Parental Love Is Not Blind: Identifying Selection into Early School Start

Ainoa Aparicio Fenoll, Nadia Campaniello, and Ignacio Monzón*

May 31, 2019

Abstract

In many countries, parents can choose whether their children start elementary school one year early. Like many other educational decisions, this choice has lasting consequences that depend on each child's characteristics. Which children are sent to school early? We propose a novel methodology to identify the sign and strength of selection into early starting. We exploit a feature of the Italian education system: only children born between January and April can start elementary school one year early. We use data on standardized tests taken by all students in Italy. We find robust evidence of positive selection: early starters would have obtained scores 0.2 standard deviations higher than the average student, had they started regularly. Additionally, we use this methodology to compare the effect of early starting on selected and average students. We find that the penalty associated with early start is lower for selected students born in March and April.

*Aparicio Fenoll: University of Turin (ESOMAS Department) and Collegio Carlo Alberto, Corso Unione Sovietica 218 Bis, 10134 Torino, Italy (ainoa.apariciofenoll@unito.it); Campaniello: University of Essex and IZA, Wivenhoe Park, CO4 3SQ, Colchester, UK (ncampa@essex.ac.uk); Monzón: Collegio Carlo Alberto, Piazza Vincenzo Arbarello, 8, 10122 Torino, Italy (ignacio@carloalberto.org). We are grateful to Sule Alan, Philip Cook, Rajeev Dehejia, Emilia Del Bono, Adeline Delavande, Raquel Fernández, Eric Hanushek, Melanie Luhrmann, Giovanni Mastrobuoni, Matthias Parey, Daniele Paserman, Birgitta Rabe, Jeff Smith, Alex Tetenov and Sarid Weisburd for their useful comments. We thank seminars participants at Collegio Carlo Alberto, IMT Lucca, University of Padova, and University of Essex, as well as participants at the Alpine Population Conference, the 2018 Meetings of the SEHO, the Comparative Law and Economics Forum Meeting, the 2018 EALE Conference, the 2018 Petralia Applied Economics Workshop, the 2018 Workshop on Perceptions and Public Policies, the First Unibg Winter Symposium in Economics, the Third Bolzano Applied Microeconomics Workshop.

1 Introduction

Parents make many decisions regarding the education of their children. They decide whether to enroll children on early formal child care, which school they attend, whether they participate in extracurricular activities, and also school starting age. These are crucial choices, with effects that may depend on each child's specific characteristics. Then, it is important to understand how parents choose. In particular, how do parents take into account their children's characteristics when making these decisions?

We focus on how parents choose the school starting age. In several countries parents can choose that their children start elementary school one year early. The literature on early child development shows that this decision has lasting consequences. On the one side, early starters enter the labor market one year early, which increases their returns from human capital (see for example Black, Devereux, and Salvanes [2011]). Moreover, children from disadvantaged backgrounds could benefit from attending school rather than staying at home. On the other side, early starters have worse academic performance (see for example Bedard and Dhuey [2006]). The relative importance of these factors and hence the effect of starting early can depend on each child's characteristics.

The main difficulty in understanding which students are selected into early start is that this decision itself affects students' outcomes. To see this, consider children who differ in their underlying unobservable ability. We would like to know whether high ability children are selected into early start. Everything else equal, a higher ability student has higher test scores. However, age also affects scores: an extra month of age leads to higher test scores. In order to compare the underlying ability of those who start one year early relative to those who do not, a naïve approach would be to compare the test scores of these two groups. Unfortunately, the difference in scores has two sources: 1) the difference in ability and 2) the difference in age. Students who start one year early are effectively twelve months younger than regular students born in the same month. Being twelve months younger has a strong negative effect on test scores. Therefore, we could mistakenly conclude that early starters are low ability students.

Our strategy to identify the characteristics of students who start early is based on

(i) a feature of the Italian education system and (ii) an empirical regularity on how age affects test scores. First, children in Italy can start elementary school at age five instead of age six *only if they are born between January and April*. Children born between May and December must start school at age six. Thus, within the same class, children can be born up to sixteen months apart. Second, there is a pattern on how age affects test scores: the increase in average test score from one extra month of age is constant. Fredriksson and Öckert [2014] show this empirical regularity for Sweden, while Crawford, Dearden, and Meghir [2010] do it for England, Black, Devereux, and Salvanes [2011] for Norway and Cook and Kang [2018] for the US. We document this also for our Italian dataset: the effect of age-in-months on average test scores is linear for children born between May and December.

Our methodology to identify which students start school early has three steps. First, we use the subsample of children born from May to December to estimate age-in-months effects on test scores. Children born in these months must start school at the age of six. Hence, differences in average test scores across months of birth are exclusively due to differences in age when taking the test. Second, we use the estimated age-in-months effects to compute the average test scores for children born between January and April, had *all* students started regularly. In practice, we extrapolate the linear trend found for children born between May and December to those born from January to April. Finally, we compute the average test scores that early starters would have obtained, had they started regularly. To do so, we only need test scores from regular starters and the proportion of early starters. If the average test scores that early starters would have obtained are higher than those of the average student in the population, we conclude that there is positive selection. Our measure of the strength of selection is the difference between the average test score of early starters, had they started regularly, and the average test score in the population, had all students started regularly.

We use data on standardized tests administered to all students in Italy. These tests cover two subjects (Italian and mathematics), are designed by an agency of the Italian Government (the National Institute for the Evaluation of the School System - INVALSI), and are mandatory for all students. Students are tested in second, fifth, eight and tenth grades of compulsory schooling. In our analysis we use information on test scores, month and year of birth, and grade for the academic years 2011–12 to 2016–17.

Our main result is that early starters are positively selected. Had they started regularly, early starters would have been at the top of the grade distribution of their cohort: they would have obtained (on average) scores 0.2 standard deviations higher than the average student. We find that this pattern of positive selection arises in all academic years, all months of birth (January to April) and in both subjects. Moreover, this result holds true even after controlling for a rich set of observable characteristics.

Our second main finding characterizes the penalty from early starting for selected and non-selected students. Early starting typically reduces test scores. For a given individual, the penalty from early starting is the magnitude of this decrease. Do selected children suffer a lower penalty than that of an average child in the population? We find that the penalty from early starting is lower for selected students born in March and April. However, selected students born in January or February suffer a penalty as large as that of an average student in the population.

Two considerations become necessary for the correct interpretation of these results. First, from a theoretical perspective, when they decide whether to send their children to school early, parents optimize some objective function, taking into account the information on their children characteristics that is available to them. The selection of children into early start is a result of this optimization problem. We observe neither the objective function of parents nor the information that they have. Instead, we use information on whether parents make their children start school one year earlier and children's test scores to learn how selection takes place. Second, the decision to send children to school early has several consequences including effects on test scores at all educational levels and labor market outcomes. In this paper we focus exclusively on test scores during compulsory education. Addressing whether the decision is optimal overall is beyond the scope of this paper.

1.1 Related Literature

A large literature studies the effect of relative age within the classroom. There is consensus that older students obtain higher scores. Bedard and Dhuey [2006] find that the oldest students score 4–12 percentiles higher than the youngest students in grade four; and 2–9 percentiles higher in grade eight. The evidence of the effects of relative age on labor market outcomes and on education attainment is mixed. Black, Devereux, and Salvanes [2011] find that older students have lower earnings at the beginning of their careers. This effect disappears by age 30. On the other side, Black, Devereux, and Salvanes do not find any effect of relative age on educational attainment. Instead, Dobkin and Ferreira [2010] find that younger individuals are more likely to complete more years of schooling. They do not find any effect on wage and employment.[1]

Recent papers focus on the effects of starting school either one year early or one year late. Ordine, Rose, and Sposato [2018] study whether students who have the option of starting early perform better than those who do not have that option, using the same data as we do. They find that those who take the option to early start perform better than the regular students who do not have the option (those born between May and December). Cook and Kang [2018] show that white males with lower academic abilities from high educated families are more likely to delay their entrance to school in North Carolina. They find that postponing school entrance decreases the male-female achievement gap by 11% for white students.

The main challenge for the identification of the effect of school starting age on outcomes is that school starting age may be endogenous. Parents may take into account their children's characteristics when making this decision. For instance, if parents perceive that their children are more mature and talented than the average child of their age, they may make their children start one year early. There are two main techniques to solve for en-

[1] There is a large literature on the effect of relative age on these outcomes. For additional evidence on the effect of relative age on test scores see Dobkin and Ferreira [2010], Strøm [2004], Fredriksson and Öckert [2014], Crawford, Dearden, and Meghir [2010], Black, Devereux, and Salvanes [2011], McEwan and Shapiro [2008], and Cook and Kang [2018]. For additional evidence on the effect of relative age on labor market outcomes and educational attainment see Dhuey and Lipscomb [2008], Bedard and Dhuey [2006], Puhani and Weber [2008], Mühlenweg and Puhani [2010], Pellizzari and Billari [2012], Zweimüller [2013], Fredriksson and Öckert [2014], and Ponzo and Scoppa [2014].

dogeneity. First, a common identification strategy is to instrument the actual starting age with the age at which children would have started school if they had followed the standard entry rule. Second, some papers use regression discontinuity designs that exploit school entry cutoff dates and compare the performance of children born at both sides of the cutoff.

Papers that estimate the effect of early school start do not focus on uncovering the pattern of selection. However, one could gain some insight about the sign of selection by comparing naïve OLS estimations (affected by selection) to other estimators which are unaffected by selection (i.e. IV or RDD estimates). This approach works when the bias of the OLS estimations can be attributed entirely to selection.[2] An IV estimate that exceeds the OLS estimate indicates the presence of negative selection into early start. However, IV estimates inform only about the sub-sample of the population of students whose school entry age behavior is affected by the rule (the "compliers"). Similarly, RDD estimates refer only to the subpopulation of students born close to the cutoff date. As a result, the comparison of OLS and IV/RDD estimates is not informative of selection in the population as a whole. What is more, this comparison provides evidence only about the direction of selection. In contrast, we measure the sign and strength of selection for the whole population and for different months of birth.

Our paper also relates to the literature about parental's perception of their children's type. Two recent papers focus on whether parents are informed about the academic performance of their children. Kinsler and Pavan [2016] use a US longitudinal survey. They find that parents beliefs about their child's skills relative to children of the same age are determined by their child's skills relative to children of the same school. They also find a positive relationship between perceived children's abilities and parents' investment in human capital (in particular they focus on remedial education). Dizon-Ross [Forthcoming] analyzes data from a field experiment in Malawi. She provides evidence that parents, especially the poorer and less educated, have distorted beliefs about their children's performance at school. In turn, these inaccurate perceptions prevent parents from mak-

[2]This assumption is appealing in our setup as reverse causality and measurement error are unlikely to affect OLS estimates.

ing the optimal investment in human capital. Once provided with the right information about their children performance at school, they update their beliefs and they invest more efficiently in their education.

The remainder of this paper is organized as follows. We present the data and institutional background in Section 2. In Section 3 we describe our methodology and in Section 4 we present our results. Section 5 discusses several applications and robustness checks. We conclude in Section 6.

2 Data and Institutional Framework

We use standardized test score data from the National Institute for the Evaluation of the School System (INVALSI). Education is compulsory in Italy between ages 6 and 16. The education system is divided in elementary school (five years), middle school (three years) and secondary school (five years).[3] Students take standardized tests in the second and fifth year of elementary school, then three years later in the third year of middle school and finally two years later in the second year of secondary school. INVALSI provides data from academic years 2009–10 to 2016–17.

The INVALSI data contains test scores from two subjects (Italian and mathematics) and indicates the number of correct answers. We standardize scores by subject, academic year, and grade to have zero mean and unit variance (as in Angrist, Battistin, and Vuri [2017]). The data set also includes students' characteristics (among them: gender and whether they are foreign born) and parental characteristics (among them: whether they are foreign born, their level of education, and labor market status).

We make a series of exclusions to arrive at the sample that we use for our analysis. First, information on students' month of birth is not available for academic years 2009–10 and 2010–11. Since this information is crucial to identify selection, we only include academic years 2011–12 to 2016–17 in our sample. Second, selection into early starting takes place right before the first year of elementary school. Since our objective is to identify selection, we focus on the second year of elementary school, the closest to this decision.

[3] We provide further institutional details in Appendix A.1.

Moreover, the effects of an extra month of age on scores are stronger in second grade than in later grades. Finally, for later grades early starters may appear as regular starters if they repeated a grade while grade repetition is extremely uncommon in second grade (See Section 5.1 for a discussion on using data from later grades).

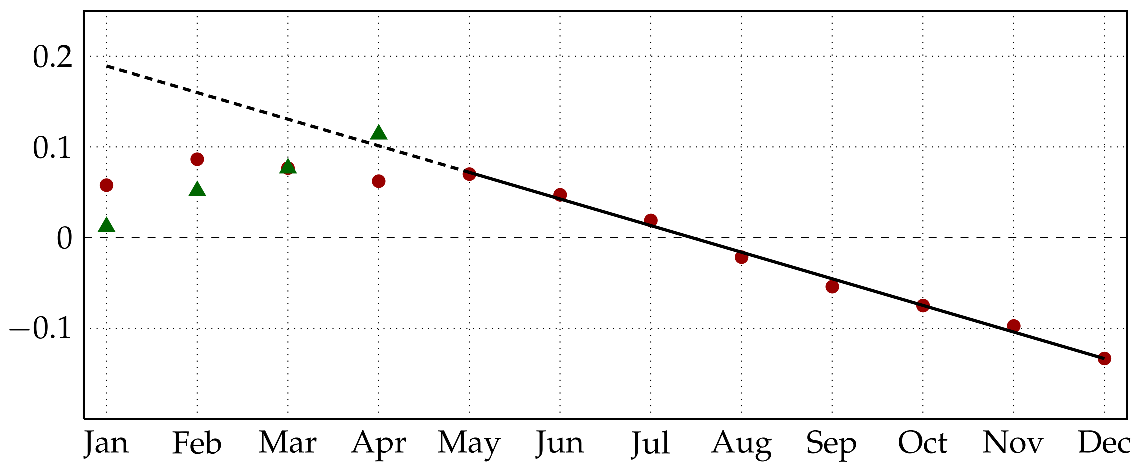
Next, we include in our sample only regular and early starters. We say a child is *regular* when he turns seven the year he starts grade 2. Instead, a child is an *early starter* if he satisfies two conditions: he turns six the year he starts grade 2 and he is born between January and April. We then exclude students from three groups. First, those who turn eight or more the year they start grade 2 (1.61% of total students in grade 2). Second, those who turn five or less the year they start grade 2 (less than 0.01% of total students in grade 2). Third, we also exclude students who turn six the year they start grade 2, but are born between May and December (0.40% of total students in grade 2).

The resulting data set includes $2,800,777$ observations for the Italian test and $2,829,095$ observations for the mathematics test. The average student answers correctly 62.3% and 61.1% of the questions in the Italian and mathematics tests, respectively. In our sample 31.8% of children are potential early starters, since they are born between January and April. Of those, only 26.2% start early, so 8.33% of the observations in our sample correspond to early starters. Of all students in grade 2 who are born in January, 43.4% of them are early starters. This proportion decreases to 28.3% for February, 19.2% for March, and finally 13.2% for April.

Early starters are more likely to be female (54% instead of 47% for regular starters born in the same months), less likely to be foreign-born (1.4% instead of 3.1%), and less likely to have foreign-born parents. They have a higher proportion of parents with university degrees (26% instead of 18% for mothers, and 21% instead of 14% for fathers). We present descriptive statistics of students and their parents in detail in Table 7 in Appendix A.4. In general, early starters have observable characteristics associated to better performance in tests. Our methodology measures the total selection, including both observable and unobservable characteristics. We discuss this point further in Section 5.2, where we adjust test scores by observed characteristics and provide strong evidence for positive selection also on unobservables.

Average test scores exhibit some common patterns for all academic years and for both subjects. To illustrate these patterns consider the test scores in Italian for the cohort 2016–17, as an example. Figure 1 presents average test scores by month both for regular and early starters. Circles (in red) represent average test scores for regular starters, while triangles (in green) represent average test scores for early starters. The thick line fits average test scores of regular starters born between May and December. Average test scores exhibit a linear decrease from May to December. However, average test scores for regular starters born between January and April lie below this linear trend. This underperformance of regular starters is preliminary evidence of the presence of positive selection into early start.

Figure 1: Average Test Scores of Regular and Early Starters. Italian. 2016–17



Notes: Circles (in red) represent average test scores for regular starters, while triangles (in green) represent average test scores for early starters. The thick line fits average test scores of regular starters born between May and December. The dashed line its extrapolation for January-April born students.

3 Methodology

Our first objective is to identify which students are selected into early starting: do higher ability students start early? In other words, had they started regularly, would selected students have obtained grades from the top of the distribution? Unfortunately, we do not observe their counterfactual scores. Moreover, the scores we do observe from these students include a strong age effect: selected students are effectively 12 months younger than non-selected students born in the same month because they start early.

We focus on counterfactual average scores, and account explicitly for the age effect. To do so, we express test scores $T^t(m, x)$ as a function of m (age-in-months) and x (any other individual characteristics that determine scores). The superscript t indexes academic years. A student with characteristics (m, x) who starts regularly obtains scores $T^t(m, x)$, while one who starts early obtains scores $T^{t-1}(m - 12, x)$.

Students may belong to one of three groups $G \in \{S, NS, U\}$, where S denotes students who are *selected* into early starting, NS denotes students who are not selected into early starting (that is, they start regularly) and $U = S \cup NS$ denotes all students. In order to construct our counterfactual scores, we consider two possible treatments $T \in \{E, R\}$, where E denotes early starting, and R denotes starting regularly. Then, expected average scores for different groups are given by:[4]

$$A^t(G, T, m) = \begin{cases} E\left[T^t(m, x_i) \mid i \in G, m\right] & \text{if } T = R \\ E\left[T^{t-1}(m - 12, x_i) \mid i \in G, m\right] & \text{if } T = E \end{cases}$$

The strength of selection is given by $A(S, R, m) - A(U, R, m)$: the difference between the average test score of early starters, had they started regularly, and the average test score in the population, had all students started regularly. Although we do not observe these magnitudes, we can indirectly infer them.

Our methodology relies on a key identifying assumption, that average test scores in the population are linear in age-in-months: $A(U, R, m) = \alpha + \beta m$. As discussed in the introduction, there is evidence for many countries with different school starting age cutoffs that this is the case. We provide further evidence that average scores are linear in age-in-months also in our data set in Appendix A.2. Our methodology follows three steps.

Estimating age-in-months effects. In our first step we estimate the linear age-in-months effect on test scores on the subsample of regular students born between May

[4]For notational simplicity we drop the superscript t from $A^t(G, T, m)$. In what follows we denote average scores as $A(G, T, m)$.

and December using the following equation:

$$T_i^{st} = \alpha^{st} + \beta^{st} m_i^t + \varepsilon_i^{st} \qquad \forall\, s, t, \text{ and for } m \in [5, 12] \tag{1}$$

where T is the standardized test score of student i in subject s (Italian or mathematics) in year t (from academic year 2011–12 to 2016–17) and m stands for age-in-months. Our coefficient of interest β measures the effect of an extra month of age on test scores. We estimate equation (1) separately for each subject s and for each academic year t.

Predicting average test scores. In our second step we compute the predicted average test scores of students born between January and April, had all students started regularly. We use the estimated coefficients $\hat{\alpha}$ and $\hat{\beta}$ from equation (1) to compute $\widehat{A}(U, R, m) = \hat{\alpha} + \hat{\beta} m$.

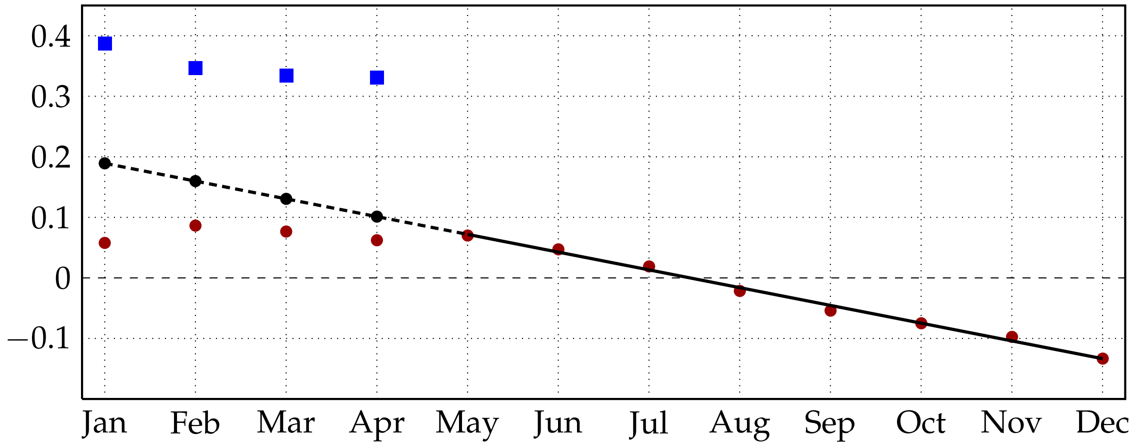
Figure 2 illustrates our methodology. This figure presents again average test scores in Italian for the cohort 2016–17. In our first step, we estimate equation (1) and obtain the thick black line in Figure 2. This line fits average test scores for regular starters born between May and December. In our second step we extrapolate this linear trend to the months between January and April. In this way we obtain the predicted average test scores $\widehat{A}(U, R, m)$, had all students started regularly. These are shown with black circles in Figure 2.

Calculating counterfactual test scores for early starters. In our third step we calculate the scores that early starters would have obtained, if they had not been selected into early starting. Our methodology allows for the indirect identification of $A(S, R, m)$. The average test score $A(U, R, m)$ of all students, had all of them started regularly, is a weighted average of the scores of those selected, and those not selected:

$$A(U, R, m) = P_S(m) A(S, R, m) + [1 - P_S(m)]\, A(NS, R, m).$$

where $P_S(m)$ denotes the proportion of students selected into early starting. We observe both $P_S(m)$ and $A(NS, R, m)$ in our sample. We compute $\widehat{A}(U, R, m)$ in our second step. Then, the predicted average test score $\widehat{A}(S, R, m)$ of early starters born in m can be easily

Figure 2: Estimated Average Test Scores of Early Starters, Had They Started Regularly. Italian. 2016–17



Notes: Circles (in red) represent average test scores for regular starters. The thick line, estimated from equation (1), fits average test scores of regular starters born between May and December. Circles over the fitted line (in black) show predicted average test scores $\widehat{A}(U, R, m)$. Squares (in blue) represent the average test scores that early starters would have obtained had they started regularly, $\widehat{A}(S, R, m)$, as computed from equation (2).

expressed as

$$\widehat{A}(S, R, m) = (P_S(m))^{-1} \left[\widehat{A}(U, R, m) - (1 - P_S(m)) A(NS, R, m)\right]. \quad (2)$$

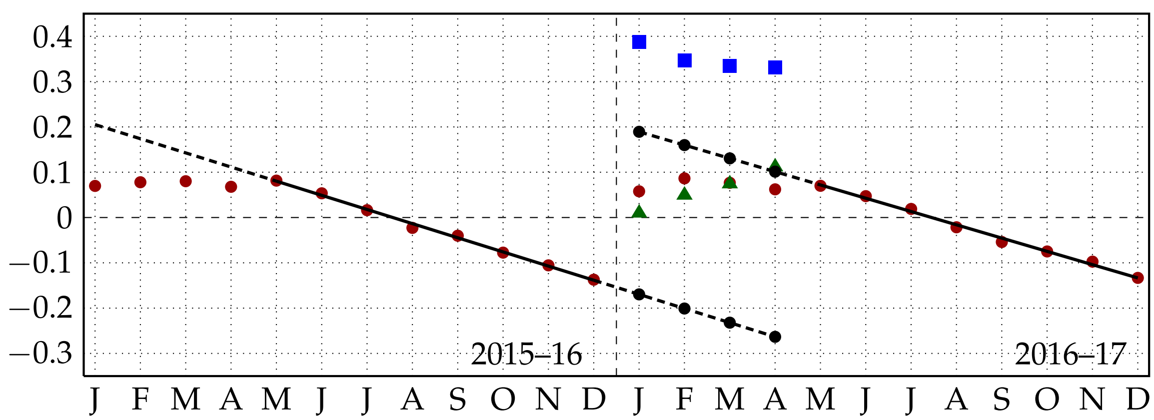
The blue squares in Figure 2 represent the predicted average test score $\widehat{A}(S, R, m)$ of early starters, as computed from equation (2).

Our measure of the strength of selection is $\widehat{A}(S, R, m) - \widehat{A}(U, R, m)$: the difference between the (predicted) average test score of early starters, had they started regularly, and the (predicted) average test score in the population, had all students started regularly. This measure is the vertical distance between the squares in blue and the circles in black in Figure 2. We compute the standard errors associated to this difference using bootstrap at the school level.

We next focus on the penalty, in terms of test scores, from starting school early. The penalty from starting early is given by $A(S, E, m) - A(S, R, m)$ for those selected into early starting. Instead, for an average student in the population, the penalty is given by $A(U, E, m) - A(U, R, m)$. We study whether the penalty for selected students differs from the penalty that an average student in the population would suffer.

Figure 3 illustrates our methodology to measure the penalty from starting early. This figure presents average test scores in Italian for two successive cohorts. Red circles represent average test scores for regular starters. Those in the left panel correspond to the academic year 2015–16, while those in the right panel correspond to the academic year 2016–17. Green triangles represent the actual scores $A(S, E, m)$ that early starters obtain. We compare them to the average test scores $\widehat{A}(S, R, m)$ of early starters, had they started regularly. As in Figure 2, these are shown with blue squares. The vertical difference between the green triangles and the blue squares represents the penalty for early starters. Instead, the effect on scores for the average student in the population is given by the difference between the (estimated) average test score $\widehat{A}(U, E, m)$, had all students started early, and the (estimated) average test score $\widehat{A}(U, R, m)$ had all students started regularly. This difference is represented by the vertical distance between black circles in Figure 3.

Figure 3: Early Starters: Actual Scores vs. Estimated Scores Had They Started Regularly. Italian. 2016–2017



Notes: The left panel shows test scores obtained by children who became 7 years old (the regular age of enrollment in 2nd grade) in 2015, i.e. children born in 2008. The right panel shows test scores obtained by children who became 7 years old in 2016, i.e. children born in 2009. Circles (in red) represent average test scores of regular starters. The thick line on the left panel fits average test scores of regular starters born between May and December 2008. The thick line on the right panel fits average test scores of regular starters born between May and December 2009. Squares (in blue) represent the average test scores that early starters born in 2009 would have obtained had they started regularly, as computed from equation (2). Triangles (in green) represent the actual average test scores of early starters born in 2009.

4 Results

We follow the three steps described in the previous section and illustrated in Figure 2. We first present our estimates for the effect of an extra month of age on test scores (our first step) from equation (1). Table 1 reports these estimates for each subject s and for each academic year t. The estimated age-in-month effect on test scores ranges between -0.29 and -0.35. The linear equation (1) fits the average test scores for regular starters extremely well in all academic years and for both mathematics and Italian.[5]

Table 1: Estimates of $\hat{\beta}$ (Linear Age-in-Month Effect on Test Scores)

	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
Mathematics	-0.030	-0.034	-0.033	-0.034	-0.031	-0.035
Italian	-0.031	-0.033	-0.031	-0.035	-0.031	-0.029

Notes: This table presents the estimates of the linear age-in-month effects $\hat{\beta}$ for all academic years, for Italian and mathematics. All shown estimates are statistically significant: p-values < 0.001. Standard errors are computed using bootstrap at the school level.

In our second step we directly compute the predicted average test score $\widehat{A}(U, R, m)$ in the population, had all students started regularly (the black circles in Figure 2). Finally, in the third step, we use equation (2) to compute the predicted average test score $\widehat{A}(S, R, m)$ of early starters, had they started regularly (the blue squares in Figure 2). Figure 2 shows that the difference $\widehat{A}(S, R, m) - \widehat{A}(U, R, m)$ is positive for test scores in Italian in the academic year 2016–17. This is evidence that early starters are positively selected for all months (January to April).

We show that there is positive selection in all academic years, for all months. Table 2 presents the strength of selection for all months (January to April), for all cohorts (2011–12 to 2016–17), and for both subjects. Estimates are positive and significant in all cases. The estimated strength of selection ranges between 0.098 and 0.318.

We next present our results on the penalty from starting early. Our estimate for the penalty for an early starter is given by the difference between $A(S, E, m)$ (the green triangles in Figure 3) and $\widehat{A}(S, R, m)$ (the blue squares in Figure 3). We represent this difference

[5] See Appendix A.2 for a thorough discussion on linearity.

Table 2: Estimates of the Strength of Selection

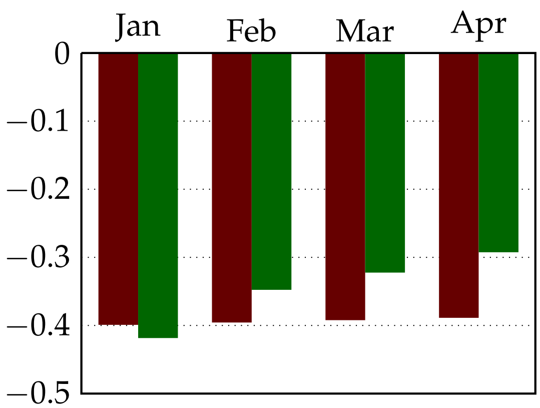
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
(A) Mathematics Scores						
January	0.187	0.178	0.179	0.191	0.209	0.204
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.24	0.281	0.172	0.215	0.259	0.207
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.213	0.21	0.138	0.18	0.238	0.212
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.156	0.246	0.098	0.191	0.258	0.226
	(0.003)	(0.000)	(0.017)	(0.000)	(0.000)	(0.000)
(B) Italian Scores						
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.156	0.172	0.168	0.209	0.191	0.198
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.168	0.287	0.178	0.252	0.243	0.187
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.151	0.271	0.149	0.224	0.232	0.204
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.136	0.318	0.121	0.248	0.267	0.23
	(0.005)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)

Notes: The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level. p-values are reported in parentheses.

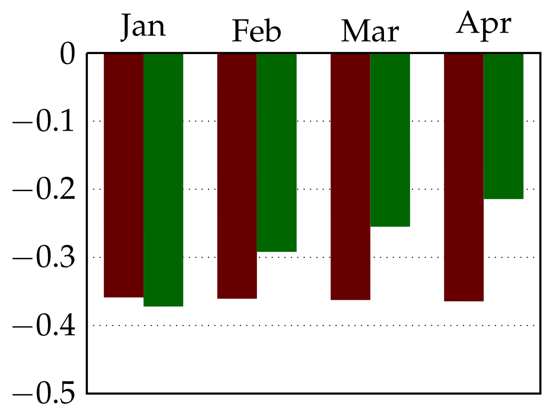
with green bars in Figure 4. Instead, our estimate for the penalty for an average student is given by the difference between $\widehat{A}(U,E,m)$ (the black circles over the left line) and $\widehat{A}(U,R,m)$ (the black circles over the right line in Figure 3). We represent this difference with red bars in Figure 4.

Table 3 presents our estimates for the differences in the effect of early starting for selected students and the average student in the population. The estimates are given by $\left[A(S,E,m)-\widehat{A}(S,R,m)\right]-\left[\widehat{A}(U,E,m)-\widehat{A}(U,R,m)\right]$. A positive difference implies that the penalty from early starting is lower for selected students. As Table 3 shows, the results are mixed for children born in January and February. There is no conclusive evidence that selected students from those months are affected by starting early differently

Figure 4: The Effect of Starting Early for Early Starters vs. the Average Student. 2016–2017



(a) Mathematics



(b) Italian

Notes: Green bars represent differences between the actual average test scores obtained by early starters born in 2009 (represented by the green triangles in Figure 3) and their (estimated) average test scores had they started earlier (blue squares in Figure 3). Red bars represent differences between the (estimated) average test scores in the population had all students started early (the black dots over the fitted line on the left panel in Figure 3) and the (estimated) average test scores in the population had all students started regularly (the black dots over the fitted line on the right panel in Figure 3).

than a random student from the population. In contrast, estimates for students born in March and April are positive and significant in 16 out of 20 cases, while non-significant in the other 4 cases. This shows that the penalty from early starting is significantly lower for selected students born in March and April who are the youngest ones.

5 Applications and Robustness Checks

5.1 Information from Other Grades

The INVALSI data contains information on grades two, five, eight and ten. We use data from grade two for several reasons. First, grade two is the closest to the decision of early starting. Second, the effect of age-in-months on scores is the largest for grade two: one extra month increases average test scores by approximately 0.032 for grade two, 0.022 for grade five, 0.015 for grade eight, and 0.008 for grade ten (See Table 4 in Appendix A.3 for all estimates).

Table 3: Estimates of the Difference in the Penalty from Early Starting for Selected and Average Students

	(A) Mathematics Scores				
	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.034**	−0.012	−0.001	−0.021	0.023
	(0.017)	(0.401)	(0.927)	(0.224)	(0.169)
February	0.005	0.091***	0.042	0.001	0.084***
	(0.842)	(0.000)	(0.119)	(0.984)	(0.001)
March	0.168***	0.209***	0.127***	0.094***	0.143***
	(0.000)	(0.000)	(0.001)	(0.006)	(0.000)
April	0.171***	0.294***	0.18***	0.119**	0.154***
	(0.005)	(0.000)	(0.001)	(0.023)	(0.001)
	(B) Italian Scores				
	2012–13	2013–14	2014–15	2015–16	2016–17
January	−0.028**	−0.031**	−0.068***	−0.017	−0.015
	(0.045)	(0.022)	(0.000)	(0.300)	(0.374)
February	−0.099***	0.053**	−0.052**	−0.008	0.039**
	(0.000)	(0.026)	(0.041)	(0.728)	(0.096)
March	0.032	0.147***	0.023	0.067**	0.066**
	(0.427)	(0.000)	(0.557)	(0.037)	(0.045)
April	−0.005	0.211***	0.061	0.094**	0.072**
	(0.935)	(0.000)	(0.224)	(0.037)	(0.083)

Notes: This table presents our estimates for the differences in the effect of early starting for selected students and the average student in the population. A positive difference implies that the penalty from early starting is lower for selected students. Standard errors are computed using bootstrap at the school level. p-values are reported in parentheses.

We introduce next a robustness exercise: we show the results of our methodology but using data from grade five instead of grade two. We do this because absent data on grade two, the best available data would be that of grade five. We also find strong evidence of positive selection. The estimated strength of selection ranges between 0.1 and 0.273 (while it is 0.098 and 0.318 using data from grade two). Table 5 in Appendix A.3 presents all estimates. Our second result compares the penalty from early starting for selected students and a randomly chosen student in the population. Using the data from grade five, we find that the penalty from early starting is lower for selected students born in March and April. Our estimates for the difference in penalty are positive for all cohorts. They are significant in 12 out of 20 cases. The results are mixed for children born in January and February. If anything, the penalty from starting early may be higher for selected students from January and February. To sum up our results using data from grade five are consistent with our findings from data on grade two. Table 6 in Appendix A.3 presents all estimates.

5.2 Selection on Unobservables

We find that early starters are positively selected in all cohorts and for all months of birth. As highlighted before, early starters have different observable characteristics than regular starters. For example, early starters are more often female and native. Moreover, their parents are more often native and more educated. Fathers of early starters are more often white-collar workers, and mothers are more often stay-at-home mothers. The observable characteristics of early starters are typically associated to better performance in tests. Then, is positive selection just a reflection of different observable characteristics?

We next study whether positive selection can be explained by observable characteristics. To do so, we first add observable characteristics as additional regressors to equation (1):

$$T_i^{st} = \alpha^{st} + \beta^{st} m_i^t + \gamma^{st} c_i^t + \varepsilon_i^{st} \qquad \forall\, s, t, \text{ and for } m \in [5, 12] \tag{3}$$

The vector c_i^t includes characteristics for student i in cohort t. These characteristics refer both to students and their parents. They include gender, whether the student or parents are foreign born, and education and labor market status of parents.[6] We estimate equation (3) using test scores for regular students born between May and December. We estimate this equation separately for each subject s and for each cohort t.

Next, we compute the test scores *adjusted* by observable characteristics: $\widehat{T}_i^{st} \equiv T_i^{st} - \widehat{\gamma}^{st} c_i^t$. The adjusted test score $\widehat{T}_i^{st}$ measures the part of the individual score not explained by observables. We follow the methodology described in Section 3 to estimate the strength of selection, using now adjusted test scores $\widehat{T}_i^{st}$, instead of T_i^{st}.

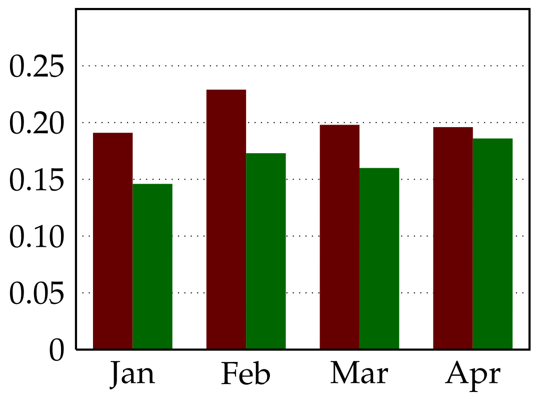
We find that there is positive selection in all academic years, even controlling for observable characteristics. We report all estimates of the strength of selection controlling

[6] We describe these variables in detail in Appendix A.4.

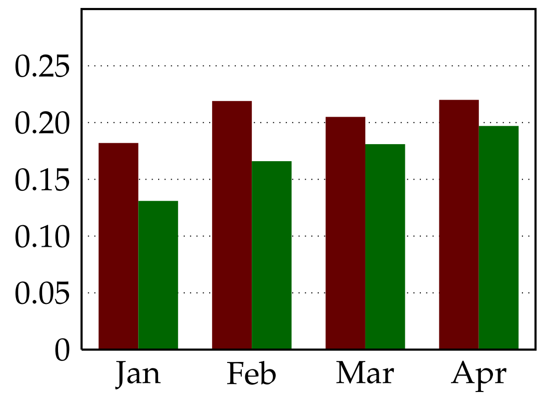
for observable characteristics in Table 9 in Appendix A.4. Our results provide strong evidence that unobservable characteristics drive positive selection into early starting.

Figure 5 compares the estimates of the strength of selection with and without controls. Each bar represents the strength of selection for a given month, averaged over all cohorts. Red bars show the strength of selection without controls, while green bars show the strength of selection with controls. The estimates of the strength of selection with controls are only slightly lower than those without controls. Observable characteristics only explain a small fraction of the strength of selection.[7]

Figure 5: Strength of Selection Controlling for Observable Characteristics



(a) Mathematics



(b) Italian

Notes: Red bars represent the strength of selection without controlling for observable characteristics. Green bars represent the strength of selection controlling for observable characteristics, that is, using adjusted test scores. Each bar reports the average strength of selection for a given month over all cohorts 2012–13 to 2016–17.

5.3 Regional Analysis

We now apply our methodology to identify the sign and the strength of selection at a disaggregated geographical level. Italy is divided into 20 regions, which are aggregated into three distinct macro regions (North, Center, and South). There is substantial heterogeneity in socioeconomic characteristics between these three macro regions. We observe significant differences in the proportion of early starters within Italy.

[7]We also compute the penalty from early starting using adjusted test scores. We report all estimates in Table 10 in Appendix A.4. The results are qualitatively similar to those reported in Table 3.

The map in Figure 6 shows the proportion of early starters in each Italian region for students born between January and April. While in the southern region of Campania 61.9% of students born between January and April start early, in the northern region of Valle d'Aosta it is only 4.6% of students born in those months who do so. In general, the proportion of early starters decreases as we move from South to North. There are differences in the proportion of early starters within each macro region. However, there is a clear pattern: in the North on average 10% of students born between January and April start early, while in the Center this proportion is 20% and in the South it is 51.8%.

We next present the pattern of selection by macro region.[8]

6 Discussion

There is a broad literature on policy evaluation. The objective in this literature is to measure the effect of a treatment (an intervention, a policy, or a program) on some outcomes of interest. Examples include the impact of job training on labor outcomes, the effect of retirement on individual well-being, and the impact of parental leave on children's outcomes. These papers are after an estimate of the effect for the average individual (or firm, schools, etc.) in the population. However, effects often depend on individual characteristics. Policy makers care about the effect of the treatment on those who actually get treated. They also care about who are the beneficiaries of the treatment. Whenever individuals can self-select into treatment, the estimate of the average treatment effect does not reflect these magnitudes. This makes the study of selection crucial in many setups.

In this paper we propose a new methodology to test for the presence of selection into treatment and measure the strength of selection. Our methodology can be applied to setups in which two requisites are satisfied. First, there is a well-known functional relationship between an exogenous running variable and the outcome of interest. Second, access to treatment depends exogenously on the running variable. Examples include: (i)

[8] We can also use our methodology at an even more disaggregated level and study selection by region. However, the number of observations decreases significantly for some regions. In Appendix **??** we apply our methodology to the three largest regions in each macro region: Lombardia (North), Lazio (Center), and Campania (South). The results are similar to those presented for the macro regions.

the impact of university scholarships awarded to students with a high-school grade over a certain threshold on the probability of university grade completion; (ii) how incentives to hire young individuals (those with ages lower than a certain threshold) affects labor outcomes; and (iii) the impact of retirement on health if retirement benefits are linked to age.

We apply our methodology to the case of school starting age in Italy. In this context, the functional relationship between students' age in months and test score is linear. Moreover, only students born in certain months can start school one year earlier. We find that early starters would have obtained scores 0.2 standard deviations higher than the average student, had they started regularly. Early start implies a penalty in scores for any student (since students who early start become effectively younger). However, we find that this effect is attenuated in practice because of selection. This corroborates that in the case of school starting age, learning about selection improves our understanding of the actual consequences of early school start.

References

ANGRIST, J. D., E. BATTISTIN, AND D. VURI (2017): "In a Small Moment: Class Size and Moral Hazard in the Italian Mezzogiorno," *American Economic Journal: Applied Economics*, 9(4), 216–49.

BEDARD, K., AND E. DHUEY (2006): "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *The Quarterly Journal of Economics*, 121(4), 1437–1472.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2011): "Too Young to Leave the Nest? The Effects of School Starting Age," *The Review of Economics and Statistics*, 93(2), 455–467.

COOK, P. J., AND S. KANG (2018): "The School-Entry-Age Rule Affects Redshirting Patterns and Resulting Disparities in Achievement," Discussion paper, National Bureau of Economic Research.

Crawford, C., L. Dearden, and C. Meghir (2010): "When You Are Born Matters: The Impact of Date of Birth on Educational Outcomes in England," Working Paper.

Dhuey, E., and S. Lipscomb (2008): "What makes a leader? Relative age and high school leadership," *Economics of Education Review*, 27(2), 173–183.

Dizon-Ross, R. (Forthcoming): "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments," *American Economic Review*.

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

Fredriksson, P., and B. Öckert (2014): "Life-cycle effects of age at school start," *The Economic Journal*, 124(579), 977–1004.

Gupta, A. (2018): "Nonparametric specification testing via the trinity of tests," *Journal of Econometrics*, 203(1), 169 – 185.

Kinsler, J., and R. Pavan (2016): "Parental Beliefs and Investment in Children: The Distortionary Impact of Schools," Working Paper.

McEwan, P. J., and J. S. Shapiro (2008): "The Benefits of Delayed Primary School Enrollment. Discontinuity Estimates Using Exact Birth Dates," *Journal of Human Resources*, 43(1), 1–29.

Mühlenweg, A. M., and P. A. Puhani (2010): "The Evolution of the School-Entry Age Effect in a School Tracking System," *Journal of Human Resources*, 45(2), 407–438.

Ordine, P., G. Rose, and D. Sposato (2018): "Parents know them better: the effect of optional early entry on pupil's schooling attainment," *Economic Inquiry*, 56(3), 1678–1705.

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

PONZO, M., AND V. SCOPPA (2014): "The long-lasting effects of school entry age: Evidence from Italian students," *Journal of Policy Modeling*, 36(3), 578–599.

PUHANI, P. A., AND A. M. WEBER (2008): "Does the Early Bird Catch the Worm?," in *The Economics of Education and Training*, pp. 105–132. Springer.

STRØM, B. (2004): "Student achievement and birthday effects," *Unpublished manuscript, Norwegian University of Science and Technology*.

ZWEIMÜLLER, M. (2013): "The effects of school entry laws on educational attainment and starting wages in an early tracking system," *Annals of Economics and Statistics*, pp. 141–169.

A Appendix

A.1 Institutional Background

The Italian education system is divided into elementary school (grades 1 to 5), middle school (grades 6 to 8) and secondary school (grades 9 to 13). Education is compulsory between the age of six (grade 1) and sixteen (grade 10). After middle school, students follow one of three tracks of secondary schooling (technical school, lyceum, or vocational school). The first two tracks lead to a high school diploma (*diploma di maturità*). Students with this diploma can then enroll in a university or other tertiary institutions.

The school year starts mid-September and finishes mid-June. The enrollment in elementary school is regulated by the Legislative Decree number 59, issued on February 2004. According to this law, children start elementary school the year they turn six. However, children born between January and April can start school one year in advance (the year they turn five). Public schools have to accept all those who opt for early starting, independently of their month of birth.

A.2 Linearity

Figure 1 shows a linear relationship between test scores and age in months for children born between May and December. In this section we present further evidence for this linear relationship in our data. We perform four tests of linearity: cross-validation, Akaike Information Criterion, Bayesian Information Criterion, and one following Gupta [2018]. The first three tests are methods used for model selection. We use them to compare two different models: 1. a linear function in age-in-months and 2. a function with a dummy variable for each month (from May to December).

Cross-validation tests the predictive ability of a model on a set of data not used in estimation. We randomly divide our sample into two sets, the training set (70% of our sample) and the testing set (30% of our sample). We use the parameters estimated in the training set to predict the dependent variable in the testing set. We find that the linear model better predicts the data in the testing set.

The Akaike Information Criterion and the Bayesian Information Criterion focus on the trade-off between the goodness of fit and the complexity of a model (in terms of number of parameters). These methods assign a score to each possible model. Then, when comparing two models, each criterion selects the one with the lowest score. The main difference between these two criteria is that the BIC imposes a larger penalty than the AIC to the number of parameters. We calculate the AIC/BIC scores for the two different models. Both criteria select the model with test scores as a liner function of age-in-months.

Gupta [2018] proposes a novel and simple methodology to test for linearity. In his approach, higher order polynomial terms are sequentially added to a linear specification. As these terms are added, the joint significance of the coefficients is tested. Following his methodology, we start from equation (1) and then we estimate up to a fifth order polynomial regression in month of birth. Results show that they are not significant at conventional levels indicating that we cannot reject the linear specification.

A.3 Information from Other Grades

Table 4 reports the estimates of equation (1) for each subject s and for each academic year t, for all grades. We next focus on data from grade five. Table 5 presents the strength of selection for all months (January to April), for all cohorts (2011–12 to 2016–17), and for both subjects. Estimates are significant and positive in all twenty-four cases but two (April 2014–15 and April 2016–17). Table 6 presents all estimates for the difference in penalty from early starting.

Table 4: Estimates of $\hat{\beta}$ (Linear Age-in-Month Effect on Test Scores). All Grades

	(A) - Mathematics Scores					
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
Grade 2	−0.030	−0.034	−0.033	−0.034	−0.031	−0.035
Grade 5	−0.019	−0.023	−0.023	−0.022	−0.021	−0.021
Grade 8	−0.008	−0.011	−0.013	−0.014	−0.012	−0.014
Grade 10	−0.005	−0.006	−0.006	−0.006	−0.008	−0.007
	(B) - Italian Scores					
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
Grade 2	−0.031	−0.033	−0.031	−0.035	−0.031	−0.029
Grade 5	−0.022	−0.023	−0.024	−0.022	−0.025	−0.023
Grade 8	−0.015	−0.017	−0.017	−0.018	−0.018	−0.020
Grade 10	−0.007	−0.009	−0.008	−0.009	−0.010	−0.010

Notes: This table presents the estimates of the linear age-in-month effects $\hat{\beta}$ for all academic years, for Italian and mathematics. Standard errors are computed using bootstrap at the school level. All shown estimates are statistically significant: p-values < 0.001.

A.4 Selection on Unobservables. Details

In Section 5.2 we identify the sign and strength of selection after controlling for observable characteristics. To do so, we include a vector of characteristics c in the regression of test scores on month of birth. This vector contains characteristics about the students and their parents. We include students' gender and whether they are foreign born. For each parent, we also include whether they are foreign born, their highest degree attained, and their labor market status.

Table 5: Estimates of the Strength of Selection. Data from Grade Five

(A) Mathematics Scores						
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.151	0.122	0.134	0.126	0.166	0.149
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.174	0.156	0.134	0.12	0.273	0.164
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.216	0.125	0.165	0.1	0.212	0.14
	(0.000)	(0.000)	(0.000)	(0.012)	(0.000)	(0.000)
April	0.217	0.179	0.142	−0.032	0.18	0.038
	(0.000)	(0.000)	(0.006)	(0.638)	(0.001)	(0.350)
(B) Italian Scores						
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.142	0.131	0.161	0.137	0.153	0.145
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.16	0.143	0.174	0.141	0.24	0.151
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.196	0.131	0.191	0.124	0.209	0.122
	(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)
April	0.163	0.198	0.206	0.083	0.185	0.057
	(0.001)	(0.000)	(0.000)	(0.156)	(0.001)	(0.169)

Notes: The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level. p-values are reported in parentheses.

Appendix A.1 describes the educational categories included in the variable highest degree attained. The possible labor market status of parents are: unemployed, stay-at-home, white-collar (*dirigente, docente universitario, funzionario, professionista dipendente, libero professionista, militare, insegnante, impiegato*), self-employed (*imprenditore o proprietario agricolo, lavoratore in proprio*), blue-collar (*operaio, addetto ai servizi o socio di cooperativa*), and retired. We only include observations that contain information for *all* these variables. These observations represent 70% of the original sample ($1,976,545$ observations for Mathematics and $1,960,580$ observations for Italian).

Table 7 provides descriptive statistics for the resulting sample. We describe separately the group of early starters—column (I), the group of regular starters born between Jan-

Table 6: Estimates of the Difference in the Penalty from Early Starting for Selected and Average Students. Data from Grade Five

	2012–13	2013–14	2014–15	2015–16	2016–17
(A) Mathematics Scores					
January	0.031**	−0.044***	−0.023	−0.044***	−0.005
	(0.020)	(0.000)	(0.155)	(0.001)	(0.727)
February	0.026	−0.011	0.015	−0.105***	0.092***
	(0.176)	(0.583)	(0.589)	(0.000)	(0.000)
March	0.143***	0.046	0.115**	0.041	0.154***
	(0.000)	(0.209)	(0.018)	(0.321)	(0.000)
April	0.141***	0.123**	0.305***	0.071	0.298***
	(0.005)	(0.034)	(0.000)	(0.212)	(0.000)
(B) Italian Scores					
	2012–13	2013–14	2014–15	2015–16	2016–17
January	−0.011	−0.097***	−0.032**	−0.029**	−0.033**
	(0.378)	(0.000)	(0.043)	(0.055)	(0.019)
February	−0.002	−0.088***	−0.004	−0.082***	0.056**
	(0.907)	(0.000)	(0.870)	(0.002)	(0.016)
March	0.089***	0.01	0.115***	0.057	0.12***
	(0.007)	(0.801)	(0.005)	(0.158)	(0.000)
April	0.054	0.019	0.213***	0.077	0.228***
	(0.297)	(0.729)	(0.001)	(0.181)	(0.000)

Notes: This table presents our estimates for the differences in the effect of early starting for selected students and the average student in the population. A positive difference implies that the penalty from early starting is lower for selected students. Standard errors are computed using bootstrap at the school level. p-values are reported in parentheses.

uary and April—column (II), and the group of regular starters born between May and December—column (IV). Column (III) pools observations form columns (I) and (II), so it describes all children born between January and April.

Table 8 reports the coefficients $\widehat{\gamma}$ from running the regression in equation (3) for the cohort 2016–17 and for both subjects.[9] The reference category for highest degree attained is elementary school. The reference category for labor market status is blue-collar.

Tables 7 and 8 show that characteristics observed more often on early starters typically display a positive correlation with test scores. Early starters are more often native-born,

[9] We also run the regression in equation (3) for all other cohorts. The results are similar, so we do not report them here.

Table 7: Observable Characteristics of Students and their Parents

Characteristics	January–April (I) Early	(II) Regular	(III) Total	May–December (IV) Regular
Male student	0.454	0.527	0.509	0.507
Foreign-born student	0.011	0.025	0.021	0.021
Foreign-born mother	0.099	0.138	0.128	0.132
Foreign-born father	0.078	0.114	0.105	0.109
Mother. Highest Degree Attained				
Elementary school	0.024	0.020	0.021	0.022
Middle School	0.236	0.261	0.254	0.258
Vocational school	0.045	0.091	0.080	0.078
High school	0.407	0.414	0.412	0.414
University	0.266	0.188	0.208	0.204
Other Tertiary Institution	0.022	0.026	0.025	0.025
Mother. Labor Market Status				
Unemployed	0.059	0.054	0.055	0.057
Stay-at-home parent	0.390	0.307	0.328	0.335
White-collar	0.385	0.395	0.392	0.385
Self-employed	0.090	0.094	0.093	0.092
Blue-collar	0.076	0.149	0.130	0.131
Retired	0.001	0.001	0.001	0.001
Father. Highest Degree Attained				
Elementary school	0.028	0.026	0.026	0.027
Middle School	0.296	0.350	0.336	0.339
Vocational school	0.054	0.101	0.089	0.089
High school	0.392	0.366	0.372	0.374
University	0.213	0.141	0.159	0.154
Other Tertiary Institution	0.018	0.017	0.017	0.017
Father. Labor Market Status				
Unemployed	0.068	0.048	0.053	0.054
Stay-at-home parent	0.005	0.006	0.005	0.005
White-collar	0.436	0.367	0.384	0.378
Self-employed	0.242	0.257	0.253	0.253
Blue-collar	0.245	0.317	0.299	0.304
Retired	0.005	0.006	0.005	0.005
Number of observations	158,749	465,220	623,969	1,336,611

Notes: This table shows the fraction of students with each characteristic. Column (I) describes the group of early starters. Column (II) describes the group of regular starters born between January and April. Column (IV) describes the group of regular starters born between May and December. Column (III) pools observations form columns (I) and (II), so it describes all children born between January and April.

Table 8: Estimates for the Coefficients $\widehat{\gamma}$ on Characteristics. Cohort 2016–17

	Mathematics	Italian
Male student	0.087^{***}	-0.070^{***}
Foreign-born student	-0.113^{***}	-0.095^{***}
Foreign-born mother	-0.187^{***}	-0.217^{***}
Foreign-born father	-0.133^{***}	-0.159^{***}
Mother. Highest Degree Attained		
Middle school	0.112^{***}	0.028^{*}
Vocational school	0.125^{***}	0.015
High school	0.294^{***}	0.179^{***}
University	0.412^{***}	0.328^{***}
Other Tertiary Institution	0.287^{***}	0.192^{***}
Mother. Labor Market Status		
Unemployed	0.061^{***}	0.093^{***}
Stay-at-home parent	0.105^{***}	0.116^{***}
White-collar	0.076^{***}	0.076^{***}
Self-employed	0.064^{***}	0.072^{***}
Retired	0.070	0.080
Father. Highest Degree Attained		
Middle school	0.104^{***}	0.090^{***}
Vocational school	0.131^{***}	0.088^{***}
High school	0.238^{***}	0.221^{***}
University	0.337^{***}	0.339^{***}
Other Tertiary Institution	0.215^{***}	0.217^{***}
Father. Labor Market Status		
Unemployed	0.044^{***}	0.046^{***}
Stay-at-home parent	0.093^{**}	0.051
White-collar	0.095^{***}	0.080^{***}
Self-employed	0.074^{***}	0.044^{***}
Retired	-0.015	0.039
Constant	-0.600^{***}	-0.398^{***}
Observations	$230,515$	$228,779$
R-squared	0.065	0.066

Notes: This table presents estimates for the coefficients on characteristics—$\widehat{\gamma}$ in equation (3). Standard errors are computed using bootstrap at the school level. The coefficients are marked with * if the level of significance is between 5% and 10%, ** if the level of significance is between 1% and 5% and *** if the level of significance is less than 1%.

and so are their parents. Parents are also more likely to have a university degree. Mothers of early starters are more often stay-at-home, while fathers are more often white-collar workers. Table 8 shows that all these characteristics are significantly and positively correlated with test scores. Finally, early starters are more likely to be female. Interestingly, this is correlated to higher test scores in Italian, but lower test scores in mathematics.

Early starters are positively selected even after controlling for observable characteristics. We use our estimates from equation (3) to adjust test scores by observable characteristics, as described in Section 5.2. Table 9 reports the estimates for the strength of selection for all months and cohorts, and for both subjects. The estimates of the strength of selection using adjusted test scores are positive and significant in all cases but one. A comparison of Tables 2 and 9 shows that, as expected, the strength of selection without controls is larger in most cases. However, the magnitude of this difference is small, also in most cases.

Finally, Table 10 reports all estimates for the penalty from early starting using adjusted test scores. The results are qualitatively similar to those reported in Table 3.

A.5 Descriptive Characteristics by Month of Birth

We use test scores from children born between May and December to construct counterfactual test scores for those born between January and April. In the first step of our methodology, we estimate age-in-month effects on test scores using only information from children born between May and December. In our second step, we compute predicted average test scores for students born between January and April using the coefficients obtained in the first step. In practice, we extrapolate the linear trend in test scores from May to December to the months from January to April. One potential concern for our identification strategy may arise if parents choose when to have their kids. In particular, parents of different characteristics may have children in different months. If so, the information contained in scores between May and December may not allow for an accurate prediction of average test scores between January and April.

We show that parents of children born between January and April have characteristics

Table 9: Estimates of the Strength of Selection. Adjusted Test Scores

(A) Mathematics Scores						
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.149	0.132	0.132	0.135	0.166	0.157
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.192	0.181	0.127	0.16	0.222	0.143
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.191	0.134	0.119	0.143	0.21	0.164
	(0.000)	(0.003)	(0.000)	(0.001)	(0.000)	(0.000)
April	0.139	0.263	0.058	0.181	0.219	0.268
	(0.021)	(0.000)	(0.209)	(0.002)	(0.000)	(0.000)
(B) Italian Scores						
	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.131	0.123	0.113	0.156	0.146	0.136
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.132	0.198	0.146	0.201	0.202	0.123
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.151	0.191	0.121	0.209	0.222	0.155
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.163	0.267	0.101	0.251	0.225	0.211
	(0.003)	(0.000)	(0.018)	(0.000)	(0.000)	(0.000)

Notes: Adjusted Test Scores represent residuals of regressing scores on controls. The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level. p-values are reported in parentheses.

Table 10: Estimates of the Difference in the Penalty from Early Starting for Selected and Average Students. Adjusted Test Scores

(A) Mathematics Scores					
	2012–13	2013–14	2014–15	2015–16	2016–17
January	0.034**	0.003	0.016	−0.01	0.056***
	(0.035)	(0.873)	(0.446)	(0.568)	(0.002)
February	0.057*	0.091***	0.054*	0.014	0.133***
	(0.066)	(0.001)	(0.083)	(0.600)	(0.000)
March	0.173***	0.155***	0.104**	0.086**	0.173***
	(0.000)	(0.000)	(0.019)	(0.018)	(0.000)
April	0.057	0.29***	0.147**	0.111**	0.092*
	(0.369)	(0.000)	(0.010)	(0.036)	(0.055)
(B) Italian Scores					
	2012–13	2013–14	2014–15	2015–16	2016–17
January	−0.019	−0.012	−0.059***	−0.012	0.017
	(0.198)	(0.420)	(0.001)	(0.456)	(0.397)
February	−0.07***	0.031	−0.055*	−0.014	0.073***
	(0.008)	(0.197)	(0.066)	(0.600)	(0.009)
March	0.039	0.088**	−0.021	0.021	0.073*
	(0.369)	(0.013)	(0.626)	(0.559)	(0.053)
April	−0.078	0.148***	−0.022	0.036	0.039
	(0.213)	(0.003)	(0.687)	(0.476)	(0.478)

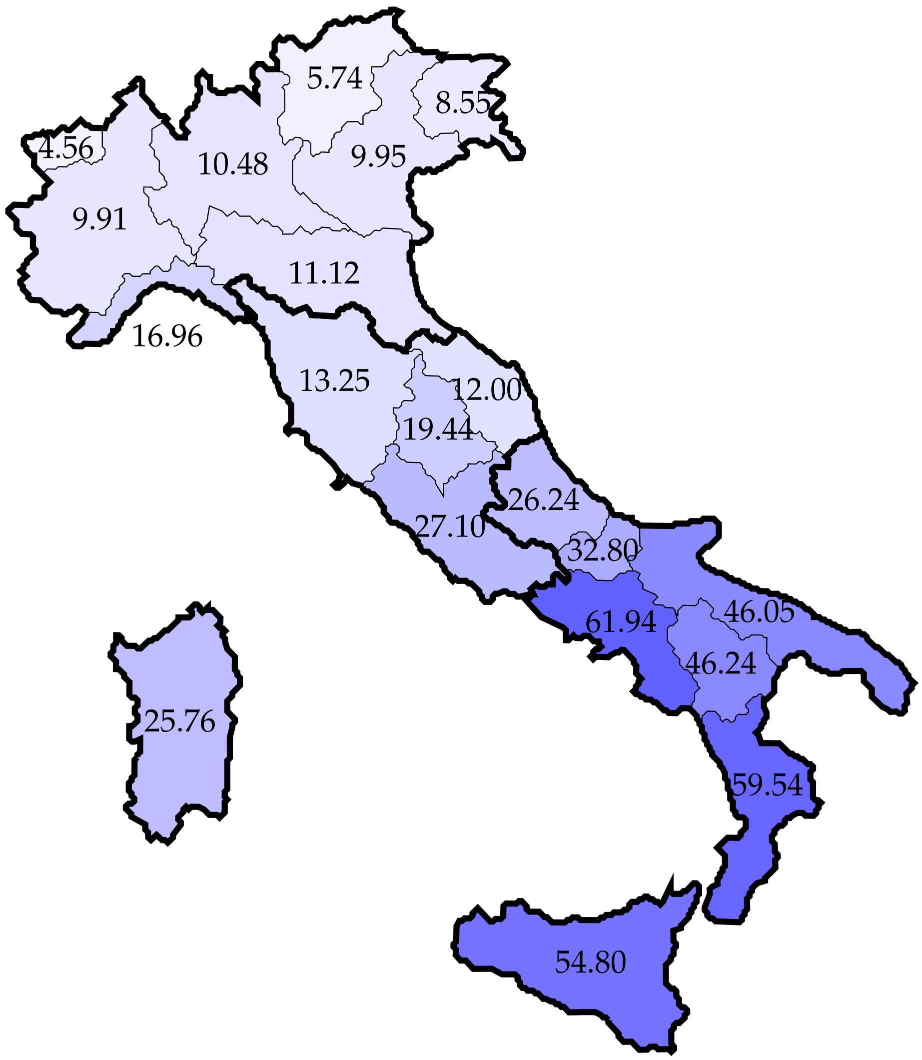
Notes: This table presents our estimates for the differences in the effect of early starting for selected students and the average student in the population. A positive difference implies that the penalty from early starting is lower for selected students. Standard errors are computed using bootstrap at the school level. p-values are reported in parentheses.

that are almost identical to those of parents of children born between May and December. To do so, we compare columns (III) and (IV) in Table 7. Column (III) summarizes the characteristics of students and parents for children born between January and April, while column (IV) summarizes these characteristics for children born between May and December. The difference between the proportions reported in columns (III) and (IV) is at most 0.7%.[10]

Finally, we perform a worst-case scenario robustness exercise. We use the tools that we develop in Section 5.2 to adjust the predicted test scores of children born between January and April by their own observable characteristics. In this way, the prediction accounts for the potential impact of the small differences in observables. We show first that the predicted average test scores change only marginally. Then, we estimate the strength of selection and we find that the results change in less than 0.1%.

[10]This dataset contains almost 2 million observations. Thus, any statistical test rejects the hypothesis that the proportions are equal, even with extremely small differences.

Figure 6: Proportion of Early Starters by Region. All Cohorts



Notes: This map reports the fraction of early starters on the total number of students born between January and April. We report proportions by region, for grade 2.