Simulating the effects of automatic promotion on early dropouts for both poor and rich students: An application to Brazilian data

Max Cardoso de Resende * Francis Carlo Petterini †

XXIII Encontro de Economia da Região Sul – ANPEC SUL 2020

Abstract

The debate regarding the relative effects of automatic grade promotion versus grade retention have long concerned education policy-makers and researchers. This work applies a random effects probit model to assess the probability of early dropout in Brazilian public schools based on a unique administrative database which permits to infer household income at the student level for more than 60 thousand pupils enrolled in 273 schools. The results suggest evidence that decisions to drop out early are significantly influenced by whether approval has taken place and age-grade distortion. Moreover, simulation exercises have suggested that automatic grade promotion policy could reduce up to 35% the probability of a low-income student dropout from secondary school, while 24% for their rich peers.

Keywords: Secondary school dropout; Automatic grade promotion; random effects probit models.

JEL Classification: C33; I21; I28.

*Corresponding author. Federal University of Santa Catarina, Brazil, e-mail: max.resende@ufsc.br.

†Federal University of Santa Catarina, Brazil, e-mail: f.petterini@ufsc.br.

1 Introduction

According to the literature, low educational achievement levels and high probability of dropping out are more often correlated with economic disadvantages, age–grade distortion, teen parenting, poorly equipped schools, frustrated teachers, and unobserved factors that contribute to them, such as, motivation, low self-esteem, and peer effects in the classroom. Moreover, there is enough evidence that exiting the school environment prematurely has significant negative consequences for the individual and society, although it is not clear what triggers the decision-making mechanism to drop out (Murnane, 2013; Bassi *et al.*, 2015). For example, these pupils are subsequently exposed to a series of risks, when compared to their peers who have graduated, such as unemployment, poverty, demand for cash transfer programs, contributing less in taxes, chemical dependency, illness or disability, and crime (De Witte *et al.*, 2013).

Social scientists, educators, and policymakers are devoting considerable attention to the dropout problem and a wide range of educational policies have been designed and implemented, that aim to promote youth engagement towards school activities, such as, automatic grade promotion and grade retention. Traditionally, grade retention is seen as a pedagogic learning rule for better academic achievement that should be taken in cases of failure (Carifio & Carey, 2010; King *et al.*, 2016), while automatic promotion is believed to be less costly in terms of educational proficiency and improve pupil's socio-emotional outcome (Brophy, 2006; UNESCO, 2012; Adelman & Szekely, 2016).

In developing countries, other issues permeate this discussion, such as, the focus of educational policies designed to universalize primary and secondary education, age–grade distortion, and fiscal restrictions, where grade retention and dropping out may be seen as obstacles to improving the efficiency of the educational system and delivering the desired academic results, especially for pupils coming from disadvantaged backgrounds (Carifio & Carey, 2010; Koppensteiner, 2014). Yet even in some countries with compulsory education – Latin America and the Caribbean – most dropouts start school late, whereas the majority of out-of-school pupils in the Arab States and Africa are likely to never enter school (UNESCO, 2012).

In Brazil, grade retention and dropping out in the secondary school play a prominent role in the educational policy agenda. Historical data indicates that there was an improvement in grade retention rates over the past 30 years, dropping from 30% in the early 1990s to about 12% in 2018, but it still remains high relative to developed countries. Despite this sharp fall, there was no reduction in the dropout rate over this period: it remained at an annual average rate of 15%. According to Busso *et al.* (2013) and OECD (2018), when comparing Brazil's educational performance with the rest of Latin America and with OECD members, the country not only has one of the worst indicators, but one of the largest shares of adults without secondary education, which raises the necessity of better understanding the effects of educational policy on its pupils.

Moreover, the majority of Brazilian public school students are in poverty, and yet are characterized by wide heterogeneity among those who are considered rich and poor students. Those from low-income families are older and with a high age–grade gap rate, perform worse than their rich peers,, and are less frequently exposed to enriching activities outside of school. Hence, they have a higher probability of giving up their studies. Therefore, this paper makes two main contributions to the literature. First, it moves forward by applying an unique data set that permits inferring household income at the student level, whereas the majority of studies concerning Brazil (Menezes-Filho *et al.*, 2008; Koppensteiner, 2014; Leighton *et al.*, 2019) are based on administrative queries which do not allow making any correlation between income and academic performance. Second, it assess the determinants of early dropping out for rich and poor students using a methodology that is less susceptible to bias from potential unobserved heterogeneity, and presents positive evidence for the effects of an automatic grade promotion policy on the students' probability of dropping out.

Thus, this study examines a different aspect of the debate about grade retention and promotion: how pupils from different social backgrounds process the information of both policies for their decisions whether to drop out early. The empirical analysis takes advantage of an administrative longitudinal database from the state of Santa Catarina (located in the southern region of Brazil and ranking 5th in GDP *per capita* in 2019) that covers a representative sample of more than 60,000 secondary students enrolled in 273 public schools, and applies a random effect probabilistic model to explore heterogeneity in terms of social background. The results provide evidence that the decision to drop out early is significantly influenced by academic achievement and age–grade distortion, regardless of social background. Moreover, an automatic grade promotion policy could reduce by up to 35% the probability of a student not continuing their schooling, and keep potential retainers on track and engaged with the school system, while shrinking the gap between high- and low-income pupils.

This paper proceeds as follows: Section 2 presents the previous relevant literature; Section 3 describes the data; Section 4 presents the random effects probit methodology employed to assess early dropping out of secondary public school in Brazil; Section 5 discuss the empirical findings and evaluates the effects of a simulated automatic grade promotion policy on students; and Section 6 presents some final remarks and the model's limitations.

2 Literature review

There is mixed evidence about the relative merits of automatic grade promotion and grade retention for explaining patterns of enrollment, dropping out, and academic performance. It has been claimed that neither policy is an effective treatment for unsatisfactory achievement, but grade retention imposes too many social and motivational costs, and students appear to get more out of a year spent in the next grade than they do out of a year spent repeating a grade, even though automatic promotion does not help low achievers to improve their skills and catch up with their age peers (Brophy, 2006; De Witte *et al.*, 2013; King *et al.*, 2016).

De Witte *et al.* (2013) and Freeman & Simonsen (2015) reviewed the literature concluded that, in general, the main indicators of higher probabilities of dropping out and retention are more often correlated with disadvantaged socio-economic backgrounds, age–grade distortion, the necessity of working, poorly equipped schools, and peer effects. The majority of studies indicate that automatic grade promotion is the most common education policy applied to prevent those issues, followed by proposals related to curriculum changes, as well as improvement of school infrastructure and teaching skills. Also, Allen *et al.* (2009) and Xu *et al.* (2020) argued that empirical evidence has shown that grade retention did not have any efficacy in improving educational outcomes when compared to students who were promoted, and indeed may lead to negative externalities on their non-repeaters peers' behavior and outcomes. Further, Jacob & Lefgren (2009) suggests that retaining low-achieving eighth grade students increases the probability that they will not graduate from high school, whereas Schwerdt *et al.* (2017) provides evidence that even if there are any benefits from grade retention in the early ages of schooling, these have dissipated over time, at least in terms of test scores and attendance. Indeed, the proponents of automatic promotion suggest that the profile of children who leave school early is similar to that of those who repeat a grade, and are more likely to be subject to potential family problems. That is, the simple maintenance of the student at school justifies some loss of academic achievement, because it reduces possible exposure to problems of juvenile delinquency and family issues (Brophy, 2006; Wu *et al.*, 2010; UNESCO, 2012).

Studies of developing countries also presents conflicting arguments. Glick & Sahn (2010) conclude that grade retention among primary school student in Senegal leads to lower attainment and raises the probability of dropping out. Moreover, Manacorda (2012) found evidence in Uruguay that these effects may last up to five years after failure. Additionally, UNESCO (2012) argues that having a wide range of pupils in the classroom tends to reduce overall outcomes and explain low attainment (Allen *et al.*, 2009; Xu *et al.*, 2020) and early dropping out of school. In this context, second-chance and catch-up programmes are more costly and more difficult in terms of learning, once former students who dropped out are more likely to remain illiterate. In contrast to these studies, Carifio & Carey (2010) reports that promoting students into grades for which they are not prepared may lead to early dropping out by lowering the parent's valuation of schooling relative to the opportunity cost of the child's time spent in school. In Pakistan, King *et al.* (2016) find that as long as parents can make the schooling decisions, promotion based on attendance and test scores has the greatest impact on whether a child continues in school, while non-merit promotion may affect only boys.

In Brazil, Gomes-Neto & Hanushek (1994) evaluates the causes and consequences of grade retention in determining enrollment patterns in primary schools. The findings did not reveal any significant gains, and non-promoted students tend to keep repeating the grade if no other educational policy is assigned to help them improve their skills. Menezes-Filho *et al.* (2008) studied the impact of a continued progression program on school approval, dropping out, and performance rates for junior high school students in urban areas. The results suggest lower rates of dropping out for urban area schools, while the gains or losses in academic performance depend on the student's age. Moreover, it reduces the negative effects induced by age variation in the classroom (De Witte *et al.*, 2013) and increases the probabilities of graduating at the right age in elementary and high school.

Investigating the long-term effects of automatic promotion policies, Leighton *et al.* (2019) indicates that the policy has no significant effect on enrollment rates and dropping out. However, the students who were pushed ahead by the policy are less likely to be left behind in their studies and are more likely to graduate, when compared to the ones who faced retention. However, the results of Koppensteiner (2014) indicated a positive achievement result from repeating a grade. Comparing the test scores of 2nd and 4th grade students prior to and after an automatic policy was implemented, the evidence suggests a negative impact on then because it removed any link with the threat of retention.

Therefore, the literature has presented a great number of studies assessing grade retention and automatic grade promotion policies and their impact on academic results and dropout figures, which rely critically on context and the age of the student. However, in Brazil, there is a lack of clear evidence regarding secondary public school students, specially when considering the heterogeneity between poor and rich students, given the impossibility of measuring household income from the available public databases. This is a contribution this paper attempts to make.

3 Data

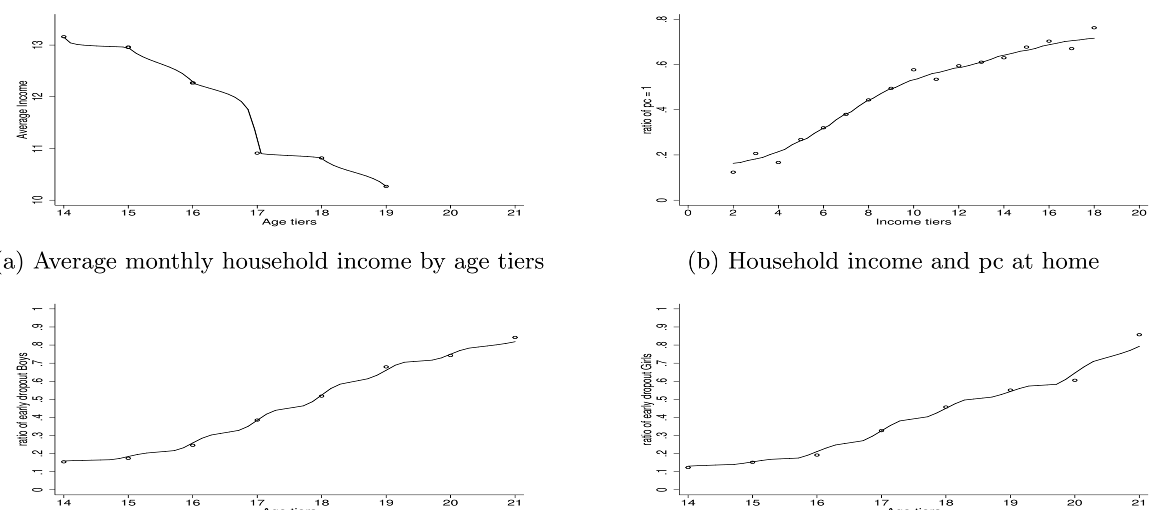
This research takes advantage of a unique longitudinal administrative data set from the state of Santa Catarina (Brazil), which combines educational outcome with information on subsequent school year enrollment and student and school characteristics of two cohorts of 30,871 and 30,853 pupils who started public secondary school in 2008 and 2009, respectively, which were monitored by 273 schools – there are repeated observations on students over the two years of cohort 2008, since there are two years of data. Brazilian secondary education consists of three school years, but the survey is restricted to pupils in grades 1 and 2, because it is not possible to identify precisely which students have completed the cycle (i.e., ended the 3rd grade with success), once the monitoring system was discontinued in 2010. Then, to better assess decisions to drop out of secondary school, this research defines *early drop out* $= 1$ for discontinued enrollment between school years, and *no early drop out* $= 0$ for continued enrollment, regardless of the academic outcome.

As mentioned by the literature (Dearden *et al.*, 2009; Glick & Sahn, 2010; De Witte *et al.*, 2013; Stark *et al.*, 2015) social background has been always a barrier to formal education. Although this research does not observe covariates for misreporting, the database contains some interesting indicators for household income. For example, upon the first enrollment in secondary school, applicants have to provide their average monthly household financial income, and 14.2% (8735 pupils) of them fulfill that query. This information is relevant because the literature shows that higher probabilities of dropping out occur among students from low-income families, which usually involves historical trends in attendance, retention in early schooling phases, which in turn may be a reflection of other institutional settings (e.g., curriculum, instruction, and/or the learning environment) and employment behavior.

Thus, Figure 1a presents a local polynomial smooth plot, which shows a negative correlation between a household's average income and the age of starting secondary school, that is, poorer families do not have the resources to provide better schooling opportunities for their children. Further, students 17 years of age or older reported being in worse environments, being more likely to manifest some of the precursors of dropping out, including low achievement rates and attendance, as well as cognitive skills linked to school success. In this context, these students are almost induced to leave the educational system without proper learning and, also, unable to promote social mobility (UNESCO, 2012; Busso *et al.*, 2013).

It is also well documented that academic performance differs by gender (Cataldi & KewalRamani, 2009; OECD, 2015). Male students make up 54% of the total student body, but 65% of repeaters and 54% of early dropouts, respectively. As shown in 1c and 1d, boys face greater risks of leaving school at the early stages of secondary school and of being older than their female peers. However, as they get older, the gap between then becomes smaller. Indeed, a boy entering secondary school today can expect to spend one year of this cycle repeating a grade. Overall, on average, 21.63% of the students do not enroll for the next school year in Santa Catarina, which is less than half of Brazil's average rate, but higher than those verified for Chile, Argentina, and Peru (Busso *et al.*, 2013).

Figure 1: Local polynomial smooth plots among household 's income and student's age tiers with ratios of early dropout and with students who had a personal computer (pc) at home



(a) Average monthly household income by age tiers

(b) Household income and pc at home

(c) Early dropout by age tiers: Boys

(d) Early dropout by age tiers: Girls

Furthermore, the database has some variables that can be proxies for household income. In particular, the schools asked which students had a personal computer (PC) at home – it was an expensive item in Brazil at that time. As suggested by Chowdry *et al.* (2011) and Helsper (2012), links between social background and digital exclusion have implied that access to and the use of informational and communication technologies (ICTs) is likely to improve educational attainment ad increase the social capital of the digitally excluded pupils.

Then, using the same structure as the previous chart, this paper attempts to integrate social and digital exclusion assessing the rates of students who have a PC by income tier, resulting in the local polynomial smooth plot in Figure 1b. There can be seen the existence of a correlation between these two variables, in which higher levels of poverty have a lower share of computer use. Following Murnane (2013), these results allow treating having a PC as a good *proxy* for household income, with the advantage that it is assigned for all pupils in the data set, which can be interpreted as a 'revealed income'. This feature is a key advantage of this research database when compared to the main data set presented at the student level in Brazil, the *Censo Escolar*, an annual census of schools in Brazil made available by the National Education Research Institute (INEP) – which does not include any variable that would allow one to infer household family income.

The database also provides an academic performance indicator per student: (i) did not attend school until end of school year, and failed the grade (retained absence); (ii) did attend class regularly until the end of the year, but failed to meet the minimum academic requirements (retained low performance); or, (iii) advanced in studies (approved). Thus, the system permits identifying if each student was enrolled in any state public school in the following year, regardless of their academic achievement. Table 1 summarizes the distribution of students' academic performance by early dropout status and social background:

Table 1: Distribution of Students academic performance by early dropout and social background (%)

				Academic Performance			
	Year	**Social Background**	**Early Dropout**	**Repeated Attendance**	**Repeated Low Performance**	**Approved**	**Total**
cohort 2008	2008	Poor	Yes	2.82 [42.19]	9.38 [62.26]	14.67 [18.74]	26.86
			No	3.87 [57.81]	5.68 [37.74]	63.59 [81.26]	73.14
		Rich	Yes	4.57 [33.11]	7.33 [60.96]	13.68 [18.44]	25.57
			No	9.23 [66.89]	4.69 [39.04]	60.51 [81.56]	74.43
	2009	Poor	Yes	1.10 [20.44]	5.66 [66.28]	7.82 [9.08]	14.58
			No	4.28 [79.56]	2.88 [[33.72]	78.26 [90.92	85.42
		Rich	Yes	1.12 [17.13]	4.88 [61.03]	7.84 [9.17]	13.84
			No	5.44 [82.87]	3.11 [38.97]	77.61 [90.83]	86.16
cohort 2009	2009	Poor	Yes	1.34 [32.71]	11.27 [58.11]	10.37 [13.56]	22.98
			No	2.75 [67.29]	8.12 [41.89]	66.15 [86.44]	77.02
		Rich	Yes	1.90 [30.03]	9.64 [54.87]	8.75 [11.50]	20.29
			No	4.43 [69.97]	7.93 [45.13]	67.35 [88.50]	79.71

Note: Number in brackets represents percentage by group. Source: Department of Education of Santa Catarina

Following Allen *et al.* (2009), Glick & Sahn (2010) and Leighton *et al.* (2019), the present paper presents evidence that dropping out early is significantly lower among promoted students. Furthermore, this negative correlation is even more dramatic for the poorest, of which, on average, 66% of repeaters did not enroll in any state public school for the next school year. On the other hand, almost 90% of the students who were approved, continued their schooling activities, results that are similar to their rich peers. Indeed, Table 1 and Figure 1 suggest that low academic performance and age are strong predictors of dropping out of school early, which are all related to a worse social background (UNESCO, 2012; Adelman & Szekely, 2016).

Furthermore, it can be seen that key variables are generally analyzed in literature reviews of dropping out of school (De Witte *et al.*, 2013; Freeman & Simonsen, 2015), which can be observed in this research data set and represents the impact of family background, individual characteristics, and school environment on academic performance. Table 2 presents the descriptive statistics of selected variables.

Table 2: Definition and Summary Statistics of selected variables (%)

Variable	Definition	Full Sample	Social background		
			Poor	**Rich**	**Difference•**
Academic outcomes					
early dropout	1 if the pupil did not enroll for the next school yea	21.63	22.14	19.42	2.72***
approved	1 if the pupil pass the school year	79.20	79.78	77.88	1.89***
retained performance	1 if the pupil fail the school year due to low performance	14.31	15.82	13.13	2.69***
retained absences	1 if the pupil fail the school year due to abscences	6.49	5.41	8.98	-3.57***
Student Characteristics					
boy	1 if the pupil is a boy	46.10	45.17	48.22	-3.05***
white	1 if the pupil is white	96.29	96.48	95.87	0.61
urb	1 if the pupil lives in an urban area	82.73	79.95	89.13	-9.18***
govaid	1 if the pupils received family aid	2.00	2.38	1.11	1.27
entry age†	Secondary school starting age	14.91	14.94	14.55	0.39***
School Characteristics					
elemntary school	1 if the school pupils is enrolled offers elementary school	88.53	89.83	85.51	4.32
lib	1 if school offers library	78.51	80.45	73.98	6.47
sci	1 if school offers science lab	24.49	25.13	22.99	2.14
sports	1 if school offers covered sports court	17.66	16.62	20.07	-3.45
size room†	Quantity of students per classroom	26.89	26.78	27.16	-0.38
teachers age†	Average School teacher age	40.75	40.61	41.04	-0.42
tecahers age deviation†	School teacher age standard deviation	10.59	10.53	10.73	0.20
student-staff ratio†	Student to Faculty Ratio	4.15	3.93	4.66	-0.72***

† - Expressed in absolute values;

• - Compare means of the same variable between two groups based on Independent sample T-test assuming unequal variances;

Source: Department of Education of Santa Catarina (Brazil).

Regarding the characteristics of the students, the database is composed mainly of female (53.9%) and white (96.3%) students who live in urban areas (82.73%) and low-income students are more likely to skew towards being older and with lower academic performance (Table 1 and Figure 1a) probably because this condition is directly related with the need to look for a job to help at home (Busso *et al.*, 2013; Murnane, 2013). As for the ethnic background, higher retention rates are more related to non-white students (Christle *et al.*, 2007), where 22% of then fail a school year due to poor academic achievement and 8% due to unexcused absences compared to 14% and 6% of their white peers. Lastly, 2.0% of the students are in families enrolled in some government aid program, but there is no statistical difference between those who are and those who are not in this condition, probably because this condition is mainly determined by the characteristics of the kind of family members whose existence determines this help, e.g., children with special needs and elders.

The last block of Table 2 addresses some of the structural characteristics of the schools. In this matter, 88.53%, the vast majority, offer elementary school as well as secondary education, 78.5% have a library, but less than one-quarter have a science lab or an indoor sports court. Also, the average class size is 28 students in the state's secondary schools and the average teacher's age is almost 41 years old. For both variables, there is no statistical difference across groups, suggesting that the impact of class size might be linear across the range of class sizes observed in the database, from roughly 15 to 35 students per class, and a similar analysis can be inferred about the teacher's age profile. As for the student–staff ratio, it express the importance of the faculty, by presenting evidence that it has different effects on different groups of students, especially on the poorest. As mentioned by Christle *et al.* (2007), most of the research on dropping out has focused on the characteristics of individuals, whereas the features of the school have rarely been considered (Allen *et al.*, 2009; De Witte *et al.*, 2013; Freeman & Simonsen, 2015). This is a contribution this paper attempts to make, by providing a deeper understanding of the the school variables affecting the decision to drop out.

4 Methodology

The factors that influence dropping out are multiple and complex, covering different perspectives, from those that portray students' socioeconomic and family status to the dynamics in the classroom between teachers and students, and the structural characteristics of the school. However, there are unobserved characteristics (for example, motivation, a student's effect on peers' performance, allocation of teachers, cognitive skills, and others), which are not captured by traditional regression models, something which could generate a bias in the estimation of the coefficient of interest (Cameron & Trivedi, 2005; Train, 2009; Clarke *et al.*, 2015).

The challenge of panel methodology is to control the impact of unobserved heterogeneity in order to obtain a valid inference for the structural parameters. As mentioned by Hsiao (2007) and Bland & Cook (2019), the advantages of random effects (RE) specifications are that the number of parameters remains constant when the sample size increases, and it also allows the estimation of the impact of time-invariant variables, such as gender, race, place of residency, and school-level facilities, among other factors that were used as exogenous variables on the matter of this research.

Letting $Y_{i,t}$ be the observable binary outcome of individual i at time t – in this study, the decision to drop out or not from a public school at the end of the school year – the latent variable representation of the RE probit model can be written as:

$$Y^*_{i,t} = X_{i,t}\beta + \varepsilon_{i,t} + u_i \qquad\qquad Y_{i,t} = \begin{cases} 1 \rightarrow Y^*_{i,t} \geq 0 \\ 0 \rightarrow Y^*_{i,t} < 0 \end{cases} \tag{1}$$

where $Y^*_{i,t}$ t is the unobserved latent variable, β is the corresponding vector of regression coefficients of each of these covariates, which are listed in Table 2. Under the RE approach, the total residual can be partitioned into two components: $\varepsilon_{i,t}$, independent errors distributed $N(0, \sigma^2_\varepsilon)$ that are orthogonal to the individual specific u_i of the panel, which assumes a parametric distribution. Then, all the parameters are jointly estimated using maximum likelihood. Thus, this methodology allows one to verify whether unobserved characteristics in the educational system as well as in the student's socioeconomic and cultural status are not correlated with any exogenous variables that were specified in the probabilistic models.

Following Bland & Cook (2019), the predicted probability can be obtained as

$$\Pr\left[Y_{i,t} = 1 | X_{i,t}\right] = \Phi\left(\frac{\alpha + X_{i,t}\beta}{\sqrt{1 + \sigma^2_u}}\right) \tag{2}$$

in which Φ is the standard normal cumulative distribution function for the Probit distribution. In the context of this paper, Equation 2 will give the probability that a randomly selected student will drop out of a secondary school in Brazil.

The random effects probit models have become essential for an educational researcher's analytical arsenal. But, an issue that may be important for whether the results are meaningful, concerns the interpretation of the estimated vector β, that is, the sign and significance provides the direction, but not the magnitude, of the exogenous variables on the binary outcome (Train, 2009; Bland & Cook, 2019). Hence, the RE probit specification should evaluate its corresponding marginal effects

$$\frac{\partial \Pr\left(Y_{i,t}=1|X_{i,t}\right)}{\partial x_k} = \frac{\beta_k}{\sqrt{1+\sigma_u^2}}\phi\left(\frac{\alpha + X_{i,t}\beta}{\sqrt{1+\sigma_u^2}}\right) \tag{3}$$

where $\phi(.)$ is the standard normal frequency function. Using these results, different RE models can be compared.

5 Results

Although there is a large literature on the determinants of dropping out early, convincing evidence for the secondary school that takes into account household income and longitudinal data is scarce, especially for developing countries where retention is widespread (UNESCO, 2012; Busso *et al.*, 2013). As described in Section 4, a naive regression of early dropout status – from this point on, referred to as 'dropout' for simplicity – would face potential heterogeneity bias and provide invalid statistical inferences. Thus, the first part of the analysis identifies the determinants of dropout based on RE probit models (Equation 2), in which two models were proposed: tthe first specification, involves only the student's characteristics plus two explanatory dummy variables to identify approved students and to differentiate cohorts; and a more general specification, which depicts all the variables just listed with school environment indicators – the variables are listed in Table 2. Furthermore, for each specification, the sampling strategy verifies the effects of these variables by social background, that is, differences between rich and poor students as defined by the revealed income criterion described earlier.

Table 5 (in the Appendix) presents the main results for all RE probit regressions. Overall, the estimates in columns (1) to (6) suggest that promotion decisions, entry age, revealed income, and gender can positively affect the probability that a student continues in school, which corroborates previous evidence found in the literature (Menezes-Filho *et al.*, 2008; Freeman & Simonsen, 2015; Schwerdt *et al.*, 2017). On the other hand, government aid and school facilities are not significant in terms of avoiding dropping out of high school. Also, the estimated values of σ_u^2 show that there is enough variability between students to favor a random effects probit regression over a standard probit specification.

However, as mentioned by Train (2009) and Bland & Cook (2019), in order to compare different RE probit models, each specification should be evaluated by its corresponding mean marginal effects. Table 3 presents the estimated average marginal effects for each variable in both specifications estimated on the first part of this analysis. Overall, the results present clear evidence that being approved is by far the most important predictor of dropout, specially among poor pupils, where the predicted probability of dropping out is 0.255 greater for failed students than for those who were approved when considering the general specification. Moreover, entry age is another crucial variable, compared to that of their rich peers.

It is worth noting that government aid and revealed income have not shown any relevance as predictors for dropping out. Indeed, the government transfer program was not statistically significant in any specification, whereas the predicted probability of dropping out is only 1.6% lower for high-income students than for their low-income peers. This brings new evidence to the literature, in which inequality is not a significant predictor of dropping out, whereas approval outcome and age are the leading indicators of the decision to drop out. This analysis focuses on relative differences across students not varying in the short term, because the neighborhoods they live in, the institutions they interact with, and the perceptions they develop about their world and future opportunities are probably formed by the semi-permanent conditions of the state and not transitory fluctuations in household income. Furthermore, countries like Brazil, where the majority of public school students live in poverty, tend to have lower levels of social mobility, which could offset any potential positive effect coming from higher income.

Table 3: Marginal Effects from Estimated RE probit models

Variable	**Full Sample (1)**	**Social Background Rich (2)**	**Poor (3)**	**Full Sample (1)**	**Social Background Rich (2)**	**Poor (3)**
approved	-0.242***	-0.217***	-0.263***	-0.239***	-0.218***	-0.255***
	(0.001)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)
boy	0.006**	0.014**	0.001	0.005**	0.012**	0.001
	(0.001	(0.005)	(0.003)	(0.002)	(0.005)	(0.003)
white	0.022***	0.014	0.026***	0.022***	0.015	0.025***
	(0.006)	(0.011)	(0.008)	(0.006)	(0.011)	(0.008)
urb	0.055***	0.063***	0.053***	0.049***	0.057***	0.048***
	(0.003)	(0.008)	(0.004)	(0.003)	(0.008)	(0.004)
govaid	-0.011	-0.008	-0.012	-0.009	-0.007	-0.011
	(0.009)	(0.023)	(0.010)	(0.009)	(0.023)	(0.010)
entry age	0.099***	0.101***	0.097***	0.095***	0.099***	0.097***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)
cohort 2009	-0.041***	-0.048***	-0.037***	-0.038***	-0.046***	-0.035***
	(0.003)	(0.005)	(0.003)	(0.003)	(0.006)	(0.003)
revealed income	-0.016***			-0.016***		
	(0.003)			(0.003)		
elementary school				-0.001	0.002	-0.001
				(0.004)	(0.007)	(0.005)
size room				-0.002***	-0.002***	-0.002***
				(0.001)	(0.001)	(0.001)
lib				-0.016***	-0.031***	-0.008**
				(0.003)	(0.005)	(0.004)
sci				-0.009***	-0.031***	-0.002
				(0.003)	(0.006)	(0.003)
sports				-0.004	-0.002	-0.007
				(0.003)	(0.006)	(0.004)
teachers age				0.003***	0.004***	0.003***
				(0.000)	(0.001)	(0.000)
teachers age desviation				-0.001	-0.002	-0.001
				(0.001)	(0.002)	(0.001)
student/staff ratio				-0.001	-0.001	-0.000
				(0.000)	(0.001)	(0.001)

Note: Marginal effects expressed at average values. Variable revealed income is omitted on models (2), (3), (5) and (6); * → p < .05; ** → p < .01; *** → p < .001.

5.1 Simulating automatic promotion effects

Among the strategies for facing the drop out crisis and age–grade distortion in high school and thus achieve the universalization of basic education and regularize school flow, one of the most debated public educational policies in the literature is automatic grade promotion (Glick & Sahn, 2010; Manacorda, 2012; Leighton *et al.*, 2019; Xu *et al.*, 2020). Thus, in order to develop insights for real-life policy changes and to make a contribution to the literature on the design of simulations, an empirical exercise was carried out in order to determine the possible effects of automatic grade promotion on the likelihood of dropping out and, also, on how it would affect the gap between high- and low-income pupils in the public school system in Santa Catarina.

The predicted probabilities of the simulated effects of an automatic promotion policy were calculated by varying the value of the explanatory characteristic in question – e.g., academic performance at the end of the school year – in the general specification. From an operational perspective, first, the predicted probability of dropping out in the grade retention regime was estimated as in Equation 2, taking into account the existing heterogeneity among students. Then, keeping all other characteristics constant, as well as their estimated coefficients, the approved condition was re-defined as equal to 1 for all students, although the value of its parameter remains the same. Hence, this simulation design ensures that the only characteristic that changes is the academic achievement, which allows the researcher to quantify the contribution of the policy employed to school dropout rates.

Table 4 provides clear evidence of the benefits of an automatic grade promotion policy on decisions to drop out against grade retention, reinforcing the evidence provided by Wu *et al.* (2010), Manacorda (2012) and UNESCO (2012). Overall, it can be seen that there is an average predicted probability of dropping out of 22% in the current grade retention regime, whereas if an automatic approval policy were adopted, this rate would be reduced by 30%. Also, Figure 2a indicates that older students are more likely to drop out: indeed, pupils aged 14–16 have a predicted probability of dropping out three times lower than those over 17 years, regardless of the policy implemented, which indicates the necessity of avoiding age–grade distortion and of keeping potential repeaters engaged with school activities, as reported by Freeman & Simonsen (2015) and Leighton *et al.* (2019).

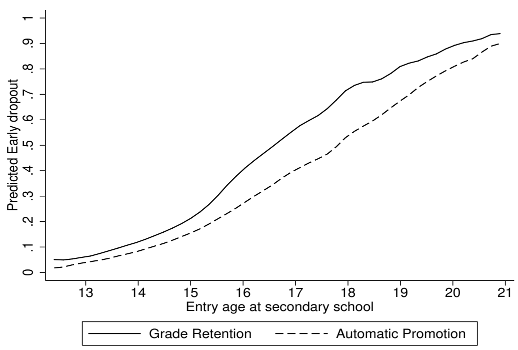
Table 4: Predicted Probabilities of dropout: Grade Retention X Simulated Automatic Promotion

		Predicted Probabilities		
	Educational Policy	**Sample**	**Mean**	**Std. Drv**
Model 1	**Grade retention**	Full Sample	0.226	0.176
		Rich	0.208	0.169
		Poor	0.240	0.181
	Automatic Promotion	Full Sample	0.160	0.089
		Rich	0.159	0.089
		Poor	0.156	0.087
Model 2	**Grade retention**	Full Sample	0.225	0.178
		Rich	0.203	0.171
		Poor	0.233	0.182
	Automatic Promotion	Full Sample	0.159	0.089
		Rich	0.151	0.092
		Poor	0.155	0.087

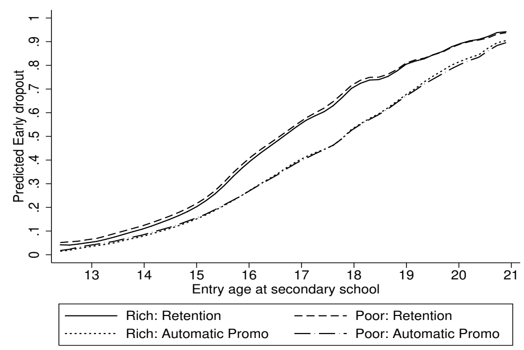
The retention rates presented in Table 1 demonstrate that those deemed most likely to suffer from low academic performance are the poorest students. Indeed, they are the ones who would be most affected by a promotion policy, as reported in Table 4, reducing their likelihood of dropping out by 35%, under both specifications. However, among older students, such a policy does not have a significant impact on dropout rates since they remain extremely high, which suggests the need of reducing problems related to age–grade distortion, as mentioned by Freeman & Simonsen (2015) and Xu *et al.* (2020). That is, regardless of the educational policy adopted and the social background, older students will always have a higher risk of dropping out, which indicates the need for treating this group of students differently, and constitutes a limitation of the scope of an automatic promotion policy.

Moreover, the simulations stress that little evidence has been found concerning the importance of school resources for both groups (Model 2), while family background and academic achievement are highly relevant for dropout decisions (Model 1). All of these aspects reinforce the positive effects of an automatic approval policy, especially in developing countries with social sector budget restrictions and where the majority of schools need infrastructure upgrades or replacements, complementing the evidences provided by Busso *et al.* (2013) and Koppensteiner (2014).

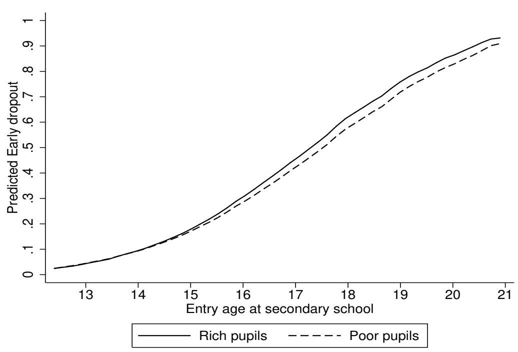
Figure 2: Local polynomial smooth plots: Comparing grade retention and simulated automatic promotion on predicted dropout rates



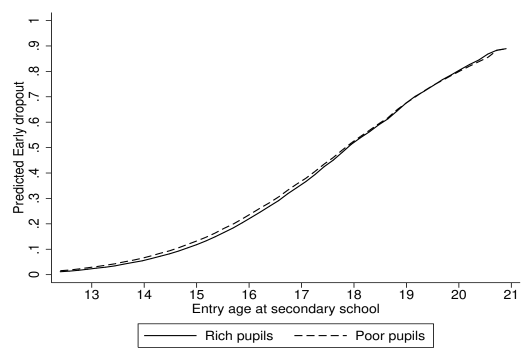
(a) Grade Promotion X Automatic Promotion



(b) Effects by social background



(c) Cohort 2008: Automatic promotion effects



(d) Cohort 2009: Automatic promotion effects

A cohort analysis such as that presented in Figures 2c and 2d helps to clarify the dynamic aspects of automatic promotion that underlie the results summarized in Tables 3 and 4. When the analysis is extended to more than one cohort, the positive impact of the automatic approval policy on dropping out of school is sustained, indicating the consistency of the results over time, since predicted dropout rates by age tier have a similar behavior for both cohorts despite the fact that there are small divergences in trends per social background group, specially after 16 years of age. This is probably due to some unobserved effect that may have affected all students in a given cohort in a homogeneous manner, such as the 2008 financial crisis, which may have driven students in the cohort of 2009 to make similar decisions regarding dropping out, regardless of their social group.

6 Final remarks

Researchers and policy makers who have focused on grade retention and automatic promotion typically conclude that neither policy is an effective treatment for unsatisfactory achievement, but if one must choose between them, automatic grade promotion is preferable. This is because grade retention imposes too many social and motivational costs, and increases the probability that these students will not graduate from high school, even though automatic promotion does not help low achievers to begin to catch up with their age peers.

Therefore, to develop a more realistic scenario for Brazil, this research assesses how pupils from different social backgrounds process the information of both policies in decisions to drop out early. The results suggest that the decision whether or not to drop out early is significantly influenced by successful academic achievement and age–grade distortion, regardless of social background, which is already in itself a indicative in favor of an automatic grade promotion policy. Moreover, the estimates have shown that the likelihood of dropping out can be reduced by 35% and 24% for low- and high-income pupils, respectively, if an automatic promotion policy were applied. However, among older students, such policy does not have any significant impact when compared to the current grade retention policy, suggesting a limitation of the scope of the automatic promotion policy. Also, the simulations stress that little evidence has been found concerning the importance of school resources, most likely due to the need of infrastructure upgrades or replacements.

The use of proxies for household incomes is highly debatable and serves as an initial exploration when precise measurements are difficult to obtain, especially in developing countries such as Brazil. Thus, based on the links between social background and digital inclusion and on the availability of student-level information, this research suggested the use of a computer as a consistent indicator for household's income. In this matter, we believe that this discussion has been helpful in showing a way to deal with partial information in school drop out researches. For future research, first, it is recommended that investigators apply this methodology to many other databases with similar observability problems, to check its capacity to predict dropping out early in other contexts, and, second assess how the costs and benefits of automatic grade promotion are evaluated by pupils form different social backgrounds.

References

Adelman, Melissa, & Szekely, Miguel. 2016. *School dropout in Central America: An overview of trends, causes, consequences, and promising interventions*. The World Bank.

Allen, Chiharu S, Chen, Qi, Willson, Victor L, & Hughes, Jan N. 2009. Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, **31**(4), 480–499.

Bassi, Marina, Busso, Matias, & Muñoz, Juan Sebastian. 2015. Enrollment, graduation, and dropout rates in Latin America: is the glass half empty or half full? *economía*, 113–156.

Bland, James R, & Cook, Amanda C. 2019. Random effects probit and logit: understanding predictions and marginal effects. *Applied Economics Letters*, **26**(2), 116–123.

Brophy, Jere. 2006. Grade repetition. *International Academy of Education*.

Busso, Matias, Bassi, Marina, & Muñoz, Juan Sebastián. 2013. Is the glass half empty or half full? School enrollment, graduation, and dropout rates in Latin America.

Cameron, A Colin, & Trivedi, Pravin K. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Carifio, James, & Carey, Theodore. 2010. Do minimum grading practices lower academic standards and produce social promotions? *educational HORIZONS*, **88**(4), 219–230.

Cataldi, Emily Forrest, & KewalRamani, Angelina. 2009. High School Dropout and Completion Rates in the United States: 2007 Compendium Report. NCES 2009-064. *National Center for Education Statistics*.

Chowdry, Haroon, Crawford, Claire, & Goodman, Alissa. 2011. The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16. *Longitudinal and Life Course Studies*, **2**(1), 59–76.

Christle, Christine A, Jolivette, Kristine, & Nelson, C Michael. 2007. School characteristics related to high school dropout rates. *Remedial and Special education*, **28**(6), 325–339.

Clarke, Paul, Crawford, Claire, Steele, Fiona, & Vignoles, Anna. 2015. Revisiting fixed-and random-effects models: some considerations for policy-relevant education research. *Education Economics*, **23**(3), 259–277.

De Witte, Kristof, Cabus, Sofie, Thyssen, Geert, Groot, Wim, & van Den Brink, Henriëtte Maassen. 2013. A critical review of the literature on school dropout. *Educational Research Review*, **10**, 13–28.

Dearden, Lorraine, Emmerson, Carl, Frayne, Christine, & Meghir, Costas. 2009. Conditional cash transfers and school dropout rates. *Journal of Human Resources*, **44**(4), 827–857.

Freeman, Jennifer, & Simonsen, Brandi. 2015. Examining the impact of policy and practice interventions on high school dropout and school completion rates: A systematic review of the literature. *Review of Educational Research*, **85**(2), 205–248.

Glick, Peter, & Sahn, David E. 2010. Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis. *The World Bank Economic Review*, **24**(1), 93–120.

Gomes-Neto, Joao Batista, & Hanushek, Eric A. 1994. Causes and consequences of grade repetition: Evidence from Brazil. *Economic Development and Cultural Change*, **43**(1), 117–148.

Helsper, Ellen Johanna. 2012. A corresponding fields model for the links between social and digital exclusion. *Communication theory*, **22**(4), 403–426.

Hsiao, Cheng. 2007. Panel data analysis—advantages and challenges. *Test*, **16**(1), 1–22.

Jacob, Brian A, & Lefgren, Lars. 2009. The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, **1**(3), 33–58.

King, Elizabeth M, Orazem, Peter F, & Paterno, Elizabeth M. 2016. Promotion with and without learning: effects on student enrollment and dropout behavior. *The World Bank Economic Review*, **30**(3), 580–602.

Koppensteiner, Martin Foureaux. 2014. Automatic grade promotion and student performance: Evidence from Brazil. *Journal of Development Economics*, **107**, 277–290.

Leighton, Margaret Alice, Souza, Priscila, & Straub, Stéphane. 2019. Social promotion in primary school: effects on grade progression. *Brazilian Review of Econometrics*.

Manacorda, Marco. 2012. The cost of grade retention. *Review of Economics and Statistics*, **94**(2), 596–606.

Menezes-Filho, N, Vasconcellos, L, Werlang, SR d C, & Biondi, RL. 2008. Evaluating the impact of the Progressão Continuada Program on student flow rates and performance in Brazil. *In: Proceedings of the 13th Lacea Annual Meeting*.

Murnane, Richard J. 2013. US high school graduation rates: Patterns and explanations. *Journal of Economic Literature*, **51**(2), 370–422.

OECD. 2015. *The ABC of gender equality in education: Aptitude, behaviour, confidence.*

OECD, Indicators. 2018. *BRAZIL - Education at a Glance.*

Schwerdt, Guido, West, Martin R, & Winters, Marcus A. 2017. The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, **152**, 154–169.

Stark, Patrick, Noel, Amber M, & McFarland, Joel. 2015. Trends in high school dropout and completion rates in the United States: 1972–2012. Washington, DC: US Department of Education. *National Center for Education Statistics.*

Train, Kenneth E. 2009. *Discrete choice methods with simulation.* Cambridge university press.

UNESCO, Institute for Statistics. 2012. *Opportunities lost: the impact of grade repetition and early school leaving.*

Wu, Wei, West, Stephen G, & Hughes, Jan N. 2010. Effect of grade retention in first grade on psychosocial outcomes. *Journal of educational psychology*, **102**(1), 135.

Xu, Di, Zhang, Qing, & Zhou, Xuehan. 2020. The Impact of Low-Ability Peers on Cognitive and Non-Cognitive Outcomes: Random Assignment Evidence on the Effects and Operating Channels. *Journal of Human Resources.*

Appendix

Table 5: Random Effects Probit parameter estimates: Probability of early dropout

Variable	Full Sample (1)	Social background Rich (2)	Social background Poor (3)	Full Sample (4)	Social background Rich (5)	Social background Poor (6)
approved	-1.184***	-1.131***	-1.205***	-1.175***	-1.132***	-1.196***
	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)
boy	0.027**	0.073**	0.008	0.025**	0.065**	0.005
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
white	0.109***	0.077	0.124***	0.109***	0.078	0.123***
	(0.03)	(0.06)	(0.04)	(0.03)	(0.06	(0.04)
urb	0.271***	0.329***	0.253***	0.243***	0.296***	0.230***
	(0.02)	(0.047)	(0.02)	(0.02)	(0.05)	(0.02)
govaid	-0.055	-0.042	-0.068	-0.048	-0.030	-0.053
	(0.05)	(0.12)	(0.051)	(0.04)	(0.12)	(0.05)
entry age	0.484***	0.527***	0.465***	0.470***	0.521***	0.459***
	(0.01)	(0.02)	0.01)	(0.01)	(0.02)	(0.01)
cohort 2009	-0.196***	-0.253***	-0.177***	-0.188***	-0.239***	-0.169***
	(0.02)	(0.03)	(0.10)	(0.02)	(0.05)	(0.05)
revealed income	-0.075***			-0.079***		
	(0.01)			(0.02)		
elementary school				-0.008	0.014	-0.009
				(0.02)	(0.01)	(0.03)
size room				-0.013***	-0.012***	-0.013***
				(0.01)	(0.01)	(0.01)
lib				-0.079***	-0.162***	-0.038**
				(0.01)	(0.03)	(0.02)
sci				-0.042**	-0.167***	0.001
				(0.02)	(0.03)	(0.02)
sports				-0.023	-0.012	-0.034
				(0.02)	(0.03)	(0.02)
teachers age				0.018***	0.022***	0.018***
				(0.01)	(0.01)	(0.01)
teachers age desviation				-0.006	-0.001	-0.008
				(0.01)	(0.01)	(0.01)
student/staff ratio				-0.001	-0.002	-0.001
				(0.01)	(0.01)	(0.02)
Constant	-7.479***	-8.328***	-7.142***	-7.604***	-8.615***	-7.295***
	(0.181)	(0.37)	(0.20)	(0.23)	(0.48)	(0.26)
σ_u^2	0.526	0.691	0.451	0.527	0.668	0.458
Number of observations	84409	25470	42837	84409	25470	42837
Log likelihood	-36919.72	-10973.164	-25930.72	-36798.27	-10911.99	-25852.94
Estimated probability of dropout	0.226	0.208	0.240	0.225	0.203	0.233
	(0.177)	(0.169)	(0.181)	(0.177)	(0.172)	(0.183)

Note: Standard deviations are in parentheses; * → p < .05; ** → p < .01; *** → p < .001