

# Lukioiden väliset erot ja paremmuusjärjestys

---

*Mika Kortelainen*

*Heikki Pursiainen*

*Jenni Pääkkönen*

# VATT TUTKIMUKSET

179

## Lukioiden väliset erot ja paremmuusjärjestys

Mika Kortelainen  
Heikki Pursiainen  
Jenni Pääkkönen

Kiitämme Tanja Kirjavaista, Antti Moisiota, Tuomas Pekkarista, Juhani Rautopuroa, Tuukka Saarimaata sekä Roope Uusitaloa hyödyllisistä kommenteista. Lisäksi haluamme kiittää Opetushallitusta sekä Ylioppilastutkintolautakuntaa oppilaskohtaisten aineistojen luovuttamisesta tutkimuskäyttöön. Mahdolliset virheet ovat omiamme.

ISBN 978-952-274-128-8 (PDF)

ISSN 1795-3340 (PDF)

Valtion taloudellinen tutkimuskeskus  
Government Institute for Economic Research  
Arkadiankatu 7, 00100 Helsinki, Finland

Edita Prima Oy  
Helsinki, marraskuu 2014

# Lukioiden väliset erot ja paremmuusjärjestys

Mika Kortelainen, Heikki Pursiainen ja Jenni Pääkkönen

19. marraskuuta 2014

Tutkimuksessa selvitetään Suomen päivälukioiden tuottamia arvonlisä ja arvonlisiin pohjautuvia paremmuusjärjestyksiä vuosille 2002-2013. Lukioiden välinen vertailu perustuu arvonlisämenetelmään, jossa kontrolloidaan oppilaiden lähtötaso ja koulujen erot oppilasaineksessa. Keskeinen ero aiempiin lukiovertailuihin on se, että arvioinnissa huomioidaan myös lähtötason kontrollointiin liittyvät ongelmat, kuten yläkoulujen toisistaan poikkeavat arviointikäytännöt. Lisäksi käytetyn menetelmän keskeinen ominaispiirre on ns. kutistusominaisuus (shrinkage), joka ottaa huomioon varsinkin pieniin lukioihin liittyvän epävarmuuden pienentämällä niiden arvonlisäarviota kohti keskiarvoa. Menetelmät, joissa ei ole tätä ominaisuutta, todennäköisesti yliarvioivat pienten lukioiden arvonlisien poikkeamista keskiarvosta. Tulosten mukaan parhaimpien ja huonoimpien lukioiden välillä on merkittäviä eroja, mutta toisaalta suuri enemmistö lukioista on sellaisia, että niitä ei voida mielekkäästi erottaa toisistaan arvonlisän perusteella. Lukioiden arvonlisissä havaitaan myös merkittävää ajallista vaihtelua, mikä vaikeuttaa huomattavasti lukioiden asettamista paremmuusjärjestykseen. Tuloksemme osoittavat, että julkisuudessa esitetyt ”lukiorankingit” ovat harhaanjohtavia, ainakin mikäli niiden tarkoituksena on asettaa lukioita paremmuusjärjestykseen laadun tai arvonlisän perusteella.

## 1. Johdanto

Joka kevät ja syksy kun uusien ylioppilaskirjoitusten tulokset julkaistaan, ilmestyy myös erilaisia lukiovertailuja. Seuratuimpia näistä lienevät Helsingin Sanomien (HS) ja Suomen Tietotoimiston (STT) tekemät vertailut [1, 12].<sup>1</sup> Näistä HS:n vertailu perustuu kunkin koulun keskimääräiseen menestykseen viimeisissä ylioppilaskirjoituksissa, joten se ei edes pyri huomioimaan lukioiden erilaista oppilasainesta. STT:n vertailussa pyritään ottamaan huomioon lukion ylioppilaiden lähtötaso vähentämällä keskimääräisistä ylioppilaiden arvosanoista sopivasti muunnettu peruskoulun päättötodistusten keskiarvojen

<sup>1</sup>Näiden lisäksi vertailuja tekevät myös MTV3, Iltasanomat ja Iltalehti.

keskiarvo. Sen enempiä HS:n kuin STT:n vertailu ei kuitenkaan perustu ylioppilaskirjoitusaineiston tilastolliseen analyysiin. On epäselvää, kuinka vakavasti yleisö suhtautuu näihin lukiovertailuihin. Kenties ne ovat pelkkää viihdettä, joka ei vaikuta koululaisten ja heidän vanhempiensa tekemiin lukiovalintoihin. Vertailujen suosio kuitenkin osoittaa, että lukioiden vertailu kiinnostaa yleisöä. Hyvälle ja luotettavalle lukioiden vertailulle olisi siis kysyntää.

Tässä tutkimuksessa selvitetään yhtäältä millaisia eroja lukioiden välillä on ja toisaalta sitä, onko lukiot mahdollista asettaa luotettavasti paremmuusjärjestykseen. Koulujen ja opettajien vertailusta on olemassa mittava tutkimuskirjallisuus (ks. esim. katsaukset [13, 29]), johon tämäkin tutkimus perustuu. Hyvältä lukiovertailulta vaaditaan tiettyjä ominaisuuksia. Ensiksi, oppilaiden lähtötaso täytyy kontrolloida tai vakioida tilastollisesti. Eri lukioiden oppilasaines vaihtelee merkittävästi, ja tämä täytyy ottaa huomioon arvioinnissa. Kirjallisuudessa lähtötason huomioon ottavista kouluvertailuista käytetäänkin yleensä nimitystä arvonlisämenetelmät (engl. *value-added methods*). Toiseksi lukion laatu voi muuttua vuosien kuluessa: aikaisemmin hyvät lukiot voivat huonontua ja vastaavasti huonot parantua. Lukion laatu ei kuitenkaan vaihtele holtittomasti yli ajan: opettajakunta pysyy samana, samoin hallintokäytännöt. Hyvä arviointimenetelmä ottaa huomioon molemmat seikat, sekä ajallisen muutoksen että pysyvyyden.

Kolmas tärkeä tekijä on epävarmuuden huomiointi. Monet lukiot ovat Suomessa varsin pieniä, mikä tarkoittaa, että niiden ylioppilaskirjoitustulokset vaihtelevat hyvinkin paljon vuodesta toiseen. Muutamalla oppilaalla voi olla suuri vaikutus vuoden tuloskeskiarvoon ja tästä syystä pienet koulut voivat olla yksittäisenä vuonna vertailun kärjessä tai hännillä pelkästään koulun pienen koon takia [16]. Luotettavassa arviointimenetelmässä tämä erityisesti pienempiä kouluja koskeva epävarmuus otetaan huomioon.

Tutkimuksessa tehdään äsken mainitut vaatimukset toteuttava vertailu Suomen päivälukioille vuosille 2002-2013. Vertailu perustuu artikkeleissa [6, 7] esitettyihin menetelmiin. Kullekin lukiolle lasketaan arvonlisä jokaiselle tarkasteluvuodelle. Laskettujen arvonlisien ominaisuuksia kuvaillaan varsin yksityiskohtaisesti ja erityisesti tarkastellaan niiden soveltuvuutta lukioiden asettamiseen paremmuusjärjestykseen. Saatuja tuloksia verrataan myös HS- ja STT-tyyppisten menetelmien tuottamiin tuloksiin.

Tulokset ovat kiinnostavia: lukioiden välillä on eroja. Kunakin vuonna parhaan ja huonoimman lukion välinen arvonlisäero on arvosanan luokkaa. Tämä tarkoittaa, että sama oppilas saisi parhaassa lukiossa arvosanan verran paremman yo-kirjoitusten keskiarvon kuin heikoimmassa. Ero on merkittävä. Mutta näin suuria eroja havaitaan vain parhaan ja heikoimman lukion välillä. Ääripäiden väliin jäävä lukioiden suuri enemmistö on sellaisia, että niitä ei voida mielekkäästi erottaa toisistaan arvonlisän perusteella. Niiden asettaminen paremmuusjärjestykseen ei siten ole tilastollisesti perusteltua. Lukioiden asettaminen paremmuusjärjestykseen muuttuu vielä hankalammaksi, kun tarkastellaan laadun muutosta. Peräkkäisten vuosien arvonlisät ovat voimakkaasti korreloituneita. Siksi pidemmällä aikavälilläkin laadussa on pysyvyyttä, vaikka muutosta tapahtuukin. Lukioiden paremmuusjärjestys on herkkä pienillekin muutoksille ja se heilahtelee vuosittain melko voimakkaasti. Vain muutamat lukiot pysyvät jatkuvasti joko kärjessä tai pohjalla. Vertailu HS- ja STT-tyyppiin menetelmiin osoittaa, että niiden tuottama paremmuusjärjestys on hyvin erilainen. STT:n menetelmä on myös huomattavan paljon epävakampi.

Raportissa kiinnitetään lisäksi huomiota yhteen kouluvertailujen keskeiseen ongelmaan, valikoitumiseen. Mikäli oppilaat valikoituvat lukioihin havaitsemattomien tekijöiden, kuten esimerkiksi opiskelumotivaation perusteella, parhaisiinkin menetelmiin perustuvat tulokset voivat olla virheellisiä. Tässä selvityksessä testattiin valikoitumista kahdella tavalla: ensiksi estimoituun malliin lisättiin yläkoulutason kiinteät vaikutukset. Niiden uskotaan kontrolloivan paitsi yläkoulujen laatueroja myös yläkoulujen erilaisia arvostelukäytäntöjä ja mahdollisia muita oppilaan taustaan liittyviä (tutkijalle) havaitsemattomia tekijöitä.

Näiden lisäksi testattiin sosioekonomisten muuttujien vaikutusta, koska oppilaat voivat valikoitua lukioihin lähtötason ohella myös perhetaustansa perusteella. Saadut tulokset osoittavat, että yläkoulutason kiinteiden vaikutusten huomioon ottaminen vaikuttaa tuloksiin merkittävästi. Sen sijaan aluetason sosioekonomiset muuttujat eivät selitä valikoitumista lainkaan silloin kun oppilaiden lähtötaso on kontrolloitu. Tämä mahdollisesti yllättävältä vaikuttava tulos voi johtua kahdesta syystä. Ensiksi, käytetyt muuttujat eivät ehkä kuvaa oppilaan sosioekonomista taustaa riittävän tarkasti. Toiseksi, mallissa jo mukana olevat muuttujat, kuten yläkoulun keskiarvo ja yläkoulutason kiinteät vaikutukset, voivat kontrolloida kaiken oleellisen oppilaan taustasta. Jälkimmäisessä tapauksessa kysymys on siitä, että oppilaan sosioekonominen tausta vaikuttaa peruskoulumenestykseen, jolloin se näkyy peruskoulun päättöarvosanassa. Yläkoulumuuttuja voi taas kuvata sosioekonomista taustaa silloin, kun oppilaan peruskoulu riippuu tämän perhetaustasta joko asuinpaikan valinnan tai vanhempien tekemän kouluvalinnan kautta.

Raportin rakenne on seuraavanlainen: luvussa 2 esitetään lyhyt katsaus arvonlisämenetelmiin ja kuvataan tutkimuksessa käytetyt menetelmät, ja luvussa 3 kuvataan käytetty tilastoaineisto. Varsinaiset tulokset esitellään luvussa 4 ja niiden luotettavuutta tutkitaan eri näkökulmista luvussa 5. Luvussa 6 keskustellaan tuloksista ja pohditaan niiden merkitystä. Viimeisessä luvussa esitetään lyhyesti johtopäätökset.

## 2. Menetelmät

### 2.1. Arvonlisämenetelmistä

Koulujen ja opettajien vertailusta on olemassa laaja kansainvälinen tutkimuskirjallisuus. Viime vuosina erityisesti opettajien vertailu on ollut suuren kiinnostuksena kohteena ainakin taloustieteilijöiden keskuudessa.<sup>2</sup> Koulujen ja opettajien arviointiin on käytetty lukuisia eri menetelmiä, mutta suurin osa vertailuista perustuu ns. arvonlisämenetelmiin (engl. *value-added methods*). Näiden menetelmien yleisenä tavoitteena on vertailla kouluja (tai opettajia) näiden aikaansaaman arvonlisän perusteella. Arvonlisällä tarkoitetaan karkeasti sitä lisähyötyä, jonka koulu oppilaalleen tuottaa. Lukioiden tapauksessa arvonlisäksi määritellään se osuus ylioppilaan opintomenestyksestä, joka voidaan kohdistaa nimenomaan lukiolle. Katsauksia kansainväliseen tutkimuskirjallisuuteen ja erityisesti

---

<sup>2</sup>Kiinnostus johtuu merkittävältä osalta Yhdysvalloissa käytävästä opettajien palkitsemiskeskustelusta, so. pitäisikö opettajien palkan ja työsuhteen keston riippua opettajan taidoista ja kuinka näitä taitoja tulisi arvioida.

arvonlisämenetelmiin tarjoavat esimerkiksi [13, 17, 19, 29].<sup>3</sup>

Kaikkien arvonlisämenetelmien yhteisenä pyrkimyksenä on vakioida mahdollisimman hyvin oppilaiden lähtötaso koulun (tai opettajan) vaikutuksen arvioimiseksi. Menetelmisä pyritään siis ottamaan huomioon se, että kouluihin valikoituu lähtökohdiltaan erilaisia oppilaita. Lähtötason vakioinnin lisäksi menetelmille yhteistä on pyrkimys ottaa huomioon arviointiin liittyvä epävarmuus. Koska opettajien ja koulujen arvonlisän arviointi perustuu kunakin vuonna tyypillisesti varsin pieneen otokseen oppilaita, arviointiin liittyy väistämättä merkittävää epävarmuutta. Tyypillinen tapa hallita tätä otoksesta johtuvaa epävarmuutta on käyttää ns. shrinkage-estimaattoreita [31, 16, 6]. Näissä epävarmoja arvonlisäarvioita “kutistetaan” kohti keskimääräistä arvonlisää. Kutistuksen määrä riippuu epävarmuuden määrästä: mitä suurempi epävarmuus, sitä suurempi kutistus. Suomen lukioiden vertailuissa *shrinkage*-ominaisuus on erityisen tärkeä, koska suuressa osassa lukiosta vuosittaisten kirjoittajien lukumäärä on pieni. Ilman *shrinkage*-estimointia pienet lukiot voidaan arvioida liian “hyviksi” tai “huonoiksi”, koska satunnaisvaihtelun merkitys korostuu pienen oppilasmäärän vuoksi [16, 6, 13].

Useat viime vuosien aikana julkaistut arvonlisämenetelmiä soveltavat tutkimukset ovat käsitelleet arvonlisäestimaattien pysyvyyttä tai muutoksia ajassa, ks. esim. [4, 6, 14, 9, 30]. Tarkemmin sanottuna kysymys on koulujen ja opettajien laatuerojen persistenssistä, eli pysyvyydestä. Perinteisissä arvonlisämenetelmissä koulujen / opettajien arvonlisä on oletettu ajassa vakioksi. Uudemmassa kirjallisuudessa tämän oletuksen paikkansapitävyyttä on tutkittu. Tuloksena on yleisesti, että vaikka laatueroissa on merkittävää ajallista pysyvyyttä, se ei suinkaan ole täydellistä, vaan laatuerot myös vaihtelevat ajassa.

Keskeinen arvonlisämenetelmiin liittyvä ongelma on valikoituminen. Mikäli oppilaat valikoituvat omasta tahdostaan tai kouluhallinnon päätöksestä opettajille / kouluihin tutkijalle havaitsemattomien ominaisuuksien perusteella, arvonlisämenetelmät voivat antaa virheellisiä tuloksia. Esimerkki tällaisesta mahdollisesti oppimistulokseen vaikuttavasta, mutta tutkijalle havaitsemattomasta ominaisuudesta on opiskelumotivaatio. Esimerkiksi Rothstein [33] pitää valikoitumisen aiheuttamaa harhaa merkittävänä virhelähteenä arvonlisien estimoinnissa. Mainittu artikkeli on tosin saanut melko paljon kritiikkiä ja uudemmat tutkimukset [8, 20, 22] ovat päätyneet päinvastaisiin tuloksiin. Valikoitumisen aiheuttamaa ongelmaa on pyritty kvantifioimaan myös satunnaiskokeiden avulla [18, 15]. Tulokset antavat jonkin verran näyttöä siitä, että arvonlisämenetelmät toimivat. Tämän selvityksen keskeisessä lähteessä [6] valikoitumisharhaa on pyritty haarukoimaan eri tavoilla, joista osaa käytetään myös tässä raportissa. Chetty ym. tutkimuksen yksi keskeisimmistä tuloksista on se, että merkittävää valikoitumisharhaa ei havaittu. Lisäksi ainakin opettajien arvonlisien tapauksessa on näyttöä siitä, että estimoidut arvonlisät ennustavat tulevaa työmarkkinamenestystä ja / tai jatko-opintoihin hakeutumista [7, 5]. Näin ollen arvonlisä näyttäisi mittaavan jotakin, jolla on todellista merkitystä.

Seuraavassa osaluvussa esitellään raportissa käytetty arvonlisämenetelmä. Menetelmä on varsin kehittynyt. Siinä otetaan huomioon esimerkiksi se, että lukion tuottama arvonlisä voi vaihdella vuodesta toiseen. On kuitenkin tärkeää pitää mielessä, että hienostu-

---

<sup>3</sup>Arvonlisämenetelmien lisäksi tässä raportissa hyödynnetään kotimaista lukioiden arviointikirjallisuutta. Tätä ei ole erityisen runsaasti, mutta esimerkkejä ovat [3, 21, 23].

neinkaan menetelmä ei voi täydellisesti ratkaista tärkeintä arvonlisämenetelmiin liittyvää ongelmaa eli oppilaiden valikoitumista. Valikoitumisongelma on kuvattu tarkemmin osaluvussa 2.3.

## 2.2. Laskentamalli

Tässä työssä käytetty menetelmä perustuu suureksi osaksi artikkeleihin [6] ja [7], joissa se on kuvattu huomattavasti tarkemmin. Lukija, joka ei ole kiinnostunut ekonometrisen mallintamisen teknisistä yksityiskohdista, voi lukea tämän luvun haluamallaan tarkkuudella. Väliin sitä ei kuitenkaan kannata jättää, sillä se sisältää tärkeitä tietoja käytetyn menetelmän vahvuuksista ja rajoituksista.

Kukin lukiolainen  $i$ , joka kirjoittaa ylioppilaaksi vuonna  $t$  on jossakin lukiossa  $j = j(i, t)$ . Ylioppilaskirjoitusten arvosana tai arvosanojen keskiarvo  $A_{i,t}^*$  määräytyy seuraavan mallin mukaisesti:

$$A_{i,t}^* = \mu_{j,t} + \beta' X_{i,t} + \varepsilon_{i,t}. \quad (1)$$

Tässä  $\mu_{j,t}$  on lukion  $j$  vaikutus vuonna  $t$ ,  $X_{i,t}$  on kokoelma vuonna  $t$  kirjoittaneen ylioppilaan  $i$  ominaisuuksia kuvaavia muuttujia joita on  $K$  kappaletta,  $\beta$  on näiden ominaisuuksien kerroinvektori ja  $\varepsilon_{i,t}$  ylioppilaan  $i$  idiosynkraattinen satunnaistermi.

Kiinnostavin osa yhtälöstä (1) on luonnollisesti lukiovaikutus  $\mu_{j,t}$ . Se kuvaa lukion aikaansaamaa parannusta tai huononnusta oppilaan menestyksessä ylioppilaskirjoituksissa. Vaikutusta voidaan kuvata seuraavan ajatusleikin avulla. Valitaan satunnaisotannalla joukko lukiolaisia ja arvotaan nämä satunnaisesti valitut henkilöt lukioihin  $A$  ja  $B$ . Tällöin näiden arvottujen henkilöiden ylioppilaskoemenestys on yhtä suuri kuin  $ero_{A,B,t} = \mu_{A,t} - \mu_{B,t}$ .

Esimerkiksi, jos lukion  $A$  todellinen lukiovaikutus vuonna 2012 on<sup>4</sup>  $\mu_{A,2012} = 0.4$  ja lukion  $B$   $\mu_{B,2012} = -0.5$ , niin se merkitsee, että satunnaisesti lukioihin arvottujen oppilaiden ylioppilaskirjoitusten keskiarvo olisi lukiossa  $A$  keskimäärin 0.9 arvosanaa parempi vuonna 2012. Lukiovaikutuksella 0 tarkoitetaan keskimääräisen lukion vaikutusta.

Lukiovaikutusten estimointi on mielekästä vain silloin, kun yhtälön (1) eri osat toteuttavat tietyt oletukset. Jotta yhtälöä (1) voidaan käyttää ennustamiseen, tarvitaan ensinnäkin seuraavat kaksi oletusta:

1. Kunkin lukion  $j$  lukiovaikutus  $\mu_{j,t}$  noudattaa samaa nollaodotusarvoista kovariansistationaarista prosessia ja on muutenkin hyvin käyttäytyvä.
2. Nollaodotusarvoisten satunnaistermien  $\varepsilon_{i,t}$  ja lukiovaikutusten yhteisjakauma säilyy ajassa samanlaisena<sup>5</sup>. Satunnaistermi ja lukiovaikutukset eivät ole korreloituneita lukioiden välillä.

<sup>4</sup>Tieteellistä käytäntöä noudattaen desimaalierottimenä käytetään raportissa pistettä.

<sup>5</sup>Tarkemmin sanottuna virhetermien varianssit ovat samoja kaikkina vuosina kaikissa kouluissa. Jos satunnaistermi on korreloituneita keskenään tai lukiovaikutuksen kanssa, nämä korrelaatiot ovat koko ajan kaikissa kouluissa samanlaisia ja korrelaation suuruus kahden eri vuoden välillä riippuu vain vuosien välillä kuluneen ajan pituudesta.



Oletuksilla 1 ja 2 on intuitiivinen tulkinta: Ne käsittelevät lukiovaikutuksen muuttumista ajassa. Ajassa tapahtuva vaihtelu voi esimerkiksi tarkoittaa sitä, että jonkin vuoden “huippulukion” vaikutus hiipuu ajan kuluessa lähemmäs lukioiden keskiarvoa. Yhtä hyvin jonakin vuonna heikosti menestyvä lukio voi ajan kuluessa parantaa laatuaan. Malli siis sallii tällaisen vaihtelun ajassa, ja se antaa tilastoaineiston kertoa kuinka nopeaa muutos on. Malli toimii yhtä hyvin silloin, kun lukioiden “hyvyys” tai “huonous” on ajassa erittäin pysyvää ja samat lukiot ovat menestyvät hyvin vuodesta toiseen kuin myös silloin, kun lukioiden laatu vaihtelee vuodesta toiseen.

Menetelmän käyttö kuitenkin edellyttää, että lukioiden laadunmuutosprosessi on hyvin käyttäytyvää. Toisin sanoen, minkään lukion laatu ei esimerkiksi voi romahtaa niin, että se pienenee koko ajan rajattomasti. Koska selitettävänä muuttujana on ylioppilaskirjoitusten arvosanat, joilla on luonnolliset rajat, tämä oletus on luonteva. Lisäksi mallin käyttö edellyttää, että laadunmuutosprosessi säilyy ajassa samanlaisena, eikä laadunvaihtelu esimerkiksi kiihdy tai hidastu tarkasteluperiodilla. Ei ole mitään erityistä syytä uskoa, etteikö tämä toteutuisi riittävällä tarkkuudella nyt tarkasteltavassa aineistossa.

Oletukset sallivat sen, että virhetermit ovat keskenään korreloituneita. Lisäksi lukiovaikutukset voivat korreloida virhetermien kanssa, jolloin korrelaatiot voivat kuvata esimerkiksi oppilaiden valikoitumista lukioihin. Valikoituminen aiheuttaa kuitenkin vaikeuksia lukiovaikutusten estimointiin. Pahimmillaan valikoituminen voi aiheuttaa lukiovaikutusten arvioimisen täysin virheellisesti. Valikoituminen on selvityksen tulosten kannalta aivan keskeinen, ja siksi siitä keskustellaan tarkemmin jäljempänä.

Nyt voidaan määritellä käytetty lukiovaikutuksen mittari eli arvonlisä. Sen laskeminen tapahtuu kolmessa vaiheessa.

Ensiksi estimoidaan regressiomalli, jossa selitetään ylioppilaskirjoitusten tuloksia  $A_{i,t}^*$  muuttujilla  $X_{i,t}$  ja lukiokohtaisilla kiinteillä vaikutuksilla<sup>6</sup>. Näin saadaan estimaatti  $\hat{\beta}$  kertoimille  $\beta$ . Kertoimien avulla voidaan laskea ylioppilaskirjoitusten residuaalisoitu eli jäännöskeskiarvo  $A_{i,t} = A_{i,t}^* - \hat{\beta}' X_{i,t}$ . Jos  $\hat{\beta}$  on estimoitu hyvin, niin likimain pätee  $A_{i,t} = \mu_{j,t} + \varepsilon_{i,t}$ . Residuaalisoidusta keskiarvosta hävitetään  $X$ -muuttujien vaikutus ja oletusten 1 ja 2 vallitessa jäännöskeskiarvon jakauma on varsin yksinkertainen.

Seuraavassa vaiheessa lasketaan jäännöskeskiarvojen keskiarvo kullekin lukiolle kunkin vuonna, mikä on  $\bar{A}_{j,t} = \frac{1}{n_{j,t}} \sum_{i=1}^{N_{j,t}} A_{i,t}$ . Tässä  $n_{j,t}$  on lukion  $j$  ylioppilaiden määrä vuonna  $t$  ja summaus kulkee yli kaikkien  $i$  joille  $j(i,t) = j$ . Tälle pätee likimain  $\bar{A}_{j,t} = \mu_{j,t} + \bar{\varepsilon}_{j,t}$ , jossa  $\bar{\varepsilon}_{j,t}$  on virhetermin keskiarvo lukiossa  $j$  vuonna  $t$ .

Nyt voidaan muodostaa lukiovaikutusta kuvaava arvonlisämittari. Artikkelin [6] suosittelema vaihtoehto on keskimääräisen jäännöskeskiarvon paras mahdollinen lineaarinen ennuste ehdolla aiemmat keskimääräiset jäännöskeskiarvot, eli  $\hat{\mu}_{j,t} = \hat{E}(\bar{A}_{j,t} | \bar{A}_{j,t-1}, \dots, \bar{A}_{j,t-s})$ .

Kaava on esitetty yksinkertaisuuden vuoksi tilanteessa, jossa lukion  $j$  tulosta vuonna

<sup>6</sup>Menetelmä tuottaa oikean arvion  $\beta$ :lle, vaikka lukiovaikutukset olisivat korreloituneet selittäjien  $X_{i,t}$  kanssa. Periaatteessa voisi olla, että myös lukiovaikutusten muutokset olisivat korreloituneita selittäjien kanssa, jolloin estimointi ei tuottaisi oikeita tuloksia. Tämä on kuitenkin epätodennäköistä, sillä edes kiinteiden lukiovaikutusten jättäminen pois ei juuri muuta seuraavassa vaiheessa laskettavia residuaalisoituja l. jäännöskeutuloja. Näin ollen kiinteät lukiovaikutukset ovat mitä luultavimmin aivan riittävä tapa huomioida lukioiden erot tässä residuaalisointivaiheessa.

$t$  ennustetaan  $s$  edellisellä vuodella<sup>7</sup>. Ennusteen laskentamenetelmä on kuvattu yksityiskohtaisemmin liitteessä A.

### 2.3. Valikoitumisongelma

Oletusten 1-2 vallitessa saatu estimaattori  $\hat{\mu}_{j,t}$  on siis paras lineaarinen ennuste  $\bar{A}_{j,t}$ :lle eli keskimääräiselle jäännöskeskisarvolle lukiossa  $j$  vuonna  $t$  ehdolla aiempien vuosien keskimääräiset keskiarvot. Mutta kiinnostuksen kohteena ei ole keskimääräinen jäännöskeskisarvo vaan lukiovaikutus. Oletukset 1-2 eivät riitä takaamaan sitä, että käytetty estimaattori on hyvä arvio todelliselle lukiovaikutukselle. Tähän tarvitaan huomattavasti vahvemmat oletukset.

Milloin lukiovaikutusestimaattori  $\hat{\mu}_{j,t}$  sitten estimoi todellista lukiovaikutusta? Seuraavat 2 keskeistä lisäoletusta ovat riittäviä:

3. Lukiovaikutus  $\mu_{j,t}$  ei ole korreloitunut satunnaistermien  $\varepsilon_{i,s}$  kanssa.
4. Satunnaistermit  $\varepsilon_{i,s}$  ovat korreloimattomia.

Jos nämä oletukset ovat voimassa aiempien lisäksi,

$$\hat{\mu}_{j,t} = \hat{E}(\bar{A}_{j,t} | \bar{A}_{j,t-1}, \dots, \bar{A}_{j,t-s}) = \hat{E}(\mu_{j,t} | \bar{A}_{j,t-1}, \dots, \bar{A}_{j,t-s}), \quad (2)$$

eli lukiovaikutusestimaatti on myös todellisen lukiovaikutuksen paras lineaarinen ennuste annetulla informaatiolla.

Näilläkin oletuksilla on intuitiivinen tulkinta: Ylioppilaat eivät saa valikoitua lukioihin havaitsemattomien ominaisuuksiensa perusteella. Suomessa oppilaat valikoituvat lukioihin koulumenestyksensä perusteella, mutta tämä ei ole ongelma koska käytetyissä malleissa kovariaattivektori  $X_{i,t}$  sisältää koulumenestystä kuvaavia keskiarvomuuttujia. Ongelma syntyy, jos oppilaat valikoituvat tiettyihin lukioihin sellaisten ylioppilaskirjoituksissa menestymiseen vaikuttavien tekijöiden perusteella, joita ei voida havaita tilastoaineistosta (esim. motivaatio).

Havainnollistetaan asiaa esimerkin avulla. Tarkastellaan samasta peruskoulusta lukioon  $A$  ja lukioon  $B$  meneviä oppilaita. Kuvitellaan, että kumpaankin lukioon menevät oppilaat ovat saaneet peruskoulussa keskimäärin samanlaisia arvosanoja. Lukio  $A$  on parempi näistä kahdesta lukiosta. Lukioon  $A$  valikoituu sen antaman paremman opetuksen vuoksi oppilaita, jotka ovat motivoituneempia. Sen sijaan lukioon  $B$  menevät oppilaat ovat vähemmän motivoituneita, ja menevät lukioon  $B$ , koska siellä pääsee helpommalla. Ylioppilaskirjoituksissa motivoituneemmat oppilaat menestyvät paremmin, jolloin lukion  $A$  tulokset ovat merkittävästi paremmat kuin lukion  $B$  tulokset. Tämä ei kuitenkaan ole kokonaan lukion  $A$  paremman opetuksen tms. ansiota, vaan johtuu osittain siitä, että sinne on valikoitunut motivoituneempia oppilaita. Mutta oppilaat *näyttävät* tutkijan näkökulmasta aivan samanlaisilta, joten lukiovaikutus tulee arvioitua virheellisesti, eli lukion  $A$  arvonnisa arvioidaan liian korkeaksi lukioon  $B$  verrattuna.

<sup>7</sup> Todellisissa laskelmissa kaikilta lukioilta ei ole saatavissa tietoja kaikilta vuosilta. Lisäksi ennusteet lasketaan paitsi viimeiselle vuodelle ehdolla edelliset, myös muille vuosille ehdolla kaikki muut saatavilla olevat, myös tulevat havainnot, kuten jäljempänä käy ilmi.

Yllä kuvatun lisäksi muunkinlaiset mekanismit voivat johtaa valikoitumisongelmaan. Valikoitumisen johdosta mallin avulla arvioitu lukiovaikutus sisältääkin lukion “todellisen” vaikutuksen lisäksi myös valikoitumisen vaikutuksen. Esimerkissä valikoituminen johtaa lukioden välisten erojen yliarviointiin. On yhtä mahdollista kuvitella valikoitumis-mekanismi, joka johtaa erojen aliarvioimiseen. Valikoitumisongelman mahdollisimman hyvä ratkaiseminen on edellytys sille, että lukiovaikutusarvionti on mielekäs. On luultavaa, että valikoitumisongelmasta on lähes mahdotonta päästä kokonaan eroon (ilman satunnaiskokeen järjestämistä), mutta sen tyydyttävä ratkaiseminen on välttämätöntä uskottavien lukiovaikutusarvioiden tuottamiseksi.

Oletukset 3-4 ovat varsin voimakkaat. Niitä voitaneen jonkin verran lieventää ilman että käytetty estimaattori menettää kaikki hyvät ominaisuutensa. Varsin voimakkaita oletuksia kuitenkin tarvitaan. Tämän tarkempi keskustelu käyttöoletuksista ei kuitenkaan ole tarkoituksenmukaista nyt käsillä olevassa raportissa.

Valikoitumisongelmaa on tässä selvityksessä lähestytty samaan tapaan kuin artikkeleissa [6]. Toisin sanoen, lukiovaikutusten estimointi / arvonlisien laskenta on suoritettu ikään kuin oletukset 3 ja 4 tai vastaavat olisivat voimassa. Näin saaduille vaikutusarvioille on sitten tehty joukko testejä, joilla on pyritty mittaamaan mahdollisen valikoitumisongelman suuruutta. Näitä testejä on kuvattu jäljempänä osaluvussa 2.6.

## 2.4. Vaihtoehtoiset mittarit

Äskeinen artikkeleista [6] ja [7] peräisin oleva arvonlisämittari käyttää vuoden  $t$  lukiovaikutuksen ennustamiseen informaatiota kaikilta muilta vuosilta kuin  $t$ . Tällainen ns. *jackknife*-tyyppinen estimaattori, joka ei sisällä vuoden  $t$  tietoja, soveltuu kyseisten tutkimusten tarkoituksiin. Artikkeleissa muodostettua estimaattia vuodelle  $t$  käytetään erilaisten vuoden  $t$  muuttujien ennustamiseen. Ennustettaessa on tärkeää, ettei vaikutusarviossa ole mukana kyseisen vuoden tietoja. Muussa tapauksessa ennusteyhtälön “kummallekin puolelle” tulee samoja estimointivirheitä ja ennusteyhtälö tuottaa virheellisiä arvioita lukion vaikutuksesta.

Tässä tutkimuksessa on tehty joitakin vastaavanlaisia ennustetarkasteluja kuin Chetty ym. [6] tekevät. Äskeisen osaluvun lukiovaikutusarvio soveltuu juuri näihin tarkasteluihin. Toisin kuin äsken mainitun artikkelin tarkoituksena, tämän raportin tavoitteena on kuvata lukioden välisiä eroja kunakin vuonna. Kun tarkoituksena ei ole käyttää lukioden vaikutusarvioita missään jatkotarkasteluissa, voidaan laskea vaikutusarvioita, jotka hyödyntävät kunakin vuonna myös kyseisen vuoden tietoja. Kun oletukset 1-4 ovat voimassa, voidaan laskea paras lineaarinen ennuste  $\tilde{\mu}_{j,t} = \hat{E}(\mu_{j,t} | \bar{A}_{j,t}, \dots, \bar{A}_{j,t-s})$ , toisin sanoen ennuste lukiovaikutukselle ehdolla kaikki käytössä oleva informaatio, myös tarkasteluvuotta  $t$  koskeva. Ennuste voidaan muodostaa samaan tapaan kuin aiempikin lukiovaikutusenuste muuttamalla hiukan ennusteessa esiintyvää kovarianssimatriisia ja -vektoria.

## 2.5. *Shrinkage*-ominaisuus

Jokainen raportissa käytetyistä estimaattoreista on ns. *shrinkage* -estimaattori, jossa residualisoituja keskiarvoja “kutistetaan” kohti nollaa. Kiinnostunut lukija voi tutustua

liitteen A yhtälöön (3) paremman intuition saamiseksi. Kutistamisen määrä riippuu käytettävissä olevan informaation määrästä: Mitä vähemmän ylioppilaita, sitä suurempaa on arviointiin liittyvä epävarmuus ja sitä voimakkaampaa kutistaminen. Jos kutistamista ei tehdä, erityisesti pienten lukioden lukiovaikutukset tulevat yli- tai aliarvioiduiksi ja ne heittelevät voimakkaasti vuodesta toiseen [16, 17]. Ilman kutistamista “parhaat” ja “huonoimmat” lukiot ovat aina pieniä lukioita, koska yksittäisten oppilaiden merkitys on niin suuri. Päättarkasteluissa käytetyillä estimaattoreilla  $\hat{\mu}_{j,t}$  ja  $\tilde{\mu}_{j,t}$  on tämä shrinkage-estimaattorien piirre, mutta lisäksi ne sallivat kouluvaikutuksen vaihtelun vuodesta toiseen.

## 2.6. Ennusteharha ja sen arviointi

Edellä todettiin, että valikoituminen voi aiheuttaa systemaattista virhettä estimointiin. Vaikka valikoitumisongelmaa ei olisi, arvonlisissä on luonnollisesti aina mukana satunnaisvaihtelun aiheuttamaa arviointivirhettä. Ilman valikoitumisongelmaa tämä virhe on kuitenkin satunnaista siinä mielessä, että arvonlisät antavat keskimäärin oikean kuvan lukiovaikutuksesta. Jos valikoituminen on ongelma, voi käydä niin, että arvonlisä ei edes keskimäärin anna oikeaa arviota lukiovaikutuksesta. Tällöin arvonlisän sanotaan olevan harhainen lukiovaikutuksen estimaatti. Harhan suuruus on se etäisyys, jolla systemaattisesti virheellinen arvonlisä keskimäärin poikkeaa oikeasta lukiovaikutuksesta.

Teknisesti harha johtuu siitä, etteivät oletukset 3 ja 4 ole voimassa. Harhaa koskevat tarkastelut voidaankin tulkita myös näiden oletusten paikkansa pitävyyden selvittämiseksi.

Valikoitumisen aiheuttaman harhan olemassaolon ja suuruuden arviointi ei ole helppoa. Tässä tutkimuksessa ei edes yritetä selvittää kaikkea valikoitumisen mahdollisesti aiheuttamaa harhaa, vaan keskitytään niin kutsuttuun ennusteharhaan (engl. *forecast bias*, [6]).

Ennusteharhan määritelmä liittyy jo esitettyyn satunnaistamisajatusleikkiin. Todelliset lukiovaikutukset kertovat, mitkä ovat oppilaiden keskimääräiset menestyserot silloin, kun oppilaat sijoitetaan lukioihin satunnaisesti. Jos arvioidut arvonlisät ennustaisivat oikein menestyseroja satunnaiskokeessa, niissä ei olisi ennusteharhaa.

Ennusteharhan puuttuminen takaa, että estimoidut arvonlisät ovat ennustemielessä oikeita. Jos jollakin toimenpiteellä saadaan lukion arvonlisä nousemaan 0.1 arvosanaa, ennuste todellisen opintomenestyksen kasvulle on 0.1. Ennusteharhaton arvonlisä on politiikkamielessä mielenkiintoinen, koska sen nostaminen keskimäärin parantaa oppimistuloksia saman verran. Ennusteharhan puuttuminen ei kuitenkaan takaa, että jokaisen lukion lukiovaikutus tulee keskimäärin oikein estimoiduksi. On periaatteessa mahdollista, että ennusteharhattomuus toteutuu, mutta silti lukiovaikutukset tulevat yli- tai aliarvioiduksi. Siksi estimoitujen lukiovaikutusten käyttämisessä lukioden arvioinnissa ja keskinäiseen paremmuusjärjestykseen asettamisessa on oltava varovainen.

Kuten motivaatioesimerkki osoittaa, valikoitumisen aiheuttaman harhan olemassaolon ja suuruuden arviointi ei ole helppoa. Motivaation mittaaminen suoraan on vaikeaa. Havaitsemattomat ominaisuudet ovat usein havaitsemattomia juuri siksi, että niiden mittaaminen on vaikeata.

Yksi mahdollinen ratkaisu on vakioida (siis sisällyttää laskelmiin) enemmän oppilas-kohtaisia ominaisuuksia. Tyypillisiä esimerkkejä tällaisista ominaisuuksista ovat oppilaan ja hänen vanhempiansa sosioekonomista asemaa kuvaavat muuttujat. Mitä enemmän havaittuja ominaisuuksia mallissa on, sitä vähemmän tilaa jää havaitsemattomille ominaisuuksille perustuvalla valikoitumisella. Esimerkiksi, jos motivaatio on korreloitunut sosioekonomisten tekijöiden kanssa, motivaation havaitsemattomuudesta johtuva ongelma pienenee, kun sosioekonomisia tekijöitä otetaan mukaan malliin.

Kuten jäljempänä todetaan, tätä tutkimusta varten oli kaikista ylioppilaista käytävissä vain yhteisvalinta-aineistossa olevia henkilötason muuttujia. Tämän lisäksi osa aineistosta sisälsi osoitteen, jonka perusteella pystyttiin yhdistämään nämä henkilöt Tilastokeskuksen ruututietokantaan. Tietokannasta saatiin tietoja vain asuinpaikan, eikä henkilöiden itsensä, keskimääräisistä sosioekonomisista muuttujista. Tätä on kuvattu tarkemmin aineiston kuvausluvussa. Kun jäljempänä puhutaan lisämuuttujista tai sosioekonomisista muuttujista, tarkoitetaan näitä muuttujia.

Kuinka sosioekonomisia muuttujia voidaan sitten käyttää ennusteharhan arvioimiseen? Äsken todettiin, että harha johtuu havaitsemattomiin tekijöihin perustuvasta valikoitumisesta. Ottamalla malliin mukaan uusia muuttujia, harha pienenee sikäli kun uudet muuttujat kuvaavat näitä aiemmin havaitsemattomia tekijöitä tai ovat edes korreloituneet näiden kanssa. Esimerkiksi jos sosioekonomiset tekijät ennustavat hyvin opiskelumotivaatiota, niiden ottaminen mukaan malliin pienentää harhaa. Ennusteharhan testaus perustuu juuri tähän. Jos uusien muuttujien lisääminen malliin muuttaa merkittävästi arvonlisäestimaatteja, viittaa se harhan olemassaoloon. Jos uusien muuttujien lisääminen malliin ei aiheuta merkittäviä muutoksia, harhan olemassaolosta ei saada viitteitä.

Ylläesitetty intuitio on mahdollista muuttaa tarkaksi ennusteharhan havaitsemisen tietyissä olosuhteissa mahdollistavaksi menetelmäksi. Menetelmä on esitetty liitteessä A. Sitä on sovellettu luvussa 5.

### 3. Aineisto

Tutkimuksessa käytetty aineisto on yhdistetty pääosin kahdesta lähteestä. Ensimmäinen on Ylioppilastutkintolautakunnalta saatu ylioppilaskokeiden arvosana-aineisto vuodesta 1990 vuoden 2013 kevään kirjoituksiin. Toinen on Opetushallituksen rekisteristä peräisin oleva yhteisvalinta-aineisto ajalta 1998-2012. Aineisto sisältää tiedot kaikista tuona aikana yhteisvalintaan osallistuneista. Aineistot yhdistettiin henkilötunnuksen perusteella. Tarkasteluperiodi haluttiin valita niin pitkäksi kuin mahdollista aineistojen kattavuus huomioiden.

Ylioppilastutkintolautakunnan aineistossa on tiedot kunkin kokelaan suorittamista kokeista, suoritusaajankohdasta ja arvosanasta / hylkäämisestä. Suoritettuja ainekohtaisia kokeita oli alkuperäisessä aineistossa noin 5 miljoonaa kappaletta. Lisäksi aineistossa on tieto kokelasajasta, so. onko kysymyksessä tavanomainen kokelas vai kokeen uusi ja jne. Aineistossa ei ole tietoa ylioppilastutkinnon myöntämisaajankohdasta eikä aineen pakollisuudesta. Ylioppilaskirjoituksia koskevat säännöt ovat lisäksi muuttuneet tarkasteluperiodilla. Näistä syistä ei ollut mahdollista havaita kunkin kokelaan tutkinnonsuo-

rittamisajankohtaa tarkasti. Sen sijaan jokaiselle kokelaalle muodostettiin laskennallinen tutkinnon suoritusajankohta ja keskiarvo aineistossa olevien tietojen perusteella.

Laskennallinen ajankohta ja keskiarvo muodostettiin seuraavasti. Ensiksi aineistosta poistettiin kaikki henkilöt, jotka olivat keskeyttäneet tutkinnon suorittamisen ja aloittaneet sen uudestaan. Näissä tapauksissa katsottiin vaikeaksi tulkita, mihin lukioon kokeilas tulisi kohdistaa. Lisäksi poistettiin henkilöt, joilla oli aineistossa ainekoodeja, joita ei pystytty kohdistamaan selväkielisiin aineisiin käytettävissä olevalla tiedolla.

Koska syksyn ylioppilaskirjoitukset todennäköisesti poikkeavat kevään kirjoituksista merkittävästi, tarkastelut suoritettiin vuositasolla eikä kirjoituskertatasolla. Tämän vuoksi kunkin kokelaan laskennalliseksi tutkinnon suorittamisajankohdaksi valittiin aina jokin kevät, riippumatta siitä, milloin henkilö todellisuudessa suoritti tutkinnon. Suoritusajankohdaksi valittiin se kevät, jolloin kokelas ensimmäisen kerran oli suorittanut hyväksyttävästi tutkintoa varten vaadittavan määrän (4 kpl) ainekohtaisia kokeita. Hyväksytyksi suoritukseksi tulkittiin sellainen suoritus, joka olisi hyväksytty yo-tutkinnon kompensaatiosääntöjen mukaisesti, mikäli nämä neljä olisivat olleet pakollisia aineita.

Ylioppilastutkinnon laskennalliseksi keskiarvoksi puolestaan määriteltiin laskennalliseen suoritusajankohtaan mennessä suoritettujen kokeiden keskiarvo. Arvosanat oli keskiarvon laskentaa varten muutettu numeerisiksi puoltoääniksi. Puoltoäänimuunnoksessa arvosana laudatur sai numeroarvon 7, eximia 6, jne. Arvosanaa approbatur vastaa 2 puoltoääntä ja improbatur 0 puoltoääntä.

Lähtötason vakioimista varten ylioppilaskoeaineistoon yhdistettiin Opetushallituksen yhteisvalinta-aineisto. Sen tietosisältö vaihtelee jonkin verran vuosittain. Kultakin vuodelta oli kuitenkin tieto kaikkien yhteisvalinnassa koulutuspaikkaa hakeneiden henkilötunnus ja lähtökoulu, so. se yläkoulu, josta hakija oli päättötodistuksen saanut. Lisäksi aineistossa oli lähes kaikilta lukioon hakeneilta tieto lukiovalinnan perusteena toimivasta peruskoulun päättötodistuksen lukuaineiden keskiarvosta<sup>8</sup>. Yhteisvalinta-aineistossa on keskiarvon lisäksi tiedot myös muista arvosanoista. Muita arvosanoja ei ole käytetty päätulosten laskemisessa, mutta ne ovat mukana eräissä täydentävissä laskelmissa.

Sama henkilö saattaa esiintyä yhteisvalinta-aineistossa useita kertoja, mikäli on osallistunut useampaan yhteisvalintaan. Tämä ei aiheuta ongelmia, jos henkilön tiedot eri kerroilla olivat samat. Joissakin tapauksissa tiedot eri hakukerroilla poikkeavat toisistaan. Tietoa lukion aloittamisvuodesta ei ollut saatavilla, joten hakutiedoista valittiin viimeisin. Pienehkölle määrälle henkilöitä jäi useita havaintoja vielä tämänkin jälkeen. Näille henkilöille valittiin se havainto, jossa hakukeskiarvo oli paras tai jos havaintoja oli useita tämänkin jälkeen, se havainnoista jossa oli eniten informaatiota hakijan ominaisuuksista.

Joihtakin vuosilta yhteisvalintatiedoissa oli myös osoite. Osoitetiedot yhdistettiin kartakoordinaatteihin. Koordinaattien avulla kokelaat, joiden osoitetieto oli saatavilla, yhdistettiin Tilastokeskuksen ruututietokantaan. Ruututietokannassa Suomi on jaettu  $250m \times 250m$  ruutuihin. Kustakin ruudusta on saatavilla tietoja laajasta valikoimasta väestöä ja alueen taloutta kuvaavia muuttujia. Koska osoitetieto oli saatavilla vain suhteellisen

---

<sup>8</sup>Keskiarvomuuttujan nimitys on ollut erilainen eri aikoina. Aineistosta ei käynyt ilmi, oliko myös keskiarvon laskutapa muuttunut. Asiaa selvitettiin laskemalla aineistosta keskiarvojen keskiarvoja eri vuosilta. Nimityksen muuttuessa vuosittainen keskiarvojen keskiarvo ei juuri muuttunut, joten päätettiin siihen, että päättötodistuksen keskiarvot ovat ajallisesti riittävän vertailukelpoisia.

Taulukko 1: Aineiston tunnuslukuja

	Keskiarvo	Keskihajonta	Min	Max	Havaintomäärä
Yo.kirj. lask. ka	4.1	0.4	2.4	5.9	4883
Perusk. ka	8.3	0.3	7.0	9.7	4883
Ruotsink.	0.1		0.0	1.0	4883
Ei suomen- tai ruotsink.	0.0		0.0	1.0	4883
Nainen	0.6		0.0	1.0	4883
Perusk. matem.	8.3	0.4	6.2	9.7	4883
Peruskoulun äidink.	8.4	0.3	6.8	9.7	4883
Lask.ylioppilaita	74.4	54.1	1.0	329.0	4883

pienelle osajoukolle, ruututietoja ei voitu käyttää varsinaisessa analyysissä, mutta niitä hyödynnettiin eräissä täydentävissä laskelmissa. Ruututietokannasta käytettiin seuraavia tietoja: työttömyysaste, omistusasuntojen osuus, väkiluku, korkeakoulutettujen osuus ja kotitalouksien mediaanitulo.

Ylioppilasaineisto ja yhteisvalinta-aineisto yhdistettiin henkilötunnuksen perusteella. Koska yhteishaun ja laskennallisen ylioppilaaksitulon välillä on yleensä vähintään 3 vuoden viive, yhdistäminen aloitettiin vuodesta 2002. Yhdistäminen onnistui hyvin. Vuosien 2002-2013 yhteensä noin 397500 laskennallisesta ylioppilaasta pystyttiin yhdistämään yhteisvalintatietoon noin 386000.

Lopullisen aineiston muodostamisessa tehtiin vielä joitakin rajauksia. Ensiksi, aikuis- ja iltalukioiden jätettiin pois aineistosta, koska niiden toiminta poikkeaa monella tavalla tavanomaisten päivälukioiden toiminnasta. Lukioita on myös lakkautettu ja yhdistetty tarkasteluperiodilla. Lakkautetut lukioiden otettiin mukaan aineistoon. Lukioiden yhdistymiset pyrittiin havaitsemaan mahdollisimman hyvin ja niiden edeltäjälukioita käsiteltiin ikään kuin osana suurempaa lukiota koko tarkasteluperiodin ajan. Tätä ei ollut aivan yksinkertaista toteuttaa, koska käytössä ei ollut kattavaa tietokantaa lakkautetuista ja yhdistetyistä lukioista eikä Tilastokeskuksen oppilaitosluetteloa ollut käytettävissä sähköisessä muodossa koko relevantilta periodilta. Ylioppilaslautakunta käyttää myös omaa oppilaitosnumeroa, joka poikkeaa Tilastokeskuksen oppilaitosnumerosta. Aikuislukioiden poistamisesta ja lukioiden yhdistämisestä käytettiin kumpaakin numerointia. Lisäksi aineistosta jätettiin pois ne henkilöt, joilta puuttui jokin olennainen tieto.

Lopullisessa aineistossa on mukana yli 360 000 laskennallista ylioppilasta, eli yli 90% laskennallisista ylioppilaista. Ruututietokantaan näistä pystyttiin yhdistämään noin 90 000.

Taulukossa 1 on esitetty aineistoa kuvailevia tunnuslukuja nille muuttujille, joita on käytetty arvonnäytteen estimointimalleissa. Tunnusluvut koskevat kunkin muuttujan kouluvuosi -keskiarvoja. Näitä koulu-vuosi -yhdistelmiä oli 4883.

## 4. Lukiovaikutusarviot

### 4.1. Lukiovaikutusten arviointi

Raportissa käytetään ilmaisua lukiovaikutus lukion aikaansaamasta todellisesta hyödystä opiskelijalle. Tätä lukiovaikutusta arvioidaan aineistosta lasketulla arvonlisällä. Näin olen ilmaissut “arvonlisä” ja “lukiovaikutusarvio” käytetään tässä tutkimuksessa tarkoittamaan samaa asiaa. Lukiovaikutusarvioita laskettiin kolmella eri tavalla. Kukin perustuu luvussa 2 esitettyihin menetelmiin. Kaikissa laskettiin ensin ylioppilaskohtaiset<sup>9</sup> jäännöskeskiarvot, sitten lukiokohtaiset kovarianssit ja näiden perusteella lukiovaikutusarviot. Ensimmäinen, tutkimuksen päätuloksissa käytetty lukiovaikutusarvio, on kaikkiin havaintovuosiin perustuva ennuste lukiovaikutukselle. Jos tekstissä puhutaan lukiovaikutusarvioista sen tarkemmin erittelemättä, kysymys on näistä.

Toinen keskeinen lukiovaikutusarvio on ns. *jackknife*-arvio, jossa kunkin vuoden lukiovaikutusta on ennustettu kaikkien muiden vuosien residualisoiduilla keskiarvoilla. Kuten luvussa 2 todettiin, tämä arvio soveltuu erilaisiin ennustetarkasteluihin, toisin kuin ensimmäinen lukiovaikutusarvio. Näitä *jackknife*-arvonlisiä käytetään erityisesti luvussa 5.

Ensiksi laskettiin ylioppilaskohtainen residualisoitu eli jäännöskeskiarvo, joka voidaan tulkita henkilön lähtötasovakioiduksi ylioppilaskirjoitusten tulokseksi. Perusmallissa jäännöskeskiarvo laskettiin mallista, jossa oli mukana lukiokohtaiset ns. kiinteät vaikutukset (dummy-muuttujat lukioille) ja selittävinä muuttujina lisäksi kolmannen kertaluvun polynomi peruskoulun lukuaineiden keskiarvosta, äidinkieli (suomi, ruotsi, muu), sukupuoli sekä lähtökoulua kuvaava muuttuja. Lisäksi mallissa oli mukana kirjoituskerta- / vuosimuuttuja<sup>10</sup>.

Lähtökoulun kontrolloimiselle on useita syitä. Ensinnäkin useat suomalaiset tutkimukset ja selvitykset ovat osoittaneet että yläkoulujen arvostelukäytännöt poikkeavat systemaattisesti toisistaan (ks. esim.[32])<sup>11</sup>. Samalla osaamistasolla voi saada hyvin erilaisia arvosanoja riippuen alueesta ja koulusta. Toisin ilmaistuna, arvosana kahdeksan kertoo yksittäisen koulun oppilailtaan edellyttämästä vaatimustasosta, eikä se välttämättä heijasta opetussuunnitelman perusteissa määritellyjä hyvän osaamisen kriteerejä. Mikäli näin on, ei peruskoulun päättötodistuksen keskiarvo riitä vakioimaan ylioppilaiden lähtötasoa. Lähtökoulumuuttujan ottaminen mukaan malliin sallii tällaisen systemaattisen arvostelun tiukkuusvaihtelun. Toinen syy ottaa lähtökoulumuuttuja mukaan malliin on yritys vakioida ylioppilaiden sosioekonomista taustaa. Tutkimusta tehtäessä ei ole ollut käytössä yksilökohtaista tietoa ylioppilaiden taustasta, jolla saattaa kuitenkin olla vaikutusta opintomenestykseen.<sup>12</sup> Esimerkiksi, jos tiettyihin lukioihin valikoituu korkeasti

<sup>9</sup>Sanoja ylioppilas ja ylioppilaskirjoitusten keskiarvo käytetään koko ajan tarkoittamaan edellisissä luvuissa kuvattuja laskennallisia käsitteitä.

<sup>10</sup>Estimoimme sekä malleja, joissa ei ollut kontrolloitu kirjoituskertaa, että malleja, joissa mukana oli kirjoituskertaindikaattori kullekin kirjoituskerralle ja lineaarinen trendimuuttuja. Näillä ei ollut vaikutusta tuloksiin.

<sup>11</sup>Koulujen antamat perusopetuksen päättöarvosanat ja Opetushallituksen seuranta-arviointitulokset eivät vastaa toisiaan useissa oppiaineissa (esim.[24, 25, 26, 27, 28, 32]) .

<sup>12</sup>On tärkeää huomata että mallissamme taustatekijöillä (kuten perhetaustalla) on vaikutusta ainoastaan silloin kun ne eivät heijastu päättötodistuksen keskiarvoon, joka on vakioitu. Koska perheeseen



koulutettujen vanhempien lapsia, tämän huomiotta jättäminen voi aiheuttaa valikoitumisharhaa lukiovaikutusarvioihin (tarkemmin luku 2.3). Lähtökoulun vakiointi saattaa helpottaa tätä ongelmaa, jos valikoituminen on edes osittain tapahtunut jo ylä- tai peruskouluun hakeuduttaessa. Luvussa 5 esitetään laskelmat, joista havaitaan, että lähtökoulun jättäminen vakioimatta muuttaisi tuloksia todella merkittävästi.

Seuraavaksi jäännöskeskiarvoista laskettiin lukiokohtaiset keskiarvot kullekin vuodelle erikseen. Käyttäen jokaisen vähintään  $k$  vuotta aineistossa olevan lukion vuosikeskiarvoa, laskettiin arviot autokovariansseille  $\sigma_{A,k} = Cov(\bar{A}_{j,t}, \bar{A}_{j,t-k})$ . Laskennassa käytettiin painona kunkin vuoden yhteenlaskettua ylioppilaiden määrää. Nämä autokovarianssit ovat keskeisiä lukiovaikutusten laskennassa, kuten liitteen A menetelmäkuvauksesta ilmenee. Autokovarianssit sisältävät kuitenkin itsessäänkin kiinnostavaa tietoa lukiovaikutusten pysyvyydestä. Suoria autokovariansseja intuitiivisempi tapa tarkastella lukioiden laadun pysyvyyttä ovat niistä lasketut autokorrelaatiot.

Jos kaikki oletukset ovat voimassa, kuvan 1 autokorrelaatiofunktio kuvaa sitä, kuinka pysyvä lukiovaikutus on. Esimerkiksi ensimmäisen viipeen kerroin 0.63 tarkoittaa sitä, että vuoden päähän ennustettu lukiovaikutus on 0.63 kertaa tämän vuoden jäännöskeskiarvo. Toisin sanoen, jos lukio tänä vuonna on yhden keskihajonnan verran keskiarvon alapuolella, niin paras ennuste on, että ensi vuonna se on enää 0.63 keskihajontaa keskiarvon alapuolella. Vastaava ennuste 9 vuoden päähän on, että lukio on enää kolmasosa keskihajontaa keskiarvon alapuolella.

Lasketut autokorrelaatiot osoittavat, että lukiovaikutukset ovat osittain ajassa muuttuvia, mutta mukana on myös pysyvä komponentti, koska korrelaatio ei häviä kokonaan pitkäaikaisena ajan kuluessa.

Kiinnostavaa on myös tietää, kuinka suuri osuus jäännöskeskiarvojen vaihtelusta on lukioiden välistä ja kuinka suuri osa lukioiden sisäistä. Toisin sanoen kysymyksessä on sen arvionti, kuinka suuri osa koulumenestyksen eroista on lukioiden sisäisiä ja kuinka suuri osuus lukioiden välisiä. Lukioiden välisen vaihtelun luonteva mittari on lukiovaikutuksen keskihajonta.

Kun oletukset 1-4 ovat voimassa, on mahdollista arvioida lukiovaikutusten varianssia ja keskihajontaa. Sen varmistamiseksi, että tulos on oikean suuntainen, lukiovaikutusten keskihajontaa voidaan haarukoida myös muilla tavoin. Kahden peräkkäisen vuoden lukiovaikutusten kovarianssi on aina pienempi kuin lukiovaikutuksen varianssi, joten yksi tapa saada alaraja keskihajonnalle on laskea yhden viipeen autokovarianssin neliöjuuri. Kolmas tapa arvioida keskihajonta on ekstrapoloida sitä autokovarianssien avulla<sup>13</sup>. Tulokset näistä laskelmista on esitetty taulukossa 2.

Jäännöskeskiarvojen kokonaiskeskihajonta on noin 0.76. Yksilötason hajonta, eli lukioiden sisäinen hajonta on 0.74 ja lukiovaikutuksen hajonnan arviot 0.15 luokkaa. Tämä tarkoittaa, että jäännöskeskiarvojen vaihtelusta vain noin 5 % on lukioiden välistä ja 95 % lukioiden sisäistä vaihtelua<sup>14</sup>.

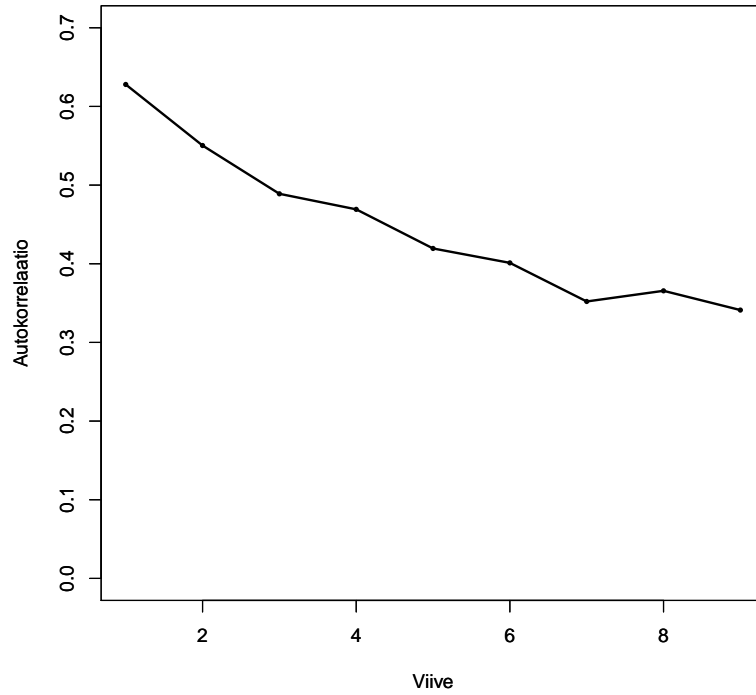
---

liittyvät taustatekijät yleensä vaikuttavat voimakkaasti opintomenestykseen peruskoulussa, niillä ei välttämättä ole omaa vaikutusta lukioiden oppimistuloksiin, kun lähtötaso kontrolloidaan.

<sup>13</sup> Autokovariansseille sovitettiin ajassa kvadraattinen malli. Arvio varianssille on mallin vakio.

<sup>14</sup> Osuudet vaihtelusta lasketaan yleensä varianssien perusteella. Luku 5 % on ns. sisäkorrelaatio (engl. *intra class correlation*), joka saadaan laskemalla lukiovaikutushajonnan neliön suhde kokonaiskeski-

Kuva 1: Autokorrelaatio



Taulukko 2: Lukiovaikutuksen keskihajonta

	Hajonta	1. viive	Kvadr.
Kokonaishajonta	0.76		
Lukioiden sisäinen	0.74		
Lukioiden välinen	0.16	0.14	0.15

Kovarianssien perusteella voidaan liitteen A menetelmillä laskea arviot arvonlisille. Arvonlisää muodostettaessa kovarianssit laskettiin yhdeksälle viipeelle. Loput viipeet asetettiin samansuuruisiksi kuin yhdeksäs, koska pidempien viipeiden kovarianssien arvio perustuisi vähille havainnoille. Päätuloksissa käytetyissä laskelmissa kunkin vuoden arvonlisä on ennuste tämän vuoden lukiovaikutukselle ehdolla kaikkien vuosien jäännöskeskiarvot, mukaanluettuna kyseinen vuosi. Arvonlisän laskemisessa tietylle vuodelle käytettiin paitsi aiempia myös myöhempiä vuosia. Laskelmat tehtiin vuosille 2002-2013. Mukaan otettiin lukio-vuosiyhdistelmät, joissa oli mukana vähintään kaksi ylioppilasta.

Ennustetarkoituksia varten laskettiin myös ns. *jackknife*-arvonlisä, jossa ennuste perustuu kaikille vuosille lukuunottamatta sitä, jota ennustettiin. Jackknife-vaikutus pystytettiin laskemaan vuosille 2002-2014, eli yhtä vuotta pidemmälle ajalle, kuin miltä aineistoa oli saatavilla.

Koko aineistoon perustuva arvonlisä ja *jackknife*-estimaatti tuottivat keskenään hyvin samanlaisia tuloksia. Joitakin eroja toki oli, mikä johtuu yksinkertaisesti siitä, että vaikutusarvioinnissa käytettiin eri informaatiota. Arvonlisäarvioiden korrelaatiokerroin oli yli 0.9. Koko aineistoon perustuvia vaikutuksia pystyttiin ennustamaan hyvin *jackknife*-vaikutuksilla, eikä ennustemalli antanut viitteitä siitä, että estimaatit olisivat systemaattisesti poikenneet toisistaan.

Jos tarkastellaan arvonlisän suuruuden sijasta lukioiden asettamista järjestykseen vuositain, kuva on monimutkaisempi. Vuotuisten sijoitusten Spearman-tyyppinen korrelaatio<sup>15</sup> on myös yli 0.9. Sen sijaan keskimääräinen vuotuinen ns. Kendall-tyyppinen järjestykskorrelaatio oli noin 0.7 eli merkittävästi alhaisempi. Tämä tulos sopii yhteen lukioiden ns. rankingia eli paremmuusjärjestykseen asettamista koskevien jäljempänä esitettävien tulosten kanssa. Näiden tulosten keskeinen sisältö on, että paremmuusjärjestykseen asettamisessa on syytä olla varovainen<sup>16</sup>.

## 4.2. Arvonlisien jakauma

Arvioitujen lukiovaikutusten eli arvonlisien jakauma on esitetty kuvassa 2. Paneelissa 2a on kuvattu kaikkien arvonlisien jakauma vuosina 2002-2013, kun taas paneelissa 2b on esitetty pelkästään vuoden 2013 arvonlisien jakauma. Vuotuiset tiedot arvonlisien maksimista, minimistä ala- ja yläkvartiilista sekä ala- ja ylädesiilistä on esitetty taulukossa 3. Aladesiilillä tarkoitetaan sitä lukua, jota pienempiä on 10 % havainnoista ja ylädesiilillä puolestaan lukua, jota suurempia on 10 % havainnoista. Vastaavasti alakvartiilia pienempiä ja yläkvartiilia suurempia havaintoja on 25 %.

Keskimääräisen lukion tuottama arvonlisä on 0. Negatiivinen arvonlisä tarkoittaa siis sitä, että lukio on keskimääräistä ”huonompi” ja positiivinen sitä, että lukio on keskimääräistä ”parempi”.

Lukioiden välillä on eroja. Ero heikoimpien ja parhaiden lukioiden välillä on kunakin vuonna noin arvosanan verran. Mikäli menetelmän käyttöoletukset ovat voimassa, par-

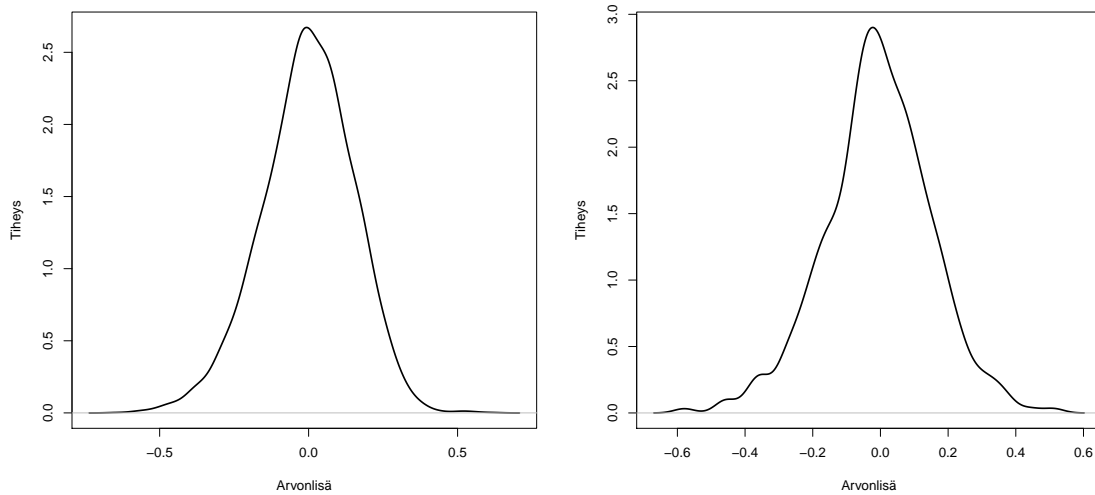
---

hajonnan neliöön.

<sup>15</sup>Kullekin lukiolle laskettiin sen sijoitus kunakin vuonna. Näistä laskettiin tavanomainen korrelaatiokerroin.

<sup>16</sup>Samantyyppisiä tuloksia on saatu esimerkiksi myös Englannista, ks. [34].

Kuva 2: Arvonlisien jakauma



(a) Arvonlisät 2002-2013

(b) Arvonlisät 2013

Taulukko 3: Arvonlisän jakaumatietoja

	2002	2003	2004	2005	2006	2007
Min	-0.50	-0.48	-0.50	-0.47	-0.52	-0.56
Alades	-0.24	-0.21	-0.19	-0.18	-0.21	-0.18
Alakv.	-0.12	-0.12	-0.09	-0.10	-0.10	-0.09
Yläkv.	0.08	0.11	0.13	0.10	0.09	0.09
Ylädes	0.17	0.19	0.20	0.20	0.16	0.18
Max	0.37	0.43	0.38	0.39	0.44	0.51
	2008	2009	2010	2011	2012	2013
Min	-0.65	-0.55	-0.59	-0.54	-0.56	-0.58
Alades	-0.19	-0.19	-0.20	-0.22	-0.22	-0.20
Alakv.	-0.08	-0.08	-0.10	-0.11	-0.10	-0.10
Yläkv.	0.11	0.11	0.09	0.10	0.09	0.09
Ylädes	0.19	0.21	0.20	0.19	0.17	0.18
Max	0.62	0.56	0.55	0.50	0.53	0.51

haassa lukiossa sattumanvaraisen oppilaan yo-kirjoitusten keskiarvo olisi arvosanaa parempi kuin heikoimmassa. Ero on merkittävä, kun muistetaan, että arvosanan parannus keskiarvossa tarkoittaa, että jokaisesta keskiarvoon laskettavasta aineesta täytyy saada keskimäärin yhden arvosanan suuruinen parannus.

Yksi arvosana kuvaa kuitenkin vain eroa ääripäiden välillä. Suurin osa lukioista on huomattavasti lähempänä toisiaan. Kunakin vuonna noin 80 % keskimmaisista lukioista on enintään 0.4 arvosanan päässä toisistaan. Keskimäinen 50 % lukioista on enää 0.2 arvosanan päässä toisistaan.

### 4.3. Lukioiden erot ja järjestys

Edellä esitetyt tulokset herättävät vähintään kaksi tärkeää kysymystä tulosten tilastollisesta merkitsevyydestä ja lukiovaikutusten pysyvyydestä. Koska suuri osa lukioista on niin lähellä toisiaan, herää kysymys voisivatko erot johtua kokonaan sattumasta ja estimointivirheestä systemaattisten erojen asemesta. Entä ovatko samat lukiot vuodesta toiseen jakauman ylä- ja alapäässä, vai vaihtuvatko sijoitukset joka vuosi?

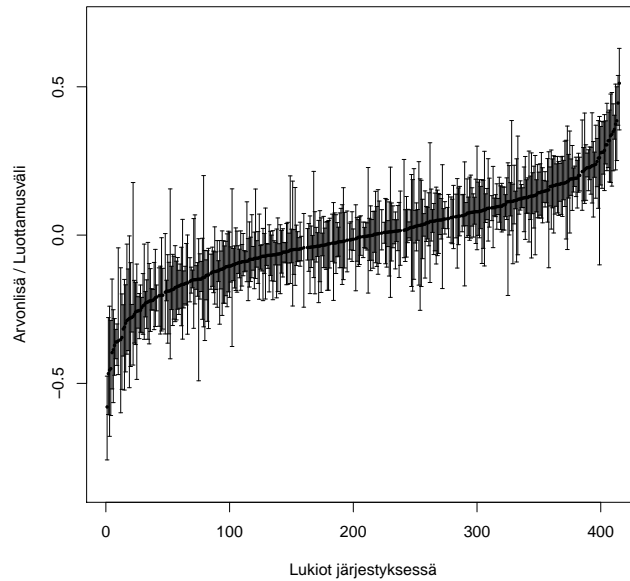
Aloitetaan ensimmäisestä kysymyksestä. Voivatko havaitut erot lukioiden välillä olla sattuman tuottamia? Tyypillisesti tilastollisessa tutkimuksessa tämääntyypisiin kysymyksiin vastataan suorittamalla tilastollisia testejä ja laskemalla sopivia luottamusvälejä. Näiden vertailujen tekeminen ei ole aivan suoraviivaista, kun vertailtavia lukioita on satoja jokaiselta vuodelta. Tämän vuoksi tarkastelu tehdään varsin epämuodollisesti. Kullekin lukiolle ja kullekin vuodelle laskettiin lukiovaikutuksen 95 % luottamusväli ns. *block bootstrap* -menetelmällä.

Lukioaineistosta poimittiin takaisinpanolla lukioita sama määrä kuin alkuperäisessä aineistossa oli. Koska otanta tapahtui takaisinpanolla, niin otokseen valikoitui osittain samoja lukioita useita kertoja, ja toisaalta osa lukioista jäi pois. Näin saadusta uudesta aineistosta laskettiin lukiovaikutukset ylläkuvatulla menetelmällä. Tätä toistettiin 1000 kertaa. Toisin sanoen menetelmä tuotti 1000 erilaista lukiovaikutusarviota kullekin vuodelle ja lukiolle (paitsi niinä vuosina kun kukin lukio jäi pois aineistosta). Näin saaduista arvioista valittiin 95 % luottamusvälin alarajaksi se luku, jonka alapuolelle jäi 2.5 % vaikutusarvioista ja ylärajaksi se luku, jonka yläpuolella oli 2.5 % arvioista.

Luottamusvälit on laskettu kaikki vuosille, mutta tulosten esitettävyyden vuoksi keskitymme pelkästään vuoteen 2013. Kyseisen vuoden luottamusvälejä havainnollistetaan kuvassa 3. kuvan vaaka-akselilla ovat lukiot lukiovaikutuksen mukaisessa paremmuusjärjestyksessä heikoimmasta (1) parhaaseen (n. 400). Pystyakselilla puolestaan ovat lukiovaikutusten luottamusvälit janoina. Monet luottamusväleistä ovat todella leveitä, eli niiden arvioon liittyy merkittävää epävarmuutta. Useilla lukioilla luottamusvälin alaraja on nollan alapuolella ja yläraja nollan yläpuolella. Nämä lukiot eivät tilastollisessa mielessä poikkea keskimääräisestä lukiosta. Niistä ei siis voida sanoa, olivatko ne arvonlisältään keskimääräistä huonompia vai parempia vuonna 2013.

Keskiarvosta eroamattomia lukioita oli vuonna 2013 yhteensä 181. Tilastollisesti keskimääräistä parempia lukioita oli 104 ja keskimääräistä huonompia oli 107. Luvut olivat samansuuntaisia kaikkina tarkasteluvuosina. Näin ollen minä tahansa vuonna todella merkittävästä osasta lukioita ei voida sanoa, ovatko ne keskiarvoa parempia vai huonom-

Kuva 3: Arvonlisät luottamusväleinen vuonna 2013



pia.

Kuvasta 3 havaitaan myös, että luottamusvälit ovat useilla kouluilla päällekkäisiä, vaikka koulut olisivat toisistaan varsin kaukana paremmuusjärjestyksessä. Luottamusvälien päällekkäisyys ei ole täysin oikea tapa tarkastella sitä, poikkeavatko lukiot toisistaan, sillä paremmuusjärjestyksen tilastolliset ominaisuudet ovat sangen monimutkaisia (ks. [10, 11]). Päällekkäisyystarkastelu antaa kuitenkin viitteitä siitä kuinka uskottavaa on, että kahden lukion arvonlisä pystytään erottamaan tilastollisesti toisistaan. Lukioiden, joiden luottamusvälit ovat osittain päällekkäiset, arvonlisät ovat tilastollisessa mielessä samat, ja niiden asettaminen paremmuusjärjestykseen ei ole mielekäästä.

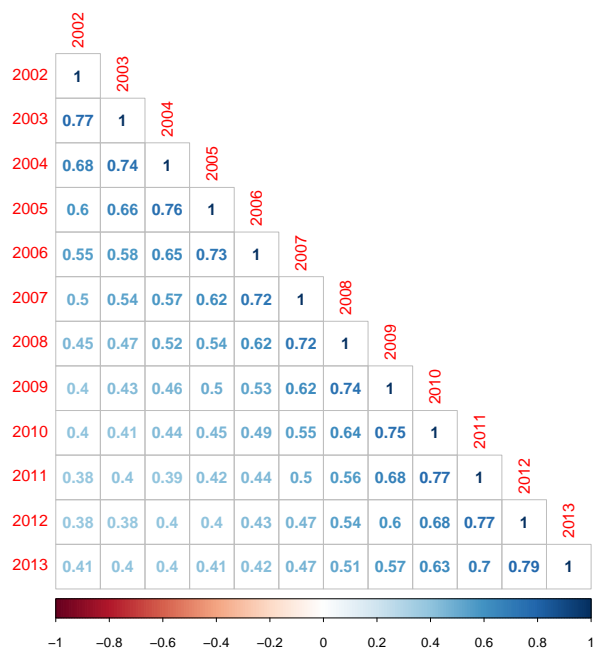
Lukioiden asettaminen yksiselitteiseen paremmuusjärjestykseen on mahdotonta. Tämä ei tarkoita sitä, etteikö joitakuita lukioita voisi asettaa järjestykseen luotettavasti: häntäpäin lukiot poikkeavat merkitsevästi parhaista lukioista.

#### 4.4. Arvonlisien ja järjestyksen pysyvyys

Lukioiden asettaminen yksiselitteiseen järjestykseen on siis vaikeaa, paitsi ääripäissä. Mielenkiintoinen kysymys onkin se, että ovatko nämä “parhaat”, “huonoimmat” ja “keskitasoiset” lukiot samoja vuodesta toiseen. Tähän liittyy kaksikin asiaa: pysyvätkö yhtäältä korkeat arvonlisät korkeina ja toisaalta korkeat sijoitukset paremmuusjärjestyksessä korkeina.

Osavastaus kysymykseen saatiin jo osaluvussa 4.1, jossa tarkasteltiin lukioiden jäännöskeskiarvojen autokorrelaatiota. Tarkasteluissa havaittiin, että jäännöskeskiarvoissa näyt-

Kuva 4: Arvonlisien korrelaatiomatriisi (Kendall)



täisi olevan pysyvä komponentti, mutta myös merkittävästi ajallista vaihtelua. Konkretisoidaan nyt hiukan näitä tuloksia.

Kenties suoraviivaisin tapa kuvata tulosten pysyvyyttä on laskea järjestyskorrelaatioita. Toisin sanoen lasketaan kuinka voimakkaasti aikaisempien vuosien sijoitus paremmuusjärjestyksessä on korreloitunut nykyisen sijoituksen kanssa.

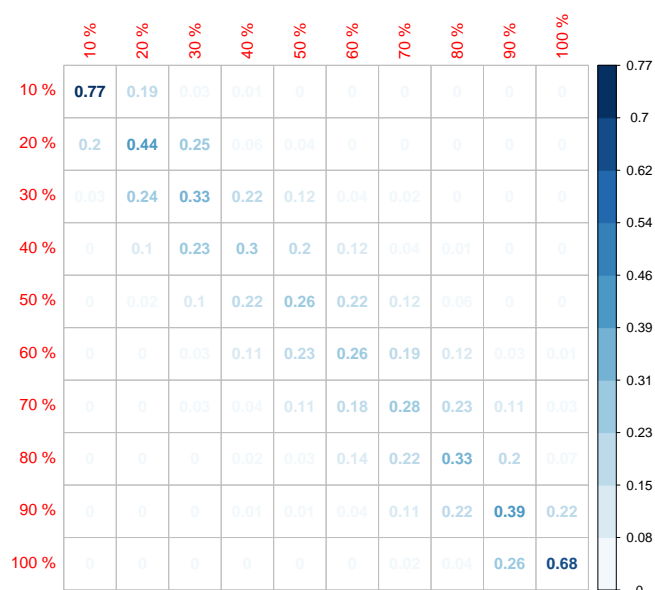
Kuvassa 4 on esitetty Kendallin järjestyskorrelaatiokertoimeen perustuva korrelaatiomatriisi arvonlisistä. Järjestyskorrelaatio osoittaa, että peräkkäisten vuosien arvonlisät vastaavat hyvin toisiaan, mutta kaukaisempien vuosienkin välillä on vielä selkeä yhteys. Esimerkiksi vuosien 2002 ja 2010 lukiojärjestysten korrelaatio on kohtalainen (0.4). Vaikka lukiovaikutuksissa on voimakasta ajallista vaihtelua, osa vaikutuksesta säilyy vuodesta toiseen.

Toinen tapa konkretisoida pysyvyyttä on tarkastella siirtymiä peräkkäisinä vuosina. Peräkkäisiä siirtymiä voidaan kuvata ns. siirtymämatriisin avulla. Siirtymämatriisi kuvaa, kuinka todennäköistä on siirtyä paremmuusjärjestyksen tietystä kohdasta toiseen kahden peräkkäisen vuoden välillä (kuva 5). Esimerkiksi jos lukio *A* on ollut vuonna 2010 parhaan 10 % joukossa, siirtymämatriisi kertoo, mikä todennäköisyys lukiolla on olla seuraavana vuonna parhaan 10 %:n joukossa, toiseksi parhaan 10 %:n joukossa jne.

Siirtymämatriisista nähdään, että lähes 80 % alimman desiilin lukioista pysyy alimmassa desiilissä myös seuraavana vuonna. Luku on melko suuri. Kuitenkin lähes viidesosa pohjimmaisista lukioista nousee seuraavaan desiiliin. Suuri osa uusista pohjaluokan lukioista tulee toiseksi alimmasta luokasta, kuten arvata saattaa.

Pysyvyys on melko samanlaista ylimmässä 10 %:ssa. Valtaosa lukioista pysyy siellä

Kuva 5: Lukioden desiilitransitiomatriisi, riveillä lähtödesiili ja sarakkeissa tulodesiili



myös seuraavana vuonna, mutta noin neljäsosa putoaa yhtä luokkaa alemmas, vastaavasti lähes samansuuruinen määrä nousee toiseksi ylimmästä 10%:sta parhaimpaan luokkaan. Kaikissa desiileissä todennäköisyys pysyä samassa tai joutua naapuridesiileihin on suuri. Pysyvyys voi näyttää varsin suurelta, mutta esimerkiksi vuosikymmenessä lukio voi kokea jopa 9 siirtymää, mikä aikaansaa varsin suurta liikkuvuutta pidemmällä aikavälillä.<sup>17</sup>

Kuva 6 vahvistaa aiemmin havaitun: peräkkäiset arvonlisät ovat voimakkaasti korreloituneita, mutta niihin perustuva paremmuusjärjestys paljon vähemmän. Yhteenvetona voidaan siis todeta, että sekä lukioden järjestys että arvonlisät vaihtelevat ajassa, mutta arvonlisien välillä on vahva aikariippuvuus. Myös lukioden paremmuusjärjestyksessä on havaittavissa pysyvyyttä ajassa, mutta vaihtelua on niin paljon, että suurinta osaa lukioista on mahdotonta laittaa selkeään paremmuusjärjestykseen.

#### 4.5. Pysyvyys ja vertailu muuntyyppisiin ranking- ja arvonlisämenetelmiin

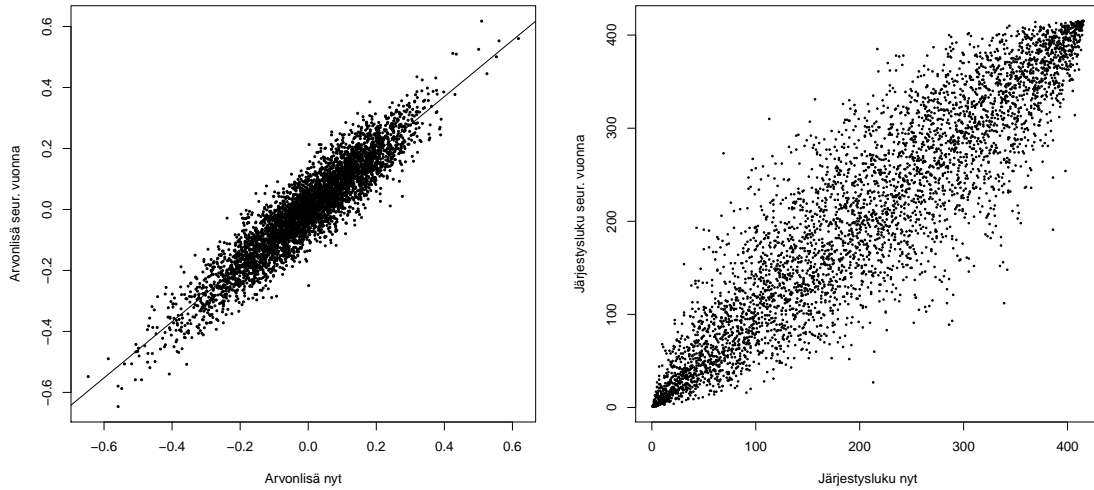
Hyvällä lukioden vertailumenetelmällä on seuraavat neljä ominaisuutta. Hyvä menetelmä:

1. Pyrkii vakioimaan oppilaiden lähtötason tilastollisesti.

<sup>17</sup>Jos tarkastellaan pelkästään siirtymiä ylimmän ja alimman 1 %:n sekä keskimmäisen 98 %:n välillä havaitaan, että parhaassa prosentissa olevalla lukiolla on 60 %:n mahdollisuus pysyä ylimmässä prosentissa myös seuraavana vuonna. Todennäköisyydellä 40 % lukio putoaa pois ylimmästä prosentista. Todennäköisyys päästä pois alimmasta prosentista on 35 %.



Kuva 6: Peräkkäiset vuodet



(a) Peräkkäiset arvonlisät

(b) Peräkkäiset sijoitukset (Huom. suurempi järjestysluku on parempi sijoitus)

2. Sallii lukion laadun vaihtelun kirjoituskertojen välillä.
3. Ottaa kuitenkin huomioon myös aikariippuvuuden lukion laadussa.
4. Ottaa huomioon varsinkin pieniin lukioihin liittyvän epävarmuuden, joka johtuu pienestä kokelaiden määrästä.

Menetelmä, jolla ei ole kaikkia neljää ominaisuutta, antaa todennäköisesti epäluotettavan kuvan lukioiden arvonlisistä ja paremmuusjärjestyksestä.

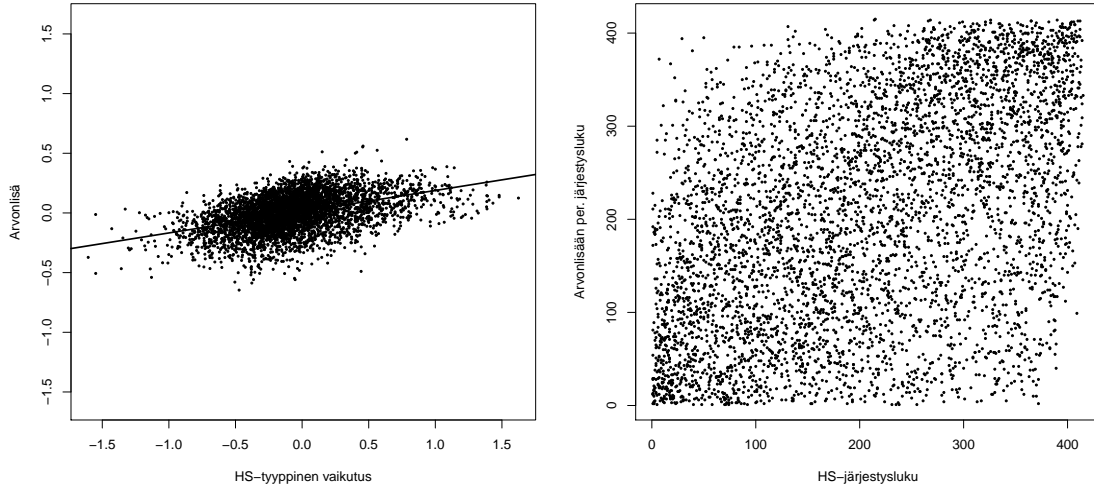
Verrataan tässä työssä käytettyä arvonlisämenetelmää muutamaan muuhun julkisuudessa ja kirjallisuudessa esitettyyn menetelmään. Kustakin menetelmästä todetaan, että koska ne eivät täytä yhtä tai useampaa äsken esitetystä kolmesta kriteeristä, ne antavat tavalla tai toisella epäluotettavia tuloksia lukiovertailussa.

Tarkastellaan ensiksi suoraviivaisinta lähestymistapaa. Siinä lukiot laitetaan paremmuusjärjestykseen yksinkertaisesti ylioppilaskirjoitusten keskiarvon perusteella kunakin vuonna (ns. Helsingin Sanomat -tyyppinen menetelmä [12]). HS-tyyppisellä menetelmällä toteutuu hyvän lukiovertailumenetelmän ominaisuuksista ainoastaan ominaisuus 2.<sup>18</sup>

Kuvassa 7 paneelissa (a) esitetään sirontakuva kaikista arvonlisistä vuosina 2002-2013 ja vastaavista HS-tyyppisistä keskistetyistä raakakeskiarvoista. Kuvasta nähdään, että arvonlisät ovat korreloituneet raakakeskiarvojen kanssa, mutta korrelaatio on varsin heikko

<sup>18</sup>HS-tyyppisen vertailun ominaisuuksien selvittämiseksi aineistosta laskettiin vuosittaiset HS-tyyppiset lukiovaikutukset. Nämä ovat yksinkertaisesti lukion laskennalliset ylioppilaskirjoitusten keskiarvot kullekin vuodelle. Vertailukelpoisuuden parantamiseksi nämä keskistettiin, so. kunkin vuoden keskiarvosta vähennettiin koko vuoden keskiarvo.

Kuva 7: Vertailu HS-tyyppiseen menetelmään



(a) HS-tyyppinen vertailu ja arvonlisät

(b) HS-järjestyslukuvertailu

(keskimäärin 0.4). Tulos on toki odotettavissa. Arvonlisämenetelmän perusajatushan on, että lähtötasolla on suuri merkitys ylioppilaskirjoitusmenestyksessä ja se täytyy ottaa huomioon laskelmissa. HS-tyyppisessä vertailussa oppilaan lähtötasoa ei vakioida. Näin ollen lukiot, joissa on paras lähtötaso, ovat tyypillisesti myös parhaita tällaisessa vertailussa.

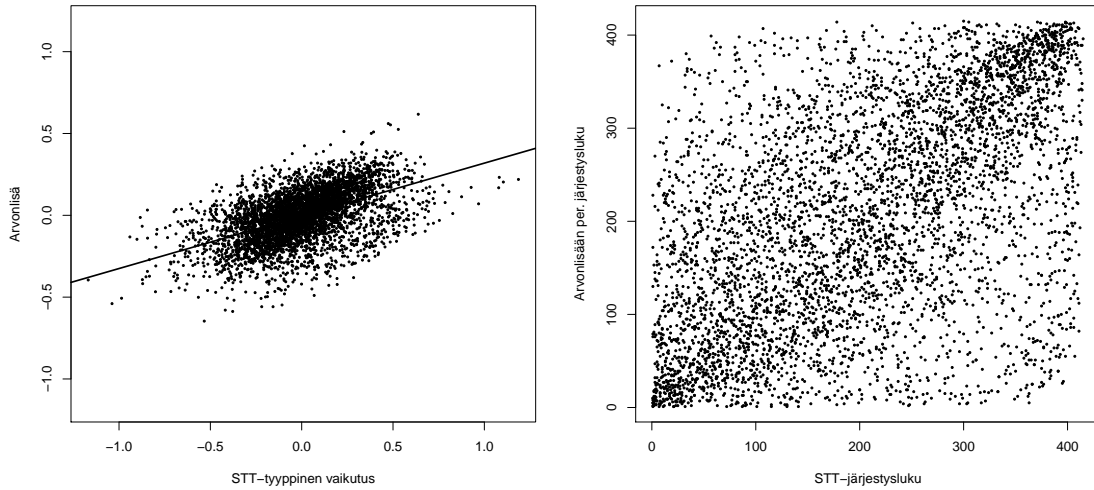
HS-vertailun puutteet näkyvät vielä selvemmin paneelin (b) kuvassa, jossa verrataan HS-tyyppisen vertailumenetelmän ja arvonlisämenetelmän perusteella laskettuja järjestyslukuja. Korrelaatio näiden kahden järjestysluvun välillä on varsin heikkoa: Vuosittaisen Kendall-järjestyskorrelaatioiden keskiarvo on noin 0.3. HS-menetelmän perusteella parhaat lukiot saattavat saada lähes minkä tahansa sijoituksen arvonlisämenetelmän perusteella lasketussa paremmuusjärjestyksessä. HS-tyyppinen vertailu antaa arvonlisämielessä täysin vääränlaisen kuvan lukioden laatueroista.

Tarkastellaan seuraavaksi menetelmää, jossa lukion arvonlisä arvioidaan vuosittain ottamalla huomioon peruskoulun keskiarvolla kuvattu lähtötaso (nk. STT-tyyppinen menetelmä [1]). STT:n käyttämä menetelmä ei kuitenkaan perustu tilastotieteeseen.<sup>19</sup>

STT-tyyppisellä menetelmällä on hyvän lukioden vertailumenetelmän ominaisuuksista kaksi ensimmäistä. Siinä pyritään vakioimaan lähtötaso tilastollisesti ja sallitaan laatu-

<sup>19</sup>Nyt tehtäviä vertailuja varten “STT-tyyppisiksi” tulkitut laskelmat pyrkivät samaan kuin alkuperäiset STT:n laskelmat, eli lähtötason vakiointiin keskiarvon avulla. Ne on kuitenkin tehty tavanomaisia tilastollisia menetelmiä käyttäen. Tarkemmin sanottuna STT-arvonlisäksi tulkitaan kunkin lukion vuosikohtaisesta aineistosta laskettu ns. kiinteä vaikutus mallissa, jossa selittäjänä on kiinteiden vaikutusten lisäksi kolmannen asteen polynomi peruskoulun keskiarvosta. Jos STT on alkuperäisissä laskelmissaan onnistunut lukioontulokeskiarvon vakioinnissa, menetelmä antaa samansuuntaisia tuloksia kuin STT:n.

Kuva 8: Vertailu STT-tyyppiseen menetelmään



(a) STT-tyyppinen vertailu ja arvonlisät

(b) STT-järjestyslukuvertailu

vaihtelu vuodesta toiseen. Menetelmä ei kuitenkaan toteuta kahta jälkimmäistä vaatimusta, ts. se ei huomioi aikariippuvuutta eikä pienten lukioiden otantavaihtelua.<sup>20</sup>

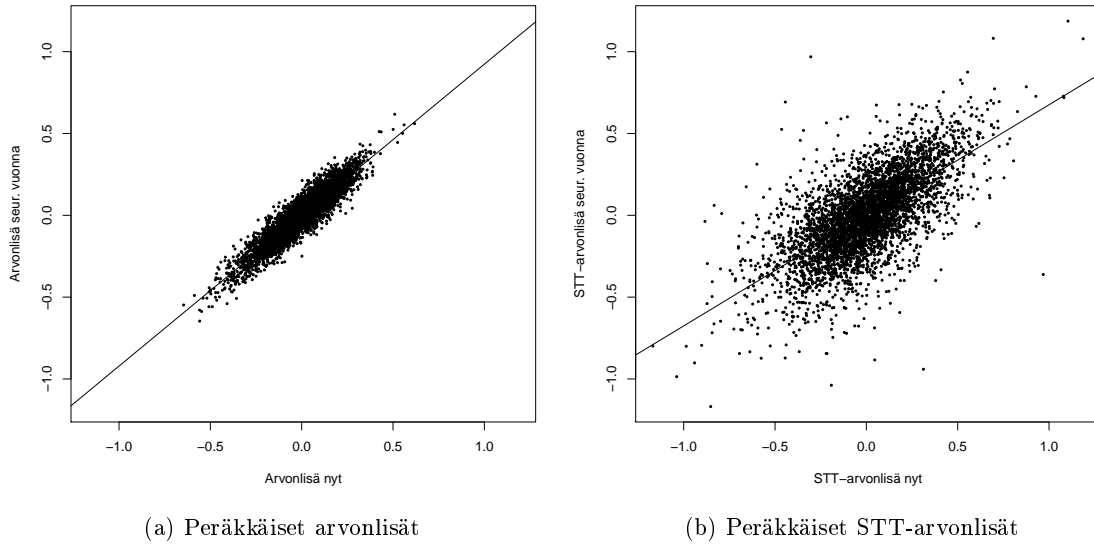
Kuvasta 8 nähdään, että korrelaatio arvonlisien ja STT-tyyppisten lukujen välillä on jonkin verran suurempi kuin arvonlisien ja HS-tyyppisten lukujen välillä. Korrelaatiot vuotuisten järjestyslukujen välillä ovat kuitenkin edelleen varsin pienet, joten STT-tyyppinen menetelmä antaa hyvin erilaisia tuloksia kuin arvonlisämenetelmä. STT-tyyppinen menetelmä on kuitenkin huomattavasti kehittyneempi kuin HS-tyyppinen menetelmä. Mistä erot sitten johtuvat?

Ensiksi, tässä tutkimuksessa käytetyssä arvonlisämenetelmässä on vakioitu peruskoulun keskiarvon lisäksi myös muita tekijöitä, merkittävimpana näistä ylioppilaaan lähtökoulu, so. peruskoulu, josta ylioppilas on lähtenyt. STT-menetelmässä tätä ei ole vakioitu. Periaatteessa olisi mahdollista tehdä STT-tyyppinen vertailu, jossa myös lähtökoulu olisi kontrolloitu. Käytännössä tällaista vertailua on mahdotonta toteuttaa, koska vuosittaiset peruskoulu-lukioyhdistelmien havaintomäärät ovat niin pieniä. Arvonlisämenetelmässä lähtökoulun vakiointi onnistuu, koska menetelmä hyödyntää koko tarkasteluperiodin yksilötason aineistoa.

Toiseksi, STT-menetelmässä jokaista vuotta tarkastellaan erikseen, eikä menetelmä täytä hyvän vertailumenetelmän kolmatta vaatimusta, eli lukiovaikutusten aikariippuvuuden huomioimista. Tämän vuoksi lukiovaikutukset vaihtelevat ajassa huomattavasti enemmän kuin arvonlisämenetelmällä arvioidut vaikutukset. Tämän osoittamiseksi kuvassa 9 on toistettu kuvan 6 peräkkäisten arvonlisien vertailu STT-tyyppisille

<sup>20</sup>Lisäksi menetelmä ei huomioi eroja yläkoulujen arvostelukäytännöissä, vaikka ne voivat merkittävästi vaikuttaa lukioiden välisiin eroihin (vrt. luku 5.2).

Kuva 9: Peräkkäiset vuodet / vertailu STT-menetelmään



arvonlisille.

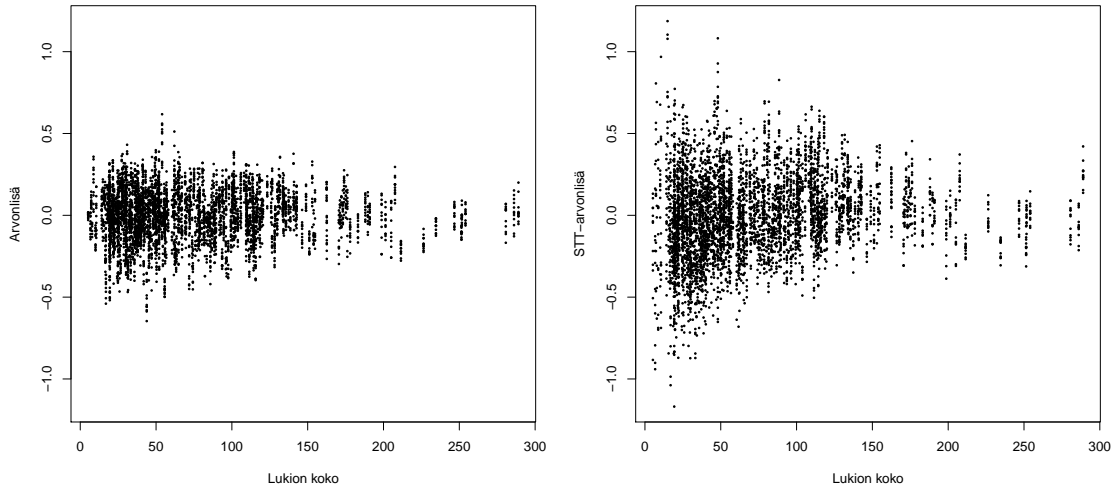
Paneelissa 9a on uudestaan kuva 6a vertailun vuoksi. Paneelissa 9b on vastaava kuva STT-menetelmälle. Kuvasta huomataan, että koska STT-menetelmä ei ota huomioon aikariippuvuutta, STT-arvonlisät vaihtelevat vuodesta toiseen huomattavasti enemmän kuin tämän tutkimuksen tuloksena saadut. Jos STT-arvonlisiä käytetään lukioiden asettamiseen paremmuusjärjestykseen, tulokset ovat vielä huomattavasti dramaattisemmat. Edellisen vuoden järjestysluku ennustaa seuraavan vuoden järjestyslukua erittäin huonosti.<sup>21</sup>

Kolmas tekijä, jonka vuoksi STT-tyyppisen menetelmän antama kuva poikkeaa tässä tutkimuksessa käytetystä arvonlisämenetelmästä on lukion koon vaikutus. Tässä tutkimuksessa käytetyssä menetelmässä arvonlisän arviota korjataan ylioppilaiden määrän mukaan. Jos ylioppilaita on vähän, arviota kutistetaan (engl. *shrink*) kohti keskimääräistä lukiovaikutusta. Tämä johtuu siitä, että kun tietoa on vähän, arvio on epävarma. Epävarmaa arviota on tilastollisessa mielessä järkevää korjata kohti keskimääräistä arvoa. Tällaista korjausta ei tehdä STT-menetelmässä, mikä merkitsee sitä, että pienten lukioiden arvonlisät arvioidaan liian suuriksi tai pieniksi ja ne vaihtelevat suuresti vuodesta toiseen. Ilmiö havaitaan selvästi kuviosta 10. Siinä on vierekkäin kaksi osakuviota, joista ensimmäisessä on kuvattu koulun koon<sup>22</sup> ja arvonlisän yhteys. Vaaka-akselilla on koulun koko ja pystyakselilla kaikki arvonlisät 2002-2013. Toinen osakuvio on samanlainen, mutta pystyakselilla on STT-tyyppinen arvio.

<sup>21</sup>Tätä kuvaava kuva on jätetty pois toiston välttämiseksi. Kuva saatavissa pyydöstä tekijöiltä.

<sup>22</sup>Koulun kooksi on määritelty vuosittaisen laskennallisten ylioppilaiden määrän keskiarvo.

Kuva 10: Lukion koko ja arvonlisä



(a) Lukion koko ja arvonlisä

(b) Lukion koko ja STT-arvonlisä

STT-tyyppisessä menetelmässä “parhaat” ja “huonoimmat” koulut ovat lähes aina pienimmät. Sama pieni koulu voi olla eri vuosina “parhaiden” ja “huonoimpien” joukossa. *Shrinkage*-ominaisuudella, siis hyvän menetelmän ominaisuudella 4, varustetussa menetelmässä ongelma on huomattavasti pienempi. Pienten lukioiden vaihtelu on huomattavasti pienempää, ja “hyviä” kouluja on jakauman eri osissa.

Sekä HS- että STT-tyyppisessä vertailussa on merkittäviä puutteita, jotka tekevät niillä lasketut lukioiden erot ja ennen kaikkea paremmuusjärjestyksen epäluotettaviksi. Kummankin ongelmat johtuvat siitä, etteivät menetelmät toteuta hyvän vertailumenetelmän ominaisuuksia 1-4. Lisäksi ne tai muutkaan aiemmat lukiovertailut eivät huomioi eroja yläkoulujen arvostelukäytännöissä toisin kuin tässä raportissa esitetty menetelmä.

## 5. Luotettavuustarkasteluja

### 5.1. Menetelmän ennusteominaisuudet

Tässä luvussa pyritään selvittämään tulosten luotettavuutta. Ensiksi tarkastellaan menetelmän kykyä ennustaa lukioiden ylioppilaskirjoitusten tuloksia (ns. *out-of-sample* -ennustamiskykyä). Sen jälkeen pyritään selvittämään, millaista oppilaiden valikoitumiseen perustuvaa harhaa tuloksissa voi olla.

Luvussa 2 esiteltiin arvonlisämenetelmän käyttöedellytykset. Ne jakautuivat kahteen ryhmään. Ensimmäiset kaksi oletusta takaavat sen, että menetelmä tuottaa parhaan lineaarisen ennusteen lukion lähtötasovakioidulle yo-kirjoitusten tulokselle ehdolla käytettävissä oleva informaatio. Jälkimmäiset kaksi oletusta ovat huomattavasti voimakkaampia

Taulukko 4: Lukiovaikutuksen keskihajonta

	Estimaatti	Keskivirhe	t-arvo	p-arvo
Vakio	-0.001	0.007	-0.177	0.860
Arvonlisä	1.017	0.022	45.751	0.000

ja takaavat sen, että laskettu arvonlisä on myös aidon lukiovaikutuksen paras ennuste.

Jos siis ensimmäiset kaksi oletusta ovat voimassa, arvonlisän pitäisi olla lähtötasovakioidun tuloksen paras lineaarinen ennuste. Sitä, pitääkö tämä paikkansa voidaan testata yksinkertaisesti ennustamalla jäännöskeskiarvoja lasketuilla arvonlisillä. Kuten luvussa 2 todettiin, menetelmän ennustamiskykyä ei voida arvioida arvonlisillä, joiden laskemisessa on käytetty ennustevuoden tuloksia. Tätä tarkoitusta varten laskettiin, ns. *jackknife*-arvonlisät, joissa kunkin vuoden tuloksia on ennustettu kaikkien muiden vuosien aineistoilla. Tässä luvussa arvonlisä viittaa luvusta 4 poiketen *jackknife*-arvonlisään.

Mikäli arvonlisä on todella lukion lähtötasovakioidun tuloksen paras ennuste, pitäisi ennustemallissa, jossa ylioppilaiden jäännöskeskiarvoa ennustetaan lukion arvonlisällä, arvonlisän kertoimeksi tulla 1. Ennustemallin estimointitulokset on esitetty taulukossa 4.

Arvonlisän kerroinestimaatti on 1.017 ja sen keskivirhe-estimaatti<sup>23</sup> on noin 0.022. Kerroinestimaatti ei poikkea tilastollisesti merkitsevästi ykkösestä. Näin ollen näyttöä siitä, etteikö arvonlisä olisi lähtötasovakioidun tuloksen paras lineaarinen ennuste, ei saada<sup>24</sup>. Käytetyn menetelmän ennusteominaisuudet ovat siten oletusten mukaiset.

## 5.2. Harhatarkasteluja

Arvonlisän tulkitsemisessa todellisen lukiovaikutuksen arvioksi keskeinen ongelma on valikoituminen. Todelliset lukiovaikutuksethan määriteltiin niin, että niiden erotus on odotettu ero ylioppilaskoemenestyksessä, kun oppilaat arvotaan satunnaisesti lukioihin. Koska todellisuudessa oppilaita ei jaeta lukioihin satunnaisesti, voi arvonlisä antaa virheellisen kuvan todellisesta lukiovaikutuksesta esimerkiksi, jos eri lukioihin valikoituneiden oppilaiden välillä on systemaattisia motivaatioeroja.<sup>25</sup>

Ensimmäinen harhatarkastelu koskee lähtökoulun vakioimista. Tutkimuksessa laskettujen arvonlisien taustalla olevassa lähtötason vakiointimallissa mukana oli lähtökoulumuuttuja. Muuttujan tarkoituksena oli yhtäältä ottaa huomioon erot koulujen arvosteluperiaatteissa ja toisaalta vähentää valikoitumisongelmaa. Mikään tähänastisista julkisuudessa esitetyistä lukiovertailuista ei ole huomionnut lähtökoulun vaikutusta. Onkin kiinnostava kysymys, millainen vaikutus tuloksiin lähtökoulun vakioinnilla on. Harhatestauksessa otettiin lähtöpisteeksi malli, jossa oli muuten samat selittäjät kuin perusmallissa,

<sup>23</sup>Keskivirhe-estimaatti on ns. klusterirobusti keskivirhe-estimaatti. Klusterointi on kaksisuuntainen klustereina lukio ja vuosi.

<sup>24</sup>Myös yhteishypoteesi, jossa vakio on nolla ja arvonlisän kerroin yksi, jää voimaan.

<sup>25</sup>Valikoituminen ei ole ongelma vain tämän tutkimuksen arvonlisämenetelmälle, vaan erityisesti myös kaikille edellisessä luvussa esitellyille vaihtoehtoisille menetelmille. Toisin sanoen vaihtoehtoiset menetelmät kärsivät paitsi kaikista edellisessä luvussa todetuista puutteista, myös mahdollisesta valikoitumisongelmasta.

Taulukko 5: Harha-arvioinnin tuloksia

	Poisjätetyt muuttujat	Harhan arvio
1	Yläastemuuttujat	60.3 %
2	Muut arvosanat	0.1 %
3	Sosioekonomiset	0.7 %

mutta lähtökoulu jätettiin vakioimatta. Sitten otettiin uudeksi muuttujaksi lähtökoulu, ja harhatestiä käyttäen arvioitiin lähtökoulujen poisjättämisen aiheuttamaa harhaa.

Arvio yläkoulumuuttujien poisjättämisen aiheuttamasta ennusteharhasta on 60 % (taulukko 5). Toisin sanoen yläkoulumuuttujilla on suuri merkitys mallissa: niiden poisjättäminen aiheuttaa harhatestauksen perusteella sen, että lukioiden eroja yliarvioidaan erittäin merkittävästi. Tälle tulokselle voi olla useampia syitä, mutta todennäköisesti tärkein tekijä on koulujen väliset erot yläkoulujen arvostelukäytännöissä. Joka tapauksessa tulos antaa vahvaa tukea sille, että yläkoulumuuttujien kontrollointi tämänkaltaisessa vertailussa on erittäin tärkeää.

Seuraavaksi testattiin sosioekonomisten muuttujien merkitystä. Koska oppilaskohtaista tietoa sosioekonomisista muuttujista ei ollut käytettävissä, sosioekonomista taustaa pyrittiin vakioimaan Tilastokeskuksen ruututietokannasta saatujen tietojen avulla. Harhaa testattiin kahdella lähestymistavalla<sup>26</sup>, jotka tuottivat hyvin samanlaisen tuloksen. Tulos on taulukon 5 rivillä 3. Harhan arvioitu määrä on alle 1 %, mikä tarkoittaa sitä, että valikoituminen sosioekonomisten muuttujien perusteella ei näyttäisi olevan suuri ongelma. Täytyy kuitenkin korostaa sitä, että tietoja ei ollut saatavilla henkilötasolla ja ruututasollakin vain pienehkölle osaotokselle.

Lisäksi testattiin sitä, voisiko mallia parantaa lisäämällä lähtötasovakiointiin peruskoulun keskiarvon lisäksi myös muita peruskoulumenestystä kuvaavia muuttujia. Käytetyt muuttujat olivat kolmannen kertaluvun polynomit äidinkielen ja matematiikan arvosanoista. Taulukon 5 riviltä 2 havaitaan, että mitään näyttöä näiden vaikutuksesta ei saatu.

Harhatestauksessa käytettävissä olevilla tiedoilla ei saatu näyttöä siitä, että valikoituminen sosioekonomisen taustan tai laajemman koulumenestyksen perusteella olisi ongelma. Sen sijaan lähtökoulun vakioimisella on suuri merkitys. Tämä on mielenkiintoinen ja tärkeä tulos jo pelkästään siitä syystä, että aiemmat lukiovertailut eivät ole huomioineet eroja yläkoulujen arvostelukäytännöissä tai oppilaiden lähtökoulujen vaikutusta.

## 6. Keskustelua tuloksista

Saatujen tulosten perusteella lukioiden välillä on eroja. Ero kunkin vuoden parhaan ja huonoimman lukion arvonlisien välillä on noin yhden arvosanan suuruinen. Mikäli arvonlisiä kuvaa lukioiden välisiä eroja oikein, tämä tarkoittaa, että sama oppilas olisi parhaassa lukiossa saanut arvosanan verran paremman yo-kirjoitusten keskiarvon kuin huonoim-

<sup>26</sup>Ensimmäisessä tavassa mallintaminen tehtiin vain täydellisiä havaintoja käyttäen, toisessa puuttuvat muuttujat koodattiin nollassi ja yhtenä selittäjänä käytettiin puuttumisindikaattoria.

massa. On selvää, että tällainen ero on merkittävä paitsi oppimistulosten, esimerkiksi jatko-opintoihin hakeutumisen näkökulmasta.

Tuloksen tulkinnessa täytyy kuitenkin olla varovainen. Ensiksikin tulos koskee vain kunkin vuoden parasta ja huonointa lukiota. Suurin osa lukioista on hyvin lähellä toisiaan, niin lähellä, että niiden erottaminen toisistaan tilastollisessa mielessä on hyvin epävarmaa. Lisäksi lukioiden arvonlisät vaihtelevat vuodesta toiseen. Vaikka lukioiden laadussa on pysyvä komponentti, vaihtelu peräkkäistenkin vuosien välillä on melko suurta. Kymmenen vuoden aikana arvonlisä voi muuttua monessa lukiossa todella merkittävästi. Toinen seikka, jossa varovaisuus on tarpeen on arvonlisätarkastelujen muuttamisessa lukioiden paremmuusjärjestykseksi. Lukioiden välinen paremmuusjärjestys on vielä arvonlisiäkin epävakampi, ja peräkkäisten vuosien välillä paremmuusjärjestys vaihtuu yllättävän paljon.

Vaikka käytetty menetelmä on varsin hienostunut, on epäselvää kuinka paljon valikoituminen aiheuttaa virhettä tuloksissa. Valikoitumisongelman suuruutta on pyritty tutkimuksessa arvioimaan, mutta ei ole lainkaan varmaa, että ongelma on täysin tyydyttävästi ratkaistu. Lukioiden väliset erot voivat johtua opetuksen tai hallintokäytäntöjen eroista, mutta on myös mahdollista että eroja selittävät tutkijalle havaitsemattomat asiat kuten oppilaiden motivaatioerot. On kuitenkin jonkin verran näyttöä siitä, että esimerkiksi opettajien arvonlisä ennustaa oppilaiden työmarkkinamenestystä aikuisena [7], joten arvonlisämenetelmiä ei missään tapauksessa voida sivuuttaa suoralta kädeltä. Silti voi olla liian suoraviivaista ajatella, että siirtämällä oppilas huonoimmasta lukioista parhaaseen, hän automaattisesti kirjoittaisi yhden arvosanan verran paremmin.

Tutkimuksessa osoitettiin, että arvonlisän ja suoraan ylioppilaskirjoitusten keskiarvoon perustuvalla vertailulla, nk. HS-vertailulla, on hyvin vähän yhteistä. Esimerkiksi kirjoitusmenestykseltään parhaat Helsingin “eliittilukiot” ovat arvonlisävertailussa muista erottumatonta keskitasoa. Tämä tulos on mielenkiintoinen, mutta ei kovin yllättävä viimeaikaisen kansainvälisen tutkimuksen perusteella. Esimerkiksi Yhdysvalloissa ja Kiinassa on havaittu että eliittilukiot eivät ole arvonlisällä mitattuna parhaita kouluja, vaikka niissä saavutetaan parhaimmat oppimistulokset suotuisan oppilasaineuksen takia (ks. esim. [2, 35]). Nämä tulokset viittaavat vahvasti siihen, että ns. vertaisryhmien vaikutukset oppimistuloksiin lukiokoulutuksessa ovat vähäisiä.

Yksi eliittilukioiden arvonlisän tulkintaa mahdollisesti vaikeuttava ongelma liittyy siihen, että arvosteluasteikko ylioppilaskokeessa on ylhäältä rajoitettu. Esimerkiksi pelkkiä laudatureja sisältäviä laskennallisia koetuloksia oli aineistossa satoja. On luonnollisesti mahdotonta saada parempaa keskiarvoa kuin laudatur. Yläraja voi leikata eliittilukioiden arvonlisien suuruutta. Kouluun tuleva peruskoulussa huippuarvosanat saaneella oppilaalla “ei ole tilaa” suurelle arvonlisälle, koska häneltä odotetaan lähellä ylärajaa olevaa yo-kirjoitustulosta jo peruskoulumenestyksen perusteella.

On vaikea sanoa varmasti, kuinka vakava tämä ongelma on. Ongelma ei kuitenkaan luultavasti ole niin suuri kuin äskeisestä esimerkistä saattaisi päätellä. Esimerkissä, jossa oppilaalle ennustetaan lähelle ylärajaa olevia yo-kirjoitustuloksia, mukana ei ole selittäviä muuttujia. Sen sijaan arvonlisälaskelmissa käytetyissä malleissa niitä on. Näin ollen peruskoulun huippuoppilaallekaan ei välttämättä ennusteta lähelle laudaturia olevia yo-kirjoituskeskiarvoja. Onhan nimittäin paljon peruskoulussa hyvin menestyneitä oppilaita,



joiden yo-keskiarvo jää kauas laudaturista.

Toisaalta tulokset osoittavat myös ns. STT-tyyppisen vertailun ongelmat. STT:n käytämä yksinkertainen arvonlisämittari ei ota huomioon lukioiden laadun pysyvyyttä ja siksi sen tuottamat tulokset ovat hyvin epävakaita. Varsinkin lukioiden paremmuusjärjestys heilahtelee voimakkaasti vuodesta toiseen. Lisäksi menetelmä ei ota huomioon pieniin lukioihin liittyvää epävarmuutta, minkä johdosta pienet lukiot ovat joka vuosi “parhaita” ja “huonoimpia”. Lisäksi on syytä huomauttaa, että erityisesti eliittilukioita koskeva arvosteluasteikon ylhäältä rajoittuneisuus on merkittävämpi ongelma STT-tyyppiselle vertailulle, jossa kontrolloidaan ainoastaan lähtötaso<sup>27</sup>.

Parhaaseen mahdolliseenkin arvonlisämenetelmään perustuva lukioiden paremmuusjärjestys on varsin epävarma ja vaihtelee vuodesta toiseen. Tästä syystä emme halua esittää raportissa uutta lukiorankingia, vaikka listaus sinänsä voitaisiinkin tehdä. Yksittäisen rankinglistan esittäminen saattaisi antaa väärän kuvan lukioiden paremmuusjärjestyksestä. Voidaanko lukioiden paremmuusjärjestyksestä sitten sanoa mitään? Kenties riittävän varovainen tapa on tarkastella lukioita, jotka ovat pysyneet merkittävän osan tarkasteluperiodista jakauman ylä- tai alapäässä.

Koska tarkastelujakso on varsin pitkä, alkupään sijoitukset eivät kenties ole yhtä kiinnostavia kuin tuoreemmat. Taulukossa 6 on lueteltu ne suomenkieliset lukiot, jotka ovat olleet ylimmän tai alimman 5 %:n joukossa vuonna 2008 ja sen jälkeen. Lukiot ovat taulukossa sattumanvaraisessa järjestyksessä. Lisäksi taulukossa on erikseen lueteltu ne ruotsinkieliset lukiot, jotka ovat olleet vuodesta 2008 alkaen ylimmässä tai alimmas-  
sa kuudenneksessä. Tämä vastaa siis suurin piirtein jakauman alinta ja ylintä 17 %:a. Ruotsinkielisten eri prosenttiluku on seurausta siitä, että ruotsinkielisiä lukioita on vähän.

Suomen- ja ruotsinkielisiä lukioita tarkastellaan erikseen, koska eri kielisten lukioiden tulosten vertailukelpoisuuteen saattaa liittyä ongelmia. Yksi merkittävä syy on se, että suurin osa äidinkielenään ruotsia puhuvista opiskelijoista menee ruotsinkieliseen lukioon. Lisäksi tulostemme perusteella suomen- ja ruotsinkielisten yläkoulumuuttujien vaikutuksissa on systemaattisia eroja. Tämä todennäköisesti johtuu erilaisista yläkoulujen arvostelukäytännöistä, mutta emme voi myöskään kumota muita selityksiä.<sup>28</sup>

Taulukosta 6 havaitaan, että alimmassa 5 %:ssa on jonkin verran enemmän lukioita kuin ylimmässä 5 %:ssa. Tämä merkitsee sitä, että jakauman alapäässä pysyminen on todennäköisempää kuin yläpäässä. Toisaalta tässä yhteydessä on hyvä palauttaa mieleen, että alimman ja ylimmän desiilin väliin jäävissä luokissa havaittiin runsasta liikkuvuutta.

Suomenkielisistä lukioista Mynämäen lukio on arvonlisältään ollut ylimmässä 5 %:ssa koko tarkasteluperiodin ajan. Arvonlisältään korkeimmat ja matalimmat suomenkieliset koulut ovat kauttaaltaan pienehköillä paikkakunnilla sijaitsevia pienehköjä lukioita.

<sup>27</sup>Erityisesti alkuperäinen STT-vertailu on herkkä tälle ongelmalle, koska käytetyssä laskentamallissa päättötodistukselle annetaan suurempi (ykkösen suuruinen) paino.

<sup>28</sup>Mikäli vertailemme suomen- ja ruotsinkielisiä lukioita keskenään, yläkoulumuuttujien huomioiminen heikentää merkittävästi ruotsinkielisten lukioiden arvonlisää. Tämä vaikuttaisi johtuvan ennen kaikkea ruotsinkielisten yläkoulujen arvostelukäytännöistä, mutta asian varmistaminen vaatisi enemmän tutkimusta.

Taulukko 6: Lukiot, jotka vuodesta 2008 lähtien olivat koko ajan jakauman ylä- tai alapäässä.

2008 alk. ylin 5 %	2008 alk. alin 5 %
Leppävirran lukio	Harjavallan lukio
Mynämäen lukio	Langinkosken lukio
Reisjärven lukio	Merikarvian lukio
	Mouhijärven lukio
	Konneveden lukio
	Rautjärven lukio
	Pälkäneen lukio

(a) Suomenkieliset

	2008 ylin kuudesosa	2008 alk. alin kuudesosa
3406	Svenska Samskolan i Tammerfors	Närpes gymnasium
		Vörå samgymnasium

(b) Ruotsinkieliset

## 7. Johtopäätökset

Raportin keskeinen tulos on, että kunakin vuonna parhaiden ja huonoimpien lukioiden välillä näyttää olevan merkittäviä laatueroja. Tulokseen liittyy kuitenkin useita varauksia. Merkittäviä eroja on vain aivan jakauman ääripäässä olevien koulujen välillä. Suurinta osaa kouluista ei voida tilastollisesti erottaa toisistaan, eikä siis asettaa luotettavasti paremmuusjärjestykseen. Tuloksemme myös osoittavat, että julkisuudessa esitetyt ”lukio-rankingit” ovat harhaanjohtavia, ainakin mikäli niiden tarkoituksena on asettaa lukioita paremmuusjärjestykseen laadun tai arvonalisän perusteella. Lukiovaikutusarvioiden tulkintaa hankaloittaa lisäksi se, että osa havaituista lukioiden välisistä laatueroista voi olla valikoitumisen aiheuttamia näennäisiä eroja. Valikoitumisongelman suuruutta ei voitu täysin tyydyttävästi selvittää käytettävällä aineistolla ja sitä olisi jatkossa syytä tutkia vielä tarkemmin.

Lukioita vertailtaessa ei aina muisteta, että lukion tuottama oppimistulos ei suinkaan ole ainoa tekijä lukiovalinnassa. Esimerkiksi pääkaupunkiseudun ns. ”eliittilukioissa” verkostoituminen voi olla yhtä lailla valintaan vaikuttava tekijä kuin hyvä tulos ylioppilaskirjoituksissa. Suomessa ei ole tutkimustuloksia lukioaikaisten ystävyyssuhteiden ja verkostojen vaikutuksesta esimerkiksi työ- tai avioliittomarkkinoilla, mutta on mahdollista, että näillä on suurikin merkitys. Lisäksi tässä tutkimuksessa esitetyt tulokset kuvaavat lukioiden vaikutusta oppimistuloksiin vain lyhyellä aikavälillä. Jatkossa olisikin mielenkiintoista tutkia lukioiden vaikutuksia erilaisiin pitkän aikavälin tulemiin.

## Viitteet

- [1] AAMULEHTI (2014): “STT:n lukiovertailu: Viime vuoden ykkönen romahti sijalle 56,” .
- [2] ABDULKADIROGLU, A., J. D. ANGRIST, AND P. A. PATHAK (2014): “The elite illusion: Achievement effects at Boston and New York exam schools,” *Econometrica*, 1(82), 137–196.
- [3] ANSALA, L., AND J. PÄÄKKÖNEN (2013): “Kouluvaikutus ja tuloksellisuusrahoitus lukiokoulutuksessa,” Discussion paper, VATT Valmisteluraportit 16.
- [4] BRIGGS, D. C., AND J. P. WEEKS (2011): “The persistence of school-level value-added,” *Journal of Educational and Behavioral Statistics*, 36(5), 616–637.
- [5] CHAMBERLAIN, G. E. (2013): “Predictive effects of teachers and schools on test scores, college attendance, and earnings,” *Proceedings of the National Academy of Sciences*, p. 201315746.
- [6] CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014a): “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates,” *American Economic Review*, 9(104), 2593–2632.
- [7] ——— (2014b): “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood,” *American Economic Review*, 9(104), 2633–2679.
- [8] GOLDHABER, D., AND D. CHAPLIN (2012): “Assessing the Rothstein Test: Does It Really Show Teacher Value-Added Models Are Biased?,” Discussion paper, Mathematica Policy Research.
- [9] GOLDHABER, D., AND M. HANSEN (2013): “Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance,” *Economica*, 80(319), 589–612.
- [10] GOLDSTEIN, H., AND M. J. HEALY (1995): “The graphical presentation of a collection of means,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 175–177.
- [11] GOLDSTEIN, H., AND D. J. SPIEGELHALTER (1996): “League tables and their limitations: statistical issues in comparisons of institutional performance,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 385–443.
- [12] HELSINGIN SANOMAT (2014): “HS:n lukiovertailussa kaikki maan koulut: Etelä-Tapiolan lukio ykkönen,” .
- [13] JACKSON, C. K., J. E. ROCKOFF, AND D. O. STAIGER (2014): “Teacher Effects and Teacher-Related Policies,” *Annual Review of Economics*, 6, 801–825.
- [14] JACOB, B. A., L. LEFGREN, AND D. P. SIMS (2010): “The persistence of teacher-induced learning,” *Journal of Human Resources*, 45(4), 915–943.

- [15] KANE, T. J., D. F. MCCAFFREY, T. MILLER, AND D. O. STAIGER (2013): “Have we identified effective teachers? Validating measures of effective teaching using random assignment,” *Seattle, WA: Bill and Melinda Gates Foundation*.
- [16] KANE, T. J., AND D. O. STAIGER (2001): “Improving school accountability measures,” Discussion paper, National Bureau of Economic Research.
- [17] ——— (2002): “The promise and pitfalls of using imprecise school accountability measures,” *The Journal of Economic Perspectives*, 16(4), 91–114.
- [18] ——— (2008): “Estimating teacher impacts on student achievement: An experimental evaluation,” Discussion paper, National Bureau of Economic Research.
- [19] KIM, H., AND D. LALANCETTE (2013): “Literature Review on the Value-Added Measurement in Higher Education,” Discussion paper, OECD.
- [20] KINSLER, J. (2012): “Assessing Rothstein’s critique of teacher value-added models,” *Quantitative Economics*, 3(2), 333–362.
- [21] KIRJAVAINEN, T. (2009): *Essays on the Efficiency of Schools and Student Achievement*, no. 53 in VATT Publications. VATT.
- [22] KOEDEL, C., AND J. R. BETTS (2011): “Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique,” *Education Finance and Policy*, 6(1), 18–42.
- [23] KUUSELA, J. (2003): “Lukioiden tuloksiin vaikuttavista tekijöistä,” Discussion paper.
- [24] ——— (2006): *Temaattisia näkökulmia perusopetuksen tasa-arvoon.*, Oppimistulosten arviointi 6/2006. Helsinki: Opetushallitus.
- [25] LAPPALAINEN, H.-P. (2004): *Peruskoulun äidinkielen oppimistulosten kansallinen arviointi 9. vuosiluokalla 1999*, Oppimistulosten arviointi 1/2000. Helsinki: Opetushallitus.
- [26] ——— (2006): *Ei taito taakkana ole. Perusopetuksen äidinkielen ja kirjallisuuden oppimistulosten arviointi 9. vuosiluokalla 2005*, Oppimistulosten arviointi 1/2006. Helsinki: Opetushallitus.
- [27] ——— (2011): *Sen edestään löytää. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2010*, Koulutuksen seurantaraportti 2011:2. Helsinki: Opetushallitus.
- [28] MATTILA, L., AND J. RAUTOPURO (2013): *Koulukohtaisia tuloksia. Teoksessa Hyödyllinen pakkolasku - Matematiikan oppimistulokset peruskoulun päättövaiheessa 2012. Toim. Juhani Rautopuro*, Koulutuksen seurantaraportit 2013:3. Helsinki: Opetushallitus.

- [29] McCaffrey, D. F., J. Lockwood, D. Koretz, T. A. Louis, and L. Hamilton (2004): “Models for value-added modeling of teacher effects,” *Journal of Educational and Behavioral Statistics*, 29(1), 67–101.
- [30] McCaffrey, D. F., T. R. Sass, J. Lockwood, and K. Mihaly (2009): “The intertemporal variability of teacher effect estimates,” *Education Finance and Policy*, 4(4), 572–606.
- [31] Morris, C. N. (1983): “Parametric empirical Bayes inference: theory and applications,” *Journal of the American Statistical Association*, 78(381), 47–55.
- [32] Ouakrim-Soivio, N. (2013): *Toimivatko päättöarvioinnin kriteerit?*, Raportit ja selvitykset. University of Helsinki, Faculty of Behavioural Sciences.
- [33] Rothstein, J. (2010): “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement,” *The Quarterly Journal of Economics*, 125(1), 175–214.
- [34] Wilson, D., and A. Piebalga (2007): *Accurate Performance Measures But Meaningless Ranking Exercise?: An Analysis of the English School League Tables*. University of Bristol, Centre for Market and Public Organisation.
- [35] Zhang, H. (2013): “The Mirage of Elite Schools: Evidence from Lottery-based School Admissions in China,” Discussion paper, Mimeo.

## A. Laskentamenetelmien kuvaus

### A.1. Jackknife-estimaattori

Oletusten 1 ja 2 ollessa voimassa lineaarinen ennuste on muotoa:

$$\hat{\mu}_{j,t} = \hat{\Sigma}_{j,t}^{-1} \hat{\gamma}_{j,t},$$

jossa  $\hat{\Sigma}_{j,t}$  on muuttujien  $\bar{A}_{j,t-1}, \dots, \bar{A}_{j,t-s}$  estimoitu  $s \times s$  -kovarianssimatriisi ja  $\hat{\gamma}_{j,t}$  on estimoitu muuttujien  $\bar{A}_{j,t-1}, \dots, \bar{A}_{j,t-s}$   $s \times 1$  -kovarianssivektori muuttujan  $\bar{A}_{j,t}$  kanssa.

Oletusten vallitessa tarvittavat kovarianssit voidaan estimoida aineistosta seuraavasti. Kovarianssi  $Cov(\bar{A}_{j,t}, \bar{A}_{j,t-k}) = \sigma_{\bar{A},k}$  riippuu vain viipeestä  $k$  ja on sama kaikille lukioille  $j$ . Autokovarianssit  $\sigma_{\bar{A},k}$  voidaan estimoida ylioppilaiden määrällä painotettuna keskiarvona kunkin lukion sisäisistä keskimääriäisten jäännöskeskiarvojen autokovariansseista. Nämä estimaatit muodostavat matriisin  $\hat{\Sigma}_{j,t}$  diagonaalin ulkopuoliset alkiot siten, että  $[\hat{\Sigma}_{j,t}]_{k,l} = \hat{\sigma}_{\bar{A},|k-l|}$  kun  $k \neq l$ . Myös vektori  $\hat{\gamma}_{j,t}$  muodostuu näistä kovarianssiestimaateista, niin että  $[\hat{\gamma}_{j,t}]_k = \hat{\sigma}_{\bar{A},k}$ .

Näiden lisäksi tarvitaan arvio kullekin varianssille  $Var(\bar{A}_{j,t-k})$ .

Sopiva estimaattori on

$$\hat{\sigma}_{\bar{A},t}^2 = \hat{\sigma}_A^2 - \hat{\sigma}_\varepsilon^2 + \frac{\hat{\sigma}_\varepsilon^2}{n_{j,t-k}},$$

jossa  $\hat{\sigma}_A^2 = \frac{1}{N-K-1} \sum \sum (A_{i,t} - \bar{A})^2$  ja  $\hat{\sigma}_\varepsilon^2 = \frac{1}{N-K-C-1} \sum_j \sum_t \sum_{j(i,t)=j} (A_{i,t} - \bar{A}_{j,t})^2$ . Näissä  $K$  on kontrollimuuttujien määrä ja  $C$  on kaikkien lukio-vuosi -yhdistelmien määrä ja  $n_{j,t-k}$  ylioppilaiden määrä lukiossa  $j$  vuonna  $t-k$ . Nämä estimaatit muodostavat  $\hat{\Sigma}_{j,t}$ :n diagonaalialkiot siten, että  $[\hat{\Sigma}_{j,t}]_{k,k} = \hat{\sigma}_{\bar{A}_{j,t-k}}^2$ . Nyt kaikki  $\hat{\mu}_{j,t}$ :n laskemiseen tarvittavat osaestimaatit ovat koossa, ja ennuste voidaan laskea.

Jos kaikki oletukset ovat voimassa voidaan lukua  $\hat{\sigma}_\mu^2 = \hat{\sigma}_A^2 - \hat{\sigma}_\varepsilon^2$  käyttää lukiovaikutusten varianssin estimointiin. Näin tehdään luvussa 4.

Myös ennustevuotta käyttävä estimaattori lasketaan samaan tapaan kuin jackknife-estimaattori, kovarianssimatriisi vain on hieman erilainen.

## A.2. Shrinkage-ominaisuudesta

Jos laatu pysyisi vakiona ajassa, yllä esitettyjä estimaattoreita voitaisiin yksinkertaistaa. Esimerkiksi *jackknife*-tyyppinen (siis tarkasteluvuoden aineistoa käyttämätön) estimaattori olisi tällöin muotoa

$$\hat{\mu}_{j,t}^{SIMPLE} = \bar{A}_j^{-t} \frac{\hat{\sigma}_A^2 - \hat{\sigma}_\varepsilon^2}{\hat{\sigma}_A^2 - \hat{\sigma}_\varepsilon^2 + \frac{\hat{\sigma}_\varepsilon^2}{N_{j,t-1} + \dots + N_{j,t-s}}} \quad (3)$$

Merkintä  $\bar{A}_j^{-t}$  tarkoittaa kaikkien jäännöskeskiarvojen keskiarvoa lukiossa  $j$  lukuunottamatta tarkasteluvuoden  $t$  havaintoja. Kaavasta nähdään selvästi shrinkage-ominaisuus: kouluvaikutusestimaatti on jäännöskeskiarvo kutistettuna tekijällä, jonka suuruus riippuu koulun ylioppilaiden määrästä. Koska oletus ajassa vakioisesta laadusta ei ollut empiiristen tulosten perusteella uskottava, näitä estimaatteja ei ole sisällytetty tähän raporttiin. Yhtälö 3 on mukana vain intuition lisäämiseksi.

## A.3. Ennusteharhan testaus

Oletetaan, että käytössä on arvonlisäestimaatit  $\hat{\mu}_{j,t}$  kaikille lukioille. Tässä, kuten koko osaluvussa käytetään nimenomaan *jackknife*-arvonlisiä  $\hat{\mu}_{j,t}$ , koska kaikki tarkastelut perustuvat ennustamiseen.

Arvotaan nyt oppilaita satunnaisesti kaikkiin lukioihin ja estimoidaan regressiomalli

$$A_{i',t} = \alpha + \lambda \hat{\mu}_{j,t} + \nu_{i',t}$$

kaikille satunnaistetuille oppilaille  $i'$ . Arvonlisissä ei ole ennusteharhaa, mikäli  $\lambda = 1$ . Ennusteharhan määritelmä on  $B(\hat{\mu}) = 1 - \lambda$ . Auki kirjoitettuna

$$\lambda = \frac{Cov(A_{i',t}, \hat{\mu}_{j,t})}{Var(\hat{\mu}_{j,t})} = \frac{Cov(\mu_{j,t} + \varepsilon_{i',t}, \hat{\mu}_{j,t})}{Var(\hat{\mu}_{j,t})} = \frac{Cov(\mu_{j,t}, \hat{\mu}_{j,t})}{Var(\hat{\mu}_{j,t})},$$

jossa viimeinen yhtälö seuraa satunnaistamisesta.

Esitetään nyt harhan arviointimenetelmä, joka perustuu tekstissä esitettyyn ajatukseen sosioekonomisten muuttujien hyödyntämisestä. Oletusten 1 ja 2 vallitessa arvonlisä on paras lineaarinen ennuste koulun  $j$  oppilaiden lähtötasovakioidulle menestykselle.

Näin ollen lukiosta  $j$  vuonna  $t$  kirjoittaneille ylioppilaille  $i$  pätee

$$1 = \frac{\text{Cov}(A_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})} = \frac{\text{Cov}(\mu_{j,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})} + \frac{\text{Cov}(\varepsilon_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})},$$

joten ennusteharha on  $B(\hat{\mu}) = \frac{\text{Cov}(\varepsilon_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})}$ .

Oletetaan nyt, että käytössä on arvonlisän estimoinnissa käytettyjen selittävien muuttujien  $X_{i,t}$  lisäksi vektori lisämuuttujia  $P_{i,t}^* = (P_{i,t}^1, \dots, P_{i,t}^K)$ . Nämä voivat olla esimerkiksi mainittuja sosioekonomisia muuttujia. "Oikea" malli onkin siis  $A_{i,t}^* = \mu_{j,t} + \beta' X_{i,t} + \rho' P_{i,t}^* + u_{i,t}$ . Koska uusia muuttujia ei otettu huomioon alkuperäistä mallia estimoidessa, ovat alkuperäisestä mallista saadut jäännöskeskiarvot muotoa (kun estimointivirheet unohdetaan)  $A_{i,t} = \mu_{j,t} + \rho'(P_{i,t}^* - \hat{P}_{i,t}) + u_{i,t}$  jossa kukin  $\hat{P}_{i,t}^k = \nu^k X_{i,t}$ . Tässä  $\nu^k$  ovat kertoimet regressiosta, jossa muuttujaa  $P_{i,t}^{*k}$  selitetään kiinteillä lukiovaikutuksilla ja  $X$ -muuttujilla.

Määritellään nyt  $P_{i,t} = P_{i,t}^* - \hat{P}_{i,t}$ . Näin ollen alkuperäinen virhetermi  $\varepsilon_{i,t}$  jakautuu kahteen komponenttiin:  $\varepsilon_{i,t} = \rho' P_{i,t} + u_{i,t}$ . Ensimmäinen komponentti on uusien selittäjien kanssa korreloitu osa ja toinen komponentti  $u_{i,t}$  on uusien selittäjien kanssa korreloimaton osa.

Oletetaan nyt optimistisesti, että valikoituminen tapahtuu ainoastaan uusien muuttujien  $P_{i,t}$  perusteella. Tällöin harhan suuruus on  $B(\hat{\mu}) = \frac{\text{Cov}(\varepsilon_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})} = \frac{\text{Cov}(\rho' P_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})}$ . Tätä on mahdollista estimoida varsin suoraviivaisesti. Muodostetaan ensin muuttujat  $P_{i,t}$ . Estimoidaan kullekin  $P_{i,t}^{*k}$  malli, jossa sitä selitetään kiinteillä lukiovaikutuksilla ja muuttujilla  $X_{i,t}$ . Muodostetaan kustakin jäännöskeskiarvo  $P_{i,t}^k$  samaan tapaan kuin aiemmin  $A_{i,t}$ . Estimoidaan sitten malli, jossa jäännöskeskiarvoja  $A_{i,t}$  selitetään näillä ja kiinteillä lukiovaikutuksilla. Näin saadaan estimaatti  $\hat{\rho}$  kertoimelle  $\rho$  ja voidaan muodostaa ennuste  $A_{i,t}^p = \hat{\rho}' P_{i,t}$ . Harhan estimaatti saadaan sitten muodostamalla regressiomalli, jossa  $A_{i,t}^p$ :tä selitetään arvonlisällä  $\hat{\mu}_{j,t}$ .

Estimointi voidaan suorittaa myös toista reittiä. Muodostetaan aluksi jäännöskeskiarvot käyttäen laajennettua mallia, jossa mukana paitsi  $X_{i,t}$  myös  $P_{i,t}^*$ . Kutsutaan tästä saatuja jäännöskeskiarvoja  $A_{i,t}^N$ . "Vanhoille" ja uusille jäännöskeskiarvoille pätee likimain  $A_{i,t} - A_{i,t}^N = \rho' P_{i,t}$ . Näin ollen harha on  $B(\hat{\mu}) = \frac{\text{Cov}(\rho' P_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})} = \frac{\text{Cov}(A_{i,t}, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})} - \frac{\text{Cov}(A_{i,t}^N, \hat{\mu}_{j,t})}{\text{Var}(\hat{\mu}_{j,t})}$ . Toisin sanoen ennusteharha voidaan estimoida luvulla  $\hat{B}(\hat{\mu}) = \hat{\nu} - \hat{\nu}^N$ , jossa  $\hat{\nu}$  on estimoitu regressiokerroin mallista, jossa selitetään alkuperäisiä jäännöskeskiarvoja arvonlisällä ja  $\hat{\nu}^N$  on vastaava kerroin mallista, jossa selitettävänä on uusi jäännöskeskiarvo.

Ennusteharha voidaan siis yrittää arvioida käyttämällä lisämuuttujia. On tärkeää huomata, että äskeinen menetelmä havaitsee koko harhan ainoastaan, mikäli oletus siitä, että valikoituminen tapahtuu vain lisämuuttujien perusteella on totta. Jos oletus ei pidä paikkaansa, voi ennusteharhaa olla, vaikka menetelmä ei sitä havaitsisikaan. Toisin sanoen, menetelmä antaa välttämättä nollatuloksen mikäli harhaa ei ole. Se voi kuitenkin antaa nollatuloksen silloinkin, kun harhaa on.