Discussion Papers in Economic Behaviour

DPEB
01/20

Month of birth and academic performance: differences by gender and educational stage

VNIVERSITAT DE VALÈNCIA

>Pilar Beneito
University of Valencia and ERICES

>Pedro Javier Soria-Espín
Paris School of Economics

January, 2020

www.erices.es eri-ces

Month of birth and academic performance: differences by gender and educational stage

Pilar Beneito López[a]

Pedro Javier Soria-Espín[b]

Abstract

The month in which you were born can have a significant impact in your academic life. It is well documented that people who are born in the first months of the academic year tend to have better educational achievement than their younger peers within the same cohort. However, there is little literature addressing this relationship looking at differences by gender and educational stage. In this paper we fill this gap by studying the effect of the month of birth on academic performance of students at the University of Valencia (Spain). Using a Regression Discontinuity (RD) design we create a cut-off in 1st January to determine whether an individual is among the oldest (right to the cut-off) or among the youngest (left to the cut-off) within her cohort. We find that being relatively old has a positive effect on the access-to-university examination score for female students but not for their male peers. In addition, this effect seems to be concentrated in the upper quantiles of the entry score distribution and attenuates for university grades. We attribute this effect to a *virtuous circle* developed from early childhood, which is a recurring cycle of behavioral responses that translates into higher self-confidence for older students. Women appear to be more sensible to this effect than men.

Keywords: month of birth, academic achievement, behavioral responses, gender, sharp regression discontinuity.

---

[a] University of Valencia and ERI-CES. [b] Paris School of Economics (PSE).

We acknowledge the University of Valencia for providing the administrative data. Pilar Beneito acknowledges financial support from the Spanish Ministerio de Economía y Competitividad (ECO2017-86793-R) and Generalitat Valenciana (PROMETEO-2019-95).

1 Introduction

The month of birth can have a significant impact in your academic and professional life. It is well documented that individuals born in the first months of the academic year are more likely to have better educational attainment and professional outcomes that those born in the last months of the same academic year. In addition, this phenomenon seems to be consistent across different countries, it is generated during the primary school period and remains significant, at least, until the end of high-school and, possibly, at the professional stage. The explanations behind this evidence are not at all astrological but cognitive and psychological. Since the early stages of education the oldest students of a given cohort have a greater cognitive development that their younger peers, as in some cases they are almost a year older. These disparities in cognitive capacity matter a lot during childhood and appear to have long lasting consequences in personality traits, which maintain this *oldest premium* beyond primary school. Therefore, since all examinations and admission tests at different educational stages are taken on a fixed date, younger students may have a handicap in comparison to their older peers. Even if unintended, this is an unfair situation that may be limiting the talent of the youngest due to an arbitrary entrance-to-school cut-off. Thus, analyzing these oldest-youngest inequalities and delve into their determinants is relevant to reduce this loss of talent and therefore increase the efficiency of educational systems.

The mechanisms that may explain these results are well-known in the psychology and education literature. The differences between the oldest and the youngest of the same cohort begin during the first years of schooling. In Spain, the academic year starts in September and finishes in June or July. However, even if the course starts in September all the children born during the same natural year enter to the same cohort. For example, if the academic year begins on September 1996 all the children born from January 1990 to December 1990 are allowed to enter to the same class. This provokes that in many cohorts there are students that are almost one year older to some of their colleagues [1]. Since all the examinations are

[1]We also find this almost-one-year difference in other educational systems. The key difference is that in other countries, for instance UK, the cohorts are formed by people born from September to August. Following our example but using the UK's rule, if the academic year starts in September 1996 the cohort will be nurtured by children born between September 1990 and August 1991 instead of January-December 1990. In addition, this explains why the reader may find the month-of-birth effect in other papers named the *August handicap* or identifying the youngest as those born during the summer.

taken on a fixed date, the oldest in the cohort are more developed than the youngest and therefore have an advantage when they sit the exams. This is especially significant in the early stages of education, the years where the cognitive development process is taking place. Thus, these differences in intellectual maturity lead to disparities in academic performance, provoking that the relatively old students tend to obtain better grades.

These disparities in cognitive capacity matter a lot during childhood. However, and although relatively under-discussed in the subject literature, they may have long-lasting consequences: since the academic development of the individual is largely based on its earlier determinants, this *oldest premium* could be maintained beyond primary school, in fact, until the adult age. For example, the early advantage of older children may impact the personality development of these students, which translate in higher self-confidence and a greater valuation of their own scholastic capacity. These psychological traits may be present during the rest of the life and are very well rewarded in competitive settings like academic examinations. This logically leads to academic success which usually implies a more positive consideration of their peers and a better feedback of both family and instructors. At the same time, to the extent that this positive feedback and peers' considerations reinforce the academic self-esteem of the student, it facilitates further success, creating a *virtuous circle* that maintains the *oldest premium* during higher stages of education.

The possibility of long-lasting consequences of the month-of-birth effect opens two relevant questions, to which we aim to contribute in this paper: (a) what is the time horizon of the oldest premium? and (b) does this effect varies by gender?. As regards the first question, we provide evidence in this paper of a significant effect of the month of birth in the last steps of high school and in university, that is, a premium in academic achievement for the oldest students at least until the end of adolescence. Second, we also find significant differences by gender. More specifically, being among the oldest increases by 0.75 points the entry score and by 0.15 points the university grades for girls but there is no significant effect on boys. This last result could be indicating that the channels through which the month-of-birth effect operates are likely to be subjective to a large extent. In this regard, and as we will explain further below, female students may be more sensible to this recurring cycle of behavioral responses than men, as there are several studies indicating this gender particularity.

In brief, this paper adds to the important body of research addressing the month-of-birth impact on academic performance with special focus on differences by educational stages and by gender. In addition, we also provide evidence of heterogeneity of the results across the students ability distribution. To this end, we use administrative data of individuals studying at the Faculty of Economics and the Faculty of Medicine of the University of Valencia (Spain) during the period 2010-2014. The data allows me to investigate the effect at two educational stages for the same sample of students: (i) the access-to-university exam and (ii) university. Therefore, to measure academic performance at these stages we use the *entry score* and *university grades*, respectively. On the one hand, the entry scores result from a weighted average of the grades obtained in high-school (40%) and a final regional-level (Autonomous Communities) standardized exam (60%). In Section 3, we offer more details about the composition of this entry score. On the other hand, university grades correspond to the final result obtained in each module of the degree.

Our objective is to identify the causal effect of being among the oldest rather than among the youngest within a cohort on academic performance. In order to capture this causal impact, we apply one of the most widely used methods when natural experiments are not available: Regression Discontinuity (RD). This method analyzes the existence of a discontinuity in the conditional mean of the outcome variable (Y) at a cut-off imposed by the running variable (X), which is the variable determining eligibility into the treatment group. In our case, the outcome variables are either the (i) entry score or (ii) the university grades and the running variable is always the distance in days from the cut-off. Hence, we create a cut-off in 1st January to determine whether an individual is among the oldest (right to the cut-off) or among the youngest (left to the cut-off) within her cohort.

Our results show that there is a causal effect of being among the oldest rather than among the youngest on academic performance for women but not for men, the impact attenuates once the students enter into the university, and it is more noticeable in the upper quantiles of the distribution of both the entry score and the university grades. In particular, being among the oldest is related to an increase of 0.75 points on the entry score and 0.15 points on the university grades for girls, while we find no significant effects in the case of male students.

These results are in line with previous research but contribute to our knowledge about the differences (a) by gender, (b) educational stage and (c) across the ability distribution.

The rest of the paper is organized as follows. Section 2 reviews the related literature and explain our contributions. Section 3 presents the institutional framework of our analysis and show some descriptive statistics of our estimation sample. Section 4 explains the theoretical foundations of the Regression Discontinuity methodology, discusses the main results and examines the validity of our estimations. In Section 5 we present the main conclusions of our research.

2 Related literature and contributions

This paper relates and contributes to, at least, two strands of the literature. Primarily, this paper is related to the literature on the month-of-birth effect. Though secondarily, it is also related to the abundant recent literature documenting the gender differences in responses to external stimulus in professional and educational settings.

A considerable amount of previous research documents that individuals born during the first months of the academic year tend to have better educational achievement than those born in the last ones (McEwan and Shapiro 2008; Crawford et al. 2010; Crawford et al. 2011; Puhani and Weber 2008; Smith 2009; Lima et al. 2019). In fact, according to Crawford et al. (2014) and Pedraja-Chaparro et al.(2015) the youngest students not only report lower grades during primary school but are also more likely to repeat an academic year and to have an early drop from school.

As regards the time horizon of the month-of-birth effect, several studies find that the effect is relevant beyond the earlier educational stages of primary school. For example, Grenet (2009) finds that the *oldest premium* is significant until the last course of secondary education even if it is much lower at this point. At university level, Pellizzari and Billari (2012) find, somewhat surprisingly, that the youngest first-year students at Bocconi University perform better than their older peers, and Russell and Startup (1986) encounter the same result analyzing more than 300,000 students in the UK. Analyzing not only the academic environment but looking also into the professional life of individuals, Peña (2017) shows that, on average, the oldest students complete more years of college, are less likely to be unemployed, earn higher wages and have more employer-provided medical insurances.

Our paper contributes to this strand of research along two lines. First, we provide here fresh evidence of a positive month-of-birth effect, while adopting a gender perspective that uncovers that such effect is attributable to the female sample. Second, our paper contributes to the discussion of the time horizon of the effect by analyzing the results obtained by a same sample of individuals in two different stages of their academic lifes: the access-to-university examination score, obtained at the end of high-school, and the later university stage. Our results indicate that this effect is huge on the entry score but much more lower on the first-year university grades. This latest result might be explained by the higher number of elements that influence a college student performance, such as a new institutional setting, colleagues, professors, examinations, etc.

Which mechanisms may help explain the long-lasting nature of the month-of-birth effect?. Page et al. (2017) and Page et al. (2018), for example, discover that individuals who have been among the oldest in their cohorts show higher levels of self-confidence, greater tolerance to risk and competitive environments and tend to trust more other people. Also in this line, Hanly et al. (2019) find that the oldest students develop better academic and social skills than the youngest ones. According to Ando et al. (2019), these students seem to have even a better emotional well-being. Furthermore, Crawford et al. (2014) claim that being among the oldest is associated with a higher confidence in self-perceived scholastic capacity.

The key point is that all these personality traits are very well rewarded in academic examinations as these are based in competition and self-confidence. Therefore, the oldest students that possess these personal skills experiment higher success in school. This success leads to a more positive and enhancing feedback of professors and family, a better consideration of the peers and higher levels of self-esteem. All these external influences reinforce the aforementioned personality characteristics that are very helpful to be academically successful. In this way, what we call a *virtuous circle* is generated: higher self-confidence and believe scholastic competence improves academic outcomes, this produces success and positive feedback by the individual's environment which at the same time increase self-confidence and believe scholastic competence. This *virtuous circle* creates resilient and robust personalities that help more the oldest than the youngest during their academic and professional life.

We see that this positive recurring cycle of behavioral responses to success is well documented in the month-of-birth effect literature. However, we also want to know if this *virtuous circle* impacts differently women and men, a question that has been poorly explored. We find several reasons that may explain a higher elasticity of women to the *virtuous circle*. This is because there is a body of psychological research which argues that women are more sensible to their environment's influences. For example, Schawble and Staples (1991) show that women attach higher importance to reflected appraisals (which is the person's perception of how other see and evaluate her or him) than men. In a recent study, Berlin and Dargnies (2016) find that women react more strongly to the feedback they receive from their environment than men. In this line, Mayo et al. (2012) carried out an experiment where student had to evaluate their peers and themselves on four aspects of leadership competence. They observe that women more rapidly align their own evaluations with peer's ratings on them than men. Interestingly, Helgeson and Johnson (2002) discovered that women's self-esteem increased after positive feedback and decreased after a negative feedback in a bank employees evaluation process. Hence, we think that this higher sensibility of women to their environment feedback might explain why we find only significant results on female students: when the *virtuous circle* is generated, women seem to internalize much more its positive effects than men, who appear to attach less importance to their environment influence.

This better understanding of the heterogeneity of results by gender is, in our view, one of the most interesting contributions of the paper. Hence, in addition to the month-of-birth literature, the paper constitutes a new piece of research in the newly resurged gender literature. Concerning the location across the ability distribution, we observe that the effects are concentrated on the upper quantiles of both the entry score and the university grades distribution. This additional finding would be suggesting that more able students are also more capable to benefit from their month-of-birth advantage.

To sum up, this paper contributes to the literature on the matter along two lines. First, it opens the gender differences question as there is an important line of research that supports the higher sensibility of women to external influences. This greater elasticity of female students to the theoretical effects of the virtuous circle may be the reason behind the gender heterogeneity of the month-of-birth effect. Second, we extend the temporal bound of the month-of-birth effect. Many of the above referenced works have shown that the effect is almost not significant beyond secondary school. However, we find that the effect is still very strong just before university, and still positive and significant, although weaker, one year later, once the individuals finish their freshmen year. In addition, we show where the effects are concentrated across the ability distribution, and find that the effects are less clear in the lower part of the distribution and more so from the 40 quantile upwards.

3 Institutional framework and data

In this paper we use individual-level administrative data of students at the Faculty of Economics and the Faculty of Medicine of the University of Valencia (UV), Spain for the period 2010-2014 to analyze the effects of being among the oldest rather than among the youngest on academic performance. As we have advanced, we focus our interest in two educational stages: the (i) *entry score*, taken at the end of high-school, and the (ii) *university grades*.

Students in Spain access the university on the basis of their *entry score* and the specific admission minimum entry score established by each university for each degree and year. The entry score is formed by the weighted average of an access-to-university examination (called PAU) and the grades obtained by the student over the two last years of high school, the so-called in Spanish *Bachillerato*. The average of these last two years in high-school is worth 40%. The access-to-university examination is standardized at the regional level (Autonomous Communities), and has two parts. The first one comprises general subjects, is compulsory to enroll in any Spanish university, and is worth 60%. In the specific part, where students can complete exams related to the field of study they are looking forward to register, the access grade can increase up to 4 points. In total, students can get a maximum of 14 points and a minimum of 5 at the entry score which will determine their eligibility for a particular university degree. Once the students enter in the university, we have data about their grades in each module of the first year of their degree. The university grades ranges from 0 to 10 and the minimum grade to pass the module is 5. As we have said before, all the students in our database are enrolled either at the Faculty of Economics or the Faculty of Medicine. The first one is the public college to study business and economics-related fields at the University of Valencia, one of the largest public universities of Spain. The institution offers a wide range of well recognized four-year (new ones, called *grados*) and five-year (old ones, called *licenciaturas*) degrees, mainly in business and economics. The second one, is the Spanish equivalent of the UK-US Medical Schools at the University of Valencia. It offers degrees in medicine and dentistry, which usually have the highest entry score. In the rest of the paper, we use an estimation sample of students who (a) have entered into the university trough the *PAU* and European Baccalaureate (EB) examinations and (b) are aged 19 or 20 at the end of their first year at the university. For these students, we will count on information on their entry score and on their first-year modules at the university. We do this selection due to the following reasons. The first reason is comparability. The *PAU* exam (which has been described above) is the most typical way to get into university (65% of the students in the original database have passed it) and focusing on this exam we am comparing the score of people that have faced the same type of test. We also include those coming from European Baccalaureate as the test is similarly organized and after the conversion the scores are also bounded between 5 and 14 (but they only represent 2% of the original database). The second reason is to reduce as much as possible differences in the format of the PAU exam taken. The vast majority of those students who are present in our sample at the end of the first year at the university and who passed either the *PAU* or the *EB* are aged 19 or 20. Yet, the important point is that selecting only these two cohorts we considerably reduce the possibility that a student takes the entry exam just after high school but enroll into university several years later. This would imply a lot of variability in the entry exam (which would deteriorate the comparability among them) as its format has changed during the last 20 years due to political reforms and European integration. Furthermore, we believe that students aged 19 or 20 that have not entered by *atypical* ways (professional training, elite sport, etc) are pretty likely to have taken the exam just after the end of high school. Finally, we have chosen only the students that have just finished the first year of their undergraduate programs because we want to know whether the effects of the *virtuous circle* described in the former section are significant or not once the students finish their next academic step.

In the following Table 1 we offer some family and students' educational and economic background characteristics. We present the results dividing a year in two semesters (from

January to June, and from July to December) to show that there are no significant differences in these variables between those born in the first semester and those born in the second semester, that is, between relatively older and relatively younger students. In fact, in the table we can see that the distribution over the two parts is very similar. Furthermore, we see that the vast majority of the students have entered the university through the *PAU* exam.

Table 1 : Background and individual's characteristics

Variables	Born in first semester	Born in second semester
Mother with tertiary education (%)	39.23	39.52
Father with tertiary education (%)	39.19	40.66
Mother with high-skill job (%)	56.81	56.14
Father with high-skill job (%)	70.36	70.93
Female students (%)	56.94	55.79
Entered through *PAU* examination (%)	95.07	95.24
	100%	100%

4 Empirical strategy and main results

This section has five objectives. Firstly, we present the Sharp RD design used to identify the causal effect of being relatively older within a cohort on academic performance and explain why we use this technique. Secondly, we show the main results from this methodology and offer a benchmark standard regression with background controls. Thirdly, we use the simultaneous quantile regression (SQR) to study in which parts, if any, of the ability distribution the treatment effects are concentrated. Fourthly, we perform falsification checks to proof the validity of our RD estimations. Finally, we explore whether or not the effect of being among the oldest rather than among the youngest has a significant impact on the university grades.

4.1 Sharp RD design

The main goal here is to understand the causal effect of being among the oldest in a cohort on two differentiated educational stages: (i) the entry score and (ii) the university grades. The analysis of this kind of causal effect is straightforward when the treatment (being among the oldest of a cohort) can be randomly allocated because this fully guarantees the comparability of individuals allocated to the treatment and control groups. Nevertheless, due to the nature of the relationship at hand, it is not possible to perform a randomized control trial (RCT) and assess the treatment effect. This is because the students are born at either the beginning or end of the academic year, but not both. Therefore, we cannot randomly assign some individuals of the sample to the treatment (being among the oldest) or to the control (being among the youngest) groups because we simply cannot change the birthday of an individual. When randomized experiments cannot be carried out, one of the most credible non-experimental techniques for the analysis of causal effects is Regression Discontinuity (RD) design (Cattaneo, Idrobo and Titiunik2017). This technique studies the existence of a jump or discontinuity in the conditional mean of the outcome variable (Y) at a threshold or cut-off imposed by the running variable (X), which is the variable determining eligibility into the treatment group. In this case, the outcome variables are either the (i) entry score or (ii) the university grades, and the running variable is always the distance in days from the cut-off. Given that our objective is to capture the causal effect of being born in the first months (treatment group) of the academic year rather than in the last ones (control group), we set the cut-off in the first day of the year: 1st January.[2] An essential assumption in the standard RD analysis is that, in the absence of treatment, the relationship between the outcome and running variable is continuous (this explains why this standard approach is also known as *continuity-based RD*). Thus, in our study, individuals' entry score and university grades are assumed to be a continuous function of the distance in days from the 1st January cut-off, but the treatment (if it exists) makes them to jump at this cut-off (provoking the discontinuity). This is why this method is called regression *discontinuity*. There are two types of RD designs: *sharp* and *fuzzy*. In the first case, the treatment necessarily occurs whenever the running variable overpasses the cutoff; in the second one, instead, it is the probability of treatment that jumps at the cutoff. In the *fuzzy* RD, there exists the possibility that some individuals do not enter into the treatment group

[2]The intuition behind this specific cut-off can be easily understood with the following example. If an individual was born in January 7, her running variable equivalent would be "+6" because she was born six days after the 1st January cut-off. This would rank her among the oldest in her cohort because she would be situated to the right and very close to the cut-off. Similarly, if an individual was born in December 25, her running variable equivalent would be "-6" because she was born six days before the 1st January cut-off. This would rank her among the youngest in her cohort because she would be situated to the left and very close to the cut-off.

even if they overpass the eligibility cut-off [3]. In *sharp* RD, all the individuals with a running variable value higher than the cut-off receive the treatment. Furthermore, in this type of RD the treatment status is a discontinuous function of the running variable as no matter how close the running variable is to the cutoff, the treatment remains unchanged until this cut-off is reached (Angrist and Pischke, 2014). In this study, the RD design is *sharp* because when an individual is born after the 1st January cut-off she is automatically classified among the oldest in her cohort (which is the treatment) and if she is born before the 1st January cut-off she is classified among the youngest (which is the control group). Therefore, it can be seen that in our case the cut-off fully determines whether or not a student experiments the treatment. To formalize we follow the notation of Cattaneo, Idrobo and Titiunik (2017). We assume that there are n students, indexed by $i = 1, 2, ..., n$, each student has a running variable value X_i (distance in days from the cut-off), and the established cut-off is noted by c (1st January). Then, those individuals with $X_i \geq c$ are assigned to the treatment group and those with $X_i < c$ to the control group. Thus, this treatment assignment that we call T_i is defined as $T_i = 1(X_i \geq c)$. To illustrate the technique, consider a simple regression function as follows:

$$y_i = \alpha + f(x_i) + \tau_i + u_i \tag{1}$$

where α is the constant, y_i is the outcome variable (students' entry score and or university grades), x_i is the running variable (distance in days from the cut-off) and T_i is the variable indicating treatment which equals 1 for being the among the oldest (treatment group) and 0 for being among the youngest (control group). The treatment effect we want to analyze is τ. In order to explain how τ is calculated, firstly we have to understand the two potential

[3] A good example to understand the *fuzzy* RD is Beneito and Rosell (2019). They study the effect of belonging to a high-ability group on the university examinations scores. In their research, we see that some individuals that have an entry scores greater than the high-ability group cut-off may decide not to enter in the high-ability group and stay in the mixed-ability one. Therefore, in *fuzzy* RD designs an individual with a running variable greater than the cut-off does not necessarily receive the treatment whereas in the *sharp* RD the individual always receives the treatment whenever his running variable surpasses the cut-off.

outcomes that an individual can have:

$$Y_i(0) \quad if \quad X_i < c \tag{2}$$

$$Y_i(1) \quad if \quad X_i \geq c \tag{3}$$

where $Y_i(0)$ refers to the outcome that would be observed under the control conditions and $Y_i(1)$ represents the outcome that would be observed under the treatment conditions. The fundamental problem of causal inference arises because, even if every individual is assumed to have both $Y_i(1)$ and $Y_i(0)$, only one of them is observed for each individual (a student either receives the treatment or not, but not both). In our specific sharp RD setup, the same problem takes place as we only observe the outcome under control, $Y_i(0)$, for individuals whose running variable is smaller than the cut-off and we only observe the outcome under treatment, $Y_i(1)$, for those individuals whose running variable is greater than the cut-off. Therefore, the observed average outcome given the running variable is:

$$E[Y_i|X_i] = \begin{cases} \mathrm{E}[\mathrm{Y}_i(0)|X_i] & \text{if} \quad \mathrm{X}_i < c \\ \mathrm{E}[\mathrm{Y}_i(1)|X_i] & \text{if} \quad \mathrm{X}_i \geq c \end{cases} \tag{4}$$

where $E[Y_i(0)|X_i]$ is the observed average outcome when the running variable is smaller than the cut-off and $E[Y_i(1)|X_i]$ is the observed average outcome when the running variable is higher than the cut-off. The calculation of the RD treatment effect is based on the comparison of these two possible outcomes. To make it easier, we show in Figure 1 the graphical representation of Cattaneo, Idrobo and Titiunik (2017) who plot these observed average outcomes for both cases against the running variable. The average treatment effect at a specific value of the running variable is represented by the vertical distance between the two lines at that specific value of the running variable. The problem, as we have advanced, is that this distance cannot be estimated because we do not observe both curves for the same range of values of the running variable. In fact, the only point in which both lines are almost observed is at the cut-off c. Hence, the technique assumes that individuals whose running variable value is equal to the cut-off ($X_i = c$) or just above it (whose outcomes are observed and receive the treatment) are comparable to those whose running variable value is just below the cut-off (whose outcomes are observed and do not receive the treatment) [4]. Therefore, we can approximatively calculate the vertical distance at the cut-off represented by $\{\mu_{+}\mu_{-}\}$ in Figure 1 comparing the observed outcomes of individuals just above and just below the cut-off. This *comparability* assumption between individuals with very similar values of the running variable but on different sides of the cut-off is the key idea in which the RD design bases its treatment effect calculation.

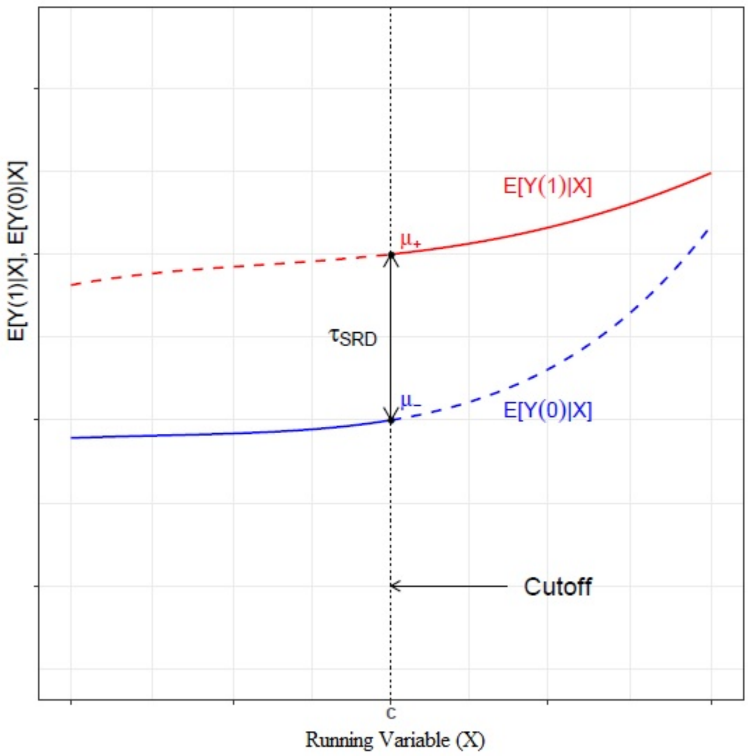

Figure 1 : *Treatment effect in Sharp RD Design (Source: Cattaneo, Idrobo and Titiunik, 2017)*

The formal support for this assumption was firstly provided by Hahn, Todd and van der Klaauw (2001). Under certain continuity assumptions, the authors proved that if regressions $E[Y_i(1)|X_i = X]$ and $E[Y_i(0)|X_i = X]$ are continuous functions at $X = c$ then we can say that average potential outcomes are continuous functions of the running variable at the cut-off. Thus, the treatment effect is equivalent to the difference between the limits of the treated and control average observed outcomes as the running variable converges to the

[4] In fact, this assumption seems very reasonable in our case: if the date of birth would not affect the academic performance, the grades of the students just before and just after the cut-off should be very close, otherwise there is something other than the date of birth that explains the discontinuity in grades near to the cut-off.

cut-off. Formally:

$$\tau = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{-x \to c} E[Y_i|X_i = X] - \lim_{+x \to c} E[Y_i|X_i = X] \tag{5}$$

An important point that should be made when it comes to the implementation of the RD technique is the distinction between *non-parametric* RD and *parametric* RD. In the first case, an optimal bandwidth determines the window encompassing the comparable individuals just above and just below the cut-off is used to calculate the treatment effect. This bandwidth is mostly data-driven and makes an optimal balance between the increase in bias caused by taking a wide window of individuals (students less and less comparable) and the loss of efficiency (due to fewer observations if the window is too narrow). Put differently, in an ideal world, the majority of the observations would be situated very close and around the cut-off in order to ensure both a small bias (because they would be very comparable) and a high efficiency (because of the high amount of observations available around the cut-off). The following Figure 2 illustrates in theoretical terms an optimal bandwidth in non-parametric RD:

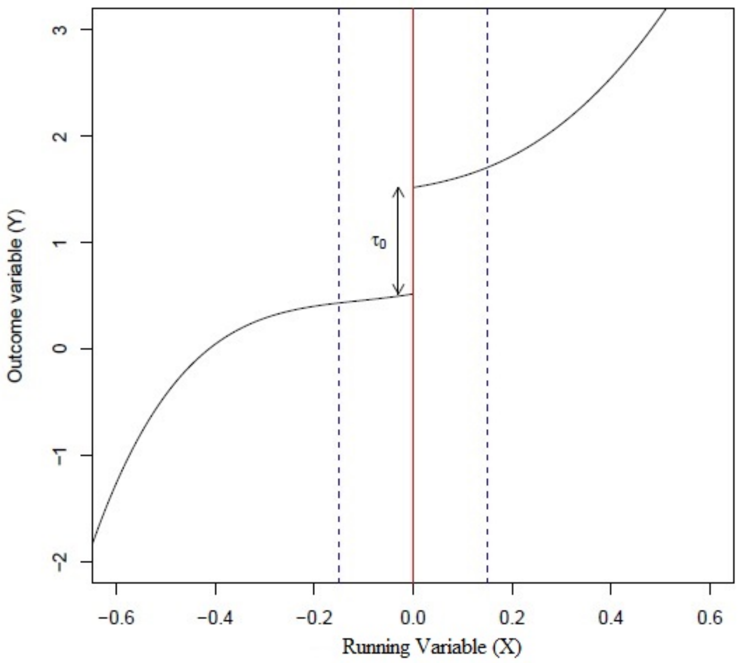

Figure 2 : *Optimal bandwidth in non-parametric Sharp RD Design (Source: Cattaneo 2015)*

Importantly, in this non-parametric RD setting the specification of Equation (1) allowing or not for non-linearities does not matter since the optimal data-driven bandwidth actually concentrates in individuals around the cut-off, where differences between the two specifications are not appreciable. However, in the parametric RD all the individuals of the sample are taken into account when estimating the treatment effect. In this later setting, controlling for eventual non-linearities of $f(x_i)$ is crucial to not mistake a non-linearity for a discontinuity.

The local character of the non-parametric RD treatment effect is an usual critique made to this technique. This criticism points out that the treatment effect estimation is carried out using only the observations within the optimal bandwidth and it is therefore not extensible beyond the optimal bandwidth. However, this critique is not at all an issue in our research question as the effect of being among the oldest rather than the youngest is actually situated in the neighborhood of the cut-off. This is because we consider that the advantages provided by the *virtuous circle* are concentrated in the individuals born in the very first months of the academic year, which are approximatively the ones taken into account by the optimal bandwidth. Therefore, in the next section 4.2 we use as a main model the non-parametric version of the Sharp RD to estimate the treatment effect. As a robustness check, we also carry out a parametric Sharp RD but setting the non-parametric optimal bandwidth to check whether or not the size and sign of the treatment effect changes. Finally, in section 4.3 we carry out three of the most widely used falsification checks to test the validity of the causal effect estimated.

4.2 Baseline results for Entry Score

In this section we provide the results of our estimations in the following order. Firstly, we estimate a benchmark standard regression to offer a hint of the relationship between the month of birth and academic performance. Secondly, we run our main model, the non-parametric Sharp RD, to study the causal effect of being among the oldest rather than among the youngest. Thirdly, we estimate a parametric Sharp RD using the optimal bandwidth of the former non-parametric version to assess whether the results vary by changing the estimation method of the Sharp RD.

4.2.1 Benchmark standard regression

Table 2 shows the effect of month of birth on entry score estimated by OLS and separated by gender [5]. We also allow for non-linearities in this relationship and offer two different specifications for each sub-sample: without any controls and with family background controls. Regarding the first case, it can be seen that the linear effect of the month of birth is negative and statistically significant for the whole sample and for female students while there are no significant effects for the male group. This means that when the month of birth increases (the individual is relatively younger) the entry score decreases, which is in line with the literature discussed above.

These results provide us an important hint about how the mechanisms discussed in the introduction work in practice: the statistical significance for the whole sample disguises one of the key results of this paper, which is that female students are more sensible to the so-called *virtuous circle*. In fact, the effect of the month of birth on entry score is strong and statistically significant only when we use the sample of female students. Furthermore, this phenomenon holds when we control for background characteristics of the student, such as parents' education attainment or economic status (that are always statistically significant for both genders and in all specifications).

Table 2 : The effect of month of birth on entry score

	(1) All: without controls	(2) All: with controls	(3) Girls: without controls	(4) Girls: with controls	(5) Boys: without controls	(6) Boys: with controls
m_birth	-0.067**	-0.069**	-0.103**	-0.094**	-0.001	-0.034
	(0.031)	(0.030)	(0.042)	(0.040)	(0.010)	(0.045)
m_birth2	0.004	0.004	0.006*	0.004		0.002
	(0.002)	(0.002)	(0.003)	(0.003)		(0.003)
studfather2		0.664***		0.637***		0.703***
		(0.063)		(0.086)		(0.091)
studmother2		0.427***		0.624***		0.207**
		(0.066)		(0.090)		(0.096)
eco_father2		0.173***		0.173**		0.156*
		(0.060)		(0.081)		(0.086)
eco_mother2		0.132**		0.223***		0.005
		(0.056)		(0.075)		(0.082)
yofbirth	0.721***	0.695***	0.691***	0.655***	0.734***	0.719***
	(0.016)	(0.016)	(0.022)	(0.021)	(0.023)	(0.023)
Constant	-1,426.534***	-1,375.909***	-1,365.898***	-1,295.826***	-1,454.208***	-1,424.408***
	(31.965)	(30.971)	(44.055)	(42.208)	(45.879)	(44.905)
Observations	5,903	5,903	3,328	3,328	2,575	2,575
R-squared	0.256	0.308	0.229	0.301	0.283	0.320

[1] Standard errors in parentheses. Statistical significance: ***$p < 0.01$, **$p < 0.05$, * $p < 0.1$.

[2] Control variables: *studfather/studmother* are dummy variables that take the value of 1 when parents holds a bachelor degree or more and 0 when they have pre-college education. *ecofather/ecomother* are dummy variables that take the value of 1 when parents have a medium-high skill paid job and 0 where the parents have a low-skill job or are unemployed.*yofbirth* represents the year of birth and controls that we compare individuals born in the same year.

[5] This division by gender is going to be used in all our estimations. "ALL" encompasses the whole selected sample, "GIRLS" only the female students within this sample and "BOYS" only the male students within this sample.

4.2.2 Non-parametric Sharp RD

In this section we present the results obtained through the application of the non-parametric Sharp RD estimation. The first outcome variable to which we apply this technique is the entry score obtained by students at the end of high school, and the running variable is the distance in days from the 1st January cut-off.

The crucial identification assumption of the RD technique is that the relationship between the outcome variable (entry score) and the running variable (when in the year the individual is born) must be continuous in absence of treatment. This implies, first, that the outcome variable would not jump at the threshold if no treatment effect exists, and, second, that the relationship is continuous (no jumps) outside the threshold. To show some evidence in this regards, and before we discuss the RD estimation results, we present in Figure 3 the entry score (Y) plotted against the distance in days from the 1st January cut-off (X) for the whole year. This figure provides interesting evidence in favor of such continuity. In the figure we see that the year is divided in two parts (semesters) on the x-axis: the half to the right approximatively corresponds to the first 6 months of the year and the half to the left to the 6 months before the end of the year. Therefore, if the relationship between the two variables is continuous, the fitted lines on both sides of the cut-off should converge to similar entry scores towards the end of June; that is, the end points to the right and to the left of the fitted lines correspond to a similar value on the vertical axis. We can see that this convergence happens in the three cases, thus providing a first piece of evidence in favor of the continuity hypothesis.

Next, in Table 3 we present the results of the RD estimation. The table displays the bias-corrected and robust non-parametric estimation for Sharp RD recommended by Calonico, Cattaneo, and Titiunik (2014a and 2014b) in addition to the conventional non-parametric coefficient.[6]. This non-parametric focuses on the observations around the cut-off, which are determined by the chosen bandwidth. In our case, the bandwidth used corresponds to the updated almost data-driven optimal bandwidth calculation proposed by Calonico, Cattaneo and Farrell (2018). The year of birth is included as a covariate to control for the potential

[6]Henceforth, all our non-parametric Sharp RD estimations are going to show these three coefficients and use this optimal bandwidth calculation.

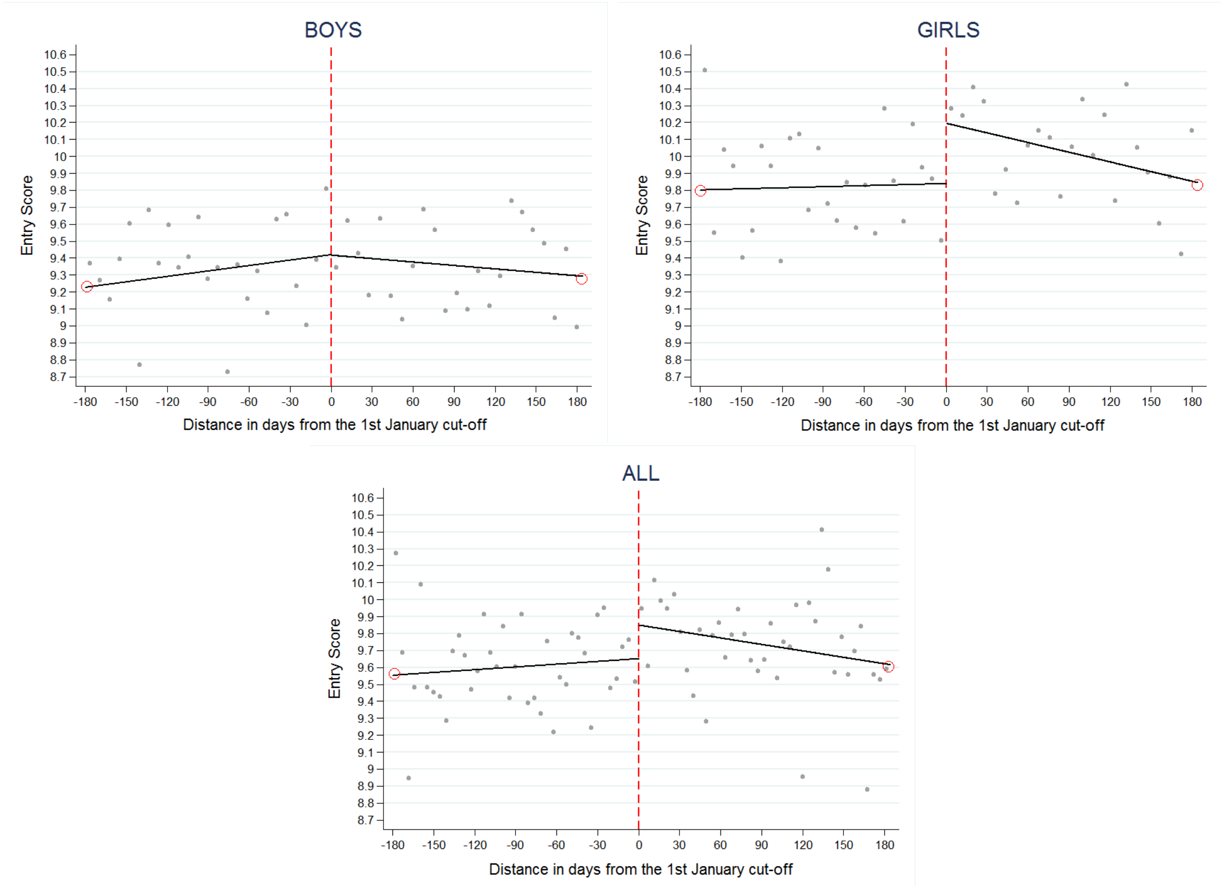

Figure 3 : *RD plot - Entry score and Distance in days from the cut-off*

specific effects that may be taking place in each cohort.

In the table we see that the treatment effect is only statistically significant for the female subsample. This confirms the results obtained in the former standard OLS regression: girls seem to be more sensible to the *virtuous circle* reinforcement than boys. From the estimation, we know that for this group the optimal bandwidth calculation has selected approximately the girls born within the 2 months before and within the 2 months after the 1st January cut-off. Therefore, here being among the oldest (treatment) means being born within the first two months of the year and being among the youngest (control) implies being born within the last two months of the year. Consequently, we can say that, on average, the fact that a girl has been among the oldest in her cohort increases her entry score by 0.671 or by 0.75 according to the bias-corrected and robust estimates, which are our preferred measures of the treatment effect.

This estimated effect can be considered sizable in quantitative terms. In fact, it can be

Table 3 : Non-parametric Sharp RD estimation for entry score

	(1) All	(2) Girls	(3) Boys
Conventional	0.235	0.671**	-0.195
	(0.178)	(0.269)	(0.265)
Bias-corrected	0.255	0.750***	-0.263
	(0.178)	(0.269)	(0.265)
Robust	0.255	0.750**	-0.263
	(0.214)	(0.316)	(0.312)
Observations	5,903	3,328	2,575

[1] Standard errors in parentheses. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the difference between getting into the desired degree or not. For instance, the average girl in the female sub-sample has an entry score of 9.92 over 14. As an example, we can take the minimum entry score of the year 2019 to get into the Business Administration and Tourism degree has been set at 10.62 over 14. This implies that, due to the treatment effect help, on average, the relatively old girl would get into this degree (9.92 + 0.75=10.67) and the relatively young (9.92) would not.

As a complement to Table 3, in Figure 4 we offer a graphical representation of the estimated treatment effects for each group, with a selection of observations (days to the right and to the left of the cut-off) that coincides with the non-parametric selected sample. The figure shows that the final estimation results are quite in line with those anticipated in Figure 3 for the whole sample.[7])

This result is one of our main contribution to the literature on the matter. As we have discussed in section 2, there is a considerable body of research that explains why being relatively old in your cohort can improve your academic achievement. Nevertheless, little has been said about the gender divergences in this relationship. We believe that this divergence in the results confirms our initial guess: girls have a greater elasticity to their environment influences than boys, for the good or the bad. Therefore, in our study, it seems that girls incorporate the positive effects of the *virtuous circle* in a greater extent than boys. Furthermore, our result is in line with Grenet (2009). The author claims that the effect of

[7] I represent the conventional coefficients because are the ones used by the *rdplot* STATA package.

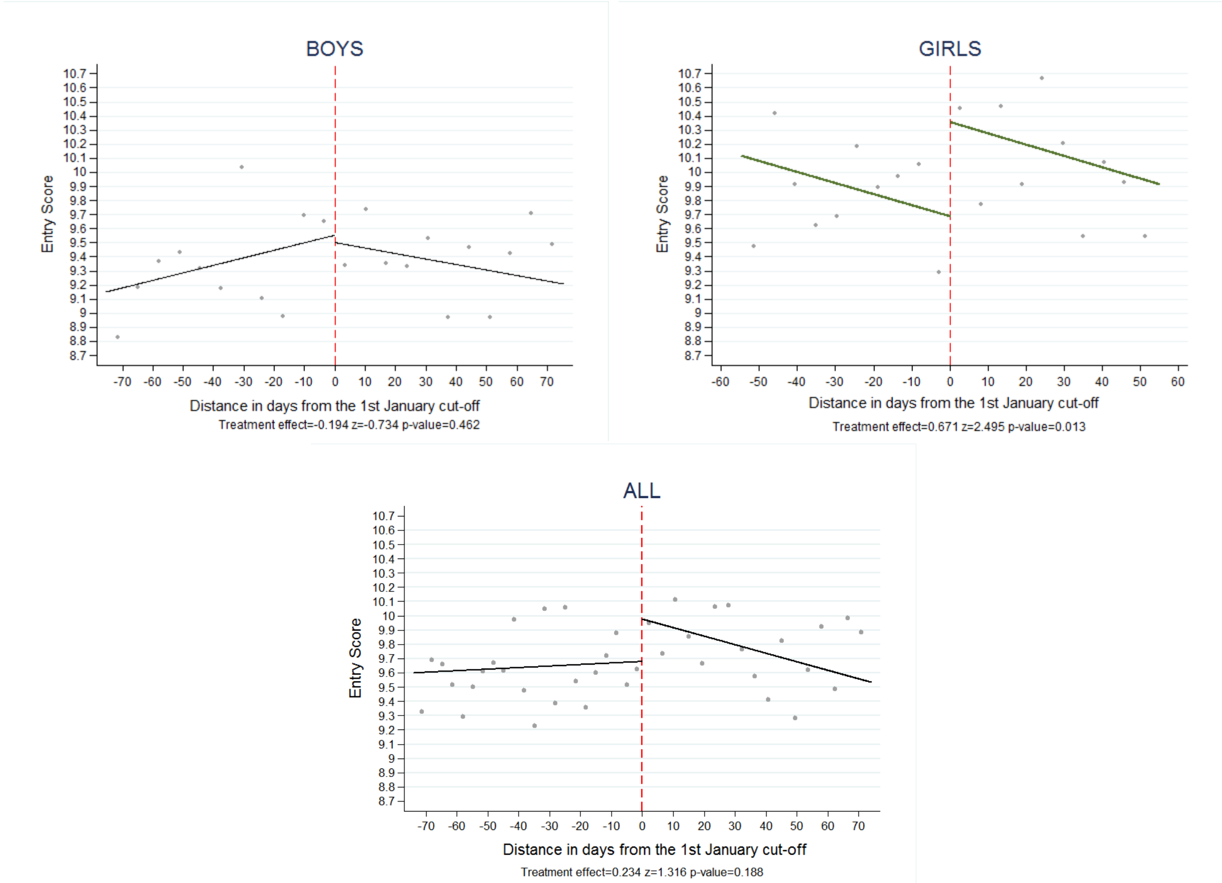

Figure 4 : *The non-parametric Sharp RD conventional treatment effect*

being among the oldest rather than the youngest on academic performance decreases when children get older, but it is still appreciable at the end of the secondary education (which is exactly the point of the academic life in which our estimations are made).

4.2.3 Parametric Sharp RD

Our goal here is to explore whether the treatment effect calculated in the former subsection varies when we switch from the non-parametric to the parametric Sharp RD. For this purpose, we take the optimal bandwidth calculated for each group in the former subsection to run a parametric Sharp RD estimation. In other words, we only take into account the effective number of observations determined by the optimal bandwidth in the non-parametric setting to produce the estimations. Although it is true that in section 4.1 we have explained that the parametric Sharp RD takes into account the whole sample, we bind the estimation to the optimal bandwidth because we want to carry out a parametric replication of our main model, the non-parametric Sharp RD.

Table 4 shows the results from the parametric Sharp RD estimation. We also offer two specifications for each group: linear and non-linear with crossed effects. The variable indicating the treatment is *oldest*, which is a dummy variable that takes the value of 1 when the distance in days from the cut-off is positive (first two months of the year) and 0 when the distance in days from the cut-off is negative (last two months of the year). The control variables are *runDay2*, which is the squared version of distance in days from the 1st January cut. *oldrunDay* which is a dummy variable that allows for different running variable coefficients to the left and to the right of the cut-off, *yofbirth* represents the year of birth and controls that we compare individuals born in the same year. In the table we see that again only the female sub-sample coefficients appear to be statistically significant. Regarding the size of both coefficients, we see that they are quite similar to those of the non-parametric conventional estimation (from 0.671 in the non-parametric to 0.663 and 0.660 in this parametric replication) [8]. Hence, we can interpret these results as a *robustness test* of our main model findings, which increase our confidence in the size of the causal treatment effect calculated.

Table 4 : Parametric Sharp RD estimations

	(1) All: linear	(2) All: crossed effects	(3) Girls: linear	(4) Girls: cross effects	(5) Boys: linear	(6) Boys: crossed effects
oldest	0.133	0.127	0.663***	0.660***	-0.313	-0.317
	(0.156)	(0.156)	(0.247)	(0.247)	(0.223)	(0.223)
runDay	0.000	0.012	-0.008**	0.001	0.005*	0.018*
	(0.002)	(0.008)	(0.004)	(0.016)	(0.003)	(0.010)
runDay2		0.000		0.000		0.000
		(0.000)		(0.000)		(0.000)
oldrunDay		-0.024		-0.017		-0.026
		(0.015)		(0.032)		(0.020)
yofbirth	0.732***	0.733***	0.737***	0.738***	0.723***	0.723***
	(0.025)	(0.025)	(0.040)	(0.040)	(0.035)	(0.035)
Constant	-1,449.488***	-1,450.775***	-1,458.180***	-1,461.414***	-1,430.737***	-1,430.815***
	(50.354)	(50.358)	(80.176)	(80.553)	(70.503)	(70.561)
Observations	2,394	2,394	985	985	1,102	1,102
R-squared	0.261	0.262	0.259	0.261	0.277	0.278

[1] Standard errors in parentheses. Statistical significance: ***$p < 0.01$, **$p < 0.05$, * $p < 0.1$.
[2] This specific estimation has been carried out only with the observations within the non-parametric optimal bandwidth calculated in the former subsection.

[8]The non-parametric bias-corrected and robust coefficients are slightly higher than the parametric ones because they are calculated in a different way to account for the possible bias and to provide robust errors.

4.3 Validity of the RD methodology: falsification checks

In this subsection we analyze the validity of the Sharp RD results. The objective of this analysis is to determine whether or not there exists causality in the effect calculated by the Sharp RD estimations. As indicated above, the crucial assumption for a causal treatment effect to be identified, is the continuity between the outcome and the running variable in absence of the treatment. This implies that (i) the jump on the outcome variable at the cutoff cannot be the response to any confounder factor also jumping at that cutoff, and that (ii) the outcome variable does not jump outside the cutoff.

Thus, two validate the method we need to rule out that unintended factors are causing the observed jump. In particular, we employ three of the most common so-called falsification checks: (a) density check, (b) background characteristics check and (c) placebo tests check. The first one controls whether the distribution of the running variable is similar at either side of cut-off or not, as big differences in densities between the two sides of the cut-off could bias the estimation. Such differences could exist if individuals could self-select into the treatment anticipating gains from it. In this case, it would imply that parents *plan* an early day of birth for their children expecting a positive impact on academic achievement. We believe that the chances that this auto-selection phenomenon is taking place in our study are rather low. The second falsification test examines whether the characteristics of individuals are similar at either side of cut-off or not. If background characteristic of individuals in both sides are similar then differences in academic performance might be mainly attributable to the relative age effect within a cohort. The third check artificially changes the cut-off to different moments of the year to analyze if there are significant treatment effects during the rest of the year.

(a) Density check. For this check we use two methods: a formal manipulation test and the visual inspection of histograms. In the first case, we perform the updated version of the widely used manipulation test proposed by Cattaneo, Jansson and Ma (2019) that employs a bandwidth selection method based on asymptotic mean squared error (MSE) minimization. This test has as null hypothesis the continuity of the density of the running variable (distance in days from the first January cut-off) at the cutoff point. As we can see in the Figure 5, where we also provide the values of the test statistics and their corresponding p-values, we do not reject the null hypothesis in any of the three groups. Therefore, we can conclude that the density of the distance in days from the first January cut-off is continuous at the cut-off. In the second case, we carry out a visual inspection of the histograms shown

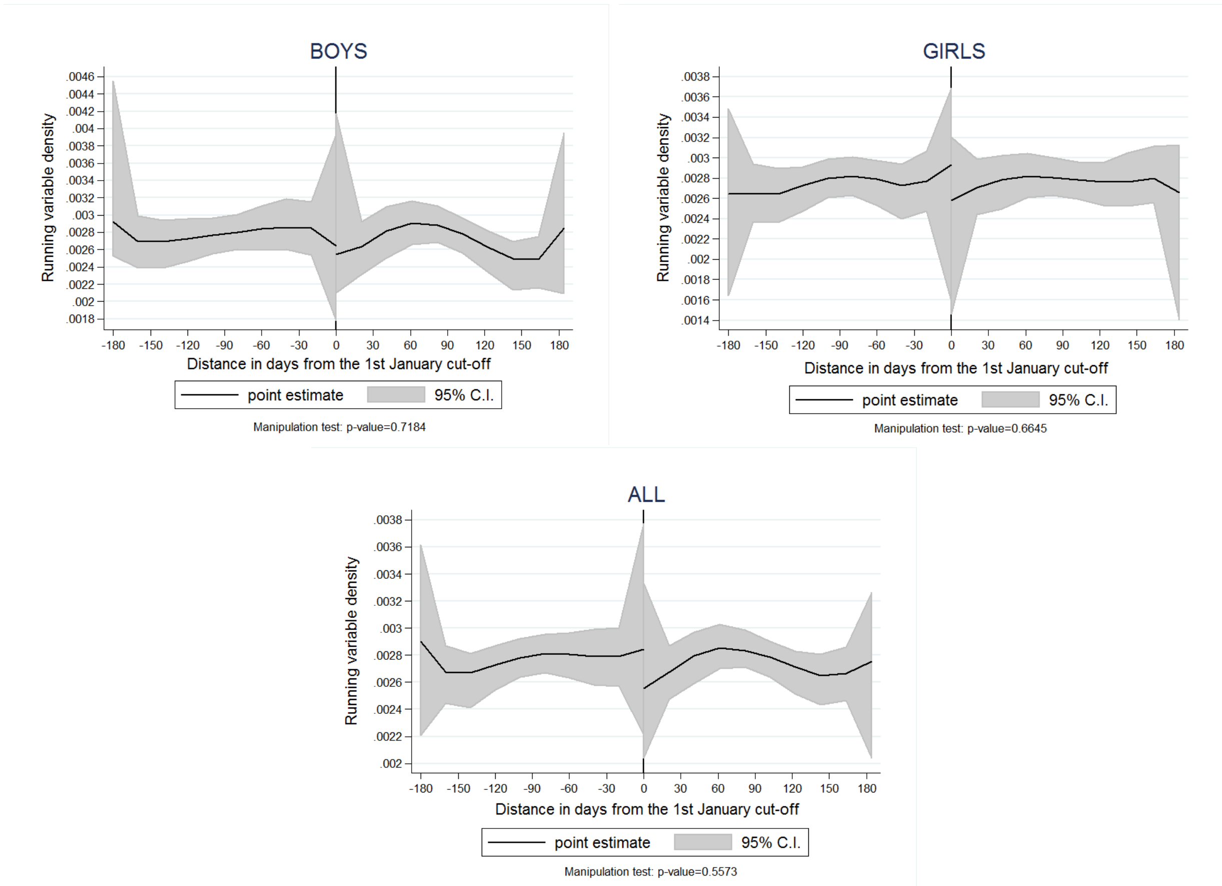

Figure 5 : *Density check - Manipulation test*

in the following Figure 6. These histograms present the frequency of students born in each month. We see that the distribution is pretty homogeneous. Then, we can say that the quantity of students born in each month is very similar, specially in the neighborhood of the cut-off. Summing up, we have proved that the density of our running variable does not show accumulation of frequencies at the cut-off.

(b) Background characteristics check. This validation check aims at controlling whether or not other relevant variables for education achievement are influencing the causal effect estimated. This could occur if such characteristics exhibited jumps at the cut-off.

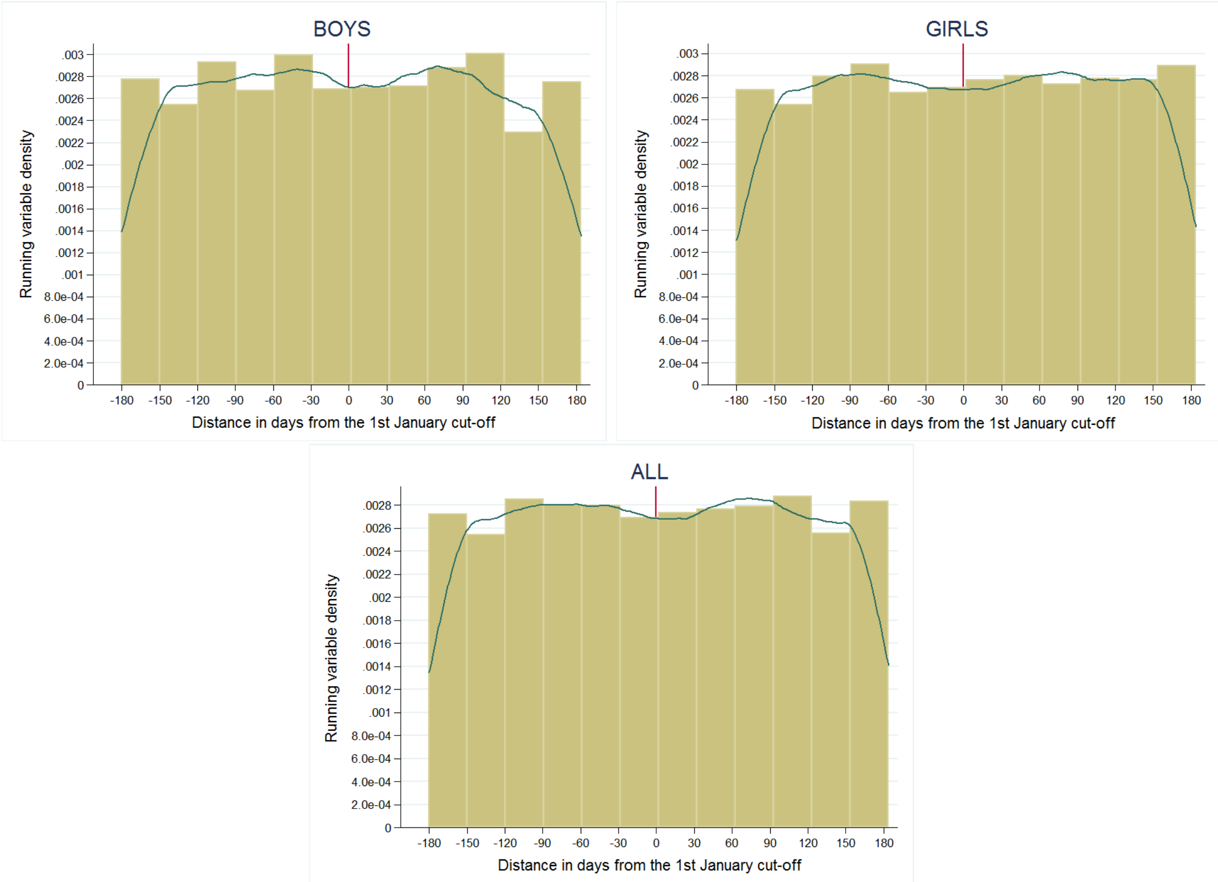

Figure 6 : *Density check - Histograms*

Hence, we run non-parametric Sharp RD estimations for each of the family background variables that are available in our dataset. These are: father education, mother education, fathers economic situation and mothers economic situation. The logic behind this procedure is easy to understand: we want to know if some of these variables are provoking that the oldest students perform better in the entry exam than their youngest peers. Table 5 provides the results of these estimations. We see that there is only one statistically significant coefficient, corresponding to mothers education in the whole sample group. Given that is the only coefficient that appear to be statistically different from zero and is not located in the female group (where the month of birth causal effects are found) we consider that the importance of this coefficient is rather anecdotal. Therefore, we discard the possibility that these family background variables are biasing the treatment effects calculated, which raises our confidence in the validity of the results of our main model.

Table 5 : Background characteristics influence

	(1) Fathers education	(2) Mothers education	(3) Fathers Econ. situation	(4) Mothers Econ. situation
ALL				
Conventional	-0.210	-0.255	0.090	-0.229
	(0.167)	(0.166)	(0.202)	(0.245)
Bias-corrected	-0.247	-0.295*	0.141	-0.245
	(0.167)	(0.166)	(0.202)	(0.245)
Robust	-0.247	-0.295	0.141	-0.245
	(0.198)	(0.197)	(0.238)	(0.294)
Observations	5,903	5,903	5,903	5,903
GIRLS				
Conventional	-0.193	-0.238	0.340	0.018
	(0.208)	(0.213)	(0.278)	(0.419)
Bias-corrected	-0.178	-0.253	0.436	0.072
	(0.208)	(0.213)	(0.278)	(0.419)
Robust	-0.178	-0.253	0.436	0.072
	(0.249)	(0.256)	(0.320)	(0.501)
Observations	3,328	3,328	3,328	3,328
BOYS				
Conventional	-0.210	-0.226	-0.211	-0.372
	(0.242)	(0.237)	(0.313)	(0.414)
Bias-corrected	-0.288	-0.285	-0.216	-0.309
	(0.242)	(0.237)	(0.313)	(0.414)
Robust	-0.288	-0.285	-0.216	-0.309
	(0.279)	(0.278)	(0.374)	(0.494)
Observations	2,575	2,575	2,575	2,575

[1] Standard errors in parentheses. Statistical significance: ***$p < 0.01$, **$p < 0.05$, * $p < 0.1$.

(c) Placebo tests. This is the last check in the validation of results process. We have argued and proved with estimations that the oldest students (born just after the 1st January cut-off) perform better in the entry exam than their younger peers (being just before the 1st January cut-off) and hence claimed that there is a causal link between the month of birth and academic performance. Then, if this is true, we should not find a significant causal effect when we artificially change the cut-off to other days of the year. Otherwise, we would doubt on the discontinuity at the original cut-off as being a casual finding. Therefore, we change the cut-off to 30, 60 and 90 days before and after the original 1st January cut-off and run again the non-parametric Sharp RD estimations. In other words, we move the cut-off 1, 2 and 3 months before and after 1st January. These new estimations at different cut-offs are the so-called placebo tests. As we can see in the footnotes of the following Figure 7 almost all the placebo tests turn to be not statistically significant. In fact, only 2 out of the 18 placebo shown are significants[9]. Finally, we conclude that our results from our main model, the non-parametric Sharp RD, can be taken as causal effects. We arrive to this conviction because (a) there are no density discontinuities of the running variable at the cut-off, (b) there is very little evidence, practically anecdotal, concerning possible confounding effects of family background variables and (c) the vast majority of the placebo tests results are satisfactory.

[9] In the exploratory work we have looked further, up to 6 months after and before 1st January. Hence, we have calculated in total 36 placebo tests but only 2 have been statistically different from zero

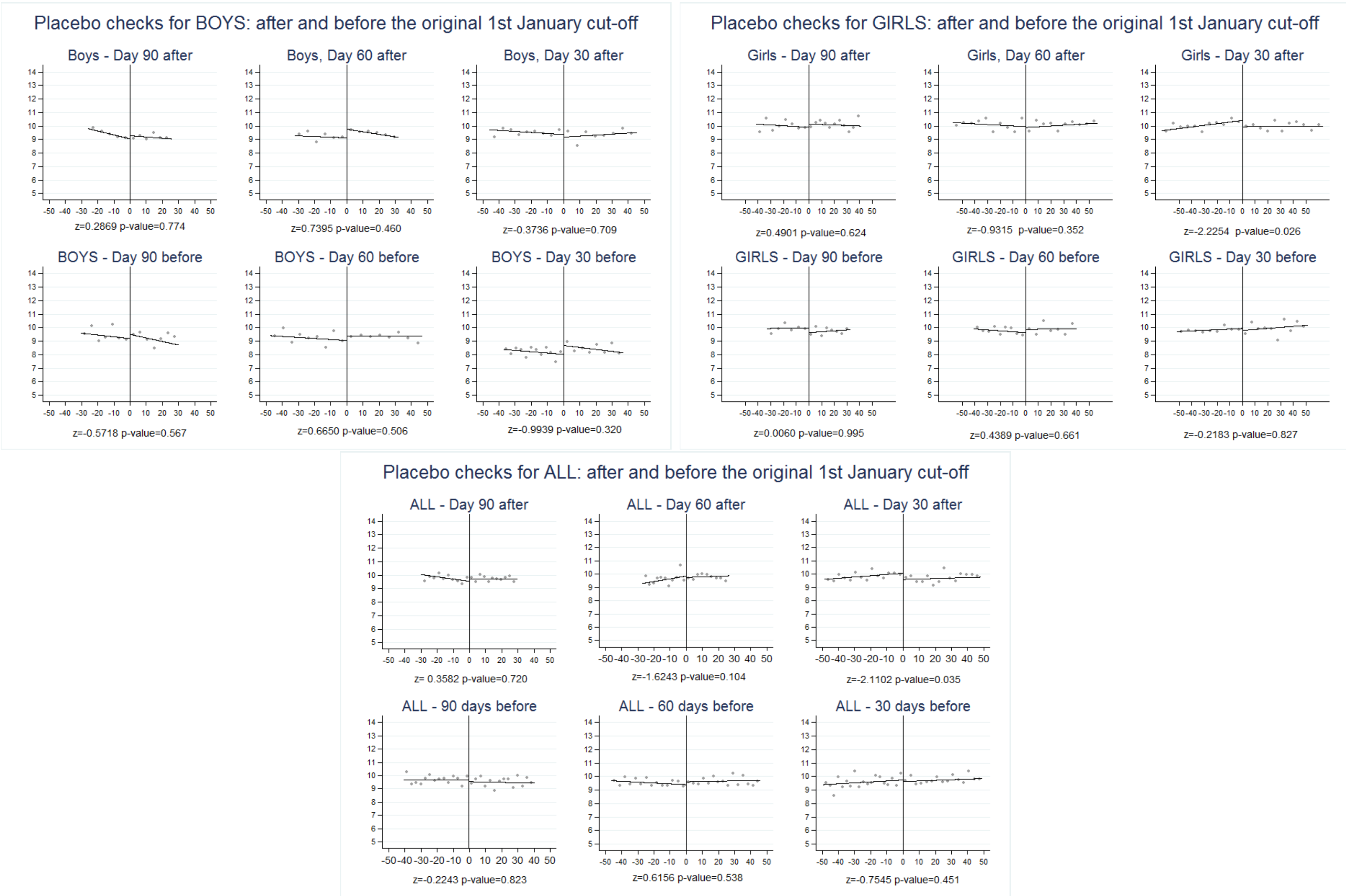

Figure 7 : *Placebo tests check*

4.4 Heterogeneity of results over the ability distribution

Once we have confirmed the validity of our baseline results, we want to explore (i) in which parts of the ability distribution (in our case, the distribution of the entry score) the effects are located and (ii) whether or not the size of the treatment changes across this distribution. For this purpose, we employ the Simultaneous Quantile Regression (SQR). This technique simultaneously carries out different estimations of the same equation putting more weight in the percentiles specified. For instance, in our case, we want to estimate the equation of the parametric Sharp RD (with crossed effects) but with special focus on the 20th (very low ability), 40th (low ability), 60th (high ability) and 80th (very high ability) quantiles of the entry score distribution. In the previous subsections, we have used the optimal bandwidth of the non-parametric Sharp RD to calculate the parametric Sharp RD and we have observed that the coefficient from this parametric approach are comparable to those of the non parametric. Therefore, this coefficient similarity, allow us to approximately estimate the distributional causal effects through the SQR using the equation of the parametric Sharp RD. Table 6 shows the results of the SQR estimation concentrated in the 20th, 40th, 60th and 80th quantiles. The variable indicating the treatment (*oldest*) is only statistically significant in the female group. This is not surprising since the significant causal treatment effects found in our main model are only significants in the female sub-sample. Regarding this group, we see that the effects are rather concentrated in the upper quantiles (40th, 60th and 80th) and with different sizes. Then, in conclusion, we can say that the treatment effect (i) is located in the upper part of the distribution and (ii) its size is homogeneous across the entry score distribution and fairly comparable with the value already shown in Table 3. These are interesting results. On the one hand, the concentration of the treatment effect for students with, say, average and above ability (quantile 40 and above) suggest that below a threshold of academic ability, students do not seem able to benefit from the month-of-birth premium.

On the other hand, the homogeneity of the effect across the upper half of the distribution enlarges the importance and scope of the treatment, being relatively old, concerning the access to university. According to our results, being among the oldest rather than among the youngest not only increases by 0.75 points the entry score for the average girl but also for the top girls. This is very relevant for those girls aiming at high-selective degrees, which are the kind of girls situated at the 80th percentile of the entry score distribution. In our sample, these girls have an entry score of 12.24 over 14. The minimum entry score of this year to get into the Dentistry degree, a very demanded top degree, has been set at 12.59. This implies that, due to the treatment effect help, the 80th-percentile relatively old girl would get into this degree (12.24 + 0.75=12.99) and the relatively young (12.24) would not [10]. Again, these findings are very relevant for the academic future of female students, enhancing the opportunities of relatively old girls and deteriorating the opportunities of relatively young girls.

[10]The information about the minimum entry score can be consulted here

Table 6 : Simultaneous Quantile Regression (SQR) estimations for entry score

	(1) 20th Quintile	(2) 40th Quintile	(3) 60th Quintile	(4) 80th Quintile
ALL				
oldest	0.190	-0.053	0.098	0.141
	(0.207)	(0.242)	(0.192)	(0.273)
runDay	0.004	0.017	0.019**	0.016
	(0.011)	(0.012)	(0.010)	(0.014)
runDay2	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
yofbirth	0.677***	0.855***	0.907***	0.760***
	(0.028)	(0.041)	(0.028)	(0.068)
oldrunDay	-0.009	-0.030	-0.038**	-0.032
	(0.020)	(0.023)	(0.019)	(0.027)
Constant	-1,341.920***	-1,694.189***	-1,796.114***	-1,502.853***
	(56.652)	(80.743)	(56.637)	(135.542)
Observations	2,394	2,394	2,394	2,394
GIRLS				
oldest	0.600	0.679*	0.533**	0.575*
	(0.396)	(0.379)	(0.244)	(0.302)
runDay	-0.001	0.013	-0.000	-0.015
	(0.023)	(0.027)	(0.018)	(0.021)
runDay2	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
yofbirth	0.729***	0.863***	0.883***	0.577***
	(0.051)	(0.058)	(0.037)	(0.090)
oldrunDay	-0.009	-0.048	-0.012	0.014
	(0.048)	(0.054)	(0.033)	(0.041)
Constant	-1,445.779***	-1,709.940***	-1,749.630***	-1,137.909***
	(101.182)	(115.570)	(73.584)	(179.125)
Observations	985	985	985	985
BOYS				
oldest	-0.162	-0.292	-0.531	-0.495
	(0.281)	(0.295)	(0.469)	(0.335)
runDay	0.017	0.006	0.032*	0.033*
	(0.012)	(0.014)	(0.017)	(0.019)
runDay2	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
yofbirth	0.643***	0.828***	0.893***	0.882***
	(0.032)	(0.045)	(0.063)	(0.057)
oldrunDay	-0.028	-0.004	-0.050	-0.054
	(0.023)	(0.026)	(0.036)	(0.035)
Constant	-1,272.976***	-1,640.499***	-1,768.053***	-1,746.446***
	(64.637)	(89.536)	(124.691)	(114.394)
Observations	1,102	1,102	1,102	1,102

[1] Standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

[2] This specific estimation has been carried out only with the observations within the non-parametric optimal bandwidth calculated in the former subsection.

[3] Control variables: *runDay* is the running variable, which is the distance in days from the 1st January cut-off and *runDay2* is just its squared version. *oldrunDay* which is a dumour variable that allows for different running variable coefficients to the left and to the right of the cut-off. *yofbirth* represents the year of birth and controls that we compare individuals born in the same year.

4.5 Is this effect still present beyond high school?

The previous three subsections have been focused on the effect of being relatively old within a cohort on the entry score. We have centered our attention on this educational stage because some of the literature discussed in Section 2 indicates that this is the last point in the academic life of a student in which this effect is still relevant. Beyond the secondary education it seems to disappear. This is usually attributed to the huge increase in the number of factors that play an important role in determining the grades at the university level (new social interactions, different institutional setting, types of examinations, etc.). Therefore, we am interested in checking whether or not the effects of the *virtuous circle* are still positive and significant once the students enter into the university. For this purpose, in this subsection we implement a non-parametric Sharp RD using the same running variable as before (distance in days from the 1st January cut-off) but using the university grades at the end of the first year as outcome variable [11]. This means that we test if the treatment effect holds in the immediate next academic step.

4.5.1 Non-parametric Sharp RD: university grades

Following the procedure described in subsection 4.1 and using the first-year university grades as the outcome variable we obtain the results shown in Table 7. The sample differs in terms of observations. In the former case we had one entry score for each student, which translates into one observation per individual (because a student cannot have two valid entry score). However, our rich administrative data provides us the grades of first-year modules for each student. Therefore, we allow each student to enter the estimation sample more than once. We let all the first-year module grades to enter in our non-parametric Sharp RD design, which appreciably multiplies the number of observations included in the estimation.

As we see in the table, the structure of the results is still the same: there is only a significant causal effect of being relatively old in your cohort only if you are a girl. Then, in contrast with much of the existing literature, the treatment effect is still present once the individuals finish their freshman year. Nevertheless, an important reduction in the size of

[11] We do not perform a parametric replication since we have shown that, taking the same optimal bandwidth, the results practically do not change

coefficients can be appreciated. In Table 3 we have seen that the treatment effect provokes an increase on the entry score of 0.75 (robust bias-corrected version) whereas for first-year university grades the treatment effect amounts to only 0.15. Therefore we conclude that, while it is true that we find significant treatment effects on the first-year university grades, it seems that the effect is gradually disappearing. This decline may be attributed to the stronger role that other variables exert on the university grades but this research question is not the one addressed in our paper.

Table 7 : Non-parametric Sharp RD estimation for first-year university grades

	(1) All	(2) Girls	(3) Boys
Conventional	0.016	0.147**	-0.171
	(0.065)	(0.071)	(0.115)
Bias-corrected	0.022	0.155**	-0.152
	(0.065)	(0.071)	(0.115)
Robust	0.022	0.155*	-0.152
	(0.079)	(0.086)	(0.139)
Observations	70,944	40,535	30,409

[1] Standard errors in parentheses. Statistical significance: ***$p < 0.01$, **$p < 0.05$, * $p < 0.1$.

4.5.2 Heterogeneity of results across the ability distribution: university grades

As in Section 4.4 we now focus our attention on the heterogeneity of results across the distribution of the first-year university grades. Using the same technique, Simultaneous quintile Regression (SQR) we want to capture the treatment effect at the 20th (very low ability), 40th (low ability), 60th (high ability) and 80th (very high ability) quantiles of the distribution. The results are presented in Table 8. Looking at the treatment variable (*oldest*) we see that the effects are accumulated in the upper part of the distribution (40th,60th and 80th), as in the case of the entry score. The surprising finding is that for the 60th percentile of the male group we find a negative significant effect. Given that we have never found significant effects on boys in any of the different specifications and techniques used, we attribute this last results to some randomly sub-group of relatively old boys performing particularly bad in their freshmen year. We also check whether the coefficients at different quantiles are significantly different or not using the same tests as in subsection 4.4.

Table 8 : Simultaneous Quintile Regression (SQR) estimations for first-year university grades

	(1) 20th Quintile	(2) 40th Quintile	(3) 60th Quintile	(4) 80th Quintile
ALL				
oldest	0.002	-0.000	-0.099	0.023
	(0.115)	(0.074)	(0.074)	(0.064)
runDay	-0.000	-0.000	0.011**	0.012***
	(0.007)	(0.005)	(0.005)	(0.004)
runDay2	0.000	-0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
yofbirth	0.260***	0.200***	0.200***	0.140***
	(0.019)	(0.010)	(0.010)	(0.010)
oldrunDay	-0.001	0.000	-0.020**	-0.025***
	(0.014)	(0.009)	(0.010)	(0.008)
Constant	-514.180***	-393.000***	-392.272***	-270.369***
	(38.786)	(20.548)	(19.577)	(19.471)
Observations	21,289	21,289	21,289	21,289
GIRLS				
oldest	0.210	0.232***	0.158**	0.121*
	(0.129)	(0.087)	(0.074)	(0.069)
runDay	-0.001	0.004	0.008**	0.001
	(0.005)	(0.004)	(0.004)	(0.003)
runDay2	0.000	0.000**	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
yofbirth	0.319***	0.243***	0.217***	0.150***
	(0.017)	(0.012)	(0.014)	(0.013)
oldrunDay	-0.003	-0.013*	-0.020***	-0.004
	(0.010)	(0.007)	(0.007)	(0.007)
Constant	-631.538***	-478.414***	-425.997***	-291.528***
	(34.111)	(23.363)	(26.990)	(25.599)
Observations	16,503	16,503	16,503	16,503
BOYS				
oldest	-0.025	0.084	-0.261**	0.047
	(0.198)	(0.131)	(0.130)	(0.146)
runDay	0.000	-0.016	-0.000	0.012
	(0.017)	(0.013)	(0.012)	(0.013)
runDay2	0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
yofbirth	0.275***	0.152***	0.175***	0.098***
	(0.028)	(0.029)	(0.021)	(0.024)
oldrunDay	-0.009	0.016	-0.004	-0.036
	(0.033)	(0.024)	(0.023)	(0.026)
Constant	-544.350***	-298.458***	-342.125***	-187.889***
	(55.192)	(57.435)	(40.950)	(48.534)
Observations	6,878	6,878	6,878	6,878

[1] Standard errors in parentheses. Statistical significance: ***$p < 0.01$, **$p < 0.05$, * $p < 0.1$.

[2] This specific estimation has been carried out only with the observations within the non-parametric optimal bandwidth calculated in the former subsection.

[3] Control variables: *runDay* is the running variable, which is the distance in days from the 1st January cut-off and *runDay2* is just its squared version. *oldrunDay* which is a dumour variable that allows for different running variable coefficients to the left and to the right of the cut-off. *yofbirth* represents the year of birth and controls that we compare individuals born in the same year.

5 Conclusions

In this paper we show that the effect on academic performance of being among the oldest rather than the youngest within a cohort of students may differ by (a) gender and (b) educational stage. In our sample, we find that being relatively old has a positive impact on the entry score and university grades for female students but not for their male peers.

We argue that the mechanisms that explain this *oldest premium* are the cognitive advantages experimented by the oldest during childhood development and their long lasting consequences on valuable personality traits for academic success, which are reinforced by the feedback and considerations of professors, family and friends. In addition, we think that a more sensible response of women to the positive effects of the *virtuous circle* may explain why we only get significant results for the female subsample. This explanation is based on an important line of psychological research that shows a greater response of women to their environment's influence.

Furthermore, our results indicate that the positive effect is still quite significant at the very end of the secondary school (+0.75 points on the entry score) but much less relevant when examined for first-year university examinations (+0.15 points on university grades). The reason behind this sudden attenuation may be the increase in the number of elements that shape a college student academic performance, as the new institutional settings, other type of examinations, colleagues, etc. Nevertheless, according to previous empirical studies, we would have expected a smaller size of the effects as it is argued that the relevance of this phenomenon dissipates beyond secondary school.

Finally, we observe that the effects are significantly identified only beyond the lower quantiles of the distribution of the entry score, thus suggesting that students under a minimum of academic ability are probably less able to benefit from the month-of-birth effect. This result holds in our female sample for both educational stages.

Our results add a different perspective to the month-of-birth effect literature, since almost no one have paid attention the three differences that we have addressed. More specifically, we believe that the concentration of our results only on female students is the most important contribution to this line of research. As in many other social phenomena, a special focus should be paid to the gender differences in the way individuals perceive, internalize and react to the interactions and influences they face, which without any doubt would improve our understanding about gender inequalities.

Even if unintended, this is an unfair situation for the youngest students, which have a handicap since early childhood due to an arbitrary cut-off set by the education authorities. In fact, this oldest-youngest inequality is one of the issues that should be solved by these authorities to improve the equity of the system. One solution may be to re-order the composition of academic cohorts by gathering individuals born in the same semester instead of the same year, which would lead to fewer age differences within a cohort and therefore smaller cognitive advantage of the relatively oldest.

References

[1] Ando, S., Usami, S., Matsubayashi, T., Ueda, M., Koike, S., Yamasaki, S., and Kasai, K. (2019). "Age relative to school class peers and emotional well-being in 10-year-olds," *PloS one*, 14(3), e0214359.

[2] Angrist, J. D. and Pischke, J. S. (2014). "Mastering' metrics: The path from cause to effect". *Princeton University Press.*

[3] Beneito,P. and Rosell, Ines. (2019). "Gender Responses to Selective Settings: A Small Fish in a Big Pond" *Discussion Papers in Economic Behaviour 0518*, University of Valencia, ERI-CES.

[4] Berlin, N., and Dargnies, M. P. (2016). "Gender differences in reactions to feedback and willingness to compete," *Journal of Economic Behavior and Organization*, 130, 320-336.

[5] Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014a). "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82(6): 2295-2326.

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

[7] Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). "Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs," *arXiv preprint* arXiv:1809.00236.

[8] Cattaneo, M. D. (2015). "Robust inference in regression-discontinuity designs," *Stata Conference 2015* Columbus, United States.

[9] Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2017). "A practical introduction to regression discontinuity designs," *Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I.*

[10] Cattaneo, M. D., Jansson, M., and Ma, X. (2019). "Simple local polynomial density estimators," *Journal of the American Statistical Association*, (just-accepted), 1-11

[11] Crawford, C., Dearden, L., and Meghir, C. (2010). "When you are born matters: the impact of date of birth on educational outcomes in England," *IFS working papers*, (No. 10, 06).

[12] Crawford, C., Dearden, L., and Greaves, E. (2011). "Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills in England," *IFS working papers*.

[13] Crawford, C., Dearden, L., and Greaves, E. (2014). "The drivers of month of birth differences in children's cognitive and non-cognitive skills," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(4), 829-860.

[14] Grenet, J. (2009). "Academic performance, educational trajectories and the persistence of date of birth effects. Evidence from France," *Unpublished manuscript*.

[15] Hahn, J., Todd, P., and van der Klaauw, W. (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201209.

[16] Hanly, M., Edwards, B., Goldfeld, S., Craven, R. G., Mooney, J., Jorm, L., and Falster, K. (2019). "School starting age and child development in a state-wide, population-level cohort of children in their first year of school in New South Wales, Australia," *Early Childhood Research Quarterly*.

[17] Johnson, M., and Helgeson, V. S. (2002). "Sex differences in response to evaluative feedback: A field study," *Psychology of Women Quarterly*, 26(3), 242-251.

[18] Lima, G., Ruivo, M. M., Reis, A. B., Nunes, L. C., and do Carmo Seabra, M. (2019). "The impact of school starting age on academic achievement," *NOVA Working Paper*, (March 12, 2019).

[19] Mayo, M., Kakarika, M., Pastor, J. C., and Brutus, S. (2012). "Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in MBA teams," *Academy of Management Learning and Education*, 11(4), 631-652.

[20] McEwan, P. J., and Shapiro, J. S. (2008). "The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates," *Journal of Human Resources*, 43(1), 1-29.

[21] Page, L., Sarkar, D., and Silva-Goncalves, J. (2017). "The older the bolder: Does relative age among peers influence childrens preference for competition?," *Journal of Economic Psychology*, 63, 43-81.

[22] Page, L., Sarkar, D., and Silva-Goncalves,J. (2018). "Long-lasting effects of relative age at school," *QuBE Working Papers 056*, QUT Business School.

[23] Pedraja-Chaparro, F., Santn, D., and Simancas, R. (2015). "Determinants of grade retention in France and Spain: Does birth month matter?," *Journal of Policy Modeling*, 37(5), 820-834.

[24] Pellizzari, M., and Billari, F. C. (2012). "The younger, the better? Age-related differences in academic performance at university," *Journal of Population Economics*, 25(2), 697-739.

[25] Pea, P. A. (2017). "Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood," *Economics of Education Review*, 56, 152-176.

[26] Puhani, P. A., and Weber, A. M. (2008). "Does the early bird catch the worm?," *The economics of education and training*, (pp. 105-132), Physica-Verlag HD.

[27] Russell, R. J. H., and Startup, M. J. (1986)." Month of birth and academic achievement," *Personality and Individual Differences*, 7(6), 839-846.

[28] Schwalbe, M. L., and Staples, C. L. (1991). "Gender differences in sources of self-esteem," *Social Psychology Quarterly*, 54(2), 158-168.

[29] Smith, J. (2009). "Can regression discontinuity help answer an age-old question in education?: the effect of age on elementary and secondary school achievement," *J. Econ. Anal. Poly*, 9, 130.