialalta vastaisi kyselyyn.

***Jotta toimipaikkojen tiedot tulisivat oikein merkityiksi, tulisi selkeyttää määritelmää sille milloin yritys on monitoimipaikkainen.*** Jotta toimipaikkojen tilinpäätöstietojen pohjalta voidaan laatia luotettavia toimiala-raportteja, on tiedettävä se millä toimialalla toimipaikka toimii. Ja kuten luvussa 8 todetaan, ei toimialaluokitus 1995 ole täysin kattava, yksikäsitteinen ja poissulkeva (mm. siksi, että toimialaluokituksen käyttö poikkeaa YTR:ssä ja rakennetilastoissa toisistaan). Olisikin yhdenmukaistettava Tilastokeskuksen sisällä toimialaluokituksen käyttötapoja. Tämä voitaisiin tehdä mm. siten, että yritysten rakenteet -yksikkö siirtyisi yritysten suhdanteet -yksikön tekemään toimialan määrittelyyn (ts. YTR:n tiedot toimialoista siirrettäisiin suoraan YKANTAAN ja TKANTAAN).

***Tulisi yhdenmukaistaa tilinpäätöstietojen määrittelytapoja Tilastokeskuksen sisällä***, koska tällä hetkellä yritysten rakenteet -yksikön käyttämät määrittelytavat poikkeavat jossain määrin yritysten suhdanteet -yksikön määrittelytavoista. Tästä on kerrottu luvussa 3.1. Toimintatapojen yhdenmukaistamista tarvitaan, koska silloin yksiköiden tuottamat tiedot olisivat vertailukelpoisempia ja niiden tuottamia rekistereitä voitaisiin paremmin yhdistää.

Toimipaikkojen tilinpäätöstietojen tuotannon parantamiseksi olisi siis tulevaisuudessa yhdenmukaistettava yritysten rakenteet -yksikön ja yritysten suhdanteet -yksikön toimintatapoja ja selkiytettävä määritelmää sille milloin yritys on monitoimipaikkainen, esim. siirtymällä yrityksen lajin määrittämiseen toimipaikkojen lukumäärän mukaan. TKANTAAN voitaisiin tällöin yhdistää YTR :n tiedot. Olisi hyvä kysellä tärkeimpiä tilinpäätöstietoja myös muiden kuin toimialojen C - F toimipaikoilta vähintään otosperusteisesti. Koska toimipaikkojen tiedoissa on aina katoa, on puuttuvat tiedot imputoitava jollain menetelmällä, joka käyttää mahdollisimman hyvin lisäinformaatiota. Tällainen menetelmä on mm. regressioimputointi, jota on käytetty tässä tutkimuksessa. Toimipaikkojen tilinpäätöstietojen jakaumat ovat vinoja ja siksi keskiarvoimputointi ei ole suositeltavaa. Eritäin tärkeää on raportoida imputointi "liputtamalla" imputoidut havaintoarvot, selostamalla käytetty imputointimenetelmä tarkkaan ja laskemalla varianssin estimaatti asetelmavarianssin ja imputointivarianssin summana.

## 8 Tutkimuksen laatuselvitys

Luku 8 on tutkimuksen erillinen laatuselvitys. Laatuselvityksessä arvioidaan erilaisten tutkimustulosten luotettavuuteen vaikuttavia tekijöitä ja arvioidaan käytettyjä rekistereitä ja tietokantoja, joita ovat Tilastokeskuksen yritysten suhdanteet -yksikön muodostama yritys- ja toimipaikkarekisteri (eli YTR) sekä yritysten rakenteet -yksikön muodostama rakennetilastojen kyselykanta (eli YK ANTA ja TKANTA).

***Tutkimuksen rajoitteita/laatua kannattaa tarkastella tutkimalla kaikkien tutkimusprosessin vaiheiden laatua, ei vain tutkimuksen tulosten (lopputuloksen) laatua/rajoitteita.*** Tutkimusprosessin laatua voidaan selvittää arvioimalla sitä, noudattaako tutkimus Suomen Virallisen Tilaston (SVT) laatukriteereitä. Se tehdään laatimalla SVT -julkaisuilla vaadittu SVT -standardin mukainen tuoteseloste tai SVT -tilaston erillinen laatuselvitys. Laatuselvityksen ja tuoteselosteen tarkoituksena on raportoida SVT -laatukriteereiden vaatimat asiat ja vastata siihen miten SVT -laatukriteerit täyttyvät. SVT -laatukriteereitä ovat tilastotietojen relevanssi, tietojen oikeellisuus ja tarkkuus, julkaistujen tietojen ajantasaisuus, tietojen saatavuus ja läpinäkyvyys/selkeys, tietojen vertailukelpoisuus, selkeys ja eheys/yhtenäisyys sekä dokumentointi. Nämä laatukriteerit on määritellyt Tilastokeskuksen (Tilastokeskus, 2002a) käsikirjassa. (Tilastokeskus, 2002a, 15-23, 101-104).

Tilastollisten ominaisuuksien kannalta SVT:n laatukriteereiden noudattaminen tarkoittaa tutkimuksessa käytetyn perusjoukon, käytetyn lisäinformaation ja sen laadun sekä otanta-asetelman raportoinnista. Otanta-asetelman raportointi sisältää käytetyn otantamenetelmän lisäksi kadon suuruuden ja merkityksen arvioinnin, käytetyn kadonkorjausmenetelmän laadun arvioinnin, estimointiproseduurin ja erilaisten virhetyyppien arvi-



oinnin (kehikkovirheet, otantavirheet, mittaus- ja tallennusvirheet, kadosta johtuvat virheet, mallintamisen virheet, eli otannasta ja otannasta riippumattomien virheiden raportointi.) Lisäksi on arvioitava tulosten tuottamisessa käytettyjen estimaattorien teoreettisia ominaisuuksia kuten harhattomuutta, täsmällisyyttä ja tarkentuvuutta. Seuraavassa arvioidaan SVT:n laatuksiteereiden toteutumista tässä tutkimuksessa.

### **1. Laatuksiteeri: Tilastotietojen relevanssi**

#### **Tietosisältö ja käyttötarkoitus**

Kyseessä on selvitys siitä miten alue- ja toimialakohtaisten tilinpäätöstietojen ja toimipaikkojen tilinpäätöstietojen estimaatteja voitaisiin tuottaa tilastollisten mallien avulla (miten voidaan imputoida malliluovuttaja-menetelmällä toimipaikkojen tilinpäätöstietoja). Nämä mallit pohjautuvat Tilastokeskuksessa tällä hetkellä olemassa oleviin tietoihin Suomen yrityksistä ja niiden toimipaikoista. Lisäksi raportoidaan tavat, joilla Tilastokeskuksen yritysten rakenteet- ja yritysten suhdanteet -yksiköt tällä hetkellä tietoja Suomen yrityksistä ja toimipaikoista tuottavat. Tutkimuksessa arvioidaan näiden tuotettujen tietojen laatua ja pyritään kehittämään vaihtoehtoisia, laskennallisesti perusteltuja tapoja tuottaa toimialakohtaisia tietoja.

Tutkimuksessa kehitettyä menetelmää yritysten tilinpäätöstietojen toimipaikkaistamiseen (eli yritystietojen disaggregointiin toimipaikkatasolle) sovellettiin jalostusarvon toimipaikkaistamiseen. Jalostusarvo valittiin, koska se on yksi merkittävimmistä tilinpäätöksistä saatavista tiedoista. Se kuvaa toimipaikan/yrityksen osuutta bruttokansantuotteesta, ja on siten yksi tärkeimmistä toimipaikan/yrityksen merkittävyyttä ja kokoa ilmaisevista tiedoista.

Tutkimuksessa muodostetaan toimialakohtaiset kokonaismäärän estimaatit jalostusarvolle. Niiden avulla saadaan arviot eri toimialojen osuuksille Suomen vuoden 2000 bruttokansantuotteesta ja imputoidaan malliluovuttaja-menetelmällä toimipaikkojen jalostusarvot (muodostetaan toimipaikkojen jalostusarvojen estimaatit). Toimiala- ja aluekohtaisten tilinpäätöstietojen (ja siten mm. toimialaraporttien) laatimisessa voidaan käyttää tutkimuksessa laadittua estimointistrategiaa. Laadittuja jalostusarvojen estimaatteja ja käytettyä estimointistrategiaa voidaan verrata kansantalouden tilinpidon tuottamiin tietoihin ja käytäntöihin. Täsmällistä vertailua ei tutkimuksessa kuitenkaan tehty.

Tutkimuksesta saa kuvan eri toimialojen merkittävydestä Suomessa vuonna 2000 ja voi arvioida sitä kuinka hyvin voidaan Tilastokeskuksessa jo olemassa olevien tietojen avulla estimoida toimipaikkojen tilinpäätöstietoja. Kaiken kaikkiaan tutkimusta voidaan käyttää toimiala- ja aluekohtaisten tilinpäätöstietojen laadun parantamisen apuna. Tässä luvussa arvioidaan myös Tilastokeskuksessa laadittavien yrityksiä ja toimipaikkoja koskevien tietojen luotettavuutta (eli käytettävien aineistojen laatua) ja vertailukelpoisuutta. Tätä tietoa voidaan käyttää yritys- ja toimipaikkarekisterin sekä rakennetilastojen kyselykannan laadun parantamiseen. Tutkimuksessa käytettäviä keskeisiä käsitteitä ovat lisäinformaation käyttö otannassa ja estimoinnissa, malliluovuttaja-imputointi, jalostusarvo, rekisteritutkimus, toimipaikkaistaminen, yleistetty regressioestimointi sekä yritystutkimus.

### **Käytetyt määritelmät ja luokitukset**

Tutkimuksessa on käytetty seuraavia luokituksia. Toimipaikan toimialan määrittelyssä toimialaluokitusta 1995 (eli TOL 95) ja toimipaikkaa vastaavan yrityksen lajia (sitä onko toimipaikkaa vastaava yritys yksi- vai monitoimipaikkainen). Toimialaluokituksista kerrotaan liitteessä 2. Luokitteluperiaatteena tulisi olla se, että havainnot luokitellaan niin, että niitä käyttävät tilastot tuottavat todellisuutta vastaavan, jäsennellyn ja ymmärrettävän kuvan yhteiskunnasta. Luokitusten tulee myös olla relevantteja käyttäjän näkökulmasta sekä olla systemaattisia ja hierarkkisia. Luokituksen luokkien tulee olla yksiselitteisiä, kattavia ja toisensa poissulkevia. Yksiselitteisyydellä tarkoitetaan siitä, että havainnot on luokiteltavissa tiettyyn luokkaan. Kattavuudella tarkoitetaan sitä, että kaikki tapaukset (esim. toimipaikat) voidaan luokitella (luokan "tuntematon" tulee olla mahdollisimman pieni (pyritään siihen ettei tällaisia tapauksia olisi lainkaan), ja poissulkevuus taas tarkoittaa sitä, että yhtä havaintoyksikköä ei voida luokitella kahteen tai useampaan luokkaan. (Tilastokeskus, 2002a).

Toimialaluokitus 1995:llä on, ainakin vuoden 2000 tietojen perusteella, hyvä kattavuus, koska tutkimusaineistossa on vain 11 sellaista toimipaikka, joiden toimiala on tuntematon. Siis sellaisia tapauksia, joita ei ole voitu luokitella, on hyvin vähän (etenkin verrattuna muihin toimialaluokituksen luokkiin). Yritystasolla ei ole yhtään yritystä, jonka toimiala olisi tuntematon. Vuonna 2000 ei toiminut yhtään yritystä tai toimipaikkaa, jonka toimiala olisi L, P tai Q. Siksi tutkimustuloksia ei voi ehkä yleistää toimialoille L, P ja Q.

Toimialaluokitus 1995:n yksiselitteisyys ja poissulkevuus ovat huonompia kuin kattavuus. Se johtuu siitä, että toimipaikan toimiala ja toimipaikkaa vastaavan yrityksen päätoimiala voivat olla Tilastokeskuksen yritysten suhdanteet -yksikön tuottamassa yritys- ja toimipaikkarekisterissä (eli YTR) ja Tilastokeskuksen yritysten rakenteet -yksikön tuottamassa rakennetilastojen kyselykannassa (eli YKANTA ja TKANTA) eriävät. Tämä on merkki siitä, että toimipaikka ja sitä vastaava yritys eivät ole toimialaluokitus 1995:n avulla luokiteltavissa vain yhteen luokkaan, vaan se mihin luokkaan toimipaikkaa tai toimipaikkaa vastaava yritys luokitellaan, riippuu luokittelutavasta. Toimipaikan toimiala ja toimipaikkaa vastaavan yrityksen toimiala eroavat YTR:ssä ja rakennetilastojen kyselykannassa toisistaan hyvin harvoin, koska n. 99 % toimipaikoista ja toimipaikkaa vastaavista yrityksistä on sellaisia, joille YTR:n ja rakennetilastojen kyselykannan tiedot toimipaikan toimialasta ja yrityksen päätoimialasta ovat yhteneväiset. Yksiselitteisyyden ja poissulkevuuden ongelmat ovat siis pieniä, mutta ei voida sanoa, että TOL 95 olisi täysin yksiselitteinen ja poissulkeva.

Koska Tilastokeskuksen yritysten suhdanteet -yksikön tuottaman yritys- ja toimipaikkarekisterin (eli YTR) ja Tilastokeskuksen yritysten rakenteet -yksikön tuottaman rakennetilastojen kyselykannan (eli YKANTA ja TKANTA) tiedot poikkeavat joskus toisistaan, eivät ne ole aina vertailukelpoisia. Olisikin tärkeää tulevaisuudessa yhdenmukaistaa YTR:n ja rakennetilastojen kyselykannan laadinnassa käytettyjä toimintatapoja. Voitaisiin ajatella, että rakennetilastojen kyselykantaan siirrettäisiin suoraan YTR:n tieto toimipaikan toimialasta ja sitä vastaavan yrityksen päätoimialasta.

Toimialaluokitus 1995 perustuu kansainvälisiin säädöksiin, koska se noudattaa pääasiallisesti kansainvälisiä taloudellisen toiminnan luokituksia ja EU:n toimialaluokituksen NACE Rev. 1:en rakennetta ja koodijärjestelmää aina nelinumerotasolle asti.

Toimipaikkaa vastaavan yrityksen lajiin perustuva luokitus on kattava, koska kaikki yritykset on aineistossa luokiteltu yksi- tai monitoimipaikkaiseksi. Kuitenkin se luokitellaanko toimipaikkaa vastaava yritys yksi- vai monitoimipaikkaiseksi, riippuu siitä millaista määritelmää käytetään monitoimipaikkaiselle yritykselle. Aineistossa yritysten jako yksi- ja monitoimipaikkaisiin yrityksiin on tehty sekä toimipaikkojen lukumäärän että yrityksen laji -tiedon mukaan. Vaikka molemmat ovat YTR:stä saatavia tietoja, ne eroavat joskus toisistaan. Ongelma ei ole kuitenkaan suuri. Yrityksen lajitiedon mukaan luokiteltaessa yritys yksitoimipaikkainen, kun sillä on ollut tilikauden aikana vain yksi toimipaikka. Yrityksen toimipaikkojen lukumäärän mukaan luokiteltaessa yritys on yksitoimipaikkainen, kun yrityksellä on tilikauden loppuessa yksi toimipaikka. On sellaisia yrityksen lajiin mukaan monitoimipaikkaisia yrityksiä, jotka ovat toimipaikkojen lukumäärän mukaan yksitoimipaikkaisia, mutta niitä on kaikilla toimialoilla enintään n. 2 % toimipaikoista. Voidaan siis sanoa, että yrityksen lajiin perustuva luokitus ei ole täysin yksiselitteinen ja poissulkeva, mutta on täysin kattava. Tulevaisuudessa voitaisiinkin siirtyä tapaan määritellä yritys yksi- tai monitoimipaikkaiseksi toimipaikkojen lukumäärän mukaan. Tämä on Tilastokeskuksesta saadun tiedon mukaan luotettavampi tieto, kuin YTR:n tieto yrityksen lajista.

Tutkimuksessa on käytetty alueellisia luokituksia toimipaikkojen ja niitä vastaavien yritysten sijainnin määrittelyyn. Nämä käytettävät luokitukset ovat lääni- ja maakuntajako. Lisäksi osittamisessa on käytetty yritysten jakoa suuriin ja PK-yrityksiin. Näiden luokitusten tiedot löytyvät vain yritys- ja toimipaikkarekisteristä, ja niiden poissulkevuudesta ja yksiselitteisyydestä ei voi tehdä tarkkaa arviota. Tietoja ei kuitenkaan puuttunut miltyään toimipaikalta, joten näitä luokituksia voidaan pitää kattavana.

## **2. Laatuksiteeri: Tietojen oikeellisuus ja tarkkuus**

### **Tavoite- ja kehikkoperusjoukon vastaavuus**

Tutkimus on rekisteritutkimus ja tietolähteinä ovat Tilastokeskuksen yritysten suhdanteet -yksikön muodostama yritys- ja toimipaikkarekisteri sekä Tilastokeskuksen yritysten rakenteet -yksikön muodostama rakennetilastojen kyselykanta vuodelta 2000 (luku 3). Tutkimuksen kohteena ovat liikeyritykset ja niiden toimipaikat. Tutkimukseen valittiin ne liikeyritykset, jotka ovat molemmissa rekisterissä. Näiden yritysten omistamat toimipaikat muodostavat tutkimuksen havaintoyksiköt. Niitä oli vuonna 2002 yhteensä 241 987.

**Tutkimuksen kohdeperusjoukko** muodostuu Suomessa vuonna 2000 toimineiden liikeyritysten toimipaikoista yritysrekisterin asettamin ehdoin (luku 3.1). **Kehikkoperusjoukon** muodostavat yritys- ja toimipaikkarekisterissä ja rakennetilastojen kyselykannassa olevat toimipaikat, jotka ovat liikeyritysten toimipaikkoja yritysrekisterin asettamin ehdoin. Kehikkoperusjoukko ei yleensä vastaa täysin kohdeperusjoukkoa kehikkoperusjoukossa olevien virheiden vuoksi. **Kehikkoperusjoukon virheitä on kolme eri tyyppiä: ylipeittoa, ali-peittoa ja kaksinkertaista listausta.** Näitä sanotaan ns. peittovirheiksi ja niitä on kuvattu luvussa 5.4. Mahdolliset virheet kehikkoperusjoukossa huonontavat tutkimustuloksia ja saattavat aiheuttaa harhaisia estimaatteja. Tärkeimmät syyt peittovirheille ovat tässä tutkimuksessa tilikauden aikana tapahtuneet muutokset yrityksissä ja toimipaikoissa. Näitä ovat fuusiot, konkurssit, uusien yritysten ja toimipaikkojen syntyminen ja vanhojen toimipaikkojen sekä yritysten toiminnan loppumiset. Peittovirheitä aiheuttava myös mittaus- ja prosessointivirheet. (Särndal ym., 1992, 13-14, Djerf, 2001, 11-12, Lehtonen ja Pahkinen, 2004).

Alipeitto ja ylipeitto johtuvat useimmiten siitä, ettei tilikauden aikana tapahtunutta fuusiota, konkurssia, syntyntä yritystä/toimipaikkaa tai vanhan toimipaikan/yrityksen toiminnan loppumista ole päivitetty rekisteriin. Tämä on voinut johtua päivituksen viivästyisestä, tai siitä ettei tapahtunutta muutosta ole ilmoitettu. Konkursin tehneet yritykset, toimintansa lopettaneet yritykset ja toimipaikat aiheuttavat rekisteriin ylipeittoa, jos niitä ei ole poistettu rekisteristä. Fuusiot aiheuttavat ylipeittoa, jos fuusioituneita yrityksiä ei ole poistettu rekisteristä ja muodostettuun tilalle uutta yritystä rekisteriin. Tällöin saattavat fuusioituneen yrityksen tiedot tulla kirjatuksi monta kertaa rekisteriin (vanhojen yritysten tietoina) ja kyse on kaksinkertaisesta listauksesta. Kaksinkertaisesta listauksesta on kyse myös, jos yritys on antanut yhden toimipaikan tietoina useamman toimipaikan tiedot.

Kehikkoperusjoukossa saattaa olla mukana sellaisia yrityksiä, jotka eivät ole todellisuudessa Tilastokeskuksen käyttämän määritelmän mukaisia liikeyrityksiä, mutta ovat sellaisiksi merkitty mittaus- tai prosessointivirheen vuoksi. Tämä on ylipeittoa. Alipeittoa aiheutuu siitä, jos yritys ei ole ilmoittanut kaikkien toimipaikkojen tietoja, tai rekisteriin ei ole päivitetty tilikauden aikana syntyntä toimipaikkaa tai yritystä, koska yritys ei ole ilmoittanut perustamisesta tai uuden toimipaikan syntymisestä. Syynä voi olla myös päivitysviive. Alipeitto aiheutuu myös siitä, että jokin yritys saattaa olla todellisuudessa Tilastokeskuksen käyttämän määritelmän mukainen liikeyritys, mutta on kirjattu yritykseksi, joka ei ole liikeyritys mittaus- tai prosessointivirheen vuoksi. (Särndal ym. 1992, 13-14).

**Mittaus- ja prosessointivirheet sekä käytettävien muuttujien laatu.** Tutkimuksessa käytettävien muuttujien havaintoarvoissa voi olla virheitä johtuen tiedon keruussa (yritys tai toimipaikka on antanut väärän tiedon) tai tallennuksessa (näppäilyvirheet) tapahtuneista virheistä eli mittaus- ja prosessointivirheistä. Havaintoarvoissa voi esiintyä virheitä myös siksi, että Tilastokeskuksen yritysten suhdanteet -yksikön yritys- ja toimipaikkarekisteriin sekä Tilastokeskuksen yritysten rakenteet -yksikön rakennetilastojen kyselykantaan vuodelta 2000 ollaan saatettu tehdä muutoksia tutkimuksen aloittamisen jälkeen. Tutkimuksessa on siirretty niiden yksitoimipaikkaisten yritysten tiedot, joiden toimipaikan toimiala on sama kuin yrityksen toimiala ja joiden tilinpäätöstietoja ei ole TKANNASSA, toimipaikan tiedoksi. Nämä havaintoarvot ovat virheellisiä, jos yritys ei tosiasiassa olekaan yksitoimipaikkainen (eli toimipaikkojen lukumäärä ei olekaan yksi, vaikka yritysrekisterissä on ilmoitettu sen olevan). Käytettävät apumuuttujat eli keskimääräinen henkilöstön määrä, toimipaikalle estimoidut palkat ja toimipaikan liikevaihto ovat hyviä muuttujia jalostusarvon selittäjänä saatujen estimointitulosten mukaan. Niiden ja jalostusarvon havaintoarvoissa on kuitenkin rajoitteita, jotka on raportoitu seuraavana.

### **Aineistossa oleva kato ja sen paikkaamisen käytettyjen menetelmien arviointi**

Käytetyssä aineistossa on vain vähän puuttuvia havaintoarvoja (vuonna 2000 vain 5:ltä toimipaikalta puuttui jalostusarvon tietoja ja YKANNASSA sekä YTR:ssä ei katoa ollut lainkaan). Kadon vähäisyys johtui kuitenkin siitä, että sekä Tilastokeskuksen yritysten suhdanteet -yksikön yritys- ja toimipaikkarekisterin että Tilastokeskuksen yritysten rakenteet -yksikön rakennetilastojen kyselykannan toimipaikkatietoja oli imputoitu jo ennen tätä tutkimusta. Tämä tarkoittaa sitä, että tutkimusaineiston sisältämiä muuttuja-arvoja oli arvioitu, eivätkä ne siis aina ole todellisia arvoja. Siksi myös jalostusarvon imputoinnissa käytettyjen apumuuttujien (henkilöstön määrä, liikevaihto ja palkat) havaintoarvoja oli imputoitu jo ennen tutkimusta. Lisäksi myös osa tulomuuttu-

jan, jalostusarvon, havaintoarvoista oli imputoitu ennen tutkimusta. Tilastokeskuksen yritysten rakenteet -yksikön toimipaikkatietojen (TKANNAN) imputoiminen on ollut Tilastokeskuksesta saadun tiedon mukaan erityisen laajaa rakentamisen toimialalla, koska rakentamisen alan toimipaikoilta on kysely vain osa niistä tiedoista, joita kysellään mineraalien kaivuun, teollisuuden ja sähkön-, kaasun-, ja vesihuollon toimipaikoilta. Tämä johtaa siihen, että tutkimuksessa laaditun imputointimenetelmän toimivuudesta tehdyt johtopäätökset eivät ole rakentamisen toimialalla niin luotettavia kuin muilla teollisuuden ja rakentamisen toimialoilla. Imputoinnista johtuvat rajoitteet eivät olisi niin suuria, jos imputointi olisi dokumentoitu tarkemmin, koska tällöin imputointi voitaisiin ottaa huomioon analyysejä tehtäessä.

***Imputoinnin dokumentointi osoittautui puutteelliseksi erityisesti siinä mitkä ovat imputoitujen havaintoarvojen luovuttajat eli imputoidut havaintoarvot vastanneiden joukossa.*** Toimipaikkarekisterissä ei ollut merkitty sitäkään mitkä havaintoarvot ovat imputoituja. Toimipaikkarekisterin tietojen (mm. apumuuttajien havaintoarvot) kohdalla ei siis tiedetä sitä mikä on vastanneiden joukko ja milloin puuttuvat tieto on korvattu imputoidulla havaintoarvolla. Ja koska estimaateille ei voida edellä mainittujen syiden vuoksi laskea imputointivarianssia, tulevat tutkimuksessa estimaattien varianssit ja variaatiokertoimet aliestimoiduiksi. Siksi se, että imputointia on käytetty rakennetilastojen kyselykantaa ja yritys- ja toimipaikkarekisteriä muodostettaessa sekä se, että käytetyistä imputointimenetelmistä on vähän tietoa saatavilla, on ollut suuri rajoitus tutkimuksen teossa.

***Tutkimuksessa käytettävien aineistojen havaintoarvojen imputointi vaikuttaa varianssiestimaattien lisäksi myös laadittuihin toimialakohtaisiin estimaatteihin.*** Koska tutkimuksessa toimipaikkojen jalostusarvoja ja apumuuttajien havaintoarvoja on imputoitu, saattaa jalostusarvo korreloida enemmän käytettävien apumuuttajien kanssa kuin jos imputointia ei olisi käytetty. Ja koska apumuuttajien havaintoarvoja on imputoitu, saattavat apumuuttajat korreloida keskenään enemmän kuin jos imputointia ei olisi. Tämä johtaa siihen, että multikollinearisuus saattaa johtua imputoinnista ja käytettyjen mallien selitysasteet saattavat olla todellista suurempia.

**Tutkimuksen menetelmäkuvaus.** Tutkimuksen tietolähteinä ovat Tilastokeskuksen yritysten suhdanteet -yksikön yritys- ja toimipaikkarekisteri (YTR) sekä Tilastokeskuksen yritysten rakenteet -yksikön rakennetilastojen kyselykanta (YKANTA ja TKANTA) vuodelta 2000. Lisäinformaatio on YTR:ssä ja tulosuuttujan havaintoarvot (eli jalostusarvon tiedot) ovat YKANNASSA ja TKANNASSA. Käytetyistä aineistoista kerrotaan luvussa 3. Tutkimuksessa käytetyt estimaattorit, otanta-asetelmat, estimointistrategiat, imputointi ja diagnostisten tarkastelujen tavat on raportoitu luvussa 5. Käytetyt muuttujat on kuvattu luvussa 3.2. Niiden laatua on arvioitu luvuissa 7 ja 8. Koska kaikille toimipaikoille ei ole tiedossa jalostusarvoa, niin joudutaan toimialakohtaisten tietojen tuotannossa käyttämään imputointia. Imputointi tuo oman epävarmuutensa tuloksiin (luku 5.5).

**3. Laatukriteeri: Julkaistujen tietojen ajantasaisuus ja oikea-aikaisuus.** Tutkimuksessa on käytetty tietoa vuonna 2000 toimineista yrityksistä ja toimipaikoista. Tällöin toimialan määrittelyssä käytettiin toimialaluokitusta 1995 (eli TOL 95). Vuodesta 2003 on siirrytty käyttämään toimialaluokitusta 2002 (eli TOL 2002). Se on päivitetty versio TOL 95:stä ja TOL 95:n mukainen toimialan määrittely on muutettavissa TOL 2002 mu-

kaiseksi. TOL 2002 ja TOL 95:n vertailukelpoisuudesta on kerrottu liitteessä 2. Muuten yritysten ja toimipaikkojen tietojen tuotantotavat ovat säilyneet ennallaan. Rahayksikkönä tilinpäätöstietojen tuotannossa on nykyään euro, mutta tutkimuksessa on käytetty markkatietoja (YTR:ssä tiedot ovat tuhansina markkoina ja rakennetilastoistakin markkatiedot olivat helposti saatavissa). Markkatiedot on kuitenkin helppo muuttaa euroiksi. Voidaan sanoa, että tutkimuksessa tuotettu menetelmä toimialakohtaisten tietojen tuottamiseen ja arviot menetelmän toimivuudesta ovat sovellettavissa tulevienkin vuosien aineistoihinkin.

**4. Laatuksiteeri: Tietojen saatavuus ja läpinäkyvyys/selkeys.** Tutkimuksessa käytetyt aineistot ovat saatavissa yritysten rakenteet -yksiköstä. Ne sisältävät estimoidut toimipaikkojen jalostusarvot. Tässä tutkimusraportissa on raportoitu saadut tulokset.

**5. Laatuksiteeri: Tietojen vertailukelpoisuus.** Tutkimuksen vertailukelpoisuuteen vaikuttavat tietojen keräämistavat ja käytettävät luokitukset, käsitteet ja määritelmät. Käytetyistä luokituksista toimialaluokitus 1995 (eli TOL 95) ei välttämättä ole yksikäsitteinen ja poissulkeva, koska se mikä on toimipaikan toimiala, riippuu siitä käytetäänkö yritysten rakenteet -yksikön tapaa vai yritysten suhdanteet -yksikön tapaa käyttää toimialaluokitus 1995:tä. Tästä on kerrottu kohdassa käytettävät määritelmät ja luokitukset. Myös yrityksen lajin kohdalla ei luokitus ole välttämättä yksikäsitteinen ja poissulkeva. Kaiken kaikkiaan yritysten suhdanteet -yksikön käytännöt poikkeavat yritysten rakenteet -yksikön käytännöistä, eivätkä YTR ja rakennetilastojen kyselykanta ole aina yhdenmukaisia (esim. saman toimipaikan henkilöstön määrä voi olla eri YTR:ssä kuin TKANNASA). YTR:n ja rakennetilastojen kyselykannan välisistä eroista on kerrottu luvussa 3.1. Tutkimus noudattaa yritysten rakenteet -yksikön tapaa määrittellä toimiala, jos toimipaikka on TKANNASSA, mutta muutoin YTR:n tapaa määrittellä toimiala. Lisäinformaatio on YTR:stä poimittu ja noudattaa siksi yritysten suhdanteet -yksikön käytäntöjä. Jalostusarvon määrittäminen noudattaa yritysten rakenteet -yksikön käytäntöjä. Tulosten vertailukelpoisuudesta yli ajan on kerrottu kohdassa julkaistujen tietojen ajantasaisuus ja oikea-aikaisuus.

**6. Laatuksiteeri: Tuotettujen tietojen eheys/yhtenäisyys.** Jos tutkimusaineistoa verrataan rakennetilastojen kyselykantaan, niin eivät tutkimusaineisto ja rakennetilastojen kyselykanta ole yhdenmukaiset, koska tutkimusaineisto sisältää vain ne yritykset, jotka ovat YTR:n mukaan liikeyrityksiä (222 817 vuonna 2000). Aineisto ei siis sisällä kaikkia YKANNAN yrityksiä ja TKANNAN toimipaikkoja. Tutkimusaineisto ei myöskään sisällä kaikkia toimipaikkarekisterin liiketoimipaikkoja, eivätkä YTR ja tutkimusaineisto ole siinä mielessä yhdenmukaisia. Tutkimuksen tarkoituksena onkin ollut selvittää sitä miten yritysten rakenteet -yksikön tuottamat tiedot ja yritysten suhdanteet -yksikön tuottamat tiedot voitaisiin yhdistää. Huomattiin, että samoista asioista kertovat tiedot eivät aina ole yhdenmukaisia (kuten toimialan määrittely YTR:n ja YKANNAN mukaan). Tulisiikin yhdenmukaistaa yritysten rakenteet -yksikön ja yritysten suhdanteet -yksikön käyttämiä käytäntöjä esim. siirtymällä samoihin määritelmiin henkilöstön määrän laskemisessa.

**7. Laatuksiteeri: Dokumentointi** Tutkimuksen tulokset ja niiden analysointi ovat tutkimusraportin luvuissa 4-6. Yhteenveto tuloksista on luvussa 7.1, laatuselvitys luvussa 8 ja suosituksia toimipaikkojen tietojen tuotantolle jatkossa on luvussa 7.2. Luvut 7 ja 8 muodostavat oikeastaan tutkimuksen ns. metatiedon.

## Lähteet

Alhola Kari, Koivikko Aarne, Rätty Päivi, ja Tuominen Olavi (2001). *Tilinpäätöksen sisältö ja toteutus*. Porvoo: WS Bookwell Oy

Autio Arvo (1995). *Taloustieto*. Porvoo: WSOY

Djerf Kari (2001). Tilastokeskus: Tutkimuksia 231. *Properties of Some Estimators Under Unit Nonresponse*. Helsinki: Hakapaino Oy

Estevao V., Hidiroglou, and Särndal C.E.(1995). *Methodological Principles for a Generalized Estimation System at Statistics Canada*. Journal of Official Statistics, Vol 11. No 2 1995, s. 181-204

Euroopan yhteisöjen komissio(2002). *Alustava esitys komission suositus, pienten ja keski suurten yritysten määritelmästä annetun suosituksen 96/280/EY muuttamisesta*.

[http://europa.eu.int/comm/enterprise/consultations/sme\\_definition/consultation2/153\\_sme\\_definiton\\_25\\_6\\_2002\\_pp1\\_10.fi.pdf](http://europa.eu.int/comm/enterprise/consultations/sme_definition/consultation2/153_sme_definiton_25_6_2002_pp1_10.fi.pdf). 17.9.2002

Finnvera Oy (2002). *Finnvera*. [www.finnvera.fi](http://www.finnvera.fi). 3.10.2002

Heikkilä Tarja (1998). *Tilastollinen tutkimus*. Helsinki: Oy Edita Ab

Holopainen Martti, ja Pulkkinen Pekka (1995). *Tilastolliset menetelmät – perusteet 1-2*. painos. Porvoo: WSOY

Karttunen Tero (2002). *Henkilöstömääriä kuvaavat muuttujat, tilikausi ja liikevaihto yritysrekisterissä*. Helsinki: Tilastokeskus, yritysten suhdanteet -yksikkö, muistio 29.5.2002, (julkaisematon)

Kinkki Seppo (1995). *Yritysavain – Yrityksen perustoiminnot (3. painos)*. Porvoo: WSOY

Kinnunen Juha, Leppiniemi Jarmo, Martikainen Teppo, Virtanen Kalervo (2000). *Yrityksen taloushallinnon perusteet*. Keuruu: Otavan Kirjapaino Oy

Koivuniemi, Päivi (2002). *Alue- ja toimialatilastojen kehittäminen rakennetilastoissa – hankkeen(YA52K) esisuunnitelman loppuraportti*. Helsinki: Tilastokeskus, yritysten rakenteet -yksikkö, muistio 8.5.2002 (julkaisematon).

Laitinen Erkki K., ja Luotonen Eero (1996). *Mitä tilinpäätös kertoo*. Vantaa: Tummauoren Kirjapaino Oy

Lehtonen Risto (1996). *Interviewer Attitudes and Unit Nonresponse in Two Different Interviewing Schemes*. In Laaksonen Seppo (eds.) Statistics Finland: Research Reports 219. International Perspectives on Nonresponse. Statistics Finland: Helsinki.

Lehtonen Risto, and Pahkinen Erkki (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Second Edition. Chichester : Wiley. Painossa tutkimuksen teko hetkellä, lähteenä on luettu käsikirjoitusta.

Leppiniemi Jarmo, ja Leppiniemi Raili (2000). *Tilinpäätöksen tulkinta (3.uudistettu painos)*. Porvoo: WS Bookwell Oy

Maliranta Mika (1996). Tilastokeskus, tutkimuksia 215. *Suomen tehdateollisuuden tuottavuus – kansainvälinen tasoverailu*. Helsinki: Oy Edita AB

Pahkinen Erkki, ja Lehtonen Risto (1989). *Otanta-asetelmat ja tilastollinen analyysi*. Helsinki: Gaudeamus.

Rantala Juha (1999). *Toimipaikkojen rekrytointikäyttäytyminen: analyysi toimipaikkojen rekrytoinneista 1997*. Helsinki: Työministeriö

Särndal Carl-Erik (2001). *Design- Based methodologies for Domain Estimation*. In Lehtonen Risto, and Djerf, Kari (eds.) *Lecture Notes on Estimation for Population Domains and Small Areas*. Helsinki: Statistics Finland, Reviews 2001/ 5, 5 – 49.

Särndal C.E., Swensson B., and Wretman J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag

Taloustieto Oy (1994). *Taloussanasto – Yritys- ja kansantaloutta suomeksi, ruotsiksi, englanniksi, saksaksi ja ranskaksi* (6. uudistettu painos). Tampere: Tammer-paino Oy

Tilastokeskus(2002 a). Käsikirjoja 43: *Laatua tilastoissa*. Helsinki: Hakapaino Oy

Tilastokeskus (2002b). *Teollisuuden ja rakentamisen alue- ja toimialatilasto 2000*.  
[http://www.tilastokeskus.fi/yr/rake\\_tejulk00\\_etusivu.html](http://www.tilastokeskus.fi/yr/rake_tejulk00_etusivu.html). 1.7.2002

Tilastokeskus (2002 c). Käsikirjoja 4: *Toimialaluokitus 2002*. Helsinki: Hakapaino Oy

Tilastokeskus (1993). Käsikirjoja 4: *Toimialaluokitus 1995*. Helsinki: Painatuskeskus Oy

Tilastokeskus (2002 d). *Yritysrekisteri*. <http://www.tilastokeskus.fi/yritysrekisteri>. 4.7.2002

Tilastokeskus (2002 e). *Yritykset 2002:3. Yritystasku*. Paino-Center Oy

Virkkala Seija (1987). *Monitoimipaikkaiset teollisuusyritykset ja alueiden välinen työnjako Suomessa*. Vammala: Vammalan Kirjapaino Oy

Wonnacott, and Wonnacott (1979). *Econometrics* (Second Edition). Singapore: John Wiley & Sons

Yritystutkimusneuvottelukunta (1999). *Yritystutkimuksen tilinpäätösanalyysi* (6. uudistettu painos). Helsinki: Gaudeamus



## Liite 1. Yrityksen toiminta

*Tutkimuksen havaintoyksiköitä ovat toimipaikat, jotka ovat yritysten yksiköitä. Tästä johtuen on kyse yritystutkimuksesta ja on tarpeen selvittää yrityksen toimintaa. Yritys on talousyksikkö, joka ostaa erilaisia tuotannontekijöitä, yhdistää ne tuotantoprosessissa ja myy tuottamansa tuotteet talousyksiköille. Yritys siis tuottaa markkinoille hyödykkeitä sekä pyrkii tällä voittoon. Tämän määritelmän mukaan yrityksiä ovat luonnolliset henkilöt, jotka harjoittavat yritystoimintaa joko omalla nimellään tai rekisteröidyllä toiminimellä. Tähän ryhmään kuuluvat ammatin- ja liikkeenharjoittajat, useimmat maatilatalouden harjoittajat sekä useimmat liikettä harjoittavat luonnolliset henkilöt. Yritystoimintaa voivat harjoittaa lisäksi oikeushenkilöt (mm. osakeyhtiöt, osuuskunnat, keskinäiset yhtiöt, säästöpankit, taloudelliset yhdistykset ja kommandiittiyhtiöt), julkiset rahoituslaitokset, valtion liikelaitokset, kuntien yleisöä palvelevat liikelaitokset ja kuntayhtymät, jotka toimivat kuten liikelaitokset. Tilastoinnissa yritys on usein tilastoyksikkö, koska kaikista sen taloustoimista on saatavissa täydelliset tiedot. Näitä ovat mm. tilinpäätös-, rahoitus- ja tuotantotiedot. (Autio, 1995, Tilastokeskus, 2002 c, 23).*

***Talousyksiköllä** tarkoitetaan yhtä tai useampaa henkilöä, jotka harjoittavat yhdessä taloudellista toimintaa. Taloudellista toimintaa ovat tuotanto ja kulutus. Talousyksiköitä (ja kansantalouden sektoreitakin) ovat kotitaloudet, yritykset ja julkinen valta (eli kunnat, kuntaliitot, valtio sekä seurakunnat). **Kansantalous** koostuu tietyn valtion sisällä toimivien talousyksiköiden toiminnasta. Yritysten toiminta on siis osa kansantaloutta. Yritykset vaikuttavat muihin talousyksiköihin ja muiden talousyksiköiden toiminta vaikuttaa niihin, siis yritykset ja muut talousyksiköt ovat vuorovaikutuksessa keskenään. (Taloustieto Oy, 1994, Autio, 1995, ja Kinkki, 1995).*

***Hyödykkeillä** tarkoitetaan tavaroita tai palveluksia, joilla tyydytetään erilaisia tarpeita. Ne ovat tuotantoprosessin tuloksia ja ne jaetaan kulutus- ja tuotantohyödykkeisiin. Kulutushyödykkeiden avulla tyydytetään kotitalouksien tarpeita. Ne ovat joko aineellisia tai aineettomia. Aineellisia kulutushyödykkeitä ovat kertatavarat (kuten ruoka ja juoma), puolikestotavarat (esim. vaatteet) ja kestokulutushyödykkeet (esim. autot, polkupyörät, pesukoneet ja huonekalut). Aineettomia kulutushyödykkeitä ovat erilaiset palvelut, (mm. opettajan, sairaanhoitajan, hierojan, kosmetologin, sosiaalityöntekijän, poliisin ja oikeuslaitoksen palvelut). Tuotantohyödykkeillä tarkoitetaan välineitä, joilla tuotetaan uusia kulutus- ja tuotantohyödykkeitä. Tuotantohyödykkeitä ovat mm. rakennukset, koneet, kuljetusvälineet, ravintolat, pankkiautomaatit ja raaka-aineet. Yrityksen näkökulmasta hyödykkeitä kutsutaan **suoritteiksi**. Yritys siis tuottaa suoritteita ja asiakas ostaa hyödykkeitä. (Autio, 1995).*

***Tuotannontekijöiksi** kutsutaan tekijöitä, joita tarvitaan tuotantoon (hyödykkeiden tuottamiseen) ja hyödykkeiden saattamiseksi ihmisten käyttöön. **Perustuotannontekijöitä** ovat luonnonvarat, työ ja pääoma. Luonnonvaroja ovat mm. viljelykelpoinen maa sekä metsä- ja malmivarat ja energiavarat (vesivoima, puu, turve, öljy, kivihiili ja uraani). Yleensä hyödykkeiden tuottamiseksi tarvittavat raaka-aineet pohjautuvat luonnonvaroihin. Työ (eli työvoima) muuntaa luonnonvarat käyttökelpoisiksi hyödykkeiksi. Pääomaa jaetaan kahteen osaan: reaali-pääomaan ja henkiseen pääomaan. Reaali-pääomalla tarkoitetaan rakennuksia, koneita, laitteita,*

toimitiloja, raaka-aineita sekä rahapääomaa. Henkisellä pääomalla tarkoitetaan erilaisia työntekijöiden ja yrittäjien ominaisuuksia, kuten koulutusta. (Taloustieto Oy, 1994, Kinkki 1995, 15-17).

Tuotantoon tarvitaan kaikkia tuotannontekijöitä. Yritys tarvitsee toimialastaan ja toiminnan laajuudestaan riippuen eri määrän tuotannontekijöitä. Esim. kiinteistönvälitystoimisto tarvitsee lähinnä vain työvoimaa (eli henkilökunnan, toimitilat ja konttoritarvikkeita), mutta teollisuusyritys tarvitsee suuret tilat ja paljon erilaisia koneita sekä enemmän henkilökuntaa kuin kiinteistönvälitysyrittäjä. Siksi teollisuusyritys joutuu tekemään suurempia hankintoja kuin kiinteistönvälitysyrittäjä ja sitomaan paljon pääomaa sekä ennen hyödykkeidensä valmistuksen aloittamista että myös varsinaisen prosessin aikana. (Kinkki 1995, 16-17).

Tilastokeskuksesta saadun tiedon mukaan joissakin määritelmässä tuotannontekijöihin luetaan vain työ ja pääoma. Täten on tuotannontekijät määritelty tunnetuimmassa tuotantofunktiossa, joka on Cobb-Douglasin tuotantofunktio. Työtä ja pääomaa sanotaan myös primaarituotannontekijöiksi ja niiden lisäksi varsinkin yritys- ja toimipaikkatasolla tärkeitä ovat ns. **välituotteet**. Ne ovat tuotannontekijöitä, jotka on tuotettu toisissa yrityksissä/toimipaikoissa tai toisilla toimialoilla. Suuri osa yrityksen kustannuksista on välituotteiden hankintaa ja ne ovatkin kustannusosuuksilla mitattuna yrityksen tärkein tuotannontekijä. Tuotannontekijöiden määritelmiä on siis monenlaisia.

Yritys saa toiminnastaan tuloja ja niiden hankkimisesta koituu aina sille menoja. **Tulalla** tarkoitetaan suoritteiden myynnistä saatavia rahamääräisiä etuja (Taloustieto Oy, 1994). Kaikille yrityksille ja yritystyyppisille yksiköille koituu maksusitoumuksia suoritteiden tuottamisesta, koska yritykset joutuvat ostamaan tuotannontekijöitä. Niitä kutsutaan Laitisen ja Luotosen (1996) mukaan **menoiksi**. Menoja on kahdenlaisia, pitkävaikutteisia ja lyhytvaikutteisia. **Pitkävaikutteiset menot** ovat pitkävaikutteisten tuotannontekijöiden hankkimisesta aiheutuvia kuluja ja ne ovat toiselta nimeltään investointeja. Niitä ovat mm. koneiden hankinnat. **Lyhytvaikutteisilla menoilla** taas hankitaan lyhytvaikutteisia tuotannontekijöitä. Niitä ovat mm. palkat, raaka-aineet ja vuokrat.

Yritys tähtää **voittoon**. Tämä tarkoittaa sitä, että yritys pyrkii siihen, että sen tulot olisivat menoja suuremmat myymällä tuottamiensa erilaisia tavaroita ja palveluksia. Jos yritys tuottaa voittoa, voi se kattaa menonsa ainakin pääosin myynnistään saamalla tuloilla. Yrityksen toiminnalla on toki muitakin tavoitteita. Niitä ovat taloudelliset tavoitteet (mm. kannattavuus-, taloudellisuus- ja rahoitustavoitteet), kasvu- ja markkinaosuustavoitteet, suoritteiden ja toiminnan laatuun liittyvät tavoitteet, henkilöstötavoitteet sekä ympäristönsuojeluun liittyviä tavoitteita. Yrityksen toiminnan erilaiset tavoitteet saattavat olla toistensa kanssa ristiriidassa. (Kinnunen ym., 2000, 245-247).

Yritys ei synny ilman, että sen perustaja(t) muodostavat ns. yrityksen **liikeidean**. Liikeidea sisältää: 1. **Suoritteet**, joita yritys tuottaa, 2. keskeiset **kohderyhmät**, eli ne kuluttajat, joille suoritteet pyritään myymään ja joihin markkinointi suunnataan (esim. nuoret, lapset), 3. **keskeiset toimintatavat** (se miten yrityksen tuottamat suoritteet pyritään saamaan myydyiksi) ja 4. sen yrityskuvan eli **imagon**, jolla yritys tuottaa kohderyhmälle

markkinoi. Imagolla tarkoitetaan sitä mielikuvaa, mikä asiakkailta ja suurella yleisöllä on yrityksen toiminnasta. (Kinkki, 1995).

Yritystoimintaan liittyy aina ns. *liikeriski*, silloinkin kun yrityksen liikeidea on mahdollisimman hyvä. Liikeriskillä tarkoitetaan yritystoiminnan epäonnistumisen mahdollisuutta, eli sitä etteivät myyntitulot riitä kattamaan kaikkia toiminnasta aiheutuvia menoja. Se johtuu siitä ettei yrittäjä voi koskaan varmasti tietää, että asiakkaat ostaisivat yrityksen valmistamia hyödykkeitä niin paljon ja siihen hintaan, että yritys saisi menojensa suuremmat myyntitulot. Liikeriski liittyy tulorahoituksen epävarmuuteen ja suuruuden vaihteluun. (Kinkki, 1995, Kinnunen ym., 2000, 148-153).

**Tulorahoituksella** tarkoitetaan yrityksen toiminnan rahoittamista yrityksen toiminnastaan saamien tulojen avulla. Se ei aina riitä yrityksen toiminnan rahoittamiseen, etenkin, jos yritys on juuri perustettu tai se on tehnyt investointeja. Investoinnit ovatkin usein suuri menoerä, eikä yrityksen myynnistä saama voitto (tulorahoitus) aina kata niitä. Niinpä yrityksen toimintaa rahoitetaan myös pääomarahoituksella. **Pääomarahoituksella** tarkoitetaan toiminnan rahoittamista vieraan tai/ja oman pääoman avulla. Oma tai/ja vierasta pääomaa hankitaan yleensä silloin kun yritys perustetaan, sitä laajennetaan tai sen toiminta on väliaikaisesti kannattamatonta. Oma ja vieras pääoma ovat vaihtoehtoja ja on myös olemassa erilaisia oman sekä vieraan pääoman väli- muotoja. Yleensä yrityksen pääomarahoitusta muodostuu sekä vieraasta että omasta pääomasta. Pääoma- ja tulorahoituksen lisäksi yritys voi rahoittaa toimintaansa **avustuksilla**. Esimerkiksi juuri perustettu yritys saa avustuksia valtiolta ja kehitysalueella toimiva yritys Finnveralta (luku 2.2). (Kinkki, 1995, Autio, 1995).

Rahoitukseen liittyviä päätöksiä ovat yrityksen johdon päätökset investointikohteista ja siitä miten investoinnit rahoitetaan. Yrityksen tulee varmistaa pitkän aikavälin rahoituskyvyn (välttämättömien investointien rahoituskyvyn) lisäksi lyhyen aikavälin rahoituskyky eli se, että laskut pystytään maksamaan ajallaan ja velallisilta saatavat maksut tulevat ajallaan. Tätä varten yritys tekee rahoituksen suunnittelua laatimalla ns. rahoitusstrategian. (Kinnunen ym., 2000, 173-176).

**Vieras pääoma** tarkoittaa yleensä velkaa ja siihen liittyy yleensä sovittu takaisinmaksuaika ja korko. Tämä velka voi olla lainaa luottolaitokselta, pankilta tai yhteisöiltä tai se voi olla ns. ostovelkaa, jolla yritys voi hankkia tuotannon tekijöitä tietyllä maksuajalla (ts. yritys ei maksa koko maksua heti). Yleensä vieraan pääoman antajat eli rahoittajat vaativat pääomalleen ns. vakuuden eli pantin, kiinnityksen tai takauksen. Vieras pääoma aiheuttaa yritykselle kuluja, jotka ovat riippumattomia yrityksen toiminnan kannattavuudesta, koska lyhennyksen määrä ja aikataulu sekä korko on sovittu etukäteen. Jos yritys ei selviä koroista ja lyhennyksistä ajallaan, voi se ajautua jopa konkurssiin. Voi käydä niinkin, että vakuudet loppuvat ennen kuin vieraan pääoman tarve. (Kinkki, 1995, Autio, 1995).

**Oman pääoman** avulla toiminnan rahoittaminen tarkoittaa sitä, että yrityksen omistajat sijoittavat rahaa yrityksen toimintaan ilman sovittua korkoa tai takaisinmaksuohjelmaa. Se, joka luovuttaa yritykselle omaa pääomaa, on myös yrityksen omistaja. Oma pääoma pysyy yrityksessä koko yrityksen olemassaoloajan. Omistaja voi myydä sijoituksensa toiselle henkilölle, jolloin yrityksellä vaihtuu omistaja, mutta oman pääoman määrä ei

muutu. Vastineeksi sijoituksilleen omistajat odottavat, että sijoitetulle pääomalle maksetaan osinkoa yrityksen kannattavuuden mukaan. Osingolla tarkoitetaan korkoa, jonka yritykseen sijoittaneet saavat sijoittamalleen pääomalle ja korvausta riskistä, joka yritystoimintaan aina liittyy. Osingon määrää tai maksuajankohtaa ei ole sovittu etukäteen, vaan se on riippuvainen yrityksen toiminnan kannattavuudesta. Osinko maksetaan vain silloin kun yrityksen toiminta on kannattavaa. (Kinkki 1995, Autio, 1995).

**Yrityksen taloudellisia toimintaedellytyksiä** ovat kannattavuus, maksuvalmius ja vakavaraisuus. Edellä mainittuja asioita kutsutaan myös tilinpäätösanalyysin taloudellisiksi toimintaedellytyksiksi. **Maksuvalmiudella** tarkoitetaan yrityksen kykyä selviytyä maksuista niiden erääntyessä. Siitä käytetään myös nimitystä likviditeetti, ja se kuvaa yrityksen lyhyen aikavälin toiminnanrahoituskykyä. Sitä kuvaa tilinpäätöksessä tase. Se ilmenee yrityksen kassavaroina, välittömästi nostettavissa olevina pankkitalletuksina ja helposti rahaksi muutettavana varallisuutena (esim. saamiset, vaihto-omaisuus ja pörssinoteeratut arvopaperit). Maksuvalmiutta on myös mahdollisuus saada velkaa, ts. vierasta pääomaa. **Vakavaraisuus** on yrityksen pitkän aikavälin toiminnanrahoituskyky ja toiselta nimeltään soliditeetti. Sillä tarkoitetaan vieraan ja oman pääoman suhdetta, eli yrityksen omavaraisuusastetta. Jos yrityksen oma pääoma on vierasta pääomaa suurempi, on yritys vakavarainen. Tilinpäätöksestä saadaan suoraan ja tunnuslukuja laskemalla tietoja yrityksen rahoituskyvystä ja kannattavuudesta. (Leppiniemi ja Leppiniemi, 2000, Taloustieto Oy, 1994).

**Kannattavuudella** tarkoitetaan yrityksen tulojen ja tuotantokustannusten välisiä eroja. Yrityksen toiminta on kannattavaa, jos se tuottaa voittoa, eli tulot ovat menoja suuremmat. Tällöin yrityksen myyntituotot riittävät sen tarvitsemien tuotantotekijöiden ostoon ja rahoituskulujen (mm. lainojen lyhennysten, korkojen ja ostovelkojen) maksuun. Jos tuotantokustannukset yrityksessä ovat suuremmat kuin tulot, tuottaa yritys tappiota. Tappion jatkuessa yrityksellä ei enää ole varaa toimintansa ylläpitoon mm. maksaa palkkoja, tehdä investointeja ja hankkia raaka-aineita. (Kinnunen ym., 2000, 246, Kinkki, 1995 ja Taloustieto Oy, 1994).

Kannattavuuteen vaikuttavat ns. katetuottoajattelussa myyntimäärä- ja hinta sekä muuttuvat ja kiinteät yksikkökustannukset (eli kustannukset tuotettavaa suoritetta kohti). Maliranta (1996,15) pitää myös tuottavuutta kannattavuuteen vaikuttavana tekijänä. Kannattavuuden ylläpitämiseksi tulee yrityksen pitää asiakkaansa tyytyväisenä, jotta asiakkaat jatkaisivat yrityksen tuottamien hyödykkeiden ostamista (asiakkaat kun määräävät yrityksen myyntituottojen määrän määrittämällä myynnin määrää ja myytävien suoritteiden hintoja) ja pitää tuotantokustannukset kurissa. Yrityksen on osattava arvioida oikein myyntituottojen, tuotettujen suoritteiden määrän ja tuotantokustannusten väliset riippuvuudet. Jotta yrityksen tuotanto olisi kannattavaa, tulee yrityksen saada kaikki tuottamansa tavarat tai palvelut myydyksi ja saada niistä vielä sellainen hinta, että ne ylittävät tuotantokustannukset. Tämä ehto täyttyy kun yritys tuottaa tuotteita ja palveluita sen verran kun on kysyntää voittoa tuottavaan hintaan. (Kinnunen ym. 2000, 246-254, 282, Maliranta, 1996, 15).

Kannattavuutta mitataan ns. absoluuttisella mittarilla eli tilikauden tuloksella (jos se on positiivinen, on yritys kannattava) ja ns. suhteellisen kannattavuuden mittareilla. Suhteellisen kannattavuuden mittareita ovat joita (tilikauden tulos)/ pääoma = pääoman tuottoaste tai (tilikauden tulos)/myyntituotot = voittoprosentti. Suoritteiden tuottamisesta aiheutuneita kustannuksia ei pystytä tarkkaan määrittämään, koska osa kustannuksista on

ns. yleiskustannuksia tai välillisiä kustannuksia. Yleiskustannukset ovat kustannuksia, joita ei voi kohdistaa jollekin suoritteelle, vaan niiden avulla tuotetaan kaikkia yrityksen tuottamia suoritteita (mm. tilojen vuokrat, koneiden ja laitteiden hankinnat, tutkimus- ja kehittämistoiminnan kustannukset ja johdon palkat). Ne ovat yleensä ns. kiinteitä kustannuksia, eli riippumattomia tuotettujen suoritteiden määrästä. Lisäksi suoritteen tuottamisesta aiheutuu ns. välittömiä kustannuksia. Ne ovat kustannuksia, jotka voidaan kohdistaa johonkin suoritteeseen, ja ne riippuvat tuotettujen suoritteiden määrästä, eli ovat ns. muuttuvia kustannuksia. (Kinnunen ym. 2000, 246, 256-262).

Riittävän kannattavuustason ylläpito on välttämätöntä yrityksen toiminnan jatkumisen takaamiseksi. Kinnunen ym. (2000, 245-247) pitääkin yrityksen tavoitetta toimia kannattavasti ns. vaatimustasotyypisenä tavoitteena, eli ehtona, sille, että muut yrityksen tavoitteet toteutuisivat. Vastuu riittävästä kannattavuustasosta ja muistakin yrityksen tavoitteista kuuluu yrityksen johdolle, ja joskus johdon palkkakin määräytyy yrityksen tavoitteiden saavuttamisen mukaan.

Yrityksen toiminta saattaa olla väliaikaisesti kannattamatonta, eli yrityksen tuotantokustannukset voivat olla väliaikaisesti tuloja suuremmat. Tämä tapahtuu yleensä silloin kun yritys perustetaan tai yritys laajentaa toimintaansa. Pitkällä aikavälillä yrityksen toiminnan tulee kuitenkin olla kannattavaa. Lyhyellä aikavälillä yrityksen on oltava ns. maksuvalmiina, eli pystyttävä selviytymään joka päivä sille tulevista erilaisista maksuista, mm. veroista, velkojen koroista, lyhennyksistä ja tuotannon tekijöiden ostoista. Maksuvalmius voidaan turvata pääomarahoituksella. Yritys kestää sen avulla tappiollista toimintaa jonkin aikaa. Tappionsietokyky on riippuvainen yrityksen pääomansaantimahdollisuuksista, tappioiden suuruudesta ja yrityksen mahdollisuudesta rationalisoida eli järkiperäistää toimintaansa.

Kannattavuuden lisäksi yrityksen ja sen toimipaikkojen tavoitteena on mahdollisimman suuri tuottavuus ja tuottavuuden jatkuva nousu. **Kokonaistuottavuus** ilmaisee miten tehokkaasti tuotannon tekijäpanokset (tuotannon tekijöitä, joita tarvitaan tuotannossa) on yhdistetty tuotoksen aikaansaamiseksi (Taloustieto Oy, 1994). Sillä tarkoitetaan Malirannan (1996, 15) mukaan "...tuotannon määrän ja sen tuottamiseen käytettyjen panosten suhdetta.". Yrityksen tuotanto koostuu erilaisista tuotteista. Eri tuotteita ei voi verrata suoraan toisiinsa, ja siksi tuotannon määrä ja sen tuottamiseen käytetyt panokset ilmaistaan usein hintojen avulla. Tällainen tuottavuuden määritelmä kertoo siitä miten tehokkaasti tuotantokustannukset on käytetty, eli miten paljon sillä on saatu suoritteita aikaan. Panoksena voidaan käyttää myös mm. aikaa, henkilömäärää ja työvoimakustannuksia, eli ns. työpanoksia (Autio 1995). Lisäksi panokset voivat olla ns. fyysiseen pääomaan (tähän on Malirannan kuvauksessa lisätty luonnonvarat) tai välituotteisiin liittyviä (Maliranta 1996, 37). Työtä ja pääomaa kutsutaan ns. primaaripanoksiksi ja muita ns. sekundaaripanoksiksi (Maliranta 1996, 37). Kokonaistuotantoa laskettaessa on otettava mukaan kaikki edellä mainitut panokset (Maliranta 1996, 37). Tuottavuutta voidaan tarkastella mm. kansantalouden tasolla, toimialoittain, yrityskohtaisesti ja yrityksen sisällä toimintoittain tai yksilökohtaisesti. Elintaso on riippuvainen koko kansakunnan tuottavuuden tasosta. Tähän vaikuttaa tuottavuus yrityksissä, toimipaikoilla ja eri toimialoilla (Maliranta 1996, 9).

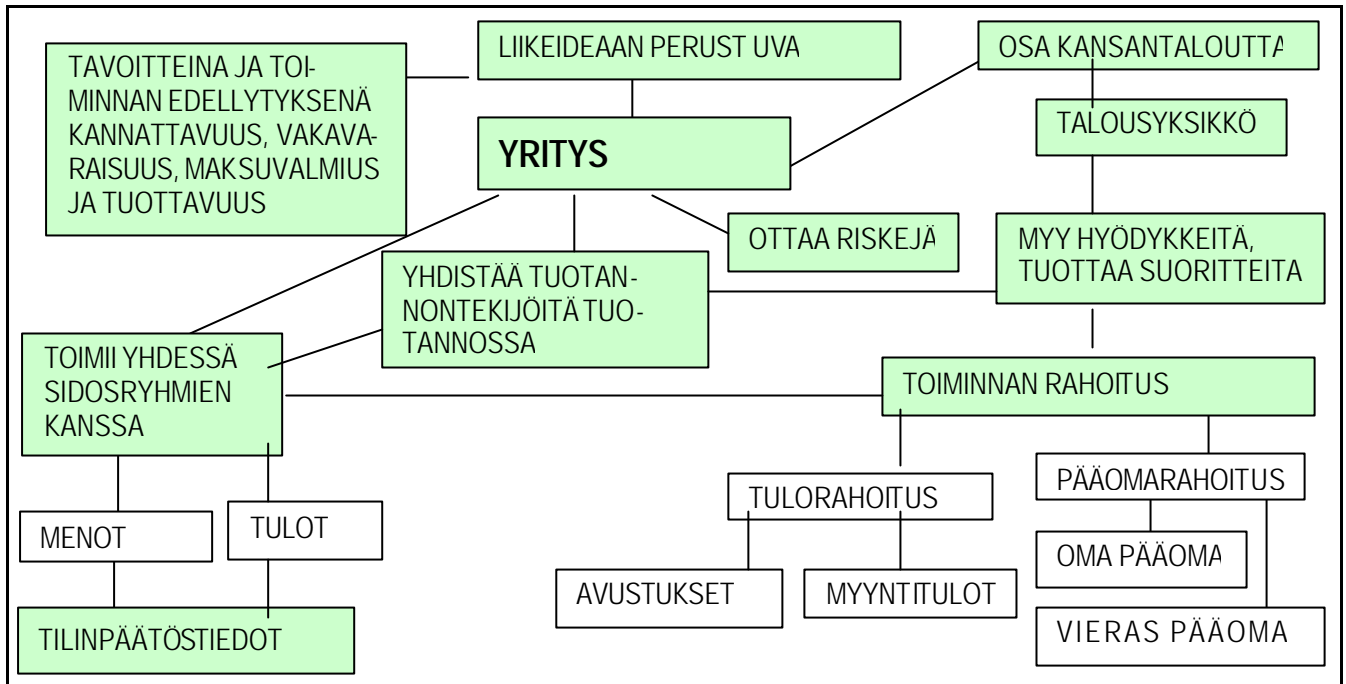
**Työn tuottavuudella** puolestaan tarkoitetaan tuotannon ja työpanoksien suhdetta (esim. tuotannon määrän ja tehtyjen työtuntien suhdetta). Sitä mitataan tilinpäätöksen luvuin jakamalla jalostusarvo henkilöstön määrällä, tehdyillä työtunneilla tai henkilöstökuluilla (Leena Storgårdsin esitelmä Tilastokeskuksen yritysaineistot aluerakenteen- ja kehityksen tutkimuksessa -seminaarissa 17.10.2002). Tutkimuksessa tuotettavien toimipaikkojen jalostusarvojen ja henkilöstökulujen estimaattien avulla voidaan arvioida toimipaikkojen ja toimialojen työn tuottavuutta.

**Yrityksen sidosryhmät** ovat talousyksiköitä, jotka ovat välittömässä vuorovaikutuksessa yrityksen kanssa, ja joilla on omat tavoitteensa liittyen yrityksen kanssa tekemisissä olemiseen. Sidosryhmät muodostuvat yksityisistä ihmisistä, toisista yrityksistä sekä yhteisöistä. Sidosryhmiin eivät kuulu sellaiset ryhmät, jotka eivät ole välittömästi tekemisissä yrityksen kanssa. Tällaisia ovat mm. toisten toimialojen yritykset ja sellaiset kunnat, joiden alueella yritys ei toimi. Sidosryhmiä ovat **omistajat, johto, työntekijät (eli henkilöstö), asiakkaat, tavarantoimittajat, rahoittajat, julkinen valta, kilpailijat, erilaiset etujärjestöt** (etenkin työntekijä- ja työnantajajärjestöt kuten AKAVA ja SAK), **oppilaitokset, tutkimuslaitokset ja media**. (Kinkki, 1995).

**Itse asiassa yrityksen toiminta muodostuu yrityksen ja sen sidosryhmien yhteistyöstä.** Sidosryhmien ja yrityksen yhteistyö perustuu sille, että kumpikin osapuoli hyötyy yhteistoiminnasta. Yritys saa sidosryhmiltään erilaisia panoksia kuten pääomia, työvoimaa, tietoa ja taitoa. Sidosryhmät taas saavat yritykseltä vastikkeita, kuten maksuja, tuotteita ja palveluita, palkkaa, veroja, korkoja ja osinkoja. Yhteistoiminta perustuu sidosryhmän ja yrityksen välille tehtyyn sopimukseen (esim. työntekijän ja työnantajan väliseen työsopimukseen). Jos jokin sidosryhmä sanoutuu irti yhteistyöstä, voi se merkitä jopa yrityksen konkurssia. Yrityksen sidosryhmät toimivat sekä yrityksen sisällä että sen ulkopuolella ja vaikuttavat yrityksen toimintaan. Yritykset taas puolestaan vaikuttavat sidosryhmiensä toimintaan, mutta kaikki sidosryhmät eivät ole riippuvaisia vain yhdestä yrityksestä (esim. rahoittajat). Niiden rooli on yrityksen toimintaa tukeva tai rajoittava. Mitään sidosryhmää ei voida yrityksen kannalta pitää toista tärkeämpänä, vaan niitä kaikkia tarvitaan yrityksen toiminnan ylläpitämisen. Sidosryhmien ja yrityksen tilinpäätöksen osan (kululajipohjaisen tuloslaskelman) välisistä yhteisistä on liitteessä 3. (Kinkki, 1995).

**Yrityksen henkilöstö** koostuu Tilastokeskuksen (2002b) mukaan yrityksessä tai yrityksen lukuun työskentelevistä työntekijöistä ja toimihenkilöistä, joille yritys maksaa ennakkipäätöksen alaista palkkaa ja sen perusteella määräytyviä maksuja. Lisäksi henkilöstöön kuuluvat yrityksessä toimivat itsenäiset yrittäjät. **Itsenäisiä yrittäjiä** ovat yksityisyrittäjät, avoimien ja kommandiittiyritysten osakkaat sekä näiden avustavat perheenjäsenet, jotka työskentelevät yrityksessä päivittäin. Osakeyhtiöiden tai vastaavan osakkaat ja heidän perheenjäsenensä luetaan heidän suorittamansa työn luonteen perusteella joko toimihenkilöihin tai työntekijöihin. **Työntekijöitä** ovat välittömästi valmistusprosessiin osallistuvat henkilöt, kuten apuosastojen työntekijät, huolto- ja kunnossapitotyöntekijät, rakennusosastojen työntekijät, koneenkäyttäjät, lämmittäjät, siivoojat. Myös työnjohtajat, jotka osallistuvat varsinaiseen valmistustyöhön, luetaan työntekijöiksi. **Toimihenkilöitä** ovat muut kuin edellä mainittu henkilöstö.

Kuvio 1 kuvaa yrityksen toimintaa. Sen on kirjoittaja laatinut kirjallisuuden perusteella (Kinkki, 1995, Kinnunen ym., 2000, Tilastokeskus, 1993).



Kuvio 1. Yrityksen toiminta liitteen 1 käsittein.

## Liite 2. Toimialaluokitus TOL95 ja yrityksen jako yksiköihin

Tutkimuksessa laadittavaa yritysten tilinpäätöstietojen toimipaikkaistamismenetelmää on tarkoitus käyttää toimiala- ja aluetilastojen laatimisessa. Lisäksi tutkimuksessa käytetään otanta-asetelmassa toimialaluokitus 1995:n (TOL 95) mukaisia pääluokkia yhtenä osituskriteerinä ja kiinnostuksen kohteena olevana osajoukkoina estimoinnissa. Näistä syistä onkin tarpeen määritellä toimiala ja toimialaluokitus 1995. Koska toimipaikka on tutkimuksen havaintoyksikkö, on tarpeen määritellä toimipaikka ja raportoida sen eroja muihin yrityksen yksiköihin nähden.

**Toimiala** on ala tai sektori, jolla jokin yhteisö (mm. yritys) toimii (Tilastokeskus, 1993). Esim. metalliteollisuus ja vähittäiskauppa ovat toimialoja. Ilman työnjakoa ei olisi toimialojakaan. Voidaankin ajatella, että yritysten ja toimipaikkojen erikoistuminen erilaiseen tuotantoon (eri toimialoilla toimiminen) on osoitus sekä yritysten välisestä että yritysten sisäisestä työnjaosta (luku 2.2).

**Toimialaluokitus** on yleinen järjestelmä tuotantotoimintaa harjoittavien yksiköiden (talousyksiköiden), kuten yritysten, yritysten kaltaisten yksiköiden, toimipaikkojen ja toimialayksiköiden, tuotantotoiminnan lajin luokittelua varten. Näitä kutsutaan toimialaluokiksi tai toimialoiksi. Samaan toimialaan kuuluvien talousyksiköiden tuotantotoiminnot ovat samankaltaisia. Tuotantotoiminnot ovat samankaltaisia, kun tuotanto on tuotettujen hyödykkeiden, tuotantopanosten ja tuotantoprosessinsa mukaan samankaltainen. Käytännössä toimialaluokitusta käytetään luokiteltaessa toimipaikkoja eri toimialaluokkiin. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

Yritys- ja toimipaikkarekisterissä sekä rakennetilastojen kyselykannassa käytettiin vuodesta 1995 vuoteen 2002 ns. **toimialaluokitusta 95 eli TOL 95**. TOL tarkoittaa kansallista toimialaluokitusta ja vuosiluku vuotta, jolloin toimialaluokitus on vahvistettu tai otettu käyttöön. TOL 95:tä käytettiin eri tilastoissa ja hallinnollisissa järjestelmissä julkisella sektorilla, yksityisen sektorin tietojärjestelmissä ja tutkimuksissa. Vuoden 2000 vuositilastoissa, jota tutkimus käsittelee, on siis käytetty TOL 95:tä. Vuodesta 2003 on käytössä ollut **toimialaluokitus 2002 eli TOL 2002**. TOL 2002 on päivitetty painos TOL 95:stä. Nämä toimialaluokitukset ovat tilastotoimen yleisimmin käyttämiä ja laajimmalle levinneitä luokitusstandardeja (mm. Tilastokeskuksessa käytössä). (Tilastokeskus, 1993, Tilastokeskus, 2002 c, 3-6).

Toimialaluokituksessa (sekä TOL 95 että TOL 2002) on viisi tarkkustasoa, joista ei kaikissa tutkimuksissa käytetä kaikkia, eivätkä kaikki toimialat aina ole mukana. Tässä tutkimuksessa käytetään ensimmäisen tason toimialaluokitusta ja mukana ovat kaikki muut toimialat kuin L, P ja Q, koska millään yrityksellä tai toimipaikalla ei vuonna 2000 ollut toimialana L, P tai Q (ainakin tutkimusaineistossa). Sekä TOL 95 että TOL 2002 sisältävät 17 varsinaista pääluokkaa, joiden tunnuksat ovat A-Q. Lisäksi mukana on luokka X (toimiala tuntematon). TOL 2002 ja TOL 95 sisältävät samat toimialojen pääluokat lukuun ottamatta pääluokkaa C, joka on TOL 2002:ssa nimetty kaivostoiminnaksi ja louhinnaksi. Muuten TOL 2002 ei poikkea TOL 95:stä. Toimialaluokitusten TOL 95 ja TOL 2002 välisistä eroista kerrotaan Tilastokeskuksen julkaisussa (Tilastokeskus, 2002 c). Se sisältää myös liitteenä 1 TOL 95 :n ja TOL 2002 :n väliset muutokset. Liitteen 1 avulla TOL 95:n mukaisen toimialatiedon voi muuttaa TOL 2002:en mukaiseksi. (Tilastokeskus 1993, Tilastokeskus, 2002c, 27-54, Tilastokeskus, 2002 e, 21).

Kaksi pääluokkaa (C ja D) on TOL 95:ssä ja TOL 2002:ssa jaettu osaluokkiin, joiden tunnus on kaksi kirjainta (CA, CB jne.). Näitä kaksikirjaimisia luokkia on 16. Kaksinumeroisella on TOL 95:ssä 61 luokkaa ja TOL 2002:ssa 63 luokkaa. Nämä kaksinumeroiset luokat on jaettu kolminumeroisiin (kolminumerotaso) ja edelleen nelinumeroisiin (nelinumerotaso) luokkiin. Kaksinumeroisesta lähtien luokitus on hierarkinen siten, että esim. viidennen tason numerokoodi kertoo suoraan mihin luokkaan kyseinen luokka kuuluu neljännellä, kolmannella tai toisella tasolla. Kirjaintunnuksen ensimmäinen kirjain taas ilmoittaa sen mihin pääluokkaan luokka kuuluu ja toinen kirjain sen mihin välitasoon kyseinen luokka kuuluu. Toimialaluokituksen 1995:n toimialojen määritelmät kaikilla viidellä tarkkuustasolla ovat Tilastokeskuksen (1993) käsikirjassa. (Tilastokeskus, 1993, Tilastokeskus, 2002 e).

Yhteiskunnassa tapahtuvaa tuotantotoimintaa on perinteisesti kuvattu jakamalla tuotanto kolmeen ryhmään: **alkutuotantoon, jalostukseen ja palveluihin**. Tätä sanotaan myös elinkeinotoiminnan kolmijaoksi ja sen osia **elinkeinoiksi**. Alkutuotannon tehtävänä on erottaa raaka-aineet luonnosta ja valmistaa hyödykkeistä luonnonvaroja ja maan hedelmällisyyttä hyväksi käyttäen. Alkutuotantoa ovat maa- ja metsätalous, metsästyys ja kalastus sekä kaivannaisteollisuus. Jalostus on tuotannonhaara, jossa alkutuotannosta saadut hyödykkeet saatetaan sellaiseen muotoon, että ne ovat valmiit otettavaksi käyttöön. Jalostukseen kuuluvat teollisuus, käsityö, rakennustoiminta sekä voimatalous. Palveluilla tarkoitetaan aineettomia hyödykkeitä ja aineettomien hyödykkeitä tuottava yksikkö harjoittaa palvelualan toimintaa. (Autio, 1995).



Alkutuotantoa ovat TOL 95:n ja TOL 2002:en pääluokat A-C (maatalous, riistatalous, metsätalous, kalatalous ja mineraalien kaivuu/kaivostoiminta ja louhinta), jalostusta TOL 95:n pääluokat D-F (teollisuus, sähkön-, kaasun- ja vesihuolto sekä rakentaminen) ja palveluita TOL 95 pääluokat G-Q (muut). Joskus pääluokka C sijoitetaan myös jalostukseen. Palvelut jaetaan Tilastokeskuksessa vielä kauppaan ja kuljetukseen (päätoimialat G-I), liike-elämän palveluihin (J-K) ja yhteiskunnallisiin palveluihin (L-P). (Tilastokeskus 1993, Tilastokeskus, 2002 c, 26).

TOL 95 noudattaa pääasiassa kansainvälisiä taloudellisen toiminnan luokituksia (mm. ISIC) ja EU:n toimialaluokituksen NACE Rev. 1:en rakennetta ja koodijärjestelmää aina nelinumerotasolle asti. Itse asiassa viisinumero-taso (eli tarkin taso) on kansallinen lisäys EU:n toimialaluokitukseen. Tarkimman tason luokkia muodostettaessa on pyritty välttämään tapauksia, joissa luokkaan kuuluisi vähemmän kuin kolme yritystä tai toimipaikkaa. Sähkön ja lämmön tuotantoon on lisätty NACE Rev.1:een nelinumerotasolla muutama luokka, koska Suomen tuotantorakenne ei sopinut luokitukseen. Lisäksi koodeja on muutettu muutamassa erikoistapauksessa, mutta muuten TOL 95 noudattaa EU:n toimialaluokituksen rakennetta. EU:n jäsenvaltioiden tilastot, jotka laaditaan toimialoittain, tuleekin laatia käyttäen NACE Rev.1:stä tai siitä johdettua kansallista luokitusta asetuksessa mainittujen sääntöjen mukaan. TOL 2002 on syntynyt kun TOL 95:teen on tehty EU:n vaatimat muutokset (NACE Rev. 1:en päivitys eli NACE 2002) ja tehty viisinumero-tasolla tarkistuksia. (Tilastokeskus, 1993 ja Tilastokeskus, 2002 c).

Monella toimialalla toimivan yrityksen **päätoimiala** määritellään siten, että se on toimiala, josta pääosa yrityksen arvonlisäyksestä muodostuu. **Arvonlisäyksellä** tarkoitetaan perushintaista bruttoarvonlisäystä. Se on perushintaisen tuotoksen ja ostajahintaisen välituotekäytön erotus, eli se on tuotteen hinnan ja sen tuottamiseen menneiden kulujen erotus. Lyhyesti sanottuna se on kunkin taloudellista toimintaa harjoittavan yksikön bruttokansantuotteeseen lisäämä osuus. Jos arvonlisäystietoa ei ole saatavilla, käytetään esim. palkkoja, henkilöstön määrää, tuotannon bruttoarvoa tai liikevaihtoa. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

**Toimialoittaiset arvonlisäykset** saadaan Tilastokeskuksen yritysrekisterissä kahden tekijän tulona. Ensiksi lasketaan kunkin toimialan arvonlisäys henkilöä kohden jakamalla toimialalla toimivien yritysten yhteenlaskettu arvonlisäys niiden yhteenlasketulla henkilöstömäärällä. Toiseksi lasketaan monitoimipaikkaisen yrityksen toimialoittainen henkilöstö, joka on samaan toimialaluokkaan kuuluvien toimipaikkojen henkilöstön summa. Tämän jälkeen yrityksen **toimialoittaiset arvonlisäykset** (tai tarkemmin sanottuna estimaatit tai arviot yrityksen toimialoittaisille arvonlisäyksille) saadaan em. tekijöiden tulona. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

**Monella toimialalla toimivan yrityksen sivutoimialoja** ovat sellaiset toimialat, joihin liittyvää toimintaa yritys harjoittaa, mutta joka ei ole yrityksen päätoimiala. **Yrityksen toiminnan puhtausasteella** tarkoitetaan sitä kuinka monta prosenttia yrityksen arvonlisäyksestä muodostuu sen päätoimialan toiminnasta. Ennen tätä tutkimusta toimialakohtaisten tilinpäätöstietojen ja tunnuslukujen estimaatit laadittiin suoraan käyttämällä yritysten tilinpäätöstietoja, yritysten toiminnan puhtausasteita ja tietoa yritysten päätoimialoista. Toimialakohtainen estimaatti muodostettiin Tilastokeskuksessa siten, että laskettiin yhteen sellaisten yritysten tilinpäätöstie-

dot, joiden päätoimiala oli se, jolle estimaatti laadittiin, ja joiden toiminnan puhtausaste oli yli 90 %. (Tilastokeskus, 1993, Tilastokeskus, 2002 c, 14).

**Yritys voidaan jakaa erilaisiin yksiköihin toimialan ja sijainnin perusteella. Niitä ovat toimipaikat, apu-toimipaikat sekä toimiala- ja paikallisyksiköt. Toimipaikalla** tai paikallisella toimialayksiköllä tarkoitetaan yhden yrityksen omistamaa (saman omistajuuden alla tai saman yrityksen valvonnan alaisuudessa toimivaa), yhdessä paikassa sijaitsevaa ja pääasiassa yhdellä toimialalla toimivaa talousyksikköä. **Yritykset ovat yksi- tai monitoimipaikkaisia.** Monitoimipaikkainen yritys on sellainen, joka toimii useammalla toimialalla ja/tai eri osoitteissa. Monitoimipaikkaisia yrityksiä on siten kolmenlaisia: 1. yritykset, jotka toimivat useammassa osoitteessa, mutta niissä kaikissa osoitteissa on yhden ja saman toimialan toimintoja, 2. yritykset, jotka toimivat vain yhdessä osoitteessa, mutta tuossa osoitteessa on useamman toimialan toimintoja ja 3. yritykset, jotka toimivat useammassa osoitteessa ja niissä harjoitetaan useamman toimialan toimintoja. Yksitoimipaikkainen yritys toimii yhdessä osoitteessa ja vain yhdellä toimialalla. Tällöin yritys muodostuu yhdestä toimipaikasta ja yrityksen tiedot = toimipaikan tiedot. Tilastokeskuksessa on tietoja sekä yritys- että toimipaikkatasolla. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

Toimipaikkatieto on keskeinen yksikkö kun sovelletaan toimialaluokitusta, koska yrityksen toimipaikkakohtaiset tiedot antavat parhaan kuvan yrityksen talouden rakenteesta, yritystoiminnasta voidaan toimipaikkojen avulla kerätä tietoja ja laatia tilastoja maantieteellisten ja hallinnollisten alueiden mukaisesti, monitoimialaisesta yrityksestä saadaan toimialoittain eriteltyä tietoa, monet tuotantoon liittyvät perustiedot (kuten hyödykkeiden hinnat, kuukausittaiset myyntitiedot jne.) on nopeinta ja taloudellisinta kerätä suoraan toimipaikoilta ja kaikki henkilöperusteiset tilastot kuvaavat väestön toimiala- eli elinkeinonjakaumaa toimipaikkojen kautta. Lyhyesti sanottuna, vain toimipaikkakohtaisten tietojen kautta saadaan toimiala- ja aluekohtaisia tietoja ja niiden yhdistelmiä. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

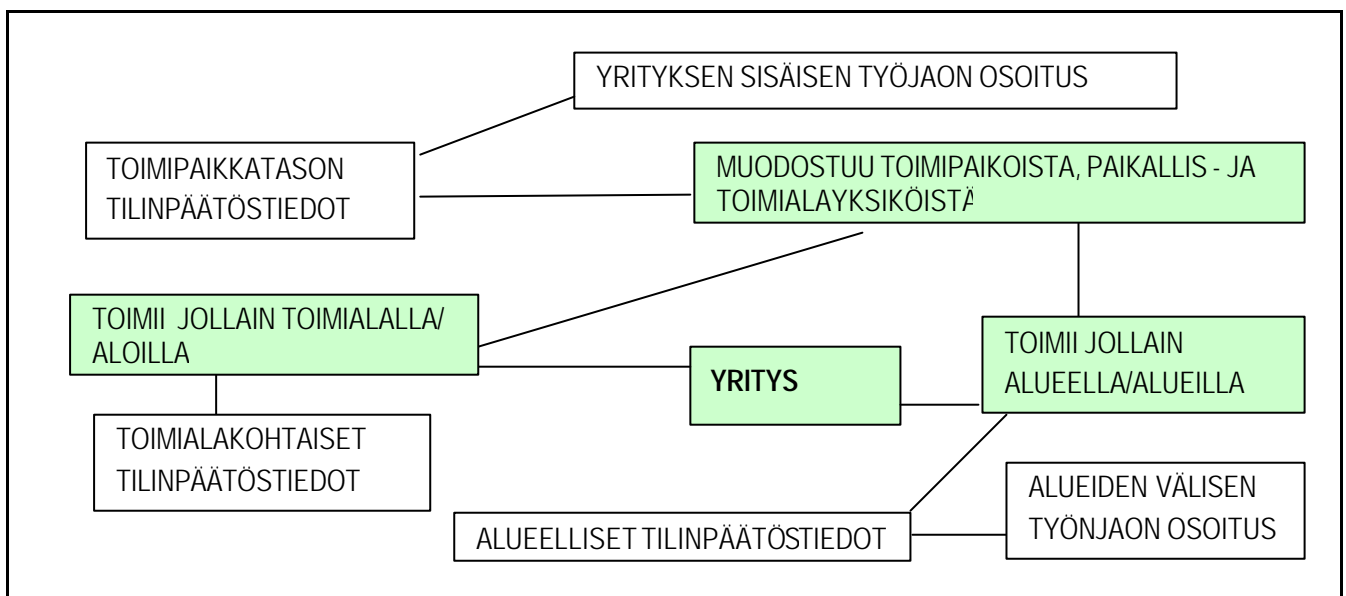
**Toimipaikalla tarkoitetaan kuitenkin eri asiaa kuin aputoimipaikka, paikallisyksikkö ja toimialayksikkö. Aputoimipaikalla** tarkoitetaan yhden yrityksen omistamaa talousyksikköä, joka tuottaa hyödykkeitä (erityisesti palveluksia) pääasiassa yritykselle tai yritystyyppiselle yksikölle itselleen. Miltei aina aputoiminnot ovat välttämättömiä pää- ja sivutoimintojen suorittamiseksi. Tästä syystä aputoimipaikkaa ei voida pitää varsinaisena toimipaikkana, vaan se liitetään yrityksen toimialaa määriteltäessä päätoimintaa suorittavaan yksikköön. Aputoimipaikkoja ovat mm. pää-, piiri- tms. hallinnollinen toimisto, erillinen kuljetusyksikkö, yrityksen omien ajoneuvojen korjaamo, varasto-, ja tietojenkäsittely -yksikkö. Jos tällaiset toiminnot on järjestetty normaalin tuotannon kanssa samaan toimipaikkaan, niitä ei yleensä pidetä erillisinä aputoimipaikkoina. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

**Toimialayksikkö** on yhden yrityksen omistama yksikkö, joka toimii pääasiallisesti yhdellä toimialalla. Sitä ei rajoiteta alueellisesti yhteen paikkaan. Jokaisella yrityksellä on yksi tai useampia toimialayksikköjä. Jos yrityksellä on enemmän kuin yksi toimialayksikkö, toimii kyseinen yritys vähintään kahdella toimialalla ja sillä on vähintään kaksi toimipaikkaa. Jos taas yrityksellä on yksi toimialayksikkö, toimii se vain yhdellä toimialalla, mutta sillä saattaa olla yksi tai useampi toimipaikka. Jos yrityksellä on yksi toimialayksikkö, muodostaa tuo

toimialayksikkö yrityksen. Mutta se, että yrityksellä on yksi toimialayksikkö ei tarkoita, että kyseinen yritys olisi yksitoimipaikkainen. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

**Paikallisyksiköllä tai paikallisella yksiköllä** tarkoitetaan yhden yrityksen omistamaa ja yhdessä paikassa, osoitteessa, (maantieteellisesti) sijaitsevaa yrityksen osaa. Paikallisyksikkö ei riipu toimialasta, eli yhdessä paikallisyksikössä voi olla useamman toimialan toimintoja. Yhdessä paikallisyksikössä voi olla monta toimipaikkaa, jos paikallisyksikössä on vähintään kahden toimialan toimintoja. Jos yritys on yksitoimipaikkainen, on sillä myös yksi paikallisyksikkö mutta, jos yrityksellä on yksi paikallisyksikkö voi se olla monitoimipaikkainen. Paikallisyksikköjen tiedot antavat tietoa siitä, miten yritys on alueellisesti jakautunut, eli millä alueilla yrityksellä on toimintaa. Paikallisyksikköä voidaan käyttää alueellisissa tarkasteluissa, joissa ei tarvita yksityiskohtaista toimialatietoa. (Tilastokeskus, 1993, Tilastokeskus, 2002 c).

Kuvio 2 kuvaa yrityksen jakoa erilaisiin yksiköihin. Kirjoittaja on sen itse laatinut kirjallisuuden perusteella (Kinkki, 1995, Kinnunen ym., 2000, Tilastokeskus, 1993, Tilastokeskus, 2002c).



**Kuvio 2.** Yrityksen jako yksiköihin.

## Liite 3. Kululajipohjainen tuloslaskelma

Tilastokeskuksessa laadittavat toimipaikkojen tilinpäätöstiedot (ja tutkimuksen toimipaikkatason muuttajat) ovat toimipaikan laskennallisen käyttökateen eriä, niistä lasketta via tunnuslukuja tai toimipaikan tilinpäätöksen liitetietoja (kuten henkilöstön määrä). Toimipaikan laskennallista käyttökateetta vastaa yritystasolla tuloslaskelman liikevoittoa/tappiota edeltävä osa. Luvussa 2.1 on kuvattu toimipaikan laskennallisen käyttökateen ja tuloslaskelman eroja. Tässä liitteessä kuvataan kululajipohjaisen tuloslaskelman liikevoittoa/tappiota edeltäviä eriä sekä raportoidaan tuloslaskelman erien käyttötapoja yritystutkimuksessa. Lisäksi esitetään kululajipohjaisen tuloslaskelman kaava ja kerrotaan tuloslaskelman erien ja yrityksen sidosryhmien välisistä yhteyksistä.

Seuraavassa on **kululajipohjaisen tuloslaskelman kaava**. Kululajipohjaista tuloslaskelmaa käytetään Tilastokeskuksessa. Lähteenä on mm. julkaisu Yritystutkimusneuvottelukunta (1999), Kinnunen ym. (2000) sekä Leppiniemi ja Leppiniemi (2000), joissa on myös määritelty kaikki kululajipohjaisen tuloslaskelman erät.

|   |     |       |
|---|-----|-------|
| <b>LIKEVAIHTO</b>   | +   | _____ |
| <b>VALMIIDEN JA KESKENERÄISTEN TUOTTEIDEN VARASTOJEN MUUTOS</b> | +/- | _____ |
| <b>VALMISTUS OMAAN KÄYTTÖÖN</b>                                 | +   | _____ |
| <b>LIKETOIMINNAN MUUT TUOTOT</b>                                | +   | _____ |
| <b>MATERIAALIT JA PALVELUT:</b>                                 | -   | _____ |
| Ostot tilikauden aikana   | -   | _____ |
| Varastojen muutos   | +/- | _____ |
| Ulkopuoliset palvelut   | -   | _____ |
| <b>HENKILÖSTÖKULUT:</b>   | -   | _____ |
| Palkat ja palkkiot  | -   | _____ |
| <b>henkilösivukulut:</b>  | -   | _____ |
| eläkekulut  | -   | _____ |
| muut henkilösivukulut   | -   | _____ |
| <b>POISTOT JA ARVONALENTAMISET:</b>                             | -   | _____ |
| Suunnitelman mukaiset poistot                                   | -   | _____ |
| Arvonalentamiset pysyvien vastaavien hyödykkeistä               | -   | _____ |
| Vaihtuvien vastaavien poikkeukselliset arvonalentamiset         | -   | _____ |
| <b>LIKETOIMINNAN MUUT KULUT</b>                                 | -   | _____ |
| <b>LIKEVOITTO/TAPPIO</b>  | +/- | _____ |
| <b>RAHOITUSTUOTOT JA KULUT:</b>                                 | +/- | _____ |
| Tuotot osuuksista saman konsernin yrityksissä                   | +   | _____ |
| Tuotot osuuksista omistusyhteyserityksissä                      | +   | _____ |
| Tuotot muista pysyvien vastaavien sijoituksista                 | +   | _____ |
| Muut korko- ja rahoitustuotot                                   | +   | _____ |
| Arvonalentamiset pysyvien vastaavien sijoituksista              | -   | _____ |
| Arvonalentamiset vaihtuvien vastaavien rahoitusarvopapereista   | -   | _____ |
| Korkokulut ja muut rahoituskulut                                | -   | _____ |
| <b>VOITTO (TAPPIO) ENNEN SATUNNAISIA ERIÄ</b>                   | +/- | _____ |
| <b>SATUNNAISET ERÄT:</b>  | +/- | _____ |
| Satunnaiset tuotot  | +   | _____ |
| Satunnaiset kulut   | -   | _____ |
| <b>VOITTO (TAPPIO) ENNEN TILINPÄÄTÖSSIIRTOJA JA VEROJA</b>      | +/- | _____ |
| <b>TILINPÄÄTÖSSIIRROT</b>                                       | +/- | _____ |
| Poistoeron muutos   | +/- | _____ |
| Vapaaehtoisten varausten muutos                                 | +/- | _____ |
| <b>TULOVEROT</b>  | -   | _____ |
| <b>MUUT VÄLITTÖMÄT VEROT</b>                                    | -   | _____ |
| <b>TILIKAUDEN VOITTO (TAPPIO)</b>                               | +/- | _____ |

Kulupohjaisen tuloslaskelman liikevoittoa/tappiota edeltävien erien määritelmät (mm. Yritystutkimusneuvotte-  
lukunta, 1999, Leppiniemi ja Leppiniemi, 2000):

**Liikevaihto** koostuu kirjanpitolain mukaan yrityksen varsinaiseen toimintaan (sen toimialan toimintaa, jolla yritys toimii) kuuluvien tuotteiden tai palvelujen myynituotoista, joista vähennetään annetut arvonalennukset sekä arvonalisävero ja muut välittömästi myynnin määrään perustuvat verot (mm. arvonalisävero). Se on yleisemmin käytetty yrityksen toiminnan laajuuden, merkittävyyden ja markkinaosuuden mitta (yrityksen liikevaihtotietoa verrataan mm. muiden yritysten liikevaihtoon tai muiden saman toimialan yritysten liikevaihtoon) ja se on osatekijänä monissa yritystä kuvaavissa tunnusluvuissa. Esimerkiksi eri kuluerien suuruutta tutkitaan usein suhteuttamalla ne liikevaihtoon. Liikevaihdon muutosten perusteella tehdään Leppiniemen ja Leppiniemen (2000) mukaan päätelmiä yritysten riskistä, mm. erittäin suurta liikevaihdon kasvua on pidetty konkurssiuhan merkinä. Samoin konkurssiuhan merkinä on pidetty myös liikevaihdon kasvun tyrehtymistä.

Seuraavaksi liikevaihtoon lisätään tai siitä vähennetään **valmiiden ja keskeneräisten tuotteiden varastojen muutos**. Se koskee vain valmistustoimintaa harjoittavia yrityksiä, koska se sisältää vain yrityksen itsensä valmistaman varaston muutoksen (ei siis muilta ostettuja tuotannon tekijöitä). Se sisältää sekä sellaiset tuotteet, jotka ovat jo valmistuneet että sellaiset, jotka ovat vielä kesken, mutta joiden valmistuksen yritys on jo aloittanut (puolivalmisteet). Jos valmiiden ja keskeneräisten tuotteiden varasto on kasvanut, esitetään se tuloslaskelmassa plusmerkkisenä, jos taas vähentynyt, merkitään se miinusmerkkisenä. Seuraavaksi edellä olevista eristä vähennetään **valmistus omaan käyttöön**. Valmistus omaan käyttöön tarkoittaa kuluja, joita on koitunut yrityksen valmistaessa tuotteita, joita se valmistaa vain omaan käyttöönsä, eli ei myy. Niitä voivat olla mm. myytävien tuotteiden valmistamisessa tarvittavien koneiden valmistus ja niiden korjauspalvelut. Niitä ei ole tarpeen merkitä erikseen, joten erän puuttumisesta ei voi päätellä, että yritys ei olisi valmistanut tuotteita/palveluita omaan käyttöön.

**Liiketoiminnan muihin tuottoihin** sijoitetaan tulot, jotka eivät sisälly liikevaihtoon, mutta on saatu yrityksen varsinaisesta toiminnasta ja ovat yrityksen toimialan mukaisia tuloja. Liiketoiminnan muita tuottoja ovat mm. vuokratuotot, saadut osingot, muilta yrityksiltä perittävät hallinto-, tietojenkäsittely yms. korvaukset, käyttöomaisuuden myynnistä saadut tuotot ja pääosa saaduista avustuksista. Liiketoiminnan muut tuotot -erään kuuluvia tuottoja on saatettu merkitä satunnaisiin tuottoihin tai jopa liikevaihtoon. Liiketoiminnan muut tuotot eivät vaikuta yrityksen liikevoittoon, eivätkä sen avulla muodostettuihin tunnuslukuihin.

**Materiaaleihin ja palveluihin** merkitään kulut, jotka koituvat materiaalien ja palvelujen (yrityksen vaihtomaisuuden), eli lyhytaikaisten tuotannon tekijöiden, hankkimisesta. Se jaetaan kahteen alaryhmään: aineet, tarvikkeet ja tavarat sekä ulkopuoliset palvelut. Kulut, jotka ovat koituneet aineiden, tarvikkeiden ja tavaroiden hankkimisesta merkitään kohtaan aineet, tarvikkeet ja tavarat, riippumatta siitä onko niillä valmistetut hyödykkeet jo myyty, vai ovatko ne varastossa. Ulkoisiin palveluihin merkitään taas lähinnä työsuorituksista maksettuja korvauksia. Ne voivat olla esimerkiksi alihankkijoiden, suunnittelu- ja konsulttitoimistojen sekä huoltoyhtiöiden suorittamista palveluista maksettuja korvauksia. Työvoiman vuokrauskulut ovat myös ulkoisia palveluita. Ulkoisiin palveluihin eivät kuulu hallintoon liittyvistä palveluista maksetut korvaukset. Ulkoisiin palveluihin siis merkitään kaikki työsuorituksista maksetut korvaukset muille kuin yrityksen työntekijöille, paitsi ei hallintoon liittyviä kuluja.

**Palkalla** tarkoitetaan työntekijän ja työnantajan välisen työsuhteen perusteella työnantajan maksamaa korvausta työntekijälle tekemästään työstä (Taloustieto Oy, 1994). Se määräytyy yleensä työhön käytettävän ajan, työn tuloksen tai muun sovituksen perusteella. Tavallisesti palkka maksetaan rahana, mutta voidaan sitä myös sopimuksen mukaan maksaa ns. luontoisetuina, joita ovat esim. vapaa, työsuhteasunto, työsuhdepuhelin tai työsuhdeauto ovat luontoisetuja. Luontoisedut rinnastetaan verotuksessa palkkaan, ne ovat siis verotettavaa tuloa työntekijälle. Luontoisetuja ei kuitenkaan merkitä tuloslaskelmassa erään palkat. Yrittäjäkin voi maksaa itselleen palkkaa määräämänsä summan verran. Yksityisyrietykset eivät maksa lainkaan palkkaa, vaan niiden omistajat (yrittäjät) maksavat itselleen korvauksia suoraan omasta pääomastaan (eli omaisuudestaan).

**Sosiaalivakuutusmaksuilla** tarkoitetaan Tilastokeskuksessa (2002 b) työnantajien lakisääteisiä sekä lakisääteisten luonteisia vapaaehtoisia sosiaalivakuutusmaksuja. Ne ilmoitetaan erikseen toimihenkilöistä ja työntekijöistä. Lakisääteisiä sosiaalivakuutusmaksuja ovat sosiaaliturvamaksu, johon sisältyy kansaneläke-, lapsilisä- ja sairausvakuutusmaksu, työnantajan työttömyysvakuutusmaksu, LEL- ja TEL- vakuutusmaksu, tapaturmavakuutusmaksu ja ryhmähenkivakuutusmaksu. Vapaaehtoisia sosiaalivakuutusmaksuja

ovat TEL- ja LEL- lisäturvamaksut, henki- ja eläkevakuutusyhtiöille maksetut ryhmähenki- ja ryhmäeläkevakuutusmaksut sekä työpaikkojen avustuskassoihin maksetut kannatusmaksut.

Työnantaja maksaa palkan lisäksi palkan suuruuden mukaan erilaisia maksuja kuten sosiaaliturvamaksun, tapaturmavakuutusmaksun ja eläkevakuutusmaksun, vuosilomapalkan, arkipyhäkorvauksia ja eräitä muita korvauksia. Näitä työnantajan maksamia maksuja sanotaan ns. **henkilösivukuiksi** tilinpäätöksessä. (Tilastokeskus, 2002 b).

**Henkilöstökuluihin** merkitään ennakkipidätyksenalaiset yrityksen maksamat palkat ja niihin verrattavat kulut (erä palkat ja palkkiot) sekä välittömästi palkkojen perusteella määräytyvät kulut (erä henkilösivukulut), mm. sosiaaliturvamaksut (työnantajan maksamat maksut verohallinnolle), pakolliset ja vapaaehtoiset henkilövakuutusmaksut sekä eläkekulut (erä eläkekulut). Yrityksen maksamia luontoisetuja ei merkitä henkilöstökuluihin, vaan ne merkitään luonnettaan vastaavaan kulueraan, yleensä poistoihin tai liiketoiminnan muihin kuluihin. Eläkekuluja ovat yrityksen maksamat sekä vapaaehtoiset että pakolliset eläkevakuutusmaksut.

**Varastojen muutoksella** tarkoitetaan yrityksen varastojen muutoksen rahallista arvoa. Varastojen muutokseen sisältyvät sellaisenaan yrityksen vaihto-omaisuuden muutokset mm. varsinaiseen toimintaan liittyvät arvonalentamiset, kuten tuotteiden tai raaka-aineiden teknisestä vanhenemisestä tai pilaantumisesta johtuvat arvonalentamiset. Varastojen muutos on positiivinen, jos varastojen arvo on kasvanut edellisestä tilikaudesta ja negatiivinen, jos varastojen arvo on laskenut edellisestä tilikaudesta.

**Poistojen** tekemisellä tarkoitetaan menettelyä, jolla käyttöomaisuuden hankintamenot ja muut pitkävaikutteiset menot (investoinnit) jaksetaan vähennettäväksi tuotoista niiden käyttöaikana. Poistoa tehdään siksi, että muuten voi olla, että sinä tilikautena, jona tehdään jokin/joitakin suuria investointeja, tilikausi muodostuu tappiolliseksi investointien vuoksi. Toisaalta investoinnit tuottavat tuloja (tai ainakin niiden ajatellaan tuottavan) useamman tilikauden aikana, joten ne eivät käytännössä kohdistu vain siihen tilikauteen, jona ne on tehty.

Poistot tehdään tilinpäätöksessä poistosuunnitelman mukaisina. Poistosuunnitelma kertoo siitä miten poistot tehdään yrityksen kirjanpidossa. Siinä on myös mainittu poistoaika (maksimissaan hankitun tuotteen käyttöaika, johon poisto kohdistuu) ja poistomenetelmä. Tilinpäätökseen merkitään tehdyn poistosuunnitelman mukaiset poistot (eli suunnitelman mukaiset poistot). Suunnitelman mukaiset poistot siis kertovat tilikaudelle kohdistetut pitkävaikutteisista menoista aiheutuva maksusuoritukset suunnitelman mukaan. Lyhyesti sanottuna poistolla tarkoitetaan tilikauteen kohdistettua investoinneista aiheutuvaa kuluosuutta. On huomattava, että toimipaikkakohtainen tuloslaskelma ei sisällä poistoa, koska niitä ei tehdä toimipaikoilla, vaan yrityksissä. Suunnitelman mukaiset poistot ovat verotuksessa vähennettäviä, ja niille on määrätty enimmäismäärä (Leppiniemi ja Leppiniemi, 2000), mutta poistoeron muutos ei ole verotuksessa vähennettävissä.

**Käyttöomaisuus (eli pysyvien vastaavien hyödykkeet)** muodostuvat pitkävaikutteisista tuotannontekijämenoista, joiden uskotaan tuottavan tuloja, ja joita ei ole vielä vähennetty poistoina. Niillä tarkoitetaan aineettomia hyödykkeitä ( tutkimusmenoja, kehittämismenoja, liikearvoa ja muita aineettomia hyödykkeitä), aineellisia hyödykkeitä (maa- ja vesialueet, rakennukset ja rakennelmat, koneet ja kalusto sekä muut aineelliset hyödykkeet) sekä sijoituksia ja pitkäaikaisia saamisia (sisäiset osakkeet ja osuudet, muut osakkeet ja osuudet, sisäiset saamiset ja muut saamiset sekä sijoitukset).

**Arvonalentamisilla pysyvien vastaavien hyödykkeistä** tarkoitetaan arvonalentamisia käyttöomaisuudesta. Pysyvien vastaavien arvonalentamisia saattaa aiheutua mm. siitä, että koneeseen tulee jokin vika tai osakkeiden ja osuuksien arvo laskee.

**Vaihtuvien vastaavien poikkeuksellisiin arvonalentamisiin** merkitään merkittävät poikkeuksellisista syistä johtuvat yrityksen vaihto-omaisuuden arvonalentamiset. Yrityksen vaihto-omaisuudella tarkoitetaan yrityksen sellaisenaan tai jalostettuina luovutettavaksi tai kulutettavaksi tarkoitetut hyödykkeet sekä niistä maksetut ennakot. Vaihto-omaisuus siis koostuu lyhytvaikutteisista tuotannontekijämenoista, joista tuloa ei ole vielä saatu, ja lyhytvaikutteisista tuotannontekijämenoista, joita yritys ei ole vielä luovuttanut, mutta joista maksu on saatu. Vaihtuvien vastaavien poikkeukselliset arvonalentamiset ovat yksi osa varastojen muutosta valmiiden ja keskeneräisten tuotteiden varastojen sekä aine-, tarvike- ja tavaravarastojen muutosten

lisäksi. Vaihtuvien vastaavien poikkeuksellisia arvonalentamisia saattaa aiheutua esim. tulipalosta tai varkaudesta.

**Liiketoiminnan muihin kuluihin** kuuluvat myyntiprovisiot, tekijäpalkkiot, rahtikulut, syntyneet luottotappiot, käyttöomaisuuden myyntitappiot, vuokrat ja leasing-vuokrat. Myös kiinteistövero ja niihin verrattavat verot ovat liiketoiminnan muita kuluja. Ylipäätänsä liiketoiminnan muihin kuluihin sisällytetään kaikki ne menot, joille ei tuloslaskelmassa ole varattu omaa erää. Ainoastaan poikkeukselliset, kertaluonteiset ja olenaiset erät merkitään satunnaisiin kuluihin.

**Liikevoitto/tappio = liikevaihto + valmiiden ja keskeneräisten tuotteiden varastojen muutos + valmistus omaan käyttöön + liiketoiminnan muut tuotot - materiaalit ja palvelut - henkilöstökulut - poistot ja arvonalentamiset - liiketoiminnan muut kulut.** Liikevoitto kuvaa yrityksen liiketoiminnan onnistumista ja kannattavuuskehitystä. Siinä on liikevaihdosta vähennetty välittömästi yrityksen toimintaan liittyvät kulut. Jos yrityksen toimintaa on rahoitettu vieraalla pääomalla (liite 1), niin osa liikevoitosta menee rahoituskuluihin (korot ja lyhennykset) (Kinnunen ym., 2000).

Jos tilikauden tulos on ollut voitollinen, jaetaan siitä osa **omistajille** erilaisina voittoina ja osinkoina. Loput menevät säästöön turvaamaan yrityksen kasvua. **Asiakkaat** eivät aina maksa koko hintaa ostamistaan tuotteista, vaan saavat erilaisia alennuksia, ja ne eivät näy tuloslaskelmassa suoraan, vaan ovat erä, joka vähennetään myyntituotoista liikevaihtoa laskettaessa. Jossakin tapauksessa asiakkaat jättävät kokonaan maksamatta ostamastaan hyödykkeestä. Tämä on luottotappiota yritykselle, mutta on luettavaksi tuloksi asiakkaalle. Luottotappiot sisältyvät liikevaihtoon ja merkitään kuluksi erään liiketoiminnan muut kulut. Siis osa liiketoiminnan muista kuluista on tuloa asiakkaille. **Tavarantoimittajille** maksetuista maksuista koostuu erä materiaalit ja palvelut, eli tuo erä ilmoittaa sen miten paljon rahaa eri tavarantoimittajat ovat yritykseltä saaneet. Henkilöstökulut ovat työstä niiden tekijöille maksettuja korvauksia ja ovat siten tuloa **yrityksen henkilökunnalle**. Korkokulut ja muut rahoituskulut, arvonalentamiset pysyvien vastaavien sijoituksista ja arvonalentamiset vaihtuvien vastaavien rahoitusarvopapereista, ovat kuluja, jotka aiheutuvat vieraasta pääomasta, ja ovat niin ollen tuloa yrityksen **rahoittajille**. Lainanlyhennykset ovat pääomanpalautusta, jotka maksetaan yritykselle jäävästä osuudesta, eivätkä näyt vähennyksinä -kuluina- tuloslaskelmassa. Tuloverot, muut välittömät verot sekä henkilösivukulut muodostavat julkiselle vallalle maksettavat maksut, ja ovat siten tuloja **julkiselle vallalle**. Satunnaiset kulut ovat tuottoja erilaisille sidosryhmille tai vain kuluja yritykselle itselleen. (Kinkki, 1995).

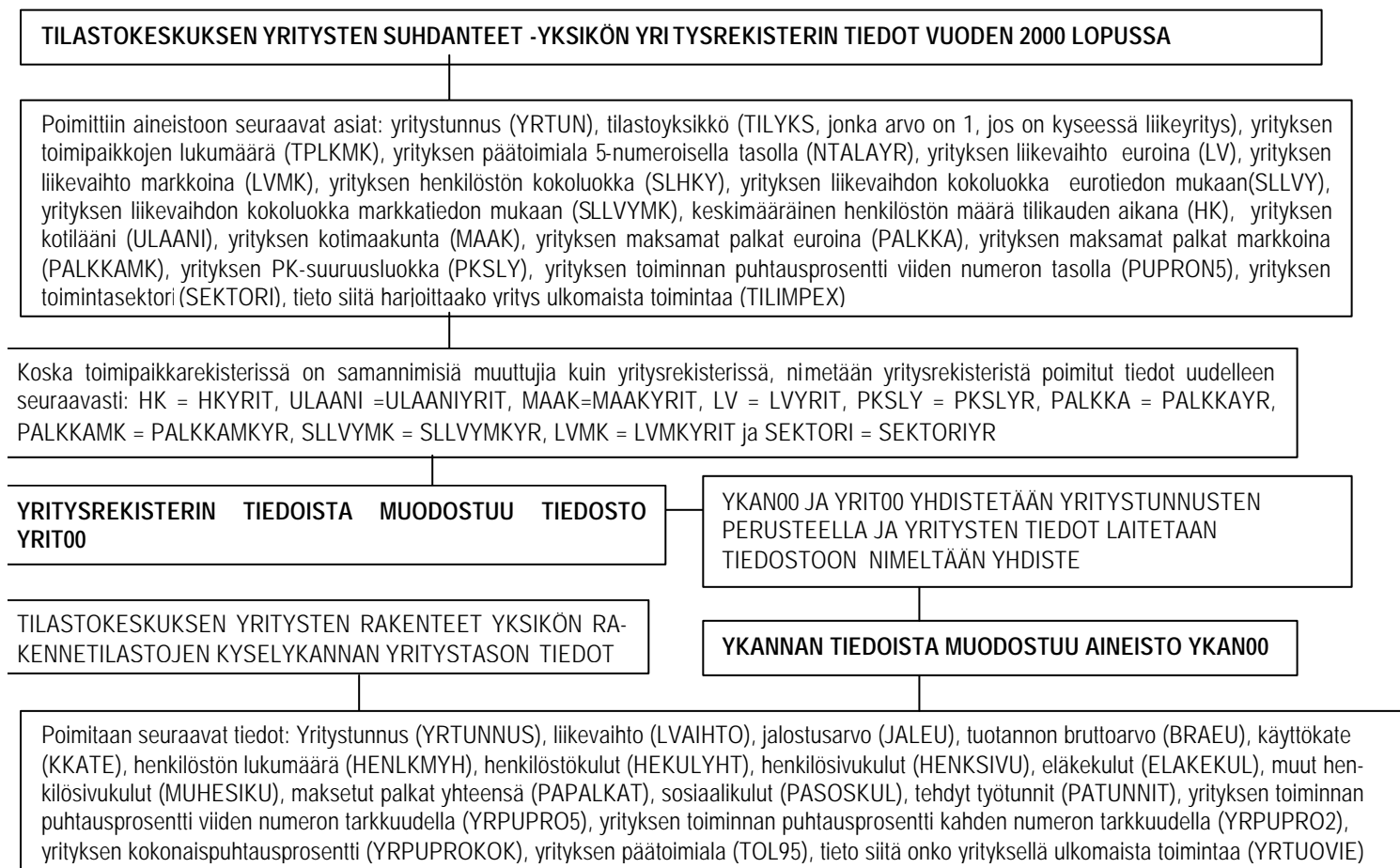
**Yrityksen toiminnan laajuutta** kuvaavat erät liikevaihto, valmiiden ja keskeneräisten tuotteiden varastojen muutos, valmistus omaan käyttöön ja liiketoiminnan muut tuotot. Ne siis muodostavat kululajipohjaisen tuloslaskelman neljä ensimmäistä riviä. Yrityksen toiminnan laajentumista tai supistumista kuvataan liikevaihdon muutosprosentilla =  $100 * (\text{Tilikauden liikevaihto} - \text{Edellisen tilikauden liikevaihto}) / (\text{Edellisen tilikauden liikevaihto})$ . (Leppiniemi ja Leppiniemi 2000).

**Yrityksen kokoa** on ollut tapana mitata liikevaihdolla, henkilöstömäärällä, taseen loppusummalla ja yrityksen osakkeiden pörssinoteeraukseen perustuvalla oman pääoman markkina-arvolla. Yrityksen koon kasvua voidaan mitata liikevaihdon prosentuaalisella muutoksella edellisestä tilikaudesta. Voidaan toki myös tutkia yrityksen koon kasvua pitemmältä ajalta. Se minkä kokoisena yritystä pidetään ja miten suurta yrityksen koon kasvun ajatellaan olevan, saattaa riippua käytettävästä mittarista ja tarkasteltavasta aikavälistä (esim. käytetäänkö liikevaihtoa vai henkilöstömäärää). (Kinnunen ym., 2000, 126-128).

**Yrityksen kannattavuutta** kuvaa yrityksen tulo (voitto tai tappio, tuloslaskelman lopussa). Jos yrityksen tulos on positiivinen, on yritys ollut tilikautena kannattava, jos yrityksen tulos on negatiivinen, ei yritys ole ollut tilikautena kannattava. Jotta yrityksen toiminta olisi pitemmällä ajalla kannattavaa, on tulojen oltava riittävän paljon menoja suuremmat, pelkkä yrityksen tuloksen positiivisuus ei riitä toiminnan jatkumisen takaamiseen. Yritysten välisissä vertailuissa käytetään yleensä yrityksen tuloksen jakamista liikevaihdolla tai yrityksen tuloksen jakaminen sen käytettävissä oleville resursseilla, esim. pääomalla. (Leppiniemi ja Leppiniemi, 2000, Kinnunen ym., 2000).

## Liite 4. Aineistojen muodostaminen SAS-ohjelmistolla yhdistämällä YTR ja rakennetilastojen kyselykanta

Tässä liitteessä on esitetty diagrammin muodossa se miten Tilastokeskuksen yritysten suhdanteet -yksikön yritys- ja toimipaikkarekisterin sekä Tilastokeskuksen yritysten rakenteet -yksikön rakennetilastojen kyselykannan yritys- ja toimipaikkatason tiedot tutkimuksessa yhdistettiin, eli miten muodostettiin tutkimusaineisto. Varsinainen muodostamiseen käytetty SAS-koodi on varsin pitkä.





**TILASTOKESKUKSEN YRITYSTEN SUHDANTEET YKSIKÖN TOIMIPAIKKAREKISTERIN TIEDOT VUODEN 2000 LOPUSSA**

Poimitaan seuraavat muuttujat: toimipaikkatunnus (YKSTUN), toimipaikkaa vastaavan yrityksen tunnus (YRTUN), toimipaikkaa vastaavan yrityksen laji (YLAJIYR), tilastoyksikkö (TILYKS, jos on 1, niin kyseessä on liikeyrityksen toimipaikka), toimipaikan sijaintilääni (ULAANI), toimipaikan sijaintimaakunta (MAAK), toimipaikan toimiala (NTALATP), toimipaikan henkilöstön suuruusluokka (SLHKT), toimipaikan keskimääräinen henkilöstön määrä tilikauden aikana (HK), toimipaikan liikevaihto euroina (LV), toimipaikan liikevaihto markkoina (LVMK), toimipaikan liikevaihdon suuruusluokka eurotiedon mukaan (SLLVT), toimipaikan liikevaihdon suuruusluokka markkatiedon mukaan (SLLVTMK), toimipaikan maksamat palkat euroina (PALKKA), toimipaikan maksamat palkat markkoina (PALKKAMK), toimipaikan PK-suuruusluokka (PKSLY), TOTY ja SEKTORI

TOIMIPAIKKAREKISTERIN TIEDOT MUODOSTAVAT TIEDOSTON YREKOOTP

YRITYSTASON TIEDOSTOON (DATAAN YHDISTE) LIITETÄÄN SEKÄ TKANNAN ETTÄ TOIMIPAIKKAREKISTERIN TIEDOT YRITYSTUNNUSTA KÄYTTÄMÄLLÄ. KAIKKI TIEDOT OVAT SIIS TUOSSA TIEDOSTOSSA JA SE NIMETÄÄN LIIKEYRITYKSET - NIMISEKSI TIEDOSTOKSI

TILASTOKESKUKSEN YRITYSTEN RAKENTEET YKSIKÖN RAKENNETILASTOJEN KYSELYKANNAN TOIMIPAIKKATASON TIEDOT VUODELTA 2000

TKANNAN tiedot muodostavat tiedoston TKANTA00

Poimitaan seuraavat muuttujat: yritystunnus (YRTUNNUS), toimipaikkatunnus (TPTUNN), toimipaikkatyyppi (TPTYYP), jalostusarvo (JALEU), tuotannon bruttoarvo (BRAEU), käyttökate (KKATE), henkilöstön lukumäärä (HENLKMYH), henkilöstönkulut (HEKULYHT), henkilösivukulut (HENKSIVU), eläkekulut (ELAKEKUL), muut henkilösivukulut (MUHESIKU), palkat yhteensä (YHTEEPAL), PAPALKAT, sosiaalkulut (PASOSKUL), tehdyt työtunnit (PATUNNIT) ja toimipaikan toimiala (TOL95TMP)

Koska rakennetilastojen kyselykannan yritystason tiedostossa eli YKANNASSA on saman nimisiä muuttujia kuin TKANNASSA nimetään osa muuttujista uudelleen seuraavasti: JALEU = JALEUTP, BRAEU = BRAEUTP, LVAIHTO = LVAIHTOTP, KKATE = KKATETP, PATUNNIT = PATUNNITP, YHTEEPAL = YHTEEPALTP, ELAKEKUL = ELAKEKULTP, MUHESIKU = MUHESIKUTP, HENKSIVU = HENKSIVUTP, HEKULYHT = HEKULYHTTP, HENLKMYH = HENLKMYHTTP, PAPALKAT = PAPALKATTP, PASOSKUL = PASOSKULTP

**TIEDOSTO LIIKEYRITYKSET SIIS SISÄLTÄÄ TOIMIPAIKKA- JA YRITYSTIETOJA JA SE ON MUODOSTETTU YHDISTÄMÄLLÄ YKANTA, TKANTA JA YTR.**

LIITETÄÄN SUORAAN REKISTEREISTÄ JA TIETOKANNOISTA HAETTUIHIN TIETOIHIN (TS. AINEISTOON LIIKEYRITYKSET) SEURAAVAT JO OLEMASSA OLEVISTA MUUTTUJISTA MUODOSTETTAVAT TIEDOT: yrityksen ja toimipaikan toimialatiedot (YTR, TKANTA ja YKANTA) kahden numeron tasolla (muuttujat TOL952TMP, TOL952, NTALA2 ja NTALATP2), yrityksen henkilöstön lukumäärä tiedot YTR:n mukaan, muokatut (muuttujat HK2 ja HK2YRIT), toimipaikan ja sitä vastaavan yrityksen välituotepanokset (VALTUOPA, VALTUOPATP), yrityksen yksimonitoimipaikkaisuus toimipaikkojen lukumäärän mukaan (YKSITOIMIPAIKKAINEN), toimialatiedot kahden kirjaimen tasolla (muuttujat TOLKIRJTMP, TOLKIRJTMP2, TOLKIRJ, TOLKIRJ2), yrityksen ja toimipaikan toimialatiedot yhden numeron/kirjaimen tasolla YTR:n, YKANNAN ja TKANNAN mukaan sekä yhdistetty tieto koskien toimipaikkoja (muuttujat PAALUO, PAALUO2, PAALUO3, PAALUOYR, PAALUOYR2). Yhdistetty tieto on muodostettu siten, että toimipaikan toimiala määritellään TKANNAN mukaan, jos toimipaikka on TKANNASSA ja jos ei niin YTR:n mukaan. Lisäksi muodostetaan muuttuja, joka kertoo onko toimipaikka suuri vai pieni tai keski-suuri toimipaikka (muuttuja PKSLY2).

Muutetaan yritys- ja toimipaikkarekisterin numeeristen tietojen esitysmuoto merkkimuotoisesta numeeriseksi sekä laitetaan YKANNAN tiedot TKANNAN tiedoiksi mikäli se on mahdollista ja TKANNAN tieto puuttuu. Yritystieto on sama kuin toimipaikkatieto, jos yrityksellä on yksi toimipaikka ja toimipaikan toimiala on sama kuin yrityksen päätoimiala.

**AINEISTO, JOTA TUTKIMUKSESSA KÄSITELLÄÄN ELI AINEISTO LIIKEYRITYKSET. SE SISÄLTÄÄ YTR:N, YKANNAN JA TKANNAN TIETOJEN LISÄKSI NIIDEN POHJALTA MUODOSTETTUJA TIETOJA.**

## Liite 5. Tutkimusaineiston tarkastelua SAS-ohjelmistolla

```
/* Käytettävät aineistot talletetaan kirjastoon otosain*/ libname otosain 'd:/elina'; run;
```

```
/* Asetetaan kaikki muuttujat sellaisiksi, että ne on kirjoitettu isolla kirjaimella*/  
OPTIONS VALIDVARNAME=UPCASE run;
```

```
/* Makrolla sort voidaan lajitella jokin aineisto nousevaan järjestykseen yhden tai kahden  
muuttujan suhteen. Lajittelua tarvitaan, jos proseduurin sisältää by -lauseen*/  
%macro sort(data,x,y); proc sort data=&data; by &x &y; %mend; run;
```

```
/* Makro survevmeans kertoo mm. haluttujen muuttujien summat, keskiarvot asetelmaberusteisesti koko  
joukolla ja toimialoittain ja/tai toimipaikkaa vastaavan yrityksen lajeittain. Tulokset eroavat jonkin verran  
univariate-proseduurin tuloksista*/
```

```
%macro survevmeans(data, x1.x2.x3.x4.x5.x6.x7.x8.x9.x10.z1.z2);  
title1 "Survevmeans - parametrien arvot muuttujien &z1 &z2 luokissa"; title2"Aineisto &data";  
proc survevmeans data =&data nobs mean sum cvsum;  
var &x1 &x2 &x3 &x4 &x5 &x6 &x7 &x8 &x9 &x10;  
by &z1 &z2; %mend; run;
```

```
/* Makrolla univariate saadaan selville erilaisia tunnuslukuja halutulle aineistolle ja halutuille muuttujille*/
```

```
%macro univariate(data,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,z1,z2);  
title1 "Aineisto &data";  
proc univariate data=&data;  
var &x1 &x2 &x3 &x4 &x5 &x6 &x7 &x8 &x9 &x10;  
by &z1 &z2; %mend; run;
```

```
/*Makrolla corr saadaan laskettua halutulle aineistolle haluttujen muuttujien välisiä Pearsonin korrelaatio-  
kertoimia koko ioukossa ja esim. toimialoittain tai/ja toimipaikkaa vastaavan yrityksen lajin mukaan*/
```

```
%macro corr(data,x1.x2.x3.x4.x5.x6.x7.x8.x9.x10,x11,x12,z1,z2);  
title1"PEARSONIN KORRELAATIOKERTOIMIA";  
title2"AINEISTO &DATA"; proc corr data=&data;  
var &x1 &x2 &x3 &x4 &x5 &x6 &x7 &x8 &x9 &x10 &x11 &x12;  
by &z1 &z2; %mend; run;
```

```
/*ESIMERKKI MAKROJEN KÄYTÖSTÄ AINEISTOSSA OTOSAIN.PERUSJOUKKO*/
```

```
%sort(otosain.perusioukko.paaluo); run;  
%freq2(otosain.perusioukko.paaluo); run;  
%survevmeans(otosain.perusioukko.HK2,LVMK,PALKKAMK,JALEUTP.paaluo); run;  
%univariate(otosain.perusioukko.HK2,LVMK,PALKKAMK,JALEUTP.paaluo); run;  
%corr(otosain.perusjoukko, HK2,LVMK,PALKKAMK,JALEUTP ,paaluo); run;
```

## Liite 6. Otoksien poiminta ja otoksien ominaisuuksia

OTOS 1= Otos monitoimipaikkaisten liikeyritysten teollisuuden ja rakentamisen toimipaikoista.  
 OTOS 2 = Otos yksitoimipaikkaisista liikeyrityksistä, joiden toimiala on TKANNAN mukaan teollisuus tai rakentaminen. OTOS 3 = Otos yksitoimipaikkaisista liikeyrityksistä, joiden toimiala ei ole TKANNAN mukaan teollisuus tai rakentaminen. OTOS 4= Otos kaikista toimipaikoista, joiden toimiala on TKANNAN mukaan teollisuus tai rakentaminen. OTOS 5= Otos kaikista YTR:ssä ja TKANNASSA olevista monitoimipaikkaisten yritysten toimipaikoista, joiden toimiala on teollisuus ja rakentaminen sekä yksitoimipaikkaisista liikeyrityksistä. OTOS 6= Otos kaikista YTR:n ja YKANNAN vuoden 2000 yksitoimipaikkaisista liikeyrityksistä, joille tilinpäätöstiedot ovat olemassa. Jos yritys on TKANNASSA, on se otoksessa, ja, jos ei, niin se tulee otokseen, jos toimipaikan toimiala = yrityksen toimiala YTR:n mukaan. Näille yrityksille on toimipaikan tiedot saatu siirtämällä YKANNAN tiedot toimipaikkojen tiedoiksi (koskee rakennetilastojen tilinpäätöstietoja, ei muita).

### Otoksien poimintaan tarvittu SAS-koodi:

```
/* Kirjasto, jossa tarvittavat aineistot ovat, ja johon muodostettavat aineistot talletetaan*/
libname otosain 'd:/elina'; run;
```

```
*****
```

```
OTOKSEN POIMINEN MAKROJEN AVULLA. KUN KÄYTETÄÄN OSITETTUA OTANTAA
SUHTEELLISILLA KIINTIÖINNILLÄ JA JOKAISESTA OSITTEESTA POIMITAAN OTOS
YKSINKERTAISELLA SATUNNAISOTANNALLA
```

```
*****
```

```
/* Makrolla sort voidaan lajitella iokin aineisto nousevaan järjestykseen yhden tai kahden
muuttujan suhteen. Lajittelua tarvitaan, jos proseduri sisältää by -lauseen*/
```

```
%macro sort(data.x.v);
proc sort data=&data; by &x &y; %mend; run;
```

```
%macro otos(data.out. samprate. seed, osite);
proc survevselect data=&data out=&out
samprate = &samprate
method=srs seed=&seed stats;
title1 "DATA &OUT"; title2 "STRSRS otos aineistosta &data";
strata &osite; %mend; run;
```

```
*** ESIMERKKINÄ OTOKSEN POIMINEN YKSITOIMIPAikkaisten LIIKEYRITYSTEN JOUKOSTA;
%sort(otosain.vksitoimipaikkaiset2, osite);run;
%otos(ottosain.yksitoimipaikkaiset2, otosain.sample2, 0.4, 7983241, osite);run;
```

```
/******
```

```
MUODOSTETAAN OTOKSET, JOTKA ON YHDISTETTY OTOKSISTA 1, 2 ja 3. DATA OTOSAIN. TEOLLISUUS
SISÄLTÄÄ OTOKSEN TEOLLISUUDEN JA RAKENTAMISEN TOIMIPAIKOISTA (TOIMIALA MÄÄRITELTY
RAKENNETILASTOJEN MUKAAN) =OTOS 4 JA DATA OTOSAIN.YHDISTETTY ON OTOS KAIKISTA
TOIMIPAIKOISTA, JOIDEN JALOSTUSARVO ON TIEDOSSA = OTOS 5. OTOS 6 SISÄLTÄÄ OTOKSEN
KAIKISTA YKSITOIMIPAikkaisista LIIKEYRITYKSISTÄ ELI SE ON OTOKSIEN 2 JA 3 YHDISTELMÄ.
YHDISTÄMINEN TEHDÄÄN MAKROILLA.
```

```
*****/
```

```
%macro yhdistaminen(uusi. data1. data2. bv);
data &uusi; merge &data1 &data2 ; by &by; %mend; run;
```

```
*** ESIMERKKINÄ OTOKSEN 4 MUODOSTAMINEN;
%sort(otosain.sample, vrtun);run; %sort(otosain.sample2, vrtun); run;
%yhdistaminen(otosain.teollisuus, otosain.sample1, otosain.sample2, vrtun);run;
```

Seuraavassa taulukossa ovat otoskorrelaatiot (ne toimipaikat, joiden jalostusarvo on tiedossa). Suluissa ovat vastaavat korrelaatiokertoimet perusjoukossa (ne toimipaikat, joiden jalostusarvo on tiedossa, N=213443).

| <i>TOIMIALA</i>                | <i>OTOS</i> | <i>MUUTTUJA</i> | <i>HK2</i>   | <i>LVMK</i> | <i>PALKKAMK</i> | <i>JALEUTP</i> |
|--------------------------------|-------------|-----------------|--------------|-------------|-----------------|----------------|
| <b>C</b>                       | <b>1</b>    | HK2             | 1            | 0.71(0.73)  | 0.98(0.97)      | 0.39(0.25)     |
| Mineraalien kaivuu             |             | LVMK            | 0.71(0.73)   | 1           | 0.69(0.70)      | 0.68(0.33)     |
|                                |             | PALKKAMK        | 0.98(0.97)   | 0.69(0.70)  | 1               | 0.33(0.36)     |
|                                |             | JALEUTP         | 0.39(0.25)   | 0.68(0.33)  | 0.33(0.36)      | 1              |
| <b>D</b>                       | <b>1</b>    | HK2             | 1            | 0.74(0.64)  | 0.99(0.99)      | 0.80(0.76)     |
| Teollisuus                     |             | LVMK            | 0.74(0.64)   | 1           | 0.73(0.64)      | 0.94(0.89)     |
|                                |             | PALKKAMK        | 0.99(0.99)   | 0.73(0.64)  | 1               | 0.80(0.77)     |
|                                |             | JALEUTP         | 0.80(0.76)   | 0.94(0.89)  | 0.80(0.77)      | 1              |
| <b>E</b>                       | <b>1</b>    | HK2             | 1            | 0.61(0.51)  | 0.99(0.99)      | 0.64(0.74)     |
| Sähkön-, kaasun- ja vesihuolto |             | LVMK            | 0.61(0.51)   | 1           | 0.70(0.54)      | 0.60(0.47)     |
|                                |             | PALKKAMK        | 0.99(0.99)   | 0.70(0.54)  | 1               | 0.65(0.78)     |
|                                |             | JALEUTP         | 0.64(0.74)   | 0.60(0.47)  | 0.65(0.78)      | 1              |
| <b>F</b>                       | <b>1</b>    | HK2             | 1            | 0.92(0.91)  | 0.998(0.998)    | 0.96(0.84)     |
| Rakentaminen                   |             | LVMK            | 0.92(0.91)   | 1           | 0.93(0.92)      | 0.89(0.79)     |
|                                |             | PALKKAMK        | 0.998(0.998) | 0.93(0.92)  | 1               | 0.96(0.84)     |
|                                |             | JALEUTP         | 0.96(0.84)   | 0.89(0.79)  | 0.96(0.84)      | 1              |
| <b>C - F</b>                   | <b>1</b>    | HK2             | 1            | 0.73(0.64)  | 0.99(0.99)      | 0.80(0.74)     |
| Teollisuus ja rakentaminen     |             | LVMK            | 0.73(0.64)   | 1           | 0.73(0.64)      | 0.93(0.89)     |
|                                |             | PALKKAMK        | 0.99(0.99)   | 0.73(0.64)  | 1               | 0.80(0.77)     |
|                                |             | JALEUTP         | 0.80(0.74)   | 0.94(0.89)  | 0.80(0.77)      | 1              |
| <b>C -F</b>                    | <b>2</b>    | HK2             | 1            | 0.81(0.77)  | 0.99(0.98)      | 0.80(0.73)     |
| Teollisuus ja rakentaminen     |             | LVMK            | 0.81(0.77)   | 1           | 0.84(0.82)      | 0.94(0.88)     |
|                                |             | PALKKAMK        | 0.99(0.98)   | 0.84(0.82)  | 1               | 0.81(0.75)     |
|                                |             | JALEUTP         | 0.80(0.73)   | 0.93(0.88)  | 0.81(0.75)      | 1              |
| <b>C</b>                       | <b>2,6</b>  | HK2             | 1            | 0.90(0.92)  | 0.99(0.99)      | 0.48(0.77)     |
| Mineraalien kaivuu             |             | LVMK            | 0.90(0.92)   | 1           | 0.90(0.91)      | 0.51(0.76)     |
|                                |             | PALKKAMK        | 0.99(0.99)   | 0.89(0.91)  | 1               | 0.48(0.77)     |
|                                |             | JALEUTP         | 0.48(0.77)   | 0.51(0.76)  | 0.48(0.77)      | 1              |
| <b>D</b>                       | <b>2</b>    | HK2             | 1            | 0.81(0.78)  | 0.99(0.99)      | 0.81(0.74)     |
| Teollisuus                     |             | LVMK            | 0.81(0.78)   | 1           | 0.84(0.82)      | 0.94(0.89)     |
|                                |             | PALKKAMK        | 0.99(0.99)   | 0.84(0.82)  | 1               | 0.83(0.76)     |
|                                |             | JALEUTP         | 0.81(0.74)   | 0.94(0.89)  | 0.83(0.76)      | 1              |
| <b>E</b>                       | <b>2</b>    | HK2             | 1            | 0.58(0.68)  | 0.98(0.995)     | 0.93(0.89)     |
| Sähkön-, kaasun- ja vesihuolto |             | LVMK            | 0.58(0.68)   | 1           | 0.67(0.71)      | 0.56(0.72)     |
|                                |             | PALKKAMK        | 0.98(0.995)  | 0.67(0.71)  | 1               | 0.92(0.89)     |
|                                |             | JALEUTP         | 0.93(0.89)   | 0.56(0.72)  | 0.92(0.89)      | 1              |
| <b>F</b>                       | <b>2</b>    | HK2             | 1            | 0.85(0.72)  | 0.96(0.96)      | 0.92(0.89)     |
| Rakentaminen                   |             | LVMK            | 0.85(0.72)   | 1           | 0.93(0.80)      | 0.71(0.77)     |
|                                |             | PALKKAMK        | 0.96(0.96)   | 0.93(0.80)  | 1               | 0.84(0.85)     |
|                                |             | JALEUTP         | 0.92(0.89)   | 0.71(0.77)  | 0.84(0.85)      | 1              |
| <b>KAIKKI TOIMIALAT</b>        | <b>3</b>    | HK2             | 1            | 0.35(0.36)  | 0.93(0.92)      | 0.48(0.35)     |
| paitsi mineraalien kaivuu      |             | LVMK            | 0.35(0.36)   | 1           | 0.35(0.38)      | 0.42(0.17)     |
|                                |             | PALKKAMK        | 0.93(0.92)   | 0.35(0.38)  | 1               | 0.56(0.41)     |
|                                |             | JALEUTP         | 0.48(0.35)   | 0.42(0.17)  | 0.56(0.41)      | 1              |
| <b>D</b>                       | <b>3</b>    | HK2             | 1            | 0.92(0.85)  | 0.96(0.77)      | 0.49(0.45)     |
| Teollisuus                     |             | LVMK            | 0.92(0.85)   | 1           | 0.88(0.69)      | 0.44(0.45)     |
|                                |             | PALKKAMK        | 0.96(0.77)   | 0.88(0.69)  | 1               | 0.66(0.84)     |
|                                |             | JALEUTP         | 0.49(0.45)   | 0.44(0.45)  | 0.66(0.84)      | 1              |
| <b>E</b>                       | <b>3</b>    | HK2             | 1            | 0.31(0.24)  | 0.98(0.997)     | 0.79(0.24)     |
| Sähkön-, kaasun- ja vesihuolto |             | LVMK            | 0.31(0.24)   | 1           | 0.21(0.28)      | -0.16(0.74)    |
|                                |             | PALKKAMK        | 0.98(0.997)  | 0.21(0.28)  | 1               | 0.73(0.29)     |
|                                |             | JALEUTP         | 0.79(0.24)   | -0.16(0.74) | 0.73(0.29)      | 1              |
| <b>F</b>                       | <b>3</b>    | HK2             | 1            | 0.73(0.62)  | 0.90(0.94)      | 0.75(0.60)     |
| Rakentaminen                   |             | LVMK            | 0.73(0.62)   | 1           | 0.94(0.84)      | 0.998(0.99)    |
|                                |             | PALKKAMK        | 0.90(0.94)   | 0.94(0.84)  | 1               | 0.95(0.83)     |
|                                |             | JALEUTP         | 0.75(0.60)   | 0.998(0.99) | 0.95(0.83)      | 1              |
| <b>C - F</b>                   | <b>4</b>    | HK2             | 1            | 0.71(0.63)  | 0.99(0.99)      | 0.77(0.74)     |
| Teollisuus ja rakentaminen     |             | LVMK            | 0.71(0.63)   | 1           | 0.71(0.64)      | 0.94(0.89)     |
|                                |             | PALKKAMK        | 0.99(0.99)   | 0.71(0.64)  | 1               | 0.78(0.76)     |
|                                |             | JALEUTP         | 0.77(0.74)   | 0.94(0.89)  | 0.78(0.76)      | 1              |
| <b>C</b>                       | <b>4,5</b>  | HK2             | 1            | 0.78(0.78)  | 0.98(0.98)      | 0.49(0.38)     |
| Mineraalien kaivuu             |             | LVMK            | 0.78(0.78)   | 1           | 0.76(0.75)      | 0.71(0.43)     |
|                                |             | PALKKAMK        | 0.98(0.98)   | 0.76(0.75)  | 1               | 0.44(0.45)     |
|                                |             | JALEUTP         | 0.49(0.38)   | 0.71(0.43)  | 0.44(0.45)      | 1              |

| <b>TOIMIALA</b>                                   | <b>OTOS</b>      | <b>MUUTTUJA</b> | <b>HK2</b>  | <b>LVMK</b> | <b>PALKKAMK</b> | <b>JALEUTP</b> |
|---|------------------|-----------------|-------------|-------------|-----------------|----------------|
| <b>D</b>  | <b>4</b>         | HK2             | 1           | 0.71(0.63)  | 0.99(0.99)      | 0.78(0.74)     |
| Teollisuus  |                  | LVMK            | 0.71(0.63)  | 1           | 0.72(0.64)      | 0.94(0.89)     |
|   |                  | PALKKAMK        | 0.99(0.99)  | 0.72(0.64)  | 1               | 0.79(0.77)     |
|   |                  | JALEUTP         | 0.78(0.74)  | 0.94(0.89)  | 0.79(0.77)      | 1              |
| <b>E</b>  | <b>4</b>         | HK2             | 1           | 0.60(0.55)  | 0.99(0.99)      | 0.70(0.77)     |
| Sähkön-, kaasun- ja vesihuolto                    |                  | LVMK            | 0.60(0.55)  | 1           | 0.69(0.58)      | 0.62(0.52)     |
|   |                  | PALKKAMK        | 0.99(0.99)  | 0.69(0.58)  | 1               | 0.70(0.80)     |
|   |                  | JALEUTP         | 0.70(0.77)  | 0.62(0.52)  | 0.70(0.80)      | 1              |
| <b>F</b>  | <b>4</b>         | HK2             | 1           | 0.91(0.85)  | 0.99(0.99)      | 0.94(0.84)     |
| Rakentaminen                                      |                  | LVMK            | 0.91(0.85)  | 1           | 0.92(0.88)      | 0.87(0.78)     |
|   |                  | PALKKAMK        | 0.99(0.99)  | 0.92(0.88)  | 1               | 0.90(0.82)     |
|   |                  | JALEUTP         | 0.94(0.84)  | 0.87(0.78)  | 0.90(0.82)      | 1              |
| <b>KAIKKI TOIMIALAT</b>                           | <b>6</b>         | HK2             | 1           | 0.66(0.57)  | 0.97(0.96)      | 0.73(0.52)     |
|   |                  | LVMK            | 0.66(0.57)  | 1           | 0.68(0.60)      | 0.76(0.44)     |
|   |                  | PALKKAMK        | 0.97(0.96)  | 0.68(0.60)  | 1               | 0.75(0.56)     |
|   |                  | JALEUTP         | 0.73(0.52)  | 0.76(0.44)  | 0.75(0.56)      | 1              |
| <b>D</b>  | <b>6</b>         | HK2             | 1           | 0.82(0.78)  | 0.99(0.97)      | 0.80(0.73)     |
| Teollisuus  |                  | LVMK            | 0.82(0.78)  | 1           | 0.84(0.81)      | 0.93(0.88)     |
|   |                  | PALKKAMK        | 0.99(0.97)  | 0.84(0.81)  | 1               | 0.82(0.77)     |
|   |                  | JALEUTP         | 0.80(0.73)  | 0.93(0.88)  | 0.82(0.77)      | 1              |
| <b>E</b>  | <b>6</b>         | HK2             | 1           | 0.51(0.52)  | 0.98(0.995)     | 0.76(0.47)     |
| Sähkön-, kaasun- ja vesihuolto                    |                  | LVMK            | 0.51(0.52)  | 1           | 0.54(0.55)      | 0.17(0.74)     |
|   |                  | PALKKAMK        | 0.98(0.995) | 0.54(0.55)  | 1               | 0.74(0.51)     |
|   |                  | JALEUTP         | 0.76(0.47)  | 0.17(0.74)  | 0.74(0.51)      | 1              |
| <b>F</b>  | <b>6</b>         | HK2             | 1           | 0.84(0.72)  | 0.96(0.96)      | 0.87(0.88)     |
| Rakentaminen                                      |                  | LVMK            | 0.84(0.72)  | 1           | 0.93(0.80)      | 0.73(0.78)     |
|   |                  | PALKKAMK        | 0.96(0.96)  | 0.93(0.80)  | 1               | 0.80(0.84)     |
|   |                  | JALEUTP         | 0.87(0.88)  | 0.73(0.78)  | 0.80(0.84)      | 1              |
| <b>KAIKKI TOIMIALAT</b>                           | <b>5</b>         | HK2             | 1           | 0.70(0.62)  | 0.99(0.98)      | 0.76(0.72)     |
|   |                  | LVMK            | 0.70(0.62)  | 1           | 0.70(0.63)      | 0.93(0.85)     |
|   |                  | PALKKAMK        | 0.99(0.98)  | 0.70(0.63)  | 1               | 0.77(0.74)     |
|   |                  | JALEUTP         | 0.76(0.72)  | 0.93(0.85)  | 0.77(0.74)      | 1              |
| <b>D</b>  | <b>5</b>         | HK2             | 1           | 0.71(0.63)  | 0.99(0.99)      | 0.78(0.74)     |
| Teollisuus  |                  | LVMK            | 0.71(0.63)  | 1           | 0.72(0.64)      | 0.94(0.89)     |
|   |                  | PALKKAMK        | 0.99(0.99)  | 0.72(0.64)  | 1               | 0.79(0.76)     |
|   |                  | JALEUTP         | 0.78(0.74)  | 0.94(0.89)  | 0.79(0.76)      | 1              |
| <b>E</b>  | <b>5</b>         | HK2             | 1           | 0.59(0.52)  | 0.99(0.99)      | 0.70(0.65)     |
| Sähkön-, kaasun- ja vesihuolto                    |                  | LVMK            | 0.59(0.52)  | 1           | 0.67(0.56)      | 0.57(0.55)     |
|   |                  | PALKKAMK        | 0.99(0.99)  | 0.67(0.56)  | 1               | 0.70(0.69)     |
|   |                  | JALEUTP         | 0.70(0.65)  | 0.57(0.55)  | 0.70(0.69)      | 1              |
| <b>F</b>  | <b>5</b>         | HK2             | 1           | 0.90(0.85)  | 0.99(0.99)      | 0.94(0.84)     |
| Rakentaminen                                      |                  | LVMK            | 0.90(0.85)  | 1           | 0.92(0.88)      | 0.87(0.78)     |
|   |                  | PALKKAMK        | 0.99(0.99)  | 0.92(0.88)  | 1               | 0.90(0.82)     |
|   |                  | JALEUTP         | 0.94(0.84)  | 0.87(0.78)  | 0.90(0.82)      | 1              |
| <b>A</b>  | <b>3, 5 ja 6</b> | HK2             | 1           | 0.78(0.75)  | 0.98(0.98)      | 0.86(0.82)     |
| Maatalous   |                  | LVMK            | 0.78(0.75)  | 1           | 0.74(0.71)      | 0.86(0.84)     |
|   |                  | PALKKAMK        | 0.98(0.98)  | 0.74(0.71)  | 1               | 0.85(0.81)     |
|   |                  | JALEUTP         | 0.86(0.82)  | 0.86(0.84)  | 0.85(0.81)      | 1              |
| <b>B</b>  | <b>3, 5 ja 6</b> | HK2             | 1           | 0.89(0.82)  | 0.89(0.92)      | 0.85(0.80)     |
| Metsätalous                                       |                  | LVMK            | 0.89(0.82)  | 1           | 0.80(0.77)      | 0.91(0.86)     |
|   |                  | PALKKAMK        | 0.89(0.92)  | 0.80(0.77)  | 1               | 0.81(0.79)     |
|   |                  | JALEUTP         | 0.85(0.80)  | 0.91(0.86)  | 0.81(0.79)      | 1              |
| <b>G</b>  | <b>3, 5 ja 6</b> | HK2             | 1           | 0.42(0.48)  | 0.94(0.94)      | 0.57(0.67)     |
| Tukku- ja vähittäiskauppa                         |                  | LVMK            | 0.42(0.48)  | 1           | 0.47(0.56)      | 0.78(0.64)     |
|   |                  | PALKKAMK        | 0.94(0.94)  | 0.47(0.56)  | 1               | 0.65(0.73)     |
|   |                  | JALEUTP         | 0.57(0.67)  | 0.78(0.64)  | 0.65(0.73)      | 1              |
| <b>H</b>  | <b>3, 5 ja 6</b> | HK2             | 1           | 0.92(0.89)  | 0.98(0.99)      | 0.96(0.95)     |
| Majoitus- ja ravitsemistoiminta                   |                  | LVMK            | 0.92(0.89)  | 1           | 0.93(0.90)      | 0.95(0.93)     |
|   |                  | PALKKAMK        | 0.98(0.99)  | 0.93(0.90)  | 1               | 0.96(0.96)     |
|   |                  | JALEUTP         | 0.96(0.95)  | 0.95(0.93)  | 0.96(0.96)      | 1              |
| <b>I</b> Kuljetus-, varastointi- ja tietoliikenne | <b>3, 5 ja 6</b> | HK2             | 1           | 0.60(0.64)  | 0.99(0.98)      | 0.41(0.64)     |
|   |                  | LVMK            | 0.60(0.64)  | 1           | 0.62(0.70)      | 0.29(0.53)     |
|   |                  | PALKKAMK        | 0.99(0.98)  | 0.62(0.70)  | 1               | 0.46(0.70)     |
|   |                  | JALEUTP         | 0.41(0.64)  | 0.29(0.53)  | 0.46(0.70)      | 1              |

| <b>TOIMIALA</b>                                    | <b>OTOS</b> | <b>MUUTTUJA</b> | <b>HK2</b>  | <b>LVMK</b> | <b>PALKKAMK</b> | <b>JALEUTP</b> |
|--|-------------|-----------------|-------------|-------------|-----------------|----------------|
| <b>J</b><br>Rahoitustoiminta                       | 3,5 ja 6    | HK2             | 1           | .           | 0.86(0.90)      | 0.46(0.80)     |
|  |             | PALKKAMK        | 0.86(0.90)  | 1           | 1               | 0.49(0.68)     |
|  |             | JALEUTP         | 0.46(0.80)  | .           | 0.49(0.68)      | 1              |
| <b>K</b><br>Kiinteistö-, vuokraus- ja tutkimus.    | 3, 5 ja 6   | HK2             | 1           | 0.58(0.58)  | 0.93(0.92)      | 0.67(0.51)     |
|  |             | LVMK            | 0.58(0.58)  | 1           | 0.64(0.66)      | 0.69(0.62)     |
|  |             | PALKKAMK        | 0.93(0.92)  | 0.64(0.66)  | 1               | 0.68(0.56)     |
| <b>M</b><br>Koulutus                               | 3, 5 ja 6   | HK2             | 1           | 0.63(0.51)  | 0.97(0.97)      | 0.82(0.92)     |
|  |             | LVMK            | 0.63(0.51)  | 1           | 0.70(0.55)      | 0.80(0.59)     |
|  |             | PALKKAMK        | 0.97(0.97)  | 0.70(0.55)  | 1               | 0.82(0.94)     |
| <b>N</b><br>Terveystieteiden- ja sosiaalipalvelut. | 3, 5 ja 6   | HK2             | 1           | 0.82(0.82)  | 0.97(0.96)      | 0.91(0.90)     |
|  |             | LVMK            | 0.82(0.82)  | 1           | 0.86(0.86)      | 0.85(0.88)     |
|  |             | PALKKAMK        | 0.97(0.96)  | 0.86(0.86)  | 1               | 0.94(0.93)     |
| <b>O</b><br>Muut yht. k ja henk. koht. palvelut    | 3, 5 ja 6   | HK2             | 1           | 0.49(0.52)  | 0.95(0.94)      | 0.85(0.83)     |
|  |             | LVMK            | 0.49(0.52)  | 1           | 0.61(0.63)      | 0.54(0.60)     |
|  |             | PALKKAMK        | 0.95(0.94)  | 0.61(0.63)  | 1               | 0.87(0.86)     |
| <b>X</b><br>Toimiala tuntematon                    | 3, 5 ja 6   | HK2             | 1           | 0.66(0.68)  | 0.52(0.36)      | 0.51(-0.05)    |
|  |             | LVMK            | 0.66(0.68)  | 1           | 0.86(0.25)      | 0.57(0.35)     |
|  |             | PALKKAMK        | 0.52(0.36)  | 0.86(0.25)  | 1               | 0.38(0.12)     |
|  |             | JALEUTP         | 0.51(-0.05) | 0.57(0.35)  | 0.38(0.12)      | 1              |

Seuraavassa taulukossa ovat toimipaikkojen jakaumat toimialoittain ja lääneittäin otoksissa (toimiala- ja aluekohtaiset jakaumat prosentteina). Suluissa ovat vastaavat luvut perusjoukossa.

| <i>TOIMIALA</i>    | <i>OTOS</i> | <i>ETELÄ-SUOMEN LÄÄNI n, %</i> | <i>LÄNSI-SUOMEN LÄÄNI n, %</i> | <i>ITÄ-SUOMEN LÄÄNI n, %</i> | <i>OULUN LÄÄNI n, %</i>   | <i>LAPIN LÄÄNI n, %</i>  | <i>AHVENANMAA n, %</i>  |
|--------------------|-------------|--------------------------------|--------------------------------|------------------------------|---------------------------|--------------------------|-------------------------|
| <b>C</b>           | 1           | 15, 30 % (29, 29.3 %)          | 22, 44 % (43, 43.4 %)          | 6, 12 % (14, 14.1 %)         | 6, 12 % (9, 9.1 %)        | 1, 2% (4, 4 %)           | 0, 0 % (0, 0 %)         |
| C                  | 2,6         | 60, 13.4 % (146, 13.1 %)       | 198, 44.3 % (473, 42.4 %)      | 68, 15.2 % (170, 15.2 %)     | 103, 23 % (263, 23.6 %)   | 17, 3.8 % (61, 5.5 %)    | 1, 0.2 % (3, 0.3 %)     |
| C                  | 4,5         | 75, 15.1 % (175, 14.4 %)       | 220, 44.3 % (516, 42.5 %)      | 74, 14.9 % (184, 15.1 %)     | 109, 21.9 % (272, 22.4 %) | 18, 3.6 % (65, 5.4 %)    | 1, 0.2 % (3, 0.3 %)     |
| <b>D</b>           | 1           | 389, 38.1 % (779, 38.2 %)      | 404, 39.6 % (796, 39 %)        | 95, 9.3 % (233, 11.4 %)      | 99, 9.7 % (175, 8.6 %)    | 27, 2.7 % (48, 2.4 %)    | 6, 0.6 % (9, 0.4 %)     |
| D                  | 2           | 3315, 34.2 % (8293, 34.2 %)    | 4430, 45.7 % (11022, 45.5 %)   | 999, 10.3% (2466, 10.2 %)    | 587, 6.1 % (1565, 6.5%)   | 304, 3.1 % (749, 3.1 %)  | 63, 0.7 % (146, 0.6 %)  |
| D                  | 3           | 29, 40.9 % (72, 40.9 %)        | 30, 42.3 % (75, 42.6 %)        | 9, 12.7 % (13, 7.4 %)        | 2, 2.8 % (14, 8 %)        | 1, 1.4 % (2, 1.1 %)      | 0, 0% (0, 0 %)          |
| D                  | 4           | 3704, 34.6 % (9072, 34.5 %)    | 4834, 45.1 % (11818, 45 %)     | 1094, 10.2 % (2699, 10.3%)   | 686, 6.4 % (1740, 6.6 %)  | 331, 3.1 % (797, 3%)     | 69, 0.6 % (155, 0.6%)   |
| D                  | 5           | 3733, 34.6 % (9144, 34.6 %)    | 4864, 45.1 % (11893, 45 %)     | 1103, 10.2 % (2712, 10.3 %)  | 688, 6.4 % (1754, 6.6%)   | 332, 3.1 % (799, 3 %)    | 69, 0.6 % (155, 0.6 %)  |
| D                  | 6           | 3344, 34.2 % (8365, 34.3 %)    | 4460, 45.7 % (11097, 45.5 %)   | 1008, 10.3 % (2479, 10.2 %)  | 589, 6 % (1579, 6.5 %)    | 305, 3.1 % (751, 3.1 %)  | 63, 0.6 % (146, 0.6 %)  |
| <b>E</b>           | 1           | 30, 30 % (61, 30.8 %)          | 23, 23 % (54, 27. 3 %)         | 10, 10 % (20, 10.1 %)        | 25, 25 % (40, 20.2 %)     | 12, 12% (23, 11.6 %)     | 0, 0 % (0, 0 %)         |
| E                  | 2           | 63, 21.3 % (151, 20.4 %)       | 137, 46.3 % (337, 45.6 %)      | 34, 11.5 % (96, 13 %)        | 34, 11.5 % (88, 11.9 %)   | 21, 7.1 % (56, 7.6 %)    | 7, 2.4 % (11, 1.5 %)    |
| E                  | 3           | 4, 66.7 % (7, 53. 9 %)         | 0, 0% (1, 7.7 %)               | 0, 0% (2, 15.4 %)            | 1, 16.7 % (2, 15.4 %)     | 0, 0 % (0, 0 %)          | 1, 16.7 % (1, 7.7 %)    |
| E                  | 4           | 93, 23.5 % (212, 22.6 %)       | 160, 40.4 % (391, 41.7%)       | 44, 11.1 % (116, 12.4%)      | 59, 14.9 % (128, 13.7%)   | 33, 8.3 % (79, 8.4%)     | 7, 1.8 % (11, 1.2%)     |
| E                  | 5           | 97, 24.1 % (219, 23.1 %)       | 160, 39.8 % (392, 41.3 %)      | 44, 11 % (118, 12.4 %)       | 60, 14.9 % (130, 13.7%)   | 33, 8.2 % (79, 8.3 %)    | 8, 2 % (12, 1.3 %)      |
| E                  | 6           | 67, 22.2 % (158, 21 %)         | 137, 45.4 % (338, 45 %)        | 34, 11.3 % (98, 13 %)        | 35, 11.6 % (90, 12 %)     | 21, 7 % (56, 7.5 %)      | 8, 2.7 % (12, 1.6 %)    |
| <b>F</b>           | 1           | 99, 44 % (187, 41.7 %)         | 71, 31.6 % (147, 32.7 %)       | 25, 11.1 % (53, 11.8 %)      | 21, 9.3 % (45, 10 %)      | 8, 3.6 % (15, 3.3 %)     | 1, 0.4 % (2, 0.5 %)     |
| F                  | 2           | 4581, 40.6 % (11412, 40.5 %)   | 4300, 38.1 % (10824, 38.4 %)   | 1078, 9.6 % (2658, 9.4 %)    | 813, 7.2 % (1999, 7.1%)   | 386, 3.4 % (1027, 3.6 %) | 126, 1.1 % (289, 1 %)   |
| F                  | 3           | 7, 63.6 % (18, 69.2 %)         | 3, 27.3 % (5, 19.2 %)          | 0, 0% (0, 0 %)               | 1, 9.1 % (2, 7.7 %)       | 0, 0% (1, 3.9 %)         | 0, 0 % (0, 0 %)         |
| F                  | 4           | 4680, 40.7% (11599, 40.5%)     | 4371, 38 % (10971, 38.3 %)     | 1103, 9.6 % (2711, 9.5 %)    | 834, 7.3 % (2044, 7.1 %)  | 394, 3.4% (1042, 3.6%)   | 127, 1.1 % (291, 1 %)   |
| F                  | 5           | 4687, 40. 7 % (11617, 40.5 %)  | 4374, 38 % (10976, 38.3 %)     | 1103, 9.6 % (2711, 9.5 %)    | 835, 7.3 % (2046, 7.1%)   | 394, 3.4 % (1043, 3.6 %) | 127, 1.1 % (291, 1 %)   |
| F                  | 6           | 4588, 40.6 % (11430, 40.5 %)   | 4303, 38.1 % (10829, 38.4 %)   | 1078, 9.5 % (2658, 9.4 %)    | 814, 7.2 % (2001, 7.1%)   | 386, 3.4 % (1028, 3.6 %) | 126, 1.1 % (289, 1 %)   |
| <b>A</b>           | 3,5,6       | 422, 17 % (1070, 17.2 %)       | 1307, 52.5 % (3257, 52.4 %)    | 400, 16.1 % (1001, 16.1 %)   | 239, 9.6 % (589, 9.5 %)   | 108, 4.3 % (262, 4.2 %)  | 14, 0.6 % (43, 0.7 %)   |
| <b>B</b>           | 3,5,6       | 45, 17.9 % (112, 17.9 %)       | 110, 43.8 % (280, 44.7 %)      | 18, 7.2 % (55, 8.8 %)        | 40, 15.9 % (99, 15.8 %)   | 19, 7.6 % (37, 5.9 %)    | 19, 7.6 % (43, 6.9 %)   |
| <b>G</b>           | 3,5,6       | 7927, 43.9 % (19650, 43.6 %)   | 6604, 36.6 % (16377, 36.3 %)   | 1657, 9.2 % (4328, 9.6 %)    | 1144, 6.4 % (2973, 6.6%)  | 556, 3.1 % (1396, 3.1 %) | 141, 0.8 % (347, 0.8 %) |
| <b>H</b>           | 3,5,6       | 1428, 37.9 % (3555, 37.8 %)    | 1358, 36.1 % (3412, 36.2 %)    | 441, 11.7 % (1047, 11.1 %)   | 264, 7 % (727, 7.7 %)     | 216, 5.7 % (536, 5.7 %)  | 60, 1.6 % (139, 1.5 %)  |
| <b>I</b>           | 3,5,6       | 3318, 36.4 % (8357, 36.7 %)    | 3354, 36.8 % (8293, 36.4 %)    | 1152, 12.6 % (2833, 12.4 %)  | 768, 8.4 % (1979, 8.9%)   | 465, 5.1 % (1156, 5.1 %) | 61, 0.7 % (176, 0.8 %)  |
| <b>J</b>           | 3,5,6       | 590, 62.4 % (1474, 62.3 %)     | 235, 24.8 % (604, 25.5 %)      | 49, 5.2 % (115, 4.9 %)       | 35, 3.7 % (84, 3.6 %)     | 22, 2.3 % (48, 2 %)      | 15, 1.6 % (40, 1.7 %)   |
| <b>K</b>           | 3,5,6       | 8298, 52.4 % (20906, 52.8 %)   | 4880, 30.8 % (12187, 30.8 %)   | 1217, 7.7 % (2965, 7.5 %)    | 889, 5.6 % (2156, 5.5%)   | 459, 2.9 % (1139, 2.9 %) | 82, 0.5 % (208, 0.5 %)  |
| <b>M</b>           | 3,5,6       | 292, 49.7 % (732, 49.9 %)      | 206, 35 % (487, 33.2 %)        | 34, 5.8 % (104, 7.1 %)       | 32, 5.4 % (89, 6.1 %)     | 20, 3.4 % (46, 3.1 %)    | 4, 0.7 % (10, 0.7 %)    |
| <b>N</b>           | 3,5,6       | 2154, 44.8 % (5314, 44.3 %)    | 1593, 33.2 % (4038, 33.6 %)    | 513, 10.7 % (1281, 10.7 %)   | 391, 8.1 % (997, 8.3 %)   | 138, 2.9 % (339, 2.8 %)  | 15, 0.3 % (39, 0.3 %)   |
| <b>O</b>           | 3,5,6       | 2841, 42.8 % (6980, 42.1 %)    | 2337, 35.2 % (5885, 35.5 %)    | 687, 10.4 % (1780, 10.7 %)   | 476, 7.2 % (1215, 7.3%)   | 260, 3.9 % (642, 3.9 %)  | 38, 0.6 % (93, 0.6 %)   |
| <b>X</b>           | 3,5,6       | 3,60 % (5, 45.5 %)             | 1, 20 % (3, 27.3 %)            | 0, 0% (0, 0 %)               | 0, 0% (2, 18.2 %)         | 1, 20 % (1, 9.1 %)       | 0, 0% (0, 0 %)          |
| <b>YKSITOIMIP.</b> | 6           | 35377,42 % (88254, 41.9 %)     | 31083, 36.9 % (77560, 36.8 %)  | 8356, 9.9 % (20914, 9.9 %)   | 5819, 6.9 % (14843, 7%)   | 2993, 3.6% (7498, 3.6 %) | 647, 0.8 % (1588, 0.8%) |
| <b>MONITOIMIP.</b> | 1           | 533, 38.2 % (1056, 37.9 %)     | 520, 37.3 % (1040, 37.3%)      | 136, 9.8 % (320, 11.5 %)     | 151, 10.8 % (269, 9.7%)   | 48, 3.4 % (90, 3.2 %)    | 7, 0.5 % (11, 0.4 %)    |
| <b>KAIKKI</b>      | 5           | 35910, 41.9 % (89310, 41.8%)   | 31603, 36.9 % (78600, 36.8%)   | 8492, 9.9 % (21234, 10%)     | 5970, 7 % (15112, 7.1%)   | 3041, 3.6 % (7588, 3.6%) | 654, 0.8 % (1599, 0.8%) |

## Liite 7. Toimialakohtaisten tilinpäätöstietojen estimaattien muodostaminen SAS-ohjelmistolla

```
/* Kiriasto. iossa aineistot ovat*/
libname otosain 'd:/elina'; run;
```

```
/* Asetetaan kaikki muuttujat sellaisiksi, että ne on kirjoitettu isolla kirjaimella, ja
säädetään tulostussivujen asetuksia*/
OPTIONS VALIDVARNAME=UPCASE nodate nonumber; run;
```

```
** SUORAT HT-ESTIMAATIT;
%macro ht(data.t. v. osite);
proc surveymeans data=&data nobs mean sum sumwat cvsum total=&t;
title1 "SAS/SURVEYMEANS / HT estimaattorit, aineisto &data";
var &v. weight samplinaweight;
strata &osite; %mend; run;
```

```
%macro ht2(data.t. v. osite, domain);
libname library 'd:/elina';
proc surveymeans data=&data nobs mean sum sumwat cvsum total=&t;
title1 "SAS/SURVEYMEANS / HT estimaattorit, aineisto &data";
var &v. weight samplinaweight;
strata &osite; domain &domain;
%mend; run;
```

```
/*Esimerkki makrojen käytöstä*/
%ht(otosain.sample.2786, jaleuto, osite); run;
%ht2(otosain.sample.2786, jaleuto, osite, paaluo); run;
```

```
/* GREG-ESTIMAATTIEN LASKEMINEN SURVEYREG-PROSEDUURILLA*/
```

```
/* Toimialakohtaisia estimaatteja laskeva makro */
%macro req2(data.paaluo.otos.v.t.z1.Tz1.z2.Tz2.z3.Tz3.z4.tz4.z5.Tz5);
title1 "SAS SURVEYREG / Tulosuuttuja y=&y/ Regressioestimointi";
title2 "Aineisto &data. STRSRS otos";
title3 "Tulosuuttuja y=&y totaalin regressioestimointi toimialalla &paaluo otoksessa &otos,
N&paaluo&otos=&t";
title4 "Riippumattomat muuttujat z1=&z1,z2=&z2,z3=&z3,z4=&z4,z5=&z5";
title5 "Riippumattomien muuttujien totaalit perusjoukossa Tz1=&Tz1 Tz2=&Tz2";
title6 "Tz3=&Tz3,Tz4=&Tz4,Tz5=&Tz5";
ods output ParameterEstimates=Toimiala&paaluo&otos;
proc surveyreg data=&data total= &t;
where paaluo = &paaluo AND otos=&otos;
model &v = &z1 &z2 &z3 &z4 &z5/ solution;
weight samplinaweight; strata osite;
estimate "&v:n totaali &paaluo toimialalla otoksessa &otos "
Intercept &t &z1 &Tz1 &z2 &Tz2 &z3 &Tz3 &z4 &Tz4 &z5 &Tz5 / E;
%mend; run;
```

```
/* Otokohtaisia estimaatteja laskeva makro*/
%macro req3(data.otos.v.t.z1.Tz1.z2.Tz2.z3.Tz3.z4.tz4.z5.Tz5);
title1 "Totaalin estimointi, aineisto &data";
title2 "SAS SURVEYREG / Tulosuuttuja y=&y/ Regressioestimointi";
title3 "STRSRS otos, otos &otos";
title4 "Regressioestimointi koko joukossa, N=&t";
title5 "Tulosuuttuja y=&y." ;
```



```

title6 "Riippumattomat muuttujat z1 = &z1 ,z2 = &z2, z3 = &z3, z4 = &z4, z5 = &z5 ";
title7 "Riippumattomien muuttuuien totaalit perusjoukossa Tz1 = &Tz1, Tz2 = &Tz2";
title8 "Tz3 = &Tz3,Tz4 = &Tz4,Tz5 = &Tz5";
ods output ParameterEstimates=Totaalit&otos;
proc surveyreg data=&data total= &t;
where otos=&otos;
model &v=&z1 &z2 &z3 &z4 &z5/ solution;
weight samplingweight;
strata osite;
estimate "&y:n totaali"
Intercept &t &z1 &Tz1 &z2 &Tz2 &z3 &Tz3 &z4 &Tz4 &z5 &Tz5 / E; %mend; run;

/* Esimerkki makrojen käytöstä*/
%reg2(otosain.sample,1,JALEUTP,99,HK2,2226,LVMK,2378911,PALKKAMK,406308); run;

/* VAPAUASASTEKORJATTUJEN SELITYSASTEIDEN JA MUIDEN OMINAISUUKSIEN LASKEMINEN
GREG-PROSEDUURILLA*/

%macro req6(data.paaluo,otos,v.z1.z2.z3.z4.z5);
title1 "SAS REG / Tulomuuttujia v=&v";
title2 "/Tavanomainen regressioanalyysi";
title3 "Aineisto &data. STRSRS otos";
title4 "Tulomuuttujia v=&v totaalin regressioestimointi toimialalla &paaluo ia otoksessa &otos";
title5 "Riippumattomat muuttujat z1=&Z1,z2=&z2,z3=&z3,z4=&z4,z5=&z5";
proc reg data=&data;
weight samplinweight;
where paaluo = &paaluo AND otos=&otos ;
model &v = &z1 &z2 &z3 &z4 &z5/clb ss1 ss2 stb collin tol vif collinoint acov spec covb corrb;
output out=Estimaatit&paaluo&otos
predicted=yhat r=yresid; plot r.*p.; %mend; run;

%macro req7(data.otos,v.z1.z2.z3.z4.z5);
title1 "SAS REG /Tulomuuttujia v=&v";
title2"/Tavanomainen regressioanalyysi";
title3 "Aineisto &data. STRSRS otos";
title4 "Tulomuuttujia v=&v totaalin regressioestimointi otoksessa &otos";
title5 "Riippumattomat muuttujat z1=&Z1.z2=&z2,z3=&z3,z4=&z4,z5=&z5";
proc reg data=&data; where otos=&otos;
weight samplinweight;
model &v = &z1 &z2 &z3 &z4 &z5/clb ss1 ss2 stb collin tol vif collinoint acov spec covb corrb;
output out=Analyysi&otos
predicted=vhat r=vresid;
plot r.*p.; %mend; run;

/* Esimerkki makroien käytöstä*/
%reg7(otosain.sample,1,JALEUTP,HK2,LVMK,PALKKAMK); run;

/* EPÄSUORIEN SYNTEETTISTEN (suorien synteettisten estimaattien laskeminen tapahtuu samaan
tapaan) JA GREG-ESTIMAATTIEN LASKEMINEN*/
/* Parametrien siirtäminen kukin omaksi datakseen*/
%macro datat(tulos.data,x);
data &tulos;set &data;
keep &x; %mend; run;

/*Datoien transponointi*/
%macro transpoosi(data.out.x);
proc transpose data = &data out=&out;
var &x; %mend; run;

```

*/\*Luodaan data iohon parametrit siirretään ja siirretään siihen halutun aineiston tieto\*/*

```
%macro data(data,x);
data &data; set &x;
%mend; run;
```

*/\*Tietojen siirtäminen ja nimien muuttaminen\*/*

```
%macro siirto(x,a1,z,g,t,y);
data &x;
if n =1 then set &a1; set &x;
rename COL1=&z;
rename COL2=&q;
rename COL3=&t;
rename COL4=&y;
%mend; run;
```

```
%macro siirto2(x,a1,z,g);
data &x;
if n =1 then set &a1; set &x;
rename COL1=&z; rename COL2=&g;
%mend; run;
```

*/\*Turhien tietojen poisto\*/*

```
%macro poisto(data,x,y,z,w);
data &data; set &data;
drop &x &y &z &w; %mend; run;
```

*/\*ESTIMAATTIEN JA JÄÄNNÖSTERMIEN LASKEMINEN\*/*

```
%macro estimaatit(tulos,data,otos,estimaatti,v,intercept,estimaatti1,z1,
estimaatti2,z2,estimaatti3,z3,jaannos,painotettu);
data &tulos; set &data;
if otos =&otos then &estimaatti = sum(&intercept + &estimaatti1*&z1
+ &estimaatti2*&z2 + &estimaatti3*&z3);
if otos = &otos then &jaannos= sum(&v -1*&estimaatti);
if otos=&otos then &painotettu=sum(samplingweight*&jaannos);
%mend; run;
```

```
%macro estimaatit2(tulos,data,otos,estimaatti,intercept,estimaatti1,z1,estimaatti2,z2,
v,jaannos,painotettu);
data &tulos; set &data;
if otos =&otos then &estimaatti = sum(&intercept + &estimaatti1*&z1
+ &estimaatti2*&z2);
if otos = &otos then &jaannos= sum(&v -1*&estimaatti);
if otos=&otos then &painotettu=sum(samplingweight*&jaannos);
%mend; run;
```

```
%macro estimaatit3(tulos,data,otos,estimaatti,intercept,estimaatti2,z1,
v,jaannos,painotettu);
data &tulos; set &data;
if otos =&otos then &estimaatti = sum(&intercept + &estimaatti2*&z1);
if otos = &otos then &jaannos= sum(&v -1*&estimaatti);
if otos=&otos then &painotettu=sum(samplingweight*&jaannos);
%mend; run;
```

*/\* Makrolla sort voidaan lajitella iokin aineisto nousevaan järjestykseen yhden tai kahden muuttujan suhteen. Lajittelua tarvitaan jos proseduri sisältää by-lauseen\*/*

```
%macro sort(data.x.v);
proc sort data=&data;
by &x &y; %mend; run;
```

*/\*Estimaatit otos- ja toimialakohtaisesti\*/*

```
%macro estimaatit4(data,otos,paaluo,y,estimaatti,jaannostermi,painotettu,
z1,z2,z3);
title1 "Tietoja aineistosta &data";
title2 "Apumuuttujat &z1 &z2 &z3"; title3 "Tulosmuuttuja &y.";
title4 "Epäsuoran mallin tuottamat estimaatit tilinpäätöstiedoille ja mallin jäännöstermit sekä niiden ominaisuuksia";
title5 "&estimaatti=&v:n estimaatti ia &jaannostermi=&v:n estimaatin jäännöstermi ja";
title6 "&painotettu = painotettu jäännöstermi otoksessa ";
proc univariate data = &data normal;
var &estimaatti &jaannostermi &painotettu;
by &otos &paaluo; %mend; run;
```

*/\*Estimaatit otoskohtaisesti\*/*

```
%macro estimaatit5(data.otos.v,estimaatti,jaannostermi,painotettu,z1,z2,z3);
title1 "Tietoja otoksessa &otos";
title2 "Apumuuttujat &z1 &z2 &z3";
title3 "Tulosmuuttuja &y. Aineisto &data.";
title4 "Epäsuoran mallin tuottamat estimaatit tilinpäätöstiedoille ja mallin jäännöstermit sekä niiden ominaisuuksia";
title5 "&estimaatti=&v:n estimaatti ia &jaannostermi=&v:n estimaatin jäännöstermi";
title6 "&painotettu=&v:n estimaatin jäännöstermi painotettuna otantapainolla. Ts. harhankorjaustermi.";
proc univariate data = &data normal; where otos=&otos;
var &estimaatti &jaannostermi &painotettu;
%mend; run;
```

*/\*Makro, joka laskee harhankorjaustermiä otoksille\*/*

```
%macro estimaatit6(data,otos,paaluo,y,painotettu,
z1,z2,z3);
title1 "Tietoja aineistosta &data";
title2 "Apumuuttujat &z1 &z2 &z3";
title3 "Tulosmuuttuja &y.";
title4 "&painotettu = painotettu jäännöstermi otoksessa ";
proc univariate data = &data normal;
var &painotettu; by &otos &paaluo;
%mend; run;
```

```
%macro estimaatit7(data,otos.v,painotettu,z1,z2,z3);
```

```
title1 "Tietoja otoksessa &otos";
title2 "Apumuuttujat &z1 &z2 &z3";
title3 "Tulosmuuttuja &y. Aineisto &data.";
title6 "&painotettu=&v:n estimaatin jäännöstermi painotettuna otantapainolla. Ts. harhankorjaustermi.";
proc univariate data = &data normal;
where otos=&otos; var &painotettu;
%mend; run;
```

*/\*HT -vertailu ja keskivirheet\*/*

```
%macro HT(data.x.v.a.otos.paaluo.t.z.x1,x2,x3);
proc surveymeans data=&data nobs mean sum sumwat cvsum total=&t;
title1 "SAS/SURVEYMEANS / Otantapainoilla painotetut estimaatit otoksesta";
title2 "Aineisto &data. otos = &otos";
title3 "Tulosmuuttujat &z.apumuuttujat &x2 &x2 &x3";
where otos =&otos; var &x &y &a;
weight samplinqweight; by &paaluo;
ods output Statistics = HT&otos&paaluo; %mend; run;
```

```
%macro HT2(data.x.v.a.otos.t.z.x1,x2,x3);
```

```
proc surveymeans data=&data nobs mean sum sumwat cvsum total=&t;
title1 "SAS/SURVEYMEANS / Otantapainoilla painotetut estimaatit otoksesta";
```

```

title2 "Aineisto &data, otos = &otos";
title3 "Tulosmuuttuina &z, apumuuttujat &x1 &x2 &x3";
where otos = &otos;
var &x &y &q; weight samplinqweight;
ods output Statistics = HT&otos; %mend; run;

```

```
/*GREG-estimaattien keskivirheet*/
```

```

%macro keskivirhe(data,x,y,rsquare);
data &data;
set &data; rsquare=&rsquare;
&y =sum(&x*(1-1*sqrt(&rsquare))); %mend; run;

```

```

%macro print(data,otos,z1,z2,z3);
proc print data=&data;
title1 "Epäsuorien HT-estimaattien keskivirheet";
title2 "Otos = &otos, apumuuttujat &z1 &z2 &z3";
var paaluo rsquare jaleutp_stddev std5; %mend; run;

```

```

%data(estimointi5,totaalit5,estimate); run;
%data(estimointi51.totaalit5.stderr); run;
%data(estimointi52,totaalit5,dendf);run;

```

```

%transpoosi(estimointi5,transpoosi5,estimate);run;
%transpoosi(estimointi51,transpoosi51.stderr);run;
%transpoosi(estimointi52,transpoosi52,dendf);run;

```

```

%data(estimaatti5,otosain.perusioukko);run;
%data(otos5,otosain.yhdistetty);run;

```

```

%siirto(estimaatti5,transpoosi5,INTERCEPT,PHK2,PLVMK,PPALKKAMK);run;
%siirto(otos5,transpoosi5,INTERCEPT,PHK2,PLVMK,PPALKKAMK); run;
%siirto(estimaatti5,transpoosi51,SINTERCEPT,SHK2,SLVMK,SPALKKAMK); run;
%siirto(otos5,transpoosi51.SINTERCEPT,SHK2,SLVMK,SPALKKAMK); run;
%siirto2(estimaatti5,transpoosi52,DENDF,DF); run;
%siirto2(otos5,transpoosi52,DENDF,DF); run;
%poisto(estimaatti5. name .DF. LABEL_); run;
%poisto(otos5,_name_,DF,_LABEL_); run;

```

```
/*Estimaattien laatiminen*/
```

```

%estimaatit (estimaatti5,estimaatti5.5,EJALEUTP,JALEUTP,INTERCEPT,PHK2,HK2,PLVMK,LVMK,
PPALKKAMK,PALKKAMK,VJALEUTP,PVJALEUTP); run;
%estimaatit(otos5,otos5.5,EJALEUTP,JALEUTP,INTERCEPT,PHK2,HK2,PLVMK,LVMK,PPALKKAMK,
PALKKAMK, VJALEUTP,PVJALEUTP); run;
%sort(estimaatti5.paaluo,otos); run;
%sort(otos5.paaluo,otos); run;
%estimaatit5(estimaatti5.5,JALEUTP,EJALEUTP,VJALEUTP,PVJALEUTP,HK2,LVMK,PALKKAMK);run;
%estimaatit4(estimaatti5,otos,paaluo,JALEUTP,EJALEUTP,VJALEUTP,PVJALEUTP,HK2,LVMK,
PALKKAMK); run;
%estimaatit5(otos5.5,JALEUTP,PVJALEUTP,HK2,LVMK,PALKKAMK); run;
%estimaatit4(otos5,otos,paaluo,JALEUTP,PVJALEUTP,HK2,LVMK,PALKKAMK); run;
%HT(otos5.EJALEUTP.VJALEUTP.JALEUTP.5.paaluo.213443.JALEUTP.HK2.LVMK.PALKKAMK); run;
%HT2(otos5,EJALEUTP,VJALEUTP,JALEUTP.5.213443.JALEUTP,HK2,LVMK,PALKKAMK); run;
%keskivirhe(HT5paaluo,JALEUTP STDDEV,STD5,0.8959); run;
%print(HT5paaluo,5,HK2,LVMK,PALKKAMK); run;

```

## Liite 8. Valitut toimialakohtaiset estimaatit

### SUORAT HORVITZ-THOMPSON ESTIMAATIT TOIMIALAKOHTAISILLE JALOSTUSARVOILLE (1000 MK):

| TOIMI-<br>ALA | OTOS  | N      | n     | HT-ESTI-<br>MAATTI | TOD.ARVO (TOD. ARVON<br>JA ESTIMAATIN EROTUS) | ESTIMAATIN<br>KESKIVIRHE | VARIAATIO-<br>KERROIN | 95% LUOTTAMUSVÄLI JA<br>LOSTUSARVOLLE |
|---------------|-------|--------|-------|--------------------|---|--------------------------|-----------------------|---------------------------------------|
| A             | 3,5,6 | 6222   | 2490  | 3377891            | 3221604 (-156287)                             | 128355                   | 0.037999              | 3126315.2,3629466.8                   |
| B             | 3,5,6 | 626    | 251   | 137261             | 155239 (17978)                                | 17771                    | 0.129470              | 102429.84,172092.16                   |
| C             | 1     | 99     | 50    | 604366             | 682423.39(78057.39)                           | 193294                   | 0.319830              | 225509.76,983222.24                   |
| C             | 2,6   | 1116   | 447   | 432010             | 522633 (90623)                                | 88317                    | 0.204432              | 258908.68,605111.32                   |
| C             | 4,5   | 1215   | 497   | 1036377            | 1205056.39 (168679.39)                        | 213195                   | 0.205712              | 618514.8,1454239.2                    |
| D             | 1     | 2039   | 1019  | 1345650548         | 123567032(-122208356)                         | 22991690                 | 0.170751              | 1300586835.6, 1390714260.4            |
| D             | 2     | 24241  | 9698  | 48752168           | 47269169.4(-1482998.6)                        | 5697187                  | 0.116860              | 37585681.48, 59918654.52              |
| D             | 3     | 176    | 71    | 3132185            | 3237112.6 (104927.6)                          | 935226                   | 0.298586              | 1299142.04, 4965227.96                |
| D             | 4     | 26280  | 10717 | 183402716          | 170836201(-12566515)                          | 26082831                 | 0.142216              | 132280367.24, 224525064.76            |
| D             | 5     | 26456  | 10788 | 186534901          | 174073314(-12461587)                          | 26189054                 | 0.140398              | 135204355.16, 237865446.84            |
| D             | 6     | 24417  | 9769  | 51884353           | 50506282(-1378071)                            | 5791612                  | 0.111625              | 40532793.48, 63235912.52              |
| E             | 1     | 196    | 98    | 4659639            | 5267269.98(607630.98)                         | 811692                   | 0.174234              | 3068722.68, 6250555.32                |
| E             | 2     | 739    | 296   | 1042043            | 1936780.58 (894737.58)                        | 167218                   | 0.160472              | 714295.72,1369790.28                  |
| E             | 3     | 13     | 6     | 244502             | 1158552.56 (914050.56)                        | 305550                   | 1.249683              | -354376, 843380                       |
| E             | 4     | 935    | 394   | 5700682            | 7204050.56 (15033678.56)                      | 838747                   | 0.147131              | 4056737.88, 7344626.12                |
| E             | 5     | 948    | 400   | 5945184            | 8362603.12 (2417419.12)                       | 893046                   | 0.150213              | 4194813.84, 7695554.16                |
| E             | 6     | 752    | 302   | 1286545            | 3095333.14 (1808788.14)                       | 348385                   | 0.270791              | 603710.4,1969379.6                    |
| F             | 1     | 448    | 225   | 9292447            | 8569223.34(-723223.66)                        | 1379055                  | 0.148406              | 6589499.2,11995394.8                  |
| F             | 2     | 28209  | 11284 | 19698208           | 19835230 (137022)                             | 448413                   | 0.022764              | 18819318.52, 20577097.48              |
| F             | 3     | 26     | 11    | 298816             | 263503 (-35313)                               | 169776                   | 0.568161              | -33944.96, 631576.96                  |
| F             | 4     | 28657  | 11509 | 28990655           | 28404453.4 (-586201.6)                        | 1470966                  | 0.050739              | 26107561.64, 31873748.36              |
| F             | 5     | 28683  | 11520 | 29289471           | 28667956.4 (621514.6)                         | 1489818                  | 0.050865              | 26369427.72, 32209514.28              |
| F             | 6     | 28235  | 11295 | 19997024           | 20098733 (101709)                             | 506243                   | 0.025316              | 19004787.72, 20989260.28              |
| G             | 3,5,6 | 45071  | 18029 | 30257855           | 28778094.9(-1479760)                          | 1252459                  | 0.041393              | 27803035.36, 32712674.64              |
| H             | 3,5,6 | 9416   | 3767  | 4094279            | 4157426.88(63147.88)                          | 152428                   | 0.037230              | 3795520.12, 4393037.88                |
| I             | 3,5,6 | 22794  | 9118  | 15041115           | 15679787.4(638672.4)                          | 101459                   | 0.067422              | 14842255.36, 15239974.64              |
| J             | 3,5,6 | 2365   | 946   | 7664448            | 12596235(4931787)                             | 1819814                  | 0.237436              | 4097612.56,11231283.44                |
| K             | 3,5,6 | 39561  | 15825 | 28981602           | 31125353.9(2143751.9)                         | 1136335                  | 0.039209              | 26754385.4,31208818.6                 |
| M             | 3,5,6 | 1468   | 588   | 759091             | 894704.107(135613.11)                         | 110373                   | 0.145402              | 542759.92,975422.08                   |
| N             | 3,5,6 | 12008  | 4804  | 4205913            | 4048848.17(-157064.83)                        | 159472                   | 0.037916              | 3893347.88, 4518478.12                |
| O             | 3,5,6 | 16595  | 6639  | 3916898            | 3930011.3(13113.3)                            | 173892                   | 0.044395              | 3576069.68,4 257726.32                |
| X             | 3,5,6 | 11     | 5     | 435.6              | 766(330.4)                                    | 162.651234               | 0.373396              | 116.8, 754.4                          |
| KAIKKI        | 1     | 2782   | 1392  | 149206001          | 138085948(-11120053)                          | 23048120                 | 0.154472              | 104031685.8, 194380316.2              |
| KAIKKI        | 2     | 54305  | 21725 | 69924429           | 69563813 (-360616)                            | 5717935                  | 0.081773              | 58717276.4, 81131581.6                |
| KAIKKI        | 3     | 156352 | 62550 | 102112292          | 109247239(7134947)                            | 2881795                  | 0.028222              | 96463973.8,107760610.2                |
| KAIKKI        | 4     | 57087  | 23117 | 219130430          | 207649761(-11480669)                          | 26138607                 | 0.119283              | 167898760.28, 270362099.72            |
| KAIKKI        | 5     | 213439 | 85667 | 321242722          | 316897000(-4345722)                           | 26392320                 | 0.082157              | 269513774.83, 72971669.2              |
| KAIKKI        | 6     | 210657 | 84275 | 172036721          | 178811052(6774331)                            | 6446103                  | 0.037469              | 159402359.12,184671082.88             |

## b) Epäsuorat GREG-estimaatit ja epäsuorat synteettiset estimaatit.

Pienimmät keskivirheet antaa "järkevästä" malleista se, jossa apumuuttuja on LVMK. Taustalla oleva malli:

| OTOS | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTE KORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N (n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|------|--|-----------|--------------|-----------------|---|---------------------------------------|----------------------------|
| 5    | 0.8678   | MALLI     | <.0001       | 310658974       | 310659217.233774                                      | 316897000                             | 9528252.80                 |
| 5    | 14938  | VAKIO     | <.0001       | 345.638943      | -243.23377  | 213443(85670)                         | 67.6918806                 |
| 5    | 0.8678, 0.03067  | LVMK      | <.0001       | 0.239325        | -6238026  | 28544                                 | 0.0153777                  |

Estimaatit:

| TOIMIALA | SYNTEETTINEN ESTIMAATTI | HARHANKORJAUS-TERMI | GREG-ESTIMAATTI | KESKIVIRHE    | VARIATIOKERROIN |
|----------|-------------------------|---------------------|-----------------|---------------|-----------------|
| C        | 1298466.06000           | -196111.19000       | 1102354.870000  | 15606.510     | 0.014157428     |
| D        | 138198890.00000         | 39942916.90000      | 178141806.9000  | 1783239.070   | 0.010010222     |
| E        | 7441718.34000           | -704931.81000       | 6736786.530000  | 65184.40      | 0.009675889     |
| F        | 28582690.10000          | -758689.41000       | 27824000.6900   | 112630.670    | 0.004047968     |
| A        | 3648466.59000           | -317999.71000       | 3330466.88000   | 8932.20000000 | 0.002681966     |
| B        | 310961.65100            | -161820.56000       | 149141.0910000  | 1218.40000000 | 0.008169445     |
| G        | 64594013.70000          | -35983137.00000     | 28610876.7000   | 89321.820     | 0.003121953     |
| H        | 6098679.32000           | -1967838.50000      | 4130840.820000  | 10894.6800    | 0.0026374       |
| I        | 16931098.40000          | -1982778.00000      | 14948320.40000  | 73659.380     | 0.004927602     |
| K        | 28359559.30000          | 1058931.04000       | 29418490.3400   | 80170.1300    | 0.002725161     |
| M        | 805545.39600            | -28299.33800        | 777246.0580000  | 7885.89000000 | 0.010145938     |
| N        | 5714289.74000           | -1602944.80000      | 4111344.94000   | 11705.18000   | 0.002847044     |
| O        | 7853039.33000           | -4140582.50000      | 3712456.83000   | 12789.11000   | 0.003444918     |
| X        | 4119.85165              | -3726.03780         | 393.81385000    | 11.13000000   | 0.028262084     |

## c) Suorat GREG-estimaatit ja synteettiset estimaatit. Imputoinnin pohjalla olevat mallit:

Toimiala A. Lisäinformaatio: LVMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTE KORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|--|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.7429   | MALLI     | <.0001       | 3293444.28      | 3293443.541526  | 3221604                              | 52536.9247                 |
| 3,5,6 | 535.50   | VAKIO     | 0.0008       | 100.639437      | 0.73847   | 6222(2490)                           | 30.0728005                 |
| 3,5,6 | 0.7428, 0.01595  | LVMK      | <.0001       | 0.426158        | 71840.28  | 884                                  | 0.0328772                  |

Toimiala B: Lisäinformaatio: LVMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTE KORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N (n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|--|-----------|--------------|-----------------|---|---------------------------------------|----------------------------|
| 3,5,6 | 0.8318   | MALLI     | <.0001       | 157336.900      | 157336.88122640                                       | 155239                                | 7557.93038                 |
| 3,5,6 | 185.27   | VAKIO     | 0.8204       | -4.0127882      | 0.01877   | 626 (251)                             | 17.6605506                 |
| 3,5,6 | 0.8311, 0.04804  | LVMK      | <.0001       | 0.4044309       | 2097.90000  | 43                                    | 0.0439895                  |

**Toimiala C:** Lisäinformaatio:LVMK

| OTOS | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N (n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|------|---|-----------|--------------|-----------------|---|---------------------------------------|----------------------------|
| 4,5  | 0.4981  | MALLI     | <.0001       | 1108780.78      | 1108781.812509  | 1205056                               | 159981.063                 |
| 4,5  | 3276.19   | VAKIO     | 0.3763       | 119.097407      | -1.03251  | 1215 (497)                            | 134.487937                 |
| 4,5  | 0.4971, 0.14429   | LVMK      | <.0001       | 0.262634        | -96275.22000  | 91                                    | 0.027522                   |

**Toimiala D :** Apumuuttuja: LVMK

| OTOS | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 1    | 0.8858  | MALLI     | <. 0001      | 127028591       | 127045194.0824  | 123567032                            | 7293634.95                 |
| 1    | 136615  | VAKIO     | <. 0001      | 15994.2104      | -16603.08240  | 2040(1020)                           | 3449.10628                 |
| 1    | 0.8857, 0.05742   | LVMK      | <.0001       | 0.2413          | 3461559   | 2                                    | 0.01349                    |
| 6    | 0.8695  | MALLI     | <. 0001      | 50937243        | 50937285.260223                                       | 50506282                             | 2074133.28                 |
| 6    | 8718.74   | VAKIO     | 0.0015       | 371.716215      | -42.26022   | 24417(9769)                          | 116.979518                 |
| 6    | 0.8695, 0.04072   | LVMK      | <. 0001      | 0.282924        | 430961  | 1580                                 | 0.026283                   |

**Toimiala E:** Apumuuttuja: Monitoimipaikkaiset PALKKAMK, yksitoimipaikkaiset HK2

| OTOS | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 1    | 0.4279  | MALLI     | <. 0001      | 6511002.41      | 6511007.22602   | 5267270                              | 674294.426                 |
| 1    | 31713   | VAKIO     | 0.0029       | 8356.08167      | -4.81602  | 198(100)                             | 2737.99272                 |
| 1    | 0.4220, 0.1035  | PALKKAMK  | <. 0001      | 4.08736         | 1243732.41  | 235                                  | 0.59320                    |
| 6    | 0.5758  | MALLI     | <. 0001      | 1843320.97      | 1843320.95553   | 3095333                              | 304430.916                 |
| 6    | 5343.50   | VAKIO     | 0.1496       | -425.56836      | 0.01447   | 752(320)                             | 294.573079                 |
| 6    | 0.5744, 0.16515   | HK2       | <. 0001      | 591.54750       | -1252012.03   | 0                                    | 109.203139                 |

**Toimiala F:** Apumuuttujat: HK2

| OTOS | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 5    | 0.8650  | MALLI     | <.0001       | 27718409.4      | 27718409.78172  | 28667956                             | 509498.396                 |
| 5    | 2574.64   | VAKIO     | 0.0015       | -359.05827      | -0.38172  | 28684(11520)                         | 113.205048                 |
| 5    | 0.8650, 0.01838   | HK2       | <.0001       | 369.51584       | -949546.6   | 807                                  | 34.783066                  |

**Toimiala G:** Apumuuttujat: LVMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|--|--------------------------------------|----------------------------|
| 3,5,6 | 0.6050  | MALLI     | <. 0001      | 29793040.9      | 29793084.720996                          | 28778095                             | 731430.376                 |
| 3,5,6 | 2559.77   | VAKIO     | <. 0001      | 354.101544      | -43.82100                                | 45071(18029)                         | 51.2003200                 |
| 3,5,6 | 0.6050, 0.02455   | LVMK      | <. 0001      | 0.067543        | 1014945.9                                | 12440                                | 0.0136164                  |

**Toimiala H:** Apumuuttujat: PALKKAMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.9278  | MALLI     | <. 0001      | 4053612.78      | 4053612.8695185                                       | 4157427                              | 34391.9470                 |
| 3,5,6 | 281.65  | VAKIO     | <. 0001      | 55.7629045      | -0.08952  | 9416(3767)                           | 7.81303431                 |
| 3,5,6 | 0.9277, 0.00848   | PALKKAMK  | <. 0001      | 1.5277321       | -103814.22  | 3504                                 | 0.03857775                 |

**Toimiala I:** Apumuuttujat: PALKKAMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSSASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.2146  | MALLI     | 0.0005       | 15286748.5      | 15286750.659327                                       | 15679787                             | 851227.389                 |
| 3,5,6 | 4086.01   | VAKIO     | 0.0010       | 275.500304      | -2.15933  | 22794(9118)                          | 84.0265910                 |
| 3,5,6 | 0.2145, 0.05568   | PALKKAMK  | 0.0005       | 1.339323        | -393038.5   | 3236                                 | 0.3852561                  |



**Toimiala J:** Apumuuttujat: PALKKAMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUASASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.2372  | MALLI     | <. 0001      | 7719553.61      | 7719548.78222   | 12596235                             | 1514316.04                 |
| 3,5,6 | 20956   | VAKIO     | 0.0291       | 1377.37439      | 4.82778   | 2365(946)                            | 630.164196                 |
| 3,5,6 | 0.2364, 0.19617   | PALKKAMK  | <. 0001      | 3.92689         | -4876681.39   | 429                                  | 0.276308                   |

**Toimiala K:** Apumuuttujat: LVMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUASASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.4724  | MALLI     | <. 0001      | 29584886        | 29584865.523819                                       | 31125354                             | 881144.512                 |
| 3,5,6 | 2812.32   | VAKIO     | 0.0003       | 235.228029      | 20.47618  | 39561(15825)                         | 64.3948900                 |
| 3,5,6 | 0.4723, 0.02978   | LVMK      | <. 0001      | 0.330475        | -1540468  | 3326                                 | 0.0520308                  |

**Toimiala M:** Apumuuttujat: HK2

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUASASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.6681  | MALLI     | <. 0001      | 838731.988      | 838731.9884186  | 894704                               | 66939.4970                 |
| 3,5,6 | 1102.55   | VAKIO     | 0.7826       | -20.613535      | -0.00042  | 1468(588)                            | 74.6876002                 |
| 3,5,6 | 0.6676, 0.07981   | HK2       | <. 0001      | 230.300442      | -55972.012  | 219                                  | 45.6256308                 |

**Toimiala N:** Apumuuttujat: PALKKAMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUASASTEKORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.8838  | MALLI     | <. 0001      | 4068121.52      | 4068121.330821  | 4048848                              | 42685.0032                 |
| 3,5,6 | 341.74  | VAKIO     | <. 0001      | 156.771489      | 0.18918   | 12008(4804)                          | 6.26950300                 |
| 3,5,6 | 0.8837, 0.01049   | PALKKAMK  | <. 0001      | 1.500591        | 19273.52  | 977                                  | 0.06381190                 |

**Toimiala O** Apumuuttujat: PALKKAMK

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSASTE KORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.7586  | MALLI     | <. 0001      | 3851352.52      | 3851352.4868305                                       | 3930011                              | 75854.3337                 |
| 3,5,6 | 459.39  | VAKIO     | <. 0001      | 92.2001196      | 0.03317   | 16595(6639)                          | 5.57624202                 |
| 3,5,6 | 0.7586, 0.01970   | PALKKAMK  | <. 0001      | 1.4733485       | -78658.48   | 771                                  | 0.08974474                 |

**Toimiala X** Apumuuttujat: HK2

| OTOS  | SELITYSASTE<br>GREG-ESTIMAATIN RMSE<br>VAPAUSASTE KORJATTU SELITYSASTE, VARIATIOKERROIN | PARAMETRI | MERKITSEVYYS | GREG-ESTIMAATTI | SYNTEETTINEN ESTIMAATTI<br>HARHANKORJAUSTERMI "HARHA" | TODELLINEN ARVO<br>N(n)<br>PUUTTUVIA | GREG-ESTIMAATIN KESKIVIRHE |
|-------|---|-----------|--------------|-----------------|---|--------------------------------------|----------------------------|
| 3,5,6 | 0.2547  | MALLI     | 0.0401       | 338.432432      | 338.43243212  | 766                                  | 98.5472859                 |
| 3,5,6 | 32.9608   | VAKIO     | 0.1914       | 11.8378378      | 0.00000   | 11(5)                                | 7.5383826                  |
| 3,5,6 | 0.0063, 0.29119   | HK2       | 0.0401       | 86.7567568      | -427.56757  | 0                                    | 28.9576472                 |

## Liite 9. Yritysten tilinpäätöstietojen toimipaikkaistaminen SAS-ohjelmistolla

/\* Kirjasto, jossa aineistot ovat\*/

```
libname otosain 'd:/elina'; run;
OPTIONS VALIDVARNAME=UPCASE pageno=1; run;
```

/\*Toimialakohtaisten totaalin estimaattien laskeminen\*/

```
%macro req2(data.paaluo.otos.v.t.z1.Tz1.z2.Tz2.z3.Tz3.z4.tz4.z5.Tz5);
title1 "SAS SURVEYREG /Tulosmuuttuja y=&y/ Regressioestimointi";
title2 "Aineisto &data. STRSRS otos";
title3 "Tulosmuuttujia y=&y totaalin regressioestimointi toimialalla &paaluo otoksessa &otos, N&paaluo&otos=&t";
title4 "Riippumattomat muuttujat z1=&z1,z2=&z2,z3=&z3,z4=&z4,z5=&z5";
title5 "Riippumattomien muuttujien totaalit perusjoukossa Tz1=&Tz1 Tz2=&Tz2";
title6 "Tz3=&Tz3,Tz4=&Tz4,Tz5=&Tz5";
ods output ParameterEstimates=Toimiala&paaluo&otos;
proc survevreg data=&data total= &t;
where paaluo = &paaluo AND otos=&otos;
model &y = &z1 &z2 &z3 &z4 &z5/ solution;
weight samplingweight;
strata osite;
estimate "&y:n totaali &paaluo toimialalla otoksessa &otos "
Intercept &t &z1 &Tz1 &z2 &Tz2 &z3 &Tz3 &z4 &Tz4 &z5 &Tz5 / E; %mend; run;
```

*/\*Makro, joka lajittelee aineiston haluttujen muuttujien mukaan\*/*

```
%macro sort(data,x,v);
proc sort data=&data;
by &x &y;
%mend; run;
```

```
%macro datat(tulos,data,x);
data &tulos; set &data;
keep &x; %mend; run;
```

*/\*Datojen transponointi\*/*

```
%macro transpoosi(data,out,x);
proc transpose data = &data out=&out; var &x; %mend; run;
```

*/\*Luodaan data, iohon parametrit siirretään, ja siirretään siihen halutun aineiston tieto\*/*

```
%macro data(data,x); data &data; set &x; %mend; run;
```

*/\*Tietojen siirtäminen ja nimien muuttaminen\*/*

```
%macro siirto2(x.a1.paaluo.z.d);
data &x; if n =1 then set &a1;set &x;
rename COL1=&z;rename COL2=&q;
where paaluo=&paaluo; %mend; run;
```

```
%macro siirto3(x,a1.paaluo.yksitoimipaikkainen,z,g);
data &x; if n =1 then set &a1;set &x;
rename COL1=&z; rename COL2=&q;
where paaluo=&paaluo AND yksitoimipaikkainen=&yksitoimipaikkainen; %mend; run;
```

*/\*Turhien tietojen poisto\*/*

```
%macro poisto(data,x,y,z,w);
data &data; set &data;
drop &x &y &z &w; %mend; run;
```

*/\*Imputoitavien havaintoarvojen liputus\*/*

```
%macro liputus(data,tulos,v):data &tulos;
set &data; if &y=. then flag=1; else flag=0; %mend; run;
```

*/\* Imputoidut havaintoarvot \*/*

```
%macro estimaatit4(tulos,data,estimaatti,intercept,estimaatti1,z1,y);
data &tulos; set &data;
if flag=1 then &estimaatti = sum(&intercept + &estimaatti1*&z1);
else &estimaatti=&y; %mend; run;
```

*/\* Estimaatit vastanneiden ioukossa, tarvitaan imputointivarianssin laskemiseen\*/*

```
%macro estimaatit5(tulos,data,estimaatti,intercept,estimaatti1,z1,y);
data &tulos; set &data;
&estimaatti = sum(&intercept + &estimaatti1*&z1); %mend; run;
```

```
%macro estimaatit6(tulos,data,estimaatti,y,jaannos);
data &tulos; set &data;
&jaannos= sum(&y -1*&estimaatti); %mend; run;
```

*/\*Kalibrointi\*/*

```
%macro summat(data.class.out,estimaatti,y);
proc summarv data=&data; class &class;
output out=&out sum(&estimaatti)=&y; %mend; run;
```

```
%macro asetus(data,out);
data &data; set &out; %mend; run;
```

```
%macro yhdistaminen(data,out,class); data &data;
merge &data(in=i) &out(in=j);
by &class; %mend; run;
```

```
%macro yhdistaminen2(data,out,class,yksitoimipaikkainen);
data &data; merge &data(in=i) &out(in=i);
by &class &yksitoimipaikkainen; %mend; run;
```

```
%macro deletion(data);
data &data; set &data; if _type_=0 then delete; %mend; run;
```

```
%macro kalibrointi(data,x,w,v,d);
data &data; set &data; if flag=1 AND &y ne 0 then &d = sum(&x*&w/&y); %mend; run;
```

```
%macro kalibrointi2(data,x,w,v,d);
data &data; set &data; if &y ne 0 then &d = sum(&x*&w/&y) ; %mend; run;
```

*/\*Estimaatit toimialakohtaisesti\*/*

```
%macro estimaatit(data.paaluo.v.estimaatti);
title1 "Tietoja aineistosta &data"; title2"Tulosmuuttuja &v.";
title2 "Suoran mallin tuottamat estimaatit sekä niiden ominaisuuksia";
title3 "&estimaatti=&v:n imputoitu havaintoarvo, on kalibroitu";
proc univariate data = &data normal; var &estimaatti;
by &paaluo; %mend; run;
```

```
%macro iaannostermi(data.paaluo.v.iaannostermi);
title1 "Tietoja aineistosta &data"; title2"Tulosmuuttuja &v.";
title2 "Suoran mallin tuottamat estimaatit sekä niiden ominaisuuksia";
title3 "&iaannostermi=&v:n estimaatin iaännostermi";
proc univariate data = &data normal; var &iaannostermi;
by &paaluo; %mend; run;
```

*/\*Estimaatit koko ioukossa\*/*

```
%macro estimaatit2(data.v.estimaatti);
title1 "Tietoja aineistosta &data"; title2"Tulosmuuttuja &v.";
title2 "Suoran mallin tuottamat estimaatit sekä niiden ominaisuuksia";
title3 "&estimaatti=&v:n estimaatti";
proc univariate data = &data normal; var &estimaatti; %mend; run;
```

```
%macro iaannostermi2(data.v.iaannostermi);
title1 "Tietoja aineistosta &data"; title2"Tulosmuuttuja &v. ";
title2 "Suoran mallin iaännostermi sekä niiden ominaisuuksia";
title3 "&iaannostermi=&v:n estimaatin iaännostermi";
proc univariate data = &data normal; var &iaannostermi; %mend; run;
```

*/\*Asetelmavarianssin laskeminen\*/*

```
%macro asetelmavar(data);
title1 "Jalostusarvon totaalinen asetelmavarianssi"; title2"Data &data";
proc surveymeans data=&data nobs nmiss sum var varsum cvsum mean min max ;
where flag=0; var jaleutp; by paaluo; %mend; run;
```

```
%macro asetelmavar2(data);
title1 "Jalostusarvon totaalinen asetelmavarianssi"; title2"Data &data";
proc surveymeans data=&data nobs nmiss sum var varsum cvsum mean min max ; where flag=0;
var jaleutp; %mend; run;
```

*/\* Imputointivarianssin laskeminen\*/*

```
%macro imputointivar(data,iaannos,paaluo);
title1 "Jalostusarvon totaalimputointivarianssi"; title2 "Data &data";
proc surveymeans data= &data nobs nmiss sum var varsum cvsum mean min max ; var &jaannos;
by &paaluo; %mend; run;
```

```
%macro imputointivar2(data,iaannos);
title1 "Jalostusarvon totaalimputointivarianssi"; title2 "Data &data";
proc surveymeans data= &data nobs nmiss sum var varsum cvsum mean min max;
var &jaannos; %mend; run;
```

*/\* Esimerkki markojen käytöstä, lisätään aineistoon toimialan B tiedot\*/*

```
%req2(otosain.sample3,6,3,JALEUTP,626,LVMK,395244);run ;
%dataf(estimointi6,toimiala63.estimate):run;
%dataf(estimointi61,toimiala63.stderr):run;
%dataf(estimointi62,toimiala63.dendf):run;
%transpoosi(estimointi6,transpoosi6.estimate):run;
%transpoosi(estimointi61,transpoosi61.stderr):run;
%transpoosi(estimointi62,transpoosi62.dendf):run;
%data(metsatalous.otosain.kehikkoperusioukko):run;
%siirto2(metsatalous,transpoosi6,6.INTERCEPT,PLVMK):run;
%siirto2(metsatalous,transpoosi61,6.SINTERCEPT,SLVMK):run;
%siirto2(metsatalous,transpoosi62,6.DENDF,DF):run;
%poisto(metsatalous, name ,DF, LABEL );run;
%liutus(metsatalous.metsatalous,JALEUTP):run;
%estimaatit4(metsatalous.metsatalous.EJALEUTP, INTERCEPT,PLVMK,LVMK,JALEUTP):run;
%estimaatit5(metsatalous.metsatalous.EJALEUTP2, INTERCEPT,PLVMK,LVMK,JALEUTP); run;
%estimaatit6(metsatalous.metsatalous.EJALEUTP2,JALEUTP,VJALEUTP):run;
%yhdistaminen(otosain.suoraimp,metsatalous,paaluo);run;
```

*/\*Kalibrointi\*/*

```
%asetus(suoraimp,otosain.suoraimp); run;
%sort(suoraimp,vrtunnus); run;
%summat(suoraimp,vrtunnus.summat.EJALEUTP,SJALEUTP); run;
%summat(suoraimp,vrtunnus.summat2,EJALEUTP2,SJALEUTP2); run;
%yhdistaminen(suoraimp,summat,vrtunnus); run;
%yhdistaminen(suoraimp,summat2,vrtunnus); run;
%deletion(suoraimp); run;
%kalibrointi(suoraimp,EJALEUTP,JALEU,SJALEUTP,EJALEUTP); run;
%kalibrointi2(suoraimp.EJALEUTP2..JALEU.SJALEUTP2.EJALEUTP2); run;
%estimaatit6(suoraimp,suoraimp,EJALEUTP2,JALEUTP,VJALEUTP); run;
```

*/\* Estimaatit kalibroidussa datassa\*/*

```
%sort(suoraimp,paaluo.vksitoimipaikkainen); run;
%estimaatit2(suoraimp,JALEUTP,EJALEUTP); run;
%estimaatit(suoraimp,paaluo,JALEUTP,EJALEUTP); run;
%iaannostermi2(suoraimp..JALEUTP,VJALEUTP); run;
%iaannostermi(suoraimp,paaluo,JALEUTP,VJALEUTP); run;
%asetelmavar2(suoraimp);run;
%imputointivar2(suoraimp,VJALEUTP); run;
%asetelmavar(suoraimp); run;
%imputointivar(suoraimp,VJALEUTP,paaluo); run;
```

*/\* Siirretään kalibroitu data talteen\*/*

```
%asetus(otosain.suoraimp,suoraimp); run;
```