

## *Vuokratietojen imputointi SISU-aineistoon*

### SISÄLLYS

1. Aineiston lähtötilanne ja imputointi .....	2
1.1 Aineiston asutokunnat ja asumistukirekisteri .....	2
1.2 Tietojen imputointimenetelmä .....	2
2. Menetelmän arviointi .....	3
2.1 Luovuttajien ja vastaanottajien tarkastelu .....	3
2.2 Vuokratilastoon vertaaminen .....	4
3. Yhteenveto .....	4
LIITETAULUKOT .....	6

# 1. Aineiston lähtötilanne ja imputointi

## 1.1 Aineiston asuntokunnat ja asumistukirekisteri

Vuokratietojen imputointimenetelmän arvioinnissa on käytetty vuoden 2010 rekisteriaineistoa. Mikrosimuloinnin rekisteriaineiston vuoden 2010 kehitysversiona oli ongelma-asuntokuntien hajottamisen jälkeen<sup>1</sup> yhteensä 400 347 asuntokuntaa. Näistä 117 992 olivat väestötilastojen tietojen mukaan vuokralla asuvia asuntokuntia, 16 806 laitospäästä ja loput 265 549 pääasiassa omistusasunnoissa asuvia asuntokuntia. Asumistukirekisteriin liittyvien korjausten ja kiinteistötietojen tarkastusten jälkeen vuokralla asuvista asuntokunnista 45 008:lle saatiin tieto vuokrasta asumistukirekisteristä. Näin ollen tieto vuokrasta täytyi imputoida 72 984 asuntokunnalle.

## 1.2 Tietojen imputointimenetelmä

Tietojen luovuttamiseen käytettiin asumistukirekisterin vuoden lopun poikkileikkaustilannetta. Kiinteistötietojen tarkastuksiin perustuvien korjausten jälkeen tietoja luovuttavia asuntokuntia oli 295 008 kappaletta.

Imputointimenetelmänä käytettiin matchingiin perustuvaa lähimmän naapurin hot deck -menetelmää. Kullekin vastaanottavalle havainnolle laskettiin yksinkertainen etäisyys kustakin luovuttajahavainnosta, ja luovuttajaksi valittiin tämän etäisyysfunktion mukainen lähin naapuri. Mikäli samalla etäisyydellä olevia luovuttajia oli useita, valittiin koordinaatteihin perustuvan metrisen etäisyyden perusteella lähin, ja tämän jälkeen satunnainen luovuttaja tarvittaessa. Etäisyysfunktiossa käytettiin kolmea parametria: asunnon pinta-alaa sekä yhtenäiskoordinaatiston pituus- ja leveyskoordinaatteja. Muuttujat standardoitiin niin, että niiden keskiarvo oli 0 ja keskihajonta 10 ennen etäisyyksien laskemista. Pinta-alalle annettiin koordinaatteja korkeampi paino etäisyysfunktiossa. Luovutuskertoja ei rajoitettu.

Menetelmä suosii pääasiassa mahdollisimman lähellä vastaanottavaa asuntoa olevia asuntokuntia, joiden pinta-ala on jotakuinkin sama luovuttajan kanssa. Luovuttajia ei rajattu esimerkiksi kuntien tai kuntien osa-alueiden perusteella, sillä tämä olisi voinut aiheuttaa ongelmia pienimpien kuntien- ja osa-alueiden luovuttajajoukkojen ollessa pienehköjä. Toisaalta voidaan myös ajatella, että koordinaatteihin perustuva metrisen etäisyys on kuntarajoja merkittävämpi tekijä.

Menetelmän positiivisena ominaisuutena voidaan nähdä sen kyky huomioida kaupunkisuunnittelun tyypillisiä ominaisuuksia. Esimerkiksi kaupungin vuokra-asuntojen ja tukiasuntojen voidaan tyypillisesti ajatella olevan lähellä toisiaan tai keskitetyksi samalla alueella, jolloin metrisen etäisyys preferoisi tällaisia tapauksia keskenään. Toisaalta vuokria voidaan pitää varsin alueellisia ilmiöinä, ja eri kaupunkien sisällä vuokratasot voivat vaihdella alueittain. Etäisyydeltään lähellä olevien luovuttajien voidaan siis olettaa muistuttavan hyvin vastaanottajiaan. Lähellä toisiaan sijaitsevien asuntojen voidaan myös olettaa olevan rakennettu pitkälti samoihin aikoihin.

<sup>1</sup> Otoksessa oli alun perin 396 767 asuntokuntaa. Ongelmallisten asuntokuntien hajottaminen tuotti otokseen hieman yli 3000 asuntokuntaa lisää, ja lopulta aineistossa oli 400 347 asuntokuntaa.

Menetelmän mahdollisena ongelmana voidaan nähdä asumistukirekisterin mahdollinen vinous (mikäli luovuttajajoukko ei kuvaa tavoitepopulaatiota riittävän hyvin, vääristymä siirtyy tämän mukana) sekä äärihavainnot (etenkin suurempia vuokra-asuntoja on asumistukirekisterissä varsin vähän: toisaalta nämä ovat todellisuudessaakin melko harvinaisia, ja eivätkä välttämättä ole mikrosimuloinnin näkökulmasta todennäköisimpiä asumistuen tai vastaavan etuuden saajia edes reformeja ajatellen).

## 2. Menetelmän arviointi

Menetelmän toimivuutta voidaan arvioida eri näkökulmista. Ensinnäkin tuloksia voidaan verrata asuntojen vuokrat -tilaston vuositilastoon vastaavilla alue- ja asuntoluokituksilla. Toiseksi menetelmää voidaan arvioida sen suorien tulosten perusteella eli miten läheisiä luovuttajia järjestelmä tuotti. Tätä tulosta voidaan arvioida yksinkertaisesti vertailemalla luovuttajien ja vastaanottajien neliökokojen eroa sekä etäisyyttä toisistaan.

Koska luovutuskertoja ei rajattu, on syytä tarkastella myös hieman luovuttajien jakaumaa. Lopulta 72 984 asuntokunnan luovuttajiksi valikoitui vain 39 331 asuntokunnan joukko asumistukirekisterin yli 295 000 asuntokunnasta. 283 asuntokuntaa luovutti tietonsa vähintään kymmenen kertaa ja yksi yksittäinen asuntokunta luovutti tietonsa jopa 60 kertaa. Yli 50 % luovuttajista luovutti tietonsa vain kerran ja yli 99 % korkeintaan kahdeksan kertaa. Huomattavan määrä saavuttaneita luovuttajia on varsin vähäinen joukko.

### 2.1 Luovuttajien ja vastaanottajien tarkastelu

Vastaanottaja-asuntokunnista 21 892 löysi luovuttajan metriseltä etäisyydeltä nolla eli käytännössä samasta asuinrakennuksesta. Näistä yli 99 %:lla vastaanottajan ja luovuttajan pinta-alat poikkesivat toisistaan korkeintaan yhdellä neliömetrillä. Noin 30 %:n osalta luovutusmenetelmää voidaan pitää erittäin luotettavana. Yli 95 % vastaanottajista löysi luovuttajan korkeintaan noin viiden kilometrin päästä, ja 57 591 (eli noin 78,9 %) korkeintaan kilometrin päästä. Suuret pinta-alaerot olivat tässäkin alle kilometrin päästä luovuttajan löytäneiden joukossa harvinaisia: yli 99 % tapauksista sai jälleen korkeintaan neliömetrin pinta-alaerolla olevan luovuttajan.

Pinta-alan painottaminen etäisyysfunktiossa vaikuttaisi johtaneen siihen, että suuret pinta-alaerot olivat yleensäkin hyvin harvinaisia. Yli 99 % vastaanottaja-asuntokunnista sai luovuttajan, jonka pinta-ala poikkesi korkeintaan kolme neliömetriä vastaanottajan pinta-alasta. Niissä tapauksissa, joissa luovuttajan pinta-ala poikkesi jonkin verran vastaanottajan pinta-alasta (ero yli kolme neliömetriä), myös maantieteelliset etäisyydet olivat korkeita. Yli 50 %:lla näistä 728 tapauksesta maantieteellinen etäisyys luovuttajan ja vastaanottajan välillä oli yli noin kymmenen kilometrin. Näitä tapauksia (noin 1 % havainnoista) voidaan siis pitää jossain määrin poikkeavina, ja nämä havainnot tyypillisesti olivat joko erittäin kookkaita (yli 100 neliömetriä) tai erittäin pieniä (paikoittain asuntokuntien hajoituksista seuranneita) asuntokuntia.

Kunta- ja osa-alueetasolla tarkasteltuna 63,43 % vastaanottajista sai luovuttajan saman kunnan samalta osa-alueelta ja 93,95 % samasta kunnasta.

## 2.2 Vuokratilastoon vertaaminen

Vuoden 2010 vuokratilaston vuosijulkaisun mukaan keskimääräinen neliövuokra yhden huoneen asunnoissa oli 11,73 euroa, kahden huoneen asunnoissa 9,82 euroa ja kolmen tai useamman huoneen asunnoissa 9,05 euroa. Vastaavat luvut vastaanottajajoukolla olivat 12,60 euroa, 9,52 euroa ja 8,84 euroa. Keskiarvo kokonaisuudessaan oli tilastossa 9,86 euroa per neliömetri ja imputoiduissa tiedoissa 10,28 euroa per neliömetri (eli noin 4 % korkeampi).

Kuntakohtaisia eroja on mahdollista tarkastella laajemminkin: liitetaulukossa 1 ja 2 on esitetty asuntojen vuokrat -tilaston koko vuoden julkaisun mukaiset kuntatason neliövuokratiedot luokiteltuna huonelukumäärän mukaisesti yhdessä imputoitujen tietojen vastaavan jakauman kanssa. Taulukkoa voidaan nopeasti tulkita niin, että etenkin suurissa kaupungeissa yksioiden neliövuokrat vaikuttaisivat yliestimoituvan menetelmän johdosta, vaikka kokonaisuudessaan keskineliövuokrat olisivatkin lähellä toisiaan. Noin euron ero keskimääräisessä yhden huoneen asunnon neliövuokrassa kasvaa esimerkiksi Vantaan tapauksessa miltei puoleentoista euroon.

Koska imputoidut tiedot edustavat vain noin 61,8 % mallidatasta (mikrosimulointiaineiston asutokunnista), on tarkastelu syytä tehdä vielä koko mallidatan tasolla yhdistäen imputoidut ja asumistukirekisteristä suoraan saatavat tiedot. Koko mallidatan (117 992 vuokralla asuvaa asutokuntaa) keskineliövuokrat olivat huonelukumäärittäin yhden huoneen asunnoille 12,64 euroa, kahden huoneen asunnoille 9,52 euroa ja kolmen tai useamman huoneen asunnoille 8,84 euroa per neliömetri. Yhdistetyn aitojen asumistukirekisterin tietojen sekä matchingillä imputoitujen tietojen kuntakohtaiset erot (liitetaulukot 2 ja 3) ovat hyvin pieniä. Näin voidaan todeta, ettei imputointimenetelmä juurikaan muuta ainakaan keskineliövuokrien jakaumaa.

## 3. Yhteenveto

Esitellyt menettely tuottaa mikrosimulointiaineistoon vuokratietojen jakauman, joka seuraa pitkälti (väestötilastojen sekä kiinteistörekisterin tietojen avulla korjatun) asumistukirekisterin tietojen jakaumaa, ainakin kunnallisella tasolla huonelukumäärittäin. Huonelukumäärien kontrolloimattomuus luovutusvaiheessa voi potentiaalisesti olla osasyynä yhden- ja kahden huoneiden asuntojen neliöhintojen eroavaisuuksiin tilaston tietojen ja asumistukirekisteriperusteisen imputoinnin välillä. Tämä vaikuttaisi kuitenkin epätodennäköiseltä, koska tarkasteltaessa eroja suurempien kaupunkien tapauksissa, luovuttajien määrä on suuri, ja lisäksi luovuttajien ja vastaanottajien asuinneliömetrien erot ovat pääosin pieniä.

Menetelmän ongelmana voidaan pitää poikkeavia havaintoja, joita on noin 1 % vastaanottavista havainnoista. Nämä poikkeavat havainnot jakautuvat kahteen ryhmään: asutokuntiin, jotka on hajotettu malliaineistoa varten (noin 5 000 kappaletta) sekä pinta-alaltaan merkittävän suuriin asuntoihin, joissa luovuttajien valikoima on paikoin huomattavan kapea.

Ehkä hieman yllättävämpi ongelma liittyy neliövuokrien tasoon: imputointimenetelmä tuottaa etenkin yhden huoneen asunnoille suurissa kunnissa

(poislukien Helsinki) korkeamman neliövuokran keskiarvon. Liitetaulukossa 4 on Asuntojen vuokrat 2010 -vuositilaston keskineliövuokra huonelukumäärittäin ja kunnittain jaettuna mikrosimulaation datan vastaavalla tiedolla.

Lyhyenä yhteenvedona tilaston ja imputointimenetelmän erona voidaan sanoa, että imputointimenetelmä tuottaa yksöiden osalta hieman korkeamman keskineliövuokran ja suurempien asuntojen osalta hieman matalamman keskineliövuokran kuin tilaston julkaisemat luvut.

## Liitetaulukot

**Taulukko 1: Keskineliövuokrat huonelukumäärittäin, asuntojen vuokrat -tilasto**

	1h	2h	3h+
Espoo	12,89	11,85	10,91
Helsinki	15,91	12,19	10,92
Hämeenlinna	11	9,59	8,66
Joensuu	11,73	9,16	8,51
Kajaani	11,3	9,32	9,05
Kokkola	10,73	8,73	7,77
Kotka	10,71	8,75	8,41
Kouvola	9,48	8,33	7,61
Kuopio	11,2	9,58	8,65
Lahti	11,26	9,61	9,08
Lappeenranta	11,21	9,27	8,67
Mikkeli	10,47	9,48	8,54
Oulu	11,47	9,7	8,86
Pori	10,79	8,57	7,74
Porvoo	11,83	9,81	9,19
Rauma	9,76	8,3	7,54
Rovaniemi	11,54	9,52	8,62
Seinäjoki	10,43	8,49	7,95
Tampere	12,32	10,53	9,34
Turku	12,05	9,75	8,95
Vaasa	11,06	9,23	8,44
Vantaa	12,47	11,25	10,31

**Taulukko 2: Keskineliövuokrat huonelukumäärittäin, imputoidut havainnot**

	1h	2h	3h+
Espoo	14,12	10,88	10,26
Helsinki	15,38	11,07	10
Hämeenlinna	11,89	9,54	8,81
Joensuu	12,43	8,92	8,11
Kajaani	11,36	8,88	8,3
Kokkola	12,57	8,61	7,83
Kotka	10,87	9,23	8,29
Kouvola	10,42	8,17	7,69
Kuopio	12	9,03	8,55
Lahti	12,09	9,73	8,89
Lappeenranta	12	9,3	8,57
Mikkeli	11,6	9,47	8,24
Oulu	12,04	9,03	8,3
Pori	11,57	8,87	8,14
Porvoo	11,6	9,47	8,9
Rauma	10,78	8,42	7,07
Rovaniemi	12,81	9,26	8,9
Seinäjoki	12,31	8,92	7,76
Tampere	12,99	9,3	9
Turku	12,45	9,35	8,43
Vaasa	12,67	9,73	8,23
Vantaa	13,92	10,76	9,92

**Taulukko 3: Neliövuokrat, imputoidut- ja asumistukirekisterin asunnot mikrosimulaation datassa**

	1h	2h	3h+
Espoo	14,13	10,89	10,27
Helsinki	15,38	11,08	10
Hämeenlinna	11,89	9,54	8,81
Joensuu	12,43	8,92	8,11
Kajaani	11,36	8,88	8,31
Kokkola	12,57	8,61	7,84
Kotka	10,88	9,23	8,29
Kouvola	10,42	8,17	7,69
Kuopio	11,96	9,03	8,55
Lahti	12,09	9,73	8,89
Lappeenranta	12	9,3	8,57
Mikkeli	11,6	9,47	8,25
Oulu	12,04	9,04	8,3
Pori	11,57	8,87	8,15
Porvoo	11,6	9,47	8,91
Rauma	10,78	8,42	7,07
Rovaniemi	12,81	9,25	8,9
Seinäjoki	12,31	8,92	7,76
Tampere	12,99	9,3	9
Turku	12,45	9,35	8,43
Vaasa	12,67	9,73	8,23
Vantaa	13,92	10,76	9,92



**Taulukko 4: Vuokratilaston keskineliövuokra jaettuna mallidatan keskineliövuokralla, huonelukumäärittäin ja kunnittain**

	1h	2h	3h+
Espoo	0,91	1,09	1,06
Helsinki	1,03	1,10	1,09
Hämeenlinna	0,93	1,01	0,98
Joensuu	0,94	1,03	1,05
Kajaani	0,99	1,05	1,09
Kokkola	0,85	1,01	0,99
Kotka	0,98	0,95	1,01
Kouvola	0,91	1,02	0,99
Kuopio	0,94	1,06	1,01
Lahti	0,93	0,99	1,02
Lappeenranta	0,93	1,00	1,01
Mikkeli	0,90	1,00	1,04
Oulu	0,95	1,07	1,07
Pori	0,93	0,97	0,95
Porvoo	1,02	1,04	1,03
Rauma	0,91	0,99	1,07
Rovaniemi	0,90	1,03	0,97
Seinäjoki	0,85	0,95	1,02
Tampere	0,95	1,13	1,04
Turku	0,97	1,04	1,06
Vaasa	0,87	0,95	1,03
Vantaa	0,90	1,05	1,04