Is grade repetition one of the causes of early school dropout? :Evidence from Senegalese primary schools.

André, Pierre

THEMA, Université de Cergy-Pontoise

December 2009

Is grade repetition one of the causes of early school dropout? Evidence from Senegalese primary schools.*

Pierre André†

December 4, 2009

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

Abstract

This chapter 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. This chapter uses two original identification strategies to identify the effect of grade repetition on school dropout. The first instrumental strategy uses the differences among teacher attitude to repetition as an instrument for grade repetition. The second strategy uses the discontinuity in the probability of grade repetition between pupils whose test score is just lower and just higher than the target achievement. Both 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, Paul Glewwe, Marco Manacorda, 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: (1) 202.258.7022. E-mail: pierre.andre01@polytechnique.org

1 Introduction

Primary education in many (principally French-speaking) Subsaharan African countries 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] whereas the African average is 15%, and the maximum is Gabon with 35% (both in 2002[2]). Manacorda (2008) remarks that this practice is more widespread in the countries where gross enrolment rates in secondary school is low, raising the question of the causality behind this correlation.

Grade repetition 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 face 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. This paper inquires whether frequent early school dropout is in part a consequence of high repetition rates. School dropout is a very prevalent phenomenon in many developing countries: some 32% of the African children enrolled in the first grade of primary school (and 40% of Senegalese pupils) do not achieve the last one.[1,2] The returns to education have been extensively studied by economists (see Duflo (2001), for the most cited study in developing countries), which emphasize the potential economic loss from early school dropout. Manski (1993b) has pushed the idea that enrollment decisions depend on the expectations on the returns to schooling, and these expectations may be different from the actual returns. Manski and Dominitz (1996) have shown that the expectations on the returns to schooling are not uniform among American high school students. Two recent controlled experiments in developping countries (Nguyen, 2008 and Jensen, 2007) have shown that further information on the returns to schooling affect the school investment decisions. All this literature emphasize the fact that endogenous school dropouts may be inefficient. In addition to its economic returns, education provides the individuals with basic capabilities, and this is the reason why the Millenium Develoment Goals include universal completion of primary school. In the end, both political commitment and economic efficiency make it necessary to fight against primary school dropout.

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. The psychologists and the pedagogical profession share a widespread view that grade repetition does not improve learning achievement. However, some 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, and McCoy and Reynolds (1999) for a more recent study). However, teachers probably use their private information on pupils to decide whether they will repeat. If low motivation at school causes grade repetition, these

[1] Ministry of Education, Senegal (2005)

[2] 2004, www.poledakar.org

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. Doing the same with a similar retention policy in Florida, Greene and Winters (2007) find a positive effect of grade repetition in third grade on reading ability after two years.

Some psychologists as Jimerson, Carlson, Rotert, Egeland, and Sourie (1997) consider that early grade repetition has a negative effect on socio-emotional adjustment. Hence, one can expect grade repetition to be discouraging and cause dropout. 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 (2008) 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 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. Manacorda (2008) uses a change in the retention policy in Uruguayan primary schools to estimate the effect of grade repetition on dropout. Grade repetition was automatic when a pupil had missed more than 25 days for school years 1996 and 1997, but not for school year 1998. This change seems to have been unanticipated by the pupils and parents, since the distribution of the number of school days missed is the broadly the same in 1998 than in 1996 - 1997. Using a diff in diff strategy, he finds that grade repetition decreases school achievement by 0.6 grades on average.

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. Glick and Sahn (2009) combines PASEC[3] data with retrospective information on school enrolment.[4] They use a control strategy to estimate the effect of grade repetition in second grade on the probability to attend to grade 5. However, control strategies may be systematically biased when measuring the effect of grade repetition on dropout. In fact, some variables like the motivation at school may cause both repetition and dropout and are hardly observable.[5] In addition, their observation of grade repetition suffers from selection, since they infer grade repetition from the school trajectory. In fact, if the grade repetition decision causes immediate dropout, actual grade repetition is censored.[6]

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

[4] This information is from EBMS survey, described in section 2

[5] Motivation at school is not observed at all in PASEC data

[6] An additional regression would show that running the same regression than Glick and Sahn (2009) with the same instrumental strategy than this paper does not lead to any interesting information on the effect of grade repetition on

This paper estimates the effect of grade repetition decision on immediate school dropout, i.e. on the probability to be enrolled at school the next school year. It controls for the potential correlation between the children's unobservable characteristics and grade repetition with two original instrumental variables strategies, and controls for the selection of the information on grade repetition decisions.

The first instrumental strategy is based on teacher attitude to repetition: grade repetition is based on the teacher's decision, and those decisions are partly based on teacher's idiosyncrasies. These indiosyncrasies are unobservable, so for each child, grade repetition by peers is used to proxy for teacher attitude.

The second instrumental strategy is based on the widespread idea that a child needs to reach a certain learning achievement to pass to the next grade. The grade repetition probability may be discontinuous between pupils whose learning achievement are just above and just below this "target achievement". This paper tries to exploit this discontinuity to identify the effect of grade repetition on school dropout.

Both results reveal a negative 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. Sections 3 and 4 presents the strategies used here for identifying this effect while 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 among primary school pupils of 98 Senegalese 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,[7] 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.

Whenever a child took a PASEC test in a given school year, the information includes his current grade. The information for grade repetition is inferred from this longitudinal information on the school careers. The pupil questionnaire also included some information about living conditions. In particular, the household wealth index used in this paper is based from the PASEC information.

dropout (coefficients are not significantly different from 0 with large standard errors).

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

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

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

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

2.2 EBMS Survey[8]

The EBMS survey provides additional information about certain PASEC pupils in 2003. It includes some of the pupils from 59 of the schools surveyed between 1995 and 2000. The objective was to resurvey households in each community (village or urban districts) with children who had been in the PASEC panel. Of the 1177 pupils attending the 59 schools surveyed by PASEC, 921 are in EBMS data after deletion of questionable matches. Information was collected about the living conditions 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). However, the school information from EBMS does not give much information on grade repetitions.[9] In addition, the EBMS data include the parent's education of the PASEC pupils and retrospective information about living conditions includes self-reported shocks on harvests.

2.3 Aggregate dataset

Both datasets provide reliable retrospective information about enrollment. Together they give enough information to reconstruct most instances of grade repetition. This information is necessary 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.

[8]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).

[9]It includes the number of grade repetitions in primary school and the number of grade repetitions in the last grade of primary school for each child living in a household surveyed by EBMS. This paper requires longitudinal information not included in the EBMS data.

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

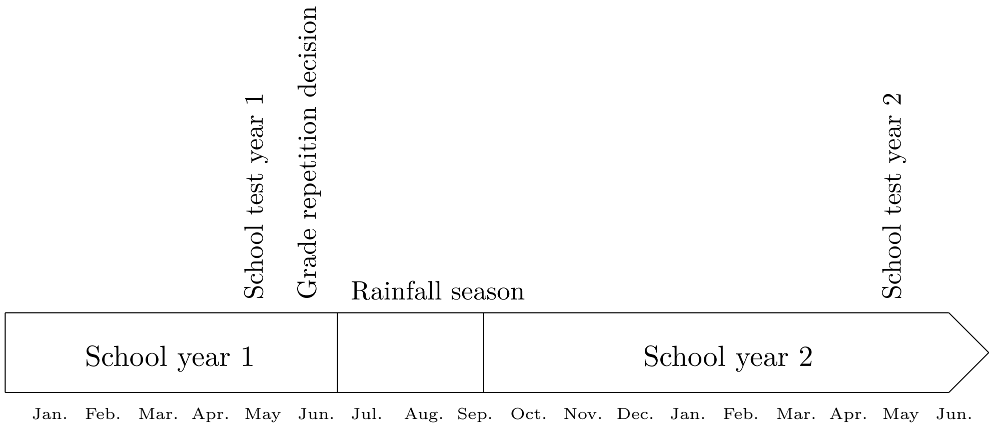

Table 2: Observation of grade repetition decision

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

2.4 Selection on grade repetition observation

Not all grade repetition decisions can be observed in the EBMS-PASEC data. The information for grade repetition is mostly inferred from this longitudinal information on the school careers. The Figure 1 summarizes the timing of the PASEC panel survey. Information on grade repetition decision at the end of school year t is known if a child took the tests in school year t and school year $t+1$.[10] Grade repetition decisions are not known for the children who dropped out immediately after this decision: if a child dropped out before the tests of school year $t+1$, there is no way of knowing what the repetition decision was at the end of school year t, as grade repetition is inferred from the school career. The structure of the data is therefore summarized in Table 2.

This selection problem makes questionable the identification of the effect of grade repetition decisions on school dropout: if grade repetition causes dropout, then it causes its own selection. However, this paper claims it is possible to control for the selection and hence to identify the determinants of grade repetition and the effect of grade repetition on school dropout in model (1):[11]

$$\begin{cases} E_{ik,t+1} = \mathbb{1}[\beta_{e1} S_{ik} + \beta_{e2} Z_s + X_{ik}\beta_{e3} + \gamma R_{ik} + u_{ik} > 0] \\ R_{ik} = \mathbb{1}[S_{ik} + \alpha Z_R + X_{ik}\beta_r + \epsilon_{ik} < 0] \\ selection = \mathbb{1}[\beta_{s1} S_{ik} + \beta_{s2} Z_s + X_{ik}\beta_{s3} + \gamma_s R_{ik} + v_{ik} > 0] \end{cases} \tag{1}$$

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

[11] The grade repetition of a child is denoted R_{ik} for child i in class k. His enrolment during the next school year is denoted $E_{ik,t+1}$ (at date $t+1$). *selection* takes value 1 if R_{ik} is known, and 0 otherwise. S_{ik} denotes the test score of child i in class k. Z_s is an instrument for the selection, which is discussed in section 3.5. Z_R is an instrument for grade repetition which is discussed through all the sections 3 and 4. The other determinants in each equation are a vector of covariates X_{ik} and unobservables u_{ik}, ϵ_{ik} and v_{ik}.

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

- If $(\epsilon_{ik}, u_{ik}, v_{ik})$ is independent of $(S_{ik}, Z_R, Z_s, X_{ik})$
- If $\lambda_2 \neq 0$ and $\beta_{s3} \neq 0$
- Under certain technical assumptions[12]

all the coefficients of model (1) are identified without any parametric assumption about the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$. 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 and γ_s are not identified by this system, since R_{ik} is binary. However, a simple adaptation of Vytlacil and Yildiz (2007) show the coefficient for the endogenous variable is identified.

Appendix C.2 even shows that under much simplier hypotheses and without Z_s, the sign of the effect of grade repetition on dropout is still identified. The intuition for that is rather simple. Indeed, the derivative of the probability of grade repetition towards Z_R gives the sign of α regardless of selection. Therefore the effect of grade repetition on enrollment is positive if the derivatives of the probability of grade repetition and of the probability of enrolment towards Z_R have the same sign, and negative if they have opposite signs.

This paper does not intend to identify model (1) semiparametrically. All the models in this paper are estimated using a maximum likelihood method. However, this result shows that there is enough information to identify the effect grade repetition on dropout in the EBMS-PASEC data without parametric assumption. Hence the results in this paper do probably not only rely on the parametric structure of the models but also on the information from the data.

3 Identification of the effect of grade repetition on dropout using teacher attitude to repetition as an instrument for grade repetition

This paper seeks to identify the effect of grade repetition, denoted R_{ik}, on school dropout (enrolment during the next school year is denoted $E_{ik,t+1}$ for child i of group k[13], at date $t+1$), which is the coefficient γ in the equation (2) below. The other determinants of dropout are test score S_{ik}, and a vector of covariates X_{ik}.[14]

$$E_{ik,t+1} = \mathbb{1}\left[\beta_{e1} S_{ik} + X_{ik}\beta_{e2} + \gamma R_{ik} + u_{ik} > 0\right] \tag{2}$$

The main difficulty in identifying γ is to control for the potential endogeneity of grade repetition. Indeed, there is endogeneity of grade repetition if the unobservable term causing dropout u_{ik} is correlated with R_{ik} conditionnal on the other covariates. The unobservable term causing dropout includes all the causes of dropout non included in the model, and all the imprecisions of the model. In particular, parental (or pupil's) preferences for schooling are included in this term. In Senegal, the grade repetition decisions are taken by the teachers. The teachers probably use their private information about their pupils to do so. In particular, they may take into account motivation at school, and motivation at school is possibly correlated with future dropouts. So grade repetition may be endogenous.

[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.

[13]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.

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

In this case, for a given learning achievement at the end of the current school year, children with a higher dropout probability are more likely to repeat their grades.

The measurement error is another concern: we do not observe the teacher's evaluation of the children in our data. Instead, we have our own evaluation based on centrally corrected tests. We denote $S^*_{ik} = S_{ik} + \mu(S_{ik})$ the test scores we observe. $\mu(S_{ik})$ is the difference between the measurement error of the test scores and the measurement error of a teacher. Hence one could expect that if a child "looks smart", $\mu(S_{ik})$ is on average negative. We can easily rewrite (2):

$$E_{ik,t+1} = \mathbb{1}\left[\beta_{e1} S^*_{ik} + X_{ik}\beta_{e2} + \gamma R_{ik} + u^*_{ik} > 0\right] \tag{3}$$

u^*_{ik} includes the measurement error of S_{ik} : $u^*_{ik} = u_{ik} - \beta_{e1}\mu(S_{ik})$.

In this paper, two original instrumental variables strategies are used to control for the potential correlation between the children's unobservable characteristics (and the measurement error) and grade repetition, and this section presents the first one.

3.1 Identification strategy

Equation (4) 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 called hereafter "target achievement".

$$R_{ik} = \mathbb{1}\left[S_{ik} - t_k + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{4}$$

However, grade repetition in the model is not determined solely by whether or not S_{ik} is greater than t_k. Equation (4) takes this into account by including other factors (X_{ik}), such as household wealth or parents' education, which may affect grade repetition. The error term ϵ_{ik} include all the causes all factors causing dropout that are not in the equation. In particular, pupil's motivation at school may affect grade repetition. In addition, it includes the imprecisions of the model, notably in the measure of learning achievement.

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

$$t_k = \lambda \overline{S_k} + \nu_k \tag{5}$$

Teacher attitude to repetition is a characteristic of the teachers, not of the pupils. Accordingly, using ν_k as an instrument for grade repetition controls for the main potential source of endogeneity: the correlation between the children's unobservables and grade repetition decisions.[15] Of course, ν_k is not observable, so it is necessary to proxy for it. This paper uses the peer's repetition rate to proxy for teacher attitude to repetition,[16] and uses this proxy as an instrument for grade repetition. The peers are the other pupils than the child himself in his group k, so that the repetition rate of the peers is:

[15] Section 3.4 assesses in details the exogeneity of ν_k.

[16] Other proxies for teacher attitude (LP_{ik} and FR_{ik}, description in section 4) were used in previous versions of this paper, giving the same results. The results are available upon request.

Repetition rate of the peers The repetition rate of the peers is written as:[17]

$$\widetilde{R_{ik}} = \frac{1}{n_k - 1}\Sigma_{j \neq i} R_{jk} \tag{6}$$

Rearranging (4) and (5) gives :

$$R_{ik} = \mathbb{1}\left[S_{ik} - \lambda\overline{S_k} - \nu_k + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{7}$$

In equation (7), 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, but not on absolute learning achievement. 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. In the Senegalese case, λ is probably close to 1.[18] However, if $\lambda \neq 1$, the group repetition rate to appropriately proxies for ν_k once controlled for $\overline{S_k}$ only. All the regressions in this paper control for $\overline{S_k}$. We use $\widetilde{R_{ik}}$ as a proxy for ν_k : $\widetilde{R_{ik}} = \nu_k + \mu(\nu_k)$. We rewrite (7)

$$R_{ik} = \mathbb{1}\left[S^*_{ik} - \lambda\overline{S_k} + \alpha\widetilde{R_{ik}} + X_{ik}\beta_r + \epsilon^*_{ik} < 0\right] \tag{8}$$

The error term in this equation is: $\epsilon^*_{ik} = \epsilon_{ik} - \mu(S_{ik}) - \alpha\mu(\nu_k)$[19]

In the first specification of this paper, (2) is estimated jointly with (8). The instrument $\widetilde{R_{ik}}$ is used to control for the potential endogeneity of grade repetition. Because of the potential relationship between $\widetilde{R_{ik}}$ and $\overline{S_k}$ in (7), $\overline{S_k}$ is controlled for in equation (2).

Controlling for school fixed effects in non-linear equations like (2) or (8) leads to non-convergent estimates. Chamberlain (1980) proposes a solution to overcome this problem: controlling for the school average of the explanatory variables. The coefficients of the explanatory variables are then identified by the difference between the individual explanatory variable and the school average, like in linear fixed effects. This specification includes a control for the school average of repetition rates among peers. This controls for the potential endogeneity of school average repetition rates: the effect of grade repetition on school dropout is identified on the difference between the repetition rates among peers and the school average of repetition rates.[20]

3.2 The results

$$\left\{\begin{array}{rcllllll} E_{ik,t+1} & = & \mathbb{1}[\beta_{e1}S^*_{ik} & +\beta_{e2}\overline{S_k} & & +X_{ik}\beta_{e3} & +\gamma R_{ik} & +u^*_{ik} > 0] \\ R_{ik} & = & \mathbb{1}[S^*_{ik} & -\lambda\overline{S_k} & +\alpha\widetilde{R_{ik}} & +X_{ik}\beta_r & & +\epsilon^*_{ik} < 0] \end{array}\right. \tag{9}$$

Model (9) addresses the endogeneity of grade repetition. It is estimated by the maximum likelihood method in Table 3. The specification is very close to the bivariate probit model, unless some observations on grade repetition are missing.[21] If the information about repetition is missing, the likelihood is $\mathbb{P}(R_{ik} = 1, E_{ik,t+1}|S_{ik}, \overline{S_k}, \widetilde{R_{ik}}, X_{ik}; \beta, \delta, \gamma, \lambda) + \mathbb{P}(R_{ik} = 0, E_{ik,t+1}|S_{ik}, \overline{S_k}, \widetilde{R_{ik}}, X_{ik}; \beta, \delta, \gamma, \lambda)$. Hence this model neglects the selection on grade repetition observation (see section 2.4). Section 3.5 nevertheless controls for this selection and shows the subsequent bias is moderate.

The two columns of Table 3 correspond to the model's two equations. 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.

[17] n_k is the number of observations per group

[18] The regressions showing it are available upon request

[19] It assumes there is no measurement error in $\overline{S_k}$. This is not expected to be costly, as the models in section 3.3

Table 3: Joint estimation of the determinants of grade repetition and school dropout (model (9))

	repetition		*enrolled*$_{t+1}$
	(1)		(2)
School mean of grade repetition rates among peers	1.456 (.487)***		.701 (.751)
Repetition rate of the peers	1.854 (.302)***		
Grade repetition			-.908 (.331)***
(Average marginal effect of grade repetition)			*-.045 (.023)***
Test score and other covariates	Yes		Yes
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.

The determinants of grade repetition Grade repetition is correlated with the peer's grade repetition. This correlation is positive, which is expected. In fact, in our model, the correlation between a child's repetition and his peer's repetition is uniquely due to teacher attitude to repetition, so it should be positive. This correlation is significant, and the χ^2 test for its significance has a value of 37. The reasons why I believe this correlation is uniquely due to teacher attitude to repetition are explained in sections 3.3 and 3.4.

The effect of grade repetition on dropout In this specification of model (9), 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 4.5 percentage points. 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 (2) cannot be estimated using probit regression.

include another source of correlation between the unobservables of different pupils from the same class.

[20] Again, the school average of repetition rates includes the schoolmates of a given pupil, and not individual repetitions.

[21] Section 3.5 shows a specification controlling for this selection

3.3 Is the peer's repetition rate a reliable proxy for teacher attitude to repetition ?

This section examines whether the connexion between a child's repetition and the peer's repetition rate can be ascribed exclusively to teacher attitude to repetition. If the peer's repetitions are related to the child's unobservables, then the proxies may be endogenous because of measurement error. In equations, $\widetilde{R_{ik}} = \nu_k + \mu(\nu_k)$ could be correlated with u_{ik}, the error term of equation predicting dropout (2) because of $\mu(\nu_k)$. So as to clarify the problem, it is useful to linearize (7):

$$\mathbb{P}(R_{ik}) = -S_{ik} + \lambda \overline{S_k} + \nu_k - X_{jk}\beta - \mathbb{E}(\epsilon_{ik}|k) \tag{10}$$

Hence we can compute the expectation of $\widetilde{R_{ik}}$:[22]

$$\begin{aligned}
\mathbb{E}(\widetilde{R_{ik}}) &= \frac{1}{n_k - 1}\Sigma_{j\neq i}\left[-S_{ik} + \lambda \overline{S_k} + \nu_k - X_{jk}\beta - \mathbb{E}(\epsilon_{ik}|k)\right] \\
&= -\widetilde{S_{ik}} + \lambda \overline{S_k} + \nu_k - \widetilde{X_{ik}}\beta - \mathbb{E}(\epsilon_{ik}|k) \\
&= \left(\lambda - \frac{n_k}{n_k - 1}\right)\overline{S_k} + \frac{1}{n_k - 1}S_{ik} + \nu_k - \widetilde{X_{ik}}\beta - \mathbb{E}(\epsilon_{ik}|k)
\end{aligned}$$

This expression includes measurement errors:

$$\mathbb{E}(\widetilde{R_{ik}}) = \left(\lambda - \frac{n_k}{n_k - 1}\right)\overline{S_k} + \frac{1}{n_k - 1}\left(S^*_{ik} - \mu(S_{ik})\right) + \nu_k - \widetilde{X_{ik}}\beta - \mathbb{E}(\epsilon_{ik}|k) \tag{11}$$

The potential correlation between $\widetilde{R_{ik}}$ and ϵ_{ik} needs to be eliminated once the observables have been controlled for. Most of the terms in equation (11) are observable, so let us focus on the unobservables.

$\mathbb{E}(\epsilon_{ik}|k)$ refers to the "correlation in unobservables", à la Manski (1993a). It means that the ϵ_{ik} could be correlated between different observations of the same group. Such correlation is theoretically plausible, either because the unobservables of children are expected to be correlated within each school or because they are expected to be correlated between different classes within a school. 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's dropout. Hence $\widetilde{R_{ik}}$ would be correlated with u_{ik}, the error term in the enrollment equation (equation (2)), and there would be endogeneity.

Concerning the endogenous placement in schools, the biprobit model in Table 3 includes a control for the school mean of grade repetition rates among peers. This term is supposed to rule out any endogenous placement of children in schools causing differences in grade repetition rates between schools, as suggested by Chamberlain (1980). The school mean of grade repetition rates in Table 3 is uncorrelated with dropout, and using this covariate does not change the effect of grade repetition on school dropout. In addition, regressions (available upon request) of repetition rate on community-level characteristics do not indicate any correlation between the community-level observable characteristics and the repetition rate. It seems therefore unplausible that endogenous placement in schools biases the effect of grade repetition on dropout in Table 3.

Concerning the endogenous placement within schools, an additionnal regression (available upon request) indicates grade repetition rate is not significantly correlated with household wealth or parents' education, once controlled for school fixed effects, grade-year fixed effects and test scores. This

[22] In this paper, $\tilde{}$ denotes the mean of a variable among the peers. So $\widetilde{S_{ik}}$ is the mean of the test scores among the peers, $\widetilde{X_{ik}}$ is the mean of the covariates.

(partly) rules out an endogenous placement of pupils correlated with teacher attitude to repetition within schools. Indeed, endogenous placement within schools may be based on unobservables that are uncorrelated with parental wealth or education.

Another concern is Manski (1993a)'s "reflexion problem". This means that if a child's grade repetition is affected by his peers, the peer's grade repetitions are affected by him. In equation (11), this term is $S^*_{ik} - \mu(S_{ik})$. It means that when a child's learning achievement is high, his peers are more likely to repeat. It is partly observed, as we observe S^*_{ik}, and partly unobserved. The crucial part is the unobservable part $\mu(S_{ik})$, as S^*_{ik} is controlled for in all the estimations of this paper. $\mu(S_{ik})$ is included in u^*_{ik}, the error term in the enrollment equation, so that the "reflexion problem" may biase the estimates in Table 3. The sign for this bias is known, as $\mu(S_{ik})$ is negative in both equations. It should biase upwards (to 0) the effect of grade repetition on enrolment in Table 3.

The importance of this bias may be overestimated by readers having in mind the debate on the peer effects in education economics (see Hoxby and Building (2000) or Angrist and Lang (2004) for examples). This literature shows that the acquisition of knowledge of a child may be affected by his peers. This is presumably not a problem here. First, all the regressions in this paper control for individual test scores and group mean test score, taking into account most of this problem. In addition, the literature usually shows that a pupil leans more when the learning achievement of his peers is high (if something). In the EBMS data, the group mean test score affects negatively grade repetition. Hence a child is more likely to repeat when the learning achievement of the peers is high. The intuition behind the reflection problem is therefore probably the following: when a child has "good" unobservable characteristics, his peers may repeat because they are compared to him. I tend to believe this effect is probably small.

3.4 Is teacher attitude to repetition exogenous?

This session assesses the exogeneity of ν_k. It is exogenous if it is uncorrelated with ϵ_{ik}, the error term of equation predicting grade repetition (2). 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}. Teacher placement is centralized in Senegal. Hence the characteristics of the schools (and the grade) are potentially crucial determinants of teacher placement. However, due to the centralization, it appears unlikely that the placement is correlated with the characteristics of a given grade in a given school. As a result, school fixed effects à la Chamberlain (1980) (and grade-year dummies) control for most of the potential endogeneity bias linked with teacher placement.

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. For example, if elder teachers favor grade repetition, and have higher pedagogic ability, then there may be a correlation between the dropout rate and teacher attitude regardless of grade repetition. Besides, teacher attitude to repetition may change the incentives in the class, so it may affect the acquisition of knowledge during the school year. Jacob (2005) studies the test-based grade repetition policy in Chicago and the simultaneous accountability policy, which made teacher and schools accountable for student achievement. Using a diff in diff strategy, he shows that the policy increased learning achievement in classes where a lot of pupils where likely to repeat ex-ante, and increases learning achievement for at-risk students. This is consistent with the fact that the grade repetition policy changes the incentives in the class, though it is impossible in this case to disentangle the effect of incentives caused by grade repetition from the effect of the incentives caused by the accountability policy. If grade repetition changes the incentives in the class and hence the learning achivement of the pupils, it may affect directly dropout.

If either or both of these two arguments are true, then the first identification strategy fails to identify the effect of grade repetition on school dropout. This is the reason why a second identification is proposed in section 4 of this paper.

3.5 Selection on grade repetition

As stated in section 2.4, not all grade repetition decisions are observed. However, section 2.4 shows that all the coefficients of model (1) can be identified without any parametric assumption about the distribution of $(\epsilon_{ik}, u_{ik}, v_{ik})$.

In Table B.9, model (1) 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 trivariate normal distribution, approximated with a GHK simulator. 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.

Table B.9 is not the main specification in this section for convergence reasons, and because the correction for selection does not change the results very much in practice. However, it is reassuring that the results of Tables 3 and B.9 are very similar: the effect of grade repetition on school dropout is quantitatively similar (-4.9 percentage points) and significant.

The determinants of selection The estimation of *selection* in model (1) 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 (see Figure 1). Accordingly, grade 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 6.6.

4 Identification of the effect of grade repetition on dropout using the discontinuity of the grade repetition probability

This section presents the second indentification strategy of this paper. This stragy is based on the widespread idea that a certain learning achievement is required to pass to the next grade. This "target achievement" is denoted t_k in equation (4). This identification strategy is based on the fact that the grade repetition probability may be discontinuous when $S_k = t_k$. In that case, equation (4) can be rewritten:

$$R_{ik} = \mathbb{1}\left[S_{ik} - t_k + \delta \mathbb{1}(S_{ik} > t_k) + X_{ik}\beta_r + \epsilon_{ik} < 0\right] \tag{12}$$

However, the "target achievement" is not observed, so this paper tries to proxy for it. It uses two proxies for t_k: the test score of the last passer and the test score of the first repeater. "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:

$$LP_{ik} = min_{\{j \neq i | R_{jk}=0\}}(S_{jk}) \tag{13}$$

Similarly, the "repeaters" are those peers of a given pupil who are not admitted to the next grade. Among the repeaters, the pupil with the highest test score is the first repeater. His test score is another proxy for t_k:

$$FR_{ik} = max_{\{j \neq i | R_{jk}=1\}}(S_{jk}) \tag{14}$$

So as to proxy for the fact that a pupil's learning achievement is higher than the "target achievement", this paper compares his test score with LP_{ik} and FR_{ik}. The "Rank relative to first repeater and last passer" is the sum of the two corresponding dummies. It is used as a proxy for $\mathbb{1}(S_{ik} > t_k)$:

$$RR_{ik} = \mathbb{1}(S^*_{ik} > LP_{ik}) + \mathbb{1}(S^*_{ik} > FR_{ik}) = \mathbb{1}(S^*_{ik} > t^*_{k1}) + \mathbb{1}(S^*_{ik} > t^*_{k2}) \tag{15}$$

The "Rank relative to first repeater and last passer" takes value 0 if the test score of a child is lower than LP_{ik} and FR_{ik}, 1 if it is higher than either of the two, and 2 if it is higher than both.[23] It is a proxy for $\mathbb{1}(S_{ik} > t_k)$, so it is used as an instrument in model (16):

$$\begin{cases} E_{ik,t+1} & = \mathbb{1}[& \beta_{e1} S^*_{ik} & +\beta_{e3a} LP_{ik} & +\beta_{e3b}\widetilde{R_{ik}} & & & +\gamma R_{ik} & +X_{ik}\beta_{e4} + u^{**}_{ik} > 0] \\ R_{ik} & = \mathbb{1}[& S^*_{ik} & -\lambda LP_{ik} & +\alpha\widetilde{R_{ik}} & +\delta RR_{ik} & & & +X_{ik}\beta_r + \epsilon^{**}_{ik} < 0] \\ selection & = \mathbb{1}[& \beta_{s1} S^*_{ik} & +\beta_{s2} LP_{ik} & +\beta_{s3}\widetilde{R_{ik}} & +\beta_{s4} Z_s & +\gamma_s R_{ik} & & +X_{ik}\beta_{s5} + v^{**}_{ik} > 0] \end{cases} \tag{16}$$

Model (16) is the model estimated in the main regression of this section. The error term in the enrollment equation writes $u^{**}_{ik} = u_{ik} - \beta_{e1}\mu(S_{ik}) - \beta_{e3b}\mu(t_k)$. Section 4.1 gives the corresponding first stage and reduced form estimates, and section 4.2 gives the results of its identification.

4.1 First stage and reduced form estimates

This section presents some semiparametric fist stage and reduced form estimates corresponding to the model (16). For simplicity, it focuses on one of the proxies for t_k: the last passer's test score. In equation (15), it focuses on $\mathbb{1}(S^*_{ik} > LP_{ik})$ and hence neglects $\mathbb{1}(S^*_{ik} > FR_{ik})$. In Table 4, columns 1 and 2 present the correlation between grade repetition probability and the difference between own test score and the test score of the last passer, $S^*_{ik} - LP_{ik}$. The first column presents the coefficient of the OLS regression of grade repetition probability on a set of dummy variables dichotomizing $S^*_{ik} - LP_{ik}$. It shows that the grade repetition is not perfectly determined by the sign of this difference ($\# \mathbb{1}(S^*_{ik} > LP_{ik})$). Indeed, grade repetition probability seems to be a continuously decreasing function of $S^*_{ik} - LP_{ik}$. The function is strongly decreasing, as the estimated grade repetition probability is 68% when $S^*_{ik} - LP_{ik} < -1$, and only 2% when $1.5 < S^*_{ik} - LP_{ik}$.

Column 2 presents the coefficients of the probit regression of grade repetition probability on the same dummy variables and covariates. The coefficient for the dummies is still decreasing. The estimates are however rather imprecise, and rarely significantly different from 0. Once controlled for the test scores (among other covariates), the grade repetition probability seems approximately contstant when $S^*_{ik} - LP_{ik}$ is negative and when $S^*_{ik} - LP_{ik}$ is positive. There is a fairly big jump in the grade repetition probability at $S^*_{ik} - LP_{ik} = 0$. Because of the imprecision of this first stage, these results

[23] When there is no repeater among the peers, $\mathbb{1}(S_{ik} > FR_{ik})$ takes arbitrarily value 1.

Table 4: Grade repetition and school dropouts as a function of a difference between own test score and the test score of the last passer ($S^*_{ik} - LP_{ik}$)

	repetition		$enrolled_{t+1}$
	OLS	Probit	Probit
	(1)	(2)	(3)
Repetition rate of the peers		1.91**	0.48
		(0.33)	(0.47)
Last passer's test score		-0.06	0.21
		(0.27)	(0.26)
Test score		-0.52+	-0.04
		(0.30)	(0.22)
$S^*_{ik} - LP_{ik} < -1$	0.68**	0.16	0.02
	(0.06)	(0.43)	(0.51)
$-1 < S^*_{ik} - LP_{ik} < -0.5$	0.65**	0.28	-0.61*
	(0.06)	(0.25)	(0.29)
$-0.5 < S^*_{ik} - LP_{ik} < 0$	0.38**	Ref.	Ref.
	(0.04)		
$0 < S^*_{ik} - LP_{ik} < 0.5$	0.19**	-0.40*	0.01
	(0.02)	(0.20)	(0.25)
$0.5 < S^*_{ik} - LP_{ik} < 1$	0.09**	-0.68*	0.05
	(0.02)	(0.31)	(0.31)
$1 < S^*_{ik} - LP_{ik} < 1.5$	0.06**	-0.52	0.85
	(0.01)	(0.44)	(0.52)
$1.5 < S^*_{ik} - LP_{ik}$	0.02**	-0.71	0.53
	(0.01)	(0.60)	(0.49)
Covariates	No	Yes	Yes
Observations	1580	1465	1678

Notes: $S^*_{ik} - LP_{ik}$ stands for “Difference between own test score and last passer's test score”.
Covariates in column 2 and 3: group mean test score, previous year's test score, household wealth, parents' education, grade-year dummies, and the constant.
***, ** 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.

have nevertheless to be interpreted with caution. This justifies that it is necessary to aggregate information as in equation (15) to build a relevant first stage in the estimation of the effect of grade repetition on dropout. It also raises the question of the estimation of the discontinuity regressions when there is measurement error on the threshold (on t_k in the model) and on the switching variable (on S_{ik} in the model). This issue is discussed in section 4.3.

In Table 4, column 3 presents the correlation between grade repetition and the difference $S^*_{ik} - LP_{ik}$. Given the precision of the first stage in column 2 (and given that the reduced form estimates in non-linear settings are not necessarily convergent to any simple function of the structural parameters), one should not expect too much precision for these estimates. There is actually no such thing as a discontinuity of grade repetition probability at $S^*_{ik} - LP_{ik} = 0$. The estimates are sometimes negative when $S^*_{ik} - LP_{ik} < 0$, sometimes positive when $S^*_{ik} - LP_{ik} > 0$, and rarely significantly different from 0.

4.2 Main results

Table 5 shows the estimation of model (16). The model is estimated with a maximum likelihood method, as a "trivariate probit" specification. The distribution of the error terms follow a trivariate normal distribution, simulated with a GHK simulator. The three columns of Table 5 correspond to the model's three equations. The data are pooled for the various grades and years. Each specification includes grade-year dummies in each equation and the χ^2 statistics for their joint significance is reported.

This model controls for a potential correlation between teacher attitude to repetition and school dropout. However, 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.

Determinants of selection Table 5 includes a correction for selection. This correction appears necessary, as the coefficient of the instrument for grade repetition in model (16) is affected by the correction for selection.[24]

In Table 5, the negative shocks on harvests are used as an exclusion restriction in the repetition equation. Like in Table B.9, this coefficient is positive and significant in the selection equation.

Determinants of grade repetition This specification includes two proxies for teacher attitude to grade repetition: the repetition rate of the peers and the test score of the last passer. The test score of the last passer is a proxy for t_k, but once controlled for $\overline{S_k}$, a proxy for t_k proxies for ν_k as well according to equation (5). The coefficient for these proxies are both positive and significant, as expected.

The coefficient for "Rank relative to first repeater and last passer" is negative and significant. This is expected: if the learning achievement is higher than the "target achievement", grade repetition is less likely. This coefficient is significant, the χ^2 test for is significance is 8.9.

The effect of grade repetition on school dropout The repetition rate of the peers is positively correlated with the probability of being enrolled at school the next year. Several explanations can be given for this coefficient. First teacher attitude to repetition may be correlated with some other educational method causing less dropouts. Second, teacher attitude to repetition has other repercussions than repetitions, and those repercussions affect the school dropouts of passers. This effect is credible: 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

[24]The specification without control for selection is available upon request.

Table 5: Joint estimation of the determinants of grade repetition, selection and school dropouts corresponding to the model (16)

	repetition	*selection*	$enrolled_{t+1}$
	(1)	(2)	(3)
Repetition rate of the peers	.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)***
Test score and other covariates	Yes	Yes	Yes
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.

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 teachers 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; and passing the grade may be a signal for high learning ability only when grade repetition rates are high.

Whatever the reason why the repetition of the peers is correlated with dropouts in Table 5, the effect of grade repetition on school dropouts in Table 3 potentially underestimates the number of dropouts caused by grade repetition. However, the coefficient for grade repetition is still negative and significant in this specification. The estimated marginal effect (6.2%) is significantly different from 0 and close to the corresponding coefficient in Table 3. Overall, the results in Table 5 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.

4.3 Sensibility to the Measurement error

RR_{ik} is a proxy for $\mathbb{1}(S_{ik} > t_k)$. One could therefore be concerned that the measurement error in RR_{ik} may biase the estimates in Table 5. Indeed, the measurement errors $\mu(S_{ik})$ and $\mu(t_k)$ are included in u^{**}_{ik}, the error term in the enrollment equation of model (16).

To make it clear, $RR_{ik} = \mathbb{1}(S^*_{ik} > t^*_{k1}) + \mathbb{1}(S^*_{ik} > t^*_{k2})$. For $j = 1, 2$, this equation introduces the measurement errors and real values of S_{ik} and t_k: $\mathbb{1}(S^*_{ik} > t^*_{kj}) = \mathbb{1}\left(S_{ik} - t_k > \mu(S_{ik}) - \mu(t_{kj})\right)$. Hence RR_{ik} is on a theoretical basis very likely to be correlated with the error term in the enrollment equation $u^{**}_{ik} = u_{ik} - \beta_{e1}\mu(S_{ik}) - \beta_{e3b}\mu(t_k)$, because both include $\mu(S_{ik})$ and $\mu(t_k)$.

This is of course undesirable, and Table 6 tries to assess to what extent the estimates in Table 5 are likely to be biased by measurement error. Indeed, this Table tries to estimate model (16) with flawed controls for S_{ik} and t_k, and assesses whether the effect of grade repetition on dropout changes. If the effect of grade repetition on school dropout is the same when the control for test score is flawed, it probably means that it is not crucial to control very carefully for the child's learning achievement. Then the endogeneity due to the measurement error in test scores is probably negligible. This seems to be actually the case in Table 6. In this Table, column 1 recalls the effect of grade repetition on enrollment from Table 5. Column 2 presents the same specification without control for the test scores in any equation. Column 3 presents a specification when test scores have been replaced by a noisier variable, which is the sum of test scores and a random noise.[25] The estimates of the effect of grade repetition on dropout are very similar in each equation, and it seems therefore unlikely that the measurement error in test scores biases seriously the effect of grade repetition on enrollment in Table 5.

Columns 4, 5 and 6 in Table 6 assess the sensibility of the effect of grade repetition on the measurement error of t_k. Table 5 and column 1 include two controls for t_k: the last passer's test score and the peer's repetition rate. Column 4 gives the results of a specification with a single control for t_k: the last passer's test score. Column 5 gives the results without any control for t_k. Column 6 gives the results when the last passer's test score has been repaced by a noisier variable which is the sum of this variable and a random noise.[26] Overall, the results seem much more sensible to the measurement in t_k than in the measurement in test scores. The probit coefficients are indeed very affected by the absence of control for t_k. However, the corresponding marginal effects are approximately the same in columns 1, 4, 5 and 6.

[25]This noise has a normal distribution with standard deviation 1. The standard deviation of the test scores is 1.

[26]This noise has a normal distribution with standard deviation 1. The standard deviation of the last passer's test score is 0.84.

Table 6: Effect of grade repetition on school enrollment corresponding to the model (16), sensibility to measurement error

	Tab. 5	No test score	Noisy test score	No repetition rate of peers	No last passer's test score	Noisy last passer's test score
	(1)	(2)	(3)	(4)	(5)	(6)
Repetition rate of the peers	.950 (.501)*	1.134 (.470)**	1.148 (.488)**			
Last passer's test score	.073 (.149)	.057 (.150)	.052 (.153)	.207 (.128)		.095$^{(1)}$ (.066)
Test score	.138 (.163)		-.018$^{(1)}$ (.073)	.189 (.155)	.193 (.151)	.232 (.153)
Grade repetition	-1.607 (.820)**	-1.837 (.597)***	-1.786 (.616)***	-1.724 (.557)***	-2.665 (1.425)*	-2.128 (1.042)**
Average marginal effect of grade repetition	*-.062* *(.029)***	*-.073* *(.033)***	*-.074* *(.036)***	*-.055* *(.021)****	*-.055* *(.17)****	*-.047* *(.13)****
Obs.	1818	1818	1818	1818	1823	1818
χ^2 first stage	8.884	99.15	71.18	22.12	78.63	60.23
corresponding p value	.003	.000	.000	.000	.000	.000

Notes: The Table only reports the coefficients of the enrollment equation.
Additional covariates in each equation of model (16): group mean test score, previous year's test score, household wealth, parents' education, grade-year dummies.
(1): A noise has been added to the variable. This noise follows a normal distribution and has a standard deviation of 1. The standard deviation of the initial variable is 1 for test scores and 0.84 for last passer's test score.
***, ** 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.

The measurement error of t_k supposedly affects the effect of grade repetition on dropout only if t_k affects dropout. Indeed, the error term in the enrollment equation is $u^{**}_{ik} = u_{ik} - \beta_{e1}\mu(S_{ik}) - \beta_{e3b}\mu(t_k)$, hence if $\beta_{e3b} = 0$ there is no bias, as $\mu(t_k)$ is not in the error term of the enrolment equation. So it is worth noting that in Table 5, the repetition rate of the peers is the only proxy for t_k affecting dropout. Dropping this variable from the sample should probably take into account most of the bias due to the measurement error in t_k. One could argue that dropping the last passer's score only increases the standard error of the estimators and that the estimates are less precise in columns 5 and 6. The correlation between last passer's score and the repetition rate of the peers is only 0.4. Hence it is difficult to argue that the last passer's score replaces the control for the repetition rate of the peers in column 4 of Table 6. It is therefore possible to believe that the estimates in Table 5 are not affected by the measurement error of t_k.

5 Conclusion

This chapter uses two different identification strategies to estimate the effect of grade repetition on school dropout. In the first one, a proxy for the differences between teachers attitude to repetition is used a an instrument for identifying the effect of grade repetition on dropout. With this instrument, a negative effect of grade repetition on school dropout is estimated. The specification includes school fixed effects, ruling out potential endogeneity from teacher or pupil placement. However, in this specification a causal effect of teacher attitude to repetition on school dropout is a potential source of bias. To control for this, the second specification uses the fact that the grade repetition probability may be discontinuous with respect to learning achievement, if a certain learning achievement is required to pass. So the second specification uses this discontinuity to identify the effect of grade repetition on dropout. In this second specification, the causal effect of grade repetition on school dropout is negative as well and quantatively similar to the benchmark specification.

This chapter 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. In fact, this evaluation suffers from empirical pitfalls due to small sample sizes: the panel dimension of these data cannot be fully exploited to analyse long-term effects of grade repetition, as the panel only lasted five years.

Finally, teacher attitude to grade repetition is likely to have a direct effect on school dropout, as shown in the last specification of this chapter. To my knowledge, this effect has never been fully identified in the literature, and this may strongly limit the conclusions drawn from the literature on grade repetition. Overall, despite endogeneity concerns in many papers, the literature on the effects of grade repetition have drawn a pretty pessimistic view on the consequences of grade repetition for repeaters. The analysis of the way grade repetition rules affect the other pupils may both mitigate these results and explain why the teaching staffs are sometimes reluctant to decrease repetition rates. This analysis raises nevertheless some identification issues, which may only be fully adressed with experimental or quasi-experimental data.

References

Angrist, J., Lang, K., 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. The American Economic Review 94, 1613–1634.

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.

Duflo, E., 2001. Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual experiment. American Economic Review 91, 795–813.

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

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

Holmes, C. T., 1989. Flunking Grades: Research and Policies on Retention, chap. Grade level retention effects: A meta-analysis of research studies. London: Falmer Press, 16–33.

Hoxby, C., Building, L., 2000. Peer effects in the classroom: Learning from gender and race variation. NBER working paper .

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

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

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

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

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

Manacorda, M., 2008. The cost of grade retention. CEPR Discussion Paper 878.

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

Manski, C., 1993a. Identification of endogenous social effects: The reflection problem. The Review of Economic Studies , 531–542.

Manski, C., Dominitz, J., 1996. Eliciting student expectations of the returns to schooling. The Journal of Human Resources , 1–26.

Manski, C. F., 1993b. Studies of supply and demand in higher education, chap. Adolescent Econometricians: How do youth infer the returns to schooling. University of Chicago press, 43–60.

McCoy, A. R., Reynolds, A. J., 1999. Grade retention and school performance: An extended investigation. Journal of School Psychology 37, 273–298.

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

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

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

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

Table A.7: 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.8: 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 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.8 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.[27] 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 during the current calendar year or the next. 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 of the peers 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.[28] The peers of a child are the other pupils of his group than himself. 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.

[27] 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.

[28] 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.

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.

B Results with model(1)

In table B.9, model (1) 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.9. 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.9: 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 of the peers	1.646		
	(.291)***		
Grade repetition		-.978	-1.425
		(.556)*	(.599)**
(Average marginal effect of grade repetition)			*-.049*
			*(-.016)****
Test score and other covariates	Yes	Yes	Yes
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.

C Proofs for the semiparametric identification of model (1)

C.1 model (1)

This section proves that model (1) can be semiparametrically identified.

The model (1) is :

$$\left\{\begin{array}{lclll} r & = & \mathbb{1}(X\beta_r + \gamma_r Z_1 & & +\varepsilon_r > 0) \\ s & = & \mathbb{1}(X\beta_s + \gamma_s Z_2 & +\alpha_s r & +\varepsilon_s > 0) \\ e & = & \mathbb{1}(X\beta_e + \gamma_e Z_2 & +\alpha_e r & +\varepsilon_e > 0) \end{array}\right. \tag{17}$$

(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 (1) 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 (1), 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 (1) are identified, second, it is shown that the coefficients β and γ of the last equation are identified, and finally, it is shown that the α are identified.

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

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

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

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

$$\frac{dP^{(11)}}{dX_1} = F_1'^{(11)} + F_2'^{(11)} \tag{18}$$

$$\frac{dP^{(11)}}{dX_i} = \beta_{ri} F_1'^{(11)} + \beta_{si} F_2'^{(11)} \ (\forall i \in \{1..K\}) \tag{19}$$

$$\frac{dP^{(11)}}{dZ_1} = \gamma_r F_1'^{(11)} \tag{20}$$

$$\frac{dP^{(11)}}{dZ_2} = \gamma_s F_2'^{(11)} \tag{21}$$

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

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

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

$$
\begin{aligned}
\frac{dP^{(01)}}{dX_1} &= F_1'^{(01)} + F_2'^{(01)} &(22)\\
\frac{dP^{(01)}}{dX_i} &= \beta_{ri}F_1'^{(01)} + \beta_{si}F_2'^{(01)} &(23)\\
\frac{dP^{(01)}}{dZ_1} &= \gamma_r F_1'^{(01)} &(24)\\
\frac{dP^{(01)}}{dZ_2} &= \gamma_s F_2'^{(01)} &(25)
\end{aligned}
$$

From equation (18) rearranged with (20) and (21), and (22) rearranged with (24) and (25), we get the two equations system:

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

Under assumptions 4b and 2, this identifies γ_s and γ_r. We can then easily compute $F_1'^{(11)}$, $F_2'^{(11)}$, $F_1'^{(01)}$ and $F_2'^{(01)}$ with (20), (21), (24) and (25). The system:

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

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

- **Identification of the third equation**

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

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

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

$$\begin{aligned}
\frac{dP^{(1)}}{dX_1} &= F_1'^{(1)} + F_2'^{(1)} && (26)\\
\frac{dP^{(1)}}{dX_i} &= \beta_{ri}F_1'^{(1)} + \beta_{si}F_2'^{(1)} && (27)\\
\frac{dP^{(1)}}{dZ_1} &= \gamma_r F_1'^{(1)} && (28)\\
\frac{dP^{(1)}}{dZ_2} &= \gamma_e F_2'^{(1)} && (29)
\end{aligned}$$

γ_r is known, so that $F_1'^{(1)}$ can be easily computed with (28). It is then possible to compute $F_2'^{(1)}$ with (26). 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 (29). Knowledge of β_{ri}, $F_1'^{(1)}$ and $F_2'^{(1)}$ identifies β_{si} in (27).

- **Identification of α_s.**

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

  If $\exists\,((X_a, Z_{1a}, Z_{2a}), (X_b, Z_{1b}, Z_{2b}), (X_c, Z_{1c}, Z_{2c}), (X_d, Z_{1d}, Z_{2d})) \in \Theta^4$ so that[29]

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

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

  Then

$$\left(\begin{array}{rl} & \mathbb{P}(r=1, s=1|a) - \mathbb{P}(r=1, s=1|c) \\ = - & [\mathbb{P}(r=0, s=1|b) - \mathbb{P}(r=0, s=1|d)] \end{array}\right) \Rightarrow \kappa_{s1} + \alpha_s = \kappa_{s2} \tag{31}$$

  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

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

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

(30) 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 (31):

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 (31):

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

We can now rewrite the first term of (31):

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

$f(\varepsilon_r,\varepsilon_s,\varepsilon_e)>0$ 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 (30) and (31) exist in Θ. In fact, points a and b in assumption 5 verify (30) and the second term of (31). 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 (30) 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

$$\left\{\begin{array}{lcl} X_a\beta_r + \gamma_r Z_{1a} & = & X_b\beta_r + \gamma_r Z_{1b} = \kappa_{r1} \\ X_c\beta_r + \gamma_r Z_{1c} & = & X_d\beta_r + \gamma_r Z_{1d} = \kappa_{r2} \\ X_a\beta_s + \gamma_s Z_{1a} & = & X_c\beta_s + \gamma_s Z_{1c} = \kappa_{s1} \\ X_b\beta_s + \gamma_s Z_{1b} & = & X_d\beta_s + \gamma_s Z_{1d} = \kappa_{s2} \\ X_a\beta_e + \gamma_e Z_{2a} & = & X_c\beta_e + \gamma_e Z_{2c} = \kappa_{e1} \\ X_b\beta_e + \gamma_e Z_{2b} & = & X_d\beta_e + \gamma_e Z_{2d} = \kappa_{e2} \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} \mathbb{P}(r|a) & = & \mathbb{P}(r|b) \\ \mathbb{P}(r|c) & = & \mathbb{P}(r|d) \\ \hat{\mathbb{P}}(s|a) & = & \hat{\mathbb{P}}(s|c) \\ \hat{\mathbb{P}}(s|b) & = & \hat{\mathbb{P}}(s|d) \\ \hat{\mathbb{P}}(e|a) & = & \hat{\mathbb{P}}(e|c) \\ \hat{\mathbb{P}}(e|b) & = & \hat{\mathbb{P}}(e|d) \end{array}\right. \quad (32)$$

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

$$\left(\begin{array}{cc} & \mathbb{P}(r=1,s=1,e=1|a) - \mathbb{P}(r=1,s=1,e=1|c) \\ = - & [\mathbb{P}(r=0,s=1,e=1|b) - \mathbb{P}(r=0,s=1,e=1|d)] \end{array}\right) \Rightarrow \kappa_{e1} + \alpha_e = \kappa_{e2} \quad (33)$$

For the same reason as for the identification of α_s, the converse of 33 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 (33):

We write the probabilities:

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

Then we can easily compute the differences of (33):

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

We can now rewrite the first term of (31):

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

$f(\varepsilon_r,\varepsilon_s,\varepsilon_e) > 0$ 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 (1) 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:

$$
\left\{
\begin{array}{lcllll}
r & = & \mathbb{1}(X\beta_r + \gamma_r Z & & +\varepsilon_r > 0) \\
s & = & \mathbb{1}(X\beta_s & +\alpha_s r & +\varepsilon_s > 0) \\
e & = & \mathbb{1}(X\beta_e & +\alpha_e r & +\varepsilon_e > 0)
\end{array}
\right.
\tag{34}
$$

(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 (34) 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.