Does grade repetition increase school dropout rates? Evidence from Senegalese primary schools.

June 18, 2014

JEL Classification: D12, I28, O12
Key words: Grade repetition, Demand for Education, School dropout, Senegal

Abstract

This paper investigates the connection between grade repetition and school outcomes. Pupils need to meet class-specific standards to pass to the next grade, and the specifications measure the differences in the link between learning achievement and grade repetition between classes with different standards. This difference in differences identifies the effect of grade repetition. The results show a negative effect of the grade repetition decision on the probability to be enrolled at school the next year, and on the probability to start secondary school.

This paper also assesses the global consequences of grade repetition policies, which might also affect non-repeating pupils. Classes with tough grade repetition policies have on average more pupils having started secondary school or still enrolled in the follow-up survey. These classes do not seem to be have particularly favorable circumstances for this. So lenient grade repetition policies can plausibly do more harm than good.

1 Introduction

Many Sub-Saharan African pupils drop out before completion of primary school – this is the case of 40% of Senegalese and 32% of African pupils.[1,2]. These figures matter: primary school dropouts lack key skills for their entire lives (e.g. reading or writing). This could be an efficient response to extreme poverty; however, two recent studies show that families take school enrollment decisions under imperfect information (Nguyen, 2008 and Jensen, 2007), so school enrollment decisions are probably inefficient.

Sub-Saharan African countries also face among the highest repetition rates, albeit decreasing : in 2009, 13% of African pupils were repeating their grades.[2] Manacorda (2012) shows that countries with low enrolment rates in secondary school often have more grade repetitions. This paper addresses the question of the causality behind this correlation with Senegalese data. It estimates whether failing a grade increases the chances of school dropout, and assesses whether lenient grade repetition policies can improve primary school completion rates.

High repetition rates in primary schools can decrease national school attainment rates because grade repetition is very expensive for the state and households alike. Indeed, for a given final grade, a grade repetition increases the time spent at school. Hence both private and public costs of schooling increase with a grade repetition, and grade repetition decreases the returns to human capital investments.

Grade repetition can also deter children's self-confidence. Some psychologists consider that early grade repetition affects adversely socio-emotional adjustment. (Jimerson, Carlson, Rotert, Egeland, and Sourie, 1997) In economic terms, grade repetition may be a negative signal about a child's ability. If the parents observe their children's ability noisily, then grade repetition diminishes parents' belief in their children's ability (and/or the children's beliefs on their own abilities).

The pedagogic effect of grade repetition can mitigate its discouraging effect. The pedagogic benefits of grade repetition are nevertheless uncertain. When children repeat grades, they may consolidate the skills taught at those grades. However, it is unclear whether this offsets their failure to acquire the skills taught at the next grade. The net effect of grade repetition on the acquisition of knowledge is therefore ambiguous.

Jacob and Lefgren (2004) measure this effect using a discontinuity in school policy in Chicago. Pupils there took standardized tests at the end of grades 3, 6 and 8. They were promoted if their test score was higher than a minimum score. Regression-discontinuity analysis revealed a small and positive effect of grade repetition on academic achievement after one year. Doing the same with a similar retention policy in Florida, Greene and Winters (2007) find a positive effect of grade repetition in third grade on reading ability after two years. Still in Florida, Schwerdt and West (2012) find this effect nearly fades out after six years. However, they also find the treatment fades out. Indeed, due to subsequent grade repetitions, students retained in grade 6 are on average only 0.74 grade behind similar non-retained peers after six years.

[1] Ministry of Education, Senegal (2005)

[2] www.poledakar.org,

Grade repetition also has consequences until the end of schooling. Jacob and Lefgren (2009) measure this effect with the Chicago test-based policy. They find that grade repetition in grade 8 increases the probability of high-school dropout, while grade repetition in grade 6 does not affect high school dropout.

Grade repetition policies may affect all pupils – including those who do not repeat. Firstly, the threat of grade failure may be an incentive to learn for low-achievement pupils. In addition, when repetition rates are low, repeating may be particularly discouraging for the few pupils who fail to meet the standards – which may increase dropout rates when grade repetition policies are lenient. Similarly, passing may signal a good learning ability only when grade repetition rates are high – which may decrease dropout rates when grade repetition policies are tough.

The literature does not provide with any convincing identification of consequences of grade repetition policies. Jacob (2005) studies the test-based grade repetition policy in Chicago. An accountability policy has simultaneously been implemented, and made teacher and schools accountable for student achievement. He uses a diff in diff strategy, and shows that the policy increased learning achievement in classes where a lot of pupils where likely to repeat ex-ante, and increases learning achievement for at-risk students. This is consistent with the fact that the grade repetition policy changes the incentives in the class. Unfortunately, it is impossible in this case to disentangle the effect of incentives for pupils caused by grade repetition from the effect of the incentives for teachers caused by the accountability policy.

This paper inquires whether frequent early school dropout in Senegal is in part a consequence of high repetition rates. The effect of grade repetition on school attainment in developing countries has been extensively studied with control-based identification strategies.[3] However, conditional on a test-score prior to the grade repetition decision, teacher's grade repetition decisions are probably not random.[4] Manacorda (2012) uses the Uruguayan grade retention policy in junior high schools to estimate the effect of grade repetition on dropout. Pupils automatically repeated when they failed more than 4 subjects. He uses a regression discontinuity design based on the number of failed subjects, and finds that grade repetition decreases school achievement by 1 grade on average.[5]

This paper estimates the effect of grade repetition decision on dropouts in Senegalese primary schools. The context is rather different from Manacorda (2012), as we study primary schools in one of the least developed countries. In contrary to emerging or developed countries, the least developed countries face massive dropout rates in primary schools. In addition, according to Banerjee and Duflo (2011), parents and teachers's beliefs on education are overly elitist in these countries. This can obviously change the consequences of grade repetition: elitist parents may seek for a signal on their children's learning ability, and grade repetition may provide with this signal.

This paper estimates the effect of grade repetition decision on immediate school dropout, and mid-

[3]See, for example, PASEC (2004) or Glick and Sahn (2009) with the same data than this paper or King, Orazem, and Paterno (2008).

[4]Glick and Sahn (2009) claim that for a given test score, differences in grade repetition decisions depend on variations across schools in test score thresholds for promotion. Grade repetition may nevertheless also depend on pupils' motivation at school for a given test score. Motivation at school can be correlated with parental preferences for education.

[5]The number of failed subjects being an integer, one can still discuss the validity of regression discontinuity designs in this context.

term academic achievement, in Senegalese primary schools. It controls for the potential correlation between the children's unobservable characteristics and grade repetition with an original instrumental variables strategy. This strategy is based on a strongly non-linearity of the grade repetition probability between pupils whose learning achievement are above and below the learning achievement to pass to the next grade. I use a difference in differences strategy, between learning achievements and between classes with tough or lenient grade repetition policies, to identify the effect of grade repetition on school dropout.

The results reveal a negative effect of grade repetition on the probability of enrollment at school the next year. The estimated effect is fairly high: the estimations show that grade repetition increases the probability of school dropout by approximately 13 percentage points on average. In the mid-term, the results show that grade repetition decreases the probability to reach secondary school by the follow-up survey.

This paper's second result is that schools with tough grade repetition policies are relatively successful: pupils from these schools are on average more likely to be enrolled during the follow-up survey and to start secondary school. These schools are not located in places with particularly favorable observable characteristics. So tough grade repetition are not necessarily harmful for school outcomes.

Section 2 presents the dataset used to identify the causal effect of grade repetition on school dropout. Section 3 presents the strategies used here for identifying this effect. Section 4 estimates the mid-term effects of grade repetition, and brief remarks are made by way of conclusion.

2 The data

This paper uses two datasets, PASEC[6] and EBMS.[7] Both contain detailed information about schooling and are combined here to estimate the effect of grade repetition.

2.1 The PASEC panel

The PASEC conducted a panel survey among primary school pupils of 98 Senegalese schools between 1995 and 2000. In each school, at the beginning of the 1995-1996 school year, twenty second-grade students were randomly drawn in a randomly drawn second-grade class. The PASEC monitored them throughout their school careers – including grade repetitions – until the first of them finished primary school (sixth grade, in 2000). They took learning achievement tests at the beginning of the first school year and at the end of each school year.

Pupils took the tests when they were at school the day of the tests. So many pupils missed the tests, either because they missed school that day, or because they had dropped out. Grade repetitions are inferred from the tests (which provide information on the grade attended). This information is missing when a pupil missed the tests – section 3.3 explains how the paper deals with this issue.

[6] *Programme d'analyse des systèmes éducatifs* set up by CONFEMEN *Conférence des ministres de l'éducation des pays ayant le français en partage.*

[7] *Education et Bien-être des Ménages au Sénégal.* This survey was designed in collaboration by a team from from LEA-INRA, France and from Cornell University, USA. It was implemented in association with the *Centre de Recherche en Economie Appliquée* (Dakar, Senegal). I thank Sylvie Lambert and Christelle Dumas for having made the data available.

Table 1: Number of children attending the tests during the panel, by grade and school year

					214	**Sixth grade (CM2)**
				357	236	**Fifth grade (CM1)**
			412	204	86	**Fourth grade (CE2)**
		594	154	53	15	**Third grade (CE1)**
789	817	102	no test			**Second grade (CP)**
789	817	696	566	614	551	**Total attendance**
Initial tests (1995)	school year 1995 - 1996	school year 1996 - 1997	school year 1997 - 1998	school year 1998 - 1999	school year 1999 - 2000	

Note: This table reports the attendance among the 921 children of PASEC sample resurveyed by EBMS

Table 1 shows the number of children attending each test in the sample and reveals children often missed a test even though still enrolled. All 921 children were enrolled in school year 1995-1996 although only 817 attended the test.

The PASEC tested the pupils in Math and French. The tests were based on the official curricula, so some questions changed between grades. But the common questions make it possible to compare achievement levels between grades. The PASEC team marked the tests, so teachers cannot influence their pupils' scores.

The pupil questionnaire also included some information about living conditions. In particular, the household wealth index used in this paper is based on this questionnaire.

2.2 EBMS Survey

The EBMS survey provides additional information about a sample of PASEC pupils in 2003. It includes some of the pupils from 59 of the schools surveyed between 1995 and 2000. The objective was to resurvey households in each community (village or urban districts) with children who had been in the PASEC panel. PASEC surveyed 1177 pupils in these 59 schools, 921 are in EBMS data after deletion of questionable matches. Information was collected about the living conditions and educational levels of the household members. EBMS includes retrospective data about the school careers of the PASEC pupils, which discriminate school dropouts from other causes of attrition in the PASEC panel. Consequently, school-leaving dates are known for almost every child re-surveyed (if they had left in 2003). In addition, the EBMS data include parental education. The retrospective information about living conditions includes self-reported shocks on harvests.

2.3 Aggregate dataset

Both datasets provide reliable retrospective information about enrollment. Together they give enough information to reconstruct most instances of grade repetition. This information is necessary to evaluate the impact of repetition on drop out. Another advantage of the aggregate dataset is that it evaluates the individual learning achievement (test scores), which is a crucial determinant of grade repetition. Appendix A gives the definitions of the variables used in this paper and the descriptive statistics table.

3 Empirical strategy and results

This paper seeks to identify the effect of grade repetition – denoted R_{ik} – on school dropout, which is the coefficient γ in the equation (1) below – enrolment during the next school year is denoted $E_{ik,t+1}$ for child i of group k[8], at date $t+1$. The other determinants of dropout are the PASEC test score S_{ik}, and a vector of covariates X_{ik}.[9]

$$E_{ik,t+1} = \mathbb{1}\left[\beta_{e1} S_{ik} + X_{ik}\beta_{e2} + \gamma R_{ik} + u_{ik} > 0\right] \tag{1}$$

The main difficulty in identifying γ is to control for the potential endogeneity of grade repetition.

3.1 Identification strategy

This paper uses an original instrumental variables strategy to control for the potential correlation between the children's unobservable characteristics and grade repetition. This strategy uses the widespread idea that a certain learning achievement is required to pass to the next grade. In Senegal, teachers (and the school hierarchy) take grade repetition decisions, so these "teacher's standards" vary between classes.

Unfortunately, teacher's standards are not observed, and this paper uses the "test score of the last passer" to proxy for them. "Passers" are those peers of a given pupil in a given year who are admitted to the next grade. Among the passers, the pupil with the lowest test score is called the last passer. Her test score is denoted LP_{-ik}:

$$LP_{-ik} = min_{\{j \neq i, R_{jk}=0\}}(S_{jk}) \tag{2}$$

So a pupil supposedly meets the standards when $S_{ik} - LP_{-ik} > 0$. But the sign of $S_{ik} - LP_{-ik}$ explains only partially grade repetition decisions: absenteeism and motivation at school can affect grade repetition decisions, S_{ik} differs from the teacher' evaluation of the pupil's learning achievement, and LP_{-ik} is probably a noisy measure of the teacher's standards. In the end, grade repetition is a non-linear function of the difference between own achievement and teacher's standards. The grade repetition equation writes:

$$R_{ik} = \mathbb{1}\left[f_r(S_{ik} - LP_{-ik}) + \beta_{r1} S_{ik} - \beta_{r2} LP_{-ik} + X_{ik}\beta_{r4} + \epsilon_{ik} < 0\right] \tag{3}$$

f_r is a non-parametric function. In practice, it is approximated with dummy variables for intervals of $S_{ik} - LP_{-ik}$ (piecewise linear splines give the same results). This paper uses the difference between

[8] A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

[9] This vector includes grade-year dummies, household wealth parents' education, and group mean test score when not included in the model.

own achievement and teacher's standards as an instrument for grade repetition, and measures the effect of grade repetition on dropout with model (4):

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[\quad \gamma R_{ik} + \beta_{e1} S_{ik} + \beta_{e2} LP_{-ik} + X_{ik}\beta_{e4} + u_{ik} > 0] \\ R_{ik} & = \mathbb{1}[\quad f_r(S_{ik} - LP_{-ik}) + \beta_{r1} S_{ik} - \beta_{r2} LP_{-ik} + X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \end{cases} \tag{4}$$

The identification of the parameters in the $E_{ik,t+1}$ equation is assured by the simple non-linearity of the two equations. However, this is better handled via an exclusion restriction, i.e. a variable that influences the R_{ik} equation but not the $E_{ik,t+1}$ equation. In model (4), $f_r(S_{ik} - LP_{-ik})$ is this exclusion restriction; we assume $S_{ik} - LP_{-ik}$ does not affect directly school dropout. This assumption is obviously crucial, but it is useful to start the discussion with a few words on the main control variables.

The model controls for S_{ik} and compares pupils with similar learning achievement, and with presumably similar background in terms of unobservable characteristics. The model controls for LP_{-ik} and includes an additive effect of teacher's standards on dropout (and on grade repetition): teacher's standards have pedagogic consequences and may in turn affect school dropout. These effects are linear in model (4), but this assumption can be relaxed – see the comparison of Tables C.2 and C.3 in the end of section 3.4.

Model (4) controls additively for S_{ik} and LP_{-ik}, so the function $f_r(S_{ik} - LP_{-ik})$ is identified with the interaction between the two: the differences in the link between S_{ik} and grade repetition between different levels of LP_{-ik}. In sum, we measure a difference in differences, between the high-achievement pupils and the low-achievement pupils and between classes with different teacher's standards. In other words, the paper uses the fact that low-achievement pupils are much more vulnerable to "tough" teachers (teachers with high standards).

In the equation of interest, the effect of grade repetition on dropout in model (4) is identified with the same variations of grade repetition probability. The paper is based on the comparison of the correlation between last passer's score and enrollment for low-achievement and high-achievement pupils. This difference in differences identifies the effect of grade repetition on dropout in model (4).

This difference in differences measures the (non-linear) effect of the distance to teacher's standards on grade repetition and on dropout (through $f_r(S_{ik} - LP_{-ik})$). It calls the latter "effect of grade repetition on dropout". The control for test scores ensures that we measure an effect of the distance to teacher's standards on dropout, and not an effect of learning achievement on dropout (or an effect of teacher's standards on dropout).

However, if the distance to teacher's standards affects dropout – e.g. if the parents observe it independently from grade repetition –, this threatens the identification assumption. This would change the interpretation of the results marginally: the results would identify an effect of the distance to teacher's standards on dropout, rather than the effect of grade repetition on dropout. The main message of the paper would remain virtually unchanged: for a given learning achievement, school failure hurts pupils' outcomes – instead of grade repetition hurts pupils' outcomes (the distinction is that school failure can be a continuous variable while grade repetition is binary).

3.2 Identification questions

Peer effects The extensive literature on peer effects in education economics emphasizes the potential interactions between peers' unobservable characteristics. Interaction between peers can be problematic here: a characteristic of the peers – the last passer's score LP_{-ik} – measures teacher's standards. I believe the sign of this bias is predictible: peer effects can only attenuate the estimated effect of grade repetition on dropout estimated here.

Grade repetition decisions are probably partly based on a relative evaluation of pupils in Senegal (see the discussion of Table C.1 in the end of this section). So when i has favorable unobservable characteristics, i's peers are more likely to repeat, and on average, LP_{-ik} increases (and overestimates teacher's standards). Besides, teachers may compare similar pupils before taking grade repetition decisions. So pupil's favorable unobservables may increase LP_{-ik} especially when the pupil is at risk of grade repetition.

The estimations presented in this paper measure a difference in difference: a difference in the link between LP_{-ik} and grade repetition between pupils with high and low test-scores. They measure the effect of the ability to meet teacher's standards on school outcomes. With peer effects, the specifications may underestimate the ability to meet teacher's standards for pupils with favorable characteristics (if favorable unobservables increase LP_{-ik} when grade repetition is plausible). This can decrease the dropout rates of pupils seemingly unable to meet teacher's standards. I observe that pupils seemingly unable to meet teacher's standards are the most likely to dropout; the effect of unobservable characteristics on LP_{-ik} might attenuate this relationship.

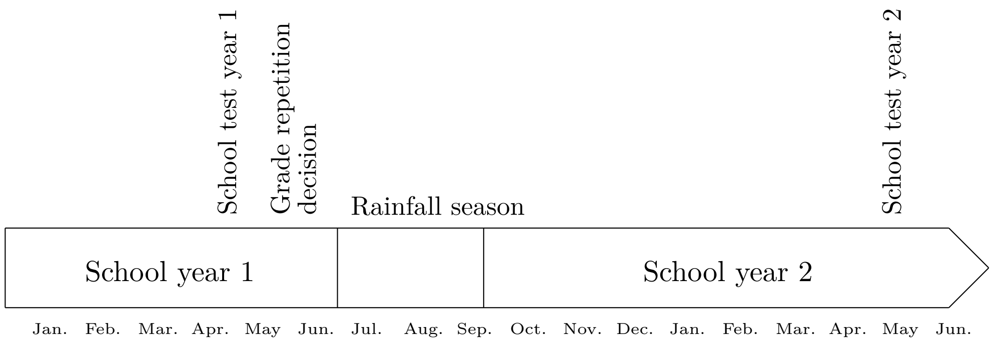
Is it possible to measure the effect of teacher's standards? This paper mainly studies the effect of grade repetition on school dropout. However, grade repetition policies (measured by teacher's standards in this paper) can affect all the pupils; and this section discusses whether this paper can measure the consequences of grade repetition policies in general.

The direct effect of LP_{-ik} measures the direct effect of teacher's standards in model (4). In this paper, it is not possible to claim the variations in LP_{-ik} are due to a well identified and exogenous source; however, no observable mechanism makes grade repetition practices endogeneous.

Conditionally on test scores, grade repetition practices seems uncorrelated with individual and location characteristics. In appendix, Table C.1 shows that the proxy for teacher's standards strongly depends on group mean test score. The coefficient in the linear regression is 1: teacher's standards increase by one point when the class average score increases by one point – grade repetition rates are hardly correlated with average learning achievement. Teacher's standards are insignificantly correlated with average progression, observable community characteristics and observable household characteristics. So grade repetition practices seem uncorrelated with the quality of teachers or with the context.

However, all the traits of teacher's pedagogy are likely to be correlated with each other, and to cause dropout. Conditional on test scores, grade repetition practices may be one of these traits, and hence correlated with other determinants of dropouts. To illustrate this, assume dropout rates are lower with tough teachers. It is hard to say whether this is due to their grade repetition practices or to some other trait of tough teachers. In the general case, it is hard to give the sign of the potential bias.

Figure 1: Sequence of the main events during the PASEC panel



In the end, it is not possible to ensure the coefficients of LP_{-ik} in the estimations are causal. The paper still comment these coefficients: nothing prove their endogeneity, and the literature mostly ignores the consequences of grade repetitions policies.

3.3 Selection issues

Selection on test participation In 1995, children were randomly selected among the second grade pupils of the schools. However, there is attrition in the panel, and this attrition may cause a selection bias. This paper does not correct this potential bias, but is sign is predictable: it would attenuate the estimated effect of grade repetition on dropout.

Our regressions are based on test scores: attrition in the panel is due to pupils who did not take the PASEC tests. Each school year, the PASEC team organized the tests on a given day in each school. Children missing school that day or no longer attending the surveyed school were not tested. So attrition in the panel is probably mostly due to school dropout or irregular school attendance.

The paper finds that pupils lagging behind teacher's standards are more likely to drop out. Pupils lagging behind teacher's standard can be less likely to take the tests: pupils who expect to fail may anticipate a school dropout and miss school the day of the test. Hence some school dropouts due to grade repetition may be selected from the sample because they have anticipated their dropout. The magnitude of the effect of grade repetition on dropout may therefore be underestimated.

Selection on grade repetition observation Some grade repetition decisions cannot be observed in the EBMS-PASEC data. This selection can nevertheless be controlled for.

The information for grade repetition is mostly inferred from this longitudinal information on the school careers. Figure 1 summarizes the timing of the PASEC panel survey. We know the grade repetition decision at the end of school year t when a child took the tests in school years t and $t+1$.[10] Grade repetition decisions are unknown when a child dropped out immediately after this decision: grade repetition is inferred from the school career, so there is no way of knowing the grade repetition decision of school year t when a child dropped out before the tests of school year $t+1$. The structure of the data is therefore summarized in Table 2.

[10]The details and other cases are explained in appendix A.

Table 2: Observation of grade repetition decision

date t	date $t+1$	
Enrolled	Enrolled	} Grade repetition decision is observed
	Enrolled	} Grade repetition decision is not observed: did not take the tests of year t+1
	Drops out	

This selection problem threatens the identification of the effect of grade repetition on school dropout: in case of school dropout, grade repetition is not observed. Model (5) shows it is possible to control for this selection, and to measure the effect of grade repetition on school dropout.

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[\ \beta_{e1}S_{ik} & +\beta_{e2}LP_{-ik} & +\beta_{e3}Z_s & +\gamma R_{ik} & +X_{ik}\beta_{e4}+u_{ik}>0] \\ R_{ik} & = \mathbb{1}[\ \beta_{r1}S_{ik} & -\beta_{r2}LP_{-ik} & +\beta_{r3}Z_R & & +X_{ik}\beta_{r4}+\epsilon_{ik}<0] \\ & & & & & \\ selection & = \mathbb{1}[\ \beta_{s1}S_{ik} & +\beta_{s2}LP_{-ik} & +\beta_{s3}Z_s & +\gamma_s R_{ik} & +X_{ik}\beta_{s4}+v_{ik}>0] \end{cases} \quad (5)$$

The timing of the PASEC panel survey provides an exclusion restriction for selection: rainfall shocks, denoted Z_s in model (5). Z_s is a dummy taking value 1 when the household head reported negative shocks on harvests during the calendar year. Indeed, the rainfall season happens to be after the end of the school year. Hence rainfalls cannot affect either grade repetition decisions or socioeconomic characteristics at the date of the grade repetition decision. Rainfalls can obviously affect subsequent grade repetition decisions and socioeconomic characteristics, but this exclusion restriction is unnecessary in model (5).[11] To simplify the notation, Z_R denotes $f_r(S_{ik} - LP_{-ik})$ in this section.

Appendix D.1 proves that in model (5):

- If $(\epsilon_{ik}, u_{ik}, v_{ik})$ is independent of $(S_{ik}, LP_{-ik}, Z_R, Z_s, X_{ik})$
- If f_r is not constant and $\beta_{s3} \neq 0$
- Under technical assumptions[12]

the coefficients of model (5) could be identified without assumption on the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$.

Appendix D.2 even shows that under much simplier hypotheses and without Z_s, the sign of the effect of grade repetition on dropout is still identified. The intuition for that is rather simple. Indeed, the derivative of the grade repetition probability with respects to Z_R gives the sign of β_{r3} regardless of selection. The derivative of the grade repetition probability with respects to Z_R gives the sign of $\beta_{r3}\gamma$. So the effect of grade repetition on enrollment is positive when the two derivatives have the same sign, and negative if they have opposite signs.

This paper does not intend to identify model (5) semiparametrically. All the models in this paper are estimated using a standard maximum likelihood method. However, this result shows that there is enough information to identify the effect of grade repetition on dropout in the EBMS-PASEC

[11] The panel structure of the data is not handled via fixed effects, so strict exogeneity is not required. Instead, clustering corrects the correlation between the different observations of the same child.

[12] Hypotheses about the support of the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$ and of the distribution of the observables.

data without parametric assumption. Hence appendix D.1 suggests the results rely essentially on the information from the data, and not so much on the distributional assumptions of the models.

3.4 Main results

Table 3 gives the estimation of model (5). The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator. The three columns of Table 3 correspond to the model's three equations. The data are pooled for the various grades and years. Each specification includes grade-year dummies in each equation.

Determinants of grade repetition In the grade repetition equation, the difference between own learning achievement and last passer's score is strongly correlated with grade repetition. The corresponding variables are strongly significant, the χ^2 test for the significance is about 30. The magnitude of the effect is substantial. The average grade repetition risk predicted by the model is 70% if all the pupils are allocated to the group $-1 < S_{ik} - LP_{-ik} < -0.75$. The average grade repetition risk predicted by the model is 10% if all the pupils are allocated to the group $1.5 < S_{ik} - LP_{-ik}$.

The effect of test score on enrollment is not significant in the grade repetition equation. This means that the effect of learning achievement on grade repetition is entirely captured by the other coefficients in the regression: grade repetition probability is not a function of learning achievement *per se*, but a function of the difference between learning achievement and teacher's standards. Similarly, last passer's test score does not seem to affect grade repetition likelihood, which means that its effect is captured by the difference between learning achievement and last passer's score.

The effect of grade repetition on school dropout Table 3 finds a negative effect of grade repetition on dropout. The average marginal effect is -14%, which is impressive given that the average dropout rate is 2% in our sample. The specification broadly predicts that all the dropouts have failed to pass their grade. Indeed, in the fitted model, the sample average of the probability of dropout without grade repetition is 0.15% and the sample average of the probability of dropout with grade repetition is 14%.

The magnitude of the effect of grade repetition on dropout is credible. First, all the dropouts observed here took place during primary school, and not between primary school and secondary school. School dropouts during school cycles are the most likely to be due to grade repetition. In addition, during the follow-up EMBS survey in 2003, household members told the reason of their dropout. Most of the dropouts in our sample[13] told why they dropped out. According to them, 60% dropped out because of school failure, and 27% for "other reasons". Only 13% of them gave another answer. It is therefore credible that dropout within a school cycle is massively associated with grade repetition.

The proofs on identification of model (5) given in Appendix D.1 suggest the results presented in Table 3 rely on the information from the data, and not so much on the distributional assumptions of the error terms. So as to make it even clearer, it is possible to identify the reduced-form results

[13] Those who lived in 2003 in the same household than during the PASEC panel starting in 1995.

Table 3: Joint estimation of the determinants of grade repetition, selection and school dropouts corresponding to the model (5)

	repetition	$enrolled_{t+1}$	*selection*
	(1)	(2)	(3)
Test score	-.223	-.023	-.238
	(.223)	(.149)	(.123)*
LP_{-ik}	.074	.379	.305
	(.207)	(.137)***	(.095)***
$S_{ik} - LP_{-ik} < -1$	.695		
	(.407)*		
$-1 < S_{ik} - LP_{-ik} < -0.75$	1.131		
	(.347)***		
$-0.75 < S_{ik} - LP_{-ik} < -0.5$	.631		
	(.262)**		
$-0.5 < S_{ik} - LP_{-ik} < -0.25$	.485		
	(.193)**		
$-0.25 < S_{ik} - LP_{-ik} < 0$	.269		
	(.155)*		
$0 < S_{ik} - LP_{-ik} < 0.25$	Ref.		
$0.25 < S_{ik} - LP_{-ik} < 0.5$	-.379		
	(.159)**		
$0.5 < S_{ik} - LP_{-ik} < 0.75$	-.451		
	(.192)**		
$0.75 < S_{ik} - LP_{-ik} < 1$	-.401		
	(.226)*		
$1 < S_{ik} - LP_{-ik} < 1.5$	-.492		
	(.274)*		
$1.5 < S_{ik} - LP_{-ik}$	-.815		
	(.451)*		
Negative shock on harvests		.150	.569
		(.254)	(.207)***
Grade repetition		-2.259	-2.740
		(.337)***	(.463)***
Average marginal effect of grade repetition		*-.138*	
		*(.051)****	
Household wealth and Parents' education, Previous year's test score, Group[a] mean test score	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes
Obs.	1818	1818	1818
log likelihood	-1258.516	-1258.516	-1258.516
χ^2 exclusion restrictions	30.290		7.575
corresponding p value	.0008		.006

The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator (25 iterations). Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.

a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

corresponding to Table 3. Appendix C.2 estimates this model, and shows there is a link between school dropout and the position relative to the last passer in a reduced-form estimation. This link mirrors the link between the position relative to the last passer and grade repetition (see Figure C.1).

Group affected by the instrument and LATE Instrumental variable estimations measure an effect of the treatment for those individuals whose endogeneous variable is affected by the instrument. Here, the estimations measure the effect of grade repetition on dropout for pupils whose grade repetition is affected by teacher's standards. It excludes the best pupils, who pass regardless of the teacher. It probably excludes some pupils with very low learning achievement, who would repeat anyways. So our estimates implicitly measure the consequences of homogenizing teachers' standards to the most lenient ones. This is close to a policy-relevant average treatment effect, which would be to make teachers' standards more lenient. There is nevertheless a distinction, as we cannot know what happens when the most lenient teachers make their standards even more lenient.

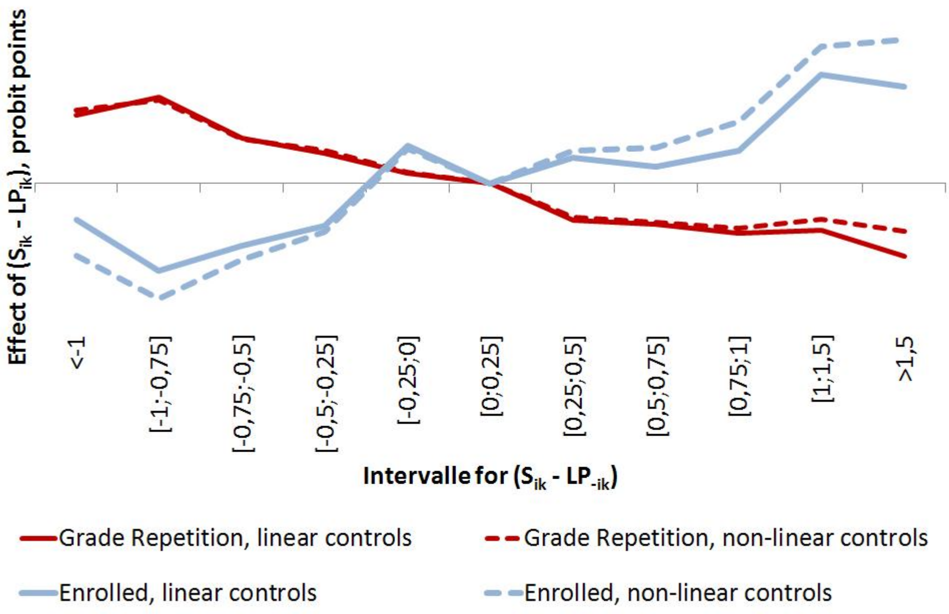
Potential direct effect of grade repetition practices Teacher's standards may directly affect dropout. In the enrollment equation, LP_{-ik} – the proxy for teacher's standards – increases the probability of being enrolled at school the next year. It is necessary to be cautious with a causal interpretation of this coefficient. However, should this interpretation be valid, this changes the consequences of a limitation of grade repetition. Indeed, the dropout rate may decrease when the number of grade repetitions decreases, but the dropout rate would increase when LP_{-ik} increases. Simulations presented in section 5 show the two effects seem to have the same magnitude.

Determinants in the selection equation The estimation of selection in model (5) is intended to control for selection bias in the observation of R_{ik}. The determinants of selection may be the determinants of moving or missing school the day of the tests in addition to the determinants of dropout. Accordingly there is no particular interpretation of these coefficients.

Nevertheless, it is necessary to focus on the effect of the negative shocks on harvests, since this variable is the exclusion restriction in the equation for R_{ik}. These shocks positively affect selection: when there is a negative shock, the child is more likely to take the test the next year. Negative shocks on harvests may decrease opportunity costs, so children may be more likely to take the tests when there is a shock. The F-test for the significance of this instrument is 7.2.

Non-linearities in the coefficients of other variables In Table 3, the effect of the difference $S_{ik} - LP_{-ik}$ is non-linear, and we assume that the effect of all the other explanatory is linear (in the latent variable). Hence, our measure of the effect of $S_{ik} - LP_{-ik}$ may catch some other non-linearities of the model. It is nevertheless possible to check this is not the case. In Table C.3, the effect of some of the other explanatory variables is treated as non-linear in a reduced-form estimation. The results are plotted in Figure 2, and compared to the estimates of Table C.2 (which is a reduced-form estimation with linear controls in the latent variable). Both estimations are quasi-identical in Figure 2. In addition, the non-linearities added in Table C.3 are not statistically significant.

Figure 2: Comparison between the reduced-form effect of $S_{ik} - LP_{-ik}$ on grade repetition and on dropout, with and without non-linear control variables (respectively Table C.2 and Table C.3)



Notes: Linear means linear in the latent variable. Plot of the estimates of Table C.2 and Table C.3. The specification with non-linear controls controls for a non-linear effect of own test score, group mean test score, difference between own test score and group mean test score, last passer's test score.

4 Mid-term consequences of grade repetition

Section 3 identifies the short-term effect of grade repetition on dropout. This effect is of limited interest if not persistent. For example, grade repetition could push out of school only pupils who intended to stay in school for a single additional school year. The retrospective information on education in the EBMS data can give an insight on this. Precisely, grade repetitions studied in this paper take place between 1996 and 2000, and the EBMS can give information on the retrospective school trajectory as of 2003. This section assesses the mid-term consequences of grade repetition. Table 4 uses the same specifications as Table 3, changing the dependent variable.

$$\begin{cases} S_{ik,t+\delta} = \mathbb{1}[\ \beta_{e1}S_{ik} + \beta_{e2}LP_{-ik} + \beta_{e3}Z_s + \gamma R_{ik} + X_{ik}\beta_{e4} + u_{ik} > 0] \\ \\ R_{ik} = \mathbb{1}[\ \beta_{r1}S_{ik} - \beta_{r2}LP_{-ik} + f_r(S_{ik} - LP_{-ik}) + X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ selection = \mathbb{1}[\ \beta_{s1}S_{ik} + \beta_{s2}LP_{-ik} + \beta_{s3}Z_s + \gamma_s R_{ik} + X_{ik}\beta_{s4} + v_{ik} > 0] \end{cases} \tag{6}$$

In equation (6), $S_{ik,t+\delta}$ is an outcome related to school enrollment of child ik at date $t+\delta$, δ years after the grade repetition decision. Table 4 estimates this model with 8 variables. $enrolled_{t+1}$ is the same variable as in Table 3, and Table 4, column 1 recalls the main quantitative results of Table 3.

Table 4 measures the effect of grade repetition on enrollment 2 years, 3 years and 4 years after the

Table 4: Effect of grade repetition on short-term and mid-term outcomes - Full Sample

	$enrolled_{t+1}$	$enrolled_{t+2}$	$enrolled_{t+3}$	$enrolled_{t+4}$	*Still enrolled* (2003)	*Last Grade Attended* ≥ 6	*Last Grade Attended* ≥ 7	*Last Grade Attended* ≥ 8
	(1)[a]	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LP_{-ik}	.379	.128	.087	.135	.249	.176	.296	.246
(enrollment equation)	(.137)***	(.187)	(.148)	(.124)	(.075)***	(.092)*	(.076)***	(.080)***
Grade repetition	-2.259	-.609	.193	.072	-.442	-.423	-.668	-.525
(enrollment equation)	(.337)***	(.880)	(.516)	(.442)	(.270)	(.360)	(.296)**	(.270)*
Average marginal effect of	*-.138*	*-.055*	*.027*	*.013*	*-.156*	*-.118*	*-.189*	*-.119*
repetition (enrollment eq.)	*(.051)****	*(.098)*	*(.073)*	*(.082)*	*(.098)*	*(.107)*	*(.083)***	*(.058)***
Obs.	1818	1449	1449	1449	1789	1777	1777	1777
log-likelihood	-1258.516	-1117.741	-1267.688	-1382.891	-2197.15	-1942.983	-2023.004	-1889.06
χ^2 instruments	30.290	21.903	25.626	20.998	28.988	28.196	27.124	28.886
corresponding p value	.0008	.016	.004	.021	.001	.002	.002	.001
Simulation: LP_{ik}decreases by 0.25 pts, no "direct effect of grade repetition policy"[b]	.005	.002	-.001	-.0006	.006	.005	.007	.004
	(.002)***	(.004)	(.004)	(.004)	(.004)	(.004)	(.003)**	(.002)**
Simulation: LP_{ik}decreases by 0.25 pts, with "direct effect of grade repetition policy"[b]	.0004	-6.06e-06	-.005	-.007	-.015	-.006	-.014	-.011
	(.001)	(.002)	(.003)	(.004)*	(.005)***	(.005)	(.004)***	(.004)***

Notes: The Table reports the results of trivariate specifications as in Table 3 (model (5)), with different dependent variables in the enrollment equation.
***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.
a: Recalls the estimates of Table 3.
b: The simulations assess the consequences of decreasing LP_{-ik} on grade repetition, and measures the indirect effect on enrollment due to "the effect of grade repetition on enrollment"
c: The simulations assess the consequences of decreasing LP_{-ik} on grade repetition. It measures the sum of the direct effect of LP_{-ik} measured in the enrollment equation and of the indirect effect on enrollment due to "the effect of grade repetition on enrollment"

grade repetition decision and finds no effect. It finds no effect of grade repetition on the probability to be enrolled at school during the follow-up survey in 2003.

A grade repetition increases the age for grade. Hence, if grade repetition does not affect dropout date in the mid-term, it may still affect the last grade attended. Table 4 finds that grade repetition decreases the likelihood to reach grade 7 (the first grade of secondary school) and grade 8 until 2003 by 19 and 12 percentage points. However, these estimates may be polluted by completions of grade 7 and 8 posterior to 2003.

The specifications in Table 4 also measure whether our proxy for grade repetition practices (LP_{-ik}) is correlated with long-term achievement (conditionally on grade repetition). Indeed, it is positively correlated with the likelihood to be enrolled in 2003, and with the likelihood to reach grade 5, 6 and 7. Again, there is no strict evidence that this is a causal effect. Should it be a causal effect, the estimated negative effect of lenient grade repetition practices would override the estimated positive effects, as shown in section 5.

5 Simulations

In the enrollment equations, LP_{-ik}, the proxy for grade repetition practices, is associated with lower school outcomes. This coefficient has to be interpreted with caution, as it is potentially endogenous. This section nevertheless assesses whether this coefficient can change the consequences of lenient grade repetition practices. As shown below, this seems to be the case: if the coefficient on LP_{-ik} is exogeneous, lenient grade repetition policies deteriorate school outcomes.

The simulations in the last 2 lines of Table 4 estimate the consequences of lenient grade repetition policies, with a decrease of LP_{-ik} by 0.25 pt. This represents a decrease in the share of failing students from 26% to 22%.

The first row of simulations shows the benefits of lenient grade repetition practices due to the effect of grade repetition. In equation (6), the simulation assumes lenient grade repetition policies affect the grade repetition probability ($f_r(S_{ik} - LP_{-ik})$ and $\beta_{r2}LP_{-ik}$). On the other hand, $\beta_{e2}LP_{-ik}$ is held constant. (Details are given in appendix B)

The simulations find that lenient grade repetition policies would increase the probability to be enrolled the next year by 0.5 percentage points, and the probability to reach the first and second grade of secondary school (grade 7 and 8) by 0.7 and 0.4 percentage points. This is coherent with the effect of grade repetition on dropout measured by the corresponding specifications.

The second row of simulations in Table 4 shows the consequences of lenient grade repetition practices with a direct effect of grade repetition practices. In equation (6), these simulation assumes lenient grade repetition policies affect the grade repetition probability ($f_r(S_{ik} - LP_{-ik})$ and $\beta_{r2}LP_{-ik}$) and directly affect the enrollment equation (via $\beta_{e2}LP_{-ik}$). They therefore assume the direct effect of lenient grade repetition practices is measured by the coefficients of LP_{-ik} in the first row of Table 4.

The results are in the opposite direction: lenient grade repetition practices would decrease the share of pupils enrolled in 2003 by 1.5 percentage points, and decrease the share of pupils reaching grades 7 and 8 by 1.4 and 1.1 percentage points.

Table C.4 checks the reduced-form counterpart of these simulations. In other words, it checks the correlation between mid-term outcomes and grade repetition practices (with probit specifications). It finds the same results: conditionally on test scores, tough grade repetition practices (high LP_{-ik}) are correlated with better long-term outcomes.

These simulations are based on fragile assumptions. Indeed, they assume the exogeneity of teacher's standards. This assumption is hard to test, but the estimations show that the benefits associated to lenient grade repetition policies are uncertain. Indeed, the fact that schools with lenient grade repetition policies have lower school outcomes on average is not encouraging.

6 Conclusion

This paper assesses whether grade repetition can deteriorate school outcomes. Its instrumental strategy measures the differences in the link between learning achievement and grade repetition between classes with different standards to pass to the next grade. This difference in differences identifies the effect of grade repetition, and shows grade repetition negatively affects school outcomes: grade repetition decreases the probability to be enrolled at school the next year and the probability to start secondary school.

However, grade repetition policies can have other consequences than affecting repeating pupils, and it is hard (and probably not achieved in this paper) to find a convincing statistical identification of these effects. Schools with tough grade repetition policies emphasize similar or better mid-term outcomes than other schools. I did not find any evidence that their environment is particularly favorable to these outcomes.

These results are interesting for several reasons common to many of the least developed countries. First, it is the first paper using an instrumental strategy to measure the effects of grade repetition in the least developed countries. In countries with massive primary school dropout rates, grade repetition can strongly affect individual trajectories, and in turn monetary or non-monetary poverty.

This is totally coherent with the fact that lenient grade repetition policies could remain average dropout rates unchanged. The relevance of grade repetition policies depends these two channels: whether grade repetitions hurt pupils, and whether grade repetition policies change aggregate outcomes. A statistical identification of the latter would fill an important gap in the literature.

References

Banerjee, A. V., Duflo, E., 2011. Poor economics: A radical rethinking of the way to fight global poverty. PublicAffairs.

Filmer, D., Pritchett, L., 2001. Estimating wealth effects without expenditure data - or tears: en application of educational enrollment in states of india. Demography 38, 115–132.

Glick, P., Sahn, D., 2009. Ability, grade repetition and school attainment in Senegal: A panel data analysis, SAGA Working Paper, Cornell University.

Greene, J. P., Winters, M. A., 2007. Revisiting grade retention: an evaluation of florida's test-based retention policy. Journal of School Psychology 2, 319–340.

Jacob, B., Lefgren, L., 2004. Remedial education and student achievement: A regression - discontinuity analysis. Review of economics and statistics 86, 226–244.

Jacob, B. A., 2005. Accountability, incentives and behavior: the impact of high-stakes testing in chicago public schools. Journal of Public Economics 89, 761–796.

Jacob, B. A., Lefgren, L., 2009. The effect of grade retention on high school completion. American Economic Journal: Applied Economics , 33–58.

Jensen, R., 2008. The perceived returns to education and the demand for schooling, mimeo, Watson Institute for International Studies, Brown University.

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., Sourie, L. A., 1997. A prospective, longitudinal study of the correlates and consequences of early grade retention. Journal of School Psyqhology 35, 3–25.

King, E. M., Orazem, P. F., Paterno, E. M., 2008. Promotion with and without learning: Effects on student dropout. Working paper, World Bank.

Manacorda, M., 2012. The cost of grade retention. Review of Economics and Statistics 94, 596–606.

Manski, C., 1988. Identification of binary response models. Journal of the American Statistical Association 83, 729–738.

Ministry of Education, Senegal, 2005. Situation des indicateurs de l'education 2000 - 2005. Tech. rep., Ministry of education, Senegal.

Nguyen, T., 2008. Information, role models and perceived returns to education: Experimental evidence from Madagascar, mimeo, Massachussets Institute of Technology.

PASEC, 2004. Le redoublement : pratiques et conséquences dans l'enseignement primaire au sénégal. Tech. rep., Conferency of the Ministers of Education of French-speaking countries.

Schwerdt, G., West, M. R., 2012. The effects of early grade retention on student outcomes over time: Regression discontinuity evidence from Florida .

Vytlacil, E., Yildiz, N., 2007. Dummy endogenous variables in weakly separable models. Econometrica 75, 759–779.

A Variables

A.1 Dependant variables

Enrolled is the fact that the child is still enrolled at school in a given year. The information is inferred from the EBMS dataset so as to distinguish attrition in the panel from school dropout.

Last grade is the last grade attended. Grade 6 is the last grade of primary school. The information is inferred from the EBMS dataset.

Repetition is a dummy taking value 1 if the child repeated the grade, and 0 otherwise. Information is from the PASEC panel. In each case, I tried to infer each year whether the child passed at the end of the school year. Table A.2 sums up the various possible cases in the PASEC data and specifies whether anything can be learned about the child's progression. Case 1 is the basic case: the child took all the tests. She repeated after school year 1995 - 1996, and has passed all the subsequent grades. In case 2, the child did not take the tests in 1996 - 1997. The reason why she did not take the test is not reported. Consequently, whether she repeated the second or the third grade is unknown. In case 3, the child dropped out in 1996. Consequently whether she was admitted to third grade after school year 1995 - 1996 is unknown. In case 4, the child is not in the sample after 1997 - 1998, so whether she repeated during the subsequent grades remains unknown. In cases 5 and 6, grade repetitions are not ambiguous: we know the child repeated twice (case 6) or passed twice (case 5) when she was not observed.

A.2 Test scores

Test scores are a proxy for learning achievement at the end of the current school year. In fact the PASEC panel contains school tests at the end of each academic year until the end of the survey.[14] The tests were marked by the PASEC team. Consequently, test scores could not be influenced by teachers. Table 1 reports the number of children taking each test.

The tests were designed to ensure easy comparisons within grade-years. They nevertheless differed between different grades and years of the panel. The test scores have a mean of 0 and a standard deviation of 1 within each grade-year.

Group mean test score A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Last passer's test score is a proxy for teacher's standards, i.e. the learning achievement is required to pass to the next grade. "Passers" are those peers of a given pupil in a given year who are admitted to the next grade. Among the passers, the pupil with the lowest test score is called the last passer. Her test score is denoted LP_{-ik}:

[14] The second grade classes were not surveyed from 1997 - 1998, so pupils still in this grade at that time were not surveyed until they passed the third grade.

Table A.1: Descriptive statistics

	N	mean	standard deviation	min.	max.
Enrolled next year	1823	.979		0	1
$enrolled_{t+2}$	1454	.959		0	1
$enrolled_{t+3}$	1454	.920		0	1
$enrolled_{t+4}$	1454	.882		0	1
Still enrolled (2003)	1794	.673		0	1
Last Grade > 5	1782	.749		0	1
Last Grade > 6	1782	.437		0	1
Last Grade > 7	1782	.291		0	1
Grade repetition	1823	.148		0	1
Selection (on grade repetition observation)	1823	.867		0	1
Negative shocks on harvests	1823	.087	.308	0	2
Test score	1823	-.055	.953	-3.20	3.34
LP_{-ik}	1818	-.770	.849	-3.20	2.63
Previous year's test score	1823	.00481	1.01	-2.34	3.81
Group[a] mean test score	1823	-.0709	.536	-1.58	1.91
Group[a] mean progression	1823	-.0840	.433	-1.55	1.09
Household wealth	1823	-.591	2.07	-3.12	4.38
Parent's education	1823	2.05	1.47	1	8
Head is not Muslim (Christian or Animist)	1820	.0335		0	1
Ethnic group of the head: Wolof	1813	.398		0	1
Ethnic group of the head: Pulaar-Halpulaar	1813	.187		0	1
Ethnic group of the head: Serere	1813	.250		0	1
Ethnic group of the head: Dioola	1813	.0281		0	1
Ethnic group of the head: Mandingue-Sose	1813	.102		0	1
Ethnic group of the head: Others	1813	.0342		0	1
Community: mean household wealth index	1823	-.642	1.61	-2.93	2.94
Community: mean education index	1823	2.20	.730	1.19	4
log(Village or city population)	1714	10.0	2.81	5.60	14.6
Community main activity: trade (ref:agri.)	1823	.404		0	1
Community has electricity	1823	.813		0	1
Rural	1823	.535		0	1
Distance to health center	1823	.0730	.260	0	1
Distance to hospital	1823	1.63	1.26	0	3

Notes: Standard deviations are not reported for binary variables.

a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Table A.2: Grade attended during the PASEC panel for six imaginary cases

case 1	case 2	case 3	case 4	case 5	case 6	
2	**2**	**2**	**2**	**2**	**2**	school year 1995 - 1996
2	**2,3**	**drop.**	**3**	**3**	**3**	school year 1996 - 1997
3	**3**		**3,4**	**4**	**3**	school year 1997 - 1998
4	**4**		**3,4,5**	**5**	**3**	school year 1998 - 1999
5	**5**		**3,4,5,6**	**6**	**4**	school year 1999 - 2000

(When the child did not take the tests, the possible grades are in grey)

$$LP_{-ik} = min_{\{j \neq i, R_{jk}=0\}}(S_{jk})$$

Previous year's test scores are a proxy for learning achievement prior to the current school year. During the panel, the children took tests at the end of each school year. In each grade-year of the panel, most of the children had been in the preceding grade the year before. The others had been in the same grade the year before, and were currently repeating their grade. The tests for currently repeating children and others had been different. Yet, some items had been common to both, and those items are used to compare the knowledge of the pupils prior to the current school year. Again, this variable has a mean of 0 and a standard deviation of 1 within each grade-year. This comparison relies exclusively on skills acquired in the preceding grade, since the tests never included items about the skills supposed to be acquired in the following grades.

Group mean progression Difference between the group mean test score and the group mean previous year's test score. Given that both test scores are standardized, the average of this variable is 0.

A.3 Other explanatory variables in main regression

Household wealth is a composite indicator for possession of durable goods, obtained by a principal component analysis (see Filmer and Pritchett, 2001). It is based on children's declarations in 1995, and so avoids reverse causality due to the children's education.

Negative shocks on harvests is a dummy taking value 1 if the head of the household reports a negative shock on harvests during the current calendar year or the next. These shocks are taken into account if the child or her parents were still in the household visited by EBMS in 2003. Otherwise this dummy equals 0, because the child was not really affected by these shocks. (140 cases out of 1823)

Parents' education is the mean of both parents' education. The education of an individual is 1 if the individual never went to school, 2 if the person began but did not finish primary school, 3 if she finished primary school but did not begin secondary school, etc. It takes the highest value, 8, if the individual attended to higher education. If information about the father's education or the mother's education was missing, it is replaced by the mean education of the other adults (aged more than 25 in 1995) in the household.

A.4 Other variables

Community has electricity is from the community questionnaire of the EBMS survey

Community main activity is from the community questionnaire of the EBMS survey

Distance to health center is from the community questionnaire of the EBMS survey

Distance to hospital is from the community questionnaire of the EBMS survey

Religion, ethnic group of the head are taken from the EBMS survey

Village or city population is taken from the EBMS survey for rural area, and from the national census for cities.

B Simulations : Formulas

The model of interest is:

$$\begin{cases} E_{ik,t+1} = \mathbb{1}[\ \beta_{e1}S_{ik} + \beta_{e2}LP_{-ik} + \gamma R_{ik} + X_{ik}\beta_{e4} + u_{ik} > 0] \\ \\ R_{ik} = \mathbb{1}[\ \beta_{r1}S_{ik} - \beta_{r2}LP_{-ik} + f_r(S_{ik} - LP_{-ik}) + X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ selection = \mathbb{1}[\ \beta_{s1}S_{ik} + \beta_{s2}LP_{-ik} + \beta_{s3}Z_s + \gamma_s R_{ik} + X_{ik}\beta_{s4} + v_{ik} > 0] \end{cases} \tag{B.1}$$

Grade repetition risk For each observation, it is possible to compute the grade repetition risk: $P_{red} = \Phi\left(\beta_{r1}S_{ik} - \beta_{r2}LP_{-ik} + f_r(S_{ik} - LP_{-ik}) + X_{ik}\beta_{r4}\right)$. This risk can easily be adapted to speculative situations with different LP_{-ik}. $\tilde{P}_{red} = \Phi\left(\beta_{r1}S_{ik} - \beta_{r2}\widetilde{LP}_{ik} + f_r(S_{ik} - \widetilde{LP}_{ik}) + X_{ik}\beta_{r4}\right)$ gives individual grade repetition risks. The simulations presented here give their sample average (and sub-sample averages).

Dropout risk, no "direct effect of grade repetition policy" To simplify the algebra, $E_{ik,t+1}$ and R_{ik} are assumed independent. Hence the probability of $E_{ik,t+1}$ writes
$P_{enr} = P_{red}\Phi\left(\beta_{e1}S_{ik} + \beta_{e2}LP_{-ik} + \gamma + X_{ik}\beta_{e4}\right) + (1 - P_{red})\Phi\left(S_{ik} + \beta_{e2}LP_{-ik} + X_{ik}\beta_{e4}\right)$. The simulations compute the consequences of a speculative change in LP_{-ik} on P_{red}. They change dropout risk through the change in P_{red}, but not through $\beta_{e2}LP_{-ik}$. The new probability of $E_{ik,t+1}$ writes $\tilde{P}_{enr} = \tilde{P}_{red}\Phi\left(\beta_{e1}S_{ik} + \beta_{e2}LP_{-ik} + \gamma + X_{ik}\beta_{e4}\right) + (1 - \tilde{P}_{red})\Phi\left(S_{ik} + \beta_{e2}LP_{-ik} + X_{ik}\beta_{e4}\right)$, where $\tilde{P}_{red}$ is the new P_{red}.

Dropout risk, with "direct effect of grade repetition policy" $E_{ik,t+1}$ and R_{ik} are still assumed independent. The simulations compute the consequences of a speculative change in LP_{-ik} on P_{red}. They change dropout risk through the change in P_{red}, and through $\beta_{e2}LP_{-ik}$. The new probability of $E_{ik,t+1}$ writes $\tilde{P}_{enr} = \tilde{P}_{red}\Phi\left(\beta_{e1}S_{ik} + \beta_{e2}\widetilde{LP}_{ik} + \gamma + X_{ik}\beta_{e4}\right) + (1 - \tilde{P}_{red})\Phi\left(S_{ik} + \beta_{e2}\widetilde{LP}_{ik} + X_{ik}\beta_{e4}\right)$, where $\tilde{P}_{red}$ is the new P_{red}, and $\widetilde{LP}_{ik}$ is the speculative LP_{-ik}.

Precision of the estimates When presented in the paper, they derive from a Delta-method not detailed here.

C Additional tables

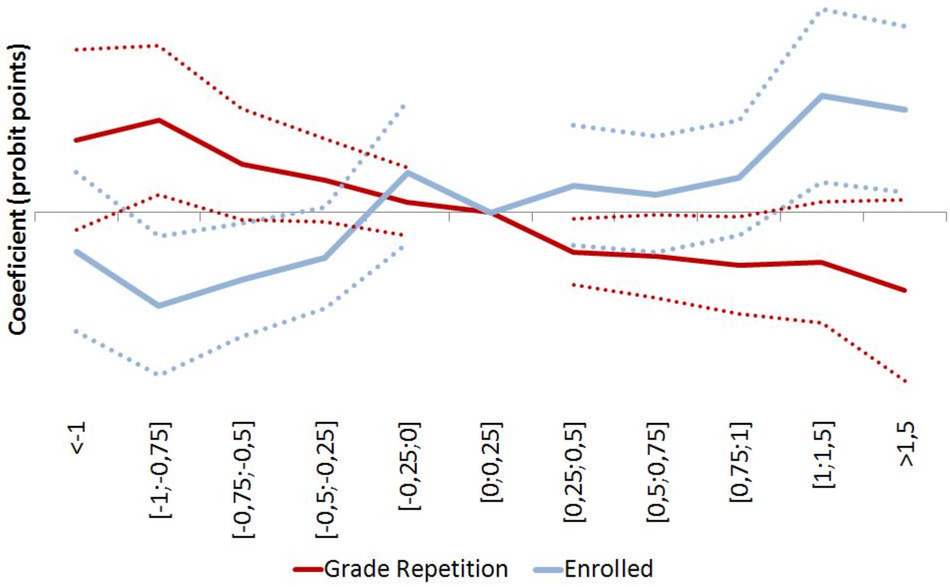
C.1 Determinants of LP_{-ik}

Table C.1: Determinants of LP_{-ik}

	Community characteristics	Household characteristics	All characteristics
	(1)	(2)	(3)
Group[a] mean test score	1.042 (.107)***	1.012 (.090)***	1.049 (.109)***
Group[a] mean progression	.074 (.092)	.083 (.088)	.076 (.098)
Community mean household wealth index	-.138 (.095)		-.150 (.094)
Community mean education index	.183 (.174)		.234 (.184)
ln(population)	.026 (.036)		.025 (.037)
Community main occupation: trade (ref: agriculture)	.145 (.205)		.169 (.195)
Electricity in the community	-.001 (.154)		-.030 (.161)
Rural	.012 (.239)		.017 (.233)
Distance to the next health center	.203 (.217)		.210 (.239)
Distance to the next hospital	-.039 (.055)		-.041 (.055)
household wealth index		.013 (.021)	-.012 (.009)
Parent's education		.006 (.015)	-.018 (.012)
Household head: non-muslim		.073 (.127)	.120 (.123)
Household head: Pulaar, halpulaar (ref: wolof)		.093 (.060)	-.0004 (.052)
Household head: Serere (ref: wolof)		.017 (.094)	-.098 (.097)
Household head: Dioola (ref: wolof)		.083 (.092)	-.143 (.112)
Household head: Mandingue-Sose (ref: wolof)		.009 (.086)	-.115 (.087)
Household head: others (ref: wolof)		.028 (.072)	-.109 (.069)
Grade*year dummies	Yes	Yes	Yes
Obs.	1709	1805	1696
R^2	.55	.519	.554
Joint significance community and/or hh. variables	7.094	.537	21.975
P-value	.527	.823	.144

Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same community.
a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Figure C.1: Non-linear effect of the difference between test score and last passer's score on grade repetition and enrollment



Notes: Plot of the estimates of Table C.2. Dotted lines give the confidence intervals at the 5% level.

C.2 First stage and reduced form estimates

This section presents the equivalent of the first stage and reduced form estimates corresponding to model (5):

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[\quad \beta_{e1} S_{ik} \quad +\beta_{e2a} LP_{-ik} \quad +\beta_{e3} Z_s \quad +f_e(S_{ik} - LP_{-ik}) \quad +X_{ik}\beta_{e4} + u_{ik} > 0] \\ \\ R_{ik} & = \mathbb{1}[\quad \beta_{r1} S_{ik} \quad -\beta_{r2} LP_{-ik} \quad \quad +f_r(S_{ik} - LP_{-ik}) \quad +X_{ik}\beta_{r4} + \epsilon_{ik} < 0] \\ selection & = \mathbb{1}[\quad \beta_{s1} S_{ik} \quad +\beta_{s2} LP_{-ik} \quad +\beta_{s3} Z_s \quad +f_s(S_{ik} - LP_{-ik}) \quad +X_{ik}\beta_{s4} + v_{ik} > 0] \end{cases} \tag{C.1}$$

In this specification, we do not measure the effect of grade repetition on dropout. Instead, we measure the effect of the distance between test scores and teacher's standards on grade repetition, and on dropout. The effect of the distance to teacher's standards on dropout is assumed to be an indirect effect of grade repetition on dropout in our main estimations.

Table C.2 gives the estimation of model (C.1), and Figure C.1 plots the coefficients of the difference between individual test score and last passer's score. The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator. The three columns of Table 3 correspond to the model's three equations. The data are pooled for the various grades and years. Each specification includes grade-year dummies in each equation.

Table C.2: Grade repetition and school dropouts as a function of a difference between own test score and the test score of the last passer ($S_{ik} - LP_{-ik}$)

	repetition	$enrolled_{t+1}$	*selection*
	(1)	(2)	(3)
Test score	-.231	-.200	.070
	(.273)	(.197)	(.148)
LP_{-ik}	.083	.468	.047
	(.242)	(.201)**	(.138)
$S_{ik} - LP_{-ik} < -1$	.763	-.409	-.157
	(.481)	(.424)	(.324)
$-1 < S_{ik} - LP_{-ik} < -0.75$	.966	-.979	-.610
	(.399)**	(.372)***	(.270)**
$-0.75 < S_{ik} - LP_{-ik} < -0.5$	.506	-.700	-.399
	(.297)*	(.302)**	(.236)*
$-0.5 < S_{ik} - LP_{-ik} < -0.25$	.339	-.471	-.404
	(.224)	(.271)*	(.187)**
$-0.25 < S_{ik} - LP_{-ik} < 0$	.113	.421	-.329
	(.182)	(.380)	(.165)**
$0 < S_{ik} - LP_{-ik} < 0.25$	Ref.	Ref.	Ref.
$0.25 < S_{ik} - LP_{-ik} < 0.5$	-.412	.287	.205
	(.177)**	(.323)	(.172)
$0.5 < S_{ik} - LP_{-ik} < 0.75$	-.454	.190	.275
	(.222)**	(.311)	(.200)
$0.75 < S_{ik} - LP_{-ik} < 1$	-.554	.366	.181
	(.258)**	(.309)	(.199)
$1 < S_{ik} - LP_{-ik} < 1.5$	-.520	1.226	.273
	(.323)	(.464)***	(.216)
$1.5 < S_{ik} - LP_{-ik}$	-.812	1.085	.469
	(.484)*	(.445)**	(.318)
Group[a] mean test score	.268	-.043	-.015
	(.130)**	(.205)	(.116)
Negative shock on harvests		.256	.414
		(.195)	(.154)***
Household wealth and Parents' education, Previous year's test score	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes
Obs.	1818	1818	1818
log likelihood	-1262.328	-1262.328	-1262.328
χ^2 exclusion restriction			7.211
corresponding p value			.007

Notes: The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator (25 iterations). ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. Standard errors clustered between different observations of the same child.

$S_{ik} - LP_{-ik}$ stands for "Difference between own test score and last passer's test score".

a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Determinants of grade repetition and of selection Similarly to Table 3, pupils are much less likely to repeat the grade when their test score is higher than the last passer's score in Table C.2. The corresponding coefficients are plotted in Figure C.1, in dark red. Besides, the negative shocks on harvests are used as an exclusion restriction in the repetition equation. Like in Table C.2, this coefficient is positive and significant in the selection equation.

Determinants of enrollment In the enrollment equation, the coefficients for the difference between own test score and last passer's score is given in Table C.2, and plotted in Figure C.1, in light blue. They tend to mirror the coefficients of the repetition equation: pupils with a learning achievement greater than teacher's standards are the least likely to drop out. In Figure C.1, the curve is grossly symmetric to the curve of the grade repetition equation.

C.3 Reduced form with non-linear controls

Table C.3 and Figure 2 gives compares the estimates when several variables based on test-scores are treated non-linearly. In sum, we add in the specification dummies for levels of own test score, of group mean test score, of difference to group mean, and of last passer's score. Hence these variables are treated non-linearly. The results in Table C.3 are similar to Table C.2. Neither of the dummies set is jointly significant. If something, Figure 2 shows the effect of distance to teacher's standards on dropout is a bit greater with the dummies set control.

Table C.3: Modification of Table C.2 with non-linear treatment of the control variables based on test scores

	repetition	$enrolled_{t+1}$	*selection*
	(1)	(2)	(3)
Test score	.093	-.282	.072
	(.495)	(.311)	(.255)
LP_{-ik}	-.098	.613	.110
	(.235)	(.269)**	(.188)
$S_{ik} - LP_{-ik} < -1$	.823	-.804	-.189
	(.492)*	(.547)	(.346)
$-1 < S_{ik} - LP_{-ik} < -0.75$	.938	-1.290	-.595
	(.400)**	(.447)***	(.289)**
$-0.75 < S_{ik} - LP_{-ik} < -0.5$	.495	-.858	-.402
	(.309)	(.355)**	(.247)
$-0.5 < S_{ik} - LP_{-ik} < -0.25$	.391	-.320	
	(.184)	(.386)	(.166)*
$-0.25 < S_{ik} - LP_{-ik} < 0$	.127	.391	-.320
	(.184)	(.386)	(.166)*
$0 < S_{ik} - LP_{-ik} < 0.25$	Ref.	Ref.	Ref.
$0.25 < S_{ik} - LP_{-ik} < 0.5$	-.381	.375	.230
	(.185)**	(.326)	(.175)
$0.5 < S_{ik} - LP_{-ik} < 0.75$	-.431	.403	.303
	(.231)*	(.287)	(.205)
$0.75 < S_{ik} - LP_{-ik} < 1$	-.500	.696	.222
	(.275)*	(.323)**	(.219)
$1 < S_{ik} - LP_{-ik} < 1.5$	-.400	1.541	.307
	(.333)	(.527)***	(.242)
$1.5 < S_{ik} - LP_{-ik}$	-.537	1.619	.457
	(.507)	(.546)***	(.343)
Group[e] mean test score	.419	-.671	-.240
	(.378)	(.478)	(.314)
Negative shock on harvests		.156	.396
		(.196)	(.147)***
χ^2 test score dummies[a]	6.982	7.742	.592
corresponding p value	.137	.101	.964
χ^2 difference to group[e] mean dummies[b]	1.416	5.245	1.412
corresponding p value	.841	.263	.842
χ^2 group[e] mean dummies[c]	5.695	6.128	2.189
corresponding p value	.127	.106	.534
χ^2 last passer's score dummies[d]	7.539	1.984	7.539
corresponding p value	.184	.851	.184
Household wealth and Parents' education, Previous year's test score	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes
Obs.	1818	1818	1818
log likelihood	-1237.971	-1237.971	-1237.971
χ^2 grade year dummies	4.958	29.208	9.239
corresponding p value	.292	7.09e-06	.055
χ^2 exclusion restriction			7.281
corresponding p value			.007

Additional covariates in each equation: test score, group mean test score previous year's test score, household wealth, parents' education, grade-year dummies.
Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.
a: Dummies for: test score<-1, -1<test score<-0.5, 0<test score<0.5, 0.5<test score<1, 1<test score. -0.5<test score<0 omitted
b: Difference to group mean: difference between own test score and the group mean. Dummies for: difference<-1, -1<difference<-0.5, -0.5<difference<0, 0.5<difference. 0<difference<0.5 omitted
c: Dummies for: group mean<-0.5, -0.5<group mean<0, 0.5<group mean. 0<group mean<0.5 omitted
d: Dummies for: last passer's score<-1.5, -1<last passer's score<-0.5, -0.5<last passer's score<0, 0<last passer's score<0.5, 0.5<last passer's score. -1.5<group mean<-1 omitted
e: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

Table C.4: Estimation of the link between last passer's test score and school outcomes in reduced form (probit models)

	$enrolled_{t+1}$	$enrolled_{t+2}$	$enrolled_{t+3}$	$enrolled_{t+4}$	*Still enrolled* (2003)	*Last Grade* > 5	*Last Grade* > 6	*Last Grade* > 7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test score	.275	.162	.346	.432	.377	.492	.759	.781
	(.116)**	(.099)	(.084)***	(.076)***	(.055)***	(.061)***	(.062)***	(.071)***
LP_{-ik}	-.009	.053	.152	.165	.184	.102	.200	.177
	(.084)	(.090)	(.086)*	(.079)**	(.062)***	(.069)	(.063)***	(.070)**
Group[a] mean test score	-.160	-.048	-.281	-.369	-.453	-.322	-.747	-.690
	(.192)	(.196)	(.173)	(.159)**	(.123)***	(.131)**	(.138)***	(.153)***
Negative shock on harvests	.180	.355	.191	.130	.166	-.053	.082	-.006
	(.211)	(.244)	(.220)	(.208)	(.151)	(.144)	(.146)	(.162)
Household wealth and Parents' education, Previous year's test score	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade*year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1818	1449	1449	1449	1789	1777	1777	1777
log-likelihood	-158.914	-224.825	-371.157	-487.188	-1060.973	-833.129	-892.445	-756.374

Additional covariates in each equation: previous year's test score, household wealth, parents' education, grade-year dummies.
Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.
a: A group is composed of all the observations from the same school, the same year and the same grade. This is an approximation of a class, since in some schools, there are several classes per grade.

D Proofs for the semiparametric identification of model (5)

D.1 model (5)

This section proves that model (5) can be semiparametrically identified.

The model (5) is :

$$\left\{\begin{array}{lcllll} r & = & \mathbb{1}(X\beta_r + \gamma_r Z_1 & & +\varepsilon_r > 0) \\ s & = & \mathbb{1}(X\beta_s + \gamma_s Z_2 & +\alpha_s r & +\varepsilon_s > 0) \\ e & = & \mathbb{1}(X\beta_e + \gamma_e Z_2 & +\alpha_e r & +\varepsilon_e > 0) \end{array}\right. \tag{D.1}$$

(For simplicity r is *repetition*, s is *selection*, and e is $enrolled_{t+1}$. For the same reason, the equations have been written in a simple form $X\beta + \gamma Z + \varepsilon$.)

Let us recall r is observed if and only if $s = 1$. $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ is the distribution function of $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$. Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the distribution function. This idea is used to show that all the parameters of model (5) are identified without any parametric assumption on $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$.

Θ is the support of (X, Z_1, Z_2). Let us make the following assumptions:

1. The distribution of $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ is independent of (X, Z_1, Z_2).

2. $\gamma_r \neq 0$ and $\gamma_s \neq 0$

3. $\forall j \in \{r, s, e\}$, $\beta_{j1} = 1$

4. $\exists (X_0, Z_{10}, Z_{20}) \in \Theta$ verifying :

   (a) In the neighborhood of (X_0, Z_{10}, Z_{20}), $(X, Z_1, Z_2) \in \Theta$

   (b) $\begin{pmatrix} \frac{d\mathbb{P}(r=1,s=1)}{dZ_1}(X_0, Z_{10}, Z_{20}) & \frac{d\mathbb{P}(r=1,s=1)}{dZ_2}(X_0, Z_{10}, Z_{20}) \\ \frac{d\mathbb{P}(r=0,s=1)}{dZ_1}(X_0, Z_{10}, Z_{20}) & \frac{d\mathbb{P}(r=0,s=1)}{dZ_2}(X_0, Z_{10}, Z_{20}) \end{pmatrix}$ has full rank

   (c) $\forall (X, Z_1, Z_2)$ in the neighborhood of (X_0, Z_{10}, Z_{20}), $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$

5. $\exists (a = (X_a, Z_{1a}, Z_{2a}), b = (X_b, Z_{1b}, Z_{2b})) \in \Theta^2$

   (a) $\left\{\begin{array}{l} X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} \\ X_a\beta_s + \gamma_s Z_{2a} + \alpha_s = X_b\beta_s + \gamma_s Z_{2b} \\ X_a\beta_e + \gamma_e Z_{2a} + \alpha_e = X_b\beta_e + \gamma_e Z_{2b} \end{array}\right.$

   (b) In the neighborhood of a and b, $(X, Z_1, Z_2) \in \Theta$ and $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$

Assumption 1 is necessary in Manski (1988) and is still necessary here. It ensures that the derivatives of the probability functions with respect to X, Z_1 or Z_2 are not caused by variations of $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$.

Assumption 2 ensures the instruments have a real causal effect on the endogenous variables.

In model (5), only the signs of the latent variables ($X\beta_r + \gamma_r Z_1 + \varepsilon_r$, $X\beta_s + \gamma_s Z_2 + \alpha_s r + \varepsilon_s$ and $X\beta_e + \gamma_e Z_2 + \alpha_e r + \varepsilon_e$) are observed. Accordingly, the parameters are identified up to the scale of the parameter vector. Assumption 3 easily fixes that scale.

Assumption 4a ensures it is possible to compute the derivatives of the probability functions with the data since the points in the neighborhood of (X_0, Z_0) are in the support of (X, Z). It is certainly possible to extend the identification result when X contains some binary variables.

Assumption 4b ensures some of the derivatives of the probability functions are not all zero and that they are not collinear, so that the systems are fully identified in (X_0, Z_{10}, Z_{20}).

Assumption 4c ensures the other derivatives of the probability functions with respect to the covariates are not null in (X_0, Z_{10}, Z_{20}).

Assumption 5 ensures the support Θ is large enough to contain a pair of points with similar characteristics for s and e when the former has $r = 1$ and the latter has $r = 0$.

This proof has three steps: first, it is shown that the coefficients β and γ of the first two equations of model (5) are identified, second, it is shown that the coefficients β and γ of the last equation are identified, and finally, it is shown that the α are identified.

- **Identification of the first two equations of the model**

  Let us compute the derivatives of $\mathbb{P}(r = 1, s = 1|X, Z_1, Z_2)$. This probability and its derivatives can be estimated with the data in (X_0, Z_{10}, Z_{20}) if assumption 4a is true:

$$\begin{aligned} P^{(11)} &= \mathbb{P}(r = 1, s = 1|X, Z_1, Z_2) \\ &= \int_{-X\beta_r-\gamma_r Z_1}^{\infty} \int_{-X\beta_s-\gamma_s Z_2-\alpha_s}^{\infty} \int_{\mathbb{R}} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\ &= F^{(11)}(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2 - \alpha_s) \end{aligned}$$

  We note $F_1'^{(11)}$ and $F_2'^{(11)}$ the derivatives of $F^{(11)}$ with respect to its two arguments. The derivatives are:

$$\frac{dP^{(11)}}{dX_1} = F_1'^{(11)} + F_2'^{(11)} \qquad \text{(D.2)}$$

$$\frac{dP^{(11)}}{dX_i} = \beta_{ri} F_1'^{(11)} + \beta_{si} F_2'^{(11)} \ (\forall i \in \{1..K\}) \qquad \text{(D.3)}$$

$$\frac{dP^{(11)}}{dZ_1} = \gamma_r F_1'^{(11)} \qquad \text{(D.4)}$$

$$\frac{dP^{(11)}}{dZ_2} = \gamma_s F_2'^{(11)} \qquad \text{(D.5)}$$

  This is clearly not sufficient to identify β and γ. In fact, these four equations contain six unknown parameters, since $F_1'^{(11)}$ and $F_2'^{(11)}$ are unknown. So the derivatives of $\mathbb{P}(r = 0, o = 1|X, Z_1, Z_2)$ are necessary to identify γ and β.

$$\begin{aligned} P^{(01)} &= \mathbb{P}(r=0, s=1|X, Z_1, Z_2) \\ &= \int_{-\infty}^{X\beta_r - \gamma_r Z_1} \int_{-X\beta_s-\gamma_s Z_2}^{\infty} \int_{\mathbb{R}} f(\varepsilon_r, \varepsilon_s, \varepsilon_e) d\varepsilon_r d\varepsilon_s d\varepsilon_e \\ &= F^{(01)}(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2) \end{aligned}$$

We note $F_1'^{(01)}$ and $F_2'^{(01)}$ the derivatives of $F^{(01)}$ towards its two arguments.

$$\frac{dP^{(01)}}{dX_1} = F_1'^{(01)} + F_2'^{(01)} \tag{D.6}$$

$$\frac{dP^{(01)}}{dX_i} = \beta_{ri} F_1'^{(01)} + \beta_{si} F_2'^{(01)} \tag{D.7}$$

$$\frac{dP^{(01)}}{dZ_1} = \gamma_r F_1'^{(01)} \tag{D.8}$$

$$\frac{dP^{(01)}}{dZ_2} = \gamma_s F_2'^{(01)} \tag{D.9}$$

From equation (D.2) rearranged with (D.4) and (D.5), and (D.6) rearranged with (D.8) and (D.9), we get the two equations system:

$$\left\{ \begin{array}{lcl} \frac{dP^{(11)}}{dX_1} &=& \frac{1}{\gamma_r}\frac{dP^{(11)}}{dZ_1} + \frac{1}{\gamma_s}\frac{dP^{(11)}}{dZ_2} \\ \frac{dP^{(01)}}{dX_1} &=& \frac{1}{\gamma_r}\frac{dP^{(01)}}{dZ_1} + \frac{1}{\gamma_s}\frac{dP^{(01)}}{dZ_2} \end{array} \right.$$

Under assumptions 4b and 2, this identifies γ_s and γ_r. We can then easily compute $F_1'^{(11)}$, $F_2'^{(11)}$, $F_1'^{(01)}$ and $F_2'^{(01)}$ with (D.4), (D.5), (D.8) and (D.9). The system:

$$\left\{ \begin{array}{lcl} \frac{dP^{(11)}}{dX_i} &=& \beta_{ri} F_1'^{(11)} + \beta_{si} F_2'^{(11)} \\ \frac{dP^{(01)}}{dX_i} &=& \beta_{ri} F_1'^{(01)} + \beta_{si} F_2'^{(01)} \end{array} \right.$$

identifies β_{ri} and β_{si}. In fact, assumption 2 ensures that $\begin{pmatrix} \gamma_r F_1'^{(11)} & \gamma_r F_1'^{(01)} \\ \gamma_s F_2'^{(11)} & \gamma_s F_2'^{(01)} \end{pmatrix}$ has full rank, that $\begin{pmatrix} F_1'^{(11)} & F_1'^{(01)} \\ F_2'^{(11)} & F_2'^{(01)} \end{pmatrix}$ has full rank.

- **Identification of the third equation**

  We compute the derivatives of $\mathbb{P}(e = 1|X, Z_1, Z_2)$:

$$
\begin{aligned}
P^{(1)} &= \mathbb{P}(e=1|X,Z_1,Z_2)\\
&= \int_{-X\beta_r-\gamma_r Z_1}^{\infty}\int_{\mathbb{R}}\int_{-X\beta_e-\gamma_e Z_2-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\
&+ \int_{-\infty}^{X\beta_r-\gamma_r Z_1}\int_{\mathbb{R}}\int_{-X\beta_e-\gamma_e Z_2}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\
&= F^{(1)}(-X\beta_r-\gamma_r Z_1,-X\beta_e-\gamma_e Z_2,-\alpha_e)
\end{aligned}
$$

We call $F_1'^{(1)}$, $F_2'^{(1)}$ and $F_3'^{(1)}$ the derivatives of $F^{(1)}$ with respect to its arguments. We compute the derivatives of $P^{(1)}$:

$$
\begin{aligned}
\frac{dP^{(1)}}{dX_1} &= F_1'^{(1)}+F_2'^{(1)} && \text{(D.10)}\\
\frac{dP^{(1)}}{dX_i} &= \beta_{ri}F_1'^{(1)}+\beta_{si}F_2'^{(1)} && \text{(D.11)}\\
\frac{dP^{(1)}}{dZ_1} &= \gamma_r F_1'^{(1)} && \text{(D.12)}\\
\frac{dP^{(1)}}{dZ_2} &= \gamma_e F_2'^{(1)} && \text{(D.13)}
\end{aligned}
$$

γ_r is known, so that $F_1'^{(1)}$ can be easily computed with (D.12). It is then possible to compute $F_2'^{(1)}$ with (D.10). Under assumption 4c, $F_2'^{(1)}$ is not null in $(X,Z_1,Z_2)\in\Theta$. That is why γ_e is identified by (D.13). Knowledge of β_{ri}, $F_1'^{(1)}$ and $F_2'^{(1)}$ identifies β_{si} in (D.11).

- **Identification of α_s.**

  Adapting Vytlacil and Yildiz (2007), it is easy to show that:

  If $\exists\,((X_a,Z_{1a},Z_{2a}),(X_b,Z_{1b},Z_{2b}),(X_c,Z_{1c},Z_{2c}),(X_d,Z_{1d},Z_{2d}))\in\Theta^4$ so that[15]

$$
\left\{\begin{array}{l}
X_a\beta_r+\gamma_r Z_{1a}=X_b\beta_r+\gamma_r Z_{1b}=\kappa_{r1}\\
X_c\beta_r+\gamma_r Z_{1c}=X_d\beta_r+\gamma_r Z_{1d}=\kappa_{r2}\\
X_a\beta_s+\gamma_s Z_{2c}=X_c\beta_s+\gamma_s Z_{2c}=\kappa_{s1}\\
X_b\beta_s+\gamma_s Z_{2b}=X_d\beta_s+\gamma_s Z_{2d}=\kappa_{s2}
\end{array}\right.
\Leftrightarrow
\left\{\begin{array}{l}
\mathbb{P}(r|a)=\mathbb{P}(r|b)\\
\mathbb{P}(r|c)=\mathbb{P}(r|d)\\
\hat{\mathbb{P}}(s|a)=\hat{\mathbb{P}}(s|c)\\
\hat{\mathbb{P}}(s|b)=\hat{\mathbb{P}}(s|d)
\end{array}\right.
\qquad \text{(D.14)}
$$

  $0<f(\varepsilon_r,\varepsilon_s,\varepsilon_e)<\infty$ in the neighborhood of a and of b and $\kappa_{r1}\neq\kappa_{r2}$.

  Then

$$
\left(\begin{array}{rc}
 & \mathbb{P}(r=1,s=1|a)-\mathbb{P}(r=1,s=1|c)\\
=- & [\mathbb{P}(r=0,s=1|b)-\mathbb{P}(r=0,s=1|d)]
\end{array}\right)\Rightarrow\kappa_{s1}+\alpha_s=\kappa_{s2} \qquad \text{(D.15)}
$$

[15] $\hat{\mathbb{P}}$ means that the probability is net of the effect of r on o.

It is obvious that the converse is true. In fact, if $\kappa_{s1} + \alpha_s = \kappa_{s2}$, then:

$$\begin{aligned}\mathbb{P}(r=1,s=1|a) + \mathbb{P}(r=0,s=1|b) &= \hat{\mathbb{P}}(s=1|b)\\ \mathbb{P}(r=1,s=1|c) + \mathbb{P}(r=0,s=1|d) &= \hat{\mathbb{P}}(s=1|d)\end{aligned}$$

because

$$\begin{aligned}\mathbb{P}(r=1,s=1|a) + \mathbb{P}(r=0,s=1|b) &= \int_{-\infty}^{\kappa_{r1}}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\ &+ \int_{-\kappa_{r1}}^{\infty}\int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\ &= \int_{\mathbb{R}}\int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\ &= \hat{\mathbb{P}}(s=1|b)\end{aligned}$$

(D.14) ensures that $\hat{\mathbb{P}}(s=1|b) = \hat{\mathbb{P}}(s=1|d)$. Finally:

$$\begin{aligned}\mathbb{P}(r=1,s=1|a) + \mathbb{P}(r=0,s=1|b) &= \mathbb{P}(r=1,s=1|c) + \mathbb{P}(r=0,s=1|d)\\ \Leftrightarrow \mathbb{P}(r=1,s=1|a) - \mathbb{P}(r=1,s=1|c) &= -[\mathbb{P}(r=0,s=1|b) - \mathbb{P}(r=0,s=1|d)]\end{aligned}$$

Proof of equation (D.15):

We write the probabilities:

$$\begin{aligned}\mathbb{P}(r=1,s=1|\kappa_r,\kappa_s) &= \int_{-\kappa_r}^{\infty}\int_{-\kappa_s-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\ \mathbb{P}(r=0,s=1|\kappa_r,\kappa_s) &= \int_{-\infty}^{-\kappa_r}\int_{-\kappa_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\end{aligned}$$

Then we can easily compute the differences of (D.15):

$$\begin{aligned}\mathbb{P}(r=1,s=1|a) - \mathbb{P}(r=1,s=1|c) &= \int_{-\kappa_{r1}}^{-\kappa_{r2}}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\ \mathbb{P}(r=0,s=1|b) - \mathbb{P}(r=0,s=1|d) &= \int_{-\kappa_{r2}}^{-\kappa_{r1}}\int_{-\kappa_{s2}}^{\infty}\int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\end{aligned}$$

We can now rewrite the first term of (D.15):

$$
\begin{aligned}
& \mathbb{P}(r=1,s=1|a) - \mathbb{P}(r=1,s=1|c) = -[\mathbb{P}(r=0,s=1|b) - \mathbb{P}(r=0,s=1|d)] \\
\Leftrightarrow\ & \int_{-\kappa_{r1}}^{-\kappa_{r2}} \left(\int_{-\kappa_{s1}-\alpha_s}^{\infty} \int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e) d\varepsilon_s d\varepsilon_e - \int_{-\kappa_{s2}}^{\infty} \int_{\mathbb{R}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e) d\varepsilon_s d\varepsilon_e \right) d\varepsilon_r = 0 \\
\Leftrightarrow\ & \int_{-\kappa_{r1}}^{-\kappa_{r2}} \int_{\mathbb{R}} \left(\int_{-\kappa_{s1}-\alpha_s}^{-\kappa_{s2}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e) d\varepsilon_s \right) d\varepsilon_r d\varepsilon_e = 0
\end{aligned}
$$

$f(\varepsilon_r,\varepsilon_s,\varepsilon_e) > 0$ in the neighborhood of a and b. As a consequence, it is strictly positive in a subset of the integration interval with a strictly positive Lebesgue measure if $\kappa_{s1}+\alpha_s \neq \kappa_{s2}$. So $\kappa_{s1}+\alpha_s = \kappa_{s2}$, QED.

Assumption 5 ensures that some points verifying (D.14) and (D.15) exist in Θ. In fact, points a and b in assumption 5 verify (D.14) and the second term of (D.15). c can be found in the neighborhood of a and d in the neighborhood of b: the hyperplanes $\hat{\mathbb{P}}(s|(X,Z_1,Z_2) = \hat{\mathbb{P}}(s|a)$ and $\hat{\mathbb{P}}(s|(X,Z_1,Z_2) = \hat{\mathbb{P}}(s|b)$ necessarily contain pairs of points that have the same $P(r)$, since $P(r|a) = P(r|b)$.

These points can be recognized because the validity of (D.14) and

$$
\mathbb{P}(r=1,s=1|a) - \mathbb{P}(r=1,s=1|b) = -[\mathbb{P}(r=0,s=1|c) - \mathbb{P}(r=0,s=1|d)]
$$

can be evaluated with the data and previous results.

- **Identification of α_e.**

If $\exists\,((X_a,Z_{1a},Z_{2a}),(X_b,Z_{1b},Z_{2b}),(X_c,Z_{1c},Z_{2c}),(X_d,Z_{1d},Z_{2d})) \in \Theta^4$ so that

$$
\left\{ \begin{array}{lcl}
X_a\beta_r + \gamma_r Z_{1a} & = & X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\
X_c\beta_r + \gamma_r Z_{1c} & = & X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\
X_a\beta_s + \gamma_s Z_{1a} & = & X_c\beta_s + \gamma_s Z_{1c} = \kappa_{s1} \\
X_b\beta_s + \gamma_s Z_{1b} & = & X_d\beta_s + \gamma_s Z_{1d} = \kappa_{s2} \\
X_a\beta_e + \gamma_e Z_{2a} & = & X_c\beta_e + \gamma_e Z_{2c} = \kappa_{e1} \\
X_b\beta_e + \gamma_e Z_{2b} & = & X_d\beta_e + \gamma_e Z_{2d} = \kappa_{e2}
\end{array} \right.
\Leftrightarrow
\left\{ \begin{array}{lcl}
\mathbb{P}(r|a) & = & \mathbb{P}(r|b) \\
\mathbb{P}(r|c) & = & \mathbb{P}(r|d) \\
\hat{\mathbb{P}}(s|a) & = & \hat{\mathbb{P}}(s|c) \\
\hat{\mathbb{P}}(s|b) & = & \hat{\mathbb{P}}(s|d) \\
\hat{\mathbb{P}}(e|a) & = & \hat{\mathbb{P}}(e|c) \\
\hat{\mathbb{P}}(e|b) & = & \hat{\mathbb{P}}(e|d)
\end{array} \right. \tag{D.16}
$$

and $\left\{ \begin{array}{l} \kappa_{r1} \neq \kappa_{r2} \\ \kappa_{s1}+\alpha_s = \kappa_{r2} \end{array} \right.$ and $0 < f(\varepsilon_r,\varepsilon_s,\varepsilon_e) < \infty$ in the neighborhood of a and of b.

Then

$$
\left(\begin{array}{ll}
 & \mathbb{P}(r=1,s=1,e=1|a) - \mathbb{P}(r=1,s=1,e=1|c) \\
= - & [\mathbb{P}(r=0,s=1,e=1|b) - \mathbb{P}(r=0,s=1,e=1|d)]
\end{array} \right) \Rightarrow \kappa_{e1}+\alpha_e = \kappa_{e2} \tag{D.17}
$$

For the same reason as for the identification of α_s, the converse of D.17 is true. In fact, if $\kappa_{e1}+\alpha_e=\kappa_{e2}$, then:

$$\begin{aligned}
\mathbb{P}(r=1,s=1,e=1|a)+\mathbb{P}(r=0,s=1,e=1|b) &= \hat{\mathbb{P}}(s=1,c=1|b)\\
\mathbb{P}(r=1,s=1,e=1|c)+\mathbb{P}(r=0,s=1,e=1|d) &= \hat{\mathbb{P}}(s=1,c=1|d)
\end{aligned}$$

Proof of equation (D.17):

We write the probabilities:

$$\begin{aligned}
\mathbb{P}(r=1,s=1,e=1|a) &= \int_{-\kappa_{r1}}^{\infty}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{-\kappa_{e1}-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\
\mathbb{P}(r=1,s=1,e=1|c) &= \int_{-\kappa_{r2}}^{\infty}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{-\kappa_{e1}-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\
\mathbb{P}(r=0,s=1,e=1|b) &= \int_{-\infty}^{-\kappa_{r1}}\int_{-\kappa_{s2}}^{\infty}\int_{-\kappa_{e2}}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\
\mathbb{P}(r=0,s=1,e=1|d) &= \int_{-\infty}^{-\kappa_{r2}}\int_{-\kappa_{s2}}^{\infty}\int_{-\kappa_{e2}}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e
\end{aligned}$$

Then we can easily compute the differences of (D.17):

$$\begin{aligned}
& \mathbb{P}(r=1,s=1,e=1|a)-\mathbb{P}(r=1,s=1,e=1|c)\\
= & \int_{-\kappa_{r1}}^{-\kappa_{r2}}\int_{-\kappa_{s1}-\alpha_s}^{\infty}\int_{-\kappa_{e1}-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e\\
& \mathbb{P}(r=0,s=1,e=1|b)-\mathbb{P}(r=0,s=1,e=1|d)\\
= & \int_{-\kappa_{r2}}^{-\kappa_{r1}}\int_{-\kappa_{s2}}^{\infty}\int_{-\kappa_{e2}}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e
\end{aligned}$$

We can now rewrite the first term of (D.15):

$$\begin{aligned}
& \mathbb{P}(r=1,s=1|a)-\mathbb{P}(r=1,s=1|c) = -[\mathbb{P}(r=0,s=1|b)-\mathbb{P}(r=0,s=1|d)]\\
\Leftrightarrow & \int_{-\kappa_{r1}}^{-\kappa_{r2}}\int_{-\kappa_{s2}}^{\infty}\int_{-\kappa_{e1}-\alpha_e}^{-\kappa_{e2}} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e = 0
\end{aligned}$$

$f(\varepsilon_r,\varepsilon_s,\varepsilon_e)>0$ in the neighborhood of any point of Θ (assumption 4c). As a consequence, it is strictly positive in a subset of the integration interval with a strictly positive Lebesgue measure if $\kappa_{e1}+\alpha_e \neq \kappa_{e2}$. That is why $\kappa_{e1}+\alpha_s=\kappa_{e2}$. Assumption 5 ensures that those points exist, so α_e can be identified.

D.2 Model (5) without Z_2

This appendix proves that Z_2 is unnecessary for identifying the sign of α_e. Accordingly, it is theoretically not necessary to control for selection to identify the sign of α_e semiparametrically. The corresponding model is:

$$\begin{cases} r &= \mathbb{1}(X\beta_r + \gamma_r Z & & +\varepsilon_r > 0) \\ s &= \mathbb{1}(X\beta_s & +\alpha_s r & +\varepsilon_s > 0) \\ e &= \mathbb{1}(X\beta_e & +\alpha_e r & +\varepsilon_e > 0) \end{cases} \tag{D.18}$$

(For simplicity r is *repetition*, s is *selection*, and e is *enrolled*$_{t+1}$. For the same reason, the equations have been written in a simple form $X\beta+\gamma Z+\varepsilon$)

Let us recall that r is observed if and only if $s=1$. $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$ is the distribution function of $(\varepsilon_r,\varepsilon_s,\varepsilon_e)$. Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the probability function of the dependent variable. This idea is used to show that the sign of α_e is identified in model (D.18) without any parametric assumption on $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$. Θ is the support of (X,Z). We make the following assumptions:

1. The distribution of $(\varepsilon_r,\varepsilon_s,\varepsilon_e)$ is independent of (X,Z).
2. $\gamma_r \neq 0$
3. $\exists (X_0,Z_0) \in \Theta$ verifying :
   (a) In the neighborhood of (X_0,Z_0), $(X,Z)\in\Theta$
   (b) $\int_{\mathbb{R}}\int_{\mathbb{R}} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e < \infty$
   (c) $f(\varepsilon_r,\varepsilon_s,\varepsilon_e) > 0$ in the neighborhood of $(-X_0\beta_r-\gamma_r Z_0, -X_0\beta_s-\alpha_s, -X_0\beta_s-\alpha_e)$, called Γ

Assumption 1 is necessary in Manski (1988) and is still necessary in this case. It ensures that the derivatives of the probability functions with respect to X or Z are not caused by variations of $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$.

Assumption 2 ensures that the instrument has a causal effect on r.

Assumption 3a ensures that it is possible to compute the derivatives of the probability functions with the data since the points in the neighborhood of (X_0,Z_0) are in the support of (X,Z). It is certainly possible to extend the identification result in the case where X contains some binary variables.

Assumption 3b ensures that the density of ε_r in $-X_0\beta_r-\gamma_r Z_0$ is finite, so that the derivatives of the probabilities with respect to Z are finite.

Assumption 3c ensures that the derivatives of the probability functions with respect to Z are not null.

– **Proof that the sign of γ_r is identified**

We write $\mathbb{P}(r=1, s=1, e=1|X,Z)$, which is identified by the data in (X_0, Z_0) because of assumption 3a:

$$\begin{aligned} \mathbb{P}(r=1,s=1,e=1|X,Z) &= \int_{-X\beta_r-\gamma_r Z}^{\infty}\int_{-X\beta_s-\alpha_s}^{\infty}\int_{-X\beta_e-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\ \Rightarrow d\mathbb{P}(r=1,s=1,s=1|X,Z)/dZ &= \gamma_r \int_{-X\beta_s-\alpha_s}^{\infty}\int_{-X\beta_e-\alpha_e}^{\infty} f(-X\beta_r-\gamma_r Z,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e \\ 0 &\leq \int_{-X\beta_s-\alpha_s}^{\infty}\int_{-X\beta_e-\alpha_e}^{\infty} f(-X\beta_r-\gamma_r Z,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e \end{aligned}$$

Assumption 3b ensures that:

$$\int_{-X_0\beta_s-\alpha_s}^{\infty}\int_{-X_0\beta_e-\alpha_e}^{\infty} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e \leq \int_{\mathbb{R}}\int_{\mathbb{R}} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e < \infty$$

And assumption 3c ensures that:

$$\begin{aligned} &\int_{[-X_0\beta_s-\alpha_s,\infty]\times[-X_0\beta_e-\alpha_e,\infty]} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e \\ \geq &\int_{([-X_0\beta_s-\alpha_s,\infty]\times[-X_0\beta_e-\alpha_e,\infty])\cap\Gamma} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e > 0 \end{aligned}$$

That is why

$$0 < \int_{-X_0\beta_s-\alpha_s}^{\infty}\int_{-X_0\beta_e-\alpha_e}^{\infty} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e < \infty$$

so that $\frac{d\mathbb{P}(r=1,s=1,e=1|X,Z)}{dZ}(X_0, Z_0)$ has the same sign as γ_r.

– **Proof that the sign of α_e is identified**

Now, let us focus on $\mathbb{P}(e=1|X,Z)$:

$$\begin{aligned}
\mathbb{P}(e=1|X,Z) &= \mathbb{P}(e=1,r=1|X,Z)+\mathbb{P}(e=1,r=0|X,Z) \\
&= \int_{-X\beta_r-\gamma_r Z}^{\infty}\int_{\mathbb{R}}\int_{-X\beta_e-\alpha_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&+ \int_{-\infty}^{-X\beta_r-\gamma_r Z}\int_{\mathbb{R}}\int_{-X\beta_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{-X\beta_e}^{\infty} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
&+ \int_{-X\beta_r-\gamma_r Z}^{\infty}\int_{\mathbb{R}}\int_{-X\beta_e-\alpha_e}^{-X\beta_e} f(\varepsilon_r,\varepsilon_s,\varepsilon_e)d\varepsilon_r d\varepsilon_s d\varepsilon_e \\
\Rightarrow d\mathbb{P}(e=1|X,Z)/dZ &= \gamma_r\int_{\mathbb{R}}\int_{-X\beta_e-\alpha_e}^{-X\beta_e} f(-X\beta_r-\gamma_r Z,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e
\end{aligned}$$

Again, if $\alpha_e > 0$, then $0 < \int_{\mathbb{R}}\int_{-X_0\beta_e-\alpha_e}^{-X_0\beta_e} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e < \infty$, because of hypotheses 3b and 3c. For the same reasons, if $\alpha_e < 0$, then $-\infty < \int_{\mathbb{R}}\int_{-X_0\beta_e-\alpha_e}^{-X_0\beta_e} f(-X_0\beta_r-\gamma_r Z_0,\varepsilon_s,\varepsilon_e)d\varepsilon_s d\varepsilon_e < 0$. This shows that $d\mathbb{P}(e=1|X,Z)/dZ$ and $\alpha_e\gamma_r$ have the same sign. The sign of γ_r is identified, so the sign of α_e is identified.