

Ekstrapolointi ja estimointitaajuuden valinta regressioepäjatkuvuusasetelmassa

Miikka Rokkanen

Regressioepäjatkuvuusasetelmassa yksilöiden jakautuminen koe- ja kontrolliryhmään määrittyy sen mukaan, ylittääkö jonkin ryhmäjaossa käytettävän muuttujan arvo ennalta asetetun jakokynnyksen. Yksilö kuuluu koeryhmään, kun muuttujan arvo ylittää kynnyksen. Muussa tapauksessa hän kuuluu kontrolliryhmään. Mikäli ryhmäjaossa käytettävä muuttuja on yhteydessä mielenkiinnon kohteena olevaan vaste-muuttujaan, koe- ja kontrolliryhmä eivät ole vertailukelpoisia käsittelyn vaikuttavuutta ajatellen. Asetelma kuitenkin mahdollistaa hyvin lievien oletusten vallitessa toimenpiteen kausaalivaikutusten identifioimisen vertailemalla yksilöitä, jotka ovat ryhmäjaossa käytetyn muuttujan suhteen hyvin lähellä toisiaan juuri jakokynnyksen ala- ja yläpuolella (Hahn, Todd ja van der Klaauw 2001).

Oppilaansa valikoivan koulun valintaprosessi on yksi esimerkki regressioepäjatkuvuusasetelmasta. Opiskelijavalintaprosessissa hakijan sisäänpääsy määrittyy sen mukaan, ylittääkö hänen pääsykoepistemääränsä koulun asettaman sisäänpääsyrajan. Mikäli opiskelijan pääsykoepistemäärä on yhteydessä hänen myöhempiin testituloksiinsa, kouluun valitut ja

valitsematta jätetyt hakijat eivät ole vertailukelpoisia koulun vaikuttavuuden arviointia ajatellen. Koulun kausaalivaikutus myöhempiin testituloksiin voidaan kuitenkin identifioida vertailemalla hakijoita, joiden pääsykoepistemäärät ovat lähellä toisiaan juuri sisäänpääsyrajan ala- ja yläpuolella.

Tarkastelen väitöskirjassani kahta regressioepäjatkuvuusasetelmaan liittyvää menetelmällistä ongelmaa. Väitöskirjan kaksi ensimmäistä lukua käsittelevät regressioepäjatkuvuusasetelman avulla identifioitujen kausaalivaikutusten yleistettävyyttä. Ilman lisäoletuksia nämä kausaalivaikutukset pätevät ainoastaan yksilöille, jotka sijaitsevat juuri jakokynnyksellä. Usein olisi kuitenkin kiinnostava tietää, kuinka toimenpide vaikuttaa yksilöihin, jotka sijaitsevat kauempana jakokynnyksestä. Tämä on oleellinen kysymys muun muassa merkittävästi jakokynnykseen vaikuttavaa politiikkamuutosta suunniteltaessa. Oppilaansa valikoivan koulun tapauksessa esimerkiksi rotu- tai sosioekonomisten preferenssien käyttöönotto opiskelijavalinnoissa muuttaisi huomattavasti eri hakijaryhmien kohtaamia sisäänpääsyrajoja. Tämän politiikkamuutoksen vaikutusten arvi-

Kirjoitus perustuu Massachusetts Institute of Technologyssa 15.5.2014 hyväksytyyn väitöskirjaan ”Extrapolation and Bandwidth Choice in the Regression Discontinuity Design”. Väitöskirjan ohjaajina toimivat professorit Joshua Angrist and Parag Pathak. Miikka Rokkanen (mr3454@columbia.edu) toimii apulaisprofessorina Columbia Universityssä.

oimiseksi on välttämätöntä tietää koulun vaikutukset myös muille kuin juuri nykyisellä sisäänpääsyrajalla sijaitseville hakijoille.

Väitöskirjani ensimmäisessä luvussa kehitän havaitsemattomiin tekijöihin perustuvan ekstrapolaatiomenetelmän regressioepäjatkuvuusasetelmaa varten. Menetelmä perustuu ajatukselle, että ryhmäjaossa käytettävä muuttuja on epätäydellinen mittari jollekin havaitsemattomalle tekijälle, ja että tämä havaitsematon tekijä on ainoa syy ryhmäjaossa käytettävän muuttujan ja kiinnostuksen kohteena olevan vastemuuttujan väliselle yhteydelle. Tarkastellen, millaisten oletusten vallitessa tämä asetelma mahdollistaa kausaalivaikutusten epäparametrisen identifiointia kaikilla ryhmäjaossa käytettävän muuttujan jakaumassa, kun tutkijalla on käytössään myös vaihtoehtoisia epätäydellisiä mittareita. Hyödynnän menetelmässä kirjallisuutta ei-klassisista mittausvirhemalleista (Hu ja Schennach 2008) ja epäparametrisistä instrumenttimuuttujamalleista (Newey ja Powell 2003).

Sovellan kehittämäni menetelmää tutkimalla, kuinka bostonilaiset oppilaansa valikoivat julkiset koulut vaikuttavat koko seitsemännelle luokalle hakevan populaation koulumesteykseen sisäänpääsyrajalla olevien marginaalisten hakijoiden sijaan (Abdulkadiroglu, Angrist ja Pathak 2014). Tässä asetelmassa ajatuksena on, että hakijoiden pääsykoepistemäärät ovat epätäydellisiä mittareita heidän havaitsemattomalle akateemiselle kyvykkyydelleen, ja että hakijoiden myöhempi testimenestys on yhteydessä pääsykoepistemääriin ainoastaan akateemisen kyvykkyyden kautta. Tulosteni perusteella bostonilaisten valikoivien julkisten koulujen positiiviset vaikutukset testituloksiin keskittyvät heikommin menestyviin hakijoihin, jotka eivät tule tällä hetkellä hyväksytyksi näi-

hin kouluihin. Lisäksi hyödynnän estimoimaani mallia ennustaakseni rotu- ja sosioekonomisten preferenssien käyttöönoton potentiaalisia vaikutuksia. Simulointieni perusteella rotu- ja sosioekonomisten preferenssien käyttöönoton aiheuttama opiskelupaikkojen uudelleenjakosta nostaisi hakijoiden myöhempiä testituloksia.

Väitöskirjani toisessa luvussa, joka on kirjoitettu yhdessä Joshua Angristin kanssa, kehitämme havaittujen taustamuuttujien kontrollointiin perustuvan ekstrapolointimenetelmän. Tässä menetelmässä ajatuksena on, että ryhmäjaossa käytettävän muuttujan ja mielenkiinnon kohteena olevan vastemuuttujan välinen yhteys voidaan poistaa kontrolloimalla riittävästi havaittuja taustamuuttujia. Tässä mielessä kehittelemämme menetelmä vastaa ajatusta kaltaistamisesta (Heckman, Ichimura ja Todd 1998). Regressioepäjatkuvuusasetelman yhteydessä asettamillamme identifiointioletuksilla on kuitenkin perinteisestä kaltaistamisesta poiketen testattavia seurauksia, jotka mahdollistavat tulosten luotettavuuden paremman arvioimisen.

Väitöskirjani ensimmäisen luvun tapaan havainnollistamme kehittämämme menetelmän toimivuutta tutkimalla bostonilaisten valikoivien julkisten koulujen vaikutuksia hakijoihin, joiden pääsykoepistemäärät ovat koulujen käyttämien sisäänpääsyrajojen ala- ja yläpuolella. Asetelman ajatuksena on, että aiemman testimenestyksen ja demografisten tekijöiden kontrolloinnin jälkeen hakijoiden pääsykoepisteet eivät ole enää yhteydessä heidän myöhempiin testituloksiinsa. Seitsemännelle luokalle hakevien opiskelijoiden osalta havaitsemiemme taustamuuttujien kontrollointi ei ole riittävä poistamaan tätä yhteyttä. Kehittämämme menetelmä toimii kuitenkin hyvin yhdeksännelle luokalle hakeville opiskelijoille, joille havait-

semme kattavammin heidän aiemmat testituloksensa. Tulostemme perusteella bostonilaisen valikoivien julkisten koulujen positiiviset vaikutukset myöhempiin testituloksiin keskittyvät hakijoihin, jotka jäävät tällä hetkellä sisäänpääsyrajojen alapuolelle.

Väitöskirjani kolmannessa luvussa tarkastelen toista regressioepäjatkuvuusasetelmaan liittyvää menetelmällistä ongelmaa. Hyvin tyyppillinen tapa kausaalivaikutuksia estimoitaessa regressioepäjatkuvuusasetelmaa hyödyntäen on käyttää epäparametrisia menetelmiä, kuten paikallista polynomiregressiota (Fan ja Gijbels 1996). Epäparametristen menetelmien suosio perustuu niiden vähäisiin oletuksiin, mutta tyyppillisesti nämä menetelmät vaativat tutkijalta jonkin estimointiparametrin valintaa. Paikallisen polynomiregression tapauksessa tutkijan on valittava mallissa käytettävä estimointitajuus, joka karkeasti sanoen kertoo, kuinka kaukana ryhmäjakokynnystä olevia havaintoja tutkija käyttää estimoidessaan hoidon vaikutavuutta. Isomman estimointitajuuden käytön etuna on useamman havainnon hyödyntäminen, mikä vähentää estimaattorin otosvaihtelua. Toisaalta pienempää estimointitajuutta käytettäessä tutkija hyödyntää ainoastaan lähempänä ryhmäjakokynnystä olevia havaintoja, mikä vähentää estimaattorin harhaisuutta. Tässä luvussa kehitän uuden menetelmän näiden

kahden hyödyn ja haitan, estimaattorin tarkkuuden ja harhaisuuden, optimaaliseen tasapainottamiseen estimointitajuutta valittaessa. Lisäksi tarkastelen kehittämäni menetelmän ominaisuuksia sekä teoreettisesti että Monte Carlo -simulointien avulla. □

Kirjallisuus

- Abdulkadiroglu, A., Angrist, J. ja Pathak, P. (2014), “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools”, *Econometrica* 82: 137–196.
- Fan, J. ja Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman & Hall, London.
- Hahn, J., Todd, P. ja van der Klaauw, W. (2001), “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design”, *Econometrica* 69: 201–209.
- Heckman, J., Ichimura, H. ja Todd, P. (1998), “Matching as an Econometric Evaluation Estimator”, *Review of Economic Studies* 65: 261–294.
- Hu, Y. ja Schennach, S. (2008), “Instrumental Variable Treatment of Nonclassical Measurement Error Models”, *Econometrica* 76: 195–216.
- Newey, W. ja Powell, J. (2003), “Instrumental Variable Estimation of Nonparametric Models”, *Econometrica* 71: 1565–1578.