

VATT Working Papers 84

The Effects of Relative School Starting Age on Educational Outcomes in Finland

Martti Kaila

VATT WORKING PAPERS

84

The Effects of Relative School Starting Age
on Educational Outcomes in Finland

Martti Kaila

Martti Kaila, Valtion taloudellinen tutkimuskeskus, martti.kaila@vatt.fi

This paper is based on my Master's thesis at the University of Helsinki. I would like to thank Markus Jäntti, Mika Kortelainen, Miika Päällysaho, Mikko Silliman and Isa Kuosmanen for valuable comments.

ISBN 978-952-274-190-5 (PDF)

ISSN 1798-0291 (PDF)

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

Helsinki, April 2017

The Effects of Relative School Starting Age on Educational Outcomes in Finland

VATT Institute for Economic Research
VATT Working Papers 84/2017

Martti Kaila

Abstract

In Finland, children start school during the calendar year they turn seven years old. This creates a discontinuous jump in school starting age. I utilize a regression discontinuity design and rich register data to study whether this discontinuous jump in the school starting age affects educational outcomes. I find that the school starting age law generates a significant jump in the school starting age at the turn of the year, which in turn affects educational outcomes. According to my results, those who are born just after new year have on average a 0.15 grade points higher GPA and are significantly more likely to be admitted to and graduate from general upper secondary school.

In addition, I study heterogeneity in the results and find that the effect is significantly stronger for females than males. The findings may be taken as a causal effect of relative school starting age. To support this, I show that the density of assignment variable and various background variables evolve continuously in the vicinity of New Year. Theoretical literature offers three potential mechanisms that could explain the effects of school starting age. Firstly, the deviation may arise from the optimal school starting age. Secondly, the gaps may be caused by peer effects and lastly, relatively older children may perform better since they take the exams at an older age. I cannot distinguish between the different channels, and hence my results should be taken as a combined effect of all mechanisms.

Key words: education, school starting age, regression discontinuity design

JEL classes: I21, I28, J13

Tiivistelmä

Suomessa oppivelvollisuus alkaa sinä vuonna, kun lapsi täyttää seitsemän vuotta. Tästä seuraa se, että alkuvuodesta syntyneet lapset aloittavat koulun keskimäärin vuoden vanhempana kuin loppuvuodesta syntyneet. Tässä tutkimuksessa hyödynnän regressioepäjatkuvuusasetelmaa ja tutkin, miten tämä suhteellinen ikäero vaikuttaa koulutusmuuttujiin Suomessa. Tutkimuksen tuloksien perusteella juuri alkuvuodesta syntyneiden yksilöiden keskiarvo on keskimäärin 0,15 arvosanaa korkeampi peruskoulun lopussa kuin juuri loppuvuodesta syntyneiden lapsien. Tämän ohella juuri alkuvuodesta syntyneet lapset pääsevät lukioon ja valmistuvat lukiosta suuremmalla todennäköisyydellä kuin juuri loppuvuodesta syntyneet lapset. Havaitseen myös, että vaikutus on suurempi tytöille kuin pojille.

Tutkimuksen tulokset voidaan tulkita niin, että suhteellisella koulunaloituserolla on kausaalinen vaikutus koulutusmuuttujiin. Tämän tueksi tutkimuksessa osoitetaan, että havaintojen määrä on samansuuruinen vuodenvaihteen molemmin puolin. Lisäksi näytetään, että erilaisten taustamuuttujien jakaumat ovat jatkuvia leikkauspisteen ympäristössä. Kolme mekanismia voi selittää havaitun eron. Ensiksi vanhempana koulun aloittavat lapset saattavat olla optimaalisemmassa iässä oppimisen kannalta. Toiseksi havaittuun eroon voi vaikuttaa suhteellisesta ikäerosta kumpuavat vertaisryhmävaikutukset. Kolmanneksi eron voi aiheuttaa se, että myöhemmin koulun aloittavat lapset tekevät kokeet vanhemmalla iällä. Koska en pysty erottelemaan eri mekanismeja toisistaan, tulokset tulee tulkita näiden kaikkien mekanismien yhteisvaikutuksena.

Asiasanat: koulutus, koulunaloitusikä, regressioepäjatkuvuusmenetelmä

JEL-luokat: I21, I28, J13

Contents

1	Introduction	1
2	Literature Review	3
3	The Education System in Finland	8
4	Data	11
5	Methodology	16
5.1	The Potential Outcomes Framework	16
5.2	Sharp Regression Discontinuity Design	17
5.3	Fuzzy Regression Discontinuity Design and LATE	20
5.4	Estimation	25
6	Validity of the Setting	27
7	Results	32
7.1	Graphical Evidence	32
7.2	Estimation Results	33
7.3	Sensitivity and Robustness Inspections	39
7.4	Heterogeneous Effects	43
8	Conclusion	50
	References	52

1 Introduction

In Finland, children start school during the year they turn seven, which is fairly late compared to most EU countries.¹ Probably partly because of this there is frequently discussion on whether the school starting age should be lower.² Arguments exist for and against lowering the school starting age. For instance, some argue that if children would start school at a younger age, they would enter the work force earlier, which could have a positive impact on their lifetime earnings. However, if the school starting age is lowered and number of years of compulsory education is held constant, children leave school at a younger age. This again may be harmful for children who have difficulties to transfer from comprehensive school to secondary schools or working life. Another, and maybe more relevant, point stated by supporters of a lower school starting age is that starting school at a younger age would be especially beneficial for the children with lower socioeconomic background, who are more likely to suffer from worse childhood environment. On the other hand, some researchers argue that play is essential for the child's development and if formal schooling is started too early, the school may disturb the development process.

As the list of possible mechanisms is long and contains factors that have opposite effects, it is impossible to theoretically ascertain what is the optimal age to start school. Thus, empirical work is needed to assist policy making. However, identifying impacts of earlier or later school starting age is empirically challenging due to endogeneity issues. In this particular case, endogeneity issues emerge from the fact that school starting age is correlated with learning disabilities because parents tend to postpone the school start of children with learning disabilities. Therefore, in order to provide a causal evidence for decision makers, researchers have to find ways to overcome this selection issue. One solution is to make use of exogenous variation in school entry age created by school starting rules. If birthdays are independent of the school starting rule, it follows that some children are as if randomly assigned to start school at different ages. For example in Finland, children born just after or before the New Year are on average similar, but start school with a one year age difference.

Although a study taking advantage of school starting rules may be able to extract causal evidence, the policy advice still depends on the actual mechanisms. The relative school starting age may affect educational outcomes in at least three different ways. Firstly, a child who is older than other children in her class due to the school entry rules may do better because she has started school at the optimal age in terms of human development. Alternatively, being older than the peer group may add confidence and hence improve the schooling outcomes. The third possibility is that relatively older children achieve better results just because they do the exams at the older age. What can be said about the

¹For a good list, see for example EU (2016).

²For the most recent example, see for instance Semkina (2016).

school starting age depends on whether researchers are able to distinguish between different channels. Nonetheless, a study which estimates the combined effect of all channels, should be used to inform parents who are considering whether they should postpone their child's school start. But in the case where the researchers are able to separate the channels from each other, the results may give us a hint about the optimal school starting age.

There exists a rich literature using school starting age rules to study the effects of school starting age on various outcomes. A general finding is that children who start school at an older age tend to do better in exams (Bedard and Dhuey, 2006). Less is known about the actual mechanisms. A few studies are able to distinguish the effects of school starting age and the age-at-test effect from each other with the result that the age-at-test effect seems to explain most of the deviation (Black et al., 2011; Crawford et al., 2010). This is supported by studies which follow individuals for a relatively long time period or even over their life-cycle and show that school starting age tends to have small effects on educational attainment and earnings (Fredriksson and Öckert, 2013).

Similarly as in some previous studies, I use exogenous variation in school starting age arising from school starting rules to study the effects of school starting age (SSA) on educational outcomes. In Finland, children start school during the calendar year they turn 7. Therefore, children born just before and just after New Year are on average very similar but start school at very different ages. I use a regression discontinuity design to study whether the discontinuous jump in the school starting age has an effect on the grade point average (GPA) at the end of comprehensive school (*peruskoulu*). Furthermore, I examine if the school starting age affects the probability of admittance and graduation from general upper secondary school (*lukio*). My main data source is the joint application register of the Finnish National Board of Education, which includes each individual who has applied to upper secondary school. I restrict my analysis to the individuals who have applied to upper secondary school during the year they graduate from comprehensive school. I link the sample to the Finnish Longitudinal Employer-Employee database, from which I acquire information on whether the individuals have graduated from general upper secondary school.

I find that the discontinuous jump in the school starting age has a significant effect on educational outcomes at the end of comprehensive school in Finland. I estimate the effect using a bandwidth of 30 days on both sides of New Year and show that those who are born just after the cutoff have on average a 0.15 points higher GPA compared to those born just before. I do not observe the exact school starting year, but I approximate the school starting age from the graduation year. When non-compliance is taken into account, the estimate is even larger. In addition, a later school starting age increases the probability of admittance and graduation from general upper secondary school. The results can be taken as a causal effect of school starting age rule. To back this up, I show that density of assignment

variable and children's background variables behave smoothly at the cutoff. Finally, I study the effect conditional on various background characteristics and show that the effect is significantly larger for girls than boys. I cannot confirm the source of deviation, but it may be because parents postpone boys' school start more easily. In any case, the finding is in the line what Fredriksson and Öckert (2013) document.

This paper contributes to the literature at least two ways. First, I show that the magnitude of the effect of school starting age on GPA is similar to what has been observed in other countries (Bedard and Dhuey, 2006; Elder and Lubotsky, 2009). This differs from what Bedard and Dhuey (2006) and Pehkonen et al. (2015) find, who document smaller effects in Finland. As my data set is extensive and covers most of the individuals graduating from comprehensive school, my results should give a better picture of the true effects than previous studies regarding Finland. Moreover, although I cannot distinguish between different channels, my estimates suggest that the relative school starting age may generate persistent gaps between children in Finland. However, future research which looks over the individual's life-cycle is required.

The structure of this article is the following. In the second section, I review previous studies regarding school starting age. Section 3 briefly introduces the Finnish education system and highlights some features which are especially relevant in my setting. Then in section 4, I explain how the data set and the main variables of the analysis are constructed. Section 5 provides motivation for the regression discontinuity design and explains in detail how the method is applied in this paper. In section 6, I examine the validity of my setting. Section 7 reports the results from the graphical analysis and the model presented in section 5. In addition, I examine the robustness of the findings and heterogeneity in the results. Lastly, in section 8 I present concluding remarks.

2 Literature Review

In this section, I introduce the literature studying the effects of school starting age (SSA). This article examines the effects of relative school starting age on educational outcomes and hence the main emphasis of the review is on similar research. Since one fundamental reason to acquire education is to influence other outcomes, I shall briefly present studies exploring the effects of school starting age on earnings and crime as well.

In Finland, a child who is born in January is on average absolutely and relatively more older than other children in the class. The literature lists many mechanisms through which the age difference could affect educational outcomes. For instance, children who are older in absolute terms may be ahead in

the development process and have better self control, which may contribute positively to educational outcomes. Alternatively, some suggest older students do not learn at a faster rate, but they do better in exams just because they have had more time to accumulate knowledge. According to this so-called age-at-test hypothesis, gaps in learning outcomes are mainly related to factors outside of school. Thus, the proponents of age-at-test hypothesis state that the deviation should be large during the first grades and diminish as the children get older.

Relative maturity explanations rely on peer group effects. In which direction the peer group pushes educational outcomes is not clear. Being the oldest child in class might improve the child's confidence and hence have a positive effect on educational outcomes. On the other hand, the success of older children may encourage younger children to work more, which may have a positive impact on long-run outcomes. Although some studies (Black et al., 2011; Crawford et al., 2010) are able to isolate these channels from each other, doing so is difficult. Hence, researchers usually estimate the combined effect of all channels.

Because the different mechanisms might have contrasting effects on the future outcomes of the child, it would be important to know which of the channels dominates. Cunha and Heckman (2007) have built a model of skill formation based on empirical observations. In the model, adulthood skills are the product of natural abilities and investments in skills that are made at different stages of childhood. Essential aspects of the model are self-productivity and dynamic complementarity of the skills. Self-productivity of skills means that the accumulation of skills in early periods increases the accumulation of skills in later periods. From dynamic complementarity it follows that early and late investments complement each other. In order to make productive investments at later stages, the individual needs a sufficient level of early investments. Similarly, early investments are more productive if the investments in later periods are large enough. In the model by Cunha and Heckman, the peer effects or differences in learning rates stemming from the age difference may generate skill gaps early. Furthermore, if the complementarity of early and later investments is high, it could be difficult to close the gap in skills at older ages.

In the case where observed deviations in educational outcomes are not caused by differences in learning rates or peer effects but are instead the result of the age-at-test effect, the implications of school starting age rules should not be that concerning. What is crucial is how quickly the relatively younger children catch up with the older children. For instance, in Finland school achievements play a major role in the selection process to upper secondary schools, which works as a route to academic studies. Therefore it would be worrisome if younger children were still behind at the stage where they seek into different tracks.

Up to now, several studies have investigated the effects of the school starting age on educational

outcomes or skills. The general conclusion is that children who start school older due to the school starting rules perform better. In one of the first quasi-experimental studies, Bedard and Dhuey (2006) use an instrumental variable approach to study the causal effects of the school starting age in OECD countries. They find that relatively older students perform better in the standardized math and science tests taken at the fourth grade in most of the countries. Relatively older students score 0.2-0.4 standard deviations higher than the youngest in the math test. In the science test, the corresponding effect is 0.2-0.4 standard deviations. The effect is still present in the eight grade, though the magnitude is smaller and not statistically significant in each country. In the countries where the effect is statistically significant, the difference between the relatively oldest and youngest is 0.13-0.39 standard deviations in math and 0.16-0.4 standard deviations in science.

For this paper, the results regarding Finland are especially interesting. Bedard and Dhuey (2006) find that compared to most of the countries, the effect is much smaller and not statistically significant in Finland. The gap between the youngest and the oldest children is 0.06 standard deviations in the math test and 0.13 standard deviations in the science test at the eight grade level. The results are somewhat similar to what Pehkonen et al. (2015) document. They use data from the Cardiovascular Young Finns Study and compare children who are born in December and January. They find that children who are born in January have higher self-reported GPA in the sixth grade. However, the effect is smaller and not statistically significant in the ninth grade.

Studies focusing on a single countries seem to support the findings of Bedard and Dhuey (2006). Elder and Lubotsky (2009) use an instrumental variable strategy with U.S. data and show that children who start school later achieve better test scores in math and reading, but that the effect decreases as the children proceed to higher grades. At the eight grade level, the oldest children score 0.22 standard deviations higher in reading and 0.16 standard deviations higher in math than the youngest children. Furthermore, they examine the magnitude of the effect conditional on socioeconomic status and find that the difference between the youngest and oldest children is larger among those with high-income parents. In another study, Puhani and Weber (2007) investigate the effects of the school starting age on educational outcomes in Germany. Germany is an interesting case because they start tracking students already after the fourth grade. The authors document that the oldest students score 0.4 standard deviations higher than the youngest in an international reading study at the fourth grade level.

It is interesting that although Bedard and Dhuey (2006) and Elder and Lubotsky (2009) are not able to distinguish between different mechanisms and the results are similar, the authors interpret their findings a bit differently. Bedard and Dhuey (2006) also find that relatively older students are more likely to attend university in the U.S. and Canada. Because the school starting age affects adulthood

outcomes, they argue that the complementarity of skills may partly explain the differences in learning outcomes. By contrast, Elder and Lubotsky (2009) interpret their results to support the age-at-test hypothesis. They state that since the difference shrinks relatively quickly and is larger among the children of high-income parents, the gap is mostly caused by knowledge accumulated before school.

As I stated earlier there exist a few studies which are able to separate the age-at-test and school starting age effects from each other. Black et al. (2011) conduct a study using data from the IQ test of the Norwegian armed forces. The test is taken when the individuals are approximately 18 years old and each year there is a certain threshold which defines when an individuals should do the test. Importantly, the cutoff is different from the one which indicates when an individual should start school. The authors take advantage of this variation in the cutoffs to distinguish between different mechanisms. According to their findings, when absolute age is controlled for, the children who start school younger perform slightly better in the test. When both mechanisms are taken into account, the late starters tend to score 0.08 standard deviations higher. Crawford et al. (2010) provide similar results using English data. In England, local authorities make decisions regarding the school admission rules, which creates regional variation in the school starting age. The researchers use this regional variation to study how different kinds of channels affect the results at national Key Stage tests, which are taken at specific ages. They use a regression discontinuity design and find that children who start school younger perform worse in the national test and that the age-at-test effect seems to explain most of the difference.

Since many studies report that later the school starting age increases test scores, the logical next step is to present studies investigating the effect of school starting age on educational attainment. I find that the literature is somewhat mixed and not as rich as in the case of school starting age and test scores. One reason is that, in order to obtain credible estimates, the individuals have to be followed for a relatively long time. Fredriksson and Öckert (2013) use a regression discontinuity design and Swedish register data. They are able to track individuals over their life cycle and show that on average late starters accumulate 0.159 more years of schooling. The magnitude of the effect is bigger for females and individuals with low educated parents. With a similar approach, Black et al. (2011) discover a very small positive but statistically insignificant effect in Norway. By contrast, Dobkin and Ferreira (2010) find that in Texas and California individuals who start school younger tend to acquire more education.

Fredriksson and Öckert (2013) offer a plausible explanation for why the results differ between the studies. Their study uses data from the period when the Swedish school system used to track students early, which may increase the effect. By contrast, Black et al. (2011) include cohorts who started school after the Norwegian comprehensive school reform, which removed early tracking. The reason

why Dobkin and Ferreira (2010) find an effect with a different sign could be an implication of the school leaving rules. In Nordic countries, compulsory education ties individuals for a long time. As a consequence, most students have completed comprehensive school before they even have an option to drop out. As Angrist and Krueger (1991) show. By contrast, in U.S. individuals who start school older may drop out relatively earlier. Thus, it is unclear whether the result of Dobkin and Ferreira (2010) is caused by the school starting or leaving rule. Altogether, these studies point that the institutional context may have a big effect on how the relative school starting age affects long-term outcomes.

The evidence regarding the link between school starting age and educational outcomes is relevant, but it would be even more important to know, whether these differences in educational outcomes influence other factors such as crimes and earnings. There are a few studies which investigate the relationship between school starting age and youth crime. Landersø et al. (2016b) study the effect of school starting age on the probability to commit crime in youth. They use a regression discontinuity design with Danish data and find that higher relative school starting age decreases the probability to commit crime before the age of 19 by 2 percentage points for girls and by 5 percentage points for boys. Cook and Kang (2016) use a very similar approach to study effects of school starting age on juvenile delinquency and adult crime in the U.S. The authors also use a regression discontinuity design and find that children who start school older are 2.8 percentage points less likely to commit crime at the age of 13-15. However, the effect seems to reverse as the individuals reach adulthood. Cook and Kang estimate that SSA increases the probability to commit crime by 0.80 percentage points between ages 17 and 19. Again, the differences in the results may be due to the school leaving rules. Both Landersø et al. (2016b) and Cook and Kang (2016) report that a higher school starting age improves educational outcomes, which probably has a negative impact on youth crime. But as I mentioned earlier, a higher school starting age increases the probability to drop out of school in the U.S, which again may have a positive effect on youth crime. So as Cook and Kang (2016) reason, the effect on drop out probability may outweigh other desirable effects of a higher school starting age, which may explain the increase in crime rates at ages of 17-19.

Last, I present studies exploring the relationship between school starting age and earnings. Black et al. (2011) follow Norwegian individuals from age 24 to 35 to examine the effects of school starting age on earnings. They show that a higher school starting age decreases earnings until age 30 and after that the gap disappears. Fredriksson and Öckert (2013) go further with Swedish data by tracking individuals over their life cycle. Interestingly, the school starting age affects mainly how individuals allocate their labour supply over their life cycle. During the first years of working life, higher school starting age decreases earnings, but the effect on prime-age earnings is negligible. However, from age 55 to retirement, the effect is reversed compared to early years; individuals who have started school

later tend to earn more. To put it differently, it seems that people who start school later tend to the enter workforce later and thus have less experience, which decreases earnings at the beginning of the career. For some reason, they stay longer in the workforce and earn more at older ages. On average, a higher school starting age decreases earnings over life the cycle although the effect is small.

To sum up, the evidence indicates that there exist positive links between relative school starting age and educational outcomes. Nevertheless, it is difficult to form a coherent conclusion regarding the mechanism behind the results. Some studies (Black et al., 2011; Crawford et al., 2010) are able to isolate the different mechanisms from each other and the results suggest that the age-at-test effect is behind the differences. In addition, Elder and Lubotsky (2009) argue that since the test scores converge, the main cause is the skills learned prior to school. By contrast, Bedard and Dhuey (2006) underline the fact that people who start school at an older age due to school starting rules still do better during later grades and are more likely to attend college. This indicates that the differences in learning rates or peer effects are also relevant.

3 The Education System in Finland

In Finland, compulsory education begins during the calendar year a child turns seven and ends either when the nine years of basic education have been completed or after ten years have passed since the start of compulsory education. However, the law is flexible in this regard as the child may start school later or earlier. The decision about a different school start is made by the local school provider and is based on psychological and medical examinations.³ (Basic Education Act, 1998.)⁴ Similarly to other Nordic countries, one implication of the school admission law is that, in the beginning of the first grade, children who are born just before New Year are on average one year younger than children born just after New Year.

Before comprehensive school, a majority of children attend day care, which is offered in the day care centers or family day cares. Heavily subsidized day care is provided by private and public operators. Although day care is not free, every child has a right to a place in day care. The last year of early childhood education is usually spent in preschool, but it has been mandatory only since 2015. Basically, day care attendance may have a substantial effect on the estimates. For example, proponents of the age-at-test effect argue that the differences in test scores are mainly due to factors outside of the school. Thus, it would be interesting to know children's early childhood education histories and examine whether the behaviour of children born just before and after the cutoff differ prior to school.

³Before 1999, the decision about earlier or later school start was made by the municipal school board.

⁴Perusopetuslaki in Finnish.

Unfortunately, I do not observe whether the child has participated in day care or preschool.

Finnish basic education consists of nine years of comprehensive school. Usually, the first six grades are spent with the same class teacher while during the last three grades, each subject has its own teacher. The content of education is set by the Finnish Government and the Finnish National Board of Education, though local authorities, schools and teachers have reasonably much freedom in planning and arranging the school day and year. There are no tuition fees at any stage of education and, in addition, in comprehensive school all the materials, transportation and lunch are free as well. Almost all children in Finland complete basic education. For instance, according to the Finnish National Board of Education, currently only 0,1 % of cohorts do not graduate from comprehensive school. At the end of ninth grade, each student who has passed all subjects receives a basic education certificate, which contains a numerical or verbal assessment in each subject. An important part of the certificate is the grade point average (GPA) for theoretical subjects,⁵ because it is used as one of the main criteria in the selection process to the upper secondary schools.

After basic education most individuals continue to general or vocational upper secondary school.⁶ The selection process to general upper secondary schools is based on the GPA of the basic education certificate, whereas the process to vocational schools may also include additional factors. General upper secondary school (*lukio*) that is comparable to high school is a more academic track and usually lasts three years. At the end of general upper secondary school, the student takes part in national matriculation examinations, where she is tested in at least four different subjects. Vocational upper secondary school (*ammattikoulu*) offers more practical training in various fields. The curricula are designed to take three years to complete as well.

Higher education is offered in universities and applied universities. A person who has completed general or vocational upper secondary school is eligible to apply to university, but most of the accepted students come through general upper secondary school. Depending on the subject and university, students are chosen through an entrance exam and based on their matriculation examination GPA. Usually the students are directly admitted to study for both their Bachelor's degrees and Master's Degrees, which should last a total of 5 years. However, on average it takes around six years to complete a Master's Degree in Finland. Universities of applied sciences (*ammattikorkeakoulu*) are aimed at providing tertiary education more directly addressing the necessities of labor markets. Similarly as in the case of traditional universities, the selection is based on entrance exam and the matriculation examination GPA but other factors like work experience may be taken into account as well. Degrees are structured

⁵Theoretical subjects consist of mathematics, mother tongue, second native language, first foreign language, physics, chemistry, biology, geography, health education, religion, history and social studies.

⁶Between years 2000-2007, on average, 91,5 % of the students who graduated from comprehensive school continued to upper secondary school.

to take 3-4 years to complete with a possibility to continue to university to obtain a Master's Degree.

A few features make Finnish education system favourable for a study investigating the effects of relative school starting age. First, student retention during basic education is highly uncommon. For instance, according to the Finnish National Board of Education, only 0,49 % of students were held back in comprehensive school annually between years 2000-2007. This should guarantee that most of the individuals have spent the same number of years in comprehensive school. Another good aspect is that postponing school start or earlier enrollments to school are relatively rare. Unfortunately, there does not seem to be any exact information about the total number of individuals who do not start the school during the year they should. However, I have obtained information about the number of children who postpone school start from Statistics Finland. Based on their statistics, around 2,1% percent of individuals in each cohort have delayed the start of school between the years 1995 and 1999. Since people obey school starting rules well, it implies that it is meaningful to compare children born at the end and start of the year. Lastly, as compulsory education extends to a relatively high age in Finland, my results should be caused by the school starting rule and not by the school leaving rule.

Although the Finnish education system offers a good setting, I want to highlight a detail, which should be noted when interpreting the results. According to the Finnish National Board of Education, the students' final assessment should be nationally comparable and equal, but there are a couple reasons why the grades in the basic education certificate should be treated with caution. Firstly, there are no national final exams in comprehensive school, and hence teachers evaluate students based on the guidelines provided by the authorities. The relatively loose final assessment guidelines only define the competence level the student should possess to obtain grade 8 (good). But as the grade scale is from 4 to 10, it is not clear on what basis teachers assign grades or how much cognitive or non-cognitive skills contribute to the grades.

Moreover, the assessment guidelines have varied a lot over time. When comprehensive school was established in 1972, the general instruction was to compare children inside the class and base the grading on the distribution of performance of children in the class. Later, this was seen to contradict with the objectives of comprehensive school and hence the practice was abandoned. According to the next set of guidelines, the aim of the assessment was to compare the individual student to national and regional targets. But it was only in 1999 that the final assessment was separated from the continuous evaluation during the studies. At the same time, the national criteria for a grade of 8 were set. In brief, since instructions have varied a lot and are still imprecise, there is reason to doubt the national comparability of grades.⁷

These concerns are supported by empirical evidence. Ouakrim-Soivio (2013) uses data from the na-

⁷This paragraph is based on article located on the webpage of Finnish National Board of Education (2016b).

tional assessment of student learning achievements conducted by the Finnish National Board of Education to study how results in the achievement tests are related to school grades in social sciences.⁸ She finds that students who achieve very similar scores in the national tests may be given very different school grades in social science in different schools. In another study, Harju-Luukkainen et al. (2016) inspect how self-reported school grades in mother tongue predict reading literacy performance in the Pisa 2009 test. They show that school grades explain on average quite decently the variation in the Pisa scores, but there is substantial regional variation in the explanatory power. In some regions students tend to notably underperform and in some regions overperform in the Pisa test relative to their school grades.

To sum up, the Finnish comprehensive education system offers a suitable context for this paper, because school starting rules are obeyed diligently and retention rates are low. However, the lack of national standardized exams and loose grading instructions makes it difficult to interpret the results for GPAs. On the other hand, GPAs plays a major role in selection process to upper secondary school. There is evidence which shows that being rejected or not receiving place to preferred option in the selection to upper secondary school, has significant negative effect on the probability to receive any degree after comprehensive school (Virtanen, 2016). Thus, any observed deviations between children born at the end and start of the year are important.

4 Data

Next, I describe how I have constructed the data I use. In addition, I explain in detail how the main variables of the analysis are defined. Lastly, I discuss limitations of the data.

My main data source is the Application Register of the Finnish National Board of Education (2016a), which contains information on each individual who participates in the joint application process to upper secondary school. From this register, I obtain information on each individual's GPA from the basic education certificate, where each individual applies after comprehensive school, and to which school the individual is admitted. I restrict my analysis to the individuals who take part in the joint application process in the same year they finish comprehensive school during the years 1991–2007.⁹ The application register does not contain information on birthdays, but Statistics Finland has provided a data set which contains the exact birthday of each individual. I link the individuals in my application register sample to the Finnish Longitudinal Employer-Employee Data (FLEED) (Statistics Finland, 2016)

⁸Nowadays, these national assessments are carried out by Finnish Education Evaluation Center. Normally the sample covers around 5 to 10 percent of the cohort.

⁹The register also includes individuals who graduate from comprehensive school but do not take part in the joint application process. However, I exclude these individuals from the sample because I do not observe their GPA.

Table 1: Descriptive Statistics

Variable	Discontinuity sample		Total sample	
	Mean	St. deviation	Mean	St.deviation
GPA	7.638	1.142	7.632	1.142
High school graduate	0.531	0.499	0.533	0.499
Father's education	0.147	0.354	0.146	0.353
Mother's education	0.121	0.326	0.121	0.326
Father's earnings	5.488	2.874	5.507	2.864
Mother's earnings	5.491	2.880	5.515	2.867
Female	0.492	0.500	0.490	0.500
Observations	155,359		1,008,297	

Notes: Discontinuity sample refers to the individuals who are born ± 30 from New Year.

using individual-specific unique encrypted identification numbers. The FLEED provides information on educational attainment and gender. Furthermore, using FLEED, I can link individuals to their parents and obtain information about parents' earnings, parents' educational attainment and families' housing arrangements.

I use children's redefined birthday as an assignment variable, which indicates whether the child belongs to the treatment or control group. This assignment or running variable R_i , measuring the distance of the birthday from New Year, is constructed by normalizing the birthday variable to be around zero. It takes non-negative values if the person is born between January and June and negative values otherwise. For instance, for the individuals who were born on the first of January, the assignment variable is 0 and for those who are born on the 31th of December the variable is -1. Notice that the reason birthday is normalized to be around zero, should become apparent in methodology section 5.

Unfortunately, I do not have information on the exact school starting age. However, I have information on the age at which the individual has graduated from comprehensive school. Thus, I approximate the school starting age in a manner similar to Landersø et al. (2016b). I define an indicator variable which measures whether the individual is "old" in the class. An individual is defined as old if she was born between January and June and the observed graduation year that is at least 9 years after the hypothesized school starting year. Another group defined as old are the people who are born between July and and December, and have an observed graduation year that is at least 10 years after the theoretical school starting year. It is clear that individuals who are held back during comprehensive school, cause a measurement error in my school starting age approximation, but because grade retention is relatively rare in Finland, it should be a good approximation. In any case, I suppose that my

approximation overestimates the number of people who start the school older. This may lead to an upward bias in the instrumental variable estimates.

The analysis has three main outcome variables. First is the GPA of theoretical subjects in the final school report, which I observe directly in the joint register. The scale of the grades is from 4 to 10, where 4 stands for failed and 10 for “excellent”. The second variable measures admission to general upper secondary school. The joint application register contains information on whether the applicant has received admission to upper secondary school. Based on this, I build an indicator variable which takes value 1 if the individual has received a place in a general upper secondary school and zero otherwise. Lastly, the FLEED provides information on whether the individual is an upper secondary school graduate. Again, I define an indicator variable that measures if the individual is an upper secondary school graduate.

There is one notable difference between the first two outcome variables and the last outcome variable. GPA and admission to general upper secondary school are measured in the year the individual graduates from comprehensive school. In principle, individuals could stay in comprehensive school for an extra year to improve grades or apply again to general upper secondary school, but this is not captured by the first two variables. By contrast, for the graduation variable I go through every year in the FLEED between 1991-2012 to check whether the individual has completed the matriculation examination and received the general upper secondary school certificate. Thus, the third variable measures more reliably any persistent effects the school starting age may generate.

To inspect the validity of my setup, I construct five background variables using information from the FLEED. The first two variables measure whether the parents have completed higher education. The third variable indicates if both parents are still living with the child. Finally, I study how the parents’ earnings evolve around the cutoff. The first three background variables are measured when the individual is 15 years old. Variables representing mother’s and father’s earnings are defined a bit differently. I observe the annual earnings for parents in the years the child was 14, 15 and 16 years old. To mitigate the effects of transitory variation in the earnings, I first calculate the average of these annual earnings. Then I use this average to specify to which income decile group the parents belong in the year the child graduates from comprehensive school.

There are a few problems with the data that are mostly related to missing variables. In the joint application register, some observations have missing values and some individuals appear multiple times. Each year around 1500 observations have GPA missing. Because the GPA is the main variable of my analysis, I drop these observations. Conveniently, most of the duplicates disappear at the same time. Yet, there are still around 7 duplicate observations in each year. With these observations, my procedure is the following: If the duplicates have the same GPA, I drop the observations which have

more missing values. In the case where duplicates have different GPAs, I drop both observations. Furthermore, I exclude immigrants, since I do not know in which kind of system they have started school.¹⁰ Finally, there are three observations, which I am not able to link to FLEED and I leave these observations out from the analysis. To sum up, after the steps I have described above, there are 1,050,378 individuals in my data set. Column 3 of table 2 presents the number of individuals in each cohort, after the procedures I described above.

In addition, some observations have missing values regarding parents. I also exclude these observations from the analysis. This decreases my sample size by 4 percents and leaves me with 1,008,297 observations. Column 4 of table 2 shows the number of individuals in each cohort in the final sample. From this final sample, my analysis uses those who are born within 30 days of New Year. Table 1 provides descriptive statistics of this discontinuity sample. In brief, based on the background characteristics presented in table 1, the total sample and discontinuity sample are very similar.

A final concern stems from the timing of the measurement of background characteristics variables. The background characteristics are from the years the individuals were 14, 15 and 16 years old. Angrist and Pischke (2009, 64-68) discuss the use of the control variables and causal interpretation. They define a bad control variable as a one that is determined after the main explanatory variable and could also work as an explanatory variable. In my case, bad controls are worrisome because I use background characteristics to evaluate if individuals who start school at different ages are, on average similar. Thus, bad controls may lead to misjudgment. For instance, Landersø et al. (2016a) present some evidence suggesting that the school starting age may affect the timing of parental divorce, which means that my variable regarding parents' housing arrangements could be troublesome. On the other hand, if the earnings were measured before the child's birth, it is possible that potential earning differences are not even visible yet. In this light, parents' educational attainment variables should be the most reliable ones, since education levels are usually determined before the child's birth.

Although I have to drop some variables, the data set I use is comprehensive and covers most of the individuals who graduate from comprehensive school. Of course, if missing information is correlated with the socioeconomic status, the exclusion of observations may bias my results. To rule this out, I perform sensitivity checks with the sample which include individuals who have missing information on parents in section 7.3. Also, a small portion of individuals never take part in selection process or even graduate from comprehensive school.¹¹ Based on previous findings on relative school starting age, this could be more likely for individuals born at the end of year, which may decrease my estimates. Luckily

¹⁰I do not directly observe if a person is immigrant. However, I use the child's first language as a proxy and exclude the observations whose first language is not Finnish or Swedish.

¹¹According to Virtanen (2016), the number of individuals who do not take part in selection process during the year they graduate is around 2%.

Table 2: Size of the Cohorts

Cohort (1)	Graduates (2)	Graduates who apply (3)	Final sample (4)
1991	64 175	59 579	55 538
1992	65 634	62 868	59 152
1993	65 483	64 595	61 375
1994	64 297	63 111	60 152
1995	63 756	62 546	59 382
1996	63 514	62 038	59 154
1997	64 247	60 634	58 079
1998	66 726	64 881	62 223
1999	67 043	65 079	62 603
2000	65 937	62 978	60 581
2001	63 747	60 528	58 285
2002	61 450	58 309	56 170
2003	60 831	57 626	55 628
2004	63 828	60 421	58 961
2005	63 755	60 488	59 091
2006	65 783	62 467	61 099
2007	65 568	62 230	60 824
Total	1,095,774	1,050,378	1,008,297

Notes: Column 2 presents the number of individuals graduating from the comprehensive school annually. The numbers are obtained from the Statistics Finland (2009). Column 3 shows the number of individuals who have graduate from comprehensive school and take part in the joint application process in my dataset. Column 4 presents my final sample and consists of remaining individuals after excluding those who have missing information on parents.

most of the individuals complete comprehensive school and take part in the selection process.

5 Methodology

Like most previous studies of this topic, I use a regression discontinuity design to study the causal effects of school starting age. In this section, I first provide the motivation for the regression discontinuity design and explain generally how it is used in empirical economics. I then present the model I use to estimate the effects.

5.1 The Potential Outcomes Framework

In causal research, we would like to know what would have happened to the people who were affected by some policy if they had not been affected by the policy. For instance, a study examining the effects of school starting age would like to investigate what would have happened to a child who started school at the age of seven had she started school at the age of six instead. One way to carry out the analysis would be just to regress an educational outcome on the observed school starting age. But are those with a higher observed school starting age a good counterfactual for those with a lower observed school starting age? Or in other words, would those who were treated perform similarly as the non-treated in the absence of treatment? Probably not, since children who postpone school start are more likely to have learning disabilities which could contribute to the educational outcomes. Thus, a naive analysis comparing children with different observed school starting ages, cannot distinguish whether the differences in the educational outcomes are caused by school starting age, or by learning disabilities. As the analysis cannot separate these two effects from each other, naive regression is very likely to provide biased estimates on the effect of school starting age. This bias is called selection bias.

The issues regarding selection bias can be illustrated more formally using a similar approach to Angrist and Pischke (2009, 12-15). Let us assume we are studying the effects of school starting age on educational outcomes Y_i with observational data and we have a binary explanatory variable $D_i = \{0, 1\}$ which takes value one if individual i starts school at the age of seven or older and zero otherwise. For each child, there are two potential outcomes

$$\begin{aligned} \text{potential outcome } Y_i &= \begin{cases} Y_{1i}, & \text{if } D_i = 1 \\ Y_{0i}, & \text{if } D_i = 0 \end{cases} \\ &= Y_{0i} + (Y_{1i} - Y_{0i})D_i, \end{aligned} \tag{1}$$

where Y_{1i} is the educational outcome if the child i starts school at the age of 7. If the same child would start school younger than at the age of 7, the outcome variable would take on value Y_{0i} . Of course, for

each individual, we only observe one outcome. Thus, if we want to study the effect of school starting age on educational outcomes, we have to compare the means of the groups with different treatment statuses.

The simple comparison of averages of the two groups can be written as follows

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0], \quad (2)$$

where the first part of the right hand side equation, $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$, represents the average causal effect of a higher school starting age on the treated, and the second part of the right hand side, $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$, refers to selection bias. According to equation 2, the average observed differences in educational outcomes are caused by the school starting age and selection bias which stems from the differences in the background characteristics. Depending on the setup, selection bias may increase or decrease the estimate. In this case, the effect is probably negative, since as I mentioned earlier, children who have learning disabilities are more likely to postpone school enrollment. Thus, those who started school at an older age would on average as a group do worse in the case they were not treated, $E[Y_{0i}|D_i = 1]$, than those who were not treated in the first place, $E[Y_{0i}|D_i = 0]$, which implies that the selection bias, $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$, is negative and hence decreases the observed average difference.

The best way to deal with selection bias from the perspective of internal validity would be a randomized controlled trial. In a randomized controlled trial, individuals are randomly assigned to a treatment and a control group. Because the assignment is random, the treatment status should be independent of the background characteristics of the groups. This implies that, if the process is done properly with a large sample, due to the law of large numbers, the groups should on average be similar. Therefore, if the treatment group starts school at the age of seven, the best guess of what would have happened to them had they not received the treatment would be provided by studying how the control group performs. Thus, differences in the average educational outcomes between groups can be interpreted as the average causal effect of a later school starting age. Although randomized controlled trials are nowadays quite common, a study that would force some people to start school later or earlier would most likely be considered both unethical and excessively expensive.

5.2 Sharp Regression Discontinuity Design

Because of the budget constraints and ethical issues facing them, researchers have to look for so-called natural experiments to overcome the selection issues in causal social science research. Natural experiments are situations where individuals are randomly assigned to treatment and control groups by a

natural process. As the name suggests, the source of assignment may literally be nature. For instance, Angrist and Evans (1998) use exogenous variation in number of children created by twins or children’s gender mix to study the causal effect of having children on parents’ labor supply. Alternatively, variation in treatment status can emerge from the actions of people or society. In another influential paper, Card and Krueger (2000) examine how employment evolved in the state which increased minimum wage compared to the states which did not increase it. The use of natural experiments and quasi-experimental methods has been in a vital part in the so-called credibility revolution in empirical economics.¹² One of these methods, which utilizes natural experiments, is the so-called regression discontinuity design.

The basic idea of a regression discontinuity design is to look for rules which force some individuals to be treated and some untreated by a policy. For instance in Finland, a child’s compulsory education begins during the year the child turns seven. This creates a discontinuous jump in the school starting age at the end of the year. Children born at the end of December start school at the age of 6.7 whereas children born at the beginning of January start at the age of 7.7. To some extent, birthdays should be randomly determined, which implies that people born close to the threshold are as if randomly assigned to start school at a different age compared to those on the other side of the threshold. If the assignment is approximately random, predetermined background characteristics should evolve continuously around New Year. Thus, individuals just below the cutoff provide a good counterfactual to those just above the cutoff. By comparing these groups, we can study the causal effects of the discontinuous jump in school starting age on educational outcomes.

The effect of the discontinuous jump can be estimated using a regression model. The model may take various forms, but I present just a simple linear form in this introductory section. Let us for a moment assume that the law is completely binding. Based on the reasoning above, let us define a variable that indicates whether the individual is on average older or younger than other children in the class. The variable takes the form of

$$D_i = \begin{cases} 1 & \text{if } D_i = R_i \geq c \\ 0 & \text{if } D_i = R_i < c \end{cases}, \quad (3)$$

where the continuous assignment variable R_i stands for the redefined birthday for individual i , c is the cutoff value and treatment D_i is defined as a discontinuous function of R . We can estimate the effect of a discontinuous jump in school starting age on educational outcomes Y_i using the following linear

¹²Angrist and Pischke (2010) provide a good introduction to the factors contributing to this so-called credibility revolution.

regression model

$$Y_i = \alpha + f(R_i) + \theta_1 D_i + e_i, \quad (4)$$

where e_i is the error term and $f(R_i)$ describes the relationship between R_i and educational outcomes. The main parameter of interest is θ_1 , which captures the effect on the educational outcomes caused by the discontinuous jump in the school starting age. It has a causal interpretation if the following conditions are met.

The first condition is that the assumption regarding manipulation has to hold. According to this assumption, individuals must not be able to accurately define the value of running variable (Lee and Lemieux, 2010). In the setting of my article, the assumption is not violated if parents are unable to precisely determine the birthday of the child, which seems quite reasonable. When the no-manipulation assumption holds, individuals near the cutoff have equal chances to end up just below or just above the threshold and thus the treatment is approximately locally randomized. As a result of local randomization, all other predetermined background characteristics should behave smoothly around the cutoff and hence the groups just below and above the threshold are on average similar, but are exposed to different kinds of treatment. In other words, as Lee and Lemieux (2010) state, if there is no manipulation present, a regression discontinuity design is comparable to a randomized controlled trial and solves the problem of selection bias. Thus the parameter θ_1 can be regarded as an average treatment effect.

Secondly, to find an unbiased true causal relationship, the researchers have to specify the functional form of the model correctly. The literature offers two approaches. In the parametric global strategy, the researcher utilizes all available data points to estimate outcomes for the individuals near the cutoff. As the analysis includes all observations, the global strategy might increase the precision of the estimates. Since some observations are far away from the cutoff, it is likely that the underlying relationship is non-linear. Thus, the researcher has to decide whether $f(R_i)$ takes, for instance, a quadratic or cubic form. Because it is often difficult to guess the correct functional form, bias in the estimation might increase with the precision as a trade-off.

Another, and a more relevant method from the point of view of this study, is the local approach. The local approach only uses observations close to the cutoff point, where the effect of the treatment is likely to be linear. In the local approach, the simplest option would be to use a regression equation without the $f(R_i)$ term to obtain the difference in average outcomes between individuals on the left side and right side of the cutoff. Unfortunately, in the cases where the true relationship between the outcome and assignment variables is upward or downward sloping, a simple regression comparing means would offer biased estimates. In order to diminish this bias, it is proposed to estimate a local

linear regression model where $f(R_i)$ takes the form of R_i (Hahn et al., 2001).

Local linear regressions are often estimated using kernel regression, where the kernel function puts more weight on observations in the neighborhood of the cutoff point and zero weight on observations outside the bandwidth of interest. The main challenge of the local approach is to choose the optimal width for the window. With a very narrow bandwidth, the precision of the estimates might be poor, but with a wider bandwidth, the linearity assumption might not hold, which increases the bias. Lee and Lemieux (2010) list methods to find optimal the balance between bias and precision. In addition, they suggest that the researcher should present the result with varying windows to demonstrate the robustness of the findings.

A final general remark on the regression discontinuity design is that the RDD estimates are always local. In RDD, we are interested in individuals who are located close to the threshold, since they are most likely to be, on average, similar. But at the same time, individuals close to the cutoff who are at the center of our analysis may, on average, considerably differ from those who are located further away from the threshold. Thus, it is not clear that the observed effect would be similar for those who are not included in the analysis. In other words, if the key assumption of the setting is met, RDD offers high internal validity, but one should always be cautious with the external validity of the estimates.

5.3 Fuzzy Regression Discontinuity Design and LATE

So far, we have considered the so-called sharp regression discontinuity design model, where the probability of treatment jumps from 0 to 100 percent after the threshold. Often the actual compliance is not perfect and thus we have individuals who are treated and untreated on both sides of the cutoff. For instance in Finland, because the school starting rule is not completely binding, parents may postpone the child's school start or enroll her earlier than the law stipulates. As Lee and Lemieux (2010) explain, if the compliance of the rule is not perfect, the estimates of equation 4 should not be taken as an average treatment effect on treated, because non-compliance dilutes the effect.

To understand, why non-compliance dilutes the effect, it is useful to think my setting through a randomized controlled trial. Let us assume that the SSA rule as if randomly assigns individuals to the treatment and control groups. Individuals born just after New Year are in treatment group, and hence start school at an older age than the control group which includes individuals born just before New Year. Because the groups are otherwise, on average, similar, by comparing the groups, we can study the causal effect of treatment on educational outcomes. Based on the previous studies presented in section 2, it is reasonable to expect that this treatment should have a positive effect. However, let us further assume that some individuals in the treatment group do not take the treatment and, in

addition, some individuals in the control group have accesses to the treatment. Now, if we compare the groups, it is reasonable to expect that the observed difference would be smaller than in the case of perfect compliance. This is caused by two factors. First, some individuals in the treatment group enroll in school earlier, and hence they perform worse compared to if they had started later. Thus, the observed average effect on treatment group is smaller compared to the situation where all individuals would be treated. Second, some individuals in the control group postpone their school start which implies that they probably perform better compared to if they had started at a younger age. Hence, the control group performs, on average, better compared to the situation where no one was treated. In conclusion, compared to sharp setting, the observed effect on individuals in the treatment group is smaller and the observed effect on control group is higher. This creates downward bias. The approach that takes this kind of imperfect compliance into account is called fuzzy regression discontinuity design.

The fuzzy regression discontinuity design (FRD) setting is similar to an instrumental variable (IV) approach. In FRD, we can think of the rule as an instrumental variable: if the individual is born after the cutoff, there is a jump in the treatment probability but not complete separation. Therefore, as Hahn et al. (2001) demonstrate, the treatment effect or IV estimate can be obtained using two-stage least squares (2SLS) or by calculating the Wald estimator. Thus, in the language of instrumental variables, equation 4 is the reduced form equation. The reduced form equation gives us the intention-to-treat (ITT) estimate, but it should not be considered as an unbiased effect of the treatment because of the non-compliance problem mentioned above. Non-compliance can be taken into account by dividing the reduced form estimate by the first stage estimate. The first stage relation can be written as

$$T_i = \alpha + f(R_i) + \theta_2 D_i + v_i, \tag{5}$$

where T_i is as a treatment indicator which denotes if the individual i is treated, $f(R_i)$ controls for the relationship between the treatment and assignment variable R_i , and v_i is the error term. D_i is defined as in equation 3. In the first stage equation, we are interested in parameter θ_2 , which measures the magnitude of the jump in the treatment propensity. In this exactly identified case, where there are as many binary regressors as binary instruments, the FRD estimate or IV-estimate ρ can be calculated using the Wald estimator formula $\rho = \theta_1/\theta_2$. The two-stage least squares regression would give exactly the same number.

The idea of two-stage least squares method is to first use the first stage equation 5 to find the fitted values \widehat{T}_i of treatment indicator T_i . Then the fitted values, which seize the variation in the treatment deriving from the exogenous instrument, are used in the second stage equation

$$Y_i = \alpha + \theta_3 \widehat{T}_i + f(R_i) + u_i, \quad (6)$$

to obtain the 2SLS estimate θ_3 of the school starting age. In this simple case, this is exactly the same as the IV-estimate ρ given by the Wald estimator. Two-stage least squares is usually used instead of the Wald estimator, because the statistical software automatically gives the correct standard errors.

Because the fuzzy regression discontinuity design is comparable to an instrumental variables approach, the treatment effect estimate should be understood through the framework of the local average treatment (LATE) theorem introduced by Imbens and Angrist (1994). In the LATE framework, $\rho = \theta_1/\theta_2$ has a specific causal interpretation if certain assumptions hold. First, the instrument has to be as good as randomly assigned, which means that the treatment must be independent of potential outcomes and potential assignments. Second, an exclusion restriction states that the instrument affects the outcome only through the treatment. Third, the first stage relationship has to be strong enough. Lastly, according to the monotonicity assumption, the instrument has to move everyone in the same direction. In other words, it is possible that people’s behavior is not influenced by the instrument, but if individuals react, they all react in a similar manner. This assumption rules out people who would only accept the treatment if they were not assigned to treatment by instrument and the other way around they would only turn down the treatment if they were assigned to the treatment.

The rationale for these assumptions becomes clearer if we divide the individuals to four subgroups based the way they react to the instrument. As shown in the table 3, we think there are four different subgroups in the LATE framework. The group we are interested in is *compliers*, who are the people who would change their behavior if they were affected by the instrument. This means that the complier i would start school older than the others ($T_i = 1$) if she was born just after New Year ($D_i = 1$). In the case where a complier was born just before New Year ($D_i = 0$) she would start school younger than other children in the class ($T_i = 0$). Alternatively, *never-takers* and *always-takers* are not influenced by the instrument. Always-takers are the group who find a way to be treated in any event. Hence, the always-taker i delays school start to be old in the class ($T_i = 1$) if she was born in December ($D_i = 0$). By contrast, never-takers always dodge the treatment, which implies that a never-taker would start school earlier than the law specifies ($T_i = 0$) if she was born in January ($D_i = 1$). The last group is the *defiers*, who always move opposite to the direction what is instructed. They would start the school earlier if they were born in January and postpone starting school if they were born in December. However, a valid instrumental variable setup should not contain defiers as they violate the monotonicity assumption. (Angrist and Pischke 2009, 158-161.)

Now, if we consider the reduced from equation 4 in terms of the LATE subgroups, it is clear that the

effect I estimate for the individuals who are treated is same as a weighted average treatment effect on compliers and always-takers. Within this LATE framework, the effect I obtain on individuals who were not treated corresponds to a weighted average effect on never-takers and compliers. Because I have assumed that the instrument is independent of the potential outcomes, equation 4 has a causal interpretation, but as mentioned earlier, it is diluted since there exist never-takers and always-takers. However, the independence assumption also states that the instrument is independent of potential treatment assignment. Thus, the first stage relationship measures how many individuals react in the treatment group when the instrument is as if randomly turned on. Therefore, by using the first stage, I can extract the effect of treatment on compliers from the intention-to-treat estimate. In addition, because of the exclusion restriction assumption, that the instrument affects outcomes only through the treatment, the LATE estimate captures the causal effect of treatment on compliers. (Angrist and Pischke 2009, 158-161.)

The setting of this study should satisfy the LATE assumptions. I demonstrate in section 6 that an individual's birthday is as good as randomly assigned in the vicinity of the cutoff. Furthermore, I show that my first stage relationship is strong enough in section 7. The exclusion restriction and monotonicity assumptions are always more cumbersome in the sense that there do not exist explicit ways to test them. Some researchers have expressed concerns about the monotonicity assumption when school starting rule is used as an instrument. For instance, Barua and Lang (2016) use U.S. data from 1950s and show that children born in the first quarter of the year started school on average at an older age than the children born during the last quarter. However, among the individuals born in the beginning of the year, there were more children who started school at an especially young age and fewer children who started the school at an especially old age compared to the individuals born in the end of the year. I cannot rule out this kind of behavior in Finland, but as Landersø et al. (2016b) note in the case of Denmark, it would be very inconsistent for the parents to prefer the child to be one of the oldest in the peer group if she was born in December and choose the opposite if the child was born in January.

With regard to the exclusion restriction, it is possible that parents are aware of the effect of relative school starting age, and hence give more support for children born at the end of year. This would bias my results downwards. On the other hand, many sports use the same thresholds as the school system to divide children to age group teams. Thus, relatively older children may benefit from the maturity and seem more talented than the relatively younger team members. If this increases child's confidence, it may spillover on the educational outcomes. It is also possible that if relatively younger children are more likely to drop out from sports, they may put excess effort on studying. This again may decrease gap in educational outcomes. It is beyond the scope of this article to examine these issues, but the

Table 3: Compliance Type

	Born before New Year		Born after New Year	
	$D_i = 0$		$D_i = 1$	
Compliers	Non-treated	$T_i = 0$	Treated	$T_i = 1$
Always-takers	Treated	$T_i = 1$	Treated	$T_i = 1$
Never-takers	Non-treated	$T_i = 0$	Non-treated	$T_i = 0$
Defiers	Treated	$T_i = 1$	Non-treated	$T_i = 0$

reader should keep these in her mind.

Lastly, one should note that the LATE estimate is a heterogeneous estimate. In this case, the heterogeneity comes from the fact that the estimate measures the effect on a particular group, namely on compliers. Thus, the IV estimate is the best guess for what would have happened to the compliers, who were born in December had they been born in January. However, the estimate $\rho = \theta_1/\theta_2$ is likely to be non-informative on the effect on always-takers and never-takers. Always-takers and never-takers may on average substantially differ from compliers and hence researcher should be careful, when generalizing the findings from the LATE analysis.

To summarize, since the school starting rule is not completely binding, I also use the fuzzy regression discontinuity design to estimate the causal effect of school starting age. The fuzzy regression discontinuity design setting is similar to the instrumental variable approach, and hence the FRD estimates should be interpreted through LATE framework. In appropriate LATE analysis, there are three subgroups. Always-takers and never-takers are people whose decisions on school starting age are not influenced by the school starting rule. Thus, regardless of their birthday always-takers are always among oldest in the class whereas never-takers are always on average younger than other children. In the center of LATE analysis is third group called compliers who change their behavior according to the school starting rule. Hence, compliers born just after New Year start school are on average one year older than compliers born just before New Year. The FRD or IV estimates from the fuzzy regression discontinuity design measure the effect on compliers and are most likely to be larger than ITT estimates.

5.4 Estimation

I now turn to present in greater detail how I apply the regression discontinuity design in this study. As stated earlier, the school starting age rules generate a jump in school starting ages in Finland. I utilize a regression discontinuity design to study if the jump in school starting age affects educational outcomes.

To estimate the discontinuous jumps in school starting age on educational outcomes, I use a weighted local linear regression and observations in the window of ± 30 days around the threshold. The observations are weighted using a triangular kernel function, which takes the form

$$K_h(D_i, R_i) = \max\left(0, 1 - \left|\frac{R_i}{h}\right|\right), \quad (7)$$

where R_i is the running variable and h is the bandwidth. The triangular kernel function puts more weight on the observations that are close the cutoff and zero weight on the observations that are outside the selected window h . I use the triangular kernel function because there is evidence that it is the least biased (Fan and Gijebels, 1992). In any case, the functional form of the kernel function should not have a big effect on the estimates. More crucial is to select the bandwidth h , which balances between bias and precision. My choice of bandwidth is guided by the approach of Calonico et al. (2014).¹³ In addition, I demonstrate later in section 7.3, that the estimates are relatively insensitive to the different bandwidths.

The reduced form equation is

$$Y_i = \alpha_1 + \beta_1 R_i + \theta_1 D_i + \delta_1 D_i R_i + \lambda_1 X_i + e_{i1}, \quad (8)$$

where Y_i is the educational outcome, α_1 are the year fixed effects, R_i represents i 's redefined birthday, X_i contains background characteristic and e_{i1} is error term. The birthday variable R_i has been redefined for each individual i in the way that it measures distance from New Year.¹⁴ For the individuals who are born between July and December R_i takes negative values and consequently for the individuals who are born between January and June, R_i takes non-negative values. D_i is an indicator variable defined as

$$D = \begin{cases} D_i = 1 & \text{if } R_i \geq 0 \\ D_i = 0 & \text{if } R_i < 0 \end{cases}. \quad (9)$$

Hence, D_i points out whether the person is born between January - June or July - December.

¹³The optimal bandwidths suggested by the method of Calonico et al. (2014) are available in appendix in table A3.

¹⁴Hence, starting from the first of January, R_i takes values 0,1,2,3,4.. until the end of June. Correspondingly, from the 31st of December towards the beginning of July, the values are -1,-2-3...

Although I do not know the exact school starting year, I approximate the school starting age to estimate the first stage relationship. I define the binary variable OLD_i to be one if the student is born between January and June and has graduated in the theoretical time or later and zero otherwise. The variable is also one if the individual is born between July and December and has graduated from comprehensive school one year later than she was supposed to. The first stage regression equation is

$$OLD_i = \alpha_2 + \beta_2 R_i + \theta_2 D_i + \delta_2 D_i R_i + \lambda_2 X_i + e_{i2}, \quad (10)$$

where e_{i2} denotes the error term for individual i and R_i , D_i and X_i are defined in the similar way as in the reduced form equation. The first stage equation measures the effect of the instrumental variable on the treatment indicator and is sometimes used to test the relevance assumption of the instrumental variable.¹⁵

I choose equations 8 and 10 to have the functional forms presented above, since it has been shown that the formulations should diminish boundary problems related to more simple models (Fan and Gijbels, 1992). These boundary problems arise from the situations briefly described in section 5.2, where the relationship between the assignment variable and the outcome slopes upwards or downwards and a simple average comparison would offer a biased estimate. In addition, the model of equations 8 and 10 get lower Akaike information criteria values than more simple or complex models.

As in the previous literature, the identification strategy here relies on the assumption that birthdays are as good as randomly assigned in the vicinity of the threshold. Hence, individuals just under the cutoff provide good counterfactual to those just above the cutoff. The main interest of the analysis is on the parameter θ_1 , which measures the jump in the educational outcomes at the cutoff. Since the law regarding the school starting age is not completely binding, there are some individuals who have started the school earlier or later. Therefore, θ_1 is the intention-to-treat estimate, which tells us something about the causal effect. However, it cannot be taken at face value for the causal effect of the school starting age, because there are some never-takers and always-takers who dilute the effect. I estimate the effect on compliers using the two-stage least squares model, but the Wald estimator $\hat{\rho} = \hat{\theta}_1 / \hat{\theta}_2$ would give the same estimate in this simple case. As I have explained earlier, $\hat{\rho}$ measures the effect of treatment on compliers, who are the individuals who react as the instrument is as if randomly turned on.

¹⁵The relevance assumption states that the instruments must have a significant effect on the endogenous explanatory variables. Violation of the assumption may lead to the bias, especially in the case where analysis incorporates multiple instruments for one endogenous instrument. As shown in the result section 7, this is not a concern in my case.

6 Validity of the Setting

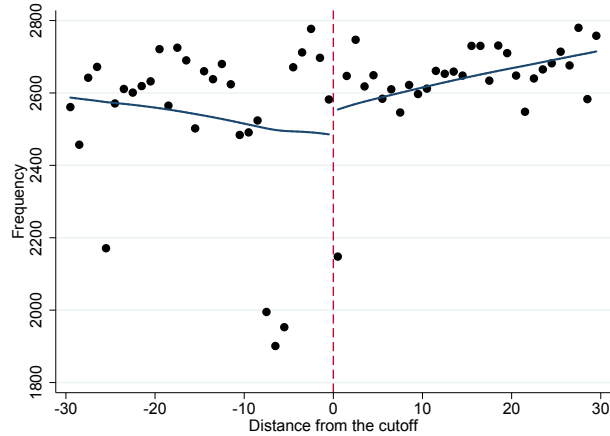
Before moving to the main results, I examine whether the key assumption of the regression discontinuity design holds. In principle, parents have incentives to manipulate their children’s birthdays. The literature shows that relatively older children achieve better results in exams, which could encourage parents to delay birth to give an edge in comprehensive school. By contrast, if a child enters school earlier, parents may save in childcare fees. But strategic behavior has its costs as well, because shifting the timing of birth may increase the risks to the mother and child. Hence, since parents may enroll a child in school at an older or younger age than the rule specifies, and childcare is reasonably affordable, it is difficult to see why parents would take the additional risk by delaying or advancing the birth. In any case, there is some evidence which suggests that parents may behave strategically. For instance, Shigeoka (2015) studies the manipulation of births in Japan, where school starting rules are followed especially strictly. Shigeoka finds that there is a significant discontinuous jump in the number of births just after the cutoff, which is a clear sign of manipulation. Thus, the possibility of manipulation should be taken seriously.

It is basically impossible to detect whether certain individuals manipulate the assignment variable. Fortunately, the assumption has several implications, which can be studied to assess the validity of the setup. Firstly, if the individuals do not manipulate the assignment variable, the density of the assignment variable should be smooth in the vicinity of the cutoff point. Secondly, if the assumption holds, individuals who are born near the cutoff should have equal chances to end up just above or below the cutoff. Consequently, background characteristics should evolve smoothly around the cutoff.

I study the density of the assignment variable using a test similar to McCrary’s (2008). The basic idea is that if people manipulate the assignment variable, we should observe heaps of observations just after or before the cutoff. I construct the test in the following way: first I split the running variable into one-day bins and calculate the number of observations in each bin. Then I use the number of observation in each bin as a outcome variable in a local linear regression, which takes form of equation 8. In this case, the parameter θ_1 should capture a discontinuous jump or drop in the number of observations, which in turn would be a sign of manipulation. McCrary provides computer code to perform the test. I use this as well, but the results should be treated with the concern, because the program is designed to be used with a continuous assignment variable.

Figure 1 shows the relationship between the density of observations and the assignment variable. The distribution of birthdays seems to evolve well, although there are notable drops in the number of observations around Christmas and on New Year’s Day. The evidence from the local linear regression supports the graphical inspection. I present the results from more formal analysis in table 4. According

Figure 1: Density of the Running variable



Notes: Each dot corresponds to the average number of births in a bin of one day. The lines on top of the dots are estimated using the equation 8.

to the test, there is a small jump in the number of observations, but the result is not statistically significant. Thus, it seems that the density test does not provide evidence to reject the no-manipulation hypothesis. In addition, I perform McCrary’s test using his original code. The result is shown in column 2 of table 4. This test indicates that there exists a small and barely significant discontinuous jump. However, the results should be treated critically, because the assignment variable is discrete instead of continuous.

Another implication of the no-manipulation assumption is that the predetermined background covariates should behave smoothly around the threshold (Lee and Lemieux, 2010). The idea is that if certain kinds of individuals are manipulating the assignment variable, we should detect a discontinuous jump in certain background covariates. Unfortunately, as I stated earlier, the background variables are observed when the individuals are 14, 15 and 16 years old. Therefore there is a possibility that their school starting age could affect their parents’ relationship through their school success. In any case, the behavior of the background covariates can be studied using equation 8, where the background covariate is treated as an outcome variable and the parameter θ_1 reveals if there is evidence regarding discontinuity.

Figure 2 shows plots of the relationships between the background covariates and the running variable. All the graphs are relatively noisy, but there is evidence that mother’s earnings and parent’s housing arrangements might behave discontinuously at the threshold. Table 5 shows the estimation results.

Table 4: Density Tests

	(1)	(2)
Discontinuity	62.40 (147.8)	0.022 (0.011)
Constant	2502 (117.8)	

Notes: The estimates in the first column have been obtained using local linear regression presented in equation 8. The dummy variable measures whether there is a discontinuity in the number of observation at New Year. The estimate in the second column has been obtained by computer code of McCrary. In McCrary's test, the estimate is obtained using a local linear regression similar to equation 8, where densities are transformed on logarithmic scale.

Table 5: Background Covariates

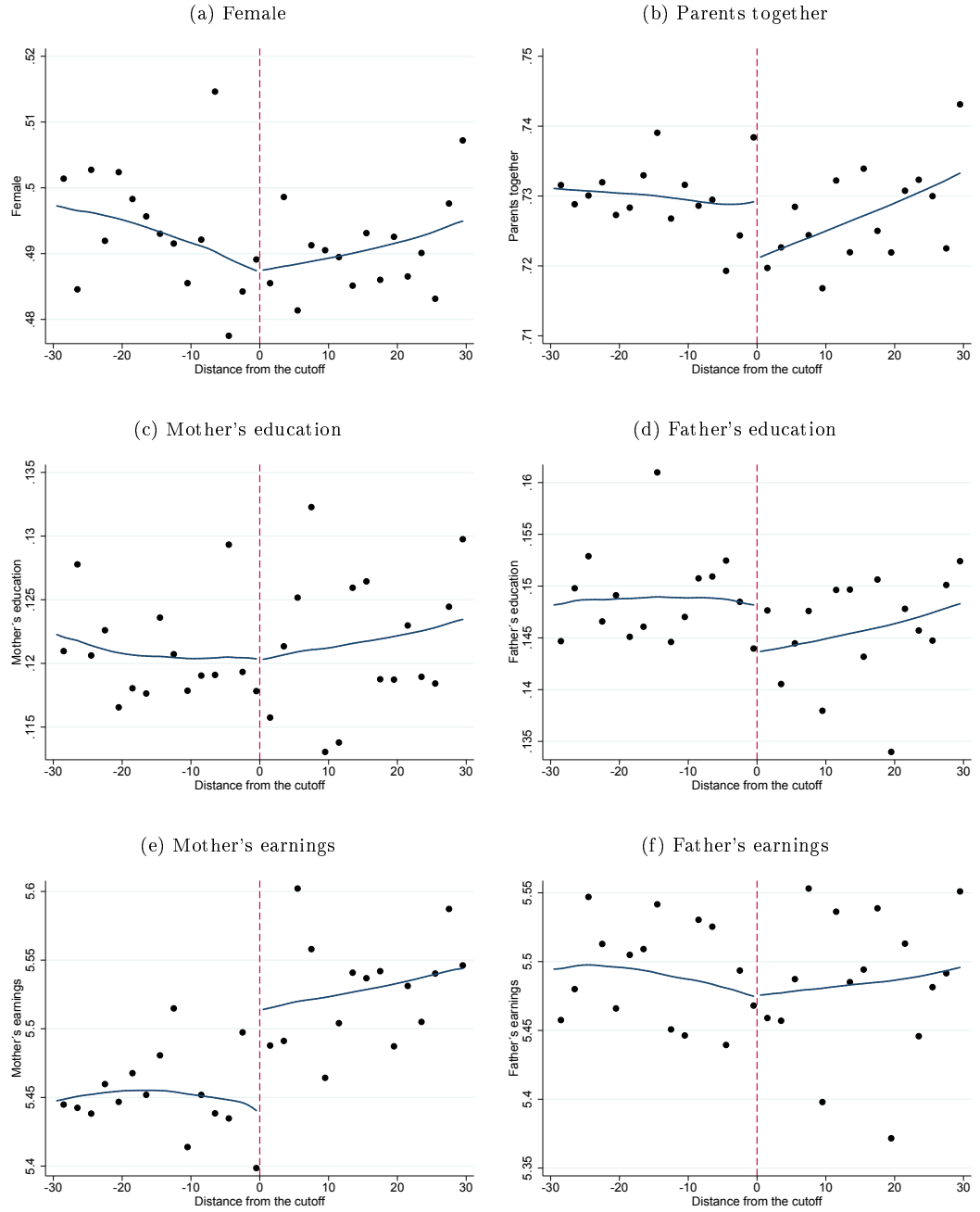
	Education		Earnings		Female (5)	Parents together (6)
	Mother (1)	Father (2)	Mother (3)	Father (4)		
Discontinuity	0.0003 (0.0036)	-0.0037 (0.0040)	0.0783 (0.0323)	0.0063 (0.0321)	0.0021 (0.0322)	-0.0073 (0.0050)
Constant	0.120 (0.003)	0.148 (0.003)	5.439 (0.023)	5.497 (0.023)	0.478 (0.004)	0.729 (0.004)
Observations	155,359	155,359	155,359	155,359	155,359	155,359

Notes: The estimates are obtained using the local linear regression presented in equation 8 and a window of 30 days around the cutoff. Education is measured by a dummy variable, which takes value one if a parent has a tertiary degree. Robust standard errors are presented in parentheses.

The coefficients measuring parents' education, housing arrangements and father's earnings are very small and statistically indistinguishable from zero. However, the analysis indicates there is a positive and statistically significant jump in mother's earnings. The estimate of the coefficient is 0.078, which means that the average position in the income distribution jumps 0.078 decile groups at the cutoff. Overall, the covariate tests do not raise notable concerns. The difference in mother's earnings is statistically significant, but probably not economically.

In conclusion, I do not find evidence of a notable jump in the density of the assignment variable. The relatively small jump in mother's earnings should not pose a threat to the validity of the setup. Altogether, these results suggest that the no-manipulation assumption holds in this setting. Thus, I can assume that people who are born in the proximity of the cutoff are locally randomized into the treatment and control groups. In the next section, I exploit this variation and study the effects of school starting age using the models presented in section 5.

Figure 2: Background Covariates



Notes: Each dot corresponds to the average number of births in a bin of one day. The lines on top of the dots are estimated using the 8.

7 Results

I start my results section by showing graphical evidence regarding the effects of school starting age on educational outcomes. Then I present the actual estimation results obtained by using the regression discontinuity design approach presented above. Finally, I move to inspect the robustness of the results and study if there is heterogeneity in the results.

7.1 Graphical Evidence

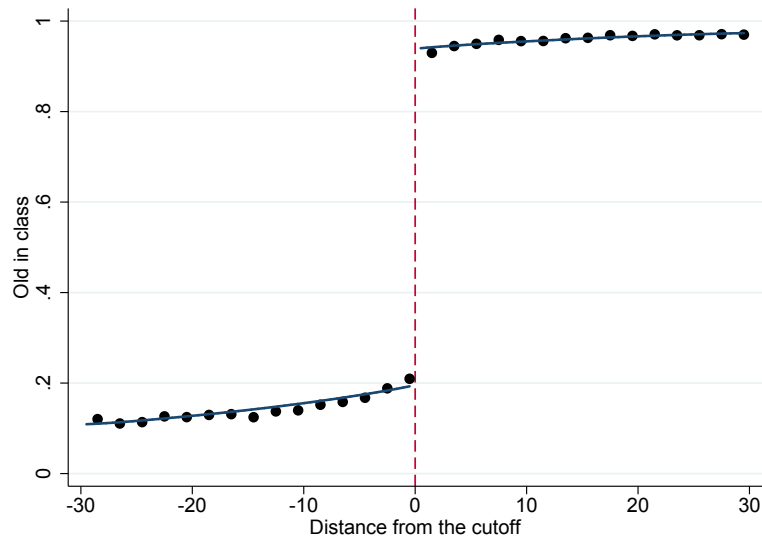
A strength of the regression discontinuity design is the possibility to illustrate the results graphically. Lee and Lemieux (2010) list several important reasons to graph the results. First, graphical inspection is an easy way to check whether the rule creates discontinuity at certain threshold. If the jump or drop is not distinguishable in a graph, standard estimation procedure is unlikely to detect it. In addition, a graph may expose whether there are discontinuity points in other parts of the distribution. Second, as we do not know the actual functional form, graphing the relationship between the assignment variable and the outcome variable may offer information about the true model. Last, as I just demonstrated, it is a convenient way to examine whether background characteristics behave smoothly around the cutoff.

In each graph in this section, a dot represents the average of the outcome variable in a 2-day bin. In addition, I fit local linear regression lines on top of the dots. The lines are estimated using equation 8 separately for the observations on both sides of the cutoff. The window around the cutoff is the same as in the main analysis and is selected using method of Calonico et al. (2014) as a guidance.

Figure 3 presents the relationship between birthday and school starting age around New Year. There is a notable jump in the share of individuals who start school older when we move from December to January. I just demonstrated that the no-manipulation assumption should be met in this setting and hence this discontinuous jump can be utilized to study effects of SSA on educational outcomes. Clearly, the jump is not 100 percent which suggests that the sharp regression discontinuity design estimates are diluted. Thus, I apply the fuzzy regression discontinuity design also.

Figure 4 illustrates how the outcome variables of my analysis evolve around the threshold. Each graph indicates that there exists a similar jump in educational outcomes as has been documented in other countries. Panel 4 a shows the relationship between birthdays and GPA. I find that the GPA seems to jump substantially at the cutoff. Furthermore, apart from the cutoff, GPA evolves nicely around the estimated local liner regression lines without notable evidence on non-linearities or discontinuous jumps in other parts of the distribution.

Figure 3: First stage



Notes: Each dot corresponds to the average number of births in a bin of one day. The lines on top of the dots are estimated using equation 8.

Panel b and c of figure 4 present similar evidence on the probability of admission and graduation from general upper secondary school. In both graphs, I detect notable jumps in the share of individuals at the cutoff, which indicates that children born after the cutoff are more likely to be admitted and graduate from general upper secondary school. Compared to panel 4 a, both relationships seem a bit noisier, but again there is little evidence of non-linearities.

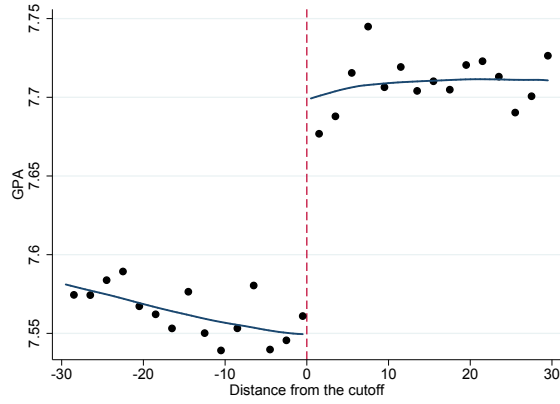
7.2 Estimation Results

In this subsection, I report my main estimation results. Graphical inspection clearly reveals that school starting age rules do have an effect on educational outcomes. Next, I provide the OLS estimates from the simple regression analysis to motivate the use of the regression discontinuity design. I then apply the regression discontinuity design introduced in section 5. Intention-to-treat estimates are obtained using local linear regression, which takes the form of equation 8. I calculate IV estimates using two-stage least squares, but in this simple case, the IV estimates can also be found by calculating the Wald estimates.

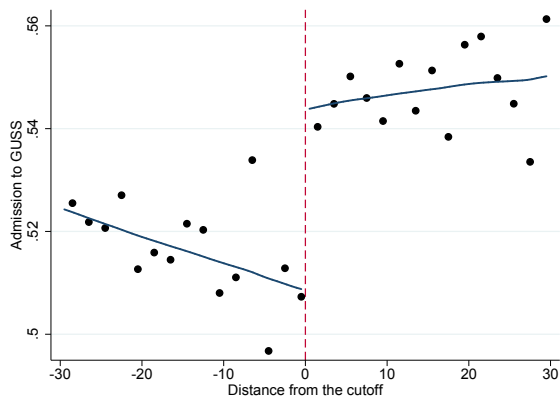
To understand why I exploit the regression discontinuity design to overcome the selection issues, I start

Figure 4: The Effect of School Starting Age on Educational Outcomes

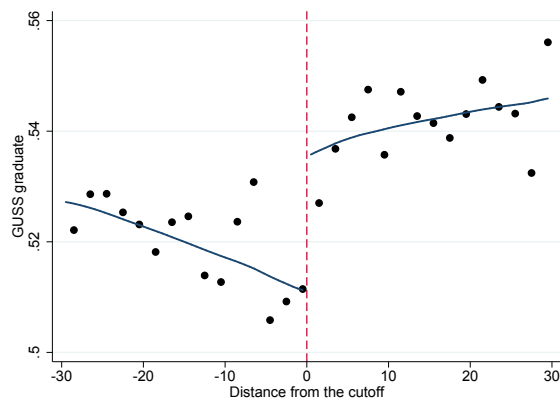
(a) School starting age and GPA



(b) School starting age and admission to GUSS



(c) School starting age and graduation from GUSS



Notes: Each dot corresponds to the average value of the outcome variable in 2-day bin. The lines on top of the dots are estimated using 8.

Table 6: Simple OLS Regression Estimates

	GPA	Admission to GUSS	GUSS graduate
SSA_i	-0.164 (0.003)	-0.098 (0.001)	-0.108 (0.001)
OLD_i	0.010 (0.002)	-0.015 (0.001)	-0.021 (0.001)
Observations	1,008,297	1,008,297	1,008,297

Notes: All estimates have been obtained using equation 11 with cohort fixed effects and control variables. Variable SSA_i captures the effect of approximated school starting age on educational outcomes. The control variables include gender, parents' educational attainment, parents' earnings and parents' housing statuses. Robust standard errors are shown in the parentheses.

this section by estimating two simple ordinary least squares (OLS) regressions using the total sample. The first OLS regression takes following form

$$Y_i = \alpha + \theta_4 SSA_i + \lambda_4 X_i + e_i, \quad (11)$$

where Y_i denotes the educational outcome, SSA_i is the approximated school starting age,¹⁶ α are the year fixed effects, X_i contains control variables and e_i is error term. The second regression is otherwise similar, but I use indicator OLD_i as an explanatory variable instead of SSA_i . I present the results from the simple OLS regressions in table 6. The OLS regressions suggests that there is a negligible or a negative correlation between the school starting age and educational outcomes. But as I stated in the previous section, the estimates are probably biased since school starting age is correlated with learning disabilities, and hence the estimates cannot be taken as a causal effect. In addition, notice that there is a measurement error in my school starting age approximation. Thus, one should not put too much emphasis on the magnitude of the estimates.

I now turn to show the results obtained using the regression discontinuity design. I present the estimated effect of school starting age on GPA in table 7. Column 1 shows the first stage, the intention-to-treat, and the instrumental variable estimates. I find that the individuals who are born just after the cutoff have on average a 0.15 grade points higher GPA than individuals born just before the cutoff. Since some individuals start school later or earlier, this intention-to-treat estimate is probably biased downwards. The non-compliance is taken into account in the IV-estimate, which measures the effect

¹⁶Notice, SSA_i is not binary variable. Thus, it measures school starting age in the way that, if you are born in January and start school at time the variables is 7.7 While for individual who enrolls in correct time and is born in December, the variable takes value 6.7

on compliers. I find that a higher school starting age increases the GPA by 0.2 grade points among compliers.

Notice that the IV estimates may be biased upwards, because I approximate the school starting age. More precisely, IV estimates are biased as there are two groups whose treatment status I cannot define precisely. The first group consists of individuals who are born in December and start school as instructed but are held back during comprehensive school. Thus, they should be defined as compliers, but I recognize them as always-takers. The second group group includes children who are born in January and start school earlier, but are also held back. Alternatively, the second group should be considered as never-takers, but in my approximation they belong to compliers. The size and direction of the bias depends on the relative sizes of these two groups. It is possible that the two groups cancel out each other, which implies that my first stage estimate matches the true effect. However, I suspect there are more representatives of the first group than the second, which implies that first stage estimate is biased downwards. This again means that my IV estimates over-estimate the true effect.

Columns 2, 3 and 4 in table 7 report the estimates from models that include year fixed effects and background covariates. Because the treatment is independent of background characteristics in the regression discontinuity design, the inclusion of control variables should not affect the coefficient estimate. Thus, the addition of control variables could be considered as a robustness check. I observe that the intention-to-treat and IV estimates are insensitive to the inclusion of fixed effects and background covariates. An especially salient result is shown in Column 4 where I add mother's earnings separately to the group of background variables. Mother's earnings are added separately, because I did observe a small but significant jump in mother's earnings around the cutoff. This could threaten the validity of the setting. But as column 4 shows, when mother's earnings are added, the coefficient hardly moves, which adds further credibility to my setup.

Since there is a discontinuous jump in the GPA of final school report, a reasonable guess would be that the jump affects admission to and graduating rates from general upper secondary school. I report the effect of school starting age on the probability of admission to general upper secondary school in table 8. As column 1 shows, this probability jumps by 3.5 percentage points at the cutoff. Again, when non-compliance is taken into account, the effect is larger. I find that the IV estimate is 4.8. This means that children who follow the school starting rules and are born just after New Year are 4.8 percentage points more likely to receive admission to general upper secondary school than children who follow the school starting rules and are born just before New Year. In addition, as shown in columns 2, 3, and 4 of table 8, the results again move very little when year fixed effects and background covariates are added.

Finally, I present the relationship between school starting age and general upper school graduation in

Table 7: The Effect of School Starting Age on GPA

	(1)	(2)	(3)	(4)
First stage	0.743 (0.004)	0.743 (0.004)	0.742 (0.004)	0.743 (0.004)
Reduced form	0.150 (0.013)	0.151 (0.013)	0.154 (0.012)	0.151 (0.012)
IV	0.202 (0.018)	0.204 (0.018)	0.208 (0.016)	0.204 (0.016)
Year fixed effects		yes	yes	yes
Background covariates			yes	yes
Mother's earnings				yes
Observations	155,359	155,359	155,359	155,359

Notes: The estimates are obtained using a window of ± 30 days around the cutoff. The first stage results are estimated using equation 10, reduced form estimates are from equation 8 and IV estimates are calculated using two-stage least squares method. Robust standard errors are presented in parentheses.

Table 8: The Effect of School Starting Age on Admission to General Upper Secondary School

	(1)	(2)	(3)	(4)
First stage	0.743 (0.004)	0.743 (0.004)	0.742 (0.004)	0.743 (0.004)
Reduced form	0.035 (0.006)	0.036 (0.006)	0.037 (0.005)	0.036 (0.005)
IV	0.048 (0.008)	0.048 (0.008)	0.050 (0.007)	0.048 (0.007)
Year fixed effects		yes	yes	yes
Background covariates			yes	yes
Mother's earnings				yes
Observations	155,359	155,359	155,359	155,359

Notes: The estimates are obtained using a window of ± 30 days around the cutoff. The first stage results are estimated using equation 10, reduced form estimates are from equation 8 and IV estimates are calculated using two-stage least squares method. Robust standard errors are presented in parentheses.

Table 9: The Effect of SSA on Probability to Graduate from General Upper Secondary School

	(1)	(2)	(3)	(4)
First stage	0.743 (0.004)	0.743 (0.004)	0.742 (0.004)	0.743 (0.004)
Reduced form	0.025 (0.006)	0.025 (0.006)	0.026 (0.005)	0.024 (0.005)
IV	0.033 (0.008)	0.033 (0.008)	0.035 (0.007)	0.033 (0.007)
Year fixed effects		yes	yes	yes
Background covariates			yes	yes
Mother's earnings				yes
Observations	155,359	155,359	155,359	155,359

Notes: The estimates are obtained using a window of ± 30 days around the cutoff. The First stage results are estimated using equation 10, reduced form estimates are from equation 8 and IV estimates are calculated using two-stage least squares method. Robust standard errors are presented in parentheses.

table 9. The magnitude of the effect on the probability to graduate from general upper secondary school is smaller but still significant. According to the results in column 1, the group who are born just after the cutoff have a 2.5 percentage points higher probability to graduate from general upper secondary school. On compliers, I observe that the school starting age increases the probability to graduate from general upper secondary school by 3.3 percentage points. Again, the results are similar with or without the year fixed effects and background covariates as the columns 2-4 in table 9 demonstrate.

How do these results compare to earlier studies? The magnitude of the effect is comparable to what has been found in other countries, but in the light of previous studies regarding Finland, the effect is surprisingly large. I estimate that the GPA jumps 0.13-0.18 standard deviations whereas Bedard and Dhuey (2006) find that that one year difference in the school starting age increases test scores by 0.06-0.13 standard deviations in Finland. One factor that could explain why my estimates differ from those obtained by Bedard and Dhuey (2006) is that they use data from an international standardized test. By contrast, as explained in section 3, the Finnish comprehensive school does not have standardized tests and it is not clear how teachers assign grades. This may imply that some non-cognitive skills, which are related to age differences, could explain the deviations in results. For instance, relatively older students may be more mature which would contribute positively to grades through better behavior.

In conclusion, the evidence above indicates that the school starting age has a causal effect on educational outcomes. Children who are born at the beginning of the year are more likely to be older in

the class and hence have higher GPA at the end of comprehensive school than those born at the end of the year. In addition, because selection into general upper secondary school is mainly based on GPA, it is unsurprising that this age difference affects the probability of admission to and graduation from general upper secondary school. I cannot distinguish between different mechanisms and thus my findings should be considered as the total effect of all channels. However, regardless of the actual mechanisms, my results on general upper secondary school suggest that the relative school starting age may have persistent effects in Finland.

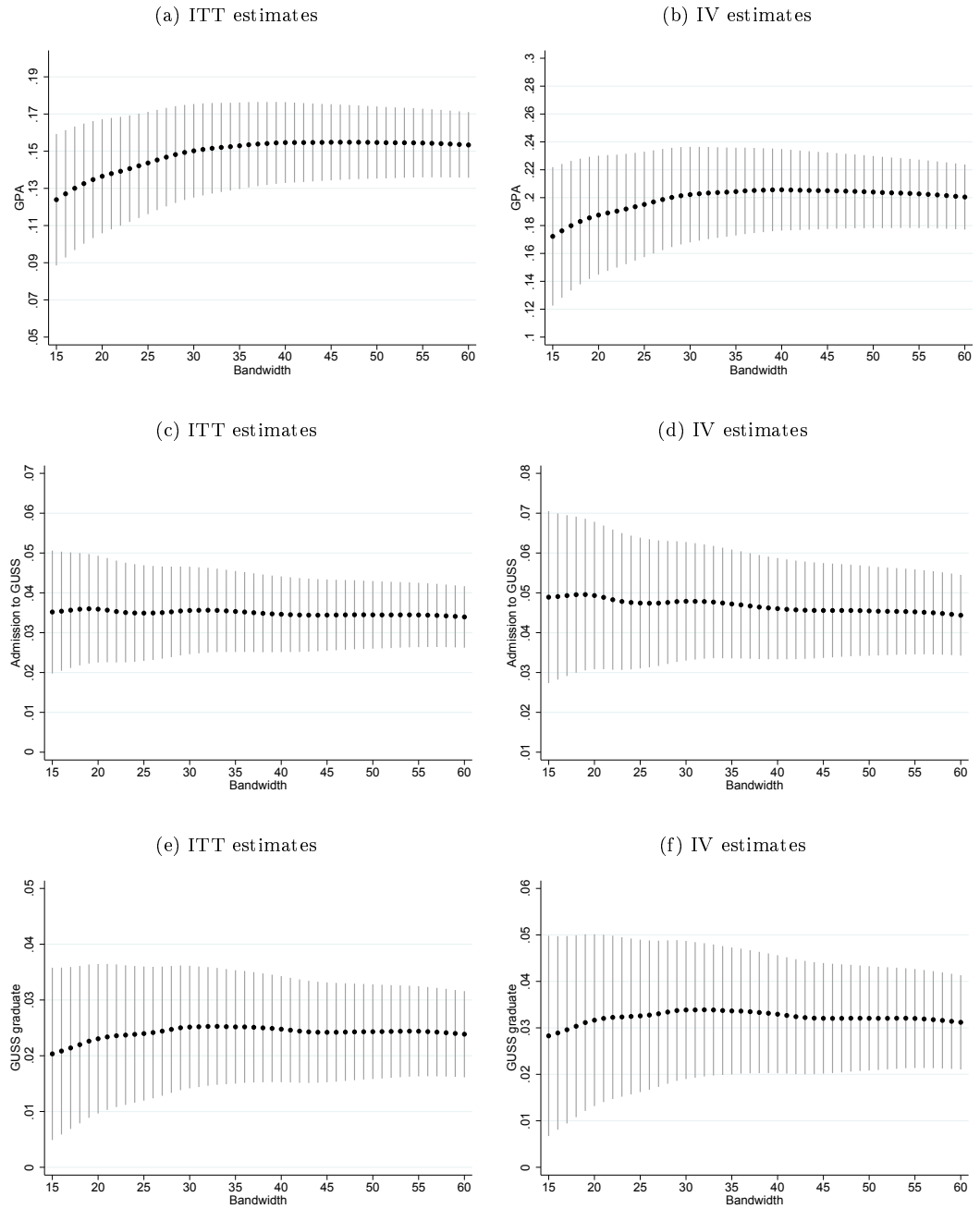
7.3 Sensitivity and Robustness Inspections

The results presented above should capture the causal effect of school starting age on educational outcomes if parents are not manipulating their children's birthdays, I have used the correct functional form, and selected an appropriate window around the cutoff. I have already provided evidence pointing that the key assumption regarding the absence of manipulation is not violated, and I shall next show that the other conditions hold as well.

I start the robustness inspections by showing that the results are relatively insensitive to the bandwidth choice. In a local approach, it is necessary to pick a bandwidth, which offers suitable balance between bias and precision. Although my choice of window was guided by a formal process, it is advisable to demonstrate that the point estimates do not move too much as the window shrinks or increases. I show how RDD estimates change with window width in figure 5. In each graph, a dot represents an ITT or IV estimate with a different bandwidth. Vertical lines show the 95% confidence intervals around point estimates. Both the IV and ITT estimates are relatively insensitive to the bandwidth choice. The magnitude of the effect gets smaller as the bandwidth shrinks but so does the precision of the estimates. A possible explanation for why we observe smaller estimates with tighter windows could be that the parents are aware of the relative age effect and hence are more likely to postpone the school start closer the child's birthday is to New Year. Thus, the compliers born just before New Year may be a very selective group and hence the effect is small.

Another standard robustness check is to study whether there are notable discontinuous jumps at placebo cutoffs. Motivated by the instructions of Imbens and Lemieux (2008), I conduct a placebo analysis using fake cutoff points. First, I divide the sample into two parts in such that the first part includes observations below the initial threshold and the second individuals above the initial threshold. Then I carry out placebo analyses separately in these two sub-samples using alternative cutoffs which are located 15 and 30 days above and below the initial threshold. The results are obtained using equation 8. With the placebo thresholds that are closer to the initial cutoff, I use a window of 15 days

Figure 5: Sensitivity to Bandwidth Choice



Notes: Each dot corresponds to an intention-to-treat estimate or IV estimate obtained using local linear regression with varying bandwidth. Vertical lines are 95 % confidence intervals calculated using robust standard errors.

Table 10: The ITT Estimates with Different Cutoffs

	(1)	(2)	(3)	(4)
GPA	0.014 (0.013)	0.011 (0.018)	-0.016 (0.018)	-0.007 (0.013)
Admission to GUSS	0.014 (0.006)	0.009 (0.008)	-0.002 (0.008)	-0.006 (0.005)
GUSS graduate	0.010 (0.007)	0.002 (0.008)	-0.009 (0.008)	-0.007 (0.004)
Cutoff	c - 30	c - 15	c + 15	c + 30
Observations	152,400	76,065	79,230	162,410

Notes: All the estimates have been obtained using local linear regression model shown in equation 8. The character c stands for the initial cutoff. The cutoff row indicates where the placebo cutoff is located. Robust standard errors are shown in parentheses.

and with the placebo thresholds that are located further away from the New Year, I use the initial bandwidth of 30 days. The results are presented in table 10. All the estimates are substantially smaller than the ones I find with the main setup. Few of them are statistically significant, but the findings do not seem to raise any doubts regarding my setup.

It is also possible that the results are sensitive to functional form. In a similar way as with the bandwidth, I picked the local linear regression model using a formal procedure. However, to address the possible bias stemming from an incorrect functional form, I estimate the effect with several different functional forms. As I show in table 11, most of the different functional forms give similar estimates. Only the estimates obtained with the quadratic linear interaction models deviate a bit but are still very close to the main estimates.

Finally, there is a concern that the observations I was forced to exclude from the analysis are systematically correlated with the educational outcomes. Thus, as a last sensitivity analysis I examine how the results change if the excluded observations are included in the analysis. The results are shown in appendix in table A1. I find that the estimates are very similar with a sample which includes the individuals with missing information on their parents. Hence, the exclusion of the observations should not cause a threat for the setup.

Table 11: The Estimates with Different Functional Forms

a) GPA	(1)	(2)	(3)	(4)
First stage	0.798 (0.002)	0.744 (0.004)	0.742 (0.006)	0.715 (0.004)
Reduced form	0.150 (0.007)	0.149 (0.013)	0.151 (0.013)	0.124 (0.019)
IV	0.188 (0.009)	0.201 (0.017)	0.204 (0.018)	0.173 (0.027)
b) Admission to GUSS				
Reduced form	0.033 (0.003)	0.035 (0.006)	0.036 (0.006)	0.035 (0.008)
IV	0.041 (0.004)	0.047 (0.008)	0.048 (0.008)	0.048 (0.012)
c) GUSS graduate				
Reduced form	0.023 (0.003)	0.025 (0.006)	0.026 (0.006)	0.021 (0.008)
IV	0.029 (0.004)	0.033 (0.008)	0.035 (0.008)	0.029 (0.012)
Simple averages	yes			
Linear		yes		
Quadratic			yes	
Quadratic interaction				yes
Observations	155,359	155,359	155,359	155,359

Notes: Each column in the table corresponds to a different alternative functional form. The alternative functional forms are presented in appendix in table A2. All estimates are estimated using a window of 30 days around the cutoff. Robust standard errors are shown in parentheses.

7.4 Heterogeneous Effects

I shall end the result section by studying to what extent results differ across groups, i.e., the heterogeneity in the results. The school starting age clearly has an effect on educational outcomes in the total sample, but it is possible that for particular groups, the effect is smaller or larger. Thus, let us turn to investigate the heterogeneity of the effect of school starting age. More precisely, in this subsection, I examine the effect of school starting age conditional on gender, parental education and region.

Columns 1 and 2 of table 12 present the effect of school starting age on educational outcomes separately for women and men. A bit unexpectedly, the effect is significantly larger for women than for men.¹⁷ According to my findings shown in column 1 of table 12, females who are born just after the threshold have 0.2 grade points higher GPA, on average, than females who are born just before. For males, the equivalent intention-to-treat estimate presented in column 2 is 0.098. Similar differences are present in the IV estimates shown in columns 1 and 2. I find that the female compliers who are born at the beginning of the year have on average 0.253 grade points higher higher GPA than the female compliers born at the end of the year. Corresponding estimate for males shown in column 2 is 0.142.

Similar differences are visible in the outcomes related to general upper secondary school shown in panels A and B of table 12. I find that women who are born after New Year are 4.7 percentage points more likely to be admitted to general upper secondary school during the year they graduate from comprehensive school than women who are born just before New Year. Over time, the effect decreases a bit but is still highly significant. On average, females in the treatment group are 3.6 percentage points more likely to graduate from upper secondary school than females in the control group. Again, the discontinuous jumps are substantially smaller for men. I estimate that men born after the cutoff are 2.4 percentage points more likely to be admitted to and 1.4 percentage points more likely to eventually graduate from general upper secondary school than men just before the cutoff. The point estimates for men are not only smaller but also more imprecise. For instance, according to column 3 of panel C in table 12, I cannot argue that the affect on graduation probability is statistically different from zero at the 95 % confidence level.

Upon initial inspection, the results regarding men and women may seem a bit counter-intuitive. However, the findings are in line with previous research regarding the school starting age. As I mentioned in section 2, Fredriksson and Öckert (2013) show that the school starting age has a larger positive effect on educational attainment for women in Sweden. Based on the observation that parents are more likely to enroll boys born in January earlier in school than girls, Fredriksson and Öckert hypothesize that the observed difference in the effect between men and women may be explained by the underperformance

¹⁷I use chi-squared test to study whether the regression coefficients for females and males are significantly different from each other. The results are shown in appendix in table A4.

Table 12: The Heterogeneous Effects of SSA

	Gender		Parental education		Mother's education	
	Female (1)	Male (2)	Low (3)	High (4)	Low (5)	High (6)
A) GPA						
First stage	0.798 (0.005)	0.690 (0.005)	0.745 (0.004)	0.707 (0.015)	0.746 (0.004)	0.720 (0.011)
Reduced form	0.202 (0.017)	0.098 (0.018)	0.162 (0.014)	0.145 (0.040)	0.150 (0.014)	0.150 (0.032)
IV	0.253 (0.022)	0.142 (0.026)	0.218 (0.019)	0.206 (0.057)	0.201 (0.018)	0.208 (0.045)
B) Admission to GUSS						
First stage	0.798 (0.005)	0.690 (0.006)	0.745 (0.004)	0.707 (0.015)	0.746 (0.004)	0.720 (0.011)
Reduced form	0.047 (0.008)	0.024 (0.008)	0.043 (0.006)	0.002 (0.016)	0.037 (0.006)	0.021 (0.013)
IV	0.059 (0.010)	0.034 (0.011)	0.058 (0.009)	0.003 (0.023)	0.050 (0.008)	0.029 (0.019)
C) GUSS graduate						
First stage	0.798 (0.005)	0.690 (0.006)	0.745 (0.004)	0.707 (0.015)	0.746 (0.004)	0.720 (0.011)
Reduced form	0.036 (0.008)	0.014 (0.008)	0.031 (0.006)	0.024 (0.013)	0.026 (0.006)	0.018 (0.012)
IV	0.045 (0.010)	0.020 (0.011)	0.041 (0.008)	0.035 (0.019)	0.035 (0.008)	0.025 (0.017)
Observations	76,426	78,933	123,975	10,322	136,525	18,834

Notes: Reduced form estimates are obtained using equation 8. First stage estimates are from equation 10. IV-estimates are calculated using two-stage least squares methods. I use same bandwidth of ± 30 days around the cutoff with all estimates. Robust standard errors are presented in the parentheses. Parents are defined as high educated if both of them have a tertiary degree and low educated if none of them have.

of men who suffer from the earlier school start. Something related may be behind the differences in Finland. I report the approximated first stage relationship for women and men in columns 1 and 2 of table 12. If my approximation is correct, the results indicate that girls are more likely to follow the school starting rules, which could be a consequence of parents' overoptimism on girls' maturity or carefulness regarding boys' school readiness.

The findings I present above show that there may exist interesting results beyond the averages. Motivated by this, let us move further to inspect the effects of SSA conditional on the parents' educational attainment listed in columns 3-6 of table 12.

A general observation is that the importance of parents' educational attainment seems to depend on outcome, even though the differences between groups are relatively small.¹⁸ For instance, as shown in the columns 3-6 in section a of table 12, I find that the effect of SSA on GPA for the individuals with high- and low-educated parents are close to each other. By contrast, differences in estimates measuring the probability of admission to general upper secondary school are larger. The point estimate for individuals with high-educated parents is close to zero, whereas the corresponding number for the children with low-educated parents is notably larger. But if we look on the effect on the probability to graduate from general upper secondary school presented in columns 3-6 of B panel, a bit surprisingly, the difference is again small. This is mainly due to the estimate for individuals with the high-educated parents which is again substantially larger than zero.

The result for children with high-educated parents is a bit puzzling. What could explain why first the school starting age does not have affect admission to general upper secondary school, but then later has an effect on the probability of graduation from general upper secondary school? One explanation could be that regardless of the success in comprehensive or school starting age, high-educated parents may have more influence on the decision of whether to attend general upper secondary school. Thus, we do not observe differences in the probability to be admitted to general upper secondary school between individuals born just before or after New Year. However, as individuals reach adulthood during the general upper secondary, they become more independent. This could imply that the skill or maturity differences arising from school starting age differences may matter more, and hence the gap may appear again.

Previous studies do not give a coherent picture of how parents background should affect the estimates. For instance, Elder and Lubotsky (2009) report that the effect of relative school starting age is notably larger for children from higher socioeconomic backgrounds. By comparison, Fredriksson and Öckert (2013) find that the relative school starting age has a larger effect for individuals with low-educated

¹⁸I use chi-squared test to study whether the regression coefficients for females and males are significantly different from each other. The results are shown in appendix in table A4.

parents. Although the differences between groups are small, my findings seem to point in the direction that Fredriksson and Öckert document. However, it should be emphasized that my definition of low-educated maybe a bit too extensive in the sense that it includes parents who have graduated from general upper secondary school. It is possible that for children whose parents do not have an upper secondary degree, the effect could be different, but unfortunately I am not able to identify whether the parents have graduated from upper secondary school.

Lastly, I examine how the effect of school starting age varies by region. Finland is divided into 19 regions (*maakunta*). Local decisions on education are made at the municipality level and, hence the regions themselves do not share coordinated education policies which would generate differences between areas. In any case, they offer a convenient way to examine whether the effect varies by region. Figure 6 presents the estimates separately for 18 regions in Finland.¹⁹ In each graph of figure 6, a dot represents an intention-to-treat estimate obtained using equation 8 for a certain region. The IV estimates are shown in figure 7. The vertical lines on top of the dots represent 95 % confidence intervals. I find that most of the estimates are in the vicinity of each other, and in addition, since the point estimates are relatively imprecise, it cannot be argued there exist significant differences between regions. Nevertheless, it is worth mentioning that in some regions, the point estimates are notably larger or smaller than the average effect. For example in Keski-Suomi, children born just after the threshold have 0.258 grade points higher GPA than individuals born just before. By contrast, in some regions the effect seems to be completely absent. I estimate that the effect of SSA on GPA is very close to zero in Pohjanmaa and Keski-Pohjanmaa. Moreover, in the same regions, the point estimates measuring the relationship between school starting age and probability of admission or graduate from general upper secondary school are even negative.

Although it is beyond the scope of this paper to figure out why the point estimates differ notably in some regions, I shall list a few possible explanations. First, it could be that the reason is related to the loose grading instructions in comprehensive schools. As I explained briefly in section 3, when comprehensive school was established, teachers were guided to base grading on relative comparisons. The regime was later officially abolished, but there is evidence suggesting that teachers are still comparing students within the class (Harju-Luukkainen et al., 2016). In addition, since there are no national exams at the end of comprehensive school, teachers have much freedom in assessment. Thus, it is possible that in some regions, teachers are more inclined to compare children relative to each other, which could amplify the effect. However, the findings of Harju-Luukkainen et al. (2016) seem to contradict this hypotheses. They find that students tend to underperform considerably in the Pisa literacy test compared to what is predicted by their mother tongue grades especially in the regions Pohjanmaa and Keski-Pohjanmaa.

¹⁹There are 19 regions, but I exclude Ahvenanmaa from this analysis, since there are so few observations.

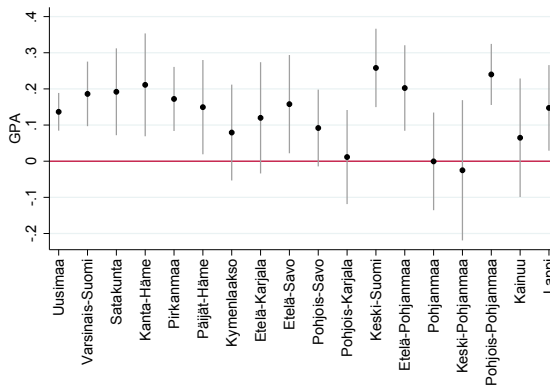
I would expect that the effect would be larger in these regions, but the point estimates I shown in figure 6 are close to zero.

On the other hand, the differences may be related to culture or parental background. Maybe in some regions, education is not valued much and children do not gain considerably from the relative age difference. It is also possible that the effect interacts with socioeconomic background, and hence the variations I observe between regions are caused by socioeconomic differences. This would conflict with the result I show above, which indicates that parents' education does not seem to have a significant effect on results. However, I also mentioned that my definition of low educated may be too broad.

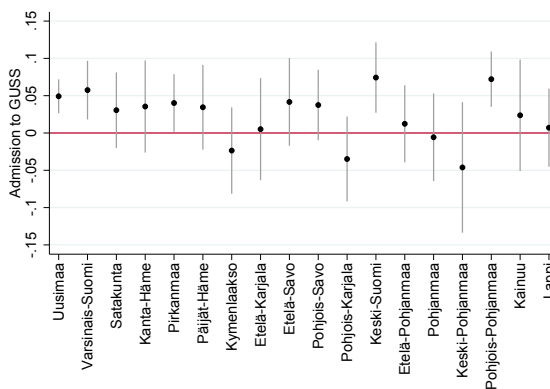
To conclude, based on my findings, there is some heterogeneity in the results. The effect of school starting age is significantly larger for women than for men which may be due to the differing enrollment behaviors. I also examine the effect by parental education and find that there do not exist notable differences between groups. However, note that I am not able to identify group whose parents are not general upper secondary school graduates, who might be the group benefiting most from a higher school starting age. Last, I show how the point estimates vary by region. I detect some notable differences between regions but all estimates are relatively imprecise.

Figure 6: The ITT Estimates by Region

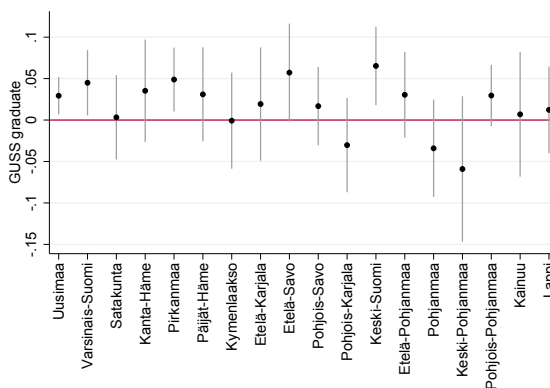
(a) GPA



(b) Admission to GUSS



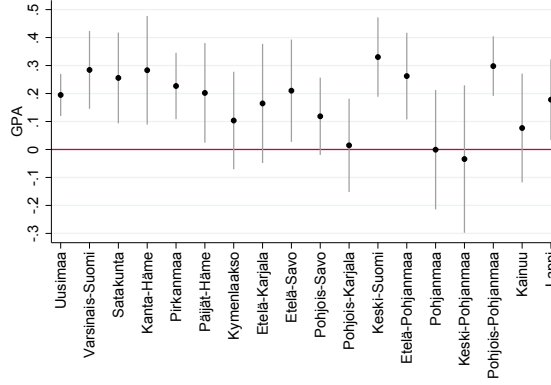
(c) GUSS graduate



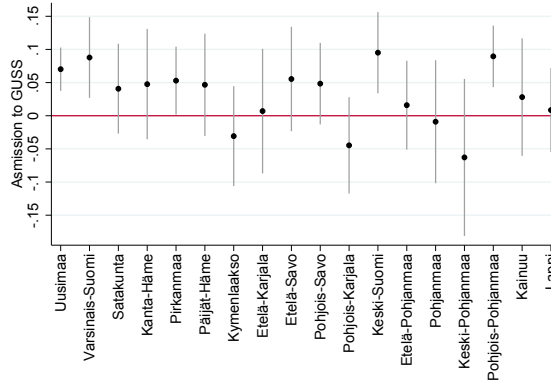
Notes: Each dot corresponds to a ITT estimate for specific region. Vertical lines represents 95% confidence intervals. The estimates are also shown in appendix in tables A6-A8.

Figure 7: The IV Estimates by Region

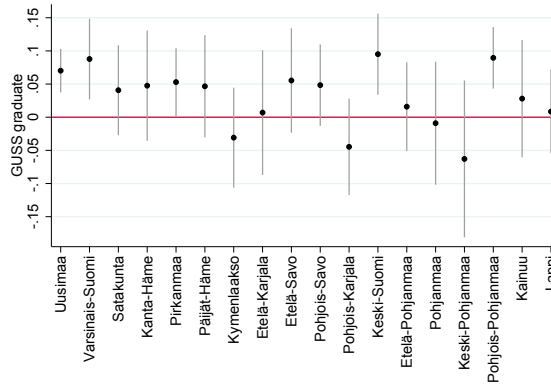
(a) GPA



(b) Admission to GUSS



(c) GUSS graduate



Notes: Each dot corresponds to a IV estimate for specific region. Vertical lines represents 95% confidence intervals. The estimates are also shown in appendix in tables A6-A8.

8 Conclusion

In this paper, I study the causal effect of school starting age on educational outcomes in Finland, using the school starting rule as a source of exogenous variation. In Finland, as a rule, individuals should start school during the calendar year they turn seven. As a consequence, there is a discontinuous jump in the school starting age at New Year, but otherwise children born just before and after New Year have on average very similar background characteristics.

I use a regression discontinuity design to investigate how the discontinuous jump in the school starting age affects the grade point average at the end comprehensive school, the probability of admission and to graduate from general upper secondary school. My main data source is the joint application register of the Finnish National Board of Education which is linked to the Finnish Longitudinal Employer-Employee Data. I restrict the analysis to individuals who take part in joint application process during the year they graduate from comprehensive school. As I do not observe the exact school starting age and the school starting rule is not binding, the main focus is on intention-to-treat estimation. In addition, I approximate the school starting age using the graduation year from comprehensive school to perform fuzzy regression discontinuity design analysis.

I estimate the effects using a window of ± 30 days around New Year. According to my results, individuals born just after New Year have on average 0.15 grade points higher GPA than individuals born just before. Furthermore, I show there are significant jumps in the probability of admission and graduation from general upper secondary school. When the non-compliance is taken into account, I find the effect to be even larger. However, I suspect my school starting age approximation suffers from measurement error and hence the IV estimates are biased upwards. Using the density of the running variable and individual's background covariates I demonstrate that these findings may be taken as causal evidence. In addition, I show that the results are insensitive to bandwidth and functional form. Lastly, I examine the effect of school starting age conditional on gender, family background and region and show that the effect is significantly larger for females than for males. The effect also varies by parent's educational attainment and regions, but the deviations between the groups are not significant.

My findings are comparable to what has been documented in other countries. I estimate that the magnitude of the jump in GPA is around 0.13-0.18 standard deviations, which is similar what Bedard and Dhuey (2006) document for most OECD countries. However, considering the previous evidence regarding Finland, my results are a bit surprising. For instance, Bedard and Dhuey (2006) and Pehkonen et al. (2015) find effects of a smaller magnitude. The difference between my findings and what Bedard and Dhuey present may be due the fact that they use standardized test scores whereas in Finland GPAs are based on teachers' personal assessment.

There are at least three different mechanisms through which the relative school starting age may contribute to educational outcomes. First, older children may have started school at more optimal age and, hence learn at a faster rate. Second, the peer effects may give older children additional confidence, which improves their results. It is also possible that the performances of older students may increase the motivation of younger students and nudge them to try harder. Last, it is suggested that older students achieve better exam results since they do the exams at an older age. I cannot distinguish between these mechanisms. Thus, my findings should be regarded as the combined effect of these channels.

My results give parents whose children are born at the end of the year reason to think carefully about whether they should postpone their children's school start. To provide more precise policy advice, more information on which of the mechanisms dominate is needed. Notice that even though postponing school start may seem beneficial from the parents' perspective, it is unclear that it is the optimal policy for the whole society. For instance, I show that a higher relative school starting age increases the probability of admission and graduation from general upper secondary school. Thus, parents may find it optimal to postpone their child's school start. Whether or not this increases welfare from the point of view of the society depends on the actual causal channel. If the difference I observe is due to the more optimal school starting in terms of human development, later enrollment may increase total welfare. On the other hand, if deviations in test scores are caused by peer effects or age-at-test effect, postponed school enrollment may have negligible effects on the individual and even negative effect on other individuals. If relatively older children achieve better results only because of the age-at-test effect, it follows that children who have started school older have similar skill sets, but fewer years in the workforce, which again may decrease their life-time earnings. When deviations are caused by peer effects, the decision may be regarded as a zero-sum game. For instance, if a child benefits from being older than the peer group, postponing school start may make the individual better off, but at the same time make the peer group worse off.

Thus, it would be important to distinguish between the mechanisms. The studies by Black et al. (2011) and Crawford et al. (2010) suggest that the deviations are mainly caused by the age-at-test effect. In that case, postponing school start is not an optimal policy unless the child has serious learning disabilities. However, since I show that a child born just after New Year has a significantly higher probability of admission and graduation from general upper secondary school, it is possible that school starting age may create persistent deviations between individuals born just after and before the cutoff. Therefore, future research which like Fredriksson and Öckert (2013), examines individual's life-cycle outcomes, is required.

References

- Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3):450–477.
- Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.
- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.
- Barua, R. and Lang, K. (2016). School entry, educational attainment, and quarter of birth: A cautionary tale of a local average treatment effect. *Journal of Human Capital*, 10(3):347–376.
- Basic Education Act (1998). Basic education act 21.8.1998/628.
- Bedard, K. and Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 121(4):1437–1472.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2):455–467.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Card, D. and Krueger, A. B. (2000). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania: Reply. *American Economic Review*, 90(5):1397–1420.
- Cook, P. J. and Kang, S. (2016). Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *American Economic Journal: Applied Economics*, 8(1):33–57.
- Crawford, C., Dearden, L., and Meghir, C. (2010). When you are born matters: The impact of date of birth on educational outcomes in england. Working paper.
- Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.
- Dobkin, C. and Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review*, 29(1):40 – 54.

- Elder, T. E. and Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers. *Journal of Human Resources*, 44(3):641–683.
- EU (2016). Compulsory education in europe - 2016/17. *Eurydice Facts and Figures*. Luxembourg: Publications Office of the European Union.
- Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20(4):2008–2036.
- Finnish National Board of Education (2016a). Joint application register.
- Finnish National Board of Education (2016b). Oppilaan arviointi menneinä vuosikymmeninä. http://www.edu.fi/perusopetus/oppilaan_arviointi/artikkeleita/oppilaan_arviointi_menneina_vuosikymmenina/. Accessed: 2017-02-19.
- Fredriksson, P. and Öckert, B. (2013). Life-cycle effects of age at school start. *The Economic Journal*, 124(579):977–1004.
- Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.
- Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N., and Bernelius, V. (2016). Differences between students' pisa reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the finnish pisa 2009 data. *Education Inquiry*, 7(4).
- Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.
- Landersø, R., Nielsen, H. S., and Simonsen, M. (2016a). Family responses to major life events: Evidence from linking school starting age and family outcomes. Working paper.
- Landersø, R., Nielsen, H. S., and Simonsen, M. (2016b). School starting age and the crime-age profile. *The Economic Journal*.
- Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698 – 714.

- Ouakrim-Soivio, N. (2013). Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina. Opetushallitus.
- Pehkonen, J., Viinikainen, J., Böckerman, P., Pulkki-Råback, L., Keltikangas-Järvinen, L., and Raitakari, O. (2015). Relative age at school entry, school performance and long-term labour market outcomes. *Applied Economics Letters*, 22(16):1345–1348.
- Puhani, P. A. and Weber, A. M. (2007). Does the early bird catch the worm? *Empirical Economics*, 32(2):359–386.
- Semkina, S. (2016). Vm: Koulun aloitus 6-vuotiaana säästäisi miljardin. *Kauppalehti (June 17, 2016)*.
- Shigeoka, H. (2015). School entry cutoff date and the timing of births. Working Paper 21402, National Bureau of Economic Research.
- Statistics Finland (2009). Peruskoulun oppilaat ja päättötodistuksen saaneet 1990-2009.
- Statistics Finland (2016). Finnish longitudinal employer-employee data.
- Virtanen, H. (2016). *Essays on Post-Compulsory Education Attainment in Finland*. PhD thesis, Aalto University.

Appendix

Table A1: Results with the Full Discontinuity Sample

	(FT)	(RF)	(IV)
GPA	0.739 (0.004)	0.148 (0.013)	0.200 (0.017)
Admission to GUSS	0.739 (0.004)	0.035 (0.006)	0.047 (0.008)
GUSS graduate	0.739 (0.004)	0.024 (0.006)	0.032 (0.008)
Observations	156,746	156,746	156,746

Notes: The estimates are obtained using a sample which includes observations that were excluded from the main analysis. The width of the window, is same ± 30 days around the cutoff, as in the main analysis. The first stage results are estimated using equation 10, reduced form estimates are from equation 8 and IV estimates are calculated using two-stage least squares method. Robust standard errors are presented in parentheses.

Table A2: Functional Forms

Simple averages	$Y_i = \alpha + \theta_1 D_i + e_i$
Linear	$Y_i = \alpha + \theta_1 D_i + \beta_1 R_i + e_i$
Quadratic	$Y_i = \alpha + \theta_1 D_i + \beta_1 R_i + \beta_2 R_i^2 + e_i$
Quadratic interaction	$Y_i = \alpha + \theta_1 D_i + \beta_1 R_i + \beta_2 R_i^2 + \beta_3 R_i D_1 + \beta_4 R_i^2 D_1 + e_i$

Notes: These functional forms are used to obtain the results presented in table 11.

Table A3: Optimal Bandwidths

	Sharp RDD (1)	Fuzzy RDD (2)
GPA	57	23
Admission to GUSS	65	26
Graduation from GUSS	57	23

Notes: Table presents the optimal windows for the different outcomes obtained using method of Calonico et al. (2014). Window is measured in days.

Table A4: Significance Tests for Heterogeneous ITT Estimates

	H_0 (1)	Chi-squared statistic (2)	P-value (3)
A. Gender	$\theta_{1f} = \theta_{1m}$		
GPA		17.65	0.001
Admission to GUSS		4.51	0.034
GUSS graduate		3.89	0.049
B. Parental educational attainment	$\theta_{1hp} = \theta_{1lp}$		
GPA		0.16	0.691
Admission to GUSS		0.20	0.656
GUSS graduate		5.63	0.018
C. Mother's educational attainment	$\theta_{1hm} = \theta_{1lm}$		
GPA		0.00	0.995
Admission to GUSS		0.36	0.549
GUSS graduate		1.25	0.263

Notes: The table presents results from chi-squared test which evaluates whether the differences in ITT estimates between groups presented in table 12 are statistically significant.

Table A5: Significance Tests for Heterogeneous First stage estimates

	H_0 (1)	Chi-squared statistic (2)	P-value (3)
A. Gender	$\theta_{2f} = \theta_{2m}$		
OLD_i		222.83	0.001
B. Parental educational attainment	$\theta_{2hp} = \theta_{2lp}$		
OLD_i		5.88	0.0153
C. Mother's educational attainment	$\theta_{2hm} = \theta_{2lm}$		
OLD_i		5.11	0.0238

Notes: The table presents results from chi-squared test which evaluates whether the differences in first stage estimates between groups presented in table 12 are statistically significant.

Table A6: The Effect on GPA by Region

	FT (1)	ITT (2)	IV (3)	Observations (4)
Uusimaa	0.701 (0.008)	0.137 (0.027)	0.195 (0.039)	35,723
Varsinais-Suomi	0.655 (0.015)	0.186 (0.046)	0.284 (0.071)	12,416
Satakunta	0.751 (0.017)	0.192 (0.061)	0.256 (0.083)	7,192
Kanta-Häme	0.745 (0.021)	0.211 (0.073)	0.283 (0.099)	4,787
Pirkanmaa	0.759 (0.013)	0.172 (0.045)	0.227 (0.060)	12,753
Päijät-Häme	0.739 (0.019)	0.154 (0.067)	0.202 (0.091)	5,911
Kymenlaakso	0.765 (0.019)	0.0792 (0.068)	0.104 (0.089)	5,513
Etelä-Karjala	0.729 (0.023)	0.120 (0.076)	0.165 (0.109)	3,994
Etelä-Savo	0.751 (0.019)	0.158 (0.069)	0.210 (0.093)	5,369
Pohjois-Savo	0.774 (0.015)	0.0918 (0.054)	0.119 (0.071)	8,376
Pohjois-Karjala	0.781 (0.018)	0.012 (0.066)	0.015 (0.085)	5,904
Keski-Suomi	0.781 (0.015)	0.258 (0.055)	0.330 (0.072)	8,362
Etelä-Pohjanmaa	0.772 (0.016)	0.202 (0.060)	0.262 (0.079)	6,981
Pohjanmaa	0.633 (0.022)	-0.001 (0.069)	-0.002 (0.109)	5,553
Keski-Pohjanmaa	0.734 (0.030)	-0.025 (0.099)	-0.034 (0.134)	2,504
Pohjois-Pohjanmaa	0.805 (0.011)	0.240 (0.043)	0.298 (0.054)	13,652
Kainuu	0.845 (0.020)	0.065 (0.084)	0.077 (0.099)	3,295
Lappi	0.828 (0.020)	0.147 (0.0605)	0.178 (0.073)	6,811

Notes: The ITT estimates are obtained using equation 8. The FT estimates are from equation 10 and the IV estimates are obtained using two-stage least squares. Robust standard errors are presented in parentheses.

Table A7: The Effect on Probability of Admission to GUSS by Region

	FT (1)	ITT (2)	IV (3)	Observations (4)
Uusimaa	0.701 (0.008)	0.049 (0.012)	0.070 (0.017)	35,723
Varsinais-Suomi	0.655 (0.015)	0.058 (0.020)	0.088 (0.031)	12,416
Satakunta	0.751 (0.017)	0.031 (0.026)	0.041 (0.035)	7,192
Kanta-Häme	0.745 (0.021)	0.036 (0.032)	0.048 (0.043)	4,787
Pirkanmaa	0.759 (0.013)	0.040 (0.019)	0.0530 (0.026)	12,753
Päijät-Häme	0.739 (0.019)	0.035 (0.029)	0.047 (0.039)	5,911
Kymenlaakso	0.765 (0.019)	-0.024 (0.030)	-0.031 (0.038)	5,513
Etelä-Karjala	0.729 (0.023)	0.005 (0.035)	0.007 (0.048)	3,994
Etelä-Savo	0.751 (0.019)	0.042 (0.030)	0.055 (0.040)	5,369
Pohjois-Savo	0.774 (0.015)	0.038 (0.024)	0.048 (0.031)	8,376
Pohjois-Karjala	0.781 (0.018)	-0.035 (0.029)	-0.045 (0.037)	5,904
Keski-Suomi	0.781 (0.015)	0.074 (0.024)	0.095 (0.031)	8,362
Etelä-Pohjanmaa	0.772 (0.016)	0.012 (0.026)	0.016 (0.034)	6,981
Pohjanmaa	0.633 (0.022)	-0.006 (0.030)	-0.009 (0.047)	5,553
Keski-Pohjanmaa	0.734 (0.030)	-0.046 (0.045)	-0.063 (0.060)	2,504
Pohjois-Pohjanmaa	0.805 (0.011)	0.072 (0.019)	0.089 (0.024)	13,652
Kainuu	0.845 (0.020)	0.024 (0.038)	0.028 (0.045)	3,295
Lappi	0.828 (0.020)	0.007 (0.027)	0.009 (0.032)	6,811

Notes: The ITT estimates are obtained using equation 8. The FT estimates are from equation 10 and the IV estimates are obtained using two-stage least squares. Robust standard errors are presented in parentheses.

Table A8: The Effect on Probability of Graduation from GUSS by Region

	FT (1)	ITT (2)	IV (3)	Observations (4)
Uusimaa	0.701 (0.008)	0.029 (0.012)	0.042 (0.017)	35,723
Varsinais-Suomi	0.655 (0.015)	0.045 (0.020)	0.069 (0.031)	12,416
Satakunta	0.751 (0.017)	0.003 (0.026)	0.004 (0.035)	7,192
Kanta-Häme	0.745 (0.021)	0.035 (0.032)	0.047 (0.042)	4,787
Pirkanmaa	0.759 (0.013)	0.049 (0.020)	0.064 (0.026)	12,753
Päijät-Häme	0.739 (0.019)	0.031 (0.029)	0.042 (0.039)	5,911
Kymenlaakso	0.765 (0.019)	-0.001 (0.030)	-0.001 (0.039)	5,513
Etelä-Karjala	0.729 (0.023)	0.019 (0.035)	0.0267 (0.0480)	3,994
Etelä-Savo	0.751 (0.019)	0.057 (0.030)	0.076 (0.040)	5,369
Pohjois-Savo	0.774 (0.015)	0.017 (0.024)	0.022 (0.0313)	8,376
Pohjois-Karjala	0.781 (0.018)	-0.030 (0.029)	-0.039 (0.037)	5,904
Keski-Suomi	0.781 (0.015)	0.065 (0.024)	0.084 (0.031)	8,362
Etelä-Pohjanmaa	0.772 (0.016)	0.030 (0.026)	0.039 (0.034)	6,981
Pohjanmaa	0.633 (0.022)	-0.034 (0.030)	-0.054 (0.047)	5,553
Keski-Pohjanmaa	0.734 (0.030)	-0.059 (0.044)	-0.080 (0.061)	2,504
Pohjois-Pohjanmaa	0.805 (0.011)	0.029 (0.019)	0.037 (0.024)	13,652
Kainuu	0.845 (0.020)	0.007 (0.038)	0.008 (0.045)	3,295
Lappi	0.828 (0.020)	0.012 (0.027)	0.015 (0.032)	6,811

Notes: The ITT estimates are obtained using equation 8. The FT estimates are from equation 10 and the IV estimates are obtained using two-stage least squares. Robust standard errors are presented in parentheses.