



HENRI LAUKKANEN

VALIKOITUMINEN JA ENDOGEENISUUS NAAPURUSTOVAIKUTUSTUTKIMUKSEN METODOLOGISINA HAASTEINA

4

TYÖPAPEREITA 2015



Helsingin kaupunki
Tietokeskus

TYÖPAPEREITA
ARBETSPAPPER
WORKING PAPERS

2015:4

SISÄLLYS

1 Naapurustovaikutukset	3
2 Metodologiset haasteet	5
3 Empiiriset tutkimusasetelmat	7
3.1 Koejärjestelyn kaltaiset tutkimusasetelmat, joiden aineiston taustalta löytyy interventio.....	8
3.2 Koejärjestelyn kaltaiset tutkimusasetelmat, jotka käyttävät luonnollisesti syntyneitä aineistoja.....	8
3.3 Havaintoaineistoa käyttävät tutkimusasetelmat	11
Lopuksi	17
Liitetaulukko 1. Empiiriset tutkimusasetelmat	19
Lähteet	21

Kvantitatiivisessa naapurustovaikutustutkimuksessa mitataan naapuruston ja jonkin yksilö- tai yhteisötason ilmiön, esimerkiksi yksilön työmarkkinamenestyksen, välistä yhteyttä. Tutkimuksen lopullinen päämäärä on esittää väitteitä naapuruston mahdollisesta kausaalivaikutuksesta. Valitettavasti tutkimusasetelmat ovat harvoin sellaisia, että mutkattomat kausaalitulkinnat ovat mahdollisia (vrt. Angrist & Pischke 2008). Keskeisimmät ongelmat lienevät puuttuvien (omitted) tai endogeenisuudesta kärsivien muuttujien estimaateille aiheuttamat harhat (bias) (ks. esim. Galster 2008).

Tässä paperissa luodaan katsaus empiirisestä tutkimuskirjallisuudesta löytyviin tapoihin kontrolloida kyseisiä harhoja. Katsauksen tarkoitus on palvella tulevien empiiristen tutkimusasetelmien suunnittelua. Vaikka toistaiseksi ei ole onnistuttu luomaan mitään täysin tyydyttävää ratkaisua harhoista eroon pääsemiseksi, tarjolla on joka tapauksessa useita vaihtoehtoja, joista on ainakin jonkin verran hyötyä. Sitä paitsi myös menetelmien heikkouksien tunteminen on hyödyksi empiiristä tutkimusta tehtäessä.

Paperi etenee siten, että ensin määritellään, mistä kvantitatiivisessa naapurustovaikutustutkimuksessa ja sitä tavallisesti rasittavissa puuttuvien tai endogeenisuudesta kärsivien muuttujien aiheuttamisissa harhoissa on kyse. Taustoittavan osuuden jälkeen esitetään paperin varsinainen ydin, katsaus empiirisessä naapurustovaikutustutkimuksessa kehiteltyihin mainittuja harhoja tavalla tai toisella kontrolloiviin tutkimusasetelmiin. Lukujen yhteenvedoissa pyritään hahmottamaan vaatimatonta ohjenuoraa tulevien tutkimusten metodologisille valinnoille.

Esityksessä tilastolliset mallit esitetään lähes poikkeuksetta luonnollisella kielellä. Paperin tarkoitus ei ole opettaa käyttämään kyseisiä menetelmiä, vaan ainoastaan tekemään lukija tietoiseksi niihin liittyvistä mahdollisuuksista. Siksi riittänee, että tyydytään valaisemaan niiden taustalta löytyviä intuitioita menemättä teknisiin yksityiskohtiin. Formaalisimmasta ja seikkaperäisemmästä esityksestä kiinnostunutta kehotetaan tutustumaan lähdekirjallisuuteen.

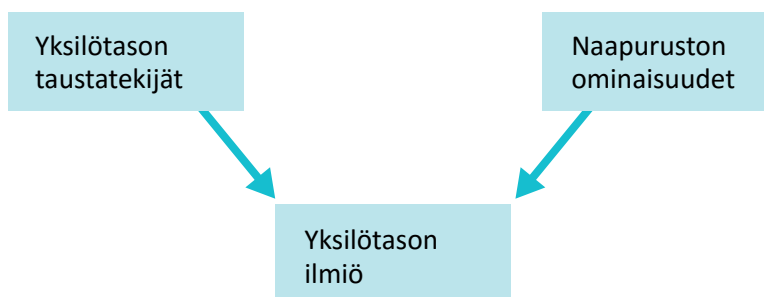
Katsaus on suuressa kiitollisuudenvelassa Galsterin 2008, Galsterin ja Hedmanin (2013) sekä Galsterin ja kumppaneiden (2015) papereissa tehdyille katsauksille ja jäsenyksille.

1 NAAPURUSTOVAIKUTUKSET

Naapurustovaikutukset (neighbourhood effect) on mahdollista luokitella kahteen luokkaan sen perusteella, mikä on tutkimuksen kohteena olevan yksikkö. Yhtäältä voidaan tutkia naapuruston ominaisuuksien vaikutusta siellä asuvaan yksilöön. Toisaalta voidaan tarkastella, kuinka naapuruston ominaisuudet vaikuttavat sen sosiaaliseen organisaatioon, missä tapauksessa tutkimuksen yksikkö on yhteisö. (Sampson 2008.) Tässä paperissa keskitytään edelliseen tapaukseen. Havainnollisuuden nimissä seuraavassa tutustutaan kuitenkin kumpaankin naapurustovaikutustyyppiin.

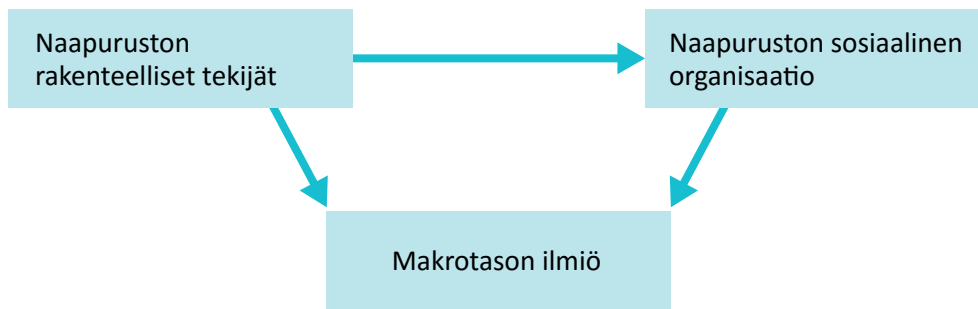
Yksilöä tutkittaessa naapurustovaikutuksen käsite viittaa tilanteeseen, jossa yksilön asuinympäristö, erityisesti sen sosiaalinen ulottuvuus, vaikuttaa yksilöön ylitse ja ohi tämän yksilöllisten taustatekijöiden (Bernelius 2013, 33) (Kuvio 1). Yksilötason naapurustovaikutusten tutkiminen on erottomasti liittynyt kaupungeissa esiintyvään sosiaaliseen eriytymiseen, ja erityisesti mielenkiinto on kohdistunut huono-osaisuuden keskittymisen mahdollisiin kielteisiin vaikutuksiin (vrt. Bernelius 2013, 33). Keskeinen kysymys on ollut, onko deprivoituneessa naapurustossa asumiselle haitallisia vaikutuksia yksilön elämään ylitse ja ohi tämän yksilöllisten ominaisuuksien (Van Ham & Manley 2010). Suosittu tutkimuskohde on ollut esimerkiksi naapuruston vaikutus yksilön työmarkkinamenestykseen.

Kuvio 1. Naapuruston vaikutus yksilötason ilmiöön



Kun tutkitaan naapuruston vaikutusta yhteisötason ilmiöihin, etsitään yhteyksiä naapuruston rakenteellisten ominaisuuksien, kuten esimerkiksi väestön sosioekonomisen koostumuksen, ja naapuruston sosiaalisen organisaation kulttuurillisten piirteiden välillä (Kuvio 2). (Sampson 2008.) Viime vuosien merkittävin tutkimuskohde tällä saralla lienee ollut naapuruston ns. yhteispystyvyys (collective efficacy). Käsite viittaa naapuruston sosiaaliseen koheesioon ja kontrolliin. Empiirisissä tutkimuksissa naapurustossa vallitsevan yhteispystyvyyden tason on havaittu olevan yhteydessä esimerkiksi rikollisuuden esiintymiseen. (Ks. esim. Sampson 2012.)

Kuvio 2. Naapuruston vaikutus yhteisötason ilmiöihin



Empiirisen naapurustovaikutustutkimuksen haasteet ovat osin hieman erilaisia riippuen siitä, tutkitaanko yksilöä vai yhteisöä, vaikka totta kai myös yhtymäkohtia löytyy paljon. Valitettavasti tämän paperin puitteissa ei ole mahdollista tutustua koko kiehtovaan ongelmakenttään, vaan seuraavassa katsauksessa rajaudutaan yksilöä käsittelevään tutkimukseen.

2 METODOLOGISET HAASTEET

Tavallisimmin kvantitatiivinen naapurustovaikutustutkimus perustuu jonkinlaiseen regressioanalyysiin, jossa kiinnostuksen kohteena olevaa yksilötason ilmiön (outcome) vaihtelua selitetään naapuruston ominaisuuksilla. Naapuruston vaikutuksia koskevien päätelmien tekemiseksi on välttämätöntä vakioida sellaiset yksilötason taustatekijät, jotka ovat mahdollisesti yhteydessä sekä selitettävään muuttujaan että naapurustoon valikoitumiseen. Yksinkertaisimmillaan ajatus voidaan ilmaista seuraavalla kaavalla.

$$Y_i = \beta_0 + \beta X_i + \alpha Z_{ij} + e_i$$

Missä Y_i on selitettävä muuttuja yksilölle i , β_0 on vakio-termi, X yksilötason ominaisuuksia, Z naapuruston ominaisuuksia ja e_i virhetermi yksilölle i .

Naapurustovaikutustutkimuksissa ei ole yleensä pystytty havaitsemaan suoraan prosesseja, jotka tuottavat naapurustovaikutuksia, vaan niiden kohdalla on jouduttu tyytymään korvikemuuttujiin (proxy). Usein ne ovat olleet melko karkeitakin. Esimerkiksi jos ollaan oltu kiinnostuneita sosiaaliseen vuorovaikutukseen liittyvistä naapurustovaikutuksista tuottavista mekanismeista, on mitattu naapuruston depriavaatioastetta.

Tällaisessa tilanteessa, jossa kausaalipolkuja ei havainnoida suoraan, on selvää, että naapuruston vaikutusta koskevien väitteiden esittämiseen liittyy monia epävarmuustekijöitä. Seuraavaksi esitetään kaksi ehkä keskeisintä tilastollisen päättelyn kohtaamaa haastetta. Niistä ensimmäinen johtuu siitä, että mallissa ei ole vakioitu kaikkia relevantteja yksilötason taustatekijöistä, toinen siitä, että naapurusto on vaikuttanut mallin selittäviin muuttujiin.

Puuttuvat muuttujat, eli valikoitumisen ongelma

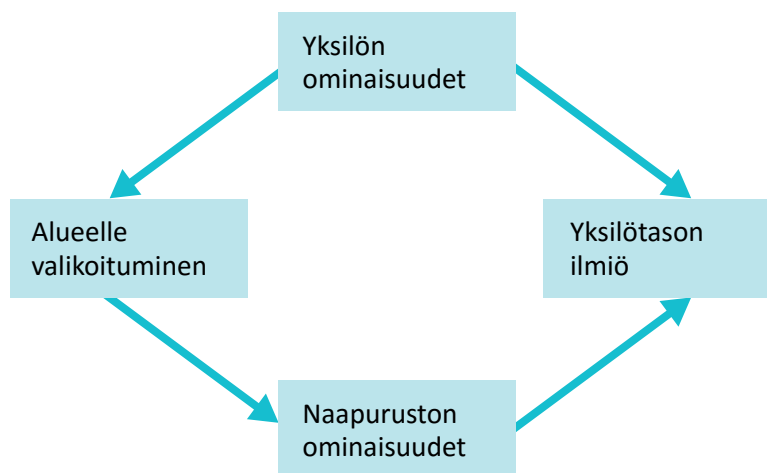
Edellä kuvatun kaltaisessa regressioanalyysissä naapuruston vaikutusta tutkimuksen kohteena olevaan yksilötason ilmiöön arvioidaan vertaamalla yksilötason ilmiön arvojen vaihtelua sellaisten yksilöiden välillä, jotka ovat asuinpaikan ominaisuuksia lukuun ottamatta mahdollisimman paljon toistensa kaltaisia. Vertailtavuus pyritään saavuttamaan yksilötason taustatekijöitä kontrolloimalla. Se, kuinka uskottavasti naapuruston ominaisuuksien ja tutkittavan yksilötason ilmiön tilastollisten yhteyksien voi tulkita kertovan kausaalisuhteesta, riippuu oleellisesti siitä, ovatko vertailtavat ryhmät relevanttien taustatekijöiden suhteen todella riittävän samanlaisia. (Vrt. Angrist & Pischke 2015.)

Puuttuvien muuttujien (omitted variables) ongelmassa on kyse siitä, että tilastollisessa mallissa on jäänyt huomioimatta joitakin sellaisia yksilötason tekijöitä, jotka ovat yhteydessä sekä naapurustoon valikoitumiseen että tutkimuksen kohteena

olevaan yksilötason ilmiöön (Kuvio 3). Tilanne johtaa harhaisiin estimaatteihin, jos tarkasteltavien alueominaisuuksien suhteen erilaisiin naapurustoihin on valikoitunut tutkittavan yksilötason ilmiön suhteen erilaisia yksilöitä. (Galster 2008.) Koska harha perustuu valikoivaan muuttoliikkeeseen, metodologista haastetta kutsutaan usein valikoitumisen ongelmaksi.¹

Esimerkki valikoitumisen ongelmasta on tilanne, jossa yksilön mittaamatta jäänyt taipumus tavoitella henkilökohtaista prestiisiä on saanut hänet sekä muuttamaan korkeaa prestiisiä nauttivalle asuinalueelle että tavoittelemaan uutterasti suuria tuloja. Nyt tutkijan on vaikea selvittää asuinalueen prestiisin itsenäinen vaikutus yksilön tulokehitykseen. (Hedman & Galster 2013.)

Kuvio 3. Valikoitumisen ongelma



Endogeenisuuden ongelma

Naapurustovaikutustutkimuksessa muuttujien endogeenisuudella on kaksi ulottuvuutta, joista ensimmäisessä on kyse siitä, että jotkut mallissa vakioitavat yksilötason tekijät ja naapuruston ominaisuudet ovat keskenään suhteessa, jossa molemmat määräytyvät yhdellä kertaa. Esimerkiksi Galsterin (2003) mukaan päätös naapurustoon muuttamisesta tehdään usein yhdessä eräiden muiden päätösten, kuten asunnon hallintaperustaa ja asettumisen kestoja, koskevien päätösten, kanssa. Tällaisessa tilanteessa estimaatit naapuruston vaikutuksesta kärsivät endogeenisuuden harhasta, koska kyseisten yksilötason taustatekijöiden vakioimisesta seuraa ylikontrolloimista (over-controlling) (Kuvio 4). (Galster 2003.)

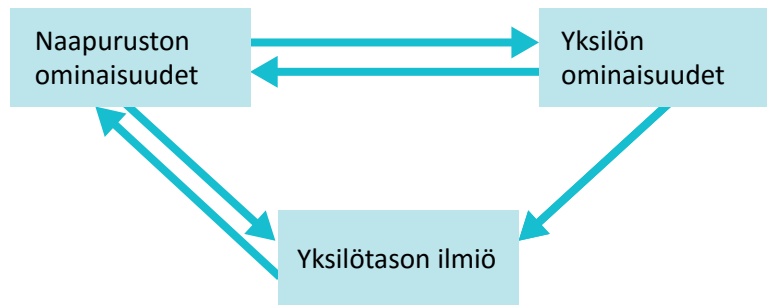
Endogeenisuuden toisessa ulottuvuudessa on kyse siitä, että naapurustossa asuminen on vaikuttanut tutkimuksen kannalta relevantteihin havaittuihin yksilötason

¹ Sampson (esim. 2011) on kritisoinut naapurustovaikutustutkimusta siitä, että sen piirissä valikoiva muuttoliike on yleensä ymmärretty pelkkänä ekonometrisenä kiusankappaleena. Hänen mielestään ilmiöllä on itsessään teoreettista mielenkiintoa pyrittäessä ymmärtämään naapurustovaikutuksista. Naapurustoon valikoituminen kannattaisikin käsittää sosiaalisesti prosessiksi, joka valikoi yksilön tilanteeseen, jossa hän altistuu naapurustolle. Esimerkiksi Chicagossa etnisiin ryhmiin liittyvillä kulttuurillisilla merkityksillä on suuri merkitys naapurustojen väestörakenteiden kehkeytyemisessä. (Sampson 2011.) Vaikka Sampsonin kanssa on vaikea olla eri mieltä, tässä katsauksessa valikoitumista lähestytään nimenomaan ekonometrisenä ongelmana.

ominaisuuksiin ja aiheuttanut näin kyseisiin ominaisuuksiin liittyvän ylikontrolloimisen riskin (Kuvio 4) (Sampson 2008). Esimerkiksi naapuruston vaikutusta yksilön työmarkkinamenestykseen tarkastelevassa tutkimuksessa tällainen tilanne syntyy, jos naapurusto on vaikuttanut yksilön koulutustasoon. Kiinnostavana erityistapauksena voidaan mainita tilanne, jossa yksilön nykyinen tai aikaisempi naapurusto on vaikuttanut naapurustoon valikoitumiseen muuttoliikkeeseen vaikuttavien resursien, kuten ansiotulojen, kautta (Hedman & Galster 2013).

Erityisen pirullista endogeenisuuden aiheuttamassa harhasta on, että se vaikuttaisi heikentävän mallien kykyä paljastaa epälineaarisia suhteita naapuruston ominaisuuksien ja tutkimuksen kohteena olevan yksilötason ilmiön välillä (Hedman & Galster 2013). Epälineaarisuuksien tunteminen on arvokasta tietoa, koska kynnyksien paikantaminen auttaa sekä naapurustovaikutuksia tuottavia mekanismeja koskevien päätelmien tekemisessä että politiikkasuositusten laatimisessa (Galster ym. 2014).

Kuvio 4. Endogeenisuuden ongelma



3 EMPIIRISET TUTKIMUSASETELMAT

Empiirisessä naapurustovaikutustutkimuskirjallisuudessa on esitetty useita tutkimusasetelmia, jotka tavalla tai toisella kontrolloivat valikoitumiseen tai endogeenisuuteen liittyviä harhoja. Karkeasti ottaen ne voidaan jakaa kolmeen ryhmään niiden aineiston ja lähestymistavan mukaan.

Ensimmäisessä on kyse koejärjestelyn kaltaisista tutkimusasetelmista, joissa aineisto on syntynyt asuntopoliittisen intervention seurauksen. Asetelmien vahvuus on tutkittavien yksilöiden satunnaistamisessa. Toisen ryhmän tutkimusasetelmat pyrkivät nekin koejärjestelyä muistuttavaan tilanteeseen, mutta toisin kuin edelliset ne perustuvat ilman interventiota syntyneeseen havaintoaineistoon. Näissä asetelmissä satunnaisuus on pyritty varmistamaan kekseliäillä perusjoukon valinnoilla. Kolmannen ryhmän tutkimusasetelmissä analysoidaan ilman interventiota syntyneitä havaintoaineistoja ja valikoitumisen ja endogeenisuuden ongelmia kontrolloidaan erinäisin ekonometristen menetelmien avulla.

Seuraavan katsauksen kokoava taulukko löytyy paperin lopusta.

3.1 Koejärjestelyn kaltaiset tutkimusasetelmat, joiden aineiston taustalta löytyy interventio

Eräässä mielessä kausaalipäättelyn kannalta vahvimpia tämän katsauksen tutkimusasetelmista ovat ne, jotka muistuttavat mahdollisimman pitkälle koejärjestelyjä, siis ne joiden aineisto on syntynyt asuntopoliittisen intervention seurauksena (vrt. Angrist & Pischke 2015). Siksi ne esitetään ensimmäiseksi. Jossain määrin niitä on mahdollista pitää esikuvana, jota vasten muita tutkimusasetelmia verrataan.

Kyseisissä tutkimusasetelmissä havaitsemattomat yksilötason tekijät eivät vaivaa yksilöiden vertailua ja edelleen naapuruston vaikutusten estimointia, koska aineiston yksilöt on satunnaistettu. Naapurustolle altistumisen vaikutusta tutkitaan tarkastelemalla uuteen naapurustoon muuttamisen (treatment) vaikutuksia, minä ansiosta myös endogeenisuuden ongelma on vältettävissä. (Galster ym. 2015, Sampson 2008.)

Kuuluisimmat tällaiset tutkimukset perustuvat Yhdysvalloissa toteutettuihin Gautreaux, HOPE IV, ja Moving to opportunity -nimisiin asuntopoliittisiin ohjelmiin, joissa deprivoituneiden naapurustojen lapsiperheitä on avustettu muuttamaan parempiin naapurustoihin. Moving to opportunity -ohjelmaan pohjautuvia tutkimuksia pidetään satunnaistamisen suhteen onnistuneimpana (Galster ym. 2015). Niihin tutustumisen voi aloittaa vaikka papereista Kling ym. (2007) ja Ludwig ym. (2013).

Ei ole tiedossa, että vastaavia ohjelmia olisi toteutettu Suomessa, joten toistaiseksi tällaiset kvasikokeelliset tutkimusasetelmat eivät tule kysymykseen täkäläisessä naapurustovaikutustutkimuksessa.

Lopuksi lienee paikallaan huomauttaa, että edes kvasikokeelliset tutkimukset eivät ole jääneet osattomiksi metodologiakritiikistä. Keskeisimmät tutkimusasetelmien epäkohdat lienevät se, että niiden avulla ei ole ollut mahdollista tutkia erilaisia naapurustovaikutuksia kattavasti, ja se, että niiden pohjalta saadut tulokset eivät ole yleistettävissä kovin hyvin (ks. esim Sampson 2008).

3.2 Koejärjestelyn kaltaiset tutkimusasetelmat, jotka käyttävät luonnollisesti syntyneitä aineistoja

Tutkijat ovat etsineet satunnaisuutta myös aineistoista, jotka ovat syntyneet ilman varsinaista interventiota. Tällaisissa tutkimusasetelmissä naapuruston vaikutuksen estimoimista häiritsevät havaitsemattomat ominaisuudet on yritetty kiertää kekseliällä perusjoukon rajauksilla. Seuraavaksi esitellään näistä muutamia esimerkkejä. Lukijan kannattaa kiinnittää huomiota siihen, että ne soveltuvat hyvin erityyppisiin tutkimuskysymyksiin. Esimerkiksi sisarusvertailuun perustuva tutkimusasetelma lienee mielekäs lähinnä lapsia tai nuoria koskevien ilmiöiden tutkimisessa.

Saman naapuruston asukkaat

Bayer ja kumppanit (2008) uskovat, että valikoitumisen ongelma on jossakin määrin mahdollista kiertää rajaamalla tarkastelu saman naapuruston asukkaisiin. Tutkijat olettavat, että vaikka yksilöt valikoituisivat naapurustoihin havaitsemattomien yksilötason taustatekijöiden perusteella, merkittävää valikoitumista ei tapahdu enää korttelin sisällä. Jos se pitää paikkansa, korttelin sisällä tapahtuvan sosiaalisen vuorovaikutuksen seurauksia tutkittavalle yksilötason ilmiölle on mahdollista tutkia vertailemalla naapuruston erilaisissa kortteleissa asuvia yksilöitä toisiinsa.

Menetelmän varjopuolina voidaan pitää edellä mainitun oletuksen uskottavuutta sekä sitä, että sen puitteissa vain tietynlaisten, nimittäin sosiaaliseen vuorovaikutukseen perustuvien, naapurustovaikutusmekanismien tutkiminen on mahdollista (Galster ym. 2015). Sitä paitsi tutkimusasetelma ei huomioi kattavasti endogeenisuuden ongelmaa.

Sosiaalisten vuokra-asuntojen asukkaat

Eräissä tutkimuksissa regressiomallit on ajettu erikseen sosiaalisen vuokra-asuntokannan asujaimistolle, koska heidän kohdalla naapurustoihin sijoittuminen on oletettavasti satunnaisempaa (esim. Sari 2012, Oreopoulos 2003 Van Ham & Manley 2010, Manley & Van Ham 2012). Ongelma asetelmissa on se, että niiden tekemä oletus satunnaisuudesta ei välttämättä pidä paikkaansa. Ainakaan pääkaupunkiseudulla sosiaalisten vuokra-asuntojen asukkaat eivät ole sijoittuneet lähelläkään satunnaisesti (ks. Lankinen 2006). Sitä paitsi asetelma eivät kontrolloi endogeenisuutta kuin joiltakin osin.

Lasten koulumenestystä käsittelevässä naapurustovaikutustutkimuksessaan Weinhardt (2014) on esittänyt jatkokehityksen samasta ideasta. Tutkimusasetelmassaan hän oletti, että pitkistä jonotusajoista johtuen sosiaaliseen vuokra-asuntoon muuttamisen ajankohta on Englannissa kvasi-satunnainen. Tällöin sosiaaliseen vuokra-asuntoon muuttaminen oli mahdollista käsittää asunnon hakijoiden suhteen satunnaisesti ilmeneväksi naapurustokontekstille altistumiseksi. Tutkimuksessa selvitettiin naapuruston vaikutusta lasten menestykseen kansallisessa osaamistestissä vertailemalla toisiinsa lapsia, jotka erosivat toisistaan siinä, että osa oli muuttanut sosiaaliseen vuokra-asuntoon ennen testiä ja osa vasta testin jälkeen. (Weinhardt 2014.)

Mikäli oletus satunnaisuudesta pitää paikkansa, asetelma on lupaava. Periaatteessa lienee selvitettävissä, miten asian laita on suomalaisilla asuntomarkkinoilla. Paljon asetelman käyttökelpoisuudesta riippunee myös siitä, kuinka uskottavasti muuttamisen itsenäiset vaikutukset onnistutaan kontrolloimaan (vrt. Sampson 2008). Endogeenisuuden ei pitäisi olla ongelma, koska aineiston yksilöt on mitattu ennen naapurustolle altistumista.

Sisarukset

Eräissä tutkimuksissa *ceteris paribus* -asetelma naapuruston ominaisuuksille altistumiselle on rakennettu vertailemalla keskenään saman perheen sisaruksia. Aineiston perheet ovat muuttaneet naapurustosta toiseen siten, että perheen eri-ikäiset sisarukset ovat eläneet tietyssä iässä erilaisissa naapurustoissa: vanhempi sisarus naapurustossa A ja nuorempi sisarus naapurustossa B. Asetelman etuna on, oletettavasti, että sisarusten vertaileminen ei kärsi havaitsemattomista perheeseen liittyvistä ominaisuuksista. Samalla oletetaan, että monilapsisten perheiden valikoituminen naapurustoon ei riipu lasten havaitsemattomista ominaisuuksista ja että sisarukset eivät poikkea toisistaan relevanteilta havaitsemattomilta ominaisuuksiltaan. Esimerkki tällaisesta tutkimuksesta on Aaronson (1998).

Edellä mainittujen oletusten kiistanalaisuuden lisäksi sisarus-asetelmat ovat kärsineet vähäisestä vaihtelusta sekä selitettävien muuttujien että naapurustotason selittävien muuttujien kohdalla. Ongelma on myös pieniksi jääneet otoskoot. (Galster ym. 2010.) Endogeenisuutta asetelma ei kontrolloi tyydyttävästi, esimerkiksi naapurustokonteksti ja vanhempien työmarkkinastatus voivat määräytyä samanaikaisesti.

Yhteenveto koejärjestelyn kaltaisista tutkimusasetelmista, jotka käyttävät luonnollisesti syntyneitä aineistoa

Edellä on esitelty joukko tutkimusasetelmia, joissa valikoivan muuttoliikkeen mahdollisesti aiheuttama harha on pyritty kiertämään kekseliäällä perusjoukon rajauksella. Saman naapuruston asukkaiden, sosiaalisten vuokra-asuntojen asukkaiden ja perheen sisarusten tapauksessa naapuruston valintaan liittyvien havaitsemattomien taustatekijöiden vaikutuksen ei oleteta häiritsevän pahasti estimaatteja naapurustovaikutuksista.

Mutta mikään taikatemppu tällaiset niin sanonut luonnollisesti syntyneet koeasetelmat eivät kuitenkaan ole. Heikkoutena voidaan pitää ennen kaikkea sitä, että niissä joudutaan tekemään sellaisia satunnaisuutta koskevia oletuksia, joiden paikkansapitävyys on usein helposti kyseenalaistettavissa. Myös endogeenisuuden kontrolloiminen jättää toivomisen varaa. Sitä paitsi asetelmien edellyttämät perusjoukon valinnat asettavat yleensä rajoittavia ehtoja sille, mitä tutkimusasetelmassa voidaan kysyä.

Parhaimmillaan, siis silloin kun satunnaisuutta koskevat oletukset ovat riittävän uskottavia, luonnollisesti syntyneet koeasetelmat lienevät joka tapauksessa käyttökelpoisia. Edelleen kannattaa pohtia niiden hyödyntämistä yhdessä seuraavaksi esiteltävien ekonometristen mallien kanssa.

3.3 Havaintoaineistoa käyttävät tutkimusasetelmat

Leijonan osa naapurustovaikutustutkimuksista on tehty aineistoilla, jotka eivät salli koejärjestelyn kaltaisen tutkimusasetelman rakentamista. Niissä helpotusta valikoitumisen ja endogeenisuuden ongelmiin on etsitty erinäisistä ekonometrisista menetelmistä. Differences-in-differences, fixed effects, instrumental variables, ja random effects -mallit kontrolloivat mahdollisia harhoja kukin omalla tavallaan. Vastaavasti jokaiseen liittyy omat vahvuutensa ja heikkoutensa.

Mainittujen ekonometristen mallien lisäksi seuraavassa katsauksessa tutustutaan sensitivity analysis -nimiseen menetelmään. Vaikka se ei kontrolloi valikoitumista tai endogeenisuutta, siitä voi kaikesta huolimatta olla hyötyä naapurustovaikutustutkimuksessa, koska sen avulla on mahdollista arvioida mahdollisten havaitsemattomien yksilötason taustatekijöiden vaikutusta mallien tuloksille. Luvun lopussa käsitellään aineiston virhetermien mahdollisen keskinäisriippuvuuden malleille asettamaa vaatimusta. Kyse on pienestä mutta tärkeästä yksityiskohdasta naapurustovaikutuksia mallinnettaessa.

Differences-in-differences-mallit

Differences-in-differences-malleilla pyritään poistamaan havaitsemattomista ajan suhteen muuttumattomista yksilötason taustatekijöistä aiheutuva harha. Perusajatuksena on rakentaa regressioyhtälö, jossa selitetään selitettävässä muuttujassa toisiaan seuraavien ajanjaksojen välillä tapahtuneita muutoksia selittävässä muuttujissa tapahtuneiden vastaavana aikana tapahtuneiden muutosten avulla. Tällaisen mallin regressiokertoimet kuvaavat muutosten vaikutusta. Ajan suhteen muuttumattomien ominaisuuksien, myös niiden joita ei havaita, vaikutusta niissä ei havaita. (Ks. Angrist & Pischke 2015, 187–188; Wooldridge 2012, 459–461).

Differences-in-differencec-mallit tuskin takaa naapurustovaikutusestimaattien harhattomuuden. Ensinnäkin, malli ei kontrolloi havaitsemattomia yksilötason ominaisuuksia, jotka muuttuvat ajassa (Galster ym. 2015). Toiseksi, menetelmä ei kykene ratkaisemaan endogeenisuuden asettamaa haastetta. Erityisesti on huomattava, että difference-malleissa tehtävä oletus muuttujien muutosten riippumattomuudesta suhteessa lähtötilanteeseen ei välttämättä pidä paikkaansa (Galster & Hedman 2013). Esimerkiksi kohdenaapurustoon valikoituminen voi olla yhteydessä lähtönaapuruston vaikutuksiin.

Difference-malleihin liittyy muitakin rajoituksia. Ensinnäkin, vain ajan suhteen muuttuvien muuttujien regressiokertoimet on mahdollista estimoida, siis esimerkiksi sukupuolen vaikutus ei tule estimoiduksi. Toiseksi, kun malleissa tarkastellaan ajanjaksojen välillä tapahtunutta muutosta, selitettävän muuttujan vaihtelu käy helpoksi pieneksi, mikä johtaa keskivirheen kasvamiseen ja sitä kautta tilastollisen voiman pienenemiseen (Wooldridge 2012, 462).

Kolmanneksi, difference-mallit asettavat reunaehdoja tutkimuskysymysten esittämiselle. Suuri rajoitus nousee siitä, että mikäli naapurustot muuttuvat aineistossa

suhteellisen hitaasti, tieto naapurustotason muuttujien vaikutuksesta tulee käytännöllisesti katsoen yksilöiltä, jotka ovat muuttaneet tarkasteltavien ajanjaksojen välissä ominaisuuksiltaan lähtönaapurustoista poikkeaviin naapurustoihin. Jos tutkimuksessa halutaan kohdella naapurustotason muuttujia kategorisina, tulee tieto pelkästään muuttajilta.

Miksi tämä olisi rajoittavaa? Ensinnäkin, kategoriset luokittelut ovat tärkeitä etsittäessä mahdollisia epälineaarisuuksia naapuruston ominaisuuksien ja yksilötason ilmiön välisessä suhteessa (ks. esim. Galster 2008). Toiseksi, muuttajat eivät välttämättä edusta kovin hyvin koko väestöä. Kolmanneksi, muuttajiin rajauduttaessa on jotenkin kyettävä erottamaan muuttamisen itsenäiset vaikutukset mahdollisista naapuruston vaikutuksista (vrt. Sampson 2008).

Difference-malleja ovat käyttäneet esimerkiksi Bolster ja kumppanit (2007) sekä Galster ja kumppanit (2008, 2010). Mielenkiintoisena sovelluksena voidaan mainita, että Galsterin ja kumppaneiden (2008) ratkaisu laskea difference-mallit erikseen henkilöille, jotka eivät muuttaneet kahden tarkastelujakson aikana. Tällaisessa tilanteessa naapuruston ominaisuuksien muutoksien voidaan olettaa olevan eksogeenisiä, minkä seurauksena selitettävässä muuttujassa tapahtuneet muutokset on jonkin verran luotettavammin tulkittavissa naapurustosta johtuviksi. (Galster ym. 2008.)

Rajautumisessa yksilöihin, jotka eivät muuttaneet, voi olla omat ongelmansa. Ensinnäkin, yksilöt ovat voineet valikoitua naapurustoon silmällä pitäen sen tulevia muutoksia. Toiseksi, yksilöt, jotka eivät muuta, eivät välttämättä edusta perusjoukkoa kovin hyvin havaitsemattomien ominaisuuksien suhteen. Kolmanneksi, otoskoko jää tällaisella rajauksella helposti pieneksi, mistä kärsii tilastollinen voima. (Galster & Hedman 2013.)

Fixed effects -mallit

Difference-mallien tapaan myös fixed effects -malleissa hyödynnetään paneeliaineiston ominaisuuksia ajan suhteen muuttumattomien havaitsemattomien yksilötason ominaisuuksien kontrolloimiseen. Perusajatus on lisätä regressiomalliin kullekin yksilölle henkilökohtainen dummy-muuttuja, ns. fixed effect -termi. Kyseinen muuttuja kontrolloi kaikki yksilön ajan suhteen muuttumattomat ominaisuudet, mukaan lukien ne, joita ei ole havaittu aineistossa. Mallien laskeminen edellyttää paneeliaineistoa, joka tarjoaa tiedot havainnoista vähintään kahdelta ajanhetkeltä. (Wooldridge 2012, 488–489.) Käytännössä naapurustovaikutustutkimuksissa tiedot on kuitenkin yleensä haettu useammalta ajanhetkeltä. Se on ero differences-in-differences-malleihin, joissa tutkimusalan piirissä on tapana ollut tyytyä kahteen ajanhetkeen.

Fixed effects -mallit jakavat monet difference-mallien heikkouksista. Ensinnäkin, ne eivät kontrolloi endogeenisuutta eivätkä, toiseksi, ajan suhteen muuttuvia havaitsemattomia ominaisuuksia (Galster ja Hedman 2013). Kolmanneksi, niissä regressiokertoimet voidaan estimoida ainoastaan muuttujille, jotka muuttuvat

ajassa (Wooldridge 2012, 488). Neljänneksi, jälleen tieto naapurustoista tulee ai-noastaan niiltä yksilöiltä, joiden naapurustoissa on tapahtunut muutoksia paneelin mittausten välillä.

Muillakin osin fixed effects -mallien kritisoiminen on mahdollista. Ensinnäkin, niissä joudutaan tekemään vahva oletus, että ajansuhteen muuttumattomien ha-vaitsemattomien yksilötason ominaisuuksien vaikutus pysyy samanlaisena ajan kuluessa (Galster ym. 2015). Toiseksi, lukuisten henkilökohtaisten dummy-muut-tujien lisääminen kuluttaa mallin vapausastetta ja sitä kautta tilastollista voimaa (Galster & Hedman 2013).

Fixed effects -malleja ovat käyttäneet esimerkiksi Weinberg ja kumppanit (2004) sekä Musterd ja kumppanit (2012). Mainittakoon myös, että difference-mallien yhteydessä käsitelty idea tarkastelun rajaamisesta yksilöihin, jotka eivät muuta tarkastelujakson aikana, on mahdollista myös fixed effects -malleilla.

Instrumental variables -mallit

Instrumental variables -menetelmällä on teoriassa mahdollista laskea sekä vali-koitumisen että endogeenisuuden harhoista puhtaita estimaatteja. Valikoitumisen ongelman tapauksessa menetelmän logiikka on seuraavanlainen. Mallinnuksen ensimmäisessä vaiheessa lasketaan malli, jossa yksilön naapuruston ominaisuuksia ennustetaan hyödyntäen yhtä tai useampaa instrumentaalimuuttujaa. Tässä instrumenttimuuttujalla tarkoitetaan sellaista eksogeenistä muuttujaa, joka on yhteydessä naapuruston ominaisuuksiin ja joka ei ole yhteydessä naapurustoon valikoitumiseen liittyviin havaittuihin tai havaitsemattomiin yksilötason taustate-kijöihin eikä tutkimuksessa pääasiallisen mielenkiinnon kohteena olevaan yksilö-tason ilmiöön (outcome). Mallinnuksen toisessa vaiheessa lasketaan tutkimuksen pääasiallisen kiinnostuksen kohteena olevaa yksilötason ilmiötä selittävä malli, jossa alkuperäisten naapurustotason selittävien muuttujien sijaan käytetään niille edellä laskettuja estimaatteja. Ideaalitapauksessa näin saatavat estimaatit ovat puh-distettu havaitsemattomiin yksilötason muuttujien aiheuttamasta harhasta. (Ks. Angrist & Pischke 2015, 131–133.) Endogeenisuus mallin selittävästä muuttujista on mahdollista puhdistaa samanlaisella logiikalla.

Teoriassa instrumental variables -mallien käyttäminen on hyvin tehokas keino harhojen poistamiseksi, semminkin kun ei-muuttajiin rajautuvien difference tai fixed effects -mallien ohella se on ainoa ekonometrinen menetelmä, joka kontrolloi endogeenisuuden aiheuttamaa harhaa. Mutta käytännössä asia ei ole aivan niin yksinkertainen. Menetelmän ongelmaksi nousee vaikeus löytää sellaisia instru-menttimuuttujia, jotka olisivat sekä valideja että vahvoja. (Galster & Hedman 2013).

Instrumenttimuuttujan validiudelle on määritelty kolme ehtoa, jotka esimerkiksi aluetason selittävän muuttujan instrumenttimuuttujan kohdalla olisivat seuraa-vat: 1) Se on yhteydessä aluetason selittävään muuttujaan. 2) Se ei ole yhteydessä havaitsemattomiin yksilötason taustatekijöihin. 3) Se on yhteydessä selitettävään yksilötason ilmiöön nimenomaan selittävän aluetason muuttujan kautta, ei jota-

kin muuta kausaalipolkua pitkin. (Ks. Angrist & Pischke 2015, 106–107; Galster & Hedman 2013.) Esimerkin instrumenttimuuttujan vahvuus tarkoittaa puolestaan sitä, että se on voimakkaasti yhteydessä selittävään aluetason muuttujaan (Galster & Hedman 2013). Laadukkaiden instrumenttimuuttujien puutteessa estimointeihin syntyy helposti harhoja (Murray 2006).

Instrumental variables -malleja ovat käyttäneet esimerkiksi Damm (2014) ja Sari (2012). Galster ja Hedman (2013) ovat yhdistäneet menetelmän fixed effects -mallin kanssa.

Random effects -mallit

Random effects -malleissa havaitsemattomia yksilötason ominaisuuksia on mahdollista kontrolloida erottamalla virhetermistä komponentti, jonka estimaatti ilmaisee yksilöiden välisiä ajan suhteen pysyviä eroja² (Wooldridge 2012, 492–494). Toisella tapaa ajateltuna menetelmässä on kyse siitä, että paneeliaineiston yksilöille estimoidaan henkilökohtaiset vakiotermit.

Fixed effects ja difference-in-differences -malleihin nähden random effects -mallien etuna on se, että niissä myös muuttujille, jotka eivät vaihtele ajan mukaan, on mahdollista estimoida regressiokertoimet (Wooldridge 2012, 492). Hyvää on myös se, että toisin kuin fixed effects -malleissa vapausastetta eivät rokota yksilökohtaiset dummy-muuttujat. Asian käänttöpuoli on, että ajan suhteen muuttumattomat havaitsemattomat yksilötason ominaisuudet eivät tule yhtä hyvin kontrolloiduksi. Endogeenisuutta malli ei kontrolloi.³ (Galster & Hedman 2013.)

Random effects -malleja ovat käyttäneet esimerkiksi Kauppinen ja kumppanit (2011), Propper ja kumppanit (2005) sekä Wheaton ja Clarck (2003).

Sensitivity analysis

Edellä on esitelty naapurustovaikutustutkimuksen piirissä käytetyt valikoitumista tai endogeenisuutta kontrolloivat ekonometriset mallit. Ennen yhteenvetoa kiinnitetään huomiota vielä pariin seikkaan, joiden tuntemisesta voi olla apua naapurustovaikutusmalleja arvioidessa. Niistä ensimmäinen koskee sensitivity analysis -menetelmän käyttöä osana analyysiä.

2 Lähtökohtaisesti mallissa oletetaan, että havaitsemattomat yksilötason ominaisuudet eivät korreloi selittävän muuttujan kanssa (Wooldridge 2012, 492). Ehdosta on kuitenkin mahdollista joustaa, jos regressiomalliin tuodaan selittävän muuttujan ryhmäkeskiarvoa kuvaava muuttuja (Snijders & Boskers 2012, 48).

3 Kannattaa ehkä mainita, että myös sellaiset random effects -mallien ominaisuudet, jotka eivät liity harhohen kontrolloimiseen, voivat perustella niiden käyttämistä naapurustovaikutustutkimuksessa. Esimerkiksi niin sanotussa random intercept -mallissa on mahdollista arvioida, kuinka virhetermit jakaantuvat yksilö- ja naapurustotasoille. Kun tällaisessa mallissa on kontrolloitu kaikki relevantit yksilötason ominaisuudet, naapurustotason varianssi on teorian valossa mahdollista tulkita viitteiksi naapurustovaikutusten suuruudesta. (Kauppinen 2004, 77; Snijders & Bosker 2012, 41–54.)

Tilanteessa jossa tilastollinen malli ei kontrolloi puuttuvien muuttujien mahdollisesti aiheuttamaa harhaa täysin pomminvarmasti, tutkija voi olla kiinnostunut tietämään, kuinka robusteja hänen mallinsa tulokset ovat sen suhteen. Tätä varten on kehitetty erinäisiä tekniikoita, jotka kulkevat nimellä sensitivity analysis. Perusajatus on kaikissa sama. Koska puuttuvien muuttujien todellisen vaikutuksen suuruus ei ole tunnettu, niiden vaikutusta mallin tuloksille, kuten esimerkiksi regressiokertoimien estimaateille tai p-arvoille, tarkastellaan antamalla niille hypoteettisia arvoja. (Hosman ym. 2010.) Sensitivity analysis -menetelmiä naapurustovaikutustutkimuksessa ovat käyttäneet esim. Harding (2003) sekä Lovasi ja kumppanit (2009).

Vaikuttaisi siltä, että sensitivity analysis - menetelmiä ei ole käytetty kovin monissa naapurustovaikutustutkimuksissa. Olisi mielenkiintoista tietää, miksi on näin. Yksi mahdollinen selitys on, että tutkijat eivät pidä sen tuottamia arvioita kovinkaan käyttökelpoisina. Menetelmän mahdollisesti osakseen saama kritiikki menee kuitenkin hieman ohi tämän paperin aiheesta, eikä siihen tutustumiseen ole siksi mahdollista lähteä. Sensitivity analysis - menetelmät onkin mainittu tässä varsinaista mallintamista mahdollisesti palvelevina apukeinoina, joiden käyttökelpoisuudesta kannattaa ottaa selvää ryhdyttäessä empiiriseen tutkimukseen.

Naapurustovaikutukset ja havaintojen keskinäisriippuvaisuudet

Edeltävää ekonometrisiä menetelmiä koskevaa keskustelua kannattaa vielä täydentää huomiolla eräästä tutkimuskohteen erityispiirteestä. Jos tutkimusaineistossa esiintyy naapurustovaikutuksia, yksilöiden välille ei voi olettaa riippumattomuutta. Tämä johtuu siitä, että asukkaiden altistuessa vaikuttavalle naapuruston ominaisuudelle, he alkavat tietyltä osin muistuttaa toisiaan. (vrt. Ellonen 2006). Tilastollisen mallinnuksen näkökulmasta ilmiö tarkoittaa, että naapuruston asukkaiden virhetermit eivät ole toisistaan riippumattomia (Kauppinen 2004). Se on huomioitava regressiomalleja rakennettaessa, koska muuten niiden tuottamat keskivirheet eivät ole luotettavia (UCLA).

Onneksi mallin virhetermien keskinäisriippuvuutta ei tarvitse olettaa tai arvuutella, vaan se voidaan laskea. Sitä kuvaa niin sanottu sisäkorrelaatioluku (intra-class correlation). Jos sisäkorrelaatiota esiintyy, se voidaan huomioida ainakin kahdella tavalla. Yksi mahdollisuus on käyttää jo edellä mainittuja random effects -malleja. Toisen vaihtoehdon tarjoavat niin sanotut clustered robust standard errors -mallit (UCLA). Random effects ja clustered robust standard errors -malleilla on kummallakin omat vahvuutensa, ja riippuu osin verran aineiston ominaisuuksista, kumman valitseminen on tarkoituksenmukaisempaa (ks. esim. Snijders & Bosker 2012, 199–200; UCLA). Clustered robust standard errors -mallien ansioksi voidaan katsoa, että ne ovat yhdistettävissä ainakin fixed effects ja differences-in-differences -malleihin (Nichols & Staffer 2007). On vaikea nähdä syytä, mikseivät ne toimisi myös instrumental variables -mallien kanssa.

Jostakin syystä monista naapurustovaikutuksia käsittelevissä tutkimusartikkeleissa artikkeleissa virhetermien keskinäisriippuvuudesta ei ole huolehdittu millään tavalla. On vaikea sanoa, johtuuko se tietämättömyydestä vai siitä, että tutkijat eivät pidä ongelmaa kovin suurena. Papereita, joissa asia on huomioitu keskivirheitä laskettaessa, ovat esimerkiksi Galster ym. (2015), Van Ham & Manley (2010), Weinberg ym. (2004). Edellisten lisäksi tietysti myös aikaisemmin mainitut random effects -malleja käyttävät tutkimukset huolehtivat asiasta.

Yhteenveto ekonometrisistä menetelmistä

Luonnollisesti syntyneeseen havaintoaineistoon perustuvat yksinkertaiset regressioanalyysit ovat yleensä alttiita valikoitumisesta tai endogeenisuudesta aiheutuille harhoille. Differences-in-differences, fixed effects, instrumental variables ja random effects -mallit korjaavat tilannetta kukin tavallaan ja omine rajoituksineen.

Menetelmien asettaminen paremmuusjärjestykseen ei ole helppoa, koska se, mikä on hyvää teoriassa, ei ole sitä välttämättä käytännössä. Esimerkiksi instrumental variables -malli on sikäli ylivoimainen, että valikoitumisen lisäksi se kontrolloi myös endogeenisuuden ongelmaa. Mallin ominaisuuksista ei ole kuitenkaan apua, niin kauan kuin tarkoituksenmukaisia instrumentteja ei ole saatavilla. Jos kuitenkin hetkeksi unohdetaan olemassa olevien aineistojen mallien käyttämiselle mahdollisesti asettamat rajoitteet, Galsterin ja Hedmanin (2013) käsitys mallien paremmuusjärjestyksestä lienee oikeansuuntainen. Vahvin on fixed effects -malli instrumenttimuuttujilla. Sen jälkeen tulevat fixed effects -malli henkilöille, jotka eivät muuta tarkastelujaksolla, fixed effects -malli, differences-in-differences -malli ja random effects -malli.⁴ Vertailu perustuu mallien kykyihin kontrolloida valikoitumisen ja endogeenisuuden harhoja, eikä ota kantaa niiden muihin ominaisuuksiin.

Empiirisessä tutkimuksessa käyttökelpoisimman mallin ratkaisevat lopulta tutkimuskysymykset ja käytössä oleva aineisto. Kenties tapauksessa kuin tapauksessa tavoiteltavana voidaan pitää Galsterin ja Hedmanin (2013) ajatusta siitä, että tutkimuksessa on hyödynnettävä rinnan erilaisia ekonometrisia tekniikoita. Siitä huolimatta, että jokaiseen mallintamisen tapaan liittyy puutteita, niiden estimaatteja vertailemalla voidaan ehkä jossakin määrin arvioida, millaisessa haarukassa todellisten naapurustovaikutusten arvot liikkuvat (Galster ja Hedman 2013).

⁴ Fixed effects -malleja Galster ja Hedman suosivat difference-mallien kustannuksella ilmeisesti siitä syystä, että edelliset käyttävät tehokkaammin aineiston informaatiota hyväkseen. Naapurustovaikutustutkimuksessahan differences-mallit on ollut tapana laskea ainoastaan kahden ajanhetken tai -jakson perusteella.

LOPUKSI

Naapurustovaikutustutkijaa kiinnostaa, voivatko jotkin naapuruston ominaisuudet, kuten esimerkiksi keskittynyt huono-osaisuus, vaikuttaa sen asukkaiden elämään. Tällaisiin kausaalisuhdetta tarkasteleviin kysymyksiin vastaaminen on osoittautunut hyvin haastavaksi. Tutkimusaineistojen avulla on yleensä ollut hyvin vaikea osoittaa, että mahdollisesti havaitut tilastolliset yhteydet ovat seurausta juuri naapuruston vaikutuksesta. Erityisesti kausaalipäättelyä ovat vaivanneet puuttuvista tai endogeenisuudesta kärsivistä muuttujista mahdollisesti johtuvat harhat.

Tehty katsaus on pyrkinyt luomaan mahdollisimman kattavan kuvan empiirisessä tutkimuskirjallisuudessa tehdyistä yrityksistä kontrolloida puuttuvista tai endogeenisuudesta kärsivistä muuttujista mahdollisesti aiheutuvat harhat. Valitettavasti mikään ei takaa, että syntynyt kuva on täydellinen. Naapurustovaikutuskirjallisuus on niin monilukuinen, että kaikkien tutkimusten läpi käyminen ei ole yhdelle ihmiselle kohtuullisessa ajassa mahdollista. Siksi joitakin kiinnostavia tutkimusasetelmia on voinut jäädä huomiotta.

Oli miten oli, puuttuviin tai endogeenisuudesta kärsiviin muuttujiin juurensa juontavia metodologisia haasteita on ratkottu empiirisissä tutkimusasetelmissä useilla erilaisilla tavoilla. Yksi tapa on ollut valita analyysin kohteeksi asuntopoliittisessa interventiossa syntynyt aineisto, jossa aineiston yksilöiden sijoittumista naapurustoihin voidaan pitää havaitsemattomien yksilötason tekijöiden suhteen satunnaisena. Tietynlaisiin naapurustovaikutuskysymyksiin vastaamisessa tällaiset kvasikokeelliset tutkimusasetelmat ovat periaatteessa parhaita mahdollisia. Valitettavasti niiden edellyttämät interventiot ovat ani harvassa, esimerkiksi Suomessa niitä ei tiedetä tehdyn.

Toinen tapa on ollut etsiä niin sanotusti luonnollisesti syntyneitä koejärjestelyn kaltaisia tutkimusasetelmia, joissa aineiston yksilöt ovat oletettavasti sijoittuneet naapurustoihin havaitsemattomien taustatekijöiden suhteen melko satunnaisesti. Esimerkiksi sosiaalisten vuokra-asuntojen jakamisen on ehdotettu olevan yksi tällainen satunnaisuutta luova mekanismi. Lähestymistavan haaste on löytää tutkimusasetelmia, joissa oletuksia satunnaisuudesta voidaan pitää riittävän uskottavina.

Kolmannen tavan poistaa, tai ainakin pienentää, estimaatteja uhkaavaa harhan riskiä ovat tarjonneet erinäiset ekonometriset menetelmät. Katsauksessa niistä on käsitelty differences-in-differences, fixed effects, instrumental variables ja random effects -malleja. Menetelmistä voi olla apua, mutta on epätodennäköistä, että niiden avulla harhan riski poistuu kokonaan.

Puuttuvien tai endogeenisuudesta kärsivien muuttujien aiheuttamiin harhoihin keskittyvän paperin lopussa on kenties paikallaan todeta, että vaikka kyseisten harhojen huomioiminen on välttämätöntä uskottavien kausaaliväitteiden esittämiselle, se ei vielä yksistään riitä. Lisäksi tarvitaan hyvää, empiirisesti perusteltua teoriaa naapurustovaikutuksia tuottavista mekanismeista.

Tilastollisten yhteyksien tulkitseminen syistä vaikutuksiksi kulkeviksi prosesseiksi tulee mahdolliseksi vasta vasten teoriaa. Useat tutkijat ovat valittaneet, että ylenmääräisestä keskittymisestä tilastollisiin harhoihin on seurannut mekanismeja koskevan ymmärryksen laiminlyönti (ks. esim. Sampson 2011, Small & Feldman 2012, Van Ham ym. 2012). Jatkotutkimuksessa uskottavalta naapuruston kausaalivaikutusta koskevien väitteiden esittämiseltä edellytetään vahvaa teoreettista ymmärrystä sekä ekonometriaan että tutkimusalan substanssiin kuuluvista asioista.

Liitetaulukko 1. Empiiriset tutkimusasetelmat (Huom. Kattavat tiedot heikkouksista löytyvät leipätekstistä)

Tutkimusasetelma	Vahvuudet	Heikkoudet
Koejärjestelyn kaltaiset tutkimusasetelmat, joiden aineiston taustalta löytyy interventio	<ul style="list-style-type: none"> Satunnaistaminen Eivät kärsi valikoitumisesta Naapuruston vaihtaminen interventiona Eivät kärsi endogeenisuudesta 	<ul style="list-style-type: none"> Aineistoja ei Suomessa saatavilla
Koejärjestelyn kaltaiset tutkimusasetelmat, jotka käyttävät luonnollisesti syntyneitä aineistoja		
Samanaapuruston asukkaat	<ul style="list-style-type: none"> Jonkinasteinen satunnaistaminen Ei oletettavasti kärsi valikoitumisesta 	<ul style="list-style-type: none"> Oletukset aineiston satunnaisuudesta eivät välttämättä pidä paikkaansa Endogeenisuus
Sosiaalisten vuokra-asuntojen asukkaat	<ul style="list-style-type: none"> Jonkinasteinen satunnaistaminen Ei oletettavasti kärsi valikoitumisesta 	<ul style="list-style-type: none"> Oletukset aineiston satunnaisuudesta eivät välttämättä pidä paikkaansa Endogeenisuus (paitsi asetelmissä, joissa muuttamisen ajankohta on kvasisatunnainen)
Sisarukset	<ul style="list-style-type: none"> Jonkinasteinen satunnaistaminen Ei oletettavasti kärsi valikoitumisesta 	<ul style="list-style-type: none"> Perheen ja lapsen havaitsemattomiin ominaisuuksiin liittyvät oletukset eivät välttämättä pidä paikkaansa Endogeenisuus Muuttajien vähäinen vaihtelu
Havaintoaineistoa käyttävät tutkimusasetelmat		
Difference-in-difference-mallit	<ul style="list-style-type: none"> Kontrolloi havaitsemattomista ajan suhteen muuttumattomista muuttujista johtuvan harhan 	<ul style="list-style-type: none"> Ei kontrolloi havaitsemattomia muuttujia, jotka muuttuvat ajassa Tieto etupäässä yksilöistä, jotka ovat muuttaneet Estimaatit vain ajan suhteen muuttaville muuttujille Endogeenisuus (paitsi kun tarkastelu rajataan ei-muuttaneisiin) Selitettävän muuttujan vaihtelun pienentyessä tilastollinen voima heikkenee

Fixed effects -mallit	<ul style="list-style-type: none"> • Kontrolloii havaitsemattomista ajan suhteen muuttumattomista muuttujista johtuvan harhan • Käyttää paneeliaineiston informaatiota tehokkaammin kuin kahden ajanhetken eroa tarkastelevat difference-mallit 	<ul style="list-style-type: none"> • Ei kontrolloii havaitsemattomia muuttujia, jotka muuttuvat ajassa • Tieto etupäässä yksilöistä, jotka ovat muuttaneet • Estimaatit vain ajan suhteen muuttuville muuttujille • Endogeenisuus (paitsi kun tarkastelu rajataan ei-muuttaneisiin) • lukuisat fixed effect -termit heikentävät tilastollista voimaa
Instrumental variables -mallit	<ul style="list-style-type: none"> • Kontrolloii valikoitumisesta johtuvan harhan • Kontrolloii endogeenisuudesta johtuvan harhan 	<ul style="list-style-type: none"> • Validien ja vahvojen instrumenttimuuttujien löytäminen
Random effects -mallit	<ul style="list-style-type: none"> • Kontrolloii havaitsemattomia ajan suhteen muuttumattomia muuttujia • Sallii estimoida ajan suhteen muuttumattomien muuttujien regressiokertoimia • Mahdollista jakaa virhetermi yksilö- ja aluetason komponenteiksi 	<ul style="list-style-type: none"> • Ei kontrolloii havaitsemattomia ajan suhteen muuttumattomia muuttujia täydellisesti • Endogeenisuus muuttujien regressiokertoimia • Mahdollista jakaa virhetermi yksilö- ja aluetason komponenteiksi

LÄHTEET

- Aaronson, D. (1998).** *Using sibling data to estimate the impact of neighborhoods on children's educational outcomes.* Journal of Human Resources, 33(4), 915-946.
- Angrist, J. D., & Pischke, J. (2008).** *Mostly harmless econometrics: An empiricist's companion.* Princeton: Princeton university press.
- Angrist, J. D., & Pischke, J. (2015).** *Mastering' metrics: The path from cause to effect.* Princeton: Princeton University Press.
- Bernelius, V. (2013).** *Eriytyvät kaupunkikoulut: Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä.* Tutkimuksia 2013:1. Helsinki: Helsingin kaupungin tietokeskus.
- Bolster, A., Burgess, S., Johnston, R., Jones, K., Propper, C., & Sarker, R. (2007).** *Neighbourhoods, households and income dynamics: A semi-parametric investigation of neighbourhood effects.* Journal of Economic Geography, 7(1), 1-38.
- Damm, A. P. (2014).** *Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants.* Journal of Urban Economics, 79, 139-166.
- Ellonen, N. (2006).** *Monitasoanalyysit ja niiden soveltaminen sosiaalitieteissä.* Janus, 14(2), 127-138.
- Galster, G. (2003).** *Investigating behavioural impacts of poor neighbourhoods: Towards new data and analytic strategies.* Housing Studies, 18(6), 893-914. doi:10.1080/0267303032000135492.
- Galster, G. (2008).** *Quantifying the effect of neighbourhood on individuals: Challenges, alternative approaches, and promising directions.* Schmollers Jahrbuch, 128(1), 7-48.
- Galster, G., & Hedman, L. (2013).** *Measuring neighbourhood effects non-experimentally: How much do alternative methods matter?* Housing Studies, 28(3), 473-498.
- Galster, G., Andersson, R., & Musterd, S. (2014).** *Are males' incomes influenced by the income mix of their male neighbors? Explorations into nonlinear and threshold effects in Stockholm.* Housing Studies, (ahead-of-print), 1-29.
- Galster, G., Santiago, A., Lucero, J., & Cutsinger, J. (2015).** *Adolescent neighborhood context and young adult economic outcomes for low-income african americans and latinos.* Journal of Economic Geography, 15, 1-33. doi:10.1093/jeg/lbv004.
- Harding, D. J. (2003).** *Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy.* American Journal of Sociology, 109(3), 676-719.
- Heck, R. H., & Thomas, S. L. (2009).** *An introduction to multilevel modeling techniques* (2nd ed.). New York: Routledge.
- Hedman, L., & Galster, G. (2013).** *Neighbourhood income sorting and the effects of neighbourhood income mix on income: A holistic empirical exploration.* Urban Studies, 50(1), 107-27.

- Hosman, C. A., Hansen, B. B., & Holland, P. W. (2010).** *The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder.* *The Annals of Applied Statistics*, , 849–870.
- Kauppinen, T. M. (2004).** *Asuinalueen ja perhetaustan vaikutukset helsinkiläisnuorten keskiasteen tutkintojen suorittamiseen.* *Tutkimuksia* 2004:6. Helsinki: Helsingin kaupungin tietokeskus.
- Kling, J. R., Liebman, J. B., & Katz, L. F. (2007).** *Experimental analysis of neighborhood effects.* *Econometrica*, 75(1), 83–119.
- Lankinen, M. (2006).** *Sosiaalisen vuokra-asumisen asema kaupunkikentässä.* *Tutkimuksia* 2006, 7. Helsinki: Helsingin kaupungin tietokeskus.
- Lovasi, G. S., Neckerman, K. M., Quinn, J. W., Weiss, C. C., & Rundle, A. (2009).** *Effect of individual or neighborhood disadvantage on the association between neighborhood walkability and body mass index.* *American Journal of Public Health*, 99(2), 279–84. doi:10.2105/AJPH.2008.138230 [doi].
- Ludwig, J., Duncan, G. J., Genetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013).** *Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity.*
- Manley, D., & Van Ham, M. (2012).** Neighbourhood effects, housing tenure and individual employment outcomes. Teoksessa **Van Ham, M., Manley, D., Bailey, N., Simpson, L. & Maclennan, D. (toim.)**, *Neighbourhood effects research: New perspectives.* New York: Springer, 147–173.
- Musterd, S., Galster, G., & Andersson, R. (2012).** *Temporal dimensions and measurement of neighbourhood effects.* *Environment and Planning-Part A*, 44(3), 605.
- Murray, M. P. (2006).** *Avoiding invalid instruments and coping with weak instruments.* *The Journal of Economic Perspectives*, 20(4), 111–132.
- Nichols, A. & Staffer, M. 2007.** *Clustered errors in stata.* Luentokalvot. [http://www.stata.com/meeting/13uk/nichols_crse.pdf]. Katsottu 10.11.2015.
- Oreopoulos, P. (2003).** *The long-run consequences of living in a poor neighborhood.* *The Quarterly Journal of Economics*, 1533–575.
- Overman, H. G. (2002).** *Neighbourhood effects in large and small neighbourhoods.* *Urban Studies*, 39(1), 117–130.
- Propper, C., Jones, K., Bolster, A., Burgess, S., Johnston, R., & Sarker, R. (2005).** *Local neighbourhood and mental health: Evidence from the UK.* *Social Science & Medicine*, 61(10), 2065–2083.
- Sampson, R. J. (2008).** *Moving to inequality: Neighborhood effects and experiments meet structure.* *AJS; American Journal of Sociology*, 114(11), 189.
- Sampson, R. J. (2012).** *Great American city: Chicago and the enduring neighborhood effect* University of Chicago Press.
- Sari, F. (2012).** *Analysis of neighbourhood effects and work behaviour: Evidence from paris.* *Housing Studies*, 27(1), 45–76.
- Small, M. L., & Feldman, J. (2012).** *Ethnographic evidence, heterogeneity, and neighbourhood effects after moving to opportunity.* *Neighbourhood effects research: New perspectives* (57–77). New York: Springer.

- Van Ham, M., & Manley, D. (2010).** *The effect of neighbourhood housing tenure mix on labour market outcomes: A longitudinal investigation of neighbourhood effects.* Journal of Economic Geography, 10, 257-282.
- Van Ham, M., Manley, D., Bailey, N., Simpson, L., & Maclennan, D. (2012).** Neighbourhood effects research: New perspectives. In **M. Van Ham, D. Manley, N. Bailey, L. Simpson & D. Maclennan (Eds.),** *Neighbourhood effects research: New perspectives* (1-21). New York: Springer.
- Weinberg, B., Reagan, P. B., & Yankow, J. J. (2004).** *Do neighborhoods affect work behavior? Evidence from the NLSY79.* Journal of Labor Economics, 22(4), 891-924.
- Wheaton, B., & Clarke, P. (2003).** *Space meets time: Integrating temporal and contextual influences on mental health in early adulthood.* American Sociological Review, 68(5), 680-706.
- Wooldridge, J. (2012).** *Introductory econometrics: A modern approach.* Mason: Cengage Learning.

Verkko-lähteet:

UCLA. [<http://www.ats.ucla.edu/stat/stata/library/cpsu.htm>] Katsottu 6.11.2015.

TIEDUSTELUT

Henri Laukkanen
p. 09 310 36366
etunimi.sukunimi@hel.fi

KUVIOT JA TAITTO

Lotta Haglund

KANNET

Tarja Sundström-Alku

VALOKUVA

Visit Helsinki / Timo Santala

JULKAISIJA

Helsingin kaupungin tietokeskus
PL 5500 (Siltasaarenkatu 18–20 A)
00099 Helsingin kaupunki

PUHELINVAIHDE

09 310 1612

INTERNET

www.hel.fi/tietokeskus/

ISSN 2342-6438
(verkkojulkaisu)