

# The effect of grade repetition on school dropout

## An identification based on the differences between teachers\*

Pierre André<sup>†</sup>

January 22, 2008

JEL Classification: D12, I28, O12

Key words: Grade repetition, School demand, School dropouts, Senegal

### Abstract

This paper studies the link between grade repetition and school dropouts. We use an original dataset that matches household data with a panel of test scores. With these datasets, we infer the school trajectory of each child. We correct for the potential endogeneity of grade repetition, using the differences between teachers' attitude towards repetition as an instrument for grade repetition. Our results show a negative effect of the grade repetition decision on the probability of being enrolled at school the next year.

---

\*I am grateful to Sylvie Lambert, Marc Gurgand, Luc Behagel and Christelle Dumas for their very useful comments and help on earlier versions of this paper.

<sup>†</sup>Paris School of economics and Lea-Inra 48, bd Jourdan, 75014 Paris, France. Tel: (33) 1.43.13.63.64. Fax: (33) 1.43.13.63.62. E-mail: pierre.andre01@polytechnique.org

## 1 Introduction

Primary education in Senegal is characterized by particularly high repetition rates and frequent dropout before completion of grade 6 (the last grade of primary school).

12% of the pupils being enrolled in Senegalese primary schools in 2004 were repeating their grade in 2005<sup>1</sup>. This widespread practice is very expensive, since all the costs of schooling increase with schooling duration, and each repeating child needs to be enrolled one more year to achieve a given grade. That is why the consequences of grade repetition should be seriously evaluated. Besides, among the Senegalese children enrolled in the first grade of primary school, 40% drop out before being enrolled in the 6<sup>th</sup> grade<sup>1</sup>.

This paper addresses the issue of whether frequent school dropout is partly a consequence of high repetition rates.

Grade repetition impacts schooling decisions through a variety of mechanisms.

First, grade repetition modifies the acquisition of knowledge at a given date, and the parent's beliefs about the future acquisition of knowledge may be a determinant of dropouts. In fact, when a child repeats a grade, he may consolidate the skills corresponding to his grade. However, it is unclear whether it compensates for the fact that he does not acquire the skills corresponding to the next grade. That is why the net effect of grade repetition on the acquisition of knowledge is ambiguous. The empirical evaluation of the net effect of grade repetition on learning achievement has suffered serious difficulties. Most studies try to control for test scores as a proxy for school ability and initial learning achievement (see Holmes (1989) for a meta analysis of many of those studies). Yet, the teachers probably use their information on the child's school ability in order to decide whether he will repeat; and it is likely that teacher observe the child's school ability better than any econometrician. Consequently, these studies may suffer from an endogeneity bias. Jacob and Lefgren (2004) control for this potential bias, using a discontinuity in school policy in Chicago. In fact, the pupils passed standardized tests at the end of grades 3, 6 and 8. They were promoted if their test score was higher than a minimum score. Using a regression discontinuity design, they find a small and positive effect of grade repetition on the acquisition of knowledge at a given date.

Second, grade repetition acts for the parents as a signal for ability. If the parents observe their child's ability noisily, then grade repetition decreases their beliefs on ability. As a result, grade repetition possibly causes school dropouts.

Finally, grade repetition may increase the cost of schooling. It increases the time needed to reach a given grade. For a given last grade attended, the opportunity costs increase by one year when a child has to repeat once, and the job market benefits of schooling are postponed by one year. Then, grade repetition would cause school dropouts because it increases the costs of schooling.

Overall, the sign of the effect of these three mechanisms is ambiguous.

Very few studies tried to estimate this effect in developing countries. King, Orazem, and Paterno (1999) find that grade repetition cause school dropouts in Pakistan. Yet, their identification strategy does not include any control either for the acquisition of knowledge or for parental preferences for schooling. Yet, both are certainly correlated and low parental preferences for schooling cause grade repetition. Consequently, the effect of grade repetition on school dropouts certainly suffers from an endogeneity bias. PASEC (2004) uses a unique panel of test scores in Senegal and finds that grade repetition in the early years of the panel is correlated with attrition at the end of the panel. They control for many covariates and use test scores as a proxy for the acquisition of knowledge and for ability. However, it is not certain that the remaining unobservable variables that cause grade repetition are uncorrelated with future school dropouts. Furthermore, attrition in the last years of the panel may be a poor proxy for school dropouts.

---

<sup>1</sup>Ministry of Education, Senegal (2005)

In this paper, using the same data than PASEC (2004) with new information about school dropouts, we intend to evaluate the effect of grade repetition on school dropouts. In order to control for the potential correlation between the unobservable characteristics of the children and grade repetition, we use an original instrumental variables strategy.

Our instruments are based on the teacher specific attitude towards repetition: we use the fact that grade repetition is based on the teacher's decision, and that this decision is partly based on his idiosyncratic attitude. This attitude is unobservable, and for each child, we use his peer's repetition to proxy for the specific attitude of his teacher.

Our results show a negative and significant effect of grade repetition on the probability of being enrolled at school the next year. The estimated effect is fairly high: the estimated average marginal effect of grade repetition on school dropout is approximately  $-5\%$ , whereas the average dropout rate in the sample is  $2\%$ .

In section 2 of this paper, we present the dataset used to identify the causal effect of grade repetition on school dropouts. In section 3, we present our strategy to identify this effect, and give the benchmark results in section 4. In section 5, we give some specification checks. Finally, we conclude with some brief remarks.

## 2 The data

In order to estimate the effect of grade repetition, we combine two datasets containing detailed information about schooling.

### 2.1 The PASEC panel

The PASEC (CONFEMEN<sup>2</sup>'s analysis program of educational systems) conducted a panel survey in 98 Senegalese primary schools between 1995 and 2001. In a second grade class randomly chosen in each school, 20 second grade students were randomly chosen at the beginning of the school year 1995 - 1996. They were followed during their school trajectory (including grade repetitions) until the first of them finished their 6th grade - which is the last one in primary school - in 2000. As a result, the children are randomly selected among the second grade pupils of the schools in 1995. However, the children in the same grade-year (for example the children being in third grade in 1998 - 1999) are more and more selected as the time passes : due to attrition and due to grade repetitions.

The reason for attrition in this panel is twofold. First, there are school dropouts in Senegal, and when a child drops out, he does not take the PASEC tests. Second, the PASEC team organizes the tests and collects the data in each of the schools on a given day in each school year. If a child missed school this day, or moved before this day, he did not take the test.

### 2.2 EBMS Survey

This survey gives complementary information about some of the PASEC pupils in 2003. 59 of the schools surveyed between 1995 and 2000 are in the EBMS sample. In each of the communities (i.e. village, or quarter in urban areas), the objective was to resurvey households with children belonging to the PASEC panel. Information on the living conditions (wealth, health) and on schooling of the people in the household was collected. In particular, retrospective data about the school career of the children surveyed by PASEC allows us to distinguish dropouts from other causes of attrition. Consequently, we know for nearly each children re-surveyed when he left school (if he had left in 2003). Out of the 1177 pupils of the 59 schools surveyed by PASEC, 921 children are in EBMS data after deletion of questionable matches.

---

<sup>2</sup>Conference of education ministers of French speaking countries

					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 atten- dance
Initial tests (1995)	school year 1995 - 1996	school year 1996 - 1997	school year 1997 - 1998	school year 1998 - 1999	school year 1999 - 2000	

Table 1: Test attendance in the panel

## 2.3 Global dataset

Both datasets give reliable retrospective information about enrollment. Furthermore, together they give enough information to reconstruct grade repetitions in most cases. However, grade repetitions are inferred from the school trajectory, so that in some cases the grade repetition is indetermined<sup>3</sup>.

This information is necessary for any evaluation of the impact of repetition on schooling. The global dataset has another unique advantage: it evaluates individual learning achievement (with the test scores), which is a crucial determinant of grade repetition. Nevertheless, we report in table 1 the number of children attending each test in our sample and observe that it is frequent that a child did not take a test even if he was still enrolled at that date. In school year 1995 - 1996, all the 921 children were enrolled.

We give in appendix A the definition of all the variables used in this paper.

## 3 The instrumental strategy

In this paper, we intend to identify the effect of grade repetition on school dropouts. The main difficulty in identifying the causal effect of grade repetition is to control for its potential endogeneity. In fact, in Senegal, the teacher decides whether each pupil passes to the next grade or has to repeat. As long as he observes the school ability of his pupils, he probably uses this information. Then, for a given learning achievement at the end of the current school year, as measured by the test score, a child with lower school ability is more likely to repeat his grade the next school year.

### 3.1 Model for the determinants of grade repetition and enrollment

In the equation (1), we modelize the causes of grade repetition. Three factors may affect the repetition of the child  $i$  of group  $k$ : first, his learning achievement  $S_{ik}$ , second, a vector of covariates  $X_{ik}$ <sup>4</sup>, and third, unobservables  $\epsilon_{ik}$ . 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. The learning achievement is compared to  $t_k$ , which is the learning achievement required to pass in group  $k$ . The difference between learning achievement and  $t_k$  affects grade repetition, and there is a discontinuity of  $\mathbb{P}(R_{ik} = 1)$  when  $S_{ik} = t_k$ .

<sup>3</sup>The details are explained in appendix A

<sup>4</sup>This vector includes grade-year dummies, wealth of the household and education of the parents

$$R_{ik} = \mathbb{1} [S_{ik} - t_k + \delta \mathbb{1}(S_{ik} - t_k > 0) + X_{ik}\beta_r + \epsilon_{ik} < 0] \quad (1)$$

The unobservables  $\epsilon_{ik}$  may be correlated with unobservable characteristics of the child causing dropouts. First, the teacher's beliefs on learning achievement may be correlated with the parents' beliefs conditionnaly to test scores. If in addition the parents' beliefs on learning achievement affect the schooling decision, then grade repetition is probably endogenous. Second, the parental preferences for schooling may have an effect on the acquisition of knowledge. Then, when parents have strong preferences for schooling, the teacher may consider that the children are more likely to improve learning achievement during the following years, and do not need to repeat. Again, this would generate endogeneity of grade repetition.

In this paper, we intend to use proxies for the teacher's specific attitude towards repetition as instruments for grade repetition. Teacher's specific attitude towards repetition relies on the teacher and not on the pupil himself, that is why it controls for the main potential source of endogeneity: the correlation between the child's unobservables and the grade repetition decision. In our model, the teacher's attitude towards repetition does affect grade repetition through  $t_k$ . Equation (2) shows that  $t_k$  is possibly caused by the average learning achievement in the group ( $\overline{S}_k$ ), and by teacher's attitude towards repetition ( $\nu_k$ ).

$$t_k = \lambda \overline{S}_k + \nu_k \quad (2)$$

Of course,  $\nu_k$  is not observable. For that reason, we will proxy for it. Rearranging (1) and (2), we can observe that as long as  $\lambda = 1$ , the individual probability of grade repetition depends on the difference between a child's learning achivement and his peers' average learning achievement, and on the teacher's attitude towards repetition. In that case, the grade repetition decision is not directly based on learning achivement: it is based on comparisons in learning achievement within each class. We will observe farther that  $\lambda$  is probably close to 1.

$$R_{ik} = \mathbb{1} [S_{ik} - \lambda \overline{S}_k - \nu_k + \delta \mathbb{1}(S_{ik} - \lambda \overline{S}_k - \nu_k > 0) + X_{ik}\beta_r + \epsilon_{ik} < 0] \quad (3)$$

In the case  $\lambda = 1$ , the repetition rate in the group depends on the distribution of the learning achievement in the group and on teacher's attitude towards repetition, and does not depend on average learning achievement in the group. For that reason, we will use the group average repetition  $\widetilde{R}_{ik}$  rate as a proxy for teacher's attitude towards repetition.

$$\widetilde{R}_{ik} = \Sigma_{j \neq i} \mathbb{1} [S_{jk} - t_k + \delta \mathbb{1}(S_{jk} - t_k > 0) + X_{jk}\beta_r + \epsilon_{jk} < 0] \quad (4)$$

In that case, the repetition of a child is given by equation (5). We do not include the term  $\delta \mathbb{1}(S_{ik} - t_k > 0)$  in this equation. In fact,  $\lambda \overline{S}_k - \alpha \widetilde{R}_{ik}$  is a poor proxy for  $t_k$ . For that reason, it is empirically impossible to evaluate  $\delta$ .

$$R_{ik} = \mathbb{1} [S_{ik} - \lambda \overline{S}_k + \alpha \widetilde{R}_{ik} + X_{ik}\beta_r + \epsilon_{ik} < 0] \quad (5)$$

In this paper, we use an additional proxy for teacher's attitude towards repetition. Among the peers of a given child a given year, some are admitted to the next grade and we call them the "passers". Among the passers, the child with the lower test score is called the last passer. We use his test score as a proxy for  $t_k$ , and note it  $LP_{ik}$ :

$$LP_{ik} = \min_{\{j \neq i | R_{jk}=0\}} (S_{jk}) \quad (6)$$

We still control for  $\overline{S}_k$ , because  $LP_{ik}$  is a proxy for  $t_k$ , and not  $t_k$ :

$$R_{ik} = \mathbb{1} [S_{ik} - \lambda_1 \overline{S}_k - \lambda_2 LP_{ik} + \delta \mathbb{1}(S_{ik} - LP_{ik} > 0) + X_{ik} \beta_r + \epsilon_{ik} < 0] \quad (7)$$

In this paper, we intend to infer whether grade repetition affects school dropout. For that reason, we modelize enrollment after the grade repetition ( $E_{ik,t+1}$ ) in a reduced form:

$$E_{ik,t+1} = \mathbb{1} [\beta_{e1} S_{ik} + \beta_{e2} \overline{S}_k + X_{ik} \beta_{e3} + \gamma R_{ik} + u_{ik} > 0] \quad (8)$$

In the benchmark specification of this paper, we will estimate (8) jointly with (7). We use the instruments  $LP_{ik}$  and  $\mathbb{1}(S_{ik} - LP_{ik} > 0)$  to control for the potential endogeneity of grade repetition.

### 3.2 Estimation of the model of grade repetition

In the previous section, we have set up the theoretical framework on which our identification rely, and then which proxies for teacher's attitude towards repetition are used in this paper. Before discussing the exogeneity of our proxies, we estimate our model of the determinants of grade repetition.

Table 2 estimates various specifications of a probit model estimating the determinants of grade repetition. The data are pooled between grades and years. The standard errors of the estimators are corrected for the correlation of the residuals between different observations of the same child. Each specification includes grade-year dummies, the  $\chi^2$  statistics for their joint significance is reported.

Grade repetition is inferred from the school trajectory of each child. For that reason, even conditionnaly to enrollment at the end of the school year (and attendance the day of the test), grade repetition potentially suffers from a selection bias. In fact, we need to reobserve a child to infer his grade repetition. However, we will observe farther that correcting from this bias barely affects the coefficients.

The column 1 of table 2 regress the grade repetition decision on the observables of the equation (3) and on household characteristics. We observe that in this table, the test scores and group mean test scores affect strongly the grade repetition probability. The sign of the coefficients are opposed, which is predicted by equation (3). The absolute value of both coefficients is nearly the same, and the difference between the absolute values is not significantly different from 0. In our model,  $\lambda$  is the ratio between the two absolute values. In that case, we cannot reject  $\lambda = 1$ .

Conditionnaly to the present learning achievement of the child, the education and the wealth of his household do not seem to be correlated with his probability of repetition.

In the columns 2 and 3, we include the proxies for grade repetition of equations (5) and (7). In the column 2, the coefficient of the grade repetition rate among the peers is positive and undoubtedly different from 0 (at the  $10^{-15}\%$  level). It increases the log-likelihood of the model by nearly 50.

In column 3, the last passer's test score is a proxy for  $t_k$ . In our model, once controlled for  $t_k$ ,  $\overline{S}_k$  is not a determinant of grade repetition. For that reason, in column 3, the coefficient for  $\overline{S}_k$  decreases and is half the absolute value of the coefficient for  $S_{ik}$ , but it is still significant.

Specification 3 includes a dummy taking value 1 if the child's test score is higher than the last passer's score ( $\mathbb{1}(S_{ik} > LP_{ik})$ ). The coefficient for this dummy is negative and significant. We expected that a non-linearity in the latent variable could cause this effect, and observe that this effect is significant.

Among the peers of a child, we call repeaters the ones who are not admitted to the next grade. Among the repeaters, the one with the highest test score is the "first repeater". Our specifications do not include the test score of the first repeater as a proxy for  $t_k$ : we do not observe repeaters in each group. However, we include a dummy taking value 1 if the test score is higher than the first repeater's score. When we do not observe any repeater in the group, this dummy takes value 1. The coefficient for this dummy is negative and significant.

In this section, we have shown that our model is (at least partly) consistant with the datas. In particular, we tested these predictions of our model:

	equation (3)	equation (5)	equation (7)
	(1)	(2)	(3)
Test score	-.898 (.068)***	-.985 (.077)***	-.670 (.097)***
Group mean test score	.912 (.096)***	1.045 (.105)***	.385 (.109)***
Household's wealth	-.034 (.022)	-.026 (.022)	-.043 (.022)*
Parental mean education	-.025 (.030)	-.020 (.031)	-.014 (.032)
Repetition rate in the group		1.941 (.229)***	
Last passer's test score			.297 (.080)***
The test score is higher than the last passer's score			-.309 (.131)**
The test score is higher than the first repeater's score			-.446 (.093)***
Obs.	1785	1768	1768
log-likelihood	-678.618	-629.282	-625.584
$\chi^2$ grade year dummies	36.875	10.815	13.476
corresponding p value	$< 10^{-5}$	.029	.009
$\chi^2$ teacher's attitude		71.896	82.822
corresponding p value		$< 10^{-15}$	$< 10^{-15}$

Additional covariates in each specification: grade-year dummies

Note: \*\*\*, \*\* and \* mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard errors of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Table 2: Estimation of the determinants of grade repetition

- The grade repetition probability depends on the learning achievement. The learning achievement required to pass strongly depends on the peer’s learning achievement.
- So as to predict a given child’s repetitions, his peers’ repetitions give relevant information.

Now, we will try to convince you that the second prediction of our model is uniquely caused by teacher’s attitude towards repetition, and that teacher’s attitude towards repetition is exogenous. If both assumptions are true, then our proxies are exogenous. The next sections answer those 2 questions.

### 3.3 Do $LP_{ik}$ and $\widetilde{R}_{ik}$ really proxy for teacher’s attitude towards repetition ?

In this section, we assess whether the link between a child’s repetition and our proxies is solely due to teacher’s attitude towards repetition. Our proxies are based on the repetitions and the test scores of the peers. For that reason, if the repetitions or the test scores of the peers are linked with the child’s unobservables, then our proxies may be endogenous.

For each child  $i$  of the group  $k$ , we write  $j$  the name of one of his peers. Both proxies can be written in the form  $\Phi(\{R_{jk}, S_{jk}\}_{j \neq i})$ , which means that they are a function of the peer’s test scores and of their repetitions. We want to rule out the fact that  $\Phi(\{R_{jk}, S_{jk}\}_{j \neq i})$  and  $\epsilon_{ik}$  could be correlated once controlled for the observables. We control for the child  $i$ ’s test score and for his group average test score, and assume that it controls for the correlation between a child  $i$ ’s unobservables and his peer’s (child  $j$ ’s) test score  $S_{jk}$ . For that reason, we will focus on the potential correlation between a child  $i$ ’s unobservables and the repetition of his peers  $R_{jk}$ . We rewrite the determinants of the repetition of the peer  $j$  of equation (3):

$$R_{jk} = \mathbb{1} [f(S_{jk} - \lambda \overline{S}_k - \nu_k) + X_{jk}\beta + \epsilon_{jk} < 0] \quad (9)$$

In equation (9), we still assume that we control for the correlation between child  $i$ ’s unobservables and the test score  $S_{jk}$ . However, the unobservables  $\epsilon_{jk}$  could be correlated with the unobservables of the child  $i$ :  $\epsilon_{jk}$  and  $\epsilon_{ik}$  are correlated if different observations of  $\epsilon$  in the same group are correlated, conditionnaly to observable variables. In our model, the correlation between the unobservables of the children from the same group could cause an endogenous measurement error on the teacher’s attitude towards repetition.

This correlation is theoretically plausible: for example, we could expect that (the lack of) motivation at school cause grade repetition. Besides, the lack of motivation causes school dropouts. If the distribution of motivation is not the same between different schools and motivation causes grade repetition, then  $\epsilon_{jk}$  is probably correlated with  $\epsilon_{ik}$ , and in that case  $\epsilon_{jk}$  is correlated to the dropouts of the child  $i$ . For that reason,  $\widetilde{R}_{ik}$  would be correlated with  $u_{ik}$ , and then it is endogenous. We have two empirical arguments to reject this spurious correlation between the unobservables of the child  $i$  and the unobservables of his peer  $j$ .

The first empirical argument is that we expect that any unobservable that has different distribution between groups is correlated with the observables of the school. For example, we expect that in wealthy communities, or in communities where the average education is high, the motivation at school is higher. If the lack of motivation causes grade repetition and motivation is higher in wealthy communities, then the repetition rate is lower in wealthy communities.

In table 3, we regress the repetition rate and the test score of the last passer on community-level characteristics. We do not observe that our proxies for teacher’s attitude toward repetition are correlated with any of our community-level variables<sup>5</sup>.

---

<sup>5</sup>An additionnal regression indicates that the grade repetition rate is not significantly correlated with household’s wealth ou household’s education, once controlled for school fixed effects, grade-year fixed effects and test scores. This rules out an endogenous placement of pupils correlated with teacher’s attitude towards repetition within schools.

	last passer's score			Repetition rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean test score	.983 (.093)***	.977 (.111)***	.975 (.101)***	-.049 (.036)	-.072 (.038)*	-.051 (.036)
Standard deviation of test scores	-.555 (.204)***	-.519 (.226)**	-.579 (.215)***	-.043 (.063)	-.038 (.073)	-.061 (.068)
Mean previous year test score	-.066 (.069)	-.082 (.067)	-.094 (.070)	-.052 (.028)*	-.037 (.029)	-.053 (.028)*
Community mean wealth		-.120 (.091)			-.052 (.029)*	
Community mean education		.005 (.121)			.053 (.040)	
log (city or village population)		.027 (.035)			.009 (.012)	
Electricity in community		.181 (.198)	-.030 (.109)		.074 (.063)	.005 (.041)
Rural community		.120 (.207)			.041 (.066)	
Distance to health centre		-.112 (.162)	-.155 (.137)		-.048 (.047)	-.053 (.043)
Distance to hospital		-.065 (.041)	-.028 (.037)		-.0006 (.014)	.009 (.013)
Agricultural community						
Commercial community		.223 (.117)*	.062 (.108)		.009 (.047)	-.023 (.035)
Obs.	280	267	280	298	283	298
$R^2$	.506	.518	.51	.358	.372	.366
F-test grade-year dum	7.366	8.638	6.621	5.170	6.206	5.290
corresponding p-value	$< 10^{-5}$	$< 10^{-5}$	.00002	.0006	.0001	.0005
F-test community variables		.893	.780		.553	.659
corresponding p-value		.529	.543		.811	.623

OLS corrected for clustering by school. Additional covariates in each equation: grade-year dummies

Note: \*\*\*, \*\* and \* mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard errors of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Table 3: Determinants of teacher's specific attitude towards repetition

The column 1, 2 and 3 run OLS regressions of the last passer's score of each group on some characteristics of the groups and schools. Obviously, the last passer's score is correlated with group mean test score. The coefficient is approximately 1, which is compatible with  $\lambda = 1$ <sup>6</sup>.

A high standard deviation in the group is associated with a lower last passer's score. In fact, suppose that in each group, the 15% of children with the lowest test scores repeat. In that case, for a given group mean test score, the higher the standard deviation of the test scores in the group is, the lower the last passer's test score is expected to be.

The previous year mean test score is not significantly linked with the last passer's score.

In column 2, the specification includes a large set of community variables. We can observe that none of them is significantly different from 0 at the 5% level. The F-test for their joint significance does not indicate that any of them is significantly correlated with the last passer's score. Yet, the community variables in the specification are correlated, which generates multicollinearity. We keep some of these variables in column 3, chosen to decrease multicollinearity. Again, nothing indicates that the last passer's score is correlated with the remaining community variables.

In the columns 4, 5 and 6 of table 3, we run OLS regressions of group repetition rate on the same covariates. The repetition rate is negatively correlated with group mean test score and group mean previous year test score. Even if we observe in table 2 that  $\lambda$  is probably close to 1, the repetition rate is lower when the learning achievement in the group is higher. This indicates that  $\lambda$  is probably slightly smaller than 1.

The standard deviation of the test scores is not correlated with the group repetition rate.

In columns 5 and 6, we do not find any evidence that the group repetition rate is correlated with the community characteristics.

However, we could be worried that the F-tests in table 3 are not powerful enough to detect a correlation between the observables of the community and repetition rates. For example, we could be worried that the estimated differences of 2.2% in repetition rate between agricultural and commercial communities in column 6 of table 3 is not significantly different from 0.

A second empirical argument rejects the spurious correlation between the unobservables of the child  $i$  and the unobservables of his peers. In fact, we expect that the correlation between the  $\epsilon$  between different peers is caused by endogenous placement in schools. Chamberlain (1980) explains how it is possible to control for fixed effects in a probit regression. We adapt this methodology to our case in section 5.1; and the results are similar to the benchmark.

### 3.4 Is teacher's attitude towards repetition exogenous?

The previous section assessed whether the difference between our proxies and the teacher's attitude towards repetition ( $\nu_k$ ) is exogenous. In this section, we assess whether  $\nu_k$  is exogenous. Two reasons could cause that  $\nu_k$  is correlated with the error term. First, the placement of the teacher could be endogenous. Second, the teacher's attitude towards repetition is random, but it is correlated with another characteristic causing dropouts.

If the placement of the teacher is endogenous, and the causes of his placement (teacher's school attainment, experience ...) are correlated with  $\nu_k$ , then  $\nu_k$  may be correlated with the unobservables causing dropouts  $u_{ik}$ . For example, if elder teacher are placed in urban schools, and  $\nu_k$  is correlated with the teacher's age, then  $\nu_k$  may be correlated with parental preferences for schooling. In that case,  $\nu_k$  would be endogenous, so that our proxies  $LP_{ik}$  and  $\widetilde{R}_{ik}$  would not control for the endogeneity of  $R_{ik}$ . However, we would also expect that the repetition rate is higher in urban schools, which is not the case in table 3.

---

<sup>6</sup> $LP_{ik}$  is a proxy for  $t_k$ . If  $\lambda = 1$ , then  $t_k = \overline{S}_k + \nu_k$

date $t$	date $t + 1$	
Enrolled	Enrolled	} Grade repetition decision is observed
	Enrolled	
	Drops out	} Grade repetition decision is not observed

Table 4: Observation of grade repetition decision

Moreover, the observables characteristics of the schools are probably the main determinants of the teacher's placement<sup>7</sup>. As a result, if the repetition rate is not correlated with the observables characteristics of the schools, which is the case in table 3, we do not expect that endogenous placement generates the endogeneity of  $\nu_k$ .

Besides, once controlled for the grade, the determinants of the teacher's placement are probably the characteristics of the schools. In that case, the modification of our identification hypotheses in order to control for school fixed effects à la Chamberlain (1980) in section 5.1 controls for this potential endogeneity bias.

If the teacher's attitude towards repetition is correlated with another characteristic causing dropouts, then our proxies do not control for the endogeneity of grade repetition. In section 5.2, we use the non-linearity of the function  $f$  to assess whether this effect is plausible. In that case, our theoretical model takes into account a potential effect of  $\nu_k$  on  $E_{ik,t+1}$ :

$$\begin{cases} R_{ik} = \mathbb{1} [S_{ik} - \lambda \bar{S}_k - \nu_k + \delta \mathbb{1}(S_{ik} - \lambda \bar{S}_k - \nu_k > 0) + X_{ik} \beta_r + \epsilon_{ik} < 0] \\ E_{ik,t+1} = \mathbb{1} [\beta_{e1} S_{ik} + \beta_{e2} \bar{S}_k + \beta_{e3} \nu_k + X_{ik} \beta_{e4} + \gamma R_{ik} + u_{ik} > 0] \end{cases} \quad (10)$$

In equation (10), the identification relies on the discontinuity of the probability of grade repetition when  $S_{ik} = t_k$ . However, we measure  $t_k$  noisily, so that it does not make sense to estimate the effect of grade repetition on dropouts with a regression discontinuity design.

In section 5.2, we observe that the estimation of this model finds a correlation between teacher's attitude towards repetition and the average dropout rate. However, the estimated effect of grade repetition on school dropouts remains negative and significant. Depending on the specification, the marginal effect is bigger and very unprecisely estimated, or close to the benchmark and significantly different from 0.

### 3.5 Selection on grade repetition

As mentioned in section 2, we cannot observe all the grade repetition decisions. In particular, if a child drops out just before a test, we do not know which the repetition decision was the year before this test. We sum up the structure of the data in table 4.

Because of this selection problem, the question arises as to the identification of the effect of grade repetition decision on school dropouts. In fact, if grade repetition cause dropouts, then it causes its own selection. It is nevertheless possible to control for the selection on repetition and hence identify the determinants of grade repetition.

In appendix C.1, we prove that the semiparametric identification of model (11) is theoretically possible. We give a very simple intuition for that: we have an instrument for grade repetition and an instrument for the selection. In that case, the system of all the derivatives of the probability functions

<sup>7</sup>Teachers' placement is centralized in Senegal. That is why observable informations are probably the most important determinant of the teachers' placement.

	probit model	joint estimation	
	<i>repetition</i>	<i>repetition</i>	<i>selection</i>
Test score	-.452 (.121)***	-.489 (.109)***	.143 (.110)
Group mean test score	.306 (.124)**	.262 (.122)**	.010 (.147)
Previous year's test score	-.312 (.066)***	-.293 (.061)***	-.038 (.078)
Household's wealth	-.019 (.024)	-.029 (.023)	.057 (.026)**
Parental mean education	-.023 (.036)	.002 (.034)	-.082 (.034)**
Last passer's test score	.339 (.098)***	.340 (.092)***	
The test score is higher than the last passer's score	-.406 (.148)***	-.454 (.138)***	
The test score is higher than the first repeater's score	-.479 (.107)***	-.348 (.112)***	
Negative shock on harvests this calendar year or the next			.473 (.173)***
Grade repetition			-.988 (.424)**
Obs.	1580	1818	
log-likelihood	-521.523	-1175.590	
$\chi^2$ grade year dummies	4.881	5.661	11.554
corresponding p value	.300	.226	.021
$\chi^2$ instruments	97.610	89.962	7.455
corresponding p value	$< 10^{-20}$	$< 10^{-15}$	.006

Additional covariates in each equation: grade-year dummies

Note: \*\*\*, \*\* and \* mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard errors of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Table 5: Joint estimation of the determinants of grade repetition and selection (model (11))

has an unique solution.  $\gamma_s$  is not identified by this system, since  $R_{ik}$  is binary. However, Vytlačil and Yildiz (2006) shows that the coefficient for the endogenous variable is identified.

$$\begin{cases} R_{ik} = \mathbb{1}[S_{ik} - \lambda_1 \overline{S}_k - \lambda_2 LP_{ik} + \delta \mathbb{1}(S_{ik} - LP_{ik} > 0) + X_{ik} \beta_r + \epsilon_{ik} < 0] \\ selection = \mathbb{1}[\beta_{s1} S_{ik} + \beta_{s2} \overline{S}_k + \beta_{s3} Z_s + X_{ik} \beta_{s4} + \gamma_s R_{ik} + v_{ik} > 0] \end{cases} \quad (11)$$

In table 5, the column (1) reports the determinants of grade repetition with a probit specification, with no control for selection. In the columns (2) and (3) we estimate model (11) with a maximum likelihood method. Accordingly, we control for the selection on  $R_{ik}$  in that model. The error terms follow a bivariate normal law. The data are pooled between grades and years. The standard errors of the estimators are corrected for the correlation of the residuals between different observations of the same child. Each specification includes grade-year dummies in each equation, the  $\chi^2$  statistics for their joint significance is reported.

**The determinants of grade repetition** In table 5, the first column corresponds to the determinants of grade repetition. In this table, we can observe that the coefficients are similar in columns (1) and (2), which means that they are not really affected by the correction for selection.

Most of the coefficients are similar to the coefficients of table 2 and will not be commented in this section.

A high previous year test score is associated with a lower probability of grade repetition, and the coefficient for the test score is closer to 0 than in table 2. If the test score is a noisy proxy for current learning achievement, then we expect that previous year test score is another proxy for current learning achievement<sup>8</sup>.

**The determinants of selection** The estimation of *selection* in model (11) aims at controlling for the selection bias in the estimation of  $R_{ik}$ . Accordingly, there is no particular interpretation to those coefficients. The determinants of *selection* may be the determinants of moving or missing school the day of the tests, as well as the determinants of dropouts.

We will nevertheless 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 are not expected to be a determinant of grade repetitions because the rainfall seasons in Senegal begins in July and ends in September. We can consequently theoretically rule out the possibility that teachers could use this information for grade repetitions.

The effect of this shock on selection is positive: 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 that the children may be more likely to take the tests when there is a shock. The F-test for the significance of this instrument is 7.5.

We have shown that the control for the selection barely affect the coefficients of the determinants of grade repetition. For that reason, we do not control for the selection bias in our benchmark specification. However, we will show suggestive evidence that controlling for the selection would not affect the effect of grade repetition on dropouts<sup>9</sup>.

## 4 The effect of grade repetition on school dropouts

In the section 3, we have investigated our instrumental strategy. In this section, we will use this instrumental strategy to identify the effect of grade repetition on school dropouts.

The model (12) addresses the endogeneity problem. This model neglects the selection problem, which seems reasonable after the compareason of columns 1 and 2 of table 5. In addition, appendix C.2 proves that the sign of the effect of grade repetition on school dropouts can be semiparametrically identified without taking into account the selection on  $R_{ik}$ .

The idea for that is very simple: the sign of the derivatives of the probability of  $R_{ik}$  towards the instruments are identified. In fact, if an increase in  $LP_{ik}$  is associated with an increase of  $\mathbb{P}(R_{ik} = 1)$ , this is due to the effect of  $LP_{ik}$  on  $R_{ik}$  and not to selection (if the instrument is valid, and because of the exclusion restriction). Because of the exclusion restriction, if the derivative of  $\mathbb{P}(E_{ik,t+1} = 1)$  towards the instruments is different from 0, it is due to the effect of grade repetition on school dropouts. We identify the sign of the effect of  $LP_{ik}$  on  $R_{ik}$ , and the sign of the reduced form effect of  $LP_{ik}$  on  $E_{ik,t+1} = 1$ , so that the sign of the effect of  $LP_{ik}$  on  $E_{ik,t+1} = 1$  is identified.

Equation (12) is estimated with the maximum likelihood method. If the information about repetition is missing, the likelihood is  $\mathbb{P}(R_{ik} = 1, E_{ik,t+1} | S_{ik}, \overline{S}_k, LP_{ik}, X_{ik}; \beta, \delta, \gamma, \lambda) + \mathbb{P}(R_{ik} = 0, E_{ik,t+1} | S_{ik}, \overline{S}_k, LP_{ik}, X_{ik}; \beta, \delta, \gamma, \lambda)$ .

---

<sup>8</sup>For that reason, previous year test score is not included in table 2: table 2 tests for  $\lambda = 1$

<sup>9</sup>except in

	<i>repetition</i>	<i>enrolled<sub>t+1</sub></i>
	(1)	(2)
Test score	-.456 (.124)***	.099 (.157)
Group mean test score	.298 (.121)**	.036 (.217)
Previous year's test score	-.303 (.063)***	-.088 (.136)
Household's wealth	-.031 (.024)	.150 (.055)***
Parental mean education	-.028 (.035)	.060 (.074)
Last passer's test score	.326 (.103)***	
The test score is higher than the last passer's score	-.401 (.148)***	
The test score is higher than the first repeater's score	-.473 (.104)***	
Grade repetition		-1.126 (.330)***
<i>(Average marginal effect of grade repetition)</i>		-.053 (.026)**
Obs.	1818	
log-likelihood	-677.162	
$\chi^2$ grade year dummies	7.765	14.920
corresponding p value	.101	.005
$\chi^2$ instruments	97.100	
corresponding p value	$< 10^{-20}$	

Additional covariates in each equation: grade-year dummies

Note: \*\*\*, \*\* and \* mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard errors of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Table 6: Joint estimation of the determinants of grade repetition and school dropouts (model (12))

$$\begin{cases} R_{ik} = \mathbb{1}[S_{ik} - \lambda_1 \overline{S_k} - \lambda_2 LP_{ik} + \delta \mathbb{1}(S_{ik} - LP_{ik} > 0) + X_{ik}\beta_r + \epsilon_{ik} < 0] \\ E_{ik,t+1} = \mathbb{1}[\beta_{e1} S_{ik} + \beta_{e2} \overline{S_k} + X_{ik}\beta_{e3} + \gamma R_{ik} + u_{ik} > 0] \end{cases} \quad (12)$$

In this paper, we cannot compare the estimation of the model with the estimation of a simple probit model: when a child drops out, the information about grade repetition is not available.

In table 6, we estimate model (12). The two columns of table 6 correspond to the two equations of the model. Again, the data are pooled between grades and years. Each specification includes grade-year dummies in each equation, the  $\chi^2$  statistics for their joint significance is reported.

In this section, we do not discuss the determinants of grade repetition, which are the same than in table 5.

**The determinants of school dropouts** Household wealth is positively associated with continuing schooling, which is not surprising. Test scores or parental education are not correlated with dropouts.

In this specification of model (12), the estimated effect of grade repetition on school dropout is negative. The coefficient is different from 0 at the 1% level. It corresponds to an average marginal effect of  $-5.3\%$ . The mean dropout rate being 2% in the sample, the magnitude of the estimated effect is fairly high.

**Controlling for the selection** We estimate the 3-equation model (13) in table 7. This model addresses both the selection problem and the endogeneity of grade repetition.

$$\left\{ \begin{array}{l} R_{ik} = \mathbb{1}[S_{ik} - \lambda_1 \overline{S}_k - \lambda_2 LP_{ik} + \delta \mathbb{1}(S_{ik} - LP_{ik} > 0) + X_{ik} \beta_r + \epsilon_{ik} < 0] \\ E_{ik,t+1} = \mathbb{1}[\beta_{e1} S_{ik} + \beta_{e2} \overline{S}_k + \beta_{e3} Z_s + X_{ik} \beta_{e4} + \gamma R_{ik} + u_{ik} > 0] \\ selection = \mathbb{1}[\beta_{s1} S_{ik} + \beta_{s2} \overline{S}_k + \beta_{s3} Z_s + X_{ik} \beta_{s4} + \gamma_s R_{ik} + v_{ik} > 0] \end{array} \right. \quad (13)$$

In appendix C.1, we prove that in model (13):

- If  $(\epsilon_{ik}, u_{ik}, v_{ik})$  is independent of  $(S_{ik}, \overline{S}_k, LP_{ik}, Z_s, X_{ik})$
- If  $\lambda_2 \neq 0$  and  $\beta_{s3} \neq 0$
- Under some technical hypotheses about points where the distribution of  $(\epsilon_{ik}, u_{ik}, v_{ik})$  should be positive and finite
- If the support  $\Theta$  contains a pair of points  $a = (S_a, \overline{S}_a, LP_a, X_a, Z_{s,a})$  and  $b = (S_b, \overline{S}_b, LP_b, X_b, Z_{s,b})$  verifying:

$$\left\{ \begin{array}{l} S_a - \lambda_1 \overline{S}_a - \lambda_2 LP_a + \delta \mathbb{1}(\overline{S}_a - LP_a > 0) + X_a \beta_r = S_b - \lambda_1 \overline{S}_b - \lambda_2 LP_b + \delta \mathbb{1}(S_b - LP_b > 0) + X_b \beta_r \\ \beta_{s1} S_a + \beta_{s2} \overline{S}_a + \beta_{s3} Z_{s,a} + X_a \beta_{s4} = \beta_{s1} S_b + \beta_{s2} \overline{S}_b + \beta_{s3} Z_{s,b} + X_b \beta_{s4} + \gamma_s \\ \beta_{e1} S_a + \beta_{e2} \overline{S}_a + \beta_{e3} Z_{s,a} + X_a \beta_{e4} = \beta_{e1} S_b + \beta_{e2} \overline{S}_b + \beta_{e3} Z_{s,b} + X_b \beta_{e4} + \gamma \end{array} \right.$$

and if  $\Theta$  includes the neighborhood of  $a$  and  $b$ .

Then all the coefficients of the model (13) are identified without parametric assumption on the distribution of  $(\epsilon_{ik}, u_{ik}, v_{ik})$ .

In the table 7, we estimate the model (13) parametrically. However, this is not the benchmark specification for convergence reasons. The error terms  $(\epsilon_{ik}, u_{ik}, v_{ik})$  follow a trivariate normal law, approximated with a GHK simulator, with 25 iterations in table 7. We observe that the maximum likelihood does not converge with more iterations in the simulator. In those cases, the model generates  $P(selection = 1) < P(schooled_{t+1} = 1)$  in some parts of the support, and the datas are constrained to  $selection = 0$  if  $schooled_{t+1} = 0$ , which explains why the maximization of the likelihood fails. For that reason, our benchmark specification is model (12). However, we find reassuring that the results are very similar between tables 6 and 7: the effect of grade repetition on school dropouts is quantitatively similar:  $-4.9\%$ .

**Partial conclusion** In this section, we have described our methodology to identify the effect of grade repetition on school dropouts. With this methodology, we find a negative and significant effect of grade repetition on school dropouts. However, we need to do some specification check to rule out potential causes of endogeneity of the instrument. Accordingly, we do those checks in section 5.

	<i>repetition</i>	<i>selection</i>	<i>enrolled<sub>t+1</sub></i>
	(1)	(2)	(3)
Test score	-.513 (.101)***	-.104 (.120)	.202 (.141)
Group mean test score	.230 (.114)**	.257 (.145)*	.097 (.199)
Previous year's test score	-.233 (.064)***	-.133 (.087)	-.203 (.138)
Household's wealth	-.032 (.022)	.053 (.031)*	.170 (.055)***
Parental mean education	.019 (.031)	-.100 (.040)**	.095 (.085)
Negative shock on harvests this calendar year or the next		.478 (.202)**	.143 (.261)
Last passer's test score	.323 (.094)***		
The test score is higher than the last passer's score	-.503 (.129)***		
Grade repetition		-2.244 (.652)***	-1.595 (.549)***
<i>(Average marginal effect of grade repetition)</i>			-.049 (.015)***
Obs.	1818	1818	1818
$\chi^2$ grade year dummies	7.249	8.729	22.574
corresponding p value	.123	.068	.0002
$\chi^2$ instruments	84.981	5.607	
corresponding p value	$< 10^{-15}$	.018	

Note: Additionnal covariates in each equation: grade-year dummies

\*\*\*, \*\* 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.

Table 7: Joint estimation of the determinants of grade repetition, selection and school dropouts (model (13))

## 5 Specification checks

### 5.1 Does variation of repetition rate within school identify the same causal effect?

In this paper, the hypothesis that grade repetition rate is independant from the unobservables of the community is crucial. Else, it is possible that the unobservables of the children are correlated within each school; or that teacher's endogenous placement generates the endogeneity of teacher's attitude towards repetition. Table 3 shows that our proxies are not correlated with the observables of the school. This is reassuring, because most of the unobservables of the community are expected to be correlated with observable characteristics, but one can be worried of the power of this test.

For that reason, we can modify our identification hypotheses in order to control for school fixed effects à la Chamberlain (1980). In fact, we can identify our model on the differences of teacher's attitude towards repetition within schools. We rewrite equation (5) and (8) and control for  $\overline{R_{iks}}$ , the

	<i>repetition</i>	<i>enrolled<sub>t+1</sub></i>
	(1)	(2)
School mean of grade repetition rates among the peers	1.456 (.487)***	.701 (.751)
Repetition rate in the group	1.854 (.302)***	
Grade repetition		-.908 (.331)***
<i>(Average marginal effect of grade repetition)</i>		-.045 (.023)**
Obs.	1823	
log-likelihood	-675.133	
$\chi^2$ grade year dummies	8.752	16.639
corresponding p value	.068	.002
$\chi^2$ instruments	37.819	
corresponding p value	$< 10^{-5}$	

Additional covariates in each equation: test score, group mean test score previous year's test score, household's wealth, parental 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 errors of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Table 8: Joint estimation of the determinants of grade repetition and school dropouts (model (14))

school average repetition rate:

$$\begin{cases} R_{ik} = \mathbb{1} \left[ S_{ik} - \lambda \overline{S}_k + \gamma \overline{R_{iks}} + \alpha \widetilde{R}_{ik} + X_{ik} \beta + \epsilon_{ik} < 0 \right] \\ E_{ik,t+1} = \mathbb{1} \left[ \beta_{e1} S_{ik} + \beta_{e2} \overline{S}_k + \beta_{e2b} \overline{R_{iks}} + X_{ik} \beta_{e3} + \gamma R_{ik} + u_{ik} > 0 \right] \end{cases} \quad (14)$$

Estimating the 2 equations of (14), we use  $\widetilde{R}_{ik}$  as an instrument, and control for  $\overline{R_{iks}}$ . We find more coherent to use the group repetition rate as proxy for teacher's attitude towards repetition in that case: we control for its school average. Accordingly, the effect of grade repetition decision is identified on the differences within schools in teacher's attitude towards repetition. This controls a potential correlation between the  $\epsilon$  from the same school.

In table 8, we estimate the model (14). We only report the coefficients for the new covariates, for the instruments and for the effect of grade repetition. However, the other covariates in the estimation are the same than in table 6. In this specification, the effect of grade repetition is identified by the differences in teacher's specific attitude towards repetition within schools. We observe that in the schools with a higher repetition rate, the repetition probability is higher, even conditionally to the group repetition rate.

The effect of grade repetition on school dropouts is still negative and significant. Its marginal effect is nearly the same (here  $-4.5\%$ ). In table 11 in appendix B, we identify the corresponding specification of model (13), and find the same results (the average marginal effect is  $-4.9\%$ ).

## 5.2 Is there an effect of teacher's attitude towards repetition on non-repeaters ?

Now, let us suppose that teacher's specific attitude towards repetition is random: the teachers' allocation is independant of their attitude towards repetition. It is still possible that this attitude is correlated with other educational methods. In that case, school dropouts may be spuriously correlated

	<i>repetition</i>	<i>enrolled<sub>t+1</sub></i>
	(1)	(2)
Repetition rate in the group	1.292 (.362)***	1.628 (.566)***
Last passer's test score	.154 (.116)	.096 (.172)
The test score is higher than the first repeater's score	-.245 (.124)**	
The test score is higher than the last passer's score	-.397 (.153)***	
Grade repetition		-2.344 (.796)***
<i>(Average marginal effect of grade repetition)</i>		-.217 (.138)
Obs.	1818	
log-likelihood	-665.079	
$\chi^2$ grade year dummies	4.745	14.485
corresponding p value	.314	.006
$\chi^2$ instruments	10.416	
corresponding p value	.005	

Additional covariates in each equation: test score, group mean test score previous year's test score, household's wealth, parental 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 errors of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Table 9: Joint estimation of the determinants of grade repetition and school dropouts (model (15))

with grade repetition. However, we can control for the correlation between grade repetition and teacher's attitude towards repetition using the discontinuity of  $P(R_{ik} = 1 \text{ when } S_{ik} = t_k)$ . In that case, our empirical model is model (15):

$$\begin{cases} R_{ik} = \mathbb{1} \left[ S_{ik} - \lambda_1 \overline{S}_k - \lambda_2 LP_{ik} + \alpha \widetilde{R}_{ik} + \delta \mathbb{1}(S_{ik} - LP_{ik} > 0) + X_{ik} \beta_r + \epsilon_{ik} < 0 \right] \\ E_{ik,t+1} = \mathbb{1} \left[ \beta_{e1} S_{ik} + \beta_{e2} \overline{S}_k + \beta_{e3a} LP_{ik} + \beta_{e3b} \widetilde{R}_{ik} + X_{ik} \beta_{e4} + \gamma R_{ik} + u_{ik} > 0 \right] \end{cases} \quad (15)$$

In table 9, we estimate the model (15), which is the empirical counterpart of model (10). We control for a potential correlation between teacher's attitude towards repetition and school dropouts. However, this correction relies on the hypothesis that the coefficient of teacher's attitude towards repetition in the equation of dropouts is the same for all children. This estimation is very parametric, since it relies strongly on the non-linearity of the effect of  $t_k$  on grade repetition. In that case, the instrument for grade repetition is the rank relative to first repeater and last passer.

We can observe that our proxies for the teacher's attitude towards repetition are positively correlated with the probability of being enrolled at school the next year. We can give two explanations for this coefficient. First, the teacher's attitude towards repetition may be correlated with any other educational method causing dropouts. Second, the teacher's attitude towards repetition has other consequences than repetitions, and those consequences affect the school dropouts of the passers. In both cases, the effect of grade repetition on school dropouts in tables 6 and 8 is potentially biased.

However, the coefficient for grade repetition is still negative and significant in this specification. The estimated marginal effect (-22%) is largely smaller, but not significantly different from 0. The

marginal effect booms whereas the probit coefficient is multiplied by 2, because of the non-linearity in the probit model. Yet, the marginal effect is very unprecisely estimated.

In the table 12 in appendix B, we identify the corresponding specification of model (13). In that case, the effect of  $\nu_k$  on dropouts is smaller (in terms of probit coefficient) and only significant at the 10% level (the p-value is below 6%). The effect of grade repetition on school dropouts is still negative and significant. In table 12, the marginal effect is  $-6.2\%$ . On the contrary to table 9, the marginal effect is significantly negative and close to the other specifications.

This would confirm that the grade repetition has a negative effect on schooling, our result being robust to the potential causal link between the teacher's attitude towards repetition and dropouts.

In this section, we checked whether a correlation between teacher's attitude and dropouts is likely to bias our results. The table 9 shows that this correlation is probably possible. However, this correlation does not change the sign and significance of the effect of grade repetition on dropouts. The magnitude of the effect of grade repetition on school dropouts is strongly affected by this correction in table 9, but it is not the case in table 12.

## 6 Conclusion

In this paper, we use proxies for the differences between teachers's attitude towards repetition as instruments to identify the effect of grade repetition on school dropouts. With these instruments, we find a negative effect of grade repetition on school dropouts.

The differences in our proxies are not correlated with geographic observable characteristics of the location of the school, which rules out the potential endogeneity due to placement of the teachers or to the correlation between unobserved characteristics of the peers. In addition, we modify our benchmark to take into account potential school fixed effects. With this specification, the results are very similar to the benchmark. Both empirical tests indicate that our result is not biased by school unobservable characteristics.

In these two specifications, a causal effect of teacher's attitude towards repetition on school dropout is a potential source of bias. So as to control for this bias, we use the fact that the effect of teacher's attitude towards repetition on children's grade repetition depends on the child's ranking in his class. In this third specification, the causal effect of grade repetition on school dropouts is negative and slightly stronger than in the benchmark specification.

This paper focuses on the effect of grade repetition on short-term dropout, and grade repetition may have other consequences. First it has a direct effect on the acquisition of knowledge. However, as long as grade repetition causes school dropouts, the evaluation of this effect faces a serious selection problem. In addition, schooling decisions and the acquisition of knowledge are closely linked, and it is doubtful that any conceptually acceptable instrument for this selection can be found.

Second, it focuses on short term consequences of grade repetition, and grade repetition may have long term consequences. The evaluation of the long term effect of grade repetition on school dropouts is a priori possible with the same data, though it is not treated in this paper.

Finally, the teacher's attitude towards repetition is likely to have direct effect on school dropouts. Yet, this effect is poorly evaluated with our data, since the teacher's attitude towards repetition may be correlated with other educational methods. However, we estimate that conditionally to grade repetition, the dropout probability is lower when teacher's attitude favours grade repetition.

## References

- Chamberlain, G., 1980. Analysis of covariance with qualitative datas. *Review of economics and statistics* 47, 225–238.
- Holmes, C. T., 1989. Grade level retention effects: A meta-analysis of research studies. in Shepard and Smith (1989) .
- Jacob, B., Lefgren, L., 2004. Remedial education and student achievement: A regression - discontinuity analysis. *Review of economics and statistics* 86, 226–244.
- King, E. M., Orazem, P. F., Paterno, E. M., 1999. Promotion with and without learning: Effects on student dropout. Working paper, World Bank.
- 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'éducation 2000 - 2005. Tech. rep., Ministry of education, Senegal.
- 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.

Shepard, L. A., Smith, M. L., 1989. Flunking Grades: Research and Policies on Retention. London: Falmer Press.

Vytlacil, E., Yildiz, N., 2006. Dummy endogenous variables in weakly separable models, forthcoming in *Econometrica*.

case 1	case 2	case 3	case 4	case 5	case 6	school year
<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	1995 - 1996
<b>2</b>	<b>2,3</b>	drop.	<b>3</b>	<b>3</b>	<b>3</b>	1996 - 1997
<b>3</b>	<b>3</b>		<b>3,4</b>	<b>4</b>	<b>3</b>	1997 - 1998
<b>4</b>	<b>4</b>		<b>3,4,5</b>	<b>5</b>	<b>3</b>	1998 - 1999
<b>5</b>	<b>5</b>		<b>3,4,5,6</b>	<b>6</b>	<b>4</b>	1999 - 2000

(When the child did not attend to the tests, the possible grades are written in grey)

Table 10: Grade attended during the PASEC panel for 6 imaginary cases

## A The variables

**Repetition** is a dummy taking value 1 if the child repeated the grade, and 0 otherwise. Information is obtained from the PASEC panel. In each case, we tried to infer each year whether the child did pass at the end of the school year. Table 10 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. We know that he has 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 he did not take the test is not reported. Consequently, we cannot know whether he has repeated the second or the third grade. In case 3, the child drops out in 1996. We can consequently not know whether he was admitted in the third grade after school year 1995 - 1996. In case 4, the child is not in the sample after 1997 - 1998, so that we do not know whether he has repeated during the subsequent grades. In cases 5 and 6, we can know whether the child repeated since the child repeated twice (case 6) or passed twice (case 5) when he was not observed, so that grade repetitions are not ambiguous.

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

**Test scores** are a proxy for learning achievement at the end of the current school year. In fact, PASEC panel contains school tests at the end of each academic year until the end of the survey<sup>10</sup>. The marking of the tests is done by the PASEC team. Consequently, the test scores cannot be influenced by the teachers. We reported in table 1 the number of children attending each test.

The tests were designed to ensure easy comparisons within grade - years. They are nevertheless different 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.

**Previous year's test scores** are a proxy for learning achievement prior to the current school year. We use the fact that during the panel, the children took tests at the end of the school year before. However, in each grade-year of the panel, some of the children were in the grade before the year before. The others were in the same grade the year before, and are currently repeating their grade. The tests are different between currently repeating children and others. Yet, some items were common between the two tests, and we use those items to compare the knowledge of the pupils prior to the current

<sup>10</sup>The second grade classes were not surveyed from 1997 - 1998, so that pupils still in this grade at that time were not surveyed until they passed to the third grade.

school year. Again, this variable has a mean of 0 and a standard deviation of 1 within each grade-year. This comparison relies only on skills of the grade before, since the tests never include items about the skills supposed to be acquired in the following grades.

**Parental education** is the mean of both parents' education. The education of an individual equals 1 if the individual never went to school, 2 if the person began but did not finish primary school, 3 if he finished primary school but did not begin secondary school, etc. If the information about the father's education or the mother's education is missing, it is replaced by the mean education of the other adults (aged more than 25 in 1995) in the household.

**Household wealth** is a composite indicator for possession of durable goods, obtained with a principal component analysis. It is based on child's declaration in 1995, and hence avoids reverse causality due to the child's education.

**Negative shocks on harvests** is a dummy taking value 1 if the head of the household reports a negative shock on harvests. We take these shocks into account if the child or his parents are still in the household visited by EBMS in 2003. Else this dummy equals 0, because the child is not really affected by these shocks. The cases when neither the child nor his parents are in the household in 2003 are treated differently. However, for all the specifications presented, we checked that including a dummy for those cases does not change the effect of grade repetition on school dropouts.

**Repetition rate in the group** is a proxy for the teacher's specific attitude towards repetition. A group is defined by all the children being in the same school and the same grade in a given school year<sup>11</sup>. Among the peers of a given child a given year, some are admitted to the next grade and we call them the "passers". Others must repeat their grade if they don't drop out and are called "repeaters". The repetition rate in the group is the proportion of "repeaters" among the peers. It is calculated among the peers that are unambiguously passers or repeaters.

Among the peers, we single out the passer with the lowest test score, and call him the "last passer".

**Last passer's test score** is another proxy for the teacher's specific attitude towards repetition. In fact, if the last passer's score is high, it is supposed to be more frequent that a given child repeats.

**The test score is higher than the last passer's score** is a dummy taking value 1 if the child's test score is higher than the last passer's score, and 0 otherwise. The idea that a child has to repeat if his learning achievement is below a certain threshold level is widespread. If there are differences between teachers in the attitude towards repetition, we suppose that this level of learning achievement may change between teachers. That is why we use the test score of the last passer as a proxy for it. Accordingly, the dummy is a proxy for the fact that the child's achievement is above the threshold.

Among those not admitted to the next grade, the one with the highest test score is the "first repeater".

**The test score is higher than the first repeater's score** is a dummy taking value 1 if the child's test score is higher than the last passer's score, and 0 otherwise. If there is no repeater in the group, the dummy for the "test score higher than the first repeater's score" equals 1 for every child.

---

<sup>11</sup>A group is an approximation of a class: there may be several classes per group in some cases. In fact, there may be several classes per grade in some schools. In that case, although all the pupils are in the same class in the first year of the panel, in the following years they may be in the same grade and in different classes.

## B Results with the model(13)

	<i>repetition</i>	<i>selection</i>	<i>enrolled<sub>t+1</sub></i>
	(1)	(2)	(3)
School mean of grade repetition rates among the peers	.866 (.453)*	1.267 (.572)**	.974 (.713)
Negative shock on harvests this calendar year or the next		.465 (.181)**	.193 (.265)
Repetition rate in the group	1.646 (.291)***		
Grade repetition		-.978 (.556)*	-1.425 (.599)**
<i>(Average marginal effect of grade repetition)</i>			-.049 (-.016)***
Obs.	1823	1823	1823
$\chi^2$ grade year dummies	9.585	10.999	19.887
corresponding p value	.048	.027	.0005
$\chi^2$ instruments	32.046	6.598	
corresponding p value	$< 10^{-5}$	.010	

Additional covariates in each equation: test score, group mean test score previous year's test score, household's wealth, parental 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.

Table 11: Joint estimation of the determinants of grade repetition, selection and school dropouts with Chamberlain (1980) fixed effects

	<i>repetition</i>	<i>selection</i>	<i>enrolled<sub>t+1</sub></i>
	(1)	(2)	(3)
Repetition rate in the group	.837 (.502)*	1.094 (.622)*	.950 (.501)*
Last passer's test score	.279 (.092)***	-.115 (.160)	.073 (.149)
Negative shock on harvests this calendar year or the next		.492 (.192)**	.161 (.268)
Rank relative to first passer and last repeater	-.551 (.185)***		
Grade repetition		-1.627 (1.670)	-1.607 (.820)**
<i>(Average marginal effect of grade repetition)</i>			-.062 (.029)**
Obs.	1818	1818	1818
$\chi^2$ grade year dummies	3.829	7.985	19.843
corresponding p value	.430	.092	.0005
$\chi^2$ instruments	8.884	6.526	
corresponding p value	.003	.011	

Additional covariates in each equation: test score, group mean test score previous year's test score, household's wealth, parental 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.

Table 12: Joint estimation of the determinants of grade repetition, selection and school dropouts corresponding to the model (15)

**Rank relative to first passer and last repeater** denotes the fact that a child's test score is bigger than the last passer's score and the first repeater's score. It takes value 2 if the child's score is higher than both comparison scores (i.e. the last passer's score or the first repeater's score). It takes value 1 if the child's score is higher than one of the two comparison scores. It is 0 otherwise.

## C Proofs for the semiparametric identification of model (13)

### C.1 model (13)

In this section, we prove that model (13) can be semiparametrically identified. In addition, it proves that the models (11) can be semiparametrically identified: the equation for  $e$  is necessary to identify neither the coefficients of  $r$  nor the coefficients of  $s$ .

The model (13) is :

$$\begin{cases} r &= \mathbb{1}(X\beta_r + \gamma_r Z_1 + \varepsilon_r > 0) \\ s &= \mathbb{1}(X\beta_s + \gamma_s Z_2 + \alpha_s r + \varepsilon_s > 0) \\ e &= \mathbb{1}(X\beta_e + \gamma_e Z_2 + \alpha_e r + \varepsilon_e > 0) \end{cases} \quad (16)$$

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

We recall that  $r$  is observed if and only if  $s = 1$ . We write  $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$  the repartition function of  $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ . In that case, Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the repartition function. We use this idea to show that all the parameters of model (13) are identified without any parametric assumption on  $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ .

We note  $\Theta$  the support of  $(X, Z_1, Z_2)$  and make the following assumptions:

1. The distribution of  $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$  is independant 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)  $\begin{cases} X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} \\ X_a\beta_s + \gamma_s Z_{2a} + \alpha_s = X_b\beta_s + \gamma_s Z_{2b} \\ X_a\beta_e + \gamma_e Z_{2a} + \alpha_e = X_b\beta_e + \gamma_e Z_{2b} \end{cases}$
  - (b) In the neighborhood of  $a$  and  $b, (X, Z_1, Z_2) \in \Theta$  and  $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$

The assumption 1 is necessary in Manski (1988) and is still necessary in this case. It ensures that the derivatives of the probabilities functions towards  $X, Z_1$  or  $Z_2$  are not caused by variations of  $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ .

The assumption 2 ensures that the instruments have a real causal effect on the endogenous variables.

In the model (13), we observe 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  $X\beta_e + \gamma_e Z_2 + \alpha_e r + \varepsilon_e)$ . Accordingly, the parameters are identified up to the scale of the parameters vector. The assumption 3 easily fixes that scale.

The assumption 4a ensures that it is possible to compute the derivatives of the probabilities 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 when  $X$  contains some binary variables.

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

The assumption 4c ensures in  $(X_0, Z_{10}, Z_{20})$ , the other derivatives of the probability functions towards the covariates are not null.

The assumption 5 ensures that the support  $\Theta$  is large enough to contain pair of points that similar characteristics for  $s$  and  $e$  when the first has  $r = 1$  and the latter has  $r = 0$ .

This proof has 3 steps: first, we show that the coefficients  $\beta$  and  $\gamma$  of the 2 first equations of model (13) are identified, then we show that the coefficients  $\beta$  and  $\gamma$  of the last equation are identified, and finally, we show that the  $\alpha$  are identified.

### • Identification of the 2 first equations of the model

We 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 datas in  $(X_0, Z_{10}, Z_{20})$  if the 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  $F^{(11)}$  towards its 2 arguments. The derivatives are:

$$\frac{dP^{(11)}}{dX_1} = F_1^{(11)} + F_2^{(11)} \quad (17)$$

$$\frac{dP^{(11)}}{dX_i} = \beta_{ri} F_1^{(11)} + \beta_{si} F_2^{(11)} \quad (\forall i \in \{1..K\}) \quad (18)$$

$$\frac{dP^{(11)}}{dZ_1} = \gamma_r F_1^{(11)} \quad (19)$$

$$\frac{dP^{(11)}}{dZ_2} = \gamma_s F_2^{(11)} \quad (20)$$

This is clearly not sufficient to identify the  $\beta$  and  $\gamma$ . In fact, we have in these 4 equations 6 unknown parameters, since  $F_1^{(11)}$  and  $F_2^{(11)}$  are unknown. That is why we use the derivatives of  $\mathbb{P}(r = 0, s = 1|X, Z_1, Z_2)$  to identify  $\gamma$  and  $\beta$ .

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

We note  $F_1^{\prime(01)}$  and  $F_2^{\prime(01)}$  the derivatives of  $F^{(01)}$  towards its 2 arguments.

$$\frac{dP^{(01)}}{dX_1} = F_1^{\prime(01)} + F_2^{\prime(01)} \quad (21)$$

$$\frac{dP^{(01)}}{dX_i} = \beta_{ri}F_1^{\prime(01)} + \beta_{si}F_2^{\prime(01)} \quad (22)$$

$$\frac{dP^{(01)}}{dZ_1} = \gamma_r F_1^{\prime(01)} \quad (23)$$

$$\frac{dP^{(01)}}{dZ_2} = \gamma_s F_2^{\prime(01)} \quad (24)$$

From the equations (17) rearranged with (19) and (20), and (21) rearranged with (23) and (24), we get the 2 equations system:

$$\begin{cases} \frac{dP^{(11)}}{dX_1} = \frac{1}{\gamma_r} \frac{dP^{(11)}}{dZ_1} + \frac{1}{\gamma_s} \frac{dP^{(11)}}{dZ_2} \\ \frac{dP^{(01)}}{dX_1} = \frac{1}{\gamma_r} \frac{dP^{(01)}}{dZ_1} + \frac{1}{\gamma_s} \frac{dP^{(01)}}{dZ_2} \end{cases}$$

Under assumptions 4b and 2, This identifies  $\gamma_s$  and  $\gamma_r$ . Then we can compute easily  $F_1^{\prime(11)}$ ,  $F_2^{\prime(11)}$ ,  $F_1^{\prime(01)}$  and  $F_2^{\prime(01)}$  with (19), (20), (23) and (24). The system:

$$\begin{cases} \frac{dP^{(11)}}{dX_i} = \beta_{ri}F_1^{\prime(11)} + \beta_{si}F_2^{\prime(11)} \\ \frac{dP^{(01)}}{dX_i} = \beta_{ri}F_1^{\prime(01)} + \beta_{si}F_2^{\prime(01)} \end{cases}$$

identifies the  $\beta_{ri}$  and the  $\beta_{si}$ .

In fact, assumption 2 ensures that  $\begin{pmatrix} \gamma_r F_1^{\prime(11)} & \gamma_r F_1^{\prime(01)} \\ \gamma_s F_2^{\prime(11)} & \gamma_s F_2^{\prime(01)} \end{pmatrix}$  has full rank, so that  $\begin{pmatrix} F_1^{\prime(11)} & F_1^{\prime(01)} \\ F_2^{\prime(11)} & F_2^{\prime(01)} \end{pmatrix}$  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  $F_1^{\prime(1)}$ ,  $F_2^{\prime(1)}$  and  $F_3^{\prime(1)}$  the derivatives of  $F^{(1)}$  towards its arguments. We compute the derivatives of  $P^{(1)}$ :

$$\frac{dP^{(1)}}{dX_1} = F_1'^{(1)} + F_2'^{(1)} \quad (25)$$

$$\frac{dP^{(1)}}{dX_i} = \beta_{ri}F_1'^{(1)} + \beta_{si}F_2'^{(1)} \quad (26)$$

$$\frac{dP^{(1)}}{dZ_1} = \gamma_r F_1'^{(1)} \quad (27)$$

$$\frac{dP^{(1)}}{dZ_2} = \gamma_e F_2'^{(1)} \quad (28)$$

$\gamma_r$  is known, so that  $F_1'^{(1)}$  can be easily computed with (27). Then we can compute  $F_2'^{(1)}$  with (25). Under assumption 4c,  $F_2'^{(1)}$  is not null in  $(X, Z_1, Z_2) \in \Theta$ . That is why  $\gamma_e$  is identified by (28). The knowledge of the  $\beta_{ri}$ ,  $F_1'^{(1)}$  and  $F_2'^{(1)}$  identifies the  $\beta_{si}$  in (26).

- **Identification of  $\alpha_s$ .**

Adapting Vytlacil and Yildiz (2006), we can easily 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<sup>12</sup>

$$\begin{cases} X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\ X_c\beta_r + \gamma_r Z_{1c} = X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\ X_a\beta_s + \gamma_s Z_{2c} = X_c\beta_s + \gamma_s Z_{2c} = \kappa_{s1} \\ X_b\beta_s + \gamma_s Z_{2b} = X_d\beta_s + \gamma_s Z_{2d} = \kappa_{s2} \end{cases} \Leftrightarrow \begin{cases} \mathbb{P}(r|a) = \mathbb{P}(r|b) \\ \mathbb{P}(r|c) = \mathbb{P}(r|d) \\ \hat{\mathbb{P}}(s|a) = \hat{\mathbb{P}}(s|c) \\ \hat{\mathbb{P}}(s|b) = \hat{\mathbb{P}}(s|d) \end{cases} \quad (29)$$

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

Then

$$\left( \begin{array}{c} \mathbb{P}(r = 1, s = 1|a) - \mathbb{P}(r = 1, s = 1|c) \\ = - [\mathbb{P}(r = 0, s = 1|b) - \mathbb{P}(r = 0, s = 1|d)] \end{array} \right) \Rightarrow \kappa_{s1} + \alpha_s = \kappa_{s2} \quad (30)$$

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

---

<sup>12</sup> $\hat{\mathbb{P}}$  means that the probability is net of the effect of  $r$  on  $o$

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

(29) 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 (30):

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 (30):

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

We can now rewrite the first term of (30):

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

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

Assumption 5 ensures that some points verifying (29) and (30) exist in  $\Theta$ . In fact, the points  $a$  and  $b$  in the assumption 5 verify (29) and the second term of (30). We can find  $c$  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 points that have the same  $P(r)$ , since  $P(r|a) = P(r|b)$ .

We can recognize these points because the validity of (29) 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  $\alpha_e$ .**

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

$$\left\{ \begin{array}{l} X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\ X_c\beta_r + \gamma_r Z_{1c} = X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\ X_a\beta_s + \gamma_s Z_{1a} = X_c\beta_s + \gamma_s Z_{1c} = \kappa_{s1} \\ X_b\beta_s + \gamma_s Z_{1b} = X_d\beta_s + \gamma_s Z_{1d} = \kappa_{s2} \\ X_a\beta_e + \gamma_e Z_{2a} = X_c\beta_e + \gamma_e Z_{2c} = \kappa_{e1} \\ X_b\beta_e + \gamma_e Z_{2b} = X_d\beta_e + \gamma_e Z_{2d} = \kappa_{e2} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \mathbb{P}(r|a) = \mathbb{P}(r|b) \\ \mathbb{P}(r|c) = \mathbb{P}(r|d) \\ \hat{\mathbb{P}}(s|a) = \hat{\mathbb{P}}(s|c) \\ \hat{\mathbb{P}}(s|b) = \hat{\mathbb{P}}(s|d) \\ \hat{\mathbb{P}}(e|a) = \hat{\mathbb{P}}(e|c) \\ \hat{\mathbb{P}}(e|b) = \hat{\mathbb{P}}(e|d) \end{array} \right. \quad (31)$$

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

$$\left( \begin{array}{l} \mathbb{P}(r = 1, s = 1, e = 1|a) - \mathbb{P}(r = 1, s = 1, e = 1|c) \\ = - [\mathbb{P}(r = 0, s = 1, e = 1|b) - \mathbb{P}(r = 0, s = 1, e = 1|d)] \end{array} \right) \Rightarrow \kappa_{e1} + \alpha_e = \kappa_{e2} \quad (32)$$

For the same reason that for the identification of  $\alpha_s$ , the converse of 32 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 (32):**

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 (32):

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

We can now rewrite the first term of (30):

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

$f(\varepsilon_r, \varepsilon_s, \varepsilon_e) > 0$  in the neighborhood of any point of  $\Theta$  (assumption 4c). As a consequence, it is strictly positive in a subset of the integration intervalle which strictly positive Lebesgue measure if  $\kappa_{e1} + \alpha_e \neq \kappa_{e2}$ . That is why  $\kappa_{e1} + \alpha_s = \kappa_{e2}$ .

Assumption 5 ensure that those points exists, so that we can identify  $\alpha_e$

## C.2 Model (13) without $Z_2$

In this appendix, we prove that  $Z_2$  is not necessary to identify the sign of  $\alpha_e$ . Accordingly, it is theoretically not necessary to control for the selection to identify semiparametrically the sign of  $\alpha_e$ . The corresponding model is:

$$\begin{cases} r = \mathbb{1}(X\beta_r + \gamma_r Z & +\varepsilon_r > 0) \\ s = \mathbb{1}(X\beta_s & +\alpha_s r & +\varepsilon_s > 0) \\ e = \mathbb{1}(X\beta_e & +\alpha_e r & +\varepsilon_e > 0) \end{cases} \quad (33)$$

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

We recall that  $r$  is observed if and only if  $s = 1$ . We write  $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$  the repartition 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 dependant variable. We use this idea to show that the sign of  $\alpha_e$  is identified in model (33) without any parametric assumption on  $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ .

We note  $\Theta$  the support of  $(X, Z)$  and make the following assumptions:

1. The distribution of  $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$  is independant 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  $(-X_0\beta_r - \gamma_r Z_0, -X_0\beta_s - \alpha_s, -X_0\beta_e - \alpha_e)$ , called  $\Gamma$

The assumption 1 is necessary in Manski (1988) and is still necessary in this case. It ensures that the derivatives of the probabilities functions towards  $X$  or  $Z$  are not caused by variations of  $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ .

The assumption 2 ensures that the instrument has a causal effect on  $r$ .

The assumption 3a ensures that it is possible to compute the derivatives of the probabilities 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 when  $X$  contains some binary variables.

The assumption 3b ensures that the density of  $\varepsilon_r$  in  $-X_0\beta_r - \gamma_r Z_0$  is finite, so that the derivatives of the probabilities towards  $Z$  are finite.

The assumption 3c ensures that the derivatives of the probabilities functions towards  $Z$  are not null.

#### – Proof that the sign of $\gamma_r$ is identified

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

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

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 than  $\gamma_r$ .

– **Proof that the sign of  $\alpha_e$  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  $\gamma_r$  is identified, so that the sign of  $\alpha_e$  is identified.