

## *Mikrosimuloinnin rekisteripohjainen tutkimusaineisto*

### SISÄLLYS

1. REKISTERIPOHJAISEN TUTKIMUSAINIESTON YLEISKUVAUS .....	2
2. OTANTA, VÄESTÖ JA PAINOTUS.....	3
2.1 Otanta .....	3
2.2 Perusjoukko.....	3
2.3 Luonne ja suhde tilastoihin .....	4
3. AINEISTOON LIITTYVÄT MUOKKAUKSET .....	4
3.1 Asuntokuntiin liittyvät tiedot .....	5
3.1.1 Uudelleenjärjestellyt asuntokunnat .....	5
3.1.2 Puolisot verotuksessa .....	5
3.1.3 Suhdematriisi .....	5
3.1.4 Kiinteistörekisterin tietopohjan puutteet .....	6
3.2 Talouksien rakennetiedot ja henkilöiden sosioekonomiset tiedot ..	6
3.2.1 Talouden viitehenkilö .....	7
3.2.2 Suhde viitehenkilöön.....	7
3.2.3 Perhenumerointi.....	7
3.2.4 Talouden rakenne ja elinvaihe .....	8
3.2.5 Talouden kulutusyksikköjen määrä.....	8
3.2.6 Sosioekonominen asema sekä toimintatiedot .....	8
3.3 Menetelmällisesti tuotetut tiedot .....	9
3.3.1 Systemaattiset nollatiedot.....	9
3.3.2 Asumiskustannukset.....	9
3.3.2.1 Vuokratiedot .....	10
3.3.2.2 Sähkölaskut.....	11
3.3.2.3 Lämmityskustannukset .....	11
3.3.2.4 Hoito- ja yhtiövastikkeet .....	11
3.3.3 Elatusmaksuihin liittyvät tiedot .....	11
4. Aineiston järjestys (sort) ja suhde simulointiin .....	12
LIITTEET .....	13
LIITE1. Vuokratietojen imputointi .....	13
LIITE2. Pohja-aineistojen muuttujakuvaukset .....	13

## 1. Rekisteripohjaisen tutkimusaineiston yleiskuvaus

Mikrosimuloinnin rekisteriaineisto on SISU-mikrosimulointimallia varten koottu tutkimusaineisto. Aineisto kattaa vuosittain noin 15 prosenttia Suomen asuntokunnista ja näin ollen myös maassa vuoden lopussa vakituisesti asuneista henkilöistä. Aineisto ei ole tilastoaineisto, eikä sillä pystytä kaikin osin pääsemään tilastojen kanssa yhtenäisiin tuloksiin. Poikkeamat tilastoissa esitettyjen tulosten ja mikrosimuloinnin rekisteriaineiston välillä johtuvat pääasiassa korjauksista, joita rekisteripohjaisiin tietoihin on tehty mikrosimuloinnin tarkkuuden parantamiseksi.

Mikrosimuloinnin rekisteriaineisto on koottu erilaisista hallinnollisista aineistoista sekä rekistereistä. Näistä merkittävimmät lähteet ovat:

- Väestörekisterikeskuksen väestötietojärjestelmä ja Tilastokeskuksen henkilö- ja kiinteistötilastojen tietokannat
- Verohallituksen verotietokanta
- Kansaneläkelaitoksen rekisteri- ja maksutietojärjestelmäpohjaiset tiedot (eläkevakuutus-, sairauskorvaus-, kuntoutus-, opintotuki- ja asumistuki-rekisteri sekä erilaisiin etuuksiin liittyvät tiedot maksujärjestelmästä)
- Terveiden ja hyvinvoinnin laitoksen toimeentulotukirekisteri
- Eläketurvakeskuksen eläketapahtumarekisteri
- Tilastokeskuksen tutkintorekisteri
- Valtiokonttorin sotilasvammakorvausjärjestelmän tietokanta
- Koulutusrahaston tiedot

Aineisto koostuu ainoastaan hallinnollisista aineistoista ja rekistereistä sekä menetelmällisin keinoin tuotetuista tiedoista. Menetelmällisesti tuotetut tiedot on tavanomaisimmin tuotettu aineistoon imputoimalla Tilastokeskuksen vastaavan aihealueen tilastotietoja rekistereissä olevilla ominaisuuksilla luokiteltuina.

Rekisteriaineiston yksikkötasot ovat henkilö ja asuntokunta (sekä perhe). Asuntokunnan muodostavat samassa asuinhuoneistossa vakinaisesti asuvat henkilöt. Vakinaisesti laitoksessa kirjoilla olevat, asunnottomat ja sijainniltaan tuntemattomat henkilöt eivät tilastoissa muodosta asuntokuntia, ja ovat näin ollen asuntokuntaväestön ulkopuolella. Mikrosimuloinnin rekisteriaineistossa nämä henkilöt muodostavat yhden hengen asuntokuntia. Asuntokuntakäsite mikrosimuloinnin rekisteriaineistossa poikkeaa siis hieman tilastojen käsitteestä. Tämä poikkeama mahdollistaa kuitenkin edustavan kuvauksen Suomen väestöstä kokonaisuudessaan, myös niiltä osin, kun henkilöt eivät kuulu asuntokuntaväestöön. Näkyvin hyöty tästä ilmenee mm. eläkkeisiin liittyvien tulojen tarkempana jakaumana aineistossa (esimerkiksi pysyvässä laitoshoidossa olevat vanhukset muodostavat osan eläketulojakaumasta). Aineistossa perheen muodostavat yhdessä asuvat avio- tai avopuolisot ja heidän lapsensa, jompikumpi vanhemmista lapsineen sekä lapsettomat avio- tai avopuolisot.

Rekisteriaineiston tietosisältö on kuvattu muuttujatasolla liitteessä 2. Kaikkien luotujen muuttujien täydellisiä päättelysääntöjä ei kuitenkaan ole koko-

naisuudessaan kuvattu. Näistä saa tarvittaessa lisätietoa ottamalla yhteyttä Tilastokeskuksen mikrosimulointitiimiin.

## 2. Otanta, väestö ja painotus

### 2.1 Otanta

Mikrosimuloinnin rekisteriaineisto on poikkileikkausaineisto. Otokskehikoon kuuluvat kaikki tilastovuoden lopussa Suomessa tietyvästi olleet henkilöt. Väestöllä tarkoitetaan alueella (esim. koko maa, lääni, kunta) vakinaisesti asuvaa väestöä. Ne henkilöt, joilla väestötietojärjestelmän mukaan oli kotipaikka Suomessa vuodenvaihteessa 31.12., kuuluvat ko. vuoden väestöön kansalaisuudestaan riippumatta; samoin ne Suomen kansalaiset, jotka asuvat tilapäisesti ulkomailla.

Otokskehikossa yksikkö on asutokunta. Asutokuntaväestön ulkopuolinen väestö on kehitossa edustettuna yhden hengen asutokuntina. Kehikosta tehdään vuosittain noin 15 prosentin yksinkertainen satunnaisotanta, jonka seurauksena otokseen valikoituu noin 400 000 asutokuntaa. Kaikki näiden asutokuntien jäsenet kuuluvat aineistoon, jolloin aineiston vuosittainen henkilölukumäärä on noin 800 000 henkeä. Otantaa ei ole ositettu tai kontrolloitu.

Seurauksena yksinkertaisesta satunnaisotannasta aineisto tuottaa varsin luotettavat jakaumat jo asetelmapiainojen tilanteessa. Otosta voidaan periaatteessa kuvata enemmänkin mallipopulaationa. Kaikilla havainnoilla on lähtökohdaltaan identtinen asetelmapiaino eikä perusvuoden aineistoa ole tarvetta kalibroida.

### 2.2 Perusjoukko

Mikrosimuloinnin rekisteriaineisto on otos Suomen väestöstä. Väestöllä tarkoitetaan tässä yhteydessä Suomessa vakinaisesti asuvia henkilöitä, joilla oli tilastovuoden viimeisenä päivänä kotipaikka Suomessa. Suomen kansalaiset, jotka asuvat tilapäisesti ulkomailla, kuuluvat väestöön. Ulkomaalaisella henkilöllä on kotipaikka Suomessa, jos hänen oleskelunsa on tarkoitettu kestämään tai on kestänyt vähintään vuoden. Turvapaikanhakija saa kotipaikan vasta, kun hänen turvapaikkahakemuksensa on hyväksytty. Suomessa asuvat vieraan valtion lähetystöön tai vastaavaan kuuluvat henkilöt ja heidän perheenjäsenensä eivät kuulu Suomen väestöön, elleivät he ole Suomen kansalaisia. Suomen ulkomailla asuvat lähetystöjen tai vastaavien henkilöt kuuluvat Suomessa asuvaan väestöön. Näin ollen tilapäisesti tai laittomasti maassa olevat henkilöt eivät kuulu väestöön.

Vaikka mikrosimuloinnin rekisteriaineiston talousyksikkönä käytetään asutokuntaa, kattaa aineisto myös asutokunnan ulkopuolisen väestön. Esimerkiksi vuoden 2011 väestö- ja asutokuntatilastoja tarkasteltaessa voidaan huomata, että asutokuntaväestöön kuului 5 282 429 henkeä, kun koko väestö saman vuoden lopussa oli 5 401 267 henkeä. Noin 118 800 henkeä, jotka eivät siis kuulu syystä tai toisesta asutokuntaväestöön, kuuluvat kuitenkin edellä mainitulla tavalla mikrosimuloinnin rekisteriaineistoon. Heille ei kui-

tenkaan ole olemassa asumiskiinteistötietoja, ja tästä syystä esimerkiksi harkinnanvaraisten etuuksien simuloinnin osalta on syytä noudattaa varovaisuutta asuntokuntaväestön ulkopuolisen väestön osalta.

Vuoden 2011 lopussa varsinaista laitospääväestöä oli noin 81 400 henkeä. Laitospääväestö koostuu mm. vailla vakinaista asuntoa olevista, sairaalaan tai huoltolaitokseen sijoitetuista, huostaan otetuista, olinpaikaltaan tuntemattomista, tilapäisesti ulkomailla olevista sekä vankilassa ja rangaistuslaitoksissa asuvista henkilöistä. Laitospääväestön lisäksi asuntokuntaväestön ulkopuolelle jää eräitä väestöryhmiä, joiden asumismuoto karsitaan tilastossa. Tällaisia ovat mm. asuntolarakennuksissa asuvat henkilöt, joiden asunto ei täytä asunnon määritelmää (keittiöllä, keittokomerolla, tai keittotilalla varustettu vähintään 7 neliömetrin ympärivuotiseen asumiseen tarkoitettu kokonaisuus) sekä ylisuuret asuntokunnat (yli 10 henkeä, ei lapsia tai yli 25 henkeä, sisältäen lapsia).

Asuntokuntaväestön ulkopuolinen väestö voidaan rajata pois aineistosta tarvittaessa asuntokuntaväestötunnusta (asmu) käyttämällä. Tällöin mallipopulaation painokertoimet korottavat otoksen edustamaan asuntokuntaväestöä.

Muutamista mikrosimuloinnin rekisteriaineiston asuntokuntiin liittyvistä poikkeamista on kerrottu luvussa 3.

## 2.3 Luonne ja suhde tilastoihin

Mikrosimuloinnin rekisteriaineisto on tutkimusaineisto, joka on koottu ja muokattu erilaisista rekistereistä sekä hallinnollisista aineistoista. Sen perusjoukko ei itsessään vastaa tilastojen perusjoukkoja. Mikäli asuntokuntaväestöön kuulumattomat henkilöt (muuttujan asmu arvon nollan perusteella) poistetaan aineistosta, vastaa aineisto tällöin perusjoukoltaan tulonjaon kokonaistilastoa, joka siis kuvaa asuntokuntaväestöä. Muutamat poikkeukset tässä kohtaa liittyvät asuntokunnan rakenteisiin tehtyihin pieniin muutoksiin. Aineisto tuottaa ilman laitospääväestöä kuitenkin varsin tarkasti tulonjaon kokonaistilaston lukuja vastaavia tuloksia (tulo-osuudet, gini-kertoimet).

## 3. Aineistoon liittyvät muokkaukset

Aineistoon ei ole tehty mittavia manuaalisia korjauksia. Tehdyt editoinnit ovat lähinnä koneellisia ratkaisuja. Käytännössä rekisteritietoihin liittyvät muutokset jakautuvat kolmeen osaan: asuntokuntayksikköön ja asuinkiinteistöön liittyviin tietoihin tehtäviin korjauksiin, olemassa olevien hallinnollisten ja rekisteriaineistojen pohjalta päätettäviin tietoihin sekä menetelmällisesti tuotettuihin tietoihin.

Näistä osittain tai kokonaan menetelmällisesti tuotetut tiedot (mm. osa asumiskustannuksista) edustavat luonnollisesti aineiston karkeinta osaa. Tällaisia tietoja on tuotettu menetelmällisesti siinä tapauksessa, että niiden on katsottu olevan merkittävä edellytys jonkin tietyn etuusjärjestelmän mallintamisen takia. Tietoja, jotka eivät vaikuta merkittävästi varsinaiseen simulointiin vaan ainoastaan talouksien tuloihin (mm. kotitalouksien puolesta maksetut laskut), ei ole pääasiassa estimoitu. Tällaisia tietoja on kuitenkin mahdolli-

suuksien mukaan johdettu rekisteriaineistosta (mm. maksetut ja saadut elatusavut).

### 3.1 Asuntokuntiin liittyvät tiedot

Asuntokuntatason mikrosimulointiin liittyy rekisteripohjaisesta tiedosta johtuen kaksi ongelmaa. Ensimmäinen liittyy kiinteistörekisterin tietojen puutteellisiin ja toinen asuntokuntien luonteeseen. Näitä ongelmia on korjattu mikrosimuloinnin rekisteriaineistossa imputoimalla tietoja ja järjestelmällä asuntokuntia uudelleen.

#### 3.1.1 Uudelleenjärjestellyt asuntokunnat

Eräät poikkeavat asumismuodot muodostavat tietyissä tapauksissa asuntokuntia, jotka eivät välttämättä muodosta lainsäädännön kannalta koherentteja kokonaisuuksia. Tällaisia ovat muun muassa eräät opiskelija- sekä vanhusasunnot. Tällaisissa tapauksissa on mahdollista, että yhteen asuntokuntaan kohdistuu lainsäädännön näkökulmasta useita erillisiä talousyksikköjä. Esimerkiksi kuuden eläkeläisen (jotka eivät ole sukua toisilleen) muodostamassa asuntokunnassa viisi eläkeläistä voi tulojen ja asumiskustannustensa puolesta nauttia eläkkeensaajan asumistukea. Asuntokuntamielessä he ovat kuitenkin (esimerkiksi palveluasuntolan tai vastaavan johdosta) yhtä asuntokuntaa.

Tällaiset asuntokunnat on uudelleenjärjestetty mikrosimuloinnin rekisteriaineistossa. Asuntokunnat, jotka koostuvat kolmesta tai useammasta jäsenestä (eivät sisällä lapsia) ja jotka eivät ole sukua toisilleen, on jaettu uudelleen. Jokaisesta asuntokunnan jäsenestä on tällöin muodostettu asuntokuntaväestöön kuuluva yhden hengen asuntokunta. Alkuperäiset kiinteistötiedot on tässä kohtaa jyvitetty henkilöille: esimerkiksi kahdeksan eläkeläisen 400 asuinneliön asunto on jaettu tasan kahdeksaksi viidenkymmenen neliön asunnoksi. Myös huonelukumäärä on jaoteltu. Muut asuntokunnan ominaisuudet (lämmitystyyppit, hallintaperusteet) on säilytetty ennallaan. Tällaiset tapaukset muodostavat aineistoon muutaman tuhatta uutta asuntokuntaa. Tämä aiheuttaa poikkeamia tilaston lukuihin verrattuna. Toisaalta esimerkiksi etuuksien allokointia ja tulonjakoa (mikrosimuloinnin keskeisiä kiinnostuksen kohteita) ajatellen tämä antaa realistisemman kuvan tilanteesta.

#### 3.1.2 Puolisot verotuksessa

Otoshenkilöiden verorekisterin tietojen mukaiset puolisot on merkattu omaan tietuekenttäänsä otoksen henkilönumerin. Näin henkilönumero, joka otoshenkilöllä on verotuksen puolisoa kuvaavassa tietuekentässä, viittaa siihen henkilöön, joka verotuksessa on tämän henkilön puoliso.

#### 3.1.3 Suhdematriisi

Asuntokunnan suhdematriisi on erillistaulukko, joka on rakennettu mahdollisten perhe- ja suhderakenteiden tarkempaa tarkastelua varten. Käytännön tasolla sen päättelyt on rakennettu samalla tapaa kuin suhde viitehenkilöön -muuttujan sisältö, mutta se on tehty jokaiselle asuntokunnan henkilöl-

le. Näin JSKO -muuttujia on suurimman asutokunnan jäsenmäärää vastaava luku. JSKO1 -muuttuja vastaa ASKO -muuttujaa, ja loput JSKO -muuttujat ovat asutokunnan muiden jäsenten suhteita kyseiseen jäseneseen. Esimerkiksi JSKO5 -muuttuja kertoo asutokunnan jäsenten suhteen asutokunnan viidenteen jäseneseen. Suhdematriisi on toistaiseksi koeluontoinen aineiston lisäosa, ja sitä voidaan käyttää tarvittaessa asutokunta- ja perherakenteiden tarkempaan tutkimiseen.

### *3.1.4 Kiinteistörekisterin tietopohjan puutteet*

Kiinteistörekisterin tiedot asuinkiinteistöistä eivät ole kaikin paikoin täysin kattavat. Puutteita esiintyy asuineliöissä, hallintaperusteissa, rakennusvuosissa sekä asuinhuoneiden lukumäärissä. Puuttuvat tiedot on imputoitu aineistoon.

Mikäli asutokunnan tiedoista on puuttunut sekä pinta-ala että asuinhuoneiden lukumäärä, on näistä korjattu ensin asuinpinta-ala. Asuinpinta-ala on mallinnettu yksinkertaisella regressiomallilla populaation tiedoista asutokunnan jäsenmäärän perusteella, käyttäen kuntaa ositteena. Tämän jälkeen huoneiden lukumäärä on imputoitu vastaavalla tavalla. Mikäli vain asuinneiliöt tai huoneiden lukumäärä on puuttunut, ne ovat vastaavasti imputoitu toistensa perusteella jälleen populaation tietojen perusteella, kuntia ositteina käyttäen.

Puuttuvat rakennuksen valmistumis- tai peruskorjausvuositiedot on imputoitu aineistoon populaatiotason kuntakeskiarvojen perusteella. Mikäli asunnon hallintaperuste on puuttunut, se on imputoitu asunnossa asuvien henkilöiden tietojen perusteella seuraavasti: mikäli henkilöillä on ollut asutolainaa, asunto on merkattu sen tyyppin perusteella joko osakeomistus- tai omistusasunnoksi. Mikäli asutolainaa ei ole ollut, asunto on merkattu vuokra-asunnoksi.

Asutokuntaväestön ulkopuolisen väestön mahdolliset (hyvin harvinaiset) asuinkiinteistötiedot on poistettu niiden epäluotettavuuden vuoksi. Asutokuntaväestön ulkopuolisella väestöllä ei siis ole asuinkiinteistötietoja. Näin ollen mm. harkinnanvaraisia etuuksia simuloitaessa (asumistuki, toimeentulotuki) tämä on syytä huomioida.

## *3.2 Talouksien rakennetiedot ja henkilöiden sosioekonomiset tiedot*

Talouksien rakennetta sekä henkilön sosioekonomista asemaa kuvaavat tiedot (toimintatiedot, sosioekonominen asema vuosikuvauksena) on muodostettu aineistoon käyttäen hyväksi olemassa olevia rekisteriaineistoja. Tärkeimmät lähteet liittyvät tässä kohtaa rakennetietojen osalta Tilastokeskuksessa olevaan tietoon lapsien ja vanhempien välisestä linkistä, talouden henkilöiden tulotietoihin sekä erinäisiin jaksotietoihin (työttömyysjaksot työ- ja elinkeinoministeriön tiedoista, eläkejaksot eläketurvakeskuksen tiedoista, ym.).

### 3.2.1 Talouden viitehenkilö

Talouden viitehenkilöksi on mikrosimuloinnin rekisteriaineistossa katsottu se henkilö, jonka nettotulot ovat vuoden aikana olleet korkeimmat. Nettotuloissa huomioidaan palkkatulot, yrittäjätulot, omaisuustulot sekä saadut tulonsiirrot. Tulojen jälkeen on priorisoitu henkilöitä iän mukaan niin, että vanhempi henkilö on viitehenkilö. Asuntokuntaväestön ulkopuolinen väestö muodostaa aineistossa yhden hengen asuntokuntia, josta johtuen he ovat aina oman taloutensa viitehenkilöitä. Tämä aiheuttaa aineistoon jonkin verran tapauksia, joissa talouden viitehenkilön ikä ja sosioekonomiset tiedot eivät vaikuta järkeviltä. Nämä tapaukset ovat seurausta muun muassa huostanotettujen lasten (laitosväestöä) käsittelystä yhden hengen asuntokuntina.

### 3.2.2 Suhde viitehenkilöön

Suhde talouden viitehenkilöön on päätelty lasten ja vanhempien välisten linkkitietojen perusteella. Viitehenkilölle on ensin haettu tämän puoliso perheistolaston tiedoista. Tämän jälkeen talouden viitehenkilölle ja mahdolliselle puolisolalle on haettu tietoa lapsista, vanhemmista ja sisaruksista. Nämä tiedot on koodattu ASKO -muuttujaan. Tietolähde on pääsääntöisesti varsin luotettava: ongelmat liittyvät lähinnä sellaisiin tapauksiin, joissa henkilö on muuttanut pois kotoaan ennen vuoden 1970 väestölaskentaa. Tällaisessa tapauksessa henkilölle ei välttämättä löydy sukulaisia. Tapaukset, joissa tällaiset henkilöt asuisivat vielä samassa asuntokunnassa, lienevät kuitenkin nykyisellään harvinaisia.

Puolison määrittäminen perustuu perhekäsitteen mukaiseen päättelyyn (luku 3.2.3). Tämä saattaa aiheuttaa tiettyjen eri sukupuolta olevien samanikäisten henkilöiden merkinnän avopariksi, vaikkei näin todellisuudessa olisi. Toisaalta, tämä voi myös jättää huomioimatta sellaiset todelliset avoparit, joiden ikäero on suuri. Ainakin jälkimmäisiä tapauksia on havaittu verrattaessa rekisteritietoa haastattelutietoon. Lapsiin ja vanhempiin liittyvät suhteet eivät kaikissa tapauksissa ole täydellisiä.

Asuntokuntien jäsenten puoliset, äidit ja isät on rakennettu aineistoon suhdematriisin pohjalta. Kunkin henkilön vastaavaan kenttään on viety jäsennumero viitaten siihen asuntokunnan jäseneseen, joka on kyseisen henkilön äiti, isä tai puoliso.

### 3.2.3 Perhenumerointi

Perhenumeroinnilla pyritään tuottamaan aineistoon tieto sellaisista tapauksista, joissa sama asuntokunta sisältää useita perheyksiköitä. Perheet on muodostettu Tilastokeskuksen väestötilastojen tietojen perusteella. Käsitteellisesti perhe koostuu yhdessä asuvista rekisteröidyssä parisuhteessa, avio- tai avoliitossa elävistä henkilöistä sekä heidän lapsistaan. Myös yksinhuoltajat ja heidän lapsensa sekä lapsettomat parit muodostavat perheitä.

Henkilöiden sisarukset tai muut sukulaiset eivät kuitenkaan muodosta perheitä: näin ollen esimerkiksi sisarusten yhteisasunto muodostaa yhden asuntokunnan, jonka sisällä on kaksi perheyksikköä. Perheessä voi olla peräkkäin korkeintaan kaksi peräkkäistä sukupolvea: useamman sukupolven perheissä



perhe muodostetaan nuorimmasta sukupolvesta lähtien. Esimerkiksi toisen puolison vanhempien kanssa lapsineen asuva pari muodostaa yhden asuntokunnan, joka käsittää kaksi erillistä perheyksikköä (nuoremman parin, joilla lapsia sekä vanhemman parin).

Avopareiksi on päätelty sellaiset tapaukset, jossa on kaksi vastakkaista sukupuolta olevaa täysi-ikäistä henkilöä (eivät ole sisaruksia), joiden ikäero on korkeintaan 15 vuotta, ja jotka asuvat yhdessä vakituisesti samassa asunnossa.

Perhenumero on muodostettu juoksevasti asuntokuntien sisälle niin, että sitä on mahdollista käyttää asuntokunnan numeron sijaan simuloinnissa talouden yksilöivänä yksikkönä. Kullakin aineiston perheellä on siis uniikki perhetunnus. Näin ollen perhetunnuksia on hieman enemmän kuin asuntokuntatunnuksia.

### *3.2.4 Talouden rakenne ja elinvaihe*

Talouden rakenteet on tuotettu aineistoon iän, sukupuolen ja väestötilastojen perhetietojen perusteella. Lapseksi on luettu jommankumman vanhemman (tai molempien) kanssa asuvat biologiset sekä adoptoidut lapset. Tämän jälkeen lapsille on muodostettu ikäluokitus. Talouden elinvaihe ja rakennetiedot ovat sen jälkeen koodattu näiden tietojen taloustasojen määrien perusteella. Esimerkiksi yksin asuvaa alaikäistä henkilöä ei kuitenkaan ole näin ollen laskettu lapseksi.

### *3.2.5 Talouden kulutusyksikköjen määrä*

Talouksille on muodostettu kaksi kulutusyksikköskaalaa - modoeed -skaala sekä kulutusyksikkö -skaala. Näissä skaaloissa yli 13-vuotiaat henkilöt on katsottu aikuisiksi (kulutusyksiköinä) ja sitä nuoremmat lapsiksi (kulutusyksiköinä).

Skaalan painot ovat modoeed -kulutusyksiköillä 1 ensimmäiselle aikuiselle, 0,5 muille aikuisille ja 0,3 lapsille. Kuluyks -kulutusyksiköillä skaalan painot ovat 1 ensimmäiselle aikuiselle, 0,7 muille aikuisille ja 0,5 lapsille. Lisäksi aineistossa on skaala aikuisen asemassa oleville taloudessa.

Asuntokuntaväestön ulkopuolella on jonkin verran yhden hengen asuntokuntia, joissa ei välttämättä ole yli 13-vuotiaita henkilöitä. Näille talouksille on asetettu kulutusyksikköjen määräksi 1. Nämä talousyksiköt ovat seurausta laitosväestöstä sekä asuntokuntaväestön ulkopuolisesta väestöstä, johon kuuluu jonkin verran alaikäisiä henkilöitä.

### *3.2.6 Sosioekonominen asema sekä toimintatiedot*

Mikrosimuloinnin rekisteriaineistossa on kaksi sosioekonomista asemaa kuvaavaa muuttujaa: työssäkäyntitilaston sosioekonominen asema, joka kuvaa henkilön toimintaa vuoden viimeisenä päivänä, sekä mikrosimuloinnin sosioekonominen asema, joka pyrkii kuvaamaan henkilön vuoden yli jatkunutta pääasiallista toimintaa. Työssäkäyntitilaston sosioekonominen asema perustuu priorisoituun päättelyyn erilaisista hallinnollisista aineistoista. Mik-



rosimuloinnin sosioekonomisessa asemassa hyödynnetään henkilön tulo- ja jaksotietoja aseman määrittelemiseksi.

Toimintatiedot ovat rekisteriaineiston osalta hieman vaillinaiset. Ne toimintakuukausitiedot, jotka aineistoon on kuitenkin kuvattu, perustuvat pääasiassa hallinnollisten jaksoaineistojen kautta laskettuihin aikamääriin (ETK:n tapahtumarekisteri, työministeriön kautta tulevat työttömyystiedot, KELAn kotihoidon tuen tiedot, ym.). Tämän lisäksi näihin tietoihin on tehty eräitä loogisuuskorjauksia (esimerkit loogisuuskorjauksista: yli 65-vuotias henkilö, jolle ei ole merkattu mitään toimintaa on päätelty tulomuuttujien valossa eläkeläiseksi; korkeat ansiotulot omaavat henkilöt, joiden työssäkäyntikuukausitiedot eivät ole loogisia, on päätelty täyden vuoden työtä tehneiksi, olettaen että heillä ei ole eräitä työttömyyteen ja perusturvaan liittyviä etuustuloja).

### 3.3 Menetelmällisesti tuotetut tiedot

Rekisteriaineiston menetelmällisesti tuotetut tiedot ovat simulointimallien kannalta olennaisia tietoja, joita rekisteri- tai hallinnolliset aineistot eivät tarjoa joko osittain tai lainkaan. Näitä tietoja on pyritty imputoimaan pääasiassa erilaisista tilastolähteistä käyttäen hyväksi mahdollisuuksien mukaan hyväksi rekisteri- ja lähdeaineiston tietoja. Menetelmällisesti tuotetut tiedot on merkattu aineiston imputoinnit kartoittavaan erityistauluun.

#### 3.3.1 Systemaattiset nollatiedot

Eräitä tietoja, joilla ei ole välttämättä vaikutusta mikrosimulointiin, ja joiden estimoiminen olisi haastavaa, on jätetty tyhjiksi. Näitä edustavat muun muassa kotitalouksien välisiä rahasiirtoja kuvaavat tiedot, joiden vaikutus on pääasiassa toimeentulotuen mallintamisessa. Nämä tiedot ovat kuitenkin pääasiassa luonteeltaan sellaisia, että niiden vaikutusta simuloinnin tuloksiin voidaan pitää vähäisinä. Tällaisten tietojen muuttujakuvauksessa lukee merkintä (NOLLA). Ne esiintyvät muuttujina aineistossa, jotta ohjelmakoodiin, jonka tarkoitus on toimia sekä tulonjaon palveluaineistolla että mikrosimuloinnin rekisteriaineistolla, ei tarvitse tehdä muutoksia näiden muuttujien puuttumisen johdosta.

Mallin tuottamien tulonjakoindikaattoreiden ja laajemmin tulokäsitteiden osalta on erityisen tärkeää huomioida, että rekisteriaineisto ei sisällä tietoja laskennallisesta asuntotulosta, jonka vuoksi ne puuttuvat käytettävissä olevien tuloista rekisteriaineistoa käytettäessä. Lisäksi rekisteriaineisto ei sisällä toistaiseksi tarvittavia tietoja päivähoitomaksujen simuloimiseksi, jonka vuoksi tuloksena syntyy tyhjiä tietoja.

#### 3.3.2 Asumiskustannukset

Asumiskustannukset ovat mahdollisesti suurin erä, jota ei saada kattavasti hallinnollisista- ja rekisteritiedoista. Asumiskustannukset kuitenkin koskevat kolmea (neljää; mikäli opintotuki huomioidaan) SISUn osamallia: yleistä asumistukea, eläkkeensaajan asumistukea ja toimeentulotukea. Toistaiseksi ainoa rekisterilähde näille tiedoille on KELAn asumistukirekisteri, joka kattaa KELAn myöntämien etuuksien piirissä olevien henkilöiden tietoja. Tästä

aineistosta saadaan poikkileikkaus, joka kuvaa vuoden vaihteen tilannetta eli mallipopulaation leikkaushetkeä. Asumistukirekisteri ei kuitenkaan ole kovin kattava lähde, ja se kuvaa luotettavasti lähinnä vuokratietoja.

### 3.3.2.1 Vuokratiedot

Asumistukirekisteristä on valikoitu tiedot rekisteriaineistoon vertailemalla asumistukirekisterin tietoja kiinteistörekisterin tietoihin. Neliömääräisesti tai sijainniltaan merkittävästi poikkeavat tiedot on jätetty pois. Tiedot, jotka vaikuttavat kohdentuvan korrektiin talouteen, on otettu käyttöön. Periaatteessa asumistukirekisterin tulisi sisältää vain yhden rivin per ruokakuntaan kuuluva henkilö, koska tukea haetaan ruokakuntaa edustavana. Poikkeuksia tähän kuitenkin löytyy mahdollisesti käsite-erojen takia. Tapaukset, joissa asutokunnan varsinaisia tuensajia on ollut vaikea loogisesti päätellä, on jätetty pois, sillä ne aiheuttavat muutoin paikoittain huomattavankin matalia vuokratasoja. Esimerkiksi kaksi opiskelijaa, jotka saavat samaan asutokuntaan tukea, muodostavat loogisen yksikön vuokran kannalta, mutta kaksi henkilöä, joista toinen saa tukea ja toinen ei, eivät välttämättä muodosta loogista yksikköä samassa mielessä.

Vuokratieto on siis luovutettu sellaisenaan niille asutokunnille, joissa yhdistäminen on toiminut loogisesti. Tämä kattaa pääosan asumistuen piirissä olevista asutokunnista. Puutteelliseksi jää siis osa asumistuen piirissä olevista asutokunnista sekä muut vuokralla asuvat taloudet.

Tieto kuukausivuokrasta on imputoitu näille havainnoille matching -menetelmällä asumistukirekisterin tietojen perusteella. Asumistukirekisteristä on valittu aiemmin loogisen päättelyn läpäisseet tiedot ja korjatut, joista on poistettu poikkeavimmat havainnot. Tämä joukko muodostaa tietoja luovuttavat havainnot. Tiedot on imputoitu matching -menetelmällä, käyttäen etäisyyden mittarina yksinkertaista euklidista etäisyysfunktiota standardoidusta pinta-alasta sekä asuinhuoneiston rakennuksen itä- ja pohjoiskoordinaateista. Näin ollen malli preferoi ensisijaisesti samassa asuinrakennuksessa olevia samankokoisia huoneistoja, ja sen jälkeen lähiseudulla olevia samankokoisia huoneistoja. Tapauksissa, joissa asunnon ominaisuudet ovat huomattavan erityiset (mm. yli 300 neliön vuokra-asunnot), imputointi voi tuottaa kaukaisiakin matcheja. Koska naapuruutta ei ole rajattu joukoittain (esimerkiksi kunnittain), vain se on laskettu koko maan tasolla, voi luovuttaja olla metrisesti kaukanakin vastaanottajasta, mikäli kiinteistön luonne on erikoinen. Kuntarajauksen poisjättämistä puoltaa kuitenkin pienten kuntien joukkojen vähäisyys. Antamalla mallin hakea luovuttajia koko populaatiosta etäisyysfunktio preferoi lähistöllä olevia asuinhuoneistoja, joten esimerkiksi kaksi eri kunnissa olevaa mutta fyysisesti lähellä toisiaan olevaa asuntoa voivat olla luovuttaja ja vastaanottaja siitä huolimatta, että ne ovat eri kunnissa.

Matching -menetelmää on arvioitu tarkemmin erillisessä dokumentissa liitteessä 1.

### 3.3.2.2 Sähkölaskut

Sähkölaskuista ei ole olemassa selkeää tietoa. Luotettavimmat lähteet ovat tyyppikäyttäjämalleja. Sähkölaskuja imputoitaessa on käytetty sähköyhtiön tyyppikäyttäjämallia, joka luokittelee asutokunnat jäsenmäärän, lämmitystyyppin, saunan ja huonelukumäärän perusteella. Keskimääräinen vuosikulutus on imputoitu näitä tietoja käyttäen, jonka jälkeen sähkölasku on laskettu tästä vuosikulutuksesta käyttäen energiamarkkinaviraston tietoa keskimääräisestä sähkön hinnasta tilastovuoden aikana.

### 3.3.2.3 Lämmityskustannukset

Lämmityskustannukset ovat asumiskustannusten ongelmallisinta osaa. Nämä tiedot on imputoitu ainoastaan erilliskiinteistöissä asuville talouksille (ei siis asunto-osaketalouksille). Tiedot on imputoitu vastaavan vuoden palveluaineiston tietojen avulla käyttäen yksinkertaista lineaarista regressiomallia. Sähkölämmitteisten talojen lämmityskustannukset sisältyvät sähköön, jolloin omamaks -muuttuja sisältää vain muut kustannukset (vesi, jäte, ym.). Muille kuin sähkölämmitteisille taloille muut kustannukset ja lämmityskustannukset on mallinnettu yhteen muuttujaan omalamm. Tiedot on viety ainoastaan omistusmuotoisissa pientaloissa asuville.

### 3.3.2.4 Hoito- ja yhtiövastikkeet

Hoito- ja yhtiövastikkeet muodostavat asumiskustannusten erän sellaisille talouksille, jotka asuvat omistamassaan asunnossa asunto-osakeyhtiössä. Yhtiövastikkeen rahoitusvastikkeen osuuden arvioiminen on käytännössä hyvin vaikeaa sen suuren vaihtelevuuden vuoksi. Näin ollen se puuttuu aineistossa yhtiövastikkeesta. Hoitovastikkeeseen liittyvät tiedot on imputoitu kaikille asunto-osakeasujille asunto-osakeyhtiöiden taloutta käsittelevän tilaston tietojen perusteella mahdollisimman tarkalla tasolla (asuntotyypeittäin, alueittain). Asunto-osakeasujien lämmityskustannusten on katsottu sisältyvän tähän erään.

### 3.3.3 Elatusmaksuihin liittyvät tiedot

Elatusmaksut ovat yksi talouksien välisten tulonsiirtojen muoto, joka voidaan jossain määrin johtaa rekisteriaineistosta. Se vaikuttaa ainoastaan toimeentulotuen simulointimalliin. Elatusmaksut on laskettu talouksille verorekisterin tietojen perusteella, elatusmaksuvelvollisen vähennysoikeuden kautta.

Verorekisterin tiedoista on ensin laskettu vuodessa maksettujen elatusmaksujen määrä. Tämän jälkeen maksajalle on haettu lapset, jotka eivät asu samassa asutokunnassa tämän kanssa. Elatusmaksu on kohdistettu niihin kotitalouksiin, joissa nämä lapset asuvat (tyypillisesti vain yksi kotitalous), ja maksu on jyvitetty lapsille heidän lukumääränsä mukaisesti. Tulonjaon palveluaineiston tietoon verrattuna tämä tuottaa käytännössä kolmenlaisia tuloksia: sellaisia joissa linkkaus on onnistunut euromäärällisesti täydellisesti, sellaisia joissa linkkaus on tuottanut tiedon taloudelle, joilla ei ole haastattelutietoa, ja sellaisia, joissa haastattelutieto on, mutta linkkaus ei tällaista tuottanut. Tämä voi implikoida paitsi haastattelutietojen osittaista puutteellisuutta.

suutta (item nonresponse) myös verovähennysoikeuden alikäyttöä, joskin kyseessä on verrattain pieni verovähennyksen määrä.

#### *4. Aineiston järjestys (sort) ja suhde simulointiin*

Mikrosimuloinnin tutkimusaineiston identiteettitunnukset on luotu niin, että henkilönumeron, asuntokuntanumeron ja perhenumeron sort -järjestys on yhtenevä. Tämä tarkoittaa, että henkilönumeron ollessa oikeassa järjestyksessä, kaksi muutakin tunnusta ovat. Tämä on simuloinnin kannalta olennaista, sillä aineistoa jouduttaisiin muutoin erilaisia järjestelyjä varten järjestämään yhä uudelleen ja uudelleen, joka hidastaisi simulointia olennaisesti, koska aineisto on varsin suuri.

Asuntokuntien sisäinen hierarkia perustuu jäsenten viitehenkilö -suhteeseen. Ensimmäisenä asuntokuntajärjestyksessä on viitehenkilö, jota seuraa hänen puolisonsa, joita seuraavat heidän lapsensa, ja niin edelleen. Tästä seuraten useimmiten myös kohdehenkilön puoliso verotuksessa on hänen viereinen havaintonsa aineiston järjestyksessä.

Aineiston järjestys tulee aina olla mainitun mukainen ennen kuin simulointi aloitetaan. Muutoin simuloinnin eteneminen pysähtyy ennen pitkää.

## **LIITTEET**

**LIITE1.** [Vuokratietojen imputointi](#)

**LIITE2.** [Pohja-aineistojen muuttujakuvaukset](#)