Lancaster University Management School Working Paper 2011/017

Estimates of the Causal Effects of Education on Earnings over the Lifecycle with Cohort Effects and Endogenous Education

Giuseppe Migali and Ian Walker

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Estimates of the Causal Effects of Education on Earnings over the Lifecycle with Cohort Effects and Endogenous Education*

Giuseppe Migali† and Ian Walker‡

8 July, 2011

Abstract

This paper acknowledges that the relationship between log wages and schooling is considerably more complex than the simple human capital earnings function suggests and that schooling is endogenous. We estimate a model where educational attainment is discrete and ordered and log wages are determined by a simple function of work experience for each level of attainment. We distinguish between lifecycle and cohort effects by exploiting the fact that we have a short panel. We strongly reject both the usual separability assumption and exogeneity of educational attainment.

Keywords: Returns to education, Selection, Lifecycle, Cohort
JEL Classification: I21, J31, C32

1 Introduction

Estimating the returns to schooling is a major industry for applied economists. The methodological difficulty in estimating the causal effect of education is well known: bias (due to ability, school quality, non-cognitive skill) arising

*The authors would like to thank the participants to the Society of Labor Economics world meeting 2010, Royal Economic Society annual conference 2010, International Workshop on Applied Economics of Education 2010, European Association of Labor Economics Annual conference 2009 and to the Lancaster University Economics seminar for useful comments.

†g.migali@lancaster.ac.uk, Department of Economics, Lancaster University Management School, UK and Dopes Universita' Magna Grecia, Catanzaro, Italy.

‡Department of Economics, Lancaster University Management School.

from the correlation between wages and the unobservable determinants of schooling contaminates least squares estimates which can then only be interpreted as an upper bound. There is a large literature that attempts to address this problem using instrumental variable methods by exploiting potential exclusion restrictions. In this context the simplicity of the workhorse empirical specification of the human capital earnings function is extremely convenient. This specification has log wages being determined by an additively separable and linear function of schooling and a quadratic function of some measure of experience (usually age). A convenient property of this simple specification is that there is one rate of return and it can, under certain assumptions, be equated to the coefficient on schooling. Moreover, there is econometric convenience associated with such a simple specification: there is just one variable, schooling, that is endogenous so the search for exclusion restrictions, that has so taxed the ingenuity of researchers in this area, need not continue beyond just one. Within the confines of studies that attempt to deal with endogenous schooling, there have been some attempts to depart from the simple specification. For example, Willis and Rosen (1979) in their structural model consider schooling to be a college education endogenous dummy variable and allow for selection on the interaction between experience and schooling. Kenny et al. (1979) allow for the minimum schooling level and use a Tobit specification; and Harmon and Walker (1995) use an extension of the Heckman two-step approach where the latent variable for years of schooling is treated as an ordered probit. To the best of our knowledge, all such studies that estimate the effect of endogenous education do so within a model where schooling is some univariate function of years of schooling and assumes that the effect of schooling does not vary across experience - that is, separability between schooling and experience is a maintained hypothesis. This limitation extends to twins studies (for example Ashenfelter and Rouse, 1998) where identification is invariably assisted by estimating an assumed linear relationship between within twin pair earnings difference and their schooling difference. Here linearity is crucial because there is rather little variance in the within difference in education levels.

Adopting this maintained hypothesis seems increasingly perverse since there is considerable evidence (e.g., Heckman et al.,1996, Jaeger and Page, 1996, Hungerford and Solon, 1987) against it. In particular, many studies suggest that the effect of schooling is not linear, that schooling itself is not univariate but rather is, at the very least, best thought of as a succession of levels of achievement that is not simply college vs no college. Moreover, many studies show that age earnings profiles are certainly not parallel across education levels (e.g. Neal, 2004 and Heckman et al., 2006). While, one might argue that such evidence, coming as it does from a framework where schooling is treated as exogenous, is subject to some bias - but, nonetheless, it seems cavalier to ignore its findings altogether. This paper is an attempt to incorporate the suggestion that the relationship between log wages and schooling is considerably more complex than the simple human capital earnings function suggests and yet schooling is endogenous.

Our work complements that of Heckman et al. (2008)[1]: they compute the internal rate of return (IRR) to the investment in education for different levels of schooling. They start from a general non parametric approach to the estimation of the determinants of log earnings but do not explicitly allow for endogenous schooling. In contrast to that work we adopt a parametric model but allow for the selection associated with endogenous schooling. Leaving aside the issue of endogenous schooling, parametric models do have some advantages over non parametric: they converge faster, they do not require the estimation of smoothing parameters, they are easy to interpret, and parametric estimates can be used to extrapolate out of sample. On the other hand, estimates of parametric models are conditional on the maintained functional form assumptions. Here we implement what we think of a useful compromise between generality and tractability. In our selection model education is captured by achievement measures that are ordered: from the lower secondary level of education associated with the minimum school leaving age (that the US literature thinks of as High School drop-outs), through High School graduation (around the age of 18), through an undergraduate college degree (around the age of 21), and up to postgraduate qualifications (less common in the UK than the US). We estimate the probability that individuals have a particular level of education by exploiting the fact that they are mutually exclusive and ordered. We then estimate age earnings profiles for each education achievement group separately, controlling for selection into each level of educational achievement. Therefore we do not impose separability, nor do we impose that the schooling has a linear effect on log wages. Even this simple departure from the usual separable linear framework comes, of course, at some cost. The assumption that education is a latent variable that can be captured by an ordered probit is not entirely innocuous: while the assumption that levels of education of achievement are ordered seems like a natural one, the assumption that the distribution of the residuals in the equation that determines academic progression through the education system are distributed normally is essentially arbitrary. It is only as justifiable as any other parametric assumption and a long way from the fully non-parametric (or, even, semi-parametric) approach that might be possible

[1] A broader extension of this work is included in a authors' chapter in the Handbook of Economics of Education (2006), where they focus on many of the issues treated in our work.

to estimate.[2] Of course, adopting normality makes an important contribution towards identifying the parameters of interest in our selection model and sacrificing such a contribution would place a correspondingly greater burden on the validity of the exclusion restrictions. Given that fully non-parametric selection models are still in their infancy (see Lanot and Walker, 1998, in the context of the effect of unions on wages and Das et al, 2003, in the context of returns to education[3]), and that studies of the returns to education inevitably have to rely on observational data where identification comes from naturally occurring experiments, we view our own approach as a practical compromise that could be implemented with many datasets. Here, we use the UK Labour Force Survey a very large and flexible dataset that is the UK equivalent of the US CPS data, and has been used extensively elsewhere to study the returns to education (see, for example, Walker and Zhu, 2008).

A second issue is how to interpret the estimates in a model where endogeneity is accounted for. If the effect of education on log wages is subject to unobservable heterogeneity, as seems reasonable, then it is well known that IV estimates (and those from other methods that rely on specific exclusion restrictions) have to be interpreted with care. In particular, it has been shown that the coefficient on schooling can no longer be interpreted as an average treatment effect and such methods identify the effect of education on those individuals whose schooling is affected by the exclusion restrictions (see Moffitt (1999) for a clear exposition of the issues). In most cases, exclusion restrictions rely on education reforms that affect one cohort of individuals in an area but not another cohort in the same region, or the same cohort in a different region. Thus, interpretation is not a trivial problem: in general we have no way of knowing how a particular exclusion restriction affects the education decisions on individuals in our data sets. In some cases we might have strong suspicions that particular types of individuals are affected (for example, Kling (2001) shows that Card's proximity IV affected low income youth more) and, if we are lucky, the estimates might then be informative about particular policy issues. For example, changes in minimum school leaving ages have been used quite extensively (see, e.g. Harmon and Walker, 1995) and in this case one might imagine that the reform affects those who intended to leave education at an age below the new minimum. Thus, such estimates may be relevant for policy that is directed towards low ability and/or credit constrained individuals (although, Cameron and Taber (2004) using several methodologies do not find any evidence that borrowing con-

[2]See, for example, the STATA sneop command of Stewart (2004), and Blundell and Powell (2004) for a survey of semi-parametric selection models.

[3]Das et al (2003) develop a non parametric estimator for sample selection models. This tests for departure from Normality by including higher order terms in the Mills ratio.

straints affect schooling decisions). Only exceptionally will we be able to identify the average treatment effect using such methods.[4] In practice, the only exception that is likely to arise is when the exclusion restriction is a random variable: for example (through a lottery) and even here one needs to be sure that such a variable has no indirect effects on wages. In general, in such models, there will be a distribution of treatment effects and what is actually identified will depend on the exclusion restriction. However, in the case where the treatment is discrete (as in our ordered probit case) the distributional assumptions allow the researcher, in principle, to identify the whole distribution of treatment effects. Thus, while the parametric restriction is strong it buys the researcher a lot of information.

Given that our aim is to estimate the internal rate of return associated with educational attainments we focus on the lifecycle pattern of earnings at different educational attainment levels. Thus, given our aim, it is important that we estimate the true lifecycle effect net of any cohort and/or calendar time effects. It is, of course, impossible to make such distinctions when using cross-section data. The data we use here is a sequence of pooled cross-sections over a period of 13 years which offers better prospects but even here we should expect significant collinearity problems. Here, we exploit the fact that our data has a short panel element to it: individuals are interviewed over 5 quarters and earnings data is collected in the first and final waves - an interval of approximately one year.

Estimation exploits the availability of both data sets to the maximum. In particular, we estimate the lifecycle earnings profiles for each education group using the 1997-2009 pooled longitudinal data, and then, controlling for the lifecycle, estimate the cohort and year effects, together with the impact of education, on the level of wages using the pooled cross sections that are also available from 1997 to 2009. We allow for non-random selection into each education level.

The distinctive features of this work are: we separately identify lifecycle and cohort effects, we control for selectivity and our model of earnings and its growth does not impose the restriction that the age profiles (and the effects of other observables) are the same at each level of education. We correct for unobserved ability using, as exclusion restrictions: the raising of the school leaving age reform (known as RoSLA) that was imposed on the English (and Welsh) education system from 1973; and the month of birth. The use of the former is well known. Crawford et al. (2007), analyzing the English education system, show that children born later in the school year perform

[4]See Carneiro et al (2010) discussion and implementation of the marginal treatment effect estimator.

significantly worse in exams than those born earlier in the school year. There are no strong reasons for thinking that both RoSLA and month of birth are correlated with the unobservable determinants of earnings. However, Buckles and Hungerman (2008) find that season of birth is correlated to family background in the US. Therefore, it's an empirical question, whether it is true in the UK, that the LFS cannot eliminate.

This paper is structured as follows: Section 2 describes the data, in Section 3 we explain the estimation method of the returns to education, Section 4 presents the results and Section 5 concludes.

2 Data description

The LFS is a quarterly sample survey of households living at private addresses. Its purpose is to provide information on the UK labour market that can then be used to evaluate labour market and educational policies. The survey seeks information on respondents' personal circumstances and their labour market status during a specific reference period, normally a period of one week or four weeks (depending on the topic) immediately prior to the interview.

The survey has been conducted on a quarterly basis, with each sample household retained for five consecutive quarters, and a fifth of the sample replaced each quarter. This is known as Quarterly LFS (QLFS) and it was designed to produce cross-sectional data, such that in any one quarter, one wave will be receiving their first interview, one wave their second, and so on, with one wave receiving their fifth and final interview. Thus there is an 80% overlap in the samples for each successive quarter. The UK LFS has existed since the mid 1970's but it is only since 1993 that data on gross earnings has been collected, and only since 1997 has earnings been recorded in both waves 1 and 5.

In recent years it has been recognised that linking together data on each individual across quarters would produce a rich source of longitudinal data, therefore five-quarter longitudinal datasets have also been produced for the same period, for example linking spring 1998 with spring 1999 and containing data from all five waves of the survey. This is known as Longitudinal LFS (LLFS), and because of the resources involved in production and the size of the resultant datasets, the longitudinal datasets include only a subset of the full LFS variable set. Since the focus of analyses of these datasets was expected to be the population of working age, the datasets has also been restricted to women aged 15 to 59 at the first quarter and men aged 15 to 64 at the first quarter. The reduced cross-sectional datasets have been matched, and all unmatched cases are dropped, as are all cases where there are no data on economic activity, in any of the quarters. In our analysis we consider the period 1997-2009 inclusive and we use the LLFS to exploit the panel feature of the data.

Our procedure is the following. We first append all five-quarter LLFS datasets from 1997 to 2009 and we obtain a total sample of 415,893 observations.[5] The proportion of employees is around 64% (265,063), the self-employed are around 7.5% (31,033), there is a small percentage of people (less than 1%) in government training programs, the remaining people are inactive in the labour market. We restrict the sample to be employees. We drop individuals observed only in either the first or the fifth quarter, and our sample size falls to 223,339. We drop missing wages and our remaining sample is 211,038; however there is also a large number (around 60,000) of individuals for whom the earnings variable is "no answer/does not apply", this restricts our sample to 150,353.[6]

We stack all QLFS datasets, from 1997 to 2009, which include around 125,000 individuals per quarter and in each quarter five waves of on average 25,036 individuals, and we obtain a total sample of 6,259,177 people wave observations. If we restrict on the same age range used in the LLFS we get around 18,000 individuals per wave in each quarter, and the proportion of employees in each wave is around 61% (11,000) and self-employed are 7.8%. Earnings are collected only in the first and fifth wave and the proportion of employes reporting a positive wage is on average 73%, on average 7500 individuals in the first wave and a little less in the fifth wave. We therefore keep only employees reporting a positive wage in the first wave and the new sample is 369,652.[7]

Now we apply similar restrictions to both datasets. We only consider people aged from 25 to 60 years old and born between 1940 and 1984 (we loose 15% of the observations in QLFS and around 11% in LLFS), and the new sample sizes are 314,169 and 134,412 in QLFS and LLFS, respectively. The following groups were dropped from the analysis: residents of Northern Ireland (3% in QLFS, 2% in LLFS) and Scotland (9%, in both QLFS and LLFS), people born outside UK (3% in LLFS, 8% in QLFS) people still in full-time education or never had education, people that completed their

[5] LFS is a panel of addresses not people. Movers are not followed so attrition between waves 1 and 5 accounts for the lower number of cases available for linking and higher attrition, the original five-quarter datasets always contain fewer observations than the QLFS datasets.

[6] This implies that the proportion of employes reporting positive earnings in both first and fifth wave is 67%.

[7] This sample size corresponds to 1 wave per quarter for 50 quarters from 1997 to 2009, but the last two quarters of 2009 were not available.

education younger than 14 and older than 30 (0.3% in QLFS, there is no equivalent variable in LLFS). The final sample size is 250,538 in QLFS and 111,522 in LLFS. The main variables of interest for our analysis are earnings, education and individual characteristics. We constructed our variables in the following ways.

Average gross hourly pay[8] is provided in the LFS raw data. We further restrict the total number of hours worked in the reference pay period to lie in the range [0...94] (loose less than 0.05%). The resulting hourly pay rate is transformed into a real wage rate by dividing by the Retail Price Index (All items) with September 2009 as the base period. The top and bottom 1% of the wage distribution by category of highest academic qualifications were trimmed to avoid outliers arising from measurement error in the wage rate influencing the results unduly. Finally, it is generally more useful to consider the proportional differences in earnings across different groups of individuals, rather than the absolute difference. That is, as usual, we consider the log of the wage rate rather than the wage rate itself.

Our analysis concentrates on education qualifications, rather than the age at which individuals leave education. In England, compulsory education is from the age of 5 to 16 with 5 to 10 being spent in primary school, and 11 to 16 spent in lower secondary education. Students undertake national examinations, typically in five to ten subjects, known as the General Certificate of Secondary education[9] (GCSEs) at age 16. After the age of 16 they can enter the labour market or continue into post compulsory upper secondary education. Students can choose between academic and vocational qualifications. The academic track consists of GCSEs at 16, followed by A-levels at 18 and university undergraduate degree usually from the age of 19 to 22, possibly followed by a postgraduate degree. There is a very clear ordered progression along this academic education track. The vocational track is less easy to characterize: typically students would leave formal education at the age of 16 and engage in some occupational training perhaps on a part-time basis while at work, or attend some further education college on a full-time basis for approximately two years gaining vocational qualifications before entering

[8]It is a derived variable defined as the ratio of usual earnings to usual hours (from main job) including paid overtime. Usual earnings are obtained using information asked directly to all employees and those on schemes, e.g. gross pay before deductions (self-assessed), expected gross earnings (self-assessed). The proportion of non-response to the earnings question is similar by education level and across each QLFS data set. Therefore there is no concern for non-random non-response.

[9]We refer here to the education system after 1973. Prior to 1973 it was common for tracking to start earlier and there was a distinction between the examinations that vocational track pupils took. We convert these older qualifications into their modern equivalents using conventional criteria.

work. Many vocational qualifications are specialised and taken by rather small group of individuals. Fortunately there is a well-developed method of grouping equivalent qualifications into levels known as National Vocational Qualifications (NVQ) equivalents. These are defined in Table 1 and divided into five NVQ levels: from NVQ1 (below GCSE qualifications) to NVQ5 (postgraduate level qualifications)[10]. Table 1 gives examples of the vocational qualifications and their associated NVQ levels, as well as the most common non-vocational ones. We follow established practice in how NVQs are defined with the exception that we pool together the NVQ4 and NVQ5 qualifications due to the small number of observations at NVQ5 level[11]. Our omitted category is no qualifications. We generated the educational variables separately and we observed the same proportion in both QLFS and LLFS. Here we report the descriptive statistics that refer to the QLFS sample which is larger and includes all of the individuals in the LLFS. The total sample size of 250,538 comprises 120,387 males and 130,151 females.

Summary statistics and the distribution of the earnings given the NVQ levels are provided in Table 2. Only those individuals earning a positive wage are included in the sample. We observe that the largest group of individuals (34% of males and 32% of females) have an NVQ4, whereas those with NVQ1 are very small. We also notice that males with NVQ3 are almost the double of females. This probably due to the fact that we include both vocational and academic qualifications and the percentage of females taking vocational courses is very small compared to males.[12] Comparing the wages at NVQ3 and NVQ2 males with the higher qualification earn 19% more than those with NVQ2, this percentage drops to 10% for females. The wage differential between NVQ4 (and 5) and NVQ3 is around 37% for males and 40% for females, and it broadly corresponds to the "college premium" in the US literature.

3 Estimation method

Belzil (2007) surveys the empirical literature concerned with estimating the returns to schooling and show that since 1970 more than 200 published articles and working papers have been devoted to this topic. It is well known that OLS methods produce unbiased estimates only if realized schooling and unobservable attributes that affect earnings are uncorrelated. The presence

[10]See Makepeace et al. (2003). In general NVQ3 corresponds to high school graduates and below NVQ3 high school drop-out.

[11]We group NVQ4 and NVQ5 and we refer as NVQ4 hereafter.

[12]For a more detailed picture of the problem see Walker and Zhu (2007) who show the NVQ distributions disaggregated by academic and vocational paths.

of ability bias implies that estimates of the effect of education on earnings reflects both the unobservable differences across individuals as well as the effect of education per se, so that OLS are biased upwards. It is commonly felt that more highly educated individuals might earn more because they have unobservable attributes, like ability, that are valued in the labour market and which are correlated with education. As stressed by Walker and Zhu (2008), we need to distinguish the average effect due to an additional year of schooling, from the differences in earnings that occur, on average, across individuals that have different levels of education.

The common solution to the empirical problem of estimating the true causal effect of education is to exploit that part of the variation in education levels across individuals that is not due to self-selection. In the literature many studies use the IV approach, and Card (2001) surveys a number of studies which use a variety of instruments: parental background (Willis and Rosen, 1979); quarter of birth (Angrist and Krueger, 1991); college proximity (Card, 1995); raising of the school leaving age (Harmon and Walker, 1995); and WWII (Ichino and Winter-Ebmer, 1999).

IV methods seem to find, in general, returns to schooling between 20% and 40% above the corresponding OLS estimate, which are typically slightly larger than the estimates from selection models. One explanation for IV estimates systematically exceeding the estimates of least sqaures, is the combination of small ability bias (in IV), with downward bias in OLS estimates due to measurement error in reported schooling (see Griliches, 1977; Angrist and Krueger, 1991; Walker and Zhu, 2008). However it seems unlikely that measurement errors in qualifications will be very large and we suspect that this source of bias will be small.

An alternative explanation is that IVs provide, in terms of the Angrist, Imbens and Rubin (1996) causal model, a Local Average Treatment Effect (LATE) estimator, which estimates the effect for those whose treatment status is affected by the instrument. LATE is the average effect of the treatment for those who change treatment status because of a change in the value of the instrument. An intervention does not affect all individuals in the same way. Typically there is heterogeneity in the response to the treatment across individuals. Consequently, there are different potential questions that evaluation methods attempt to answer, the most common is the average effect on individuals of a certain type. Heckman and Vytlacil (1998) and Blundell and Dias (2009) distinguish between: the population average treatment effect (ATE), which would be the impact if individuals were assigned at random to treatment; the average effect on individuals that were assigned to treatment (TT); the effect of treatment on agents that are indifferent to participation, which is the marginal version of the local average treatment effect (LATE) discussed above; or the effect of treatment on the untreated (TU) which is typically an interesting measure for decisions about extending some treatment to a group that was formerly excluded from treatment. Except for the case of homogeneous treatment effect[13] all these measures are conceptually different and each has a different set of conditions for identification.

Moffit (1999) gives a simple graphical interpretation of these estimators. Figure 1 shows a hypothetical density of α_i (with mean $\bar{\alpha}$) in the population, where α_i is the heterogenous treatment response for individual i, and it is also known as random coefficient. Moffit assumes that selectivity takes place strictly on α_i and that selection is positive. This means that individuals with higher α_i are more likely to participate in the treatment - the monotonicity assumption. More precisely those with α_i above the cutoff point α^* participate and the other individuals do not. In Figure 1, the parameter $\bar{\alpha}$ is the ATE. Whereas α^{TT} is the TT, the average gain of those who participate $E(\alpha|\alpha_i > \alpha^*)$, and α^{TU} is the mean gain of non participants in case of treatment $E(\alpha|\alpha_i < \alpha^*)$. To represent the LATE we need to consider a discrete changes in α^* to $\alpha^{*'}$, which changes α^{TT} to $\alpha^{TT'}$. The LATE is given by the the difference $\alpha^{TT} - \alpha^{TT'}$ divided by the area under the curve between α^* and $\alpha^{*'}$, which is the change in the probability of participating.

While Blundell et al. (2005) estimate a multiple treatment model in a separable framework, here we estimate the effect of multiple treatment without separability by estimating one equation per qualification. Since the difference in estimated constant terms gives a rate of return, and the constant term has an additive error, our model effectively allow for heterogenous treatment effects.

Rather than use IV estimation and face the problem of interpretation, we adopt the more restrictive Heckman selection model, extended to allow for an ordered choice across several education levels. Of course, a selection model is more restrictive than IV since it assumes that the distribution of unobservables in both wages and schooling are jointly normal distributed. Our educational variables can be ordered - so that NVQ4 is higher than 3, and higher than 2 and so on. Therefore, we can estimate the probability that individuals have an NVQ at any particular level and exploit the fact that they are mutually exclusive - so that the probabilities of having each level sum to one across levels. The main difference from IV methods, is that by making an assumption about the distribution of the unobservable determinants of earnings we estimate the effect of qualifications on hourly wages across the whole distribution of unobservables, in particular at the mean. Thus, the selection model method, unlike IV, yields estimates that are comparable to

[13]Under this assumption all these measures are identical.

the much simpler least squares regression method[14].

In our analysis we use a first difference model of hourly wage to estimate the lifecycle effects for separately each level of education, exploiting the fact that our data have a short panel element. Estimating in first difference using only the panel element of the 1997-2009 data allows us to identify the lifecycle pattern of wages independently of any cohort trends providing those cohort effects are fixed effects in the data. That is, provided that cohort effects in the cross-section equation are additively separable from lifecycle (age) effects. We estimate this difference equation separately for each education level to avoid imposing separability. Implicit in estimating in differences using OLS is the presumption that unobserved ability affects wages only through the level and not through an effect on growth. This is a common assumption in the literature, but not an innocuous one because ability that is unobserved early in life may become revealed to employers later in life (e.g. Altonji and Pierret, 2001). Our attempts to correct for selection effect in the wage growth equation, in the same way as we approach selection in levels, did not suggest that this was a statistically significant issue.

Moreover, our model does not impose the restriction that the selectivity works in the same way at all NVQ levels. We do not impose the restriction that the age profiles (and the region and race effects) are the same at each level of NVQ. So we allow for non separability in the earnings function between the schooling effect and the age effect.

3.1 Model

Our baseline model is a conventional human capital earnings function except that it is separable in education level. Thus equation 1 is written for each of 5 levels of education, and is modeled as the sum of quadratic age effects, cubic cohort effects (c captures additively separable cohort effects through a cubic year of birth), time effects and individual error term[15]. Equation 2 simply differences equation 1. Equation 3 imposes the estimates of the lifecycle parameters from 2.

$$W_{isq} = c_s + \delta_s age_{isq} + \rho_s age_{isq}^2 + \mu_s t + u_{isq} \tag{1}$$

$$\Delta W_{isq} = (\delta_s + \mu_s) + 2\rho_s age_{isq} + \nu_{isq} \tag{2}$$

[14]We estimate the first difference and wage level equations in two steps. This is equivalent to a single step SUR but less efficient.

[15]The standard Mincerian model assumes that log earnings are quadratic in experience. We do not have a good measure of experience, people use age minus schooling which puts an interaction term.

$$W_{is} = (\widehat{\delta_s + \mu_s})t + \widehat{\rho}_s age_{is}^2 + \ c_s + \beta_s S_{is} + \gamma_s X_{is} + \epsilon_{is} \tag{3}$$

$$S_{is}^* = \omega_s + \psi_s Z_{is} + v_{is} \tag{4}$$

where W is the log of wages, $i = 1 \ldots N$ individuals, $s = 0 \ldots 5$ educational levels, $q = 1, 5$ LFS quarter, $c_s = c_{0s} + c_{1s} yob + c_{2s} yob^2 + c_{3s} yob^3$, $t =$ LFS survey year and $S_{is} = 1$ if $S_{is}^* = s$. We also assume $E[u_{is}] = E[v_{is}] = E[\epsilon_{is}] = 0$, but $cov(v_{is}, \epsilon_{is}) \neq 0$, then $E[S_{is}\epsilon_{is}] \neq 0$, v_{is} and ϵ_{is} follow a bivariate Normal distribution.

The estimation of 1 is difficult because individuals' age added to their birth year gives the current year, so that there is an exact linear relationship between the age, cohort, and time effects.

However, the interval period between the first and the fifth LFS wave is around 1-year, we can take the first difference of 1 and remove all the time invariant effects, thus controlling for time-invariant unobserved individual heterogeneity. In equation 2 wage growth is linear in age (our measure of experience) for each education classification, therefore we are not imposing separability between age and schooling. Lifecycle effects on earnings are given by the constant term, which is a cumulative effect of age and time, and ρ which corresponds to the effect of age squared in 1. We can obtain consistent estimates of the parameters ρ and $\delta + \mu$ from 2 providing selection bias is driven only by fixed effects and cohorts effects are additively separable in 1. We impose these consistent estimates on 1 and then estimate the remaining parameters using the selection model represented in equations 3 and 4. We can recover all the coefficients except we cannot separately identify μ from δ.

The selection model estimates an ordered probit as the first step, equation 4, where the vector **Z** includes at least some variables that are not contained in **X**. As discussed by Heckman (1990) and Card (1994) identification in selection models (as for IV) has to be able to include variables that affect education and do not also affect earnings directly, the so-called exclusion restrictions.

We use the estimated probit coefficients from equation 4, to generate the relevant Inverse Mills Ratios (IMR), which capture the likelihood that an individual has a particular level of qualification. We include these IMRs in the wage equation 3 to correct for the fact that individuals with a particular level of education will have a particular unobserved component to earnings. We include an ethnic variable, also used in the ordered probit, and we assume a cohort effect for each education classification. Thus, our final specification allows the intercept and the coefficients on the controls to vary by schooling qualification.

We estimate log earnings for each NVQ level using regressions, which allows heterogeneity in the effect of qualifications on earnings[16], of the form

$$W_{is} = \theta_{is} + \widehat{\rho}_s age_{is}^2 + (\widehat{\delta_s + \mu_s})t + \lambda_s \widehat{IMR}_{is} + \beta_s X_{is} + \epsilon_{is} \tag{5}$$

where $\theta_{is} = c_{0s} + (c_{1s} - \delta_s)yob_i + c_{2s}yob_i^2 + c_{3s}yob_i^3$ for $i = 0 \dots N$ and $s = NVQ0 \dots NVQ4$.

As mentioned above, the identification strategy requires an exclusion restriction in the vector $\mathbf{X}$. In the absence of such, the selection variables will coincide with the independent variables, i.e. $\mathbf{Z} = \mathbf{X}$, and, while we could still use the two-step procedure and obtain the estimates of the IMR, their identification would come only from the distributional assumptions. It is well known that such estimates would be sensitive and would rely exclusively on the assumption that the IMR is a non-linear transformation of the same regressors $\mathbf{X}$ as in the outcome equations (Heckman, 1979).

In a traditional two-step selection model with only two outcomes in the participation equation, a standard $t-$test on the coefficient, λ, of the IMR is a valid test of the null hypothesis of no selection bias. We would expect a positive λ, which means that more highly educated individuals might earn more because they have unobservable attributes, like ability and perseverance, that are valued in the labour market and which are positive correlated with education. In our model, with multiple treatments, the IMR represents the correlation between a particular level of education compared to all the others, therefore a significant coefficient can be interpreted as evidence of selectivity but the sign of this coefficient does not have as clear an interpretation as in the binary model.

3.2 Exclusion restrictions

Our identification strategy is based on the exclusion in the wage equation 3 of two variables that we think can reasonably be considered to be exogenous and affect wages only through education. The first variable is the raising of the school leaving age reform (RoSLA): those born before August 1958 faced a minimum school leaving age of 15 years, those born after that date were required to stay in school until at least 16 years of age. As shown by Harmon and Walker (1995) RoSLA can be used both as IV and as an exclusion restriction in a selection model. Oreopoulos and Page (2006) estimates the

[16] We use the Stata routine "oheckman" written by Chiburis and Lokshin (2007).

LATE for secondary schooling exploiting the high drop out rates in the UK and using RoSLA as an IV. They find large gains from compulsory schooling. These estimates are not very different from those of US and Canada, although the proportion of people affected by the change in compulsory schooling in the UK was much higher than that in North America.

In Figure 3 we show for both males and females, born before and after the RoSLa, the percentage difference in the probability of obtaining a given NVQ qualification conditional on having a lower NVQ. As expected the effect of RoSLA is bigger for low levels of education, the probability of having an NVQ1 and NVQ2 is very high conditional on having no qualification or an NVQ1, respectively. This was indeed the purpose of the reform, to increase participation for those that would otherwise have not remained at school. If we consider just those born 5 years either side of the reform, to reduce the extent to which there are cohort trends, we find that the reform immediately reduced the probability of leaving school at the old minimum, 15, from approximately 30% to zero and that the probability of leaving school, at the new minimum rose immediately from approximately 30% to approximately 60%. The distribution above leaving at 16 remained unchanged (see Figure 2 and Chevalier et al., 2004).

The second exclusion restriction is month of birth. In the literature it is well documented that there is an impact of date of birth on cognitive test scores, with the youngest children in each academic cohort year performing poorer, on average, than the older members of their cohort. Puhani and Weber (2005) use a sample of German children and investigate the impact of age at school entry on test scores at the end of primary school (age 10). They find that children who start school aged 7 rather than aged 6 have test scores that are 0.42 standard deviations higher at the end of primary school. Bedard and Dhuey (2006) use internationally comparable data for OECD countries to estimate the impact of relative age on test scores at ages 9 and 13. They find that children being one month older get higher test scores at the age of 9 than at age 13. Ashworth and Heyndels (2007) consider the effects of month of birth in soccer education programs. They find systematic differences in players' performance depending on the months in which they are born. These differences could conceivably produce productivity and wage differences in adulthood. Crawford et al. (2007) is a recent example that notes the relationship between month of birth and educational attainment in the UK.[17] They show that children born later in the school year

[17] The English rule for admission says that children have to start school at the beginning of the term following their 5th birthday. There are three terms: start September, start January, start April. However what happens in all Local Education Authorities is that children start at the beginning of the academic year during which they will turn 5. So

perform significantly worse in exams than those born earlier in the school year, even up to GCSE level (NVQ2 level). A child born in September will, on average, perform better in academic tests than a child born in the following August, simply because they start school (and sit the tests) up to a year younger[18]. This means that access to further and higher education, and hence future success in the labour market, is likely to be affected by the month in which they are born. Indeed, Figure 4 shows, for males and females respectively, the percentage difference in the probability of obtaining a given NVQ qualification conditional on having a lower NVQ, for those born in September and August. The probability of having an NVQ1 conditional on having no qualification is lower for the oldest males and females in a school cohort (September born children). This because they have a higher probability to get a higher qualifications compared to the youngest children in the same school cohort (August born children). Formal tests of the validity and strength of our instruments suggest that we do not have a weak instrument problem. We assume the effect of month of birth on education is linear to improve the precision of our first stage estimates. That is, we set month of birth = [1...12] for September-August.

4 Estimation Results

4.1 Ordered probit first step results

We exclude month of birth and RoSLA from the wage equations. Both variables are in principle uncorrelated with the unobservable determinant of the earnings. The two exclusion restrictions satisfy the condition of the random assignment to treatment, in terms of the Angrist et al. (1996) causal model. We think of RoSLA as being a regression discontinuity once we control for social changes through the inclusion of a cubic in the year of birth. Indeed, it seems that there was almost complete compliance with RoSLA in England and Wales. And in the same way, month of birth of an individual, according to recent studies (e.g. Hoogerheide et al., 2007 and Kleibergen, 2002) appears to be uncorrelated with other covariates, unconditionally. Unlike

almost all children start school in September whilst aged 4, in what is called the Reception (kindergarten) class. Then they will be aged 5 by the time the school year ends in August 31st and at the start of Year 1. If children do not start until age 5, then they will start in Year 1 rather than Reception. And if they start Reception (kindergarten) in January or April, the only adjustment is in how much time they have in Reception. As this class is not so different from nursery school etc, this should not be an issue to use month of birth as exclusion restriction.

[18]Crawford et al. (2007) show that September born children have on average 0.2 year more education than August born children.

Buckles and Hungerman (2008) we cannot analyze the relationship between month of birth and family background with our data. However, we feel that the weight of evidence suggests that month of birth only has indirect effects on log wages - through the level of educational achievement. We also include in both the wage and education equations self-reported ethnicity (grouped simply into white and non-white).

In Table 3, we report the results of the first step of our estimation, i.e the ordered probit selection equation. We include, in this stage, a cohort effect represented by a cubic function of the year of birth to capture long run social changes, as distinct from the sharp effect of RoSLA. This social change turns out to be quite significant for educational attainment. The month of birth effect is captured by including a continuous month variable, where the omitted month is September which corresponds to the oldest children in each class cohort. We find, as expected, that its sign is negative and highly significant indicating that the oldest children in each class do better.[19].

We show, in Table 4, the marginal effects of RoSLA and month of birth evaluated by each level of education. The coefficient of RoSLA is significantly positive from NVQ0 to NVQ3, and it has its highest values at NVQ0 and NVQ2. The intention of the government and the consequent effect of the policy was to increase the participation at the lower secondary levels of education, and it is here where we find the strongest effect of the RoSLA. Note that if we had used RoSLA as an IV then the estimated effect of the NVQ would be weighted towards those at the bottom of the education distribution - e.g. for those who wanted to leave at 15 and the policy forces to remain in school until 16. In contrast, in our selection model, we estimate the effect on the mean. Looking at the month of birth, we notice that the coefficient is significantly positive and decreasing from NVQ0 to NVQ3, as expected, while it is significantly negative at NVQ3 and NVQ4: youngest children have a higher probability of achieving only lower levels of education, while older children in each class have a higher probability of attaining higher levels.

[19]To assess the stability of our exclusion restrictions we estimated the selection model applying one restriction at a time, and we have found that the results are substantially unchanged when using only RoSLA, while the model does not converge when using only month of birth. We have also performed pairwise comparison of different education levels by estimating separate probit models. The results show a stronger of effect of RoSLA and month of birth at lower levels of education, around 10 times the marginal effects obtained with the ordered probit; while the effects are much smaller at higher NVQ levels, also lower than ordered probit marginal effects.

4.2 Heckman estimation

To better interpret our results we show in Table 5 the estimation of the second step of the selection model where the dependent variable is raw log wage (i.e. without any age effect correction from a first difference model) and we allow for age effects and no cohort effects. This can be seen as a more conventional approach for the estimation of the selection model.

In Table 6 we report the results of the estimation of equation 2 and in the last column we present the estimates when we include the education variables, the purpose is to perform a test of non separability and, again, we reject the null of separability, so we are confident that the conventional separable model can be rejected in favour of of our non-separable one[20]. We also include the OLS estimates of log hourly earnings on quadratic age, using QLFS data only. In particular, we want to contrast the lifecycle earnings profile estimated without cohort effects from the pooled cross section data with the estimates of lifecycle earnings profiles estimated when we allow for additive cohort effects using the panel data. The estimates from the first difference model are immune from cohort effects, that is the effect of experience from being employed 1 year. The traditional estimates of the model without cohort effects generate lifecycle effects that are contaminated by the omitted cohort effects. Notice that age is not as good a proxy for experience for women as it is for men.

Finally, in Table 7 we show the estimation with the wages corrected for lifecycle effects and we include cubic year of births that capture additive cohort effects. The first important difference is the coefficient of the IMR, λ, which in Table 5 is always significant for females (except NVQ2), and significant at NVQ0, NVQ3 and NVQ4 for males.[21] This means that our exclusion restrictions are detecting and correcting for the presence of selection bias. In Table 7, we notice that λ is never significant for males, this suggests that the selection bias, in terms of unobserved fixed effects, has already been removed in the first difference estimation of the log wages (i.e. equation 2). For females, we still find evidence of sample selection, although less strong than in Table 5.[22] The coefficients of the cubic year of births are not reported, but the cohort effects are always jointly significant (we tested these effects both by and across NVQ levels), and this confirms that when we separately estimate the life cycle effects we identify cohort effects unconfounded by collinearity.

[20]We have also tested the model using cubic age effects, however this effect is mostly insignificant and an F test of the significance of the cubic in all NVQ levels rejects this extension.

[21]See tests on joint significance of λ at the bottom of Table 5.

[22]See tests on joint significance of λ at the bottom of Table 7.

4.3 Predicted Wages and Age-Earnings Profiles

In our ordered probit selection models (that assume that educational qualifications are correlated with the unknown factors that determine earnings) the returns to schooling do not appear among the estimated coefficients of the wage equation. They can be obtained from the estimated increments to predicted earnings that are associated with the successive levels of NVQ - that is, from the constant terms. In Table 8 we compare the average predicted wages obtained from the two selection models estimated.

In the top panel of Table 8 we report the average predicted wages (from the estimation in Table 5), which include age effects but no cohort effects. In the middle panel we show the predicted wages (from the estimation in Table 7) which do not include age effects, since the log wages have been explicitly corrected for them, but allow for cohort effects. We notice that the predicted wages from the more conventional approach are higher, for both males and females, compared to our new approach. In the bottom panel of Table 8, we show the wage differences by NVQ levels, under the two selection models[23]. Looking at the college premium, it is line with the literature: females get a higher premium of around 42% while for males the premium is around 37% in model *a*, while the premiums are 29% and 36% in model *b* for males and females, respectively. The returns to NVQ3 versus NVQ2 are much higher for males than females, in model *b*. This is consistent with our raw data, since we have more males with vocational qualifications than females. As we stated above, in the selection model we are estimating the effect of education on earnings across the whole distribution of unobservables, in particular at the mean. In fact, if we add back to the fitted wages in model *b* the age effects estimated in equation 2, we obtain returns to education in the two selection models that are practically the same. However, the fact that the at mean the predicted earnings are similar does not imply the absence of differences across the entire distribution. Indeed, to highlight these differences we compute the age-earnings profiles.

Figure 5 shows the profiles obtained from the raw pooled data, that is from the OLS estimation (see Table 6) using quadratic age with discrete schooling groups and no cohort effects. These profiles are identical[24] to those obtained from the estimation (see Table 5) of the conventional selection model with age effects only. In Figure 5, we observe the well-know convex shape of the profiles, where for males the peak is at age 45 with a college premium of around 40% whilst for females the peak is at 46 years old with a college

[23]For simplicity, we call model *a* the conventional selection model and model *b* the new approach.

[24]For this reason we do not report them.

premium higher than 40%. It is evident that the age-earnings profile for NVQ4 is higher than NVQ3 at all ages, and steeper than NVQ3 at early ages for both males and females. Notice that we find that women with low qualification levels have flatter lifecyle wage profiles.

Figure 6 shows the profiles obtained from the estimation in Table 7 when controlling for life cycle effects but assuming no cohort differences. We observe increasing profiles, which implies strong age effects throughout lifecycle for all educational levels. This clearly contrasts Figure 5, because now we have profiles where age effects are immune from cohort effects (see Table 6). Finally, in Figure 7 we show our last set of profiles which combine all our extensions to the simple workhorse specification used in the literature. We consider discrete groups of educational qualifications, we control for lifecycle effects and we separately allow for cohort differences. We find two clear results, the age earnings profiles are flat and younger cohorts have smaller college premiums compared to older cohorts, for both males and females. In Figure 8 we report the returns to education for high school (NVQ3 minus NVQ2) and higher education (NVQ4 minus NVQ3) for the age groups in the overlapping cohorts showed in Figure 7. For example, a 44-year-old graduate male (female) from the cohort 1950-55 has a college premium of around 50% (43%) while at the same age a graduate male (female) from the cohort 1960-65 has a premium of 36% (35%). Whereas, a 36-year-old graduate male (female) from the cohort 1960-65 has a college premium of around 33% (34%) while at the same age a graduate male (female) from the cohort 1970-75 has a premium of 17% (29%). Looking at the gender differences, males from the older cohort have higher college premiums than females; the premiums are almost similar for the middle cohort, while females from the younger cohort have a bigger premium than males.

5 Conclusion

This paper has proposed and implemented a simple methodology, to estimate the returns to education, that is sufficiently tractable that it could used with many datasets, and yet provides a significant generalisation of the usual additively separable and linear human capital earnings function. We separately estimated lifecycle and cohort effects, and identification was achieved through exploiting an education reform and month of birth. While the former is well known there are few examples of the latter. Both have significant effects on educational attainment. Our results amount to a strong rejection of the simple workhorse specification that is commonly used in this literature and we hope that our own specification will, ultimately, replace this workhorse to provide the departure point for further extensions.

References

Altonji J.G. and Pierret, C.R. (2001), Employer Learning And Statistical Discrimination, *The Quarterly Journal of Economics*, MIT Press, Vol. 116(1), pp. 313-350, February.

Angrist, J. and Inbens, G. and Rubin, D. (1996) Identification of Causal effects Using Instrumental Variables, *Journal of Econometrics*, Vol. 71, No. 1-2, pp. 145-160.

Angrist, J. and Krueger A. (1991), Does Compulsory Schooling Attendance Affect Schooling and Earnings?, *Quarterly Journal of Economics*, Vol. 106, pp. 979-1014.

Ashenfelter, O. and Rouse, C. (1998), Income, schooling and ability: Evidence from a new sample of identical twins, *Quarterly Journal of Economics*, Vol. 113, pp. 253-284.

Ashworth J. and Heyndels B. (2007) Selection Bias and Peer Effects in Team Sports, *Journal of Sport Economics*, Vol. 8, No. 4, pp. 355-377.

Bedard, K. and Dhuey, E. (2006), The persistence of early childhood maturity: international evidence of long-run age effects, *Quarterly Journal of Economics*, Vol. 121, pp. 1437-72.

Belzil, C. (2007), The return to schooling in structural dynamic models: a survey, *European Economic Review*, Elsevier, vol. 51(5), pages 1059-1105.

Blundell, R. Dearden, L. and Sianesi, B. (2005) Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey, *Journal of Royal Statististics Society A*, 168, Part 3, pp. 473-512.

Blundell, R. and Costa Dias, M. (2009), Alternative approaches to evaluation in empirical microeconomics, *Journal of Human Resources*, Vol. 44, No. 3, pp. 565-640, Summer 2009

Blundell, R. and Powell, J. L. (2004) Endogeneity in Semiparametric Binary Response Models, *Review of Economic Studies*, Vol. 71, No. 3, pp. 655-679, July.

Buckles, K. and Hungerman, D. M. (2008) Season of Birth and Later Outcomes: Old Questions, New Answers, *NBER Working Papers 14573*.

Cameron, S. V. and Taber, C. (2004) Estimation of Educational Borrowing Constraints Using Returns to Schooling, *Journal of Political Economy*, University of Chicago Press, vol. 112(1), pp. 132-182, February.

Card, D. (1994), Earnings, Schooling, and Ability Revisited, *NBER Working Papers 4832*, National Bureau of Economic Research.

Card, D. (1995), Using Geographic Variation in College Proximity to Estimate the Returns to Schooling, in: Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp. University of Toronto Press, pp. 201-222.

Card, D. (2001), Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica*, vol.69 n.5, pp. 1127-1160

Carneiro, P., Heckman, J.J and Vytlacil, E. (2010), Estimating Marginal Returns to Education, Forthcoming, *American Economic Review*.

Chevalier, A. (2004), Parental Education and Child's Education: A Natural Experiment, IZA Discussion Papers 1153, Institute for the Study of Labor (IZA).

Chiburis R. and Lokshin M. (2007) Maximum likelihood and two-step estimation of an ordered-probit selection model, *Stata Journal*, Vol. 7, issue 2, pp. 167-182

Crawford C., Dearden, L. and Meghir, C. (2007) When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England, *The Institute for Fiscal Studies*.

Das, M. Newey, W.K. and Vella, F. (2003), Nonparametric Estimation of Sample Selection Models, *Review of Economic Studies*, Blackwell Publishing, vol. 70(1), pages 33-58, January.

Griliches, Z. (1977), Estimating the Returns to Schooling: Some Econometric Problems, *Econometrica*, Vol. 45, pp. 1-22.

Harmon, C. and Walker, I. (1995), Estimates of the Economic Returns to Schooling for the United Kingdom, *American Economic Review*, Vol. 85, N. 5, pp. 1278-1286.

Heckman, J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, Vol. 47, No. 1, pp. 153-161.

Heckman, J. (1990), Varieties of Selection Bias, *AEA Papers and Proceedings*, Vol. 80(2), pp. 313-318.

Heckman, J.J., Layne-Farrar, A., Todd, P.E. (1996). Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *Review of Economics and Statistics* 78 (4), 562-610. November.

Heckman, Lochner L. and Todd P. (2006), Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, *Handbook of Education Economics*, Vol. 1, ed. E. Hanushek and F. Welch, chapter 7, Elsevier.

Heckman, Lochner L. and Todd P. (2008) Earnings Functions and Rates of Return, *Journal of Human Capital*, Vol. 2, no. 1

Heckman, J. and Vytlacil, E. (1998), Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is correlated with Schooling, *The*

Journal of Human Resources, Vol. 33, No. 4 , pp. 974-987.

Hoogerheide, L., Kleibergen, F. and Van Dijk., H. (2007), Natural Conjugate Priors for the Instrumental Variables Regression Model Applied to the Angrist-Krueger Data, *Journal of Econometrics*, 138, 63-103.

Hungerford, T., Solon, G. (1987). Sheepskin effects in the returns to education. *Review of Economics and Statistics* 69 (1), 175-177. February.

Ichino, A. and Winter-Ebmer, R. (1999), Lower and Upper Bounds of Returns to Schooling: An Exercise in IV Estimation with Different Instruments, *European Economic Review*, Vol. 43, pp. 889-901

Jaeger, D.A., Page, M.E. (1996), Degrees matter: New evidence on sheepskin effects in the returns to education. *Review of Economics and Statistics* 78 (4), 733-740. November.

Kenny, L.W. and Lee, L.-F. and Maddala, G.S. and Trost, R.P. (1979) Returns to college education: An investigation of self-selection bias based on the project talent data, *International Economic Review* Vol. 20, pp. 775-789.

Kleibergen, F. (2002), Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, *Econometrica*, 70, 1781-1803.

Kling, J. R. (2001), Interpreting Instrumental Variables Estimates of the Returns to Schooling *Journal of Business and Economic Statistics* 19(3), 358-364, July.

Lanot, G. and Walker, I. (1998), The union/non-union wage differential: An application of semi-parametric methods, *Journal of Econometrics*, Elsevier, vol. 84, no.2, pp. 327-349, June.

Makepeace G., Woods L., Joshi H., Dolton P. and Galinda-Rueda F. (2003), From School to the Labour Market, (eds) Changing Britain, Changing Lives Three Generations at the Turn of the Century, Institute of Education.

Moffit, R. A. (1999), Models of treatment effects when responses are heterogenous, Commentary *Proceeding National Academic Science* vol.96, pp.6575-6576, USA.

Neal, D. A. (2004), The Measured Black-White Wage Gap among Women Is Too Small *Journal of Political Economy*, 112(1), pp. S1-28.

Oreopoulos, P. and Page, M.E. (2006). The Intergenerational Effects of Compulsory Schooling, *Journal of Labor Economics*, University of Chicago Press, vol. 24(4), pages 729-760, October.

Puhani, P. and Weber, A. (2005), Does the early bird catch the worm? Instrumental variable estimates of educational effects of age of school entry in Germany, *IZA Discussion Paper no. 1827*.

Stewart, M.B. (2004), Semi-Nonparametric Estimation of Extended Ordered Probit Models, Stata Journal, 4, 27-39.

Walker, I. and Zhu, Y. (2007), The Labour Market Effects of Qualifications with Special Reference to Scotland, *Report* N.1 for the Government of Scotland.

Walker I. and Zhu, Y. (2008), The college wage premium, over education and the expansion of higher education in the UK, Scandinavian Journal of Economics, vol.10, 4.

Willis, R. J. and Rosen, S. (1979), Education and Self-Selection, Journal of Political Economy, vol. 87, no. 5, pt. 2, pp.S7-36.

Table 1: NVQ Equivalent Qualifications

NVQ level	Examples of academic qualifications	Examples of vocational qualifications
5	PhD Masters degree	PGCE (Teaching) Non-masters postgraduate quals
4	Undergraduate degree	Other teaching quals, HE below degree, RSA Higher, HNC/HND, Nursing
3	2+ A-levels 3+ SCE Highers AS-level or equivalent (4+)	GNVQ - advanced, RSA advanced dip/cert, OND, ONC, BTEC, City and Guilds Advanced Craft , Scot. Cert 6th year Studies, Higher national qualification or equivalent (Scotland) , ,(other) NVQ Level 3
2	1 A-level, 1-3 AS levels, 1-2 Highers 5+ GCSEs at A-C	GNVQ – intermediate, RSA diploma, City and Guilds - Craft , BTEC, SCOTVEC etc. first or general diploma , Scottish CSYS , Intermediate 2 Higher qualification (Scotland), (other) NVQ Level 2
1	1-4 GCSEs/SCE passes (but below Level 2)	GNVQ, GSVQ foundation level, CSE below grade1, GCSE below grade C, BTEC, SCOTVEC first or general certificate, SCOTVEC modules, RSA other, City & Guilds other , YT/YTP certificate , Key Skills, Basic Skills
Apprenticeships	n.a.	Modern and Traditional
Other	n.a.	Unspecified in the data

Source: Walker and Zhu (2007)

Table 2: Summary statistics

Males

	Mean	Std. Dev.	Min.	Max.	N
log hourly wage (Sept09 prices)	2.474	0.487	0.786	4.060	116913
RoSLA	0.627	0.484	0	1	119700
age	41.364	9.577	25	60	119700
year of birth	60.688	9.888	40	84	119700
month of birth	6.392	3.42	1	12	119700
non white	0.982	0.134	0	1	119700

Females

	Mean	Std. Dev.	Min.	Max.	N
log hourly wage (Sept09 prices)	2.198	0.486	0.781	4.059	127696
RoSLA	0.626	0.484	0	1	129874
age	41.43	9.497	25	60	129874
date of birth year	60.715	9.838	40	84	129874
date of birth month	6.394	3.415	1	12	129874
non white	0.98	0.142	0	1	129874

log hourly wage and NVQ percentage

Males

	Mean	Std. Dev.	%
NVQ0	2.105	0.360	10.35
NVQ1	2.183	0.358	5.23
NVQ2	2.343	0.417	19.16
NVQ3	2.434	0.413	30.89
NVQ4	2.807	0.458	34.37

Females

	Mean	Std. Dev.	Perc.
NVQ0	1.829	0.329	13.82
NVQ1	1.925	0.333	7.30
NVQ2	2.075	0.385	30.27
NVQ3	2.174	0.418	16.17
NVQ4	2.589	0.441	32.44

Table 3: First Step - Selection model: ordered probit

Dep var: NVQ levels		
	Males	Females
month of birth	-0.00164*	-0.00371***
	(0.00095)	(0.00090)
non white	-0.15139***	-0.25877***
	(0.02464)	(0.02247)
RoSLA	-0.05980***	-0.07458***
	(0.01523)	(0.01424)
year of birth	0.39362***	0.48996***
	(0.03444)	(0.03353)
(year of birth)^2	-0.00631***	-0.00753***
	(0.00059)	(0.00057)
(year of birth)^3	0.00003***	0.00004***
	(0.00000)	(0.00000)
cut1		
constant	6.75712***	9.54308***
	(0.66443)	(0.64890)
cut2		
constant	7.00923***	9.84784***
	(0.66444)	(0.64892)
cut3		
constant	7.63021***	10.71901***
	(0.66449)	(0.64898)
cut4		
constant	8.42595***	11.14824***
	(0.66454)	(0.64901)
N	105088	117140

Significance levels : $*$ 10% $**$ 5% $***$ 1%

std. err. in brackets.

Ho: yob=yob^2=yob^3=0, rej at 1%

LR $\chi^2(3)$=535.66 for males

LR $\chi^2(3)$=3365.56 females.

Ho: $rosla$=$month\,of\,birth$=0, rej at 1%

LR $\chi^2(2)$=17.72 for males

LR $\chi^2(2)$=41.83 for females.

Table 4: First Step - ordered probit - marginal effects

Dep var: NVQ levels	Males				Females			
	RoSLA			month of birth	RoSLA			month of birth
	Before	After	Δ		Before	After	Δ	
NVQ0	0.09730	0.10793	0.01063	0.00029	0.12937	0.14501	0.01564	0.00078
	(0.00177)	(0.00147)	(0.00268)	(0.000177)	(0.00174)	(0.00178)	(0.00295)	(0.00019)
NVQ1	0.05036	0.05388	0.00352	0.00010	0.07165	0.07663	0.00498	0.00027
	(0.00086)	(0.00076)	(0.00089)	(0.00006)	(0.00096)	(0.00083)	(0.00093)	(0.00007)
NVQ2	0.18622	0.19397	0.00775	0.00021	0.29734	0.30520	0.00786	0.00043
	(0.00177)	(0.00135)	(0.00197)	(0.00012)	(0.00192)	(0.00131)	(0.00150)	(0.00010)
NVQ3	0.30740	0.30747	-0.00007	-0.00001	0.15970	0.15707	-0.00263	-0.00016
	(0.00141)	(0.00140)	(0.00008)	(0.00000)	(0.00104)	(0.00110)	(0.00046)	(0.00004)
NVQ4	0.35869	0.33672	-0.02197	-0.00060	0.34191	0.31606	-0.02585	-0.00132
	(0.00397)	(0.00237)	(0.00560)	(0.00035)	(0.00371)	(0.00194)	(0.00493)	(0.00032)

All estimates are significant at 1% (for simplicity we omit the stars to indicate significance). Std. err. in brackets.

Table 5: Estimates Heckman model - no cohort effects

Dep var: log earnings

Males

	NVQ0	NVQ1	NVQ2	NVQ3	NVQ4
λ	0.84652***	0.41899	-0.23196	1.43493***	1.24955***
	(0.31514)	(0.35749)	(0.20949)	(0.35164)	(0.34213)
age	0.04829***	0.05772***	0.06048***	0.08089***	0.10276***
	(0.00542)	(0.00597)	(0.00354)	(0.00642)	(0.00476)
age^2	-0.00058***	-0.00067***	-0.00063***	-0.00104***	-0.00119***
	(0.00008)	(0.00009)	(0.00005)	(0.00009)	(0.00007)
non white	-0.04467	-0.02572	0.08839**	-0.19812***	-0.12682***
	(0.05323)	(0.06687)	(0.03886)	(0.06705)	(0.04700)
constant	2.66633***	1.50969***	0.74473***	1.14093***	-0.50060
	(0.55700)	(0.45248)	(0.16755)	(0.10345)	(0.36733)
N	105086				

Females

	NVQ0	NVQ1	NVQ2	NVQ3	NVQ4
λ	-0.45159***	-0.04212	-0.41407***	-0.81721***	-0.24505***
	(0.09318)	(0.09766)	(0.07309)	(0.11931)	(0.08964)
age	0.01695***	0.00012	0.01500***	0.03733***	0.04109***
	(0.00359)	(0.00378)	(0.00241)	(0.00400)	(0.00217)
age^2	-0.00006*	0.00007*	-0.00004	-0.00023***	-0.00040***
	(0.00003)	(0.00004)	(0.00003)	(0.00005)	(0.00003)
non white	0.00921	-0.10179***	-0.01622	0.11228***	0.02980
	(0.03561)	(0.03869)	(0.02619)	(0.04329)	(0.02168)
constant	0.48442*	1.85177***	1.39843***	1.14213***	1.85605***
	(0.26261)	(0.21604)	(0.10942)	(0.12122)	(0.05805)
N	117136				

Significance levels : $*$ 10% $**$ 5% $***$ 1% Std. err. in brackets.

Dependent Variable in 2009 prices.

Ho: $\lambda_0 = \cdots = \lambda_5 = 0$ $\chi^2_5 = 39.81$ for males; $\chi^2_5 = 110.16$ for females;

Table 6: First difference model from LLFS and OLS from QLFS

	Males					
	NVQ0	NVQ1	NVQ2	NVQ3	NVQ4	NVQ tot
Dep var: first difference log hourly earnings - LLFS						
2×age	-0.00053***	-0.00024	-0.00046***	-0.00052***	-0.00077***	-0.00059***
	(0.00020)	(0.00030)	(0.00014)	(0.00011)	(0.00010)	(0.00006)
NVQ1						0.00522
						(0.00607)
NVQ2						0.00409
						(0.00446)
NVQ3						0.00429
						(0.00408)
NVQ4						0.01599***
						(0.00406)
constant	0.09120***	0.07553***	0.09042***	0.09468***	0.12758***	0.09636***
	(0.01872)	(0.02302)	(0.01156)	(0.00933)	(0.00869)	(0.00667)
N	4889	2688	9394	16309	17739	51019
Dep var: log hourly earnings - QLFS - OLS model[a]						
age	0.03848***	0.05703***	0.06584***	0.06643***	0.09829***	0.07566***
	(0.00339)	(0.00495)	(0.00276)	(0.00212)	(0.00220)	(0.00120)
age^2	-0.00041***	-0.00062***	-0.00070***	-0.00076***	-0.00107***	-0.00083***
	(0.00004)	(0.00006)	(0.00003)	(0.00003)	(0.00003)	(0.00001)
N	11047	5612	20464	33233	36882	107238

	Females					
	NVQ0	NVQ1	NVQ2	NVQ3	NVQ4	NVQ tot
2×age	-0.00012	-0.00039*	-0.00040***	-0.00055***	-0.00061***	-0.00046***
	(0.00019)	(0.00021)	(0.00011)	(0.00015)	(0.00011)	(0.00006)
NVQ1						0.01524***
						(0.00488)
NVQ2						0.00722**
						(0.00360)
NVQ3						0.00815**
						(0.00409)
NVQ4						0.01860***
						(0.00357)
constant	0.05293***	0.09432***	0.08722***	0.10067***	0.11566***	0.08534***
	(0.01801)	(0.01844)	(0.00921)	(0.01217)	(0.00913)	(0.00654)
N	7935	4532	17973	8992	18774	58206
OLS model						
age	0.00458	0.00285	0.01320***	0.03051***	0.04300***	0.02318***
	(0.00290)	(0.00362)	(0.00201)	(0.00285)	(0.00214)	(0.00111)
age^2	-0.00002	0.00003	-0.00013***	-0.00037***	-0.00047***	-0.00024***
	(0.00003)	(0.00004)	(0.00002)	(0.00003)	(0.00003)	(0.00001)
N	16531	8691	36198	19262	38852	119534

Significance levels : $*$ 10% $**$ 5% $***$ 1% Std. err. in brackets. Dependent Variable in 2009 prices.
LLFS: Test on separability for males rej. Ho at 5%, Fstat=5.744; for females rej. Ho at 5%, Fstat=6.013
NVQ tot is a categorical variable for each NVQ level.

[a]Basic OLS model with no cohort effects

Table 7: Estimates Heckman model - with cohort effects

Dep var: log earnings corrected for lifecycle effects

Males

	NVQ0	NVQ1	NVQ2	NVQ3	NVQ4
λ	-0.17056	-0.03264	-0.00466	0.22527	0.32310
	(0.34337)	(0.39332)	(0.24308)	(0.19223)	(0.30438)
non white	0.08216	0.04533	0.04851	-0.00744	0.00195
	(0.05829)	(0.06748)	(0.04123)	(0.03477)	(0.03622)
constant	10.28976***	8.86501***	8.22799***	7.47121***	9.79283***
	(2.27805)	(3.27755)	(2.11201)	(1.81581)	(2.47798)
N	105088				

Females

	NVQ0	NVQ1	NVQ2	NVQ3	NVQ4
λ	-0.30541*	-0.69032***	-0.22152*	0.14598	-0.52381**
	(0.18251)	(0.29960)	(0.14273)	(0.18735)	(0.24716)
non white	-0.00096	0.08612	-0.06634*	-0.13125***	0.09897***
	(0.05135)	(0.09033)	(0.03957)	(0.05313)	(0.04576)
constant	7.62182***	19.72844***	15.21160***	11.63619***	21.77756***
	(1.76624)	(3.92310)	(1.8370)	(2.59009)	(2.76227)
N	117140				

Significance levels : $*$ 10% $**$ 5% $***$ 1% Bootstrapped Std. err. (500 reps) in brackets.
Dependent Variable in 2009 prices. All equations include cubic year of births that capture additive cohort effects. Age effects are imposed from Table 6, according to equation 5.
Ho: $\lambda_0 = \cdots = \lambda_5 = 0$ $\chi^2_5 = 3.80$ for males; $\chi^2_5 = 26.88$ for females;

Table 8: Predicted wages

Selection Model [a] with age effects and no cohort effects

	Males		Females	
	Mean	Std. dev	Mean	Std. dev
NVQ0	2.10580	0.04469	1.82674	0.02884
NVQ1	2.18321	0.07158	1.92295	0.04664
NVQ2	2.34460	0.08971	2.07269	0.03076
NVQ3	2.43373	0.07026	2.17044	0.04195
NVQ4	2.80956	0.12422	2.58901	0.05654

Selection Model [b] with cohort effects and lifecycle correction

	Males		Females	
	Mean	Std. dev	Mean	Std. dev
NVQ0	2.10297	0.27735	1.65489	0.0838
NVQ1	1.89639	0.15235	1.72606	0.21022
NVQ2	2.10852	0.27909	1.8362	0.19759
NVQ3	2.31797	0.27186	1.92789	0.27214
NVQ4	2.61111	0.45962	2.28441	0.35534

Wage differences by NVQ

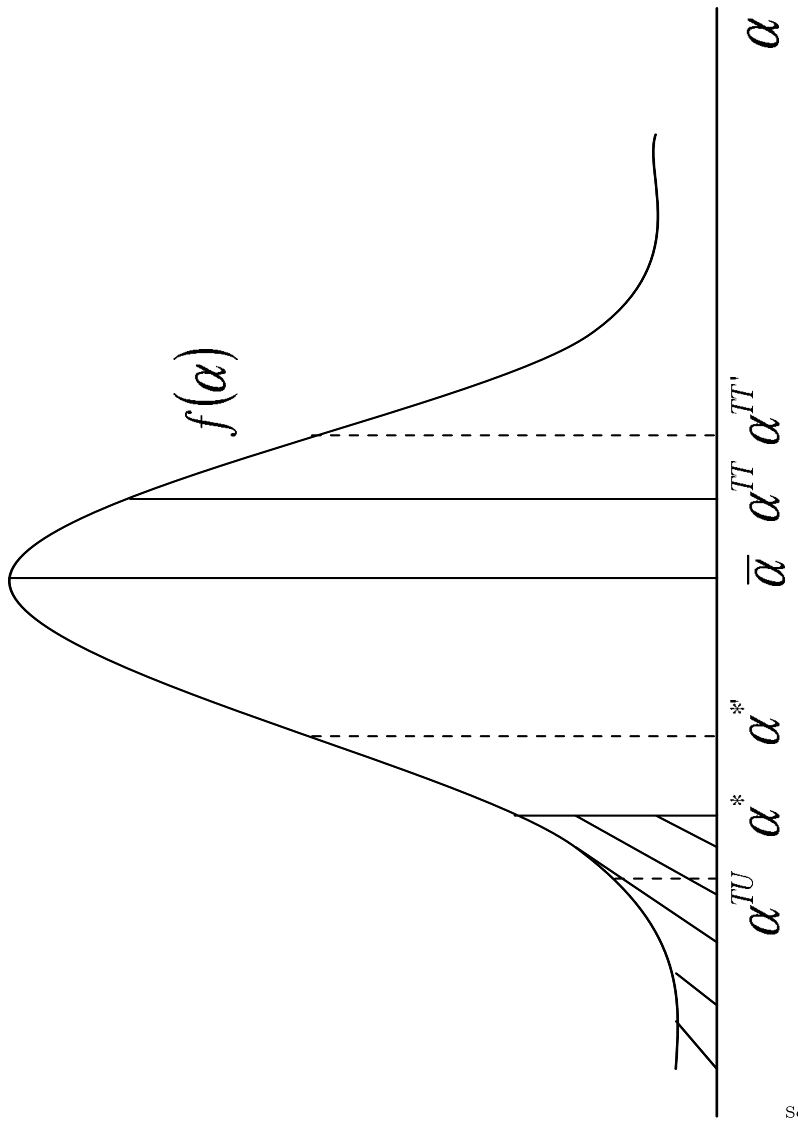
	Males		Females	
	SM [a]	SM [b]	SM [a]	SM [b]
NVQ1-NVQ0	0.07741	-0.20657	0.0962	0.07117
NVQ2-NVQ1	0.16139	0.21212	0.14975	0.11014
NVQ3-NVQ2	0.08913	0.20945	0.09775	0.0917
NVQ4-NVQ3	0.37583	0.29314	0.41857	0.35651

[a] Predicted wages and differences from Table 5.

[b] Predicted wages and differences from Table 7.

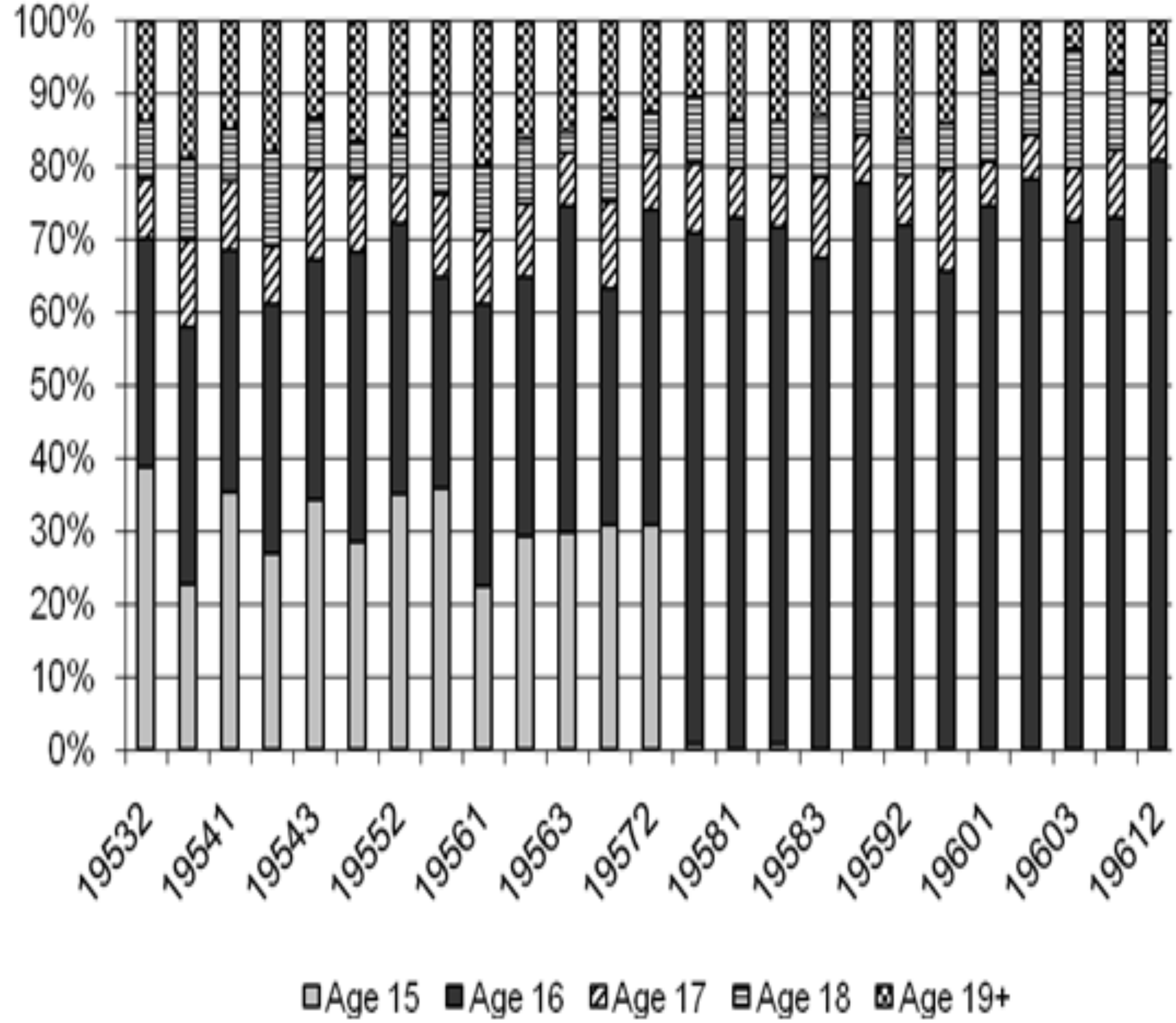
Note: If we add the age effects to the predictions in model b we obtain average wage predictions similar to model a.

Figure 1: Density of Treatment gains in population



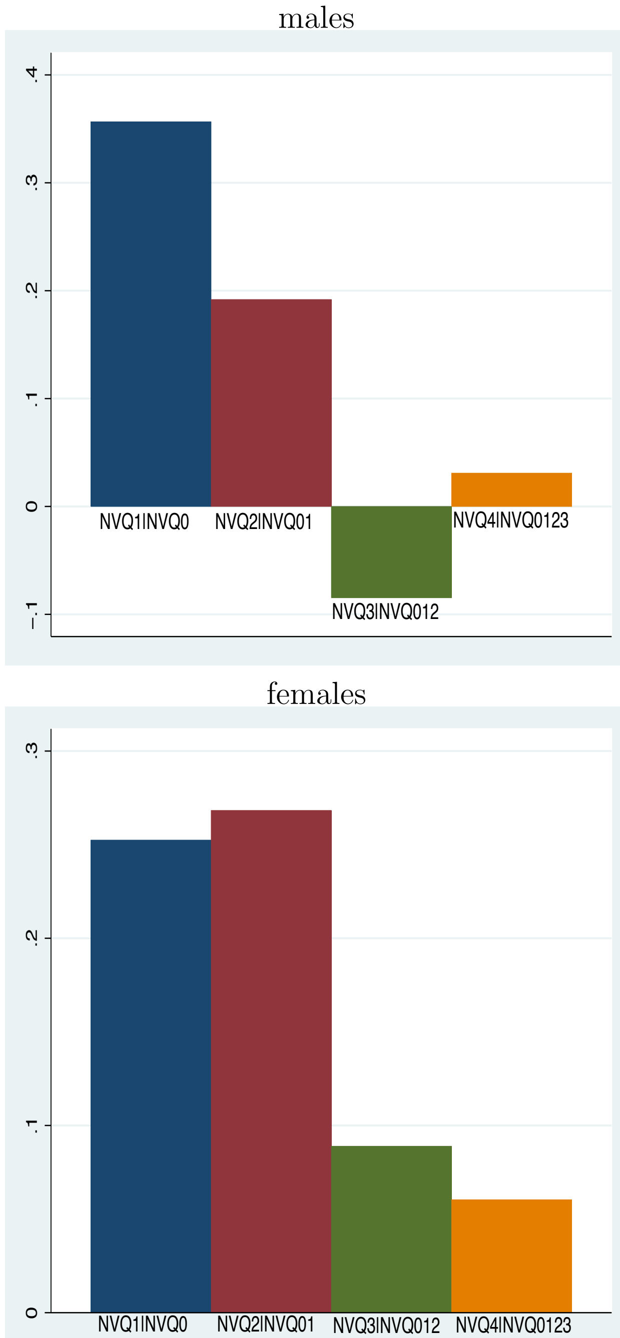
Source: Moffit (1999)

Figure 2: RoSLA affects only bottom of S distribution



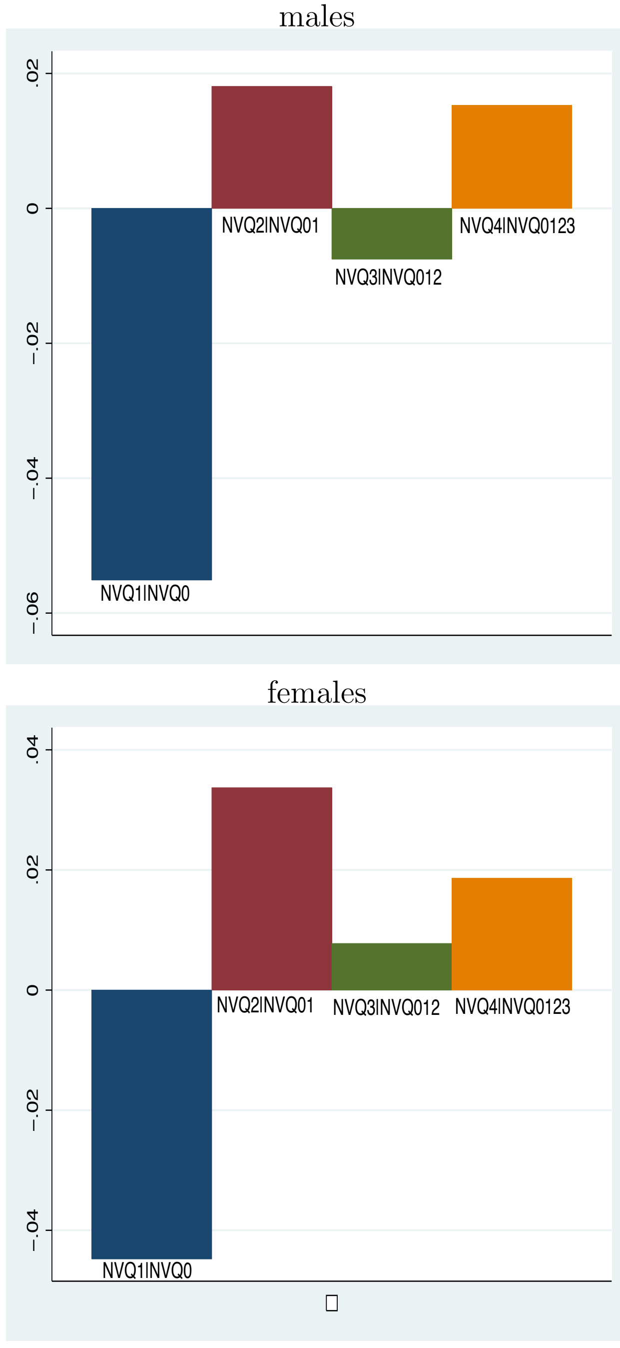
Source: Chevalier et al. (2004)

Figure 3: Percentage Difference of highest NVQ by RoSLA



Note: Each NVQ level is conditional to previous levels achieved.

Figure 4: Proportion September-August born by NVQ



Note: Each NVQ level is conditional to previous levels achieved.

Figure 5: Raw LFS pooled data - OLS quadratics with discrete S groups and no cohort effects

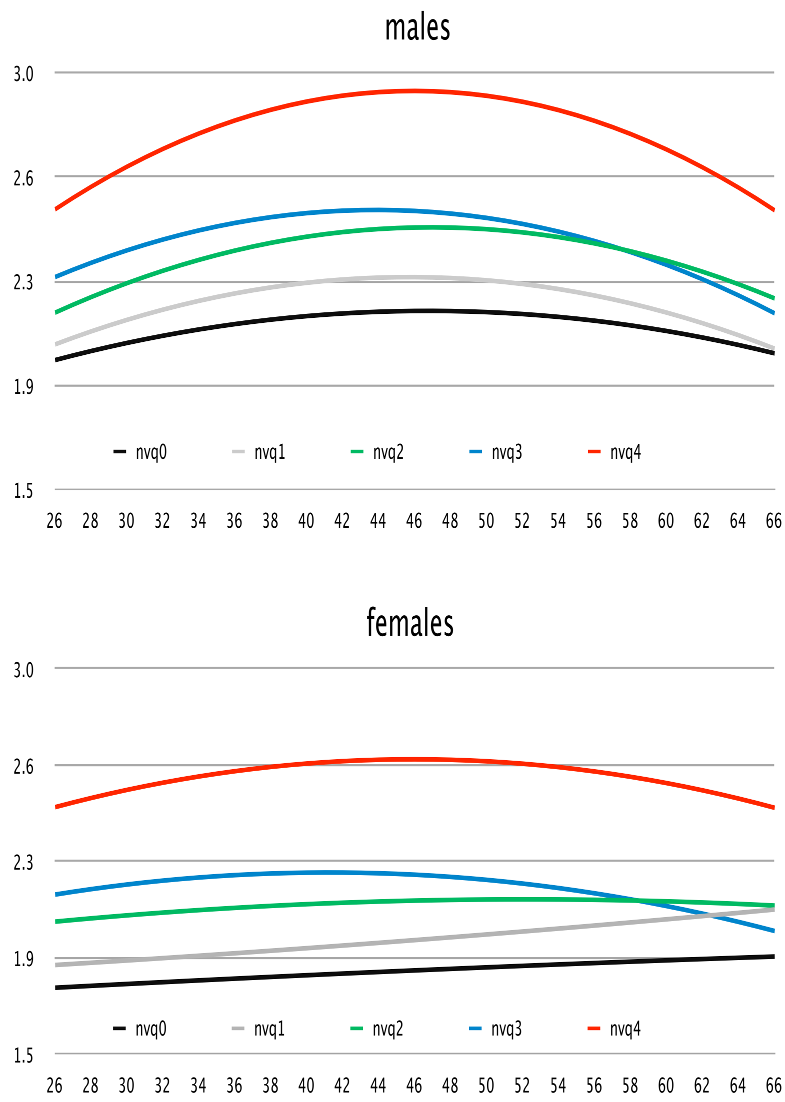
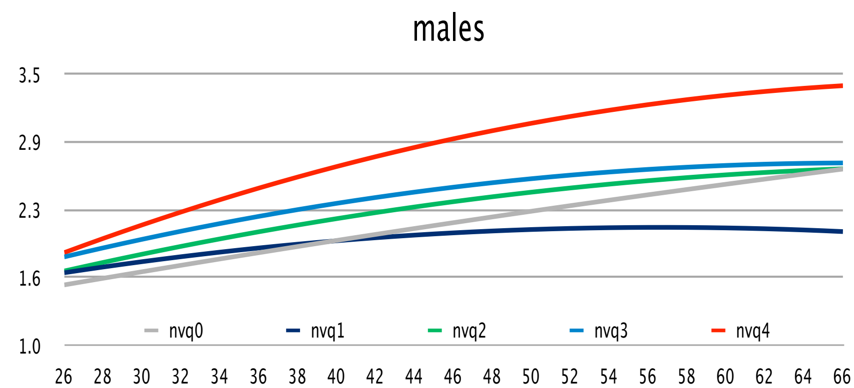


Figure 6: Lifecycle effects - Quadratic in age, assuming no cohort differences



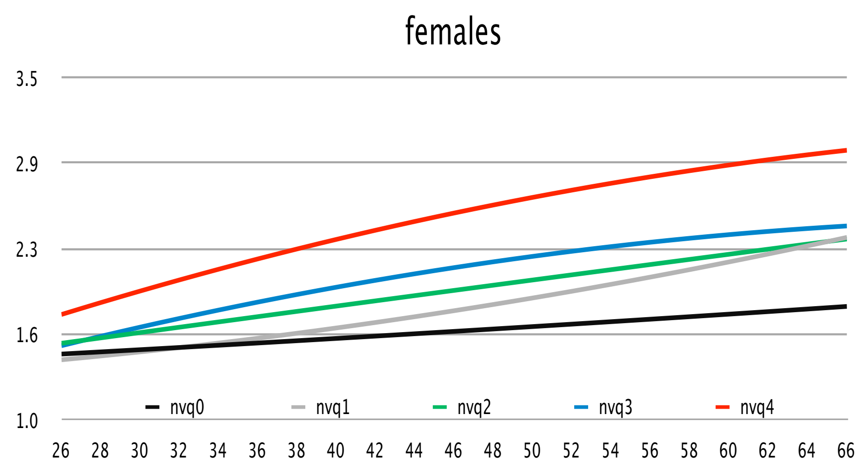


Figure 7: Lifecycle effects - Allowing for cohort differences

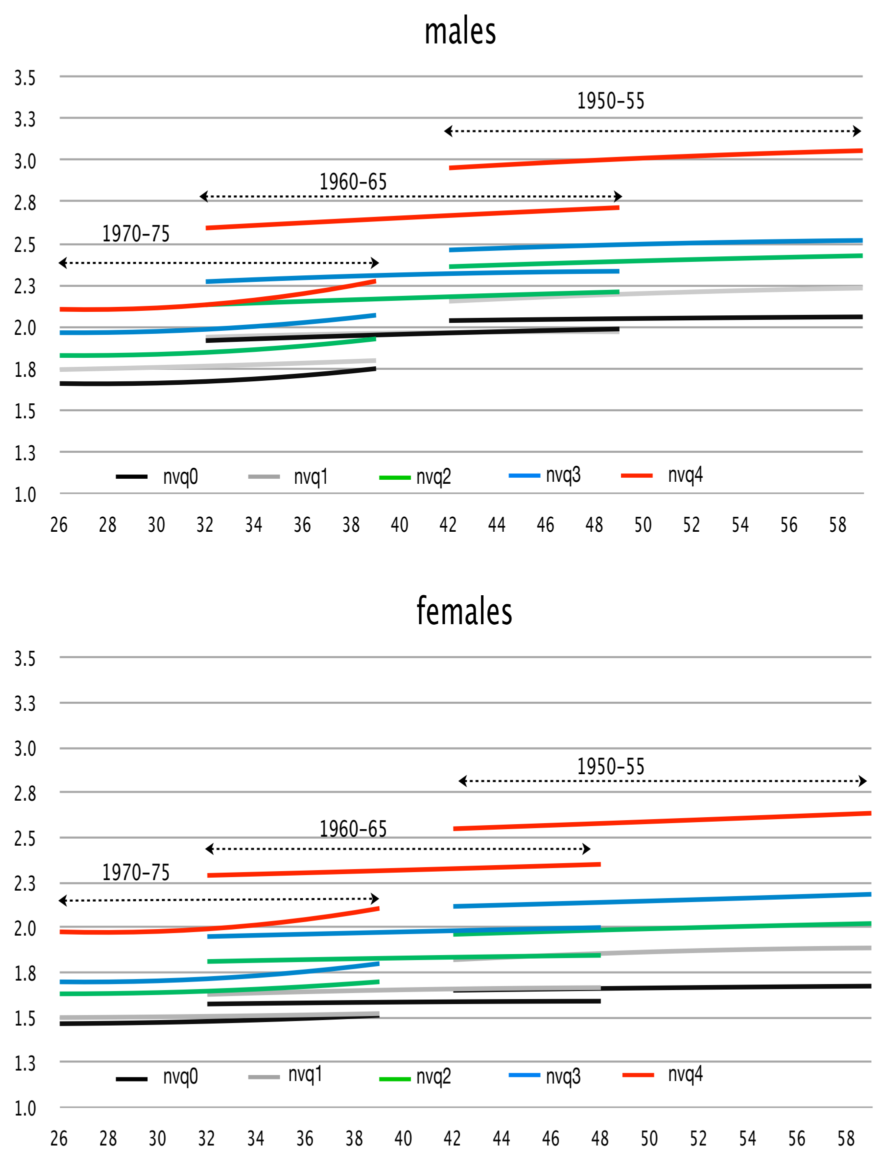


Figure 8: Returns to education by cohort - High school and Higher education