Determinants and consequences of educational choices in the UK

by

Massimiliano Bratti

Thesis

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

University of Warwick, Department of Economics

December 2004

THE UNIVERSITY OF
WARWICK

Contents

List of Figures

List of Tables

Acknowledgments

I wish to thank my supervisors Prof. Mark Stewart and Dr. Jeremy Smith for their helpful guidance. I also wish to thank a number of other people who gave comments on the work included in this Thesis.

Early versions of chapter two benefited from presentations at the University of Warwick, the *ZEW Summer Workshop 2002 on Human Capital* (Mannheim, special thanks to the discussant Charlotte Lauer), the *European Society for Population Economics* 2002 Conference (Bilbao), the *European Economic Association* 2002 Conference (Venice) and from comments from Lorenzo Cappellari and two anonymous referees. The BCS70 data were kindly provided by and used with permission of the UK Data Archive (UKDA, University of Essex).

As to chapter three, I acknowledge both the Universities' Statistical Record (USR), as the original depositor, and the UK Data Archive for the use of the data set SN: 3456 Universities' Statistical Record. Part of this chapter was written when the author was visiting the Center for Economic Research (CentER) at Tilburg University as a Marie Curie Fellow within the ENTER exchange program. Funding from the European Commission is gratefully acknowledged. This chapter benefited from presentations at the University of Warwick, the *2003 IZA Summer School in Labor Economics* (Amersee), the *II Mediterranean Summer School in Theoretical and Empirical Economics 2003* (Palma de Mallorca), the *European Association of Labour Economists* 2004 Conference (Lisbon) and the *Third Labour Economics Workshop "Brucchi Luchino"* (Florence). I am particularly indebted to Jan van Ours, Qin Tu, Marton Csillag and Ian Walker for useful comments on early drafts of this chapter.

Chapter four represents joint work with Prof. Robin Naylor and Dr. Jeremy Smith and was partly written at CentER (Tilburg). Also in this case funding from the European Commission is gratefully acknowledged. I would like to thank Jan van Ours and seminar participants at CentER and the University of Pavia for helpful comments on an early draft of the chapter and the UKDA for data provision.

Funding from the ESRC is gratefully acknowledged.

The usual disclaimers apply.

Declaration

This Thesis is my own work and has not been submitted for a degree at another university. The fourth chapter represents joint work with my supervisor Dr. Jeremy Smith and with Prof. Robin Naylor. I contributed more than 50% to the analyses and the content of the fourth chapter.

Abstract

In this Thesis we study some aspects related to the determinants and the consequences of acquiring education in the UK. Chapter one outlines the structure of the Thesis.

In chapter two we analyse the probability of staying-on at school at age 16 in England and Wales using data from the 1970 British Cohort Study. A primary focus is on the effect of household income. The issue of income endogeneity is addressed using instrumental variable techniques. We find a statistically significant, but small, positive effect of income and stronger effects of long-term family characteristics, such as parental education, on the staying-on decision.

In chapter three we investigate social class influences on the probability of enrolling in different degree subjects in the UK during the period 1981-1991 using Universities' Statistical Record data. We consider three broad subject groups and estimate a trinomial probit model. Our results show the absence of social class differences in the period under study. Moreover, the analysis turns out to be robust to the use of a finer disaggregation of subjects and of a different econometric model (i.e. a flexible-thresholds ordered probit model).

In chapter four we estimate the log-wage premium to a first degree using data from the British Cohort Study 1970. We replicate the analysis in Blundell et al. (2000), who used data on the 1958 British birth cohort, and find evidence of declining returns to a first degree for women. We also investigate differences in premia by degree class and degree subject. We find evidence supporting the presence of such differences, although in many cases degree class and subject premia are not very precisely estimated, probably due to small sample size. We also consider the robustness of our results when taking account of the endogeneity of educational outcomes and of the possibility of heterogeneous treatment effects.

Chapter five briefly summarises the main findings of the Thesis.

Chapter 1

Introduction

Educational research has a long tradition in the UK and has often been used as an analytical support for policy interventions aimed at improving or reforming the educational system. There are many examples. In the 1980s several studies were published on the factors responsible for the low participation rates in post-compulsory education in the UK (see, for example, Pissarides, 1981, 1982, Rice, 1987, Micklewright, 1989, among others). The topic was considered of primary interest since it was a common thought that the performance of the UK economy was adversely affected by the low levels of skills and qualifications of its workforce (see, for instance, Finegold, 1993). More recently, the introduction of home students tuition fees and the replacement of means-tested maintenance grants with student loans followed the recommendations of the Dearing Report (Dearing, 1997). There is currently a lively debate on the opportunity of universities charging students top-up fees (which will be in place in England starting from 2006) and of introducing differential fees by subject (see Greenaway and Haynes, 2003). The main worry is that these recently introduced measures might adversely affect mainly poor students' participation in Higher Education (see Machin and Gregg, 2003) and reduce intergenerational mobility (see Dearden et al. 1997). This Thesis aims at contributing to this debate and to applied educational research in the UK by analysing some of the causes and consequences of individual educational choices.

We start in chapter two with an analysis of the individual decision to go on to post-compulsory education in England and Wales. In the UK an ambitious target of 50% participation in Higher Education (HE) for the 18-30 age group has been set for 2010. It is clear that to reach this goal it is necessary to increase the fraction of the population continuing beyond post-compulsory schooling. This analysis is important in order to assess the role of family background variables on students' school continuation decisions in a period of relatively low participation rates beyond post-compulsory education in the UK, and in particular the relative impact of household pecuniary and non-pecuniary characteristics. With respect to the existing literature we use a less exploited data set (the 1970 British Cohort Study) and seek to address several issues often neglected by previous studies, among which are survey non response and the endogeneity of family income. In more detail, Chapter two presents an empirical analysis of the factors affecting the choice of staying on at school at age 16 in England and Wales. The choice at 16 represents the first choice concerning enrollment in education that British individuals have to make. This decision also strongly affects their future academic and working lives. The issue was particularly topical in the UK where the staying-on rate at 16 in the late 1980s was relatively low (about 49% in 1987-88, DfEE 1999) compared with other OECD countries, and given the consequent effort of UK governments in the direction of widening access to both further and higher education. In this chapter we are mainly concerned with the influence of family background on the staying-on decision with a particular emphasis on the effect of parental income. The influence of parental income has a central importance since family finances, unlike child ability or household non-pecuniary characteristics, are the main determinant of children's education that policy makers can affect in the short-run through economic policy. In the US there has been a lively debate on the effect of household income and credit constraints on the demand for education. The findings of the most recent literature (see for instance Cameron and Heckman, 1998, 2001 and Carneiro and Heckman, 2002, for the US and Chevalier and Lanot, 2002, and Blanden and Gregg, 2004, for the UK) seem to attribute a marginal role to household income and point to children's ability and long-term household characteristics, such as parental education and socio-economic status, as the main forces driving educational decisions. The same conclusion has been reached by UK studies, although only a very limited number of them have tried to address the numerous problems arising in the empirical work (among which Blanden and Gregg, 2004). The main problems are the endogeneity of family income and the need to control for other household characteristics which may be correlated with income and affect a child's education. We seek to address this and other issues using data from the British Cohort Study

1970 (BCS70) to estimate a binary probit model of the school continuation decision at age 16. We seek to address the endogeneity problem using instrumental variables (IV). To the best of our knowledge there is only one other paper about the UK trying to tackle the issue of income endogeneity, Blanden and Gregg (2004).[1] Our chapter differs from Blanden and Gregg (2004) since while they exploit the longitudinal dimension of the BCS70 to identify "temporary" variations in household income, we try to identify exogenous, and not necessarily temporary, sources of variations in income using variables such as parents' industry of employment or grandfathers' social class, following the approach in Shea (2000) and Maurin (2002). Moreover, unlike Blanden and Gregg (2004), who matched income data from the Family Expenditure Survey, we transform the grouped income variable available in the BCS70 into a continuous variable using interval regression techniques and avoid the problems caused by imputation from an external source. Last but not least, we also investigate the problems potentially generated by non-response to the income question in the BCS70 (the missing rate was about 16% in the age 10 follow-up survey.). We obtain the following results. Firstly, in line with the previous UK literature, also when the endogeneity problem is tackled, the positive effect of income is statistically significant but of limited magnitude. The effect turns out to be higher for women than for men confirming early results by Rice (1987) and Micklewright (1989), who used different data sets. By contrast, non-pecuniary influences of long-term family

[1] Chevalier and Lanot (2002) simply control for a child's ability in the education equations but do not try to address the endogeneity problem.

characteristics such as parental education and social class or parental interest in a child's education, a measure of parenting quality, appear to be some important determinants of the choice of staying-on at 16. Secondly, income non-response does not seem to be an important issue for our analysis.

In chapter three, we contribute to the current debate on Higher Education (HE hereafter) by analysing the determinants of choice of field of study at the undergraduate level conditional on being enrolled at university. We consider the period 1981-1991. The period largely pre-dates the boom in university student numbers which took place during the early 1990s, the progressive replacement of means-tested maintenance grants with student loans and the introduction of home students tuition fees and offers therefore an interesting benchmark with respect to the actual design of the UK system of HE. We argue that the past characteristics of the UK HE system were likely to attenuate social class differences in subject choice, while the recent innovations, such as the substitution of maintenance grants with student loans, the introduction of top-up fees, and the eventual introduction of differential fees across subjects might have a different impact on students with different social class origin. We outline here the content of the chapter in more detail. After having studied family background influences on the probability of staying-on at 16 in the previous chapter, we study here the probability of entering different undergraduate fields of study conditional on enrolling in HE. The issue is especially important given the evidence of marked differences in graduates' labour market outcomes by degree subject. In this chapter, after giving some motivation for the importance of the topic, we review some of the reasons as to why social class may or may not be an important determinant of the choice of degree subject. The main hypotheses, put forward by both economists and sociologists, which can generate social class differences are that low social class students are more sensitive to economic returns and enroll in high return subjects or tend to enroll in technical subjects compared with high social class students, who are more likely to enroll in subjects providing 'cultural capital' or in which they have a comparative advantage in the probability of finding a good job. The main factors which would tend to have an 'equalising' effect are the absence of subject preferences heterogeneity by social class, and the equality of costs and benefits from enrolling in different degree subjects across social classes. The data set used is the Universities' Statistical Record. We report in the chapter a behavioural model that motivates the specific econometric model used, a trinomial probit, which allows for a non-zero correlation between the unobservables affecting the utilities of the different groups of subjects considered. We firstly consider a broad aggregation of subjects: Quantitative, Non-Quantitative and Medicine and Law. Our analysis on the 1981-1991 period does not generally support the existence of significant differences across social classes in the probability of enrolling in different subject groups, showing that in the period under study the design of the HE system in the UK appeared to grant equal opportunities in terms of access to different study fields (conditional on enrolling in HE) to individuals with different social class origins. In this chapter we also include a sensitivity analysis using a different econometric model and a finer grouping of subjects, which confirms the robustness of our findings.

In the final part of the Thesis we provide some evidence on the returns to education, in the narrow sense of log-wage premia, from the 1970 British cohort. The analysis is important since most of the evidence on the high private returns to HE in the UK is based on the 1958 British cohort. It is clear then that judgments on the opportunity of the introduction and the amount of top-up fees should be made on the basis of the private economic returns estimated from a more recent cohort of individuals. We also investigate differences by degree performance and degree subject. The analysis gives then some useful insights into the opportunity of charging students for a risky investment, given the ex-ante uncertainty on future degree performance which might deter more risk-adverse, typically less wealthy, individuals from enrolling in HE or in certain fields. In this case our analysis is also useful for the debate on fee differentiation by subject since it indicates what subjects ensure the highest wage returns. In particular, in chapter four we estimate the economic rewards to a first degree in terms of log-wage premia using the 1970 British Cohort Study (BCS70). We also consider the extent of variation of the wage returns by degree class and degree subjects. This turns out to be important to update the estimates on the return to a first degree obtained from the 1958 British cohort by Blundell *et al.* (2000) and to assess the magnitude of any variation in a first degree log-wage premium according to observed degree characteristics, such as performance and subject. Firstly, we replicate the analysis in Blundell *et al.* (2000), who used data from the National Child Development Study (NCDS), which refers to the 1958 British birth cohort, on BCS70 data and find evidence of a fall in the returns to a first degree for women. Secondly, we investigate differences in the return to a first degree by degree class and degree subject. We find some evidence of a higher wage premium to 'good' degrees compared with lower degree classifications and of differences in the returns to different degree subjects. Both for men and women, the subject group that ensures the highest return is Social Sciences while Art and Humanities ranks at the bottom. However, in many cases the effects are not very precisely estimated, probably due to small cell size. In this chapter we also assess the robustness of our results to considering educational outcomes endogenous, using the control function approach, and to allowing for heterogeneous treatment effects, using a propensity score matching-average treatment effect on the treated framework.

Chapter 2

Parents' income and children's staying-on at school in England and Wales: Evidence from the 1970 British Cohort Study

in the BCS70. In this respect, our analysis differs from other papers using the BCS70, such as Chevalier and Lanot (2002) who use dummies for income groups, and Blanden and Gregg (2004) who impute income from the Family Expenditure Survey (FES). Second, we address the issue of the endogeneity of income by using instrumental variables techniques in order to identify exogenous sources of household income variation. Our method differs from Chevalier and Lanot (2002), who simply control for child's ability, and Blanden and Gregg (2004), who use transitory variations in income to isolate the effect of household finances. Last but not least, we investigate the potential effect of income non-response on our estimates.

The chapter is organised as follows. The next section introduces the problems arising in empirical studies of the effect of family income on children's education. Section 2.3 briefly reviews past UK empirical literature. Section 2.4 describes the main features of our empirical analysis and our main findings. Section 2.5 concludes.

2.2 Empirical assessment of the effect of family income

There are several reasons why parental income may have a direct effect on children's educational attainment.[1] Following Becker's (1975) human capital investment theory, rich parents can invest more financial resources in their children's quantity of education. Parents' money can also be used to

[1] See Haveman and Wolfe (1995) for a comprehensive survey.

in the BCS70. In this respect, our analysis differs from other papers using the BCS70, such as Chevalier and Lanot (2002) who use dummies for income groups, and Blanden and Gregg (2004) who impute income from the Family Expenditure Survey (FES). Second, we address the issue of the endogeneity of income by using instrumental variables techniques in order to identify exogenous sources of household income variation. Our method differs from Chevalier and Lanot (2002), who simply control for child's ability, and Blanden and Gregg (2004), who use transitory variations in income to isolate the effect of household finances. Last but not least, we investigate the potential effect of income non-response on our estimates.

The chapter is organised as follows. The next section introduces the problems arising in empirical studies of the effect of family income on children's education. Section 2.3 briefly reviews past UK empirical literature. Section 2.4 describes the main features of our empirical analysis and our main findings. Section 2.5 concludes.

2.2 Empirical assessment of the effect of family income

There are several reasons why parental income may have a direct effect on children's educational attainment.[1] Following Becker's (1975) human capital investment theory, rich parents can invest more financial resources in their children's quantity of education. Parents' money can also be used to

[1] See Haveman and Wolfe (1995) for a comprehensive survey.

buy better educational quality, which may affect both current educational performance and children's future demand for education. Moreover, low income parents might push their children towards work in the labour market in order to contribute to family finances; in the absence of sufficient money transfers from their parents children from low income families may decide to work while studying (see Dustmann and Micklewright 2001), with possible negative effects on their school performance, or decide to quit education at the minimum leaving age to earn money and finance their own consumption.

When one wants to investigate the role of family income it is necessary to address several issues. The main problem is that family income may be endogenous with respect to a child's education. The endogeneity may arise from the correlation of some unobserved family factors with both family income and a child's education. Two such factors might be a child's inherited ability, which causes the so-called *ability bias* (see Griliches, 1977), and unobserved parental traits, such as parenting quality, which affect a child's education and are likely to be correlated with household income (*simultaneity bias*). The general 'solution' to these problems is the use of instrumental variables (IV) techniques which require the analyst to find variables uncorrelated with the unobserved child's or parental traits but correlated with parental income. In both cases, Shea (2000) and Maurin (2002), for instance, have suggested indicators of parents' industry of employment and financial capital as possible sources of valid instruments. Another problem is that income may be affected by a substantial measurement error and also in this case a possible remedy is the use of IV.

In this chapter we rely on IV techniques to find some exogenous sources of variation in household income which can allow us to identify its 'causal' impact on a child's education. Our analysis is therefore in the tradition of the empirical studies which have attempted to assess whether family income has a direct 'causal' effect on child's education and the magnitude of this effect using non-experimental data, some of which will be reviewed in section 2.3. As we estimate a reduced form model, we might only be able to identify the 'causal' impact of an unconditional income transfer but not the effect of conditional income transfers, such as Educational Maintenance Allowances (EMAs), which are transferred only to individuals who decide to go on in post-compulsory education. The assessment of the effect of EMAs in a regression framework requires a structural model that takes account of all costs and benefits related to the educational decisions, whose estimation would require much more information, which we do not have, and many assumptions, which we want to avoid. Moreover, the potential impact of EMAs has been the object of recent evaluation using experimental data, which has shown their effectiveness. Ashworth *et al.* (2002) using data from the first two years of implementation of the EMA pilot scheme find that in pilot areas EMAs have increased men's participation by 4.3 percentage points and women's participation by 3 percentage points when considering the whole population (i.e. both the eligible and ineligible population).

In this chapter, we only consider the British cohort born in 1970 using data from the BCS70 while we do not perform a comparison between the 1970 and the 1958 British cohorts. Such a comparison has already been undertaken by Blanden and Machin (2004) and Blanden and Gregg (2004) among others. In both articles the authors find an increase in the association between family income and a child's educational attainment between the 1958 and the 1970 British cohorts.

In our empirical analysis we include among the determinants of a child's education family income along with a set of characteristics relating to the family background, such as parents' education and social class. The rationale for including them is to separate out the non-pecuniary effects of long-term family characteristics and family income. The problem of *ability bias* should be already attenuated by the inclusion of a measure of child's ability at age 10. The *simultaneity bias* should instead be mitigated by the inclusion of a proxy for 'parenting quality' (see Mayer 1997), i.e. the interest of the parents in a child's education assessed by the teacher when the child was 10 years old, which Feinstein and Symons (1999) find as the most important 'parenting' variable. However, to fully address both problems, we use IV techniques in order to identify exogenous sources of variations in income uncorrelated with a child's ability and parenting quality. On the grounds that the chosen instruments are uncorrelated with measurement error in parental income, IV also help to address the problem of measurement error.

2.3 Past empirical literature

This section presents a brief survey of past empirical work concerning the role of family income and other family characteristics on a child's education using UK micro-data.[2] As we said, a primary focus of this chapter is on the assessment of the causal effect of parental income. In this respect, all studies that we review here have included family income among the explanatory variables although some have not attempted to identify the causal impact of income and have only estimated simple correlations. Here we report only the main findings of these papers concerning the effect of household income and other household characteristics of particular interest.

Rice (1987) estimates the demand for post-compulsory education in the UK using data from the 1976 Family Expenditure Survey (FES). She includes, among the other factors, controls for family income and social class, but due to data unavailability, she does not include either parental education or child ability among the explanatory variables. For this reason, due to the effect of unobserved innate ability and omitted proxies for parenting quality, we expect an overestimate of the effects of income and social class. Her findings show marked differences in the probability of undertaking further education by social class but a significant effect of income only for females, with the effect being greatest for individuals with low income.

Micklewright (1989) analyses the school continuation decision in England and Wales using the National Child Development Study (NCDS), which collects data on the 1958 British cohort. The decision to go on to further edu-

[2]There are also some papers studying staying-on decisions and demand for university education using macro data, such as Pissarides (1981, 1982). Moreover, this topic has been widely studied in the US. Recent contributions includes Mayer (1997), Cameron and Heckman (1998, 2001), and Carneiro and Heckman (2002), among others. All these papers find a small effect of family income on children's education.

cation after reaching the minimum leaving age was taken by individuals from the 1958 cohort in 1974. A major advantage of the NCDS with respect to the FES is that the former has an indicator of a child's ability which may contribute to reducing the problem of ability bias. Micklewright (1989) includes among the explanatory variables the sum of age 16 maths and comprehension test scores. The author finds an important role for parents' education, social class, and child ability for the demand of post-compulsory education for both men and women. He also finds an insignificant effect of income for boys irrespective of whether or not controls for ability and school type are included. By contrast, the effect of income for girls is statistically significant but small: a one-standard deviation reduction in gross weekly income (20 pounds) reduces the probability of continuing by 4.5 percentage points from an average value of around 50%. Since Micklewright (1989) builds a continuous measure of income by taking the mid-points of the 11 bounded ranges reported in the NCDS and a measure of median income for the top unbounded range, he states that his income measure may be problematical and suffer from considerable measurement error. For this reason he replicates the analysis using a dummy for household's bad financial conditions. However, in the specification controlling for child ability and school type the dummy variable for bad financial conditions is not statistically significant neither for men nor for women.

Dearden (1999a) estimates an education equation (the years of full-time education) for men using the NCDS. The author includes controls for child ability and some proxies for parenting quality, such as parental interest in child's education. Dearden (1999) finds a strong effect of parental education and social class and a statistically significant negative effect of the proxy for financial difficulties on children's educational achievement.

Ermisch and Francesconi (2001a) estimate the determinants of educational attainment (the highest educational qualification achieved) using data from the first seven waves of the British Household Panel Study (BHPS) and an ordered logit model. The authors do not estimate separate models by gender. They find a significant effect of parental education, with a stronger effect for mother's education. As to family income, there are only statistically significant differences between individuals with income in the bottom quartile and the other quartiles. The probability gap between the first and the fourth quartile is for instance 3.2 percentage points. The effect is, therefore, small in magnitude. It must be noted that unlike the previously reviewed articles, Ermisch and Francesconi (2001a), due to data unavailability, do not include controls for either parental social class or a child's ability. Then, the estimated effects of parents' income and education may partly reflect non-pecuniary aspects related to their socio-economic status and unobserved ability.

Chevalier and Lanot (2002) estimate the determinants of the age at leaving education using both the NCDS and the BCS70 and ordered probit models. The authors try to address the issue of income endogeneity by controlling for a child's ability and family contextual factors, such as parental interest in children's education. Chevalier and Lanot (2002) do not include a continuous measure of income but include dummies for income groups. They find a strong effect of parental education and social class in both cohorts and for both genders. Also the effect of parental income is statistically significant, although the marginal effects computed at the sample means are rather small, especially for the BCS70. The effect of family financial resources is therefore dominated by that of other non-pecuniary family characteristics.

Blanden and Gregg (2004) estimate the effect of family income on children's educational outcomes using several data sets, namely the NCDS, the BCS70 and the BHPS. A special emphasis is given to the estimates from the BCS70 since two measures of parental income are available, at age 10 and age 16, and the data set also includes measures of a child's ability. The identification strategy of the authors relies on exploiting the longitudinal dimension of the BCS70 by simultaneously using information on family income at age 10 and age 16, which enables them to identify the effect of transitory components of income. The authors use two identifying strategies: the first includes both measures of income in the regression, while the second includes only the change in income between age 10 and age 16. The estimated effects of a one third reduction in the level of income are of reducing the probability of staying-on by 3.9 and 0.9 percentage points, respectively, when applying the two strategies above. As observed by the authors the small effect found when using the second method may be due to the measurement error that affects short run variations in income. Blanden and Gregg's (2004) identifying strategy rests on the assumption that changes in income between age 10 and age 16 are exogenous with respect to a child's education. This means that the education regressions must control for all factors which may simultaneously affect family income changes between age 10 and 16 and a child's education. Some examples of such factors are a child's health or behavioural problems that might induce some parents to withdraw from the labour force to devote more time to their children, affecting, therefore, both family income and a child's education. In their study the authors build a continuous measure of family income from the grouped variable by matching families in the BSC70 with similar families in the FES and taking the median income within each band. This procedure has the advantage of reducing the potential problems of income endogeneity since the unobserved factors affecting both family income and a child's education are 'averaged' across families with similar observed characteristics coming from another data source. However, the accurateness of the imputation procedure cannot be checked and its effect on the estimates of the effect of family income on a child's education is unknown. For this reason it may be useful to compare Blanden and Gregg's (2004) findings with estimates based on a different procedure that we implement in the current chapter.

2.4 The empirical analysis

The following subsections describe the data, the explanatory variables and the econometric model used in the empirical analysis to estimate the staying-on decision model.

2.4.1 Data description

In this chapter, we use data drawn from the BCS70. The BCS70 began in 1970 when data were collected on the births and families of 17,198 babies born in England, Wales, Scotland and Northern Ireland from the 5^{th} to the 11^{th} of April.[3] There are currently five complete follow-up surveys available: 5-year, 10-year, 16-year, 26-year and 30-year. In order to carry out our analysis we need the family income and the age at leaving continuous full-time education of individuals included in the BCS70. For this reason we use a matched sample of individuals who replied to both the 10-year and the 30-year follow-up surveys. The choice of the two waves needs some explanations. We focus on the 10-year follow-up for income and other family and individual characteristics since the 16-year follow-up has a much higher rate of income non-response.[4] We need to use data from the 10-year follow-up in any case, for a child's ability and would have therefore a high number of unmatched individuals between the 10-year and the 16-year follow-ups and, therefore, with missing ability scores. The dependent variable, a dummy for being enrolled in education at age 16, is derived from the age at leaving full-time

[3] Subjects from Northern Ireland were included in the birth survey, but have been excluded from all subsequent sweeps.

[4] In a previous version of the chapter we used family income and background variables at age 16. However, in the present version we follow the suggestion of an anonymous referee and use the 10-year follow-up. The rate of income non-response is 15.7% in the 10-year follow-up and 33.9% in the 16-year follow-up. Data from the 10-year follow-up are also used in the analysis of Chevalier and Lanot (2002).

education coming from the 30-year follow-up.[5] Since educational data are retrospective, the dependent variable may be affected by measurement error. In order to have an idea of the impact of panel attrition on the composition of the sample used in the estimates, we report in Table 2.1 for various samples the means of some family background variables that we consider particularly important in our analysis, father's and mother's years at leaving education and father's social class from the first wave. The table shows that panel attrition seems to be random with respect to some important aspects of family background. In the empirical analysis we focus only on individuals residing in England and Wales and drop individuals residing in Scotland on the grounds that the Scottish educational system is substantially different.[6]

2.4.2 Explanatory variables

In this section, we describe some of the explanatory variables included in the econometric model. In order to ensure comparability of results, for the spec-

[5]Some alternative sources for this piece of information are the 16-year follow-up that contains a question on whether the teenager is (or about to be) continuing his/her education. However, the response rate is only 61% (the information is available only for 7,073 individuals on a total sample of 11,615 individuals). By contrast the information on the age at leaving full-time education from the 30-year follow-up is available for 99.80% of the individuals not still in education, i.e. for 10,930 individuals on a total sample of 10,952 units. Also the 26-year follow-up provides these information, but the main problem is that the wave suffered from a high rate of panel attrition and collected information only on 9,003 individuals.

[6]The same sample selection criterion has been followed also by Micklewright (1989) and Chevalier and Lanot (2002) among others.

ification of the education equation we generally follow the previous literature and adopt a specification similar to those reviewed in section 2.3. When not stated otherwise, all variables are measured when the child was 10 years old.

Among the covariates in the education equation we include some proxies for quantitative and verbal ability (British Ability Scales score), personal characteristics (ethnicity, region of residence), long-term family characteristics (parents' education and social class, mother's type of employment), other household characteristics (home ownership), a proxy for parenting quality (parental interest in child's education) and family income. We describe, here, only some variables of special interest:

- *British ability scales (BAS).*[7] This is an indicator of 'child quality'.[8] The BAS score is computed according to performance on a questionnaire that assesses both verbal and mathematical ability. The BAS score is computed at age 10, so as to reflect mostly innate ability and early parental inputs. We include a dummy for individuals with a missing BAS score. The inclusion of ability scores enables the researcher to reduce the problem of so-called *ability bias*. Indeed, in the case that the ability of a child is unobserved, the positive relationship between family income and a child's education may be only spurious and driven by the correlation between a child's ability and his/her parents' ability, which affects their income.

[7]See Elliot *et al.* (1979).

[8]According to Becker (1975), for instance, mature individuals decide on the total amount of investment in education on the basis of their parents' earlier choice of investment in 'child quality'.

- *Parents' education* and *social class.* We include both mother's and father's education and an individual's social class determined as the social class of the father, if he is present, or that of the mother if the father is not present. These are long-term family factors. The impact of parents' education and social class on children's educational outcomes may operate through several channels: it may affect the taste for education (non-pecuniary effects), the 'quality of parenting',[9] or be a proxy for permanent income, affecting the consumption and investment demand for education. By simultaneously including both long-term family characteristics and family income we estimate the effects of these factors free from cross-correlations. In this sense the effect of long-term family factors captures non-pecuniary effects, while that of income the pecuniary effect net of the impact of long-term family factors such as parental education and social class that are likely to affect parents' permanent income.

- *Parents' interest in children's education.* We include dummies for the level of parental interest in children's education as assessed by the teacher when the child was 10 years old. There are five possible levels: interested, moderately interested, scarcely interested, not interested, cannot say. We aggregate the second and third categories, due to small cell size, and build an indicator of parental interest in child's education

[9] See for instance Datcher-Loury (1988), who finds that greater time devoted to child care by highly educated mothers raises children's years of schooling.

by taking the maximum between father's and mother's interest.[10] Feinstein and Symons (1999), for instance, find in their study of secondary school attainment that the quality of 'parenting' is more important than parental education, and that the most important parental input is interest in children's education.

We also include dummies for the region of residence to assess the impact of regional factors such as school quality and local labour market conditions, and home property as a proxy of parental wealth.

The matched sample includes 3,888 men and 4,186 women. Tables 2.2 and 2.3 report means and standard deviations for the samples and the variables that we use in our preferred specification reported in section 2.4.3. It is immediate to see the high percentage of students in our samples who quit full-time education at age 16 (i.e. in 1986), 61% of men and 50% of women.

Blau (1999) claims that some of the explanatory variables included in previous studies, such as parents' education, should be excluded from a reduced form model of the demand for education because, although predetermined, they are potentially endogenous. However, our personal point of view is that in the context of the estimation of education equations the main source of endogeneity which may affect parental variables such as parents' education and social class is the correlation between parents' and children's academic abilities and that once the latter is included or proxied in the education equation the problem is greatly attenuated. For this reason and in order to make our results comparable with the past UK literature, which generally controls for

[10]To avoid multicollinearity problems since the two variables are highly correlated.

parental education and socio-economic status, we include in the staying-on model these long-term parental characteristics. A way to fully address the problem would be to use IV methods also for these variables, which would require finding other valid instruments in addition to those used for parents' income.[11] However, following Blau (1999) and Maurin (2002), we do not include in the education equation variables related to the family structure.[12] The reason is that we want to exclude choice variables that might be determined by parents jointly with children's education. Indeed, the economic theory of fertility (see Becker, 1981) states that the quantity of children is jointly determined with their quality, e.g. their level of education, and both are influenced by family financial resources. For the same reason we exclude from the regressors school type at age 10.

[11] Chevalier (2004), for instance, uses changes in the minimum school leaving age in the UK as an instrument to identify the effect of parents' education on children's education. He finds that both mother's and father's education have a positive effect on a child's education and that this effect is direct and robust to the inclusion of several control variables. Moreover, in most cases in his analysis the endogeneity of parental education is rejected.

[12] The role of family structure on children's achievement has been investigated in detail in Ermisch and Francesconi (2001b).

Family income

The BCS70 reports gross weekly parents' income in categorical form.[13] It may be convenient to use a continuous measure of household income rather than the grouped variable for at least two reasons, one statistical and the other practical. Firstly, the use of a continuous income variable rather than income groups increases the precision of the estimate of the effect of income. Secondly, the BCS70 groups weekly household income into classes of 50 pounds sterling. This means that if income groups are used, differences in educational attainment can only be assessed between individuals falling into different groups. However, policy makers are likely to be interested in finer policy interventions. Indeed, for individuals with household income at the lower bound of each group a 50 pounds increase in income would be necessary to change group. 50 pounds at the 1980 value are equivalent to about 124 pounds at the 1999 value. The Educational Maintenance Allowance (EMA) pilot scheme started in September 1999 in the UK provided 16-18 years old students with a financial allowance of 30 or 40 pounds per week (for annual family income less than 13,000 pounds), depending on the piloting area, if they remained in full-time education after year 11 (see Chevalier and Lanot 2002). Then, it is evident that a transfer of 50 pounds per week at the 1980 value, for instance, would largely exceed the amount of income transferred within the EMA pilot scheme.

[13]Parents income is combined gross income of child's mother and father (excludes child benefits but includes all other earned and unearned income before deductions of tax, national insurance, etc.).

In order to obtain a continuous income variable, we follow Stewart (1983), using interval regression techniques to predict parents' income. Appendix A briefly reviews the method used to estimate the income equation. Hence, our study differs from the other studies that have used the BCS70, such as Chevalier and Lanot (2002) who use grouped income and Blanden and Gregg (2004) who use income from an external source (the FES). Moreover, in order to address the issue of income endogeneity through IV we need to include in the income regression some variables which are excluded from the education equation, i.e. that do not have a direct effect on the staying-on decision. We follow some exclusion restrictions used in the previous literature: parents' industry (Shea, 2000, Maurin, 2002), and grandfathers' socio-economic status (Maurin, 2002). According to Shea (2000) industry wage premia reflect rents rather than unobserved ability differences. Evidence in favour of the validity of this exclusion restriction is provided by some of the literature on interindustry wage differences. First, estimates of the amount of the differences which can be accounted for by unmeasured workers' ability generally leave a substantial part of the differences unexplained (see for instance Krueger and Summers 1988, Katz and Summers 1989). Moreover, the literature has shown that wage premia are higher in industries with higher profits, and that industries that pay higher wages do so in all occupations, findings that are difficult to reconcile with the *unobserved ability hypothesis*.[14] However, as Shea (2000) acknowledges there are other studies supporting the unob-

[14]Which states that some industries pay higher wages to attract workers with higher ability. For a brief review see Romer (1996).

served ability view, such as Murphy and Topel (1990) among others. Maurin (2002) uses another instrument, grandfathers' socio-economic status. The variable is expected to be highly correlated with family income, due to the well established intergenerational correlation between parents' and children's incomes or socio-economic status (SES). In this case the identifying assumption is that grandfathers' SES has no direct effect on children's educational outcomes but only an indirect effect through the impact on family income. Studies which have used IV have generally assumed the validity of one instrument and tested the validity of the other instruments. Shea (2000, p. 162) assumes that his instruments (industry and union wage premia, job loss) are uncorrelated with unobserved ability. Maurin (2002, p. 322) assumes that grandfathers' past socio-economic status is a valid instrument and tests the validity of father's industry, which turns out to be also a valid instrument. In the current chapter we follow a different approach, we test for the potential validity of our candidate instrumental variables by implementing the tests suggested by Bound *et al.* (1995), i.e. testing the significance of all instruments in both the income equation and the education equation that uses instrumented income. In the model for parental income we include as explanatory variables: parental ethnic groups, education, social classes and industries (one-digit SIC 80), the region of residence and grandfathers' social class.[15] The results of the parental income interval regression are reported in Table 2.4. The variables included are generally highly significant. Our estimates show significant variation in parental income by industry and

[15]This last piece of information comes from the 5-year follow-up survey.

grandparents' social class. The Wald tests reported in Table 2.5 show that father's and mother's industries and grandfathers' social classes are all statistically significant determinants of parental income at the 1% level. The estimated effects of the explanatory variables are generally in the expected direction: parental income increases with parental education and social class, is lower for ethnic minorities and is positively related to grandfathers' social class. The Pseudo R^2 of the model is 17.75%.

2.4.3 Econometric model and results

Although the percentage of observations with missing income in the 10-year follow-up is much lower than that in the 16-year follow-up, our estimates may still be affected by the so-called sample selection bias (see Heckman, 1979). For this reason, we decide to model non-response to the income question. The estimated model is therefore described by the following two equations:

$$S_i^* = \alpha ln(\hat{y}_i) + X_i\beta + u_i \tag{2.1}$$

where S_i^* is a latent variable measuring an individual's demand for post-compulsory education. We only observe the binary variable S_i assuming value 1 if the individual i leaves education at age 16 and zero otherwise;[16] X_i

[16]Like most of the previous literature, in this chapter we model the school continuation decision as a dichotomous variable. In our case, we are mainly interested in the role of family background variables on the probability of staying on at 16. However, these variables may have a different impact on the different choices available at sixteen. For instance, Andrews and Bradley (1997) analysed the transition from school by modeling six possible choices (two types of schooling, youth training, two forms of employment

are observable characteristics and u_i unobservables affecting the staying-on decision; $\hat{y}_i$ is family income, which has been predicted using in the first stage the interval regression,[17] and α our parameter of interest; while the second equation is

$$R_i^* = W_i\gamma + v_i \tag{2.2}$$

where R_i^* is a latent variable measuring the propensity to answer to the income question. We only observe the binary variable R_i assuming value 1 if family income is non-missing and zero otherwise. W_i is the vector of observable characteristics affecting the choice to answer to the income question. We assume that these factors are the same factors affecting the level of income that are included in the interval regression estimated in the previous section. v_i is a vector of unobservables affecting the probability of answering to the income question.

The sample selection bias arises from the potential correlation between the error terms of the two equations u_i and v_i. Since in income surveys higher non-response is sometimes observed at high income levels, one possible source of selection could be due to unobserved parents' earnings ability that is not controlled for by our included covariates. If parents' earnings ability

---

and unemployment) and using a multinomial logit model. They found that the different outcomes could not be aggregated. In our case it is not possible to model the choice in a similar way given that data on the type of transition are provided only in the 16-year follow-up and contain a high number of missings (39%).

[17]We include the natural logarithm of income rather the level of income following the suggestion of an anonymous referee and using a specification close to Blanden and Gregg (2004).

positively affects income response and negatively affects a child's probability to leave education, for instance, we would observe a negative correlation between the error terms of the two equations.

Assuming that u_i and v_i are jointly normally distributed with correlation ρ, the model constituted by equations (2.1) and (2.2) can be estimated with a probit model with selection (see van de Ven and van Praag 1981) using maximum likelihood.[18] The effect of income in the education equation is estimated through a two-stage procedure and, therefore, to identify the effect of parental income we need some variables affecting schooling decisions only through their influence on income.[19] We have already seen in section 2.4.2 that our potential identifying instruments are parents' industry and grandfathers' social class. For this reason to have a basic idea of their validity we estimate the probit model with selection without exclusion restrictions and perform some Wald tests for the significance of the instruments in the education equation using instrumented income.[20] The tests are reported in

[18]If the correlation ρ is not statistically significant the education equation, i.e. equation (2.1), can be consistently and efficiently estimated using a simple binary probit model.

[19]More precisely the model is formally identified also in the absence of exclusion restrictions because of the non-linearity of predicted income in the explanatory variables (see Appendix A). However, using exclusion restrictions is important to avoid that identification rests only on functional form.

[20]As we said the model using predicted income from the interval regression is virtually identified also in the absence of exclusion restrictions (due to the non-linearity of the income prediction in the explanatory variables, see Appendix A.). The same is true for the probit model with selection. Standard errors have been computed with bootstrap since the income variable has been generated from the first stage income interval regression for household income.

Table 2.5.[21] The results show that for men father's and mother's industries and mother's father social class can be dropped from the education equation. A possible reason for the significance of father's father social class is that it may be a proxy for wealth that will be inherited by the household, or it may reflect some residual child's unobserved ability uncorrelated with parents' ability since some genetic characters may appear only in some generations. For women grandfathers' social classes can be dropped. In this case parents' industry does not appear to be a valid instrument. Since the same variable is not significant for men, we think that the effect on children's education should not reflect unobserved ability but parental tastes towards female work. For instance, parents working in female dominated sectors might be more favourable to female work and female education.[22] We follow the indications of these tests to specify the models with exclusion restrictions. For men, the estimates of the probit model with selection are reported in Tables 2.6 and 2.7. Since the correlation in Table 2.7 is not significant we observe that there is no evidence of a sample selection bias and estimate accordingly a simple probit model. Table 2.10 reports the results for the probit model. The coefficient on parents' income is negative and statistically significant at the 5% level. Therefore, parental income has a significant negative effect on quitting full-time education at age 16.

[21]We do not report here the full set of estimates for the models without exclusion restrictions but they are available upon request from the author.

[22]Indeed, from Table 2.11 we will observe that a father working in sectors such as Distribution, Banking and Other services where the presence of female workers is higher than in the reference group (Metal goods) raises the staying-on probability.

Also for women the correlation between the error terms of the two equations is not statistically significant, i.e. we find no evidence of a sample selection bias (see Tables 2.8 and 2.9). The estimates of the simple probit model are reported in Table 2.11, which shows that the coefficient of parental income is negative and significant at the 1% level.

Although parental income has a significant effect on the school continuation decision for both males and females it is necessary to compute marginal effects to have an idea of the magnitude of its effect on the probability of quitting full-time education at the minimum leaving age. For this purpose we report in Table 2.12 the effect of a one standard deviation increase in income (58.4 pounds per week) and an increase equivalent to the 1999 maximum of the EMA scheme on the probability of quitting education at the sample mean and at various levels of the quitting probability. The marginal effects show a limited impact of parental income on the schooling decision. Due to the shape of the normal density function the marginal effect is the highest at the 50% probability of quitting education, where a one-standard deviation increase in income reduces the probability of leaving education at 16 by 2.8 percentage points for men. The effect of income is lower for 'marginal' and advantaged students, i.e. for the students who have very high and very low probabilities of quitting education at 16, respectively. The same is observed for women for which the corresponding marginal effect at a 50% probability of quitting is 3.5 percentage points. Therefore, our analysis, in line with Rice (1987) and Micklewright (1989), shows a stronger effect of parental income on daughters' than on sons' probability of staying-on. The marginal effects for the 1980 equivalent of the 40 pounds, which represents the maximum amount of the EMA in 1999 (about 16 pounds at the 1980 value), on the probability of staying-on are much smaller. However, as we said, it must be noted that we are considering the effect of a non-contingent income transfer while EMAs are conditioned on the decision to continue in education and their effect is likely to be much stronger, especially on less advantaged students. Therefore, our results are not necessarily in contrast with the recent evidence provided on the effectiveness of the EMAs (Ashworth *et al.* 2002).

In line with the previous literature, our analysis shows a substantial effect of family long-term characteristics, such as education and social class, and early parental and school inputs, such as parents' interest in a child's education on children's staying-on probability. Men (women) with fathers with a degree are for instance about 17 (15) percentage points more likely to stay-on than individuals whose fathers have less than O-level qualifications,[23] while those with mothers with a degree have a probability premium of 30 (21) percentage points. Individuals from social class I (professionals) are about 17 (10) percentage points more likely to continue education than those from social class IIIM (skilled manual workers). Men (women) whose parents showed very little or no interest in education are about 13 (23) percentage points more likely to quit full-time education at the minimum leaving age than individuals whose parents were very interested in education. Since in the probit models we control for parents' income and home property (as a proxy for wealth) these effects are free from the correlation with pecuniary

[23]Ordinary level is the lower secondary education qualification in the UK.

factors and have a non-pecuniary nature.

Also the effect of child ability is remarkable. A one-standard deviation increase in the BAS quantitative score reduces the quitting probability by 5.4 percentage points for men and 6.2 for women. The same variation in the BAS verbal score reduces the stopping probability by 6.3 percentage points for men and 6.1 for women.

2.5 Concluding remarks

In this chapter we have investigated the effect of parental income on children's staying-on probability at school at age 16 in England and Wales using data from the 1970 British Cohort Study (BCS70). With respect to the previous literature the main innovations we have introduced are the following. Firstly, unlike previous authors who have used the BCS70 to study the school staying-on probability, we do not use dummies for income groups (such as Chevalier and Lanot 2002), or impute income from an external source (such as Blanden and Gregg 2004) but we obtain a continuous income measure from the grouped income in the BCS70 using interval regression. This is important since it increases the precision of the estimates and avoids the potential bias of the imputation technique, respectively. Secondly, in this chapter we seek to address the issue of income endogeneity using Instrumental Variables techniques. In particular, unlike Chevalier and Lanot (2002), who try to attenuate the endogeneity problem by simply including in their regressions a measure of child ability, we seek to identify an exogenous source of variation in family income in the spirit of Shea (2000) and Maurin (2002). For this purpose we use parents' industries and grandfathers' social class. We find that mother's father social class and mother's and father's industries are valid instruments for men and grandfathers' social classes are valid instruments for women. Our identifying strategy differs from Blanden and Gregg (2004), who use the longitudinal dimension of the BCS70 to identify the transitory component of parental income. Thirdly, unlike previous studies we address the issue of income non response in the BCS70 by estimating a probit model with selection, but we find no evidence of a sample selection bias. Our main results are as follows. Like the other two studies that used the BCS70 we find a statistically significant positive effect of parental income on a child's staying-on probability at age 16. However, the magnitude of the effect is relatively small, and the impact of income is stronger for women. This last result on gender differences confirms early findings by Rice (1987) and Micklewright (1989) who both used NCDS data. By contrast, other family characteristics such as social class, parental education and early parental interest in a child's education have much stronger effects on the probability of staying in full-time education at age 16. Our results confirms, therefore, the findings of the US literature and those of past UK studies of the predominance of family non-pecuniary over pecuniary influences on children's education.

Appendix A: From grouped to continuous family income

In the BCS70, family income (i.e. parents' income) is observed in a certain interval on a continuous scale. We want to transform the grouped variable into a continuous one. The procedure has been investigated by Stewart (1983). We summarise here only the main features of the problem and the proposed solution. The latent structure of the model under consideration is given by:

$$y_i = z_i'\delta + p_i'\phi + \epsilon_i \tag{2.3}$$

where y_i is the latent family income of individual i, which falls within a certain interval of the real line (A_{k-1}, A_k). z_i and p_i are vectors of regressors affecting family income and δ and ϕ vectors of unknown parameters to be estimated, respectively. z_i represents the variables excluded from the education equation (i.e. the identifying instruments) that we estimate in a second stage using predicted income from equation (2.3). ϵ_i's are i.i.d. normally distributed random disturbances with zero mean and variance σ^2 and are assumed to be independent of z_i and p_i.

Ad hoc procedures, such as assigning to each individual the midpoint of her income group, do not in general result in consistent estimates of the parameters δ and ϕ, while consistent estimates can be obtained by assigning to each observation its conditional expectation:

$$\hat{y}_i = E(y_i|A_{k-1} < y_i < A_K, z_i) = z_i'\delta + p_i'\phi + \sigma\left[\frac{\phi(Z_{k-1}) - \phi(Z_k)}{\Phi(Z_k) - \Phi(Z_{k-1})}\right] \quad (2.4)$$

where $Z_k = (A_k - z_i'\delta - p_i'\phi)/\sigma$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative distribution functions.

Stewart (1983) suggests several ways to estimate the parameters of interest δ, ϕ and σ.

In our specific case the parameters are estimated using a maximum likelihood estimator.[24]

After estimating δ, ϕ and σ consistently, it is possible to obtain predicted values for y_i, i.e. a continuous measure of family income.

This measure is used in a second stage for the estimation of the education equations (the probit models of staying-on).

[24]See 'methods and formulas' for the TOBIT command in the Stata Reference Manual (Stata 2003).

Tables

Table 2.1: Some descriptive statistics for different samples (BCS70)

Variable	first wave	10-year follow-up	30-year follow-up
Father's age at leaving education	15.84	15.84	15.94
Mother's age at leaving education	15.59	15.60	15.67
Father's social class	%	%	%
Not known	0.56	0.55	0.47
I	4.77	4.74	5.21
II	11.08	11.16	11.86
IIINM	11.19	11.64	12.55
IIIM	43.87	45.25	44.77
IV	14.38	14.35	13.75
V	6.44	6.07	5.32
other	2.92	2.60	2.75
not applicable	4.79	3.64	3.32
Wave size	17,197	13,700	10,389

Note. This table reports some descriptive statistics for the variables of the first wave in the sample at birth and in the age 10 and age 30 follow-ups. Some abbreviations commonly used in the UK for social classes are: I (professional), II (intermediate), IIINM (skilled non manual), IIIM (skilled manual), IV (partly skilled), V (unskilled).

Table 2.2: Sample summary statistics, men (BCS70)

Variable	Mean	s.d.
Quits at 16	0.603	0.489
Ln(income)	4.771	0.486
Child's ethnicity (white)		
Non-white	0.035	0.183
Father's education (O-level)		
< O-level	0.151	0.358
A-level	0.085	0.279
Professional or other	0.031	0.174
Degree	0.123	0.329
Missing	0.026	0.160
Father not present	0.069	0.253
Mother's education (O-level)		
< O-level	0.197	0.397
A-level	0.038	0.191
Professional or other	0.052	0.221
Degree	0.030	0.171
Missing	0.007	0.082
Mother not present	0.060	0.237
Social class (IIIM)		
I	0.060	0.237
II	0.224	0.417
IIINM	0.088	0.284
IV	0.106	0.308
V	0.028	0.164
OLFU[a]	0.034	0.182
Missing	0.051	0.221
Parental interest in child's education (very interested)		
No parental figure	0.002	0.045
Missing or cannot say	0.247	0.431
Moderately interested	0.290	0.454
Very little or uninterested	0.074	0.262
Home ownership (owned)		
Mortgage	0.545	0.498
Rented or other	0.339	0.474
Missing	0.005	0.072
Region (South East)		
Missing	0.002	0.045
North East	0.064	0.244
North West	0.145	0.352
Yorkshire	0.109	0.311
East Midlands	0.064	0.245
West Midlands	0.114	0.318
East	0.072	0.258
London	0.076	0.264
South West	0.057	0.232
Wales	0.074	0.262

continued

Variable	Mean	s.d.
Father's father social class (IIIM)		
Not matched	0.132	0.339
Not known	0.164	0.371
Not applicable	0.033	0.179
I	0.018	0.132
II	0.113	0.317
IIINM	0.056	0.230
IV	0.115	0.320
V	0.049	0.216
Ability scores		
BAS quantitative	29.821	16.986
BAS verbal	22.650	12.347
BAS missing	0.211	0.408
Mother's employment (OLF)[b]		
Missing	0.047	0.212
Regular employment	0.505	0.500
Occasional employment	0.141	.348

Note. This table reports means and standard deviations for the the variables used in our preferred specification (the binary probit model reported in Table 2.10). The sample includes 3,888 individuals. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force.

Table 2.3: Sample summary statistics, women (BCS70)

Variable	Mean	s.d.
Quits at 16	0.505	0.500
Ln(income)	4.761	0.482
Child's ethnicity (white)		
Non-white	0.031	0.173
Father's education (O-level)		
< O-level	0.152	0.359
A-level	0.083	0.276
Professional or other	0.030	0.171
Degree	0.121	0.326
Missing	0.037	0.189
Father not present	0.075	0.263
Mother's education (O-level)		
< O-level	0.184	0.388
A-level	0.037	0.190
Professional or other	0.054	0.226
Degree	0.026	0.160
Missing	0.059	0.235
Social class (IIIM)		
I	0.056	0.229
II	0.220	0.414
IIINM	0.084	0.278
IV	0.102	0.302
V	0.025	0.156
OLFU[a]	0.042	0.200
Missing	0.064	0.244
Parental interest in child's education (very interested)		
No parental figure	0.003	0.056
Missing or cannot say	0.240	0.427
Moderately interested	0.281	0.450
Very little or uninterested	0.063	0.243
Home ownership (owned)		
Mortgage	0.549	0.498
Rented or other	0.338	0.473
Missing	0.007	0.083
Region (South East)		
Missing	0.003	0.058
North East	0.060	0.238
North West	0.139	0.346
Yorkshire	0.117	0.321
East Midlands	0.061	0.240
West Midlands	0.108	0.310
East	0.070	0.255
London	0.079	0.270
South West	0.072	0.259
Wales	0.067	0.250

continued

Variable	Mean	s.d.
Mother's industry (other services)		
Missing	0.323	0.468
Agriculture	0.013	0.113
Extraction	0.017	0.128
Metal goods	0.039	0.194
Other manufacturing	0.069	0.254
Construction	0.007	0.086
Distribution	0.141	0.348
Transport	0.010	0.100
Banking	0.088	0.283
Father's industry (metal goods)		
Missing	0.167	0.373
Agriculture	0.021	0.142
Energy	0.054	0.225
Extraction	0.058	0.234
Other manufacturing	0.083	0.277
Construction	0.102	0.303
Distribution	0.085	0.279
Transport	0.078	0.268
Banking	0.045	0.207
Other services	0.128	0.335
Ability scores		
BAS quantitative	30.845	16.969
BAS verbal	22.335	11.824
BAS missing	0.201	0.401
Mother's employment (OLF)		
Missing	0.041	0.198
Mother not present	0.008	0.087
Regular employment	0.518	0.500
Occasional employment	0.135	0.342

Note. This table reports means and standard deviations for the the variables used in our preferred specification (the binary probit model reported in Table 2.11). The sample includes 4,168 individuals. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force.

Table 2.4: Family income regression (BCS70)

Variable	Coef.		s.e.
Father not present	-59.117	***	2.041
Mother not present	-8.345		5.735
Father's ethnicity (white)			
Missing	-56.375	***	2.983
Non-white	-8.262	**	4.159
Mother's ethnicity (white)			
Missing	6.695		7.008
Non-white	-3.136		3.927
Father's education (O-level)			
< O-level	8.800	***	1.363
A-level	14.014	***	1.761
Professional	22.348	***	3.383
Degree	35.563	***	2.456
Other	3.045		9.564
Missing	0.025		1.915
Mother's education (O-level)			
< O-level	8.781	***	1.283
A-level	13.379	***	2.747
Professional	21.542	***	2.656
Degree	31.507	***	3.600
Other	23.954	*	13.209
Missing	-3.048	*	1.793
Father's social class (IIIM)			
I	32.323	***	3.177
II	27.178	***	1.599
IIINM	6.966	***	1.778
IV	-5.440	***	1.263
V	-15.830	***	1.870
Unemployed	-46.633	***	1.648
OLF[a]	-37.936	***	1.973
Missing	-2.969		2.635
Mother's social class (IIINM)			
I and II	5.610	***	2.012
IIIM	-7.501	***	2.023
IV	-12.083	***	1.433
V	-18.813	***	1.773
Unemployed	-25.536	***	3.114
OLF	-20.707	***	1.540
Missing	-15.933	***	2.320
Region (South East)			
Missing	32.781	***	10.201
North East	-12.901	***	1.895
North West	-11.907	***	1.513
Yorkshire	-13.435	***	1.616
East Midlands	-11.699	***	1.901
West Midlands	-10.958	***	1.613
East	-7.427	***	2.007
London	3.595	*	1.938
South West	-17.431	***	2.038
Wales	-16.503	***	1.939
Scotland	-13.095	***	1.611

continued

Variable	Coef.		s.e.
Father's industry (metal goods)			
Missing	-3.441	**	1.509
Agriculture	-31.161	***	3.134
Energy	12.097	***	2.099
Extraction	-0.197		1.848
Other manufacturing	2.839		1.756
Construction	-0.292		1.622
Distribution	-13.447	***	1.950
Transport	-0.978		1.821
Banking	12.367	***	2.805
Other services	-7.322	***	1.797
Mother's industry (other services)			
Missing	1.189		1.441
Agriculture	4.624		4.262
Extraction	15.621	***	3.616
Metal goods	11.175	***	2.264
Other manufacturing	6.837	***	1.611
Construction	-4.643		5.665
Distribution	-5.844	***	1.437
Transport	10.758	***	4.191
Banking	5.016	***	1.734
Father's father social class (IIIM)			
Not matched	1.579		1.404
Not known	-2.965	**	1.231
Not applicable	4.930	*	2.564
I	9.744	**	4.472
II	1.438		1.700
IIINM	-2.559		2.058
IV	-2.664	**	1.332
V	-1.437		1.719
Mother's father social class (IIIM)			
Not known	-3.082	**	1.503
Not applicable	1.212		2.247
I	6.997	*	4.169
II	5.119	***	1.646
IIINM	2.295		1.978
IV	-1.528		1.240
V	-2.727	*	1.640
Constant	137.804	***	2.082
Pseudo R^2	0.178		
N. observations	12,542		
Wald tests			
Father's industry	$Chi^2(10)=274.21$ (0.00)		
Mother's industry	$Chi^2(9)=113.00$ (0.00)		
Father's father social class	$Chi^2(8)=26.64$ (0.00)		
Mother's father social class	$Chi^2(7)=25.37$ (0.00)		

Note. The family income regression has been estimated using interval regression (Stata 2003). The Wald test are the tests for the exclusions of the candidate instruments from the income equation. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.5: Wald tests for the exclusion of candidate instruments from the education equation, probit model with selection

Wald tests	Men	Women
1 Father's industry	$Chi^2(10) = 13.72$ (0.1863)	$Chi^2(9) = 19.80$ (0.019)
2 Mother's industry	$Chi^2(9) = 12.63$ (0.1801)	$Chi^2(10) = 15.80$ (0.106)
3 Father's father social class	$Chi^2(8) = 17.22$ (0.0279)	$Chi^2(8) = 6.95$ (0.541)
4 Mother's father social class	$Chi^2(7) = 4.77$ (0.6881)	$Chi^2(7) = 6.91$ (0.438)
1+2+4	$Chi^2(26) = 32.87$ (0.1660)	-
2+3+4	-	$Chi^2(25) = 34.64$ (0.095)
3+4	-	$Chi^2(15) = 18.41$ (0.242)

Note. This table reports Wald tests for the exclusion of the candidate instruments from the education equation in the probit model with selection estimated using instrumented income and without exclusion restrictions.

Table 2.6: Probit model with selection of quitting at 16, men (BCS70)

Variable	Coef.		s.e.
Ln(income)	-0.150	**	0.063
Child's ethnicity (white)			
Non-white	-0.939	***	0.141
Father's education (O-level)			
< O-level	-0.299	***	0.067
A-level	-0.148	*	0.084
Professional or other	-0.690	***	0.150
Degree	-0.452	***	0.106
Father not present	-0.365	**	0.158
Missing	-0.107		0.122
Mother's education (O-level)			
< O-level	-0.326	***	0.060
A-level	-0.273	**	0.121
Professional or other	-0.551	***	0.125
Degree	-0.772	***	0.176
Mother not present	0.273		0.362
Social class (IIIM)			
Missing	-0.025		0.120
I	-0.417	***	0.144
II	-0.303	***	0.092
IIINM	-0.278	***	0.082
IV	0.061		0.085
V	0.002		0.159
OLFU	-0.189		0.140
Parental interest in child's education (very interested)			
Missing	-0.176		0.122
No parental figure	-0.393		0.588
Missing or cannot say	0.257	***	0.068
Moderately interested	0.226	***	0.057
Very little or uninterested	0.350	***	0.102
Home ownership (owned)			
Mortgage	0.102		0.075
Rented or other	0.405	***	0.082
Missing	0.392		0.325
Region (South East)			
Missing	-0.927		0.565
North East	0.278	**	0.112
North West	0.168	**	0.077
Yorkshire	0.077		0.082
East Midlands	-0.008		0.098
West Midlands	-0.012		0.088
East	-0.039		0.101
London	0.072		0.096
South West	-0.002		0.108
Wales	-0.257	***	0.100

continued

Variable	Coef.		s.e.
Father's father social class (IIIM)			
Not matched	-0.045		0.093
Not known	-0.031		0.072
Not applicable	-0.008		0.135
I	-0.486	**	0.205
II	-0.167	**	0.079
IIINM	0.086		0.101
IV	0.160	**	0.079
V	-0.070		0.112
Ability scores			
BAS quantitative	-0.023	***	0.004
BAS verbal	-0.049	***	0.008
BAS missing	-2.258	***	0.231
Mother's employment (OLF)			
Missing	0.274	**	0.122
Regular employment	0.063		0.059
Occasional employment, unemployed	0.016		0.077
Constant	3.360	***	0.386

Note. Bootstrapped standard errors (since parental income is predicted using an interval regression), 100 replications. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.7: Selection equation: income response, men (BCS70)

Variable	Coef.		s.e.
Father's ethnicity (white)			
No father figure	-0.118	*	0.066
Missing	0.127		0.091
Non-white	0.068		0.156
Mother's ethnicity (white)			
No mother figure	0.658	***	0.114
Missing	-1.705	***	0.162
Non-white	-0.161		0.143
Father's education (O-level)			
< O-level	0.048		0.033
A-level	-0.041		0.039
Professional	0.306	***	0.063
Degree	0.162	***	0.045
Missing	-0.470	***	0.048
Mother's education (O-level)			
< O-level	0.125	***	0.028
A-level	0.182	***	0.053
Professional or other	-0.185	***	0.052
Degree	0.095		0.069
Missing	-0.232	***	0.048
Father's social class (IIIM)			
I	-0.281	***	0.056
II	-0.288	***	0.036
IIINM	-0.043		0.043
IV	0.098	***	0.038
V	-0.303	***	0.062
U	-0.096		0.067
OLF	0.593	***	0.107
Missing	-0.228	***	0.068
Mother's social class (IIINM)			
I and II	0.004		0.042
IIIM	0.154	***	0.046
IV	0.039		0.037
V	0.282	***	0.062
U	-0.062		0.094
OLF	-0.058		0.040
Missing	-0.024		0.059
Region (South East)			
Missing	-1.386	***	0.166
North East	0.041		0.043
North West	-0.020		0.034
Yorkshire	0.043		0.038
East Midlands	0.029		0.047
West Midlands	-0.214	***	0.039
East	-0.203	***	0.042
London	-0.092	**	0.043
South West	0.002		0.047
Wales	-0.002		0.046

continued

Variable	Coef.		s.e.
Father's industry (metal goods)			
Missing	-0.212	***	0.038
Agriculture	-0.340	***	0.074
Energy	0.060		0.050
Extraction	-0.083	*	0.048
Other manufacturing	-0.234	***	0.052
Construction	-0.204	***	0.057
Distribution	-0.293	***	0.045
Transport	-0.030		0.051
Banking	-0.220	***	0.057
Other services	-0.021		0.043
Mother's industry (other services)			
Missing	-0.079	**	0.037
Agriculture	0.067		0.080
Extraction	-0.056		0.096
Metal goods	-0.133	**	0.062
Other manufacturing	0.061		0.048
Construction	-0.181		0.142
Distribution	-0.085	*	0.052
Transport	-0.099		0.095
Banking	0.022		0.040
Father's father social class (IIIM)			
Not matched	-0.348	***	0.035
Not known	-0.095	***	0.030
Not applicable	-0.186	***	0.064
I	0.081		0.085
II	-0.019		0.035
IIINM	-0.099	**	0.046
IV	-0.057	*	0.035
V	0.037		0.051
Mother's father social class (IIIM)			
Not known	-0.046		0.039
Not applicable	0.156	***	0.050
I	-0.052		0.083
II	-0.129	***	0.034
IIINM	0.083	*	0.046
IV	-0.046		0.033
V	-0.062		0.050
Constant	1.606	***	0.063
N. observations		5,085	
ρ	-0.157		0.478

Note. Bootstrapped standard errors (since parental income is predicted using an interval regression), 100 replications. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.8: Probit model with selection of quitting at 16, women (BCS70)

Variable	Coef.		s.e.
Ln(income)	-0.187	***	0.064
Child's ethnicity (white)			
Non-white	-0.853	***	0.198
Father's education (O-level)			
< O-level	-0.177	***	0.063
A-level	-0.159	*	0.085
Professional or other	-0.258	*	0.149
Degree	-0.367	***	0.099
Father not present	-0.183		0.138
Missing	-0.125		0.111
Mother's education (O-level)			
< O-level	-0.278	***	0.062
A-level	-0.402	***	0.120
Professional or other	-0.387	***	0.122
Degree	-0.570	***	0.172
Mother not present	0.317		0.326
Missing	0.185	*	0.095
Social class (IIIM)			
I	-0.177		0.181
II	-0.170	*	0.091
IIINM	-0.081		0.085
IV	0.088		0.078
V	0.077		0.148
OLFU	0.012		0.117
Parental interest in child's education (very interested)			
Missing	-0.349	***	0.104
No parental figure	-0.354		0.456
Missing or cannot say	0.158	**	0.064
Moderately interested	0.285	***	0.055
Very little or uninterested	0.586	***	0.106
Home ownership (owned)			
Mortgage	0.058		0.072
Rented or other	0.343	***	0.083
Missing	0.170		0.234
Region (South East)			
North East	-0.128		0.093
North West	0.058		0.081
Yorkshire	0.162	**	0.077
East Midlands	0.239	**	0.095
West Midlands	0.160	*	0.084
East	0.020		0.090
London	-0.122		0.118
South West	0.174	*	0.091
Wales	-0.309	***	0.101

continued

Variable	Coef.		s.e.
Ability scores			
BAS quantitative	-0.020	***	0.004
BAS verbal	-0.036	***	0.007
BAS missing	-1.710	***	0.220
Mother's industry (other services)			
Missing	0.039		0.083
Agriculture	0.001		0.182
Extraction	0.501	***	0.184
Metal goods	0.128		0.121
Other manufacturing	0.159	*	0.091
Construction	-0.021		0.257
Distribution	0.177	**	0.071
Transport	0.372	*	0.216
Banking	0.172	**	0.085
Father's industry (metal goods)			
Missing	-0.090		0.089
Agriculture	-0.199		0.213
Energy	-0.065		0.109
Extraction	-0.027		0.102
Other manufacturing	-0.195	*	0.104
Construction	-0.148		0.094
Distribution	-0.204		0.108
Transport	-0.109		0.094
Banking	-0.229	*	0.125
Other services	-0.273	***	0.089
Mother's employment (OLF)[b]			
Missing	0.110		0.124
Regular employment	0.068		0.068
Occasional employment, unemployed	0.003		0.079
Constant	2.703	***	0.389

Note. Bootstrapped standard errors (since parental income is predicted using an interval regression), 100 replications. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.9: Selection equation: income response, women (BCS70)

Variable	Coef.		s.e.
Father's ethnicity (white)			
No father figure	0.340	***	0.060
Missing	0.645	***	0.145
Non-white	0.506	***	0.113
Mother's ethnicity (white)			
No mother figure	-0.308	***	0.114
Missing	-0.157		0.195
Non-white	-0.721	***	0.115
Father's education (O-level)			
< O-level	0.017		0.030
A-level	-0.117	***	0.038
Professional	-0.238	***	0.064
Degree	0.016		0.045
Missing	-0.350	***	0.055
Mother's education (O-level)			
< O-level	0.250	***	0.029
A-level	0.099	*	0.059
Professional or other	-0.083		0.073
Degree	0.430	***	0.093
Missing	-0.141	***	0.047
Father's social class (IIIM)			
I	-0.441	***	0.060
II	-0.230	***	0.037
IIINM	-0.008		0.050
IV	0.141	*	0.035
V	0.145	**	0.063
U	0.083		0.060
OLF	-0.092		0.095
Missing	-0.066		0.070
Mother's social class (IIINM)			
I and II	0.190	***	0.059
IIIM	-0.016		0.052
IV	0.074	**	0.035
V	-0.078		0.082
U	-0.369	***	0.096
OLF	-0.157	***	0.035
Missing	-0.444	***	0.058
Region (South East)			
North East	0.056		0.044
North West	-0.218	***	0.033
Yorkshire	0.019		0.035
East Midlands	-0.199	***	0.047
West Midlands	-0.257	***	0.037
East	-0.076	*	0.041
London	-0.274	***	0.042
South West	0.001		0.047
Wales	-0.083	**	0.042

continued

Variable	Coef.		s.e.
Father's industry (metal goods)			
Missing	-0.246	***	0.036
Agriculture	-0.644	***	0.079
Energy	0.169	***	0.048
Extraction	0.020		0.047
Other manufacturing	-0.230	***	0.043
Construction	-0.187	***	0.040
Distribution	-0.232	***	0.043
Transport	0.093	**	0.043
Banking	-0.115	*	0.063
Other services	0.130	***	0.044
Mother's industry (other services)			
Missing	-0.061	*	0.037
Agriculture	-0.033		0.108
Extraction	0.525	***	0.075
Metal goods	0.310	***	0.053
Other manufacturing	-0.139	***	0.042
Construction	-0.151		0.118
Distribution	-0.133	***	0.034
Transport	-0.198	**	0.100
Banking	0.160	***	0.047
Father's father social class (IIIM)			
Not matched	-0.169	***	0.036
Not known	-0.154	***	0.035
Not applicable	-0.091		0.071
I	-0.061		0.111
II	-0.095	**	0.047
IIINM	-0.034	**	0.051
IV	-0.007		0.036
V	0.055		0.051
Mother's father social class (IIIM)			
Not known	-0.023		0.040
Not applicable	-0.126	*	0.071
I	-0.272	***	0.092
II	-0.265	***	0.044
IIINM	-0.007		0.042
IV	0.188	***	0.036
V	-0.226	***	0.047
Constant	1.724	***	0.050
N. observations		4,671	
ρ	-0.384		0.628

Note. Bootstrapped standard errors (since parental income is predicted using an interval regression), 100 replications. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.10: Probit model of quitting at 16, men (BCS70)

Variable	Coef.		s.e.	m.e.
Ln(income)	-0.151	**	0.064	-0.057
Child's ethnicity (white)				
Non-white	-0.949	***	0.140	-0.362
Father's education (O-level)				
< O-level	-0.298	***	0.068	-0.113
A-level	-0.149	*	0.087	-0.055
Professional or other	-0.675	***	0.147	-0.262
Degree	-0.441	***	0.104	-0.170
Missing	-0.133		0.097	-0.049
Father not present	-0.365	**	0.165	-0.139
Mother's education (O-level)				
< O-level	-0.322	***	0.060	-0.123
A-level	-0.265	**	0.130	-0.101
Professional or other	-0.564	***	0.109	-0.219
Degree	-0.774	***	0.191	-0.301
Missing	-0.044		0.108	-0.016
Mother not present	0.304		1.413	0.102
Social class (IIIM)				
I	-0.433	***	0.133	-0.167
II	-0.323	***	0.069	-0.123
IIINM	-0.280	***	0.086	-0.106
IV	0.065		0.091	0.023
V	-0.009		0.153	-0.003
OLFU[a]	-0.181		0.136	-0.068
Missing	-0.190	*	0.113	-0.071
Parental interest in child's education (very interested)				
No parental figure	-0.399		1.505	-0.158
Missing or cannot say	0.258	***	0.068	0.098
Moderately interested	0.228	***	0.059	0.087
Very little or uninterested	0.351	***	0.107	0.132
Home ownership (owned)				
Mortgage	0.104		0.079	0.041
Rented or other	0.408	***	0.082	0.154
Missing	0.398		0.370	0.150
Region (South East)				
Missing	-1.023	**	0.407	-0.382
North East	0.282	**	0.115	0.103
North West	0.168	**	0.080	0.062
Yorkshire	0.078		0.090	0.030
East Midlands	-0.005		0.099	-0.002
West Midlands	-0.021		0.081	-0.008
East	-0.048		0.101	-0.019
London	0.068		0.094	0.026
South West	-0.001		0.113	0.000
Wales	-0.258	**	0.105	-0.101

continued

Variable	Coef.		s.e.	m.e.
Father's father social class (IIIM)				
Not matched	-0.062		0.077	-0.024
Not known	-0.037		0.071	-0.014
Not applicable	-0.018		0.136	-0.007
I	-0.488	**	0.232	-0.192
II	-0.170	**	0.076	-0.066
IIINM	0.082		0.099	0.030
IV	0.157	**	0.079	0.058
V	-0.070		0.116	-0.027
Ability scores				
BAS quantitative	-0.023	***	0.004	-0.007
BAS verbal	-0.049	***	0.008	-0.014
BAS missing	-2.266	***	0.226	-0.369
Mother's employment (OLF)[b]				
Missing	0.266	**	0.123	0.098
Regular employment	0.071		0.051	0.027
Occasional employment, unemployed	0.023		0.079	0.009
Constant	3.342	***	0.422	-
N. observations			3,888	
Pseudo R^2			0.225	

Note. Bootstrapped standard errors (since parental income is predicted using an interval regression), 500 replications. For a continuous variable, e.g. x_i, the marginal effect (m.e.) is computed as $\phi(\beta\overline{Z})\beta_i$ where $\beta\overline{Z}$ is the linear predictor excluding the variable x_i, computed at the sample mean, β_i the coefficient on x_i and $\phi(\cdot)$ the standard normal density function. For a categorical variable (including dummies), e.g. C_i, the marginal effect is computed as $\Phi(\beta\overline{Z} + \beta_i) - \Phi(\beta\overline{Z})$, where $\beta\overline{Z}$ is the linear predictor excluding all the categorical variables of the same group as C_i (e.g. mother's level education) computed at the sample mean, β_i is the coefficient on C_i and $\Phi(\cdot)$ the standard normal distribution function. Reference characteristics for each group of variables are indicated in brackets. [a] Out of the labour force or unemployed; [b] Out of the labour force. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.11: Probit model of quitting at 16, women (BCS70)

Variable	Coef.		s.e.	m.e.
Ln(income)	-0.191	***	0.062	-0.076
Child's ethnicity (white)				
Non-white	-0.916	***	0.133	-0.324
Father's education (O-level)				
< O-level	-0.174	***	0.061	-0.069
A-level	-0.172	**	0.080	-0.068
Professional or other	-0.289	**	0.135	-0.115
Degree	-0.369	***	0.100	-0.146
Missing	-0.129		0.102	-0.069
Father not present	-0.174		0.149	-0.051
Mother's education (O-level)				
< O-level	-0.251	***	0.060	-0.100
A-level	-0.404	***	0.126	-0.159
Professional or other	-0.406	***	0.113	-0.160
Degree	-0.537	***	0.181	-0.209
Missing	0.167	*	0.097	0.066
Social class (IIIM)				
I	-0.251	*	0.134	-0.100
II	-0.206	***	0.072	-0.082
IIINM	-0.098		0.090	-0.039
IV	0.101		0.073	0.040
V	0.081		0.148	0.032
OLFU	0.002		0.118	0.001
Missing	-0.356	***	0.109	-0.141
Parental interest in child's education (very interested)				
No parental figure	-0.352		0.518	-0.133
Missing or cannot say	0.161	**	0.068	0.064
Moderately interested	0.285	***	0.053	0.113
Very little or uninterested	0.584	***	0.107	0.227
Home ownership (owned)				
Mortgage	0.063		0.073	0.025
Rented or other	0.358	***	0.081	0.142
Missing	0.186		0.261	0.074
Region (South East)				
Missing	-0.371		0.456	-0.144
North East	-0.127		0.098	-0.050
North West	0.035		0.077	0.014
Yorkshire	0.167	**	0.077	0.067
East Midlands	0.223	**	0.097	0.088
West Midlands	0.138		0.081	0.055
East	0.013		0.095	0.005
London	-0.165	*	0.093	-0.065
South West	0.177	*	0.094	0.070
Wales	-0.324	***	0.097	-0.126

continued

Variable	Coef.		s.e.	m.e.
Mother's industry (other services)				
Missing	0.014		0.069	0.006
Agriculture	-0.008		0.194	-0.003
Extraction	0.535	***	0.186	0.207
Metal goods	0.149		0.116	0.060
Other manufacturing	0.147		0.096	0.059
Construction	-0.045		0.277	-0.018
Distribution	0.166	**	0.074	0.066
Transport	0.355		0.231	0.140
Banking	0.185	**	0.083	0.074
Father's industry (metal goods)				
Missing	-0.115		0.084	-0.046
Agriculture	-0.284	*	0.164	-0.113
Energy	-0.053		0.115	-0.021
Extraction	-0.025		0.097	-0.010
Other manufacturing	-0.223	**	0.095	-0.089
Construction	-0.168	*	0.089	-0.067
Distribution	-0.231		0.094	-0.092
Transport	-0.104		0.098	-0.041
Banking	-0.242	*	0.128	-0.096
Other services	-0.261	***	0.094	-0.104
Ability scores				
BAS quantitative	-0.021	***	0.004	-0.008
BAS verbal	-0.037	***	0.007	-0.014
BAS missing	-1.746	***	0.205	-0.437
Mother's employment (OLF)[b]				
Missing	0.064		0.114	0.026
Mother not present	0.299		0.528	0.118
Regular employment	0.073		0.066	0.029
Occasional employment, unemployed	-0.002		0.084	-0.001
Constant	2.716		0.381	-
N. observations		4,168		
Pseudo R^2		0.175		

Note. Bootstrapped standard errors (since parental income is predicted using an interval regression), 500 replications. For a continuous variable, e.g. x_i, the marginal effect (m.e.) is computed as $\phi(\beta\overline{Z})\beta_i$, where $\beta\overline{Z}$ is the linear predictor excluding the variable x_i, computed at the sample mean, β_i the coefficient on x_i and $\phi(\cdot)$ the standard normal density function. For a categorical variable (including dummies), e.g. C_i, the marginal effect is computed as $\Phi(\beta\overline{Z}+\beta_i)-\Phi(\beta\overline{Z})$ where $\beta\overline{Z}$ is the linear predictor excluding all the categorical variables of the same group as C_i (e.g. mother's level education) computed at the sample mean, β_i is the coefficient on C_i and $\Phi(\cdot)$ the standard normal distribution function. Reference characteristics for each group of variables are indicated in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.12: Marginal effects of an increase in parental income from the probit models of staying-on at 16

Quitting probability	Men		Women	
p	one-s.d. (£58)	£16	one-s.d. (£58)	£16
0.1	-0.012	-0.003	-0.015	-0.004
0.3	-0.024	-0.007	-0.030	-0.008
0.5	-0.027	-0.008	-0.035	-0.010

Note. Marginal effects of income variations are computed using the following expression: $\phi(\Phi^{-1}(p)) \cdot \alpha \cdot \frac{\Delta y}{\overline{y}}$, where $\overline{y}$ is average income and α the coefficient of ln(income) in the staying-on equation, and $\phi(\cdot)$ and $\Phi(\cdot)$ the standard normal density and distribution functions, respectively.

Chapter 3

Social class and undergraduate degree subject in the UK

3.1 Introduction

Widening access to education is commonly viewed as an effective way of promoting higher intergenerational mobility and the study of educational attainment has received growing attention by economists. This is especially true for the UK where the empirical analysis of educational attainment using microdata has a long tradition. Several studies have investigated the determinants of the level of education achieved (the number of years of schooling or the highest educational qualification) using cross-section or longitudinal microdata. Some examples include Rice (1987), Micklewright (1989), Ermisch and Francesconi (2000, 2001a), Feinstein (2000) and Chevalier and Lanot (2002) among others. The vast majority of these studies finds that long-term family characteristics, such as parental education and social class, are of paramount importance for children's educational attainment suggesting, therefore, the presence of factors which may reduce intergenerational mobility. Moreover, Blanden and Machin (2004) find that the recent expansion of the Higher Education (HE) system acted to widen the participation gap between rich and poor children.

However, in a period of increasing access to education, a great deal of the variation in individuals' labour market outcomes (employment opportunities and earnings) may be determined by the *type* in addition to the *level* of education achieved. Hence, in this chapter we are interested in factors influencing the field of study in which individuals enroll at the (university) undergraduate level, with a specific focus on the effect of social class. Empirical evidence that supports the importance of field of study as one of the main determinants of graduates' performance in the labour market is provided by several studies. Smith *et al.* (2000) and Bratti *et al.* (2004), for instance, report significant differences in first destinations of graduates from different subject fields. Large differences also exist in graduates' earnings by degree subject, as shown by Blackaby *et al.* (1999), Walker and Zhu (2001), Chevalier *et al.* (2002), Bratti and Mancini (2003), and Sloane and O'Leary (2004), among others.

Despite the potential interest of the topic, to the best of our knowledge, to date there exists only one empirical study on undergraduate field of graduation in the UK, van de Werfhorst et al. (2003), which analyses survey data for the 1958 British cohort. In the current chapter, we aim to contribute to the existing literature on subject choice in the UK by extending the analysis to several cohorts of university students (from 1981 until 1991), using administrative individual-level data. Unlike many previous studies, we use cohorts of entrant students rather than cohorts of students leaving with a university qualification and model subject choice allowing for a non-zero correlation across the unobserved factors which might simultaneously affect the utilities received from studying different disciplines. Furthermore, using several cohorts of university students we also analyse the changes in social class effects over time.

The chapter is organised as follows. The next section provides some motivation for the study of the choice of undergraduate field of enrollment. Section 3.3 gives an outline of the reasons why social class should matter for the choice of degree subject. Section 3.4 briefly surveys the existing literature.

Section 3.5 introduces some econometric issues and section 3.6 describes the econometric model. Section 3.7 discusses the main features of the data set and the sample used in our estimates and section 3.8 reports the empirical results. Section 3.9 concludes.

3.2 Motivation

The process through which students enroll in a particular subject is worth studying for several reasons. Firstly, Higher Education (HE) Institutions are interested in the determinants of subject choice. Indeed, once HE institutions have determined the potential 'target population' for specific courses (that is the background of the students enrolled in specific fields of study), they can focus their 'marketing policies' on this target, to consolidate their position, or alternatively act so as to attract new segments of 'potential customers' (i.e. students).

Secondly, the issue is of interest also to society as a whole, since there is some concern about the lack of workers in high demand fields, such as graduates in computer sciences and IT.[1] Hence, a deeper understanding of the mechanism driving students' choices is important also to explain some apparent inefficiencies in the labour market and to forecast future labour market trends.

Thirdly, the topic is also central to the policy makers involved in poverty reduction and HE reform. The former are interested in factors promoting in-

[1]See for instance Mason (1999).

tergenerational mobility and the reduction of poverty. While previous studies have shown a substantial amount of correlation between parents' and children's incomes and education,[2] there is much less empirical evidence on the effect of family and social background on the choice of subject at tertiary level.[3] However, this is a very important issue since, as already stated, previous research has shown that graduates' labour market outcomes differ according to the subject studied at university. As to HE reform, one relatively recent innovation introduced in the UK educational system is home student tuition fees for cohorts entering university from the autumn 1998. Some people have argued in favour of the need to differentiate tuition fees according to the different returns that subjects command in the labour market in terms of both earnings and employment prospects and to the differences in the cost of teaching and the quality of infrastructure used.[4] However, the introduction of 'top-up' fees might have especially adverse effects on students from more deprived family backgrounds. If higher fees are not counter balanced by the introduction of new scholarships and loans made available to students, the innovation could mainly hit the pockets of middle and low social class students discouraging them from enrolling in HE, or from enrolling in the

[2]For some studies related to intergenerational mobility in Britain see Dearden *et al.* (1997), Blanden *et al.* (2003) and Machin and Gregg (2003), among others. Machin and Gregg (2003) observe, for instance, that the educational expansion of the late 1980s early 1990s benefited especially high social class students, contributing to a decrease in intergenerational mobility.

[3]See Chevalier (2002) for a recent analysis of the role of social class influences on graduates' occupational choices.

[4]See the discussion in Greenaway and Haynes (2003).

most remunerative fields.[5] In a recent article Greenaway and Haynes (2003) cite the case of Australia, where a balanced combination of fees and income loans did not damage access. From this evidence and the fact that in the last two decades in the UK the participation of low income students in HE did not rise as fast as the overall number of students when tuition was free, the authors infer that a system of differential fees and loans will not result in a decrease in participation of low social class students. In their view, the objective of providing more resources for HE without damaging access can be achieved through income contingent loans, which avoid up-front charges and make education free at the point of consumption.[6] Although it could be true that a carefully designed system of differential fees and loans may not impact adversely on low income students' access, the potential impact on degree subject choice (or choice of institution) is less clear. The provision of income contingent student loans *per se* might not be sufficient to deter low income students from enrolling in high fee subjects. Indeed, the introduction of differential fees by subject would change the relative structure of costs and benefits from enrolling in different subjects.[7] These costs and benefits may differ across social classes. In particular, we may expect students

[5] Abbott and Leslie (2004), for istance, observe that most universities saw a fall in applications and enrollments following the introduction of tuition fees.

[6] At the moment in the UK the interest rate on student loans only covers the rate of inflation. Loans are income contingent (Pay As You Earn - PAYE - system): students start to repay the debt when their annual incomes overcome the threshold of £ 10,000 per year.

[7] This may especially happen whether or not fees are determined on the basis of the future private returns to the different subjects.

coming from wealthy families to have lower costs from enrolling in tertiary education than individuals with a low social class background[8] and higher expected returns with respect to the latter, since they are more likely to find 'good jobs' because of family networks.[9] For these reasons, many low social class students may be discouraged from enrolling in high fees subjects. Moreover, if the most remunerative subjects are also those with higher uncertainty (e.g., with higher variance in earnings) individuals from more disadvantaged backgrounds, who are likely to use student loans, may prefer to enroll in less remunerative, but relatively riskless (in terms of earnings variance) subjects, since they may perceive that if they end up in low income jobs it would be particularly hard to pay off the debt. Last but not least, the pressure to earn an income might push some students to accept the first job offer they re-

[8]We are thinking about situations in which high income families pay fees and living costs for their children enrolled in HE without asking them to reimburse the expenses borne. In this case wealthy students' decision will not be affected by the introduction of differential fees. The reasons why families can make financial transfers to their progeny might be purely altruistic or strategically altruistic (see Cigno, 1991) and they can choose between inter-vivo transfers or bequests. Moreover, high income families may decide to use income contingent student loans to finance their children's education, even though they do not need it, since it may be cheaper than using their own resources that are invested in more profitable financial assets, and to refund to their children the expenses borne to repay the debt once they start working.

[9]Hansen (2001) using data from Norway finds that individuals from high social classes earn higher incomes even when education and field are controlled for. Moreover, the advantage tends to be largest in soft educational fields (such as humanities and social studies). Previous research on students' expectations formation has also found a positive relation between social class and the expected returns to education (see Betts, 1996).

ceive without carefully searching for better employment opportunities, with adverse effects on their future labour market outcomes.[10]

3.3 Social class influences on the choice of field of study

The case for strong social class influences on educational choices has been put forward by both sociologists and economists. Although the theoretical work has almost exclusively focused on the level of education, the existing analytical framework can also be easily applied to the choice of field of study.

Starting from the sociological literature, Boudon's (1974) model of 'rational action' states that educational choices depend on the perception of the costs and benefits of each educational alternative available. For the choice of subject field, a related hypothesis elaborated by Kelsall *et al.* (1972) is that low social class students may be more inclined to choose subjects that offer better labour market prospects. This could happen because future labour market outcomes depend more on subject studied for low social class than for high social class students. The latter are likely to enjoy good labour market outcomes once they get a university degree irrespective of their field of graduation, thanks to 'family networks'. Moreover, Kelsall *et al.* (1972) also maintain that low social class students may tend to choose technical fields of study, which are closer to the occupational experience of many man-

[10] Stewart and Swaffield (1999), for instance, using UK data find that the probability of being low paid depends on low pay in the previous year.

ual working class parents. Boudon's (1974) model closely resembles the one commonly used by economists: Becker's (1975) human capital theory. According to Becker (1975) the costs and the returns of education are the main factors driving educational choices. Some fields might be more closely linked to professions for which the presence of 'social networks' (to which high social class students are typically better connected) is more important to ensure labour market success and a higher economic return of the educational investment. Then, the expected return from different educational fields may differ accordingly over social classes. Previous research has shown the existence of family networks effects. Hansen (2001), for instance, found that the ex-post impact of social class on the economic rewards of education varies across educational fields and tends to be largest in 'soft' educational fields (such as social studies and humanities). Moreover, since in the presence of capital market imperfections low social class individuals might have higher costs of enrolling in HE,[11] standard economic theory predicts that these individuals will require a higher return from their investment in university education. The higher return can be obtained by enrolling in subjects highly rewarded in the labour market. Therefore, low social class students will choose relatively 'high performing' subjects (in terms of earnings or employment prospects), a prediction very similar to that elaborated by Kelsall *et al.* (1972). On the grounds of these considerations we advance the following hypotheses:

Hypothesis 1. Low social class individuals are relatively more inclined

[11] Because they need to borrow and pay back student loans while high social class students usually have access to cheaper, even free, family resources to finance higher education.

to enroll in 'technical' degree subjects than middle and high social class students (Kelsall *et al.* 1972);

Hypothesis 2. Low social class individuals are more likely to enroll in subjects which offer better labour market prospects, i.e. higher wages or better employment opportunities, than middle and high social class students (Kelsall *et al.* 1972 and Becker 1975).

Bourdieu's (1984) 'cultural reproduction hypothesis' emphasizes the role of education as a means of reproducing social class. In this context education is the instrument through which the high and middle social classes prevent individuals from lower social backgrounds from accessing the highest positions in society. Two very important concepts for educational choices are those of 'economic capital' and 'cultural capital'. As stressed by van de Werfhorst *et al.* (2001), according to this hypothesis people from the economic elites prefer lucrative fields, which can ensure a comfortable life, while people from the cultural elites are less interested in economic returns to education and prefer fields in which they can acquire 'cultural capital'. Then the following hypothesis can be put forward:

Hypothesis 3. High social class students are relatively more likely to enroll in subjects which provide cultural capital, such as arts and humanities (Bordieu 1984).

This last hypothesis is also in line with hypothesis 2, since arts and humanities are usually not very highly remunerated subjects.

We may expect that with an increased access to higher education the advantages of having a degree for the upper class are progressively lost, and the type of degree possessed becomes an important distinguishing factor for labour market and social success. Highly prestigious professions, such as the legal and medical ones, offer socially and economically advantageous positions in society, and control over these professions becomes a valuable asset for the upper class. A theory leading to similar conclusions exists also in the domain of economics: the 'social networks' model elaborated by Montgomery (1991). Montgomery (1991) builds a theoretical model for explaining the large use of employee referrals as a device for screening job applicants, starting from the observation that workers tend to refer others who are similar to themselves (see Doeringer and Piore, 1971). His model explains "why workers who are well connected (possessing social ties to those in high-paying jobs) might fare better than those who are poorly connected and why firms hiring through referral might earn higher profits" (Montgomery 1991, p. 1414). We may expect that individuals from high social classes are better connected to people working in high-paying jobs, their parents *in primis*. Then we derive the following hypothesis:

Hypothesis 4. High social class students are relatively more likely to enroll in prestigious subjects, i.e. subjects leading to highly paid and often entry-regulated professions, such as medicine and law. In the latter, the comparative advantage of high social class individuals may stem from a direct control over the entry of related professions (Bordieu 1984), or from the existence of social networks effects (Montgomery 1991).[12]

[12]Using Universities' Statistical Record data for the UK, Mancini (2003) finds, for instance, that working class graduates in law are significantly more likely to be unemployed than their wealthier peers.

It is worth noting that the last hypothesis is in apparent contrast[13] with hypothesis 2 since graduates in medicine or law are usually very well paid.

In the empirical analysis, we shall compare the empirical evidence with the hypotheses outlined above and see which ones seem to better fit the data. Although the hypotheses above suggest why social class may be important for the choice of degree subject, they offer at the same time some counter arguments on the reasons why we could find that social class differences are not significant. Some of these reasons could be:

1. individual preferences (or non-pecuniary costs) are not shaped by social class but by factors unrelated to it. One such factor might be the performance in specific subjects at secondary school;

2. given the increasing demand for graduates in the period under study, labour market outcomes did not depend on social class, i.e. family networks were not important and a university degree was sufficient to ensure a good labour market outcome to graduates (absence of heterogeneous returns to degree subjects according to social class);

3. individuals had in the period under study the same (pecuniary) costs of enrolling in different subjects. This might be the case since there were no tuition fees for domestic students and student financial assistance was based on means-tested maintenance grants for low-income students.[14]

[13] We say in "apparent" contrast since in reality students may enroll in the subjects which ensure them the highest expected returns conditional on social class origin.

[14] See Blanden and Machin (2004) for an outline of recent changes in the UK system

Thus all factors above may have contributed to making very similar the behaviour of individuals with different social origins.

3.4 Previous empirical literature

In this section, we report a brief survey of the empirical literature investigating university students' choice of field of study at the undergraduate level and which has also analysed the role of social class influences. To the best of our knowledge, there are only a handful of studies which have investigated social class effects using individual-level data.

Oosterbeek and Webbink (1997) used data from the Netherlands in order to analyse the decision whether or not to attend technical studies. The authors found that children from high income families were less likely to enroll in technical fields, but more likely to persist in their choice once they had undertaken a technical education.

Davies and Guppy (1997) analysed the choice of field of study using US microdata. They found that males were more likely than females to enroll in lucrative fields of study. Moreover, high ability individuals and low social

of Higher Education. In brief, the expansion in the UK education system was partly implemented by reducing the generosity in student support. In particular, the major changes were the freezing of maintenance grants in 1990 and their progressive replacement with subsidised loans; the introduction of home students undergraduate fees of 1,000 pounds per year, the increase in the maximum loan and the introduction of an income-contingent repayment system following the 1997 Dearing Report; the recent discussion over the introduction of top-up fees which may reach a maximum of 3,000 pounds and the reintroduction of a maximum grant of around 1,000 pounds for low-income students.

class individuals were more likely to enter high-return fields.

Van de Werfhorst *et al.* (2001) using Dutch data found that children of the cultural élite tended to choose fields where they could acquire 'cultural capital', i.e. non technical fields, while students from the economic élite were under-represented in cultural fields (such as arts and humanities). By contrast, low social class individuals were over-represented in economics and engineering, i.e. lucrative fields.

Rochat and Demeulemeester (2001) investigated the process of study field choice using Belgian data. They found that students with fathers in 'élite' occupations, such as managers, civil servants or professionals, were relatively more likely to enroll in short cycle artistic and pedagogical studies and long cycle curricula in engineering and less likely to enroll in long cycle business, economics and social studies.

Montmarquette *et al.* (2002) estimated a multinomial logit model of subject choice using Canadian microdata. They found no effect of having a parent in a professional occupation, but that students supported by an educational loan were more likely to choose those fields (education or liberal arts) in which the probability of success was higher on average.

We are only aware of one paper investigating field of study at the undergraduate level in the UK with an emphasis on social class and that is by van de Werfhorst *et al.* (2003).[15] The authors analysed the educational choices and the educational performance of the 1958 British Cohort using

[15] We found only one other recent study analysing subject choice at the university level for the UK, Ashworth and Evans (2001) whose emphasis is, however, on the decision of studying economics and on the effect of secondary school variables.

data from the National Child Development Study. Since there is only this study for the UK, it is worthwhile to devote some space to summarizing the main findings. The focus of van de Werfhorst *et al.* (2003) is on the role of social class, cultural and economic capital and ability on the subject choice in secondary and tertiary education in Britain. We comment here only on the results relating to university education. Van de Werfhorst *et al.* (2003) estimated a multinomial logit model of subject of graduation considering six broad subject categories and including among the explanatory variables family background variables (such as parental social class and measures of 'economic' and 'cultural capital'), ability (verbal and mathematical ability), and measures of comparative advantage (based on O-level subjects choice and performance). The authors found that children from professional backgrounds preferred faculties of medicine and law, even after controlling for ability at age 11 and exam performance at age 16. However, they did not find other social class differences, which, as they pointed out, is not due to the controls for various sorts of school attainments since a model without age-11 and age-16 attainments also shows no other social class effects. However, the authors themselves stated that the lack of a strong social class effect might be due to the specific characteristics of the cohort studied. In fact, at the time of the study only a very small minority of the working class entered HE, and this could be considered as a very particular and selected group (e.g., in terms of academic ability).

3.5 Some econometric issues

In a recent article Todd and Wolpin (2003) discuss the problems arising when researchers specify and estimate Educational Production Functions (EPFs), especially that of missing data on relevant factors that are likely to affect student performance, and which usually constrain researchers to the use of proxy variables. In particular, the authors maintain that: 'the relationship of proxy variables to measured and unmeasured inputs must be understood in the context of a behavioural decision model in order to analyse their likely impact on biases' (p. F15). We follow Todd and Wolpin's suggestion and put our empirical analysis in the context of a model of student behaviour, which, we believe, can give some useful suggestions for the specification of the econometric model and some insights into the potential biases of our estimates.

Let us assume that at time t an individual has to decide the type of tertiary education j, i.e. the university degree subject, among a set of J available alternatives. All the following analysis is conditional on the decision to enter HE. For the sake of simplicity we use here, like in the bulk of the literature, linear functional forms. Let the utility of a student depend on her educational performance, her ability endowment and a stochastic term in the following way:

$$U_{jt} = T_{jt} + \gamma_j \mu_0 + u_{jt} \tag{3.1}$$

where we have omitted for simplicity the subscript for the individual,[16] j is

[16]The weight given to T_{jt} is assumed to be constant across alternatives and has been

the subscript for the subject, T_{jt} is performance in subject j at time t, μ_0 is student ability that does not vary over time and u_{jt} a taste shifter,[17] i.e. an idiosyncratic stochastic term affecting the utility of subject j and unobservable to the econometrician. We have chosen here to let utility depend on educational performance and ability only. However, the argument remains the same if we assume that utility depends on income and that the latter is a function of educational performance, or that utility depends on both income and performance, although the notation becomes more involved.

We assume an educational production function of the following form:

$$T_{jt} = \tau_{0jt} + \tau_{1jt}T_{jt-1} + \tau_{2jt}F_{jt} + \tau_{3jt}\mu_0 \tag{3.2}$$

where T_{jt-1} is performance in the subject chosen at the previous educational stage $jt-1$, and F_{jt} some family (or educational institutions) inputs.[18] Here, we assume a 'value added approach' also for the 'true' educational production function. We posit that there is a technology (EPF) which transforms educational inputs into an output represented by educational performance or 'knowledge'. In this sense, past knowledge (i.e. performance) is combined with current inputs in order to obtain new knowledge, which is measured by

normalized to one for simplicity.

[17]This specification in which ability enters as a direct determinant of subject-related utilities is empirically supported by Arcidiacono (2004), who found that almost all ability sorting across subjects is due to differences in preferences for particular majors at university or in the workplace.

[18]We use here a non-stochastic specification for the performance function. However, the term involving μ_0 is usually not observed and enters the error term in empirical applications.

T_{jt}. The crucial assumption is that only knowledge acquired in the immediately previous educational stage and measured by the relative educational performance is used to produce new knowledge (or performance) at each new stage, which is tantamount to assuming that the amount of knowledge acquired by individuals at stages $1,...,s-2$ is embedded in the knowledge acquired at stage $s-1$, where s is the current educational stage. Here, we consider two stages only: A-levels[19] and undergraduate university education. To make an example, in our case we assume that only the knowledge acquired through A-levels is useful for degree performance while GCSE 0-levels[20] performance, for instance, does not give any additional benefit over and above A-levels. In this respect, we assume an EPF different from Todd and Wolpin (2003) who not only assume that the 'true' EPF depends at each stage on the complete flow of inputs up to the educational stage under study, which is also true in our specification,[21] but also that past inputs contribute to the production of current cognitive achievement over and above the effect acting through past achievement (see equation (9), p. F20 in Todd and Wolpin 2003).[22] To go back to our previous example, in Todd and Wolpin's specification past educational inputs, such as the primary or secondary school attended or the resources devoted to education by a student's family when

[19]That is GCSE 'Advanced' levels, a type of upper secondary education usually undertaken after passing GSCE O-levels (Ordinary levels) exams, normally at age 16, by individuals who want to enroll at a university.

[20]i.e. GCSE 'Ordinary' levels, which are part of compulsory secondary schooling.

[21]This can be seen by substituting for T_{jt-1} in T_{jt}.

[22]Todd and Wolpin (2003) discuss in their article the conditions under which the two specifications are equivalent.

she was a child, have a direct effect on current degree performance over and above that exerted through A-level performance. Thus, in our approach, unlike Todd and Wolpin (2003), T_{jt-1} is a sufficient statistic for all educational inputs used in previous educational stages. It is difficult to say which one of the two different views of the cognitive achievement process is closer to reality since little is known about the process through which 'knowledge' is formed. In the future, a major interaction between educational economists, psychologists and educational researchers could give useful insights for a correct specification of EPFs.[23]

Plugging equation (3.2) into (3.1) we obtain:

$$U_{jt} = \tau_{0jt} + \tau_{1jt}T_{jt-1} + \tau_{2jt}F_{jt} + (\tau_{3jt} + \gamma_j)\mu_0 + u_{jt}. \tag{3.3}$$

Unfortunately, this specification still includes some variables unobservable to the econometrician. Typically the family inputs F_{jt} are missing in the administrative data commonly used to estimate EPFs. However, we can suppose that family inputs are in turn the outcome of an optimizing process and a function of both observable and unobserved (i.e. missing) family exogenous characteristics. In particular:

$$F_{jt} = \phi_{0jt} + \phi_{1jt}C + \phi_{2jt}M + \phi_{3jt}\mu_0 \tag{3.4}$$

[23]However, the fact that university selection procedures are largely based on secondary school performance, and particularly A-levels, suggests that university selectors consider secondary school (or A-levels) curriculum as a 'sufficient statistic' or as the main determinant of potential degree performance.

where C are observed family characteristics,[24] such as social class, while M are family unobserved characteristics.[25] By plugging equation (3.4) into (3.3) we obtain the following expression for a student's utility:

$$U_{jt} = \tau_{0jt}+\tau_{2jt}\phi_{0jt}+\tau_{1jt}T_{jt-1}+\tau_{2jt}\phi_{1jt}C+\tau_{2jt}\phi_{2jt}M+(\tau_{2jt}\phi_{3jt}+\tau_{3jt}+\gamma_j)\mu_0+u_{jt} \tag{3.5}$$

Let us define $\nu_{jt} \equiv \tau_{2jt}\phi_{2jt}M + (\tau_{2jt}\phi_{3jt} + \tau_{3jt} + \gamma_j)\mu_0$, the error component due to potentially observable but missing variables.

This gives us some useful insights into possible causes of bias. First, as observed by Todd and Wolpin (2003) conditioning on past education performance (T_{jt-1}) makes the model susceptible to endogeneity bias. The endogeneity is due to the correlation between T_{jt-1} and the error component ν_{jt}. This correlation arises both directly from μ_0, the individual unobserved ability endowment, and indirectly through past family (or school) inputs which both enter T_{jt-1} and are likely to be correlated with current family inputs (through μ_0 and M). All these reasons explain why the estimate of τ_{1jt} is likely to be biased. However, the bias may also extend to the estimate of $\tau_{2jt}\phi_{1jt}$, the effect of observed family characteristics. This may happen through the correlation between M and C, i.e. between observed and missing family characteristics, or that between C and μ_0, i.e between observed family characteristics and a student's unobserved ability. In our case, among

[24]The formulation may be extended also to include university/school observed and unobserved characteristics.

[25]For simplicity we consider here only time-invariant family characteristics, at least in the period between enrollment in secondary school and the choice of the degree subject.

family observed characteristics the primary focus is on parents' social class. Thus, the effect of social class is likely to be biased if other unobserved family characteristics are correlated with the former and affect students' subject choice or if students' unobserved cognitive ability endowment is correlated with parents' social class. However, on the grounds that A-level results also act as a good proxy for contemporaneous unobserved family inputs and students' ability, the 'value added specification' may help to attenuate the bias in the estimation of the social class effects on degree subject choice in the spirit of the 'proxying and matching' method in Blundell *et al.* (2000). We think that including A-level results is also important in order to attenuate possible problems of sample selection bias due to the fact that our analysis is conditional on entering HE. Indeed, if individuals with different social class origins who enroll in HE systematically have different levels of ability the effect of social class might reflect the effect of the latter factor.

We can also translate in our context the 'contemporaneous specification' described in Todd and Wolpin (2003). This can be obtained by noting that:

$$F_{jt-1} = \phi_{0jt-1} + \phi_{1jt-1}C + \phi_{2jt-1}M + \phi_{3jt-1}\mu_0 \tag{3.6}$$

$$T_{jt-1} = \tau_{0jt-1} + \tau_{1jt-1}F_{jt-1} + \tau_{2jt-1}\mu_0 + u_{jt-1}, \tag{3.7}$$

where we have assumed for simplicity that secondary school performance is only determined by contemporaneous inputs.[26] Plugging these equations into equation (3.5) and collecting terms we obtain:

[26] The inclusion of past performance in the right-hand-side of equation (3.7) does not alter the essence of the discussion although adds unnecessary complexity to the notation.

$$
\begin{aligned}
U_{jt} = &(\tau_{0jt} + \tau_{1jt}\tau_{0jt-1} + \tau_{1jt}\tau_{1jt-1}\phi_{0jt-1} + \tau_{2jt}\phi_{0jt})+ \\
&+ (\tau_{1jt}\tau_{1jt-1}\phi_{1jt-1} + \tau_{2jt}\phi_{1jt})C+ \\
&+ (\tau_{1jt}\tau_{1jt-1}\phi_{2jt-1} + \tau_{2jt}\phi_{2jt})M+ \\
&+ (\tau_{1jt}(\tau_{1jt-1}\phi_{3jt-1} + \tau_{2jt-1}) + \tau_{2jt}\phi_{3jt} + \tau_{3jt} + \gamma_j)\mu_0+ \\
&+ \tau_{1jt}u_{jt-1} + u_{jt}.
\end{aligned}
\tag{3.8}
$$

Thus, with respect to the 'value added specification', when using the 'contemporaneous specification' two additional terms involving M and μ_0 enter the error term and increase the possible sources of estimation bias. These terms are $(\tau_{1jt}\tau_{1jt-1}\phi_{2jt-1})M$ and $\tau_{1jt}(\tau_{1jt-1}\phi_{3jt-1} + \tau_{2jt-1})\mu_0$, respectively. In words, now that A-level results are not included in the regression, part of the effects of unobserved family inputs (past and present) and students' ability are likely to be picked up by social class. This explains why 'value added specifications' are often preferred by researchers to 'contemporaneous specifications' on the basis of the claim that they reduce the bias due to omitted variables and also why the latter usually have a reduced performance in terms of explanatory power of the included variables. It is also worth noting that the correlation between all the unobserved factors, i.e. the error terms, in the U_{jt}'s both in (3.5) and (3.8) will be in general different from zero since all involve M and μ_0, and will depend on family inputs parameters, EPF parameters and preference parameters.

3.6 The econometric model

In equation (3.5) we define $\epsilon_j \equiv \nu_j + u_j$ and refer to it as the 'error term', on the grounds that it is not observed. We assume that an individual can choose a subject group j among three different alternatives, $j \in \{0, 1, 2\}$, which will be defined in section 3.7.2, each of them providing a utility of:

$$
\begin{aligned}
U_{0i} &= \beta_0' X_i + \epsilon_{0i} && (3.9)\\
U_{1i} &= \beta_1' X_i + \epsilon_{1i} && (3.10)\\
U_{2i} &= \beta_2' X_i + \epsilon_{2i} && (3.11)
\end{aligned}
$$

where we have reintroduced the subscript i for individuals. X_i is the vector of all individual observed characteristics affecting the utility of each group and ϵ_{0i}, ϵ_{1i}, ϵ_{2i} the unobserved components (errors) in these utilities. A possible way of modeling the choice is to use a multinomial logit model (MNL). However, a strong assumption of the MNL is the independence of irrelevant alternatives, i.e. that the error terms of the utilities associated with the alternatives are uncorrelated, an assumption which contrasts with what we have seen in the previous section. Therefore, it is evidently advantageous to use a trinomial probit model (TNP) instead, since the covariance matrix of the error terms is unrestricted. As suggested by Bunch (1991) and Dansie (1985), among others, a convenient way of achieving identification of the TNP is by normalizing one of the utilities to zero. This reduces the dimensionality of the problem. By normalising to zero the utility of the first alternative (U_{0i}), our model becomes:

$$\begin{aligned} U_{0i} &= 0 && (3.12) \\ U_{1i} &= \beta_1' X_i + \epsilon_{1i} && (3.13) \\ U_{2i} &= \beta_2' X_i + \epsilon_{2i} && (3.14) \end{aligned}$$

where:

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix} \right).$$

This model represents a formally identified TNP.[27] As observed by Heckman and Sedlacek (1985) the TNP is identified so long as X_i contains a single regressor that varies across individuals and no exclusion restrictions are required for formal identification. However, as stated by Keane (1992) the TNP may suffer from 'tenuous' identification, and exclusion restrictions may contribute to improving the model identification. The problem is likely to arise especially when considering the choice among a number of alternatives higher than three. However, as often happens in labour economics applications, our data set does not contain alternative-specific variables, so no natural exclusion restrictions are available. Since theory does not suggest any obvious exclusion restrictions for non-alternative-specific variables, i.e. variables affecting the utility of a specific alternative only, we can only estimate the formally identified TNP without exclusion restrictions.[28]

[27]The assumption that also the variance of the second error term is one is not strictly necessary, but is often found in empirical applications. In this case we assume that the error terms in the equations 3.13 and 3.14 are standard normal.

[28]For a review on the multinomial probit model see also Weeks (1997).

The probabilities of the different outcomes are:

$$\begin{aligned} P(Y=0) &= P(U_0 > U_1, U_0 > U_2) \\ &= P(\epsilon_1 < -\beta_1' X, \epsilon_2 < -\beta_2' X) \\ &= \Phi_2(-\beta_1' X, -\beta_2' X | \rho_{12}) \end{aligned} \tag{3.15}$$

$$\begin{aligned} P(Y=1) &= P(U_1 > U_0, U_1 > U_2) \\ &= P(\epsilon_1 < \beta_1' X, \epsilon_1 - \epsilon_2 < -\beta_1' X - \beta_2' X) \\ &= \Phi_2\left(\beta_1' X, \frac{\beta_1' X - \beta_2' X}{\sqrt{2-2\rho_{12}}} | \frac{1-\rho_{12}}{\sqrt{2-2\rho_{12}}}\right) \end{aligned} \tag{3.16}$$

$$\begin{aligned} P(Y=2) &= P(U_2 > U_0, U_2 > U_1) \\ &= P(\epsilon_2 < \beta_2' X, \epsilon_2 - \epsilon_1 < -\beta_2' X - \beta_1' X) \\ &= \Phi_2\left(\beta_2' X, \frac{\beta_2' X - \beta_1' X}{\sqrt{2-2\rho_{12}}} | \frac{1-\rho_{12}}{\sqrt{2-2\rho_{12}}}\right) \end{aligned} \tag{3.17}$$

where we have omitted the subscript for the individual and $\Phi_2(x_1, x_2|\rho_{12})$ is the bivariate standard normal distribution of the two normal random variables ϵ_1 and ϵ_2 computed at the values x_1 and x_2, respectively, with correlation coefficient ρ_{12}.

3.7 Empirical analysis

In the present section we describe the data set used, the choice of explanatory variables and the econometric model.

3.7.1 The data set

In this chapter we use individual-level data from the Universities' Statistical Record (USR). The USR was the institution in charge of the collection of the statistical returns from all university institutions in Great Britain which formerly received Exchequer grants from the University Funding Council (UFC), together with corresponding institutions for the Queen's University of Belfast and the University of Ulster. The USR has stored data from the academic year 1972/1973 until 1993/1994 when it was replaced by the Higher Education Statistics Agency (HESA).[29] The USR data are rich in information concerning the academic life and prior educational qualifications of students and include entire cohorts of students leaving Universities each year.[30]

Using different cohorts of university leavers, it is possible to re-construct the cohorts of entrant students in each academic year. Unlike other papers which use samples of university leavers, in particular students leaving with a university qualification, in this chapter we use cohorts of entrant students. We think that this is more appropriate to investigate factors related to the choice of degree subject. Indeed, obtaining a degree in a specific subject is only the final outcome of several processes, namely the choice to enroll in a

[29]HESA data are generally comparable to USR data. However, the new data do not include performance by type of A-level but only A-level score in the best three passes and is not publicly available. The analysis in this chapter is therefore limited to the 'old' universities, i.e. to the institutions with a university status before the abolition of the binary divide between universities and polytechnics that took place in 1992.

[30]Some previous papers using USR data are, among others, Smith *et al.* (2000), Smith and Naylor (2001a, 2001b), Naylor *et al.* (2002), Bratti (2002b).

certain field, that of remaining in the same field along the course, and that of students' progression. Therefore, it is difficult to isolate the effect of the explanatory variables on each of these single processes by analysing cohorts of leaving students. In this chapter, we aim to analyse the first subject in which students enrolled and accordingly use cohorts of entrant students.

We are aware of the fact that observing an individual enrolled in a certain field implies that he/she has received an offer by a university, and therefore that also the supply side is important. However, we have no individual-level data on the subject preferences stated by students at the application stage. We observe only students' revealed preferences. In this sense, we assume that all students have the same initial choice set (the complete set of subject fields), that they apply for their preferred subset of fields and/or institutions and receive offers by one or more universities. Then they make their final decision based on this restricted set of offers. Hence, although the final choice is the student's one, the process leading to it is complex and is the result of the interactions between students and universities. This should be kept in mind every time we talk of students' choices in this chapter. However, we would like to add that there is some evidence suggesting that our analysis of students' enrolled subject is very close to one of students' subject choice. In a recent article Leslie (2003) uses Universities' College Admissions Service (UCAS) data to build an indicator of quality of subjects. The author uses UCAS data for 1996-2001 and observes that 'each applicant is permitted to make up to six applications (except in medicine, which is restricted to four). Usually these six applications are in a well-defined subject area, but they need not to be so.' (p. 330-331). Another possible criticism to our analysis is that it might confound the effect of social class on subject choice with that on the probability of receiving an offer and accepting it. However, as Leslie (2003) observes, entry qualifications, especially A-levels score, are the key determinant of applicants' success and no other family or social background effects emerge. Therefore, on the basis of this evidence we can argue that in our model we are mainly estimating the effect of social class on applications rather than on offers and acceptances.

It must be noted that since the USR gathers information on university students only, all the empirical analysis that follows is conditional on enrollment in HE[31] and seeks to answer the following question: although there are social class differences in access to HE, once individuals from different social classes manage to enter HE, do they enroll in similar subjects?

3.7.2 Sample definition and descriptive statistics

From the cohort of students in each year 1981-1991, we select only non-mature students (students less than 21 when they entered HE), studying full time for a degree qualification and we exclude overseas and married students. Moreover, since in our specification we want to control for the type and the level of performance at secondary school, we consider only students with A-level qualifications.[32] For the definition of the subject groups we take into

[31]Such as all the literature reported in section 3.4.

[32]Since we want to focus on the choice of a typical student and investigate the effect of his/her parental background, the decision to restrict the analysis to non-mature students, studying full-time, i.e. individuals for which study is the main activity and who are likely

account the predictions of the different hypotheses outlined in section 3.3 and the need to keep the econometric model estimable,[33] and aggregate all subjects in the following three broad subject areas:

1. 'Non-quantitative subjects' (abbreviated as NQS hereafter): Social Studies (excluding Economics), Mass Communications and Documentation, Languages and Related, Humanities, Creative Arts, Education, Combined degrees not included in the following category;

2. 'Quantitative subjects' (abbreviated as QS hereafter): Biological Sciences, Agriculture and Related, Physical Sciences, Mathematical Sciences and Informatics, Engineering and Technology, Architecture, Building and Planning, Economics, Business and Administration Studies, General Sciences Combined degrees;

3. Law and Medicine (L&M hereafter): Law, Medicine and Dentistry, Subjects Allied to Medicine.

Figures 3.1 and 3.2 show the distribution of students by undergraduate field of study and social class, respectively, for all 11 years under study. In

to be more affected by their parental and social background is natural. Mature students may have enrolled in HE after working for a period and have accumulated the financial resources necessary to enroll in HE, in any case they are likely to be more independent of their families. Moreover, the USR data do not provide family background information for mature students. We exclude students with non-traditional entry qualifications into HE since the level of secondary school performance, which we consider as a control for students' ability, is not available. However, A-level entrants represent the vast majority of university students in the period studied (1981-1991).

[33]See section 3.6.

the period 1981-1991, the number of students satisfying our sample selection criteria rose by about 19%, from 48,024 to 57,096 units. However, the rise has been unevenly distributed across social classes, with students from social classes I and V,[34] for instance, rising only by 3.5% and decreasing by 1.2%, respectively, and those from social classes II, IIINM, IIIM and IV, rising by 25.2, 39.1, 9.15 and 52.3 percentage points,[35] respectively. The increase in the number of students was more equally distributed across subject groups. Both QS and NQS experienced an increase of around 20 percentage points, while the increase in L&M was about 7 percentage points lower.[36]

Table 3.1 reports some descriptive statistics on mean A-level score of entrant students by subject,[37] which can be interpreted as a raw measure of subject selectivity. L&M always ranked first in terms of mean A-level score of the students intake, while the second place was occupied by quantitative subjects until 1986 and by non-quantitative subjects from 1987 onwards.

[34]Some abbreviations commonly used in the UK for social classes are: I (professional), II (intermediate), IIINM (skilled non manual), IIIM (skilled manual), IV (partly skilled), V (unskilled). In the USR data social class was built using parents' occupation. We are grateful to Abigail McNight for providing us with the mapping information.

[35]The high increase in social class IV is, however, partly determined by the low initial number of students with this social background in 1981.

[36]Although the figure for this group reflects the slower dynamic for Medicine, since the number of medical students is determined by the Government.

[37]A-level scores are computed according to the UCAS method: A=10, B=8, C=6, D=4, E=2.

Table 3.2 shows average gross weekly occupational earnings of cohorts of student leavers since 1985.[38] For the whole period L&M ranked first in terms of average earnings, followed by quantitative and non-quantitative subjects, respectively. The coefficients of variation of earnings by study field generally show in the period of study a lower variation within QS, and a similar amount of earnings variation within the other subjects.

3.7.3 Choice of explanatory variables

The primary focus of this chapter is on the effect of social class on the choice of undergraduate degree subject. Previous research has identified secondary school curriculum (Polachek, 1978), gender (Polacheck, 1978, Blakemore and Low, 1984) and forward-looking factors (Berger, 1988, Rochat and Demeulemeester, 2001, Montmarquette *et al.*, 2002) as the main determinants of undergraduate field of study.

In the present chapter, we estimate a 'value added specification' of the subject choice model (see equation (3.5)) and do not consider the effect of forward-looking factors, such as expected incomes and academic performance. We decide to do so for several reasons. Firstly, past research has shown that the expected life-time flow of earnings is much more important than starting earnings for students' subject choice (see Berger, 1988) and that this flow is highly uncertain to students (see Wolter and Zbinden, 2002),

[38] The author wishes to thank Abigail McNight, Robin Naylor and Jeremy Smith for providing data on earnings. Weekly occupational earnings are obtained by matching First Destination Supplement data, in the USR, with data from the New Earnings Survey. See Bratti and Mancini (2003) for a more detailed description of the matching procedure.

while from USR data it is possible to have information on graduates' early occupational earnings only (i.e. six months after graduation, using the First Destination Supplement). In the absence of data on subjective earnings expectations by students, the construction of life-time subject specific expected earnings would require a substantial amount of discretion and assumptions on the part of the researcher.[39] Given that the inclusion of expected performance raises similar problems, the latter is also excluded from the present analysis. It must also be noted that our model is a reduced form model and therefore of descriptive nature. We want only to investigate whether there are statistically significant social class differences in the probability of enrolling in different subjects. Then, these differences may be originated by very different factors such as differences in preferences or in the expected economic returns and costs of enrolling in different subjects, but unfortunately we do not have enough information to identify the various channels through

[39]One possibility would be to estimate individuals' expected earnings using individual actual earnings data. Unfortunately, individual earnings data by subject are not available for the period considered (e.g., the UK Labour Force Survey does not contain any information on subject of degree for the period under study). Even when the data are available model identification requires the exclusion of at least one variable affecting earnings but not the degree subject choice directly, whose selection is far from being obvious. Moreover, some studies, such as Dominitz and Manski (1996), Betts (1996), and Brunello *et al.* (2001), cast doubts on students' ability to predict their life-time future earnings. These studies generally show a large heterogeneity in students' beliefs about current earnings, which reflects a large variation in students' information. Finally, as noticed by Dominitz and Manski (1996): 'incorrect assumptions can yield incorrect inferences about the way students make schooling decisions' (p. 3).

which social class may exert its influence.

In detail, we include among the explanatory variables: gender, age at enrollment, secondary school type (not known, grammar, independent, comprehensive, 6th form college, other type), score in A-levels in specific subjects (biology, chemistry, economics, English, French, general studies, geography, history, mathematics, physics),[40] number of A-levels, best three A-level passes score, region of residence prior to university enrollment (inner London, outer London, other England, Scotland, Wales, Northern Ireland) and social class (I, II, IIINM, IIIM, IV, V, armed forces, non-workers, inadequately described).[41]

3.7.4 Models' fit

Table 3.3 reports some statistics for the TNP models estimated for the period 1981-1991. For all years, the Wald test for the null hypothesis that the coefficients on all regressors but the constant are equal to zero is strongly rejected. The Wald tests for the omission of the variables related to social class and pre-university school curriculum show that both sets of variables cannot be omitted from the model. However, it is the latter group of variables which accounts for most of the explanatory power of the model as the pseudo R^2 of the models with and without pre-university school variables show. Last but not least, in all years the estimated correlation ρ_{12} between the error components of the utilities (see section 3.6) of the QS and L&M groups is

[40] In order to keep the model tractable we choose to include only the most popular subjects for which a score as well as a pass indicator is available.

[41] For the explanation of the abbreviations for social classes see footnote 34.

significantly different from zero, suggesting that the TNP model has to be preferred to a MNL model. We have already seen that a non-zero correlation between the random components of the utilities of the different alternatives is what is also suggested by the behavioural model outlined in section 3.5.[42]

3.8 Empirical results on the effect of social class

In this section we comment on the estimated probabilities obtained from the TNP model.

The predicted probabilities of enrolling in the three different subject areas by social class are reported for each year in Table 3.4 and are computed as the means of the individual predicted probabilities. Standard errors and 95% confidence intervals are computed using the delta method and Z critical values. We focus here only on the differences by social class.

Firstly, although Table 3.3 shows that social class variables are jointly highly significant, we observe in Table 3.4 that the predicted probabilities of enrolling in the different subject groups are generally not statistically different across social classes. Apart from statistical differences, we observe that some predictions of the theory are met by the data. In all years considered, except 1982, individuals from social class I had the smallest probability to

[42]Another fact worth noting is the constant decline of this correlation over time, which seems to suggest that the unobservables entering the utilities of the different alternatives are diverging over time.

enroll in QS (cf. hypothesis I in section 3.3). By contrast, the same individuals had in all years but 1991, the highest probability to enroll in Law and Medicine (cf. hypothesis IV in section 3.3). However, the differences are very low in magnitude and this influences their statistical significance. Moreover, there does not appear to be any dramatic change in the likelihood of enrolling in the different subject groups in the period under study.

Thus, our analysis appears to show that the structure of the UK higher education system during the period 1981-1991 was able to ensure that individuals with different social backgrounds had equal opportunities of accessing different subjects at tertiary level (conditional on accessing HE). Recall that the system was characterized by the absence of undergraduate tuition fees and by the provision of means-tested maintenance grants for economically disadvantaged students. Both these features of the UK higher education system were likely to attenuate differences across social classes in the probability of enrolling in the different subjects by making the pecuniary costs of enrolling in the different fields very similar across social classes and subjects.[43] Moreover, in a period in which the number of graduates was not very high, possessing a degree was probably sufficient per-se to ensure high earnings in the labour market and more important than social class origin.[44]

[43]We have already said that in the presence of imperfect capital markets less wealthy students may need to borrow money since they do not have the financial support of their families, and have therefore higher enrollment costs, and, following an economic approach, they are more likely to enroll in high return subjects.

[44]Compared to a situation in which the supply of graduates is high and 'family networks' may be important for employers to screen among graduates.

On the basis of these results it might be interesting to replicate our analysis for more recent years, since the 1990s were characterized by a gradual substitution of the maintenance grants with repayable loans. As stated by Callender (2003) the replacement of student maintenance grants with subsidised loans marked a switch from a system granting a large subsidy to lower income students to a less generous system benefiting all students. Although student loans might be very close to implementing perfect capital markets, since students can borrow against their future incomes at a zero real interest rate, they are surely less generous than maintenance grants for low income students and may have had some consequences not only on student access to HE but also in terms of differentiating students' choices across social classes. Low income students, who are more risk adverse, might have preferred to enroll in less selective and "easier" subjects, in which the probability of failure is lower or that of achieving a 'good' (first or upper second) degree class higher, given the growing importance of degree class over time (see Naylor *et al.* 2003), or to enroll in the subjects in which there is less earnings dispersion. And, of course, it would also be interesting to study the consequences of differentiating the fees by subject and imposing top-up fees for the most popular subjects, which have the potential of producing further unequalising effects on the choices of students from different social classes.

The bulk of explanatory power of our model of subject choice can be ascribed to the type of pre-university school curriculum and performance, proxied by the A-levels score, number of A-levels, school type and type of A-levels with the relative performance. The drop in the pseudo R^2 when these variables are omitted (see Table 3.3) shows that they have a high explanatory power over and above social class (as the models with full controls show), which in turn has only a limited influence on pre-university school curriculum. Indeed, the explanatory power of social class remains low also when secondary school variables, on which the former may have an influence, are not controlled for. This is relatively good news in terms of intergenerational mobility and equal educational opportunities of the UK university system in the 1980s as far as subject choice is concerned: social class did not appear to be the main determinant of students' differences of undergraduate subject studied. Our analysis appears to attribute a major role to other individual or family characteristics, such as parenting quality or students' ability and motivation, or to quality of schools and teachers, which affected the type of curriculum and A-level performance. However, we do not exclude that there might have been other forms of educational inequalities across social classes, for instance, in terms of the type of institution enrolled (polytechnics vs universities, Russel group institution vs other universities).[45]

3.8.1 Sensitivity analysis

We have seen that our econometric analysis excludes the presence of statistically significant differences across individuals from different social classes on the probability of enrolling in different subject groups. However, we may wonder whether there are differences within broadly defined subject groups,

[45]This may happen since there are positive economic returns to attending prestigious Universities as shown by Chevalier and Conlon (2003).

and whether the results are determined by the specific econometric model chosen.

We address both issues by considering a more detailed definition of subject groups. In particular: Medicine, Law, Sciences, Technical, Economics and Business, Mathemathics, Soft Social Sciences, Art and Humanities.[46]. Since the high number of subject groups considered does not allow the estimation of a multinomial probit model, we use instead a flexible-thresholds ordered probit model (see Appendix B). However, estimation of such a model requires an ordering of the discrete dependent variable. We ranked the subjects in ascending order of occupational weekly earnings. In particular, we matched 1985-1991 university students leavers data from the USR with occupational earnings from the NES survey and computed an average of the occupational earnings by subject group. On the basis of the average weekly occupational earnings in each year the subjects were ordered to estimate the flexible-thresholds ordered probit model. The ranking of the subjects is shown in Table 3.5 and is very similar across years. We report in Tables 3.6-3.7 the estimates for a benchmark year, 1985.[47] Concerning 1985, it must be noted that the ordering by weekly occupational earnings of the broader sub-

[46]The composition is: Medicine, Law, Sciences (Biological Sciences, Agriculture, Physical Sciences, General Sciences Combined degrees), Technical (Computing, Engineering, Technology, Architecture), Economics and Business, Mathematics, Soft Social Sciences (Social studies excluding economics, Politics, Mass Communications, General Social Sciences Combined degrees), Art and Humanities (Classics and Literature, Modern Euro Languages, Other Languages, Humanities, Creative Arts, Education, Other combined degrees).

[47]The same sensitivity analysis was also replicated for 1991 and showed the same results.

ject groups shown in Table 3.2 is generally preserved (i.e. Law and Medicine graduates had higher earnings than all QS graduates who in turn earned more than NQS graduates), the sole exception being Sciences, whose graduates were at the bottom of the earnings distribution. Table 3.6 shows that by considering a finer disaggregation of subject groups, imposing an ordering and using a different econometric model our results do not change: differences across social classes are never statistically significant. Moreover, Table 3.6 shows that when we compute the aggregate probabilities of the broader subjects considered in the previous section (QS, NQS, Law and Medicine) by summing the probabilities of the finer subjects, they are very similar to those obtained from the TNP model. In summary, our results do not appear to be driven by the aggregation of subjects or by the type of econometric model chosen.

3.9 Concluding remarks

In this chapter we investigated social class influences on degree subject choice at the undergraduate level, conditional on enrolling in HE in the UK. We have used data for several cohorts of university students in the 1980s to estimate a trinomial probit model of subject choice (Quantitative Subjects, Non-Quantitative Subjects, Law and Medicine). We have also considered a finer disaggregation of subjects and an alternative econometric model (a flexible-thresholds ordered probit) and showed the robustness of our results. From our empirical analysis:

1. We do not find statistically significant differences among social classes in the probability of enrolling in different subjects in the period 1981-1991.

2. In the 'value added specification' of our model (i.e., controlling for secondary school variables) social class explains only a small part of the variation in subject enrolled, while secondary school curriculum (school type, A-level score and number and performance in specific A-levels) has a much higher explanatory power. The explanatory power of social class remains low also in a 'contemporaneous specification' of the model of subject enrolled, where secondary school variables are omitted, suggesting that its role may be limited also at early stages of the educational process.

Both findings can be interpreted as good news in terms of intergenerational mobility of the UK university system during the 1980s as far as subject choice is concerned, in the sense that a student's choice of study field was made on the basis of characteristics generally unrelated to social class.

For future research, it would be interesting to replicate the analysis of this chapter for more recent cohorts of university students. Indeed, the recent changes introduced in the UK university system, such as the gradual replacement of student maintenance grants with student loans, may have contributed to differentiating the degree subject choices of students with different social class backgrounds and may give some useful insights into the potential effects of introducing differential fees by subject.

Appendix B: The flexible-thresholds ordered probit model

In order to check the robustness of our results to different modeling strategies we estimated the subject choice model for the 1985 and 1991 using a flexible-thresholds ordered probit model. The latter represents a straightforward extension of the standard ordered probit model. In what follows we give a brief outline of both the standard ordered probit model and of the flexible-thresholds ordered probit model.

Ordered probit model

Following Cameron and Heckman (1998), let us assume that the direct costs of undertaking undergraduate education in field j, given the characteristics of individual i, $X = x_i$, are $c(j|x_i)$, which are assumed to be weakly convex and increasing in j. They might reflect, for instance, the disutility produced by the time devoted to study. The underlying idea is that since we order the subjects according to the level of income earned, high return subjects are also those 'tougher' in terms of academic selectivity and effort required *ceteris paribus* of students.[48] The X's, being individuals' characteristics, are the same across all university fields. Let us also assume that the discounted lifetime return to undergraduate field j is $R(j|x_i, \epsilon_i)$, concave and increasing in j (which directly derives from our ordering of subjects), where ϵ_i is a person

[48] At this stage the model does not allow for heterogeneity across students in the cost of subjects. Such a possibility will be discussed later on when presenting the flexible-thresholds ordered probit model.

specific shifter of the return to field j. We define the utility for individual i in subject j, i.e. V_{ij}, as the difference between the return and the cost of acquiring education in field j. The optimal field of study is determined for each individual by solving the problem:

$$\underset{j}{Max}[R(j|x_i, \epsilon_i) - c(j|x_i)] \tag{3.18}$$

where $j = 1, ..., J$, and J is the field with the highest economic return (which does not necessarily imply that it is also the highest utility field for individual i).

Let us assume that ϵ_i is stochastic, such that $\epsilon_i \perp X_i$ and:

$$R(j|x_i, \epsilon_i) = R(j)\psi(x_i)\epsilon_i \tag{3.19}$$

where $E(\epsilon_i) = 1$, $\epsilon_i \geq 0$, while

$$c(j|x_i) = c(j) \tag{3.20}$$

i.e. costs do not depend on individual characteristics.

If s is the optimal undergraduate field for individual i then:

$$\frac{[c(s) - c(s-1)]}{(R(s) - R(s-1))\psi(x_i)} \leq \epsilon_i \leq \frac{[c(s+1) - c(s)]}{(R(s+1) - R(s))\psi(x_i)}. \tag{3.21}$$

ϵ_i is therefore bounded by the ratios of marginal return to the marginal cost of the different undergraduate fields. If ϵ_i is continuously distributed and defining:

$$\exp[l(j)] = \frac{c(j) - c(j-1)}{R(j) - R(j-1)} \tag{3.22}$$

then:

$$Pr(j = s|X = x_i) = Pr[\exp[l(s)]\psi(x_i)^{-1} \leq \epsilon_i \leq \exp[l(s+1)]\psi(x_i)^{-1}]. \tag{3.23}$$

If we further assume that $\psi(x_i) = \exp(x_i\beta)$ and that ϵ_i is log-normally distributed, the expression above takes the more familiar form of the standard *ordered probit model*:

$$Pr(j = s|X = x_i) = \Phi[l(s) - x_i\beta \leq \mu \leq l(s+1) - x_i\beta] \tag{3.24}$$

where $\mu = ln(\epsilon_i)$. The parameters to be estimated are β and the thresholds $l(j)'s$. In this framework individual and parental attributes increase or decrease the return to education and therefore affect the probability of enrolling in the different fields. The specification can be adjusted so that it can allow for both the return and costs to education to depend on individual and family characteristics (see, for instance, Lauer, 2003). However, this has no empirical relevance, since only the effect of the covariates on the ratio of the marginal return to the marginal cost of the educational qualifications (see equation (3.24)), and not on the single components (return and cost), can be identified.

Flexible-thresholds ordered probit model

In this section, we offer an economic rationalization for the flexible-thresholds ordered probit (FT-OP, hereafter) model, introduced by Pradhan and van Soest (1995), in the spirit of Cameron and Heckman (1998). Let us start from the same theoretical framework introduced for the simple ordered probit model and define the cost functions for the different fields as follows:

$$c(s|x_i) = \sum_{j=1}^{s} \exp(\delta_1) \prod_{z=2}^{j} \exp(\delta_j(x_i)), \tag{3.25}$$

and the return functions as follows:

$$R(s|x_i, \epsilon_i) = \epsilon_i \sum_{j=1}^{s} \exp(\phi_1)\psi(x_i) \prod_{z=2}^{j} \exp(\phi_j(x_i)). \tag{3.26}$$

Hence, both cost and return depend on individuals characteristics. We assume that students enroll in the field which maximize their utility. If $1 < s < J$ is the optimal undergraduate field for individual i, then in our specific case:

$$\frac{\exp(\delta_1) \prod_{j=2}^{s} \exp(\delta_j(x_i))}{\exp(\phi_1) \prod_{j=2}^{s} \exp(\phi_j(x_i))\psi(x)} \leq \epsilon_i \leq \frac{\exp(\delta_1) \prod_{j=2}^{s+1} \exp(\delta_j(x_i))}{\exp(\phi_1) \prod_{j=2}^{s+1} \exp(\phi_j(x_i))\psi(x)}. \tag{3.27}$$

By further assuming that $\phi_j(x_i) = \exp(\phi_j x_i)$ and $\delta_j(x_i) = \exp(\delta_j x_i)$ and $\psi(x_i) = \exp(x_i\beta)$ the last expression can be rewritten as:

$$\frac{\exp(\delta_1) \prod_{j=2}^{s} \exp(\exp(\delta_j x_i))}{\exp(\phi_1)\exp(x_i\beta) \prod_{j=2}^{s} \exp(\exp(\phi_j x_i))} \leq \epsilon_i \leq \frac{\exp(\delta_1) \prod_{j=2}^{s+1} \exp(\exp(\delta_j x_i))}{\exp(\phi_1)\exp(x_i\beta) \prod_{j=2}^{s+1} \exp(\exp(\phi_j(x_i)))}. \tag{3.28}$$

If we define $\exp(l(s,x_i)) \equiv \exp(\delta_1-\phi_1)\prod_{j=2}^{s}\exp((\delta_j-\phi_j)x_i)$, and $\delta_j-\phi_j \equiv \lambda_j$ and assume that ϵ_i is log-normally distributed,

$$Pr(j=s|X=x_i)=\Phi\big[\lambda_1+\sum_{j=2}^{s}\exp(\lambda_j x_i)-\beta x_i \leq \mu_i \leq \lambda_1+\sum_{j=2}^{s+1}\exp(\lambda_j x_i)-\beta x_i\big]. \tag{3.29}$$

where $\mu_i = ln(\epsilon_i)$, we obtain the flexible-thresholds ordered probit model introduced by Pradhan and van Soest (1995). The thresholds $l(s,x_i)$ are allowed to depend on the variables x_i. As Pradhan and van Soest (1995) observed, this model allows greater flexibility compared to the standard ordered probit model. Indeed, model identification only requires one threshold to be fixed. Therefore the generality of the model can be increased by letting the other thresholds depend on individual characteristics. In particular, while the choice of the lowest return field depends on the index $x_i\beta$ only, the choices of the other subjects also depend on the $x_i\lambda_j$'s indexes. Despite being more flexible than the ordered probit model the flexible-thresholds ordered probit model requires the ordering of the outcome variable and it is therefore less general than a multinomial probit model although much easier to estimate.

Tables

Table 3.1: Average A-level score of entrant students by subject (USR data)

Year	QS		NQS		L&M	
	mean	c.v.	mean	c.v.	mean	c.v.
1981	20.418	0.308	20.220	0.292	23.774	0.195
1982	21.130	0.281	21.071	0.264	24.214	0.179
1983	22.020	0.247	21.699	0.237	24.663	0.161
1984	22.123	0.250	21.849	0.237	25.027	0.155
1985	22.076	0.255	21.657	0.239	25.055	0.157
1986	21.947	0.259	21.774	0.234	25.140	0.153
1987	21.728	0.266	21.897	0.231	24.910	0.164
1988	21.794	0.268	21.971	0.231	25.015	0.161
1989	21.883	0.270	22.324	0.214	25.199	0.158
1990	21.939	0.269	22.631	0.208	25.097	0.164
1991	21.710	0.277	23.129	0.197	25.110	0.172

Note. QS: 'Quantitative Subjects'; NQS: 'Non-Quantitative Subjects'; L&M: Law and Medicine (see section 3.7.2). The table reports average A-level scores of entrant students by subject and the coefficient of variation (c.v.) within subject groups. The average A-level score are computed on the estimation samples (i.e. non-mature, non-overseas, unmarried students studying for a first degree qualification and with A-levels).

Table 3.2: Average gross weekly occupational earnings of leaving students by subject - 3 groups

Year	QS		NQS		L&M	
	mean	c.v.	mean	c.v.	mean	c.v.
1985	445.5	0.2	430.5	0.2	564.6	0.2
1986	487.1	0.2	468.0	0.2	612.1	0.2
1987	537.4	0.2	524.8	0.2	699.8	0.2
1988	598.7	0.2	567.9	0.2	785.6	0.2
1989	656.6	0.2	626.9	0.2	871.6	0.2
1990	677.1	0.2	661.3	0.2	980.9	0.2
1991	710.7	0.2	690.8	0.3	1044.7	0.3
1992	735.7	0.2	716.2	0.3	1071.4	0.3
1993	761.4	0.2	747.1	0.3	1085.3	0.3

Note. QS: 'Quantitative Subjects'; NQS: 'Non-Quantitative Subjects'; L&M: Law and Medicine (see section 3.7.2). The table reports average gross weekly occupational earnings in current pounds (from the New Earnings Survey) of leaving students by subject and the coefficient of variation (c.v.) within subjects. Average earnings are computed on the estimation samples (i.e. non-mature, non-overseas, unmarried students studying for a first degree qualification and with A-levels).

Table 3.3: Model diagnostics

Statistics	Years						
	1981	1982	1983	1984	1985	1986	1987
Log-likelihood	-27,978.0	-26,887.6	-26,231.0	-26,826.2	-27,787.3	-27,678.0	-28,038.7
Wald test: overall significance[(a)]	13,752.1	13,718.4	14,206.0	14,677.4	16,015.6	16,027.0	15,872.1
(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wald test: social class[(b)]	80.80	203.98	97.76	52.66	98.29	113.92	95.19
(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wald test: pre-university variables[(c)]	12,224.1	12,455.7	13,035.9	13,357.4	14,645.2	14,590.8	14,424.0
(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wald test: school type[(d)]	120.3	88.6	46.8	88.8	78.8	113.3	110.4
(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo R^2:[(e)]							
i. model with secondary education	0.41	0.41	0.41	0.41	0.41	0.40	0.40
ii. model withouth secondary education	0.05	0.05	0.04	0.04	0.05	0.04	0.04
ρ_{12}[(f)]	0.70	0.73	0.74	0.70	0.63	0.61	0.61
(s.e.)	0.03	0.03	0.03	0.03	0.03	0.03	0.03
N. observations	48,024	45,837	44,412	45,336	47,601	46,568	46,914

continued

Statistics	Years			
	1988	1989	1990	1991
Log-likelihood	-28,905.1	-31,239.1	-33,443.6	-34,060.5
Wald test: overall significance[a]	17,243.3	19,522.9	20,236.2	22,006.7
(p-value)	0.00	0.00	0.00	0.00
Wald test: social class[b]	86.91	77.74	98.75	88.28
(p-value)	0.00	0.00	0.00	0.00
Wald test: pre-university variables[c]	15,668.5	17,767.9	18,547.6	20,280.8
(p-value)	0.00	0.00	0.00	0.00
Wald test: school type[d]	116.9	131.7	175.6	166.0
(p-value)	0.00	0.00	0.00	0.00
Pseudo R^2:[e]				
i. model with secondary education	0.39	0.40	0.39	0.39
ii. model withouth secondary education	0.04	0.04	0.04	0.05
ρ_{12}[f]	0.56	0.48	0.41	0.39
(s.e.)	0.04	0.04	0.04	0.05
N. observations	48,303	52,871	55,782	57,096

Note. [a] $Chi^2(64)$; [b] $Chi^2(16)$; [c] $Chi^2(24)$. Pre-university variables include A-level number, A-level score (UCAS formula) and A-level performance by subject; [d] $Chi^2(10)$; [e] The Pseudo R^2 is computed as: 1-(loglikelihood only-constant model/loglikelihood full model); [f] Coefficient of correlation between the alternatives Law and Medicine and Quantitative Subjects.

Table 3.4: Predicted probabilities (%), with standard errors and 95% confidence intervals, of enrolling in the different study fields by social class

Social Class	Subject group QS Prob	QS left	QS right	NQS Prob	NQS left	NQS right	L&M Prob	L&M left	L&M right
1981									
I	0.48	0.46	0.51	0.36	0.34	0.38	0.16	0.13	0.18
II	0.50	0.48	0.52	0.37	0.36	0.39	0.13	0.11	0.15
IIINM	0.49	0.47	0.51	0.37	0.36	0.39	0.13	0.11	0.15
IIIM	0.50	0.48	0.53	0.37	0.36	0.39	0.12	0.10	0.14
IV	0.49	0.46	0.51	0.39	0.37	0.41	0.12	0.10	0.15
V	0.52	0.48	0.57	0.36	0.33	0.40	0.12	0.08	0.15
Armed forces	0.48	0.44	0.52	0.40	0.37	0.43	0.12	0.09	0.15
Inadequately described	0.50	0.47	0.53	0.37	0.35	0.40	0.12	0.10	0.15
Non workers	0.49	0.45	0.52	0.39	0.36	0.41	0.12	0.10	0.15
1982									
I	0.47	0.45	0.49	0.36	0.34	0.38	0.17	0.14	0.19
II	0.49	0.47	0.51	0.37	0.35	0.39	0.14	0.12	0.16
IIINM	0.51	0.48	0.53	0.36	0.34	0.38	0.13	0.11	0.15
IIIM	0.51	0.49	0.54	0.36	0.35	0.38	0.12	0.10	0.14
IV	0.50	0.47	0.53	0.36	0.34	0.38	0.14	0.11	0.16
V	0.47	0.43	0.52	0.36	0.33	0.40	0.16	0.12	0.21
Armed forces	0.49	0.45	0.53	0.37	0.34	0.40	0.14	0.11	0.17
Inadequately described	0.49	0.46	0.52	0.38	0.36	0.41	0.13	0.10	0.15
Non workers	0.37	0.34	0.41	0.37	0.34	0.39	0.26	0.23	0.29
1983									
I	0.46	0.44	0.49	0.36	0.34	0.38	0.18	0.16	0.20
II	0.49	0.47	0.52	0.36	0.35	0.38	0.15	0.12	0.17
IIINM	0.50	0.47	0.52	0.36	0.34	0.37	0.15	0.12	0.17
IIIM	0.51	0.48	0.53	0.36	0.34	0.38	0.13	0.11	0.16
IV	0.51	0.48	0.53	0.36	0.34	0.38	0.14	0.11	0.16
V	0.52	0.48	0.57	0.38	0.34	0.41	0.10	0.07	0.14
Armed forces	0.47	0.43	0.51	0.38	0.34	0.41	0.15	0.12	0.19
Inadequately described	0.48	0.45	0.51	0.36	0.34	0.38	0.16	0.13	0.19
Non workers	0.48	0.45	0.52	0.37	0.35	0.40	0.14	0.12	0.17
1984									
I	0.47	0.45	0.50	0.36	0.34	0.37	0.17	0.15	0.19
II	0.49	0.47	0.51	0.36	0.35	0.38	0.15	0.13	0.16
IIINM	0.49	0.47	0.52	0.36	0.34	0.38	0.15	0.13	0.17
IIIM	0.50	0.48	0.52	0.36	0.34	0.38	0.14	0.12	0.16
IV	0.49	0.47	0.52	0.36	0.34	0.38	0.15	0.13	0.17
V	0.48	0.43	0.53	0.38	0.35	0.41	0.14	0.10	0.18
Armed forces	0.47	0.43	0.51	0.38	0.35	0.41	0.15	0.12	0.19
Inadequately described	0.48	0.45	0.51	0.36	0.34	0.39	0.15	0.13	0.18
Non workers	0.47	0.43	0.50	0.39	0.36	0.41	0.15	0.12	0.18
1985									
I	0.49	0.47	0.51	0.35	0.33	0.36	0.17	0.15	0.19
II	0.51	0.49	0.53	0.36	0.34	0.37	0.14	0.12	0.15
IIINM	0.51	0.49	0.54	0.35	0.33	0.37	0.13	0.11	0.15
IIIM	0.52	0.50	0.54	0.35	0.33	0.36	0.13	0.11	0.15
IV	0.51	0.48	0.53	0.35	0.33	0.37	0.14	0.12	0.16
V	0.52	0.48	0.56	0.35	0.32	0.39	0.12	0.09	0.15
Armed forces	0.49	0.45	0.53	0.38	0.35	0.41	0.12	0.09	0.15
Inadequately described	0.52	0.49	0.55	0.35	0.32	0.37	0.13	0.11	0.16
Non workers	0.49	0.46	0.52	0.38	0.35	0.40	0.13	0.11	0.16

continued

Social Class	QS Prob	QS left	QS right	NQS Prob	NQS left	NQS right	L&M Prob	L&M left	L&M right
1986									
I	0.48	0.46	0.50	0.35	0.33	0.37	0.17	0.15	0.19
II	0.51	0.49	0.53	0.36	0.34	0.37	0.14	0.12	0.16
IIINM	0.51	0.49	0.54	0.35	0.33	0.37	0.14	0.12	0.16
IIIM	0.53	0.50	0.55	0.34	0.32	0.36	0.13	0.11	0.15
IV	0.52	0.49	0.54	0.34	0.32	0.36	0.14	0.12	0.16
V	0.46	0.42	0.50	0.37	0.33	0.40	0.17	0.13	0.21
Armed forces	0.52	0.48	0.56	0.36	0.33	0.39	0.12	0.09	0.15
Inadequately described	0.51	0.48	0.54	0.34	0.32	0.37	0.15	0.12	0.17
Non workers	0.50	0.46	0.53	0.36	0.33	0.39	0.14	0.12	0.17
1987									
I	0.48	0.46	0.50	0.35	0.33	0.37	0.17	0.15	0.19
II	0.51	0.49	0.53	0.35	0.33	0.37	0.14	0.12	0.16
IIINM	0.53	0.50	0.55	0.34	0.32	0.36	0.13	0.12	0.15
IIIM	0.52	0.50	0.55	0.34	0.32	0.35	0.14	0.12	0.16
IV	0.51	0.49	0.54	0.34	0.32	0.36	0.14	0.12	0.17
V	0.53	0.48	0.57	0.33	0.29	0.36	0.14	0.11	0.18
Armed forces	0.51	0.47	0.55	0.34	0.31	0.37	0.15	0.11	0.18
Inadequately described	0.51	0.48	0.54	0.34	0.32	0.36	0.15	0.13	0.18
Non workers	0.54	0.50	0.57	0.35	0.32	0.37	0.12	0.09	0.14
1989									
I	0.49	0.47	0.51	0.37	0.35	0.38	0.15	0.13	0.16
II	0.51	0.49	0.53	0.36	0.34	0.37	0.13	0.11	0.14
IIINM	0.53	0.51	0.55	0.35	0.34	0.37	0.12	0.10	0.14
IIIM	0.52	0.50	0.54	0.35	0.34	0.37	0.13	0.11	0.14
IV	0.51	0.48	0.53	0.35	0.34	0.37	0.14	0.12	0.16
V	0.52	0.47	0.56	0.34	0.31	0.38	0.14	0.11	0.18
Armed forces	0.51	0.47	0.55	0.38	0.34	0.41	0.12	0.09	0.14
Inadequately described	0.52	0.50	0.55	0.34	0.32	0.37	0.13	0.11	0.16
Non workers	0.53	0.50	0.56	0.36	0.33	0.38	0.12	0.09	0.14
1990									
I	0.48	0.46	0.50	0.38	0.36	0.39	0.15	0.13	0.16
II	0.50	0.48	0.52	0.38	0.36	0.39	0.12	0.11	0.14
IIINM	0.51	0.49	0.53	0.36	0.35	0.38	0.13	0.11	0.15
IIIM	0.51	0.49	0.53	0.36	0.35	0.38	0.13	0.11	0.15
IV	0.50	0.48	0.53	0.36	0.34	0.38	0.13	0.11	0.15
V	0.51	0.46	0.55	0.35	0.31	0.38	0.14	0.11	0.18
Armed forces	0.50	0.46	0.54	0.40	0.37	0.43	0.10	0.08	0.13
Inadequately described	0.49	0.46	0.52	0.36	0.33	0.38	0.15	0.13	0.18
Non workers	0.51	0.49	0.54	0.35	0.33	0.38	0.13	0.11	0.16
1991									
I	0.48	0.46	0.50	0.38	0.36	0.39	0.14	0.13	0.16
II	0.50	0.48	0.52	0.38	0.36	0.39	0.12	0.11	0.14
IIINM	0.51	0.49	0.53	0.36	0.35	0.38	0.12	0.11	0.14
IIIM	0.51	0.49	0.53	0.37	0.35	0.38	0.13	0.11	0.14
IV	0.51	0.48	0.53	0.36	0.34	0.38	0.14	0.12	0.16
V	0.50	0.46	0.55	0.34	0.31	0.38	0.16	0.12	0.20
Armed forces	0.50	0.46	0.54	0.38	0.35	0.41	0.12	0.09	0.15
Inadequately described	0.49	0.46	0.51	0.38	0.36	0.40	0.13	0.11	0.15
Non workers	0.52	0.48	0.55	0.35	0.32	0.37	0.14	0.11	0.17

Note. QS: 'Quantitative Subjects'; NQS: 'Non-Quantitative Subjects'; L&M: Law and Medicine (see section 3.7.2). Social Classes: I (professional), II (intermediate), IIINM (skilled non manual), IIIM (skilled manual), IV (partly skilled), V (unskilled). Standard errors and confidence intervals were computed using the delta method and Z critical values.

Table 3.5: Average gross weekly occupational earnings of leaving students by subject - 8 groups

Cohorts of leaving students	QS								NQS				L&M			
	Sciences		Maths		Technical		Economics and Business		Soft SS		Art & Humanities		Law		Medicine	
	mean	rank	mean	rank	mean	rank	mean	rank	mean	rank	mean	rank	mean	rank	mean	rank
1985	422	1	476	6	443	4	475	5	429	2	431	3	504	7	605	8
1986	458	1	519	5	482	4	528	6	465	2	469	3	553	7	650	8
1987	506	1	571	5	530	4	589	6	523	2	525	3	647	7	730	8
1988	565	2	641	6	601	4	635	5	563	1	569	3	704	7	832	8
1989	617	1	700	6	656	4	699	5	628	3	627	2	829	7	899	8
1990	647	1	722	6	678	4	701	5	656	2	663	3	988	8	977	7
1991	682	2	743	6	720	4	725	5	675	1	696	3	984	7	1082	8
1992	706	2	766	6	744	4	754	5	701	1	722	3	1019	8	1103	7
1993	717	1	794	5	775	4	796	6	729	2	754	3	1027	7	1122	8

Note. The table reports average gross weekly occupational earnings (from the New Earnings Survey) of leaving students by subject and the rank for each year. Average earnings are computed on the estimation samples (i.e. non-mature, non-overseas, unmarried students studying for a first degree qualification and with A-levels). For the definition of the subject groups see section 3.8.1.

Table 3.6: Predicted probabilities, with standard errors and 95% confidence intervals, of enrolling in the different study fields by social class - flexible-threshold ordered probit model, 1985 USR data

		Social class								
Subject		I	II	IIINM	IIIM	IV	V	Armed forces	I.D.	Non workers
	Prob	0.22	0.24	0.23	0.24	0.24	0.26	0.22	0.25	0.23
Sciences	left	0.21	0.22	0.22	0.22	0.22	0.22	0.19	0.23	0.20
	right	0.24	0.25	0.25	0.26	0.26	0.30	0.26	0.27	0.26
	Prob	0.09	0.10	0.10	0.10	0.10	0.10	0.12	0.10	0.11
Soft Social Sciences	left	0.08	0.08	0.08	0.08	0.09	0.07	0.09	0.08	0.09
	right	0.11	0.11	0.11	0.11	0.12	0.12	0.14	0.11	0.13
	Prob	0.25	0.26	0.25	0.25	0.26	0.25	0.26	0.25	0.26
Art and Humanities	left	0.24	0.24	0.24	0.23	0.24	0.22	0.23	0.23	0.24
	right	0.27	0.27	0.27	0.27	0.27	0.28	0.29	0.27	0.29
	Prob	0.12	0.11	0.12	0.12	0.12	0.10	0.14	0.12	0.11
Technical	left	0.11	0.10	0.10	0.11	0.10	0.07	0.11	0.10	0.09
	right	0.13	0.13	0.13	0.13	0.13	0.13	0.16	0.14	0.13
	Prob	0.08	0.09	0.10	0.09	0.09	0.09	0.08	0.09	0.08
Economics and Business	left	0.07	0.08	0.08	0.07	0.07	0.07	0.06	0.07	0.06
	right	0.10	0.10	0.11	0.10	0.10	0.12	0.10	0.11	0.10
	Prob	0.05	0.05	0.05	0.06	0.05	0.07	0.05	0.05	0.07
Maths	left	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.05
	right	0.06	0.06	0.06	0.07	0.06	0.09	0.07	0.06	0.09
	Prob	0.07	0.06	0.05	0.06	0.05	0.05	0.04	0.06	0.06
Law	left	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.04	0.03
	right	0.08	0.07	0.07	0.08	0.06	0.07	0.06	0.08	0.08
	Prob	0.11	0.09	0.10	0.09	0.10	0.09	0.09	0.09	0.09
Medicine	left	0.09	0.08	0.08	0.08	0.08	0.06	0.07	0.07	0.07
	right	0.12	0.11	0.11	0.11	0.12	0.11	0.11	0.10	0.11

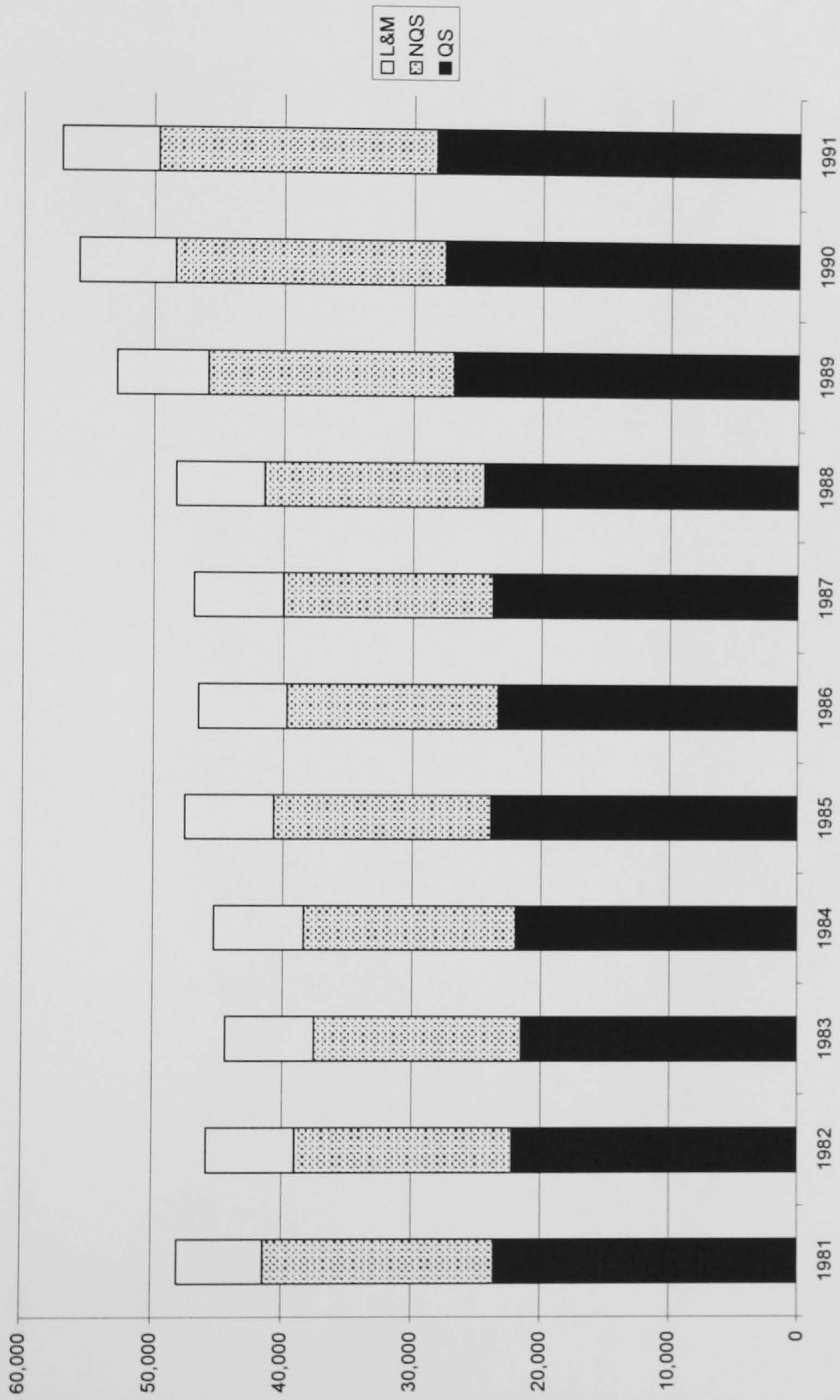
Note. QS: 'Quantitative Subjects'; NQS: 'Non-Quantitative Subjects'; L&M: Law and Medicine (see section 3.7.2). Social Classes: I (professional), II (intermediate), IIINM (skilled non manual), IIIM (skilled manual), IV (partly skilled), V (unskilled). Standard errors and confidence intervals were computed using the delta method and Z critical values.

Table 3.7: Predicted probabilities, with standard errors and 95% confidence intervals, of enrolling in the broader subject groups by social class - flexible-threshold ordered probit model, 1985 USR data

Social Class	Subject group								
	QS			NQS			L&M		
	Prob	left	right	Prob	left	right	Prob	left	right
I	0.48	0.46	0.50	0.35	0.33	0.36	0.17	0.15	0.20
II	0.50	0.48	0.52	0.35	0.34	0.37	0.15	0.13	0.17
IIINM	0.50	0.47	0.52	0.35	0.33	0.37	0.15	0.13	0.18
IIIM	0.50	0.48	0.52	0.35	0.33	0.37	0.15	0.13	0.18
IV	0.49	0.47	0.52	0.36	0.34	0.38	0.15	0.12	0.17
V	0.52	0.47	0.57	0.35	0.31	0.38	0.13	0.09	0.18
Armed forces	0.49	0.45	0.53	0.38	0.35	0.41	0.13	0.09	0.17
Inadequately described	0.51	0.48	0.54	0.34	0.32	0.37	0.15	0.12	0.18
Non workers	0.49	0.45	0.52	0.37	0.34	0.40	0.14	0.11	0.18

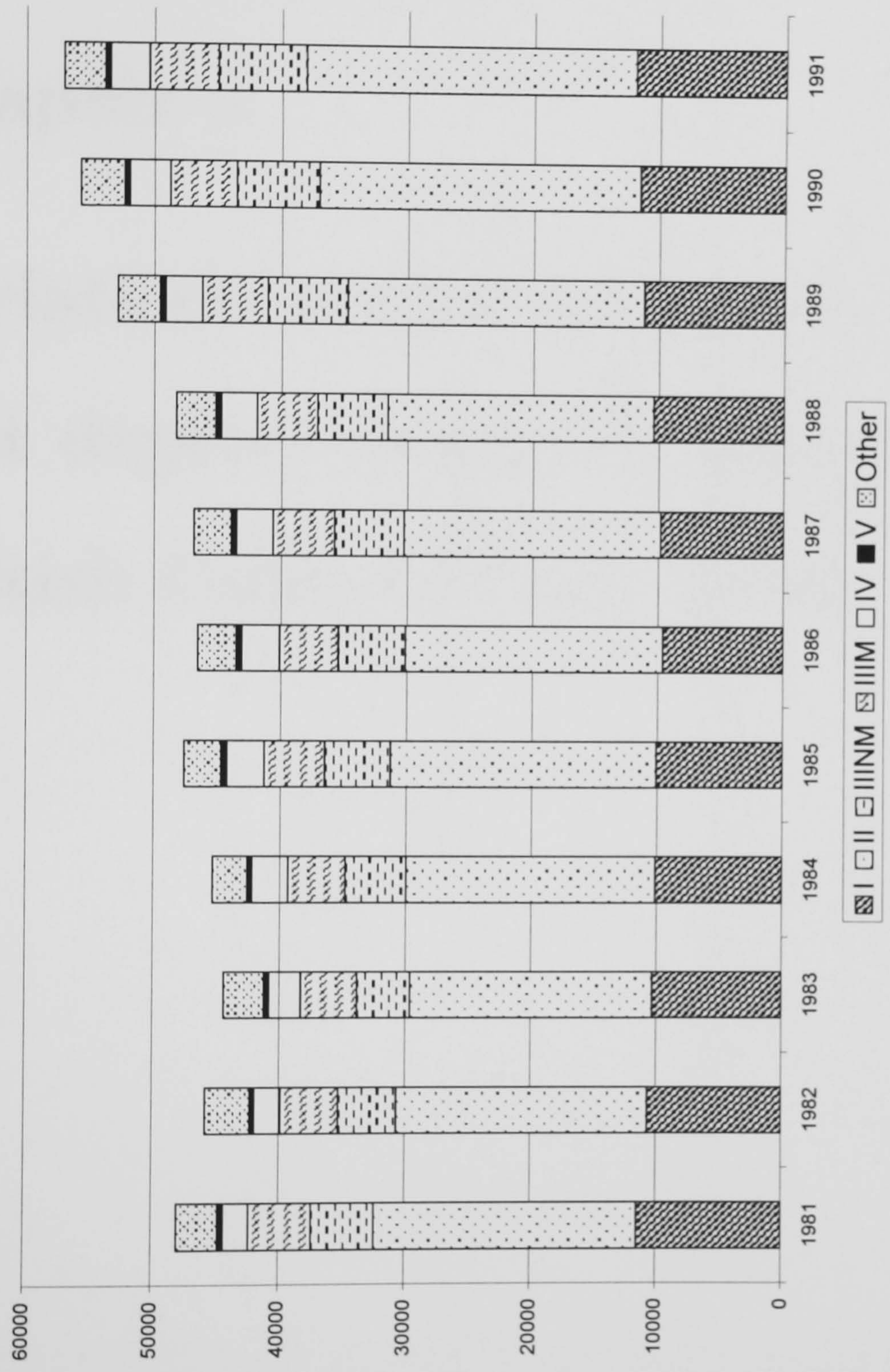
Note. QS: 'Quantitative Subjects'; NQS: 'Non-Quantitative Subjects'; L&M: Law and Medicine (see section 3.7.2). Social Classes: I (professional), II (intermediate), IIINM (skilled non manual), IIIM (skilled manual), IV (partly skilled), V (unskilled). Standard errors and confidence intervals were computed using the delta method and Z critical values.

Figure 3.1: Distribution of students by broad field of study, 1981-1991 USR data



Note. QS: 'quantitative subjects'; NQS: 'non-quantitative subjects'; L&M: law and medicine (see section 3.7.2). The graph refers to full-time, unmarried, non-mature, home students with A-levels and aiming at a first degree qualification (i.e. the estimation sample).

Figure 3.2: Distribution of students by social class, 1981-1991 USR data



Note. Social Classes: I (professional), II (intermediate), IIINM (skilled non manual), IIIM (skilled manual), IV (partly skilled), V (unskilled), ID (inadequately described). The graph refers to full-time, unmarried, non-mature, home students with A-levels and aiming at a first degree qualification (i.e. the estimation sample).

Chapter 4

Variations in the returns to a first degree: evidence from the British Cohort Study 1970

4.1 Introduction

Higher education policy in Europe is in flux, not least in the UK which has witnessed considerable and ongoing policy change over the last half-century. One aspect of the UK experience has been a steady shift in the burden of funding higher education (HE, hereafter) away from the taxpayer and towards students and their families. Maintenance grant provision has been removed substantially and has been replaced by a system of repayable loans. Furthermore, since 1999, uniform university tuition fees have been paid by all full-time UK university students from within the European Union. The ceiling on fees has been recently increased to £3,000 and English universities will be free to charge 'top-up fees' from September 2006. Contemporaneously, there has been a significant expansion in the HE participation rate since the early 1980s, associated both with a reduction in the prior academic performance required for university admission and in the unit of resource in the teaching of university undergraduates.

In this context of ongoing policy change, it is important to examine the magnitude of private returns to HE and the extent to which they have changed over time. Using data on the 1958 British birth cohort from the National Child Development Study (NCDS), Blundell *et al.* (2000) report an estimated return to a degree of around 17% for men and 37% for women, relative to a control group who obtained one or more A-levels (the highest secondary school qualification) but who did not proceed into HE. In part, estimates of sizeable private returns to university degrees have been cited as evidence in support of policies shifting the burden of costs on to students.

Graduates in the cohort analysed by Blundell *et al.* (2000) would most typically have graduated circa 1979, at just about the time that public sector financial support to university students began to decline significantly. Also at this time, UK government policy changes sought to raise substantially the HE participation rate. Rapid expansion of student numbers since the early 1980s is likely to have exerted downward pressure on average returns to a degree. Against this, skill-biased technical change (SBTC) during the last two decades of the twentieth century is likely to have increased the demand for graduate labour. The direction of the net effect of these changes on graduate returns is ambiguous. It seems timely, therefore, to update the estimates obtained by Blundell *et al.* (2000) with the application of their analysis to data for a more recent cohort. We also note that, while research has concentrated on average returns to qualifications, the issue of variations according to level of performance, given qualifications, is under-explored. In the current chapter, we examine both the average returns to a degree and also variations by specific factors. In particular, we address the argument that over a period in which the graduate population has expanded, better-performing graduates might have experienced a relative increase in the premium for a 'good' degree performance (see Naylor and Smith, 2004), thereby raising the returns for such students.[1]

Section 4.2 of this chapter provides a brief review of evidence on trends in the returns to a degree in the UK. The subsequent analysis conducted in

[1]In the UK undergraduate degrees are classified, according to degree performance, in descending order as first, upper second, lower second, third, non-honour degree, fail. First and upper second class degrees are often referred to as 'good' degrees.

this chapter is based on data from the 1970 British Cohort Study (BCS70). This cohort would, typically, have graduated circa 1991, some 12 years after the 1958 cohort analysed in Blundell *et al.* (2000). In view of the various supply and demand-side changes occurring between the late 1970s and the early 1990s and their likely effect on graduate returns, this 1970 cohort is interesting to contrast with the 1958 cohort. It is also of particular interest to address the question of how returns to higher education vary by specific observed characteristics. In the current chapter, we focus on the issue of how returns to a degree vary with both (i) the class of degree awarded and (ii) the subject of the degree studied.

Variation in returns by class of degree has received relatively little attention in the literature. This is largely a consequence of the fact that few data-sets contain adequate information on class of degree awarded. The issue is of interest, however, for two reasons. First, if there is significant variation by degree class around the average return to a degree, then the investment in HE could yield a low return to poor-performing students. Shifting the burden of university fees further towards students then risks generating a greater disincentive to HE participation than would be the case with relatively little variation around the average: a narrow focus on the average return is inadequate for policy purposes. Second, it is of general interest to examine the extent to which the labour market rewards the graduate's class of degree. Estimates of returns to education have tended to focus on years of schooling or on levels of qualifications. Yet, as there is substantial clustering of labour market entrants on both these criteria, one would expect employers to discriminate between candidates on factors such as grades achieved: that is, on degree class awarded in the context of higher education in the UK. This itself is likely to vary with the proportion of a cohort investing in a university degree.

Variation in returns by degree subject has received more attention, as we discuss in more detail below. Since the introduction of flat-rate fees, a number of authors have argued that there is a theoretical case for differentiating fees by subject (see, for example, Greenaway and Haynes, 2003). The strength of the case for differentiating fees depends in part on the strength of evidence that the return to a degree varies by subject studied and/or by institution attended. At the time of writing, legislation to introduce differential fees has just passed through the UK parliament. The proposed legislation would permit fees to vary by university and by course. In the current chapter, we provide a brief review of the literature on this and present new estimates on log-wage premia by subject studied. Our data do not enable us to estimate *ceteris paribus* variations in returns by institution of study. On this issue, see Chevalier and Conlon (2003).

The rest of the chapter is organised as follows. Following a brief survey in section 4.2 of recent evidence on returns to HE in the UK, section 4.3 provides a description of the data set and the sample selection procedure used in our analysis. In section 4.4, we discuss the issue of the endogeneity of educational qualifications and describe a way of addressing it, the so-called *proxying and matching method.* In section 4.5, we present a replication study of Blundell *et al.* (2000) comparing results for the 1958 and the 1970 birth cohorts using the *proxying and matching method.* Section 4.6 reports estimates of the return to a first degree based on our most preferred specifications for the 1970 cohort and provides evidence on variations in returns according to degree class awarded and area of subject studied. Section 4.7 describes an alternative way of addressing the issue of endogeneity, using the so-called control function approach, while section 4.8 explores the possibility of heterogeneity in the returns to a first degree according to observed characteristics in a heterogeneous treatment framework. Finally, section 4.9 summarises the main findings and concludes.

4.2 Evidence on the returns to a degree in the UK

One of the most influential analyses estimating the returns to a degree in Britain is that of Blundell *et al.* (2000). This study used data from the National Child Development Study (NCDS), an ongoing survey of all individuals born in Britain in the week 3^{rd}-9^{th} March 1958. The data are particularly rich in information useful in the estimation of returns to education, such as ability measures, educational qualifications and family background characteristics. The approach adopted in Blundell *et al.* (2000) is to compare individuals with HE qualifications with those individuals who did not go on to HE but whose secondary school qualifications (A-levels) would have permitted them admission to HE. Hence, the estimated returns are to be interpreted as conditional on having performed well at secondary school.

As is well known, the estimation of educational returns is potentially susceptible to problems of endogeneity bias arising from the fact that typically unobserved - and hence omitted - characteristics (such as ability) affecting educational outcomes are also correlated with subsequent labour market outcomes. Blundell *et al.* (2000) exploit the fact that the NCDS data are rich in information on typically unobserved characteristics and include these as regressors in their log-wage equations. This *proxy and matching* approach assumes that HE decisions are made on the basis of (i) observed and included characteristics and/or (ii) unobserved characteristics which are well proxied by the included observed variables.[2] If these assumptions are valid, then the wage returns estimated using OLS are consistent. We discuss these issues in more detail in section 4.4 of the chapter.

Blundell *et al.* (2000) estimate the impact of different levels of HE on gross hourly wages at age 33. They estimate the raw returns to a first degree[3] to be 21% for men and 39% for women, relative to the control group of cohort

[2]Wooldridge (2002) discusses two assumptions which make a proxy variable a perfect proxy variable. In our case, the first is that the proxy variable must be uncorrelated with the error term of the 'true wage equation', i.e. the model including the unobserved characteristics, but highly correlated with the unobserved characteristics. The second condition is that when running a regression of the unobserved variable on the proxy variable and the other control variables included in the wage regression, the coefficients on the latter converge to zero.

[3]Heckman *et al.* (2003) stress that in estimating rates of return it is necessary to take account of, among other factors, the direct and indirect costs of schooling, taxes, and the length of working life. In what follows, we often use the term wage 'return' although it should be interpreted in the narrow sense of a log-wage premium.

members with A-level qualifications but without HE. When the full set of controls is included in the estimation, the estimated returns to a first degree fall substantially in the case of men - to only 12% - and only slightly in the case of women - to 34%. Without controls for ability at age 16 or A-level score, the estimated returns are 17% for men and 37% for women. The ceteris paribus returns to higher degrees (such as Master and Doctoral degrees) are estimated to be 8% for men and 32% for women, relative to those with just A-levels, when all controls are included.

Blundell *et al.* (2000) also report different returns estimated for different courses studied at HE, finding that returns for men tend be relatively low in Biology, Chemistry, Environmental Sciences, and Geography and for women tend to be relatively high in Education, Economics, Accountancy and Law and in 'Other social sciences'. As the authors acknowledge, splitting the NCDS graduate sample by subject leads to relatively small cell sizes and hence produces poor precision in the estimates at the subject level.

As Blundell *et al.* (2000) also acknowledge, analysis of wage returns to an undergraduate degree based on the NCDS cohort refers to individuals who, typically, were making HE decisions in the late 1970s and graduating around 1980. As we noted earlier, there have been substantial changes in the HE sector and in the graduate labour market in the last two decades and hence it is interesting to compare and contrast the results obtained for the 1958 birth cohort with results based on the analysis of the more recent 1970 birth cohort. This forms a principal focus of the current chapter. We are also particularly interested in how returns to a degree might vary with the class of degree awarded and with the HE subject studied.

There have been a number of studies using a variety of data sources in order to estimate the private return to a university first degree in the UK. Dearden (1999b), also using NCDS, reports an estimated wage return to a degree of 17% for men and of 32% for women, based on OLS, and also finds that the conventional OLS estimates are reasonable approximations of the true causal impact of higher education on wages. Harkness and Machin (1999) examine changes in wage premia to education in the UK between 1974 and 1995 using data from the General Household Survey (GHS). They report time-varying estimates of the wage premium associated with various educational qualifications. For the period 1979-81, the estimated wage premia to a first degree, relative to A-level qualifications, is 14% for men and 21% for women. By the period 1993-95, these estimated premia have risen to 20% and 26%, respectively. Harkness and Machin (1999) conclude that despite a rise in the relative supply of workers who have a degree in the UK, the fact that the return to a degree was rising in the 1980s and 1990s suggests that relative demand - for example induced by SBTC - rose faster than relative supply. Walker and Zhu (2001), using Labour Force Survey (LFS) data from 1993-2000, estimate the average return to a degree over A-level to be approximately 25% for men and 30% for women. The return to a first degree was 20% for men in 1993 and about 26% in 2000, while for women it was 33% in 1993 and about 25% in 2000. These figures suggest, therefore, an increase over time in the return to HE for men and a decrease for women.

The differences in the estimates from different studies referring to the same period often stem from the specification adopted which in turn depends on the nature of the data used. Longitudinal studies, such as those based on the NCDS or BCS70, are rich in information on family background, ability-related and past educational variables, which are important to address the issue of ability bias and whose inclusion often results in reducing the estimated return to education (see Card, 1999, and Blundell *et al.*, 2003, among others). For the same reason, the studies using other data sources, where these variables are not available (such as the LFS) estimate higher returns. Moreover, Heckman *et al.* (2003) discussing the differences between cross-sectional and cohort-based estimates of the return to education, suggest that the latter should be used when the purpose is to estimate historical returns and make comparisons over time, since cohort changes are likely to affect the cross-section estimates slowly as more and more individuals from the new cohorts enter the labour market.

In addition to the study of Blundell *et al.* (2000), a number of other studies have also investigated the extent to which returns to a university degree vary by subject studied. Because of problems of small cell size, most studies consider broad subject groups. Lissenburgh and Bryson (1996) using the Youth Cohort Study estimate returns of 9% for Science relative to Arts and Social Sciences for both males and females combined. Harkness and Machin (1999) find that for men Social Sciences always give the highest wage premium with respect to A-level (24.6% in 1995) while Science ensured the highest premium for women (44.8%).[4] It should be observed that while male

[4]Including controls for age, age squared, dummies for degree subject, teacher status,

graduates generally did not have statistically significant wage premia from UG degrees in Arts, female Arts graduates earned significantly higher wages in all years considered, especially in 1995 when the wage of Arts graduates was higher than that of Social Sciences graduates (with premia of 31% and 23.4%, respectively).

Walker and Zhu (2001) use quite disaggregated definition of subjects (13 in total), but based on their disaggregated estimates, for males (females) in 1999 the returns with respect to A-level are 19% (41.6%), 23.7% (45.8%) and 4.3% (20.8%) for Science, Social Sciences and Art and Humanities, respectively.[5] Therefore, both males and females appear to obtain higher returns for Social Sciences degrees. Moreover, women have higher returns than men in all degree subjects, and in particular in Arts and Humanities. Neither Harkness and Machin (1999) nor Walker and Zhu (2001) control for family background variables, and this may have inflated their estimates of the return to undergraduate degrees.

One of the problems facing estimates of returns to degrees by subject studied - where degree subject information even exists in the data - is that of small cell sizes. This problem is overcome when data from graduate surveys are used. There are two sources of such data. First, there are follow-up surveys of samples of graduates from the graduating cohorts of 1960, 1970,

region and industry.

[5]Science includes Health, Nursing, Science, Maths, Engineering, Arthitecture. Social Sciences includes Economics, Law and Social Studies. Art and Humanities includes Language, Education, Art and Combined degrees. Their specification includes controls for age, age squared and dummies for marital status, race, union status and region.

1980, 1985, 1990 and 1995. Each of the cohort samples generates, on average, several thousands of observations on graduates and their early labour market experiences, including earnings. With these data one can estimate earnings premia by degree subject studied. A second source of graduate cohort data comes from matching administrative data on the entire population of UK university students - as collected formerly by Universities' Statistical Record (USR) and now by the Higher Education Statistics Agency (HESA) - to the information contained in the responses to the first destination survey of all graduates. This matched data-set is very rich in information on graduates' education (including university, course, degree class, pre-university schooling and attainment), personal characteristics and family background. The post-university outcome provides information on the graduate's occupation six months after graduation.

Using the follow-up surveys of samples of graduates, Dolton and Makepeace (1990) and Dolton *et al.* (1990) analyse earnings data from the 1986 survey of one in six of the 1980 UK university graduates. Unlike the HESA/USR data, the information on degree, other qualifications and family background do not come from administrative data but from personal recall. The 1986 survey generated earnings information for 5,002 graduates. Dolton *et al.* (1990) find significant earnings premia for Science and Social Sciences students compared to Humanities or Education students. A positive wage premium for Mathematics-related degree courses is a common finding in studies using the graduate sample follow-ups: see Chevalier *et al.* (2002), Belfield *et al.* (1997), and Battu *et al.* (1999) for results pertaining to the 1996 follow-ups of the 1985 and 1990 graduate cohorts. Chevalier *et al.* (2002) analyse 1998 earnings data for a sample of 8,264 graduates from the 1995 graduate cohort. They report that relative returns are highest for Mathematics (at 29% for men and 19% for women), compared to the field of Education studies. They make the important point that differences in relative returns across cohorts are to be interpreted with care given differences across cohorts in the method of classifying degree subjects. Chevalier *et al.* (2002) provide a comprehensive survey of estimates of returns to HE.

With respect to differences in returns to a degree according to the degree class obtained, Battu *et al.* (1999), using graduate cohort data, estimate a significant log-wage premium associated with a first class over a second class degree. Naylor *et al.* (2003), using HESA-USR data for 39,454 individuals responding to the first destination survey of all 1993 graduates, estimate an occupational earnings premium of 3.9% (3.6%) for a first class degree relative to an upper second class degree for men (women). The premium for a first over a third class degree is estimated to be 13.8% for men and 8.9% for women. Replicating the analysis on earlier graduate cohorts, Naylor *et al.* (2003) find that the premium for a first class degree has been growing over time. It is noticeable that while for the earlier cohorts there was no statistically significant premium associated with the class of degree awarded, a significant differential has developed and grown over time. One hypothesis for this is that as the population of graduates has grown, greater importance is attached by employers to the signal emitted by a graduate who has performed well at university. For a more formal treatment of this hypothesis, see Naylor and Smith (2004).

One focus of the current chapter is to test for corroborating evidence on the extent of any degree class premium from a different data source. Using BCS70, our attention focuses on a cohort of young people who, typically, would have been graduating in the very early 1990s - the period of time for which Naylor *et al.* (2003) estimate significant relative premia for a 'good' degree performance.

Finally, we note that there has been some work on the extent to which returns to degrees vary by institution attended. From first destination survey data, Naylor *et al.* (2003) estimate statistically significant differences, ceteris paribus, in occupational earnings across universities. Chevalier and Conlon (2003), using data from the follow-up surveys of samples of graduates for 1985, 1990 and 1995, conclude that graduating from one of the more highly regarded UK universities (that is, a Russell Group institution) is associated with a wage premium of up to 6% for men, compared to the default case of having graduated from a new university. From this estimate, Chevalier and Conlon (2003) infer justification for a policy of differentiating fees by institution.

4.3 Data and sample selection

In this chapter we use data drawn from the BCS70. The BCS70 began in 1970 when data were collected on the births and families of 17,198 babies born in England, Wales, Scotland and Northern Ireland between the 5^{th} and the 11^{th} of April of that year. There are currently five complete follow-up surveys available: at 5, 10, 16, 26 and 30 years after the original survey. In this chapter we use data on gross hourly wages collected in the 30-year follow-up survey, while family background and individual characteristics come from the 10-year follow-up survey. Based on the sample of respondents to the 30-year follow-up survey (11,261 individuals), and in analogy with Blundell *et al.* (2000, p. F84), we select only individuals who have obtained at least one A-level, which is our population of interest, and analyse the return to HE qualifications with respect to those individuals who did not complete any form of HE. As observed by Blundell *et al.* (2000), individuals with at least one A-level but who did not continue into HE are probably a better comparison group for students undertaking undergraduate degree courses than for those enrolling in courses leading to a non-degree HE qualification, since the latter group usually has non-traditional entry qualifications (i.e. different from the standard A-level qualification). The same can be said for individuals undertaking postgraduate studies, as they must possess a first degree, and for whom individuals with A-levels only are not a very good comparison group. This should be kept in mind when interpreting our empirical results on, for example, postgraduate students. For these reasons in the chapter we focus on the return to undergraduate degrees only.

From our sample of individuals who have at least one A-level (a total of 2,755 cases), we primarily focus on those who also replied to the 10-year follow-up survey (2,553). This is done since in the estimation of the log-wage regressions we will include individual and family background control variables which are provided by the 10-year follow-up. Among this sub-set we choose individuals who worked as employees full-time or part-time (2,092 individuals) and exclude those individuals with missing data on their gross hourly wage (114 cases dropped). In order to maintain the sample size, and in analogy with Blundell *et al.* (2000), individuals with missing values for other variables were kept in the data set and missing value dummy variables were included in the regressions. The final sample includes 1,978 individuals.

Since our analysis examines a sample derived from matching the 10-year and 30-year follow-up surveys of the BCS70, it is useful to show how our sample compares with respect to the 30-year follow-up survey. In Table 4.1 we report the distribution of educational qualifications and the sample mean of gross hourly wages in natural logarithm for various samples. From Table 4.1 the number of matches between the 30-year and the 10-year follow-up surveys is higher than that between the 30-year and the 16-year follow-up surveys by about 8.5 percentage points. Together with a higher incidence of item-non response in the 16-year wave, this is the main reason for our use of family background variables at age 10, along with the availability of an indicator of 'innate' or 'early' ability (the British Ability Scale score, see Elliot *et al.*, 1979) at this age. As can be seen from Table 4.1 the distribution of educational qualifications is very similar across samples. This evidence suggests that survey non-response and panel attrition, although of non-negligible size, might be random with respect to educational qualifications. In relation to wages, means are very similar across samples. In particular, the average of the natural logarithm of gross hourly wages is 2.270 in the 30-year follow-up, 2.267 in the 30-year and 10-year matched sample, 2.267 in the 30-year and 16-year matched sample and 2.265 in the sample of individuals matched in all three waves.

4.4 OLS, endogeneity and the proxying and matching method

When we estimate the returns to different educational qualifications, we consider the effect of a multiple treatment, namely educational qualifications, denoted as $j = 1, ..J$, on individual wages, w_i . We consider four different educational qualifications: A-level only ($j = 1$, the reference group), Non-degree Higher Education ($j = 2$), undergraduate (UG, or first) degrees ($j = 3$) and postgraduate (PG, or higher) degrees ($j = 4$). If we define as w_i the gross hourly wage of individual i, our model can then be written as follows:

$$\ln w_i = mX_i + \sum_{j=2}^{J} b_j Q_{ij} + u_i. \tag{4.1}$$

where mX_i is a linear function of the observed variables X_i, which we will refer to as the no-treatment outcome, Q_{ij} are dichotomous variables assuming value 1 if individual i has as her/his highest educational qualification a qualification j and 0 otherwise, and the b_j's are the effects of these educational qualifications on log-wages, i.e. they are our parameters of interest. We abstract for the moment from problems concerning the correct specification of the no-treatment outcome and assume that a linear function is an appropriate representation of the log-wage data generating process, as is the usual assumption in most of the existing empirical literature on the returns to education. In the case $E(u_i|X_i, Q_{ij}) = 0$, the b_j parameters can be estimated without bias using ordinary least squares (OLS, hereafter). Assuming no heterogeneity in the returns to education, the Average Treatment on the Treated (ATT), the Average Treatment on the Non-Treated (ATNT), and the Average Treatment Effect (ATE) all coincide and are recovered by the b_j's.

However, there are several reasons why we may expect a non-zero correlation between educational qualifications and the error term in the log-wage equation. These are outlined, for instance, in Blundell *et al.* (2003) and include:

1. *Ability bias (absolute advantage).* We might assume that the error term u_i in equation (4.1) consists of two components, i.e. $u_i = \alpha_i + \epsilon_i$, one reflecting unobserved earnings capacity (α_i), with $E(\alpha_i|X_i, Q_{ij}) \neq 0$ and the other some unobserved factors uncorrelated with all covariates included in the wage regression $E(\epsilon_i|X_i, Q_{ij}) = 0$. It is the non-zero correlation between unobserved earnings capacity (also referred to in the literature as ability) and education which causes the so-called 'ability bias'. In particular, we may expect high ability individuals both to acquire more education and to earn higher wages. Earnings capacity, is observed by the individual but not by the analyst;

2. *Return bias (comparative advantage).* The returns to the different educational qualifications may not be the homogeneous across individuals.

In this case, we will have a distribution of b_{ij}'s, where j is the subscript for the educational qualification and i that for the individual. There is a return bias when $E(b_{ij}|X_i, Q_{ij} = 1) \neq 0$, i.e. individuals self-select into the different educational qualifications according to their idiosyncratic returns, which depend in turn on characteristics that are observable to the individual but not to the researcher;

3. *Measurement error bias.* The educational variables may be measured with error. In our case, where education is a categorical variable, measurement error is non-classical and in general it is not possible to say anything on the direction and magnitude of the bias (see Kane *et al.*, 1999).

In our analysis in the current chapter, we focus only on the first source of bias, i.e. ability bias, and assume the absence of a return bias. Although heterogeneous returns according to unobserved characteristics may exist, there is a return bias only if individuals are able to correctly predict their idyosioncratic gains in the return distribution, that is they know b_{ij}, and use this information to choose their level or type of educational qualification, which is a strong assumption. In this regard, there is an interesting stream of literature on students' income expectations. Betts (1996) using US data finds that students can predict their starting salaries quite well and better than life-time earnings profiles and tend to underestimate wages in fields outside their own. He also finds that the most widely used source of information for wages are newspapers and magazines, which would suggest a substantial homogeneity in income expectations. Dominitz and Manski (1996) find using

US data that students are very uncertain about their own future earnings, both at ages 30 and 40 and tend to be more uncertain about their earnings with a university degree than about earnings with only secondary school. The authors also find substantial heterogeneity in students' beliefs about the actual earnings distribution. Wolter and Zbinden (2002) use Swiss data and find that students' expectations are much closer to actual wages at the time of graduation while their prediction errors are higher when considering the pattern of wage increase during the first 10 years of their careers.

Therefore, most studies show that individuals are able to predict more accurately their starting wages, while their predictions are much less precise for earnings later on in the life-cycle, which we consider in this chapter since individuals from the BCS70 with a first degree typically have in 2000 about 9 years of labour market experience. Blundell *et al.* (2003) using NCDS data find the absence of both an ability and a return bias when interactions between educational qualifications and individuals' observed characteristics are included in the log-wage equation estimated through OLS. Unfortunately, they are able to provide this evidence only in the case of a single treatment model (first degree vs. other educational qualification) since the size of their sample, which is very similar to ours, does not allow the exploration of heterogeneity of returns with respect to observed characteristics when considering a multi-treatment model in a regression framework.[6] Finally, we think that the third source of bias should be less severe when including educational

[6]We tried to interact the educational qualifications with social class, however probably due to small cell size the interactions generally turned out not to be statistically significant.

qualifications, as we do, rather than the number of years of schooling.[7]

A possible approach to tackle endogeneity issues when the data set is particularly rich, as in our case, is the so-called *proxy and matching method.* This is the approach followed in Blundell *et al.* (2000) and in the replication of their analysis on BCS70 data reported in section 4.5. This consists of including among the individual characteristics X_i factors which might affect both the educational qualification achieved and wages, and by proxying the unobserved component α_i with observed factors highly correlated with it, so that $u_i = \epsilon_i$. As observed in Blundell *et al.* (2000) equation (4.1) can be viewed as a form of regression-based linear matching. Thus, the estimates presented in sections 4.5 and 4.6 can be argued to have been obtained using a method which addresses the issue of ability bias.

4.5 Comparison with the 1958 British Cohort

In this section we compare BCS70 and NCDS data, the latter referring to the 1958 British cohort - analysed in Blundell *et al.* (2000). Comparing the first column of Table 4.1 - with the descriptive statistics reported in Blundell *et al.* (2000, p. F86)[8] we observe a reduction in both the proportion of people with A-levels not completing any form of HE, and in the proportion of individuals

[7]For the simple fact that recall errors should be only minor for the highest educational qualification achieved by age 30.

[8]The percentage of men (women) completing non-degree HE, a first degree or a higher degree are 21% (25%), 38% (33%) and 17% (14%), respectively, in the sub-sample of the NCDS considered by Blundell *et al.* (2000).

taking non-degree HE qualifications. We also note a large increase in the proportion of individuals completing first degrees, and a slight increase in the proportion of people with postgraduate degrees. These figures are consistent with the widening access to HE that took place in the UK during the 1980s, when graduates of the 1970 British cohort typically entered HE. Our data show a particularly marked increase in the supply of female graduates with respect to the 1958 cohort. On the basis of the observed increase in the number of graduates, we would expect a reduction in the return to a first degree, *ceteris paribus*. However, demand-side forces, working, for example, through skill-biased technological change (SBTC) might have counteracted this tendency and brought about an increase in the return to a first degree.

In order to gauge the extent of any change in the return to a first degree between the 1958 and the 1970 cohorts, we attempt to replicate the analysis in Blundell *et al.* (2000). We have included the same kind of explanatory variables and used the same classification of educational qualifications as those underlying the log-wage equations estimated in Blundell *et al.* (2000),[9] in order to ensure as high a degree of comparability as possible. Inevitably, however, there are some coding and other data differences across the two surveyed cohorts. For example, it should be noted that the wage data in the BCS70 refers to age 30, while Blundell *et al.*'s (2000) analysis of NCDS refers

[9] PG degree: all higher degree qualifications; UG degree: first degree; Non-degree HE: non-degree NVQ Level 4, HNC/HND, BEC/TEC higher, university diploma or certificate, professional or nursing qualifications, HE diploma or certificate, C&G full technology certificate or insignia award in technology; A-level: A-level qualification, Scottish Higher or Scottish Six form college.

to age 33.

Following Blundell *et al.* (2000) we estimate a gross hourly wage regression in natural logarithms using three specifications:[10]

1. including only educational qualification dummies;

2. including educational qualification dummies, plus British Ability Scales score in the non-verbal and verbal questions separately, regional dummies and school type, all at age 10;

3. including all the variables in the previous specification, plus family background variables at age 10 (parental education, parental social class, number of older siblings, number of younger siblings, house ownership), socio-demographic variables (% of children's fathers in social class I and % of children's fathers in social classes IV-V in the child's school at age 10), school attendance at age 10 (number of missed days of school), employer characteristics in 2000 (firm size, union membership dummy, private sector dummy).

The estimates from the three specifications conducted on the BCS70 are reported in Tables 4.2 and 4.3, for men and women respectively, alongside the equivalent estimates for the NCDS 1958 cohort reported in Blundell *et al.* (2000), indicated with 1', 2' and 3', respectively. From specification 1, males with a first degree earn significantly more than those with just A-levels: the

[10]We are not able to consider specification 4 in Blundell *et al.* (2000, p. F91, Box 1) since we cannot compute the UCAS score, as A-level grades are not available in the BCS70 data set.

wage return is 0.20. This is remarkably close to the point estimate of 0.21 for the 1958 cohort. In general, the returns to the other types of HE appear to have decreased over time. For females, the coefficients of all educational dummies are statistically significant. In specification 1, the estimated wage return associated with an undergraduate degree is 0.27, compared to an estimate of 0.39 for the 1958 cohort.

Typically, adding control variables leads to reductions in the estimated returns to the different HE qualifications. In the most complete specification, specification 3, the return to an undergraduate degree turns out to be 0.15 for males and 0.23 for females. The equivalent figures from the corresponding specification in Blundell *et al.* (2000, p. F90) are 0.17 for men and 0.37 for women.

It appears, then, that while the private return to a first degree for men is remarkably similar in the two British cohorts (1958 and 1970), there is a striking fall in the return to a first degree for women in the 1970 cohort. For men, the results are consistent with the idea that the opposing supply-side and demand-side pressures on returns to a first degree have approximately balanced each other. We note that the 1970 birth cohort would, typically, have graduated in the early 1990s and that their earnings are observed in 1999/2000. The 1958 cohort would have graduated around 1980 and their earnings were observed in 1991. Hence, our results call in to question the findings based on LFS and GHS data of a rise in the return to a degree for males during the 1980s and 1990s. Our estimate of a log-wage premium of 0.20 for males, based on specification 1, is very similar to those reported in our survey of UK evidence in section 2. However, the fact that when we introduce other control variables - not typically observed in other datasets - in specification 3 the estimate falls to 0.15 suggests that other studies have tended to over-estimate the return to a degree for men in the UK.

For women, our finding that the estimated return to an undergraduate degree has fallen between the two cohorts is consistent with results reported in Walker and Zhu (2001). There are various interpretations which one could put on this. First, we note that the very high estimated returns for female graduates in the earlier (NCDS) cohort could be interpreted as arising in part from greater gender pay discrimination at lower education levels. Some evidence in this direction is provided by Makepeace *et al.* (1999), who analysed the 1948 and the 1958 British cohorts and found a positive effect of the Equal Pay Legislation in reducing gender pay gaps and that the reduction was relatively higher at the bottom quantiles of the earnings distribution. Hence, the legislation was particularly effective in reducing the gender pay differential for low paid women, who were also likely to be low educated women. If this tendence has continued in recent years, then this would have contributed to reducing the estimated return to education also for women in the 1970 cohort. Second, we note that the expansion in the number of graduates between the two cohorts was a particularly female phenomenon. Thus, to the extent that there was a particular shortage of female graduates among the 1958 birth cohort, this was likely to have been less true for the 1970 cohort. Hence, the supply-side shift in the number of graduates was much greater for females than for males. To the extent that the graduate labour market is 'gender-neutral', this should not have implications for the magnitude of any gender difference in returns to a degree. However, to the extent that male and female graduates are either not perfect substitutes or do not have equal preferences, then it is likely that the relative increase in the supply of female graduates will have been associated with the observed reduction in the gender difference in the return to an undergraduate degree.

The supply-side argument is sometimes presented as associated with a form of over-education: see, for example, Dolton and Vignoles (2000), Hartog (2000) and Chevalier (2003). Similar findings of an underutilisation of graduates' skills in a context of expanding student numbers are reported in Rigg *et al.* (1990), Mason (1995) and Green *et al.* (1999). We note however that the very similar estimated return for men in the 1958 and the 1970 birth cohorts, runs counter to a strong version of the over- education hypothesis.

4.6 Preferred estimates using the proxying and matching method

4.6.1 Return to a first degree

We have already said that the application of the *proxying and matching method* requires the availability and inclusion among the X_i's of a wide set of individual characteristics affecting education and wages. In section 4.5, our choice of specification was dictated by our attempt to replicate the analysis of Blundell *et al.* (2000). In this section, we present estimates of the return to a degree based on our most preferred specifications for the BCS70 dataset, under a proxying and matching approach.

In particular, we include among the X_i's in the starting specification:

1. personal characteristics: region of residence at age 10, ethnicity.

2. family background variables: father's education and social class, mother's education and social class, presence of the mother, presence of the father, home ownership, family income, number of younger siblings, number of elder siblings, parental interest in child's education, all at age 10.

3. age 10 school variables: school type at age 10, school attendance at age 10, % of children whose father is in social class I (professionals)and % of children whose father is in social classes IV-V (partly skilled and unskilled, respectively) in the school attended at age 10.

4. ability at age 10: score in the verbal and non-verbal sections of the British Ability Scales questionnaire, as proxies for verbal and quantitative innate (or early) ability.

Therefore, we included most of the variables already included in the replication reported in the previous section but unlike Blundell *et al.* (2000) we did not include employer characteristics for two main reasons. Firstly, they may be endogenous, i.e. choice variables for the individual and jointly determined with wages. Secondly, employer characteristics may be affected by educational qualifications, and by excluding them we estimate the 'overall' effect of education, both on wages and on the likelihood of working for certain types of employers (see for instance Blundell *et al.*, 2003, and Pereira and Martins, 2004).

We firstly estimated the model including all factors 1-4 above in the sample of matched individuals from the 30-year and the 10-year follow-up surveys and performed some F-tests for the significance of the various groups of variables. After selecting a 'parsimonious' specification (specification I), by keeping the groups of regressors statistically significant or only marginally not significant at the 10% level, we also estimated additional specifications which include past educational variables available from the 30-year follow-up survey.[11] Specifically, specification II adds:

- S (Supplementary), A (Advanced) and AS (Advanced Supplementary) level information (i.e. age 18 educational variables): including the number of S-, A- and AS-levels in various categories (A-C and D-E for S- and A-levels and A-C and D-G for AS-levels), obtained from the 30-year follow-up survey.

- O (Ordinary) levels, CSE's (Certificate of Secondary Education), GCSE's (General Certificate of Secondary Education) information (i.e. age 16 educational variables): including the number of O-levels, CSE's and GCSE's levels divided by grades (A-C, D-E for O-levels and GCSE's and 1, 2-5 for CSE's), obtained from the 30-year follow-up survey.

[11]The results of the F-tests are available upon request.

Including past educational variables may be important, since secondary school educational performance is likely to affect both individuals' demand for HE, i.e. the selection into educational qualifications, and wages. However, these variables, being based on retrospective questions, may be affected by a recall bias.

All results are shown in Table 4.4. In the parsimonious specification the wage return to a UG degree is about 0.19 for men. Adding past educational variables (specification II) reduces the return to 0.14. For men, the estimate of 0.19 is slightly higher that that reported in Table 2 for specification 3. However, as we have commented, specification 3 controls for employer characteristics whose effect is generally to reduce the coefficient on the educational qualifications. For women, the parsimonious specification, specification I, produces an estimate of the return to a degree of 0.23 - the same as that reported under specification 3 in Table 2. Thus, for women, there is little impact from the inclusion of employer characteristics in the estimated equation. Specification II, however, produces an estimated return of 0.18. We conclude from our most parsimonious specification that the return for an undergraduate degree in the UK for the 1970 birth cohort was 0.19 for men and 0.23 for women. The inclusion of additional secondary education variables causes these estimates to fall to 0.14 and 0.18 for men and women, respectively.

Thus, in general, including past educational variables reduces the estimated return to a first degree. This comes as little surprise since individuals are likely to decide whether to continue in HE in part on the basis of their past educational performance. Moreover, HE institutions also select potential students on the basis of their pre-university educational performance. In both cases, therefore, we expect a positive correlation between secondary school performance variables and the highest educational qualification achieved, and accordingly a reduction in the return to a first degree. A surprising thing is perhaps the fact that both HE dummies and past educational variables are statistically significant, representing therefore distinct sources of wage variation. However, this may be an artifact of the specific functional form chosen for the no-treatment outcome, in particular the constraint that the X_i's have the same effect on the outcome (log-wage) for both the treated and the non-treated.[12] We will return to this issue in section 4.8. It must be noted that, as with the HE qualifications dummies, past educational qualifications may be affected by the same problems of endogeneity, i.e. correlation with individuals' unobserved earnings capacity, but this is the main reason why they are included in the log-wage regression following the *proxy and matching* approach.

4.6.2 Differences by degree class

In the previous section we have considered UG degree education as a homogeneous good. However, students may be more or less successful in completing their UG studies. In particular, previous work has shown the positive effect of a 'good' degree performance on graduates' earnings. Battu *et al.* (1999)

[12] For instance the effect of past educational variables on wages may be significant only for individuals with only A-level qualifications.

using data on two cohorts of graduates (from 1985 and 1990) found a positive log-wage premium of a first class honours degree with respect to upper and lower second class degrees (for females in 1985 and 1990 and males in 1990) on graduates' earnings one year after graduation. Furthermore, the degree premium associated with a 'first' turned out to be always significant six years after graduation, for both males and females and for both cohorts of graduates.[13]

Similar evidence of an important role of degree class is also obtained by Naylor *et al.* (2003), using USR data for several cohorts of university leavers (from 1985/6 to 1993/4). The authors found, for instance, a significant positive premium, growing over time, associated with a first class degree on first destination occupational earnings of UK graduates. Neither the Battu *et al.* (1999) nor Naylor *et al.* (2003) papers are able to address the issue of returns to degrees relative to non-graduate outcomes as these studies are based on graduate data only, with no control group of non-graduates.

The BCS70 provides degree class for UG degrees, and here we are able to investigate differences in the return to an UG degree according to the class of degree awarded. In particular, in order to avoid small cell size problems we consider only two broad degree classes: 'good' degrees (first class or upper second class honour degrees) and 'lower' degree classes. This distinction is also suggested by the common practice of some employers of conditioning job offers on the attainment of a 'good' degree result.

[13]Battu *et al.* (1999) control for several individual characteristics which include the change of region, degree class, firm size, being self-employed and others.

The estimation results are shown in Table 4.5.[14] We base the analysis on specifications I and II from the previous section of the chapter, replacing the dummy variable for obtaining an undergraduate degree with two dummy variables: one for obtaining a 'good' degree class and one for obtaining a lower degree class. For males, compared to an average return to an undergraduate degree of 0.19 (see Table 4.4) the estimated returns to a 'good' degree and lower degree classifications are 0.24 and 0.15, respectively - when controls for past educational qualifications are not included - a difference which is marginally significant at the 10% level. Adding past educational control variables again reduces the return to a 'good' degree to 0.19 and to 0.11 for a lower degree classification. An F-test for the equality of the return to different degree classes cannot now be rejected at conventional significance levels. Hence, our point estimates show a substantial distance between the log-wage premia to 'good' and to lower degree classifications, although the effects are not very precisely estimated, probably due to small cell sizes. Again, the change in the point estimates when including secondary school variables is in the direction we would have expected given the positive correlation between secondary school performance and performance in HE.

For females, compared to an average return to a first degree of 0.23 (see Table 4.4) the returns to a 'good' degree and a lower degree class are 0.26 and 0.18, respectively, when past educational variables are not included - a difference which is significant at the 5% level. Inclusion of past educational variables reduces the size of the wage return to 'good' and a lower degree

[14]The sample size falls to 957 for men since 4 individuals did not report degree class.

class to 0.21 and 0.14, a difference which remains significant at the 10% level. Thus, our evidence supports findings by Battu *et al.* (1999) and Naylor *et al.* (2003) of variation in the returns to degrees according to the graduate's level of academic performance at university.

4.6.3 Differences by degree subject

In this section, we consider another possible source of heterogeneity in the return to UG degrees: the degree subject studied. Here, we are only able to consider some broad aggregations of subjects studied because of the size of our sample. We focus on the following aggregation of subjects: Sciences (includes: Medicine and Dentistry, Subjects Allied to Medicine, Biological Sciences, Agriculture, Physical Sciences, Mathematical Sciences, Computing, Engineering, Technology and Architecture), Social Sciences (includes Social Studies, Law and Politics, Business and Mass Communications), Art and Humanities (includes Classics and Literature, Modern European Languages, Other Languages, Creative Arts, Education and Other)

Table 4.6 shows our estimates.[15] For men our estimated returns for the different subjects are not very different from those of Walker and Zhu (2001). Compared to an average return to a first degree of 0.19 estimated in Table 4.4, Social Sciences have the highest wage return (0.26), and Art and Humanities the lowest wage return (0.12), which is not statistically different from zero at the 5% level. Adding controls for past educational variables reduces the

[15]The sample size falls to 930 for men and 996 for women since we dropped the individuals who did not report degree subject.

estimated return to a first degree irrespective of subjects. However, their relative positions in terms of return do not change. F-tests for the equality of returns across all degree subjects cannot be rejected at the conventional statistical levels of significance. However, when considering Social Sciences vs Art and Humanities, the difference is statistically significant at the 10% level.

For women, we observe the same ordering of subjects as for men, although the spread of the estimates around the average return of 0.23 estimated in Table 4.4 is much tighter, with Social Sciences having the highest wage return (0.25), and Art and Humanities the lowest return (0.18). When including secondary school variables the relative order of subjects remains unchanged, however the fall in the estimated returns is generally bigger than in the case of men. In neither case are the subject returns estimated precisely enough to make the differences statistically significant. However, as for males, the return to Social Sciences seems to be some distance apart from that of the other two subject groups. It is interesting to observe that unlike for men, women with a first degree in Art and Humanities earn significantly more than those with A-levels only, findings that is in line with Harkness and Machin (1999) and seems to suggest that women might self-select into Art and Humanities since they have comparative advantages with respect to men.

4.7 An alternative approach to endogeneity: the control function approach

In the previous sections, we addressed the issue of potential endogeneity bias in estimating the return to a degree and used an approach based on a *proxying and matching* method. The idea underlying the *proxying and matching* method is to proxy the unobserved characteristics, which may cause endogeneity problems, with highly correlated observed characteristics. Although the BSC70 data set is very rich in information concerning family background, past education and ability-related variables, nothing really prevents the possibility that even after controlling for these characteristics there might still be some other omitted variables responsible for some residual correlation between the educational qualification dummies and the error term in the log-wage equation.

A possible alternative approach to the issue of endogeneity, associated with the selection into educational qualifications through unobservables, is the *control function approach* (CFA hereafter, see Heckman and Robb, 1985). The CFA consists of simultaneously modelling both the process of educational attainment and the process of generating wages. In particular, let us assume that educational qualifications are achieved according to the following process:

$$\begin{aligned} Q_{i1} =& 1(Q_{ij}^* < \mu_1) \\ Q_{i2} =& 1(\mu_1 \leq Q_{ij}^* < \mu_2) \\ Q_{i3} =& 1(\mu_2 \leq Q_{ij}^* < \mu_3) \\ Q_{i4} =& 1(Q_{ij}^* \geq \mu_3) \end{aligned} \tag{4.2}$$

with $Q_{ij}^* = BV_i + \nu_i$, where Q_{ij}^* is a latent educational variable and V_i some covariates affecting educational attainment, with $E(\nu_i|V_i) = 0$. The μ_j's are some parameters (thresholds) to be estimated.

If we assume that:

$$(\nu_i, u_i) \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix})$$

then it is straightforward to show that by including conditional mean terms or 'control functions' in equation (4.1) it is possible to obtain consistent estimates of the b_j's by running OLS on the following augmented regression:

$$\begin{aligned} \ln w_i =& mX_i + \sum_{j=2}^{J} b_j Q_{ij} + \sum_{j=1}^{J} E(u_i|Q_{ij} = 1) + \omega_i \\ =& mX_i + \sum_{j=2}^{J} b_j Q_{ij} + \sum_{j=1}^{J} \rho\sigma\lambda_{ij} Q_{ij} + \omega_i \\ =& mX_i + \sum_{j=2}^{J} b_j Q_{ij} + \rho\sigma \sum_{j=1}^{J} \lambda_{ij} Q_{ij} + \omega_i \end{aligned} \tag{4.3}$$

where

$$E(u_i|Q_{ij}=1)=\begin{cases} E(u_i|\nu_i<c_{1i})=\rho\sigma\frac{-\phi(c_{1i})}{\Phi(c_{1i})}\equiv\rho\sigma\lambda_{1i} & \text{if } Q_{i1}=1\\ E(u_i|c_{1i}\le\nu_i<c_{2i})=\rho\sigma\frac{\phi(c_{1i})-\phi(c_{2i})}{\Phi(c_{2i})-\Phi(c_{1i})}\equiv\rho\sigma\lambda_{2i} & \text{if } Q_{i2}=1\\ E(u_i|c_{2i}\le\nu_i<c_{3i})=\rho\sigma\frac{\phi(c_{2i})-\phi(c_{3i})}{\Phi(c_{3i})-\Phi(c_{2i})}\equiv\rho\sigma\lambda_{3i} & \text{if } Q_{i3}=1\\ E(u_i|\nu_i\ge c_{3i})=\rho\sigma\frac{\phi(c_{3i})}{1-\Phi(c_{3i})}\equiv\rho\sigma\lambda_{4i} & \text{if } Q_{i4}=1\end{cases}$$

and $c_{1i}=\mu_1-BV_i$, $c_{2i}=\mu_2-BV_i$ and $c_{3i}=\mu_3-BV_i$ and $E(\omega_i|X_i,Q_{ij},\lambda_i)=0$.[16]

Given the coefficient restriction on the artificial regressors λ_{ij}'s, estimating equation (4.3) is equivalent to estimating the following:

$$\ln w_i = mX_i+\sum_{j=2}^{J}b_jQ_{ij}+\rho\sigma\lambda_i+\omega_i \tag{4.4}$$

where $\lambda_i=\sum_{j=1}^{J}\lambda_{ij}Q_{ij}$. Hence in equation (4.4) it is sufficient to include only one conditional mean term. The 'generalised residual' λ_i is estimated from an ordered probit model of equation (4.2). Although the model is formally identified by the non-linearity of the education equation, it is usually thought that exclusion restrictions are necessary in order not to rely exclusively on functional form. Therefore, an 'economic' identification requires that at least one variable included in the education equation (i.e., in V_i) is excluded from the log-wage equation (i.e., from X_i).

In the following sections we use the CFA to estimate the log-wage premia to different educational qualifications and to different levels of degree performance.

[16]See also Vella and Verbeek (1999) on how to model and estimate endogenous treatment effects.

We use as identifying variables some variables that were not significant in the log-wage equation estimated with OLS but that previous research has shown to be highly correlated with children's education: parents' educational qualifications (see Ermisch and Francesconi, 2001a, and Chavalier and Lanot, 2002, among the others).[17] In particular, in the light of the fact the father's education turned out to be significant in women's wages, we use father's education for men and mother's education for women following therefore a 'gender role model' approach. Some recent evidence supporting a stronger parents' effect on children's education for same-sex parents is provided by Chevalier (2004), for instance. For both men and women we use a parsimonious specification of the ordered probit model including only age 18 and age 16 educational variables and the indentifying variables (father's highest educational qualification for men and mother's highest educational qualification for women).

The CFA offers a direct test for the endogeneity (or self-selection through unobservables), which can also be interpreted as a specification test in the spirit of Heckman's (1979) seminal paper. In particular, the absence of endogeneity can be tested by testing whether the coefficient of the generalised

[17] However, if education is endogenous the coefficients on the educational dummies as well as those on the other regressors, including the candidate 'instruments', will be affected by the 'endogeneity bias'. For this reason we replicate the test of significance of the instruments on the log-wage equation estimated with the CFA and without exclusion restrictions and report them in Table 4.8. Since the model using the CFA is identified without exclusion restrictions, through the functional form, the endogeneity problem is formally addressed and the coefficients on the instruments are consistent.

residual (i.e. $\rho\sigma$) equals zero. Implicitly, what the test says is whether the omitted variables in the log-wage equation, entering u_i, and in the education equation, entering ν_i, are correlated or not, and therefore whether or not the educational qualifications dummies are correlated with u_i. If they are not correlated the log-wage equation (4.1) should be estimated using OLS, otherwise OLS should be applied on equation (4.4). Thus the t-test on the coefficient of the generalised residual can also be interpreted as a test for the omission of variables correlated with educational qualifications from the log-wage equation, i.e. as a test for omitted variables.

Table 4.7 shows the estimation results from the CFA for both males and females. In the light of the significance of the past educational qualifications on the log-wage shown in tables 4.4 and 4.5, we estimated only specification II, including both age 18 and age 16 educational variables. Table 4.7 reports the results of F-tests for the significance of the identifying variables. Our 'instruments' do not appear to be 'weak':[18] parents' educational qualifications are generally highly significant in the selection equation (the ordered probit) but not in the log-wage equation. In all cases the coefficient of the generalised residual is not significant, showing the absence of an endogeneity problem.

We used the same framework (i.e. two-stage estimation of the log-wage equation with a selection equation estimated through an ordered probit) to test the potential endogeneity of degree class. In particular, the generalised residual was estimated using an ordered probit with five categories (in in-

[18]Such a test is suggested in Bound *et al.* (1995).

creasing order: A-level, Non-degree HE, lower degree class, 'good' degree, PG degree). Table 4.8 reports the results. We used the same identifying variables employed to address the endogeneity of the educational qualifications, and found them to be generally valid, as the F-tests suggest. As we also saw for the educational qualifications, there is no evidence of an endogeneity problem: the coefficient of the generalised residual is never statistically significant.

Unfortunately, our data are not sufficiently rich to permit convincing tests for the potential endogeneity of degree subjects, along the lines suggested by Lee (1983). We would need variables affecting subject choice but not wages. Parents' education appears to be a much less appropriate instrument for the choice of degree subject than for the level of education or degree performance. Moreover, past educational variables, which are likely to affect the type of subject chosen appear to have a significant impact also on wages and therefore are not suitable instruments.

Hence, we cannot exclude the possibility that our estimates of degree subject returns may be affected by a selection bias. However, we do conclude that there is no evidence of ability bias in the estimation of either the return to educational qualifications or to degree class. Individuals do not seem to self-select into educational qualifications according to their earnings capacity; in other words, individuals who expect to earn more do not necessarily acquire more education or have a better degree performance.

4.8 The case of heterogeneous returns

Our previous analysis suggests the absence of an ability bias both in the estimation of the returns to a degree and in that for returns by class of degree. However, as in the case of selection exclusively on observables, OLS estimates will recover the unbiased ATT only if the no-treatment outcome has been correctly specified (i.e. the model is correctly specified in terms of the covariates included and the functional form chosen) and if the treatment effect is homogeneous across individuals with different observed characteristics (i.e. treatment has only an intercept and not a slope effect). A method that allows us to relax these assumptions is the estimation of average treatment effects based on propensity score matching (PSM). A description of the method can be found in Becker and Ichino (2002). In this section we estimate the return to a first degree using propensity score matching.

Let us define X_i as a vector of variables affecting both educational qualifications and wages, Q_i the treatment variable that equals one for the treated and zero for the non-treated (in our case it will be the dummy for having a first degree) and w_{1i} and w_{0i} the log-wage for individual i in the case of treatment and no-treatment, respectively. Following Rosenbaum and Rubin (1983) the propensity score is defined as:

$$p(X_i) \equiv Pr\{Q_i = 1|X_i\} = E\{Q_i|X_i\}, \tag{4.5}$$

i.e. the conditional probability of receiving a treatment given pre-treatment characteristics. Rosenbaum and Rubin (1983) show that if the following two hypotheses hold:

1. *Balancing hypothesis*: If $p(X_i)$ is the propensity score, then $Q_i \perp X_i | p(X_i)$;

2. *Unconfoundedness hypothesis*: Suppose that assignment to treatment is unconfounded,[19] i.e. $w_{1i}, w_{0i} \perp Q_i | X_i$. Then assignment to treatment is unconfounded given the propensity score, i.e. $w_{1i}, w_{0i} \perp Q_i | p(X_i)$;

then the Average Treatment effect on the Treated (ATT) can be estimated as follows:

$$\begin{aligned} ATT &= E\{w_{1i} - w_{0i} | Q_i = 1\} \\ &= E\{E\{w_{1i} - w_{0i} | Q_i = 1, p(X_i)\}\} \\ &= E\{E\{w_{1i} | Q_i = 1, p(X_i)\} - E\{w_{0i} | Q_i = 0, p(X_i)\} | Q_i = 1\}. \end{aligned} \tag{4.6}$$

In our case propensity score matching and ATT are implemented using the procedures `pscore` and `attnd` created by Becker and Ichino (2002), the latter using as options *Nearest Neighbour Matching* with replacement and a probit model to compute propensity scores. The procedure `pscore` also offers some diagnostics for the balancing property. Both for males and females, the balancing property was satisfied for all variables used for the computation of the propensity score. Standard errors for the ATT were computed using bootstrap and 500 replications. Table 4.9 reports the OLS estimates of the return to a first degree using the specification II in section 4.4 for the computation of the propensity scores, on the common support sample, in order

[19]This hypothesis is also called the Conditional Independence Assumption, i.e. selection only on observables, and cannot be tested within the PSM-ATT framework.

to assess the impact of the potential lack of common support. For males the ATT computed is 0.12, not statistically significant at the conventional levels and lower than that computed in section 4.6.1 using OLS that was 0.14. Lack of common support does not seem to be a problem, since more than 99% of observations fall in the common support.[20] The OLS estimates in the common support are very similar to the ATT computed using propensity score matching showing that for males a linear specification of the no-treatment outcome and the hypothesis of homogeneity of treatment effects are reasonable assumptions. The effect estimated is slightly lower than that reported in section 4.6.1. The difference is probably due to the fact that here we are considering only the sample of individuals with A-levels or a first degree, i.e. the two groups of individuals who are more directly comparable (see section 4.3). This suggests a careful choice of the control group when assessing the impact of the treatment outcome in order to ensure that the treated and control groups are as similar as possible. Furthermore, for women, the lack of common support also appears not to be a major problem, and again the estimated effect is lower than that in section 4.6.1. For females the estimated effect from the PSM-ATT procedure is 0.15, statistically significant at 5%, and lower than that obtained using OLS that was 0.18.

[20]This is the percentage of observations whose propensity score belongs to the intersection of the supports of the propensity score of treated and control observations. The problem of 'lack of common support' arises when the intersection of the two supports, of the treated and untreated groups, is very small, suggesting that treated and untreated individuals have very different observed characteristics and are not easily comparable.

4.9 Concluding remarks

In this chapter we have estimated the return to a first degree using data from the 1970 British cohort (BCS70). In order to tackle the issue of potential endogeneity of educational qualifications we have used the *proxying and matching method* that consists of including in the log-wage equation factors affecting both educational attainment and wages. This approach may be a viable solution to the possible endogeneity problem given that the data set used is very rich in information related to family background, ability related variables and past educational performance. We have replicated the analysis of Blundell *et al.* (2000) on BCS70 and have shown that while the return to a first degree is largely unchanged for men belonging to the 1958 and the 1970 cohorts, the return for women has declined substantially over the two cohorts. We estimate the wage return to be 0.15 for men and 0.23 for women in the 1970 cohort compared to 0.17 and 0.37, respectively, for the 1958 cohort. Our own preferred specifications for the 1970 cohort leads to an estimated return to a degree of 0.19 for men and 0.23 for women not controlling for secondary school performance, and 0.14 and 0.18 for men and women, respectively when controlling for it.

We have also analysed differences in returns according to both degree class and degree subjects. Our estimates show the existence of a positive additional log-wage premium for 'good' degrees compared to lower degree classifications. For both men and women, the premium for a 'good' degree over a poor degree is about 8 percentage points. However, our estimates are not very precise, probably because of relatively small cell sizes, and in our samples the hypothesis of no difference between degree classes can be rejected at the 10% level only for women when the full set of controls is included. Our results qualitatively confirm previous findings by Battu *et al.* (1999) and Naylor *et al.* (2003) who also found wage premia for a 'good' degree performance, using larger samples where the effects can be more precisely estimated, albeit without information on a non-graduate control group. Our analysis of log-wage differences by degree subjects also confirms findings from related work. As far as the ranking of subjects is concerned, for instance, we have in decreasing order: Social Sciences, Sciences and Art and Humanities, both for men and women (compared with Walker and Zhu, 2001). Moreover, Art and Humanities degrees give a positive return (relative to workers with A-levels) only to women (compared with Harkness and Machin, 1999). Although our estimates suggest the presence of differences in degree subjects, the effects are not precisely estimated and only the difference between Social Sciences and Art and Humanities appear statistically significant for males.

We have also tested for the presence of endogeneity in our estimates of the return to a first degree and to degree class using the control function approach. In both cases the hypothesis of an absence of endogeneity could not be rejected by our data. This is perhaps not a very surprising result since in the present chapter we consider a sample of individuals who have attained at least an A-level education. Once we have controlled for several family and individual characteristics including early academic ability, this sample is likely to be relatively homogeneous with respect to unobserved characteristics.

Finally, we have explored the issue of the heterogeneity of returns to a first degree by observed household and individual characteristics, other than degree class and degree subject, and the adequacy of the linear specification using a Propensity Score Matching-Average Treatment Effect approach. Our results suggest that when estimating the return to a first degree and considering as the control group individuals with A-levels only, the absence of common support is not an issue and that the linear specification and the homogeneity of treatment effects does not seem to be too strong an assumption in our sample.

In conclusion, our estimates suggest that for males, the return to a university degree in the UK is remarkably similar across the 1970 birth cohort - typically graduating in the early 1990s - and the 1958 birth cohort, typically graduating in 1979 or 1980. This is despite the significant changes taking place in UK HE during the 1980s. In contrast, our estimates suggest that the return to a degree for females fell considerably across the two cohorts.

Tables

Table 4.1: Descriptive statistics for different samples (BCS70)

Educational qualifications	30-year follow-up		Matched 30-year and 10-year follow-up		Matched 30-year and 16-year follow-up		Matched 30-year, 16-year and 10-year follow-up	
	N.	%	N.	%	N.	%	N.	%
A-levels	677	24.57	640	25.07	555	23.92	530	24.22
Non degree HE	418	15.17	378	14.81	351	15.13	329	15.04
UG degree	1,215	44.10	1,125	44.07	1,035	44.61	971	44.38
PG degree	445	16.15	410	16.06	379	16.34	358	16.36
Total	2,755	100	2,553	100	2,320	100	2,188	100
	N	mean	N	mean	N	mean	N	mean
ln(gross hourly wage)	2,144	2.270	1,982	2.267	1,821	2.267	1,713	2.265

Table 4.2: Estimates of the log-wage premia (wage 'returns') to HE qualifications, replication of Blundell *et al.* (2000) on the BCS70 - Men - OLS

	Men Specification					
	1	1'	2	2'	3	3'
Non-degree HE	0.006	0.150***	0.025	0.155***	0.007	0.144***
	(0.050)	(0.039)	(0.047)	(0.039)	(0.047)	(0.039)
UG degree	0.202***	0.208***	0.198***	0.184***	0.154***	0.171***
	(0.040)	(0.034)	(0.038)	(0.035)	(0.041)	(0.035)
PG degree	0.077	0.156***	0.077	0.144***	0.097	0.141***
	(0.063)	(0.042)	(0.063)	(0.042)	(0.056)	(0.042)
No. of observations	961	1,006	961	1,006	961	1,006
R^2	0.028	0.033	0.078	0.052	0.192	0.091

Note. The dependent variable is the natural logarithm of gross hourly wages. Robust standard errors for the presence of heteroskedasticity in parentheses. Specification 1 includes only controls for HE qualifications. Specification 2 also includes controls for ability scores, regional dummies and school type. Specification 3 adds controls for family background, socio-demographic variables, school attendance and employer charateristics. Specifications 1', 2' and 3' are the equivalent specifications reported in Blundell *et al.* (2000), Table 3, p. F90. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4.3: Estimates of the log-wage premia (wage 'returns') to HE qualifications, replication of Blundell *et al.* (2000) on the BCS70 - Women - OLS

	Women					
	Specification					
	1	1'	2	2'	3	3'
Non-degree HE	0.124***	0.261***	0.132***	0.272***	0.106***	0.223***
	(0.040)	(0.040)	(0.039)	(0.040)	(0.038)	(0.040)
UG degree	0.274***	0.391***	0.245***	0.384***	0.232***	0.368***
	(0.033)	(0.038)	(0.034)	(0.039)	(0.033)	(0.039)
PG degree	0.169***	0.427***	0.141***	0.408***	0.136***	0.368***
	(0.038)	(0.046)	(0.040)	(0.046)	(0.040)	(0.048)
No. of observations	1,017	832	1,017	832	1,017	832
R^2	0.068	0.133	0.106	0.14	0.196	0.194

Note. The dependent variable is the natural logarithm of gross hourly wages. Robust standard errors for the presence of heteroskedasticity in parentheses. Specification 1 includes only controls for HE qualifications. Specification 2 also includes controls for ability scores, regional dummies and school type. Specification 3 adds controls for family background, socio-demographic variables, school attendance and employer charateristics. Specifications 1', 2' and 3' are the equivalent specifications reported in Blundell *et al.* (2000), Table 3, p. F90. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4.4: Estimates of the log-wage premia (wage 'returns') to a first degree (BCS70) - OLS

Return to UG degree	Specification I Coef.		s.e.	Specification II Coef.		s.e.
Men						
UG degree	0.189	***	0.039	0.142	***	0.042
F-test A-level info. (p-value)		-			0.004	
F-test O-level info. (p-value)		-			0.003	
N.obs.		961			961	
R^2		0.092			0.125	
Women						
UG degree	0.229	***	0.033	0.180	***	0.035
F-test A-level info. (p-value)		-			0.056	
F-test O-level info. (p-value)		-			0.009	
N.obs.		1,017			1,017	
R^2		0.133			0.165	

Note. The dependent variable is the natural logarithm of gross hourly wages. Specifications I also includes for men: region of residence, British ability scales scores (both quantitative and non quantitative), school type and father's social class. Specification I includes for women: region of residence, British ability scales scores (both quantitative and non quantitative), school type and father's education. Specification II includes all the controls of specification I and controls for age 16 and age 18 secondary school performance. The regressors included in specification I were chosen by performing F-tests on the general specification including all controls listed in section 4.6.1. Robust standard errors for the presence of heteroskedasticity in parentheses. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4.5: Estimates of the log-wage premia (wage 'returns') by degree class (BCS70) - OLS

Return to UG degree class	Specification I Coef.		s.e.	Specification II Coef.		s.e.
Men						
'Good' degree	0.239	***	0.050	0.187	***	0.051
Lower degree class	0.153	***	0.044	0.112	**	0.047
F-test A-level info. (p-value)		-			0.004	
F-test O-level info. (p-value)		-			0.005	
F-test 'good'=lower (p-value)		0.101			0.145	
N.obs		957			957	
R^2		0.095			0.127	
Women						
'Good' degree	0.262	***	0.037	0.211	***	0.038
Lower degree class	0.183	***	0.041	0.142	***	0.041
F-test A-level info. (p-value)		-			0.061	
F-test O-level info. (p-value)		-			0.010	
F-test 'good'=lower (p-value)		0.046			0.071	
N.obs		1,017			1,017	
R^2		0.137			0.168	

Note. The dependent variable is the natural logarithm of gross hourly wages. Specifications I also includes for men: region of residence, British ability scales scores (both quantitative and non quantitative), school type and father's social class. Specification I includes for women: region of residence, British ability scales scores (both quantitative and non quantitative), school type and father's education. Specification II includes all the controls of specification I and controls for age 16 and age 18 secondary school performance. The regressors included in specification I were chosen by performing F-tests on the general specification including all controls listed in section 4.6.1. Robust standard errors for the presence of heteroskedasticity in parentheses. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4.6: Estimates of the log-wage premia (wage 'returns') by degree subject (BCS70) - OLS

Return to UG degree subject	Specification I Coef.		s.e.	Specification II Coef.		s.e.
Men						
Sciences	0.192	***	0.046	0.131	***	0.050
Social sciences	0.259	***	0.059	0.212	***	0.060
Art and Humanities	0.117	*	0.066	0.082		0.065
F-test A-level info. (p-value)		-			0.004	
F-test O-level info. (p-value)		-			0.005	
F-test Science = Social Sciences (p-value)		0.272			0.179	
F-test Science = Art and Hum. (p-value)		0.269			0.449	
F-test Social Sciences = Art and Hum. (p-value)		0.067			0.084	
F-test all subjects equal (p-value)		0.184			0.205	
N.obs		930			930	
R^2		0.099			0.132	
Women						
Sciences	0.208	***	0.039	0.136	***	0.040
Social Sciences	0.246	***	0.052	0.185	***	0.053
Art and Humanities	0.181	***	0.043	0.121	***	0.043
F-test A-level info. (p-value)		-			0.018	
F-test O-level info. (p-value)		-			0.005	
F-test Science = Social Sciences (p-value)		0.456			0.343	
F-test Science = Art and Hum. (p-value)		0.536			0.231	
F-test Social Sciences = Art and Hum.(p-value)		0.221			0.724	
F-test all subjects equal (p-value)		0.472			0.479	
N.obs		996			996	
R^2		0.136			0.174	

Note. The dependent variable is the natural logarithm of gross hourly wages. Specifications I also includes for men: region of residence, British ability scales scores (both quantitative and non quantitative), school type and father's social class. Specification I includes for women: region of residence, British ability scales scores (both quantitative and non quantitative), school type and father's education. Specification II includes all the controls of specification I and controls for age 16 and age 18 secondary school performance. The regressors included in specification I were chosen by performing F-tests on the general specification including all controls listed in section 4.6.1. Robust standard errors for the presence of heteroskedasticity in parentheses. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4.7: Estimates of the log-wage premia (wage 'returns') for an UG degree (BCS70) - CFA

	Specification II					
Return to	Men			Women		
UG degree	Coef.		s.e.	Coef.		s.e.
UG degree	0.143	***	0.044	0.176	***	0.035
$\rho\sigma$	-0.0003		0.0069	0.0001		0.0004
N.obs.	961			1,017		
R^2	0.126			0.165		
F-test father's education (p-value)						
- Education equation	0.015			-		
- Log-wage equation	0.726			-		
F-test mother's education (p-value)						
- Education equation	-			0.000		
- Log-wage equation	-			0.756		

Note. The dependent variable is the natural logarithm of gross hourly wages. Specifications II includes all the controls listed in section 4.6.1. Standard errors are bootstrapped with 500 replications since the model is estimated in two stages. ***Significant at the 1% level; **significant at the 5% level; * significant at the 10% level.

Table 4.8: Estimates of the log-wage premia (wage 'returns') by degree class (BCS70) - CFA

	Specification II					
Return to	Men			Women		
UG degree	Coef.		s.e.	Coef.		s.e.
'Good' degree	0.186	***	0.052	0.211	***	0.037
Lower degree class	0.112	**	0.044	0.142	***	0.043
$\rho\sigma$	0.0002		0.0035	0.0014		0.0017
N.obs.		957			1,017	
R^2		0.128			0.1685	
F-test father's education (p-value)						
- Education equation		0.014			-	
- Log-wage equation		0.719			-	
F-test mother's education (p-value)						
- Education equation		-			0.000	
- Log-wage equation		-			0.707	

Note. The dependent variable is the natural logarithm of gross hourly wages. Standard errors are bootstrapped with 500 replications since the model is estimated in two stages. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4.9: Estimates of the log-wage premia (wage 'returns') using PSM-ATT (BCS70)

	Specification II								
Return to		Men				Women			
UG degree	N.obs.	Coef.		s.e.	N.obs	Coef.		s.e.
PSM-ATT[a]	672	0.123	*	0.074	664	0.149	**	0.071
OLS on common support	672	0.125	***	0.043	664	0.158	***	0.036
OLS	675	0.124	***	0.043	669	0.161	***	0.036
PSM diagnostics								
% obs. in common support		99.56				99.25		
Balancing property[b]		OK				OK		
Probit pseudo R^2 [c]		0.170				0.244		

Note. The outcome variable is the natural logarithm of gross hourly wages. [a] Average Treatment on the Treated (ATT) computed using Propensity Score Matching (PSM), in particular *Nearest Neighbour Matching* with replacement, see Becker and Ichino (2002). Bootstrapped standard errors (since propensity scores are estimated using a probit model), 500 replications. [b] test for the *balancing property*, see Becker and Ichino (2002), in particular the balancing property is not rejected only in the case it holds for every single variable using to compute the propensity score. [c] Pseudo R^2 of the probit model used to compute the propensity scores, which includes all the explanatory variables listed in specification II in section 4.6.1. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Chapter 5

Conclusions

5.1 Main findings

In this Thesis we investigate the determinants of two distinct educational choices in the UK, the choice to enroll in post-compulsory education and that of enrolling in different degree subjects at the undergraduate level, and one of the consequences of the choice to acquire a university education, namely the wage return to a first degree, to different degree subjects and to different levels of degree performance.

The decision to enroll in post-compulsory education is the first educational choice that an individual faces in the UK. This choice has important consequences in terms of future labour market outcomes. In chapter two we investigate the role of an individual's family background, particularly his/her parents' income, on the decision of staying-on at school at age 16 using data from the 1970 British Cohort Study (BCS70). With respect to the previous literature our main innovations are: 1) the use of the grouped income variable provided by the BCS70, which has been converted into a continuous variable using interval regression techniques, while the other paper using the BCS70 and a continuous measure of income we are aware of, Blanden and Gregg (2004), uses income imputed from the Family Expenditure Survey; 2) an attempt to address the issue of endogeneity of family income using Instrumental Variables techniques, in particular we seek to identify exogenous variations of family income using parents' industry of employment and grandfathers' social class, following Shea (2000) and Maurin (2002); 3) an investigation of the issue of the effects of income non response on the estimate of the effect of family income. Our main findings suggest that even when all these issues are addressed, family income has only a very limited impact on children's staying-on decisions, while other family influences, such as those running through parents' education, social class and proxies of parenting quality (e.g. parental interest in a child's education) are much stronger. We also observe that our findings do not necessarily contrast with the recent experimental evidence supporting the effectiveness of the Education Maintenance Allowances (EMAs, such as Ashworth *et al.* 2002), since we estimate the effect of unconditional transfers in family income that are very different from the EMAs.

Chapter three contains an analysis of the decision to enroll in different degree subjects in the UK. In this chapter we use data from the Universities' Statistical Record (USR) 1981-1991. This topic is very important given the differences in graduates' labour market outcomes by degree subject. We introduce a behavioural model that suggests that the unobserved factors affecting the utilities that an individual derives from different degree subjects are likely to be correlated and that an econometric model allowing for this correlation, such as a multinomial probit model, is likely to offer several advantages with respect to models based on the Independence of Irrelevant Alternatives (IIA) assumption. We consider the decision to enroll in three groups of subjects: Quantitative, Non-Quantitative and Law and Medicine and focus in particular on the effect of social class. Both economists and sociologists postulate important social class influences on the choice of degree subject that might have important implications in terms of intergenerational mobility. We estimated a trinomial probit model for each of the eleven years considered, and do not find any statistically significant social class difference in the likelihood of enrolling in the different subject groups. Our sensitivity analysis suggests that this is not the result of the particular grouping of subjects adopted or of the econometric model chosen. The main explanation for the absence of social class differences may stem from the specific characteristics of the UK university system during the period studied, when there were no undergraduate tuition fees and poor students were supported by means-tested maintenance grants. This suggests that the innovations recently introduced in the UK university system, such as the introduction of undergraduate fees or the gradual replacement of student maintenance grants with student loans might have changed the earlier situation and increased the differences across social classes. Moreover, the possible introduction of differential fees could also exacerbate these differences.

In chapter four we switch from the analysis of the determinants of the educational choices to that of some economic consequences of the decision to acquire a first degree. In particular, we analyse the log-wage premium related to a first degree with respect to workers possessing only A-level qualifications. Also in this chapter, like in chapter two, we use data from the BSC70. We first compare the estimates of the wage return to a first degree in the BCS70 with those from the National Child Development Study (NCDS) data, relating to the 1958 British cohort and reported in Blundell *et al.* (2000), who used a 'proxy and matching method' approach. We observe that while the return for men has remained virtually unchanged, that for women has registered a remarkable fall. Then, we analyse the differences in the return to a first degree by degree class. Our estimates suggest differential wage premia by class of degree, graduates with a 'good' degree (first class or upper second class honours degrees) earned more than those with lower degree classes. The differences, although non negligible both for men and women, turn out to be statistically significant at conventional levels only for the latter. Also our estimates of the differences of returns by degree subjects appear to be remarkable, though they are not statistically significant. We think that this may mainly be the result of small cell size, since the number of graduates in the different fields is low in the BCS70. In this chapter we also seek to address the problem of endogeneity of education for the estimates of the return to a first degree and to degree class using a different approach, the so called Control Function Approach (CFA). The estimates using the CFA do not suggest the existence of an endogeneity problem, a result which is probably due to the high homogeneity of our sample, once we have controlled for several individual and household characteristics, which includes only individuals with at least one A-level. We also analyse the issue of heterogeneous returns using propensity score matching and find that even when allowing for this heterogeneity the estimated returns are very similar to that obtained with the 'proxy and matching method'.

5.2 Further research

There are several ways in which the analysis in this Thesis could be extended.

As to the analysis of staying on at 16, one possibility is to use data from the Youth Cohort Study (YCS) and model the transition from school to work as a polycothomous variable (see Andrews and Bradley, 1997) distinguishing, for instance, between school continuation, youth training, employment and unemployment. Although the YCS does not supply data on household income it is nonetheless possible to analyse the effect of other family background characteristics. Bradley and Nguyen (2004), for instance, use YCS data for 1992, 1994, 1996 and 1998 and a multinomial logit model with six outcomes to investigate the school-to-work transition in the UK. It would be possible to update their analysis using more recent cohorts of the YCS and also using a multinomial model which is not based on the restrictive IIA (independence of irrelevant alternatives) assumption, such as a multinomial probit model.

The analysis of subject of enrollment at the undergraduate level of chapter three can be extended in several directions. Firstly, we have already said that our analysis using USR data relates only to 'old' universities and is limited to the period 1981-1991. Given the relatively low number of low social class students enrolled in HE in the period under study, they might be considered as relatively more motivated or abler, charateristics that are generally unobservable in administrative data (our analysis assume that A-level information is a sufficient control for individual ability and academic motivation) which may have some consequences in terms of observing social class differences. Moreover, we have argued that in the period analysed the design of the UK HE system was likely to attenuate social class differences. Therefore, it would be interesting, were they publicly available, to consider recent administrative university data from the HESA. These data would allow the analysis of the 'new' universities (i.e. former-polytechnics), which are likely to have a different subject-mix with respect to 'old' universities and recruit individuals from relatively lower social class backgrounds. Both these combined factors along with the recent changes of the UK HE system might have contributed to increasing the social inequality of the system in terms of subject choice. Secondly, it is also possible to perform an analysis of undergraduate subject choice using data from the YCS. By comparing both analyses it would be possible to analyse the bias in the estimates obtained using administrative HESA data produced by the fact that many family background variables are not available, unlike in longitudinal data such as the YCS. Last but not least, the YCS also allows the possibility of analysing subject choice at secondary level and social class influences at that educational stage. This is important in order to address the question of whether social class exerts its effect mainly at early stages of the educational process.

As to the analysis of the returns to a first degree by degree subject and degree class (with respect to A-level only), we have already said that it would be useful to have access to data sets containing a higher number of graduates, which have to provide also information on non-graduates and individuals' family background and, of course, wages. However, at present there are data sets which safisty only some of these requirements. The Labour Force Survey, for instance, has data on a high number of graduates but does not contain a set of background variables comparable to longitudinal studies such as the NCDS or the BCS70. The BCS70, that we have used in chapter four, gathers a wealth of information on family background variables, individual educational variables and wages, but has a limited number of graduates. The YCS contains many information on background variables, a sufficient number of graduates but does not have data on wages of graduates except for cohort three which was also observed at age 23. Hence, a possible extension of our work is the analysis of 'early' wages of graduates using the latter cohort, which refers to individuals born in the early 1970s and therefore highly comparable with the individuals from the 1970 cohort observed in the BCS70.

Bibliography

Abbott, A. and Leslie, D. (2004), 'Recent trends in higher education applications and acceptances', *Education Economics*, 12: 67-86.

Andrews, M. and Bradley, S. (1997), 'Modelling the transition from school and the demand for training in the United Kingdom', *Economica*, 64: 387-413.

Arcidiacono, P. (2004), 'Ability sorting and the returns to college major', *Journal of Econometrics*, 121: 343-375.

Ashworth, J. and Evans, J. L. (2001), 'Modeling student subject choice at secondary and tertiary level: a cross-section study', *Journal of Economic Education*, 32: 311-320.

Ashworth, K., Hardman, J., Hartfree, Y., Maguire, S., Middleton, S., Smith, D., Dearden, L., Emmerson, C., Frayne, C. and Meghir, C. (2002), 'Education maintenance allowance: the first two years. A quantitative evaluation', Department for Education and Skills Research Report No. 352, DfEE, London.

Battu, H., Belfield, C. and Sloane, P. (1999), 'Overeducation among graduates: a cohort view', *Education Economics*, 7: 21-38.

Becker, G. (1975), *Human capital.* Chicago: Chicago University Press.

Becker, G. (1981), *A treatise on the family.* Cambridge (Mass.): Harvard University Press.

Becker, S. and Ichino, A., (2002), 'Estimation of average treatment effects based on propensity scores', *The Stata Journal*, 2: 358-377.

Belfield, C., Bullock, A., Chevalier, A., Fielding, A., Siebert, W. S. and Thomas, H. (1997), *Mapping the careers of highly qualified workers.* Bristol: Higher Education Funding Council for England.

Berger, M. C. (1988), 'Predicted future earnings and choice of college major', *Industrial and Labor Relations Review*, 41: 418-429.

Betts, J. R. (1996), 'What do students know about wages? Evidence from a survey of undergraduates', *Journal of Human Resources*, 31: 27-56.

Blackaby, D. H., Murphy, P. D. and O'Leary, N. C. (1999), 'Graduate earnings in Britain: a matter of degree?', *Applied Economics Letters*, 6: 311-315.

Blakemore, A. E. and Low, S. A. (1984), 'Sex differences in occupational selection: the case of college majors', *Review of Economics and Statistics*, 66: 157-163.

Blanden, J., Goodman, A., Gregg, P. and Machin, S. (2003), 'Changes in intergenerational mobility in Britain', CEPR Discussion Paper n. 517, LSE, London.

Blanden, J. and Gregg, P. (2004), 'Family income and educational attainment: a review of approaches and evidence for Britain', *Oxford Review of Economic Policy*, 2: 245-263.

Blanden, J., Machin, S. (2004), 'Educational inequality and the expansion of UK higher education', *Scottish Journal of Political Economy*, 51: 230-249.

Blau, D. M. (1999), 'The effect of income on child development', *Review of Economics and Statistics*, 81: 261-276.

Blundell. R., Dearden, L., Goodman, A. and Reed, H. (2000), 'The returns to higher education in Britain: evidence from a British cohort', *Economic Journal*, 110: F82-F89.

Blundell, R., Dearden, L. and Sianesi, B. (2003), 'Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS', IFS Working Paper no. 03/20, The Institute for Fiscal Studies, London.

Boudon, R. (1974), *Education, opportunity and social inequality*. New York: Wiley.

Bound, J., Jaeger, D.A. and Baker, R.M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association*, 90: 443-450.

Bourdieu, P. (1984), *Distinction: a social critique of the judgement of taste.* London: Routledge and Kegan Paul.

Bradley, S. and Nguyen, A. N. (2004), 'The school-to-work transition' in G. Johnes and J. Johnes (eds), *International Handbook on the Economics of Education*, 484-521. Cheltenham: Edward Elgar.

Bratti, M. (2002b), 'Does the choice of university matter? A study of the differences across UK universities in life sciences students' degree performance', *Economics of Education Review*, 21: 431-444.

Bratti, M. and Mancini, L. (2003), 'Differences in early occupational earnings of UK male graduates by degree subject: evidence from the USR, 1980-1993', IZA Discussion Paper No. 890, IZA, Bonn.

Bratti, M., McKnight, A., Naylor, R. and Smith, J. (2004), 'Higher education outcomes, graduate employment and university performance indicators" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167: 475-496.

Brunello, G., Lucifora, C. and Winter-Ebmer, R. (2001), 'The wage expectations of European college students', IZA Discussion Paper No. 299, IZA, Bonn, forthcoming in *Journal of Human Resources.*

Bunch, D. S. (1991), 'Estimability in the multinomial probit model', *Transportation Research B*, 25: 1-12.

Callender, C. (2003), 'Student financial support in higher education: access and exclusion. In M. Tight (ed.), *Access and exclusion: international perspectives on higher education research.* London: Elsevier Science.

Cameron, S. and Heckman, J. (1998), 'Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males', *Journal of Political Economy*, 106: 262-333.

Cameron, S. and Heckman, J. (2001) 'The dynamics of educational attainment for black, hispanic and white males', *Journal of Political Economy*, 109: 455-499.

Card, D. (1999), 'The causal effect of education on earnings' in O. Ashenfelter and D. Card (eds), *Handbook of labor economics, Volume 3A*, 1801-1863. Amsterdam: Elsevier.

Carneiro, P. and Heckman, J. (2002), 'The evidence on credit constraints in post-secondary schooling', *Economic Journal* 112: 705-734.

Chevalier, A. (2002), 'Just like daddy: the occupational choice of UK graduates', mimeo, LSE, London.

Chevalier, A. (2003), 'Measuring over-education', *Economica*, 70: 509-531.

Chevalier, A. (2004), 'Parental education and child's education: a natural experiment', IZA Discussion Paper no. 1153, IZA, Bonn.

Chevalier, A. and Conlon, G. (2003), 'Does it pay to attend a prestigious university?', IZA Discussion Paper No. 848, IZA, Bonn.

Chevalier, A., Conlon, G., Galindo-Rueda, F. and Mcnally, S. (2002), 'The returns to higher education teaching', Report to the Department of Education and Skills, Center for the Economics of Education, LSE, London.

Chevalier, A. and Lanot, G. (2002), 'The relative effect of family characteristics and financial situation on educational achievement', *Education Economics*, 10: 165-182.

Cigno, A. (1991), *Economics of the family*. Oxford: Clarendon Press.

Dansie, B. (1985), 'Parameter estimability in the multinomial probit model', *Transportation Research B*, 19: 526-528.

Datcher-Loury, L. (1988) 'Effects of mother's home time on children's schooling', *Review of Economics and Statistics*, 70: 367-373.

Davies, S. and Guppy, N. (1997), 'Fields of study, college selectivity, and student inequalities in higher education', *Social Forces*, 75: 1417-1438.

Dearden, L. (1999a) 'The Effects of families and ability on men's education and earnings in Britain', *Labour Economics*, 6: 551-567.

Dearden, L. (1999b), 'Qualifications and earnings in Britain: how reliable are conventional OLS estimates of the returns to education?' IFS Working Paper no. 99/7, The Institute for Fiscal Studies, London.

Dearden, L., Machin, S. and Reed, L. (1997), 'Intergenerational mobility in Britain', *Economic Journal*, 107: 47-64.

Dearing, R. (1997), *Higher education in the learning society.* National Committee of Inquiry into Higher Education, London.

DfEE (1999), *Statistical first release*, Department for Education and Employment, London.

Doeringer, P. and Piore, M. (1971), *Internal labor makets and manpower analysis.* Lexington, MA: D.C. Health.

Dolton, P. and Makepeace, G. H. (1990). 'The earnings of economics graduates', *Economic Journal*, 100: 237-250.

Dolton P., Makepeace, G. and Inchley, G. D. (1990), 'The early careers of 1980 graduates: earnings, earnings differentials and postgraduate study', Employment Department Group Research Paper No.78, London: HMSO.

Dolton, P. and Vignoles, A. (2000) 'The incidence and effects of overeducation in the UK graduate labour market', *Economics of Education Review*, 19: 179-98.

Dominitz, J. and Manski, C. F. (1996), 'Eliciting student expectations of the returns to schooling', *Journal of Human Resources*, 31: 1-26.

Dustmann, C. and Micklewright, J. (2001), 'Intra-household transfers and the part-time work of children', CEPR Discussion Paper No 2796, LSE, London.

Elliot, C. D., Murray, D. J. and Pearson, L. S. (1979), *British ability scales, manual 4: tables of abilities and norms.* Windsor: NFER.

Ermisch, J. and Francesconi, M. (2000), 'Educational choice, families, and young people's earnings', *Journal of Human Resources*, 35: 143-176.

Ermisch, J. and Francesconi, M. (2001a), 'Family matters: impacts of family background on educational attainments', *Economica*, 68: 137-156.

Ermisch, J. and Francesconi, M. (2001b), 'Family structure and children's achievements', *Journal of Population Economics*, 14: 249-270.

Feinstein, L. (2000), 'The relative economic importance of academic, psychological and behavioural attributes developed in childhood', Center for Economic Performance Discussion Papers n. 443, LSE, London.

Feinstein, L. and Symons, J. (1999) 'Attainment in secondary school', *Oxford Economic Papers*, 51: 300-321.

Finegold, D. (1993), 'Breaking out the low skilled equilibrium', *Education Economics*, 1: 77-83.

Green, F., McIntosh, S. and Vignoles, A. (1999), 'Overeducation and skills - clarifying the concepts', CEP Discussion Paper no. 453, London.

Greenaway, D. and Haynes, M. (2003), 'Funding higher education in the UK: the role of fees and loans', *Economic Journal*, 113: F150-F166.

Griliches, Z. (1977), 'Estimating the returns to schooling: some econometric problems', *Econometrica*, 45: 1-22.

Hansen, M. N. (2001), 'Education and economic rewards. Variations by social-class origin and income measures', *European Sociological Review*, 17: 209-231.

Harkness, S. and Machin, S. (1999), 'Graduate earnings in Britain, 1974-95', Department for Education and Employment, Research Report RR95, DfEE, London.

Hartog, J. (2000), 'Over-education and earnings: where are we, where should we go?', *Economics of Education Review*, 19: 131-147

Haveman, R. and Wolfe, B. (1995) 'The determinants of children's attainments: a review of methods and findings', *Journal of Economic Literature*, 33: 1829-1878.

Heckman, J. (1979), 'Sample selection bias as a specification error', *Econometrica*, 47: 153-161.

Heckman, J., Lochner, L. and Todd, P. (2003), 'Fifty years of Mincer earnings regressions', IZA Discussion Paper, No. 775, IZA, Bonn.

Heckman, J. and Robb, R. (1985), 'Alternative identifying assumptions in econometric models of selection bias', *Advances in Econometrics*, 5: 243-287.

Heckman, J. and Sedlacek, G. (1985), 'Heterogeneity, aggregation, and market wage functions: an empirical model of self selection in the labor market', *Journal of Political Economy*, 93: 1077-1125.

Kane, T., Rouse, C.E. and Staiger, D. (1999), 'Estimating the returns to schooling when schooling is misreported', National Bureau of Economic Research, Working Paper no. 7235.

Katz, L.F. and Summers, L.H. (1989) 'Industry rents: evidence and implications', *Brooking Papers on Economic Activity*, Microeconomics, 209-275.

Keane, M. P. (1992), 'A note on identification in the multinomial probit model', *Journal of Business and Economics Statistics*, 10: 193-200.

Kelsall, R. K., Poole, A. and Kuhn, A. (1972), *Graduates: the sociology of an elite.* London: Methuen.

Krueger, A. and Summers, L. (1988) 'Efficiency wages and the inter-industry wage structure', *Econometrica*, 56: 259-193.

Lauer, C. (2003), 'Family background, cohort and education. A French-German comparison based on a multivariate ordered probit model of educational attainment', *Labour Economics*, 10: 231-251.

Lee, L.-F. (1983), 'Generalized econometric models with selectivity', *Econometrica*, 51: 507-512.

Leslie, D. (2003) 'Using the success to measure quality in British higher education: which subjects attract the best-qualified students?, *Journal of The Royal Statistical Society, Series A*, 166: 329-347.

Lissenburgh, S. and Bryson, A. (1996), 'The returns to graduation', Research Studies RS15, Department for Education and Employment. London.

Machin, S. and Gregg, P. (2003), 'A lesson for education: university expansion and falling income mobility', *New Economy*, 10: 194-198.

Makepeace, G., Paci, P., Joshi, H. and Dolton, P. (1999), 'How unequally has equal pay progressed since the 1970s? A study of two British cohorts', *Journal of Human Resouces*, 34: 534-556.

Mancini, L. (2003), *Higher education in the UK and the market for labour: evidence from the Universities' Statistical Record*, Ph.D. Thesis, University of Warwick, Coventry, December 2003.

Mason, G. (1995), 'The new graduate supply-shock. Recruitment and utilisation of graduates in British industry', National Institute of Economic Research, Report Series no. 9, London.

Mason, G. (1999), 'The labour market for engineering, science and IT graduates: are there mismatches between supply and demand', Research Brief No. 112, Department for Education and Employment, Sheffield.

Maurin, E. (2002), 'The impact of parental income on early schooling transitions: a re-examination using data over three generations', *Journal of Public Economics*, 85: 301-332.

Mayer, S. (1997), *What money can't buy: family income and children's life chances*. Camdridge (MA): Harvard University Press.

Micklewright, J. (1989) 'Choice at sixteen', *Economica*, 56: 25-39.

Montgomery, J. D. (1991), 'Social networks and labor-market outcomes: toward an economic analysis', *American Economic Review*, 81: 1408-1418.

Montmarquette, C., Cannings, K. and Mahseredjian, S. (2002), 'How do young people choose college majors?', *Economics of Education Review*, 21: 543-556.

Murphy, K. and Topel, R. (1990) 'Efficiency wages reconsidered: theory and evidence'. In: Weiss, Y., Fishelson, G. (Eds.), *Advances in the theory of measurement of unemployment*, MacMillan, London, pp. 204-242.

Naylor, R. and Smith, J. (2004), 'Returns to education: a signaling approach', mimeo, University of Warwick, Coventry.

Naylor, R., Smith, J. and McKnight, A. (2002), 'Why is there a graduate earnings premium for students from independent schools?', *Bulletin of Economic Research*, 54: 315-339.

Naylor, R., Smith, J. and McKnight, A. (2003), 'Returns to educational performance: evidence from UK graduates' first destination labour market outcomes', mimeo, University of Warwick, Coventry.

Oosterbeek, H. and Webbink, D. (1997), 'Is there a hidden technical potential?', *De Economist*, 145: 159-177.

Pereira, P.T. and Martins, P.S. (2004), 'Returns to education and wage equations', *Applied Economics*, 36: 525-531.

Pissarides, C. (1981), 'Staying-on at school in England and Wales', *Economica*, 48: 345-363.

Pissarides, C. (1982), 'From school to university: the demand for post-compulsory education in Britain', *Economic Journal*, 92: 654-667.

Polachek, S. W. (1978), 'Sex differences in college major', *Industrial and Labor Relations Review*, 31: 498-508.

Pradhan, M. and van Soest, A. (1995), 'Formal and informal sector employment in urban areas of Bolivia', *Labour Economics*, 2: 275-297.

Rice, P. G. (1987) 'The demand for post-compulsory education in the UK and the effects of educational maintenance allowances', *Economica*, 54: 465-75.

Rigg, M., Elias, P., White, M. and Johnson, S. (1990), *An overview of the demand for graduates*, HMSO, London.

Rochat, D. and Demeulemeester, J.-L. (2001), 'Rational choice under unequal constraints: the example of Belgian higher education', *Economics of Education Review*, 20: 15-26.

Romer, D. (1996) *Advanced macroeconomics*. New York: McGraw-Hill.

Rosenbaum, P. and Rubin, D.B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, 70: 41-55.

Shea, J. (2000) 'Does parents' money matter?', *Journal of Public Economics*, 77: 155-84.

Sloane, P. J. and O'Leary, N. C. (2004), 'The return to a university education in Great Britain'. IZA Discussion Paper no. 1119, IZA, Bonn.

Smith, J., McKnight, A. and Naylor, R. A. (2000), 'Graduate employability: policy and performance in higher education in the UK, *Economic Journal*, 110: F382-411.

Smith, J. and Naylor, R. A. (2001a), 'Determinants of degree performance in UK universities: a statistical analysis of the 1993 student cohort', *Oxford Bulletin of Economics and Statistics*, 63: 29-60.

Smith, J. and Naylor, R. A. (2001b), 'Dropping-out of university: a statistical analysis of the probability of withdrawal for UK university students', *Journal of the Royal Statistical Society, Series A*, 164: 389-405.

Stata (2003), *Reference S-Z. Release 8.* College Station, TX: Stata Corporation, pp. 248-263.

Stewart, M. B. (1983) 'On least squares estimation when the dependent variable is grouped', *Review of Economic Studies*, 50: 737-753.

Stewart, M. B. and Swaffield, J. K. (1999), 'Low pay dynamics and transition probabilities', *Economica*, 66: 23-42.

Todd, P. E. and Wolpin, K. I. (2003), 'On the specification and estimation of the production function for cognitive achievement', *Economic Journal*, 113: F3-F33.

Van de Ven, W. P. and van Praag, B. M. (1981), "The demand for deductibles in private health insurance: A probit model with sample selection", *Journal of Econometrics*, 17: 229-252.

Van de Werfhorst, H. G., de Graaf, N. D. and Kraaykamp, G. (2001), 'Intergenerational resemblence in field of study in the Netherlands', *European Sociological Review*, 17: 275-293.

Van de Werfhorst, H. G., Sullivan, A. and Cheung, S. Y., (2003), 'Social class, ability and choice of subject in secondary and tertiary education in Britain', *British Educational Research Journal*, 29: 41-62.

Vella, F. and Verbeek, M. (1999), 'Estimating and interpreting models with endogenous treatment effects', *Journal of Business and Economic Statistics*, 17: 473-478.

Walker, I. and Zhu, Y. (2001), *The returns to education. Evidence from the Labour Force Surveys.* Department for Education and Skills, Research Report 313, DfES, London.

Weeks, M. (1997), 'The multinomial probit model revisited: a discussion of parameter estimability, identification and specification testing', *Journal of Economic Surveys*, 11: 297-320.

Wolter, S. C. and Zbinden, A. (2002), 'Labour market expectations of Swiss university students', *International Journal of Manpower*, 23: 458-470.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross-section and Panel Data.* Cambridge (MA): The MIT Press.