

NO COUNTRY FOR YOUNG KIDS?

The effects of school starting age throughout childhood and beyond

PLEASE DO NOT CITE WITHOUT PERMISSION

Latest version: November 21, 2019

Gonçalo Lima	Miguel Madeira Ruivo	Ana Balcão Reis
goncalo.lima@novasbe.pt	22182@novasbe.pt	ana.balcao.reis@novasbe.pt

Luís Catela Nunes	Maria do Carmo Seabra
lcnunes@novasbe.pt	carmo.seabra@novasbe.pt

The authors would like to thank Diogo Pereira for the excellent research assistance provided in the early stage of this work. We would also like to thank participants of the 5th Lisbon Economics and Statistics of Education research workshop, the 27th Meeting of the Economics of Education Association, the 10th International Workshop on Applied Economics of Education, the 13th Annual Meeting of the Portuguese Economic Journal, the 2019 International Conference on Public Economic Theory and the 2019 Education Research Oxford Symposium whose comments greatly benefited the current version of this work. Without the anonymized administrative data provided by the Portuguese Directorate General for Education and Science Statistics (DGEEC) this work would have not been possible. We also gratefully acknowledge the support of Portugal's Foundation for Science and Technology (FCT) through grant Nova SBE/BIM/2018/06. All authors affiliated with Nova School of Business and Economics, at Rua da Holanda, n.1, 2775-405, Carcavelos, Portugal.

Extended Abstract

Children start school at different stages of their social, emotional and cognitive development. That these differences play a significant role in explaining academic success is a well-established empirical fact. Less clear is how certain institutional features may help amplify or mute the extent of these effects throughout a child's schooling career. We develop a simple theoretical model tapping into how may school starting age rules, when coupled with given grade repetition policies, educational tracking, compulsory schooling laws and higher education access requirements, throw individuals into different human capital accumulation paths. To test our hypotheses we use de-identified longitudinal administrative records of every student enrolled in public schools in Portugal. Our identification strategy relies on exploiting small differences in birth dates around a binding school entrance cutoff. Local regression discontinuity estimates show that being 1-year older when starting school leads to gains in perceived cognitive capacity in Language and Math at grades 4, 6 and 9. Nonetheless, the gains on test scores from starting school later fade quickly from over 0.3 standard deviations in grade 4 to around 0.15 s.d. by grade 9. The rate at which local average treatment effects fall suggests that the findings are consistent with the hypothesis that differences in cognitive maturity when taking the test - rather than school entry postponement - are driving the results. However, effects on individuals persist through institutional features, well beyond elementary education. Older entrants have a lower probability of repeating grades (6 p.p.) and dropping out of school (2 p.p.). Significantly, reduced form estimates also show that older students are more likely to enroll in the academic track (2 p.p.) and scientific curricula (3 p.p.) on upper secondary education, have higher application scores to access public higher education (0.1 s.d.) and enroll in more selective undergraduate courses (0.12 s.d.). We find no evidence of differences on the demand for college seats, enrollment in STEM courses, or first-choice application success. Our results provide support to policies that, without promoting the postponement of a child's school starting age, incentivize educators to take into account cognitive maturity differences attributable to age, when considering retaining a child in grade or pondering on preventive pedagogical strategies.

JEL Classification: H75, I21, J13.

1. Introduction

Every year millions of children around the world enter school for the first time. However, school age regulations make them initiate formal schooling at very different stages of their social, emotional and cognitive development. That these differences play a significant role in explaining individual outcome, and academic success in particular, is a well-established empirical fact¹. Individuals starting school older typically reveal higher cognitive capacity, as measured by standardized achievement tests (e.g. Bedard and Dhuey, 2006; Puhani and Weber, 2007; McEwan and Shapiro, 2008; Elder and Lubotsky, 2009; Cascio and Schanzenbach, 2016; Attar and Cohen-Zada, 2018).

Less clear is how given institutional features change the extent of these effects throughout a child's schooling career and onto adulthood. Measured cognitive differences tend to fade as children age, but effects may persist through other mechanisms. In countries where grade retention is a typical strategy, younger entrants are likelier to repeat grades (e.g. McEwan and Shapiro, 2008). In systems that track students into different curricular offers, older students are likelier to choose or be tracked into an academically-oriented offer (Puhani and Weber, 2007; Schneeweis and Zweimüller, 2014; Attar and Cohen-Zada, 2018). We start by building a simple theoretical model of human capital accumulation, dependent on age and education system rules. In particular, we attempt to capture how school starting age rules, when coupled with given grade retention policies, educational tracking, compulsory schooling laws and higher education access requirements, may throw individuals into different human capital accumulation paths. Does entering school have an impact beyond cognitive capacity, or are younger students disadvantaged in other ways?

Recent literature shows that age differences impact on individual and social well-being through mechanisms other than academic success. Younger students are more likely to be classified as having learning disabilities and attention deficit disorders (Dhuey and Lipscomb, 2010; Elder and Lubotsky, 2009; Evans et al., 2010; Mühlenweg et al., 2012), are less persistent and more irritable (Mühlenweg et al., 2012), are significantly more likely to suffer from bullying or victimization (Mühlenweg and Puhani, 2010a) and less likely to hold leadership positions in high-school (Dhuey and Lipscomb, 2008). Younger entrants are also shown to have a higher propensity to commit crimes as teenagers (Landersø et al., 2017), as well as a higher likelihood of juvenile delinquency (Cook and Kang, 2016)

¹The literature on the effects of school starting age emerges from such diverse contexts as the United States (e.g. Dhuey and Lipscomb, 2010; Elder and Lubotsky, 2009; Evans et al., 2010; Dobkin and Ferreira, 2010; Cascio and Schanzenbach, 2016; Cook and Kang, 2018), Sweden (Fredriksson and Öckert, 2014; Carlsson et al., 2015), Norway (Black et al., 2011), Denmark (Landersø et al., 2017), Germany (Fertig and Kluve, 2005; Puhani and Weber, 2007; Mühlenweg and Puhani, 2010a,b; Mühlenweg et al., 2012), Austria (Schneeweis and Zweimüller, 2014), the UK (Crawford et al., 2007), Poland (Herbst and Strawiński, 2016), Hungary (Altwicker-Hámori and Köllö, 2012), Chile (McEwan and Shapiro, 2008), Canada (Smith, 2009), China (Zhang et al., 2017), Japan (Kawaguchi, 2011), Australia (Suziedelyte and Zhu, 2015), Israel (Attar and Cohen-Zada, 2018), Italy (Ponzo and Scoppa, 2014) or Mexico (Peña, 2017). Bedard and Dhuey (2006) present cross-country evidence.

or of being incarcerated for juvenile crime (Dhuey et al., 2017). The impact on long-term outcomes is somewhat more ambiguous. While some find a causal link between starting school later and higher wages (Fredriksson and Öckert, 2014; Kawaguchi, 2011) or the likelihood of becoming a corporate CEO (Du et al., 2012), others do not find long-term effects on prime-age earnings (Black et al., 2011; Dobkin and Ferreira, 2010)².

To test our hypotheses, we empirically estimate the impact of school starting age on a wide range of individual outcomes, from early childhood until the end of upper secondary education. For that purpose, we use de-identified longitudinal administrative records of every student enrolled in public schools in Portugal. To the best of our knowledge, we provide the first plausibly causal evidence about the impact of school starting age on student outcomes in Portugal³, a school system with idiosyncratic characteristics: contrary to most countries, beyond a binding enrollment cutoff at 1 January, parents whose children are born as early as 16 September have leeway to legally postpone their child's entrance in school.

Our identification strategy exploits variation in school starting age around cutoff discontinuities using exact birth dates (such as in McEwan and Shapiro, 2008; Evans et al., 2010; Dobkin and Ferreira, 2010; Peña, 2017; Attar and Cohen-Zada, 2018), as these enable us to avoid biases induced by seasonal patterns present in coarser measures, such as quarter or month of birth (as noted in Buckles and Hungerman, 2013). Resting on a well-identified set of falsifiable assumptions, our design is analogous to a local randomized experiment (Lee and Lemieux, 2010). Our estimates rely on local polynomial methods in accordance with the growing methodological consensus for their adequacy in regression discontinuity designs (Gelman and Imbens, 2017; Imbens and Lemieux, 2008)⁴. Given the longitudinal nature of our samples, and in order to allay concerns with non-compliance, we estimate local average treatment effects within the context of a fuzzy regression discontinuity design, going beyond simple intent-to-treat effects for most of our estimates.

We find that being 1-year older when entering school leads to significant gains in perceived cognitive capacity in Language and Math at grades 4, 6 and 9. In grade 4 the age premium is of sizable .3 standard deviations (σ) in Math and $.38\sigma$ in Language exam scores. Nonetheless, the gains on test scores from starting school later fade quickly. By the end of Grade 9 we estimate the impact to be

²For Norway, Black et al. (2011) find that the effect of school starting age becomes insignificant from 30-years of age onwards, culminating in a negative discount present value of lifelong earnings gain from starting school 1-year later. In Sweden, Fredriksson and Öckert (2014) find that being 1-year older increases educational attainment and changes labor supply over the life-cycle, with initial losses due to later entry into the labor market being offset by increased earnings after the age of 55. However, contrary to the average, individuals with low-educated parents and women tend to have higher prime-age earnings in response to school starting age. All-in-all, discounted life-time earnings are shown to fall by almost 1 percent in response to being 1-year older when enrolling in first grade.

³The exception being the estimates in Bedard and Dhuey (2006), in a cross-country context.

⁴Most recently, for instance, Cook and Kang (2016) use a local-linear estimator to uncover the causal effect of school starting age on crime outcomes. However, most studies on the impact of school starting age on academic outcomes have not relied on these type of methods. Most have focussed on exploring global – rather than local – polynomial specifications (e.g. McEwan and Shapiro, 2008; Black et al., 2011).

of about $.13\sigma$ in Language and only $.09\sigma$ in Math. The rate at which local average treatment effects fall suggests that the findings are consistent with the hypothesis that differences in cognitive maturity when taking the test - rather than school entry postponement - are driving the results. As we only have metrics of individual outcomes measured at the same time – not at the exact same age – the estimates combine the effects of starting school age, with that of age-at-test. Unfortunately, we cannot access a second source of exogenous variation to separately estimate each of these effects. However, a simple model of children’s cognitive development suggests that, at the rate of decline predicted by our estimates, the orthogonal impact of delaying entrance may be null, if not negative (such as in Crawford and Iriberry, 2007; Black et al., 2011; Peña, 2017).

Despite the apparent decline in the cognitive relative premium, we also find that school starting age differences persist through institutional features, well beyond elementary education. In a country where grade retention remains ubiquitous as a remedial strategy, older entrants have a 5 percentage points (pp) lower probability of repeating at least one year in primary education, and 4pp by the end of grade 9. Older students also have a lower probability (-2pp) of dropping out from school. Significantly, reduced-form estimates also show that older students are more likely to enroll in the academic track (2pp) and scientific curricula (3pp) on upper secondary education, have higher application scores to access public higher education ($.1\sigma$) and enroll in more selective undergraduate courses ($.12\sigma$). On the other hand, we find no evidence of differences on demand for college seats, enrollment in STEM courses, or first-choice application success.

The remainder of this paper is structured as follows. Section 2 presents the conceptual framework. Section 3 highlights the empirical context, and section 4 describes the empirical strategy given the institutional background in Portugal. Section 5 details the data used in the analysis. Section 6 presents the results. Section 7 concludes.

2. Conceptual Framework

2.1. A simple model of human capital accumulation

We begin with a simple model of a child’s cognitive development. As in Elder and Lubotsky (2009), we assume a child accumulates human capital (h) in school and out-of-school settings. Prior to enter the first grade of school, the child accumulates human capital mainly through family investments at home or exposure to formal pre-schooling or kindergarten environments. Furthermore, a family endowment function (E) maps a vector of observable and unobservable family background characteristics, as well as innate child abilities (\mathbf{X}) onto a given scalar in each relevant period t of a child’s life, both before and after starting school.

When a child starts her formal schooling path, her human capital accumulation function depends on both her readiness to learn – independent of home endowments – as well as exposure to schooling.

Readiness to learn is captured by the age at which the child starts school (A), whereas schooling is given by the number of years in school (g) and the match between student characteristics and schooling. Human capital accumulated until a give period t is thus given by:

$$h_t = \rho h_{t-1} + E_t(\mathbf{X}) + S_t(g, A, \mathbf{X}) \quad (2.1)$$

With $1 - \rho$ being the rate at which human capital accumulated until period $t-1$ (h_{t-1}) depreciates and S_t defining the impact of school year g for a student entering school at age A on human capital. As a first order condition we define the vector of student characteristics \mathbf{X} in such a way that $E_t(\mathbf{X})$ is strictly increasing on each element of \mathbf{X} , observable or otherwise: $\partial E_t(\mathbf{X})/\partial \mathbf{X} > 0$.

As an hypothesis – grounded on empirical evidence – we pose that an additional year of exposure to formal schooling has non-negative returns to human capital, in any given period t and grade g .

$$\frac{\partial h_t}{\partial S_t} \cdot \frac{\partial S_t(g, A, \mathbf{X})}{\partial g} \geq 0 \quad (2.2)$$

Importantly, the marginal impact of schooling to human capital accumulation in this framework can differ according to student characteristics and the age at which the child enters school. The effect of being exposed to an additional grade of schooling may be stronger for children from higher socio-economic status or those that have entered school at a later stage of development. Therefore, it does not only matter how much exposure to school has a child received, but also the moment at which such exposure started. We explore the impact of school starting age on human capital.

Within our framework, as in [Elder and Lubotsky \(2009\)](#), the human capital of a child one school year ($g = 1$) after the beginning of her formal schooling is given by the human capital accumulated in pre-school environments, human capital endowed out-of-school and that acquired in school environments:

$$h_{A+1} = \rho h_A + E_{A+1}(\mathbf{X}) + S_{A+1}(1, A, \mathbf{X}) \quad (2.3)$$

Generalizing for any given grade $g = k$:

$$h_{A+k} = \rho^k h_A + \sum_{j=1}^k \rho^{k-j} \{E_{A+j}(\mathbf{X}) + S_{A+j}(j, A, \mathbf{X})\} \quad (2.4)$$

The marginal return of a 1-year increase in A on human capital k -grades after starting school is thus given by:

$$\frac{\partial h_{A+k}}{\partial A} = \rho^k E_{A-1}(\mathbf{X}) + \sum_{j=1}^k \rho^{k-j} \left\{ \frac{\partial E_{A+j}(\mathbf{X})}{\partial A} + \frac{\partial S_{A+j}(j, A, \mathbf{X})}{\partial A} \right\} \quad (2.5)$$

The total effect of school starting age on human capital depends on the rate at which this is depleted $(1 - \rho)$, relevant skills acquired prior to entering school, as well as the separate marginal effects of differences in parental investment and school starting age. Our empirical strategy is thus focused in identifying $\partial h_{A+k}/\partial A$ as the cognitive premium for being 1-year older when starting school, for each grade k .

2.2. Human capital in context

We now extend the analysis to include the policy environment in which the process of human capital accumulation is embedded. A given set of exogenous rules Ω shapes the means by which a child can accumulate skills throughout formal schooling, namely by hindering or accelerating the rate at which human capital is accumulated. Depending on their timing, some institutional features can project different children into different human capital accumulation paths.

As our main empirical identification strategy, we explore school starting age rules (r^A) that provoke random variation in the age at which children are first exposed to school. In the context of multiple education systems, regulations also allow to halt the progress of children through grade retention (r^\ominus) – a strategy to compensate for potentially inadequate human capital acquisition. Educational tracking (r^V) also exposes children of dissimilar ability and preferences to different curricula. Compulsory schooling laws (r^C) mandate until which age are children obliged to attend school. Finally, rules concerning access to higher education institutions (r^{HE}) may constrain the choice set of individuals willing to continue investing in their human capital.

We explore how the combination of school starting age laws with each of the other rules impacts on human capital as well as other individual outcomes. Each of the rules affect children in different stages of schooling.

Therefore, we extend our human capital accumulation model in Equation 2.1 as follows:

$$\begin{aligned} h_{A+g} = & \rho h_{A+g-1} + E_t(\mathbf{X}) + \mathbb{1}(g \leq \iota_{r^V}) \cdot S_t(g, A(r^A), \mathbf{X}, r^\ominus) + \\ & + \mathbb{1}(\iota_{r^V} < g \leq \iota_{r^{HE}}) \cdot S_t^U(g, A(r^A), \mathbf{X}, U(r^V, r^C, h_{t-1}, \mathbf{X}, r^\ominus)) + \\ & + \mathbb{1}(g \geq \iota_{r^{HE}}) \cdot S_t^{HE}(g, A(r^A), \mathbf{X}, HE(r^{HE}, h_{t-1}, \mathbf{X}), r^\ominus) \end{aligned} \quad (2.6)$$

Where ι_\bullet represents the grade at which each rule from set Ω binds, $S_t(\bullet)$ is the additional human capital accumulated at period t , if such period is until the moment of tracking decisions, $S_t^U(\bullet)$ is if the individual is exposed to a given track U of upper secondary education and $S_t^{HE}(\bullet)$ if the additional human capital is accumulated when exposed to a given higher education course HE . The choice of track and higher education course depends in turn on the binding rules, accumulated human capital

and other student characteristics.

3. Institutional background

The Portuguese school system is organized in three sequential levels: early childhood education and care, basic education and secondary education. Basic education covers the initial nine grades of schooling and is divided in three studying cycles, of various lengths. First cycle comprises the four initial grades of primary education and teaching of most subjects is under the purview of a single teacher. Second cycle has a length of two academic years. The third cycle of basic education – comprising grades 7 through 9 – corresponds to lower secondary education. At completion of basic education, typically at age 15, students transition to upper secondary education. Upper secondary education offer is divided in a general academic and vocational pathways. In the general academic track, students can select on of four concentrations: science and technology, social and economic sciences, languages and humanities, or visual arts. On the other hand, the vocational track offers a plethora of denominations, with curricula geared toward earlier integration in the labor market. Compulsory schooling laws in the country determine that students should be enrolled in school until finishing the academic year when they turn 18 or until completion of upper secondary education before the age of 18.

Students in the Portuguese school system are evaluated through teacher testing and national exams or assessments. Barring accommodations to specific student needs, national exams in Portuguese Language and Math are performed by every student in the system, by the end of fourth grade (until 2015) and ninth grade. Children sit through the exam at exactly the same time, facing the same questions. Exams are then evaluated by a randomly allocated evaluator teachers, from schools other than the school in which the student is enrolled, in an anonymous fashion. By the end of grade 6, national assessment tests follow a similar procedure. In order to complete the general academic track of upper secondary education, students must also sit through national exams – typically completing two course-specific exams in grade 11 and another two in grade 12, in most cases Language and Math. Students can only gain admission to tertiary education if they have a passing grade in both grade 12 exams.

Admission to public higher education in Portugal is centrally governed. Candidates are publicly listed by the government according to candidate's ranked preferences, application scores and available capacity. Application scores combine high school GPA and exam scores. The application score depends on the tertiary educational institution and the department to which the student applies. In the final application score, high school GPA must weigh a minimum of 50 percent in the admission decision. However, each tertiary education institution can set the weight of exam scores within a band of 35 to 50 percent of the total application score.

Most relevant to our identification strategy, Portugal's compulsory education laws dictate enroll-

ment in first grade to be mandatory for every child who is at least 6-years old by 15 September. Yet considerable leeway is given to parents wanting to enroll children after this date. Children born between 16 September and 31 December of a given calendar year are deemed conditional and can still enroll if parents so require and there are available places in already created classes in school¹. In fact, the existing rules implicitly generate a second – more binding – at 1 January, as children born in the beginning of the next calendar year enroll in the following school year. Children must thus be at least 5.7 years-old by 15 September, when starting school. Since most conditional students in Portugal are not deferred to enroll the following year², a child born in the beginning of January typically enrolls in first grade 1-year later than their peers born in the end of December.

4. Empirical strategy

4.1. Identification and estimation

We start with the basic relationship of interest captured by the following linear model:

$$Y_{ig} = \beta_0 + \beta A_i + \mathbf{X}_i \delta + \epsilon_i \quad (4.1)$$

Where Y_{ig} is the outcome of interest of student i measured by the end of school year g , A_i is the age of student i measured in decimal years as of 15 September in the year she first enrolled in first grade, \mathbf{X}_i is a vector of individual and family background characteristics measured during the year student i first enrolled in first grade, and ϵ_i is an idiosyncratic error term. In this setting, β represents the marginal effect of delaying enrollment by 1-year. Nonetheless, Equation 4.1 does not account for school starting age being likely correlated with characteristics of students and their families that are not typically observed in the data, such as learning maturity.

To overcome endogeneity concerns, we exploit exogenous variation induced by the school starting age regulations described in section 3. In section ?? we present evidence that exists a sharp discontinuity at 1 January, and a kink at the 16 September cutoff. Therefore, our main identification strategy relies on comparing outcomes of students that are born before the cutoff of 1 January and those that are born on or after that same date and are induced to enroll only in the following school year.

We interpret our regression discontinuity results in light of a potential outcomes framework (Hahn et al., 2001). Provided that other characteristics associated with the outcomes of interest are continuous at the cutoff and birth dates around the cutoff are as good as randomly assigned, the outcomes of those born before the cutoff provide reasonable counterfactual outcomes of those that are born after the cutoff had they enrolled one year earlier.

¹Importantly, enrollment for conditionals is not the default option; an enrollment requirement by parents is a necessary condition to starting the school year before turning 6-years old. However, it is not a sufficient one, as it also depends on school capacity constraints.

²In our dataset only 14% of eligible (conditional) students postpone enrollment.

In order to estimate our coefficients we first construct a continuous variable $B_i \in \mathbb{Z}$ with 366 unique integer values (allowing for leap years) that identify the birthday of student i in the calendar year, as in [McEwan and Shapiro \(2008\)](#). We standardize B_i as a distance (in days) to the cutoff of 1 January ($B_i = 0$) and make it run from 1 July ($B_i = -184$) to 30 June ($B_i = 181$), so that we have the discontinuity at about mid-range of the running variable. Based on it we define an indicator variable, $\tau_i = \mathbb{1}(B_i \geq \bar{B}_i)$, which indicates the values of B_i that are equal to or exceed the enrollment cutoff of 1 January ($\bar{B}_i = 0$).

We estimate the discontinuity using both local-linear and local-quadratic regression techniques, through the following reduced form weighted least squares specification¹:

$$\min \sum_{i=1}^{N(h)} (Y_{ig} - \alpha_0 - \alpha\tau_i - \mathbf{f}(B_i))^2 \mathbf{K}_h(\tau_i, B_i) \quad (4.2)$$

Where $\mathbf{f}(B_i)$ is, depending on the specification, a piecewise linear ($\phi_1 B_i + \phi_2 \tau_i B_i$) or a piecewise quadratic ($\sum_{p=1}^2 \phi_p B_i^p + \sum_{p=1}^2 \phi_{pp} \tau_i B_i^p$) function of the running variable interacted with the cutoff. $\mathbf{K}_h(\tau_i, B_i)$ is a triangular weighting kernel function with bandwidth h , given by:

$$\mathbf{K}_h(\tau_i, B_i) = \max\left(0, 1 - \left|\frac{B_i}{h}\right|\right) \quad (4.3)$$

The bandwidth h here denotes the window of birth dates (B_i) to the left and to the right of the cutoff, used to estimate the coefficient of interest α , and $N(h) : \mathbb{1}\{-h \leq B_i \leq h\}$. The triangular kernel assigns zero weight to all observations outside the interval defined by the bandwidth $[-h, h]$ and positive weights to all observations within, with weights declining symmetrically and linearly as the value of the running variable gets farther away from the cutoff. In order to avoid imposing an ad hoc bandwidth length, we use recently developed data-driven methods to select optimal bandwidth for each regression. In particular, we use an upgraded version of the mean square error (MSE) optimal bandwidth selectors discussed in [Imbens and Kalyanaraman \(2012\)](#) (for the linear case) and [Calonico et al. \(2014b\)](#) (for the quadratic case) and derived in [Calonico et al. \(2018\)](#)².

Nonetheless, simply estimating treatment effects at the cutoff does not provide the true impact of school starting age on outcomes. Since existing rules in Portugal provide leeway for parents to delay children's enrollment between the 16 September and 31 December (covering our control group), and since some parents may not comply with the more binding cutoff of 1 January, α captures at best an intent-to-treat effect. Therefore, we estimate local average treatment effects (LATE) through a fuzzy regression discontinuity design, using the indicator of being born after the cutoff (τ_i) as an instrument

¹[Gelman and Imbens \(2017\)](#) show why regression discontinuity estimation through local low-order polynomial approximations should be preferred to global polynomial regressions.

²Estimation, including optimal bandwidth selection, is implemented through the software package `rdrobust` developed in [Calonico et al. \(2014a\)](#) and [Calonico et al. \(2017\)](#).

for school starting age (A_i):

$$\min \sum_{i=1}^{N(h)} (A_i - \theta_0 - \theta\tau_i - \mathbf{f}(B_i))^2 \mathbf{K}_h(\tau_i, B_i) \quad (4.4)$$

$$\min \sum_{i=1}^{N(h)} (Y_{ig} - \beta_0 - \beta\hat{A}_i - \mathbf{f}(B_i))^2 \mathbf{K}_h(\tau_i, B_i) \quad (4.5)$$

First stage Equation 4.4 allows to attest if there is sharp variation at the cutoff, as well the ability of the cutoff to predict school starting age. Estimates of β , in Equation 4.5, are thus the local average treatment effects for those complying with the cutoff (Imbens and Angrist, 1994). For precision, we also estimate models controlling for a vector of individual and family background characteristics (\mathbf{X}_i) and cohort fixed effects (φ_c):

$$\min \sum_{i=1}^{N(h)} (A_i - \theta_0 - \theta\tau_i - \mathbf{f}(B_i) - \mathbf{X}_i\delta - \varphi_c)^2 \mathbf{K}_h(\tau_i, B_i) \quad (4.6)$$

$$\min \sum_{i=1}^{N(h)} (Y_{ig} - \beta_0 - \beta\hat{A}_i - \mathbf{f}(B_i) - \mathbf{X}_i\psi - \varphi_c)^2 \mathbf{K}_h(\tau_i, B_i) \quad (4.7)$$

Importantly, the plausibly causal interpretation of the effects rests on the assumption that parents do not systematically time births relative to the cutoff. In the next section we provide evidence that there does not seem to be systematic manipulation of the running variable close to the cutoff. Additionally, we run balancing analysis of observable socioeconomic characteristics of the students around the cutoff. Finally, our working assumption is that precise birth timing around the cutoffs does not introduce sharp differences in unobserved characteristics that affect our outcomes of interest.

4.2. Differential exposure to schooling and age-at-test

An important concern with our identification strategy is the timing at which individual outcomes are measured. Because in our data we can only observe performance in tests sat by children at the same time – not at the exact same age – estimated β 's will capture the net result of two effects: an 'exposure to schooling' effect and an 'age-at-test' effects. Those that are induced by the cutoff to start school 1-year later are also older when performing the test.

Based on the typical physical and cognitive development of children, some ages are better than others to start learning in school. As learning effort in formal school environments may be higher for younger children, cognitive development may be hindered by a too early exposure to school which can persist throughout the individual's life. On the other hand, the youngest children in each cohort may tend to perform poorly simply because they are at an earlier stage of cognitive development than their

peers.

Given data limitations, few studies managed to convincingly separate each effect (Cornelissen and Dustmann, 2019; Peña, 2017; Black and Devereux, 2011; Elder and Lubotsky, 2009; Crawford et al., 2007), consistently finding that the ‘age-at-test’ effect dominates and that delaying the moment of first exposure to schooling may be actually detrimental in cognitive terms³. Not being able to use students’ cognitive abilities at exactly the same age as a second source of exogenous variability, our identification strategy is not able to disentangle differential cognitive maturity – simply stemming from the fact that students are older – from differences in the age at which they began being exposed to formal schooling. It means that the margin in equation 2.5 is not identifiable in our empirical context. However, we extend the model of human capital accumulation presented in section 2 in a way that may shed light on the potential influence of each of these effects in the Portuguese education system.

As additional theoretical constraints to our model in equations 2.1 and 2.6, we assume that the stock of human capital (h_t) of a given individual age t grows at a weakly decreasing rate, such that:

$$\frac{\partial h_t}{\partial t} \geq 0 \quad \wedge \quad \frac{\partial^2 h_t}{\partial t^2} \leq 0 \quad (4.8)$$

According to empirical literature, we argue that these first-order conditions are not too stringent. Independently of the magnitude of $\partial h_t / \partial t$ simply conforms to the fact that children typically accrue greater gains in cognition in earlier stages of their development. We assume that these derivatives are, nonetheless, proportional to the age growth of at a given age t (in decimal years), so that:

$$\frac{\partial h_t}{\partial t} \propto \pi(t) = \frac{1}{t-1}, \quad t > 1 \quad (4.9)$$

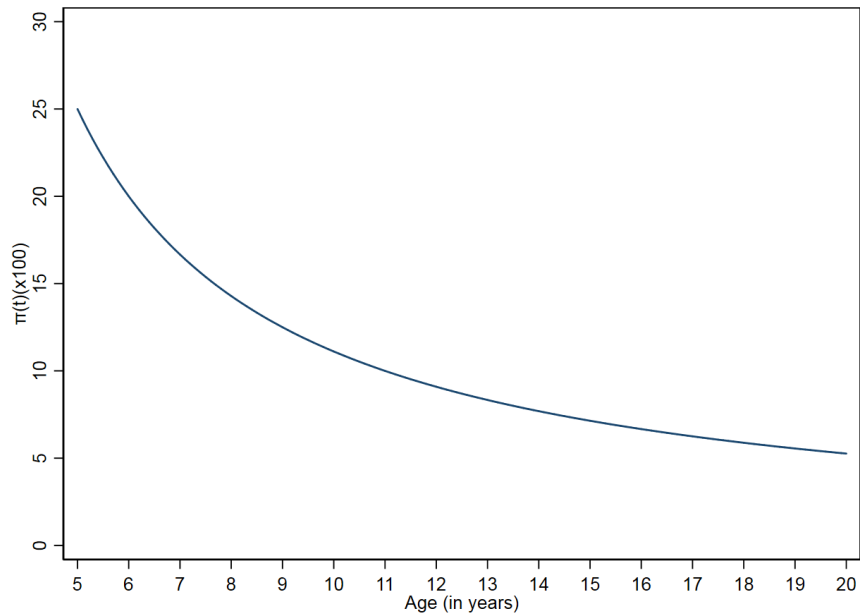
Given these theoretical conditions, human capital will accumulate at a rate that will closely track the growth rate in age. As shown in Figure 4.1, a complier child born in January will, by the end of Grade 1 (in June), be approximately 7.5 years-old and about 15% older than a complier child born

³Black et al. (2011) exploit both variation in the cutoff dates for entering first grade in Norway and in the age at which individuals take IQ tests to access mandatory military service to independently identify each of the effects. Results show that being 1-year older leads to 10 percent of a standard deviation gain in IQ tests, but that entering school one year older actually has an independent negative impact of 2 percent of a standard deviation, implying that the net effect of being 1-year older is only of 8 percent of a standard deviation. Crawford and Iriberry (2007) leverage variation in school starting age across local education authorities in England to separate school starting age and age-at-test effects similarly find that the latter dominate. On the other hand, Cornelissen and Dustmann (2019) finds that for England, earlier exposure to schooling contributes to closing more than 4/5 of the gap between high-SES and low-SES children early on, even if the independent effects of school starting age dissipate after age 11. Identically, Peña (2017) uses regional variation in changes of cutoff dates in Mexico and, through a difference-in-differences approach, finds that the independent effect of school starting age is negative, implying that if younger students were tested at the exact same age of their older peers they would outperform them. Other studies, on the other hand, estimate the combined effect of exposure to schooling effect and school starting age. Elder and Lubotsky (2009) find that when kindergarten children are tested at the same age (not at the same time), they score lower in in-school tests if they had counter-factually started school older and thus had less exposure to formal education.

in December. On the other hand, at 15.5 years of age, when a January-born child with no repetitions typically ends Grade 9, she will only be about 7% older than a child one-year younger.

If $\partial h_t / \partial t$ is proportional to $\pi(t)$, then we predict human capital, orthogonally explained by age (say, cognitive maturity), to decline at a similar rate as $\pi(t)$ declines, if the impact of differential exposure to schooling is null. Although we do not have an absolute measure of human capital, by measuring the rate at which predicted marginal effects $\partial h_{A+g} / \partial A$ change throughout g , we can infer whether the effect of delaying exposure to schooling is likely positive or negative. If the second derivative of h_{A+g} is smaller than the first derivative of $\pi(t)$, then our simple model suggests – in line with evidence from other countries – that late-entrants, if sitting the exam at the exact *same age*, not at the *same time* as compliers born before the cutoff, would actually perform worse.

Figure 4.1: Proportional age growth



Notes: Figure depicts age growth according to $\pi(t)$. $\pi(t)$ gives, in percentage terms, how much more older is someone t -years old, compared to someone 1-year younger.

5. Data

5.1. Data description

We use a de-identified administrative dataset (MISI)¹ containing detailed information on every student enrolled in public schools in mainland Portugal from 2007 to 2016. We focus on non-adult students enrolled in the regular public system of basic education².

¹MISI data is collected and maintained by the Directorate General of Education and Science Statistics (*Direcção-Geral de Estatísticas da Educação e Ciência* - DGEEC), a department under the indirect administration of the Ministry of Education in Portugal.

²Therefore, we exclude from the analysis all students enrolled in second-chance programmes, adult, vocational education and training or artistic courses outside the scope of the regular curriculum between Grades 1 and 9.

Originating from an information management system, MISI contains relevant data on birthday and socioeconomic characteristics of each student – such as parents’ education, country of origin, home neighborhood, eligibility for social support, access to computer or Internet at home – with minimum measurement error or missing information. A unique student identifier allows us to track students throughout grades and gather additional information about their educational pathway. We thus have a panel dataset of students since they are first observed in the public education system. A student’s track is lost when she moves abroad, drops from the education system altogether, or dies. We may also lose track of students if these move to a different public or private school and the matching algorithm is unable to correctly assign the unique identifier to new instances of the same student in the system.

We merge MISI data with a two other administrative datasets (ENEB and ENES) containing comprehensive information on student achievement in Mathematics and Portuguese Language basic education national exams, as well as track- and course-specific exams undertaken by every eligible student in the country. Standardized achievement tests in Portugal are not a stable policy, so we can only recover a few years of outcome data. During the period for which there is available data we can gather Grade 4 national exam scores sat between 2013 and 2015, as well as Grade 6 scores for the period 2012-2015 and Grade 9 information for the entire period of the dataset (2008-2017). Since, for basic education outcomes, we estimate local average treatment effects, we must observe all students when in Grade 1, in order to confirm they actually complied with the virtual assignment to control and treatment groups.

All information in period 2013-2015 for both Grades 4 and Grade 6 students is kept, whereas achievement information for Grade 9 comes from exams sat during 2016 and 2017. For Grade 4 students sitting exams in 2013-2015 we manage to follow seven different cohorts, starting school between 2006 and 2012. For each of these cohorts we can follow over 95 percent of the children between Grade 1 and Grade 4. We follow five cohorts of Grade 6 sample students entering school in the period 2006-2010, with attrition rates lower than 10 percent. Finally, given our lower bound on the data, we can follow more than 90 percent of two cohorts students until they sit Grade 9 exams in either 2016 or 2017. As we track students from their first school year to later periods in our dataset we lose more students either to private schools, a foreign country, registration failures, other non-regular education programs, or idiosyncratic errors in the matching algorithm. Annex 9.1 documents attrition rates for each of the different cohorts.

Our first analytical dataset contains student-level information at Grade 1 as well as student outcomes for those that sat at least one Math or Language Grade 4, Grade 6 or Grade 9 national exam. We also trim the dataset for students with outlier ages by only keeping children that are at least 5-years old and had at most 8-years old when first enrolled in Grade 1. The first analytical dataset has

a total of 660 494 individual students³, born between 1998 and 2008, having started school between calendar years 2006 and 2013.

A second analytical dataset is constructed for reduced form estimates of upper secondary outcomes and out-of-sample estimation of basic education outcomes. Since, in this case, we do not require to observe each student at Grade 1, we estimate intent-to-treat effects from a total of 2 227 679 individuals born between 1983 and 2010. For an analysis of upper secondary and post-secondary outcomes, we further restrict this dataset to one where each student is followed from Grade 9 until the end of Grade 12. The restricted sample follows a total of 634 827 individuals born between 1988 and 2000, from which 171 319 (27%) applied to public higher education after graduating from high school. In the case of college application outcomes, we benefited from an hand-collected dataset, that was available on-line for a given period of time and allowed to link the data with the existing administrative datasets⁴.

5.2. Main variables

Achievement (Y_{ig}): The main outcome variables are constructed from score points of students in Portuguese Language and Math national exams by the end of Grades 4, 6 and 9, as well as other course-specific and track-specific national exams by the end of Grades 11 and 12⁵. In order to account for differences in exam difficulty across years, we standardize our measure of achievement by subtracting the mean and dividing by the standard deviation within exam year. The outcome variable is thus interpretable in standard deviation units of the exam scores in a 0-20 scale.

Grade retention (Y_{ig}): School grade repetition still is a widely used strategy to compensate for students' under-performance in Portugal. We thus consider grade retention as an outcome variable. In particular, we construct dummy variables switched on for if the student was retained in the same school year at least once until each of Grades 4, 6, 9, 10, 11 and 12.

Attainment outcomes (Y_{ig}): We measure attainment outcomes in basic education as an indicator variable switched on for if the student dropped out of school before finishing her basic education, and a graduation variable indicating whether the student successfully completed basic education.

Tracking outcomes (Y_{ig}): Tracking outcomes include two variables. A first variable indicates

³The distribution of information on their outcomes is as follows. Grade 4 sample: 229 661 (35%). Grade 6 sample: 300 182 (45%). Grade 9 sample: 188 648 (29%). 10% of the students have scores for both Grades 4 and 6. 28% have scores for both Grades 6 and 9. None has observed scores for all grades.

⁴Students in the college application datasets were all anonymized under the algorithmically set unique identifier code.

⁵Each school year, the the exam for the Grade is identical for all eligible students in the country. Grade 4 exams were discontinued from the beginning of the 2015/16 school year onwards. Additionally, the score scale at which ability was measured has changed from school year 2012/2013 onwards, with the previous discrete scale (0-5) being insufficient to retain relevant variation across students. Given these constraints, we retain three years of observations of Grade 4 outcomes (2013-2015) as measured by exam scores (0-100 scale) by the end of the school year. In order to allay grade manipulation concerns we first re-scale every exam score into a 0-20 scale. Refer to [Annex A.2](#) for a full description of the construction of these variables. Course- and track-specific exams include the Grade 11 subject exams of Physics, Biology, Geometry, Economics, Philosophy and Geography, as well as Grade 12 subject exams of Language, Math A, Physics, Biology, History, Geometry, Drawing and Economics.

whether the student, conditional on having completed grade 9, decided to enroll in the general academic track. A second dummy indicates whether the student, conditional on having opted for the general academic track, decided to pursue a science-oriented curricula in high school.

Post-secondary outcomes (Y_{ig}): We also observe the outcomes of students that have successfully graduated from high school, in particular: *i*) if the student applied to a public higher education degree; *ii*) if the student enrolled in an academic degree in higher education; *iii*) if, conditional on having been accepted in an academic undergraduate degree, enrolled in a STEM course; *iv*) if the student was rejected from public higher education; *v*) if the student in the first phase of applications; *vi*) the application score of each higher education candidate, standardized to have a mean of zero and standard deviation of one; and *vii*) the level of selectivity of the higher education course in which each successful candidate enrolled. Course selectivity is measured as an index of *i*) percentile rank of the pair higher education institution course in terms of the application scores of accepted candidates, *ii*) percentile rank of the standard deviation of application scores of accepted candidates, and *iii*) acceptance rate of applications. The values of the latent variable are predicted through principal factor analysis, with results being later standardized to have mean zero and standard deviation of one.

School starting age (A_i): We define school starting age – our main regressor of interest – as the exact student age as of 15 September of each school year. It measures the age at which the student is first observed in first grade. By the way it is constructed, unit variations in this variable represent a one-year variation in the age at which the student is first enrolled in Grade 1.

Birth date (B_i): The exact date of birth is measured as a continuous variable (measured in days) representing all possible 366 days of the calendar year (see Section 4.1). It is the running variable in our regression discontinuity strategy, from which we also extract the relevant post-cutoff indicator for the analysis (τ_i).

Student characteristics (X_i): We construct a vector of variables for several observable characteristics of students such as: *i*) gender (switched on for female); *ii*) an indicator for first generation immigrant (switched on for if the student is not born in Portugal); *iii*) an indicator for if the student has access to a personal computer at home; *iv*) an indicator of reciprocity of school social support (*Acção Social Escolar - ASE*) as a proxy variable for household financial constraints; *v*) unemployment status of the child’s father; and *vi*) a proxy for the level of education in the household. To construct this last variable we minimize problems with missing values by including the maximum level of education (in terms of years of study required) in the household (from either the father, the mother, or the legal guardian of the child, in case information for any of the parents is not available). All these characteristics are measured as of the school year in which the student is first enrolled in Grade 1.

5.3. Descriptive statistics

Table 5.1 documents summary statistics of the outcomes and covariates for each full Grade sample, and separately for students born 60-days before and after the 1 January cutoff⁶. The groups born before and after the cutoff seem to be relatively homogeneous with respect to their observable individual characteristics, at least for a statistical significance of 1 percent. Prevalence of female students, students with access to computer at home, first generation immigrants, recipients of school social support or those whose dad were in an unemployment situation when entering school is similar across samples, as expected were the students randomly allocated around the cutoff. For higher levels of confidence, however, statistical differences may be found in some of the covariates. Were we estimating intent-to-treat effects and these differences could potentially bias causal claims through our strategy. In Table 6.2, however, we provide estimates that confirm the continuity of each of these covariates at the cutoff, within a local average treatment effects framework, and controlling for each of the other covariates. The Portuguese education system is characterized by a relatively low number of immigrants (2.3%) and a relatively low proportion of households in which one of the parents or the guardian as at least some sort of higher education (Bachelor degree or above, 18.5%). A relatively high proportion of students also received some sort of school social support⁷ when in Grade 1 (39% for Grade 4 sample). Differences across Grade samples are explained by a steady increase in the rate of identification of students in need of social support in the most recent years of the dataset.

On the other hand, for all outcome variables and the main regressor, differences in means between the 60-days before and 60-days after cutoff samples are always statistically significant at a 1 percent confidence level. The size of these differences already help anticipate the size of some of our estimates. Students born 60-days after the cutoff have, on average more 0.65 years of age when starting school, relative to those born 60-days before the cutoff. Likewise, after-cutoff children have a .23 standard deviation points (σ) advantage in Math and .26 σ in Language by the end of Grade 4. Even if these differences tend to quickly dissipate from grade to grade, we still reject a null difference in unconditional achievement in both subjects by the end of Grade 9. Likewise, children born just after the cutoff are less likely to repeat in all of the observed grades.

⁶The choice of a 60-days window to each side of the cutoff is justified by full coverage of the optimal bandwidths used for estimation later on.

⁷School social support (ASE) in Portugal is tied to household financial constraints. Students identified as ASE have half- to fully-reduced price meals at school, textbooks and school material.

Table 5.1. Descriptive statistics - longitudinal datasets

Sample:	Full sample			60-days before cutoff			60-days after cutoff			Difference	
	Obs.	%	SD	Obs.	%	SD	Obs.	%	SD	Diff (p.p)	p-value
Student characteristics											
<i>Grade 4 sample</i>											
Female	229,661	48.61	49.98	36,220	48.56	49.98	37,145	48.40	49.98	-0.16	0.67
First generation immigrant	229,661	2.31	15.03	36,220	2.26	14.86	37,145	2.27	14.89	0.01	0.92
Access to computer at home	229,661	55.39	49.71	36,220	55.10	49.74	37,145	55.15	49.73	0.05	0.89
School social support (ASE)	229,661	38.76	48.72	36,220	39.18	48.82	37,145	39.70	48.93	0.52	0.15
Dad unemployed	229,661	6.93	25.40	36,220	7.16	25.78	37,145	7.05	25.59	-0.11	0.61
Household with higher education	229,661	21.67	41.20	36,220	21.10	40.80	37,145	21.57	41.13	0.47	0.15
<i>Grade 6 sample</i>											
Female	300,182	49.37	50.00	46,198	49.96	50.00	49,066	49.17	49.99	-0.79	0.01
First generation immigrant	300,182	2.49	15.58	46,198	2.33	15.08	49,066	2.58	15.85	0.25	0.01
Access to computer at home	300,182	46.09	49.85	46,198	45.63	49.81	49,066	46.40	49.87	0.77	0.02
School social support (ASE)	300,182	23.77	42.56	46,198	23.73	42.54	49,066	24.10	42.77	0.36	0.19
Dad unemployed	300,182	4.66	21.08	46,198	4.87	21.52	49,066	4.63	21.02	-0.24	0.12
Household with higher education	300,182	17.85	38.29	46,198	17.69	38.16	49,066	17.85	38.29	0.16	0.56
<i>Grade 9 sample</i>											
Female	188,648	51.51	49.98	28,676	52.21	49.95	31,125	51.38	49.98	-0.83	0.04
First generation immigrant	188,648	2.27	14.90	28,676	2.07	14.24	31,125	2.53	15.70	0.46	0.00
Access to computer at home	188,648	47.91	49.96	28,676	47.41	49.93	31,125	48.32	49.97	0.91	0.03
School social support (ASE)	188,648	16.24	36.89	28,676	16.28	36.92	31,125	16.33	36.97	0.05	0.87
Dad unemployed	188,648	3.67	18.81	28,676	3.64	18.73	31,125	3.72	18.93	0.08	0.63
Household with higher education	188,648	20.04	40.03	28,676	20.08	40.06	31,125	19.73	39.80	-0.35	0.32
	Obs.	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD	Diff.	p-value
Main variables											
School starting age (A)											
<i>Grade 4 sample</i>	229,661	6.25	0.31	36,220	5.97	0.39	37,145	6.62	0.14	0.65	0.00
<i>Grade 6 sample</i>	300,182	6.24	0.31	46,198	5.95	0.37	49,066	6.61	0.15	0.67	0.00
<i>Grade 9 sample</i>	188,648	6.24	0.30	28,676	5.93	0.35	31,125	6.61	0.15	0.68	0.00
Grade retention (%)											
<i>Grade 4 sample</i>	229,661	10.73	30.95	36,220	13.71	34.40	37,145	8.015	27.15	-5.70	0.00
<i>Grade 6 sample</i>	300,182	14.10	34.80	46,198	16.75	37.34	49,066	11.71	32.16	-5.04	0.00
<i>Grade 9 sample</i>	188,648	13.90	34.59	28,676	15.70	36.38	31,125	12.37	32.92	-3.33	0.00
Math performance (s.d.)											
<i>Grade 4 sample</i>	229,141	-0.05	0.99	35,906	-0.17	0.98	36,698	0.06	0.99	0.23	0.00
<i>Grade 6 sample</i>	299,156	-0.01	0.99	46,058	-0.09	0.97	48,880	0.06	1.01	0.15	0.00
<i>Grade 9 sample</i>	187,590	-0.06	0.99	28,532	-0.09	0.98	30,917	-0.04	1.00	0.05	0.00
Language performance (s.d.)											
<i>Grade 4 sample</i>	229,348	-0.05	0.99	36,173	-0.17	0.98	37,078	0.08	0.99	0.26	0.00
<i>Grade 6 sample</i>	299,663	-0.01	0.99	46,111	-0.10	0.98	48,972	0.09	1.00	0.19	0.00
<i>Grade 9 sample</i>	188,520	-0.05	0.99	28,653	-0.10	0.97	31,093	0.01	1.00	0.11	0.00

Notes: Full sample includes cohorts of students that entered regular education on a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. 60-days samples are subsamples of students born until 60 days before 1 January or 60 days after 1 January.

6. Results

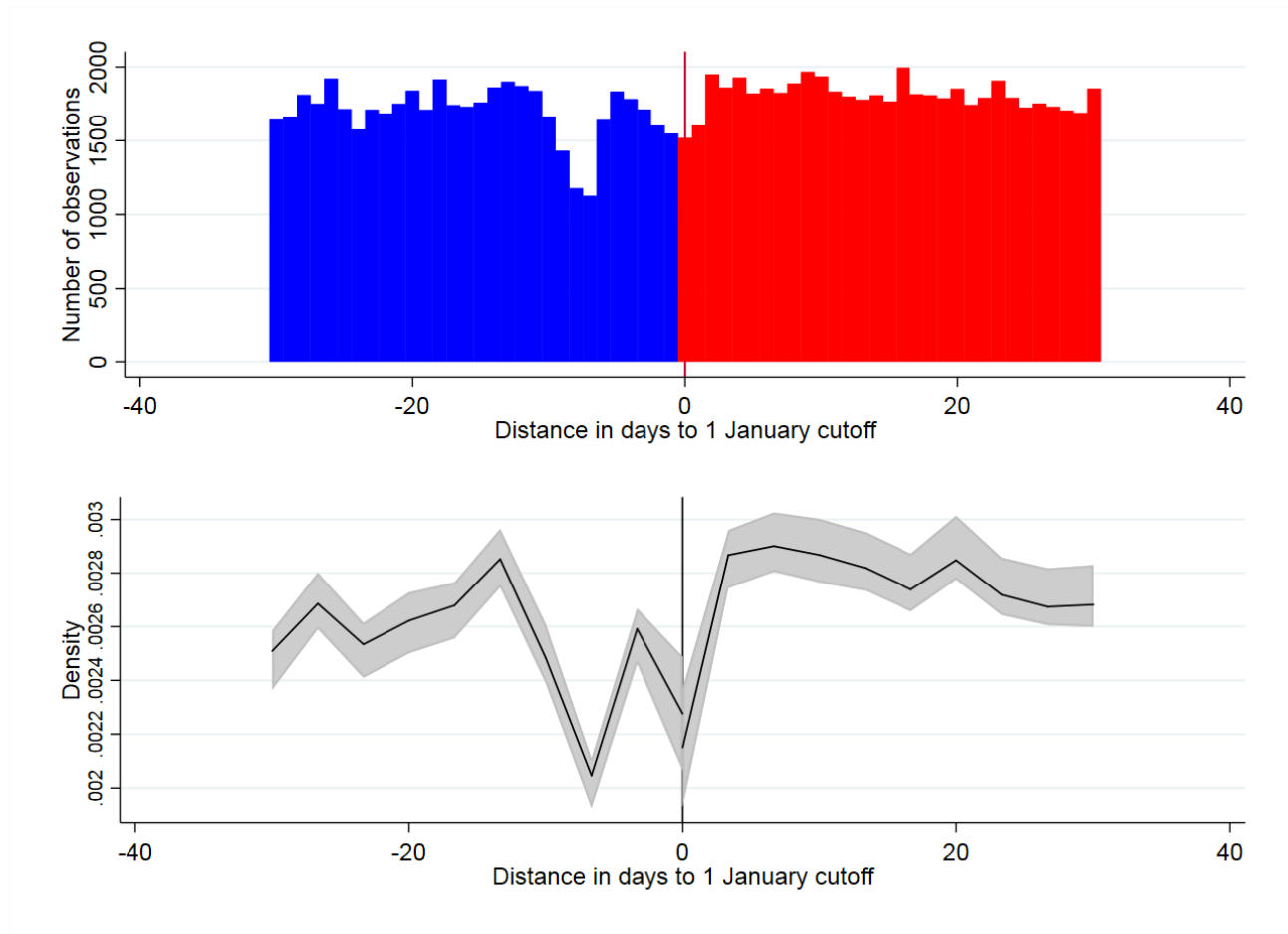
6.1. Do parents plan births strategically and how do they comply?

We begin by presenting evidence in support of our identification strategy. It could be the case that assignment to our virtual control and treatment groups is not as good as random if strategic parents – aware benefits – plan births to occur before or after the cutoff. Figure 6.1 presents in the top panel an histogram of the distribution of birth dates 30-days before and after the 1 January cutoff. Despite the downward trend in births around 7 to 10 days before the end of the year (Christmas period) the density of the running variable seems relatively continuous at the cutoff of 1 January, with no strong suspicion of parents timing births because of the cutoff. Due to the their timing, downward and upward trends seem to be driven by Christmas holidays, possibly given expected reduction in medical delivery and scheduled births for the period. The bottom panel of Figure 6.1 confirms this through a local cubic density manipulation test (Cattaneo et al., 2018; McCrary, 2008), where we can observe the shaded confidence intervals intersecting at the cutoff (p -value = 0.256, for a null hypothesis of dissimilar density 4-days before and after the cutoff)¹. Given this evidence, our virtual assignment mechanism seems to hold its’ validity.

What is the age at which students start school in Portugal? Figure 6.2 depicts the average school starting age (SSA) of students within each birth date bin, measured as the distance in days relative to the 1 January cutoff ($B_i = 0$). The solid lines are fitted values from a piecewise quadratic spline. In the figure, it can be observed a small discrete jump in school starting age at the 16 September ($B_i = -107$). As discussed in section 3, there is leeway for parents to delay their child’s SSA from this date onwards. SSA jumps slightly and, throughout the end of the calendar year, decreases at a relatively slower rate than before 16 September. A sharp discontinuity follows at the 1 January cutoff. Therefore, students born on or just after the latter cutoff average about 0.7 more years when starting school, relative to those born just before the cutoff. The fact that the difference in school starting age at the cutoff is less than 1 year between the two groups confirms imperfect compliance with the quasi-experimental design and motivates regression discontinuity estimates that take into account left-side non-compliance. About 14% of the students born on or after the 16 September defer entrance to the following school year. Figure 10.1, in Annex B reveals the declining rates of compliance for our different samples. Compliant students are those that start school on the year they turn 6-years old and do not defer entrance before reaching the 1 January cutoff. On the other hand, our strategy does

¹The relatively sharp decline in density before the cutoff occurs during the Christmas period, with parents appearing to opt-out from having children during the period. Despite making a full analysis of the parental preferences for birth dates being out of our scope, an analysis of the histogram of births across the whole year reveals clear seasonality in the data. In particular, we find that parents prefer to have children during September, 20 September being the most common birthday in our data.

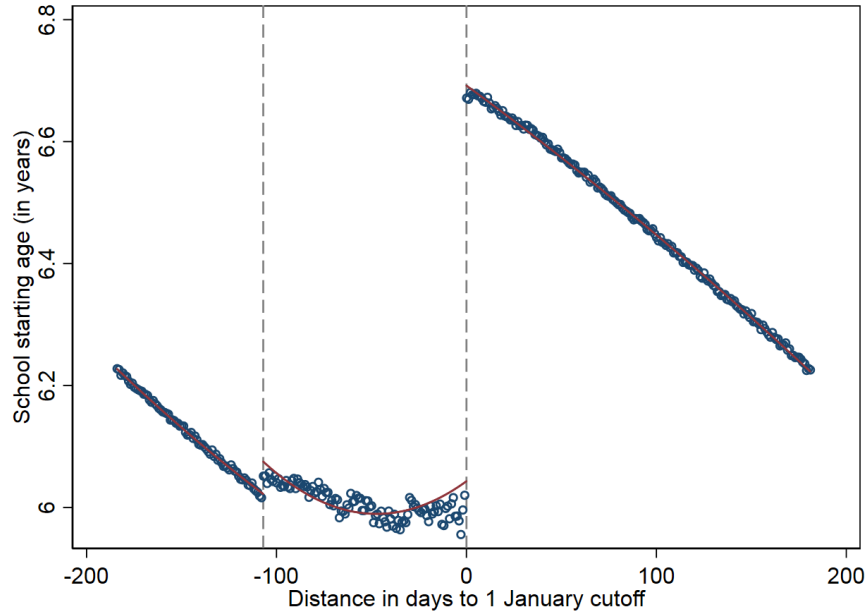
Figure 6.1: Density of birth dates close to 1 January



Notes: Top panel: Histogram of birth dates 30 days before and after the cutoff of 1 January. Bottom panel: Local polynomial density estimation at the cutoff, within a window of 30 days before and after 1 January. Density estimation performed through *rddensity* software package, described in [Cattaneo et al. \(2018\)](#). Figure is based on cohorts of students that entered a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in 1st grade.

not hold on the 16 September cutoff as compliance with such a quasi-experiment would be extremely low to render powered estimates (i.e., the vast majority of parents does not delay their child's entrance at school).

Figure 6.2: Discontinuity of school starting age at 16 September and 1 January



Notes: Figure is based on cohorts of students that entered a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in 1st grade. Horizontal axis represents the birth date relative to the cutoff date of 1 January. Each hollow circle represents within birth day cell averages of school starting age. Solid lines represent fitted values from a piecewise quadratic spline. Vertical dashed lines identify 16 September ($B_i = -107$) and 1 January ($B_i = 0$).

Table 6.1 presents point estimates of the school starting age discontinuity at the cutoff for 30- and 60-days fixed bandwidths, as well as for a changing MSE-optimal bandwidth, for each of the grade samples (per row). The results are analogous to the first stage of our two-stage fuzzy discontinuity design (Equation 4.4). The first column reports estimates for a local baseline specification, where school starting age is regressed on a quadratic function of the running variable and a triangular kernel, for a bandwidth of 30-days before and after the cutoff, separately for each grade sample. In accordance with Figure 6.2, we find a sharp and precisely estimated discontinuity at the cutoff. Column 2 reports results from a regression controlling for all student covariates presented in Table 5.1 and cohort fixed effects. Consistent with the evidence that student characteristics are continuous at the cutoff, there is no sizable change in the coefficient. For a longer 60-days bandwidth, results are more precise but tell a similar story, with no significant changes to the point estimates. Columns 5 and 6 display results for bandwidths selected through data-driven methods. Results are the most precise, with coefficients varying between 0.69 and 0.72 across specifications and samples. Local polynomial estimates thus corroborate the raw differences in means on Table 5.1: students born just after the cutoff are, on average, about 8 months older when starting school. However, among the compliant subpopulation

Table 6.1. Effect of delayed school entry eligibility at 1 January on school starting age

Outcome: school starting age	30-days bandwidth		60-days bandwidth		MSE-optimal bandwidth	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Grade 4</i>						
τ_4	0.674 (0.038)	0.670 (0.038)	0.709 (0.026)	0.706 (0.026)	0.686 (0.013)	0.685 (0.014)
Observations	36124	36124	72807	72807	113709	108656
Bandwidth (in days)	30	30	60	60	93	89
<i>Grade 6</i>						
τ_6	0.730 (0.014)	0.729 (0.014)	0.740 (0.010)	0.739 (0.011)	0.733 (0.008)	0.731 (0.008)
Observations	47125	47125	94573	94573	105815	104133
Bandwidth (in days)	30	30	60	60	66	65
<i>Grade 9</i>						
τ_9	0.737 (0.023)	0.736 (0.023)	0.748 (0.015)	0.746 (0.016)	0.742 (0.010)	0.741 (0.010)
Observations	29428	29428	59345	59345	84777	82809
Bandwidth (in days)	30	30	60	60	84	83
Polynomial order	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Student controls	NO	YES	NO	YES	NO	YES
Cohort FEs	NO	YES	NO	YES	NO	YES

Notes: All coefficients are first stage estimates of per-grade local-quadratic regressions which include a post-cutoff indicator (τ) and a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. Where indicated, a triangular kernel with a 30-day, 60-day or MSE-optimal bandwidths is used. Regressions on Columns 2, 4 and 6 also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

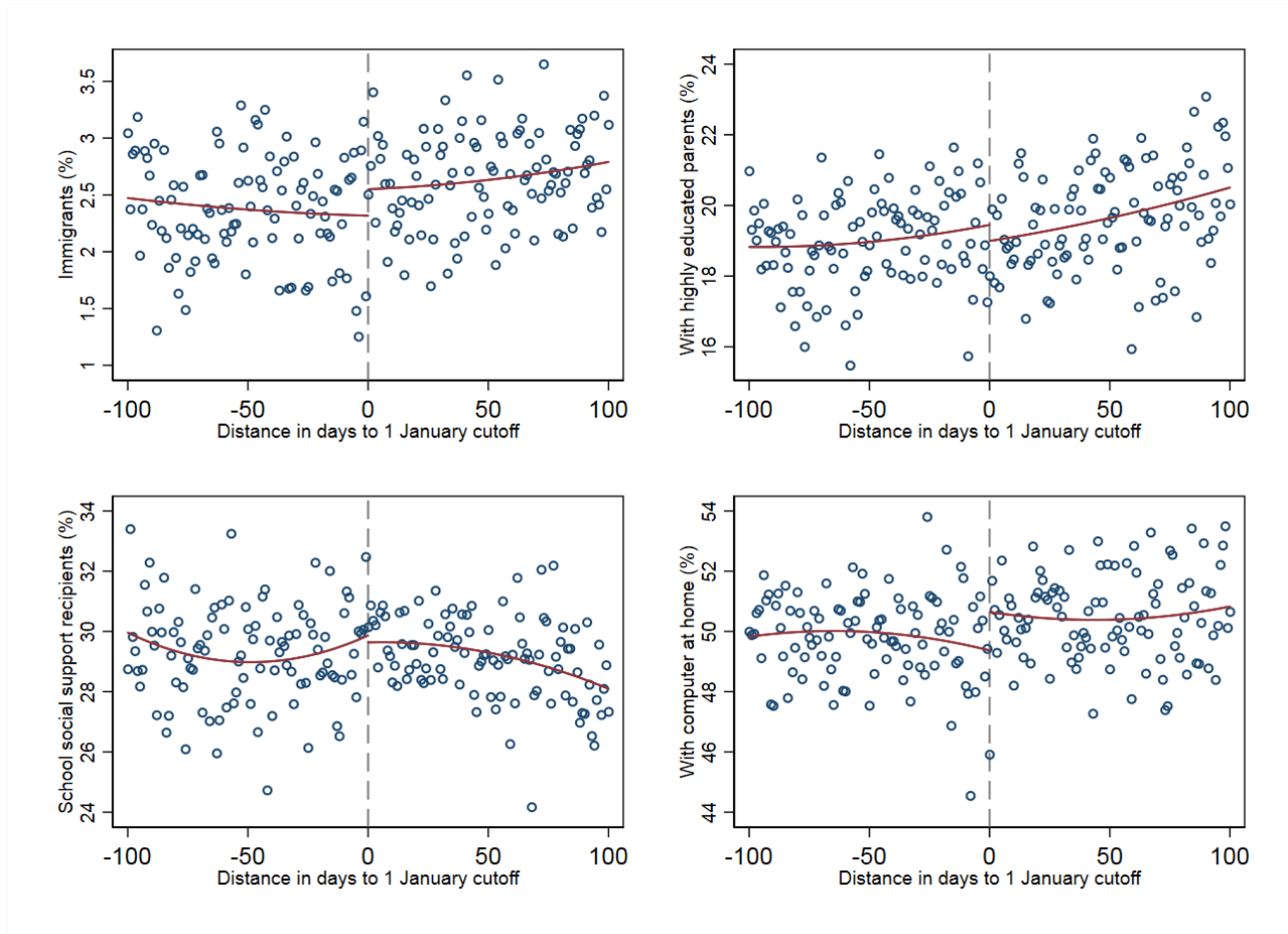
this difference will evidently be 12 months, which allows us to identify the impact of starting school 1-year later.

But is our virtual control group a reasonable counterfactual to older entrants? If the group of students used as control is not identical to the treatment group, it is not reasonable to assume that this is so. In Table 5.1 above, we presented suggestive evidence that student characteristics of those born before and after 1 January are identical close to the cutoff. The fact that point estimates in 6.1 do not change considerably after controlling for covariates further suggests that the groups are balanced – at least in terms of observable characteristics. Figure 6.3 provides the graphical representation of the per-sample distribution of student characteristics within each birth date cell, showing no apparent significant differences at the cutoff.

Despite suggestive, these are yet not sufficient to prove covariates' continuity at the cutoff. As discussed in Section 3, compulsory schooling rules in Portugal provide considerable leeway for parents to delay children's entrance for students born between 16 September and 31 December. If parents that delay students at 16 September significantly differ in their characteristics from those that opt to not defer, then if not our intent-to-treat, our local average treatment effect (LATE) estimates could be biased by compositional effects.

We test if our identification strategy survives continuity concerns along predetermined character-

Figure 6.3: Distributions of student characteristics close to 1 January



Notes: Figure is based on cohorts of students that entered a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Horizontal axis represents the birth date relative to the cutoff date of 1 January. Each hollow circle represents within birth day cell averages of immigrant status (top left panel), students with highly educated parents (top right panel), school support recipients (bottom left panel) and students with access to a computer at home (bottom right panel). Solid lines represent fitted values from a piecewise quadratic spline. Vertical dashed lines identify 1 January ($B_i = 0$).

istics of treated children by running regressions analogous to those in Equations 4.4 and 4.5. Alternatively, though, we regress each observable covariate on school starting age, having the cutoff as the excluded instrument. Table 6.2 summarizes per-sample two-stage least square estimates for local-quadratic specifications, with and without student controls, and a 60-days fixed bandwidth. We find precisely estimated null differences between complier groups, except for Grade 4 sample students that receive school social support (for $p\text{-value} < .01$). However, the qualitative interpretation of these results is unequivocal: since Grade 4 students born on or after the 1 January cutoff are likelier to be recipients of school social support – a characteristic predictive of lower achievement – our estimated LATE, if positive, will be at most underestimated relative to the true average treatment effect at the cutoff.

6.2. Impact on achievement outcomes throughout basic education

Coupled with a sharp discontinuity in SSA, we also observe sharp discontinuities in student achievement. Figure 6.4 depicts the average student achievement in Math and Language within each birth date cell, as well as each Grade sample. Fourth graders born right after the cutoff are expected to score substantially higher in each of the exams. However, differences tend to vanish as students become older.

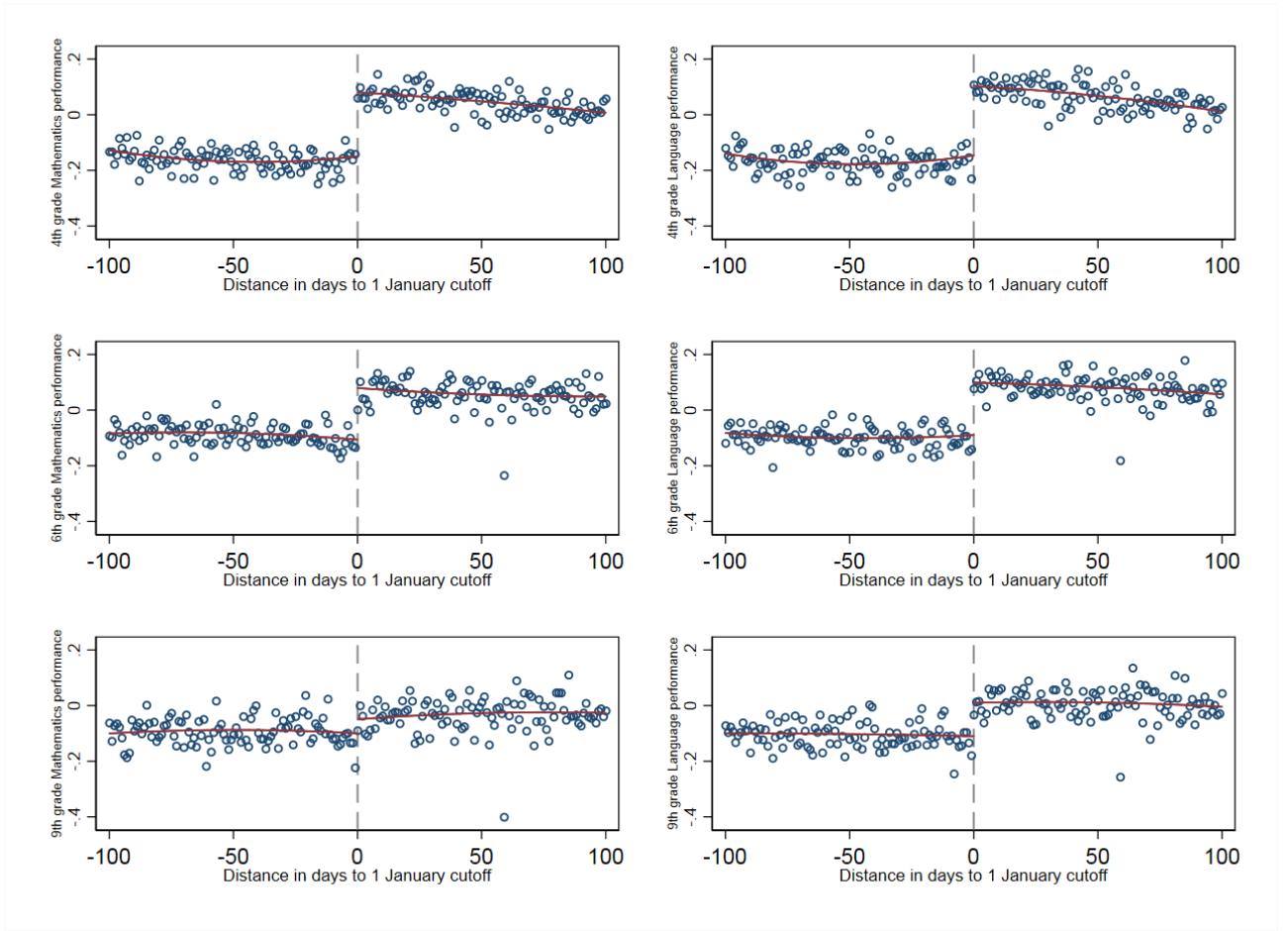
Table 6.4 presents the reduced form estimates depicted in Figure 6.4. Results originate from local-quadratic regressions, controlling for student characteristics and cohort fixed effects, applying alternative bandwidths. On average, students assigned to treatment have consistently better performance. Subject-specific differences seem to be important. By the end of Grade 9, while being assigned to treatment confers an estimated $.09\sigma$ premium in Math relative to being born just before the cutoff, in Language the difference is $.13\sigma$ in our MSE-optimal bandwidth specification.

Table 6.2. Impact of school starting age on student characteristics - Compliers

Outcome:	Female		Immigrant		School social support		Unemployed dad		Computer at home		Higer education in HH	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Grade 4</i>												
School starting age	-0.016 (0.022)	-0.016 (0.022)	0.015 (0.008)	0.015 (0.008)	0.059 (0.020)	0.057 (0.018)	-0.008 (0.013)	-0.012 (0.013)	-0.014 (0.016)	-0.006 (0.015)	-0.007 (0.018)	0.007 (0.017)
Observations	72807	72807	72807	72807	72807	72807	72807	72807	72807	72807	72807	72807
<i>Grade 6</i>												
School starting age	0.004 (0.017)	0.005 (0.017)	0.007 (0.006)	0.007 (0.006)	-0.012 (0.011)	-0.014 (0.011)	0.002 (0.005)	0.003 (0.005)	0.004 (0.021)	0.005 (0.020)	-0.008 (0.014)	-0.010 (0.013)
Observations	94573	94573	94573	94573	94573	94573	94573	94573	94573	94573	94573	94573
<i>Grade 9</i>												
School starting age	0.019 (0.022)	0.020 (0.022)	0.008 (0.008)	0.008 (0.008)	-0.024 (0.012)	-0.027 (0.011)	0.009 (0.010)	0.010 (0.010)	0.018 (0.019)	0.019 (0.019)	-0.009 (0.012)	-0.015 (0.011)
Observations	59345	59345	59345	59345	59345	59345	59345	59345	59345	59345	59345	59345
Student controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Cohort Fes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are estimates of local-quadratic regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a fixed 60-days bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as other individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes) or if there is at least one child guardian with higher education (1 if yes). Robust standard errors clustered at the birthday level are presented in parentheses.

Figure 6.4: Discontinuities on Math and Language student achievement at 1 January (per grade)



Notes: Figure is based on cohorts of students that entered a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Horizontal axis represents the birth date relative to the cutoff of 1 January. Each hollow circle represents within birth day cell averages of Math performance (left panels) and Language performance (right panels) in Grade 4 (top panels), Grade 6 (middle panel) and Grade 9 (bottom panel). Solid lines represent fitted values from a piecewise quadratic spline. Vertical dashed lines identify 1 January ($B_i = 0$).

Since estimation of intent-to-treat effects is less demanding on the panel structure of the data – as we do not actually need to observe students’ first year of school – we also compute reduced form estimation in all 829 876 Grade 9 students that sat national exams in our data. In this out-of-sample estimations, we largely reproduce the same results: gains of $.08\sigma$ in Math and about $.14\sigma$ in Language, with standard errors as small as $.01\sigma^2$. These out-of-sample findings give us confidence to assume that our in-sample reduced form estimates have limited sample or attrition biases.

Nonetheless, as estimates in Table 6.1 demonstrate, not every student assigned to the control group complies with the virtual assignment mechanism. Left-side non-compliance is thus likely downward-biasing the causal impact estimated through our reduced form model. In order to overcome this bias and better disentangle those actually treated from those merely assigned to treatment we next estimate local average treatment effects (LATE) for each of our Grade samples.

Table 6.5 summarizes LATE from regressions of SSA on Math and Language student achievement.

²Out-of-sample estimates are not reported here yet. Full results are available at request.

Table 6.4. Reduced form estimates of assignment to treatment on student performance

Outcome: student performance at	Math			Language		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Grade 4</i>						
τ_4	0.194 (0.016)	0.192 (0.015)	0.190 (0.016)	0.265 (0.039)	0.252 (0.025)	0.263 (0.017)
Observations	36042	72648	44652	36065	72693	103724
Bandwidth (in days)	30	60	37	30	60	85
<i>Grade 6</i>						
τ_6	0.141 (0.028)	0.177 (0.021)	0.186 (0.020)	0.237 (0.021)	0.223 (0.022)	0.212 (0.017)
Observations	46958	94251	88560	47043	94393	133310
Bandwidth (in days)	30	60	56	30	60	84
<i>Grade 9</i>						
τ_9	0.124 (0.041)	0.115 (0.027)	0.086 (0.022)	0.112 (0.033)	0.137 (0.025)	0.134 (0.022)
Observations	29245	58995	72086	29407	59290	67356
Bandwidth (in days)	30	60	73	30	60	67

Notes: All coefficients are estimates of local-quadratic regressions. The variable of interest is the post-cutoff indicator (τ). All regressions include a quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a 30-days, 60-days or MSE-optimal bandwidth is used. All regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

School starting age is instrumented by the cutoff and, depending on the specification, includes a piecewise linear or quadratic function of birth dates, our running variable. We also allow the estimation bandwidth to vary across specifications, according to a data-driven method, as described in Section 4.1. In accordance with the literature, we adjust standard errors for clustering within birth date cells. Column 1 presents per-grade local-linear estimations without any covariates. According to this baseline specification, entering school 1-year older entails an average benefit of 30 percent of a standard deviation in Grade 4 Math performance. When including student covariates and cohort fixed effects in the linear specification, the effect size is slightly reduced to $.29\sigma$, as precision is improved through smaller standard errors (Column 2). Columns 3 and 4 present the effects estimated through a local-quadratic specification with and without controls, respectively. Comparison with the two first columns suggests that the estimated LATE is stable across specifications, at least for Grade 4. The coefficient is again slightly smaller, but identical, albeit less precise. In the quadratic case, however, the optimal-bandwidth method allows a wider window of days used to estimate the coefficient. Columns 5-8 of Table 6.5 present analogous results for achievement in Language exams. The local-quadratic specification states a causal effect of $.39\sigma$, with clustered standard errors at the birthday-level smaller than $.04\sigma$.

As the focus turns to differences in achievement at later ages, LATE decline at a relatively fast rate, especially between the end of Grades 6 and 9. Whereas the estimated coefficient in Column 4 drops to $.25\sigma$ in Math and $.29\sigma$ in Language, by the end of Grade 9 the local-quadratic estimate is

as small as $.15\sigma$ for Math and $.19\sigma$. Compared to the other specifications, Grade 9 Math LATE are less stable across specifications and relatively less precise. Nevertheless all estimates are statistically significant at a 1 percent level.

Table 6.5. Impact of school starting age on student performance - LATE

Outcome: student performance at	Math				Language			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Grade 4</i>								
School starting age	0.300 (0.027)	0.288 (0.022)	0.277 (0.029)	0.272 (0.024)	0.385 (0.035)	0.364 (0.033)	0.367 (0.047)	0.359 (0.041)
Observations	38430	34866	50840	48282	53330	42296	69379	64468
Bandwidth (in days)	35	31	46	43	43	37	58	55
<i>Grade 6</i>								
School starting age	0.254 (0.030)	0.263 (0.025)	0.247 (0.033)	0.254 (0.028)	0.282 (0.025)	0.293 (0.023)	0.288 (0.027)	0.291 (0.023)
Observations	66075	66075	93353	90264	64514	66175	92018	131748
Bandwidth (in days)	42	41	58	56	41	41	57	83
<i>Grade 9</i>								
School starting age	0.104 (0.031)	0.094 (0.022)	0.147 (0.043)	0.153 (0.038)	0.177 (0.028)	0.193 (0.025)	0.175 (0.033)	0.189 (0.030)
Observations	40317	68110	37230	41273	45503	57803	55766	77666
Bandwidth (in days)	40	69	38	42	45	58	55	78
Polynomial order	Linear	Linear	Quadratic	Quadratic	Linear	Linear	Quadratic	Quadratic
Student controls	No	Yes	No	Yes	No	Yes	No	Yes
Cohort Fes	No	Yes	No	Yes	No	Yes	No	Yes

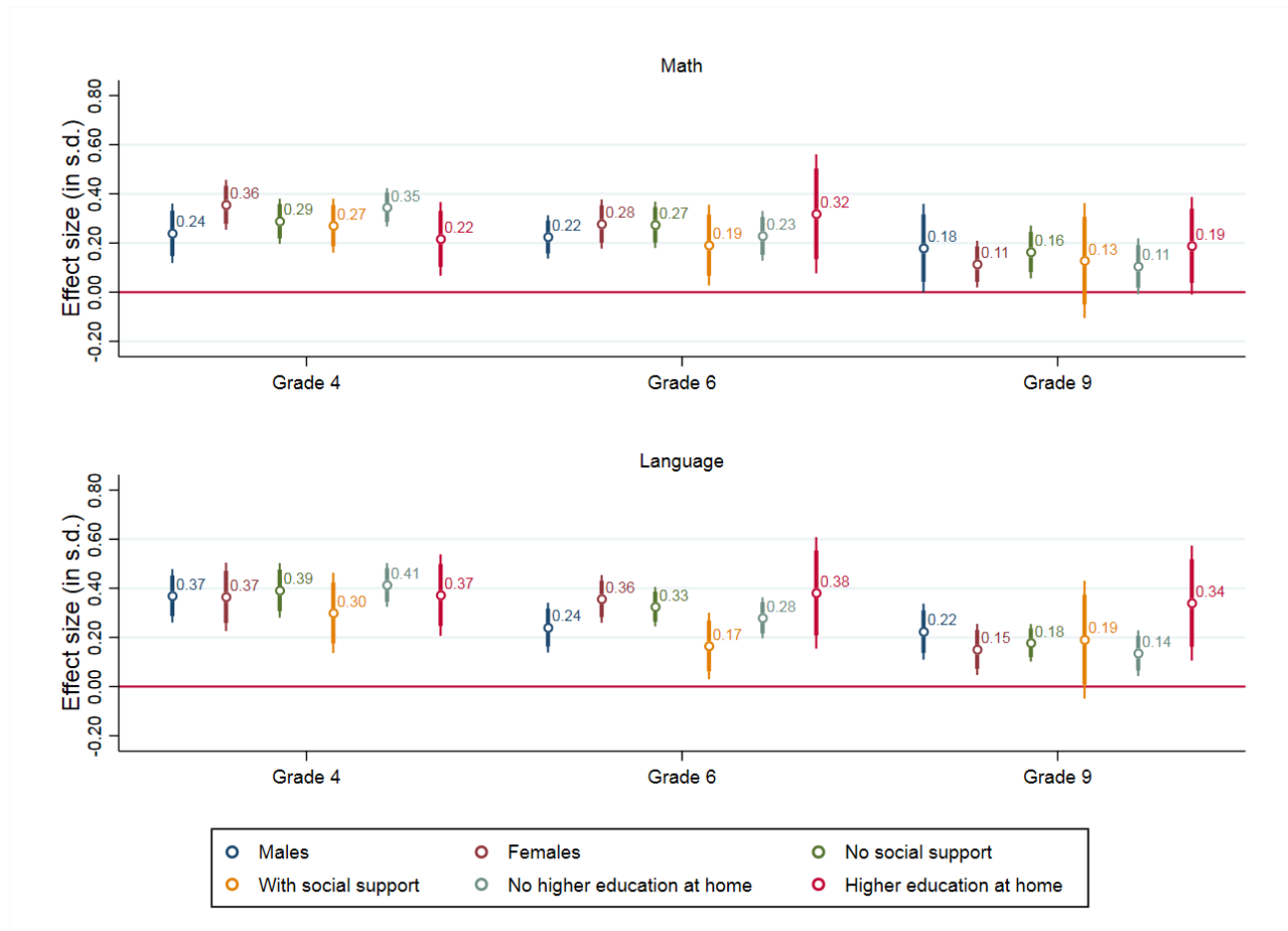
Notes: All coefficients are estimates of local polynomial regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with an indicated data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

But is the cognitive premium of being an older entrant different across students with different characteristics? We apply the same strategy to subsamples of our data in order to investigate heterogeneity. Figure 6.5 summarizes LATE estimates for different subsamples of a local-quadratic specification with student controls and cohort fixed effects, comparing it to LATE estimates in Table 6.5. Top panel and bottom panel show the effect and 95% confidence intervals on Grade 4, 6 and 9 Math and Language exam scores, respectively.

Gains in Grade 4 Math are larger for girls by $.08\sigma$ relative to boys. As comparison, the gap between boys and girls in the Math exam in our Grade 4 sample of $.13\sigma$ in favor of the former. In Language, on the other hand, being older by 1-year does not provide a relative advantage to any of the genders, as the confidence intervals almost entirely overlap. By Grade 6, the relative gains for girls in Math become a symmetric relative loss, even if not sufficient to reject a statistically null effect on the difference. In Language, the SSA effect for female students falls on the upper tail of the confidence interval of boys' SSA. Regarding Language performance, statistically significant benefits by grade 9 seem to be driven by students that are male and with from more advantage socioeconomic

backgrounds, i.e., by those that do not benefit from social support and live in households where at least one of the parents or legal guardians as an higher education degree. In summary, the results suggest that – regarding achievement – relatively well-off male students tend to benefit slightly more in Language, for entering school 1-year later.

Figure 6.5: Heterogeneous impacts of school starting age on student achievement



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each hollow circle represents a point estimate of the impact of school starting age on student performance in Grades 4, 6 and 9 Math exams (top panel) and Grades 4, 6 and 9 Language exams (bottom panel) for each indicated subsample of students. Point estimates are coefficients of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates. Thicker (thinner) vertical bars represent 95% (99%) confidence intervals for clustered standard errors at the birthday-level.

We now investigate the implications of our results in a dynamic perspective and discuss some of its policy implications. The LATE effects of entering 1-year later in school are not stable across basic education in Portugal. Others in the literature have also identified declining SSA effects throughout compulsory education (e.g. [Bedard and Dhuey, 2006](#); [Puhani and Weber, 2007](#); [McEwan and Shapiro, 2008](#); [Elder and Lubotsky, 2009](#); [Cascio and Schanzenbach, 2016](#); [Attar and Cohen-Zada, 2018](#)), a pattern that seems to extend until adulthood ([Fredriksson and Öckert, 2014](#); [Black et al., 2011](#); [Dobkin and Ferreira, 2010](#)), at least in countries where data is available.

Why does this decline occur? Delayed entrance may become less beneficial at later ages. Alternatively, or even simultaneously, decreasing gains throughout compulsory education may actually mimic the declining importance of differences in the cognitive maturity and brain development of children.

As described in section 4.2, we do not have a second source of exogenous variation to separately identify the independent effects of differential exposure to schooling from age-at-test. Nonetheless, with a minimal-assumption model of cognitive development we can likely infer the expected effect of delaying school entry, by contrasting it with the evolution of cognitive premiums as individuals age. In order to retain the maximum amount of information possible, we take reduced-forms for each grade in which we have available achievement information, including upper secondary achievement in Math and Language. Based on available estimates we then predict the cognitive premium of being born just after the cutoff by the end of Grade 1. To allow comparability, we standardize the measures of $\pi(t)$ (the proportional age difference between treatment and control subjects by age t) and the function of reduced form estimates by age to take a value of 100 by the end of Grade 1.

Figure 6.6 depicts the evolution of standardized $\pi(t)$ and the decline of estimated reduced form SSA effects as a proportion of predicted Grade 1 cognitive gain, as students age. The left panel showcases the decline in Math cognitive premium for students born just after the cutoff of 1 January. The right panel depicts the same pattern for Language performance. As is clear by the figure, the effects fall quicker than what a simple model of cognitive maturity differences would predict, such that:

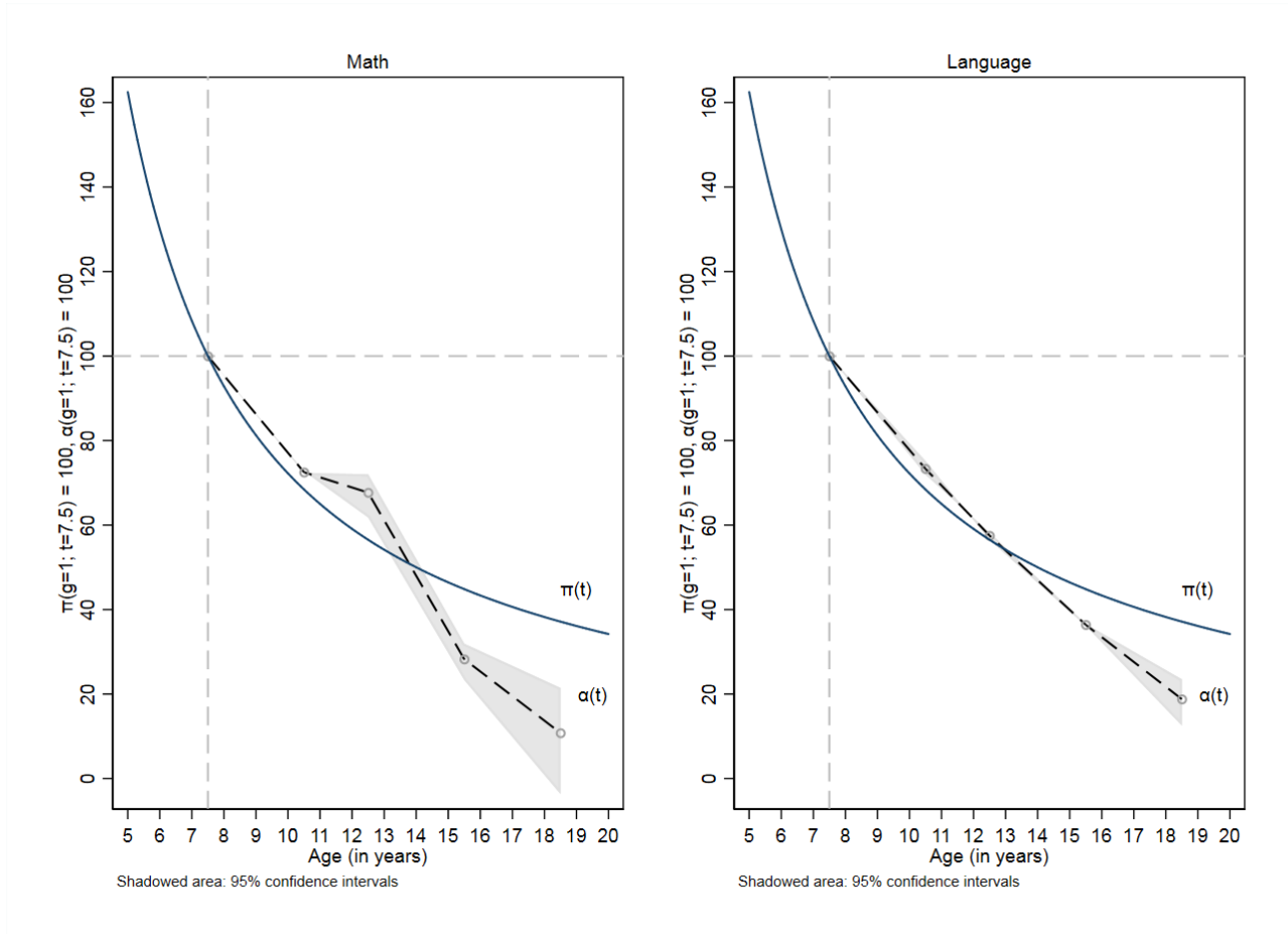
$$\frac{\partial^2 h_t}{\partial t^2} < \frac{\partial \pi(t)}{\partial t} \quad (6.1)$$

For instance, by 18.5 years of age (for a school pathway without repetitions) a student born just after the cutoff would have a cognitive premium in Math relative to students in the virtual control group of just statistically insignificant 10 percent of the predicted cognitive gain in Grade 1. On the other hand, our model predicts that – conditional on there being no effects from differential exposure to schooling – the cognitive gain should have been about 40 percent. Even restricting for students aged to be in basic education, those that complete Grade 9 by 15.5 years of age, have a cognitive advantage of just 36% of the predicted effect in Grade 1, whereas our model would predict a proportion of 44% of that effect.

Due to the rate at which cognitive gains decline as students age, we have reasons to argue – in line with plausibly causal evidence from other countries (Cornelissen and Dustmann, 2019; Peña, 2017; Black and Devereux, 2011; Elder and Lubotsky, 2009; Crawford et al., 2007) – that the orthogonal effect of enrolling later in school is *negative*.

Because ‘age-at-test’ effects are likely to dominate, these findings offer a cautionary tale to policymaking. Educational policy response – if intended at improving social well-being – may be at odds

Figure 6.6: School starting age impacts and proportional age growth



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal, were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1 and were born between 1983 and 2010. The left panel refers to reduced form estimates of achievement in Math at Grades 4, 6, 9 and 12 as the dependent variable. The right panel refers to reduced form estimates of achievement in Language at Grades 4, 6, 9 and 12 as the dependent variable. The solid line represents $\pi(t)$, standardized to have a value of 100 when students born just after the 1 January cutoff are 7.5 years and typically finishing Grade 1. $\alpha(t)$ is represented by the dashed line and shaded 95% confidence intervals, and is a function of reduced form estimates, standardized to have a value of 100 when students born just after the 1 January cutoff are 7.5 years and typically finishing Grade 1. The cognitive gain by the end of Grade 1 for these students is predicted from a quadratic trend. The regressor of interest in the reduced form specification is the post-cutoff indicator (τ). All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates presented in Table 5.1.

with the optimal choice by parents. Strategic parents will tend to respond to evidence of benefits of being relatively older through delaying enrollment of their children. For the individual child this can signify an advantage in her school success relative to her peers that may (or may not) spillover into adult outcomes. However, delayed students – by being exposed to schooling at later ages – may well be not reaching their full cognitive potential. For policy, then, evidence that school starting age leads to cognitive gains for older students should not be understood as a justification to move cutoffs to earlier in the year. Likewise, it is not reasonable to assume that, from a policy perspective, delaying enrollment is a desirable behavior to promote.

Nonetheless, human capital is accumulated within a rules-based context. As captured by equation 2.6, the policy environment in a given empirical context may set otherwise similar individuals into very different accumulation paths. Although cognitive gains fade quickly as students move through grades, the observed differences in achievement often impact on other outcomes relevant to the individual. The next two sections investigate in which other ways SSA impacts the individual.

6.3. Impact on other basic education outcomes

We now investigate the impact of SSA on outcomes that – according to our conceptual framework – may shift the function of human capital accumulation for each student. Although, as discussed above, the cognitive premium from starting school later persists throughout basic education, even if quickly dissipating as children get older. Nonetheless, cognitive capacity differences may impact on other outcomes relevant for the individual.

First, we look into grade retention. Table 6.7 describes intent-to-treat (ITT) and local average treatment effects (LATE) of starting school 1-year older on the probability of repeating a grade. As with achievement, older entrants also benefit with respect to grade retention. Columns 1-3 present ITT for different specifications, and each Grade sample. The estimates show that students born after the cutoff have a significantly lower probability of repeating a grade at least once. Restricting our attention to the compliers (Columns 4-7), the benefit becomes even larger. According to the specification in Column 7, compliant students born just after the cutoff are 5.3 percentage points (pp) less likely to have repeated at least once until Grade 4, for a sample average of 10.7 percent of repeaters (Table 5.1). Table 6.7 also shows that ITT and LATE on grade retention display a different persistence pattern throughout time than impacts on student achievement do: instead of continually decreasing, effects become slightly larger in magnitude by Grade 6 (–8pp), quickly fading by Grade 9 (–4.3pp).

Given the estimated effects on student achievement, results on grade retention are not surprising. Indeed, decisions to retain students strongly rely on the achievement of cognitive capacity of students. As older students perform better, they are also less likely to be penalized by retention.

However, are younger children more likely to repeat in spite of lower achievement, or mostly because

Table 6.7. Impact of school starting age on grade retention - ITT and LATE

Outcome: grade retention	ITT				LATE		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Until grade 4</i>							
τ_4 or School starting age	-0.027 (0.012)	-0.037 (0.009)	-0.038 (0.008)	-0.093 (0.013)	-0.060 (0.009)	-0.084 (0.018)	-0.053 (0.011)
Observations	36124	72807	76490	72020	58332	65724	69484
Bandwidth (in days)	30	60	62	58	48	54	57
<i>Until grade 6</i>							
τ_6 or School starting age	-0.052 (0.010)	-0.058 (0.009)	-0.059 (0.008)	-0.078 (0.009)	-0.080 (0.010)	-0.077 (0.010)	-0.080 (0.011)
Observations	76490	94573	90575	61467	66301	85535	93673
Bandwidth (in days)	30	60	57	38	42	53	58
<i>Until grade 9</i>							
τ_9 or School starting age	-0.023 (0.009)	-0.027 (0.008)	-0.028 (0.008)	-0.048 (0.008)	-0.047 (0.009)	-0.036 (0.010)	-0.043 (0.010)
Observations	29428	59345	48604	54750	42520	50638	63328
Bandwidth (in days)	30	60	48	54	43	51	63
Polynomial order	Linear	Quadratic	Quadratic	Linear	Linear	Quadratic	Quadratic
Student controls	Yes	No	Yes	No	Yes	No	Yes
Cohort Fes	Yes	No	Yes	No	Yes	No	Yes

Notes: All coefficients of reduced form specifications (Columns 1-3) are estimates of local-quadratic regressions. For these regressions the variable of interest is the post-cutoff indicator (τ), indicating treatment assignment. All regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a 30-days, 60-days or an MSE-optimal bandwidth is used. All coefficients are estimates of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with an indicated data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

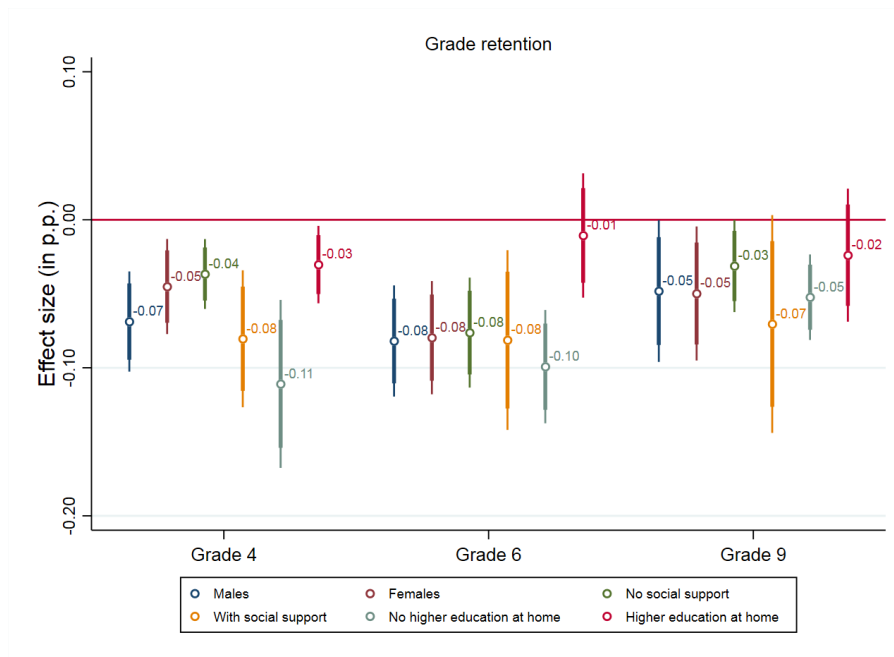
Table 6.8. Impact of school starting age on grade retention conditional on achievement - LATE

Outcome: grade retention				
	(1)	(2)	(3)	(4)
<i>Until grade 4</i>				
School starting age	-3.415 (0.869)	-2.956 (0.927)	-2.922 (1.095)	-2.413 (1.162)
Observations	60751	58241	70587	72693
Bandwidth (in days)	50	48	58	59
<i>Until grade 6</i>				
School starting age	-4.257 (0.889)	-4.118 (0.895)	-4.293 (1.024)	-3.948 (0.987)
Observations	59626	61354	82137	88687
Bandwidth (in days)	38	39	51	55
<i>Until grade 9</i>				
School starting age	-3.005 (0.897)	-3.007 (0.835)	-1.835 (0.985)	-2.867 (0.892)
Observations	39324	46523	48321	73457
Bandwidth (in days)	39	47	48	73
Polynomial order	Linear	Linear	Quadratic	Quadratic
Math achievement control	Yes	No	Yes	No
Language achievement control	No	Yes	No	Yes
Cohort FEs	Yes	Yes	Yes	Yes

Notes: All coefficients are estimates of local-quadratic regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with an indicated data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses, student achievement in Math or Language for the relevant grade.

of it? Table 6.8 shows that, even controlling for Math or Language achievement, older entrants are still significantly less likely to repeat a grade, even if the effect quickly approaches zero as students age. The results suggest that the impact of SSA on grade retention operate through other mechanisms. Younger students are thus doubly penalized, through lower achievement in Math and Language as well as a higher likelihood of retention.

Figure 6.7: Heterogeneous impacts of school starting age on likelihood of grade retention



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each hollow circle represents a point estimate of the impact of school starting age on student performance in Grades 4, 6 and 9 Math exams (top panel) and Grades 4, 6 and 9 Language exams (bottom panel) for each indicated subsample of students. Point estimates are coefficients of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates. Thicker (thinner) vertical bars represent 95% (99%) confidence intervals for clustered standard errors at the birthday-level.

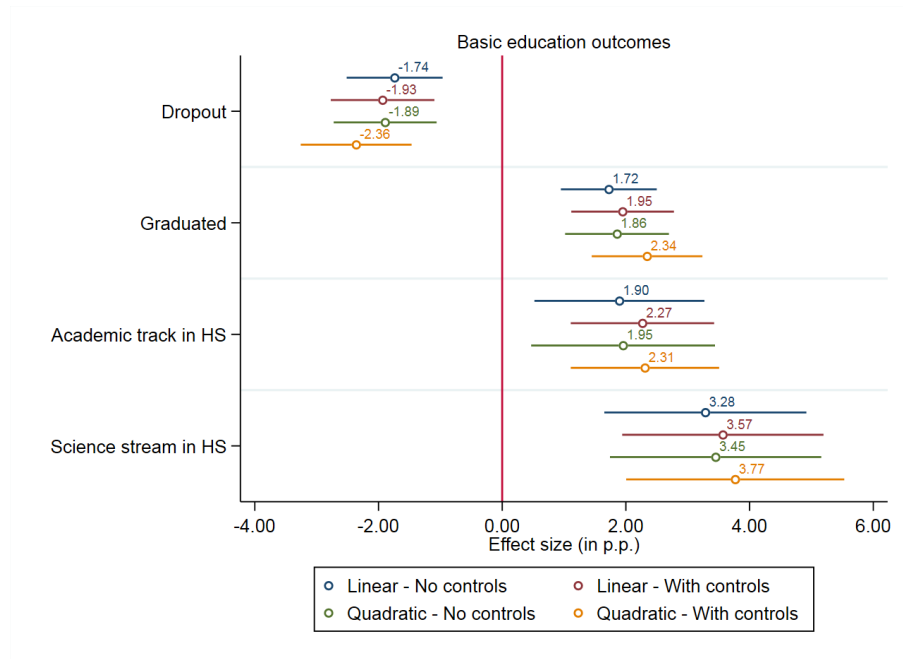
How does repetition probability change across students with different characteristics? Figure 6.7 depicts point estimates and 95% confidence intervals of the impact of SSA on the probability of grade retention in samples parsed by student observable characteristics. Reductions in repetition rates are larger for children from less educated parents (-11pp) and receivers of school social support (-8pp), whereas we cannot reject a null effect of SSA on the likelihood of repetition for children from highly educated parents in any of the grades³.

On the other hand, students from more disadvantaged backgrounds seem to benefit slightly more from delayed school entrance at the cutoff of 1 January in terms of their progression throughout basic education. These results are particularly relevant for the context of a country with high levels of

³Grade retention results are not presented here yet.

repetition and large asymmetries in retention patterns across socioeconomic groups⁴.

Figure 6.8: Heterogeneous impacts of school starting age on likelihood of grade retention



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each hollow circle represents a point estimate of the impact of school starting age on the probabilities of dropping out from school, graduating from basic education, opting for an academic track in high school and, conditional on having opted for an academic track in high school, having selected the science and technology stream. Point estimates are coefficients of local regressions, where the post-cutoff indicator (τ) is the regressor of interest. Regressions include a piecewise linear or a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates reported in Table 5.1, where indicated. Horizontal bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

Differences in school starting age also have an impact on basic education attainment outcomes. Figure 6.8 depicts reduced form estimates and 95% confidence intervals for the estimated impact of SSA on a series of attainment outcomes, across different specifications. Because our data does not allow to observe whether high school students complied with the virtual assignment mechanism, we can only restrict our interpretation to intent-to-treat, a lower bound of the true LATE. Children eligible to enter school 1-year later are less likely (–2pp) to dropout by the end of Grade 9, as well as likelier to graduate from basic education (2pp).

SSA significantly impacts enrollment patterns in high school too. Older entrants are likelier to select a general academic track in high school (2pp) – a pathway that typically leads to higher education. Figure 6.8 also shows that – conditional on having selected a general academic track in high school – students induced by the cutoff to start school 1-year later are also likelier (3.5pp) to enroll in the science and technology stream, a pathway that typically leads to STEM higher education courses.

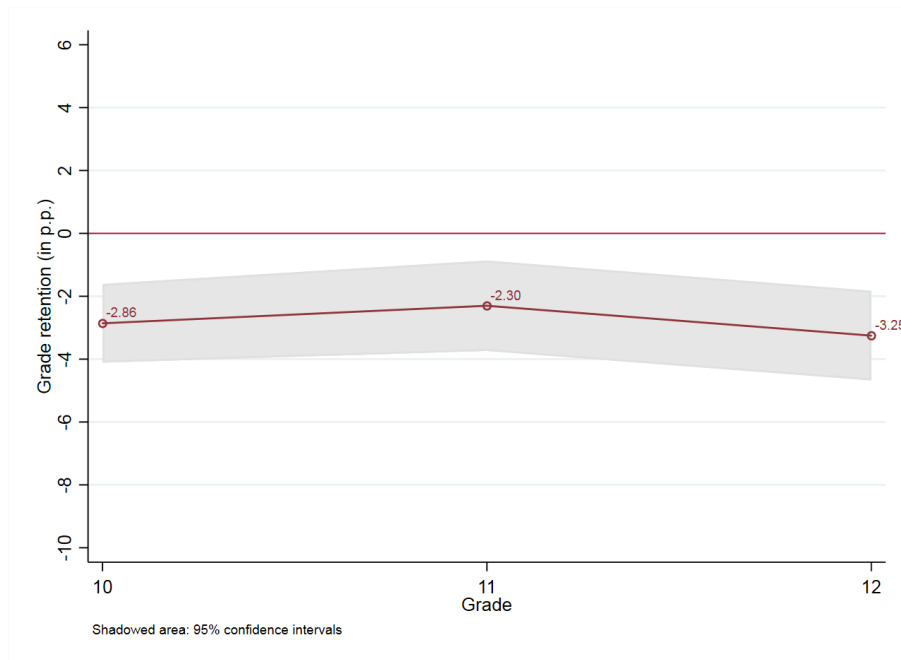
⁴For instance, in our analytical Grade 4 sample about 13% of the children born to parents without higher education are retained in grade at least once until Grade 4, while only about 1% of the children of highly educated households ever repeat a grade.

6.4. Impact on upper secondary and post-secondary outcomes

Do the effects observe in basic education persist throughout and after upper secondary graduation? **Annex C** presents evidence on achievement in national exams at both Grade 11 and Grade 12 for different subjects. Although in most subjects we do not have enough statistical power to estimate precise nulls, for most subjects achievement gains tend to be statistically insignificant. However, we can safely affirm that older entrants have higher achievement in Physics and Biology by the end of Grade 11, and Language and Biology of the order of between 5 to 10 percent of a standard deviation. Importantly, these results, albeit deriving from reduced forms, are lower bounds on the true LATE, as by upper secondary education the proportion of compliers in the control group that was not lost to grade retention, dropped out or moved to the vocational track is smaller than the same proportion in the treatment group.

Analogous to the impact on basic education outcomes, for students that have followed to upper secondary education, older entrants are less likely to be retained in grade. Figure 6.9 depicts reduced form estimates of the impact of SSA on grade retention throughout upper secondary education. Students induced to start school 1-year later due to the 1 January cutoff will be less likely to be retained in Grade.

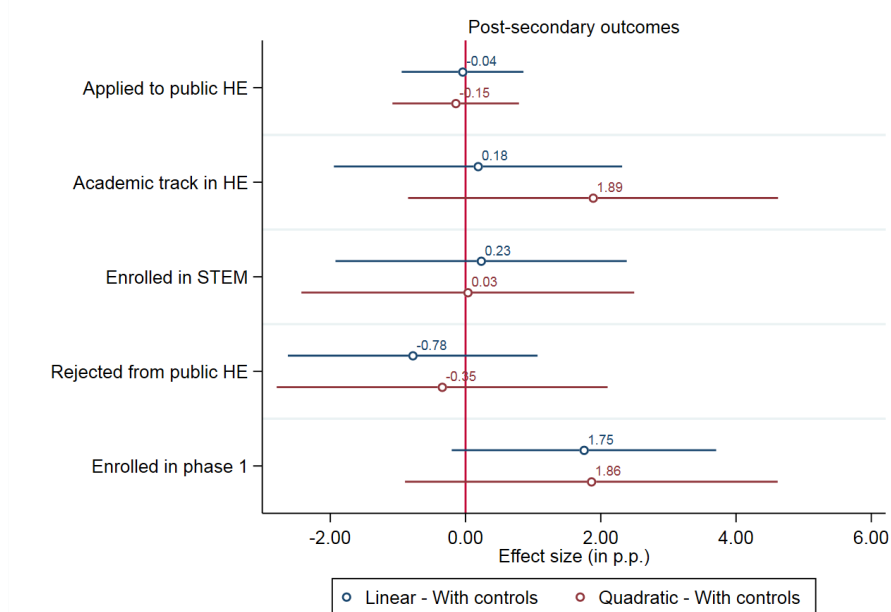
Figure 6.9: Impact of school starting age on grade retention throughout high school - ITT



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1 and enrolled in upper secondary education. Point estimates are coefficients of local regressions. The regressor of interest is the post-cutoff indicator (τ). All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates. The shaded area represents 95% confidence intervals for clustered standard errors at the birthday-level.

Do the effects of SSA persist along other margins even after high school graduation? We look into multiple outcomes on the applications of high school graduates to public colleges in Portugal. Figure 6.10 depicts reduced form estimates and their respective 95% confidence intervals for a series of binary type of outcomes. SSA does not seem to impact on college seat demand. Students born just after the 1 January cutoff are not more or less likely to apply to higher education. Although with slightly less precision, we also fail to find significant effects in rejection rates, as well as success rates in the first of the three application phases to public higher education in Portugal. Although, by the end of basic education differences in SSA seem to play a role in preferences for scientific subjects, we fail to find such differences by the end of upper secondary education. Our intent-to-treat effects suggest that older entrants are not more likely than their younger peers to be accepted into academic universities vis-à-vis polytechnic institutions. Likewise, no significant differences are uncovered for enrollment patterns in Science, Technology, Engineering and Mathematics (STEM) majors, between groups.

Figure 6.10: Impact of school starting age on demand for college seats - ITT

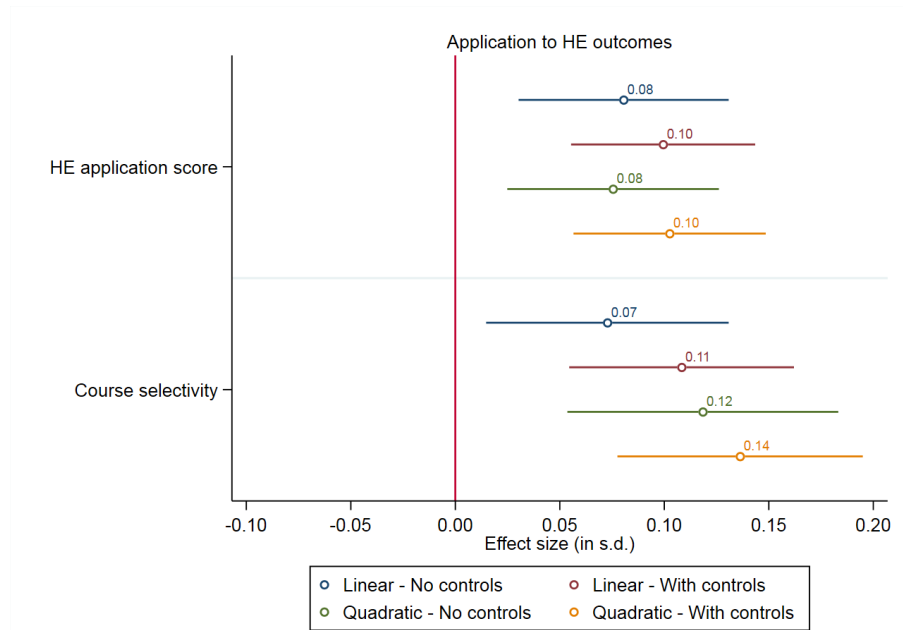


Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1 and graduated from upper secondary education. Point estimates are coefficients of local polynomial regressions. The regressor of interest is the post-cutoff indicator (τ). The dependent variables, where indicated, are dummies switched on for if the students applied to a public college, enrolled in the academic track in college, enrolled in a STEM course, failed to enroll in any course due to rejection and was enrolled in application phase 1. All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates in Table 5.1. The horizontal bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

But in which ways may older and younger college applicants be different? Figure 6.11 depicts other margins through which SSA effects may persist after high school graduation. Significantly, we find that higher education candidates born just after the 1 January cutoff have higher application scores ($.1\sigma$). Higher application scores enlarge the option set of candidates, as well as the chances of being

admitted. Evidence of this same phenomenon is the fact that, because SSA effects on the high school GPA and some national exam scores in upper secondary education are still prevalent, older entrants also enroll in more selective courses. In our most conservative point estimates, students born just after the cutoff enroll in courses $.07\sigma$ more selective than others. Our index of course selectiveness takes into account courses in each higher education institution. It is constructed from the percentile rank of the mean of the application scores of accepted applicants, the percentile rank of the standard deviation of application scores of accepted candidates and acceptance rate of applications of each course in each higher education institution.

Figure 6.11: Impact of school starting age on college application scores and index of course selectivity - ITT



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1 and graduated from upper secondary education. Point estimates are coefficients of local polynomial regressions. The regressor of interest is the post-cutoff indicator (τ). The dependent variables, where indicated, are standardized variables for college application scores and an index of course selectivity. Course selectivity is measured as an index of *i*) percentile rank of the pair higher education institution course in terms of the application scores of accepted candidates, *ii*) percentile rank of the standard deviation of application scores of accepted candidates, and *iii*) acceptance rate of applications of each course in each higher education institution. The values of the latent variable are predicted through principal factor analysis, with results being later standardized to have mean zero and standard deviation of one. All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates in Table 5.1. The horizontal bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

6.5. Robustness and placebos

An important concern about our estimated LATE in Table 6.5 is that – despite controlling for cohort fixed effects – results may be driven by students that are retained in grade, clustered just before the cutoff. As shown before repeaters are disproportionately concentrated before the enrollment cutoff, which can bias estimates by introducing compositional effects. In order to overcome such concern,

our first set of robustness checks is to restrict the main regressions to students that never repeated a grade. Table 6.9 presents results analogous to those in Table 6.5, only considering students that never repeated a grade. There are no statistically significant quantitative changes in the size of the effects, and certainly no changes in the qualitative interpretations the restricted model. For each Grade sample, even for the subset of relatively higher achieving never-repeaters, LATE estimates are relatively similar, allowing to allay concerns with significant sample and attrition biases introduced by grade retention patterns.

Table 6.9. Impact of school starting age on student performance for non-repeaters

Outcome: student performance at	Math				Language			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Grade 4</i>								
School starting age	0.251 (0.029)	0.280 (0.025)	0.245 (0.034)	0.275 (0.029)	0.342 (0.029)	0.370 (0.022)	0.325 (0.051)	0.356 (0.045)
Observations	31138	30064	45349	44192	63021	78135	57492	60766
Bandwidth (in days)	29	28	41	41	57	72	53	56
<i>Grade 6</i>								
School starting age	0.195 (0.031)	0.218 (0.024)	0.179 (0.035)	0.206 (0.026)	0.235 (0.029)	0.255 (0.025)	0.242 (0.031)	0.261 (0.027)
Observations	58094	56681	67816	73176	59572	60964	86475	86475
Bandwidth (in days)	43	42	49	54	43	45	64	64
<i>Grade 9</i>								
School starting age	0.073 (0.035)	0.105 (0.032)	0.150 (0.045)	0.152 (0.039)	0.170 (0.032)	0.182 (0.030)	0.180 (0.034)	0.187 (0.031)
Observations	38128	32081	32991	36401	34910	40944	51910	71246
Bandwidth (in days)	44	38	38	43	41	47	61	83
Polynomial order	Linear	Linear	Quadratic	Quadratic	Linear	Linear	Quadratic	Quadratic
Student controls	NO	YES	NO	YES	NO	YES	NO	YES
Cohort Fes	NO	YES	YES	YES	NO	YES	YES	YES

Notes: All coefficients are estimates of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

A second concern is with patterns in birth dates reflecting parental characteristics that are not perceivable solely by inspecting the distribution of births across the calendar year. As in other countries, scheduled birth-giving and hospital service adjustments cause the frequency of births to decrease during weekends in Portugal. If the enrollment cutoff coincidentally falls close to weekends then differences at the cutoff may introduce some correlation with the characteristics of parents. However, controlling for weekday indicators produces no changes in the point estimates, suggesting this is not a problem in our analysis (see Table 12.1 in Annex D).

A third source of concern relates to the method itself. Since local regression estimates are sensitive to the choice of bandwidth, the optimal data-driven bandwidth could be systematically biasing our results. The choice of bandwidth typically entails contemplating a trade-off: opting for a larger

bandwidth includes more valid observations and increases precision, however if it is too wide our local specification may not be adequate. Table 6.10 presents a sensitivity analysis, replicating results for our preferred specification with alternative bandwidths. Columns 1, 4 and 7 show coefficients for a 30-day bandwidth on both sides of the cutoff. Columns 2, 5 and 8 present the same local-quadratic regression results for a larger 60-day bandwidth. Columns 3, 6 and 9 show results from employing an alternative optimal bandwidth selector, allowing bandwidths to the left and right of the cutoff differ from each other (Calonico et al., 2018)⁵. Reassuringly, both point estimates and standard errors are stable across alternative bandwidths and identical to the ones presented in Tables 6.5 and 6.7⁶. Our estimated LATE are thus relatively robust to different bandwidth selection methods.

⁵Optimal bandwidth selectors are deployed through the `rdrobust` software package, developed in Calonico et al. (2014a) and Calonico et al. (2017).

⁶Despite not reported, all first-stage and reduced form estimates are also identical to those presented in Tables 6.1, 6.4 and 6.7.

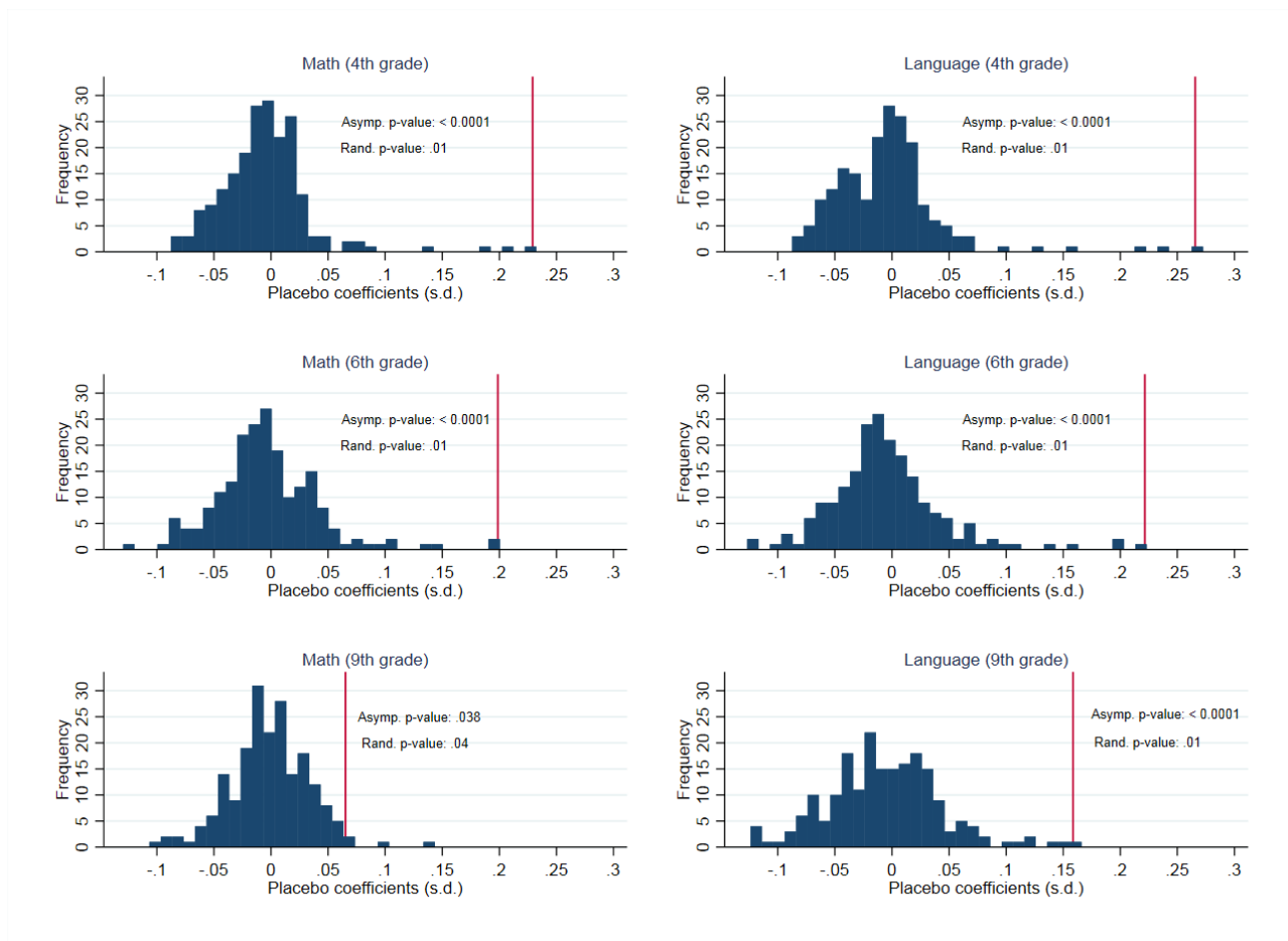
Table 6.10. Sensitivity analysis: impact of school starting age on student outcomes for alternative bandwidths

Outcome:	Math performance		Language performance		Grade retention	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>(Until) Grade 4</i>						
School starting age	0.289 (0.022)	0.274 (0.024)	0.364 (0.034)	0.360 (0.040)	-0.060 (0.009)	-0.054 (0.011)
Observations	34866	53287	41015	65624	57032	66960
Bandwidth (in days)	29	44	34	53	47	54
<i>(Until) Grade 6</i>						
School starting age	0.262 (0.025)	0.255 (0.028)	0.294 (0.023)	0.291 (0.024)	-0.081 (0.010)	-0.079 (0.011)
Observations	64417	100530	61354	115385	61467	93673
Bandwidth (in days)	41	64	38	72	39	58
<i>(Until) Grade 9</i>						
School starting age	0.095 (0.024)	0.150 (0.038)	0.192 (0.024)	0.179 (0.032)	-0.047 (0.009)	-0.048 (0.010)
Observations	56529	39324	66312	69474	43488	76707
Bandwidth (in days)	57	39	66	70	43	76
Polynomial order	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic

Notes: All coefficients are estimates of local-quadratic regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. As indicated, a triangular kernel with a 30-days, 60-days and a data-driven MSE-optimal choice of bandwidth, allowing bandwidths before and after the cutoff to differ, are used. All regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

Finally, we examine if our reduced form results are not an artifact of our polynomial specifications. Calonico et al. (2014b) show that in the presence of misspecification, computed standard errors can uncover actually spurious effects. In order to test alternative p -values we run permutation tests in the spirit of Fisher (1935) and according to recently developed methods (Ganong and Jäger, 2018). In particular, we assume that the cutoff is drawn from a random distribution of 200 potential cutoffs. For each of the placebo cutoffs and the true cutoff we estimate intent-to-treat effects, as LATE would give us a non-negligible portion of meaningless under-powered estimates for most placebo cutoffs. We then compute a randomization-based p -value based on the distribution of these estimates. The strategy allows us to assess what is the likelihood of the true intent-to-treat effect being due to chance.

Figure 6.12: Distributions of placebo and true intent-to-treat effects by subject and grade



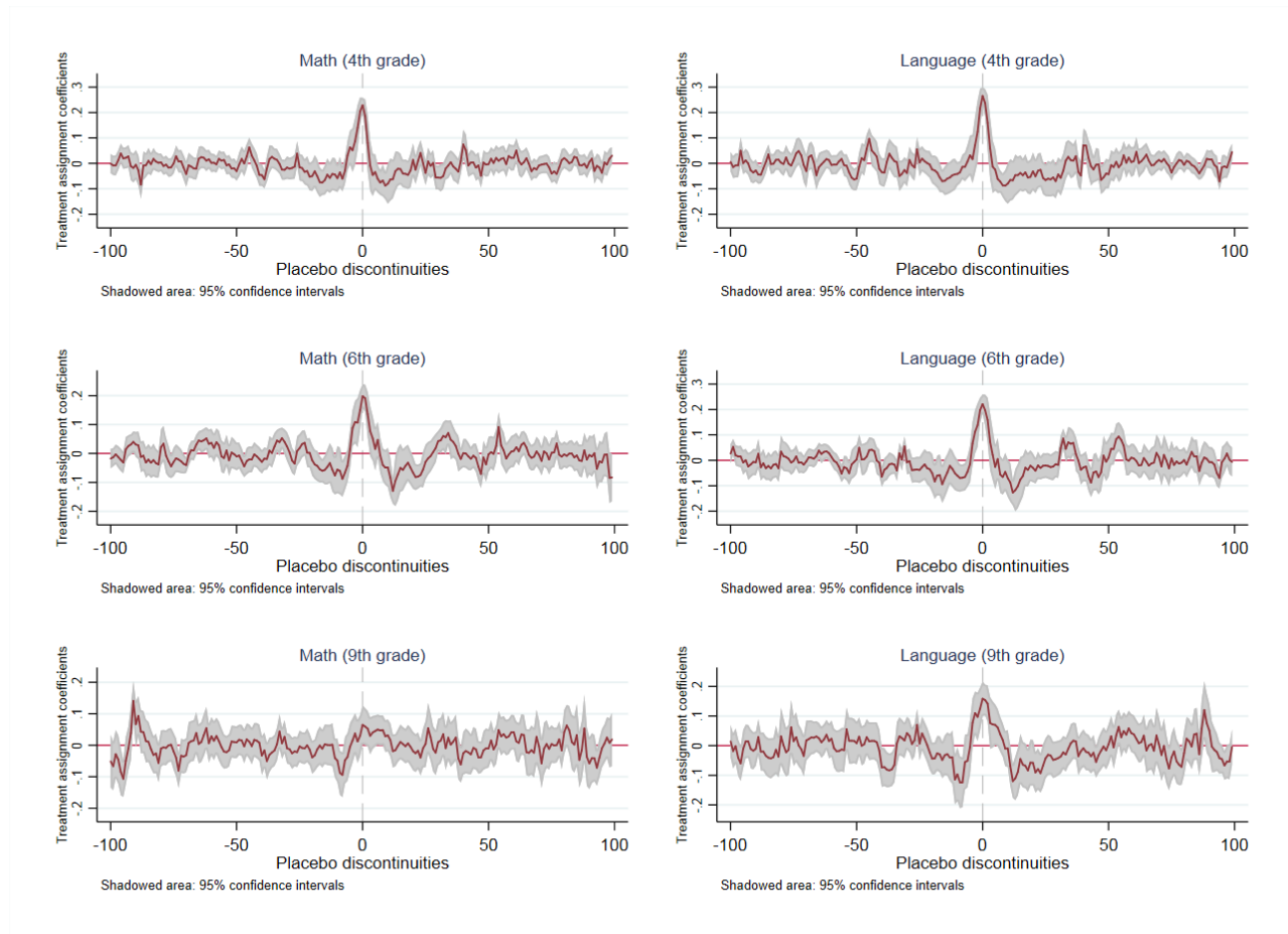
Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel is an histogram of estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one). Cutoffs considered are all where $B_i = \{-100, 99\}$. For each placebo cutoff the specification estimated is the one in Column 3 (for Math) and Column 6 (for Language) in Table 6.4, with varying MSE-optimal bandwidths. The vertical red line represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs. Randomization-based p -values, computed through the software package `rdpermut` and according to Ganong and Jäger (2018), are presented under the asymptotic p -values for the preferred specification at the true cutoff.

Figure 6.12 shows that – through this method – our estimates are unlikely due to chance. Except for Grade 9 Math, for which we cannot safely exclude a null effect, all true estimates are at the end

of the right-tail of the placebo distributions and are in line with the asymptotic p -values returned by our estimations. Figure 13.1, in Annex E does the same exercise for the specifications in which grade retention is the outcome of interest.

Figure 6.13 presents the estimated coefficients with the respective 95% confidence intervals for each of the placebo and true cutoffs. As it is also perceivable through this depiction, the coefficients at the true cutoff (represented by the gray dash line), except at Grade 9 Math, are a clear outlier among the placebos (spread out along each of the x -axis).

Figure 6.13: Intent-to-treat effects by potential cutoff, subject and grade



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel shows estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one) along the y -axis. Cutoffs considered are all where $B_i = \{-100, 99\}$, spread along the x -axis. For each placebo cutoff the specification estimated is the one in Column 3 (for Math) and Column 6 (for Language) in Table 6.4, with varying MSE-optimal bandwidths. The vertical dashed line represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs.

7. Conclusion

An exogenous one-year variation in school starting age has significant effects on primary level student outcomes. Students that are induced to delay enrollment in first grade for one-year improve student performance in 4th grade national exams by about 0.3 standard deviations (σ) in Math and almost 0.38σ in Language in Portugal¹. Heterogeneity across groups is also limited. Students from more disadvantaged backgrounds – i.e. that receive school social support and have less educated parents – seem to benefit slightly more in terms of achievement in Math. Older girls also benefit slightly more than older boys. In any case, delayed entrance is homogeneously beneficial to students across identified socioeconomic groups, with overlapping confidence intervals precluding us from taking further conclusions about these patterns. Importantly, we find that the cognitive premium by the end of elementary education persists across all groups, but quickly fades throughout lower and upper secondary education.

Through which the causal mechanism do our local average treatment effects fade as students age. Since we do not have a second source of exogenous variation, we cannot separate the ‘age-at-test’ effects, reflecting cognitive maturity differences, from differential ‘exposure to schooling’ effects. Our results are thus best interpreted as absolute age effects. If underlying causal mechanisms in Portugal are no different than in other contexts (e.g. Crawford et al., 2007; Black and Devereux, 2011; Fredriksson and Öckert, 2014; Peña, 2017; Cornelissen and Dustmann, 2019), ‘age-at-test’ effects may tend to dominate and lead to null or even negative impacts of delayed school entrance in the long-run. A simple model of human capital accumulation gives us some confidence that this is also the case in the Portuguese education system. By comparing the rate at which our marginal impacts diminish as students age, our interpretation is that absolute age differences in outcomes are being mostly driven by ‘age-at-test’ effects.

However, certain institutional mechanisms make school starting age matter to the individual through other margins besides measurable achievement differences. Students that enter school one-year later are less likely to repeat a grade, a pattern that persists well into upper secondary education. Conditional on achievement – arguably the most determinant factor in retention decisions – older entrants still are less likely to repeat. Likewise, we find that older entrants are less likely to dropout and more likely to complete basic education. SSA may be important in yet other ways. Students

¹The local average treatment effect of school starting age on test scores here presented are consistent with comparable evidence from other countries. Cross-country evidence from Bedard and Dhuey (2006) find effect sizes ranging from 0.2σ to 0.5σ . McEwan and Shapiro (2008) – with an identical empirical strategy, but different estimation procedure – find increases of 4th grade test scores of 0.3 - 0.4σ , in Chile. For fifth grade students in Israel, Attar and Cohen-Zada (2018) estimate somewhat smaller 0.29σ gains for Language, and 0.16σ for Math. For fourth graders in Germany, Puhani and Weber (2007) find an increase of 0.4σ .

predicted to be older entrants into school – even if being exposed to schooling later – are more likely to enroll in a general academic track in high school and, conditional on it, to opt for high school concentrations dominated by scientific courses.

Our intent-to-treat effects also show that older students have higher application scores to access public higher education (0.1 s.d.) and enroll in more selective undergraduate courses (0.12 s.d.). However, we find no evidence of differences on the demand for college seats, enrollment in STEM courses, or first-choice application success.

But how relevant are our findings for policy or individual decision-making? Our empirical strategy – besides evidence that birth dates are not manipulated around the cutoff and that covariates are balanced independently of treatment assignment – gives us confidence that our LATE and ITT are at least internally valid. However, it is well known that RD estimates are local to the cutoff and that direct extrapolation requires relatively strong assumptions about the homogeneity of treatment (Imbens and Lemieux, 2008). In our empirical context, it would be relevant to attest if treatment can be extended to a period during the calendar year in which parents have more leeway to defer student entrance – in particular during the period in which the Portuguese law allows for students to be easily delayed (16 September - 31 December). However, as shown in section 6.1, the cutoff at 16 September does not predict a very large jump in school starting age, with selection bias and low percentage of compliers rendering local estimates at that cutoff unreliable. We can affirm that, due to their reliability, our estimates fall with a high degree of confidence within a short interval of true estimates of the causal effects of delayed school entrance in regions of the running variable (in this case, birth dates) where parents can more easily choose to delay children. In this sense, parents can be relatively safeguarded that delaying school entrance of their children within reasonably close distance of the cutoff – even if not exactly at it – will, on average, yield the described benefits.

Nevertheless, choice prescription needs be nuanced. Even if restricting our attention to short-run benefits, the policy response – if intended at improving social well-being – may be at odds with the optimal choice by parents. Strategic parents will tend to respond to evidence of benefits of being relatively older through delaying enrollment of their children. For the individual child this can signify an advantage in her school success that may (or may not) spillover into adult outcomes. However, variance-increasing effects of delaying entrance on social welfare may lead to relatively more unequal outcomes across children from different backgrounds. In the case of Portugal, if strategic parents become more responsive to evidence of relative gains to older children, this may lead to an advantage that can be perceived as unfair to those constrained in their choice. As the legal option to defer entrance is mostly granted to those whose children are born between the 16 September and 31 December, this leads to an unequal distribution of choice. Moreover, if even for parents of conditional children access

to information and good professional judgment is unequally distributed, early enrollees – i.e. those that do not delay school entrance – may be disproportionately penalized. Taking our evidence at face-value, early enrollees will have lower achievement and will also be more likely to repeat at least once during primary education, offsetting the potential future gain of entering one-year earlier into the labor market.

On the other hand, if school capacity constraints force parents to delay children entrance into formal schooling – another mechanism through which students may be delayed in the Portuguese education system – this too may have unintended consequences for ensuring equal opportunities. Children who start school a year later will typically remain in pre-school environments whose quality for learning will be more strongly correlated with family background. Many have been arguing that, insofar pre-schooling conditions are unequal, delaying public schooling may well reproduce and amplify such initial conditions (Deming and Dynarski, 2008). Both parents and policymakers should thus appropriately weigh costs associated with an additional year of childcare outside formal schooling environments and shorter work careers. Alternatively to changing policy in school entry laws, policymakers can also consider other ex ante measures, namely early childhood interventions aimed at addressing school readiness gaps across children from different socioeconomic groups.

8. References

- Altwickler-Hámori, S. and J. Köllö (2012, jul). Whose children gain from starting school later? - evidence from Hungary. *Educational Research and Evaluation* 18(5), 459–488.
- Attar, I. and D. Cohen-Zada (2018). The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth. *Journal of Economic Behavior and Organization* 153(10568), 38–57.
- Bedard, K. and E. Dhuey (2006, nov). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics* 121(4), 1437–1472.
- Black, S. E. and P. J. Devereux (2011). Recent Developments in Intergenerational Mobility. In *Handbook of Labor Economics (Volume 4, Part B)*, pp. 1487–1541.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2011, may). Too Young to Leave the Nest? The Effects of School Starting Age. *Review of Economics and Statistics* 93(2), 455–467.
- Buckles, K. S. and D. M. Hungerman (2013, jul). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics* 95(3), 711–724.
- Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018). Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. *Unpublished*.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17(2), 372–404.

- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust data-driven inference in the regression-discontinuity design. *Stata Journal* 14(4), 909–946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6), 2295–2326.
- Carlsson, M., G. B. Dahl, B. Öckert, and D. O. Rooth (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics* 97(3), 533–547.
- Cascio, E. U. and D. W. Schanzenbach (2016, jul). First in the Class? Age and the Education Production Function. *Education Finance and Policy* 11(3), 225–250.
- Cattaneo, M. D., A. Arbor, M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal* 18(1), 234–261.
- Cook, P. and S. Kang (2018). The School-Entry-Age Rule Affects Redshirting Patterns and Resulting Disparities in Achievement.
- Cook, P. J. and S. Kang (2016). Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *American Economic Journal: Applied Economics* 8(1), 33–57.
- Cornelissen, T. and C. Dustmann (2019). Early School Exposure, Test Scores, and Noncognitive Outcomes. *American Economic Journal: Economic Policy* 11(2), 35–63.
- Crawford, C., L. Dearden, and C. Meghir (2007). When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England.
- Crawford, V. P. and N. Iriberri (2007). Level-K Auctions : Can a Nonequilibrium Model of Strategic Thinking Explain the Winner ’ s Curse in Private-Value Auctions? *Econometrica* 75(6), 1721–1770.
- Deming, D. and S. Dynarski (2008). The Lengthening of Childhood. *Journal of Economic Perspectives* 22(3), 71–92.
- Dhuey, E., D. Figlio, K. Karbownik, and J. Roth (2017). Age and Cognitive Development.
- Dhuey, E. and S. Lipscomb (2008, apr). What makes a leader? Relative age and high school leadership. *Economics of Education Review* 27(2), 173–183.
- Dhuey, E. and S. Lipscomb (2010, oct). Disabled or young? Relative age and special education diagnoses in schools. *Economics of Education Review* 29(5), 857–872.
- Dobkin, C. and F. Ferreira (2010, feb). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review* 29(1), 40–54.
- Du, Q., H. Gao, and M. D. Levi (2012, dec). The relative-age effect and career success: Evidence from corporate CEOs. *Economics Letters* 117(3), 660–662.
- Elder, T. E. and D. H. Lubotsky (2009). Kindergarten Entrance Age and Children’s Achievement:

- Impacts of State Policies, Family Background, and Peers. *The Journal of Human Resources* 44(3), 641–683.
- Evans, W. N., M. S. Morrill, and S. T. Parente (2010). Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children. *Journal of Health Economics* 29(5), 657–673.
- Fertig, M. and J. Kluve (2005). The Effect of Age at School Entry on Educational Attainment in Germany. *IZA Discussion Paper No.1507*.
- Fisher, R. (1935). *The Design of Experiments*. Oxford: Oliver and Boyd.
- Fredriksson, P. and B. Öckert (2014, sep). Life-cycle effects of age at school start. *Economic Journal* 124(579), 977–1004.
- Ganong, P. and S. Jäger (2018). A Permutation Test for the Regression Kink Design. *Journal of the American Statistical Association* 113(522), 494–504.
- Gelman, A. and G. Imbens (2017, aug). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, Published online.
- Hahn, J., P. Todd, and W. Van Der Klaauw (2001, jan). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Herbst, M. and P. Strawiński (2016). Early effects of an early start: Evidence from lowering the school starting age in Poland. *Journal of Policy Modeling* 38(2), 256–271.
- Imbens, G. and K. Kalyanaraman (2012, jul). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3), 933–959.
- Imbens, G. W. and J. D. Angrist (1994, mar). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2), 467.
- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.
- Kawaguchi, D. (2011, jun). Actual age at school entry, educational outcomes, and earnings. *Journal of the Japanese and International Economies* 25(2), 64–80.
- Landersø, R., H. S. Nielsen, and M. Simonsen (2017, jun). School Starting Age and the Crime-age Profile. *Economic Journal* 127(602), 1096–1118.
- Lee, D. S. and T. Lemieux (2010, jun). Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2), 281–355.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2), 698–714.
- McEwan, P. J. and J. S. Shapiro (2008). The Benefits of Delayed Primary School Enrollment: Dis-

- continuity Estimates Using Exact Birth Dates The Benefits of Delayed Primary School Enrollment Discontinuity Estimates Using Exact Birth Dates. *The Journal of Human Resources* 43(1), 1–29.
- Mühlenweg, A., D. Blomeyer, H. Stichnoth, and M. Laucht (2012, jun). Effects of age at school entry (ASE) on the development of non-cognitive skills: Evidence from psychometric data. *Economics of Education Review* 31(3), 68–76.
- Mühlenweg, A. M. and P. A. Puhani (2010a). The Evolution of the School-Entry Age Effect in a School Tracking System. *Journal of Human Resources* 45(2), 407–438.
- Mühlenweg, A. M. and P. A. Puhani (2010b, mar). The Evolution of the School-Entry Age Effect in a School Tracking System. *Journal of Human Resources* 45(2), 407–438.
- Peña, P. A. (2017, feb). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review* 56, 152–176.
- 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 (2007, may). Does the early bird catch the worm? *Empirical Economics* 32(2-3), 359–386.
- Schneeweis, N. and M. Zweimüller (2014, apr). Early Tracking and the Misfortune of Being Young. *The Scandinavian Journal of Economics* 116(2), 394–428.
- Smith, J. (2009, jan). Can regression discontinuity help answer an age-old question in education? the effect of age on elementary and secondary school achievement. *B.E. Journal of Economic Analysis and Policy* 9(1).
- Suziedelyte, A. and A. Zhu (2015, apr). Does early schooling narrow outcome gaps for advantaged and disadvantaged children? *Economics of Education Review* 45, 76–88.
- Zhang, S., R. Zhong, and J. Zhang (2017). School starting age and academic achievement: Evidence from China’s junior high schools. *China Economic Review* 44(2016), 343–354.

9. Annex A: Data treatment

9.1. Attrition in underlying data

Figure 9.1: Panel attrition rates and number of tracked students by entry cohort (Grade 4 Sample, LATE Dataset)

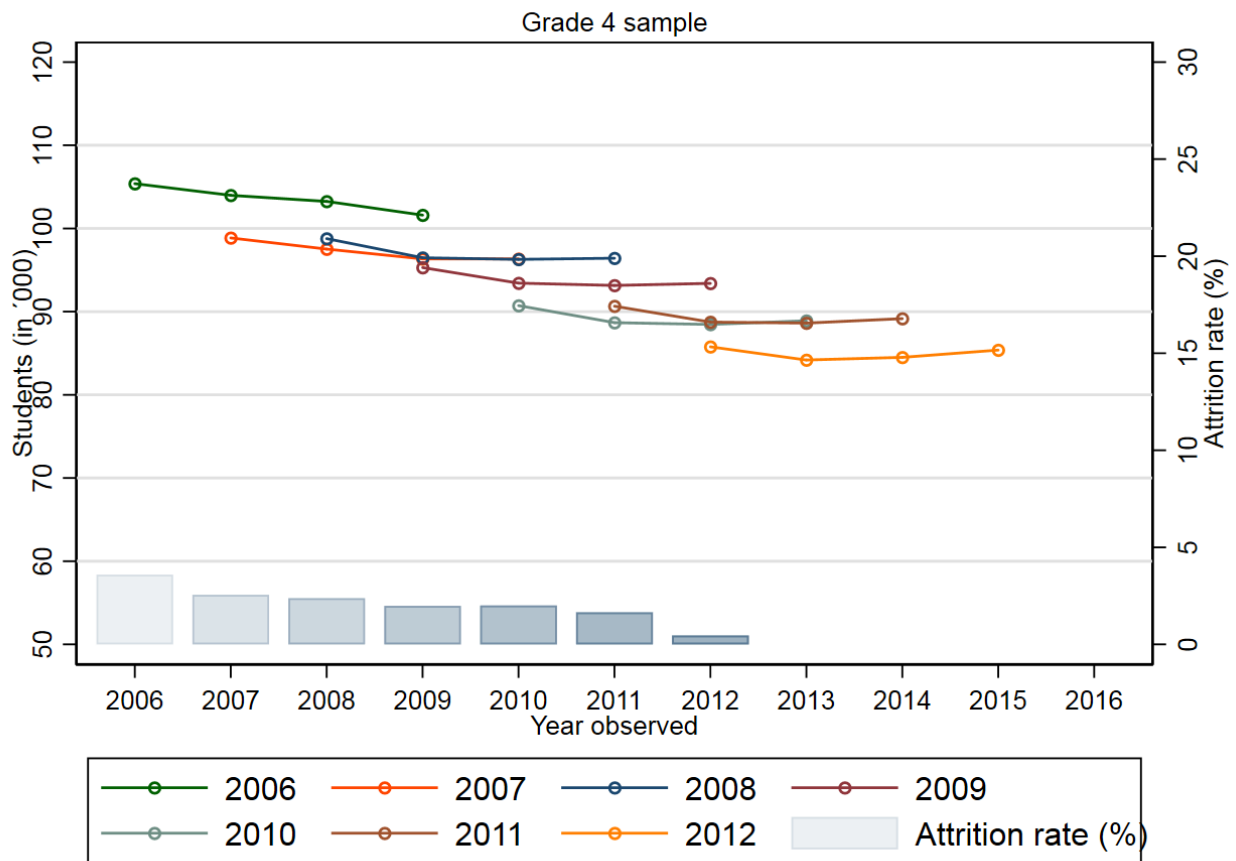
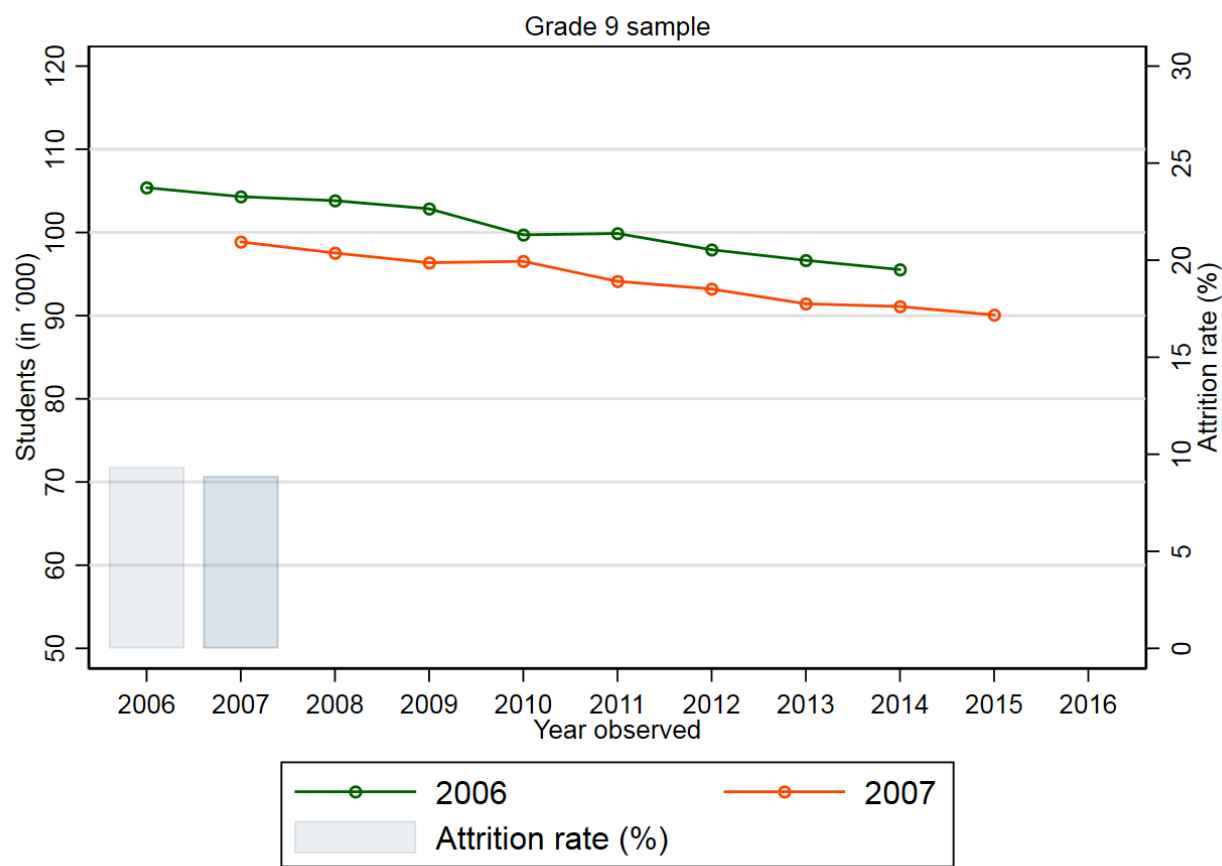


Figure 9.2: Panel attrition rates and number of tracked students by entry cohort (Grade 6 Sample, LATE Dataset)



Figure 9.3: Panel attrition rates and number of tracked students by entry cohort (Grade 9 Sample, LATE Dataset)



9.2. Solving for score manipulation concerns

The main outcome variables in our regressions is constructed from the score points of students in Grades 4, 6 and 9 Math and Language national exams, as a proxy for student cognitive ability. Students sit – or sat – national exams or standardized achievement tests at the end of each of these grades. To the best of our knowledge, this is the only reliable assessment of cognitive ability that was systematically performed to a large cohort of children in Portugal. Exams are anonymized and scored by randomly selected evaluating teachers, which are not teachers of the students whose achievement is being scored.

The major advantage is that these tests were sat by the universe of eligible children by Grades 4, 6 and 9. Due to policy changes, Grade 4 exams were discontinued from the beginning of the 2015/16 school year onwards. Additionally, the score scale at which ability was measured has changed from school year 2012/2013 onwards, with the previous discrete scale (0-5) being insufficient to retain relevant variation across students. Given the constraints, we retain three years of observations of student outcomes (2013-2015) for Grade 4, as measured by exam scores (0-100 scale) by the end of the school year.

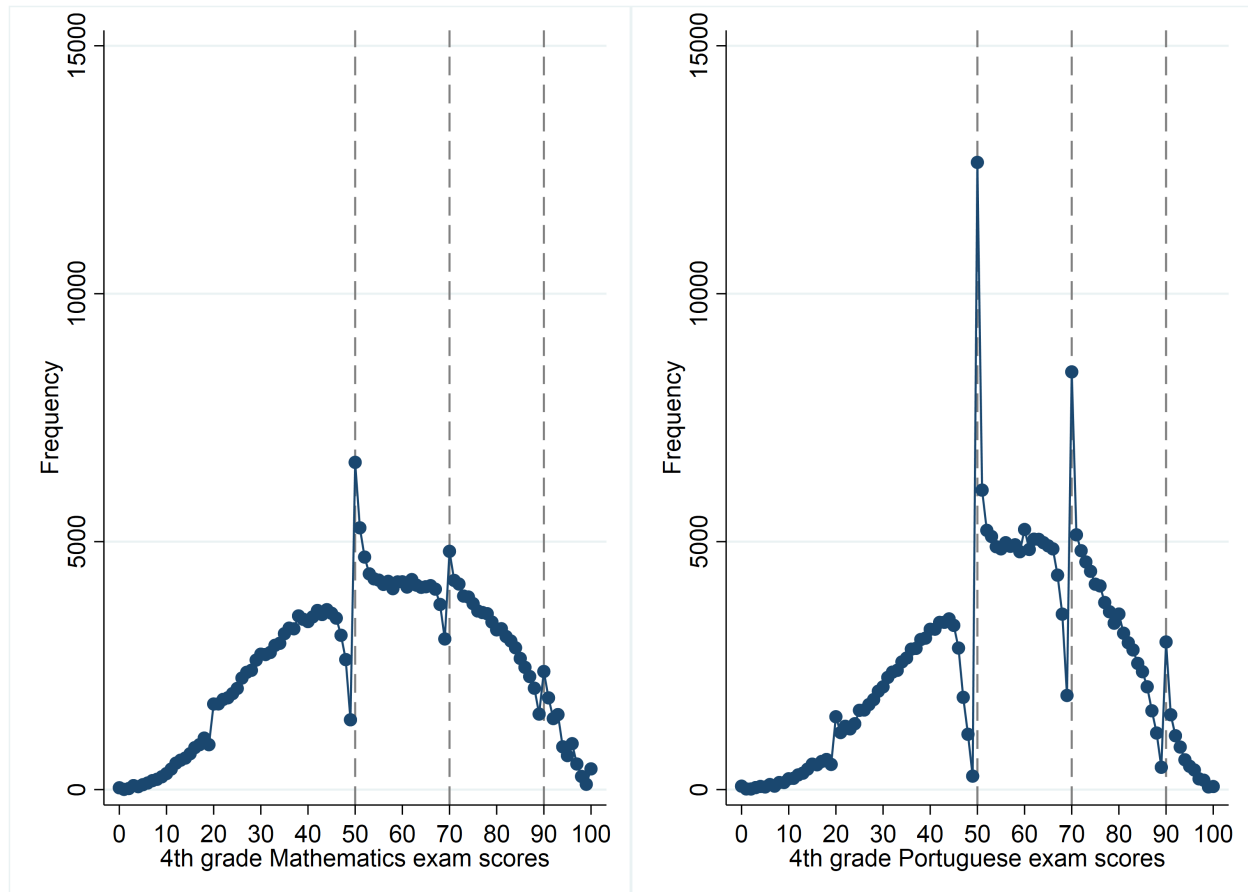
A simple analysis of the discrete distribution of exam scores shows significant manipulation of exam scores (see Figure 9.4 for this manipulation at Grade 4). The kinks in the distribution (at 20, 50, 75, and 90) coincide with threshold grade levels. The internal grades of K-12 education students are measured on a scale from 0-5 (from now on described as levels). These correspond to the threshold grade levels where kinks can be observed. Level 2, whose threshold is surpassed at 50 in the score scale is the most relevant here and the one where the kinks are most prominent. Exam scores below 50 are considered a fail, and students get a level 2 – a “negative” level, as it can lead to retention in the same grade. The raw distribution of exam scores strongly suggests that evaluators tend to upgrade scores that fall within a region of 5 points below the relevant level threshold. For instance, most students that would have 49 points as score are upgraded to 50. Such evidence of manipulation is visible in intervals of 5 points.

While this sort of manipulation may be beneficial to students at the margin of threshold levels, it can significantly bias the analysis if one is to take unit changes in the exam score as informative of cognitive differences between individuals. In order to circumvent this concern, we collapse the 0-100 scale into a 0-20 scale. Such a scale has a couple of advantages. First, it still retains informative variation across students. Second, it almost entirely eliminates score manipulation bias and reduces noise. In the new scale, 5 – instead of 1 – underlying exam points are now considered informative of student abilities differences. In other words, students that have 47, 48, 49, or 50 are considered within the same range of cognitive ability. The distribution of scores is thus smoother by being less prone to

bias introduced by score manipulation in the original scale.

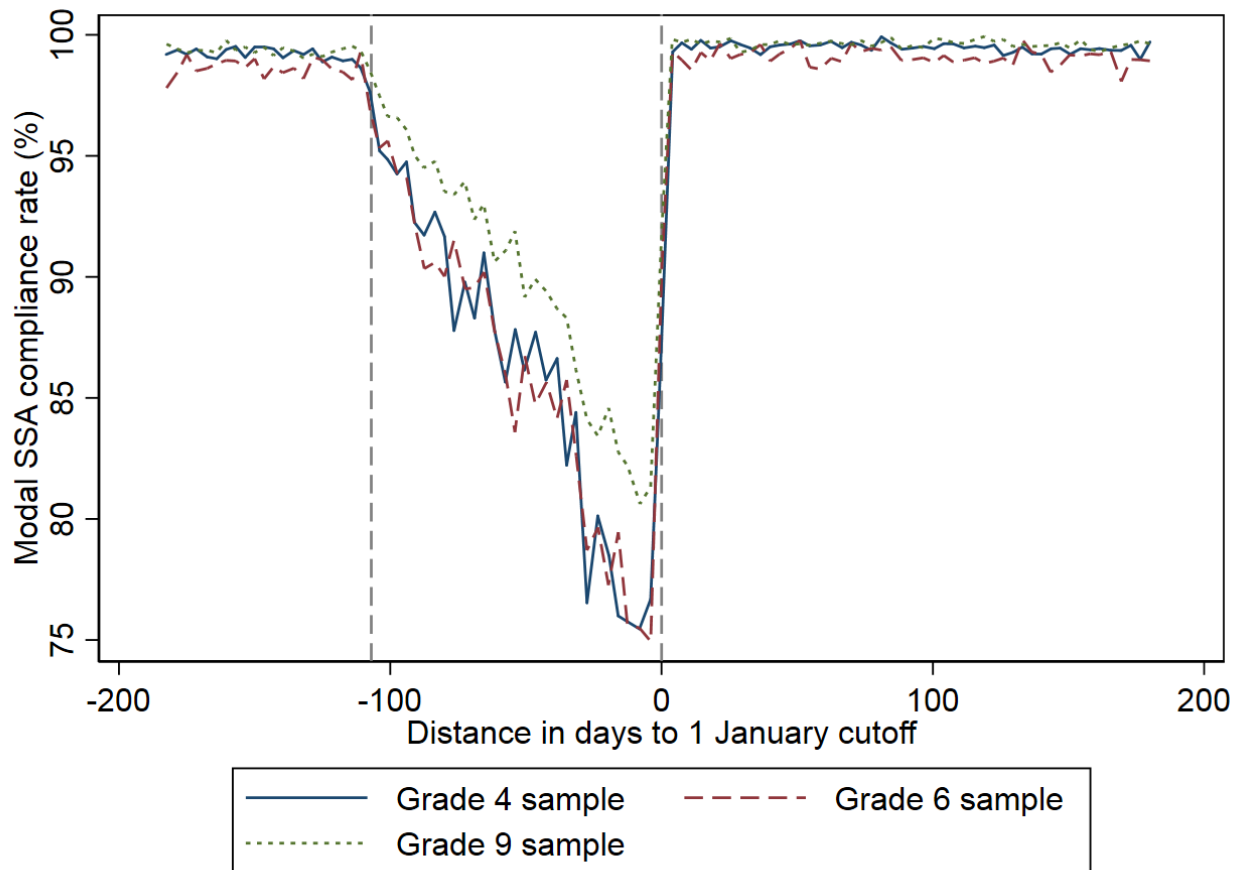
National exams in Portugal are not standardized across years, making inter-temporal comparisons biased. In order to control for the difficulty of the exam in each year, we standardize the scores in the new scale (0-20) by subtracting the mean and dividing by the standard deviation of subject scores in each exam year and by grade. The outcome variable is thus interpretable in standard deviation units of the exam scores in a 0-20 scale.

Figure 9.4: Evidence of exam score manipulation



10. Annex B: Compliance rates

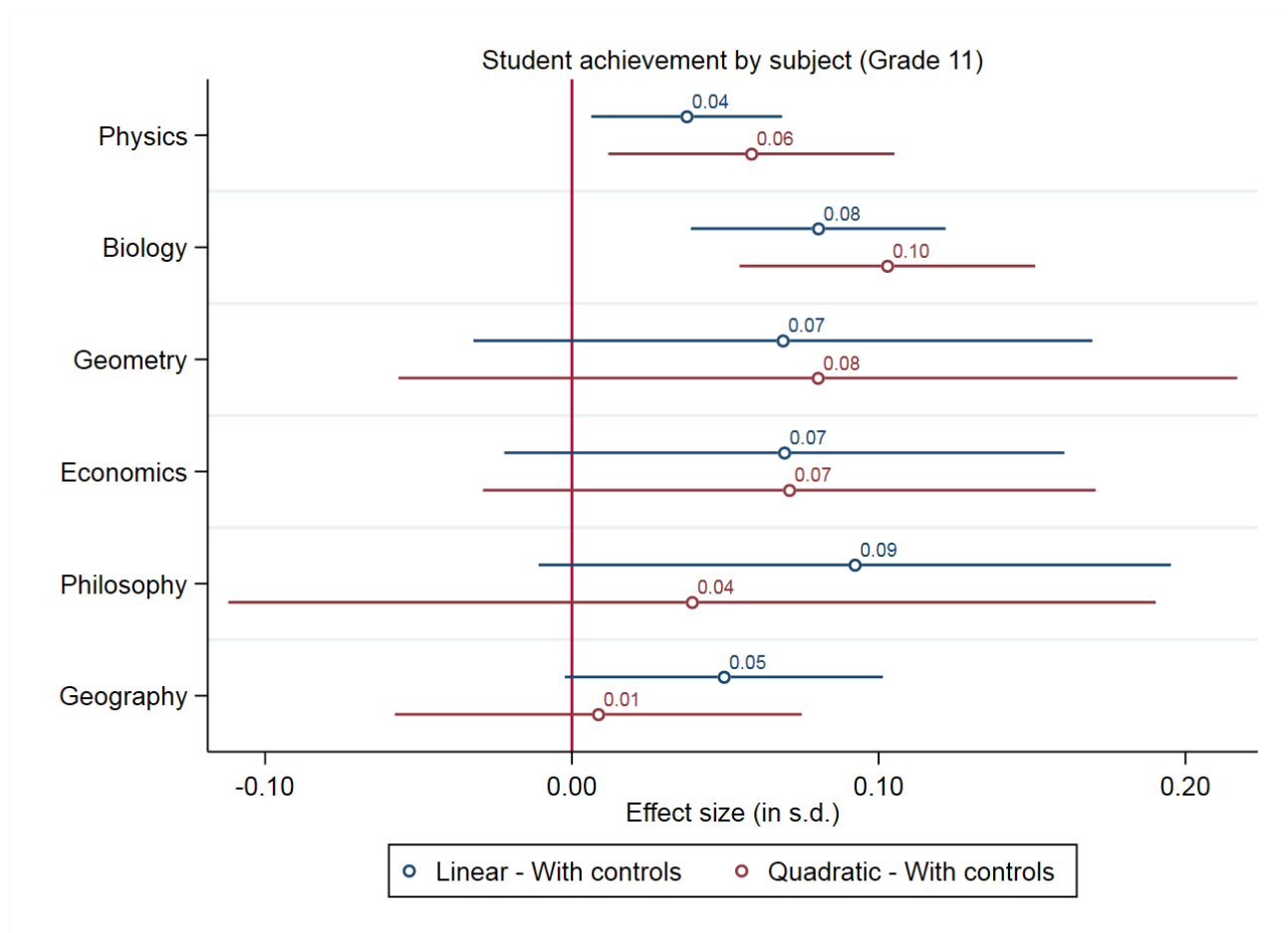
Figure 10.1: Compliance rates with the 1 January cutoff by grade sample



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Lines represent the trend (by day of birth) in compliance rates, i.e., the ratio of students that did not differ school entrance to the next school year given the modal school starting age for students born in a given day and the number of students born in each birth date bins, by grade sample.

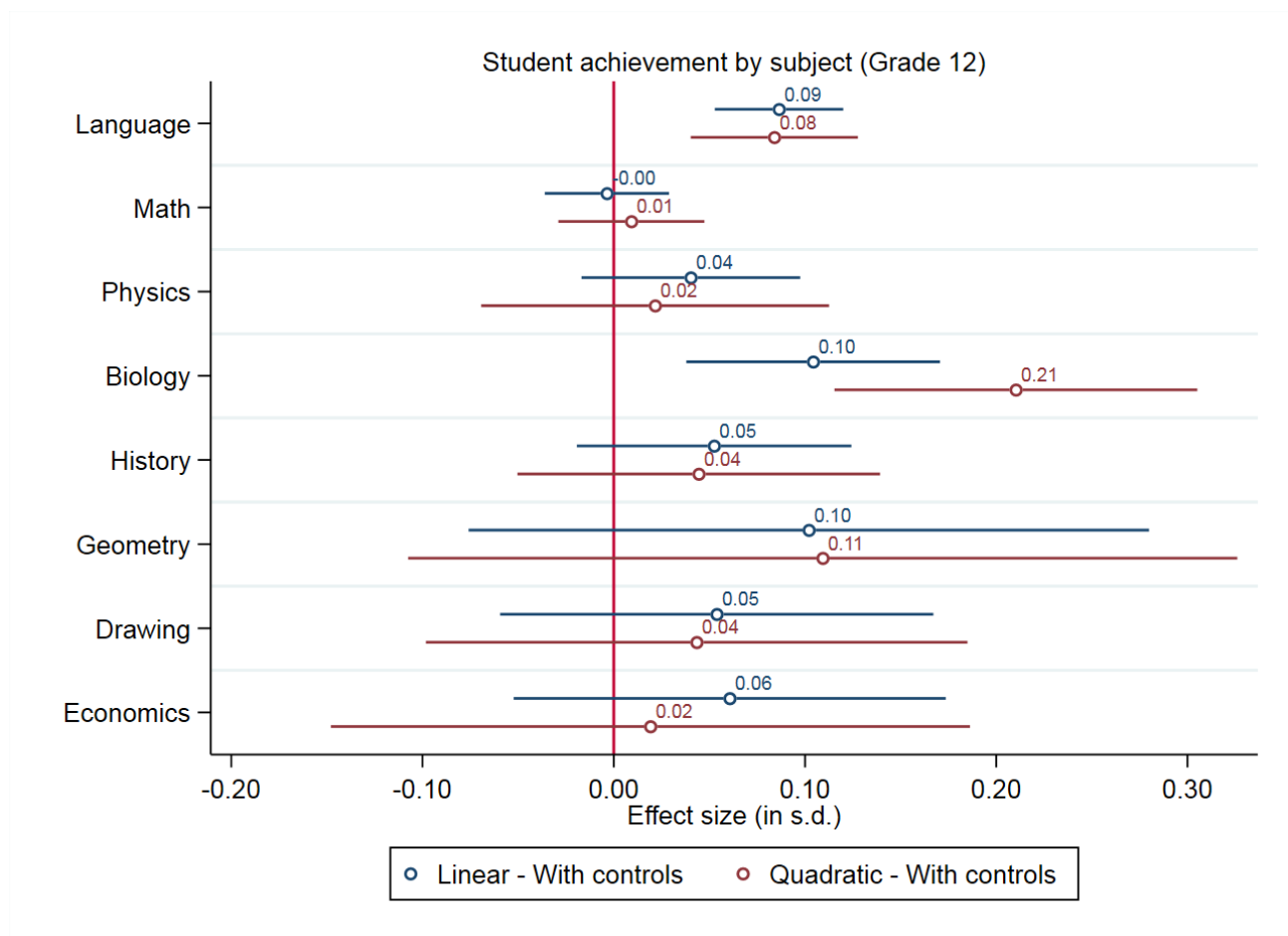
11. Annex C: Achievement premium by subject in Grades 11 and 12

Figure 11.1: Impact of school starting age on student achievement in Grade 11 national exams by subject



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal, were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1 and attended upper secondary education. Lines represent the trend (by day of birth) in compliance rates, i.e., the ratio of students that did not differ school entrance to the next school year given the modal school starting age for students born in a given day and the number of students born in each birth date bins, by grade sample.

Figure 11.2: Impact of school starting age on student achievement in Grade 12 national exams by subject



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Lines represent the trend (by day of birth) in compliance rates, i.e., the ratio of students that did not differ school entrance to the next school year given the modal school starting age for students born in a given day and the number of students born in each birth date bins, by grade sample.

12. Annex D: Testing for inclusion of weekday controls

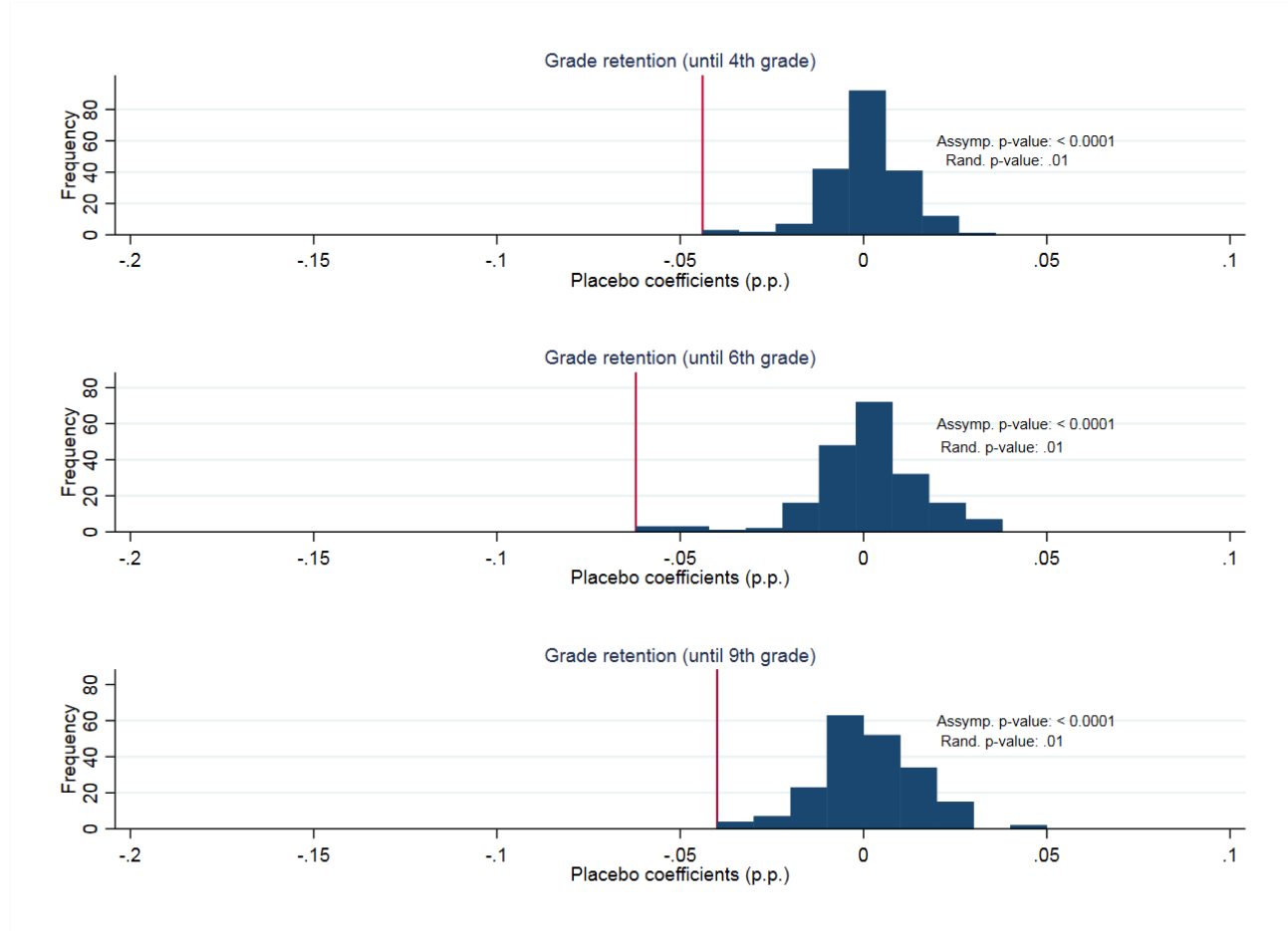
Table 12.1. Impact of school starting age on student outcomes including controls for day of the week

Outcome:	Math performance			Language performance			Grade retention		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>(Until) Grade 4</i>									
School starting age	0.288 (0.028)	0.272 (0.024)	0.274 (0.024)	0.393 (0.075)	0.356 (0.048)	0.373 (0.035)	-0.041 (0.017)	-0.052 (0.013)	-0.053 (0.011)
Observations	35742	72001	61461	35767	72040	71283	35826	72155	68302
Left bandwidth (in days)	30	60	53	30	60	66	30	60	54
Right bandwidth (in days)	30	60	47	30	60	50	30	60	64
<i>(Until) Grade 6</i>									
School starting age	0.194 (0.042)	0.239 (0.030)	0.247 (0.028)	0.326 (0.031)	0.302 (0.031)	0.291 (0.024)	-0.071 (0.014)	-0.079 (0.013)	-0.078 (0.011)
Observations	46958	94251	87299	47043	94393	123601	47125	94573	101436
Left bandwidth (in days)	30	60	63	30	60	73	30	60	57
Right bandwidth (in days)	30	60	48	30	60	82	30	60	69
<i>(Until) Grade 9</i>									
School starting age	0.168 (0.060)	0.155 (0.038)	0.153 (0.037)	0.152 (0.049)	0.184 (0.036)	0.175 (0.031)	-0.032 (0.013)	-0.036 (0.011)	-0.039 (0.010)
Observations	29245	58995	49042	29407	59290	65373	29428	59345	54603
Left bandwidth (in days)	30	60	42	30	60	70	30	60	49
Right bandwidth (in days)	30	60	56	30	60	61	30	60	61

Notes: All coefficients are estimates of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with an data-driven MSE-optimal bandwidth is used. All regressions also control for cohort fixed effects, day of the week in which the student is born, as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

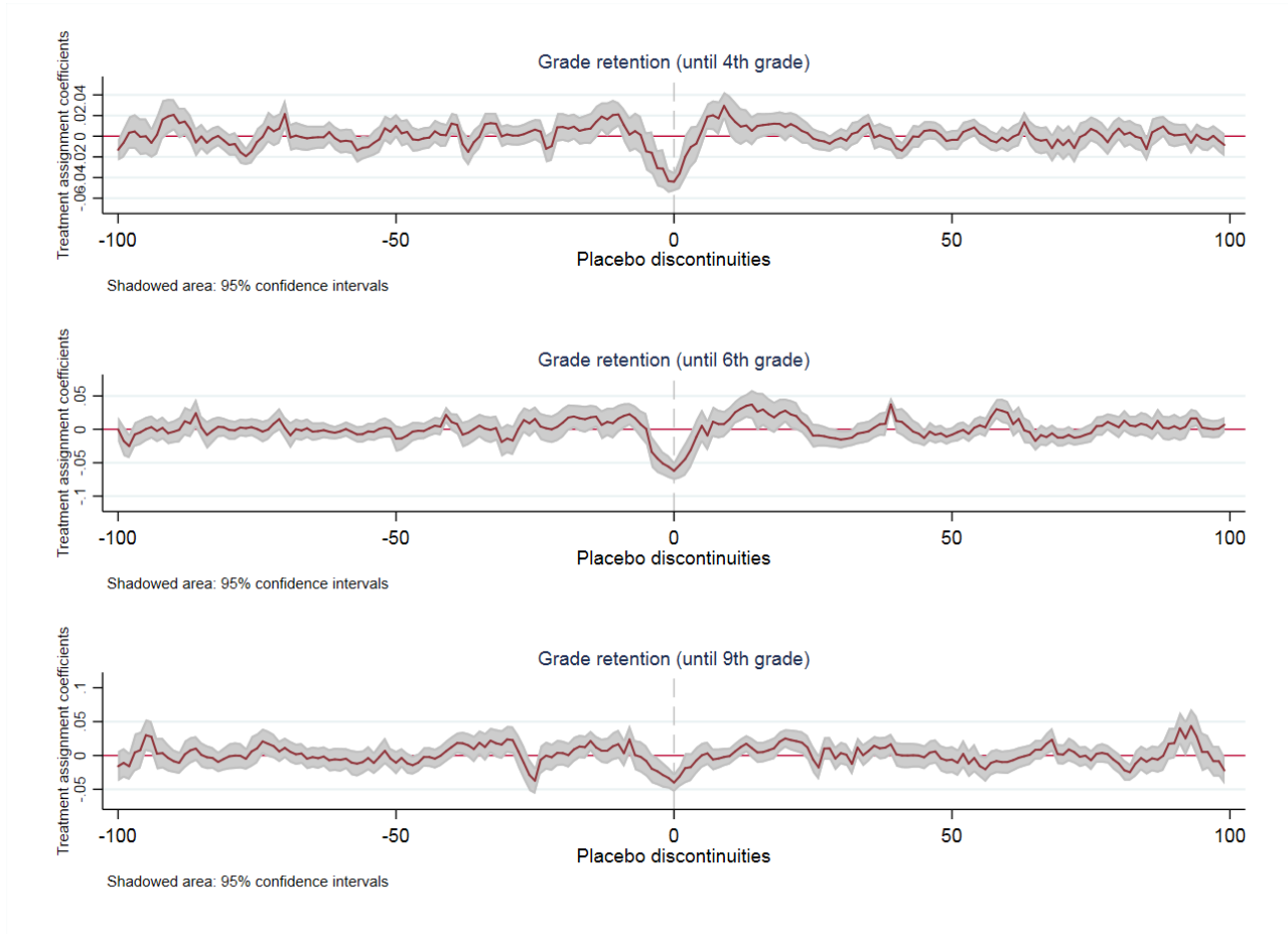
13. Annex D: Placebo tests on grade retention outcomes

Figure 13.1: Distributions of placebo and true intent-to-treat effects by subject and grade



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel is an histogram of estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one). Cutoffs considered are all where $B_i = \{-100, 99\}$. For each placebo cutoff the specification estimated is the one in Column 3, Table ??, with varying MSE-optimal bandwidths. The vertical red bar represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs. Randomization-based p -values, computed through the software package `rdpermut` and according to [Ganong and Jäger \(2018\)](#), are presented under the asymptotic p -values for the preferred specification at the true cutoff.

Figure 13.2: Intent-to-treat effects by potential cutoff, subject and grade



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel shows estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one) along the y -axis. Cutoffs considered are all where $B_i = \{-100, 99\}$, spread along the x -axis. For each placebo cutoff the specification estimated is the one in Column 3, Table ??, with varying MSE-optimal bandwidths. The vertical dashed line represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs.