

Selittävätkö hyvinvointierot odotettua alhaisempaa poismuuttoa eräiltä korkean työttömyyden alueilta?

OLLI LEHTONEN & MARKKU TYKKYLÄINEN

Artikkeli on julkaistu Yhteiskuntapolitiikka-lehden numerossa 2/2013.

Liite 1. Samanlaisuusindeksin määrittely ja käyttö

Määrittelemme samanlaisuusindeksin kunnassa i seuraavasti. Olkoon L_i kahden muuttujan y ja x spatiaalista rakennetta kuvaava samanlaisuusindeksi. Tällöin

$$L_i = f(y_{1,i}, y_{2,i}, x_{1,i}, x_{2,i}) \quad (1),$$

jossa $y_{1,i}$ ovat työttömyyden havaintoarvot kunnassa i ja $y_{2,i}$ ja työttömyyden havaintoarvot kunnan i viereisissä kunnissa ja $x_{1,i}$ ovat nettomuuton havaintoarvot kunnassa i ja $x_{2,i}$ nettomuuton havaintoarvot sen viereisillä alueilla. Molemmat muuttujat ovat näin spatiaalisesti viivästettyjä. Sekä työttömyyden että nettomuuton havaintoarvot ovat standardisoitu ennen indeksin laskentaa, jotta niiden hajontojen erot eivät vaikuttaisi tuloksiin. Lisäksi työttömyyden standardisoidut arvot käännettiin vastaluvuikseen, jolloin molempien muuttujien negatiiviset arvot kuvasivat samansuuntaisesti epätasapainossa olevien työntö–tappioominaisuuksia eli keskimääräistä korkeampaa työttömyyttä ja suurempaa muuttotappiota. Globaaleissa ja lokaaleissa spatiaalisissa autokorrelaatioindekseissä havaintoarvojen standardisointi on yleinen käytäntö, millä eliminoidaan muuttujien erilaisista mittakaavoista aiheutuvat vääristymät (Anselin 1995, 95). Kunnan i viereisten kuntien arvot eli spatiaalisesti viivästetyt työttömyysasteen ja nettomuuton arvot lasketaan spatiaalisella painomatriisin $W = [w_{ij}]$ avulla. Samanlaisuusindeksissä spatiaalisen painomatriisi voidaan määritellä, kuten paikallisen spatiaalisen autokorrelaation laskennassa yleensä, usealla eri menetelmällä. Muodostimme painomatriisin queen-menetelmällä, jossa viereisiksi kunniksi määrittyvät kunnat, joilla on yhteistä rajaa vähintään yhden pisteen verran (Robinson 1998, 275). Spatiaalista painomatriisia käytetään tulkinnallisesta syystä johtuen rivistandardoituna, jolloin kukin y :n ja x :n havaintoarvo muodostuu alueen i ja sen viereisten alueiden havaintojen keskiarvona (Anselin 1995, 95).

Kun merkitään paikkavektorit nettomuuttoliikkeelle $a = (x_{1,i}, x_{2,i})$ ja työttömyysasteelle $b = (y_{1,i}, y_{2,i})$, voidaan tarkemmin samanlaisuusindeksi (θ_i) määritellä seuraavasti:

$$\theta_i = \text{acos} \frac{a \cdot b}{\|a\| \|b\|} \cdot \left(\|a\| - \|b\| \right) \quad (2),$$

missä $\text{acos} \frac{a \cdot b}{\|a\| \|b\|}$ on kahden paikkavektorin a (työttömyysaste) ja b (nettomuutto) välinen kulma, ja $\|a\| - \|b\|$ on itseisarvo paikkavektorien a ja b euklidisten origo-ettäisyyksien erotuksesta. Edellistä indeksin osaa voidaan nimittää kulmaksi ja jälkimmäistä ilman itseisarvo otamista intensiteetiksi. Merkinnät $\|a\|$ ja $\|b\|$ ovat paikkavektorin normit kunnassa i , ja $a \cdot b$ on vektoreiden a ja b pistetulo myös kunnassa i . Samanlaisuusindeksi saa pienimmän arvon 0 kun $\cos \frac{a \cdot b}{\|a\| \|b\|} = 0$ tai kun $\|a\| - \|b\| = 0$, koska molemmissa tapauksissa paikkavektorit a ja b olisivat samat. Indeksille ei ole määritely suurinta arvoa. Kirjoittamalla auki paikkavektorit a ja b , voidaan indeksi kirjoittaa myös seuraavasti:

$$\theta_i = \text{acos} \frac{x_{1,i}y_{1,i} + x_{2,i}y_{2,i}}{\sqrt{x_{1,i}^2 + x_{2,i}^2} \sqrt{y_{1,i}^2 + y_{2,i}^2}} \cdot \left| \sqrt{y_{2,i}^2 + x_{2,i}^2} - \sqrt{y_{1,i}^2 + x_{1,i}^2} \right| \quad (3).$$

Kuten yhtälöstä 2 nähdään, kahden muuttujan samanlaisuusindeksi koostuu kahden osatekijän tulosta: intensiteetistä eli spatiaalisesti viivästettyjen muuttujien erotuksesta ja kilmasta eli autokorrelatiivisten rakenteiden samankaltaisuudesta. Intensiteetti tarkoittaa kaa-
vassa 2 paikkavektorien a (nettomuuttoliike) ja b (työttömyysaste) origoetäisyyksien erotusta ilman itseisarvoa. Erotuksen negatiiviset arvot tarkoittavat, että työttömyys on odotettua korkeampaa suhteessa muuttoliikkeeseen. Tällöin muuttoliike ei ole riittävää vähentämään korkeaa työttömyyttä. Mitä heikommin muuttoliike vastaa työmarkkinoiden epätasapainoon, sitä enemmän muuttoliikkeellä oletamme olevan sosioekonomisia ja institutionaalisia esteitä. Tasapainotilanne edellyttäisi korkean työttömyyden kunnassa myös korkeaa poismuuttoa. Samanlaisuusindeksin toinen osatekijä, aluerakenteiden samankaltaisuus, kuvaa kuinka samankaltaisia työttömyyden ja muuttoliikkeen spatiaaliset muodostumat ovat kunnassa ja sitä ympäröivissä kunnissa.

Samanlaisuusindeksin tilastollinen testaus perustuu ehdolliseen satunnaistamiseen, jossa luodaan permutaatiojakauma vale-merkitsevyyden (*pseudo-significance*) laskemista varten. Satunnaistaminen on ehdollinen ja laskennallisesti raskas, koska muuttujien arvot $x_{1,i}$ ja $y_{1,i}$ ovat vakioitu ja loput arvot ovat satunnaisesti permutoitu näiden ympärille (Anselin 1995). Permutaatioista muodostuvalla jakaumalla voidaan arvioida havaitun indeksin harvinaisuutta verrattuna satunnaistettuun tilanteeseen. Työntö-tappio -kunnille määritimme satunnaistamisella permutaatiojakaumasta rajat työttömyysasteen ja muuttoliikkeen tasapainolle. Kun nämä olivat tasapainossa, samanlaisuusindeksin havaitut arvot saivat poiketa korkeintaan 5 prosentin merkitsevyydellä permutoimalla saadusta nollahypoteesin mukaisesta jakaumasta.

Liite 2: Yleistetty additiivinen malli ja siinä käytetyt muuttujat

Yleistetyssä additiivisessa mallissa selitettävän muuttujan ja selittävien muuttujien välistä yhteyttä mallinnetaan tasoittavilla funktioilla, jotka estimoivat parametrittoman epälineaarisen sovituksen selittävän muuttujan ja selitettävän muuttujan välille. Mallia sanotaan additiiviseksi, koska sen muuttujien yksittäiset vaikutukset lisätään toisiinsa. Additiivisten mallien vahvuudet liittyvät niiden kykyyn mallintaa myös epälineaarisia ja epämonotonisia yhteyksiä selittävän ja selitettävän muuttujan välillä (Wood 2006). Työntö-tappio-kuntia mallinnettaessa mallin hyödyt liittyvät myös epäparametrisuuteen, koska mallinnettavien kuntien lukumäärä on alhainen. Tulosten vertailtavuuden vuoksi estimoimme myös perinteisen lineaarisen regressiomallin pienimmän neliösumman menetelmällä.

Käyttämämme yleistetty additiivinen malli voidaan kirjoittaa seuraavasti

$$y = \beta_0 + \sum_{j=1}^4 f_j(\text{tulot}_j) + \sum_{j=1}^4 f_j(\text{terveys}_j) + \sum_{j=1}^4 f_j(\text{päähteet}_j) + \sum_{j=1}^4 f_j(\text{rikosuhka}_j) + \varepsilon \quad (4),$$

missä selitettävä muuttuja y tarkoittaa kunnan samanlaisuusindeksiä ja termit f_1, \dots, f_4 ovat selittävien muuttujien tulot, terveys, päähteet ja rikosuhka, niin sanottuja tasoitettuja funktioita. Tasoittavien funktioiden tasoitusparametrit estimoitiin GCV-menetelmällä (*generalized cross validation*) (Wood 2004).

Toimeentuloa eli aineellista hyvinvointia kuvaava muuttujaryhmä eli rivivektori sisältää seuraavat muuttujat:

- *toimeentulotuki*: muuttuja ilmaisee kalenterivuoden aikana toimeentulotukea saaneiden 25–64-vuotiaiden pitkäaikaisasiakkaiden osuuden prosentteina vastaavanikäisestä väestöstä. Muuttuja ilmaisee, kuinka suuri osuus parhaassa työiässä olevasta väestöstä on toimeentulotuen piirissä pitkäaikaisesti.
- *pienituloiset kotitaloudet*: muuttuja ilmaisee pienituloisiin kotitalouksiin kuuluvien henkilöiden osuuden prosentteina kaikista alueella asuvista henkilöistä. Pienituloisuuden rajana käytetään 60 % suomalaisten kotitalouksien käytettävissä olevan ekvivalentin rahatulon (uudistetulla OECD-skaalalla laskien) mediaanista kunakin vuonna.

Terveystilaa kuvaaviin muuttujiin otettiin mukaan sairastavuusindeksi, erityiskorvattaviin lääkkeisiin oikeutetut ja masennuslääkkeiden käyttö. Nämä muuttujat määritellään seuraavasti:

- *sairastavuusindeksi*: muuttuja ilmaisee jokaiselle Suomen kunnalle lasketun indeksin avulla miten tervettä tai sairasta väestö on suhteessa koko maan väestön keskiarvoon (= 100). Indeksillä perustuu kolmeen rekisterimuuttuajaan: kuolleisuuteen, työkyvyttömyyseläkkeellä olevien osuuteen työikäisistä ja erityiskorvattaviin lääkkeisiin oikeutettujen osuuteen väestöstä ja se on laskettu ikävakioiduna.
- *erityiskorvattaviin lääkkeisiin oikeutetut*: muuttuja ilmaisee vuoden lopussa erityiskorvattaviin lääkkeisiin oikeutettujen henkilöiden osuuden prosentteina väestöstä. Oikeus erityiskorvattavaan lääkitykseen toimii kohtalaisen hyvänä pitkäaikaissairastavuuden indikaattorina.
- *masennuslääkkeiden käyttö*: muuttuja ilmaisee vuoden aikana depressiolääkkeistä korvautta saaneiden 25–64-vuotiaiden osuuden prosentteina vastaavanikäisestä väestöstä.

Kolmas muuttujaryhmä muodostui päihteiden käyttöä kuvaavista muuttujista:

- *Alkoholimyynti*: muuttuja ilmaisee alkoholijuomien myynnin litroina asukasta kohti 100 %:n alkoholina. Muuttuja kuvaa epäsuorasti alkoholin kulutusta kunnissa.
- *Päihdehoitokustannukset*: muuttuja ilmaisee kuntien kustantamien päihdehuollon erityispalveluihin sekä raittiusryhmiin suunnattujen palveluiden nettokustannuksia asukasta kohden. Muuttuja kuvaa epäsuorasti alkoholin ja huumeiden käytön aiheuttamia ongelmien määrää ja niiden yleisyyttä.

Rikosten uhriksi joutumisen todennäköisyyttä ja samalla turvattomuutta kuvaavat tiedot poliisin tietoon tulleista rikoksista:

- *omaisuusrikokset*: muuttuja ilmaisee poliisin tietoon tulleiden omaisuusrikosten osuuden tuhatta asukasta kohti.
- *henkirikokset*: muuttuja ilmaisee poliisin tietoon tulleiden väkivaltarikosten eli henkeen ja terveyteen kohdistuneiden rikosten osuuden tuhatta asukasta kohti.

Selittävät muuttujat valikoituvat tästä muuttujajoukosta.

KIRJALLISUUS

- Anselin, Luc: Local indicator of spatial association: LISA. *Geographical Analysis* 27 (1995): 2, 93–115.
- Robinson, Guy: *Methods and Techniques in Human Geography*. Chichester: John Wiley, 1998.
- Wood, Simon: Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association* 99 (2004), 673–686.
- Wood, Simon: *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC, 2006.