Università Politecnica delle Marche
DIPARTIMENTO DI ECONOMIA

Differences in Early Occupational Earnings of UK Male Graduates by Degree Subject: Evidence from the 1980-1993 USR

Massimiliano Bratti and Luca Mancini

QUADERNI DI RICERCA n. 189

Settembre 2003

Abstract

This paper investigates the differences in early occupational earnings of UK male graduates by degree subject during the period 1980-1993. We match administrative student-level data from the Universities' Statistical Record (USR) and occupational earnings information from the New Earnings Survey (NES). The paper estimates relative earnings premia by degree subject using three alternative modelling approaches to control for student self-selection into university courses: i) a proxying and matching method, ii) a propensity score matching method, and iii) a simultaneous equations model of subject choice and earnings determination. Our analysis shows that there is a substantial amount of sample selection originating from unobservable student characteristics. The ranking of university subjects based on relative earnings premia is sensitive not only to the modelling approach but also to time, showing that analyses focusing on single-year data may not generalise to other periods.

JEL Class.: C34, C35, J21, J31
Keywords: degree subject, earnings, graduates, self-selection, university

Indirizzo: **Massimiliano Bratti**, Dipartimento di Economia, Università Politecnica delle Marche, Piazzale Martelli 8, I-60121 Ancona, Italy and University of Warwick, Coventry, UK. email: massib@dea.unian.it
Luca Mancini, Department of Economics, University of Warwick, CV4 7AL, Coventry, UK. email: luca.mancini@warwick.ac.uk

Differences in Early Occupational Earnings of UK Male Graduates by Degree Subject: Evidence from the 1980-1993 USR*

Massimiliano Bratti and Luca Mancini

* *The paper benefited from presentations at the Westminster Business School, at the Irish Economic Association Conference 2003 and at the European Economic Association Conference 2003. We acknowledge both the USR, as the original depositor, and the UK Data Archive for the use of the data set SN: 3456 Universities' Statistical Record. The authors wish to thank Robin Naylor, Jeremy Smith, Ian Walker and an anonymous referee for useful comments. The responsibility for the content of the paper is the authors' only.*

1. Introduction

The economic literature has shown that there are substantial positive returns to an undergraduate university degree in the UK. Blundell *et al.* (2000), for instance, using National Child Development Survey (NCDS) data found that the average return to an undergraduate degree, in terms of wages, with respect to individuals aged 33 with two or more A-level passes who did not continue into higher education, was 17% for men and 37% for women in 1991. However, the vast majority of the Mincer-type earnings regressions estimate only an average rate of return to a university degree and do not control for field of study, while in a recent review of the literature Chevalier *et al.* (2002) show that the private rate of return to a university degree is likely to differ substantially by degree subject. Moreover, to the best of our knowledge, we are not aware of any UK studies that also attempt to model directly student self-selection into university subjects, despite acknowledging the importance of the potential endogeneity of subject choice.[1]

This paper aims to contribute to the empirical return-to-education literature by estimating early occupational earnings *premia* by subject studied using alternative methods to control for the potential endogeneity of subject choice. Firstly, we use OLS estimation techniques widely used in related research for the UK. This standard approach is not only interesting for comparison purposes with previous research, but also represents a useful benchmark to assess the existence and size of a potential selection bias. Secondly, this first set of results is contrasted with estimates obtained from the propensity score matching method that has become an increasingly popular technique in the evaluation literature (Rosenbaum and Rubin, 1983). Although substantially different, both methods hinge on the assumption that selection is driven solely by observable factors. Finally, we introduce a third approach estimating a simultaneous model of graduate earnings and subject choice (Lee, 1983), which allows for self-selection through unobservable attributes.

The concept of heterogeneous returns across degree courses is particularly relevant over time. As more individuals experience higher education, just holding a university degree becomes a weaker distinguishing mark for students and a less informative screening device for the talent at the disposal of employers, if not supplemented by information on the graduates' awarding university, field of study, or degree class obtained. On the grounds that the economic return to a degree in a certain subject depends on the demand and supply for that specific university specialisation, our multi-cohort analysis over

[1] Blundell *et al.* (2000) acknowledge the issue and use a 'proxying and matching' method to reduce the impact of self-selection. However, this methodology relies on rather restrictive assumptions, namely that the wealth of information available is sufficient to control for sample selectiviction.

the period 1980-1993 is also expected to provide useful information on the past trends of the graduate labour market in the UK.
The outline of the paper is as follows. In section 2 we report the findings of previous studies on the differences in graduate earnings by degree subject in the UK. Section 3 illustrates three alternative modelling strategies. Section 4 discusses some issues regarding occupational earnings data, while section 5 describes the salient features of the sample. Section 6 presents the three sets of results for male graduates obtained from OLS, propensity score matching, and the simultaneous equation model, respectively. Section 7 concludes summarising the main results.

2. Previous literature

In comparison to the rich literature investigating the return to education in the UK (see for instance Harmon and Walker 1995, 1999, 2000 and Blundell *et al.,* 2000), there are only few studies which analyse differences in graduate earnings by degree course.
Dolton and Makepeace (1990) using data from the Survey on 1980 Graduates and Diplomates conducted by the Department for Employment in the UK find that the average earnings of Commerce graduates (including Accounting, Business & Management, Economics and Law) were higher than the earnings of graduates in other disciplines.
Belfield *et al.* (1997) use survey data on the 1985 and 1990 graduate cohorts to investigate differences in 1996 average salaries by subject of degree. The authors find that the relative ranking of degree subjects based on average male salaries remained unchanged for the two cohorts, with Social Sciences ranked first followed in order by Science, Humanities and Education.
Chevalier (2000) estimates *premia* by degree subject using pooled survey data on the 1985 and 1990 graduate cohorts and finds that graduates from Mathematics and Social Science earned respectively 6% and 2% more than graduates in Education, while Humanities graduates earned 12% less.
Chevalier *et al.* (2002) use a similar specification on the 1980 graduate cohort and find earnings differences with respect to Education of +8.4%, +11.6% and +6.8% for Science, Social Science and Humanities, respectively. The corresponding figures for the 1995 cohort (using average salaries in 1999) were +17.9%, +16.8% and +5.4%, showing a further advancement of Science and Social Sciences with respect to Education.
Naylor *et al.* (2002) use USR individualised data on the 1993 graduate cohort and find significant differences in inter-occupational earnings across degree subjects. The most economically rewarding subjects were Law, Computer

Science, Medical Related and Economics and Business, while the least economically rewarding were Agriculture, Humanities, and Classics & Literature.

Lissenburgh and Bryson (1996) use data from the 4th wave of the Youth Cohort Study 3 (YCS) and find that graduates from Science, Mathematics and Engineering earned 9% more than other graduates.

Blundell *et al.* (2000) using UK data from the National Child Development Study (NCDS) find that graduates in Economics, Accountancy and Law performed significantly better in terms of hourly wages (at age 33) than those with undergraduate degrees in other subjects (such as Arts, Chemistry & Biology and Other, the residual category).

Harkness and Machin (1999) use data from the General Household Survey (GHS) between 1980 and 1995 focusing on the return to degree subject for full-time workers. They consider four broad subject groups (Arts, Science, Social Science and Other) and find, *inter alia*, that the return to a degree in Arts for males was about 10% lower than in the other fields.

Blackaby et al. (1999) using UK data from the 1993-1995 Labour Force Surveys (LFS) find that male graduates' earnings vary significantly across degree subjects. In particular, after controlling for a number of personal, job and demographic characteristics, the authors find that graduates from Economics, Accountancy, Law and Management did better than their peers in other subjects, especially compared to Other Social Sciences and Arts.

Walker and Zhu (2001) using LFS data for the period 1993-1999 find that "there are no systematic trends in returns by subject nor is there any tendency for them to converge" (p. 37). Their study shows marked differences in the return to an undergraduate degree across subjects. Graduates in Economics, Law, "Health" (i.e. medical related) and Mathematics ranked at the top of the earnings scale while graduates in Arts performed substantially worse (with a negative mark-up with respect to students with at least two A-level passes who did not continue in higher education).

3. Methodology

This section describes three alternative models used to estimate graduate occupational earnings premia[2] by subject of degree: i) the 'proxying and matching' method (OLS); ii) the propensity score matching-average treatment on the treated method (PSM-ATT); iii) a simultaneous equations model of earnings determination and subject choice (MNL-OLS).

[2] In the paper, by 'earnings premia' we refer to 'log-earnings premia'.

a. *Selection on observable factors: the 'proxying and matching method' (OLS)*

A common method to ascertain earnings differences by degree subject is to estimate by OLS an earnings function specified as: [3]

$$y_{ij} = \sum_{j=1}^{J} S_{ij}\theta_j + X_i\beta + \varepsilon_i \qquad (1)$$

where y_{ij} is the natural logarithm of the earnings of individual *i* who studied subject *j*, X_i is a vector of individual attributes which may affect both subject choice and occupational earnings, S_{ij} is a dummy variable which takes value one if the individual *i* graduated in the subject *j* and zero otherwise, and θ_j is the earnings premium of graduating from subject *j* relative to the default case. As observed by Blundell *et al.* (2000), this is tantamount to matching individuals on the basis of the index $X_i\beta$ and to assuming equality of θ_j's across individuals. The OLS model does not require any distributional assumption on ε_i, but it does require orthogonality between ε_i and X_i.

b. *Selection on observable factors: PSM-ATT method*

An alternative method to estimate subject *premia* is to compare occupational earnings for individuals who graduated in one subject with 'matched' individuals who studied for a different degree course. This framework considers the subject of study as the treatment that the individual receives and aims to assess the causal effect of this treatment on the outcome variable, namely occupational earnings. The direct comparison between individuals in different treatment groups may be misleading because they may differ systematically in their observable and unobservable characteristics. In their seminal paper, Rosenbaum and Rubin (1983) suggested the use of propensity scores matching procedures to solve the issue of sorting due to observable factors. The propensity score is defined as the conditional probability of receiving the treatment given an individual's characteristics:

$$p(X_i) \equiv \Pr\{S_{ij} = 1 \mid X_i\} \qquad (2)$$

[3] This is the method used in all the studies reviewed in section 2.

where S_{ij} are dummy variables which take value one if individuals graduated in the subject j and zero otherwise, and X_i is the vector of conditioning factors that we observe. In our case, the propensity scores are computed by estimating binary logit models of subject choice for each of the broad course categories defined in Section 4 using Economics and Business graduates as the reference group.[4]

Under the assumption that differences between individuals affecting the outcome are entirely captured by their observed characteristics X_i,[5] the average treatment effect on the treated (ATT) can be estimated as follows:

$$ATT = E\{y_{i1} \mid P(X_i), S_{ij} = 1\} - E\{y_{i0} \mid P(X_i), S_{ij} = 1\} \tag{3}$$

where y_{i1} and y_{i0} are the occupational earnings of graduates in subjects 1 and 0, respectively. In words, individuals with the same value of the propensity score $P(X_i)$ but different treatments S_{ij} , act as controls for each other and the average difference between their earnings equals the ATT.

Compared to the *proxying and matching* method (OLS), the assumption of equality of the subject *premia* θ_j's across individuals is relaxed. In fact, here the earnings *premia* are computed as the average of the earnings differences between 'matched' pairs of treated-untreated individuals. This method is non-parametric and does not require any distributional assumption on the unobservables.

c. *Selection on observable and unobservable factors: a simultaneous equations model of earnings determination and subject choice*

The two methods illustrated in sections 3.*a* and 3.*b* rely on the assumption that treated and untreated individuals differ only with respect to observable attributes. Hence, these approaches neglect the possibility of *self-selection* with

---

[4] This implies the breakdown of our sample into *J-1* sub-samples of graduates in each year. Each sub-sample includes the 'treated' individuals, that is the individuals who graduated in a specific subject and the 'untreated' individuals, i.e. the individuals who graduated in the reference subject (Economics and Business) and, therefore, received a different 'treatment'. Strictly speaking, we are evaluating the differential impact of alternative treatments.

[5] This is known as the Conditional Independence Assumption (CIA), formally: $y_{0i} \perp S_{ij} \mid P(X_i)$. The other necessary assumption is the so-called 'common support' assumption: '[All treated agents have a counterpart on the non-treated population and anyone constitutes a possible participant: $0 < P(S_{ij} = 1 \mid X_i) < 1$]' (Blundell and Costa Dias, 2002, p. 22).

respect to unobservable characteristics. If individuals choose the degree subject by maximising their individual utility, the students enrolled in the different subjects are those for whom the choice turns out to be optimal. If this sorting effect is not fully accounted for by observable attributes, both the OLS and the PSM-ATT estimates of the earnings *premia* by subject are likely to be biased.

The econometric framework we use to address the self-selection on unobservables in a polycotomous choice model was originally developed by Lee (1983). Below, we report the main features of the model.

Let us assume that the utility of the student i in the subject j (V_{ij}), with j=1,…J, depends on individual's characteristics (Z_i) and an idionsyncratic unobservable term u_{ij} , reflecting for instance individual's preferences over degree subjects, in the following way:

$$V_{ij} = Z_i \delta_j + u_{ij} \qquad \text{j=1,…,J.} \tag{4}$$

Occupational earnings are generated according to the following process:

$$y_{ij} = \sum_{j=1}^{J} S_{ij} \theta_j + X_i \beta + \varepsilon_i \tag{5}$$

where y_{ij} are the log-earnings of individual i who read subject j, X_i is a vector of individual attributes, S_{ij} is a dummy variable which takes value one if the individual i graduated in the subject j and zero otherwise, and θ_j's are the subjects earnings *premia*, which are the primary focus of our analysis. The *selection bias* arises from the correlation (ρ_j) between the stochastic components u_{ij}'s and ε_i, that is, between the unobserved individual's characteristics affecting subject choice and those affecting occupational earnings. If the model does not account for this correlation, the subject dummies may simply pick up the effect of the individual's unobserved characteristics rather than the 'true' earnings *premium* associated with the subject studied. For instance, because the type of occupation is typically related to the subject studied at university, individuals with a preference for certain jobs will be more likely to choose those subjects more related to their preferred occupation. After graduation, an individual will be more likely to be observed in his/her preferred occupation, which in turn will affect his/her earnings.[6] Therefore, the earnings *premia* by subject studied may also capture the effect of

[6] Arcidiacono (2002), for instance, shows that ability sorting observed across majors is mainly determined by the different preferences for certain jobs and certain subjects rather than by expected performance or expected earnings.

idiosyncratic occupational preferences. Another possible source of selection bias could be the higher preference for some non-pecuniary characteristics of the job by graduates in certain subjects. This may explain, for instance, why graduates in Humanities are at the bottom of the earnings scale. [7]
Lee (1983) suggests that the model above can be estimated using maximum likelihood methods, under some specific assumptions on the distributions of the stochastic terms u_{ij}'s and ε_i. Here we follow Lee (1983) and assume that the u_{ij}*'s* are independent and identically Gumbel-distributed, while $\varepsilon_i \sim N(0,\sigma^2_\varepsilon)$. The form of the log likelihood function in this specific case is shown in the Appendix B. The attractiveness of estimating simultaneously a Multinomial Logit-OLS model (MNL-OLS hereafter) is that the model does not impose restrictions on the correlations between the unobservables affecting the subject choice and the individual's earnings, which are jointly estimated along with the other parameters of the model.[8]
In brief, we estimate simultaneously a MNL model for the subject choice and an earnings regression in which the degree subject appears as one of the explanatory variables. In this way, we aim to estimate the differences in graduate earnings by degree subject corrected for *self-selection bias*.

Model identification. Even when the vectors Z_i and X_i coincide, the different functional forms of Equations (4) and (5) (non-linear *vs* linear) are sufficient to identify the simultaneous equations model. To ensure that identification does not rely exclusively on the specific functional forms adopted, it is necessary to find variables that affect subject choice but do not influence earnings (i.e. we look for an 'economic' identification). We restricted our focus on A-level curriculum and age.[9] From a theoretical point of view, the type of secondary school curriculum is a pre-requisite (in terms of type of pre-university knowledge or entry requirements) for some university courses, and should therefore affect subject choice (for instance, see Altonji, 1993). Van de Werfhorst *et al.* (2002) provide some empirical evidence for the UK. By

[7] Daymont and Andrisani (1984), for instance, find that students in Humanities have weaker preferences for pecuniary job characteristics (earnings).
[8] Therefore, we do not impose the independence of u_{ij}*'s* and ε_i, which would allow the separate estimation of the earnings equation and the subject choice model.
[9] We did not focus on social class since it might affect earnings through family networks or other effects (see for instance Hansen, 2001), residence prior to entry university since it might pick up school quality effects, type of school since there is evidence showing that it affects earnings (see Naylor *et al.* 2002), marital status as there are studies providing evidence of the existence of an earning premium for married workers (see Ginther and Zavodny 2001) and A-level score because it is often considered as a proxy for individual ability.

contrast, we do not have a strong *a priori* about its importance for graduates' earnings, once controlled for the degree subject.[10] As for student's age, Becker's (1993) human capital theory predicts that younger individuals, who have a longer expected working life, have higher returns to education and also to more selective and lucrative subjects requiring higher effort. Davies and Guppy (1997), for instance, found that older students were less likely to enter more lucrative fields. Therefore, we considered age as a potential candidate for the identification of our simultaneous equation model. Although we acknowledge that age may affect earnings through work experience accumulated prior to university enrolment and, therefore, may be a weak identifying variable, we expect this problem to be less severe when occupational earnings (rather than actual salaries) are used.

Once potential 'candidates' were identified, we tested whether the chosen variables actually affect subject choice but not earnings determination. OLS earnings regressions were run separately by year of graduation. In any year, the set of identifying variables (ID.Vs henceforth) were selected from the set of 'candidates' by excluding the least significant (the one with the highest p-value above the threshold of 0.10), and performing a likelihood ratio (LR hereafter) test for the validity of the restriction. If the test passed, the second least significant variable (with p-value greater than 0.10) was excluded and a new cumulative LR test performed. This procedure was reiterated until the cumulative LR test for the joint omission was rejected. We investigated the sensitivity of the estimated *premia* to alternative identification strategies (functional form *vs* functional form and ID.Vs). The results are reported in Appendix C.

4. Earnings data and control variables

The analysis presented in this paper is based on USR data on individual university students who graduated from 'pre-1992' UK universities between 1980 and 1993. Unfortunately, the USR does not include information on employed graduates' individual salaries. However, the First Destination Record does contain detailed information on the (self-reported) type of occupation held by graduates six months after graduation.[11] As in Naylor *et al.* (2002), we were able to match the individual's reported occupation with the corresponding (gender-specific) 3-digit SOC of the New Earnings Survey

[10] A-level Mathematics is an exception and, therefore, is not included among the potential identifying variables.

[11] In the First Destination Record of the USR, occupations are classified into more than 120 categories contained into 6 "major" groups and 28 "minor" groups.

(NES). [12] Occupational earnings were then computed as the average gross weekly pay of individuals employed full-time (in the same occupation) in the year following graduation. [13]

The length of time since it was first conducted in 1975 makes the NES an ideal source to study long-term trends in pay. However, the change in the NES occupational classifications that took place in 1990 has required a significant coding effort to ensure consistency/continuity over time.[14] Prior to 1990, the coding scheme used was the Key List of Occupations for Statistical Purposes (KOS), which consisted of 404 occupations arranged into 18 main groups. From 1990 onwards occupational data were coded to the Standard Occupational Classification (SOC), consisting of 371 unit groups contained into 77 "minor", 22 "sub-major", and 9 "major" groups. To help bridge these two coding schemes, individuals were classified under both schemes in 1990. However, the match was fairly imprecise. Some KOS occupations were scattered across a number of SOC occupations and vice versa. This, in turn, meant that for some occupations there was a jump in the average earnings series in 1990 due to this reclassification process (Bell and Elias, 2000).

Notwithstanding the reclassification, the existence of a dual coding in 1990 and the fact that the USR classification remained substantially unchanged over the period 1980-1993, enabled us to achieve a satisfactory level of consistency in the earnings series before and after 1990. However, the quality of the match was significantly higher for males. For female graduates, the repercussions of the 1990 change in the NES coding schemes on the coherence of the SOC-to-USR mapping of occupations over time, and ultimately on the calculation of average earnings, have been more serious. This was largely due to the generally smaller sample size of occupational groups for females both in the NES and in the USR, which caused average earnings to be more volatile over time. [15] To some extent, changes in occupational segregation by gender and in women's participation over time may have contributed to exacerbating the

[12] The NES is an annual survey of pay and hours of work, by far the largest of its kind in the UK, producing 2 million observations between 1975 and 1998 (Bell and Elias, 2000). Unlike most surveys of earnings, the information is collected from employers rather than from employees. This is generally accepted as producing more accurate estimates of earnings, since employers are perhaps less inclined to misrepresent employees' earnings than employees themselves.

[13] The age range considered in the computation of average earnings is 18-63 for men and 18-59 for women.

[14] The SOC-to-USR and the KOS-to-SOC mappings and the calculation of average occupational earnings were kindly provided by Abigail McKnight. We are grateful to her for making the data available to us.

[15] A smoothing of the break in the female earnings in 1990 would have required aggregating occupational groups up to a level which did not guarantee enough inter-occupation variation to estimate our earnings equations efficiently.

breaks in the female occupational earnings series. For these reasons, in what follows we will restrict our empirical analysis to male graduates only.

The use of occupational earnings has advantages and disadvantages compared to individual starting salaries. A clear disadvantage is the loss of any intra-occupational variation in pay. On the other hand, occupational earnings have the advantage of being a better proxy for career earnings and, therefore, a better measure of the lifetime rate of return to a university degree, compared to starting salaries. We are aware that occupational earnings based on graduates' first destination information collected six months after graduation may be only weakly correlated to later career earnings. However, there is some evidence suggesting that this is not necessarily the case. Dolton and Makepeace (1992) find that there is little career mobility between graduates' occupations six months and six years after leaving university. Purcell and Pitcher (1996) report that nearly 60% of students had already started seeking employment by the last term of their final year, and more than half planned to embark upon career-related job. Moreover, as a more recent study by McKnight (1999) suggests, unemployment 6 months after graduation is a surprisingly good predictor of longer-term difficulties in the labour market.

We focus on five broadly defined subject areas: [16]

1. *Science* (including Life Sciences, Physical Sciences and Mathematical Sciences);
2. *Hi-Tech* (including Computer Science, Engineering and Technology);
3. *Eco-Bus* (Economics and Business);
4. *HSS* (Humanities and Other Social Sciences);
5. *Other* (a residual and rather heterogeneous category).[17]

The purpose of this paper is to estimate relative earnings *premia* by subject studied over time. Therefore, it is crucial to control for a number of individual factors, which are also expected to affect graduate earnings. They are:

(i) *Family background*. Graduates were grouped into six social classes: SC I (Professionals), SC II (intermediate), SC IIINM (skilled non manual), SC IIIM (skilled manual), SC IV-V (partly skilled and unskilled), SC OTH (other);

[16] See Appendix A for a detailed definition of the five subject groups and the complete list of the variables used in the regressions. Due to the complexity of the model and the number of parameters to be estimated we were not able to consider a finer definition of academic subjects. A similar level of aggregation is used both in the articles reviewed in section 2, and in international studies on college major's choice correcting for sample selection (see for instance Berger, 1988, and Rochat and Demeulemeester, 2001).

[17] For the pooled 1980-1993 sample the composition of this category is as follows: Law (27.45%), Arts (8.49%), Subjects Allied to Medicine (15.80%), Education (5.92%), Combined (16.90%), Other (25.44%).

(ii) *Schooling background*. This includes controls for A-level grades, number of A-level passes by broad subject field, curriculum breadth, and the type of school attended (LEA comprehensive, LEA selective, independent, other);

(iii) *Personal characteristics*. These include age (mature student status), marital status, and residence prior to university.

Our dependent variable is the natural logarithm of average nominal gross weekly occupational earnings. In the next section we present the main features of the sample.

5. Sample and summary statistics

Given the focus of the paper, we only consider those graduates who reported an occupation six months after graduation. [18] Amongst the employed, we excluded non-UK students, medical students, and individuals from non-traditional (outside A-level) entry routes to higher education. We further excluded part-time graduates.[19] After selection, cohort size ranges between 17,100 and 21,300 for males in the period considered. We focus in the present analysis only on male graduates, because of the superior quality of earnings data.

Figure 1 shows the pattern of average occupational earnings by broad subject over time for male graduates.

It is also interesting to look at the variability of occupational earnings in each subject. Table 1 shows that during the entire sample period, Hi-Tech graduates were those with the lowest variance in occupational earnings, generally followed by Science graduates. This may be simply due to the fact that Hi-Tech graduates find employment in occupations with similar pay levels. However, it can also be the case that Hi-Tech degrees are more 'specialistic', in the sense that graduates with this specialisation are observed in a relatively narrow range of occupations. Consequently, there might be little variation in occupational earnings across individuals. As a crude test for the latter hypothesis, Table 2 shows for each subject the distribution of graduates

[18] Like all the studies reviewed in section 2, in this paper we do not address the issue of the potential biases due to self-selection into employment and survey non-response.

[19] The reasons for their exclusion are, respectively: non-UK students since they may represent a very self-selected population, medical students since there is not enough variation in their occupations and non-A level students, since we use A levels information to model the degree subject choice. Part-time students are also excluded from the analysis, since they have previous working experience and probably different early career outcomes with respect to full-time students. They generally represent less than 2% of the annual samples.

across broad occupation and sector categories. According to this criterion, Eco-Bus subjects are the most ‘specialistic’ because 40.8% of male graduates go into Accounting occupations. The corresponding proportions in the modal occupations for Science, Hi-Tech, and HSS graduates are 10.2%, 26.3%, and 12.5%, respectively. These results are in line with the findings reported in Dolton and Makepeace (1990). [20]

With respect to the sector of employment, the ranking of subjects varies considerably by gender. Hi-Tech graduates are the most ‘specialistic’, with the Engineering & Construction (EC) sector attracting over 50% of the employed. In the other subject groups, the modal sector attracts 15%, 37% and 21% of Science, Eco-Bus, and HSS graduates, respectively.

The evidence presented in Table 2 seems to confirm the intuition that the degree of specialisation of university subjects is negatively correlated with the dispersion of occupational earnings.

Figure 2 shows the proportion of graduates by broad degree subject and year. For males, the proportion of graduates in Hi-Tech degrees increased steadily in the first half of the 1980s, and remained rather stable until 1990, when numbers started to decline. The proportion of HSS graduates fell in the early 1980s but has since increased, especially in the 1990s. After a decline in the period 1980-1982, the proportion of Science graduates has generally increased during the 1980s. In 1989 figures started to fall but during the 1990s trends remained stable. Finally, the proportion of Eco-Bus graduates had slightly increased throughout the sample period.

The next section presents and compares the estimation results from the three alternative methods discussed in section 3.

6. Results

Table 3 reports, for each cohort, three sets of relative subject *premia* (with standard errors) estimated from equations (1), (2)-(3), and (6), in Appendix B, respectively. To aid interpretation, and particularly to help understand the dynamics of relative earnings *premia*, the results are also shown graphically in Figures 3 (OLS), 4 (PSM-ATT), and 5 (MNL-OLS). The discussion that follows examines the results obtained using each method in turn and is largely based on the graphical analysis.

---

[20] Dolton and Makepeace (1990) construct an ‘entropy’ index measuring the degree of specialisation of academic subjects in terms of first-job destinations. The most ‘specialistic’ degrees were found to be, in decreasing order, Education, Law, Health, Engineering, Economics & Accounting.

a. *OLS*

A first important result is that in all years the differences in occupational earnings by degree subject are highly statistically significant (Table 3, part I).[21] Consequently, the return to a university education estimated by standard Mincerian earnings regressions, which typically do not control for subject of study, is only an average measure and it fails to capture the marked differences in returns that exist across broadly defined subjects.

Second, Figure 3 shows that the relative rank of degree subjects is stable over time, except for Science and Hi-Tech whose ranks swap position in 1980 and 1990.

Third, it is clear that over the whole period Eco-Bus graduates (the reference category) had positive earnings *premia* with respect to graduates in other disciplines.[22] This result is in line with the findings of most of the UK literature reviewed in section 2. [23]

Fourth, it is also evident that the average return to a university degree was much more similar across subjects in 1980. Relative to Eco-Bus, the (negative) *premium* of graduating in Science, Hi-tech or HSS was -2.4%, -3.1% and -3.4%, respectively. The size of these *premia* widened between 1981 and 1987 when the gaps reached -8.3%,-8% and -9.7% (for Science, Hi-tech and HSS). [24] Whilst Science and Hi-Tech gained ground on Eco-Bus between 1987 and 1991, in the last two years the gap widened again. By contrast, the relative rank of graduates in HSS worsened in the first half of the 1980s when the negative *premium* stabilised at about -10% with respect to Eco-Bus.

[21] Only in 1982, 1983, and 1989 are the relative subject *premia* of *Other* subjects not statistically significant.

[22] The only exception is represented by the category labeled as "other", whose relative *premia* are positive from 1989 onwards. Given the high level of heterogeneity of this group (see footnote 17), we do not comment this result.

[23] Our estimates are not directly comparable with the findings of most of the literature reviewed in section 2, since we use a different definition of earnings and only observe university students. However, the qualitative results, and especially the relative ranking of degree subjects, closely replicate the findings of those studies.

[24] Here, we are mainly concerned with the description of the trend in earnings mark-ups due to degree subjects and do not have the ambition to explain their causes. We do not comment on the relative premium to "Other" subjects as this category is highly heterogeneous and must be interpreted as a residual group, including subjects that do not fit into the other four categories considered in the analysis.

b. *PSM-ATT*

Table 3 (part II) shows the average treatment effects on the treated (with standard errors).[25] In the matching procedure, we used the single nearest neighbours matching.[26] The relative (to Eco-Bus) earnings *premia* of Science, Hi-Tech and HSS are all statistically significant and negative, in line with the OLS results. Figure 4 shows that the subject *premia* for Science and Hi-Tech run parallel to, but are systematically lower than, their OLS counterparts. Therefore, relative to OLS, PSM-ATT yields higher (in absolute value) negative earnings *premia* for Science and Hi-Tech graduates. On the contrary, with the exception of 1984 and 1987, the negative relative earnings *premium* associated to HSS graduates is lower (in absolute value) when PSM-ATT techniques are used. These trends have a bearing on the dynamics of the relative ranking of subjects over time. The most striking result is the change in the relative position of HSS, which is no longer always at the bottom of the earnings scale. In some years, HSS graduates have enjoyed earnings *premia* with respect to both Science and Hi-Tech.

c. *MNL-OLS model*

The selected ID.Vs by year, selected following the procedure indicated in section 3.*c*, are reported in Table 4, which shows the results of the LR test for their exclusion from the MNL model and the earnings equations (estimated with OLS).[27] It is worth noting that the set of ID.Vs is rather stable over time. Furthermore, the high significance in the MNL model and the lack of statistical significance in the earnings equations, suggests that they are potentially good instruments.[28]

The relative subject *premia* estimated by MNL-OLS are shown in Table 3 (part III) and in Figure 5. The results look very different from those obtained from the OLS and the PSM-ATT methods. For instance, there is a high and statistically significant correlation between the unobservables u_{ij} affecting subject choice and the ε_i influencing occupational earnings for Hi-Tech

---

[25] Standard errors are bootstrapped using 500 replications.

[26] Diagnostics on the matching procedure for each year are available upon request from the authors. The average percentage of observations out of the common support for the full period is 0.53%.

[27] See Bound *et al.* (1995).

[28] In Appendix C we investigate the sensitivity of the estimated *premia* to alternative identifying strategies (functional form *vs* functional form and exclusion restrictions). Our results show that the estimated earnings *premia* are generally robust to changes in the identifying strategy.

graduates (Table 5 shows that ρ is always greater than 0.74 in absolute value). With the noticeable exceptions of the 1981, 1988, and 1991-93 cohorts, the correlation is generally positive suggesting that those factors inducing enrolment into Hi-Tech courses also tend to command higher earnings in the labour market. As a consequence, the OLS and the PSM-ATT estimates of the Hi-Tech earnings *premium* are biased upwards ('positive selection bias'). However, as noted above, in 1981, 1988, and 1991-93 the direction of the bias is reversed (i.e. 'negative selection'). Since ρ is the correlation between two sets of unobservables, it is difficult to offer an economic interpretation of the direction and magnitude of the bias. [29] A tentative explanation for the 1981 and 1991-93 peaks is the 'specialistic' nature of Hi-Tech courses. In fact, in these two periods the UK manufacturing sector, which employs the bulk of Hi-Tech graduates, suffered from a severe crisis. In 1981 the sector experienced its worst crisis since the start of the economic recession in the late 1970s, also contributing, to an extent, to accelerate the secular expansion of the service sector. Similarly, in the period 1990-92, the engineering industry experienced a 10% contraction. [30] These negative sector trends may have had the effect of magnifying the negative OLS *premia* of Hi-Tech relative to Eco-Bus, because the negative sector effect is wrongly ascribed to the subject studied. Clearly, the other 'outlier' found in 1988 is more difficult to justify under this line of reasoning, since the late 1980s were a period of steady expansion for the manufacturing sector. A second explanation is linked to evidence produced by Nicholson and Souleles (2002) in a study of physicians' income. They find that income prediction errors might be very different according to the speciality undertaken. These errors may depend on unanticipated market and practice changes. For instance physicians practising in a market where the demand for their services or the payments from health insurers increased, each earned about \$29,000 more than expected. This shows how market factors can change the actual realisations of income, and individuals who expected a substantially lower (higher) income might turn out to receive big unexpected gains (losses) from a specific occupation. This might also explain why the sign of ρ changes dramatically year-on-year. Finally, the use of occupational earnings based on first occupation could magnify the size of subject's *premia* and exacerbate their volatility over time. In fact, on one hand, the distribution of new graduates across occupations and jobs is likely to be more sensitive to economic fluctuations or to sector-

---

[29] We tried to explain these peaks using business cycle indicators both at an aggregate and sectoral level. However, the results were inconclusive, in the sense that we found no correlation between the estimated *premia* and the cyclical components of the real GDP and employment obtained using Hodrick-Prescott filtering techniques.

[30] D. Grow, 'Recession in Engineering worse than 1990' The Guardian, Wednesday, 3 October 2001.

specific shocks compared to the whole stock of graduates in the labour market. On the other hand, occupational earnings do not account for intra-occupational differences in pay by level of experience. This aspect of the data is expected to magnify inter-occupational differences vis-à-vis starting salaries, because the gradient of pay levels to work experience can be very different across occupations (e.g. teaching and engineering).

To a certain extent, the time profile of the Science *premium* mimics the pattern of Hi-Tech. The smoother time profile of the earnings *premium* for the former may be explained by the fact that Science graduates generally find employment in both the service and manufacturing sectors[31] and the service sector has been less sensitive to economic fluctuations than the manufacturing sector. Consequently, Science graduates were less exposed to the effects of economic fluctuations than Hi-Tech graduates. Unlike OLS, MNL-OLS results suggest that Science graduates earned on average more than Eco-Bus graduates in some years. Table 5 shows the existence of 'negative selection' for Science graduates (i.e. $\rho < 0$).

We observe a negative *premium* associated to HSS relative to Eco-Bus throughout the sample period, except for the 1980. However, the magnitude of the *premium* is smaller (the average for the whole period falls from –8.8% to –3.7%). This is due to the existence of a 'positive selection' bias for Eco-Bus courses and the absence of a significant selection effect for HSS graduates (see Table 5). It is also interesting to note that HSS graduates are not systematically at the bottom of the earnings distribution, as generally indicated by the OLS results.

Finally, ρ is always positive and generally significant for Eco-Bus degrees. In addition, the size of the bias is particularly high in 1980-81, 1988, and 1991. Again, the concomitant positive selection bias for Eco-Bus and the negative selection bias for Hi-Tech commented above, may explain the widening of the relative earnings *premium* associated with the latter degrees in these periods.

Given the volatility of the MNL-OLS *premia*[32] in Figure 5 we no longer observe any systematic ranking of subjects over time.[33] This result is in line with the evidence found for a more recent period by Walker and Zhu (2001).

[31] This is due to the more 'generalistic' nature of Science degrees (see Table 3).

[32] In Appendix D we show a graphical analysis comparing OLS, PSM-ATT and MNL-OLS earnings *premia* for each subject over time.

[33] In Appendix E we performed likelihood ratio tests for the equality of earnings *premia* by subject between consecutive years. The results reject at the 1% statistical level the null hypothesis of equality for every pair of consecutive years.

7. Concluding remarks

This paper has presented alternative estimates of the occupational earnings of UK university graduates by degree subject using USR and NES data from 1980 to 1993. The analysis is innovative because it does not limit itself to recognize that subject choice may be endogenous to the determination of earnings *premia*, but attempts to correct directly for student self-selection into degree courses. The results obtained from a standard OLS approach are contrasted with estimates from propensity score matching techniques, which correct for selectivity through observable characteristics only, and simultaneous equation models of earnings determination and subject choice, which also account for selectivity through unobservables. We find that, irrespective of the estimation technique used, the differences in occupational earnings by degree subject are in general statistically significant. This confirms that the return to a university degree estimated by standard earnings regressions controlling only for the level of educational attainment, and not for the subject studied, has to be considered only as the average return to a university degree, with marked differences even across broadly defined subjects. Our main findings from the three approaches are:

(a) OLS: graduates in Economics and Business subjects had positive earnings *premia* with respect to graduates in other subjects for the whole period. This result is consistent with the existing literature for the UK. The relative ranking of subjects remained unchanged between 1980 and 1993, with Economics & Business followed by Hi-Tech (including Engineering, Technology, and Computer Science), Science, and Humanities & other Social Sciences.
(b) Propensity score matching: using a semi-parametric matching approach to compare the earnings of individuals from different subjects we still find that Eco-Bus graduates rank first in the earnings scale. However, we observe that the relative penalty associated with graduating in Science and Hi-Tech is higher throughout the sample period compared to OLS. By contrast, HSS graduates are no longer ranked last.
(c) Maximum likelihood: when taking into account the potential self-selection of students into field of study based also on unobservable individual characteristics, no stable ranking of subjects by relative earnings *premia* emerges over time. Furthermore, we generally observe 'positive selection' for graduates in Economics & Business and Hi-Tech (except for periods of economic downturn for the latter group), 'negative selection' for Science graduates, while no selection was generally found for graduates in Humanities & other Social Sciences. This last result of nonzero correlation between unobservable factors driving subject choice and occupational

earnings cast serious doubts on the reliability of estimates based on methods where selectivity runs only through observable characteristics. In fact, earnings differences due to individual unobserved characteristics may be wrongly ascribed to the subject of degree. Moreover, earnings *premia* are likely to change over time, thus affecting the relative ranking of subjects even between consecutive years. As a consequence, studies focusing on specific cohorts of graduates may give only a very short-term account of the relative economic return to different degree subjects.

References

Altonji, J. G. (1993). "The Demand for and the Return to Education when Education Outcomes are Uncertain." Journal of Labour Economics **11**(1): 48-83.

Arcidiacono, P. (2002). "Ability Sorting and the Returns to College Major". Forthcoming in Journal of Econometrics.

Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago and London: Chicago University Press.

Bell, D., and Elias, P. (2000). *Analysis of Pay Trends. A Teaching Profession for the 21st Century. A Report for the McCrone Inquiry*. Edinburgh: McCrone Commision.

Berger, M. C. (1988). "Predicted Future Earnings and Choice of College Major". Industrial and Labour Relations Review **41**(3): 418-429.

Blackaby, D. H., Murphy, P. D., and O'Leary, N. C. (1999). "Graduate Earnings in Britain: A Matter of Degree?". Applied Economic Letters **6**(5): 311-315.

Blundell, R. and Costa Dias, M. (2002). Alternative Approaches to Evaluation in Empirical Microeconomics. Cemmap Working Paper n. 10/02. Center for Microdata Methods and Practice, UCL, London.

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**(464): F82-F89.

Bound, J., Jaeger, D. A. and Backer, R. M. (1995). "Problems with Instrumental Variables Estimation when the Correlation between the

Instruments and the Endogeneous Explanatory Variable Is Weak". Journal of the American Statistical Association **90**(430): 443-450.

Chevalier, A. (2000). Graduate Over-Education in the UK. CEE Discussion Papers no. 8. Centre for the Economics of Education, LSE, London.

Chevalier, A., Conlon, G., Galindo-Rueda. F. and Mcnally, S. (2002). *The Returns to Higher Education Teaching*. Research Report to the Department of Education and Skills. London: Centre for the Economics of Education, LSE.

Davies, S. and Guppy, N. (1997). "Fields of Study, College Selectivity, and Student Inequalities in Higher Education". Social Forces **75**(4): 1417-1438.

Daymont, T. N. and Andrisani, P. J. (1984). "Job Preferences, College Major, and the Gender Gap in Earnings". Journal of Human Resources **19**(3): 408-428.

Dolton, P. J. and Makepeace, G. H. (1990). "The Earnings of Economic Graduates". Economic Journal **100**(399): 237-250.

Dolton, P. J. and Makepeace, G. H. (1992). "The Early Careers of 1980 Graduates. Work Histories, Occupational Choice and Job Tenure", The Department of Employment Research Paper no. 79. Department of Employment, London.

Ginther, D. K. and Zavodny, M. (2001). "Is the Male Marriage Premium Due to Selection? The Effect of Shotgun Weddings on the Return to Marriage". Journal of Population Economics, 14(2): 313-328.

Hansen, M. N. (2001). "Education and Economic Rewards. Variations by Social Class Origin and Income Measures". European Sociological Review, 17(3): 209-231.

Harkness, S. and Machin, S. (1999). *Graduate Earnings in Britain, 1974-1995*. Department for Education and Employment Research Report. Nottingham: Department for Education and Employment.

Harmon, C. and Walker, I. (1995). "Estimates of the Economic Return to Schooling for the United Kingdom". American Economic Review **85**(5): 1278-1286.

Harmon, C. and Walker, I. (1999). "The Marginal and Average Return to

Schooling in the UK". European Economic Review **43**(4-6): 879-887.

Harmon, C. and Walker, I. (2000). "The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales". Economica **67**(265): 19-35.

Lee, L. F. (1983). "Generalized Econometric Models with Selectivity". Econometrica **51**(2): 507-512.

Lissenburgh, S. and Bryson, A. (1996). *The Returns to Graduation*. London: Department for Education and Employment.

McKnight, A. (1999). "Graduate Employability and Performance Indicators: First Destination and Beyond" in P. Elias and A. McKnight (eds) *Moving On - Graduate Careers Three Years after Graduation*, 24-50. Manchester: Higher Education Careers Services Unit (CSU).

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**(4): 315-339.

Nicholson, S. and Souleles, N. S. (2001). "Physician Income Expectations and Specialty Choice". NBER Working paper n. 8536. Chicago: NBER.

Purcell, K. and Pitcher, J. (1996*). Great Expectations: The New Diversity of Graduate Skills and Aspirations*. Manchester: Careers Services Unit.

Rochat, D. and Demeulemeester, J. L. (2001). "Rational Choice under Unequal Constraints: the Example of Belgian Higher Education". Economics of Education Review **20**(1): 15-26.

Rosenbaum, P. R. and Rubin, D. B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". Biometrika **70**(1): 41-55.

Sianesi, B. (2001). "Implementing Propensity Score Matching Estimators with STATA". Presented at the UK Stata Users Group, VII Meeting, London.

Van de Werfhorst, H. G., Sullivan, A., and Cheung, S. Y. (2002). "Social Class, Ability, and Choice of Subject in Secondary and Tertiary Education in Britain". Forthcoming in British Educational Research Journal.

Walker, I. and Zhu, Y. (2001). *The Returns to Education: Evidence from the Labour Force Surveys*. London: Department of Education and Skills.

Figure 1. Male gross weekly average occupational earnings (pounds sterling) by subject and year

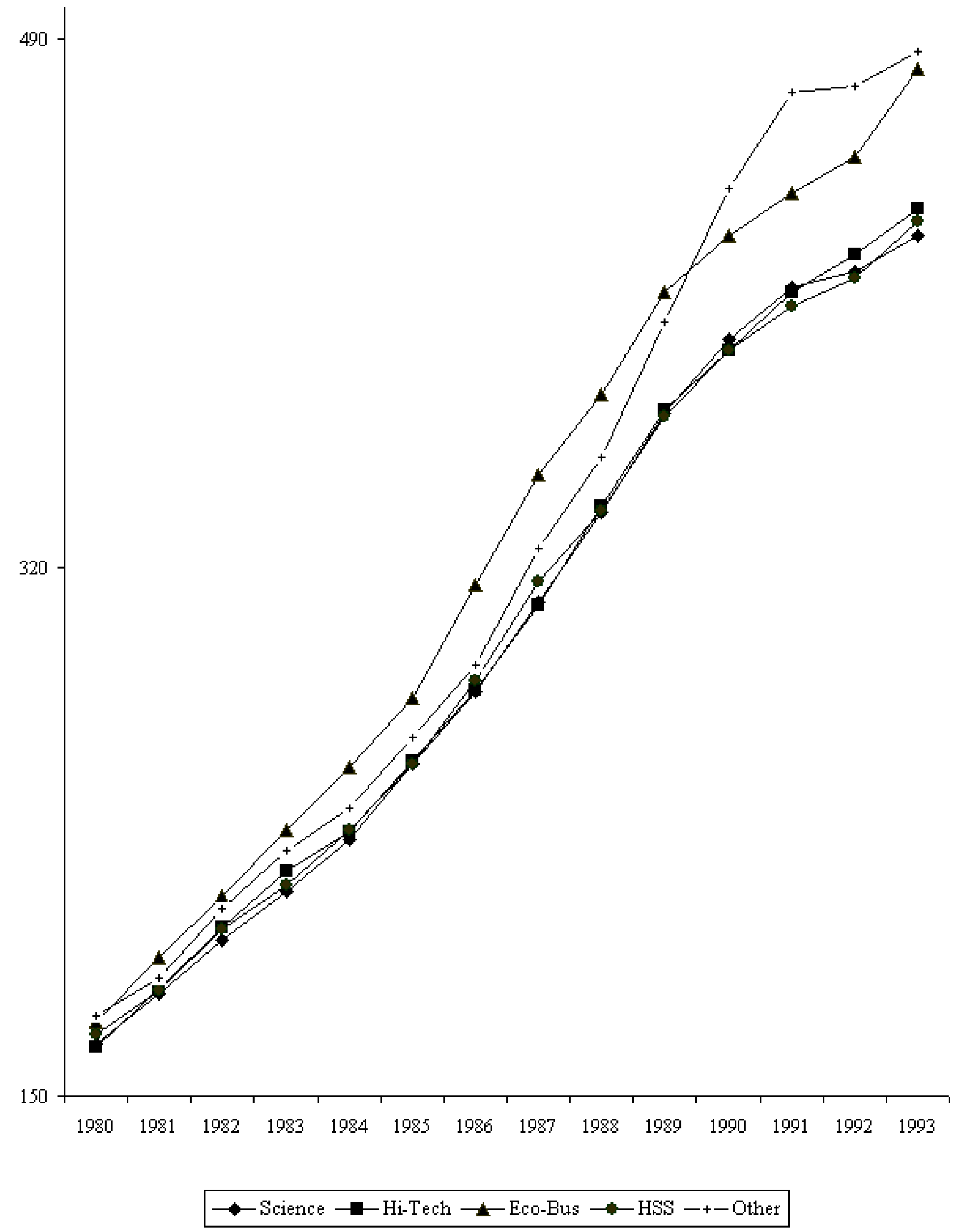


Figure 2. Proportion (%) of male graduates by 'broad' degree subject and year

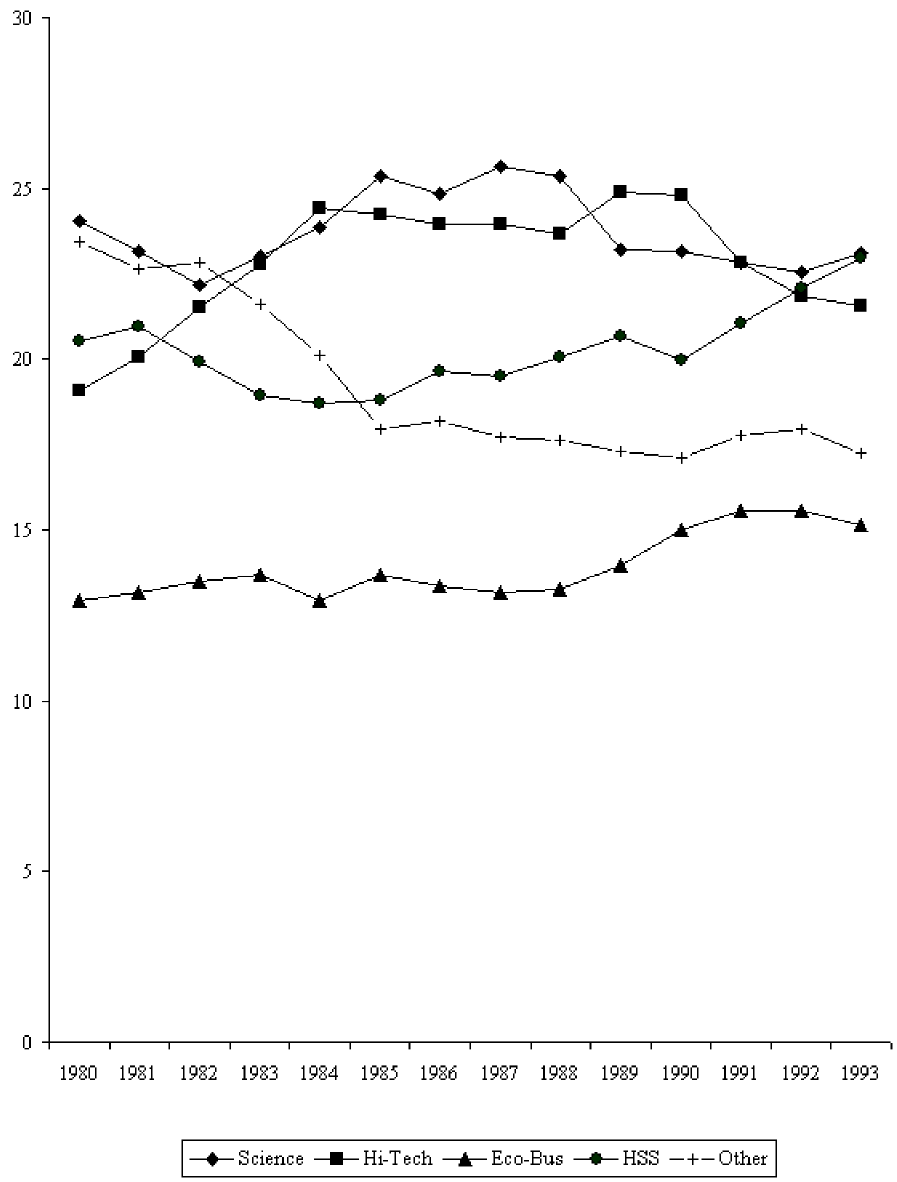


Figure 3. Relative male graduates earnings premia by degree subject and year: OLS

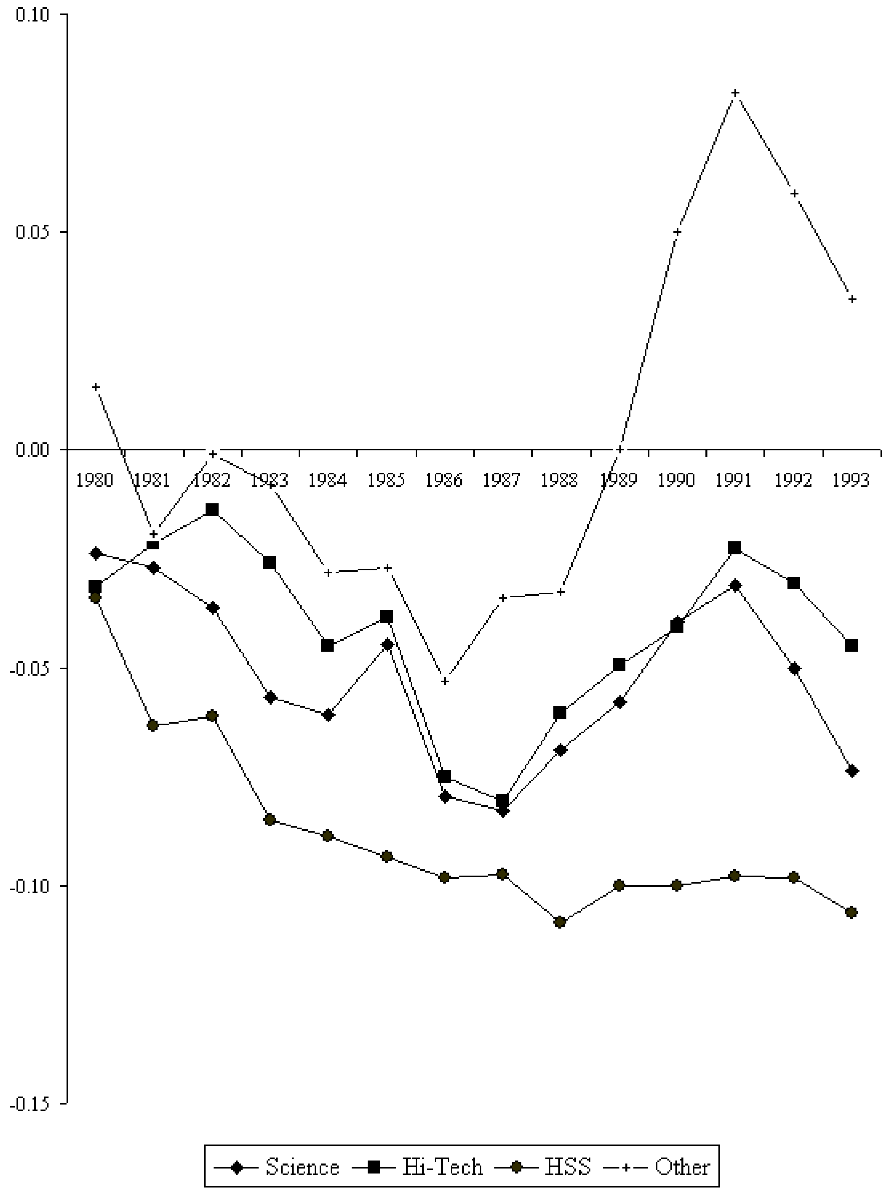


Figure 4. Relative male graduates earnings premia by degree subject and year: PSM-ATT

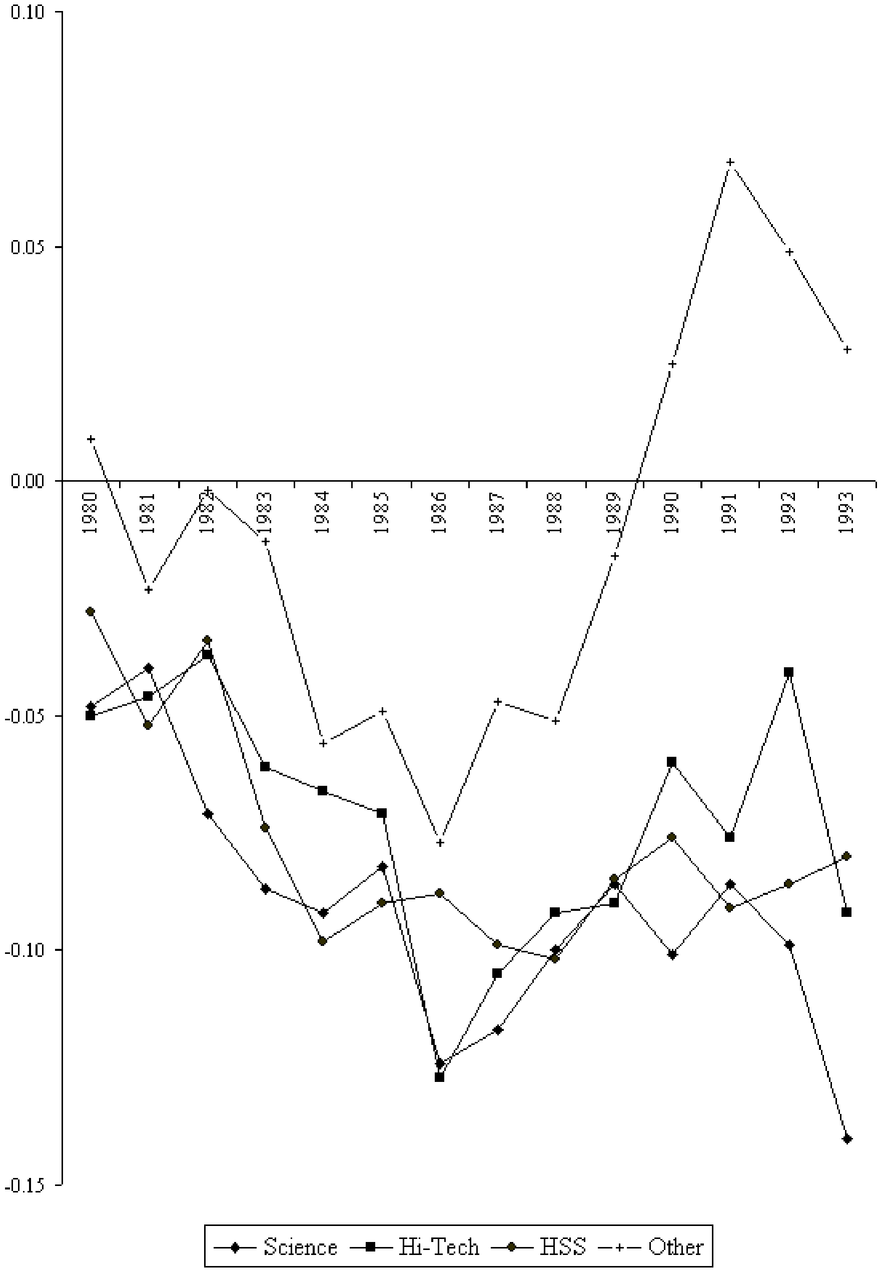


Figure 5. Relative male graduates earnings premia by degree subject and year: MNL-OLS

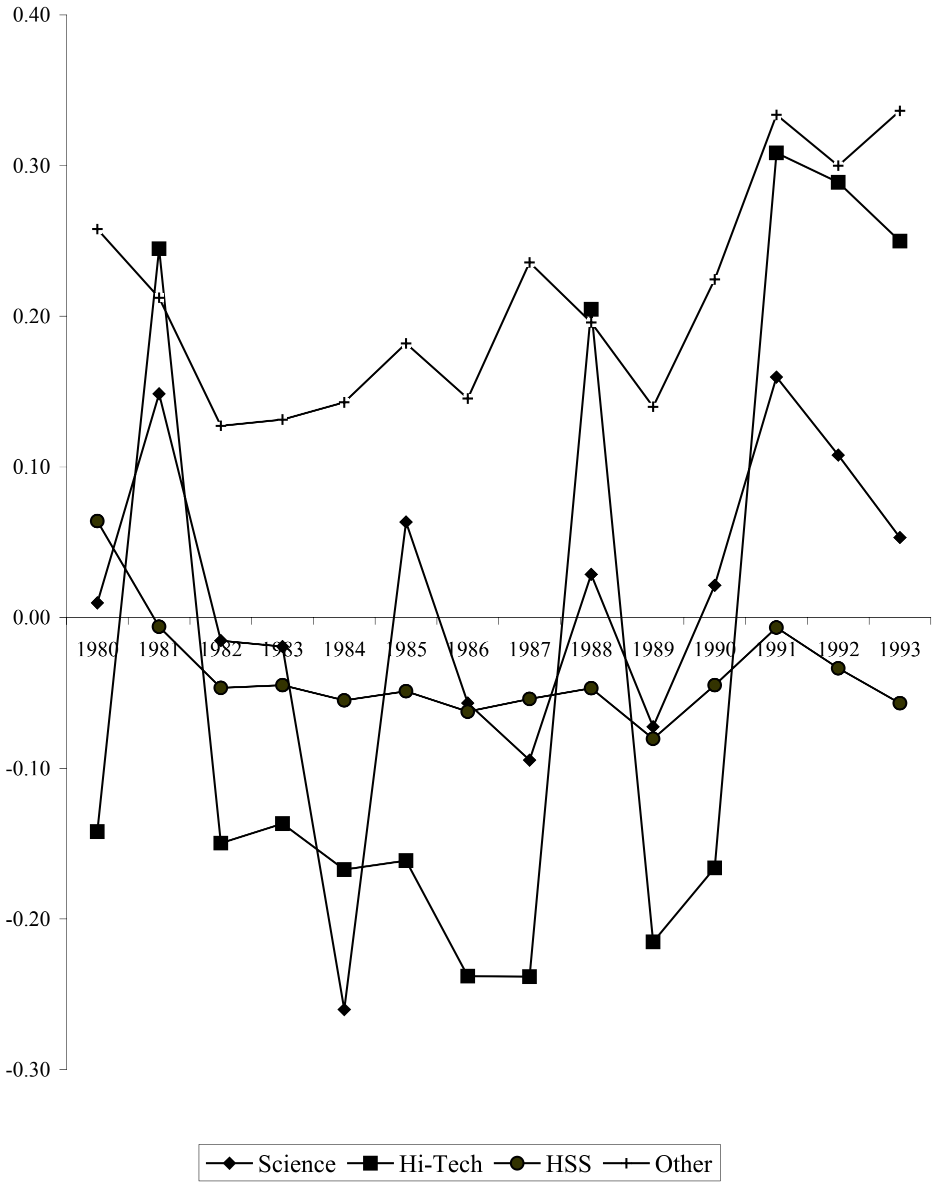


Table 1. Mean, standard deviation and coefficient of variation of male graduates' nominal gross weekly earnings by degree subject and year

Year		MALES					
		Degree subjects					TOTAL
		Science	Hi-tech	Eco-Bus	HSS	Other	
1980	mean	166.82	165.82	173.89	169.70	175.58	170.19
	s.d.	24.90	14.48	24.30	30.14	27.07	25.26
	c.of v.	0.15	0.09	0.14	0.18	0.15	0.15
	N	5030	3994	2703	4291	4899	20917
1981	mean	183.05	184.60	194.39	184.04	187.70	186.11
	s.d.	28.58	17.56	28.68	34.49	28.32	28.34
	c.of v.	0.16	0.10	0.15	0.19	0.15	0.15
	N	4655	4040	2647	4215	4558	20115
1982	mean	200.08	203.90	214.19	203.63	210.37	205.87
	s.d.	34.35	19.8	35.01	40.93	34.85	33.85
	c.of v.	0.17	0.10	0.16	0.20	0.17	0.16
	N	4353	4221	2653	3913	4479	19619
1983	mean	215.99	222.11	235.49	217.64	229.01	223.18
	s.d.	41.16	25.11	40.10	47.89	40.95	39.94
	c.of v.	0.19	0.11	0.17	0.22	0.18	0.18
	N	4856	4800	2885	3990	4556	21087
1984	mean	232.22	235.04	255.59	235.68	242.38	238.62
	s.d.	41.52	24.31	45.48	52.32	40.62	41.48
	c.of v.	0.18	0.10	0.18	0.22	0.17	0.17
	N	5076	5202	2753	3981	4279	21291
1985	mean	256.69	257.66	278.12	256.86	265.47	261.46
	s.d.	44.93	26.61	51.36	59.61	45.76	46.28
	c.of v.	0.18	0.10	0.18	0.23	0.17	0.18
	N	5200	4973	2806	3859	3684	20522
1986	mean	280.00	280.49	314.17	283.32	288.73	286.93
	s.d.	59.82	36.79	67.53	71.65	53.18	58.90
	c.of v.	0.21	0.13	0.21	0.25	0.18	0.21
	N	4966	4794	2677	3926	3640	20003
1987	mean	308.82	308.05	349.68	315.54	326.26	318.42
	s.d.	67.20	44.00	79.90	83.44	61.21	68.38
	c.of v.	0.22	0.14	0.23	0.26	0.19	0.21
	N	5258	4907	2701	4002	3631	20499
1988	mean	337.69	339.88	375.36	338.34	355.26	346.44
	s.d.	72.77	45.30	78.56	85.82	71.13	72.00
	c.of v.	0.22	0.13	0.21	0.25	0.20	0.21
	N	5231	4886	2739	4135	3637	20628
1989	mean	369.43	370.64	408.43	368.26	398.96	380.04
	s.d.	79.88	47.39	83.98	89.84	82.14	78.15
	c.of v.	0.22	0.13	0.21	0.24	0.21	0.21
	N	4602	4937	2767	4097	3429	19832
1990	mean	393.51	389.78	426.78	389.69	441.63	405.04
	s.d.	89.80	59.14	87.10	100.54	101.76	89.96
	c.of v.	0.23	0.15	0.20	0.26	0.23	0.22
	N	4305	4604	2790	3706	3176	18581
1991	mean	409.90	408.48	440.25	403.93	472.92	424.23
	s.d.	92.59	60.99	95.71	110.46	112.19	98.40
	c.of v.	0.23	0.15	0.22	0.27	0.24	0.23
	N	3907	3908	2662	3608	3043	17128
1992	mean	414.80	420.72	451.76	412.77	474.90	432.18
	s.d.	100.38	72.04	106.29	115.80	117.87	105.75
	c.of v.	0.24	0.17	0.24	0.28	0.25	0.24
	N	4023	3898	2777	3939	3199	17836
1993	mean	426.54	435.29	480.21	431.46	486.07	447.95
	s.d.	111.67	84.48	122.92	124.11	118.15	115.14
	c.of v.	0.26	0.19	0.26	0.29	0.24	0.26
	N	4302	4017	2818	4277	3216	18630

Table 2. Index of subjects' specialisation by occupation and sector: males

Subject	Occupation			Sector		
	modal occupation	%	relative rank	modal sector	%	relative rank
Science	ACC	10.2	4	EC	15.1	4
Hi-Tech	EC	26.3	2	EC	50.6	1
Eco-Bus	ACC	40.8	1	ACC	37.4	2
HSS	Teaching	12.5	3	PHE	21.3	3

Note. ACC: accounting; EC: engineering and construction; PHE: public administration, health and education

Table 3. Relative male graduate earnings premia

Year	I. OLS estimates				II. PSM-ATT estimates				III. MNL-OLS estimates			
	Degree subjects				*Degree subjects*				*Degree subjects*			
	Science	Hi-Tech	HSS	Other	Science	Hi-Tech	HSS	Other	Science	Hi-Tech	HSS	Other
1980 Coeff.	-0.024 **	-0.031 **	-0.034 **	0.015 **	-0.048 **	-0.044 **	-0.036 **	0.017 **	0.010	-0.142 **	0.064 **	0.258 **
s.e.	-0.004	0.004	0.004	0.004	0.009	0.007	0.009	0.007	0.019	0.013	0.016	0.031
1981 Coeff.	-0.027 **	-0.021 **	-0.063 **	-0.019 **	-0.054 **	-0.047 **	-0.050 **	-0.022 **	0.149 **	0.245 **	-0.006	0.212 **
s.e.	0.005	0.004	0.005	0.004	0.011	0.009	0.011	0.008	0.040	0.025	0.020	0.052
1982 Coeff.	-0.036 **	-0.014 **	-0.061 **	-0.001	-0.071 **	-0.040 **	-0.028 **	-0.007	-0.015	-0.150 **	-0.047 **	0.127 **
s.e.	0.005	0.005	0.006	0.005	0.010	0.011	0.011	0.009	0.025	0.014	0.014	0.033
1983 Coeff.	-0.057 **	-0.026 **	-0.085 **	-0.008	-0.081 **	-0.063 **	-0.063 **	-0.022 **	-0.019	-0.137 **	-0.045 **	0.131 **
s.e.	0.005	0.005	0.006	0.005	0.010	0.012	0.012	0.009	0.020	0.015	0.016	0.023
1984 Coeff.	-0.061 **	-0.045 **	-0.089 **	-0.028 **	-0.089 **	-0.071 **	-0.099 **	-0.057 **	-0.026	-0.167 **	-0.054 **	0.143 **
s.e.	0.005	0.005	0.006	0.005	0.012	0.011	0.013	0.009	0.027	0.012	0.012	0.025
1985 Coeff.	-0.045 **	-0.039 **	-0.094 **	-0.027 **	-0.081 **	-0.070 **	-0.085 **	-0.049 **	0.063 **	-0.161 **	-0.049 **	0.182 **
s.e.	0.005	0.005	0.006	0.005	0.011	0.014	0.013	0.010	0.023	0.012	0.013	0.028
1986 Coeff.	-0.080 **	-0.075 **	-0.098 **	-0.053 **	-0.125 **	-0.105 **	-0.102 **	-0.086 **	-0.057 **	-0.238 **	-0.062 **	0.145 **
s.e.	0.006	0.006	0.007	0.006	0.012	0.015	0.017	0.011	0.021	0.013	0.013	0.021
1987 Coeff.	-0.083 **	-0.080 **	-0.097 **	-0.034 **	-0.102 **	-0.130 **	-0.100 **	-0.047 **	-0.095 **	-0.238 **	-0.054 **	0.236 **
s.e.	0.006	0.006	0.007	0.006	0.013	0.014	0.017	0.012	0.035	0.014	0.013	0.019
1988 Coeff.	-0.069 **	-0.061 **	-0.109 **	-0.032 **	-0.092 **	-0.090 **	-0.091 **	-0.052 **	0.029	0.205 **	-0.047 **	0.196 **
s.e.	0.006	0.006	0.007	0.006	0.014	0.013	0.016	0.012	0.024	0.013	0.013	0.030
1989 Coeff.	-0.058 **	-0.050 **	-0.100 **	0.000	-0.085 **	-0.071 **	-0.083 **	-0.017	-0.072 **	-0.215 **	-0.080 **	0.140 **
s.e.	0.006	0.005	0.007	0.006	0.013	0.014	0.016	0.011	0.018	0.014	0.014	0.027
1990 Coeff.	-0.040 **	-0.041 **	-0.100 **	0.050 **	-0.087 **	-0.067 **	-0.088 **	0.039 **	0.021	-0.166 **	-0.045 **	0.224 **
s.e.	0.006	0.006	0.007	0.006	0.014	0.016	0.015	0.011	0.027	0.019	0.019	0.030
1991 Coeff.	-0.031 **	-0.023 **	-0.098 **	0.082 **	-0.080 **	-0.066 **	-0.091	0.057 **	0.160 **	0.308 **	-0.007	0.334 **
s.e.	0.007	0.007	0.008	0.007	0.014	0.016	0.017	0.013	0.039	0.017	0.015	0.043
1992 Coeff.	-0.050 **	-0.031 **	-0.098 **	0.059 **	-0.088 **	-0.072 **	-0.079 **	0.045 **	0.108 **	0.289 **	-0.034 **	0.300 **
s.e.	0.007	0.007	0.008	0.007	0.019	0.016	0.018	0.014	0.039	0.020	0.016	0.044
1993 Coeff.	-0.074 **	-0.045 **	-0.106 **	0.035 **	-0.142 **	-0.079 **	-0.083 **	0.025 **	0.053 **	0.250 **	-0.057 **	0.336 **
s.e.	0.008	0.008	0.008	0.008	0.018	0.019	0.020	0.014	0.031	0.022	0.015	0.042

Note. ** statistically significant at the 5% level

Table 4. Choice of identifying variables

Year	"Identifying variables" (ID.Vs)	Earnings regression (OLS) LR-test (a)	p-value	Subject choice (MNL) LR-test (b)	p-value
1980	altechno, alart, mature	$Chi^2(3)=1.27$	0.74	$Chi^2(12)=531.72$	0.00
1981	altechno, aloth	$Chi^2(2)=1.62$	0.45	$Chi^2(8)=495.15$	0.00
1982	altechno, alart, aloth	$Chi^2(3)=1.48$	0.69	$Chi^2(12)=551.61$	0.00
1983	altechno, aloth	$Chi^2(2)=0.74$	0.69	$Chi^2(8)=585.65$	0.00
1984	mature	t=-1.22	0.22	$Chi^2(4)=48.14$	0.00
1985	altechno, mature	$Chi^2(2)=0.61$	0.74	$Chi^2(8)=422.46$	0.00
1986	altechno, alart, aloth	$Chi^2(3)=1.61$	0.66	$Chi^2(12)=795.00$	0.00
1987	altechno	t=0.35	0.72	$Chi^2(4)=433.37$	0.00
1988	aloth, mature	$Chi^2(2)=0.95$	0.62	$Chi^2(8)=334.92$	0.00
1989	altechno	t=0.25	0.80	$Chi^2(4)=646.07$	0.00
1990	altechno, aloth	$Chi^2(2)=1.15$	0.56	$Chi^2(8)=758.20$	0.00
1991	altechno, aloth, mature	$Chi^2(3)=0.94$	0.81	$Chi^2(12)=914.27$	0.00
1992	altechno, aloth, mature	$Chi^2(3)=5.11$	0.16	$Chi^2(12)=1172.33$	0.00
1993	altechno	t=1.30	0.19	$Chi^2(4)=743.98$	0.00

(a) Likelihood ratio test for the exclusion of the ID.Vs from the earnings regressions
(b) Likelihood ratio test for the exclusion of the ID.Vs from the MNL model of subject choice

Table 5. Estimated coefficients of correlation (ϱ) from the MNL-OLS model

Year	*Degree subjects* Science	Hi-Tech	Eco-Bus	HSS	Other
1980	0.2574 **	0.8654 **	0.3682 **	-0.0067	-0.6910 **
1981	-0.4639 **	-0.8972 **	0.2858 **	0.0567 **	-0.6937 **
1982	-0.0932	0.7670 **	0.0507	0.0206	-0.5136 **
1983	-0.0634	0.7407 **	0.1336 **	-0.0119	-0.4399 **
1984	-0.0612	0.8138 **	0.1376 **	0.0312	-0.5694 **
1985	-0.4065 **	0.8116 **	0.1393 **	-0.0015	-0.6394 **
1986	-0.0623	0.8318 **	0.0873 **	-0.0220	-0.6038 **
1987	0.1278	0.7962 **	0.1065 **	0.0000	-0.7475 **
1988	-0.0636	-0.8207 **	0.2263 **	0.0300	-0.4897 **
1989	0.0505	0.8072 **	0.0429	-0.0007	-0.4313 **
1990	-0.0616	0.8005 **	0.1908 **	0.0294	-0.3738 **
1991	-0.2578 **	-0.8478 **	0.3021 **	0.0334	-0.4060 **
1992	-0.2131 **	-0.8153 **	0.2134 **	0.0392	-0.4169 **
1993	-0.1511 **	-0.7480 **	0.1675 **	0.0455	-0.5689 **

Note. ** statistically significant at the 5% level

Appendix A. Variables definition

- Dependent variable:

<u>Subject studied</u>

Science: Biology, Chemistry, Physical Sciences, Mathematics
Hi-Tech: Engineering, Computer Sciences
Eco-Bus: Economics, Business Studies
HSS: Sociology, Politics, Other Social Sciences, Classics, Modern Euro Languages, Humanities
Other: Law, Arts, Allied to Medicine, Education, Combined Subjects, Other.

- Independent variables:

<u>Number of A-level passes by broad subject area</u>

Mathematics
Science: Biology, Chemistry, Physics, Other Sciences, and Statistics
Hi-Tech: Computer Studies, Electronics, Mechanics, and Engineering
Eco-Bus: Economics, Business
HSS: English, French, German, Italian, Spanish, Other Languages, Law, Politics, Classics, Geography, and History
Other: All remaining A-levels.

<u>Social class</u>

SC I (parent has a professional occupation)
SC II (intermediate, including managerial and technical occupations)
SC IIINM (skilled non-manual)
SC IIIM (skilled manual)
SC IV-V (partly skilled & unskilled)
SC OTH (other, including individuals whose parental occupation is either inadequately described or unknown).

<u>A-level point score</u>

The 'point score' is an aggregated measure of an individual's A-level grades, calculated as follows: A=10, B=8, C=6, D=4, E=2 giving a possible maximum score of 30 points in the best three GCE A-level passes, and A=3, B=2, C=1 giving a possible maximum score of 15 points in the best five SCE Higher passes. The conversion Higher-to-A-level point score to create unique bands

of grades is: A (14+→28+), B (12-13→23-27), C (10-11→19-23), D (9-→18). We consider the following groups:
SCOREA (28 pts or higher)
SCOREB (23-27 pts)
SCOREC (19-23 pts)
SCORED (18 pts or less)

A-level curriculum

The A-level (Higher) subjects defining the subject areas are: *Science* (Biology, Chemistry, Physics, Other Sciences, Design, Electronics, Mechanics, Computers, Mathematics, Statistics); *Social Sciences* (Economics, Politics, Law); *Humanities* (Classics, English, Geography, History, French, German, Italian, Other Languages); *Other* (Art, Business, General Studies, Other). We consider the following groups:
CURR0 (no A-level qualifications)
CURR1 (one subject area of specialisation)
CURR2 (two subject areas)
CURR3 (three subject areas).
ALGENS (=1 if student took A-level General Studies, 0 otherwise),
ALMATH (=1 if student took A-level Mathematics, 0 otherwise)
ALCOUNT (total number of A-level passes), HCOUNT (total number of Higher passes).

Residence prior to entry university

RESID0-RESID10 (Non-resident UK nationals, Scotland and Northern Ireland, Wales, London, North-West, North & Yorkshire, West Midlands, East Midlands, South-West, South-East, East Anglia)

Marital status

MARRIED ('1' if student is married, '0' otherwise)

Age

MATURE ('1' is student is aged 21 or older at the date of enrolment, '0' otherwise)

Appendix B. The MNL-OLS model

Here we report only the expression for the log-likelihood function of our polychotomous choice model. In particular we follow Lee (1983) and estimate a simultaneous MNL-OLS model. The interested reader can find the details in Lee (1983, p. 511).
The log likelihood function for the MNL-OLS model has the following form:

$$\begin{aligned}\ln L = \; & \sum_{i=1}^{N}\sum_{j=1}^{J}\{\, S_{ij}\ln\Phi((J_j(Z_i\delta_j)+(\rho_j/\sigma_\varepsilon)(y_{ij}-\sum_{j=1}^{J}S_{ij}\theta_j - X_i\beta))/(1-\rho_j{}^2)^{1/2} + \\ & + S_{ij}\ln\phi((y_{ij}-\sum_{j=1}^{J}S_{ij}\theta_j - X_i\beta)/\sigma_\varepsilon) - S_{ij}\ln\sigma_\varepsilon\}\end{aligned} \tag{6}$$

where N is the number of individuals. Following Lee (1983) we assume

$$\begin{bmatrix} J_1(u_{i1}) \\ J_2(u_{i2}) \\ J_3(u_{i3}) \\ J_4(u_{i4}) \\ J_5(u_{i5}) \\ J_6(\varepsilon_i) \end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \rho_1 \\ 0 & 1 & 0 & 0 & 0 & \rho_2 \\ 0 & 0 & 1 & 0 & 0 & \rho_3 \\ 0 & 0 & 0 & 1 & 0 & \rho_4 \\ 0 & 0 & 0 & 0 & 1 & \rho_5 \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 & 1 \end{bmatrix}\right)$$

where $J_j(u_{ij})$ for j=1,…,5 is the transformation of the non-normal stochastic term u_{ij} into standard normal, while $J_6(\varepsilon_i)$ is the transformation of $\varepsilon_i \sim N(0,\sigma^2_\varepsilon)$ into standard normal and

$$J_j(Z_i\delta_j) = \Phi^{-1}\Lambda_{ij}(Z_i\delta_j)$$

$$\Lambda_{ij}(Z_i\delta_j) = \exp(Z_i\delta_j)/(1+\sum_{j=1}^{J}\exp(Z_i\delta_j),$$

$\phi(.)$ and $\Phi(.)$ are the standard normal density and distribution functions. The main difference with respect to the original Lee's (1983) specification is that we do not estimate separate earnings regressions by subject but only one earnings regression in which some endogenous dummies for degree subjects appear among the regressors. We preferred this specification given the specific focus of our paper: to assess the differences of the earnings *premia* across degree subjects in a standard Mincerian specification of the earnings equation augmented with controls for the degree subject, followed in the literature reviewed in section 2, allowing for the endogeneity of the subject dummies.

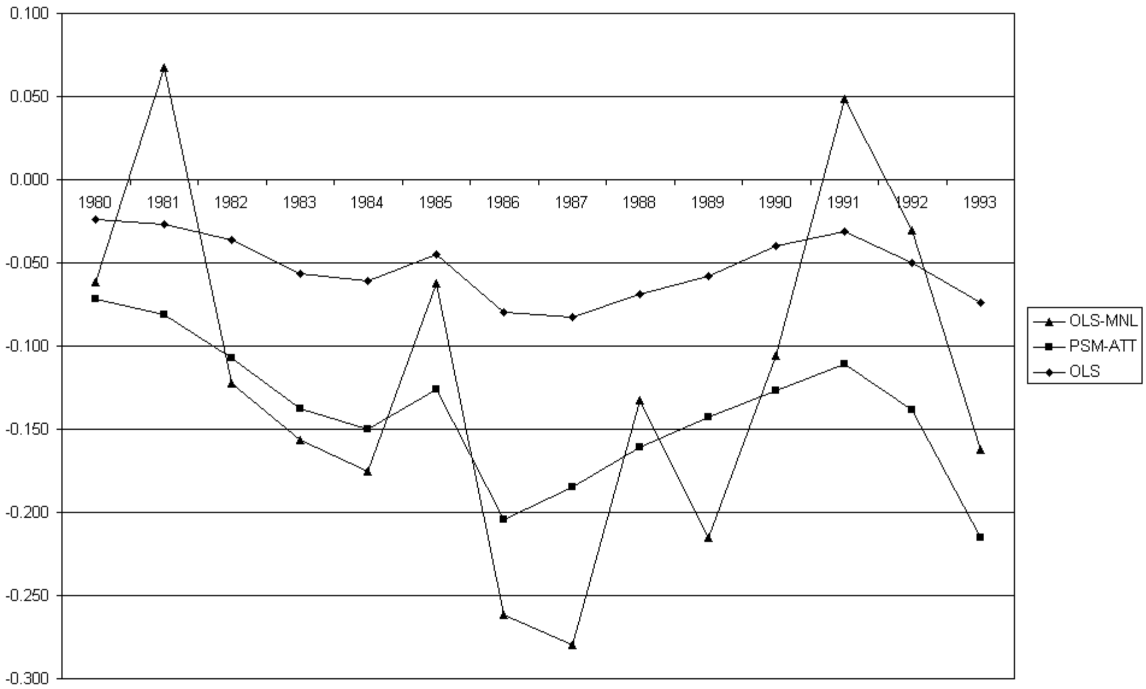
Appendix C. Robustness check to different model identification strategies

year	Science		Hi-Tech		HSS		Other	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1980	0.009	0.010	-0.139	-0.142	0.073	0.064	0.267	0.258
1981	0.138	0.149	0.243	0.245	-0.011	-0.006	0.199	0.212
1982	-0.015	-0.015	-0.157	-0.150	-0.046	-0.047	0.131	0.127
1983	-0.019	-0.019	-0.145	-0.137	-0.043	-0.045	0.136	0.131
1984	-0.024	-0.022	-0.172	-0.172	-0.052	-0.053	0.146	0.146
1985	0.064	0.063	-0.171	-0.161	-0.050	-0.049	0.181	0.182
1986	-0.052	-0.057	-0.244	-0.238	-0.059	-0.062	0.152	0.145
1987	-0.083	-0.095	-0.248	-0.238	-0.055	-0.054	0.232	0.236
1988	-0.015	0.029	-0.216	0.205	-0.064	-0.047	0.167	0.196
1989	-0.071	-0.072	-0.230	-0.215	-0.081	-0.080	0.138	0.140
1990	0.025	0.021	-0.184	-0.166	-0.046	-0.045	0.223	0.224
1991	0.153	0.160	0.327	0.308	-0.008	-0.007	0.335	0.334
1992	0.099	0.108	0.316	0.289	-0.037	-0.034	0.299	0.300
1993	0.052	0.053	0.287	0.250	-0.058	-0.057	0.350	0.336

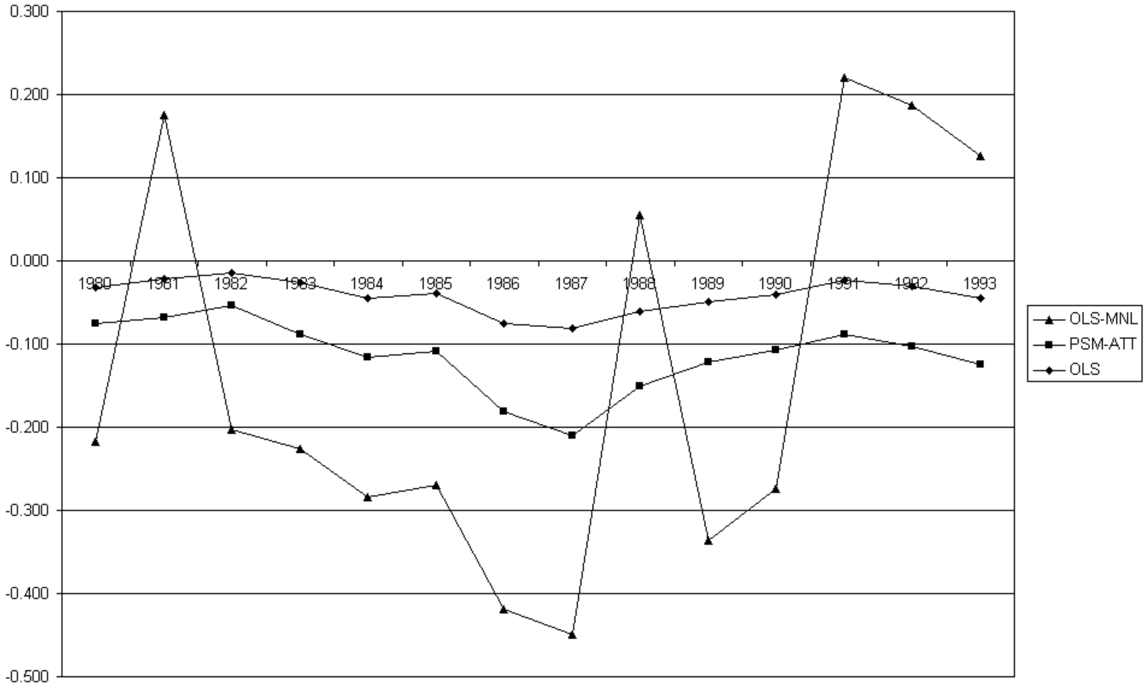
(a) MNL-OLS model identification is through functional forms only.
(b) MNL-OLS model identification is through functional forms and exclusion restrictions.

Appendix D. Graphical comparisons between relative earnings *premia* estimated by OLS, PSM-ATT, and MNL-OLS

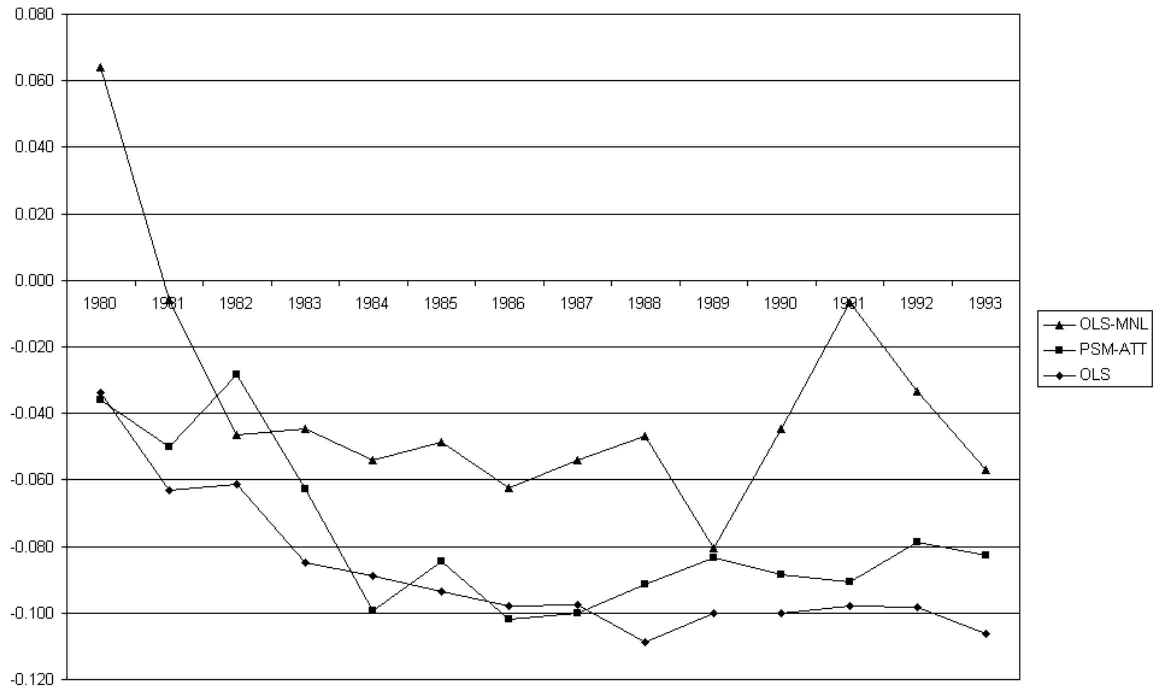
a. Science



b. Hi-tech


0.300
0.200
0.100
0.000
-0.100
-0.200
-0.300
-0.400
-0.500
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
OLS-MNL
PSM-ATT
OLS

c. HSS


0.080
0.060
0.040
0.020
0.000
-0.020
-0.040
-0.060
-0.080
-0.100
-0.120
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
OLS-MNL
PSM-ATT
OLS

Appendix E. Likelihood ratio tests for the equality of subject *premia* over time

year	Chi2(4)	p-value
1981	911.08	0.00
1982	584.7	0.00
1983	518.28	0.00
1984	817.79	0.00
1985	863.38	0.00
1986	1044.41	0.00
1987	919.57	0.00
1988	751.47	0.00
1989	778.84	0.00
1990	739.67	0.00
1991	833.62	0.00
1992	443.46	0.00
1993	322.79	0.00

Note: The constraints are imposed only on the coefficients of the subject dummies θ_j's, that is:

H_0: $\theta_j(t) = \theta_j(t-1)$

H_1: $\theta_j(t) \neq \theta_j(t-1)$ for j=1, 2, 3, 4

where $\theta_j(t-1)$ is taken as given. No inter-cohort equality restrictions are imposed on the coefficients of the other explanatory variables.