Do children in private schools learn more than in public schools? Evidence from Mexico

Florian Wendelspiess Chávez Juárez
University of Geneva

January 6, 2010

Abstract

In this study I try to answer the question whether private schools do better in the human capital accumulation process than public schools in Mexico. The analysis is based on panel data including out-of-school cognitive skill tests, which allows dealing with some potential endogeneity problems due to the selection process into private schools. The absolute advantage of private school graduates in cognitive skills disappears once controlling for the selection bias, where no positive effect is found anymore.

Keywords: education, Mexico, private schooling JEL-Classification: I21,L33

1 Introduction

Coming from Switzerland and living in Mexico, one can be quite surprised about the sharp division of the Mexican educational system in private and public schooling. That is what happened to me some years ago and I asked myself all the time if private schools perform really better than public institutions, or if the obvious selection advantage of private school graduates is due to other factors.

The question whether private schools are better than their public competitors is quite old and widely discussed in the economic literature. Finger and Schlesser (1963) for instance analyzed this issue in the sixties by comparing some standardized test scores of both, public and private school graduates. At this time, they found that private school graduates did actually worse than public school pupils and argue that this might be due to lower scholastic aptitudes and motivation problems. Using similar tests, Horowitz and Spector (2005) find opposite results for the year 2002. They study the performance of more than 15.000 undergraduate students at Bell State University and find out that graduates from private high schools perform slightly better than graduates from public or religious schools, however, the effect is only present during the first years at college and is not persistent to the end of the college studies. Angrist et al. (2002) study a lottery-like voucher program for private schools in Colombia to estimate the differences in cognitive skills, since this natural experiment solves part of the estimation problems due to endogeneity. They find a positive effect for lottery winner, thus for private schools. Rouse (1998) analyzes a similar program in the US and finds positive effect of private schools for mathematical skills, whereas no effect for reading skills. Hanushek (2002) provides a very complete discussion of the private-public school issue and sacrifices also an important discussion on

The author is Master student at the University of Geneva in economics and former research assistant for the United Nations Development Programme in Mexico. I want to thank Cyril Pasche for his assistance with the polychoric factor analysis. All errors are mine.

the religious schools. He tends also to underline the good performance of the private sector in the educational system.

In this paper I try to investigate these questions by analyzing the progress of pupils during 3 years on a standardized cognitive test score in Mexico. The paper deals with several empirical problems, going from the proper definitions of the test score to the correct methodology when analyzing the private school advantage.

The remaining of this paper is organized as follows. In section 2 I introduce the methods used to estimate the effect of private schooling and the related econometric issues. Section 3 introduces the data and the methodology used to create the test score, section 4 presents the estimation strategy, whereas section 5 presents the main results of the analysis. Finally section 6 discusses the results and the limits of the analysis and section 7 concludes.

2 Measuring the effect of private schooling

The measuring of the effect of private schooling on the outcome of students is not easy at all. Clearly, one could make life easy by just comparing the average scores of students coming from private school to those of students coming from public schools. In a regression from, this model would be

$$S_i = \alpha + \delta P_i + \epsilon_i \qquad (1)$$

where S_i is the cognitive test score, α is the average test score of public school students P_i a dummy variable for private schooling and ϵ_i the error term. In this case, δ would indicate the additional gain from being at a private school in terms of test score.

What would such a measure tell us? Actually we could say something about the relative performance of both types of students, but we would be far form a causal inference of the effect of private schools on the knowledge accumulation. The main problem is that people going to private schools are commonly not a random sample of the whole population. If that would be the case, e.g. the only difference of the two types is the school type, then this easy method would yield to correct results. However, the assumption of identical populations in both types of schools is certainly not satisfied in reality. Children from relatively wealthy parents are certainly much more likely to go to a private school than poor children. Given that public schooling is for free and private schooling may be very expensive in Mexico, it is not hard to imagine, that the population in public schools differs quite substantially from the population in private schools.

Now, one could argue that by controlling for some family background characteristics, such as income, parents' education etc, we could get unbiased estimates of the private school effect. The model would then look like

$$S_i = X_i\beta + \delta P_i + \epsilon_i \qquad (2)$$

where S_i is the cognitive test score, X_i the set of control variables such as family income, P_i a dummy variable for private schooling and ϵ_i the error term. Estimating this equation by OLS would yield to unbiased estimated of δ according to Vandenberghe and Robin (2004), if the vector X_i perfectly controls for all other determinants of achievement. This is generally not the case, since we do not observe very important determinants such as motivation, ability and commitment to school. We can partially reduce this problem, when we have at least two measures of the cognitive test scores. This allows us to replace the dependent variable by the difference of the test score as proxy for the added value of knowledge, which then no longer depends on ability.

$$\Delta S_i = X_i\beta + \delta P_i + \epsilon_i \qquad (3)$$

where $\Delta S_i = S_{i,t} - S_{i,t-1}$. Alternatively, we could also include $S_{i,t-1}$ on the right hand side of the regression as a proxy of the initial ability. This permits $S_{i,t-1}$ not to have an elasticity of 1 to $S_{i,t}$ and is therefore less restrictive. For this reason the model would write

$$S_{i,t} = X_i\beta + \delta P_i + \eta S_{i,t-1} + \epsilon_i \tag{4}$$

We can take this model as our "benchmark" model, but we should not forget, that this yields only to unbiased estimates and the strong assumption of perfect description of all determinants of achievement by the X_i vector and the initial test score $S_{i,t-1}$. Therefore, additional models must be included in order to control for potential biases. I use in addition to the OLS estimation of equation 4 two more estimation procedures, the instrumental variable and the Heckman two-stage estimator. Both methods allow reducing the bias due to the endogeneity of the private school participation by instrumenting the dummy. The main concern by doing this is to find a valid instrument which reasonably well explains the decision of going to a private school on one hand, and which is not explaining the test scores on the other hand. As such instrument I use in this study the geographical location, mainly the division in urban, less urban and rural areas. We have reasons to believe that private school supply is mainly concentrated in urban areas and that the fact of living in a rural or urban area does not directly explain the cognitive skills of people. I discuss the validity of the instrument in the result section of the paper.

3 Data

The data I use in this study comes from the *Mexican Family Life Survey* (MXFLS) which is a two-period panel of a very complete household survey, carried out in 2002 and 2005. The whole survey includes around 8500 households in almost 20 states of Mexico. Respondents were interviewed about very different topics, such as labor, income, consumption, education, health and cognitive skills. The latter one was assessed by cognitive tests included in the survey, which were separately applied to adult members of the household and member below 15 years. The big advantage of this data is its panel structure, which allows us to get two cognitive skill measures for each individual. Moreover, the cognitive test was not applied at school, therefore special preparation of some school classes for the test do not matter in this case, differently to surveys carried out in school. In the following section, I explain how the test score indicator was obtained from raw data, thereafter I explore with quite some details the explanatory variables I use in the study.

A third period will be added with values of the year 2008.

3.1 The cognitive test scores

Respondents were asked to complete a relatively short cognitive test where they had to complete the missing part of an abstract picture. They had the choice among 6 different responses. The test for household members below 15 years consists of 18 questions, while the adult test is limited to 12 pictures. Figure 1 shows an example of a question taken from the youth questionnaire.

Given that only one answer can be right, the variable describing their performance on each question is reduced to a dummy variable, indicating 1 if the answer was correct and 0 otherwise. From these 18 respectively 12 dummy variables, I had to construct a cognitive test score. A very easy way would be to average just all the questions, which would yield to an index on the interval 0 to 1. The problem is that the different questions do not have the same level of difficulty and such an index would give the same weight to all questions, which would then lead to a wrong approximation of the cognitive skills. Therefore I use in this study two different methods of aggregation. The first method is an *ad-hoc* method, where I take a weighted average of the dummy variables

$$S_i = \sum_{q=1}^{N} w_q D_{iq} \tag{5}$$

where w_q is the weight of each question and D_{iq} is the q^{th} dummy variable of individual i. The weights are taken such that they are related to the difficulty

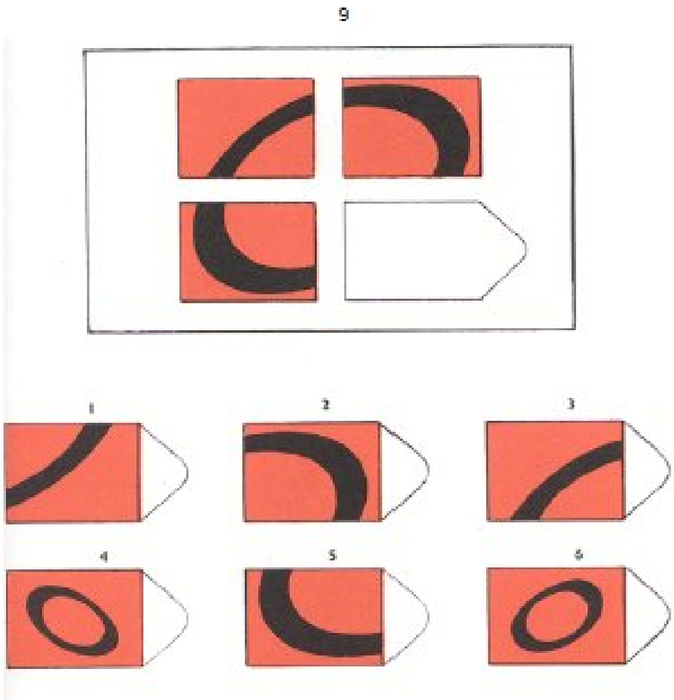

Figure 1: Example of a question in the youth questionnaire

of the question, giving more weight to the difficult pictures. The difficulty of a question is approximated by the percentage of wrong answers in the population. This gives the following definition:

$$w_q = \frac{p_q}{\sum_{q=1}^{N} p_q} \tag{6}$$

where p_q is the percentage of wrong answers to question q in the population. As I mentioned before, this method is *ad-hoc*, however, it might be reasonable to give more weight to the harder questions and the way I do it here is one of the simplest.

A second way I define the test score is applying a polychoric factor analysis on the set of dummies coming from the test. I then use only the first factor and its loadings in order to construct the test scores. Scores are normalized to the interval [0,1] in order to have comparable results with the other test score measure. The problem encountered in the application of this, a priori, more sophisticated method, is that there seem to be two factors with eigenvalues above 1. However, since I am only interested in one dimensions, supposed to proxy cognitive abilities, I retain only one factor.

Comparing the two methods allows identifying differences and similarities. As one can observe in figure 2, the two measures are strongly correlated, having a linear correlation of 0.9202.

Throughout the analysis, I present the results for both cognitive skills indices and I denote them with S_i^a for the *ad-hoc* measure and S_i^f for the factor analysis index.

The score index is computed for both periods, however, using the same weights coming from the first observation. This allows a direct comparison and given that exactly the same test was applied, this method seems to be justifiable. One would imagine that there is a strong relation of the two years.

As one can see in table 1 the average test score increases from 2002 to 2005 in both measures quite substantially. This is not surprising, since only people having done both tests are considered, therefore the increase in the indices reflects an increase in cognitive skills.

However, there seems to be a lot of noise in the

This is generally the threshold to retain a factor

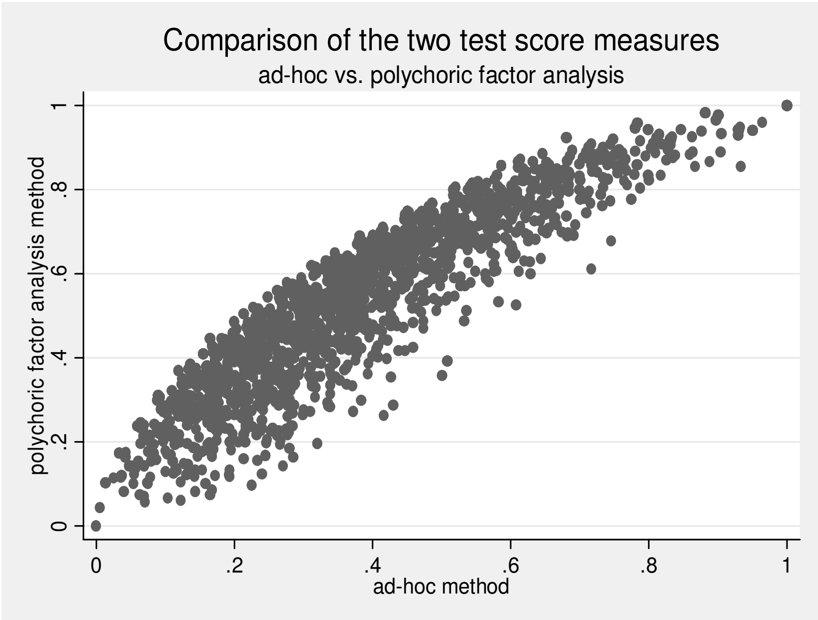

Figure 2: Correlation of the two aggregation methods

Table 1: Average test scores in both periods

Year	Ad-hoc	Factor Analysis
2002	0.39	0.56
	(0.21)	(0.22)
2005	0.55	0.72
	(0.23)	0.20)

Standard errors in parenthesis

measure, probably due to different interviewers or circumstances. Even the motivation can play a crucial role. Since the analysis would be probably biased, I decided to restrict the sample to plausible values, which is indeed highly arbitrary. At some point, it is necessary to decide whether an observation should be taken into account or not. It seems to be quite implausible that a student got an index near 1 in the first period and close to zero in the second, which would indicate a huge loss of cognitive skills. Therefore a first sample is constructed in such a way that the bottom and top 5% in terms of differences in the two measures are excluded. This eliminates most probably those not paying attention to the test in one of the two periods. Given that the measure is arbitrary, I propose a second way to define the sample, which is to take only those children that did better in the second period, excluding again the same top 5%. This may be justified by the fact that there is no obvious reason for a child to loose cognitive skills when going to school.

Moreover, both indices are taken in logs in order to get elasticities, rather than absolute values.

3.2 Explanatory variables

Besides the not straightforward definition of the dependent variable, the set of explanatory variables is somewhat challenging in several aspects. First and in contrast to the prior impression, there are many missing data in the different variables, this is particularly true since I use as well data of the children as of their parents.

I include a set of constant characteristics of the students, such as gender and their ethnic background. In the case of Mexico, it is interesting to distinguish between indigenous and non indigenous

The analysis does not yield to substancially different results when taking other samples people. A student is considered to be indigenous if he or she declares to speak an indigenous language. Age is computed based on the birth date and the date of the interview, the unit I retain is month, rather than years. In addition, a dummy variable is computed which indicates if the student suffers of underweight, according to the criteria of the (WHO, 2009). This variable is taken into account, because the cognitive capacity seems to be influenced by undernourishment according to the literature (see for example Behrman and Rosenzweig (2004); Grantham-McGregor (1995)).

Besides the characteristics of the student, some indicators of his or her family background are taken into account. First, I use the log consumption per capita in the household, rather than income, since it may reflect closer the average wealth of the family. The cognitive test score of the mother, computed in the same way as for children, is used in order to proxy the cognitive skills of the mother and to take into account some genetical transmission of abilities. The highest education of the parents is computed based on their schooling achievement and considered to proxy the social status and family's affinity to schooling.

A set of dummy variables is constructed to describe the geographical location of a family, according to the size of the location. This variable is used as instrument in the IV-regression and the Heckman method.

Finally a set of variables describing the school of the student are considered. In first instance, a categorical variable containing information about the private schools is considered. Since there are two school years involved between the two tests, both are taken into account, simply by averaging the two private school dummy variables. Additionally information regarding the repetition of a grade is taken into account. Both variables are self-declaration of the students or their parents. Finally I decided to renounce using information about the class size or the number of teachers at school, since many missing data were present, which would have reduced the set of usable observations by a lot. This is actually a problem throughout the analysis, given than many observations were lost due to non response. This may induce a bias in my estimates. I will come back to this in the discussion of the results.

4 Estimation strategy

According to what was said in section 2 I use the different methods in order to see the differences and to check if the theoretical changes in the coefficients are satisfied in reality. When using the very simple model described by equation 1 I would expect a positive effect of the private schooling, since it includes as well the positive selection bias, as a potential real effect of private schooling. Hence, by estimating equation 2 where the set of control variables is included, we would expect lower coefficients, but still a potential upward bias coming from the positive selection due to unobserved abilities. Therefore, by incorporating the ability proxy as described in equation 4 we could expect a further decrease of the coefficients. Under the assumption that by including this information, the endogeneity issues are eliminated, we could expect an unbiased estimation.

Finally I do present some IV-regression and Heckman estimates. However, the key issue by doing this is that the instrument is valid. The data did not allow me to find a better instrument than the geographical location, meaning the size of the location. Moreover, I do have only information in a categorical way. This might be a weak instrument, but in absence of a better one, I have to present the results using this one.

5 Results

In this section, I present the results according to the method presented in the previous section. It might be interesting to present the results in such a chronological order, which allows getting an idea of the relative importance of the potential biases. Therefore, first I present the very naive estimates

Table 2: Naive estimates

	Ad-hoc		Factor analysis	
	Sample 1	Sample 2	Sample 1	Sample 2
	(1)	(2)	(3)	(4)
Private school	0.206***	0.205***	0.119***	0.123***
	(0.048)	(0.042)	(0.035)	(0.028)
Constant	-0.699***	-0.579***	-0.373***	-0.298***
	(0.014)	(0.012)	(0.010)	(0.008)
R-squared	0.005	0.007	0.003	0.005
N	1558	1172	1558	1172

Source: Authors calculation. Std. Errors in parenthesis.
Significance levels at 10% (*), 5% (**) and 1% (***).

according to equation 1. Table 2 presents the results from this simple regression. Depending on the measure of cognitive skills, private school students perform around 20 respectively 12 log points better than students from public schools. This difference is relatively big and in all estimation highly significant at a 1% level. It is important to remind that these results do not permit any conclusion regarding causality, since they have just a descriptive character. We observe indeed that private school students have higher cognitive skills, but we are not able to say if that is due to the private school or rather that they are in the private school due to their higher cognitive skills. Although there is no causal analysis at this step, it might explain why graduates from private schools have easier labor market access.

In a second step, I include a set of background variables as described in section 3.2. One can expect that the coefficient of private schools goes down sharply when doing that, given that this set of characteristics explains part of the selection process of private schools. Table 3 presents the results according to equation 2. As we can see, the positive and highly significant effect of private schooling almost completely disappeared. Only using sample 2 and the ad-hoc aggregation method leads still to a positive and significant effect, although much smaller than in the previous results. This sharp decrease in the coefficients is due to the fact that the included variables explain part of the performance and part of the selection process of private schools. One can easily imagine that students coming from richer families have better access to private schooling and in the same time, it seems to be true that the wealth of the family matters in the production of cognitive skills (see for example Plug and Vijverberg (2005)).

Regarding the background characteristics it seems to be true that underweight has a negative and highly significant effect on cognitive abilities of children. In the same biological way, mother's abilities have a positive effect, which might support the genetical transmission theory of cognitive skills or be due to the possibility to help children at home. The fact of repeating a grade is directly linked to the cognitive scores as well, which is not very surprising. It will be interesting to see the behavior of this variable when including the proxy of abilities in the set of explanatory variables.

For the student related variables I find a negative effect of indigenous children, however the effect is not stable. The same is true for girls, who seem to have on average a slightly lower performance than boys. This finding might be due to the nature of the test. However, the differences between girls and boys are not the main issue of this paper and goes far beyond the scope of the analysis.

Now, let's turn to the estimation of equation 4, which is my benchmark model, since it is able to identify the prior abilities of children. The results

Table 3: OLS including the set of control variables

	Ad-hoc		Factor analysis	
	Sample 1	Sample 2	Sample 1	Sample 2
	(1)	(2)	(3)	(4)
Private school	0.050	0.086**	0.003	0.039
	(0.047)	(0.038)	(0.033)	(0.025)
Indigenous	-0.079**	-0.042	-0.073**	-0.023
	(0.040)	(0.035)	(0.030)	(0.026)
Female	-0.023	-0.043**	-0.012	-0.020
	(0.023)	(0.021)	(0.016)	(0.015)
UW	-0.168***	-0.154***	-0.151***	-0.125***
	(0.042)	(0.042)	(0.037)	(0.034)
Log consumption per capita	0.018	0.020*	0.021**	0.017**
	(0.013)	(0.011)	(0.009)	(0.007)
Mother's score	0.311***	0.230***	0.158***	0.120***
	(0.053)	(0.048)	(0.036)	(0.032)
Parents education	0.017***	0.012***	0.014***	0.009***
	(0.004)	(0.004)	(0.003)	(0.003)
Repeated grade	-0.126***	-0.130***	-0.077***	-0.081***
	(0.036)	(0.036)	(0.026)	(0.025)
Age in month	0.002	0.002	0.001	0.003
	(0.005)	(0.005)	(0.003)	(0.003)
Age in month (squared)	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-1.392***	-1.173***	-0.886***	-0.775***
	(0.235)	(0.236)	(0.158)	(0.162)
R-squared	0.157	0.152	0.161	0.142
N	1558	1172	1558	1172

Source: Authors calculation. Std. Errors in parenthesis.

Significance levels at 10% (*), 5% (**) and 1% (***).

are presented in table 4.

The first observation one can make is a further decrease in the coefficient of private schools, as we could expect it to be. Now, private schooling does no longer present any positive and significant effect, apparently once we control for abilities and background variables, private schooling does not have an effect on the accumulation of cognitive skills. The newly introduced variable is highly significant at a 1% level. However, it's important to see that the elasticity is far from being 1, which supports the inclusion on the right hand side rather than as part of the dependant variable. Moreover, the effect is surprisingly robust across aggregation methods and samples. As mentioned before, the evolution of the coefficient related to the dummy capturing a repetition of a grade is interesting. It is much smaller than before, which is obvious, since the repetition of a grade is supposed to be highly correlated with the initial cognitive skills. The rest of the explanatory variables show persistent effects regarding their significance. Regarding the size of the effect, we can observe several small changes, but overall the observations made before remain valid. Moreover, looking at the R^2 of model (2) we can see that it is relatively high, considering the high amount of noise in the data.

Finally I tried to estimate an IV-regression and a

Table 4: OLS including the set of control variables and ability proxy

	Ad-hoc		Factor analysis	
	Sample 1	Sample 2	Sample 1	Sample 2
	(1)	(2)	(3)	(4)
Private school	0.011	0.036	-0.019	0.020
	(0.045)	(0.034)	(0.033)	(0.026)
Indigenous	-0.067*	-0.020	-0.060**	-0.005
	(0.038)	(0.029)	(0.030)	(0.023)
Female	-0.004	-0.029*	-0.003	-0.014
	(0.022)	(0.017)	(0.015)	(0.013)
UW	-0.150***	-0.096**	-0.119***	-0.060**
	(0.042)	(0.038)	(0.036)	(0.030)
Log consumption per capita	0.007	0.007	0.012	0.007
	(0.013)	(0.010)	(0.009)	(0.007)
Mother's score	0.180***	0.083**	0.080**	0.045
	(0.048)	(0.039)	(0.034)	(0.028)
Parents education	0.012***	0.004	0.011***	0.005**
	(0.004)	(0.003)	(0.003)	(0.002)
Repeated grade	-0.082**	-0.069**	-0.047*	-0.041*
	(0.033)	(0.028)	(0.024)	(0.021)
Age in month	0.004	0.005	0.001	0.002
	(0.004)	(0.004)	(0.003)	(0.003)
Age in month (squared)	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Score 2002 (ad-hoc)	0.269***	0.356***		
	(0.021)	(0.021)		
Score 2002 (FA)			0.228***	0.269***
			(0.020)	(0.018)
Constant	-0.938***	-0.534***	-0.556***	-0.354***
	(0.216)	(0.181)	(0.142)	(0.127)
R-squared	0.251	0.416	0.240	0.326
N	1558	1172	1558	1172

Source: Authors calculation. Std. Errors in parenthesis.

Significance levels at 10% (*), 5% (**) and 1% (***).

Heckman two-steps model, using the geographical location as an instrument. Tables 5 and 6 present the results for the ad-hoc and the factor analysis aggregation method respectively. If there is still a bias in the previously presented results, then estimates should be even smaller in the case of IV and Heckman. In the case of the ad-hoc aggregation method this is partially true, however, the conclusion remains exactly the same. For the case of the factor analysis aggregation index we can observe a sharp increase of the coefficient to quite unreasonable values, however, still insignificant. This result supports my concerns about the validity of the instrument I use. For this reason, I would personally prefer the results of table 4.

6 Discussion

The results of this study suggest therefore that the better performance of private school students is due

Table 5: IV and Heckman estimates for the ad-hoc index

	Heckman two-steps		IV-Regression	
	Sample 1	Sample 2	Sample 1	Sample 2
	(1)	(2)	(3)	(4)
Score 2002 (ad-hoc)	0.269***	0.357***	0.271***	0.363***
	(0.021)	(0.021)	(0.024)	(0.025)
Indigenous	-0.066*	-0.022	-0.067*	-0.027
	(0.039)	(0.029)	(0.038)	(0.033)
Female	-0.004	-0.029*	-0.004	-0.035*
	(0.022)	(0.017)	(0.022)	(0.020)
UW	-0.149***	-0.097**	-0.150***	-0.101**
	(0.042)	(0.038)	(0.042)	(0.039)
Log consumption per capita	0.006	0.008	0.010	0.012
	(0.013)	(0.010)	(0.017)	(0.014)
Mother's score	0.180***	0.083**	0.181***	0.078*
	(0.048)	(0.039)	(0.049)	(0.042)
Parents education	0.012***	0.004	0.013**	0.007
	(0.004)	(0.003)	(0.006)	(0.006)
Repeated grade	-0.082**	-0.070**	-0.085**	-0.081**
	(0.033)	(0.028)	(0.035)	(0.034)
Age in month	0.004	0.005	0.004	0.006
	(0.004)	(0.004)	(0.005)	(0.004)
Age in month (squared)	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Private school (est)	0.047	-0.115		
	(0.355)	(0.264)		
Private school			-0.165	-0.521
			(0.653)	(0.952)
Constant	-0.938***	-0.547***	-0.981***	-0.611***
	(0.217)	(0.182)	(0.274)	(0.227)
R-squared	0.251	0.415	0.247	0.357
N	1558	1172	1558	1172

Source: Authors calculation. Std. Errors in parenthesis.

Significance levels at 10% (*), 5% (**) and 1% (***).

to the positive self-selection process into private school and not the fruit of a better education. Although, this does not mean that labor markets prefer without a reason private school graduates, since as we saw in table 2, they have higher cognitive skills on average. Hence, for the labor market decisions, such a simple analysis might be sufficient, if only the current cognitive skill level matters. However, the results from the slightly more sophisticated analysis suggest, that the net return to education in terms of knowledge accumulation is not statistically different in private and public schools. Most of the observed differences in the simple analysis seem to be due to the background variables, such as the education of the parents, the gender or even the cognitive skill level of the mother. The bias due to the self-selection based on different abilities, does not seem to be as big as one might expect, however, it is present.

In general we can therefore take two main conclu-

Table 6: IV and Heckman estimates for the factor analysis index

	Heckman two-steps		IV-Regression	
	Sample 1	Sample 2	Sample 1	Sample 2
	(1)	(2)	(3)	(4)
Score 2002 (FA)	0.227***	0.268***	0.218***	0.263***
	(0.020)	(0.018)	(0.022)	(0.020)
Indigenous	-0.058*	-0.005	-0.059*	0.000
	(0.030)	(0.023)	(0.031)	(0.024)
Female	-0.004	-0.015	-0.002	-0.009
	(0.015)	(0.013)	(0.016)	(0.016)
UW	-0.117***	-0.059**	-0.119***	-0.056*
	(0.036)	(0.030)	(0.037)	(0.030)
Log consumption per capita	0.008	0.006	0.002	0.002
	(0.009)	(0.007)	(0.013)	(0.010)
Mother's score	0.078**	0.043	0.075**	0.048
	(0.034)	(0.028)	(0.036)	(0.030)
Parents education	0.010***	0.005*	0.007*	0.002
	(0.003)	(0.002)	(0.004)	(0.004)
Repeated grade	-0.047*	-0.042**	-0.035	-0.030
	(0.024)	(0.021)	(0.026)	(0.026)
Age in month	0.001	0.002	-0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Age in month (squared)	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Private school (est)	0.330	0.180		
	(0.236)	(0.198)		
Private school			0.562	0.522
			(0.486)	(0.656)
Constant	-0.529***	-0.344***	-0.414**	-0.281*
	(0.143)	(0.128)	(0.190)	(0.158)
R-squared	0.241	0.326	0.148	0.230
N	1558	1172	1558	1172

Source: Authors calculation. Std. Errors in parenthesis.

Significance levels at 10% (*), 5% (**) and 1% (***).

sions. First, there is indeed a higher cognitive skill level observable among students from private institutions. Second, this difference seems to be due to the non randomness of students in private schools, not because private schools would perform better.

However, all these results are drawn from an analysis which encounters several practical problems. First, the data I use in the study is actually not as good as it seemed to be at the beginning. Especially the relatively high number of observations I had to exclude due to missing data might cause a bias. Moreover the proportion of students actually going to a private school is quite low and it would be certainly good to run a similar analysis based on a bigger sample of student. Finally a more technical problem I found is the instrument used in the study, which does not seem to be very convincing.

Despite all these issues, the results seem to be relatively robust and the behavior of the coefficients of private schooling behave as expected throughout the chronological application of the methods, going from the naive estimation to more realistic methods.

Regarding the best choice of the aggregation to the cognitive skill index, I would probably prefer the use of the "ad-hoc" method, since it generates more plausible results. Especially for its simplicity it has some advantages over the polychoric factor analysis method.

7 Conclusion

In this analysis I used data from the *Mexican Family Life Survey* to estimate the effect of private schooling on the accumulation process of cognitive skill. Different methods are used, going from a very simplistic to more sophisticated. The educational outcome is measured using a cognitive ability test applied to the respondents, from which I then compute an index of cognitive abilities. The results suggest that students from private school indeed present higher average cognitive skill, but that these differences are not due to a better education in private school, but to the selection process of students into private school. Not only the self-selection matters, also external determinants such as gender, ethnicity and parental education.

References

Angrist, Joshua, Eric Bettinger, Elizabeth King, and Michael Kremer, "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment," *The American Economic Review*, December 2002, *92* (5), 1535–1558.

Behrman, J.R. and M.R. Rosenzweig, "Returns to Birthweight," *Review of Economics and Statistics*, 2004, *86* (2), 586–601.

Finger, John A. and George E. Schlesser, "Academic Performance of public and private school students," *Journal of Educational Psychology*, 1963, *54* (2), 118–122.

Grantham-McGregor, Sally, "A review of studies of the effect of severe malnutrition on mental development," *Journal of Nutrition*, 1995, *125* (8), 2233–2238.

Hanushek, Eric A., "Publicly provided education," in A.J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics - Volume 4*, Elsevier Science B.V., 2002, chapter 30, pp. 2046–2125.

Horowitz, John B and Lee Spector, "Is there a difference between private and public education on college performance?," *Economics of Education Review*, 2005, *24*, 189–195.

Plug, Erik and Wim Vijverberg, "Does Family Income Matter For Schooling Outcomes? Using Adoptees As A Natural Experiment," *The Economic Journal*, October 2005, *115*, 879–906.

Rouse, Cecilia Elena, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *The Quarterly Journal of Economics*, 1998, *113* (2), 553–602.

Vandenberghe, V. and S. Robin, "Evaluating the effectiveness of private education across countries: a comparison of methods," *Labor Economics 11*, April 2004, *11* (4), 487–506.

WHO, "Growth reference 5-19 years," World Health Organization, online at: http://www.who.int/growthref/en/ 2009.