DISCUSSION PAPER SERIES

IZA DP No. 4419

The Causal Effect of Education on Wages Revisited

Matt Dickson

September 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Causal Effect of Education on Wages Revisited

Matt Dickson
CMPO, University of Bristol and IZA

Discussion Paper No. 4419
September 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4419
September 2009

ABSTRACT

The Causal Effect of Education on Wages Revisited*

This paper estimates the return to education using two alternative instrumental variable estimators: one exploits variation in schooling associated with early smoking behaviour, the other uses the raising of the minimum school leaving age. Each instrument estimates a 'local average treatment effect' and my motivation is to analyse the extent to which these differ and which is more appropriate for drawing conclusions about the return to education in Britain. I implement each instrument on the same data from the British Household Panel Survey, and use the over-identification to test the validity of my instruments. I find that the instrument constructed using early smoking behaviour is valid as well as being strong, and argue that it provides a better estimate of the average effect of additional education, akin to ordinary least squares but corrected for endogeneity. I also exploit the dual sources of exogenous variation in schooling to derive a further IV estimate of the return to schooling. I find the OLS estimate to be considerably downward biased (around 4.6%) compared with the IV estimates of 12.9% (early smoking), 10.2% (RoSLA) and 12.5% (both instruments).

JEL Classification: I20 J30

Keywords: human capital, endogeneity, local average treatment effect

Corresponding author:

Matt Dickson
Centre for Market and Public Organisation
University of Bristol
2 Priory Road
Bristol BS8 1TX
United Kingdom
E-mail: m.r.dickson@bristol.ac.uk

* Many thanks for useful comments to Ian Walker, Mark Stewart, Simon Burgess, Sarah Smith, Colm Harmon, Jennifer Smith, Paul Gregg and seminar participants at University of Warwick and the Centre for Market and Public Organisation.

1 Introduction

This paper estimates the causal effect of education on wages using two alternative methods of instrumentation. I compare estimates that are derived using variations in schooling associated with early smoking behaviour, with estimates derived by exploiting the impact on schooling of the raising of the minimum school leaving age. The latter instrument follows in the tradition of Card (1995) and similar papers[1], which use institutional factors or elements of the budget constraint to create instruments. This earlier research using instrumental variable methods covers a wide range and my work here is motivated by the worry that these instrumental variable methods identify a 'local average treatment effect' which might be rather different to the average effect on the treated and that will differ across instruments. These IV estimates isolate the return to education for the group whose education decision is most affected by the institutional feature exploited or the change in their own budget constraint, which may be quite a specific and unrepresentative group. The raising of the minimum school leaving age affected only those who that had wanted to leave school early and therefore, in this case, IV estimates the effect of additional schooling for those at the bottom of the schooling distribution who were forced to stay longer. In contrast, I find that early smoking affects the schooling decisions of individuals across the whole of the distribution – that is, it is not only individuals at a certain point in the schooling distribution who are affected. I interpret the estimates from this latter exercise as closer to an average effect of additional schooling akin to least squares but corrected for endogeneity. My contribution is to investigate the extent to which this effect differs from the local effect at the bottom of the distribution, implementing the alternative instrumental variables strategies using that same data from the British Household Panel Survey. In addition, as I have multiple instruments I am able to test the validity of the exclusion restrictions, something that is rarely possible to do, and also to simultaneously exploit two differing sources of exogenous variation in order

[1]The first notable paper to use instrumental variables to estimate the return to education was Angrist and Krueger (1991). A UK study by Harmon and Walker (1995 *inter alia*) also exploited the minimum school leaving age change.

to derive a further estimate of the return to education. The next section introduces the problem of estimating the return to education, section 3 then discusses potential solutions. Section 4 proposes early smoking as an instrument for education, before section 5 describes the data. Section 6 explains the estimation procedure, section 7 the results and section 8 analyses these results and considers various tests of the instrument. Section 9 then compares the smoking instrument estimates with ones derived from the raising of the school leaving age, before section 10 exploits the presence of two instruments to formally test the validity of these instruments. Section 11 offers some concluding remarks.

2 The Problem of Estimating the Return to Education

The foundation of the education returns literature has been Mincer's (1974) human capital earnings function:

$$\ln w_i = X_i'\varphi + \beta S_i + \epsilon_i \tag{1}$$

in which w_i is the wage, X_i is a vector of the individual's characteristics, including experience and experience-squared, and S_i is the number of years of schooling, determined by:

$$S_i = X_i'\gamma + u_i \tag{2}$$

This human capital earnings function tells us the expected (log) wage that an individual will earn given his/her observable characteristics and years of education. It is well known that if this relationship in equation (1) is estimated by least squares the estimate of the parameter β can only be interpretable as the causal effect on wages of one additional year of schooling if $\mathrm{E}(X_i\ \epsilon_i) = 0$ and $\mathrm{E}(S_i\ \epsilon_i) = 0$. If however $\mathrm{E}(S_i\ \epsilon_i) \neq 0$, though we can still interpret the equation as the conditional expectation of ln w_i given X_i and S_i, we cannot interpret β as the *causal* effect of education on wages since education is endogenous with respect to the causal effect β. The potential for the unobserved characteristics that determine schooling choice to also be correlated with wage, has for a long time been a concern to labour economists. If we are to draw valid conclusions regarding the economic return to education we must isolate the *causal* effect of education on wages. Clearly this is not straightforward because of this concern: we anticipate that factors affecting the education choice an individual makes will also independently affect their earned wage, we expect $\mathrm{E}(u_i\ \epsilon_i) \neq 0$.

Earlier research concentrated on the issue of 'ability bias' which suggested that $\mathrm{E}(S_i\ \epsilon_i) > 0$ because the residual picks up ability which is positively correlated with both wages and schooling. This ability bias explanation suggested that OLS was unambiguously biased upwards. In contrast, in his influential paper of 1977, Griliches proposed that measurement error in the schooling variable would lead to an attenuation of the OLS coefficient on schooling, biasing it towards zero. Griliches concluded that 'ability bias' was in reality small and was overwhelmed by the bias introduced by measurement error, with the result that OLS under-estimated the actual return to education. Card (1994) reported that in the micro-survey data commonly used by labour economists, measurement error in the schooling variable accounts for approximately 10 percent of the variance in observed schooling. This would lead to a 10 percent attenuation bias in the OLS coefficient – and even more if other covariates in the regression are correlated with the *real* level of schooling (Card, 1994). This estimate of around 10% concurs with evidence from Ashenfelter and Kreuger (1995) (using data from twins and reporting the measurement error to be between 8% and 12%), and furthermore, studies in which the education variable is deemed to be much more reliably measured (for example Uusitalo (1999), in which the schooling information comes directly from school records) still find the IV estimates to be considerably higher than the OLS.

At the start of the 1990s, a number of economists suggested that OLS estimates of the return to education may suffer from a further bias – 'discount rate bias' (see Lang, 1993; Card, 1994). In Becker's model of human capital formation, with standard assumptions[2], an individual will accumulate human capital to the point where the marginal rate of return on the last unit of education is equal to his/her discount rate. To illustrate this: (see Kling

[2] i) workers maximize the discounted present value of lifetime wealth; ii) time in school is independent of time in work, or alternatively lifetimes are infinitely lived; iii) there are no direct costs of education; iv) the effect of experience on earnings is multiplicative.

2000) assume that the individual's earnings opportunities are summarized by the function y=$g(S)$ which specifies the earnings available for each level of education, S. Further assume that individuals earn nothing whilst in school, and discount the future at a constant rate r. Then in deciding upon the level of education to acquire, individuals will maximise the present discounted value of future earnings:

$$\int_S^\infty g(S)e^{-rt}dt = \frac{g(S)e^{-rS}}{r} \tag{3}$$

As standard in the literature, taking the log of this to be the individual's utility function over (S), having substituted y out of the utility function, gives:

$$U^*(S) = log(g(S)) - rS - log(r) \equiv log(g(S)) - \phi(S) \tag{4}$$

where $\phi(S) = log(r) + rS$. The optimal level of schooling is determined where the marginal benefit of an additional year of schooling is equal to the marginal cost, which is explicit in the first order condition:

$$\frac{g'(S)}{g(S)} = \phi'(S) \Rightarrow \frac{g'(S)}{g(S)} = r \tag{5}$$

If we further assume that $g(S)$ is log-concave then this solution equates the marginal rate of return to schooling with the individual's discount rate.

An individual's discount rate reflects both his/her access to finance to fund current investment in education whilst deferring earnings and also his/her rate of time preference. If individuals differ in their preferences and in their financial resources, this will result in different discount rates and lead to variation in the point at which they stop acquiring education – a higher discount rate resulting in a lower optimal level of education. Therefore schooling level choice may differ amongst individuals of the *same* ability because of differences in individual discount rates (Lang, 1993). The natural question to ask is: what effect will discount rate variation have on the OLS estimates of the return to schooling – does the unobserved discount rate that affects education also affect wages?

Intuition tells us that there is reason to believe that it might. It could be the case that individuals who have a higher discount rate because of their rate of time preference, have more ambition or determination to get into the labour market and earn money. This drive is rewarded in higher wages and also these individuals are more likely to choose career paths with steep wage curves. Consequently a higher discount rate is associated with lower education but also a higher wage controlling for education, thus $\mathrm{E}(u_i\ \epsilon_i) < 0$. In this case the OLS estimation of the return to education is negatively biased. However, it may be that the opposite is true: Munasinghe and Sicherman (2000) present strong evidence from the NLSY[3] that smoking can proxy for rate of time preference, and that after controlling for a rich set of covariates, smokers (high discount rate individuals) experience lower initial wages and lower wage growth than non-smokers (low discount rate), which would suggest that high discount rate individuals are not selecting into steep wage growth occupations. If the wages of high discount rate individuals are lower (conditional on education) and grow more slowly then OLS estimates will be upward biased.

Discount rate and ability are both sources of variation in levels of schooling, moreover these two sources of variation interact in a complex way. Momentarily ignoring the demographic and background characteristics in X that affect schooling, the demand for schooling function is $S{=}S(a,r)$: schooling level choice depends positively on the individual's innate ability (a) and negatively on their discount rate (r). We can invert this function to get innate ability as a function of schooling and the discount rate: $a{=}a(S,r)$. So "...even if the discount rate and innate ability are uncorrelated, they are correlated once we condition on the level of schooling. For a given level of schooling, individuals with higher discount rates will have more innate ability" (Lang, 1993, p10). While a higher discount rate reduces an individual's level of schooling, when we hold that level of schooling constant, those with higher discount rates will have higher ability and this will be rewarded with a higher wage. Recalling the model, this makes sense: we know that if two individuals have chosen the same level of schooling it means that for each, at that point, the marginal return to schooling is equal to their discount rate. Thus the individual with the higher discount rate has a higher

[3]National Longitudinal Survey of Youth, US data

marginal return at that level of education, indicating that they have higher ability. Therefore a higher value of discount rate will reduce schooling, but conditional on schooling in the wage equation, a higher discount rate will mean higher ability and a higher wage: thus $E(u_i \epsilon_i) < 0$. Therefore this potential mechanism through which discount rate affects the joint process of education and earnings again suggests a negative bias in the OLS estimates. If both ability bias and discount rate bias affect the OLS estimate of the return to education but work in opposite directions, then *a priori* we cannot determine what the net bias in the coefficients will be. It is possible that the OLS is higher or lower than the 'true' return to education.

3 Solving the Endogeneity Problem

Over many years, economists have attempted solve the problem of the endogeneity of education in a number of ways. Firstly, a number of studies attempt to control for the effect of ability bias directly by including measures of ability such as IQ and other test scores in the model. However, aside from concerns over whether these types of variables are a good proxy for wage earning ability, Lang (1993) demonstrates that depending on the functional form chosen for the earnings equation, adding 'ability' variables to the model may not necessarily improve their explanatory power and in fact may result in perverse signs for the these variables. The variety of findings in the empirical literature (see Lang, 1993) for the signs and significance of these variables justified Lang's concerns. Moreover this 'ability' variable 'solution' does nothing to counter problem of discount rate bias.

Another approach is to use twins or siblings and exploit differences in their education levels and earnings under the assumption that using twins (especially identical twins) or siblings, eliminates differences in innate ability, and provides an unbiased estimator of the return to education. However, Bound and Solon (1998) argue forcefully that the twins methodology is problematic, highlighting a number of non-trivial issues. Moreover, this strategy also constrains us to the assumption that twins/siblings are identical with respect to discount rates – which unlike ability (which is arguably genetic) is a taste parameter and so this would appear to be an even stronger assumption. A further concern for this approach is that when identification relies on differences in education, there are two points at which measurement error can occur, consequently identifying the return to education through differences in education is likely to be subject to greater measurement error (Harmon and Walker, 1995). Therefore it is far from certain that twin studies can offer a solution and return an unbiased estimate of the return to education.

An alternative strategy which has been the focus of much of the literature, is to identify a variable (or ideally a set of variables) which affect schooling but do not independently enter into the earnings equation and are uncorrelated with the error term in the wage equation. If such variable(s) can be found, then they can be used to construct instrumental variables estimates of the return to education. We will only arrive at a consistent estimator for the return to education if the model is statistically identified. Recalling the model from the start of this section, the moment conditions that we want to impose:

$$\mathrm{E}(X_i \epsilon_i) = 0 \Rightarrow \mathrm{E}(X_i(\ln w_i - X_i' \varphi - \beta S_i)) = 0 \tag{6}$$

$$\mathrm{E}(S_i \epsilon_i) = 0 \Rightarrow \mathrm{E}(S_i(\ln w_i - X_i' \varphi - \beta S_i)) = 0 \tag{7}$$

would be sufficient to identify the model's parameters – providing us with a consistent estimator for β. The corresponding sample moments provide K equations to estimate K parameters, therefore we can estimate $\hat{\varphi}$ and $\hat{\beta}$. However, when we know $\mathrm{E}(S_i\ \epsilon_i) \neq 0$ equation (7) no longer holds and we do not have enough equations to solve for the number of parameters to be estimated. The instrumental variables solution is to use the instrument to derive an additional moment condition that does hold, and replace $\mathrm{E}(S_i\ \epsilon_i) = 0$ and its corresponding sample moment condition with the new condition.

If such an instrument, $\hat{S}_i$, can be found then the violated moment condition can be replaced with $\mathrm{E}(\hat{S}_i(\ln\ w_i$ - $X_i'\varphi$ - $\beta\ S_i)) = 0$. Provided $\hat{S}_i$ is not a linear combination of the $X_i s$ then the corresponding sample moment condition along with the other non-violated moment conditions will be sufficient to identify the parameters $\hat{\varphi}$ and $\hat{\beta}$.

Allowing heterogeneity across individuals in marginal costs of education (due to differences in discount rates) and in marginal returns to education, according to individual characteristics in the vectors X and Z, we can write:

$$\frac{g'(S)}{g(S)} = b_i; \qquad b_i = X_i\gamma_1 + u_{1i} \tag{8}$$

$$\phi'(S) = r = r_i + kS; \qquad k \geq 0; \qquad r_i = X_i\gamma_2 + Z_i\pi + u_{2i} \tag{9}$$

Marginal returns to schooling are constant (within individual), whereas the marginal costs of schooling are increasing in the level of schooling. This is plausible if individuals can finance education initially from family resources, then perhaps from government funding and later only through their own private sources, and if the time and psychic costs of education increase with the level of the qualification/education in question. Equating equations (8) and (9) gives an explicit solution for the optimal level of schooling:

$$S_i^* = \frac{b_i - r_i}{k} = X_i\tilde{\gamma} + Z_i\tilde{\pi} + u_i \tag{10}$$

We can get back to the Mincerian specification of the human capital earnings function (equation (1)) by integrating the marginal benefits of education over the years of education (and here we specify explicitly the heterogeneity in returns across individuals by allowing the β to vary with i):

$$\int_0^{s_i} \frac{g_i'(s)}{g_i(s)} ds = log(w_i) = a_i + b_i s_i = X_i\varphi + S_i\beta_i + \epsilon_i \tag{11}$$

In this model, we can have ability influencing individual earnings both through the individual intercept term a_i (this is the 'unobserved ability' that has been the focus of much of the literature), and through the marginal benefit of an additional year of education captured in b_i, which varies according to the individual's characteristics. Any candidate instrument must be independent of the individual ability intercept term a_i, which means that Z_i must be orthogonal to ϵ_i (and indeed to u_i). The IV estimate – based on 2SLS in which the first stage is estimated by (10) and the second stage is estimated by (11) – of the schooling coefficient β is a weighted average of the marginal returns to education (the β_i) for those whose schooling choice is influenced by the instrument, conditional on X. In order to give this 'local average treatment effect' (LATE) interpretation, there is a monotonicity requirement that all individuals have the same signed response to the instrument i.e. in the case of RoSLA this is that $\tilde{\pi}$ is greater than or equal to zero for all individuals i.e. no-one chooses *less* education as a result of the change in the minimum school leaving age.

There is a large literature in this area in which a number of instruments have been used. Many studies are reviewed in Card (2000). Some studies exploit institutional features or policy changes while others rely on variations in costs across individuals (in each case these instruments alter the marginal cost functions r_i). The latter includes instrumenting using college proximity (for example, Card, 1995), while the former group includes the seminal Angrist and Krueger (1991) paper exploiting differences in schooling owing to the interaction of quarter-of-birth and state variation in when children have to commence compulsory schooling.

While IV has the advantage that we can potentially derive estimates purged of the biases discussed above, it also has some shortcomings. Weak instruments (that is, those that although uncorrelated with wages are hardly correlated with schooling) and invalid instruments (those that although correlated with schooling, may also be correlated with wages) may be worse than no instruments at all – as Bound *et al.* (1993) put it "the cure can be worse than the disease".

A number of authors (Staiger and Stock, 1997, and Bound *et al.*, 1995) have highlighted that many existing instrumental variables studies have been undermined by a lack of precision in their first stage estimates. If the instrument used is only weakly correlated with the endogenous regressor (schooling) then the IV estimates are potentially as biased as the OLS estimates. Bound and Jaeger (1996) show how quarter-of-birth interactions with state and year, used in Angrist and Krueger (1991), form weak instruments that cause IV to be *more* biased than OLS.

Much attention has been given to the weak instruments issue in the econometrics literature of the last 15 years and it is now well established (see for example, Baum *et al.* (2007), Murray (2006a,b)) that two-stage least squares performs very poorly in the presence of weak instruments: not only are point estimates biased, the estimated standard errors of parameters are too small such that confidence intervals are too narrow. Consequently null hypotheses are too readily rejected, and inference can be wildly incorrect.

Further, Bound *et al.* (1995) show that even a small correlation between the instrument and the error term in the wage equation can result in a large bias in the IV estimates even in large samples. This problem is compounded if the instrument is weak, the magnitude of the bias in the IV approaches the bias in the OLS as the R^2 from the first stage regression of the endogenous explanatory variable on the instruments approaches zero.

While this first stage R^2 statistic has previously not been routinely reported, the problem of weak instruments has been quite prevalent since most of the IV studies surveyed in Card (2000) suffer from imprecision and the IV returns are not significantly different to those from OLS. Since the work of Staiger and Stock (1997), Bound *et al.* (1995) and more recently Stock and Yogo (2005), it has become more common to report the first stage R^2 and the F-statistic on the exclusion of the instruments from the first stage, which help to confirm the relevance of a candidate instrument. However, the above named authors have helped to establish that even when an instrument is significant at conventional levels, it may still be weak and lead to the problems of bias and unreliable inference outlined above. As a result, Stock and Yogo (2005) have developed a number of tests for the presence of weak instruments, tabulating critical values depending on whether we use 2SLS, the limited information maximum likelihood (LIML) estimator or Fuller's modified LIML estimator.

Thus it is crucial to establish that there is a strong relationship between the instrument and the endogenous regressor (schooling) i.e. that the instrument is relevant; and that it passes the various tests to establish that it is not a weak instrument.

It is not routinely possible however to test an instrument for correlation with the error term in the wage equation (i.e. test the validity) as to do that we would first need to estimate the wage equation to give us a valid error term which requires a consistent estimator for φ and β, but we can only find a consistent estimator if we have an alternative instrument that we know is valid and strong in the first place. The advantage in having multiple instruments – as I have in this study – is that this allows me to determine the validity of the preferred instrument (early smoking), exploiting the validity of the other instrument available (RoSLA). In addition to this formal econometric test of the instrument's validity, I am also able to provide further supportive evidence for the validity of the early smoking instrument from the reduced forms, from intuition and from the consistency of results estimated with different instruments. As Murray (2006b) points out, every candidate instrument arrives on the scene with "a dark cloud of invalidity overhead" (p. 114). While this cloud can rarely be completely chased away, I believe that there is very strong evidence in favour of the validity of early smoking as an instrument.

An additional problem with the IV strategies is that what they capture is a 'local average treatment effect' (LATE), as outlined above in the formal modelling[4]. The basic problem is that while OLS provides an estimate of the average marginal return to another year of schooling, the IV estimator provides a weighted average marginal return to another year of schooling with the weighting determined by the extent to which individuals' behaviour is changed by the 'treatment' (Angrist and Imbens, 1995). Card (1998) notes that depending on whether the marginal returns to education for individuals in the 'treatment' group are higher or lower than the average marginal return to education, the IV estimator may over- or under-estimate the average marginal return to education for the population as a whole. In these circumstances it is not possible to generalise from the IV estimates to all individuals. Prior to Angrist and Imbens formalisation of LATE reasoning, Lang's (1993) paper – in which the term 'discount rate bias' was first used – criticised Angrist and Krueger (1991)

[4]As the endogenous variable is not binary, technically the IV estimates a 'local average partial effect', see Wooldridge (2002) ch. 18.

on the basis that what they were identifying was in fact a LATE, though Lang termed it 'discount rate bias'. Kling (2000) has demonstrated how Card's 1995 paper using the proximity of a four-year college to instrument for education does indeed capture the return for less advantaged families whose schooling decisions were most effected by the reduced cost associated with a college being nearby. This was Card's intuition in the paper, and Kling has formally shown that Card's estimates do indeed capture a LATE. This is not necessarily a problem, the estimate is not invalid, however it does affect the interpretation. In this case Card captures a LATE which from a policy perspective is an important LATE to know.

I have already outlined the argument that, for a given level of education, those with higher discount rates will have higher ability. Therefore when we take a given level of education – for example the 10 years education that was the minimum prior to the date when the school leaving age in England was raised from 15 to 16 – those with high discount rates will have greater ability than those who choose to leave at 15 because of low returns to education. Thus to the extent that individuals in the low education group have high discount rates because of higher than average costs of education rather than lower than average returns to education, LATE reasoning suggests that IV estimates that isolate this group will find returns that are higher than the average marginal return to education, and may be higher than the OLS estimates (Lang, 1993; Card, 2000).

Alternatively, one could argue that the majority of individuals in this group whose behaviour is affected by the raising of the school leaving age, are low discount rate, low ability and would have located at the minimum prior to the raising of the school leaving age because their return to schooling has already fallen to the same (low) level as their discount rate. In this case, we would expect that the IV estimates of the return to education would be below the average marginal return to an additional year of education. Figure 1 shows the education leaving age density when the minimum school leaving age is 15 compared with when it is 16. It is clear that in the upper ranges the densities are very similar, and that the increase in minimum school leaving age affects only the lower part of the distribution of leaving ages. This concurs with the evidence of Chevalier *et al.* (2004) who use a large sample of data from the General Households Survey (GHS) and find – using a number of tests of the equality of distributions – that RoSLA only affected the attainment of those at the bottom of the schooling distribution, there was not a ripple effect further up. Similarly, Oreopoulos (2006) concludes that the earlier RoSLA (in 1947 raising the minimum age from 14 to 15) only affected the lower part of the distribution, and Harmon and Walker (1995) using both the 1947 and 1973 RoSLA find that only the lower portion of the distribution is affected. Whether these individuals affected by the policy are predominantly high discount rate or predominantly low ability will determine whether we expect the IV estimate from the raising of the school leaving age to be higher or lower than OLS.

Therefore it is important to identify an instrument that avoids these three prominent problems: being correlated with the structural equation error term, being only weakly correlated with the endogenous regressor or capturing a LATE that is not informative when it comes to answering the question we want to ask – what Murray (2006a) terms the bad, the weak and the ugly instruments.

4 Instrumenting Education Using Early Smoking

4.1 Theory

Evans and Montgomery (1994) proposed using whether or not an individual smoked when they were young as an instrument for schooling[5]. The intuition behind the instrument starts from the observation that just as schooling is not randomly assigned across the population, the decision to engage in (un)healthy habits is not randomly distributed. Evans and Montgomery note that "one of the most persistent relationships in health economics is that more educated people have better health and better health habits" (1994, p1). This view is supported by a number of reviews of the empirical evidence on the link between health and education by Grossman (see Grossman, 2005). After extensively reviewing the evidence

[5]This IV strategy has also been pursued by Chevalier and Walker (1999) using GHS and National Child Development Study (NCDS) data, and by Fersterer and Winter-Ebmer (2002) for Austrian data.

Grossman concludes that that completed years of formal schooling is the most important correlate of good health, and this statement applies whether health is being measured by mortality rates, morbidity rates, self-evaluated health status or psychological well being (Grossman, 2000). In the UK, Oreopoulos (2006) uses data from the General Household Survey (GHS) which asks individuals to self-report their health status, and finds that an additional year of schooling increases the chance that an individual will report good health by 6.0% points, and reduces the chance of reporting poor health by 3.2% points. There remains a debate as to whether or not this education-health relationship is causal i.e. through more education people learn the dangers of poor health habits and are thus less likely to engage in them, with Evans and Montgomery citing a quite different explanation for the relationship due to Victor Fuchs (1982). Fuchs argues that unobserved differences in the rate of time preference determine both the number of years schooling that an individual attains and their investments in health, as both decisions involve a trade off between current costs and the discounted value of future benefits.[6]

As with Becker's model of human capital accumulation, in a health accumulation model individuals invest in health until the marginal return to health investment equals their discount rate. If an individual has a higher discount rate because of her rate of time preference, he/she cares less about the future and more about the present and will therefore *ceteris paribus* quit formal education at a younger age and be less likely to invest in good health habits (and be more likely to engage in unhealthy habits). If the correlation between health habits, such as smoking, and education is driven by a common unobserved factor (time-preference) then some health habits could potentially be used as in instrument for education.

Not all health habits can be used as an instrument for two reasons. Firstly, some health habits have consumption as well as investment value. Going to the gym or play-

[6]It is worth noting that the explanations of the health/education correlation as being causal or driven by unobserved time preference are not mutually exclusive: it may be that education promotes better health habits or improves the efficiency of health inputs but individuals may still choose to act differently in light of this education according to their rate of time preference.

ing squash for example, have consumption value and are likely to be correlated with family income/background and possibly correlated with the unobserved component of earnings. Secondly, some health habits such as heavy drinking or drug abuse would be unsuitable as they are likely to have an effect on current wage through their effect on productivity. I follow Evans and Montgomery in arguing that smoking as a teenager is a health habit that can be used as a valid instrument for education.

The decision that an individual makes at age 16 as to whether to continue in education or not is likely to be significantly affected by his/her discount rate – whether that is because of access to financial resources or because of the individual's rate of time preference. In the UK this is the first point at which individuals can choose to leave education, moreover it remains the case that staying in school post-16 and taking A-levels is still the major route into university, therefore the decision to remain at school at 16 is likely to be affected by the individual's discount rate. Moreover, whether an individual chooses to smoke at 16 is also likely to be determined in large part by their rate of time preference. Whether I look at the largest sample of working age men available in the BHPS or my estimation sample it is the case that of the individuals who have ever smoked, approximately 61% were smoking when age 16, and approximately 80% were smoking when age 18[7]. Therefore it is clear that the majority of individuals who ever smoke, first take that decision at around the same time that they are making decisions over the continuation of their education. Evans and Montgomery find that the concurrence in the timing of the smoking and school leaving decisions generates a statistically precise and quantitatively large correlation between years of education and early smoking and, unsurprisingly, the same relation is found in UK data. Thus smoking at 16 satisfies the first criterion for an instrument: it is relevant as it is strongly correlated with completed education. Moreover, as will be illustrated below, the effect of early smoking on years of schooling is sizeable (just under one year less education is completed on average by those who smoke when 16 *ceteris paribus*), therefore the instrument

[7]The precise figures for the estimation sample (largest possible sample) are 60.47% (61.00%) smoking at age 16, 81.11% (79.73%) smoking at age 18.

works through a substantial variation in education (Angrist and Krueger (1991) in particular has been criticized on the basis not only that the correlation between their instrument and education is low – i.e. low t-statistic(s) on the instrument(s) – but also that it induces only a very small variation in education attained, approximately only 0.1 years of education).

In addition to looking at the reduced form for years of schooling – which shows that early smoking has a quantitatively large and statistically significant effect on years of schooling (see Table 5, column 3) – looking at the reduced form for the dependent variable of interest (log hourly wage), supports the argument that early smoking can be used to instrument for education. As pointed out in Murray (2006b), if the candidate instrumental variable does not appear significantly in the reduced form for the structural equation dependent variable, or does but with the 'wrong' sign, then this seriously undermines the case for the instrument. Appendix Table D-1 shows that the smoker-at-16 indicator has a significant coefficient in this second reduced form regression, and is negative as the intuition would tell us: those who smoked when 16 have lower wages that those who did not, with the argument being that this is driven wholly by the difference in average years of schooling between the two groups.

The second criterion is validity: the instrument must not be correlated with wage. As I am using a past health habit, smoking at age 16, to instrument for education in the equation for *current* wage, there should not be a correlation via an income effect: the contemporary wage can have no impact on the disposable income of 16 year old deciding whether or not to smoke. Moreover, theoretically whether one smoked at 16 should have no independent direct effect on *current* wage. It is by no means certain that current smoking affects current wage via a productivity effect, thus a link between smoking at 16 and current wage would be even more speculative. So there is no reason to think that smoking at 16 would affect current wage – and as individuals age and move further away from being 16 this is even more so the case. Moreover, there is a good degree of movement between smoking and non-smoking amongst my sample of men, with 42.0% of men who *did* smoke when they were 16 having stopped by the time they are first observed in the data, and 38.4% of the men who are smokers when first observed in the data *were not* smokers at age 16. In light of these arguments, I believe that smoking at 16 can legitimately be excluded from the wage equation.

However, due to the very nature of the unobservables in the wage equation, it is not possible *a priori* to rule out a correlation between smoking at 16 and the unobservables that do affect wage. If the rate of time preference that characterises early smokers does lead them into higher than average wage jobs (as one part of the discount rate bias story suggests) then this would invalidate the instrument and the estimates derived would continue to be biased. Alternatively, it may be the case that discount rates affects human capital accumulation but once human capital is controlled for in the wage equation, there is no further affect of discount rate on earnings. Whether or not the instrument is valid is an empirical point, and usually it is not possible to formally test for the validity of an instrument.

Fortunately, given I have more than one instrument I have an over identified system and can therefore test the validity of the instruments. In section 10 I test the validity of both instruments and cannot reject the null hypothesis that the instruments are indeed valid. Moreover, I can use the RoSLA instrument to just identify the system and also include early smoking as an explanatory variable and find that it does not have a significant coefficient in the wage equation, which again indicates that it can be excluded from the structural equation. Both of these tests are predicated on the assumption that the RoSLA instrument is valid, which I do not believe is a strong assumption given that the raising of the school leaving age was an exogenous policy change. In addition, in section 10 I discuss the various different robustness tests that I employ when using each instrument separately and when using both together, in line with what is considered current best practice with instrumental variables, in order to make the results and inference robust. In all cases both the qualitative and quantitative nature of the results remains unchanged, and the formal tests support the strength and validity of the instruments.

If we accept that early smoking satisfies these two criteria of relevance (and non-weakness) and validity then an indicator for early smoking can be used as an instrument: it can be the

Z_i in equation (10), influencing schooling through changing the marginal costs of schooling in a way which is uncorrelated with ability.

4.2 Is it a spurious relationship?

This observed relationship between smoking at age 16 and educational attainment could be driven by something other than rate of time preference, something that also affects wages and therefore makes the instrument invalid. It could be argued for example, that poorer socio-economic background lowers education and increases the likelihood of smoking – i.e. smoking at 16 is more a reflection of socio-economic background than discount rate. Clearly socio-economic background may influence the decision to smoke at 16, however, my preferred specification of the model includes variables to control for background characteristics at the time that the individual was a teenager and therefore should take this effect out of the coefficient on the early smoking indicator. If it is the case that smoking at 16 is channelling the effects of such characteristics then adding background characteristics into the schooling demand equation would seriously reduce the impact and significance of the smoker at 16 variable. As it is, the coefficient on smoker at 16 changes only from -1.08 (with a standard error of 0.11) to -0.88 (s.e. 0.11) when we add in the background characteristics. The background characteristics that I am able to include are dummies for the occupational class of each parent when the individual was 14, and a dummy to indicate whether the person lived with both natural parents from birth up until the age of 16. These variables should do a very good job of capturing the individual's background socio-economic circumstances at the time when they are making decisions over education (and whether or not to smoke). Therefore the fact that when they are included in the model, the smoker at 16 indicator still has a quantitatively large effect on schooling and is precisely estimated suggests that it is not socio-economic background that is picked up in the early smoker indicator.

Like Fuchs, in their work on rational addition Becker and Murphy (1988) posit that the decision to smoke reflects discount rate in that it indicates the rate of time preference and this is what I argue – that smoking at 16 reflects rate of time preference. One way in which Fuchs supported his hypothesis was to show that education at age 24 when education levels vary considerably, is as important a predictor of smoking at 17 – when most individuals have the same level of education – as it is a predictor of smoking at 24 (see Farrell and Fuchs, 1982). Using a larger dataset than my actual estimation sample, I implement a probit of current smoking using completed years of schooling amongst the explanatory variables, and repeat the probit for smoking at age 16. The marginal effects estimated at the mean of the explanatory variables suggest that for each additional year of schooling the probability of being a current smoker falls by 2.7% (significant at below the 1% level). In the probit for smoking at 16, it is estimated that each additional year of completed education reduces the probability of having smoked at age 16 by 3.8% (significant at the 1% level, see Table 1). Thus completed education is a significant determinant of early smoking – suggesting that it is not greater education that determines the decision (not) to smoke – education predicts early smoking as well as later smoking, suggesting that another underlying factor (time preference) is determining both.

Moreover, with regard to the question of whether it is a knowledge effect, it is less likely to be the case that the education-smoking link is causal, to the extent that formal schooling is not the main avenue through with knowledge of the detrimental, indeed potentially fatal, health consequences of smoking are disseminated. Since the mid-1960s, the negative effects of smoking on health have been known and increasingly communicated to the public via various awareness campaigns and successive governments have been increasingly direct in their discouragement to smoke both via taxation and the media. As a result, it is decreasingly likely to be the case that only through continued education (past the compulsory level) that individuals are made aware of the negative health effects of smoking. The hypothesis that the relation between education and smoking is in fact driven by the time preference of the individual rather than being a causal or knowledge effect can be tested and this is something that I return to in section 8.

The correlation between smoking and education is also consistent with an alternative hypothesis: that those with lower unobserved ability will acquire less education and are more likely to smoke. I have outlined how ability and discount rate bias interact in a complex fashion thus it is difficult to completely disentangle the different effects. However, if it is the case that we are primarily picking up some measure of ability then we would expect that – by definition – smoking at 16 only affects the education of individuals at the lower end of the ability distribution. If we assume that the residual from the OLS log wage regression is a reasonable proxy for ability, we can divide this residual wage distribution into quintiles and examine whether smoking at 16 is a feature only of low ability (low residual wage) individuals or if it is something that individuals of all abilities engage in.

Table 2 shows the numbers who smoke at age 16 in each quintile of this residual log wage distribution. The left-side panel of the table shows that in the lowest quintile approximately 44% of the males smoked at 16. This figure falls to approximately 39% in the next quintile up and the next after that (30%) before rising again in the fourth quintile (34%). Despite a fall in the last quintile, the figure for the percentage of individuals who smoked at age 16 is still as high as 23% in the highest quintile of the residual log wage distribution. There are fewer smokers at 16 in the higher quintiles of the distribution but that is to be expected, given that smoking at 16 is likely to be in some part be correlated with lower ability. Nevertheless there remain substantial numbers of smokers at 16 in the highest quintiles of the residual log wage distribution which indicate the highest ability individuals. To futher illustrate this point, Figure 2 shows the density of the mean residual log wage for both the smokers and non-smokers at age 16. While the distribution for non-smokers at 16 is slightly to the right of that for smokers at 16, we can see that there is a great deal of common support: there are large numbers of smokers at 16 who have high values of residual log wage.

In addition, Figure 3 plots the density of education leaving age for smokers at 16 and non-smokers at 16. If it was only low educated, low ability individuals who smoke at 16 then we would expect the densities to look very different with very little mass in the upper ranges for the early smokers. However, while the non-smokers at 16 density does have a greater mass around 21 and less around 15/16 suggesting more non-smokers go to university, it is quite close to being a general right-ward shift of the distribution compared with the smokers at 16. This is consistent with the idea that A-levels are the main route into university – we would expect more lower discount rate individuals to remain in school at 16 and the result of this is the lower percentage leaving at 16 and the resulting higher percentage leaving at around 21. Elsewhere the picture is very similar but with the smokers at 16 distribution to the left of the non-smokers. This is consistent with the discount rate hypothesis which says that there are smokers and non-smokers at 16 of all abilities and that smoking at 16 has an effect to reduce education at all points of the ability distribution.

It is certainly true that younger cohorts have consistently acquired more education, and for the men in my sample, smoking at 16 has generally been decreasing: 39.8% of the cohort born in the 1940s smoked when 16, this fell in successive cohorts to 30.0% (those born in the 1950s), 27.8% (60s) before rising again amongst those born in the 1970s, of whom 36.3% smoked when 16. This general pattern would also lead to a shift of the curve to the right for non-smokers at 16, therefore to be sure that it is the case that smokers at 16 do get less education than non-smokers at 16, Figure 4 produces the same plot for the cohorts born in the 40s, 50s, 60s and 70s (which accounts for 88.0% of the men in my sample)[8]. For each cohort the picture broadly follows the pattern of Figure 3: the density for non-smokers is a rightward shift of the smokers at 16 density, illustrating that for all cohorts there are smokers at age 16 across the entire distribution of education levels, but that smokers at 16 acquire less education on average[9].

Therefore in answer to the criteria for a suitable instrument: early smoking is not "bad", there is no reason to suspect that smoker status at 16 should violate the exclusion restriction (and this is something that I test, see section 10, to ensure the instrument is valid); it is

[8]The corresponding graph for individuals born in the 1930s reflects a similar pattern but only accounts for 9.4% of the sample

[9]The cohort born in the 1970s have a restricted education leaving age in that the majority of this cohort are 22 years old or younger, hence their distribution is slightly truncated.

not "weak" as there is a strong, very significant and sizeable *ceteris paribus* effect of early smoking on years of schooling; and it is not "ugly", though it captures a LATE – the group of individuals who have lower education because of a higher than average discount rate – this is a group comprised of individuals of all abilities and is therefore an informative group to consider the return to education for, arguably more representative of the population as a whole than groups identified by other IV estimation strategies.

5 Data

I use the British Household Panel Survey (BHPS) which is a nationally representative survey of the population which began in 1991 and follows the sample individuals each year. In 1999 in addition to the core survey there was a supplementary component in which questions were asked regarding previous health habits. I have constructed an 15-wave pooled-panel dataset containing variables describing individuals' characteristics, a dummy to indicate whether the individual smoked when 16, education, and current hourly wage rate. Since the previous health habits question was only asked in wave 9, I only have observations from individuals present in wave 9, but I have all waves of observations for these individuals. I include males who are in full-time employment (30+hours per week), are not self-employed and are in the age range 19 to 65 inclusive[10].

There are issues of measurement error when using number of years of schooling as the measure of education, however in order to make my results comparable with the majority in the literature I use the observed number of years of schooling as my education variable[11]. The BHPS does not ask how many years education an individual has nor when the individual first left full-time education, rather it asks the age at which the individual left school and age at which he/she left further education. As I construct my years of schooling variable

[10]This age range captures 'prime-age' males and ensures that smoking at 18 is not the same as current smoking for any individuals, as smoking at 18 will be used as an instrument as evidence in support of the rationale behind the early smoking instrument.

[11]Formally: Years-of-schooling = (age left education - 5); thus I assume a school start age of 5, which is the compulsory school start age in the UK.

from age when left school or age when left further education if the individual went on to further education, I encounter problems when people return to full-time education after a number of years away. If an individual completes GCSEs, A-levels, a standard 3-year degree, then a Masters degree and then a PhD (3 years) this would equate to 21 years of education, therefore I exclude any individual with more than 21 years recorded education. This excludes observations from just 84 individuals (3.6% of those with years of schooling calculated)[12].

With respect to earnings, it is standard to use the log of hourly earnings and so again for comparability this is what I have constructed – the log of real wage (using 2006 pounds as the base)[13]. I trim the log wage distribution such that the top and bottom 1% within each year are excluded.

The dataset constructed contains 21,256 observations from 2,266 males with each individual having between 1 and 15 observations; the mean number of observations per individual is 9.38, median 10[14]. Table 3 contains summary statistics for the estimation sample, with the breakdown by early smoking status in Table 4.

6 Estimation

I cannot exploit the panel to eliminate unobserved ability since completed years of education is a fixed effect but I can use the repeated observations to improve precision – although I need then to adjust the standard errors to take account of there being repeated observations of the same individuals at different times[15]. I do this by allowing clustering for each individual

[12]The results are robust to an alternative assumption of recoding such that anyone with education greater than 21 years education is recorded as having 21 years of education.

[13]Current hourly wage is not explicitly recorded, however following other BHPS users (for example Booth and Frank (1999)) I constructed the natural log of hourly wage rate by constructing hourly wage as: $w_i = \text{PAYGU}_i \; / \{4.33(\text{JBHRS}_i + 1.5\text{JBOT}_i)\}$ where PAYGU_i is gross monthly earnings before tax and other deductions in current main job; JBHRS_i is standard weekly hours worked; and JBOT_i is overtime hours worked each week. It is assumed overtime is paid at 1.5 times the normal hourly wage, $4.33 \approx$ no. weeks per month. Therefore w_i = (Monthly Gross Earnings/No. hours worked per month) = Hourly wage rate.

[14]I order to avoid issues around differential attrition, I have re-estimated the models using both inverse probability weighting and also including in the regressions a variable indicating the number of observations that each individual has, and in each case the results remain, available from the author.

[15]As the first stage involves regression of years-of-schooling – which is time-invariant– on characteristics, I re-estimate the model using just one observation (their first) for each member of the sample but then all of the observations in the second stage, bootstrapping to get the correct standard errors in each stage. The

in the variance-covariance matrix which allows for there to be a correlation between the error terms for each individual but no correlation between the error terms of different individuals. The robust standard errors generated do not impose any assumptions on the functional form of the potential correlations and heteroskedasticity controlled for in the error.

I aim to produce estimates that are comparable with other research so I begin by estimating a conventional human capital earnings function where the dependent variable is the natural log of real hourly wage, and the explanatory variables are age, age-squared, and years-of-schooling. I also include controls for ethnicity, for region (using the 13 standard regions) in order to pick up regional effects such as real wage differentials, year-of-birth[16] and its square to pick up cohort effects[17] and dummies for parental characteristics. As discussed, I include parental characteristics because in their absence, the smoking at 16 variable could be picking up background characteristics correlated with education and smoking at 16. The parental characteristics variables that I have are the standard occupational classification of the job of both the individual's father and mother when the individual is 14 years of age, and a dummy to indicate that the individual lived with both natural parents from birth up to the age of 16. Including year dummies in the model would be problematic since I include both age and year-of-birth, however I do include controls for whether it was the early-, mid-, late-1990s or post-2000 to allow for business cycle effects [18]. Mincer's specification of the human capital earnings function, included experience and experience-squared. In the absence of information on labour market experience, Mincer suggested potential experience i.e. age minus schooling minus six (assuming individuals begin schooling aged six), could be used as an approximation. However, using this approximation would mean that mea-

results for the early smoker instrument and for the RoSLA instrument are in the appendix Tables B-1 and B-2 respectively. There is no substantive change in the conclusions. Similarly the models can be estimated on any single wave and the nature of the results does not change, available from the author.

[16]Year-of-birth is rescaled such that 1897=1,..., 1989=93, since in the range 1897-1989 the birth years in my total dataset, year-of-birth and year-of-birth-squared are perfectly collinear.

[17]Including a higher order polynomial in a suitably rescaled year-of-birth does not alter the results nor add to precision in the estimates and so in the interests of parsimony only a quadratic is used.

[18]These dummies are significant in the wage equation, though their inclusion/exclusion does not alter the coefficient on the instrument (1st stage) or $\hat{S}_i$ in the second stage.

surement error in the education variable would necessarily transmit into the experience and experience-squared variables and moreover, the endogeneity of schooling (our main concern) will lead to potential experience and its square being endogenous, resulting in three endogenous regressors. Age and age-squared are the standard candidates to use as instruments for experience and its square, and are widely used as such, therefore this is the approach that I have taken.

I estimate the model first by OLS. I then implement the IV regression using the smoker at 16 indicator as the instrument generating the variation in years-of-schooling.

7 Results

The first column of Table 5 reports the OLS estimate of the human capital earnings function, the second column reports the IV results using smoking at 16 as the instrument. The third column reports the results from the reduced form equation for years of schooling. Looking at the third column of Table 5 we can see that individuals who smoke when they are 16 have on average 0.88 fewer years of schooling than those who do not smoke when they are 16. The robust standard error is 0.108 giving an absolute value of the t-statistic of 8.13. Therefore smoking when 16 is strongly significant for education, and the parameter precisely estimated. This is encouraging given the concerns raised by *inter alia* Staiger and Stock (1997) and Bound *et al.* (1995) concerning the precision of first stage estimates. The R^2 of 0.246 is higher than the R^2 for first stage regressions in some other IV studies[19], and the F-statistic of 66.17 suggests a very strong instrument. The partial-R^2 of the effect of the instrument on years-of-schooling having partialled out the effect of the other covariates is 0.0289 which is high relative to the guidelines given by Bound *et al.* (1995). In terms of formal tests for weak identification, when using 2SLS-IV (as opposed to LIML or Fuller's modified LIML) one of Stock and Yogo's (2005) test statistics can be constructed. The test is based on the Wald test statistic for β: under weak identification, the Wald test rejects

[19]Harmon and Walker (1995) for example have a first stage R^2 of 0.147.

too frequently. The test statistic centres on the rejection rate that the researcher is willing to tolerate if the true rejection rate should be 5%. The test statistic when standard errors are clustered is the Kleibergen-Paap *rk* Wald F-statistic[20]. Critical values relevant when standard errors are clustered have not (at time of writing) been tabulated, however stata's `ivreg2` routine reports the critical values for the *i.i.d.* errors case, which Baum *et al.* (2007) suggest applying though with caution (or alternatively falling back on the original Staiger and Stock (1997) rule-of-thumb that the F-statistic should be 10 or more). If we are willing to accept an actual rejection rate of 10% (the lowest tabulated value) when it should be 5%, the critical value is 16.38: therefore the Kleibergen-Paap *rk* Wald F-statistic of 66.167 that I get, overwhelmingly indicates that there is not a problem of weak identification introducing bias to the coefficient on years of schooling.

Therefore controlling for parental characteristics and year-of-birth, smoking at 16 reduces education by almost 1 year and is precisely estimated. The coefficients on year-of-birth and year-of-birth-squared suggest that from the 1920s onwards, later year of birth is associated with a greater number of years of schooling until the mid-1950s at which point this levels off for a decade before starting to decrease. Turning to the parental occupation dummies, we can see some significant effects on years of schooling[21], particularly for the father's occupational class. As we might expect almost all of the higher occupational strata dummies (the lower numbers) are associated with sizeable positive effect on an individual's education and are precisely estimated. This is particularly true of management (1), professional occupations (2) and associate professional/technical occupations (3), increasing education by 1.1 and 2.3 and 1.5 years respectively. Much fewer of the mother's occupation variables are significant, though a mother in a professional occupation (2) has sizeable positive and significant effect on education (associated with 1.4 years more education). The fact that these parental characteristics dummies are strongly significant in the schooling equation but then

[20] In the special case, as we have here, of a single endogenous regressor, the Kleibergen-Paap *rk* Wald F-statistic reduces to the standard F-statistic on the exclusion of the instruments from the first stage.

[21] The omitted category are plant or machine operatives.

not significant in the IV wage equation suggests that parental characteristics have a strong influence on education controlling for discount rate, but then controlling for education these parental characteristics do not influence wage.

Turning to columns 1 and 2, the OLS estimate suggests that an additional year of schooling increases wage by 4.6% whereas the IV estimate suggests the return is 12.9%. We expect that the IV results will be less precisely estimated than the OLS, and while the robust standard error on years of schooling in the instrumented regression is higher at 0.020 compared to 0.003 in the OLS regression, this still gives a t-statistic of 6.31 and is therefore still precisely estimated and significant at all conventional levels. The dramatic difference in the estimated coefficients suggests that years of schooling is an endogenous variable, and this conclusion is strengthened if I include the residual from the first stage reduced form equation as a regressor in the OLS regression, providing a Hausman test of the endogeneity of schooling. The absolute value of the t-statistic on this residual is 4.78[22].

There is nothing unexpected in the coefficients on the other variables. The dummy for the South-East region is significant in both the OLS and IV wage regressions, and is precisely estimated in each. Since the South East region contains London, it is expected that there will be a positive coefficient on wages given the London weighting. The R^2 for the OLS regression of 0.265 is comparable to other IV studies[23]where it is usually in the range 0.25 to 0.35. Though the R^2 for the instrumented regression is lower at 0.072 the fact that I am using instrumental variables suggests that goodness of fit is not what I am primarily seeking, my main concern is to find a consistent estimator of the causal effect of education on earnings and that is what the instrumented regressions allow me to estimate[24].

Estimation of the IV using the Fuller-LIML estimator rather than standard 2SLS-IV, in order to be as robust as possible to any potential bias in the IV estimates, does not result

[22]Using the endogeneity test built into stata's `ivreg2` routine provides a similarly emphatic confirmation of the endogeneity of years-of-schooling: the null that the variable is exogenous is strongly rejected, the C-test statistic is 22.78 which has a p-value of 0.0000.

[23]Card (1995); Angrist and Krueger (1991); Harmon and Walker (1995).

[24]Moreover, in the context of IV, the reported R^2 has no natural interpretation, and can in fact be negative.

in any substantive change to the estimated coefficients or standard errors: the return to schooling in the IV estimation remains 12.9, st. err. of 0.020 (see Appendix Table C-1 for the results of the Fuller(1) estimation[25]). Moreover, since when the Fuller-LIML estimator is used the finite moments of the IV estimator exist, the Stock-Yogo (2005) test for the maximal relative (to OLS) bias in the IV coefficient can be performed: again the Kleibergen-Paap *rk* Wald F-statistic is compared with the critical values tabluated by Stock and Yogo: as the F-statistic is 66.167 again the null of bias is strongly rejected.

8 Analysis

The results that I find are in line with those found in other studies. Angrist and Krueger (1991) find a return to schooling of 7.0% by OLS rising to 10% by IV (quarter of birth and state interactions). Card (1995) finds an increase in the estimated return to schooling from 7.3% by OLS to 13.2% by IV (college proximity). In studies using UK data, Harmon and Walker have consistently found results similar to my findings: using Family Expenditure Survey (FES) data for 1978-1986 they find estimates of 6.1% by OLS and 15.3% by IV (RoSLA, 1995), using the NCDS[26] they find estimates of 5.0% by OLS and 9.9% by IV (peer effects and education system level effects, Harmon and Walker, 2000), and using the GHS data they find results of 4.9% by OLS rising to 14.0% by IV (RoSLA and educational reforms, Harmon and Walker, 1999). Chevalier and Walker (1999) find using an earlier smaller sample of BHPS men (using just 6 waves) an OLS estimate of 6.4% rising to 20.5% using IV (RoSLA). Chevalier and Walker also construct estimates using smoking status at 16 and NCDS data, estimating a return of 6.1% by OLS, rising to 8.0% by IV; and using GHS data they estimate an OLS return of 6.4% rising to 9.5% when instrumenting using smoking status at 14/16/18.

More recently Oreopoulos (2006) has used the 1947 raising of the school leaving age in

[25]The modified LIML estimator introduced by Fuller, with the Fuller parameter (a) set to 1 is regarded as most robust to any potential weakness of the instrument.

[26]National Child Development Study.

Britain, from 14 to 15[27], and GHS data to compute a standard IV estimate and a regression discontinuity IV estimate of the return to schooling, arguing that as this was a reform which affected around half of the population, the estimated LATE is closer to the average treatment effect (ATE). Oreopoulos estimates the return for British males (aged 32-64) to an additional year of education to be 5.5% by OLS, rising to 9.4% by IV, though the IV estimate is imprecisely estimated. Implementing a regression discontinuity design, Oreopoulos estimates a return of 15.0% for men, though again the estimate is rather imprecise[28].

Therefore my results of 4.6% by OLS rising to 12.9% by IV are of similar magnitude to the studies above, particularly the Harmon and Walker (1995).

8.1 Testing for a spurious relationship

8.1.1 Is it a background effect?

As outlined in section 4.2, it may be argued that the early smoker indicator is picking up differences in background characteristics between those who do and do not smoke at 16, and that these characteristics also affect wage. Hence the need to control as much as possible for socio-economic characteristics of the individuals at the time that they are making their decisions over education and smoking. As a robustness check I also estimate the model without the parental characteristic variables included, see Table 6. As alluded to in 4.2, the effect of removing the background characteristic variables is that the coefficient on the smoker at 16 indicator in the reduced form schooling demand equation increases to -1.08 (robust s.e. 0.113, t =-9.61). The F-statistic on the exclusion of the instrument is 92.39, with the partial-R^2 of the instrument of 0.0400, again both suggesting a strong instrument[29], with a total R^2 for the first stage of 0.143, again comparable with similar studies. The estimated return to education in the wage equation is 12.1% (robust s.e. 0.016, t =7.56). Thus with the nature

[27] Harmon and Walker (1995) exploit both this 1947 raising of the school leaving age, and the later increase from 15 to 16 in 1973 to derive their IV estimates.

[28] When Oreopoulos implements his IV and RD models for *all* individuals – i.e. including females as well as males – the estimated returns are precisely estimated (1% level) and suggest returns of 14.7% (RD) and 15.8% (IV).

[29] Clearly with the F-statistic even higher than before, the Stock-Yogo formal tests of weak identification continue to overwhelmingly suggest a strong instrument.

of the result remaining unchanged, it is clear that the inclusion of parental characteristics is not driving the result.

8.1.2 Is it a knowledge effect?

As discussed in section 4.2 it could be argued that the correlation between health and education is indeed a causal relationship: individuals with more education have more knowledge of the health consequences of certain habits and are less likely to engage in them. However, as outlined above, to the extent that public information campaigns have made the health risks of a particular habit known to the majority of the population, the correlation between education and that habit is more likely to be due to variations in unobserved factors such as discount rates. In the 1940s and 1950s we would expect less of a correlation between smoking and education, because smoking was not known then to be an investment in health choice. However, given the vast amount of information available to the public since the 1960s concerning the risks of smoking, it is fair to deduce that compared to other health habits, the correlation between smoking and schooling is more likely to reflect individuals' investment choices driven by time preference rather than knowledge.

Following Evans and Montgomery (1994), to test this hypothesis, we can postulate that if there has been an increase in the general availability to the public of information on the risks of smoking, then we would expect that decisions to smoke at age 16 taken *after* the effects were widely known are more likely to reflect differences in discount rates, thus the negative correlation between schooling and smoking at 16 should be higher for individuals who reach 16 after the effects of smoking were widely known. If however the link between smoking and education is due to knowledge effects, after the knowledge of the consequences of smoking are widely known, the correlation should disappear. The first Surgeon General's report highlighting the health effects of smoking was published in 1964, therefore I have repeated the estimation and rather than including smoking at 16 alone as an instrument, I interacted this variable with a dummy indicating that the individual turned 16 before the report was published and a dummy indicating that the individual turned 16 after the report was published (i.e. in 1965 or later). If the relationship becomes stronger i.e. if the t-statistic on the smoking term interacted with the turned 16 post-1964 indicator is greater in absolute value than the turned 16 pre-1964 interaction term this would suggest that the relationship is reflecting differences in discount rates.

In Table 7, I report the first stage regression coefficients on these interaction terms when we use these terms rather than just smoking at 16. We can see that both the interaction terms are precisely estimated, significant and that the term for individuals who turned 16 in 1965 or later has a coefficient which is larger in absolute value by 0.1 years of education and has a substantially lower standard error, thus suggesting a stronger relationship post-1964.

A further test of the hypothesis that there is a causal link between education and smoking is to remove individuals who have less than the 11 years of education that the majority of individuals should have by the time that they are 16 and make the decision over whether or not to smoke and whether to continue in education[30]. Re-estimating on this smaller sample produces the results in the Appendix Table A-1. As can be seen, there are no substantive changes to the results in either the first or second stage regressions: smoking at 16 has an almost identical effect on years of education on this sample as it does the full sample.

8.1.3 Is it an ability effect?

Another issue is the question of whether smoking at 16 is just picking up differences in ability. As already discussed, if smoking at 16 was picking up (lack of) ability, we would not expect that smoking at 16 would occur across the whole wage residual distribution as we have seen that it does – significant numbers smoked at 16 in the upper quintiles of the log wage residual distribution. If we continue to use the wage residual distribution as a proxy for ability and, again dividing it into five quintiles, look at the first stage reduced form schooling equations, we can see that the effect of smoking at 16 is actually increasing as we move up the distribution. The left side of Table 8 shows that in the lowest quintile, schooling is reduced by 0.77 years, this is equivalent to a reduction of 6.21% of the mean number of

[30]This removes 527 (23.3%) of the men from the data and 4271 (20.1%) of the observations.

years of education in this group. In the second and third quintiles the reduction in education associated with early smoking is even greater both in absolute terms and relative to mean education in these quintiles. The fourth quintile is affected the least by early smoking but still it is associated with three-quarters of a year less education, and in the highest quintile the estimated reduction is 0.88 years, 6.9% of mean education in this quintile. We can see in the Table 2 that there are significant numbers of individuals who smoke at 16 in all of the quintiles thus these results are not due to small numbers of smokers at 16, and the coefficient on smoking at 16 is significant at the 1% level in all quintiles. Far from only affecting the low ability individuals, this evidence indicates that smoking at 16 has a greater absolute and relative effect on the highest ability individuals. This supports the hypothesis that individuals of all abilities smoke at 16 because of their rate of time preference.

To further pursue the hypothesis that individuals who have lower ability are likely to get less education and are more likely to smoke, I have replicated my results using smoking at age 18 rather than smoking at age 16. Age 18 is the point at which individuals in the UK have to decide whether to remain in education and go to university, and this decision is likely to be affected by their rate of time preference. Moreover, it is more difficult to argue that smokers at 18 are more likely to be lower ability than higher ability individuals. The right panel of Table 2 shows the numbers who smoke at age 18 in the quintiles of the log wage residual distribution. The table illustrates that in the lowest quintile the smokers at 18 out number non-smokers (54% v 46%), and this remains the case in the next quintile up (52% smokers v 48% non). As with smoking at 16, the numbers who did smoke are generally lower as we move up the quintiles yet in the highest quintile, still as much as 35% of the individuals smoked at 18. There are a higher number of individuals who smoked at 18 in the upper quintiles than in the corresponding table for smoking at 16, indeed in each quintile there are more smokers at 18 than there were at 16, at least a 10%-point swing to smokers from non-smokers compared with the age 16 measure. This further supports the idea that teenage smoking is a habit that high discount rate individuals of all abilities engage in.

Using smoking at 18 as the instrument, I obtain the results in Table 9. Looking at the third column, the reduced form equation for schooling, smoking at 18 reduces education by 0.75 years. This is lower than the corresponding reduction associated with smoking at 16 but this is consistent with the time preference story: smokers at 18 have a higher discount rate than *non-smokers* at 18 but *ceteris paribus smokers at 16* will have a higher discount rate than *smokers at 18*. If smokers at 18 have a lower discount rate relative to those who smoke at 16, they will remain in education longer thus we expect that the reduction in education for smoking at 18 is not as much as it is for smoking at 16. The robust standard error on smoking at 18 is 0.108, giving a t-statistic with an absolute value of 6.93, therefore the parameter remains precisely estimated. The first stage regression is very similar to first stage regression when using smoking at 16. The R^2 for this first stage regression is 0.242 so again high relative to other studies' findings and the Kleibergen-Paap *rk* Wald F-statistic of 48.025 again rejects even a hint of weak identification.

Turning to column 2, the estimated return to schooling when we instrument with smoking at 18, is slightly higher at 13.5% than the corresponding figure using smoking at 16 (12.9%), but not by very much. The parameter remains precisely estimated, robust standard error of 0.023 giving a t-statistic of 5.76. Of the other covariates in the model, each has a coefficient and standard error very close to the estimate when I use smoking at 16.

As I get very similar results with smoking at 18 as I do using smoking at 16, and given the distribution of smokers at 16 and 18 throughout the wage distribution, I believe that this is evidence to support the hypothesis that early smoking is picking up the discount rate of the individual rather than being a proxy for ability. Estimates using smoking at 17 rather than 16 or 18 give similar results.

8.1.4 Is it a work effect?

An alternative explanation for the observed relationship between early smoking and lower education, could be that some of the individuals who get a low level of education leave school *before* they are 16 as *non-smokers* and enter work. Then finding themselves in the more adult environment of work rather than school, and perhaps influenced by older colleagues, these low educated men then start to smoke. This reverse causation from low education to smoking at 16 would change the interpretation of the LATE. We would effectively be identifying the return to education for early school leavers who then start to smoke at work – a group much less representative than the discount rate hypothesis would suggest. One way in which to explore this "started smoking at work" hypothesis, is to instrument using smoking status at age 15 rather than 16. Almost the entire sample[31] of men would have been in school when aged 15, even if leaving at the minimum age, therefore if they were a smoker at 15 they will likely have started smoking whilst at school rather than in work. This would suggest that it is something (i.e. discount rate) other than adult work environment which is driving the decision to commence smoking and also the decision to finish school. Table 10 illustrates the results of the IV regression when we use the smoker at 15 indicator as the instrument. The second column shows that the estimated return to education in this new instrumented regression is almost identical to the case when the instrument is smoker at 16 status: the estimate falls to 12.8% from 12.9%. Moreoever, looking at the first stage regression (column 3) we see that smoking at age 15 reduces the average number of years of education by 0.95 years ($t = -7.76$) – which is a greater reduction than we find with the smoker at 16 indicator (0.88 years) and the smoker at 18 indicator (0.75 years), and is highly significant. This is again entirely consistent with the discount rate hypothesis: smokers at 15 have a greater discount rate than non-smokers at 15 ***and*** have a greater discount rate than smokers at 16 (or 18), hence the greater associated reduction in years of education. There are substantial numbers who do smoke at age 15: 334 of the 2266 men in the sample (14.7%), though as would be expected, many fewer than the number who smoke at age 16 (765 out of 2266 men, 33.8%). This evidence therefore adds weight to the discount rate hypothesis, as opposed to the alternative "started smoking at work". Moreover, Table 11 shows the results when using smoking status at 14 as the instrument. Again the instrument is associated with a large

[31]There are 73 out of 2266 men in the sample for whom the minimum leaving age was 14 rather than 15 or 16

reduction in years of schooling (0.91 years, $t = -6.17$) and the estimated return to education is 15.0%[32]. These results again support the discount rate hypothesis, especially considering that all of the 216 men in the sample (9.5%) who did smoke at 14 faced a minimum school leaving age of at least 15, which completely rules out the proposed alternative explanation for the smoking/education correlation. Though it is noted that the numbers who smoke when 14 are lower than for the other ages, taken with the results for smokers at 15, 16 and 18, these results add weight to the discount rate hypothesis.

8.2 Testing for the discount rate hypothesis

One final test of whether early smoking is picking up differences in time preference is to test whether early smoking is correlated with other future oriented behaviours such as saving, investing and taking precautionary health measures. Home-ownership is one such measure of future orientated behaviour, and Table 12 presents a probit of home-ownership in which the explanatory variables are those included in the wage equation (bar years-of-schooling)[33], plus log wage itself and the early smoking indicator. The marginal effects estimated at the means of the explanatory variables suggest that smoking at 16 is associated with a 4.4% reduction in the probability of being a home owner, and is significant at the 1% level. Thus, controlling for human capital and other background characterisitcs to capture heterogeneity, early smoking is associated with a significantly lower probability of being a homeowner, supporting the idea that early smoking is revealing something of the individual's discount rate.

There is an obvious problem in looking at health measures when early smoking is an explanatory variable in that there may be direct consequences of the early smoking on the health outcome, hence the need to look at health related behaviours rather than outcomes. Table 13 contains the results of probit regression of having a dental check up in the past year,

[32]As with smoker at 16 or 18, both the smoker at 15 and smoker at 14 instruments are strong using the Stock-Yogo criteria.

[33]I exclude years-of-schooling, including log wage instead, if years-of-schooling is included it is not significant and alters the smoking coefficient very slightly.

and having an eye check in the past year, using the same explanatory variables as in the home-ownership probit. Having regular dental and eye check-ups involve trading off future benefits (preventing ill health and associated costs) for current costs (time and expense of appointments) and thus should be influenced by the individual's rate of time preference. As can be seen in these tables, controlling for characteristics and log wage, individuals who were early smokers are 4.0% less likely to have had a dental check up and 2.9% less likely to have had an opticians check up in the past year, each significant at the 1% level. Though these are not perfect indicator measures, with potential problems in each case, they do add to the evidence that the early smoking-education link is capturing the effect of the individual's rate of time preference.

Given all of the tests I have conducted, I am satisfied that smoking at age 16 is a valid instrument for education, and conclude therefore that the OLS estimates are underestimating the return to education. I am not claiming to recover the 'true' return to education and the underlying schooling demand equation. What I have done is estimate the return to education, negating the discount rate bias present in OLS by using smoking at 16 in the schooling equation to generate some variation in schooling which is uncorrelated with the wage equation error term – something that the dual instruments allow me to test (more in section (10)). Moreover, I am removing the ability bias that is present in OLS estimates, as the instrument is uncorrelated with ability – individuals of all abilities can have a high discount rate because of their rate of time preference. Therefore I am confident that the instrumental variables estimation has removed the bias from the OLS, allowing a consistent estimate of the return to education.

My estimate ***is*** a local average treatment effect. However, I argue that smoking at 16 demonstrates that the individual has a high discount rate because of their rate of time preference. Thus when I estimate the return to education using smoking at 16 as an instrument, what I am recovering is the average marginal return to education for the group of individuals who have high discount rates not because they have poor access to finance, but because they have a rate of time preference that reflects that they favour the present.

The natural 'local average treatment effect' question is whether I should expect the average marginal return to education to be higher or lower for individuals in this group than the average marginal return to education in the population as a whole? Since individuals of all abilities have rates of time preference that are reflected in a high discount rate, and we have seen that smoking at 16 affects all across the (log wage residual proxying for) ability distribution, we do not have the 'problem' that estimates using compulsory schooling laws are subject to: that they identify returns for individuals with low education and who are (arguably) disproportionately of low ability. If ability is distributed amongst the early smokers group in the same way that it is amongst the population then these early smoker IV estimates are more appropriate for making inferences about the return to education in the population as a whole than similar estimates from IV studies which isolate minimum age school leavers. However that is not to say that estimates derived from the raising of the school leaving age are unsound – only that they are less useful in drawing inference on the average marginal return to education in the population as a whole. What the RoSLA estimates do provide is an estimate of the return to education for those individuals who wanted to leave full-time education at the minimum age – and from a policy point of view this is an important parameter, especially as the Government has recently raised the education leaving age to 17 (from 2013) and it is later to be raised to 18 (by 2015).

The return that I recover is purged of the effects of ability bias and discount rate bias. Both Card (1994, 1998) and Lang (1993) conclude from looking at the broad literature on the effect of ability bias, that ability bias if it is present has only a small biasing effect, Lang suggesting that discount rate bias dominates such that OLS estimates are biased substantially downwards and Card similarly concludes that the OLS are at least 10-to-30% biased downwards. For the body of UK estimates detailed earlier, the IV estimate is between 1.7 and 3.2 times (average 2.5) the OLS. My early smoker IV evidence is consistent with these results – estimating the return to education controlling for ability bias and discount rate bias, I get an estimate that is 2.8 times the OLS estimate. Furthermore, if we believe that ability has the same distribution amongst the high discount rate group as it is in the population as a whole, it is more valid to generalise to the population as a whole than perhaps is the case with using estimates recovered from instrumental variables that affect only the low educated.

9 Instrumenting Using the Raising of the School Leaving Age (RoSLA)

Now to pursue this line of enquiry further, I will compare the estimate using the early smoking instrument with an IV estimate derived using the raising of the minimum school leaving age. The school leaving age was raised in England and Wales from 15 to 16 in 1973 such that if an individual was 16 by the end of August 1973 he/she was allowed to leave school in the June of 1973, while if the individual was only 15 at the end of August 1973 he/she would have to remain another year at school. This means that those born after August 1957, face a minimum school leaving age of 16. In Scotland this reform took place in August 1976 therefore individuals born after August 1960 face a minimum school leaving age of 16.

This information, plus an individual's date of birth and country of residence, allows the alternative IV estimate to be constructed. Rather than including the smoker at 16 indicator in the first stage regression, I include a dummy to indicate whether the individual faced the minimum school leaving age of 16[34]. As I am controlling for a quadratic in year-of-birth, the smooth changes in schooling as a result of younger cohorts generally gaining more education is controlled for, while the identification derives from the discontinuity induced by the RoSLA. Figure 5 shows the proportion of individuals who have left school at or before age 15, by year of birth, for the majority of men in my sample[35]. As the figure shows, there

[34]The minimum school leaving age was raised from 14 to 15, in 1947 for England and Wales, 1946 for Scotland, however, in the sample of men that I use, there are only 73 individuals (3.22%) who face a minimum school leaving age of 14 so I have concentrated on the later change to create an instrument.

[35]I have trimmed the sample to remove the small number of men born before 1931 and after 1970 due to

is a steady decline in the proportion of men who have left education at 15 or before, and though the relatively small number of men born in any single year in my data means that it is slightly volatile[36], the pattern of steady decline is evident. In year-of-birth 1958, when the policy is in effect for all individuals, we can see that there is a drop from 17.4% to 1.9% of men leaving at or before 15. The figure remains low for the years thereafter, though with some volatility remaining. Contrasting this is the upper line on the graph which shows the proportion of individuals who have left at age 16 or earlier. While similarly showing a decline as younger cohorts gain more education, the proportion who have left by or at 16 continues to show volatility after the RoSLA, rising and falling quite sharply in places. So while the small numbers of men born in any particular year leads to volatility in each graph, it is evident that the RoSLA results in a discontinuity at the point in which is was implemented, and it is from this discontinuity that I am able to construct the IV estimates using RoSLA. This is a well established instrument, and the reduced form for log wage, including an indicator for 16 being the minimum school leaving age faced by the individual, shows that the raising of the school leaving age is associated with a statistically significant increase in log wage, see Appendix Table D-1.

Table 14 contains the results for the RoSLA IV along with the OLS estimates (from Table 5). Column 1 contains the OLS results, column 2 is the result from the IV using RoSLA, while column 3 contains the first stage regression result using the raising of the school leaving age as the instrument.

The main columns of interest are columns 2 and 3. Looking first at column 3, the raising of the school leaving age is associated with an increase in education of 0.564 years and the coefficient is precisely estimated with a robust standard error of 0.206 giving a t-statistic of 2.74. Again, it is noticeable that the R^2 (0.227) is higher than has been found in similar

the small cell sizes, the graph contains the information for 83.9% of the English men in the sample. I have excluded the small number of Scottish men for the purpose of this illustration as the RoSLA occured later for Scotland.

[36]As year-of-birth increases the cell sizes increase and for the years relevant to the RoSLA the numbers are larger.

studies. The partial-R^2 for the instrument in the first stage is 0.0044 which is smaller than for the early smoker instrument but is exactly the same as that found by Harmon and Walker (1995) for their first stage, and compares well with Bound *et al.* (1995). The F-statistic on the exclusion of the instrument from the first stage is 7.49. While this is below Staiger and Stock's (1997) rule-of-thumb guide of 10, taken with the partial R^2, the overall picture is not of a weak instrument. Moreover, using the Fuller(1) estimator – which is the most robust to the presence of a potentially weak instrument introducing bias to the coefficient on the endogenous variable – the result is almost identical (see Appendix Table C-2). The size of the average increase in education, controlling for other covariates in the first stage, is comparable with that found by Harmon and Walker (1995) (0.54 years for the 1947 RoSLA), and slightly larger than that found by Oreopoulos (2006)(0.44 years for the 1947 RoSLA).

Turning to column 2, we see that the estimated return to schooling is 10.2% when we instrument using RoSLA. This is more than double the size of the OLS return though below the other IV estimate. However it is not as precisely estimated, the robust standard error is 0.051 giving a t-statistic of 1.99, the p-value of this t-statistic is 0.046 thus it is significant at the 5% level.

Again, as a robustness check to verify that the inclusion of the parental characteristics variables are not driving the result, Table 15 displays the results for the more basic specification excluding these background variables. In this more basic specification, the instrument is actually strengthened, the F-statistic on the exclusion of the instrument from the first stage increasing to 9.98 (much closer to Staiger and Stock's rule-of-thumb of 10) and the partial R^2 of the instrument is 0.0058 (increased from 0.0044 in the main specification), and the overall first stage R^2 is 0.113. The effect on the estimated return to education is minor – reducing from 10.2% to 10.0%, with a robust standard error of 0.042 giving a t-statistic of 2.41, making the estimate signficant at the 5% level (p-value 0.016). Thus again the inclusion of parental characteristic variables is not driving the result. More importantly, in this specification the instrument is almost exactly attaining Staiger and Stock's threshold for a non-weak instrument and the estimated coefficient on years of schooling is almost identical to the main specification case, when the F-statistic was only 7.49. This suggests that there is no bias in the estimated coefficient on years of schooling in the main specification.

The question is whether this is evidence that using an institutional change – such as the raising of the school leaving age – to form an instrument isolates the return to schooling for only a specific group that is heavily weighted towards the low ability or those with high discount rate particularly because of financial constraints?

If the group whose return is identified by the RoSLA instrument (which is by definition a low education group) is comprised mainly of individuals of low ability rather than those who have high discount rates because of poor access to finance, then we would expect that the return for this group would be lower than the return we find with the smoker at 16 instrument – as I have demonstrated that individuals of all abilities are in the early smokers group. The imprecision of the estimate using RoSLA does not allow me to conclude that the estimate is definitely smaller than the smoking at 16 IV estimate, however one test of the extent to which RoSLA affects individuals of different abilities is to repeat the first stage regressions by quintile of the log wage residual distribution that I used to illustrate the effect of smoking at 16 on educational attainment in all quintiles of the distribution. The results from these regressions are in right hand section of Table 8. If the contention is that RoSLA affects primarily low ability individuals then we would expect that the effect would be quantitatively larger for the lowest quintiles of the log wage residual distribution but falling in size and significance as we move up the distribution.

Table 8 illustrates that the raising of the minimum school leaving age increases the number of years of schooling by 1.04 years in the lowest quintile, which is 8.4% of the mean number of years schooling for this group. Being almost exactly 1 year extra education this suggests that in this lower quintile of the (proxy) ability distribution, all the individuals wished to leave school at the minimum age. In the second lowest quintile RoSLA increases the number of years of schooling by 0.84 years which is 6.9% of the mean for this group.

In the three quintiles above this the increase in education associated with RoSLA is much smaller in absolute and relative terms than in both of the lowest two quintiles but in none of these higher quintiles is the dummy for minimum school leaving age of 16 close to being statistically significant.

This evidence is consistent with the hypothesis that the low education group affected by RoSLA are generally lower ability – if they were mainly high discount rate then we would expect to see a similar effect across the log wage residual distribution.

The contention that the RoSLA group is weighted more towards low ability rather than high discount rate individuals is supported by Carneiro and Heckman (2002). They find that in the US, only 8% of American youths are credit constrained to the point that it affects their post-secondary schooling. Moreover, they find that when ability is controlled for responses to tuition costs are uniform across income groups. Low family income at the time when decisions over post-secondary education are made does not appear to be a major constraint in the US. Two recent studies in the UK have indicated that credit constraints do not prevent individuals from participating in higher education. Chowdry *et al.* (2008) use a unique dataset from a cohort comprising all state school pupils who were in the final year of compulsory schooling in England in 2001-2002. These students have been followed from age 11 through to their higher education participation decision at age 18 (in 2004-05) or age 19 (2005-06). The results indicate that conditional on prior attainment, there is no difference in higher (university) education participation rates between children of higher and lower socio-economic status (SES) – illustrating for the UK, what Carneiro and Heckman find for the US. Similarly, Dearden *et al.* (2008) study the effect of alterations to the funding of higher education in England – with the introduction of fees and indeed top-up fees. They find that participation rates among the lower SES groups have not declined following the introduction of tuition fees (due to the provision of loans by the government to pay the fees), which again supports the contention that the RoSLA group in this country are not credit constrained.

If it was the case that those affected by RoSLA are high discount rate rather than low ability, the IV results which use RoSLA could well be higher than the OLS estimates. However, the evidence above and these conclusions from the Carneiro and Heckman, Chowdry *et al.* and Dearden *et al.* papers suggest that it is more likely to be the case that the group identified by RoSLA are individuals of low ability rather than high discount rate. Though the imprecision of the RoSLA IV estimate prevents a concrete conclusion that it is indeed lower, comparing the RoSLA IV result with the early smoking IV estimate suggests that the RoSLA group are lower ability as the RoSLA IV estimates a lower return. This, and the results from looking at where in the proxy ability distribution each instrument is working, supports the contention that it is more appropriate to generalise from the early smoking IV estimate to the rest of the population: as unlike RoSLA, the estimate is not capturing a LATE that is primarily a lower ability group.

10 Testing of the Instruments

Having more than one instrument means that I have an over-identified system – more moment conditions than are necessary to identify the parameters of the model – which means that I can test the instruments to establish whether the exclusion restrictions are valid. In other studies, such as Angrist and Krueger (1991) and Evans and Montgomery (1994), multiple instruments are used and tested. In each of these cases however, they essentially only have one mechanism to generate the exogenous variation in education: including interactions of that mechanism (the instrument) with other variables does not entail genuinely having multiple instruments. If the mechanism is not valid then none of the 'instruments' are valid, the problem being that the Hansen J-test of the exclusion restrictions involves assuming one of the instruments is valid in order to test the others.

On the contrary, I have two independent sources of exogenous variation in education and so can genuinely test the validity of the exclusion restrictions. As Murray (2006a,b) points out, the Hansen test is more compelling when one of the instruments is thought to be definitely valid, and I believe that I am in this situation: there is a strong argument to suggest that the RoSLA instrument is valid as it was an exogenous (to the individual) policy change.

Instrumenting using both the early smoking instrument and the minimum school leaving age instrument and then performing the Hansen J-test results in a test statistic of 0.202, p-value 0.6529, which is a comprehensive failure to reject the null hypothesis that the instruments are valid[37]. The first stage R^2 is high at 0.250 and the F-statistic on the exclusion of the instruments is 36.83 with a partial R^2 on the instruments of 0.0332, all of which suggests that the instruments are strong as well as valid. The Kleibergen-Paap *rk* Wald F-statistic indicates that the Stock-Yogo tests of weak identification are easily passed (i.e. no weak instrument problem)[38]. Furthermore, the high F and R^2 statistics suggest that the bias inherent in IV estimation in finite samples will be smaller than the OLS bias[39]. Using the Fuller(1) LIML estimator, the results are almost identical (see Table C-3), and again all weak instrument tests are comprehensively passed.

The Hansen J test provides compelling statistical evidence for the validity of the early smoking instrument, which earlier evidence has shown to be a strong instrument. Furthermore, in order to re-enforce the evidence of the Hansen test, it can be decomposed to illustrate directly the validity of the early smoker instrument specifically: by using the RoSLA instrument to just identify the system of equations and then taking these valid estimates of the error from the structural equation and regressing them on the early smoker instrument. The results of such an exercise are contained in Appendix Table D-2. As can be seen, there is no relationship between the residuals from the structural equation and the early smoker

[37]Moreover it is well known that the Hansen test rejects too often i.e. it rejects the null that the instruments are valid in cases where it should not, thus such a strong failure to reject suggests we are far from the rejection region, re-enforcing the validity of the instruments.

[38]Though the correct critical values for this test are not tabulated in the case where standard errors are clustered, using the critical values for the *i.i.d.* case or the Staiger and Stock rule-of-thumb indicates strong instruments.

[39]The ratio of the finite sample biases of 2SLS and OLS is $\approx \frac{l}{nR_1^2}$ where l is the number of instruments and R_1^2 is the R^2 from the first stage of the 2SLS (see Murray, 2006b). In my estimation l =2, n =21256 and R^2 =0.250, such that the 2SLS finite sample bias is a fraction of the OLS finite sample bias.

indicator[40].

Finally, an alternative IV regression can be run in which RoSLA is used as the identifying instrument, while the early smoker indicator is included as one of the X variables. Appendix Table D-3 shows the coefficient estimates when this exercise is carried out. As can be seen, while early smoking affects education in the first stage (with a coefficient almost identical to the other specifications in which it is used as an instrument) it is completely insignificant in the structural equation. This evidence supports the contention that early smoking affects choice of education, conditional on the other variables in X, but then has no further independent effect on log wage. All of these results suggest that early smoking is both a strong ***and*** valid instrument. The evidence indicates that discount rate, as captured by early smoking, affects human capital accumulation, however once that has been controlled for in the structural equation, there is no remaining effect of discount rate on wage.

Returning to the estimation results when using both RoSLA and early smoking as instruments, Table 16 shows that the coefficient on each instrument in the first stage is almost identical to the case when the instruments are used separately, and the estimated return to education using both instruments together is 12.5% with a robust standard error of 0.019 giving a t-statistic of 6.66. The standard error is lower than is the case when either of the instruments are used singly, so the extra variation in schooling that comes with using both instruments results in a more precise estimate of the IV return to education, as we would expect.

The problem with this strategy is that using both instruments makes the interpretation 'ugly', to borrow Murray's parlance. Though I am exploiting two sources of exogenous variation in years-of-schooling, which is good for identification, the problem is interpreting exactly whose return the resulting LATE estimator is capturing. It is not as straightforward as in the individual instruments cases in which we identify the low ability/high discount rate

[40]It is worth noting that strong instruments that are 'almost valid' bias 2SLS estimates only a little, thus even if there was any remaining doubt regarding even a small correlation between the early smoker instrument and the structural equation error term, the overwhelming strength of the instrument would suggest any bias would be very small, see Murray (2006b).

individuals' return – using RoSLA – or the high discount rate (because of time preference) individuals' return – using early smoking. Given that the effects of each instrument in the first stage are similar to their impacts when used separately, and that the early smoking instrument is the stronger and the resulting IV estimate of the return is very close to the early smoking IV estimate, it appears that this instrument is doing most of the work. In interpretation this would suggest the estimate is more heavily weighted towards the return for the individuals who have high discount rates because of their rate of time preference.

11 Conclusions

I have presented three IV estimates: the RoSLA estimate of 10.2%, the combined estimate of 12.5%, and the early smoking estimate of 12.9%, all of which whilst being statistically significant are sufficiently imprecise for me to be unable to conclude are actually different from each other. My analysis, looking at the effects on different quintiles of the proxy ability distribution, suggests that the RoSLA estimate captures the return for the individuals who wanted to leave at the minimum leaving age but were forced to stay longer – concurring with the earlier evidence of Oreopoulos, Chevalier *et al.* and Harmon and Walker. I have argued that early smoking is a behaviour engaged in by individuals of all abilities who have high discount rates due to their rate of time preference, thus the IV estimate derived from this instrument is closer to an average marginal return to education, purged of the bias of OLS. Importantly, exploiting the over-identification, I have demonstrated that using early smoking behaviour allows the construction of a **valid** instrumental variables estimate of the return to education.

That both the RoSLA and early smoking IV estimates are not statistically different to each other suggests that the RoSLA LATE is also close to an average marginal return to education i.e. that the returns at the lower part of the distribution are similar to the average return. This follows Oreopoulos who finds a return substantially higher than the estimated OLS return when implementing IV esitmates based on RoSLA, and a RoSLA that affected a large proportion of the population.

This leaves a question of *why* we get a similar estimated return for the RoSLA and early smoking groups, despite the fact that the groups have differing distributions of ability and levels of education i.e. they are capturing different LATEs. I believe that the results that this and other IV studies find can be reconciled when we consider the assumptions imposed by Mincer's human capital earnings function as I (and others) have estimated it. Implicit in this specification is the assumption that each additional year of schooling has the same proportional effect on earnings i.e. concavity in the schooling-wage profile is not modelled. Moreover, in interpreting IV estimates we need explicitly recognise that returns to education vary across the population depending on individual characteristics (the β_i vary). If different individuals have different returns to schooling at the same level of schooling and if each individual's return to schooling is strictly decreasing in their level of schooling, then there is no **unique** causal effect of schooling.

While some authors[41] have concentrated on "sheep-skin" effects creating non-linearities in the returns to education, Lang (1993) finds a diminishing marginal product of education i.e. concavity in the education-wage profile. The individuals affected by RoSLA may be of lower ability, however, if all individuals have a higher marginal return to schooling at lower levels of schooling then this is consistent with the estimate from the RoSLA IV being higher than the OLS estimate. Similarly, though the smoking at 16 group have all levels of education, some higher than the minimum that the RoSLA individuals have by definition, there is more weight in the lower part of the schooling distribution among early smokers and so the the average marginal return across these individuals will be weighted towards the RoSLA estimate. Thus in this light it is perhaps unsurprising that both the smoking instrument and the RoSLA instrument result in estimates of the return to education that are similar to each other.

More generally there is the question of why the OLS estimates are consistently found to

[41]For example, Park (1999) has looked at "sheep-skin" effects in the US.

be below IV estimates – irrespective of the instrument chosen – when, as noted above, measurement error in standard micro surveys could only sensibly account for a relatively small attenuation in the OLS coefficient and moreover it appears from this study that 'discount rate bias' is not a major factor biasing the OLS estimates downwards. The 'discount rate bias' story suggests that the effect of discount rate to reduce education also independently increases wages. However, when I test for the correlation between the discount rate (as captured by early smoking) and the wage error the instrument is shown to be valid. Hence I do not believe that 'discount rate bias' is the major factor biasing the OLS estimates downwards. Given that all instruments estimate a 'local average treatment effect', which may or may not be different to the average effect on the treated, it appears that the instruments that have commonly been used – and the two that I use here – isolate the treatment effect for groups of individuals who are located at point(s) in the education distribution at which there is a higher average return to education than the global average estimated by OLS. Support for this conclusion also comes from Oreopoulos (2006) who estimates that when the OLS is carried out only for those who left school at 16 or less, the estimated coefficient is similar to his IV estimates which use RoSLA. If I replicate this approach and estimate the OLS regression only for those who left school at the minimum age the estimated return is 19.7%. Whilst acknowledging that the endogeneity of years of schooling in this regression is not dealt with, the much greater coefficient on years of schooling does suggest that the linearity in returns assumption of the OLS when estimated over the entire range of education levels contributes significantly to the lowering of the OLS coefficient.

One conclusion is that in modelling the returns to education, while the endogeneity of schooling is clearly a problem, it is important to recognize that there are also issues regarding the appropriateness of the linearity assumption and the reality of heterogeneous returns to education across individuals. Thus for policy purposes in particular, it may not even be appropriate to refer to ***the*** causal effect of education on earnings. In answering the question of the return, we may need to focus on the individuals in question and the margin in question before we can arrive at a valid answer.

References

[1] Angrist, J. and G. Imbens (1995) “Two-stage least squares estimation of average causal effects in models with variable treatment intensity” *Journal of the American Statistical Association*, vol. 90, pp. 431-442.

[2] Angrist, J. and A. Krueger (1991) “Does compulsory schooling attendance affect schooling and earnings?” *Quarterly Journal of Economics*, vol.106, no. 4, pp. 979-1014.

[3] Ashenfelter, O. and A. Krueger (1994) “Estimates of the Economic Return to Schooling for a New Sample of Twins” *American Economic Review*, vol. 84, no. 5, pp. 1157-1173.

[4] Baum, C., M. Schaffer and S. Stillman (2007) “Enhanced Routines for Instrumental Variables/GMM Estimation and Testing” *Boston College Economics Working Paper* no. 667.

[5] Becker, G. (1964) *Human Capital: A Theoretical & Empirical Analysis with special reference to Education*. New York: National Bureau of Economic Research.

[6] Becker, G. (1967) *Human Capital and the Personal Distribution of Income*. Ann Arbor: University of Michigan Press.

[7] Becker, G. and K. Murphy (1988) “A Theory of Rational Addiction” *Journal of Political Economy*, vol. 96, no. 4, pp. 675-700.

[8] Booth, A. and J. Frank (1999) “Earnings, Productivity and PRP” *Journal of Economic Literature*, vol. 17, no.3, pp. 447-463.

[9] Bound, J., D. Jaeger and R. Baker (1993) “The Cure Can Be Worse Than The Disease: A Cautionary Tale Regarding Instrumental Variables” *National Bureau of Economic Research Technical Working Paper* no. 0137.

[10] Bound, J., D. Jaeger and R. Baker (1995) "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak" *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 443-450.

[11] Bound, J. and D. Jaeger (1996) "On the validity of season of birth as an instrument in wage equations: a comment on Angrist and Kreuger's 'Does compulsory school attendance affect schooling and earnings?"' *National Bureau of Economic Research Working Paper* no. 5835

[12] Bound, J. and G. Solon (1998) "Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling" *National Bureau of Economic Research Working Paper* no. W6721

[13] Card, D (1994) "Earnings, schooling and ability revisited" *National Bureau of Economic Research Working Paper* no. 4832.

[14] Card, D (1995) "Using geographic variation in college proximity to estimate the return to schooling" in Christofides, L., E. Grant and R. Swidinsky (eds), *Aspects of Labour Market Behaviour: essays in honour of John Vanderkamp*, Toronto: University of Toronto Press, pp. 201-222.

[15] Card, D (1998) "The causal effect of education on earnings" in Ashenfleter, O. and D. Card (eds) *Handbook of Labor Economics*, vol. 3A, Amsterdam: Elsevier, pp. 1801-1863.

[16] Card, D (2000), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems" *National Bureau of Economic Research Working Paper* no. 7769.

[17] Carneiro, P. and J. Heckman (2002) "The Evidence on Credit Constraints in Post-secondary Schooling" *Economic Journal*, vol. 112, no. 482, pp. 705-734.

[18] Chevalier, A., C. Harmon, I. Walker and Y. Zhu (2004) "Does Education Raise Productivity, or Just Reflect It?" *Economic Journal*, vol. 114, F499-F517.

[19] Chevalier, A. and I. Walker (1999) "Further Results on The Returns to Education in the UK" *Mimeo*, London: Royal Holloway, University of London.

[20] Chowdry, H., C. Crawford, L. Dearden, A. Goodman and A. Vignoles (2008) "Widening Participation in Higher Education: analysis using linked administrative data" *Institute for Fiscal Studies Report* no. R69.

[21] Dearden, L., E. Fitzsimons and G. Wyness (2009) "Estimating the Impact of Up-front Fees and Student Support on University Participation" *Mimeo*, London: Institute of Education.

[22] Evans, W. and E. Montgomery (1994) "Education and Health: Where there's smoke there's an instrument" *National Bureau of Economic Research Working Paper* no. 4949.

[23] Farrell, P. and V. Fuchs (1982) "Schooling and Health: the Cigarette Connection" *Journal of Health Economics*, vol. 1, pp. 217-30.

[24] Fersterer, J. and R. Winter-Ebmer (2000) "Smoking, Discount Rates and the Return to Education" *IZA Discussion Paper* no. 126.

[25] Fuchs, V. (1982) "Time Preferences and Health: An Exploratory Study" in Fuchs, V. (ed) *Economic Aspects of Health*, Chicago: University of Chicago Press.

[26] Griliches, Z. (1977) "Estimating the returns to schooling: some econometric problems" *Econometrica*, vol. 45, no. 1, pp. 1-22.

[27] Grossman, M. (2005) "Education and Nonmarket Outcomes" *National Bureau of Economic Research Working Paper* no. 11582.

[28] Harmon, C. and I. Walker (1995) "Estimates of the Economic Return to Schooling for the UK" *American Economic Review*, vol. 85, no. 5, pp. 1278-1286.

[29] Harmon, C. and I. Walker (1999) "The Marginal and Average Return to Schooling in the UK" *European Economic Review*, vol. 43, no. 4-6, pp. 879-887.

[30] Harmon, C. and I. Walker (2000) "The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales" *Economica*, vol. 67, no. 265., pp. 19-35.

[31] Kling, J. (2000) "Interpreting Instrumental Variables Estimates of the Returns to Schooling" *National Bureau of Economic Research Working Paper* no. 7989.

[32] Lang, K. (1993) "Ability bias, discount rate bias and the return to education" *Mimeo*, Boston, MA: Boston University.

[33] Mincer, J. (1974) *Schooling, Experience and Earnings*, New York: Columbia University Press.

[34] Munasinghe, L. and N. Sicherman (2000) "Why Do Dancers Smoke? Time Preference, Occupational Choice and Wage Growth" *National Bureau of Economic Research Working Paper* no. 7542.

[35] Murray, M. (2006a) "The Bad, the Weak and the Ugly: Avoiding the Pitfalls of Instrumental Variables Estimation" *Bates College Working Paper*.

[36] Murray, M. (2006b) "Avoiding Invalid Instruments and Coping with Weak Instruments" *Journal of Economic Perspectives*, vol. 20, no. 4, pp. 111-132.

[37] Oreopoulos, P. (2006) "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter" *American Economic Review*, vol. 96, no. 1, pp. 152-175.

[38] Park, J. (1999) "Estimation of sheepskin effects using the old and the new measures of educational attainment in the Current Population Survey" *Economics Letters*, vol. 62, no. 2, pp.237-240 .

[39] Staiger, D. and J. Stock (1997) "Instrumental variables regression with weak instruments" *Econometrica*, vol. 65, no. 3, pp. 557-586.

[40] Stock, J. and M. Yogo (2005) "Testing for Weak Instruments in IV regression" in Andrews, D. and J. Stock (eds), *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press, pp. 80-108.

[41] Uusitalo, R. (1999) "Return to Education in Finland" *Labour Economics*, vol. 6, pp. 569-580.

[42] Wooldrige, J. (2002) *Econometric Analysis of Cross Sectional and Panel Data*. Cambridge, US: The MIT Press.

12 Tables

Table 1: Effect of schooling on probability of Current and Early Smoking

	Current Smoker		Smoker at 16		
	marginal fx	**z**	**marginal fx**	**z**	**x-bar**
years of schooling	−0.027∗∗∗	−7.24	−0.038∗∗∗	−8.38	12.306
age	0.005	1.35	0.002	0.50	42.374
age^2	0.000∗∗∗	−2.99	0.000	−0.36	1939.210
year-of-birth	−0.016∗∗∗	−2.66	−0.002	−0.30	59.190 (=1955)
year-of-birth^2	0.000∗∗∗	2.84	0.000	0.25	3647.300
region: North	−0.032	−0.76	−0.033	−0.66	0.066
region: Yorkshire	0.037	0.99	0.031	0.69	0.098
region: North West	0.037	0.97	0.031	0.69	0.104
region: East Midlands	0.056	1.46	0.060	1.30	0.094
region: East Anglia	0.047	1.05	0.112∗	1.94	0.043
region: South East	0.055∗	1.72	0.047	1.23	0.285
region: South West	−0.005	−0.12	0.059	1.29	0.097
region: Wales	0.062	1.36	0.006	0.11	0.053
region: Scotland	0.065	1.57	0.063	1.28	0.078
ethnicity: Black	−0.121	−1.27	−0.181	−1.39	0.006
ethnicity: Asian	0.225∗∗∗	3.18	−0.176∗∗	−2.29	0.016
ethnicity: Other	−0.016	−0.16	−0.174	−1.55	0.008
father's occ class: 1	−0.022	−0.74	−0.041	−1.17	0.141
father's occ class: 2	−0.094∗∗	−2.39	−0.094∗	−1.88	0.058
father's occ class: 3	0.013	0.28	−0.099∗	−1.81	0.035
father's occ class: 4	−0.055	−1.28	−0.104∗∗	−2.10	0.047
father's occ class: 5	0.012	0.46	−0.010	−0.33	0.236
father's occ class: 6	0.022	0.49	−0.089∗	−1.79	0.042
father's occ class: 7	0.006	0.13	−0.002	−0.03	0.032
father's occ class: 9	0.009	0.29	0.016	0.43	0.094
father's occ class: 10	−0.012	−0.39	−0.056∗	−1.67	0.151
mother's occ class: 1	−0.017	−0.33	−0.026	−0.39	0.037
mother's occ class: 2	0.070	1.07	0.003	0.04	0.027
mother's occ class: 3	−0.050	−0.87	−0.036	−0.51	0.030
mother's occ class: 4	−0.036	−0.84	−0.023	−0.42	0.089
mother's occ class: 5	−0.023	−0.41	0.048	0.67	0.028
mother's occ class: 6	0.025	0.51	0.036	0.62	0.068
mother's occ class: 7	−0.054	−1.16	−0.083	−1.48	0.060
mother's occ class: 9	−0.017	−0.39	−0.057	−1.08	0.083
mother's occ class: 10	−0.012	−0.33	−0.060	−1.29	0.532
'nuclear family' to 16	−0.062∗∗∗	−2.86	−0.099∗∗∗	−3.73	0.820
mid 1990s	0.002	0.22	0.000	0.06	0.223
late 1990s	0.037∗∗∗	2.90	0.035∗∗∗	2.85	0.200
post 2000	−0.006	−0.30	0.002	0.08	0.371
# individuals	2805		2805		
# observations	33298		33298		
obs. prob.	0.287		0.344		
pred. prob. (at x-bar)	0.276		0.331		

Notes: Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 2: Smokers at 16/18 by quintile of the mean log wage residual distribution

quintile	**Non-smoker at 16**	**Smoker at 16**	**Total**	**Non-smoker at 18**	**Smoker at 18**	**Total**
1	256	198	454	209	245	454
	56.39%	43.61%	100.00%	46.04%	53.96%	100.00%
2	278	175	453	216	237	453
	61.37%	38.63%	100.00%	47.68%	52.32%	100.00%
3	319	134	453	265	188	453
	70.42%	29.58%	100.00%	58.50%	41.50%	100.00%
4	299	154	453	255	198	453
	66.00%	34.00%	100.00%	56.29%	43.71%	100.00%
5	349	104	453	295	158	453
	77.04%	22.96%	100.00%	65.12%	34.88%	100.00%
Total	1501	765	2266	1240	1026	2266
	66.24%	33.76%	100.00%	54.72%	45.28%	100.00%

Notes: OLS log wage regression (Table 5 column 1) run on pooled panel dataset, residuals are taken and the mean residual for each individual is calculated. These are then ranked into 5 quintiles as a measure of unobserved ability.

Table 3: Sample Summary Statistics

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
log wage	21256	2.214	0.437	0.848	3.407
years of schooling	21256	12.507	2.646	7	21
smoker at age 16	21256	0.315	0.465	0	1
minimum school leaving age was 16	21256	0.537	0.499	0	1
age	21256	39.991	10.757	19	65
cohort: born in the 1920s	21256	0.003	0.052	0	1
cohort: born in the 1930s	21256	0.050	0.219	0	1
cohort: born in the 1940s	21256	0.203	0.402	0	1
cohort: born in the 1950s	21256	0.274	0.446	0	1
cohort: born in the 1960s	21256	0.319	0.466	0	1
cohort: born in the 1970s	21256	0.146	0.354	0	1
cohort: born in the 1980s	21256	0.005	0.069	0	1
region: North	21256	0.068	0.252	0	1
region: Yorkshire	21256	0.097	0.296	0	1
region: North West	21256	0.106	0.308	0	1
region: East Midlands	21256	0.092	0.290	0	1
region: East Anglia	21256	0.043	0.202	0	1
region: South East	21256	0.280	0.449	0	1
region: South West	21256	0.100	0.300	0	1
region: Wales	21256	0.051	0.221	0	1
region: Scotland	21256	0.076	0.265	0	1
ethnicity: Black	21256	0.004	0.062	0	1
ethnicity: Asian	21256	0.016	0.124	0	1
ethnicity: Other	21256	0.007	0.083	0	1
father's occ class: 1	21256	0.139	0.346	0	1
father's occ class: 2	21256	0.064	0.244	0	1
father's occ class: 3	21256	0.038	0.191	0	1
father's occ class: 4	21256	0.049	0.216	0	1
father's occ class: 5	21256	0.234	0.423	0	1
father's occ class: 6	21256	0.044	0.205	0	1
father's occ class: 7	21256	0.032	0.177	0	1
father's occ class: 8	21256	0.171	0.377	0	1
father's occ class: 9	21256	0.086	0.280	0	1
father's occ class: 10	21256	0.143	0.350	0	1
mother's occ class: 1	21256	0.037	0.188	0	1
mother's occ class: 2	21256	0.026	0.159	0	1
mother's occ class: 3	21256	0.032	0.175	0	1
mother's occ class: 4	21256	0.098	0.297	0	1
mother's occ class: 5	21256	0.029	0.168	0	1
mother's occ class: 6	21256	0.073	0.260	0	1
mother's occ class: 7	21256	0.066	0.248	0	1
mother's occ class: 8	21256	0.051	0.220	0	1
mother's occ class: 9	21256	0.084	0.277	0	1
mother's occ class: 10	21256	0.505	0.500	0	1
'nuclear family' to 16	21256	0.831	0.375	0	1
early 1990s	21256	0.195	0.396	0	1
mid 1990s	21256	0.213	0.409	0	1
late 1990s	21256	0.221	0.415	0	1
post 2000	21256	0.371	0.483	0	1
number of observations per person	2266	9.380	4.516	1	15

Notes: 'nuclear family' to 16 means lived with both natural parents from birth to age 16. Occupational class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (8) plant/machine operative, (9) other, (10) self-emp/unemp.

Table 4: Sample Summary Statistics, by Early Smoking Status

	Smoker at 16					Non-Smoker at 16				
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
log wage	6696	2.120	0.423	0.878	3.395	14560	2.257	0.437	0.848	3.407
years of schooling	6696	11.646	2.081	8	21	14560	12.902	2.781	7	21
smoker at age 16	6696	1.000	0.000	1	1	14560	0.000	0.000	0	0
min. sch leaving age 16	6696	0.486	0.500	0	1	14560	0.561	0.496	0	1
age	6696	41.230	11.337	19	65	14560	39.421	10.431	19	65
cohort: born in the 1920s	6696	0.004	0.063	0	1	14560	0.002	0.046	0	1
cohort: born in the 1930s	6696	0.063	0.244	0	1	14560	0.044	0.206	0	1
cohort: born in the 1940s	6696	0.258	0.438	0	1	14560	0.177	0.382	0	1
cohort: born in the 1950s	6696	0.250	0.433	0	1	14560	0.284	0.451	0	1
cohort: born in the 1960s	6696	0.266	0.442	0	1	14560	0.344	0.475	0	1
cohort: born in the 1970s	6696	0.150	0.357	0	1	14560	0.145	0.352	0	1
cohort: born in the 1980s	6696	0.008	0.091	0	1	14560	0.003	0.057	0	1
region: North	6696	0.057	0.231	0	1	14560	0.073	0.260	0	1
region: Yorkshire	6696	0.103	0.304	0	1	14560	0.095	0.293	0	1
region: North West	6696	0.103	0.303	0	1	14560	0.107	0.309	0	1
region: East Midlands	6696	0.095	0.293	0	1	14560	0.091	0.288	0	1
region: East Anglia	6696	0.050	0.219	0	1	14560	0.039	0.193	0	1
region: South East	6696	0.272	0.445	0	1	14560	0.284	0.451	0	1
region: South West	6696	0.119	0.324	0	1	14560	0.091	0.288	0	1
region: Wales	6696	0.044	0.204	0	1	14560	0.055	0.228	0	1
region: Scotland	6696	0.082	0.274	0	1	14560	0.074	0.261	0	1
ethnicity: Black	6696	0.001	0.037	0	1	14560	0.005	0.070	0	1
ethnicity: Asian	6696	0.006	0.075	0	1	14560	0.020	0.140	0	1
ethnicity: Other	6696	0.002	0.049	0	1	14560	0.009	0.094	0	1
father's occ class: 1	6696	0.123	0.329	0	1	14560	0.146	0.353	0	1
father's occ class: 2	6696	0.036	0.185	0	1	14560	0.077	0.266	0	1
father's occ class: 3	6696	0.027	0.161	0	1	14560	0.043	0.203	0	1
father's occ class: 4	6696	0.032	0.177	0	1	14560	0.057	0.231	0	1
father's occ class: 5	6696	0.257	0.437	0	1	14560	0.223	0.416	0	1
father's occ class: 6	6696	0.036	0.187	0	1	14560	0.048	0.213	0	1
father's occ class: 7	6696	0.036	0.187	0	1	14560	0.031	0.172	0	1
father's occ class: 8	6696	0.194	0.396	0	1	14560	0.161	0.367	0	1
father's occ class: 9	6696	0.113	0.316	0	1	14560	0.074	0.261	0	1
father's occ class: 10	6696	0.146	0.353	0	1	14560	0.142	0.349	0	1
mother's occ class: 1	6696	0.031	0.173	0	1	14560	0.039	0.195	0	1
mother's occ class: 2	6696	0.018	0.134	0	1	14560	0.030	0.169	0	1
mother's occ class: 3	6696	0.028	0.166	0	1	14560	0.033	0.179	0	1
mother's occ class: 4	6696	0.085	0.279	0	1	14560	0.104	0.305	0	1
mother's occ class: 5	6696	0.043	0.204	0	1	14560	0.023	0.148	0	1
mother's occ class: 6	6696	0.098	0.297	0	1	14560	0.062	0.240	0	1
mother's occ class: 7	6696	0.051	0.221	0	1	14560	0.072	0.259	0	1
mother's occ class: 8	6696	0.053	0.223	0	1	14560	0.050	0.218	0	1
mother's occ class: 9	6696	0.092	0.288	0	1	14560	0.080	0.272	0	1
mother's occ class: 10	6696	0.501	0.500	0	1	14560	0.507	0.500	0	1
'nuclear family' to 16	6696	0.795	0.404	0	1	14560	0.848	0.359	0	1
early 1990s	6696	0.196	0.397	0	1	14560	0.194	0.396	0	1
mid 1990s	6696	0.212	0.409	0	1	14560	0.214	0.410	0	1
late 1990s	6696	0.219	0.414	0	1	14560	0.222	0.415	0	1
post 2000	6696	0.373	0.484	0	1	14560	0.371	0.483	0	1
# obs. per person	765	8.753	4.650	1	15	1501	9.700	4.414	1	15

Notes: 'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Occupational class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (8) plant/machine operative, (9) other, (10) self-emp/unemp.

Table 5: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status

	OLS		IV: smoker at 16		IV: first stage	
		Robust		Robust		Robust
Dep. Var: log hourly wage	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**
constant	−0.754***	0.250	−0.607**	0.287	−0.471	1.664
years of schooling	0.046***	0.003	0.129***	0.020	— —	— —
smoker at 16 indicator	— —	— —	— —	— —	−0.876***	0.108
age	0.099***	0.004	0.094***	0.005	0.056***	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.052***	0.011	0.398***	0.041
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.003***	0.000
region: North	0.047	0.038	0.054	0.044	−0.103	0.272
region: Yorkshire	0.003	0.033	−0.022	0.041	0.331	0.253
region: North West	0.054*	0.032	0.023	0.040	0.402	0.253
region: East Midlands	−0.010	0.032	−0.005	0.038	−0.034	0.235
region: East Anglia	0.015	0.039	−0.009	0.048	0.366	0.324
region: South East	0.142***	0.028	0.082**	0.037	0.757***	0.206
region: South West	0.023	0.034	0.015	0.041	0.175	0.237
region: Wales	−0.012	0.040	−0.019	0.045	0.081	0.285
region: Scotland	0.028	0.036	−0.021	0.044	0.643**	0.262
ethnicity: Black	0.114	0.105	0.115	0.117	−0.164	0.779
ethnicity: Asian	−0.136*	0.071	−0.312***	0.105	1.965***	0.485
ethnicity: Other	−0.048	0.103	−0.234**	0.119	2.067*	1.111
father's occ class: 1	0.116***	0.028	0.020	0.041	1.122***	0.214
father's occ class: 2	0.121***	0.038	−0.077	0.065	2.268***	0.291
father's occ class: 3	0.089**	0.043	−0.043	0.058	1.499***	0.321
father's occ class: 4	0.065*	0.036	−0.053	0.051	1.320***	0.305
father's occ class: 5	0.038*	0.023	0.011	0.028	0.335**	0.170
father's occ class: 6	0.014	0.035	−0.074	0.048	0.991***	0.305
father's occ class: 7	0.103***	0.040	0.066	0.049	0.467	0.330
father's occ class: 9	−0.021	0.029	0.028	0.035	−0.551***	0.197
father's occ class: 10	0.029	0.027	0.027	0.030	−0.012	0.186
mother's occ class: 1	0.047	0.049	0.035	0.061	0.112	0.411
mother's occ class: 2	0.015	0.054	−0.103	0.070	1.433***	0.439
mother's occ class: 3	0.056	0.048	0.053	0.057	0.046	0.387
mother's occ class: 4	0.055	0.040	0.014	0.048	0.485	0.307
mother's occ class: 5	0.010	0.049	0.031	0.058	−0.117	0.417
mother's occ class: 6	0.025	0.040	0.029	0.045	0.054	0.311
mother's occ class: 7	0.055	0.041	0.057	0.048	−0.083	0.312
mother's occ class: 9	−0.004	0.038	0.034	0.044	−0.461	0.284
mother's occ class: 10	0.004	0.032	−0.006	0.036	0.115	0.253
'nuclear family' to 16	0.028	0.019	0.001	0.022	0.247*	0.136
mid 1990s	−0.045***	0.009	−0.050***	0.010	0.067	0.046
late 1990s	−0.065***	0.014	−0.070***	0.016	0.080	0.081
post 2000	−0.033	0.021	−0.040*	0.023	0.108	0.126
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.072		0.246	

F-test on exclusion of smoking at 16 from first stage: 66.17; Partial R^2 of instrument = 0.0289

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 6: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status, Basic Specification

	OLS		**IV: smoker at 16**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.849***	0.247	−0.590**	0.280	−2.204	1.697
years of schooling	0.052***	0.003	0.121***	0.016	— —	— —
smoker at 16 indicator	— —	— —	— —	— —	−1.087***	0.113
age	0.098***	0.004	0.094***	0.005	0.067***	0.023
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.014**	0.007	−0.048***	0.011	0.466***	0.042
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.004	0.000
region: North	0.041	0.038	0.048	0.044	−0.124	0.286
region: Yorkshire	−0.003	0.033	−0.018	0.039	0.269	0.265
region: North West	0.050	0.033	0.022	0.039	0.432	0.270
region: East Midlands	−0.016	0.032	−0.006	0.037	−0.122	0.258
region: East Anglia	0.010	0.040	−0.006	0.047	0.318	0.338
region: South East	0.143***	0.028	0.080***	0.036	0.946***	0.219
region: South West	0.023	0.034	0.010	0.040	0.271	0.251
region: Wales	−0.018	0.040	−0.019	0.044	0.003	0.305
region: Scotland	0.020	0.036	−0.023	0.043	0.688**	0.283
ethnicity: Black	0.117	0.093	0.104	0.113	−0.015	0.751
ethnicity: Asian	−0.150**	0.070	−0.290***	0.098	1.844***	0.519
ethnicity: Other	−0.042	0.095	−0.221*	0.119	2.406**	0.996
mid 1990s	−0.047***	0.009	−0.049***	0.010	0.040	0.050
late 1990s	−0.068***	0.014	−0.068***	0.016	0.019	0.086
post 2000	−0.038*	0.021	−0.037	0.023	0.014	0.133
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.251		0.098		0.143	
F-test on exclusion of instrument from first stage: 92.39; Partial R^2 of the instrument = 0.0400						

Notes: *** significant at 1% level, ** significant at 5% level, * significant at 10% level.
Standard errors are clustered at the level of the individual and robust.

Table 7: First Stage IV regression coefficients using Smoker at 16 indicator interacted with year turned 16 indicator

	Coeff.	**Robust Std. Err.**	***t***	***p***
Smoker at 16 × turned 16 pre-1965	−0.797***	0.209	−3.82	0.000
Smoker at 16 × turned 16 post-1965	−0.904***	0.120	−7.51	0.000
# observations	21256			
R^2	0.247			

Notes: *** significant at 1% level; standard errors clustered at individual level and robust.
Turned 16 post-1965 includes those turning 16 from January 1965 onwards.
Other covariates included in these first stage regressions are those in Table 5.

Table 8: First Stage IV Regression coefficients on Smoker at 16 indicator and on Minimum School Leaving Age of 16 indicator, by quintile of the mean log wage residual distribution

	IV first stage, Early Smoking			**IV first stage, RoSLA**		
quintile	**Coeff. on smoker 16**	**Robust Std. Err.**	**R^2**	**Coeff. on MSLA=16**	**Robust Std. Err.**	**R^2**
1 #obs = 3684 mean years of schooling 12.41	-0.773***	0.265	0.268	1.044**	0.510	0.262
2 #obs = 4285 mean years of schooling 12.09	-1.044***	0.227	0.317	0.837*	0.458	0.292
3 #obs = 4461 mean years of schooling 12.30	-0.950***	0.249	0.329	0.315	0.496	0.309
4 #obs = 4496 mean years of schooling 12.28	-0.747***	0.213	0.257	0.398	0.388	0.240
5 #obs = 4330 mean years of schooling 12.65	-0.879***	0.241	0.341	0.080	0.435	0.321

Notes: *** signficant at 1% level, ** significant at 5% level, * significant at 10% level
Standard errors are clustered at the level of the individual and robust.
Other covariates included in regressions are as Table 5.

Table 9: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 18 Status

	OLS		**IV: smoker at 18**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.596**	0.293	−0.399	1.675
years of schooling	0.046***	0.003	0.135***	0.023	— —	— —
smoker at 18 indicator	— —	— —	— —	— —	−0.745***	0.108
age	0.099***	0.004	0.093***	0.005	0.054**	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.054***	0.012	0.399***	0.041
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.003***	0.000
region: North	0.047	0.038	0.054	0.045	−0.121	0.274
region: Yorkshire	0.003	0.033	−0.024	0.042	0.319	0.254
region: North West	0.054*	0.032	0.020	0.041	0.414	0.253
region: East Midlands	−0.010	0.032	−0.005	0.038	−0.041	0.236
region: East Anglia	0.015	0.039	−0.011	0.050	0.392	0.324
region: South East	0.142***	0.028	0.077**	0.039	0.760***	0.209
region: South West	0.023	0.034	0.014	0.042	0.138	0.238
region: Wales	−0.012	0.040	−0.020	0.046	0.063	0.287
region: Scotland	0.028	0.036	−0.025	0.046	0.600**	0.264
ethnicity: Black	0.114	0.105	0.115	0.119	−0.212	0.774
ethnicity: Asian	−0.136*	0.071	−0.325***	0.112	2.081***	0.511
ethnicity: Other	−0.048	0.103	−0.248**	0.124	2.112*	1.091
father's occ class: 1	0.116***	0.028	0.013	0.044	1.160***	0.213
father's occ class: 2	0.121***	0.038	−0.092	0.071	2.327***	0.292
father's occ class: 3	0.089**	0.043	−0.053	0.063	1.514***	0.326
father's occ class: 4	0.065*	0.036	−0.062	0.054	1.362***	0.309
father's occ class: 5	0.038*	0.023	0.009	0.029	0.340**	0.170
father's occ class: 6	0.014	0.035	−0.081	0.051	0.983***	0.308
father's occ class: 7	0.103***	0.040	0.063	0.050	0.493	0.329
father's occ class: 9	−0.021	0.029	0.032	0.036	−0.551***	0.196
father's occ class: 10	0.029	0.027	0.026	0.031	−0.008	0.186
mother's occ class: 1	0.047	0.049	0.034	0.063	0.062	0.411
mother's occ class: 2	0.015	0.054	−0.112	0.074	1.352***	0.443
mother's occ class: 3	0.056	0.048	0.053	0.059	0.074	0.388
mother's occ class: 4	0.055	0.040	0.011	0.050	0.471	0.310
mother's occ class: 5	0.010	0.049	0.033	0.060	−0.199	0.420
mother's occ class: 6	0.025	0.040	0.029	0.046	0.021	0.313
mother's occ class: 7	0.055	0.041	0.058	0.049	−0.113	0.313
mother's occ class: 9	−0.004	0.038	0.037	0.046	−0.489*	0.287
mother's occ class: 10	0.004	0.032	−0.007	0.037	0.088	0.256
'nuclear family' to 16	0.028	0.019	−0.001	0.022	0.258*	0.135
mid 1990s	−0.045***	0.009	−0.050***	0.010	0.066	0.046
late 1990s	−0.065***	0.014	−0.070***	0.016	0.079	0.082
post 2000	−0.033	0.021	−0.040*	0.024	0.100	0.127
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.042		0.242	

F-test on exclusion of smoking at 18 from first stage: 48.02; Partial R^2 of instrument = 0.0236

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 10: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 15 Status

	OLS		**IV: smoker at 15**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.609**	0.287	−1.636	1.663
years of schooling	0.046***	0.003	0.128***	0.023	— —	— —
smoker at 15 indicator	— —	— —	— —	— —	−0.947***	0.122
age	0.099***	0.004	0.094***	0.005	0.050**	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.051***	0.012	0.435***	0.041
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.004***	0.000
region: North	0.047	0.038	0.054	0.044	−0.116	0.268
region: Yorkshire	0.003	0.033	−0.021	0.041	0.330	0.255
region: North West	0.054*	0.032	0.023	0.040	0.366	0.253
region: East Midlands	−0.010	0.032	−0.005	0.038	−0.036	0.234
region: East Anglia	0.015	0.039	−0.009	0.048	0.305	0.323
region: South East	0.142***	0.028	0.083**	0.037	0.739***	0.206
region: South West	0.023	0.034	0.015	0.040	0.150	0.237
region: Wales	−0.012	0.040	−0.019	0.045	0.087	0.285
region: Scotland	0.028	0.036	−0.020	0.045	0.600**	0.262
ethnicity: Black	0.114	0.105	0.115	0.117	−0.154	0.739
ethnicity: Asian	−0.136*	0.071	−0.310***	0.107	2.050***	0.487
ethnicity: Other	−0.048	0.103	−0.231*	0.120	2.184**	1.111
father's occ class: 1	0.116***	0.028	0.022	0.043	1.136***	0.214
father's occ class: 2	0.121***	0.038	−0.074	0.070	2.328***	0.294
father's occ class: 3	0.089**	0.043	−0.041	0.061	1.507***	0.328
father's occ class: 4	0.065*	0.036	−0.052	0.052	1.370***	0.303
father's occ class: 5	0.038*	0.023	0.011	0.028	0.344**	0.170
father's occ class: 6	0.014	0.035	−0.073	0.049	0.987***	0.309
father's occ class: 7	0.103***	0.040	0.066	0.049	0.512	0.329
father's occ class: 9	−0.021	0.029	0.028	0.035	−0.509***	0.197
father's occ class: 10	0.029	0.027	0.027	0.030	0.010	0.187
mother's occ class: 1	0.047	0.049	0.035	0.061	0.134	0.412
mother's occ class: 2	0.015	0.054	−0.101	0.072	1.412***	0.448
mother's occ class: 3	0.056	0.048	0.053	0.057	0.024	0.391
mother's occ class: 4	0.055	0.040	0.015	0.049	0.472	0.308
mother's occ class: 5	0.010	0.049	0.031	0.058	−0.178	0.424
mother's occ class: 6	0.025	0.040	0.029	0.045	0.051	0.311
mother's occ class: 7	0.055	0.041	0.057	0.048	−0.102	0.312
mother's occ class: 9	−0.004	0.038	0.034	0.045	−0.502*	0.284
mother's occ class: 10	0.004	0.032	−0.006	0.036	0.085	0.254
'nuclear family' to 16	0.028	0.019	0.002	0.022	0.277**	0.136
mid 1990s	−0.045***	0.009	−0.049***	0.010	0.072	0.047
late 1990s	−0.065***	0.014	−0.070***	0.016	0.102	0.082
post 2000	−0.033	0.021	−0.040*	0.023	0.123	0.127
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.077		0.242	

F-test on exclusion of smoking at 15 from first stage: 60.17; Partial R^2 of instrument = 0.0229

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 11: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 14 Status

	OLS		**IV: smoker at 14**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.569*	0.309	−1.693	1.670
years of schooling	0.046***	0.003	0.150***	0.030	— —	— —
smoker at 14 indicator	— —	— —	— —	— —	−0.911***	0.148
age	0.099***	0.004	0.093***	0.005	0.053**	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.061***	0.015	0.435***	0.041
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.004***	0.000
region: North	0.047	0.038	0.056	0.048	−0.130	0.269
region: Yorkshire	0.003	0.033	−0.028	0.045	0.320	0.254
region: North West	0.054*	0.032	0.015	0.044	0.360	0.252
region: East Midlands	−0.010	0.032	−0.004	0.040	−0.047	0.233
region: East Anglia	0.015	0.039	−0.015	0.053	0.254	0.323
region: South East	0.142***	0.028	0.066	0.043	0.716***	0.205
region: South West	0.023	0.034	0.013	0.044	0.098	0.237
region: Wales	−0.012	0.040	−0.021	0.049	0.093	0.286
region: Scotland	0.028	0.036	−0.034	0.050	0.586**	0.261
ethnicity: Black	0.114	0.105	0.115	0.125	−0.111	0.737
ethnicity: Asian	−0.136*	0.071	−0.357***	0.120	2.073***	0.494
ethnicity: Other	−0.048	0.103	−0.282**	0.140	2.149*	1.096
father's occ class: 1	0.116***	0.028	−0.004	0.051	1.132***	0.215
father's occ class: 2	0.121***	0.038	−0.128	0.087	2.328***	0.298
father's occ class: 3	0.089**	0.043	−0.077	0.071	1.551***	0.328
father's occ class: 4	0.065*	0.036	−0.084	0.062	1.383***	0.306
father's occ class: 5	0.038*	0.023	0.004	0.031	0.324*	0.171
father's occ class: 6	0.014	0.035	−0.097*	0.056	0.984***	0.313
father's occ class: 7	0.103***	0.040	0.056	0.055	0.431	0.334
father's occ class: 9	−0.021	0.029	0.041	0.038	−0.519***	0.198
father's occ class: 10	0.029	0.027	0.026	0.032	0.019	0.187
mother's occ class: 1	0.047	0.049	0.032	0.067	0.092	0.413
mother's occ class: 2	0.015	0.054	−0.133	0.081	1.381***	0.454
mother's occ class: 3	0.056	0.048	0.052	0.062	−0.011	0.393
mother's occ class: 4	0.055	0.040	0.004	0.054	0.408	0.310
mother's occ class: 5	0.010	0.049	0.037	0.063	−0.262	0.432
mother's occ class: 6	0.025	0.040	0.030	0.049	0.004	0.312
mother's occ class: 7	0.055	0.041	0.058	0.052	−0.114	0.314
mother's occ class: 9	−0.004	0.038	0.044	0.048	−0.566**	0.286
mother's occ class: 10	0.004	0.032	−0.008	0.040	0.038	0.255
'nuclear family' to 16	0.028	0.019	−0.006	0.024	0.268**	0.135
mid 1990s	−0.045***	0.009	−0.051***	0.010	0.061	0.047
late 1990s	−0.065***	0.014	−0.071***	0.017	0.082	0.082
post 2000	−0.033	0.021	−0.041*	0.025	0.100	0.128
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.191		0.235	
F-test on exclusion of smoking at 14 from first stage: 38.10; Partial R^2 of instrument = 0.0148						

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 12: Effect of Early Smoking on Probability of Being a Home-owner, Probit Model

	Home Owner		
	marginal fx	**z**	**x-bar**
log of hourly wage	0.204***	16.15	2.196
smoker at 16 indicator	−0.044***	−3.49	0.313
age	−0.004	−1.03	39.430
age^2	0.000***	2.93	1682.770
year-of-birth	0.022***	3.60	62.39 = 1958
year-of-birth^2	0.000***	−3.20	4025.480
region: North	−0.013	−0.42	0.069
region: Yorkshire	−0.067**	−2.22	0.096
region: North West	0.009	0.35	0.105
region: East Midlands	−0.043	−1.48	0.090
region: East Anglia	−0.038	−1.09	0.043
region: South East	−0.127***	−4.96	0.287
region: South West	−0.100***	−3.11	0.098
region: Wales	−0.075**	−2.00	0.050
region: Scotland	−0.107***	−3.17	0.075
ethnicity: Black	−0.094	−1.57	0.006
ethnicity: Asian	0.075**	2.25	0.017
ethnicity: Other	0.045	0.72	0.007
father's occ class: 1	0.040*	1.93	0.134
father's occ class: 2	−0.004	−0.16	0.064
father's occ class: 3	0.021	0.65	0.038
father's occ class: 4	0.026	0.76	0.047
father's occ class: 5	0.015	0.82	0.222
father's occ class: 6	−0.028	−0.94	0.044
father's occ class: 7	−0.013	−0.35	0.032
father's occ class: 9	−0.087***	−3.08	0.080
father's occ class: 10	−0.002	−0.12	0.177
mother's occ class: 1	0.038	0.95	0.034
mother's occ class: 2	−0.021	−0.47	0.029
mother's occ class: 3	−0.046	−1.05	0.031
mother's occ class: 4	−0.007	−0.21	0.094
mother's occ class: 5	0.016	0.38	0.026
mother's occ class: 6	0.017	0.52	0.073
mother's occ class: 7	0.020	0.60	0.063
mother's occ class: 9	−0.017	−0.49	0.078
mother's occ class: 10	−0.034	−1.27	0.525
'nuclear family' to 16	0.017	1.21	0.825
mid 1990s	−0.004	−0.43	0.207
late 1990s	−0.024*	−1.66	0.224
post 2000	−0.057***	−2.86	0.385
observed prob.	0.829		
predicted prob. (at x-bar)	0.863		
# observations	24034		
# individuals	2615		

Notes: Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'.
Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 13: Effect of Early Smoking on Probability of Having Had a Dental or Optician Check-up in the Last Year, Probit Models

	Dental Check		**Opticians Check**		
	marginal fx	**z**	**marginal fx**	**z**	**x-bar**
log of hourly wage	0.132∗∗∗	8.72	0.070∗∗∗	5.92	2.196
smoker at 16 indicator	−0.040∗∗∗	−2.67	−0.029∗∗∗	−2.59	0.313
age	0.001	0.11	−0.007	−1.59	39.415
age^2	0.000	1.46	0.000∗∗∗	3.55	1681.620
year-of-birth	0.039∗∗∗	5.49	0.006	1.03	62.41 = 1958
year-of-birth^2	0.000∗∗∗	−5.05	0.000	−0.88	4028.120
region: North	0.014	0.40	−0.004	−0.15	0.069
region: Yorkshire	−0.006	−0.20	−0.007	−0.30	0.096
region: North West	−0.034	−1.07	−0.015	−0.59	0.105
region: East Midlands	−0.039	−1.21	−0.012	−0.49	0.090
region: East Anglia	0.110∗∗∗	2.79	−0.017	−0.57	0.043
region: South East	−0.048∗	−1.81	−0.021	−1.02	0.287
region: South West	−0.003	−0.10	0.009	0.39	0.098
region: Wales	−0.058	−1.47	−0.008	−0.28	0.050
region: Scotland	−0.045	−1.25	0.005	0.20	0.075
ethnicity: Black	0.001	0.01	−0.014	−0.19	0.006
ethnicity: Asian	−0.151∗∗∗	−2.71	0.051	1.13	0.017
ethnicity: Other	−0.042	−0.52	0.128	1.62	0.007
father's occ class: 1	0.037	1.40	0.072∗∗∗	3.24	0.135
father's occ class: 2	0.047	1.46	0.080∗∗∗	2.87	0.064
father's occ class: 3	0.027	0.63	0.088∗∗	2.57	0.038
father's occ class: 4	0.047	1.24	0.010	0.37	0.047
father's occ class: 5	0.013	0.54	0.022	1.16	0.222
father's occ class: 6	0.026	0.75	0.044	1.42	0.044
father's occ class: 7	0.058	1.30	0.040	1.07	0.032
father's occ class: 9	−0.011	−0.35	0.016	0.66	0.080
father's occ class: 10	0.026	1.02	0.022	1.01	0.178
mother's occ class: 1	0.049	1.03	−0.016	−0.42	0.034
mother's occ class: 2	0.059	1.15	−0.048	−1.10	0.029
mother's occ class: 3	0.031	0.59	0.018	0.43	0.031
mother's occ class: 4	0.061	1.63	−0.016	−0.49	0.094
mother's occ class: 5	−0.016	−0.31	−0.018	−0.40	0.026
mother's occ class: 6	0.008	0.22	−0.013	−0.41	0.072
mother's occ class: 7	0.105	2.68	0.049	1.38	0.063
mother's occ class: 9	−0.019	−0.49	−0.026	−0.85	0.078
mother's occ class: 10	0.024	0.74	−0.014	−0.53	0.526
'nuclear family' to 16	0.035∗∗∗	1.94	0.019	1.27	0.825
mid 1990s	0.003	0.26	0.010	0.85	0.207
late 1990s	0.010	0.56	0.028∗	1.69	0.224
post 2000	0.018	0.67	0.023	0.94	0.386
observed prob.	0.631		0.307		
predicted prob. (at x-bar)	0.636		0.302		
# observations	24086		24086		
# individuals	2615		2615		

Notes: Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 14: Human Capital Earnings Function Estimations, OLS and IV using RoSLA

Dep. Var: log hourly wage	OLS Coeff.	OLS Robust Std. Err.	IV: RoSLA Coeff.	IV: RoSLA Robust Std. Err.	IV: first stage Coeff.	IV: first stage Robust Std. Err.
constant	−0.754***	0.250	−0.655**	0.280	−1.459	1.681
years of schooling	0.046***	0.003	0.102**	0.051	— —	— —
min. school LA=16	— —	— —	— —	— —	0.564***	0.206
age	0.099***	0.004	0.095***	0.005	0.056**	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.040*	0.023	0.427***	0.041
year-of-birth^2	0.000***	0.000	0.000**	0.000	−0.004***	0.000
region: North	0.047	0.038	0.051	0.041	−0.080	0.272
region: Yorkshire	0.003	0.033	−0.014	0.040	0.320	0.256
region: North West	0.054*	0.032	0.033	0.041	0.386	0.255
region: East Midlands	−0.010	0.032	−0.007	0.035	−0.035	0.234
region: East Anglia	0.015	0.039	−0.001	0.047	0.324	0.327
region: South East	0.142***	0.028	0.101**	0.051	0.741	0.208
region: South West	0.023	0.034	0.017	0.038	0.114***	0.240
region: Wales	−0.012	0.040	−0.017	0.043	0.093	0.290
region: Scotland	0.028	0.036	−0.005	0.050	0.658**	0.266
ethnicity: Black	0.114	0.105	0.114	0.110	0.037	0.746
ethnicity: Asian	−0.136*	0.071	−0.255*	0.139	2.146***	0.515
ethnicity: Other	−0.048	0.103	−0.174	0.152	2.214**	1.074
father's occ class: 1	0.116***	0.028	0.051	0.069	1.162***	0.216
father's occ class: 2	0.121***	0.038	−0.013	0.128	2.404***	0.298
father's occ class: 3	0.089**	0.043	0.000	0.093	1.585***	0.333
father's occ class: 4	0.065*	0.036	−0.015	0.083	1.440***	0.308
father's occ class: 5	0.038*	0.023	0.020	0.029	0.322*	0.172
father's occ class: 6	0.014	0.035	−0.046	0.064	1.046***	0.313
father's occ class: 7	0.103***	0.040	0.078	0.049	0.484	0.339
father's occ class: 9	−0.021	0.029	0.012	0.044	−0.592***	0.196
father's occ class: 10	0.029	0.027	0.028	0.028	0.043	0.186
mother's occ class: 1	0.047	0.049	0.039	0.056	0.107	0.426
mother's occ class: 2	0.015	0.054	−0.065	0.094	1.378***	0.454
mother's occ class: 3	0.056	0.048	0.054	0.053	0.007	0.395
mother's occ class: 4	0.055	0.040	0.027	0.050	0.453	0.317
mother's occ class: 5	0.010	0.049	0.025	0.054	−0.240	0.430
mother's occ class: 6	0.025	0.040	0.027	0.042	−0.070	0.322
mother's occ class: 7	0.055	0.041	0.057	0.044	−0.053	0.324
mother's occ class: 9	−0.004	0.038	0.022	0.047	−0.491*	0.293
mother's occ class: 10	0.004	0.032	−0.003	0.034	0.103	0.264
'nuclear family' to 16	0.028	0.019	0.010	0.026	0.330**	0.137
mid 1990s	−0.045***	0.009	−0.048***	0.010	0.063	0.047
late 1990s	−0.065***	0.014	−0.068***	0.015	0.075	0.083
post 2000	−0.033	0.021	−0.038*	0.023	0.094	0.129
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.177		0.227	

F-test on exclusion of min. school LA=16 from first stage: 7.49; Partial R^2 of the instrument = 0.0044

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table 15: Human Capital Earnings Function Estimations, OLS and IV using RoSLA, Basic Specification

	OLS		IV: RoSLA		IV: first stage	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.849***	0.247	−0.668**	0.303	−3.375*	1.727
years of schooling	0.052***	0.003	0.100**	0.042	— —	— —
min. school LA=16	— —	— —	— —	— —	0.691***	0.219
age	0.098***	0.004	0.095***	0.005	0.068***	0.023
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.014**	0.007	−0.038*	0.022	0.501***	0.042
$year\text{-}of\text{-}birth^2$	0.000***	0.000	0.000**	0.000	−0.004***	0.000
region: North	0.041	0.038	0.046	0.041	−0.089	0.286
region: Yorkshire	−0.003	0.033	−0.014	0.038	0.246	0.268
region: North West	0.050	0.033	0.030	0.041	0.418	0.273
region: East Midlands	−0.016	0.032	−0.009	0.035	−0.129	0.258
region: East Anglia	0.010	0.040	−0.001	0.045	0.277	0.343
region: South East	0.143***	0.028	0.099*	0.052	0.936***	0.223
region: South West	0.023	0.034	0.014	0.038	0.202	0.256
region: Wales	−0.018	0.040	−0.019	0.042	0.022	0.314
region: Scotland	0.020	0.036	−0.010	0.048	0.710**	0.288
ethnicity: Black	0.117	0.093	0.108	0.105	0.251	0.700
ethnicity: Asian	−0.150**	0.070	−0.248**	0.120	2.075***	0.560
ethnicity: Other	−0.042	0.095	−0.167	0.149	2.566***	0.939
mid 1990s	−0.047***	0.009	−0.048***	0.010	0.036	0.050
late 1990s	−0.068***	0.014	−0.068***	0.015	0.013	0.089
post 2000	−0.038*	0.021	−0.037*	0.022	−0.004	0.137
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.251		0.176		0.113	
F-test on exclusion of instrument from first stage: 9.98; Partial R^2 of the instrument = 0.0058						

Notes: *** significant at 1% level, ** significant at 5% level, * significant at 10% level.
Standard errors are clustered at the level of the individual and robust.

Table 16: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status and RoSLA

	OLS		**IV: both**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754∗∗∗	0.250	−0.613∗∗	0.283	−0.157	1.663
years of schooling	0.046∗∗∗	0.003	0.125∗∗∗	0.019	— —	— —
smoker at 16 indicator	— —	— —	— —	— —	−0.874∗∗∗	0.107
min. school LA=16	— —	— —	— —	— —	0.556∗∗∗	0.202
age	0.099∗∗∗	0.004	0.094∗∗∗	0.005	0.054∗∗	0.022
age^2	−0.001∗∗∗	0.000	−0.001∗∗∗	0.000	−0.001∗∗∗	0.000
year-of-birth	−0.016∗∗	0.007	−0.050∗∗∗	0.011	0.399∗∗∗	0.041
year-of-birth^2	0.000∗∗∗	0.000	0.001∗∗∗	0.000	−0.004∗∗∗	0.000
region: North	0.047	0.038	0.053	0.044	−0.097	0.272
region: Yorkshire	0.003	0.033	−0.021	0.040	0.347	0.253
region: North West	0.054∗	0.032	0.024	0.039	0.409	0.253
region: East Midlands	−0.010	0.032	−0.006	0.037	−0.014	0.235
region: East Anglia	0.015	0.039	−0.008	0.048	0.398	0.325
region: South East	0.142∗∗∗	0.028	0.084∗∗	0.036	0.767∗∗∗	0.207
region: South West	0.023	0.034	0.015	0.040	0.192	0.236
region: Wales	−0.012	0.040	−0.019	0.045	0.082	0.286
region: Scotland	0.028	0.036	−0.019	0.043	0.705∗∗∗	0.263
ethnicity: Black	0.114	0.105	0.115	0.116	−0.114	0.788
ethnicity: Asian	−0.136∗	0.071	−0.305∗∗∗	0.103	1.975∗∗∗	0.493
ethnicity: Other	−0.048	0.103	−0.226∗	0.116	2.021∗	1.080
father's occ class: 1	0.116∗∗∗	0.028	0.024	0.040	1.118∗∗∗	0.213
father's occ class: 2	0.121∗∗∗	0.038	−0.068	0.062	2.271∗∗∗	0.290
father's occ class: 3	0.089∗∗	0.043	−0.038	0.056	1.485∗∗∗	0.319
father's occ class: 4	0.065∗	0.036	−0.048	0.050	1.324∗∗∗	0.303
father's occ class: 5	0.038∗	0.023	0.012	0.027	0.322∗	0.170
father's occ class: 6	0.014	0.035	−0.070	0.046	0.968∗∗∗	0.303
father's occ class: 7	0.103∗∗∗	0.040	0.067	0.048	0.501	0.330
father's occ class: 9	−0.021	0.029	0.026	0.034	−0.542∗∗∗	0.194
father's occ class: 10	0.029	0.027	0.027	0.030	0.000	0.185
mother's occ class: 1	0.047	0.049	0.036	0.060	0.079	0.414
mother's occ class: 2	0.015	0.054	−0.098	0.069	1.379∗∗∗	0.442
mother's occ class: 3	0.056	0.048	0.053	0.057	0.018	0.388
mother's occ class: 4	0.055	0.040	0.016	0.048	0.451	0.310
mother's occ class: 5	0.010	0.049	0.030	0.057	−0.104	0.414
mother's occ class: 6	0.025	0.040	0.028	0.045	0.030	0.313
mother's occ class: 7	0.055	0.041	0.057	0.047	−0.111	0.316
mother's occ class: 9	−0.004	0.038	0.033	0.044	−0.488∗	0.285
mother's occ class: 10	0.004	0.032	−0.005	0.036	0.099	0.256
'nuclear family' to 16	0.028	0.019	0.002	0.022	0.251∗	0.136
mid 1990s	−0.045∗∗∗	0.009	−0.049∗∗∗	0.010	0.073	0.046
late 1990s	−0.065∗∗∗	0.014	−0.070∗∗∗	0.016	0.092	0.081
post 2000	−0.033	0.021	−0.039∗	0.023	0.120	0.126
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.088		0.250	

F-test on exclusion of instruments from first stage: 36.83; Partial R^2 of the instrument = 0.0332
Hansen's J-test of overidentification = 0.202, p-value = 0.6529

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

13 Figures

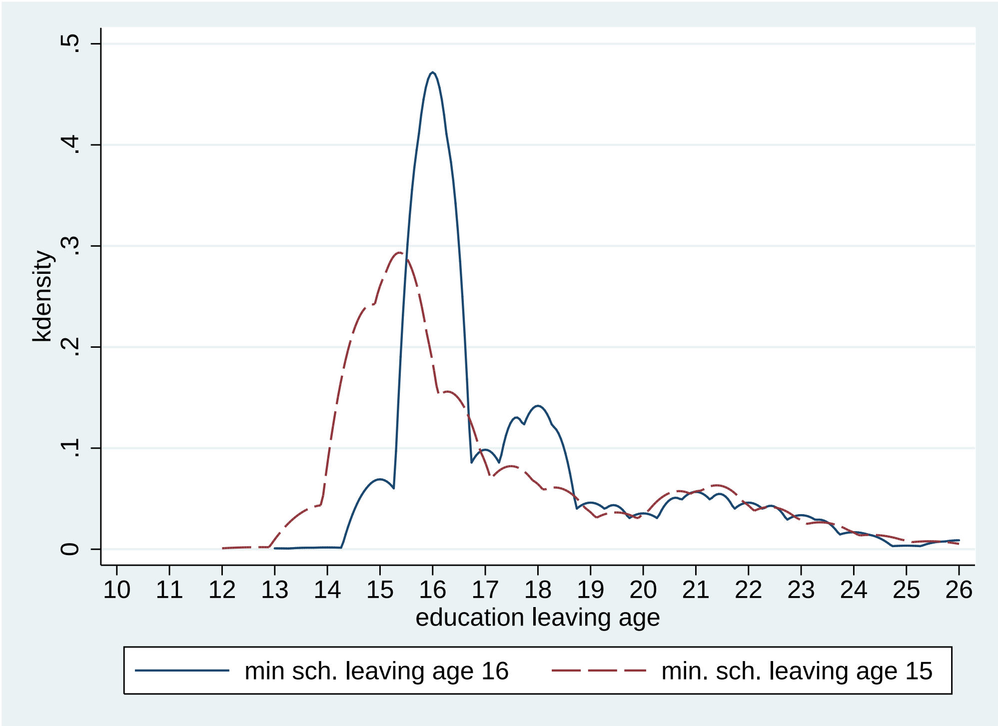

Figure 1: Education Leaving Age Density, by Minimum School Leaving Age

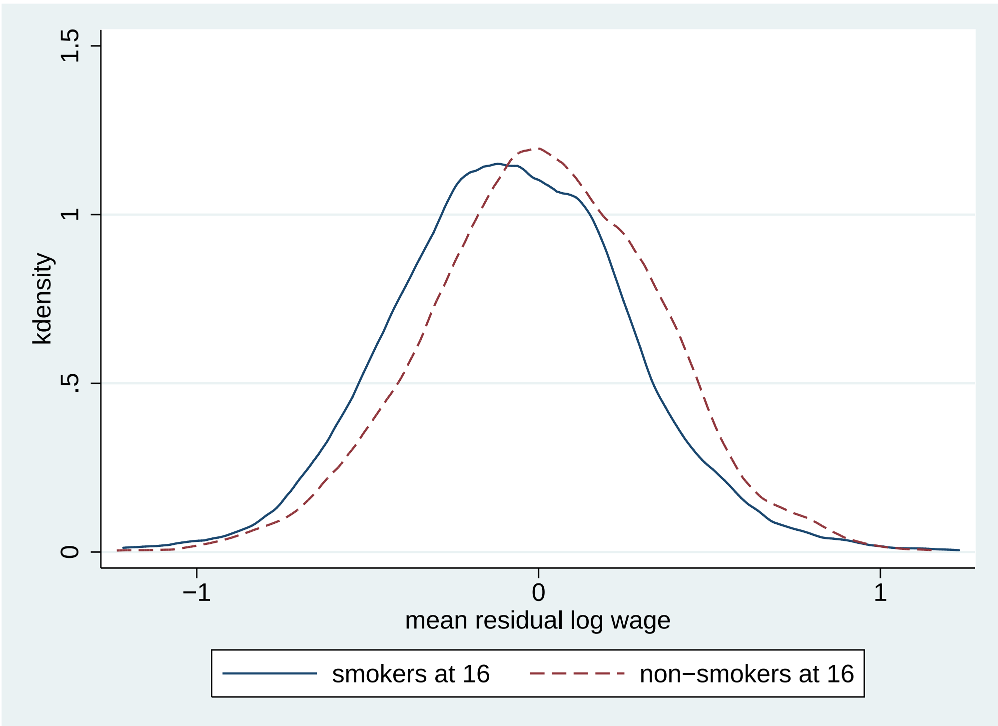

Figure 2: Residual Log Wage Density, by Smoker at 16 Status

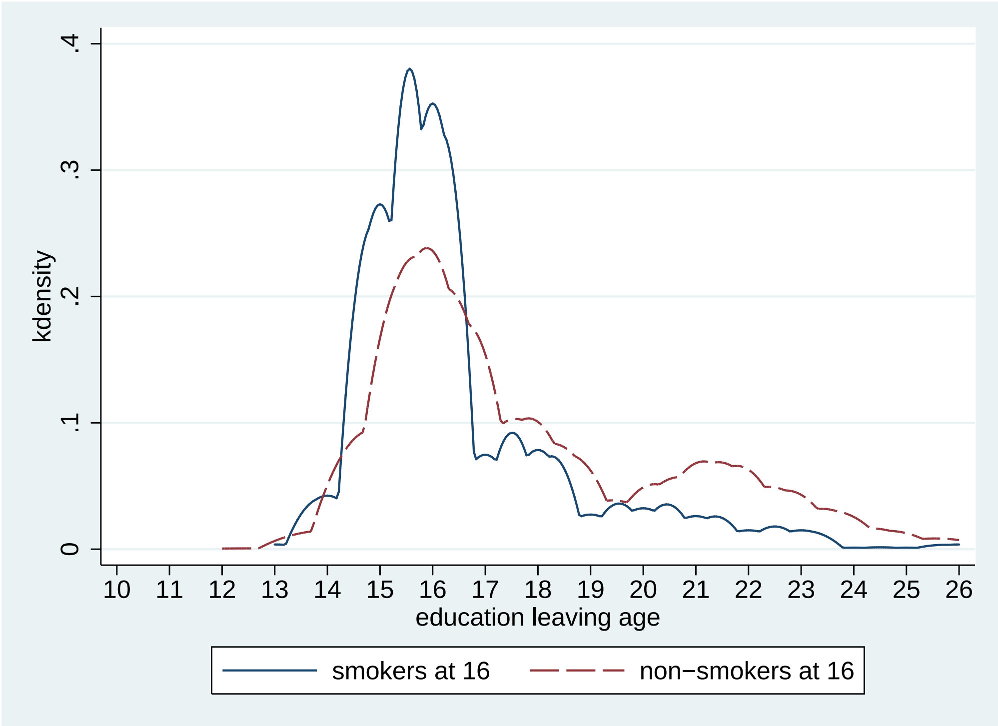

Figure 3: Education Leaving Age Density, by Smoker at 16 Status

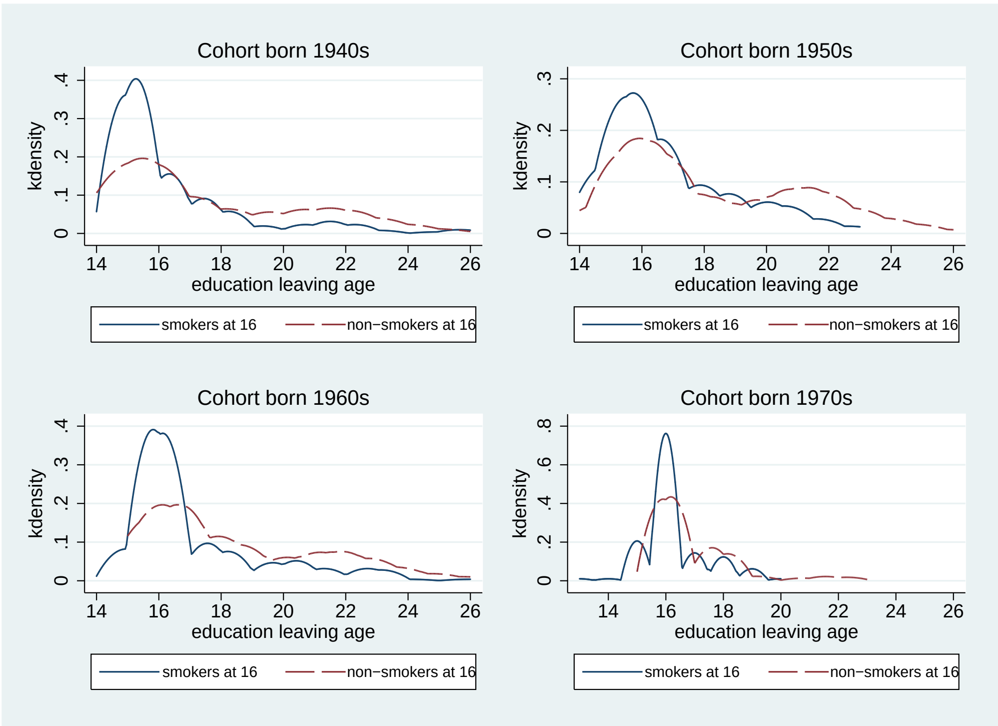

Figure 4: Education Leaving Age Density, by Smoker at 16 Status and Cohort

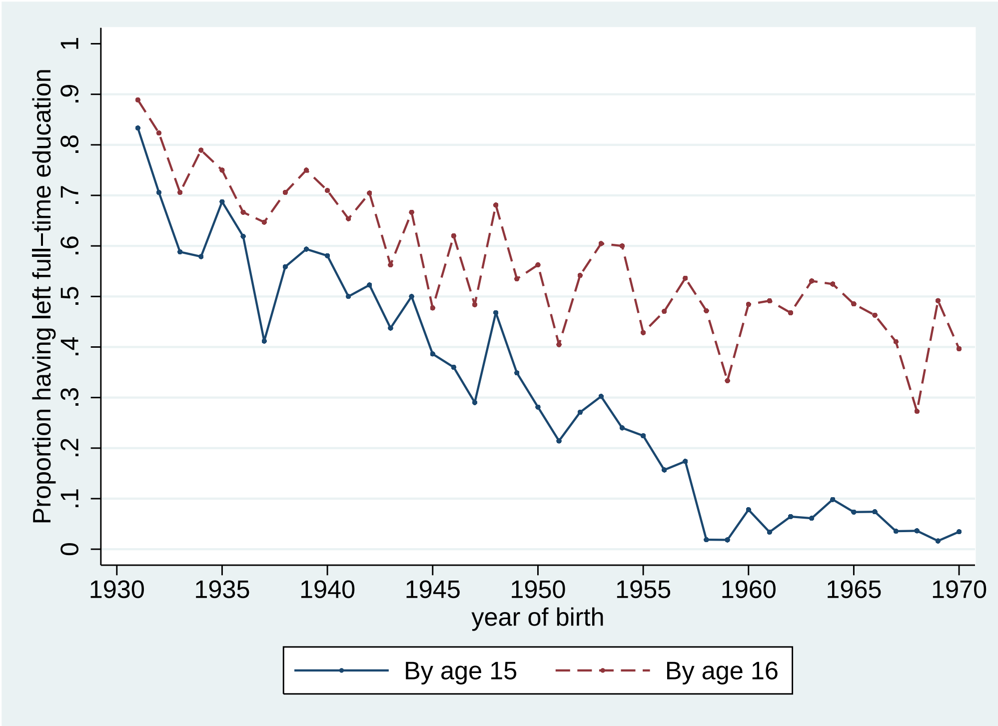

Figure 5: Proportion Left Full-Time Education at age 15 and at age 16

APPENDIX

A Estimating the HCEF using only those with 11 or more years education

Table A-1: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status

	OLS		**IV: smoker at 16**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.318***	0.331	−0.845**	0.419	6.778	2.227
years of schooling	0.038***	0.003	0.132***	0.026	— —	— —
smoker at 16 indicator	— —	— —	— —	— —	−0.833***	0.125
age	0.106***	0.005	0.104***	0.006	0.021***	0.026
age^2	−0.001***	0.000	−0.001***	0.000	0.000***	0.000
year-of-birth	−0.026**	0.009	−0.052***	0.013	0.247***	0.059
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.002***	0.000
region: North	0.061**	0.042	0.070	0.052	−0.081	0.298
region: Yorkshire	0.017	0.038	−0.020	0.049	0.440	0.285
region: North West	0.057	0.038	0.002	0.048	0.611	0.283
region: East Midlands	−0.005	0.037	−0.011	0.045	0.107	0.268
region: East Anglia	0.008	0.044	−0.015	0.055	0.308	0.361
region: South East	0.149***	0.032	0.082**	0.043	0.762***	0.223
region: South West	0.023	0.038	0.014	0.047	0.175	0.261
region: Wales	0.008	0.046	−0.017	0.053	0.241	0.311
region: Scotland	0.039	0.040	−0.032	0.052	0.799**	0.292
ethnicity: Black	0.132	0.113	0.115	0.129	0.055	0.716
ethnicity: Asian	−0.165*	0.070	−0.340***	0.107	1.733***	0.484
ethnicity: Other	−0.041	0.112	−0.279**	0.134	2.392*	1.103
father's occ class: 1	0.125***	0.031	0.036	0.045	0.909***	0.231
father's occ class: 2	0.144***	0.039	−0.049	0.071	1.935***	0.297
father's occ class: 3	0.082**	0.044	−0.036	0.060	1.162***	0.331
father's occ class: 4	0.085*	0.040	−0.020	0.057	1.006***	0.324
father's occ class: 5	0.038*	0.026	0.016	0.033	0.227***	0.200
father's occ class: 6	0.020	0.038	−0.072	0.054	0.873***	0.332
father's occ class: 7	0.107***	0.043	0.098	0.055	0.131	0.358
father's occ class: 9	0.002	0.037	0.058	0.046	−0.585***	0.252
father's occ class: 10	0.035	0.031	0.034	0.037	−0.047	0.223
mother's occ class: 1	0.019	0.056	0.007	0.073	0.098	0.454
mother's occ class: 2	−0.017	0.057	−0.111	0.073	1.033***	0.451
mother's occ class: 3	0.025	0.052	0.064	0.064	−0.367	0.427
mother's occ class: 4	0.024	0.043	0.005	0.055	0.238	0.345
mother's occ class: 5	−0.018	0.057	0.004	0.072	−0.097	0.489
mother's occ class: 6	0.003	0.044	0.036	0.055	−0.227	0.357
mother's occ class: 7	0.026	0.046	0.067	0.058	−0.472	0.354
mother's occ class: 9	−0.065	0.044	−0.011	0.055	−0.572	0.346
mother's occ class: 10	−0.015	0.036	−0.004	0.046	−0.105	0.305
'nuclear family' to 16	0.016	0.022	−0.009	0.026	0.180*	0.147
mid 1990s	−0.049***	0.010	−0.057***	0.011	0.086	0.051
late 1990s	−0.068***	0.016	−0.077***	0.018	0.111	0.088
post 2000	−0.033	0.023	−0.045*	0.027	0.144	0.136
# observations	16985		16985		16985	
# individuals	1739		1739		1739	
R^2	0.278		0.040		0.218	

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

B Estimating the HCEF using only one observation per person in the first stage

Table B-1: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status

	OLS		IV: smoker at 16		IV: first stage	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.541	0.351	−0.137	6.170
years of schooling	0.046***	0.003	0.133***	0.021	— —	— —
smoker at 16 indicator	— —	— —	— —	— —	−0.876***	0.097
age	0.099***	0.004	0.101***	0.006	−0.004	0.083
age^2	−0.001***	0.000	−0.001***	0.000	0.000	0.001
year-of-birth	−0.016***	0.007	−0.058***	0.012	0.411***	0.109
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.003***	0.001
region: North	0.047	0.038	0.061	0.044	−0.187	0.238
region: Yorkshire	0.003	0.033	0.006	0.039	0.136	0.223
region: North West	0.054	0.032	0.048	0.037	0.207	0.226
region: East Midlands	−0.010	0.032	−0.001	0.037	−0.104	0.224
region: East Anglia	0.015	0.039	0.010	0.049	0.180	0.302
region: South East	0.142***	0.028	0.102***	0.035	0.610***	0.186
region: South West	0.023	0.034	0.030	0.037	0.037	0.217
region: Wales	−0.012	0.040	−0.009	0.044	−0.042	0.270
region: Scotland	0.028	0.036	0.014	0.040	0.362	0.229
ethnicity: Black	0.114	0.105	0.096	0.166	−0.034	0.881
ethnicity: Asian	−0.136	0.071	−0.251***	0.092	1.385***	0.458
ethnicity: Other	−0.048	0.103	−0.180	0.140	1.615*	0.841
father's occ class: 1	0.116***	0.028	0.010	0.039	1.163***	0.146
father's occ class: 2	0.121***	0.038	−0.094	0.069	2.314***	0.209
father's occ class: 3	0.089**	0.043	−0.033	0.056	1.369***	0.236
father's occ class: 4	0.065*	0.036	−0.077	0.056	1.439***	0.219
father's occ class: 5	0.038*	0.023	0.007	0.027	0.359***	0.116
father's occ class: 6	0.014	0.035	−0.073	0.046	0.933***	0.189
father's occ class: 7	0.103***	0.04	0.057	0.048	0.490**	0.219
father's occ class: 9	−0.021	0.029	0.025	0.037	−0.494***	0.132
father's occ class: 10	0.029	0.027	0.007	0.030	0.152	0.126
mother's occ class: 1	0.047	0.049	0.025	0.064	0.159	0.296
mother's occ class: 2	0.015	0.054	−0.108	0.071	1.379***	0.313
mother's occ class: 3	0.056	0.048	0.032	0.063	0.179	0.312
mother's occ class: 4	0.055	0.04	0.008	0.049	0.491***	0.225
mother's occ class: 5	0.01	0.049	0.019	0.062	−0.044	0.263
mother's occ class: 6	0.025	0.04	0.030	0.049	0.031	0.228
mother's occ class: 7	0.055	0.041	0.054	0.049	−0.082	0.227
mother's occ class: 9	−0.004	0.038	0.030	0.049	−0.439***	0.201
mother's occ class: 10	0.004	0.032	−0.010	0.039	0.138	0.176
'nuclear family' to 16	0.028	0.019	−0.006	0.023	0.290	0.097
mid 1990s	−0.045***	0.009	−0.039	0.015	−0.130	0.211
late 1990s	−0.065***	0.014	−0.056	0.023	−0.144	0.347
post 2000	−0.033	0.021	−0.023	0.034	−0.055	0.533
# observations	21256		13498		1432	
# individuals	2266		1398		1432	
R^2	0.265		0.220		0.250	

F-test on exclusion of instrument from first stage: 51.50; Partial R^2 of the instrument = 0.0302

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.
IV second stage standard errors obtained by bootstrapping.

Table B-2: Human Capital Earnings Function Estimations, OLS and IV using RoSLA

Dep. Var: log hourly wage	OLS Coeff.	OLS Robust Std. Err.	IV: RoSLA Coeff.	IV: RoSLA Robust Std. Err.	IV: first stage Coeff.	IV: first stage Robust Std. Err.
constant	−0.754***	0.250	−0.533	0.377	−0.564	5.011
years of schooling	0.046***	0.003	0.143**	0.058	— —	— —
min. school LA=16	— —	— —	— —	— —	0.487***	0.153
age	0.099***	0.004	0.101***	0.006	0.020	0.070
age^2	−0.001***	0.000	−0.001***	0.000	0.000	0.001
year-of-birth	−0.016***	0.007	−0.062**	0.028	0.408***	0.091
year-of-$birth^2$	0.000***	0.000	0.001***	0.000	−0.004***	0.001
region: North	0.047	0.038	0.058	0.048	−0.132	0.192
region: Yorkshire	0.003	0.033	0.000	0.045	0.142	0.178
region: North West	0.054	0.032	0.047	0.041	0.176	0.175
region: East Midlands	−0.010	0.032	0.004	0.039	−0.139	0.174
region: East Anglia	0.015	0.039	0.007	0.054	0.121	0.243
region: South East	0.142***	0.028	0.093*	0.049	0.609***	0.141
region: South West	0.023	0.034	0.026	0.039	−0.005	0.171
region: Wales	−0.012	0.040	−0.006	0.047	−0.023	0.211
region: Scotland	0.028	0.036	0.007	0.048	0.388**	0.185
ethnicity: Black	0.114	0.105	0.102	0.181	0.123	0.788
ethnicity: Asian	−0.136	0.071	−0.270**	0.134	1.571***	0.377
ethnicity: Other	−0.048	0.103	−0.202	0.173	1.756**	0.695
father's occ class: 1	0.116***	0.028	−0.004	0.078	1.217***	0.149
father's occ class: 2	0.121***	0.038	−0.120	0.155	2.443***	0.213
father's occ class: 3	0.089**	0.043	−0.052	0.101	1.474***	0.244
father's occ class: 4	0.065*	0.036	−0.093	0.102	1.566***	0.222
father's occ class: 5	0.038*	0.023	0.001	0.035	0.359***	0.118
father's occ class: 6	0.014	0.035	−0.084	0.074	0.994***	0.196
father's occ class: 7	0.103***	0.040	0.050	0.057	0.535**	0.223
father's occ class: 9	−0.021	0.029	0.029	0.048	−0.522***	0.133
father's occ class: 10	0.029	0.027	0.003	0.034	0.210	0.128
mother's occ class: 1	0.047	0.049	0.025	0.073	0.132	0.303
mother's occ class: 2	0.015	0.054	−0.130	0.108	1.373***	0.321
mother's occ class: 3	0.056	0.048	0.030	0.068	0.150	0.317
mother's occ class: 4	0.055	0.040	−0.001	0.058	0.485**	0.234
mother's occ class: 5	0.010	0.049	0.014	0.069	−0.112	0.274
mother's occ class: 6	0.025	0.040	0.027	0.052	−0.056	0.234
mother's occ class: 7	0.055	0.041	0.059	0.054	−0.080	0.233
mother's occ class: 9	−0.004	0.038	0.037	0.061	−0.471**	0.208
mother's occ class: 10	0.004	0.032	−0.013	0.042	0.135	0.184
'nuclear family' to 16	0.028	0.019	−0.011	0.033	0.376***	0.097
mid 1990s	−0.045***	0.009	−0.039**	0.015	−0.161	0.209
late 1990s	−0.065***	0.014	−0.056**	0.023	−0.137	0.349
post 2000	−0.033	0.021	−0.024	0.034	−0.030	0.539
# observations	21256		13498		1398	
# individuals	2266		1398		1398	
R^2	0.265		0.160		0.229	

F-test on exclusion of instrument from first stage: 4.06; Partial R^2 of the instrument = 0.0029

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.
IV second stage standard errors obtained by bootstrapping.

C HCEF Estimates, OLS and IV using Fuller(1) LIML estimator

Table C-1: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status

	OLS		**IV: smoker at 16**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.607**	0.287	−0.471	1.664
years of schooling	0.046***	0.003	0.129***	0.020	— —	— —
smoker at 16 indicator	— —	— —	— —	— —	−0.876***	0.108
age	0.099***	0.004	0.094***	0.005	0.056***	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.052***	0.011	0.398***	0.041
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.003***	0.000
region: North	0.047	0.038	0.054	0.044	−0.103	0.272
region: Yorkshire	0.003	0.033	−0.022	0.041	0.331	0.253
region: North West	0.054*	0.032	0.023	0.040	0.402	0.253
region: East Midlands	−0.010	0.032	−0.005	0.038	−0.034	0.235
region: East Anglia	0.015	0.039	−0.009	0.048	0.366	0.324
region: South East	0.142***	0.028	0.082**	0.037	0.757***	0.206
region: South West	0.023	0.034	0.015	0.041	0.175	0.237
region: Wales	−0.012	0.040	−0.019	0.045	0.081	0.285
region: Scotland	0.028	0.036	−0.021	0.044	0.643**	0.262
ethnicity: Black	0.114	0.105	0.115	0.117	−0.164	0.779
ethnicity: Asian	−0.136*	0.071	−0.312***	0.105	1.965***	0.485
ethnicity: Other	−0.048	0.103	−0.234**	0.119	2.067*	1.111
father's occ class: 1	0.116***	0.028	0.020	0.041	1.122***	0.214
father's occ class: 2	0.121***	0.038	−0.076	0.065	2.268***	0.291
father's occ class: 3	0.089**	0.043	−0.043	0.058	1.499***	0.321
father's occ class: 4	0.065*	0.036	−0.053	0.051	1.320***	0.305
father's occ class: 5	0.038*	0.023	0.011	0.028	0.335**	0.170
father's occ class: 6	0.014	0.035	−0.074	0.048	0.991***	0.305
father's occ class: 7	0.103***	0.040	0.066	0.049	0.467	0.330
father's occ class: 9	−0.021	0.029	0.028	0.035	−0.551***	0.197
father's occ class: 10	0.029	0.027	0.027	0.030	−0.012	0.186
mother's occ class: 1	0.047	0.049	0.035	0.061	0.112	0.411
mother's occ class: 2	0.015	0.054	−0.103	0.070	1.433***	0.439
mother's occ class: 3	0.056	0.048	0.053	0.057	0.046	0.387
mother's occ class: 4	0.055	0.040	0.015	0.048	0.485	0.307
mother's occ class: 5	0.010	0.049	0.031	0.058	−0.117	0.417
mother's occ class: 6	0.025	0.040	0.029	0.045	0.054	0.311
mother's occ class: 7	0.055	0.041	0.057	0.048	−0.083	0.312
mother's occ class: 9	−0.004	0.038	0.034	0.044	−0.461	0.284
mother's occ class: 10	0.004	0.032	−0.006	0.036	0.115	0.253
'nuclear family' to 16	0.028	0.019	0.001	0.022	0.247*	0.136
mid 1990s	−0.045***	0.009	−0.050***	0.010	0.067	0.046
late 1990s	−0.065***	0.014	−0.070***	0.016	0.080	0.081
post 2000	−0.033	0.021	−0.040*	0.023	0.108	0.126
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.073		0.246	

F-test on exclusion of smoking at 16 from first stage: 66.17; Partial R^2 of instrument = 0.0289

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table C-2: Human Capital Earnings Function Estimations, OLS and IV using RoSLA

	OLS		IV: RoSLA		IV: first stage	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.656**	0.279	−1.459	1.681
years of schooling	0.046***	0.003	0.101**	0.051	— —	— —
min. school LA=16	— —	— —	— —	— —	0.564***	0.206
age	0.099***	0.004	0.095***	0.005	0.056**	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.040*	0.023	0.427***	0.041
year-of-birth^2	0.000***	0.000	0.000**	0.000	−0.004***	0.000
region: North	0.047	0.038	0.051	0.041	−0.080	0.272
region: Yorkshire	0.003	0.033	−0.013	0.040	0.320	0.256
region: North West	0.054*	0.032	0.033	0.041	0.386	0.255
region: East Midlands	−0.010	0.032	−0.007	0.035	−0.035	0.234
region: East Anglia	0.015	0.039	−0.001	0.046	0.324	0.327
region: South East	0.142***	0.028	0.102**	0.051	0.741***	0.208
region: South West	0.023	0.034	0.017	0.038	0.114	0.240
region: Wales	−0.012	0.040	−0.017	0.042	0.093	0.290
region: Scotland	0.028	0.036	−0.005	0.050	0.658**	0.266
ethnicity: Black	0.114	0.105	0.114	0.110	0.037	0.746
ethnicity: Asian	−0.136*	0.071	−0.254*	0.138	2.146***	0.515
ethnicity: Other	−0.048	0.103	−0.172	0.151	2.214**	1.074
father's occ class: 1	0.116***	0.028	0.052	0.068	1.162***	0.216
father's occ class: 2	0.121***	0.038	−0.011	0.127	2.404***	0.298
father's occ class: 3	0.089**	0.043	0.001	0.092	1.585***	0.333
father's occ class: 4	0.065*	0.036	−0.014	0.082	1.440***	0.308
father's occ class: 5	0.038*	0.023	0.020	0.029	0.322*	0.172
father's occ class: 6	0.014	0.035	−0.045	0.063	1.046***	0.313
father's occ class: 7	0.103***	0.040	0.078	0.049	0.484	0.339
father's occ class: 9	−0.021	0.029	0.012	0.044	−0.592***	0.196
father's occ class: 10	0.029	0.027	0.028	0.028	0.043	0.186
mother's occ class: 1	0.047	0.049	0.039	0.056	0.107	0.426
mother's occ class: 2	0.015	0.054	−0.064	0.093	1.378***	0.454
mother's occ class: 3	0.056	0.048	0.054	0.052	0.007	0.395
mother's occ class: 4	0.055	0.040	0.028	0.050	0.453	0.317
mother's occ class: 5	0.010	0.049	0.024	0.054	−0.240	0.430
mother's occ class: 6	0.025	0.040	0.027	0.042	−0.070	0.322
mother's occ class: 7	0.055	0.041	0.057	0.044	−0.053	0.324
mother's occ class: 9	−0.004	0.038	0.022	0.047	−0.491*	0.293
mother's occ class: 10	0.004	0.032	−0.003	0.034	0.103	0.264
'nuclear family' to 16	0.028	0.019	0.010	0.026	0.330**	0.137
mid 1990s	−0.045***	0.009	−0.048***	0.010	0.063	0.047
late 1990s	−0.065***	0.014	−0.068***	0.015	0.075	0.083
post 2000	−0.033	0.021	−0.038*	0.022	0.094	0.129
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.178		0.227	
F-test on exclusion of min. sch. LA=16 from first stage: 7.49; Partial R^2 of instrument = 0.0044						

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table C-3: Human Capital Earnings Function Estimations, OLS and IV using Smoker at 16 Status and RoSLA

	OLS		**IV: both**		**IV: first stage**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.754***	0.250	−0.613**	0.283	−0.157	1.663
years of schooling	0.046***	0.003	0.125***	0.019	— —	— —
smoker at age 16	— —	— —	— —	— —	−0.874***	0.107
min. school LA=16	— —	— —	— —	— —	0.556***	0.202
age	0.099***	0.004	0.094***	0.005	0.054**	0.022
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.016**	0.007	−0.050***	0.011	0.399***	0.041
year-of-birth^2	0.000***	0.000	0.001***	0.000	−0.004***	0.000
region: North	0.047	0.038	0.053	0.044	−0.097	0.272
region: Yorkshire	0.003	0.033	−0.021	0.040	0.347	0.253
region: North West	0.054*	0.032	0.024	0.039	0.409	0.253
region: East Midlands	−0.010	0.032	−0.006	0.037	−0.014	0.235
region: East Anglia	0.015	0.039	−0.008	0.048	0.398	0.325
region: South East	0.142***	0.028	0.084**	0.036	0.767***	0.207
region: South West	0.023	0.034	0.015	0.040	0.192	0.236
region: Wales	−0.012	0.040	−0.019	0.045	0.082	0.286
region: Scotland	0.028	0.036	−0.019	0.043	0.705***	0.263
ethnicity: Black	0.114	0.105	0.115	0.116	−0.114	0.788
ethnicity: Asian	−0.136*	0.071	−0.305***	0.103	1.975***	0.493
ethnicity: Other	−0.048	0.103	−0.226*	0.116	2.021*	1.080
father's occ class: 1	0.116***	0.028	0.024	0.040	1.118***	0.213
father's occ class: 2	0.121***	0.038	−0.069	0.062	2.271***	0.290
father's occ class: 3	0.089**	0.043	−0.038	0.056	1.485***	0.319
father's occ class: 4	0.065*	0.036	−0.049	0.050	1.324***	0.303
father's occ class: 5	0.038*	0.023	0.012	0.027	0.322*	0.170
father's occ class: 6	0.014	0.035	−0.071	0.046	0.968***	0.303
father's occ class: 7	0.103***	0.040	0.067	0.048	0.501	0.330
father's occ class: 9	−0.021	0.029	0.026	0.034	−0.542***	0.194
father's occ class: 10	0.029	0.027	0.027	0.030	0.000	0.185
mother's occ class: 1	0.047	0.049	0.036	0.060	0.079	0.414
mother's occ class: 2	0.015	0.054	−0.098	0.069	1.379***	0.442
mother's occ class: 3	0.056	0.048	0.053	0.057	0.018	0.388
mother's occ class: 4	0.055	0.040	0.016	0.048	0.451	0.310
mother's occ class: 5	0.010	0.049	0.030	0.057	−0.104	0.414
mother's occ class: 6	0.025	0.040	0.028	0.045	0.030	0.313
mother's occ class: 7	0.055	0.041	0.057	0.047	−0.111	0.316
mother's occ class: 9	−0.004	0.038	0.033	0.044	−0.488*	0.285
mother's occ class: 10	0.004	0.032	−0.005	0.036	0.099	0.256
'nuclear family' to 16	0.028	0.019	0.002	0.022	0.251*	0.136
mid 1990s	−0.045***	0.009	−0.049***	0.010	0.073	0.046
late 1990s	−0.065***	0.014	−0.070***	0.016	0.092	0.081
post 2000	−0.033	0.021	−0.039*	0.023	0.120	0.126
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.265		0.087		0.250	

F-test on exclusion of both instruments from first stage: 36.83; Partial R^2 of instrument = 0.0332

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table C-4: Human Capital Earnings Function Estimations, OLS and IV using RoSLA, basic specification

	OLS		IV: RoSLA		IV: first stage	
		Robust		Robust		Robust
Dep. Var: log hourly wage	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**
constant	−0.849∗∗∗	0.247	−0.669∗∗	0.302	−3.375∗	1.727
years of schooling	0.052∗∗∗	0.003	0.100∗∗	0.041	— —	— —
min. school LA=16	— —	— —	— —	— —	0.691∗∗∗	0.219
age	0.098∗∗∗	0.004	0.095∗∗∗	0.005	0.068∗∗∗	0.023
age^2	−0.001∗∗∗	0.000	−0.001∗∗∗	0.000	−0.001∗∗∗	0.000
year-of-birth	−0.014∗∗	0.007	−0.037∗	0.022	0.501∗∗∗	0.042
year-of-birth^2	0.000∗∗∗	0.000	0.000∗∗	0.000	−0.004∗∗∗	0.000
region: North	0.041	0.038	0.046	0.041	−0.089	0.286
region: Yorkshire	−0.003	0.033	−0.013	0.038	0.246	0.268
region: North West	0.050	0.033	0.030	0.040	0.418	0.273
region: East Midlands	−0.016	0.032	−0.009	0.035	−0.129	0.258
region: East Anglia	0.010	0.040	−0.001	0.045	0.277	0.343
region: South East	0.143∗∗∗	0.028	0.100∗	0.052	0.936∗∗∗	0.223
region: South West	0.023	0.034	0.014	0.038	0.202	0.256
region: Wales	−0.018	0.040	−0.019	0.042	0.022	0.314
region: Scotland	0.020	0.036	−0.010	0.047	0.710∗∗	0.288
ethnicity: Black	0.117	0.093	0.108	0.105	0.251	0.700
ethnicity: Asian	−0.150∗∗	0.070	−0.247∗∗	0.119	2.075∗∗∗	0.560
ethnicity: Other	−0.042	0.095	−0.166	0.148	2.566∗∗∗	0.939
mid 1990s	−0.047∗∗∗	0.009	−0.048∗∗∗	0.010	0.036	0.050
late 1990s	−0.068∗∗∗	0.014	−0.068∗∗∗	0.015	0.013	0.089
post 2000	−0.038∗	0.021	−0.037∗	0.022	−0.004	0.137
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.251		0.177		0.113	

F-test on exclusion of min. sch. LA=16 from first stage: 9.98; Partial R^2 of instrument = 0.0058

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.

D In Support of Early Smoking as a Valid Instrument

Table D-1: Reduced Form for Log Hourly Wage: Smoker at 16 instrument, RoSLA instrument and Both instruments

	Smoker at 16		**Min. School LA=16**		**Both**	
Dep. Var: log hourly wage	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**	**Coeff.**	**Robust Std. Err.**
constant	−0.668**	0.261	−0.803***	0.263	−0.636**	0.261
smoker at 16 indicator	−0.113***	0.016	— —	— —	−0.113***	0.016
min. school LA=16	— —	— —	0.058**	0.028	0.057**	0.027
age	0.101***	0.004	0.101***	0.004	0.101***	0.004
age^2	−0.001***	0.000	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.000	0.007	0.003	0.007	−0.000	0.007
year-of-birth^2	0.000**	0.000	0.000	0.000	0.000*	0.000
region: North	0.040	0.040	0.043	0.040	0.041	0.039
region: Yorkshire	0.021	0.035	0.019	0.035	0.023	0.035
region: North West	0.075**	0.034	0.072**	0.034	0.075**	0.034
region: East Midlands	−0.010	0.033	−0.010	0.034	−0.008	0.033
region: East Anglia	0.038	0.040	0.032	0.041	0.041	0.040
region: South East	0.179***	0.029	0.177***	0.029	0.180***	0.029
region: South West	0.037	0.034	0.029	0.034	0.039	0.034
region: Wales	−0.009	0.042	−0.007	0.043	−0.009	0.042
region: Scotland	0.062*	0.037	0.062*	0.037	0.068*	0.037
ethnicity: Black	0.094	0.115	0.118	0.113	0.099	0.115
ethnicity: Asian	−0.059	0.066	−0.036	0.066	−0.058	0.066
ethnicity: Other	0.032	0.132	0.052	0.129	0.028	0.130
father's occ class: 1	0.165***	0.029	0.170***	0.029	0.164***	0.029
father's occ class: 2	0.215***	0.040	0.233***	0.041	0.216***	0.041
father's occ class: 3	0.150***	0.046	0.162***	0.047	0.149***	0.047
father's occ class: 4	0.117***	0.040	0.132***	0.041	0.117***	0.040
father's occ class: 5	0.054**	0.024	0.053**	0.024	0.053**	0.024
father's occ class: 6	0.054	0.039	0.061	0.038	0.051	0.038
father's occ class: 7	0.126***	0.042	0.127***	0.043	0.129***	0.042
father's occ class: 9	−0.043	0.031	−0.048	0.031	−0.042	0.031
father's occ class: 10	0.025	0.028	0.032	0.028	0.026	0.028
mother's occ class: 1	0.050	0.052	0.050	0.053	0.046	0.052
mother's occ class: 2	0.082	0.060	0.076	0.060	0.076	0.061
mother's occ class: 3	0.059	0.051	0.055	0.052	0.056	0.051
mother's occ class: 4	0.077*	0.042	0.074*	0.043	0.073*	0.042
mother's occ class: 5	0.016	0.052	−0.000	0.054	0.018	0.052
mother's occ class: 6	0.035	0.043	0.020	0.044	0.033	0.043
mother's occ class: 7	0.047	0.043	0.051	0.044	0.044	0.043
mother's occ class: 9	−0.025	0.040	−0.028	0.042	−0.028	0.040
mother's occ class: 10	0.009	0.034	0.008	0.036	0.007	0.034
'nuclear family' to 16	0.033	0.021	0.044**	0.021	0.033	0.021
mid 1990s	−0.041***	0.009	−0.042***	0.009	−0.040***	0.009
late 1990s	−0.060***	0.015	−0.061***	0.015	−0.058***	0.015
post 2000	−0.026	0.021	−0.028	0.022	−0.025	0.021
# observations	21256		21256		21256	
# individuals	2266		2266		2266	
R^2	0.217		0.205		0.218	

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.

Table D-2: Regression of Residuals from Structural Equation when using the RoSLA IV on the Smoker at 16 indicator

Dep. Var: $\hat{\epsilon}_i$	**Coeff.**	**Robust Std. Err.**
constant	0.007	0.009
smoker at 16 indicator	-0.022	0.016
# observations	21256	
# individuals	2266	
R^2	0.001	

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.

Construction of the dependent variable $\hat{\epsilon_i}$:
Log hourly wage estimated by IV regression, first stage equation (10) estimated:
$\hat{S}_i = X_i\tilde{\gamma} + Z_i\tilde{\pi} + u_i$ where Z_i is the min. school LA=16 indicator.
Second stage equation (11) estimated:
$log(w_i) = X_i\varphi + \hat{S}_i\beta_i + \epsilon_i$. The residuals are recovered and these $\hat{\epsilon_i}$ are the dependent variable.

Table D-3: HCEF using RoSLA IV, including Smoker at 16 Status as an Explanatory Variable

	IV: RoSLA		IV: first stage	
Dep. Var: log hourly wage	Coeff.	Robust Std. Err.	Coeff.	Robust Std. Err.
constant	−0.620**	0.267	−0.157	1.663
years of schooling	0.102*	0.052	— —	— —
min. school LA=16	— —	— —	0.556***	0.202
smoker at 16 indicator	−0.024	0.049	−0.874***	0.107
age	0.095***	0.005	0.054**	0.022
age^2	−0.001***	0.000	−0.001***	0.000
year-of-birth	−0.041*	0.022	0.399***	0.041
year-of-birth^2	0.000**	0.000	−0.004***	0.000
region: North	0.051	0.041	−0.097	0.272
region: Yorkshire	−0.013	0.041	0.347	0.253
region: North West	0.034	0.042	0.409	0.253
region: East Midlands	−0.006	0.035	−0.014	0.235
region: East Anglia	0.001	0.048	0.398	0.325
region: South East	0.102*	0.053	0.767***	0.207
region: South West	0.019	0.038	0.192	0.236
region: Wales	−0.017	0.042	0.082	0.286
region: Scotland	−0.004	0.052	0.705***	0.263
ethnicity: Black	0.110	0.109	−0.114	0.788
ethnicity: Asian	−0.259*	0.133	1.975***	0.493
ethnicity: Other	−0.178	0.146	2.021*	1.080
father's occ class: 1	0.051	0.068	1.118***	0.213
father's occ class: 2	−0.015	0.123	2.271***	0.290
father's occ class: 3	−0.002	0.089	1.485***	0.319
father's occ class: 4	−0.018	0.078	1.324***	0.303
father's occ class: 5	0.020	0.029	0.322*	0.170
father's occ class: 6	−0.047	0.061	0.968***	0.303
father's occ class: 7	0.078	0.050	0.501	0.330
father's occ class: 9	0.013	0.042	−0.542***	0.194
father's occ class: 10	0.026	0.028	0.000	0.185
mother's occ class: 1	0.038	0.056	0.079	0.414
mother's occ class: 2	−0.064	0.095	1.379***	0.442
mother's occ class: 3	0.054	0.052	0.018	0.388
mother's occ class: 4	0.028	0.050	0.451	0.310
mother's occ class: 5	0.028	0.053	−0.104	0.414
mother's occ class: 6	0.030	0.042	0.030	0.313
mother's occ class: 7	0.055	0.044	−0.111	0.316
mother's occ class: 9	0.022	0.047	−0.488*	0.285
mother's occ class: 10	−0.003	0.034	0.099	0.256
'nuclear family' to 16	0.008	0.023	0.251*	0.136
mid 1990s	−0.048***	0.010	0.073	0.046
late 1990s	−0.068***	0.016	0.092	0.081
post 2000	−0.037	0.023	0.120	0.126
# observations	21256		21256	
# individuals	2266		2266	
R^2	0.178		0.250	

Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level
Standard errors are clustered at the level of the individual and robust.
'nuclear family' to 16 means lived with both natural parents from birth to age 16.
Reference categories: West Midlands, white, did not live with both natural parents to 16, father/mother occupational class 'plant/machine operative'. Occupational Class dummies: (1) management, (2) professional, (3) associate professional/technical, (4) clerical/secretarial, (5) craft and related, (6) personal/protective services, (7) sales, (9) other, (10) self-emp/unemp.