IFS

An ESRC Research Centre

QUALIFICATIONS AND EARNINGS IN BRITAIN: HOW RELIABLE ARE CONVENTIONAL OLS ESTIMATES OF THE RETURNS TO EDUCATION?

Lorraine Dearden

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

THE INSTITUTE FOR FISCAL STUDIES
Working Paper Series No. W99/7

Qualifications and Earnings in Britain: How reliable are conventional OLS estimates of the returns to education?*

Lorraine Dearden
Institute for Fiscal Studies
7 Ridgmount St
London WC1E 7AE
Email: ldearden@ifs.org.uk

*This is a substantially revised version of Dearden [13]. I would like to thank Orley Ashenfelter, Richard Blundell, Richard Freeman, Colm Harmon, Steve Machin, Costas Meghir, Hessel Oosterbeek, Hans van Ophem, Arthur van Soest, Frank Vella and Ian Walker for comments and suggestions on earlier versions of this work. The author also received valuable comments from participants at the Tinbergen Institute Workshop on the "Economic Returns to Education: New Evidence" held in Amsterdam on October 10-11 1997 and seminar participants at the London School of Economics and Warwick University. This work was funded by the Leverhulme Trust, Grant No. F386/G and the ESRC Research Center for Microeconomic Analysis at the IFS. The usual disclaimer applies.

Executive Summary

The paper estimates the returns to education for a cohort of individuals born in Britain in March 1958 who have been followed since birth until the age of 33. The data used has a wealth of information on family background including parental education, social class and interest shown in the child's education as well as measures of ability. The nature of our data allows us to directly assess the relative importance of omitted ability and family background bias as well as biases arising from measurement error in education qualification variables which have been found to be important in other studies. The paper also looks at possible biases arising from compositional differences between individuals in work and those out of work. This 'composition bias' arising from self-selection into employment is generally ignored in the returns to schooling literature and is why most studies focus only on men (for whom it is assumed this is much less of a problem). The paper also examines whether there is evidence of heterogeneity in the returns to education as well as the impact of education on gender wage differentials.

The paper finds that conventional OLS estimates, which assume that education is exogenous, are reasonable estimates of the true causal impact of education on wages. In the UK it would appear that the effects of measurement error bias and composition bias directly offset the countervailing effect of unobserved ability and family background bias for most qualifications. The results from the paper suggest that conventional OLS estimates of the returns to education can generally be relied upon for policy decisions.

The paper also finds evidence of heterogeneity in the returns to education in Britain. The results from the paper suggest that individuals undertaking schooling involving some sort of formal qualification have significantly larger rates of return than individuals who complete the same number of years of schooling but who obtain no formal qualifications. There is also some evidence that individuals with lower tastes for education, have significantly higher marginal returns to certain education qualifications. We also find that post-school qualifications, particularly degree qualifications, play an important role in reducing gender wage differentials.

JEL Classification: J31, I21, J24
Keywords: Education, Qualifications, Earnings

1. Introduction

Estimates of the returns to education can be upward or downward biased if no account is taken of the fact that education is not randomly determined. Educational attainment depends on individual choices, attributes and circumstances and if we do not control for these factors, then the measured differences in the wages of individuals with different levels of education may over- or under- estimate the true causal effect of education on wage outcomes. These biases arise because of correlation between unobserved individual attributes which determine individual's education decisions, as well as employment and wage outcomes. They can also arise if education is measured with error.

A lot of the recent literature on the returns to education has focused on devising ways of correcting for these potential biases using a number of different techniques including proxy, fixed effect and/or IV methods. Proxy methods rely on having extremely good data with good proxies of innate ability (see for example Blackburn and Neumark [5]). This kind of data is rarely available. Fixed effect methods typical involve twin studies which rely on the fact that the important unobservable individual characteristics are identical among twins, and hence an unbiased estimate of the return to education can be obtained by exploiting the differences between twin's levels of schooling and earnings (i.e. by using fixed effect methods). The fixed effect estimator, however, has the disadvantage of introducing far greater measurement error bias (see for example Ashenfelter and Krueger[3]). The IV methods generally rely on some exogenous change or intervention which affects educational choices but not earnings, controlling for education and use this as an instrument for education. The UK study by Harmon and Walker[18], for example, used changes in the compulsory school leaving age in England and Wales in 1947 and 1974 as an instrument for education.

However, for a large number of data sets which cover a nationally representative sample of the population, these techniques are generally not available and the question remains - how reliable are conventional OLS estimates of the returns to education and more importantly, can they be relied upon for policy decisions? This is the issue addressed in this paper.

It is commonly assumed that the most important unobserved component is unobserved ability and that conventional OLS estimates of the returns to education *overstate* the true returns because of this "omitted ability bias". This arises because the estimation procedure is unable to separate the contribution of unobserved ability to productivity from that made by education and ascribes it all to education. A similar problem may arise with regard to missing family background information.

A number of recent empirical studies looking at this question such as Butcher and Case[9], Ashenfelter and Krueger[3], Card[10], and Harmon and Walker[18] and [19], have found evidence that conventional OLS estimates *understate* the returns to education, once account is taken of the correlation between unobserved components of education and wages. This can arise if education is measured with error. As Card [11] points out, however, it may also arise if the IV estimation procedure being used relies on "interventions" that affect the schooling choices of children from relatively disadvantaged family backgrounds with high discount rates rather than low ability, as their marginal return to schooling will exceed the average return to schooling for the population as a whole. Thus if there is heterogeneity in the returns to schooling, then these IV studies will provide inconsistent estimates of the average return for the population as a whole (see for example Imbens and Angrist[1]). Of course, if the instruments used are not truly exogenous, then this may also bias IV estimates upwards (see Bound et. al. [8]).

A further issue, which is generally ignored in the returns to schooling literature, is the effect of self-selection into employment. If this occurs, then estimates of the returns to education which do not take this selection process into account, may also be biased. This arises because the characteristics of those in work and out of work will differ and this potential 'composition bias' needs to be taken into account to produce consistent estimates of the returns to education. This is

generally thought to be particularly important for women among whom there is a much larger proportion of non-workers. This is why a majority of studies looking at the returns to education only look at men and ignore problems associated with self-selection. We specifically look at this issue in the paper for *both* men and women.

The reason we can attempt to look at all these issues is because we have an extremely rich British panel data set, the National Child Development Survey (NCDS). The NCDS survey is a continuing longitudinal survey of persons living in Great Britain who were born between 3 and 9 March 1958. Our data has a wealth of family background information which has not generally been available in previous studies looking at the returns to schooling. This includes variables which measure parents' education, social class and interest in the child's education (as assessed by the child's teacher), as well as the families financial circumstances and composition. We also utilise the results of ability tests administered at the age of 7. The importance of controlling for these typically unobserved characteristics can therefore be directly assessed.

In this paper we concentrate solely on education undertaken before individuals entered the labour market[1]. We directly assess the relative importance of ability and family background bias, measurement error bias and composition bias. The methodological approach used to deal with 'ability bias' essentially involves using proxy or matching methods[2]. The paper, however, also uses IV methods to deal with possible measurement error in schooling variables which has shown to be important in studies such as Ashenfelter and Krueger [3]. We do not use IV methods when looking at the importance of biases arising from compositional differences between individuals in work and those out of work. This is because we do not have suitable instruments. Instead, we look at the sensitivity of our

[1] The returns to subsequent education and training are the focus of the paper by Blundell, Dearden and Meghir [7].

[2] This type of approach has also been used in related papers using the NCDS on the returns to higher education (see Blundell, Dearden, Goodman and Reed [6]) and the impact of school quality on education and earnings (see Dearden, Ferrier and Meghir [14]).

estimated returns under a number of different assumptions regarding self-selection into employment. We conclude by looking at whether there is any evidence of heterogeneity in the returns to different education qualifications in Britain and the relationship between qualifications and gender wage differentials.

The paper finds that in the UK, if workers are a representative sample of the entire population (i.e. there is no composition bias), when we control for the effects of omitted ability and family background bias, and measurement error bias, our corrected estimates of the returns to qualifications are not statistically different from conventional OLS estimates for all qualifications except (possibly) the highest two schooling qualifications (5 or more O levels and A levels). For these qualifications there is some evidence that the effect of omitted ability and family background outweighs biases arising from measurement error in qualifications. We also find that measurement error problems are much more serious the longer the elapsed time between completing the qualification and being surveyed. If, however, individuals self-select into employment by comparative advantage, then the estimated returns to qualifications which also have a positive effect on the participation decision, will be *downward* biased. We find that self-selection into employment is potentially important for *both* men and women. When we consider the combined effects of measurement error bias, ability bias *and* composition bias, we find that the conventional OLS estimate, which ignore all these factors, appear to be a good approximation of the true causal effect of education on wage outcomes.

The paper also finds that there is considerable heterogeneity in the estimated return to an additional year of education, and that it is crucial in the UK context, to look at qualifications and not years of education. Our results suggest that individuals undertaking school qualifications have significantly larger returns to education than individuals who have undertaken the same number of years of education, but who have obtained no formal school qualifications. Individuals who

leave school at 16 with 5 or more O levels[3], receive on average between 20 to 26 per cent higher wages at 33, compared to individuals who left school at the same time but with no school qualifications. There is also some evidence that individuals with lower tastes for education, have significantly higher marginal returns to certain educational qualifications. We also find that obtaining post-school educational qualifications, particularly degree qualifications, plays an important role in reducing gender wage differentials.

The structure of the paper is as follows. In section 2 we look more closely at the NCDS data used in our analysis. In section 3 we outline our methodological approach and in section 4 the results of our analysis are discussed. Conclusions are offered in section 5.

2. The NCDS Data

2.1. Introduction

The National Child Development Survey (NCDS) is a continuing longitudinal survey of persons living in Great Britain who were born between 3 and 9 March, 1958. There have been 5 waves of the NCDS, the last survey having been undertaken in 1991 when the cohort members were 33 years of age. In this paper we focus on a sample of individuals who participated in waves 4 and 5 of the NCDS in 1981 and 1991 respectively.

2.2. Variables used in the analysis

2.2.1. Education and Ability Variables

The NCDS has information on the individuals highest school qualification and post-school qualification as at 1981 which we view as "education" or "schooling". It also has the results from reading and mathematical ability tests undertaken

[3] This is what is generally required in order to undertake A level qualifications at the age of 18. CSEs and O level qualifications (which are now called GCSEs) are all intermediate level schooling qualifications generally taken at the age of 16.

when the person was seven, eleven and sixteen as well as the information on the years of full-time education.

In this paper we concentrate on the returns to qualifications[4]. The NCDS has information on the persons highest school and post-school qualification as at 1981. We construct this measure using information from NCDS4 and a 1978 exams file obtained from the individual's school which contains detailed high school examination results. We use this information to identify a person's highest school and post-school educational qualification[5]. Both our school and post school qualification measures are ordered and a full description of these variables is contained in Table 2.1.

Most individuals who have no school qualifications or whose highest school qualification was a CSE or O level qualification will have left school at the minimum school leaving age of 16[6]. Those who completed at least on A level will generally have completed school at the age of 18. Hence the difference in years of full-time schooling between individuals with A level qualifications and those with other school qualification is on average 2 years. If we turn to post-school qualifications, the average difference in years of full-time post-school education between those individuals with no post-school qualifications and those with lower and middle vocational qualifications will be on average somewhere between 6 months and a year and those with higher vocational qualifications around 2 years. Finally individuals with degrees will have on average around 3 years more education than

[4]The returns to years of full-time education has been considered in Dearden [13]. This paper, however, does not consider the effects of composition bias.

[5]In earlier versions of this paper we combined information on a person's highest school and post-school qualification into one variable which identified their highest qualification. While this is methodologically convenient, it clouds the important differences in returns to post-school qualifications for individuals with different schooling backgrounds. For example the approach involves classifying individual's with sub-degree level qualifications (classified as higher vocational) as having identical qualifications regardless of their schooling background. Blundell, Dearden, Goodman and Reed [7] found that for this group, it was very important to distinguish between individuals with A level qualifications and those with lower schooling qualifications.

[6]These qualifications are now called GCSEs.

Table 2.1: Description of Highest Education Qualification Variables

Variable	Description
Highest Post School Qualification in 1981:	
Degree	University or CNAA first degree, CNAA Post-graduate Diploma, or University or CNAA Higher Degree.
Higher Vocational	Highest Vocational: Full professional qualification, part of a professional qualification, Polytechnic Diploma or Certificate (not CNAA validated), University or CNAA Diploma or Certificate, Nursing qualification including nursery qualification, non-graduate teaching qualifications, Higher National Certificate (HNC) or Diploma (HND), BEC/TEC Higher Certificate or Higher Diploma, City and Guilds Full Technological Certificate.
Middle Vocational	Middle Vocational: City and Guilds Advanced or Final, Ordinary National Certificate (ONC) or Diploma (OND), BEC/TEC National, General or Ordinary.
Lower Vocational	Lower Vocational: City and Guilds Craft or Ordinary, a Royal Society of Arts (RSA) awards, stage 1, 2 or 3, other commercial or clerical qualifications and all other courses leading to some sort of qualification which are not identified above including miscellaneous apprenticeship qualifications.
None	No post school qualifications.
Highest School Qualification in 1981:	
A Levels	At least one GCE A Level, Scottish Leaving Certificate (SLC), Scottish Certificate of Education (SCE), Scottish University Preliminary Examination (SUPE) at Higher Grade, Certificate of Sixth Year Studies.
5+ O Levels	At least five GCE O Level passes or grades A–C, or CSE Grade 1 or equivalent.
O Levels	Between one and four GCE O Level passes or grades A–C, or CSE Grade 1 or equivalent.
CSEs	At least one CSE grade 2–5 or equivalent.
None	No school qualifications including individuals with no formal schooling.

individuals with only A levels and no other post-school qualifications[7].

Similar information on an individual's highest education qualifications (as at March 1981 and 1991) can also be obtained from the 1991 NCDS5 survey. This allows us to also construct another set of variables which identify the person's highest school and post-school qualification as at March 1981 based solely on responses in the 1991 survey. It is these variables which we exploit in looking at possible measurement error in our qualification variables. This is discussed in more detail below.

We also construct measures of reading and mathematics ability which are based on ability tests undertaken when the child was aged seven. We use the seven year old test results, as these are much less likely to be affected by knowledge gained at school. From these verbal and mathematics ability tests we construct 10 dummy variables which rank the individual's results in each of the tests by quintiles[8].

2.2.2. School and Family Background Variables

We use data from the first wave of the NCDS to construct dummy variables identifying the teacher's assessment of the interest shown by the mother and father in the education of the child at that age. From the third wave of the survey we construct dummy variables identifying the type of school the individual attended in 1974 (government comprehensive, government grammar (selective), government secondary modern, private or special). We ignore other school quality variables which are available in the data such as teacher/pupil ratios. The effects of these other measured school quality variables on education and earnings was found to be small in the paper by Dearden, Ferrier and Meghir [14].

We also use the data from the third wave of the survey to construct variables identifying fathers' social class; the years of full-time education undertaken by the

[7]For this cohort, A levels were generally a pre-requisite for entry into degree courses.

[8]We choose quintiles, as 20 per cent of individuals in 1965 when the tests were undertaken obtained maximum marks in the reading ability test. The quintiles refer to quintiles at the time the test was taken and not in our final sample.

child's mother and father at that age[9]; variables identifying individuals who had no father figure at that age; whether the family was experiencing financial difficulties in 1969 or 1974[10]; the number of siblings and older siblings the respondent had; and finally whether the respondent had only brothers or sisters[11].

2.2.3. Wage, Demographic, Employer and Regional Variables

We use data from the NCDS5 survey to construct gross hourly wage data. We limit our employed sample to individuals who are employees at the time of the 1991 survey[12]. Since all individuals in the sample are born in the same week of March 1958 age (or potential labour market experience) is controlled for in all of models. We also have information on actual labour market experience. We also use the NCDS5 data to identify whether the individual was working in a large firm (more than 500 workers), in the private sector and whether they were a member of a trade union in 1991. We use both the NCDS3 and NCDS5 surveys to construct 11 regional dummy variables for both 1974 and 1991. We also use 1971 local area Census information to control for local authority[13] demographics. This Census information has been mapped into the local authority in which the individual lived in 1974. The variables we use in the paper measure the proportion of households in the local authority: with an unemployed or sick head of household, with a head of household in the top social class (professional/managerial) and bottom social

[9]The variable measures the years of full-time education undertaken by the child's mother and father figure at the age of 16. This is constructed from a variable which identifies the age at which the parent's left full-time education, assuming they started school at the age of five. If there is no mother or father figure or parental education is missing, then parental years of education are set to zero. We separately identify individual's who have no father figures as well as those with missing parental education information.

[10]Following Micklewright[25], this identifies individual's who received free school meals in 1969 or 1974 or whose parents were seriously troubled financially in the year prior to the 1969 or 1974 survey.

[11]Dearden [12] looked at the effects of various family composition variables on education and earnings and found that these four composition variables were the most important.

[12]We exclude self-employed individuals because of the unavailability of reliable wage data.

[13]There were approximately 140 local authorities in Britain in 1974. Local authorities are responsible for schools in their area, although they received the majority of their funding for schools from central government.

class (unskilled), and who are council tenants and owner-occupiers.

2.2.4. The Final Sample

We drop individuals who are employed and have missing observations on wages and those for whom we don't have details of highest school and post-school qualifications and ability at 7[14]. We also drop individuals who are self employed[15]. This leaves us with a final sample of 3048 males of whom 2597 are employed in 1991 and 3894 females of whom 2363 are employed in 1991. Summary Statistics for these individuals are given in Table .1 in the Appendix. These show that the sample used in the paper under-represents individuals in the bottom quintiles of the reading and arithmetic ability tests undertaken when the child was 7.

3. Methodology

3.1. Estimation Methodology

Our methodological approach assumes that education decisions are made on the basis of variables that are observable (or well proxied by variables) in our NCDS data[16]. To estimate the returns to our school and post-school qualifications we estimate the following wage equation

$$\begin{aligned} w_i &= \beta_1' Q_i^1 + \beta_2' Q_i^2 + \alpha' X_i + \varepsilon_i \\ &= \beta' Q_i + \alpha' X_i + \varepsilon_i \end{aligned} \quad (3.1)$$

where Q_i is a vector of dummy variables identifying the person's highest school qualification Q_i^1 and highest post-school qualification Q_i^2, w_i is the log of the real hourly wage rate, X_i is a vector of exogenous observed individual characteristics,

[14] Rather than dropping individuals who have missing information on other variables of interest we include missing variable dummies.

[15] The NCDS does not have reliable wage information for the self-employed.

[16] This kind of approach has also been taken in the papers by Dearden, Blundell, Goodman and Reed [6] looking at Higher Education and Dearden, Ferrier and Meghir [14] in looking at school quality.

β measures the returns to school and post-school qualifications and ε_i is a residual term.

OLS estimation of equation (3.1) gives rise to a unbiased estimate of the return to education if both Q_i^1 and Q_i^2 are exogenous in equation (3.1) (i.e. $E(Q_i^1, \varepsilon_i) = E(Q_i^2, \varepsilon_i) = 0$). This will arise if conditioning on the observable variables (X_i) is sufficient to control for the endogenous choice of individual's school and post-school qualifications. We assume that individuals who are the same in the observable dimension X_i but choose different levels of schooling and further education do not differ on average in the unobserved dimension ε_i. Formally this means that $E(\varepsilon_i|Q_i, X_i) = E(\varepsilon_i| X_i)$. The arguments used here are similar to the arguments made for the matching estimators (see Heckman, Ichimura and Todd [23] and Dearden, Ferrier and Meghir[14] for more details). If, however, there are unobserved determinants of wages which are correlated with education choices then OLS will produce biased estimates of the returns to schooling.

In equation (3.1) we assume that there is a constant return to different qualifications. The model could be extended to allow the returns to education to be heterogeneous (i.e. $\beta_i = \beta + e_i$ where $Var(e_i) > 0$). If we assume that only the average population value of e_i, conditional on the observables is known by the person undertaking the choice of Q_i then $E(e_i|Q_i, X_i)Q_i = E(e_i| X_i)Q_i$. Hence the average effect β can be identified by the regression

$$w_i = \beta' Q_i + \alpha' X_i + \delta' (X_i \otimes Q_i) + u_i \quad (3.2)$$

where $E(u_i|Q_i, X_i) = 0$. In equation (3.2) the coefficients δ reflect the heterogeneity in the returns to Q_i. Given the above assumptions the model can be estimated by Ordinary Least Squares (OLS). The standard errors must be computed using White's (1982) adjustment for heteroskedasticity, because the heterogeneous returns imply that the variance of u_i will depend on Q_i.

3.2. Controlling for measurement error in schooling

Clearly OLS estimation of equation (3.1) or equation (3.2) will only be consistent if there are no other unobserved individual effects correlated with qualifications (or indeed any right hand side variable), that is if $E(Q_i\varepsilon_i) = 0$. If our qualifications are measured with error (or our methodological approach does not appropriately control for the endogeneity of education) then our estimates of the returns to school and post-school qualifications will still be biased. The biases associated with measurement error in education are discussed in detail in Ashenfelter and Krueger [3] and Card[11]. If this is the case then we have to rely on instrumental variable techniques. This requires finding at least one instrument which is correlated with the true measure of each of our qualification variables and uncorrelated with the measurement error. For each individual in our data we have a number of measures of their qualification outcomes by the age of 23 in 1981. In an attempt to correct for possible measurement error we use measures of the persons highest school and post-school qualification as at March 1981 *reported by the individual in 1991*, as instruments for their earlier qualification outcomes reported in 1981. If the measurement errors in the 1991 reports of educational outcomes are uncorrelated with the measurement errors in the 1981 variables, this IV procedure should eliminate any downward bias associated with measurement error. This is an open question, but we feel our attempt may give us some ball park figures on the extent of measurement error in our data. As a check on the robustness of our IV procedure we also compare results obtained when we instead use our 1981 survey measures of qualifications as instruments for our 1991 survey education variables[17].

One problem, however, is that our qualification measures are discrete. We adopt two approaches in carrying out our instrumental variable technique. The first exploits the fact that our highest school and post-school qualification variables are ordered and estimates two reduced form ordered probits for our two

[17]The author would like to thank Arthur van Soest for making this suggestion.

qualification variables. We then use the parameter estimates to calculate the usual Heckman [20] selection adjustment term for our ordered qualification variables which we denote $\widehat{\lambda}_{Q_i^1}$ and $\widehat{\lambda}_{Q_i^2}$. We can then estimate the following model

$$w_i = \beta' Q_i + \alpha' X_i + \varphi' \widehat{\lambda}_i + \varepsilon_i \tag{3.3}$$

where $\widehat{\lambda}_i' = (\widehat{\lambda}_{Q_i^1}, \widehat{\lambda}_{Q_i^2})$[18]. A Hausman test of no measurement error in our qualification variables can be obtained by testing whether $\varphi = 0$ in equation (3.3) (see Smith and Blundell[30]). In this formulation our standard errors are corrected to take account of the generated regressors ($\widehat{\lambda}_i$) in the equation[19]. As a check on the robustness of this procedure we also estimate a standard IV model which, by definition, uses linear probability models for each of the different qualifications rather than ordered probits in the first stage estimation. This IV procedure does not exploit the ordering of our education qualification variables[20].

3.3. Composition bias

One problem with our approach is that we only look at the returns to qualifications for men and women who are in work in 1991. In particular, a sizable proportion of women are not in work in our 1991 sample, which may bias our estimates of the returns to education. This arises because the characteristics of those women in work and out of work may differ and we need to take this potential 'composition bias' into account. Ideally, we should estimate a structural model correcting for selection into work, in order to obtain corrected estimates of the returns to qualifications. Unfortunately this strategy requires finding at least one suitable identifying restriction, namely a variable which determines the work participation

[18]A similar IV approach was taken in the papers of Vella and Gregory [31] and Harmon and Walker [18]

[19]See for example, Pagan [27] and Arellano and Meghir [2].

[20]In fact, we should estimate probit models for each of our qualification variables and include inverse mills ratios from this estimation procedure and appropriately adjust the standard errors. This is computationally burdemsome and is complicated by the fact that a lot of our 1991 dummy variables completely determine not obtaining a certain 1981 qualification level (and vice versa) and hence have to be dropped from the particular probit model. These problems are avoided when using a linear probability model.

decision, but not earnings for those women in work. First, we have argued that family background variables are potentially important determinants of wages and hence not suitable instruments. The presence of children has been used in some studies as an instrument for the participation decision. Iacovou [24] has shown, however, that the number of children is endogenous for female participation in the NCDS cohort. When she controls for the endogeneity of children she finds that they have no significant effect on labour supply. Thus the strategy used in the paper will be to discuss the possible size of this bias under different assumptions.

Conditional on participation (P_i) the wage equation (3.1) becomes:

$$E(w_i|Z_i, P_i = 1) = \gamma' Z_i + g(Z_i) \tag{3.4}$$

where $Z_i' = (Q_i', X_i')$ and $g(Z_i) = E(\varepsilon_i|Z_i, P_i = 1)$ and where ε_i summarises the unobservables in the wage equation. Under the assumption of bivariate normality of log wages and the variable inducing participation, then $g(Z_i) = \zeta\lambda(\psi' Z_i)$ where $\lambda(\psi' Z_i) = \phi(\psi' Z_i)/\Phi(\psi' Z_i)$ is the inverse Mill's ratio and where ϕ and Φ are the standard normal density and distribution functions respectively. The latter represents the probability of participation for those with characteristics Z_i. Under joint normality $\zeta = \rho\sigma$ where ρ is the correlation between the participation equation and the log wage equation and σ is the standard deviation of the wage residual term ε_i. To a first order approximation we can write $\lambda(\psi' Z_i) \approx \omega\psi'(Z_i - \overline{Z})$, where $\omega = \partial\lambda/\partial(\psi' Z_i)|_{Z_i=\overline{Z}}$ and where $\overline{Z}$ are the characteristics of individuals at the means.

Under the assumption of normality the slope coefficients ($\widehat{\gamma}$) using a first order approximation are given by

$$\widehat{\gamma} = \gamma + \rho\sigma\omega\psi \tag{3.5}$$

where γ is a vector of the true population returns. Clearly if $\rho = 0$ the workers are a representative sample of the entire population and the estimated coefficients of the wage equation do not suffer from composition bias. If individuals self-select into employment by comparative advantage then we would expect $\rho > 0$.

The value of ρ can not be identified without further identifying restrictions (unless one uses the nonlinear restrictions from the normality assumption). In general $\omega = \partial\lambda(\psi' Z_i)/\partial(\psi' Z_i)|_{Z_i=\overline{Z}}$ may be positive or negative and unbounded. However, under normality $\omega = -\lambda(\psi'\overline{Z}) \times (\psi'\overline{Z} + \lambda(\psi'\overline{Z})) < 0$ i.e. is unambiguously negative[21]. This term can easily be easily calculated from our employment probits. In this case the coefficients in the wage equation for any variable which has a positive effect on participation decision, will be downward biased. Thus if our education qualification variables have a positive effect on the probability of participation, then our estimates of the returns to these qualifications will be downward biased. This issue will be investigated further in our empirical work.

3.4. Gender Wage Differentials

The mean difference in the observed wages of men and women in terms of log differences, or gender gap ($\widehat{g}$) is given by

$$\widehat{g} = \widehat{\gamma}_m'\overline{Z}_m - \widehat{\gamma}_f'\overline{Z}_f \tag{3.6}$$

where $\overline{Z}_m$ and $\overline{Z}_f$ are vectors containing the means of all the explanatory variables in our male and female wage equations and $\widehat{\gamma}_m'$ and $\widehat{\gamma}_f'$ are the corresponding estimated coefficient vectors. Following the approach of Oaxaca [26] and Harkness [17], we can rewrite this expression as

$$\widehat{g} = \widehat{\gamma}_m'(\overline{Z}_m - \overline{Z}_f) + (\widehat{\gamma}_m - \widehat{\gamma}_f)'\overline{Z}_f = \widehat{g}_c + \widehat{g}_p \tag{3.7}$$

which decomposes the observed raw gender wage differential into two effects[22]. The first ($\widehat{g}_c$) is the difference in observed wages which arises because men and women have different observed characteristics, for instance education and labour market experience. The second ($\widehat{g}_p$) is the differences in observed wages which is

[21] This will be true for any log concave distributions and is not limited to the normal distribution.

[22] An alternative way of doing this decomposition is as $\widehat{g} = \widehat{\gamma}_f'(\overline{Z}_m - \overline{Z}_f) + (\widehat{\gamma}_m - \widehat{\gamma}_f)'\overline{Z}_m = \widehat{g}_c + \widehat{g}_p$. It is, of course, an arbitrary decision which one we choose.

a result of men and women being "paid" differently for a given set of characteristics. This is the estimated differential which exists once background has been controlled for or the *ceteris paribus* gender wage differential. If observed gender wage differentials primarily reflect differences in observed characteristics then the policy response will be different than if they primarily reflect differences in the "price" paid for the observed characteristics of women. This latter term, is often interpreted as that part of the wage gap resulting from discrimination, but can also be due to differences in unobserved characteristics (see Harkness [17] for a fuller discussion of these issues). The mean gender wage gap of any subgroup s, of our sample, for instance individuals with a particular set of educational qualifications, can be approximately calculated by replacing the mean characteristics of males and females with those of the subgroup s of interest, $\overline{Z}_{sm}$ and $\overline{Z}_{sf}$ in equation (3.7)[23].

4. Results

4.1. Estimates of the Returns to Qualifications

4.1.1. How important is unobserved ability and family background bias?

We begin by comparing 'conventional' OLS estimates of the returns to qualifications with those that explicitly control for ability and family background. These variables are typically not available in studies looking at the returns to education. Our estimates of the returns to highest school and post-school qualifications for men are given in Table 4.1. The base group in these equations are individuals with no school or post-school qualifications by the age of 23. In the first column, headed 'OLS - conventional specification' we control for region of residence in 1974 and 1991 only. This column is taken as a benchmark of typical OLS estimates of the returns to education when only gender, age and region are included

[23]For different subgroups this will not be exact as the mean of the residual will not generally be zero.

as explanatory variables[24]. In the next column, headed 'OLS - preferred specification' we also include our measures of ability, school type variables, demographic variables, family background and composition variables, and variables identifying what we term "employer characteristics" (whether the firm employed more than 500 workers, whether it was in the private sector and whether the individual was a union member)[25]. A detailed set of results for this specification is given in Table .2 in the Appendix. We have carried out this exercise using both our 1981 and 1991 measures of qualification outcomes. The results reported in Table 4.1 are based on the 1981 measures. The results based on the 1991 measures are given in Table .3 in the Appendix.

The OLS estimates presented in the table suggest that there are significant returns to all types of qualifications for men in our sample. We see from column 1 that when we use conventional OLS controls, the estimated return to an A level qualification is 43.3 per cent compared to 29.2 per cent for 5 or more O levels. This suggests an annual return of around 7.0 per cent for undertaking an A level qualification assuming this takes 2 additional years to 5 or more O levels. The estimated return to a CSE qualification is 9.3 per cent and less than 5 O levels 20.1 per cent. In the UK both CSE and O Level qualifications are usually completed by the age of 16, the minimum school leaving age. Thus individuals whose highest schooling is a CSE or O level qualification have significantly higher returns to education than those leaving school at the same time, but with no schooling qualifications. This provides clear evidence of considerable heterogeneity in the return to years of schooling. Finally, the conventional OLS estimate of a return to a degree is 19.5 per cent, which suggests an annual return of around 6.5 per cent (assuming a degree takes 3 years on average). The annual return to under-

[24]In some cases ethnicity is also used as a control. In the NCDS sample over 99% of the sample is white.

[25]In previous versions of the paper we experimented with a number of other possible control variables, including more extensive family background variables from wave 1 of the NCDS. This made no significant difference to the estimates of the returns to qualifications. The crucial background variables appear to be ability, parental education, parental interest in the child's education and father's social class.

taking a higher vocational qualification is substantially higher than for a degree (assuming these qualifications take around 2 years to complete). This in part reflects returns to experience, as individuals undertaking higher vocational qualifications have more actual labour market experience than those who undertook degree qualifications[26].

When we also control for ability and family background variables the estimated returns become smaller, however, these differences are only significant for O level and A level school qualifications. In particular our ability variables are positive and significant for men. Father's social class is also an important determinant of male wages. When we control for these ability and family background variables, our estimated return to an A level qualification is now 33.2 percent and for 5 or more O levels, 20.5 per cent. This suggests an annual return of 6.4 per cent to undertaking an A level, compared to our estimate of 7.0 per cent using a conventional OLS specification. Similarly, our estimated return to a degree is now 16.5 per cent, which suggests an annual return of 5.5 per cent compared to 6.5 per cent using a conventional OLS specification, however this latter difference is not significant at conventional levels.

The corresponding results for women are given in Table 4.2 and those based on the 1991 measures in Table .4 in the Appendix. For women, there are also clear returns to ability, particularly reading ability (see Table .2 in the Appendix). This once again results in a downward revision of our conventional OLS estimates of the returns to qualifications, but once again these differences are only significant for O and A level school qualifications and higher vocational qualifications. It is clear from Table .2 in the Appendix that family background variables, such as father's social class and family composition variables are also important determinants of the level of women's wages. The returns to schooling qualifications are very similar to men, however, the returns to post-school qualifications are higher for

[26]For men in our sample the average actual labour market experience of those with higher vocational qualifications is 14.6 years compared with 10.6 years for men with degrees. The corresponding figures for women are 13.1 and 9.9 years respectively.

Table 4.1: The Returns to Qualifications: Males

Variable	OLS				IV- Preferred Specification			
	Conventional Specification		Preferred Specification		Ordered Probit		Linear Probability	
	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)
Constant	**1.916**	**(0.035)**	**1.676**	**(0.126)**	**1.671**	**(0.126)**	**1.665**	**(0.127)**
Highest School Qualification 1981:								
CSEs	0.093	(0.024)	0.065	(0.024)	0.072	(0.033)	0.085	(0.046)
<5 O Levels	0.201	(0.024)	0.143	(0.024)	0.150	(0.038)	0.166	(0.033)
5+ O Levels	0.292	(0.027)	0.205	(0.029)	0.208	(0.047)	0.218	(0.037)
A levels	0.433	(0.030)	0.332	(0.033)	0.336	(0.048)	0.330	(0.042)
Highest Post-School Qualification 1981:								
Lower vocational	0.091	(0.019)	0.081	(0.019)	0.090	(0.024)	0.066	(0.065)
Middle vocational	0.099	(0.022)	0.088	(0.022)	0.102	(0.030)	0.102	(0.042)
Higher vocational	0.186	(0.026)	0.178	(0.026)	0.196	(0.039)	0.187	(0.048)
Degree	0.195	(0.030)	0.165	(0.030)	0.182	(0.044)	0.198	(0.041)
$\widehat{\lambda}_{Q_i^1}$					-0.008	(0.019)		
$\widehat{\lambda}_{Q_i^2}$					-0.012	(0.014)		
Hausman Tests (P-value):								
School Qualifications							7.817	(0.099)
Post-school Qualifications							4.848	(0.435)
Number of observations	2597		2597		2597		2597	
P-value 1991 regional dummies	0.000		0.000		0.000		0.000	
P-value 1974 regional dummies	0.023		0.015		0.015		0.015	
P-value ability variables			0.000		0.000		0.000	
P-value school type variables			0.325		0.323		0.327	
P-value family variables			0.174		0.199		0.209	
P-value parental interest			0.599		0.643		0.692	
P-value demographics			0.241		0.232		0.259	
P-value employer characteristics			0.000		0.000		0.000	
R^2	0.2973		0.3411		0.3415		0.3400	

Table 4.2: The Returns to Qualifications: Females

Variable	OLS				IV - Preferred Specification			
	Conventional Specification		Preferred Specification		Ordered Probit		Linear Probability	
	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)
Constant	1.620	(0.040)	1.617	(0.133)	1.613	(0.143)	1.557	(0.136)
Highest School Qualification 1981:								
CSEs	0.060	(0.028)	0.037	(0.027)	0.069	(0.073)	0.187	(0.048)
<5 O Levels	0.135	(0.026)	0.075	(0.026)	0.112	(0.077)	0.171	(0.035)
5+ O Levels	0.347	(0.031)	0.239	(0.033)	0.258	(0.149)	0.316	(0.041)
A levels	0.463	(0.034)	0.339	(0.036)	0.370	(0.148)	0.409	(0.048)
Highest Post-School Qualification 1981:								
Lower vocational	0.052	(0.022)	0.036	(0.022)	0.063	(0.056)	0.063	(0.049)
Middle vocational	0.121	(0.040)	0.073	(0.039)	0.105	(0.081)	0.132	(0.082)
Higher vocational	0.316	(0.030)	0.227	(0.030)	0.264	(0.079)	0.319	(0.045)
Degree	0.369	(0.036)	0.317	(0.037)	0.354	(0.108)	0.382	(0.051)
$\widehat{\lambda}_{Q_i^1}$					-0.038	(0.038)		
$\widehat{\lambda}_{Q_i^2}$					-0.031	(0.024)		
Hausman Tests (P-value):								
School Qualifications							22.521	(0.000)
Post-school Qualifications							8.149	(0.086)
Number of observations	2363		2363		2363		2363	
P-value 1991 regional dummies	0.000		0.000		0.000		0.000	
P-value 1974 regional dummies	0.099		0.075		0.088		0.090	
P-value ability variables			0.033		0.081		0.137	
P-value school type variables			0.934		0.941		0.824	
P-value family variables			0.009		0.011		0.016	
P-value parental interest			0.404		0.482		0.521	
P-value demographics			0.603		0.657		0.724	
P-value employer characteristics			0.000		0.000		0.000	
R^2	0.2973		0.4487		0.4523		0.4397	

women than for men. Again it is clear from the Table that all qualifications are associated with significant annual returns compared to the base group who have no qualifications. For women, the annual return to an A level qualification compared to an individual with 5 or more O levels is 5.8 per cent in our conventional OLS specification and 5 per cent when we control for ability and family background variables. The annual return to a degree is 12.3 per cent in our conventional OLS specification and 10.6 when we control for ability and family background, but this difference is not significantly different at conventional levels. The annual return to a higher vocational qualification is around 15 per cent in our conventional OLS specification but only 11 per cent when we have a full set of controls (assuming such qualifications take 2 years on average).

When we compare these estimates with those based on our 1991 measures (see columns 1 and 2 of Tables .3 and .4 in the Appendix), we see that there are some important differences in the estimated returns when we use these different measures. For example, the return to an A level for women using the 1991 measure is significantly higher than the estimate based on the 1981 survey measures. For men the returns to post-school qualifications based on the 1991 measures are somewhat lower than those based on the 1981 measures in both specifications. This difference may in part be due to problems associated with measurement error and this issue is explored in more detail below. The results in these tables however, confirm our findings based on our 1981 measures, that including family background and ability variables significantly reduces the returns to O and A level qualifications for both men and women, as well as the returns to higher vocational qualifications for women.

One important point to emerge from the results is that there is considerable heterogeneity in the returns to education for those individuals who leave school at the minimum school leaving age of 16. Those who obtain CSE and/or O level qualifications by this age have considerably higher returns to education than those who leave at the same age, but obtain no formal school qualifications. This heterogeneity would not be picked up if we just focused on returns to an additional year of education[27]. But there may not only be heterogeneity in the return to an additional year of education, but also heterogeneity in the returns to different qualifications. This latter issue is explored in more detail below.

The results obtained to this point suggest that there are significant returns to ability and other factors such as family background variables and that estimates of the returns to O and A level qualifications (and higher vocational qualifications for women), which do not take this into account over-estimate the returns to such qualifications. However, this is not the end of the story as we have ignored biases associated with measurement error and composition. These issues are explored in more detail below.

4.1.2. Is there evidence of measurement error bias?

In this section we use instrumental variable methods in an attempt to correct for possible measurement error which may be biasing our estimates of the returns to qualifications. If the measurement errors in the 1991 reports of educational outcomes are uncorrelated with the measurement errors in the 1981 variables, then these variables can also be used as instruments to correct for possible measurement error. The results of doing this are reported in columns 3 and 4 of Table 4.1 for men and Table 4.2 for women. The IV results reported in column 3 are based on using an ordered probit model for the highest school and highest post-school qualification in the first stage of the IV procedure[28]. We undertook two checks of the robustness of these results. Firstly we also estimated a standard IV model which, by definition, uses linear probability models for each of the different qualifications rather than ordered probits in the first stage estimation. The results of doing this are presented in column 4 of Tables 4.1 and 4.2. As a second check,

[27]If we only looked at years of education (or years of full-time education which is usually the measure used in the UK), those in this group who undertook no further education would all be treated identically and effectively be the base group.

[28]Unsurprisingly, all our instruments were highly significant in our ordered probit models. These results are available from the author.

we compared the results obtained when we instead use our 1981 survey measures of education as instruments for our 1991 qualification measures and these are reported in columns 3 and 4 of Tables .3 and .4 in the Appendix.

Our IV estimates of the returns to different qualifications reported in column 3 are slightly above our OLS estimates for men, however, these differences are not statistically significant, either individually or jointly. There is more evidence of biases associated with measurement error when we carry out traditional IV estimation as seen from the results from Column 4 of Table 4.1. There appears to be more of a problem with regard to school qualifications than post-school qualifications. These results suggest that there may be some measurement error bias when using our 1981 survey measures, though the evidence is mixed. It is clear from Table .3 in the Appendix that there is more of a measurement error problem with our 1991 measures and our corrected estimates shown in columns 3 and 4 of this Table are now quite close to those of Table 4.1. Our fully corrected estimates of the returns to qualifications are very close to our conventional OLS estimates, for all qualifications except 5 or more O levels and A levels. For these qualifications the effect of omitted ability bias dominates. The results from Table .3 in the appendix suggests that for men, there are more severe problems of measurement error - the longer the time between completing the qualifications and the time of the survey. This is an important point, because for a lot of nationally representative surveys, individuals are asked about the highest qualifications they have obtained and for a lot of these individuals these qualifications will have been completed at least 10 years earlier. If there is a similar recall problem then it is likely that estimates which do not take this into account may be too low.

For women as for men, there is some indications of biases associated with measurement error. From column 3, it appears that the biggest bias is associated with the estimated returns to some post-school qualifications although the estimates are not precisely determined. When we use a standard IV model (see column 4) this result is confirmed for higher vocational and degree qualifications. As was the case with men, our fully corrected estimates of the returns to qualifications are very close to our conventional OLS estimates, for all qualifications except possibly 5 or more O levels and A levels (depending on which model we adopt). For these qualifications the effect of omitted ability and family background bias dominates. When we use our 1981 measures of outcomes as instruments for our 1991 measures we again see that there are significant measurement error problems associated with post-school educational qualifications when we use a standard IV procedure. These fully corrected estimates are again broadly consistent with our fully corrected estimates contained in Table 4.2.

The results from this section suggest that measurement error in our education variables can result in a significant downward bias in our OLS estimates of the returns to some qualifications. This is much more of a problem the longer the elapsed time between completing the qualification and being surveyed. If workers are a representative sample of the entire population (i.e. there is no composition bias), the effect of omitted ability and family background bias outweighs the countervailing bias arising from measurement error for 5 or more O levels and A level qualifications. For all other qualifications, however, the effects are roughly identical, though not precisely determined. It is, however, unlikely that workers are representative of the entire population and we look at this issue in more detail below.

4.1.3. How important is composition bias?

We use the methodology described in the previous section to get an idea of the direction of the biases involved because of possible composition bias. This involves estimating male and female participation equations at age 33 using a probit model[29]. The results of doing this show that qualifications are an important positive determinant of male and female participation at 33. For example, a man with A levels increases his probability of employment by 10.1 percentage

[29]The full set of results from these employment probits are available from the author.

points compared to a man with no school qualifications. If he also has a degree, his probability of employment increases by a further 3.9 percentage points. For women with A levels, the probability of being employed is 6.7 percentage points higher on average than women with no school qualifications although this effect is only marginally significant.. Women with degrees, however, are 10.6 percentage points more likely to be employed than women with no post-school qualifications. The coefficient estimates, which we require to work out the composition bias are presented in Tables 4.3 and 4.4 below. All qualifications with the exception of middle vocational qualifications for women have a positive and significant impact on the probability of being employed. We also saw from the previous sections, that all qualification levels impact positively on wages. Moreover, if individuals self-select into employment by comparative advantage, we would expect the correlation between the participation equation and the log wage equation to be positive (i.e.$\rho > 0$ in the model of section 3.3).

If we think these assumption are valid, then the results of Table 4.3 imply that the returns to all qualifications for men are underestimated because of self-selection into employment. For women this is true for all qualifications except middle vocational qualifications[30]. To assess the possible magnitude of the bias note that under normality for men, $\omega_m \approx -0.294$ and $\sigma_m \approx 0.352$. Thus an approximation of the bias (b_m) of our estimated coefficients in the male wage equations is given by $b_m = (-(0.352 \times 0.294)\rho_m\psi_m = -0.104\rho_m\psi_m$ where ψ_m are the coefficients from our male employment probit maximum likelihood procedure. Note that if there is measurement error in our education variables then these probit coefficients may also be downward biased and hence our estimates of the possible bias may be on the low side.

In Table 4.3 we look at our estimate of the bias under the assumption that $\rho = 0.1$ and $\rho = 0.66$. In the first column, we present our earlier conventional OLS estimates of the returns to qualifications, whilst in column 2 we present the

[30]There are very few women in our sample with middle vocational qualifications.

Table 4.3: Composition Bias and the Returns to Qualifications: Males

Variable	Conventional	IV -	Employment		Returns adjusting for composition bias			
					$\rho_m = 0.10$		$\rho_m = 0.66$	
Qualification 1981:	OLS Coef.	Ordered Probit Coef.	Probit Coef.(ψ_m)	(S.E.)	Bias	Corrected IV Coef.	Bias	Corrected IV Coef.
Highest School Qualification 1981:								
CSEs	0.093	0.071	0.505	(0.097)	-0.005	0.076	-0.035	0.106
<5 O Levels	0.201	0.148	0.488	(0.096)	-0.005	0.153	-0.033	0.181
5+ O Levels	0.292	0.206	0.528	(0.123)	-0.005	0.211	-0.036	0.242
A levels	0.433	0.334	0.612	(0.137)	-0.006	0.340	-0.042	0.376
Highest Post-School Qualification 1981:								
Lower vocational	0.091	0.090	0.225	(0.081)	-0.002	0.092	-0.015	0.105
Middle vocational	0.099	0.102	0.518	(0.110)	-0.005	0.107	-0.036	0.138
Higher vocational	0.186	0.196	0.802	(0.171)	-0.008	0.204	-0.055	0.251
Degree	0.195	0.183	0.334	(0.161)	-0.003	0.186	-0.023	0.206

results from our IV ordered probit model. In column 3 of the table we present the coefficient estimates and standard errors from our employment probit model. In the final columns we present estimates of the bias and the corrected IV coefficients (obtained by adding the bias to the estimates in column 2) under the assumption that $\rho = 0.1$ and $\rho = 0.66$. It is easy to calculate the bias for other values of ρ. The table shows that if our assumptions hold, then the bias in our estimate of the return to an A level qualification is approximately $b_m = -0.104 \times 0.612\rho = -0.064\rho$. Thus the degree of composition bias ranges between 0 and -6.4 percentage points and is -0.6 percentage points if $\rho = 0.1$ and -4.2 percentage points if $\rho = 0.66$ (see Table 4.3). Potentially large effects also exist for other qualifications such as middle and higher vocational qualifications and all other school qualifications. This suggests that self-selection into employment may results in downward biased estimates of the returns to education for men. This is also true when we use our 1991 measures[31].

For women $\omega_f \approx -0.570$ and $\sigma_f \approx 0.369$[32]. Thus an approximation of the bias (b_f) for women is given by $b_f = (-(0.369 \times 0.570)\rho_f\psi_f = -0.21\rho_f\psi_f$. The results

[31] When we use our 1991 measures of qualifications the corresponding figures are $\omega_m \approx -0.290$ and $\sigma_m \approx 0.353$.

[32] When we use our 1991 measures of qualifications the corresponding figures are $\omega_f \approx -0.566$ and $\sigma_f \approx 0.370$.

from Table 4.4 show that the order of composition bias for women is very similar to that of men (*assuming* ρ is similar for men and women). This arises because education, on average has larger marginal effects (i.e. the coefficients on the education qualifications in the male employment probit are generally larger for men than women) and this counterbalances the fact that $\omega = \partial\lambda(\psi' Z_i)/\partial(\psi' Z_i)|_{Z_i=\overline{Z}}$ is larger for women than men. The exact order of composition bias depends on the value of ρ (the correlation between the participation equation and the log wage equation) and we have no way of correctly identifying the magnitude of this parameter or establishing whether it is larger or smaller for men compared to women. This of course also assumes that the linearisation as well as the computations made under the normality assumption are an adequate approximation of self-selection into employment. These issues clearly needs to be investigated in future research.

Table 4.4: Composition Bias and the Returns to Qualifications: Females

Variable	Conventional	IV -	Employment		Returns adjusting for composition bias			
					$\rho_f = 0.10$		$\rho_f = 0.66$	
Qualification 1981:	OLS Coef.	Ordered Probit Coef.	Probit Coef.(ψ_f)	(S.E.)	Bias	Corrected IV Coef.	Bias	Corrected IV Coef.
Highest School Qualification 1981:								
CSEs	0.060	0.069	0.247	(0.077)	-0.005	0.074	-0.034	0.093
<5 O Levels	0.135	0.112	0.317	(0.071)	-0.007	0.119	-0.044	0.156
5+ O Levels	0.347	0.258	0.254	(0.083)	-0.005	0.263	-0.035	0.293
A levels	0.463	0.369	0.068	(0.095)	-0.001	0.370	-0.009	0.378
Highest Post-School Qualification 1981:								
Lower vocational	0.052	0.063	-0.061	(0.058)	0.001	0.062	0.008	0.055
Middle vocational	0.121	0.105	0.240	(0.130)	-0.005	0.110	-0.033	0.138
Higher vocational	0.316	0.265	0.469	(0.087)	-0.010	0.275	-0.065	0.330
Degree	0.369	0.354	0.452	(0.102)	-0.009	0.363	-0.063	0.417

4.2. Is there heterogeneity in the returns to qualifications?

In the next section of the paper we take a further look at whether there is any evidence of heterogeneity in the returns to education by interacting our qualification variables with ability and family background variables. We split firstly split our sample into high ability and low ability groups. A person is taken to be of high ability if they are in the top two quintiles of either the mathematics or reading ability tests. We then interact all our education variables with this high ability dummy variable. We also interacted our education variable with three of our family background variables: the dummy variable identifying individuals coming from families with financial difficulties, a dummy variable identifying children whose mother was interested in their education at an early age[33] and the father's years of education variable. Card[11] has speculated that children from relatively disadvantaged family backgrounds (which should be picked up from our financial difficulties dummy variable) and/or with relatively low tastes for education (possibly individual's whose father has low levels of education or whose mother take little interest in their education) may choose low levels of education because they have high discount rates rather than low ability. If this is the case then the marginal return to schooling for these individuals will exceed the average return to schooling for the population as a whole.

We find no evidence of heterogeneity in the returns to education according to ability and family financial circumstances as a child. There is, however, some evidence that the returns to education decrease as father's education increases. Father's education, however, has a large and generally significant positive effect on the overall *level* of wages received by individuals[34].We also find some evidence the estimated return to different qualifications is generally lower for individual's whose mother showed interest in their child's education at a young age. The results of interacting qualifications with mother's interest in the child's education is given in Table 4.5.

There is some evidence that the marginal return to some qualifications are lower for men whose mother took an active interest in their education at the age of 7. From Table 4.5 we see that the return to an A level is 10.0 per cent lower for such men and the return to a middle vocational qualification 14.4 per cent lower.

[33]Our dummy variable identified children whose mother expected too much or were very interested in their child's education at the age of 7.

[34]See Dearden [13] for more details.

Table 4.5: Heterogeneity and Returns to Qualifications

Variable	Men				Women			
	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)
Constant	1.676	(0.126)	1.678	(0.126)	1.617	(0.133)	1.590	(0.133)
Highest School								
Qualification 1981:								
CSEs	0.065	(0.024)	0.078	(0.027)	0.037	(0.027)	0.051	(0.029)
<5 O Levels	0.143	(0.024)	0.110	(0.027)	0.075	(0.026)	0.063	(0.028)
5+ O Levels	0.205	(0.029)	0.204	(0.036)	0.239	(0.033)	0.303	(0.041)
A levels	0.332	(0.033)	0.382	(0.039)	0.339	(0.036)	0.305	(0.050)
Highest Post-School								
Qualification 1981:								
Lower vocational	0.081	(0.019)	0.091	(0.023)	0.036	(0.022)	0.053	(0.028)
Middle vocational	0.088	(0.022)	0.147	(0.027)	0.073	(0.039)	0.063	(0.059)
Higher vocational	0.178	(0.026)	0.195	(0.034)	0.227	(0.030)	0.261	(0.040)
Degree	0.165	(0.030)	0.149	(0.043)	0.317	(0.037)	0.396	(0.063)
Mother very interested×								
Highest School								
Qualification 1981:								
CSEs			-0.064	(0.059)			-0.060	(0.070)
<5 O Levels			0.066	(0.059)			0.011	(0.065)
5+ O Levels			-0.014	(0.065)			-0.145	(0.070)
A levels			-0.100	(0.067)			0.019	(0.077)
Highest Post-School								
Qualification 1981:								
Lower vocational			-0.022	(0.039)			-0.041	(0.045)
Middle vocational			-0.148	(0.044)			0.018	(0.079)
Higher vocational			-0.046	(0.052)			-0.063	(0.055)
Degree			0.017	(0.060)			-0.118	(0.075)
Mother's interest in edn:								
Expects too much	-0.013	(0.050)	0.035	(0.070)			0.106	(0.072)
Very interested	0.026	(0.025)	0.077	(0.056)			0.109	(0.057)
Some interest	0.037	(0.020)	0.033	(0.020)			0.032	(0.022)
Number of observations	2597		2597		2363		2363	
P-value Interaction Variables			0.002				0.034	
P-value 1991 regional dummies	0.000		0.000		0.000		0.000	
P-value 1974 regional dummies	0.015		0.011		0.075		0.081	
P-value ability variables	0.000		0.000		0.033		0.047	
P-value school type variables	0.325		0.337		0.934		0.892	
P-value family variables	0.174		0.181		0.009		0.017	
P-value parental interest	0.599		0.541		0.404		0.433	
P-value demographics	0.241		0.234		0.603		0.571	
P-value employer characteristics	0.000		0.000		0.000		0.000	
R^2	0.3411		0.3476		0.4487		0.4529	

A joint test that the coefficients on the interaction variables are equal to zero is clearly rejected. For women, we see from that the average return to 5 or more O levels is 14.5 per cent lower for women whose parent's took an active interest in their education and again a joint test that the coefficients on the interaction variables are equal to zero, is rejected. However, for women whose mother took an active interest in their child's education at the age of 7, the overall *level* of wages is just over 10 per cent higher.

The results from this part of the paper provide further evidence of heterogeneity in the returns to education and some support for the idea that individual's with less taste for education may gave higher average marginal returns to certain education qualifications than the population as a whole. This suggests that IV procedures which rely on interventions that affect schooling choices of children with low tastes for education, may overestimate the true average marginal return to education. This evidence, together with our earlier results showing considerable heterogeneity in the returns to education for those leaving school at the minimum school leaving age, might in part explain why relatively higher returns to education were obtained in the earlier UK study of Harmon and Walker[18] which relied on IV techniques. In a more recent paper Harmon and Walker[18] allow non-linearities in the schooling earnings relationship by including the number of years of post-18 schooling in addition to the total number of years of schooling. They find that there is marked sensitivity to the instruments used once they use this non-linear specification. However, their design does not allow for heterogeneity for those leaving education at the minimum school leaving age which we have found to be important in this paper.

4.3. Qualifications and Gender Wage Differentials

In the final section of the paper we look at gender wage differentials and in particular how these wage differentials vary for different qualification group. These gender wage gaps are decomposed into differences that can be explained in terms

of observed average differences in the characteristics of men and women ($\widehat{g}_c$) and that attributable to women's characteristics being valued differently to those of men ($\widehat{g}_p$). The estimates are based on our ordered probit IV model of column 3 in Tables 4.1 and 4.2 and hence do not take account of possible composition bias. One important difference between men and women is obviously going to be actual labour market experience and this is not specifically controlled for in our wage models (because of endogeneity issues which we cannot satisfactorily control for). In estimating our gender wage differentials we also include experience in our IV ordered probit model wage models, to see what impact this has in decomposing gender wage differentials across different qualification groups. The results of doing this are also reported in Table 4.6. We have chosen 11 mutually exclusive qualification groups. These identify individuals by their highest school qualification and whether or not they have also undertaken post-school qualifications. For those whose highest school qualification is A levels, we also distinguish between sub-degree post-school qualifications and degree qualifications.

The results show that even when we control for actual labour market experience, less than one-third of the observed wage differential between men and women is attributable to differences in observed characteristics. Focusing on the results when we control for experience, we see that after controlling for differences in background, the largest differential is between men and women with less than 5 O levels and no post-school qualifications (32.8 per cent). Obtaining post-school qualifications reduces this *ceteris paribus* gender wage differential for individuals with O and A level qualifications. In general gender wage differentials initially rise and then fall with schooling qualifications, but even after controlling for observed differences in characteristics, these differentials are highly significant. The lowest differential is among women and men with A level and degree qualifications, suggesting that university education plays an important role in reducing the gender earnings gap. While these differences may be partly due to discrimination in the labour market, they may also be due to differences in other unobserved characteristics. One important characteristic, which has been ignored in this paper, is access to work-related training. Dearden [12] shows that this may account, on average, for up to 3 percentage points of the observed difference in male and female earnings.

Table 4.6: Qualifications and Gender Wage Differentials

Highest Qualification	$\widehat{g}_c = \widehat{\gamma}_m(\overline{Z}_m - \overline{Z}_f)$ Estimate	(S.E.)	$\widehat{g}_p = (\widehat{\gamma}_m - \widehat{\gamma}_f)'\overline{Z}_f$ Estimate	(S.E.)	$\widehat{g} = \widehat{g}_c + \widehat{g}_p$ Coef.	(S.E.)
IGNORING EXPERIENCE:						
Qualifications:						
No school or post-school	0.011	(0.007)	0.380	(0.030)	0.391	(0.029)
No school but post-school	-0.043	(0.014)	0.378	(0.044)	0.336	(0.042)
CSEs and no post-school	0.019	(0.006)	0.394	(0.026)	0.413	(0.025)
CSEs and post-school	-0.003	(0.011)	0.424	(0.032)	0.420	(0.030)
<5 O Levels and no post-school	0.020	(0.006)	0.424	(0.028)	0.445	(0.027)
<5 O Levels and post-school	-0.007	(0.009)	0.406	(0.027)	0.399	(0.026)
5+ O Levels and no post-school	0.018	(0.007)	0.332	(0.035)	0.350	(0.035)
5+ O Levels and post-school	-0.011	(0.009)	0.275	(0.031)	0.264	(0.030)
A Levels and no post-school	0.042	(0.009)	0.316	(0.044)	0.358	(0.043)
A Levels and sub-degree	0.007	(0.009)	0.286	(0.035)	0.292	(0.034)
A Levels and degree	0.018	(0.008)	0.143	(0.032)	0.161	(0.031)
All Qualifications	0.028	(0.006)	0.342	(0.012)	0.371	(0.010)
CONTROLLING FOR EXPERIENCE:						
Qualifications:						
No school or post-school	0.145	(0.016)	0.233	(0.033)	0.378	(0.029)
No school but post-school	0.074	(0.019)	0.234	(0.044)	0.309	(0.040)
CSEs and no post-school	0.143	(0.015)	0.274	(0.029)	0.417	(0.024)
CSEs and post-school	0.108	(0.016)	0.294	(0.033)	0.403	(0.029)
<5 O Levels and no post-school	0.127	(0.013)	0.328	(0.029)	0.456	(0.026)
<5 O Levels and post-school	0.082	(0.013)	0.301	(0.028)	0.383	(0.025)
5+ O Levels and no post-school	0.086	(0.010)	0.267	(0.034)	0.353	(0.033)
5+ O Levels and post-school	0.049	(0.011)	0.209	(0.031)	0.258	(0.029)
A Levels and no post-school	0.065	(0.009)	0.300	(0.043)	0.365	(0.042)
A Levels and sub-degree	0.040	(0.010)	0.264	(0.034)	0.305	(0.033)
A Levels and degree	0.039	(0.008)	0.115	(0.031)	0.155	(0.029)
All Qualifications	0.112	(0.011)	0.258	(0.015)	0.371	(0.010)

5. Conclusion

The paper has attempted to estimate the returns to school and post-school qualifications for a sample of individuals born in Britain in March 1958 who have been followed since birth. The data used has a wealth of information on family background including parental education, social class and interest shown in the child's education as well as measures of ability. These variables are typically missing in

studies looking at the returns to schooling.

The paper finds that conventional OLS estimates of the returns to education which typically control for age, gender, ethnicity and region of residence and ignore the endogeneity of education, are not a bad approximation of the true causal impact of education on wages. This is because the effects of omitted ability and family background bias are generally completely offset by the effects of measurement error bias and composition bias arising from self-selection into employment. We conclude that it is entirely reasonable to use simple conventional estimates of the returns to qualifications as a basis for policy. This suggests that it is possible to use of a number of nationally representative surveys which only have limited or indeed no information on ability or family background to estimate the returns to qualifications for different cohorts of individuals over different periods of time. Based on our work with the NCDS cohort, such conventional OLS estimates will be close to the average causal impact of different qualifications on wages.

The paper also presents evidence that the returns to an additional year of schooling in the UK are heterogeneous. In particular, individuals leaving school at the minimum school leaving age with some formal qualification (either CSE or O level qualifications) have significantly higher returns to education than those leaving with no formal qualifications. In the UK it is clear that it is important to look at the returns to qualifications rather than an additional year of education as the relationship is far from linear. Moreover, there is also some evidence that individuals with lower tastes for education, have significantly higher marginal returns to certain qualifications. This suggests that IV procedures which rely on interventions that affect schooling choices of children with low tastes for education, may overestimate the true average marginal return to education in the UK. Finally we find that post-school qualifications, particularly degree qualifications, play an important role in reducing gender wage differentials.

References

[1] Angrist, J.D. and Imbens, G., (1995), 'Two-stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity', *Journal of the American Statistical Association,* 90, pp. 431–442.

[2] Arrelano, M. and Meghir, C. (1992), 'Female labour supply and on–the–job search: an empirical model estimated using complementary data sets', *Review of Economic Studies,* 59, pp. 537–57.

[3] Ashenfelter and Krueger, (1994), 'Estimates of the Economic Returns to Schooling from a New Sample of Twins', *American Economic Review,* vol. 84, pp. 1157–73.

[4] Ashenfelter, O. and Zimmerman, D. (1993), 'Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons and Brothers', Working Paper no. 318, Industrial Relations Section, Princeton University, September.

[5] Blackburn, M.L. and Neumark, D. (1993), 'Omitted Ability Bias and the Increase in the Return to Schooling', *Journal of Labor Economics,* vol. 11, pp. 521–43.

[6] Blundell, R., Dearden, L., Goodman, A. and Reed, H. (1997), *Higher Education, Employment and Earnings in Britain,* Institute for Fiscal Studies, London.

[7] Blundell, R., Dearden, L. and Meghir, C. (1996), *The Determinants and Effects of Work-related Training in Britain,* Institute for Fiscal Studies, London.

[8] Bound, J., Jaeger, D.A. and Baker, P. (1993), 'The Cure Can be Worse than the Disease - A Cautionary Tale Regarding Instrumental Variables', NBER Technical Paper No. 139.

[9] Butcher, K. and Case, A. (1994), 'The Effect of Sibling Sex Composition on Women's Education and Earnings', *Quarterly Journal of Economics,* vol 109, pp. 531-63.

[10] Card, D. (1993), 'Using geographic variation in college proximity to estimate the return to schooling', National Bureau of Economic Research, Working Paper no. 4832.

[11] Card, D. (1999), 'The Causal Effect of Education on Earnings', in Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics, vol 3A,* North Holland, Amsterdam.

[12] Dearden, L. (1995), 'Education, training and earnings in Australia and Britain', unpublished Ph.D. thesis, University College London.

[13] Dearden, L. (1998), 'Ability, Families, Education and Earnings in Britain', IFS Working Paper no. W98/14.

[14] Dearden, L., Ferrier, J., and Meghir, C. (1999), 'The Effect of School Quality on Educational Attainment and Wages', IFS Working Paper no. W98/4.

[15] Griliches, Z. (1977), 'Estimating the Returns to Schooling: Some Econometric Problems', *Econometrica,* vol. 45, 1pp. –22

[16] Griliches, Z. and Mason, W. (1972), 'Education, Income and Ability', *Journal of Political Economy, vol.* 80, pp. S74–S103.

[17] Harkness, S. (1996), 'The gender earnings gap: evidence from the UK', *Fiscal Studies,* vol. 2, pp.1–36.

[18] Harmon, C. and Walker, I. (1995), 'Estimates of the economic return to schooling for the UK', *American Economic Review,* vol. 85, pp. 1278–86.

[19] Harmon, C. and Walker, I. (1999), 'The Marginal and Average Returns to Schooling in the UK', forthcoming *European Economic Review.*

[20] Heckman, J.J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica, vol.* 47, pp. 153–61.

[21] Heckman, J.J. (1995a), 'Instrumental variables: a cautionary tale', unpublished manuscript, University of Chicago.

[22] Heckman, J.J. (1995b),'Identification and estimation in the model, "Earnings, ability and schooling"', unpublished manuscript, University of Chicago.

[23] Heckman, J.J., Ichimura, H. and Todd, P. (1997), 'Matching as an Econometric evaluation Estimator: Evidence from Evaluating a Job Training Program', forthcoming *Review of Economic Studies.*

[24] Iacovou, M. (1996), 'Fertility and Female Labour Force Participation', unpublished manuscript, University College London, July.

[25] Micklewright, J. (1988), "Schooling Choice, Educational Maintenance Allowances and Panel Attrition", Department of Economics Queen Mary College Paper No. 185, July.

[26] Oaxaca, R. (1973), 'Male female wage differentials in urban labor markets', *International Economic Review,* pp. 693–703.

[27] Pagan, A.(1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review,* 25, 221–247.

[28] Meghir, C. and Palme, M (1999), 'Assessing the Effect of Schooling on Earnings using a Social Experiment', forthcoming, IFS Working Paper.

[29] Schmitt, J. (1993), "The Changing Structure of Male Earnings in Britain, 1974–88", Centre for Economic Performance, London School of Economics, Discussion Paper No. 122, March.

[30] Smith, R. and Blundell, R.W. (1986), "An Exogeneity Test for a Simultaneous Tobit Model with an Application to Labour Supply", *Econometrica,* 54, 679–685.

[31] Vella, F. and Gregory, R.G. (1992), "Are We Over-educating Young Australians? Human Capital Investment and Self-Selection" in Gregory, R.G. and Karmel, T. (eds), *Youth in the Eighties,* Centre for Economic Policy Research, Australian National University, 152–175.

Appendix

Table .1: Summary Statistics

Variable	Males				Females			
	Employed 2597 Obs		Total 3048 Obs		Employed 2363 Obs		Total 3894 Obs	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Real log hourly wage	2.053	(0.428)			1.682	(0.491)		
Labour market experience	14.767	(2.689)			12.278	(3.542)		
Education measures from 1981 survey:								
Years of full-time education by 1981	12.260	(2.076)			12.248	(2.014)		
Highest School Qualification 1981:								
None	0.139	(0.346)	0.189	(0.391)	0.121	(0.326)	0.168	(0.374)
CSEs	0.190	(0.392)	0.186	(0.389)	0.159	(0.366)	0.164	(0.370)
<5 O Levels	0.249	(0.432)	0.239	(0.426)	0.290	(0.454)	0.276	(0.447)
5+ O Levels	0.157	(0.364)	0.146	(0.353)	0.191	(0.393)	0.177	(0.381)
A levels	0.265	(0.442)	0.241	(0.428)	0.239	(0.427)	0.215	(0.411)
Highest Post-School Qualification 1981:								
None	0.386	(0.487)	0.431	(0.495)	0.554	(0.497)	0.600	(0.490)
Lower vocational	0.216	(0.412)	0.211	(0.408)	0.181	(0.385)	0.185	(0.388)
Middle vocational	0.162	(0.368)	0.148	(0.355)	0.037	(0.188)	0.032	(0.176)
Higher vocational	0.105	(0.306)	0.092	(0.289)	0.115	(0.319)	0.091	(0.288)
Degree	0.131	(0.338)	0.118	(0.322)	0.113	(0.317)	0.092	(0.289)
Education measures from 1991 survey:								
Years of full-time education by 1981	12.168	(2.000)			12.205	(1.974)		
Highest School Qualification 1981:								
None	0.114	(0.318)			0.110	(0.312)		
CSEs	0.171	(0.376)			0.129	(0.336)		
<5 O Levels	0.291	(0.454)			0.330	(0.470)		
5+ O Levels	0.165	(0.371)			0.190	(0.392)		
A levels	0.260	(0.438)			0.242	(0.428)		
Highest Post-School Qualification 1981:								
None	0.459	(0.498)			0.526	(0.499)		
Lower vocational	0.186	(0.389)			0.216	(0.411)		
Middle vocational	0.139	(0.346)			0.028	(0.165)		
Higher vocational	0.106	(0.308)			0.132	(0.339)		
Degree	0.111	(0.314)			0.098	(0.297)		
Maths ability at 7:								
5th quintile (highest)	0.243	(0.429)	0.223	(0.417)	0.215	(0.411)	0.188	(0.391)
4th quintile	0.211	(0.408)	0.204	(0.403)	0.212	(0.409)	0.194	(0.396)
3rd quintile	0.213	(0.409)	0.205	(0.404)	0.197	(0.398)	0.190	(0.392)
2nd quintile	0.175	(0.380)	0.177	(0.382)	0.209	(0.406)	0.201	(0.400)
1st quintile (lowest)	0.158	(0.365)	0.173	(0.378)	0.167	(0.373)	0.179	(0.383)
Reading ability at 7:								
5th quintile (highest)	0.186	(0.389)	0.175	(0.380)	0.278	(0.448)	0.242	(0.428)
4th quintile	0.218	(0.413)	0.204	(0.403)	0.234	(0.423)	0.218	(0.413)
3rd quintile	0.209	(0.407)	0.196	(0.397)	0.208	(0.406)	0.196	(0.397)
2nd quintile	0.210	(0.408)	0.209	(0.407)	0.169	(0.375)	0.168	(0.374)
1st quintile (lowest)	0.177	(0.382)	0.199	(0.399)	0.111	(0.314)	0.127	(0.333)

Continued next page....

Table .1 continued

Variable	Males				Females			
	Employed 2597 Obs		Total 3048 Obs		Employed 2363 Obs		Total 3894 Obs	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Father's social class 1974:								
Professional	0.045	(0.207)	0.041	(0.198)	0.042	(0.200)	0.040	(0.197)
Intermediate	0.150	(0.357)	0.142	(0.350)	0.146	(0.353)	0.144	(0.351)
Skilled non-manual	0.085	(0.279)	0.079	(0.270)	0.072	(0.258)	0.067	(0.250)
Skilled manual	0.315	(0.465)	0.318	(0.466)	0.314	(0.464)	0.305	(0.461)
Semi-skilled non-manual	0.011	(0.105)	0.010	(0.102)	0.012	(0.108)	0.011	(0.106)
Missing	0.089	(0.284)	0.092	(0.288)	0.088	(0.284)	0.096	(0.295)
No father figure 1974	0.037	(0.190)	0.041	(0.199)	0.055	(0.228)	0.054	(0.225)
Mother employed 1974	0.538	(0.499)	0.521	(0.500)	0.536	(0.499)	0.506	(0.500)
Bad finances 1969 or 1974	0.149	(0.356)	0.168	(0.374)	0.179	(0.384)	0.183	(0.387)
Bad finances missing	0.019	(0.137)	0.020	(0.141)	0.021	(0.143)	0.023	(0.149)
Father's interest in edn:								
Expects too much	0.014	(0.119)	0.014	(0.118)	0.008	(0.087)	0.008	(0.087)
Very interested	0.291	(0.455)	0.268	(0.443)	0.278	(0.448)	0.252	(0.434)
Some interest	0.243	(0.429)	0.235	(0.424)	0.222	(0.416)	0.209	(0.407)
Mother's interest in edn:								
Expects too much	0.035	(0.183)	0.032	(0.176)	0.024	(0.153)	0.024	(0.152)
Very interested	0.397	(0.489)	0.368	(0.482)	0.423	(0.494)	0.384	(0.487)
Some interest	0.389	(0.488)	0.384	(0.486)	0.375	(0.484)	0.362	(0.481)
Large employer 1991	0.231	(0.422)	0.211	(0.408)	0.183	(0.387)	0.157	(0.364)
Union member 1991	0.447	(0.497)	0.392	(0.488)	0.359	(0.480)	0.229	(0.420)
Private sector firm 1991	0.698	(0.459)	0.675	(0.469)	0.568	(0.495)	0.546	(0.498)
Local Authority Census demographics:								
% Unemployed/sick	4.609	(2.666)	4.574	(2.764)	4.722	(2.734)	4.488	(2.819)
% Professional/managerial	11.446	(6.278)	11.087	(6.407)	11.557	(6.528)	11.009	(6.743)
% Unskilled manual	6.783	(3.586)	6.700	(3.725)	6.864	(3.598)	6.562	(3.800)
% Owner occupiers	29.219	(19.626)	28.727	(19.860)	29.536	(19.689)	27.945	(19.964)
% Council tennants	42.779	(21.423)	41.551	(22.072)	42.712	(21.539)	40.861	(22.706)
Missing	0.107	(0.310)	0.127	(0.333)	0.103	(0.304)	0.144	(0.351)

Continued next page....

Table .1 continued

Variable	Males				Females			
	Employed 2597 Obs		Total 3048 Obs		Employed 2363 Obs		Total 3894 Obs	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Region 1991:								
North	0.061	(0.239)	0.061	(0.240)	0.056	(0.231)	0.055	(0.228)
North West	0.104	(0.305)	0.109	(0.312)	0.111	(0.314)	0.114	(0.318)
Yorkshire and Humberside	0.097	(0.296)	0.105	(0.307)	0.098	(0.297)	0.106	(0.308)
West Midlands	0.094	(0.291)	0.088	(0.284)	0.100	(0.300)	0.084	(0.278)
East Midlands	0.083	(0.276)	0.081	(0.273)	0.063	(0.242)	0.064	(0.244)
East Anglia	0.037	(0.189)	0.035	(0.184)	0.043	(0.203)	0.040	(0.197)
South West	0.077	(0.267)	0.075	(0.264)	0.090	(0.286)	0.083	(0.276)
South East	0.239	(0.427)	0.236	(0.425)	0.220	(0.414)	0.241	(0.428)
London	0.058	(0.233)	0.058	(0.233)	0.060	(0.238)	0.060	(0.238)
Wales	0.055	(0.229)	0.058	(0.235)	0.047	(0.212)	0.055	(0.228)
Scotland	0.095	(0.293)	0.093	(0.290)	0.113	(0.316)	0.097	(0.296)
Region 1974:								
North Western	0.102	(0.302)	0.106	(0.308)	0.117	(0.321)	0.115	(0.319)
North	0.074	(0.261)	0.076	(0.266)	0.068	(0.252)	0.070	(0.255)
East and West Riding	0.077	(0.266)	0.080	(0.272)	0.078	(0.268)	0.079	(0.270)
North Midlands	0.079	(0.270)	0.077	(0.267)	0.066	(0.248)	0.069	(0.253)
Eastern	0.079	(0.270)	0.076	(0.266)	0.074	(0.262)	0.079	(0.270)
London and South East	0.134	(0.341)	0.136	(0.343)	0.138	(0.345)	0.147	(0.354)
Southern	0.061	(0.239)	0.058	(0.234)	0.056	(0.231)	0.061	(0.240)
South Western	0.063	(0.243)	0.059	(0.236)	0.060	(0.237)	0.058	(0.234)
Midlands	0.089	(0.284)	0.083	(0.276)	0.094	(0.292)	0.075	(0.264)
Wales	0.056	(0.230)	0.058	(0.235)-	0.049	(0.216)	0.054	(0.226)
Scotland	0.099	(0.299)	0.097	(0.296)	0.112	(0.316)	0.097	(0.296)

Table .2: Detailed Qualification Wage Equations

Variable	Males Preferred OLS Specification Coef.	(S.E.)	Females Preferred OLS Specification Coef.	(S.E.)
Constant	1.676	(0.126)	1.617	(0.133)
Highest School Qualification 1981:				
CSEs	0.065	(0.024)	0.037	(0.027)
<5 O Levels	0.143	(0.024)	0.075	(0.026)
5+ O Levels	0.205	(0.029)	0.239	(0.033)
A levels	0.332	(0.033)	0.339	(0.036)
Highest Post-School Qualification 1981:				
Lower vocational	0.081	(0.019)	0.036	(0.022)
Middle vocational	0.088	(0.022)	0.073	(0.039)
Higher vocational	0.178	(0.026)	0.227	(0.030)
Degree	0.165	(0.030)	0.317	(0.037)
Maths ability at 7:				
5th quintile (highest)	0.122	(0.026)	0.007	(0.030)
4th quintile	0.070	(0.025)	0.010	(0.028)
3rd quintile	0.075	(0.024)	-0.006	(0.028)
2nd quintile	0.046	(0.024)	0.007	(0.026)
Reading ability at 7:				
5th quintile (highest)	0.081	(0.028)	0.115	(0.032)
4th quintile	0.084	(0.026)	0.094	(0.032)
3rd quintile	0.071	(0.024)	0.101	(0.030)
2nd quintile	0.067	(0.022)	0.083	(0.029)
Type of school 1974:				
Secondary modern	-0.003	(0.020)	0.009	(0.023)
Grammar school	-0.001	(0.027)	0.021	(0.029)
Private school	0.078	(0.038)	0.022	(0.043)
Other	-0.001	(0.067)	0.024	(0.068)
Father's years of education	0.009	(0.006)	-0.001	(0.006)
Father's education missing	0.104	(0.080)	0.055	(0.079)
Mother's years of education	-0.014	(0.007)	-0.014	(0.007)
Mother's education missing	-0.113	(0.090)	-0.158	(0.094)
Number of siblings	0.001	(0.007)	-0.011	(0.007)
Number of older siblings	0.000	(0.008)	0.019	(0.009)
Sisters only	0.004	(0.021)	0.015	(0.022)
Brothers only	0.020	(0.020)	-0.006	(0.023)
Father's social class 1974:				
Professional	0.049	(0.050)	0.212	(0.052)
Intermediate	0.063	(0.028)	0.065	(0.031)
Skilled non-manual	0.063	(0.030)	0.069	(0.034)
Skilled manual	0.026	(0.021)	0.038	(0.022)
Semi-skilled non-manual	0.099	(0.091)	0.042	(0.079)
Missing	-0.043	(0.063)	-0.106	(0.084)
No father figure 1974	0.048	(0.043)	0.011	(0.049)
Mother employed 1974	0.028	(0.017)	0.011	(0.019)
Bad finances 1969 or 1974	-0.024	(0.021)	-0.020	(0.023)
Bad finances missing	-0.034	(0.058)	-0.034	(0.061)
Father's interest in edu:				
Expects too much	0.030	(0.076)	-0.094	(0.120)
Very interested	0.003	(0.025)	-0.014	(0.025)
Some interest	0.003	(0.019)	-0.024	(0.021)
Mother's interest in edn:				
Expects too much	-0.013	(0.050)	0.053	(0.059)
Very interested	0.026	(0.025)	0.055	(0.026)
Some interest	0.037	(0.020)	0.036	(0.022)

Continued next page....

Table .2 continued

Variable	Males Preferred OLS Specification Coef.	(S.E.)	Females Preferred OLS Specification Coef.	(S.E.)
Large employer 1991	0.117	(0.016)	0.145	(0.020)
Union member 1991	0.034	(0.015)	0.187	(0.017)
Private sector firm 1991	0.024	(0.015)	-0.040	(0.018)
Local Authority Census demographics:				
% Unemployed/sick	-0.002	(0.005)	-0.001	(0.005)
% Professional/managerial	0.005	(0.002)	0.004	(0.002)
% Unskilled manual	0.007	(0.004)	0.003	(0.005)
% Owner occupiers	0.001	(0.001)	0.000	(0.001)
% Council tenants	-0.001	(0.001)	-0.001	(0.001)
Missing	0.122	(0.109)	0.071	(0.118)
Region 1991:				
North	-0.286	(0.054)	-0.334	(0.066)
North West	-0.214	(0.046)	-0.244	(0.052)
Yorkshire and Humberside	-0.235	(0.045)	-0.305	(0.049)
West Midlands	-0.189	(0.050)	-0.183	(0.056)
East Midlands	-0.200	(0.046)	-0.228	(0.058)
East Anglia	-0.173	(0.049)	-0.190	(0.054)
South West	-0.166	(0.046)	-0.274	(0.053)
South East	0.007	(0.034)	-0.129	(0.039)
Wales	-0.264	(0.063)	-0.278	(0.100)
Scotland	-0.160	(0.056)	-0.276	(0.058)
Region 1974:				
North Western	0.000	(0.045)	-0.002	(0.052)
North	-0.001	(0.049)	-0.055	(0.060)
East and West Riding	-0.056	(0.044)	-0.009	(0.051)
North Midlands	-0.013	(0.044)	-0.072	(0.056)
Eastern	-0.058	(0.039)	-0.127	(0.046)
Southern	-0.112	(0.035)	-0.119	(0.040)
South Western	-0.078	(0.045)	-0.087	(0.061)
Midlands	-0.068	(0.047)	-0.101	(0.056)
Wales	-0.001	(0.060)	-0.034	(0.104)
Scotland	-0.159	(0.056)	-0.081	(0.060)
Number of observations	2597		2363	
R^2	0.3411		0.4487	

Table .3: Returns to Qualifications using 1991 Measures: Males

Variable	OLS				IV			
	Typical Specification		Preferred Specification		Ordered Probit		Linear Probability	
	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)
Constant	1.927	(0.036)	1.676	(0.126)	1.637	(0.126)	1.639	(0.132)
Highest School Qualification 1981:								
CSEs	0.094	(0.027)	0.056	(0.026)	0.101	(0.067)	0.105	(0.049)
<5 O Levels	0.210	(0.024)	0.145	(0.024)	0.196	(0.048)	0.184	(0.038)
5+ O Levels	0.324	(0.028)	0.226	(0.029)	0.255	(0.195)	0.248	(0.043)
A levels	0.452	(0.029)	0.332	(0.032)	0.384	(0.122)	0.396	(0.047)
Highest Post-School Qualification 1981:								
Lower vocational	0.045	(0.020)	0.041	(0.020)	0.079	(0.064)	0.177	(0.066)
Middle vocational	0.075	(0.020)	0.063	(0.021)	0.120	(0.086)	0.072	(0.051)
Higher vocational	0.129	(0.026)	0.116	(0.026)	0.188	(0.111)	0.273	(0.047)
Degree	0.180	(0.030)	0.155	(0.030)	0.224	(0.121)	0.210	(0.042)
$\widehat{\lambda}_{Q_i^1}$					-0.053	(0.023)		
$\widehat{\lambda}_{Q_i^2}$					-0.052	(0.038)		
Hausman Tests (P-value):								
School Qualifications							5.483	(0.241)
Post-school Qualifications							23.215	(0.000)
Number of observations	2597		2597		2597		2597	
P-value 1991 regional dummies	0.000		0.000		0.000		0.000	
P-value 1974 regional dummies	0.079		0.046		0.035		0.085	
P-value ability variables			0.000		0.000		0.000	
P-value school type variables			0.375		0.383		0.395	
P-value family variables			0.184		0.320		0.394	
P-value parental interest			0.485		0.626		0.692	
P-value demographics			0.379		0.362		0.314	
P-value employer characteristics			0.000		0.000		0.000	
R^2	0.2824		0.3279		0.3386		0.3051	

Table .4: Returns to Qualifications using 1991 Measures: Females

Variable	OLS				IV			
	Typical Specification		Preferred Specification		Ordered Probit		Linear Probability	
	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)
Constant	1.580	(0.039)	1.573	(0.134)	1.580	(0.147)	1.636	(0.137)
Highest School Qualification 1981:								
CSEs	0.123	(0.029)	0.114	(0.027)	0.129	(0.044)	0.061	(0.052)
<5 O Levels	0.174	(0.026)	0.132	(0.025)	0.140	(0.042)	0.091	(0.038)
5+ O Levels	0.400	(0.031)	0.301	(0.033)	0.291	(0.123)	0.244	(0.044)
A levels	0.566	(0.033)	0.430	(0.036)	0.414	(0.089)	0.368	(0.048)
Highest Post-School Qualification 1981:								
Lower vocational	0.026	(0.021)	0.020	(0.020)	0.085	(0.107)	0.010	(0.047)
Middle vocational	0.109	(0.051)	0.061	(0.050)	0.141	(0.149)	0.103	(0.089)
Higher vocational	0.250	(0.027)	0.183	(0.027)	0.270	(0.145)	0.300	(0.044)
Degree	0.310	(0.037)	0.276	(0.037)	0.380	(0.139)	0.391	(0.052)
$\widehat{\lambda}_{Q_i^1}$					-0.020	(0.018)		
$\widehat{\lambda}_{Q_i^2}$					-0.062	(0.048)		
Hausman Tests (P-value):								
School Qualifications							8.317	(0.081)
Post-school Qualifications							19.678	(0.001)
Number of observations	2363		2363		2363		2363	
P-value 1991 regional dummies	0.000		0.000		0.000		0.000	
P-value 1974 regional dummies	0.239		0.112		0.132		0.168	
P-value ability variables			0.105		0.249		0.086	
P-value school type variables			0.833		0.850		0.926	
P-value family variables			0.006		0.009		0.005	
P-value parental interest			0.647		0.702		0.710	
P-value demographics			0.715		0.800		0.786	
P-value employer characteristics			0.000		0.000		0.000	
R^2	0.3660		0.4457		0.4508		0.4377	