The effect of grade repetition on school dropout. An identification based on differences among teachers.*

Pierre André†

September 10, 2008

JEL Classification: D12, I28, O12
Key words: Grade repetition, School demand, School dropouts, Senegal

Abstract

This paper investigates the connection between grade repetition and school dropout. Household data is matched against a panel of academic test scores and the school career of each child inferred from the combined dataset. The potential endogeneity of grade repetition is corrected for using the differences among teacher attitude to repetition as an instrument for grade repetition. The 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 Behaghel and Christelle Dumas for their very useful comments and help on earlier versions of this paper; and to Christopher Sutcliffe for grammatical corrections.

†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: some 12% of the pupils enrolled in Senegalese primary schools in 2004 were repeating their grades in 2005.[1] This widespread practice is very expensive for the state and households alike since both private and public costs of schooling increase with the duration of schooling.

Whether the costs of grade repetition are compatible with universal primary education in developing countries is seriously debated in multilateral institutions. The World Bank's publication Bruns, Mingat, and Rakotomalala (2003) observes that the developing countries with high primary completion rates faces relatively low repetition rates. On the basis of cross county OLS, the authors conclude that average repetition rate has a strong negative effect on primary completion rate, suspecting it is due to the state's budget constraint. The average repetition rate is included in the "Education For All indicative framework", which is the benchmark for getting EFA Fast Track Initiative fincancing for primary education.

However, the cost of grade repetition has to be compared with its potential consequences: positive if it improves learning achievement, negative if it causes dropout. Indeed, dropout before completion of final (sixth) grade is very frequent in Senegal: some 40% of Senegalese children enrolled in first grade drop out before reaching sixth grade[1]. This paper inquires whether frequent school dropout is not in part a consequence of high repetition rates.

Grade repetition affects schooling decisions through a variety of mechanisms. On the one hand, it has an effect on the acquisition of knowledge. If grade repetition is pedagogically effective, it may prevent dropout. On the other hand, grade repetition may be discouraging.

Grade repetition modifies the learning achievement at a given date. When children repeat grades they may consolidate the skills expected at those grades. However, it is unclear whether this offsets their failure to acquire the skills taught at the next grade. The net effect of grade repetition on the acquisition of knowledge is ambiguous, then. Empirical evaluations of the net effect of grade repetition on learning achievement have serious shortcomings. 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). However, teachers probably use their private information on pupils to decide whether they will repeat. If low motivation at school causes grade repetition, these studies probably suffer from an endogeneity bias: low motivation at school deters future acquisition of knowledge. Jacob and Lefgren (2004) control for this potential bias using a discontinuity in school policy in Chicago. Pupils there took standardized tests at the end of grades 3, 6 and 8. They were promoted if their test score was higher than a minimum score. Regression-discontinuity analysis revealed a small and positive effect of grade repetition on academic achievement at a given date.

Grade repetition may be discouraging for a least two reasons. First, it extends the time needed to achieve a given final grade and get the benefits from education. So grade repetition may increase the cost of schooling: 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. Second, grade repetition may be a negative signal about a child's ability. If the parents observe their children's ability noisily, then grade repetition diminishes parents' belief in their children's ability. Grade repetition possibly causes school dropout for these two reasons.

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

Very few studies have tried to estimate this effect in developing countries. King, Orazem, and Paterno (1999) report that grade repetition causes school dropout in Pakistan. Yet, their identification strategy does not include any control either for the acquisition of knowledge or for parental preferences

[1] Ministry of Education, Senegal (2005)

for schooling. The two are certainly correlated and low parental preferences for schooling possibly cause grade repetition. Consequently, the effect of grade repetition on school dropout 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. Many covariates are controlled for and test scores used as a proxy for the acquisition of knowledge and for ability. However, it is not certain that the remaining unobservable variables causing grade repetition are uncorrelated with future school dropout. Furthermore, attrition in the last years of the panel may be a poor proxy for school dropout. Children may still be enrolled but not have taken the tests because of illness or because they changed schools.

This paper combines PASEC [2] data with fresh survey information to evaluate the effect of grade repetition on school dropout. An original instrumental variables strategy is used to control for the potential correlation between the children's unobservable characteristics and grade repetition.

These instruments are based on teacher attitude to repetition: grade repetition is based on the teacher's decision, and controlling for the learning achievement, those decisions are partly based on teacher's idiosyncrasies. These are unobservable so, for each child, grade repetition by peers is used to proxy for teacher attitude.

The results reveal a negative and significant effect of grade repetition on the probability of enrolment at school the next year. The estimated effect is fairly high: the estimations show that grade repetition increases the probability of school dropout by approximately 5 percentage points on average, whereas the average dropout rate in the sample is 2%.

Section 2 presents the dataset used to identify the causal effect of grade repetition on school dropout. Section 3 presents the strategy used here for identifying this effect while section 4 gives the benchmark results. Section 5 provides some specification checks. Finally, brief remarks are made by way of conclusion.

2 The data

PASEC and EBMS datasets both contain detailed information about schooling and are combined here to estimate the effect of grade repetition.

2.1 The PASEC panel

The PASEC conducted a panel survey in 98 Senegalese primary schools between 1995 and 2001. Twenty second grade students were chosen at random in randomly chosen second grade classes in each school at the beginning of the 1995-1996 school year. They passed learning achievement tests at the end of each school year,[3] and were monitored throughout their school careers (including grade repetitions) until the first of them finished primary school (sixth grade) in 2000. Although children were randomly selected among the second grade pupils of the schools in 1995, attrition and grade repetition meant that the children in the same grade-year were increasingly selected as time elapsed.

There were two causes for attrition in this panel. First, dropouts did not take the PASEC tests. Second, the PASEC team organized the tests and collected the data in each of the schools on a given day in each school year. Children missing school that day or no longer attending the surveyed school were not tested.

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

[3] The tests were marked by the PASEC team. Consequently, test scores could not be influenced by teachers.

Table 1: Test attendance in the panel

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

2.2 EBMS Survey[4]

The EBMS survey provides additional information about certain PASEC pupils in 2003. It includes 59 of the schools surveyed between 1995 and 2000. The objective was to resurvey households in each community (village or urban districts) with children who had been in the PASEC panel. Information was collected about the living conditions (wealth, health) and educational levels of the household members. Retrospective data about the school careers of the children surveyed by PASEC meant dropout could be differentiated from other causes of attrition. Consequently, school-leaving dates are known for almost every child re-surveyed (if they had left in 2003). Of the 1177 pupils attending the 59 schools surveyed by PASEC, 921 are in EBMS data after deletion of questionable matches.

2.3 Aggregate dataset

Both datasets provide reliable retrospective information about enrollment. Together they give enough information to reconstruct most instances of grade repetition.[5] This information is necessary for evaluating the impact of repetition on drop out. Another advantage of the aggregate dataset is that it evaluates the individual learning achievement (test scores), which is a crucial determinant of grade repetition. Table 1 shows the number of children attending each test in the sample and reveals children often missed a test even though still enrolled. All 921 children were enrolled in school year 1995-1996 although only 817 attended the test. Definition of all the variables used in this paper can be found in appendix A.

3 The Empirical strategy

This paper seeks to identify the effect of grade repetition, denoted R_{ik}, on school dropout (enrolment during the next school year is denoted $E_{ik,t+1}$), which is the coefficient γ in the equation (1) below. The other determinants of dropout are test score S_{ik}, and a vector of covariates X_{ik}.[6]

[4]Education et Bien-être des Ménages au Sénégal. This survey was designed by a team composed of Peter Glick, David Sahn, and Léopold Sarr (Cornell University, USA), and Christelle Dumas and Sylvie Lambert (LEA-INRA, France), and implemented in association with the Centre de Recherche en Economie Appliquée (Dakar, Senegal).

[5]The details are explained in appendix A

[6]This vector includes grade-year dummies, household wealth parents' education, and group mean test score[7] when not included in the model.

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

The main difficulty in identifying γ is to control for the potential endogeneity of grade repetition. In Senegal, teachers decide whether pupils pass to the next grade or repeat. They probably use their private information about their pupils' school ability. First, teachers' beliefs about learning achievement may be correlated with the parents' beliefs conditionally on test scores. If in addition parents' beliefs about learning achievement affect the schooling decision, then grade repetition is probably endogenous. Second, parental preferences for schooling may affect knowledge acquisition. Teachers may think children whose parents have strong preferences for schooling are more likely to improve their learning achievement in the next years, and need not repeat grades. Again, this would generate endogeneity of grade repetition. In both cases, for a given learning achievement at the end of the current school year as measured by the test score, children with a higher dropout probability are more likely to repeat their grades.

3.1 Modeling grade repetition determinants

Equation (2) below models the determinants of grade repetition. Learning achievement is compared to t_k, which is the learning achievement required to pass in group k.

$$R_{ik} = \mathbb{1}\left[S_{ik} - t_k + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{2}$$

However, grade repetition in the model is not determined solely by whether or not S_{ik} is greater than t_k. Equation (2) takes this into account by including other factors (X_{ik}), such as household wealth or parents' education, which may affect grade repetition. We suspect X_{ik} to affect grade repetition only if S_{ik} is close to t_k, since very high or very low learning achievements will drive grade repetition whatever the individual characteristics are. In that case, the coefficient β_r will be small ($|\beta_r| << 1$).

In the model, the teacher attitude to repetition affects grade repetition through t_k. However, we have to keep in mind that t_k may be determined by the group average learning achievement[7] ($\overline{S_k}$), and by teacher attitude to repetition (ν_k):

$$t_k = \lambda \overline{S_k} + \nu_k \tag{3}$$

Proxies for t_k and ν_k are used here as instruments for grade repetition. Teacher attitude to repetition depends on teachers, not pupils. Accordingly these instruments control for the main potential source of endogeneity: the correlation between the children's unobservables and grade repetition decisions. Of course, teacher attitude to repetition is not observable and proxies are required. This paper uses two different strategies to proxy for it, both using peer repetition.

Repetition rate in the group The first proxy is repetition rate in the group, written as:

$$\widetilde{R_{ik}} = \frac{1}{n_k - 1} \Sigma_{j \neq i} R_{jk} \tag{4}$$

Rearranging (2) and (3) gives :

$$R_{ik} = \mathbb{1}\left[S_{ik} - \lambda \overline{S_k} - \nu_k + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{5}$$

[7]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.

In equation (5), grade repetition probability depends on $S_{ik} - \lambda\overline{S_k}$. If $\lambda = 1$, the probability of a child repeating his current grade depends on the difference between his test scores and the group's average test score. Grade repetition is relative, then: for a given ν_k, children do not repeat grades because their learning achievements are low but because their learning achievements are lower than those of their peers. We will see farther that λ is probably close to 1.

If $\lambda = 1$, the repetition rate in the group does not depend on the group average learning achievement but on teacher attitude to repetition and the distribution of learning achievement in the group. The repetition rate in the group depends on the variance of test-scores. In this case, the group repetition rate $\widetilde{R_{ik}}$ proxies for ν_k. If $\lambda \neq 1$, it is necessary to control for $\overline{S_k}$ for the group repetition rate to appropriately proxy for ν_k. Replacing $\widetilde{R_{ik}}$ as a proxy for ν_k gives:

$$R_{ik} = \mathbb{1}\left[S_{ik} - \lambda\overline{S_k} + \alpha\widetilde{R_{ik}} + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{6}$$

Last passer's test score This paper uses a second proxy for teacher attitude to repetition. "Passers" are those peers of a given pupil in a given year who are admitted to the next grade. Among the passers, the pupil with the lowest test score is called the last passer. His test score, LP_{ik}, is used as a proxy for t_k:

$$LP_{ik} = min_{\{j \neq i | R_{jk}=0\}}(S_{jk}) \tag{7}$$

$$R_{ik} = \mathbb{1}\left[S_{ik} - \lambda LP_{ik} + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{8}$$

In the benchmark specification of this paper presented in section 4, (1) is estimated jointly with (8). Because of the potential relationship between t_k and $\overline{S_k}$ in (3), $\overline{S_k}$ is controlled for. 2 instruments, LP_{ik} and the non-linear function $\mathbb{1}(S_{ik} > LP_{ik})$ (see section 3.2 below), are used to control for the potential endogeneity of grade repetition.

3.2 Estimating the grade repetition model

The previous section sets out the theoretical framework for identifying the proxies for teacher attitude to repetition used here. Before discussing the exogeneity of these proxies, the model of the determinants of grade repetition is estimated. Two crucial predictions of the model are tested:

- Grade repetition probability depends on learning achievement. The learning achievement required to pass depends strongly on peer learning achievement.
- Peer repetitions provide relevant information for predicting a given pupil's repetitions.

Table 2 shows various specifications of a probit model estimating the determinants of grade repetition. The data are pooled for the various 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 χ^2 statistics for their joint significance are reported.

Grade repetition is inferred from each child's school career. Accordingly, even conditionally on enrolment at the end of the school year (and attendance the day of the test), grade repetition suffers potentially from selection bias. Children must be re-observed to infer whether they repeated grades. However, it shall be seen that correcting from this bias barely affects the coefficients.

Column 1 of Table 2 regresses the grade repetition decision on the observables of equation (5) and on household characteristics. The test scores and group mean test scores strongly affect the grade repetition probability. The coefficients have opposite signs as predicted by equation (5). Absolute

Table 2: Estimation of the determinants of grade repetition

	equation (5)	equation (6)	equation (8)
	(1)	(2)	(3)
Test score	-.898	-.985	-.670
	(.068)***	(.077)***	(.097)***
Group mean test score	.912	1.045	.385
	(.096)***	(.105)***	(.109)***
Household wealth	-.034	-.026	-.043
	(.022)	(.022)	(.022)*
Parents' education	-.025	-.020	-.014
	(.030)	(.031)	(.032)
Repetition rate in group		1.941	
		(.229)***	
Last passer's test score			.297
			(.080)***
Test score higher than last passer's score			-.309
			(.131)**
Test score higher than first repeater's score			-.446
			(.093)***
Obs.	1785	1768	1768
log-likelihood	-678.618	-629.282	-625.584
χ^2 grade year dummies	36.875	10.815	13.476
corresponding p value	$< 10^{-5}$	.029	.009
χ^2 teacher 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.

values of both coefficients are similar, and the difference between them is not significantly different from 0. In the model, λ is the ratio between the two absolute values, so $\lambda = 1$ cannot be rejected. Conditionally on pupils' present learning achievement, household education levels and wealth do not seem to be correlated with their probability of repetition.

Columns 2 and 3 include the proxies for grade repetition of equations (6) and (8). In column 2, the coefficient of grade repetition rate among peers is positive and undoubtedly different from 0. 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. Its coefficient has the expected sign (positive) and is significant. The coefficient for $\overline{S_k}$ decreases and is half the absolute value of the coefficient for S_{ik}, but is still significant. This is coherent with the model: once controlled for t_k, $\overline{S_k}$ is not a determinant of grade repetition.

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 latent variable for R_{ik} is suspected to be non-linear, and in particular to be discontinuous when $S_{ik} = t_k$. $\mathbb{1}(S_{ik} > LP_{ik})$ proxies for this discontinuity, and in this estimation the coefficient for this discontinuity is significant. It has the expected sign: the probability of grade repetition is lower when $S_{ik} > LP_{ik}$.

"Repeaters" are a child's peers who are not admitted to the next grade. The repeater with the highest test score is the "first repeater". The specifications here do not include the first repeater's test score as a proxy for t_k: repeaters are not observed in each group. However, a dummy taking value 1 if the test score is higher than the first repeater's score is included. Where no repeater is observed in the group, the dummy takes value 1. The coefficient for this dummy is negative and significant.

Equation (8) (column 3) is the first stage in the benchmark of this paper because the discontinuity of the probability of grade repetition when $S_{ik} = t_k$ can be identified in this equation. However, equation (6) (column 2) is required as a first stage in one of the specification checks (see section 5).

So far, the model has proved consistent with the determinants of grade repetition. For this grade repetition model to be a relevant first stage in estimating equation (1), it has to be ascertained that the impact of peer repetition on a child's school career is attributable exclusively to teacher attitude to repetition, and that such attitudes are exogenous. The next sections address these two issues.

3.3 Are these reliable proxies for teacher attitude to repetition ?

This section examines whether the connexion between a child's repetition and the proxies used here can be ascribed exclusively to teacher attitude to repetition. The proxies are based on peer repetition and peer test scores. If either or both of these are related to the child's unobservables, then the proxies may be endogenous.

For each child i in group k, one of his peers is denoted j. Both proxies can be written in the form $\Phi(\{R_{jk}, S_{jk}\}_{j \neq i})$, meaning that they are a function of peer repetitions and of peer test scores. The potential correlation between $\Phi(\{R_{jk}, S_{jk}\}_{j \neq i})$ and ϵ_{ik} needs to be eliminated once the observables have been controlled for. Child i's test score and his group average test score are controlled for, and the assumption that this controls for the correlation between child i's unobservables and his peer's (child j's) test score S_{jk} is made. For that reason, let us focus on the potential correlation between child i's unobservables and his peers' repetition R_{jk}, and rewrite the determinants of repetition of peer j of equation (5):

$$R_{jk} = \mathbb{1}\left[S_{jk} - \lambda \overline{S_k} - \nu_k + X_{jk}\beta + \epsilon_{jk} < 0\right]$$

In equation (5), it is still assumed the correlation between child i's unobservables and the test score S_{jk} is controlled for. However, the unobservables ϵ_{jk} could be correlated with child i's unobsevables:

ϵ_{jk} and ϵ_{ik} are correlated if different observations of ϵ in the same group are correlated, conditionally on observable variables. In the model, correlation between the unobservables of children from the same group could cause an endogenous measurement error of teacher attitude to repetition.

Such correlation is theoretically plausible. For example, lack of motivation at school might cause grade repetition. Lack of motivation at school causes dropout. If motivation differs among schools and if motivation causes grade repetition then ϵ_{jk} is probably correlated with ϵ_{ik}, in which case ϵ_{jk} is correlated with child i dropout. Hence $\widetilde{R_{ik}}$ would be correlated with u_{ik}, the error term in the enrollment equation (equation (1)), and there would be endogeneity. This paper gives two empirical arguments for rejecting this spurious correlation between child i's unobservables and peer j's unobservables.

The first argument is that any unobservable having different distributions between groups can be expected to be correlated with the school's observables. For example, motivation is expected to be higher in wealthy communities or in communities where average education is high. If lack of motivation causes grade repetition and if motivation is higher in wealthy communities, then the repetition rate will be lower in wealthy communities.

Table 3 regresses repetition rate and last passer's test score on community-level characteristics. Proxies for teacher attitude to repetition are not correlated with any community-level variables.[8]

Columns 1, 2 and 3 run OLS regressions of the last passer's test scores of each group on certain characteristics of 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 $\lambda = 1$.[9] A high standard deviation in the group is associated with a lower last passer's score. Suppose that the 15% of children with the lowest test scores in each group repeat their grades. Then 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's 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. 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. Some of these variables, chosen to decrease multicollinearity, are maintained in column 3. Again, nothing indicates that the last passer's score is correlated with the remaining community variables.

Table 3 columns 4, 5 and 6 runs OLS regressions of group repetition rate on the same covariates. The repetition rate is negatively correlated with the group mean test score and with the previous year's group mean test score. Even if Table 2 shows that λ is probably close to 1, the repetition rate is lower when the group learning achievement is higher. This indicates that λ is probably slightly less than 1. The standard deviation of test scores is not correlated with group repetition rate.

Columns 5 and 6 provide no evidence that the group repetition rate is correlated with community characteristics. However, the F-tests in Table 3 might not be powerful enough to detect a correlation between the community's observables and repetition rates. For example, the estimated differences of 2.2% in repetition rate between agricultural and commercial communities in Table 3, column 6 are not significantly different from 0. But it might be that the estimation is not powerful enough.

A second empirical argument rejects the spurious correlation between child i's unobservables and those of his peers. In fact, I expect that the correlation between the ϵ of different peers could be caused by endogenous placement in schools. Chamberlain (1980) explains how it is possible to control for

[8] An additionnal regression indicates grade repetition rate is not significantly correlated with household wealth or parents' education, once school fixed effects, grade-year fixed effects and test scores are controlled for. This rules out an endogenous placement of pupils correlated with teacher attitude to repetition within schools.

[9] LP_{ik} is a proxy for t_k. If $\lambda = 1$, then $t_k = \overline{S_k} + \nu_k$

Table 3: Determinants of teacher specific attitude to repetition

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

fixed effects in a probit regression. This method is adopted in section 5.1 and the results are similar to the benchmark.

3.4 Is teacher attitude to repetition exogenous?

There are two reasons why ν_k could be correlated with the error term. First, teacher placement could be endogenous. Second, teacher attitude to repetition may be random, but correlated with another characteristic causing dropout.

If teacher placement is endogenous and reasons for their placement (teacher qualification, experience...) are correlated with ν_k, then ν_k may be correlated with the unobservables causing dropout u_{ik}. For example, if older teachers are appointed to urban schools and ν_k is correlated with teacher age, then ν_k may be correlated with parental preferences for schooling. Hence ν_k would be endogenous and the proxies LP_{ik} and $\widetilde{R_{ik}}$ would not control for the endogeneity of R_{ik}. However, higher repetition rates would be expected in urban schools, which is not the case in Table 3.

Schools' observable characteristics are likely to be the main determinants of teacher placement.[10] As a result, if the repetition rate is not correlated with these, which is the case in Table 3, endogenous placement may not generate the endogeneity of ν_k.

Once controlled for grades, the characteristics of the schools are probably what determines of teacher placement. So modifying the identification hypotheses to control for school fixed effects à la Chamberlain (1980) in section 5.1 controls for this potential endogeneity bias.

If teacher attitude to repetition is correlated with some other characteristic causing dropout, then the proxies fail to control for the endogeneity of grade repetition. In section 5.2 the non-linearity of the function f is used to assess whether this effect is plausible. The theoretical model takes into account a potential effect of ν_k on $E_{ik,t+1}$:

$$\left\{ \begin{array}{l} E_{ik,t+1} = \mathbb{1}\left[\beta_{e1} S_{ik} + \beta_{e2} t_k + X_{ik}\beta_{e3} + \gamma R_{ik} + u_{ik} > 0\right] \\ R_{ik} = \mathbb{1}\left[S_{ik} - t_k + \delta \mathbb{1}(S_{ik} > t_k) + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \end{array} \right. \tag{9}$$

The identification in equation (9) relies on the probability of grade repetition being discontinuous when $S_{ik} = t_k$. However, t_k is measured noisily, so that it is pointless estimating the effect of grade repetition on dropouts with a regression discontinuity design.

In section 5.2 the estimation of this model reveals a correlation between teacher attitude to repetition and the average dropout rate. However, the estimated effect of grade repetition on school dropout remains negative and significant. Depending on the specification the marginal effect may be greater and very imprecisely estimated or close to the benchmark and significantly different from 0.

3.5 Selection on grade repetition

As stated in section 2, not all grade repetition decisions can be observed. If a child dropped out before a test there is no way of knowing what the repetition decision was the year before this test. The structure of the data is summarized in Table 4.

The selection problem makes it difficult to identify the effect of grade repetition decisions on school dropout. If grade repetition causes dropout, then it causes its own selection. However, it is possible to control for the selection and hence to identify the determinants of grade repetition in model (10):

$$\left\{ \begin{array}{rcllllll} R_{ik} & = & \mathbb{1}[S_{ik} & -\lambda LP_{ik} & +X_{ik}\beta_r & & +\epsilon_{ik} < 0] \\ selection & = & \mathbb{1}[\beta_{s1} S_{ik} & +\beta_{s2} Z_s & +X_{ik}\beta_{s3} & +\gamma_s R_{ik} & +\upsilon_{ik} > 0] \end{array} \right. \tag{10}$$

[10] Teacher placement is centralized in Senegal, so observable information is probably the most important determinant of teacher placement.

Table 4: Observation of grade repetition decision

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

Appendix C.1 shows that model (10) can in theory be semiparametrically identified. This is based on a simple intuition: there is an instrument for grade repetition and an instrument for selection. In this case the system of all the probability function derivatives has a single solution. γ_s is not identified by this system, since R_{ik} is binary. However, Vytlacil and Yildiz (2006) show the coefficient for the endogenous variable is identified.

Column 1 in Table 5 reports the determinants of grade repetition with a probit specification, with no control for selection. In columns 2 and 3, model (10) is estimated using a maximum likelihood method. Accordingly, this estimation controls for the selection on R_{ik}. The error terms are assumed to follow a bivariate normal distribution. The data are pooled for the various grades and years. The standard errors of the estimators are corrected for the correlation of residuals between different observations of the same child. Each specification includes grade-year dummies in each equation. The χ^2 statistics for their joint significance is reported.

The determinants of grade repetition Column 1 in Table 5 lists the determinants of grade repetition. The coefficients in columns 1 and 2 are similar, which means they are not really affected by the correction for selection. Most of the coefficients are similar to the coefficients of Table 2 and require no further comment here. A coefficient is slightly affected by the correction for selection: the coefficient of the dummy variable for a test score higher than the first repeater's score diminishes by a third in absolute value. A high previous year's test score is associated with a lower likelihood of grade repetition, and the coefficient for the test score is closer to 0 than in Table 2. If test score is a noisy proxy for current learning achievement, then previous year's test score is expected to be another proxy for current learning achievement.[11]

The determinants of selection The estimation of *selection* in model (10) is intended to control for selection bias in the estimation of R_{ik}. The determinants of *selection* may be the determinants of moving or missing school the day of the tests in addition to the determinants of dropout. Accordingly there is no particular interpretation of these coefficients.

Nevertheless, it is necessary to focus on the effect of the negative shocks on harvests, since this variable is the exclusion restriction in the equation for R_{ik}. These shocks are not expected to be a determinant of grade repetition because the rainfall season in Senegal is from July to September, during the school vacations. Accordingly, repetition is known when the rainfall season begins. Theoretically, then, it can be ruled out that teachers might use this information for grade repetitions.

These shocks positively affect selection: when there is a negative shock, the child is more likely to take the test the next year. Negative shocks on harvests may decrease opportunity costs, so children may be more likely to take the tests when there is a shock. The F-test for the significance of this instrument is 7.5.

It has been seen that the control for selection barely affects the coefficients of grade repetition determinants. Accordingly selection bias is not controlled for in the benchmark specification. However,

[11] For that reason, previous year's test score is omitted from Table 2. Table 2 tests for $\lambda = 1$.

Table 5: Joint estimation of the determinants of grade repetition and selection (model (10))

	probit model	joint estimation	
	repetition	*repetition*	*selection*
Test score	-.452	-.489	.143
	(.121)***	(.109)***	(.110)
Class mean test score	.306	.262	.010
	(.124)**	(.122)**	(.147)
Previous year's test score	-.312	-.293	-.038
	(.066)***	(.061)***	(.078)
Household's wealth	-.019	-.029	.057
	(.024)	(.023)	(.026)**
Parental mean education	-.023	.002	-.082
	(.036)	(.034)	(.034)**
Last passer's test score	.339	.340	
	(.098)***	(.092)***	
The test score is higher than the last passer's score	-.406	-.454	
	(.148)***	(.138)***	
The test score is higher than the first repeater's score	-.479	-.348	
	(.107)***	(.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	
χ^2 grade year dummies	4.881	5.661	11.554
corresponding p value	.300	.226	.021
χ^2 instruments	97.610	89.962	7.455
corresponding p value	5.07e-21	2.23e-19	.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.

evidence will be shown that controlling for the selection would not alter the effect of grade repetition on dropouts.

4 The effect of grade repetition on school dropout

Section 3, investigated the instrumental strategy. This section uses that strategy to identify the effect of grade repetition on school dropout.

$$\left\{\begin{array}{rcllll} E_{ik,t+1} & = & \mathbb{1}[\beta_{e1}S_{ik} & & +X_{ik}\beta_{e3} & +\gamma R_{ik} & +u_{ik}>0] \\ R_{ik} & = & \mathbb{1}[S_{ik} & -\lambda_2 LP_{ik} & +X_{ik}\beta_r & & +\epsilon_{ik}<0] \end{array}\right. \quad (11)$$

Model (11) addresses the endogeneity problem. From the results of the previous section the selection issue can be ignored in the benchmark empirical specification. In addition, appendix C.2 proves that the sign of the effect of grade repetition on school dropout can be semiparametrically identified without taking into account the selection on R_{ik}.

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

Equation (11) is estimated by the maximum likelihood method. If the information about repetition is missing, the likelihood is $\mathbb{P}(R_{ik}=1, E_{ik,t+1}|S_{ik}, \overline{S_k}, LP_{ik}, X_{ik}; \beta, \delta, \gamma, \lambda)+$
$\mathbb{P}(R_{ik}=0, E_{ik,t+1}|S_{ik}, \overline{S_k}, LP_{ik}, X_{ik}; \beta, \delta, \gamma, \lambda)$.

Model (11) is estimated in Table 6. The two columns of Table 6 correspond to the model's two equations. Again the data are pooled for the various grades and years. Each specification includes grade-year dummies in each equation and the χ^2 statistics for their joint significance is reported. Column 1 is identical to column 1 in Table 5, and needs no further comments.

The effect of grade repetition on dropout In this specification of model (11), the estimated effect of grade repetition on school dropout is negative and significant. 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: grade repetition apparently increases the probability of dropout approximately threefold.

Suppose that repeaters are 12% of the pupils, that the dropout rate of repeaters is 7% (5% due to grade repetition, 2% due to other reasons), and that the dropout rate of other pupils is 2%. Then repeaters make up 30% of dropouts and grade repetition accounts for 21% of all dropouts.

This back-of-the-envelope calculation suggests grade repetition is an important determinant of dropout. Although dropout is obviously caused by other factors, it can be estimated that the proportion (partly) due to grade repetition is not negligible.

The IV coefficient for the effect of grade repetition on dropout cannot be compared with the coefficient for a simple probit model. In fact, there is no information on grade repetition for the pupils who drop out, so that model (1) cannot be estimated using probit regression.

Other determinants of school dropout Unsurprisingly, household wealth is positively associated with continuing schooling. Test scores or parental education are not correlated with dropout.

Table 6: Joint estimation of the determinants of grade repetition and school dropout (model (11))

	repetition		$enrolled_{t+1}$
	(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 wealth	-.031 (.024)		.150 (.055)***
Parents' education	-.028 (.035)		.060 (.074)
Last passer's test score	.326 (.103)***		
Test score higher than last passer's score	-.401 (.148)***		
Test score higher than first repeater's score	-.473 (.104)***		
Grade repetition			-1.126 (.330)***
(Average marginal effect of grade repetition)			*-.053* *(.026)***
Obs.		1818	
log-likelihood		-677.162	
χ^2 grade year dummies	7.765		14.920
corresponding p value	.101		.005
χ^2 instruments	97.100		
corresponding p value	6.53e-21		

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.

Controlling for selection Model (12) below adresses both the selection problem and the endogeneity of grade repetition:

$$\left\{\begin{array}{rclllll} E_{ik,t+1} & = & \mathbb{1}[\beta_{e1}S_{ik} & +\beta_{e2}Z_s & +X_{ik}\beta_{e3} & +\gamma R_{ik} & +u_{ik}>0] \\ R_{ik} & = & \mathbb{1}[S_{ik} & -\lambda LP_{ik} & +X_{ik}\beta_r & & +\epsilon_{ik}<0] \\ selection & = & \mathbb{1}[\beta_{s1}S_{ik} & +\beta_{s2}Z_s & +X_{ik}\beta_{s3} & +\gamma_s R_{ik} & +v_{ik}>0] \end{array}\right. \tag{12}$$

Appendix C.1 proves that in model (12):

- 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 certain technical assumptions[12]

all the coefficients of model (12) are identified without any parametric assumption about the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$.

The 3-equation model (12) is estimated in Table B.11. This is not the benchmark specification for convergence reasons. However, it is reassuring that the results of Tables 6 and B.11 are very similar: the effect of grade repetition on school dropout is quantitatively similar (−4.9%) and significant.

5 Specification checks

5.1 Does the effect of grade repetition on dropout persist when school fixed-effects are taken into account?

Crucially here, this paper assumes grade repetition rate is independent of the unobservables of the community. If not, pupils' unobservables may be correlated within each school, or endogenous teacher placement may generate endogeneity of teacher attitude to repetition. Table 3 shows the proxies for teacher attitude to repetition are not correlated with school observables. This is reassuring because most community unobservables are expected to be correlated with observable characteristics, but the power of this test is questionable.

Accordingly, the identification hypotheses can be modified in order to control for school fixed effects à la Chamberlain (1980). The model is then identified on the differences in teacher attitude to repetition within schools. Let us rewrite equations (6) and (1) and control for $\overline{\widetilde{R_{iks}}}$, the school average repetition rate:

$$\left\{\begin{array}{l} E_{ik,t+1} = \mathbb{1}\left[\beta_{e1}S_{ik} + \beta_{e2}\overline{S_k} + \beta_{e3}\overline{\widetilde{R_{iks}}} + X_{ik}\beta_{e4} + \gamma R_{ik} + u_{ik} > 0\right] \\ R_{ik} = \mathbb{1}\left[S_{ik} - \lambda\overline{S_k} + \gamma\overline{\widetilde{R_{iks}}} + \alpha\widetilde{R_{ik}} + X_{ik}\beta + \epsilon_{ik} < 0\right] \end{array}\right. \tag{13}$$

The 2 equations of (13) are estimated to control for $\overline{\widetilde{R_{iks}}}$. For the sake of consistency group repetition rate is used as a proxy for teacher attitude to repetition. The effect of the grade repetition decision is identified on the differences within schools in teacher attitude to repetition. This controls for potential correlation between ϵ values from the same school.

Model (13) is estimated in Table 7. The coefficients of the new covariates, of the instruments and of the effect of grade repetition are reported in this table. The other covariates in the estimation

[12] Hypotheses about points where the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$ should be positive and finite, and about the support of the distribution of the observables.

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

	repetition		$enrolled_{t+1}$
	(1)		(2)
School mean of grade repetition rates among peers	1.456		.701
	(.487)***		(.751)
Repetition rate in group	1.854		
	(.302)***		
Grade repetition			-.908
			(.331)***
(Average marginal effect of grade repetition)			*-.045*
			*(.023)***
Obs.		1823	
log-likelihood		-675.133	
χ^2 grade year dummies	8.752		16.639
corresponding p value	.068		.002
χ^2 instruments	37.819		
corresponding p value	7.76e-10		

Additional covariates in each equation: test score, group mean test score previous year's test score, household wealth, parents' education, grade-year dummies.

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

are the same as in Table 6. In this specification, the effect of grade repetition is identified by the differences in teacher attitude to repetition within schools. In schools with higher repetition rates, the probability of repetition is higher even conditionally on the group repetition rate.

The effect of grade repetition on school dropout is still negative and significant. Its marginal effect is nearly the same (here -4.5%). Table B.12 in Appendix B shows the corresponding specification of model (12), and gives the same results (the average marginal effect is -4.9%).

5.2 Does teacher attitude to repetition affect non-repeaters?

Let us suppose teacher attitude to repetition is random. Teacher placement is independent of attitude to repetition. Attitude may still be correlated with other educational methods, in which case school dropout may be spuriously correlated with grade repetition. The correlation between grade repetition and teacher attitude to repetition can be controlled for using the discontinuity of $P(R_{ik} = 1)$ when $S_{ik} = t_k$. In this case, the instrument for grade repetition is the rank relative to first repeater and last passer, and the empirical model is model (14):

$$\begin{cases} E_{ik,t+1} = \mathbb{1}\left[\beta_{e1}S_{ik} + \beta_{e3a}LP_{ik} + \beta_{e3b}\widetilde{R_{ik}} + X_{ik}\beta_{e4} + \gamma R_{ik} + u_{ik} > 0\right] \\ R_{ik} = \mathbb{1}\left[S_{ik} - \lambda LP_{ik} + \alpha\widetilde{R_{ik}} + \delta\mathbb{1}(S_{ik} > LP_{ik}) + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \end{cases} \tag{14}$$

Table 8 shows the estimation of model (14), which is the empirical counterpart of model (9). This model controls for a potential correlation between teacher attitude to repetition and school dropout. This correction relies on the assumption that the coefficient of teacher attitude to repetition in the dropout equation is the same for all children. This estimation is highly parametric, since it relies strongly on the non-linearity of the effect of t_k on grade repetition.

The proxies for teacher attitude to repetition are positively correlated with the probability of being enrolled at school the next year. Two explanations can be given for this coefficient. First

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

	repetition (1)		*enrolled*$_{t+1}$ (2)
Repetition rate in group	1.292		1.628
	(.362)***		(.566)***
Last passer's test score	.154		.096
	(.116)		(.172)
Test score higher than first repeater's score	-.245		
	(.124)**		
Test score higher than last passer's score	-.397		
	(.153)***		
Grade repetition			-2.344
			(.796)***
(Average marginal effect of grade repetition)			*-.217*
			(.138)
Obs.		1818	
log-likelihood		-665.079	
χ^2 grade year dummies	4.745		14.485
corresponding p value	.314		.006
χ^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 wealth, parents' education, grade-year dummies.

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

teacher attitude to repetition may be correlated with some other educational method causing dropouts. Second, teacher attitude to repetition has other repercussions than repetitions, and those repercussions affect the school dropouts of passers. In both cases, the effect of grade repetition on school dropouts in Tables 6 and 7 is potentially biased.

However, the coefficient for grade repetition is still negative and significant in this specification. The estimated marginal effect (-22%) is much smaller but imprecisely estimated and not significantly different from 0. The marginal effect increases whereas the probit coefficient is multiplied by 2, because of the non-linearity in the probit model. Yet, the marginal effect is very imprecisely estimated.

Table B.13 in Appendix B identifies the corresponding specification of model (12). The direct effect of ν_k on dropout is smaller (in terms of probit coefficient) and only significant at the 10% level (the p-value is less than 6%). The effect of grade repetition on school dropout is again negative and significant. The marginal effect is -6.2% in Table B.13. By contrast to Table 8, the marginal effect is significantly negative and close to the other specifications.

This difference between Table 8 and Table B.13 is certainly explained by the correction for selection. The coefficient for test score higher than the first repeater's score was affected by the correction for selection in Table 2. In model (14), this instrument is crucial, since there are only two dummies as instruments in this model whereas our benchmark, model (11), included a third continuous instrument. Accordingly the correction for selection alters the effect of grade repetition on dropout here and not in the benchmark. Hence this correction for selection is probably necessary here. Overall, the results in Table B.13 are probably more reliable, and they seemingly confirm that grade repetition has a negative effect on schooling, this result being robust to the potential causal link between teacher attitude to repetition and dropout.

This section checks whether a correlation between teacher attitude to repetition and dropout is likely to bias the result. Table 8 shows such a correlation is probable. However, the existence of this correlation does not change the sign and significance of the effect of grade repetition on dropout. The magnitude of the effect of grade repetition on school dropout is strongly affected by this correction in Table 8, but not so in Table B.13, which controls for the selection.

6 Conclusion

Proxies for the differences between teacher attitude to repetition are used here as instruments for identifying the effect of grade repetition on dropout. With these instruments, a negative effect of grade repetition on school dropout is estimated.

The differences in the proxies are not correlated with observable characteristics of school geographic location, ruling out potential endogeneity from teacher placement or from the correlation between unobserved characteristics of peers. In addition, the main estimation is modified to allow for school fixed effects. With this specification the results are very similar to the benchmark. Both empirical tests indicate that the result is not biased by school unobservable characteristics.

In both specifications a causal effect of teacher attitude to repetition on school dropout is a potential source of bias. To control for this, the procedure uses the fact that the effect of teacher attitude to repetition on children's grade repetition depends on children's ranking in their class. In this third specification, the causal effect of grade repetition on school dropout is negative and slightly stronger than in the benchmark specification.

This paper focuses on the effect of grade repetition on short-term dropout but 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 dropout, evaluation of this effect raises a serious selection problem. In addition, schooling decisions and knowledge acquisition are closely interlinked, and it is doubtful any conceptually acceptable instrument can be found for this selection.

Second, grade repetition may have long-term consequences. It is possible a priori to evaluate the long term effect from the same data, but this is not addressed here.

Finally, teacher attitude to grade repetition is likely to have a direct effect on school dropout. As shown in the last specification check on this paper, conditionally on grade repetition, the probability of children dropping out is lower when teacher attitude favors grade repetition. In fact, it is credible that teacher attitude to repetition directly affects dropout: it may increase motivation for pupils willing to avoid grade repetition, or may decrease the standard deviation of test scores in the class. In both cases, grade repetition may encourage the acquisition of knowledge so it may discourage dropout. Further, it is also credible that teacher attitude to repetition is perceived by the parents as a signal for the school quality. In that case teacher favoring grade repetition face lower dropout rates because their pupils are selected and have a higher demand for schooling, not because the grade repetition of their peers has any positive impact on their acquisition of knowledge. Finally, a grade repetition may be less discouraging when grade repetition rates are low.

References

Bruns, B., Mingat, A., Rakotomalala, R., 2003. Achieving Aniversal Primary Education by 2015: A Chance for Every Child. The World Bank, Washington. D.C.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. 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'education 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.

Table A.9: Descriptive statistics for the variables of this paper

	N	mean	standard deviation	min.	max.
Grade repetition	2176	0.173	0.379	0	1
Enrolled next year	2820	0.976	0.152	0	1
Test score	2380	-0.066	0.983	-3.20	3.34
Previous year's test score	2286	0.009	1.00	-2.34	3.81
Group mean test score	2513	-0.065	0.590	-1.63	1.91
Negative shocks on harvests	2818	0.101	0.328	0	2
Repetition rate in the group	2503	0.172	0.180	0	1
Last passer's test score	2466	-0.754	0.901	-3.20	4.69
Test score higher than last passer's score	2393	0.730	0.444	0	1
Test score higher than first repeater's score	2393	0.717	0.451	0	1
Parent's education	839	1.93	1.42	1	8
Household wealth	823	-0.88	2.01	-3.12	4.38

Notes: The last school year of the panel is dropped because repetition is not observed. Once attrition is taken into account, 2825 observations for time-variant variables remain, and 921 individuals for time-constant variables

Table A.10: Grade attended during the PASEC panel for six imaginary cases

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

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

A The variables

Repetition is a dummy taking value 1 if the child repeated the grade, and 0 otherwise. Information is from the PASEC panel. In each case, I tried to infer each year whether the child passed at the end of the school year. Table A.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. He 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, whether he repeated the second or the third grade is unknown. In case 3, the child dropped out in 1996. Consequently whether he was admitted to third grade after school year 1995 - 1996 is unknown. In case 4, the child is not in the sample after 1997 - 1998, so whether he repeated during the subsequent grades remains unknown. In cases 5 and 6, grade repetitions are not ambiguous: we know the child repeated twice (case 6) or passed twice (case 5) when he was not observed.

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

Test scores are a proxy for learning achievement at the end of the current school year. In fact the PASEC panel contains school tests at the end of each academic year until the end of the survey.[13] The tests were marked by the PASEC team. Consequently, test scores could not be influenced by teachers. Table 1 reports the number of children taking each test.

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

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

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

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

Negative shocks on harvests is a dummy taking value 1 if the head of the household reports a negative shock on harvests. These shocks are taken into account if the child or his parents were still in the household visited by EBMS in 2003. Otherwise this dummy equals 0, because the child was not really affected by these shocks. (140 cases out of 1823) However, for all the specifications presented, including a dummy for those cases did not change the effect of grade repetition on school dropouts.

Repetition rate in the group is a proxy for teacher attitude to repetition. A group is defined by all the children being in the same school and the same grade in a given school year.[14] Among the peers of a given child a given year, "passers" are those admitted to the next grade. Others must repeat their grade if they do not 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 passers, the "last passer" is the passer with the lowest test score.

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

[14] 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.

Last passer's test score is another proxy for the teacher specific attitude to repetition. In fact, if the last passer's score is high, a given child is expected to repeat more frequently.

Test score higher than 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 among teachers in their attitudes to repetition, this level of learning achievement may change among teachers. That is why the test score of the last passer is used 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".

Test score higher than 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 first repeater's score" equals 1 for every child.

B Results with model(12)

In table B.11, model (12) is estimated parametrically. This is not the benchmark specification for convergence reasons. The error terms $(\epsilon_{ik}, u_{ik}, v_{ik})$ follow a trivariate normal distribution, approximated with a GHK simulator, with 25 iterations in Table B.11. The maximum likelihood does not converge with more iterations in the simulator.

When maximization fails, the coefficient vector generates $\hat{P}(selection = 1) > \hat{P}(E_{ik,t+1} = 1)$ for many observations. It would consequently be expected that for some of these observations, $selection = 1$ and $E_{ik,t+1} = 0$. The data are constrained to $selection = 0$ if $E_{ik,t+1} = 0$, and I suspect that this incoherence between the data and the predictions of the model causes the failure of the maximization process.

Table B.11: Joint estimation of the determinants of grade repetition, selection, and school dropout (model (12))

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

Note: Additional 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 B.12: Joint estimation of the determinants of grade repetition, selection, and school dropout with Chamberlain (1980) fixed effects

	repetition	*selection*	$enrolled_{t+1}$
	(1)	(2)	(3)
School mean of grade repetition rates among peers	.866	1.267	.974
	(.453)*	(.572)**	(.713)
Negative shock on harvests this calendar year or next		.465	.193
		(.181)**	(.265)
Repetition rate in group	1.646		
	(.291)***		
Grade repetition		-.978	-1.425
		(.556)*	(.599)**
(Average marginal effect of grade repetition)			*-.049*
			*(-.016)****
Obs.	1823	1823	1823
χ^2 grade year dummies	9.585	10.999	19.887
corresponding p value	.048	.027	.0005
χ^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 wealth, parents' education, grade-year dummies.

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

Table B.13: Joint estimation of the determinants of grade repetition, selection and school dropouts corresponding to the model (14)

	repetition	*selection*	$enrolled_{t+1}$
	(1)	(2)	(3)
Repetition rate in the group	.837	1.094	.950
	(.502)*	(.622)*	(.501)*
Last passer's test score	.279	-.115	.073
	(.092)***	(.160)	(.149)
Negative shock on harvests this calendar year or next		.492	.161
		(.192)**	(.268)
Rank relative to first passer and last repeater	-.551		
	(.185)***		
Grade repetition		-1.627	-1.607
		(1.670)	(.820)**
(Average marginal effect of grade repetition)			*-.062*
			*(.029)***
Obs.	1818	1818	1818
χ^2 grade year dummies	3.829	7.985	19.843
corresponding p value	.430	.092	.0005
χ^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 wealth, parents' education, grade-year dummies.
Note: ***, ** and * mean respectively that the coefficient is significantly different from 0 at the 1%, 5% and 10% level. The standard deviations of the estimators are corrected for the correlation of the residuals between different observations of the same child.

Rank relative to first passer and last repeater compares a child's test score with 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 (12)

C.1 model (12)

This section proves that model (12) can be semiparametrically identified. It also proves that model (10) can be semiparametrically identified: the equation for e is not necessary to identify either the coefficients of r or the coefficients of s.

The model (12) is :

$$\left\{\begin{array}{lcllll} r & = & \mathbb{1}(X\beta_r + \gamma_r Z_1 & & +\varepsilon_r > 0) \\ s & = & \mathbb{1}(X\beta_s + \gamma_s Z_2 & +\alpha_s r & +\varepsilon_s > 0) \\ e & = & \mathbb{1}(X\beta_e + \gamma_e Z_2 & +\alpha_e r & +\varepsilon_e > 0) \end{array}\right. \tag{15}$$

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

Let us recall r is observed if and only if $s = 1$. $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ is the distribution function of $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$. Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the distribution function. This idea is used to show that all the parameters of model (12) are identified without any parametric assumption on $f(\varepsilon_r, \varepsilon_s, \varepsilon_e)$.

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

1. The distribution of $(\varepsilon_r, \varepsilon_s, \varepsilon_e)$ is independent of (X, Z_1, Z_2).
2. $\gamma_r \neq 0$ and $\gamma_s \neq 0$
3. $\forall j \in \{r, s, e\}$, $\beta_{j1} = 1$
4. $\exists (X_0, Z_{10}, Z_{20}) \in \Theta$ verifying :
   (a) In the neighborhood of (X_0, Z_{10}, Z_{20}), $(X, Z_1, Z_2) \in \Theta$
   (b) $\begin{pmatrix} \frac{d\mathbb{P}(r=1,s=1)}{dZ_1}(X_0, Z_{10}, Z_{20}) & \frac{d\mathbb{P}(r=1,s=1)}{dZ_2}(X_0, Z_{10}, Z_{20}) \\ \frac{d\mathbb{P}(r=0,s=1)}{dZ_1}(X_0, Z_{10}, Z_{20}) & \frac{d\mathbb{P}(r=0,s=1)}{dZ_2}(X_0, Z_{10}, Z_{20}) \end{pmatrix}$ has full rank
   (c) $\forall (X, Z_1, Z_2)$ in the neighborhood of (X_0, Z_{10}, Z_{20}), $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$
5. $\exists (a = (X_a, Z_{1a}, Z_{2a}), b = (X_b, Z_{1b}, Z_{2b})) \in \Theta^2$
   (a) $\left\{\begin{array}{l} X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} \\ X_a\beta_s + \gamma_s Z_{2a} + \alpha_s = X_b\beta_s + \gamma_s Z_{2b} \\ X_a\beta_e + \gamma_e Z_{2a} + \alpha_e = X_b\beta_e + \gamma_e Z_{2b} \end{array}\right.$
   (b) In the neighborhood of a and b, $(X, Z_1, Z_2) \in \Theta$ and $0 < f(-X\beta_r - \gamma_r Z_1, -X\beta_s - \gamma_s Z_2, -X\beta_e - \gamma_e Z_2) < \infty$

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

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

In model (12), only the signs of the latent variables ($X\beta_r + \gamma_r Z_1 + \varepsilon_r$, $X\beta_s + \gamma_s Z_2 + \alpha_s r + \varepsilon_s$ and $X\beta_e + \gamma_e Z_2 + \alpha_e r + \varepsilon_e$) are observed. Accordingly, the parameters are identified up to the scale of the parameter vector. Assumption 3 easily fixes that scale.

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

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

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

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

This proof has three steps: first, it is shown that the coefficients β and γ of the first two equations of model (12) are identified, second, it is shown that the coefficients β and γ of the last equation are identified, and finally, it is shown that the α are identified.

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

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

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

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

$$
\begin{aligned}
\frac{dP^{(11)}}{dX_1} &= F_1'^{(11)} + F_2'^{(11)} && (16) \\
\frac{dP^{(11)}}{dX_i} &= \beta_{ri} F_1'^{(11)} + \beta_{si} F_2'^{(11)} \ (\forall i \in \{1..K\}) && (17) \\
\frac{dP^{(11)}}{dZ_1} &= \gamma_r F_1'^{(11)} && (18) \\
\frac{dP^{(11)}}{dZ_2} &= \gamma_s F_2'^{(11)} && (19)
\end{aligned}
$$

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

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

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

$$\frac{dP^{(01)}}{dX_1} = F_1^{\prime(01)} + F_2^{\prime(01)} \tag{20}$$
$$\frac{dP^{(01)}}{dX_i} = \beta_{ri}F_1^{\prime(01)} + \beta_{si}F_2^{\prime(01)} \tag{21}$$
$$\frac{dP^{(01)}}{dZ_1} = \gamma_r F_1^{\prime(01)} \tag{22}$$
$$\frac{dP^{(01)}}{dZ_2} = \gamma_s F_2^{\prime(01)} \tag{23}$$

From equation (16) rearranged with (18) and (19), and (20) rearranged with (22) and (23), we get the two equations system:

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

Under assumptions 4b and 2, this identifies γ_s and γ_r. We can then easily compute $F_1^{\prime(11)}$, $F_2^{\prime(11)}$, $F_1^{\prime(01)}$ and $F_2^{\prime(01)}$ with (18), (19), (22) and (23). The system:

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

identifies β_{ri} and β_{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, 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 $\mathbb{P}(e = 1|X, Z_1, Z_2)$:

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

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

$$\begin{aligned}
\frac{dP^{(1)}}{dX_1} &= F_1'^{(1)} + F_2'^{(1)} &(24)\\
\frac{dP^{(1)}}{dX_i} &= \beta_{ri}F_1'^{(1)} + \beta_{si}F_2'^{(1)} &(25)\\
\frac{dP^{(1)}}{dZ_1} &= \gamma_r F_1'^{(1)} &(26)\\
\frac{dP^{(1)}}{dZ_2} &= \gamma_e F_2'^{(1)} &(27)
\end{aligned}$$

γ_r is known, so that $F_1'^{(1)}$ can be easily computed with (26). It is then possible to compute $F_2'^{(1)}$ with (24). Under assumption 4c, $F_2'^{(1)}$ is not null in $(X, Z_1, Z_2) \in \Theta$. That is why γ_e is identified by (27). Knowledge of β_{ri}, $F_1'^{(1)}$ and $F_2'^{(1)}$ identifies β_{si} in (25).

- **Identification of α_s.**

  Adapting Vytlacil and Yildiz (2006), it is easy to show that:

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

$$\left\{\begin{array}{l}
X_a\beta_r + \gamma_r Z_{1a} = X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1}\\
X_c\beta_r + \gamma_r Z_{1c} = X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2}\\
X_a\beta_s + \gamma_s Z_{2c} = X_c\beta_s + \gamma_s Z_{2c} = \kappa_{s1}\\
X_b\beta_s + \gamma_s Z_{2b} = X_d\beta_s + \gamma_s Z_{2d} = \kappa_{s2}
\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}
\mathbb{P}(r|a) = \mathbb{P}(r|b)\\
\mathbb{P}(r|c) = \mathbb{P}(r|d)\\
\hat{\mathbb{P}}(s|a) = \hat{\mathbb{P}}(s|c)\\
\hat{\mathbb{P}}(s|b) = \hat{\mathbb{P}}(s|d)
\end{array}\right. \quad (28)$$

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

  Then

$$\left(\begin{array}{rl}
 & \mathbb{P}(r=1, s=1|a) - \mathbb{P}(r=1, s=1|c)\\
= - & [\mathbb{P}(r=0, s=1|b) - \mathbb{P}(r=0, s=1|d)]
\end{array}\right) \Rightarrow \kappa_{s1} + \alpha_s = \kappa_{s2} \quad (29)$$

  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

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

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

(28) 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 (29):

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 (29):

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

We can now rewrite the first term of (29):

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

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

Assumption 5 ensures that some points verifying (28) and (29) exist in Θ. In fact, points a and b in assumption 5 verify (28) and the second term of (29). c can be found in the neighborhood of a and d in the neighborhood of b: the hyperplanes $\hat{\mathbb{P}}(s|(X, Z_1, Z_2) = \hat{\mathbb{P}}(s|a)$ and $\hat{\mathbb{P}}(s|(X, Z_1, Z_2) = \hat{\mathbb{P}}(s|b)$ necessarily contain pairs of points that have the same $P(r)$, since $P(r|a) = P(r|b)$.

These points can be recognized because the validity of (28) and

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

can be evaluated with the data and previous results.

- **Identification of α_e.**

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

$$\left\{\begin{array}{lcl} X_a\beta_r + \gamma_r Z_{1a} & = & X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\ X_c\beta_r + \gamma_r Z_{1c} & = & X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\ X_a\beta_s + \gamma_s Z_{1a} & = & X_c\beta_s + \gamma_s Z_{1c} = \kappa_{s1} \\ X_b\beta_s + \gamma_s Z_{1b} & = & X_d\beta_s + \gamma_s Z_{1d} = \kappa_{s2} \\ X_a\beta_e + \gamma_e Z_{2a} & = & X_c\beta_e + \gamma_e Z_{2c} = \kappa_{e1} \\ X_b\beta_e + \gamma_e Z_{2b} & = & X_d\beta_e + \gamma_e Z_{2d} = \kappa_{e2} \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} \mathbb{P}(r|a) & = & \mathbb{P}(r|b) \\ \mathbb{P}(r|c) & = & \mathbb{P}(r|d) \\ \hat{\mathbb{P}}(s|a) & = & \hat{\mathbb{P}}(s|c) \\ \hat{\mathbb{P}}(s|b) & = & \hat{\mathbb{P}}(s|d) \\ \hat{\mathbb{P}}(e|a) & = & \hat{\mathbb{P}}(e|c) \\ \hat{\mathbb{P}}(e|b) & = & \hat{\mathbb{P}}(e|d) \end{array}\right. \qquad (30)$$

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

$$\left(\begin{array}{rl} & \mathbb{P}(r = 1, s = 1, e = 1|a) - \mathbb{P}(r = 1, s = 1, e = 1|c) \\ = - & [\mathbb{P}(r = 0, s = 1, e = 1|b) - \mathbb{P}(r = 0, s = 1, e = 1|d)] \end{array}\right) \Rightarrow \kappa_{e1} + \alpha_e = \kappa_{e2} \qquad (31)$$

For the same reason as for the identification of α_s, the converse of 31 is true. In fact, if $\kappa_{e1}+\alpha_e = \kappa_{e2}$, then:

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

Proof of equation (31):

We write the probabilities:

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

Then we can easily compute the differences of (31):

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

We can now rewrite the first term of (29):

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

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

C.2 Model (12) without Z_2

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

$$\left\{\begin{array}{lcllll}
r & = & \mathbb{1}(X\beta_r+\gamma_r Z & & +\varepsilon_r > 0) \\
s & = & \mathbb{1}(X\beta_s & +\alpha_s r & +\varepsilon_s > 0) \\
e & = & \mathbb{1}(X\beta_e & +\alpha_e r & +\varepsilon_e > 0)
\end{array}\right. \tag{32}$$

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

Let us recall that r is observed if and only if $s=1$. $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$ is the distribution function of $(\varepsilon_r,\varepsilon_s,\varepsilon_e)$. Manski (1988) shows that in the one-dimensional binary model case, the parameters are identified by the derivatives of the probability function of the dependent variable. This idea is used to show that the sign of α_e is identified in model (32) without any parametric assumption on $f(\varepsilon_r,\varepsilon_s,\varepsilon_e)$. Θ is the support of (X,Z). We make the following assumptions:

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

(c) $f(\varepsilon_r, \varepsilon_s, \varepsilon_e) > 0$ in the neighborhood of $(-X_0\beta_r - \gamma_r Z_0, -X_0\beta_s - \alpha_s, -X_0\beta_s - \alpha_e)$, called Γ

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

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

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

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

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

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

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

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

Assumption 3b ensures that:

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

And assumption 3c ensures that:

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

That is why

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

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

– **Proof that the sign of α_e is identified**
Now, let us focus on $\mathbb{P}(e=1|X,Z)$:

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

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