

---

TAMPEREEN YLIOPISTO  
Taloustiede Pro gradu -tutkielma

---

Tiina Kuuppelomäki

Selektiivisestä yliopistosta  
valmistumisen vaikutus alku-uran  
tuloihin teknillisellä alalla

---

Johtamiskorkeakoulu  
Taloustiede  
Huhtikuu 2016

---

Tampereen yliopisto  
Johtamiskorkeakoulu

Kuuppelomäki, Tiina: Selektiivisestä yliopistosta valmistumisen vaikutus alkuvuorantuloihin teknillisellä alalla.

Pro gradu -tutkielma: 50 sivua, 12 liitesivua.

Taloustiede

Huhtikuu 2016

Ohjaaja: Hannu Laurila

---

## Tiivistelmä

Koulutuksen laadun vaikutuksia yksilöiden työmarkkinatulemiin on tutkittu maailmalla paljon, mutta Suomessa aihealueen tutkimus on toistaiseksi ollut vähäistä. Tutkielma pyrkii paikkaamaan tätä aukkoa tutkimalla asiaa Suomen teknillisten yliopistojen saralla. Käytännössä tarkastellaan eroavatko selektiivisestä metropolialueen teknillisestä yliopistosta valmistuneiden alkuvuorantulo muista teknillisistä yliopistoista valmistuneiden tuloista.

Tutkimuskysymyksen tarkasteluun käytetään aineistoa, joka muodostuu vuosina 2000–2004 DIA-yhteishaussa hakeneista sekä heidän tulotiedoistaan vuoteen 2012 asti. Se sisältää yhteensä 223628 valintaryhmä-hakukohdevuosi-havaintoa. Käytössä oleva aineisto mahdollistaa tutkimuskysymyksen tarkastelun käyttäen fuzzy regressioepäjatkuvuusmenetelmää (FRD). Siinä hyödynnetään selektiivisen yliopiston sisäänpääsyrajoilla esiintyvää epäjatkuvuutta kyseisestä yliopistosta valmistumisen todennäköisyydessä.

Saatujen tulosten perusteella selektiivisestä yliopistosta valmistumisen tulovaikutukset ovat suuria (noin 20%) ja positiivisia, mutta eivät tilastollisesti merkitseviä. Verrattuna muihin pohjoismaisiin tutkimustuloksiin saatu premio on erittäin suuri, mutta toisaalta yhdenpitävä siinä suhteessa, ettei se ole tilastollisesti merkitsevä.

Erot muihin pohjoismaisiin tutkimustuloksiin saattavat selittyä sillä, että tutkielma on ensimmäisiä, jossa koulutuksen laadun vaikutuksia tutkitaan alan sisällä. Muita mahdollisia syitä ansiopremion suuruuteen on esimerkiksi se, että selektiivisestä teknillisestä yliopistosta valmistuneet päätyvät tulosten perusteella muuta todennäköisemmin asumaan Uudellemaalle, jossa tuottavuus ja myös tulot ovat muutenkin muuta maata korkeammat. Tulosten perusteella selektiivisestä teknillisestä yliopistosta valmistuneet myös työskentelevät sekä opintojensa aikana, että tulojen havaintohetkellä muista teknillisistä yliopistoista valmistuneita enemmän.

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Koulutuksen laadun määritelmät ja aiempi kirjallisuus</b>	<b>4</b>
2.1	Laadun määritelmät ja taustamekanismit . . . . .	4
2.2	Aiemmat tutkimustulokset ja menetelmät . . . . .	5
<b>3</b>	<b>DIA-valinta</b>	<b>10</b>
3.1	Historia . . . . .	10
3.2	Valintaprosessi 2000–2004 . . . . .	10
<b>4</b>	<b>Aineiston kuvailu ja rajaus</b>	<b>14</b>
4.1	Tutkimusaineisto ja otoksen rajaus . . . . .	14
4.2	Aineiston kuvaus ja OLS-regressio . . . . .	17
<b>5</b>	<b>Empiirinen menetelmä</b>	<b>21</b>
5.1	Sharp ja fuzzy RD . . . . .	21
5.2	Fuzzy RD . . . . .	23
<b>6</b>	<b>Tulokset</b>	<b>26</b>
6.1	Selektiivisestä yliopistosta valmistumisen vaikutus tuloihin . . . . .	26
6.2	Validius- ja robustisuustarkastelut . . . . .	28
6.3	Muita tuloksia . . . . .	36
<b>7</b>	<b>Lopuksi</b>	<b>41</b>
	<b>Viitteet</b>	<b>43</b>
	<b>Liitteet</b>	<b>48</b>

# 1 Johdanto

Yhteiskunnan näkökulmasta korkeakoulutuksen laadun tulovaikutusten ja vaikutuskanavien tutkiminen on tärkeää ja mielenkiintoista, etenkin viimeaikaisten koulutukseen kohdistuvien leikkausten valossa. Inhimillisen pääoman teorian mukaan koulutuksen laatu nimittäin vaikuttaa työntekijöiden kyvykkyyteen ja tuottavuuteen, jotka taas ovat yhteydessä yritysten tuottavuuteen ja siten talouskasvuun. Jos tällöin koulutuksen laadun havaitaan nostavan yksilöiden tuloja ja tämä selittyy edes osittain inhimillisen pääoman teorialla, voi koulutuksen laatuun investoiminen hyödyttää talouskasvua. Koulutuksen laadun heikentäminen sen sijaan saattaa heikentää talouskasvua.

Kysymys korkeakoulun tulovaikutuksista on mielenkiintoinen myös yksilön näkökulmasta. Yksilö joutuu vaikean kysymyksen eteen tehdessään päätöstä siitä, mille alalle ja mihin korkeakouluun hakeutuu. Vaikka Pohjoismaissa korkeakoulussa opiskelu onkin opiskelijoille lähes maksutonta, ei se kuitenkaan ole yksilölle kustannuksetonta. Kun huomioidaan korkeakouluopintojen aiheuttamat vaihtoehtoiskustannukset menetettyjen tulojen muodossa, voivat yksilötason kustannukset muodostua suuriksi. Tällöin on tärkeää, että nuori pystyy arvioimaan jo hakuvaiheessa mahdollisimman tarkkaan eri alojen ja eri korkeakoulujen vaikutuksen tulevaisuuden tuloihinsa. Tämän tutkielman eräs motivaatio onkin tarjota lisäinformaatiota korkeakoulujen välisistä tulovaikutuksista.

Vuosituhanne vaihteen jälkeen tutkijoiden parissa on herännyt kiinnostus siitä, mikä on koulutuksen laadun vaikutus yksilöiden työmarkkina- ja opintotulemiin. Etenkin Yhdysvalloissa ja Pohjoismaissa kysymystä on jo ehditty tutkia melko paljon, mutta Suomessa tutkimus on toistaiseksi ollut melko vähäistä. Yhdysvaltalais tutkimusten perusteella koulutuksen laadulla vaikuttaisi olevan positiivinen vaikutus myöhempään tuloihin (Behrman, Rosenzweig & Taubman 1996; Hoekstra 2009). Sen sijaan Pohjoismaissa tehdyt tutkimukset antavat tästä vähemmän todisteita (Borgen 2014; Öckert 2010).

Koulutuksen laadun tulovaikutuksia tutkitaan tässä tutkielmassa käyttäen Valtion taloudellisen tutkimuskeskuksen (VATT) hallussa olevaa diplomi-insinööri- ja arkkitehtikoulutuksen (DIA) yhteisvalinta-aineistoa vuosilta 2000–2004. Aineisto sisältää kaiken kaikkiaan 223628 valintaryhmä-hakukohdevuosi-havaintoa. DIA-aineistoon on yhdistetty Tilastokeskukselta saatu FLEED-aineisto vuosille 1999–2012 käyttäen salattuja henkilötunnuksia. Kyseiset aineistot mahdollistavat teknillisen yliopiston laadun vaikutusten tutkimisen yksilöiden alku-uran ansio- ja pääomatuloihin. Käytännössä verrataan selektiivisestä teknillisestä yliopistosta valmistuneiden tuloja muista teknillisistä yliopistoista valmistuneiden tuloihin.

Pääasiallisena tutkimusmenetelmänä käytetään Thistlethwaiten ja Campbellin (1960) esittämää regressioepäjatkuvuusmenetelmää (regression discontinuity, RD). Siinä verrataan selektiivisen teknillisen yliopiston sisäänpääsy-

rajan marginaalisesti ylittäneitä rajan marginaalisesti alittaneisiin.<sup>1</sup> Menetelmän vahvuutena on, että rajan ympäristössä olevat yksilöt voidaan olettaa keskimäärin samankaltaisiksi niin havaittujen kuin havaitsemattomienkin taustamuuttujien suhteen. Tällöin rajalla tapahtuva hyppäys selitettävässä muuttujassa johtuu ainoastaan rajan ylittämisestä ja sen aikaansaamasta käsittelystä. Jos kaikki rajan ylittäneet saavat jonkin käsittelyn, eikä kukaan rajan alittanut saa sitä, on käytössä RD:n muoto nimeltä sharp RD. Jos sen sijaan osa sisäänpääsyrajan ylittäneistä ei saa käsittelyä ja osa rajan alittaneista saa sen, käytetään RD:n muotoa nimeltä fuzzy RD.

Koska selektiivisistä teknillisistä yliopistosta valmistumisen todennäköisyydessä tapahtuu alle sadan prosentin hyppäys rajalla, on käytössä fuzzy RD. Se toteutetaan käytännössä instrumenttimuuttujamenetelmänä, jossa selektiivisistä teknillisistä yliopistosta valmistumista instrumentoidaan kyseisen yliopiston sisäänpääsyrajan ylittämisellä.<sup>2</sup>

RD:tä ja sen eri muotoja on käytetty maailmalla paljon ja sitä pidetäänkin satunnaistetun kenttäkokeen jälkeen varminpana tapana tarkastella koulutuksen laadun kausaali vaikutuksia (esimerkiksi Suhonen 2015, s.155). Suomessa RD-menetelmää ei kuitenkaan ole vielä juuri sovellettu koulutuksen laatuvaikutuksiin liittyvissä tutkimuksissa, Kannisen (2013) tutkimusta lukuun ottamatta. Sitä on kuitenkin käytetty esimerkiksi vaalitulosten vaikutusten selvittämisessä (Meriläinen 2013).

Menetelmä kuuluu niin kutsuttuihin selection-on-unobservables-lähestymistapoihin (Heckman & Robb 1985), joissa kouluun valikoitumisen oletetaan tapahtuvan paitsi havaittujen myös havaitsemattomien tekijöiden perusteella. Kyseiseen lähestymistapaan kuuluvia menetelmiä pidetään kausaali vaikutuksen estimoinnin kannalta luotettavampina kuin niin sanotun selection-on-observables-lähestymistavan menetelmiä. Tästä huolimatta suuri osa koulutuksen laadun vaikutuksia tarkastelevista tutkimuksista käyttää kyseiseen lähestymistapaan kuuluvia naiiveja regressiomalleja. Näitä menetelmiä pidetäänkin aiheeseen liittyvässä kirjallisuudessa benchmark-lähestymistapana, mistä johtuen tässä tutkielmassa raportoidaan myös pienimmän neliösumman (PNS/OLS) menetelmällä saadut tulokset.

Aiemmasta kirjallisuudesta tätä tutkielmaa lähimpänä ovat Abdulkadiroglu, Angrist ja Pathak (2014), Hoekstra (2009) sekä Kirkeboen, Leuven ja Mogstad (2015). Myös he ovat tutkimuksissaan käyttäneet RD:tä erilaisiin koulutusvalintoihin liittyviin analyyseihin.

Tämä tutkielma kuitenkin eroaa aiemmista tutkimuksista muutamain tavoin. Ensinnäkin tutkielma on eräs ensimmäisiä, jossa tutkitaan koulutuksen laadun vaikutuksia tietyn alan sisällä. Tällöin on mahdollista erottaa alavalinnan tulovaikutukset instituution tulovaikutuksista. Aiemmissa RD:tä

---

<sup>1</sup> Sisäänpääsyrajalla tarkoitetaan jotakin pisterajaa, jonka ylittäminen ainakin periaatteessa mahdollistaa opintojen aloittamisen kyseisessä korkeakoulussa.

<sup>2</sup>RD:stä ja sen muodoista tarjoavat lisätietoa esimerkiksi Angrist ja Pischke (2008, s.189–203).

käytävissä amerikkalaistutkimuksissa tämä ei useinkaan ole ollut mahdollista, sillä opiskelijat valitsevat siellä alan vasta opintojensa aikana. Toiseksi aiemmissa RD:tä hyödyntävissä tutkimuksissa on tyydytty lähes aina tutkimaan selektiivisessä korkeakoulussa aloittamisen vaikutuksia myöhempisiin työmarkkinatulemiin. Tässä tutkielmassa käytetty aineisto kuitenkin mahdollistaa valmistumisen vaikutusten tarkastelun aloittamisen sijaan. Myös aineiston laatu erottaa tämän tutkielman useimmista aikaisemmista tutkimuksista, sillä käytössä olevat tiedot ovat esimerkiksi useiden amerikkalais-tutkimusten käyttämiä tietoja tarkempia. Tämä johtuu siitä, että Tilastokeskukselta saatu FLEED-aineisto,<sup>3</sup> on muodostettu kyselyaineistojen sijaan verotietojen perusteella.

Saatujen tulosten perusteella selektiivisestä teknillisestä yliopistosta valmistumisella ei ole tilastollisesti merkitsevää tulovaikutusta 10% -riskitasolla. Tuloksissa on kuitenkin jonkin verran vaihtelua, mistä johtuen kausaalisuhteen olemassaoloa ei voida täysin poissulkea. Lisäksi saatu estimaatti on positiivinen ja erittäin suuri (20%), varsinkin verrattuna muista Pohjoismaissa saatuihin tuloksiin.

Estimoidun tulopreemion suuruutta selittää ainakin osittain se, että selektiivinen yliopisto sijaitsee metropolialueella ja sieltä valmistuneet asuvat tulosten perusteella muita todennäköisemmin Uudellamaalla. Tällä alueella palkat ovat muuta maata suuremmat ja työttömyysaste eräs Suomen alhaisimmista (EVA 2013; VATT-työryhmä 2015). Selektiivisestä yliopistosta valmistuneet myös työskentelevät tulosten perusteella opintojensa aikana sekä tulojen havaintovuonna muista teknillisistä yliopistoista valmistuneita enemmän. RD:llä saadut tulokset ovat myös lokaaleja, jolloin on mahdollista että sisäänpääsyrajan marginaalisesti ylittäneet hyötyvät laadukkaasta koulutuksesta keskimääräistä enemmän. Tätä tukee havainto, että saadut OLS-regressiotulokset (noin 9,7%) ovat RD-estimaatteja (jopa 20,3%) pienempiä.

Tutkielma rakentuu siten, että luvussa 2 esitellään koulutuksen laadun määritelmiä ja laatuvaikutusten taustamekanismeja. Luvussa käydään myös läpi aiheeseen liittyviä aikaisempia tutkimustuloksia ja käytettyjä menetelmiä. Luvussa 3 esitellään diplomi-insinööri- ja arkkitehtikoulutuksen (DIA) yhteisvalinnan taustaa sekä vuosien 2000–2004 valintamekanismeja. Luvussa 4 käydään läpi otoksen rajaukset ja kuvaillaan käytössä olevaa aineistoa. Luvussa 5 esitetään käytetty empirinen menetelmä. Luvussa 6 esitellään saadut tulokset ja tarkastellaan muita tuloihin vaikuttavia tekijöitä. Kyseisessä luvussa käydään lisäksi läpi RD:hen liittyvät robustisuus- ja validiustarkastelut. Luku 7 sisältää yhteenvedon sekä johtopäätelmät saaduista tuloksista. Liitteissä on lisätietoja DIA-valinnasta, analyysissä käytettyjen muuttujien kuvaukset sekä lisätaulukoita ja -kuvioita.

---

<sup>3</sup> FLEED-aineisto pitää sisällään muun muassa tulotiedot.

## 2 Koulutuksen laadun määritelmät ja aiempi kirjallisuus

### 2.1 Laadun määritelmät ja taustamekanismit

Korkeakoulun laadusta on olemassa useita määritelmiä. Yhdysvaltalaisista tutkimuksista suurin osa (esimerkiksi Hoekstra 2009; Dale ja Krueger 2002) käyttää kuitenkin laadun mittarina korkeakoulun selektiivisyyttä. Luultavasti tämä johtuu ainakin osittain siitä, että selektiivisyyden mittaaminen on helppoa esimerkiksi standardoitujen testien (PSAT) tai hyväksytyt/hakijat-suhteen avulla.

Tässä tutkielmassa keskitytään amerikkalaistutkimusten tapaan selektiivisyyteen laadun mittarina. Muitakin laadun mittareita on toki esitetty, ja selektiivisyydsmittareita kritisoitu siitä, että ne kuvaavat huonosti oppilaiden koulutuksesta saamaa hyötyä (Kuh & Pascarella 2004). Muita koulutuksen laadun mittareita ovat esimerkiksi Suhosen (2015, s.144) mainitsemat opetusresurssit tai korkeakoulun sijainti hyvällä työmarkkina-alueella.<sup>4</sup> Luvussa 4 tarkastellaankin selektiivisyyden lisäksi hieman myös muita koulutuksen laadun mittareita.

Taloustieteen kirjallisuuden parissa korkeakoulun laadun ja työmarkkinamenestyksen välistä syy-seuraussuhdetta on perinteisesti selitetty inhimillisen pääoman teorian (Schultz 1961; Becker 1962) sekä signaalointiteorian avulla (Spence 1973). Inhimillisen pääoman teoria selittää syy-seuraussuhdetta siten, että laadukkaaseen korkeakouluun meneminen kasvattaa yksilön tuottavuutta ja siten aikaansaa esimerkiksi opiskelujen jälkeisten tulojen kasvun. Laatuefektin voidaan tällöin ajatella olevan seurausta esimerkiksi paremmista resursseista tai kyvykkäämmistä koulukavereista.

Selektiivisyys yhdistetään usein korkeakoulun maineeseen, joten on mahdollista, että koulutuksen laadun tulovaikutus selittyy myös ainakin osittain signaalointiteorialla. Teorian mukaan korkeakoulun laadun ja työmarkkinamenestyksen syy-seuraussuhteen syntyyn riittää, että työnantajat kykenevät havainnoimaan jollain tapaa korkeakoulujen väliset laatuerot. Tämä havainnointi saattaa perustua esimerkiksi positiiviseen uutisointiin tai omiin kokemuksiin. Työnantaja tulkitsee tällöin nämä havainnot signaaleiksi työnhakijan tuottavuudesta. (Hershbein 2013).

Jos suurin osa syy-seuraussuhteesta selittyy inhimillisen pääoman teorialla, on koulutuksen korkeampaan laatuun investoiminen yhteiskunnan kannalta järkevää, sillä se kasvattaa työntekijöiden tuottavuutta ja luo siten talouskasvulle aiempaa paremmat edellytykset. Sen sijaan jos korkeakoulun laadun ja työmarkkinamenestyksen syy-seuraussuhde selittyy lähes kokonaan signaalointiteorian kautta, ei koulutuksen laadun parantaminen saa aikaan mainit-

---

<sup>4</sup>Tässä tutkielmassa korkeakoulun laatu määrittyy osittain myös sijainnin sekä opetusresurssien kautta.

tavia tuottavuushyötyjä yhteiskunnassa.

Käytännössä on vaikeata selvittää kuinka suuri osa syy-seuraussuhteesta selittyy signaalointiteorialla ja kuinka suuri osa inhimillisen pääoman teorialla. Erottelu on hankalaa muun muassa siksi, että ne vaikuttavat myös toisiinsa. Esimerkiksi opiskelijoiden tuottavuutta hyvin kasvattava korkeakoulu saa ajan myötä mainetta laadukkaana korkeakouluna, joka kouluttaa ammattitaitoisia työntekijöitä. Tällöin työmarkkinat alkavat tulkita kyseisessä korkeakoulussa opiskelun signaaliksi tuottavuudesta.

Taustamekanismien erottelun hankaluudesta huolimatta muutamat koulutuksen laadun tulovaikutuksia käsittelevät tutkimukset ottavat kantaa myös niiden takana vaikuttaviin mekanismeihin. Esimerkiksi Lang ja Siniver (2011) havaitsivat tutkimuksessaan, että huippuyliopistosta valmistumisen jälkeinen ansiopreemion nousu vähenee ajan kuluessa. Tätä he selittävät signaalointiteorialla siten, että signaalin vaikutus vähenee ajan kuluessa työnantajien oppiessa tuntemaan työntekijöiden todellisen kyvykkyyden. Myös Hershbeinin (2013) tekemän tutkimuksen mukaan koulutuksen laadun tulovaikutukset selittyvät lähinnä signaalointiteorialla siinä tapauksessa, että koulutuksen laatua mitataan selektiivisyydellä.

Sen sijaan Suhosen (2014) tekemässä tutkimuksessa havaittu opettaja/opiskelija -suhteen positiivinen vaikutus naisten työmarkkinamenestykseen on tuskin selitettävissä signaalointiteorialla (Suhonen 2015). Kyseisessä tutkimuksessa havaitaan negatiivinen suhde korkeakoulun opetusresurssien ja selektiivisyydsmittarien välillä, joista selektiivisyyden vaikutusta pidetään yleensä signaalointivaikutuksena. Myös Hämäläinen ja Uusitalo (2008) tutkivat taustamekanismeja hyödyntäen vuosina 1992–2000 Suomessa tapahtunutta koulutusreformia. Tuolloin osa ammattiopistotason kouluja muutettiin ammattikorkeakouluiksi pidentämällä niiden kestoa ja parantamalla niiden tasoja. Heidän saamiensa tulosten perusteella koulutuksen laatu vaikuttaa tuloihin sekä signaalointiteorian että inhimillisen pääoman teorian kautta.

Tässä tutkielmassa ei erotella missä määrin eri teorit selittävät saatua tulopreemiota. Aikaisempien tutkimusten perusteella vaikuttaa kuitenkin uskottavalta, että koulutuksen laatu vaikuttaa tuloihin sekä inhimillisen pääoman teorian että signaalointiteorian kautta.

## 2.2 Aiemmat tutkimustulokset ja menetelmät

Korkeakoulun laadun vaikutuksia työmarkkinamenestykseen on tutkittu paljon viimeisen kahdenkymmenen vuoden aikana. Olemassa oleva tutkimuskirjallisuus on pitkälti painottunut Yhdysvalloissa ja Pohjoismaissa tehtyihin tutkimuksiin (esimerkiksi Dale & Krueger 2002; Hoekstra 2009; Suhonen 2014; Öckert 2010). Muutamia aihetta käsitteleviä tutkimuksia on kuitenkin löydettävissä myös muualta maailmasta (esimerkiksi Saavedra 2009; Lang & Siniver 2011).

Suurin osa amerikkalaistutkimusten tuloksista puoltaa näkemystä, jonka



mukaan korkeakoulun laatu vaikuttaa opintojen jälkeisiin tuloihin (Behrman, Rosenzweig & Taubman 1996; Hoekstra 2009), joskin myös vastakkaisia tuloksia on saatu (Dale & Krueger 2002). Pohjoismaissa tehdyt tutkimukset sen sijaan löytävät amerikkalaistutkimuksia vähemmän todisteita siitä, että korkeakoulun laadulla olisi kovinkaan suurta positiivista vaikutusta työmarkkinamestykseen (Borgen 2014; Eliasson 2006; Suhonen 2013;2014; Öckert 2010). Yhtenä syynä tähän on todennäköisesti se, että Pohjoismaissa tulojen keskihajonta on pienempi kuin Yhdysvalloissa. Toisena mahdollisena syynä on, että Pohjoismaissa korkeakoulujen laatueroit maiden sisällä ovat vähäiset. Esimerkiksi yksityisiä korkeakouluja onkin Pohjoismaissa vain vähän.<sup>5</sup>

Saavedran (2009) Kolumbiassa tekemien tutkimusten perusteella maan parhaasta yliopistosta (Los Andes) valmistuminen kasvattaa valmistumisen jälkeisenä vuonna ansioita noin 35% ja työllistymisen todennäköisyyttä 16%. Yleisesti ottaen muualta maailmasta saadut tulokset kuitenkin antavat pohjoismaisten tutkimusten tavoin yhdysvaltalaisista tutkimuksista vähemmän todisteita korkeakoulun laadun tulovaikutuksista.

Korkeakoulujen laadun vaikutusten tutkimisessa käytetyt lähestymistavat vaihtelevat melko paljon. Nämä lähestymistavat voidaan jakaa karkeasti kahteen luokkaan (Suhonen 2015; Heckman & Robb 1985). Nämä luokat ovat valikoituminen havaittujen tekijöiden mukaan (selection-on-observables) sekä valikoituminen havaitsemattomien tekijöiden mukaan (selection-on-unobservables). Ensimmäisessä luokassa korkeakouluun valikoitumisen oletetaan tapahtuvan pelkästään havaittujen tekijöiden perusteella. Toisessa luokassa valikoitumisen oletetaan riippuvan myös havaitsemattomista tekijöistä.

Ensimmäiseen luokkaan kuuluvat niin kutsutut naiivit regressiomallit<sup>6</sup> sekä propensity score matching. Naiiveja regressiomalleja käytetään jossakin muodossa valtaosassa tutkimuskirjallisuutta (Black & Smith 2004; Eliasson 2006; Holmlund 2009; Lang & Siniver 2011; Monks 2000). Tämä suosio selittyy regressiomallien yksinkertaisuudella sekä niiden estimointitehokkuudella. Niitä pidetäänkin kirjallisuudessa benchmark-lähestymistapana.

Naiivien regressiomallien tapauksessa korkeakoulun laadun vaikutus yksilöiden työmarkkinatulemiin pyritään saamaan selville selittämällä työmarkkinamenestystä esimerkiksi korkeakoulun valinnalla ja lisäämällä kontrolleiksi havaittavia taustamuuttujia. Ongelmana ja huolena tällöin on, että harvoin havaitaan kaikkia korkeakoulun valinnan kanssa korreloivia taustateki-

---

<sup>5</sup> Esimerkiksi Ruotsissa on vain kolme yksityistä yliopistoa, joilla on oikeus myöntää kolmannen asteen korkeakoulututkinto ja vain yhdeksän yksityistä korkeakoulua, joilla on oikeus myöntää alempi korkeakoulututkinto (UKÄ 2014, s.14). Yksityisen korkeakoulun määritelmä on Pohjoismaissa myös hieman erilainen kuin esimerkiksi Amerikassa. Täällä yliopistot saavat usein valtaosan rahoituksestaan valtiolta, eivätkä esimerkiksi lukukausimaksuista.

<sup>6</sup> Tässä tutkielmassa käytetään Suhosen (2015) määritelmää naiiveista regressiomalleista. Siihen kuuluvat hänen määritelmänsä mukaan pienimmän neliösumman menetelmä (PNS/OLS), kvantiiliregressio, probit- ja logit-mallit sekä duraatiomallit.

jöitä. Tällöin tulos saattaa olla harhainen siksi, että kyseiset havaitsemattomat tekijät selittävät usein myös työmarkkinamenestystä. Toisin sanoen näitä regressiomalleja käyttämällä ei varmuudella kyetä sanomaan, johtuvatko suuremmat tulot koulutuksen korkeammasta laadusta vai olisivatko laadukkaampaan kouluun hyväksytyt muutoinkin menestyneet paremmin. Tutkittaessa koulutuksen laadun vaikutuksia naiivien regressiomallien avulla ongelmaksi saattaa muodostua myös se, ettei laadukkaassa korkeakoulussa opiskeleville kyetä havaitsemaan vähemmän laadukkaassa korkeakoulussa kyvykkyydeltään samankaltaista kontrolliryhmää. Tätä nimitetään kirjallisuudessa common support-ongelmaksi (esimerkiksi Angrist & Pischke 2008, s.57).

Propensity score matching-menetelmää pidetään naiivia regressiomallia läpinäkyvämpänä tapana tehdä havaittujen tekijöiden vakiointi (Black & Smith 2004; Eliasson 2006).<sup>7</sup> Jos taustaoletus pätee ja valikoituminen tapahtuu pelkästään havaittujen muuttujien perusteella, ovat tällä menetelmällä saadut tulokset naiivia regressiomallia robustimpia. Toisaalta jos taustaoletus ei päde, niin naiivia regressiomallia hieman monimutkaisempi propensity score matching ei välttämättä anna lisäarvoa analyysiin (Suhonen 2015, s.150).

Havaitsemattomien tekijöiden mukaan valikoitumisen luokkaan kuuluvat itsepaljastus- ja kaltaistetun hakijan mallit (self-revelation model ja matched applicant model), instrumenttimuuttujamallit, kaksosten- ja sisarusten kiinteitä vaikutuksia (fixed effects) hyödyntävät mallit sekä regressioepäjatkuvuusmenetelmä (RD). Tähän luokkaan kuuluvia malleja pidetään kirjallisuudessa selection-on-observables lähestymistapaa hyödyntäviä malleja parempina, sillä ne eivät oleta, että kaikki korkeakoulun valintaan vaikuttavat tekijät kyettäisiin havaitsemaan. Tällöin niiden avulla on mahdollista päästä suuremmalla varmuudella käsiksi koulutuksen laadun kausaalivaikutuksiin korrelaation sijasta.

Itsepaljastus- ja kaltaistetun hakijan mallit esittivät Dale ja Krueger (2002).<sup>8</sup> Menetelmällä saatuja tuloksia on vertailtu regressiomalleilla saatuihin tuloksiin (Dale & Krueger 2002;2011; Long 2008; Borgen 2014; Chen, Grove ja Hussey 2012). Saadut tulokset eivät ole olleet yksiselitteisiä. Dalen ja Kruegerin (2002;2011) saamat tulokset erosivat naiiveilla regressiomalleilla saaduista tuloksista, kun taas muut eivät ole tutkimuksissaan (esimerkiksi Borgen 2014; Chen ym. 2012) löytäneet kovinkaan suuria eroja näiden mallien tuottamien estimaattien välillä.

Kaksosten- ja sisarusten kiinteitä vaikutuksia (fixed effects) hyödyntä-

---

<sup>7</sup> Menetelmässä vertaillaan yksilöitä joilla odotettu korkeakouluun valikoitumistodennäköisyys on sama. Lisätietoa menetelmästä ja sen soveltamisesta tarjoavat esimerkiksi Murnane ja Willet (2008, s.310–328).

<sup>8</sup>Mallien ideana on vertailla joko laadultaan samanlaisiin korkeakouluihin hakeneita tai samanlaisten korkeakoulujen hyväksymiä ja hylkäämiä hakijoita (Suhonen 2015, s.151–152).

mällä on mahdollista välttyä ongelmilta, jotka aiheutuvat havaitsemattomista tekijöistä. Näiden menetelmien ongelmana on kuitenkin usein aineistojen pieni koko. Tämä selittää sen, ettei kovinkaan moni tutkimus ole hyödynnänyt niitä korkeakoulun laadun vaikutusten arvioinnissa. Näillä menetelmillä saadut estimaatit ovat pääsääntöisesti hiukan pienempiä kuin naiiveilla regressiomalleilla saadut tulokset (Behrman ym. 1996; Borgen 2014). Tämä viittaa siihen, että naiiveihin regressiomalleihin sisältyy kyvykkyysharhaa.

Myös instrumenttimuuttujamenetelmiä on käytetty jonkin verran korkeakoulun laadun vaikutusten tutkimisessa (Black & Smith 2006; Long 2008; Suhonen 2013;2014; Borgen 2014). Tämän menetelmän etuna on, että sen avulla voidaan korjata paitsi havaitsemattomista taustatekijöistä johtuvaa kyvykkyysharhaa niin myös mittausvirheistä aiheutuvaa harhaa. Instrumenttimuuttujamenetelmän ongelmana on kuitenkin validien ja relevanttien instrumenttien löytämisen vaikeus.

Regressioepäjatkuvuusmenetelmä on nousemassa suureen suosioon koulutuksen laadun vaikutusten tutkimuksessa (esimerkiksi Abdulkadiroglu ym. 2014; Hoekstra 2009; Saavedra 2009; Öckert 2010; Kanninen 2013). Suosiota selittää muun muassa menetelmän läpinäkyvyys ja luotettavuus. Siinä käytetään hyväksi laadukkaan koulun sisäänpääsyrajoilla esiintyviä epäjatkuvuuksia selitettävissä muuttujissa. Toisin sanoen siinä vertaillaan sisäänpääsyrajan läheisyydessä olevia yksilöitä, jotka voidaan olettaa samankaltaisiksi niin havaituilta kuin havaitsemattomiltakin ominaisuuksiltaan. Kirjallisuudessa RD:tä pidetäänkin satunnaistetun kenttäkokeen jälkeen varminpana tapana tarkastella koulutuksen laadun kausaalivaikutuksia (esimerkiksi Suhonen 2015, s.155). Ongelmana tässä menetelmässä on se, että sillä saadut tulokset ovat lokaaleja, eli koskevat vain sisäänpääsyrajalla olevia yksilöitä. Tuloksia ei voida näin ollen kovinkaan helposti yleistää koskemaan kauempana rajasta olevia yksilöitä. Sama ongelma koskee tosin myös esimerkiksi instrumenttimuuttujamenetelmiä.<sup>9</sup>

Myös RD:llä saatuja estimaatteja on verrattu naiiveihin regressioestimaatteihin. Esimerkiksi Öckert (2010) vertasi ruotsalaisaineiston avulla näillä menetelmillä saatuja estimaatteja toisiinsa. Tulosten perusteella estimaatit eivät eronneet tilastollisesti merkitsevästi toisistaan.

Suomessa koulutuksen laatuvaikutuksiin liittyvissä julkaistuissa tutkimuksissa RD:tä ei ole vielä juurikaan sovellettu. Poikkeuksena tästä on Kannisen (2013) tekemä tutkimus. Muutamia julkaisuja, joissa sitä sovelletaan koulutuksen alalla, on kuitenkin tulossa. Esimerkiksi Virtanen (2016) tutkii tulevassa väitöskirjassaan toisen asteen koulutuksen preferoidussa koulutusvaihtoehdossa aloittamisen vaikutuksia valmistumistodennäköisyyteen. Myös tämä tutkielma käyttää päätutkimusmenetelmänään RD:tä.

Suurin osa varsinkin RD:llä tehdyistä tutkimuksista keskittyy ladukkaa-

---

<sup>9</sup> RD:stä ja sen kahdesta muodosta/lähestymistavasta keskustellaan tarkemmin luvussa 5.

massa korkeakoulussa aloittamisen vaikutuksiin valmistumisen sijasta (Abdulkadiroglu ym. 2014, Hoekstra 2009). Syynä tähän lienevät usein puutteet aineistossa. Aloittamisen vaikutuksen tutkiminen on mielekästä ainakin tilanteessa, jossa se korreloi erittäin voimakkaasti valmistumisen kanssa. Muutamia RD:tä hyödyntäviä tutkimuksia, joissa tutkitaan valmistumisen vaikutuksia, on kuitenkin tehty. Näistä tunnetuimpana on Langin ja Siniverin (2011) tutkimus. Myös Saavedra (2009) sekä Kirkeboen ym. (2015) ovat tutkimuksissaan tutkineet valmistumisen vaikutuksia.

Tässä tutkielmassa käytetyllä aineistolla pystytään luotettavasti tutkiemaan alemman tai ylemmän korkeakoulututkinnon suorittamisen vaikutuksia tuloihin. Tämä on aloittamisen vaikutusten tutkimista järkevämpi lähestymistapa, sillä Suomessa teknillisissä yliopistoissa aloittavista moni vaihtaa alaa jo vuoden jälkeen. Tällöin mahdolliset erot tuloissa tuskin kuvastaisivat hyvin teknillisen korkeakoulutuksen laadun eroja. Lisäksi valmistumisen vaikutuksia tutkimalla sisältyy preemioon kaikille myös tutkinnon suorittamisesta aiheutuva signalointivaikutus eli niin sanottu sheepskin-vaikutus (Hungerford & Solon 1987).

Aikaisemmasta tutkimuskirjallisuudesta tätä tutkielmaa lähinnä ovat Abdulkadiroglun ym. (2011;2014), Hoekstran (2009) sekä Kirkoboenin ym. (2014) ja Saavedran (2009) tekemät tutkimukset. Abdulkadiroglu ym. (2011;2014) tutkivat Bostonin ja New Yorkin lukioissa vertaisryhmän vaikutuksia myöhempiin koetuloksiin käyttäen RD:tä. Yhteistä tämän tutkielman kanssa on esimerkiksi se, että hakijat valitsevat useasta koulusta, jotka eroavat lähinnä laadun suhteen. Hoekstran (2009) tutkimus taas on ensimmäisiä, jossa tutkitiin selektiivisen korkeakoulun vaikutuksia tuloihin käyttäen RD:tä. Hänen ei tutkimuksessaan kyetty havaitsemaan kaikkia hakuun liittyviä tekijöitä, mukaan lukien preferenssejä yli kaikkien koulutusvaihtoehtojen. Myöskään tässä tutkielmassa kaikkia hakijoiden preferenssejä ei tunneta.<sup>10</sup>

Tällä tutkielmalla on paljon yhtymäkohtia myös Kirkoboenin ym. (2014)<sup>11</sup> ja Saavedran (2009) tutkimuksiin. Yhteistä on muun muassa se, että vuosittain haettavana oli useita hakukohteita yliopistoa kohden. Lisäksi, kuten tässäkin tutkielmassa, näissä tutkimuksissa kiinnostuksen kohteena on valmistumisen vaikutus työmarkkinatulemiin.

---

<sup>10</sup> Vuosina 2000–2004 DIA-haku oli erillään muista korkeakouluhauista. Tutkimusaineiston perusteella saadaan näin ollen tietoa hakijoiden preferensseistä vain DIA-hakukohteisiin liittyen.

<sup>11</sup>Kirkeboenin ym. tutkimus eroaa muista tutkimuksista (myös tästä) siinä mielessä, että he eivät tutki korkeakoulun laadun vaikutuksia. Sen sijaan siinä tutkitaan kaikkien yhteishaussa hakeneiden preferenssejä ja verrataan preferoidulta koulutusosalta valmistumisen tulovaikutuksia hakijoiden seuraavaksi preferoidulta koulutusosalta valmistumiseen.

## 3 DIA-valinta

### 3.1 Historia

Tässä tutkielmassa pyritään estimoimaan selektiivisestä teknillisestä yliopistosta valmistumisen tulovaikutuksia käyttäen RD:tä. Vertailuryhmänä toimivat muista teknillisistä yliopistoista valmistuneet. Tästä johtuen on tärkeää tietää mitä teknillisiä yliopistoja Suomessa on, miten niihin haetaan, millainen valintaprosessi on ja miten sisäänpääsyrajat muodostuvat.

Teknillisiin yliopistoihin hakeminen tapahtuu DIA-valinnan, eli diplomi-insinööri- ja arkkitehtikoulutuksen yhteisvalinnan, kautta. DIA-valinta järjestettiin ensimmäistä kertaa vuonna 1974, jolloin mukana oli viisi erillistä teknillistä yliopistoa (DIA-valintaoppaat 2000–2004).<sup>12</sup> Nämä olivat Espoossa sijaitseva Teknillinen korkeakoulu, Tampereen teknillinen yliopisto, Lappeenrannan teknillinen yliopisto, Oulun yliopiston teknillinen tiedekunta sekä Åbo Akademin kemiallis-teknillinen tiedekunta. Näistä vanhin ja suurin oli Teknillinen korkeakoulu. Se toimi vuoden ajan nimellä Aalto-yliopiston teknillinen korkeakoulu, ennen kuin se vuonna 2011 jaettiin neljäksi tekniikan alan korkeakouluksi (Nykänen 2014). Näitä korkeakouluja ovat Insinöritieteiden korkeakoulu, Kemianteekniikan korkeakoulu, Perustieteiden korkeakoulu ja Sähkötekniikan korkeakoulu. Vuonna 2004 DIA-valintaan tuli mukaan myös Turun yliopiston matemaattis-luonnontieteellinen tiedekunta sekä Vaasan teknillinen tiedekunta.

Joissakin teknillisissä yliopistoissa tarjolla on vuodesta toiseen ollut vain muutamia koulutusohjelmia, kun taas joissakin tarjonta on ollut erittäin runsasta. Monipuolisin on ollut Helsingin Teknillinen korkeakoulu, jossa esimerkiksi vuonna 2000 oli tarjolla peräti 17 hakukohdetta. Åbo Akademiassa sen sijaan tarjolla oli vuonna 2000 vain kaksi hakukohdetta.

### 3.2 Valintaprosessi 2000–2004

DIA-valinnan valintaprosessi on hieman monimutkaisempi kuin useimmissa aikaisemmissa RD:tä hyödyntäneissä tutkimuksissa. Tästä johtuen valintaprosessi pyritään tässä luvussa käymään tarkasti läpi.

Vuosina 2000–2004 teknillisiin yliopistoihin haettiin yhdellä ja samalla lomakkeella. Hakijat listasivat lomakkeelle preferoimaansa järjestykseen maksimissaan viisi teknillistä hakukohdetta (toisin sanoen yliopisto-linja-yhdistelmää). Hakutoiveiden järjestys oli sitova. Haku ajoittui maaliskuulle ja tulokset julkaistiin heinäkuussa. Valituksi saattoi tulla hakuvuonna vain yhteen DIA-hakukohteeseen (DIA-valintaoppaat 2000-2004) eikä haku-

---

<sup>12</sup>Teknillisillä yliopistoilla viitataan tässä tutkielmassa myös yliopistoihin, joissa on teknillinen tiedekunta

kohteisiin voinut jonottaa.<sup>13</sup>

Vuosina 2000–2004 DIA-valinnassa tapahtui useita muutoksia, mutta pääpiirteet pysyivät samoina. Jokaiseen hakutoiveeseen saattoi esimerkiksi tulla valituksi useissa valintaryhmissä. Niin kutsutussa paperivalinnassa (valintaryhmä 1) valituiksi saattoivat tulla ainoastaan hakuvuonna ylioppilastutkinnon suorittaneet tai ruotsinkielisen ylioppilastutkinnon (RP) suorittaneet (joskin vasta vuodesta 2003 eteenpäin). Paperivalinta oli lisäksi rajattu koskemaan ainoastaan ensimmäistä (vuosina 2001–2004 myös toista) diplomi-insinöörihakukohdetta. Paperivalinnan ratkaisivat loppupisteet, jotka muodostuivat lisäpisteistä<sup>14</sup> sekä alkupisteistä. Alkupisteet muodostuivat ylioppilasarvosanojen sekä lukion päättötodistuksen fysiikan tai kemian arvosanojen perusteella.<sup>15</sup> Niiden muodostus on esitetty tarkemmin liitteen taulukoissa A.1–A.3.

Paperivalinnassa valittiin vuosina 2000–2004 maksimissaan 30% hakukohdekohtaisesta kiintiöstä. Jos hakukohteen paperivalinnan sisäänpääsyrajalla oli vähintään kaksi hakijaa samoissa pisteissä, tilanteen ratkaisivat heidän pitkän matematiikan ylioppilastutkinnon arvosanansa. Jokaisessa valintaryhmässä oli vuosina 2000–2004 myös yksi tai useampi kynnysehto, joka hakijan oli täytettävä, jotta hän voisi tulla valituksi. Paperivalinnassa tämän ehdon muodosti tarpeeksi korkea ylioppilastutkinnon pitkän matematiikan arvosana. Kyseisessä valintaryhmässä valituille postitettiin tieto valinnasta toukokuun lopulla ja mikäli hakija tuli valituksi paperivalinnassa, ei hän voinut enää osallistua valintakokeisiin. Toisin sanoen valittu hakija ei voinut valita missä valintaryhmässä hänet valittiin vaan se tehtiin automaattisesti. Kussakin valintaryhmässä sisäänpääsyrajan muodostivat viimeisen hyväksytyen loppupisteet.

Ne hakijat, jotka eivät tulleet valituiksi/eivät olleet hakukelpoisia paperivalinnassa, osallistuivat kaikkien hakukohteidensa mukaisiin valintakokeisiin.<sup>16</sup> Valintaryhmässä 2 huomioitiin nyt hakijan alkupisteistä, valinta-koepisteistä sekä lisäpisteistä koostuvat loppupisteet (maksimipistemäärä oli 66). Siinä saattoivat tulla valituiksi hakijat, jotka olivat suorittaneet ylioppilastutkinnon jonain hakua edeltävänä vuonna. Alkupisteet määräytyivät samalla tavalla kuin valintaryhmässä 1 ja niitä saattoi maksimissaan saa-

---

<sup>13</sup> Hakukelpoisia olivat kaikki, jotka olivat suorittaneet yhden seuraavista tutkinnoista: ylioppilastutkinto, RP-, IB-, EB-tutkinto tai jokin vähintään 3-vuotta kestänyt ammatillinen tutkinto. Hakukelpoisia olivat myös muut korkeakoulukelpoisuuden saavuttaneet hakijat. Åbo Akademia edellytti muiden ehtojen lisäksi, että hakija osoittaa kyvykkyytensä opiskella ruotsin kielellä. Arkkitehdeille oli erilliset, omat valintamenettelynsä. Niistä lisätietoa löytyy liitteestä A.

<sup>14</sup> Ensimmäisestä hakukohteesta sai 1–3 lisäpistettä jokaisessa valintaryhmässä, lukuunottamatta vuoden 2000 paperivalintaa.

<sup>15</sup> Vuonna 2000 katsottiin ainoastaan ylioppilastodistuksen matematiikan, fysiikan ja kemian arvosanoja.

<sup>16</sup> Näitä olivat yleensä matematiikan, fysiikan ja kemian sekä muutamissa hakukohteissa yhteiskuntatiedon koe.

da 23 pistettä. Valintakokeista annettiin 0–20 pistettä kustakin. Kokeesta saadut pisteet kerrottiin luvulla 5/9 (yhteiskuntatiedon koe luvulla 1/3) ja pyöristettiin kahden desimaalin tarkkuudella ennen kuin kahden kokeen valintakoepisteet summattiin yhteen. Tästä syystä hakijat, joiden valintakoepisteiden pyöristämätön summa oli sama, saattoivat saada erillaiset loppupisteet.<sup>17</sup> Tämä aiheutti loppupisteisiin vaihtelua myös sisäänpääsyräjälle ja saattoi määrittää sen, ketkä tulivat ylittivät sisäänpääsyräjän ja ketkä eivät. Vuosina 2001–2004 valintaryhmässä 2 valittiin 60% paperivalinnan jälkeisestä kiintiöstä, kun taas vuonna 2000 valittiin kyseisestä kiintiöstä jopa 80%.

Mikäli hakija ei tullut valituksi tai ei ollut hakukelpoinen valintaryhmissä 1 tai 2 oli hänellä vielä mahdollisuus tulla valituksi pelkkien valintakoepisteiden ja lisäpisteiden avulla valintaryhmässä 3. Tässä valintaryhmässä huomioitiin kahden valintakokeen tulokset ja ensisijaisuuspisteet (maksimipistemäärä oli 43). Kokeiden pisteet skaalattiin ja pyöristettiin jälleen samalla tavalla kuin valintaryhmässä 2. Vuonna 2000 tämä valintaryhmä oli jaoteltu kahtia. Tuolloin siinä valittiin 20% paperivalinnan jälkeisestä kiintiöstä. Valintaryhmässä 3a saattoivat tulla valituiksi vain ei-ylioppilaat. Heille oli varattu maksimissaan 15% kiintiö. Loput 20:stä prosentista oli varattu ylioppilaille, jotka eivät tulleet valituksi valintaryhmissä 1 tai 2 (3b). 3b kiintiö saattoi kuitenkin kasvaa jos kynnysehdot täyttyviä ei-ylioppilaita ei löytynyt tarpeeksi.<sup>18</sup> Vuosina 2001–2004 ei 3-valintaryhmää enää jaoteltu kahtia, vaan ei-ylioppilaat ja ylioppilaat olivat samassa kiintiössä.

Yhteenvetona yhteisvalintaprosessi eteni seuraavasti: valitsijat katsoivat kaikkia hakukohteita ja täyttivät paikat henkilöiden loppupisteiden ja listatun preferenssijärjestyksen mukaan. Ensin täytettiin valintaryhmän yksi kiintiö, sitten valintaryhmän kaksi kiintiö ja viimeisenä valintaryhmän kolme kiintiö. Jos hakija ylitti sisäänpääsyräjän preferenssilistansa mieluisammasa hakukohteessa (missä tahansa valintaryhmässä) ei hän voinut enää tulla valituksi listansa vähemmän mieluisiin kohteisiin. Poikkeuksen tästä muodosti tilanne, jossa hakijat ilmoittivat osallistuvansa paperivalintaan myös toisessa hakutoiveessaan ja tulivat siinä valituiksi. Tällöin he eivät menneet tekemään valintakokeita eivätkä niin ollen voineet tulla valituiksi ensimmäiseen hakukohteeseensa. Tällaisia hakijoita ei ole aineistossa kovinkaan paljoa

---

<sup>17</sup> Esimerkiksi, jos hakija sai matematiikan kokeesta 19 ja fysiikan kokeesta 11 pistettä, niin hänen lopulliset valintakoepisteensä olisivat olleet erikseen pyöristettyinä ja skaalattuina 10,56 ja 6,11, eli yhdessä 16,67 pistettä. Samaan aikaan, jos toinen hakija olisi saanut molemmista kokeista 15 pistettä, niin valintakoepisteet olisivat erikseen molemmista kokeista skaalauksen ja pyöristyksen jälkeen 8,33 ja yhdessä 16,66 pistettä. Toisin sanoen, vaikka molemmat hakijat saivat yhteensä 30 pistettä kokeista, olisivat heidän lopulliset valintakoepisteensä olleet erilaiset.

<sup>18</sup> Valintaryhmissä 2 ja 3 kynnysehtona oli, että hakija sai hakukohteen edellyttämistä kokeista vähintään 0,4 x kokeiden keskiarvojen summan (laskettiin koko maan tuloksista). Valintaryhmässä 3b hakijoiden oli lisäksi saatava valintakokeista 0,8 x edellisenä vuonna kyseiseen hakukohteeseen viimeisenä valintaryhmässä 3 hyväksytyyn ylioppilaan pistemäärä.

(noin 5,2% vuonna 2000–2004 ensikertaa hakeneista). Myös yhteisvalinnan ulkopuolelta oli mahdollista tulla valituksi (esimerkiksi erillisvalinnan kautta). Liitteestä löytyvät tiedot siitä, ketkä saattoivat hakea vuosina 2000–2004 yhteisvalinnan ulkopuolelta.



## 4 Aineiston kuvailu ja rajausta

### 4.1 Tutkimusaineisto ja otoksen rajausta

Tässä tutkielmassa käytetään hyväksi Valtion taloudellisen tutkimuskeskuksen (VATT) hallussa olevaa aineistoa diplomi-insinööri- ja arkkitehtikoulutuksen yhteisvalinnasta. Kyseinen aineisto kattaa koko populaation vuosina 2000–2004 DIA-valintaan osallistuneista hakijoista, käsittäen yhteensä 223628 havaintoa. Aineistoon kuuluvat muun muassa tiedot hakijoiden hakemista hakukohteista, niiden preferenssijärjestyksestä, saaduista alku-, lisä- ja valintakoepisteistä, loppupisteistä, rajoittavista ehdoista valintaryhmästä sekä tieto siitä, onko hakija tullut valituksi hakemaansa kohteeseen. Aineistosta löytyy myös muuttuja, jonka avulla sisäänpääsyrajalla tasapisteissä olevat hakijat on laitettu paremmuusjärjestykseen. Heillä tuo muuttuja ratkaisee siis sen, ketkä tulevat valituksi hakukohteeseen ja ketkä eivät. Aineisto sisältää myös muun muassa tiedot hakijoiden sukupuolesta, ylioppilasarvosanoista, äidinkielestä, kotikunnasta ja syntymävuodesta.

DIA-aineistoon yhdistetään Tilastokeskukselta (FLEED)-aineiston osia, jotka sisältävät tietoja hakijoiden työmarkkinamenestyksestä ja opintojen kuluista sekä vanhempien taustasta. Aineistot liitetään toisiinsa käyttäen salattuja henkilötunnuksia. FLEED-aineiston tiedot loppuvat vuoteen 2012, joten maksimaaliseksi tarkasteluperiodiksi saadaan kahdeksan vuotta.

Kuten on jo aiemmin mainittu, tutkielmassa käytetään RD:tä päämenetelmänä. Jotta sitä voitaisiin käyttää, rajataan aineistosta aluksi pois ne havainnot, joilta puuttuu valintatieto, tiedot tarvittavista loppupisteistä, tai joilla kynnysehto esti valituksi tulemisen hakukohteeseen. Ainoan poikkeuksen muodostaa vuonna 2000 valintaryhmä 3a, jossa kaikki rajan alle jääneet tulevat poistetuiksi sen takia, että tähän ryhmään kuuluvat eivät ole täyttäneet rajoittavia ehtoja. Tuossa valintaryhmässä päädytään pitämään kaikki havainnot, joilla loppupisteet olivat maksimissaan 5 pistettä rajan alla. Muutoin jouduttaisiin poistamaan kaikki rajan vuonna 2000 alittaneet ei-ylioppilaat, jolloin RD:tä ei voitaisi käyttää. Aineistosta poistuu rajauksen myötä 34% havainnoista.

Koska eri valintaryhmien loppupisteet ovat eri skaaloissa, luodaan ennen muita rajauksia jokaiselle valintaryhmä–hakukohde–hakuvuosi-yhdistelmälle erillinen, niin kutsuttu sijoitusmuuttuja. Tämän muuttujan luonnissa käytetään loppupisteitä siten, että suuremmilla loppupisteillä oleva hakija saa isomman sijoituksen. Lisäksi täsmälleen samoilla loppupisteillä (samassa valintaryhmässä, hakukohteessa ja samana hakuvuotena) olevat yksilöt saavat saman sijoituksen.<sup>19</sup> Tämä sijoitus siis kertoo sen, monennellako sijalla ha-

---

<sup>19</sup>Lukuunottamatta tilanteita, joissa sisäänpääsyrajalla on samoilla pisteillä useita yksilöitä. Tällöin käytetään sijoitus-muuttujan luomiseen myös loppupisteitä, jotka riippuvat insinöörikohteiden tapauksessa pitkän matematiikan arvosanasta ja arkkitehtikohteiden tapauksessa kaikkien kokeiden tuloksista (katso luku 3).

kija oli kyseisenä vuotena tietyssä hakukohteessa ja valintaryhmässä. Tämä muuttuja korvaa nyt siis loppupisteet ja sen avulla pystytään selvittämään riittivätkö hakijan pisteet tiettyyn hakukohteeseen vai eivät. Nyt kaikkien valintaryhmien ja hakukohteiden sisäänpääsyraajat määräävä muuttuja on siis yhteismitallinen. Vastaavaa tapaa on käytetty Abdulkadiroglun ym. (2014) tutkimuksessa, sillä erotuksella ettei tässä jaeta sijoitusmuuttujaa hakijoiden lukumäärällä.

Myöhempien tarkasteluiden helpottamiseksi keskitetään tämä sijoitusmuuttuja vielä siten, että se hakija, joka pienimmällä pisteillä ylittää sisäänpääsyrajan, saa sijoituksen nolla. Tämä onnistuu seuraavasti. Olkoon yksilön henkilökohtainen (sekä valintaryhmä-, hakukohde- ja hakuvuositahainen) sijoitus  $C_{ikvt}$  ja olkoon viimeisenä sisääntulleen sijoitus  $T_{kvt}$ , missä  $i$ =hakija,  $k$ =hakukohde,  $v$ =valintaryhmä ja  $t$ =hakuvuosi. Tällöin niillä hakijoilla, joilla  $C_{ikvt} \geq T_{kvt}$  pisteet riittivät hakukohteeseen sisäänpääsyyn tietyssä valintaryhmässä ensimmäisenä hakuvuonna. Vastaavasti niille, joilla  $C_{ikvt} < T_{kvt}$  pisteet eivät riittäneet hakukohteeseen sisäänpääsyyn. Tällöin yksilön hakuvuosi-, hakukohde- ja valintaryhmäkohtainen keskistetty sijoitusmuuttuja  $r_{ikvt}$  on

$$(4.1) \quad r_{ikvt} = C_{ikv} - T_{kvt}.$$

Jos  $r_{ikvt}$  on negatiivinen, eivät yksilön pisteet riittäneet hakukohteeseen valituksi tulemiseen, jos taas  $r_{ikvt}$  on epänegatiivinen, ovat yksilön pisteet ylittäneet hakukohteen sisäänpääsyrajan. Tämä muuttuja on nyt niin sanottu juokseva muuttuja (running variable, esiintyy kirjallisuudessa myös nimellä forcing variable).

RD:tä käytettäessä pelkona on, että osa hakijoista saattaa kyetä vaikuttamaan siihen ylittävätkö sisäänpääsyrajan vai eivät. Saattaisi olla esimerkiksi mahdollista, että osa hakijoista hakee, kunnes tulee valituksi. Kirjallisuudessa (RD:n yhteydessä) tätä kutsutaan taktikointiongelmaksi (gaming). Jotta tämä pelko pienentyisi, rajataan aineistosta pois kaikki hakijat, jotka ovat hakeneet aiemmin DIA-haussa tai joilla on aikaisempia korkeakouluopintoja. Tällöin poistuu noin 12% havainnoista. Valitettavasti ei ole mahdollista täydellä varmuudella sanoa, ovatko varsinkaan vuonna 2000 DIA-haussa mukana olevat hakijat olleet haussa mukana jo vuonna 1999 tai aiemmin. Tästä syystä poistetaan aineistosta myös hakijat, jotka aloittivat ensimmäisenä hakuvuotenaan kahdessa eri korkeakoulussa (noin 0,1% havainnoista).<sup>20</sup> Samasta syystä tehdään myös ikärajaus, jossa pudotetaan aineistosta pois kaikki yli 20-vuotiaat (noin 6% havainnoista). Näiden rajausten jälkeen aineisto koostuu ensimmäistä kertaa DIA-haussa mukana olleista nuorista.

<sup>20</sup> Vuosina 2000–2004 opiskelupaikan pystyi ottamaan vastaan vain yhdestä koulusta per vuosi. Tällöin hakijat, jotka aloittivat kahdessa koulussa samana vuonna, olivat luultavasti tulleet toiseen kouluun valituiksi jo aiempina vuosina ja saattaneet myös tuolloin olla myös teknillisessä haussa mukana.

Toisaalta taktikoinnin huoli on ikärajausten jälkeenkin varteenotettava, etenkin ensimmäisessä valintaryhmässä, jossa valinnat tehdään puhtaasti ylioppilasarvosanojen ja lisäpisteiden mukaan. Tällöin on mahdollista, että hakijat, tietoisina aiempien vuosien pisterajoista ja omista ylioppilasarvosanoistaan, taktikoivat hakupreferenssejään, jotta tulisivat varmasti valituiksi. Tästä johtuen pudotetaan valintaryhmä yksi ja kaikki siinä valituiksi tulleet pois otoksesta (noin 11% havainnoista). Valintaryhmän yksi poistaminen on järkevää myös siitä syystä, että suurin osa rajan alle jääneistä pääsi rajan yli muissa valintaryhmissä. Muissa valintaryhmissä taktikoinnin huoli on pienempi, sillä hakijoiden oli vaikea ennakoita miten he valintakokeissa menestyvät. Lisäksi kynnysehdot vaikeuttivat taktikointia. Taktikointiongelmia ja sen todennäköisyyttä tässä aineistossa käsitellään enemmän luvussa 6.2.

Tässä tutkielmassa tutkitaan selektiivisestä teknillisestä yliopistosta valmistumisen vaikutuksia pääoma- ja ansiotuloihin kahdeksan vuotta ensimmäisestä hakukerrasta, verrattuna muista teknillisistä yliopistoista valmistuneisiin. Tästä johtuen rajataan aineistosta pois myös kaikki ne yksilöt, jotka eivät valmistu hakua seuraavina kahdeksana vuotena mistään teknillisestä yliopistosta (54% havainnoista). Jotta RD:tä voidaan käyttää, pudotetaan aineistosta pois myös kaikki yksilöt, jotka eivät hakeneet selektiiviseen teknilliseen yliopistoon heidän ensimmäisenä havaittuna hakuvuotenaan (18% havainnoista).

Aineistosta pudotetaan pois myös kaikki arkkitehtikohteisiin valituksi tulleet, niistä valmistuneet sekä hakijoiden arkkitehtikohteet (kaikkiaan noin 3% havainnoista putoaa pois). Näihin rajauksiin on useita syitä. Ensinnäkin valintamenettely oli niissä hyvin erilainen kuin muissa DIA-valinnan hakukohteissa. Toiseksi, arkkitehtikohteista valmistuminen vie kauemmin kuin tavallisista diplomi-insinöörikohteista valmistuminen.<sup>21</sup> Kun rajoitutaan tarkastelemaan ensimmäisestä hakukerrasta kahdeksassa vuodessa valmistuneiden tuloja, on tällöin mahdollista, että otos olisi ilman näitä pudotuksia valikoitunut ja tulokset harhaisia.

Näiden rajausten jälkeen valitaan jokaiselle hakijalle se selektiivisen yliopiston valintaryhmä–hakukohde-yhdistelmä, jossa hakija sai suurimman keskitetyn sijoituksensa. Hakijoiden muut valintaryhmät ja hakukohteet pudotetaan aineistosta (88% havainnoista). Tällöin jäljelle jää hakijan helpoin selektiivisen yliopiston valintaryhmä–hakukohde-yhdistelmä. Virtanen (2016) käyttää samantyyppistä menetelmää.<sup>22</sup>

Koska tutkielmassa käytetään pääoma- ja ansiotulojen logaritmeja, raja-

---

<sup>21</sup> Vuosina 2008–2012 arkkitehdiksi valmistumisen mediaani oli vuodesta riippuen 8–10 vuotta, kun diplomi-insinööreillä vastaava luku oli 6,5–8 vuotta (Tilastokeskus, a 2013).

<sup>22</sup> Vaihtoehtona olisi tehdä kuten Abdulkadiroglu (2014) ja käyttää kaikkia hakijan selektiivisiä hakukohteita, määrittää niiden interaktio sisäänpääsyrajan ylittämisen kanssa ja klusteroida keskivirheet opiskelijatasolla. Johtuen hakukohteiden suuresta määrästä ja Virtasen lähestymistavalla saatujen tulosten helposta tulkinnasta, käytetään tässä tutkielmassa Virtasen lähestymistapaa.

taan aineistosta pois myös kaikki yksilöt, joille logaritmoitua tuloja ei havaita kahdeksan vuoden kuluttua ensimmäisestä hakukerrasta (2% havainnoista). Aineistossa on myös muutamia omituisuuksia, kuten henkilöitä jotka ovat kirjoittaneet ylioppilaiksi, mutta joille äidinkielen ylioppilasarvosanaa ei havaita (0,03%). Kyseiset hakijat poistetaan aineistosta. Joinakin vuosina joissakin koulutusohjelma–valintaryhmä-yhdistelmässä on se tilanne, että kaikki hakijat ovat tulleet valituiksi. On myös muutamia hakukohde–valintaryhmä-yhdistelmiä, joissa ei ole valittu ketään tietynä vuotena. Regressioepäjatkuvuusmenetelmässä on ideana vertailla yksilöitä sisäänpääsyrajan molemmin puolin, joten molemmissa edellä mainituissa tapauksissa RD ei ole käyttökelpoinen. Tästä syystä poistetaan myös nämä havainnot aineistosta (0,4% havainnoista). Lopullisessa aineistossa on 2875 havaintoa. Sijoitusmuuttujan (juoksevan muuttujan) minimiarvo on  $-458$  ja maksimiarvo on  $485$ .

## 4.2 Aineiston kuvaus ja OLS-regressio

Metropolialueen teknillinen yliopisto on useiden mittareiden näkökulmasta muita laadukkaampi. Taulukosta 1 nähdään vuosille 2000–2004 keskimääräiset minimipisteet (eli sisäänpääsyrajat) eri valintaryhmissä alueittain. Taulukon perusteella metropolialueen teknillinen yliopisto on huomattavasti muita selektiivisempi, sillä sen sisäänpääsyrajan ylitys vaatii keskimäärin muita suuremmat loppupisteet jokaisessa valintaryhmässä. Se on siis keskimäärin muita laadukkaampi tämän määritelmän mukaan.

Taulukko 1: Teknillisten yliopistojen selektiivisyys eri alueilla ja eri valintaryhmissä, mitattuna hakukohteiden keskimääräisillä minimipisteillä,

Yliopisto	min vry1	min vry2	min vry3
Metropolialue	28,408	35,102	17,804
Tampere	21,602	30,803	15,591
Lappeenranta	15,565	19,497	8,520
Oulu	16,517	21,524	8,962
Åbo	17,217	19,620	4,391
Turku	15,125	18,485	4,288
Vaasa	14,675	11,769	13,652

Taulukossa 2 on esitetty Vipunen–tietokannasta saatujen tietojen perusteella tehty taulukko, joka kertoo keskimääräisen opettaja/opiskelija-suhteen eri alueiden teknillisissä yliopistoissa vuosina 2010–2014.<sup>23</sup> Taulukosta näh-

<sup>23</sup> Vipusessa opettajien määrät löytyvät vasta vuodesta 2010 alkaen (Vipunen, Yliopistojen opetus- ja tutkimushenkilökunta). Opiskelijat on laskettu käyttäen alemman ja ylempään korkeakouluopiskelijat (Vipunen, Yliopistojen opiskelijat).

dään, että metropolialueen teknillisissä yliopistossa on toiseksi paras opettaja/opiskelija-suhde. Se häviää vain Åbo Akademin kemiallis-teknilliselle tiedekunnalle.<sup>24</sup> Toisin sanoen myös tämän laadun mittarin mukaan metropolialueen teknillinen yliopisto on laadukkaimpia teknillisiä yliopistoja Suomessa. Myös sijaintia metropolialueella voidaan pitää työmarkkinoiden kannalta merkinä laadukkaasta koulutuksesta (Suhonen 2015, s.144).

Taulukko 2: Opettajat/oppilaat-suhde teknillisissä yliopistoissa

Teknillinen yliopisto	Opettajat/oppilaat
Metropolialue	0,217
Tampere	0,14
Lappeenranta	0,15
Oulu	0,167
Åbo	0,391
Turku	0,096
Vaasa	0,104

Liitteen taulukossa C.1 on esitetty muutamien otoksen muuttujien keskiarvoja ja suhteita. Taulukosta huomataan, että lähes kaikki selektiiviseen teknilliseen yliopistoon hakeneet ovat suomenkielisiä (93%) ja ylioppilaita (99%). Huomataan myös, että hakijoista yli 75% on miehiä. Tämä ei sinällään ole yllättävää, sillä teknilliset alat ovat perinteisesti hyvin miesvaltaisia. Taulukosta havaitaan myös, että metropolialueen teknilliseen yliopistoon hakeneista noin 25% asuu hakuhetkellä pääkaupunkiseudulla. Lisäksi hakijoiden alku-uran keskimääräiset vuositulot ovat taulukon perusteella lähes 38000 euroa. Taulukosta nähdään myös, että selektiiviseen teknilliseen yliopistoon hakeneiden keskimääräiset pitkän matematiikan ja äidinkielen ylioppilasarvosanat ovat selvästi maan mediaanin (3) yläpuolella (Ylioppilas-tutkintolautakunta, tietoja tutkinnosta).

Taulukossa C.2 on esitetty samojen muuttujien keskiarvoja ja suhteita jaoteltuna niihin, jotka ylittivät selektiivisen yliopiston sisäänkäyntirajan ja niihin, jotka jäivät kyseisen rajan alle. Taulukon perusteella nähdään, että sisäänkäyntirajan ylittäneiden pitkän matematiikan ylioppilasarvosana on keskimäärin jopa 4,5 (eli M:n ja E:n välillä). Rajan alittaneilla se taas on keskimäärin 3,5. Taulukon perusteella sisäänkäyntirajan ylittäneistä keskimäärin noin 27% asuu hakuhetkellä pääkaupunkiseudulla, kun taas rajan alittaneista siellä asuu vain noin 13%.

<sup>24</sup>Opettajien määrän kasvaessa suhteessa opiskelijoiden määrään voidaan ajatella, että opettajat kykenevät paremmin keskittymään yksittäisiin opiskelijoihin, jolloin oppiminen saattaa helpottua. Luokkakoon ollessa pieni opettajan on myös helpompaa räätälöidä kurssi paremmin opiskelijoiden tarpeisiin ja osaamiseen sopivaksi.

Kirjallisuudessa naiiveja regressiomalleja pidetään bechmark-lähestymistapana, mistä johtuen myös tässä esitetään OLS-regressiolla saadut tulokset. Koska DIA-valinnassa mukana olevat hakijat ovat valmiiksi melko samanlaisia taustatekijöiltään, saattaa OLS tässä tuottaa tavallista harhattomampia tuloksia. Formaalisti käytetty OLS-regressio-lauseke on seuraavanlainen:

$$(4.2) \quad Y_i = \alpha_0 + \alpha_1 * V_i + \alpha_2 * U_i + \alpha_3 * H_i + \sum_j \alpha_j * X_{ij} + \sum_k \alpha_k * Z_{ik} + \varepsilon_i,$$

missä  $\alpha$ :t ovat regressiokertoimia ja  $Y_i$  on henkilön  $i$  vuositason pääoma- ja ansiotulojen luonnollinen logaritmi mitattuna kahdeksan vuotta ensimmäisestä hakukerrasta. Tulot on logaritmoinnin lisäksi korjattu inflaation suhteen käyttäen rahanarvonkertoimia (Tilastokeskus, b 2013). Perusvuotena toimii vuosi 2012.  $V_i$  on dummy-muuttuja, joka saa arvon yksi, jos hakija on valmistunut selektiivisestä teknillisestä yliopistosta ja arvon nolla muulloin.  $U_i$  kertoo valmistumisvuoden ja  $H_i$  tutkinnon, josta hakija valmistui.

Joihinkin tarkasteluihin on lisätty myös muita kontrolleja. Yhtälössä 4.2  $X_{ij}$  kuvaa tätä muuttujajoukkoa. Siihen kuuluvat hakuvuosi, sukupuoli, kieli, ikä, iän neliö sekä äidinkielen ja matematiikan ylioppilasarvosanat, tieto siitä onko hakija kotoisin pääkaupunkiseudulta, vanhempien tulot, isän ja äidin tuoreimmat tutkinnot sekä heidän päätoimensa, yhtälössä 4.2 muuttuja  $X_{ij}$ .<sup>25</sup> Lisäksi viimeisiin tarkasteluihin, on myös lisätty muuttujajoukko  $Z_{ik}$ , joka sisältää hakijoiden hakupreferenssit. Yhtälössä 4.2 esiintyvä  $\varepsilon_i$  on virhetermi.

OLS-regressiomenetelmällä saadut tulokset on esitetty taulukossa 3. Taulukon ensimmäisessä sarakkeessa on estimoitu selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus tuloihin ilman mitään kontrolleja. Tällöin valmistumisen vaikutus tuloihin on pieni, eikä se ole tilastollisesti merkitsevä. Toisessa sarakkeessa on kontrolloitu valmistumisvuosi sekä valmistumiskoodi. Tällöin estimaatit kasvavat noin 7,8%:iin ja tulevat myös tilastollisesti merkitseviksi. Kolmannessa sarakkeessa lisätään muut peruskontrollit (X funktiossa 4.2), tulokset eivät tällöin juurikaan muutu.

Viimeisessä sarakkeessa kontrolloidaan vielä hakijoiden preferenssejä yliopistotasolla (7 teknillistä yliopistoa), tällöin tulokset (samoin kuin keskivirheet) kasvavat noin 9,7%:iin. OLS-regressiolla saatujen tulosten perusteella vaikuttaa siis siltä, että selektiivisestä teknillisestä yliopistosta valmistuminen todellakin vaikuttaa tuloihin ja tuo tulopremio olisi jopa 9,7 prosenttia.

Saatuihin regressiotuloksiin on kaikesta huolimatta syytä suhtautua varauksella, sillä on edelleen mahdollista että ne pitävät sisällään puuttuvien muuttujien harhaa. Tulokset saattavat myös olla harhaisia johtuen common support-ongelmasta.

---

<sup>25</sup> Analyseissä käytetyt muuttujat on kuvattu tarkemmin liitteessä B.

Taulukko 3: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin käyttäen regressiomenetelmää.

	ln(tulot)			
Valmistui selektiivisestä yliopistosta	0,0331 (0,0324)	0,0777** (0,0359)	0,0770** (0,0386)	0,0966* (0,0581)
N	2875	2875	2875	2875
Kontrollit		kyllä	kyllä	kyllä
Lisäkontrollit			kyllä	kyllä
Preferenssit				kyllä

*Robustit keskivirheet suluisissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .*

*Kontrollit: valmistumisvuosi ja -koodi. Lisäkontrollit: sukupuoli, kieli, ikä, äidinkielen ja matematiikan arvosanat, onko hakija kotoisin pääkaupunkiseudulta, vanhempien tulot, -tuoreimmat tutkinnot sekä heidän päätoimensa.*

## 5 Empiirinen menetelmä

### 5.1 Sharp ja fuzzy RD

OLS-regressiota koskevat ongelmat voidaan onneksi välttää käyttämällä hyväksi regressioepäjatkuvuusmenetelmää (RD). RD:ssä oletetaan, että ihmiset, jotka ovat marginaalisesti sisäänpääsyrajan alla, ovat niin havaituilta, kuin havaitsemattomilta ominaisuuksiltaan samanlaisia kuin ihmiset, jotka ovat marginaalisesti tuon sisäänpääsyrajan yläpuolella. Toisin sanoen RD:n sisäinen validiteetti on vahva. Tällöin taustaoletusten pätiessä voidaan olettaa, että saadut tulokset kuvastavat kausaalisuutta eivätkä pelkästään korrelaatiota, toisin kuin tavallista OLS-regressiomallia käytettäessä. Taustaoletuksista ja niiden testaamisesta keskustellaan enemmän luvussa 6.

RD:n käytön kannalta olennaista on, että kyetään määrittämään niin kutsuttu juokseva muuttuja. Tämä juokseva muuttuja määrää ainakin pääasiallisesti sen, onko yksilö toimenpiteen kohteena vai ei. Tässä tutkielmassa toimenpiteenä on tutkinnon suorittaminen (valmistuminen) metropolialueen selektiivisestä teknillisestä yliopistosta ja juoksevana muuttujana toimii luvussa 4.1 kuvailtu keskitetty sijoitusmuuttuja. Ideana siis on, että juoksevala muuttujalla on jokin arvo (sisäänpääsyraja), jonka alapuolelle jääneet eivät todennäköisesti suorita tutkintoa selektiivisestä teknillisestä yliopistosta, mutta jonka yläpuolella olevilla on suuri todennäköisyys valmistua kyseisestä koulusta.

Jos rajalla tapahtuu sadan prosentin hyppäys esimerkiksi tutkinnon suorittamisen todennäköisyydessä, niin puhutaan sharp RD:stä (SRD). Tämä tarkoittaa käytännössä, että kaikki juoksevan muuttujan rajan marginaalisesti ylittäneet suorittivat tutkinnon eikä kukaan kyseisen rajan marginaalisesti alittaneista suorittanut sitä.<sup>26</sup> Tällöin rajalla tapahtuva hyppäys selitettävässä muuttujassa voi johtua ainoastaan tutkinnon suorittamisesta, sillä sisäänpääsyrajan ympäristössä yksilöt voidaan olettaa samanlaisiksi ominaisuuksiltaan. Liitteen kuvassa C.1 on esitetty alku-uran logaritmoidut tulot selektiivisen teknillisen yliopiston sisäänpääsyrajan tuntumassa ja sen alla olevassa taulukossa C.3 optimaalisella havaintovälillä saatu estimaatti. Toisin sanoen nämä kuvastavat SRD:llä saatuja tuloksia, jossa tuloja selitettäisiin rajan ylittämällä.<sup>27</sup> Kuvan ja taulukon perusteella sisäänpääsyrajan ylittämällä ei ole 10%-riskitasolla tilastollisesti merkitsevää vaikutusta alku-uran tuloihin. Estimaatti on kuitenkin positiivinen ja melko iso (noin 10%).

On kuitenkin olemassa myös toinen RD:tä hieman eri tavalla hyödyntävä tutkimusasetelma/muoto, jota kutsutaan fuzzy RD:ksi (FRD). Se eroaa SRD:stä siten, että juoksevan muuttujan rajalla tapahtuva hyppäys ei ole

---

<sup>26</sup> Yleensä SRD:tä käytetään esimerkiksi korkeakouluun valituksi tuleminen vaikutusten estimoinnissa.

<sup>27</sup> Käytetty funktio vastaa funktiota 5.3, sillä erotuksella, että valmistumisen sijaan selitettävänä muuttujana ovat logaritmoidut tulot.



nyt sata prosenttia, vaan vähemmän (toisin sanoen vaikutus on ”fuzzy”). On esimerkiksi mahdollista, että sisäänpääsyrajan ylittäneet aloittavat opinnot, mutta jättävät ne kesken. On myös mahdollista, että jotkut havaittuna vuonna rajan alle jääneet pyrkivät myöhemmin vuosina kouluun uudelleen ja pääsevät silloin rajan yli ja valmistuvat selektiivisestä yliopistosta. Tällöin sisäänpääsyrajalla tapahtuva hyppäys valmistumistodennäköisyydessä on alle sadan prosentin eikä SRD ei tule kysymykseen. Sen sijaan FRD mahdollistaa vaikutuksen estimoimisen.

FRD perustuu siihen, että sisäänpääsyrajan kohdalla tapahtuva epäjatkuvuus rajan ylittämisessä aiheuttaa eksogeenista vaihtelua selektiivisestä yliopistosta valmistumiseen. Käytännössä FRD suoritetaan kaksivaiheisena pienimmän neliösumman estimointina (2SLS), jossa tutkinnon suorittamista instrumentoidaan rajan ylittämisellä. Malleissa kontrolloidaan myös keskistetyin juoksevan muuttujan suoraa vaikutusta endogeeniseen muuttujaan (valmistumiseen). Regressioepäjatkuvuusmenetelmistä löytyy tarkemmin tietoa esimerkiksi Angristin ja Pischken kirjasta ”Mostly harmless econometrics” (2008, s.189–203).

FRD:n avulla saadut tulokset ovat LATE -estimaatteja (local average treatment effect). Eli ne pätevät vain lokaalisti sisäänpääsyrajan ympäristössä. Tämä tarkoittaa myös sitä, että selektiivisestä teknillisestä yliopistosta valmistumisen tulovaikutus pystytään selvittämään niille, joilla sisäänpääsyrajan ylittäminen vaikutti valmistumiseen. Kirjallisuudessa heitä kutsutaan nimellä ”compliers”. Sen sijaan saatu tulos ei päde niille ihmisille, jotka olisivat joka tapauksessa joskus valmistuneet selektiivisestä teknillisestä yliopistosta (riippumatta siitä ylittivätkö rajan vai eivät). Tulos ei myöskään päde niille jotka eivät valmistuisi noista hakukohteista riippumatta siitä ylittivätkö rajan vai eivät. Näitä ryhmiä kutsutaan kirjallisuudessa nimillä ”always-takers” ja ”never-takers”. (Angrist & Pischke 2008, s.112–119).

FRD:tä käytettäessä on myös tehtävä seuraavat oletukset (Angrist & Pischke 2008, s.112–114). Ensinnäkin instrumentin (sisäänpääsyrajan ylittäminen) tulee olla riippumaton tuloista ja valmistumisesta. Toiseksi oletetaan, että instrumentti vaikuttaa tuloihin vain valmistumisen kautta. Kolmantena oletuksena on, että selektiivisestä yliopistosta valmistumisen todennäköisyydessä tapahtuu hyppäys sisäänpääsyrajalla (ns. First-stage oletus/ehto). Viimeisenä oletetaan, että rajan ylittäminen ei kasvata todennäköisyyttä olla valmistumatta selektiivisestä yliopistosta. Tätä sanotaan kirjallisuudessa monotonisuusehdoksi/oletukseksi, ja se on yllä olevista ainoa, joka ei välttämättä toteudu.

On esimerkiksi mahdollista, että jotkut hakijat preferoivat jotakin muuta DIA-hakukohdetta verrattuna metropolialueen hakukohteeseen. Tällöin vaikka kyseiset hakijat ylittivätkin selektiivisen yliopiston sisäänpääsyrajan, saattoivat he tulla valituiksi johonkin muuhun teknilliseen yliopistoon ja valmistuivat sieltä. Jos kyseiset hakijat olisivat jääneet rajan alle, on mahdollista, etteivät he olisi tuona vuonna tulleet valituiksi mihinkään kouluun.

Tällöin he olisivat todennäköisesti hakeneet seuraavana vuonna uudelleen ja kenties silloin tulleet valituiksi selektiiviseen teknilliseen yliopistoon ja valmistuneet sieltä. Rajan ylittäminen on saattanut heidän kohdallaan tällöin pienentää todennäköisyyttä valmistua selektiivisestä teknillisestä yliopistosta.

Monotonisuusoletuksen rikkoontumiselta on mahdollista välttyä rajaamalla aineistosta pois ne hakijat, joilla metropolialueen hakukohteet eivät olleet etusijalla verrattuna muihin hakukohteisiin. Aineiston koko kuitenkin pienentyisi tällöin huomattavasti, eikä RD:n käyttö olisi suositeltavaa. Tästä johtuen kyseistä rajausta ei tässä tutkielmassa tehdä. Lisäksi tulokset pätevät muutenkin vain niille, joilla sisäänpääsyrajan ylittäminen vaikutti valmistumiseen (compliers). Tällöin ei oletuksen mahdollinen rikkoontuminen todennäköisesti aiheuta ongelmaa. Monotonisuusoletus rikkoontuu myös useissa aikaisemmissa tutkimuksissa (esimerkiksi Hoekstra 2009; Saavedra 2009; Kirkeboen ym. 2015).

## 5.2 Fuzzy RD

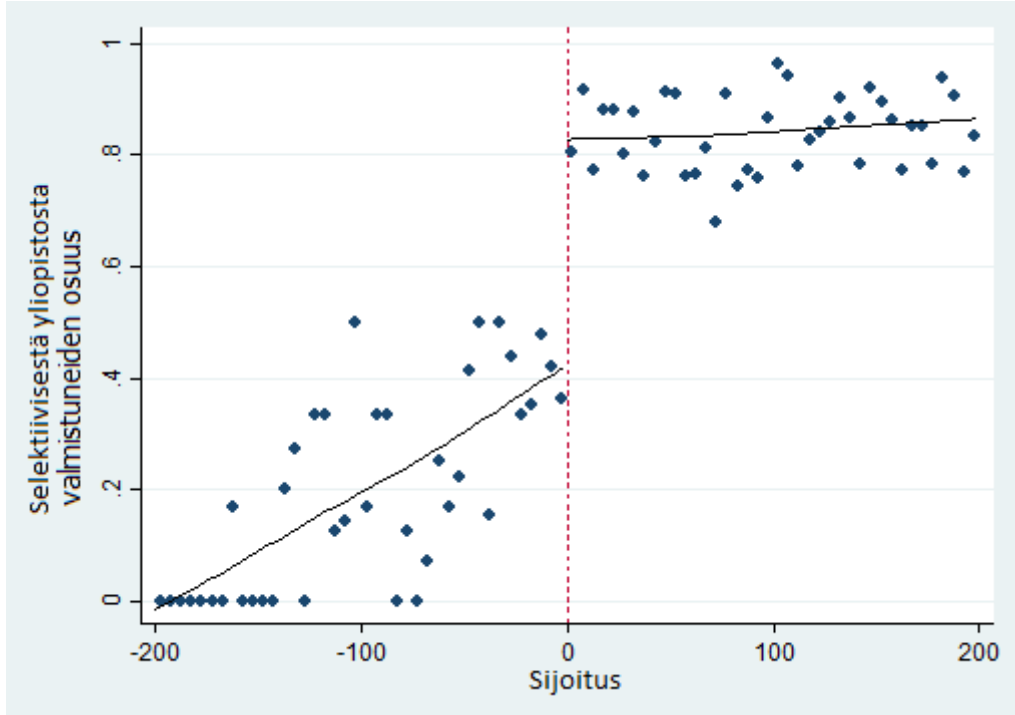
Tässä tutkielmassa käytetään FRD:tä, koska selektiivisestä teknillisestä yliopistosta valmistumisen todennäköisyydessä tapahtuu noin 40 prosenttiyksikön hyppäys sisäänpääsyrajalla (katso kuva 1). Sisäänpääsyrajana toimii nyt aiemmin luodun, keskistetyyn sijoitusmuuttujan (tästä eteenpäin puhutaan sijoituksesta tai juoksevasta muuttujasta) arvo nolla. Kuvassa 1 otos on rajattu juoksevan muuttujan arvojen -200 ja +200 välille. Pisteparvessa jokainen piste kuvaa keskimääräistä selektiivisestä yliopistosta valmistuneiden osuutta viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaletta lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä. Myös myöhemmissä kuvissa pisteparvi ja on esitetty vastaavasti.

Metropolialueen teknillisestä yliopistosta valmistumisen tulovaikutusten estimoinnissa instrumentoidaan valmistumista dummy-muuttujalla, joka saa arvon yksi, jos hakijan pisteet ylittivät helpoimman selektiivisen teknillisen yliopiston hakukohteen sisäänpääsyrajan. Muutoin dummy-muuttuja saa arvon nolla, eli hakijan pisteet eivät riittäneet helpoimpaan haettuun hakukohteeseen valituksi tulemiseen. Analyyseissa FRD tehdään käyttäen epäparametrista mallia. Tämä tarkoittaa yksinkertaisesti sitä, että aineistoa rajataan sisäänpääsyrajan ympäristöön.<sup>28</sup> Tähän käytetään Imbensin ja Kalanyaramanin (2012) versiota optimaalisesta havaintovälistä, jossa keskikohtana on juoksevan muuttujan nollakohta (katso kuva 1).

Optimaalisen havaintovälin valinta on kompromissi harhan ja tarkkuuden välillä. Mitä suurempi tarkasteltava havaintoväli on, sitä tarkempia tulokset ovat, mutta sitä suurempi on myös harhan mahdollisuus. Imbensin ja Kala-

---

<sup>28</sup>Tämä on tutkimuskirjallisuudessa yleisimmin käytetty malli. Vaihtoehtoisesti oltaisiin voitu käyttää parametrista mallia.



Kuva 1: Selektiivisestä teknillisestä yliopistosta valmistuneiden osuus sisäänpääsyrajalla.

nyaramanin ehdottamassa optimihavaintovälin valinnassa tarkoituksena on valita sellainen  $h$  (havaintoväli), joka toteuttaa ehdon

$$(5.1) \quad h = \operatorname{argmin}MSE(h).$$

Eli valitaan sellainen  $h$ , jolla keskineliövirhe on pienin mahdollinen. Optimihavaintovälin valinnan jälkeen otoksen ulkopuolelle rajataan kaikki havainnot, joilla juoksevan muuttujan arvo on  $> h$  tai  $< -h$ . Tarkasteluissa käytetään lisäksi painoina kolmio-kerneliä, koska se on kirjallisuudessa useimmin käytetty kernel-funktio ja lisäksi sen on osoitettu olevan paras lokaaleissa lineaarisissa regressioissa (Lee & Lemieux 2010). Formaalisti se on muotoa:

$$(5.2) \quad K_h(r_i) = \left(1 - \left|\frac{r_i}{h}\right|\right) * 1\left(\left|\frac{r_i}{h}\right| \leq 1\right),$$

missä  $h$  on edelleen optimihavaintoväli ja  $r_i$  on yksilötason juokseva muuttuja.  $K_h(r_i)$  on tällöin hakijaa vastaavan havainnon paino-arvo analyysissä. Tämä tarkoittaa kolmio-kernelin 5.2 tapauksessa sitä, että mitä lähempänä havainnot ovat sisäänpääsyrajaa sitä suuremman painon ne saavat, ja paino vähenee lineaarisesti etäisyyden kasvaessa. Funktiomuodon valinnassa käytetään Akaiken informaatiokriteeriä (AIC). AIC-kriteerin perusteella optimifunktio on ensimmäisen asteen polynomi.

Koska FRD toteutetaan käyttämällä 2SLS:ää, saadaan valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin estimoimalla seuraavat yhtälöt. Ensimmäisen vaiheen yhtälö on muotoa

$$(5.3) \quad D_i = \alpha_0 + \alpha_1 * r_i + \alpha_3 * V_i + \alpha_4 * r_i * V_i + \delta_i$$

ja toisen vaiheen yhtälö on muotoa

$$(5.4) \quad Y_i = \beta_0 + \beta_1 * r_i + \beta_3 * D_i + \beta_4 * r_i * D_i + \varepsilon_i.$$

Yhtälöissä 5.3 ja 5.4  $Y_i$  kuvaa jälleen yksilötason ansio- ja pääomatuloja logaritmituna,  $\alpha$  ja  $\beta$  ovat regressiokertoimia ja  $r_i$  on edelleen hakijakohtainen juokseva muuttuja.  $V_i$  on hakijakohtainen, selektiivisen teknillisen yliopiston sisäänpääsyrajan ylittämistä kuvaava dummy-muuttuja, joka saa arvon yksi, kun hakijan sijoitus on epänegatiivinen ja muutoin arvon nolla. Juoksevan muuttujan ja sisäänpääsyrajan ylittämistä kuvaavan muuttujan interaktion  $r_i * V_i$  lisääminen malliin sallii juoksevan muuttujan kertoimen vaihtelun sisäänpääsyrajan eri puolilla.  $D_i$  on selektiivisestä teknillisestä yliopistosta valmistumista kuvaava hakijakohtainen dummy-muuttuja, joka saa arvon yksi jos yksilö valmistui selektiivisestä yliopistosta ja muutoin arvon nolla. Yhtälöissä  $\delta_i$  ja  $\varepsilon_i$  ovat virhetermejä. Osassa tarkasteluja malliin lisätään myös hakuvuosi, hakukohde ja valintaryhmä kontrollit.

FRD:ssä kiinnostuksen kohteena on toisen asteen yhtälön muuttujan  $D_i$  saama parametri. Formaalisti valmistumisen vaikutus tuloihin voidaan esittää seuraavasti (esimerkiksi Imbens & Lemieux 2007):

$$(5.5) \quad \frac{\lim_{r \downarrow 0} \mathbf{E}[Y_i | r_i = r] - \lim_{r \uparrow 0} \mathbf{E}[Y_i | r_i = r]}{\lim_{r \downarrow 0} \mathbf{E}[D_i | r_i = r] - \lim_{r \uparrow 0} \mathbf{E}[D_i | r_i = r]}.$$

Yhtälössä 5.5  $\mathbf{E}$  kuvastaa odotusarvoa ja  $\lim$  raja-arvoa, kun sisäänpääsyrajaa lähestytään alapuolelta tai yläpuolelta. Lisätietoa sharp ja FRD:stä antavat myös Angrist ja Pischke (2008, s.189–203).

## 6 Tulokset

### 6.1 Selektiivisestä yliopistosta valmistumisen vaikutus tuloihin

Taulukossa 4 on raportoitu selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran tuloihin käyttäen FRD:tä. Selektiivisestä teknillisestä yliopistosta valmistumista instrumentoidaan tällöin sisäänpääsyrajan ylittämällä, kuten kuvattu yhtälöissä 5.3 ja 5.4. Taulukossa on esitetty tulokset käyttäen heteroskedastisuuden suhteen korjattuja keskivirheitä. Optimihavaintovälin lisäksi taulukosta löytyvät myös puolikkaalla ja kaksinkertaisella havaintovälillä saadut tulokset. Imbensin ja Kalanyaramanin optimihavaintoväli on nyt 261. Optimihavaintovälin kaksinkertaistaminen johtaisi tällöin otoksen vaihteluväliä suurempaan havaintoväliin. Tämän estämiseksi käytetään kaksinkertaisena havaintovälinä 450:ä. Taulukosta nähdään, että selektiivisestä teknillisestä yliopistosta valmistuminen nostaa optimihavaintovälillä alku-uran pääoma- ja ansiotuloja noin 18% verrattuna muista DIA-kohteista valmistumiseen. Tulokset eivät kuitenkaan ole tilastollisesti merkitseviä.<sup>29</sup>

Taulukko 4: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin eri havaintoväleillä käyttäen FRD:tä.

	ln(tulot)		
Valmistui selektiivisestä yliopistosta	0,180 (0,123)	0,111 (0,177)	0,140 (0,0897)
Ensimmäisen vaiheen F-arvot	207,311	109,167	263,588
N	1876	1035	2823
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet sulussa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ . Ei kontroleita.*

Useissa koulutuksen laatua RD:llä tutkivissa papereissa päätuloksina pidetään niitä RD-tuloksia, joihin on lisätty joitakin hakuhetkellä määräytyviä

<sup>29</sup>Analyysi oltaisiin myös voitu toteuttaa käyttäen päällekkäin kasattuja valintaryhmiä. Vaihtoehtoisesti otos oltaisiin voitu myös rajata koskemaan hakijoita, jotka hakivat vähintään kahteen teknilliseen yliopistoon, tällöin hakijat olisivat myös preferensseiltään hyvin samankaltaisia. Analyysit on tehty molemmilla tavoilla, joista jälkimmäisen tulokset on esitetty taulukossa 10. Saadut tulokset ovat hyvin samanlaisia kuin taulukossa 4 raportoivat estimaatit.

muuttujia (esimerkiksi hakuvuosi ja kohortti, jossa hakija haki). Tästä johtuen myös tässä lisätään FRD:hen kontrolleiksi hakuvuosi, hakukohde sekä valintaryhmä. Näin saatuja estimaatteja pidetään nyt päätuloksina. Taulukosta 5 nähdään, että kaksinkertaisella sekä optimihavaintovälillä saadut estimaatit nousevat kontrollien lisäämisen jälkeen hieman. Optimihavaintovälillä estimaatti nousee 20%:iin ja kaksinkertaisella havaintovälillä 17,5%:iin. Kontrollien lisäys tehtiin käytännössä yksitellen, jolloin huomattiin, että havaittu estimaattien muutos johtuu pääasiassa hakukohteiden lisäämisestä. Olisikin mielenkiintoista nähdä miten tulokset vaihtelevat, jos tutkittaisiin yksittäisiä hakukohteita. Valitettavasti aineiston pieni koko kuitenkin estää tämän, joten hakukohteita tyydytään kontrolloimaan.

Taulukko 5: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin eri havaintoväleillä käyttäen FRD:tä.

	ln(tulot)		
Valmistui selektiivisestä yliopistosta	0,203 (0,124)	0,0977 (0,181)	0,175* (0,0915)
Ensimmäisen vaiheen F-arvot	231,938	118,412	291,588
N	1876	1035	2823
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet suluissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ . Kontrolleina hakuvuosi, hakukohde ja valintaryhmä.*

Taulukosta 5 havaitaan myös, että vaikka saadut estimaatit ovat suuria, ei optimihavaintovälillä saatu tulos ole suurista keskivirheistä johtuen edelleenkin tilastollisesti merkitsevä. Sen sijaan kaksinkertaisella havaintovälillä saatu tulos on, mistä johtuen ei merkittävien tulovaikutusten olemassaoloa voida poissulkea.

Yhtenä syynä optimaalisella havaintovälillä saatujen tulosten suuriin keskivirheisiin saattaa olla havaintojen suhteellisen pieni määrä. RD:n käyttö vaatii erittäin ison otoskoon, joten on mahdollista, ettei optimihavaintovälin alle 2000 havainnon otoskoko ole riittävä. Liitteen kuvasta C.5 nähdään myös toinen mahdollinen selitys suurille keskivirheille. Kuvan perusteella sisänpääsyrajan alapuolella tulojen vaihtelu on huomattavasti suurempaa kuin rajan yläpuolella. Tämä ei ole sinällään yllättävää, sillä rajan alapuolella olevat hakijat valmistuvat useista eri yliopistoista ja eri alueilta, kun taas rajan yläpuolella olevat hakijat valmistuvat todennäköisesti metropolialueen teknillisestä yliopistosta. Tällöin on mahdollista, että tulopremio on erittäin suuri verrattuna joihinkin tiettyihin yliopistoihin ja olematon verrattuna toisiin.

Havaintomäärät jäisivät yksittäisiä yliopistoja vertaillessa kuitenkin melko pieniksi, mistä johtuen näitä estimointeja ei tehdä tässä tutkielmassa.

Taulukoista 4 ja 5 huomataan myös, että niissä esitetyt estimaatit ovat huomattavasti suurempia, kuin taulukossa 3 esitetyt OLS-regressiotulokset. Tämä on hieman yllättävä tulos, sillä OLS:n ajatellaan usein yliarvoivan tuloksia. Toisaalta, ero johtuu luultavasti tässä tapauksessa siitä, että FRD:llä saadut tulokset ovat lokaaleja, kun taas OLS:llä saadut tulokset globaaleja. Toisin sanoen on mahdollista, että heikoimmilla pisteillä sisääntulleet hyötyvät keskimääräistä enemmän valmistumisesta selektiivisestä yliopistosta esimerkiksi vertaisvaikutusten kautta (peer-effects).

Taulukoissa 4 ja 5 on esitetty myös analyysiin liittyvät ensimmäisen vaiheen F-estimaatin arvot. Instrumenttimuuttujien tapauksessa ongelmaksi usein muodostuu, ettei instrumentti ole kovin vahva (eli on selityskyvyltään heikko). Peukalosääntönä ekonometriassa pidetään F-estimaatin arvoa kymmenen (esimerkiksi Staiger & Stock 1997). Toisin sanoen sitä pienemmät F-arvot saava 2SLS tulkitaan selityskyvyltään heikoksi. Tulosten perusteella instrumenttien huono selityskyky ei tule olemaan ongelma, sillä molempien taulukoiden kaikilla havaintoväleillä F-arvot ovat yli sadan.

## 6.2 Validius- ja robustisuustarkastelut

Aloitetaan validiustarkastelut tutkimalla onko havaituissa taustamuuttujissa systemaattisia epäjatkuvuuksia sisäänpääsyräjällä. Toisin sanoen tutkitaan pitääkö se taustaoletus paikkansa, että rajan marginaalisesti ylittäneet ovat ominaisuuksiltaan samanlaisia kuin marginaalisesti rajan alle jääneet. Tätä voidaan tutkia helposti SRD:llä, jossa funktio vastaa funktiota 5.3, sillä erotuksella, että selitettävä muuttuja (valmistuminen) on korvattu mielenkiinnon kohteena olevalla taustamuuttujalla. Tässä tarkastelussa ei tavallisesti käytetä lisäkontrolleita.

Testattuja taustamuuttujia ovat sukupuoli, ikä, kieli, ylioppilasstatus, dummy-muuttuja, joka kertoo asuiko hakija pääkaupunkiseudulla hakuhetkellä, äidinkielen ja pitkän matematiikan arvosanat sekä vanhempien tulot (jatkuvana muuttujana), koulutus ja sosioekonominen asema. IK -optimihavaintovälin lisäksi tarkastelut on esitetty myös puolikkaalla ja kaksinkertaisella havaintovälillä. Nämä formaalit tarkastelut on esitetty taulukossa 6. Taulukossa estimaatit on tummennettu. Niiden alta löytyvät heteroskedastisuuden suhteen korjatut robustit keskivirheet ja havaintomäärät. Optimaalinen havaintoväli on laskettu jokaiselle muuttujalle erikseen ja kaksinkertaisen havaintovälin ylittäessä otoksen vaihteluvälin, käytetään sen sijaan havaintoväliä 450.

Taulukon perusteella taustamuuttujissa ei vaikuta olevan systemaattisia epäjatkuvuuksia. Ainoa muuttuja, joka on optimihavaintovälillä (5% riskitasolla) tilastollisesti merkitsevä on muuttuja, joka kertoo onko äidin tuorein tutkinto korkeakoulututkinto. Taustamuuttujien epäjatkuvuuksia voi-

Taulukko 6: Taustamuuttujien keskiarvojen epäjatkuvuus sisään pääsyrajalla käyttäen eri havaintovälejä.

Kontrollit	Havaintoväli		
	Optimi	Puoli	Kaksinkertainen
Mies	<b>-0,0313</b> (0,0440)	<b>0,0223</b> (0,0590)	<b>-0,0505</b> (0,0369)
N	1930	1082	2823
Ikä	<b>-0,0202</b> (0,0339)	<b>-0,0202</b> (0,0430)	<b>-0,0322</b> (0,0297)
N	2022	1124	2823
Suomen kieli	<b>0,0194</b> (0,0245)	<b>0,0451</b> (0,0323)	<b>0,0140</b> (0,0232)
N	2536	1426	2823
Ylioppilas	<b>-0,0171</b> (0,0148)	<b>-0,0175</b> (0,0212)	<b>-0,0148</b> (0,0129)
N	2251	1267	2823
Äidinkielen arvosana	<b>0,151*</b> (0,0845)	<b>0,165</b> (0,105)	<b>0,158*</b> (0,0870)
N	2832	1832	2780
Pitkänmatikan arvosana	<b>-0,0248</b> (0,103)	<b>-0,0950</b> (0,137)	<b>-0,0212</b> (0,0929)
N	2233	1250	2768
Pääkaupunkiseutu	<b>0,0613</b> (0,0382)	<b>0,0380</b> (0,0503)	<b>0,0825**</b> (0,0331)
N	2067	1148	2823
Vanhempien tulot	<b>4,834</b> (6,966)	<b>1,937</b> (8,991)	<b>5,075</b> (6,318)
N	2301	1296	2814
Äiti korkeakoulutettu	<b>0,119***</b> (0,0384)	<b>0,163***</b> (0,0495)	<b>0,116***</b> (0,0376)
N	2731	1547	2824
Isä korkeakoulutettu	<b>0,0389</b> (0,0366)	<b>0,0429</b> (0,0487)	<b>0,0287</b> (0,0331)
N	2251	1267	2823
Äiti työtön	<b>-0,00735</b> (0,0176)	<b>0,00506</b> (0,0245)	<b>-0,0175</b> (0,0149)
N	1890	1041	2823
Isä työtön	<b>0,00500</b> (0,0168)	<b>0,0326</b> (0,0222)	<b>0,00289</b> (0,0141)
N	2045	1132	2823

*Robustit keskivirheet suluissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Estimaatit tummennettu.*



daan myös helposti tarkastella graafisesti ja niitä esittävät kuvat löytyvätkin liitteen kuvista C.2–C.4.

Seuraavaksi tutkitaan aiheuttaako näiden taustamuuttujien lisääminen malliin suurta vaihtelua estimaatissa. Jos näin on, saattaa syynä olla se, että hakijat kykenevät täydellisesti taktikoimaan itsensä rajan yli. Tarkastelun tulokset on raportoitu taulukossa 7.<sup>30</sup> Tuloksia tulee verrata taulukon 5 tuloksiin, koska myös näissä on kontrolloitu hakuvuosi, -kohde ja valintaryhmä. Taulukosta huomataan, että tulokset pysyvät melko samanlaisina, varsinkin isommilla havaintoväleillä. Keskivirheet kasvavat myös vain marginaalisesti. Pienemmillä havaintoväleillä tulokset sen sijaan muuttuvat jonkin verran. Tämä saattaa johtua siitä, että pienillä havaintoväleillä ei yksinkertaisesti ole tarpeeksi havaintoja suuren kontrollimäärän tukemiseen. Kaikkiaan nämä tulokset antavat tukea päätulosten validiudelle.

Taulukko 7: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin eri havaintoväleillä käyttäen FRD:tä.

	ln(tulot)		
Valmistui selektiivisestä yliopistosta	0,214 (0,131)	0,188 (0,193)	0,171* (0,0947)
N	1876	1035	2823
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet suluissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .*

*Kontrolleina hakuvuosi, hakukohde ja valintaryhmä sekä kaikki taulukossa 6 luetellut kontrollit.*

Jatketaan validiustarkasteluja tutkimalla esiintyykö muualla aineistossa epäjatkuvuutta tuloissa. Tätä kutsutaan placebo-testiksi. Perinteisesti placebo-testi tehdään sisäänpääsyrajan ylä- ja alapuolelle. Esimerkiksi sisäänpääsyrajan yläpuolelle tehtäessä poistetaan kaikki varsinaisen rajan alle jäävät havainnot ja luodaan uusi, täysin synteettinen sisäänpääsyraja juoksevan muuttujan mediaanin kohtaan. Tämän jälkeen tutkitaan tulojen epäjatkuvuutta kyseisen rajan ympäristössä käyttäen FRD:tä. Jotta tämä testi läpäistäisiin, ei epäjatkuvuutta tuloissa saisi esiintyä synteettisen rajan ympäristössä. Testi tehdään vastaavasti oikean sisäänpääsyrajan alapuolelle. Placebo-testin tulokset on esitetty taulukossa 8. Estimaattien perusteella epäjatkuvuuksia ei esiinny synteettisten rajojen ympäristössä. Eli asetelma

<sup>30</sup> Jotta havaintoja olisi sama määrä kuin taulukossa 5, lisätään matematiikan arvosanoihin myös lyhyen matematiikan arvosanat ja muutetaan vanhempien tulot-muuttuja kategoriseksi muuttujaksi, kuten kuvattu liitteessä B.

Taulukko 8: Placebo-testi.

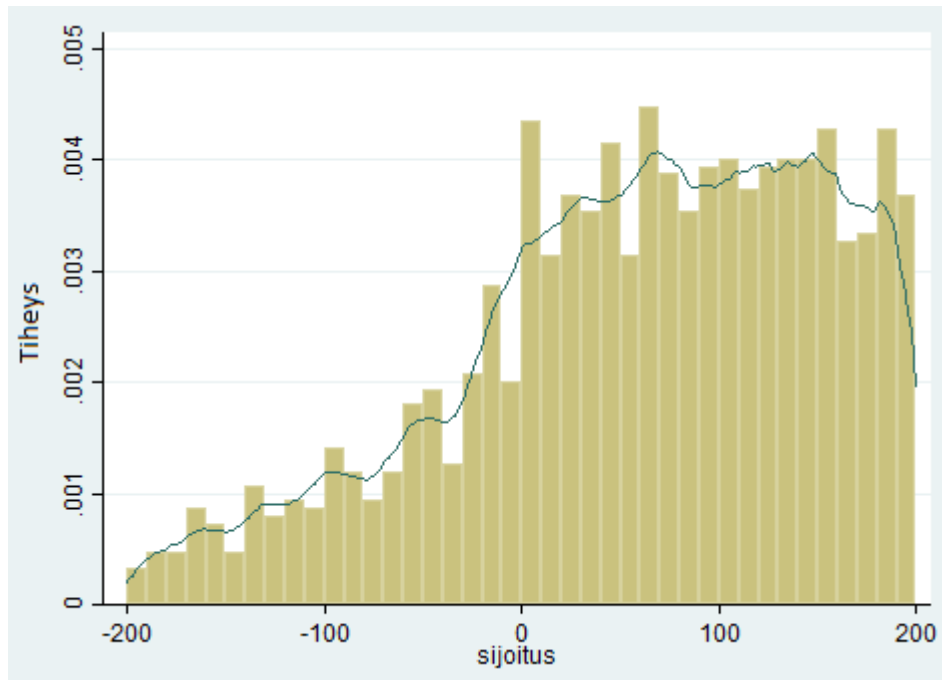
	ln(tulot)	
Valmistui selektiivisestä yliopistosta	1,861 (1,836)	3,292 (4,207)
N	284	2190
Sisäänpääsyraja	Rajan alapuoli	Rajan yläpuoli

*Robustit keskivirheet sulussa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran  
pääoma- ja ansiotuloihin optimihavaintovälillä, kun muodostettu uusi  
sisäänpääsyraja rajan ala- ja yläpuolisiin mediaanikohtiin.*

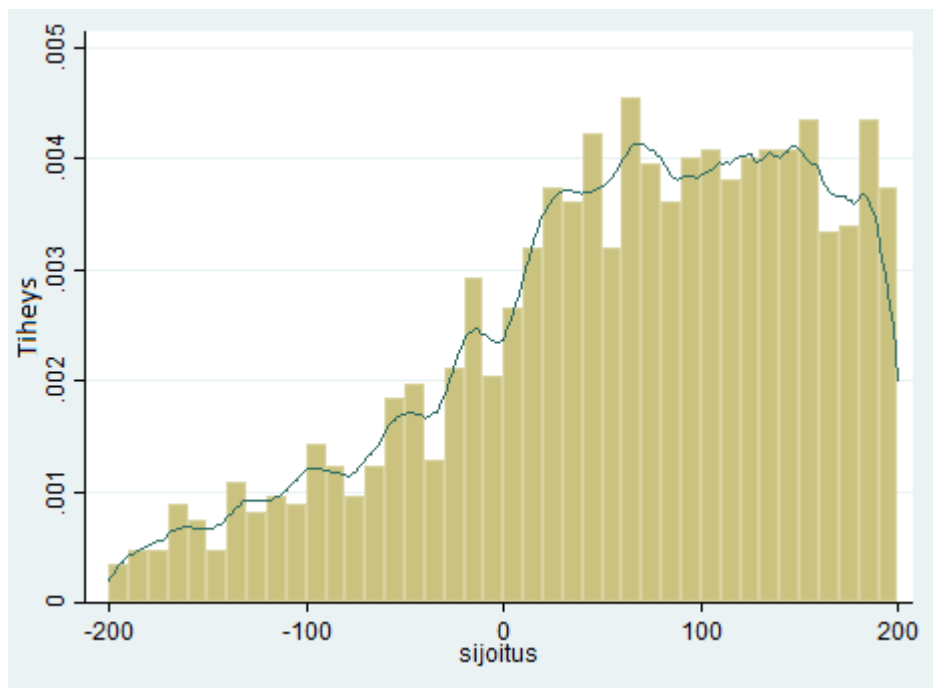
läpäisee tämän placebo-testin.

Seuraavaksi tutkitaan esiintyykö sisäänpääsyrajalla epäjatkuvuutta juoksevan muuttujan jakauman tiheydessä. Tämä liittyy aiemmin mainittuun taktikointiongelmaan, joka siis tarkoittaa sitä, että hakijat pystyisivät kontrolloimaan valituksi tulemistaan (esimerkiksi tietäisivät täysin ennalta pisterajat, hakisivat kunnes pääsisivät tai olisivat voineet lahjoa päättäjät). Lisäpisteiden ja edellisten vuoden pisterajojen julkisuuden vuoksi taktikointia voi esiintyä. Tässä tutkimuksessa prosessin monimutkaisuudesta, valintaryhmän 1 pudottamisesta ja pisterajojen vuosittaisesta vaihtelusta johtuen taktikointi tuskin kuitenkaan muodostuu suureksi ongelmaksi. Myös aineiston raja-ensimmäistä kertaa DIA-haussa mukana oleviin nuoriin pienentää taktikointiongelman esiintymisen todennäköisyyttä. Lisäksi kuten Lee (2008) osoittaa RD on validi, kunhan hakijoiden harjoittama kontrolli ei ole täydellistä. Toisin sanoen tulos on validi, kunhan hakijat eivät pysty täysin määräämään lopputulosta, vaan siihen liittyy myös jokin satunnainen elementti.

Yleensä taktikointiongelmaa tarkastellaan käyttäen McCraryn (2008) testiä, joka testaa epäjatkuvuutta juoksevan muuttujan tiheydessä sisäänpääsyrajalla. Jos rajalla tapahtuisi positiivinen hyppäys, voisi se olla merkki taktikointiongelma. Tämä testi ei kuitenkaan ole tässä tilanteessa kovin käyttökelpoinen, sillä jokaista valintaryhmä–hakukohde–vuosi-yhdistelmää vastaa vähintään yksi täsmälleen juoksevan muuttujan arvon nolla saava havainto. Tällöin juoksevan muuttujan tiheydessä tapahtuu väistämättä hyppäys rajalla. Kuvissa 2 ja 3 esitetyt histogrammit kuvaavat käytetyn otoksen juoksevan muuttujan jakaumaa, kuviin on piirretty myös juoksevan muuttujan jakauman kernel-tiheysfunktio (täsmällisesti ottaen kolmio-kernel). Histogrammien pylväiden väli on 10 sijoitusta. Ylemmässä kuvassa näkyy sisäänpääsyrajalla selkeä hyppäys. Kuva 3 vastaa täysin kuvaa 2, sillä erotuksella, että siitä on poistettu kaikki ne havainnot, joissa juokseva muuttuja saa täsmälleen arvon nolla. Kuten huomataan, ylemmässä histogrammissa havaittua



Kuva 2: Juoksevan muuttujan jakauma sekä kernel-tiheysfunktio.



Kuva 3: Juoksevan muuttujan jakauma sekä kernel-tiheysfunktio. Muutoin sama kuin 2, sillä erotuksella, että on poistettu täsmälleen arvon nolla saaneet hakijat.

terävää hyppäystä sisäänpääsyrajalla ei enää tapahdu. Tämän perusteella taktikointiongelmia ei esiinny.

Valmistumisen vaikutusten tutkimisessa yhtenä huolenaiheena on lisäksi valikoituminen. Koska aineistosta puuttuvat ulkomailla asuvien tulotiedot, on esimerkiksi mahdollista, että alku-uran logaritmoitujen tulojen havaitsemistodennäköisyydet eroavat sisäänpääsyrajan ylä- ja alapuolella. Jos näin olisi, saattaisi se olla merkki siitä, että otos on valikoitunut. Tarkastellaan asiaa käyttäen SRD:tä samaan tapaan kuin taustamuuttujien tarkasteluissa taulukossa 6. Estimoinnin tulokset ovat esitetty taulukossa 9. Niiden perusteella hyppäystä rajalla ei tapahdu, eli tulojen havaitsemistodennäköisyydet ovat rajan tuntumassa keskimäärin samat. Toisin sanoen otos ei ole tulojen havaitsemistodennäköisyyden suhteen valikoitunut.

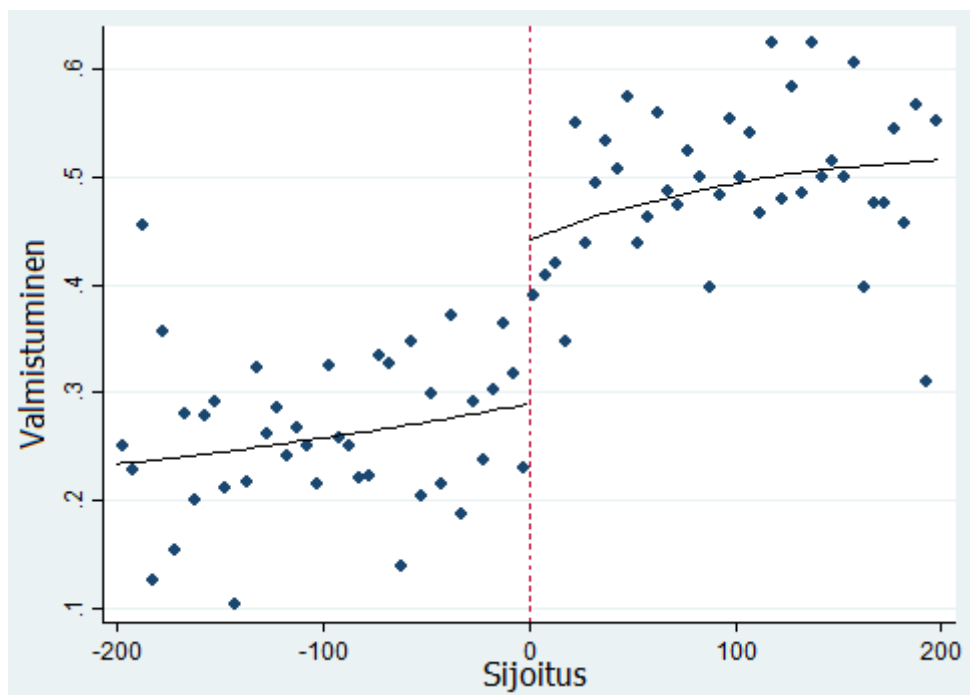
Taulukko 9: Logaritmoitujen tulojen havaitsemistodennäköisyys sisäänpääsyrajalla kahdeksan vuoden kuluttua ensimmäisestä hakuhetkestä.

Tulojen havaitsemistodennäköisyys			
Ylitti sisäänpääsyrajan	0,00445 (0,0125)	-0,00543 (0,0165)	0,00451 (0,0123)
N	2824	1607	2882
Havaintoväli	Optimi	Puoli	Kaksinkertainen
<i>Robustit keskivirheet suluisissa, *** <math>p &lt; 0,01</math>, ** <math>p &lt; 0,05</math>, * <math>p &lt; 0,1</math>.</i>			

On myös mahdollista, että selektiivisen hakukohteen rajan ylittäneistä valmistuvat vain kaikkein kyvykkäimmät. Jos näin tapahtuisi, otos olisi valikoitunut ja saadut tulokset todennäköisesti ylöspäin harhaisia. Asian tutkimiseen käytetään otosta, jota ei ole rajattu teknillisistä hakukohteista valmistuneisiin.

Tällöin tarkastellaan tapahtuuko selektiivisen hakukohteen sisäänpääsyrajalla hyppäystä alaspäin tutkittaessa teknillisistä hakukohteista valmistumisen todennäköisyyttä. Toisin sanoen tutkitaan valmistuvatko selektiiviseen yliopistoon marginaalisesti valituiksi tulleet teknillisistä yliopistoista pienemällä todennäköisyydellä, kuin rajan marginaalisesti alittaneet. Kuten kuvasta 4 huomataan, rajalla tapahtuu kylläkin hyppäys, mutta se on positiivinen, ei negatiivinen. Toisin sanoen oletetun kaltaista valikoitumista ei näytä tapahtuvan. Sen sijaan vastakkaisen kaltaista valikoitumista saattaa kuvan perusteella tapahtua. On tosin mahdollista, että havaittu hyppäys kertoo vain selektiivisen teknillisen yliopiston korkeammasta laadusta, ei valikoitumisesta.

Yhtenä keinona tämän potentiaalisen valikoitumisen poistamiseksi olisi rajata aineisto hakijoihin, joilla on hyvin samankaltaiset preferenssit. Toise-



*Huomioita: Kuvassa jokainen piste kuvaa keskimääräistä valmistuneiden osuutta viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä.*

Kuva 4: Teknillisistä yliopistoista valmistuneiden osuus sisäänpääsyrajalla.

na mahdollisuutena on tehdä kuten Kirkeboen ym. (2015) eli kontrolloida hakijoiden preferenssejä. Näiden preferenssien lisääminen malliin on kuitenkin tässä tapauksessa hieman ongelmallista muutamista syistä. Ensinnäkin, koska hakukohteita oli useita kymmeniä, kasvattaisi preferenssien lisääminen analyysiin kontrollien määrää huomattavasti, mikä ei suhteellisen pienillä otoskoilla ole välttämättä haluttavaa. Toisekseen koska DIA-valinnan ulkopuolisia hakupreferenssejä ei tunneta, olisi havaittujen preferenssien kontrolloimisen jälkeenkin mahdollista, että tuloksissa on harhaa.

Tarkastellaan tätä mahdollista valikoitumista vielä hieman rajaamalla aineisto hakijoihin, joilla on hieman samankaltaisemmat preferenssit. Tämä tehdään rajaamalla aineistosta pois ne, jotka hakevat vain selektiiviseen teknilliseen yliopistoon. Estimoinnin tulokset löytyvät taulukosta 10 ja liitteen kuvasta C.6. Aiemman kaltaista hyppyä valmistumisen todennäköisyydessä ei enää havaita, mutta piste-estimaatit selektiivisestä yliopistosta valmistumisen tulovaikutuksista ovat hyvin samankaltaiset kuin taulukossa 4 esitetyt estimaatit. Keskivirheet sen sijaan kasvavat johtuen luultavasti havaintomäärän pienentymisestä. Estimaatit ovat hyvin samankaltaisia kuin aiemmin saadut estimaatit, mikä myös antaa tukea tulosten validiudelle.

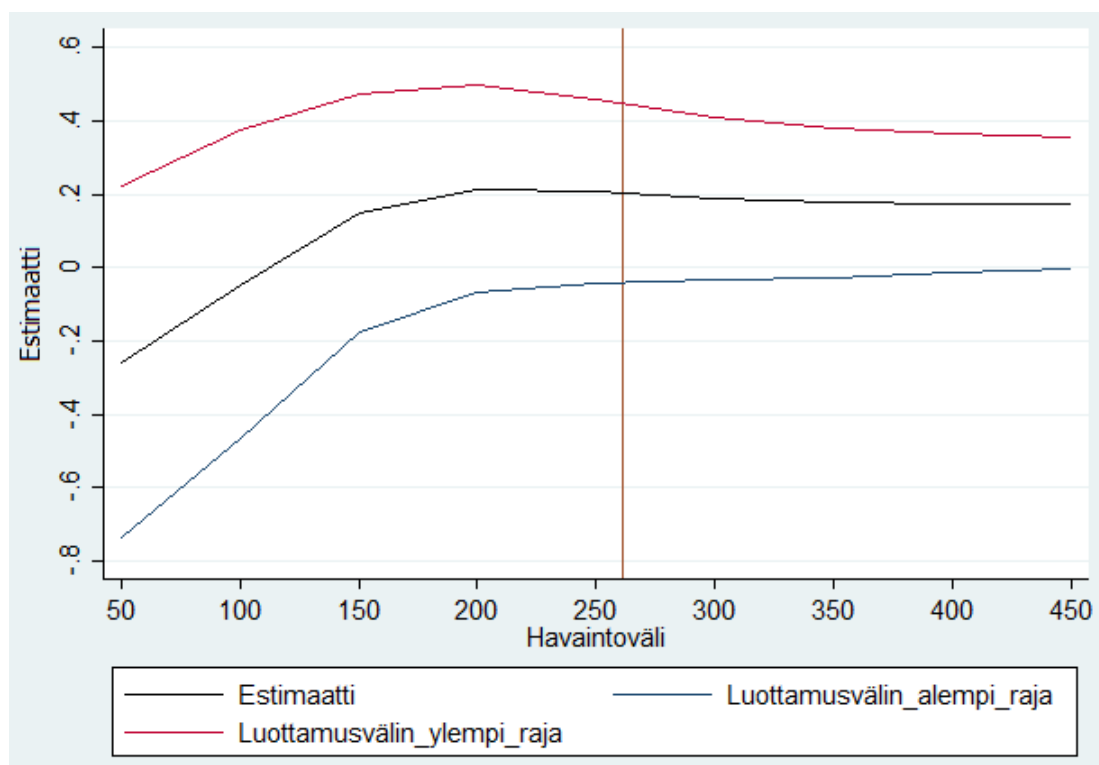
Aloitetaan robustituustarkastelut estimoimalla tulopremio useilla eri ha-

Taulukko 10: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin.

	ln(tulot)		
Valmistui selektiivisestä yliopistosta	0,151 (0,141)	0,0609 (0,195)	0,148 (0,115)
N	1102	667	1314
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet sulussa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Otos rajattu niihin, jotka hakivat vähintään kahteen teknilliseen yliopistoon. Ei ylimääräisiä kontrolloita.*

vaintoväleillä. Tällöin tarkastellaan saatujen estimaattien sensitiivisyyttä havaintovälin valintaan. Kuvassa 5 on esitetty saadut tulokset sekä niiden 95% luottamusvälit käyttäen havaintovälejä 50 sijoituksen välein, välillä 50–450. Kuvan esittämässä tuloksissa on kontrolloitu hakijoiden hakuvuosi, hakukohde sekä valintaryhmä. Kuvaan on piirretty pystyviiva optimihavaintovälin



Kuva 5: Selektiivisestä yliopistosta valmistumisen vaikutus pääoma- ja ansiotuloihin eri havaintoväleillä sekä 95% luottamusvälit.

(261) kohdalle. Kuvasta nähdään, että estimaatin tarkkuus lisääntyy havaintovälin kasvaessa. Tämä on RD:lle hyvin tyypillistä. Estimaatit stabiloituvat havaintovälin kasvaessa yli 200:n noin 20%:in tienoille.

Toisena robustisuustarkasteluna analyysit on myös tehty käyttäen muutamia erilaisia funktiomuotoja. Tässä käytetään ensimmäisen asteen polynomin lisäksi myös toisen ja kolmannen asteen polynomeja. Nämä tulokset on esitetty taulukossa 11. Taulukosta nähdään, että toisen asteen polynomin käyttö nostaa optimihavaintovälin estimaatin yli 30%:n, sen sijaan kolmannen asteen polynomin käyttö heilauttaa tuloksia takaisin alaspäin, noin 25%:iin. Keskiarvot suurenevät optimihavaintoväleillä molemmissa tapauksissa verrattuna taulukon 5 keskiarvoihin. Huolimatta estimaattien suurista heilahteluista tulokset ovat kuitenkin optimaalisella ja kaksinkertaisella havaintovälillä aina positiivisia ja estimaatit yli 16%.

Taulukko 11: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus alku-uran pääoma- ja ansiotuloihin käyttäen toisen ja kolmannen asteen polynomia.

Polynomi	Havaintoväli		
	Optimi	Puoli	Kaksinkertainen
Toisen asteen polynomi	<b>0,334</b>	<b>0,0880</b>	<b>0,334*</b>
	(0,214)	(0,251)	(0,173)
N	1876	1035	2823
Kolmannen asteen polynomi	<b>0,247*</b>	<b>-0,449</b>	<b>0,169</b>
	(0,137)	(0,207)	(0,114)
N	1876	1035	2823

*Robustit keskiarvot sulussa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Kontrollina hakuvuosi, -kohde ja valintaryhmä. Estimaatit tummennettu.*

### 6.3 Muita tuloksia

Alaluvussa 6.1 saatu tulopremio on erittäin iso verrattuna muista Pohjoismaista saatuihin tuloksiin. Tästä johtuen tarkastellaan seuraavaksi muutamia potentiaalisia, alku-uran tuloihin vaikuttavia tekijöitä. Tarkastelut aloitetaan analysoimalla päätyivätkö metropolialueen selektiivisestä teknillisestä yliopistosta valmistuneet muita useammin Uudellemaalle asumaan. Uudella maalla asuvilla on muualla Suomessa työskenteleviä suuremmat tulot, jo-

ten on mahdollista, että se osaltaan selittää tulopreemion suuruutta.<sup>31</sup> Tutkitaan tätä käyttäen FRD:tä, joten tulokset pätevät jälleen vain lokaalisti sisäänpääsyrajan ympäristössä. Tulokset on raportoitu taulukossa 12 optimihavaintovälin lisäksi puolikkaalle ja kaksinkertaiselle havaintovälille. Taulukosta käy ilmi, että selektiivisestä teknillisestä yliopistosta valmistuminen nostaa sisäänpääsyrajoilla todennäköisyyttä sijoittua Uudellemaalle yli 46 prosenttiyksikköä. On siis mahdollista, että tämä selittää osaltaan havaittua tulopreemiota.

Taulukko 12: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus Uudellamaalla asumisen todennäköisyyteen tulojen havainnointitihetkellä.

	Uusimaa		
Valmistui selektiivisestä yliopistosta	0,458*** (0,0695)	0,385*** (0,107)	0,522*** (0,0478)
N	1740	958	2823
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet suluissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Kontrolleina hakuvuosi, -kohde ja valintaryhmä.*

VATT-työryhmän (2015) tutkimuksen mukaan Helsingin aluepalkka on kuitenkin vain noin 5% korkeampi kuin Suomessa keskimäärin. Tämän perusteella ei havaittua tulopreemion suurta kokoa voida kokonaisuudessaan selittää Uudenmaan suuremmalla ansiopreemiolla. Myös Suhonen (2014) tutkii eroavatko metropolialueella asuvien alku-urien tulot niiden tuloista jotka asuvat muualla. Hän pystyy tutkimuksessaan myös erottelemaan tulokset eri aloilla. Tulosten mukaan metropolialueella asuminen ei vaikuta alku-uran tuloihin teknillisellä alalla.

Yliopiston sijainti metropolialueella voi vaikuttaa tuloihin myös muita kanavia pitkin. On esimerkiksi mahdollista, että selektiivisestä teknillisestä yliopistosta valmistuneet ovat työskennelleet opintojensa ohessa enemmän kuin muista teknillisistä yliopistoista valmistuneet. Tällöin heidän suurempi työkokemuksensa selittäisi osaltaan tulopreemion suuruutta. Tätä mahdollisuutta tutkitaan niin ikään käyttäen FRD:tä. Saadut tulokset on raportoitu taulukossa 13. Niiden perusteella metropolialueen selektiivisestä teknillisestä yliopistosta valmistuneet ovat työskennelleet opintojensa aikana noin 12,5 kuukautta enemmän kuin muista teknillisistä yliopistoista valmistuneet.

<sup>31</sup> Yhtenä mahdollisena syynä Uudenmaan korkeampiin tuloihin on, että siellä toimivat yritykset ovat muuta maata tuottavampia ja pystyvät tällöin maksamaan muuta maata korkeampaa palkkaa.



Mahdollisia syitä tähän on useita. Esimerkiksi on mahdollista, että metropolialueella opiskeleminen ja asuminen on muuta maata kalliimpaa, jolloin opintojen ohessa työskentely on ollut monille välttämättömyys. On myös mahdollista, että metropolialueella on muuta maata enemmän opintojen aikaisia työmahdollisuuksia.

Taulukko 13: Selektiivisestä teknillisestä yliopistosta valmistuneiden opintojen aikaisten työkuukausien määrä verrattuna muista teknillisistä yliopistoista valmistuneisiin.

	Opintojen aikaiset työkuukaudet		
Valmistui selektiivisestä yliopistosta	12,47*** (2,920)	9,754** (4,568)	11,53*** (2,243)
N	2067	1148	2823
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet suluisissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Kontrolleina hakuvuosi, -kohde ja valintaryhmä.*

Tutkitaan myös, onko valmistumisen nopeuksissa eroja teknillisten yliopistojen välillä. Tätä tutkitaan käyttäen jälleen FRD:tä. Saadut tulokset on esitetty taulukossa 14. Taulukon mukaan selektiivisestä teknillisestä yliopistosta valmistuminen kestää noin yhdeksän kuukautta (eli 0,76 vuotta) pidempään kuin muista teknillisistä yliopistoista valmistuminen. Tämä antaa viitteitä siitä, että tulopreemio saattaisi myöhemmin mitattuna olla vielä tässä havaittuakin preemiota suurempi. Toisaalta on mahdollista, että opinto-

Taulukko 14: Selektiivisestä teknillisestä yliopistosta valmistuneiden valmistumisnopeus verrattuna muista teknillisistä yliopistoista valmistuneisiin.

	Valmistumisnopeus		
Valmistui selektiivisestä yliopistosta	0,755*** (0,160)	0,780*** (0,239)	0,683*** (0,134)
N	2254	1273	2823
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet suluisissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ . Kontrolleina hakuvuosi, -kohde ja valintaryhmä. Valmistumisnopeus mitattu vuosissa.*

jen pidempi kesto selittyy täysin aiemmin havaitulla suuremmalla opintojen aikaisella työssäkäynnillä. Tällöin opinnoille jäi vähemmän aikaa. On myös mahdollista, että kurssit olivat metropolialueen hakukohteissa vaativampia kuin muualla maassa, jolloin niiden suorittamiseen kului kauemmin.

Tarkastellaan seuraavaksi työskentelevätkö selektiivisestä teknillisestä yliopistosta valmistuneet muita enemmän tulojen havaintovuonna käyttäen jälleen FRD:tä. Nämä tulokset on esitetty taulukossa 15. Tulosten perusteella vaikuttaa siltä, että selektiivisestä teknillisestä yliopistosta valmistuminen kasvattaa tulojen havaintovuoden työkuukausia noin 0,73 kuukaudella. Tämä saattaa myös osaltaan selittää havaitun tulopreemion suuruutta. Alle kuukauden työssäkäyntipremio tuskin kuitenkaan selittää kokonaisuudessaan estimoidun tulopremio suuruutta.

Taulukko 15: Selektiivisestä teknillisestä yliopistosta valmistumisen vaikutus havaintovuoden työkuukausiin.

	Työkuukaudet havaintovuonna		
Valmistui selektiivisestä yliopistosta	0,729** (0,311)	0,954** (0,466)	0,682** (0,265)
N	2240	1248	2733
Havaintoväli	Optimi	Puoli	Kaksinkertainen

*Robustit keskivirheet suluissa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Kontrolleina hakuvuosi, -kohde ja valintaryhmä.*

Estimoidaan lopuksi tulopremio vielä erikseen miehille ja naisille, sillä on mahdollista, että alku-uran tulopremio vaihtelee riippuen sukupuolesta. Naiset esimerkiksi saavat usein ensimmäiset lapsensa alku-uransa aikana, mistä johtuen tulopremio saattaa olla heillä huomattavasti pienempi.<sup>32</sup> Taulukossa 16 on raportoitu selektiivisestä yliopistosta valmistumisen tulovaikutukset erikseen miehille ja naisille käyttäen FRD:tä. Taulukosta havaitaan, etteivät saadut estimaatit eroa miehillä ja naisilla juurikaan toisistaan (noin 20%). Keskivirheet sen sijaan ovat naisilla suuremmat, mikä johtuu luultavasti huomattavasti pienemmästä otoskoosta. Luultavasti juuri suuremmasta keskivirheestä johtuen naisten kaksinkertaisella havaintovälillä saatu estimaatti ei ole tilastollisesti merkitsevä, mutta miesten vastaava estimaatti on. Taulukosta huomataan myös, että saadut tulopremiot ovat hyvin samankaltaisia kuin taulukossa 5 esitetyt tulokset. Näiden havaintojen perusteella sukupuolten väliset erot eivät vaikuta ajavan tuloksia.

<sup>32</sup> Esimerkiksi Tilastokeskuksen (2009) mukaan vuonna 2008 äideiksi tultiin keskimäärin 28,2 vuotiaina.

Taulukko 16: Selektiivisestä yliopistosta valmistumisen vaikutus alku-uran tuloihin miehille ja naisille käyttäen FRD:tä.

Sukupuoli	Havaintoväli		
	Optimi	Puoli	Kaksinkertainen
Miehet	<b>0.205</b> (0,146)	<b>0.113</b> (0,262)	<b>0,197*</b> (0,109)
N	1443	786	2085
Naiset	<b>0.200</b> (0,172)	<b>0.239</b> (0,212)	<b>0.194</b> (0,170)
N	535	333	540

*Robustit keskivirheet sulussa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ . Kontrolleina hakuvuosi, -kohde ja valintaryhmä. Estimaatit tummennettu.*

## 7 Lopuksi

Tässä tutkielmassa tutkitaan koulutuksen laadun vaikutuksia alku-uran pääoma- ja ansiotuloihin teknillisen alan sisällä. Käytännössä vertaillaan selektiivisestä teknillisestä yliopistosta valmistuneiden tuloja muista teknillisistä yliopistoista valmistuneiden tuloihin kahdeksan vuoden kuluttua ensimmäisestä hakukerrasta. Tutkielma on eräs ensimmäisiä, jossa koulutuksen laadun vaikutuksia tarkastellaan tietyn alan sisällä. Tästä johtuen saadut tulokset kuvastavat nimenomaan kiinnostuksen kohteena olevan korkeakoulun vaikutusta tuloihin eivätkä esimerkiksi eri alojen tuloeroja. Tulokset ovat kiinnostavia paitsi yhteiskunnan myös yksilön näkökulmasta, sillä ne tarjoavat lisää informaatiota yksilön kouluvalinnan tueksi.

Tutkielmassa koulutuksen laadun mittarina käytetään teknillisen yliopiston selektiivisyyttä. Se onkin tutkimuskirjallisuudessa eräs käytetyimpiä koulutuksen laadun mittareita, johtuen luultavasti sen mittaamisen helppoudesta. Selektiivinen teknillinen yliopisto on kuitenkin muillakin mittareilla mitattuna erittäin laadukas yliopisto. Se sijaitsee erittäin hyvällä työmarkkina-alueella ja siellä on teknillisistä yliopistoista maan toiseksi suurin opettaja/opilas-suhde.

Päämenetelmänä tutkielmassa käytetään FRD:tä (fuzzy regression epäjatkuvuusmenetelmä), jossa hyödynnetään selektiivisen yliopiston sisäänpääsyrajoilla esiintyvää epäjatkuvuutta valmistumistodennäköisyydessä. Menetelmän etu on se, että sen avulla kyetään luotettavasti estimoimaan yliopistosta valmistumisen tulovaikutus pelkän korrelaation sijaan. Huono puoli menetelmässä taas on se, että tulokset pätevät vain sisäänpääsyrajan ympäristössä. Käytännössä selektiivisestä teknillisestä yliopistosta valmistumista instrumentoidaan tutkielmassa kyseisen yliopiston sisäänpääsyrajan ylittämistä kuvaavalla dummy-muuttujalla.

Koska niin sanottua naiivia regressiomallia pidetään tutkimuskirjallisuudessa benchmark-lähestymistapana, raportoidaan tulokset myös käyttäen pienimmän neliösumman (PNS/ OLS) menetelmää. OLS:llä saatujen tulosten perusteella selektiivisestä teknillisestä yliopistosta valmistuminen nostaa alku-uran tuloja noin 9,7%. Sen sijaan FRD:n tulosten mukaan tulopreemio voi sisäänpääsyrajan ympäristössä olla jopa 20%. Kesquivirheet ovat kuitenkin suuria, eikä optimaalisella havaintovälillä saatu tulos eroa tilastollisesti merkittävästi nollassa (edes 10% riskitasolla). Tästä huolimatta ei merkittävien tulovaikutusten olemassaoloa voida poissulkea, sillä kaksinkertaisella havaintovälillä saatu tulos on tilastollisesti merkittävä.

FRD tulosten keskivirheiden suuruus voi johtua siitä, ettei havaintoja ole riittävästi. Myöhemmissä tutkimuksissa olisikin mielenkiintoista toistaa analyysi aineistolla, jossa tulot havaitaan pidemmälle ajanjaksolle ja jossa olisi siten enemmän havaintoja. Toinen kiinnostava myöhempi tutkimuskohde olisi tulosten estimointi siten, että kontrolliryhmäksi otettaisiin vain yksi

yliopisto kerrallaan. Oletettavaa on, että myös tällöin keskivirheet saataisiin laskemaan ja tuloksista siten tarkempia. Olisi myös kiinnostavaa estimoida tulokset hakukohteittain. Hakukohteiden suuren määrän vuoksi tähän vaa-dittaisiin kuitenkin nykyistä huomattavasti suurempi aineisto.

FRD:llä saadut estimaatit ovat tarkasteluiden perusteella valideja ja melko robusteja. Ne ovat kuitenkin huomattavasti suurempia kuin useimmissa aikaisemmissa pohjoismaisissa tutkimuksissa saadut tulokset. Tästä johtuen tutkielmassa tarkastellaan myös muutamia tekijöitä, jotka saattavat vaikuttaa tuloihin. Myös näihin estimointeihin käytetään FRD:tä, joten saadut tulokset pätevät vain lokaalisti sisäänpääsyrajan ympäristössä.

Tulosten perusteella selektiivisestä yliopistosta valmistuneet päätyvät asumaan muita todennäköisemmin Uudellemaalle, jossa esimerkiksi palkat ovat muuta maata suuremmat. VATT:in työryhmän ja Suhosen (2014) saamien tulosten perusteella ei saatua preemiota voida kuitenkaan täysin selittää pääkaupunkiseudun korkeammalla palkkatasolla. Selektiivisestä teknillisestä yliopistosta valmistuneet ovat myös työskennelleet noin 12 kuukautta muita enemmän opintojensa aikana. Heidän opintonsa ovat myös kestäneet noin yhdeksän kuukautta muita pidempään. Lisäksi teknillisestä yliopistosta valmistuminen nostaa tulosten perusteella tulojen havaintovuoden työkuukausien määrää lähes kuukaudella.

Etenkin tulojen havaintovuoden suurempi työkuukausien määrä saattaa-kin osaltaan selittää tulopreemion suuruutta, mutta on epätodennäköistä, että koko havaittu premio selittyisi kokonaisuudessaan alle kuukauden työ-määrän erolla. Tutkielman tulosten erot muihin pohjoismaisiin tuloksiin ver-rattuna voivat johtua myös siitä, että tämä tutkimus on ensimmäisiä, jossa koulutuksen laadun tulovaikutuksia tutkitaan tietyn alan sisällä.

Aikaisempien tutkimusten tulosten perusteella koulutuksen laadun tu-lovaikutusten vaikutuskanavana vaikuttavat toimivan sekä inhimillisen pää-oman teoria että signaalointiteoria. Johtuen tästä sekä siitä, ettei saatujen tulosten perusteella voida täysin poissulkea kausaalivaikutuksen olemassaoloa, voi yliopistojen laatuun investoiminen ainakin teknillisellä alalla olla yhteiskunnan kannalta järkevää. Laatuun voitaisiin investoida esimerkiksi opettaja/opiskelija-suhdetta kasvattamalla. Lisäksi, koska selektiivinen yliopisto sijaitsee hyvällä työmarkkina-alueella (metropolialueella) on myös syy-tä pohtia, kannattaisiko muitakin teknillisiä yliopistoja keskittää tulevaisuu-  
dessa nykyistä enemmän hyville työmarkkina-alueille.

Toisaalta huonoina aikoina, jolloin myös koulutuksen rahoituksesta joudutaan leikkaamaan, on tulosten perusteella säästökohteet ainakin teknillisten yliopistojen tapauksessa valittava varoen. Leikkausten kohdistuessa esimer-kiksi opetushenkilöstön määrään, on mahdollista, että Suomen talouskasvu hidastuu tulevaisuudessa entisestään. Sen sijaan esimerkiksi päällekkäisyyk-siä karsimalla ja koulujen välistä yhteistyötä lisäämällä saattaa olla mah-dollista paitsi välttää koulutuksen laadun heikentyminen, myös luoda uusia innovaatioita eri alojen yhteistyön kautta.

## Viitteet

- [1] Abdulkadiroglu, Atila & Angrist, Joshua & Pathak, Parag (2011). *The elite illusion: Achievement effects at Boston and New York exam schools*. National Bureau of Economic Research, Working Paper No.17264
- [2] Abdulkadiroglu, Atila & Angrist, Joshua & Pathak, Parag (2014). *The elite illusion: Achievement effects at Boston and New York exam schools*. *Econometrica*, 82(1): 137–196.
- [3] Angrist, Joshua & Pischke, Jörn-Steffen (2008). *Mostly Harmless Econometrics*. Princeton University Press: New Jersey: 112–119 & 189–203.
- [4] Becker, Gary (1962). *Investment in Human Capital: A Theoretical Analysis*. *Journal of Political Economy* 70: 9–49.
- [5] Behrman, Jere & Rosenzweig, Mark & Taubman, Paul (1996). *College Choice and Wages: Estimates Using Data on Female Twins*. *Journal of Labour Economics*, 24: 701–728.
- [6] Black, Dan & Smith, Jeffrey (2004). *How Robust is the Evidence on the effects of College Quality? Evidence from matching*. *Journal of Econometrics*, 121: 99–124.
- [7] Black, Dan & Smith, Jeffrey (2006). *Estimating the Returns to College Quality with multiple proxies for Quality*. *Review of Economics and Statistics* 78: 672–685.
- [8] Borgen, Nicolai (2014). *College Quality and Hourly Wages: Evidence from the Self-Revelation Model, Sibling Models and Instrumental Variables*. *Social Science Research* 48: 121–134.
- [9] Chen, Weiwei & Grove, Wayne & Hussey, Andrew (2012). *The Payoff to School Selectivity: An Application of Dale and Krueger’s Method to MBA Programs*. *Economics Letters* 116: 247–249.
- [10] Dale, Stacy & Krueger, Alan (2002). *Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables*. *Quarterly Journal of Economics* 117: 1491–1527.
- [11] Dale, Stacy & Krueger, Alan (2011). *Estimating the Return to College Selectivity Over the Career Using Administrative Earnings Data*. National Bureau of Economic Research, Working Paper No. 17159
- [12] DIA-valintaoppaat, 2000–2004.
- [13] Eliasson, Kent (2006). *College Choice and Earnings Among University Graduates in Sweden*. Umeå Economic Studies No. 693, Umeå University.

- [14] Heckman, James & Robb, Richard (1985). *Alternative Methods for Estimating The Impact of Interventions*. Journal of Econometrics 30: 239–26.
- [15] Hershbein, Brad (2013). *Worker Signals Among New College Graduates: The Role of Selectivity and GPA*. Upjohn Institute Working Paper: 13–190.
- [16] Hoekstra, Mark (2009). *The effect of attending the flagship state university on earnings: A discontinuity-based approach*. The Review of Economics and Statistics, 91(4): 717–724.
- [17] Holmlund, Linda (2009). *The effect of College quality on Earnings: Evidence from Sweden*. Umeå Economic Studies No. 781, Umeå University.
- [18] Hungerford, Thomas & Solon, Gary (1987). *Sheepskin Effects in the Returns to Education..* Review of Economics and Statistics 69 (February 1987): 175–77.
- [19] Hämäläinen, Ulla & Uusitalo Roope (2008). *Signalling or Human Capital: Evidence from the Finnish Polytechnic Reform*. Scandinavian Journal of economics 110: 755–775.
- [20] Imbens, Guido & Kalyanaraman, Karthik (2012). *Optimal bandwidth choice for the regression discontinuity estimator*. The Review of Economic Studies 79(3): 933–959.
- [21] Imbens, Guido & Lemieux, Thomas (2007). *Regression Discontinuity Designs: Guide to practice*. National Bureau of Economic Research, Working Paper No. 13039
- [22] Kanninen, Ohto (2013). *Five Essays on Economics of Education*. European University Institute DOI: 10.2870/93297: 61–88.
- [23] Kirkeboen, Lars & Leuven, Edwin & Mogstad, Magne (2015). *Field of study, earnings, and self-selection*. Julkaisematon tutkimus.
- [24] Kuh, George & Pascarella, Ernest (2004). *What Does Institutional Selectivity Tell Us About Educational Quality*. Change 36: 52–58.
- [25] Lee, David (2008). *Randomized experiments from non-random selection in U.S. house elections*. Journal of Econometrics, 142(2): 675–697.
- [26] Lee, David & Lemieux, Thomas (2010). *Regression Discontinuity Designs in Economics*. Journal of Economic Literature 48: 281–355.
- [27] Lang, Kevin & Siniver, Erez (2011). *Why is Elite Undergraduate Education Valuable? Evidence from Israel*. Labour Economics, 18: 767–777.

- [28] Long, Mark (2008). *College Quality and Early Adult Outcomes*. *Economics of Education Review*, 27: 588–602.
- [29] McCrary, Joshua (2008). *Manipulation of the running variable in the regression discontinuity design; A density test*. *Journal of econometrics*, 142(2): 698–714.
- [30] Meriläinen, Jaakko (2013). *Do Single-Party and Coalition Governments Differ in Economic Outcomes? Evidence from Finnish Municipalities*. VATT Working Papers, 51/2013.
- [31] Monks, James (2000). *The Returns to Individual and College Characteristics: Evidence from the National Longitudinal Survey of Youth*. *Economics of Education Review*, 19: 279–289.
- [32] Murnane, Richard & Willet, John (2008). *Methods Matter: Improving Causal Inference in Education and Social Science Research*. Oxford University Press: New York: 310–328.
- [33] Saavedra, Juan (2009). *The Learning and Early Labor Market Effects of College Quality: A Regression Discontinuity Analysis*. Julkaisematon tutkimus.
- [34] Schultz, Theodore 1961. *Investment in Human Capital*. *American Economic Review* 1: 1–17.
- [35] Staiger, Doug and Stock, James (1997). *Instrumental Variables Regression with Weak Instruments*. *Econometrica*, 65: 557–86.
- [36] Suhonen, Tuomo (2013). *Are There Returns from University Location in a State-Funded University System?*. *Regional Science and Urban Economics*, 43: 465–478.
- [37] Suhonen, Tuomo (2014). *Quality of Higher Education and Earnings: Evidence from Finland Using Field-of-Study-Level Quality Measures*. *International Review of Applied Economics*, 28: 22–44.
- [38] Suhonen, Tuomo (2015). *Korkeakoulun laadun vaikutus opiskelijoiden työmarkkinamenestykseen*. Opetus- ja kulttuuriministeriön julkaisuja 2015:13, 142–167.
- [39] Spence, Michael (1973). *Job Market Signaling*. *Quarterly Journal of Economics* 87: 355–374.
- [40] Tilastokeskus. *FLEED 1999-2012*.
- [41] Thistlethwaite, Donald & Campbell, Donald (1960). *Regression Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment*. *Journal of Educational Psychology*, 51:309–317.



- [42] UKÄ (2014). *Higher Education in Sweden. 2014 status report*. Swedish Higher Education Authority Report 2014:10, s.14.
- [43] VATT-työryhmä (2015). *Hyvän valtionosuusjärjestelmän periaatteet*. VATT Julkaisut 70.
- [44] Virtanen, Hanna (2016). *Failure in admission and the process of completing a post-compulsory degree*. Julkaisematon tutkimus.
- [45] Öckert, Björn (2009). *What's the value of an acceptance letter? Using admissions data to estimate the return to college*. *Economics of Education Review* 29: 504–516.

### WWW-sivut:

- [46] Nykänen (2014) *Teknillinen korkeakoulu*. Viitattu 22.03.2016. Saantitapa: <http://www.aalto.fi/fi/about/history/tkk/>.
- [47] Elinkeinoelämän Valtuuskunta (EVA) (2013). *Työttömyysaste alueittain 2013*. Viitattu 29.3.2016. Saantitapa: <http://www.eva.fi/tyotjatekijat/tyottomyysaste-alueittain-2013/>.
- [48] Tilastokeskus, a (2013). *Ylemmän korkeakoulututkinnon keskiarvitus aika (mediaani) tutkinnoittain 2001-2012, Liitetaulukko 4*. Viitattu: 29.3.2016. Saantitapa: [http://www.stat.fi/til/yop/2012/02/yop\\_2012\\_02\\_2013-06-19\\_tau\\_004\\_fi.html](http://www.stat.fi/til/yop/2012/02/yop_2012_02_2013-06-19_tau_004_fi.html)
- [49] Tilastokeskus, b (2013). *Rahanarvonkerroin 1860–2012*. Viitattu: 18.02.2016. Saantitapa: [http://www.stat.fi/til/khi/2012/khi\\_2012\\_2013-01-15\\_tau\\_001.html](http://www.stat.fi/til/khi/2012/khi_2012_2013-01-15_tau_001.html)
- [50] Tilastokeskus (2009). *Syntyneet 2008*. Viitattu: 29.3.2016. Saantitapa: [http://www.stat.fi/til/synt/2008/synt\\_2008\\_2009-04-22\\_fi.pdf](http://www.stat.fi/til/synt/2008/synt_2008_2009-04-22_fi.pdf)
- [51] Vipunen, tilastovuodet 2010-2014. *Yliopistojen opetus- ja tutkimushenkilökunta*. Viitattu 27.01.2016. Saantitapa: [https://vipunen.fi/fi-fi/\\_layouts/15/xlviewer.aspx?id=/fi-fi/Raportit/Yliopistojen%20opetus-%20ja%20tutkimushenkil%C3%B6kunta%20-%20yliopisto.xlsb](https://vipunen.fi/fi-fi/_layouts/15/xlviewer.aspx?id=/fi-fi/Raportit/Yliopistojen%20opetus-%20ja%20tutkimushenkil%C3%B6kunta%20-%20yliopisto.xlsb).
- [52] Vipunen, tilastovuodet 2010-2014. *Yliopistojen opiskelijat*. Viitattu 27.01.2016. Saantitapa: [https://vipunen.fi/fi-fi/\\_layouts/15/xlviewer.aspx?id=/fi-fi/Raportit/Yliopistokoulutuksen%20opiskelijat-n%C3%A4k%C3%B6kulma-yliopisto.xlsb](https://vipunen.fi/fi-fi/_layouts/15/xlviewer.aspx?id=/fi-fi/Raportit/Yliopistokoulutuksen%20opiskelijat-n%C3%A4k%C3%B6kulma-yliopisto.xlsb).

- [53] Ylioppilastutkintolautakunta. Ylioppilastutkinto Suomessa. *Tietoja tutkinnosta*. Viitattu 18.4.2016 Saantitapa: <https://www.ylioppilastutkinto.fi/fi/tilastot/tietoja-tutkinnosta>

# Liitteet

## A Lisätietoa DIA-valinnasta

### **DIA-valinnan arkkitehtivalinta:**

Arkkitehtivalinta oli vuosina 2000–2004 kolmivaiheinen: Hakijoiden oli tehtävä ennakkotehtäviä, jotka läpäistyään he pääsivät tekemään matematiikan kokeen. Ne jotka selviytyivät näistä kahdesta kokeesta, pääsivät tekemään piirustus- ja suunnittelukokeen. Arkkitehtivalinnassa valituksi saattoi tulla kahdessa eri valintaryhmässä. Valintaryhmässä 1 lopulliset pisteet, joiden avulla päätökset valinnoista tehtiin, muodostuivat alkupisteistä (maksimipistemäärä oli 15), valintakoe pisteistä sekä yhdestä lisäpisteestä. Valintaryhmässä 2 valittiin ei-ylioppilaita ja ylioppilaita pelkkien valintakokeiden perusteella, siten että ensin valittiin kynnysehdot täyttävät ei-ylioppilaat ja heidän jälkeensä ylioppilaita, jotka olivat alittaneet rajan edellisessä valintaryhmässä. Kynnysehtoina arkkitehtivalinnassa olivat: 0,4 x matematiikan kokeen maan keskiarvo, piirustus- ja suunnittelukokeessa 5,5/12, luonnontieteen kokeessa 1/3 x maan keskiarvo (maisema-arkkitechdeille). Tilanteessa, jossa useiden hakijoiden pisteet olivat kiintiöiden alarajalla samat, valinta tehtiin kokonaispisteiden (myös matematiikka) avulla. Jos hakijalla oli jo ennestään arkkitehdin opiskelupaikka, menetti hän tällöin lisäpisteen.

### **DIA-yhteisvalinnan ulkopuolelta hakeneet:**

DIA-hakukohteisiin tulivat suoraan valituiksi muun muassa Suomen Akatemian tiedekilpailussa palkitut heidän suoritettuaan hyväksytysti lukion oppimäärän ja yo-tutkinnon. Samoin lukion viimeisille luokille valintavuonna järjestetyn matematiikka-, fysiikka-, ja kemiakilpailujen toisen vaiheen kymmenen parasta hyväksyttiin DI-tutkintoon johtaviin koulutusohjelmiin teknillisiin yliopistoihin ja tiedekuntiin ilman valintakoetta. Myös EB-, IB- ja vuoteen 2002 asti myöskin RP-tutkinnon suorittaneet hakivat erillishaussa, jos olivat ylittäneet tietyn pisterajan. Joihinkin teknillisiin kohteisiin myös haettiin kokonaan DIA-valinnan ulkopuolella, esimerkiksi vuonna 2003 tällainen kohde oli Teknillisen korkeakoulun informaatioverkostojen koulutusohjelma. Tällaisia hakijoita on vuosittain vain muutamia ja lopullisessa aineistossakin heitä on vain 14.

Taulukko A.1: Fysiikan kurssien pisteytys

Fysiikan kurssien lkm	Arvosana				
	6	7	8	9	10
8	1	2	3	4	5
7		1	2	3	4
6			1	2	3

Taulukko A.2: Kemian kurssien pisteytys

Kemian kurssien lkm	Arvosana				
	6	7	8	9	10
4	1	2	3	4	5
3		1	2	3	4

Taulukko A.3: Ylioppilastodistuksen pisteytys

Oppiaine	Yo-todistuksen arv.					
	A	B	C	M	E	L
Äidinkieli			1	2	3	4
2.kotimainen kieli, pitkä			1	2	3	4
2.kotimainen kieli, lyhyt/keskipitkä				1	2	3
Vieras kieli, pitkä			1	2	3	4
Vieras kieli, lyhyt/keskipitkä				1	2	3
Matematiikka, pitkä	1	2	3	4	5	6
Matematiikka, lyhyt			1	4	3	4
Reaali			1	2	3	4

## B Analyyseissa käytettyjen muuttujien selitykset

*ln(tulot)*. Kertoo hakijan pääoma- ja ansiotulojen (euroissa) luonnollisen logaritmin kahdeksan vuoden kuluttua ensimmäisestä hakukerrasta.

*Tulot*. Kertoo hakijan pääoma- ja ansiotulot (euroissa) kahdeksan vuoden kuluttua ensimmäisestä hakukerrasta.

*Valmistumisvuosi*. Kertoo ensimmäisen vuoden jolloin hakija valmistui DIA-kohteista kandidiksi tai maisteriksi. Muuttuja on kategorinen (saa arvot 2003–2012).

*Valmistumiskoodi*. Kertoo ensimmäisen DIA-koulutusalan, jolta hakija valmistui kandidiksi tai maisteriksi. Muuttuja koostuu 34 kategoriasta.

*Hakukohde*. Kertoo mihin selektiivisen teknillisen yliopiston hakukohteseen hakija oli lähinnä tulla valituksi. Muuttuja koostuu 18 kategoriasta.

*Valintaryhmä*. Kertoo mihin selektiivisen yliopiston valintaryhmään hakija oli lähinnä tulla valituksi. Koostuu neljästä valintaryhmästä. Valintaryhmien muodostus esitetty luvussa 3.

*Hakuvuosi*. Kertoo ensimmäisen vuoden jolloin hakija oli mukana DIA-valinnassa. Muuttuja koostuu viidestä kategoriasta (2000–2004).

*Preferenssit*. Kyseinen muuttujaryhmä kertoo hakijan koulukohtaisen preferenssijärjestyksen. Jokainen maksimissaan viidestä preferenssistä koostuu seitsemästä eri koulusta (Aalto, Tampere, Lappeenranta, Oulu, Turku, Åbo ja Vaasa) ja saa arvon nolla jos kyseinen preferenssi puuttuu.

*Sijoitus*. Kertoo kuinka kaukana (sijoituksissa mitattuna) hakija oli sisäänpääsyrajasta. Muuttuja on muodostettu loppupisteiden avulla. Se saa epänegatiivisia arvoja, jos hakijan pisteet riittivät sisäänpääsyyn ja negatiivisia muutoin.

*Ikä*. Kertoo hakijan iän ensimmäisenä hakuvuotena (vuosina).

*Kieli*. Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakijan hakuhetken äidinkieli on suomi ja arvon nolla, jos se on jokin muu. Esiintyy taulukoissa myös nimityksellä suomen kieli.

*Sukupuoli*. Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakija on mies ja arvon nolla, jos hän on nainen. Esiintyy taulukoissa myös nimityksellä mies.

*Pääkaupunkiseutu*. Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakijan hakuhetken kotikunta on Espoo, Vantaa, Kauniainen tai Helsinki ja arvon nolla muutoin.

*Uusimaa*. Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakija asui Uudellamaalla tulojen havainnointihetkellä ja arvon nolla muutoin.

*Ylitti sisäänpääsyrajan*. Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakijan sijoitus oli epänegatiivinen ja arvon nolla muutoin.

*Vanhempien tulot.* Muuttuja kertoo hakijan vanhempien pääoma- ja ansiotulot hakuvuotta edeltävänä vuotena (eli 1999-2003). Tulot muutetaan vuoden 2003 tuloiksi käyttäen tilastokeskuksen rahanarvonkerrointa (joka on johdettu kuluttajahintaindeksistä), tällöin on tehty myös markkojen muunto euroihin. Taustamuuttuja-tarkastelussa muuttujaa käytetään jatkuvana, muissa yhteyksissä se jaetaan kahdeksaan kategoriaan: tuloja ei havaita; tulot alle 1000 euroa; tulot välillä [1000;2000]; tulot välillä [2000;3000]; tulot välillä [3000;4000]; tulot välillä [4000;5000]; tulot välillä [5000;6000]; tulot vähintään 6000 euroa.

*Pitkän matematiikan arvosana.* Kertoo hakijan pitkän matematiikan ylioppilasarvosanan hakuhetkellä. Koostuu seitsemästä kategoriasta: improbatur (I); approbatur (A); lubenter approbatur (B); cum laude approbatur (C); magna cum laude approbatur (M); Eximia cum laude approbatur (E); laudatur (L).

*Matematiikan arvosana.* Kertoo hakijan pitkän tai lyhyen matematiikan ylioppilasarvosanan. Koostuu 13 kategoriasta; ei matematiikkaa; lyhyen matematiikan approbatur (A); lyhyen matematiikan lubenter approbatur (B); lyhyen matematiikan cum laude approbatur (C); lyhyen matematiikan magna cum laude approbatur (M); lyhyen matematiikan eximia cum laude approbatur (E); lyhyen matematiikan laudatur (L); pitkän matematiikan approbatur (A); pitkän matematiikan lubenter approbatur (B); pitkän matematiikan cum laude approbatur (C); pitkän matematiikan magna cum laude approbatur (M); pitkän matematiikan eximia cum laude approbatur (E); pitkän matematiikan laudatur (L).

*Äidinkielen ylioppilasarvosana.* Kertoo hakijan äidinkielen ylioppilasarvosanan hakuhetkellä. Koostuu seitsemästä kategoriasta: improbatur (I); approbatur (A); lubenter approbatur (B); cum laude approbatur (C); magna cum laude approbatur (M); Eximia cum laude approbatur (E); laudatur (L).

*Äidin/isän päätoimi.* Muuttujat on muodostettu hakua edeltävän vuoden pisimmän työsuhteen perusteella. Otoksesta havaitaan kuusi kategoriaa: työllinen; työtön; opiskelija; eläkeläinen; muu työvoiman ulkopuolella oleva; tieto puuttuu.

*Äidin/isän tuorein tutkinto.* Kertovat äidin/isän tuoreimman tutkinnon tason hakuvuotta edeltävänä vuonna. Ne muodostuvat 13-kategoriasta: yleisivistävä koulutus; muu keskiasteen koulutus; alin korkea-aste; korkeamman asteen (kandin tai maisterin) tutkinto kasvatustieteistä; korkeamman asteen tutkinto humanistisella alalta (mukaan lukien taideala); korkeamman asteen tutkinto yhteiskuntatieteiden alalta (mukaan lukien kaupallinen ala); korkeamman asteen tutkinto luonnontieteiden alalta; korkeamman asteen tutkinto teknilliseltä alalta; korkeamman asteen tutkinto maa- ja metsätalousalalta; korkeamman asteen tutkinto terveys- ja sosiaialialalta; korkeamman asteen tutkinto palvelualalta; tutkijakoulutusasteen tutkinto; tieto puuttuu.

*Äiti/isä työtön.* Kyseessä on kaksi dummy-muuttujaa, jotka saavat arvon yksi, jos hakijan äiti/isä oli työtön vuotta ennen ensimmäistä hakukertaa ja arvon nolla muutoin.

*Äiti/isä korkeakoulutettu.* Kyseessä on kaksi dummy-muuttujaa, jotka saavat arvon yksi, jos hakijan äiti/isä oli suorittanut korkeamman asteen tutkinnon vuotta ennen ensimmäistä hakukertaa ja muulloin arvon nolla.

*Työkuukaudet havaintovuonna.* Kertoo työssäolokuukausien määrän tulojen havaintohetkellä (saa arvot 1–12).

*Opintojen aikaiset työkuukaudet.* Kertoo opintojen aikaisten työssäolokuukausien määrän (saa arvot 0–108).

*Valmistumisen nopeus.* Kertoo kandidiksi tai maisteriksi valmistumisen nopeuden teknillisestä valmistumisvuoden ja aloitusvuoden erotuksena (saa arvot 2–8).

*Tulojen havaitsemistodennäköisyys.* Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakijalle havaitaan  $\ln(\text{tulot})$ , muutoin se saa arvon nolla.

*Valmistui selektiivisestä yliopistosta.* Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakija valmistui selektiivisestä teknillisestä yliopistosta kandidaatiksi/maisteriksi kahdeksan vuoden kuluessa ensimmäisestä hakukerrasta. Muutoin muuttuja saa arvon nolla.

*Valmistuminen.* Kyseessä on dummy-muuttuja, joka saa arvon yksi, jos hakija valmistui mistä tahansa teknillisestä yliopistosta ja arvon nolla muutoin.

## C Lisätaulukoita ja -kuvioita

Taulukko C.1: Otoksen tunnusluvut.

Muuttuja	Keskiarvo/osuus	sd	N
Tulot	37923	15828	2875
Mies	0,761	0,427	2875
Ikä	19,106	0,368	2875
Suomen kieli	0,929	0,257	2875
Ylioppilas	0,985	0,121	2875
Pääkaupunkiseutu	0,251	0,434	2875
Pitkän matematiikan yo-arvosana	4,331	1,139	2820
Äidinkielen yo-arvosana	3,691	1,113	2832
Vanhempien tulot	48026	71919	2866
Äiti korkeakoulutettu	0,758	0,429	2875
Isä korkeakoulutettu	0,825	0,380	2875
Äiti työtön	0,031	0,172	2875
Isä työtön	0,028	0,165	2875

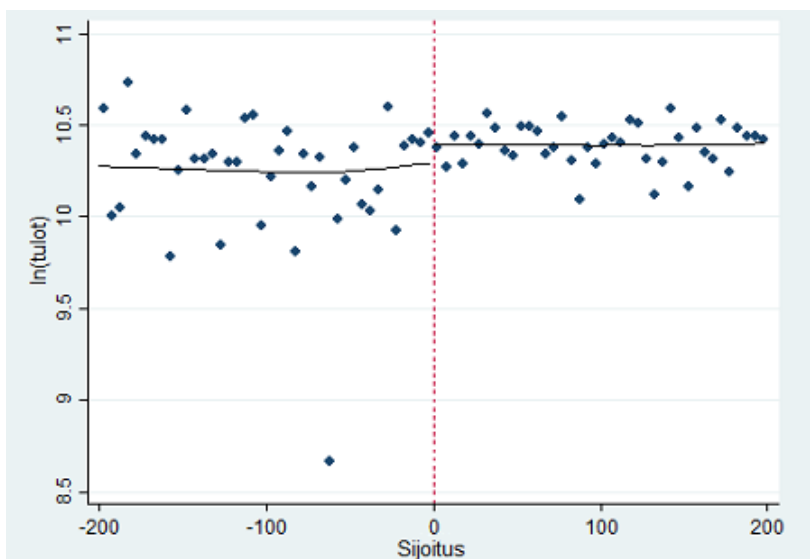
*Taulukossa on esitetty muuttujien keskiarvot tai osuudet. Sd tarkoittaa keskivirhettä ja N kertoo havaintojen määrän.*



Taulukko C.2: Tunnusluvut, eriteltynä niihin jotka jäivät sisäänpääsyrajan alle ja niihin jotka pääsivät sen yli.

Muuttuja	Alle_raján			Yli_raján		
	Keskiarvo/osuus	sd	N	Keskiarvo/osuus	sd	N
Tulot	34761	15327	411	38451	15851	2464
Mies	0,771	0,421	411	0,759	0,428	2464
Ikä	19,146	0,374	411	19,099	0,367	2464
Suomen kieli	0,912	0,283	411	0,932	0,252	2464
Ylioppilas	0,978	0,147	411	0,986	0,117	2464
Pääkaupunkiseutu	0,127	0,333	411	0,272	0,445	2464
Pitkän matematiikan yo-arvosana	3,451	1,068	395	4,475	1,084	2425
Äidinkielen yo-arvosana	3,303	1,012	402	3,755	1,116	2430
Vanhempien tulot	41209	68515	411	49167	72424	2455
Äiti korkeakoulutettu	0,657	0,475	411	0,774	0,418	2464
Isä korkeakoulutettu	0,759	0,428	411	0,836	0,370	2464
Äiti työtön	0,056	0,230	411	0,026	0,160	2464
Isä työtön	0,039	0,194	411	0,026	0,159	2464

*Taulukossa on esitetty muuttujien keskiarvot tai osuudet. Sd tarkoittaa keskivirhettä ja N kertoo havaintojen määrän.*



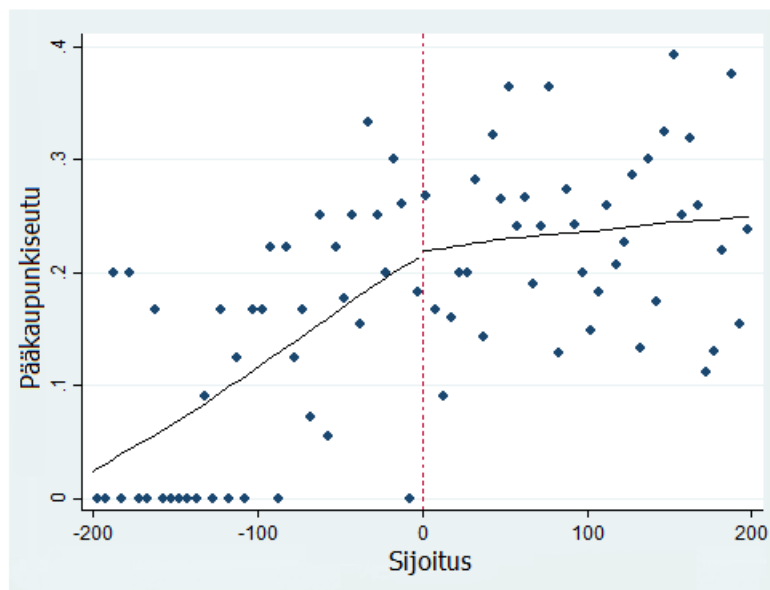
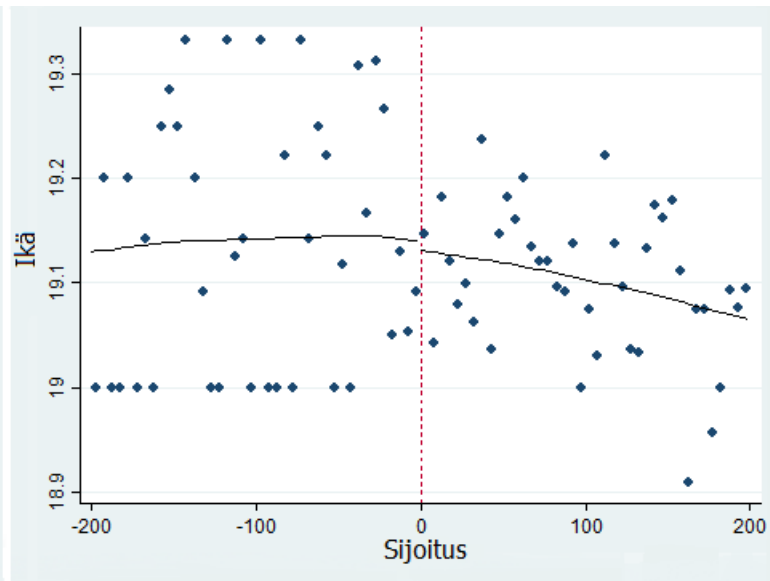
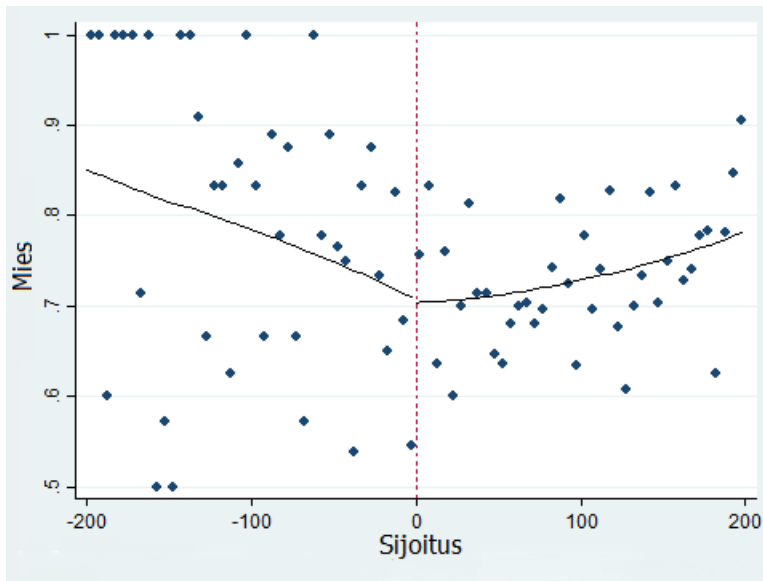
*Huomioita: Kuvassa jokainen piste kuvaa logaritmoitujen tulojen keskiarvoa viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä.*

Kuva C.1: Logaritmoidut tulot sisäänpääsyrajan ympäristössä.

Taulukko C.3: Sisäänpääsyrajan ylittämisen vaikutus alku-uran pääoma- ja ansiotuloihin optimihavaintovälillä käyttäen SRD:tä.

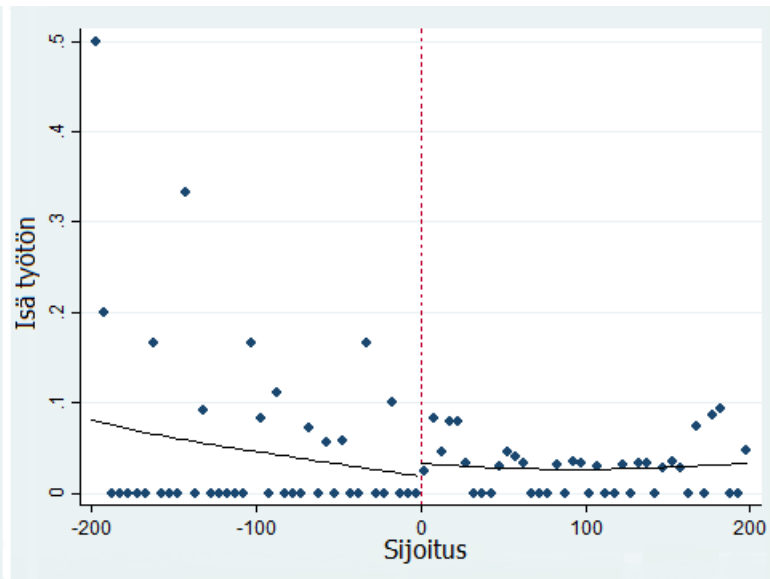
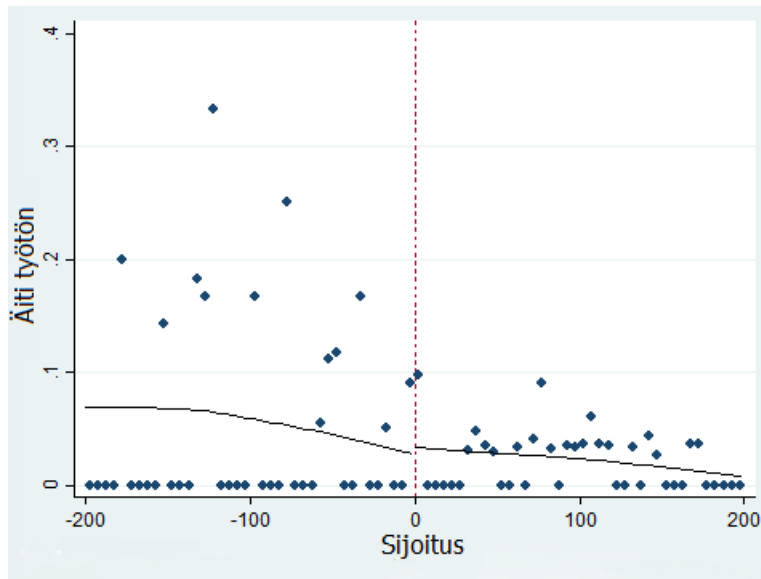
	ln(tulot)
Ylitti rajan	0,104 (0,123)
N	1705

*Robustit keskivirheet sulussa, \*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .  
Estimoinnissa ei käytetty kontroleita.*



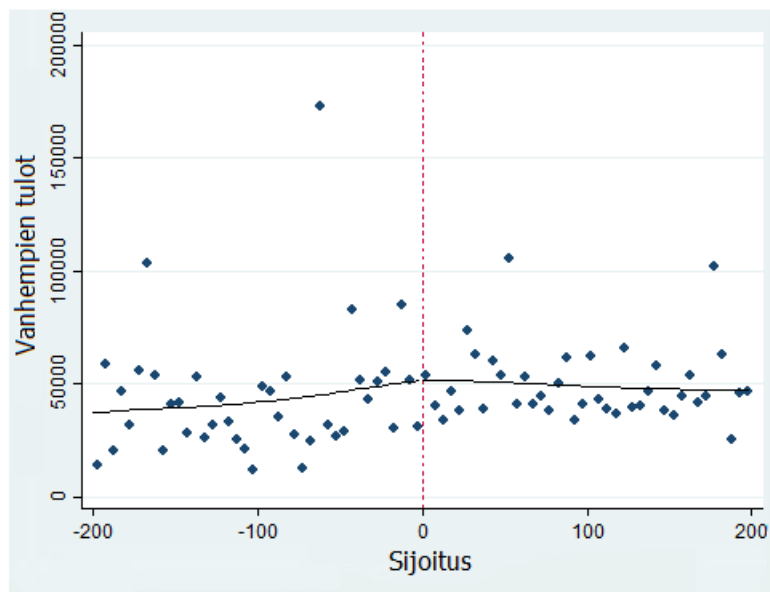
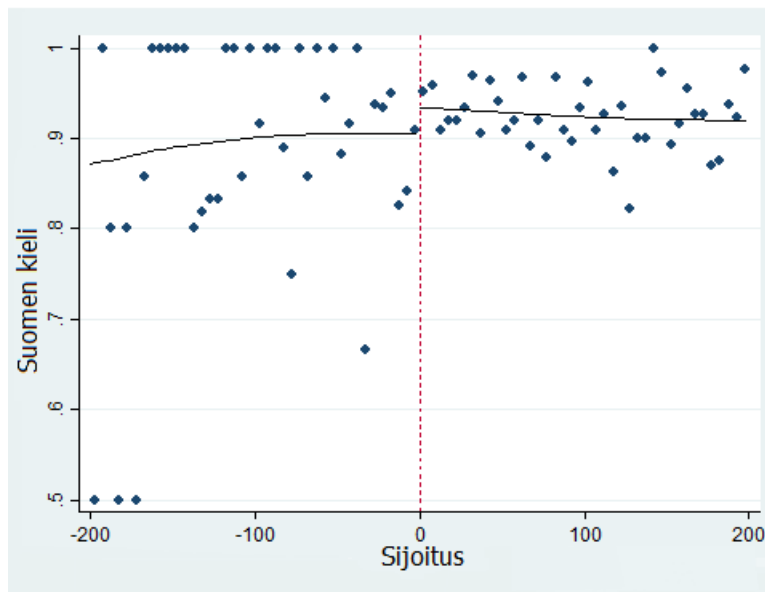
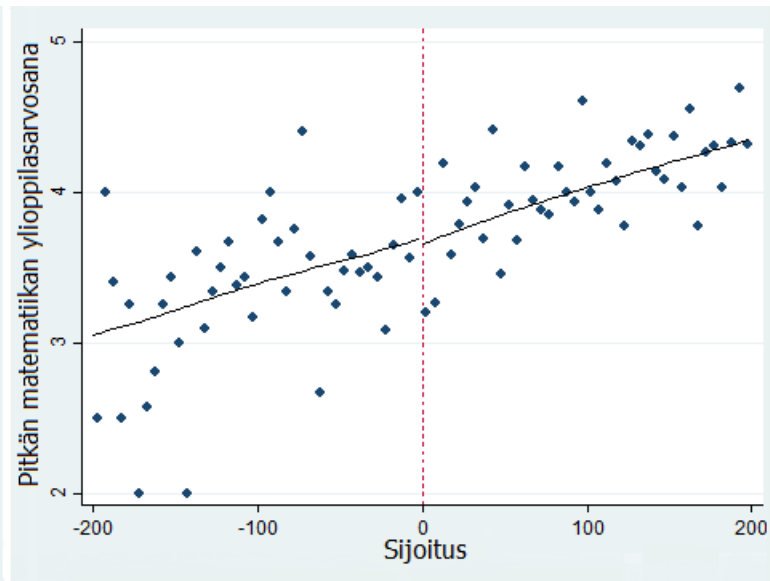
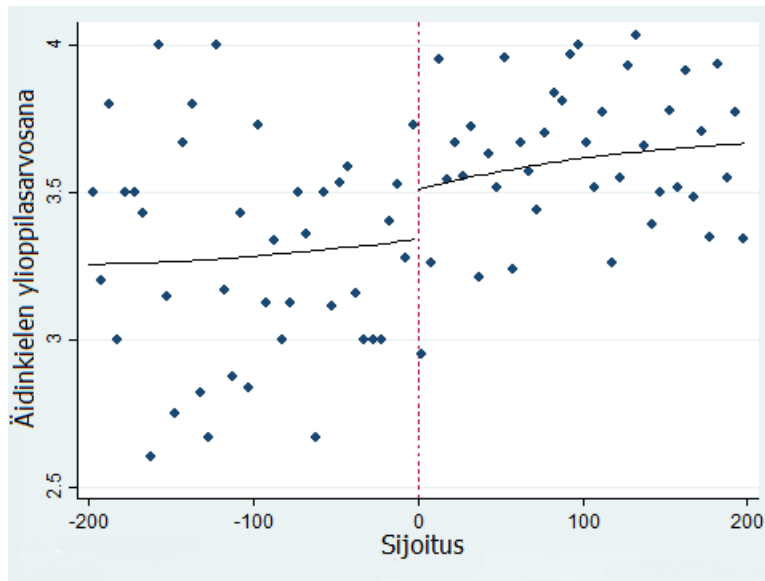
*Huomioita: Pisteparvissa jokainen piste kuvaa y-akselin muuttujan keskiarvoa viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä. Dummy-muuttujien tapauksessa Y-akselit kuvaavat nyt niissä mainittujen muuttujien osuutta (esim. miesten osuus).*

Kuva C.2: Taustamuuttujien graafiset tarkastelut, 1/3.



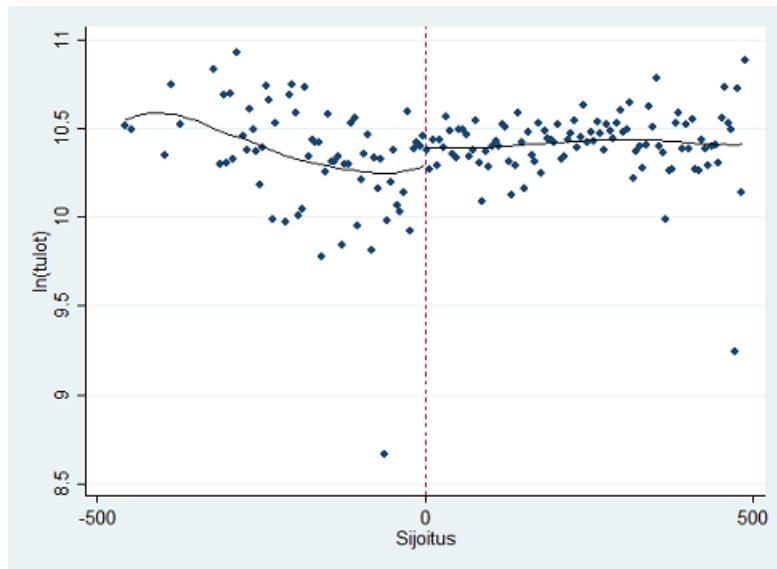
*Huomioita: Pisteparvissa jokainen piste kuvaa y-akselin muuttujan keskiarvoa viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä. Y-akselit kuvaavat nyt niissä mainittujen muuttujien osuutta (esim. niiden osuus joiden isä oli korkeakoulutettu).*

Kuva C.3: Taustamuuttujien graafiset tarkastelut, 2/3.



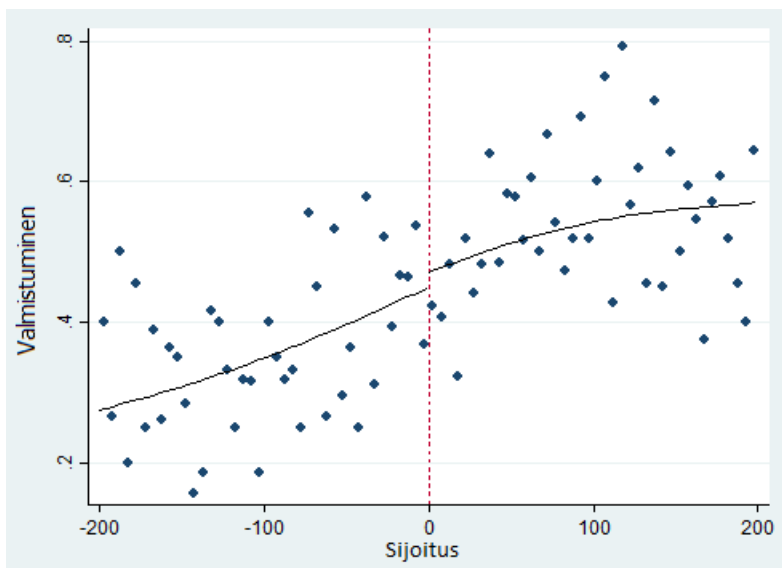
*Huomioita: Pisteparvissa jokainen piste kuvaa y-akselin muuttujan keskiarvoa viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä. Suomen kieli tarkoittaa tässä suomenkielisten osuutta.*

Kuva C.4: Taustamuuttujien graafiset tarkastelut, 3/3.



*Huomioita: Kuvassa jokainen piste kuvaa logaritmoitujen tulojen keskiarvoa viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä.*

Kuva C.5: Logaritmoidut tulot koko aineistossa.



*Huomioita: Kuvassa jokainen piste kuvaa keskimääräistä valmistuneiden osuutta viiden peräkkäisen sijoituksen klusterissa. Jatkuvat viivat kuvaavat lokaaleja lineaarisia regressiosuoria, jotka on painotettu käyttäen kolmio-kerneliä.*

Kuva C.6: Teknisestä yliopistosta valmistuneiden osuus sisäänpääsyrajalla. Otos rajattu niihin, jotka hakivat vähintään kahteen teknilliseen yliopistoon.