The Intergenerational Effect of Parental Schooling: Evidence from the British 1947 School Leaving Age Reform*

Fernando Galindo-Rueda †

First and preliminary draft: December 2003

Abstract

This paper studies the causal impact of an additional year of parental schooling on children's educational and labour market outcomes. This study combines detailed longitudinal information from a cohort of British children and their parents, exploiting the natural experiment that stems from the 1947 school leaving age (SLA) reform, in the form of a *fuzzy* regression discontinuity design. Although there are strong impacts on individual (parental) earnings, the evidence on intergenerational effects is less robust, suggesting that in many cases OLS estimates are upward biased due to the intergenerational transmission of other unobserved characteristics. However, there are instances of positive effects of father's schooling on the attainment by sons and mother's schooling influencing child bearing decisions by daughters. Besides, the impacts are shown not to be even across the distribution of outcomes. A model is constructed to explain how policies such as SLAs can improve intergenerational outcomes even in the absence of credit constraints and spillovers outside the family.

*Please do not quote or circulate without the author's permission. Funding from the Treasury's Evidence Based Policy Fund and the EC Framework 5 EDWIN (Education and Wage Inequality in Europe) grant, and comments from Steve Machin and Anna Vignoles are gratefully acknowledged. The usual disclaimer applies.

†Centre for the Economics of Education, Centre for Economic Performance, London School of Economics; Centre for Research into Business Activity (CeRiBA) and IZA. Correspondence address: Centre for Economic Peformance, London School of Economics, Houghton Street, WC2A 2AE. Tel: ++44(0)2079556646. E-mail: F.Galindo-Rueda@lse.ac.uk

1 Introduction

Parents make all sorts of transfers to their children. These include market-valued resources such as time and efforts spent and other financial resources aimed at providing for basic needs, leisure and, in many instances, more and better quality education. In this context, any public policy aimed at improving the educational outcomes of children that does not account for the role played by parents is particularly vulnerable to failure.

The extent to which a child's level of educational attainment and further life outcomes depend on her parent's level of educational attainment captures some of the most passionately debated issues in economics, psychology, biology and sociology.[1] This debate has far-reaching consequences as it bears several implications regarding the scope for policy interventions that have an impact on a child's environment.[2]

Related questions arise with the topic of intergenerational mobility and the intergenerational transmission of human capital. Human capital transfers from one generation to the next mostly occur within families through a series of mechanisms, some of them just mentioned above. The persistence of poverty across generations is a socially undesirable problem which has no unambiguously optimal answer from a policy point of view. Only a better knowledge of the mediating mechanisms can help identify where interventions can be most effective.

In the empirical literature in economics, substantial efforts have been devoted trying to sort out sources of variation in parental education which are unrelated to the biological linkage that exists between parents and children. A branch in this literature, exemplified by Sacerdote (2002), has looked at the association between adoptive family income and education on children's outcomes. Adoption is considered there as a natural experiment which randomly –hopefully, from the empirical researcher's point of view– assigns parental levels of schooling to children.[3] An alternative to this approach is to consider differences within pairs of twins who happen to reach different levels of schooling, as in Ashenfelter and Krueger (1994). Behrman and Rosenzweig (2002) employ a male survey of identical twins (and their children) from the Minnesota Registry of Twins to investigate whether mother's schooling raises the schooling of the next generation, finding little evidence supporting this, if anything the opposite. A third, though far less common and generalizable approach, is to examine pairs of twins reared apart so that one can identify the effect of different environments for genetically identical children. The psychology literature has extensively studied these natural experiments but has concentrated on IQ scores rather than more usual measures of skills in the economics literature.

Policy initiatives have also been used as a source of valuable independent variation in schooling levels to look at in empirical analysis. Amongst this class of natural experiments, reforms raising

[1]At their core lies the old inquiry on the relative importance of genes and environment, as parents transmit a combination of their genetic endowments to their children.

[2]In the extreme, hypothetical case of a world entirely ridden by genes, there would be hardly any room left for effective interventions. This would also be the likely outcome if other non-genetic transmission mechanisms exerted a similar type of influence.

[3]For another example in this literature, see Plug (2002).

the statutory minimum school leaving age (SLA) are probably the most influential, alongside with measures affecting the price that individuals have to pay for education. The applicability of the reform is the instrument that leads to variation in schooling levels that is suitable for identification of its intra- and inter-generational effects. Besides, policy-based instruments are additionally valuable in that they also provide a guidance on the particular effects of the reform itself. There are instances in which this outcome can guide related reforms elsewhere or in the future.

This paper is concerned about the intergenerational impact of raising parental schooling and the methodology behind its identification and estimation, as driven by SLA reforms. In particular, I address the following questions:

1. How do SLA reforms affect the distribution of schooling in the population?
2. What is the effect of exogenous changes in an individual's schooling on the educational outcomes of her children? How do they compare with the intra-generational impacts?
3. How can one exploit SLA reforms to identify and estimate the inter-generational effect of an extra year of schooling?
4. Under which circumstances can policy promote better social outcomes if there are no intra-generational market failures?

SLA reforms are most commonly implemented on a nationwide basis, leaving no variation in the applicability of the reform apart from that arising from time series. In that case, confounding cohort and age effects may distort estimates. This paper main methodological contribution is to shed some light into this problem by thoroughly discussing the theoretical and practical features in using this type of SLA reform as a fuzzy regression discontinuity design (RDD).

For estimation purposes, I examine the reform that took place in Great Britain at the end of World War II which, among other things, raised the age of compulsory schooling from 14 to 15. Using extremely rich data from the National Child Development Study, a longitudinal study of the cohort born in Britain in a week of 1958, it is possible to relate parental education, in many cases influenced by the above mentioned reform, with the cohort member's own education and other adult outcomes. Moreover, data from the 1991 survey provides information about certain features of the educational attainment of the cohort members' own children, bringing three different generations together into a unique data set.

The empirical findings in this paper can be summarised as follows: Although parental schooling is robustly associated with children's outcomes, as inferred from OLS estimates with diverse controls, this conclusion cannot be drawn under more stringent identification constraints. There are counted though significant exceptions to this, emphasizing the importance of father's schooling towards mathematics and O-level attainment and that of mother's schooling regarding daughter's child bearing behaviour. This research highlights the importance of examining not only impacts on mean values, but elsewhere in the distribution of outcomes subject to investigation too. This is natural for identification methods that are based on natural experiments that stem from policies aimed at reducing inequality in society. As a result, measured effects will tend to underestimate the impact on the target population.

The plan of the paper is as follows: First of all, in section 2 I discuss the policy questions purely raised by intergenerational concerns. To this effect, I build a model that explains why individuals can fail to take into account gains from current schooling for their own family's future generations. In this context, policy measures normally thought of as correcting intra-generational market failures are also shown to improve social outcomes. Next, the NCDS58 data set is described in section 3, including some basic estimates of the association between parental characteristics and children's outcomes. Section 4 studies the identification of mean effects and partial quantile differences due to higher parental schooling. The estimation methods are discussed in section 5. After describing some of the features of the SLA reform in 1947 in section 6, estimation results are provided in section 7. Section 8 concludes.

2 Policy and intergenerational spillovers: An overlapping generations model

Arguments supporting the economic rationale for policy interventions on the sole basis of intergenerational education spillovers are not straightforward. Provided these spillovers are private to the decision making unit -i.e. the family-, nothing should in principle prevent it from making optimal decisions whenever the standard conditions for efficiency are met. In this section I provide a relatively simple model which characterises a rather plausible framework for inefficient decisions to arise within a private, intergenerational setting.

If exogenous changes in parents education are capable of inducing higher schooling for their offspring by making learning easier or more attractive, it is not always the case that future parents will internalise this type of outcome when making their own education decisions. Limited altruism can be considered as part of the answer, but this on its own will not conflict with general optimality. Irrational behaviour can be thought of as an alternative explanation, particularly since teenagers are often criticised for not caring enough about their future. My approach here is to model the intergenerational education spillovers in a framework which restricts the need for ad-hoc assumptions on the roots of behaviour. For this purpose, I combine the presence of forward looking young individuals making education (investment) decisions with the existence of altruistic parents. The link between both stems from the fact that young individuals, upon becoming adults, realise the welfare of their own children matters to them. Thus, young individuals are allowed to be forward looking but in a strictly individualistic fashion. When young, schooling effort decisions are made taking into account future financial gains (indeed, teenagers appear to have clear future earnings expectations) and balance them against more immediate costs. However, they fail to take into account that in a second stage, when adults, they will themselves be parents and their education will not only increase their own child's consumption through higher earnings, but also improve their efficiency at becoming educated.[4] More concisely, what this effectively means is that, when young, individuals cannot anticipate they will be altruistic towards their own offspring.

Let us formulate this type of model in a more precise way. Consider a society consisting only of overlapping generations of parents and their children where individuals live only for two broadly defined periods. In the first they are children and in the second they grow up and become parents. For simplicity, there are only pairs of gender-neutral parent and child. Denote x_p as the level of schooling of the parent (p), which the child (k, for kid) takes as given as no adult learning is considered here. When young, he will seek to maximise the following utility:

$$V_k(x_p, c_k) = \max_x \{u(c_k) - v(x) \cdot \phi(x_p) + \beta \cdot V_{kp}(x)\}, \tag{1}$$

where $u(c_k)$ denotes the utility derived when young from consumption c_k, determined by the parent, $v(x)$ denotes the child's present disutility from exerting effort so as to acquire schooling level x. The latter is affected by the factor $\phi(x_p)$, that is, depends on the level of parent's schooling. The thinking behind many theories of intergenerational transmission of human capital is that more educated parents find it easier to get their children educated either by being able to help with learning or by transmitting a series of values which place a higher value on education, making it more attractive. Optimisation takes into account future utility $V_{kp}(x)$ of the child as an adult, discounted by time preference factor β. From the point of view of a child, the future implies work which provides an income of $\psi(x)$, which is increasing in own schooling x, but awareness of the need to spend resources on their future own son is non-existent. All this implies that the young individual expects to use all his income for consumption when adult $c_{kp} = \psi(x)$ so that $V_{kp}(x) = u(\psi(x))$. The 'internally' optimal schooling decision will thus satisfy:

$$v'(x_k) \cdot \phi(x_p) = \beta \cdot u'(\psi(x_k)) \cdot \psi'(x_k). \tag{2}$$

Under standard assumptions, we can establish that a child's schooling choice is a function of the parent's $x_k = f(x_p)$. If $\phi(x_p)$ is decreasing in x_p, this implies that higher levels of paternal schooling reduce the disutility of schooling for the child, and thus schooling will be higher if everything else is equal.

When the individual reaches adulthood, the world now looks very different. A child is automatically born and altruism suddenly becomes an issue for the individual. The new child needs to be maintained while young, while he faces schooling decisions in the same way the parent did in the past. The new parent's only choice now is how much to consume for himself and how much to give to his own child. Consider the simplest possible specification for an altruist parent's

[4]This can be justified in a number of ways. For example, people can argue when young that they are not interested in having children and behave accordingly. Biological mechanisms (explained for by selection, e.g.) could explain the behavioural change upon growing older. Borrowing on Dawkins' terminology, one could argue that the 'selfish' gene takes over the 'selfish' individual at a given stage, turning 'selfish' individuals into 'altruistic' ones for the sake of gene preservation.

utility:

$$\begin{aligned}
V_p(x_p) &= \max_{\{c_p,c_k\}} \left\{u(c_p) + \delta \cdot V_k(x_p, c_k)\right\}, \\
\text{subject to } \psi(x_p) &= c_p + c_k \\
\text{and } x_k &= f(x_p). \qquad (3)
\end{aligned}$$

The parameter δ denotes the degree of altruism of the parent towards the child. When $\delta = 1$, the parent weighs the importance of his child's utility equally to his own. There are some aspects that merit further attention. First, I have assumed away the possibility of financial bequests. Parents only provide consumption and education opportunities. This would be the case with a 100 percent inheritance tax. The solution is quite straightforward, with parental earnings being split between parent and child according to relative marginal utilities and the altruism coefficient:

$$\begin{aligned}
u'(c_p) &= \delta \cdot u'(\psi(x_p) - c_p), \\
c_k &= \psi(x_p) - c_p. \qquad (4)
\end{aligned}$$

Besides, in this specification the parent is not extremely paternalistic in the sense that he cares only about making his child happier *as a child*. The alternative is to model a parent who cares about what his child should care about. This is basically the idea of *'one day you'll be a parent too'*, which is formalised below.

$$\begin{aligned}
V_p(x_p) &= \max_{c_p} \left\{u(c_p) + \delta\left(u(\psi(x_p) - c_p) - v(x_k)\cdot\phi(x_p) + \beta \cdot V_p(x_k)\right)\right\}, \\
\text{subject to } x_k &= f(x_p). \qquad (5)
\end{aligned}$$

This recursive formulation for a stationary problem basically states the problem in a fashion common to rational economic models, with parents making full forward-looking decisions, only to face a situation of *remorse* about their past actions knowing their own children will behave as they previously did. The child's limited rationality provides the basis for inefficient outcomes. In either of the cases considered above, the consumption split solution will be identical because there is nothing in the parent's set of feasible actions that can improve the mechanism of intergenerational transmission towards his own child and forthcoming generations.

A steady state solution in schooling levels will be found when $x = f(x)$, i.e. when $v'(x)\cdot\phi(x) = \beta \cdot u'(\psi(x)) \cdot \psi'(x)$. In this case, all generations in a family attain identical levels of schooling. There need not exist such type of equilibrium, and even if it does, the fixed point x may not be unique. This raises the question of whether policy can favour certain types of equilibria relative to others according to some notion of intertemporal social desirability.[5] For example, by fixing a minimum level of schooling $\underline{\mathrm{x}}$ that would rewriting the child's schooling decision as $x_k = \max\{\underline{\mathrm{x}}, f(x_k)\}$. This would involve a negative shock for the child's utility, but would

[5]For example, by making a current generation marginally worse off to produce a sizeable improvements for all future generations' welfare.

provide a boost to his own utility as a parent in both of the models, with the addition of a persistent boost to the family's capacity to produce human capital.

We can compare the solution to the problem with selfish sons with the problem in which parents can choose themselves the level of education of their own children. With altruism parameter $\delta = 1$, this would yield in a steady state: $\phi(x) \cdot v'(x) = \beta \cdot [u'(\psi(x)/2) \cdot \psi'(x) - \phi'(x) \cdot v(x)]$, which under the assumptions stated above will provide a higher level of schooling than the equilibrium in the 'selfish' model: $\phi(x) \cdot v'(x) = \beta \cdot u'(\psi(x)/2) \cdot \psi'(x)$. This is so because when the parent makes all choices, he takes into account the reduction in disutility experienced by his grandchild brought about by the child's higher level of schooling. This increases marginal benefits and thus raises the equilibrium level of schooling. In this model, there is nothing that policy can do to improve outcomes even if there are positive intergenerational spillovers. Thinking about young individuals who make investments without realising about the positive spillovers of their current decisions helps understand under which circumstances policy can improve welfare.

There are multiple avenues for improving the characterisation of intergenerational choices by building on this simple model. One would imply integrating this framework with standard credit constraint arguments and other intragenerational efficiencies. Another would pursue to integrate this model within a general equilibrium framework where the function $\phi(x)$ is affected by the supply of educated individuals. In the appendix, I relax the assumption that that parents do not effectively make purposeful intergenerational investments in their offspring.

Traditionally, the education problem has been considered in a framework in which parents' altruism is limited and they cannot borrow against their son's future earnings to pay for their education. This is just a reformulation of the intragenerational credit constraint problem, with a change in the perspective on who makes investment and the nature of the non-contractibility. In the intergenerational case, parents cannot draw contracts with their children, whereas in the intragenerational case, children cannot sign contracts with lenders. To this effect, a non altruistic father *is* just another potential bank, whereas if altruistic, the phenomenon is purely intragenerational as he is the credit constrained party.

Standard education policy instruments[6] have been traditionally thought of as mechanisms for dealing with the standard problem of borrowing against future earnings to pay for current investments in education. This model has argued that such policy tools can also influence the mechanism of intergenerational transmission of human capital that runs at a much deeper level than the credit constraint framework[7] and is suspected as responsible for a large share of existing educational inequalities in developed economies.

Having argued why the transmission of human capital from parents to children can be relevant for economists and policy makers alike, I now turn to the description of the data that will assist

[6]For example, policy interventions that distort the incentives to exert effort by sons –such as Education Maintenance Allowances (EMAs) paid directly to the children, School leaving age reforms– and the incentives to invest in children's education –such as EMAs paid to parents, taxation of schooling-related expenditure, taxation of properties located closer to better schools, charitable status of private educational institutions, etc.

[7]For example, see Cameron and Heckman (1998).

in identifying and estimating such effects.[8]

Central to my analysis is the equation $x_k = f(x_p)$. In the theoretical model presented above, nothing prevented different family units from having different parameterisations for their utility functions. For example, ability was not explicitly considered but it could be easily embedded in the function $v()$, which determines the disutility of educational investments. More able individuals may have a lower marginal disutility from schooling and therefore tend to become more educated. In a cross-section of families, any estimate of the derivative $f'(x_p)$ might be biased as a result of the unobserved correlation between ability and both parent and child education levels if there is some degree of intergenerational correlation in ability.[9] As a result of this, an empirical researcher faces limitations as to what can be said about the nature of $\phi'(x)$ and $\psi'(x)$ from a data consisting of paired observations of (x_p, x_k). This paper is concerned with addressing the policy relevant question of how much will the attainment of a new generation will respond to a policy-driven increase in the current generation's level of schooling.

3 Data description

This study uses data from the The National Child Development Study (NCDS). This is a continuing, multi-disciplinary longitudinal study which takes as its subjects all those living in Great Britain who were born between 3 and 9 March, 1958. Following the initial birth survey in 1958 - the Perinatal Mortality Survey (PMS) - there have been several attempts to trace all members of the birth cohort in order to monitor their physical, educational and social development. Data were collected in 1958 (birth), 1965 (age 7), 1969 (age 11), 1974 (age 16), 1978 (exam results obtained), 1981 (age 23), 1991 (age 33), 1999 (age 41).

The rationale for relying on this data is basically twofold. First, the NCDS is one of the few available studies that provides information on children over their lifetime as well as their parents. In our case, NCDS is particularly helpful because its rich information not only provides clues about the basic correlation between parental education and child's qualification attainment, but also illustrates how parental education affects numerous mediating factors, including living conditions, early educational outcomes as inferred from test scores on a range of subjects, child's social adjustment and measures of parental interest in child's education at various stages. It is also possible to check whether there exist long-term benefits for the child from parents' education in the form of higher earnings as adults or many other attributes. This wealth of information is further complemented by the fact that at the 1991 survey of the cohort members, a random one-third of their biological children were also surveyed and administered most of the cognitive and social-emotional tests that were given to the Children of the NLSY sample, including reading and mathematics PIAT scores.

[8]This being a gender neutral model, there are not explicit predictions about differences in the effect of paternal and maternal education.

[9]This argument's relevance increases the degree of intergenerational transmission of the unobserved ability variable, which conceptually is not a feasible object of policy. The model could easily be extended to include a Markov transition process of ability from parents to children.

Given the existence of a data set that contains information about individuals and their parents, the second advantage of looking at the NCDS cohort is the fact that the population of parents for this cohort can be clearly divided according to their exposure to different levels of compulsory schooling (SLAs).[10] In practice, this means that all parents aged less than 24 were subject to the SLA of 15, as opposed to their older counterparts who could leave as early as 14. I exploit this variation as a source of identification of the relevant intergenerational transmission mechanism.

NCDS variables

One of the key advantages of NCDS is the ability to investigate a whole series of educational outcomes measured at different ages. I focus on standardised tests that were administered to subject children in their schools, by their teachers. At age 7, a reading (Southgate Reading Test) and an arithmetics test (Pringle Problem Arithmetic) were used. At the ages of 11 and 16, specially constructed tests by the National Foundation of Educational Research (NFER) on mathematics and reading comprehension were administered. Another test on general ability with verbal and non-verbal components, trying to capture a measure of IQ, was obtained from the age 11 survey. This information is complemented with data on their academic qualifications attainment and information on a measure of behavioural adjustment problems, the Bristol Social Adjustment Guide, available at 7 and 11. I look at adult individual outcomes such as number of children at 23 and 33, as well as earnings at 33 and 42.

Parental information on schooling and their own background is derived from questions in the birth survey. Schooling is derived as the age (integer) at which parents left full time education. A broad measure of occupational classification of the father and each parent's father are available. These classes are: Professional, intermediate/supervisory, skilled non-manual, skilled manual, semi-skilled non-manual, semi-skilled manual and unskilled. Unfortunately, there is no information about grandparent's level of schooling though we can safely assume that -at the time- this was mostly captured by social class. Parental gross weekly earnings and income are obtained from the 1974 survey, when the children were 16. The number of siblings provides an indication of the level of competition for resources in the household. Finally, teacher's reports provide information on the degree of interest of both father and mother in the child's education at 7, 11 and 16.

Table 1 summarily describes some of the key variables mentioned by exploring their pairwise correlations, all of them statistically significant. The correlation between parental schooling and ability scores appears to increase with age, suggesting some degree of cumulative effects as the child grows up. Correlations with learned abilities (maths and reading) are only slightly higher than correlations with the general ability proxy for IQ. Parental interest is not more correlated with parental education than with child's ability, suggesting that parents may exert a greater effort on their children if they see some potential, but this argument is certainly subject to a

[10]More details on the intervening SLA reform in 1947 will be provided in section 6.

Table 1: Simple Correlations

	Fat' Sch Sch.	Mat' Sch.	lnInc. (1974)	Maths (7)	Read (7)	GenAb (11)	Maths (11)	Read (11)	Maths (16)	Read (16)	Fa'Int	Ma'Int	Sibs
Father's Sch.	1.00												
Mother's Sch.	0.56	1.00											
Log income	0.27	0.26	1.00										
Maths(7)	0.17	0.17	0.12	1.00									
Read(7)	0.20	0.19	0.15	0.53	1.00								
Gen.Ab.(11)	0.27	0.25	0.19	0.52	0.64	1.00							
Maths (11)	0.30	0.28	0.19	0.57	0.61	0.81	1.00						
Read (11)	0.29	0.28	0.18	0.48	0.62	0.74	0.74	1.00					
Maths (16)	0.32	0.30	0.20	0.48	0.49	0.68	0.76	0.64	1.00				
Read (16)	0.27	0.27	0.19	0.47	0.61	0.72	0.70	0.79	0.65	1.00			
Fa's interest	0.27	0.23	0.19	0.23	0.27	0.36	0.39	0.37	0.38	0.34	1.00		
Ma's interest	0.22	0.22	0.15	0.23	0.28	0.37	0.39	0.37	0.36	0.35	0.66	1.00	
No.Siblings	-0.12	-0.12	-0.09	-0.06	-0.12	-0.14	-0.13	-0.16	-0.14	-0.21	-0.12	-0.12	1.00

NOTES: Pairwise correlation coefficients of child's attainment scores, parental schooling, income, level of interest in child's education and number of siblings.

circularity criticism. Children of more educated parents are also likely to have less siblings and more likely to have higher income. The correlation between parental income and child ability is substantially lower than that between parental education and child ability.

I also investigate the statistical association between parental schooling and parental interest on child's education, by child's age and gender. Table 2 presents the estimated marginal effects for a probit model of parental interest (whether very interested in child's education, approx. 30% of sample) on both parents' schooling and age, family income, father's social class and number of siblings.

In general, mother's schooling is associated not only with a higher interest on her side, but also on the father's. The same applies to the effect of father's schooling, which is a far more robust result given the inclusion of the social class variable. There are only two instances in which the effects are not significant, as it occurs for mother's effect on mother's interest at 16 in boys and for father's effect on mother's interest at 11.

The logical next step is to investigate how all these characteristics combine to produce academic qualifications. In particular, I examine the association between educational attainment of NCDS cohort members and their parents', and observe how this association changes as one controls for other observed characteristics. No causal interpretation can be yet given to these coefficients as all evidence points out that parental schooling is endogenous to the outcome of interest.

In table 3 I examine two types of outcomes. Columns (1) to (7) depict results for an ordered logit regression of highest educational attainment on the set covariates. Columns (8) to (12) reproduce estimation results for the probability of attaining a level 3 or higher qualification, which theoretically correspond to the period of post-compulsory schooling. Table 3 shows, when

Table 2: Parental schooling and parental interest in child's education

	Par. Interest in Boys				Par. Interest in Girls			
	Interest at 11		Interest at 16		Interest at 11		Interest at 16	
	Mother	Father	Mother	Father	Mother	Father	Mother	Father
Mother's schooling	0.036	0.024	0.014	0.038	0.041	0.027	0.041	0.065
St. Error	*0.011*	*0.010*	*0.012*	*0.012*	*0.012*	*0.010*	*0.011*	*0.012*
Father's schooling	0.010	0.016	0.026	0.039	0.023	0.035	0.030	0.022
St. Error	*0.010*	*0.009*	*0.010*	*0.010*	*0.010*	*0.009*	*0.009*	*0.010*

NOTES: Probit marginal effects with standard errors reported. Dependent variable=whether teacher thinks parent very interested in child's education, at ages of 11 and 16. Observations: Boys=1629. Girls=1685. Additional controls detailed in main text.

Table 3: Conditional Estimates of Association between Parental Education and Offspring's Qualifications

		Ordered logit coef. for highest qual.						Prob of attain. post-16 qualification					
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Boys	Father's schooling	0.134	0.147	0.138	0.138	0.190	0.215	0.037	0.039	0.037	0.038	0.046	0.054
	St. Error	0.041	0.040	0.040	0.041	0.041	0.041	0.010	0.009	0.009	0.010	0.010	0.010
	Mother's schooling	0.035	0.043	0.049	0.055	0.168	0.191	0.017	0.018	0.019	0.023	0.041	0.047
	St. Error	0.050	0.049	0.048	0.048	0.045	0.044	0.012	0.012	0.012	0.012	0.012	0.012
Girls	Father's schooling	0.058	0.087	0.070	0.078	0.130	0.153	0.014	0.019	0.016	0.021	0.032	0.037
	St. Error	0.040	0.037	0.039	0.041	0.038	0.037	0.008	0.008	0.008	0.009	0.009	0.010
	Mother's schooling	0.122	0.130	0.149	0.127	0.216	0.265	0.022	0.024	0.030	0.023	0.041	0.054
	St. Error	0.046	0.046	0.045	0.047	0.044	0.043	0.010	0.010	0.010	0.011	0.011	0.011
Controls													
Ability scores at 16		√	√	√	x	x	x	√	√	√	x	x	x
Ability scores at 11		√	√	√	√	x	x	√	√	√	√	x	x
Parent's age and household size		√	√	√	√	√	√	√	√	√	√	√	√
Parental income		√	√	√	√	√	√	√	√	√	√	√	√
Father's SEG		√	x	√	√	√	√	√	x	√	√	√	√
Parent's interest in child's education		√	√	x	√	√	x	√	√	x	√	√	x

NOTES: Columns (1) to (6): Ordered logit estimates and st.errors of coefficients on father's and mother's schooling, with different sets of controls. Dependent variable=highest educational attainment. Columns (7) to (12): Probit marginal effects and st.errors of coefficients on father's and mother's schooling. Dependent variable=whether offspring attained qualification at A-level or higher. Observations: Boys=1629. Girls=1685.

I control for child ability and parents' characteristics, that the effect of mother schooling on boys' attainment is not statistically significant, whereas the opposite is true for the effect of father's schooling. This picture is reversed if one looks at the effect on girls' attainment. Given that father's SES is also included, this may tend to underestimate the importance of father's schooling relative to the mother's.

Controlling for parental interest appears to reduce marginally the estimated effect of parental education, as one compares columns (7) and (9), for example. The inclusion of child ability in the specifications provides a lower bound on the effect parental schooling by implicitly attributing all of child's ability to genes. The difference between estimates (col(7) vs col(11), for example, is particularly striking for the coefficient on mother's schooling, which basically doubles in size as one removes the ability control. This is also true for the effect of father's schooling on girls' attainment. Without controlling for either ability or parental interest, the effect of an extra year of parental schooling appears to be associated with a 5 % higher probability of obtaining a level 3 or higher qualification (2% in worst case scenario). This effect may be moderate, but if one considers the economic return to this type of qualification (assume 20%), an extra year for both parents will imply an expected extra 2% earnings per child when adult. I now proceed to investigate whether these estimates be given a causal interpretation.

4 School leaving age reforms as a regression discontinuity design

There are abundant examples in the literature on how changes in the MSLA can be used as a source of exogenous variation in the levels of schooling so as to estimate the effect of schooling on a number of variables. For example, Harmon and Walker (1995) use two reforms raising in SLA in Great Britain to estimate the wage returns to schooling.[11] Chevalier (2003) uses one of these reforms, the change in SLA from 15 to 16 for the cohorts born after 1957, to identify the effect of parental education on child's staying on at school after the -still unchanged- SLA of 16 in England and Wales. Black *et al.*(2003) have discussed the shortcomings in relying on purely time series variation in the SLA instrument, as well as some other features in the implementation of the IV estimator in Chevalier (2003). The problem with variation in the SLA being restricted to time series stems from the need to impose that circumstances and attributes of cohorts prior to the reform are identical to those affected by the reform relative to their impact on the unobserved heterogeneity in the outcome of interest. The restricted variation in the instrument becomes more accentuated with the need to make arbitrary restrictions on the functional form for the way cohort patterns are supposed to enter as controls in the equation being estimated.

Let us consider an extreme case, in which only a cross section of offspring is available. Parent's year of birth fully determines exposure to a given SLA, but children born to older parents are also likely to differ from children born to younger parents in a large number of features, some of them potentially correlated with the outcome of interest, namely, their own educational attainment. Additionally, child's age is likely to influence educational outcomes, hence that variable should be also adequately controlled for. Chevalier (2003) pools seven cross sections of families from the Family Resource Survey, but excludes all information about the parent's birth date from the second stage. As suggested above, this imposes several restrictions on the model, which turn out to be more evident when, by reducing the width of the window for parent's year of birth considered in the sample (around year of reform), estimates considerably shrink down to the

[11] Oreopoulos (2002) uses changes in minimum school-leaving laws in Great Britain and Ireland to measure pecuniary and non-pecuniary gains from education. Clark and Hsieh (2000) use instead a school expansion program in Taiwan.

point of becoming statistically insignificant.

Black *et al.* (2003) provide instead evidence based on a change in the SLA in Norway which occurred over a period in which different municipalities were able to decide the timing of the increase in the SLA. In this case, the identifying variation in SLA coming from within municipalities time variation in the applicable SLA. This allows them to include both municipality and year of birth fixed effects, thus accounting for some of the problems just mentioned. Another relevant study which uses regional and time series variation in schooling laws is that by Oreopoulos *et al.* (2003), who estimate the effect of parental education on children's grade retention and staying-on behaviour looking at children and their parents in the 1960, 1970 and 1980 US Censuses.[12]

In this paper, I argue that country-wide SLA reforms can be used in a meaningful way to locally identify and estimate intra- and intergenerational effects of education, provided reasonable sample sizes can be obtained. Special emphasis is made on the intergenerational impact on children's outcomes. The basic point is that the introduction of the SLA induces a discontinuity in the rules that influence parent's educational attainment. This variation in schooling can be thought of as independent of the unobservables that also drive the outcome of interest (child's outcomes) at least for a sufficiently close set of birthdates to the discontinuity point. This strategy, known as Regression Discontinuity Design (RDD) has been proposed and used by Thistlethwaite and Campbell (1960) on scholarships and career aspirations, Angrist and Lavy (1999) for the effects of classroom size on attainment, Van der Klaauw (1997) for financial aid offers on college attendance, Hahn *et al.* (1999,2001), for the effects of discrimination laws on minorities employment, Black (1999) for the effect of school quality on house prices, among many others. Whereas the idea of thinking of SLA reforms as regression discontinuities is not new, little has been to fully exploit the full range of available econometric methods to the task of identifying and estimating the effect of parental schooling on children's outcomes. That is this paper's main methodological contribution.

4.1 Identification of expected treatment effects of parental schooling

The goal of this study is to determine the causal effect of the pseudo-continuous treatment X, which denotes the age at which a parent left education, on a particular outcome Y. The model for the observed outcome can be described as $Y = \alpha + \beta X$, which allows for heterogeneous returns to the treatment X by treating β as potentially random, as α and Y. In our case, there is a continuous variable B, denoting parent's date of birth, which fully determines the minimum school leaving age applicable to the child's father. The treatment X (schooling) does not depend in a deterministic way on B for a number of reasons. Above the minimum school leaving age (MSLA), completed years of schooling are determined by many observed and unobserved characteristics. There may be also realizations of X below the MLSA threshold because of limited

[12]This study is close in spirit to that by Currie and Moretti (2002), who exploit regional in college openings (as opposed to schooling laws) as an instrument for maternal schooling so as to estimate the impact of the latter on subsequent birth outcomes.

enforceability of the MLSA laws. A feature of the discontinuity design -known as fuzzy due to the afore mentioned reasons- is the fact that the expected value of the treatment $E[X|B=b]$[13] experiences a discontinuity at a value $b=\bar{B}$, which in this case denotes that parents born before 1933 were subject to a MSLA of 14 whereas those born in 1934 and after had to stay on in school at least until the age of 15, due to the imposition of the new MSLA in 1947. We can then assume the existence of left and right limits at $\bar{B}$ such that:

$$x^{+} \equiv \lim_{b \to \bar{B}^{+}} E[X|b] > x^{-} \equiv \lim_{b \to \bar{B}^{-}} E[X|b].$$

The data are assumed to provide a sample of independent N realizations of the vector (Y,X,B) given by $\{y_i, x_i, b_i\}_{i=1}^{N}$. We can consider now the difference between expected outcomes for individuals with birth dates arbitrarily close to $\bar{B}$, after and before respectively.

$$E[Y|\bar{B}+\epsilon] - E[Y|\bar{B}-\epsilon] = E[\alpha|\bar{B}+\epsilon] - E[\alpha|\bar{B}-\epsilon] + E[\beta X|\bar{B}+\epsilon] - E[\beta X|\bar{B}-\epsilon].$$

If we assume that $E[\alpha|b]$ is a continuous function of b at $\bar{B}$ and the return to the schooling treatment is homogeneous ($Var(\beta)=0$), the limit of this difference when $\epsilon \to 0$ will yield:

$$y^{+} - y^{-} \equiv \lim_{\epsilon \to 0} E[Y|\bar{B}+\epsilon] - \lim_{\epsilon \to 0} E[Y|\bar{B}-\epsilon] = \beta(x^{+} - x^{-}),$$

because by continuity $\lim_{\epsilon \to 0} E[\alpha|\bar{B}+\epsilon] - \lim_{\epsilon \to 0} E[\alpha|\bar{B}-\epsilon] = 0$. Therefore, the parameter of interest $E[\beta] \equiv \beta_0$ is identified by:

$$\beta_0 = \frac{y^{+} - y^{-}}{x^{+} - x^{-}}, \tag{6}$$

which immediately suggests the Wald Estimator $\hat{\beta}_0 = \frac{\hat{y}^{+} - \hat{y}^{-}}{\hat{x}^{+} - \hat{x}^{-}}$, based on consistent estimators $\hat{y}^{+}, \hat{y}^{-}, \hat{x}^{+}\hat{x}^{-}$ of the four one-sided limits.

If we allow instead β to be heterogeneous, but impose that X is independent of β in the vicinity of $\bar{B}$ and that the expectation $E[\beta|b]$ is continuous at $b=\bar{B}$, then $E[\beta X|\bar{B}+\epsilon] = E[\beta|\bar{B}+\epsilon] \cdot E[X|\bar{B}+\epsilon]$. This implies that: $\lim_{\epsilon \to 0}(E[\beta X|\bar{B}+\epsilon] - E[\beta X|\bar{B}-\epsilon]) = \lim_{\epsilon \to 0}(E[\beta|\bar{B}+\epsilon] \cdot E[X|\bar{B}+\epsilon] - E[\beta|\bar{B}-\epsilon] \cdot E[X|\bar{B}-\epsilon]) = E[\beta|\bar{B}] \cdot (x^{+} - x^{-})$. Then:

$$(y^{+} - y^{-}) = E[\beta|\bar{B}] \cdot (x^{+} - x^{-}),$$

which identifies the average treatment effect of X on Y at $B=\bar{B}$.

4.2 Identification of partial quantile differences in a discontinuity design

It is often the case that applied analysis tends to focus on mean impacts and neglect potential effects arising in particular moments of the distribution of outcomes. Indeed, SLA reforms are

[13]For notational simplicity, I will write $E[X|b]$ instead of the more appropriate $E[X|B=b]$.

not devised for just average pupils, but instead seek to improve the performance of the majority or at least those doing less well. It is worth examining under which conditions one can identify changes in specific quantiles of the distribution of outcomes.

Following Chesher's (2002a,2002b) analysis of instrumental values, we can think of our RDD setting as one in which two values arbitrarily close to $\bar{B}$ are the only instrumental values available, namely, $\bar{B}-\epsilon$ and $\bar{B}+\epsilon$. It is possible to present a more general model of the effects of schooling X on Y:[14]

$$Y = h_1(X, B, \nu_1, \nu_2) \quad (7)$$
$$X = h_2(B, \nu_2) \quad (8)$$

This is a triangular model in which some monotonicity assumptions are imposed, namely, h_1 being non-increasing (or non-decreasing) in ν_1, and h_2 being strictly increasing (or decreasing) in ν_2.[15] The standard limitations on the conditional distribution of the unobserved ν_1 and ν_2 are assumed in this case to hold approximately for instrumental values $b'' = \bar{B}+\epsilon$ and $b' = \bar{B}-\epsilon$. The only difference from the traditional framework stems from the fact that the statement refers to conditional quantiles. For example, $Q_{\nu|B}(b,\tau)$ denotes the τ quintile of variable ν conditional on the variable B adopting the value b, i.e. $Pr[\nu \leq Q|B=b] = \tau$. Thus:

$$\lim_{\epsilon\to 0}\left(Q_{\nu_2|B}(\bar{B}+\epsilon,\tau_2^*) - Q_{\nu_2|B}(\bar{B}-\epsilon,\tau_2^*)\right) = 0 \quad (9)$$
$$\lim_{\epsilon\to 0}\left(Q_{\nu_1|\nu_2,B}(\bar{B}+\epsilon,\tau_1^*,Q_{\nu_2|B}(\bar{B}+\epsilon,\tau_2^*)) - Q_{\nu_1|\nu_2,B}(\bar{B}-\epsilon,\tau_1^*,Q_{\nu_2|B}(\bar{B}-\epsilon,\tau_2^*))\right) = 0 \quad (10)$$

For this to be true it is sufficient that the conditional quantiles are continuous functions of B at $\bar{B}$. The asterisk denotes the particular quantile at which the invariance assumption holds. This means that the statement does not need apply for the whole distribution, but it suffices for it to be true at a particular quantile of interest. For simplicity, denote the conditional quantiles $v_2^*(b,\tau_2^*) \equiv Q_{\nu_2|B}(b,\tau_2^*)$ and $v_1^*(b,\tau_1^*) \equiv Q_{\nu_1|\nu_2,B}(b,\tau_1^*,v_2^*(b,\tau_2^*))$.

Additionally, the equivalent of an order condition is satisfied when variation within the set of instrumental values (only b' and b'') does not lead to variation in the outcome Y if the intermediate outcome X is kept fixed.

$$\lim_{\epsilon\to 0}\left(h_1(x,\bar{B}+\epsilon,v_1^*(\bar{B}+\epsilon,\tau_1^*),v_2^*(\bar{B}+\epsilon,\tau_2^*)) - h_1(x,\bar{B}-\epsilon,v_1^*(\bar{B}-\epsilon,\tau_1^*),v_2^*(\bar{B}-\epsilon,\tau_2^*))\right) = 0 \quad (11)$$

Under these conditions,[16] the following functional of conditional distributions of observed characteristics (Y,X) at $\bar{B}-\epsilon$ and $\bar{B}+\epsilon$ can be constructed for ϵ sufficiently small

$$Q_{Y|X,B}(\tau_1^*,Q_{X|B}(\tau_2^*,\bar{B}+\epsilon),\bar{B}+\epsilon) - Q_{Y|X,B}(\tau_1^*,Q_{X|B}(\tau_2^*,\bar{B}-\epsilon),\bar{B}-\epsilon), \quad (12)$$

[14]One could also follow the approach of Chernozhukov and Hansen (2001), who develop a general model of quantile treatment effects with treatment endogeneity.

[15]These conditions can be normalised to non decreasing h_1 and strictly increasing h_2.

[16]Together with a completeness assumption about the existence and uniqueness of a solution to $Y = h_1(X,b,v_1^*(b),v_2^*(b))$ and $Y = h_2(b,v_2^*(b))$ denoted by $y^*(b)$ and $x^*(b)$ respectively.

making it possible to identify the partial difference:

$$h_1(x^*(\bar{B}+\epsilon),\bar{B}+\epsilon,v_1^*(\bar{B}+\epsilon),v_2^*(\bar{B}+\epsilon))-h_1(x^*(\bar{B}-\epsilon),\bar{B}-\epsilon,v_1^*(\bar{B}-\epsilon),v_2^*(\bar{B}+\epsilon)), \quad (13)$$

which in itself is of little interest, but as $\epsilon \to 0$ provides:

$$h_1(x_+^*(\bar{B}),\bar{B},v_1^*(\bar{B}),v_2^*(\bar{B}))-h_1(x_-^*(\bar{B}),\bar{B},v_1^*(\bar{B}),v_2^*(\bar{B})). \quad (14)$$

This has a potentially interesting causal interpretation only when $x_+^*(\bar{B}) \equiv lim_{\epsilon\to 0}x^*(\bar{B}+\epsilon) \neq x_-^*(\bar{B}) \equiv lim_{\epsilon\to 0}x^*(\bar{B}-\epsilon)$, for otherwise the quantile impact function is nil by definition (no change in any of the arguments). This "rank" condition implies that inference is only meaningful for cases in which a conditional quantile of X conditional on B experiences a discontinuity at $\bar{B}$. In the absence of spillovers, an increase in the MSLA may imply no change in the upper quantiles of the schooling distribution. In that case, it will be impossible to identify partial differences involving higher levels of X.

Since in our empirical application (discontinuity framework) only two instrumental values are available ($\bar{B}-\epsilon$ and $\bar{B}+\epsilon$), for a given percentile τ_2^* only two values of schooling can be generated $(Q_{X|B}(\tau_2^*,\bar{B}+\epsilon),Q_{X|B}(\tau_2^*,\bar{B}-\epsilon))$, and the X-chord of the structural function h_1 can be calculated as:

$$\frac{h_1(x_+^*(\bar{B}),\bar{B}+\epsilon,v_1^*(\bar{B}),v_2^*(\bar{B}))-h_1(x_-^*(\bar{B}),\bar{B},v_1^*(\bar{B}),v_2^*(\bar{B}))}{x_+^*(\bar{B})-x_-^*(\bar{B})}. \quad (15)$$

Thus, beyond the local nature of the identification (at $\bar{B}$), further restrictions arise because of the limitations induced in the possible choice of τ_2^* such that the rank condition holds. Identification is only feasible for quantiles of the schooling distribution which have discontinuously changed as a result of the introduction of a higher MSLA.

4.3 Other problems

There are other aspects of practical importance. Unfortunately, our data do not provide precise birth dates for parents, (the questionnaire only asks the age integer for both parents when the cohort member child is born). This implies that instead of observing the continuous B, we only observe realizations $\{j_i\}$ of the birth year J with integer values in $\{1920,...,1939\}$ such that $j_i=j$ if and only if $j-1<B\leq j$. To simplify things, it is possible to assume that the discontinuity in the MSLA occurs at a particular integer, i.e. $int(\bar{B})=\bar{B}$.[17] We can consider the following estimator:

$$\tilde{\beta}=\frac{\hat{E}[Y|J=\bar{B}+1]-\hat{E}[Y|J=\bar{B}]}{\hat{E}[X|J=\bar{B}+1]-\hat{E}[X|J=\bar{B}]}. \quad (16)$$

[17] In fact, parental age is obtained when the child is born in a week of March, not distant from the 1st April date set to determine who was to be affected by the raising MSLA.

The conditional expectations are all identified from the data and can be consistently estimated with simple means. Under the common treatment effect assumption $E[Y|J = \bar{B}+1] - E[Y|J = \bar{B}] = \beta(E[X|J = \bar{B}+1] - E[X|J = \bar{B}]) + E[\alpha|\bar{B} < B \leq \bar{B}+1] - E[\alpha|\bar{B}-1 < B \leq \bar{B}]$, which implies the existence of a bias given by:

$$plim[\tilde{\beta} - \beta_0] = \frac{E[\alpha|\bar{B} < B \leq \bar{B}+1] - E[\alpha|\bar{B}-1 < B \leq \bar{B}]}{E[X|\bar{B} < B \leq \bar{B}+1] - E[X|\bar{B}-1 < B \leq \bar{B}]}.$$

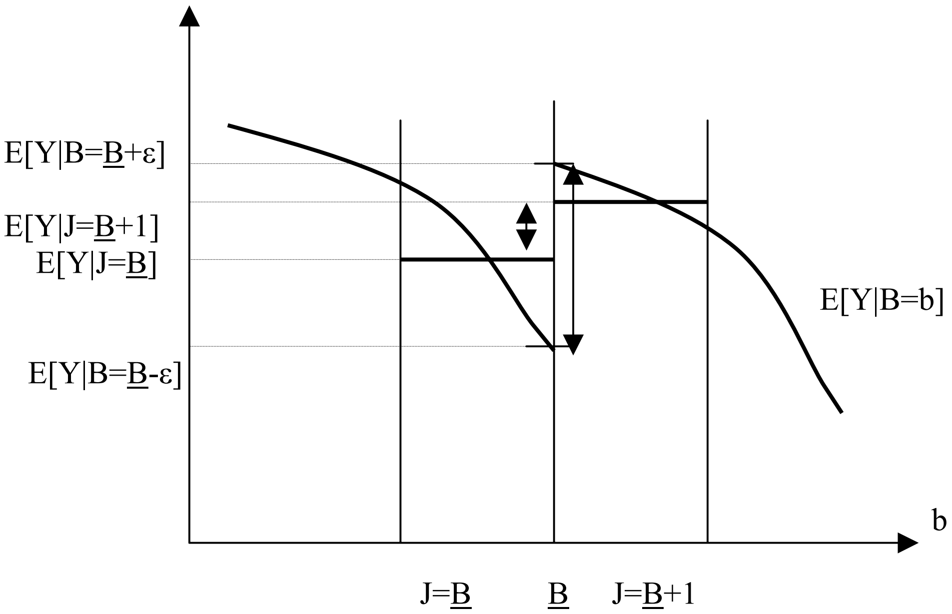
Selection into our parent's data is determined by the requirement of having a child born in a given week in 1958. Besides cohort patterns in the relationship between characteristics and year of birth, it is also important to take into account that younger parents have different attributes than older parents. In this case, for example, teenage parents lie on the right hand side of the discontinuity, benefiting from a higher MSLA, but also subject to worse socioeconomic (and possibly also inherited ability) conditions. Thus, the basic comparison refers to parents aged 25 in 1958 (MSLA=14) and parents aged 24 in that particular year (MSLA=15). If we assume that $E[\alpha|b] = \rho_\alpha \cdot b$, then we can rewrite the numerator in the bias expression as $\rho_\alpha(E[B|\bar{B} < B \leq \bar{B}+1] - E[B|\bar{B}-1 < B \leq \bar{B}])$ and if B is uniformly distributed then $E[B|\bar{B} < B \leq \bar{B}+1] - E[B|\bar{B}-1 < B \leq \bar{B}] = \bar{B}+1/2 - \bar{B}+1/2 = 1$ the bias would become ρ_α. We can think of ρ_α as an approximate correlation between the baseline outcome and the individual birth date. For 'positive' outcomes, if younger parents in the sample are more likely to have lower values, then $\rho_\alpha < 0$. Regarding the denominator, it is important that the underlying discontinuity due to the higher SLA exceeds the effect of the declining ability for younger parents. If the difference is small, even if positive, any existing bias will be further accentuated. Thus, in general, the estimator $\tilde{\beta}$ will systematically underestimate the genuine causal effect of the treatment x.

This problem of missing information on the detailed parental date of birth date b implies a source of bias for estimates. However, it must be also noted that for a finite sample of continuous observations, some bias will follow from the need to calculate estimates for the limits using observations (averaging) that lie not so close to the discontinuity point, as seen in figure 1. If in our data we expect a monotonically decreasing profile around the discontinuity due to parent's age association with ability, the value of the potentially biased estimates stems from its role as lower (upper) bound estimates for the impact of schooling on positive (negative) outcomes.

5 Estimation methods

As explained in the previous section, our ability to identify the impact of an additional year of education on individual and offspring's outcomes relies on the capacity to consistently estimate a series of limits for conditional means or quantiles. This section explains the empirical methods used for this purpose.

Figure 1: Bias in Regression Discontinuity Design



NOTES: Illustration of practical problems with RDD. Averages on the left and right hand side of discontinuity lead to underestimation of discontinuity when the profile is monotonically increasing around the discontinuity point.

5.1 Estimation of average impacts

Consider the solution for the following local linear regression of Y on B for a value $b \leq \bar{B}$:

$$\min_{\{\alpha_-(b),\beta_-(b)\}} \sum_{i=1}^{N} (y_i - \alpha(b) - \beta(b) \cdot b_i)^2 \cdot \mathbf{1}(b_i \leq \bar{B}) \cdot \mathbf{K}\left(\frac{b - b_i}{h}\right), \tag{17}$$

where $\mathbf{K}(s)$ is a symmetric, compactly supported, univariate probability kernel (non negative and satisfying $\int \mathbf{K}(s)ds = 1$) and h is the standard bandwidth. Applications of local linear regression has been discussed by Fan and Gijbels (1996). This type of local modelling approach such as this is very convenient because it lets data choose a suitable function that fits well the given data. In our setting, year of birth can affect outcomes in many possible ways. With global modelling, a large bias may occur when the underlying regression curve is not in the parametric family. Continuity everywhere apart from the reform year is the only constraint imposed on the data. This is the only requirement imposed by the theory of regression discontinuity design.

If we denote the weights $LK_i(b) \equiv \mathbf{1}(b_i \leq \bar{B}) \cdot \mathbf{K}(\frac{b-b_i}{h})$, then the minimand can be written as:

$$\sum_{i=1}^{N}(LK_i(b)^{1/2}y_i - \alpha(b) \cdot LK_i(b)^{1/2} - \beta(b)LK_i(b)^{1/2} \cdot b_i)^2, \tag{18}$$

which is equivalent to the objective of an OLS regression of y_i on $LK_i(b)^{1/2}$ and $LK_i(b)^{1/2} \cdot b_i$, thus giving positive weight only to values less or equal $\bar{B}$ and giving more importance to those values closer to b, according to the specified bandwidth and the kernel function. For the purposes of the RDD, one can examine the estimate for $b = \bar{B}$, that is, right at the boundary. Our interest in boundaries provides another justification for local linear regression as opposed to other non-parametric methods. Hahn, Todd and Van der Klaauw (1999) recall Fan's (1992) assertion that standard kernel estimators exhibit a poor performance at boundaries and suggest instead the use of local linear regression which is more rate-efficient and whose bias does not depend on the design density of the data.[18]

Similarly, one can solve a similar problem for values of $b > \bar{B}$.

$$\min_{\{\alpha_+(b),\beta_+(b)\}} \sum_{i=1}^{N}(y_i - \alpha(b) - \beta(b)b_i)^2 \cdot \mathbf{1}(b_i > \hat{B}) \cdot \mathbf{K}\left(\frac{b - b_i}{h}\right) \tag{19}$$

The solutions to these problems, evaluated at $b = \bar{B}$, can be used to calculate $\hat{y}^+ = \alpha_+(\bar{B}) + \beta_+(\bar{B}) \cdot \bar{B}$ and $\hat{y}^- = \alpha_-(\bar{B}) + \beta_-(\bar{B}) \cdot \bar{B}$. By considering outcome x instead of y, one can do the same to estimate $\hat{x}^+$ and $\hat{x}^-$, so that the Wald estimator in equation 6 can be calculated.

5.2 Estimation of quantile impacts

Similarly, we can attempt to estimate the conditional τ-quantiles through local linear quantile regression. Local linear quantile regression has been discussed by Yu and Jones (1998). This methodology combines the convenience of local methods discussed in the previous sub-section, with the shifted emphasis towards quantiles of the distribution of outcomes conditional on year of birth. The latter is of particular importance for the analysis of a SLA reform, which has a very distinct impact on the lower segments of the outcome's distribution. Yu and Jones (1998) build on the seminal work of Koenker and Bassett (1978) to study nonparametric regression quantile estimation by kernel-weighted local fitting. I reproduce this approach by solving:

$$\min_{\{\alpha^\tau_-(b),\beta^\tau_-(b)\}} \sum_{i=1}^{N}(y_i - \alpha^\tau - \beta^\tau b_i) \cdot (\tau - \mathbf{1}(y_i - \alpha^\tau - \beta^\tau b_i \leq 0)) \cdot \mathbf{1}(b_i \leq \bar{B}) \cdot \mathbf{K}\left(\frac{b - b_i}{h}\right), \tag{20}$$

[18] Porter (1993) has also analysed the bias problem in estimation of $y^+ - y^-$. Nadaraya-Watson based estimates have a bias at the boundary of $O(h)$, as opposed to $O(h^2)$ in the interior of the covariate support. He suggests to approach this difference as the partially linear estimator or Robinson (1988), i.e $y(b) = m(b) + (y^+ - y^-) \cdot 1(b > \bar{B}) + \mu$ so that one can estimate $y(b) - E[y|b] = (y^+ - y^-) \cdot 1(b > \bar{B})$. He also considers how polynomial correction (of order equal or higher than one) in local polynomial regression can accomplish bias reduction.

where the minimand can be rewritten as:

$$\sum_{i=1}^{N}(LK_i(b)^{1/2}\cdot y_i-\alpha^{\tau}(b)\cdot LK_i^{1/2}(b)-\beta^{\tau}(b)\cdot LK_i(b)^{1/2}\cdot b_i)$$
$$\cdot(\tau-\mathbf{1}(y_i-\alpha^{\tau}(b)\cdot LK_i(b)^{1/2}-\beta^{\tau}(b)\cdot LK_i(b)^{1/2}\cdot b_i<0)). \tag{21}$$

Solving this problem for $b=\bar{B}$ produces a simple weighted quantile regression estimator that can be used to construct $\lim_{\epsilon\to 0}Q_{Y|B}(\tau,\bar{B}-\epsilon)$. A similar problem can be stated to get to $\lim_{\epsilon\to 0}Q_{Y|B}(\tau,\bar{B}+\epsilon)$ and eventually construct the quantile impact $\lim_{\epsilon\to 0}(Q_{Y|B}(\tau,\bar{B}+\epsilon)-Q_{Y|B}(\tau,\bar{B}-\epsilon))$. The X-chord of interest can be approximated by the quotient:

$$\frac{\lim_{\epsilon\to 0}(Q_{Y|B}(\tau,\bar{B}+\epsilon)-Q_{Y|B}(\tau,\bar{B}-\epsilon))}{\lim_{\epsilon\to 0}(Q_{X|B}(\tau,\bar{B}+\epsilon)-Q_{X|B}(\tau,\bar{B}-\epsilon))}.$$

5.3 Simple IV estimation and model overidentification

A simple strategy to estimate the average impact of an additional year of schooling on outcomes is to run simple IV regressions using an indicator for the new SLA being applicable. In order to approximate as much as possible the desirable features of the discontinuity design, one can select the sample of (b_i,x_i,y_i) observations to those which satisfy $b_i=1933$ or $b_i=1934$. IV estimation, based on the moment condition $E[b_i\cdot(y_i-\alpha-\beta\cdot x_i)]=0$ in the close neighbourhood of $\bar{B}$ will provide:

$$\hat{\beta}_{IV}=\left(\sum_{b_i\in C}(x_i-\bar{x})(b_i-\bar{b})\right)^{-1}\cdot\left(\sum_{b_i\in C}(y_i-\bar{y})(b_i-\bar{b})\right), \tag{22}$$

where $C=\{1933,1934\}$ is the 'discontinuity sample'.

We can think of this estimator as an unsophisticated RDD for the case with discrete birth years, which does not utilise information about other values of b_i other than the strictly necessary, though at the cost of ignoring valuable information elsewhere.

The existence of a natural experiment such as the raising of the SLA provides a good opportunity to assess the performance of more readily available, though less plausible instruments 'a priori'. Parental background characteristics have been often used as instruments for an individual's schooling level, regarding estimation of the effect of exogenous changes in her schooling on subsequent labour market outcomes such as wages. Carneiro and Heckman (2002) have discussed the problems with this type of assumption and how correlation with unobserved ability can explain why instrumental variable estimates based on invalid instruments can produce higher estimates than OLS. In our case, one might be induced to think of grandparent's background data such as social class (education is unfortunately unavailable for grandparents of the NCDS cohort members) to instrument parental education.

To keep discussion as simple as possible, let us focus on the homogeneous effects case. Under the assumption that the *grand-parent's* social class measure z_i is a good instrument, x_i and z_i are correlated and $E[y_i - \beta x_i | z_i] = 0$. Consider again the RDD sample C which only includes values of b_i located arbitrarily close to the discontinuity point $\bar{B}$. For the relevant subpopulation, it is plausible to assume that the variables B and Z are mutually independent. [19] In this case, the estimator that uses both B and Z as instruments (in matrix notation) delivers: $\hat{\beta} = \pi\beta_B + (1-\pi)\beta_Z$. This depicts the estimator as a weighted average of two separate IV estimators, $\hat{\beta}_B = (B'X)^{-1}B'Y$ and $\hat{\beta}_Z = (Z'X)^{-1}Z'Y$, which correspond to the alternative estimates of β using each instrument separately. The weight given to the estimator based on the discontinuity design is given by $\pi = \frac{X'M_BX}{X'M_BX+X'M_ZX}$, where $M_B = B(B'B)^{-1}B'$ and $M_Z = Z(Z'Z)^{-1}Z'$. The estimator based on the overidentified model exploits a great more deal of information than the RDD one, basically giving a larger weight to the instrument that is more correlated with the X education variable. However, the strict RDD estimator is known to be consistent under the basic assumptions but less efficient. The formula for the asymptotic bias will be $plim\hat{\beta} - \beta = (1-\pi)(plim\hat{\beta}_Z - \beta)$. It is then possible to test the identifying restriction $E[Y - X\beta|Z] = 0$ using a Hausman-type test based on the comparison of $\hat{\beta}$ (consistent and efficient under the null) and $\hat{\beta}_B$ (less efficient, but consistent also under the alternative). Another possible approach is to carry out a Sargan-type test of the overidentifying restrictions. Rejection must imply invalidity of Z as an instrument.

6 A brief history of the 1947 school leaving age reform

The number of pupils in state maintained schools in England and Wales increased by two million between the end of the Second World War and the 1970s. The Education Act of 1944 was instrumental to this massive change. This Act's main ambition was to ensure that every child would benefit from secondary level education by ensuring that the much-postponed raising of the school leaving age would become a reality, as opposed to earlier measures enacted by the Parliament in 1918, 1921 and 1936. This section details some of the key features of this legislation and the increase in the SLA it brought about.

The Act's basic principle was that of a secondary education free for all, though insisting that this should be made available in different types of school among which selective grammar school should remain pre-eminent vs secondary modern schools (the idea being that each type of school was needed to match the 'nature' of the child). The raising of the SLA was a cornerstone of this project.

It could be argued that the Education Act of 1944 was the most socially progressive measure carried through actually during the war, building on the (at the time) radical proposals of the Beveridge Report. The Conservative majority in the wartime coalition government, headed by Winston Churchill, had indeed committed itself to action for social reform but only once hostilities ended. It has been argued that the Conservative support for the Education Act

[19] This is just for expositional clarity, but it is also the natural outcome of the RDD setting.

(Simon, 1991) was partly due to the wish to divert the public from Beveridge's proposed social security reforms. Although the war had led to a weakened economy, with an acute shortage for building materials and an extraordinary demand for housing, it also contributed to strengthen the political will to reconstruct and expand the school system.

As a result of this public pressure, the wartime all-party Coalition Government decided that the school-leaving age would be raised from 14 to 15 no later than 1 April 1947, leading to an increase in the number of pupils in schools of about 400,000. However, special power was also given to the Minister of Education to delay the raise in SLA for two years. Clause 35 also provided for it to be raised to 16 by Order in Council 'as soon as it has become practicable'.[20]

The Ministerial Committee on Economic Planning, reported that raising the school age on 1 April would mean a 370,000 strong direct loss to the national labour force, a serious matter at a time of overstrained resources and labour shortages.[21] The White Paper of 1943 estimated a total additional expenditure of 67.4 million GBP. Additionally, direct implementation costs of the reform had to be added to the short-term opportunity costs arising from a reduced workforce. In preparation for the raising of the SLA, Local Education Authorities (LEAs) were requested to complete returns specifying the investments required to accommodate the additional age group. Following this consultation, the new government presided by Clement Attlee set out the Hutting Operation for the Raising of the School Age (HORSA),[22] but soon faced the dilemma of whether to devote resources to wider reorganization issues or focusing instead or just meet the pressing short term objective of raising the SLA.

By the summer of 1946, full implementation of the Act seemed more unlikely than ever, particularly regarding the target of reducing class sizes below 30, the construction of further education colleges and the raising of the SLA to 16.[23] Instead, the focus turned to the more immediate targets of free-for-all secondary education, free milk and meals service, together with the more modest raising of the SLA to 15. At the time, Ellen Wilkinson, the first Labour minister for Education, had argued that a further postponement would particularly affect children whose education had been most seriously interrupted by the war, particularly those born to working class parents.

Finally, after internal cabinet discussions, the leaving age was raised to 15 in April 1947. However, 1947 turned out to be a particularly difficult year on a number of fronts[24], leading to

[20]Besides, the parent's legal duty was changed from that of causing his child to receive 'efficient elementary instruction in reading, writing and arithmetic' to a duty to cause his child to receive 'efficient full-time education suitable to his age, aptitude and ability either by regular attendance at school or otherwise' (Maclure, 1965).

[21]This aspect has been carefully analysed by O'Keeffe (1975). He documents a substantial increase in the weekly wage rates of juvenile workers coinciding with the raising of the SLA.

[22]Huts, provided by the Ministry of Works, were finally accepted in spite of criticism by LEAs and architects (on the grounds of substandard in accommodation, heating costs, being erected on playing space) because of the fact that rejection would imply gross overcrowding (LEAs had plans to hire church halls). A total of 5000 HORSA classrooms were erected, being equipped with 'close on half a million SFORSA -School Furniture Operation for the Raising of the School Age- chairs, tables, desks, stoves...' (Dent, 1970).

[23]Although the Education Act of 1944 targeted the age of 16 as the desirable leaving age, indefinite postponement was possible without even asking parliament for legislation. This further reform was only carried out a whole generation -in 1973- after the passage of the 1944 Act.

[24]A crisis in coal production, exceptionally freezing winter, sterling convertibility crisis and the financial hardship

a shift in spending priorities and financial cuts that limited progress in the full implementation agenda. On the other hand, the fact that most of those who attained the age of 14 just before April 1947 left school before Easter helped ensure that enough places were finally provided so that no children would be excluded from school. Besides, due to the declining fertility during the thirties as seen in table 4, 1947 turned out to be an optimal year in which to raise the SLA from the government's cost perspective.

Table 4: Live births in England and Wales

Year	Cohort's MSLA	No.Births
1925	14	710,582
1930	14	648,811
1931	14	632,081
1932	14	613,972
1933	14	580,413
1934	15	597,642
1935	15	598,756
1936	15	605,292
1937	15	610,557

Source: Number of births from ONS historical statistics.

One should interpret this phenomenon with caution, as other cohort effects can affect the interpretation of the estimates if authorities decide to enforce a higher SLA when the cohort size is considered sufficiently small, as cohort size may have an independent effect on lifetime earnings.

Table 5: Pupils and teachers in secondary schools

	(1)	(2)	(3)	(4)	(5)	(6)
Year	Pupils* (Total)	Teachers	Pupil/Teacher	Modern*	Grammar*	(6) Technical*
1946	1,268	58,455	21.7	719	488	59
1947	1,334	65,476	20.4	763	504	66
1948	1,544	71,112	21.7	960	511	71
1949	1,654	76,999	21.5	1058	523	72

Source: DES/DfES Annual Reports and Statistics of Education, 1946-1977. (1) Pupils in maintained secondary schools in England and Wales. (2) Number of teachers as in (1). (3) Pupil/teacher ratio. (4,5,6) Number of pupils as in (1) in Modern Secondary, Grammar and Technical schools, respectively. * indicates thousands of students.

As table 5 shows, the reform brought about a large increase in the number of students in secondary schools, largely concentrated amongst the school types for which high ability was not

caused by American postwar loan.

an entry requirement (modern secondary and technical schools). Thus, increased participation took place mostly at the bottom of the ability distribution, as one would expect from this type of reform. This increase in participation was parallel to an increase in the number of teachers, implying a relatively stable pupil/teacher ratio over the period considered.[25]

7 Estimation results

7.1 The 1947 Reform and changes in schooling

Table 6 describes changes in the distribution of the ages at which fathers and mothers of NCDS cohort members finished their education, according to their year of birth. Since fathers tend to be older than mothers, the distribution of birth years is concentrated in the period prior to the reform for the former, whereas the latter are more evenly spread across the 1925-1939 range depicted here, though it is still true that the mode, located in 1931, precedes the cohort that was first affected by the reform.

For fathers, the shift in the schooling distribution between the 1933 and 1934 cohorts is most striking, with a considerable increase in mean and median levels. The change for the first quartile only becomes evident in 1935, probably suggesting limited reform enforceability at first or perhaps misreporting of either age or schooling.[26] The third quartile remains constant over the period, suggesting that the raising of the SLA had reduced spillover effects in the sense of inducing more schooling (as measured by years) by those at the top of the distribution. Following 1935, we can see how the reform led to at least three quarters of the father's population leaving school no earlier than 15. Everything combines to produce a lower degree of inequality in schooling inputs as measured by years. For mothers, the picture is even simpler, with the post-reform cohort (1934) increasing its mean, first quartile and median schooling levels with respect to the the pre-reform cohort (1933).

Figure 2 analyses the changes in NCDS mean and median father's schooling induced by the raising of the SLA. Estimates for mean schooling are based on local linear regressions with a three-year bandwidth and an Epanechnikov kernel function.[27] For the 1933 right limit ($\hat{x}^{+}$), this effectively implies an out of sample prediction based only on values corresponding to birth years equal or higher than 1934.

The relationship between father's year of birth and father's schooling is clearly neither monotonic nor continuous. Schooling increases with year of birth for earlier birth year values as younger cohorts tend to get more education. However, this profile reaches a maximum in 1929, possibly reflecting idiosyncracies of younger fathers (more likely to be less academically able). This composition effect counteracts the secular schooling-cohort pattern in the population and

[25] However, a blip can be found at the exact year of the reform.

[26] Unfortunately, lacking the parents' precise age in months makes some degree of error unavoidable. This combines with the fact that schooling is only available as an integer.

[27] Estimates for the conditional median are also based on local linear quantile regression, with weights implied by the above mentioned kernels.

Table 6: Parent's year of birth and schooling

	Fathers						Mothers					
Birth year	Mean	S.D.	Freq.	P25	P50	P75	Mean	S.D.	Freq.	P25	P50	P75
1925	14.86	1.62	472	14	14	15	14.77	1.37	385	14	14	15
1926	14.77	1.68	576	14	14	15	14.70	1.49	440	14	14	15
1927	14.85	1.79	569	14	14	15	14.77	1.59	444	14	14	15
1928	14.86	1.68	658	14	14	15	14.78	1.48	553	14	14	15
1929	14.85	1.71	690	14	14	15	14.79	1.51	671	14	14	15
1930	14.93	1.74	728	14	14	15	14.83	1.50	660	14	14	15
1931	14.90	1.67	679	14	14	15	14.86	1.55	805	14	14	15
1932	14.78	1.44	673	14	14	15	14.91	1.57	721	14	14	16
1933	14.73	1.41	606	14	14	15	14.80	1.32	730	14	14	15
1934	15.07	1.23	509	14	15	15	15.19	1.18	715	15	15	15
1935	15.04	0.85	445	15	15	15	15.23	1.01	705	15	15	15
1936	15.11	0.99	333	15	15	15	15.15	0.91	653	15	15	15
1937	15.11	0.96	200	15	15	15	15.20	0.91	582	15	15	15
1938	15.00	0.72	113	15	15	15	15.11	0.73	472	15	15	15
1939	15.03	0.74	59	15	15	15	15.08	0.64	282	15	15	15

NOTES: Selected statistics for NCDS member's father's and mother's age at which finished schooling, by year of birth.

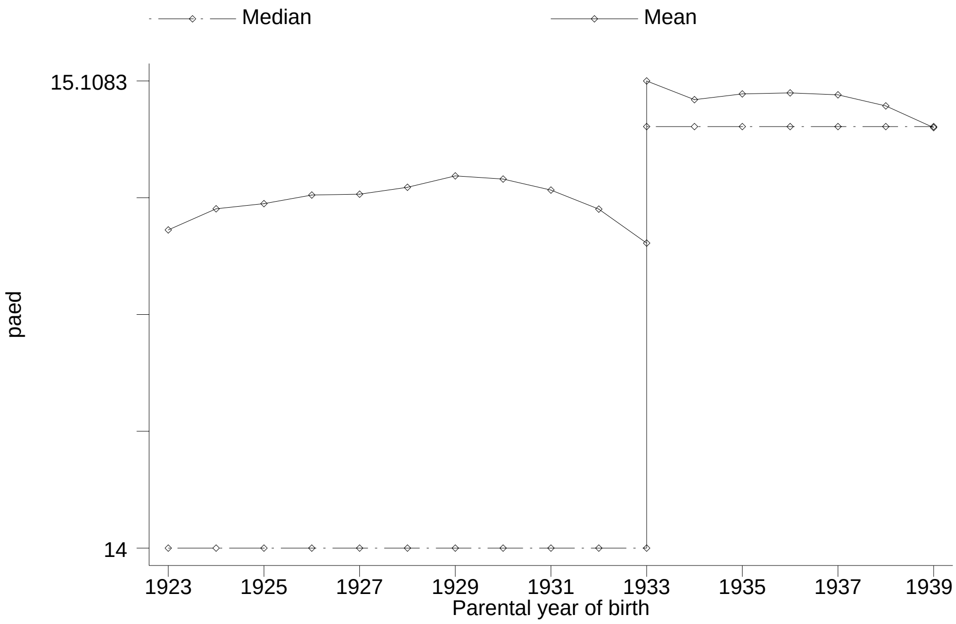
persists within the sample of parents of children born in 1958, only interrupted by the clear discontinuity brought about by the SLA reform. After that, mean schooling levels continue to decrease with year of birth.

It is important to note how larger bandwidths in this sample will tend to produce lower estimates of the effect of the SLA reform on schooling ($\hat{x}^+ - \hat{x}^-$), given the declining profile of schooling for the birth years surrounding the reform. Regarding the effect on the median level of schooling, we can see how the reform brought about a radical one-year shift after 1933 on an unchanged level of the previous years. To put this in perspective, if all these individuals ended up benefitting in the long term as a result of higher schooling, a ballot among the relevant population would have received majority support. Reality is obviously more complicated, and one could argue why a policy based on compulsory schooling –that is, imposing a restriction on behaviour– could ever improve individual outcomes. Credit constraints are often alluded as partly responsible for this divergence, although alternative explanations such as extreme forms of discounting or limited rationality have been put forward.[28] In section 2 I discussed a form of bounded rationality which drew on unforeseen altruism towards future generations when making schooling decisions as a young person.

For mothers, the estimated profile of schooling by year of birth shown in figure 3 parallels that of fathers with the difference that the decline in schooling appears to occur only in the years surrounding the reform, as would correspond to the fact that mothers tend to be younger than fathers. Therefore, any compositional effects induced by selection based on having a child born in 1958 become visible only afterwards. Again, estimates are suggestive of a substantial increase

[28]See Oreopoulos (2002) for an analysis of potential reasons for why dropouts drop out 'too soon'.

Figure 2: Father's year of birth and father's schooling.



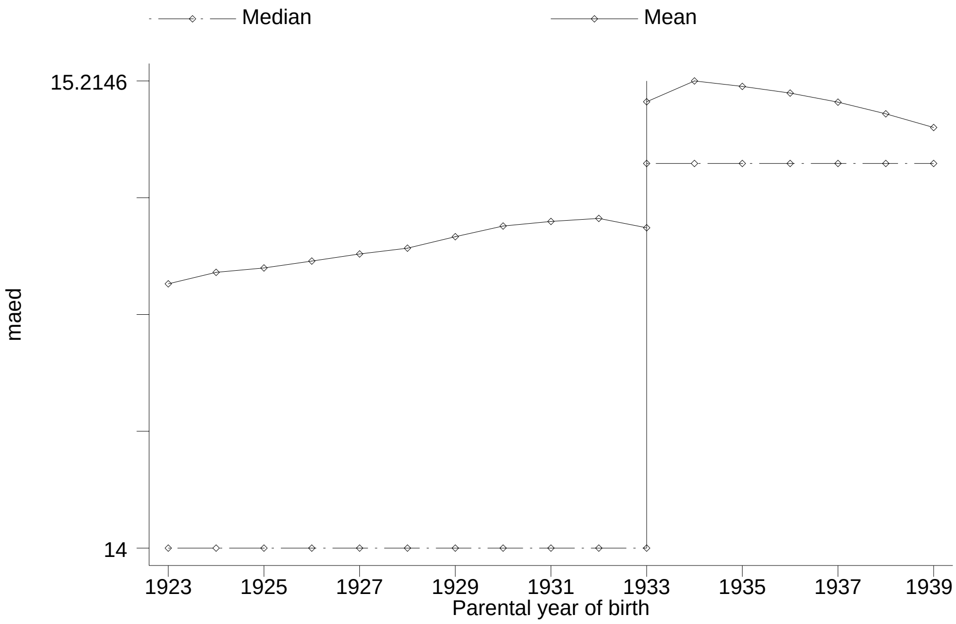
NOTES: Regressions of age father completed schooling on father's year of birth. Local linear regression for the mean. Local linear quantile regression for median. Weights given by Epanechnikov kernels accounting for discontinuity in 1933, bandwidth=3. $[1923, 1933], (1933, 1939]$.

in mean schooling levels, which appear to be massively accounted for an increase in the lower half of the distribution, as depicted by the median.

I complement this evidence with OLS estimates of the effect of the SLA in table 7 using the smallest possible bandwidth, by restricting the sample to those parents born either in 1933 or 1934. Regressions suggest that the SLA induced an average increase of 0.36 (s.e.=0.09) years of schooling for men and 0.43 (s.e.=0.07) for women. These estimates are very robust to the inclusion of additional controls, including grandfather's socio-economic status (neither income nor education is available for them), region of residence or number of siblings (only available for mothers). Certainly, these variables improve the explanatory power of the schooling regressions, but the fact that the coefficient on the post-reform birth cohort (1934) hardly varies suggests a large degree of independence of the reform/year of birth instrument with respect to other potential covariates.

Another aspect worth investigating is how individual responses to the raising of the SLA vary according to observable characteristics, in other words, whether the reform has heterogeneous effects on schooling. This is important for understanding for which subgroup in the parents' population the RDD-IV estimator identifies the causal effect of education. Above all, we would like to carry out this exercise looking at the role of parental ability, but this is not observable in

Figure 3: Mother's year of birth and mother's schooling.



NOTES: Regressions of age mother completed schooling on mother's year of birth. Local linear regression for the mean. Local linear quantile regression for median. Weights given by Epanechnikov kernels accounting for discontinuity in 1933, bandwidth=3. $[1923, 1933], (1933, 1939]$.

Table 7: The effects of the SLA on schooling levels.

	Father's schooling					Mother's schooling					
Pa's Birth=1934	0.359	0.366	0.384	0.342	0.364	0.429	0.425	0.426	0.436	0.387	0.388
St. Error	*0.088*	*0.084*	*0.083*	*0.079*	*0.078*	*0.069*	*0.065*	*0.064*	*0.063*	*0.066*	*0.066*
Grandfather's SES	x	√	√	x	√	x	√	√	√	x	√
Parent's region	x	x	√	x	x	x	x	√	√	x	x
Parent's siblings	na	na	na	na	na	x	x	x	√	x	x
R-Squared	0.018	0.12	0.13	0.016	0.12	0.03	0.14	0.15	0.19	0.02	0.17
Observations	875	875	875	1115	1033	1261	1261	1261	1261	1445	1445
Largest sample	x	x	√	√	√	x	x	x	√	x	√

NOTES: Sample of NCDS parents born in 1933 or 1934. OLS regressions of parental schooling on parent's background information.

the data. Table 8 presents the coefficients on the interaction between the SLA reform indicator with parent's background characteristics.

I look at the interactions with grandparent's socioeconomic group and, only for mothers, with parent's number of siblings. The results indicate that none of the interactions are significant for

Table 8: Differentiated schooling responses to the SLA reform

	Father's schooling		Mother's schooling	
	Coef.	SE	Coef.	SE
$1(b_i = 1934)$*(SEG I)	n.s.		n.s.	
$1(b_i = 1934)$*(SEG II)	n.s.		n.s.	
$1(b_i = 1934)$*(SEG III)	n.s.		0.627	(0.216)
$1(b_i = 1934)$*(SEG IV)	n.s.		0.339	(0.179)
$1(b_i = 1934)$*(Mother's No.siblings)	n.a.		-0.049	(0.024)

NOTES: Regression of parental education on post-reform dummy interacted with background variables (grandparent's SEG and number of siblings (only available for mother): $x_i = \gamma_b b_i + \gamma_z z_i + \gamma_{bz} b_i \cdot z_i + \eta_i$. Coefficients and standard errors for interactions displayed when significant at 10 percent (n.s. otherwise). Discontinuity sample of NCDS58 parents born in 1933 or 1934.

men, suggesting the SLA reform had a very uniform impact across the male population. One could argue that this finding would be very difficult to reproduce with an increase in the current SLA from 16 to 17.

The picture for mothers is quite different. In terms of grandfather's social class interactions, it is those with skilled and semi-skilled parents who are most likely to increase their schooling levels. The negative coefficient on the interaction with number of siblings indicates that those mothers in larger families were less likely to increase their schooling levels, compared to the rest. Thus, it looks as if it was able women in middle class smaller families who experienced a more acute rise in schooling as a result of the higher SLA.

These findings contrast with those in Meghir and Palme (2003), who show that the Swedish SLA reform mostly increased the attainment of those with unskilled fathers, but theirs is a more recent cohort, born in 1948, suggesting that there would be fewer children of parents with intermediate backgrounds not staying on after the statutory SLA. In the British case, even though the largest impact was to shift those who would have stopped at the old SLA, at the time of the reform there was still a large group of relatively able and middle class women in this group. In the case of men, both effects appear to cancel out though we should keep in mind that in our sample, the population of fathers differs from that of mothers for the simple reason that the latter tend to be younger and sampling is based on a single cohort of children.

7.2 Parent's earnings effects

Before turning to the effects of parental schooling on children's outcomes, it is worthwhile to examine whether we find any type of intragenerational impact in the form of higher earnings for fathers. This provides us both with a check on our methodology *vis a vis* existing studies on the returns to schooling and also informs us about a crucial mediating mechanism in the intergenerational transmission of human capital.

Table 9 shows the mean and standard deviation for the log of weekly earnings of fathers, as measured in 1974, by year of birth. Earnings appear to increase between 1933 and 1934 by 2.3 percent, whereas the dispersion of wages decreases. This could well be a result of the SLA reform, which reduces the education inequality and raises the earnings capacity of those at the bottom of the distribution.

Table 9: Father's year of birth and log earnings

Birth year	Mean	S.D.	Freq.
1923	3.507	.363	286
1924	3.524	.380	336
1925	3.540	.359	368
1926	3.550	.381	440
1927	3.549	.342	435
1928	3.573	.391	511
1929	3.588	.358	552
1930	3.561	.359	571
1931	3.546	.382	543
1932	3.540	.346	520
1933: $b_i = 0$	3.553	.351	479
1934: $b_i = 1$	3.576	.328	386
1935	3.580	.310	370
1936	3.511	.321	259
1937	3.581	.337	159
1938	3.511	.371	86
1939	3.458	.472	46

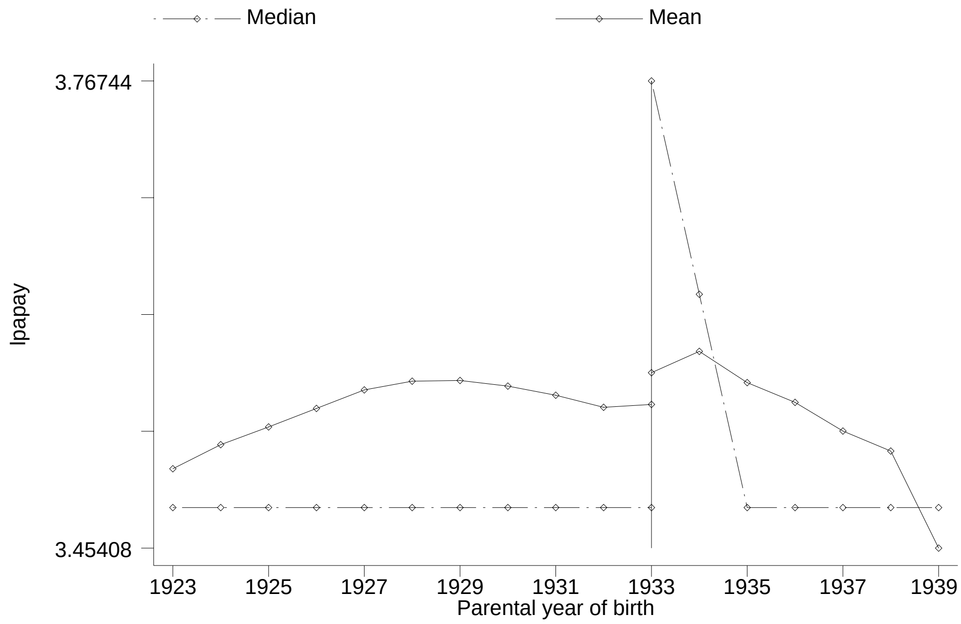
NOTE: Summary statistics of father's log weekly earnings in 1974, by year of birth.

Figure 4 depicts the estimated average log earnings profile by year of birth, as well as the median log earnings. Because the earnings information is elicited from earnings intervals, the smooth variation in the mean by birth year is not reproduced by the median. However, whereas the estimated discontinuity is moderate (and positive) for the mean, it is far more accentuated for the median earnings.[29] With a positively skewed distribution, there are reasons why the median may be a more interesting moment to look at. Overall, this result suggests that the SLA reform has important distributional effects which may go neglected by only investigating the mean.

These results are reinforced by estimates based on the discontinuity sample in table 10, in which potential overidentifying restrictions are also explored. Identification is rejected in the specifications which use all background information as additional instruments (region, SES background and number of siblings). For fathers, if we do not exclude father's region, the combined use of SES and reform do not lead to rejection of the identification assumption, leading to a relatively precise estimate of the average return to a year of schooling of 0.104 (col. 3,

[29]The value for right limit in 1933 is indeed an out-of-sample estimate, based on 1934 and later years, but even if we take the 1934 estimate, the discontinuities are still positive and large.

Figure 4: Father's year of birth and father's earnings.



NOTES: Regressions of log father's earnings in 1974 on father's age completed schooling. Local linear regression for the mean. Local linear quantile regression for median. Weights given by Epanechnikov kernels accounting for discontinuity in 1933, bandwidth=3. $[1923, 1933], (1933, 1939]$. Earnings data available in brackets, with middle point used in regression.

s.e.=0.026). However, if we only rely on the reform variation estimates are lower and more imprecise (cols 3 and 4). The OLS estimate without additional controls in column 5 provides a significant figure – 0.071 (0.009) – which lies halfway in between the (not rejected) overidentified estimate and the exactly identified one. This is reassuring in the sense that OLS estimates with a reasonable degree of controls are not far from being an average of IV estimates that capture the impact of the reform on different subpopulations.

Regarding the estimates for mothers, the qualitative results are almost identical, although we should be cautious about the low degree of female participation in the labour market in this cohort. Point estimates are, as often found, higher than for men.

Finally, the estimate for the earnings return to schooling based on the local linear regression approach to estimating the boundaries at the discontinuity point gives a value of 0.10. This is reported in the first row of table 11, alongside other intergenerational effects. I have calculated the bootstrap for this estimator and the resulting standard error is considerably large (0.25). The corresponding estimate for the median impact on log earnings relative to the change in median schooling is 0.29, and the standard error of the bootstrap is 0.13, suggesting a significant effect on median earnings. This is consistent with findings in figure 4, leaving little doubt about the

Table 10: Earnings returns to schooling: Parents.

	Father's earnings					Mother's earnings				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Schooling	0.085	0.104	0.032	0.051	0.071	0.099	0.116	0.085	0.132	0.096
St. Error	*0.024*	*0.026*	*0.079*	*0.093*	*0.009*	*0.052*	*0.065*	*0.118*	*0.125*	*0.018*
Controls										
SES	x	x	√	x	x	x	x	√	x	√
Region	x	√	√	x	x	x	√	√	x	√
Siblings	na	na	na	na	na	x	√	√	x	√
Instruments										
Reform	√	√	√	√	x	√	√	√	√	x
SES	√	√	x	x	x	√	√	√	x	x
Region	√	√	√	x	x	√	√	√	x	x
Siblings	na	na	na	na	na	√	√	√	x	x
Overid. test: p-value	0.003	0.477	Ex.Id	Ex.Id	na	0.0004	0.215	Ex.Id	Ex.Id	na
Observations	666	666	666	666	666	803	803	803	803	803

NOTES: Sample of NCDS parents born in 1933 or 1934. IV regressions of log earnings in 1974 on schooling (and other variables as denoted in panel of controls). Schooling instrumented with variables in instrument panel as specified. Cols. (5) denote OLS estimates without additional controls. Overidentification test p-values for null hypothesis of valid instruments.

presence of positive intragenerational effects from higher schooling as induced by the SLA.

7.3 Intergenerational effects of schooling

Turning our attention now to intergenerational effects, tables 13 and 14 in the appendix summarise the estimation results for the effects of father and mother schooling, respectively, on their offspring's outcomes for a range of variables. Estimates are run separately for boys and girls. In all cases the sample is restricted to the smallest possible bandwidth around the year the SLA was raised (1933 and 1934). The first of results presents the standard OLS estimates, followed by the exactly identified IV estimates using the post-reform indicator as sole instrument. A final set of results produces the IV estimates from the overidentified model using grandparent's social class, accompanied by the p-value resulting from an overidentification test. Whereas most OLS produce statistically significant results of the expected sign, the exactly identified IV hardly ever yields such result. In fact, this is never the case for father's set of estimates, and only child bearing at 33 appears to be significantly affected by mother's schooling changes induced by the reform. In this case, boys appear to be more likely to have children whereas the opposite is true for girls. Interestingly, this effect was not captured in simple OLS regressions.

The estimates based on the overidentified IV are only sound for sufficiently large values of the p-value. When this is the case, we tend to observe the typical result with IV estimates providing larger point estimates than OLS, and in most cases larger than the exactly identified IV. Given that the overidentified IV estimator is a mixture of IV estimates based on the different instruments, one could argue that different instruments induce different types of changes in schooling and can therefore produce non-identical estimates. In that case, each IV would identify the average impact of schooling within a particular subpopulation for whom changes in the instrument induce higher levels of schooling. For example, the SLA reform mainly induced higher schooling by those at the bottom of the ability distribution, which is a very different group from that induced by having grandparents with a higher socioeconomic status.

One should regard the grandfather's social class instrument with caution. For father's schooling, overidentification cannot be rejected in very counted occasions. For mother's schooling, rejection is slightly less frequent, suggesting some positive impacts on general ability at 11 for girls, reading scores at 16 for boys, interest in boys and girls at 11, better qualifications performance for both and higher wages for boys.

In table 11 I provide some selected results for the mean and median Wald estimator for the impact of schooling based on semi-parametric estimation of the boundary points, as opposed to simple IVs using only the discontinuity sample. The extent of statistically significant mean impacts is very reduced. Standard errors, as mentioned above, are calculated using bootstrapping.[30]

Table 11: RDD-Wald estimates for effect of parental schooling

Dependent variable	Individual	Regressor	Mean Coef	Mean SE		Median Coef	Median SE		Observations
Father's log earnings	Father	Father's School	0.10	0.25		0.29	0.13	*	2266
Mathematics score at 16	Son	Father's School	4.78	7.27		4.00	2.40		1249
Reading score at 16	Son	Father's School	1.63	6.83		2.00	1.55		1253
	Son	Mother's School	2.16	3.42		2.00	1.19		1545
Arithmetic score at 7	Son	Father's School	0.94	2.98		2.00	1.03	*	1362
Social adjustment at 7	Son	Father's School	-3.77	3.27		-4.00	2.25		1363
	Son	Mother's School	3.34	4.12		4.00	1.48	*	1719
Number of O-levels	Son	Father's School	2.02	1.72		4.00	1.75	*	909
	Son	Mother's School	0.25	0.85		1.00	1.10		1163
Social class=Professional	Son	Father's School	0.21	0.68		na	na		910
	Son	Mother's School	0.22	0.11	*	na	na		1163
Log hourly earnings at 33	Son	Father's School	0.17	6.51		0.19	0.13		781
Social adjustment at 11	Daughter	Mother's School	6.73	12.35		3.00	1.20	*	2061

NOTES: Selected estimates of the effect of parental schooling on specified outcomes. Components of Wald estimator for impact on the mean calculated through weighted local linear regression, with bandwidth=3. Median regressions used for median impact estimates. Not applicable for dummy variable SC=Professional. Standard errors calculated by bootstrapping with 50 replications. Omitted estimates were insignificant. A higher value for social adjustment denotes a higher incidence of adjustment problems.

There is some evidence of the impact of father's schooling on median mathematics scores of the child at 7 and 16. This positive impact can also be found in the number of O-level

[30] In most cases, the bootstrap suggested the existence of a negative bias for *positive* outcomes, which strengthens the need to interpret these as worst-case estimates.

examinations passed by sons. Mother's schooling appears to induce some better median reading results, but these are not entirely significant. However, sons with mothers with an extra year of schooling are 22 percent more likely to have a professional occupation than their counterparts.

It is perhaps puzzling to find that mother schooling is associated with higher scores in the Social Adjustment Index, which implies a lower adjustability to social interactions. This occurs for sons at the age of 7 and daughters at the age of 11. This index is based on a questionnaire in which the child's teacher underlines descriptions most closely representing the child's behaviour in terms of potential emotional disturbance or social maladjustment in the school environment. It is impossible to tell whether this reflects a genuine negative impact of maternal schooling or just reflects a higher level of awareness by teachers of problems concerned the children of more educated mothers.

There is no clear conclusion regarding the possible existence of long term effects. For example, one cannot fully reject the hypothesis that father's schooling does not increase son's earnings at the age of 33.

7.4 Additional evidence: The next generation

The National Child Development Study provides yet another opportunity to investigate the intergenerational effects of schooling. In this case, identification is based on the detailed information which is available from NCDS cohort members, including their ability scores, and their own children. This is possible because when NCDS cohort members were interviewed at the age of 33, an attempt was made to assess their own children's performance in a series of tests as part of a wider set of measures of physical and mental well-being.[31]

Since children of NCDS cohort members are of different ages, it is likely that younger children appear to be less able in absolute terms than older ones. However, this effect can be attenuated by the fact that older children are more likely to be born to parents who became so at a much younger age, thus being more likely to display unfavourable characteristics. I try to control for this *unknown* effect of age with a sufficiently large order polynomial. Preliminary checks showed a relative uniformity of the effects of ability and qualifications by different ages, and are therefore not interacted in the preferred specification.

Table 12 displays the results from regressions of the reading and maths scores on parental attributes.[32] The correlation between mother's and children's scores is much higher than between father's and children. There are numerous interpretations consistent with this finding, ranging from theories about genetic transmission of traits to views about the 'quality' of the human resource that spends more time with child, traditionally the mother. The counterpart of this finding is how education, after controlling for ability, appears to matter much more for the father figure rather than the mother's. Again, there is a plethora of possible interpretations, but from

[31] Gregg and Machin (1997) have also examined the Peabody Individual Achievement Test Scores on maths and reading available from the children of the NCDS cohort.

[32] Only the attributes of the parent who is a cohort member are included, because the quality of information on the other parent is quite irregular.

Table 12: Cohort members education and child's attainment

	Mathematics		Reading		Mathematics		Reading	
Cohort member	Father	Mother	Father	Mother	Father	Mother	Father	Mother
Child=Male	2.117	2.086	13.552	13.205	-7.612	4.134	16.807	10.270
	7.503	5.194	10.054	7.652	8.411	6.012	12.301	8.912
Ability index (age 11)	1.073	2.325	1.682	3.746	1.393	2.404	1.887	3.733
(Parent)	0.621	0.453	0.873	0.562	0.776	0.473	1.059	0.670
Highest school qualification								
CSEs	2.911	1.039	3.144	1.572	2.400	1.492	1.618	0.936
	1.270	1.275	1.682	1.605	1.564	1.257	2.140	1.900
5- O-levels	2.365	2.797	4.931	2.892	2.666	2.508	4.751	2.636
	1.335	1.119	1.812	1.522	1.651	1.214	2.090	1.867
5+ O-levels	4.747	3.821	5.486	4.624	3.833	3.315	4.466	3.644
	1.729	1.290	2.110	1.729	2.128	1.375	2.596	2.135
A-level	4.590	3.885	7.749	5.633	5.608	2.989	7.481	4.296
	1.998	1.479	2.626	1.952	2.375	1.527	3.182	2.356
Highest post-school qualification								
Vocational: Low	1.934	-0.218	4.504	0.877	2.415	0.409	5.287	1.337
	1.041	0.761	1.520	0.967	1.201	0.842	1.765	1.138
Vocational: Mid	-0.620	-2.074	1.959	0.838	-0.819	-1.248	2.348	1.642
	1.243	1.428	1.356	2.172	1.407	1.752	1.579	2.460
Vocational: High	-1.133	0.895	2.111	1.616	-0.792	0.802	3.305	1.129
	1.282	0.939	1.869	1.241	1.465	1.001	2.306	1.353
Degree/Higher	1.283	1.299	2.002	-1.617	1.379	1.277	1.456	-1.414
	2.703	1.920	3.084	2.246	3.136	2.243	3.485	2.539
Gr.Fath Sch.					-0.306	0.103	0.556	-0.330
(Parent's father)					0.335	0.219	0.454	0.252
Gr.Moth Sch.					0.350	0.114	-0.051	0.888
(Parent's mother)					0.447	0.245	0.536	0.317
Observations	744	1433	738	1427	535	1071	532	1068
Number of households	514	954	510	949	362	715	364	713
R-Squared	0.76	0.75	0.69	0.71	0.77	0.76	0.70	0.72

NOTES: Ordinary least squares regression of PIAT maths and reading scores from cohort members' children tested in 1991. Scores range from 0 to 80 (sd approx 20). Other controls included: Square polynomial in child's age, birth order, parent's social adjustment index and all interactions with offspring's gender. Robust standard errors adjusted for family clusters.

an economic point of view, it is important to highlight that the transmission mechanisms for fathers may pass through higher earnings. I checked for this by running the same regressions including parental log hourly wages at the age of 33, with a coefficient on mathematics of 2.26 (s.e.=0.95) for fathers and 1.91 (s.e.=0.78) for mothers. The coefficients for regressions of reading scores were insignificant, and in all cases did the coefficients on schooling diminished, though still remained largely significant.

Some additional sensitivity analyses where carried out. When ability was excluded from the regressions, maternal schooling became relatively more important than paternal qualifications. I also studied the effect of qualifications on children being breastfed by mother. Interestingly, the effect of parental education and ability did not depend on whose ability or qualifications (father or mother) were considered. One standard deviation in the ability index brought about a 5.8 (s.e.=2) point increase in the probability of breastfeeding and holding more than 5 O-levels implied a 25 (s.e.=6) point higher probability relative to a parent without any school qualification.[33]

Given the unusual opportunity of holding together information on the educational attainment of three different generations in family, I tested whether grandparent's (NCDS members' parents) schooling had any visible impact on children (NCDS members' children) after controlling for parental (NCDS members) qualifications and ability. The last four columns of table 12 document the estimation results for the same specification including grandparent's schooling. Only mother's mother schooling is significant regarding the determination of reading scores, whereas none of the variables plays any role for mathematics scores. This points out to a strong mother-child transmission mechanism which appears to persist after two generations and is robust to the inclusion of ability.

8 Conclusions

This paper has analysed the problem of estimating the causal impact of schooling on intra and inter-generational outcomes using a reform raising the school leaving age as the main source of identifying variation. The SLA reform in Britain at the end of World War II triggered a massive change in schooling levels, both in absolute terms and relative to more recent SLA reforms. This paper has combined this natural experiment with a very unique data set to look at whether exogenous changes in schooling levels, such as those induced by the reform, led to any type of intergenerational impact in addition to other better documented intragenerational effects. In spite of the restrictions placed by the fact that, as many other reforms of its kind, the 1947 SLA reform was implemented on a nationwide scale, the features of a regression discontinuity design still apply. This paper has discussed at length how RDD techiques can be exploited to identify and estimate mean and quantile impacts of schooling on outcomes of interest. The key findings can be summarised as follows:

1. Increased schooling is indeed robustly associated with higher earnings. Besides, the reform appears to have brought about lower education and earnings inequality.

2. There is limited trace of mean intergenerational effects of parental schooling, if one takes conservative RDD estimates at face value. There are a few, though significant exceptions.

    - There seems to be some robust evidence for mother schooling reducing the extent of childbearing by daughters and raising a son's socioeconomic status and reading scores. However, mother's schooling also appears to increase the incidence of children's social

[33]All these results are available on request. Same controls as above.

adjustment problems, although this could possibly just differences in awareness by level of education.

- Father schooling has some impact on the median scores in mathematics for sons and also appears to improve attainment at the O-level stage.
- There is no clear effect of parental schooling on parental preferences towards the education of their children as measured from teacher's reports. This makes higher earnings resulting from higher schooling a prime suspect for explaining observed positive effects, most robustly associated to fathers.

3. Regarding the NCDS members and their own children, even after controlling for ability, parental qualification attainment at school has a significant, positive association with their children's mathematics and reading scores. Maternal grandmother's schooling is also linked to higher reading scores.

4. Additional OLS evidence on NCDS members and their parents presents a complex picture of how parental schooling is associated with children's attainment, through intermediate associations with income, fertility, interest in education and child's ability development. After controlling for all, a significant association remains and is stronger within same gender pairs (father-son and mother-daughter).

This paper has also exposed some of the problems involved in using more standard instruments derived from parental background, as suggested by rejection of the overidentification hypothesis in most occasions. It has also emphasized the importance of examining moments of the distribution of outcomes other than the mean. This a rather reasonable claim since the identifying source of variation, a SLA reform, has (and to some extent, is intended to have) an uneven impact on the population.

In addition to empirical and methodological findings, this paper has also proposed a framework in which to consider mechanisms of intergenerational spillovers of parental education for the purposes of policy design. This has brought attention to the problem that such spillovers are unlikely to be internalised by individuals at the time of making schooling decisions even if they are forward looking, benefits are private to families and there are no credit constraints, the key aspect being a rather plausible phenomenon of unanticipated altruism. Therefore, this opens a discussion about which are most effective policies to bring about welfare improvements. The conservative RDD estimates in this paper provide a lower bound for such positive impacts and the resulting *intergenerational multiplier*. It is left for further research a more detailed investigation of whether such impacts are explained by a larger availability of financial resources for children of more educated parents, or whether parental education improves overall efficiency in the family's human capital production technology.

References

[1] Angrist, J. and Lavy, V. (1999). "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 114, pp. 533-575.

[2] Ashenfelter, O. and Krueger, A. (1994). "Estimates of the Economic Returns to Schooling from a New Sample of Twins", *American Economic Review*, LXXXIV, pp. 1157-1173.

[3] Behrman, J. and Rosenzweig, M. (2002). "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" *American Economic Review*, 91 (1), pp. 323-334.

[4] Black, S. (1999). "Do Better Schools Matter? Parental Valuation of Elementary Education", *Quarterly Journal of Economics*, pp. 578-599.

[5] Black, S., Devereux, P. and Salvanes, K. (2003). "Is Education Inherited? Understanding Intergenerational Transmission of Human Capital". Unpublished manuscript.

[6] Cameron, S. and Heckman, J. (1998). "Life Cycle Schooling and Dynamic Selection Bias : Models and Evidence for Five Cohort of American Males", *Journal of Political Economy*, 106 (2), 262-333.

[7] Carneiro, P. and Heckman, J. (2002). "The Evidence on Credit Constraints in Post-Secondary Schooling", IZA Discussion Paper No. 518.

[8] Chernozhukov, V. and Hansen, C. (2001). "An IV Model of Quantile Treatment Effects", MIT Department of Economics Working Paper.

[9] Chesher, A. (2002a). "Instrumental Values", Cemmap Working Paper, CWP17/02.

[10] Chesher, A. (2002b). "Local Identification in Nonseparable Models", Cemmap Working Paper, CWP05/02. Forthcoming in *Econometrica*.

[11] Chevalier, A. (2003). "Parental Education and Childs Education: A Natural Experiment". Mimeo, University College Dublin.

[12] Clark, D. and Hsieh, C.-T. (2000). "Schooling and Labor Market Impact of the 1968 Nine-Year Education Program in Taiwan", Hong Kong Institute of Economics and Business Strategy Working Paper No. 1009.

[13] Currie, J. and Moretti, E. (2002). "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings and Longitudinal Data," NBER Working Paper No. 9360.

[14] Dent, H.C. (1970): *1870-1970. A Century of Growth in English Education*, Longman (London)

[15] Fan, Y. (1992). "Design Adaptive Nonparametric Regression", *Journal of the American Statistical Association*, 87, 998-1004.

[16] Fan, Y. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

[17] Gregg, P. and Machin, S. (1997). "Child Development and Success or Failure in the Youth Labour Market", in Blanchflower, D. and Freeman, R. (eds.) *Youth Employment and Joblessness in Advanced Countries*, National Bureau of Economic Research, Cambridge: MA.

[18] Hahn, J., Todd, P. and Van der Klaauw, W. (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, vol. 69 (1), pp. 201-209.

[19] Hahn, J., Todd, P. and Van der Klaauw, W. (1999). "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design." NBER Working Paper No 7131.

[20] Harmon, C. and Walker, I. (1995). "Estimates of the Economic Return to Schooling for the United Kingdom", *American Economic Review*, 85, pp. 1278-1296.

[21] Koenker, R. and Bassett, G. (1978). "Regression Quantiles", *Econometrica*, 46, 33-50.

[22] Maclure, J.S. (1965): *Educational Documents. England and Wales. 1816-1963.* Chapman and Hall (London).

[23] Meghir, C. and Palme, M. (2003). "Ability, Parental Background and Education Policy: Empirical Evidence from a Social Experiment", Institute of Fiscal Studies Working Paper WP03/05.

[24] O'Keefe, D. (1975). "Some Economic Aspects of Raising the School Leaving Age in England and Wales in 1947". *The Economic History Review*, Vol. 28, No. 3, 500-516.

[25] Oreopoulos, P. (2003). "Do Dropouts Drop Out too Soon?", mimeo, University of Toronto.

[26] Oreopoulos, P., Page, M. and Stevens, A. (2003). "Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling", NBER Working Paper No. 10164.

[27] Plug, E. (2002). "How do Parents Raise the Educational Attainment of Future Generations?", IZA Discussion Paper No. 652.

[28] Porter, J. (2003). "Estimation in the Regression Discontinuity Model", unpublished manuscript.

[29] Robinson, P. (1988). "Root-N-Consistent Semiparametric Regression", *Econometrica*, 56, 931-954.

[30] Sacerdote, B. (2002). "The Nature and Nurture of Economic Outcomes", *American Economic Review Papers and Proceedings*, 92, 344-348.

[31] Simon, B. (1991): *Education and the Social Order: 1940-1990.* Lawrence and Wishart (London).

[32] Thistlethwaite, D. and Campbell, D. (1960). "Regression Discontinuity Analysis: An Alternative to the Ex-Post Factor Experiment", *Journal of Educational Psychology*, 51, pp. 309-317.

[33] Van der Klaauw, W. (2002). "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach", *International Economic Review*, Vol. 43(4).

[34] Yu, K. and Jones, M. (1998). "Local Linear Quantile Regression", *Journal of the American Statistical Association*, No. 93, Vol. 441, 228-237.

Appendix

A Extensions to the basic model: Parent-child interactions

The natural extension of the model with unexpected altruism would be to allow for a level of investment q on the child's education by an adult. For example, consider that schooling x can be produced according to a combination of of children's effort e and parental investment q given by $x = \pi(q) \cdot e$ with disutility of effort now determined by $v(e) \cdot \phi(x_p)$. The child's problem can now be stated as:

$$V_k(x_p, c_k, q) = \max_e \left\{ u(c_k) - v(e) \cdot \phi(x_p) + \beta \cdot u(\psi(\pi(q) \cdot e)) \right\}, \tag{23}$$

which gives the first order condition:

$$v'(e_k) \cdot \phi(x_p) = \beta \cdot u'(\pi(q) \cdot e_k) \cdot \psi'(\pi(q) \cdot e_k) \cdot \pi(q), \tag{24}$$

and defines an implicit equation of the form $e_s = g_e(x_f, q)$. If the disutility of effort is an increasing, convex function, the marginal disutility will be an increasing function in effort. The marginal benefit, on the contrary, can be expected to be a non-increasing function of effort as the marginal utility of consumption is non increasing and $\psi(x)$ is concave. Raising parental education, under a decreasing function ϕ, implies a lower marginal cost of effort. The effect of a higher level of investment q can be thought of increasing the productivity of effort, and hence it is likely to increase the choice of e_k. However, if we assume a higher degree of substitutability in the production of schooling, for example, by considering $x_k = e_k + \pi(q)$, the first order condition becomes $v'(e_k) \cdot \phi(x_f) = \beta \cdot u'(\pi(q) + e_s) \cdot \psi'(\pi(q) + e_s)$. If q is raised in this case, the marginal benefit of additional effort is lower because the complementarity effect disappears and marginal utility is lower because the level of consumption is higher.

Timing is now bound to become a very important issue as we can think of a game being played between two generations in a family. This model shares features of a typical principal-agent model, where the principal, that is, the father, tries to induce a given level of effort by the agent, i.e. the son. When the father's choice on the investment q precedes the son's choice of e, the former's problem in the non-paternalistic case now becomes:[34]

$$\begin{aligned} V_p(x_p) &= \max_{\{c_p, c_k, q\}} u(c_p) + \delta \cdot [u(c_k) - v(e_k) \cdot \phi(x_p) + \beta \cdot u(\psi(\pi(q) \cdot e_k))], \\ \text{subject to } \psi(x_p) &= c_p + c_k + q \\ \text{and } e_k &= g_e(x_p, q). \end{aligned} \tag{25}$$

[34] This is the case where the father caring about the son only takes into account the son's utility from being a son, without accounting for his welfare as a father.

In this problem, the parent takes into account what the child's level of effort exerted will be, conditional on his own level of investment and schooling. Another inefficiency may arise in this case as the child could free-ride on the parent's investment and over-investment by the latter could follow. On the other hand, this may not be the case because a direct consequence of free riding can be a reduced level of consumption for the child, so this perverse incentive may disappear to some extent.

After some algebra, the first order conditions of the parent's problem, using the properties of the son's best response, yield:

$$\begin{aligned} u'(c_p) &= \delta \cdot u'(\psi(x_p) - c_p - q), \\ u'(\psi(x_p) - c_p - q) &= \beta \cdot u'(\pi(q) \cdot g_e(x_p, q)) \cdot \psi'(\pi(q) \cdot g_e(x_p, q)) \cdot g_e(x_p, q). \end{aligned} \tag{26}$$

The second order condition has a substantially simplified expression because, in practice, the parent does not really need to take into account when optimising how the child will respond to his own decisions. Indeed, it is possible to observe how the term in square brackets in the parent's optimisation problem coincides with the objective in the child's problem.

In a context of simultaneous decisions, it is natural to postulate an equilibrium in which son and father draw their best responses to each other's decisions. In equilibrium, none of them would like to deviate unilaterally from their current actions. Thus, for a given level of child's effort a parent will solve:

$$\begin{aligned} V_p(x_p) &= \max_{\{c_p, c_k, q\}} \left\{ u(c_p) + \delta \cdot \left[u(c_k) - v(e) \cdot \phi(x_p) + \beta \cdot u(\psi(\pi(q) \cdot e)) \right] \right\}, \\ \text{subject to } \psi(x_p) &= c_p + c_k + q, \end{aligned} \tag{27}$$

which leads to:

$$u'(c_p) = \delta \cdot u'(\psi(x_p) - c_p - q) \tag{28}$$

$$u'(\psi(x_p) - c_p - q) = \beta \cdot u'(\pi(q) \cdot e) \cdot \psi'(\pi(q) \cdot e) \cdot e. \tag{29}$$

This result, together with the previously stated child's best response function produce exactly the same result as the game in which the parent's decision precedes the child's. As anticipated above, this occurs simply because there is no real conflict between father and son apart from the relative weight given to the consumption by each one.

Let us see now what happens when the father cares about what the son should care about if he were sufficiently farsighted about the latter's own future as a father. In the simultaneous game, the father makes his own investment choice taking the son's choice of e as given:

$$\begin{aligned} V_p(x_p) &= \max_{\{c_p, c_k, q\}} \left\{ u(c_p) + \delta \left[u(\psi(x_p) - c_p - q) - v(e) \cdot \phi(x_p) + \beta \cdot V_p(x_k) \right] \right\}, \\ \text{subject to } \psi(x_p) &= c_p + c_k + q \\ \text{and } x_k &= e \cdot \pi(q). \end{aligned} \tag{30}$$

The first order conditions of this problem are:

$$
\begin{aligned}
u'(c_p) &= \delta \cdot u'(\psi(x_p) - c_p - q), && (31)\\
u'(\psi(x_p) - c_p - q) &= \beta \cdot V_p'(e \cdot \pi(q)) \cdot e \cdot \pi'(q), && (32)
\end{aligned}
$$

and the envelope theorem states:

$$
V_p'(x_p) = \delta \cdot \left[u'(\psi(x_p) - c_p - q) \cdot \psi'(x_p) - v(e) \cdot \phi'(x_p)\right]. \tag{33}
$$

Combining the last two expressions, this gives:

$$
u'(\psi(x_p) - c_p - q) = \beta\delta \cdot e\pi'(q) \cdot \left[u'(\psi(e \cdot \pi(q)) - c_p - q)\psi'(e \cdot \pi(q)) - v(e) \cdot \phi'(e \cdot \pi(q))\right] \tag{34}
$$

This provides a solution of the form $q = q(x_p, e)$. Investment in the child's education will be such that the current marginal disutility caused to the child by investing an extra monetary unit in his education as opposed to current consumption will be exactly offset by the gains from increased consumption for the child when adult through higher earnings plus the grandchild's induced reduction in learning effort, properly discounted. In this model, the parent only cares about the grandchild because he is altruistic about his own child, who will in turn be altruistic when adult.

Notice that in this setting the father does not take into account how investment can also encourage/disincentive current efforts by the child because that choice is made simultaneously. Alternatively, if parents anticipate how their children will react to their investment in q (sequential game), they will solve:

$$
\begin{aligned}
V_p(x_p) &= \max_{\{c_p,c_k,q\}} \left\{u(c_p) + \delta \cdot [u(\psi(x) - c_p - q) - v(e) \cdot \phi(x_p) + \beta \cdot V_p(x_k)]\right\},\\
\text{subject to } \psi(x_p) &= c_p + c_k + q,\\
e_k &= g_e(x_p, q)\\
\text{and } x_k &= e_k \cdot \pi(q), && (35)
\end{aligned}
$$

which yields a more complicated set of first order conditions:

$$
\begin{aligned}
u'(c_p) &= \delta \cdot u'(\psi(x_p) - c_p - q), && (36)\\
u'(\psi(x_p) - c_p - q) &= -\phi(x_p) \cdot v'(e) \cdot \frac{\partial e}{\partial q}\\
&+ \beta \cdot V_p'(e \cdot \pi(q)) \left(\frac{\partial e}{\partial q} \cdot \pi(q) + \pi'(q) \cdot e\right), && (37)\\
v'(e) \cdot \phi(x_p) &= \beta \cdot u'(\psi(\pi(q) \cdot e)) \cdot \psi'(\pi(q) \cdot e) \cdot \pi(q), && (38)
\end{aligned}
$$

and the envelope theorem states:

$$\begin{aligned} V_p'(x_p) &= \delta \cdot u'(\psi(x_p) - c_p - q) \cdot \psi'(x_p) - v(e) \cdot \phi'(x_p) - v'(e) \cdot \phi(x_p) \cdot \frac{\partial e}{\partial x_p} \\ &+ \beta \cdot V_p'(\pi(q) \cdot e) \cdot \pi(q) \cdot \frac{\partial e}{\partial x_p}. \end{aligned} \tag{39}$$

In steady-state, $x_p = x_k = x^* = e^* \cdot \pi(q^*)$. The parent's optimal choice of q will be characterised by:

$$\begin{aligned} u'(\psi(x^*) - c^* - q^*) &= -\phi(x^*) \cdot v'(e^*) \cdot \frac{\partial e}{\partial q} + \frac{\beta \cdot \pi'(q^*)}{1 - \beta \cdot \pi'(q^*)} \\ &\cdot \left[u'(\psi(x^*) - c^* - q^*) \cdot \psi'(x^*) - v(e^*) \cdot \phi'(x^*) - v'(e^*) \frac{\partial e}{\partial x_p} \right]. \end{aligned} \tag{40}$$

It is beyond the scope of this paper to carry out a more detailed analysis of this solution. The key feature of the last problem's solution is the fact that the terms including the partial derivatives of child's effort with respect to parent's schooling and investment do not cancel out. The solution of the simultaneous game (Nash equilibrium) generally differs from the sequential game (Subgame perfect Nash equilibrium) because, unlike the case of parents strictly concerned about the current (myopic) happiness of their children, there is a conflict of interest between them regarding the optimal choices of effort and investment. Conflict arises as a result fathers wanting to invest to improve the situation for their child's as a future parent, whereas children only invest to improve their own situation (current and future) as singletons. In the sequential game, parents predict that children will underinvest in the future, so they have to compensate for their children's 'insufficient' level of effort and their behavioural response to their investments. If complementarity prevails, higher investments induce higher effort by the sons. If, on the contrary, production of schooling is characterised by strong substitutability of parent's and child's inputs, free riding behaviour will arise.

B Intergenerational effects estimates: Discontinuity sample

Table 13: Inter-generational effects of father's schooling

		OLS			Ex.Id.IV		Ov.Id.IV			Overid.
Dependent variable		Coef	SE		Coef	SE	Coef	SE		P-Value
Arithmetics score at 7	Boys	0.143	0.075	*	0.009	0.595	0.136	0.183		0.029
Arithmetics score at 7	Girls	0.273	0.120	*	-0.828	0.887	0.128	0.307		0.010
Reading score at 7	Boys	0.886	0.196	*	-1.623	1.838	0.370	0.375		0.244
Reading score at 7	Girls	0.726	0.187	*	-1.000	1.968	1.515	0.695	*	0.063
Maths score at 11	Boys	2.211	0.353	*	1.296	2.256	2.156	0.735	*	0.013
Maths score at 11	Girls	2.052	0.361	*	-4.091	3.222	4.068	1.134	*	0.012
Reading score at 11	Boys	1.086	0.238	*	-0.322	1.406	0.400	0.435		0.460
Reading score at 11	Girls	1.201	0.245	*	-1.586	1.707	2.279	0.672	*	0.012
Gen. ability score at 11	Boys	2.836	0.581	*	0.202	3.461	1.503	1.362		0.309
Gen. ability score at 11	Girls	2.353	0.581	*	-8.298	5.390	5.221	1.808	*	0.009
Maths score at 16	Boys	1.649	0.231	*	1.423	1.900	1.348	0.436	*	0.193
Maths score at 16	Girls	1.232	0.327	*	-4.430	2.846	1.883	0.855	*	0.019
Reading score at 16	Boys	0.955	0.232	*	-0.878	2.081	0.683	0.466		0.436
Reading score at 16	Girls	1.281	0.231	*	-2.267	2.374	2.511	0.770	*	0.021
Father's interest at 7	Boys	0.054	0.018	*	0.074	0.099	0.095	0.044	*	0.230
Father's interest at 7	Girls	0.052	0.020	*	-0.150	0.145	0.099	0.052	*	0.184
Father's interest at 11	Boys	0.051	0.018	*	0.063	0.104	0.094	0.045	*	0.051
Father's interest at 11	Girls	0.050	0.020	*	-0.209	0.154	0.135	0.057	*	0.029
Mother's interest at 7	Boys	0.049	0.018	*	0.041	0.110	0.071	0.042		0.696
Mother's interest at 7	Girls	0.063	0.020	*	-0.181	0.166	0.146	0.056	*	0.220
Mother's interest at 11	Boys	0.071	0.017	*	0.115	0.111	0.045	0.040		0.206
Mother's interest at 11	Girls	0.050	0.021	*	-0.078	0.147	0.097	0.058		0.113
Mother's interest at 16	Boys	0.054	0.017	*	-0.074	0.110	0.066	0.040		0.008
Mother's interest at 16	Girls	0.056	0.020	*	-0.327	0.191	0.139	0.061	*	0.001
Any O-levels	Boys	0.881	0.124	*	0.612	0.796	0.603	0.148	*	0.012
Any O-levels	Girls	0.637	0.143	*	-2.502	2.045	0.778	0.434		0.004
Number O-levels	Boys	0.891	0.149	*	-0.157	0.959	0.514	0.168	*	0.030
Number O-levels	Girls	0.785	0.158	*	-0.024	1.436	1.122	0.433	*	0.779
Any A-levels	Boys	0.340	0.061	*	0.088	0.310	0.130	0.064	*	0.231
Any A-levels	Girls	0.273	0.071	*	-0.771	0.655	0.448	0.150	*	0.053
Number A-levels	Boys	0.390	0.063	*	0.421	0.599	0.121	0.086		0.489
Number A-levels	Girls	0.333	0.075	*	-0.458	0.538	0.444	0.139	*	0.194
Highest qualification indicator	Boys	0.376	0.061	*	0.039	0.624	0.209	0.108	*	0.166
Highest qualification indicator	Girls	0.298	0.073	*	-2.063	1.622	0.630	0.231	*	0.023
Log hourly wages at 33	Boys	0.053	0.025	*	0.029	0.221	-0.041	0.038		0.172
Log hourly wages at 33	Girls	0.077	0.022	*	-0.230	0.264	0.049	0.057		0.058
Log hourly wages at 42	Boys	0.072	0.025	*	0.016	0.259	0.031	0.039		0.637
Log hourly wages at 42	Girls	0.069	0.027	*	-0.224	0.193	0.037	0.066		0.110
Any Child at 23	Boys	-0.007	0.017		0.097	0.109	-0.019	0.017		0.690
Any Child at 23	Girls	-0.023	0.023		0.241	0.235	-0.034	0.058		0.287
Number children at 23	Boys	0.008	0.024		0.214	0.166	-0.029	0.023		0.567
Number children at 23	Girls	-0.058	0.031		0.661	0.490	0.023	0.096		0.103
Any Child at 33	Boys	-0.016	0.020		-0.274	0.227	-0.103	0.044	*	0.994
Any Child at 33	Girls	0.000	0.023		0.152	0.217	-0.091	0.070		0.510
Number children at 33	Boys	-0.015	0.054		-0.794	0.625	-0.246	0.101	*	0.899
Number children at 33	Girls	-0.013	0.061		0.858	0.732	0.025	0.182		0.320

NOTES: Children with father born in 1933 or 1934. Coefs. of dep.var on schooling.

Table 14: Inter-generational effects of mother's schooling

		OLS			IV1			IV2			Sargan
Dependent variable		Coef	SE		Coef	SE		Coef	SE		P-Value
Arithmetics score at 7	Boys	0.360	0.081	*	-0.474	0.450		0.576	0.184	*	0.036
Arithmetics score at 7	Girls	0.048	0.083		-0.775	0.634		0.601	0.236	*	0.069
Reading score at 7	Boys	0.887	0.179	*	0.970	1.113		1.391	0.482	*	0.497
Reading score at 7	Girls	0.738	0.179	*	-2.272	1.783		1.232	0.513	*	0.064
Maths score at 11	Boys	1.869	0.332	*	-0.564	1.936		3.981	0.986	*	0.016
Maths score at 11	Girls	1.627	0.332	*	-2.094	2.575		2.626	0.891	*	0.105
Reading score at 11	Boys	1.185	0.204	*	0.545	1.085		2.683	0.542	*	0.022
Reading score at 11	Girls	0.948	0.191	*	-1.183	1.445		1.316	0.508	*	0.017
Gen. ability score at 11	Boys	2.963	0.499	*	2.278	2.690		5.162	1.336	*	0.065
Gen. ability score at 11	Girls	1.830	0.540	*	-1.976	3.772		4.192	1.268	*	0.343
Maths score at 16	Boys	1.646	0.221	*	0.758	1.205		2.854	0.514	*	0.031
Maths score at 16	Girls	1.264	0.230	*	-1.062	1.635		2.028	0.761	*	0.076
Reading score at 16	Boys	1.325	0.211	*	0.764	1.213		2.356	0.508	*	0.199
Reading score at 16	Girls	0.977	0.196	*	0.298	1.402		2.844	0.719	*	0.010
Father's interest at 7	Boys	0.078	0.014	*	-0.123	0.078		0.054	0.035		0.033
Father's interest at 7	Girls	0.058	0.016	*	-0.190	0.121		0.034	0.044		0.145
Father's interest at 11	Boys	0.050	0.015	*	-0.095	0.073		0.084	0.040	*	0.064
Father's interest at 11	Girls	0.031	0.016	*	-0.148	0.117		0.156	0.046	*	0.014
Father's interest at 16	Boys	0.042	0.016	*	0.013	0.072		0.154	0.040	*	0.175
Father's interest at 16	Girls	0.056	0.017	*	-0.106	0.114		0.074	0.047		0.170
Mother's interest at 7	Boys	0.081	0.014	*	-0.101	0.081		0.069	0.038		0.080
Mother's interest at 7	Girls	0.060	0.017	*	-0.230	0.134		-0.003	0.048		0.194
Mother's interest at 11	Boys	0.053	0.015	*	0.011	0.073		0.088	0.042	*	0.559
Mother's interest at 11	Girls	0.030	0.017		-0.108	0.117		0.093	0.047	*	0.278
Mother's interest at 16	Boys	0.041	0.015	*	-0.049	0.073		0.120	0.040	*	0.037
Mother's interest at 16	Girls	0.061	0.017	*	-0.088	0.107		0.080	0.046		0.267
Any O-levels	Boys	0.804	0.150	*	0.204	0.588		0.787	0.292	*	0.252
Any O-levels	Girls	0.742	0.116	*	-1.285	1.059		1.042	0.303	*	0.006
Number O-levels	Boys	0.729	0.159	*	0.762	0.572		0.917	0.292	*	0.263
Number O-levels	Girls	0.747	0.123	*	-0.453	0.821		0.662	0.312	*	0.294
Any A-levels	Boys	0.371	0.075	*	0.310	0.240		0.361	0.168	*	0.302
Any A-levels	Girls	0.321	0.047	*	0.018	0.286		0.247	0.122	*	0.653
Number A-levels	Boys	0.345	0.072	*	0.013	0.286		0.160	0.134		0.068
Number A-levels	Girls	0.281	0.071	*	0.004	0.264		0.153	0.117		0.703
Highest qualification indicator	Boys	0.366	0.064	*	0.151	0.305		0.471	0.148	*	0.146
Highest qualification indicator	Girls	0.396	0.054	*	-0.950	0.604		0.397	0.165	*	0.003
Log hourly wages at 33	Boys	0.063	0.030	*	0.039	0.097		0.087	0.044	*	0.690
Log hourly wages at 33	Girls	0.056	0.021	*	-0.800	0.759		0.032	0.058		0.058
Log hourly wages at 42	Boys	0.044	0.022	*	0.058	0.084		0.117	0.045	*	0.655
Log hourly wages at 42	Girls	0.072	0.019	*	-0.169	0.186		0.100	0.065		0.052
Any Child at 23	Boys	-0.024	0.011	*	-0.038	0.062		-0.071	0.030	*	0.352
Any Child at 23	Girls	-0.052	0.017	*	0.007	0.121		-0.083	0.046		0.668
Number children at 23	Boys	-0.032	0.016	*	-0.122	0.088		-0.122	0.041	*	0.487
Number children at 23	Girls	-0.078	0.027	*	0.128	0.212		-0.099	0.072		0.271
Any Child at 33	Boys	0.002	0.020		0.223	0.098	*	0.027	0.048		0.116
Any Child at 33	Girls	-0.005	0.018		-0.324	0.153	*	-0.083	0.053		0.052
Number children at 33	Boys	-0.015	0.050		0.380	0.244		0.038	0.138		0.321
Number children at 33	Girls	-0.031	0.047		-0.908	0.406	*	-0.232	0.139		0.061

NOTES: As in table 13.