Accounting for Unobservables in Comparing Selective and Comprehensive Schooling [1]

Stéphane Bonhomme[2]
CEMFI, Madrid

Ulrich Sauder[3]
University of Warwick

This draft: February 2007

PRELIMINARY AND INCOMPLETE
PLEASE DO NOT QUOTE

[1] We thank Martin Browning, Per Johansson and participants at Uppsala university and RTN conference in Madrid for helpful coments. We also thank Peter Shepherd, Director of Survey Operations of the Centre for Longitudinal Studies.
[2] CEMFI, Casado del Alisal, 5, 28014 Madrid, Spain; bonhomme@cemfi.es
[3] Department of Economics, Coventry CV47AL, England; u.m.sauder@warwick.ac.uk

Abstract

We compare the effects of selective and non selective education on various educational outcomes, using British data from the National Child Development Study (NCDS). Our modelling framework allows the unobserved and possibly multidimensional child endowment to simultaneously affect treatment and outcomes. Using pre-treatment test scores also affected by the endowment, we develop a quasi-differencing method to correct the matching estimand for the bias created by the presence of the unobserved confounding factors. We find that, controlling for pre-treatment differences, pupils' selection has a positive but small and often insignificant effect on the means of outcomes. In contrast, selective schools show significantly more variance than non selective ones. We also find evidence of a premium associated with being admitted to a more ability demanding grammar school. We conclude that the selective system is mainly characterized by a higher risk, which provides a rationale for risk-averse parents to push towards comprehensivisation.

JEL codes:
Keywords:

1 Introduction

In this paper, we compare the effects of selective and non selective education on various educational and labor market outcomes, using data from the National Child Development Study (NCDS). The recent history of British secondary education has been characterized by the shift from a selective towards a non selective– or "comprehensive" –system. After the second world war and starting with the Butler Education act (1945) the education system in the U.K. was mostly selective. Namely, children were assigned to two different types of schools depending on their results to a test score at age 11, called the *11-plus* exam. Successful children went to grammar schools, while the others attended less demanding secondary modern schools. Then, an important change was initiated in 1965 by the Crosland Circular, by which a possibility of comprehensivisation was introduced, creating schools where children of different ability levels were pooled together. As the circular did not force schools to be comprehensive, the timing of the shift towards comprehensivisation has been heterogeneous. Still in 1965, less than 5% of all public schools were comprehensive. By 1975 this proportion had reached 60%. So in the period when the children of the NCDS, all born in 1958, were at school the two systems coexisted. This makes these data a potential laboratory for comparing the two systems.

There have been several recent attempts at measuring the effect of pupils' selection on outcomes using the NCDS data. In an early study, Kerckhoff (1986) measures the effects of school type on later educational outcomes. He finds that attending a grammar or secondary modern school is associated with higher and lower educational attainment, respectively, than attending a comprehensive school. Dearden *et al.* (2002) define "selective schools" as either grammar or private schools. Using propensity score matching techniques to control for a large set of covariates, they find that attending either of the two types is associated with higher outcomes. More recently, Galindo-Rueda and Vignolles (2005) focus on the differences between school systems: selective (either grammar or secondary modern) and comprehensive. They find that attending a selective school yields better educational outcomes. Moreover, the gain is found to be concentrated on high-ability pupils.

In an important methodological criticism to these findings, Manning and Pischke (2006) question the exogeneity of the comprehensivisation reform. They measure the effect of attending a selective school on test scores at age 11, that is before entering a secondary school. They find positive effects, of similar magnitude or larger than the effects on subsequent outcomes. They interpret this exercise as a falsification test, which suggests that attending a selective school is likely to be affected by unobservables that in turn affect later outcomes. They conclude that "we don't know very much about the effects of comprehensive schooling in Britain" (p.19).

The point raised by Manning and Pischke is the main motivation of this paper. We start by focusing on the differences between the selective and comprehensive schooling systems. Attending a comprehensive school is the determinant, or "treatment", of interest. Given the non exogeneity of the treatment, consistent estimation of treatment effects on the mean and variance of outcomes is a difficult task. Our strategy is inspired from recent advances in the education production function literature. We build a model where test scores depend on parental and school inputs as well as on the child's endowment, as in Todd and Wolpin (2003, 2004). The endowment is unobserved to the econometrician, and can be multidimensional. We embed this model into a framework *à la* Rubin (1974), where we do not simultaneously observe the two potential post-treatment outcomes. We assume that the same unobserved endowment affects both pre- and post-treatment outcomes as well as the probability of attending a selective or comprehensive school.

In the presence of selection-on-unobservables, usual matching estimators are biased. To correct the matching estimand for the bias created by the presence of the endowment, we develop a quasi-differencing approach inspired from the "within" approach to eliminate fixed effects in linear panel data models (*e.g.* Holtz-Eakin *et al.*, 1988). For this purpose, we make use of the availability of several test scores at age 11 (maths, reading, verbal), which are also affected by the endowment. Then, estimating the Average Treatment Effect (ATE) in our model is simple and transparent. Under the main assumption that pre-treatment outcomes are not affected by the treatment other than through the child's endowment, we obtain that the bias of the matching estimand on post-treatment outcomes is proportional to the difference in pre-treatment test scores between the two schooling systems. Moreover, the coefficient of proportionality can be estimated by using other test scores (such as the reading score at age 11) as alternative measures of the endowment.

Following this strategy one is able to consistently estimate the ATE as well as other treatment effects (AT for the Treated, ATE and ATT on the variances). It is to be noted that we do not assume anything on the correlation between the endowment and observed covariates, such as parental inputs. This fixed effects approach is especially appealing in the education production function perspective, where parents take decisions based on their child's ability. Moreover, the model implies testable restrictions on the data. This property is important in order to check if the assumed structure is correct.

Our identification strategy is closely linked to a class of models introduced by James Heckman and coauthors. Starting with Carneiro *et al.* (2003) and Hansen *et al.* (2004), these authors use factor models to restrict the correlation between measurements and achieve identification. Compared to the recent models in this literature (*e.g.* Cunha and Heckman, 2006, and Cunha *et al.*, 2006) our framework is more restrictive as the identifying content of the model is limited. In particular, we are only able to identify effects of the treatment on the mean and the variance, as we make no distributional assumptions (such as independence) on the errors. Another restriction compared to recent advances in that literature is the linear way the endowment enters the model. One virtue of our approach is that it does not restrict the correlation between the endowment and the covariates, and that it yields easily testable implications.

Applying our methodology to the NCDS data, we find evidence of a bi-dimensional endowment. We find that, correcting for the differences in endowment at age 11, attending a comprehensive rather than a selective school has a negative effect on means of outcomes, but small and insignificant in many cases. Stronger are the results we obtain for variances, as the selective system is found to present significantly more dispersion than the comprehensive one.

We then focus on the variance within the selective system. We argue that one important part of this variance comes from the differences between achievements at grammar and secondary modern schools. In a tentative attempt to measure the premium associated to the fact of passing the *11-plus* exam, we find strong and often significant effects. We interpret these results as a potential explanation of risk-averse parents being likely to have pushed towards comprehensivisation.

The effects of pupils' selection and "ability tracking" on outcomes are currently attracting interest ouside the U.K. too. In the U.S., there is already a large literature on the subject, with mixed evidence (see Figlio and Page, 2002, and references therein). Comparable to the British case, one of the main features of the Swedish reform that took place in the 1950's was to abolish selection by ability into academic and non-academic streams. Meghir and Palme (2004) exploit the fact that the reform was preceded by a social experiment to estimate its effects on educational and labor market outcomes. They find positive effects, concentrated on children with lower parental background. Using country-level data and a difference-in-difference methodology, Hanushek and Woessman (2006) find little evidence of positive effects of ability tracking on mean outcomes. However, they also find that the variance of educational attainment is larger in countries with selective education (but see also Waldinger, 2006). Maurin and McNally (2006) study the effects of a shift in the admission criteria in grammar schools in Northern Ireland. They find that the probability of attending a grammar school and of achieving higher educational qualifications both strongly increased after the reform.

The outline of the paper is as follows. In Section 2 we present the NCDS data and compute some preliminary estimates of the effect of selection on various outcomes. We also present some evidence that the treatment consisting in attending a comprehensive school is not exogenous, building on Manning and Pischke (2006). We present our model in Section 3 and list and motivate the assumptions we make. Section 4 is devoted to the identification of causal parameters, starting with the ATE. In that section we also provide estimators of the parameters of interest, and derive some testable implications of the model. In Section 5 we apply our methodology to the NCDS data, and compare and contrast the two schooling systems. In Section 6 we focus on the differences that exist within the selective system, between school types. Lastly, Section 7 concludes.

2 A first look at the data

The NCDS is an ongoing longitudinal survey of a British birth cohort born between March 3 and March 9 of 1958. The initial sample consisted of 17,634 individuals which were resurveyed on six further occasions in order to monitor their changing health, education, social and economic circumstances. Attrition has led the sample size to shrink in the subsequent waves:[1] in 1965 at age 7, 1969 at age 11, 1974 at age 16, 1981 at age 23, 1991 at age 33, and 1999/2000 when the cohort had reached the age of 42. We will be looking at schooling achievements of the NCDS cohort members and will use information from all five waves. The NCDS children received secondary schooling between 1969 and 1974, and thus lived through the change of the British education system.

Sample selection. We are interested in the relative effects of selective and comprehensive schooling on outcomes. We use a NCDS variable that gives the type of the school attended by the child at age 16, when finishing secondary education. We do not use observations for which this variable is missing. We also exclude other types of schools, such as technical or *public* (*i.e.* private) schools. These schools are likely to be very different from the ones we consider. Moreover, they represent a small percentage of all schools (less than 10%). Note that the school type variable also indicates which kind of

[1]There has been a large amount of attrition in the NCDS. Current practice has treated attrition as exogenous. Conolly *et al.* (1992) show that attrition is somewhat stronger among children from lower parental background. Dearden *et al.* (2002) argue that the results are not necessarily biased, as they dispose of many indicators of parental characteristics, as we also do.

selective school, grammar or secondary modern, the child was attending at age 16. We will use this information in Section 6.

For the purpose of comparing children that went through fully selective or fully comprehensive secondary education, we keep children who stayed during five years in the same schooling system. To select the sample, we use a variable which gives the teacher's opinion on the type of school she thought the pupil was about to enter at age 11. We classify as comprehensive the observations for which the teacher's opinion at 11 and the school type at 16 both indicate that the child's school was comprehensive, and as selective when the two sources agree that it was selective. See the data Appendix for details about sample construction. We obtain a sample of 5155 observations in the sample, 2474 (48%) of which are comprehensive.

Variables. We are interested in the effect of the schooling system on several educational outcomes. We shall consider two outcomes at age 16: the maths and reading test scores given by the NCDS staff. It is important to notice that, unlike the math test, the reading test was the same at age 11 and age 16. For this reason, the distribution of the reading score at 16 is concentrated at high values. The presence of threshold effects (Kerckhoff, 1986) might explain why we often find different results for this outcome variable, especially in the case of variances. In addition, we will look at later educational outcomes: age when the child left school, and number of O-levels and A-levels obtained by the child. We are currently working on the construction of more complete educational measures.

We will also use indicators of educational level before secondary schooling. In particular, we will use the results to the NCDS test scores at 7 and 11, before the *11-plus* examination took place. Maths and reading tests were administered at ages 7 and 11, and an additional verbal test was administered at age 11 only.

We will also use parental background, such as the social class and education levels of the parents measured when the child was 7 or 11 years old, and school characteristics, such as class size variables, at age 7, 11 and 16. The complete list of these variables is given in the Appendix. Finally, we complement the data by constructing a set of local controls. For this purpose, we merge the 1971 Census statistics to the NCDS data. We then obtain characteristics of the enumeration district in which the child lives. In addition, we construct a political variables set. As the education reform was first implemented on a voluntary basis, political control of the Local Authority e.g. City or County Borough Council seems an important variable to explain the shift towards comprehensive schooling. We aggregated the results of the 1970 general election for each constituency. See the Data Appendix for a more precise description.

Descriptive statistics. Table 1 shows some descriptive statistics for the two groups of children in the sample, attending a comprehensive or a selective system. Educational outcomes are higher for children attending selective schools, who score on average 2.3 points higher in mathematics and 1.7 points in reading than children at comprehensive schools, about 30% of one standard deviation in each case. Moreover, they leave school on average half a year later, and obtain more O and A-levels. There is no marked difference for hourly wages. Now, the two groups are also very different in terms of intake, as children at selective schools score better at all tests at age 7 and 11. For instance, they score 4.1 points higher in mathematics at age 11, 40% of one standard deviation. We find similar discrepancies for the dispersion of test scores, selective schools showing higher variance of pre-treatment variables at age 11, and post-treatment outcomes except the reading score at 16.

	Comprehensive			**Selective**		
Variable	**Mean**	**Std.Dev.**	**N**	**Mean**	**Std. Dev.**	**N**
Mathematics score age 16	11.6	6.1	2258	13.9	7.1	2571
Reading score age 16	24.9	6.5	2269	26.6	6.4	2579
Age when left full-time education	16.8	1.7	1697	17.3	2.2	1946
Number of O-levels obtained	2.51	3.0	1931	3.38	3.4	2236
Number of A-levels obtained	.58	1.1	1102	.86	1.3	1471
Mathematics score age 11	14.6	9.2	2440	18.7	10.5	2682
Reading score age 11	15.2	5.8	2439	17.0	6.1	2683
Verbal score age 11	20.0	8.4	2440	24.3	9.1	2683
Mathematics score age 7	5.0	2.4	2224	5.4	2.4	2484
Reading score age 7	22.7	7.0	2228	24.4	6.4	2497
Log Hourly real gross wage 1981	.569	.32	890	.575	.32	988
Log Hourly real gross wage 1991	.888	.46	1281	.942	.47	1450
Log Hourly real gross wage 2000	1.20	.49	699	1.23	.52	830

Table 1: Comparing outcomes in selective and comprehensive schools.

Another way to illustrate the differences pre-secondary education is to summarize the information contained in the test scores previous to the *11-plus* examination into one scalar indicator. It is common practice to compute the first principal component of the 7 and 11 test scores and to label it as “ability” (*e.g.* Galindo-Rueda and Vignoles, 2005).

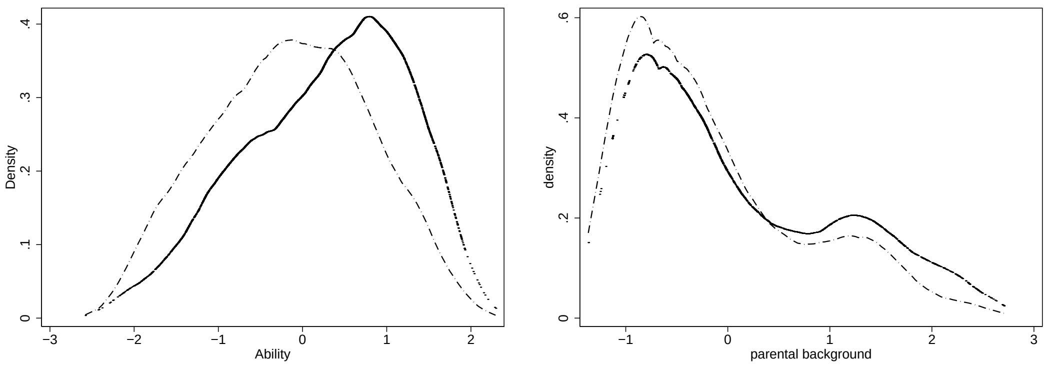

Figure 1: The "ability" and "parental background" principal components for pupils attending comprehensive (dash-dotted line) and selective schools (solid line).

Drawing the densities of the "ability" variable for the two groups shows strong differences, as the left panel of Figure 1 illustrates. On the right panel, we draw the density of a "parental background" variable, similarly computed as the first principal component of the many NCDS variables related to the social class of parents or their education. Figure 1 shows that there are less marked but still noticeable differences between the two groups, children in the selective system coming from better background on average.

Raw effects. Next, we regress the maths score at age 16 on the indicator of comprehensive schooling, with various controls. Table 2 presents the results. The unconditional effect reported in column (1) is negative. In columns (2)-(5), we adopt a value-added specification and use as controls lagged test scores, at age 11 and 7. Such specifications have been widely used in the education production function literature. The results show a consistently negative effect of attending a comprehensive school on the math score. The introduction of lagged test scores as controls lowers the effect by a large amount. Still, the effect, although not especially large (about 6% of one standard deviation) is significant at the 5% level in all the specifications. Lastly, in column (6), we replace the lagged test scores by their first principal component (the "ability" variable). Columns (5) and (6), comparable in terms of other included covariates, are very similar, with a slightly lower R-squared in column (6).

We present propensity-score matching estimates in Table 3. We use the inverse probability weighting of Hirano *et al.* (2003) to compute the effects of comprehensivisation on the various outcomes. Details on the estimation will be given in the next sections. In the first row of the table, we use as controls the parental, school and local covariates used in column (6) of table 2. We find negative and significant effects of attending a comprehensive school on all outcomes. These effects remain negative when we include

	Outcome: Mathematics score at 16					
	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	−2.378 (.192)	−.366 (.130)	−.303 (.129)	−.284 (.127)	−.301 (.133)	−.330 (.136)
Math 11		.505 (.006)	.400 (.011)	.363 (.011)	.362 (.011)	
Read 11			.144 (.017)	.101 (.016)	.100 (.016)	
Verbal 11			.048 (.012)	.079 (.012)	.079 (.012)	
Child female				−1.57 (.127)	−1.57 (.128)	−2.10 (.128)
Father's education				.105 (.046)	.114 (.046)	.129 (.047)
Mother's education				.234 (.058)	.208 (.059)	.240 (.060)
Ability (factor)						4.78 (.071)
Constant	13.9 (.13)	4.42 (.15)	3.10 (.25)	3.15 (.75)	3.04 (.98)	12.5 (.99)
R-squared	.030	.577	.593	.616	.620	.595
N	4829	4827	4808	4808	4808	4829
Controls						
Individual	no	no	yes	yes	yes	yes
Test scores	no	no	yes	yes	yes	no
Parental background	no	no	no	yes	yes	yes
Regional & political	no	no	no	yes	yes	yes
Census & local stats	no	no	no	no	yes	yes

Table 2: The effect of attending a comprehensive school on the math score at 16, conditional on increasing sets of covariates.

the "ability" factor as a covariate, but they turn mostly insignificant.

	Math16	**Read16**	**Years educ.**	**# O-levels**	**# A-levels**
(1)	−2.388	−2.116	−.990	−.783	−.289
	(.239)	(.322)	(.402)	(.126)	(.051)
(2)	−.809	−.427	−.477	−.152	−.076
	(.218)	(.320)	(.383)	(.110)	(.047)

Table 3: The propensity score matching results of the effect of attending a comprehensive school on all outcomes; (1): controls without "ability" factor; (2): controls include the "ability" factor.

To summarize, results conditional on school, parental and local covariates suggest that attending a comprehensive school is associated with lower educational achievement. Controlling for lagged test scores, or alternatively for their first principal component, yields also negative estimates, though insignificant except for the math score at 16. However, the "value-added" strategy, which consists in directly controlling for lagged test scores in the outcome equations, is not without problems. Todd and Wolpin (2003) show that value-added specifications impose strong restrictions on the parameters of the production function of education outcomes. In a more general model, later test scores do not depend directly on lagged ones, but on the child's endowment that is unobserved to the econometrician. In this perspective, the test scores used in the value-added specification can be seen as informal proxies for the child's endowment at age 11. If the restrictions that this specification imposes are not satisfied, it is not clear what we are estimating by controlling for these variables as we did in Tables 2 and 3.

"Falsification test". Moreover, even within the logic of value-added specifications, there are reasons not to believe the above estimates of the effect of attending a comprehensive school. To see why, we follow Manning and Pischke (2006), and estimate the effect of attending a comprehensive school between age 11 and 16 on outcomes *prior to* entering secondary education. Table 4 shows the results of the regressions of the test scores at age 11, controlling for an increasing number of covariates.[2] We find strong effects, indeed even stronger than on outcomes at age 16. For instance, the effect on the math score is never lower than 2 in absolute value, that is 20% of one standard deviation. Estimates obtained using propensity score matching give a similar picture.

This evidence casts serious doubts on the exogeneity of the school system variable in

[2]Column (1) corresponds to the case without controls. Then we add: (2) the maths score at 7; (3) the reading score at 7; (4) individual and family characteristics until 1969; and (5) regional controls.

	Outcome: Mathematics score at 11				
	(1)	(2)	(3)	(4)	(5)
Comprehensive	−4.102	−3.208	−2.544	−2.227	−2.278
	(.277)	(.244)	(.222)	(.218)	(.225)
Math 7		2.221	1.397	1.317	1.307
		(.051)	(.053)	(.052)	(.052)
Read 7			.612	.560	.562
			(.019)	(.019)	(.019)
Female child				−1.094	−1.089
				(.220)	(.217)
Father's education				.303	.292
				(.085)	(.084)
Mother's education				.490	.470
				(.105)	(.104)
Constant	18.7	6.82	−3.61	−8.78	−6.74
	(.191)	(.323)	(.432)	(.970)	(.873)
R-squared	.041	.319	.445	.487	.485
N	5122	4706	4689	4572	4689

Table 4: The effect of attending a comprehensive school on the math score at 11, conditional on increasing sets of covariates.

outcome equations. Of course, this "falsification test" needs very special conditions in order to be a proper test of exogeneity. For the effect on age 11 test scores to be zero, one would in particular need a stationarity condition to be satisfied (Imbens, 2004), requiring that the effect of age 7 test scores on scores at age 11 be the same as the effect of age 11 test scores on subsequent outcomes. In our data this condition is not satisfied, as is clear from the difference in R^2 between Tables 2 and 4. In any case, this results point at the presence of unobserved variables that cannot be properly controlled for by the many covariates present in the NCDS data.

Alternative strategies. In this paper, we explicitely address the concern, raised by Table 4, that attending a selective or comprehensive school may be endogenous in test score equations. Our approach is motivated by the failure of standard methods to provide credible inference on the parameters of interest. A natural approach would be to find an instrument for the schooling system. Galindo-Rueda and Vignoles (2005) propose to use the political colour of the LEA for this purpose. Indeed, in the period schools in Labour dominated LEA's are more predominantly comprehensive than conservative ones. Although a natural idea, using political colour as an instrument fails to remove the bias as Manning and Pischke (2006) show. This is because Labour dominated LEA's have on average children of lower ability and parental background. Using the IV strategy in the age 11 test scores equations yields a large effect of attending a selective school (see

Table 21 in Appendix). Another approach would be to derive bounds for the effect of comprehensive schooling, as the effect obtained by OLS or matching can be interpreted as an upper bound (in absolute value). To find a lower bound, we would need to find a variable that influences positively the probability of attending a selective school, yet is negatively correlated with the unobservables in the outcome equations. One example of this strategy is the use of the affirmative action legislation in Rothstein and Yoon (2006). In our case, it is not clear that such a variable exists.

Lastly, one might want to check the sensitivity of the results to departures from exogeneity. Commonly used techniques assume the existence of an unobserved "confounding" regressor conditional on which the treatment, here the type of school attended, is randomly assigned. See *e.g.* Rosenbaum and Rubin (1983), Imbens (2003), and recently Ichino *et al.* (2006). In this approach, one checks the sensitivity of treatment effect estimates under various assumptions on the unobserved confounder. One weakness of the approach is that it is not always clear how to calibrate the magnitude of the sensitivity check. Altonji *et al.* (2005) propose to use *observed* covariates as a guide to calibrate the sensitivity parameters. However, in our case, the central role played by the child's ability, that is unobserved to the analyst, makes this method difficult to apply.[3]

Our strategy is close in spirit to the sentivity approach. We also assume that there exists an unobserved confounder. In addition, we use information on the test scores prior to secondary education to remove the bias created by the presence of the confounder on the treatment effect estimate. The structure we assume has the advantage of being in line with the recent literature on the education production function. We now formally present our approach.

3 A model accounting for selection on unobservables

We start with the Rubin (1974) framework. Attending a comprehensive school is the treatment the effects of which we want to study. We denote by $T = 0$ the fact of attending a selective schooling system, while the comprehensive alternative is denoted by $T = 1$. We consider an outcome that has two potential values, in the presence (Y_1) and the absence (Y_0) of treatment. To fix ideas, we think of the outcome as the math test score at age 16. We shall also use the reading score at 16, the age when left school, and the

[3]If one does not include lagged test scores in the set of covariates, then the estimates seem very robust. But if we include them, then they become close to zero. One source of the problem could be that the child's ability is a very special regressor, unfit to the framework in Altonji *et al.* (2005), where observed and unobserved variables are assumed randomly drawn from the same pool of covariates.

number of O and A-levels obtained. The two potential outcomes are not simultaneously observed by the econometrician, only the combination $Y = TY_1 + (1-T)Y_0$.

In addition, we consider two outcome variables, W and Z, that we think of as pre-treatment outcomes. For instance, W is the math score at age 11 and Z is another test score (*e.g.* reading or verbal) at the same age. These variables are not causally affected by the treatment and are always observed.

Our modelling of the four outcomes variables reads as follows:

$$\begin{cases} Y_0 &= f_0(X) + \alpha_0 \eta + \varepsilon_0, \\ Y_1 &= f_1(X) + \alpha_1 \eta + \varepsilon_1, \\ W &= g(X) + \beta \eta + U, \\ Z &= h(X) + \gamma \eta + V. \end{cases} \tag{1}$$

In this model, X are covariates that include children's characteristics as well as parental and school inputs. Their effect on the outcomes is modelled by the unknown functions f_0, f_1, g and h. Then, η is an unobserved variable that represents the child's endowment at the time the treatment is received, that is at age 11. The endowment, that includes child and school unobserved characteristics, has a distinct effect on the various outcomes, through the parameters α_0, α_1, β and γ.

We make the following assumptions.

Assumption 1 *(selection on unobservables)*

$$\mathbb{E}(\varepsilon_0|X,T) = \mathbb{E}(\varepsilon_0|X); \quad \mathbb{E}(\varepsilon_1|X,T) = \mathbb{E}(\varepsilon_1|X).$$

Assumption 1 is reminiscent of the "selection-on-observables" assumption often used in policy evaluation, where the treatment is independent of the potential outcomes conditional on observed covariates:

$$(Y_0, Y_1) \perp\!\!\!\perp T| \quad X, \tag{2}$$

where ". $\perp\!\!\!\perp$.| ." represents conditional independence. A weaker assumption requires that treatment and outcomes be independent conditional on observed and unobserved covariates:

$$(Y_0, Y_1) \perp\!\!\!\perp T| \quad X, \eta. \tag{3}$$

Assumption (3) is typically made when checking the sensitivity of treatment effects estimates to departure from exogeneity, as in Imbens (2003). In that approach, one calibrates the contribution of η to the model and check by how much the estimates vary.

Assumption 1 is slightly weaker than (3). In particular, mean independence is assumed instead of full independence.

Assumption 2 *(pre-treatment outcome)*

$$\mathbb{E}(U|X,T) = \mathbb{E}(U|X).$$

Assumption 2 requires that W be not affected by the treatment. Hence differences between $\mathbb{E}(W|T=1,X)$ and $\mathbb{E}(W|T=0,X)$ simply reflect composition effects as the composition in terms of η is different in the treated and non treated groups. We have in mind pre-treatment variables that are often used in order to test for the selection-on-observables assumption (2).

Assumption 3 *(additivity)*

$$\eta \perp (\varepsilon_0, \varepsilon_1, U, V) \mid \quad X, T,$$

Assumption 3 emphasizes the additive structure of the model. In the assumption, the notation ". $\perp$.| ." stands for conditional uncorrelatedness. A sufficient condition for this assumption to hold is:

$$\mathbb{E}(\varepsilon_0|X,\eta,T) = \mathbb{E}(\varepsilon_1|X,\eta,T) = \mathbb{E}(U|X,\eta,T) = \mathbb{E}(V|X,\eta,T) = 0.$$

Note that these assumptions do not assume any structure for the correlation between η and X. η is analogous to a "fixed" effect in a linear panel data model of the form:

$$y_{it} = x_{it}'\theta + \eta_i + u_{it},$$

where $\mathrm{Cov}(\eta_i, u_{it}) = 0$ for all t, but $\mathrm{Cov}(\eta_i, x_{it})$ is not restricted.

Assumption 4 *(instrument)*

$$V \perp (\varepsilon_0, \varepsilon_1, U) \mid \quad X, T,$$

Assumption 4 requires that Z be an "instrument" in the following sense. We assume that Z is uncorrelated with the residuals of the other outcome variables, conditional on observed and unobserved variables. Of course, the presence of the unobserved η creates a correlation between the variable Z and the potential outcomes Y_0 and Y_1. Hence, Z cannot be used directly as a proper instrumental variable.

Assumption 5 *(rank conditions)*

$$\beta \neq 0; \qquad \gamma \neq 0; \qquad \mathrm{Var}(\eta|X,T) \neq 0.$$

Lastly, Assumption 5 requires that η presents variation conditional on X and T and has an effect on the pre-treatment outcome and the instrument. It means that W and Z have an unobservable in common.

As presented here, the model allows for a one-dimensional endowment. It can immediately be extended in order to allow for a multivariate endowment $\eta \in \mathbb{R}^K$, and vectors of pre-treatment outcomes and instruments W and Z. The model is presented in this general version in the Appendix. We require the same assumptions, except that Assumption 5 is replaced by the condition that β, γ and $\text{Var}\,(\eta|X,T)$, that are now matrices, have all rank K, the number of dimensions of the endowment. In particular, we need at least K pre-treatment outcomes and K instruments for this assumption to be satisfied. As is intuitive, data requirements increase if we want to account for a multivariate structure of unobservables.

The model is in line with recent models of the education production function literature, as in Todd and Wolpin (2003, 2004). In that literature, test scores are modelled as combinations of present and past parental and school inputs, and of the child's endowment. We allow that parental and school decisions be based on the endowment, as the correlation between η and X is left unrestricted. Moreover, the residuals of the test score equations, ε_0 and ε_1, can include unobserved parental and school inputs, as they can also be correlated with X.

In the model, the child's endowment at age 11, that may include individual and school effects, can be correlated to attending a selective system. The endowment η is the only variable that correlates the treatment with the outcomes. In particular the residuals $(U, \varepsilon_0, \varepsilon_1)$ are mean independent of the treatment. In contrast, they can be freely correlated with each other. This correlation can represent a special ability that the child posseses for maths, compared to other subjects. This special ability is imposed not to influence the treatment, the endogeneity of the latter coming only from the endowment.

As emphasized by Todd and Wolpin (2003), it is important that η be possibly correlated to the parental and school variables. Two common strategies to cancel the child's endowment out are the use of data on siblings, and the use of time series of test scores. In both cases, a "within" estimator provides consistent estimates of the model's parameters. We do not have information on siblings, and the time series we have are rather short (age 7, 11 and 16). Moreover, it is not clear how to adopt a "within" approach in the presence of a treatment that can be correlated to the endowment. Lastly, approaches such as the ones advocated by Todd and Wolpin are not applicable if the endowment is multidimensional.

4 Identification of policy parameters

In this section, we show how to identify and estimate several treatment effects of interest if Assumptions 1-5 hold. The discussion is conducted in the case of the model in its simplest form, where the endowment η is scalar. A treatment of the multidimensional endowment case is provided in the Appendix, as well as a proof of all statements appearing in this section. Let us start with the Average Treatment Effect (ATE) parameter: $\Delta = \mathbb{E}(Y_1 - Y_0)$. We have:

$$\Delta = \mathbb{E}[\mathbb{E}(Y_1|X) - \mathbb{E}(Y_0|X)].$$

In contrast with the case where there is no selection on unobservables (e.g. Rosenbaum and Rubin, 1983), this quantity has no direct counterpart in the data. Formally, for $j = 0, 1$ we have:

$$\mathbb{E}(Y_j|X) \neq \mathbb{E}(Y_j|X, T = j).$$

To illustrate how our method works, consider the case where $\alpha_j = \beta = 1$. Then it follows immediately from Assumptions 1 and 2 that:

$$\mathbb{E}(Y_j - W|X) = \mathbb{E}(Y_j - W|X, T = j).$$

The selection on the differenced outcomes is only driven by the covariates X. We can thus estimate the effect of the treatment on $Y - W$, as the endowment η has been differenced out. This strategy, inspired from the "within" estimation of linear panel data models with fixed effects, is used in studies where data on siblings is available (Altonji and Dunn, 1996).

In our case, we have several test scores taken at different years, but the individual time series is rather short (three dates), and we see no *a priori* reason to assume that the weight of the ability is the same in every of them. We propose a generalization of the "within" approach which is based on the following property, that is also immediate based on Assumptions 1 and 2.

$$\mathbb{E}\left(Y_j - \frac{\alpha_j}{\beta}W|X\right) = \mathbb{E}\left(Y_j - \frac{\alpha_j}{\beta}W|X, T = j\right). \tag{4}$$

A first interpretation of (4) is obtained by remarking that we have, equivalently:

$$\mathbb{E}(Y_j|X, T = j) - \mathbb{E}(Y_j|X) = \frac{\alpha_j}{\beta}\left[\mathbb{E}(W|X, T = j) - \mathbb{E}(W|X)\right]. \tag{5}$$

The term on the left-hand side of (5) can be interpreted as one of the two parts of the bias of the matching estimand, for given covariates X. If the selection-on-observables assumption holds, then this bias term is zero and the ATE parameter is equal to the matching estimand. If not, then the bias is proportional to the difference in *pre*-treatment outcomes W between the treated and non-treated. Hence, under the model's assumptions, "falsification tests" à la Manning and Pischke (2006) are useful, as they indicate if matching on X will provide a consistent estimate of the ATE parameter.

A second interpretation of (4) is related to the panel data literature. Indeed, our approach can be seen as a "quasi-differencing" approach, close to the method used by Holz-Eakin *et al.* (1988) to estimate panel data models with interactive fixed effects of the form:

$$y_{it} = x_{it}'\theta + \alpha_t \eta_i + u_{it},$$

where α_t are time-varying parameters.

Our approach requires the identification of the ratios α_j/β, for $j = 0, 1$. The necessary information for that purpose is provided by the availability of other pre-treatment outcomes Z ("instruments"). We show in the Appendix that, if the model's assumptions hold then this ratio is given by:

$$\frac{\alpha_j}{\beta} = \frac{\mathbb{E}\left[\mathrm{Cov}\left(Z, Y|X, T=j\right)|T=j\right]}{\mathbb{E}\left[\mathrm{Cov}\left(Z, W|X, T=j\right)|T=j\right]}, \tag{6}$$

$$= \frac{\mathrm{Cov}\left[Z - \mathbb{E}\left(Z|X, T=j\right), Y - \mathbb{E}\left(Y|X, T=j\right)|T=j\right]}{\mathrm{Cov}\left[Z - \mathbb{E}\left(Z|X, T=j\right), W - \mathbb{E}\left(W|X, T=j\right)|T=j\right]}. \tag{7}$$

It follows from equation (6), or alternatively (7), that the weight of the endowment in post-treatment relative to pre-treatment outcomes can be consistently estimated from the data. It can be interpreted as the IV estimand in the quasi-differenced equation. A first implication of this result is that the ratio α_1/α_0 can also be estimated from the data. This is especially interesting in the case of selective schooling, as the literature has shown interest in the different returns to the child's ability in the two education systems.

A second implication of the identification of α_j/β concerns the estimation of policy parameters. Indeed, from (4) the ATE on $Y - \frac{\alpha_j}{\beta}W$ can be estimated by matching. It follows that the ATE on Y is also identified, as:

$$\Delta = \mathbb{E}\left[\mathbb{E}\left(Y - \frac{\alpha_1}{\beta}W|X, T=1\right) - \mathbb{E}\left(Y - \frac{\alpha_0}{\beta}W|X, T=0\right)\right] + \frac{\alpha_1 - \alpha_0}{\beta}\mathbb{E}\left(W\right). \tag{8}$$

See the Appendix for a straightforward derivation of (8).

Extending this approach, we can also identify the Average Treatment Effect for the Treated (ATT): $\Delta^{TT} = \mathbb{E}(Y_1 - Y_0|T=1)$. We show in the Appendix that the following identity holds:

$$\Delta^{TT} = \mathbb{E}[Y|T=1] - \mathbb{E}\left[\mathbb{E}\left(Y - \frac{\alpha_0}{\beta}W|X, T=0\right)|T=1\right] - \frac{\alpha_0}{\beta}\mathbb{E}(W|T=1). \quad (9)$$

In Section 6 we will be interested in the AT on the Non Treated (ATNT), that is obtained in a similar way.

A last parameter we are interested in is the ATE on variances:

$$\Delta^V = \mathrm{Var}(Y_1) - \mathrm{Var}(Y_0).$$

Comparing the variances of (potential) outcomes in the two schooling systems is interesting, as it reflects within-system dispersion and inequality in educational performance. In order to achieve identification we augment Assumptions 1 and 2 with the two following assumptions.

Assumption 6 *(selection on unobservables, variances)*

$$\mathrm{Var}(\varepsilon_0|T=1, X) = \mathrm{Var}(\varepsilon_0|X); \quad \mathrm{Var}(\varepsilon_1|T=1, X) = \mathrm{Var}(\varepsilon_1|X).$$

Assumption 7 *(pre-treatment outcomes, variances)*

$$\mathrm{Var}(U|T=1, X) = \mathrm{Var}(U|X).$$

Assumptions 6-7 rule out effects of the treatment on the variances of pre- and post-treatment outcomes. Under Assumptions 1-7 we show in Appendix that

$$\mathrm{Var}(Y|T=j, X) - \mathrm{Var}(Y_j|X) = \left(\frac{\alpha_1}{\beta}\right)^2 [\mathrm{Var}(W|T=j, X) - \mathrm{Var}(W|X)]. \quad (10)$$

Here also, as in equation (5), the bias of the matching estimand (on the variance) of the post-treatment outcome is proportional to the difference in pre-treatment variable between the two groups. We thus obtain the within-X variance. The between-X variance is easily obtained using the results for the ATE case. Likewise we can also derive the expression of the ATT on variances $\Delta^V = \mathrm{Var}(Y_1|T=1) - \mathrm{Var}(Y_0|T=1)$. See the Appendix for details.

In this paper, we shall focus on the above parameters: ATE and ATT on means and variances. One could be interested in many other policy-relevant effects of selective schooling. Their identification would require more assumptions, such as ones of conditional independence between the residuals. Carneiro *et al.* (2003) show how to identify Marginal Treatment Effects using a flexible independent factor analytic structure for the unobserved variables.

Multidimensional endowment. Allowing for multivariate η does not complicate the analysis very much. To simplify the presentation, we select K linearly independent rows of β and K linearly independent rows of γ. This is possible as the two matrices have rank K. With abuse of notation, we still call the resulting vectors of pre-treatment outcomes W and instruments Z as W and Z, respectively, and similarly still denote the matrices of coefficients as β and γ. Note that these are now $K \times K$ non singular matrices.

With this notation we show in the Appendix that the ratio of the return to endowment in post-treatment and pre-treatment outcomes, respectively, is given by:

$$\begin{aligned} \alpha_1\beta^{-1} &= \mathbb{E}\left[\mathrm{Cov}\left(Y, Z|X, T=j\right)\right]\left\{\mathbb{E}\left[\mathrm{Cov}\left(W, \boldsymbol{Z}|X, T=j\right)\right]\right\}^{-1}, \qquad (11)\\ &= \mathrm{Cov}\left[Y - \mathbb{E}\left(Y|X, T=j\right), Z - \mathbb{E}\left(Z|X, T=j\right)|T=j\right] \\ &\qquad \times \left\{\mathrm{Cov}\left[W - \mathbb{E}\left(W|X, T=j\right), Z - \mathbb{E}\left(Z|X, T=j\right)|T=j\right]\right\}^{-1}. \qquad (12) \end{aligned}$$

Quasi-differencing approach yields:

$$\mathbb{E}\left(Y_j - \alpha_j\beta^{-1}W|X\right) = \mathbb{E}\left(Y_j - \alpha_j\beta^{-1}W|X, T=j\right), \qquad (13)$$

and we immediately obtain the identification of the ATE parameter, as well as the ATT and the ATE on variances. An important difference with the univariate endowment case, however, is that the relative returns to the various components of η for the treated and non treated children (α_j/β in the univariate case) are not identified if $K > 1$.

Practical issues. Given a i.i.d. sample $\{Y_i, W_i, Z_i, X_i\}$, $i = 1, ..., N$, we proceed in two steps for the estimation of the ATE, ATT and ATE on variances parameters. In the first step, we estimate the ratio α_1/β by replacing expectations by empirical means in (7). This is straightforward to do if there are few discrete covariates. In our model of schooling, we want to condition on many covariates, discrete and continuous, as parental and school characteristics. One option is to adopt a flexible form for the conditional expectations, using for instance series estimators. We go half way in that direction, and adopt a linear specification. In practice, this amounts to first computing residuals of Y, Z and W on covariates X, and then estimating the covariances of these residuals, for the treated or non treated individuals. Including interaction terms had little effect on the results, suggesting that the linear specification is somewhat satisfying.

Note that in the estimation we have as many overidentifying restrictions as instruments. We use these restrictions for specification testing. Let τ be the vector of two-by-two equalities of the ratios given by (7). Under the model's assumptions τ is zero. The statistic we consider is $\widehat{\tau}'\left(\mathrm{diag}\,\widehat{\Omega}\right)^{-1}\widehat{\tau}$, where $\widehat{\tau}$ is an analog estimator of τ and $\widehat{\Omega}$

is a consistent estimate of the variance-covariance matrix of $\widehat{\tau}$. We take the diagonal as in Altonji and Segall (1996). We estimate $\widehat{\Omega}$ by nonparametric bootstrap. We compute critical values from the bootstrap distribution of $\widehat{\tau}'\left(\operatorname{diag}\widehat{\Omega}\right)^{-1}\widehat{\tau}$.

We also use the overidentifying restrictions in order to improve the efficiency of the estimator. To do so, we weight the various estimates of α_j/β by the inverse of their bootstrap variance.

Then, given first stage estimates, say $\frac{\widehat{\alpha}_1}{\widehat{\beta}}$ and $\frac{\widehat{\alpha}_2}{\widehat{\beta}}$, we estimate the ATE by matching on the propensity score. We use the inverse probability weighting of Hirano *et al.* (2003), and remark that:

$$\Delta \;=\; \mathbb{E}\left[\frac{T\left(Y-\frac{\alpha_1}{\beta}W\right)}{\pi(X)}-\frac{(1-T)\left(Y-\frac{\alpha_0}{\beta}W\right)}{1-\pi(X)}\right]+\frac{\alpha_1-\alpha_0}{\beta}\mathbb{E}\left(W\right), \tag{14}$$

where $\pi(X)=\mathrm{P}\left(T=1|X\right)$ is the propensity score with respect to X. Note that it is not possible to match on $\pi(X)$ directly to estimate the ATE on levels Y. For this, knowledge of the propensity score with respect to X and η would be necessary. However, matching on $\pi(X)$ is feasible to estimate the ATE on quasi-differences $Y_j-(\alpha_j/\beta)W$.

We plug our estimates $\frac{\widehat{\alpha}_1}{\widehat{\beta}}$, $\frac{\widehat{\alpha}_2}{\widehat{\beta}}$ and $\widehat{\pi}(X)$ into (14), and replace the (unconditional) expectations by empirical means. To estimate the propensity score we use Logit ML, which can be seen as a first approximation to the series Logit estimator proposed by Hirano *et al.* (2003). Again, including interaction terms had little effect on the results. Lastly, note that, for (14) to be valid, we need that $\pi(X)$ be strictly comprised between 0 and 1. In practice, we compute the expectation ouside of the tails of the propensity score distribution, selecting the portion between the 5^{th} and 95^{th} percentiles.[4]

We proceed similarly to estimate the ATT parameter. Estimation of the ATE on variances is slightly different. Details on the estimation, including the case where η is multidimensional, are provided in the Appendix. Lastly, standard errors are computed by 500 bootstrap iterations.

5 Mean and variance of educational outcomes

We present in Table 5 the correlation matrix of the various test scores at age 7 and 11, for children attending selective and comprehensive schools. The estimates are all significant, and remain so when we include covariates (not shown). This suggests that

[4] We also restricted the estimation to the "thick support" (percentiles 33^{rd} and 67^{th}) as an informal check of robustness of the effect to the presence of other unobservables (see Black and Smith, 2004). Doing so had some effects on the results, but did not modify the qualitative conclusions below.

	Math 11	Read 11	Verbal 11	Math 7	Read 7
Comprehensive					
Math 11	1.00				
Read 11	.69	1.00			
Verbal 11	.74	.71	1.00		
Math 7	.52	.41	.45	1.00	
Read 7	.55	.56	.62	.49	1.00
Selective					
Math 11	1.00				
Read 11	.74	1.00			
Verbal 11	.78	.74	1.00		
Math 7	.55	.46	.48	1.00	
Read 7	.60	.60	.66	.48	1.00

Table 5: Correlation matrices of pre-treatment outcomes, by schooling system.

Assumption 5, which requires intuitively that the pre-treatment variables have at least one observable in common, is satisfied. Moreover, correlations between test scores are higher in selective schools, with the exception of the age 7 math and reading scores. Most of these differences are significant at 5%, and also robust to the inclusion of covariates. This is consistent with children at selective schools having a greater initial endowment, which in turn affects positively all educational outcomes.

In Table 6 we report the ATE estimates of the effect of attending a comprehensive school on the five outcomes of interest: the math and reading test scores at 16, age when the child left school, and the nuber of O and A-levels obtained. We present the original ATE estimate and then the corrected estimate with its standard error. The corrected estimate is obtained by the quasi-differencing method assuming a single scalar endowment η. As a pre-treatment variable W, we have chosen the age 11 math score for all outcomes but the age 16 reading score outcome. For the latter, we have chosen the age 11 reading score. The variables Z used as "instruments" are all the other pre-treatment outcomes. The estimates of α_j/β were weighted in inverse proportion to their bootstrap variance. For the uncorrected and corrected estimates, the three columns correspond to (1) no covariates, (2) parental controls, and (3) a large set of parental, school and local controls.

We find that accounting for selection on unobservables matters, as it leads to a nearly ten-fold reduction of the ATE estimates. The corrected ATE is still negative in four cases out of five. The effect is insignificant for the reading score and the number of A-levels, and marginally significant for the math score. The effect on years of education is more substantial, as attending a selective school is associated to an increase of more than half a year of schooling. Lastly, the effect on the number of O-levels obtained is positive and insignificant.

	ATE			ATE (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−2.872 (.203)	−2.222 (.216)	−2.388 (.239)	−.297 (.140)	−.270 (.147)	−.269 (.159)
Read 16	−2.633 (.259)	−1.970 (.248)	−2.116 (.322)	−.142 (.163)	−.125 (.134)	−.124 (.170)
Years educ.	−1.242 (.304)	−.839 (.299)	−.990 (.402)	−.732 (.255)	−.517 (.261)	−.682 (.358)
# O-Levels	−1.015 (.115)	−.643 (.104)	−.763 (.104)	.141 (.080)	.146 (.085)	.040 (.090)
# A-Levels	−.420 (.053)	−.255 (.043)	−.290 (.050)	−.036 (.042)	−.013 (.041)	−.026 (.046)

Table 6: ATE and bias corrected ATE for various specifications of covariates; one-dimensional η.

In Table 7 we report the estimates of the returns to the endowment η, in the post-treatment relative to the pre-treatment outcomes. The variables Z used as "instruments" are ordered as in Table 5. For instance, the first figure in the first column of Table 7 refers to the age 11 math score used as W and the the age 11 reading score used as Z. The fourth figure in the same column corresponds to W being the age 11 math, and Z being the age 7 reading test scores. In the upper part of the table, we present the estimates of α_1/β, the relative return in the comprehensive system. In the lower part we report α_0/β, the relative return in the selective one. We also report the weighted means of the estimates, that we used in order to estimate the corrected ATE in Table 6. Lastly, the included covariates correspond to specification (3) of Table 6.

Several interesting insights can be obtained from Table 7. First, the magnitude of the effects is informative. The endowment at age 11, η, has a lower effect on later outcomes than on the age 11 test score. The reading test score at 16 is an exception, as the coefficient is about one, consistently with the fact that the reading tests at 11 and 16 are identical. Finding coefficients very different from one for the other outcomes motivates our quasi-differencing approach, which allows for different returns for various test scores. Second, for the math score and the years of education, we find some evidence that the return to η is higher in the selective system. However, the differences are only weakly significant. For instance, the p-value of the difference of the weighted means estimates is .18 for the math score, and .20 for the years of education measure. For the reading

	Math 16	Read 16	Years educ.	# O-levels	# A-levels
Comprehensive	.536 (.014)	1.002 (.026)	.090 (.006)	.245 (.010)	.060 (.005)
	.509 (.012)	1.032 (.025)	.070 (.005)	.198 (.010)	.043 (.004)
	.542 (.019)	1.162 (.042)	.080 (.009)	.206 (.014)	.049 (.007)
	.502 (.018)	1.087 (.033)	.068 (.007)	.206 (.013)	.034 (.004)
weighted mean	.520 (.013)	1.049 (.015)	.076 (.006)	.213 (.009)	.046 (.004)
Selective	.546 (.017)	.947 (.024)	.096 (.008)	.235 (.007)	.056 (.004)
	.535 (.143)	.990 (.024)	.077 (.006)	.215 (.006)	.043 (.003)
	.548 (.023)	.982 (.053)	.082 (.009)	.200 (.011)	.051 (.006)
	.539 (.020)	.992 (.041)	.079 (.008)	.219 (.009)	.032 (.004)
weighted mean	.541 (.012)	.973 (.019)	.083 (.005)	.219 (.006)	.044 (.003)

Table 7: Relative returns to the endowment; one-dimensional η; specification (3) of Table 6.

score, the differences are in the opposite direction, and significantly so. In any cases the differences are small. This result challenges the intuition that there be a higher reward to "ability" in selective schools, if we are ready to interpret the unobservable η as ability. At the same time, this result is in line with the results of Table 6, which show only few differences between the two schooling systems.

	Math 16	Read 16	Years educ.	# O levels	# A levels
One-dimensional η					
Comprehensive	.07	.01	.01	.03	.00
Selective	.77	.08	.01	.00	.00
Two-dimensional η					
Comprehensive	.93	.14	.98	.45	.32
Selective	.54	.93	.67	.74	.45

Table 8: pvalues of the test of overidentifying restrictions, $K = 1$ and 2; specification (3) of Table 6.

Under the null that the model's assumption are satisfied for one single η, the estimands obtained using different variables as instruments should be equal. Indeed, the estimates reported in Table 7 are of the same order of magnitude, suggesting that a single-η structure is a reasonable assumption, less so in the case of years of education and the number of A-levels obtained. To check the statistical validity of this assumption, we then report in Table 8 the p-values of the tests that all returns are equal. In the first two rows, equalities of the estimates α_1/β (comprehensive) and α_0/β (selective), for different variables used as "instruments", are tested. Included covariates are the same as in the previous table. At conventional levels, the tests do not reject the equality in the case of the age 16 math score, but they do reject it for the four other outcomes. We interpret this rejection as evidence that, though not a bad first-order approximation (see Table 7), a uni-dimensional η fails to capture a statistically significant part of the endowments of children. We then investigate if a bi-dimensional η would do better. The two last rows of Table 8 shows that it does, with much higher p-values.

For this reason, we then report the results using the quasi-differencing approach able to deal with a bi-dimensional η. In Table 9, we report the corrected ATE estimates, where selection on a bivariate unobservable is accounted for. The results indicate a very similar picture as in table (6). For all outcomes but the number of O-levels, attending a comprehensive school is associated with slightly lower later outcomes. The effects are very smilar in magnitude to those in the case of one single η. Only the significance of the results is marginally affected.

	ATE (corrected)		
	(1)	(2)	(3)
Math 16	−.273 (.169)	−.239 (.173)	−.269 (.159)
Read 16	−.056 (.144)	−.038 (.137)	−.020 (.188)
Years educ.	−.726 (.329)	−.491 (.317)	−.660 (.378)
# O-Levels	.126 (.091)	.135 (.082)	.015 (.087)
# A-Levels	−.163 (.063)	−.042 (.043)	−.079 (.071)

Table 9: Bias corrected ATE for various specifications of covariates; bi-dimensional η.

We then report the estimates of the ATE on variances. The results we obtain in Table 10 are much more clear-cut than in the case of the ATE on means. We find that, even controlling for a bivariate unobserved endowment, selective schools are associated with a larger variance. This is true in the case of the math score, for which the difference in variance between the two systems is about 10 percent. This is even more true in the case of later educational outcomes, where the difference in variances is about 25 percent. The result for the reading score is surprising, showing a difference in variance of 10 percent in the opposite direction.

The results in Table 9 and 10 suggest that comprehensive shools do differ from se-

	ATE on variances			ATE on variances (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−12.959	−11.321	−11.126	−4.722	−3.819	−4.084
	(1.436)	(1.645)	(1.551)	(1.564)	(1.664)	(1.516)
Read 16	1.824	.201	−.363	7.179	4.685	4.381
	(1.646)	(1.851)	(2.202)	(1.611)	(2.055)	(2.439)
Years educ.	−1.651	−1.141	−1.264	−1.395	−.936	−.994
	(.283)	(.299)	(.328)	(.296)	(.318)	(.341)
# O-Levels	−2.299	−1.560	−2.038	−1.170	−.532	−.807
	(.2834)	(.339)	(.414)	(.299)	(.390)	(.360)
# A-Levels	−.474	−.378	−.461	−.415	−.313	−.386
	(.085)	(.088)	(.112)	(.084)	(.092)	(.108)

Table 10: ATE and bias corrected ATE on variances for various specifications of covariates; bi-dimensional η.

lective schools. Moreover, differences appear not to be so much related to the average achievement of children, but to their dispersion. As an implication, children at selective schools have slightly better educational results, but they face more inequality. We view this result as a candidate to explain the push towards comprehensivisation. In the next section, we shall investigate this issue more thoroughly.

We also performed a series of robustness checks. First, restricting the sample to children that stayed in the same school during the five years yielded very similar results. Then, we estimated the effects of attending a comprehensive school, restricting the sample to purely selective and purely comprehensive LEA's, as proposed by Manning and Pischke (2006). Indeed, if for example grammar schools coexist alongside comprehensive schools, they might attract the best students (the "cream-skimming" effect). By excluding not fully selective LEA's, we might expect to lower the differences in the age 11 test scores. In that smaller sample (638 children in the comprehensive system, 729 in the selective one), we were unable to allow for the presence of a bi-dimensional unobserved endowment, as we obtained very imprecise estimates. Using one factor, the results are in line with the previous ones, though less precisely estimated. Lastly, we estimated the effect of attending a comprehensive school for a sample of children that were directly affected by the comprehensivisation reform. We found a somewhat more negative effect, suggesting that there might have been a short-run impact of the reform (caused *e.g.* by reorganization costs at the school level). These additional results are available from the authors.

6 The two sides of selective schooling (preliminary)

Selective schools can be of two types, either grammar or secondary modern. By construction, these two types of schools are very different in many respects. In this section, we modify the previous analysis in order to account for these differences. This part is under revision. We are currently working on a more tightly parameterized model, that should allow us to measure the risk faced by parents in the two schooling systems.

In Table 11, we report the mean outcomes for children attending grammar and secondary modern schools. We see that differences within the selective system are much stronger than between the selective and comprehensive systems. Indeed, children at grammar school score on average 10.2 points more than the ones at secondary schools at mathematics at age 16. Differences previous to entry are also enormous, as children at selective schools score on average 25.2 points more at maths at age 11.

	Grammar			**Secondary Modern**		
Variable	**Mean**	**Std.Dev.**	**N**	**Mean**	**Std. Dev.**	**N**
Mathematics score age 16	20.5	5.3	1113	10.3	5.2	2061
Reading score age 16	31.2	3.0	1115	24.2	6.3	2070
Age left full-time education	18.8	2.5	836	16.6	1.4	1556
Number of O-levels obtained	6.42	2.7	783	1.74	2.5	1453
Number of A-levels obtained	1.42	1.4	757	.26	.73	714
Mathematics score age 11	28.5	6.5	1141	13.3	8.2	2179
Reading score age 11	22.1	4.4	1141	14.4	5.1	2180
Verbal score age 11	31.8	4.9	1141	20.1	8.2	2180
Mathematics score age 7	6.9	2.1	1062	4.7	2.3	1979
Reading score age 7	28.5	2.4	1066	22.2	6.7	1992
Log Hourly real gross wage 1981	.651	.30	399	.549	.33	806
Log Hourly real gross wage 1991	1.12	.43	681	.835	.45	1113
Log Hourly real gross wage 2000	1.39	.51	406	1.14	.50	621

Table 11: Comparing outcomes in grammar and secondary modern schools

Consider the following modification of Model (1), where the potential outcome in the selective system is replaced by two potential outcomes:

$$\begin{cases} Y_{0G} &= f_{0G}(X) + \alpha_{0G}\eta + \varepsilon_{0G}, \\ Y_{0S} &= f_{0S}(X) + \alpha_{0S}\eta + \varepsilon_{0S}. \end{cases}$$

In the model, Y_{0G} is realized if the child belongs to the selective system ($T = 0$), and if she is assigned to a grammar school (that we denote as $G = 1$). In turn, Y_{0S} is realized if $T = 0$ and the child is assigned to a secondary modern school ($G = 0$). The assumptions of Model (1) are readily extended to this model.

In the context where grammar and secondary modern schools are different, the treatment effects calculated in the previous section have no simple interpretation. Indeed, it is difficult to give a precise meaning to Y_0, the potential outcome in the selective system. One possibility would be to define $Y_0 = G^* Y_{0G} + (1 - G^*) Y_{0S}$, and define G^* as a latent assignment to a grammar school for children attending the comprehensive system, while $G^* = G$ for the ones in the selective system. However, taking this route does not appear straightforward given our lack of data on pupils' ability and the precise nature of the assignment process.

Instead, we report in Tables 12-13 the estimates of the treatment effects of the fact of attending a comprehensive school on the outcomes, calculated for children attending a selective school. The effects reported in Table 12 are formally Average Treatment effects for the Non Treated (ATNT):

$$\Delta^{TNT} = \mathbb{E}(Y_1 - Y | T = 0).$$

The effects given in Table 13 are ATNT on the variances:

$$\Delta^{V,TNT} = \mathrm{Var}(Y_1 | T = 0) - \mathrm{Var}(Y | T = 0).$$

	ATNT			**ATNT (corrected)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−2.625	−2.345	−2.392	−.314	−.276	−.250
	(.202)	(.187)	(.208)	(.190)	(.181)	(.217)
Read 16	−2.188	−1.851	−1.861	.229	.262	.388
	(.202)	(.189)	(.214)	(.143)	(.252)	(.183)
Years educ.	−.872	−.664	−.667	−.481	−.190	−.195
	(.174)	(.175)	(.205)	(.179)	(.198)	(.293)
# O-Levels	−.932	−.752	−.801	−.007	.083	−.028
	(.095)	(.010)	(.105)	(.596)	(.564)	(.517)
# A-Levels	−.339	.047	−.283	−.279	−.158	−.211
	(.048)	(−.258)	(.044)	(.259)	(.256)	(.299)

Table 12: ATNT and bias corrected ATNT for various specifications of covariates; bidimensional η.

The results strengthen the ones obtained in Tables 9 and 10. For the math score and years of education, we find negative effects on mean outcomes. These effects are smaller than in Table 9 and become insignificant when controlling for parental characteristics. However, effects on variances are still large and significant. The effects on the reading score are again not in line with the ones for the two other outcomes.

	ATNT on variances			ATNT on variances (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−12.979	−11.173	−10.819	−4.926	−3.818	−3.613
	(1.337)	(1.669)	(1.625)	(1.883)	(1.774)	(2.197)
Read 16	1.809	.498	.281	8.257	5.921	5.975
	(1.804)	(1.752)	(1.859)	(2.079)	(1.921)	(1.952)
Years educ.	−1.654	−1.196	−1.165	−1.412	−.981	−.922
	(.322)	(.309)	(.344)	(.345)	(.335)	(.381)
# O-Levels	−2.302	−1.761	−1.835	−1.048	−.515	−.621
	(.350)	(.329)	(.342)	(.387)	(.418)	(.516)
# A-Levels	−.475	−.450	−.449	−.419	−.357	−.341
	(.087)	(.066)	(.108)	(.096)	(.101)	(.133)

Table 13: ATNT and corrected ATNT on variances for various specifications of covariates; bi-dimensional η.

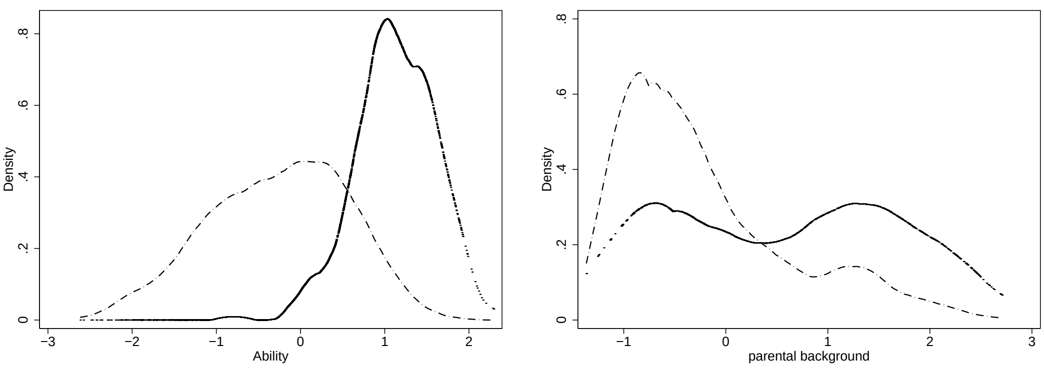

Figure 2: The "ability" and "parental background" principal components for pupils attending secondary modern (dash-dotted line) and grammar schools (solid line).

As in the previous section, the strongest result we obtain is the fact that selective schools show more dispersion in educational outcomes than comprehensive ones. The effect on means, when significant, is of a smaller order of magnitude. In the rest of this section, we intend to better understand the sources of this greater dispersion.

One source of the higher variance shown by the selective sytem is the variance between grammar and secondary schools. To investigate if these differences are robust to the inclusion of covariates we would ideally like to estimate the ATE parameter:

$$\Delta_{GS} = \mathbb{E}\left(Y_{0G} - Y_{0S}|T = 0\right).$$

Still, this overall average is meaningless as, by construction, there are very few children in grammar schools with a low η, say in the lower quartile of the distribution. The low-η children in secondary modern schools have no counterparts in grammar schools to whom they could be compared. Note that the lack of overlap occurs because of the unobserved regressor η, for which we have no direct measure. In Figure 2 we draw the densities of the “ability” and “parental background” components in grammar and secondary modern schools. The left panel, which represents “ability”, illustrates what is likely to be the case of the distribution of η, showing little room for comparison.

For this reason, we choose to compute averages on children that have an intermediate level of the “ability” factor, between the percentiles .40 and .80. These children have significant probabilities of attending the two school types. As there is some evidence that the scalar endowment assumption is incorrect (see Table 22 in Appendix), we report in Table 14 the corrected ATE estimate, which allows for a bi-dimensional η. We find strong positive effects of the fact of attending a grammar school on four outcomes out of five. When corrected by the quasi-differencing method allowing for one scalar endowment, the effects are much reduced but still strongly significant for the math and reading scores and for the number of O-levels passed. When allowing for covariates, the effect on the years of education is insignificantly different from zero. Lastly, the effect on the number of A-levels obtained is negative and insignificant.

These results shed some light on the differences between selective and comprehensive secondary schooling. Comparing the orders of magnitude in Tables 12 and 14 shows that there are few differences in means between the two systems, but there are strong differences between the two school types of the selective system. This has interesting implications for the understanding of the schooling reform in Britain. As any test, the *11-plus* exam involves several factors: the ability of the child, her specific preparation and, importantly, a certain amount of luck. By comparing children with similar ability who

	ATE			ATE (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	9.414	7.887	7.599	5.570	3.023	3.112
	(.210)	(.373)	(.373)	(1.684)	(1.321)	(1.409)
Read 16	10.136	7.725	7.382	8.682	5.386	5.371
	(.268)	(.453)	(.517)	(1.497)	(1.140)	(1.338)
Years of education	4.511	3.019	2.249	3.634	2.095	1.457
	(.415)	(.507)	(.478)	(1.101)	(1.207)	(1.134)
# O-Levels	3.731	2.710	2.526	2.741	1.743	1.636
	(.134)	(.144)	(.148)	(.551)	(.690)	(.537)
# A-Levels	1.028	.609	.590	−.002	−.250	−.163
	(.116)	(.121)	(.141)	(1.485)	(1.285)	(.834)

Table 14: Bias corrected ATE of attending a grammar rather than a secondary modern school, for various specifications of covariates; bi-dimensional η.

ended up at a grammar or a secondary modern school, we measure the "premium" that is associated with being sucessful at the exam. The estimates we obtain suggest that this premium is large, and explains a part of the variance in outcomes in the selective system. The conclusions of this exercise are related to the findings of Maurin and McNally (2006), who show that relaxing the admission criteria to grammar schools in Northern Ireland has led to an increase in total educational attainment. However, to our knowledge no such reform was undertaken in Britain. From the perspective of parents, the variance between school types represents a risk associated with the probability of failing the *11-plus* exam. As we find small differences in means between the two schooling systems, it may have been rational for risk-averse parents to support the push towards comprehensivisation.

7 Conclusion

Several recent papers in the literature insist on the fact that it is difficult to separate the effect of school type on educational outcomes from other factors, most importantly the child's ability. Possible answers to this problem are to use "natural experiments" (Meghir and Palme, 2004, and Maurin and McNally, 2006), or to exploit cross-country and time variation (Hanushek and Woessman, 2006). The absence of such data motivates our quasi-differencing method, that uses the availability of various test scores previous to secondary schooling.

It is often thought that when choosing between selective and comprehensive education one faces a trade-off between efficiency and equity. We find little evidence of such a trade-off. The raw data from the NCDS, as well as regressions using a large set of covariates, suggest that children attending selective schools have better results on average. We show that this is mainly the result of composition effects in terms of unobservables, selective schools having a better intake. When we difference the unobservables out, the effect of attending a selective school is still positive, but small and most often insignificantly different from zero. In contrast, the difference in variance remains, at least for the maths and years of education outcomes. In addition, we give some evidence on the sources of the higher variance in the selective system, showing that there is a large premium to passing the admission exam and attending a grammar school.

APPENDIX

A Model identification

In its most general form, the model reads as follows:

$$\begin{cases} Y_0 &= f_0(X) + \boldsymbol{\alpha}_0'\boldsymbol{\eta} + \varepsilon_0, \\ Y_1 &= f_1(X) + \boldsymbol{\alpha}_1'\boldsymbol{\eta} + \varepsilon_1, \\ \boldsymbol{W} &= \boldsymbol{g}(X) + \boldsymbol{\beta\eta} + \boldsymbol{U}, \\ \boldsymbol{Z} &= \boldsymbol{h}(X) + \boldsymbol{\gamma\eta} + \boldsymbol{V}, \end{cases}$$

where $\boldsymbol{\eta}$ is a $K \times 1$ vector of endowments, $\boldsymbol{W}$ is a $L_1 \times 1$ vector of pre-treatment outcomes, $\boldsymbol{Z}$ is a $L_2 \times 1$ vector of instruments, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $L_1 \times K$ and $L_2 \times K$ matrices of parameters.

We require the following set of assumptions:

$$\mathbb{E}(\varepsilon_0|X,T) = \mathbb{E}(\varepsilon_0|X)\,; \quad \mathbb{E}(\varepsilon_1|X,T) = \mathbb{E}(\varepsilon_1|X)\,, \tag{A1}$$

$$\mathbb{E}(\boldsymbol{U}|X,T) = \mathbb{E}(\boldsymbol{U}|X)\,, \tag{A2}$$

$$\boldsymbol{\eta} \perp (\varepsilon_0, \varepsilon_1, \boldsymbol{U}, \boldsymbol{V}) \mid \quad X, T, \tag{A3}$$

$$\boldsymbol{V} \perp (\varepsilon_0, \varepsilon_1, \boldsymbol{U}) \mid \quad X, T, \tag{A4}$$

$$\text{Rank}\,(\boldsymbol{\beta}) = \text{Rank}\,(\boldsymbol{\gamma}) = \text{Rank}\,(\text{Var}(\eta|X,T)) = K, \tag{A5}$$

$$\text{Var}\,(\varepsilon_0|X,T) = \text{Var}\,(\varepsilon_0|X)\,; \quad \text{Var}\,(\varepsilon_1|X,T) = \text{Var}\,(\varepsilon_1|X)\,, \tag{A6}$$

$$\text{Var}\,(\boldsymbol{U}|X,T) = \text{Var}\,(\boldsymbol{U}|X)\,. \tag{A7}$$

We start by selecting K linearly independent rows of $\boldsymbol{\beta}$, and K linearly independent rows of $\boldsymbol{\gamma}$. For notational simplicity we still denote the subsets of K pre-treatment outcomes as $\boldsymbol{W}$, the matrix of factor loadings as $\boldsymbol{\beta}$, and similarly for the instruments. In the submodel, Assumptions (A1)-(A7) are satisfied, $L_1 = L_2 = K$, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are invertible.

We have, using (A3) and (A4):

$$\begin{aligned} \text{Cov}\,(Y, \boldsymbol{Z}|X, T=j) &= \boldsymbol{\alpha}_j'\,\text{Var}\,(\boldsymbol{\eta}|X,T=j)\,\boldsymbol{\gamma}', \\ \text{Cov}\,(\boldsymbol{W}, \boldsymbol{Z}|X, T=j) &= \boldsymbol{\beta}\,\text{Var}\,(\boldsymbol{\eta}|X,T=j)\,\boldsymbol{\gamma}'. \end{aligned}$$

Hence, making use of (A5) we obtain:

$$\begin{aligned} \boldsymbol{\alpha}_j'\boldsymbol{\beta}^{-1} &= \boldsymbol{\alpha}_j'\mathbb{E}[\text{Var}\,(\boldsymbol{\eta}|X,T=j)\,|T=j]\,\boldsymbol{\gamma}'\left\{\boldsymbol{\gamma}'\right\}^{-1}\{\mathbb{E}[\text{Var}\,(\boldsymbol{\eta}|X,T=j)\,|T=j]\}^{-1}\boldsymbol{\beta}^{-1}, \\ &= \boldsymbol{\alpha}_j'\mathbb{E}[\text{Var}\,(\boldsymbol{\eta}|X,T=j)\,|T=j]\,\boldsymbol{\gamma}'\left\{\boldsymbol{\beta}\mathbb{E}[\text{Var}\,(\boldsymbol{\eta}|X,T=j)\,|T=j]\,\boldsymbol{\gamma}'\right\}^{-1} \\ &= \mathbb{E}[\text{Cov}\,(Y,\boldsymbol{Z}|X,T=j)\,|T=j]\,\{\mathbb{E}[\text{Cov}\,(\boldsymbol{W},\boldsymbol{Z}|X,T=j)\,|T=j]\}^{-1}. \end{aligned} \tag{A8}$$

Then we also have, using (A1) and (A2):

$$\mathbb{E}\left(Y_j - \boldsymbol{\alpha}_j' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X\right) = \mathbb{E}\left(Y_j - \boldsymbol{\alpha}_j' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = j\right), \tag{A9}$$

from which is follows that:

$$\begin{aligned}
\Delta &= \mathbb{E}(Y_1 - Y_0) \\
&= \mathbb{E}\left(Y_1 - \boldsymbol{\alpha}_1' \boldsymbol{\beta}^{-1} \boldsymbol{W} - Y_0 + \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W}\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W}) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y_1 - \boldsymbol{\alpha}_1' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X\right) - \mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X\right)\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W}) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y_1 - \boldsymbol{\alpha}_1' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 1\right) - \mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 0\right)\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W}) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y - \boldsymbol{\alpha}_1' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 1\right) - \mathbb{E}\left(Y - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 0\right)\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W}),
\end{aligned} \tag{A10}$$

where we have used (A9) to show the fourth equality.

Similarly we have:

$$\begin{aligned}
\Delta^{TT} &= \mathbb{E}(Y_1 - Y_0 | T = 1) \\
&= \mathbb{E}(Y_1 | T = 1) - \mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | T = 1\right) - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W} | T = 1) \\
&= \mathbb{E}(Y_1 | T = 1) - \mathbb{E}\left(\mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 1\right) | T = 1\right) - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W} | T = 1) \\
&= \mathbb{E}(Y_1 | T = 1) - \mathbb{E}\left(\mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 0\right) | T = 1\right) - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W} | T = 1) \\
&= \mathbb{E}(Y | T = 1) - \mathbb{E}\left(\mathbb{E}\left(Y - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \boldsymbol{W} | X, T = 0\right) | T = 1\right) - \boldsymbol{\alpha}_0' \boldsymbol{\beta}^{-1} \mathbb{E}(\boldsymbol{W} | T = 1).
\end{aligned} \tag{A11}$$

The formula for Δ^{TNT} is obtained by interverting $T = 1$ and $T = 0$ in this expression.

Then, the ATE on variances is obtained as follows. Let

$$\Delta^V = \mathrm{Var}(Y_1) - \mathrm{Var}(Y_0).$$

Let $j = 1, 2$. We have:

$$\mathrm{Var}(Y_j) = \mathrm{Var}(\mathbb{E}(Y_j | X)) + \mathbb{E}(\mathrm{Var}(Y_j | X)).$$

We have, using (A3):

$$\mathrm{Var}(\boldsymbol{W} | X) = \boldsymbol{\beta} \,\mathrm{Var}(\boldsymbol{\eta} | X)\, \boldsymbol{\beta}' + \mathrm{Var}(\boldsymbol{U} | X),$$

and, using (A3) and (A2):

$$\mathrm{Var}(\boldsymbol{W} | X, T = j) = \boldsymbol{\beta} \,\mathrm{Var}(\boldsymbol{\eta} | X, T = j)\, \boldsymbol{\beta}' + \mathrm{Var}(\boldsymbol{U} | X).$$

We obtain:

$$\mathrm{Var}(\boldsymbol{W}|X,T=j)-\mathrm{Var}(\boldsymbol{W}|X) \;\; = \;\; \boldsymbol{\beta}\left[\mathrm{Var}\left(\boldsymbol{\eta}|X,T=j\right)-\mathrm{Var}\left(\boldsymbol{\eta}|X\right)\right]\boldsymbol{\beta}'.$$

Likewise, making use of (A1) we have:

$$\mathrm{Var}(Y_j|X,T=j)-\mathrm{Var}(Y_j|X) \;\; = \;\; \boldsymbol{\alpha}_j\left(\mathrm{Var}\left[\boldsymbol{\eta}|X,T=j\right)-\mathrm{Var}\left(\boldsymbol{\eta}|X\right)\right]\boldsymbol{\alpha}_j'.$$

It follows that:

$$\mathrm{Var}(Y_j|X,T=j)-\mathrm{Var}(Y_j|X) \;\; = \;\; \boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathrm{Var}(\boldsymbol{W}|X,T=j)-\mathrm{Var}(\boldsymbol{W}|X)\right]\left\{\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\right\}'.$$

Hence the within-X component of the variance:

$$\mathbb{E}(\mathrm{Var}(Y_j|X)) \;\; = \;\; \mathbb{E}\Big\{\mathrm{Var}(Y|T=j,X)-\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathrm{Var}(\boldsymbol{W}|X,T=j)-\mathrm{Var}(\boldsymbol{W}|X)\right]\left\{\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\right\}'\Big\}. \tag{A12}$$

Lastly, using (A9) we obtain the between-X part, as:

$$\mathrm{Var}\left(\mathbb{E}(Y_j|X)\right)=\mathrm{Var}\left\{\mathbb{E}(Y|X,T=j)-\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathbb{E}(\boldsymbol{W}|X,T=j)-\mathbb{E}(\boldsymbol{W}|X)\right]\right\}. \tag{A13}$$

The within-X and between-X components of $\mathrm{Var}\left(Y_0|T=1\right)$, that intervenes in $\Delta^{V,TT}$, are obtained similarly as:

$$\begin{aligned}\mathbb{E}(\mathrm{Var}(Y_0|T=1,X)|T=1) \;\; = \;\; & \mathbb{E}\Big\{\mathrm{Var}(Y|T=0,X)\\ & -\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathrm{Var}(\boldsymbol{W}|X,T=0)-\mathrm{Var}(\boldsymbol{W}|X,T=1)\right]\left\{\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\right\}'\Big|T=1\Big\},\end{aligned} \tag{A14}$$

and

$$\mathrm{Var}\left(\mathbb{E}(Y_0|X,T=1)\,|T=1\right)=\mathrm{Var}\left\{\mathbb{E}(Y|X,T=0)-\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathbb{E}(\boldsymbol{W}|X,T=0)-\mathbb{E}(\boldsymbol{W}|X,T=1)\right]\Big|T=1\right\}. \tag{A15}$$

B Parameter estimation

We rewrite (A8) as:

$$\begin{aligned}\boldsymbol{\alpha}_j'\boldsymbol{\beta}^{-1} \;\; = \;\; & \mathrm{Cov}\left[Y-\mathbb{E}(Y|X,T=j)\,,\boldsymbol{Z}-\mathbb{E}(\boldsymbol{Z}|X,T=j)\,|T=j\right]\times\\ & \left\{\mathrm{Cov}\left[\boldsymbol{W}-\mathbb{E}(\boldsymbol{W}|X,T=j)\,,\boldsymbol{Z}-\mathbb{E}(\boldsymbol{Z}|X,T=j)\,|T=j\right]\right\}^{-1}.\end{aligned} \tag{B16}$$

Then, we write the ATE and ATT estimands as:

$$\Delta \;\; = \;\; \mathbb{E}\left(\frac{T\left(Y-\boldsymbol{\alpha}_1'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right)}{\pi(X)}-\frac{(1-T)\left(Y-\boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right)}{1-\pi(X)}\right)+(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_0)'\boldsymbol{\beta}^{-1}\mathbb{E}(\boldsymbol{W}), \tag{B17}$$

and:

$$\Delta^{TT} = \mathbb{E}(Y|T=1) - \mathbb{E}\left(\frac{\pi(X)}{\mathrm{P}(T=1)}\frac{(1-T)\left(Y-\boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right)}{1-\pi(X)}\right) - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\mathbb{E}(\boldsymbol{W}|T=1). \tag{B18}$$

To derive these expressions, see Hirano *et al.* (2003).

Lastly, we estimate the ATE on variances on the basis of (A12) and (A13), replacing conditional expectations by linear projections. We proceed similarly for the ATT on variances.

C Data Appendix

C.1 Construction of samples

Samples We here explain how we constructed our sample, and how it relates to other samples used in the literature. It is useful to look at the raw data (see Table C.1). We only keep either comprehensive,grammar or secondary Modern pupils. All others (notably private schools) are dropped.

	Frequency	Percentage	Cumulative
Non Lea school[a]	6,287	38.78	38.78
Comprehensive	5,895	36.36	75.15
Grammar	1,294	7.98	83.13
Secondary modern	2,582	15.93	99.06
Technical	65	0.40	99.46
All-age	1	0.01	99.46
Day spec.non esn	1	0.01	99.47
Other	85	0.52	99.99
Res.esn school	1	0.01	100.00
Total	16,211	100.00	

[a]This includes also schools for which information is missing, which explains the large percentage of this school type

Table 15: The raw distribution of schools at age 16

A first possibility is to consider as comprehensive all observations for which the children attended a comprehensive school in 1974, and consider all other observations as selective. This is what Kerckhoff (1986) and Galindo-Rueda and Vignolles (2005) probably did to construct their sample. Manning and Pischke (2006) state that they do not intend to reproduce any specific dataset. However it is not clear how they constructed their data. We believe that they used some information from 1969 to subdivide their data.

The information in 1969 is given by three variables. One variable contains the response of the teacher to the question: "What type of school is the NCDS child likely to attend in the next year ?" The answer can be 1) non-LEA (Local Education Authority) school, 2) selective school, 3) maintained (comprehensive) school. Another variable in 1969 indicates the year a school went comprehensive. A third variable, reported in 1974, indicates for how long the child has been attending the same school.

Thus, from those variables one should be able to construct at least an indicator of how long one child had been in the comprehensive system, since we know the date his school switched to comprehensive. Thus we are theoretically able to distinguish if a child had been exposed 5 full years to the comprehensive education system or only some years (1-4 years) within the period

1969-1974. Unfortunately, these variables conflict in some instances, indicating either that the child switched school or that the teacher's assessment was wrong. The NCDS team told us that the teacher's answer is likely to be the more reliable source, since the other information comes from school questionaires. For this reason, we use the teacher's answer to construct our comprehensivisation variable.

We also construct a subsample keeping only schools in either purely selective LEA's or purely comprehensive LEA's. We use the list given by Manning and Pischke (2006) and select only those LEA's from the NCDS.

To make sure that our assignment rule is reasonable, we contrast the subjective opinion of the teacher with the date the school became comprehensive. Inconsistencies between this variable and the teacher's answer may reflect errors, or mobility between schools of different types. Reconstructing the two samples using the comprehensivisation date variable, we obtained two other samples. Raw statistics on these two samples are similar to the ones in our benchmark sample, and so are the estimation results. We also obtained qualitatively similar results using the number of years the child attended the school in order to classify observations as "comprehensive" and "selective".

Political variables Until now, we have not been allowed to use the full local information in the NCDS and we did not have access to local election results. Thus, we constructed our political controls in an indirect fashion. We would like to approximate the political colour of an LEA with information from constituency level election results.

There can be an overlap between sets of constituencies and LEA's in the following sense. LEA's are larger administrative units than constituencies, since they are bodies of the local authorities, which may regroup one or more constituencies. LEA's are defined as a collection of electoral wards and these electoral wards may or may not coincide with the boundaries of a set of constituencies. Furthermore constituency boundaries are subject to change in the period we are considering. We only have information on constituency level election results. We matched as closely as possible constituencies to LEA's, that is we checked with various sources, such as the Constituency boundary commission and local authority websites to establish which constituencies belonged to which LEA's in the end of the 1970's. A majority of constituencies can be aggregated into an LEA. Some constituencies have some electoral wards in one LEA and other wards in the other LEA because of redrafting of boundaries. In this case, we aggregate the results using the "redrafted constituency" results for both LEA's.

The NCDS has only confidential information on constituencies, so a first step was to assign each constituency an LEA identifier. Then we aggregated the results to the LEA level taking simple means over the constituencies forming the LEA.

Missing values Some of the variables we use as regressors suffer from a missing values problem. This is especially true for social class indicators. We checked roughly if the missing values are mising at random. For this, we replaced missing values with the sample means of the variables and run some of our regression equations. If the coefficients did not change drastically, we concluded that the missing values are at random. Doing this, we found very little change in our estimates. In a next step, we replaced the missing values by dummy variables, taking the value 0 if the value is not missing and 1 if it is missing. Using this method, we preserved most of our sample intact.

Different specifications used The specifications labelled (1), (2) and (3) used in the ATE/ATE variance tables are as follows:

- (1) no covariates
- (2) child female dummy, number of older siblings, father's and mother's years of education and the corresponding missing value indicators.
- (3) is (2) augmented by: pupil teacher ratio (1969), if child in junior school in 1969, if child gets free meals in 1969, sex of the headteacher (1969), Age of the main school buildings (1969), How many pupils stay on at the school after the minimun age (1965), Number of children on school roll (1965), the number of schools attended since 1965, the proportion of comprehensive schools in the area, dummies for an ability streamed class in 1969, The proportion of unemployed and sick in the census ED, the proportion of mining workers in the census ED, the proportion of working mothers (census ED), the proportion of skilled, semi-skilled, managerial and unskilled workers in census ED, the proportion of households with indoor WC in census ED, the proportion of immigrants in census ED, and the proportion of labour voters in the 1970's general election. Notice that we only include variables which intervene before or at the date of treatment.

C.2 List of variables used

Table 16: Family and child characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Hourly real gross wage 2000 (1981 prices)	1.215	0.507	-1.128	3.919	1536
Hourly real gross wage 1991 (1981 prices)	0.917	0.468	-1.394	5.146	2744
Hourly real gross wage 1981 (1981 prices)	0.572	0.321	-0.903	2.651	1888
Age when left education	17.09	1.994	14	28	3662
Number O-levels obtained	2.98	3.254	0	9	4167
Number A-levels obtained	.737	1.209	0	5	2573
Math 16	12.82	6.779	0	31	4855
Read 16	25.801	6.499	1	35	4875
Math 11	16.727	10.115	0	40	5122
Read 11	16.139	6.028	0	35	5122
Verbal 11	22.249	9.051	0	40	5123
Total score all syndroms BSAG	8.006	8.571	0	70	5132
Math7	5.226	2.405	0	10	4733
Read7	23.562	6.756	0	30	4749
Female dummy	0.5	0.5	0	1	5155
Father's completed education, years	3.932	1.73	1	11	4109
Mother's completed education, years	3.876	1.296	1	11	4179
Social Class 1958 I	0.125	0.331	0	1	4707
Social Class 1958 II	0.105	0.307	0	1	4707
Social Class 1958 III	0.529	0.499	0	1	4707
Social Class 1958 IV	0.124	0.33	0	1	4707
Social Class 1965 I	0.038	0.19	0	1	5155
Social Class 1965 II	0.117	0.321	0	1	5155
Social Class 1965 III	0.097	0.296	0	1	5155
Social Class 1965 IV	0.414	0.493	0	1	5155
Social Class 1965 V	0.015	0.123	0	1	5155
Social Class 1965 VI	0.142	0.349	0	1	5155
Social Class 1965 V	0.015	0.123	0	1	5155
Social Class 1969 I	0.038	0.191	0	1	4626
Social Class 1969 II	0.102	0.302	0	1	4626
Social Class 1969 III	0.004	0.064	0	1	4626
Social Class 1969 IV	0.039	0.194	0	1	4626
Social Class 1969 V	0.052	0.222	0	1	4626
Social Class 1969 VI	0.1	0.3	0	1	4626
Social Class 1969 VII	0.004	0.066	0	1	4626
Social Class 1969 VIII	0.064	0.245	0	1	4626
Social Class 1969 IX	0.333	0.471	0	1	4626
Social Class 1969 X	0.136	0.342	0	1	4626
Social Class 1969 XI	0.039	0.194	0	1	4626
Social Class 1969 XII	0.009	0.096	0	1	4626
Social Class 1969 XIII	0.01	0.099	0	1	4626
Social Class 1969 XIV	0.011	0.102	0	1	4626
Social Class 1969 XV	0.008	0.087	0	1	4626

Table 17: Family and child characteristics - continued

Variable	Mean	Std. Dev.	Min.	Max.	N
Total number of older siblings (girls/boys) of NCDS child	0.468	0.661	0	8	4232
Do you wish you had left school earlier, NCDS child answer	0.293	0.455	0	1	4204
Age you like to leave school, NCDS child answer	1.527	0.847	1	3	4502
Will child benefit staying on ? Teacher answer	0.423	0.494	0	1	5052
How long (in years) has child been at this school	4.572	0.945	0	9	4903
Total school attendance	87.701	15.721	0.704	100	4787
Child has a room to study	0.883	0.322	0	1	4836
Father's completed education, years	3.932	1.73	1	11	4109
Mother's completed education, years	3.876	1.296	1	11	4179
Natural father takes care of child in 1974	0.872	0.334	0	1	4245
Natural mother takes care of child in 1974	0.953	0.211	0	1	4245
Father born in British Isles	0.911	0.284	0	1	4632
Mother born overseas	0.026	0.159	0	1	4632
Mother born in British Isles	0.924	0.265	0	1	4632
Child has no father figure, 1965	0.029	0.168	0	1	4694
Child has no mother figure, 1965	0.003	0.05	0	1	4695
Natural mother in household, 1965	0.002	0.048	0	1	4695
Natural mother in household at age 7	0.013	0.111	0	1	4694
Persons per room in 1974	1.352	0.626	1	4	4201
Accomodation type in 1974	1.625	0.893	1	6	4231

Table 18: Parental characteristics and preferences

Variable	Mean	Std. Dev.	Min.	Max.	N
Parents' wish child left at 15	0.246	0.431	0	1	4130
Parents' wishes child leaves at 18	0.276	0.447	0	1	3957
Parents' wishes child completes further educ.	0.338	0.473	0	1	3957
Father's interest in child education	3.129	1.117	1	5	4365
Mother's interest in child education	3.003	1.049	1	5	4696
Mother's banded income, 1974	3.33	1.559	1	12	2516
Father's banded income, 1974	5.8	2.935	1	12	2764
Mother works since child at school, 1965	0.291	0.454	0	1	4635

Table 19: School characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Enrolled in a grammar school 1974	0.18	0.385	0	1	5155
Enrolled in a secondary modern school in 1974	0.343	0.475	0	1	5155
Years spent in comprehensive between 1969 and 1974	0.467	0.499	0	1	2455
Proportion of LEA comprehensive in 1974	63.628	25.347	8	99.900	4709
Comprehensive school formed out of a Grammar school	0.175	0.38	0	1	2439
Comprehensive school formed out of secondary modern	0.216	0.411	0	1	2439
Comprehensive school purpose built	0.248	0.432	0	1	2439
Comprehensive school formed by amalgamtion	0.326	0.469	0	1	2439
Child is not in an ability streamed class, 1969	0.646	0.478	0	1	5155
Child is in low ability streamed class, 1969	0.095	0.293	0	1	5155
Child is in average ability streamed class, 1969	0.105	0.306	0	1	5155
Child is in high ability streamed class, 1969	0.144	0.351	0	1	5155
Number of pupils on school roll, 1965	247.578	117.248	6	903	4654
Number of children in child's class, 1965	35.825	6.991	2	59	4763
Number of schools attended since age of 5, 1965	1.239	0.606	0	11	4679
Number of children on school roll, 1969	328.698	138.832	10	999	5000
In junior and infant school in 1965	0.416	0.493	0	1	5093
Free school meals at school, 1969	0.089	0.285	0	1	5082
Sex of headteacher, 1969	0.173	0.379	0	1	5137
Age of the main school buildings, 1969	47.021	36.701	0	193	5028

Table 20: Local area and political variables

Variable	Mean	Std. Dev.	Min.	Max.	N
RE161	0.13	0.336	0	1	5155
RE162	0.086	0.28	0	1	5155
RE163	0.1	0.3	0	1	5155
RE164	0.089	0.284	0	1	5155
RE165	0.096	0.294	0	1	5155
RE166	0.181	0.385	0	1	5155
RE167	0.066	0.249	0	1	5155
RE168	0.081	0.273	0	1	5155
RE169	0.105	0.307	0	1	5155
% Unemployed and sick, 1971 Census ED	4.836	5.79	0	42.85	2735
% Married women working, 1971 Census ED	40.152	17.18	0	80	2674
% Mining & Manuf. Employment, 1971 Census ED	39.18	18.975	0	100	2736
% Managment Employment, 1971 Census ED	12.409	12.33	0	78.570	2734
% Skilled manual Employment, 1971 Census ED	29.126	13.127	0	78.570	2733
% Semi-skilled manual Employment, 1971 Census ED	19.216	11.023	0	70	2734
Persons per room, 1971 Census ED	0.613	0.115	0.38	1.18	2736
Households lacking inside WC, 1971 Census ED	9.716	15.785	0	96.37	2736
% New commonwealth immigrants, 1971 Census ED	1.52	4.782	0	56.37	2734
Mean labour win in 1970 GE per constituency	0.306	0.082	0.101	0.509	4941

D Complementary tables and figures

	Dependent variable:	
	Maths score at 16	Maths score at 11
Comprehensive	-1.96468	-3.51956
	(1.37783)	(2.06817)
Ability (factor)	4.33572	
	(0.14757)	
Ability at 7 (factor)		5.18235
		(0.12565)
Female dummy	-1.37463	-0.94888
	(0.13770)	(0.22698)
Father's education	0.59102	0.60056
	(0.06022)	(0.09535)
Mother's education	0.61307	0.63362
	(0.06809)	(0.10910)
Cons	11.68595	11.78768
	(1.88529)	(1.12432)
R-squared	.607481	.4661144
N	4109	4381

Table 21: Instrumental variable regression using political colour of the LEA as instrument for the indicator of comprehensive schooling

	Math 16	Read 16	Years educ.
One-dimensional η			
Grammar	.22	.15	.10
S. Modern	.37	.84	.21
Two-dimensional η			
Grammar	.59	.82	.81
S. Modern	.94	.86	.91

Table 22: pvalues of the test of overidentifying restrictions, $K = 1$ and 2. Specification (3).

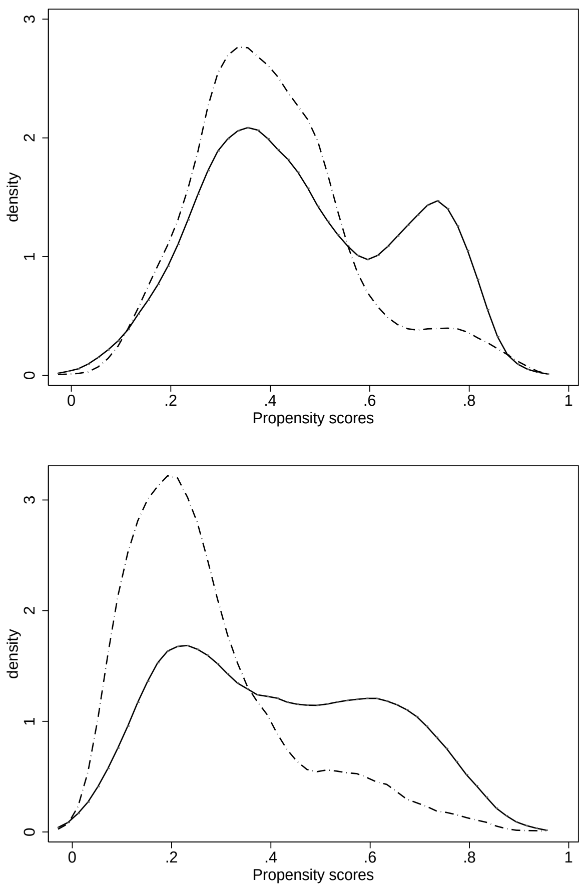

Figure 3: The estimated propensity scores of attending a comprehensive vs a selective school (up), and a grammar vs a secondary modern school (bottom). In both panels the density for the treated group is given in solid line.

References

[1] ALTONJI, J.F. and DUNN, T. (1996): "Using siblings to estimate the effects of school quality on wages ," *Review of Economics and Statistics*, 78, 665-671.

[2] ALTONJI, J.F., ELDER, T. and TABER, C. (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151-184.

[3] ALTONJI, J.F., and L.M. SEGAL (1996): "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14, 353-366.

[4] BLACK, D.A., and J.A. SMITH (2004): "How Robust in the Evidence on the Effect of College Quality? Evidence from Matching," *Journal of Econometrics*, 121, 99-124.

[5] BLANDEN, J., P. GREGG, and S. MACHIN (2003): "Changes in Educational Inequality," *CMPO Working Paper,* Nb.03/079.

[6] CARNEIRO P. , HANSEN, K. and HECKMAN, J. (2003):"Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44(2), 361-422.

[7] CUNHA, F. , and J. HECKMAN (2006): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation ," *Mimeo.*

[8] CUNHA, F. , J. HECKMAN, and S. SCHENNACH (2006): "Estimating the Technology of Cognitive and Noncognitive Skill Formation ," *Mimeo.*

[9] DEARDEN, L., J. FERRI, and C. MEGHIR (2002): "The Effect of School Quality on Educational Attainment and Wages," *Review of Economics and Statistics*, 84, 1-20.

[10] DEARDEN, L., S. MACHIN, and H. REED (1997): "Intergenerational Mobility in Britain," *Economic Journal*, 107, 47-66.

[11] DUSTMANN, C., N. RAJAH, and A. VAN SOEST (2003): "Class Size, Education, and Wages," *Economic Journal*, 113, F99-F120.

[12] FIGLIO, D.N. and PAGE, M.E. (2002): "School Choice and the Distributional Aspects of Ability Tracking: Does Separation Increase Equality ?," *Journal of Urban Economics*, 51(3), p. 497-514 .

[13] GALINDO-RUEDA, F., and A. VIGNOLES (2005): "The Heterogeneous Effect of Selection In Secondary Schools: Understanding the Changing Role of Ability," *Working Paper, Center for the Economics of Education.*

[14] HANSEN, K., J. HECKMAN, and K. MULLEN (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121, 39-98.

[15] HANUSHEK, E. A. and WOESSMAN, L. (2006): "Does Educational Tracking Affect Performance and Inequality ? Differences-in-Differences Evidence Across Countries," *Economic Journal*, 116, C36-C76.

[16] HARMON, C. and WALKER, I. (2000) "The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales", *Economica*, 67(265).

[17] HIRANO, K., IMBENS, G. and RIDDER, G. (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score ", *Econometrica*, 71(4), 1161.

[18] HOLTZ-EAKIN, D., W. NEWEY and H. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56(6), 1371-1395.

[19] IMBENS, G. (2003): "Sensitivity to Exogeneity Assumptions in Program Evaluation ," *American Economic Review*, *Papers and Proceedings*, 93(2), 126-132.

[20] IMBENS, G. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review ," *Review of Economics and Statistics*, *Papers and Proceedings*, 86, 4-30.

[21] ICHINO, A. , MEALLI, F. and NANNICINI, T. (2006): "From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? ," *IZA Discussion Paper n. 2149.*

[22] KERCKHOFF, Alan C. (1986): "Effects of Ability Grouping in British Secondary Schools " *American Sociological Review*, 51, 842-858.

[23] MANNING, A., and J.S. PISCHKE (2006): "Comprehensive versus Selective Schooling in England and Wales: What do we Know?," *NBER Working Paper,* n.12176.

[24] MAURIN, E., and S. McNALLY (2006): "Selective Schooling," *Mimeo.*

[25] MEGHIR, C. and PALME, M. (2005): "Educational Reform, Ability and Family Background," *American Economic Review*, 95(1) , 414-424.

[26] ROBERTSON, D., and J. SYMONS (2003): "Do Peer Groups Matter? Peer Group versus Schooling Effect on Academic Attainment," *Economica*, 70, 31-53.

[27] ROSENBAUM, P. R. and RUBIN, D. B. (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects. ," *Biometrika*,70 , 41-55.

[28] ROTHSTEIN, J., and A. YOON (2006): "Mismatch in Law School," *Mimeo.*

[29] RUBIN, D. (1974) "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*,66 , 688-701.

[30] TODD, P.E. and WOLPIN, K.I. (2003) "," On the Specification and Estimation of the Production Function of Cognitive Achievement *Economic Journal*,113, F3-F33.

[31] TODD, P.E. and WOLPIN, K.I. (2004) "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," *University of Pennsylvania Working Paper.*

[32] WALDINGER, F. (2006) "Does Tracking Affect the Importance of family Background on Students' Test Scores?," *Mimeo.*