FACULTAD ECONOMÍA Y NEGOCIOS UNIVERSIDAD DE CHILE
DEPARTAMENTO DE ECONOMÍA

SERIE DE DOCUMENTOS DE TRABAJO

SDT 429

The Impact of Grade Retention on Juvenile Crime

Autores:
Juan Díaz
Nicolás Grau
Tatiana Reyes
Jorge Rivera

Santiago, Diciembre de 2016

sdt@econ.uchile.cl
econ.uchile.cl/publicaciones

THE IMPACT OF GRADE RETENTION ON JUVENILE CRIME

Juan Díaz
Harvard University

Nicolás Grau
University of Chile

Tatiana Reyes
University of Chile

Jorge Rivera
University of Chile

This version: December, 2016

Abstract

Implementing a fuzzy regression discontinuity design, we estimate a local causal effect of grade retention on juvenile crime. We assemble a novel data set that merges administrative information on schooling and juvenile crime for the entire population of students during the period 2007 – 2014 in Chile. Our main finding shows robust evidence that repeating a grade in school increases the probability of juvenile delinquency by 1.8 percentage points (pp), an increase of 37.5% of that probability. This effect is higher for males, and twice the indicated value for students of low socioeconomic status. We also show that grade retention increases the probability of dropping out of school by 1.5 pp. Regarding mechanisms, our findings suggest that the effect of grade retention on crime does not only manifest itself indirectly as a result of its effect on dropping out. We also show that the effect of grade retention on crime is greater when students switch schools right after failing a grade.

Keywords: Juvenile Crime, Grade Retention, Regression Discontinuity, Matching. **JEL Classification: I21, K42, and C26** .

We thank Aureo De Paula, Claudio Ferraz, Guido Imbens, Jose Zubizarreta, Miguel Urquiola, Zelda Brutti and seminar participants at Universidad de Chile, Sao Paulo School of Economics, UDP, Workshop on Human Capital (RIDGE Forum 2016), LACEA–LAMES 2016, Workshop on Applied Economics of Education (Catanzaro, 2016), and the 2016 North American Summer Meeting of the Econometric Society for valuable comments and suggestions. We thank the staff at Defensoría Penal Pública and Ministerio de Educación de Chile. Jorge Rivera thanks the partial financial support of Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816). Juan Díaz and Nicolás Grau thank the Centre for Social Conflict and Cohesion Studies (CONICYT/FONDAP/15130009) for financial support. All remaining errors are our own.

1 Introduction

Does grade retention in school make it more likely that young people will engage in criminal activity? From an opportunity cost point of view, it seems reasonable that students are more prone to pursue non-educational activities when they are not promoted to the next grade in school (Lochner (2004)). Conversely, repeating a grade might strengthen a student's knowledge and discipline, with potential positive effects on her outcomes. Thus, instead of representing a "cost" for students, not being promoted to the next grade could be viewed as an "opportunity" that may help them become more competitive in the classroom, discouraging the divergence to non-educational activities (Jacob (2005)). This ambiguity on the potential effect of retention on crime is at the core of the well-known "grade retention controversy".[1]

An empirical settlement of this controversy should be of particular importance in developing countries, where both the rate of students repeating grades and the rate of juvenile crime are much higher than those observed in developed countries. While in 2012 the average rate of grade retention in primary education was 5.1% in developing countries, that figure was 1.4% for developed countries (Institute for Statistics, UNESCO). In Chile, although its rate was below the average of developing countries, there has been an increase in the last decade, when the rate went from 3.1% in 1999 to 3.8% in 2012. Regarding crime, while the incarceration rate is 145.5 inmates per $100,000$ in the OECD countries,[2] in Chile this number is 266.[3]

Despite the vast literature linking grade retention and youth crime,[4] the evidence of a causal effect between them is scarce, and does not exist for developing economies, as

[1]The "grade retention controversy" exists because of ambiguous, and even contradictory, evidence of the effect that this measure has on some academic and socio-emotional outcomes of students. See Holmes et al. (1989) and Jimerson (2001); see also Reschly and Christenson (2013) for a fresh look at this controversy.

[2]An exception among developed countries seems to be the US, both in grade retention and crime rate. For instance, Wu, West, and Hughes (2010) state that in Texas, during the 2003-2004 school year, retention in first grade was 6.4%. Moreover, its incarceration rate is 710 inmates per $100,000$.

[3]European Institute for Crime Prevention and Control, affiliated with the United Nations (2010).

[4]For instance, Burdick-Will (2013), Fagan and Pabon (1990) and Hirschfield (2009), among others, have shown how criminal activities affect some schooling outcomes, a sort of inverse of the problem studied here. The effect of compulsory schooling laws on crime has been investigated by Lochner and Moretti (2004), Brugård and Torberg (2013) and Machin, Marie, and Vujić (2011), among others. Other contributions have investigated how school starting age may affect crime (see Landersø, Nielsen, and Simonsen (2016), and references therein).

discussed later on this section. It is rarely possible to find a proper empirical setting and dataset to overcome the potential endogeneity due to the fact that the latent outcome –crime activity that would be observed in the absence of grade retention– and the propensity to fail a grade are simultaneously determined.

To fill this gap, this paper estimates the causal effect of grade retention on juvenile crime by using an exceptional database from Chile, which matches individual academic records for all students (1st to 12th grade) with youth penal prosecution information, also on an individual basis, during the period 2007–2014. We find robust evidence on a (positive) local causal effect of grade retention on youth delinquency in Chile. Furthermore, as a byproduct of our main investigation, new findings on the relationship between retention and dropping out of school are also obtained. The results provide information about the mechanisms through which repeating a grade impacts the likelihood of committing a crime during youth.

Our identification strategy relies on a discontinuity in the probability of grade retention generated by the two grade retention rules commonly applied in Chile. They are based on the student's scores, which range from 1 to 7, with an increment of 0.1. The most prevalent rule employed by far, which we call "Rule I", applies when a student scores below 4 on two subjects (≤ 3.9) and has an average score across all subjects lower than or equal to 4.9. Our main results are based on the use of this rule. In addition, for the sake of completeness and to provide additional evidence supporting our main findings, we also estimate the causal effect using the second-most prevalent retention rule, "Rule II", which applies when a student scores below 4 on one subject and has an average score across all subjects lower than or equal to 4.4.

For estimation purposes, because there is evidence of manipulation around the threshold in the case of the forcing variable of Rule I (the student's second-lowest score), we follow Barreca, Guldi, Lindo, and Waddell (2011) in implementing a *donut-hole* fuzzy regression discontinuity design (FRD), where the observations in the immediate vicinity of the threshold for grade repetition are removed. To correct the potential bias due to differences in observable variables created by removing data, we complement this *donut-hole* approach by slightly extending the method developed by Keele, Titiunik, and Zubizarreta (2015) to implement an estimation procedure that combines matching with FRD. In the case of retention Rule II, because there is no evidence of manipulation of the forcing variable, we follow a more standard approach to estimate the effect, which is the FRD procedure developed by Calonico, Cattaneo, and Titiunik (2014b). Because of its stability to different specifications and the precision of its estimations, we focus our conclusions on the estimation magnitudes delivered by retention Rule I. However, the results from retention Rule II are in the same direction.

Our estimates show that repeating a grade between 4th and 8th grade increases the probability of juvenile crime by 1.8 pp, i.e., an increase of 37.5%. Moreover, we also find that this effect is larger for students of low socioeconomic status (SES), with an estimated effect twice that obtained for the entire population (the probability increases 3.7 pp). In terms of gender, the most affected by the policy are males, with an estimated increase of 2.5 pp in the probability of committing a crime during the juvenile period.[5]

As a secondary topic, we also examine the effect of grade retention on future grade retention and dropping out of school, both issues already studied in a series of recent contributions (Roderick (1994) and Manacorda (2012); see also King, Orazem, and Paterno (2015) for a comprehensive literature review). Using retention Rule I, we find that grade retention in primary school decreases the probability of grade retention in subsequent years by 5.9 pp (10.7%), but increases the probability of dropping out by 1.6 pp (23.8%).

Lastly, we also empirically study some mechanisms through which repeating a grade impacts the likelihood of committing a crime during youth. First, we find that the effect of grade retention on crime occurring before (or simultaneously to) dropping out is more pronounced than the effect on crime that occurs after dropping out. Second, we show that the effect of grade retention on crime is increased when students switch schools right after failing a grade. The first mechanism implies that our results add value to what we would have concluded if we had just put together two results already established in the literature, namely, the positive effect of grade retention on dropping out (Manacorda (2012)) and the positive effect of dropping out on juvenile crime (see, e.g., Anderson (2014), Fagan and Pabon (1990), and Thornberry, Moore, and Christenson (1985)).

As mentioned before, the literature on the causal effect of retention on crime is scarce.

[5] To study the robustness of our main findings, we implement a placebo test by replicating the "donut-hole" FRD and the FRD-matching estimations. In the last case, we consider only those students who did not repeat the grade, comparing those who scored below the threshold with those who scored above. These two methods do not deliver a statistically significant effect in this placebo test for any of the three outcomes considered, namely, juvenile crime, dropping out and future grade retention.

To the best of our knowledge, the closest paper to our investigation was provided by Depew and Eren (2015), who estimate the impact of grade retention (with summer school) on juvenile delinquency (and dropping out) in Louisiana. They assemble a novel dataset after merging administrative information on educational outcomes with the criminal records of students attending schools in Louisiana. Then, taking advantage of the test-based grade promotion policy that has been applied in Louisiana as of a decade ago, the authors build an RD design, where the forcing variable is the score on a standardized test which determines whether or not a student is promoted. Their principal conclusion is that, for students attending eighth grade, the test-based grade retention policy decreases the likelihood of being involved in felony offences during their youth. Although the authors make a remarkable effort in identifying a causal effect of grade retention on juvenile delinquency, they do not correct the latent manipulation that the forcing variable suffers close to the cut-off. Indeed, the key assumption in their RD is that teachers (or someone else in charge) do not exercise precise control over the score in the standardized test near the cut-off point. If this holds, the variation in scores obtained at the threshold is as good as randomized (Imbens and Lemieux (2008) and Lee and Lemieux (2010)).

Another similar study is by Cook and Kang (2016), who merge administrative data of academic performance with the criminal records of students attending public schools in North Carolina. They exploit the sharp RD design generated by the specific date that establishes the minimum age for school enrollment (the cut date) and assess its effect on a number of educational outcomes, as well as on juvenile crimes committed. Their main findings are, first, during middle school, students born just after the cut date (the oldest) are more likely to outperform (in math and reading) those born just before (the youngest), and are less prone to be involved in juvenile delinquency; second, those born before the cut date are more likely to drop out of school and commit a severe offence. Finally, Depew and Eren (2016), exploiting the same discontinuity as in op.cit., as well as using the aforementioned data for students attending school in Louisiana, find that late school entry by one year decreases the frequency of juvenile delinquency for young black females.

In summary, this paper makes three main contributions. First, together with Depew and Eren (2015), it is one of the first that estimates a causal effect of grade retention on juvenile crime and it is the first such evidence for a developing country, where the retention rates are higher. Second, by extending the method developed by Keele, Titiunik, and Zubizarreta (2015) to the *fuzzy RD* case, we present a method that might be useful in contexts where there is some evidence of manipulation in the forcing variable. Third, it sheds light on the mechanisms that may explain the impact of grade retention on juvenile crime.

This paper is organized as follows. In Section 2, we begin by describing the main features of both the educational and criminal data. The evidence about the discontinuity created by retention Rules I and II is also presented in this section. In Section 3, we present the main estimation strategy using retention Rule I, namely, a methodology that combines FRD with matching techniques. Section 4 presents our alternative strategy using grade retention under Rule II. In Sections 5 and 6, we show the results obtained from these two strategies. In Section 7, we present two exercises that shed light on how grade retention may impact juvenile crime. Finally, Section 8 concludes.

2 Data and Retention's Rules

In this section, we first describe the characteristics of our dataset and then explain how the grade retention rules operate.

2.1 Data

We assemble administrative dataset from the Ministry of Education and the Public Defender's Office (*Defensoría Penal Pública*, DPP). Among others, the DPP is the institution in Chile which provides free legal representation for all youths who have been accused of committing a crime.

The information collected from the Ministry of Education is an administrative panel dataset from 2002 to 2015, which, for every student in the country, indicates the school attended every year, the grade level (and whether the student has repeated the grade), the student's attendance rate, some basic demographic information, and (only for 2007) the annual average score for all subjects taken by the student (her cumulative grade point average).[6] The latter is needed to establish which students are close to the threshold for

[6]Hereinafter, to avoid confusion of the concept of grade (level) with grade (performance), the latter will be referred to as "score".

grade repetition and it also allows us to build a more continuous measurement for the average across all subjects.[7] We merge this panel with the information on performance on the national standardized test (*Sistema de Medición de Calidad de la Educación*, SIMCE), which is taken annually by all students in the 4th grade and every other year by all 8th grade students.

When students take the SIMCE, a survey is administered to their parents. From these surveys, we obtain information about the mother's and father's education level and family income. We focus our attention on the students who, in 2007, were in 4th to 8th grade, attending public or subsidized schools.[8] Due to their high SES and low criminal rate, we excluded students attending private schools, which represent 8% of the national enrollment.

The DPP's records contain information on all youth defendants in criminal cases tried in Chile during the period of January, 2006 to December, 2014. This database includes information on the time of the accusation, the type of offence, and the verdict (including the length of the sentence). In this study, we consider only juvenile criminal cases. In order to focus on crimes that can be thought of as motivated by a cost-benefit analysis, we omit individuals who committed the most severe crimes, such as murder or rape.[9] Given that our "treatment" is grade retention in 2007, we also exclude students who were prosecuted before 2008. Thus, in all our estimations, the students who committed crimes are those who were prosecuted, between 2008 and 2014, for an offence with an *economic motivation*.

The final dataset includes close to 640,000 students, their primary and secondary school records and their criminal records (ages 12 to 18). This information is linked to a large set of demographic characteristics about their families. In Group A of Table 1 we give in Section 3.1.3, it is possible to see the information at individual level that we

[7]For the other years, we only have the average across all subjects officially reported by the Ministry of Education. The problem with this measurement is that it is approximated. Thus, if we had this level of aggregation, in our estimation we would have to compare students with an average of 4.4 with students with an average of 4.5. Then, since students have around 10 subjects, this implies comparing students whose scores are different in all these subjects by 0.1 points on average, which would imply facing two different groups (treated versus control). Instead, our empirical strategy, which exploits Grade Retention Rule I, compares students who have differences in only one subject.

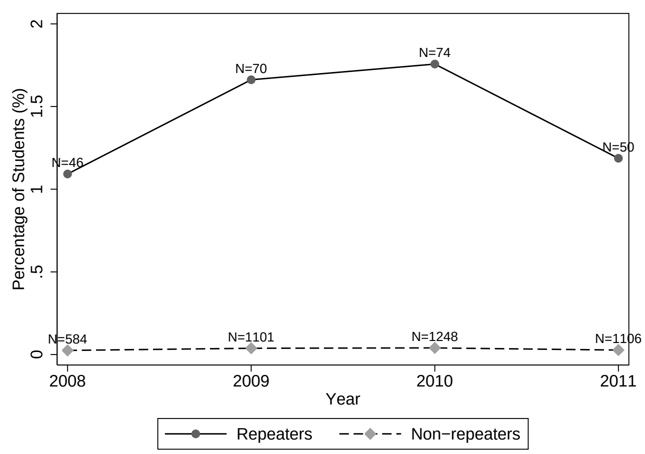
[8]We do not have SIMCE information for those students who were attending 7th grade in 2007. Thus, most of our estimations do not consider this group.

[9]We do not consider as crime the juvenile criminal cases where the verdict was *not guilty*.

use as independent variables, namely, whether they had repeated a grade before, their attendance rate pre-2007, their scores on standardized tests, the level of education of their parents, the family income, and gender. This table also shows the basic statistics of repeaters and non-repeaters in 2007; as we discuss later, these groups are rather different.

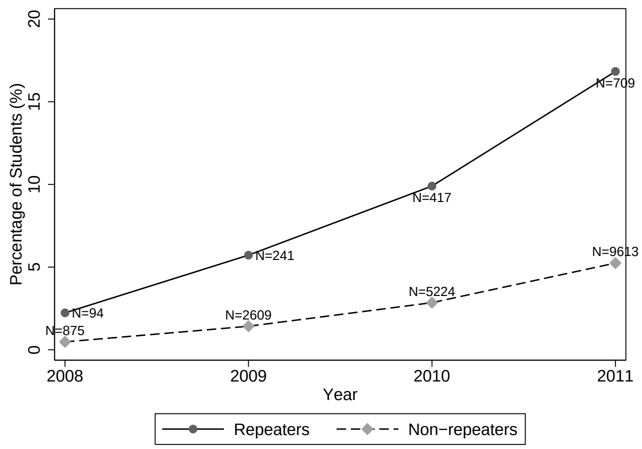
Regarding our dependent variables, Figures 1 and 2 show the evolution of juvenile crime and drop-out rates across years. To simplify interpretation of the dynamics, we only focus on the cohort of students who were attending 8th grade in 2007. To motivate the study on the effect of grade retention on crime and dropping out, we compare the dynamics between repeaters and non-repeaters of 2007. In the case of crime (Figure 1), each dot represents the fraction of youth who were criminally prosecuted each year for a first offense. Considering the 2008-2011 period, $4,279$ were prosecuted at least once, which represents 2.3% of this cohort. A higher fraction of repeaters were prosecuted, at 5.7% ($240/4,212$). In the case of dropping out (Figure 2), while $19,782$ dropped out at some point between 2008 and 2011, which represents 10.5% of this cohort, this percentage increases to 34.7% (1.461/4.212) among those who repeated in 2007.

Figure 1: Percentages of students committing a crime by year



Note: This figure considers students who were in 8th grade in 2007, for whom we have all the individual covariates used in our estimations. These criteria leave us with $187,611$ students, $4,212$ of whom repeated in 2007. In each dot, N represents the number of youths who were criminally prosecuted for a first offense in that year.

Figure 2: Percentages of students dropping out by year



Note: This figure considers students who were in 8th grade in 2007, for whom we have all the individual covariates used in our estimations. These criteria leave us with 187, 611 students, 4, 212 of whom repeated in 2007. In each dot, N represents the number of youths who dropped out of school that year.

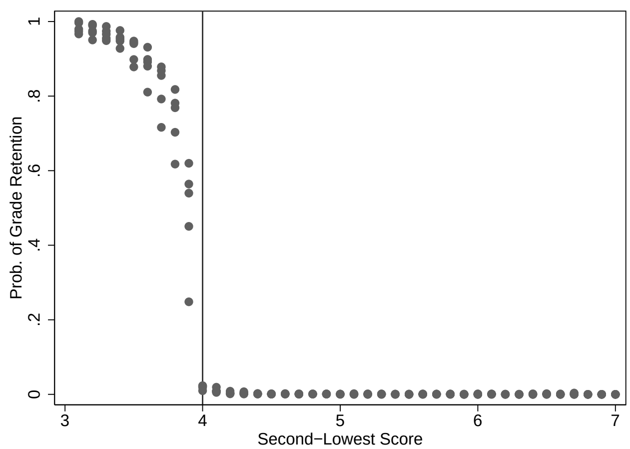
2.2 Retention Rules and Discontinuity

In order to present the grade retention rules employed in Chile, we recall that students' scores range from 1 to 7, with an increment of 0.1.[10] Using these records, the two most prevalent rules employed to determine grade retention are scoring below 4 on two subjects (≤ 3.9) and having an average score across all subjects lower than or equal to 4.9 (we call this grade retention "Rule I"), and scoring below 4 on one subject and having an average score across all subjects lower than or equal to 4.4 (grade retention "Rule II"). To have a sense about the relevance of these rules, among the 639, 092 students for whom we have all information considered in the application of the two rules, 9, 229 should repeat the grade because of Rule I (but they do not meet all conditions of Rule II), 479 should repeat the grade because of Rule II (but they do not meet all conditions of Rule I), and 7, 435 should repeat the grade because they meet the conditions for Rule I and II.

[10] At this stage, an important clarification is necessary. Although all schools must apply the 1-7 grading scale, they are free to set their own grading standards, which means that scores are not comparable across schools. This explains why in all of our estimations we compare students – below and above the threshold – who attend the same school and the same grade. This is also why, in all the plots that we present, we consider only students attending schools with at least one student scoring below and at least one student scoring above the specific threshold.

The retention rules suggest the possibility of implementing a regression discontinuity approach to study the causal effect of grade retention on crime. To explore the discontinuity due to grade retention Rule I, in Figure 3 each dot represents the grade retention rate of all the students in a particular grade who have a specific value of the second-lowest score. As the rule specifies, there is a strong discontinuity between scores 3.9 and 4.[11] Regarding the second rule, Figure 4 shows the discontinuity due to grade retention Rule II. In this case, each dot represents the grade retention rate of all the students in a particular grade who have a specific value of the average scores across all subjects. As the rule determines, there is an important discontinuity around 4.5.

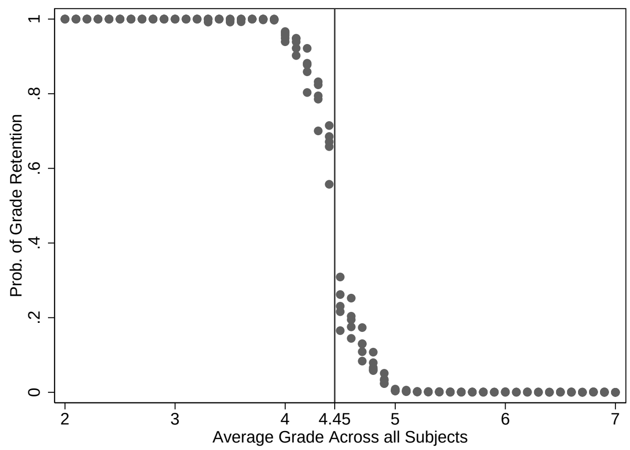
Figure 3: Discontinuity due to retention Rule I



Note: This figure considers only the performance in 2007 of 4th to 8th grade students attending schools which have at least one student scoring 4 or 4.1 and at least one student scoring 3.9 or 3.8 in their second-lowest score.

[11]The main reason why this is not a sharp discontinuity is that – as discussed – students who have two scores below 4 can pass the grade as long as their average across all subjects is greater than or equal to 5.

Figure 4: Discontinuity due to retention Rule II



Note: This figure considers only the performance in 2007 of 4th to 8th grade students attending schools which have at least one student scoring more than 4.4 and at least one student scoring less than 4.5 in their average score across all subjects.

3 Estimation Strategy Using Retention Rule I

In this section, we explore the validity of implementing an FRD design that exploits grade retention Rule I. Taking into account the the evidence of manipulation around the threshold in the case of the forcing variable of Rule I (the student's second-lowest score), we present a methodology that combines FRD with matching techniques.

3.1 Validity of the RD Design: Evidence of Local Manipulation

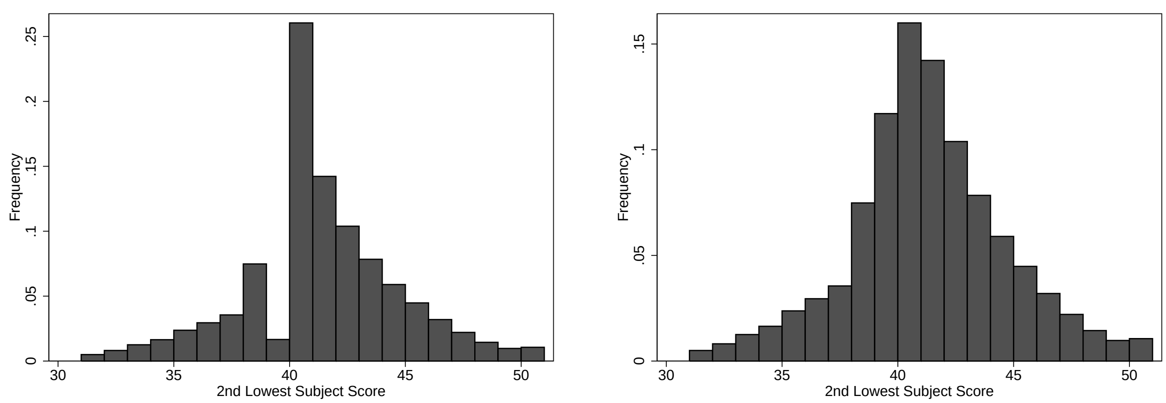
As we discuss in the following paragraphs, there are institutional reasons and empirical evidence to support the idea that the second-lowest score – the forcing variable – is manipulated around the threshold. However, we present evidence that this problem could be restricted to the scores closest to the threshold. The existence of this manipulation problem, and its local nature, is what determines our empirical strategies to estimate causal effects.

3.1.1 The Density of the Forcing Variable

Figure 5 shows two histograms for the second-lowest score. Panel (a) presents the real histogram (derived from data), while in Panel (b) a hypothetical histogram is introduced, which is created from Panel (a) by moving a number of students from scoring 4 to 3.9. There are two lessons to be gleaned from these plots. First, there is a remarkable discontinuity in the histogram for the second-lowest score, around the threshold $(3.9-4)$. Second, the discontinuity (and possibly the manipulation) seems to be limited to the scores closest to the threshold $(3.9-4)$; in fact, the histogram shown in Panel (b) does not show any evidence of discontinuity.

Figure 5: Histograms for the 2nd Lowest Score

(a) Actual (b) Hypothetical



The first point raises reasonable doubts about the internal validity of an RD estimator (see Lee and Lemieux (2010)), because it is arguable that teachers' grading decisions at the margin of repetition may not sort students randomly. The second point, which addresses local manipulation, is in line with the incentives that teachers face. In fact, even though the anecdotal evidence suggests that school leaders promote an upper bound for the rate of grade retentions, and, therefore, teachers may be *forced* to pass students who have a *real* score lower than 4, there is no reason to raise that score to a value higher than 4.[12] Moreover, if a student's real score (a latent variable) was 3.9 and her teacher manipulates that score to 3.8 to avoid any complaint from her parents (who might ask

[12] Teachers' grading behavior is not audited to find evidence of manipulation in their grading.

for a small increase from 3.9 to 4 to pass the grade), that would make our treatment and control groups more comparable. The relevance of this point will be more clear when we introduce our empirical strategy.

3.1.2 Graphical Test for Local Manipulation

We present direct evidence of local manipulation by taking advantage of the richness of our database. Intuitively, local manipulation should imply that the mapping from knowledge (a latent variable) to scores should be discontinuous around the threshold. In our framework, this means there should be a discontinuity in such a mapping between scores 3.9 and 4.

Fortunately, besides students' grade point average (GPA) at school, we have information on their standardized test scores (the SIMCE), where the latter can be thought of as unbiased proxies of students' knowledge. Thus, we can test manipulation by studying the behavior of the mapping from SIMCE to GPA around the threshold, at each primary school. We do so in the following steps:

1. To have the closest possible relationship between standardized tests and grade scores, we focus on the math SIMCE and math GPA for 8th grade students.[13]

2. Let i index students. We run the following OLS regression for each school s:

$$MathSimce_{is} = \mu_0^s + \mu_1^s * MathGPA_{is} + v_{is}.$$

3. We allow for a different mapping for each school, because schools may have different standards to evaluate their students.[14] Furthermore, to have enough precision in our estimated parameters, we exclude schools with fewer than 20 students. By doing so, we drop 1625 schools, keeping 4125 for our OLS estimations.

[13] To be clear, this means that the sample that we use to show local manipulation is different from the sample that we consider in our estimations of the effect of grade retention on crime. While in the former we only use 8th grade students and their math performance, in the latter we consider data from 4th to 8th grade and their average performance across subjects. Because we do not have standardized tests measuring all subjects, we are constrained to show local manipulation in math and to assume that it is also local for the other subjects.

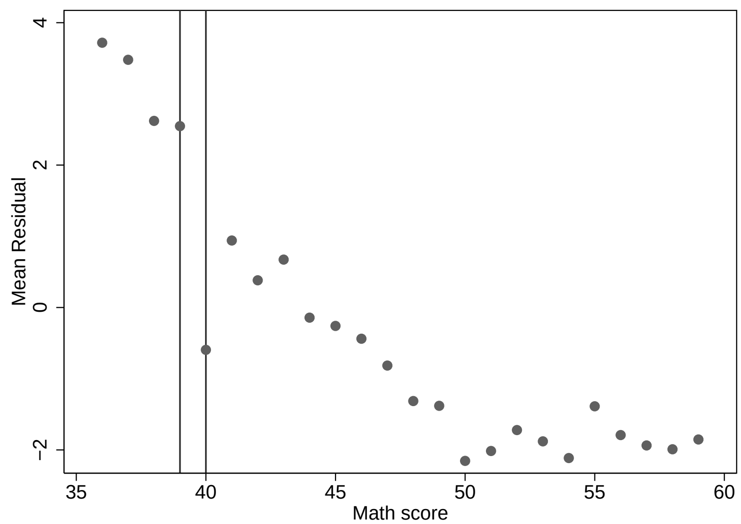
[14] In Figure 12 of Appendix A, we show how μ_0 (constant) and μ_1 (slope) are different across schools.

4. Given 4125 pairs of OLS estimations for μ_0 and μ_1, we calculate the residual for each student i, such that:

$$Residual_{is} = MathSimce_{is} - \hat{\mu}_0^s - \hat{\mu}_1^s * MathGPA_{is}.$$

In Figure 6, we present the mean of these residuals for each value of $MathGPA$. As can be seen in this figure, and even though there are other, smaller jumps in other parts of the math score range, there is a clear discontinuity between 3.9 and 4.

Figure 6: Test for Local Manipulation (second-lowest score)



This simple test has clear limitations. The most important one is the assumption of a linear relationship between knowledge (measured by the SIMCE) and school GPA.[15] Indeed, this assumption is what determines the negative slope of the mean of the residuals. That said, it is remarkable that, even imposing a linear relationship, the figure shows a clear jump only between 3.9 and 4.

[15] Another potential limitation is to assume that the SIMCE is an unbiased measure of the student's knowledge. In our opinion, given the way in which the SIMCE is taken (where regular teachers are not in the classroom during the test), there is no reason to think that SIMCE scores are manipulated.

3.1.3 Tests Involving Covariates

To study the extent to which this manipulation could be a problem and how useful it is to use a FRD approach in this context, Table 1 shows the differences in observables among different groups. In Group A, we compare students who were retained in 2007 with students who were not. In this selected sample, the normalized differences in the means of the independent variables are all economically relevant, ranging from 1.49 to 0.29.[16] Moreover, all of these differences are in the same direction: the repeaters are students with characteristics highly correlated with future criminal behavior. They come from lower socioeconomic groups (measured by income and parents' education), they have lower levels of academic performance, their attendance rate is lower, and males are overrepresented in this group.

This story contrasts to that of Group B, where we compare students who scored 3.9 in 2007 with those who scored 4, in their second-lowest subject. The stories from these two samples differ in two ways. First, the magnitudes of the normalized differences are remarkably smaller in Group B, where the largest normalized difference is 0.1. Second, in Group B, the signs of the differences in observables – between the highly probable repeaters and the rest – are in some cases in the opposite direction of those in Group A. For instance, students scoring 3.9 have a lower mean in repetition before 2007, and higher means in attendance in 2006 and in family income.

The comparison between these two selected samples (Groups A and B) illustrates how much we gain by taking advantage of the discontinuity. Without an RD approach, the initial differences between the treated and the control groups – presented in Group A – would be too large to implement an empirical method based on controlling observables (*e.g.,* a type of matching) as a credible approach to estimate a causal effect. That said, as was anticipated in the density analysis and in our test for local manipulation, Group B shows some evidence of manipulation around the threshold, because, without manipulation, students scoring 3.9 are expected to have worse performance and lower socioeconomic status on average than students scoring 4, which is not always the case with our data. Of particular note is the difference in the fraction of students who have previously repeated. A reasonable explanation for this difference is that teachers are less

[16] The normalized difference in the mean is equal to $\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(Sd(X_1)^2 + Sd(X_2)^2)/2}}$, where $\bar{X}_i$ is the sample mean for group i and $Sd(X_i)^2$ is the estimated variance for group i.

demanding with students who have previously failed a grade, which creates a non-random sorting around the threshold.

To address the sorting of students around the threshold, in Group C we compare students who scored 3.8 with those who scored 4.1 in their second-lowest subject. This selected sample has advantages and disadvantages, when compared to Group B. In the case of Group C, all of the mean differences in observables between students below and above the threshold have the expected signs, which is consistent with the evidence that the data is free of manipulation beyond 3.9 and 4. Regarding the disadvantages, we lose comparability between the groups below and above the threshold, particularly with respect to student performance. In sum, the remaining differences observed in Group C are much smaller than the differences observed in Group A and are arguably free of manipulation. However, they are large enough to suggest the need to complement the RD design with another approach, to at least control for the differences in observables.

Table 1: Differences in covariates among different treatments and control groups

			Group A: All				
Variable	Non Repeaters	Repeaters	Norm. Dif.	Statistic	p-value	N (Non Rep.)	N (Rep.)
Repeated before 2007	0.09	0.29	-0.54	-65.25	0.000	683,972	21,293
Attendance 2006	94.6	92.3	0.38	49.71	0.000	683,972	21,293
Math SIMCE	0.01	-0.89	1.00	154.64	0.000	683,972	21,293
Language SIMCE	0.01	-0.88	0.99	151.55	0.000	683,972	21,293
Mother Education	10.78	9.50	0.37	52.61	0.000	683,972	21,293
Father Education	10.86	9.70	0.32	45.76	0.000	683,972	21,293
Family Income	98,960	74,412	0.28	43.81	0.000	683,972	21,293
Male	0.50	0.64	-0.30	-42.24	0.000	683,972	21,293
			Group B: second lowest subject score $\in \{3.9, 4.0\}$				
Variable	Mean (= 4.0)	Mean (= 3.9)	Norm. Dif.	Statistic	p-value	N (= 4.0)	N (= 3.9)
Repeated before 2007	0.25	0.20	0.10	2.65	0.008	2,496	885
Attendance 2006	93.5	93.6	-0.02	-0.63	0.528	2,496	885
Math SIMCE	-0.58	-0.65	0.08	2.09	0.037	2,496	885
Language SIMCE	-0.58	-0.62	0.05	1.28	0.200	2,496	885
Mother Education	10.43	10.29	0.04	1.07	0.283	2,496	885
Father Education	10.54	10.39	0.04	1.08	0.282	2,496	885
Family Income	91,620	95,597	-0.04	-1.08	0.280	2,496	885
Male	0.56	0.60	-0.07	-1.68	0.093	2,496	885
			Group C: second lowest subject score $\in \{3.8, 4.1\}$				
Variable	Mean (= 4.1)	Mean (= 3.8)	Norm. Dif.	Statistic	p-value	N (= 4.1)	N (= 3.8)
Repeated before 2007	0.21	0.23	-0.05	-2.71	0.007	7,463	3,889
Attendance 2006	93.2	93.1	0.01	0.61	0.540	7,463	3,889
Math SIMCE	-0.54	-0.73	0.23	11.89	0.000	7,463	3,889
Language SIMCE	-0.57	-0.73	0.19	9.89	0.000	7,463	3,889
Mother Education	10.36	10.17	0.06	2.82	0.005	7,463	3,889
Father Education	10.52	10.40	0.04	1.84	0.065	7,463	3,889
Family Income	88,325	88,132	0.00	0.11	0.913	7,463	3,889
Male	0.57	0.58	-0.00	-0.17	0.865	7,463	3,889

Note: Norm. Dif. is the normalized differences in the means.

3.2 Empirical Approach Using Retention Rule I

By the evidence we have shown about the manipulation, we implement two different strategies to estimate the effect of grade retention on juvenile crime using Rule I for grade retention. In the first approach, which takes advantage of the local nature of the manipulation, we implement a standard fuzzy regression discontinuity design (FRD), but only using the students who scored 3.8 or 4.1 in their second-lowest score (Group C sample, Table 1). This method is known in the literature as the "donut-hole regression discontinuity"; see Barreca, Guldi, Lindo, and Waddell (2011). In the second approach (*FRD-matching*), which addresses the differences in covariates observed in the Group C sample, we combine FRD with the matching approach, named *design matching*, developed by Zubizarreta (2012).[17] As can be seen, the key difference between our two procedures is the definition of the sample used.

Let Y_i be a variable that takes the value one if the student committed a crime after 2007 and zero otherwise; Z_i a variable that takes the value one if the student's second-lowest subject score in 2007 is below the threshold and zero otherwise; W_i a variable that takes the value one if the student repeats the grade, and zero otherwise; and X_i a set of covariates of student i. Hence, as is shown in Hahn, Todd, and der Klaauw (2001), when the sample considered is close to the threshold, the identification of the Local Average Treatment Effect (LATE) is given by a type of Wald estimator, such that:

$$\tau_{FRD} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[W|Z=1] - E[W|Z=0]} \tag{1}$$

Furthermore, as pointed out by Imbens and Lemieux (2008), it is possible to obtain this Wald estimator by implementing a Two Stage Least Square method, where the first and second stages are described by:

$$First\ Stage: \quad W_i = \alpha_c^w + \alpha_z^w Z_i + \alpha_x^w X_i + \varepsilon_i^w, \tag{2}$$

$$Second\ Stage: \quad Y_i = \alpha_c^y + \tau_{FRD}\widehat{W}_i + \alpha_x^y X_i + \varepsilon_i^y. \tag{3}$$

In this context, the instrumental variable estimator of τ_{FRD} can be interpreted as the estimation of the LATE under the assumption of monotonicity (Angrist, Imbens and Rubin (1994)), which certainly holds in our case [18]. In our framework, the monotonicity assumption would be violated if a student would be promoted to the next grade if his second-lowest subject score was below the cut-off, but would be retained in the same grade if his second-lowest subject score was above the cut-off. Thus, using the second-

[17] This method is an extension of Keele, Titiunik, and Zubizarreta (2015), where the authors combine sharp regression discontinuity design with matching.

[18] This method is implemented in Stata using the command *ivreg*, with robust standard errors. It should be noted that this method of calculating robust standard errors does not take into account that the sample to implement the FRD estimation is built using a matching procedure.

lowest subject score threshold indicator as an instrumental variable produces a consistent estimation of the effect of Rule I on youth crime for the compliers (see Angrist, Imbens and Rubin (1994) and Imbens and Rubin (2015)), i.e. the students whose promotion to the next grade is affected by Rule I of grade retention.

The first empirical approach, the "donut-hole" FRD, is the standard FRD but excludes students whose second-lowest score is 3.9 or 4. Specifically, the sample consists of all students who score 3.8 or 4.1 in their second-lowest subject score, and who belong to a school-cohort with at least one student at each side of the threshold (Group C, Table 1).[19] Thus, given this restricted sample, the LATE is obtained by regressing Equations (2) and (3).

The second empirical approach, the FRD-matching method, has as its starting point the same sample as the first approach (Group C sample). The difference is that, in order to address the imbalance in observables between students scoring below and above the threshold, we use the *design matching* estimator to build similar groups. Unlike the standard matching methods, which attempt to achieve covariate balance by minimizing the total sum of distances between treated units and matched controls, this method achieves covariate balance directly by minimizing the total sum of distances while constraining the measures of imbalance to be less than or equal to certain tolerances. In our implementation of this matching, we optimally find a pair for each student scoring 3.8, selected from those who are attending the same school-cohort and score 4.1,[20] by minimizing the weighted distance in math and language standardized test scores, parents' education, previous grade repetitions, attendance during the past year, an income variable and gender, subject to mean balance on the same set of variables. Details of this matching approach are described in Appendix B. Then, using this matched sample of $2 * N_{bt}$ students,[21] we estimate the LATE by implementing the 2SLS estimator described by equations (2) and (3).

Table 2 presents the balance achieved by this matching procedure on the mentioned covariates. Comparing the differences observed in Table 2 with the differences presented

[19]This means a cohort within a school with at least one student scoring 3.8 and one student scoring 4.1 in the subject with the second-lowest score.

[20]In one specification, we also implement an exact match in gender.

[21]N_{bt} is the number of students who score below the threshold and who have a match – above the threshold – found by the design matching procedure.

in Group C in Table 1, it is clear that there is an improvement in terms of balance in observables.[22] However, there is an important reduction in the sample size (from 3889 to 2931 individuals below the threshold).

Table 2: Post-matching differences in covariates

Variable	4.1	3.8	Norm. Dif.	Statistic	p-value	N (= 4.1)	N (= 3.8)
Repeated before 2007	0.20	0.20	-0.01	-0.55	0.582	2,959	2,959
Attendance 2006	93.3	93.2	0.02	0.70	0.481	2,959	2,959
Math SIMCE	-0.65	-0.68	0.03	1.15	0.248	2,959	2,959
Language SIMCE	-0.65	-0.68	0.03	1.28	0.202	2,959	2,959
Mother Education	10.33	10.26	0.02	0.79	0.430	2,959	2,959
Father Education	10.51	10.48	0.01	0.34	0.731	2,959	2,959
Family Income	89,012	88,476	0.01	0.23	0.821	2,959	2,959
Male	0.57	0.58	-0.01	-0.39	0.693	2,959	2,959

Note: Norm. Dif. is the normalized differences in the means.

4 Estimation Strategy Using Retention Rule II

In this section, we explore the validity of implementing an RD design that exploits grade retention Rule II. Because in this case we do not find evidence of manipulation, we present procedure developed by Calonico, Cattaneo, and Titiunik (2014b), which implements a FRD that includes a correction for potential bias.

4.1 Validity of the RD Design: Evidence of no Manipulation

To corroborate that there is no evidence of manipulation in retention when using Retention Rule II, we implement the same graphical test that we discussed and implemented in studying Retention Rule I in the previous section. As before, this test takes advantage of the existence of a standardized test (the SIMCE) that is a non-manipulated measure of student knowledge. In this case, because we want to study the discontinuity in the relationship between the standardized test and the GPA, we use a simple average of the three tests for which we have scores: science, math and Spanish. We define this average

[22] We also tried to achieve this balance by implementing a more standard matching approach (*e.g.*, minimizing the mahalanobis distance). However, in that case, the improvement was only partial, probably due to the important number of school-cohort clusters for which there are few students scoring 4.1 who qualify as a match for those scoring 3.8.

as $AveSimce$. Thus, letting i index students, we run the following OLS regression for each school s:

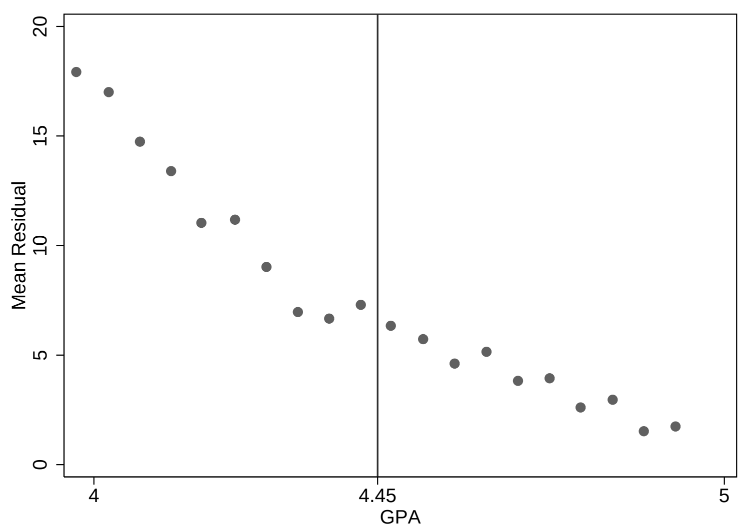
$$AveSimce_{is} = \mu_0^s + \mu_1^s * GPA_{is} + \upsilon_{is}.$$

As before, to have enough precision in our estimated parameters, we exclude schools with less than 20 students. We calculate the residual for each student i, such that:

$$Residual_{is} = AveSimce_{is} - \hat{\mu}_0^s - \hat{\mu}_1^s * GPA_{is}.$$

In Figure 7, we present the mean of these residuals for each value of GPA. As can be seen in this figure, as opposed to what we observe using the second-lowest score, there is no clear discontinuity around 4.45. Because of this, we do not need to exclude the data that is closest to the threshold as we did when we presented the *donut-hole* approach.

Figure 7: Test for Local Manipulation (average grade across all subjects)



Given that we can directly test the existence of manipulation, due to the richness of our data, we do not need to implement the commonly used indirect test of manipulation that studies the density of the running variable around the threshold.[23]

[23]In the next section, which discusses the results, we also present evidence of the robustness of our

4.2 Empirical Approach Using Retention Rule II

We implement a standard Fuzzy RD to estimate the effect of grade retention on juvenile crime using Rule II of grade retention. Before briefly presenting the methodology, some notation is introduced to define the estimator. Although we cannot use exactly the same notation as that used in the previous section because of the differences in the grade retention rules, we introduce notation as similar as possible. The running variable we employ is the GPA score in 2007, which is denoted by Z_i for the i-th individual. Its cut-off level is denoted by $\bar{z}$ (which in our scenario is 4.45). As before, the treatment indicator is W_i, which takes the value one if the i-th student repeats the grade in 2007, and zero otherwise; and Y_i is the observed outcome variable, which takes the value one if the i-th student commits a crime after 2007 and zero otherwise. Finally, X_i denotes a vector of covariates of the i-th student.

Thus, as is shown in Imbens and Lemieux (2008), the fuzzy RD estimand is given by:

$$\tau_{FRD} = \frac{\lim_{z\uparrow\bar{z}} \mathbb{E}[Y|Z=z] - \lim_{z\downarrow\bar{z}} \mathbb{E}[Y|Z=z]}{\lim_{z\uparrow\bar{z}} \mathbb{E}[W|Z=z] - \lim_{z\downarrow\bar{z}} \mathbb{E}[W|Z=z]} = \frac{\tau_Y}{\tau_W}.$$

As mentioned, we follow Calonico, Cattaneo, and Titiunik (2014b) to estimate τ_{FRD} as[24]:

$$\widehat{\tau}_{FRD}(h) = \frac{\widehat{\tau}_Y(h)}{\widehat{\tau}_W(h)},$$

$$\widehat{\tau}_Y(h) = \widehat{\alpha}_{Y,-}(h) - \widehat{\alpha}_{Y,+}(h), \qquad \widehat{\tau}_W(h) = \widehat{\alpha}_{W,-}(h) - \widehat{\alpha}_{W,+}(h),$$

for a positive bandwidth h, where, for $J = Y, W$ the estimators $\widehat{\alpha}_{J,-}$ and $\widehat{\alpha}_{J,+}$ come from

approach by running the procedure developed by Calonico, Cattaneo, and Titiunik (2014), but instead using our control variables as dependent variables.

[24]To estimate the fuzzy RD and draw inferences about the parameter of interest, we use the Stata routine called *rdrobust* developed by Calonico, Cattaneo, and Titiunik (2014a).

a standard local-linear RD estimator:

$$\begin{pmatrix} \hat{\alpha}_{J,-} \\ \hat{\alpha}_{J,+} \\ \hat{\beta}_{J,-} \\ \hat{\beta}_{J,+} \\ \hat{\gamma}_J \end{pmatrix} = \underset{\alpha_{J,-},\alpha_{J,+},\beta_{J,-},\beta_{J,+},\gamma_J}{\arg\min} \sum_{i=1}^{n} [J_i - \mathbb{1}_{\{Z_i<\bar{z}\}} \cdot (\alpha_{J,-} + \beta_{J,-} \cdot (Z_i - \bar{z})) -$$

$$\mathbb{1}_{\{Z_i\geq\bar{z}\}} \cdot (\alpha_{J,+} + \beta_{J,+} \cdot (Z_i - \bar{z})) - \gamma_J \cdot X_i]^2 \cdot \frac{K\left(\frac{Z_i-\bar{z}}{h}\right)}{h},$$

where $K(\cdot)$ is a kernel function. As mentioned, we are able to identify the effect because the monotonicity assumption holds, and because we are consistently estimating the causal effect on the compliers. It is also worth mentioning that we have added covariates in the specification and that, even though they are not required for the identification of τ_{FRD}, they increase the efficiency of the estimator. In the next section, we present results assuming a triangular kernel, as well as different choices of h equal to 0.1, 0.15, and 0.2.[25]

Intuitively, the estimator corrects for the potential misspecification bias of $\widehat{\tau}_{FRD}(h)$, which may be more significant when including covariates. The bias-corrected estimator is obtained after removing the estimator of the bias, which is computed through local polynomials. To draw inferences and calculate confidence intervals for this parameter, we use both the conventional and the robust nonparametric bias-correction procedures developed by Calonico, Cattaneo, and Titiunik (2014b) and Calonico, Cattaneo, and Titiunik (2016). In the latter, the inference is made after calculating the variance of the bias-corrected estimator: a combination of the variance of the point estimator of τ_{FRD} and the variance of the estimator of the bias.

5 Results Exploiting Retention Rule I

In this section, we present our findings on the impact of grade retention on juvenile crime, dropping out, and future grade retention. These results come from our first empirical strategy, which exploits Rule I of grade retention.

[25]We also used $h = 0.05$, which delivered point estimators that were even higher but not precise.

5.1 Effect of Grade Retention on Crime

The main results of this section are presented in Table 3. Focusing on the first two columns, which summarize the results of our first empirical strategy, we find that the effect of grade retention on crime ranges from 1.6 to 3.7 pp, and in almost all specifications the effect is statistically significant.[26]

Table 3: Effect of grade retention on juvenile crime

Sample	(1) Donut-Hole FRD	(2) FRD-Matching	(3) OLS
All	0.016**	0.018**	0.035***
	(0.0068)	(0.0082)	(0.0019)
	N = 9,681	N = 5,130	N = 705,261
Low SES	0.024**	0.037***	0.044***
	(0.0120)	(0.0142)	(0.0027)
	N = 4,527	N = 2,330	N = 359,021
Males	0.024**	0.025*	0.042***
	(0.0114)	(0.0147)	(0.0026)
	N = 4,187	N = 2,176	N = 353,552
First Repetition	0.011	0.015	0.034***
	(0.0079)	(0.0097)	(0.0021)
	N = 6,630	N = 3,338	N = 638,582

Note: In the case of FRD-Matching there are $N_{bt}/2$ students with their second-lowest score equal to 3.8 and $N_{bt}/2$ students with that score equal to 4.1. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The effect is heterogeneous and economically significant. The effect is larger for students from low SES, with an estimated effect twice that obtained for the entire population of students (the probability increases 3.7 pp).[27] In terms of gender, the most affected by the policy are males, with an estimated increase of 2.5 pp for the probability of committing a crime during the school period. Regarding the magnitudes, given that the crime rate for the students in this sample is about 4.8%, the estimates range from an effect of 33% to 77%.[28] Lastly, the OLS estimation delivers larger effects (biased

[26]The difference between the sample size of Columns 1 and 2 is due to the fact that, in Column 1, most of the time, there is more than one student scoring 4.1 for each student scoring 3.8, which is not the case in Column 2 (by construction).

[27]Low SES is defined as the group of students attending schools which fall below the median for a school's average income.

[28]A precautionary note about this range: these population groups also have different crime rates. For

upward), due to unobservable variables that affect the probability of committing a crime and that are correlated with repeating a grade (Column (3)). This issue further supports the soundness of the empirical method developed in this paper (*i.e.,* the FRD-matching).

5.2 Effects of Grade Retention on Other Outcomes

Given our rich panel dataset, we can also examine the effect of grade retention on other outcomes.[29] Specifically, Table 4 shows the effect of repeating the grade in 2007 on the probability of future grade retention.[30] In particular, not being promoted to the next grade in 2007 decreases the probability of future grade retention from 2.3 to 10.4 pp (Column (1)). Given that, in the estimation sample, 55% of the students repeat at least one grade after 2007, our finding represents a decrease ranging from 4.1 to 18.9%.[31]

---

example, the male rate is 6.7% (the female rate is 2.2%) and the crime rate for students attending low SES schools is 6.8%.

[29]We focus on the *donut-Hole* RD method, as opposed to FRD-matching, given that this approach presents the smaller point estimates in the placebo analysis, and it also delivers the smaller effects in all the estimations.

[30]Given that there are dropouts, there is a potential selection bias problem that we do not address in this paper.

[31]In the estimation sample, 55% of the students repeat at least one grade after 2007, which reflects two features of the data. First, the grade retention rate is remarkably high in Chile; in fact, the percentage for the entire population is 39%. Second, low-performing students are overrepresented in the estimation sample.

Table 4: Effect of grade retention on future grade retention

Sample	(1) Donut-Hole FRD	(2) FRD-Matching	(3) OLS
All	-0.082***	-0.059***	0.116***
	(0.0155)	(0.0191)	(0.0035)
	N = 9,681	N = 5,130	N = 705,261
Low SES	-0.050**	-0.023	0.107***
	(0.0221)	(0.0272)	(0.0047)
	N = 4,527	N = 2,330	N = 359,021
Males	-0.104***	-0.096***	0.111***
	(0.0215)	(0.0278)	(0.0044)
	N = 4,187	N = 2,176	N = 353,552
First Repetition	-0.068***	-0.048**	0.152***
	(0.0189)	(0.0238)	(0.0041)
	N = 6,630	N = 3,338	N = 638,582

Note: In the case of FRD-Matching there are $N_{bt}/2$ students with their second-lowest score equal to 3.8 and $N_{bt}/2$ students with that score equal to 4.1. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We define dropping out as a situation in which the student does not attend school in the years corresponding to 11th and 12th grade. For instance, we say that a student who was attending 4th grade in 2007 dropped out of school if she did not attend school in 2014 and 2015. Table 5 shows the effects of grade retention on dropping out of school. As can be seen, repeating a grade in 2007 increases the probability of dropping out of school by 1.2 to 3.2 pp (Column (1)).[32] According to the school dropout measure used in this paper, 6.3% of the students dropped out after 2007. Thus, our finding represents an increase ranging from 19 to 51%. We find no effects for those students who repeated for the first time in 2007.

[32]These results are along the same lines as the findings of Manacorda (2012) and Jacob and Lefgren (2009).

Table 5: Effect of grade retention on dropping out

Sample	(1) Donut-Hole FRD	(2) FRD-Matching	(3) OLS
All	0.015*	0.016*	0.059***
	(0.0079)	(0.0096)	(0.0023)
	N = 9,681	N = 5,130	N = 705,261
Low SES	0.030**	0.032*	0.081***
	(0.0142)	(0.0174)	(0.0035)
	N = 4,527	N = 2,330	N = 359,021
Males	0.010	0.019	0.057***
	(0.0118)	(0.0141)	(0.0029)
	N = 4,187	N = 2,176	N = 353,552
First Repetition	0.006	0.003	0.039***
	(0.0076)	(0.0093)	(0.0022)
	N = 6,630	N = 3,338	N = 638,582

Note: In the case of FRD-Matching there are $N_{bt}/2$ students with their second-lowest score equal to 3.8 and $N_{bt}/2$ students with that score equal to 4.1. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In addition to the discussion on the magnitudes of the effects, there are several aspects of these results that are important to highlight. First, as in the crime estimation, the OLS estimation delivers upward-biased effects for the reasons discussed above. Second, the effect of grade retention on dropping out suggests a relevant mechanism through which not being promoted to the next grade in school may affect juvenile crime: repeating a grade has an impact on dropping out, and dropping out has an impact on crime. Third, if we assume that grade retention in higher grades also has an impact on juvenile crime, then the results of Table 4 suggest that we are finding a lower bound of the effect, since those who did not repeat in 2007 (who are *non-treated* in our estimation) had a higher probability of repeating a grade in the future, which also has an impact on crime.

5.3 Robustness Analysis

To examine the robustness of our results, we perform two empirical exercises. In the first one, we re-estimate the "donut-hole" RD and the RD-matching, but now we restrict the sample to the students whose final status at school is consistent with the retention rule. In practice, this is equivalent to re-estimating Columns (1) and (2) of Table 3, but now imposing a sharp RD design. To be clear, we re-estimate the "donut-hole" RD specification in two steps: (1) among all students whose second-lowest score is 3.8 or 4.1, we keep only those students whose final status at school is consistent with the retention rule: below the threshold, we drop the students who pass the grade, and, above the threshold, we drop the students who repeat the grade; (2) given this sample, we estimate a standard sharp design RD, by regressing the following equation:[33]

$$Y_i = \alpha_c^y + \tau W_i + \alpha_x^y X_i + \varepsilon_i^y. \tag{4}$$

Along the same lines, we re-estimate the RD-matching in two steps: (1) among all students whose second-lowest score is 3.8 or 4.1, as before, we keep only those students whose final status at school is consistent with the retention rule; (2) given the matched sample, the LATE parameter (τ) is estimated by regressing Equation 4.[34]

The second empirical exercise to review the robustness of our results is to implement a placebo test. In this case, we replicate the "donut-hole" RD and the RD-matching estimations, but now we compare only students scoring below and above the threshold who did not repeat the grade.[35] For instance, in the case of the "donut-hole" RD, we proceed with the following two steps: (1) among all students whose second-lowest score is 3.8 or 4.1, we keep only those students whose final status at school is *pass the grade*; (2) given this sample, we estimate $E[Y_i|Z_i = 0, W_i = 0, X_i]$ and $E[Y_i|Z_i = 1, W_i = 0, X_i]$ by regressing Equation 4. If our empirical approach is valid, we will find that $E[Y_i|Z_i = 0, W_i = 0, X_i] = E[Y_i|Z_i = 1, W_i = 0, X_i]$.[36]

The results of these empirical exercises are presented in Table 6. In short, the figures in the first two columns, coming from the re-estimation of the "donut-hole" RD and the RD-matching (but now imposing a sharp design), are remarkably similar to the results presented in Table 3. More importantly, the results of the placebo exercises (Columns (3) and (4)) show no statistical significance. Regarding the magnitudes, although none of the estimates are statistically significant, Column (4) shows better (closer to zero) point

[33] Given Step (1), this sample does not require a 2SLS estimator. Indeed, it is a sharp design RD.

[34] We are using the matched sample described in Table 2, as opposed to finding a new matched sample, given the smaller number of students scoring below the threshold. These samples would be different due to the fact that design matching involves constraining the measures of imbalance to be less than or equal to certain tolerances.

[35] In principle, we could do the same by comparing those who are below and above the threshold and repeated the grade. However, we do not have a sufficiently large sample size to do that.

[36] See Imbens and Rubin (2015).

estimates compared to Column (3), i.e., the RD-matching seems more robust than the "donut-hole" RD. Overall, the placebo results are important because they reinforce our belief that the numbers presented in Columns (1) and (2) of Table 3 can be interpreted as (local) causal effects.[37]

Table 6: Effect of grade retention on juvenile crime (sharp design and placebo)

	Sharp Design		Placebo	
	(1)	(2)	(3)	(4)
Sample	Donut-Hole RD	RD-Matching	Donut-Hole RD	RD-Matching
All	0.015***	0.018***	0.002	-0.001
	(0.0058)	(0.0067)	(0.0082)	(0.0081)
	$N = 8{,}694$	$N = 4{,}421$	$N = 7{,}054$	$N = 3{,}236$
Low SES	0.018*	0.033***	0.015	0.014
	(0.0102)	(0.0118)	(0.0165)	(0.0166)
	$N = 4{,}096$	$N = 2{,}040$	$N = 3{,}259$	$N = 1{,}435$
Males	0.021**	0.026**	0.022	0.005
	(0.0101)	(0.0123)	(0.0158)	(0.0168)
	$N = 3{,}783$	$N = 1{,}910$	$N = 2{,}891$	$N = 1{,}340$
First repetition	0.013	0.016**	-0.003	-0.004
	(0.0067)	(0.0080)	(0.0097)	(0.0092)
	$N = 5{,}934$	$N = 2{,}878$	$N = 4{,}782$	$N = 2{,}103$

Note: In the case of FRD-Matching there are $N_{bt}/2$ students with their second lowest score equal to 3.8 and $N_{bt}/2$ students with that score equal to 4.1. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Finally, in Appendix C.1, we present the robustness analysis for dropping out and grade retention after 2007 (Tables 9 and 10). As in the case of crime, in the placebo test all parameters are statistically insignificant in the case of future grade retention and dropping out. These results confirm the soundness of our empirical strategy to find causal estimates.

[37] That said, it is important to consider that the robustness of our approaches critically depends on the level of the initial imbalance in observables. For instance, we ran the same placebo approaches, but compared students who scored 4.1 to students who scored 4.4, and we found differences that were statistically significant. However, the differences in observables between these two groups (scoring 4.1 and 4.4) were much higher than the differences between the groups that were used in our estimation.

6 Results Exploiting Retention Rule II

In this section, we present our findings on the impact of grade retention on juvenile crime, dropping out, and future grade retention. These results are based on an empirical strategy that implements the fuzzy RD developed by Calonico, Cattaneo, and Titiunik (described in Subsection 4.2), which exploits the grade retention rule that specifies that the student has to repeat the grade when she scores below 4 on one subject (≤ 3.9) and has an average score across all subjects lower than or equal to 4.5. To support our interpretation as a causal effect, we also show the results of our estimation procedure when, instead of having crime, dropping out, or future grade retention as our dependent variable, we consider our control variables as dependent variables.

As in the previous section, we have run our model for different sample groups, namely, with and without previous grade retention, low SES, and only men. For each of these sample groups (and their combinations), we have a plot that shows the results for three different bandwidths (0.1, 0.15, and 0.2). Therefore, we focus our attention on the sample group that presents more robust results. In particular, we focus on the estimation that only considers men who did not repeat before 2007. This is the most robust estimation in the sense that, for this group, we do not find differences in the control variables below and above the threshold.[38]

6.1 Impacts on Crime, Future Grade Retention, and Dropping Out

In Figure 8, we present our estimations for the impact of grade retention on juvenile crime, for males who did not repeat before 2007. The effect goes from 13 pp to 6 pp, and in two out of three cases is significantly different from zero. The confidence intervals (CI) are calculated using the conventional approach. In Appendix D, we show the same plots but with the robust estimations of the CI developed by Calonico, Cattaneo, and Titiunik (2014b).[39]

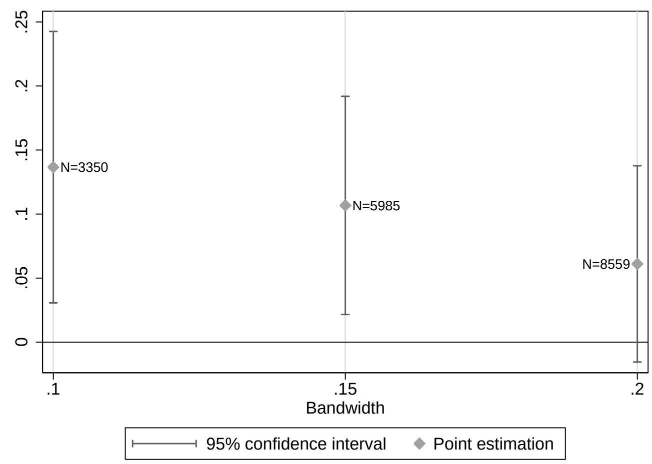
In Figures 9 and 10, we show the effects of grade retention on future grade retentions

[38]Most of the results for the other groups are presented in the appendix; the rest can be shared upon request.

[39]When the CI are calculated using the robust formula, the significance is lost in some cases; however, the p-values, although higher, are rather similar.

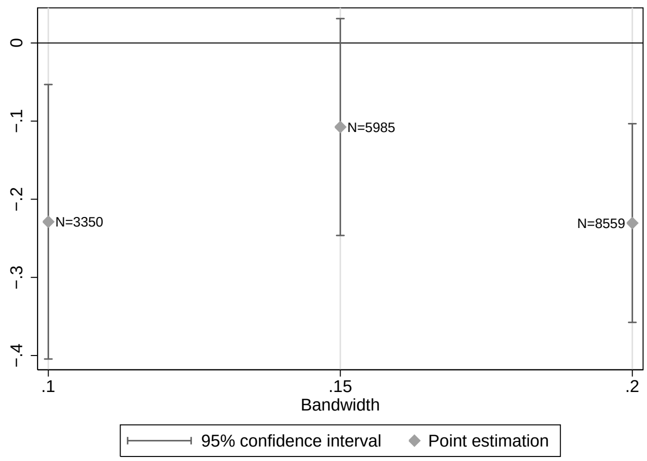
and on students dropping out. In the case of future grade retention, we do observe that grade retention decreases the probability of grade retention in a range of 10 pp to 23 pp, even though, for one bandwidth, the effect is not significantly different from zero, for a small margin. However, we do not observe the same strong evidence in the case of the effect on dropping out.

Figure 8: Effect of grade retention on juvenile crime (Men who did not repeat before 2007)



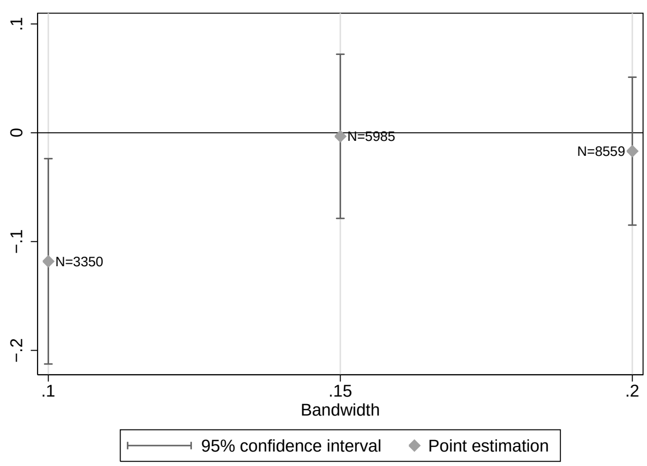
Note: The plot shows the effect of grade retention on crime using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 9: Effect of grade retention on future grade retention
(Men who did not repeat before 2007)



Note: The plot shows the effect of grade retention on future grade retention using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 10: Effect of grade retention on droppping out
(Men who did not repeat before 2007)



Note: The plot shows the effect of grade retention on dropping out using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented, together with the number of individuals considered in the estimation.

Overall, the results go in the same direction as the estimations using grade retention Rule I, with the exception of students dropping out, where the existence of a causal effect is less clear. However, the current estimations of the effects on crime and on future grade retention present larger magnitudes and they are less precise in terms of variance. Due to the latter, we prefer to base our analysis and conclusions on the magnitudes of the first empirical strategy (exploiting grade retention Rule I). That said, we should emphasize that the differences in magnitudes between these results and the results based on Rule I are smaller if we focus our attention on percentage change instead of the change in percentage points. This is because those students who only repeat because of Rule I have better outcomes than those who only repeat because of Rule II. Indeed, while the crime rate of the first group is 4.5%, this figure is 7.3% for the second group.

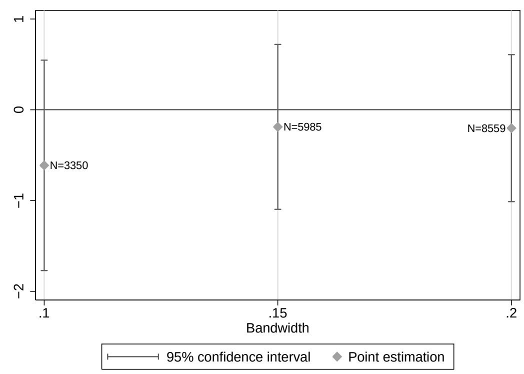
6.2 Robustness Analysis

To the extent that testing differences in observables around the threshold is an indirect way to test differences in unobservable variables (Lee and Lemieux (2010)), we study the robustness of our empirical approach by running the same FRD model but considering the covariates as dependent variables. To have a more demanding placebo test, in these cases the estimation is run without the other covariates as control variables. In order to support our claim about causality, we need to show that our model does not deliver a statistically significant relationship between grade retention and our covariates, namely, father's and mother's education, math and Spanish test scores, belonging to the low socioeconomic group, and school attendance in 2006.

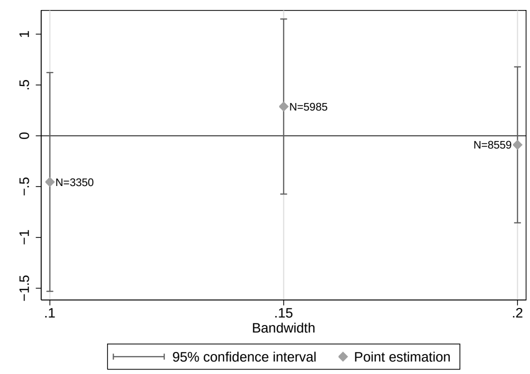
The results of this placebo test are presented in Figure 11. Fortunately, in most cases the estimated parameter is not significantly different from zero, with the exception of one bandwidth for each test score. The latter is not a surprise because there should be a close and linear relationship between scores at school and standardized test scores. What is a surprise, however, is that the stronger relationship shows up at the smaller bandwidth. We address this issues below. Something to highlight is the fact that, in almost all cases, the point estimation does not have a monotonic relationship with respect to the bandwidth. This reinforces the idea that our empirical approach is not finding a relationship between grade retention and the covariates.

Figure 11: Placebo estimations with control as dependent variables (Men who did not repeat before 2007)

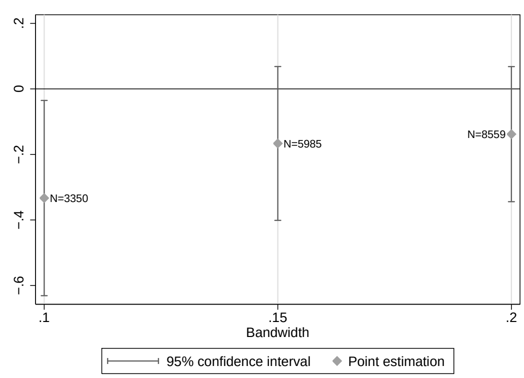
(a) Father's Education



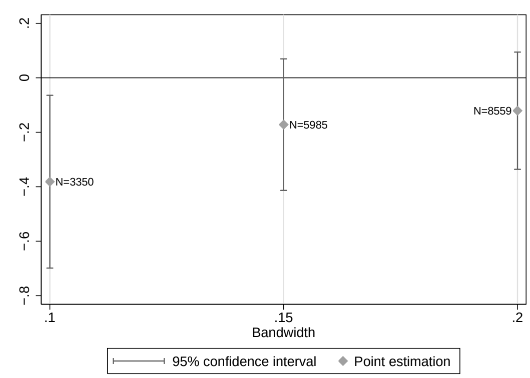
(b) Mother's Education



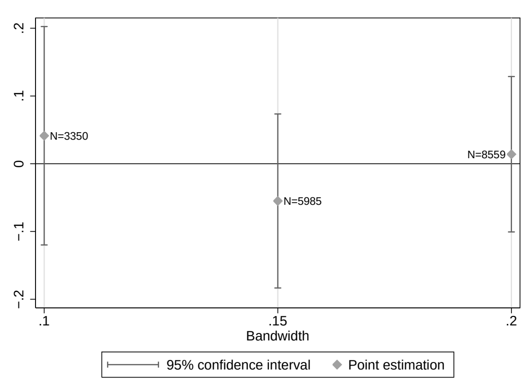
(c) Math Test Scores



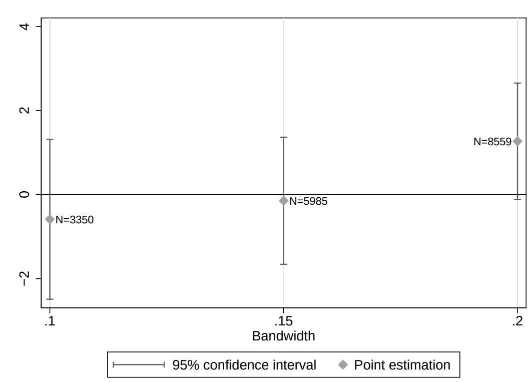
(d) Spanish Test Scores



(e) Fraction of Low SES



(f) 2006 School Attendance



Note: The plots show the estimation of the FRD procedure but using all the control variables as the dependent variable. The estimation is run without the other control variables as covariates.

It is not easy to explain why, in the case of the smallest bandwidth, the effects of grade retention on all of the outcomes are higher and the placebo test delivers some statistically significant results. If, as expected, the unobserved variables are more influential as we move away from the threshold of the running variable and the existence of these variables biases our estimation upward, we should observe the opposite tendency across the bandwidths. One way to rationalize these results is thinking about manipulation and sorting around the threshold. However, as we show in Figure 7, we do not find evidence supporting that. Although we do not have a clear explanation for this phenomenon, given these results, we think it is fair to say that our estimations using bandwidths 0.15 and 0.2 are robust and reliable.

In Appendix D, we show the same plots but for different groups and specifications. Indeed, we present our estimations for different sample groups, namely, men and women who did not repeat before 2007 and low SES males who did not repeat before 2007. Moreover, we also present the estimation for all these groups but calculate the CI using the robust formula. Overall, the results are similar in terms of signs and magnitudes; however, the statistical significance is less robust. That said, it is remarkable that the placebo tests are also consistent with the claim of causality.

7 Mechanisms

In this section, we present two exercises to shed some light on what may explain the impact of grade retention on juvenile crime. First, we explore how grade retention increases the probability of the occurrence of negative trajectories after 2007. Second, we discuss the relevance of school switching in reinforcing the negative effect of grade retention on a student's future. We do so by focusing our analysis on retention Rule I.

7.1 Interaction Between Crime and Dropping Out of School

The causal effect of grade retention on juvenile crime, documented in the previous sections, could have operated through different mechanisms. For instance, grade retention could have had its effect only indirectly, through the effect of repetition on dropping out, and the subsequent effect of dropping out on crime. To explore what happens after not being promoted to the next grade and how this event affects students' trajectories, we run a multinomial logit with four trajectories as possible outcomes. The results are presented in Table 7. The possible trajectories after 2007 are: attending school in all periods of our sample, without committing a crime during those years (Column (1)); dropping out of school in a year t (after 2007), without committing a crime in a year $t + a$, with $a > 0$ (Column (2)); dropping out of school in a year t (after 2007) and committing a crime in a year $t + a$ (Column (3)); and committing a crime in a year t (after 2007) and dropping out of school after that, simultaneously, or never (Column (4)).

Given the non-linearity of this model, we avoid the use of the score in the second-lowest subject as the instrument in a fuzzy RD design, and instead we run a multinomial logit as if we had a sharp RD design. We do so by following the approach described in Section 5.3, namely, among all students whose second-lowest score is 3.8 or 4.1, we keep only those students whose final status at school is consistent with the retention rule. Given this sample, we implement the design matching to optimally find a match for each student scoring 3.8 among those who score 4.1. Therefore, given this matched sample, the variable of interest is a dummy that takes the value one if the student was not promoted to the next grade in 2007, and scored 3.8 in her second-lowest score, and the value zero if the student was promoted to the next grade in 2007, and scored 4.1 in the second-lowest score.[40] Besides this variable, the model includes the same controls as the models in the previous section.[41]

Table 7 shows the marginal effects of this multinomial logit. It can be observed that grade retention increases the probability of committing a crime after 2007 before dropping out of school (if the student dropped out) by 1 pp. Overall, the results show that grade retention increases the probabilities of "bad trajectories" (involving either crime or dropping out of school) and that the effect of grade retention on crime is not only through its effects on dropping out. In fact, the effect on delinquency occurring before (or simultaneously to) dropping out is more relevant than the effect on crime that occurs after dropping out.

[40]Even though this approach does not allow for a discussion on causality, because the right approach would be a FRD, the analysis of the effects of grade retention on crime, dropping out, and grade retention after 2007 (presented in the previous sections) shows that this *fake* sharp design RD delivers rather similar results to the fuzzy RD estimators.

[41]These are: gender, father's education, mother's education, math and language SIMCE, family income, attendance in 2006, and previous grade retentions.

Table 7: Effect of grade retention on the probability of different trajectories

	Always at School, no crime	Dropout, no crime	First Dropout, then crime	First crime, then (if so), or simultaneously, dropout
Grade Retention, in 2007	-0.0435*** (0.0118)	0.0282*** (0.0108)	0.0053*** (0.0020)	0.0100** (0.0048)
			Observations = 4,961	Pseudo R2 = 0.11

Note: This is a multinomial logit model with a dependent variable with four categories. The model includes the following controls: gender, father's education, mother's education, math and language SIMCE, family income, attendance in 2006, previous grade retentions. The table presents the marginal effects. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7.2 Switching Schools After Grade Retention

As documented in Hanushek, Kain, and Rivkin (2004), students may experience a substantial pedagogical cost as a result of switching schools.[42] If so, it is possible that part of the effect of grade retention on crime is due to the fact that not being promoted to the next grade may increase the probability of switching schools.

To explore the relevance of this mechanism, we run an OLS regression among all the students who repeated in 2007 and scored between 3.0 and 4.5 on the second-lowest score.[43] We study the correlation between switching schools (between 2007 and 2008) and juvenile crime, controlling for the same variables as in the models in the previous sections, and including school-grade fixed effects. Considering that 8th grade presents a higher rate of students changing schools, we show the results of this model including and excluding this grade.

Table 8 shows that, for those students who repeated a grade, switching schools increases the probability of crime by 2.2 pp, a result that does not change when 8th grade is excluded. Thus, in line with the literature, grade retention could be particularly negative for a student's future when directly followed by a change in school.

[42]In the case of Chile, Grau, Hojman, and Mizala (2016) find that a school closing, which always implies switching schools, increases the probability of grade retention and dropping out.

[43]We use this set of students to have both a large enough sample size and a group which is similar to the sample used in Section 6.1.

Table 8: CRIME AND SWITCHING SCHOOLS

Variables	4th to 8th grade		Excluding 8th grade	
Switching school	0.0220***	(0.0068)	0.0224***	(0.0080)
Attendance 2006	-0.0012**	(0.0005)	-0.0014**	(0.0006)
Mother Education	-0.0018	(0.0012)	-0.0022*	(0.0013)
Father Education	-0.0019*	(0.0011)	-0.0023*	(0.0012)
Family Income	-0.0000	(0.0000)	-0.0000	(0.0000)
Male	0.0546***	(0.0061)	0.0591***	(0.0070)
Math SIMCE	0.0008	(0.0047)	0.0009	(0.0057)
Language SIMCE	-0.0011	(0.0044)	-0.0031	(0.0053)
Constant	0.1794***	(0.0476)	0.2066***	(0.0554)
N		18,946		15,171
R2		0.086		0.093

Note: These two estimations include school-grade fixed effects. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

8 Conclusion

In this paper, we present strong and robust evidence of a causal relationship between grade retention and juvenile crime in the case of Chile. To do so, we exploit two discontinuities in the probability of not being promoted to the next grade, which are the results of the grade retention rules of the Chilean educational system. In the case of the retention rule that focuses on the second-lowest subject score, due to clear evidence about local manipulation on the forcing variable, we depart from standard RD methods. First, we follow Barreca, Guldi, Lindo, and Waddell (2011) to implement a *donut-hole* FRD, where, after removing observations in the immediate vicinity of the threshold for grade repetition, we run a standard FRD. Second, we extend the method developed by Keele, Titiunik, and Zubizarreta (2015) to combine matching with a fuzzy regression discontinuity design. In the case of the retention rule that focuses on the average grade across all subjects, we follow a more standard FRD approach, implementing the methodology developed by Calonico, Cattaneo, and Titiunik (2014b).

This paper makes three main contributions. First, together with a recent paper (Depew and Eren (2015)), it is the first one that estimates a causal effect of grade retention on juvenile crime and it is the first such evidence for a developing country. This causal evidence calls into question the appropriateness of grade repetition as a public policy, a concern that is even more relevant in the context of Chile, a developing country with a high rate of grade retention.[44] That said, the interpretation of our findings should consider that we are not taking into account other aspects of this policy; for example, the threat of retention could serve as an incentive for all students to exert more effort. Second, by extending the method developed by Keele, Titiunik, and Zubizarreta (2015) to the *fuzzy RD* case, we present an empirical approach that can be useful in many other contexts in which there is some evidence of manipulation in the forcing variable. Third, the paper sheds light on the mechanisms that could explain the impact of grade retention on juvenile crime. From this analysis, it is possible to infer relevant insights for public policy debates.

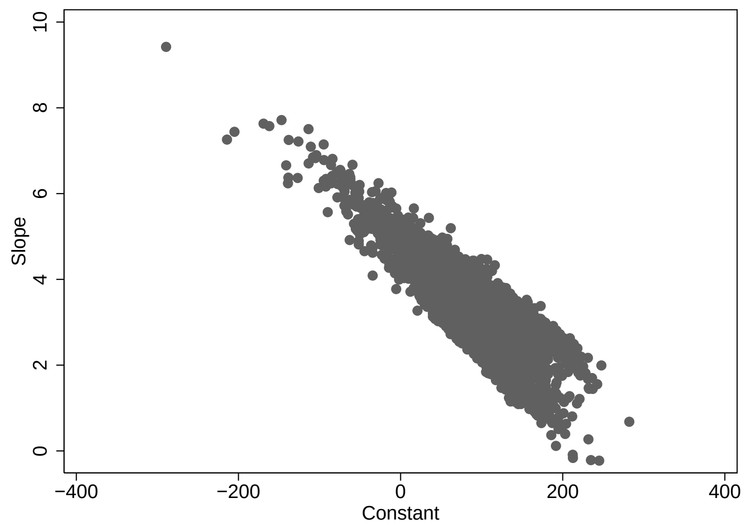
Because the impact of grade retention on crime does not only operate through its effect on dropping out of school, it is not possible to argue that, rather than concern ourselves with grade retention, we should only start to worry when the student who has been held back drops out. The evidence presented in this paper implies that policymakers should be concerned about high levels of grade retention on its own merit. However, the relevance of the interaction between grade retention and school switching in the determination of juvenile crime gives important clues about a possible avenue to attenuate the negative effects of grade retention, in case policymakers decide to continue supporting this practice. In particular, because repetition is a negative response from the education system, it may discourage students' commitment to their educational process. Thus, it is essential to design policies that will counteract this negative effect, breaking the connection between grade repetition and other causes of dropping out and juvenile crime. This is particularly important for those students who have repeated more than once.

[44]In fact, in our sample, 13.1% of the students repeated at least one grade between 1st and 8th grade.

Appendix

A Grading Standards

Figure 12: Heterogenous Grading Standards Across Schools



B Design Matching

Let Z_1 denote the group of students whose second-lowest score in 2007 is below the threshold (*i.e.,* equal to 3.8), and let Z_0 denote the group of students whose second-lowest score is above the threshold (equal to 4.1).[45] Let j_1 index the members of group Z_1 and j_0 index the members of group Z_0. Define d_{j_1,j_0} as the covariate distances (in math and language standardized test scores, parents' education, previous repetitions, attendance during the previous year, per capita income, and gender) between unit j_1 and j_0. To enforce specific forms of covariate balance, define $e \in \varepsilon$ as the index of the covariate (school and grade identification) for which it is needed to match exactly, and $b_e \in B_e$ as the categories that covariate e takes, so that $x_{j_1;e}$ is the value of nominal covariate e for unit j_1 with $x_{j_1;e} \in B_e$. Finally, let $m \in M$ be the index of covariates for which it is desired to balance their means, in this case: math and language standardized

[45] We follow the notation and the description from Keele, Titiunik, and Zubizarreta (2015)

test scores, parents' education, previous retentions, attendance during the previous year, per capita income, and gender. So that $x_{j_1;m}$ is the value of covariate m for unit j_1, and $x_{j_0;m}$ is the value of covariate m for j_0.

To solve the problem optimally, the following decision variables are introduced:

$$a_{j_1;j_0} = \begin{cases} 1 & \text{if unit } j_1 \text{ is matched to unit } j_0 \\ \\ 0 & \text{otherwise,} \end{cases}$$

Then, for a given scalar λ, the objective function to minimize is equal to:[46]

$$\sum_{j_1 \in Z_1, j_0 \in Z_0} d_{j_1,j_0} a_{j_1,j_0} - \lambda \sum_{j_1 \in Z_1, j_0 \in Z_0} a_{j_1,j_0}, \tag{5}$$

subject to pair matching and covariate balancing constraints. Under this penalized match, if distance can be minimized it will be, and if it cannot be minimized in every case, it will be minimized as often as possible. In particular, the pair matching constraints require each treated and control subject to be matched at most once,

$$\sum_{j_0 \in Z_0} a_{j_1,j_0} \leq 1, \quad \forall j_1 \in Z_1, \tag{6}$$

$$\sum_{j_1 \in Z_1} a_{j_1,j_0} \leq 1, \quad \forall j_0 \in Z_0. \tag{7}$$

This implies that it matches without replacement. The covariate balancing constraints are defined as follows

$$\sum_{j_1 \in Z_1, j_0 \in Z_0} \left| \mathbb{1}_{\{x_{j_1;e} = b_e\}} x_{j_1;e} - \mathbb{1}_{\{x_{j_0;e} = b_e\}} x_{j_0;e} \right| a_{j_1,j_0} = 0, \quad \forall e \in \varepsilon, \tag{8}$$

$$\left| \sum_{j_1 \in Z_1, j_0 \in Z_0} a_{j_1;j_0} (x_{j_1;m} - x_{j_0;m}) \right| \leq \varepsilon_m \sum_{j_1 \in Z_1, j_0 \in Z_0} a_{j_1;j_0}, \quad \forall m \in M, \tag{9}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function.

These constraints enforce exact matching and mean balance, respectively. More precisely, (8) requires exact matching by matching each subject in Z_1 to a subject in Z_0

[46]We solve this optimization problem, by implementing the R package described in Zubizarreta and Kilcioglu (2016).

in the same school and grade; and (9) forces the differences in means after matching to be less than or equal to $\varepsilon_m = 0.03$ standard deviations apart for all $m \in M$, with $M =$ standardized scores in language and math, parents' education, previous retentions, attendance during the previous year, an income variable and gender.

The "Designmatch" incorporates optimal subset matching into the integer programming framework in the objective function (5) via the λ parameter. The first term in (5) is the total sum of mahalanobis distances between matched pairs, and the second term is the total number of matched pairs. Therefore, λ emphasizes the total number of matched pairs in relation to the total sum of distances and, according to (5), it is preferable to match additional pairs if on average they are at shorter distances than λ. In our application, we choose λ to be equal to the median mahalanobis distance between j_1 and j_0 subjects so, according to (5), it is preferable to match additional pairs if on average they are at a shorter distance than the typical distance (as measured by the median).[47] Subject to the pair matching constraints (6) and (7) and the covariate balancing constraints (8) and (9), this form of penalized optimization addresses the lack of common support problem in the distribution of observed covariates of subject in Z_1 and Z_0.

Due to this penalty, the Designmatch keeps the largest number of matched pairs for which distance is minimized and the balance constraints are satisfied. This implies that as we alter the distances or the balance constraints, the number of j_1 and j_0 subjects retained changes. In particular, for stricter constraints we tend to retain a smaller number of subjects.

[47] λ can be thought of as a parametrization of the trade-off between bias and variance: a higher value of it would imply a bigger sample size, but more differences between treated and controls.

C Robustness Analysis for the First Empirical Strategy (Rule I)

C.1 Other Outcomes

Table 9: Effect of grade retention on future grade retention (sharp design and placebo)

	Sharp Design		Placebo	
	(1)	(2)	(3)	(4)
Sample	Donut-Hole RD	RD-Matching	Donut-Hole RD	RD-Matching
All	-0.087***	-0.060***	0.006	-0.005
	(0.0129)	(0.0151)	(0.0205)	(0.0212)
	$N = 8{,}694$	$N = 4{,}421$	$N = 7{,}054$	$N = 3{,}236$
Low SES	-0.058***	-0.029	0.022	0.010
	(0.0190)	(0.0221)	(0.0321)	(0.0326)
	$N = 4{,}096$	$N = 2{,}040$	$N = 3{,}259$	$N = 1{,}435$
Males	-0.120***	-0.101***	0.048	0.007
	(0.0191)	(0.0227)	(0.0346)	(0.0334)
	$N = 3{,}783$	$N = 1{,}910$	$N = 2{,}891$	$N = 1{,}340$
First repetition	-0.078***	-0.045**	0.018	-0.020
	(0.0155)	(0.0187)	(0.0250)	(0.0267)
	$N = 5{,}934$	$N = 2{,}878$	$N = 4{,}782$	$N = 2{,}103$

Note: In the case of FRD-Matching there are $N_{bt}/2$ students with their second lowest score equal to 3.8 and $N_{bt}/2$ students with that score equal to 4.1. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Effect of grade retention on dropping out (sharp design and placebo)

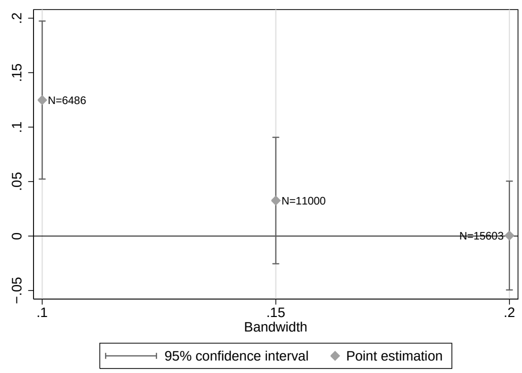
	Sharp Design		Placebo	
	(1)	(2)	(3)	(4)
Sample	Donut-Hole RD	RD-Matching	Donut-Hole RD	RD-Matching
All	0.017***	0.020**	-0.009	-0.013
	(0.0066)	(0.0079)	(0.0095)	(0.0092)
	$N = 8{,}694$	$N = 4{,}421$	$N = 7{,}054$	$N = 3{,}236$
Low SES	0.033***	0.040***	-0.016	-0.025
	(0.0121)	(0.0146)	(0.0185)	(0.0181)
	$N = 4{,}096$	$N = 2{,}040$	$N = 3{,}259$	$N = 1{,}435$
Males	0.010	0.020*	0.001	-0.006
	(0.0100)	(0.0117)	(0.0169)	(0.0161)
	$N = 3{,}783$	$N = 1{,}910$	$N = 2{,}891$	$N = 1{,}340$
First repetition	0.006	0.008	0.001	-0.014*
	(0.0065)	(0.0077)	(0.0085)	(0.0081)
	$N = 5{,}934$	$N = 2{,}878$	$N = 4{,}782$	$N = 2{,}103$

Note: In the case of FRD-Matching there are $N_{bt}/2$ students with their second lowest score equal to 3.8 and $N_{bt}/2$ students with that score equal to 4.1. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

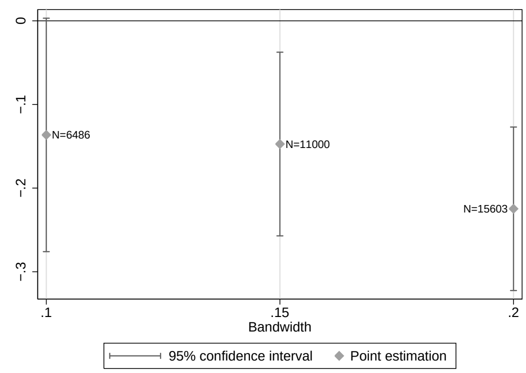
D Results for Other Sample Groups, Second Empirical Strategy (Rule II)

Figure 13: Effect of grade retention on different outcomes (Men and women who did not repeat before 2007)

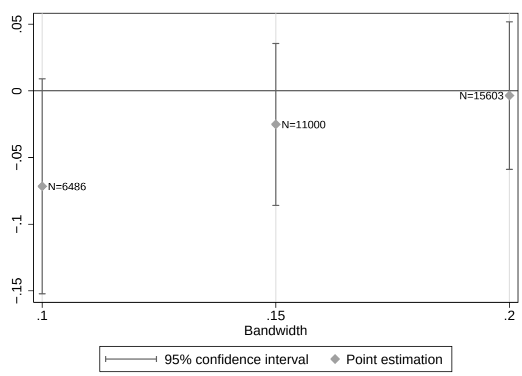
(a) Juvenile Crime



(b) Future Grade Retention



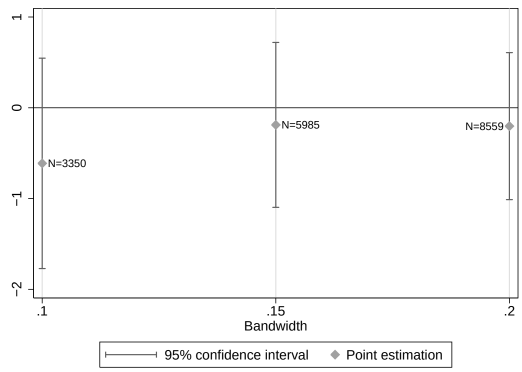
(c) Dropping Out



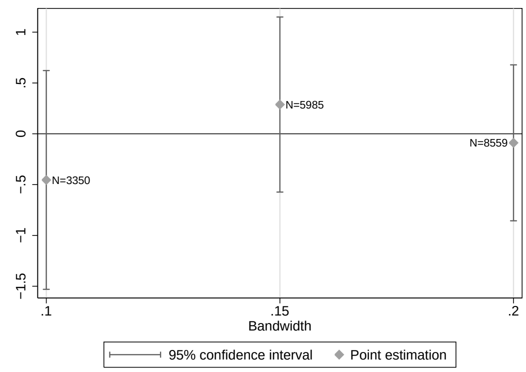
Note: The plots show the effects of grade retention on different outcomes using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 14: Placebo estimations with control as dependent variables (Men and women who did not repeat before 2007)

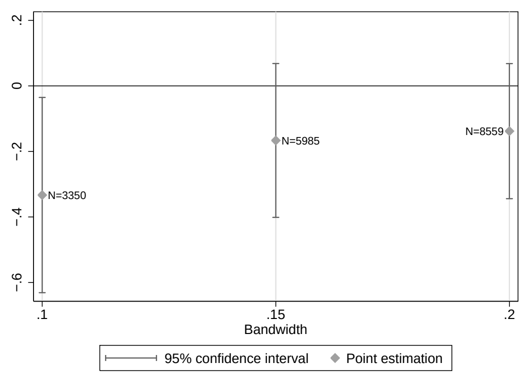
(a) Father's Education



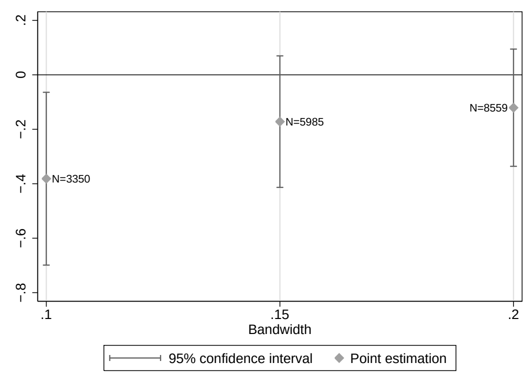
(b) Mother's Education



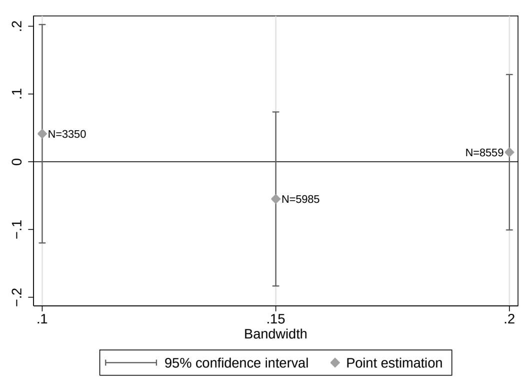
(c) Math Test Scores



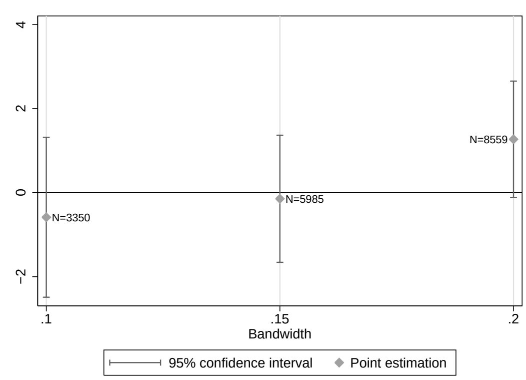
(d) Spanish Test Scores



(e) Fraction of Low SES



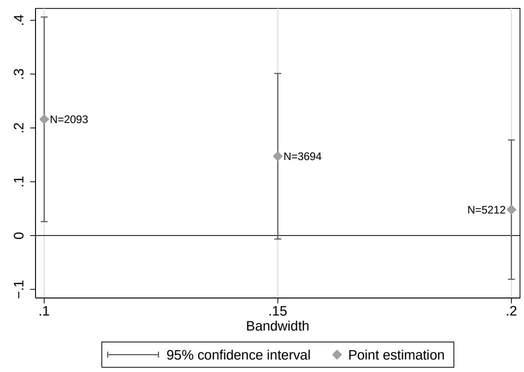
(f) 2006 School Attendance



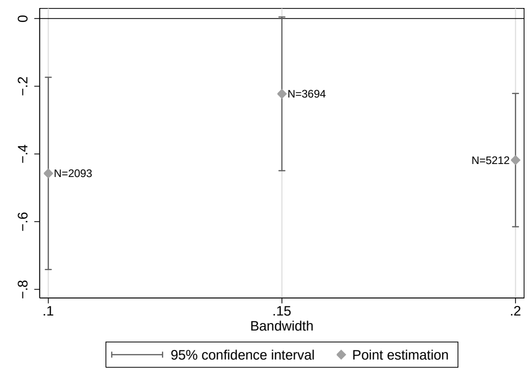
Note: The plots show the estimation of the FRD procedure but using all the control variables as the dependent variable. The estimation is run without the other control variables as covariates.

Figure 15: Effect of grade retention on different outcomes
(Low SES males who did not repeat before 2007)

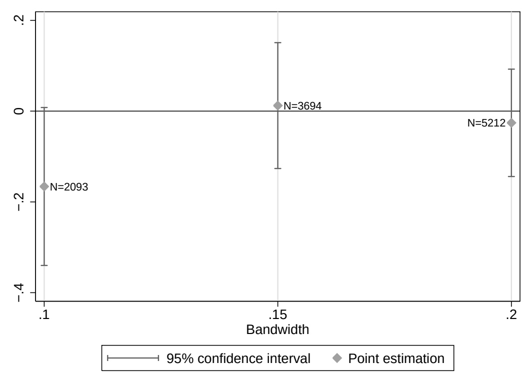
(a) Juvenile Crime



(b) Future Grade Retention



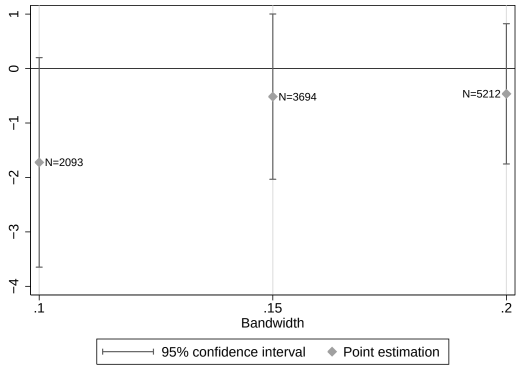
(c) Dropping Out



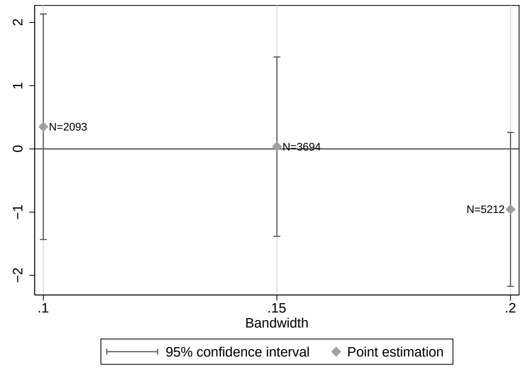
Note: The plots show the effects of grade retention on different outcomes using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 16: Placebo estimations with control as dependent variables (Low SES males who did not repeat before 2007)

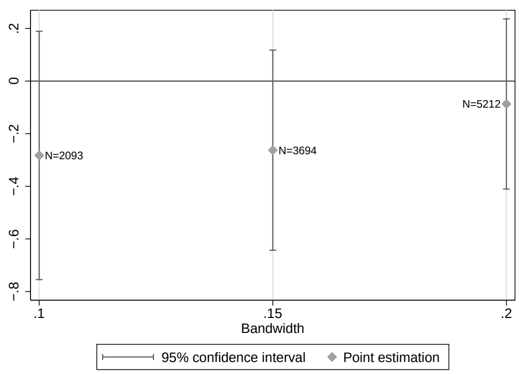
(a) Father's Education



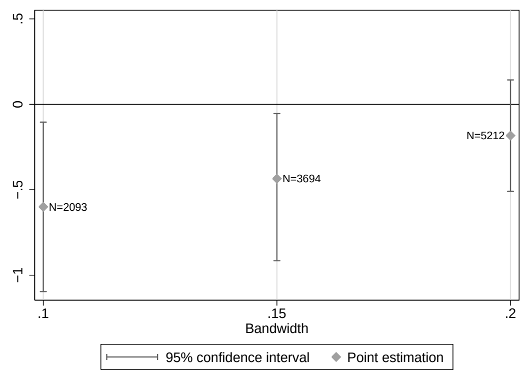
(b) Mother's Education



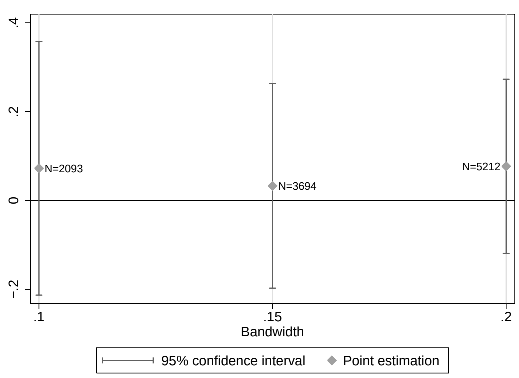
(c) Math Test Scores



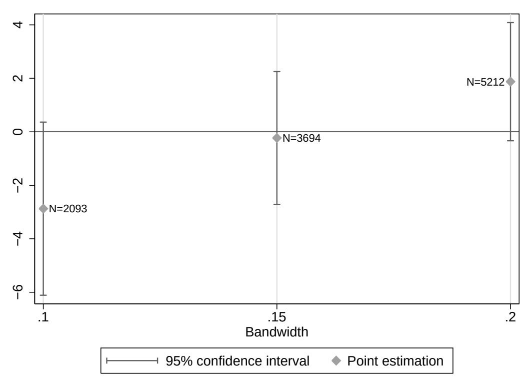
(d) Spanish Test Scores



(e) Fraction of Low SES



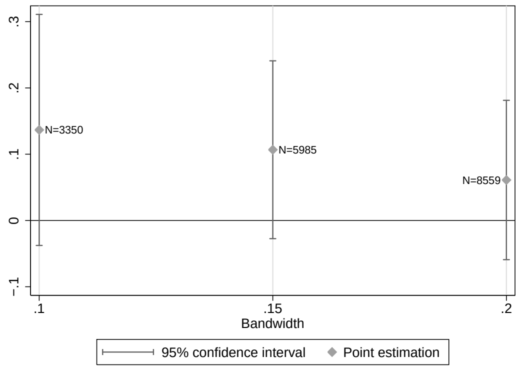
(f) 2006 School Attendance



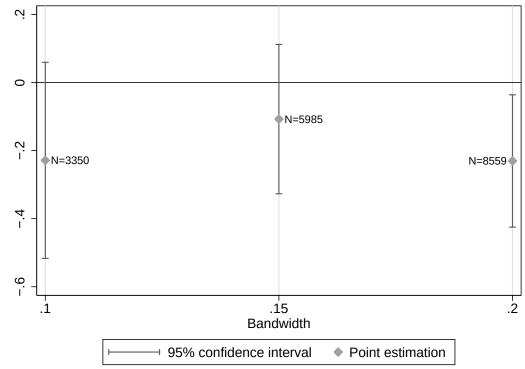
Note: The plots show the estimation of the FRD procedure but using all the control variables as the dependent variable. The estimation is run without the other control variables as covariates.

Figure 17: Effect of grade retention on different outcomes, robust estimation for C.I
(Men who did not repeat before 2007)

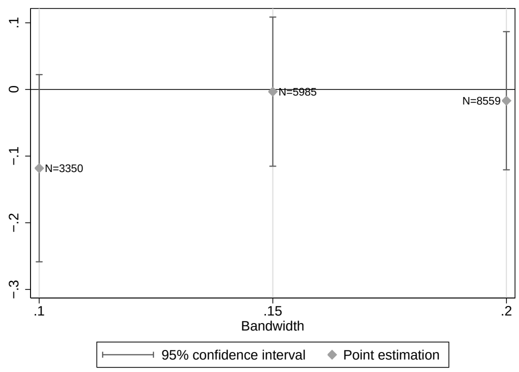
(a) Juvenile Crime



(b) Future Grade Retention



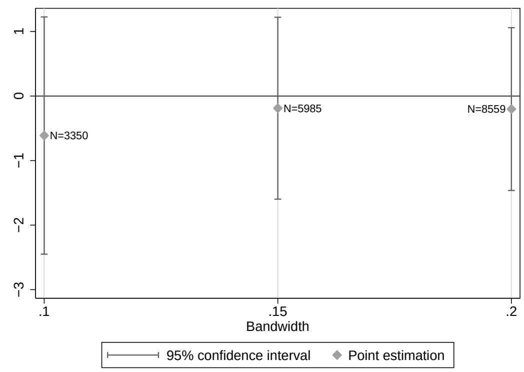
(c) Dropping Out



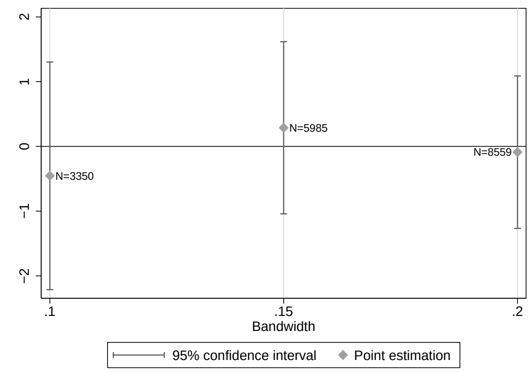
Note: The plot shows the effect of grade retention on crime using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 18: Placebo estimations with control as dependent variables, robust estimation for C.I
(Men who did not repeat before 2007)

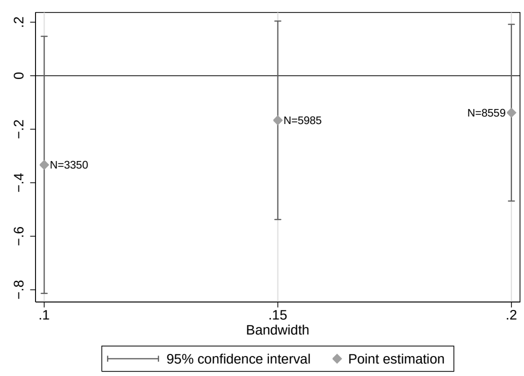
(a) Father's Education



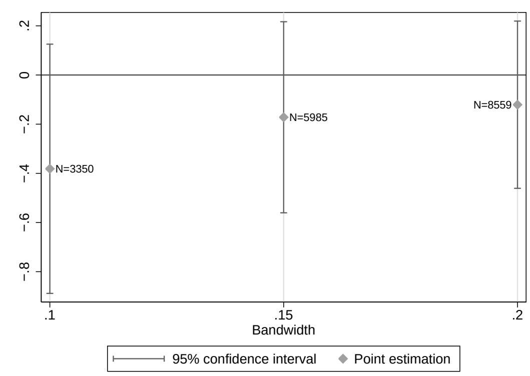
(b) Mother's Education



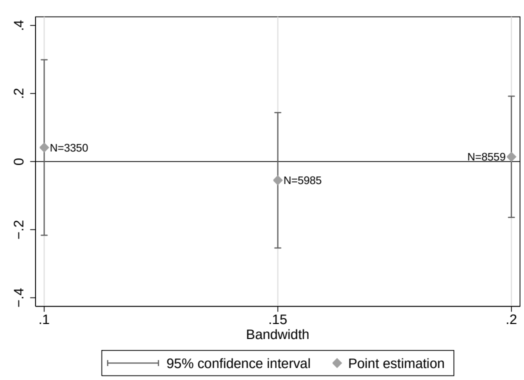
(c) Math Test Scores



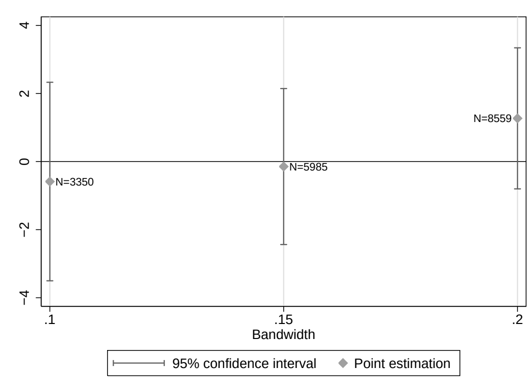
(d) Spanish Test Scores



(e) Fraction of Low SES



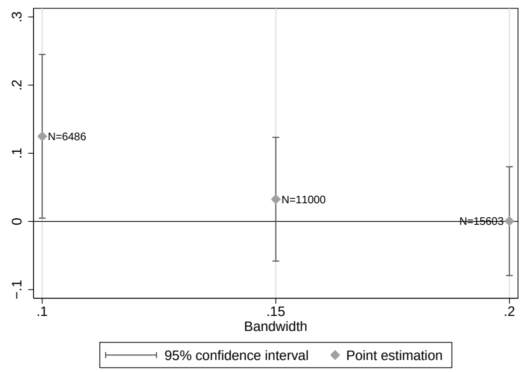
(f) 2006 School Attendance



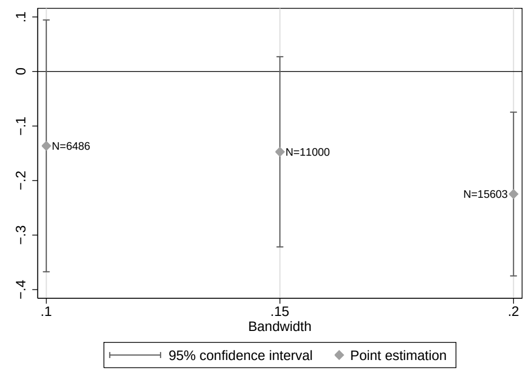
Note: The plots show the estimation of the FRD procedure but using all the control variables as the dependent variable. The estimation is run without the other control variables as covariates.

Figure 19: Effect of grade retention on different outcomes, robust estimation for C.I
(Men and women who did not repeat before 2007)

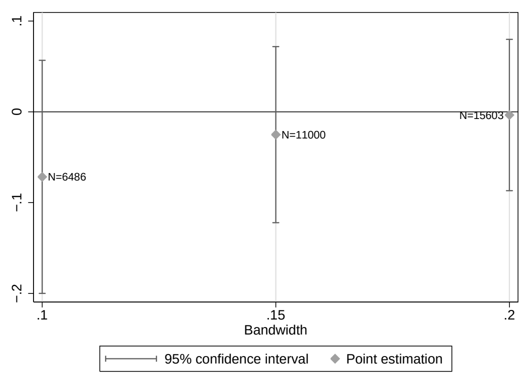
(a) Juvenile Crime



(b) Future Grade Retention



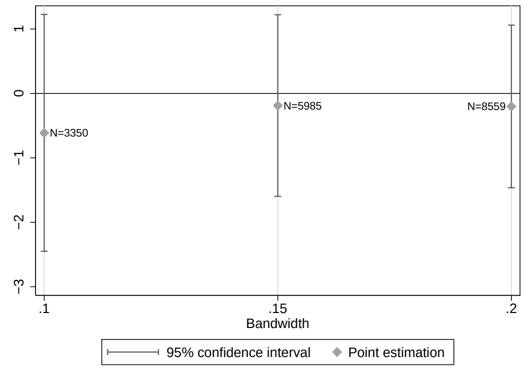
(c) Dropping Out



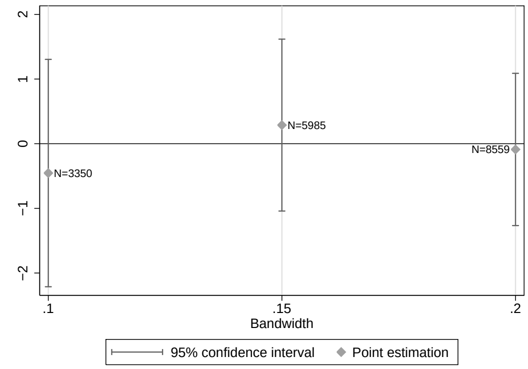
Note: The plots show the effects of grade retention on different outcomes using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 20: Placebo estimations with control as dependent variables, robust estimation for C.I.
(Men and women who did not repeat before 2007)

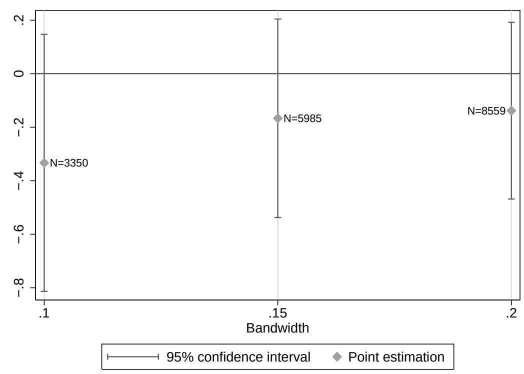
(a) Father's Education



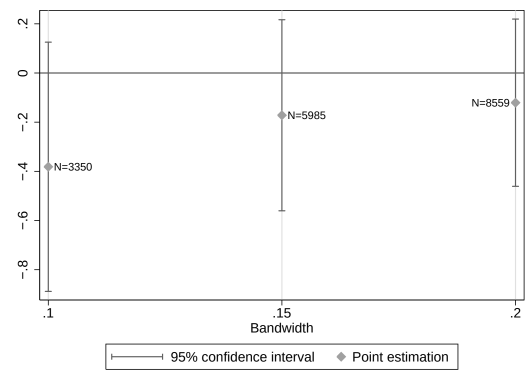
(b) Mother's Education



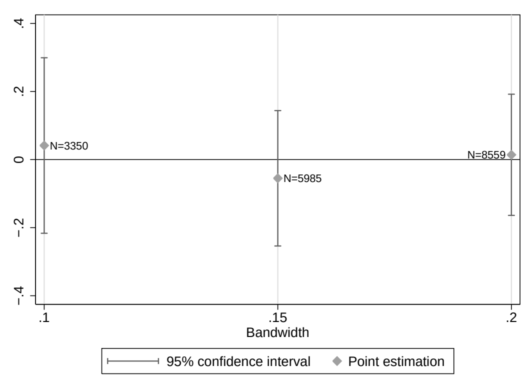
(c) Math Test Scores



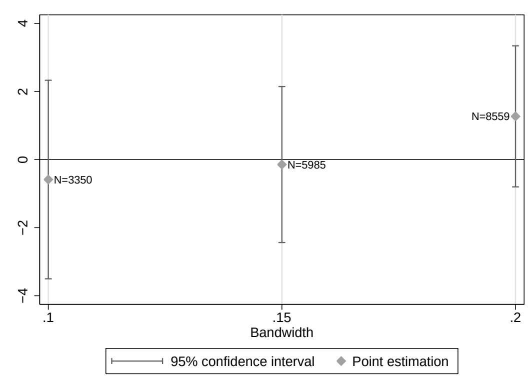
(d) Spanish Test Scores



(e) Fraction of Low SES



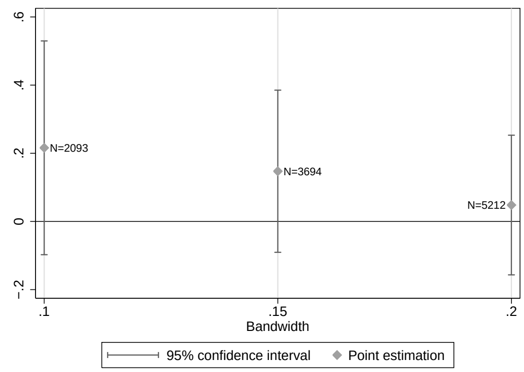
(f) 2006 School Attendance



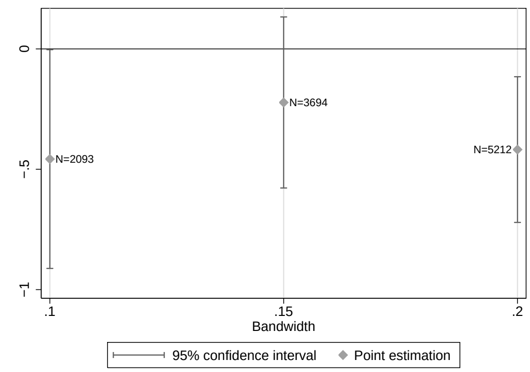
Note: The plots show the estimation of the FRD procedure but using all the control variables as the dependent variable. The estimation is run without the other control variables as covariates.

Figure 21: Effect of grade retention on different outcomes, robust estimation for C.I.
(Low SES males who did not repeat before 2007)

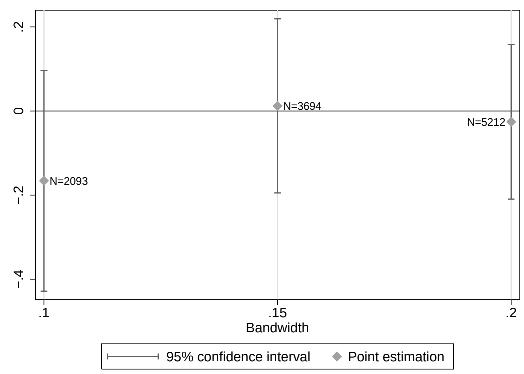
(a) Juvenile Crime



(b) Future Grade Retention



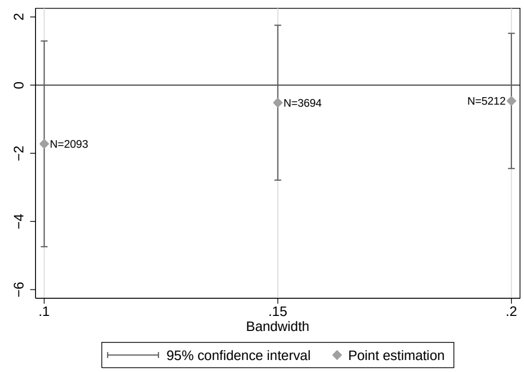
(c) Dropping Out



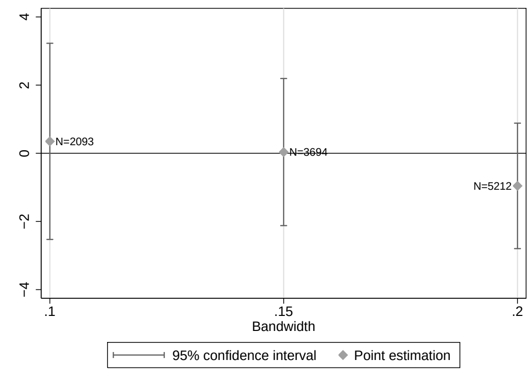
Note: The plots show the effects of grade retention on different outcomes using the FRD method that corrects for potential bias. For each bandwidth value, the point estimation is presented together with the number of individuals considered in the estimation.

Figure 22: Placebo estimations with control as dependent variables, robust estimation for C.I.
(Low SES males who did not repeat before 2007)

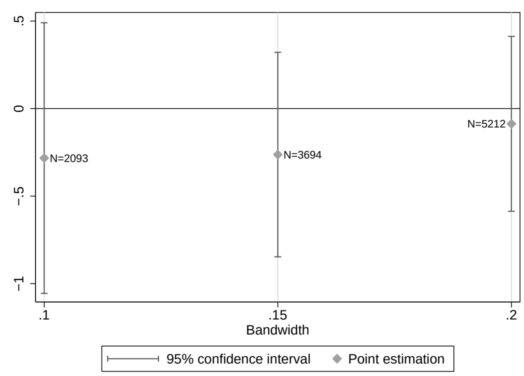
(a) Father's Education



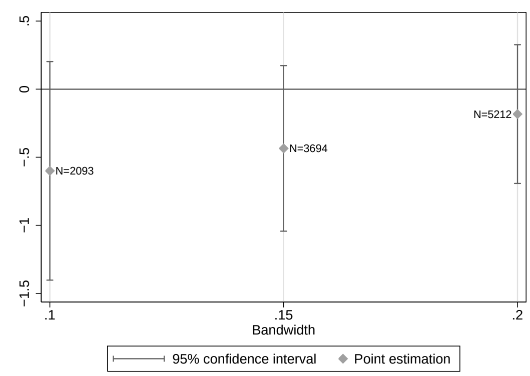
(b) Mother's Education



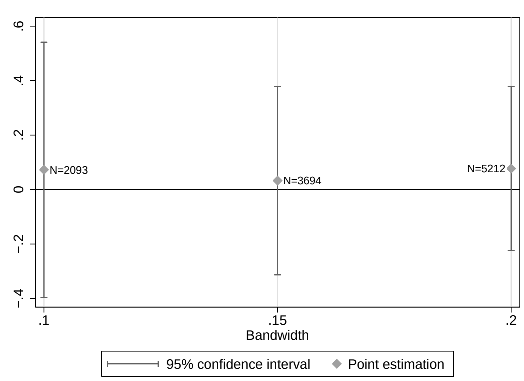
(c) Math Test Scores



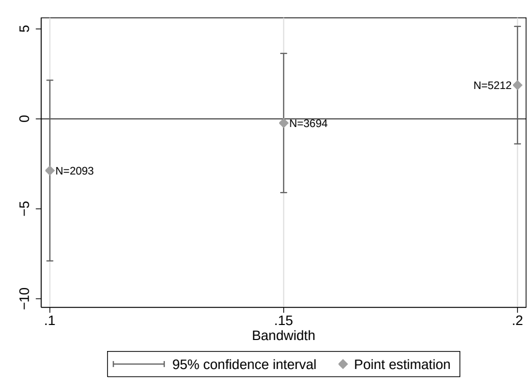
(d) Spanish Test Scores



(e) Fraction of Low SES



(f) 2006 School Attendance



Note: The plots show the estimation of the FRD procedure but using all the control variables as the dependent variable. The estimation is run without the other control variables as covariates.

References

Anderson, D. M. (2014): "In school and out of trouble? The minimum dropout age and juvenile crime," *The Review of Economics and Statistics*, 96(2), 318–331.

Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011): "Saving Babies? Revisiting the effect of very low birth weight classification," *The Quarterly Journal of Economics*, 126(4), 2117–2123.

Brugård, K. H., and F. Torberg (2013): "Post-compulsory education and imprisonment," *Labour Economics*, 97, 97–106.

Burdick-Will, J. (2013): "School violent crime and academic achievement in Chicago," *Sociology of education*, 86(4), 343–361.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a): "Robust data-driven inference in the regression-discontinuity design," *Stata Journal*, 14(4), 909–946.

——— (2014b): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

——— (2016): "Regression Discontinuity Designs Using Covariates," Working papers.

Cook, P. J., and S. Kang (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.

Depew, B., and O. Eren (2015): "Test-Based Promotion Policies, Dropping Out, and Juvenile Crime," Departmental working papers, Department of Economics, Louisiana State University.

——— (2016): "Born on the wrong day? School entry age and juvenile crime," *Journal of Urban Economics*, 96, 73–90.

European Institute for Crime Prevention and Control, affiliated with the United Nations (2010): "International Statistics on Crime and Justice," HEUNI Publication Series 64, United Nations.

Fagan, J., and E. Pabon (1990): "Contributions of delinquency and substance use to school dropout among inner-city youths," *Youth and Society*, 21(3), 306.

Grau, N., D. Hojman, and A. Mizala (2016): "Destructive Creation: School Turnover and Educational Attainment," Discussion paper.

Hahn, J., P. Todd, and W. V. der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2004): "Disruption versus Tiebout improvement: the costs and benefits of switching schools," *Journal of Public Economics*, 88(9-10), 1721–1746.

Hirschfield, P. (2009): "Another way out: The impact of juvenile arrests on high school dropout," *Sociology of Education*, 82(4), 368–393.

Holmes, C. T., et al. (1989): "Grade level retention effects: A meta-analysis of research studies," *Flunking grades: Research and policies on retention*, 16, 33.

Imbens, G. W., and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142(2), 615–635.

Imbens, G. W., and D. B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge Books. Cambridge University Press.

Jacob, B. A. (2005): "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago," *Journal of Public Economics*, 89, 761–796.

Jacob, B. A., and L. Lefgren (2009): "The Effect of Grade Retention on High School Completion," *American Economic Journal: Applied Economics*, 1(3), 33–58.

Jimerson, S. R. (2001): "Meta-analysis of grade retention research: Implications for practice in the 21st century," *School psychology review*, 30(3), 420.

Keele, L., R. Titiunik, and J. R. Zubizarreta (2015): "Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout," *Journal of the Royal Statistical Society Series A*, 178(1), 223–239.

King, E. M., P. F. Orazem, and E. M. Paterno (2015): "Promotion with and without learning: Effects on student enrollment and dropout behavior," *The World Bank Economic Review*, p. lhv049.

Landersø, R., H. S. Nielsen, and M. Simonsen (2016): "School Starting Age and the Crime-age Profile," *The Economic Journal*, DOI: 10.1111/ecoj.12325.

Lee, D. S., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

Lochner, L. (2004): "Education, Work, And Crime: A Human Capital Approach," *International Economic Review*, 45(3), 811–843.

Lochner, L., and E. Moretti (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94(1), 155–189.

Machin, S., O. Marie, and S. Vujić (2011): "The crime reducing effect of education," *The Economic Journal*, 121, 463–484.

Manacorda, M. (2012): "The Cost of Grade Retention," *The Review of Economics and Statistics*, 94(2), 596–606.

Reschly, A. L., and S. L. Christenson (2013): "Grade retention: Historical perspectives and new research," *Journal of school psychology*, 51(3), 319–322.

Roderick, M. (1994): "Grade retention and school dropout: Investigating the association," *American Educational Research Journal*, 31(4), 729–759.

Thornberry, T. P., M. Moore, and R. Christenson (1985): "The effect of dropping out of high school on subsequent criminal behavior," *Criminology*, 23(1), 3–18.

Wu, W., S. G. West, and J. N. Hughes (2010): "Effect of Grade Retention in First Grade on Psychosocial Outcomes.," *Journal of educational psychology*, 102(1), 135–152.

Zubizarreta, J. (2012): "Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure after Surgery," *Journal of the American Statistical Association*, 107, 1360–1371.

Zubizarreta, J., and C. Kilcioglu (2016): "designmatch: Construction of Optimally Matched Samples for Randomized Experiments and Observational Studies that are Balanced by Design," R package version 0.1.1.