WORKING PAPER

N° 2020-6

THE EFFECTS OF AGE ON EDUCATIONAL PERFORMANCES AT THE END OF PRIMARY SCHOOL : CROSS-SECTIONAL AND REGRESSION DISCONTINUITY APPROACH APPLICATIONS FROM REUNION ISLAND

Daniel RAKOTOMALALA

www.tepp.eu

TEPP – Theory and Evaluation of Public Policies - FR CNRS 2042

The effects of age on educational performances at the end of primary school : cross-sectional and regression discontinuity approach applications from Reunion Island

Daniel Rakotomalala

Abstract

This paper focuses on the quantitative measure of the causal relationship between age and school results of pupils at the end of primary school in Reunion Island. The effect of age is composed of at least three distinct ones : (1) age at entry effect, (2) age at test effect and (3) relative age (compared to grade peers) effect. In order to extend the knowledge about determinants of educational sucess, especially about the impact of age on scholar results and then help policy makers in their decisions about optimal policies in the education field by providing informative results ; this paper, using cross-sectional data sets, exploits an exogenous variation of the age at test within a grade induced by the date of birth to measure the causal impact of age at test on the national achievement assessment scores in grade 5 in Reunion Island. I implement additionally a regresion discontinuity design for comparison purpose. The principal findings are that the age at test have a substantial positive effect on test scores in grade 5. Also, the effects in grade 5 are heterogeneous across sex subgroups but such a pattern is difficult to draw across social category subgroups. These results would suggest at best that, in order to improve the educational results of pupils in Reunion Island meaning the age variable, policy makers could first increase the minimum age of school entry. Second, they could regulate classroom compositions such that the age distribution within a classroom does not disperse too much. Third, they could normalize national achievement assessment scores by age or making pupils with different ages within a grade pass the national assessment at different times such that they have sufficiently close ages at test to not significantly impact their results. The latter enables at the same time to correct the inequality of having a different month of birth (unchose by the pupils) which is likely to lead all else equal towards different educational outcomes. Pupils would be indeed assessed « at equal luck ».

Keywords : Age at test, Relative age, Month of birth, Educational performances

Contents

List of Tables

List of Figures

1 Introduction

1.1 Context

In the french (primary) educational system, there is a unique cutoff date of eligibility for entering school : children must have turned six by the December, 31st of the calendar year containing the school entry date (August, 18th or 19th).[1] This leads to the existence of a continuum of children ages within a grade because those born in different months of birth find themselves being in the same grade. For example, for the 2005-2006 school year, because the school entry date is the August, 18th and because of the condition mentioned above, all chidren who turn six by the December, 31st (regardless of their months of birth) are eligible to enter school for the 2005-2006 school year. Hence, by the end of 2005 (within the grade 1 of the 2005-2006 school year), there are children aged between exactly 6 year (born in the December, 31st of 1999) and almost 7 year (born

[1] A more detailed presentation of the french primary education system is provided in Section 3.

in the January, 1st of 1999). It is known that generally, this difference in ages causes differences in educational outcomes (test scores at several school grades, cognitive abilities and even later life outcomes). More precisely and generally, the youngest ones of a grade have generally some disadvantages compared to their older peers.

1.2 Underlying mechanisms of observed age-based outcome differences

Several mixed potential explanations underly this evidence. Cascio and others (2008) gives a good summary of these concerns.[2]

First, the difference in educational outcomes between the young ones and the older ones can be due to a **relative age effect**. In other words, there is peer effects such that, because of their younger peers, the older ones tend to do better.[3]

Second, the observed age effect can cover an **absolute maturity effect**. This is of special concern if the studied outcomes are of educational performance types (for example test scores or higher education participation) because the absolute maturity can be reflected in the age at which the individual sits the examination (thus there is the « absolute age » appelation in some papers of the literature). This second effect has to be distinguished from the first in the sens that on average, older an indiviual takes an examination, regardless of his rank into the age distribution of his grade, better his results will be. It means that within a grade, if the older ones and the younger ones within a grade take their respective exams at different dates such that they are equally aged at the moment of these exams, on average, they would have the same performance (if the age effect is solely composed of an absolute age effect). As matter of intepreting this second effect, borrowing Kaila (2017) terms, « older students do not learn at a faster rate, but they do better in exams just because they have had more time to accumulate knowledge ».

Third, the age-based gap in educational outcomes can be caused by a pure **age of school entry effect**. This feature is partially motivated by a child developmentalists concern : the readiness of a child to enters school (Fredriksson and Öckert 2006). In fact, since being among the youngers within a cohort implies, if no redshirting or grade repetition occured, having an inferior age at school entry, younger children could perform worse because they were not sufficiently mature when they entered school, which negatively troubled their learning skills in the following school years. Moreover, to clearly distinguish this effect from the two previous ones, consider children such that some are relatively older compared to their peers and at the same time older at test dates. Then, if there were no existence of relative age effects nor age at test effects and all children started school at the same age, that would lead to zero difference in educational outcomes at all between the olders and the youngers.

Fourth, there is potentially a **length of schooling effect**. This can be interpreted as follows : children differs in some educational outcomes because some of them spent more time in school[4]

[2] Although the paper refers to entry age differentials, the reasoning can be reported into a within grade age differentials based reasoning.

[3] An hypothetical example is that the olders gain self-confidence and motivation due to their consciousness of being more aged compared to their classmates.

[4] As explained in Grenet (2009), this is caused by the existence of multiple entry cutoff for the same school year and compulsory laws in some states.

(hence had accumulated more knowledge).

1.3 Empirical challenges

As this is all theory, and no measure of such effects can be done theoretically, measuring the causal impact of age within a grade is reported to an empirical problem (Black, Devereux, and Salvanes 2011 ; Fredriksson and Öckert 2006 ; Robertson 2011). Then, data is necessary to perform such (quantitative) measuration. Following from this, there is principally two empirical challenges that one will most likely have to face within this type of study.

First, isolating the different age effects mentioned previously is empirically difficult because of, given a unique school system, the perfect collinearity between age at test, age of school entry and length of schooling. Indeed, the age at which a pupil takes a test equals to the sum of his age at entry school and his length of schooling.
Consequently, to isolate an, age at test effect for example, there is the necessity to find a framework and data in which age at test varies independently from age at school entry and length of schooling (Crawford, Dearden, and Meghir 2007). It means that the framework allows the age at test to vary for the same age at school entry and length of schooling. Given that such conditions are hard to fulfill, only few studies managed to separate some age effects from another : Black, Devereux, and Salvanes (2011), Crawford, Dearden, and Meghir (2007), Cascio and Lewis (2006) (see the literature review in Section 2 for further details). Otherwise, most of the studies present estimates that contain more than one of these effects.

On the other hand, in a within grade comparison of educational outcomes by age, the age variable (age at school entry or age at test) is endogeneous. This is because, given a grade, there are some children that are aged at least one year more than their theoretical age (the age they would have if they entered at the first time they were eligible and if they did not repeat or skip a grade)[5] and there are some aged at least one year less. This phenomenon is not a random one, impying that the comparison mentioned above suffers from bias selection. In fact, those who have one year of delay compared to their theoretical age can be scinded into two types : those who went through grade retention and those whose parents intentionally delayed their age at school entry (this practice is called « redshirting »). The second one has two potential explanations : either the parents were awared about the advantage of the olders within a cohort and then delay their child school entry in purpose to make him one of the oldests in his grade ; either the parents observed that their child have learning disabilities compared to normal children and then decide to delay their school entry in order to give time to the child to be more ready for school environments. There is then a positive correlation between the age and grade retention (which is not accounted for in the comparison of educational outcomes by age). The grade retention, in its case, is arguably due to the poorer ability of the concerned pupils. It is then negatively correlated to educational outcomes. If these practices (redshirting and grade retention) were randomly practiced conditional on age (i.e if they were not correlated to the age – of school entry or at sitting tests –), they would cause no endogeneity problem at all. Unfortunately, it has been observed that repeaters or redshirters

[5] For instance, in France, the theoretical age for being in grade 5 is the age of 10.

are more likely to be born at the end of the year (meaning that they are more likely to be the relatively youngest ones within their cohorts, which at it turn basically means that grade retention does not occur at random conditional on age of pupils)[6]. A natural questioning follow from these purposes : while grade retention is likely to non-randomly lower average educational outcomes of the youngests of a grade[7] and redshirting is likely to do the opposite, what is then the sign of the bias ? To anticipate the answer, let us consider that since repeaters are generally more numerous than redshirers, the age effect estimation suffers from a downard bias (Bedard and Dhuey 2006 ; Grenet 2009 ; Hámori 2007 ; Hámori and Köllő 2012).
The same logic applies to those who are in advance of one year because it is most likely that children with higher abilities are enrolled (by decision of their parents) earlier in school, making them highering the average educational outcomes of the oldests of a grade. What is to be kept in mind is these advanced pupils are not numerous at all (empirical informations are given in Section 3), making the downward bias dominating.

1.4 Contributions

This paper takes into account the above discussed endogeneity by instrumenting the age at test by the assigned relative age[8] (it can be percepted by the age position at which a pupil should be if he entered school by the first time he were eligible and if he did not repeat or skip a grade) to estimate the effect of age on standardized national achievement assessment scores in the end of primary school in Reunion Island using cross-sectional data sets. I indeed exploit the arguable exogeneity of month of birth and its strong prediction power of the actual age at test to estimate by an instrumental variable and a reduced form framework the effect of age on national test scoroes at grade 5. In addition, I perform a control function approach which give interesting supplementary informations compared to the methods mentioned above. Then, I attempt to provide further investigation on the clear identification of what the estimates measure since one of the fundamental instrumental variable identification assumption appears to be, by framework, violated here : the monotonicity assumption. Last, as robustness check, I implement a regression discontinuity design which I expect to yield to comparable estimates of age effects to the previous approaches. This method measure in fact the effect of age with a different approach question : what is the effect for the average pupil of being one year older at examination due to the entry rule cutoff date ?
According to my knowledge of the existing studies about this issue or a comparable one, this is the first conducted with Reunion Island data. Also, this study is of education policy matter because of its informative causal findings combined with the possibility for policy makers to influence one variable of interest determining educational outcomes : the age. In fact, I find a substantial and positive effect of being one year older on total test scores that varies from +0.2 to +0.3 of a standard deviation. I also find heterogeneous effects by group of sex with the female pupils generally gaining

[6]See Datar (2006) for example.

[7]Because repeaters are supposed to have poorer abilities then poorer results.

[8]Which is computed as the time distance between the December, 31st of the pupil's year of birth and his month of birth. And in this paper, it is reported to year unit.

more than males pupils from being a year older. Although the estimated were expected to be clearly heterogeneous by social category, it appears that it is not what is retrieved here. The control function approach approach results are arguably similar to those of the classical instrumental variable results. Since the France educational system allow significant amount of redshirting / grade repetition, reduced form estimates, i.e estimates containing, in addition to age effects, the effects of these mentioned phenomenon (further explanations are upcoming).

The remainder of the paper is organized as follows : Section 2 presents a literature review, Section, 3 describes the data I used, the Reunion Island education system and the econometric framework. The two next sections : 4 and 5 respectively exposes the results (with discussions) and concludes.

2 Literature review

2.1 Existing literature

The relationship between age (in different forms) and educational outcomes in general is a well documented topic. The literature body was already large until 10 years earlier. In the end of the 1980's for example, Cahan and Cohen (1989) investigate the effect of age and schooling (to be distinguished) on scores obtained from ability tests in grades 5 and 6 using Jerusalem's data. Another example in the beginning of the 1990's is the study performed by Bell and Daniels (1990) where they compare within a grade automn-born children and summer-born children at their 11st, 13th and 15th years, on their APU (*Assessment of Performance Unit's Science Project*) Science Syrvey tests to assess the effect of being one year older (birthday effect) than their classmates in education.[9] More studies using number of outcomes types appear to adress the relationship between age and education in the 2000's (and before 2010) than in the 1990's. Taking an example fo them, consider Graue and DiPerna (2000). This paper performs a statistical analysis of achievement gap (promotion to the next grade) between those who delay their entry in kindergarten or are retained[10] and those who enter school as soon as they are eligible in Wisconsin. Using another type of variable of interest, Leuven et al. (2004) work on the impact of expanding school enrollment opportunities on achievement (language or math test scores) in Netherlands. In other point of view, this study estimate the effect of schooling (measured by potential months enrolled in school) on the test scores of interest using datas from the PRIMA (Primary Education and Special Education Cohort Studies) survey. Also, almost all of these papers are interested in the effect of school starting age and they come from considerable variety of states. For instance, Strøm (2004) estimates the effect of school starting age on reading test scores for 15-16 years old pupils in Norway using PISA (*Programme for International Student Assessment*) data. In addition, the same outcomes are of interest in Germany, United States, Sweden, England, Hungary and France respectively in Puhani

[9] For other references, see Langer, Kalk, and Searls (1984) ; Cahan and Davis (1987) or Mayer, Knutson, and others (1997). The first attemps to highlight the causal effect of school entry age and relative age on mathematics, science and reading test scores using National Assessment of Educational Progress (NAEP) data . The second in its case adress the measure of the effect of schooling on achivement (measured by nationa test scores) whilst the third exploits the National Longitudinal Survey of Youth (NLSY) to estimate the school entry age effect on cognitive and non-cognitive (behaviour problems) developement.

[10] In kindergarten or grade 1-3

and Weber (2005)[11], Datar (2006)[12], Fredriksson and Öckert (2006)[13], Crawford, Dearden, and Meghir (2007)[14], Hámori (2007)[15] and Grenet (2009). For studies interested in the effect of age on another outcomes, see for example Fertig and Kluve (2005) which use obtained degree and probability of retention as outcome in Germany ; or Dhuey and Lipscomb (2008) which adress the effect of relative age on high school leadership activities (being the president of a club or the captain of a team).

Even from 2010 till the present moment, the literature body continues to extend. For example, an interesting study with noticeable outcomes of study is Black, Devereux, and Salvanes (2011). This paper investigates the long-run impact of entry age on IQ scores at age 18 and teenage pregnancy (in addition of usual edumcational attainment and earings.). Additionally, in the recent years and from many states, there seems to be a set of studies that are interested in the effect of age on outcomes besides test scores type. In fact, while Mühlenweg and Puhani (2010) ; Jürges and Schneider (2011) or Schneeweis and Zweimüller (2014) are into outcomes of track choice types in, respectively Germany (for the two first) and Austria, studies like Suziedelyte and Zhu (2015) and Dhuey et al. (2017) investigate in outcomes of cognitive and non-cognitive development types in, respectively Australia and Florida.

2.2 Conceptual considerations

Almost all of these papers mentioned above are awared of the potential explanations underlying age effects (relative age effect, absolute age effect, school starting age effect and length of schooling effect). Note that in the french context, since there is generally a single school entry cutoff date considered, it makes sens to discuss issues about school entry age and length of schooling that are massively developped in the literature body, especially those of school starting age.
Consider the following questions that needs involvement of conceptual considerations. Is it better to delay school entry ? What is the optimal age at which a child should start school and what policies or interventions[16] are necessary to improve outcomes of interest without harming other children ?
To begin with, several theoretical considerations about how age would affect educational and later was made over years. The ones presented here is surely not exhaustive[17], but appears to have their importances. Councerning the debate about delaying or not school entry, this question was

[11]Another studies in Germany as Fertig and Kluve (2005) can be retrieved who are interested in the impact of school starting age on schooling and retention ; in Mühlenweg and Puhani (2010) which consider track attendance (academic versus non academic track) as outcome.

[12]See Aliprantis (2014) and Fletcher and Kim (2016) for other studies in the United States. The former estimates separately pure entry age effect and relative entry age effect on mathematics and reading item response theory test scores while the latter investigates the impact of entry cutoff changes on National Assessment of Educational Progress (NAEP) test scores.

[13]The same authors published another study from Sweden in 2013 (Fredriksson and Öckert 2013) in which they put interest in educational attainment and 25-54 years old earnings.

[14]A recently published study from England (Wales) is Hart and Moro (2017) in which the authors study the impact of quarter of birth on the probability of gaining selective school entry.

[15]See also Hámori and Köllő (2012).

[16]Beside action from policy makers, parents could decide of their children's own scholl entry age.

[17]This is because theoretical concepts are often given very implicitely in the economic literature. Also, these are more of child development concerns rather than economic concerns. Finally, several desagreements seems to persists about the theory.

largely adressed in the United States around the 2000's[18].
One key concept of interest behind the age of school entry of a children is the « readiness ». Under this latter, there are hypothetical factors that are of cognitive, as non-cognitive dimensions (Stipek 2002).[19] More concise, DeCos (1997) provides a clear classification of theories behind « readiness for kindergarten » : *maturationist*, *behaviorist*, *envrionmentalist* or *interactionist/constructivist* theories.
For the maturationists, the readiness is defined solely within the child and depend only on his biological age. Thus, education has just to provide optimal environment for the child's maturation. From this point of view , children who are suspected to be not mature enough at a certain point are given a « gift of time » (by redshirting, retention or transitional classes). On the other hand, for the behaviorist and the environmentalist, the knowledge of a child is external and skill is considered as a puzzle such that its peices are supposed to be identified and assembled by the education. Last, an interactionist/constructivist have a combined idea between the maturationist's and the behaviorist's. About the optimal age of school entry, it appears to vary from 4 to 7 years old for several countries cited in DeCos (1997). Since there appears to be no theoretical conception of this issue, empirical studies adressing relationship between age at school entry and cognitive and non-cognitive development can help for clarifications. Refering to Aliprantis (2014) which provides a stress on the importance of investing in earcly childhood education, it is important because early skill accumulation is complementary with later one then can positively later life outcomes. Hence, this is of policy matter because policy makers can act on the age of kindergarten entry. Indeed, in the United States, there was a massive rise in minimum school entry age in the last decades (Datar 2006 ; Elder and Lubotsky 2009) in response to these considerations. The most chosen kindergarten entry age appears to be around 5 years old (Elder and Lubotsky 2009). The other maneers to affect the age at which a child enters formal education are redshirting, pre-school interventions, transitional classes or grade retention. How did studies manage to highlight age effects on educational or later-life outcomes ? As Stipek (2002) managed to survey 36 papers on age effect of his time, the approaches used by the authors can generally be classified in three categories : first, a comparison in outcomes between pupils who delay their school entry and those who do not ; second, a comparison within a grade of pupils with different birth dates ; and third, the combination of the previous two. Peña (2017) greatly managed to add his survey on age effects on recent (from 2006, for e.g Bedard and Dhuey (2006) to 2014, for e.g Nam (2014)[20]) works. The most used methods appear to be the within grade approach using IV methods. Especially, those who use the theoretical relative age as instrument for age are massively used. When OLS techniques are used instead of IV approaches, the fixed effects of month of births or its variants (quarter of birth for example) are controlled for. This latter feature requires that there is sufficient independent variation between the age and the month of births. RDD appear are rarely applied, according to the survey. Additionally to the papers cited in Peña (2017), two works that perform similar RDD frameworks seem to be worth mentionning : Kaila (2017) and Matta et al. (2016). The second one is a study in Brazil that investigate, using the school entry rule to implement its RDD, the effect of school entry age on test scores, college admission and earnings.

[18] Some empirical examples are cited above

[19] Some consider wellbeing, social and language development for example.

[20] This paper studies the relationship between age and and test scores in grade 7, 8, 9 and 12 in Kora using several datas.

2.3 Existing evidences

While doing a within grade comparison of children's outcomes, it is largely demonstrated (with rare exceptions) that the older children perform better than their younger peers. See Dhuey et al. (2017) for an extensive citation of this evidence. In other words, the age have a positive causal effect on educational outcomes. This relative advantage of the olders is more pronounced in early ages and fades over time (Bedard and Dhuey 2006 ; Elder and Lubotsky 2009 ; Grenet 2009). Bedard and Dhuey (2006) which investigated the effect of school starting age on mathematics and science internationally standardized test scores in OECD countries found that the oldest children within grade 4 score from about +0.2 to +0.4 of a standard deviation. The remaining effect at grade 8 is from about +0.1 to 0.4 for mathematics and science test scores again. Similarly, Elder and Lubotsky (2009) found an advantage of +0.16 of a standard deviation in mathematics test scores at grade 8. Also, for France, Grenet (2009) found that being a month older compared to the youngest children inscrease mathematics and test scores by approximatively +0.02 of a standard deviation. When reported to a year scale, it is equivalent to an advantage of +0.24 of a standard deviation, which joins the precedent results. This effect is attenuated in the future (grade 9) as mentionned earlier by shrinking to an advantage of just +0.02 and +0.13 in, respectively mathematics scores and french scores for those who are a year older.[21] A recent study, yet using alternative method (a regression discontinuity) but leading to similar results[22] is Peña (2017) which estimates the effect of relative age in test scores in Mexico for grade 3-9 pupils. The author found differences between means of mathematics test scores on one side and on another the cutoff (a causal effect in this case) from +0.3 to +0.36 of a standard deviation. Besides, note that the comparison between test score age effect on test scores reported in unit of a standard deviation only makes sens when recalling that over a year, an average student can gain from one quarter to one third of a standard deviation in test scores (Woessmann 2016).
On another side, few studies find small or negative age effects. Concerning the first case, probably the most known example is the Black, Devereux, and Salvanes (2011)'s paper in which they find, in Norway, small effect fo school entry age on IQ scores.[23] An example for the second case is Fertig and Kluve (2005) in which they found no effect on entry age on school attainment (schooling degree and probability of retention) in Germany. Even, Mayer, Knutson, and others (1997) assessed at that period that entering school at an younger age give an advantage in cognitive and non-cognitive development. Going with, Dobkin and Ferreira (2010) found that younger school enrollment increase the education attainment level in Texas and California. These exceptional evidences make somewhat informations about age effects on educational outcomes related variables slightly inconclusive.

In the other hand, the advantage of older children compared to their same grade peers is sometimes revealed to be heterogenous with subgroups (generally with sex subgroups or social category or similar type subgroups). For instance, Bedard and Dhuey (2006) found found that the relative age of school entry effect to be higher for children at risk over OECD countries. This is as well what Grenet (2009) highlighted in France : the age effect is more important in early years for

[21] In the Table 1 of Grenet (2009), they are expressed in month scales : +0.002 and +0.011 respectively.

[22] Though higher estimates

[23] Eventhough they reported a strong effect of age at test, the advantage of entering school a year later in Germany was founded to be also +0.06 *of a stanine* (about +0.04 of a standard error).

disadvantaged (refering to the child's household's occupation) pupils.[24] It appears to be also the case with Schneeweis and Zweimüller (2014) which studied the causal effect of relative age on probability of getting higher school tracking using PISA data. Recently, Aliprantis (2014) likewise found highly heterogeneous effects on math and reading item response theory test scores by home environment in the United States. These heterogeneity evidences are relevant in policy perspectives in the sens that they sugger policy makers to point interventions towards the concerned subgroups instead of towards all types of individuals. More recently and going with the set of papers that highlighted heterogeneity of age effects, Kaila (2017), in Finland, found a greater effect[25] of females than for males.

Overall, these existing evidences should be analyzed and interpreted carefully because many mechanisms besides the considered age of study can change the resulting effect (Kaila 2017 ; Aliprantis 2014). One of these mechanism is the plurality in features of school systems all over the world. For example, in Japan, the length of schooling of a pupil is invariant with the month of birth because the Japan school system requires individuals to accomplish a fixed amount education regardless of their month of birth (Kawaguchi 2011). Thus, the author could estimate a pure school entry age effect. Another case when Kaila (2017) compared her results with that of Bedard and Dhuey (2006) within Finland. The magnitude of the estimate of school starting age on test scores reported by Kaila (2017) is smaller than that is reported in Bedard and Dhuey (2006). Moreover, an essential institutional feature that may gratly differ from a country to another and that could explain the differences in the estimated effects within different study is the amount extensive use or not of grade retention or redshirting. Bedard and Dhuey (2006), in their papers clearly highlight differences in estimates magnitudes patterns depending on the retention / redshirting rate of the country. One interpretation the former's author give is that this can be caused by the simple fact that the test score used as outcome in her study, the Grade Point Average (GPA) is based on teachers'personal assessment while in the latter the test scores were internationally standardized ones (Trends in International Mathematics and Science Study or TIMSS).

3 Methods

3.1 Institutional background and data

3.1.1 The Reunion Island educational system

Since the Reunion Island is a region of France, its educational system can be, at least in the interest of this paper, presented as the Reunion Island system. Alet, Bonnal, and Favard (2013) provide a good description of primary school features in France : primary school is composed of five years (grade 1 to grade 5). Compulsory school begins in grade 1 and the age at which a children must starts formal school is the age of 6 (at least).

Then, at the end of the grade 5, all children pass a national achievement assessment (*évaluation nationale des acquis*). Resulting from that, children receives a score between 0 and 100, with the

[24] Though, Grenet (2009) found no significance difference between males and females pupils.

[25] Effect of relative school starting age on test scores and admission in general school.

government expected score to be 50. A brief description of the national achievement assessment will be presented just next. One invariant feature is that for each assessment (each school year), two documents are established : one for the pupils that take the evaluation (*cahier de l'élève CM2*) and one for teachers who will manage to correct the assessments (*Livret enseignant*). The first is just the document in which there are the assessment questions and materials inteded for the pupils answerings. The second contains precise instructions intended for teachers when the national evaluation takes place. For example, it requires teachers to check before the examination if the children have at their disposal necessary materials such as an eraser or a serviceable pencil. The principal informations within the teachers document are : the identification of the knowledges and skills to be assessed, necessary informations for passing exercises, the time affected to each exercise and all necessary instructions concerning the correction process.

About the content of the assessment itself, there are one hundred questions called « items » (which are ranked from 1 to 100) to which pupils are supposed to give answers. These can be divided and subdivided by topics and subtopics. Indeed, the first level of division is such that the items ranked from 1 to 60 are related to french topic exercises and the items ranked from 61 to 100 are related to mathematics topic exercises. Moreover, both the french topic items and the mathematics topic items can be themselves subdivided. Indeed, the subtopics within the french items are reading, writing, vocabulary, grammar and spelling while those within the mathematics items are numbers, calculation, geomtetry, sizes and measures, and data organization and management.[26] Descriptions of each subtopics are given in the teacher document.[27]

Concerning the cutoff rules, a pupil is first eligible to the $t - t + 1$ school year for grade 1, with the entry date placed at the third week of August of the year t, if he turns 6 by the December, 31[st] of the year t. For example, the December, 01[st] of 2000 is considered as a cutoff date of birth in the sens that those who are born just before are normally, at grade 5 in the 2010 cohort while those who are born at this date or just after are normally, at grade 5 in the 2011 cohort. This imply that around this cutoff date, those who are born just before are comparable in age than those who are born just after. This latter idea is essential for an upcoming method in this paper : the regression discontinuity design. Aslo, as mentioned earlier, in Reunion Island, the school entry date turns around the August, 18[th]. It is because of the cyclonic period in Reunion Island around the beginning of the calendar year which makes the school vacation longer than the vacation in metropolitan France. This difference is then compensated by the earlier school entry date in Reunion Island.

On another hand, empirically, the retention rate in Reunion Island is about 16%, as illustrated in the third set of statistics in Table 2. This is quite large compared to the Metropolitan France's rate, where it's around 6% (Alet, Bonnal, and Favard 2013). The proportion of pupils who are in advanced compared to their theoretical age in grade 5 is very small : 2% (see again the third set of statistics in Table 2).[28] Hence, the proportion of pupils who are aged as their theoretical age in

[26]These splittings are invariant accross school years. In contrary, the rank and amount of items attributed to a subdivision do vary across the different school years. See Tables 24 and 25 in the appendix which support this affirmation.

[27]For example, the purposes of the writing items are for pupils to copy without error and with an adapted presentation a text ; and to write several types of texts of at least two paragraphs with coherence and good spelling.

[28]Note that this proportion is calculated after some observations such that the year of birth was superior of inferior more than one year compared to the normal date of birth given the grade were removed, as stated in the next paragraph.

grade 5 is around 82%.

Concerning the other institutional features, the Reunion Island educational system is made of 92% of public schools and 8% of private schools as seen in the fourth set of statistics in Table 2. What can be concluded is that education is quite homogeneous in all angles. For example the average class size is around 23 ± 5 pupils per class and this feature is stable across three school years[29] in which these statistics were observed.

In addition, from 1981, the french system introduced the concept of « *zones d'éducation prioritaires* » which are defined as areas where there are identified factors that cause school-based difficulties to pupils living in there. Hence, some school level measures are made in order to overcome these difficulties : the education placed-based policies. Later, instead of being defined by geographical limits, the priority education identification is defined by a network system[30]. In the data, schools are classified in three education placed-based policies categories : « *Hors Éducation Prioritaire (HEP)* », « *Écoles, Collèges et Lycées pour l'Ambition, l'Innovation et la Réussite (ECLAIR)* » and « *Réseau de Réussite Scolaire (RRS)* ». The first cateogry designates schools that do not benefit from these measures. While schools classified in the second type benefit from policies that are rather focused on elements concerning the education personnel, the RRS category is rather focused on social criterions. According to informations displayed in the fifth set of variables in Table 2, the proportions are stable across school years : around 52% of HEP schools, 25% of ECLAIR schools and 23% of RRS ones. Alternatively, a variable indicating if a school belongs to a network is easily computed and may be interesting, summary statistics of such a variable are displayed in the sixth set of variables in Table 2. It can be observed that the proportions of schools that belong to a priority education network are slightly lower compared to the proportions of schools that does not (48% versus 52%).

3.1.2 Generic description of the data

I use three cross-sectional administrative data sets to perform the estimations of the impact of age on test scores in grade 5. Each data set is a micro-level data corresponding to a school year. Recall that the three school years are : 2009-2010, 2010-2011 and 2011-2012. Those are sometimes labelled the 2010, the 2011 and the 2012 cohorts for simplicity. They are an administrative data directly obtained from the rectorship of Reunion Island.

Each set contains personal informations about the pupils (exact date of birth, sex and socio-professionnal category of the legally first responsible for the pupils), their educational achievements (5^{th} grade national assessment scores in great details[31]) and their schools (townships,the school status – i.e either they are public or private schools and the type of opriority education network which the school belongs to). Further details about the data variables will be given afterwards.

The pupils are born between 1998 and 2002. Those born in 1998 are observed in the 2010 cohort, and thus are repeaters / redshirters. Those born in 2002 are observed in the 2012 cohort, and thus have one year of advance comparing to their theoretical grade 5 age. This can be clearly

[29] 2009-2010, 2010-2011 and 2011-2012, refer to the next subsection for more details.
[30] *Circulaire n° 94.082 du 21/01/1994*
[31] Total score, french topic score and mathematics topic score as well as each subtopic score

Table 1: Year or birth, cohort and position relationship

	Cohort			
Year of birth	2009	2010	2011	2012
1997	Delayed	NA	NA	NA
1998	On time	Delayed	NA	NA
1999	Advanced	On time	Delayed	NA
2000	NA	Advanced	On time	Delayed
2001	NA	NA	Advanced	On time
2002	NA	NA	NA	Advanced

[a] For example, if a pupil is born in 2001 and observed in 2011 cohort, then this pupil is in advance.

[b] NA : Year of birth such that corresponding pupils can not be observed in the indicated cohort

visualized in Table 1 wich illustrates, for each cohort, the connection between year of birth and position. For example, within the 2011 cohort, one can not observe pupils born in 1998 or 2002 ; pupils born in 1999 are repeaters / redshirters (labelled « Delayed » in the table) while those born in 2000 and 2001 are respectively on time and advanced.

Children are aged between 9 and 12. Since the normal (without an advance or a delay of a year relative to the class) age to be in grade 5 is 10 year old and the test were taken in January and May of the civil year after the December of the school year entry, the maximum age of 12 (instead of 11) corresponds to the pupils born in January and had repeated a class.[32]

Moreover, considering the three cohorts successively, the number of observations were in order of 14000 in the three cohorts.[33] This approximation represents the whole population of Reunion Island grade 5 pupils in year schools 2009-2010, 2010-2011 and 2011-2012. Since there were outliers within the data in the sens that there were pupils being aged more than or less than one year compared to the normal age at which they should be in grade 5, they were removed from the observations. Note that even their proportions (in the three cohorts) are very small[34], they can create serious bias because some individuals were indicated to be aged less than 1 year old in grade 5 for example.[35]

3.1.3 Variables and summary statistics

The dependent variable is the 5th grade national assessment scores. It was originally a 100 scale integer score since it is defined as a sum of one hundred items scores. An item (as presented earlier) is a binary variable related to a specific question within the assessment. It takes the value of 1 if the pupil had the correct answer, 0 if not. For the purpose of comparing my estimates with

[32] For example, in the 2010 cohort, a repeater who was born in January, 1st, 1998 took the national test on January 2010 ; hence he is aged at least 12 at the moment of the test.

[33] Respectively 13630, 14708 and 13786 individuals within the three cohorts

[34] Resepctively 0.5%, 0.6% and 0.4%

[35] The remaining data thus have respectively 13561, 14622 and 13675 observations.

other works results, I normalized this test score to the mean of 0 and a standard deviation of 1. Recall that the one hundred items can be grouped by two main topics : french and mathematics, leading to the french topic and mathematics topic test scores. The french topic and mathematics topic test scores variables had originally, respectively, a scale of 60 and 40. In the same manner as the total score, I computed normalized versions of these two variables for comparability purpose.

A noticeable feature of test scores is their rising across cohorts, as seen in the first set of rows of Table 4. Indeed, the total test score mean varies from 47.50 to 52.89 (an increase of about 11%) between the school years 2009-2010 and 2011-2012. A similar variation in proportion is observed for the median and the third quartile.[36] The underlying pattern that explains this variation in total test scores is the significative variation of the mathematics test score accross the three cohorts (see the third set of rows in Table 4) and not of the french test score (see the second set of rows in Table 4 as the mean, median and quantiles are stable). Two by two comparisons (2010 with 2011, 2011 with 2012) of mean tests on the maths scores yield unsurprisingly to a 0 p-value. This can also be visualized within the Figure 1 which illustrates the empirical cumulative distributive function of each type of scores (total, french and mathematics). Among the french and mathematics scores, the domination of the 2012 distribution over the 2011 distribution which in its turn dominates the 2010 distribution is striking for the mathematics scores (third graphic of the Figure 1) while nothing can be percepted for the french scores (second graphic of the Figure 1). Consequently, one observes the significative evolution of total scores across cohorts (first graphic of the Figure 1).

[36] The corresponding variation for the first quartile is about 16%.

Table 2: Summary statistics

	2010	2011	2012
1 - Age at test			
Min	9.05	9.07	9.4
First quartile	10.37	10.35	10.7
Mean	10.71	10.69	11.03
Standard deviation	0.47	0.47	0.45
Median	10.67	10.66	11
Third quartile	10.95	10.94	11.28
Max	12.05	12.05	12.4
2 - Sex			
Females (%)	0.49	0.5	0.4
Males (%)	0.51	0.5	0.39
Missings values (%)	NA	NA	0.22
3 - Position			
Repeaters/redshirters (%)	0.17	0.16	0.15
On time (%)	0.81	0.82	0.83
Advanced (%)	0.02	0.02	0.02
4 - School status			
Privates (%)	0.08	0.08	0.08
Publics (%)	0.92	0.92	0.92
5 - Priority education network			
HEP (%)	0.52	0.53	0.52
ECLAIR (%)	0.25	0.25	0.26
RRS (%)	0.23	0.23	0.22
6 - Priority education network (yes/no)			
No (%)	0.52	0.53	0.52
Yes (%)	0.48	0.47	0.48
7 - Class size			
Mean	22.84	23.52	23.26
Standard deviation	5.36	5.27	5.29
N	13,561	14,622	13,675

Table 3: Mean of total, french and mathematics scores by institutional features

	2010			2011			2012		
	Total	French	Mathematics	Total	French	Mathematics	Total	French	Mathematics
1 - Sex									
Females	50.30	34.02	16.28	53.72	34.06	19.66	57.79	35.46	22.33
Males	44.79	28.93	15.86	47.84	28.87	18.97	52.70	30.82	21.88
2 - SPC									
Farmers	48.62	32.00	16.62	54.45	33.55	20.90	55.77	33.05	22.71
Entrepreneurs	54.37	35.55	18.81	59.61	36.55	23.06	59.63	35.85	23.78
Executives	63.50	41.00	22.50	69.22	42.15	27.07	69.87	42.19	27.68
Intermediates	56.32	37.03	19.29	61.62	37.89	23.73	62.38	37.49	24.89
Employees	50.88	33.68	17.21	55.85	34.59	21.26	57.29	34.59	22.69
Workers	47.36	31.38	15.97	50.91	31.53	19.38	52.62	31.44	21.18
Retired	56.25	36.76	19.48	63.04	39.26	23.78	64.26	38.26	26.00
Unemployed	43.14	28.84	14.31	46.99	29.36	17.63	47.58	28.41	19.17
Others	40.51	26.89	13.62	42.62	26.41	16.21	45.53	27.09	18.43
3 - SPC (grouped)									
Underprivileged	46.02	30.61	15.41	49.98	31.11	18.87	51.06	30.58	20.48
Privileged	58.67	38.22	20.45	64.38	39.40	24.99	64.57	38.89	25.68
Others	40.51	26.89	13.62	42.62	26.41	16.21	45.53	27.09	18.43
4 - School status									
Privates	56.43	37.24	19.18	62.12	38.33	23.79	62.78	38.00	24.78
Publics	46.72	30.92	15.80	49.80	30.87	18.93	52.15	31.22	20.93
5 - Priority education network									
HEP	50.73	33.50	17.23	53.73	33.25	20.47	55.80	33.57	22.23
ECLAIR	43.54	28.91	14.63	46.31	28.78	17.53	50.33	29.89	20.44
RRS	44.39	29.44	14.95	48.82	30.24	18.58	49.80	29.87	19.93
6 - Priority education network (yes/no)									
No	50.73	33.50	17.23	53.73	33.25	20.47	55.80	33.57	22.23
Yes	43.94	29.16	14.78	47.52	29.48	18.04	50.09	29.88	20.21

Table 4: Summary statistics of the total, french and mathematics test score variables

Cohort	Min	First quartile	Mean	Standard deviation	Median	Third quartile	Max
1 - Total							
2010	0	31	47.50	21.55	47	64	100
2011	0	34	50.79	21.79	51	67	100
2012	0	36	53.04	22.28	54	71	100
2 - French							
2010	0	21	31.43	13.64	32	42	60
2011	0	21	31.47	13.43	32	42	60
2012	0	21	31.79	14.01	32	43	60
3 - Mathematics							
2010	0	9	16.07	9.04	15	22	40
2011	0	12	19.32	9.39	19	26	40
2012	0	14	21.25	9.38	21	29	40

Arguably, this difference in results in mathematics is mainly caused by the changing structure and contents of the three assessments across cohorts. Indeed, the items are not assinged to the same exact question types through the 2009-2010, 2010-2011 and 2011-2012 school year national assessments. This is clearly demonstrated in Tables 24 and 25 in the appendix where the mathematics sub-items scoring systems appear to be unstable accross cohorts, in constrast of the french sub-items one. Although the following purpose is just a hypothesis, this evolving structure of assessments can reflect a will from policy makers to improve the test scores of Reunion Island pupils.

The independent variable of interest is the age at test of pupils. It is measured in years, taking in account month, exact day of birth and test date.[37] As completion of the information given above on the age at test, see the first set of statistics in Table 2. As can be observed, the mean age at test within the 2010 and the 2011 cohorts is about 10.7, whereas it is about 11 in the 2011-2012 cohort. Two underliying ideas can be illustrated here. First, the non negligeable amount of retention or redshirting, drives unsurprisingly the mean age to be above $\frac{9+12}{2} = 10.5$. Second, the mean age within the 2011-2012 cohorts is clearly above 10.7 because of the date of the assessment in 2012.[38] Since the age at test variable is directly related to date of birth, we can describe the latter next. From a month of birth perspective, their proportions are illustrated by Figure 2 and unconditionally to individual background charasteristics values, the month of birth appears to be uniformly attributed to pupils since there seems to be no overrepresented or underrepresented month of birth. Moreover, as data on month of birth distribution in entire France coud be extracted from the INSEE[39], a Reunion Island-France comparison could be made, as illustrated in the second panel of Figure 2. It can be observed that the two distributions are very similar, which means that there are in, in this general perspective, no month of birth distribution specificity of Reunion Island compared to France.

[37] For example, a pupil in the 2010 cohort with an age at test of 10.08 is aged 10 years and one month ($\frac{1}{12} \approx 0.08$) at the date of January, 20th 2010.

[38] It was taken on May, 25th 2012 while the two other assessments were taken on January, 20th.

[39] National Statistical and Economic Study Institute of France. INSEE stands for *Institut National de la Statistique et des Études Économiques*

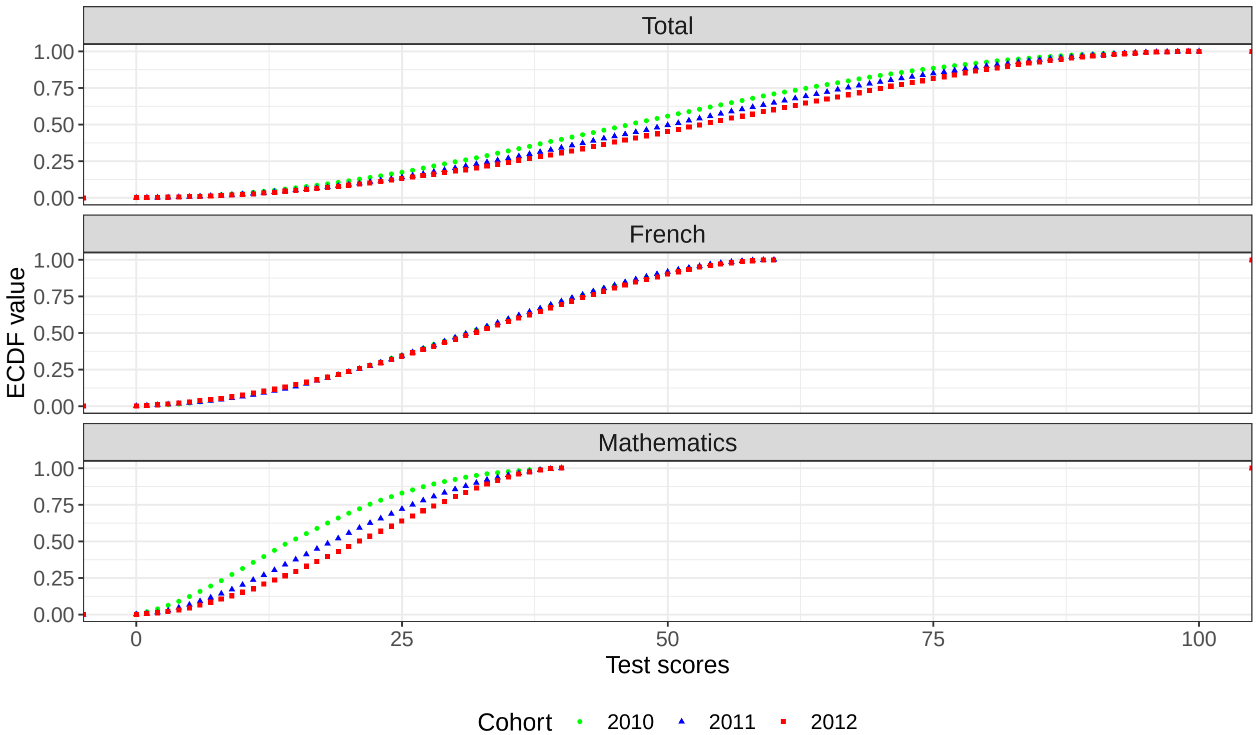

Figure 1: Empirical cumulative distribution functions of test scores

The additional independent variables are the sex and socio-professional category of the parents (it is important to point out that here this is referring either to the father, the mother or the person that is legally first responsible of the pupil, not both) of the pupils. Refer to the first set of statistics in Table 2 for informations about proportions of sex accross cohort taking into account missing values within the data. What can be concluded is that the sex proportion appears to be fair and this is invariant across cohorts except the considerable amount of missing values in the 2012 cohort. Concerning the socio-professional category variable, the proportions by cohort are illustrated by Figure 3. The pattern of the distribution of this variable appears to be stable through the three cohorts. More precisely, the unemployed are those who occupy the most of the part in each cohort (35% in the 2010 cohort, 36% in the 2011 cohort and 37% in the 2012 cohort) whereas the retired are the least numerous of the categories (0.8% in the 2010 cohort and 0.7% in the 2011 and 2012 cohort). Besides, the farmers take only around 1% of proportion in each cohort.

Similarly, a brief Reunion Island-France comparison between the structure of social category (Figure 3). It can be observed that the two portions visually differ for few social categories. For example, in the 2010 cohort, the entrepreneurs are slightly more numerous in Reunion Island in contrast of intermediates that are slighlty fewer in Reunion Island. Also, the retired parents appear to be largely fewer in Reunion Island (20% in France for 7% in Reunion Island)[40]. Moreover, the Reunion Island appears to contain slightly more unemployed parents than France. These few

[40]This observation is to be reconsiered carefully because for the Reunion Island, there are a considerable amount of social category that are unientified or that correspond to missing datas (the « Others » label). This label could contain the remaining retired parents and eventually explain the structure differences outlined for the few social categories.

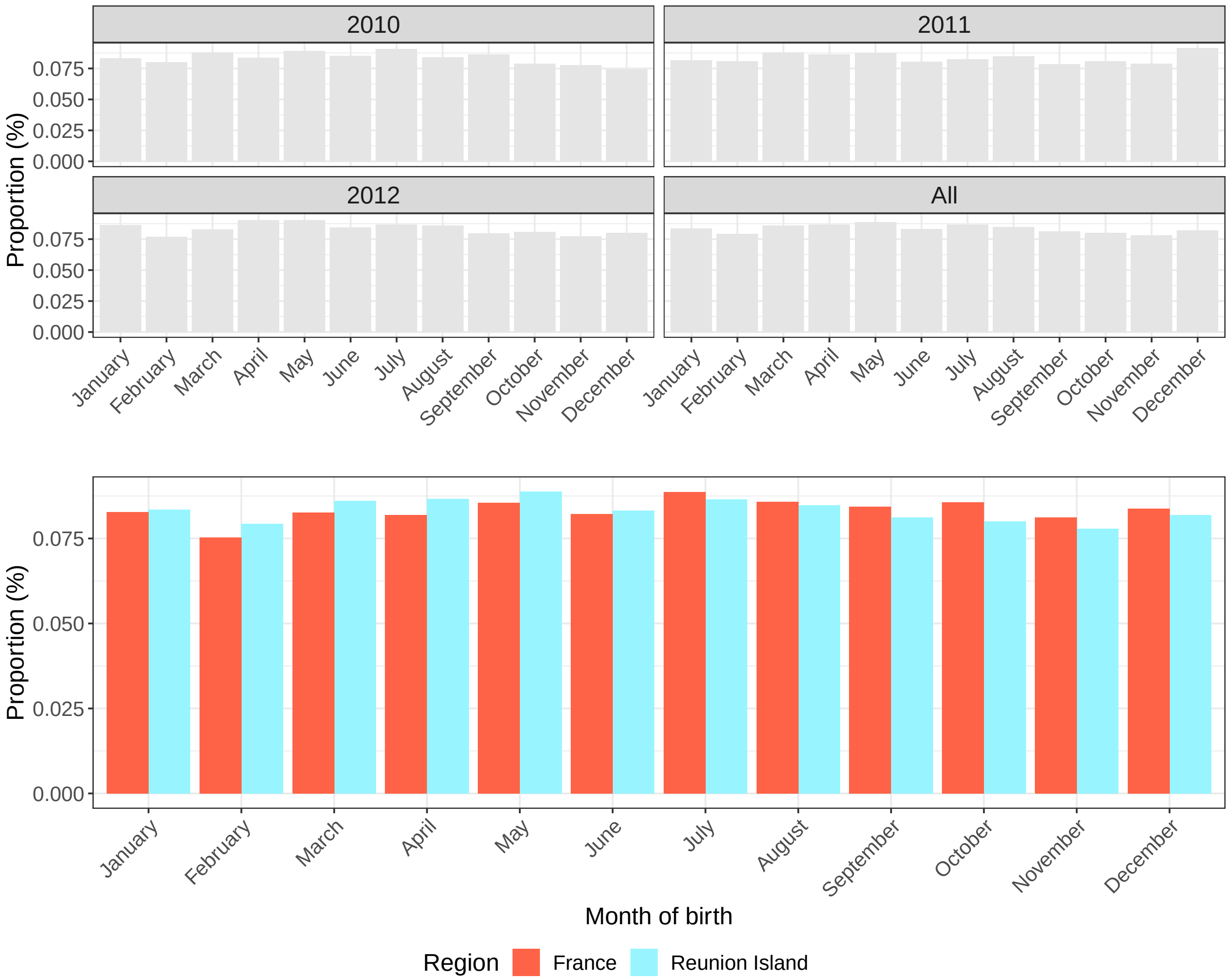

Figure 2: Month of birth proportions

patterns are observed in the three cohorts.

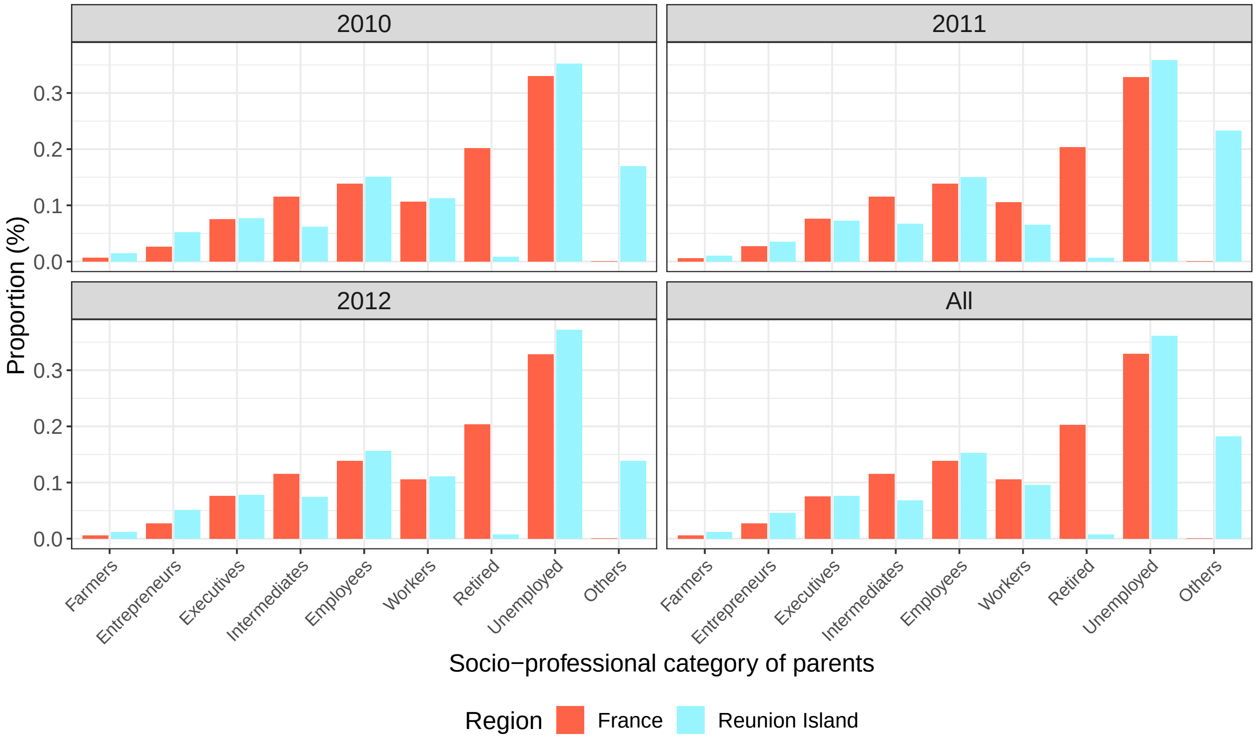

Figure 3: Proportions of socio-professional category of parents by cohort

The logical continuation of the description is to check if some of these institutional features is likely to be highly correlated with the level of pupils. In fact, these are elements that are important to account for in the upcoming econometric analysis if it is, expectingly, the case. In order to highlight this concern, refer to Table 3.

First, concerning the sex variable, males perform worst than females in total scores. Moreover, it can be drawn that this gap is due to the difference in french scores since it (the gap) is greater than the mathematics scores one.[41] This pattern is observed for the three school years.

When focusing on the socio-professional category of parents, it is revealed that the best performance is attributed to the executives children while the worst is attributed to the unemployed's children.[42] Also, if these eight categories are rearranged by the mean of total test scores in descending order, the resulting ranks are stables over the three cohorts.[43] When considering the french and mathematics scores separately, the rankings seems to be non significatively different.[44] More concise, if one

[41]For example, in 2010, the females-males gap in french scores is $34 - 28.9 = 5.1$ points while in mathematics it is only $16.3 - 15.9 = 0.4$ points.

[42]Excluding the « Others » category since it contains unkown categories as well as missing values

[43]In 2010, the ranking, from the highest total test scores is : executives > intermediates > retired > entrepreneurs > employees > farmers > workers > unemployed.
In 2011 and 2012, the intermediates and retired's rank are inverted : executives > retired > entrepreneurs > employees > farmers > workers > unemployed.

[44]With the french scores, the ranking is exactly the same as with the total scores. With the mathematics scores, the ranking is, for the three cohorts, as with the total scores for 2011 and 2012 cohort, as illustrated earlier.

aggregate these social groups in two simple indicators : priviliged[45] or underprivileged[46], we can clearly observe in the third set of rows in Table 3 that the privileged category perform better that the underprivileged category (0 p-value on means comparison tests of the different scores between the underprivileged's children and those of the privileged ones).

Concerning the average scores by school status (fourth set of rows in Table 3), it can be assessed that private schools perform better than public ones. This is true wether with total, french or mathematics scores, wether in 2010, 2011 or 2012 cohort. The private-public gap in total scores turns around 10 points, caused by a gap about 6 points in french and 4 points in mathematics. Finally, for the average scores by priority education network, schools that perform the best are those which do not belong to any network (HEP). This is true regardless of wether the total, french or mathematics scores are considered and wether within the 2010, 2011 or 2012 cohort. For the two remaing categories of school (ECLAIR and RRS), their performances seems to be similar with the RRS schools performing, in general, slightly better than the ECLAIR schools.

Further, about the test scores, the Tables 26 and 27 in the appendix expose respectively the average french and sub-items scores by the previously disciussed institutional features. Concerning the sex variable, girls on average perform better in french sub-items scores than boys do, while such patterns is not clear in the case of maths sub-items scores (lines 1 of the two tables). When we focus to the social category variable (set of lines 2), there appears to be no outstanding difference in the writing scores, while in reading and grammar the executives children perform clearly better. For the same covariate, concerning the maths sub-items scores, we can spot an obvious superiority of the executuve children solely within the calculus scores. No striking difference could be, generally, observed. If one rank these average scores by parent's social categories, similar ranking to the previous ones (with the french and maths scores) are obtained. We now move our focus to the third set of lines of the two tables. Unsurprisingly, there is no case, either within the french sub-items scores or the maths sub-items scores (either in Table 26 or Table 27), in which the underprivileged category do not score lower on average than the privileged category. The others category seems to regroup those who score the lowest of them. Then, about the fourth set of lines of the two mentioned tables, i.e about the school-level priority education network variable, the HEP schools, as in the aggregated cases (french and maths scores desagragation only), perform on average slightly better than the other two categories. Same type of pattern in retrieved in the set of lines 6 of the Tables 26 and 27.

3.2 Econometric framework

For practical reasons, let us expose first the critical notations. The dependent variable (total test score, french test score or mathematics test score in grade 5) is denoted Y. The independent variable of interest (age at test) is represented by A and individual-level independent variables (sex and socio-professional category of parents) will be compiled in the notation X.[47] Also, let the set $(1, A, X)$ be denoted by J. The instrumental variable, called *assigned relative age* and computed as

[45] Assembling executives, entrepreneurs and intermediates

[46] Assembling farmers, employees, workers, retired and unemployed

[47] This simply means X = (sex, spc of parents)

$\frac{(12-m)}{12}$ where m represents the rank of the month of birth[48] is denoted Z. The indexes c and i that will be writen with these notations depending on the necessity represent respectively class and pupil.

3.2.1 Ordinary Least Squares

For illustration purpose, I first establish a linear regression of the test score on age at test and covariates, as illustrated below :

$$Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \nu_i \tag{1}$$

ν_i is the usual error term.

Note that this equation will be estimated separately for the three cohorts.

The parameter of interest is α_1, which capture, under the assumption of exogeneity of A and X[49], the causal effect of age at test[50] on national assessment scores. However, as asserted previously, A is highly expected to be an endogeneous variable, which formally means $E[A_i.\nu_i \mid X_i] \neq 0$.[51] Consequently, the OLS estimator $\widehat{\alpha_1}(OLS) = (J^T J)^{-1} J^T \; Y$ is biaised and non-consistent.

The origin of such endogeneity is the typical existence of an omitted variable plus a selection bias. In other words, some unobservables that have effects on test scores are correlated to the age at test. For example, consider that since the « ability », which determines test scores and is expected to be correlated to the age variable[52], is unobserved, the modeller is constrained to insert it into the error term. This leads then to a correlation between the error term and the independent variables and results in a biased estimate of the parameter of interest α_1.

Additionally, the framework in equation (1) suffers from a selection problem because of the repeaters and those in advance. Indeed, being born late in the year is likely to rise the probability of repeating a year. Also, pupils who have an advance of one year are likely to be born earlier in the year (See Figure 5). Since the proportions of pupils who have a year of advance are very low within the three cohorts (see again Table 2), the main features that causes the bias selection problem is the presence of repeaters and redshirters. Thus, basically, the bias arises with the fact that the age at test variable, within the equation (1) has two effects : a direct effect (α_1) and an effect going through ν_i because retention or redshirting is included in this error term. Note that these two features have distinct correlation with the independent variable – the test score. In fact, retention is negatively correlated with test score as repeaters are very likely to have lower ability and redshirting as positively correlated with test score as redshirtirers have higher maturity compared to their peers. Nethertheless, if repeaters account more than redshirters, which is most likely the case (Bedard and Dhuey (2006), Grenet (2009))[53], $\widehat{\alpha_1}_{OLS}$ is downward biased.

[48] $m = 1$ if the pupil is born on 31st January. Thus, this m accounts for exact day of birth.

[49] No correlation between A and ν and between X and ν

[50] Although, the data don't allow to separate the different age effects, hence this still is a mixed effect of them.

[51] To support this purpose, an endogeneity test of A is performed and presented in the next subsection. That is because before performing the endogeneity test, the establishment of the instrumental variable framework is necessary.

[52] Because of the repeaters / redshirters and the advanced pupils.

[53] I don't have enough data to possess this information, hence I rely on these few papers to make this assumption.

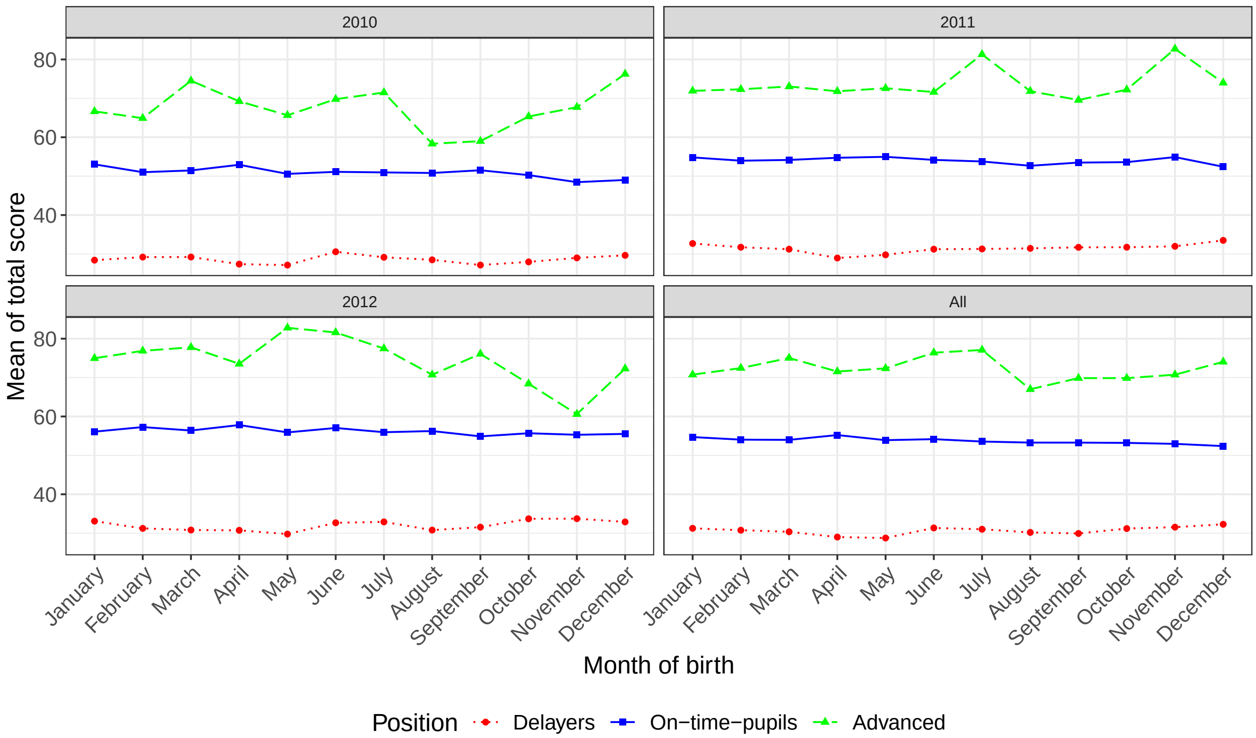

Figure 4: Mean of test score by month of birth and position

As illustration about this issue, see Figure 4. It clearly shows that, either for the three separate cohorts (the first three graphics of Figure 4) or for the three cohorts combined (the last graphic of Figure 4), on average, repeaters / redshirters perform poorer than pupils who are on time. Combined with the observation in Figure 5 which shows that this institutional feature (retention / redshirting) is not a random one (because it appears that the proportion of delayers is positively correlated to the month of birth)[54] ; one could expected the average test scores of the oldests to be decreased in a non randomly way (because the delayers are among the oldest ones within a grade), which is the source of the discussed downward bias in the estimation.

This drives me to an instrumental variable strategy[55], a widely used solution of the omitted variable problem, as presented in the upcoming paragraph.

3.2.2 Two-Stages Least Squares (2SLS)

Suggested by Bedard and Dhuey (2006) and Grenet (2009), the assigned relative age, which reflect the relative[56] age if all the pupils were on time, can be, under some conditions, used as an instrument for the endogenous age at test. This leads to the following simultaneous equations

[54] Once again, this appears to be the case when the cohorts are separated or compiled. Although, the rising pattern of the retention / redshirting rate in 2010 is not as striking as the other three cases.

[55] One alternative solution is to use a exogeneous good proxy of the ability as a regressor, as in Pellizzari and Billari (2012). Given that I don't possess such variable but possess precise month of birth instead, using instrumental variable approach seems to be the most logical solution.

[56] Relative in the sens that it is expressed as a difference in month of birth compared to the theoretical youngest (born in December) within a grade.

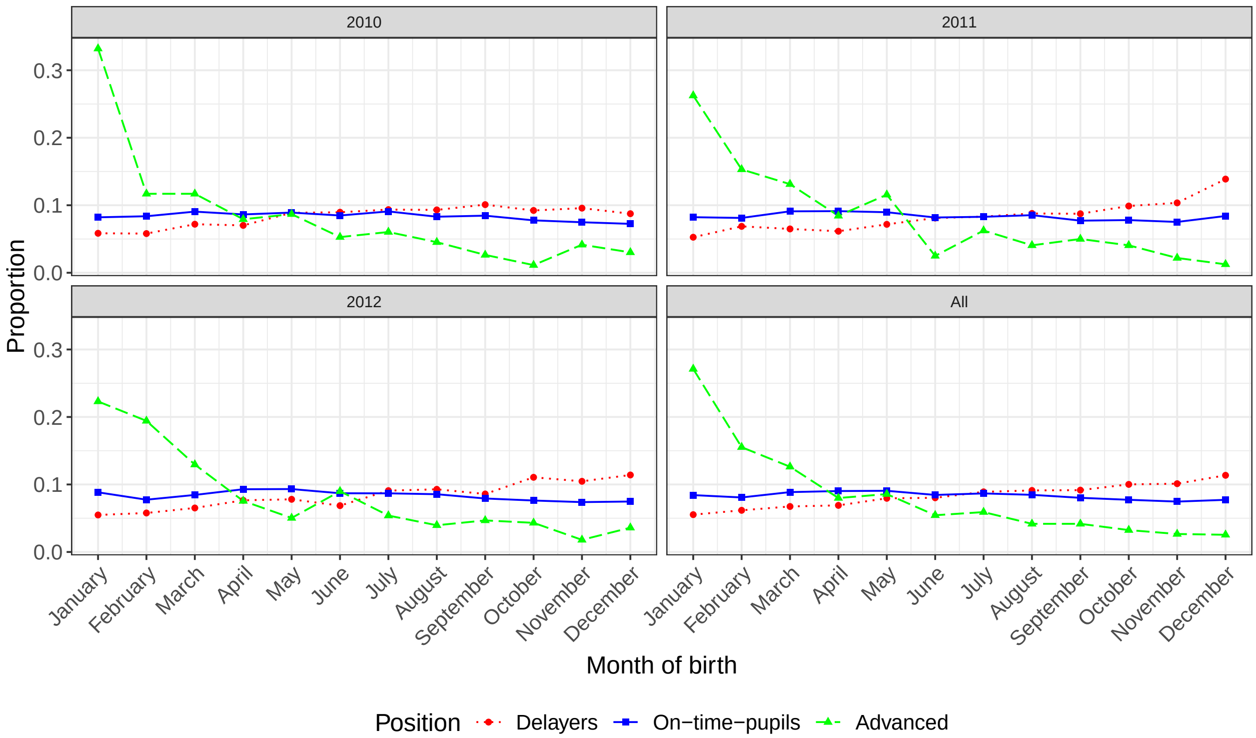

Figure 5: Month of birth, retention/redshirting and advance of pupils in grade 5

model :

$$\begin{cases} Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \nu_i \\ A_i = \gamma_0 + \gamma_1.Z_i + \gamma_2.X_i + \eta_i \end{cases} \quad (2)$$

α_1 still is the parameter of interest in the system above. It is often estimated by 2SLS.[57]

More important, following Imbens and Angrist (1994), Angrist and Imbens (1995) and Angrist, Imbens, and Rubin (1996), there are three conditions to be verified so that the instrument Z is a valid one and that the resulted estimates have a well-defined causal interpretation.

First, Z needs to be randomly assigned. Knowing that Z is a linear transformation of the month of birth, this first condition is equivalent to the random assignation of month of birth. A potential pattern that would drive to the incompletion of this condition is the existence of a seasonality in month of births across the socio-professional categories of parents. In other words, higher-category parents may tend to have their children born in a particular quarter of the year whereas lower-category-parents children in another quarter of the year (Buckles and Hungerman 2013).[58]

In order to investiage this question,I first plot and analyse month of birth proportions across different social categories, then I performed a χ^2 test of comparison of proportions of months of

[57] An alternative is to estimate it by a control function approach, as Hámori (2007) did. I also perform this approach in this paper.

[58] For France, Grenet (2009) found that the month of birth had a seasonality pattern such as those born in April-May had on average the highest earnings (thus the highest socio-professional categories) parents and those born in August had on average the lowest earnings (thus the lowest socio-professional categories) parents.

birth across the different socio-professional categories of parents. For the first procedure, see Figure 6 where each panel illustrates the proportion of month of births within the corresponding socio-professional category sample. After a simple visual inspection first, the farmers appear to be born less often in August and slightly more often in September. These observations have nevertheless to be interpreted carefully because the farmers are around 500 in number for the three cohorts (about 200, 160 and 160 in the three cohorts respectively). A similar pattern can be drawn for the retired parents. In fact, there is a notable hollow at August and this time a peak in November. One can go further by replicating the exercise of Figure 6 by cohort, as illustrated in Figure 7. In this figure, there are even more striking patterns of month of birth seasonality for farmers and retired parents. These visual results are threatening to the validity of the date of birth as instrument, a more reason to perform the χ^2 test.

Concerning the χ^2 test, it is performed to assess whether these proportions (comparing the categories) are equal or not. The former gives credit to the fulfillness of the first condition because it suggests an absence of seasonality of birth by social category. Since the null hypothesis is the equality of the proportions, and the p-value here equals to 0.113, at the 10% level, it can be assessed that the proportion of month of births does not change across the socio-professional categories of parents for the case in which the data is compiled across cohorts. Similarly, one can replicate this exercise for the three different cohorts 2010, 2011 and 2012. The p-values are respectively 0.027, 0.671 and 0.764. The null hypothesis is rejected at the 5% level for the 2010 cohort and is largely non-rejected for the two remaining cohorts. Then, in the 2009-2010 school year, the use of date of birth as instrument for age at test is seriously questionable. In order to have an idea of the bias due to this pattern, I propose to perform alternative regressions in which I exclude from the observations the August borns and in which I aggregate the « Retired » social category in the « Others » social category. If the results are comparables with the framework with the august borns and the unaggregated « Retired » value of parent's the social category, variable, then the bias should not be a grave one. The results of such regressions will be introduced and commented later, since it is not of principal interest in this section.

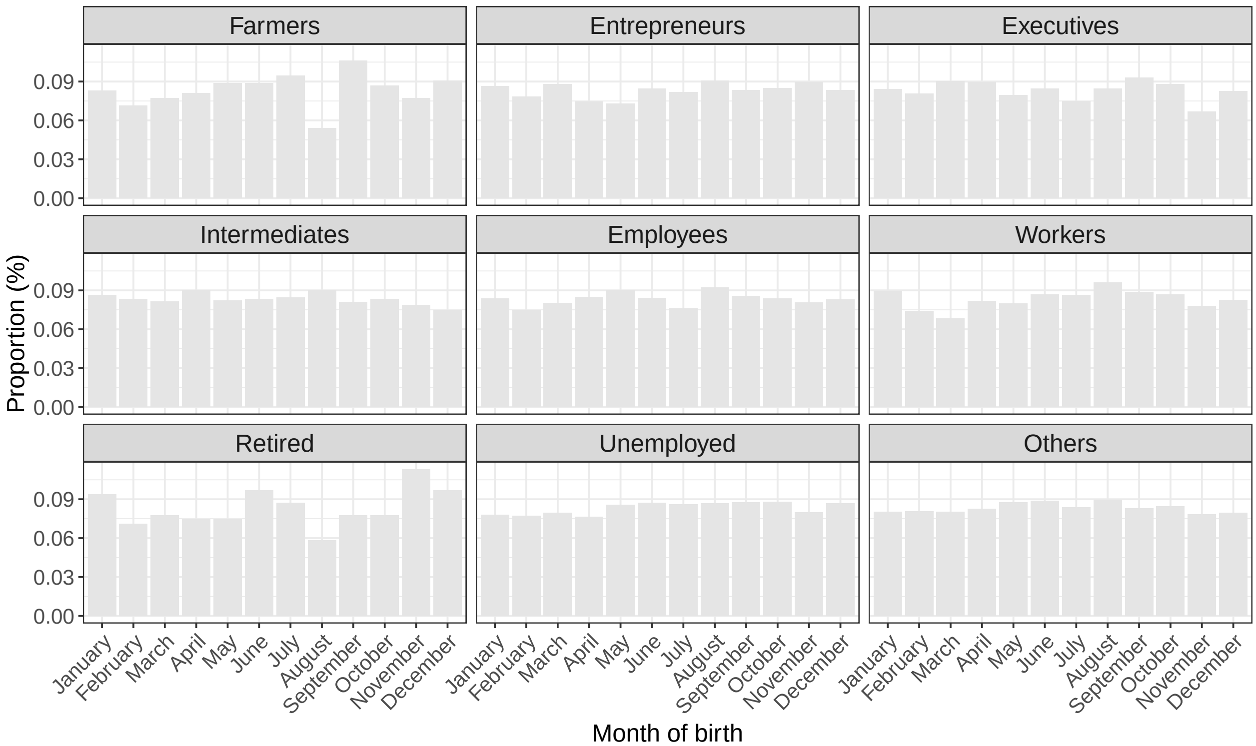

Figure 6: Proportions of month of birth by socio-professional category of parents across cohorts

Second, the instrument Z is required to have a non-zero average effect on the endogeneous variable A. Higher the effect of the assigned relative age on age at test, stronger the instrument. This condition can be investigated quite straightfully with the estimation of the parameter $\gamma 1$ which is the average causal effect of Z on A.[59]

The first two conditions previously described are required to validate the use of the instrument itself. Nevertheless, the assigned relative age is required to satisfy the monotonicity condition in order to the 2SLS estimates to have a well-defined causal interpretation : a *LATE*-type interpretation.[60] This condition requires that the effect of the instrument on age at test does not, in a counterfactual reasoning, have to change in sign over all the pupils. In our case, the effect is positive, this means that there should be no pupils such that the augmentation of Z (*i.e* being born earlier in the year) would lead to a decrease of its age at test.

More precisely, monotonicity can be defined in this framework by the following. Let $A_i(z_k)$ be the counterfactual age at the moment of the test of the indiviual i if this individual was born at a certain date k of the year such that his assigned relative age equals to z_k. It is standard to assume that all $\{A_i(z_k), k = 1, ..., 365\}$ exist for the individual i but only one is observed : the only $A_i(z_k)$ that is observed for i is the one such that k corresponds to the effective date of birth of the individual

[59]This is the first stage regression, the results shown in Table 5 demonstrate that the assigned relative age have a strong effect on the age at test. This is explained by the fact that in each cohort, most of the pupils were one time. Indeed, for these individuals, the assigned relative age variable is a linear transformation of the age at test variable.

[60]*Local Average Treatment Effect (LATE)* in the sens of Imbens and Angrist (1994). I specify in purpose that it is a LATE - « type » because the instrument here is non-binary, as well as the treatment variable – the age.

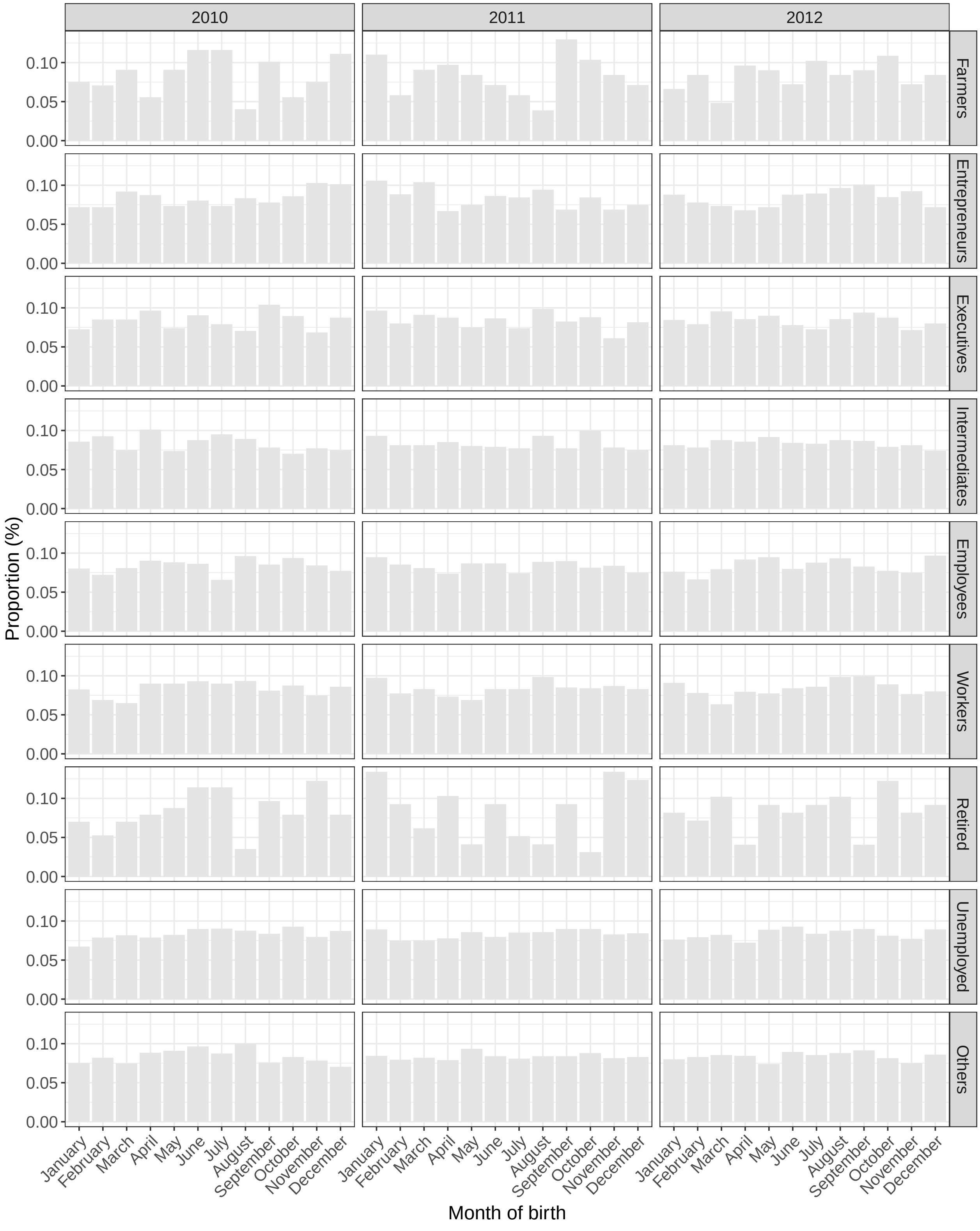

Figure 7: Proportions of month of birth by socio-professional category of parents across and by cohort

i.[61] The monotonicity can be then written as :

$$A_i(z_k) \geq A_i(z_{k-}) \quad \forall i, \quad \forall z_{k-} \geq z_k \quad (or \quad k- \leq k)$$

More comprehensively, this condition states that all individuals should be aged more at test if they were born earlier at the year. Here, it is important to note that **if the date of birth does not counterfactually affect the position, the monotonicity condition is mechanically verified**. The problem is then the effect of date of birth on the position (recall for instance the Figure 5). In fact for a pupil such that the date of birth affect counterfatually the position, monotonicity is mechanically violated. It is very likely that, if one regress the probability of being a delayer on Z and the other covariates, there would be a highly significant effect of Z. This means that there is pupils such that their positions are counterfactually affected by their date of birth.[62]

A suggestion from Angrist and Imbens (1995) to check, in a particular framework of binary instrument and multi-valued treatment, the empirical cumulative distribution functions of the treatment variable separately for the two values of the instrument. They demonstrate that a crossing between these two ECDF means that the monotonicity condition is violated. We can adapt our framework to replicate this empirical investigation : by restricting the data uniquely to those who are born in January or in December and, by taking the binary instrument for age at test which is the month of birth (this variable equals 1 if the pupil is born in January and 0 if born in December) and then by ploting the age at test ECDF's of January and December borns. See Figure 8 for the resulting plot. As we can visually observed in Figure 8, the january-borns and december borns age's ECDF cross multiple times for the three cohorts and even in the case the data is compiled.

More precisely, one can perform stochastic dominance of order 1 tests with the null hypothesis assuming that the january-borns age's ECDF is superior to the december-borns one for all age values.[63]. The computed p-values for the 2010, 2011, 2012 and all cohorts equal all zero. Thus, the stochastic dominance of oerder one of the january-borns age ecdf is rejected (meaning a violation of the monotonicity).

Similar evidences are presented in Aliprantis (2012) and Fiorini, Stevens, and others (2014) about the violation of the monotonicity condition. Any interesting and precise causal interpretation seems to be impossible given the violation of the condition. However, we will see in Section 4 what can be at best identified with 2SLS in the present framework.

3.2.3 Reduced form estimation

The reduced form equation is obtained after inserting the first stage regression equation (the regression of A on Z and X) into the structural equation (the regression of Y on A and X). Therefore,

[61] In other words, $A_i(z_k)$ is observed only if i is born at the day k.

[62] For instance, Alet, Bonnal, and Favard (2013) are partially performing this idea.

[63] The corresponding methodological paper for the test is Barret and Donald (2003) while an application to an age and educational outcomes framework can be found in Fiorini, Stevens, and others (2014).

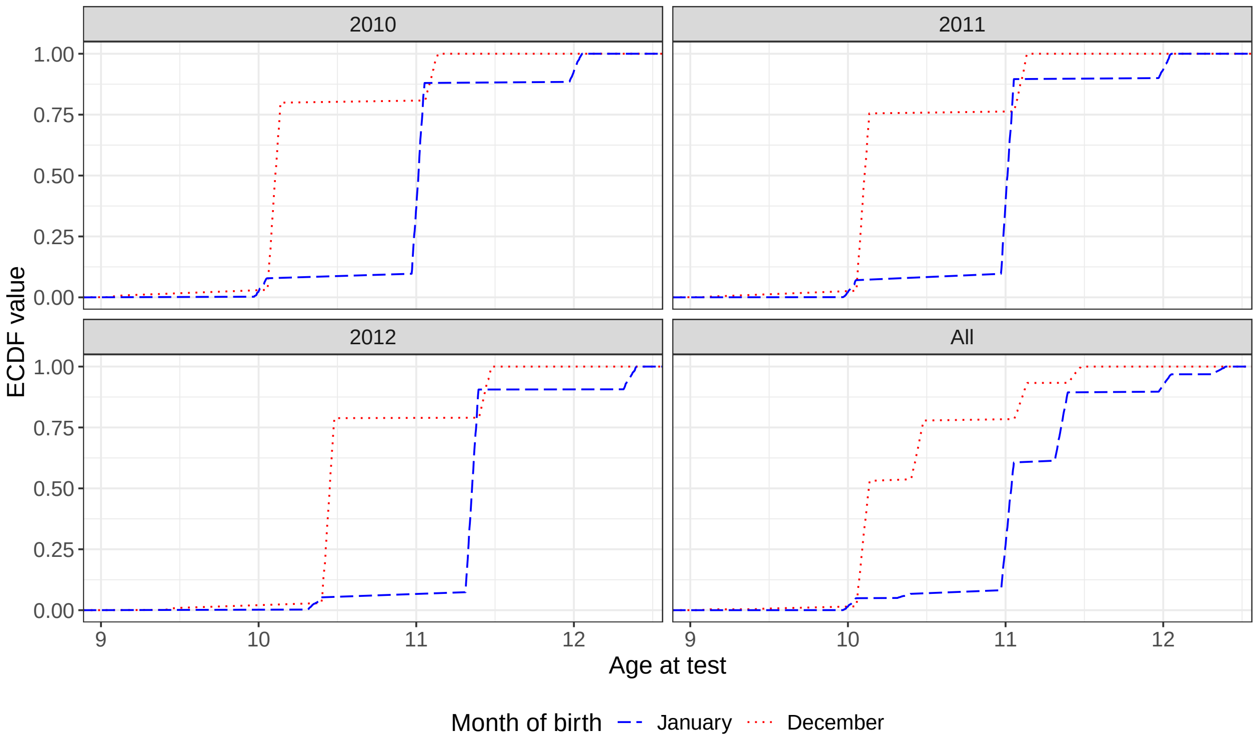

Figure 8: Empirical cumulative distribution function of the age at test by month of birth (January / December restricted sample)

the reduced-form is a regression of Y on Z and X, as shown in the following :

$$Y_i = \delta_0 + \delta_1.Z_i + \delta_2.X_i + \epsilon_i \quad (3)$$
$$\text{with} \quad \epsilon_i \equiv (\alpha_0 + \alpha_1).\eta_i + \nu_i$$

Note that the resulting error term is composed by the error terms of the first stage (η_i) and the structural equation (ν_i). The equation (3) is used to estimate an « *intention-to-treat-(type of)-effect* » of the assigned relative age on the test scores : δ_1. Furthermore, since Z is a linear transformation of the month of birth variable, the reduced form equation can be taken for estimating the effect of month of birth on test scores.

The fundamental difference between estimating α_1 and estimating γ_1 is that the former yields to the estimated effect of age net of the effect of retention / redshirting and acceleration whereas the latter captures at the same time the effect of not being one time (i.e the effect of retention / redshirting and acceleration). More detailed explanation is provided by Bedard and Dhuey (2006) in which they highlight that if the retention or redshirting had positive effect on test scores, then the reduced form estimate (here $\widehat{\gamma_1}$) should be lower than the 2SLS estimate (here $\widehat{\alpha_1}$. Another implication is that higher the proportion of redshirters / repeaters, lower the reduced form estimate compared to the 2SLS estimate. Recall the well-known link between the 2SLS estimation shown in the precedent paragraph and the reduced form estimation is that

$$\widehat{\alpha_1}(IV) = \frac{\widehat{\delta_1}}{\widehat{\gamma_1}}$$

3.2.4 Control function approach

Considering the evidence of violation of the monotonicity condition in the 2SLS approach, I propose a control function approach to the problem of measuring the effect of age on educational performances. The method, its relative advantages and its limits are well and compactly exposed by Wooldridge (2015). A paper that also seems worth mentionning is a rare one that use this present method within the study of the effect of starting age on academic performance in Hungary : Hámori (2007). I hence mainly structure the present paragraph based on these papers. First, note that the control function approach is a method used to account for endogeneity within a simultaneous equation model.[64] The key idea, and at the same time one of its limitation is that it impose a linear relationship between the error terms of the structural equation and the first stage equation, respectively ν and η (see Equation (2))[65] :

$$\nu = coef.\eta + error \tag{4}$$

with the additional construction assumption

$$E[\eta.error] = 0$$

Note that since ν and η are not correlated either with Z or X, $error$ is thus not correlated either with Z, X or A. When putting the linear relationship presented in Equation (4) within the structural equation (1), one obtains

$$Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \alpha_3.\eta_i + error_i$$

where η can be thought as a *proxy of the part of* ν *that is correlated with A*. In our framework, η could be then interpreted as a proxy for the ability that is not observed.

Since η_i is unknown, the control function process will then consist of, first, noticing that

$$\eta_i = A_i - (\gamma_0 + \gamma_1.Z_i + \gamma_2.X_i)$$

so $\widehat{\eta_i}$ is computable from the first stage equation. Then, the following equation, which could be called the control function (CF) equation, can be estimated :

$$Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \alpha_3.\widehat{\eta_i} + error_i \tag{5}$$

[64] For reader's ease, let us rewrite the simultaneous equation model here

$$\begin{cases} Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \nu_i \\ A_i = \gamma_0 + \gamma_1.Z_i + \gamma_2.X_i + \eta_i \end{cases}$$

[65] The following assumptions are implicitely maintained : (i) exogeneity of Z and X and (ii) non-zero average effect of Z on A

Within the described procedure, the estimation of α_1 should capture, in the langage of potential framework, an average treatment effect (ATE). However, recall that the underlying assumption of this identification – the homogeneity of the effect of age for all pupils – is known to be a very strong one and is most likely to be violated. For example, the difference in ability could differ at different ages, meaning that the effect of age is different for different children . If this information is known by the parents and if they decide to enroll on time, earlier or later their children based on this, this will cause another bias, typically because the age at the test of children would be non random in another way.[66].Hence, an extension of the Equation (5) is needed when one need to account for this heterogeneity of the age effect, which is more realistic. The basic reasoning that will lead to the control function equation that account for such heterogeneity will be described now.

Let us first modelize the presence of the heterogeneity of the effect of age on test scores :

$$Y_i = \alpha_0 + \alpha_{1i}.A + \alpha_2.X_i + \nu_i$$

Note the index on α_{1i} which represents the individual effect of the age on test scores. This can always be decomposed into two terms : the mean effect plus a individual random deviation :

$$\begin{cases} \alpha_{1i} = \alpha_1 + error_{2i} \\ \text{with} \\ E[error_{2i}] = 0 \end{cases}$$

Inserting this expression into the structural equation yields

$$Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \nu_i + A.error_{2i}$$

such that $\nu_i + A.error2_i$ is unobserved. Note here that formally, the problem is that the unobserved part of the structural equation is mechanically correlated with the age variable because this latter is belong to this unobserved component.

Now, similarly as within the homogeneity case, given the first stage of Equation (2), the key fo the approach is to assume a linear relationship between ν_i and η_i and between $error_2$ and ν_i i.e

$$\begin{cases} E[\nu \mid \eta] = \alpha_3.\eta \\ E[error_2] = \alpha_4.\eta \end{cases}$$

Thus, taking it into consideration :

$$E[Y \mid Z, X, A] = E[Y \mid Z, X, A, \eta] = \alpha_0 + \alpha_1.A_i + \alpha_2.X + coef_2.\eta + coef_3.\eta.A_i$$

[66] They would enroll their child at the age at which effect of age is at the maximum. This is sometimes called « selection on gains » in the litterature. The combination of heterogeneity of the effect and the selection on gains form what Heckman, Urzua, and Vytlacil (2006) call « essential heterogeneity ».

Translating this population regression into its sample analogous leads, similarly as the homogeneity case[67], to the following estimable sample regression function :

$$Y_i = \alpha_0 + \alpha_1.A_i + \alpha_2.X_i + \alpha_3.\widehat{\eta}_i + \alpha_4.\widehat{\eta}_i.A_i + \nu^h \tag{6}$$

The inclusion of the terms $\hat{\eta}_i$ and $\hat{\eta}_i.A_i$ control respectively for the endogeneity of age discussed since the beginning of this paper and the heterogeneity of age effect. In other words, respectively, the omitted variable of ability and its variation across different age levels are taken into account. Moreover, recall that in contrast of the 2SLS procedure under monotonicity, the present control function approach does not require monotonicity and estimate consistently the average treatment effect.

The control function approach possesses at least two interesting advantages : the possibility to test for the highly suspected endogeneity of age via a significativity test of the coefficients α_3 and α_4 ; and additional informations on the selection bias via the sign of these coefficients.

3.2.5 Regression discontinuity design for comparison purpose

Last, as some papers perform regression-discontinuity design (RDD henceforth) to adress similar questions as in the present paper (see for example Kaila (2017) for Finland or Dobkin and Ferreira (2010) for Texas and California), I propose to do as well for compraison purposes (with cross sectional results). The framework construction will be described before econometric modelizations.

3.2.5.1 The framework

Recall that we are interested in the effect of being one year older at the moment of examinations on test scores. In addition, the cutoff rule described earlier imply that a pupil born between January, 1st, 1999 and December, 31st, 1999 should be theoretically observed in the 2010 cohort. In the same way, a pupil born between January, 1st, 2000 and December, 31st, 2000 should be theoretically observed in the 2011 cohort (and consequently take the test one year later than if he was observed in the 2010 cohort, although he/she have comparable age in the case he would instead took the test in 2010). Thus, an individual born in January 2000 instead of December 1999 will be theoretically about one year older at the test because of the cutoff rule. This difference in age at test is highly excpected to have an effect on test scores, and that where lies the motivation of the present RDD. To avoid misunderstandings, in this particular framework, instead of considering one a cohort in which analysises are performed as in the previous models, one should consider a *year of birth* instead. Intuitively, I propose to compare pupils born in 2000 instead of 1999. The reason behind the fact that solely the 1999 and 2000 borns were retained is found in the Table 1. In fact, these two year of birth are the only two in which all the three positions (Delayers, On-time pupils and Advanced pupils) are observed.

In order to perform the analysis, I first create a data containing pupils born in 1999 and 2000, establishing that the « *cutoff date* » is the January, 1st, 2000. ˆ [Instead of December, 31st, 1999

[67] By recovering $\hat{\eta}$ from the first stage equation.

because of modelisation necessity, but the idea of a cutoff date is the same.] Then I create the necessary variables for the RDD. First, I compute $dist$, the distance (in days) between the date of birth and the cutoff date, such that a pupil having $dist = 0$ is born in January, 1^{st}, 2000. This first variable is thus centered on the cutoff, implying that a pupil born at or after the cutoff date will have $dist_i \geq 0$ and a pupil born strictly before the cutoff date will have $dist_i < 0$. Second, I define the binary variable old indicating if a pupil is born in or after the cutoff (i.e born in January, 1^{st}, 2000 or later and theoretically be in the 2011 cohort[68]). Formally, that means

$$old_i = \begin{cases} 1 & if \quad dist_i \geq 0 \\ 0 & if \quad dist_i < 0 \end{cases}$$

Third, note that the independent variable of interest is still the age at test of the pupils. Matta et al. (2016) uses a quite similar framework.[69] The standard RDD is often presented with a binary independent variable of interest. This implementation do not make much sens in the present case. The precise explanation will be given just above since some basic vocabularies should be introduced first.

Within this part of the study, I will generally follow most of these following praoctical papers guidelines : Imbens and Lemieux (2008), Lee and Lemieux (2010) and Cattaneo, Idrobo, and Titiunik (2017a). Let us now borrow some of the treatment effect literature vocabularies to better define the variables above and the parameters of interest. The treatment variable is the age at the moment of the test. Note that if all pupils were on time, $dist$ (and thus old) would be perfectly colinear with the age. In fact, if the cutoff rule were strictly followed, the age at the test in year t of pupil i would be the time length of the intervall $[birthday \quad ; \quad test\ date_t]$. Because this is not the case (there are repeaters / redshirters and pupils in advance), old will be distinguished as the assignment (to treatment or control group) variable. **Here the fact that a binary treatment variable can be explained** : this is because, if the hypothetical treatment variable w take the value of 1 for an individual that is observed in the 2010 cohort, this variable should take the value of 0 if the pupils do not belong to the 2010 cohort. This definition of the 0 value of w is problematic bacause, two other possibilites of observed cohorts can occur instead of a unique one : if the individual is in advanced and if the individual is a delayer. Moreover, not all individuals who are assigned to the 2011 cohort (i.e individuals with $old = 1$) take effectively the examination in 2011 and not all individuals who are assigned to the 2010 cohort (i.e individuals with $old = 0$) take effectively the examination in 2010. We say that there is imperfect compliance (not all individuals comply to the cutoff rule).[70] Consequently, this is a case in which a fuzzy regression discontinuity (FRD henceforth) should be implemented to measure the causal effect of age at test on test scores. In such a framework, a sharp jump in the average age at the cutoff ($dist = 0$) should be observed. It is this sharp jump that is hihgly expected to affect all else equal the test scores. From a perspective of the present study, this is the case because of the decrasing (with later month of birth) proportion

[68] One vocabulary that one can use is that such a pupil is « assigned » to the 2011 cohort, i.e assigned to take the test on January, 20^{th}, 2011.

[69] See also Smith (2009) who uses multiple cutoff dates instead.

[70] More precisely, this is a case of two-sided non-compliance because some individuals assigned to the 2010 cohort are observed in the 2011 cohort (the repeaters / redshirters) and some individuals assigned to the 2011 cohort are observed in the 2010 cohort (pupils having one year of advance).

of delayers (compared to the within month of birth frequency of the three positions) within a month of birth. Refer to Figure 9 to observe such a pattern at the cutoff date.

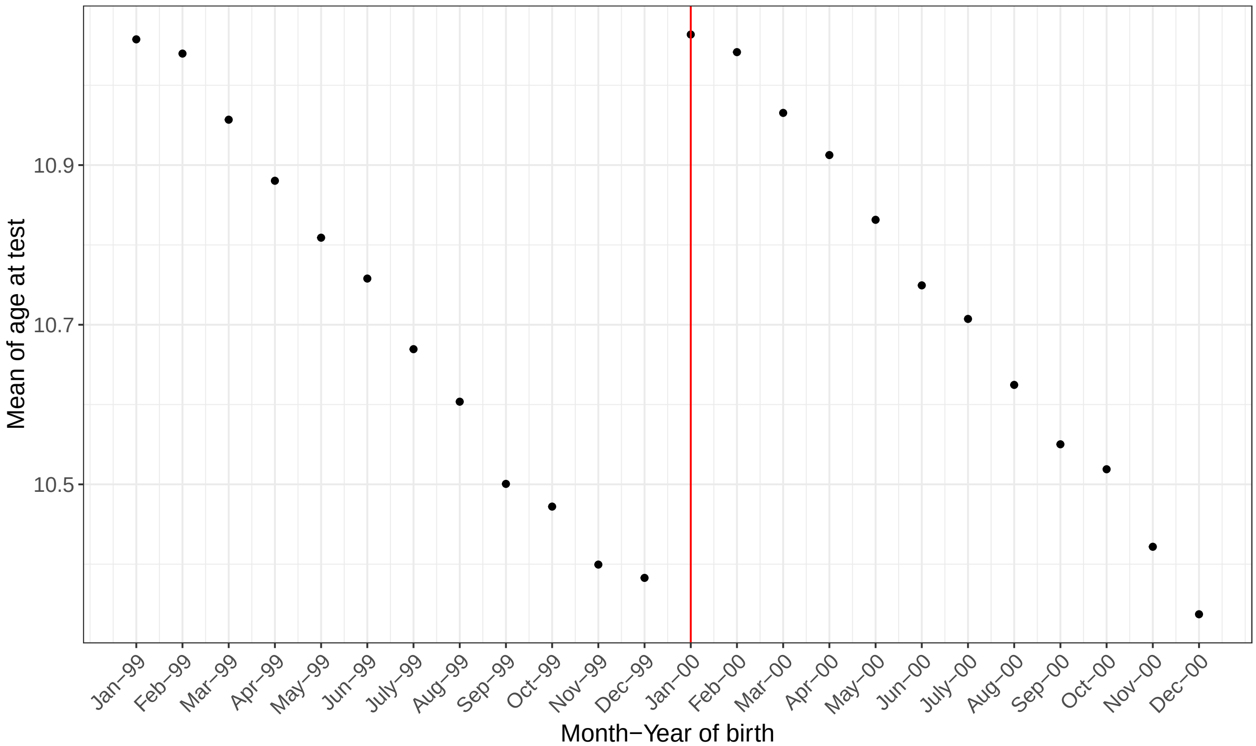

Figure 9: Average age at test by month-year of birth for 1999 and 2000 borns

Two parameters of interest can be defined now. I first use simple sentences here and only formalize them in the econometric modelization itself for conciseness. First, an intention-to-treat can be estimated. It is the effect of the assignment – old – (here being born before or after the January, 31st, 2000) on the outcome of interest Y. Second, under well known assumptions, a LATE-type of estimate can be derived : the effect of being one year older at the moment of the test for the subpopulation that would counterfactually comply to the assignment..[71] Unfortunately, fllowing similar evidences presented in the 2SLS framework, the monotonicity here is likely to be violated. More precisely, the monotonicity assumption, in the present case, requires that all individuals born in 2000 ($old = 1$) will be aged more than in the **counterfactual case** of being born in 1999 instead. Thus, due to the presence of individuals who verify $A_i(old_i = 1) < A_i(old_i = 0)$[72] where $A_i(old_i)$, $old_i = \{1, 2\}$ designs the counterfactual ages corresponding respectively to the world in which i would have $old_i = 0$ (i.e would have been born in 1999) and the world in which the same i would have $old_i = 1$ (i.e would have been born in 2000)[73] ; the monotonicity is mechanically violated. We can replicate the exercise in 8 to assess this violation. Following the idea of the locality around the cutoff, we will not use all birthdates but some restricted to be within a symetrical bandwidth around the cutoff. For instance, see that in the Figure 10, we can observe the two lines within a panel crossing (giving an evidence of violation of the monotonicity). Given this issue, an attempt to

[71] The compliers, is the sens of Imbens and Angrist (1994)
[72] In simple worlds, if some individual's age is counterfactually affected by his / her year of birth.
[73] Recall that the two values of $A()$ exist, the only one that is observed is the one that match the realised value of old_i.

discuss the identified estimand will be provided in the within the Section 4, since the classical LATE identification is not achieved here.

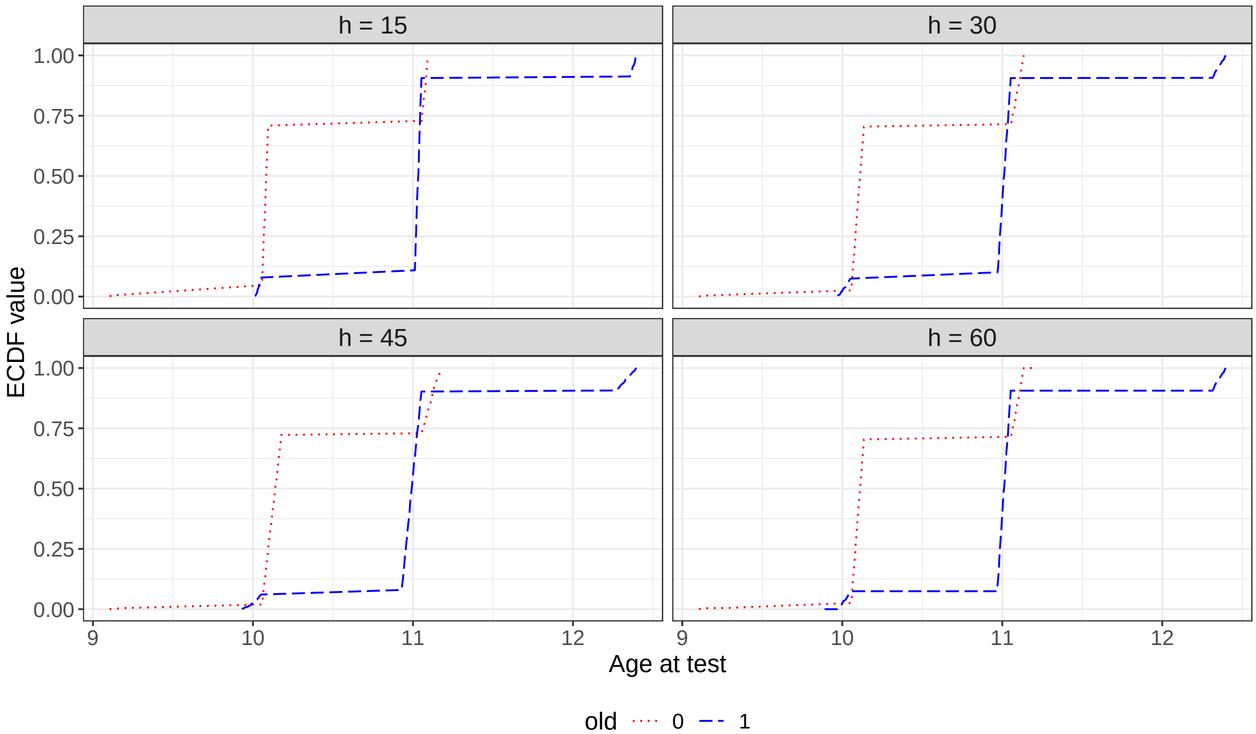

Figure 10: ECDF of age at test for each year of birth

Computing such a effect is done by performing a 2SLS estimation on two regression equations : the first stage, which consists in regressing the absolute age on old and $f_{(A)}(dist)$ with $f_{(A)}()$ being a smooth polynomial function of $dist$ and its interaction with old ; and the structural equation which consists in regressing Y on the age and $f(dist)$. $f()$ and $f_{(A)}()$ are noted differently just because of the difference in their regression coefficients.

Last, I will solely focus on the french score as outcome of interest because fo the systematic rise[74] in the mathematics scores from 2010 to 2012. This means that if I take the total score as outcome in the regressions, I will capture that systematic rise in addition to the causal effect of the age (a bias to be ruled out).

3.2.5.2 Validity checks

There are several validity threats to the RDD framework. First, it is the presence of a precise manipulation of the birthdate at the cutoff. More precisely, the running variable, here the birthdate (computed in the form of $dist$), should not contains a discontinuity at the cutoff because if such a jump is correlated to the test scores[75], then one would not recover the unique effect of age on test scores. This requirement is similar to the condition of exogeneity of the date of birth discussed

[74] Due to the difference in the structure of the examinations discussed earlier.

[75] If, more privileged parents precisely aim to give birth at January 2000 instead of December 1999 for example, knowing that having more privileged parents is likely to increase test scores.

in the 2SLS cross section model. If no such precise manipulation is present, then the density of $dist$ should be continuous at the cutoff (at $dist = 0$). This is checked via the McCrary (2008) density test in which the null hypothesis is the absence of manipulation (i.e the continuity of the density of $dist$ at $dist = 0$). Let us visualize the test result. An illustration of the estimated densities of $dist$ is reported in Figure 11. It can be easily assesed that the density is unfortunately discontinuous at the cutoff, in addition of the calculated p-value of 0.005 (the null hypothesis of continuity is strongly rejected). The reason of this unexpected feature is likely to be, after scrutiny of the position proportion patterns with different bandwidhs around the cutoff, the jump one could observe at Figure @ ref(fig:rdadvanced) for the advanced pupils. This particular plot has to be read carefully because of the y axis that represent non-intuitively the proportions of 1999 borns (among the 1999 and 2000 borns) within a type of position. At the limit of 110 days around the cutoff, the proportion of 1999 born ($old = 0$) pupils among the advanced ones were more or less around 10% while it jumps, for a 115 days bandwidth, to 20%. Since the density test use the full data at the first place, this jump is likely the reason of the reject of the null hypothesis of density continuity of the birdthdate at either side of the cutoff.

Hence, as this feature is likely to threat the validity of the framework, I pick a more logical option : I removed children that were in advanced compared to their theoretical cohort. Note that the consequent selection caused by this operation is probably not collapsing the identification because of the few frequency of pupils that are in advance. Once these pupils removed from the data, the density test appears to be more convincing to use perform a RDD. In fact, as demonstrated by Figure 12, the density is visually continuous in addition of the p-value that not permit to reject the null hypothesis of continuity anymore.

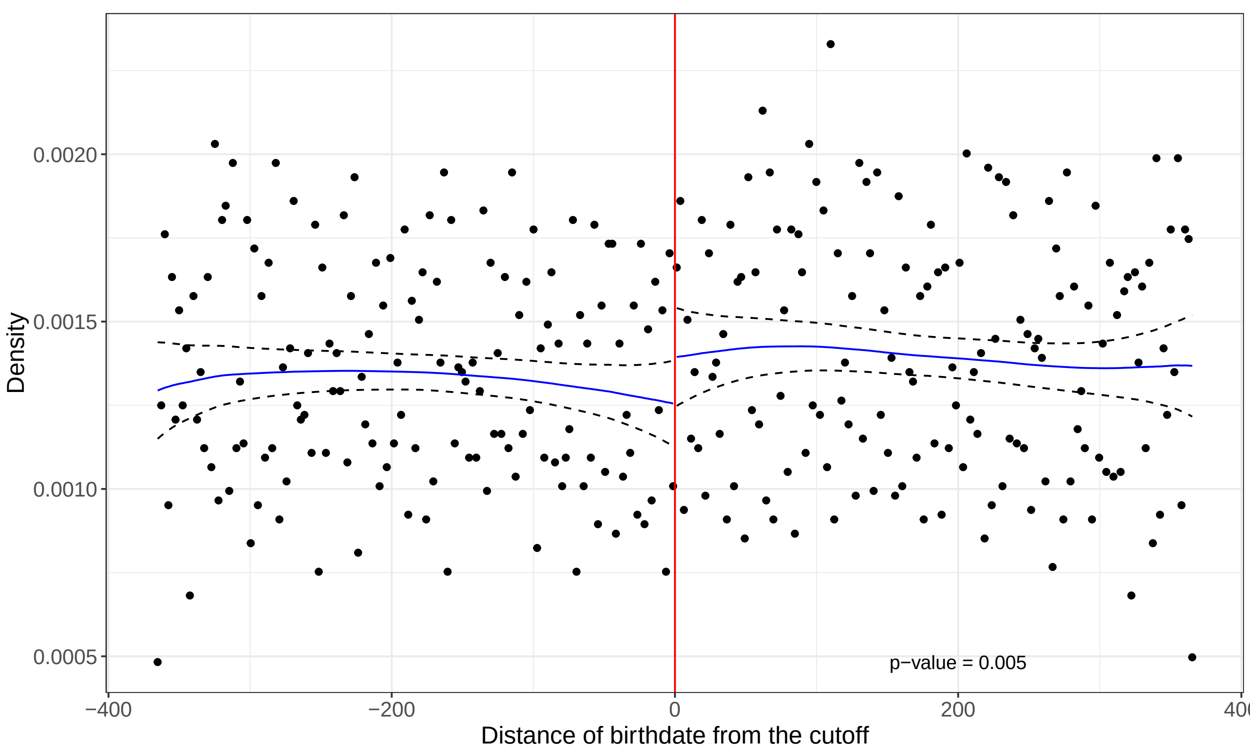

Figure 11: McCrary (2008) density continuity test of birthdates (all 1999 and 2000 borns)

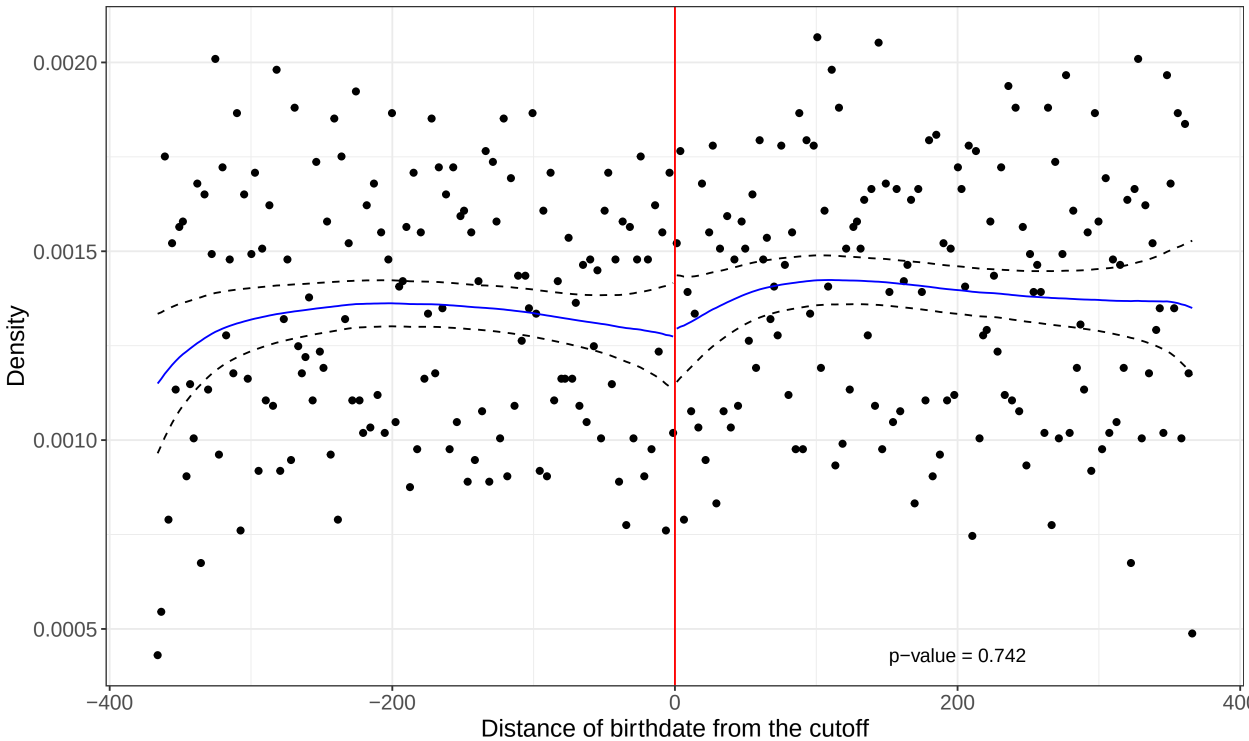

Figure 12: McCrary (2008) density continuity test of birthdates (1999 and 2000 borns without advanced ones)

Second, another standard check in RDD is to inspect if the baseline covariates (for instance, the sex and social category variables here) are « locally » balanced on either side of the cutoff.[76] It is important to check this feature to assume that pupils are « comparables » in terms of observables at either side of the cutoff. There are two practices to be implemented to do such inspection. The first practice is a graphical analysis that consists of plotting birthdate-binned proportions of the covariates for the two sides (born in 1999 and born in 2000). An observation of an obvious jump in the binned proportions at the cutoff would probably indicate that those born just after the cutoff are not locally comparable to those born just before the cutoff of birthdate. If these observables are likely to affect the dependent variable, this would cause a bias. The Figure 13 illustrate the results of the exercise for 10-day bins. The largest difference between proportions in the first 10-day bins on the left and proportions in the first 10-day bins on the right is observed for the grouped social category labelled « Underprivileged ».[77] When numerically checked, this difference is about 6 points of percentage, hence it seems not very worrying. The other exercise to perform in order to check the covariates balance is to the intention-to-treat regression as described earlier and take, instead of test scores, the covariates as dependent variable. I performed these regressions considering 1 to 4 polyomial order of $f()$ (See the next subsection to visualize the form of right hand of the equations). To ensure the local aspect of the regressions, I restrain , in this case, the regression sample to a bandwith of 30 days around the cutoff. The results are

[76] Lee and Lemieux (2010)

[77] The « Males » sex value is not plotted because it is deduced from the « Females » plot. Instead, The grouped social category can take three possible values : « Underprivileged », « Privileged » and « Others ». That is why two values are plotted for the grouped social category covariate.

presented in Table 20 in the Appendix. From this table, one can obseve that any evidence of discontinuities in covariates proportions at the cutoff are present for the specification where $f()$ is a 3 or 4 polynomial order of $dist$ and its interaction with old (column (7) and (8) respectively. Eventhough, this is not of a worry because of the very low R-squared, meaning that only 0.5% of the variation in the probability of being a child of privileged parents is explained by the variation in either the child is born on either side of the cutoff.[78] In addition, it is generally not recommended in regression discontinuity designs (Gelman and Imbens 2019). And some evidences in estimates sensitivity analysises in Section 4 seems to support that we should consider polynomials of at max, order 2.

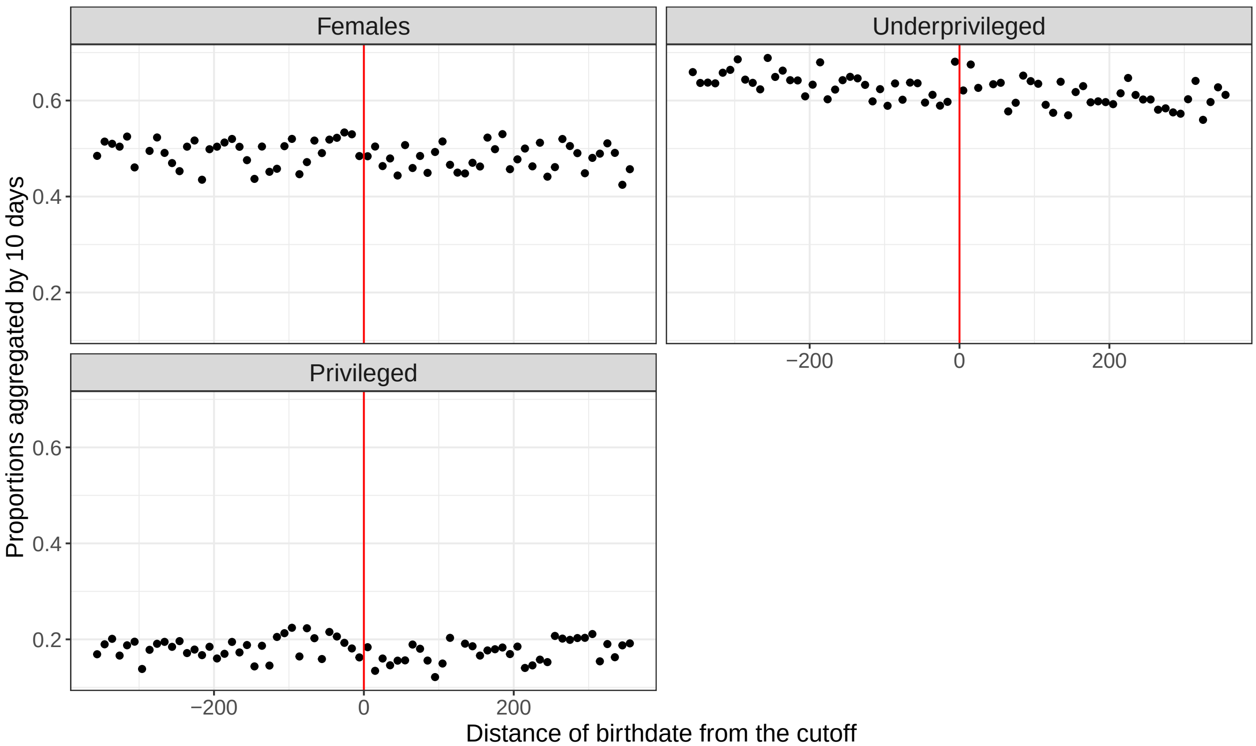

Figure 13: Covariates balance on either side of the cutoff

Third and last, one should check the presence of cutoff effects on the dependent variable at other dates than the January, 1st, 2000 (i.e at other values of $dist$). Proposed by Imbens and Lemieux (2008), I perform the same regressions as described for the pseudo-outcomes (covariates) but instead of considering a cutoff at $dist = 0$ for the whole data, I divide the data in two by the value of old and I test if either there is a cutoff effect on test scores at the medians of $dist$ on the two sides. The median on the left of the cutoff (for $dist < 0$) equals to -183 (corresponding to the July, 02nd of 1999 birthdate) while the one one the right equals to 181 (corresponding to the July, 30th of 2000 birthdate). The Figure 14 illustrates these medians with the average french test scores bined by 10 birthdays. Note that no worrying jump is observed at the left median while a slight visual jump in the average score could be at the right median. This latter requires more precise investigation, i.e the regressions with pseudo cutoff mentionned above. Also note the clear jump in the average test scores at the original cutoff $dist = 0$. The regression results with the medians

[78] Taking higher bandwidth does not change these features.

taken as cutoffs on either side of $dist = 0$ are presented in Table 16 in the appendix. As the figure suggested it, the regressions yield to a significant at the 5% level age effect at the median on the right side (column (6)) for the 2 polynomial order specification within the whole sample (-0.133) and the male pupils (-0.166). We could think that this is linked to the exclusion of the advanced pupils. If one inserts them within the sample and replicates the Table 16 exercise, the results in Table 17 indicates however significant but with much lower magnitude effect at the right median within the whole sample (-0.099) but similar effect as earlier within the male pupils (-0.154). Hence, regression results at the original cutoff should be considered carefully for male pupils because of this feature : a downard bias is possible.

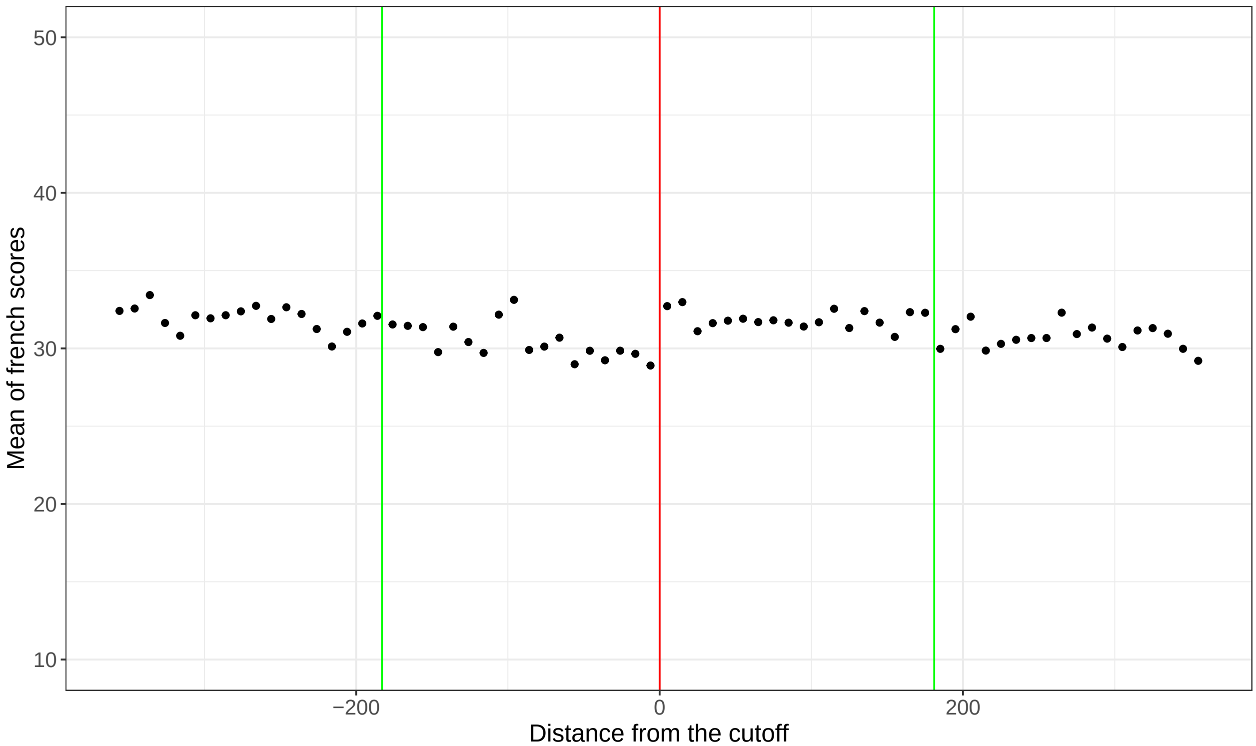

Figure 14: Medians of birthdates distrubtions on either side of the cutoff and average french scores (10 days bins)

3.2.5.3 Regression models

In this part, I will present concisely the form of the main RDD equations performed in this paper. Considering all the explanations in the presentation of the framework, the equations take the following forms. First, the Fuzzy RD is, as described earlier, a 2SLS estimation, for the p polynomial order considered[79], of the equations

$$\begin{cases} Y_i = \alpha_0^{RD} + \alpha_1^{RD}.A_i + f(dist_i) + \nu_i^{RD} \\ A_i = \gamma_0^{RD} + \gamma_1^{RD}.old_i + f_{(A)}(dist_i) + \eta_i^{RD} \end{cases} \tag{7}$$

[79] $p = \{1, 2, 3, 4\}$

with

$$f(dist_i) = \sum_{k=1}^{p} \pi_k.dist_i^k + \sum_{k=1}^{p} \pi_{p+k}.old_i.dist_i^k$$

and

$$f_{(A)}(dist_i) = \sum_{k=1}^{p} \pi_{(A)k}.dist_i^k + \sum_{k=1}^{p} \pi_{(A)p+k}.old_i.dist_i^k$$

On the other hand, the RDD reduced form estimations (itention-to-treat) are based, for the p polynomial order considered, on the equations

$$Y_i = \delta_0^{RD} + \delta_1^{RD}.old_i + f_{(RF)}(dist_i) + \epsilon_i^{RD} \tag{8}$$

with

$$f_{(RF)}(dist_i) = \sum_{k=1}^{p} \pi_{(RF)k}.dist_i^k + \sum_{k=1}^{p} \pi_{(RF)p+k}.old_i.dist_i^k$$

The coefficients of interest are α_1^{RD} for the Fuzzy RD and δ_1^{RD} for the reduced form. Theses regressions will be performed using restricted data such as the restricting variable is $dist$, taking individuals at a certain distance of birthdate (left and right symetrically) from $dist = 0$. More precisely, in RDD vocabulary, we choose only individuals satisfying $-h \leq dist \leq h$. h is labelled a bandwidth. h is calculated specifically for a pair of sample-polynomial order as described in Cattaneo, Idrobo, and Titiunik (2017b). Moreover, as suggested by Imbens and Lemieux (2008), I choose to use a rectangular kernel function to weight the observations in the $dist$ limited data, which is equivalent to use the same weigth for all observations. Such type of regression is called local linear regression, it consists simply of regressing the equations with standard methods (OLS or 2SLS) within the limited data.[80]

3.2.6 Specification issues

In this paragraph, I adress the problem of the exact forms that should have, given the data, the equations (1), (2), (3), (6), (7) and (8). This question is relevant since the data are cross sectional ones except in the RDD. The main specification issues which would be investigated in the different models are the presence of class effects and the heteroskedasticity of the error terms.

[80] Kaila (2017) and Matta et al. (2016) used triangular kernel weights in their papers instead. It is common knowledge that the estimation results are not likely to be sensitive to the kernel weighting function.

3.2.6.1 Fixed effects

Since the data are three cross-sectional data and have grouping variables,[81] it is highly suspected that there are effects of one of these variables. Besides, in the french educational context, as Piketty, Valdenaire, and others (2006) found a substantial negative effect of the class size in primary school, I focus my fixed effects analysis on the class identification grouping variable. An F-Test was implemented to detect the presence of class fixed effects in the equations (1), (2), (3) and in the equation (6). With no surprise, the presence of class fixed effects is detected in all four equations.[82] However, as the class identifiers are identical across cohorts, class fixed effects will not be included into RDD equations. As consequences, the final forms of the equations are presented below :

Ordinary least squares

$$Y_{ic} = \alpha_0 + \alpha_1.A_{ic} + \alpha_2.X_{ic} + \phi_c + \nu_{ic} \tag{9}$$

2SLS

$$\begin{cases} Y_{ic} = \alpha_0 + \alpha_1.A_{ic} + \alpha_2.X_{ic} + \phi_c + \nu_{ic} \\ A_{ic} = \gamma_0 + \gamma_1.Z_{ic} + \gamma_2.X_{ic} + \psi_c + \eta_{ic} \end{cases} \tag{10}$$

Reduced form

$$Y_{ic} = \delta_0 + \delta_1.Z_{ic} + \delta_2.X_{ic} + \omega_c + \epsilon_{ic} \tag{11}$$

Control function approach

$$Y_{ic} = \alpha_0 + \alpha_1.A_{ic} + \alpha_2.X_{ic} + \alpha_3.\widehat{\eta}_{ic} + \alpha_4.\widehat{\eta}_{ic}.A_{ic} + \phi_c + \nu_{ic} \tag{12}$$

The RDD equations remains the same. Recall that the main interest of performing an RDD in this paper in for comparison purpose. Hence, the cross-sectional regressions will also be performed additionally, and only in confrontation with RDD results, without class fixed effects[83], without the inclusion of baseline covariates in the equations and without the advanced pupils (this latter, intuitively, should reduce the magnitude of estimation results since the high ability pupils are selected into the advanced ones) in order to assure their comparability to the RDD results.

3.2.6.2 Inference procedures

Again, in the presence of cross-sectional datas, the sandard errors are likely to be heteroskedastic. Since heteroskedasticity can take several forms, I compute heteroskedasticity robust standard errors using the Arellano (1987) estimator. It is preferred over the White and others (1980) estimator because of the presence of the class effects.

[81] School township identification, school identification and class identification

[82] All of the p-values are equal to zero.

[83] This does not affect much the results.

Concerning the control function method, the standard errors of the estimates are calculated using wild bootstrap procedure as exposed in Davidson and Flachaire (2008).[84] The number of bootstrap samples chosen in the present work is chosen to equals 501 (on purpose an odd number).

Last, since the RDD models are estimated without class effects, the standard White and others (1980) estimator is used to compute the corresponding standard errors.

3.2.7 Sensitivity checks

Several sensitivity checks are performed in this paper in order to assess how convincing the methods and results are. First, for cross-sectional models, I performed a sensitivity check of the estimation results to some alternative forms of the equations and to the form of the instrument as well. More precisely, concerning the 2SLS cross section model (equation (10)), an alternative is to estimate a « fully saturated » version, in the sens of Angrist and Imbens (1995), i.e, in the first stage, adding all possible interactions between the instruments and the covariates as well as between the covariates themselves as additional control variables and in the structural equation, adding all possible interaction between the age and covariates as well as between the covariates themselves as additional control variables. The result of such manipulation should recover a weighted average of LATEs (under monotonicity assumption, which is unfortunately not verified here) in the estimated coefficient of interest (the coefficient of the age variable). Although this alternative seems more correct, the simple version of equation (10) is more parcimonious. Thus if the results are not too sensitive to this aspect of specification (which is the case, as the Figure 18 illustrates it, if one compare the 2SLS model with the « 2SLS-sat » labelled model), the equation (10) is prefered. A similar alternative specification could be applied to the reduced form estimations (equation (11)), i.e by adding all the mentioned interaction terms on the right-hand side of the equations. I do not report graphical sensitivity analysis of this option because of its extreme fluctuation, which will perturbe the visualization of the other plots. Instead, I highlight in column (7) of Table 13 such fluctuation.

Moreover, concerning the control function approach, Wooldridge (2015) propose two extensions of the equation (6). The general idea of these extensions is to account for a potential heterogeneity of effects of other independent variables across individuals. The first extension consists of adding as additional covariates all possible interactions between $\hat{\eta}$ and the baseline covariates X. It is similar to the fully saturated 2SLS equations described above. The second extension is, adding as additional regressors **to the first extension regressors**, $\hat{\eta}^2$ and all possible interactions between $(\hat{\eta}^2 - \widehat{Var(\hat{\eta})})$ and the baseline covariates. As the scope of the paper is not to scrutinize these specifications but to check how the results are sensitive to these, I refer to Wooldridge (2015) for which details their precise sens. These two extensions are respectively labelled « CFH-E1 » and « CFH-E2 » in the Figure 18. This figure is very general, hence The table 13 shows a more detailed numerical version. An interest of this latter is to assess that all the coefficients are significant at 1% level except in the column (7) which reports the saturated reduced form. In this model, the results are way too sensitive such that they probably should not be considered at all.

[84] The weighted used to generate the bootstraped samples in the procedure are the weights of Rademacher, among other possibilities.

In another perspective, the I also face the RD results to their robustness to the bandwidth choice. Since a specific bandwidth is used for a specific pair of sample-polynomial order, illsutrating varying RDD results corresponding to these bandwiths would be cumbersome. Instead, I illustrate the sensitivity analysis by the devation relative to the specific optimal bandwidth. For example, the optimal bandwidth to the whole-sample-polynomial-of-order-1-intention-to-treat regression equals to 111 days, the one corresponding to the polynomial of order 2 is different. Thus, to compactly illustrate a sensitivity check, I plot the variation of the estimates in function of their common deviations from their specific optimal bandwidth. See the Figure 20, columns labelled « Poly 1 » and « Poly 2 », line labelled « All » for a clear visualization. More analysis of these results are given in Section 4.

4 Results and discussions

4.1 Endogeneity test of the age at test and first stage regression results

The estimations resulting from equation (5) and the first stage in equation (10) are presented in Table 5. Columns (1) to (3) reports for the three cohorts respectively the endogeneity test results. What is of interest in these are the significance of the estimates. Columns (4) to (6) correspond to the first stage of the main regression results. First, as highly expected, the $\widehat{\alpha_3}$ (columns (1) to (3)) are all revealed significant at the 1% level, which means that the age at test is indeed and endogenous variable within the structural equation. On the other hand, as discussed above, a way of investigating the validity of the instrument in our case is to measure its prediction power of the independent variable of interest (the age at test). This is why the first stage results (here with covariates and class fixed effects) are of importance. It can be observed that the causal impact of the instrument on the endogenous variable for the three regressions is very strong, since it is not less than 0.8 (columns (4) to (6)) (close to 1) with significance at the 1% level. This general feature suggests that the assigned relative age does not suffer from weak prediction power because ceteris paribus, being the relatively oldest within a cohort instead of the youngest (a variation of 1 year in assigned relative age) causes on average the age at test to vary nearly about 0.8 year. Moreover, the F-statistics in the first stage regressions are all way above 10, which supports again that the instrument is not « weak » (Staiger and James 1997).

In addition, I outline the first stage results for the RDD in the appendix, Table 22. The column (1), for comparison purpose, reports a first stage similar to that in equation (10) **but without covariates and class fixed effects within the regressors and excluding from the observations used for estimations the advanced pupils**. All these operations are executed in order to ensure the comparability to the regession discontinuity framework detailed in Section 3. Columns (2) to (4) correspond to the first stage estimations of equation system (7). Concerning the label of these columns, « RD-FS-01 » designate the regression discontinuity first stage considering a 1^{st} polynomial order form of $f_{(A)}()$, « RD-FS-02 » correspond to a 2^{nd} polynomial order and so on. A first streaking feature is that when the estimation is performed using all values of the covariates (but not, recall, all the observations available in the data) or using solely the children having underprivileged parents, the first stage estimates of the cross-sectional model (column (1)) and the RDD (columns (2) to (5)) are very similar, independently of the polynomial order. Second, for the privileged's children subgroup, the prediction power of the instrument appear stronger for for the regression discontinuity design. Last, when the considered sample are the sex subgroups, the estimates are slightly weaker in the RDD.

Table 5: Endogeneity test of age and First stage regression results

	Endogeneity test (dep.var : Total test score)			First stage regressions (dep.var : Age at test)		
	2010	2011	2012	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned relative age				0.845***	0.806***	0.843***
				(0.012)	(0.012)	(0.012)
$\widehat{\eta_{ic}}$	−1.121***	−1.090***	−1.019***			
	(0.044)	(0.043)	(0.052)			
F-statistic				543.921***	578.541***	547.594***
N	13,561	14,622	10,734	13,561	14,622	10,734
R^2	0.236	0.256	0.202	0.297	0.294	0.352
Adjusted R^2	0.196	0.218	0.149	0.260	0.258	0.309

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

4.2 The effect of age at test on test scores in grade 5

Two types of results will be developed therefore : cross-sectional and regression discontinuity approach results., Recall that the first type will be based on cohort-distinct regression models while the second type is based on pupils born in 1999 and 2000 only (i.e comparable with the 2010 cohort cross-sectional results since the 1999 borns are assigned to this cohort).

4.2.1 Cross-sectional estimates

4.2.1.1 Main results

The very main results about the effect of age at test on grade 5 national assessment test scores are presented in Table 6. The dependent variable is the total test score. Four corresponding regression models are illustrated here : ordinary least squares from columns (1) to (3), 2SLS estimations – using the assigned relative age as an instrument to the age at the moment of the examination – results are reported in columns (4) to (6), regression outputs from a control function approach controlling for the heterogeneity of age effect can be read in columns (7) to (9) and finally reduced form outputs are illustrated by the columns (10) to (12). In this table, the first two lines are those of main interest since they correspond respectively to the estimated coefficients of the age and assigned relative age variables.

Unsurprisingly, the OLS outputs report a highly negative effects of age on test scores. This is common in this specific framework (age and educational performances relationship). Nevertheless, as well discussed in this paper, these results suffer from a downard bias. This latter is confirmed by the results of two following models (2SLS and CFH). In fact, the estimates of age effects vary between -0.4 and -0.5 s.d for OLS models while in the case one control for the endogeneity of age, the estimates appear to vary between $+0.26$ and $+0.3$ s.d. As a benchmark to understand the magnitude of these estimates, recall that a year of learning provides from $+0.25$ to $+0.33$ s.d « learning gains » (which correspond to test scores here).[85] These estimates on the total test scores are quite in line with this benchmark. Another feature of these main results is that the reduced form (« RF ») estimates are inferior to the 2SLS ones. This feature appear to be robust to several elements : cohort, sample choice, dependent variable or form of the instrument, as it will be illsutrated with the upcoming results. This is likely to be due to the considerable amount of repeaters / redshirters within a grade. Indeed, the 2SLS estimates do not contain retention / redshirting effects while the reduced form estimates do.[86] Hence, a lower reduced form estimate means that the retention / redshirting have positive impact on the relatively youngest pupils in grade 5. This is an interesting information from a policy perspective, yet more investigations on the effect of retention were already performed by several authors.[87]

On the other hand, while one compare these age effects across successive cohorts, these appear to slightly decrase. A possible explanation is the changing structure of the examinations from a school year to another. However, the magnitude of the decrase accross cohorts here prevent us from drawing interesting pattern since the differences are quite small even in a unit of s.d. More interesting results may be be highlighted in the upcoming alternatives (subgroups used for the regressions). Recall that we discussed a way of neutralizing the bias due to the month of birth patterns of the farmers and the retired social category's children : by excluding august births from the observations used to estimate the effects and by merging the retired social category into the « Others » category. The results are presented in Table 7. When one compares the magnitudes of this table with those of Table 6 (main results) and 9 (subgroup results), a dramatical difference, either in magnitudes or in patterns of the estimates (significance, within background characteristics comparison), is hardly spotted. Hence, we could argue that the bias due to the worrying pattern of the month of birth proportions fo the farmers and retired's children in not collapsing the results.

Moreover, Lines 3 to 11 report the estimated coefficients on the covariates included in the regressions equations. Since these are not the principal interest of this paper, only few commentary will be provided about these. First, the effect of being a boy instead of a girl provides all else equals a consistent disadvantage of the order of -0.2 s.d in total test scores. These coefficients are all significant at conventional level. Then, this pattern about the sex covariate is in contrast with the social category one. In fact, similar striking observation seems to be more rare since for one of the value of the social category variable, the magnitudes of the effects across models and cohorts (i.e if one concentrates on a unique line) are not as stable as for the sex variable. However, few observations are worth some highligths. The reference value is being a farmer children. Unsurprisingly, the largest magnitudes are attributes to the executives (line 5) while it is difficult to affirm to which category the lowest magnitudes are attributed between the unemployed (line

[85] A « rule of thumb », according to Woessmann (2016).

[86] Bedard and Dhuey (2006) provides more detailed explanations.

[87] Alet, Bonnal, and Favard (2013) in France for example ; the grade 5 timing is important to note because their paper showed that there are a negative effect of retention in later years

10) and the « Others » category (line 11). Also, the unemployed category yields negative estimates, confirming that being a child of an unemployed parent is, all else equals disadvantaged compared to one of an average child of a farmer. This disadvantage is, in grade 5 total test scores s.d scale, in order of -0.2 to -0.4. Other social category variable lines (except the line 11, which I avoid to interpret because of its content) do not present a robust significant pattern. Last, note the number of observations in the 2012 cohort being exceptionaly inferior. This is due to the 21% of missing values of the sex variable. One unexpected result is that in column (8), line 13, the coefficient between the first stage residual and the age interaction is not significant (its p-value equals to 0.11), suggesting, according to what was presented in the econometric framework subsection, that the age variable in this precise case may be not endogenous. This is unlikely the case because of the similarity of the standard error to the other cohorts : in 2011, the standard error equals to 0.034 while it equals to 0.038 in 2010 and 0.044 in 2012. Moreover, another similarity is the value of the OLS estimates of the age on test scores (line 1). Hence, this non significance of the coefficient at column (8), line 13 is more likely an eventual anomaly of the bootstraping procedure of the standard error.[88]

[88] The standard errors presented in the table, for the CFH models are individual-level wild bootstraped, in contrast of a class-clustered wild bootstrap. The two procedures yield to very similar results.

Table 6: Main regressions results

	OLS			2SLS			CFH			RF		
	2010	2011	2012	2010	2011	2012	2010	2011	2012	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1 - Age at test	−0.509***	−0.495***	−0.414***	0.304***	0.302***	0.269***	0.301***	0.300***	0.265***			
	(0.015)	(0.015)	(0.019)	(0.033)	(0.031)	(0.034)	(0.038)	(0.036)	(0.042)			
2 - Assigned relative age										0.257***	0.244***	0.226***
										(0.026)	(0.024)	(0.027)
3 - Sex - Male	−0.218***	−0.229***	−0.216***	−0.288***	−0.289***	−0.256***	−0.283***	−0.287***	−0.252***	−0.262***	−0.267***	−0.242***
	(0.015)	(0.014)	(0.016)	(0.017)	(0.015)	(0.017)	(0.018)	(0.016)	(0.020)	(0.016)	(0.014)	(0.016)
4 - SPC - Entrepreneurs	0.112*	0.163**	0.011	0.157**	0.213***	0.044	0.151*	0.213**	0.043	0.137**	0.203***	0.034
	(0.059)	(0.077)	(0.069)	(0.068)	(0.079)	(0.073)	(0.082)	(0.098)	(0.096)	(0.063)	(0.076)	(0.070)
5 - SPC - Executives	0.436***	0.484***	0.392***	0.550***	0.600***	0.491***	0.543***	0.599***	0.493***	0.508***	0.562***	0.460***
	(0.061)	(0.073)	(0.064)	(0.070)	(0.076)	(0.068)	(0.081)	(0.093)	(0.093)	(0.065)	(0.072)	(0.065)
6 - SPC - Intermediates	0.181***	0.295***	0.141**	0.265***	0.356***	0.190***	0.256***	0.355***	0.191**	0.240***	0.337***	0.177***
	(0.059)	(0.075)	(0.064)	(0.067)	(0.077)	(0.068)	(0.081)	(0.092)	(0.090)	(0.062)	(0.074)	(0.065)
7 - SPC - Employees	−0.003	0.101	−0.053	0.026	0.125*	−0.044	0.021	0.124	−0.043	0.017	0.119*	−0.046
	(0.055)	(0.071)	(0.063)	(0.063)	(0.072)	(0.067)	(0.078)	(0.087)	(0.088)	(0.058)	(0.069)	(0.064)
8 - SPC - Workers	−0.104*	−0.040	−0.153**	−0.084	−0.052	−0.162**	−0.089	−0.052	−0.159*	−0.090	−0.046	−0.154**
	(0.056)	(0.074)	(0.061)	(0.065)	(0.078)	(0.066)	(0.079)	(0.091)	(0.089)	(0.060)	(0.074)	(0.062)
9 - SPC - Retired	0.218**	0.346***	0.238**	0.235**	0.390***	0.222**	0.232**	0.393***	0.231	0.226**	0.372***	0.230**
	(0.093)	(0.104)	(0.102)	(0.108)	(0.116)	(0.110)	(0.116)	(0.135)	(0.148)	(0.100)	(0.108)	(0.104)
10 - SPC - Unemployed	−0.231***	−0.175**	−0.348***	−0.282***	−0.237***	−0.413***	−0.280***	−0.234***	−0.405***	−0.264***	−0.213***	−0.384***
	(0.054)	(0.070)	(0.062)	(0.062)	(0.072)	(0.066)	(0.077)	(0.086)	(0.088)	(0.058)	(0.068)	(0.062)
11 - SPC - Others	−0.351***	−0.373***	−0.294**	−0.430***	−0.461***	−0.369***	−0.418***	−0.455***	−0.355**	−0.397***	−0.426***	−0.324***
	(0.056)	(0.071)	(0.120)	(0.066)	(0.074)	(0.124)	(0.079)	(0.086)	(0.165)	(0.061)	(0.070)	(0.119)
12 - $\hat{\eta}$							0.413	−0.483	0.206			
							(0.430)	(0.383)	(0.505)			
13 - Age $\times\hat{\eta}$							−0.138***	−0.055	−0.107**			
							(0.038)	(0.034)	(0.044)			
N	13,561	14,622	10,734	13,561	14,622	10,734	13,561	14,622	10,734	13,561	14,622	10,734
R^2	0.168	0.194	0.145	0.026	0.051	0.047	0.238	0.257	0.203	0.105	0.135	0.109
Adjusted R^2	0.125	0.153	0.089	−0.026	0.003	−0.016	0.198	0.219	0.150	0.058	0.091	0.050

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

4.2.1.2 Alternative dependent variables and covariate subgroups results

Table 8 present more compactly regression results in the case one use french scores and maths scores as dependent variables. Column (1) to (3) correspond to the case in which the french scores are used while column (4) to (6) correspond to the maths scores as dependent variable. One interest of this exercice is to highligth a probable gap in « absolute difficulty » between the two topics. To understand this purpose, consider that the maths questions are obviously more difficult than the french ones (for example, if a considerable amount of the maths quetions require implicit mathematical formula manipulations to get the right answer, which is normally above the standard difficulty in grade 5) ; then, since a difference of one year at early ages is an important difference in terms of child intelligence development, the oldest pupils will be highly advantaged compared to the youngest ones to anwser the maths questions. Hence, the age effect will be considerably superior when one choose the maths scores instead of the french scores as the dependent variable.

The table illustrates that generally, the estimates are similar each other (very slightly lower for the french score case in column (1) to (3)). Following the reasoning above, an interesting interpretation of this feature is that the french-maths relative difficulty is balanced although the maths questions could be perceived as slightly more difficult by the pupils compared to the french questions. Such possibility could be assessed by searching for neurological studies about the brain requirements to answer french-type and mathematics-type questions. These are reserved for future investigations.

Next, it is of a standard practice to run the estimations only on pupils taking a specific value of a covariate (for example, running the estimations on the subsample of girls only). It would be informative on how heterogenous the effect is across pupils with considerably different backgroung characteristics. It is of interest since the effect may be more alarming for a specific type of pupils and then responding policies could be in priority target to this type of pupils. Another argument in favor of the mentioned practice is that, without it, policies may be unecessarly (which is equivalent to unecessary use of resources) extended to pupils within which the age effect is not economically significant. For example, suppose that the age effects are not economically significant except within the unemployed's children. Thus, if one does not have this information, one could allocate resources (material, financial or human) to make a policy for the whole population while it would produce the same results if one allocate resources solely to make a policy for the unemployed.

The Table 9 presents the results of the subgroup estimations results. Columns (1) to (3), (4) to (6) and (7) to (9) reports results corresponding to, respectively, the total score, the french score and the maths score as dependent variables. The column labels inform the regression models used. Grouping succesively the three cohorts, lines 1 to 3, 4 to 6, 7 to 9 and 10 to 12 indicate respectively : the females, males, underprivileged (social category)'s children and privileged's children.[89]. A logical way to analyse this table is to compare females individuals to males ones and underprivileged's children to privileged's ones. One could easily observe that the age effects are more accentuated within the girls than those within the boys. Nevertheless, the females-males gap in the estimated effects is not as obvious for the 2012 cohort (comparing line 3 with line 6). If this attenuation is not due to the prevalence of missing values within the 2012 cohort, a probable reason is the presence of the effect of a gender-gap reduction measure. Continuing within the sex

[89] No interaction between the variables will be report for conciseness, yet it would be interesting to keep in mind for future researches.

subgroup analysis, for boys or girls, the fact that the reduced form estimates (hence containing the retention / redshirting effect) appear to be not as inferior to the 2SLS of control function approach estimates as in the case in which the whole sample is used for regressions. This result suggest that, apart of deeper investigations, in the context of Reunion Island, the effect of retention or delaying school entry on test scores in grade 5 is less driven by the gender than the parents social category.

Concerning the underprivileged-privileged comparison, drawing conclusions about the gap is hard to perform since the gap sign is varying across the three cohorts : generally, the estimate effects are reported superior for the underprivileged category in 2010 ; it is the opposite in 2011 and in 2012, the underprivileged category appears to have again a superior age effect. This is less common in studies of age effects because these effects are more often clearly heterogenous for subgroups of pupils having less advantaged social backgrounds. The estimation results across different ungrouped social category are presented in two tables (in two parts) : Table 14 and Table 15 in the appendix. The social category subgroups concerned in the first one are the farmers, entrepreneurs, executive and intermediates while those in the second one are the employees, workers, retired and unemployed. These tables are structured similarly to the Table 9, as most of the upcoming results tables.Finding patterns in terms of significance or some trends in the magnitudes of the estimates requires detailed observation of the tables. Indeed, the only line in Table 14 that seems to report some consistently significant estimates (across models and dependent variables) are in line 11 (for the intermediates category in 2011) : the magnitudes seem not to be specially different from main results. The following patterns are retrieved in this line : superiority of the 2SLS / control function approach estimates to the reduced form estimates and superioriy of the effects for french score dependent variable (columns (4) to (6)) compared to the case of maths score as dependent variable (columns (7) to (9)), possibly explained by a perception of more difficulty in the maths questions compared to french questions. Note however the line 8 for the executives that report higher magnitudes than in the line 11 or generally than in main results. Additionally, concerning the second table, two categories yield to regular pattern of significance and magnitudes : the employees (set of lines *I*) and the unemployed (set of lines *IV*). Unfortunately, the sign of the gap in their magnitudes (comparing line 1 with line 10, line 2 with line 11 and line 3 with line 12) are not strikingly constant. However, one easily spotted difference in the feature of the estimates between these two categories are their respective gaps between 2SLS / control function and reduced form estimates. This is logically explained by the proportion of repeaters / redshirters within these two groups respectively (the higher the proportion of repeaters / redshirters, the higher should be the difference between IV approaches and reduced form) : these proportions, illustrated by the Table 19 are between 8% and 10% within the employees and the double, i.e about 20% within the unemployed.

Last in this paragraph, following the same reasoning about the difficulty perception differences between french questions and maths questions, it is possible to go even deeper and analyze the sub-items-depedent-variables regression results. These are presented in Tables 10 and 11 corresponding respectively to french sub-items and maths sub-items. As expected, they are structred similarly to the previous compact regression tables. The ungrouped social category subgroup results will not be reported for conciseness. Standard patterns of the estimates are maintained : higher effects for girls compared to those of the boys, especially in 2010 (line 4 compared to line 7 in the two tables mentioned above) ; hard to spot trend in the underprivileged-privileged comparison of the effects (set of lines *IV* compared to set of lines *V* in the two tables) and lower reduced forms estimates.

4.2.2 Regression discontinuity estimates

Estimation results from Equations (8) and (7) which correspond respectively to the intention-to-treat RD estimates and Fuzzy regression estimates are reported in Table 12. Recall that in all case the dependent variable is the french scores. Column (1) and (6) illustrate respectively the reduced form estimates in 2010 and the 2SLS (cross-sectional) estimates in 2010, without advanced pupils and without inclusion of covariates for comparability purpose : column (1) is to be compared with columns (2) to (5) while column (6) is to be compared with column (7) to (10). One important difference with cross-sectional regression tables to notice here is the highly varying number of observations used, because of the bandwidth choices that are specific for each combination of subgroup, polynomial order and type of RD model (fuzzy or not). In order to highligth interesting informations, I propose to analyse this table from several perspectives. First, a line by line analysis would be informative of how the results behave with different polynomila order, given a subgroup pupils in which the models are estimated (i.e given a line). In the case the estimates are run over the all pupils (in the sens that all covariates values are retained), the reduced form estimate are slightly lower in the cross sectional model than those of the RD models. This feature is maintained within all lines except the line 4 (the underprivileged's children). A logical reason behind this exceptional pattern for the underprivileged's children is that within those children, the proportion of repeaters / redshirters are, relatively to other social cateogries, higher around the cutoff, hence reduce the relative age gap that cause all else equal the test scores gap. When comparing the magnitude of the estimates within a line across different RD specifications, no constant pattern seems to be present. For example, the polynomial order 2 yields the highest intention to treat estimates within the female pupils while it is not the case for the male pupils. Either in the reduced form or the 2SLS regression discontinuity results, the order of the polynomial does not greatly affect the magnitudes but do affect the significance levels instead. This suggest that some polynomial orders may be not adapted to the data. In practice, it is advised to visualize graphically the results of RDD to assess how good a choice of the polynomial order is adapted to the data. I perform this practice for the columns (2) to (5), i.e the four polynomial order and the 5 estimation samples. The resulting graphic is presented in Figure 15. It can be observed at first sight that the higher the polynomial order, the larger the bandwidth.

Now, how about a column by column analysis ? i.e, analyzing, given a model, how the estimates are higher or lower relatively to each other depending on the values of covariates that are imposed to perform the regressions. First, apattern that were retrieved in the cross-sectional results is the superiority of the effects within the girls than within the boys. A known fact that is likely to be behind such robust pattern is that the girls, at early ages, mature more rapidly than boys do. Hence, a one year differential among the average girl is more « important » in terms of ability (that determines test scores) than a one year differential among the average boy. Interestingly, the RD estimates appear lower for the underprivileged's children than those for the privileged one. This could be explained by the local characteristic around the cutoff, i.e within the underprivileged ones, the age differential between those locally on the right of the cutoff and those locally on the left are not as wide as this of the privileged's children.

Table 7: Main regression results without august borns and modified social category variable

	Total score		French score		Maths score	
	2SLS (1)	RF (2)	2SLS (3)	RF (4)	2SLS (5)	RF (6)
I - All						
1 - 2010						
	0.3***	0.253***	0.283***	0.239***	0.286***	0.242***
	(0.033)	(0.026)	(0.034)	(0.027)	(0.031)	(0.025)
	N = 12422	N = 12422	N = 12422	N = 12422	N = 12422	N = 12422
2 - 2011						
	0.296***	0.238***	0.27***	0.217***	0.295***	0.238***
	(0.031)	(0.024)	(0.031)	(0.023)	(0.032)	(0.024)
	N = 13383	N = 13383	N = 13383	N = 13383	N = 13383	N = 13383
3 - 2012						
	0.265***	0.223***	0.25***	0.211***	0.253***	0.213***
	(0.034)	(0.027)	(0.035)	(0.028)	(0.034)	(0.027)
	N = 9797	N = 9797	N = 9797	N = 9797	N = 9797	N = 9797
II - Females						
4 - 2010						
	0.36***	0.304***	0.361***	0.305***	0.314***	0.265***
	(0.047)	(0.037)	(0.049)	(0.038)	(0.044)	(0.036)
	N = 6135	N = 6135	N = 6135	N = 6135	N = 6135	N = 6135
5 - 2011						
	0.351***	0.279***	0.343***	0.272***	0.318***	0.252***
	(0.045)	(0.034)	(0.045)	(0.033)	(0.046)	(0.034)
	N = 6698	N = 6698	N = 6698	N = 6698	N = 6698	N = 6698
6 - 2012						
	0.274***	0.232***	0.255***	0.216***	0.267***	0.226***
	(0.051)	(0.041)	(0.05)	(0.041)	(0.051)	(0.042)
	N = 5014	N = 5014	N = 5014	N = 5014	N = 5014	N = 5014
III - Males						
7 - 2010						
	0.239***	0.204***	0.207***	0.177***	0.255***	0.217***
	(0.049)	(0.04)	(0.05)	(0.041)	(0.046)	(0.038)
	N = 6287	N = 6287	N = 6287	N = 6287	N = 6287	N = 6287
8 - 2011						
	0.215***	0.177***	0.173***	0.143***	0.248***	0.205***
	(0.046)	(0.036)	(0.045)	(0.036)	(0.046)	(0.036)
	N = 6685	N = 6685	N = 6685	N = 6685	N = 6685	N = 6685
9 - 2012						
	0.262***	0.222***	0.255***	0.217***	0.238***	0.202***
	(0.054)	(0.044)	(0.056)	(0.046)	(0.052)	(0.043)
	N = 4783	N = 4783	N = 4783	N = 4783	N = 4783	N = 4783
IV - Underprivileged						
10 - 2010						
	0.315***	0.263***	0.301***	0.252***	0.294***	0.246***
	(0.039)	(0.031)	(0.042)	(0.033)	(0.037)	(0.029)
	N = 7925	N = 7925	N = 7925	N = 7925	N = 7925	N = 7925
11 - 2011						
	0.299***	0.237***	0.286***	0.227***	0.279***	0.221***
	(0.042)	(0.032)	(0.042)	(0.032)	(0.042)	(0.032)
	N = 7921	N = 7921	N = 7921	N = 7921	N = 7921	N = 7921
12 - 2012						
	0.255***	0.214***	0.237***	0.199***	0.248***	0.208***
	(0.04)	(0.032)	(0.041)	(0.033)	(0.04)	(0.032)
	N = 7420	N = 7420	N = 7420	N = 7420	N = 7420	N = 7420
V - Privileged						
2010						
	0.271***	0.226***	0.243***	0.203***	0.276***	0.231***
	(0.081)	(0.066)	(0.078)	(0.063)	(0.084)	(0.068)
	N = 2401	N = 2401	N = 2401	N = 2401	N = 2401	N = 2401
2011						
	0.29***	0.237***	0.267***	0.218***	0.285***	0.233***
	(0.071)	(0.057)	(0.072)	(0.058)	(0.071)	(0.056)
	N = 2367	N = 2367	N = 2367	N = 2367	N = 2367	N = 2367
2012						
	0.331***	0.286***	0.332***	0.286***	0.289***	0.249***
	(0.08)	(0.066)	(0.078)	(0.065)	(0.081)	(0.068)
	N = 2330	N = 2330	N = 2330	N = 2330	N = 2330	N = 2330

Dependent variable

Table 8: Regression results with french and maths scores as dependent variables

	Dependent variable : french scores			Dependent variable : maths scores		
	2SLS (1)	CFH (2)	RF (3)	2SLS (4)	CFH (5)	RF (6)
1 - 2010						
	0.285^{***}	0.281^{***}	0.241^{***}	0.295^{***}	0.293^{***}	0.25^{***}
	(0.033)	(0.038)	(0.027)	(0.032)	(0.038)	(0.026)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
2 - 2011						
	0.28^{***}	0.277^{***}	0.226^{***}	0.301^{***}	0.301^{***}	0.243^{***}
	(0.031)	(0.036)	(0.024)	(0.032)	(0.036)	(0.024)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
3 - 2012						
	0.252^{***}	0.248^{***}	0.213^{***}	0.261^{***}	0.259^{***}	0.22^{***}
	(0.035)	(0.042)	(0.028)	(0.035)	(0.042)	(0.028)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734

Table 9: Subgroup regression results

	Dep.var : total scores			Dep.var : french scores			Dep.var : maths scores		
	2SLS	CFH	RF	2SLS	CFH	RF	2SLS	CFH	RF
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
I - Females									
1 - 2010									
	0.365^{***}	0.364^{***}	0.308^{***}	0.361^{***}	0.361^{***}	0.305^{***}	0.324^{***}	0.324^{***}	0.274^{***}
	(0.048)	(0.052)	(0.037)	(0.048)	(0.052)	(0.038)	(0.046)	(0.052)	(0.037)
	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666
2 - 2011									
	0.355^{***}	0.353^{***}	0.282^{***}	0.351^{***}	0.348^{***}	0.278^{***}	0.322^{***}	0.322^{***}	0.256^{***}
	(0.045)	(0.05)	(0.033)	(0.045)	(0.05)	(0.033)	(0.045)	(0.05)	(0.034)
	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325
3 - 2012									
	0.278^{***}	0.276^{***}	0.236^{***}	0.26^{***}	0.257^{***}	0.22^{***}	0.272^{***}	0.271^{***}	0.231^{***}
	(0.05)	(0.06)	(0.041)	(0.049)	(0.06)	(0.04)	(0.052)	(0.06)	(0.042)
	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461
II - Males									
4 - 2010									
	0.237^{***}	0.233^{***}	0.203^{***}	0.205^{***}	0.199^{***}	0.175^{***}	0.257^{***}	0.254^{***}	0.22^{***}
	(0.048)	(0.054)	(0.039)	(0.049)	(0.054)	(0.04)	(0.048)	(0.054)	(0.039)
	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895
5 - 2011									
	0.228^{***}	0.227^{***}	0.188^{***}	0.19^{***}	0.189^{***}	0.157^{***}	0.257^{***}	0.257^{***}	0.212^{***}
	(0.046)	(0.051)	(0.036)	(0.046)	(0.051)	(0.036)	(0.047)	(0.051)	(0.037)
	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297
6 - 2012									
	0.271^{***}	0.268^{***}	0.23^{***}	0.261^{***}	0.257^{***}	0.222^{***}	0.254^{***}	0.252^{***}	0.216^{***}
	(0.054)	(0.063)	(0.043)	(0.055)	(0.063)	(0.044)	(0.053)	(0.063)	(0.043)
	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273
III - Underprivileged									
7 - 2010									
	0.325^{***}	0.32^{***}	0.271^{***}	0.309^{***}	0.303^{***}	0.258^{***}	0.309^{***}	0.305^{***}	0.258^{***}
	(0.04)	(0.048)	(0.031)	(0.041)	(0.048)	(0.033)	(0.038)	(0.048)	(0.031)
	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664
8 - 2011									
	0.298^{***}	0.298^{***}	0.237^{***}	0.287^{***}	0.285^{***}	0.229^{***}	0.28^{***}	0.283^{***}	0.223^{***}
	(0.042)	(0.049)	(0.032)	(0.042)	(0.049)	(0.032)	(0.042)	(0.049)	(0.032)
	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645
9 - 2012									
	0.263^{***}	0.257^{***}	0.221^{***}	0.242^{***}	0.235^{***}	0.203^{***}	0.263^{***}	0.26^{***}	0.22^{***}
	(0.04)	(0.048)	(0.032)	(0.04)	(0.048)	(0.032)	(0.041)	(0.048)	(0.033)
	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131
IV - Privileged									
10 - 2010									
	0.272^{***}	0.274^{***}	0.226^{***}	0.239^{***}	0.241^{***}	0.198^{***}	0.289^{***}	0.29^{***}	0.24^{***}
	(0.08)	(0.088)	(0.065)	(0.076)	(0.088)	(0.061)	(0.086)	(0.088)	(0.069)
	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592
11 - 2011									
	0.314^{***}	0.33^{***}	0.254^{***}	0.291^{***}	0.308^{***}	0.235^{***}	0.312^{***}	0.325^{***}	0.252^{***}
	(0.072)	(0.087)	(0.056)	(0.074)	(0.087)	(0.058)	(0.071)	(0.087)	(0.056)
	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564
12 - 2012									
	0.313^{***}	0.323^{***}	0.27^{***}	0.316^{***}	0.328^{***}	0.272^{***}	0.272^{***}	0.278^{***}	0.234^{***}
	(0.078)	(0.09)	(0.065)	(0.076)	(0.09)	(0.063)	(0.081)	(0.09)	(0.068)
	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554

Table 10: Regression results taking french sub-items as dependent variables

	Dependent variable									
	Writing		Reading		Grammar		Spelling		Vocabulary	
	2SLS (1)	RF (2)	2SLS (3)	RF (4)	2SLS (5)	RF (6)	2SLS (7)	RF (8)	2SLS (9)	RF (10)
I - All										
1 - 2010										
	0.234***	0.198***	0.282***	0.238***	0.223***	0.189***	0.232***	0.196***	0.25***	0.211***
	(0.034)	(0.028)	(0.034)	(0.027)	(0.032)	(0.026)	(0.035)	(0.028)	(0.031)	(0.025)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
2 - 2011										
	0.227***	0.183***	0.25***	0.202***	0.263***	0.212***	0.221***	0.178***	0.201***	0.162***
	(0.032)	(0.025)	(0.031)	(0.024)	(0.031)	(0.024)	(0.032)	(0.025)	(0.03)	(0.024)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
3 - 2012										
	0.219***	0.184***	0.224***	0.188***	0.234***	0.197***	0.201***	0.17***	0.229***	0.193***
	(0.033)	(0.027)	(0.036)	(0.029)	(0.035)	(0.028)	(0.035)	(0.029)	(0.041)	(0.033)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734
II - Females										
4 - 2010										
	0.307***	0.26***	0.354***	0.299***	0.296***	0.25***	0.292***	0.246***	0.293***	0.248***
	(0.049)	(0.039)	(0.049)	(0.039)	(0.048)	(0.039)	(0.049)	(0.039)	(0.044)	(0.035)
	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666
5 - 2011										
	0.268***	0.212***	0.304***	0.242***	0.331***	0.263***	0.299***	0.237***	0.254***	0.202***
	(0.044)	(0.034)	(0.045)	(0.034)	(0.046)	(0.034)	(0.047)	(0.036)	(0.042)	(0.032)
	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325
6 - 2012										
	0.239***	0.203***	0.234***	0.199***	0.221***	0.187***	0.233***	0.197***	0.218***	0.185***
	(0.046)	(0.038)	(0.051)	(0.042)	(0.05)	(0.041)	(0.049)	(0.04)	(0.057)	(0.048)
	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461
III - Males										
7 - 2010										
	0.173***	0.148***	0.201***	0.172***	0.139***	0.118***	0.172***	0.147***	0.203***	0.173***
	(0.051)	(0.042)	(0.049)	(0.041)	(0.045)	(0.037)	(0.051)	(0.043)	(0.046)	(0.038)
	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895
8 - 2011										
	0.168***	0.138***	0.18***	0.148***	0.175***	0.144***	0.135***	0.112***	0.132***	0.109***
	(0.048)	(0.038)	(0.044)	(0.035)	(0.046)	(0.037)	(0.046)	(0.037)	(0.044)	(0.035)
	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297
9 - 2012										
	0.21***	0.178***	0.243***	0.207***	0.269***	0.228***	0.166***	0.141***	0.246***	0.209***
	(0.052)	(0.043)	(0.055)	(0.045)	(0.055)	(0.045)	(0.056)	(0.046)	(0.062)	(0.051)
	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273
IV - Underprivileged										
10 - 2010										
	0.235***	0.197***	0.311***	0.26***	0.24***	0.2***	0.253***	0.212***	0.281***	0.235***
	(0.042)	(0.034)	(0.041)	(0.033)	(0.041)	(0.033)	(0.044)	(0.035)	(0.039)	(0.031)
	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664
11 - 2011										
	0.226***	0.18***	0.258***	0.205***	0.294***	0.234***	0.204***	0.162***	0.202***	0.161***
	(0.044)	(0.034)	(0.041)	(0.031)	(0.043)	(0.032)	(0.043)	(0.033)	(0.041)	(0.032)
	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645
12 - 2012										
	0.212***	0.178***	0.212***	0.178***	0.223***	0.187***	0.185***	0.155***	0.235***	0.197***
	(0.04)	(0.032)	(0.041)	(0.034)	(0.041)	(0.033)	(0.041)	(0.033)	(0.047)	(0.038)
	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131
V - Privileged										
13 - 2010										
	0.203**	0.168***	0.251***	0.209***	0.156**	0.13**	0.181**	0.151**	0.243***	0.202***
	(0.079)	(0.065)	(0.081)	(0.066)	(0.074)	(0.06)	(0.083)	(0.068)	(0.069)	(0.056)
	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592
14 - 2011										
	0.254***	0.205***	0.226***	0.182***	0.203***	0.164***	0.359***	0.29***	0.198***	0.16***
	(0.075)	(0.059)	(0.072)	(0.057)	(0.077)	(0.061)	(0.084)	(0.065)	(0.068)	(0.054)
	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564
15 - 2012										
	0.284***	0.245***	0.305***	0.263***	0.261***	0.225***	0.287***	0.248***	0.251***	0.216***
	(0.066)	(0.055)	(0.079)	(0.066)	(0.079)	(0.066)	(0.08)	(0.067)	(0.085)	(0.072)
	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554

Table 11: Regression results taking maths sub-items as dependent variable

	Dependent variable									
	Calculus		Geometry		Measures		Number		Data organization	
	2SLS (1)	RF (2)	2SLS (3)	RF (4)	2SLS (5)	RF (6)	2SLS (7)	RF (8)	2SLS (9)	RF (10)
I - All										
1 - 2010										
	0.205***	0.174***	0.266***	0.225***	0.215***	0.182***	0.27***	0.228***	0.21***	0.177***
	(0.029)	(0.024)	(0.035)	(0.029)	(0.026)	(0.021)	(0.035)	(0.029)	(0.029)	(0.024)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
2 - 2011										
	0.217***	0.175***	0.284***	0.229***	0.26***	0.21***	0.196***	0.158***	0.279***	0.225***
	(0.03)	(0.023)	(0.034)	(0.026)	(0.028)	(0.022)	(0.031)	(0.024)	(0.036)	(0.028)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
3 - 2012										
	0.246***	0.207***	0.144***	0.122***	0.252***	0.212***	0.134***	0.113***	0.23***	0.194***
	(0.038)	(0.031)	(0.028)	(0.023)	(0.039)	(0.032)	(0.027)	(0.023)	(0.035)	(0.029)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734
II - Females										
4 - 2010										
	0.238***	0.201***	0.272***	0.23***	0.234***	0.198***	0.313***	0.264***	0.21***	0.177***
	(0.041)	(0.034)	(0.051)	(0.042)	(0.037)	(0.03)	(0.052)	(0.043)	(0.042)	(0.034)
	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666	N = 6666
5 - 2011										
	0.261***	0.207***	0.334***	0.265***	0.251***	0.199***	0.174***	0.138***	0.284***	0.225***
	(0.043)	(0.032)	(0.049)	(0.038)	(0.039)	(0.03)	(0.043)	(0.033)	(0.052)	(0.04)
	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325	N = 7325
6 - 2012										
	0.281***	0.238***	0.107***	0.091***	0.263***	0.223***	0.125***	0.106***	0.248***	0.211***
	(0.055)	(0.045)	(0.04)	(0.033)	(0.059)	(0.049)	(0.039)	(0.033)	(0.054)	(0.044)
	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461	N = 5461
III - Males										
7 - 2010										
	0.17***	0.145***	0.255***	0.218***	0.187***	0.16***	0.221***	0.188***	0.193***	0.165***
	(0.043)	(0.036)	(0.051)	(0.042)	(0.039)	(0.032)	(0.051)	(0.042)	(0.043)	(0.035)
	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895	N = 6895
8 - 2011										
	0.162***	0.133***	0.227***	0.187***	0.239***	0.197***	0.203***	0.168***	0.245***	0.202***
	(0.043)	(0.034)	(0.05)	(0.04)	(0.04)	(0.032)	(0.045)	(0.036)	(0.052)	(0.041)
	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297	N = 7297
9 - 2012										
	0.212***	0.18***	0.181***	0.153***	0.249***	0.212***	0.133***	0.113***	0.233***	0.198***
	(0.06)	(0.05)	(0.041)	(0.034)	(0.06)	(0.05)	(0.042)	(0.035)	(0.052)	(0.043)
	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273	N = 5273
IV - Underprivileged										
10 - 2010										
	0.229***	0.191***	0.272***	0.228***	0.21***	0.176***	0.268***	0.224***	0.229***	0.191***
	(0.036)	(0.029)	(0.043)	(0.035)	(0.032)	(0.026)	(0.044)	(0.036)	(0.035)	(0.028)
	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664	N = 8664
11 - 2011										
	0.187***	0.149***	0.264***	0.21***	0.27***	0.214***	0.17***	0.135***	0.272***	0.216***
	(0.04)	(0.031)	(0.044)	(0.034)	(0.038)	(0.029)	(0.041)	(0.032)	(0.047)	(0.036)
	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645	N = 8645
12 - 2012										
	0.235***	0.197***	0.154***	0.129***	0.279***	0.234***	0.133***	0.111***	0.225***	0.188***
	(0.045)	(0.037)	(0.033)	(0.028)	(0.046)	(0.037)	(0.033)	(0.027)	(0.042)	(0.034)
	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131	N = 8131
V - Privileged										
13 - 2010										
	0.191***	0.159***	0.278***	0.231***	0.257***	0.214***	0.236***	0.196***	0.184**	0.153**
	(0.073)	(0.06)	(0.093)	(0.076)	(0.071)	(0.058)	(0.09)	(0.073)	(0.083)	(0.068)
	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592	N = 2592
14 - 2011										
	0.168**	0.136***	0.286***	0.231***	0.291***	0.235***	0.212***	0.171***	0.374***	0.302***
	(0.066)	(0.052)	(0.081)	(0.064)	(0.068)	(0.054)	(0.069)	(0.055)	(0.092)	(0.073)
	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564	N = 2564
15 - 2012										
	0.288***	0.248***	0.151***	0.13***	0.197**	0.17**	0.132**	0.114**	0.258***	0.222***
	(0.092)	(0.077)	(0.058)	(0.05)	(0.087)	(0.074)	(0.061)	(0.052)	(0.079)	(0.067)
	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554	N = 2554

Table 12: Regression Discontinuity regression results

Dependent variable : french scores									
	ITT RDD					FRD			
RF	RD-01	RD-02	RD-03	RD-04	2SLS	FRD-01	FRD-02	FRD-03	FRD-04
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1 - All									
0.201***	0.274***	0.236***	0.249***	0.25***	0.222***	0.326***	0.288***	0.304***	0.328***
(0.03)	(0.044)	(0.068)	(0.078)	(0.081)	(0.035)	(0.059)	(0.087)	(0.101)	(0.107)
N = 13296	N = 8306	N = 7379	N = 9725	N = 14322	N = 13296	N = 7471	N = 7379	N = 9725	N = 13603
2 - Females									
0.262***	0.386***	0.412***	0.354***	0.336***	0.285***	0.436***	0.45***	0.46***	0.43***
(0.041)	(0.06)	(0.075)	(0.108)	(0.12)	(0.047)	(0.096)	(0.115)	(0.152)	(0.165)
N = 6507	N = 3885	N = 5560	N = 4634	N = 5821	N = 6507	N = 2765	N = 4532	N = 4634	N = 5821
3 - Males									
0.145***	0.191***	0.253***	0.292***	0.27**	0.164***	0.32***	0.325**	0.344**	0.312**
(0.043)	(0.058)	(0.097)	(0.105)	(0.112)	(0.05)	(0.083)	(0.131)	(0.143)	(0.158)
N = 6789	N = 4610	N = 3454	N = 5283	N = 7138	N = 6789	N = 4035	N = 3454	N = 5167	N = 6070
4 - Underprivileged									
0.243***	0.23***	0.162**	0.155*	0.241**	0.276***	0.271***	0.169*	0.181	0.286**
(0.036)	(0.051)	(0.076)	(0.092)	(0.109)	(0.042)	(0.064)	(0.092)	(0.112)	(0.137)
N = 8571	N = 5298	N = 5341	N = 6169	N = 6812	N = 8571	N = 5341	N = 5216	N = 6169	N = 6812
5 - Privileged									
0.147**	0.284***	0.333**	0.313*	0.387**	0.157**	0.297***	0.359**	0.337*	0.412**
(0.063)	(0.097)	(0.143)	(0.162)	(0.185)	(0.068)	(0.11)	(0.165)	(0.184)	(0.207)
N = 2485	N = 1305	N = 1305	N = 1877	N = 2203	N = 2485	N = 1337	N = 1268	N = 1877	N = 2203

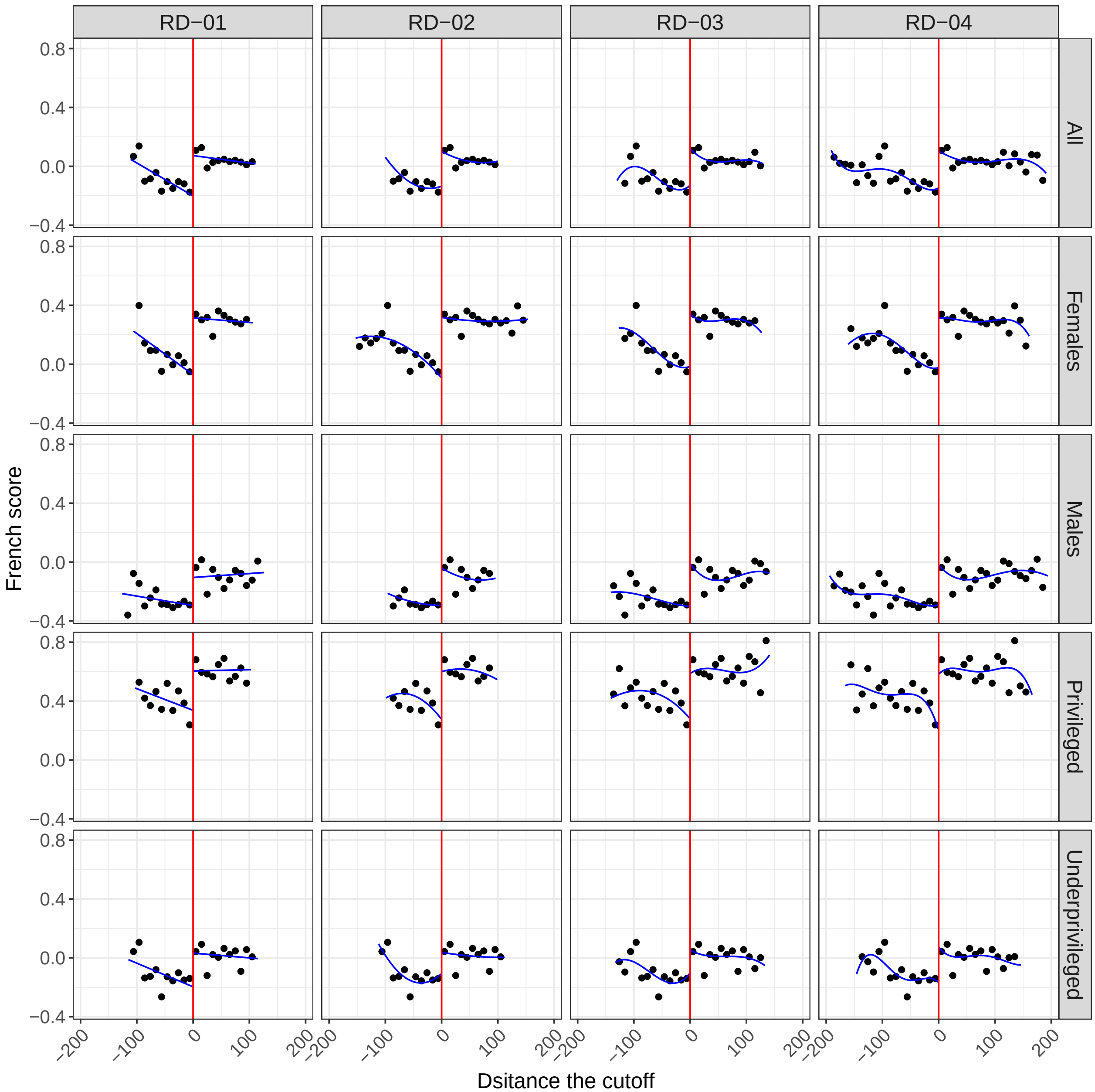

Figure 15: Regression discontinuity results plots (reduced form)

4.3 Sensitivity checks results

4.3.1 Cross-sectional cases

In this paragraph, an analysis of the estimation results when different models, different isntruments, different samples and different dependent variables is performed : first on a case specific case that is numerically informative then on a highly general case with a figure covering estimation results of all the possible combination of the aforementioned dimensions for each cohort.

The estimates resulting from alternative forms of the instruments and models are presented in Table 13. Columns (1) and (2) correspond to 2SLS estimates, (3) to (5) to control function and (6) and (7) to reduced forms ones. The five set of lines *I*, *II*, *III*, *IV* reports the three cross-sectional cohorts using respectively the assigned relative age (used to report the principal results), the fortnite of birth (which equals to 1 if born in the first half of January, 2 if born in the second half of January, 3 if born in the first half of February and so on), the month of birth, the bimester of birth and the quarter of birth as instruments. For clarification purpose, this table is limited to the cas in which the whole sample is used, and the total score is the dependent variable. A more complete illustration will be presented afterwards. First, except for the column (7), which we will not consider as discussed earlier, the coefficients are all significant at the 1% level. Then, by comparing the models with their alternatives, there is no essential difference (apart from the fact that the 2SLS-sat – column (2) alternative coefficients is very slightly lower than those of the column (1)) in the resulting estimates, which is a fact that credits our choice to use the simplest models to render the previous results. Also, no dramatical difference across the different use of instruments could be found. To be precise, let us provide some figures about the magnitudes of these estimates. On average, apart from the reduced form model, in 2010, the assigned relative age yields to the highest magnitude, the fortnite of birth yields to the lowest. Respectively, in 2011, the corresponding instruments are the assigned relative age and the quarter of birth ; in 2012, those are (surprisingly) the quarter of birth and the month of birth. For the last cohort 2012, the assigned relative age is placed after the quarter of birth if one rank the instruments by which one yields the highest magnitude. The maximum difference between the magnitudes equals to $+0.304 - (+0.191) = +0.113$ s.d. This minimum magnitude corresponds to the saturated 2SLS model which use the fortnite of birth as instrument in 2012 (column (2), line 6). Note how the 2012 often produces exceptional features. The first explanation that comes to mind is the presence of missing observations for the sex variable. If we reproduce the same instrument comparison but solely in the reduced form model, we note one interesting regularity : for each cohort, on average, it is the fortnite of birth that gives the highest estimates while the quarter of birth gives the lowest estimates. How to interpret this last remark ? It is possible that the date of birth effects are mostly driven by the variations of the fortnite of birth instead of those of the exact date of birth. Under the reserve of deeper analysis of this possibility, we could conclude that being in possession of the fortnite of birth of individuals is sufficiently good for the interested researcher.

As mentioned earlier, this table is very case-specific, a more general sensitivity analysis is illustrated in Figure 18 in the appendix, eventhough numerical observations as just above would be too cumbersome due to the number of regression performed. Let us now detail how to read this figure. It report in the y axis estimated coefficients on age estimates while the x axis represent the pair **sample-dependent variable** on which the regression is performed. The difference in the instruments are illustrated inside each panel with the different lines. Last, the cohort dimension is represented by a column of panels while the model dimension is represented by a line of panels. One could have noticed the estimates which values fall exactly intto zero : these are estimates that are non significant at least at the 5% level.

Once this figure described, we now provide some observations and interpretations on some interesting patterns. First, we could concentrate on the significance pattern : neither the 2SLS nor saturated 2SLS model provide non significant estimates across the three cohorts. The other models (line by line observation) yield in at least one cohort some non significant estimates. More interestingly, the second extension of the control function approach, in 2011 and 2012 leads to non significant estimates for all the case in which the underprivileged's children are used for the estimation. It seems that once the aspect of heterogeneity controled, there is no more age effect for the underprivileged pupils. Notice also how the assigned relative age instrument does not drive to any case in which the estimates are non significant while its nearest variant – the fortnite of birth – does not present comparable behavior (there are few cases of non significance one the fortnite of birth is used).

From a model comparison perspective now, one striking observation is the robustness of the results with the instrument and the **sample-dependent variable** within the 2SLS line of panels compared to its saturated version (second line of panels). In fact, in the first case, the values seems to be « compressed » between +0.2 and +0.35, while the it is less the case for the saturated 2SLS model. Then, another appealing feature is the apparent exact equality between the estimates of the control function approach (the CFH line of panels) and those of its first extension (CFH-E1 line of panels). Actually, there is some **sample-dependent variable** on which these two models yields different estimates, but the gap is too small to be detectable (I checked the table on which this figure is based on ; that is the reason of such affirmation). The second extension (CFH-E2 line of panels) behave quite differently from its two alternatives, except for the cas ine which the assigned relative age is used as instrument. In the control function models, except with the assigned relative age and fortnite of birth, for each value of the x axis, the estimates also appear compressed (i.e robust to the form of the instument). Concerning the reduced form estimates, however, there is no such robustness of the estimates (the magnitudes are varying more with different instruments). One pattern could be however mentioned in the reduced form model : until the **Males-Maths** case (on the x axis, from the very left), the decrasing rank of the magnitudes for each cohort are generally (with some exceptions) obtained using the instruments : fortnite of birth, month of birth, assigned relative age, bimester of birth then quarter of birth. Notice in this ranking that the assigned relative age is exactly in the « middle » ; the two less aggregating forms (month of birth and fortnite of birth) are on its left in this order (i.e yielding higher estimates in this order) and the two most aggregating forms (bimester and quarter of birth) are on its right in this order (i.e yielding lower estimates in this order).

Last, as we could often observe, the 2012 results behave diffently in the sens that from a **sample-dependent variable** to another : there is not as much variation in the value of estimates as in 2010 or 2011 (for example, when comparing the 2012 2SLS model with the 2010 and 2011 one, we could observe that the line polt is « smoother » in 2012 while it is more in a sawtooth form in

2010 and 2011). Similar remark can be made for the reduced form model.It suggests that there may be some fundamental difference that was not accounted for concerning the 2012 cohort, given the several exceptions in terms of results in this cohort. This is leaved to future research.

4.3.2 Regression discontinuity cases

Several sensitivity analysis seems necessary to check in the RDD. Most of them are related to the bandwidth choice h, i.e the data used to estimate the regression discontinuity equations are limited to individuals verifying $-h \leq dist_i \leq h$. First, recall that for the covariate balance regressions (Table 20 in the appendix), to ensure the idea of locality around the cutoff $dist = 0$, I limited the data to $h = 30$. However, as the regression results are performed using case specific bandwiths which are largers than 30 days, it is then more plausible for the framework if these case specific bandwidth were used for the covariate balance regressions. As the Table 21 demonstrates in the appendix, the corresponding results is even more crediting since no significance at the 5% level is detected (line 2).

Second, a practice that would credit the validity of the instrumentation of the age at test by the year of birth and the controlling by different polynomials of birthdates (fuzzy regression discontinuity) is to check the sensitivity of the first stage results. See the Figure 19 that provides the results of the mentioned exercise. Recall the idea that the x axis represent the deviation from the case specific optimal bandwidths instead of absolute values of h on which the estimates are based. The dashed lines in the figure which represent the 5% level confidence intervals tell us first that all the estimates are largely significant significant at the 5% level except for extreme values (i.e for a -100 days deviation from optimal bandwith for example, which leaves to very few number of observations). Moreover, the vertical width of the confidence intervals are unsurprisingly shrinking as one move to the right on the x axis (i.e as one include more and more individuals in the regressions since one extends the bandwidth). Also, when the bandwidths are extended, there are generally no dramatical change in the estimates. If one move to the left on the x axis, considerable variations in the estimates are observed at extreme deviations. Additionally, notice the outstanding robustness and proximity to the cross-section first stage estimate (in 2010) of the first stage estimates when all covariates values are included in the regression data and when a polynomial of order 1 is used (top left panel).

Overall, we can conclude that the instrumentation of the age at test with the year of birth indicator (old) and the controlling with birthdate polynomial is generally bandwidth-choice-robust in first stages, apart from extremely narrow bandwidths.

Third, and probably the most important sensitivity check in RDD, is the sensitivity of the estimates of interest themselves to the bandwidth choice. Indeed, a highly fluctuating estimates to little variations of the bandwidth choices would cast doubt on the fiability of the results and the setup itself. Similarly as in the cross-sectional models and instrument sensitivity analysis, I propose to first give numerical but limited information results (a regression table with different bandwidth choices) and a overall view of with a more general figure like the Table 13 and the Figure 18 respectively. See the Table 23 in the appendix for the first case. The regressions in this table are performed solely with all individuals with no covariate value restriction. The estimates are all significant at usual level and a vertical comparison of the estimates (i.e with different bandwidth choice given a specification) does not highlight a any great variation. Following form this, the Figure 20 in the appendix shows with a higher level of generality how the regression discontinuity reduced form estimates and its inferences are varying with the bandwidth choices. The case of the fuzzy regression discontinuity is handled by the Figure 21. Concerning the first figure, when one moves to the left of the x axis, the intention-to-treat estimates generally get higher magnitudes but also less precision. It is logical considering that less observations are used for the estimations and the age differentials within the used observations are on average wider. In the case one moves to the right i.e extending the bandwidth, the magnitude converges (especially in the « Poly 01 » column of panels, less in the « Poly 02 » column of panels) to the cross-section reduced form estimate. This is normal since as one extends the bandwith, less differentials in age (between those having $old = 1$ and those having $old = 0$) are mechanically accounted in the data. In terms of precision, the confidence intervals are getting wider as one use higher polynomial order, the most precise estimates appearing to be obtained with a polynomial of order 1 specification. Notice that, the lines formed by the magnitude estimates in the Figure 21 appear to follow the exact same form of those in the Figure 20 but placed slightly (vertically) higher, i.e are higher in magnitude. This is logical because of the identity such that a 2SLS estimate equals to the quotient of the reduced form estimate with its first stage. We can assess that the instrument is robustly strong.

An other form of sensitivity investigation one can perform, despite the invalidation of the McCrary test for the case with the advanced ones, is to nevertheless confront the two cases : the RDD estimates performed without the advanced pupils and those performed with. Note that the social category variable not available for second case. One can quickely check the age effect estimates on french test scores in grade 5 results in the Table 18 in the appendix for this purpose and confront them with the results of the principal results in Table 12. The structure of the former table follows the previous ones to ensure its easy reading. Comparing these two tables tell us that the effect estimated with the alternative (with advanced ones) yields generally higher estimates in magnitude. One possible reason of this would be that the proportion of advanced pupils at the right of the cutoff are more numerous than those at the left of the cutoff. This appear to be empirically the case **around the cutoff**. The Figure 16 is constructed to support this affirmation. In this figure, one can indeed check that poportion of born 1999 advanced pupils are much lower compared to the other positions around the cutoff (here, h is up to 195 days). More interestingly, a sharp jump in these quantity is observed for the advanced pupils between the 110 and 115 bandwidths, corresponding respectively to the birthdate intervals of **[September, 08th of 1999 ; April, 20th of 2000]** and **[September, 13rd of 1999 ; April, 20th of 2000]**. Remind that rejection of the McCrary density test in the alternative framework could be due to this feature. Especially, the alternative fuzzy regression estimates yield to estimates that go up to +0.6 s.d, for example in column (3), line 5 of the Table 18 while the no-advanced case corresponding estimates equals to +0.45 s.d (column (8) line 2 of Table 12). Last, as in the previous cases, I also propose two general sensitivity plots for the alternative cases : the RD reduced form and fuzzy regression discontintuity sensitivity to bandwidth choices in the case the advanced pupils are included into the observations used to perform the regressions are illustrated respectively by the Figures 22 and 23. The identity of the form of the lines between the first and the second figure is of course maintained. The only striking difference between these two figures and the figures for the no-advanced case is that the formers confidence intervals lower bounds are less near to zero, meaning that these produce more significant estimates. Also, figure patterns exposed earlier for the no-advanced case are retrieved in the alternative case : a most converging (to the cross sectional results) in the polynomial of order 1 specification, largest confidence intervals in the polynomial of order 4 specification.

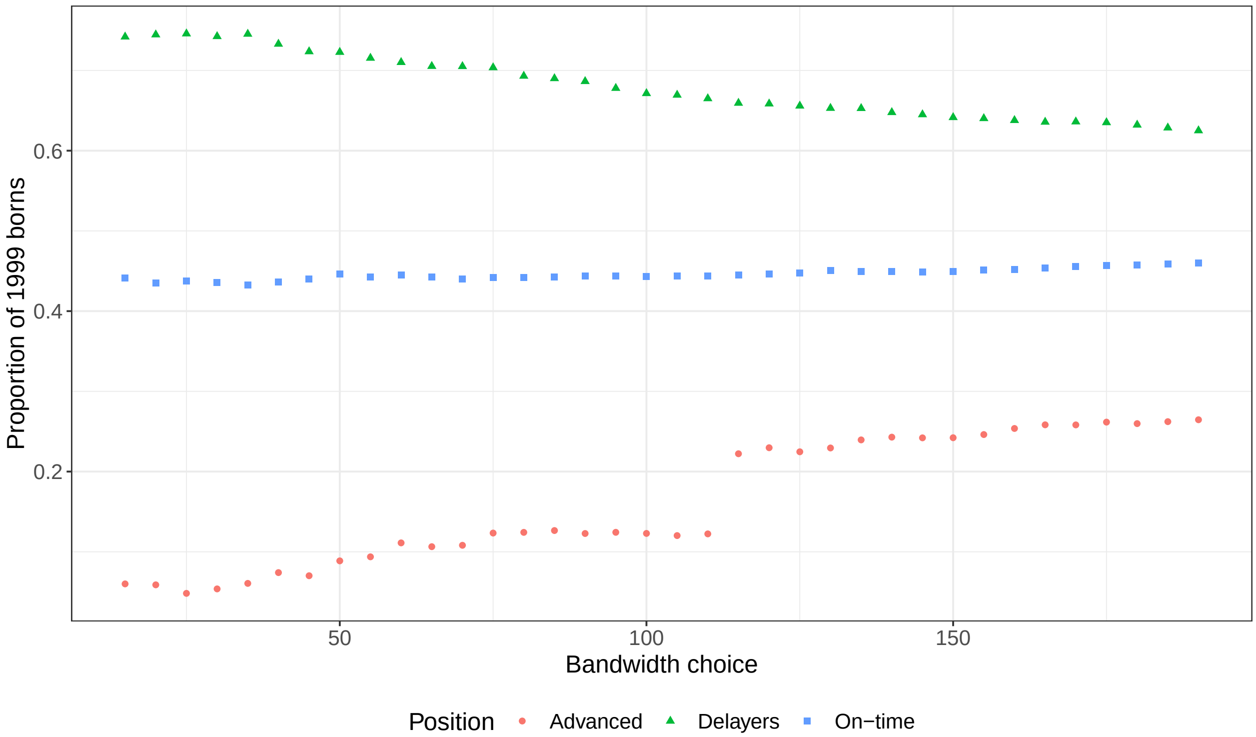

Figure 16: Proportions of 1999 borns within different bandwidths

4.4 How about the monotonicity condition being unverified ?

A first idea of discussion of the 2SLS estimates interpretations can be drawn by comparing them to the CFH results (column (4) to (6) in this order versus (7) to (9) in this order). As the beginning of the reasoning, recall that the control function approach permits to recover an average treatment effect instead of a LATE. Thus, if one compare control function estimates with 2SLS estimates and robustly similar results, one could think that the unidentified 2SLS estimation results are an approximation of an average treatment effect. We begin such comparison unconditionally to background characteristics, by confronting the column (4) to (6) in this order to the columns (7) to (9) in this order of Table 6. Notice how the 2SLS that theoretically do not identify any interesting estimates yield to very similar results to a model that recover an average treatment effect. In fact, for the two models, the magnitudes of the estimates take values between $+0.27$ s.d and 0.3 s.d. Moving deeper to ensure the robustness of such pattern, we now use the Table 8 to compare the two models. Recall that this table differ from the first mentioned in this subsection in the dependent variables used : french score and maths score. Again, in columns (1) and (2) of this table, the CFH estimates arguably do not differ from 2SLS ones. This feature is easily verified either in the french score dependent variable case or the maths score dependent variable case.

Following the same process, we can now take a look at the Table 9. This table, for reminder, present subgroups estimations results. We hardly find noticeable difference in the two type of estimates (column (1) versus (2), column (4) versus (5) and column (7) versus (8)). Even in Table 13 which we used to numerically assess estimates sensitivity to models and instruments do not report, striking differences between its column (1) and column (3) (corresponding respectively to the 2SLS and the CFH results). Overall, the Figure 18 sums up all these observations. It highlight clearly that instead of using other form of the instrument than assigned relative age, the 2SLS and the CFH models (see the line of panels that are labelled « 2SLS » and « CFH ») yields to **generally** similar quantities.

A very plausible reason of this pattern was already evoked by Black, Devereux, and Salvanes (2011) : they argue that they approximate the average treatment effect despite of the monotonicity problem because of a « high compliance ». In fact, we could think that the proportion of compliers is strongly related, in the present framework, to the proportion of on-time pupils. Note that this is limited to be hypothetical since the prortion of compliers can not be directly computed by definition.

Similarly in the RD design, monotonicity violation theoretically collapsing all interesting interpretation of the estimates. But since one argue that there is a high proportion of pupils that have $old_i = 1$ and observed at the 2011 cohort (symmetrically there is a high proportion of pupils that have $old_i = 0$) and are observed at the 2010 cohort, one could give credit to the, although untestable assumption, of a high proportion of compliers, leading the bias due to the monotonicity violation tolerable, compared to the average treatment effect. To support this idea and being in line the RDD framework, refer to the Figure 17 that illustrates the proportions, within 30-days bins, of those observed in the assigned cohort (as explained above). Note how these proportions does not go below 70%. The maximum values observed are up to 87%.

5 Conclusion

Overall, this paper performed several empirical procedures in the attempt to measure the effect of age of test on educational performances at the end of primary school in Reunion Island. It were expected that the age effects are strong because of the early age of measurement. Several models were run. First the OLS results yielded expectedly negative estimates, a feature that is in line with most of the literature. Second, by exploiting the variation of age at test induced by a specific form of date for birth – the assigned relative age, a 2SLS model is used to measure the interested effect. Estimates around $+0.2$ s.d and 0.3s.d were found. This quantitites are in line with several contemporary studies. See especially the literature review of Peña (2017) that report very clearly the results of similar studies. These quantities makes sens considering the rule of thumb of a $+0.33$ s.d representing a « learning gain » over a year. This learning gain is here expressed as test scores. The effects appeared to be stronger for girls compared to those of boys, arguably due to the fact that girls are maturing with higher rate than boys do. This heterogeneity is also inline with the literature, see Dhuey et al. (2017) for an extensive citation of the heterogeneity of age effects conditionally to background charasteristics patterns. One result that did not meet the excpectations concern the estimates performed by social category of parents subgroups. In fact, there were hardly robust patterns telling that in which of the privileged and underprivileged's

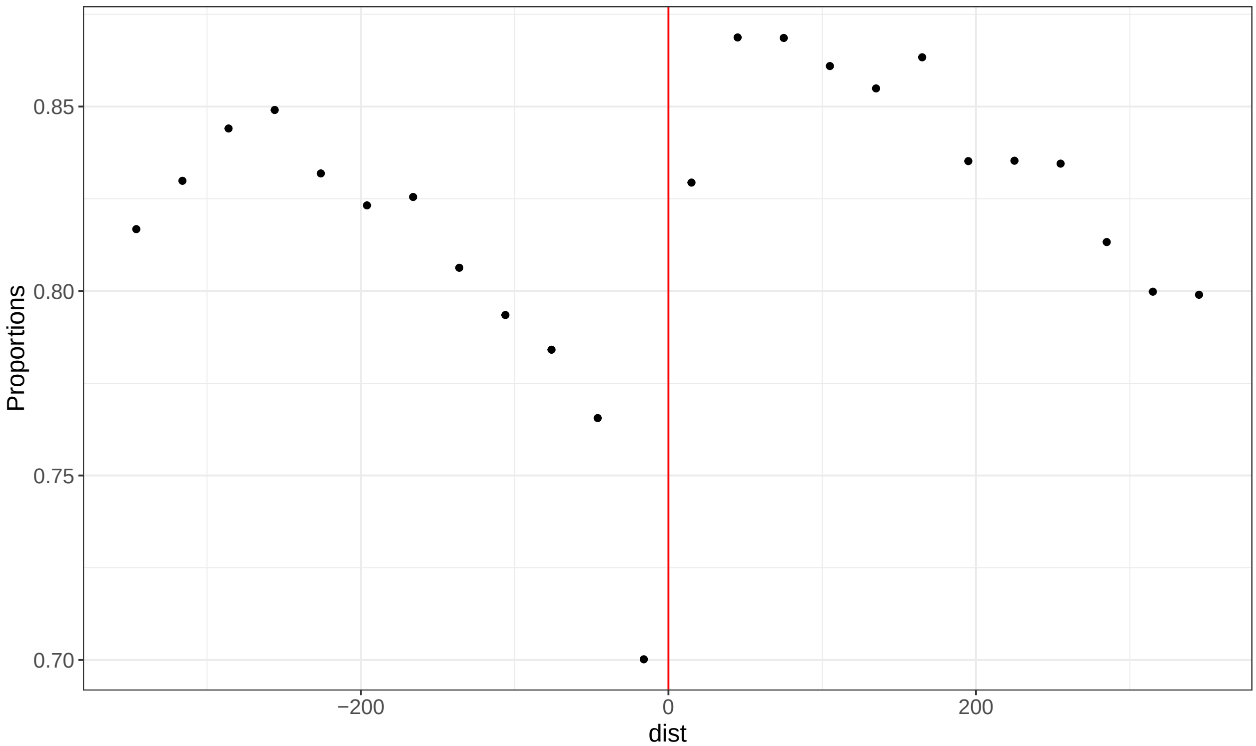

Figure 17: Proportions of pupils observed in the assigned cohort (30-days bins)

children the effects are stronger. Those are the principal 2SLS estimates. Unfortunately, many evidences cast doubts about the validity of the identification via this model. Thus, I performed a control function approach taking into account heterogeneity of age effects (yet labelled similarly, this latter does not designate the same concept as the heterogeneity conditionally to observable characteristics such as the sex), that theoretically recover an average treatment effect. With noticing that, robustly, the 2SLS results are similar to the control function one, I propose the idea that the 2SLS gives in this precise framework an approximation of an average treatment effect. The reduced forms results, geenrally lower than 2SLS results, suggest that there is a positive effect of repetition / redshirting. This affirmation is to be confronted to more laborous studies performed by several authors. To ensure that these results were not mostly as we expected by luck, I propose sensitivity analysises in several ways. A first possibility is to check how different form of the instrument and alternative model specifications affect the significance and magnitude of the estimates. The results were not generally dramatically sensitive to those paramters. Also, I implement a regression discontinuity design that yielded, in most case, comparable estimates to cross-sectional results (and I even ensure that these RDD results are quite robust). This comparability of RDD estimates is eventhough conditioned on a good choice of specification. One limit of the present paper is the lack of interpretability of the interesting patterns of the data or results and the extensive use of hypothetization rather than evidences. This has to be overcomed in the future with the acquisition of necessary informations to priorize evidences. A second limit is probably the lack of background covariates within the models, not allowing us to explain more thana about 20% of the variation of the test scores. Another worth mentionning limit of this study is the mixed characteristics of the estimates age effects in the sens that they contain, in addition to the age at test effects the age at school entry effects.

On the other hand, recall that this paper is of policy perspective interest because of the clear evidences that there is definitely age effects in Reunion Island, eventhough at early ages. This study appears to be the first to do so. With more datas and more elaborated framework, more knowledge about the age effects could be ivestigated for sure. For example, it is possible to obtain the grade 9 national assessment scores for all the grade 5 pupils within the data of this study. Such data can be exploit to assess if the early age effects last several years after the grade 5. If it is, then it gives more importance to the evidence of early age effects and the necessity to act on these (by educational policies) because some children, in addition to perform worse in grade 5, will perform worse till grade 9 just because of their date of birth. This question is reserved for future research.

As it is proposed as well as in several studies, three maneers of acting on age are implementable by policy maker : the rise of the age at school entry, the normalisation of national tests by date of birth or the regulation of classrooms age distributions.

Appendix

Table 13: Sensitivity analysis of the estimates to models and instruments

	IV		Control functions			Reduced forms	
	2SLS (1)	2SLS-sat (2)	CFH (3)	CFH-E1 (4)	CFH-E2 (5)	RF (6)	RF-sat (7)
Instrument : assigned relative age							
1 - 2010							
	0.304***	0.299***	0.301***	0.304***	0.304***	0.257***	0.158
	(0.033)	(0.033)	(0.038)	(0.038)	(0.039)	(0.026)	(0.322)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
2 - 2011							
	0.302***	0.301***	0.3***	0.301***	0.295***	0.244***	0.609**
	(0.031)	(0.031)	(0.036)	(0.036)	(0.037)	(0.024)	(0.266)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
3 - 2012							
	0.269***	0.266***	0.265***	0.264***	0.25***	0.226***	0.319
	(0.034)	(0.034)	(0.042)	(0.042)	(0.046)	(0.027)	(0.294)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734
I - Instrument : fortnite of birth							
4 - 2010							
	0.295***	0.221***	0.293***	0.297***	0.295***	0.298***	−0.94
	(0.033)	(0.03)	(0.038)	(0.039)	(0.039)	(0.051)	(0.685)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
5 - 2011							
	0.296***	0.241***	0.293***	0.295***	0.287***	0.345***	0.542
	(0.031)	(0.03)	(0.036)	(0.036)	(0.037)	(0.047)	(0.38)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
6 - 2012							
	0.261***	0.191***	0.257***	0.256***	0.236***	0.28***	0.96**
	(0.034)	(0.033)	(0.042)	(0.043)	(0.046)	(0.057)	(0.478)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734
II - Instrument : month of birth							
7 - 2010							
	0.298***	0.259***	0.296***	0.3***	0.297***	0.264***	0.092
	(0.033)	(0.032)	(0.039)	(0.039)	(0.039)	(0.038)	(0.406)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
8 - 2011							
	0.298***	0.282***	0.296***	0.297***	0.291***	0.297***	0.673*
	(0.031)	(0.031)	(0.036)	(0.036)	(0.037)	(0.034)	(0.399)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
9 - 2012							
	0.259***	0.213***	0.255***	0.254***	0.229***	0.197***	0.559
	(0.034)	(0.033)	(0.043)	(0.043)	(0.046)	(0.037)	(0.419)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734
III - Instrument : bimester of birth							
10 - 2010							
	0.299***	0.275***	0.297***	0.3***	0.297***	0.211***	0.066
	(0.033)	(0.033)	(0.039)	(0.039)	(0.039)	(0.027)	(0.333)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
11 - 2011							
	0.293***	0.285***	0.29***	0.291***	0.284***	0.204***	0.322
	(0.031)	(0.031)	(0.036)	(0.036)	(0.037)	(0.024)	(0.244)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
12 - 2012							
	0.262***	0.244***	0.258***	0.257***	0.232***	0.172***	0.106
	(0.035)	(0.034)	(0.043)	(0.043)	(0.046)	(0.027)	(0.268)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734
IV - Instrument : quarter of birth							
13 - 2010							
	0.296***	0.285***	0.293***	0.296***	0.291***	0.19***	0.134
	(0.033)	(0.033)	(0.04)	(0.04)	(0.04)	(0.021)	(0.265)
	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561	N = 13561
14 - 2011							
	0.289***	0.284***	0.286***	0.287***	0.278***	0.175***	0.457**
	(0.032)	(0.032)	(0.037)	(0.037)	(0.038)	(0.02)	(0.223)
	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622	N = 14622
15 - 2012							
	0.274***	0.259***	0.268***	0.267***	0.235***	0.163***	0.253
	(0.036)	(0.036)	(0.044)	(0.044)	(0.046)	(0.022)	(0.235)
	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734	N = 10734

Table 14: Social category subgroup results (part 1)

	Dep.var : total scores			Dep.var : french scores			Dep.var : maths scores		
	2SLS	CFH	RF	2SLS	CFH	RF	2SLS	CFH	RF
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
I - Farmers									
1 - 2010									
	0.948**	0.89	0.952***	0.698*	0.658	0.702**	1.206***	1.13*	1.211***
	(0.383)	(0.587)	(0.338)	(0.367)	(0.587)	(0.332)	(0.406)	(0.587)	(0.364)
	N = 198	N = 198	N = 198	N = 198	N = 198	N = 198	N = 198	N = 198	N = 198
2 - 2011									
	−0.147	−0.152	−0.097	−0.283	−0.279	−0.187	0.063	0.047	0.041
	(0.55)	(0.851)	(0.369)	(0.479)	(0.851)	(0.325)	(0.649)	(0.851)	(0.426)
	N = 154	N = 154	N = 154	N = 154	N = 154	N = 154	N = 154	N = 154	N = 154
3 - 2012									
	1.012*	0.498	0.869*	0.991*	0.487	0.851*	0.924	0.456	0.794
	(0.565)	(0.762)	(0.447)	(0.54)	(0.762)	(0.425)	(0.604)	(0.762)	(0.491)
	N = 156	N = 156	N = 156	N = 156	N = 156	N = 156	N = 156	N = 156	N = 156
II - Entrepreneurs									
4 - 2010									
	0.11	0.111	0.1	0.068	0.069	0.062	0.159	0.16	0.145
	(0.16)	(0.222)	(0.145)	(0.159)	(0.222)	(0.144)	(0.168)	(0.222)	(0.151)
	N = 710	N = 710	N = 710	N = 710	N = 710	N = 710	N = 710	N = 710	N = 710
5 - 2011									
	0.027	0.019	0.021	0.009	−0.001	0.007	0.05	0.045	0.039
	(0.199)	(0.307)	(0.152)	(0.206)	(0.307)	(0.158)	(0.202)	(0.307)	(0.154)
	N = 510	N = 510	N = 510	N = 510	N = 510	N = 510	N = 510	N = 510	N = 510
6 - 2012									
	0.432**	0.439*	0.394***	0.373**	0.375	0.34**	0.47**	0.482**	0.429**
	(0.177)	(0.23)	(0.152)	(0.168)	(0.23)	(0.145)	(0.193)	(0.23)	(0.166)
	N = 642	N = 642	N = 642	N = 642	N = 642	N = 642	N = 642	N = 642	N = 642
III - Executives									
7 - 2010									
	0.283**	0.283*	0.227**	0.243**	0.243	0.195**	0.309**	0.309*	0.248**
	(0.131)	(0.164)	(0.101)	(0.123)	(0.164)	(0.096)	(0.146)	(0.164)	(0.114)
	N = 1040	N = 1040	N = 1040	N = 1040	N = 1040	N = 1040	N = 1040	N = 1040	N = 1040
8 - 2011									
	0.39***	0.452***	0.293***	0.39***	0.445**	0.292***	0.348**	0.412**	0.261***
	(0.129)	(0.175)	(0.095)	(0.129)	(0.175)	(0.094)	(0.135)	(0.175)	(0.099)
	N = 1068	N = 1068	N = 1068	N = 1068	N = 1068	N = 1068	N = 1068	N = 1068	N = 1068
9 - 2012									
	0.185	0.205	0.151	0.209	0.233	0.171	0.128	0.139	0.104
	(0.132)	(0.173)	(0.105)	(0.132)	(0.173)	(0.105)	(0.133)	(0.173)	(0.107)
	N = 978	N = 978	N = 978	N = 978	N = 978	N = 978	N = 978	N = 978	N = 978
IV - Intermediates									
10 - 2010									
	0.09	0.104	0.077	0.04	0.053	0.034	0.155	0.167	0.133
	(0.159)	(0.191)	(0.135)	(0.151)	(0.191)	(0.129)	(0.175)	(0.191)	(0.148)
	N = 842	N = 842	N = 842	N = 842	N = 842	N = 842	N = 842	N = 842	N = 842
11 - 2011									
	0.351***	0.371**	0.305***	0.309**	0.33**	0.268**	0.372***	0.387**	0.324***
	(0.133)	(0.161)	(0.11)	(0.134)	(0.161)	(0.113)	(0.136)	(0.161)	(0.113)
	N = 986	N = 986	N = 986	N = 986	N = 986	N = 986	N = 986	N = 986	N = 986
12 - 2012									
	0.303*	0.31*	0.269**	0.332**	0.339*	0.295**	0.224	0.229	0.199
	(0.158)	(0.185)	(0.134)	(0.154)	(0.185)	(0.13)	(0.163)	(0.185)	(0.141)
	N = 934	N = 934	N = 934	N = 934	N = 934	N = 934	N = 934	N = 934	N = 934

Table 15: Social category subgroup results (part 2)

	2SLS (1)	CFH (2)	RF (3)	2SLS (4)	CFH (5)	RF (6)	2SLS (7)	CFH (8)	RF (9)
I - Employees									
1 - 2010	0.286*** (0.084) N = 2053	0.296*** (0.097) N = 2053	0.269*** (0.075) N = 2053	0.235*** (0.086) N = 2053	0.245** (0.097) N = 2053	0.222*** (0.078) N = 2053	0.327*** (0.082) N = 2053	0.334*** (0.097) N = 2053	0.308*** (0.073) N = 2053
2 - 2011	0.29*** (0.089) N = 2189	0.289*** (0.11) N = 2189	0.241*** (0.071) N = 2189	0.305*** (0.088) N = 2189	0.303*** (0.11) N = 2189	0.254*** (0.07) N = 2189	0.237*** (0.091) N = 2189	0.238** (0.11) N = 2189	0.197*** (0.074) N = 2189
3 - 2012	0.288*** (0.099) N = 1944	0.282** (0.115) N = 1944	0.24*** (0.078) N = 1944	0.265*** (0.099) N = 1944	0.258** (0.115) N = 1944	0.22*** (0.078) N = 1944	0.289*** (0.099) N = 1944	0.283** (0.115) N = 1944	0.241*** (0.078) N = 1944
II - Workers									
4 - 2010	0.402*** (0.116) N = 1523	0.39*** (0.149) N = 1523	0.311*** (0.084) N = 1523	0.387*** (0.119) N = 1523	0.384*** (0.149) N = 1523	0.3*** (0.086) N = 1523	0.374*** (0.115) N = 1523	0.351** (0.149) N = 1523	0.29*** (0.085) N = 1523
5 - 2011	0.246 (0.161) N = 957	0.253 (0.195) N = 957	0.192 (0.12) N = 957	0.272 (0.167) N = 957	0.276 (0.195) N = 957	0.213* (0.124) N = 957	0.18 (0.159) N = 957	0.192 (0.195) N = 957	0.141 (0.121) N = 957
6 - 2012	0.244** (0.112) N = 1403	0.24* (0.137) N = 1403	0.211** (0.092) N = 1403	0.205* (0.112) N = 1403	0.202 (0.137) N = 1403	0.177* (0.093) N = 1403	0.272** (0.114) N = 1403	0.269** (0.137) N = 1403	0.235** (0.095) N = 1403
III - Retired									
7 - 2010	2.869 (4.318) N = 114	3.348 (5.102) N = 114	0.835 (0.542) N = 114	0.717 (2.551) N = 114	1.036 (5.102) N = 114	0.209 (0.621) N = 114	5.758 (7.213) N = 114	6.419 (5.102) N = 114	1.676*** (0.512) N = 114
8 - 2011	5.289 (7.512) N = 97	5.028 (3.774) N = 97	1.302* (0.737) N = 97	2.586 (4.469) N = 97	2.395 (3.774) N = 97	0.637 (0.694) N = 97	8.57 (11.431) N = 97	8.239** (3.774) N = 97	2.111** (0.746) N = 97
9 - 2012	2.285 (5.816) N = 87	1.245 (3.797) N = 87	0.726 (1.211) N = 87	2.105 (5.69) N = 87	1.055 (3.797) N = 87	0.669 (1.197) N = 87	2.282 (5.38) N = 87	1.382 (3.797) N = 87	0.726 (1.127) N = 87
IV - Unemployed									
10 - 2010	0.356*** (0.06) N = 4776	0.351*** (0.071) N = 4776	0.287*** (0.044) N = 4776	0.36*** (0.062) N = 4776	0.353*** (0.071) N = 4776	0.29*** (0.046) N = 4776	0.306*** (0.057) N = 4776	0.302*** (0.071) N = 4776	0.246*** (0.044) N = 4776
11 - 2011	0.263*** (0.054) N = 5248	0.267*** (0.066) N = 5248	0.206*** (0.041) N = 5248	0.248*** (0.056) N = 5248	0.251*** (0.066) N = 5248	0.195*** (0.042) N = 5248	0.254*** (0.052) N = 5248	0.261*** (0.066) N = 5248	0.2*** (0.039) N = 5248
12 - 2012	0.205*** (0.052) N = 4541	0.199*** (0.069) N = 4541	0.171*** (0.042) N = 4541	0.183*** (0.052) N = 4541	0.174** (0.069) N = 4541	0.153*** (0.043) N = 4541	0.214*** (0.054) N = 4541	0.212*** (0.069) N = 4541	0.179*** (0.044) N = 4541

Table 16: Discontinuity check regression results

	old = 0				old = 1			
	spe-01 (1)	spe-02 (2)	spe-03 (3)	spe-04 (4)	spe-01 (5)	spe-02 (6)	spe-03 (7)	spe-04 (8)
1 - All								
	0.02	−0.023	0.014	−0.045	−0.056	-0.133^{***}	−0.098	-0.155^{*}
	(0.034)	(0.052)	(0.069)	(0.087)	(0.034)	(0.052)	(0.069)	(0.086)
	N = 13344	N = 13344	N = 13344	N = 13344	N = 13970	N = 13970	N = 13970	N = 13970
2 - Females								
	0.065	−0.069	−0.115	−0.183	−0.064	−0.068	−0.104	−0.148
	(0.047)	(0.07)	(0.093)	(0.115)	(0.045)	(0.068)	(0.091)	(0.112)
	N = 6577	N = 6577	N = 6577	N = 6577	N = 6705	N = 6705	N = 6705	N = 6705
3 - Males								
	−0.02	0.015	0.127	0.064	−0.064	-0.166^{**}	−0.042	−0.113
	(0.049)	(0.074)	(0.1)	(0.126)	(0.049)	(0.075)	(0.101)	(0.126)
	N = 6767	N = 6767	N = 6767	N = 6767	N = 6508	N = 6508	N = 6508	N = 6508
4 - Underprivileged								
	0.043	−0.028	−0.028	−0.042	−0.007	−0.054	−0.037	−0.155
	(0.04)	(0.06)	(0.08)	(0.099)	(0.041)	(0.062)	(0.083)	(0.104)
	N = 8482	N = 8482	N = 8482	N = 8482	N = 8512	N = 8512	N = 8512	N = 8512
5 - Privileged								
	−0.083	−0.037	0.055	−0.12	−0.052	−0.086	−0.027	−0.107
	(0.072)	(0.108)	(0.146)	(0.19)	(0.071)	(0.105)	(0.135)	(0.165)
	N = 2449	N = 2449	N = 2449	N = 2449	N = 2440	N = 2440	N = 2440	N = 2440

Table 17: Discontinuity check regression alternative results

	old = 0				old = 1			
	spe-01 (1)	spe-02 (2)	spe-03 (3)	spe-04 (4)	spe-01 (5)	spe-02 (6)	spe-03 (7)	spe-04 (8)
1 - All								
	0.071^{**}	−0.009	0.098	0.075	−0.032	-0.099^{**}	−0.031	−0.067
	(0.035)	(0.052)	(0.069)	(0.087)	(0.032)	(0.049)	(0.066)	(0.082)
	N = 13620	N = 13620	N = 13620	N = 13620	N = 14235	N = 14235	N = 14235	N = 14235
2 - Females								
	0.132^{***}	−0.038	0.017	−0.019	−0.033	−0.019	−0.012	−0.009
	(0.047)	(0.069)	(0.092)	(0.115)	(0.043)	(0.065)	(0.087)	(0.108)
	N = 6716	N = 6716	N = 6716	N = 6716	N = 6864	N = 6864	N = 6864	N = 6864
3 - Males								
	0.014	0.013	0.164	0.144	−0.053	-0.154^{**}	−0.012	−0.101
	(0.051)	(0.076)	(0.102)	(0.128)	(0.046)	(0.071)	(0.096)	(0.12)
	N = 6904	N = 6904	N = 6904	N = 6904	N = 6614	N = 6614	N = 6614	N = 6614

Table 18: Regression Discontinuity alternative results

	Dependent variable : french scores			
RF/2SLS	spe-01	spe-02	spe-03	spe-04
(1)	(2)	(3)	(4)	(5)
I - Intention-to-treat				
1 - All				
0.241^{***}	0.338^{***}	0.316^{***}	0.356^{***}	0.355^{***}
(0.027)	(0.043)	(0.067)	(0.076)	(0.082)
N = 13561	N = 8412	N = 7333	N = 10203	N = 13570
2 - Females				
0.305^{***}	0.455^{***}	0.471^{***}	0.407^{***}	0.42^{***}
(0.038)	(0.061)	(0.073)	(0.104)	(0.106)
N = 6666	N = 3674	N = 5734	N = 4983	N = 7357
3 - Males				
0.175^{***}	0.232^{***}	0.35^{***}	0.349^{***}	0.357^{***}
(0.04)	(0.058)	(0.097)	(0.102)	(0.11)
N = 6895	N = 4707	N = 3410	N = 5579	N = 7367
II - Fuzzy regression discontinuity				
4 - All				
0.285^{***}	0.407^{***}	0.446^{***}	0.424^{***}	0.435^{***}
(0.033)	(0.078)	(0.103)	(0.116)	(0.119)
N = 13561	N = 5150	N = 7085	N = 9644	N = 14017
5 - Females				
0.361^{***}	0.533^{***}	0.612^{***}	0.608^{***}	0.538^{***}
(0.048)	(0.122)	(0.138)	(0.182)	(0.187)
N = 6666	N = 2424	N = 4630	N = 4700	N = 6161
6 -Males				
0.205^{***}	0.352^{***}	0.467^{***}	0.398^{**}	0.434^{**}
(0.049)	(0.1)	(0.149)	(0.159)	(0.173)
N = 6895	N = 3057	N = 3449	N = 4782	N = 6620

Table 19: Proportions of positions by institutional features

	2010				2011				2012			
	N	Delayers	On time	Advanced	N	Delayers	On time	Advanced	N	Delayers	On time	Advanced
1 - Sex (%)												
Females	6666	13.14	84.47	2.39	7325	12.91	84.46	2.62	5461	9.49	88.41	2.11
Males	6895	20.78	77.68	1.54	7297	19.34	78.91	1.75	5273	14.22	83.90	1.88
2 - SPC (%)												
Farmers	198	15.15	82.83	2.02	154	9.74	89.61	0.65	166	9.64	89.76	0.60
Entrepreneurs	710	8.03	89.15	2.82	510	7.25	90.20	2.55	696	4.17	93.53	2.30
Executives	1040	3.17	90.58	6.25	1068	1.87	90.26	7.87	1068	1.87	89.61	8.52
Intermediates	842	5.70	91.69	2.61	986	4.56	92.80	2.64	1013	5.13	91.31	3.55
Employees	2053	10.57	88.21	1.22	2189	9.55	88.67	1.78	2138	7.81	90.27	1.92
Workers	1523	13.53	85.36	1.12	957	13.79	85.37	0.84	1518	12.52	86.43	1.05
Retired	114	12.28	85.09	2.63	97	11.34	80.41	8.25	98	14.29	81.63	4.08
Unemployed	4776	21.63	77.45	0.92	5248	19.89	79.12	0.99	5088	19.63	79.80	0.57
Others	2305	29.11	68.07	2.82	3413	24.73	72.66	2.61	1890	28.47	69.21	2.33
3 - SPC (grouped) (%)												
Underprivileged	8664	17.31	81.61	1.07	8645	16.32	82.43	1.25	9008	15.39	83.60	1.01
Privileged	2592	5.32	90.55	4.13	2564	3.98	91.22	4.80	2777	3.64	91.21	5.15
Others	2305	29.11	68.07	2.82	3413	24.73	72.66	2.61	1890	28.47	69.21	2.33
4 - School status (%)												
Privates	1086	4.79	93.09	2.12	1177	4.25	92.35	3.40	1148	3.75	93.12	3.14
Publics	12475	18.09	79.97	1.94	13445	17.16	80.76	2.08	12527	15.82	82.25	1.93
5 - Priority education network (%)												
HEP	7101	14.20	83.51	2.30	7703	13.84	83.38	2.78	7067	12.11	85.51	2.38
ECLAIR	3400	19.71	79.03	1.26	3594	19.42	78.85	1.73	3608	18.65	79.74	1.61
RRS	3060	20.62	77.45	1.93	3325	17.83	80.84	1.32	3000	16.53	81.73	1.73
6 - Priority education network (yes/no) (%)												
No	7101	14.20	83.51	2.30	7703	13.84	83.38	2.78	7067	12.11	85.51	2.38
Yes	6460	20.14	78.28	1.58	6919	18.66	79.81	1.53	6608	17.69	80.64	1.66

Table 20: Covariate balance regressions (30 days around the cutoff)

	Dep.var : Sex - Male				Dep.var : Social category - Privileged			
	spe-01	spe-02	spe-03	spe-04	spe-01	spe-02	spe-03	spe-04
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
old	−0.049	−0.037	−0.060	−0.036	0.022	0.069	0.134**	0.139*
	(0.043)	(0.065)	(0.088)	(0.116)	(0.032)	(0.045)	(0.060)	(0.078)
dist	0.002	−0.0001	0.008	−0.040	−0.002	−0.008	−0.031**	−0.056*
	(0.002)	(0.007)	(0.019)	(0.042)	(0.001)	(0.005)	(0.014)	(0.031)
dist ^{2}		−0.0001	0.001	−0.006		−0.0002	−0.002*	−0.006
		(0.0002)	(0.001)	(0.005)		(0.0002)	(0.001)	(0.004)
dist ^{3}			0.00001	−0.0003			−0.00004*	−0.0002
			(0.00003)	(0.0003)			(0.00002)	(0.0002)
dist ^{4}				−0.00001				−0.00000
				(0.00000)				(0.00000)
old × dist	−0.001	0.001	−0.008	0.088*	0.001	0.005	0.025	0.083**
	(0.002)	(0.010)	(0.025)	(0.052)	(0.002)	(0.007)	(0.018)	(0.037)
old × dist ^{2}		0.0001	−0.001	−0.001		0.0003	0.002	0.001
		(0.0003)	(0.002)	(0.007)		(0.0002)	(0.001)	(0.005)
old × dist ^{3}			−0.00001	0.001**			0.00004	0.0005*
			(0.00004)	(0.0003)			(0.00003)	(0.0002)
old × dist ^{4}				−0.00000				−0.00000
				(0.00001)				(0.00000)
Constant	0.517***	0.505***	0.529***	0.443***	0.147***	0.114***	0.052	0.006
	(0.031)	(0.049)	(0.070)	(0.097)	(0.023)	(0.032)	(0.046)	(0.066)
N	2,173	2,173	2,173	2,173	2,173	2,173	2,173	2,173
R^2	0.001	0.001	0.001	0.003	0.002	0.003	0.005	0.006

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 21: Covariate balance regressions with case specific bandwidths

	Dep.var : Sex - Male				Dep.var : Social category - Privileged			
	spe-01	spe-02	spe-03	spe-04	spe-01	spe-02	spe-03	spe-04
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 - h = 30								
	0.021	−0.001	0.016	−0.01	0.022	0.069	0.134^{**}	0.139^{*}
	(0.043)	(0.065)	(0.088)	(0.116)	(0.032)	(0.045)	(0.06)	(0.078)
	N = 2173	N = 2173	N = 2173	N = 2173	N = 2173	N = 2173	N = 2173	N = 2173
2 - h = hopt								
	-0.038^{*}	−0.009	0.014	0.014	−0.014	−0.017	−0.01	0.01
	(0.022)	(0.035)	(0.041)	(0.043)	(0.017)	(0.026)	(0.03)	(0.031)
	N = 8306	N = 7379	N = 9725	N = 14322	N = 8306	N = 7379	N = 9725	N = 14322

Table 22: RDD First stage estimates

	Depedent variable : age at test				
	FS	RD-FS-01	RD-FS-02	RD-FS-03	RD-FS-04
	(1)	(2)	(3)	(4)	(5)
1 - All					
	0.847^{***}	0.821^{***}	0.82^{***}	0.818^{***}	0.824^{***}
	(0.013)	(0.02)	(0.03)	(0.036)	(0.038)
	N = 13561	N = 7471	N = 7379	N = 9725	N = 13603
2 - Females					
	0.849^{***}	0.783^{***}	0.773^{***}	0.766^{***}	0.778^{***}
	(0.017)	(0.029)	(0.034)	(0.047)	(0.052)
	N = 6666	N = 2765	N = 4532	N = 4634	N = 5821
3 - Males					
	0.845^{***}	0.796^{***}	0.781^{***}	0.79^{***}	0.804^{***}
	(0.018)	(0.027)	(0.043)	(0.048)	(0.055)
	N = 6895	N = 4035	N = 3454	N = 5167	N = 6070
4 - Underprivileged					
	0.839^{***}	0.844^{***}	0.854^{***}	0.854^{***}	0.843^{***}
	(0.015)	(0.024)	(0.037)	(0.044)	(0.053)
	N = 8664	N = 5341	N = 5216	N = 6169	N = 6812
5 - Privileged					
	0.84^{***}	0.899^{***}	0.906^{***}	0.917^{***}	0.935^{***}
	(0.022)	(0.029)	(0.045)	(0.052)	(0.061)
	N = 2592	N = 1337	N = 1268	N = 1877	N = 2203

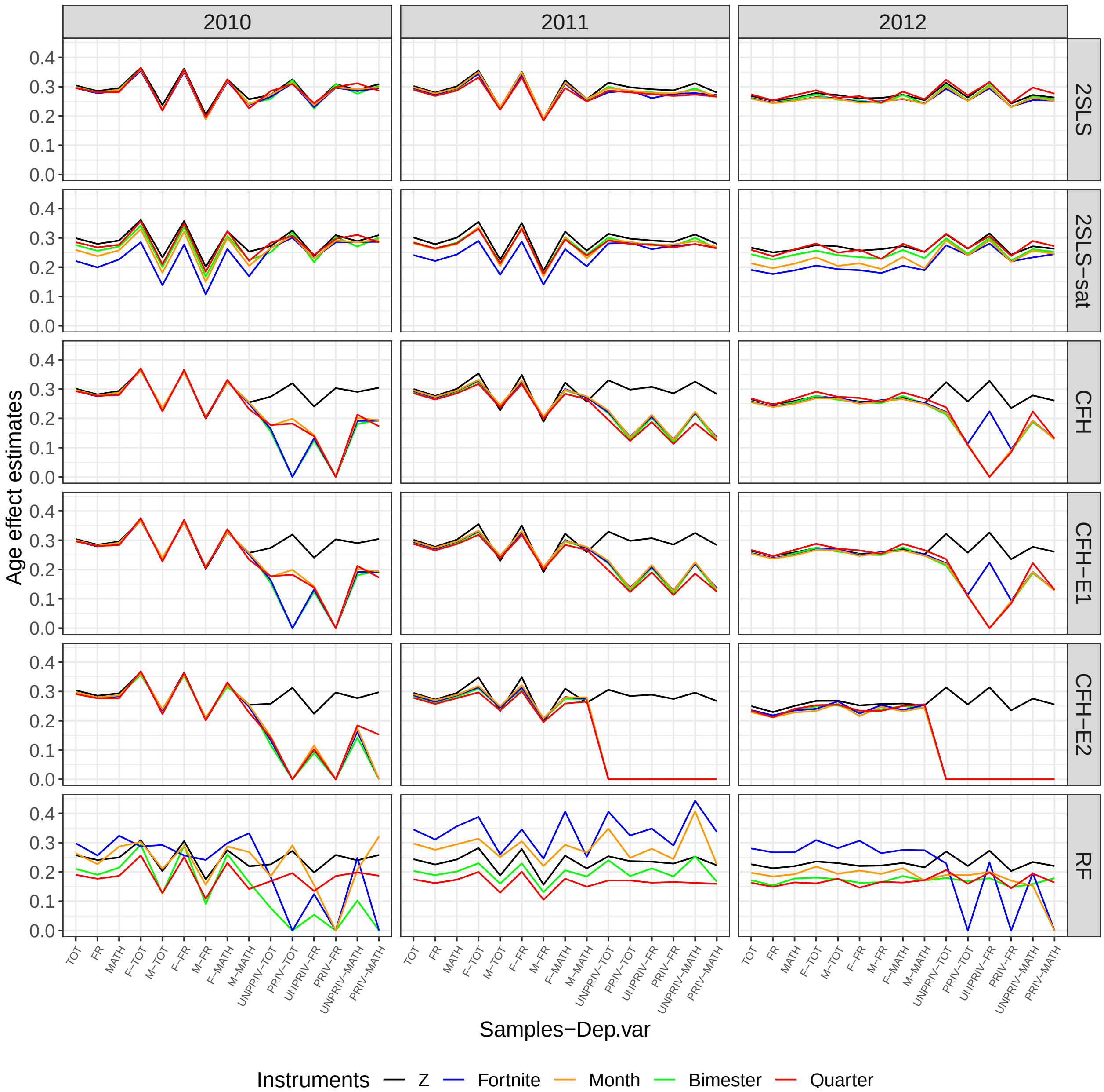

Figure 18: Sensitivity analysis illustration

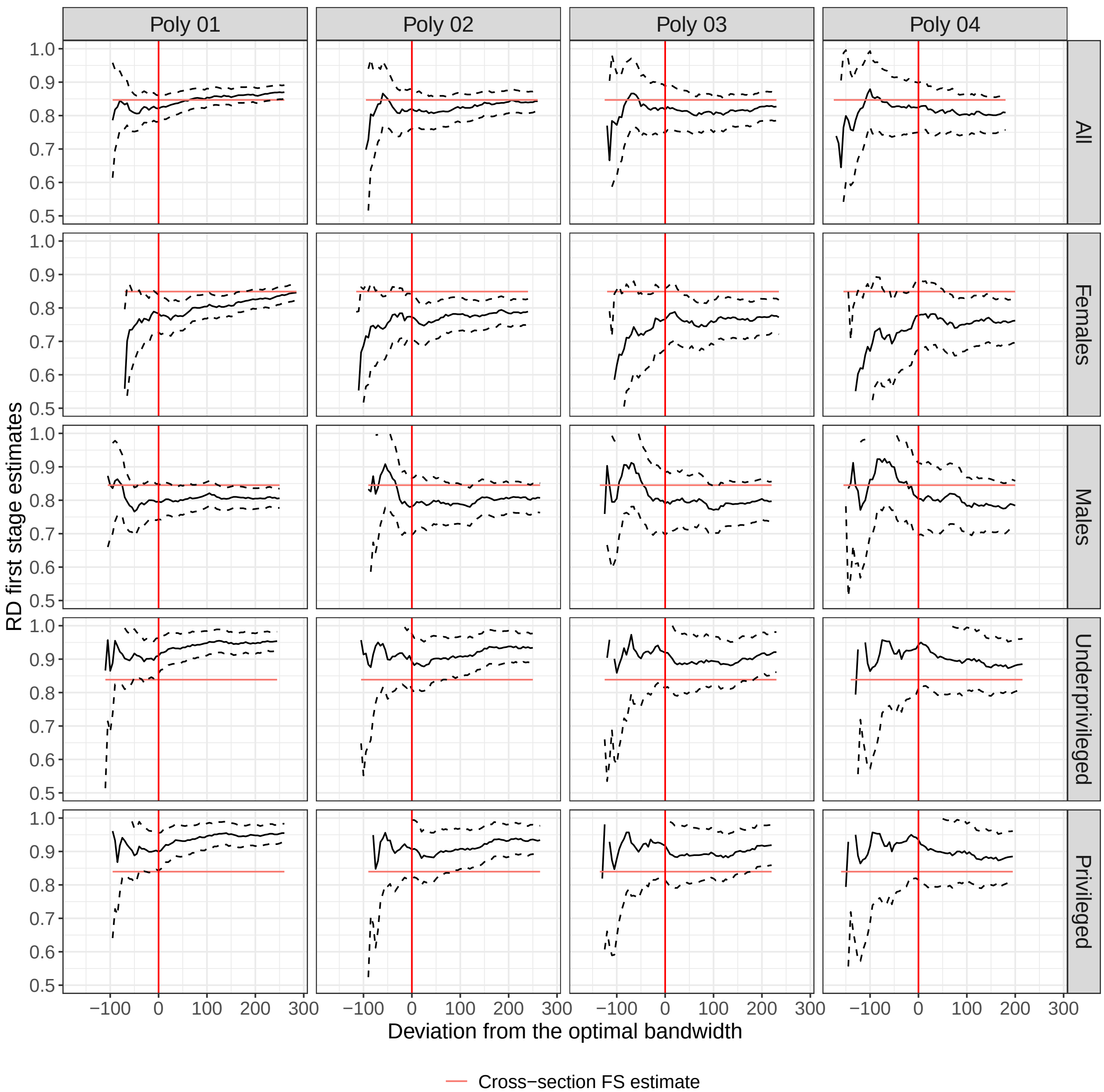

Figure 19: Sensitivity analysis of first stages in fuzzy regression design

Table 23: Sensitivity of RDD regression results to bandwidth choice

Dependent variable : french score							
ITT-RD				FRD			
RD-01	RD-02	RD-03	RD-04	FRD-01	FRD-02	FRD-03	FRD-04
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 - h = hopt - 30							
0.24^{***}	0.288^{***}	0.348^{***}	0.254^{***}	0.307^{***}	0.357^{***}	0.426^{***}	0.353^{***}
(0.051)	(0.08)	(0.089)	(0.088)	(0.069)	(0.106)	(0.118)	(0.116)
N = 5940	N = 5119	N = 7379	N = 12054	N = 5197	N = 5119	N = 7379	N = 11340
2 - h = hopt - 15							
0.271^{***}	0.288^{***}	0.264^{***}	0.266^{***}	0.299^{***}	0.355^{***}	0.324^{***}	0.327^{***}
(0.047)	(0.073)	(0.083)	(0.084)	(0.064)	(0.096)	(0.108)	(0.111)
N = 7066	N = 6219	N = 8589	N = 13212	N = 6292	N = 6219	N = 8589	N = 12497
3 - h = hopt							
0.274^{***}	0.236^{***}	0.249^{***}	0.25^{***}	0.326^{***}	0.288^{***}	0.304^{***}	0.328^{***}
(0.044)	(0.068)	(0.078)	(0.081)	(0.059)	(0.087)	(0.101)	(0.107)
N = 8306	N = 7379	N = 9725	N = 14322	N = 7471	N = 7379	N = 9725	N = 13603
4 - h = hopt + 15							
0.232^{***}	0.268^{***}	0.262^{***}	0.297^{***}	0.313^{***}	0.327^{***}	0.318^{***}	0.315^{***}
(0.041)	(0.064)	(0.074)	(0.078)	(0.055)	(0.082)	(0.096)	(0.102)
N = 9417	N = 8589	N = 10901	N = 15490	N = 8672	N = 8589	N = 10901	N = 14701
5 - h = hopt + 30							
0.221^{***}	0.293^{***}	0.293^{***}	0.311^{***}	0.275^{***}	0.359^{***}	0.359^{***}	0.383^{***}
(0.039)	(0.06)	(0.071)	(0.075)	(0.051)	(0.079)	(0.093)	(0.102)
N = 10578	N = 9725	N = 11993	N = 16597	N = 9819	N = 9725	N = 11993	N = 15852

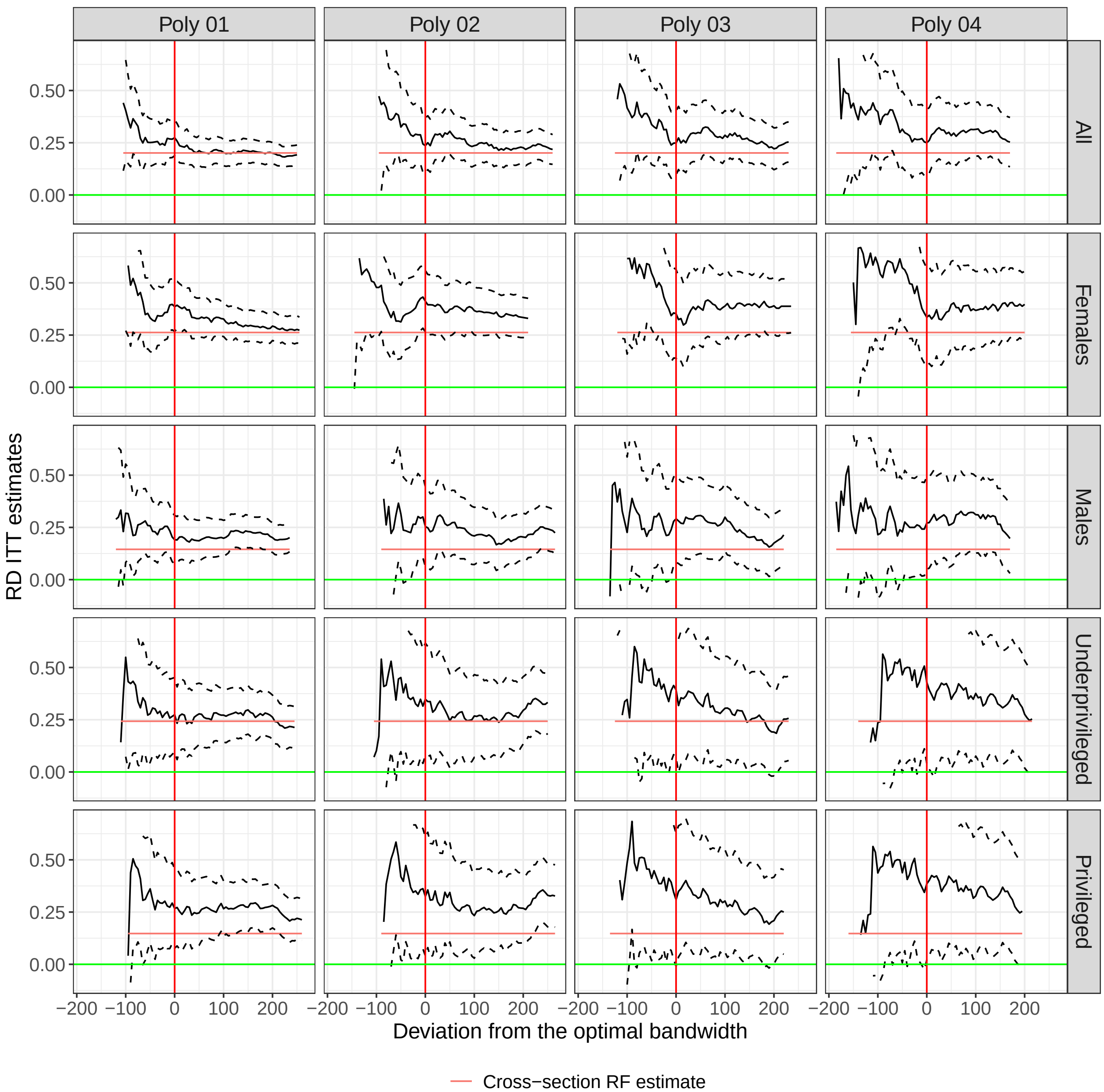

Figure 20: Illustration of sensitivity of RD estimates to bandwidth choice

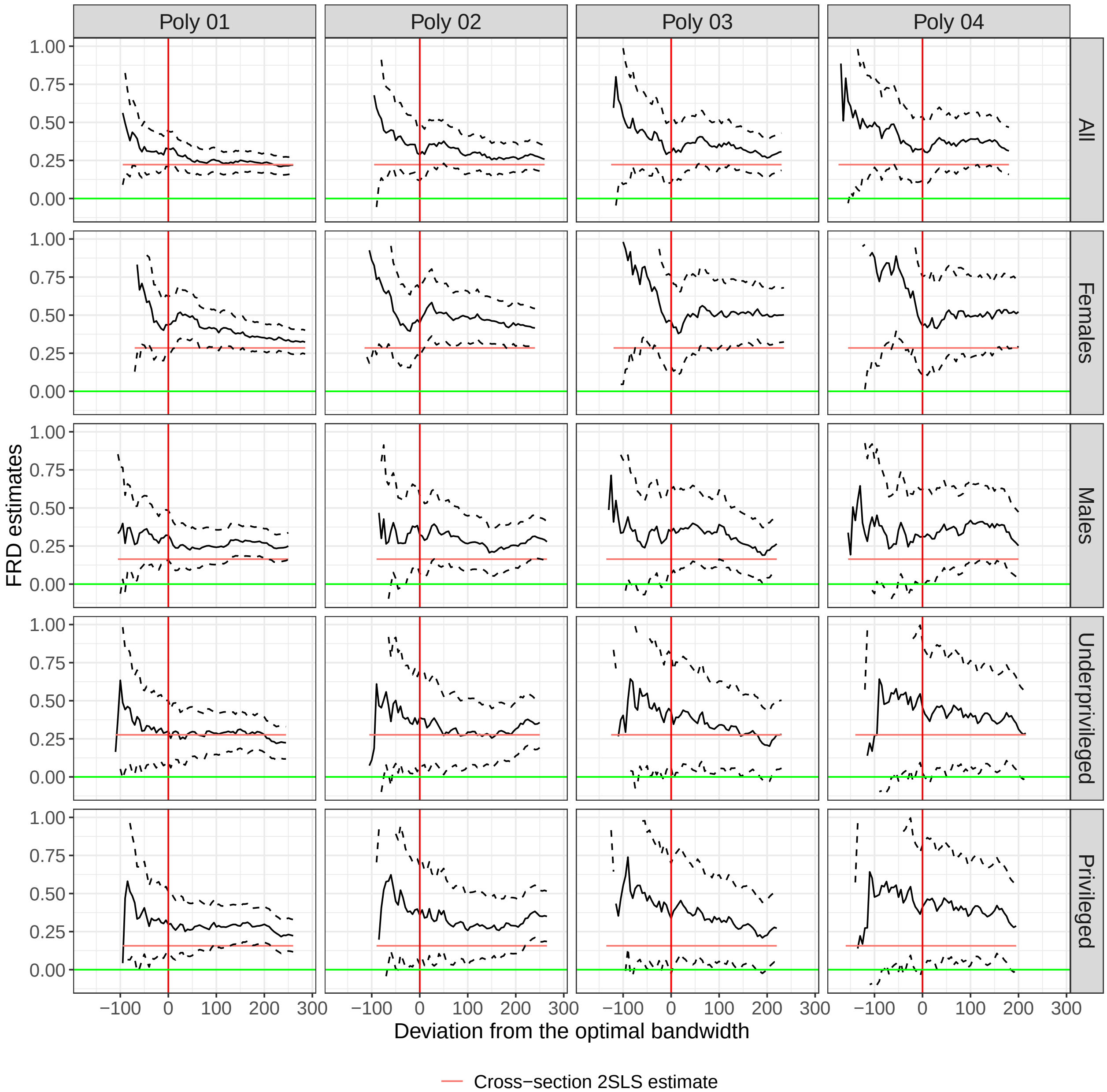

Figure 21: Sensitivity analysis illustration (FRD - no advances)

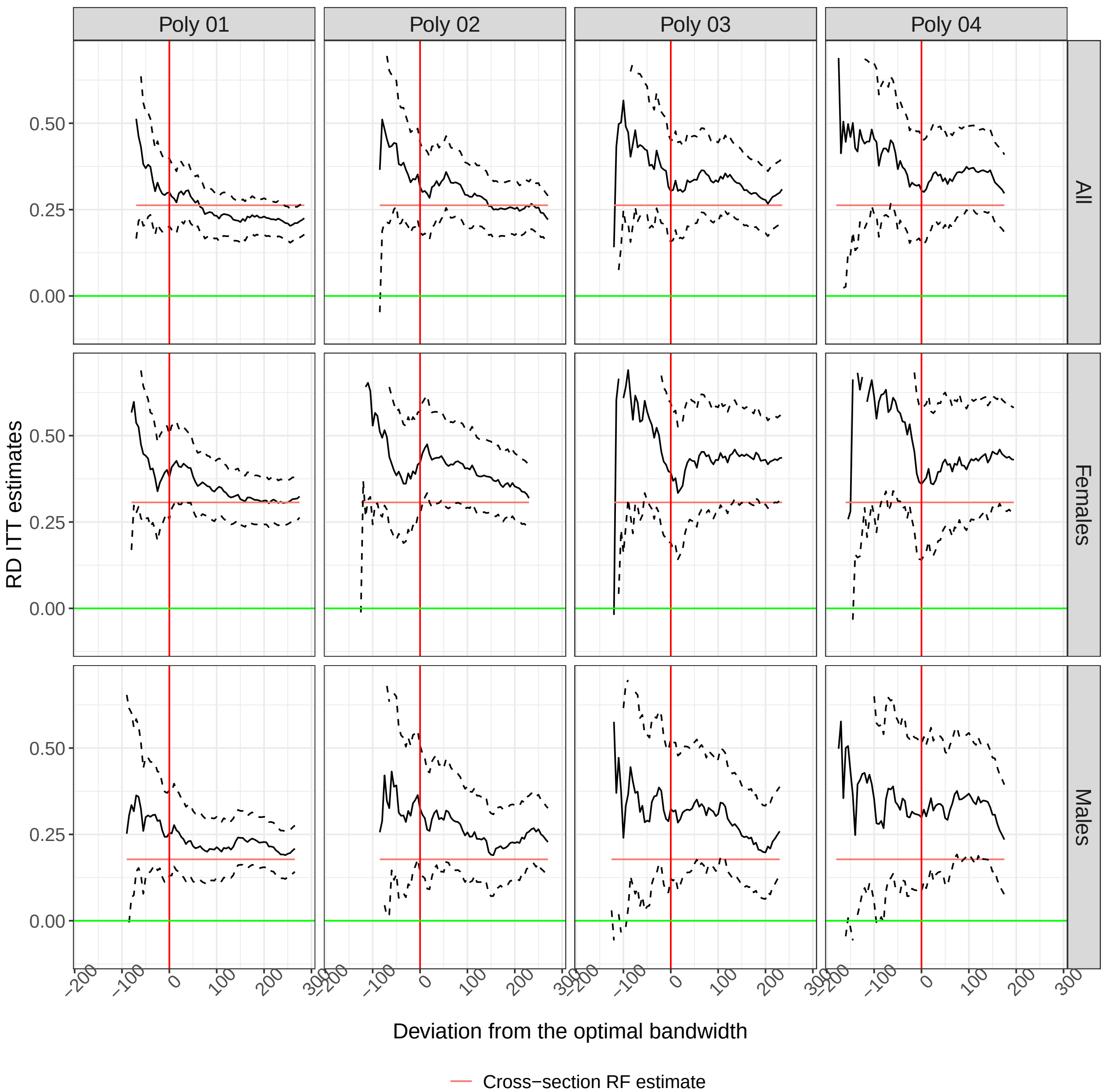

Figure 22: Illustration of sensitivity of RD estimates to bandwith choice (with advanced ones)

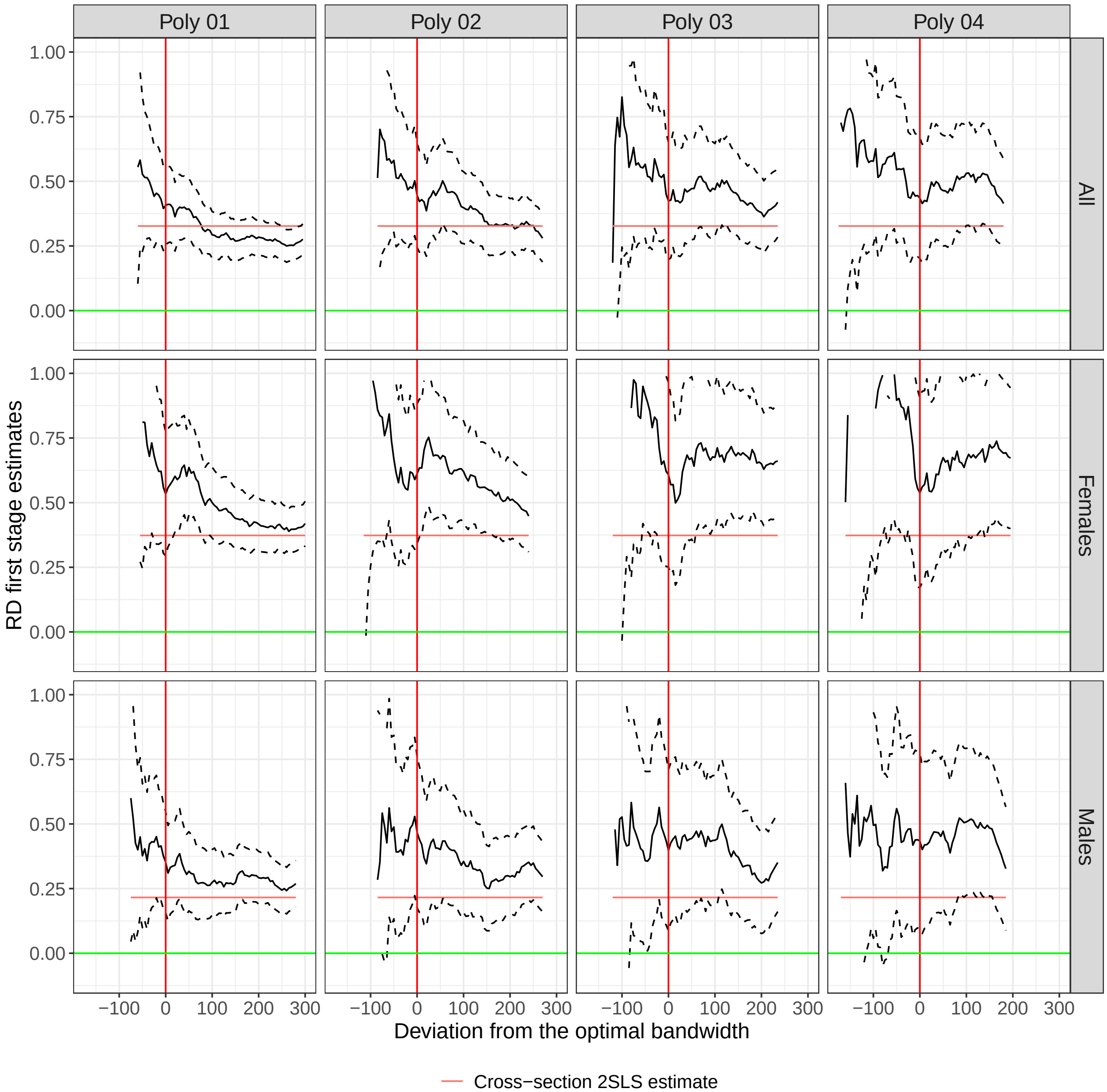

Figure 23: Sensitivity analysis illustration of the FRD resutls (with the advanced ones)

Table 24: Summary statistics of the french sub-items scores

Cohort	Min	First quartile	Mean	Standard deviation	Median	Third quartile	Max
1 - Writing							
2010	0	4	6.01	2.71	6	8	10
2011	0	4	6.05	2.70	6	8	10
2012	0	5	6.37	2.61	7	9	10
2 - Reading							
2010	0	5	8.02	3.70	8	11	15
2011	0	5	7.70	3.46	8	10	15
2012	0	5	7.84	3.60	8	11	15
3 - Grammar							
2010	0	3	6.48	3.90	6	9	15
2011	0	4	6.84	4.08	7	10	15
2012	0	4	7.29	4.09	7	11	15
4 - Spelling							
2010	0	2	4.86	3.06	5	7	10
2011	0	2	4.79	2.98	5	7	10
2012	0	3	4.98	2.99	5	7	10
5 - Vocabulary							
2010	0	4	6.07	2.43	6	8	10
2011	0	4	6.09	2.45	6	8	10
2012	0	3	5.30	2.83	5	8	10

Table 25: Summary statistics of the mathematics sub-items scores

Cohort	Min	First quartile	Mean	Standard deviation	Median	Third quartile	Max
1 - Calculus							
2010	0	3	5.64	3.37	6	8	12
2011	0	5	7.18	3.48	7	10	13
2012	0	4	7.70	4.13	8	11	15
2 - Geometry							
2010	0	2	3.65	1.94	4	5	7
2011	0	3	4.08	1.88	4	6	7
2012	0	2	3.11	1.38	3	4	5
3 - Measures							
2010	0	0	1.74	1.73	1	3	7
2011	0	1	2.36	1.75	2	4	6
2012	0	2	3.95	2.31	4	6	8
4 - Number							
2010	0	2	3.53	2.33	3	5	8
2011	0	2	3.36	2.09	3	5	7
2012	0	3	3.65	1.60	4	5	6
5 - Data organization							
2010	0	0	1.50	1.64	1	2	6
2011	0	1	2.34	2.04	2	4	7
2012	0	1	2.85	1.84	3	4	6

Table 26: Mean of french sub-items scores by institutional features

	2010					2011					2012				
	Writ.	Read.	Gram.	Spel.	Voca.	Writ.	Read.	Gram.	Spel.	Voca.	Writ.	Read.	Gram.	Spel.	Voca.
1 - Sex															
Females	6.62	8.47	7.00	5.49	6.45	6.68	8.17	7.44	5.34	6.42	7.09	8.58	8.21	5.79	5.80
Males	5.42	7.58	5.99	4.25	5.70	5.40	7.23	6.24	4.23	5.76	6.12	7.71	7.07	4.65	5.27
2 - Social category															
Farmers	6.25	8.25	6.59	4.77	6.13	6.39	8.31	7.29	5.01	6.54	6.76	7.98	7.78	4.93	5.60
Entrepreneurs	6.55	9.10	7.50	5.69	6.71	6.76	8.79	8.17	5.94	6.89	6.95	8.84	8.26	5.69	6.11
Executives	7.35	10.47	9.00	6.64	7.55	7.46	10.14	9.86	6.90	7.79	7.79	10.27	9.95	6.79	7.40
Intermediates	6.89	9.44	7.88	5.90	6.93	6.91	9.19	8.64	5.99	7.15	7.17	9.23	8.63	6.09	6.37
Employees	6.34	8.56	7.00	5.33	6.45	6.57	8.37	7.67	5.39	6.58	6.78	8.51	7.97	5.52	5.81
Workers	6.06	8.00	6.37	4.89	6.06	6.18	7.66	6.80	4.79	6.10	6.41	7.71	7.17	4.96	5.20
Retired	6.70	9.54	8.22	5.50	6.80	7.15	9.43	8.95	6.39	7.33	7.22	9.48	8.58	6.22	6.74
Unemployed	5.67	7.33	5.78	4.36	5.70	5.85	7.20	6.19	4.32	5.80	5.95	7.04	6.46	4.34	4.62
Others	5.22	6.90	5.51	3.96	5.30	5.12	6.62	5.58	3.87	5.22	5.53	6.74	6.15	4.17	4.50
3 - Social category (grouped)															
Underprivileged	5.93	7.79	6.22	4.71	5.96	6.09	7.59	6.68	4.68	6.06	6.25	7.55	6.99	4.76	5.04
Privileged	6.98	9.76	8.22	6.14	7.12	7.11	9.50	9.05	6.36	7.36	7.35	9.53	9.05	6.26	6.70
Others	5.22	6.90	5.51	3.96	5.30	5.12	6.62	5.58	3.87	5.22	5.53	6.74	6.15	4.17	4.50
4 - Priority education network															
HEP	6.26	8.54	7.02	5.27	6.42	6.26	8.15	7.32	5.13	6.40	6.62	8.28	7.71	5.33	5.63
ECLAIR	5.69	7.30	5.85	4.39	5.67	5.71	7.04	6.15	4.25	5.63	6.06	7.35	6.99	4.60	4.89
RRS	5.78	7.59	5.96	4.41	5.71	5.92	7.38	6.48	4.57	5.89	6.18	7.37	6.66	4.63	5.03
5 - Priority education network (yes/no)															
No	6.26	8.54	7.02	5.27	6.42	6.26	8.15	7.32	5.13	6.40	6.62	8.28	7.71	5.33	5.63
Yes	5.73	7.44	5.90	4.40	5.69	5.81	7.20	6.31	4.40	5.76	6.12	7.36	6.84	4.61	4.95

Table 27: Mean of maths sub-items scores by institutional features

	2010					2011					2012				
	Calc.	Geom.	Meas.	Num.	Data org.	Calc.	Geom.	Meas.	Num.	Data org.	Calc.	Geom.	Meas.	Num.	Data org.
1 - Sex															
Females	5.80	3.72	1.71	3.54	1.51	7.36	4.13	2.36	3.40	2.41	8.24	3.23	4.08	3.82	2.97
Males	5.49	3.59	1.77	3.52	1.49	6.99	4.02	2.36	3.33	2.27	7.82	3.16	4.18	3.73	2.99
2 - Social category															
Farmers	5.71	3.77	1.89	3.61	1.64	7.87	4.46	2.49	3.51	2.56	8.17	3.28	4.30	3.81	3.15
Entrepreneurs	6.48	4.10	2.15	4.12	1.97	8.52	4.46	3.07	3.96	3.05	8.57	3.39	4.57	3.93	3.33
Executives	7.55	4.60	2.79	4.91	2.65	9.47	5.17	3.68	4.76	3.99	10.10	3.70	5.49	4.45	3.93
Intermediates	6.52	4.16	2.30	4.30	2.02	8.59	4.74	3.10	4.09	3.20	9.13	3.40	4.77	4.13	3.46
Employees	6.10	3.82	1.84	3.82	1.63	7.94	4.31	2.63	3.69	2.69	8.26	3.24	4.28	3.86	3.06
Workers	5.69	3.67	1.64	3.53	1.44	7.41	4.07	2.34	3.36	2.21	7.63	3.14	3.91	3.65	2.85
Retired	6.73	4.09	2.36	4.30	2.01	8.55	4.58	3.15	4.11	3.39	9.76	3.57	5.16	4.08	3.43
Unemployed	5.13	3.40	1.46	3.13	1.19	6.66	3.87	2.06	3.09	1.95	6.92	2.92	3.44	3.40	2.49
Others	4.76	3.22	1.45	2.98	1.20	6.02	3.63	1.90	2.81	1.84	6.61	2.81	3.37	3.25	2.38
3 - Social category (grouped)															
Underprivileged	5.49	3.57	1.60	3.39	1.36	7.11	4.02	2.25	3.29	2.19	7.41	3.05	3.75	3.56	2.71
Privileged	6.92	4.32	2.45	4.50	2.26	8.94	4.86	3.33	4.35	3.50	9.36	3.51	5.00	4.20	3.61
Others	4.76	3.22	1.45	2.98	1.20	6.02	3.63	1.90	2.81	1.84	6.61	2.81	3.37	3.25	2.38
4 - Priority education network															
HEP	6.02	3.82	1.93	3.78	1.69	7.58	4.20	2.54	3.57	2.58	8.03	3.21	4.22	3.76	3.01
ECLAIR	5.18	3.43	1.52	3.24	1.26	6.54	3.87	2.08	3.03	2.01	7.42	3.01	3.70	3.60	2.71
RRS	5.29	3.51	1.55	3.28	1.33	6.94	4.01	2.25	3.24	2.14	7.24	2.98	3.64	3.44	2.62
5 - Priority education network (yes/no)															
No	6.02	3.82	1.93	3.78	1.69	7.58	4.20	2.54	3.57	2.58	8.03	3.21	4.22	3.76	3.01
Yes	5.23	3.47	1.53	3.26	1.29	6.73	3.94	2.16	3.13	2.07	7.34	3.00	3.67	3.53	2.67

References

Alet, Elodie, Liliane Bonnal, and Pascal Favard. 2013. "Repetition: Medicine for a Short-Run Remission." *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*. JSTOR, 227–50.

Aliprantis, Dionissi. 2012. "Redshirting, Compulsory Schooling Laws, and Educational Attainment." *Journal of Educational and Behavioral Statistics* 37 (2). SAGE Publications Sage CA: Los Angeles, CA: 316–38.

———. 2014. "When Should Children Start School?" *Journal of Human Capital* 8 (4). University of Chicago Press Chicago, IL: 481–536.

Angrist, Joshua D, and Guido W Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430). Taylor & Francis Group: 431–42.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434). Taylor & Francis: 444–55.

Arellano, Manuel. 1987. "PRACTITIONERS'CORNER: Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics* 49 (4). Wiley Online Library: 431–34.

Barret, Garry F, and Stephen G Donald. 2003. "Consistent Tests for Stochastic Dominance." *Econometrica* 71 (1): 71–104.

Bedard, Kelly, and Elizabeth Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics* 121 (4). MIT Press: 1437–72.

Bell, John F, and Sandra Daniels. 1990. "Are Summer-Born Children Disadvantaged? The Birthdate Effect in Education." *Oxford Review of Education* 16 (1). Taylor & Francis: 67–80.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *The Review of Economics and Statistics* 93 (2). MIT Press: 455–67.

Buckles, Kasey S, and Daniel M Hungerman. 2013. "Season of Birth and Later Outcomes: Old Questions, New Answers." *Review of Economics and Statistics* 95 (3). MIT Press: 711–24.

Cahan, Sorel, and Nora Cohen. 1989. "Age Versus Schooling Effects on Intelligence Development." *Child Development*. JSTOR, 1239–49.

Cahan, Sorel, and Daniel Davis. 1987. "A Between-Grade-Levels Approach to the Investigation of the Absolute Effects of Schooling on Achievement." *American Educational Research Journal* 24 (1). Sage Publications: 1–12.

Cascio, Elizabeth, and others. 2008. "How and Why Does Age at Kindergarten Entry Matter?" *FRBSF Economic Letter*. Federal Reserve Bank of San Francisco.

Cascio, Elizabeth U, and Ethan G Lewis. 2006. "Schooling and the Armed Forces Qualifying Test Evidence from School-Entry Laws." *Journal of Human Resources* 41 (2). University of Wisconsin Press: 294–318.

Cattaneo, Matias D, Nicolás Idrobo, and Rocı'o Titiunik. 2017a. "A Practical Introduction to Regression Discontinuity Designs." *Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I.*

———. 2017b. "A Practical Introduction to Regression Discontinuity Designs." *Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I.*

Crawford, Claire, Lorraine Dearden, and Costas Meghir. 2007. "When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England." Institute for Fiscal Studies.

Datar, Ashlesha. 2006. "Does Delaying Kindergarten Entrance Give Children a Head Start?" *Economics of Education Review* 25 (1). Elsevier: 43–62.

Davidson, Russell, and Emmanuel Flachaire. 2008. "The Wild Bootstrap, Tamed at Last." *Journal of Econometrics* 146 (1). Elsevier: 162–69.

DeCos, Patricia L. 1997. *Readiness for Kindergarten: What Does It Mean?* California State Library, California Research Bureau.

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2017. "School Starting Age and Cognitive Development." National Bureau of Economic Research.

Dhuey, Elizabeth, and Stephen Lipscomb. 2008. "What Makes a Leader? Relative Age and High School Leadership." *Economics of Education Review* 27 (2). Elsevier: 173–83.

Dobkin, Carlos, and Fernando Ferreira. 2010. "Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review* 29 (1). Elsevier: 40–54.

Elder, Todd E, and Darren H Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources* 44 (3). University of Wisconsin Press: 641–83.

Fertig, Michael, and Jochen Kluve. 2005. "The Effect of Age at School Entry on Educational Attainment in Germany."

Fiorini, Mario, Katrien Stevens, and others. 2014. "Assessing the Monotonicity Assumption in Iv and Fuzzy Rd Designs." *Economics Working Paper Series-University of Sidney* 13.

Fletcher, Jason, and Taehoon Kim. 2016. "The Effects of Changes in Kindergarten Entry Age Policies on Educational Achievement." *Economics of Education Review* 50. Elsevier: 45–62.

Fredriksson, Peter, and Björn Öckert. 2006. "Is Early Learning Really More Productive." *The Effect of School Starting Age on School and Labor Market Performance, IZA WP* 1659.

———. 2013. "Life-Cycle Effects of Age at School Start." *The Economic Journal* 124 (579). Oxford University Press Oxford, UK: 977–1004.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economics Statistics* 37 (3). Taylor &

Francis: 447–56.

Graue, M Elizabeth, and James DiPerna. 2000. “Redshirting and Early Retention: Who Gets the" Gift of Time" and What Are Its Outcomes?” *American Educational Research Journal* 37 (2). Sage Publications: 509–34.

Grenet, Julien. 2009. “Academic Performance, Educational Trajectories and the Persistence of Date of Birth Effects. Evidence from France.” *Unpublished Manuscript*.

Hart, Robert A, and Mirko Moro. 2017. “Date of Birth and Selective Schooling.”

Hámori, Szilvia. 2007. “The Effect of School Starting Age on Academic Performance in Hungary.” Budapest Working Papers on the Labour Market.

Hámori, Szilvia, and János Köllő. 2012. “Whose Children Gain from Starting School Later?– Evidence from Hungary.” *Educational Research and Evaluation* 18 (5). Taylor & Francis: 459–88.

Heckman, James J, Sergio Urzua, and Edward Vytlacil. 2006. “Understanding Instrumental Variables in Models with Essential Heterogeneity.” *The Review of Economics and Statistics* 88 (3). MIT Press: 389–432.

Imbens, Guido W., and Joshua D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62 (2). [Wiley, Econometric Society]: 467–75. http://www.jstor.org/stable/2951620.

Imbens, Guido W, and Thomas Lemieux. 2008. “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics* 142 (2). Elsevier: 615–35.

Jürges, Hendrik, and Kerstin Schneider. 2011. “Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System.” *German Economic Review* 12 (4). Wiley Online Library: 371–94.

Kaila, Martti. 2017. “The Effects of Relative School Starting Age on Educational Outcomes in Finland.”

Kawaguchi, Daiji. 2011. “Actual Age at School Entry, Educational Outcomes, and Earnings.” *Journal of the Japanese and International Economies* 25 (2). Elsevier: 64–80.

Langer, Philip, John Michael Kalk, and Donald T Searls. 1984. “Age of Admission and Trends in Achievement: A Comparison of Blacks and Caucasians.” *American Educational Research Journal* 21 (1). Sage Publications: 61–78.

Lee, David S, and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48 (2): 281–355.

Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek, and Dinand Webbink. 2004. “New evidence on the effect of time in school on early achievement.” HEW 0410001. University Library of Munich, Germany. https://ideas.repec.org/p/wpa/wuwphe/0410001.html.

Matta, Rafael, Rafael P Ribas, Breno Sampaio, and Gustavo R Sampaio. 2016. “The Effect of Age at School Entry on College Admission and Earnings: A Regression-Discontinuity Approach.” *IZA Journal of Labor Economics* 5 (1). SpringerOpen: 9.

Mayer, Susan E, David Knutson, and others. 1997. "Does Age at Enrollment in First Grade Affect Children's Cognitive Test Scores." Northwestern University/University of Chicago Joint Center for Poverty Research.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2). Elsevier: 698–714.

Mühlenweg, Andrea M, and Patrick A Puhani. 2010. "The Evolution of the School-Entry Age Effect in a School Tracking System." *Journal of Human Resources* 45 (2). University of Wisconsin Press: 407–38.

Nam, Kigon. 2014. "Until When Does the Effect of Age on Academic Achievement Persist? Evidence from Korean Data." *Economics of Education Review* 40. Elsevier: 106–22.

Pellizzari, Michele, and Francesco C Billari. 2012. "The Younger, the Better? Age-Related Differences in Academic Performance at University." *Journal of Population Economics* 25 (2). Springer: 697–739.

Peña, Pablo A. 2017. "Creating Winners and Losers: Date of Birth, Relative Age in School, and Outcomes in Childhood and Adulthood." *Economics of Education Review* 56. Elsevier: 152–76.

Piketty, Thomas, Mathieu Valdenaire, and others. 2006. *L'impact de La Taille Des Classes Sur La Réussite Scolaire Dans Les écoles, Collèges et Lycées Français: Estimations à Partir Du Panel Primaire 1997 et Du Panel Secondaire 1995.* Direction de l'évaluation et de la prospective.

Puhani, PA, and AM Weber. 2005. "Does the Early Bird Catch the Worm? Instrumental Variable Coefficients of Educational Effects of Age of School Enty in Germany." IZA Discussion Paper.

Robertson, Erin. 2011. "The Effects of Quarter of Birth on Academic Outcomes at the Elementary School Level." *Economics of Education Review* 30 (2). Elsevier: 300–311.

Schneeweis, Nicole, and Martina Zweimüller. 2014. "Early Tracking and the Misfortune of Being Young." *The Scandinavian Journal of Economics* 116 (2). Wiley Online Library: 394–428.

Smith, Justin. 2009. "Can Regression Discontinuity Help Answer an Age-Old Question in Education? The Effect of Age on Elementary and Secondary School Achievement." *The BE Journal of Economic Analysis & Policy* 9 (1). De Gruyter.

Staiger, Douglas, and H James. 1997. "Stock. 1997.'Instrumental Variables with Weak Instruments.'." *Econometrica* 65 (3): 557–86.

Stipek, Deborah. 2002. "At What Age Should Children Enter Kindergarten? A Question for Policy Makers and Parents." *Social Policy Report* 16 (2): 1–20.

Strøm, Bjarne. 2004. "Student Achievement and Birthday Effects." Citeseer.

Suziedelyte, Agne, and Anna Zhu. 2015. "Does Early Schooling Narrow Outcome Gaps for Advantaged and Disadvantaged Children?" *Economics of Education Review* 45. Elsevier: 76–88.

White, Halbert, and others. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4). Princeton: 817–38.

Woessmann, Ludger. 2016. “The Importance of School Systems: Evidence from International Differences in Student Achievement.” *Journal of Economic Perspectives* 30 (3): 3–32.

Wooldridge, Jeffrey M. 2015. “Control Function Methods in Applied Econometrics.” *Journal of Human Resources* 50 (2). University of Wisconsin Press: 420–45.

TEPP Working Papers 2020

20-5. Slowdown antitrust investigations by decentralization
Emilie Dargaud, Armel Jacques

20-4. Is international tourism responsible for the pandemic of COVID19? A preliminary cross-country analysis with a special focus on small islands
Jean-François Hoarau

20-3. Does labor income react more to income tax or means tested benefit reforms?
Michaël Sicsic

20-2. Optimal sickness benefits in a principal-agent model
Sébastien Ménard

20-1. The specific role of agriculture for economic vulnerability of small island spaces
Stéphane Blancard, Maximin Bonnet, Jean-François Hoarau

TEPP Working Papers 2019

19-8. The impact of benefit sanctions on equilibrium wage dispersion and job vacancies
Sebastien Menard

19-7. Employment fluctuations, job polarization and non-standard work : Evidence from France and the US
Olivier Charlot, Idriss Fontaine, Thepthida Sopraseuth

19-6. Counterproductive hiring discrimination against women: Evidence from French correspondence test
Emmanuel Duguet, Loïc du Parquet, Yannick L'Horty, Pascale Petit

19-5. Inefficient couples: Non-minimization of the tax burden among French cohabiting couples
Olivier Bargain, Damien Echevin, Nicolas Moreau, Adrien Pacifico

19-4. Seeking for tipping point in the housing market: evidence from a field experiment
Sylvain Chareyron, Samuel Gorohouna, Yannick L'Horty, Pascale Petit, Catherine Ris

19-3. Testing for redlining in the labor market
Yannick L’Horty, Mathieu Bunel, Pascale Petit

19-2. Labour market flows: Accounting for the public sector
Idriss Fontaine, Ismael Galvez-Iniesta, Pedro Gomes, Diego Vila-Martin

19-1. The interaction between labour force participation of older men and their wife: lessons from France
Idriss Fontaine

TEPP Working Papers 2018

18-15. Be healthy, be employed: a comparison between the US and France based on a general equilibrium model
Xavier Fairise, François Langot, Ze Zhong Shang

18-14. Immigrants' wage performance in the routine biased technological change era: France 1994-2012
Catherine Laffineur, Eva Moreno-Galbis, Jeremy Tanguy, Ahmed Tritah

18-13. Welfare cost of fluctuations when labor market search interacts with financial frictions
Elini Iliopulos, François Langot, Thepthida Sopraseuth

18-12. Accounting for labor gaps
François Langot, Alessandra Pizzo

18-11. Unemployment fluctuations over the life cycle
Jean-Olivier Hairault, François Langot, Thepthida Sopraseuth

18-10. Layoffs, Recalls and Experience Rating
Julien Albertini, Xavier Fairise

18-9. Environmental policy and health in the presence of labor market imperfections
Xavier Pautrel

18-8. Identity mistakes and the standard of proof
Marie Obidzinski, Yves Oytana

18-7. Presumption of innocence and deterrence
Marie Obidzinski, Yves Oytana

18-6. Ethnic Discrimination in Rental Housing Market: An Experiment in New Caledonia
Mathieu Bunel, Samuel Gorohouna, Yannick L'Horty, Pascale Petit, Catherine Ris

18-5. Evaluating the impact of firm tax credits. Results from the French natural experiment CICE
Fabrice Gilles, Yannick L'Horty, Ferhat Mihoubi, Xi Yang

18-4. Impact of type 2 diabetes on health expenditure: an estimation based on individual administrative data
François-Olivier Baudot , Anne-Sophie Aguadé,Thomas Barnay, Christelle Gastaldi-Ménager, Anne Fargot-Campagna

18-3. How does labour market history influence the access to hiring interviews?
Emmanuel Duguet, Rémi Le Gall, Yannick L'Horty, Pascale Petit

18-2. Occupational mobility and vocational training over the life cycle
Anthony Terriau

18-1. Retired, at last? The short-term impact of retirement on health status in France
Thomas Barnay, Eric Defebvre

TEPP Working Papers 2017

17-11. Hiring discrimination against women: distinguishing taste based discrimination from statistical discrimination
Emmanuel Duguet, Loïc du Parquet, Pascale Petit

17-10. Pension reforms, older workers' employment and the role of job separation and finding rates in France
Sarah Le Duigou, Pierre-Jean Messe

17-9. Healthier when retiring earlier? Evidence from France
Pierre-Jean Messe, François-Charles Wolff

17-8. Revisting Hopenhayn and Nicolini's optimal unemployment insurance with job search monitoring and sanctions
Sebastien Menard, Solenne Tanguy

17-7. Ethnic Gaps in Educational Attainment and Labor-Market Outcomes: Evidence from France
Gabin Langevin, David Masclet, Fabien Moizeau, Emmanuel Peterle

17-6. Identifying preference-based discrimination in rental market: a field experiment in Paris
Mathieu Bunel, Yannick L'Horty, Loïc du Parquet, Pascale Petit

17-5. Chosen or Imposed? The location strategies of households
Emilie Arnoult, Florent Sari

17-4. Optimal income taxation with composition effects
Laurence Jacquet, Etienne Lehmann

17-3. Labor Market Effects of Urban Riots: an experimental assessment
Emmanuel Duguet, David Gray, Yannick L'Horty, Loic du Parquet, Pascale Petit

17-2. Does practicing literacy skills improve academic performance in first-year university students? Results from a randomized experiment
Estelle Bellity, Fabrices Gilles, Yannick L'Horty

17-1. Raising the take-up of social assistance benefits through a simple mailing: evidence from a French field experiment
Sylvain Chareyron, David Gray, Yannick L'Horty

TEPP Working Papers 2016

16-8. Endogenous wage rigidities, human capital accumulation and growth
Ahmed Tritah

16-7. Harder, better, faster...yet stronger? Working conditions and self-declaration of chronic diseases
Eric Defebvre

16-6. The influence of mental health on job retention
Thomas Barnay, Eric Defebvre

16-5. The effects of breast cancer on individual labour market outcomes: an evaluation from an administrative panel
Thomas Barnay, Mohamed Ali Ben Halima, Emmanuel Duguet, Christine Le Clainche, Camille Regaert

16-4. Expectations, Loss Aversion, and Retirement Decisions in the Context of the 2009 Crisis in Europe
Nicolas Sirven, Thomas Barnay

16-3. How do product and labor market regulations affect aggregate employment, inequalities and job polarization? A general equilibrium approach
Julien Albertini, Jean-Olivier Hairault, François Langot, Thepthida Sopraseuth

16-2. Access to employment with age and gender: results of a controlled experiment
Laetitia Challe, Florent Fremigacci, François Langot, Yannick L'Horty, Loïc Du Parquet, Pascale Petit

16-1. An evaluation of the 1987 French Disabled Workers Act: Better paying than hiring
Thomas Barnay, Emmanuel Duguet, Christine Le Clainche, Yann Videau

TEPP Working Papers 2015

15-10. Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach
Kory Kroft, Kavan Kucko, Etienne Lehmann, Johannes Schmieder

15-9. Search frictions and (in) efficient vocational training over the life-cycle
Arnaud Chéron, Anthony Terriau

15-8. Absenteeism and productivity: the experience rating applied to employer contributions to health insurance
Sébastien Ménard, Coralia Quintero Rojas

15-7. Take up of social assistance benefits: the case of homeless
Sylvain Chareyron

15-6. Spatial mismatch through local public employment agencies. Answers from a French quasi-experiment
Mathieu Bunel, Elisabeth Tovar

15-5. Transmission of vocational skills at the end of career: horizon effect and technological or organisational change
Nathalie Greenan, Pierre-Jean Messe

15-4. Protecting biodiversity by developing bio-jobs: A multi-branch analysis with an application on French data
Jean De Beir, Céline Emond, Yannick L'Horty, Laetitia Tuffery

15-3. Profit-Sharing and Wages: An Empirical Analysis Using French Data Between 2000 and 2007
Noélie Delahaie, Richard Duhautois

15-2. A meta-regression analysis on intergenerational transmission of education: publication bias and genuine empirical effect
Nicolas Fleury, Fabrice Gilles

15-1. Why are there so many long-term unemployed in Paris?
Yannick L'Horty, Florent Sari

TEPP Working Papers 2014

14-14. Hiring discrimination based on national origin and the competition between employed and unemployed job seekers
Guillaume Pierné

14-13. Discrimination in Hiring: The curse of motorcycle women
Loïc Du Parquet, Emmanuel Duguet, Yannick L'Horty, Pascale Petit

14-12. Residential discrimination and the ethnic origin: An experimental assessment in the Paris suburbs
Emmanuel Duguet, Yannick L'Horty, Pascale Petit

14-11. Discrimination based on place of residence and access to employment
Mathieu Bunel, Yannick L'Horty, Pascale Petit

14-10. Rural Electrification and Household Labor Supply: Evidence from Nigeria
Claire Salmon, Jeremy Tanguy

14-9. Effects of immigration in frictional labor markets: theory and empirical evidence from EU countries
Eva Moreno-Galbis, Ahmed Tritah

14-8. Health, Work and Working Conditions: A Review of the European Economic Literature
Thomas Barnay

14-7. Labour mobility and the informal sector in Algeria: a cross-sectional comparison (2007-2012)
Philippe Adair, Youghourta Bellache

14-6. Does care to dependent elderly people living at home increase their mental health?
Thomas Barnay, Sandrine Juin

14-5. The Effect of Non-Work Related Health Events on Career Outcomes: An Evaluation in the French Labor Market
Emmanuel Duguet, Christine le Clainche

14-4. Retirement intentions in the presence of technological change: Theory and evidence from France
Pierre-Jean Messe, Eva Moreno-Galbis, Francois-Charles Wolff

14-3. Why is Old Workers' Labor Market more Volatile? Unemployment Fluctuations over the Life-Cycle
Jean-Olivier Hairault, François Langot, Thepthida Sopraseuth

14-2. Participation, Recruitment Selection, and the Minimum Wage
Frédéric Gavrel

14-1. Disparities in taking sick leave between sectors of activity in France: a longitudinal analysis of administrative data
Thomas Barnay, Sandrine Juin, Renaud Legal

TEPP Working Papers 2013

13-9. An evaluation of the impact of industrial restructuring on individual human capital accumulation in France (1956-1993)
Nicolas Fleury, Fabrice Gilles

13-8. On the value of partial commitment for cooperative investment in buyer-supplier relationship
José de Sousa, Xavier Fairise

13-7. Search frictions, real wage rigidities and the optimal design of unemployment insurance
Julien Albertini, Xavier Fairise

13-6. Tax me if you can! Optimal nonlinear income tax between competing governments
Etienne Lehmann, Laurent Simula, Alain Trannoy

13-5. Beyond the labour income tax wedge: The unemployment-reducing effect of tax progressivity
Etienne Lehmann, Claudio Lucifora, Simone Moriconi, Bruno Van Der Linden

13-4. Discrimination based on place of residence and access to employment
Mathieu Bunel, Emilia Ene Jones, Yannick L'Horty, Pascale Petit

13-3. The determinants of job access channels: evidence from the youth labor market in Franc
Jihan Ghrairi

13-2. Capital mobility, search unemployment and labor market policies: The case of minimum wages
Frédéric Gavrel

13-1. Effort and monetary incentives in Nonprofit et For-Profit Organizations
Joseph Lanfranchi, Mathieu Narcy

The TEPP Institute

The CNRS **Institute for Theory and Evaluation of Public Policies** (the TEPP Institute, FR n°2024 CNRS) gathers together research centres specializing in economics and sociology:

- L'**Equipe de Recherche sur l'Utilisation des Données Individuelles en lien avec la Théorie Economique** (Research Team on Use of Individuals Data in connection with economic theory), **ERUDITE**, University of Paris-Est Créteil and University of Gustave Eiffel
- Le **Centre d'Etudes des Politiques Economiques de l'université d'Evry** (Research Centre focused on the analysis of economic policy and its foundations and implications), **EPEE**, University of Evry Val d'Essonne
- Le **Centre Pierre Naville** (Research on Work and Urban Policies), **CPN**, University of Evry Val d'Essonne
- Le **Groupe d'Analyse des Itinéraires et des Niveaux Salariaux** (Group on Analysis of Wage Levels and Trajectories), **GAINS**, University of Le Mans
- Le **Centre de Recherches en Economie et en Management**, (Research centre in Economics and Management), **CREM**, University of Rennes 1 et University of Caen Basse-Normandie
- Le **Groupe de Recherche ANgevin en Économie et Management** (Angevin Research Group in Economics and Management), **GRANEM**, University of Angers
- Le **Centre de Recherche en Economie et Droit** (Research centre in Economics and Law) **CRED**, University of Paris II Panthéon-Assas
- Le **Laboratoire d'Economie et de Management Nantes-Atlantique** (Laboratory of Economics and Management of Nantes-Atlantique) LEMNA, University of Nantes
- Le **Laboratoire interdisciplinaire d'étude du politique Hannah Arendt** – Paris Est, **LIPHA-PE**
- Le **Centre d'Economie et de Management de l'Océan Indien**, **CEMOI**, University of La Réunion

TEPP brings together 230 teacher-researchers and 100 doctoral students. It is both one of the main academic operators in the evaluation of public policies in France, and the largest multidisciplinary federation of research on work and employment. It responds to the demand for impact assessment of social programs using advanced technologies combining theoretical and econometric modeling, qualitative research techniques and controlled experiences.